1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Development of a new method using hydrophilic interaction liquid chromatography (HILIC) - tandem mass spectrometry for the characterization of *O*-sialoglycopeptides from proteolytically digested caseinomacropeptide

Oswaldo Hernandez-Hernandez¹, Rosa Lebron-Aguilar², Jesus Eduardo Quintanilla-Lopez¹, Maria Luz Sanz^{1*}, and Francisco Javier Moreno³

¹Instituto de Química Orgánica General (CSIC), C/ Juan de la Cierva, 3 28006 Madrid (Spain)

²Instituto de Química-Física “Rocasolano” (CSIC), C/ Serrano, 119 28006 Madrid (Spain)

³Instituto de Fermentaciones Industriales (CSIC)- Instituto de Investigación en Ciencias de la Alimentación (CSIC-UAM), C/ Juan de la Cierva, 3 28006 Madrid (Spain)

* To whom correspondence should be addressed:

E-mail: mlsanz@iqog.csic.es

Tel: +34 91 2587559

Fax: +34 91 5644853

20 **ABSTRACT**

21 This work addresses the optimization of HILIC-ESI-MSⁿ conditions for the
22 comprehensive characterization of *O*-glycopeptides from proteolytically digested
23 caseinomacropeptide (CMP). *O*-Glycopeptides were satisfactorily analysed on a
24 zwitterionic HILIC column based on their glycan structure and amino acid sequence.
25 The contribution of ionic interactions to the retention of charged glycopeptides was
26 found to be substantial. Thus, *O*-glycopeptides carrying neutral glycans were more
27 retained than *O*-sialoglycopeptides due to that negatively charged sialic acid residues
28 were electrostatically repelled by the stationary phase. In addition, glycopeptides
29 differing only in the position of the linkage of the sialic acid moiety could be separated.
30 The same chromatographic behaviour was observed for model systems constituted by a
31 synthetic tetrapeptide covalently conjugated to neutral and sialylated carbohydrates.
32 Subsequent detection of CMP *O*-glycopeptides was carried out on an electrospray ion
33 trap tandem mass spectrometer at both positive and negative ionization modes. MS²
34 fragmentation at positive ionization mode was valid for determining the glycan structure
35 as the resulting main fragments corresponded to Y_n-type ions derived from sequential
36 glycosidic bond fragmentation, whilst the fragmentation of the peptide structure was
37 preferably obtained through the formation of b_n-type ions at the MS³ stage, allowing the
38 complete structure elucidation of the peptidic chain. Overall, the developed method
39 allowed the identification and characterization of 41 *O*-glycopeptides covering all the
40 known glycosylation sites without any previous enrichment step. These results point out
41 that HILIC coupled to multistage MS procedures can be a powerful technique for future
42 glycoproteomic applications.

43 **Keywords:** HILIC, sialic acid, *O*-glycosylation, caseinomacropeptide, ion trap,
44 tandem mass spectrometry, κ-casein, protein glycosylation.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1. INTRODUCTION

Protein glycosylation is one of the most usual post-translational modifications with important functions in all biological systems. Carbohydrates can be attached to proteins in two major ways resulting in the so-called *N*-glycans and *O*-glycans. *N*-glycans consist in a carbohydrate moiety *N*-linked to the amide nitrogen of asparagine side chains within an asparagine-x-serine/threonine motif where x corresponds to any amino acid except for proline, although there is also the option of having cysteine instead of serine/threonine. However, no unique sequence motif for *O*-glycosylation is known [1]. Moreover, the typical microheterogeneity of peptide (genetic variance) and oligosaccharide structures described for *O*-glycoproteins makes their characterization a challenging task and, hence, advanced analytical techniques are required [2].

Caseinomacropetide (CMP) is the soluble C-terminal fragment derived from the action of chymosin or pepsin cleavage on κ -casein, during the primary stage of cheesemaking or during digestion in the stomach. CMP is a heterogeneous group of acidic peptides due to extensive post-translational modifications such as *O*-glycosylation, phosphorylation and genetic variance. CMP is also considered as a multifunctional peptide with many possible biological applications [3, 4, 5]. Structure-activity studies have particularly pointed out the importance of the sugar residues for the different biological functions [6]. Glycosylated forms represent about 60% of the total bovine CMP [7], and five different mucin-type carbohydrate chains, composed of the sialic acid *N*-acetylneuraminic acid (Neu5Ac), galactose (Gal) and *N*-acetylgalactosamine (GalNAc), have been identified in bovine CMP [8]: i) monosaccharide GalNAc-O-R, ii) disaccharide Gal β 1-3GalNAc-O-R, iii) trisaccharide

Neu5Ac α 2-3Gal β 1-3GalNAc-O-R, iv) trisaccharide Gal β 1-3(Neu5Ac α 2-6)GalNAc-O-R, v) tetrasaccharide Neu5Ac α 2-3Gal β 1-3(Neu5Ac α 2-6)GalNAc-O-R.

In the last twenty years, hydrophilic interaction liquid chromatography (HILIC) has proved to be a powerful tool in separating polar compounds such as carbohydrates [9, 10], pharmaceutical compounds [11], metabolites and other small polar molecules [12, 13]. Whilst the main retention mechanism described for HILIC is the partitioning of polar analytes between the bulk eluent and a water-rich layer partially immobilized on the stationary phase [9], functional groups on the stationary phases might also have an important influence on selectivity through secondary interactions such as ionic interactions [12, 14], specific adsorption [15], hydrogen bonding [16], and even hydrophobic interactions [17].

HILIC has been also successfully used in the analysis of *N*-glycopeptides [18-22], but very little information is found about the separation of *O*-glycopeptides by this technique [18, 20]. *O*-Linked glycosylation consists of attaching the glycans to the hydroxyl oxygen of serine or threonine residues in sequence regions of high hydroxyamino acid density imparting, thus, a polar character to the peptidic chain. Therefore, HILIC coupled to mass spectrometry (MS) by an electrospray interface (ESI) could be a useful technique to separate and characterize *O*-glycopeptides.

In this work a HILIC multi-stage mass spectrometric method (HILIC-ESI-MSⁿ) has been developed to characterize *O*-sialoglycopeptides in a tryptic/chymotryptic hydrolyzate of bovine CMP. Likewise, a synthetic tetrapeptide, covalently conjugated to carbohydrates of different molecular mass and monosaccharide composition (i.e., a couple of sialylated trisaccharide isomers, one sialopentasaccharide, and three neutral carbohydrates with different degree of polymerization) were also analysed in order to gain a better understanding on the mechanism of the separation of glycopeptides by

HILIC on a zwitterionic column. Furthermore, in order to confirm the amino acidic sequence of the target CMP glycopeptides, MS³ spectra obtained from the collisionally generated ion corresponding to the unglycosylated peptide were successfully performed.

2. MATERIALS AND METHODS

Materials. All chemicals were purchased from Sigma-Aldrich and were of analytical grade unless otherwise stated. CMP was kindly provided by Davisco Foods International, Inc (Le Sueur, Minnesota, USA).

Proteolysis of CMP. CMP was subjected to a combined trypsin/chymotrypsin proteolysis (overnight at 37 °C and pH 7) at 1 : 0.05 : 0.025 CMP : trypsin : chymotrypsin ratios (w : w : w). Activities of porcine pancreas trypsin (EC 3.4.21.4, type IX-S) and bovine pancreas α -chymotrypsin (EC 3.4.21.1, Type I-S) were 13000 - 20000 U and ≥ 40 U per mg of protein, respectively.

Preparation of model glycosylated tetrapeptides. Aliquots of a solution consisting of 0.25 mg mL⁻¹ acetylS-D-K-P and 0.25 mg mL⁻¹ of the following and individual carbohydrates: i) galactose, ii) lactose, iii) maltopentaose, iv) Neu5Ac α 2-3Gal β 1-4Glc, v) Neu5Ac α 2-6Gal β 1-4Glc, and vi) Neu5Ac α 2-6Gal β 1-4GlcNAc β 1-3Gal β 1-4Glc in 10 mM ammonium acetate pH 6.8, were lyophilized. These were kept under vacuum in a desiccator at 40 °C for 1 day (for galactose) and 3 days (for the rest of carbohydrates) at water activity of 0.44 achieved with a saturated K₂CO₃ solution. The glycation occurred by condensation between the carbonyl group of carbohydrates (glucose moiety) and the ϵ -amino group of the lysine residue, leading to an *N*-glycosylamine that rearranges into the so-called Amadori product. Although MS² fragmentation would be different for glycosylated and *O*-glycosylated peptides, model systems were analysed by HILIC-ESI-MS

as described below and identified on the basis of their expected $[M+H]^+$ and $[M-H]^-$ ions for the neutral and sialylated compounds, respectively.

HILIC-ESI-MSⁿ analysis. All experiments were carried out on a Finnigan Surveyor pump with quaternary gradient system coupled to a Finnigan LCQ Deca ion trap mass spectrometer using an ESI interface. Sample injections (10 μ L) were carried out by a Finnigan Surveyor autosampler. All instruments (Thermo Fisher Scientific, San José, CA, USA), and data acquisition were managed by Xcalibur software (1.2 version; Thermo Fisher Scientific).

Chromatographic conditions. The [proteolytically digested CMP](#) (1.5 μ g) and the model [glycated](#) tetrapeptides (0.05 μ g) were injected, and their separation performed at 25 °C on a sulfoalkylbetaine zwitterionic stationary phase ZIC[®]-HILIC column (150 mm x 2.1 mm, 3.5 μ m particle size, 200 Å pore size, SeQuantTM) at a flow rate of 100 μ L min⁻¹. The separation of the CMP hydrolysate was optimized by using different linear binary gradient programmes of Milli-Q water and acetonitrile with formic acid (final content of 0.05%, v:v) as an organic modifier. Once the gradient was optimized, the total concentration of formic acid was also adjusted (from 0.005% to 1.5%, v:v) in order to obtain the best resolution for the separation of CMP glycopeptides and the model [glycated](#) tetrapeptides.

Mass spectrometric conditions. The mass spectrometer spray voltage was set at 4.5 kV, heated capillary temperature at 200 °C, nitrogen (99.5% purity) was used as sheath (0.6 L min⁻¹) and auxiliary (6 L min⁻¹) gas, and helium (99.999% purity) as the collision gas. Full scan mass spectra were recorded in the positive and negative ion modes between m/z 100 and 2,000. MS² data were acquired in the automatic data-dependent mode with a normalized collision energy of 35%, using a total cycle time of approximately 5 s, and with an ion selection threshold of 5 10^5 counts. MS³ experiments

were also carried out with a normalized collision energy of 35% but selecting the appropriated target ion transitions. Automatic gain control was used to maintain constant ion populations into the ion trap analyzer at $5 \cdot 10^7$ and $2 \cdot 10^7$ for full scan MS and MSⁿ spectra, respectively.

Data treatment. Data processing was managed by Xcalibur software (2.0 version, Thermo Fisher Scientific). The initial assignment of observed ions to the corresponding amino acid sequences was based on the known sequence of CMP by using the protein database Swiss-Prot and TrEMBL and the tools Peptide Mass and FindPept available at www.expasy.org. Parameters for the search were the following: i) monoisotopic peptide masses were indicated as $[M+H]^+$ with cysteines treated with nothing; ii) as enzymes, trypsin/chymotrypsin were chosen; iii) four missed cleavages were allowed; iv) peptides with a mass larger than 350 u were displayed; v) the mass tolerance was kept at 0.5 u.

To confirm the sequence of the glycopeptides, the MS³ spectra were compared with the theoretical fragmentation of the putative peptides, obtained from the MS-Product software program (Protein Prospector, <http://prospector.ucsf.edu/>).

pH measurement and pK_a estimation. The pH determination was carried out using a pH meter (S40 Seven Multi, Mettler-Toledo International Inc, USA) equipped with a combined glass Ag/AgCl electrode (Mettler-Toledo International Inc, USA). The pH values ($^w_p K_a$) were taken in hydro-organic solutions after calibration of the instrument with pH 4.01 and 7.00 aqueous buffer solutions. The notation $^s_p K_a$ means that the pH is directly measured in the solution (s) after the electrode was calibrated in pure water (w).

The estimation of ${}^w\text{p}K_a$ values was done by using the ADME suite 5.0 (Advanced Chemistry Development, Inc., Toronto, ON, Canada; www.acdlabs.com). The ${}^s\text{p}K_a$ values for each compound were calculated at different ratios of acetonitrile:water, according to the following equation [23]:

$${}^s\text{p}K_a = a_s {}^w\text{p}K_a + b_s + \delta \quad (1)$$

where a_s and b_s are the slope and the intercept, respectively, of the linear relationship between ${}^s\text{p}K_a$ values in acetonitrile:water and the ${}^w\text{p}K_a$ values in pure water for aliphatic carboxylic acids, and δ is a parameter that allows conversion between both $\text{p}K_a$ scales [24]. The three parameters depend on the hydro-organic composition.

3. RESULTS AND DISCUSSION

Optimization of HILIC separation of the proteolytically digested CMP. It is generally accepted that retention in HILIC is mainly driven by the interaction of analytes with the polar stationary phase through partitioning into a water-rich layer, which is generated through chemisorption phenomena at the surface of the stationary phase [9, 15, 25]. Thus, seven different binary gradients of acetonitrile (solvent A) and Milli-Q water (solvent B) both with 0.05% formic acid (v/v) were assayed. Finally, the best conditions for the separation of the hydrolysate of CMP were as follows: from 0 to 5 min a constant 30% B; from 5 to 40 min a linear gradient from 30 to 45% B; from 40 to 45 min a linear gradient from 45 to 90% B; from 45 to 75 min at 90% B isocratic; ramped to original composition in 1 min; then equilibrated for 15 min. The resulting base peak chromatogram for the proteolytically digested CMP revealed the presence of a complex mixture of peptides eluting between 4 and 30 min (Figure 1B) indicating a high rate of proteolysis of the CMP.

Since other secondary mechanisms, in addition to partitioning, such as electrostatic interactions and hydrogen bonding may also be involved in the HILIC retention process [15, 26], the effect of the final concentration of formic acid, from 0.005 to 1.5% (v/v) leading to a mobile phase s_w pH range from 3.6 to 2.1, on the separation of the CMP hydrolysate was also investigated. Changing pH should have an effect mainly on the net charge of the peptides as it has been described that the surface charge of the sulfoalkylbetaine zwitterionic stationary phase ($-\text{CH}_2\text{N}^+(\text{CH}_3)_2-\text{CH}_2-\text{CH}_2-\text{CH}_2-\text{SO}_3^-$) is not largely affected by changes in pH [15], specially between pH 3 and 8 [27]. Consequently, different separation selectivity was observed depending on the concentration of formic acid, indicating that the interactions between the peptides or the

terminal unit of the glycans and the stationary phase changed. Figure 1 shows the base peak chromatograms of the CMP hydrolysate obtained under different concentrations of formic acid. In general, a trend of longer retention with increasing pH was obtained and, consequently, addition of the lowest concentration of formic acid, 0.005% v/v (Figure 1A), seemed to provide the best chromatographic separation conditions among all assayed concentrations.

HILIC separation of the glycosylated tetrapeptide model systems. In order to rationalize the retention behaviour of glycosylated peptides, several model systems of the tetrapeptide, acetylS-D-K-P, covalently conjugated to different carbohydrates (neutral and sialylated) indicated in the experimental section, were analysed under different concentrations of formic acid (Figures 2 and 3). The retention of the unglycosylated peptide increased steadily in the studied pH range, especially between 2.6 and 3.6 (Figure 2, peak 1). At the lowest concentration of formic acid, 0.005% v/v, the unglycosylated peptide exhibited a long retention time (38.7 min) probably due to the high hydrophilicity of the positively charged peptidic chain, with three of the four residues being polar amino acids. The presence of a lysine residue, which is a basic amino acid, largely contributed to the HILIC retention as the zwitterionic stationary phase acts, in addition to the water-retaining property, as weak-cation exchange LC column between the basic amino acids and the negatively charged terminal sulfonate group [15]. The decrease in the retention time of the positively charged peptide as the concentration of formic acid increased could be attributed to a gradual suppression of the electrostatic (attraction) interactions with the zwitterionic stationary phase. In good agreement with this behaviour, a decrease in the capacity factor of peptides such as bradykinin or angiotensin II separated in a ZIC-HILIC column was observed as the formic acid concentration increased (http://www.nestgrp.com/pdf/Zp1/ZIC_02poster.pdf).

Regardless the formic acid concentration, a remarkable increase of the retention was detected following **glycation** of the tetrapeptide with neutral carbohydrates (Figure 2, peaks 2, 3 and 4), this increase being proportional to the length of the attached carbohydrate. This behaviour can be attributed to the incorporation of galactose or glucose residues into the peptidic chain that leads to an increase in peptide hydrophilicity and, therefore, to a higher partitioning (hydrogen bonding) with the water-enriched layer covering the polar hydrophilic stationary phase.

When this model peptide was **glycated** with sialylated oligosaccharides, a dramatic decrease in the retention time was found at 0.005% (v/v) of formic acid, this effect even being more considerable for the sialylated trisaccharide isomers than for the pentasaccharide (Figure 3A, peaks 8, 9 and 10). Nevertheless, increasing the formic acid concentration affected selectivity in the ZIC-HILIC column (Figures 3B, C and D). Thus, the retention of the **unglycated** peptide (peak 1) gradually decreased with the increased formic acid concentration as explained above; whilst an increase in the retention of the **glycated** peptides (peaks 8, 9 and 10) was observed when 0.05% of formic acid was assayed. Higher concentrations of formic acid (0.5 and 1.5%) have hardly any effect on the retention of the **glycated** peptides with both trisaccharides and the pentasaccharide (Figures 3C and 3D, peaks 8, 9 and 10). It is well known, that sialic acids are strongly acidic, e. g. the pK_a of Neu5Ac is between 2.6 and 2.9 [28], and the carboxylic acid group at the C-1 position is easily ionized providing, thus, a negative charge to the molecule. Therefore, it is very plausible that, in addition to hydrophilic interactions, some electrostatic repulsion between the localized negative charge on the sialylated oligosaccharides and that on the stationary phase are involved in the retention of the **glycated** model tetrapeptide. These electrostatic repulsion interactions could also explain the low retention of the free sialylated carbohydrates, non-attached to the

peptide, particularly when 0.005% formic acid was used (Figure 3, peaks 5, 6 and 7). The ${}^w pK_a$ values of the sialylated oligosaccharides were calculated by using the ADME suite programme, and their respective ${}^s pK_a$ values at 70:30 and 10:90 of acetonitrile:water (v:v) through Equation 1 (Table 1). From the calculated ${}^s pK_a$ values, it can be inferred that at 0.005% of formic acid concentration (pH of the mobile phase 3.6-3.2), all sialylated oligosaccharides can be negatively charged due to the deprotonation of the carboxylic acid group at the C-1 of the sialic acid. Whereas by increasing the amount of formic acid (i.e., lowering the pH of the mobile phase), the retention of the free sialylated oligosaccharides was also increased because of the partial protonation of the sialic acids and, thereby, minimizing the electrostatic repulsions with the negatively charged terminal sulfonate group (Figure 3, peaks 5, 6 and 7).

Finally, it should be remarkable that both sialylated trisaccharide isomers could be partially separated either in their unconjugated form (Figure 3, peaks 5 and 6) or conjugated with the tetrapeptide (Figure 3, peaks 8 and 9), the trisaccharide with the α 2-3 linked sialic acid being less retained than the isomer with the α 2-6 linked sialic acid. This fact reveals that the ZIC-HILIC column has the capacity for separation of isomeric sialylated glycans with small oligosaccharide and peptide chains. Takegawa et al. [19] described a similar behaviour, although they studied the separation of large isomeric tri- and tetra-antennary *N*-glycan structures. These authors indicated that the higher retention of the α 2-6 sialylated isomers could be due to the higher flexibility in changing its conformation/orientation as compared to the relatively rigid α 2-3 sialylated isomer so, thus, the electrostatic repulsion interaction might be minimized. In our case, according to their ${}^s pK_a$ values (Table 1), we also postulate that the α 2-6 sialylated isomer is slightly less acidic and, therefore, less negatively charged than the

277 trisaccharide with the α 2-3 linked sialic acid, resulting in a diminishing of the
278 electrostatic repulsion and leading to a higher retention. In addition, free or conjugated
279 sialylated pentasaccharide (Figure 3, peaks 7 and 10) had a higher retention than the
280 corresponding sialylated trisaccharide isomers probably due to the partitioning effect as
281 explained above under all assayed conditions.

282 Overall, it can be concluded that the best separation between the neutral and the
283 sialylated model peptides was obtained with the lowest concentration of formic acid
284 assayed (0.005%, v:v). This behaviour agrees with the separation of CMP, which
285 contains either neutral or sialylated glycans, at different concentrations of formic acid
286 shown in Figure 1. Furthermore, it should be noted that the presence of low
287 concentrations of formic acid in the mobile phase are favoured for MS detection, as
288 these tend to improve the signal intensity. Consequently, the addition of 0.005% of
289 formic acid was used for the characterization of the *O*-sialoglycopeptides from
290 proteolytically digested CMP.

291 **Characterization of *O*-sialoglycopeptides derived from CMP by HILIC-**
292 **ESI(+)-MSⁿ.** HILIC-ESI(+)-MS² analyses of *O*-glycopeptides mainly resulted in
293 glycosidic bond fragmentation, providing Y_n-type ions, according to the nomenclature
294 of Domon & Costello [29] for glycan cleavages, corresponding to intact peptidic chain
295 fragments formed by sequential losses of carbohydrate residues from the [M+H]⁺ ion.
296 While only the minor sugar oxonium ion (B_n-type) corresponding to the intact glycan
297 could be detected (i.e: *m/z* 948; Figure 4). Therefore, based on the detected Y-ions, the
298 carbohydrate sequence of the glycan could be completely elucidated. Thus, the analysis
299 by HILIC-ESI-MS² on a three-dimensional ion trap monitoring neutral losses of 291 u
300 (singly charged ion, data not shown) or 145.5 u (doubly charged ion, Figure 5),
301 corresponding to the loss of one molecule of Neu5Ac, was a suitable and direct method

to detect *O*-sialoglycopeptides. Furthermore, a wide range of retention times (from 3 to 50 min) of *O*-sialoglycopeptides was observed under these conditions, which is indicative of the importance of the nature of the amino acidic sequence (i. e., hydrophilicity and surface net charge), as well as the charge state and size of the attached glycan. Likewise, monitorization of neutral losses of 162 / 203 u (singly charged ions) and 81 / 101.5 u (doubly charged ions) corresponding to Hex / HexNAc residues, respectively, were also successfully performed in order to detect *O*-glycopeptides not containing sialic acid residues (data not shown).

Whilst the most abundant ion present in all MS² spectra corresponded to the unglycosylated (intact) peptide, the sequence informative *b*- and *y*-ions, following the nomenclature of Roepstorff & Fohlman [30], resulting from peptide backbone cleavage were normally very weak or undetected. This fact impaired the accurate and direct identification of the amino acidic sequence of target glycopeptides by HILIC-ESI(+)-MS² analyses. Nevertheless, taking advantage of multi-stage fragmentation capabilities of ion traps, analyses of the corresponding unglycosylated peptide ion with a subsequent stage of MS were successfully performed. Thus, MS³ spectra obtained from the collisionally generated ion corresponding to the unmodified peptide were preferentially characterized by the *b*-serial ions owing to the free *N*-terminus (Figure 6), allowing an accurate identification of the peptide backbone.

Table 2 summarizes the 41 *O*-glycopeptides belonging to CMP genetic variants A and B identified by HILIC-ESI(+)-MSⁿ. Three glycopeptides, ¹¹³N-N¹²³, ¹³⁸A-L¹⁴⁶, and ¹⁶¹T-V¹⁶⁹, could be characterized containing either the sialyltetra- or the sialyltrisaccharide glycans. This implies an additional advantage of using HILIC separation with a low concentration of formic acid, as it is very usual that at least one sialic acid residue can be lost during sample processing for proteomics or reverse phase

LC analysis due to its lability under acidic conditions; thus, impairing the detection of CMP glycopeptides containing the attached tetrasaccharide [31, 32].

According to their order of elution, it was observed that those glycopeptides containing the tetrasaccharide (comprised of two Neu5Ac residues) had shorter retention times than the counterpart glycopeptide containing the trisaccharides (one Neu5Ac), which in turn also had a shorter retention time than the corresponding glycopeptide modified with the (neutral) di- or monosaccharide. These data were in very good agreement with the behaviour observed above for the model tetrapeptide *glycated* with sialylated and neutral oligosaccharides. This behaviour was also observed for those *O*-sialoglycopeptides, such as $^{113}\text{N-N}^{123}$, largely retained in the column (t_R of 21 min; Table 2). In this case this high retention could be predominantly attributed to the presence of one lysine residue in the peptidic chain which may largely contribute to the HILIC retention as it was observed for the model tetrapeptide.

On the other hand, the detection of isobaric sialoglycopeptides at different t_R values was mostly observed when the attached glycan was a trisaccharide (Table 2). This fact was attributed to the occurrence of two different isomeric trisaccharides, linear ($\alpha 2$ -3) or branched ($\alpha 2$ -6), depending on the attachment of the sialic acid. In a similar way to the couple of sialylated trisaccharide isomers employed for the *glycation* of the model tetrapeptide, the branched CMP *O-linked* trisaccharide had higher $^w pK_a$ and $^s pK_a$ values than the *O-linked* trisaccharide with the $\alpha 2$ -3 linked sialic (Table 1). Thereby, it could be expected that the CMP glycopeptides containing the branched trisaccharide had a higher HILIC retention due to a low degree of electrostatic repulsion interactions than their counterparts carrying the linear isomeric trisaccharide, as it occurred for the model peptide *glycated* with the sialylated trisaccharide isomers (Figure 3, peaks 8 and 9). This behaviour was confirmed by the MS^2 detection of the

fragment $Y_{1\beta}$ -type ion corresponding to the initial neutral loss of Gal residue ($[M+H]^+ = 1399.2$ in Figure 7B), denoting the presence of this carbohydrate at the terminal position of the glycan structure and, therefore, revealing the presence of the branched trisaccharide (Figure 7). Consequently, as it is shown in Table 2, peptides having the same amino acid sequence but glycosylated with the sialyl $\alpha 2-3$ trisaccharide eluted earlier than those conjugated with the sialyl $\alpha 2-6$ isomer.

Furthermore, glycopeptides $^{124}\text{T-E}^{137}$ and $^{125}\text{I-E}^{137}$ modified with the single monosaccharide, and $^{124}\text{T-T}^{135}$ that contained the disaccharide also showed two different t_R which could be due to the attachment of the glycan to a different glycosylation site (T^{131} or T^{133}) in the peptide chain. A similar behaviour was previously described in the characterization by RP-LC of intact glycoforms of bovine [33], ovine [34] and caprine [35] CMP.

Identification of the specific glycosylation-site. Bovine CMP has up to six well-defined *O*-glycosylation sites located at T^{121} , T^{131} , T^{133} , T^{136} , T^{142} and T^{165} (residue numbering is based on the Swiss-Prot entry for the mature form of bovine κ -casein, accession number P02668) as it was determined by solid-phase Edman degradation [36]. A more recent proteomic identification of the modified sites, which involved the previous separation of the glycoforms by two-dimensional electrophoresis, revealed a distinct hierarchy in the *O*-glycosylation of κ -casein that implies an ordered addition of glycans to T^{131} , T^{142} and T^{133} [32]. Our work has permitted a comprehensive characterization of the *O*-glycopeptides of CMP because the identified structures covered all threonine residues described as potential *O*-glycosylation sites (Figure 8). Thus, in 24 of 41 glycopeptides an unambiguous characterization of the *O*-glycosylation site could be obtained based on the previously described glycosylation sites (Table 2).

Moreover, as glycosylation confers protease resistance to nearby peptide bonds, it is very plausible that the peptide $^{134}\text{P-T}^{142}$ is glycosylated at T^{136} rather than T^{142} .

Nevertheless, the specific glycosylation site of a few peptides located at the region $^{124}\text{T-E}^{137}$, which contains three potential glycosylation-sites (T^{131} , T^{133} and T^{136}), could not be straightforward assigned (Table 2). In order to overcome this drawback, several strategies were carried out. The first approach was based on the use of the fragment containing the peptidic chain attached to the GalNAc residue (which was the second most abundant ion in the positive-ion MS^2 spectra) as a precursor ion to trigger the MS^3 spectra. A similar experiment was successfully carried out for the identification of *N*-glycosylation sites [37] by using electrospray ionization on a linear ion trap time-of-flight mass spectrometer. Nevertheless, in the case of CMP *O*-glycopeptides analysed on an ion trap, this approach resulted in the rapid cleavage of the glycosidic bond (according to the high abundance of the whole peptide fragment without the GalNAc residue attached). The only detected peptide fragment which still carried the GalNAc residue attached was the *b*-ion containing all the sequence except the C-terminal amino acid residue, thus, providing no further evidence of the specific glycosylation site (data not shown).

Secondly, despite MS^2 spectra of glycopeptides showed abundant ions from glycosidic bond and very low abundant ions derived from peptide-bond fragmentation, useful data could be obtained from an extended search for the low-abundant peptide-backbone ions with minimal fragmentations of glycans. Using this approach, the T^{136} could be ruled out as a glycosylation site in the peptides $^{124}\text{T-E}^{137}$ (Figure 9A) and $^{125}\text{I-E}^{137}$ (data not shown). For example, this was possible due to the detection of the minor b_{10} ion, i. e. the fragment $^{124}\text{T-T}^{133}$ that contains the T^{131} and T^{133} residues, glycosylated either with the complete trisaccharide (m/z 1601) or the GalNAc residue (m/z 1148) in

the MS² spectrum of the glycopeptide ¹²⁴T-E¹³⁷ (Figure 9A). However, it was not feasible to distinguish between T¹³¹ and T¹³³ as glycosylation sites, since *b*₉ and *b*₈ or *y*₆ and *y*₅ ions, all of them containing only one of the two potential glycosylation sites, could not be detected. This strategy was also used to verify the glycosylation sites which were unambiguously assigned according to the described potential glycosylation sites in glycopeptides containing several threonine residues. As examples, this was the case for ¹³⁸A-L¹⁴⁶ (Figure 4) or ¹¹⁷T-N¹²³ (Figure 9B). In the latter case, the fragment *y*₄ (¹²⁰P-N¹²³), indicative of the T¹²¹ as the specific glycosylation site, was reasonably abundant due to the termed “proline effect” resulting from cleavages at the N-terminal side of proline [38, 39]. In this sense, it has been described that *y* ions that have a proline residue as their N-terminus are normally over-represented in the tandem mass spectra due to their relatively higher stability [40]. Similarly, Medzihradszky et al. [41] reported a direct mass-spectrometric approach to structurally elucidate *O*-linked glycopeptides and their glycosylation sites using high-energy tandem mass spectra obtained on four sector instrument (EBEB), equipped with a liquid secondary ion mass spectrometry source. In that work, glycosylated peptide fragment ions were also scarce and of low abundance, but sufficiently relevant for the determination of the glycosylation site.

Characterization of *O*-sialoglycopeptides derived from CMP by HILIC-ESI(-)-MSⁿ. HILIC-ESI(-)-MS² analyses were mostly dominated by the Y_n-type fragmentation derived exclusively from the losses of Neu5Ac residues. This fact is due to sialic acid is the only carbohydrate found in CMP that can provide negative charge to the glycopeptides at acidic or neutral pH because its acidic nature. Furthermore, MS² analyses at negative ionization mode also provided additional information on non-

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

glycosylated threonine residues due to the loss of C₂H₄O (44 u) from their side chains. As an example, Figure 10 shows the MS² spectrum at negative ionization mode of the glycopeptide ¹⁶¹T-V¹⁶⁹ containing the tetrasaccharide linked to T¹⁶⁵ with two additional non-glycosylated threonine residues (T¹⁶¹ and T¹⁶⁷). Additionally, neutral losses of 98 u were also detected and attributed to serine dephosphorylation (data not shown). In this sense, it has been described that S¹⁴⁹ is fully phosphorylated, being S¹²⁷ the second most important phosphorylation site in bovine CMP [7, 42].

4. CONCLUDING REMARKS

Overall, the developed HILIC-ESI(+)-MSⁿ method allowed the characterization of 41 *O*-glycopeptides from CMP genetic variants A and B without the need of a previous enrichment step. This method was suitable for i) the satisfactory separation of the glycopeptides according to their glycan structure and amino acidic sequence; and ii) for the subsequent determination of the amino acid and carbohydrate sequences of glycopeptides. Furthermore, complementary information about the degree of sialylation and the number of non-glycosylated threonine and phosphorylated serine residues was also obtained by performing analysis at negative ionization mode under the same chromatographic conditions.

To conclude, our study points out the potential of HILIC coupled to multistage MS procedures for future glycoproteomic applications. In the glycobiology field, combined advances in chromatographic and mass spectrometric techniques are being of great utility for gaining a better understanding of the molecular basis of the function of glycoproteins in biological processes.

451

452 **5. ACKNOWLEDGEMENTS**

453 The authors gratefully acknowledge financial support from the Spanish Council
454 for Scientific Research (CSIC) (PIF-SIALOBIOTIC 200870F010-1, -2) and from the
455 Spanish Ministry of Education and Science (CONSOLIDER INGENIO 2010
456 Programme; FUN-C-FOOD CSD2007-063). Oswaldo Hernández-Hernández thanks the
457 CSIC for a JAE-PreDoc PhD-grant. We thank Laurie Davis from Davisco Foods
458 International, Inc for kindly providing us with CMP.

6. REFERENCES

1. Vance, B.A., Wu, W., Ribaud, R., Segal, D.M., *et al.* CD69 heterogeneity results from usage of both typical (N-X-T) and atypical (N-X-C) N-linked glycosylation motifs. *J. Allergy Clin. Immunol.* 1997, 99, 1869-1869.
2. Seipert, R.R., Dodds, E. D., Lebrilla, C. B. Exploiting Differential Dissociation Chemistries of O-Linked Glycopeptide Ions for the Localization of Mucin-Type Protein Glycosylation. *J. Proteome Res.* 2009, 8, 493-501.
3. Abd El-Salam, M.H., El-Shibini, S., Buchheim, W. Characteristics and potential uses of the casein macropeptide. *Int. Dairy J.* 1996, 6, 327-341.
4. Manso, M. A., López-Fandiño, R. κ -casein macropeptides from cheese whey: physicochemical, biological, nutritional, and technological features for possible uses. *Food Rev. Int.* 2004, 20, 329-355.
5. Thomä-Worringer, C., Sørensen, J., López-Fandiño, R. Health effects and technological features of caseinomacropeptide. *Int. Dairy J.* 2006, 16, 1324-1333.
6. Dziuba, J., Minkiewicz, P. Influence of glycosylation on micelle-stabilizing ability and biological properties of C-terminal fragments of cow's κ -casein. *Int. Dairy J.* 1996, 6, 1017-1044.
7. Vreeman, H. J., Visser, S., Slangen, C. J., van Riel, J. A. M. Characterization of bovine κ -casein fractions and the kinetics of chymosin-induced macropeptide release from carbohydrate-free and carbohydrate containing fractions determined by high performance gel permeation chromatography. *Biochem. J.* 1986, 240, 87-97.
8. Saito, T., Itoh, T. Variations and distributions of O-glycosidically linked sugar chains in bovine κ -casein A. *J. Dairy Sci.* 1992, 75, 1768-1774.

- 1
2
3 484 9. Alpert, A.J. Hydrophilic-Interaction chromatography for the separation of
4
5 485 peptides, nucleics-acids and other polar compounds. *J. Chromatogr.* 1990, 499,
6
7 486 177-196.
8
9
10 487 10. Alpert, A. J., Shukla, M., Shukla A.K., Zieske, L. R., *et al.* Hydrophilic-
11
12 488 Interaction Chromatography of complex carbohydrates. *J. Chromatogr. A* 1994,
13
14 489 676, 191-202.
15
16
17 490 11. Guo, Y., Huang, A. H. A HILIC method for the analysis of tromethamine as the
18
19 491 counter ion in an investigational pharmaceutical salt. *J. Pharm. Biomed. Anal.*
20
21 492 2003, 31, 1191-1201.
22
23
24 493 12. Guo Y, Gaiki S. Retention behavior of small polar compounds on polar
25
26 494 stationary phases in hydrophilic interaction chromatography. *J. Chromatogr. A*
27
28 495 2005, 1074, 71-80.
29
30
31 496 13. Godejohann, M. Hydrophilic interaction chromatography coupled to nuclear
32
33 497 magnetic resonance spectroscopy and mass spectroscopy - A new approach for
34
35 498 the separation and identification of extremely polar analytes in bodyfluids. *J.*
36
37 499 *Chromatogr. A* 2007, 1156, 87-93..
38
39
40 500 14. McCalley, D. V. Study of the selectivity, retention mechanisms and performance
41
42 501 of alternative silica-based stationary phases for separation of ionised solutes in
43
44 502 hydrophilic interaction chromatography. *J. Chromatogr.* 2010, 1217, 3408-
45
46 503 3417.
47
48
49 504 15. Hemstrom, P., Irgum, K. Hydrophilic interaction chromatography. *J. Sep. Sci.*
50
51 505 2006, 29, 1784-1821.
52
53
54 506 16. Berthod, A., Chang, S. S. C., Kullman, J. P. S., Armstrong, D. W. Practice and
55
56 507 mechanism of HPLC oligosaccharide separation with a cyclodextrin bonded
57
58 508 phase. *Talanta* 1998, 47, 1001-1012.
59
60

- 1
2
3 509 17. Hao Z. G., Xiao B. M., Weng N. D. Impact of column temperature and mobile
4
5 510 phase components on selectivity of hydrophilic interaction chromatography
6
7 511 (HILIC). *J. Sep. Sci.* 2008, *31*, 1449-1464.
8
9
10 512 18. Wuhrer, M., Koeleman, C. A. M., Hokke, C. H., Deelder, A. M. Protein
11
12 513 glycosylation analyzed by normal-phase nano-liquid chromatography-mass
13
14 514 spectrometry of glycopeptides. *Anal. Chem.* 2005, *77*, 886-894.
15
16
17 515 19. Takegawa, Y., Deguchi, K., Ito H., Keira, T., *et al.* Simple separation of
18
19 516 isomeric sialylated *N*-glycopeptides by a zwitterionic type of hydrophilic
20
21 517 interaction chromatography. *J. Sep. Sci.* 2006a, *29*, 2533-2540.
22
23
24 518 20. Takegawa, Y., Deguchi, K., Ito H., Keira, T., *et al.* Separation of isomeric 2-
25
26 519 aminopyridine derivatized *N*-glycans and *N*-glycopeptides of human serum
27
28 520 immunoglobulin G by using a zwitterionic type of hydrophilic-interaction
29
30 521 chromatography. *J. Chromatogr. A* 2006b, *1113*, 177-181.
31
32
33 522 21. Takegawa, Y., Hato, M., Deguchi, K., Nakagawa, H., *et al.* Chromatographic
34
35 523 deuterium isotope effects of derivatized *N*-glycans and *N*-glycopeptides in a
36
37 524 zwitterionic type of hydrophilic interaction chromatography. *J. Sep. Sci.* 2008,
38
39 525 *31*, 1594-1597.
40
41
42 526 22. Picariello, G., Ferranti, P., Mamone, G., Roepstorff, P., *et al.* Identification of N-
43
44 527 linked glycoproteins in human milk by hydrophilic interaction liquid
45
46 528 chromatography and mass spectrometry. *Proteomics* 2008, *8*, 3833-3847.
47
48
49 529 23. Espinosa S., Bosch E., Roses, M. Retention of ionizable compounds on HPLC.
50
51 530 5. pH scales and the retention of acids and bases with acetonitrile-water mobile
52
53 531 phases. *Anal. Chem.* 2000, *72*, 5193-5200.
54
55
56
57
58
59
60

24. Gagliardi, L.G., Castells, C.B., Rafols, C., Roses, M., Bosch, E. δ Conversion
Parameter between pH Scales (s_pK_a and s_pK_a) in Acetonitrile/Water Mixtures
at Various Compositions and Temperatures. *Anal. Chem.* 2007, 79, 3180-3187.
25. Yang, Y.Z., Boysen, R.I., Hearn, M.T.W. Hydrophilic interaction
chromatography coupled to electrospray mass spectrometry for the separation of
peptides and protein digests. *J. Chromatogr. A* 2009, 1216, 5518-5524.
26. Yoshida, T. Peptide separation by Hydrophilic-Interaction Chromatography: a
review. *J. Biochem. Biophys. Methods* 2004, 60, 265-280.
27. Boersema, P.J., Divecha, N., Heck, A.J.R., Mohammed, S. Evaluation and
optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. *J. Proteome
Res.* 2007, 6, 937-946.
28. Scheinthal, B.M., Bettelheim, F.A. Multiple forms of sialic acids. *Carbohydr
Res* 1968, 6, 257-265.
29. Domon, B., Costello, C.E. A systematic nomenclature for carbohydrate
fragmentations in FAB-MS MS spectra of glycoconjugates. *Glycoconj. J.* 1988,
5, 397-409.
30. Roepstorff, P., Fohlman, J. Proposal for a common nomenclature for sequence
ions in mass-spectra of peptides. *Biomed. Mass Spectrom.* 1984, 11, 601-601.
31. Holland, J. W., Deeth, H., Alewood, P. F. Proteomic analysis of κ -casein micro-
heterogeneity. *Proteomics* 2004, 4, 743-752.
32. Holland, J. W., Deeth, H., Alewood, P. F. Analysis of O-glycosylation site
occupancy in bovine κ -casein glycoforms separated by two dimensional gel
electrophoresis. *Proteomics* 2005, 5, 990-1002.

- 1
2
3 555 33. Mollé, D., Léonil, J. Heterogeneity of the bovine κ -casein caseinomacropeptide,
4
5
6 556 resolved by liquid chromatography on-line with electrospray ionization mass
7
8 557 spectrometry. *J. Chromatogr. A* 1995, 708, 223-230.
9
10 558 34. Moreno, F. J., Recio. I., Olano, A., Lopez-Fandino, R. Chromatographic
11
12 559 characterization of ovine κ -casein macropeptide. *J. Dairy Res.* 2000, 67, 349-
13
14 560 359.
15
16
17 561 35. Moreno, F. J., Recio. I., Olano, A., Lopez-Fandino, R. Heterogeneity of caprine
18
19 562 κ -casein macropeptide. *J. Dairy Res.* 2001, 68, 197-208.
20
21
22 563 36. Pisano, A., Packer, N. H., Redmond, J. W., Williams, K. L., *et al.*
23
24 564 Characterization of O-linked glycosylation motifs in the glycopeptides domain
25
26 565 of bovine kappa-casein. *Glycobiology* 1994, 4, 837-844.
27
28
29 566 37. Deguchi, K., Ito, H., Takegawa, Y., Shinji, N., *et al.* Complementary structural
30
31 567 information of positive- and negative-ion MSⁿ spectra of glycopeptides with
32
33 568 neutral and sialylated *N*-glycans. *Rapid Commun Mass Spectrom* 2006, 20, 741-
34
35 569 746.
36
37
38 570 38. Hunt, D. F., Yates, J. R., Shabanowitz, J., Winston, S., *et al.* Protein sequencing
39
40 571 by tandem mass-spectrometry. *Proc. Natl. Acad. Sci. USA* 1986, 83, 6233-6237.
41
42
43 572 39. Vaisar, T., Urban, J. Probing the proline effect in CID of protonated peptides. *J.*
44
45 573 *Mass Spectrom.* 1996, 31, 1185-1187.
46
47
48 574 40. Schaaff, T. G., Cargile, B. J., Stephenson, J. L., McLuckey, S. A. Ion trap
49
50 575 collisional activation of the (M + 2H)²⁺ - (M + 17H)¹⁷⁺ ions of human
51
52 576 hemoglobin β -chain. *Anal. Chem.* 2000, 72, 899-907.
53
54
55 577 41. Medzihradszky, K.F., Gillece-Castro, B.L., Townsend, R.R., Burlingame, A.L.,
56
57 578 *et al.* Structural elucidation of O-linked glycopeptides by high energy collision-
58
59 579 induced dissociation. *J. Am. Soc. Mass Spectrom.* 1996, 7, 319-328.
60

1
2
3 580 42. Rasmussen, L.K., Sorensen, E.S., Petersen, T.E., Nielsen, N.C., *et al.*
4
5 581 Characterization of phosphate sites in native ovine, caprine, and bovine casein
6
7 582 micelles and their caseinomacropeptides: a solid state phosphorus-31 nuclear
8
9 583 magnetic resonance and sequence and mass spectrometric study. *J. Dairy Sci.*
10
11 584 1997, 80, 607-614.
12
13 585
14
15 586
16
17 587
18
19 588
20
21 589
22
23 590
24
25 591
26
27 592
28
29 593
30
31 594
32
33 595
34
35 596
36
37 597
38
39 598
40
41 599
42
43 600
44
45 601
46
47 602
48
49 603
50
51 604
52
53
54
55
56
57
58
59
60

Figure legends

Figure 1. HILIC-ESI-MS base peak profiles monitorized at positive ionization mode of proteolytically digested CMP using different formic acid concentrations (% v/v). A) 0.005%; B) 0.05%; C) 0.5%; D) 1.5%. s_w pH values at 70:30 and 10:90 (v:v) acetonitrile:water of mobile phase are shown in the figure.

Figure 2. HILIC-ESI-MS base peak profiles monitorized at positive ionization mode of the tetrapeptide (acetylS-D-K-P) model system conjugated with neutral carbohydrates using different formic acid concentrations (% v/v). A) 0.005%; B) 0.05%; C) 0.5%; D) 1.5%. Labelled peaks are as follows: **(1)** Tetrapeptide; **(2)** Tetrapeptide conjugated with Galactose; **(3)** Tetrapeptide conjugated with Lactose; **(4)** Tetrapeptide conjugated with Maltopentaose. s_w pH values at 70:30 and 10:90 (v:v) acetonitrile:water of mobile phase are shown in the figure.

Figure 3. HILIC-ESI-MS base peak profiles monitorized at negative ionization mode of the tetrapeptide (acetylS-D-K-P) model system conjugated with sialylated carbohydrates using different formic acid concentrations (% v/v). A) 0.005%; B) 0.05%; C) 0.5%; D) 1.5%. Labelled peaks are as follows: **(1)** Tetrapeptide; **(5)** Neu5Ac α 2-3Gal β 1-4Glc; **(6)** Neu5Ac α 2-6Gal β 1-4Glc; **(7)** Neu5Ac α 2-6Gal β 1-4GlcNAc β 1-3Gal β 1-4Glc; **(8)** Tetrapeptide conjugated with Neu5Ac α 2-3Gal β 1-4Glc; **(9)** Tetrapeptide conjugated with Neu5Ac α 2-6Gal β 1-4Glc; **(10)** Tetrapeptide conjugated with Neu5Ac α 2-6Gal β 1-4GlcNAc β 1-3Gal β 1-4Glc. s_w pH values at 70:30 and 10:90 (v:v) acetonitrile:water of mobile phase are shown in the figure.

630

Figure 4. MS² spectrum of a singly charged ion at m/z 1837.5 corresponding to the protonated CMP peptide ¹³⁸A-L¹⁴⁶ glycosylated with the sialyl-tetrasaccharide (Neu5Ac α 2-3Gal β 1-3(Neu5Ac α 2-6)GalNAc) at the T¹⁴² residue. Symbols: ■ N-Acetyl-Galactosamine; ● Galactose; ♦ Sialic acid.

635

Figure 5. Monitorization profile at positive ionization mode of neutral loss of 145.5 (doubly charged ion) corresponding to one molecule of Neu5Ac in the MS² spectrum of proteolytically digested CMP.

639

Figure 6. MS³ spectrum of a singly charged ion corresponding to the CMP glycopeptide ¹³⁸A-L¹⁴⁶ from collisionally generated unglycosylated peptide ion ([M+H]⁺ = 890.2) derived from the MS² analysis at positive ionization mode.

643

Figure 7. MS² spectra at positive ionization mode of a singly charged ion ([M+H]⁺ = 1561.40) corresponding to the CMP peptide ¹⁶¹T-V¹⁶⁹ glycosylated with A) the sialyl-trisaccharide Neu5Ac α 2-3Gal β 1-3GalNAc and B) the sialyl-trisaccharide Gal β 1-3(NeuAc α 2-6)GalNAc at the ¹⁶⁵T residue. Symbols: ■ N-Acetyl-Galactosamine; ● Galactose; ♦ Sialic acid.

649

Figure 8. Amino acid sequence of CMP (genetic variants A and B). Potential glycosylation sites are in bold and identified glycopeptides are underlined. Residue numbering is based on the Swiss-Prot entry for the mature form of bovine κ -casein (accession number P02668).

654

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 9. MS² spectra at positive ionization mode of: A) a doubly charged ion ([M+H]²⁺ = 1025.0) corresponding to the CMP peptide ¹²⁴Thr-Glu¹³⁷; and B) a singly charged ion ([M+H]⁺ = 1443.5) corresponding to the CMP peptide ¹¹⁷T-N¹²³, both glycosylated with the sialyl-trisaccharide Neu5Acα2-3Galβ1-3GalNAc.

Figure 10. MS² spectrum at negative ionization mode of a singly charged ion ([M-H]⁻ = 1850.6) corresponding to the CMP peptide ¹⁶¹T-V¹⁶⁹ glycosylated with the sialyl-tetrasaccharide (Neu5Acα2-3 Gal β1-3(Neu5Acα2-6)GalNAc) at T¹⁶⁵ residue.

Table 1. ${}^w pK_a$ and ${}^s pK_a$ values of the sialylated oligosaccharides identified in bovine CMP and of those employed for the [glycation](#) of the model tetrapeptide acetylS-D-K-P.

Sialyloligosaccharide structure	${}^w pK_a$ ¹	${}^s pK_a$ ²	
		70:30	10:90
Neu5Acα2-3Galβ1-4Glc	2.3	3.7	2.4
Neu5Acα2-6Galβ1-4Glc	2.4	3.9	2.6
Neu5Acα2-6Galβ1-4GlcNAcβ1-3Galβ1-4Glc	2.5	3.9	2.6
Neu5Acα2-3Galβ1-3GalNAc	2.3	3.7	2.4
Galβ1-3(Neu5Acα2-6)GalNAc	2.5	3.9	2.6
Neu5Acα2-3Galβ1-3(Neu5Acα2-6)GalNAc	2.5 and 2.3	3.9 and 3.7	2.6 and 2.4

666

667 ¹ Calculated by ADME suite software (see materials and methods section).

668 ² Calculated by Equation 1 at 70:30 and 10:90 acetonitrile:water (v/v) mixtures.

Table 2. O-Glycopeptides identified in proteolytically digested CMP by HILIC-ESI-MSⁿ.

t _R (min)	Glycopeptide sequence	Position ^a	Attached glycan	Exp. Mass [M+H] ^{±b}	Genetic variant	T glycosylated residue number
4.6	AVESTVATL	138-146	a	1837.4	A & B	142
7.9	AVESTVATL	138-146	b	1546.4	A & B	142
10.2	AVESTVATL	138-146	c	1546.4	A & B	142
9.9	TIASGEPTSTPTTE	124-137	b	2048.6	A	131 or 133
12.7	TIASGEPTSTPTTE	124-137	c	2048.6	A	131 or 133
46.5	TIASGEPTSTPTTE	124-137	d	1756.4	A	131 or 133
41.1 and 42.5	TIASGEPTSTPTTE	124-137	e	1594.4	A	131 or 133
4.6	TVQVTSTAV	161-169	a	1852.4	A & B	165
9.5	TVQVTSTAV	161-169	b	1561.4	A & B	165
12.5	TVQVTSTAV	161-169	c	1561.4	A & B	165
49.2	TVQVTSTAV	161-169	d	1270.3	A & B	165
46.5	TVQVTSTAV	161-169	e	1108.4	A & B	165
9.7	VTSTAV	164-169	b	1233.3	A & B	165
14.1	VTSTAV	164-169	c	1233.3	A & B	165
51.1	VTSTAV	164-169	d	942.2	A & B	165
49.9	VTSTAV	164-169	e	780.2	A & B	165
10.2	IASGEPTSTPTTE	125-137	b	1946.7	A	131 or 133
12.8	IASGEPTSTPTTE	125-137	c	1946.7	A	131 or 133
48.9	IASGEPTSTPTTE	125-137	d	1655.7	A	131 or 133
45.4 and 46.7	IASGEPTSTPTTE	125-137	e	1493.6	A	131 or 133
9.4	TIASGEPTSTPTIE	124-137	b	2059.4	B	131 or 133
45.4	TIASGEPTSTPTIE	124-137	d	1768.4	B	131 or 133
41.1	TIASGEPTSTPTIE	124-137	e	1606.5	B	131 or 133
8.6	TEIPTIN	117-123	b	1443.4	A & B	121
45.4	TEIPTIN	117-123	d	1152.3	A & B	121
40.9	TEIPTIN	117-123	e	990.3	A & B	121
10.7	IEAVESTVATL	136-146	b	1788.5	B	142
10.4	STVA	141-144	b	1033.2	A & B	142
51.0	STVA	141-144	d	742.4	A & B	142
50.1	STVA	141-144	e	580.3	A & B	142
10.3	TIASGEPTSTPT	124-135	b	1817.3	A & B	131 or 133
13.1	TIASGEPTSTPT	124-135	c	1817.3	A & B	131 or 133
47.9 and 48.6	TIASGEPTSTPT	124-135	d	1526.4	A & B	131 or 133
45.7	TIASGEPTSTPT	124-135	e	1364.3	A & B	131 or 133
15.5	STPTIE	132-137	c	1303.4	B	133
45.9	STPTIE	132-137	e	850.3	B	133
21.1	NQDKTEIPTIN	113-123	a	2220.0	A & B	121
48.9	NQDKTEIPTIN	113-123	b	1928.4	A & B	121
50.3	NQDKTEIPTIN	113-123	c	1928.4	A & B	121
50.2	PTTEAVEST	134-142	d	1299.3	A	136
48.7	PTTEAVEST	134-142	e	1137.3	A	136

- 1
2 671 **a) Tetrasaccharide:** Neu5Ac α 2-3Gal β 1-3(Neu5Ac α 2-6)GalNAc-O-R.
3 672 **b) Trisaccharide:** Neu5Ac α 2-3Gal β 1-3GalNAc-O-R
4 673 **c) Trisaccharide:** Gal β 1-3(Neu5Ac α 2-6)GalNAc-O-R
5 674 **d) Disaccharide:** Gal β 1-3GalNAc-O-R.
6 675 **e) Monosaccharide:** GalNAc-O-R.
7 676 ^a Residue numbering is based on the Swiss-Prot entry for the mature form of bovine κ -casein
8 677 (accession number P02668).
9 678 ^b Monoisotopic mass values
10 679

For Peer Review

Figure 1. Hernandez-Hernandez et al.

Figure 2. Hernandez-Hernandez et al.

Figure 3. Hernandez-Hernandez et al.

- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- 21
- 22
- 23
- 24
- 25
- 26
- 27
- 28
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39
- 40
- 41
- 42
- 43
- 44
- 45
- 46
- 47
- 48
- 49

Figure 5. Hernandez-Hernandez et al

Figure 6. Hernandez-Hernandez et al.

Figure 7. Hernandez-Hernandez et al.

Figure 8. Hernandez-Hernandez et al.

Figure 9. Hernandez-Hernandez et al.

Figure 10. Hernandez-Hernandez et al.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review