





OPEN

## Molecular characterization of sequence-driven peptide glycation

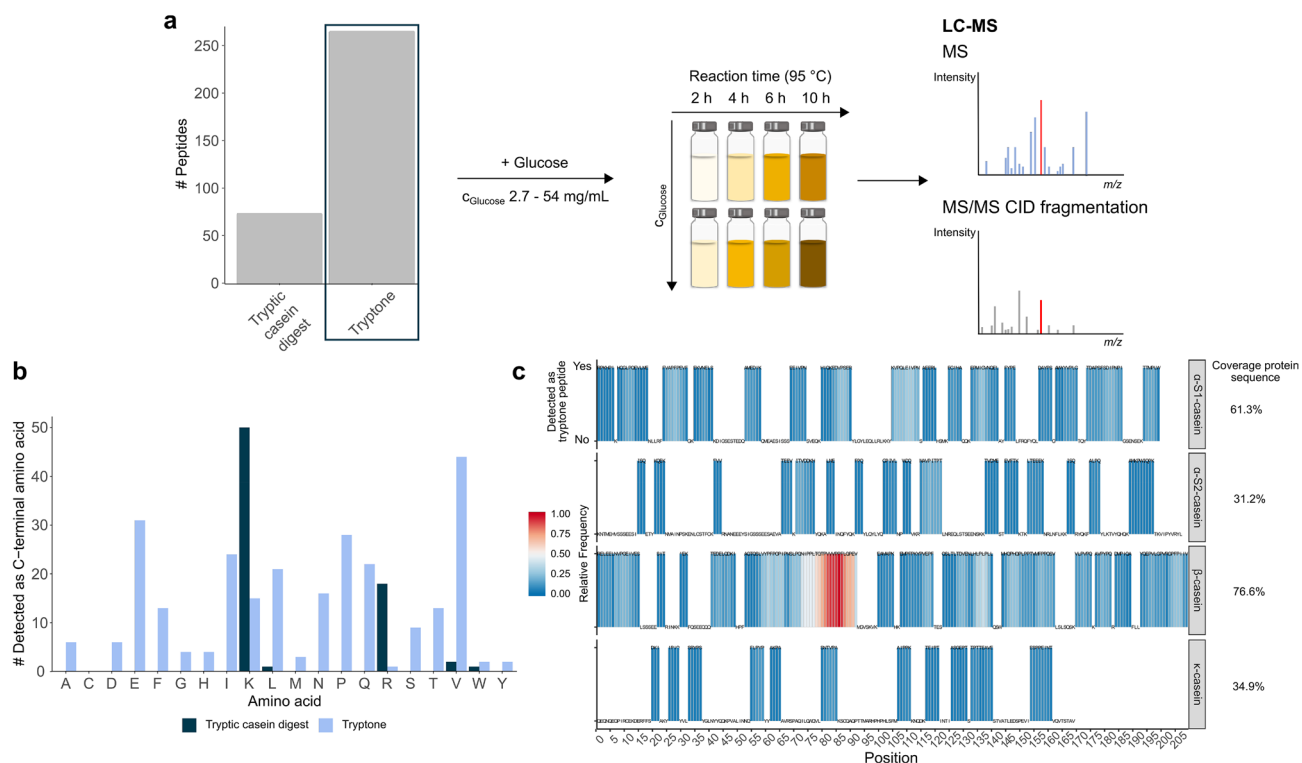
Michelle T. Berger<sup>1,2</sup>, Daniel Hemmler<sup>1,2</sup>, Alesia Walker<sup>2</sup>, Michael Rychlik<sup>1</sup>, James W. Marshall<sup>3</sup> & Philippe Schmitt-Kopplin<sup>1,2</sup>

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.

Glycation presents a ubiquitous non-enzymatic post-translational modification<sup>1,2</sup>, 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)<sup>3,4</sup>. Glycation is a multifactorial reaction, which depends on the nature of the precursors and the reaction conditions, including time and concentration<sup>5–8</sup>. 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<sup>9,10</sup>, and contribute to various pathologies, such as diabetes<sup>11,12</sup>. 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<sup>13–16</sup>. 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<sup>17</sup>. 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<sup>18–23</sup>. Glycation induced alterations in the bioactivity and improvement of sensory attributes have been described for various peptide mixtures<sup>24–32</sup>. 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<sup>26,33,34</sup> or peptide derived MRPs in specific foods<sup>35–38</sup>. 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<sup>10</sup>. 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.

<sup>1</sup>Chair of Analytical Food Chemistry, Technical University Munich, Maximus-von-Imhof-Forum 2, 85354 Freising, Germany. <sup>2</sup>Research Unit Analytical BioGeoChemistry (BGC), Helmholtz Zentrum München, Ingolstädter Landstrasse 1, 85764 Neuherberg, Germany. <sup>3</sup>The Waltham Pet Science Institute, Mars Petcare UK, Waltham-on-the-Wolds, Leicestershire LE14 4RT, UK. ✉email: michelle.berger@tum.de; schmitt-kopplin@helmholtz-muenchen.de

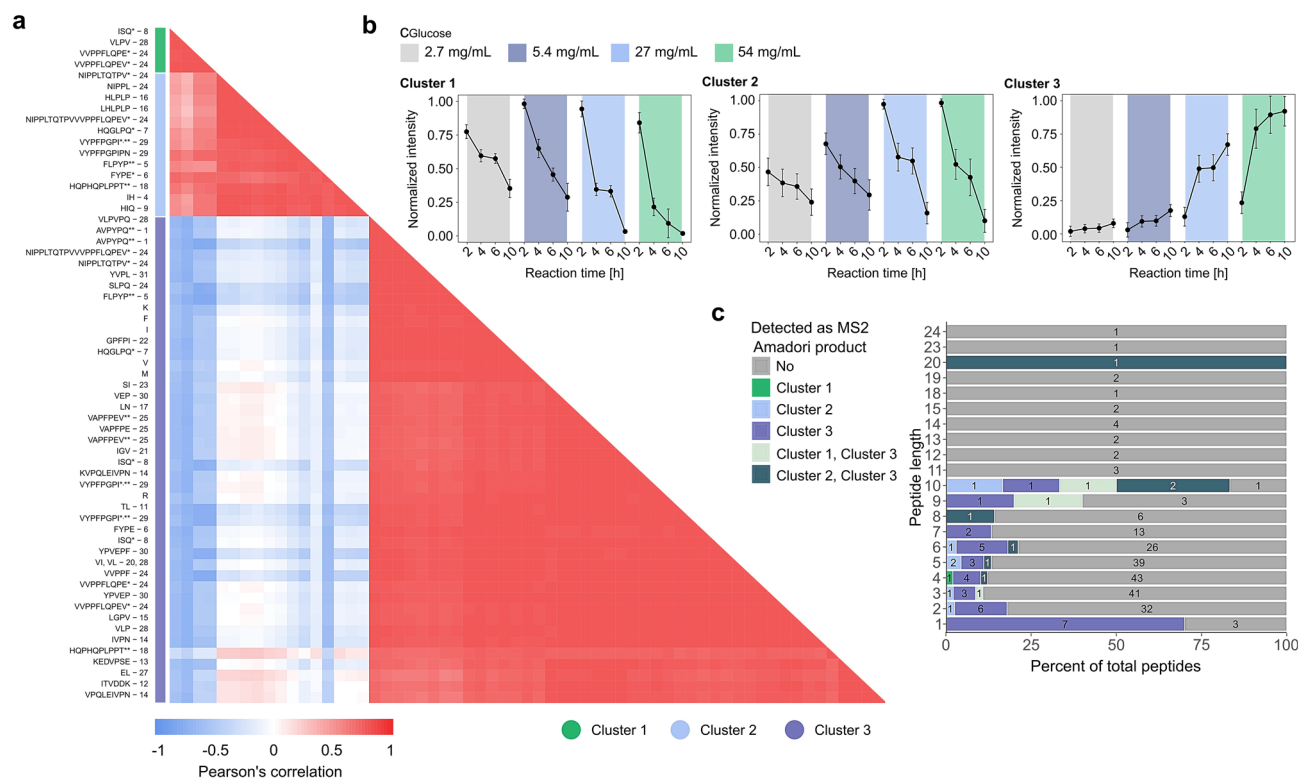


**Figure 1.** Analysis of tryptone glucose model systems by UHPLC-QTOF-MS. **(a)** Experimental design—high-resolution mass spectrometry was used to analyze tryptone glucose model systems and for the identification of site-specific non-enzymatic glycation. **(a)** The bar plot illustrates the number 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.

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<sup>39,40</sup>. Information on net chemical transformations and precursor reactivity in such systems can be gained by non-targeted experiments and generation of mass difference networks<sup>41,42</sup>. Collision-induced dissociation (CID) after electrospray ionization (ESI) MS has been successfully applied for peptide derived AP analysis<sup>43–45</sup>. 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<sup>33,34,46–48</sup>. 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.

## Results

**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 (Fig. 1a). A vast number of non-enzymatic cleavage sites generates many diverse peptides (Supplementary Fig. 1) 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



**Figure 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)**.

tryptone peptides comprise a much greater diversity than enzymatic digests can cause, which becomes apparent by comparison with an *in silico* tryptic casein digest (Fig. 1b and Supplementary Fig. 2a). Unlike tryptone, there is a bias toward lysine- (Lys) and arginine- (Arg) containing peptides for tryptic casein digestion (Supplementary Fig. 2b) and poor enzymatic cleavage of certain protein regions (Supplementary Fig. 3) 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<sup>49</sup>. With the applied method, we could nearly completely cover the casein protein sequences (Fig. 1c).

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<sup>50</sup>, which is formed via condensation between the amino compound and the reducing sugar, and subsequent rearrangement<sup>11</sup>. 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<sub>6</sub>H<sub>12</sub>O<sub>6</sub>-H<sub>2</sub>O). Tandem MS was applied to obtain structural information, which allowed confirmation of the AP chemical structure of 47 amino compounds (Supplementary Table 1). Significant correlations (p-value ≤ 0.05) were observed across normalized AP intensity profiles, and APs clustered by the influence of sugar concentration but also reaction time (Fig. 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 Fig. 2b, the representative normalized mean intensity profiles for the three clusters are shown, and Supplementary Fig. 4 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 (Fig. 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 2c compares AP formation between different peptide lengths, demonstrating that nearly all amino acids and dipeptides were found in Cluster 3, whereas larger peptides

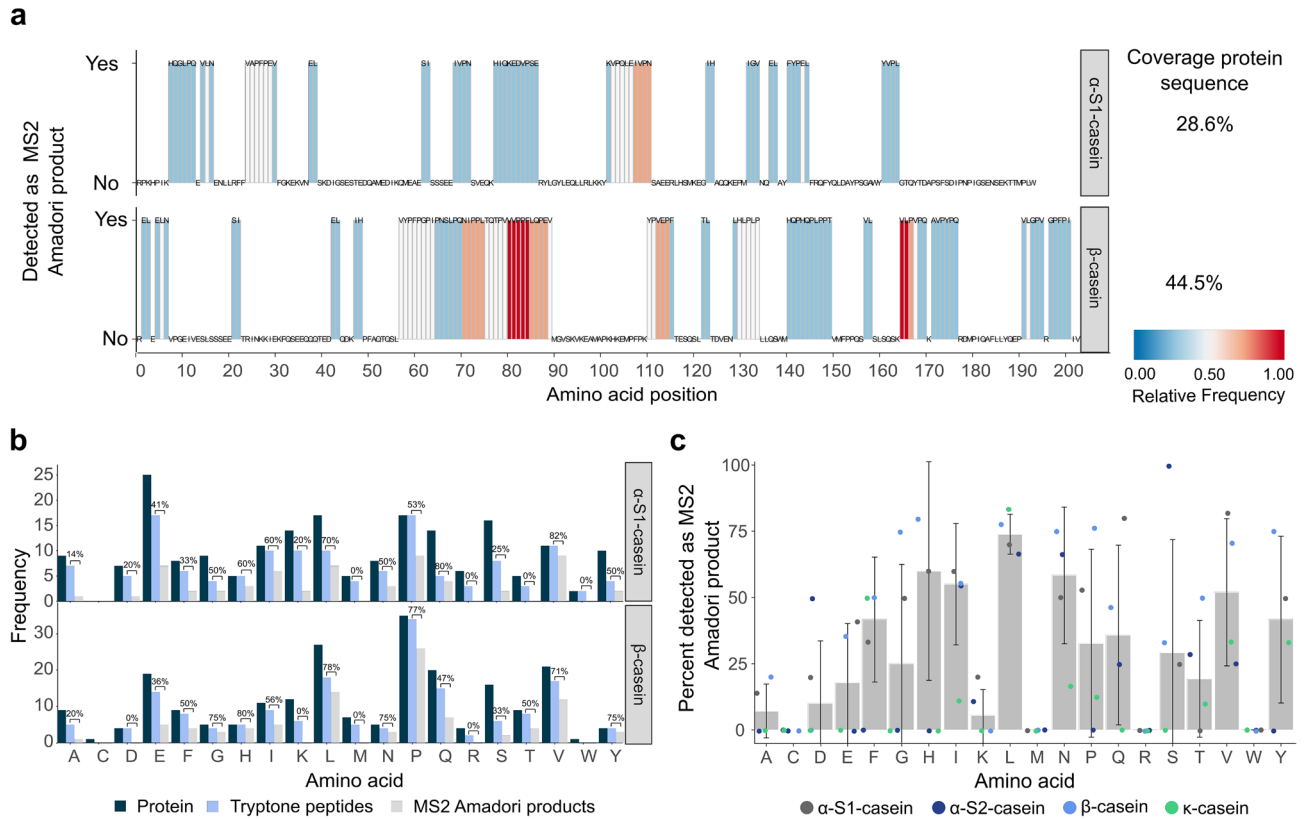
did not show uniform normalized AP intensity profiles. Moreover, the percentage observed as an AP for each peptide length is displayed, but no general correlation between peptide reactivity and sequence length could be observed. Note, this calculation is based on the total number of peptides per length, so one length may appear to be glycosylated 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<sup>33,43</sup>, 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.

All APs significantly increased after 2 h (t-test,  $p$ -value  $\leq 0.05$ ); however, different glucose levels were required (Supplementary Table 1). 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).

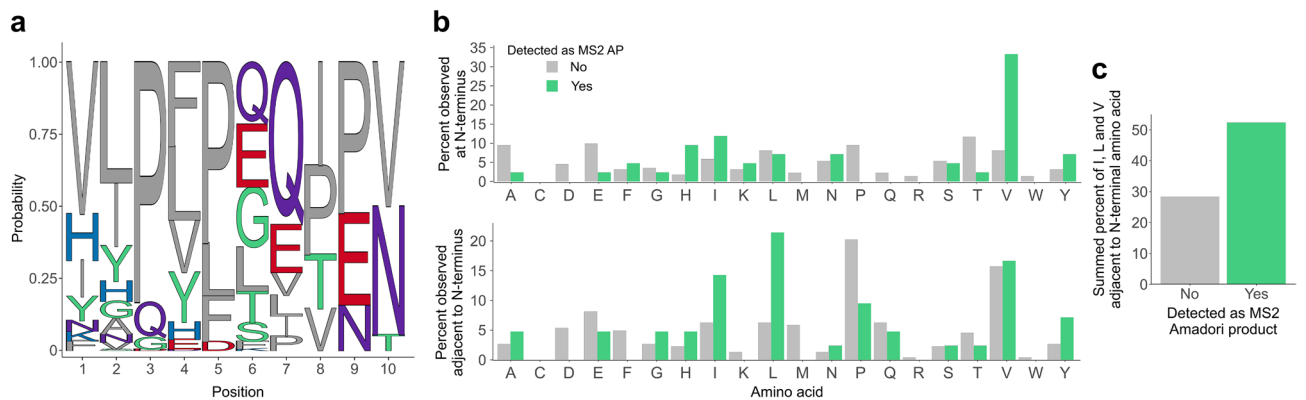
**Shedding light onto the role of peptide composition in glycation.** AP analysis uniquely facilitates characterization of site-specific glycation, and our dataset provides insight into the highly complex and yet largely unknown reaction behavior of peptides in glycation. Previous studies explored peptide derived MRPs in particular foods providing limited information from a global prospect<sup>35–38</sup>. 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<sup>33,34</sup>. Here, we aspired to approach these research questions from a general level using a large reservoir of peptides and APs. Mapping glycosylated peptides onto casein proteins revealed that AP formation was observed for only 14% of  $\alpha$ -S2- and 9% of  $\kappa$ -casein peptides (Supplementary Fig. 5), but APs related to  $\alpha$ -S1-casein (29%) and  $\beta$ -casein (45%) were detected to a larger extent (Fig. 3a). Analyzing the amino acid sequence of  $\alpha$ -S2- and  $\kappa$ -casein peptide APs, we identified protein-specific peptides such as FLPYP (F<sub>55</sub>-P<sub>59</sub> of  $\kappa$ -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 glycosylated peptides. AP formation of peptides sharing certain amino acid sequences appeared to be favored, e.g. including peptides that originated from N<sub>73</sub>-V<sub>92</sub> and V<sub>170</sub>-V<sub>173</sub> of  $\beta$ -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 3b shows amino acids, e.g. glutamic acid (Glu) and leucine (Leu), that appeared equally in peptide APs from  $\alpha$ -S1- (top) and  $\beta$ -casein (bottom). Other amino acids, such as proline (Pro) and histidine (His) showed considerable variations in their contribution to AP formation (Supplementary Fig. 6) 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<sup>46</sup>. 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<sup>26,33</sup>, 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 (Fig. 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.

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<sup>33</sup>, and on the suggestion of Lhiang Zhili and co-workers that Leu and isoleucine (Ile) promote AP formation<sup>26</sup>. 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-glycosylated peptides (Fig. 4a and Supplementary Fig. 7a). Here, amino acids enriched at certain positions in AP forming peptides are illustrated (relative abundance<sub>glycosylated</sub> - relative abundance<sub>non-glycosylated</sub> > 0). This analysis indicated preference for valine (Val), Ile and Leu at the first two positions of the amino acid sequence for glycosylated peptides (Fig. 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 (Fig. 4b, top). An illustration of the absolute amino acid abundance can be found in Supplementary Fig. 7b. 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 (Fig. 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 Fig. 4c. This result complies with previous findings that hypothesized that Ile and Leu promote N-terminal glycation based on pronounced hydrophobicity<sup>51</sup> (Ile 1.80; Leu 1.70) and polarizability<sup>52,53</sup> (Ile 91.21; Leu 91.60). Val has comparable physicochemical properties (hydrophobicity 1.22, polarizability 76.09)



**Figure 3.** Deconvolution of glycosylated 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  $\alpha$ -S1- and 44.5% of  $\beta$ -casein were detected as the corresponding MS2 AP, and the majority contains redundant sequence motifs (i.e. P-E-V, see Fig. 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.



**Figure 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 glycosylated peptides are illustrated (relative abundance<sub>glycosylated</sub> - relative abundance<sub>non-glycosylated</sub> > 0). **(b)** Comparison of amino acids at the N-terminal position and adjacent to the peptide N-terminus. Bars show the percentage of (glycosylated) 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.



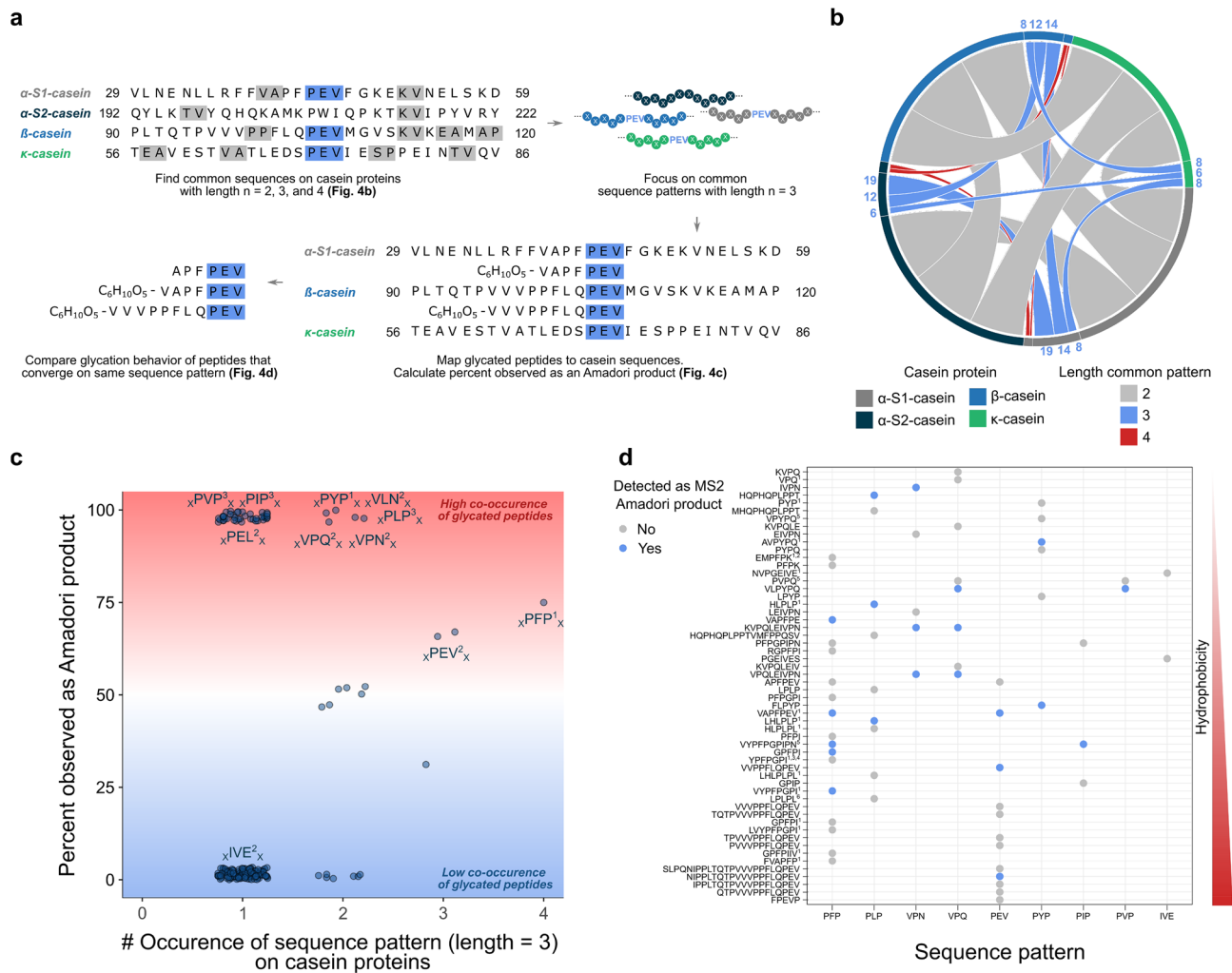
to Ile and Leu and was previously shown to exert a similar effect on the reactivity of the peptide N-terminus<sup>26</sup>. Furthermore, the percentage of aspartic acid (Asp), methionine (Met), phenylalanine (Phe) and Pro next to the N-terminus was considerably decreased in glycated peptides (Fig. 4b, top). We also observed that glycation of peptides was disfavored for Pro at the first two sequence positions (Fig. 4b and Supplementary Fig. 7a). A recent study on Pro containing dipeptides (Gly-Pro, Pro-Gly) suggested that its secondary amine may hinder Schiff base formation<sup>54</sup>. Nevertheless, we found that proline was frequently observed at the third and fifth sequence position of glycated peptides (Fig. 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 (Supplementary Fig. 7c and d).

**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<sup>26,33</sup>, 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<sup>55</sup>. 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 (Fig. 5a, top). We mapped (non-) glycated peptides onto proteins (Fig. 5a, bottom right part) and across each other (Fig. 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 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  $\kappa$ -casein. To explore sequence patterns of maximum length, tri-patterns were chosen for further investigation. Figure 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<sub>44</sub>-V<sub>46</sub> of  $\alpha$ -S1-casein and P<sub>105</sub>-V<sub>107</sub> of  $\beta$ -casein, Fig. 3a; P<sub>171</sub>-V<sub>173</sub> of  $\kappa$ -casein, Supplementary Fig. 5), two appear as subsequences in peptide APs, as well as the alternated V-E-P (V<sub>131</sub>-P<sub>133</sub> of  $\beta$ -casein, Fig. 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; Fig. 5c). These subsequences all share amino acids with a low dissimilarity score ( $D$ ), which is based on 134 categories of activity and structure<sup>56</sup>, and are partially rearranged (Supplementary Table 2). For example, P-E-V and P-E-L only differ by a single amino acid with strong physicochemical similarity ( $D(\text{Val}, \text{Leu}) = 9$ ), whereas in case of P-E-V and V-P-Q ( $D(\text{Glu}, \text{Gln}) = 14$ ) also the sequence order was changed. By comparison, peptides that contain I-V-E, which shows more pronounced sequence variations ( $D(\text{Pro}, \text{Ile}) = 24$  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<sup>33</sup>, 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<sup>56</sup>. All these patterns showed a high co-occurrence on peptide APs relative to their total abundance (displayed as percent in Fig. 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<sub>198</sub>-V<sub>200</sub> of  $\alpha$ -S1-casein) did not contribute to an AP (Fig. 3a) and shows pronounced differences in its amino acid characteristics.

A peptide-sequence pattern plot in Fig. 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. VVVPFLQPEV; YPFPGPI vs. VYPFPGPI) or not cause (VAPFPE vs. VAPFPEV; YPFPGPI vs. VYPFPGPI) 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.

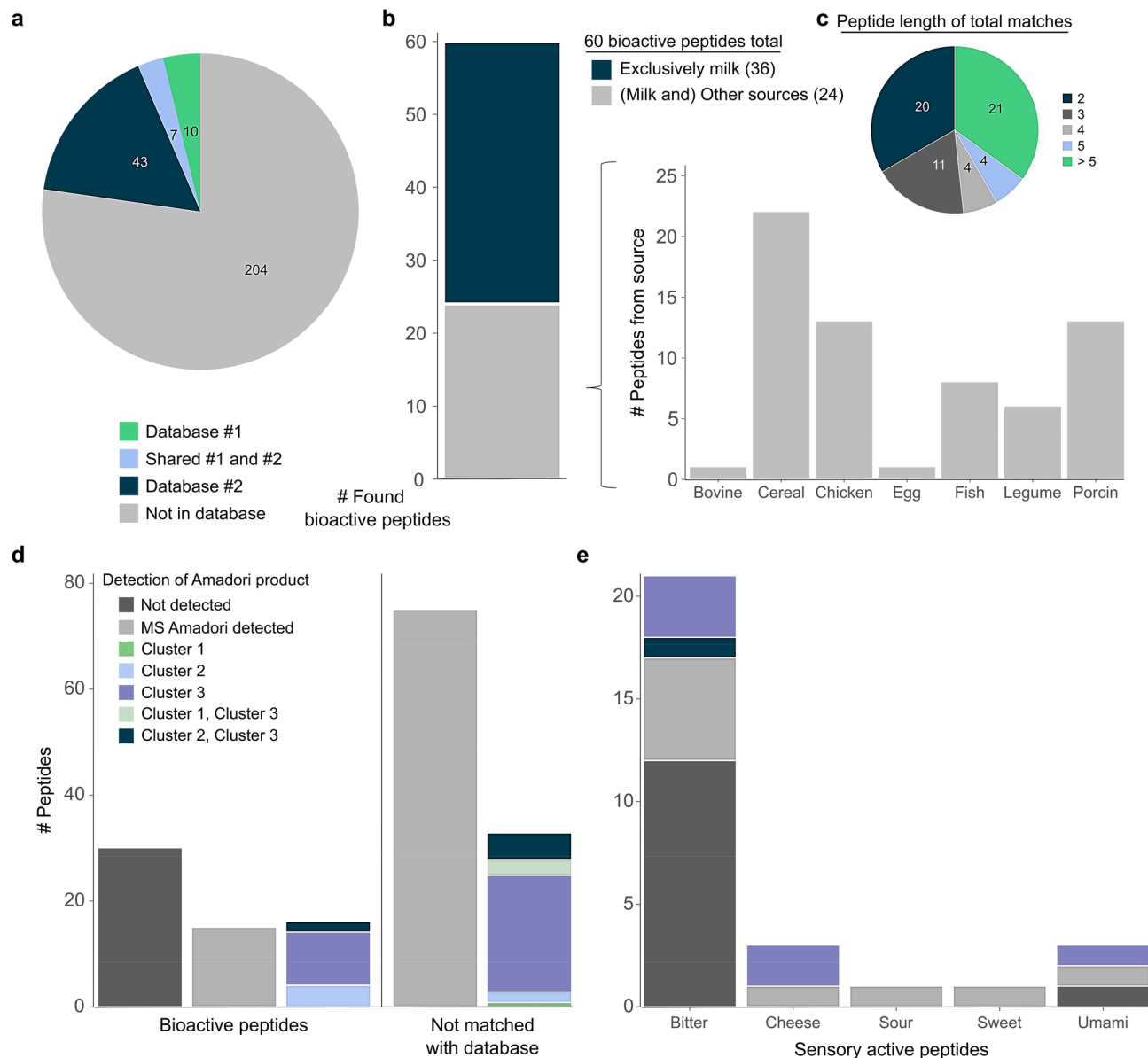
### System-wide analysis of bioactive and sensory-active peptide glycation enabled by in silico sequence mapping.

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 (Supplementary Table 3) with reported bioactivities (Fig. 6a), which were included in databases compiled from literature<sup>57,58</sup>. While 36 peptides were exclusively found in milk, 24 peptides appeared in a variety of other food sources as well (Fig. 6b). We also found that these peptides of diverse chain lengths (Fig. 6c) cover various bioactivity categories (Supplementary Fig. 8), suggesting that tryptone models may facilitate inter-disciplinary



**Figure 5.** Mapping sequence patterns to protein and peptide domains exemplifies their role in peptide glycosylation. **(a)** Schematic illustration of sequence crosstalk decoding: Effect of sequence similarity on protein and peptide glycosylation based on presence of short-chain amino acid patterns. **(b)** A chord diagram representation of overlapping 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 Supplementary Table 2. **(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).

investigation of peptide glycosylation. In our study, for 62% of the bioactive peptides the corresponding AP was detected (33% confirmed by MS/MS tandem experiment; Fig. 6d). Hence, approximately 34% of the 47 peptides, for which the AP was identified via MS2 fragmentation, were previously reported to be bioactive. Furthermore, we identified 25 peptides with sensory attributes<sup>59</sup>. Figure 6e illustrates the prevalence of AP detection for sensory-active peptides and the level of AP identification (MS and MS2). Approximately 24% of all sensory-active peptides and 19% of the bitter peptides were observed as the corresponding AP on MS2 level. As noted by Shiyuan Dong and co-workers<sup>32</sup>, bitterness of MRPs was decreased compared to original casein peptides, and further reduced with heating time and glucose concentration. Xiaohong Lan et al. previously published that bitter soybean peptides below 1000 Da decreased 28.49% after reaction with xylose at 120 °C<sup>60</sup>. 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 (Supplementary Fig. 9) and would not allow comprehensive investigation of their reaction behavior (Supplementary Fig. 10 and Supplementary Fig. 11). For example, using a tryptic casein digest, the analysis of the opioid peptides, e.g. YPF-PGPI and YPVEPE, for which N-glycosylation is known to have major consequences on the bioactivity of the parent

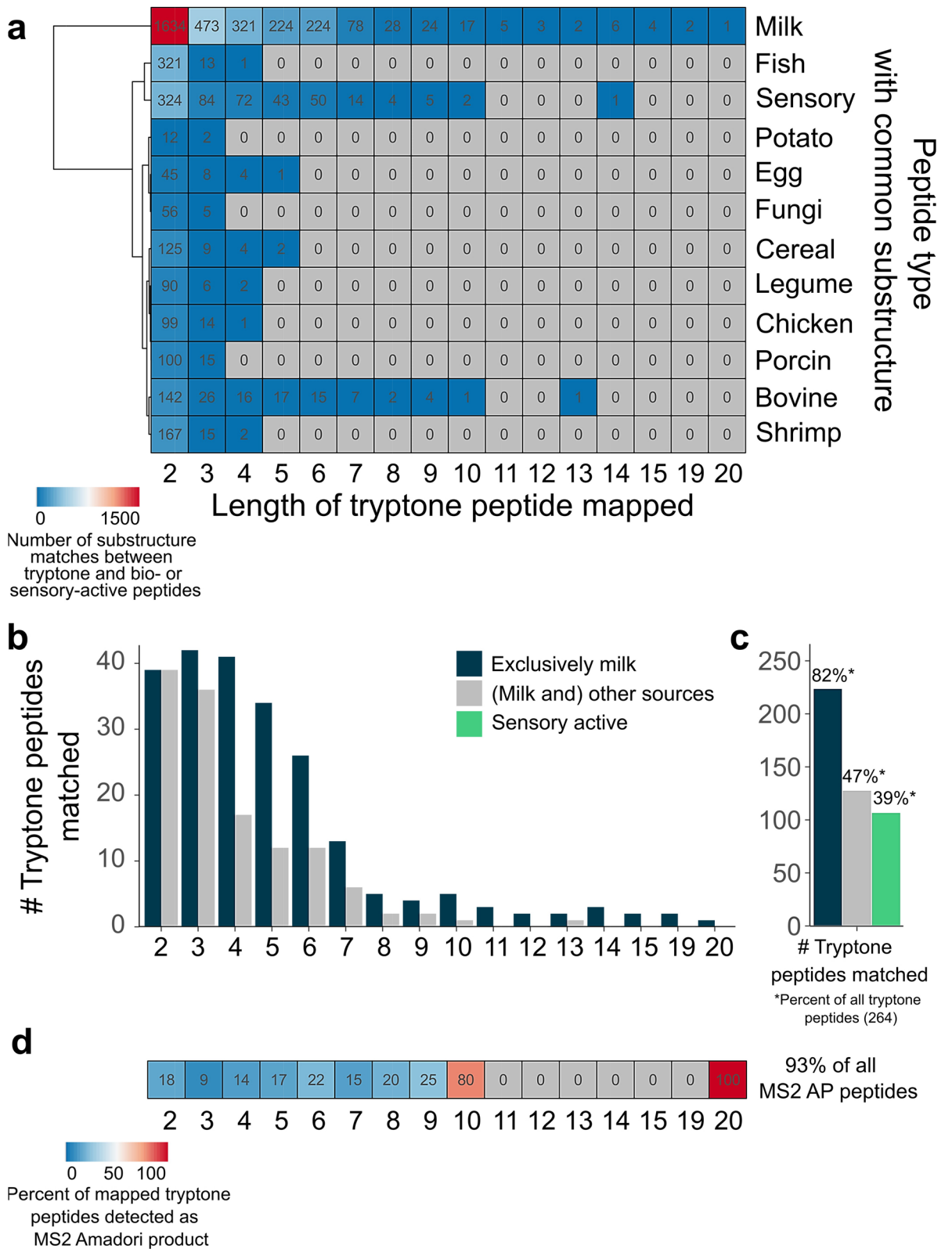


**Figure 6.** Database search uncovers overlap with miscellaneous peptide sources. **(a)** Identification of 60 peptides reported as bioactive peptides in databases for milk and other foods. **(b)** Classification of the assigned bioactive peptides according to different sources. Of these bioactive peptides, 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  $\pm 10$  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 Fig. 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.

peptide<sup>61,62</sup>, 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.

**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<sup>57–59</sup>. First, we examined the num-





**Figure 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.

ber of matches observed. In total, tryptone peptides were successfully mapped to 1172 unique bioactive peptide species (Supplementary Fig. 12). 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 (Fig. 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 (Supplementary Fig. 13). Importantly, not only small sequence commonalities were found as indicated by the length of the tryptone peptides mapped (Fig. 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 (Fig. 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. Supplementary Fig. 14 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 (Fig. 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 (Fig. 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 (Fig. 7c). Figure 7d depicts the percentage of matched tryptone peptides, for which the corresponding AP was identified. Up to 22% of the shorter substrings 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 (Fig. 7d).

## 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 a *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, Fig. 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-NH<sub>2</sub> toward oxidative degradation in aqueous model systems<sup>63</sup>. Loss of the Amadori compound of the endogenous opioid pentapeptide leucine-enkephaline (Tyr-Gly-Gly-Phe-Leu) was also noticed by Jakas and Horvat<sup>64</sup>, 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<sup>65,66</sup>. 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<sup>67</sup>. 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<sup>26,33,34</sup>. 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<sup>26,33,34</sup>. 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<sup>55</sup>. 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<sup>33</sup>, 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<sup>54</sup>, 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<sup>33</sup>. 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<sup>29,30,68</sup>. Conversely, if increased antioxidant potency<sup>26,69</sup> must be achieved, 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<sup>27,28,30</sup>, 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<sup>69</sup>, 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 were increased<sup>26</sup>. Similar findings were reported for peptides from other sources<sup>60,70,71</sup>. 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<sup>71,72</sup>, we reasoned that selection of peptides susceptible to early glycation may be promising. As bitter peptides cannot be employed above their bitter flavor threshold<sup>72</sup>, increased bioactivity accompanied concurrently by decreased bitterness may be desirable for the production of functional foods to improve health and enhance customer acceptance<sup>23</sup>. 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.

## Methods

**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 M $\Omega$ , 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<sup>42</sup>. Control samples were heated for 10 h, analogously. Sample preparation was performed in triplicate (n = 3). Model systems were stored at -20 °C.

**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 (QqTOF) mass spectrometer (MS) (maXis, Bruker Daltonics, Bremen, Germany). For reversed phase (RP) chromatography an ACQUITY UPLC BEH C18 column (100  $\times$  2.1 mm, 1.7  $\mu$ m, 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  $\mu$ L). 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  $\alpha$ -S1-,  $\alpha$ -S2-,  $\beta$ -, and  $\kappa$ -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  $\pm$  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<sup>43,73</sup> and manually validated. Monoisotopic mass tolerance was set to  $\pm$  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 (MS2 level) could be verified, were computed. The minimum consolidated score per peptide length was chosen as a threshold for peptide identification.

**Statistical analysis.** Pearson correlation coefficients were calculated in R software between intensity values for putatively assigned MS2 Amadori products (n  $\geq$  3,  $p \leq$  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 ( $p \leq$  0.05) and relevant reaction conditions are reported in Supplementary Table 1. Multiple testing correction was performed using the Benjamini-Hochberg procedure.

**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 `grep()` and `match()` base functions were applied. Based on derived sequence commonalities, we assigned all peptides to sequence groups (Supplementary Table 1). Amino acids were not assigned to sequence groups. Sequence groups, for which no Amadori product was detected, were excluded in Supplementary Table 1.

**Database search.** Bioactive peptide database search was carried out using the Milk Bioactive Peptide Database (March 13, 2020)<sup>57</sup> and the BioPepDB database (March 13, 2020)<sup>58</sup>. Sensory peptide database search was



performed using the BIOPEP database (March 13, 2020)<sup>59</sup>. 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.

## Data availability

The authors declare that the main data supporting the findings of this study are available within the article and its Supplementary Information files. Extra data are available from the corresponding authors upon request.

Received: 18 February 2021; Accepted: 7 June 2021

Published online: 24 June 2021

## References

- Priego-Capote, F. *et al.* Glycation isotopic labeling with <sup>13</sup>C-reducing sugars for quantitative analysis of glycated proteins in human plasma. *Mol. Cell. Proteom. (MCP)* **9**, 579–592. <https://doi.org/10.1074/mcp.M900439-MCP200> (2010).
- Soboleva, A., Schmidt, R., Vikhnina, M., Grishina, T. & Frolov, A. Maillard proteomics: Opening new pages. *Int. J. Mol. Sci.* **18**, 2677. <https://doi.org/10.3390/ijms18122677> (2017).
- Hemmler, D. *et al.* Evolution of complex Maillard chemical reactions. Resolved in time. *Sci. Rep.* **7**, 3227. <https://doi.org/10.1038/s41598-017-03691-z> (2017).
- Hemmler, D. & Schmitt-Kopplin, P. 2.11 - Analytical challenges and strategies to decipher the maillard reaction network. In *Comprehensive Foodomics* (eds Cifuentes, A) 155–173 (Elsevier, 2021). <https://doi.org/10.1016/B978-0-08-100596-5.22788-1>
- 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 Chem.* **107**, 1244–1252. <https://doi.org/10.1016/j.foodchem.2007.09.062> (2008).
- 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 Chem.* **267**, 271–276. <https://doi.org/10.1016/j.foodchem.2017.07.141> (2018).
- Lund, M. N. & Ray, C. A. Control of Maillard reactions in foods: Strategies and chemical mechanisms. *J. Agric. Food Chem.* **65**, 4537–4552. <https://doi.org/10.1021/acs.jafc.7b00882> (2017).
- Laroque, D. *et al.* Kinetic study on the Maillard reaction. Consideration of sugar reactivity. *Food Chem.* **111**, 1032–1042. <https://doi.org/10.1016/j.foodchem.2008.05.033> (2008).
- Tessier, F. J. & Birlouez-Aragon, I. Health effects of dietary Maillard reaction products: The results of ICARE and other studies. *Amino Acids* **42**, 1119–1131. <https://doi.org/10.1007/s00726-010-0776-z> (2012).
- Somoza, V. Five years of research on health risks and benefits of Maillard reaction products: An update. *Mol. Nutr. Food Res.* **49**, 663–672. <https://doi.org/10.1002/mnfr.200500034> (2005).
- Hellwig, M. & Henle, T. Baking, ageing, diabetes: A short history of the Maillard reaction. *Angew. Chem. Int. Ed.* **53**, 10316–10329. <https://doi.org/10.1002/anie.201308808> (2014).
- Brownlee, M., Cerami, A. & Vlassara, H. Advanced glycosylation end products in tissue and the biochemical basis of diabetic complications. *N. Engl. J. Med.* **318**, 1315–1321. <https://doi.org/10.1056/NEJM198805193182007> (1988).
- Schmidt, A. M. *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. *Proc. Natl. Acad. Sci. U S A* **91**, 8807–8811. <https://doi.org/10.1073/pnas.91.19.8807> (1994).
- Miyata, T. *et al.* 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. *J. Clin. Invest.* **98**, 1088–1094. <https://doi.org/10.1172/JCI118889> (1996).
- Schmidt, A. M. *et al.* 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. *J. Clin. Invest.* **96**, 1395–1403. <https://doi.org/10.1172/JCI118175> (1995).
- Wautier, J. L. *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. *Proc. Natl. Acad. Sci. U S A* **91**, 7742–7746. <https://doi.org/10.1073/pnas.91.16.7742> (1994).
- Münch, G. *et al.* Amino acid specificity of glycation and protein-AGE crosslinking reactivities determined with a dipeptide SPOT library. *Nat. Biotechnol.* **17**, 1006–1010. <https://doi.org/10.1038/13704> (1999).
- Chakrabarti, S., Jahandideh, F. & Wu, J. Food-derived bioactive peptides on inflammation and oxidative stress. *Biomed. Res. Int.* **2014**, 608979. <https://doi.org/10.1155/2014/608979> (2014).
- Rutherford-Markwick, K. J. Food proteins as a source of bioactive peptides with diverse functions. *Br. J. Nutr.* **108**, S149–S157. <https://doi.org/10.1017/S000711451200253X> (2012).
- Wang, H.-Y., Qian, H. & Yao, W.-R. Melanoidins produced by the Maillard reaction: Structure and biological activity. *Food Chem.* **128**, 573–584. <https://doi.org/10.1016/j.foodchem.2011.03.075> (2011).
- Borrelli, R. C. *et al.* Characterization of coloured compounds obtained by enzymatic extraction of bakery products. *Food Chem. Toxicol.* **41**, 1367–1374. [https://doi.org/10.1016/S0278-6915\(03\)00140-6](https://doi.org/10.1016/S0278-6915(03)00140-6) (2003).
- 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.* **50**, 6997–7006. <https://doi.org/10.1021/jf020618n> (2002).
- Li-Chan, E. C. Y. Bioactive peptides and protein hydrolysates: Research trends and challenges for application as nutraceuticals and functional food ingredients. *Curr. Opin. Food Sci.* **1**, 28–37. <https://doi.org/10.1016/j.cofs.2014.09.005> (2015).
- 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 Chem.* **290**, 40–46. <https://doi.org/10.1016/j.foodchem.2019.03.065> (2019).
- Zhang, Y., Wang, Y., Jiang, F. & Jin, H. Sensory characteristics of Maillard reaction products from chicken protein hydrolysates with different degrees of hydrolysis. *CyTA J. Food* **17**, 221–227. <https://doi.org/10.1080/19476337.2019.1575907> (2019).
- Liang, Z. *et al.* Formation of peptide bound pyrroline in the Maillard model systems with different lys-containing dipeptides and tripeptides. *Molecules* **21**, 463. <https://doi.org/10.3390/molecules21040463> (2016).
- Chevalier, F., Chobert, J.-M., Genot, C. & Haertlé, T. Scavenging of free radicals, antimicrobial, and cytotoxic activities of the Maillard reaction products of beta-lactoglobulin glycosylated with several sugars. *J. Agric. Food Chem.* **49**, 5031–5038. <https://doi.org/10.1021/jf010549x> (2001).
- Chuyen, N. V., Ijichi, K., Umetsu, H. & Moteki, K. In *Process-Induced Chemical Changes in Food* (eds Shahidi, F. *et al.*) 201–212 (Springer, 1998).
- 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 Res. Int.* **50**, 266–274. <https://doi.org/10.1016/j.foodres.2012.09.028> (2013).



30. 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.* **141**, 3837–3845. <https://doi.org/10.1016/j.foodchem.2013.06.041> (2013).
31. Jing, H. & Kitts, D. D. Chemical and biochemical properties of casein–sugar Maillard reaction products. *Food Chem. Toxicol.* **40**, 1007–1015. [https://doi.org/10.1016/S0278-6915\(02\)00070-4](https://doi.org/10.1016/S0278-6915(02)00070-4) (2002).
32. 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. *J. Agric. Food Chem.* **59**, 13311–13317. <https://doi.org/10.1021/jf203778z> (2011).
33. de Kok, P. M. T. & Rosing, E. A. E. In *Thermally Generated Flavors*, Vol. 543. ACS Symposium Series, Chap. 13, 158–179 (American Chemical Society, 1993).
34. Kim, J.-S. & Lee, Y.-S. Study of Maillard reaction products derived from aqueous model systems with different peptide chain lengths. *Food Chem.* **116**, 846–853. <https://doi.org/10.1016/j.foodchem.2009.03.033> (2009).
35. Meltretter, J. et al. Untargeted proteomics-based profiling for the identification of novel processing-induced protein modifications in milk. *J. Proteome Res.* **19**, 805–818. <https://doi.org/10.1021/acs.jproteome.9b00630> (2020).
36. Meltretter, J., Wüst, J. & Pischetsrieder, M. Modified peptides as indicators for thermal and nonthermal reactions in processed milk. *J. Agric. Food Chem.* **62**, 10903–10915. <https://doi.org/10.1021/jf503664y> (2014).
37. Milkovska-Stamenova, S. & Hoffmann, R. Diversity of advanced glycation end products in the bovine milk proteome. *Amino Acids* **51**, 891–901. <https://doi.org/10.1007/s00726-019-02707-4> (2019).
38. Poojary, M. M. et al. 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. *J. Chromatogr. A* **1615**, 460767. <https://doi.org/10.1016/j.chroma.2019.460767> (2020).
39. Fay, L. B. & Brevard, H. Contribution of mass spectrometry to the study of the Maillard reaction in food. *Mass Spectrom. Rev.* **24**, 487–507. <https://doi.org/10.1002/mas.20028> (2005).
40. Sibbersen, C. et al. Profiling of methylglyoxal blood metabolism and advanced glycation end-product proteome using a chemical probe. *ACS Chem. Biol.* **13**, 3294–3305. <https://doi.org/10.1021/acschembio.8b00732> (2018).
41. Hemmler, D. et al. Simulated sunlight selectively modifies Maillard reaction products in a wide array of chemical reactions. *Chem. Eur. J.* **25**, 13208–13217. <https://doi.org/10.1002/chem.201902804> (2019).
42. Hemmler, D. et al. Insights into the chemistry of non-enzymatic browning reactions in different ribose-amino acid model systems. *Sci. Rep.* **8**, 16879. <https://doi.org/10.1038/s41598-018-34335-5> (2018).
43. Frolov, A., Hoffmann, P. & Hoffmann, R. Fragmentation behavior of glycated peptides derived from D-glucose, D-fructose and D-ribose in tandem mass spectrometry. *J. Mass Spectrom.* **41**, 1459–1469. <https://doi.org/10.1002/jms.1117> (2006).
44. Sillner, N. et al. Milk-derived Amadori products in feces of formula-fed infants. *J. Agric. Food Chem.* **67**, 8061–8069. <https://doi.org/10.1021/acs.jafc.9b01889> (2019).
45. Stefanowicz, P., Kijewska, M. & Szewczuk, Z. Sequencing of peptide-derived Amadori products by the electron capture dissociation method. *J. Mass Spectrom.* **44**, 1047–1052. <https://doi.org/10.1002/jms.1580> (2009).
46. Iberg, N. & Flückiger, R. Nonenzymatic glycosylation of albumin in vivo. Identification of multiple glycosylated sites. *J. Biol. Chem.* **261**, 13542–13545 (1986).
47. Johansen, M. B., Kierner, L. & Brunak, S. Analysis and prediction of mammalian protein glycation. *Glycobiology* **16**, 844–853. <https://doi.org/10.1093/glycob/cwl009> (2006).
48. Venkatraman, J., Aggarwal, K. & Balaram, P. Helical peptide models for protein glycation: Proximity effects in catalysis of the Amadori rearrangement. *Chem. Biol.* **8**, 611–625. [https://doi.org/10.1016/S1074-5521\(01\)00036-9](https://doi.org/10.1016/S1074-5521(01)00036-9) (2001).
49. Dau, T., Bartolomucci, G. & Rappsilber, J. Proteomics using protease alternatives to trypsin benefits from sequential digestion with trypsin. *Anal. Chem.* **92**, 9523–9527. <https://doi.org/10.1021/acs.analchem.0c00478> (2020).
50. Hodge, J. E. Dehydrated foods, chemistry of browning reactions in model systems. *J. Agric. Food Chem.* **1**, 928–943. <https://doi.org/10.1021/jf60015a004> (1953).
51. Otto, S. & Engberts, J. B. Hydrophobic interactions and chemical reactivity. *Org. Biomol. Chem.* **1**, 2809–2820. <https://doi.org/10.1039/b305672d> (2003).
52. Bartoli, G. & Todesco, P. E. Nucleophilic substitution. Linear free energy relations between reactivity and physical properties of leaving groups and substrates. *Acc. Chem. Res.* **10**, 125–132. <https://doi.org/10.1021/ar50112a004> (1977).
53. Edwards, J. O. & Pearson, R. G. The factors determining nucleophilic reactivities. *J. Am. Chem. Soc.* **84**, 16–24. <https://doi.org/10.1021/ja00860a005> (1962).
54. Izzo, H. V. & Ho, C.-T. Peptide-specific Maillard reaction products: A new pathway for flavor chemistry. *Trends Food Sci. Technol.* **3**, 253–257. [https://doi.org/10.1016/S0924-2244\(10\)80004-4](https://doi.org/10.1016/S0924-2244(10)80004-4) (1992).
55. Gavel, Y. & von Heijne, G. Sequence differences between glycosylated and non-glycosylated Asn-X-Thr/Ser acceptor sites: Implications for protein engineering. *Protein Eng.* **3**, 433–442. <https://doi.org/10.1093/protein/3.5.433> (1990).
56. Sneath, P. H. A. Relations between chemical structure and biological activity in peptides. *J. Theor. Biol.* **12**, 157–195. [https://doi.org/10.1016/0022-5193\(66\)90112-3](https://doi.org/10.1016/0022-5193(66)90112-3) (1966).
57. 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 Chem.* **232**, 673–682. <https://doi.org/10.1016/j.foodchem.2017.04.056> (2017).
58. Li, Q. et al. BioPepDB: An integrated data platform for food-derived bioactive peptides. *Int. J. Food Sci. Nutr.* **69**, 963–968. <https://doi.org/10.1080/09637486.2018.1446916> (2018).
59. Minkiewicz, P., Iwaniak, A. & Darewicz, M. BIOPEP-UWM database of bioactive peptides: Current opportunities. *Int. J. Mol. Sci.* **10**, 5978. <https://doi.org/10.3390/ijms20235978> (2019).
60. Lan, X. et al. 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 Chem.* **120**, 967–972. <https://doi.org/10.1016/j.foodchem.2009.11.033> (2010).
61. Zigrović, I., Kidric, J. & Horvat, S. Influence of glycation on cis/trans isomerization and tautomerization in novel morphiceptin-related Amadori compounds. *Glycoconj. J.* **15**, 563–570. <https://doi.org/10.1023/a:1006907724915> (1998).
62. Jakas, A. & Horvat, Š. Synthesis and <sup>13</sup>C NMR investigation of novel Amadori compounds (1-amino-1-deoxy-D-fructose derivatives) related to the opioid peptide, leucine–enkephalin. *J. Chem. Soc. Perkin Trans. 2*, 789–794. <https://doi.org/10.1039/P29960000789> (1996).
63. Greifenhagen, U., Frolov, A. & Hoffmann, R. Oxidative degradation of Ne-fructosylamine-substituted peptides in heated aqueous systems. *Amino Acids* **47**, 1065–1076. <https://doi.org/10.1007/s00726-015-1940-2> (2015).
64. Jakas, A. & Horvat, Š. Study of degradation pathways of Amadori compounds obtained by glycation of opioid pentapeptide and related smaller fragments: Stability, reactions, and spectroscopic properties. *Biopolymers* **69**, 421–431. <https://doi.org/10.1002/bip.10338> (2003).
65. Morales, F. J. & van Boekel, M. A. J. S. A study on advanced Maillard reaction in heated casein/sugar solutions: Colour formation. *Int. Dairy J.* **8**, 907–915. [https://doi.org/10.1016/S0958-6946\(99\)00014-X](https://doi.org/10.1016/S0958-6946(99)00014-X) (1998).
66. Morales, F. J. & van Boekel, M. A. J. S. A study on advanced Maillard reaction in heated casein/sugar solutions: Fluorescence accumulation. *Int. Dairy J.* **7**, 675–683. [https://doi.org/10.1016/S0958-6946\(97\)00071-X](https://doi.org/10.1016/S0958-6946(97)00071-X) (1997).
67. Grillo, M. A. & Colombatto, S. Advanced glycation end-products (AGEs): Involvement in aging and in neurodegenerative diseases.

68. 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 Sci. Technol.* **59**, 35–42. <https://doi.org/10.1016/j.lwt.2014.06.013> (2014).
69. Kim, J.-S. Antioxidant activity of Maillard reaction products derived from aqueous and ethanolic glucose-glycine and its oligomer solutions. *Food Sci. Biotechnol.* **22**, 39–46. <https://doi.org/10.1007/s10068-013-0006-z> (2013).
70. Abdelhedi, O. *et al.* Effect of ultrasound pretreatment and Maillard reaction on structure and antioxidant properties of ultrafiltered smooth-hound viscera proteins-sucrose conjugates. *Food Chem.* **230**, 507–515. <https://doi.org/10.1016/j.foodchem.2017.03.053> (2017).
71. Ogasawara, M., Katsumata, T. & Egi, M. Taste properties of Maillard-reaction products prepared from 1000 to 5000 Da peptide. *Food Chem.* **99**, 600–604. <https://doi.org/10.1016/j.foodchem.2005.08.040> (2006).
72. Spellman, D., O’Cuinn, G. & FitzGerald, R. Bitterness in *Bacillus* proteinase hydrolysates of whey proteins. *Food Chem.* **114**, 440–446. <https://doi.org/10.1016/j.foodchem.2008.09.067> (2009).
73. 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. *Carbohydr. Res.* **491**, 107985. <https://doi.org/10.1016/j.carres.2020.107985> (2020).

## Acknowledgements

The authors would like to thank Martina Daubmeier for her help during sample preparation, which was very much appreciated, and Jenny Uhl for her excellent technical assistance.

## Author contributions

M.B. and D.H. designed the study. MB performed the experiments and wrote the paper. D.H. and A.W. supervised mass spectrometry data analysis. D.H. supervised bioinformatics data analysis. A.W. supervised statistical data analysis. P.S.K. provided resources. M.B., D.H., A.W., M.R., J.M., and P.S.K. discussed and interpreted results. All authors revised and approved the manuscript. Correspondence to D.H. and P.S.K.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-92413-7>.

**Correspondence** and requests for materials should be addressed to M.T.B. or P.S.-K.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021