TECHNISCHE UNIVERSITÄT MÜNCHEN

TUM School of Life Sciences

Analysis of Peptide Glycation Reactions using High-Resolution Mass Spectrometry

Michelle Tamara Berger

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Corinna Dawid

Prüfer der Dissertation:

1. apl. Prof. Dr. Philippe Schmitt-Kopplin
2. Prof. Dr. Mathias Wilhelm

Die Dissertation wurde am 10.05.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 25.09.2023 angenommen.

This page intentionally left blank.

Acknowledgements

It takes years to finish a PhD project. This thesis marks the endpoint of my journey as a doctoral student, and I realized how quickly time has passed. Time went by so fast not only because of the many fascinating projects but also because I was surrounded by a fantastic team. Completing this PhD would not have been possible without the support of many wonderful people. I would like to express my gratitude to all of you for the help that I have received during the past years.

First, I would like to thank Prof. Dr. Philippe Schmitt-Kopplin for being an extraordinary supervisor. Thank you for always *having an open door*, while letting me discover new fields of interest and realize my own ideas. I also thank you for giving me the opportunity to work on such a versatile topic. These years are very valuable to me, both personally and professionally.

I am very grateful to Prof. Dr. Michael Rychlik and Dr. James W. Marshall for the fruitful discussions, and the help during the time-consuming yet exciting publishing processes. Thanks to Prof. Dr. Fabiana Perocchi and Dr. Kenneth Dyar for the excellent collaboration. I really enjoyed our meetings leading to a spontaneous flood of great ideas and finally amazing results. A special thanks to Dr. Daniel Hemmler for supporting me *from start to finish line*. Thank you for the training, countless scientific discussions, and your valuable advice at all times.

Thank you to all my colleagues at BGC. I am incredibly thankful for this outstanding team, which made me feel welcome from day one. You made my PhD thesis a truly enjoyable and unforgettable time. I will do my best to pass on the atmosphere of this positive working environment to wherever my professional future may take me. Many thanks to Astrid Bösl and Anja Brinckmann for their support in the many administrative matters.

A special thank you goes to Dr. Alesia Walker sharing an office with me for more than four years. Even though troubleshooting and debugging sometimes wasn't easy, I could always count on you. You were one of my biggest supporters during this thesis and I want you to know that I will always appre*cat*e that. A very big thank you to Dr. Marco Matzka, who brightened many days of my thesis. You were not only my colleague, but also my neighbor and friend. This PhD thesis is no longer *unwritten*. Many thanks to Liesa Salzer for the inspiring discussions and great time during business trips. I am very especially grateful for the *boat*, which gave me comfort when things did not go as planned. Thank you to Dr. Jasmine Hertzog, Jenny Uhl, and Dr.

Marianna Lucio. One of the greatest aspects of PhD is meeting people like you. You are not only incredible scientists, but also amazing friends. A heartfelt thanks to Yingfei Yan, my glycation partner and dear friend. Thank you for all the discussions on model systems, but more importantly great moments we have shared. I will remember all the times we finally *got it*.

I would like to thank my family and friends. Thank you to my mom, I felt your support every day. Many thanks to my aunt Renate Simseker for always having an open ear. A big thank you to Sonja Unterholzner, Ameliè Collinèt, and Nathalie Hock. I am sure that your endless interest in my work made you mass spectrometry experts along the way. Thank you to Dr. Meshal Ansari for a friendship lasting from Abitur to PhD – and beyond. I am very thankful for your valuable input, and I am happy that I could share this experience with you.

Finally, I want to acknowledge Prof. Dr. Mathias Wilhelm and the committee chairwoman Prof. Dr. Corinna Dawid for their commitment to take part in my examination.

This page intentionally left blank

Abstract

Glycation describes non-enzymatic reactions between amino acids, peptides, or proteins and reducing sugars. Unlike most chemical reactions, many diverse reaction products form in glycation. To date, numerous glycation products remain unaccounted. Glycation of peptides is particularly underexplored. The large heterogeneity of peptides impeded to understand, which structures are preferred in glycation. Another constraint is the lack of methods for non-targeted analysis of peptides with known or unknown glycation modifications. Considering that the relevance of glycation is just as broad as the spectrum of possible peptide glycation modifications, we need to better understand what drives glycation of peptide reactants. Not only is glycation important for aroma, color, and taste of foods, but is also associated with a plethora of pathologies.

This cumulative thesis consists of two original publications, which address how the structure of peptides influences the formation of early-stage glycation products and propose a new computational approach for fast annotation of glycated peptides using tandem mass spectra. The first publication presents how multi-peptide model systems enable to find links between peptide reactant structural features and Amadori product formation over time. Instead of a conventional enzymatic digest, tryptone was used as a peptide source offering comparatively more heterogeneous amino acid sequences. The many and diverse peptides in tryptone allowed to detect small sequence alterations, and finally, potential sequence patterns, associated with changes in early-glycation behavior. Trends were summarized as structural checkpoints, which can help to estimate the Amadori product-forming potential of peptides.

The second publication introduces a computational strategy, which uses tandem mass spectral data to empower the concept of non-targeted analysis. Modern mass spectrometers can fragment thousands of precursors during the measurement of a single sample, which makes tandem mass spectra highly suitable for high-throughput structure elucidation. The complex fragmentation behavior of modified peptides requires consideration of different ion classes. The presented method categorizes fragments of peptide glycation products into three different ion types and relies on both the structure of the peptide backbone and the glycation modification substructure. Compared to solely using theoretical peptide fragmentation, the triple-ion strategy could cover significantly more fragment ions. The algorithm unveiled neutral losses specific for the bound glycation product, which was demonstrated for the Amadori product of H-Ile-Leu-OH

and two previously undescribed modification delta masses. Not only pre-selected but all fragmented peptide glycation products could be analyzed from (iso)leucine dipeptide glucose model systems independent of whether the glycation modification was known or not.

Zusammenfassung

Glykierung umfasst nicht-enzymatische Reaktionen zwischen Aminosäuren, Peptiden oder Proteinen und reduzierenden Zuckern. Im Gegensatz zu den meisten chemischen Reaktionen bildet sich bei der Glykierung eine große Anzahl vielfältiger Reaktionsprodukte. Die Strukturen zahlreicher Glykierungsprodukte bleiben bis heute unaufgeklärt. Glykierungsreaktionen von Peptiden wurden besonders wenig erforscht. Die große Heterogenität von Peptiden erschwerte es zu verstehen, welche Strukturen bevorzugt Glykierungsreaktionen eingehen. Weiterhin mangelt es an Methoden zur nicht zielgerichteten Analyse von Peptiden, welche bekannte oder unbekannte Glykierungsmodifikationen tragen. In Anbetracht der Tatsache, dass die Relevanz von Glykierung genauso weitreichend ist wie das Spektrum möglicher Peptidglykierungsmodifikationen, ist es notwendig, besser zu verstehen, welche Faktoren die Glykierung von Peptidreaktanten beeinflussen. Glykierung ist nicht nur von essenzieller Bedeutung für das Aroma, die Farbe und den Geschmack von Lebensmitteln, sondern spielt auch bei einer Vielzahl von Krankheiten eine wichtige Rolle.

Diese kumulative Dissertation besteht aus zwei Veröffentlichungen, die sich damit befassen, wie die Struktur von Peptiden die Bildung von Glykierungsprodukten der frühen Reaktionsphase beeinflusst, und einen Algorithmus für die schnelle Annotation glykierter Peptide unter Verwendung von Tandem-Massenspektren vorschlagen. Die erste Veröffentlichung beschreibt, wie Multi-Peptid-Modellsysteme dazu beitragen können, Zusammenhänge zwischen den strukturellen Eigenschaften von Peptidreaktanten und der zeitabhängigen Bildung der jeweiligen Amadori-Produkte zu finden. Anstelle eines herkömmlichen enzymatischen Verdaus wurde Trypton als Peptidquelle verwendet, welches Peptide mit vergleichsweise heterogeneren Aminosäuresequenzen enthält. Die vielen und unterschiedlichen Peptide in Trypton ermöglichten es, kleine Sequenzänderungen und schließlich potenzielle Sequenzmuster zu ermitteln, welche mit Änderungen im frühen Glykierungsverhalten einhergingen. Besagte Trends wurden als strukturelle Checkpoints zusammengefasst, die helfen können, das Potenzial von Peptiden abzuschätzen, Amadori Produkte zu bilden.

Die zweite Veröffentlichung stellt eine computerbasierte Strategie vor, welche Tandem-Massenspektren nutzt, um das Konzept der nicht zielgerichteten Analyse zu stärken. Moderne

Massenspektrometer können Tausende von Analyten während der Messung einer einzigen Probe fragmentieren, weshalb sich Tandem-Massenspektren für die Hochdurchsatz-Strukturaufklärung besonders eignen. Das komplexe Fragmentierungsverhalten modifizierter Peptide erfordert die Berücksichtigung unterschiedlicher Ionenklassen. Der beschriebene Algorithmus kategorisiert Fragmente von Peptidglykierungsprodukten in drei verschiedene Ionentypen und stützt sich neben der Struktur des Peptidrückgrats auch auf die Glykierungsmodifikationssubstruktur. Im Vergleich zur ausschließlichen Verwendung theoretischer Peptidfragmentierung konnte die Tripel-Ionen-Strategie signifikant mehr Fragmentionen abdecken. Der Algorithmus legte neutrale Massenverluste offen, die spezifisch für das gebundene Glykierungsprodukt sind. Dies wurde für das Amadori-Produkt von H-Ile-Leu-OH und zwei zuvor unbeschriebene Modifikations-Delta-Massen demonstriert. In (Iso)leucin Dipeptid-Glucose-Modellsystemen konnten nicht nur zuvor ausgewählte, sondern alle fragmentierten Peptidglykierungsprodukte analysiert werden, unabhängig davon, ob die Glykierungsmodifikation bekannt war oder nicht.

Table of Contents

Scientific Communications

Publications

Publications directly addressed in this thesis:

Chapter 2: Berger, M. T., Hemmler, D., Walker, A., Rychlik, M., Marshall, J. W., and Schmitt-Kopplin, P. Molecular Characterization of Sequence-Driven Peptide Glycation. *Scientific reports*, *11*(1), 13294 (2021).

Chapter 3: Berger, M. T., Hemmler, D., Diederich, P., Rychlik, M., Marshall, J. W., and Schmitt-Kopplin, P. Open Search of Peptide Glycation Products from Tandem Mass Spectra. *Analytical chemistry*, *94*(15), 5953–5961 (2022).

Work in progress

Prudente De Mello, N.[#], Berger, M.T.[#], Lehmann, K., Yan, Y., Weidner, L., Tokarz, J., Möller, G., Cheng, Y., Keipert, S., Ciciliot, S., Artati, A., Wettmarshausen, J., Brandt, D., Kutschke, M., Anja, Leimpek, Vogt, Weisenhorn D., Wurst, W., Adamski, J., Jain, M., Jastroch, M., Mandemakers, W., Bonifati, V., Schmitt-Kopplin, P., Perocchi, F., and Dyar, K.A. N-α-Glycerinyl-AGEs are Biomarkers for Human and Mouse PARK7 (DJ-1) Loss-of-Function. *Manuscript in preparation.* [#]These authors contributed equally.

Berger, M. T., Walker, A., Hemmler, D., Diederich, P., Rychlik, M., Marshall, J. W., and Schmitt-Kopplin, P. Characteristic Neutral Losses for Spectral Alignment of Peptide Modifications. *Manuscript in preparation.*

Oral presentations

metaRbolomics Hackathon, 2021, Lutherstatt-Wittenberg, Germany. Title: Non-Targeted Analysis of Peptide Glycation Products from Tandem Mass Spectra.

32. Doktorandenseminar Hohenroda, 2022, Hohenroda, Germany. Title: Open Search of Peptide Glycation Products from Tandem Mass Spectra.

Advanced Mass Spectrometry Seminar, 2022, Freising, Germany. Title: Characteristic Neutral Losses for Open Search of Peptide Glycation Modifications.

Poster presentations

2nd Food Chemistry Conference, 2019, Sevilla, Spain. Title: Non-Targeted Analytical Strategies Give Insights into the Peptide Reactivity in Maillard Model Reactions.

14th IMARS, 2021, Doha, Qatar. Title: Understanding Structural Microheterogeneity Effects on Peptide Glycation.

18th Annual Conference of the Metabolomic Society, Metabolomics 2022, 2022, Valencia, Spain. Title: Characteristic Neutral Losses for Open Search of Peptide Glycation Modifications.

Abbreviations

1-DG	1-Desoxyglucodiulose
3-DG	3-desoxyglucosone
4-DG	1-Amino-1,4-didesoxygluco-2,3-diulose
AGE	Advanced glycation end product
ARP	Amadori rearrangement product
BSA	Bovine serum albumin
CAS	Chemical abstracts service
CFM-ID	Competitive fragmentation modeling for metabolite identification
CID	Collision-induced dissociation
CML	Carboxymethyllysine
DDA	Data-dependent acquisition
DIA	Data-independent acquisition
DNA	Deoxyribonucleic acid
ESI	Electrospray ionization
GIP	Gastric inhibitory polypeptide
HbA1c	Glycated hemoglobin
HEK293	Human embryonic kidney 293
HILIC	Hydrophilic interaction chromatography
HMDB	Human metabolome database
HRP	Heyns rearrangement product
KEGG	Kyoto encyclopedia of genes and genomes
LC	Liquid chromatography
m/*z*	Mass-to-charge ratio
MR	Maillard reaction
MRM	Multiple-reaction monitoring
MS	Mass spectrometry
MS/MS or MS2	Tandem mass spectrometry
NAD^+	Nicotinamide adenine dinucleotide
NIST	National institute of standards and technology

PTM	Post-translational modification
Q1	First quadrupole
Q2	Second quadrupole
QToF	Quadrupole time-of-flight
RNAF	Ribonucleic acid
RP	Reversed phase
SWATH	Sequential acquisition of all theoretical masses
UHPLC	(Ultra) high-performance liquid chromatography

Chapter 1

General Introduction and Methods

"The important thing is not to stop questioning. Curiosity has its own reason for existence. One cannot help but be in awe when he contemplates the mysteries of eternity, of life, of the marvelous structure of reality. It is enough if one tries merely to comprehend a little of this mystery each day."

— Albert Einstein

At the beginning of the 20th century, the first successful peptide synthesis by Emil Fischer sparked Louis-Camille Maillard's curiosity.[1] Trying to develop an alternative synthesis route for peptides closer to physiological conditions, the French chemist observed brown discoloration of amino acid-glycerol mixtures upon heating. Indeed, he ascertained peptide formation but, more importantly, the mystery of non-enzymatic glycation began in 1912 with this first experiment.[2] In the last hundred years since its discovery, scientists tried to comprehend the so-called Maillard reaction (MR). A large and diverse collective of scientific publications revealed that glycation reactions are highly relevant in various fields including health-related research and food science. In recent years, substantial improvements in the quality and quantity of data led to an increased number of non-targeted studies. This shift from hypothesis-driven toward data-driven research allowed to demonstrate that glycation is not a single reaction but a reaction network with many intermediates and reaction products. The factors driving glycation of peptides, however, remain largely unknown, even though peptide modifications are also relevant in different "omics" sciences. To date, the reaction complexity and the virtually endless number of possible peptide structures hampered comprehensive analysis of peptide glycation.

Considering that the mystery of glycation could not yet be fully resolved and that peptides are an underexplored dimension of glycation studies, untargeted analysis of well-controlled model systems could help to understand the vast heterogeneity of peptide glycation products, which reflects precursor reactivity and environmental influences.

We will have to keep questioning.

1.1 The three stages of glycation reactions

The term "glycation" was coined in the 1980s. Initially, all reactions that link sugars to peptides or proteins, whether or not catalyzed by an enzyme, were defined as glycation.[3] In 1993, Lis and Sharon have addressed the need to differentiate between enzymatic and non-enzymatic modification of amino compounds introducing the current terminology of glycosylation and glycation.[4] Nowadays, the designation "glycation" usually is exclusively used to describe reactions between amino acids, peptides or proteins and reducing sugars. It should be mentioned that a wide range of other amino compounds and carbonyl precursors can react, analogously.

Already in 1953, Hodge proposed a comprehensive scheme of reaction mechanisms involved in the glycation cascade.[5] Although slight adjustments have since been made, the Hodge scheme remains fundamentally correct in its original form.[6,7] Consistent with Hodge, glycation reactions are still commonly classified in three different stages: the (i) early, (ii) advanced, and (iii) final stage. Even though classification into stages might imply a consecutive reaction process, glycation describes a complex series of both sequential and parallel reactions.[8] Initially, the reaction pool only contains the reaction precursors, however, the reactant "reservoir" rapidly becomes more complex. The nucleophilic amino reactant is modified by the carbonyl compound yet catalyzes carbonyl degradation at the same time. Both the carbonyl starter and its breakdown products may induce modification of not only the amino precursor but also of other reaction intermediates.[9] Hence, glycation may be visualized as an interactive cyclic reaction system, rather than a linear reaction process, where intermediates are continuously fed into the reaction pool and can potentially function as secondary reaction precursors (Figure 1.1).[10]

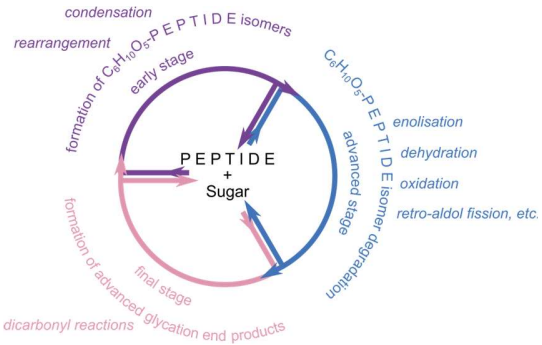
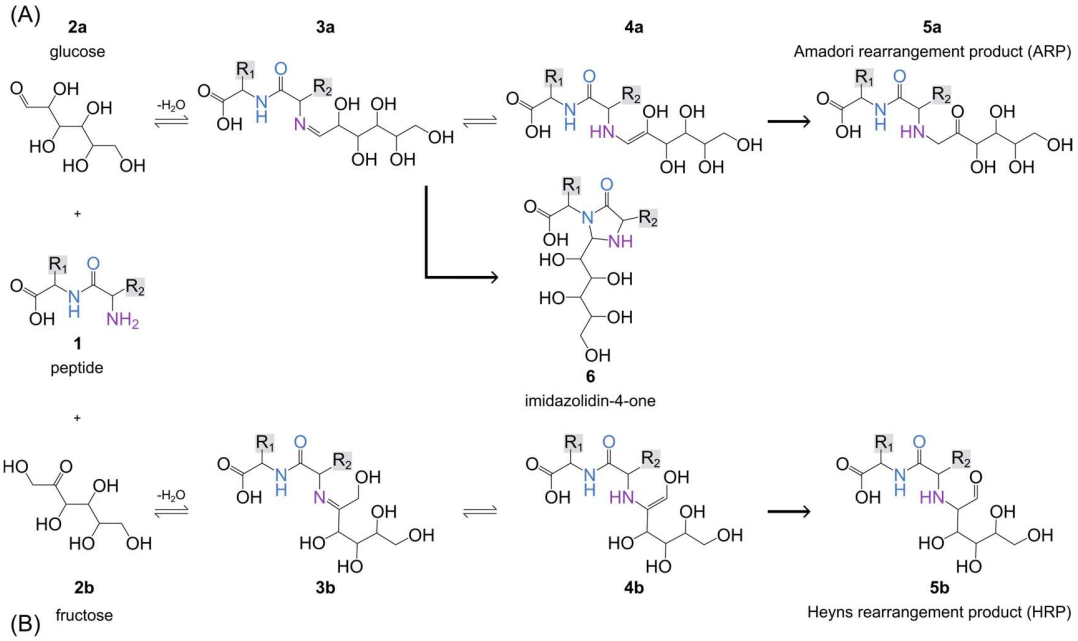


Figure 1.1: Scheme of the three different glycation reaction stages. For each reaction phase, the major pathway types are listed. Reaction products from all stages interact in the common reaction pool and may undergo secondary reactions. Adapted from Hodge, and Hemmler *et al.*[5,10]

In the following sections, I will adhere to the conventional classification into the early, advanced, and final stage, and describe representative products for each reaction phase. It is important to note that the pathways described herein depend on the nature of the amino and/or sugar reactant. Reactant dependency will be discussed for each reaction mechanism, separately. General factors, which need to be considered for the structure of the amino reactant across all reaction stages are described in section 1.1.5. For simplicity, the amino compound is illustrated as a generic dipeptide **1** (Scheme 1.1) throughout reaction schemes. Even though many glycation products have been first discovered on amino acids, peptides most often react analogously. Exceptions are marked and described.

1.1.1. The early reaction stage

Scheme 1.1 summarizes the main proposed mechanisms of early-phase glycation reactions. The glycation cascade is initiated by a condensation reaction between the amino and the carbonyl compound producing a mixture of amino-bound $C_6H_{10}O_5$-isomers.



Scheme 1.1: Early stage of glycation reactions. Amino compounds undergo spontaneous condensation with (A) aldose (e.g., glucose 2a) or (B) ketose (e.g., fructose 2b) sugars. Early glycation products, more specifically isomeric $C_6H_{10}O_5$-modifications, form *via* consecutive rearrangement (Amadori rearrangement product, 5a; Heyns rearrangement product, 5b) or cyclization (imidiazolidin-4-one, 6).

If any amino reactant **1** undergoes rapid condensation with an α-hydroxy aldehyde like glucose **2a** (Scheme 1.1A), an intermediate glycosylamine (Schiff base) **3a** is formed. Concomitant isomerization, also known as the Amadori rearrangement, *via* **4a** results in an 1-amino-1-deoxyketose **5a** (Amadori rearrangement product, ARP).[11] Both the rearrangement step and the end product are named after Mario Amadori, an Italian chemist, who provided first indications on the early phase of glycation and the ascribed $C_6H_{10}O_5$-amino isomer formation pathway in the 1920s.[12] If instead an α-hydroxy ketose (e.g., fructose **2b**) is coupled to the amino group of **1** (Scheme 1.1B), the Schiff base **3b** isomerizes *via* the Heyns rearrangement affording the corresponding 2-amino-2-deoxyaldose **5b** (Heyns rearrangement product, HRP). Kurt Heyns, the eponym for this second type of rearrangement and $C_6H_{10}O_5$-amino isomer species, was a German chemist, who documented the formation of D-glucosamines for the first time in the reaction of fructose with ammonia.[13] Many chemical structures of later-stage glycation products contain nitrogen heterocycles (see section 1.1.3). After more than 60 years since the discovery of the ARP, formation of the early-stage nitrogen heterocyclic structure **6** was reported. At first, imidazolidin-4-one modifications **6** were only documented for peptides.[14-16] In addition to the Amadori rearrangement, the Schiff base **3a** has been shown to undergo intramolecular cyclization through nucleophilic attack by the adjacent peptide bond amide-nitrogen atom.[16] Due to participation of the peptide bond in the reaction mechanism, imidazoldin-4-ones **6** were considered highly specific for peptide glycation. Later, Chu and Yaylayan showed that imidazolidin-4-one modifications form in alanine model systems through reaction of the decarboxylated amino acid and dicarbonyls.[17]

As mentioned above, amino compounds and α-hydroxy aldehydes form ARPs and imidazolidin-4-one moieties, while HRPs are produced from reaction with α-hydroxy ketones. The nature of the sugar reactant does not only determine the type of reaction pathway but also the reaction rate. The formation of early glycation products strongly depends on the sugar's tendency to occur in its reactive acyclic form. A higher proportion of the open-chain form is documented for pentoses compared to hexoses leading to superior reactivity of pentose sugars in early glycation reactions.[18] Glucose is the major free sugar and energy source in the human body.[19] Among pentoses and hexoses, glucose is considered the weakest glycation precursor barely found in its reactive acyclic state (0.002%).[20] Further, ARP formation is pH-dependent and favored in an acidic environment due to proton-catalyzed enamine-formation at low pH values.[21]

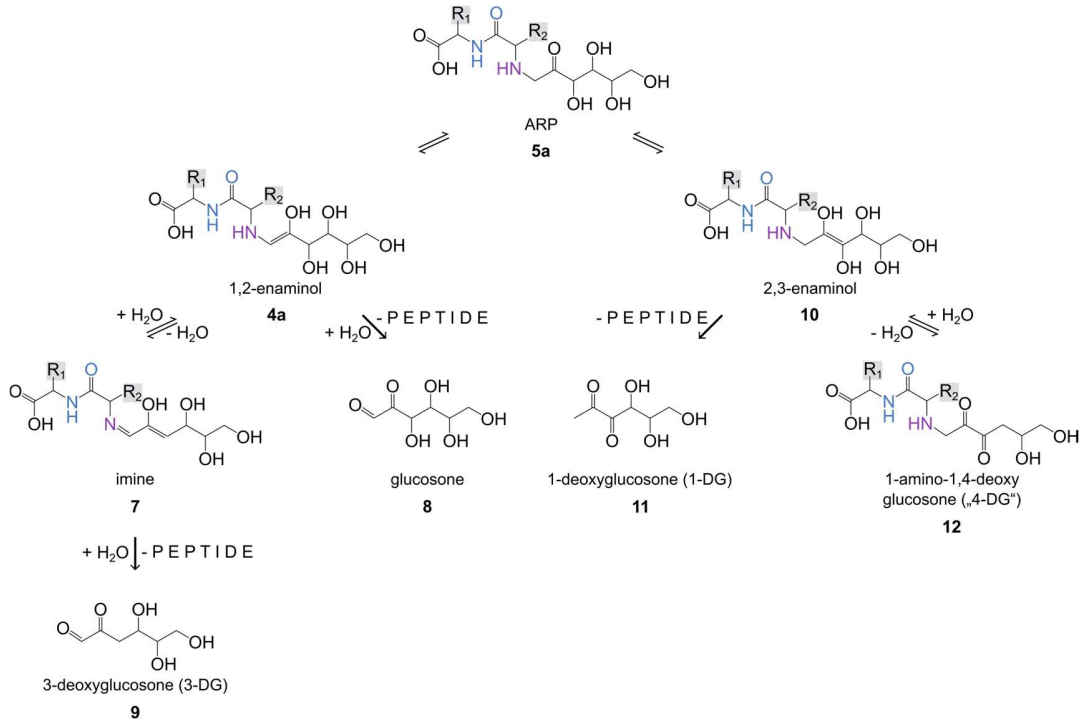
The ARP is the most studied early glycation intermediate. The breakdown of early glycation isomers drives the generation of heterogeneous degradation products, which lead to a plethora of further reactions and production of miscellaneous advanced glycation end products (AGEs). Such downstream reactions are referred to as the advanced and final phase of glycation.

1.1.2 The advanced reaction stage

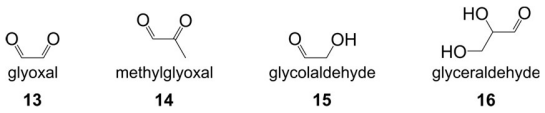
In the advanced, or intermediate, phase of glycation, ARP and Schiff base breakdown products are formed. On the basis of a review by Hodge, thorough investigations into the degradation of early glycation products were performed.[5] Degradation mainly occurs through dehydration, enolization, oxidation and retro-aldol fission reactions (Figure 1.1).

The stability of the ARP strongly depends on various factors including the environmental pH.[7,22,23] Degradation rates increase with increasing pH. The type of degradation pathway is also pH dependent, and determined by the so-called Lobry de Bruyn-Alberda van Ekenstein transformation.[24] While acidic conditions favor nitrogen protonation and consequent 1,2-enolization (Scheme 1.2, left), neutral to basic pH values promote 2,3-enolization (Scheme 1.2, right).[25,26] The 1,2-enaminol **4a** may dehydrate to form the imine **7** or oxidize leading to glucosone **8**. Hydroxylation of **7** produces 3-deoxyglucosone **9** (3-DG). From the 2,3-enaminol **10**, 1-deoxyglucosone **11** (1-DG) and 1-amino-1,4-dideoxyglucosone **12** (4-DG) can be generated, the latter forming *via* dehydration.

Note that the structures of the advanced-stage reaction products **8, 9, 11** and **12** all contain dicarbonyl moieties. Dicarbonyls are highly reactive and drive the course of final-stage glycation reactions. The reactivity of dicarbonyls depends on the position of methylene moieties relative to the dicarbonyl subunit. Adjacent methylene groups, which separate the dicarbonyl substructure from nearby hydroxyl groups, lead to increased stability (e.g., **9**). Neighboring α-hydroxyl groups create a reductone structure (α-oxo-enediol) and increase dicarbonyl reactivity (e.g., **11**).[27,28] Accordingly, studies have shown that the half-life is approximately 40 h for **9** and 0.5 h for **11** under physiological conditions.[29,30] α,β-dicarbonyls may undergo retro-aldol fission to produce short-chain carbonyl species. Dicarbonyls may also be oxidized. Representative products of such dicarbonyl follow-up reactions are shown in Scheme 1.3 including glyoxal **13**, methylglyoxal **14**, glycolaldehyde **15**, and glyceraldehyde **16**.



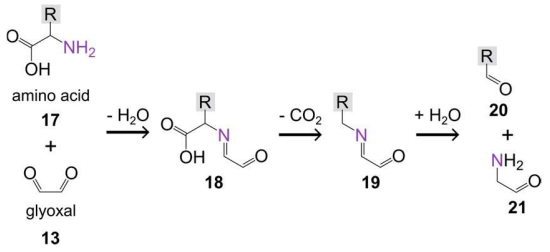
Scheme 1.2: Amadori product degradation in advanced-stage glycation reactions. Low pH values promote the 1,2-enolization pathway (left), and high pH values favor 2,3-enolization (right). Adapted from Hellwig, and Lund and Ray.[9,31]



Scheme 1.3: Representative carbonyl compounds. Shown carbonyls can be formed from higher molecular weight dicarbonyls by fission reactions.

As stated above, glycation does not follow a linear reaction sequence. Different reaction pathways from one reaction stage can lead to the same product class. Dicarbonyls are not only produced from the degradation of early glycation products in the advanced reaction stage. Intermediate-phase sugar degradation reactions are another source of dicarbonyls. In fact, some of the abovementioned dicarbonyl compounds are also formed through alternative sugar breakdown reactions.[32] Note that this means that amines are not essential for the generation of dicarbonyls. However, amines do have a catalytic effect on dicarbonyl formation. In the absence of amino compounds, more extreme temperatures and/or pH values are required to generate dicarbonyls. In the presence of amines, dicarbonyls can be produced under pH conditions normally encountered in foods and living organisms.[25]

In the advanced stage of glycation, not only the sugar but also the amino reactant may be degraded. Already in 1927, Shiro Akabori observed decarboxylation of amino acids in presence of glucose.[33] Later, it was shown that not glucose itself but its degradation products, namely *vic*-dicarbonyls, are involved in the decomposition reaction.[34] Since 1948, this reaction step is known as "Strecker degradation", after the discoverer Adolph Strecker.[34,35] Condensation between an amino acid **17** and a *vic*-dicarbonyl (e.g., **13**) leads to the formation of an α-imino carbonyl **18**. **18** readily undergoes decarboxylation to give **19**. The α-amino carbonyl **20** and the free aldehyde **21** are produced *via* hydrolysis of **19**. **21** is also called "Strecker aldehyde".[36,37] Importantly, peptides cannot undergo Strecker degradation as the carboxylic group of the N-terminal amino acid is part of the adjacent peptide bond. Peptides degrade among the peptide bonds into shorter-chain peptides and amino acids.[38,39] Higher temperatures promote peptide decomposition.[40] Further, the degradation rate strongly depends on the peptide structure.[41,42] Downstream amino acids, however, may be subjected to Strecker degradation. For modified peptides, cleavage of peptide bonds cannot only result in free amino acids. For example, breakage of the N-terminal peptide bond in α-modified peptides may lead to the formation of "free" amino acid glycation products. If lysine is positioned at the N-terminus, structures like N-α-carboxymethyllysine (N-α-CML) may be detected.

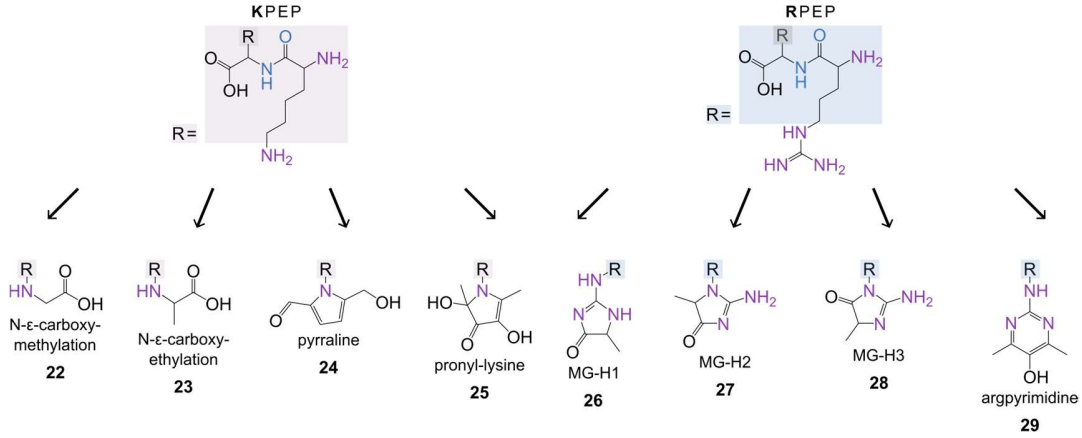


Scheme 1.4: Strecker degradation of amino acids in presence of *vic*-dicarbonyls. Adapted from Yaylayan, and Hellwig and Henle.[8,36]

1.1.3 The final reaction stage

In the final phase of glycation reactions, both high and low molecular weight compounds are formed. Heterogeneous polymers are produced by reaction of low molecular weight glycation products from the early and/or advanced stage. These high molecular weight, brown to black colored polymeric analytes are known as melanoidins.[26] The term "melanoidins" was created by Maillard himself and he postulated a heterocyclic character for their structure. However, the chemical structures and formation mechanisms of melanoidins are still largely unknown. Isolation and identification have thus far been achieved only from model systems.[43,44]

Not all final-stage reactions are polymerization reactions. Moreover, sugar degradation products, mainly abovementioned highly reactive α, β-dicarbonyls, react with N-terminal amino groups and amino acid side chain functional groups. Examples for important α-oxoaldehyde glycating agents are **9**, **13**, and **14**. Besides the amino reactant, such reactive species may also attack thus far formed reaction intermediates. Final-phase carbonyl reactions lead to a diverse set of compounds, often called advanced glycation end products (AGEs). A selection of representative AGEs is shown in Scheme 1.5. The presented AGE modifications are commonly detected and investigated for peptides but were first found for amino acids.[45,46]

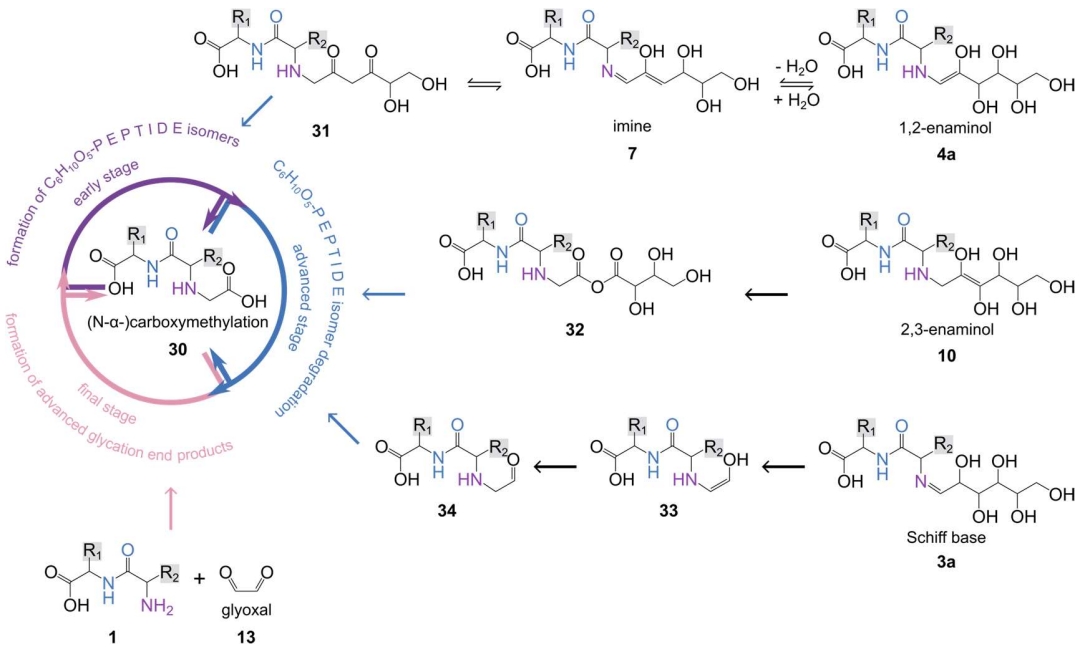


Scheme 1.5: Selection of well-known advanced glycation end products. Structures are divided into lysine side chain modifications (left) and arginine side chain modifications (right). Accordingly, lysine- and arginine-containing dipeptides were chosen as representative peptide core structures.

Recently, inconsistencies in the application of the AGE nomenclature have arisen. In contrary to the terminology, AGEs can be formed in different stages of glycation. Above, dicarbonyl formation is described through different advanced-stage pathways. Not only such "intra-stage overlaps" but also "inter-stage overlaps" between advanced- and final-stage reactions can be observed. As a representative example, both intermediate- and late-stage reactions potentially leading to N-α-carboxymethylation of peptides **30** can be seen in Scheme 1.6. Shown pathways have been originally described for the formation of N-ε-CML.

In the intermediate reaction phase, three possible reaction mechanisms can lead to carboxymethylation (Scheme 1.6, blue arrows). The 1,2-enaminol **4a** (top), the 2,3-enaminol **10** (middle) or the Schiff base **3a** (bottom) can serve as precursors for said glycation modification. Acidic conditions favor the formation of the 1,2-enaminol **4a** from the ARP **5a**, which dehydrates

and produces the imine **7**. Note that **7** contains multiple hydroxy- and oxo-moieties enabling a variety of reactions including keto-enol tautomerism and dehydration reactions. Hence, diverse reaction products may originate from **7** as documented by Monnier and co-workers.[47] The diketo-compound **31** is yet only one of the many possible downstream reaction products. Fragmentation of **31** leads to carboxymethylation as in **30**.[48] The 2,3-enaminol **10**, which is predominantly observed at increased pH values, does not only produce sugar degradation products (for sugar breakdown products from **10**, see Scheme 1.2). Baeyer-Villiger oxidation of **10** *via* reactive oxygen species (e.g., hydroxyl radicals) yields the anhydride **32**. Elgawish *et al.* proposed that hydrolysis of **32** induces the formation of erythronic acid and carboxymethylated amino compounds like **30**.[49] Alternatively, retro-aldol fission of the Schiff base **3a** produces the enaminol intermediate **33**. **33** rearranges into **34**, and oxidation of **34** leads to carboxymethyl moieties as in **30**.[50-52]

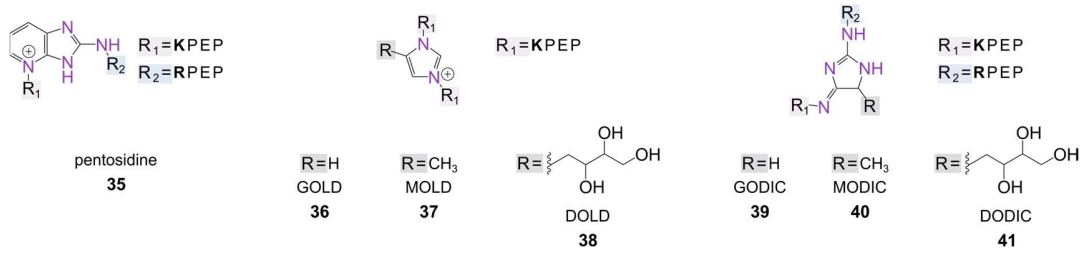


Scheme 1.6: Advanced- (blue arrows) and final-stage (pink arrow) reaction pathways yielding carboxymethylated peptides. Pathways were originally reported for N-ε-carboxymethyllysine. Adapted in part from Hellwig and Cho *et al.*[9,53]

In the final stage of glycation reactions, the so-called Cannizzaro reaction between **13** and free amino groups also leads to formation of carboxymethyl modifications.[52] There are clear intra- and inter-phase overlaps, however, the concept of early and advanced glycation products is simple and still widely accepted.

1.1.4. Crosslinks

As mentioned, reaction intermediates often undergo secondary reactions during the glycation cascade. As a product of final-stage secondary reactions, amino compound crosslinks may be formed. Crosslinking can occur at the N-terminus and at side chain functional groups. In the 1970s and 1980s, studies proposed that glycation-induced crosslinks can have detrimental effects on the structure and function of proteins.[54,55] For example, crosslinking of collagen has been shown to increase mechanical strength (for details on the relevance of glycation in health, see section 1.2.1, Health). Hence, structure identification and characterization of glycation crosslinks became a highly necessary task. Here, we discuss important crosslinks known to form *in vivo* (Scheme 1.7) with focus on their discovery and formation pathway.



Scheme 1.7: Structures of selected crosslinks.

Already in 1981, Monnier and Cerami documented glycation, more specifically glycation-induced crosslinking, as a potential cause for aging of long-lived proteins *in vivo*.[56] Later in 1989, Sell and Monnier successfully extracted and identified the first glycation crosslink, namely pentosidine **35**, from human skin collagen. They also showed that **35** forms in a mixture of arginine, lysine and pentoses upon heating.[57] However, formation of **35** is not limited to pentoses as initially proposed, but was later also observed in presence of glucose.[58] Soon after, Wells-Knecht and co-workers discovered the crosslink glyoxal-derived lysine dimer (GOLD) **36** through reaction of N-α-hippuryl-lysine and glyoxal **13**.[59] Analogous reactions of methylglyoxal **14** and 3-DG **9** give the homologs MOLD **37** and DOLD **38**.[60,61] Another group of important crosslinks are imidazolines. Through reaction of (lysine) amino groups with hydroimidazolones GODIC **39**, MODIC **40**, and DODIC **41** are produced.[62,63]

Naturally, crosslinks discovered for amino acids can also be observed for peptides, and the association of peptide crosslinking with different pathologies has been shown (e.g., Alzheimer's disease).[64] Technically, investigation of protein crosslinks often also depends on the

analysis of crosslinked peptides. Most protein crosslinking studies are performed in a "bottom-up" approach and rely on identification of covalently connected peptides.[65,66]

Peptides were also investigated as anti-glycation agents to prevent protein crosslinking. Basically, peptides should protect proteins by "absorbing" the reactive glycating species. Generally, the idea of anti-glycation agents started when Brownlee *et al*. proposed that aminoguanidine can inhibit arterial wall crosslinking in diabetes patients in 1986.[67] The naturally occurring dipeptide L-carnosine (ß-alanyl-L-histidine), which was first discovered in 1990, was suggested as an anti-protein crosslinking agent in 1995 by Hipkiss *et al.*[68-70] Later studies supported Hipkiss' findings.[71] Alhamdani *et al.* found that L-carnosine could prevent proteins from modification when exposing cells to glucose degradation products.[72] Oral supplementation in diabetic rats fully averted retinal vascular damage, which is an indicator for diabetic retinopathy.[73] Besides carnosine, peptides with aliphatic amines (e.g., lysine) or strongly nucleophilic amino acids (e.g., histidine) were proposed as potential anti-glycation agents.[71,74] It is important to consider that if peptides are used to suppress protein crosslinking, the peptides themselves are often crosslinked instead. For application as anti-glycation agents, it is thus essential to study the properties and health implications of peptide crosslinks.

1.1.5. The nature of the amino reactant

As mentioned above, the course of glycation depends on the chemical structure of the sugar reactant. The structure of the amino compound also plays a crucial role. Different structural properties need to be considered depending on whether the amino precursor is an amino acid, peptide, or protein. Amino reactant structural features, which may direct the route of glycation reactions, are summarized in Figure 1.2. The following section will provide an overview of relevant structural characteristics influencing amino compound reactivity.

Amino acids

In vitro model systems have been widely used as an efficient and relatively simple strategy to understand glycation reactions across different amino compounds (see section 1.3.2, Sample preparation). Many studies focused on amino acids as the amino reactant. Sharing a common core structure, amino acids only differ by the side chain residues (Figure 1.2A). Model system studies revealed that reactivity is considerably influenced by the amino acid side chain structure.[75,76]

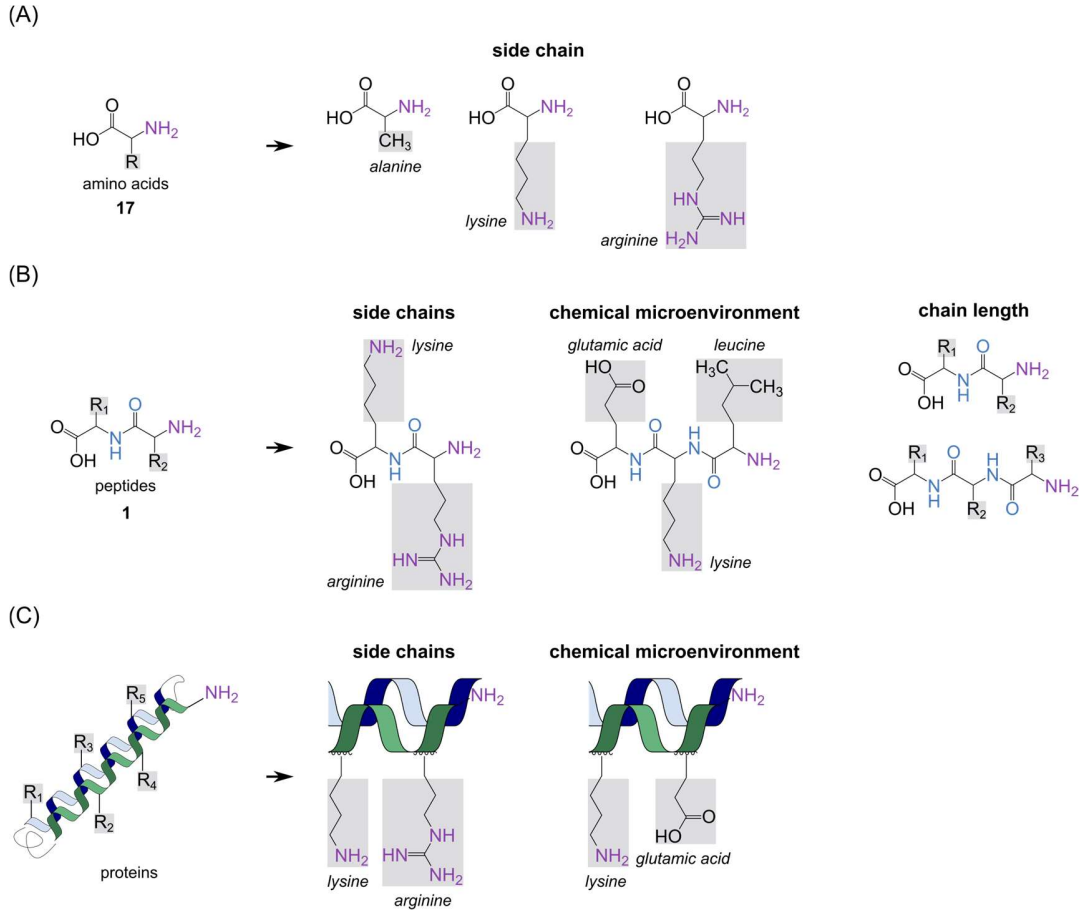


Figure 1.2: Amino compound structural features relevant in glycation reactions. Depending on whether (A) amino acids, (B) peptides, or (C) proteins act as the amino reactant, different structural characteristics need to be considered (e.g., the amino acid side chains, the amino acid sequence, and the amino acid sequence length).

Notably, only the α-amino group may be attacked by sugar reactants for most amino acids. For arginine, cysteine, histidine, and lysine, the side chain guanidino-, thiol-, imidazole- and amino-group function as an additional reaction site. For named amino acids, superior reactivity in glycation reactions was reported, and most studies focus on arginine and lysine as amino acid precursors.[76-79] The side chain glycation sites of both lysine (pK_S = 10.3) and arginine (pK_S = 12.1) are mainly protonated at food-typical (pH = 3-7) or physiological (pH = 7.4) pH values, and thus, barely react with carbonyls in real life.[26] Notably, the reaction behavior of side chain functional groups is different. The ε-amino group of lysine can generally react with practically any carbonyl compound and can initiate the glycation cascade. In contrast, the arginine δ-guanidino group cannot "start" glycation but only react with α-dicarbonyls formed in later stage

reactions.[80] If comparatively simple chemical structures amino acids already show such strong differences in their reaction behavior, the determinants driving peptide glycation appear virtually infinite.

Peptides

Amino acids are the building blocks of peptides. The number of possible permutations increases exponentially with the amino acid sequence length n. Considering all 20 gene-encoded amino acids, the number of peptides can be calculated as 20^n.[81] Hence, a substantially larger number of conceivable peptide structures exist compared to amino acids. Accordingly, more factors need to be considered for peptide reactants (Figure 1.2B). As described above, both the N-terminal amino group and side chain functional groups may act as reaction sites for glycation. In peptide structures, the number of reactive side chain residues may accumulate depending on the peptide amino acid composition.

The tendency of reactive peptide functional groups to undergo glycation differs. All peptides may be modified through glycation at the N-terminus; however, the side chain of N-terminal amino acid is a major influence factor. As described by Van Lancker *et al.*, diverse dipeptides with lysine at the N-terminus produced equivalent pyrazine levels.[82] Dipeptides with, for example, N-terminal leucine led to comparatively lower amounts of pyrazines. The authors proposed that deprotonation at the α-position of the amide moiety may be hindered by the bulkier side chain of leucine.[83] Further, the amino acid sequence, thus the chemical microenvironment of the glycation site, must be considered. Prior work has established that certain amino acid side chains adjacent to the reactive functional group influence peptide glycation. For example, de Kok and Rosing used *in vitro* glycine dipeptide model systems to discover a catalytic effect of the glutamic acid carboxylic group on glycation of the primary amino group. In this study, peptide reactivity was measured as the rate of glucose-conversion and quantified using HPLC.[84] An (iso)leucine-induced increase in lysine-peptide reactivity has been reported by Mennella *et al.* and Zhili *et al.* using a similar experimental strategy.[85,86] While Mennella *et al.* took the percentages of residual peptide at different reaction times as a direct index of peptide reactivity, Zhili *et al.* additionally monitored the levels of peptide-bound pyrraline and 3-DG. Sequence length can also influence peptide glycation. In de Kok's and Rosing's study, glycine di- and tripeptides were investigated, and reactivity decreased with peptide length.[84] Others studied more diverse di- and

tripeptides and confirmed higher reactivity of di- compared to tripeptides.[83,86] These authors suggested that the catalytic effect on the Amadori rearrangement through direct interaction of the carboxylic group and the imino nitrogen is much less likely for tripeptides compared to dipeptides.[83,84]

Each of these studies has only focused on a small number of selected short-chain peptides, and models were incubated under different conditions. Even more, each study has determined "reactivity" using different benchmarks. Early studies relied on color formation, meaning browning of the model system.[87] Already in the 1990s, it was known that browning intensity is not *per sé* proportional to the conversion of reactants.[84] Browning rather reflects, for example, the type of formed melanoidins. Especially for peptides, browning cannot properly indicate reactivity. As early as 1973, Motai *et al.* reported darker degrees of color for peptide-derived melanoidins than for melanoidins from amino acids.[88] Some studies assess peptide reactivity by quantifying the loss of the peptide and/or sugar reactant. Importantly, not only glycation but also degradation leads to decreasing concentrations of reaction precursors structures (see section 1.1.2). Degradation rates differ across peptide structures (see section 1.1.2). Further, peptide precursors can be re-liberated from glycation products.[5] Hence, reactant concentration alone cannot serve as a suitable parameter for the reactivity of peptides. Current studies often use the rate of glycation product formation to determine peptide reactivity. To use the quantity of a reaction product for tracing reactivity, a suitable glycation product needs to be chosen among all possible glycation products. Here, we need to decide, whether one reaction product can reflect the entire glycation reaction cascade.

So far, peptide glycation studies remain isolated instances. Previous results cannot be generalized and have not coalesced into a basic principle for selective glycation, which can be applied to all peptides. Finally, we do not yet understand what controls the susceptibility of certain peptide reaction sites to become glycated. To date, comprehensive investigation of peptides as scaffolds for glycation reactions remains challenging. Clearly, there is a need to understand what promotes susceptibility of peptide functional groups to undergo glycation.

Proteins

As found for peptides, the chemical microenvironment influences reactivity of the protein N-terminus and side chain functional groups in glycation processes (Figure 1.2C). Only a small proportion of all, for example, lysine or arginine residues of proteins undergo glycation.[89,90] A collective of independent studies reported selective glycation of many proteins including hemoglobin, human serum albumin, and α-lens crystallin.[89,91-94] This may also be true for glycation of peptides with multiple reaction sites. Proximity of certain amino acids seemed to play a decisive role. For proteins, the structural surrounding of the reactive functional group, however, does not only depend on primary structure, thus amino acid sequence. In contrast to peptides, the tertiary structure must always be considered in protein glycation studies.[95] Not only "sequence neighbors" can be in close proximity to reactive functional groups. More distant amino acid side chains may also interact with the reaction sites as a virtue of three-dimensional structure.[91,96,97] In human serum albumin, for example, Ahmed *et al.* identified Arg-410 as the major target for modification by methylglyoxal under physiological conditions. Superior reactivity of Arg-410 was attributed to the catalytic base Tyr-411, which is the direct sequence neighbor.[92] Near carboxylic groups were repeatedly proposed to catalyze the rearrangement of protein-ARPs. Shapiro *et al.* noted this effect for lysine residues glycated in hemoglobin, which were in close proximity to carboxylic groups in the primary or secondary structure.[96] For Ribonuclease A, Watkins *et al*. suggested that the acidic amino acids Glu-2 and Asp-38 promote glycation of the reactive lysines Lys-1 and Lys-31 indicating an effect of the primary structure on reactivity.[98] The imidazole group of histidine, which can act as an acid-base catalyst, may also account for site-specificity in the glycation of several proteins, such as liver alcohol dehydrogenase. As discussed by Shilton and Walton, both primary and higher order vicinity may enable glycation site-histidine interactions.[97] Higher order structures do not play a general role in peptide glycation. Some peptides show, for example, secondary helical structures, however, not all.[99] While peptide glycation research can learn from protein glycation studies, there are clear differences.

1.2 The relevance of glycation reactions

Maillard himself already postulated broad relevance of glycation because the reaction precursors, amino compounds and carbonyls, exist and may react practically everywhere in nature.[2] Interestingly, *in vivo* consequences of glycation reactions, such as changes in color, solubility, and

elasticity of proteins with age, were detected even before the discovery and definition of "the MR".[100] Since then, we have learned that understanding glycation is essential for human health, food production and other areas.

1.2.1 Health

Amino acid and protein glycation were the major focus of early investigations. Independent studies linked the accumulation of different glycation products to age and disease (e.g., different neurodegenerative pathologies).[101,102] The most common example for protein glycation may be glycated hemoglobin, known as HbA1c. Hb1Ac was first discovered in 1955.[103] Elevated HbA1c concentrations in diabetes patients were first noted by Huisman and Dozy in 1962, and over the decade confirmed by numerous studies.[104-106] Since the approval of HbA1c levels as a diagnostic tool by the American Diabetes Association in 2009, it has become a key measure for the diagnosis, screening, and monitoring of diabetes.[107] N-ε-CML was the first AGE, which was found *in vivo*, and is one of the most studied amino acid glycation products.[108] It was previously shown that increased N-ε-CML levels in adipose tissue contribute to the development of obesity-associated insulin resistance.[109] However, novel glycation biomarkers are yet to be discovered. For example, a recent study found that previously unknown glycation products, namely glyceroyl-modified amino acids, play an important role in hereditary Parkinson's disease.[110] By contrast, the *in vivo* implications of peptide glycation products are far less comprehensively characterized. There is profound experimental evidence that more attention needs to be devoted to glycation of peptides. Many peptides are bioactive, and so are many glycation products.[111-113] Peptides cover diverse bioactivities, and modification of peptides *via* glycation often leads to changes in bioactivity.[114,115] For example, the peptide hormone gastric inhibitory polypeptide (GIP) stimulates the release of insulin from pancreatic β-cells upon meal-induced secretion.[116] Increased levels of glycated GIP (~20% of the total GIP) were found in obese mice.[117] Glycated GIP showed increased insulin-releasing and antihyperglycemic activity in rats and *ob/ob* mice *in vivo*.[118-120] The altered bioactivity of structurally modified (e.g., glycated) peptides may be due to resistance to *in vivo* degradation or effects on receptor binding.[119,121,122] The bioactive potential of (glycated) peptides and the fact that peptide glycation products form naturally in a glucose-rich environment indicate great potential of further studies on functional peptides to assess the physiological significance of their glycated forms. Due to the vast heterogeneity of peptide glycation products and the lack of

suitable methods for high-throughput identification, we potentially miss out on disease markers and compounds with valuable bioactivities.

1.2.2 Food

The MR is of utmost importance in food production. Both early- and later-stage glycation products are frequently found in foods, and largely contribute to their color, flavor, and taste. Systematic peptide glycation could help to improve all three. As mentioned, melanoidins from peptides show more intense browning than melanoidins from amino acids.[88] An example for glycation-induced aroma are pyrazines, which have a pleasant nutty and meat-like odor.[123] Especially, peptides offer great potential for aroma formation. Reaction of glucose with dipeptides, for instance, led to more pyrazines than reaction with amino acids.[82] Trimethylpyrazine has a very low odor threshold and was produced at higher yields in model systems containing lysine-dipeptides as compared to model systems from free amino acids.[82] Many peptides themselves also show sensory activity. Indeed, peptides cover not only a wide range of bioactivities but also many taste modalities: bitter, salty, sour, sweet, and umami.[124] Bitterness and off-flavor of peptides is a critical factor, which can impede the use of, for example, bioactive peptides in functional foods. Reduction of peptide bitterness is a challenging task, though has been reported after glycation for diverse peptide mixtures.[39,125,126] Peptide glycation often did not only reduce bitterness but at the same time induced or increased bioactivity.[127] Improved flavor of food peptides was also reported independent of bitterness.[128]

In foods, ARPs account for the largest proportion of glycated amino compounds. In fact, the total content of glycation products in foods is commonly calculated as fructosyllysine (g/kg protein or g/kg food product), the ARP from glucose and lysine. For comparison, the concentration of protein-bound fructosyllysine in baked cookies was 459 mg/kg cookie, while the concentrations of protein-bound N-ε-CML, N-ε-CEL and MG-H1 were 7.9, 1.5, and 8.8 mg/kg cookie, respectively.[129] Henle estimated an ARP daily intake of 500 to 1200 mg based on the consumption of 1 L milk, 500 g bakery products and 400 mL coffee.[130] Importantly, the essential amino acid lysine is not available for nutrition, when bound in different glycation products. The human body cannot access lysine when modified as fructosyllysine or crosslinked as lysinoalanine.[131,132] Interestingly, not many studies document the detection of HRPs in food.[129,133]

Summarizing, peptides need to be studied more comprehensively in the context of glycation. Particular attention may be paid to the bio- or sensory activities of their glycation products and peptide-ARPs.

1.3 Analytical methods

1.3.1 Peptide glycation in the family of "omics"

Given the ability to holistically collect data and to study virtually any living entity, from a multicellular organism to a single cell, the diverse research field of "omics" developed.[134] The suffix "omics" implies the comprehensive assessment of a set of molecules. Different "omics" subgroups exist depending on the major biomolecule class analyzed, whether it be deoxyribonucleic acid (DNA, genomics), ribonucleic acid (RNA, transcriptomics), proteins (proteomics) or metabolites (metabolomics) (Figure 1.3).

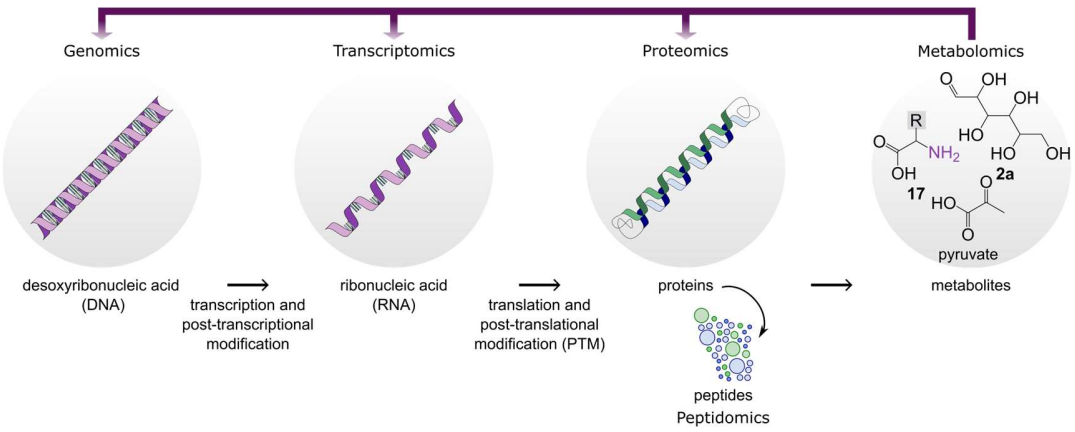


Figure 1.3: "Omics" dimensions including genomics, transcriptomics, proteomics, peptidomics and metabolomics. "Omics" fields are ordered by the flow of genetic information in a biological system. DNA is transcribed into messenger RNA. Using messenger RNA as a template, proteins are synthesized *via* translation. Proteins may undergo post-translational modifications leading to changes in structure and function. Metabolites are last in the "omics" cascade. The arrow indicates influence of metabolites on all previous layers of biological information. Adapted from Patti and Siuzdak, and Krassowski *et al.*[135,136]

Genomics is the most mature "omics" term and was first coined by the American geneticist Thomas H. Roderick in 1986.[137] Genomics measures the genotype of an organism. The initial expectations for genomic studies were high. In 2001, the human genome was successfully sequenced as part of the Human Genome Project, which was believed to be a breakthrough for

understanding and treating pathologies.[138] Indeed, genomic studies provided a very useful framework for investigation of specific genetic variants that play a role in both mendelian and complex diseases.[139] Yet, we also learned that analysis of the genome alone cannot reflect a dynamic and highly complex biological system. The downstream disciplines proteomics and metabolomics are different from genomics. Proteomics and metabolomics measure the phenotype shaped by both the genotype and the past and present environment of the organism.[140,141] Because metabolites are the substrates or products of protein-catalyzed biochemical reactions, metabolomics is broadly acknowledged to be the "omics" discipline that is closest to the phenotype.[135,141,142] To understand the mechanism of a biological process, data from two or more "omics" sciences nowadays are often combined in multi-omics.[136]

Proteomics was coined more than 20 years ago. More specifically, Marc Wilkins proposed the term proteomics in analogy to genomics in 1995.[143] Proteomics describes the study of the entire proteome expressed in a specific state of a cell population or an organism.[144] Proteins are responsible for nearly all the work within a cell and contribute to the structure, function and regulation of cells, tissues and organs.[145] Proteomics can provide information about the (i) quantity, (ii) primary structure (including post-translational modifications, PTMs), and (iii) interactions of proteins.[140,146] These three data types can be used to answer biological and medical questions. For mass spectrometry (MS)-based proteomics, two major workflows exist: "top-down" and "bottom-up".[147] In brief, intact proteins are analyzed in a "top-down" approach, typically in mixtures of limited complexity.[148] Yet, analysis of intact proteins still faces challenges (e.g., protein solubility).[149,150] In "bottom-up" proteomics, also known as shotgun proteomics, the protein is enzymatically digested prior to analysis using sequence-specific proteases (e.g., trypsin). [151]

Numerous "omics" sciences have emerged from the discipline of proteomics, which are engaged in studying narrower arrays of bioorganic compounds, such as peptidomics. Analogously, peptidomics is defined as the comprehensive analysis and quantification of peptides in a cell, organ, body fluid or organism.[152,153] Peptides can, for example, be formed during the breakdown and turnover of proteins.[154] As peptides are virtually found in any biological matrix and play an essential role in physiological functions, differential peptide expression patterns or modification states may serve as an indicator for the presence or severity of some disease state.[152,153,155]

Metabolomics is the youngest "omics" science and describes the systematic study of metabolites. The term metabolome was first introduced in 1998 as the sum of all metabolites given in a biological system under particular physiological conditions, either extra- (exometabolome) or intracellular (endometabolome).[156,157] The biological significance of metabolites is far reaching. The functions of metabolites are very diverse, as are their chemical structures. Metabolites are crucial for cellular functions including energy production and storage, and signal transduction.[158] In fact, also proteins strongly depend on metabolites. For example, acetyl-CoA and nicotinamide adenine dinucleotide (NAD+) act as co-substrates and regulate PTMs.[159,160] Some metabolomics studies include glycation products.[110,161,162] Metabolomics often relies on mass spectral library search, which excludes analysis most glycation products due to a lack of reference spectra (see section 1.4.1).

As implied by the cascadic structure in Figure 1.3 and representative examples, "omics" rather is a complementary puzzle than a collective of independent scientific fields. There is a clear co-dependency of the "omics" sciences. Even more, some analytes cannot be clearly classified into a definite "omics" category. Peptide metabolites, as studied in peptide glycation, are two-component analytes (Figure 1.4). Glycated peptide chemical structures contain a peptide core, and a metabolite-like glycation modification. Hence, peptide glycation may be seen as an intersection of "bottom-up" proteomics/peptidomics and metabolomics.

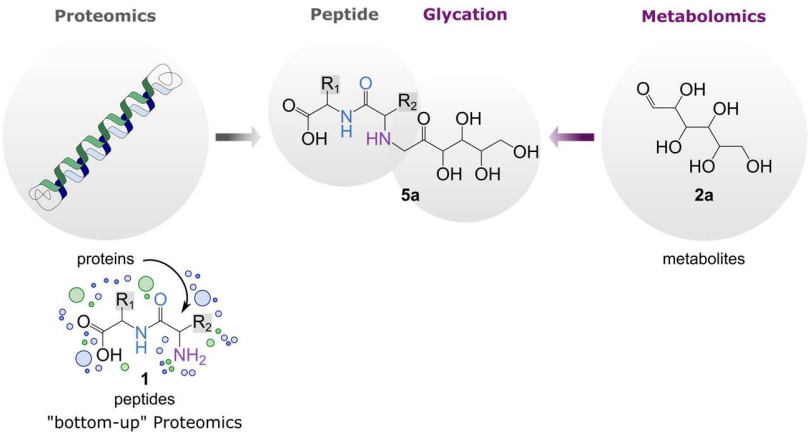


Figure 1.4: Peptide glycation as an intersection between "bottom-up" proteomics/peptidomics and metabolomics. Peptide glycation products are two-component analytes and contain both a peptide and metabolite-like substructure.

Many methods have been developed for the analysis of free glycation products.[163-165] Various different methods have also been established for the analysis of peptides.[166-168] Due to their collage-like structure, peptide glycation studies lack suitable approaches. Methods optimized for the detection of peptide-bound glycation modifications are needed. To elucidate the structure of glycated peptides, identification of their peptide backbone is necessary. Note that the performance of conventional peptide sequencing tools may be compromised by the presence of the glycation modification side chain. Besides commonly detected PTMs, such as phosphorylation and acetylation, many glycation modifications play an important role in, for example, human health (see section 1.2.1) but are not considered in proteomics. The peptide-bound glycation "metabolite" is commonly small molecule-like and often not described in literature. Identification of the glycation modification substructure holds similar challenges to non-targeted metabolomics. However, application of metabolomics approaches may be impeded by the peptide core. The following section aims to give an overview of analytical tools from proteomics/peptidomics and metabolomics, which should be considered when investigating peptide glycation. Opportunities and challenges in the application of classic "omics" methodologies for peptide glycation products will be discussed.

1.3.2 Liquid chromatography-mass spectrometry for peptide glycation studies

Due to its high speed, sensitivity and specificity, high resolution MS is currently the most prominent platform for both proteomics and metabolomics studies. Analytes are usually not introduced to the mass spectrometer all at once. Instead, samples are injected onto an (ultra) high-performance liquid chromatography (UHPLC) column that is directly coupled to, or is 'on-line' with, the MS system (Figure 1.5).[169] This enables separation of analytes by their physicochemical properties prior to mass spectrometric analysis. The mass spectrometer records the mass-to-charge ratio (m/z) of the compounds in a MS^1 full scan. For structure elucidation, the analytes are subjected to tandem mass spectrometry (MS/MS or MS^2). In the following, an overview of the components of a LC-MS/MS system is given focusing on UHPLC-quadrupole time-of-flight (QToF) MS/MS. UHPLC-QToF MS/MS was used to generate data throughout this thesis.

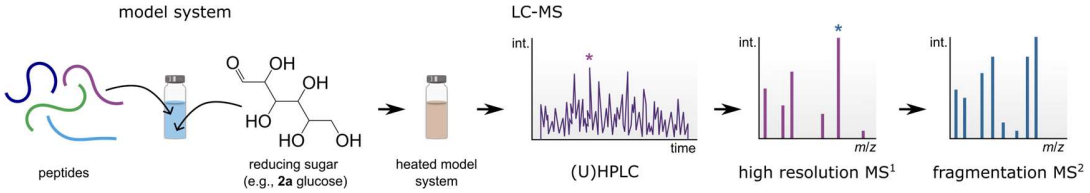


Figure 1.5: Generic LC-MS/MS workflow for the analysis of peptide glycation. Peptides are mixed with a reducing sugar. Model systems are heated to form peptide glycation products. Analytes are separated on a liquid chromatography system that is coupled to the mass spectrometer. At a given retention time, the exact mass-to-charge ratio is measured in a high-resolution MS^1 scan. Selected precursors are isolated and fragmented in MS^2 scans for structural information. Asterisks indicate the origin of the MS^1 and MS^2 scan, respectively. Adapted from Altelaar *et al*.[170]

Sample preparation

"Omics" analyses can be performed from diverse biological materials: from cells through tissues and body fluids. A critical step is the efficient, structure-wide extraction and solubilization of analytes. The immense structural heterogeneity can impede this endeavor. To improve recovery, diverse extraction protocols and combinations of agents have been described.[171-174] Equal attention must be paid to sample preparation if glycated peptides are extracted from biological matrices.[175]

Else, peptide glycation often is investigated from *in vitro* model systems (Figure 1.5). Model systems allow to detect unprecedented targets for which chemical tractability has yet to be established. Thus, models are particularly suitable for non-targeted analysis and potential detection of previously unknown glycation modifications. Historically, in fact, many first-in-class glycation products have been discovered in studies that monitored glycation reactions in model systems.[176,177] For example, **6** was first identified in peptide, namely leucine-enkephalin, hexose model systems.[14] To this date, novel glycation products continue to be found.[110] However, model systems are not only a valuable source of *hitherto* undescribed structures. The rate, extent and course of glycation depends on many factors: the structure of the sugar and amino reactant, temperature, time, pH, and water activity (see section 1.1).[18,86,178,179] It is complicated to study glycation reactions from real-life samples, such as food or tissues. Compounds that are naturally found in said matrices (e.g., vitamins) can interfere with the reaction pool.[180] Model systems, however, are much better controlled and less complex. More defined reaction conditions and reduced complexity allow to assess precursor reactivity, trace downstream reaction products, and

resolve reaction mechanisms. For instance, Chu *et al.* determined possible formation routes of peptide-imidazolidin-4-ones in this way (see section 1.1).[17]

For preparation of model systems, peptides and reducing sugars are mixed at a defined molar ratio and heated. To reduce the influence of extrinsic factors, models may be heat-treated in sealed vials or tightly capped screw-sealed tubes.[181,182] As a solvent, water or buffers are commonly used. Even though buffers may help to maintain the pH, it is known that buffering agents can influence the course of glycation.[183] Studies have uncovered a correlation between specific buffer components and, for example, the rate of protein glycation.[184] This effect appears to be site-dependent, and must also be considered for peptides.[185] Model systems are not extracted but analyzed directly or after dilution. This minimal sample work-up reduces the risk of enriching or losing analytes and helps to prevent biased analysis. In this thesis, peptide glycation was exclusively studied from aqueous model systems. Heat treatment was performed in sealed glass vials.

Liquid chromatography

Mixtures of compounds, such as peptide-based model systems, can be separated with a gradient of aqueous and organic solvent using liquid chromatography (LC). The MS can distinguish analytes by their *m*/*z*. However, their chromatographic retention time can serve as a highly specific identifier to assist structure elucidation.[186] Isobaric analytes are an obvious problem for identification using only accurate mass. There are two main reasons for such ambiguities in peptide glycation product annotation: (i) the makeup of the peptide backbone from amino acids gives rise to many isomeric permutations, (ii) the modification is most often smaller, yet covers a much wider chemical space and is mostly unknown; here, the frequent occurrence of heterocycles and unsaturated bonds further increases the odds of misidentifications.[187] Some isomers can be distinguished *via* diagnostic fragmentation (see section 1.3.6). However, sometimes even high mass accuracy and fragmentation cannot help.[188] The MS/MS spectrum may be uninformative, contain interferences or not even exist. Here, the retention time is also required to identify an isomer. Further, chromatographic separation can improve the mass spectrometric analysis by reducing disturbances, such as ion suppression.

Gradients may last from a few minutes to several hours depending on, for example, the column type, flow rate, and sample complexity. Like metabolomics gradients, LC methods for

peptide glycation analysis are often rather short. To date, reversed phase (RP) columns are the golden standard for the separation of (modified) peptides.[189-191] RP separation is also applied in metabolomics studies (e.g., for analysis of bile acids or short-chain fatty acids).[192,193] In RP chromatography, the compounds are eluted from the column using a solvent gradient with increased organic content. This strategy enables to elute analytes in order of their hydrophobicity. It is important to mention that small, polar analytes usually do not retain on such columns. Metabolites cover a broad spectrum of polarity, and their retention behavior varies, widely. For analysis of highly polar metabolites and free glycation products, hydrophilic interaction liquid chromatography (HILIC) has shown great potential.[194-196] In the proteomics field, HILIC allowed to improve the analysis of PTMs.[197] The small molecule-like structures of glycation modifications are diverse and comparable to metabolites. The polarity of peptide-bound glycation products, unlike classic metabolites, also strongly depends on the amino acid sequence and length of the peptide backbone. Glycated short-chain peptides (e.g., dipeptides) with a high proportion of hydrophilic amino acids are more probable to not retain on RP columns. Increased peptide length and content of hydrophobic amino acids promotes interaction with RP column material. The focus of this thesis was glycation of hydrophobic short-chain peptides, namely (iso)leucine dipeptides, and longer-chain peptides, which showed sufficient retention on either C8 or C18 RP columns.

Time-of-flight mass analyzers

The analytes are ionized by electrospray ionization (ESI) at the end of the LC column (Figure 1.6, left).[198] The eluent is vaporized, and the electrospray plum is generated at atmospheric pressure near the entrance of the mass spectrometer. From here, the ions are transferred into the vacuum of the MS instrument.[169] In this thesis, a QToF mass spectrometer was used. Due to their high resolving power of up to 50,000, high mass accuracy (≤ 2-10 ppm), and superb scanning rates, QToF mass analyzers are often used for non-targeted analyses.[162,199-201]

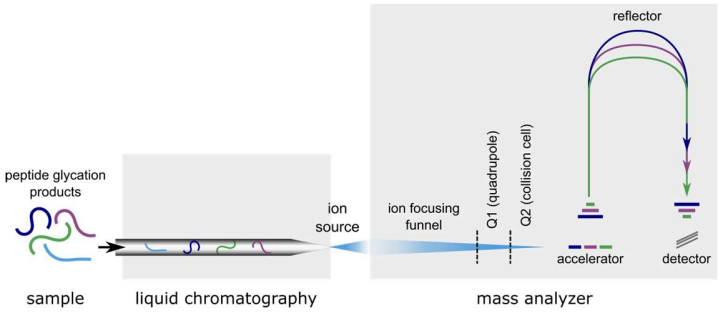


Figure 1.6: Scheme of peptide glycation product analysis by a LC-coupled hybrid QToF-MS. Adapted from the maXis™ user manual version 1.1, Bruker Daltonics GmbH).

In a QToF system (Figure 1.6, right), the first quadrupole (Q1) functions as a mass filter. The second quadrupole (Q2) is the collision cell. In full scan mode, the Q2 only serves as a transfer unit leading the ions into the flight tube. In a MS/MS experiment (see section 1.3.4), ions from the Q1 are fragmented in the Q2 *via* collision-induced dissociation (CID) with nitrogen gas. As implied by the designation, ions are separated by their flight times in a ToF instrument. An accelerator pushes the incoming ions to the reflector. The drift time of each ion is measured, starting from acceleration until reaching the detector unit. Providing their initial energies are identical, the flight time is longer for larger molecules than for smaller analytes. In brief, the lower the *m*/*z* of an ion, the shorter is its time of flight.[202] The resolving power of reflector ToF instruments is higher than for linear ones because the reflector compensates the different kinetic energies of ions with the same mass.[199,203]

1.3.3 Non-targeted mass spectrometry for comprehensive peptide glycation research

It is difficult to ask revealing questions at the start of your analysis because you do not know what insights are contained in your dataset.

— Hadley Wickham and Garrett Grolemund

The main analytical methodologies that are used for detection and identification of (peptide) metabolites are targeted and untargeted MS. Targeted MS requires *a priori* knowledge about the sample composition, which is not always available. Targeted methods, however, provide high sensitivity and selectivity for the chosen candidates.[204] For the analysis of peptide model systems, targeted MS approaches classically rely on the decrease of reactants or the enrichment of peptides with known glycation modifications by changes in reaction conditions (see 1.1.5, Peptides).[86,178,182] The LC-MS/MS methods are developed and optimized for specific analytes and reaction pathways of interest. However, science in the current century is becoming an ever more data-driven endeavor.[205] This is both due to substantial improvements in the quantity and quality of MS data, and also the vast increase in algorithms and tools used for the processing and analysis of such data sets.[140,145,206] Based on that, a dynamic field of research, namely untargeted MS, has emerged. Non-targeted analyses are of central importance in the discovery of unknown pathways and analytes. Targeted MS cannot find previously undescribed molecules because the target simply cannot be defined prior to analysis. More and more researchers came to realize not only the opportunities but also the limitations of targeted studies: while well-known reaction products are

the cornerstone of peptide glycation research, they limit our view, and we cannot see the whole picture.

In general, non-targeted MS aims to measure the broadest spectrum of analytes in a sample without any prior knowledge about its molecular composition to gain deeper insights into various aspects of living systems and chemistry. The result of untargeted measurements is a large and complex set of data. For non-targeted data interpretation, efficient computational tools are needed (see section 1.4) to identify and statistically evaluate analytes in relation to the phenotype and to determine their interconnectivity in formation pathways or molecular networks. Untargeted MS approaches have been essential in propelling our understanding many aspects of chemistry and biology; however, analytical methods and computational methods need to be further optimized to effectively assess and handle non-targeted data.

1.3.4 Tandem mass spectrometry

Tandem mass spectrometry (MS/MS or MS^2) performs multiple rounds of mass measurements (Figure 1.7). After recording a survey full scan over the full *m*/*z* range, a subset of precursor ions is automatically selected and subjected to fragmentation. A MS/MS spectrum is acquired in a second round of mass measurements measuring the product ions.[207] Ions are collected for fragmentation within the so-called precursor isolation window. Depending on the MS/MS acquisition mode, the *m*/*z* isolation window may be narrow or comparatively wide.[145]

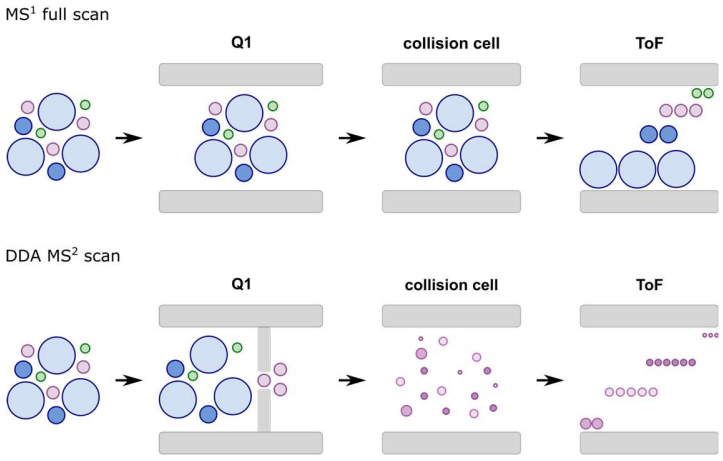


Figure 1.7: Tandem mass spectrometry in data-dependent acquisition (DDA) mode. After a precursor (MS^1) full scan (top), a set number of the most abundant ions are selected for fragmentation. Selected precursors are successively isolated and fragmented in a product

ion (MS^2) scan (bottom).

In targeted mass spectrometry, multiple reaction monitoring (MRM) or parallel reaction monitoring (PRM) experiments are commonly used to (quantitatively) monitor or identify a set of peptides or small molecules.[27,204,208,209] The analytes of interest are isolated using a small isolation window around the theoretical *m*/*z*. Evaluation of MS/MS data generated from targeted measurements is straightforward and can be routinely performed using dedicated software.[210] For time-efficient and comprehensive data acquisition in untargeted approaches, a fast and automated fragmentation mode, namely data-dependent acquisition (DDA), is implemented. In DDA mode, the instrument selects a set number of precursors from the most abundant ions in the survey full scan spectrum and immediately subjects them to fragmentation applying a narrow isolation window.[145,211] This setting can provide both quantitative and structural information for thousands of analytes from the same sample in a single measurement. Even more analytes are fragmented by modern mass spectrometers with higher scan rates allowing for improved sampling depth.[212-215] Still, many analytes remain unsampled calling for even faster MS/MS acquisition.[216] To make sense of the accumulated spectra and to not waste the valuable structural information, automized strategies and computational data evaluation tools are usually required (see section 1.4). It should be mentioned that DDA depends on the precursor intensity, thus features with low concentration or poor ionization may never be selected for MS/MS analysis. Data-independent acquisition (DIA), including sequential acquisition of all theoretical masses mass spectrometry (SWATH-MS), is not driven by the intensity or *m*/*z* of the precursor but comes with its own technical and computational challenges. To solve the latter, first approaches were provided.[217-219] For DIA, a comparatively wide isolation window is used.

Overall, MS/MS analysis adds an additional dimension and thus improves annotation confidence providing highly valuable information for structure elucidation and identification of elementary composition. Strategy development for optimized evaluation of LC-MS/MS data is still ongoing. At present, evaluation of DDA data is much more advanced and generally more straightforward compared to DIA. DDA was used throughout this thesis.

1.3.5 Peptide Sequencing

In the past, a technique known as Edman degradation was used to sequence a protein. This method, in which one amino acid after another is cleaved off the N-terminus, was developed by

Peer Edman and used to determine the primary sequence of peptides and proteins for approximately three decades.[220,221] Importantly, this approach often failed. Either no sufficiently long or an unambiguous peptide sequence could be assigned.[169] As Edman reaction requires a free N-terminus, the method could not be applied in case of amino terminus blockage (e.g., by acetylation).[222] Investigation of N-terminal glycation would have been impossible.

In 1990s, MS-based peptide sequencing largely displaced Edman's technique because of its much higher sensitivity and speed. Peptides could be fragmented within seconds instead of hours or days.[223] Their identification could even be performed in complex mixtures without any prior purification steps independent of N-terminal modification. Figure 1.8 explains how peptides fragment in MS/MS and how the C- and N-terminal fragment ions are designated (Roepstorff–Fohlmann–Biemann nomenclature).[224] Most commonly, peptide fragmentation is induced by collisions with residual gas and peptides structures break at the peptide bonds between the amino acids because bond cleavage mainly occurs through the lowest energy pathway.[169]

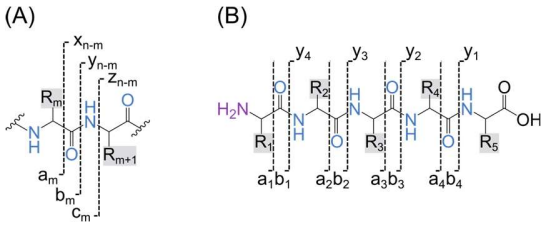


Figure 1.8: Peptide backbone fragmentation. (A) Roepstorff-Fohlmann-Biemann nomen- clature of C- and N-terminal ions. (B) Example for a-, b-, and y-ion designation of a schematic pentapeptide. Adapted from Steen and Mann.[169]

Fragment ions are labeled consecutively from the peptide N-terminus a_m, b_m and c_m. m is the number amino acid residues that the fragment contains. C-terminal ions are designated as $x_{(n-m)}$, $y_{(n-m)}$, $z_{(n-m)}$. Here, n is the number of amino acid residues, which are part of the fragment. m equals the number of amino acid residues that the counterpart N-terminal fragment ion would contain. Due to the loss of water or ammonia, so-called satellite fragment ions may form (e.g., $b_m - H_2O$).[224,225] (Almost) complete mass ladders or ion series allow for identification of the peptide structure (see section 1.4). Besides breakage of single peptide bonds, double backbone fragmentation can give rise to internal cleavage ions. For QToF instruments, a-, b-, and y-ions are usually observed, and y-ions are most abundant.[169]

Obviously, the MS/MS spectra of peptide glycation products strongly depend on the peptide core structure, which fragments according to the above-described rules. The abundance of the regular peptide backbone ions may vary from unmodified peptides. Modifications generally influence the probability of peptide fragmentation pathways.[226,227] For example, dimethylation or imidazolinylation of double charged peptides from proteolytic digestion of human embryonic kidney 293 (HEK293) proteins and bovine serum albumin (BSA) with Lys-N led to a significant increase in peptide sequence coverage, which enabled more reliable peptide identification. In contrast, nicotinylation severely suppressed fragmentation of the peptide backbone.[228] Further, the amino acid sequence plays an important role. Not each peptide bond is equally probable to break. Diverse effects of certain amino acids on peptide backbone fragmentation have been described (e.g., proline effect). Particularly abundant y-ions are, for example, observed upon CID of proline-containing peptides from preferred cleavage of the amide bond N-terminal to the proline residue.[229] All these factors need to be considered when interpreting MS/MS spectra of peptide glycation products.

1.3.6 Glycation product diagnostic fragmentation

High-resolution mass spectrometers can differentiate between analytes, including glycation products, with a single unit mass difference. Hence, the *m*/*z* of a compound is a valuable parameter for structure assignment. Preceding separation of analytes by LC increases selectivity as the retention time can be used to confirm the identity of compounds (see section 1.3.2, Liquid chromatography). Sometimes the applied LC method, however, is not capable of separating two analytes of interest. The extracted ion chromatograms (EICs) may strongly or even fully overlap. The latter becomes obvious when inspecting the MS/MS spectra along the chromatographic peak. If two isobaric compounds co-elute, *m*/*z* and retention time are not sufficient for unambiguous identification of such compound pairs. This is particularly true for isomeric structures, which share the same *m*/*z* and often elute at similar retention times. The addition of fragmentation as a third dimension can often help to ease identification and increase annotation confidence. Highly specific structure-dependent fragments can be monitored (Figure 1.9). These fragments are so-called diagnostic ions. In fact, sometimes isomeric structures can only be differentiated based on the abundance of diagnostic fragment ions or after extensive LC method development, even with reference standards. For example, Yuan *et al.* proposed diagnostic fragmentation pathways for early-stage glycation products from amino acids.[133] Accordingly, Schmutzler *et al.* suggested that

such diagnostic ions also exists for peptide ARPs and HRPs.[230] Interestingly, the diagnostic fragmentation mechanisms were not the same for amino acids and peptides. This shows once again that "rules" for amino acid glycation cannot be simply adopted for peptides but need to be individually studied and adjusted.

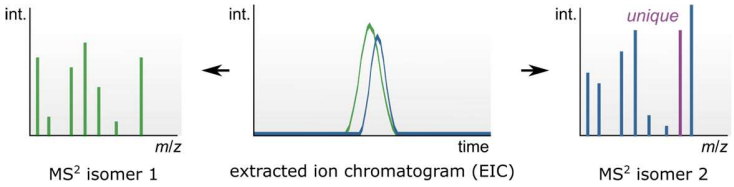


Figure 1.9: Schematic illustration of co-eluting isomers with diagnostic fragments.

If diagnostic fragmentation pathways are known, LC-MS/MS can detect, identify, and quantify isomers, simultaneously, even without baseline separation.[230] There is a need to establish bioinformatic approaches, which enable for fast determination of such diagnostic MS/MS signatures. This is especially true for peptides glycation products. While many isomeric structures are observed in peptide glycation (see section 1.3.2, Liquid chromatography), the highly complex fragmentation signatures of peptide glycation products impede efficient manual discovery of systematic patterns.

1.3.7 Stable isotope labeling

Many glycation products remain unaccounted to date. In peptide glycation, the many (unknown) modifications are combined with highly heterogenous peptide structures. The MS/MS spectra of peptide glycation products do not only contain regular peptide fragments or fragments, which are purely from the glycation modification. Mostly, parts from both the peptide and modification substructure are retained during fragmentation (see Chapter 3). Isotopic tags can help to systematically navigate through the structural maze and simplify interpretation of MS^2 spectra. Isotope labeling has been used for many years, for example, in metabolomics for identification, and the analysis and discovery of reaction and metabolic pathways.[231,232] By comparison of MS/MS spectra with and without isotopic labels, label-induced mass shifts between fragments can be detected (see Chapter 3).[17] In this way, fragmentation pathways and, thus, the structure of the glycation modification can be assigned faster and with higher confidence. Systematic isotopic labeling of reactants or intermediates can enable the unambiguous tracing of heavy elements, even through complex reaction (or metabolic) networks.[233,234]

1.4 Computational mass spectrometry

In recent years, it has become increasingly easier to generate large-scale data sets and to share results and methods across the globe. Substantial improvements in mass spectrometric data acquisition have led to a veritable explosion of data. In the field of glycation, large model system data sets across various amino compounds, different sugar precursors and conditions are now routinely generated using non-targeted high-resolution MS. These well-controlled models help to understand the influence of precursor reactivity and reaction conditions mirrored by reaction product heterogeneity (see section 1.3.2, Sample Preparation). Analyte identification is an essential step in MS-based studies. To determine the identity of compounds in complex mixtures, MS/MS spectra are widely used. However, identification of the ever-increasing number of analytes represents a major impediment to the effective use of high-quality mass spectrometric data. Manual interpretation of this deluge of data is cumbersome and time consuming. At some point, manual annotation simply becomes impossible. There is a crucial need for algorithms that distill the information masked in complex non-targeted analyses. In fact, computational MS has emerged as an own scientific field with a plethora of methods and tools. A variety of valuable bioinformatics strategies have been established in both proteomics and metabolomics, many focusing on analysis of MS/MS.[235-240] Yet, this field needs further development. Besides, most of the current MS-based computational approaches have been tailor-made for defined "omics" fields. "Inter-omics analytes", such as peptide glycation products, do not or only partially benefit from these tools. Methods are needed, which consider both the peptide core structure and glycation modification of glycated peptides rather than specialize in either peptide or small-molecule identification. In this section, common "omics" computational approaches will be described together with their opportunities and drawbacks for the analysis of peptide glycation products.

1.4.1 Mass spectral reference libraries

A conventional tool for analyte identification from MS/MS data are spectral libraries[241,242]: collections of chemical structures and their fragmentation signatures. Examples for tandem mass spectral repositories are LipidBlast, MassBank, the Human Metabolome Database (HMBD), and the National Institute of Standards and Technology (NIST) mass spectral library.[243,244] Experimental spectra are compared to reference spectra (Figure 1.10).

To reflect the likelihood that an experimental and a library spectrum were generated from the same chemical compound, a spectral similarity score is calculated. While the dot product (cosine) score is well-accepted and often used for reference library searching, different scoring methods for the calculation of MS/MS similarities are available.[239,240,245-247] For each experimental spectrum, the database match with the maximum score will be returned. Alternatively, a defined number of top database hits can be the output. As implied by the plethora of matching algorithms, there is no "one fits all" solution for all chemical structures. Development of tailored similarity algorithms, customized for specific analyte classes, can optimize spectral library search.

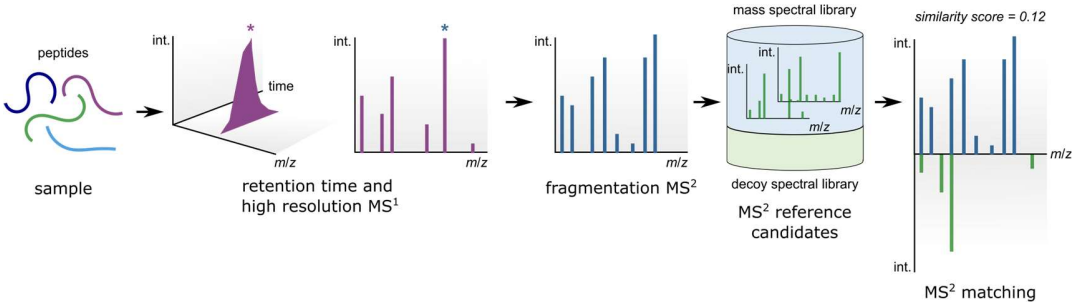


Figure 1.10: Mass spectral library search. Analytes (e.g., peptides) are measured using LC-MS/MS. Asterisks indicate, where the MS^1 and MS^2 scan are recorded, respectively. Collections of reference tandem mass spectra are compared to experimental fragmentation patterns. For each spectral comparison, a similarity score is calculated. If in-house reference libraries are used, retention times may also be matched to reduce false-positive annotations.

Mass spectral libraries allow for fast and reliable analyte identification. Naturally, spectral library search only enables matching of pre-recorded structures. Analytes without available reference spectra cannot be identified. Mass spectral libraries have steadily grown in the past and now contain spectra for hundreds of thousands of compounds.[242,243] Importantly, many spectra are from the same redundant structures. Even though NIST14 MS/MS previously contained 193,120 spectra, said database only covered 9344 compounds.[243] In comparison, known chemical structures deposited in PubChem, ChemSpider and Chemical Abstracts Service (CAS) account for more than 100 million structures combined.[243] That said, it is not surprising that most analytes remain "unknowns" in non-targeted analyses. It is important to note that, despite the continuous growth of MS/MS repositories, there are barely any glycation product reference spectra available. Even for long-known AGEs, such as CML, no recorded MS/MS exists in said libraries to my best knowledge. Especially for glycated peptides, mass spectral libraries cannot support untargeted

studies due to nonexistent reference spectra. Summarizing, the size and availability of reference MS/MS spectra are limited, and MS^2 libraries may never be complete. The lack of most reference spectra calls for development of alternative computational strategies.

1.4.2 *In silico* structure identification tools

For analytes without reference standards, which are not available in mass spectral libraries, structure assignment is a challenging task. As stated above, this is true for nearly all (peptide) glycation products. To overcome the bottleneck of the limited coverage of mass spectral libraries, researchers have developed diverse computational tools. These tools can be loosely divided in two categories (Figure 1.11): (i) compound-to-MS, and (ii) MS-to-compound matching.[236] It is time for peptide glycation to find its place within the capabilities of computational annotation strategies. For that purpose, we need to better understand the fragmentation behavior of peptides with glycation modifications (see Chapter 3).

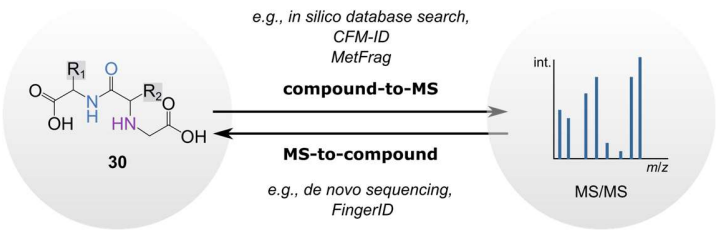


Figure 1.11: Classification of *in silico* annotation tools. *In silico* computational methods for structure annotation from MS/MS can be divided into compound-to-MS and MS-to-compound strategies.

Compound-to-MS methods

As the designation implies, compound-to-MS *in silico* methods start from an input structure and provide a predicted MS/MS spectrum as the output. To construct the MS^2 of a compound, pre-defined handmade and/or computer-learned rules are applied. It is important to consider that computer-learned rules (e.g., from machine learning) depend on the type and quality of the training data set. Tools, which are based on rules learned from metabolomics data, may not work for modified peptides and *vice versa*. Matching an experimental MS/MS spectrum against the

output *in silico* spectrum can help to identify unknown analytes. An example for a compound-to-MS approach from handmade rules is *in silico* database search (Figure 1.12), which is currently one of the most applied techniques for peptide identification in "bottom-up" proteomics.[248-252]

In brief, the search engine translates genomic database information into protein sequences. Said proteins or tailored input sequences are digested *in silico*. Depending on the enzyme chosen for digestion, different cutting rules are applied. For example, the commonly used enzyme trypsin cleaves after arginine or lysine ("Keil rules").[253,254] Besides co- or post-translational modifications, sample processing induced peptide modifications can be allowed. Based on the precursor mass, possible peptides for tandem mass spectral comparison are selected within a specified *m/z* tolerance window. According to the mode of fragmentation, candidate peptides are fragmented *in silico* (for theoretical peptide fragmentation, see section 1.3.5) and compared to the experimental spectrum.[145] A more recent example for a MS-to-compound strategy is Prosit, which predicts peptide MS/MS by deep learning.[255] The deep neutral network additionally enables prediction of chromatographic retention time and fragment ion intensities.

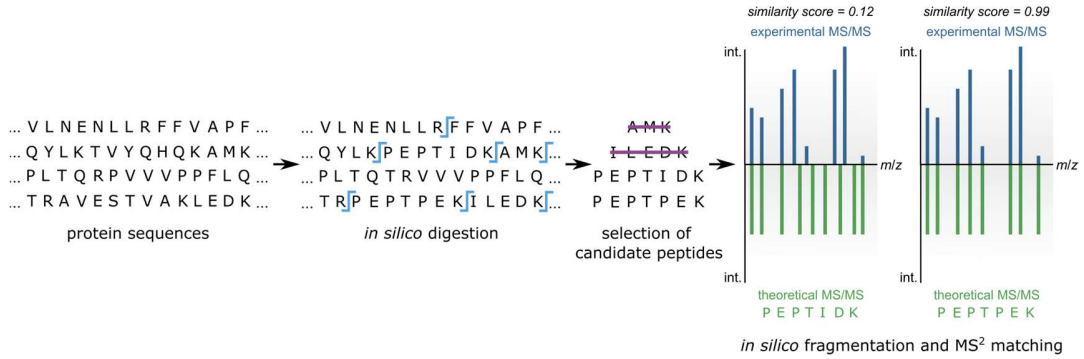


Figure 1.12: Peptide *in silico* database matching.

Peptides follow a systematic, comparatively simple mode of fragmentation unlike other compound classes (e.g., small molecules). A well-known method suitable for predicting ESI-MS/MS spectra of more diverse (metabolite) chemical structures is CFM-ID (competitive fragmentation modeling for metabolite identification):[256] a probabilistic generative model for the MS/MS fragmentation process that uses machine learning techniques to learn its parameters from data.[257] After computing all possible fragments for a molecule, CFM-ID constructs a MS/MS spectrum considering the chance of the persistence transition for the given fragment.[236] MetFrag, which was launched in 2010, is another important *in silico* fragmenter. MetFrag retrieves candidate structures from compound databases (e.g., PubChem, ChemSpider, Kyoto Encyclopedia of Genes

and Genomes [KEGG]). Alternatively, customized candidate structures can be submitted *via* so-called structure data files. After *in silico* fragmentation using a bond dissociation method, the experimental MS/MS spectrum is compared against the candidate fragmentation signatures. MetFrag uses its own scoring algorithm for selection of the best match. The score is a function of the *m/z*, intensity, and bond dissociation energy of the matched peaks. A limited number of neutral losses (5 in total) account for rearrangements.[258] In MetFrag 2.2, CFM-ID was added, and can be used as an additional scoring method, which revealed that results from MetFrag may be further improved using complementary *in silico* fragmentation approaches.[237]

MS-to-compound methods

In non-targeted analyses, there are many MS/MS, and manually finding a candidate structure for each precursor *m/z* comes with a high workload. *In silico* tools from the second category do not need any input structure but only experimental MS/MS data, which is convenient for high-throughput annotation. MS-to-compound strategies distill chemical and/or physical properties from the experimental spectrum of the unknown compound (e.g., the molecular formula, and chemical fingerprints). Said analyte features are then used to compute the candidate chemical structure. It should, however, be noted that these methods require high-quality MS/MS spectra.[259] Incomplete fragmentation patterns complicate structure identification and increase false annotation rates. The quality of a MS^2 may be decreased by contaminant peaks (e.g., created by co-fragmentation).[260]

To determine the structure of a peptide from a MS/MS spectrum, *de novo* sequencing leverages their systematic breakage between the single amino acids. Cleavage of peptides in MS/MS is described in section 1.3.5. (Nearly) complete "mass ladders" allow to reconstruct the amino acid sequence by calculating the mass differences between consecutive fragment ions. Among many algorithms for *de novo* sequencing, particular attention was devoted to the software package Lutefisk, which uses a graph theory approach.[261,262] The MS/MS is translated into a "sequence graph", where nodes correspond to peaks, and edges are constructed if two nodes differ by the mass of an amino acid.[263] The software then attempts to find the path from the N- to the C-terminus, which connects all the (b- and) y-ion nodes resembling the amino acid sequence.[262]

De novo sequencing, however, is again specific for peptides and cannot be applied for other compound classes. However, the basic principle of *de novo* annotation has also found its way

to the metabolomics community. One the state-of-the-art MS-to-compound method suitable for small molecules (e.g., metabolites) is Heinonen *et al.*'s FingerID. FingerID is a two-step pattern recognition approach based on machine learning. After (i) predicting a set of characteristic chemical fingerprints from the MS/MS using a kernel-based approach, (ii) the fingerprints are matched against large molecular databases (e.g., PubChem) to obtain a list of candidate metabolites. In this way, the identification model is capable of generalizing and covering compound structures, which are not present in reference spectral libraries.[238] FingerID was later extended to CSI:FingerID and incorporated into the webservice SIRIUS 4.[264] CSI:FingerID does not only rely on the prediction of the query compound's molecular fingerprint but also its fragmentation tree, which enabled to substantially increase the number of correct identifications.[264,265]

1.5 Motivation and aim of the thesis

Peptides with modifications can virtually be found in every biological system. Their association with numerous diseases has become increasingly evident.[266,267] Peptides can be modified enzymatically and non-enzymatically.[268] Glycation is an important non-enzymatic reaction, which can lead to modification of peptides. Among enzymatic PTMs like methylation or phosphorylation, some glycation-induced modifications are routinely analyzed in proteomics experiments, such as carboxymethylation.[269-271]

Even though broad relevance of glycation reactions is widely accepted, the structure and significance of many glycation modifications is unknown, and especially peptides are underrepresented in glycation research. Factors, which contributed to the lack of knowledge on peptide glycation are among others: (i) the immense number of theoretical peptides, (ii) the structural heterogeneity of glycation modifications, and (iii) the lack of strategies for identification of peptides modified with unknown structures. The main objective of this thesis was to improve the understanding of peptide glycation. This aim can be achieved in many ways. Development may target any step of peptide glycation studies from sample preparation to data evaluation. Detectability of analytes depends *inter alia* on concentration. By optimizing sample preparation, the concentration of AGEs can be increased. Prasanna *et al*. recently established a strategy for enrichment of peptides, which are modified with AGEs like CML **22**, CEL **23** and MG-H1 **26**, using copper(II) immobilized metal affinity chromatography.[272] In terms of improved data evaluation,

innovative methods for MS/MS similarity analysis have been developed, which may allow to include unknown glycation products into non-targeted data interpretation.[239,246]

In this thesis, this research objective was approached from two perspectives: (i) experimental setup and (ii) data analysis. In Chapter 2 (**Berger *et al.*, Scientific Reports, 2021, *11*, 13294**), we use non-traditional multi-peptide glucose model systems to describe time-dependent formation of peptide-bound ARPs. Using high-resolution MS, we investigated amino acid sequence dependent trends in the early-glycation of structurally diverse peptides. Reaction behavior of tryptone peptides with potential bioactivities was studied by integration with bioactive peptide databases.

To enable analysis of not only a single pre-defined but any peptide glycation product, we then developed an algorithm for *de novo* identification of modified peptides from tandem mass spectra (**Berger *et al.*, ACS Analytical Chemistry, 2022, *94*, 5953-5961)**. While fragmentation of well-established glycation products had been known for a long time, there is a lack of strategies for untargeted interpretation of MS/MS spectra from peptides with glycation modifications. Our approach helps with the high-throughput annotation of peptides modified with both known and unknown glycation products.

Chapter 2

Molecular Characterization of Sequence-driven Peptide Glycation

Peptide glycation is an important, yet poorly understood reaction not only found in food but also in biological systems. The enormous heterogeneity of peptides and the complexity of glycation reactions impeded large-scale analysis of peptide derived glycation products and to understand both the contributing factors and how this affects the biological activity of peptides. Analyzing time-resolved Amadori product formation, we here explored site-specific glycation for 264 peptides. Intensity profiling together with in-depth computational sequence deconvolution resolved differences in peptide glycation based on microheterogeneity and revealed particularly reactive peptide collectives. These peptides feature potentially important sequence patterns that appear in several established bio- and sensory-active peptides from independent sources, which suggests that our approach serves system-wide applicability. We generated a pattern peptide map and propose that in peptide glycation the herein identified molecular checkpoints can be used as indication of sequence reactivity.

This chapter has been published Berger, M. T., Hemmler, D., Walker, A., Rychlik, M., Marshall, J. W., and Schmitt-Kopplin, P. Molecular Characterization of Sequence-Driven Peptide Glycation. *Scientific reports*, 11(1), 13294 (2021).

Candidate's contributions: M.T.B. designed the study and conducted the experiments. M.T.B. performed the bioinformatics, and statistical data analysis. M.T.B. interpreted the data. M.T.B prepared the figures, wrote, and revised the manuscript.

2.1. Introduction

Glycation presents a ubiquitous non-enzymatic post-translational modification,[273,274] which is formed by the reaction of amino compounds and reducing sugars. It refers to a complex reaction network and produces a multitude of heterogeneous reaction products, also known as Maillard reaction products (MRPs) or advanced glycation end products (AGEs).[10,181] Glycation is a multifactorial reaction, which depends on the nature of the precursors and the reaction conditions, including time and concentration.[18,31,275,276] The Maillard reaction (MR) is one of the most common and essential reactions in food processing and determines color, flavor and taste of food. Further, its reaction products are known to affect human health,[277,278] and contribute to various pathologies, such as diabetes.[8,279] Here, glycation leads to molecular and cellular changes in a series of complicated events. Hyperglycemia drives glycation of lipids and proteins and development of vascular lesions by AGE engagement of the receptor for AGE (RAGE) in cells of the vessel wall.[280-283] Due to their broad relevance, thorough understanding of glycation reactions is indispensable.

As insights into the MR of amino acids continue to emerge, new models are needed to improve the understanding of peptide and protein glycation.[284] Many previous studies analyzing the health effects of glycation products and peptides point to their miscellaneous bioactivities and their potential as nutraceuticals and functional food ingredients.[285-290] Glycation induced alterations in the bioactivity and improvement of sensory attributes have been described for various peptide mixtures.[86,114,115,291-296] However, the specific peptides related to these changes remain largely uncharacterized and, even more important, the behavior of peptides in glycation reactions has barely been systematically analyzed and thus is unaccounted. Only a limited number of studies on peptide reactivity in the MR have been conducted and focused on synthetic peptides[84,86,182] or peptide derived MRPs in specific foods.[297-300] These approaches have revealed the relevance of both peptide length and composition in the MR and the importance of peptide glycation in various fields, not only including diverse food matrices but also biological systems and disease progression.[278] The information describing the general determinant factors of peptide glycation, however, remains elusive. Therefore, particular sequences and, thus classes of proteins that have preference to undergo glycation reactions and the final consequences, such as loss and gain of bioactivities, cannot be determined. Especially required are model systems for large-scale MRP characterization that enable general understanding of the influence of the amino acid composition, sequence and peptide length on glycation product formation.

To acquire a better understanding of site-specific peptide glycation the analysis of the Amadori product (AP), a relatively stable intermediate of the MR, and consecutive downstream reaction products MRPs is particularly well-suited. High-resolution mass spectrometry (MS) is a fast and highly sensitive method, which enables detection and identification of both early and advanced products of the MR.[301,302] Information on net chemical transformations and precursor reactivity in such systems can be gained by non-targeted experiments and generation of mass difference networks.[76,303] Collision-induced dissociation (CID) after electrospray ionization (ESI) MS has been successfully applied for peptide derived AP analysis.[161,304,305] However, non-targeted large-scale analysis and interpretation of peptide derived MRPs is limited. Only a few studies have gained insight into peptide reactivity and the influence of sequence microheterogeneity.[84,99,182,306,307] This includes glycation based on accessibility of the N-terminus and catalytic effects in some short-chain peptides. Here, we report that the combination of high-resolution ESI quadrupole time of flight (QTOF) MS, bioinformatics and multivariate statistics enables a deep and molecular-level investigation of complex peptide systems. Using this combinatorial method for large-scale AP analysis, we characterized the reaction behavior of 264 casein-derived peptides in the MR and used this data to gain insight into sequence-dependent differences in AP formation profiles and, thus, peptide reactivity. Furthermore, we discovered potentially relevant glycation-patterns and demonstrate system-wide applicability of this study to various food-peptide sources by *in silico* sequence mapping. Database search serves as a reference for investigation of bioactive and sensory-active peptide reaction behavior. This approach may be amendable to practically any type of glycation system, and it allows exploration at various levels of information, from the influence of the peptide composition to the role of specific sequence-patterns in peptide glycation.

2.2. Methods

2.2.1. Preparation of model systems

D-(+)-glucose (> 99.5%) and tryptone were purchased from Sigma-Aldrich (Steinheim, Germany). Tryptone (6% (*w*/*v*)) was mixed with glucose at different concentrations (0.015 M, 0.03 M, 0.15 M, 0.3 M) in MilliQ-purified water from a Milli-Q Integral Water Purification System (18.2 MΩ, Billerica, MA, USA). Aqueous tryptone solutions (6% (*w*/*v*)) were prepared as control samples. Model systems were heated in closed glass vials for two, four, six and ten hours at 95 °C according

to the protocol recently described.[76] Control samples were heated for 10 h, analogously. Sample preparation was performed in triplicate ($n = 3$). Model systems were stored at −20 °C.

2.2.2. LC-MS/MS

Model systems were diluted 1:6 (*v*/*v*) with an aqueous solution containing 2% acetonitrile (LC-MS grade, Merck, Darmstadt, Germany) prior to LC-MS/MS analysis. Samples were analyzed by UHPLC (Acquity, Waters, Milford, MA, USA) coupled to a quadrupole time-of-flight mass spectrometer (maXis, Bruker Daltonics, Bremen, Germany). For reversed phase (RP) chromatography an ACQUITY UPLC BEH C18 column (100 × 2.1 mm, 1.7 µm, Waters, Milford, MA, USA) was used. The column temperature was maintained at 40 °C. RP separation was run in gradient mode. The RP eluent A was a composition of 0.1% (*v*/*v*) formic acid and RP eluent B was composed of acetonitrile with 0.1% (*v*/*v*) formic acid. Pre-equilibration time was set to 2.5 min. Initial conditions were set to 95% eluent A and 5% eluent B. This composition was maintained until 1.12 min. Eluent B was increased to 99.5% within 5.29 min, maintained to the end of the run. The gradient was completed after 10.01 min. Samples were injected *via* partial-loop-injection (5 µL). Mass spectra were acquired in positive electrospray ionization mode. Internal calibration was performed using a tuning mix solution (Agilent Technologies, Waldbronn, Germany) prior to each measurement. Parameters of the ESI source were: capillary voltage 4000 V, dry gas temperature 200 °C, nebulizer pressure 2 bar, and nitrogen flow rate 10 L/min. Mass spectra were recorded with an acquisition rate of 5 Hz within a mass range of $m/z = 50–1500$. For data-dependent MS/MS acquisition, the most abundant ion of a full MS scan was subjected to MS/MS after each precursor scan. The collision energy was set to 35 eV and to change dynamically and proportionally to the mass of the precursor molecule. Raw data were post-processed using Genedata Expressionist Refiner MS 13.0 (Genedata GmbH, Basel, Switzerland) applying chemical noise subtraction, intensity cutoff filter, calibration, chromatographic peak picking, and isotope clustering. Only features detected in all three replicates were retained in the matrix.

Chemical peptide structures were confirmed by peptide mapping in Genedata Expressionist Refiner MS 13.0 (Genedata GmbH, Basel, Switzerland) with an absolute m/z tolerance of 0.005 and 0.1 for the precursor and product ions, respectively, unspecific enzyme cleavages, a minimum peptide length of 1 AA, and no fixed or variable modifications. The fragmentation type was set to ESI CID/HCD. The peptide mapping was performed using a text file containing four AA

sequences in FASTA format of bovine milk caseins including α-S1-, α-S2-, β-, and κ-casein. Top-down sequencing annotations for each of the four casein proteins were exported from Refiner MS module providing a list of the identified peptides along with their positions in the protein sequence they were successfully mapped to. Further processing was performed in R software (version 3.5.2). Amadori product precursor mass was calculated by a mass increase of 162.0528 Da. Amadori product precursor signals were computationally assigned by an algorithm within a mass tolerance of ± 10 ppm. Putatively assigned Amadori products with available MS2 spectra from our data were clustered according to their similarity in normalized intensity profiles using Pearson correlation. Product ion annotation was automatically performed in R software by *in silico* fragmentation[304,308] and manually validated. Monoisotopic mass tolerance was set to ± 0.005 Da for product ions. To separate false positive assignments, we excluded signals with a poor fit of the MS/MS spectrum to the *in silico* predicted fragments and maximum intensity in the tryptone control samples heated for 10 h. Genedata Expressionist Refiner MS 13.0 (Genedata GmbH, Basel, Switzerland) peptide mapping activity provides a consolidated score, which describes the average fit for each peptide across all MS2 spectra available. Consolidated scores of all peptides, for which the corresponding Amadori product (MS^2 level) could be verified, were computed. The minimum consolidated score per peptide length was chosen as a threshold for peptide identification.

2.2.3. Statistical analysis

Pearson correlation coefficients were calculated in R software between intensity values for putatively assigned MS^2 Amadori products ($n \geq 3$, $p < 0.05$). For this analysis, relative intensity values were used. Relative intensity values were calculated by normalizing intensity values to the maximum intensity value across all time points and for each Amadori product, respectively. Hierarchical clustering to provide the domain ordering was done using R software. Amadori product wise distances were calculated based on these correlations using the as.dist() function followed by hierarchical clustering using the hclust() function. To assess the importance of small sequence variations, pairwise two-sided t-tests were performed. Intensity values in model systems and control samples were compared. Significantly increased Amadori products ($p < 0.05$) and relevant reaction conditions are reported in Table A.S1. Multiple testing correction was performed using the Benjamini-Hochberg procedure.

2.2.4. Sequence grouping

The computational analysis of sequence groups was performed with the peptide single letter code using R software. To identify peptides with common sequences the grepl() and match() base functions were applied. Based on derived sequence commonalities, we assigned all peptides to sequence groups (Table A.S1). Amino acids were not assigned to sequence groups. Sequence groups, for which no Amadori product was detected, were excluded in Table A.S1.

2.2.5. Database search

Bioactive peptide database search was carried out using the Milk Bioactive Peptide Database (March 13, 2020)[111] and the BioPepDB database (March 13, 2020)[112]. Sensory peptide database search was performed using the BIOPEP database (March 13, 2020)[309]. Database queries and substring search of tryptone peptides in bioactive peptides were conducted with the peptide sequence single letter code using R software. Hierarchical clustering to provide the bioactive peptide source ordering was done in R software.

2.3. Results

2.3.1. A time-resolved analysis of peptide dependent Amadori product formation

To study the reaction behavior of peptides in glycation, we heated complex model systems containing glucose (2.7–54 mg/mL) and tryptone at 95 °C for 2, 4, 6 and 10 h, respectively. Compared to an *in silico* tryptic protein digest, tryptone provides approximately four times more free peptides with increased diversity (Figure 2.1A). A vast number of non-enzymatic cleavage sites generates many diverse peptides (Figure A.S1) with partially overlapping amino acid sequences. This enables characterization of site-specific microheterogeneity and, ultimately, identification of specific sequence patterns that promote glycation. Apart from that, C- and N-terminal amino acids in tryptone peptides comprise a much greater diversity than enzymatic digests can cause, which becomes apparent by comparison with an *in silico* tryptic casein digest (Figure 2.1B and Figure A.S2A). Unlike tryptone, there is a bias toward lysine- (Lys) and arginine- (Arg) containing peptides for tryptic casein digestion (Figure A.S2B) and poor enzymatic cleavage of certain protein regions (Figure A.S3) may lead to the preclusion of particularly relevant peptides. Even alternative proteases or sequential digestion would only provide a marginally

increased total number and diversity of peptides compared to trypsin.[310] With the applied method, we could nearly completely cover the casein protein sequences (Figure 2.1C).

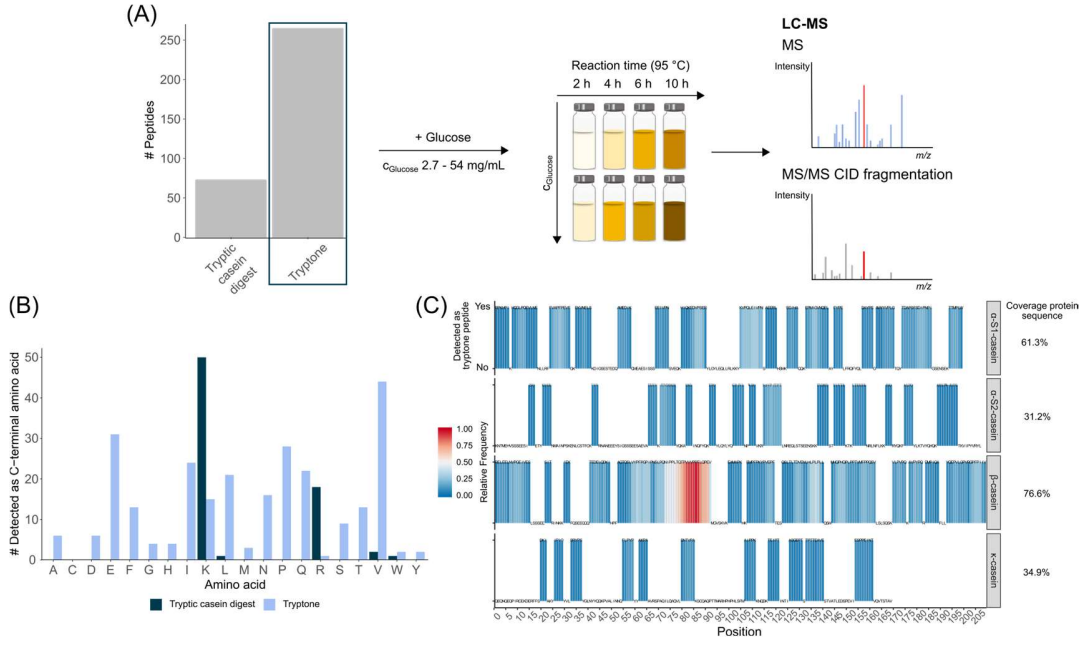


Figure 2.1: Analysis of tryptone glucose model systems by UHPLC-QTOF-MS. (A) Experimental design – high-resolution mass spectrometry was used to analyze tryptone glucose model systems and for the identification of site-specific non-enzymatic glycation. (B) The bar plot illustrates the number of peptides provided by an *in silico* tryptic casein digest (72) compared to tryptone (264). (B) A bar plot shows the number of C-terminal amino acids observed for tryptone peptides (light blue) and a theoretical casein digest by trypsin (dark blue). Tryptic digestion predominantly forms peptides with C-terminal lysine or arginine. (C) A casein protein heatmap represents how often (relative scale) detected tryptone peptides covered the same amino acid sequence in the proteins, showing which protein substructures contribute to peptide heterogeneity of the model systems. Dipeptides were removed for more sequence specificity.

With this extensive dataset in hand, we first explored the concentration- and time-dependent reaction behavior of this diverse pool of peptides based on the formation of the corresponding APs. The AP is a relatively stable intermediate of the early MR,[5] which is formed *via* condensation between the amino compound and the reducing sugar, and subsequent rearrangement11.[8] The number of hexose residues coupled to an amino acid or peptide was estimated by a mass increase of 162.0528 Da per attached monosaccharide ($C_6H_{12}O_6$-H_2O). Tandem MS was applied to obtain structural information, which allowed confirmation of the AP chemical

structure of 47 amino compounds (Table A.S1). Significant correlations ($p\text{-}value < 0.05$) were observed across normalized AP intensity profiles, and APs clustered by the influence of sugar concentration but also reaction time (Figure 2.2A). Interestingly, APs clustered that were formed from peptides with comparable amino acid sequences. Sequence similarity is highlighted by the suspended numbers indicating sequence groups, e.g. HLPLP and LHLPLP (Cluster 2, sequence group 16). Certain APs found in Cluster 1 and 2 seemed to form isomers, which caused them to also appear in Cluster 3. In Figure 2.2B, the representative normalized mean intensity profiles for the three clusters are shown, and Figure A.S4 provides the individual normalized intensity curves for each AP, separately. The intensities of APs in Cluster 1 and 2 reached their maximum after two hours for all glucose concentrations and decreased with further reaction time (Figure 2.2B). In contrast, Cluster 3 contains APs, which increased with time and reached maximum intensity after either six or ten hours of incubation. Further, enrichment of AP levels at different glucose concentrations was observed. The highest intensity of Cluster 1-APs was either detected at 5.4 or 27 mg/mL of glucose, whereas for APs in Cluster 2 (27–54 mg/mL) and 3 (54 mg/mL) higher sugar concentrations were required to reach their maximum.

Figure 2.2C compares AP formation between different peptide lengths, demonstrating that nearly all amino acids and dipeptides were found in Cluster 3, whereas larger peptides did not show uniform normalized AP intensity profiles. Moreover, the percentage observed as an AP for each peptide length is displayed, but no general correlation between peptide reactivity and sequence length could be observed. Note, this calculation is based on the total number of peptides per length, so one length may appear to be glycated to a greater extent if a minor number of peptides was identified with this length. A higher proportion of APs derived from dipeptides compared to tripeptides resembles observations from previous studies,[84,304] which suggested decreasing reactivity with increasing peptide length. Here, we confirm and extend these observations by investigating a larger range of peptide chain lengths. Interestingly, longer-chain peptide sequences were not generally associated with reduced reactivity in early glycation reactions, e.g. when comparing penta- and hexapeptides or nona- and decapeptides. Even with these observations, it is difficult to comment on the influence of peptide length on glycation on a universal level.

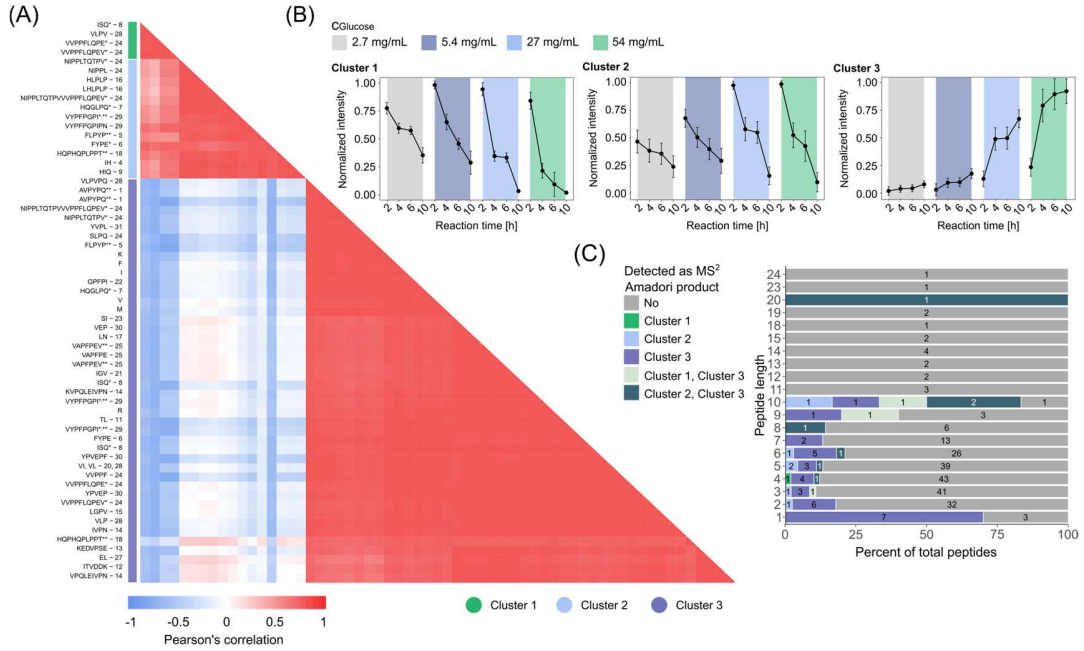


Figure 2.2: Intensity profiling enables classification of APs identified. (A) Pearson's correlation heatmap computed for the intensity profiles from the APs confirmed by MS/MS fragmentation pattern. Rows are ordered by unsupervised hierarchical clustering. Positive values (red) represent higher similarity in formation/degradation rates. Negative correlation values are colored in blue. Asterisks denote APs from the same peptide (* isomers; ** different charge states). Suspended numbers indicate peptides with similar amino acid sequences. (B) Intensity profiles visualize time-resolved AP formation depending on the glucose concentration. Intensity values were normalized towards the greatest intensity value. Colors indicate different glucose concentrations. (C) A stacked bar plot shows the percentage of the total number of peptides of a certain length that were identified as an AP (colored bars). Different colors provide information on the assigned clusters as identified in (A).

All APs significantly increased after 2 h (t-test, *p-value* < 0.05); however, different glucose levels were required (Table A.S1). Most (25) of the identified APs significantly increased at a glucose concentration of 2.7 mg/mL, while other peptides required higher glucose levels. Interestingly, different observations were made for highly similar peptides. For example, VPQLEIVPN required a glucose concentration of 27 mg/mL to increase, significantly, whereas for KVPQLEIVPN only 5.4 mg/mL of glucose were needed. Analogous behavior was observed for the APs of VAPFPE (27 mg/mL) and VAPFPEV (5.4 mg/mL).

2.3.2. Shedding light onto the role of peptide composition in glycation

AP analysis uniquely facilitates characterization of site-specific glycation, and our dataset provides insight into the highly complex and yet largely unknown reaction behavior of peptides in glycation. Previous studies explored peptide derived MRPs in particular foods providing limited information from a global prospect.[297-300] Others have investigated the reactivity of highly specific synthetic peptides depending on factors explored herein to some degree, such as peptide length (discussed above) and amino acid composition.[84,182] Here, we aspired to approach these research questions from a general level using a large reservoir of peptides and APs. Mapping glycated peptides onto casein proteins revealed that AP formation was observed for only 14% of α-S2- and 9% of κ-casein peptides (Figure A.S5), but APs related to α-S1-casein (29%) and ß-casein (45%) were detected to a larger extent (Figure 2.3A). Analyzing the amino acid sequence of α-S2- and κ-casein peptide APs, we identified protein-specific peptides such as FLPYP (F_{55}-P_{59} of κ-casein). The ability to profile glycosites at this scale provides opportunities to determine the relative susceptibility of peptide collectives with similar amino acid sequences to the early MR. Especially reactive peptide classes are captured here, as the protein heatmaps show the frequency that sequences co-occurred on glycated peptides. AP formation of peptides sharing certain amino acid sequences appeared to be favored, e.g. including peptides that originated from N_{73}–V_{92} and V_{170}–V_{173} of β-casein.

To examine the influence of the amino acid composition on peptide reactivity in the early phase of glycation, we calculated the contribution of each amino acid to AP formation, given as a percentage of the total observations in tryptone peptides. Figure 2.3B shows amino acids, e.g. glutamic acid (Glu) and leucine (Leu), that appeared equally in peptide APs from α-S1- (top) and β-casein (bottom). Other amino acids, such as proline (Pro) and histidine (His) showed considerable variations in their contribution to AP formation (Figure A.S6) depending on the source protein, meaning the overall peptide sequence, which fits with the known role of the microenvironment of amino acids in glycation. Importantly, a previous report described varying *in vivo* reactivity of lysine depending on its position in the albumin sequence and, thus, its neighboring amino acids.[306] Further, investigation of short-chain peptide model systems showed that AP formation is considerably influenced by the immediate chemical environment, hence, adjacent amino acids side chains in the peptide sequence,[84,86] which overall indicates that we may also have identified reactivity-sequence interrelation for peptide structures. We visualized the

median percentage of each amino acid in APs to explore the effect of amino acid composition and microheterogeneity over all four casein proteins (Figure 2.3C). Most amino acids showed a wide distribution of the values, again demonstrating that the type of amino acids that contribute to AP forming peptides can vary based on their immediate chemical environment. This presents a promising starting point to explore for sequence-specific glycation.

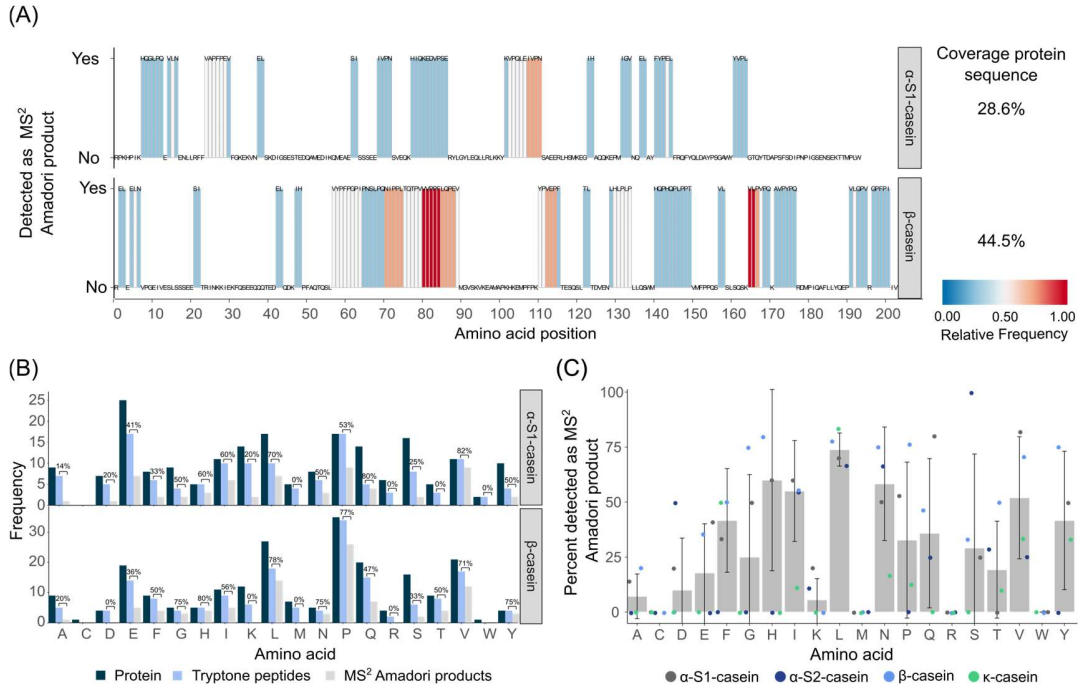


Figure 2.3: Deconvolution of glycated protein sequences reveals their compositional characteristics. (A) Contribution of protein regions and the amino acid microenvironment to peptide AP formation. Casein protein heatmaps represent the relative frequency of amino acid positions that were detected as an AP, indicating which peptides from specific protein regions contribute most to early MR and how the amino acid microenvironment influences peptide reaction behavior. Approximately 28.6% of α-S1- and 44.5% of β-casein were detected as the corresponding MS^2 AP, and the majority contains redundant sequence motifs (i.e., P-E-V, see Figure 2.4). (B) Abundance of amino acids in casein proteins (dark blue), tryptone peptides (light blue) and APs (gray). Embedded values indicate the percentage of peptides that could also be detected as an AP. (C) The bar plot depicts the median percentage of the amino acid found in APs in the different casein proteins. Overlaid points indicate the percentage for each studied casein protein. The whiskers represent the standard deviation.

To dive into this intriguing facet of peptide glycation, we examined the location of amino acids relative to the reactive peptide N-termini. This was based on de Kok's hypothesis that the

side chain carboxylic group of Glu catalyzes glycation of primary amino groups,[84] and on the suggestion of Zhili and co-workers that Leu and isoleucine (Ile) promote AP formation.[86] As short-chain peptides were investigated in these studies, they describe the influence of directly neighboring amino acid side chains on N-terminal peptide glycation. Hence, we reasoned that our dataset could allow to explore the impact of both the N-terminal amino acid and the adjacent amino acid side chain across a large number of highly diverse peptide species. To detect amino acid overrepresentation at the mentioned positions, we generated sequence logos by comparing sequences between peptide APs and non-glycated peptides (Figure 2.4A and Figure A.S7A). Here, amino acids enriched at certain positions in AP forming peptides are illustrated (relative $\text{abundance}_{\text{glycated}} - \text{relative abundance}_{\text{non-glycated}} > 0$). This analysis indicated preference for valine (Val), Ile and Leu at the first two positions of the amino acid sequence for glycated peptides (Figure 2.4A). Indeed, the percentage of N-terminal Val was considerably increased for AP forming peptides compared to peptides, for which the corresponding AP could not be identified (Figure 2.4B, top). An illustration of the absolute amino acid abundance can be found in Figure A.S7B. Interestingly, we further found substantially higher relative frequencies for Ile, Leu and Val next to the N-terminus in peptides also observed as an AP (Figure 2.4B, bottom), echoing the result from the sequence logo. To account for preferred glycation of peptides with Ile, Leu or Val at the second sequence position, their summed relative frequency depending on AP detectability is shown in Figure 2.4C. This result complies with previous findings that hypothesized that Ile and Leu promote N-terminal glycation based on pronounced hydrophobicity[311] (Ile 1.80; Leu 1.70) and polarizability[312,313] (Ile 91.21; Leu 91.60). Val has comparable physicochemical properties (hydrophobicity 1.22, polarizability 76.09) to Ile and Leu and was previously shown to exert a similar effect on the reactivity of the peptide N-terminus.[86]

Furthermore, the percentage of aspartic acid (Asp), methionine (Met), phenylalanine (Phe) and Pro next to the N-terminus was considerably decreased in glycated peptides (Figure 2.4B, top). We also observed that glycation of peptides was disfavored for Pro at the first two sequence positions (Figure 2.4B and Figure A.S7A). A recent study on Pro containing dipeptides (Gly-Pro, Pro-Gly) suggested that its secondary amine may hinder Schiff base formation.[314] Nevertheless, we found that proline was frequently observed at the third and fifth sequence position of glycated peptides (Figure 2.4A), thus raising the possibility of its involvement in increased peptide reactivity towards early glycation. This is supported by an increased relative

abundance of proline at the same locations relative to the glycation site in AP forming peptides compared to peptides, for which the corresponding AP was not detected (Figure A.S7C and A.S7D).

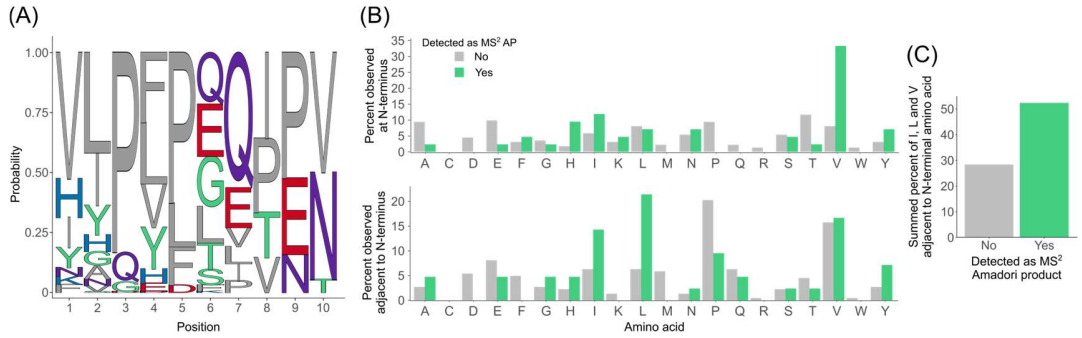


Figure 2.4: Analysis of the amino acid location relative to the glycation site reveals sequence-resolved changes in peptide reaction behavior. (A) Sequence logo representation of the first ten amino acid sequence positions of peptide APs. Amino acids with increased relative abundance in glycated peptides are illustrated (relative abundance$_{glycated}$ – relative abundance$_{non\text{-}glycated}$ > 0). (B) Comparison of amino acids at the N-terminal position and adjacent to the peptide N-terminus. Bars show the percentage of (glycated) peptides that contain a given amino acid at the first (top) and second position (bottom) of the amino acid sequence (detected as AP: green bars; not detected as AP: gray bars). (C) Summed percentage of peptides that contained isoleucine, leucine, and valine adjacent to the N-terminal amino acid.

2.3.3. Capturing relevant sequence patterns in peptide glycation

Large-scale peptide derived AP analysis enables us to identify potentially relevant glycation-patterns, and our dataset can provide an initial glimpse into this intriguing aspect of glycation. While others have explored the influence of the amino acid sequence based on di- and tripeptide glucose model systems,[84,86] we can now comment on trends across 264 peptides originating from four proteins. In enzymatic glycosylation the importance of the N-X-S/T sequon and the negative effect of Pro in X position has been shown, which may result from conformational changes.[315] However, relevant structural motifs in peptide glycation have not been identified. In this detailed analysis, we identified small regions of identical subsequences in casein proteins and, thus, the thereof arising peptides (length = 2, 3, and 4) using the amino acid one letter code (Figure 2.5A, top). We mapped (non-) glycated peptides onto proteins (Figure 2.5A, bottom right part) and across each other (Figure 2.5A, bottom left part). This provided information on the degree of co-occurrence for sequence patterns on glycated peptides with a different overall amino acid

sequence. This protocol allows to detect relevant glycation patterns that are anticipated to be important factors in determining the preference for early peptide glycation.

Analysis of common sequence patterns in casein proteins revealed several sequence overlays. Figure 2.5B provides information on the degree of sequence similarity, from which it is evident that the number of shared sequences varied for different pattern lengths and pairs of proteins. No shared tetra-sequences were found for κ-casein. To explore sequence patterns of maximum length, tri-patterns were chosen for further investigation. Figure 2.5C captures the total frequency of tri-sequence patterns in the casein protein sequences and which percentage of these differentially located subsequences was covered by peptide APs. This analysis allows to identify how different patterns contribute to glycation of peptides with shared subsequences but different overall amino acid composition as they originate from different casein protein regions. Of the three P-E-V sequons in casein proteins (P_{44}–V_{46} of α-S1-casein and P_{105}–V_{107} of ß-casein, Figure 2.3A; P_{171}–V_{173} of κ-casein, Figure A.S5), two appear as subsequences in peptide APs, as well as the alternated V-E-P (V_{131}–P_{133} of ß-casein, Figure 2.3A). Several interesting cases where substructures highly similar to P-E-V contribute to AP formation are highlighted (P-E-L, V-L-N, V-P-N, and V-P-Q; Figure 2.5C). These subsequences all share amino acids with a low dissimilarity score (*D*), which is based on 134 categories of activity and structure,[316] and are partially rearranged (Table A.S2). For example, P-E-V and P-E-L only differ by a single amino acid with strong physicochemical similarity (*D*(Val, Leu) = 9), whereas in case of P-E-V and V-P-Q (*D*(Glu, Gln) = 14) also the sequence order was changed. By comparison, peptides that contain I-V-E, which shows more pronounced sequence variations (*D*(Pro, Ile) = 24 and sequence rearrangement), do not participate in AP formation. All of these patterns, which show strong contribution to AP formation, either comprise Glu (carboxylic group), glutamine (acid amide group) or asparagine (acid amide group). A catalytic effect of the carboxylic group on AP formation was previously hypothesized,[84] which resembles the here found promoting effect of Glu-containing sequence patterns on the early MR. P-Y-P and P-F-P contain amino acids with comparable properties. The substructures P-I-P, P-L-P and P-V-P feature pronounced physicochemical similarities as well.[316] All these patterns showed a high co-occurrence on peptide APs relative to their total abundance (displayed as percent in Figure 2.5C), which indicates their contribution to glycation independent of the overall peptide composition, thus their origin in the

source protein . In contrast, P-N-P (P_{198}–V_{200} of α-S1-casein) did not contribute to an AP (Figure 2.3A) and shows pronounced differences in its amino acid characteristics.

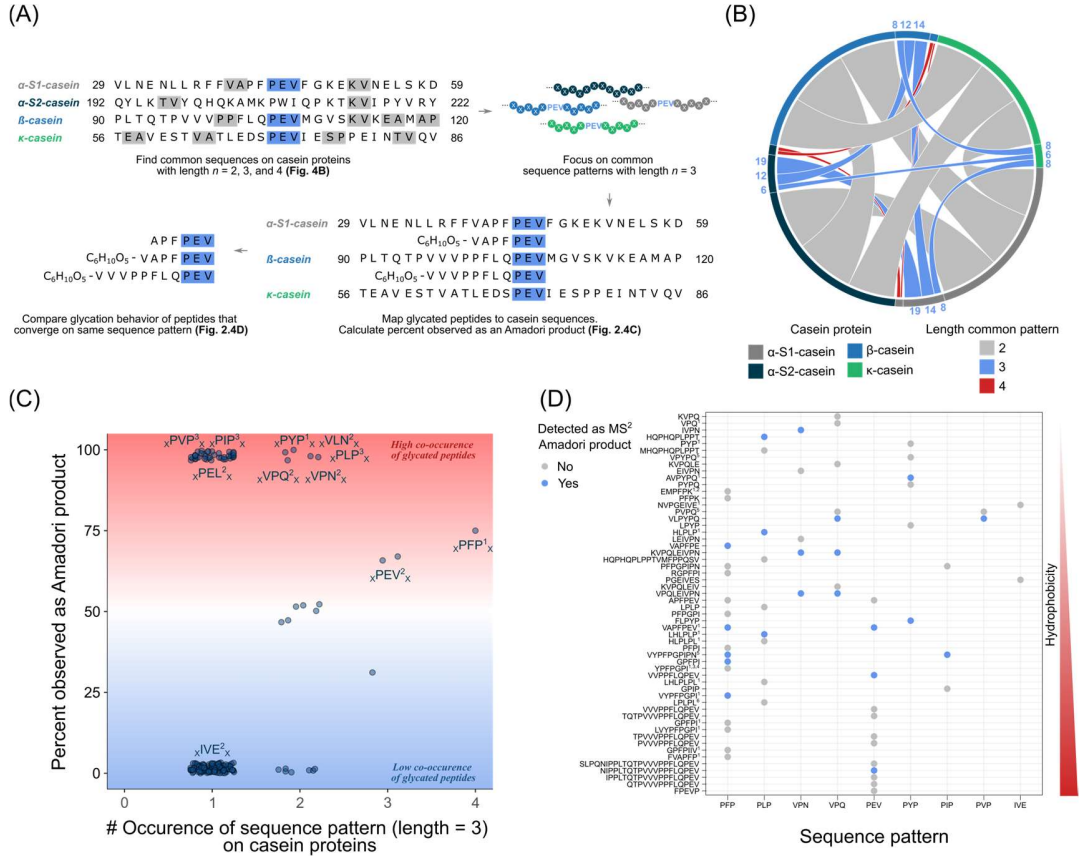


Figure 2.5: Mapping sequence patterns to protein and peptide domains exemplifies their role in peptide glycation. (A) Schematic illustration of sequence crosstalk decoding: Effect of sequence similarity on protein and peptide glycation based on presence of short-chain amino acid patterns. (B) A chord diagram representation of overlaying sequence patterns between casein proteins. The number of common di-, tri-, and tetra-sequence patterns was computed between each casein protein sequence. The size of the connections between the proteins (arcs) is relative to the number of common sequence patterns. (C) A scatterplot showing the percentage of sequence patterns with three amino acids in casein proteins covered by APs. Suspended numbers indicate sequence pattern dissimilarities as given in Table A.S2. (D) Mapping sequence patterns to peptide domains. Peptides detected as an AP are colored in blue. Peptides that were not observed as the corresponding AP are shown in gray. Peptides are ordered by hydrophobicity according to their chromatographic retention time. The numbers denote established bioactive peptides ([1] antihypertensive, [2] antimicrobial, [3] opioid, [4] immunomodulatory, [5] antioxidant, [6] DPP-IV-inhibitory).

A peptide-sequence pattern plot in Figure 2.5D maps relevant sequence patterns to different (glyco-) peptides for which they could be observed. These peptides vary in their overall composition and peptide length. The map indicates, which patterns contribute to glycation on a peptide-level, and other peptide properties that considerably affect their reaction behavior. This analysis reveals several interesting trends, such as pronounced discrepancies in glycation of peptides with the same pattern and, perhaps most striking differences in AP formation of strongly related peptides. Small variations in the peptide amino acid sequence may cause (VVPPFLQPEV vs. VVVPPFLQPEV; YPFPGPI vs. VYPFPGPI) or not cause (VAPFPE vs. VAPFPEV; VYPFPGPI vs. VYPFPGPIN) differentiated behavior in the early MR. General correlation was not observed between AP formation and peptide physicochemical properties, expressed as hydrophobicity according to their retention time.

2.3.4. System-wide analysis of bioactive and sensory active peptide glycation enabled by *in silico* sequence mapping

The complexity of glycation represents a great challenge for the identification of glycation patterns that are associated with the gain or loss of bioactivity and glycation induced changes in sensory attributes. A combination of bioinformatics and database search enabled to study the established sensory and bioactivities of peptides in our dataset and to evaluate their behavior in glycation. We matched 60 peptides (Table A.S3) with reported bioactivities (Figure 2.6A), which were included in databases compiled from literature.[111,112] While 36 peptides were exclusively found in milk, 24 peptides appeared in a variety of other food sources as well (Figure 2.6B). We also found that these peptides of diverse chain lengths (Figure 2.6C) cover various bioactivity categories (Figure A.S8), suggesting that tryptone models may facilitate inter-disciplinary investigation of peptide glycation. In our study, for 62% of the bioactive peptides the corresponding AP was detected (33% confirmed by MS/MS tandem experiment; Figure 2.6D). Hence, approximately 34% of the 47 peptides, for which the AP was identified *via* MS^2 fragmentation, were previously reported to be bioactive.

Furthermore, we identified 25 peptides with sensory attributes.[309] Figure 2.6E illustrates the prevalence of AP detection for sensory-active peptides and the level of AP identification (MS and MS^2). Approximately 24% of all sensory-active peptides and 19% of the bitter peptides were observed as the corresponding AP on MS^2 level.

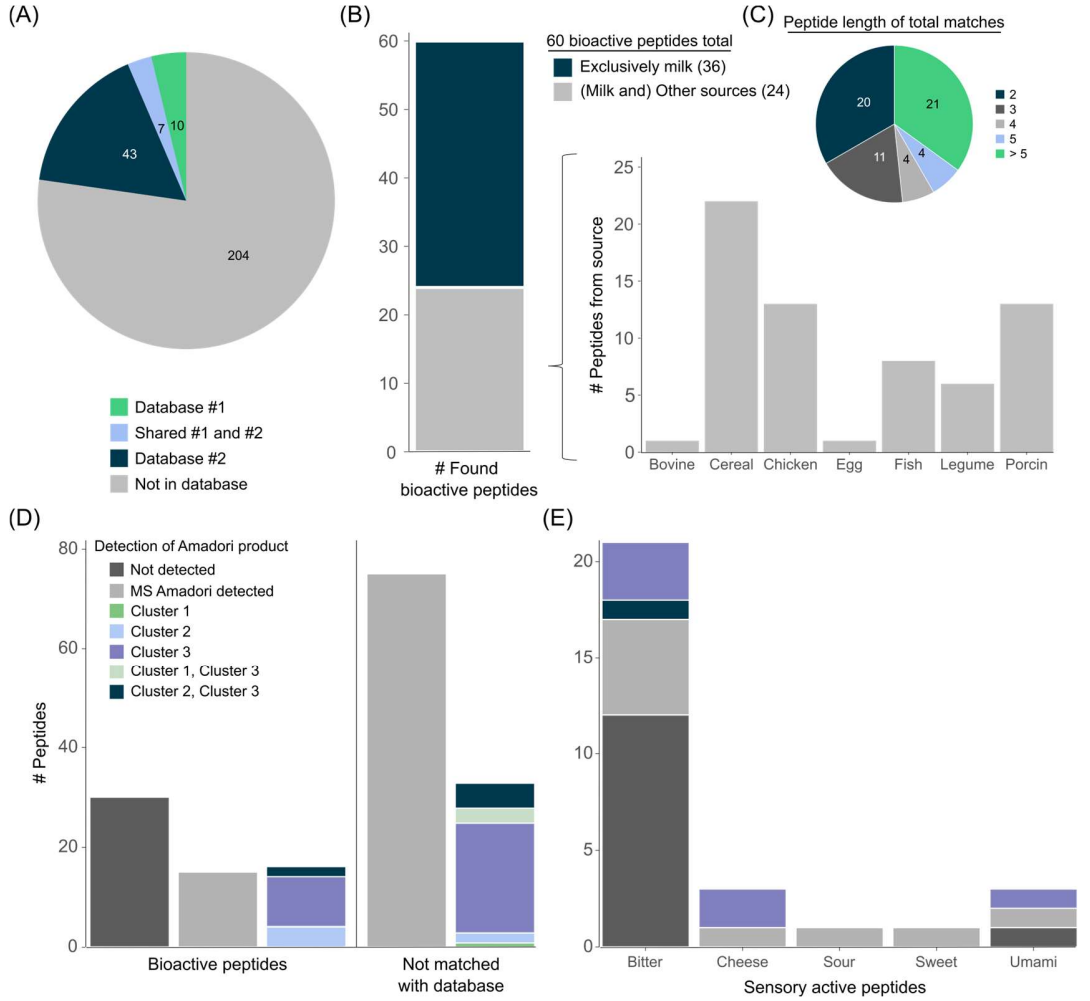


Figure 2.6: Database search uncovers overlap with miscellaneous peptide sources. (A) Identification of 60 peptides reported as bioactive peptides in databases for milk and other foods. (B) Classification of the assigned bioactive peptides according to different sources. Of these bioactive peptides, 36 are exclusively described as bioactive milk peptides (dark blue), and 24 are known bioactive peptides in other sources (gray). A bar plot represents the number of times peptides were matched with other sources than milk. Note, peptides that were found in multiple sources were counted for each source, individually. (C) Distribution of peptide lengths of those peptides which could be assigned to an established bioactivity. (D) The bar chart displays the number of bioactive peptides (left), and the number of peptides not matched with a database (right). Peptides were grouped into those, for which no AP (dark gray), an AP solely on precursor mass ± 10 ppm (light gray) or an AP also confirmed by MS/MS fragmentation pattern could be detected. APs with MS/MS fragmentation certainty were further classified into Cluster 1, Cluster 2, Cluster 3, Cluster 1 and 3 or Cluster 2 and 3 (see Figure 2.2). (E) The bar graph depicts AP formation for 25 peptides with different sensory activities. If multiple sensory activities were reported for a peptide, it was counted more than once for this calculation.

As noted by Dong and co-workers,[296] bitterness of MRPs was decreased compared to original casein peptides, and further reduced with heating time and glucose concentration. Lan *et al*. previously published that bitter soybean peptides below 1000 Da decreased 28.49% after reaction with xylose at 120 °C.[39] In contrast to the reported experiment, digesting casein with trypsin would produce large peptides, through protein cleavage after arginine and lysine, which would lead to the loss of highly interesting bioactive peptides (Figure A.S9) and would not allow comprehensive investigation of their reaction behavior (Figure A.S10 and Figure A.S11). For example, using a tryptic casein digest, the analysis of the opioid peptides, e.g. YPFPGPI and YPVEPF, for which N-glycation is known to have major consequences on the bioactivity of the parent peptide,[317,318] would not be possible. By comparison, in our experiment the large and diverse peptide spectrum of tryptone enabled us to widely predict bioactive peptide glycation.

2.3.5. Convergence of tryptone peptides and peptides with established activities into common sequences

Given that similarities in the amino acid sequence and sequence patterns may determine the reaction behavior of peptides, we reasoned our approach could provide insight into peptide reactivity from a systems level. To test this hypothesis, we searched peptides from our dataset as a pattern of bioactive and sensory-active peptide sequences reported in databases by substring matching.[111,112,309] First, we examined the number of matches observed. In total, tryptone peptides were successfully mapped to 1172 unique bioactive peptide species (Figure A.S12). Even though caseins are major proteins in milk, a considerable number of common sequences between tryptone peptides and a plethora of peptides from other sources was found (Figure 2.7A). Overall, we achieved 3046 sequence overlays with 675 bioactive milk peptides, 1350 overlays with 510 peptides from other sources, and 599 overlays with 174 sensory-active peptides (Figure A.S13).

Importantly, not only small sequence commonalities were found as indicated by the length of the tryptone peptides mapped (Figure 2.7A). Tryptone amino acid sequences, up to a length of thirteen amino acids, occurred on bioactive peptides from entirely different origins. Even larger sequences ($n = 14$) co-occurred in peptides with established sensory activity. The heatmap (Figure 2.7A) visualizes classes of peptides delineated by the number of overlays per peptide length, showing increased sequence similarities for sensory-active peptides and fish. Or, in case of potato peptides we found a lower proportion of common di- and tri-sequences compared to other sources and no matches for larger peptides ($n > 3$) were observed. We note that calculating

co-occurring sequences is straightforward and may provide information about glycation susceptibility of specific peptide classes from various protein origins.

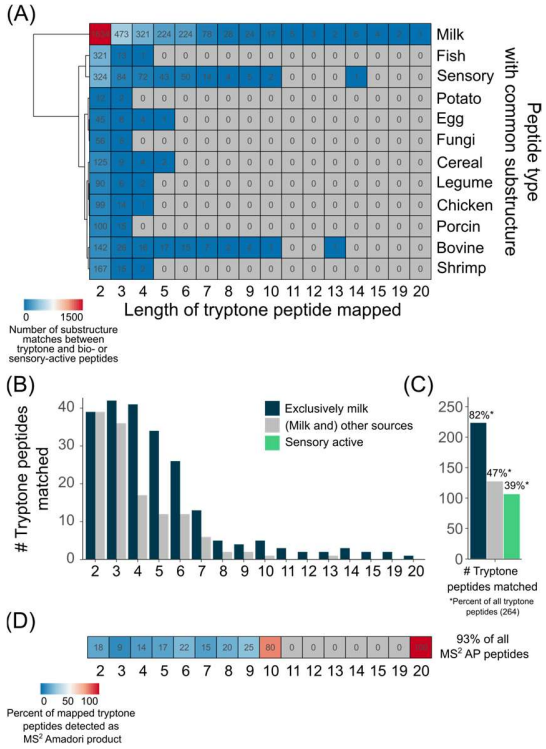


Figure 2.7: Mapping of tryptone peptides to established bioactive peptides indicates a wide scope of application. (A) Tryptone peptides were mapped on sensory-active peptides and bioactive peptides from 11 sources by substring matching to explore similarity in their amino acid sequence. A sequence co-occurrence heatmap indicates the number of sequence overlaps between tryptone peptides of a certain length and bio-/sensory-active peptides, with colors indicating the number of matches. Gray denotes zero sequence overlays. (B) Absolute number of tryptone peptides successfully mapped to bioactive peptides from milk (dark blue) or other sources (gray) depending on the peptide length. (C) Total number of tryptone peptides identified as subsequence of bioactive peptides in milk (dark blue, 82% of all tryptone peptides) or in other sources (gray, 47%) and of sensory-active peptides (green, 39%). (D) Heatmap illustrating the percentage of the matched tryptone peptides that were detected as AP. Approximately 95% of all AP forming peptides were identified as a subsequence of bioactive peptides.

Figure A.S14 demonstrates that a broad variety of bioactivities is covered by the database peptides, to which tryptone peptides were successfully mapped. Other trends arise, such as the presence of a relatively high abundance of antihypertensive peptides. Note, this is expected because of the inordinate number of antihypertensive peptides in the databases used. The same

number of dipeptides was successfully mapped to bioactive peptides independent of the source (Figure 2.7B), and the number of tripeptides found as a subsequence was comparable as well. For larger tryptone peptides, approximately half of the peptides matched for milk were found to overlay with peptides, which were (also) found in other food sources. A striking feature of this analysis are the percentages of peptides in our dataset found as a subsequence (Figure 2.7C). Considering the 264 peptides that could be mapped to a domain, 82% existed in bioactive peptides exclusively from milk and 47% in those from (milk and) other sources. Furthermore, 39% of tryptone peptides were successfully mapped to sensory-active peptides (Figure 2.7C). Figure 2.7D depicts the percentage of matched tryptone peptides, for which the corresponding AP was identified. Up to 22% of the shorter substring peptides ($n \leq 6$ amino acids) were detected as an AP by tandem MS experiments, while the majority of larger peptides appeared to be glycated. In total, 93% of the peptides detected as an AP are substructures of bioactive peptides. This represents a considerably larger proportion compared to other peptide groups, e.g. sensory-active peptides (Figure 2.7D).

2.4. Discussion

In all, here we present a straightforward approach to refine evaluation of peptide derived APs by using the power of high-resolution MS in combination with multivariate statistics and bioinformatics to access large-scale information about peptide reactivity in the MR and the influence of both reaction time and sugar concentration. Investigation of glucose-tryptone model systems enabled the most in-depth profiling of peptide APs to date. By comparison with an *in silico* tryptic casein-digest, we demonstrated considerable advantages of tryptone models, such as a notably larger coverage of (bioactive) peptides from various food sources. This strategy is amenable to virtually any type of MR model system or reactivity study with known reaction intermediates. Finally, the reaction behavior of 264 casein derived peptides was characterized by AP analysis from a single type of model system, which demonstrates that new models must be developed to unravel the glycation reaction network in its full complexity. Clearly, large-scale studies are needed to explore peptide glycation and its importance particularly in food but also biological systems and, thus, health.

A caveat of practically any MS-based experiment is that detectability can be affected by the type of ionization, analyte concentration as well as sample complexity. Thus, there may be a

bias toward specific peptides and APs to consider in this dataset. Furthermore, Figure 2.1C shows a high frequency of tryptone peptides from certain protein regions, which may arise from its production process and above-mentioned detectability issues. Our data interpretation, however, reflects on ubiquitous observations in the overall dataset, and not on specific peptide species. Despite this, we achieve in-depth characterization of a large reservoir of peptides and provide thorough information on peptide properties influencing AP formation. We relied on normalized AP intensity profiles for reaction behavior investigation, meaning there are limitations in the stability of this early reaction intermediate to consider. Greifenhagen *et al*. has noted pronounced susceptibility of the N-terminally acetylated Amadori peptide Ac-Ala-Lys-Ala-Ser-Ala-Ser-Phe-Leu-NH2 toward oxidative degradation in aqueous model systems.[319] Loss of the Amadori compound of the endogenous opioid pentapeptide leucine-enkephaline (Tyr-Gly-Gly-Phe-Leu) was also noticed by Jakas and Horvat,[320] however, markedly slower degradation behavior was observed in this study. Different reaction conditions, such as concentration of catalytically active phosphate buffer and temperature, tend to affect AP stability, but also the amino acid sequence of the peptides. Even with this, all peptide APs detected in this dataset remained above the limit of detection at all time points.

While other studies have investigated glycation using a limited number of highly specific synthetic peptides, we could simultaneously study the reaction behavior of a large pool of casein peptides. We can also see similar trends to previous studies, such as the influence of both reaction time and sugar concentration.[321,322] In contrast, we can provide detailed information on how APs derived from specific peptide sequences are affected. We show that upon reaction time the bulk of peptides differentiated into three clusters, reaching maximum AP intensity at different time points and, thus that AP formation rates likely depend on the peptide structure. Peptides forming APs that peaked at early reaction time points and low glucose concentrations may represent more reactive precursors in glycation. A consecutive decrease in relative AP peak intensity may be attributed to further rearrangement and oxidative cleave reactions yielding heterogeneous AGEs.[323] Further, we show that there is no general correlation between peptide length and reactivity. More pronounced susceptibility of dipeptides compared to tripeptides toward glycation was seen in early studies of glycine (Gly) peptide model systems (GlyGly > GlyGlyGly) and in more miscellaneous synthetic peptide studies.[84,86,182] Similarly, we found higher proportions of glycated dipeptides than tripeptides. Although our observations are in congruence

with previous studies, we are the first to investigate the impact of peptide length at this scale providing a new perspective on its influence on reactivity. We also show that there is not a general correlation between amino acid content and susceptibility towards the initial phase of glycation reactions. This suggests a strong contribution of other factors such as the amino acid sequence, thus, the amino acid microenvironment.

In peptide glycation, reaction behavior has been proposed to be driven by the amino acid sequence. An important role of the amino acid adjacent to the N-terminus has been suggested based on short-chain peptide model systems.[84,86,297] Here, we noticed strong preference of glycation for valine-starting peptides and noted more pronounced AP formation of peptides, which contain Ile, Leu and Val positioned next to the N-terminal amino acid. Further, we observed prevalence of Met, Phe, and especially Pro at the second sequence position in peptides, for which the corresponding AP could not be identified, which indicates that steric hindrance or conformational changes may prevent N-terminal glycation. Congruent observations were made for the N-X-S/T sequon, where Pro in the X position causes pronounced changes in conformation and, thus, prevention of enzymatic glycosylation.[315] We show that neighboring Glu, for example, may not always exert a catalytic effect on N-terminal glycation as a result of its carboxylic side chain. Interestingly, APs from peptides containing Asp adjacent to the N-terminus were not observed, even though its structure is closely related to Glu. Depending on the amino acid sequence and the peptide N-terminus, differentiated effects of the neighboring amino acid may be observed,[84] and here, we can cover a broad range of diverse peptide properties. Such thorough and integrated characterization of peptide APs depending on the reaction conditions is necessary for a complete understanding of peptide glycation and its impact on food and biological systems. We further found two location sites near the N-terminus with increased relative abundance of Pro in AP forming peptides, namely the third and fifth position of the amino acid sequence.

We identified peptide collectives particularly prone to early glycation reactions by mapping APs to casein sequences and across each other. Furthermore, we leverage these reactive peptide species to provide information on potentially important tri-sequence patterns and propose that glycation patterns among many other factors promote peptide glycation, which is indicated by strong connectivity in glycation susceptibility and presence of specific sequence patterns. Even though N-terminal proline may inhibit Schiff base formation,[314] here, we established several proline-rich sequence patterns, which considerably triggered AP formation.

Furthermore, Glu containing sequence patterns, such as P-E-V, may exert a catalytic effect towards early MR.[84] Depending on whether glycation is desired or not, peptides may be chosen, accordingly (discussed below). Overall, we provide a comprehensive set of molecular checkpoints for peptide reactivity estimation towards glycation.

Finally, we established system-wide applicability of tryptone model systems by mapping tryptone peptides to a plethora of bioactive and sensory-active peptides from various food sources. Depending on whether glycation of peptides is desired or not, we suggest that the amino acid sequences may be chosen, accordingly. For development of functional foods, health benefits must be preserved, thus for most bioactive peptides AP formation needs to be obviated. For example, for opioid peptides, the requirement of free N-terminal tyrosine was demonstrated and the loss of antihypertensive properties of casein peptides as a result of glycation was revealed in various model systems.[114,115,324] Conversely, if increased antioxidant potency must be achieved,[114,325] we suggest that the peptide species may be capitalized that are more prone to AP formation. The increased antioxidative properties of MRPs compared to their corresponding casein peptides has previously been determined,[115,293,294] and consequently targeted peptide glycation may enable to enhance the health benefits of peptides. Superior antioxidative properties have been established for MRPs derived from small peptides compared to larger peptide species,[325] which makes tryptone particularly suitable for identification of potential peptide candidates. In total, we observed APs from 47 amino compounds. For 34% of them, bioactivity was previously established and 93% were identified as a substructure of bioactive peptides, which suggests that bioactive peptides are particularly prone to glycation. For sensory-active peptides, others have observed reduced bitterness of MRPs compared to heated casein peptides alone, while antioxidative properties where increased.[86] Similar findings were reported for peptides from other sources.[39,127,326] Even more benefits for food production can thus be provided by choosing appropriate peptide candidates, such as enhanced sensory attributes of foods. To assess desired flavor improvement,[326,327] we reasoned that selection of peptides susceptible to early glycation may be promising. As bitter peptides cannot be employed above their bitter flavor threshold,[327] increased bioactivity accompanied concurrently by decreased bitterness may be desirable for the production of functional foods to improve health and enhance customer acceptance.[290] Taken together, our dataset allows to select suitable peptide candidates, given (1) a checklist for estimation of their reaction behavior in early glycation reactions according to the amino acid at

the N-terminus, the adjacent sequence position and presence of relevant sequence patterns, and (2) screening for established sensory attributes and bioactivity.

Future studies are required to investigate a wider range of peptides from different proteins and, thus a broader variety of amino acid sequences to gain more global information on the relevance of amino acid composition and sequence patterns. Model systems prepared from highly specific synthetic peptides have provided valuable findings in previous studies, which suggests that targeted investigation of peptides, in particular with potentially relevant sequence patterns, may be promising for identification of peptides especially prone to peptide glycation. These strategies also present an opportunity for determination of peptides less susceptible toward glycation reactions. Investigation of changes in sensory attributes and bioactivity as a result of glycation will be a worthwhile endeavor in future experiments to gain the necessary refined information for systematic use of specific peptides and their glycation products as functional food ingredients.

Chapter 3

Open Search of Peptide Glycation Products from Tandem Mass Spectra

Identification of chemically modified peptides in mass spectrometry (MS)-based glycation studies is a crucial yet challenging task. There is a need to establish a mode for matching tandem mass spectrometry (MS/MS) data, allowing for both known and unknown peptide glycation modifications. We present an open search approach that uses classic and modified peptide fragment ions. The latter are shifted by the mass delta of the modification. Both provide key structural information that can be used to assess the peptide core structure of the glycation product. We also leverage redundant neutral losses from the modification side chain, introducing a third ion class for matching referred to as characteristic fragment ions. We demonstrate that peptide glycation product MS/MS spectra contain multidimensional information and that most often, more than half of the spectral information is ignored if no attempt is made to use a multi-step matching algorithm. Compared to regular and/or modified peptide ion matching, our triple-ion strategy significantly increased the median interpretable fraction of the glycation product MS/MS spectra. For reference, we apply our approach for Amadori product characterization and identify all established diagnostic ions automatically. We further show how this method effectively applies the open search concept and allows for optimized elucidation of unknown structures by presenting two *hitherto* undescribed peptide glycation modifications with a delta mass of 102.0311 and 268.1768 Da. We characterize their fragmentation signature by integration with isotopically labeled glycation products, which provides high validity for non-targeted structure identification.

This chapter has been published as Berger, M. T., Hemmler, D., Diederich, P., Rychlik, M., Marshall, J. W., and Schmitt-Kopplin, P. Open Search of Peptide Glycation Products from Tandem Mass Spectra. *Analytical chemistry*, *94*(15), 5953–5961 (2022).

Candidate's contributions: M.T.B. designed the study and performed the experiments. M.T.B. developed and established the algorithm and analyzed the data. M.T.B. interpreted the data and prepared the figures. M.T.B. wrote and revised the manuscript.

3.1 Introduction

Non-enzymatic glycation of amino compounds, also known as the Maillard reaction (MR), is a hallmark of food quality, metabolic stress, disease, and aging.[8,328-330] After Amadori product formation by spontaneous attachment of a reducing sugar to amino or guanidino groups and subsequent rearrangement, a series of loosely understood condensation and rearrangement steps lead to a diverse set of modifications known as advanced glycation end products (AGEs).[5,181] The reaction conditions including the nature of the amino compound reactant greatly impact the type of reactions, intermediate and end products of glycation reactions.[76,331] This makes the MR certainly one of the most complex reaction networks producing a multitude of reaction products only from a few initial precursors. Due to its high specificity, sensitivity, and speed, mass spectrometry (MS) has been a cornerstone of glycation product analysis for several decades.[332-335] Many methods have been developed for the analysis of glycation products.[163-165] High-resolution MS techniques enable simultaneous detection and identification of both early and advanced glycation products even in complex mixtures.[336-338] This makes high-resolution MS particularly suitable for non-targeted glycation studies.[297] The highly accurate precursor mass and fragmentation behavior are powerful features for structure elucidation and diagnostic fragment identification.[133,304,308,339] Challenges in non-targeted analysis including the lack of commercial standards and the diversity of structures can be tackled using isotope labeling.[273,340,341] Yet, peptide glycation product tandem mass (MS/MS) spectra contain multidimensional structural information, and manual interpretation of MS/MS spectra is time-consuming and can be cumbersome. Peptides harboring the same type of chemical modification do not show common absolute fragments and are eluted at different retention times. This even impedes the integration and structure alignment of well-known glycation products across peptides. It is also important to consider that glycation reaction product mixtures are complex, which makes comprehensive structure determination a challenging task. Even given substantial improvements in the quantity and quality of MS/MS data acquired on modern mass spectrometers, a vast diversity of non-enzymatic chemical modifications has remained unidentified.

In shotgun proteomics, several computational strategies have been developed to identify modified peptides, including ion indexing methods.[342-346] Non-targeted peptide glycation product search lacks computational strategies. In conventional open searching of peptide chemical modifications, regular peptide fragment ions are used for peak matching. For peptide glycation

products, only a fraction of the experimental spectrum can be matched to all theoretical peptide fragment ions when considering only unmodified ions, leading to a poor correspondence between theoretical and experimental data. Matching shifted ions with modifications alongside regular peptide fragment ions in a dual indexing approach has been demonstrated to empower the open search concept in the proteomics field.[235] There is a need to also match fragments containing unknown modifications for fast identification of known and discovery of novel glycation products. Further, there is much valuable information in the neutral losses from the peptide modification side chains, and it is crucial to leverage these characteristic fragments for non-targeted glycation product analysis.

Here, we use three types of fragment ions for open search of peptide glycation products (Figure 3.1): (i) classic peptide fragment ions (CPFIs) to describe the theoretical peptide ion ($[Pep + H]^+$) and all theoretical N-terminal (e.g., a- or b-ions) and C-terminal ions (e.g., y-ions) without modifications, (ii) modified peptide fragment ions (MPFIs) to define peptide ions with modifications, and (iii) characteristic fragment ions (CFIs) to specify modified fragments with characteristic redundant neutral losses. We propose a computational approach that enables fast matching of experimental spectra, which allows exploration for both known and unknown peptide glycation products.

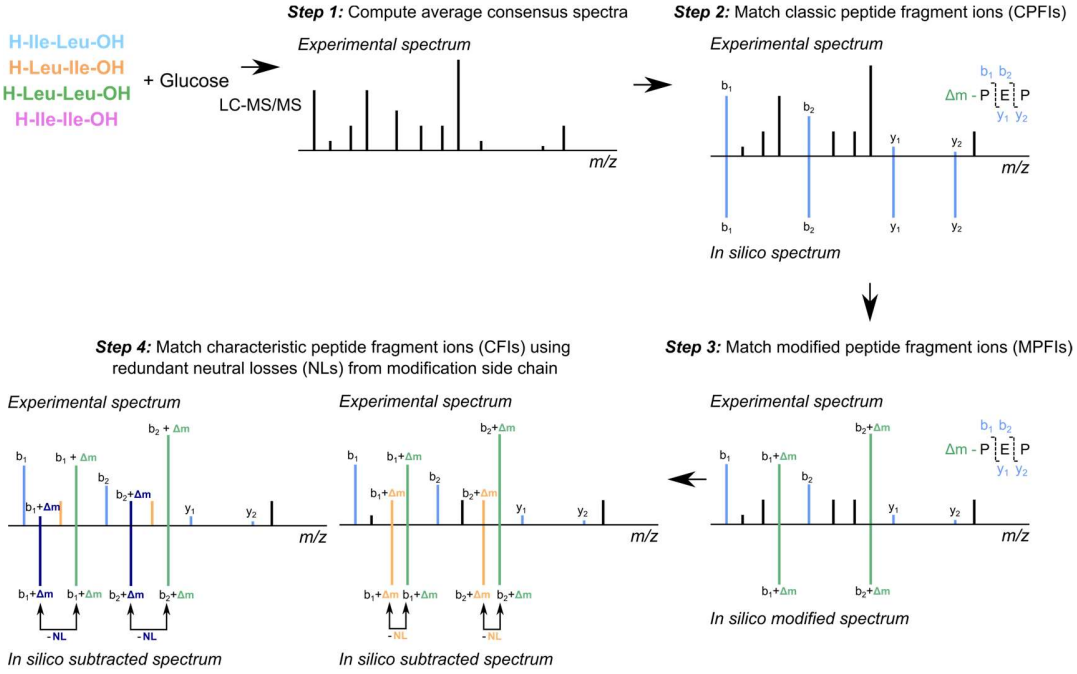


Figure 3.1: Open peptide glycation product search strategy step-by-step workflow. Step 1, consensus spectra computation of glycation products from four standard peptide model systems. Step 2, *in silico* digestion of input peptides provides a set of theoretical peptide fragment ions. Comparison of a CPFI catalogue (theoretical fragments and relevant amino acid fragmentation) against experimental spectra. The exact mass of the chemical modification (Δm) is calculated by subtraction of the corresponding peptide from the precursor mass (Δm = $mass_{glycation\ product}$ − $mass_{peptide\ matched}$). Step 3, each theoretical fragment is subtracted from each experimental peak. Fragments yielding Δm after subtraction are matched as MPFIs. Step 4, each theoretical fragment is subtracted from each experimental peak. Fragments yielding redundant Δm_n values due to characteristic neutral losses from the modification side chain are identified as CFIs.

3.2 Experimental section

3.2.1 Preparation of peptide glucose model systems

(+)-Glucose (≥99.5%) was purchased from Sigma-Aldrich (Steinheim, Germany). H-Ile-Ile-OH and H-Leu-Leu-OH were purchased from Santa Cruz Biotechnology (Texas, USA). H-Ile-Leu-OH and H-Leu-Ile-OH were purchased from Bachem (Bubendorf, Switzerland). Glucose stock solution (0.08 M) was mixed with each peptide standard stock solution (0.02 M) [both prepared in MilliQ-purified water from a Milli-Q Integral Water Purification System (18.2 MΩ, Billerica, MA, USA)] 1:1 (*v*/*v*). Aqueous peptide standard solutions (0.01 M) were prepared as control samples. Model systems were heated in closed glass vials for 2 and 10 h at 100 °C according to the protocol recently described.[347] Control samples were heated, analogously. Sample preparation was performed in triplicate (n = 3).

3.2.2 LC-MS/MS

Model systems were diluted 1:5 (*v*/*v*) with an aqueous solution containing 2% acetonitrile (LC−MS grade, Merck, Darmstadt, Germany) prior to LC−MS/MS analysis. Samples were analyzed by ultrahigh-performance liquid chromatography (UHPLC) (Acquity, Waters, Milford, MA, USA) coupled to a quadrupole time-of-flight mass spectrometer (maXis, Bruker Daltonics, Bremen, Germany). For reversed phase (RP) chromatography, an ACQUITY UPLC BEH C8 column (150 × 2.1 mm, 1.7 μm, Waters, Milford, MA, USA) was used. The column temperature was maintained at 60 °C. RP separation was run in gradient mode. The RP eluent A was a composition of 5 mmol/L NH4Ac and 0.1% acetic acid in water and RP eluent B was composed of acetonitrile. The pre-equilibration time was set to 2.5 min. Initial conditions were set to 90% eluent A and 10% eluent B. This composition was maintained until 1 min. Eluent B was increased to 22% within 2 min. Eluent B then reached 27% within 4 min and 55% after an additional 5.5 min. Eluent B was finally increased to 95% within 2.5 min and kept at 95% until the end of the run. The gradient was completed after 17 min. Samples were injected *via* full-loop injection (10 μL). Mass spectra were acquired in positive electrospray ionization (ESI) mode. Internal calibration was performed using a tuning mix solution (Agilent Technologies, Waldbronn, Germany) prior to each measurement. Parameters of the ESI source were as follows: capillary voltage 4000 V, dry gas temperature 200 °C, nebulizer pressure 2 bar, and nitrogen flow rate 10 L/min. Mass spectra were recorded with an acquisition rate of 5 Hz within a mass range of m/z = 50−1500. For data-dependent MS/MS acquisition, the most abundant ion of a full MS scan was subjected to MS/MS after each precursor

scan. The collision energy (CE) was set to 25 eV and to change dynamically and depending on the mass of the precursor molecule. For CE optimization, fixed collision energies of 15, 20, and 25 eV were applied. Raw data were post-processed using Genedata Expressionist Refiner MS 13.5 (Genedata GmbH, Basel, Switzerland) applying chemical noise subtraction, intensity cutoff filter, calibration, chromatographic peak picking, and isotope clustering. Only features detected in all three replicates were retained in the matrix.

3.2.3 Analyte classification

Analytes (retention time ≥ 2 min) were classified according to Yaylayan into two reaction pools:[348] glycation products and peptide-related analytes (e.g., degradation products). Due to poor retention on RP liquid chromatography, sugar degradation products were not considered. Ion signals reaching minimum intensity values of 1500 in all three replicates of the aqueous peptide−glucose model systems but not in control samples (peptides heated alone) were classified as peptide glycation products. Analytes reaching minimum intensity values of 1500 also in peptide blank samples were categorized as peptide-related compounds. Features found in solvent blank samples were considered as contaminants and removed by blank subtraction.

3.2.4 Consensus spectra computation

MS/MS consensus spectra computation was performed using the R package MsnBase.[349] MS/MS spectra were merged using the consensusSpectrum() function. To mitigate the abundance of noisy and non-reproducible fragments in fragmentation patterns, we retained mass peaks present in a minimal proportion of 50% of spectra in the final consensus spectra. Intensities were summed for the aggregated peaks, and the maximum m/z merge distance was set to 0.005 Da.

3.2.5 Statistics

To statistically assess the improvement of glycation product MS/MS matching by the use of different ion types, the non-parametric Wilcoxon test was applied for pairwise comparison using the wilcox.test() function of the stats R package.

3.2.6 Nuclear magnetic resonance spectroscopy

All samples were diluted 1:2 with a 1:1 H_2O/D_2O containing sodium 3- (trimethylsilyl)propionate-d_4 (TSP, 0.9 mM) as a chemical shift reagent and di-sodium hydrogen phosphate (0.75 M, pH 7) to buffer the sample at pH 7. The samples were measured in triplicates. Experiments were carried out using an 800 MHz Bruker AVANCE lll spectrometer equipped with a 5 mm QCI-probehead

at 300 K. One dimensional (1D) ^{1}H spectra were recorded using a pulse program constituted of pre-saturation during the relaxation delay followed by a 90 ° hard pulse. The overall recycling delay of each scan was set to 12 s after T1 relaxation for the peaks under investigation was determined. A 12.5 s hard pulse was used. 64 scans were acquired for every spectrum. The acquisition time was set to 4 s with a spectral width of 12,820 Hz. The assignment of the observed signals was carried out based on comparing the spectra of the unreacted control samples to the spectrum of the mixture. Relative quantification was done by integration of all isolated peaks of the peptide under investigation. The obtained areas were used to calculate the corresponding relative concentration by comparison with the TSP area.

3.3 Results and discussion

3.3.1 Collision energy optimization

As the detected ion profile of MS/MS spectra and the sensitivity are determined by the CE, this parameter was first investigated. Incubation of short-chain peptides (0.01 M) with glucose (0.04 M) at 100 °C for up to 10 h was followed by RP UHPLC-MS/MS analysis. The results were obtained under positive ESI [ESI(+)]. To obtain maximum information about the CPFI, and thus, the amino compound-derived core structure of glycation products, we applied dynamic voltages in the range 15−25 with 5 eV increments. After LC−MS/MS analysis, we computed MS/MS consensus spectra (Figure 3.1). It has been previously shown that the consensus spectrum is a superior representation compared to the best replicate, especially when measuring a small number of replicates and none of them being of particularly good quality.[350,351] Combining mediocre replicates to form a consensus spectrum allows to remove noise and to average experimental variation. Consensus spectra computation represents a much more robust approach compared to selecting any of the replicates for MS/MS matching, and its benefit for proteomics applications was previously demonstrated.[352]

Importantly, chemical modification of peptides can significantly affect the mass distribution of fragment ions generated by collisional activation.[353-355] For example, b_1 ions appear as dominant features in MS/MS spectra of amidinated peptides, while usually not observed for unmodified peptides due to their instability.[356] Equally, sufficient peptide backbone fragmentation may be observed for the unmodified peptide but cannot be achieved for peptide glycation products using the same CID voltage. CE dependency and variance in the relative

abundance of the CPFI for fragmentation of the starter peptide H-Ile-Leu-OH and the corresponding Amadori compound $C_6H_{10}O_5$-Ile-Leu-OH, a well-studied product of early glycation reactions, can be seen in Figure B.S1. At 15 eV, formation of CPFI appeared to be highest for the unmodified peptide (Figure B.S1A, left), while the Amadori product showed the highest relative CPFI intensity at 25 eV (Figure B.S1C, right). For the Amadori product, fragment diversity notably increased at the high voltages (Figure B.S1, right), suggesting that the formation of MPFIs and CFIs requires excess of energy. Considering the highest abundance of all three ion types, namely CPFIs, MPFIs, and CFIs, the CE was set to 25 eV (Figure 3.2A), changing dynamically with the precursor mass. This ensures improved certainty for full backbone coverage in peptide glycation product MS/MS spectra and facilitates proper fragmentation of higher molecular mass modifications.

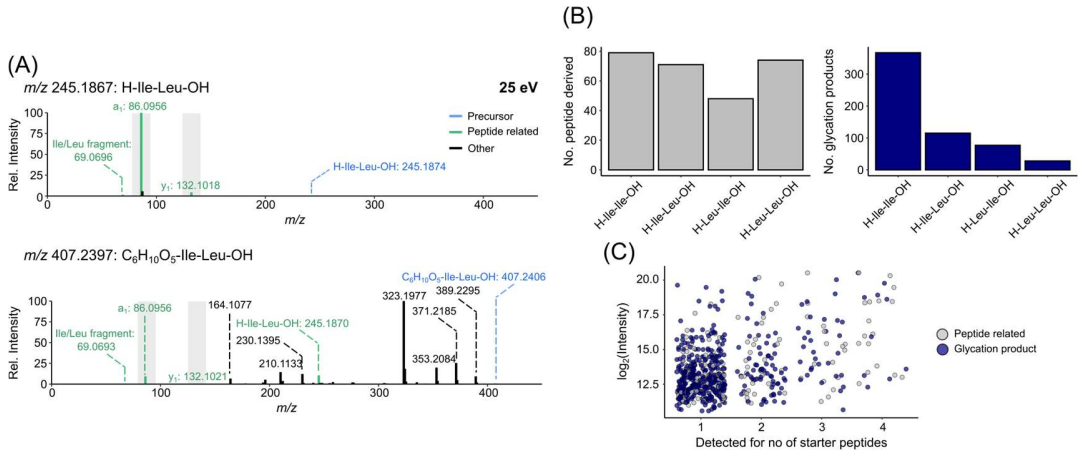


Figure 3.2: Characterization of short-chain peptide model systems. (A) Consensus MS/MS spectra (25 eV) of the standard peptide H-Ile-Leu-OH (top) and the corresponding Amadori product $C_6H_{10}O_5$-Ile-Leu-OH (bottom). The base peak is assigned to an abundance of 100%. Relative abundance of N-terminal a- and b-ions and C-terminal y-ions is marked by gray boxes. The precursor ion is highlighted in blue. (B) Bar plots of the abundance of peptide-derived analytes (left: gray) and glycation products (right: dark blue) in short-chain peptide-glucose model systems after heat treatment. (C) Maximum intensity as $\log_2$ of analytes across starter peptides. In total, model systems from four different starter peptides were analyzed. Legend: gray, peptide-derived analytes; dark blue, glycation products.

3.3.2 Short-chain peptide demonstration data sets

A full comprehensive study of glycation product fragmentation behavior was performed to showcase the algorithm and contribute to the dark matter of non-enzymatic chemical

modifications. We chose glucose model systems from four (iso)leucine peptides, namely H-Ile-Ile-OH, H-Ile-Leu-OH, H-Leu-Ile-OH, and H-Leu-Leu-OH, as demonstration data sets. It is important to note that the short-chain peptides investigated herein were selected to keep fragmentation patterns simple and to display poor comparability of glycation reactions even among peptides with highly similar amino acid sequences. To visualize the linkage between reagents and products, we show the number of peptide-derived analytes (e.g., degradation products) and glycation products for these four standard peptides (Figure 3.2B). Classification of the reaction pool into peptide degradation products and reaction products was inspired by Yaylayan's approach[348] (for details, see Methods: Analyte Classification). For peptide-derived signals, similar numbers were detected independent of the starter peptide (Figure 3.2B, left). Substitution of leucine for isoleucine, especially at the N-terminal sequence position, caused pronounced differences in glycation product diversity (Figure 3.2B, right). We performed nuclear magnetic resonance spectroscopy to quantify the peptide consumed by glycation reactions and to get additional insights into the peptide reaction behavior *via* an orthogonal measurement technology.

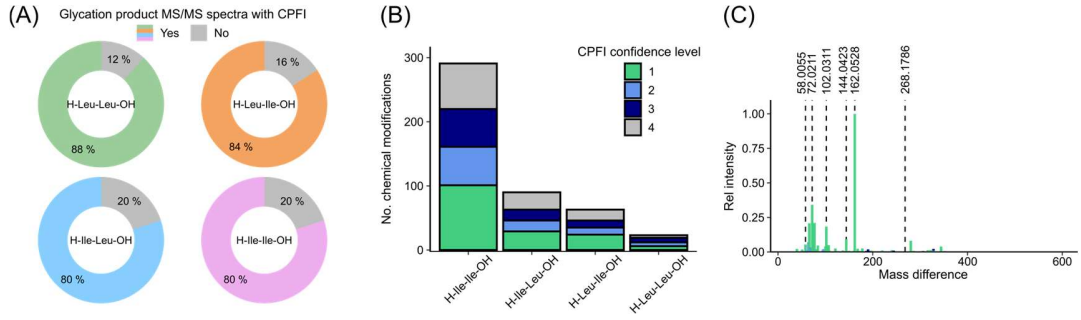


Figure 3.3: Matching of CPFIs. (A) Percent of the glycation products matched to CPFIs. (B) Distributions of CPFI matching confidence levels versus model system starter peptides. (C) Relative maximum intensity of glycation chemical modifications, shown as mass differences (Δm = $mass_{glycation\ product}$ − $mass_{peptide\ matched}$) of minimum +1.0078 for confidence levels 1−3.

As suggested by the MS results, superior reactivity of isoleucine peptides was indicated by higher levels of peptide loss (Figure B.S2) and increased model system complexity (Figures B.S3 and B.S4). High specificity of glycation products for the peptide reagent is visualized in Figure 3.2C, showing that most glycation products were unique for one single starter peptide. This may not only be based on differences in reactivity but also on marginal deviations in the starter peptide physicochemical properties, which can cause retention time shifts for the same

type of reaction product. Observation of such pronounced discrepancies for highly similar short-chain peptides demonstrates that there is a need to establish an algorithm to compare glycation modifications across peptides, comprehensively.

3.3.3 Fragmentation study of short-chain peptide glycation derivatives

An algorithm for automated search of peptide glycation products can only work well if the peptide component embedded in the cores of the reaction products is considered. Thus, we begin by matching experimental spectra against the CPFIs of the corresponding candidate peptides. For CPFI fragment matching, we perform *in silico* digestion of the four peptide standards studied in this work. This peptide fragment database was complemented by an in-house amino acid LC–MS/MS library, which is available from the Mass Bank of North America. Figure 3.3 shows the results of this first step of the multi-step algorithm. Searching acquired glycation product MS/MS spectra against the CPFI database, up to 88% of each reaction product pool shared peptide backbone fragments or had amino acid MS/MS fragmentations (Figure 3.3A). CPFI intensities may vary depending on the chemical modification, and analytes were classified into four confidence levels depending on the match quality: Level 1, complete peptide backbone detection by both N- (a- and b-ions) and C-terminal (y-) ions, respectively; detection of the ion related to the unmodified peptide (e.g., *m*/*z* 245.1860 for [H-Ile-Leu-OH + H]$^+$). Level 2, complete peptide backbone detectability considering all types of peptide backbone ions; no requirement for detection of the unmodified peptide ion. Level 3, detection of any N- or C-terminal peptide ions. Level 4, only observation of ions related to fragmentation of relevant amino acids. Most CPFI matches correspond to confidence levels 1 and 2 for all starter peptides (Figure 3.3B), meaning that sufficient fragmentation of the peptide backbone was achieved for the majority of the glycation products. Chemical modification mass differences (Δm) were calculated as follows

$$\Delta m = \text{mass}_{\text{glycation product}} - \text{mass}_{\text{peptide matched}}$$

Summarized delta masses are shown in Figure 3.3C and Table B.S1. Mass differences at least observed for confidence level 3 and of minimum 1.0078 (mass of one hydrogen atom) were included. Highly abundant examples are highlighted (e.g., delta masses 162.0528, 144.0423, 72.0211, and 58.0055 Da). A mass increase of 162.0528 Da may be putatively assigned as the Amadori product (+$C_6H_{10}O_5$, 162.0528 ± 0.005 Da). Delta masses 58.0055 and 72.0211 Da may correspond to carboxymethylation (+$C_2H_2O_2$, 58.0000 ± 0.005 Da) and carboxyethylation (+$C_3H_4O_2$, 72.0211 ± 0.005 Da). Other *hitherto* undescribed glycation modifications with high abundance were

observed (e.g., +$C_4H_6O_3$, 102.0316 ± 0.005 Da) as well as higher molecular mass shifts (e.g., +$C_{14}H_{24}N_2O_3$, 268.1787 ± 0.005 Da).

Next, experimental spectra were searched for MPFIs. MPFIs carry the glycation modification Δm and MPFI calculation is expressed as

$$m/z_{\mathrm{MPFI}} = m/z_{\mathrm{a-,b-i}} \quad + \Delta m$$

Fragment ions that remained unmatched were used for the detection of redundant neutral losses to identify CFIs. We believe that the key feature of an effective peptide glycation product search strategy is to determine not only the peptide backbone but also the ability to subtract the peptide core structure and unmask the modification structural information concealed in the MS/MS spectra. In particular, the implementation of CFIs can determine important neutral losses that do not represent core structural information and allow the comparison of glycation modifications across different peptide types.

Overall, our approach found glycation modifications predominantly producing CPFIs, MPFIs, or CFIs, respectively (Figure 3.4A). A rather poor contribution of CPFIs to experimental spectra was identified for most peptide glycation products. For glycation products with a low CPFI relative intensity, more than half of the useful spectral information is ignored when solely matching unmodified peptide ions. However, the amino compound core structure alone is also key information for structural identification of both known and unknown peptide glycation products. MPFIs are particularly useful for non-labile modifications, where peptide fragment ions retain the modification after dissociation of the peptide core structure.[235] Other glycation products are highly prone to fragmentation of the modification side chain (e.g., Amadori products) and mainly produce CFIs by redundant neutral losses (e.g., water loss), which may give additional structural information. A high abundance of CFIs is also expected for peptides with multiple modifications (e.g., $\Delta m = \Delta m_1 + \Delta m_2$) as detectability of MPFIs may be limited in such fragmentation patterns and each individual modification Δm_n will instead be assigned to be characteristic. The fraction of interpretable fragments was remarkably boosted by refined matching using MPFIs and CFIs alongside CPFIs (Figure 3.4B). With MPFIs and CFIs in conjunction with regular peptide fragment ions, we significantly increased the median MS/MS intensity matched from 7 to 71% (Wilcoxon test, $p < 0.001$, Figure 3.4C).

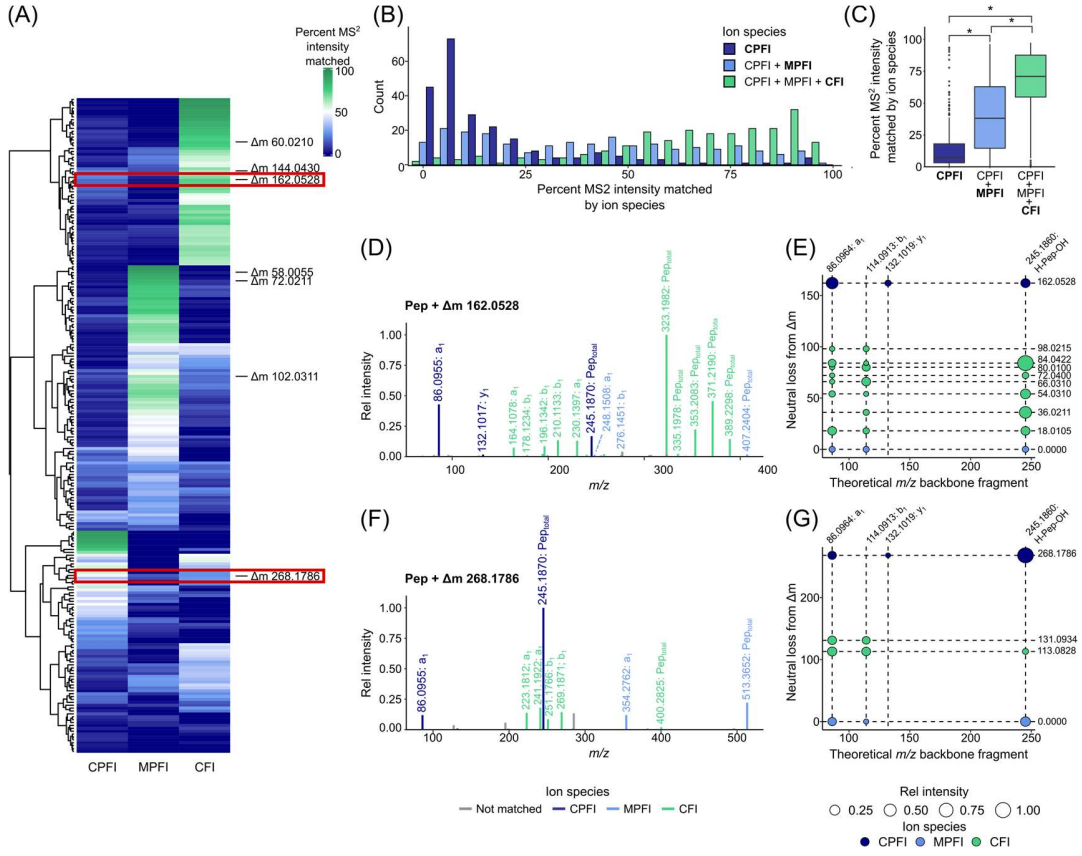


Figure 3.4: Contribution of CPFIs, MPFIs and CFIs to peptide glycation product MS/MS spectra. (A) Relative MS/MS intensity of each ion type visualized as a heatmap for peptide glycation products. For each mass difference, the analyte with the highest relative MS/MS intensity matched is shown. Row clustering is based on the results from hclust analysis ("ward.D2" method). Mass differences illustrated in (D−G) are highlighted by red rectangles. (B) The distribution shows the high or low coverage of glycation product MS/MS spectra upon use of CPFIs, MPFIs, and CFIs. (C) The boxplot shows glycation product MS/MS coverage for the different ion species. The box represents the interquartile range, the horizontal line in the box is the median, and the whiskers represent 1.5 times the interquartile range. Asterisks indicate significant changes (Wilcoxon test, $p < 0.001$). As an example, representative MS/MS consensus spectra are illustrated for two selected chemical modifications with a mass delta of 162.0528 Da (D) and 268.1786 Da (F) color-coded by ion type. Redundant neutral losses are shown for CFI characterization depending on corresponding peptide backbone ion *m*/*z* for the glycation modifications 162.0528 Da (E) and 268.1768 Da (G).

We present in Figure 3.4D and Figure 3.4F two examples for matched experimental spectra of peptide glycation modifications that show different relative intensities for the three ion types (see Table B.S2 for detailed information on matched fragment ions). The consensus spectrum

of the putatively assigned Amadori product (+$C_6H_{10}O_5$, Δm = 162.0528 ± 0.005 Da) primarily shows CFI fragments (Figure 3.4A, highlighted in red rectangle; Figure 3.4D). Here, identification sensitivity toward MPFIs is low, and such modifications do barely benefit from the inclusion of modified ions on matching. Due to its labile nature, this type of modification easily fragments. This is also true for post-translational modifications (PTMs) that are searched in the proteomics field, such as phosphorylation.[235] The main CFIs were characterized by a neutral loss of 84.0422 Da (Figure 3.4E), which can be annotated as −3H_2O−CH_2O (84.0423 ± 0.005 Da). Besides, distinct CFI neutral losses were 18.0105 Da (−H_2O, 18.0105 ± 0.005 Da), 36.0211 Da (−2H_2O, 36.0211 ± 0.005 Da), and 54.0310 Da (−3H_2O, 54.0317 ± 0.005 Da). Consequently, our approach identified all Amadori diagnostic ions described in the literature so far.[304,357] We also observed that our algorithm helped to find 66.0310, 72.0400, 80.0100, and 98.0215 Da as additional redundant neutral losses, demonstrating the potential of the open search strategy.

The delta mass 268.1786 Da was annotated as a potential peptide crosslink. It may correspond to two peptide moieties that were linked *via* an ethylene bridge (+$C_{14}H_{24}N_2O_3$, 268.1787 ± 0.005 Da). The consensus MS/MS spectrum of this modification showed a large fraction of CPFIs. The relative contribution of MPFIs and CFIs was similar (Figure 3.4A, highlighted in red rectangle). As redundant neutral losses, 113.0820 Da (−$C_6H_{11}NO$, 113.0841 ± 0.005 Da) and 131.0934 Da (−$C_6H_{13}NO_2$, 131.0940 ± 0.005 Da) were determined (Figure 3.4G). These neutral losses correspond to fragmentation of the peptide core structure.

3.3.4 Algorithm validation using stable isotope-labeling

Stable isotope-labeling was used to provide increased annotation confidence and to further demonstrate the utility of our open search method. Reliable identification of peptide glycation products, especially if previously undescribed, has been traditionally challenging with long data analysis times and increased false annotation rates. Applying our triple-ion search algorithm on LC−MS/MS data from both unlabeled and isotopically tagged analytes enables high-confidence annotation in an untargeted manner. For non-targeted profiling of ^{13}C-labeled glycation products, model systems were prepared from D-glucose-$^{13}C_6$. H-Ile-Leu-OH was used as the peptide reagent. Both well-known and *hitherto* uncharacterized glycation products were used as benchmarks. Based on chromatographic and mass spectrometric properties of the labeled reaction products, identification and characterization of selected glycation modifications were achieved. Because of their similar structures, the retention times of ^{13}C-labeled and non-labeled products

were comparable (Figure 3.5, left). The mass-to-charge ratio differences ($\Delta m/z$) were $\Delta m/z$ = 6.0197 (Figure 3.5A, 6 × $^{12}C \leftrightarrow {}^{13}C$ $\Delta m/z_{theoretical}$ = 6.0206), $\Delta m/z$ = 4.0127 (Figure 3.5B, 4 × $^{12}C \leftrightarrow {}^{13}C$ $\Delta m/z_{theoretical}$ = 4.0137) and $\Delta m/z$ = 2.0158 (Figure 3.5C, 2 × $^{12}C \leftrightarrow {}^{13}C$ $\Delta m/z_{theoretical}$ = 2.0069). These results indicate the number of glucose-derived carbon atoms that contribute to the glycation product structures (1 × $^{12}C \leftrightarrow {}^{13}C$ $\Delta m/z_{theoretical}$ = 1.0034). The fragmentation patterns of labeled glycation products allow verification of known and identification of undescribed modifications. As expected, the ^{13}C-labeled and non-labeled reaction products had highly similar MS/MS spectra (Figure 3.5, right). Detection of common fragment ions (e.g., a_1: m/z 86.0964, b_1: m/z 132.1019, [H-Ile-Leu-OH + H]$^+$: m/z 245.1860) suggests the absence of isotopic tags. This is expected as CPFIs do not contain any modification and may be detected for all glycated peptides. In ^{13}C-labeled models, for example, the fragment ions m/z 395.2497, 377.2392, and 359.2285 (Figure 3.5A right, blue) of the putatively assigned Amadori product ($C_6H_{10}O_5$-Ile-Leu-OH, m/z 407.2388 ± 0.005 Da) were mass shifts $\Delta m/z$ = 6.0206 ± 0.005 to their corresponding unlabeled ions (−H_2O: m/z 389.2299, −2H_2O: m/z 371.2192, −3H_2O: m/z 353.2085; Figure 3.5A right, black). These fragments contained whole isotopic tags, which was expected as there is no ^{13}C loss from the $^{13}C_6H_{10}O_5$ side chain by any of these fragmentation pathways. The labeled ion m/z 328.2149 only retains a part of its isotopic tags and differs from its non-labeled ion m/z 323.1983 (−3H_2O−CH_2O) by 5 × $^{12}C \leftrightarrow {}^{13}C$ ($\Delta m/z$ = 5.0172 ± 0.005). All described fragmentation patterns are consistent with the literature.

We show in Figure 3.5B (right) that for the mass delta 102.0311 Da ($C_4H_6O_3$-Ile-Leu-OH: m/z 345.2020 ± 0.005 Da), the $\Delta m/z$ between isotopically tagged ions (e.g., m/z 331.2058 and 313.1961) and the unlabeled fragments (e.g., m/z 327.1917 and 309.1825) was 4 × $^{12}C \leftrightarrow {}^{13}C$. Peptide backbone fragmentation (−$C_6H_{13}NO_2$−CO) generated the base peaks at m/z 188.1289 (black) and m/z 192.1422 (blue), leaving the labeled glycation modification intact. These results suggested that the unknown modification does indeed contain four glucose-derived carbon atoms as annotated. The isotopically labeled fragment at m/z 289.2048 was formed by a loss of $^{13}C_2H_2O_2$ from the modification side chain and corresponds to m/z 287.1991 in the unlabeled MS/MS spectrum, shifted by 2 × $^{12}C \leftrightarrow {}^{13}C$. Sequential fragmentation of the peptide backbone (−$C_6H_{13}NO_2$−CO) and loss of $C_2H_4O_2$ from the modification side chain leads to m/z 128.1071 (black) shifted by 2 × $^{12}C \leftrightarrow {}^{13}C$ in the labeled MS/MS spectrum (blue, m/z 130.1136). In the context of MR, $C_4H_6O_3$ was previously described as free 2- hydroxy-3-oxobutanal, which can isomerize into various

tautomers with methylreductone and dicarbonyl structures.[358] To the best of our knowledge, no peptide bound glycation modification with this molecular formula has been previously described.

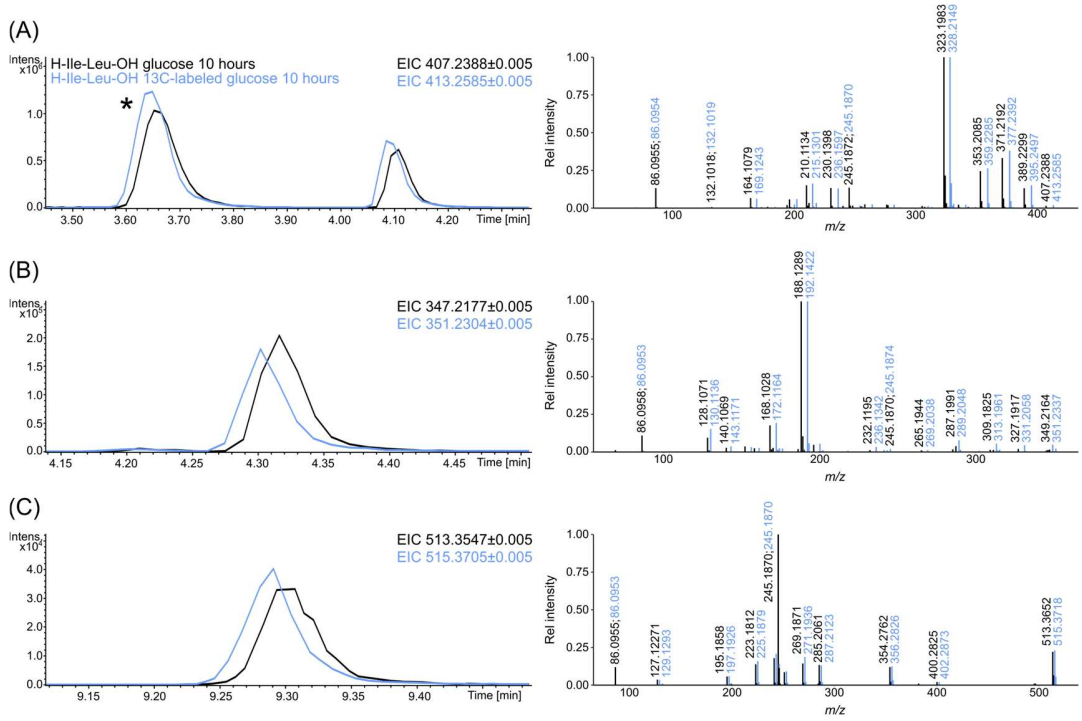


Figure 3.5: Chromatographic and mass spectrometric characterization of selected glycation modifications using isotope labeling. Representative extracted ion chromatograms (left) and collision-induced dissociation consensus MS/MS experiments (right) of peptide chemical modifications with a mass delta of 162.0528 Da (A), 102.0311 Da (B), and 268.1786 Da (C) (non-labeled: black, ^{13}C-labeled: blue).

For the *hitherto* unknown cross-linked H-Ile-Leu-OH, isotopically tagged ions (e.g., *m/z* 402.2873, 356.2826, and 287.2123) showed a $\Delta m/z$ of 2 × ^{12}C ↔ ^{13}C (*m/z* 400.2528, 354.2762, and 285.2061; Figure 3.5B, right). This indicates ^{13}C-labeling of the hypothesized ethylene cross-link coming from the sugar compound. The labeled linkage may be retained, while the peptide moieties dissociate by peptide bond fragmentation giving, for example, *m/z* 400.2825 (−$C_6H_{11}NO$) and *m/z* 354.2726 (−$C_6H_{13}NO_2$−CO) shifted by $\Delta m/z$ of 2 × ^{12}C ↔ ^{13}C in the labeled MS/MS spectrum (*m/z* 402.2872 and 356.2826).

Integrating complementary MS/MS spectra from ^{13}C-labeled glycation products, we provide high-confidence structural information. Both known and unknown peptide glycation

product annotation could be refined with orthogonal stable isotope-labeling approaches to validate the predicted structure identities.

3.4 Conclusions

In summary, understanding MS/MS spectra of glycated peptides is a crucial yet challenging task due to their complicated and peptide core structure-dependent fragmentation pattern. In this work, we revealed the labile nature of many glycation side chains, pointing out the necessity to include characteristic neutral losses for MS/MS matching. We thus developed a triple-ion strategy, also considering side-chain fragmentation, to significantly improve matching of peptide glycation product MS/MS spectra. From the study of short-chain peptide model systems, we conclude that our multi-step algorithm allows for fast, accurate, and automatic identification of well-known glycation products. By characterizing previously undescribed peptide glycation modifications, we further demonstrate that our straightforward approach is highly suitable for exploration of novel glycation products and that redundant neutral losses constitute an informative source to accurately determine their molecular identity. We also discuss the possibility to efficiently integrate isotopically labeled and label-free fragmentation patterns for increased annotation confidence. Overall, our work shows that the use of MPFIs and CFIs among conventional CPFIs provides a significant advancement in MS/MS spectra analysis, empowering the open search concept for modified peptides. We envision that our triple-ion approach, with more successful cases demonstrated, will show a great value in future non-targeted analyses of peptide modifications.

Chapter 4

General Discussion and Outlook

"The time will come when diligent research over long periods will bring to light things which now lie hidden. […] And so, this knowledge will be unfolded only through long successive ages."

— Seneca

The presented thesis reports on the analysis of peptide glycation with focus on sequence-dependent reaction behavior and strategies for data evaluation. Investigating how different peptides react in early glycation allowed to detect structural properties, which influence the formation of ARPs (Chapter 2). Glycation does not only produce ARPs but many other reaction products. For fast and simple interpretation of both known and unknown glycation modifications across all reaction stages, a custom method was established, which relies on tandem mass spectra (Chapter 3).

This thesis emphasizes that the peptide structure plays an important role in glycation. We found that even simple (iso)leucine dipeptides do not show comparable reaction behavior. When said dipeptides were incubated with glucose, we observed considerable differences in the number of glycation products. Reactivity seemed to increase in the following order: H-Leu-Leu-OH, H-Leu-Ile-OH, H-Ile-Leu-OH, H-Ile-Ile-OH. One cannot imagine how broad the spectrum of peptide reactivity is, if already slight sequence alterations like exchanging leucine for isoleucine can have such strong effects.

Most studies on peptide glycation have focused on di- or tripeptides. Many longer-chain peptides, however, contribute to real-life matrices.[359,360] One major challenge for studying glycation of (larger) peptides is the selection of a representative peptide set. Enzymatic digests of proteins can help to produce a mixture of peptides with different sequence lengths.[361] Nevertheless, enzymes follow specific cutting rules, and introduce a bias toward certain amino acid sequences. For example, only arginine or lysine can be found at the C-terminus of tryptic peptides, which cannot reflect reality. Heat treatment can break down proteins into more diverse peptides due to less systematic bond cleavages.[362,363] Tryptone is produced by spray drying a

tryptic casein digest and contains many heterogenous peptide structures. We created model systems using tryptone as a peptide source, which allowed us to study glycation of longer-chain peptides in a more realistic setting. To use heat-treated digests from diverse proteins could be an objective for future studies.

Early-stage glycation products are the source of many downstream AGEs (see section 1.1.2). The fragmentation pattern of peptide-ARPs is well-studied, which allows for reliable identification *via* MS/MS.[304,364,365] Especially in non-targeted MS, where we aim to identify many analytes at once, MS/MS spectra can enable high-throughput structure elucidation. Peptides with other $C_6H_{10}O_5$-modifications, namely HRPs and imidazolidin-4-ones, are far less intensively researched. To assess differences in peptide reactivity, monitoring the ARP over time appeared straightforward. Further, no suitable method was available to screen poly-peptide model systems for all types of glycation products at this stage of the thesis. It appeared that small sequence alterations can change the early-stage reaction behavior of both dipeptides and longer-chain peptides (e.g., YPFPGPI vs. VYPFPGPI).

Questions raised were whether there are general trends across all studied peptides and does the sequence position of a certain amino acid or the presence of a sequence "pattern" influence reactivity. Conventional dipeptide studies usually rely on variation of a single amino acid in the sequence and can only comment on effects of the first and/or second N-terminal sequence position.[85,86] We compared the sequence of ARP forming tryptone peptides to the sequence of tryptone peptides, which we could not detect as an ARP. For the first two N-terminal amino acids, our findings were consistent with a recent short-chain peptide study: if (iso)leucine or valine were positioned at the N-terminus or at the adjacent sequence position, ARP formation seemed to be generally promoted.[86] Previous peptide glycation studies, which used representative glycine dipeptides, have proposed that N-terminal proline lowers glycation efficiency.[314] While no peptide with N-terminal proline formed an ARP, proline was enriched at the third sequence position relative to the N-terminus for reactive peptides. Further, peptides with proline-rich subsequences (e.g., P-I-P, P-L-P, P-V-P, P-Y-P, and P-F-P) were more likely to produce ARPs than peptides without any of said patterns. Clearly, dipeptide studies alone cannot explain what drives peptide glycation.

To know whether structurally related peptides act similar in glycation is important in food science. Glycation products from amino reactants with comparable chemical structures often produce likewise aromas. Isoleucine and leucine model systems, for instance, both give off a caramel-like odor.[366] We suggest to investigate the aroma and flavor of glycation products with comparable peptide core structures as a part of future experiments. Health research could also profit from probing glycation of similar peptides. Peptides with common substructures often show similarities in their biological activity. As an example, the three egg-white peptides VVEVYLPR, VEVYLPR, and VYLPR all have antioxidative properties.[367] Bioactivity may be increased or lost by glycation.[100] Glycated peptides can even develop new bioactivities compared to their unreacted forms. The effect of glycation on the bioactivity of peptides depends on their structure. Future studies could investigate whether peptides with similar sequences and activities are equally reactive. Even more, the bioactivity of isolated glycation products should be tested. The information on reactivity and pre- and post-glycation activities of related peptides could then be combined. In this way, the optimum in terms of nutritional value, taste and/or bioactivity may be achieved. Peptides with biological or sensory activities were not the focus of this thesis. We, however, found that many tryptone peptides are part of relevant databases. For 60 peptides, bioactivities were documented, and 24 peptides were reported to have sensory activities.

Even with the many tryptone starter peptides, coverage of the peptidome in terms of all conceivable peptides was not feasible. There are virtually countless theoretical peptide structures. A general understanding of peptide glycation can only be achieved, if we cover the widest range of reactants. Hence, extending the data from peptide model systems as much as possible is a worthwhile endeavor for the future. Ongoing studies may pay more attention to, for example, lysine- or arginine-containing peptides, which were barely touched in this thesis. Most amino acid glycation studies focus on said amino acids because they are highly reactive (see section 1.1.5). Additionally, some glycation crosslinks and modifications can only be formed from the arginine side chain (see section 1.1.3 and 1.1.4). As stated, not many longer-chain peptide models are studied at this time. Hence, future experiments should aim to also cover more peptides with greater sequence length.

Not only the peptide reactant but also the sugar component may be varied. Glucose was the only carbonyl precursor studied in this thesis. We intentionally focused on glucose as it is the major free sugar and energy source in the human body. In foods, however, other mono- (e.g.,

fructose) and even disaccharides (e.g., lactose) are commonly found.[368,369] Instead of sugars, reactive dicarbonyls (e.g., glyoxal and methylglyoxal) may be used to prepare peptide models. Dicarbonyl models mirror real life, where we often add unreacted peptides to pre-heated (mixtures of amino and) sugar compounds.

For future *in vitro* studies, it should be considered that peptide model systems can be classified into mono- and poly-peptide models. Depending on the scientific question, the type of model system may be chosen. In this thesis, we used both model types. In real-life matrices, multiple peptides often co-exist and compete for reaction with the sugar. The most reactive peptides "absorb" the carbonyl in terms of glycation. If we want to, for example, mimic such samples, we must incubate many peptides together rather than isolate them. Using tryptone models, we showed that poly-peptide models can also be a way to efficiently detect how the amino acid sequence influences the potential of peptides to undergo glycation. If we incubate mono-peptide models, the peptides react with glucose, separately. The analyses are later merged and compared. To date, model systems, which contain only one single amino reactant, are most often used to study glycation. For Chapter 3, we aimed to choose model systems, which allowed for detection of known and unknown glycation modifications. Yet, we also wanted to keep fragmentation patterns simple. To further showcase poor comparability of glycation reactions even among short-chain peptides with highly similar amino acid sequences, we focused on (iso)leucine dipeptides. Mono-peptide model systems were most suitable for this study. All glycation products from said peptides would be isobaric. Further, similar chromatographic retention and comparable MS/MS can be expected. If incubated in multi-peptide model systems, tracing back the peptide precursor for each glycation product would have been complicated. Besides the selection of suitable models, isotopic labeling can support the discovery of previously undescribed glycation products and help to provide high confidence annotation. Future studies should consider incorporating isotopic labels to ease and improve non-targeted screening.

Already simple dipeptide glucose models produce many glycation products and offer great opportunities for detection of unknown structures. This becomes evident in Chapter 3, where we identified a previously undescribed ethylene crosslink from such models. Chapter 3 also underlines why we are missing out on new structures: there is a clear lack of suitable computational strategies for non-targeted analysis of peptide glycation. When taking a closer look at the experimental design of present peptide glycation studies, the need for open search methods

becomes apparent as well. Almost all current investigations rely on detection of one single glycation product.[82,83,86] Studies, which do analyze more than one glycation product, usually pre-select a limited number of glycation modifications for data evaluation. Usually, well-known glycation products are searched in such studies (e.g., carboxymethylated peptides).[297] To broaden our view, tools are required, which allow us to screen complex samples for any peptide glycation product.

The motivation behind the development of an open search approach can thus be summarized as a need for a unifying method to simultaneously assess known and unknown peptide glycation modifications. The algorithm in Chapter 3 is based on tandem mass spectra, as they are highly suitable for high-throughput annotation (see above). Because "bottom-up" proteomics is in fact analytics of (modified) peptides, it is an obvious option to test proteomics software for peptide glycation analysis. Proteomics tools are indeed suited to detect a wide range of PTMs.[370,371] Most proteomics tools are designed to detect well-known modifications, which must be pre-selected, and cannot help with non-targeted analysis. A limited number of proteomics strategies exist for search of peptides with unknown modifications. For example, MSFragger applies a dual-ion approach for identification of peptides and all their modified forms. Yu *et al.* demonstrated that using shifted besides regular fragment ions provides a significant boost in the number of identified modified peptides.[235] The non-targeted algorithm presented in this thesis uses three fragment categories. The third ion type is based on neutral losses, which describe the modification substructure. In Chapter 3, we show that a dual-ion strategy helps to describe many glycation products in dipeptide glucose models. However, three fragment classes were required to achieve high MS/MS coverage across all types of glycated dipeptides, which demonstrates that search methods should be expanded by CFIs. In the future, the use of CFIs could be furthered. Furthermore, a method for detection of false positives (e.g., using decoy peptides) should be established. Importantly, the current implementation is not yet a "one-stop-shop" for modified peptide search. Publishing an open-source package could be the next step. In this way, the feedback from the scientific community could help to optimize the approach.

A | Appendix Chapter 2

A.1 Supplementary figures

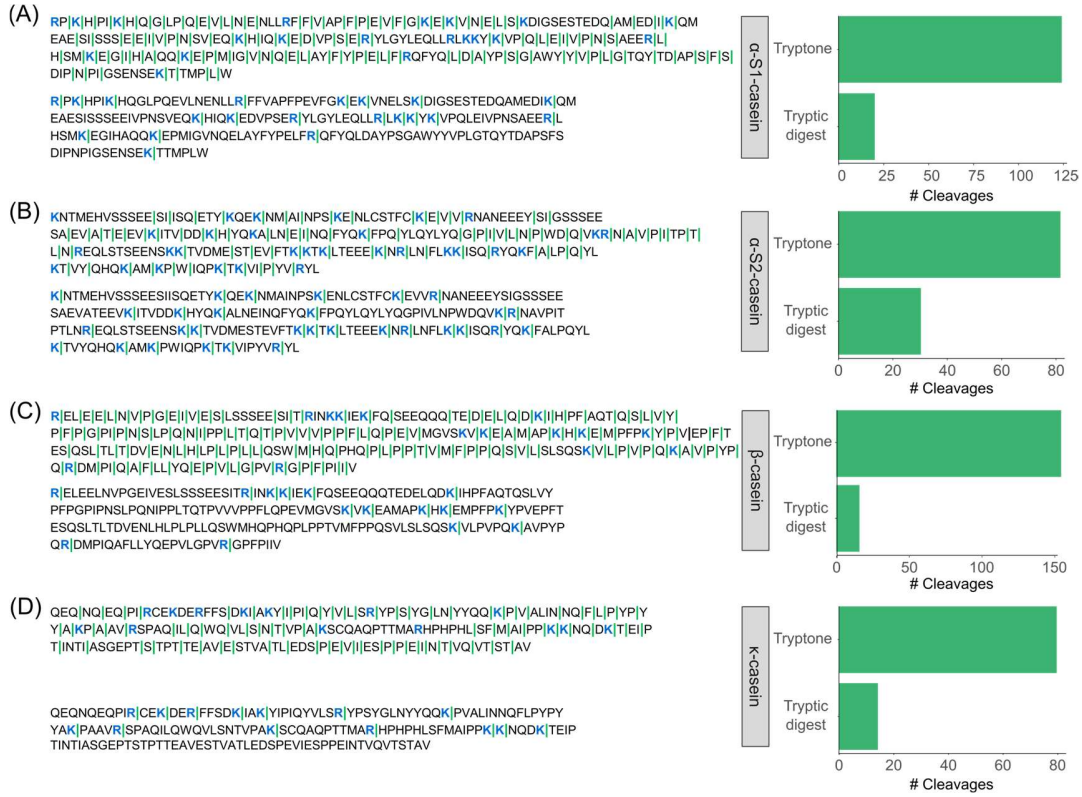


Figure A.S1: Tryptone peptides (*n* = 433 sites, upper protein sequences) provide substantially larger number of cleavage sites compared to a theoretical digest of casein proteins with trypsin (*n* = 79 sites, lower protein sequences). Detected tryptone peptides were used to compute cleavage sites on casein proteins (upper protein sequences). Theoretical cleavage sites for a tryptic casein digest are highlighted in the lower protein sequences, respectively. Vertical bars (|) indicate cleavage positions, and lysine (K) and arginine (R) are highlighted in blue to provide an eye guide for amino acids that lead to a protein cleavage by trypsin. Bar plots provide the total number of cleavage sites per protein sequence in tryptone and in a theoretical tryptic digest.

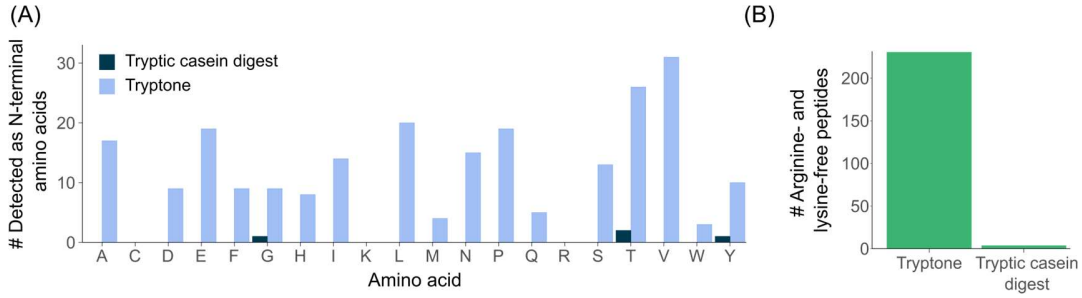


Figure A.S2: Tryptone promotes N-terminal diversity in arginine- and lysine-free peptides compared to a tryptic casein digest. (A) Bar graph shows N-terminal amino acids for arginine- and lysine-free peptides detected in tryptone (light blue) compared to arginine- and lysine-free peptides that can be expected after a tryptic digest of casein (dark blue). For the *in silico* digestion, α-S1-, α-S2-, β-, and κ-casein were considered. (B) Total number of arginine- and lysine-free peptides for tryptone and a theoretical tryptic casein digest.

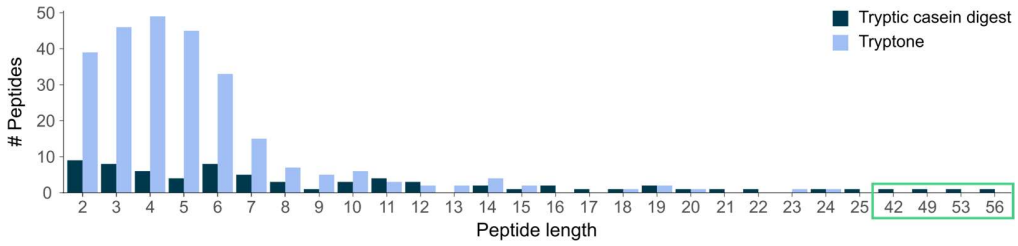


Figure A.S3: *In silico* tryptic casein digest predicts increased peptide lengths compared to experimentally determined tryptone peptides. Bar chart depicts length of peptides detected in tryptone and after a theoretical casein digest with trypsin. For this calculation, α-S1-, α-S2-, β-, and κ-casein were considered. Note the markedly large peptides generated by using trypsin (highlighted by green box).

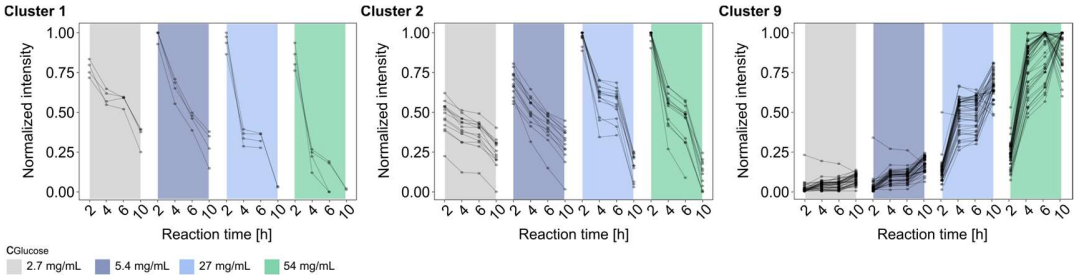


Figure A.S4: Intensity profiling enables AP classification by their formation behavior. Intensity profiles visualize time-resolved AP formation depending on the glucose concentration. Intensity values were normalized towards the greatest intensity value. Colors indicate different glucose concentrations.

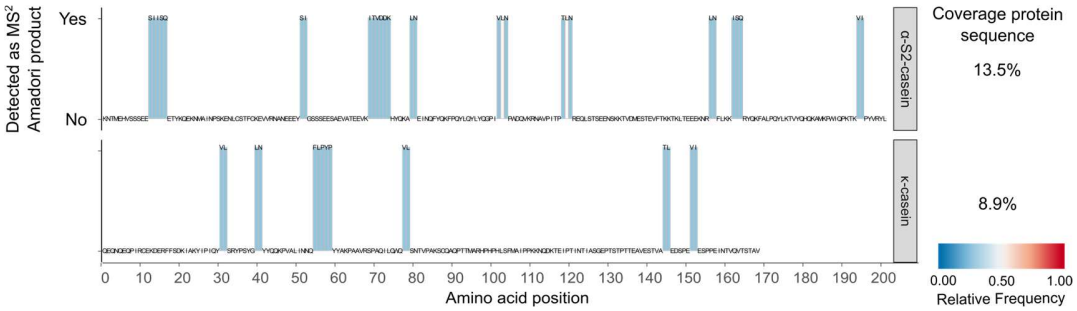


Figure A.S5: Unraveling casein protein sequences prone to AP formation. α-S2- and κ-casein protein heatmaps represent the relative frequency of amino acid positions that were detected as an AP, indicating which protein regions contribute most to early Maillard reaction and how the amino acid microenvironment influences peptide reaction behavior. Approximately 13.5% of α-S2- and 8.9% of κ-casein are detected as the corresponding AP.

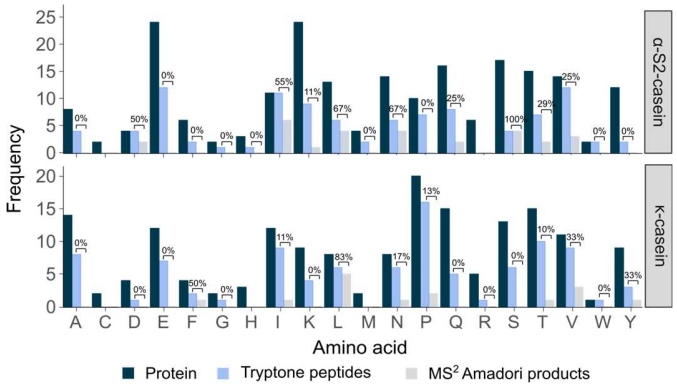


Figure A.S6: Analyzing amino acid distribution displays their contribution to peptide glycation. Abundance of amino acids in casein proteins (dark blue), tryptone peptides (light blue), and detected APs (gray). Embedded values indicate the percentage of peptides that could also be detected as an AP.

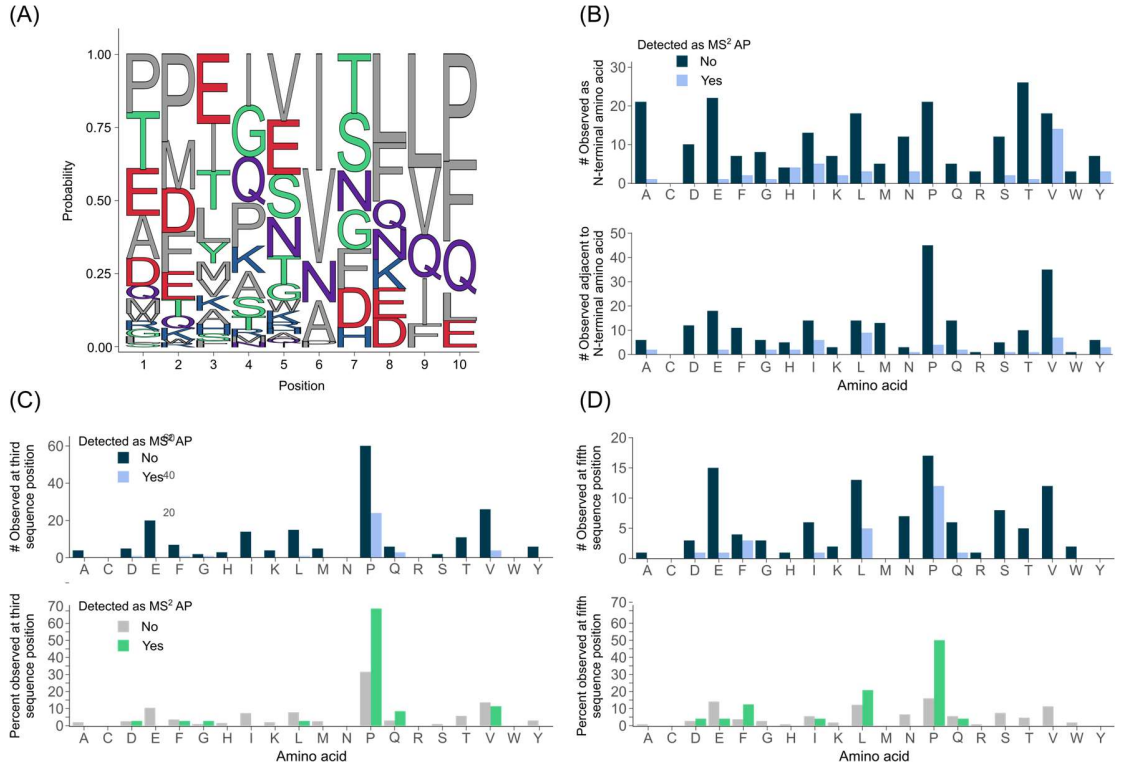


Figure A.S7: Peptide sequence analysis reveals interesting trends for the location of amino acids relative to the glycation site. (A) Sequence logo representation of the first ten amino acid sequence positions of non-glycated peptides. Amino acids with increased relative abundance in non-glycated peptides are illustrated (relative abundance$_{\text{non-glycated}}$ – relative abundance$_{\text{glycated}}$ > 0). (B) Bars represent the absolute number of amino acids detected at the first (top) and second (bottom) sequence position in peptides, which were not observed as an AP (dark blue) or for which the corresponding AP could be detected (light blue). (C) Bars show the absolute (top) and relative (bottom) frequency of (glycated) peptides that contain a given amino acid at the third position of the amino acid sequence (detected as AP: light blue (top) and green (bottom); not detected as an AP: dark blue (top) and gray (bottom)). (D) Same illustration as shown in (C) but for the fifth position of the peptide amino acid sequence.

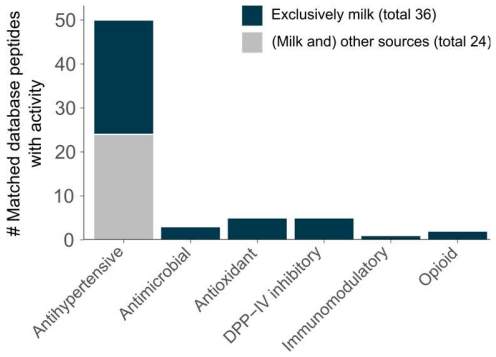


Figure A.S8: Database search reveals broad coverage of bioactivity categories by tryptone peptides. Bars show the number of matched peptides exclusively found in milk (dark blue bars) and found in other sources (gray bars) and compare different bioactivity categories.

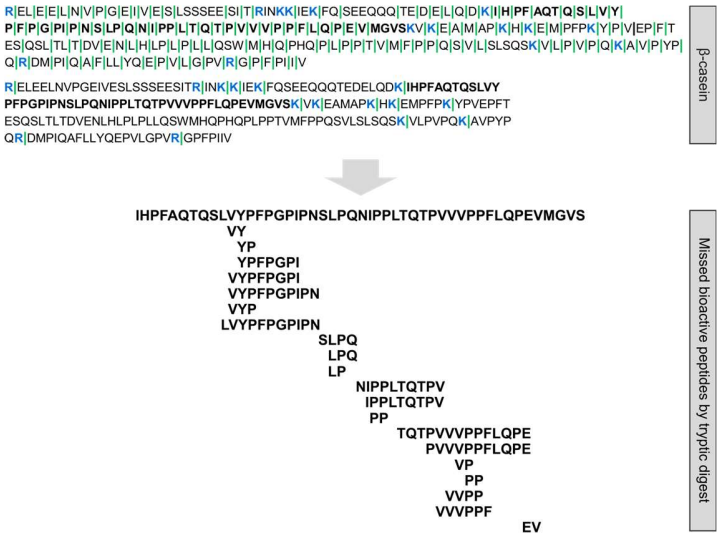


Figure A.S9: Sequence analysis of ß-casein$_{49-112}$ demonstrates omission of several bioactive tryptone peptides after *in silico* tryptic digestion. Comparison of cleavage sites on ß-casein for tryptone peptides (top) and a theoretical digest with trypsin (bottom). Vertical bars (|) indicate cleavage positions. Lysine (K) and arginine (R) are highlighted in blue to indicate trypsin cleavage positions. Bold letters denote ß-casein$_{49-112}$ as produced by trypsin digestion and the corresponding tryptone peptides with established bioactivities.

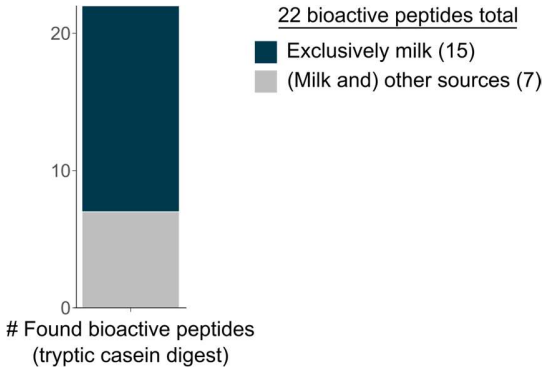


Figure A.S10: Bioactive peptides identified for a theoretical tryptic casein digest. Of 22 peptides with established bioactivities, 15 are exclusively described as bioactive milk peptides (dark blue), and 7 are known bioactive peptides in other sources (gray). For this calculation, α-S1-, α-S2-, β-, and κ-casein were considered.

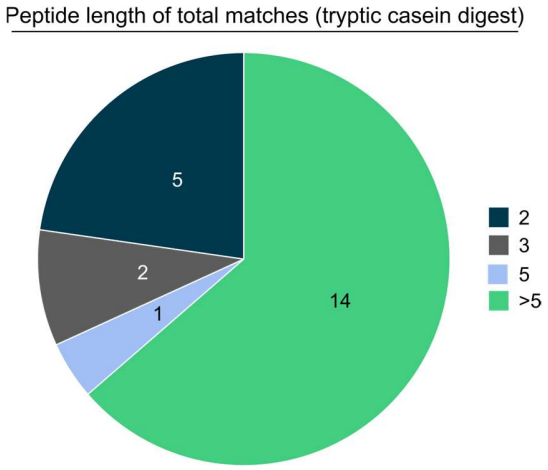


Figure A.S11: Distribution of the peptide length uncovers poor coverage of bioactive short-chain peptides for a theoretical tryptic casein digest. Pie charts show the length of peptides for which a possible bioactivity could be found.

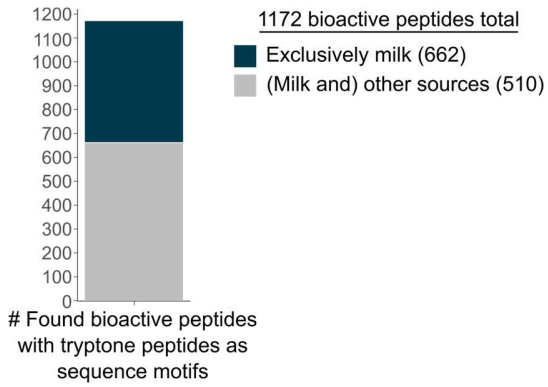


Figure A.S12: Substring matching shows substantial sequence overlap of tryptone peptides and reported bioactive peptides. Establishing sequence overlays between tryptone peptides and 1172 bioactive peptides derived from milk (675) and other sources (510). If a peptide was found in multiple sources, it was counted more than once (exclusively milk: 662; (milk and) other sources: 510).

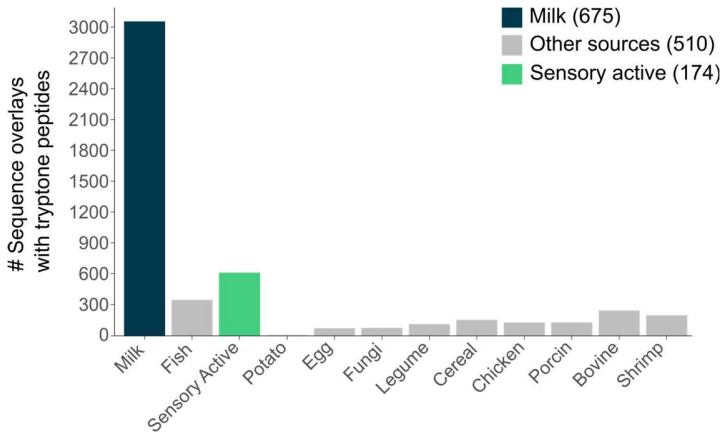


Figure A.S13: Mapping tryptone peptides to established bioactive peptides exposes considerable sequence commonalities with various sources. Number of subsequence overlays between tryptone peptides and peptides that were previously established to be bioactive. Colors indicate different peptide sources.

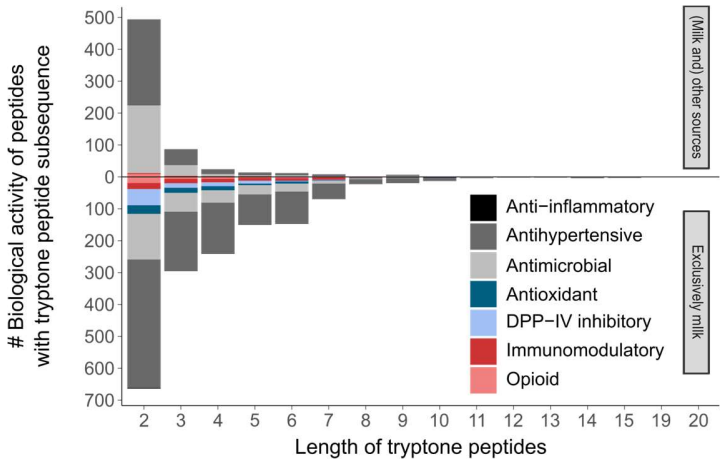


Figure A.S14: Peptides with miscellaneous bioactivities show common subsequences with tryptone peptides. The bar graph compares bioactivity heterogeneity for all the bioactive peptides containing tryptone peptides as a subsequence (exclusively milk: bottom, (also) other sources: top).

A.2 Supplementary tables

Table A.S1A (columns 1-7 and ID): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Peptide	Theoretical mass	Experimental mass	Peptide error [ppm]	Charge	Peptide identification	Detected as Amadori Product	ID
P	115.1	115.1	11.38	1	Mass and MSMS	no	135
V	117.1	117.1	11.24	1	Mass and MSMS	yes	56
K	146.1	146.1	2.93	1	Mass and MSMS	yes	58
M	149.1	149.1	3.11	1	Mass and MSMS	yes	59
F	165.1	165.1	22.08	1	Mass and MSMS	yes	60
R	174.1	174.1	-3.10	1	Mass and MSMS	yes	61
Y	181.1	181.1	4.36	1	Mass and MSMS	yes	62
AV	188.1	188.1	-4.30	1	Mass and MSMS	no	140
VP	214.1	214.1	5.42	1	Mass and MSMS	no	143
AVP	285.2	285.2	1.84	1	Mass and MSMS	yes	73
PYP	375.2	375.2	-0.30	1	Mass and MSMS	no	172
PYPQ	503.2	503.2	-0.07	1	Mass and MSMS	yes	109
VPYPQ	602.3	602.3	-0.39	1	Mass and MSMS	no	189
AVPYPQ	673.3	673.3	0.09	1	Mass and MSMS	yes	25
LP	228.1	228.1	5.12	1	Mass and MSMS	no	146
LPYP	488.3	488.3	0.01	1	Mass and MSMS	no	191

Table A.S1A (columns 1-7 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Peptide	Theoretical mass	Experimental mass	Peptide error [ppm]	Charge	Peptide identification	Detected as Amadori Product	ID
FLPYP	635.3	635.3	-1.19	1	Mass and MSMS	yes	22
PE	244.1	244.1	-0.01	1	Mass and MSMS	no	150
YPE	407.2	407.2	1.19	1	Mass and MSMS	yes	93
FYPE	554.2	554.2	0.34	1	Mass and MSMS	yes	115
LPQ	356.2	356.2	0.41	1	Mass and MSMS	no	169
GLPQ	413.2	413.2	0.83	1	Mass and MSMS	no	158
GLPQEV	641.3	641.3	-0.27	1	Mass and MSMS	yes	130
HQGLPQ	678.3	678.3	2.98	2	Mass and MSMS	yes	26
ISQ	346.2	346.2	1.15	1	Mass and MSMS	yes	81
HIQ	396.2	396.2	-0.60	1	Mass and MSMS	yes	91
IQPK	484.3	484.3	-0.38	1	Mass and MSMS	yes	104
PI	228.1	228.1	5.83	1	Mass and MSMS	no	147
TL	232.1	232.1	4.42	1	Mass and MSMS	yes	67
TPT	317.2	317.2	2.86	1	Mass and MSMS	no	160
PITPT	527.3	527.3	-0.35	1	Mass and MSMS	yes	17
KH	283.2	283.2	-4.81	2	Mass and MSMS	yes	12
ITV	331.2	331.2	0.66	1	Mass and MSMS	yes	78

Table A.S1A (columns 1-7 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Peptide	Theoretical mass	Experimental mass	Peptide error [ppm]	Charge	Peptide identification	Detected as Amadori Product	ID
ITVDD	561.3	561.3	0.03	1	Mass and MSMS	yes	118
ITVDDK	689.4	689.4	1.71	2	Mass and MSMS	yes	30
ITVDDKH	826.4	826.4	1.40	2	Mass and MSMS	yes	39
KE	275.1	275.1	1.13	1	Mass and MSMS	yes	70
VPSE	430.2	430.2	1.09	1	Mass and MSMS	no	181
DVPSE	545.2	545.2	0.11	1	Mass and MSMS	no	180
EDVPS	545.2	545.2	0.88	1	Mass and MSMS	no	199
EDVPSE	674.3	674.3	-1.03	1	Mass and MSMS	no	215
KEDVPSE	802.4	802.4	0.62	2	Mass and MSMS	yes	37
VPQLEIVPN	1007.6	1007.6	-0.36	2	Mass and MSMS	yes	45
KVPQLEIVPN	1135.7	1135.7	-0.09	2	Mass and MSMS	yes	51
EI	260.1	260.1	-2.04	1	Mass and MSMS	no	153
VPQ	342.2	342.2	1.04	1	Mass and MSMS	no	166
IVPN	441.3	441.3	1.42	1	Mass and MSMS	yes	97
KVPQ	470.3	470.3	0.39	1	Mass and MSMS	no	187
LEIV	472.3	472.3	-0.11	1	Mass and MSMS	no	170
EIVPN	570.3	570.3	-0.59	1	Mass and MSMS	yes	120

Table A.S1A (columns 1-7 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Peptide	Theoretical mass	Experimental mass	Peptide error [ppm]	Charge	Peptide identification	Detected as Amadori Product	ID
LEIVPN	683.4	683.4	-1.13	1	Mass and MSMS	yes	27
KVPQLE	712.4	712.4	0.83	2	Mass and MSMS	no	206
YQEPVLGPV	1000.5	1000.5	-2.19	1	Mass and MSMS	yes	2
PV	214.1	214.1	1.05	1	Mass and MSMS	no	142
GPV	271.2	271.2	-0.58	1	Mass and MSMS	no	156
LGPV	384.2	384.2	3.21	1	Mass and MSMS	yes	88
YQEPV	634.3	634.3	-1.25	1	Mass and MSMS	yes	129
LH	268.2	268.2	-0.72	1	Mass and MSMS	no	155
HLP	365.2	365.2	-0.48	1	Mass and MSMS	no	171
LPLP	438.3	438.3	1.06	1	Mass and MSMS	yes	96
LPLPL	551.4	551.4	0.55	1	Mass and MSMS	no	200
HLPLP	575.3	575.3	-0.09	1	Mass and MSMS	yes	19
LHLPLP	688.4	688.4	2.01	2	Mass and MSMS	yes	29
HLPLPL	688.4	688.4	-1.52	1	Mass and MSMS	no	138
LHLPLPL	801.5	801.5	-0.20	2	Mass and MSMS	yes	36
LN	245.1	245.1	1.10	1	Mass and MSMS	yes	68

Table A.S1A (columns 1-7 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Peptide	Theoretical mass	Experimental mass	Peptide error [ppm]	Charge	Peptide identification	Detected as Amadori Product	ID
LNE	374.2	374.2	-0.35	1	Mass and MSMS	yes	83
HQPHQPLPPT	1150.6	1150.6	0.39	2	Mass and MSMS	yes	3
MHQPHQPLPPT	1281.6	1281.6	-0.09	2	Mass and MSMS	yes	4
HQPHQPLPPTVMF PPQSV	2036.0	2036.0	-1.05	3	Mass and MSMS	yes	6
SV	204.1	204.1	-2.26	1	Mass and MSMS	yes	64
PHQ	380.2	380.2	-3.08	1	Mass and MSMS	yes	87
VMFP	492.2	492.2	-8.22	1	Mass and MSMS	no	192
QPHQ	508.2	508.2	-2.92	1	Mass and MSMS	no	193
QPHQP	605.3	605.3	-2.81	1	Mass and MSMS	no	202
FPPQSV	673.3	673.3	-1.12	1	Mass and MSMS	no	198
MFPPQS	705.3	705.3	-1.40	1	Mass and MSMS	yes	31
MFPPQSV	804.4	804.4	-2.85	1	Mass and MSMS	yes	38
NVPGEI	627.3	627.3	-0.97	1	Mass and MSMS	yes	21
NVPGEIV	726.4	726.4	-1.77	1	Mass and MSMS	yes	133
NVPGEIVE	855.4	855.4	-2.48	1	Mass and MSMS	yes	42
GPIVL	497.3	497.3	-1.06	1	Mass and MSMS	yes	15
NQ	260.1	260.1	14.40	1	Mass and MSMS	no	152

Table A.S1A (columns 1-7 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Peptide	Theoretical mass	Experimental mass	Peptide error [ppm]	Charge	Peptide identification	Detected as Amadori Product	ID
IGV	287.2	287.2	0.28	1	Mass	yes	74
MIG	319.2	319.2	0.43	1	Mass and MSMS	no	136
IGVN	401.2	401.2	-0.47	1	Mass and MSMS	yes	92
PMIGVN	629.3	629.3	-6.47	1	Mass and MSMS	yes	128
IGVNQE	658.3	658.3	-0.57	1	Mass and MSMS	no	195
PFPI	472.3	472.3	0.22	1	Mass and MSMS	no	188
GPFPI	529.3	529.3	2.80	1	Mass and MSMS	yes	112
GPFPII	642.4	642.4	-0.49	1	Mass and MSMS	no	208
RGPFPI	685.4	685.4	0.94	2	Mass and MSMS	yes	28
SI	218.1	218.1	1.96	1	Mass and MSMS	yes	65
VVPPFLQPE	1024.6	1024.6	0.70	2	Mass and MSMS	yes	46
NIPPLTQTPV	1078.6	1078.6	0.18	2	Mass and MSMS	yes	47
TQTPVVVPPF	1083.6	1083.6	0.32	2	Mass and MSMS	yes	48
VVPPFLQPEV	1123.6	1123.6	2.11	2	Mass and MSMS	yes	50
PVVVPPFLQPE	1220.7	1220.7	-0.24	2	Mass and MSMS	no	212
VVVPPFLQPEV	1222.7	1222.7	-0.57	2	Mass and MSMS	no	211
PVVVPPFLQPEV	1319.7	1319.7	-0.41	2	Mass and MSMS	no	214

Table A.S1A (columns 1-7 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Peptide	Theoretical mass	Experimental mass	Peptide error [ppm]	Charge	Peptide identification	Detected as Amadori Product	ID
TPVVVPPFLQPE	1321.7	1321.7	-0.35	2	Mass and MSMS	yes	52
TPVVVPPFLQPEV	1420.8	1420.8	-1.71	2	Mass and MSMS	no	213
NIPPLTQTPVVVPP	1470.8	1470.8	-1.77	2	Mass and MSMS	yes	53
SLPQNIPPLTQTPV	1503.8	1503.8	-1.30	2	Mass and MSMS	yes	54
QTPVVVPPFLQPEV	1548.9	1548.9	-2.41	2	Mass and MSMS	no	210
TQTPVVVPPFLQPE	1550.8	1550.8	-1.69	2	Mass and MSMS	no	163
NIPPLTQTPVVVPPF	1617.9	1617.9	-1.58	2	Mass and MSMS	yes	55
TQTPVVVPPFLQPE V	1649.9	1649.9	-1.44	2	Mass and MSMS	yes	5
IPPLTQTPVVVPPFL QPEV	2070.2	2070.2	-1.65	3	Mass and MSMS	yes	7
NIPPLTQTPVVVPPF LQPE	2085.2	2085.1	-1.08	3	Mass and MSMS	yes	8
PP	212.1	212.1	-1.35	1	Mass and MSMS	no	141
NIPPLTQTPVVVPPF LQPEV	2184.2	2184.2	-1.11	3	Mass and MSMS	yes	9
EV	246.1	246.1	0.53	1	Mass and MSMS	yes	69

Table A.S1A (columns 1-7 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Peptide	Theoretical mass	Experimental mass	Peptide error [ppm]	Charge	Peptide identification	Detected as Amadori Product	ID
SLPQNIPPLTQTPV VVPPFLQPE	2510.4	2510.4	-0.68	3	Mass and MSMS	yes	10
SLPQNIPPLTQTPV VVPPFLQPEV	2609.4	2609.4	-0.67	3	Mass and MSMS	yes	11
PFL	375.2	375.2	-0.52	1	Mass and MSMS	no	173
VVPP	410.3	410.3	1.19	1	Mass and MSMS	no	179
SLPQ	443.2	443.2	1.84	1	Mass and MSMS	yes	98
LQPE	485.2	485.2	0.78	1	Mass and MSMS	yes	105
TQTPV	544.3	544.3	0.41	1	Mass and MSMS	yes	113
NIPPL	552.3	552.3	-0.24	1	Mass and MSMS	yes	114
SLPQN	557.3	557.3	6.69	1	Mass and MSMS	no	194
VVPPF	557.3	557.3	-0.34	1	Mass and MSMS	yes	116
PFLQP	600.3	600.3	-0.95	1	Mass and MSMS	no	201
VVVPPF	656.4	656.4	-1.70	1	Mass and MSMS	no	204
VVPPFL	670.4	670.4	-1.32	1	Mass and MSMS	no	203
IPPLTQTPV	964.6	964.6	0.30	2	Mass and MSMS	no	207
AP	186.1	186.1	-0.85	1	Mass and MSMS	no	139
VAPF	432.2	410.3	0.41	1	Mass and MSMS	no	161
VAPFPE	658.3	658.3	-0.63	1	Mass and MSMS	yes	131

Table A.S1A (columns 1-7 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Peptide	Theoretical mass	Experimental mass	Peptide error [ppm]	Charge	Peptide identification	Detected as Amadori Product	ID
APFPEV	658.3	658.3	-1.62	1	Mass and MSMS	no	209
VAPFPEV	757.4	757.4	-0.63	1	Mass and MSMS	yes	34
KP	243.2	243.2	1.01	1	Mass and MSMS	no	149
VLP	327.2	327.2	0.61	1	Mass and MSMS	yes	77
VLPV	426.3	426.3	1.01	1	Mass and MSMS	yes	94
PVPQ	439.2	439.2	1.39	1	Mass and MSMS	no	182
VLPVPQ	651.4	651.4	-0.90	1	Mass and MSMS	yes	24
VYPFPGPIPN	1099.6	1099.6	0.24	2	Mass and MSMS	yes	49
YP	278.1	278.1	-1.89	1	Mass and MSMS	yes	71
VY	280.1	280.1	-0.53	1	Mass and MSMS	yes	72
VYP	377.2	377.2	0.02	1	Mass and MSMS	yes	86
GPIP	382.2	382.2	-1.27	1	Mass and MSMS	no	175
PFPGPI	626.3	626.3	-0.91	1	Mass and MSMS	yes	20
PFPGPIPN	837.4	837.4	-2.67	1	Mass and MSMS	yes	40
VYPFPGPI	888.5	888.5	-3.25	1	Mass and MSMS	yes	43
VEP	343.2	343.2	1.58	1	Mass and MSMS	yes	80
EPF	391.2	391.2	1.14	1	Mass and MSMS	no	177
VEPF	490.2	490.2	-0.97	1	Mass and MSMS	no	174
YPVEP	603.3	603.3	0.39	1	Mass and MSMS	yes	124

Table A.S1A (columns 1-7 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Peptide	Theoretical mass	Experimental mass	Peptide error [ppm]	Charge	Peptide identification	Detected as Amadori Product	ID
YPVEPF	750.4		750.4	-1.47	1	Mass and MSMS	yes
YV	280.1		280.1	-0.62	1	Mass and MSMS	no
YVPL	490.3		490.3	0.70	1	Mass and MSMS	yes
AI	202.1		202.1	-0.41	1	Mass and MSMS	yes
PPK	340.2		340.2	1.42	1	Mass and MSMS	no
AIPPK	524.3		524.3	0.06	1	Mass and MSMS	yes
AM	220.1		220.1	5.45	1	Mass and MSMS	no
AMED	464.2		464.2	0.86	1	Mass and MSMS	yes
AMEDI	577.2		577.2	-0.64	1	Mass and MSMS	yes
AMEDIK	705.3		705.3	2.02	2	Mass and MSMS	no
NELS	461.2		461.2	0.89	1	Mass and MSMS	yes
FQ	293.1		293.1	-0.16	1	Mass and MSMS	yes
ESPPEIN	784.4		784.4	-2.13	1	Mass and MSMS	yes
IEK	388.2		388.2	-1.38	1	Mass and MSMS	yes
DMPI	474.2		474.2	0.68	1	Mass and MSMS	yes
AKPA	385.2		385.2	-0.45	1	Mass and MSMS	yes
MAPK	445.2		445.2	0.53	1	Mass and MSMS	yes
EAMAPK	645.3		645.3	4.33	2	Mass and MSMS	yes
KQEK	531.3		531.3	-0.23	2	Mass and MSMS	yes

Table A.S1A (columns 1-7 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Peptide	Theoretical mass	Experimental mass	Peptide error [ppm]	Charge	Peptide identification	Detected as Amadori Product	ID
LTEEE	619.3	619.3	-0.26	1	Mass and MSMS	yes	127
EPM	375.1	375.1	-0.16	1	Mass and MSMS	yes	85
TTMP	448.2	448.2	0.99	1	Mass and MSMS	no	183
TMPL	460.2	460.2	-0.03	1	Mass and MSMS	no	185
TTMPL	561.3	561.3	0.68	1	Mass and MSMS	yes	119
ELEE	518.2	518.2	-0.68	1	Mass and MSMS	yes	110
AEERL	616.3	616.3	5.42	1	Mass and MSMS	yes	125
RELEE	674.3	674.3	3.46	2	Mass and MSMS	yes	132
RPK	399.3	399.3	1.73	2	Mass and MSMS	yes	13
ASGEPT	560.2	560.2	0.65	1	Mass and MSMS	yes	117
DKI	374.2	374.2	-0.69	1	Mass and MSMS	yes	84
TEDE	492.2	492.2	-0.04	1	Mass and MSMS	yes	108
DELQD	618.2	618.2	0.00	1	Mass and MSMS	no	205
TEDELQD	848.3	848.3	-1.77	1	Mass and MSMS	yes	41
TEDELQDK	976.4	976.4	-0.44	2	Mass and MSMS	yes	44
NAVPI	512.3	512.3	-0.62	1	Mass and MSMS	yes	16
EKVNE	617.3	617.3	-1.23	1	Mass and MSMS	yes	126
SLV	317.2	317.2	1.14	1	Mass and MSMS	yes	76
NTVP	429.2	429.2	1.57	1	Mass and MSMS	yes	95

Table A.S1A (columns 1-7 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Peptide	Theoretical mass	Experimental mass	Peptide error [ppm]	Charge	Peptide identification	Detected as Amadori Product	ID
TDV	333.2	333.2	2.33	1	Mass and MSMS	yes	79
DVE	361.1	361.1	1.72	1	Mass and MSMS	no	144
TDVE	462.2	462.2	-0.14	1	Mass and MSMS	yes	100
DVEN	475.2	475.2	-0.17	1	Mass and MSMS	yes	103
TDVEN	576.2	576.2	0.50	1	Mass and MSMS	yes	121
TVPA	386.2	386.2	-0.42	1	Mass and MSMS	yes	89
VNE	360.2	360.2	1.08	1	Mass and MSMS	yes	82
SF	252.1	252.1	-0.15	1	Mass and MSMS	no	151
DIPN	457.2	457.2	-0.03	1	Mass and MSMS	no	184
TDAPS	489.2	489.2	0.10	1	Mass and MSMS	yes	106
FSDIP	577.3	577.3	5.62	1	Mass and MSMS	yes	123
DIPNPI	667.4	667.4	-0.23	1	Mass and MSMS	no	197
TDAPSFS	723.3	723.3	-1.81	1	Mass and MSMS	no	186
SDIPNPI	754.4	754.4	-1.57	1	Mass and MSMS	yes	33
I, L*	131.1	131.1	14.24, -9.19	1	Mass and MSMS, Mass	yes	57
VI, VL*	230.2	230.2	3.44, 3.96	1	Mass and MSMS, Mass and MSMS	yes	66

Table A.S1A (columns 1-7 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Peptide	Theoretical mass	Experimental mass	Peptide error [ppm]	Charge	Peptide identification	Detected as Amadori Product	ID
EMPFPK, LTEEEK	747.4	747.4	-1.83, 1.34	1, 2	Mass and MSMS, Mass and MSMS	yes	32

Table A.S1B (columns 8-16 and ID): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Amadori product theoretical mass	Amadori product experimental mass	Amadori product error [ppm]	Amadori product charge	Amadori product identification	Significance time	Significance $c_{Glucose}$	p value	Sequence group	ID
NA	NA	NA	NA	NA	NA	NA	NA	Amino acid	135
279.1318	279.1311	2.5576	1	Mass and MSMS	2 h	2.7 mg/mL	6.02E-05	Amino acid	56
308.1584	308.1593	-2.9998	1	Mass and MSMS	2 h	2.7 mg/mL	6.80E-04	Amino acid	58
311.1039	311.1045	-2.0459	1	Mass and MSMS	2 h	2.7 mg/mL	4.34E-03	Amino acid	59
327.1318	327.1327	-2.8288	1	Mass and MSMS	2 h	2.7 mg/mL	1.03E-04	Amino acid	60
336.1645	336.1655	-3.0702	1	Mass and MSMS	2 h	2.7 mg/mL	4.16E-02	Amino acid	61
343.1267	343.1279	-3.4080	1	Mass	NA	NA	NA	Amino acid	62
NA	NA	NA	NA	NA	NA	NA	NA	1	140
NA	NA	NA	NA	NA	NA	NA	NA	1	143
447.2217	447.2207	2.1916	1	Mass	NA	NA	NA	1	73
NA	NA	NA	NA	NA	NA	NA	NA	1	172
665.2908	665.2902	0.9488	1	Mass	NA	NA	NA	1	109
NA	NA	NA	NA	NA	NA	NA	NA	1	189
835.3964	835.3979	-1.8834	2	Mass and MSMS	2 h	27 mg/mL	3.29E-02	1	25
NA	NA	NA	NA	NA	NA	NA	NA	5	146
NA	NA	NA	NA	NA	NA	NA	NA	5	191

Table A.S1B (columns 8-16 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Amadori product theoretical mass	Amadori product experimental mass	Amadori product error [ppm]	Amadori product charge	Amadori product identification	Significance time	Significance $c_{Glucose}$	p value	Sequence group	ID
797.3847	797.3844	0.3929	2	Mass and MSMS	2 h	2.7 mg/mL	2.00E-05	5	22
NA	NA	NA	NA	NA	NA	NA	NA	6	150
569.2221	569.2212	1.5227	1	Mass	NA	NA	NA	6	93
716.2905	716.2896	1.2888	1	Mass and MSMS	2 h	2.7 mg/mL	3.61E-02	6	115
NA	NA	NA	NA	NA	NA	NA	NA	7	169
NA	NA	NA	NA	NA	NA	NA	NA	7	158
803.3913	803.3891	2.7026	1	Mass	NA	NA	NA	7	130
840.3977	840.3995	-2.0664	2	Mass and MSMS	2 h	27 mg/mL	6.97E-05	7	26
508.2381	508.2382	-0.3027	1	Mass and MSMS	2 h	2.7 mg/mL	7.11E-03	8	81
558.2649	558.2648	0.2695	1	Mass and MSMS	2 h	2.7 mg/mL	1.15E-05	9	91
646.3538	646.3538	-0.1293	1	Mass	NA	NA	NA	10	104
NA	NA	NA	NA	NA	NA	NA	NA	11	147
394.1951	394.1950	0.3053	1	Mass and MSMS	2 h	5.4 mg/mL	8.61E-04	11	67
NA	NA	NA	NA	NA	NA	NA	NA	11	160
689.3483	689.3423	8.8090	2	Mass	NA	NA	NA	11	17
445.2173	445.2191	-4.1830	2	Mass	NA	NA	NA	12	12
493.2635	493.2631	0.8615	1	Mass	NA	NA	NA	12	78

Table A.S1B (columns 8-16 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Amadori product theoretical mass	Amadori product experimental mass	Amadori product error [ppm]	Amadori product charge	Amadori product identification	Significance time	Significance $c_{Glucose}$	p value	Sequence group	ID
723.3174	723.3151	3.2737	1	Mass	NA	NA	NA	12	118
851.4124	851.4134	-1.1811	2	Mass and MSMS	2 h	27 mg/mL	9.93E-05	12	30
988.4713	988.4708	0.4748	2	Mass	NA	NA	NA	12	39
437.2009	437.2019	-2.2361	1	Mass	NA	NA	NA	13	70
NA	NA	NA	NA	NA	NA	NA	NA	13	181
NA	NA	NA	NA	NA	NA	NA	NA	13	180
NA	NA	NA	NA	NA	NA	NA	NA	13	199
NA	NA	NA	NA	NA	NA	NA	NA	13	215
964.4237	964.4235	0.1926	2	Mass and MSMS	2 h	2.7 mg/mL	2.12E-02	13	37
1169.6180	1169.6182	-0.2161	2	Mass and MSMS	2 h	27 mg/mL	1.10E-03	14	45
1297.7129	1297.7128	0.0842	2	Mass and MSMS	2 h	5.4 mg/mL	1.04E-03	14	51
NA	NA	NA	NA	NA	NA	NA	NA	14	153
NA	NA	NA	NA	NA	NA	NA	NA	14	166
603.3116	603.3110	0.8642	1	Mass and MSMS	2 h	54 mg/mL	3.73E-03	14	97
NA	NA	NA	NA	NA	NA	NA	NA	14	187
NA	NA	NA	NA	NA	NA	NA	NA	14	170
732.3542	732.3520	2.9111	1	Mass	NA	NA	NA	14	120

Table A.S1B (columns 8-16 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Amadori product theoretical mass	Amadori product experimental mass	Amadori product error [ppm]	Amadori product charge	Amadori product identification	Significance time	Significance $c_{Glucose}$	p value	Sequence group	ID
845.4382	845.4391	-1.0130	2	Mass	NA	NA	NA	14	27
NA	NA	NA	NA	NA	NA	NA	NA	14	206
1162.5758	1162.5856	-8.4449	3	Mass	NA	NA	NA	15	2
NA	NA	NA	NA	NA	NA	NA	NA	15	142
NA	NA	NA	NA	NA	NA	NA	NA	15	156
546.2901	546.2901	-0.0618	1	Mass and MSMS	2 h	2.7 mg/mL	2.22E-02	15	88
796.3491	796.3449	5.2078	1	Mass	NA	NA	NA	15	129
NA	NA	NA	NA	NA	NA	NA	NA	16	155
NA	NA	NA	NA	NA	NA	NA	NA	16	171
600.3370	600.3365	0.8456	1	Mass	NA	NA	NA	16	96
NA	NA	NA	NA	NA	NA	NA	NA	16	200
737.3960	737.3965	-0.7183	2	Mass and MSMS	2 h	2.7 mg/mL	1.88E-06	16	19
850.4800	850.4810	-1.0941	2	Mass and MSMS	2 h	2.7 mg/mL	1.12E-05	16	29
NA	NA	NA	NA	NA	NA	NA	NA	16	138
963.5641	963.5627	1.4081	2	Mass	NA	NA	NA	16	36
407.1904	407.1905	-0.3652	1	Mass and MSMS	2 h	2.7 mg/mL	1.18E-02	17	68

Table A.S1B (columns 8-16 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Amadori product theoretical mass	Amadori product experimental mass	Amadori product error [ppm]	Amadori product charge	Amadori product identification	Significance time	Significance $c_{Glucose}$	p value	Sequence group	ID
536.2330	536.2324	1.1103	1	Mass	NA	NA	NA	17	83
1312.6412	1312.6437	-1.9006	3	Mass and MSMS	2 h	2.7 mg/mL	6.13E-05	18	3
1443.6817	1443.6822	-0.3753	3	Mass	NA	NA	NA	18	4
2198.0830	2198.0764	3.0179	3	Mass	NA	NA	NA	18	6
366.1638	366.1634	1.1798	1	Mass	NA	NA	NA	18	64
542.2336	542.2335	0.3521	1	Mass	NA	NA	NA	18	87
NA	NA	NA	NA	NA	NA	NA	NA	18	192
NA	NA	NA	NA	NA	NA	NA	NA	18	193
NA	NA	NA	NA	NA	NA	NA	NA	18	202
NA	NA	NA	NA	NA	NA	NA	NA	18	198
867.3684	867.3692	-0.8855	2	Mass	NA	NA	NA	18	31
966.4368	966.4368	0.0425	2	Mass	NA	NA	NA	18	38
789.3756	789.3715	5.2297	2	Mass	NA	NA	NA	19	21
888.4440	888.4437	0.3240	1	Mass	NA	NA	NA	19	133
1017.4866	1017.4868	-0.2067	2	Mass	NA	NA	NA	19	42
659.3742	659.3700	6.3538	2	Mass	NA	NA	NA	20	15
NA	NA	NA	NA	NA	NA	NA	NA	21	152

Table A.S1B (columns 8-16 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Amadori product theoretical mass	Amadori product experimental mass	Amadori product error [ppm]	Amadori product charge	Amadori product identification	Significance time	Significance $c_{Glucose}$	p value	Sequence group	ID
449.2373	449.2373	0.0300	1	Mass and MSMS	2 h	2.7 mg/mL	7.87E-04	21	74
NA	NA	NA	NA	NA	NA	NA	NA	21	136
563.2803	563.2801	0.3404	1	Mass	NA	NA	NA	21	92
791.3735	791.3685	6.3527	1	Mass	NA	NA	NA	21	128
NA	NA	NA	NA	NA	NA	NA	NA	21	195
NA	NA	NA	NA	NA	NA	NA	NA	22	188
691.3429	691.3419	1.3883	1	Mass and MSMS	2 h	2.7 mg/mL	2.18E-03	22	112
NA	NA	NA	NA	NA	NA	NA	NA	22	208
847.4440	847.4409	3.5959	2	Mass	NA	NA	NA	22	28
380.1795	380.1798	-0.7355	1	Mass and MSMS	2 h	5.4 mg/mL	2.33E-04	23	65
1186.6122	1186.6107	1.2452	2	Mass and MSMS	2 h	2.7 mg/mL	1.12E-03	24	46
1240.6551	1240.6547	0.3393	2	Mass and MSMS	2 h	5.4 mg/mL	2.47E-03	24	47
1245.6493	1245.6488	0.3703	2	Mass	NA	NA	NA	24	48
1285.6806	1285.6804	0.1691	2	Mass and MSMS	2 h	5.4 mg/mL	3.55E-03	24	50
NA	NA	NA	NA	NA	NA	NA	NA	24	212
NA	NA	NA	NA	NA	NA	NA	NA	24	211
NA	NA	NA	NA	NA	NA	NA	NA	24	214

Table A.S1B (columns 8-16 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Amadori product theoretical mass	Amadori product experimental mass	Amadori product error [ppm]	Amadori product charge	Amadori product identification	Significance time	Significance $c_{Glucose}$	p value	Sequence group	ID
1483.7810	1483.7780	2.0035	2	Mass	NA	NA	NA	24	52
NA	NA	NA	NA	NA	NA	NA	NA	24	213
1632.8974	1632.8916	3.5909	2	Mass	NA	NA	NA	24	53
1665.8825	1665.8788	2.2509	2	Mass	NA	NA	NA	24	54
NA	NA	NA	NA	NA	NA	NA	NA	24	210
NA	NA	NA	NA	NA	NA	NA	NA	24	163
1779.9659	1779.9599	3.3200	2	Mass	NA	NA	NA	24	55
1811.9557	1811.9513	2.4132	3	Mass	NA	NA	NA	24	5
2232.2293	2232.2248	2.0443	3	Mass	NA	NA	NA	24	7
2247.2039	2247.2006	1.4528	3	Mass	NA	NA	NA	24	8
NA	NA	NA	NA	NA	NA	NA	NA	24	141
2346.2723	2346.2680	1.8398	3	Mass and MSMS	2 h	54 mg/mL	6.96E-03	24	9
408.1744	408.1745	-0.1911	1	Mass	NA	NA	NA	24	69
2672.4313	2672.4283	1.1045	3	Mass	NA	NA	NA	24	10
2771.4997	2771.4980	0.6201	3	Mass	NA	NA	NA	24	11
NA	NA	NA	NA	NA	NA	NA	NA	24	173
NA	NA	NA	NA	NA	NA	NA	NA	24	179

Table A.S1B (columns 8-16 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Amadori product theoretical mass	Amadori product experimental mass	Amadori product error [ppm]	Amadori product charge	Amadori product identification	Significance time	Significance $c_{Glucose}$	p value	Sequence group	ID
605.2908	605.2905	0.5978	1	Mass and MSMS	2 h	27 mg/mL	4.57E-05	24	98
647.3014	647.2950	9.9026	1	Mass	NA	NA	NA	24	105
706.3385	706.3378	0.9640	1	Mass	NA	NA	NA	24	113
714.3800	714.3787	1.7153	1	Mass and MSMS	2 h	2.7 mg/mL	2.42E-05	24	114
NA	NA	NA	NA	NA	NA	NA	NA	24	194
719.3742	719.3730	1.6317	1	Mass and MSMS	2 h	2.7 mg/mL	2.17E-03	24	116
NA	NA	NA	NA	NA	NA	NA	NA	24	201
NA	NA	NA	NA	NA	NA	NA	NA	24	204
NA	NA	NA	NA	NA	NA	NA	NA	24	203
NA	NA	NA	NA	NA	NA	NA	NA	24	207
NA	NA	NA	NA	NA	NA	NA	NA	25	139
NA	NA	NA	NA	NA	NA	NA	NA	25	161
820.3855	820.3836	2.2782	1	Mass and MSMS	2 h	27 mg/mL	9.16E-04	25	131
NA	NA	NA	NA	NA	NA	NA	NA	25	209
919.4539	919.4545	-0.7222	2	Mass and MSMS	2 h	5.4 mg/mL	1.98E-03	25	34
NA	NA	NA	NA	NA	NA	NA	NA	27	149
489.2686	489.2684	0.5705	1	Mass and MSMS	2 h	27 mg/mL	3.18E-03	28	77

Table A.S1B (columns 8-16 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Amadori product theoretical mass	Amadori product experimental mass	Amadori product error [ppm]	Amadori product charge	Amadori product identification	Significance time	Significance $c_{Glucose}$	p value	Sequence group	ID
588.3370	588.3360	1.7453	1	Mass and MSMS	2 h	2.7 mg/mL	2.06E-05	28	94
NA	NA	NA	NA	NA	NA	NA	NA	28	182
813.4484	813.4482	0.2352	2	Mass and MSMS	2 h	27 mg/mL	1.03E-02	28	24
1261.6231	1261.6223	0.5989	2	Mass and MSMS	2 h	2.7 mg/mL	1.92E-07	29	49
440.1795	440.1795	0.0115	1	Mass	NA	NA	NA	29	71
442.1951	442.1951	0.1012	1	Mass	NA	NA	NA	29	72
539.2479	539.2479	-0.0496	1	Mass	NA	NA	NA	29	86
NA	NA	NA	NA	NA	NA	NA	NA	29	175
788.3956	788.3998	-5.3353	2	Mass	NA	NA	NA	29	20
999.4913	999.4909	0.4597	2	Mass	NA	NA	NA	29	40
1050.5274	1050.5285	-1.1154	2	Mass and MSMS	2 h	2.7 mg/mL	4.96E-04	29	43
505.2272	505.2269	0.5513	1	Mass and MSMS	2 h	27 mg/mL	7.58E-05	30	80
NA	NA	NA	NA	NA	NA	NA	NA	30	177
NA	NA	NA	NA	NA	NA	NA	NA	30	174
765.3433	765.3419	1.7049	1	Mass and MSMS	2 h	27 mg/mL	5.70E-04	30	124
912.4117	912.4093	2.5790	1	Mass and MSMS	2 h	5.4 mg/mL	7.06E-06	30	134
NA	NA	NA	NA	NA	NA	NA	NA	31	157

Table A.S1B (columns 8-16 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Amadori product theoretical mass	Amadori product experimental mass	Amadori product error [ppm]	Amadori product charge	Amadori product identification	Significance time	Significance $c_{Glucose}$	p value	Sequence group	ID
652.3320	652.3307	1.9261	1	Mass and MSMS	2 h	27 mg/mL	2.26E-03	31	107
364.1846	364.1810	9.8999	1	Mass	NA	NA	NA	33	63
NA	NA	NA	NA	NA	NA	NA	NA	33	165
686.3851	686.3843	1.1722	1	Mass	NA	NA	NA	33	111
NA	NA	NA	NA	NA	NA	NA	NA	34	145
626.2105	626.2100	0.9006	1	Mass	NA	NA	NA	34	101
739.2946	739.2930	2.1486	1	Mass	NA	NA	NA	34	122
NA	NA	NA	NA	NA	NA	NA	NA	34	216
623.2650	623.2636	2.2029	1	Mass	NA	NA	NA	35	99
455.1904	455.1896	1.7562	1	Mass	NA	NA	NA	36	75
946.4131	946.4109	2.3891	2	Mass	NA	NA	NA	40	35
550.2850	550.2845	0.8359	1	Mass	NA	NA	NA	41	90
636.2676	636.2667	1.5268	1	Mass	NA	NA	NA	42	102
709.3382	709.3390	-1.0398	3	Mass	NA	NA	NA	43	1
607.2887	607.2901	-2.2641	2	Mass	NA	NA	NA	43	14
807.3684	807.3693	-1.0393	2	Mass	NA	NA	NA	43	23
693.3545	693.3558	-1.8572	2	Mass	NA	NA	NA	45	18

Table A.S1B (columns 8-16 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Amadori product theoretical mass	Amadori product experimental mass	Amadori product error [ppm]	Amadori product charge	Amadori product identification	Significance time	Significance $c_{Glucose}$	p value	Sequence group	ID
781.3229	781.3215	1.7570	1	Mass	NA	NA	NA	45	127
537.1992	537.1988	0.7288	1	Mass	NA	NA	NA	46	85
NA	NA	NA	NA	NA	NA	NA	NA	47	183
NA	NA	NA	NA	NA	NA	NA	NA	47	185
723.3361	723.3343	2.3708	1	Mass	NA	NA	NA	47	119
680.2752	680.2731	3.1003	1	Mass	NA	NA	NA	48	110
778.3709	778.3729	-2.6556	1	Mass	NA	NA	NA	48	125
836.3763	836.3740	2.7883	1	Mass	NA	NA	NA	48	132
561.3122	561.3118	0.8024	2	Mass	NA	NA	NA	49	13
722.2970	722.2951	2.6576	1	Mass	NA	NA	NA	50	117
536.2694	536.2691	0.5124	1	Mass	NA	NA	NA	51	84
654.2232	654.2223	1.4277	1	Mass	NA	NA	NA	51	108
NA	NA	NA	NA	NA	NA	NA	NA	51	205
1010.3928	1010.3923	0.4898	2	Mass	NA	NA	NA	51	41
1138.4877	1138.4879	-0.1398	2	Mass	NA	NA	NA	51	44
674.3487	674.3433	7.9795	2	Mass	NA	NA	NA	53	16
779.3549	779.3525	3.0813	1	Mass	NA	NA	NA	54	126

Table A.S1B (columns 8-16 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Amadori product theoretical mass	Amadori product experimental mass	Amadori product error [ppm]	Amadori product charge	Amadori product identification	Significance time	Significance $c_{Glucose}$	p value	Sequence group	ID
479.2479	479.2478	0.2376	1	Mass	NA	NA	NA	55	76
591.2752	591.2746	0.8993	1	Mass	NA	NA	NA	56	95
495.2064	495.2059	1.0315	1	Mass	NA	NA	NA	58	79
NA	NA	NA	NA	NA	NA	NA	NA	58	144
624.2490	624.2500	-1.5387	1	Mass	NA	NA	NA	58	100
637.2443	637.2444	-0.2791	1	Mass	NA	NA	NA	58	103
738.2919	738.2900	2.6677	1	Mass	NA	NA	NA	58	121
548.2694	548.2692	0.2121	1	Mass	NA	NA	NA	59	89
522.2173	522.2174	-0.0526	1	Mass	NA	NA	NA	61	82
NA	NA	NA	NA	NA	NA	NA	NA	62	151
NA	NA	NA	NA	NA	NA	NA	NA	62	184
651.2599	651.2609	-1.4802	1	Mass	NA	NA	NA	62	106
739.3276	739.3295	-2.5755	1	Mass	NA	NA	NA	62	123
NA	NA	NA	NA	NA	NA	NA	NA	62	197
NA	NA	NA	NA	NA	NA	NA	NA	62	186
916.4389	916.4390	-0.0977	2	Mass	NA	NA	NA	62	33
293.1500	293.1468	2.21	1	Mass and MSMS	2 h	2.7 mg/mL	1.57E-03	24	57

Table A.S1B (columns 8-16 and ID, continued): Identified amino compounds (peptides and amino acids). Peptides were assigned to sequence groups based on common amino acid sequences (column: Sequence Group), and only those sequence groups are shown for which an Amadori product was detected. Further, the characteristics of the corresponding Amadori products are illustrated. The Amadori products of 47 amino compounds could be confirmed by MS/MS fragmentation (column Amadori Product Identification: Mass and MSMS). While we could unambiguously detect the corresponding Amadori product for 43 amino compounds, four amino compounds led to only two Amadori products that could not be unambiguously assigned by MS/MS (both peptides/amino acids are listed in the table and indicated by an asterisk, respectively).

Amadori product theoretical mass	Amadori product experimental mass	Amadori product error [ppm]	Amadori product charge	Amadori product identification	Significance time	Significance $c_{Glucose}$	p value	Sequence group	ID
392.2200	392.2157	0.33	1	Mass and MSMS	2 h	2.7 mg/mL	4.40E-03	20, 28	66
909.4200	909.4158	-0.52, 2.23	2	Mass	NA	NA	NA	44, 45	32

Table A.S2: Dissimilarity of important glycation patterns (length = 3) from Figure 2.4.

ID pattern group	Patterns	Sequence order change	Dissimilarity indices
1	PFP, PYP	no	*D*(Phe F; Tyr Y) = 13
2	PEV, PEL	no	*D*(Val V; Leu L) = 9
2	PEV, VPQ	yes	*D*(Gln E; Glu Q) = 14
2	PEV, VPN	yes	*D*(Gln E; Asp N) = 19
2	PEV, IVE	yes	*D*(Pro P; Ile I) = 24
2	PEV, VLN	yes	*D*(Pro P; Leu L) = 23; *D*(Gln E; Asp N) = 19
2	PEL, IVE	yes	*D*(Pro P; Ile I) = 24
2	PEL, VPN	yes	*D*(Gln E; Asp N) = 19; *D*(Leu L; Val V) = 9
2	PEL, VPQ	yes	*D*(Gln E; Glu Q) = 14; *D*(Leu L; Val V) = 9
2	PEL, VLN	yes	*D*(Gln E; Asp N) = 19; *D*(Pro P; Val V) = 20
2	VPQ, VPN	no	*D*(Glu Q; Asp N) = 10
2	VPQ, VLN	no	*D*(Glu Q; Asp N) = 10; *D*(Pro P; Leu L) = 23
2	VPQ, IVE	no	*D*(Glu Q; Gln E) = 14; *D*(Pro P; Ile I) = 24
2	VPN, VLN	no	*D*(Pro; Leu L) = 23
2	VPN, IVE	no	*D*(Pro P; Val V) = 20; *D*(Glu N; Gln E) = 19
2	VLN, IVE	no	*D*(Leu L; Ile I) = 5; *D*(Glu N; Gln E) = 19
3	PIP, PLP	no	*D*(Ile I; Leu L) = 5
3	PIP, PVP	no	*D*(Ile I; Val V) = 7
3	PLP, PVP	no	*D*(Leu L; Val V) = 9

Table A.S3: Bioactive peptide identities from Figure 2.5.

Peptide	Activity	Source
AF	Antihypertensive	Cereal, milk, fish, legume, chicken, porcin
AI	Antihypertensive	Cereal, legume, porcin
AP	Antihypertensive	Cereal, milk, fish, legume, chicken
AV	Antihypertensive	Fish
AVP	Antihypertensive	Cereal, milk
AVPYPQ	Antihypertensive	Milk
EAMAPK	Antimicrobial	Milk
EI	Antihypertensive	Cereal, milk, chicken
EL	Antioxidant	Milk
EMPFPK	Antihypertensive, Antimicrobial	Milk
EV	Antihypertensive	Cereal, milk, chicken, porcin
FVAPFP	Antihypertensive	Milk
GPFPIIV	Antihypertensive	Milk
GPV	Antihypertensive	Cereal, milk, bovine
HLPLP	Antihypertensive	Milk, Egg
HLPLPL	Antihypertensive	Milk
IPPLTQTPV	Antihypertensive	Milk
KE	Antihypertensive	Cereal, milk, chicken, porcin
KP	Antihypertensive	Cereal, milk, fish, porcin
LF	Antihypertensive	Cereal, milk, fish, chicken, porcin
LHLPLP	Antihypertensive	Milk
LHLPLPL	Antihypertensive	Milk
LN	Antihypertensive	Cereal, milk, legume, chicken, porcin
LP	Antihypertensive, DPP-IV inhibitory	Milk
LPL	DPP-IV inhibitory	Milk
LPLP	Antihypertensive	Milk
LPLPL	DPP-IV inhibitory	Milk
LPQ	DPP-IV inhibitory	Milk
LVYPFPGPI	Antihypertensive	Milk
NIPPLTQTPV	Antihypertensive	Milk
NVPGEIVE	Antihypertensive	Milk
PP	Antihypertensive	Cereal, milk, chicken, porcin
PPK	Antihypertensive	Cereal, chicken, porcin
PVVVPPFLQPE	Antimicrobial	Milk
PYP	Antihypertensive	Cereal, milk
PYPQ	Antioxidant	Milk
RELEE	Antihypertensive	Milk
RPK	Antihypertensive	Milk
SF	Antihypertensive	Cereal, milk, fish, chicken

Table A.S3 (continued): Bioactive peptide identities from Figure 2.5.

Peptide	Activity	Source
SLPQ	Antihypertensive	Milk
TQTPVVVPPFLQPE	Antioxidant	Milk
VAPFPEV	Antihypertensive	Milk
VEP	Antihypertensive	Milk
VL	Antihypertensive	Cereal
VLP	Antihypertensive	Cereal, milk, chicken, porcin
VLPVPQ	Antihypertensive	Milk
VP	Antihypertensive	Cereal, milk, fish, legume, chicken, porcin
VPQ	Antihypertensive	Milk
VPYPQ	Antioxidant	Milk
VVPP	Antihypertensive	Cereal, milk
VVVPPF	Antihypertensive	Milk
VY	Antihypertensive	Cereal, milk, fish, legume, porcin
VYP	Antihypertensive	Milk
VYPFPGPI	Antihypertensive	Milk
VYPFPGPIPN	Antioxidant	Milk
WQ	Antihypertensive	Milk
YP	Antihypertensive	Cereal, milk, chicken, porcin
YPFPGPI	Antihypertensive, opioid, Immunomodulatory, Immunomodulatory, opioid	Milk
YPVEPF	Antihypertensive, DPP-IV inhibitory, opioid	Milk
YV	Antihypertensive	Cereal

B | Appendix Chapter 3

B.1 Supplementary figures

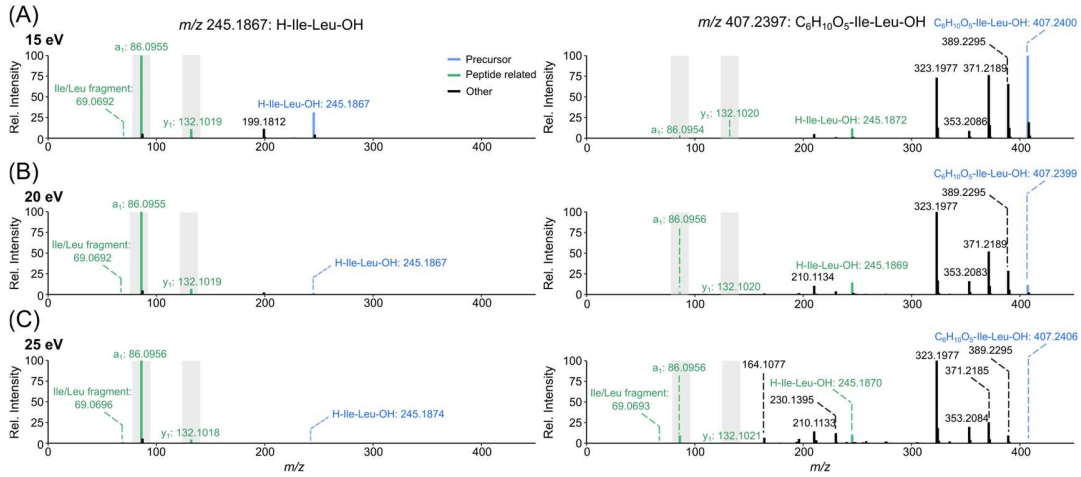


Figure B.S1: Variations in collision energy for classic peptide fragment ions generation. Collision energy dependent fragmentation behavior of the standard peptide H-Ile-Leu-OH (left) and the corresponding Amadori product $C_6H_{10}O_5$-Ile-Leu-OH (right). Changes in the relative ion intensity of (i) a-, b- and y-ions from peptide backbone fragmentation, (ii) ions derived from amino acid fragmentation, (iii) signals corresponding to the total peptide, and (iv) the precursor with the collision energy. Precursor ions were subjected to (A) 15 eV, (B) 20 eV, and (C) 25 eV.

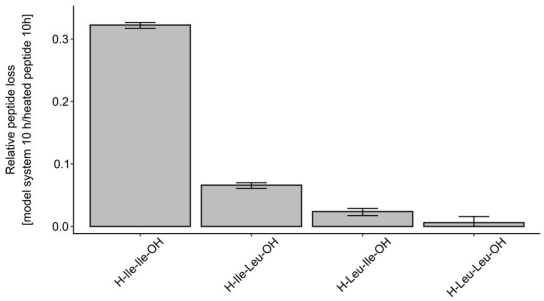


Figure B.S2: Relative concentration of starter peptide consumed by glycation reactions. Concentration of short-chain peptide reactants in glucose-model systems relative to heated peptide alone (100 °C, 10 hours). Quantification was performed by ^{1}H NMR spectroscopy. Error bars represent the standard deviation (n = 3).

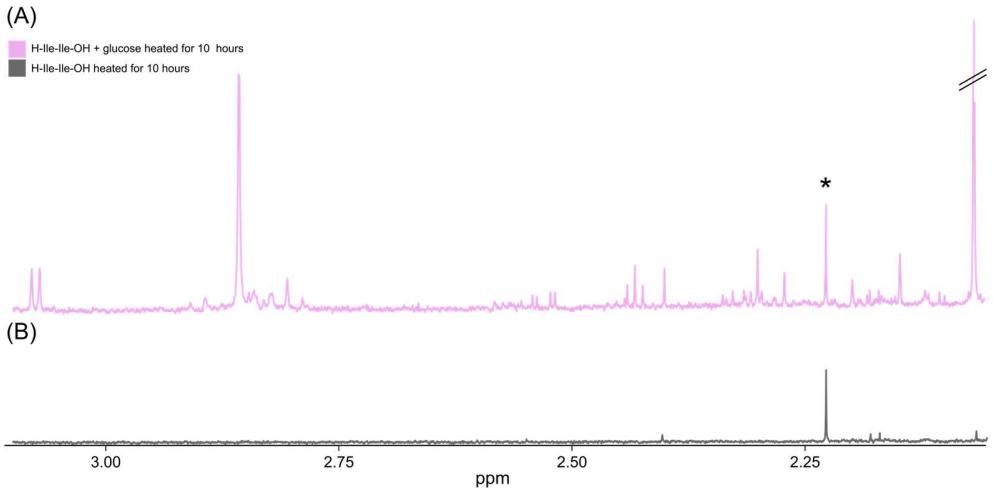


Figure B.S3: Exemplary region showing glycation reaction induced complexity in ^{1}H NMR spectra. Formation of reaction product related signals in H-Ile-Ile-OH glucose model systems (A) and absence in heated H-Ile-Ile-OH alone (B). Asterisk indicates a contaminant signal.

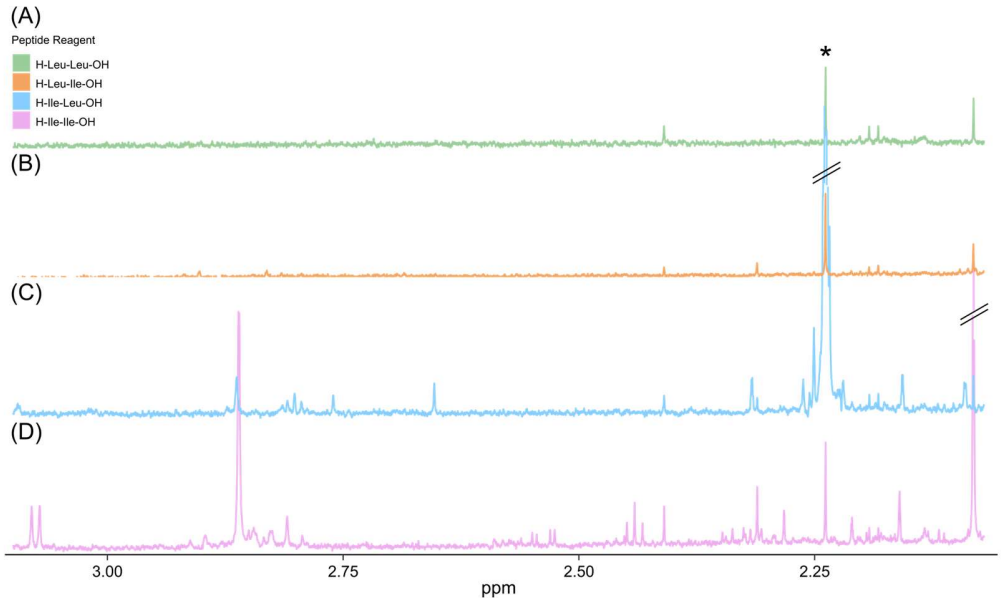


Figure B.S4: Glycation induced complexity of glucose short-chain peptide model systems prepared from (A) H-Leu-Leu-OH, (B) H-Leu-Ile-OH, (C) H-Ile-Leu-OH, and (D) H-Ile-Ile-OH measured by ^{1}H NMR spectroscopy. Asterisk indicates a contaminant signal. Specificity of this region for glycation events is shown in Figure B.S3.

B.2 Supplementary tables

Table B.S1: Glycation mass differences.

Analyte ID	*m/z*	Mass difference	Confidence level	Mass difference group ID
2429	259.1661	13.9802	2	16
2519	269.1866	24.0006	3	29
2631	279.2078	34.0219	2	42
2670	285.1818	39.9958	1	49
2734	294.1822	48.9962	1	63
2757	296.1981	51.0121	1	66
2785	299.1975	54.0116	2	70
2805	301.1769	55.9910	1	77
2844	303.1928	58.0069	1	80
2866	305.2079	60.0220	1	84
2884	307.2027	62.0167	2	89
2904	309.2184	64.0325	2	94
2907	310.2136	65.0276	1	98
2915	311.1976	66.0116	1	101
2932	312.2294	67.0435	2	102
2940	313.2133	68.0274	2	105
2949	314.2084	69.0224	2	108
2953	315.1922	70.0063	2	110
2782	299.1611	71.9857	2	116
2993	317.2082	72.0222	1	117
3038	322.2137	77.0278	1	127
3058	323.1971	78.0112	2	128
3070	324.1929	79.0069	1	132
3086	325.2133	80.0273	2	135
3096	326.2087	81.0227	1	138
3106	327.1922	82.0063	2	139
3160	329.2083	84.0223	2	142
3168	330.2035	85.0175	2	144
3218	334.2138	89.0278	2	153
3232	335.2188	90.0328	2	155
3240	336.2292	91.0433	1	158
3269	338.2085	93.0225	1	161
3285	339.1926	94.0066	2	162
3294	340.2245	95.0385	2	163
3304	341.2081	96.0221	2	165
3317	342.2038	97.0178	2	166
3330	343.2605	98.0745	1	170
3335	344.2192	99.0332	1	171
3110	327.1927	100.0173	1	173
3359	346.2135	101.0276	1	176

Table B.S1 (continued): Glycation mass differences.

Analyte ID	*m/z*	Mass difference	Confidence level	Mass difference group ID
3367	347.2187	102.0328	1	179
3392	350.2084	105.0225	1	182
3400	351.1921	106.0062	2	183
3419	352.2244	107.0384	1	187
3437	353.2085	108.0225	1	188
3474	357.2391	112.0531	3	198
3493	359.2185	114.0326	2	202
3513	361.2253	116.0394	1	206
3524	363.2138	118.0279	3	209
3532	364.2244	119.0384	2	210
3547	365.2079	120.0219	1	212
3561	366.2036	121.0177	1	214
3575	369.2028	124.0168	1	217
3594	371.2187	126.0328	2	219
3614	374.2083	129.0224	1	224
3663	382.2349	137.0489	1	235
3695	387.2135	142.0275	1	244
3699	388.2454	143.0595	1	246
3709	389.2296	144.0436	1	247
3737	393.2241	148.0382	3	250
3756	395.2248	150.0389	1	251
3783	399.2967	154.1108	1	259
3822	405.2245	160.0386	1	269
3842	407.2398	162.0539	1	271
3874	412.2449	167.0589	1	279
3885	413.2297	168.0437	1	283
3909	417.1327	171.9467	2	291
3922	418.2347	173.0487	1	293
3935	419.2398	174.0539	1	295
3949	421.2187	176.0327	3	300
3966	423.2352	178.0492	1	304
4000	428.2397	183.0538	1	316
4018	431.2396	186.0536	3	326
4040	435.2708	190.0849	3	332
4053	437.2285	192.0425	3	336
4057	437.2501	192.0642	3	337
4063	438.2606	193.0747	3	339
4088	442.2557	197.0697	1	346
4220	465.2452	220.0593	1	384
4226	465.2920	220.1061	3	385
4231	466.2917	221.1058	2	387
4242	467.2603	222.0744	3	389
4265	472.2659	227.0800	1	398

Table B.S1 (continued): Glycation mass differences.

Analyte ID	*m/z*	Mass difference	Confidence level	Mass difference group ID
4273	475.3125	230.1266	1	404
4290	479.2604	234.0744	1	409
4301	482.2855	237.0995	3	413
4319	485.3341	240.1481	1	416
4324	487.3494	242.1635	1	419
4332	489.3445	244.1585	3	421
4388	499.3490	254.1630	3	431
4414	509.2834	264.0975	3	442
4429	511.3457	266.1597	3	443
4447	513.3645	268.1785	1	447
4491	525.3655	280.1795	1	459
4517	531.3642	286.1782	1	468
4519	531.3752	286.1893	1	469
4531	533.3550	288.1691	1	472
4579	543.3758	298.1898	3	482
4632	551.3442	306.1583	3	488
4637	551.3809	306.1949	3	489
4649	553.3597	308.1738	3	493
4668	555.3745	310.1886	3	495
4682	557.3546	312.1687	3	496
4695	560.3650	315.1791	2	498
4696	561.3493	316.1634	1	500
4699	561.3866	316.2007	1	501
4746	567.3754	322.1895	3	502
4761	569.2919	324.1059	3	503
4781	571.3701	326.1841	3	504
4799	573.3649	328.1790	3	505
4805	573.3864	328.2004	1	506
4819	574.3814	329.1954	1	507
4855	579.3744	334.1885	3	510
4872	582.3496	337.1637	1	515
4886	584.3654	339.1794	3	516
4893	585.3862	340.2002	1	519
4903	586.3803	341.1943	3	520
4905	588.3752	343.1892	3	521
4916	589.3813	344.1953	1	523
4927	591.3750	346.1890	3	524
4945	593.3545	348.1686	3	526
4952	594.3858	349.1999	3	527
4963	595.3692	350.1832	3	528
4976	597.3851	352.1991	3	529
5045	605.3529	360.1669	3	534
5049	605.3901	360.2041	3	535

Table B.S1 (continued): Glycation mass differences.

Analyte ID	*m/z*	Mass difference	Confidence level	Mass difference group ID
5107	614.3769	369.1909	1	539
5130	615.3946	370.2086	1	540
5149	621.3865	376.2005	3	542
5181	627.3584	382.1724	1	547
5187	627.4062	382.2203	3	548
5211	631.3912	386.2052	3	550
5237	639.3950	394.2090	3	553
5279	651.4167	406.2308	3	560
5349	693.4285	448.2426	3	577
5427	759.4397	514.2538	1	598
5592	849.5551	604.3691	3	604

Table B.S2.1: Matched ion types for mass difference 162.0528 Da.

m/z	Intensity	MS/MS mass difference	Peptide backbone ion	Neutral loss from modification	Ion type	Mean neutral loss	Neutral loss group	Relative intensity	Theoretical m/z peptide backbone ion
86.0955	1183065	-0.0009	a_1	162.0548	CPFI	NA	NA	42.8969	86.0964
132.1017	46487	-0.0002	y_1	162.0541	CPFI	NA	NA	1.6856	132.1019
150.1282	6335	64.0318	a_1	98.0221	CFI	98.0219	135	0.2297	86.0964
164.1078	205503	78.0114	a_1	84.0425	CFI	84.0422	124	7.4514	86.0964
168.1392	6145	82.0428	a_1	80.0111	CFI	80.0113	122	0.2228	86.0964
176.1078	3409	90.0114	a_1	72.0425	CFI	72.0425	112	0.1236	86.0964
178.1234	32383	64.0320	b_1	98.0219	CFI	98.0219	135	1.1742	114.0913
182.1182	4359	96.0217	a_1	66.0321	CFI	66.0317	102	0.1581	86.0964
192.1029	2315	78.0116	b_1	84.0423	CFI	84.0422	124	0.0839	114.0913
194.1184	62879	108.0219	a_1	54.0319	CFI	54.0318	86	2.2799	86.0964
196.1342	235733	82.0428	b_1	80.0111	CFI	80.0113	122	8.5475	114.0913
210.1133	371567	96.0220	b_1	66.0319	CFI	66.0317	102	13.4727	114.0913
222.1137	3851	108.0224	b_1	54.0315	CFI	54.0318	86	0.1396	114.0913
227.1768	6251	0.0014	Total Peptide	162.0525	CPFI	NA	NA	0.2267	245.1860
228.1604	5091	0.0010	Total Peptide	162.0529	CPFI	NA	NA	0.1846	245.1860
230.1397	359641	144.0433	a_1	18.0106	CFI	18.0105	26	13.0403	86.0964
240.1243	41937	126.0330	b_1	36.0209	CFI	36.0210	65	1.5206	114.0913
245.1870	464335	0.0011	Total Peptide	162.0528	CPFI	NA	NA	16.8364	245.1860
248.1508	40363	162.0544	a_1	-0.0005	MPFI	NA	NA	1.4635	86.0964

Table B.S2.1 (continued): Matched ion types for mass difference 162.0528 Da.

m/z	**Intensity**	**MS/MS mass difference**	**Peptide backbone ion**	**Neutral loss from modification**	**Ion type**	**Mean neutral loss**	**Neutral loss group**	**Relative intensity**	**Theoretical m/z peptide backbone ion**
258.1348	54435	144.0435	b_1	18.0104	CFI	18.0105	26	1.9738	114.0913
276.1451	39135	162.0538	b_1	0.0001	MPFI	NA	NA	1.4190	114.0913
323.1982	2757927	78.0123	Total Peptide	84.0416	CFI	84.0422	124	100.0000	245.1860
335.1978	63301	90.0119	Total Peptide	72.0420	CFI	72.0425	112	2.2952	245.1860
353.2083	615857	108.0224	Total Peptide	54.0315	CFI	54.0318	86	22.3304	245.1860
371.2190	1258963	126.0330	Total Peptide	36.0209	CFI	36.0210	65	45.6489	245.1860
389.2298	397833	144.0438	Total Peptide	18.0101	CFI	18.0105	26	14.4251	245.1860
407.2404	50403	162.0544	Total Peptide	-0.0005	MPFI	NA	NA	1.8276	245.1860

Table BS2.2: Matched ion types for mass difference 268.1786 Da.

m/z	**Intensity**	**MS/MS mass difference**	**Peptide backbone ion**	**Neutral loss from modification**	**Ion type**	**Mean neutral loss**	**Neutral loss group**	**Relative intensity**	**Theoretical *m/z* peptide backbone ion**
354.2762	5331	268.1798	a_1	-0.0013	MPFI	NA	NA	12.0222	86.0964
86.0955	5293	-0.0009	a_1	268.1795	CPFI	NA	NA	11.9365	86.0964
223.1812	6153	137.0848	a_1	131.0938	CFI	131.0934	161	13.8759	86.0964
241.1922	7955	155.0958	a_1	113.0828	CFI	113.0828	146	17.9397	86.0964
251.1766	3817	137.0852	b_1	131.0933	CFI	131.0934	161	8.6079	114.0913
269.1871	6367	155.0957	b_1	113.0828	CFI	113.0828	146	14.3585	114.0913
382.2700	441	268.1787	b_1	-0.0001	MPFI	NA	NA	0.9945	114.0913
245.1870	44343	0.0010	Total Peptide	268.1775	CPFI	NA	NA	100.0000	245.1860
400.2825	929	155.0965	Total Peptide	113.0820	CFI	113.0828	146	2.0950	245.1860
513.3652	9829	268.1792	Total Peptide	-0.0007	MPFI	NA	NA	22.1658	245.1860
132.1011	409	-0.0008	y_1	268.1793	CPFI	NA	NA	0.9224	132.1019

Bibliography

(1) FOURNEAU, E. F. U. E. Über einige Derivate des Glykocolls. *Ber* **1901**, *34*, 2868-2877.

(2) Maillard, L. C. Action des acides amines sur les sucres; formation des melanoidines par voie methodique. *Comptes R. Acad. Sci.(Paris)* **1912**, *154*, 66-68.

(3) Sharon, N. IUPAC-IUB Joint Commission on Biochemical Nomenclature (JCBN). Nomenclature of glycoproteins, glycopeptides and peptidoglycans. Recommendations 1985. *Eur J Biochem* **1986**, *159*, 1-6.

(4) Lis, H.; Sharon, N. Protein glycosylation. Structural and functional aspects. *Eur J Biochem* **1993**, *218*, 1-27.

(5) Hodge, J. E. Dehydrated Foods, Chemistry of Browning Reactions in Model Systems. *Journal of Agricultural and Food Chemistry* **1953**, *1*, 928-943.

(6) Tressl, R.; Nittka, C.; Kersten, E.; Rewicki, D. Formation of Isoleucine-Specific Maillard Products from [1-13C]-D-Glucose and [1-13C]-D-Fructose. *Journal of Agricultural and Food Chemistry* **1995**, *43*, 1163-1169.

(7) Davidek, T.; Clety, N.; Aubin, S.; Blank, I. Degradation of the Amadori compound N-(1-deoxy-D-fructos-1-yl)glycine in aqueous model systems. *J Agric Food Chem* **2002**, *50*, 5472-5479.

(8) Hellwig, M.; Henle, T. Baking, Ageing, Diabetes: A Short History of the Maillard Reaction. *Angewandte Chemie International Edition* **2014**, *53*, 10316-10329.

(9) Hellwig, M. Proteolytische Freisetzung und epithelialer Transport von Maillard-Reaktionsprodukten und Crosslink-Aminosäuren. **2011**.

(10) Hemmler, D.; Schmitt-Kopplin, P. In *Reference Module in Food Science*; Elsevier, 2020.

(11) Hodge, J. E. The Amadori rearrangement. *Adv Carbohydr Chem* **1955**, *10*, 169-205.

(12) Amadori, M. The condensation product of glucose and p-anisidine. *Atti R Accad Naz Lincei* **1929**, *9*, 226-230.

(13) Heyns, K.; Eichstedt, R.; Meinecke, K.-H. Die Umsetzung von Fructose und Sorbose mit Ammoniak und Aminen. *Chemische Berichte* **1955**, *88*, 1551-1555.

(14) Horvat, Ŝ.; Varga-Defterdarović, L.; Horvat, J. Synthesis of novel imidazolidinones from hexose–peptide adducts: model studies of the Maillard reaction with possible significance in protein glycation. *Chemical Communications* **1998**, 1663-1664.

(15) Roščić, M.; Versluis, C.; Kleinnijenhuis, A. J.; Horvat, Š.; Heck, A. J. R. The early glycation products of the Maillard reaction: mass spectrometric characterization of novel imidazolidinones derived from an opioid pentapeptide and glucose. *Rapid Communications in Mass Spectrometry* **2001**, *15*, 1022-1029.

(16) Roscić, M.; Horvat, S. Transformations of bioactive peptides in the presence of sugars--characterization and stability studies of the adducts generated via the Maillard reaction. *Bioorg Med Chem* **2006**, *14*, 4933-4943.

(17) Chu, F. L.; Yaylayan, V. A. Formation of the peptide-specific imidazolidin-4-one moiety in alanine containing Maillard reaction mixtures. *Food Chemistry* **2010**, *123*, 1185-1189.

(18) Laroque, D.; Inisan, C.; Berger, C.; Vouland, É.; Dufossé, L.; Guérard, F. Kinetic study on the Maillard reaction. Consideration of sugar reactivity. *Food Chemistry* **2008**, *111*, 1032-1042.

(19) Mergenthaler, P.; Lindauer, U.; Dienel, G. A.; Meisel, A. Sugar for the brain: the role of glucose in physiological and pathological brain function. *Trends Neurosci* **2013**, *36*, 587-597.

(20) Bunn, H. F.; Higgins, P. J. Reaction of monosaccharides with proteins: possible evolutionary significance. *Science* **1981**, *213*, 222-224.

(21) Weygand, F. Über N-Glykoside, II. Mitteil.: Amadori-Umlagerungen. *Berichte der deutschen chemischen Gesellschaft (A and B Series)* **1940**, *73*, 1259-1278.

(22) Kato, Y.; Matsuda, T.; Kato, N.; Watanabe, K.; Nakamura, R. Browning and insolubilization of ovalbumin by the Maillard reaction with some aldohexoses. *Journal of Agricultural and Food Chemistry* **1986**, *34*, 351-355.

(23) Kato, Y.; Matsuda, T.; Kato, N.; Nakamura, R. Maillard reaction of ovalbumin with glucose and lactose. Browning and protein polymerization induced by amino-carbonyl reaction of ovalbumin with glucose and lactose. *Journal of Agricultural and Food Chemistry* **1988**, *36*, 806-809.

(24) Speck, J. C., Jr. The Lobry de Bruyn-Alberda van Ekenstein transformation. *Adv Carbohydr Chem* **1958**, *13*, 63-103.

(25) Yaylayan, V. A.; Huyghues-Despointes, A. Chemistry of Amadori rearrangement products: analysis, synthesis, kinetics, reactions, and spectroscopic properties. *Crit Rev Food Sci Nutr* **1994**, *34*, 321-369.

(26) Ledl, F.; Schleicher, E. New Aspects of the Maillard Reaction in Foods and in the Human Body. *Angewandte Chemie International Edition in English* **1990**, *29*, 565-594.

(27) Henning, C.; Smuda, M.; Girndt, M.; Ulrich, C.; Glomb, M. A. Molecular basis of maillard amide-advanced glycation end product (AGE) formation in vivo. *J Biol Chem* **2011**, *286*, 44350-44356.

(28) Henning, C.; Liehr, K.; Girndt, M.; Ulrich, C.; Glomb, M. A. Extending the spectrum of α-dicarbonyl compounds in vivo. *J Biol Chem* **2014**, *289*, 28676-28688.

(29) Glomb, M. A.; Tschirnich, R. Detection of alpha-dicarbonyl compounds in Maillard reaction systems and in vivo. *J Agric Food Chem* **2001**, *49*, 5543-5550.

(30) Glomb, M. A.; Pfahler, C. Synthesis of 1-deoxy-D-erythro-hexo-2,3-diulose, a major hexose Maillard intermediate. *Carbohydr Res* **2000**, *329*, 515-523.

(31) Lund, M. N.; Ray, C. A. Control of Maillard Reactions in Foods: Strategies and Chemical Mechanisms. *Journal of Agricultural and Food Chemistry* **2017**, *65*, 4537-4552.

(32) Gobert, J.; Glomb, M. A. Degradation of glucose: reinvestigation of reactive alpha-Dicarbonyl compounds. *J Agric Food Chem* **2009**, *57*, 8591-8597.

(33) Akabori, S. Oxidation of Amino-acids with Sugars. *Proceedings of the Imperial Academy* **1927**, *3*, 672-674.

(34) Schonberg, A.; Moubacher, R. The Strecker degradation of α-amino acids. *Chemical Reviews* **1952**, *50*, 261-277.

(35) Strecker, A. On a peculiar oxidation by alloxan. *Justus Liebigs Ann Chem.* **1862**, *123*, 363-367.

(36) Yaylayan, V. A. Recent Advances in the Chemistry of Strecker Degradation and Amadori Rearrangement: Implications to Aroma and Color Formation. *Food Science and Technology Research* **2003**, *9*, 1-6.

(37) Rizzi, G. P. The Strecker Degradation of Amino Acids: Newer Avenues for Flavor Formation. *Food Reviews International* **2008**, *24*, 416-435.

(38) Lu, C. Y.; Hao, Z.; Payne, R.; Ho, C. T. Effects of water content on volatile generation and peptide degradation in the maillard reaction of glycine, diglycine, and triglycine. *J Agric Food Chem* **2005**, *53*, 6443-6447.

(39) Lan, X.; Liu, P.; Xia, S.; Jia, C.; Mukunzi, D.; Zhang, X.; Xia, W.; Tian, H.; Xiao, Z. Temperature effect on the non-volatile compounds of Maillard reaction products derived from xylose–soybean peptide system: Further insights into thermal degradation and cross-linking. *Food Chemistry* **2010**, *120*, 967-972.

(40) Patel, K.; Borchardt, R. T. Chemical pathways of peptide degradation. II. Kinetics of deamidation of an asparaginyl residue in a model hexapeptide. *Pharm Res* **1990**, *7*, 703-711.

(41) Mirzaei, M.; Mirdamadi, S.; Safavi, M.; Soleymanzadeh, N. The stability of antioxidant and ACE-inhibitory peptides as influenced by peptide sequences. *LWT* **2020**, *130*, 109710.

(42) Wang, B.; Xie, N.; Li, B. Influence of peptide characteristics on their stability, intestinal transport, and in vitro bioavailability: A review. *J Food Biochem* **2019**, *43*, e12571.

(43) Yaylayan, V. A.; Kaminsky, E. Isolation and structural analysis of maillard polymers: caramel and melanoidin formation in glycine/glucose model system. *Food Chemistry* **1998**, *63*, 25-31.

(44) Tressl, R.; Wondrak, G. T.; Krüger, R. P.; Rewicki, D. New Melanoidin-like Maillard Polymers from 2-Deoxypentoses. *J Agric Food Chem* **1998**, *46*, 104-110.

(45) Hellwig, M.; Geissler, S.; Peto, A.; Knütter, I.; Brandsch, M.; Henle, T. Transport of free and peptide-bound pyrraline at intestinal and renal epithelial cells. *J Agric Food Chem* **2009**, *57*, 6474-6480.

(46) Xue, J.; Ray, R.; Singer, D.; Böhme, D.; Burz, D. S.; Rai, V.; Hoffmann, R.; Shekhtman, A. The receptor for advanced glycation end products (RAGE) specifically recognizes methylglyoxal-derived AGEs. *Biochemistry* **2014**, *53*, 3327-3335.

(47) Monnier, V. M.; Sell, D. R.; Dai, Z.; Nemet, I.; Collard, F.; Zhang, J. The role of the amadori product in the complications of diabetes. *Ann N Y Acad Sci* **2008**, *1126*, 81-88.

(48) Kasper, M.; Schieberle, P. Labeling studies on the formation pathway of Nepsilon-carboxymethyllysine in maillard-type reactions. *Ann N Y Acad Sci* **2005**, *1043*, 59-62.

(49) Elgawish, A.; Glomb, M.; Friedlander, M.; Monnier, V. M. Involvement of hydrogen peroxide in collagen cross-linking by high glucose in vitro and in vivo. *Journal of Biological Chemistry* **1996**, *271*, 12964-12971.

(50) Namiki, M.; Hayashi, T.; Ohta, Y. Novel free radicals formed by the amino-carbonyl reactions of sugars with amino acids, amines, and proteins. *Adv Exp Med Biol* **1977**, *86b*, 471-501.

(51) Hayashi, T.; Namki, M. Formation of two-carbon sugar fragment at an early stage of the browning reaction of sugar with amine. *Agricultural and Biological Chemistry* **1980**, *44*, 2575-2580.

(52) Glomb, M. A.; Monnier, V. M. Mechanism of protein modification by glyoxal and glycolaldehyde, reactive intermediates of the Maillard reaction. *J Biol Chem* **1995**, *270*, 10017-10026.

(53) Cho, S. J.; Roman, G.; Yeboah, F.; Konishi, Y. The road to advanced glycation end products: a mechanistic perspective. *Curr Med Chem* **2007**, *14*, 1653-1671.

(54) Koenig, R. J.; Peterson, C. M.; Jones, R. L.; Saudek, C.; Lehrman, M.; Cerami, A. Correlation of glucose regulation and hemoglobin AIc in diabetes mellitus. *N Engl J Med* **1976**, *295*, 417-420.

(55) Curtiss, L. K.; Witztum, J. L. Plasma apolipoproteins AI, AII, B, CI, and E are glucosylated in hyperglycemic diabetic subjects. *Diabetes* **1985**, *34*, 452-461.

(56) Monnier, V. M.; Cerami, A. Nonenzymatic browning in vivo: possible process for aging of long-lived proteins. *Science* **1981**, *211*, 491-493.

(57) Sell, D. R.; Monnier, V. M. Structure elucidation of a senescence cross-link from human extracellular matrix. Implication of pentoses in the aging process. *J Biol Chem* **1989**, *264*, 21597-21602.

(58) Wells-Knecht, K. J.; Zyzak, D. V.; Litchfield, J. E.; Thorpe, S. R.; Baynes, J. W. Mechanism of autoxidative glycosylation: identification of glyoxal and arabinose as intermediates in the autoxidative modification of proteins by glucose. *Biochemistry* **1995**, *34*, 3702-3709.

(59) Wells-Knecht, K. J.; Brinkmann, E.; Baynes, J. W. Characterization of an imidazolium salt formed from glyoxal and N. alpha.-hippuryllysine: A model for maillard reaction crosslinks in proteins. *The Journal of Organic Chemistry* **1995**, *60*, 6246-6247.

(60) Brinkmann, E.; Wells-Knecht, K. J.; Thorpe, S. R.; Baynes, J. W. Characterization of an imidazolium compound formed by reaction of methylglyoxal and N α-hippuryllysine. *Journal of the Chemical Society, Perkin Transactions 1* **1995**, 2817-2818.

(61) Skovsted, I. C.; Christensen, M.; Breinholt, J.; Mortensen, S. B. Characterisation of a novel AGE-compound derived from lysine and 3-deoxyglucosone. *Cell Mol Biol (Noisy-le-grand)* **1998**, *44*, 1159-1163.

(62) Lederer, M. O.; Klaiber, R. G. Cross-linking of proteins by Maillard processes: characterization and detection of lysine-arginine cross-links derived from glyoxal and methylglyoxal. *Bioorg Med Chem* **1999**, *7*, 2499-2507.

(63) Biemel, K. M.; Reihl, O.; Conrad, J.; Lederer, M. O. Formation pathways for lysine-arginine cross-links derived from hexoses and pentoses by Maillard processes: unraveling the structure of a pentosidine precursor. *J Biol Chem* **2001**, *276*, 23405-23412.

(64) Loske, C.; Gerdemann, A.; Schepl, W.; Wycislo, M.; Schinzel, R.; Palm, D.; Riederer, P.; Münch, G. Transition metal-mediated glycoxidation accelerates cross-linking of beta-amyloid peptide. *Eur J Biochem* **2000**, *267*, 4171-4178.

(65) Piersimoni, L.; Kastritis, P. L.; Arlt, C.; Sinz, A. Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein-Protein Interactions─A Method for All Seasons. *Chem Rev* **2022**, *122*, 7500-7531.

(66) Yılmaz, Ş.; Busch, F.; Nagaraj, N.; Cox, J. Accurate and Automated High-Coverage Identification of Chemically Cross-Linked Peptides with MaxLynx. *Anal Chem* **2022**, *94*, 1608-1617.

(67) Brownlee, M.; Vlassara, H.; Kooney, A.; Ulrich, P.; Cerami, A. Aminoguanidine prevents diabetes-induced arterial wall protein cross-linking. *Science* **1986**, *232*, 1629-1632.

(68) Gulewitsch, W.; Amiradžibi, S. Ueber das Carnosin, eine neue organische Base des Fleischextractes. *Berichte der deutschen chemischen Gesellschaft* **1900**, *33*, 1902-1903.

(69) Hipkiss, A. R.; Michaelis, J.; Syrris, P. Non-enzymatic glycosylation of the dipeptide L-carnosine, a potential anti-protein-cross-linking agent. *FEBS Lett* **1995**, *371*, 81-85.

(70) Hipkiss, A. R. Carnosine, a protective, anti-ageing peptide? *Int J Biochem Cell Biol* **1998**, *30*, 863-868.

(71) Freund, M. A.; Chen, B.; Decker, E. A. The Inhibition of Advanced Glycation End Products by Carnosine and Other Natural Dipeptides to Reduce Diabetic and Age-Related Complications. *Compr Rev Food Sci Food Saf* **2018**, *17*, 1367-1378.

(72) Alhamdani, M. S.; Al-Kassir, A. H.; Abbas, F. K.; Jaleel, N. A.; Al-Taee, M. F. Antiglycation and antioxidant effect of carnosine against glucose degradation products in peritoneal mesothelial cells. *Nephron Clin Pract* **2007**, *107*, c26-34.

(73) Pfister, F.; Riedl, E.; Wang, Q.; vom Hagen, F.; Deinzer, M.; Harmsen, M. C.; Molema, G.; Yard, B.; Feng, Y.; Hammes, H. P. Oral carnosine supplementation prevents vascular damage in experimental diabetic retinopathy. *Cell Physiol Biochem* **2011**, *28*, 125-136.

(74) Szwergold, B. S. Carnosine and anserine act as effective transglycating agents in decomposition of aldose-derived Schiff bases. *Biochem Biophys Res Commun* **2005**, *336*, 36-41.

(75) Ajandouz, E. H.; Puigserver, A. Nonenzymatic browning reaction of essential amino acids: effect of pH on caramelization and Maillard reaction kinetics. *J Agric Food Chem* **1999**, *47*, 1786-1793.

(76) Hemmler, D.; Roullier-Gall, C.; Marshall, J. W.; Rychlik, M.; Taylor, A. J.; Schmitt-Kopplin, P. Insights into the Chemistry of Non-Enzymatic Browning Reactions in Different Ribose-Amino Acid Model Systems. *Scientific Reports* **2018**, *8*, 16879.

(77) Ashoor, S.; Zent, J. Maillard browning of common amino acids and sugars. *Journal of Food Science* **1984**, *49*, 1206-1207.

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

(79) Xing, H.; Yaylayan, V. Investigation of thermo-chemical properties of mechanochemically generated glucose–histidine Maillard reaction mixtures. *European Food Research and Technology* **2021**, *247*, 111-120.

(80) Heyns, K.; Noack, H. Die Umsetzung von D-Fructose mit L-Lysin und L-Arginin und deren Beziehung zu nichtenzymatischen Bräunungsreaktionen. *Chemische Berichte* **1962**, *95*, 720-727.

(81) Zhao, P. L.; Nachbar, R. B.; Bolognese, J. A.; Chapman, K. Two new criteria for choosing sample size in combinatorial chemistry. *J Med Chem* **1996**, *39*, 350-352.

(82) Van Lancker, F.; Adams, A.; De Kimpe, N. Formation of pyrazines in Maillard model systems of lysine-containing dipeptides. *J Agric Food Chem* **2010**, *58*, 2470-2478.

(83) Van Lancker, F.; Adams, A.; De Kimpe, N. Impact of the N-terminal amino acid on the formation of pyrazines from peptides in Maillard model systems. *J Agric Food Chem* **2012**, *60*, 4697-4708.

(84) de Kok, P. M. T.; Rosing, E. A. E. In *Thermally Generated Flavors*; American Chemical Society, 1993, pp 158-179.

(85) Mennella, C.; Visciano, M.; Napolitano, A.; Del Castillo, M. D.; Fogliano, V. Glycation of lysine-containing dipeptides. *J Pept Sci* **2006**, *12*, 291-296.

(86) Liang, Z.; Li, L.; Qi, H.; Wan, L.; Cai, P.; xu, Z.; Li, B. Formation of Peptide Bound Pyrraline in the Maillard Model Systems with Different Lys-Containing Dipeptides and Tripeptides. *Molecules* **2016**, *21*, 463.

(87) Van Chuyen, N.; Kurata, T.; Fujimaki, M. Studies on the reaction of dipeptides with glyoxal. *Agricultural and Biological Chemistry* **1973**, *37*, 327-334.

(88) Motai, H. Color tone of various melanoidins produced from model systems. *Agricultural and Biological Chemistry* **1973**, *37*, 1679-1685.

(89) Gao, Y.; Wang, Y. Site-selective modifications of arginine residues in human hemoglobin induced by methylglyoxal. *Biochemistry* **2006**, *45*, 15654-15660.

(90) Hudson, D. M.; Archer, M.; King, K. B.; Eyre, D. R. Glycation of type I collagen selectively targets the same helical domain lysine sites as lysyl oxidase-mediated cross-linking. *J Biol Chem* **2018**, *293*, 15620-15627.

(91) Ito, S.; Nakahari, T.; Yamamoto, D. The structural feature surrounding glycated lysine residues in human hemoglobin. *Biomed Res* **2011**, *32*, 217-223.

(92) Ahmed, N.; Dobler, D.; Dean, M.; Thornalley, P. J. Peptide mapping identifies hotspot site of modification in human serum albumin by methylglyoxal involved in ligand binding and esterase activity. *J Biol Chem* **2005**, *280*, 5724-5732.

(93) Schmidt, R.; Böhme, D.; Singer, D.; Frolov, A. Specific tandem mass spectrometric detection of AGE-modified arginine residues in peptides. *J Mass Spectrom* **2015**, *50*, 613-624.

(94) Gangadhariah, M. H.; Wang, B.; Linetsky, M.; Henning, C.; Spanneberg, R.; Glomb, M. A.; Nagaraj, R. H. Hydroimidazolone modification of human alphaA-crystallin: Effect on the chaperone function and protein refolding ability. *Biochim Biophys Acta* **2010**, *1802*, 432-441.

(95) Whitaker, J. R.; Feeney, R. E. Chemical and physical modification of proteins by the hydroxide ion. *Crit Rev Food Sci Nutr* **1983**, *19*, 173-212.

(96) Shapiro, R.; McManus, M. J.; Zalut, C.; Bunn, H. F. Sites of nonenzymatic glycosylation of human hemoglobin A. *J Biol Chem* **1980**, *255*, 3120-3127.

(97) Shilton, B. H.; Walton, D. J. Sites of glycation of human and horse liver alcohol dehydrogenase in vivo. *J Biol Chem* **1991**, *266*, 5587-5592.

(98) Watkins, N. G.; Thorpe, S. R.; Baynes, J. W. Glycation of amino groups in protein. Studies on the specificity of modification of RNase by glucose. *J Biol Chem* **1985**, *260*, 10629-10636.

(99) Venkatraman, J.; Aggarwal, K.; Balaram, P. Helical peptide models for protein glycation: proximity effects in catalysis of the Amadori rearrangement. *Chem Biol* **2001**, *8*, 611-625.

(100) Horvat, S.; Jakas, A. Peptide and amino acid glycation: new insights into the Maillard reaction. *J Pept Sci* **2004**, *10*, 119-137.

(101) Snedeker, J. G.; Gautieri, A. The role of collagen crosslinks in ageing and diabetes - the good, the bad, and the ugly. *Muscles Ligaments Tendons J* **2014**, *4*, 303-308.

(102) Salahuddin, P.; Rabbani, G.; Khan, R. H. The role of advanced glycation end products in various types of neurodegenerative disease: a therapeutic approach. *Cell Mol Biol Lett* **2014**, *19*, 407-437.

(103) Kunkel, H. G.; Wallenius, G. New hemoglobin in normal adult blood. *Science* **1955**, *122*, 288.

(104) Huisman, T. H.; Dozy, A. M. Studies on the heterogeneity of hemoglobin. V. Binding of hemoglobin with oxidized glutathione. *J Lab Clin Med* **1962**, *60*, 302-319.

(105) Rahbar, S.; Blumenfeld, O.; Ranney, H. M. Studies of an unusual hemoglobin in patients with diabetes mellitus. *Biochem Biophys Res Commun* **1969**, *36*, 838-843.

(106) Trivelli, L. A.; Ranney, H. M.; Lai, H. T. Hemoglobin components in patients with diabetes mellitus. *N Engl J Med* **1971**, *284*, 353-357.

(107) Cox, M. E.; Edelman, D. Tests for screening and diagnosis of type 2 diabetes. *Clinical diabetes* **2009**, *27*, 132-138.

(108) Ahmed, M. U.; Thorpe, S. R.; Baynes, J. W. Identification of N epsilon-carboxymethyllysine as a degradation product of fructoselysine in glycated protein. *J Biol Chem* **1986**, *261*, 4889-4894.

(109) Gaens, K. H.; Goossens, G. H.; Niessen, P. M.; van Greevenbroek, M. M.; van der Kallen, C. J.; Niessen, H. W.; Rensen, S. S.; Buurman, W. A.; Greve, J. W.; Blaak, E. E.; van Zandvoort, M. A.; Bierhaus, A.; Stehouwer, C. D.; Schalkwijk, C. G. Nε-(carboxymethyl)lysine-receptor for advanced

glycation end product axis is a key modulator of obesity-induced dysregulation of adipokine expression and insulin resistance. *Arterioscler Thromb Vasc Biol* **2014**, *34*, 1199-1208.

(110) Heremans, I. P.; Caligiore, F.; Gerin, I.; Bury, M.; Lutz, M.; Graff, J.; Stroobant, V.; Vertommen, D.; Teleman, A. A.; Van Schaftingen, E.; Bommer, G. T. Parkinson's disease protein PARK7 prevents metabolite and protein damage caused by a glycolytic metabolite. *Proc Natl Acad Sci U S A* **2022**, *119*.

(111) Nielsen, S. D.; Beverly, R. L.; Qu, Y.; Dallas, D. C. Milk bioactive peptide database: A comprehensive database of milk protein-derived bioactive peptides and novel visualization. *Food Chemistry* **2017**, *232*, 673-682.

(112) Li, Q.; Zhang, C.; Chen, H.; Xue, J.; Guo, X.; Liang, M.; Chen, M. BioPepDB: an integrated data platform for food-derived bioactive peptides. *International Journal of Food Sciences and Nutrition* **2018**, *69*, 963-968.

(113) Pihlanto-Leppälä, A.; Antila, P.; Mäntsälä, P.; Hellman, J. Opioid peptides produced by in-vitro proteolysis of bovine caseins. *International Dairy Journal* **1994**, *4*, 291-301.

(114) Jiang, Z.; Rai, D. K.; O'Connor, P. M.; Brodkorb, A. Heat-induced Maillard reaction of the tripeptide IPP and ribose: Structural characterization and implication on bioactivity. *Food Research International* **2013**, *50*, 266-274.

(115) Jiang, Z.; Wang, L.; Wu, W.; Wang, Y. Biological activities and physicochemical properties of Maillard reaction products in sugar-bovine casein peptide model systems. *Food Chem* **2013**, *141*, 3837-3845.

(116) Walsh, J. H.; Dockray, G. J.; Mitty, R. D.; LWW, 1994, p 487.

(117) Mooney, M. H.; Abdel-Wahab, Y. H.; Morgan, L. M.; O'Harte, F. P.; Flatt, P. R. Detection of glycated gastric inhibitory polypeptide within the intestines of diabetic obese (ob/ob) mice. *Endocrine* **2001**, *16*, 167-171.

(118) O'Harte, F. P.; Abdel-Wahab, Y. H.; Conlon, J. M.; Flatt, P. R. Amino terminal glycation of gastric inhibitory polypeptide enhances its insulinotropic action on clonal pancreatic B-cells. *Biochim Biophys Acta* **1998**, *1425*, 319-327.

(119) O'Harte, F. P.; Mooney, M. H.; Kelly, C. M.; Flatt, P. R. Improved glycaemic control in obese diabetic ob/ob mice using N-terminally modified gastric inhibitory polypeptide. *J Endocrinol* **2000**, *165*, 639-648.

(120) Gault, V. A.; Flatt, P. R.; O'Harte, F. P. Glucose-dependent insulinotropic polypeptide analogues and their therapeutic potential for the treatment of obesity-diabetes. *Biochem Biophys Res Commun* **2003**, *308*, 207-213.

(121) Deacon, C. F.; Knudsen, L. B.; Madsen, K.; Wiberg, F. C.; Jacobsen, O.; Holst, J. J. Dipeptidyl peptidase IV resistant analogues of glucagon-like peptide-1 which have extended metabolic stability and improved biological activity. *Diabetologia* **1998**, *41*, 271-278.

(122) O'Harte, F. P.; Abdel-Wahab, Y. H.; Conlon, J. M.; Flatt, P. R. Glycation of glucagon-like peptide-1(7-36)amide: characterization and impaired action on rat insulin secreting cells. *Diabetologia* **1998**, *41*, 1187-1193.

(123) Mottram, D. S.; ACS Publications, 1994.

(124) Temussi, P. A. The good taste of peptides. *J Pept Sci* **2012**, *18*, 73-82.

(125) Liu, J.; Liu, M.; He, C.; Song, H.; Chen, F. Effect of thermal treatment on the flavor generation from Maillard reaction of xylose and chicken peptide. *LWT-Food Science and Technology* **2015**, *64*, 316-325.

(126) Chen, F.; Lin, L.; Zhao, M.; Zhu, Q. Modification of Cucumaria frondosa hydrolysate through maillard reaction for sea cucumber peptide based-beverage. *LWT* **2021**, *136*, 110329.

(127) Abdelhedi, O.; Mora, L.; Jemil, I.; Jridi, M.; Toldrá, F.; Nasri, M.; Nasri, R. Effect of ultrasound pretreatment and Maillard reaction on structure and antioxidant properties of ultrafiltrated smooth-hound viscera proteins-sucrose conjugates. *Food Chem* **2017**, *230*, 507-515.

(128) Fu, Y.; Zhang, Y.; Soladoye, O. P.; Aluko, R. E. Maillard reaction products derived from food protein-derived peptides: insights into flavor and bioactivity. *Crit Rev Food Sci Nutr* **2020**, *60*, 3429-3442.

(129) Treibmann, S.; Hellwig, A.; Hellwig, M.; Henle, T. Lysine-Derived Protein-Bound Heyns Compounds in Bakery Products. *J Agric Food Chem* **2017**, *65*, 10562-10570.

(130) Henle, T. AGEs in foods: do they play a role in uremia? *Kidney Int Suppl* **2003**, S145-147.

(131) Lee, K.; Erbersdobler, H. In *Maillard reactions in chemistry, food and health*; Elsevier, 2005, pp 358-363.

(132) Förster, A.; Kühne, Y.; Henle, T. Studies on absorption and elimination of dietary maillard reaction products. *Ann N Y Acad Sci* **2005**, *1043*, 474-481.

(133) Yuan, H.; Sun, L.; Chen, M.; Wang, J. The Comparison of the Contents of Sugar, Amadori, and Heyns Compounds in Fresh and Black Garlic. *J Food Sci* **2016**, *81*, C1662-1668.

(134) Yadav, S. P. The wholeness in suffix -omics, -omes, and the word om. *J Biomol Tech* **2007**, *18*, 277.

(135) Patti, G. J.; Yanes, O.; Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol* **2012**, *13*, 263-269.

(136) Krassowski, M.; Das, V.; Sahu, S. K.; Misra, B. B. State of the Field in Multi-Omics Research: From Computational Needs to Data Mining and Sharing. *Front Genet* **2020**, *11*, 610798.

(137) McKusick, V. A.; Ruddle, F. H.; Academic Press, 1987, pp 1-2.

(138) Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; Funke, R.; Gage, D.; Harris, K.; Heaford, A.; Howland, J.; Kann, L.; Lehoczky, J.; LeVine, R.; McEwan, P.; McKernan, K., et al. Initial sequencing and analysis of the human genome. *Nature* **2001**, *409*, 860-921.

(139) Manzoni, C.; Kia, D. A.; Vandrovcova, J.; Hardy, J.; Wood, N. W.; Lewis, P. A.; Ferrari, R. Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. *Brief Bioinform* **2018**, *19*, 286-302.

(140) Cox, J.; Mann, M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu Rev Biochem* **2011**, *80*, 273-299.

(141) Guijas, C.; Montenegro-Burke, J. R.; Warth, B.; Spilker, M. E.; Siuzdak, G. Metabolomics activity screening for identifying metabolites that modulate phenotype. *Nat Biotechnol* **2018**, *36*, 316-320.

(142) Lin, H.; Su, X.; He, B. Protein lysine acylation and cysteine succination by intermediates of energy metabolism. *ACS Chem Biol* **2012**, *7*, 947-960.

(143) Wasinger, V. C.; Cordwell, S. J.; Cerpa-Poljak, A.; Yan, J. X.; Gooley, A. A.; Wilkins, M. R.; Duncan, M. W.; Harris, R.; Williams, K. L.; Humphery-Smith, I. Progress with gene-product mapping of the Mollicutes: Mycoplasma genitalium. *Electrophoresis* **1995**, *16*, 1090-1094.

(144) Wilkins, M. R.; Pasquali, C.; Appel, R. D.; Ou, K.; Golaz, O.; Sanchez, J. C.; Yan, J. X.; Gooley, A. A.; Hughes, G.; Humphery-Smith, I.; Williams, K. L.; Hochstrasser, D. F. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)* **1996**, *14*, 61-65.

(145) Sinitcyn, P.; Rudolph, J. D.; Cox, J. Computational Methods for Understanding Mass Spectrometry–Based Shotgun Proteomics Data. *Annual Review of Biomedical Data Science* **2018**, *1*, 207-234.

(146) Hasin, Y.; Seldin, M.; Lusis, A. Multi-omics approaches to disease. *Genome Biol* **2017**, *18*, 83.

(147) Chait, B. T. Chemistry. Mass spectrometry: bottom-up or top-down? *Science* **2006**, *314*, 65-66.

(148) Kelleher, N. L. Top-down proteomics. *Anal Chem* **2004**, *76*, 197a-203a.

(149) Gregorich, Z. R.; Ge, Y. Top-down proteomics in health and disease: challenges and opportunities. *Proteomics* **2014**, *14*, 1195-1210.

(150) Melby, J. A.; Roberts, D. S.; Larson, E. J.; Brown, K. A.; Bayne, E. F.; Jin, S.; Ge, Y. Novel Strategies to Address the Challenges in Top-Down Proteomics. *J Am Soc Mass Spectrom* **2021**, *32*, 1278-1294.

(151) Zhang, Y.; Fonslow, B. R.; Shan, B.; Baek, M. C.; Yates, J. R., 3rd. Protein analysis by shotgun/bottom-up proteomics. *Chem Rev* **2013**, *113*, 2343-2394.

(152) Schulte, I.; Tammen, H.; Selle, H.; Schulz-Knappe, P. Peptides in body fluids and tissues as markers of disease. *Expert Rev Mol Diagn* **2005**, *5*, 145-157.

(153) Svensson, M.; Sköld, K.; Nilsson, A.; Fälth, M.; Nydahl, K.; Svenningsson, P.; Andrén, P. E. Neuropeptidomics: MS applied to the discovery of novel peptides from the brain. *Anal Chem* **2007**, *79*, 15-16, 18-21.

(154) Foreman, R. E.; George, A. L.; Reimann, F.; Gribble, F. M.; Kay, R. G. Peptidomics: A Review of Clinical Applications and Methodologies. *J Proteome Res* **2021**, *20*, 3782-3797.

(155) Petricoin, E. F.; Belluco, C.; Araujo, R. P.; Liotta, L. A. The blood peptidome: a higher dimension of information content for cancer biomarker discovery. *Nat Rev Cancer* **2006**, *6*, 961-967.

(156) Oliver, S. G.; Winson, M. K.; Kell, D. B.; Baganz, F. Systematic functional analysis of the yeast genome. *Trends Biotechnol* **1998**, *16*, 373-378.

(157) Teusink, B.; Baganz, F.; Westerhoff, H. V.; Oliver, S. G. In *Methods in Microbiology*, Brown, A. J. P.; Tuite, M., Eds.; Academic Press, 1998, pp 297-336.

(158) Johnson, C. H.; Ivanisevic, J.; Siuzdak, G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol* **2016**, *17*, 451-459.

(159) Wellen, K. E.; Hatzivassiliou, G.; Sachdeva, U. M.; Bui, T. V.; Cross, J. R.; Thompson, C. B. ATP-citrate lyase links cellular metabolism to histone acetylation. *Science* **2009**, *324*, 1076-1080.

(160) Nakahata, Y.; Kaluzova, M.; Grimaldi, B.; Sahar, S.; Hirayama, J.; Chen, D.; Guarente, L. P.; Sassone-Corsi, P. The NAD+-dependent deacetylase SIRT1 modulates CLOCK-mediated chromatin remodeling and circadian control. *Cell* **2008**, *134*, 329-340.

(161) Sillner, N.; Walker, A.; Hemmler, D.; Bazanella, M.; Heinzmann, S. S.; Haller, D.; Schmitt-Kopplin, P. Milk-Derived Amadori Products in Feces of Formula-Fed Infants. *Journal of Agricultural and Food Chemistry* **2019**, *67*, 8061-8069.

(162) Zhou, X.; Ulaszewska, M. M.; De Gobba, C.; Rinnan, Å.; Poulsen, M. W.; Chen, J.; Mattivi, F.; Hedegaard, R. V.; Skibsted, L. H.; Dragsted, L. O. New Advanced Glycation End Products Observed in Rat Urine by Untargeted Metabolomics after Feeding with Heat-Treated Skimmed Milk Powder. *Mol Nutr Food Res* **2021**, *65*, e2001049.

(163) Ahmed, N.; Thornalley, P. J. Chromatographic assay of glycation adducts in human serum albumin glycated in vitro by derivatization with 6-aminoquinolyl-N-hydroxysuccinimidyl-carbamate and intrinsic fluorescence. *Biochem J* **2002**, *364*, 15-24.

(164) Zhou, Y.; Lin, Q.; Jin, C.; Cheng, L.; Zheng, X.; Dai, M.; Zhang, Y. Simultaneous analysis of N(ε) -(carboxymethyl)Lysine and N(ε) -(carboxyethyl)lysine in foods by ultra-performance liquid chromatography-mass spectrometry with derivatization by 9-fluorenylmethyl chloroformate. *J Food Sci* **2015**, *80*, C207-217.

(165) Hashimoto, C.; Iwaihara, Y.; Chen, S. J.; Tanaka, M.; Watanabe, T.; Matsui, T. Highly-sensitive detection of free advanced glycation end-products by liquid chromatography-electrospray ionization-tandem mass spectrometry with 2,4,6-trinitrobenzene sulfonate derivatization. *Anal Chem* **2013**, *85*, 4289-4295.

(166) Mann, M.; Jensen, O. N. Proteomic analysis of post-translational modifications. *Nat Biotechnol* **2003**, *21*, 255-261.

(167) Zolg, D. P.; Wilhelm, M.; Schmidt, T.; Médard, G.; Zerweck, J.; Knaute, T.; Wenschuh, H.; Reimer, U.; Schnatbaum, K.; Kuster, B. ProteomeTools: Systematic Characterization of 21 Post-translational Protein Modifications by Liquid Chromatography Tandem Mass Spectrometry (LC-MS/MS) Using Synthetic Peptides. *Mol Cell Proteomics* **2018**, *17*, 1850-1863.

(168) Rojas Echeverri, J. C.; Volke, D.; Milkovska-Stamenova, S.; Hoffmann, R. Evaluating Peptide Fragment Ion Detection Using Traveling Wave Ion Mobility Spectrometry with Signal-Enhanced MS(E) (SEMS(E)). *Anal Chem* **2022**.

(169) Steen, H.; Mann, M. The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* **2004**, *5*, 699-711.

(170) Altelaar, A. F.; Munoz, J.; Heck, A. J. Next-generation proteomics: towards an integrative view of proteome dynamics. *Nat Rev Genet* **2013**, *14*, 35-48.

(171) Vuckovic, D. Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. *Anal Bioanal Chem* **2012**, *403*, 1523-1548.

(172) Feist, P.; Hummon, A. B. Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples. *Int J Mol Sci* **2015**, *16*, 3537-3563.

(173) Kuljanin, M.; Dieters-Castator, D. Z.; Hess, D. A.; Postovit, L. M.; Lajoie, G. A. Comparison of sample preparation techniques for large-scale proteomics. *Proteomics* **2017**, *17*.

(174) Hughes, C. S.; Moggridge, S.; Müller, T.; Sorensen, P. H.; Morin, G. B.; Krijgsveld, J. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat Protoc* **2019**, *14*, 68-85.

(175) Zhang, Q.; Monroe, M. E.; Schepmoes, A. A.; Clauss, T. R.; Gritsenko, M. A.; Meng, D.; Petyuk, V. A.; Smith, R. D.; Metz, T. O. Comprehensive identification of glycated peptides and their glycation motifs in plasma and erythrocytes of control and diabetic subjects. *J Proteome Res* **2011**, *10*, 3076-3088.

(176) Nakayama, T.; Hayase, F.; Kato, H. Formation of ε-(2-Formyl-5-hydroxy-methyl-pyrrol-1-yl)-l-norleucine in the Maillard Reaction between d-Glucose and l-Lysine. *Agricultural and Biological Chemistry* **1980**, *44*, 1201-1202.

(177) Shipanova, I. N.; Glomb, M. A.; Nagaraj, R. H. Protein modification by methylglyoxal: chemical nature and synthetic mechanism of a major fluorescent adduct. *Arch Biochem Biophys* **1997**, *344*, 29-36.

(178) Mori, N.; Bai, Y.; Ueno, H.; Manning, J. M. Sequence-dependent reactivity of model peptides with glyceraldehyde. *Carbohydr Res* **1989**, *189*, 49-63.

(179) Kim, J.-S.; Lee, Y.-S. Study of maillard reaction products in model aqueous and water/ethanol systems containing glucose and glycine, diglycine, and triglycine. *Food Science and Biotechnology* **2010**, *19*, 1471-1477.

(180) Vinson, J. A.; Howard, T. B. Inhibition of protein glycation and advanced glycation end products by ascorbic acid and other vitamins and nutrients. *The Journal of Nutritional Biochemistry* **1996**, *7*, 659-663.

(181) Hemmler, D.; Roullier-Gall, C.; Marshall, J. W.; Rychlik, M.; Taylor, A. J.; Schmitt-Kopplin, P. Evolution of Complex Maillard Chemical Reactions, Resolved in Time. *Scientific Reports* **2017**, *7*, 3227.

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

(183) Bell, L. N. Maillard reaction as influenced by buffer type and concentration. *Food Chemistry* **1997**, *59*, 143-147.

(184) Watkins, N. G.; Neglia-Fisher, C. I.; Dyer, D. G.; Thorpe, S. R.; Baynes, J. W. Effect of phosphate on the kinetics and specificity of glycation of protein. *J Biol Chem* **1987**, *262*, 7207-7212.

(185) Jacobitz, A. W.; Dykstra, A. B.; Spahr, C.; Agrawal, N. J. Effects of Buffer Composition on Site-Specific Glycation of Lysine Residues in Monoclonal Antibodies. *J Pharm Sci* **2020**, *109*, 293-300.

(186) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying small molecules via high resolution mass spectrometry: communicating confidence. *Environ Sci Technol* **2014**, *48*, 2097-2098.

(187) Aicheler, F.; Li, J.; Hoene, M.; Lehmann, R.; Xu, G.; Kohlbacher, O. Retention Time Prediction Improves Identification in Nontargeted Lipidomics Approaches. *Anal Chem* **2015**, *87*, 7698-7704.

(188) Kind, T.; Fiehn, O. Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. *BMC Bioinformatics* **2006**, *7*, 234.

(189) Xu, P.; Duong, D. M.; Peng, J. Systematical optimization of reverse-phase chromatography for shotgun proteomics. *J Proteome Res* **2009**, *8*, 3944-3950.

(190) Josic, D.; Kovac, S. Reversed-phase High Performance Liquid Chromatography of proteins. *Curr Protoc Protein Sci* **2010**, *Chapter 8*, 8.7.1-8.7.22.

(191) Meier, F.; Köhler, N. D.; Brunner, A. D.; Wanka, J. H.; Voytik, E.; Strauss, M. T.; Theis, F. J.; Mann, M. Deep learning the collisional cross sections of the peptide universe from a million experimental values. *Nat Commun* **2021**, *12*, 1185.

(192) Bazanella, M.; Maier, T. V.; Clavel, T.; Lagkouvardos, I.; Lucio, M.; Maldonado-Gòmez, M. X.; Autran, C.; Walter, J.; Bode, L.; Schmitt-Kopplin, P.; Haller, D. Randomized controlled trial on the impact of early-life intervention with bifidobacteria on the healthy infant fecal microbiota and metabolome. *Am J Clin Nutr* **2017**, *106*, 1274-1286.

(193) Hinrichsen, F.; Hamm, J.; Westermann, M.; Schröder, L.; Shima, K.; Mishra, N.; Walker, A.; Sommer, N.; Klischies, K.; Prasse, D.; Zimmermann, J.; Kaiser, S.; Bordoni, D.; Fazio, A.; Marinos, G.; Laue, G.; Imm, S.; Tremaroli, V.; Basic, M.; Häsler, R., et al. Microbial regulation of hexokinase 2 links mitochondrial metabolism and cell death in colitis. *Cell Metab* **2021**, *33*, 2355-2366.e2358.

(194) Sillner, N.; Walker, A.; Harrieder, E. M.; Schmitt-Kopplin, P.; Witting, M. Development and application of a HILIC UHPLC-MS method for polar fecal metabolome profiling. *J Chromatogr B Analyt Technol Biomed Life Sci* **2019**, *1109*, 142-148.

(195) Tang, D. Q.; Zou, L.; Yin, X. X.; Ong, C. N. HILIC-MS for metabolomics: An attractive and complementary approach to RPLC-MS. *Mass Spectrom Rev* **2016**, *35*, 574-600.

(196) Yan, Y.; Hemmler, D.; Schmitt-Kopplin, P. HILIC-MS for Untargeted Profiling of the Free Glycation Product Diversity. *Metabolites* **2022**, *12*.

(197) Boersema, P. J.; Mohammed, S.; Heck, A. J. Hydrophilic interaction liquid chromatography (HILIC) in proteomics. *Anal Bioanal Chem* **2008**, *391*, 151-159.

(198) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray ionization for mass spectrometry of large biomolecules. *Science* **1989**, *246*, 64-71.

(199) Radionova, A.; Filippov, I.; Derrick, P. J. In pursuit of resolution in time-of-flight mass spectrometry: A historical perspective. *Mass Spectrom Rev* **2016**, *35*, 738-757.

(200) He, X.; Taylor, A. M.; Jarvis, M.; Wang, A. Ultra-Fast Forensic Toxicological Screening and Quantitation under 3 Minutes using SCIEX X500R QTOF System and SCIEX OS 1.0 Software.

(201) Zhang, A.; Sun, H.; Wang, P.; Han, Y.; Wang, X. Modern analytical techniques in metabolomics analysis. *Analyst* **2012**, *137*, 293-300.

(202) Gross, J. H. *Mass spectrometry: a textbook*; Springer Science & Business Media, 2006.

(203) Frey, R.; Schlag, E. W.; Google Patents, 1988.

(204) Stahl-Zeng, J.; Lange, V.; Ossola, R.; Eckhardt, K.; Krek, W.; Aebersold, R.; Domon, B. High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol Cell Proteomics* **2007**, *6*, 1809-1817.

(205) Macklin, P. Key challenges facing data-driven multicellular systems biology. *Gigascience* **2019**, *8*.

(206) Chavan, S. S.; Shaughnessy, J. D., Jr.; Edmondson, R. D. Overview of biological database mapping services for interoperation between different 'omics' datasets. *Hum Genomics* **2011**, *5*, 703-708.

(207) Glish, G. L.; Burinsky, D. J. Hybrid mass spectrometers for tandem mass spectrometry. *J Am Soc Mass Spectrom* **2008**, *19*, 161-172.

(208) Wei, R.; Li, G.; Seymour, A. B. High-throughput and multiplexed LC/MS/MRM method for targeted metabolomics. *Anal Chem* **2010**, *82*, 5527-5533.

(209) Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol Cell Proteomics* **2012**, *11*, 1475-1488.

(210) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26*, 966-968.

(211) Davies, V.; Wandy, J.; Weidt, S.; van der Hooft, J. J. J.; Miller, A.; Daly, R.; Rogers, S. Rapid Development of Improved Data-Dependent Acquisition Strategies. *Anal Chem* **2021**, *93*, 5676-5683.

(212) Senko, M. W.; Remes, P. M.; Canterbury, J. D.; Mathur, R.; Song, Q.; Eliuk, S. M.; Mullen, C.; Earley, L.; Hardman, M.; Blethrow, J. D.; Bui, H.; Specht, A.; Lange, O.; Denisov, E.; Makarov, A.;

Horning, S.; Zabrouskov, V. Novel parallelized quadrupole/linear ion trap/Orbitrap tribrid mass spectrometer improving proteome coverage and peptide identification rates. *Anal Chem* **2013**, *85*, 11710-11714.

(213) Hebert, A. S.; Thöing, C.; Riley, N. M.; Kwiecien, N. W.; Shiskova, E.; Huguet, R.; Cardasis, H. L.; Kuehn, A.; Eliuk, S.; Zabrouskov, V.; Westphall, M. S.; McAlister, G. C.; Coon, J. J. Improved Precursor Characterization for Data-Dependent Mass Spectrometry. *Anal Chem* **2018**, *90*, 2333-2340.

(214) Riley, N. M.; Hebert, A. S.; Coon, J. J. Proteomics Moves into the Fast Lane. *Cell Syst* **2016**, *2*, 142-143.

(215) Hebert, A. S.; Richards, A. L.; Bailey, D. J.; Ulbrich, A.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. The one hour yeast proteome. *Mol Cell Proteomics* **2014**, *13*, 339-347.

(216) Trujillo, E. A.; Hebert, A. S.; Brademan, D. R.; Coon, J. J. Maximizing Tandem Mass Spectrometry Acquisition Rates for Shotgun Proteomics. *Anal Chem* **2019**, *91*, 12625-12629.

(217) Doerr, A. DIA mass spectrometry. *Nature Methods* **2015**, *12*, 35-35.

(218) Bilbao, A.; Varesio, E.; Luban, J.; Strambio-De-Castillia, C.; Hopfgartner, G.; Müller, M.; Lisacek, F. Processing strategies and software solutions for data-independent acquisition in mass spectrometry. *Proteomics* **2015**, *15*, 964-980.

(219) Röst, H. L.; Rosenberger, G.; Navarro, P.; Gillet, L.; Miladinović, S. M.; Schubert, O. T.; Wolski, W.; Collins, B. C.; Malmström, J.; Malmström, L.; Aebersold, R. Publisher Correction: OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat Biotechnol* **2020**, *38*, 374.

(220) Edman, P. A method for the determination of amino acid sequence in peptides. *Arch Biochem* **1949**, *22*, 475.

(221) Mann, M. The Rise of Mass Spectrometry and the Fall of Edman Degradation. *Clin Chem* **2016**, *62*, 293-294.

(222) Bergman, T.; Gheorghe, M. T.; Hjelmqvist, L.; Jörnvall, H. Alcoholytic deblocking of N-terminally acetylated peptides and proteins for sequence analysis. *FEBS Lett* **1996**, *390*, 199-202.

(223) Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M. Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **1996**, *379*, 466-469.

(224) Roepstorff, P.; Fohlman, J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. *Biomed Mass Spectrom* **1984**, *11*, 601.

(225) Biemann, K. Mass spectrometry of peptides and proteins. *Annu Rev Biochem* **1992**, *61*, 977-1010.

(226) Conrotto, P.; Hellman, U. Sulfonation chemistry as a powerful tool for MALDI TOF/TOF de novo sequencing and post-translational modification analysis. *J Biomol Tech* **2005**, *16*, 441-452.

(227) Warwood, S.; Mohammed, S.; Cristea, I. M.; Evans, C.; Whetton, A. D.; Gaskell, S. J. Guanidination chemistry for qualitative and quantitative proteomics. *Rapid Commun Mass Spectrom* **2006**, *20*, 3245-3256.

(228) Hennrich, M. L.; Boersema, P. J.; van den Toorn, H.; Mischerikow, N.; Heck, A. J.; Mohammed, S. Effect of chemical modifications on peptide fragmentation behavior upon electron transfer induced dissociation. *Anal Chem* **2009**, *81*, 7814-7822.

(229) Vaisar, T.; Urban, J. Probing the proline effect in CID of protonated peptides. *J Mass Spectrom* **1996**, *31*, 1185-1187.

(230) Schmutzler, S.; Wölk, M.; Hoffmann, R. Differentiation and Quantitation of Coeluting Isomeric Amadori and Heyns Peptides Using Sugar-Specific Fragment Ion Ratios. *Anal Chem* **2022**, *94*, 7909-7917.

(231) Breitling, R.; Pitt, A. R.; Barrett, M. P. Precision mapping of the metabolome. *Trends Biotechnol* **2006**, *24*, 543-548.

(232) Chokkathukalam, A.; Kim, D. H.; Barrett, M. P.; Breitling, R.; Creek, D. J. Stable isotope-labeling studies in metabolomics: new insights into structure and dynamics of metabolic networks. *Bioanalysis* **2014**, *6*, 511-524.

(233) Fan, T. W.; Lorkiewicz, P. K.; Sellers, K.; Moseley, H. N.; Higashi, R. M.; Lane, A. N. Stable isotope-resolved metabolomics and applications for drug development. *Pharmacol Ther* **2012**, *133*, 366-391.

(234) Creek, D. J.; Chokkathukalam, A.; Jankevics, A.; Burgess, K. E.; Breitling, R.; Barrett, M. P. Stable isotope-assisted metabolomics for network-wide metabolic pathway elucidation. *Anal Chem* **2012**, *84*, 8442-8447.

(235) Yu, F.; Teo, G. C.; Kong, A. T.; Haynes, S. E.; Avtonomov, D. M.; Geiszler, D. J.; Nesvizhskii, A. I. Identification of modified peptides using localization-aware open search. *Nature Communications* **2020**, *11*, 4065.

(236) Wang, F.; Liigand, J.; Tian, S.; Arndt, D.; Greiner, R.; Wishart, D. S. CFM-ID 4.0: More Accurate ESI-MS/MS Spectral Prediction and Compound Identification. *Anal Chem* **2021**, *93*, 11692-11700.

(237) Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. MetFrag relaunched: incorporating strategies beyond in silico fragmentation. *J Cheminform* **2016**, *8*, 3.

(238) Heinonen, M.; Shen, H.; Zamboni, N.; Rousu, J. Metabolite identification and molecular fingerprint prediction through machine learning. *Bioinformatics* **2012**, *28*, 2333-2341.

(239) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W. T.; Crüsemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderón, M.; Kersten, R. D., et al. Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat Biotechnol* **2016**, *34*, 828-837.

(240) Xing, S.; Hu, Y.; Yin, Z.; Liu, M.; Tang, X.; Fang, M.; Huan, T. Retrieving and Utilizing Hypothetical Neutral Losses from Tandem Mass Spectra for Spectral Similarity Analysis and Unknown Metabolite Annotation. *Anal Chem* **2020**, *92*, 14476-14483.

(241) Stein, S. Mass spectral reference libraries: an ever-expanding resource for chemical identification. *Anal Chem* **2012**, *84*, 7274-7282.

(242) Milman, B. L.; Zhurkovich, I. K. Mass spectral libraries: A statistical review of the visible use. *TrAC Trends in Analytical Chemistry* **2016**, *80*, 636-640.

(243) Kind, T.; Tsugawa, H.; Cajka, T.; Ma, Y.; Lai, Z.; Mehta, S. S.; Wohlgemuth, G.; Barupal, D. K.; Showalter, M. R.; Arita, M.; Fiehn, O. Identification of small molecules using accurate mass MS/MS search. *Mass Spectrom Rev* **2018**, *37*, 513-532.

(244) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N., et al. MassBank: a public repository for sharing mass spectral data for life sciences. *J Mass Spectrom* **2010**, *45*, 703-714.

(245) Falkner, J. A.; Falkner, J. W.; Yocum, A. K.; Andrews, P. C. A spectral clustering approach to MS/MS identification of post-translational modifications. *J Proteome Res* **2008**, *7*, 4614-4622.

(246) Moorthy, A. S.; Wallace, W. E.; Kearsley, A. J.; Tchekhovskoi, D. V.; Stein, S. E. Combining Fragment-Ion and Neutral-Loss Matching during Mass Spectral Library Searching: A New General Purpose Algorithm Applicable to Illicit Drug Identification. *Anal Chem* **2017**, *89*, 13261-13268.

(247) Stein, S. E.; Scott, D. R. Optimization and testing of mass spectral library search algorithms for compound identification. *J Am Soc Mass Spectrom* **1994**, *5*, 859-866.

(248) Eng, J. K.; McCormack, A. L.; Yates, J. R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* **1994**, *5*, 976-989.

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

(250) Geer, L. Y.; Markey, S. P.; Kowalak, J. A.; Wagner, L.; Xu, M.; Maynard, D. M.; Yang, X.; Shi, W.; Bryant, S. H. Open mass spectrometry search algorithm. *J Proteome Res* **2004**, *3*, 958-964.

(251) Craig, R.; Beavis, R. C. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics* **2004**, *20*, 1466-1467.

(252) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: a peptide search engine integrated into the MaxQuant environment. *J Proteome Res* **2011**, *10*, 1794-1805.

(253) Keil, B. *Specificity of proteolysis*; Springer Science & Business Media, 2012.

(254) Rodriguez, J.; Gupta, N.; Smith, R. D.; Pevzner, P. A. Does trypsin cut before proline? *J Proteome Res* **2008**, *7*, 300-305.

(255) Gessulat, S.; Schmidt, T.; Zolg, D. P.; Samaras, P.; Schnatbaum, K.; Zerweck, J.; Knaute, T.; Rechenberger, J.; Delanghe, B.; Huhmer, A.; Reimer, U.; Ehrlich, H. C.; Aiche, S.; Kuster, B.; Wilhelm, M. Prosit: proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat Methods* **2019**, *16*, 509-518.

(256) Allen, F.; Greiner, R.; Wishart, D. Competitive fragmentation modeling of ESI-MS/MS spectra for putative metabolite identification. *Metabolomics* **2015**, *11*, 98-110.

(257) Allen, F.; Pon, A.; Wilson, M.; Greiner, R.; Wishart, D. CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. *Nucleic Acids Res* **2014**, *42*, W94-99.

(258) Wolf, S.; Schmidt, S.; Müller-Hannemann, M.; Neumann, S. In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinformatics* **2010**, *11*, 148.

(259) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* **2003**, *17*, 2337-2342.

(260) Houel, S.; Abernathy, R.; Renganathan, K.; Meyer-Arendt, K.; Ahn, N. G.; Old, W. M. Quantifying the impact of chimera MS/MS spectra on peptide identification in large-scale proteomics studies. *J Proteome Res* **2010**, *9*, 4152-4160.

(261) Taylor, J. A.; Johnson, R. S. Sequence database searches via de novo peptide sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom* **1997**, *11*, 1067-1075.

(262) Taylor, J. A.; Johnson, R. S. Implementation and uses of automated de novo peptide sequencing by tandem mass spectrometry. *Anal Chem* **2001**, *73*, 2594-2604.

(263) Bartels, C. Fast algorithm for peptide sequencing by mass spectroscopy. *Biomed Environ Mass Spectrom* **1990**, *19*, 363-368.

(264) Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A. A.; Melnik, A. V.; Meusel, M.; Dorrestein, P. C.; Rousu, J.; Böcker, S. SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information. *Nat Methods* **2019**, *16*, 299-302.

(265) Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S. Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc Natl Acad Sci U S A* **2015**, *112*, 12580-12585.

(266) Martin, L.; Latypova, X.; Terro, F. Post-translational modifications of tau protein: implications for Alzheimer's disease. *Neurochem Int* **2011**, *58*, 458-471.

(267) Gajjala, P. R.; Fliser, D.; Speer, T.; Jankowski, V.; Jankowski, J. Emerging role of post-translational modifications in chronic kidney disease and cardiovascular disease. *Nephrol Dial Transplant* **2015**, *30*, 1814-1824.

(268) Hermann, J.; Schurgers, L.; Jankowski, V. Identification and characterization of post-translational modifications: Clinical implications. *Mol Aspects Med* **2022**, *86*, 101066.

(269) Reinders, J.; Sickmann, A. Modificomics: posttranslational modifications beyond protein phosphorylation and glycosylation. *Biomol Eng* **2007**, *24*, 169-177.

(270) Gomord, V.; Faye, L. Posttranslational modification of therapeutic proteins in plants. *Curr Opin Plant Biol* **2004**, *7*, 171-181.

(271) Di Sanzo, S.; Spengler, K.; Leheis, A.; Kirkpatrick, J. M.; Rändler, T. L.; Baldensperger, T.; Dau, T.; Henning, C.; Parca, L.; Marx, C.; Wang, Z. Q.; Glomb, M. A.; Ori, A.; Heller, R. Mapping protein carboxymethylation sites provides insights into their role in proteostasis and cell proliferation. *Nat Commun* **2021**, *12*, 6743.

(272) Prasanna, R. R.; Venkatraman, K.; Vijayalakshmi, M. Pseudoaffinity chromatography enrichment of glycated peptides for monitoring advanced glycation end products (AGEs) in metabolic disorders. *Journal of Proteins & Proteomics* **2016**, *7*.

(273) Priego-Capote, F.; Scherl, A.; Müller, M.; Waridel, P.; Lisacek, F.; Sanchez, J. C. Glycation isotopic labeling with 13C-reducing sugars for quantitative analysis of glycated proteins in human plasma. *Mol Cell Proteomics* **2010**, *9*, 579-592.

(274) Soboleva, A.; Schmidt, R.; Vikhnina, M.; Grishina, T.; Frolov, A. Maillard Proteomics: Opening New Pages. *International journal of molecular sciences* **2017**, *18*, 2677.

(275) Ajandouz, E. H.; Desseaux, V.; Tazi, S.; Puigserver, A. Effects of temperature and pH on the kinetics of caramelisation, protein cross-linking and Maillard reactions in aqueous model systems. *Food chemistry* **2008**, *107*, 1244-1252.

(276) Oh, M.-J.; Kim, Y.; Hoon Lee, S.; Lee, K.-W.; Park, H.-Y. Prediction of CML contents in the Maillard reaction products for casein-monosaccharides model. *Food Chemistry* **2018**, *267*, 271-276.

(277) Tessier, F. J.; Birlouez-Aragon, I. Health effects of dietary Maillard reaction products: the results of ICARE and other studies. *Amino Acids* **2012**, *42*, 1119-1131.

(278) Somoza, V. Five years of research on health risks and benefits of Maillard reaction products: An update. *Molecular Nutrition & Food Research* **2005**, *49*, 663-672.

(279) Brownlee, M.; Cerami, A.; Vlassara, H. Advanced Glycosylation End Products in Tissue and the Biochemical Basis of Diabetic Complications. *New England Journal of Medicine* **1988**, *318*, 1315-1321.

(280) Schmidt, A. M.; Hasu, M.; Popov, D.; Zhang, J. H.; Chen, J.; Yan, S. D.; Brett, J.; Cao, R.; Kuwabara, K.; Costache, G.; et al. Receptor for advanced glycation end products (AGEs) has a central role in vessel wall interactions and gene activation in response to circulating AGE proteins. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 8807-8811.

(281) Miyata, T.; Hori, O.; Zhang, J.; Yan, S. D.; Ferran, L.; Iida, Y.; Schmidt, A. M. The receptor for advanced glycation end products (RAGE) is a central mediator of the interaction of AGE-beta2microglobulin with human mononuclear phagocytes via an oxidant-sensitive pathway.

Implications for the pathogenesis of dialysis-related amyloidosis. *The Journal of clinical investigation* **1996**, *98*, 1088-1094.

(282) Schmidt, A. M.; Hori, O.; Chen, J. X.; Li, J. F.; Crandall, J.; Zhang, J.; Cao, R.; Yan, S. D.; Brett, J.; Stern, D. Advanced glycation endproducts interacting with their endothelial receptor induce expression of vascular cell adhesion molecule-1 (VCAM-1) in cultured human endothelial cells and in mice. A potential mechanism for the accelerated vasculopathy of diabetes. *The Journal of clinical investigation* **1995**, *96*, 1395-1403.

(283) Wautier, J. L.; Wautier, M. P.; Schmidt, A. M.; Anderson, G. M.; Hori, O.; Zoukourian, C.; Capron, L.; Chappey, O.; Yan, S. D.; Brett, J.; et al. Advanced glycation end products (AGEs) on the surface of diabetic erythrocytes bind to the vessel wall via a specific receptor inducing oxidant stress in the vasculature: a link between surface-associated AGEs and diabetic complications. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 7742-7746.

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

(285) Chakrabarti, S.; Jahandideh, F.; Wu, J. Food-derived bioactive peptides on inflammation and oxidative stress. *Biomed Res Int* **2014**, *2014*, 608979.

(286) Rutherfurd-Markwick, K. J. Food proteins as a source of bioactive peptides with diverse functions. *Br J Nutr* **2012**, *108 Suppl 2*, S149-157.

(287) Wang, H.-Y.; Qian, H.; Yao, W.-R. Melanoidins produced by the Maillard reaction: Structure and biological activity. *Food Chemistry* **2011**, *128*, 573-584.

(288) Borrelli, R. C.; Mennella, C.; Barba, F.; Russo, M.; Russo, G. L.; Krome, K.; Erbersdobler, H. F.; Faist, V.; Fogliano, V. Characterization of coloured compounds obtained by enzymatic extraction of bakery products. *Food Chem Toxicol* **2003**, *41*, 1367-1374.

(289) Lindenmeier, M.; Faist, V.; Hofmann, T. Structural and functional characterization of pronyl-lysine, a novel protein modification in bread crust melanoidins showing in vitro antioxidative and phase I/II enzyme modulating activity. *J Agric Food Chem* **2002**, *50*, 6997-7006.

(290) Li-Chan, E. C. Y. Bioactive peptides and protein hydrolysates: research trends and challenges for application as nutraceuticals and functional food ingredients. *Current Opinion in Food Science* **2015**, *1*, 28-37.

(291) Wang, W.; Zhang, L.; Wang, Z.; Wang, X.; Liu, Y. Physicochemical and sensory variables of Maillard reaction products obtained from Takifugu obscurus muscle hydrolysates. *Food Chemistry* **2019**, *290*, 40-46.

(292) Zhang, Y.; Wang, Y.; Jiang, F.; Jin, H. Sensory characteristics of Maillard reaction products from chicken protein hydrolysates with different degrees of hydrolysis. *CyTA - Journal of Food* **2019**, *17*, 221-227.

(293) Chevalier, F.; Chobert, J.-M.; Genot, C.; Haertlé, T. Scavenging of Free Radicals, Antimicrobial, and Cytotoxic Activities of the Maillard Reaction Products of β-Lactoglobulin Glycated with Several Sugars. *Journal of Agricultural and Food Chemistry* **2001**, *49*, 5031-5038.

(294) Chuyen, N. V.; Ijichi, K.; Umetsu, H.; Moteki, K. In *Process-Induced Chemical Changes in Food*, Shahidi, F.; Ho, C.-T.; van Chuyen, N., Eds.; Springer US: Boston, MA, 1998, pp 201-212.

(295) Jing, H.; Kitts, D. D. Chemical and biochemical properties of casein–sugar Maillard reaction products. *Food and Chemical Toxicology* **2002**, *40*, 1007-1015.

(296) Dong, S.; Wei, B.; Chen, B.; McClements, D. J.; Decker, E. A. Chemical and Antioxidant Properties of Casein Peptide and Its Glucose Maillard Reaction Products in Fish Oil-in-Water Emulsions. *Journal of Agricultural and Food Chemistry* **2011**, *59*, 13311-13317.

(297) Meltretter, J.; Wüst, J.; Dittrich, D.; Lach, J.; Ludwig, J.; Eichler, J.; Pischetsrieder, M. Untargeted Proteomics-Based Profiling for the Identification of Novel Processing-Induced Protein Modifications in Milk. *Journal of Proteome Research* **2020**, *19*, 805-818.

(298) Meltretter, J.; Wüst, J.; Pischetsrieder, M. Modified Peptides as Indicators for Thermal and Nonthermal Reactions in Processed Milk. *Journal of Agricultural and Food Chemistry* **2014**, *62*, 10903-10915.

(299) Milkovska-Stamenova, S.; Hoffmann, R. Diversity of advanced glycation end products in the bovine milk proteome. *Amino Acids* **2019**, *51*, 891-901.

(300) Poojary, M. M.; Zhang, W.; Greco, I.; De Gobba, C.; Olsen, K.; Lund, M. N. Liquid chromatography quadrupole-Orbitrap mass spectrometry for the simultaneous analysis of advanced glycation end products and protein-derived cross-links in food and biological matrices. *Journal of Chromatography A* **2020**, *1615*, 460767.

(301) Fay, L. B.; Brevard, H. Contribution of mass spectrometry to the study of the Maillard reaction in food. *Mass Spectrometry Reviews* **2005**, *24*, 487-507.

(302) Sibbersen, C.; Schou Oxvig, A. M.; Bisgaard Olesen, S.; Nielsen, C. B.; Galligan, J. J.; Jørgensen, K. A.; Palmfeldt, J.; Johannsen, M. Profiling of Methylglyoxal Blood Metabolism and Advanced Glycation End-Product Proteome Using a Chemical Probe. *ACS Chem Biol* **2018**, *13*, 3294-3305.

(303) Hemmler, D.; Gonsior, M.; Powers, L. C.; Marshall, J. W.; Rychlik, M.; Taylor, A. J.; Schmitt-Kopplin, P. Simulated Sunlight Selectively Modifies Maillard Reaction Products in a Wide Array of Chemical Reactions. *Chemistry – A European Journal* **2019**, *25*, 13208-13217.

(304) Frolov, A.; Hoffmann, P.; Hoffmann, R. Fragmentation behavior of glycated peptides derived from D-glucose, D-fructose and D-ribose in tandem mass spectrometry. *Journal of Mass Spectrometry* **2006**, *41*, 1459-1469.

(305) Stefanowicz, P.; Kijewska, M.; Szewczuk, Z. Sequencing of peptide-derived Amadori products by the electron capture dissociation method. *Journal of Mass Spectrometry* **2009**, *44*, 1047-1052.

(306) Iberg, N.; Flückiger, R. Nonenzymatic glycosylation of albumin in vivo. Identification of multiple glycosylated sites. *Journal of Biological Chemistry* **1986**, *261*, 13542-13545.

(307) Johansen, M. B.; Kiemer, L.; Brunak, S. Analysis and prediction of mammalian protein glycation. *Glycobiology* **2006**, *16*, 844-853.

(308) Xing, H.; Mossine, V. V.; Yaylayan, V. Diagnostic MS/MS fragmentation patterns for the discrimination between Schiff bases and their Amadori or Heyns rearrangement products. *Carbohydrate Research* **2020**, *491*, 107985.

(309) Minkiewicz, P.; Iwaniak, A.; Darewicz, M. BIOPEP-UWM Database of Bioactive Peptides: Current Opportunities. *International journal of molecular sciences* **2019**, *20*, 5978.

(310) Dau, T.; Bartolomucci, G.; Rappsilber, J. Proteomics Using Protease Alternatives to Trypsin Benefits from Sequential Digestion with Trypsin. *Analytical Chemistry* **2020**, *92*, 9523-9527.

(311) Otto, S.; Engberts, J. B. Hydrophobic interactions and chemical reactivity. *Org Biomol Chem* **2003**, *1*, 2809-2820.

(312) Bartoli, G.; Todesco, P. E. Nucleophilic substitution. Linear free energy relations between reactivity and physical properties of leaving groups and substrates. *Accounts of Chemical Research* **1977**, *10*, 125-132.

(313) Edwards, J. O.; Pearson, R. G. The Factors Determining Nucleophilic Reactivities. *Journal of the American Chemical Society* **1962**, *84*, 16-24.

(314) Izzo, H. V.; Ho, C.-T. Peptide-specific Maillard reaction products: a new pathway for flavor chemistry. *Trends in Food Science & Technology* **1992**, *3*, 253-257.

(315) Gavel, Y.; von Heijne, G. Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: implications for protein engineering. *Protein Eng* **1990**, *3*, 433-442.

(316) Sneath, P. H. A. Relations between chemical structure and biological activity in peptides. *Journal of Theoretical Biology* **1966**, *12*, 157-195.

(317) Zigrović, I.; Kidric, J.; Horvat, S. Influence of glycation on cis/trans isomerization and tautomerization in novel morphiceptin-related Amadori compounds. *Glycoconj J* **1998**, *15*, 563-570.

(318) Jakas, A.; Horvat, Š. Synthesis and 13C NMR investigation of novel Amadori compounds (1-amino-1-deoxy-D-fructose derivatives) related to the opioid peptide, leucine–enkephalin. *Journal of the Chemical Society, Perkin Transactions 2* **1996**, 789-794.

(319) Greifenhagen, U.; Frolov, A.; Hoffmann, R. Oxidative degradation of N(ε)-fructosylamine-substituted peptides in heated aqueous systems. *Amino Acids* **2015**, *47*, 1065-1076.

(320) Jakas, A.; Horvat, S. Study of degradation pathways of Amadori compounds obtained by glycation of opioid pentapeptide and related smaller fragments: stability, reactions, and spectroscopic properties. *Biopolymers* **2003**, *69*, 421-431.

(321) Morales, F. J.; van Boekel, M. A. J. S. A Study on Advanced Maillard Reaction in Heated Casein/Sugar Solutions: Colour Formation. *International Dairy Journal* **1998**, *8*, 907-915.

(322) Morales, F. J.; van Boekel, M. A. J. S. A study on advanced Maillard reaction in heated casein/sugar solutions: Fluorescence accumulation. *International Dairy Journal* **1997**, *7*, 675-683.

(323) Grillo, M. A.; Colombatto, S. Advanced glycation end-products (AGEs): involvement in aging and in neurodegenerative diseases. *Amino Acids* **2008**, *35*, 29-36.

(324) Jiang, Z.; Wang, L.; Che, H.; Tian, B. Effects of temperature and pH on angiotensin-I-converting enzyme inhibitory activity and physicochemical properties of bovine casein peptide in aqueous Maillard reaction system. *LWT - Food Science and Technology* **2014**, *59*, 35-42.

(325) Kim, J.-S. Antioxidant activity of Maillard reaction products derived from aqueous and ethanolic glucose-glycine and its oligomer solutions. *Food Science and Biotechnology* **2013**, *22*, 39-46.

(326) Ogasawara, M.; Katsumata, T.; Egi, M. Taste properties of Maillard-reaction products prepared from 1000 to 5000Da peptide. *Food Chemistry* **2006**, *99*, 600-604.

(327) Spellman, D.; O'Cuinn, G.; FitzGerald, R. Bitterness in Bacillus proteinase hydrolysates of whey proteins. *Food Chemistry* **2009**, *114*, 440-446.

(328) Harmel, R.; Fiedler, D. Features and regulation of non-enzymatic post-translational modifications. *Nat Chem Biol* **2018**, *14*, 244-252.

(329) Stadtman, E. R. Protein oxidation and aging. *Free Radic Res* **2006**, *40*, 1250-1258.

(330) Rabbani, N.; Thornalley, P. J. Dicarbonyl stress in cell and tissue dysfunction contributing to ageing and disease. *Biochem Biophys Res Commun* **2015**, *458*, 221-226.

(331) van Boekel, M. A. Kinetic aspects of the Maillard reaction: a critical review. *Nahrung* **2001**, *45*, 150-159.

(332) Mittelmaier, S.; Pischetsrieder, M. Multistep Ultrahigh Performance Liquid Chromatography/Tandem Mass Spectrometry Analysis for Untargeted Quantification of Glycating Activity and Identification of Most Relevant Glycation Products. *Analytical Chemistry* **2011**, *83*, 9660-9668.

(333) Lopez-Clavijo, A. F.; Barrow, M. P.; Rabbani, N.; Thornalley, P. J.; O'Connor, P. B. Determination of Types and Binding Sites of Advanced Glycation End Products for Substance P. *Analytical Chemistry* **2012**, *84*, 10568-10575.

(334) Redman, E. A.; Ramos-Payan, M.; Mellors, J. S.; Ramsey, J. M. Analysis of Hemoglobin Glycation Using Microfluidic CE-MS: A Rapid, Mass Spectrometry Compatible Method for Assessing Diabetes Management. *Analytical Chemistry* **2016**, *88*, 5324-5330.

(335) Lapolla, A.; Gerhardinger, C.; Baldo, L.; Fedele, D.; Keane, A.; Seraglia, R.; Catinella, S.; Traldi, P. A study on in vitro glycation processes by matrix-assisted laser desorption ionization mass spectrometry. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease* **1993**, *1225*, 33-38.

(336) Sillner, N.; Walker, A.; Lucio, M.; Maier, T. V.; Bazanella, M.; Rychlik, M.; Haller, D.; Schmitt-Kopplin, P. Longitudinal Profiles of Dietary and Microbial Metabolites in Formula- and Breastfed Infants. *Front Mol Biosci* **2021**, *8*, 660456.

(337) Schwarzenbolz, U.; Hofmann, T.; Sparmann, N.; Henle, T. Free Maillard Reaction Products in Milk Reflect Nutritional Intake of Glycated Proteins and Can Be Used to Distinguish "Organic" and "Conventionally" Produced Milk. *Journal of Agricultural and Food Chemistry* **2016**, *64*, 5071-5078.

(338) Frolova, N.; Soboleva, A.; Nguyen, V. D.; Kim, A.; Ihling, C.; Eisenschmidt-Bönn, D.; Mamontova, T.; Herfurth, U. M.; Wessjohann, L. A.; Sinz, A.; Birkemeyer, C.; Frolov, A. Probing glycation potential of dietary sugars in human blood by an integrated in vitro approach. *Food Chem* **2021**, *347*, 128951.

(339) Festring, D.; Hofmann, T. Discovery of N2-(1-Carboxyethyl)guanosine 5′-Monophosphate as an Umami-Enhancing Maillard-Modified Nucleotide in Yeast Extracts. *Journal of Agricultural and Food Chemistry* **2010**, *58*, 10614-10622.

(340) Zhang, J.; Zhang, T.; Jiang, L.; Hewitt, D.; Huang, Y.; Kao, Y.-H.; Katta, V. Rapid Identification of Low Level Glycation Sites in Recombinant Antibodies by Isotopic Labeling with 13C6-Reducing Sugars. *Analytical Chemistry* **2012**, *84*, 2313-2320.

(341) Li, S.; Gao, D.; Song, C.; Tan, C.; Jiang, Y. Isotope Labeling Strategies for Acylcarnitines Profile in Biological Samples by Liquid Chromatography–Mass Spectrometry. *Analytical Chemistry* **2019**, *91*, 1701-1705.

(342) Yu, F.; Li, N.; Yu, W. PIPI: PTM-Invariant Peptide Identification Using Coding Method. *Journal of Proteome Research* **2016**, *15*, 4423-4435.

(343) Chi, H.; Liu, C.; Yang, H.; Zeng, W.-F.; Wu, L.; Zhou, W.-J.; Wang, R.-M.; Niu, X.-N.; Ding, Y.-H.; Zhang, Y.; Wang, Z.-W.; Chen, Z.-L.; Sun, R.-X.; Liu, T.; Tan, G.-M.; Dong, M.-Q.; Xu, P.; Zhang, P.-H.; He, S.-M. Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. *Nature Biotechnology* **2018**, *36*, 1059-1061.

(344) Devabhaktuni, A.; Lin, S.; Zhang, L.; Swaminathan, K.; Gonzalez, C. G.; Olsson, N.; Pearlman, S. M.; Rawson, K.; Elias, J. E. TagGraph reveals vast protein modification landscapes from large tandem mass spectrometry datasets. *Nature Biotechnology* **2019**, *37*, 469-479.

(345) Solntsev, S. K.; Shortreed, M. R.; Frey, B. L.; Smith, L. M. Enhanced Global Post-translational Modification Discovery with MetaMorpheus. *Journal of Proteome Research* **2018**, *17*, 1844-1851.

(346) Bittremieux, W.; Meysman, P.; Noble, W. S.; Laukens, K. Fast Open Modification Spectral Library Searching through Approximate Nearest Neighbor Indexing. *Journal of Proteome Research* **2018**, *17*, 3463-3474.

(347) Berger, M. T.; Hemmler, D.; Walker, A.; Rychlik, M.; Marshall, J. W.; Schmitt-Kopplin, P. Molecular characterization of sequence-driven peptide glycation. *Sci Rep* **2021**, *11*, 13294.

(348) Yaylayan, V. A. Classification of the Maillard reaction: A conceptual approach. *Trends in Food Science & Technology* **1997**, *8*, 13-18.

(349) Gatto, L.; Lilley, K. S. MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics* **2012**, *28*, 288-289.

(350) Tabb, D. L.; Thompson, M. R.; Khalsa-Moyers, G.; VerBerkmoes, N. C.; McDonald, W. H. MS2Grouper: group assessment and synthetic replacement of duplicate proteomic tandem mass spectra. *J Am Soc Mass Spectrom* **2005**, *16*, 1250-1261.

(351) Flikka, K.; Meukens, J.; Helsens, K.; Vandekerckhove, J.; Eidhammer, I.; Gevaert, K.; Martens, L. Implementation and application of a versatile clustering tool for tandem mass spectrometry data. *Proteomics* **2007**, *7*, 3245-3258.

(352) Lam, H.; Deutsch, E. W.; Eddes, J. S.; Eng, J. K.; Stein, S. E.; Aebersold, R. Building consensus spectral libraries for peptide identification in proteomics. *Nature Methods* **2008**, *5*, 873-875.

(353) Beardsley, R. L.; Reilly, J. P. Fragmentation of amidinated peptide ions. *J Am Soc Mass Spectrom* **2004**, *15*, 158-167.

(354) Beardsley, R. L.; Sharon, L. A.; Reilly, J. P. Peptide de novo sequencing facilitated by a dual-labeling strategy. *Anal Chem* **2005**, *77*, 6300-6309.

(355) Korwar, A. M.; Vannuruswamy, G.; Jagadeeshaprasad, M. G.; Jayaramaiah, R. H.; Bhat, S.; Regin, B. S.; Ramaswamy, S.; Giri, A. P.; Mohan, V.; Balasubramanyam, M.; Kulkarni, M. J. Development of Diagnostic Fragment Ion Library for Glycated Peptides of Human Serum Albumin: Targeted Quantification in Prediabetic, Diabetic, and Microalbuminuria Plasma by Parallel Reaction Monitoring, SWATH, and MSE*. *Molecular & Cellular Proteomics* **2015**, *14*, 2150-2159.

(356) O'Hair, R. A.; Reid, G. E. The search for stable gas phase b(1) ions derived from aliphatic amino acids: a combined experimental and ab initio study. *Rapid Commun Mass Spectrom* **2000**, *14*, 1220-1225.

(357) Xing, H.; Yaylayan, V. Mechanochemical generation of Schiff bases and Amadori products and utilization of diagnostic MS/MS fragmentation patterns in negative ionization mode for their analysis. *Carbohydrate Research* **2020**, *495*, 108091.

(358) Davídek, T.; Devaud, S.; Robert, F.; Blank, I. Sugar fragmentation in the maillard reaction cascade: isotope labeling studies on the formation of acetic acid by a hydrolytic beta-dicarbonyl cleavage mechanism. *J Agric Food Chem* **2006**, *54*, 6667-6676.

(359) Korhonen, H.; Pihlanto, A. Food-derived bioactive peptides--opportunities for designing future foods. *Curr Pharm Des* **2003**, *9*, 1297-1308.

(360) Lautenbacher, L.; Samaras, P.; Muller, J.; Grafberger, A.; Shraideh, M.; Rank, J.; Fuchs, S. T.; Schmidt, T. K.; The, M.; Dallago, C.; Wittges, H.; Rost, B.; Krcmar, H.; Kuster, B.; Wilhelm, M. ProteomicsDB: toward a FAIR open-source resource for life-science research. *Nucleic Acids Res* **2022**, *50*, D1541-d1552.

(361) Russell, W. K.; Park, Z. Y.; Russell, D. H. Proteolysis in mixed organic-aqueous solvent systems: applications for peptide mass mapping using mass spectrometry. *Anal Chem* **2001**, *73*, 2682-2685.

(362) Park, Z. Y.; Russell, D. H. Thermal denaturation: a useful technique in peptide mass mapping. *Anal Chem* **2000**, *72*, 2667-2670.

(363) Park, Z. Y.; Russell, D. H. Identification of individual proteins in complex protein mixtures by high-resolution, high-mass-accuracy MALDI TOF-mass spectrometry analysis of in-solution thermal denaturation/enzymatic digestion. *Anal Chem* **2001**, *73*, 2558-2564.

(364) Stefanowicz, P.; Kapczynska, K.; Jaremko, M.; Jaremko, Ł.; Szewczuk, Z. A mechanistic study on the fragmentation of peptide-derived Amadori products. *J Mass Spectrom* **2009**, *44*, 1500-1508.

(365) Fedorova, M.; Frolov, A.; Hoffmann, R. Fragmentation behavior of Amadori-peptides obtained by non-enzymatic glycosylation of lysine residues with ADP-ribose in tandem mass spectrometry. *J Mass Spectrom* **2010**, *45*, 664-669.

(366) Wong, K. H.; Abdul Aziz, S.; Mohamed, S. Sensory aroma from Maillard reaction of individual and combinations of amino acids with glucose in acidic conditions. *International journal of food science & technology* **2008**, *43*, 1512-1519.

(367) Jakubczyk, A.; Karaś, M.; Rybczyńska-Tkaczyk, K.; Zielińska, E.; Zieliński, D. Current Trends of Bioactive Peptides-New Sources and Therapeutic Effect. *Foods* **2020**, *9*.

(368) Livesey, G.; Taylor, R. Fructose consumption and consequences for glycation, plasma triacylglycerol, and body weight: meta-analyses and meta-regression models of intervention studies. *Am J Clin Nutr* **2008**, *88*, 1419-1437.

(369) Qi, X.; Tester, R. F. Lactose, Maltose, and Sucrose in Health and Disease. *Mol Nutr Food Res* **2020**, *64*, e1901082.

(370) MacCoss, M. J.; Wu, C. C.; Yates, J. R., 3rd. Probability-based validation of protein identifications using a modified SEQUEST algorithm. *Anal Chem* **2002**, *74*, 5593-5599.

(371) Tyanova, S.; Temu, T.; Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat Protoc* **2016**, *11*, 2301-2319.

List of Tables

List of Figures

List of Schemes