Journal of Chromatography A, 1218 (2011) 5101–5107

Contents lists available at ScienceDirect

Journal of Chromatography A

journal homepage: www.elsevier.com/locate/chroma

Effect of cyclization of N-terminal glutamine and carbamidomethyl-cysteine (residues) on the chromatographic behavior of peptides in reversed-phase chromatography

Janice Reimer[a], Dmitry Shamshurin[a], Michael Harder[c], Andriy Yamchuk[c], Vic Spicer[c], Oleg V. Krokhin[a,b,*]

[a] Manitoba Centre for Proteomics and Systems Biology, 799 JBRC, 715 McDermot Avenue, Winnipeg R3E 3P4, Canada
[b] Department of Internal Medicine, University of Manitoba, 799 JBRC, 715 McDermot Avenue, Winnipeg R3E 3P4, Canada
[c] Department of Physics and Astronomy, University of Manitoba, Winnipeg R3T 2N2, Canada

ARTICLE INFO

Article history:
Received 11 January 2011
Received in revised form 6 May 2011
Accepted 19 May 2011
Available online 27 May 2011

Keywords:
Peptide reversed-phase HPLC
Chemically modified peptides
Peptide retention prediction

ABSTRACT

N-terminal loss of ammonia is a typical peptide modification chemical artifact observed in bottom-up proteomics experiments. It occurs both in vivo for N-terminal glutamine and in vitro following enzymatic cleavage for both N-terminal glutamine and cysteine alkylated with iodoacetamide. In addition to a mass change of −17.03 Da, modified peptides exhibit increased chromatographic retention in reversed-phase (RP) HPLC systems. The magnitude of this increase varies significantly depending on the peptide sequence and the chromatographic condition used. We have monitored these changes for extensive sets (more than 200 each) of tryptic Gln and Cys N-terminated species. Peptides were separated on 100 Å pore size C18 phases using identical acetonitrile gradient slopes with 3 different eluent compositions: 0.1% trifluoroacetic acid; 0.1% formic acid and 20 mM ammonium formate at pH 10 as ion-pairing modifiers. The observed effect of this modification on RP retention is the product of increased intrinsic hydrophobicity of the modified N-terminal residue, lowering or removing the effect of ion-pairing formation on the hydrophobicity of adjacent residues at acidic pHs; and possibly the increased formation of amphipathic helical structures when the positive charge is removed. Larger retention shifts were observed for Cys terminated peptides compared to Gln, and for smaller peptides. Also the size of the retention increase depends on the eluent conditions: pH 10 ≪ trifluoroacetic acid < formic acid. Different approaches for incorporation these findings in the peptide retention prediction models are discussed.

© 2011 Elsevier B.V. All rights reserved.

1. Introduction

Reversed-phase HPLC is a leading tool for peptide separation, and used globally every day in thousands of research and industrial laboratories. The study of peptide RP-LC separation selectivity was a subject of intense research and discussions for more than 30 years [1,2]. Peptides are organo-polymeric compounds consisting of amino acid residues linked via peptide bonds. There are countless possible combinations of 20 naturally occurring amino acids, depending on the number of residues in the peptide chain. The notion that not only peptide composition, but also the sequence in which these residues are linked impacts peptide retention [3] makes the study and prediction of separation selectivity a very complicated and exciting problem. While hydrophobic interactions are believed to be a major contributing factor, the ionogenic nature of analytes and the fact that separation is typically performed in the presence of ion-pairing modifiers makes the separation mechanism more complex [4,5]. Possible formation of secondary structures upon interaction with the stationary phase is another unique feature of RP-LC of peptidic compounds [6].

So far, the vast majority of studies regarding peptide retention prediction were performed on peptides consisting of naturally occurring residues. There are, however, a number of unusual amino acids, and countless post-translational/chemical modifications. The latter became the subject of intensive studies with accelerated development of mass-spectrometry proteomic techniques starting in 1990s. Since then, a vast volume of information has been acquired showing the complexity of the proteome and importance of these modifications for protein function. So far, the study of these modifications was mostly mass-spectrometry driven; chemical modifications result in peptide molecular weight changes, and can be distinguished by MS or tandem MS/MS

* Corresponding author at: Manitoba Centre for Proteomics and Systems Biology, 799 JBRC, 715 McDermot Avenue, Winnipeg R3E 3P4, Canada. Tel.: +1 204 789 3283; fax: +1 204 480 1362.
E-mail address: krokhino@cc.umanitoba.ca (O.V. Krokhin).

0021-9673/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.chroma.2011.05.079

measurements. However, these chemical modifications will cause modification of both the mass and the chromatographic properties of the molecules. There is a general understanding among chromatographers how common modifications will affect peptide hydrophobicity or retention in RP systems: methionine oxidation decreases peptide hydrophobicity [7], removal of N-terminal amino group via acetylation or N-terminal cyclization produces the opposite effect [8,9], and deamidated peptides can exhibit both lower and higher retention in RP systems [10]. However, systematic studies of chromatographic behavior of modified peptides and attempts at developing predictive approaches are still very rare. The first attempts to develop retention prediction models for modified peptides were driven by their biological importance. Recently, retention of phosphorylated peptides was a subject of chromatographic studies. It was shown that phosphorylation can both increase and decrease retention of modified peptides [11]. Kim et al. studied chromatographic retention of 33 phosphorylated peptides together with their non-modified counterparts in attempt to develop a retention prediction model [12]. The sequence-specific character of peptide retention suggests that the retention of modified peptides will be dependent not only on composition (presence) of the modified residues but also on its position within a molecule. Therefore, chromatographic properties of these species should be studied using extensive data sets to provide as many combinations of adjacent residues and positions within a peptide chain as possible.

Another application of altered chromatographic behavior for the development of proteomic procedures is COFRADIC (combined fractional diagonal chromatography). It applies the methodology of orthogonal chromatography to the analysis of complex peptide mixtures when chemical modifications were imposed following the first RP-LC separation to provide variations in the species' chromatographic behavior and their respective shift in non-populated areas of chromatographic space. COFRADIC was introduced by Gevaert et al. [13], and applied to the isolation of methionyl [7], cysteinyl [14], and amino terminal peptides [15]. It was shown that methionine oxidation reduces RP-LC retention of peptides by 1–7 min for TFA-based eluent at a 0.7% per minute linear acetonitrile gradient [7]. Such variation is further evidence of the sequence-specific character of peptide retention in RP-LC systems.

Peptide species detected in proteomic LC–MS experiments can be divided in two groups: non-modified (including those carrying applied or constant chemical modifications like protective alkylation of cysteines) and modified. In this paper we refer to all species carrying free N-terminal amino-, C-terminal carboxy-group, cysteines alkylated with iodoacetamide (IAM-Cys) and 19 remaining unaltered residues as non-modified. The modified peptides under investigation include those with N-terminally cyclized Gln and IAM-Cys (pyro-Glu and pyro-cmC shown in Fig. 1). Overall modified peptides include some with in vivo post-translational modifications (PTMs) and chemical in vitro modifications. While both modifications complicate proteomic analyses by splitting peptide signals in both the MS and the LC spaces, in vivo PTMs carry important information related to protein structure and function. Chemical in vitro modifications are unwanted artifacts of sample preparation procedures. Examples of in vivo modifications include phosphorylation, acetylation, glycosylation, etc. Some modifications occur both in vivo and in vitro: methionine oxidation, deamidation of asparagine, and N-terminal loss of ammonia for glutamine due to cyclization (Fig. 1). Some modifications occur in vitro only as a consequence of chemical treatment during sample preparation. A typical case of such an artifact chemical modification is the N-terminal loss of ammonia from Cys alkylated with iodoacetamide (Fig. 1). Both examples of N-terminal cyclization feature similar chemistry and consequences: a 17.03 Da loss in mass and an increase in chromatographic retention. These modifications are

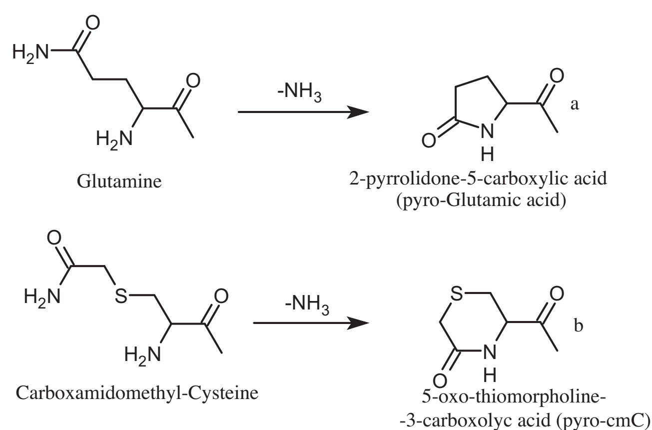

Fig. 1. Formation of pyroglutamic (a) and 5-oxo-thiomorpholine-3-carboxylic acids (pyro-cmC, b) from N-terminal Gln and IAM-Cys, respectively.

very abundant in enzymatic digestion protocols applied in bottom-up proteomics [9,16]. Thus, typical overnight digestion at 37 °C results in ~40% and 60% conversion into the products for Gln and Cys, respectively [9]. Khandke et al. [17] studied the influence of digestion conditions on the rate of pyro-Glu formation in attempt to minimize the degree of conversion. In our opinion, such an abundant modification should be studied carefully from both mass-spectrometry and chromatographic point of views. This will allow their inclusion in the list of "potential" chemical modifications to increase confidence in protein identification and characterization protocols.

While collecting peptide retention data sets for the development the Sequence Specific Retention Calculator (SSRCalc) retention prediction model [18], we have always observed these two modifications. SSRCalc exists as a family of models optimized for the most popular RP-HPLC conditions in proteomic applications: 100 Å pore size C 18 columns with trifluoroacetic (TFA) and formic acids (FA) as ion-pairing modifiers, and separation at pH 10 (ammonium formate). Subsequently, we obtained three different sets of data showing retention of both non-modified and −17 Da modified counterpart peptides. This paper presents a systematic overview of RP retention behavior, and an attempt to propose a sequence dependent approach for predicting RP-LC retention of peptides featuring N-terminal pyroglutamic and 5-oxo-thiomorpholine-3-carboxylic acids (pyro-cmC, Fig. 1).

2. Materials and methods

2.1. Materials

Deionized (18 MΩ) water and HPLC-grade acetonitrile were used for the eluents preparation. All chemicals were sourced from Sigma Aldrich (St.-Louis, MO), unless otherwise noted. Sequencing grade modified trypsin (Promega, Madison, WI) was used for digestion. A number of commercially available proteins as well as a few in-house purified (list provided in Appendix A) were used to generate mixtures of peptides for subsequent off-line LC–MALDI-MS or LC–ESI-MS/MS analysis and collection of peptide retention data sets.

2.2. Sample preparation

Stock solutions of proteins (1 mg/ml) were prepared in 100 mM NH_4HCO_3 buffer and digested with trypsin. Steps included reduction (10 mM dithiothreitol, 30 min, 57 °C), alkylation (50 mM iodoacetamide for 30 min in the dark, at room temperature),

dialysis (100 mM NH_4HCO_3, 6 h using a 7 kDa molecular weight cut-off filter; Pierce, Rockford, IL), and finally trypsin digestion (1:50 enzyme/substrate weight ratio for 12 h at 37 °C) [18]. Digests were combined to provide peptide mixtures of moderate complexity (100–200 species); ~1 to 2 pmole and ~100 to 200 fmole of each component per injection for LC–MALDI-MS and LC–ESI-MS/MS experiments, respectively. Each sample was spiked with the digest of a standard protein (horse myoglobin) or a mixture of standard peptides [19] for the retention time alignment purposes. Between 20 and 30 off-line LC–MALDI-MS runs were used to build each retention data set for three different chromatographic conditions. Two tryptic digests of bovine/human albumin, transferrin, fibrinogen, and lactotransferrin mixtures were used for the nano LC–ESI-MS/MS experiments: 250–300 peptides in total were confidently identified in these two runs as described elsewhere [20]. A detailed sample preparation protocol is provided in Appendix A.

2.3. HPLC instrumentation

A micro-Agilent 1100 Series system (Agilent Technologies, Wilmington, DE) was used for the off-line LC–MALDI-MS experiments [18] with direct sample injection (loop size 10 μl). Digests were fractionated on a 300 μm × 150 mm PepMap100 column (Dionex, Sunnyvale, CA, 0.1% formic acid and 0.1% trifluoroacetic acids as ion-pairing modifiers), and on a 150 μm × 150 mm XTerra column (Waters, Milford, MA, 20 mM ammonium formate, pH 10) for pH 10 experiments. A linear water–acetonitrile gradient of 0.75% per minute was used in all three cases, starting from 1% of organic solvent. Column effluent (3 μL/min) was mixed on-line with 2,5-dihydroxybenzoic matrix solution (150 mg/ml in water:acetonitrile 1:1; 0.5 μL/min), deposited on MALDI target at 30 s intervals, air-dried, and subjected to MALDI-MS analysis.

A split-less nano-flow Tempo LC system (Eksigent, Dublin, CA) with 20 μL sample injection via a 300 μm × 5 mm PepMap100 precolumn and a 100 μm × 150 mm analytical column packed with 5 μm Luna C18(2) (Phenomenex, Torrance, CA) was used for the LC–ESI-MS/MS analyses [20]. Both eluents A (0.5% acetonitrile in water) and B (98% acetonitrile) contained 0.1% formic acid as an ion-pairing modifier. Digest were fractionated using linear gradient 0.75% acetonitrile per minute starting from 0% B.

2.4. Mass spectrometry, peptide identification, retention time assignment

MALDI spectra of chromatographic fractions were collected using the Manitoba/Sciex prototype MALDI quadrupole/TOF (time-of-flight, QqTOF) mass spectrometer with 10 ppm mass accuracy in both MS and MS/MS modes [21]. Peptides were first tentatively assigned using our home-buildt LC–MALDI-MS search engine, which uses MAss and Retention Time (sMART) as identification constraints [22]. Peptides' identity was confirmed by MS/MS measurements. Fraction number was used as a measure of peptide's retention time. If the full intensity of a peak was contained in a single fraction, the peak was assigned a retention time equal to the fraction number. However, if that peak's signal was distributed between two (three) consecutive fractions, the retention time assigned was the intensity weighted average of the fraction numbers. Retention times in minutes for all three different chromatographic conditions are shown in Appendix B (both for non-modified and for −17.03 Da species).

A QStar Elite mass spectrometer (Applied Biosystems, Foster City, CA) was used in standard MS/MS data-dependent acquisition mode for nano LC–ESI-MS/MS. One second survey MS spectra were collected (m/z 300–1500) followed by three MS/MS measurements on the most intense parent ions (80 counts/s threshold, +2 to +4 charge state, m/z 100–1500 mass range for MS/MS). Previously targeted parent ions were excluded from repetitive MS/MS acquisition for 60 s (±50 mDa mass window). Raw spectra files were treated using the standard Mascot.dll script (Analyst QS2.0) to generate text files in Mascot generic file format. Protein identification was performed using the X!Tandem (GPM) search engine. Standard QTOF settings were used for the search: 100 ppm and 200 ppm mass tolerance for parent and fragment ions, respectively. Retention times of identified peptides were assigned manually.

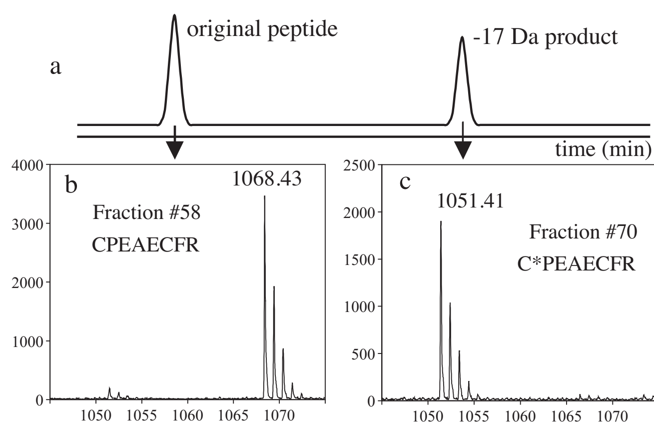

Fig. 2. The separation of non-modified and modified species in RP-HPLC. (a) Schematic representation of RP-LC–UV chromatogram; (b) and (c) MALDI-MS spectra of fractions containing the peptide CPEAECFR and products of its degradation.

3. Results and discussion

3.1. The collection of retention data sets for modified peptides

Fractions for peptide identification using off-line MALDI-MS were collected following peptide separation by micro-RP-LC by mixing the effluent with MALDI matrix and robotically spotting onto a stainless steel target at regular intervals. The resulting fraction number was used as the equivalent of a retention time for the construction of the peptide retention databases. Fig. 2a shows schematically the RP-LC–UV chromatogram of a peptide sample with N-terminal Glu or IAM-Cys, which contains both non-cyclized and cyclized species. Two analytes were injected into the RP-HPLC system, separated due to the difference in hydrophobicity, and detected in the same run. Representative MS spectra for the CPEAECFR peptide from Human α5 Integrin found in different fractions of off-line LC–MALDI-MS run are shown in insets b and c. The distinct −17.03 Da mass shift and ability to confirm a peptide's identity by MS/MS allowed us to unambiguously assign retention times for over 200 non-cyclized–cyclized peptide pairs for each of the conditions studied (Appendix B).

3.2. Chromatographic behavior of peptides with N-terminal cyclization under various RP-HPLC conditions

N-terminal cyclization leads to the loss of the positively charged hydrophilic amino group, which causes an increase in peptide hydrophobicity and retention in RP systems. This increase is a function of the chromatographic conditions used (pH, ion-pairing modifier) and peptide sequence. Fig. 3 shows the dependence of retention time shift on peptide molecular weight for all three eluent conditions under investigation. A number of conclusions can be drawn based on these results:

(i) N-terminal cyclization always increases retention in RP-HPLC; this increase is larger for N-terminal IAM-Cys compared to Gln. An average increase of 5.9, 4.2, and 1.8 min for Cys and 5.5, 3.4,

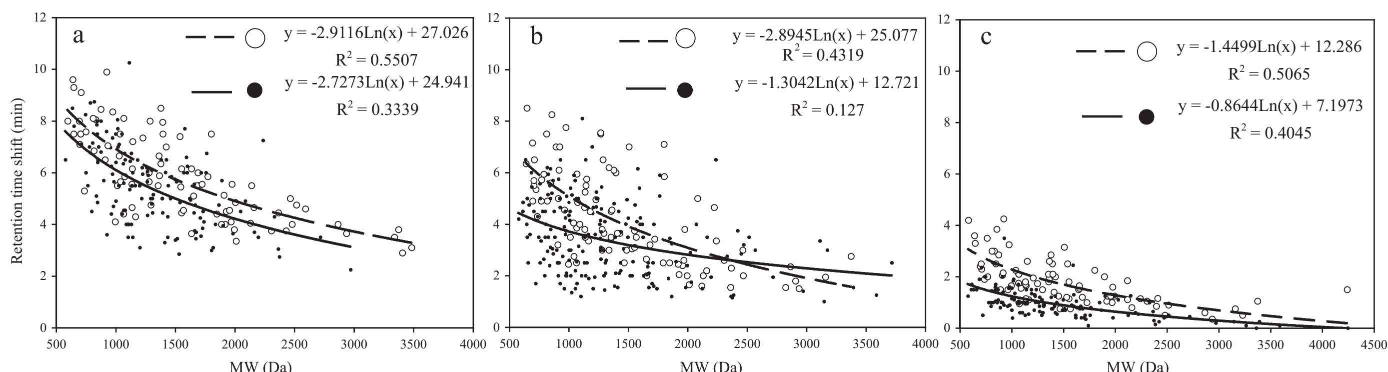

Fig. 3. Retention time shift (min) observed upon N-terminal cyclization depending on molecular weight. ● – Gln-terminated peptides; ○ – IAM-Cys – terminated peptides; eluent conditions: (a) FA; (b) TFA; and (c) pH 10. Note that in panel (b) (TFA) the trend-curves for Gln terminated peptides and IAM-Cys terminated peptides cross at approximately 2300 Da MW but this of no significance, and is an artifact of a low population of sampled peptides at higher mass.

and 1 min for Gln was found for FA, TFA and pH 10 conditions, respectively.

(ii) The amplitude of the shift decreases with peptide length; alteration of one residue gives a larger effect for small species.

(iii) The effect of this modification decreases in the following order FA > TFA ≫ pH 10. Ion-pairing formation is an important part of the RP separation process, and therefore, the hydrophobicity of the ion-pairing modifier plays an important role. Trifluoroacetate is a more hydrophobic counter-ion compared to formate, and its removal due to the loss of N-terminal charge causes a smaller effect. The magnitude of retention shift is 2.2–10.2 min for FA and 1.0–8.5 min for TFA. Since amino groups are not protonated at pH 10, an alteration in peptides' retention under these conditions is minimal; there is no change in peptide charge due to cyclization.

(iv) The effect of N-terminal cyclization is composition and sequence specific. As shown in Fig. 3, in addition to some dependence from peptide size (molecular weight), there are additional factors, which cause a significant spread in observed retention shifts (discussed in Section 3.3). Therefore, the model for retention prediction of N-terminally cyclized peptides must consider sequence specific contributions.

3.3. Sequence specific factors affecting retention of modified peptides

Some of the advanced retention prediction models, including SSRCalc [18,23], use a different retention coefficient for an amino acid dependent on its position within the peptide chain. N-terminal pyroglutamic acid and pyro-cmC are unique in this regard since their position is fixed. Therefore, these residues will have only one value of retention coefficient, and all the sequence specific factors will be determined by position and nature of neighboring amino acids. Table 1 shows typical examples of Gln-terminated peptides (this can be done for Cys-terminated as well) with respective retention time increases for all three chromatographic conditions. Analysis of the large number of sequences (Appendix B) allows segregating peptides into different groups and providing a possible explanation for such behavior:

- Group 1 (small change in retention): these peptides (1–3 in Table 1) feature hydrophilic residues in positions 2–5.
- Group 2 (large change in retention): these peptides (4–6 in Table 1) feature hydrophobic residues in positions 2, 3 and any residues in positions 4, 5.
- Group 3 (intermediate alteration of retention): these peptides (7–9 in Table 1) feature moderately hydrophobic residues in positions 2, 3 and hydrophilic in 4, 5.
- Group 4 (intermediate alteration of retention): similar to the first group, these peptides (10–12) have hydrophilic residues in positions 2 and 3 and hydrophobic ones in positions 4 and 5.

Behavior of the first three groups of peptides can be explained from the point of view of ion-pairing formation, which involves N-terminal residues at acidic pH. During development of the first SSRCalc model [4], we showed that the intrinsic hydrophobicity of N-terminal residues is affected by the ion-pairing formation at the terminal amino group. This effect is very profound for the most hydrophobic residues. This prompted us to employ separate sets of retention coefficients for amino acids in positions 1, 2, and 3 compared to internal positions. Removal of the positively charged N-terminal group due to cyclization leads to elimination of the ion-pairing "shield," and makes hydrophobic residues in positions 2 and 3 more accessible to interact with the C18 phase. The smallest effect was observed for peptides with a hydrophilic N-terminus (peptides 1–3), and an intermediate effect for the cases of intermediate hydrophobicity in Group 3.

Group 4 of peptides in Table 1 represent the special case in which positions $(i+3)$ and $(i+4)$ from the N-terminal amino acid are occupied by hydrophobic residues. We propose a possible mechanism of this effect based on the long-range influence of ion-pairing formation on the stabilization of amphipathic helical structures [6] upon the interaction with C18 phase. The helical-wheel projections of these species, both for non-cyclized and for cyclized forms, are shown in Fig. 1S Appendix C. The influence of an ion-pairing agent could decrease the intrinsic hydrophobicities for Leu (5) and Phe (4) for QSGFLSQMWIGDK (positions $(i+3)$ and $(i+4)$ from the N-terminal amino acid), and Val (5) and Leu (4) for QTALVELVK (Fig. 1S). This in-turn will affect the ability of these residues to participate in hydrophobic interactions with the C18 phase using the hydrophobic face of the amphipathic helix, as shown in Fig. 1S. This is just one of the possible explanations of observed effects. Further proof of it could be obtained using a study of chromatographic behavior for custom designed peptide sequences.

3.4. Incorporation of retention features of modified species into retention prediction models

Fig. 4a–c shows the retention time prediction using corresponding versions of SSRCalc for all three retention data sets – all containing respective pairs of original and modified peptides. Since the retention prediction was not adjusted to take into account the influence of N-terminal cyclization, t_R *vs. HI* plots for modified peptides are shifted up indicating increased retention times, where *HI* – predicted peptide hydrophobicity expressed in hydrophobicity index (or acetonitrile %) units [19]. As noticed in Section

Table 1
Retention time increase (min) for modified N-terminal Gln peptides under various eluent conditions and 0.75% acetonitrile per minute gradient.

Group	Peptide sequence	Retention time increase (min)		
		0.1% TFA	0.1% FA	pH 10
1	QEPERNECFLQHK (albumin, human)	1.6	3.55	0.35
	QNCDQFEK (albumin, bovine)	1.7	4.55	1.05
	QDGSVDFGR (fibrinogen beta chain, human)	2.1	4.7	0.9
2	QFVSSSTTVNR (fibrinogen alpha chain, bovine)	4.65	5.95	1.25
	QLLTPLR (beta-galactosidase, *E. coli*)	5.5	7.85	1.6
	QVLLHQQAK (lactotransferrin, human)	5	6.35	1.7
3	QAYPNLCQLCK (lactotransferrin, bovine)	2.1	4.45	0.75
	QGFGNIATNAEGK (fibrinogen beta chain, bovine)	2.5	5	0.9
	QEINEENVIVK (provicilin B, garden pea)	2.6	5.16	0.9
4	QSGLYFIKPLK (fibrinogen gamma chain, human)	3	5	1.1
	QTALVELVK (albumin, human)	4.85	7.5	1.55
	QSGFLSQMWIGDK (beta-galactosidase, *E. coli*)	3.5	5.8	1.05

3.2, the amplitude of the shift decreases in the following order: FA > TFA ≫ pH 10. Decrease in R^2-value correlations for modified peptides also indicates the sequence specific character of this effect. If modification would cause an identical retention increase independent from the peptide sequence, corresponding t_R *vs. HI* plots should shift up with no change in prediction accuracy. Contrary to this, R^2-values decreased 0.966–0.925, 0.981–0.956 and 0.979–0.973 for FA, TFA and pH 10 models, respectively. As we noted before, N-terminal cyclization at pH 10 conditions does not result in alteration of peptide charge state, therefore sequence-specific effects are not significant.

Developing a retention prediction model for modified species is functionally equivalent to adjusting it so that the resulting t_R *vs. HI* plots will have the same slope, intercept, and ideally the same R^2-value correlations compared to non-modified species. Fig. 4a–c shows that in all three cases the slopes of the dependencies are very close for modified and non-modified peptides. Therefore, as the first approximation, the models can be adjusted by adding a fixed hydrophobicity value to all calculated HI of modified peptides. These values were found to correspond to 4.4, 3.2, and 1.4 *HI* units (or % acetonitrile) for Cys-IAM, and 4.1, 2.6, and 0.75 *HI* units for Gln under FA, TFA and pH 10 conditions, respectively. This, however, will not take into account sequence specificity and decrease in the resulting R^2-value correlation for the combined data sets.

Operational adjustments of retention prediction models for modified species can be done in two different ways:

(i) Through adjusting hydrophobicity values calculated for parent species. This suggests creating a sequence specific model for predicting retention time shifts (as in Fig. 3). When a peptide sequence with a specified modification is submitted for calculation, the predictor will calculate the hydrophobicity for the non-modified one and adjust it using the predicted shift.
(ii) Through "*de novo*" calculation of hydrophobicity values for modified peptides. This suggests creating a prediction model,

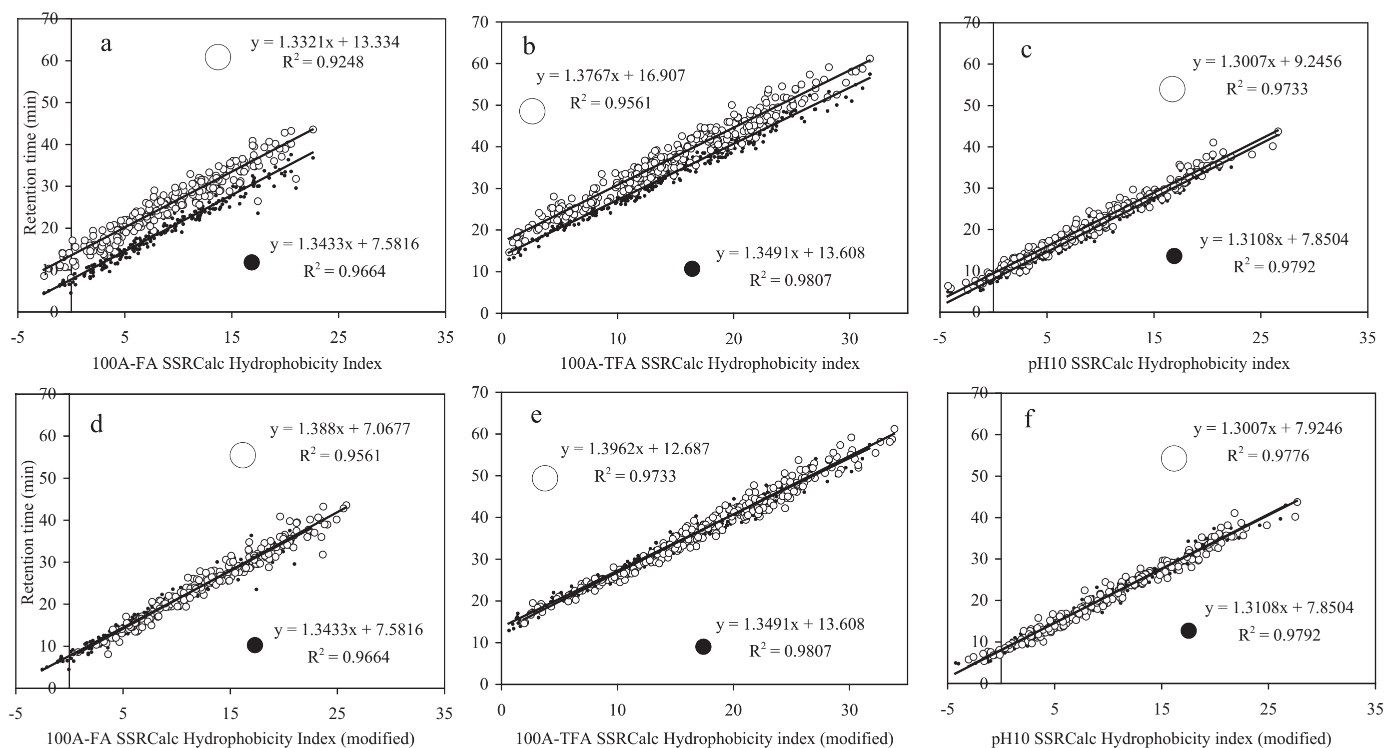

Fig. 4. SSRCalc retention time prediction before (a–c) and after (d–f) model adjustment for N-terminal cyclization. ● – non-modified species; ○ – N-terminally modified analogs; eluent conditions: (a) and (d) – formic acid; (b) and (c) – TFA, (c) and (f) – pH 10.

Table 2
Optimized retention coefficients for 100 Å-TFA SSRCalc model modified to accommodate N-terminal cyclization.

Residue	Modified SSRCalc			Original SSRCalc			
	R_{c1}	R_{c2}	R_{c3}	R_{c1}	R_{c2}	R_{c3}	R_c
W	–	15.51	14.59	11.38	12.61	14.36	13.12
F	–	12.31	12.02	7.70	9.24	11.42	11.34
L	–	10.12	9.86	5.87	7.50	9.25	9.44
I	–	8.72	8.19	5.06	6.70	7.68	7.86
M	–	7.16	6.46	4.51	5.40	5.76	6.57
V	–	4.89	5.06	2.45	3.37	4.74	4.86
Y	–	6.34	5.72	4.92	5.13	5.44	5.40
C	C^* 6.38	0.14	0.52	1.14	0.18	0.85	0.04
P	–	2.33	2.28	1.28	2.20	2.59	1.62
A	–	0.68	1.17	0.10	0.12	1.10	1.11
E	–	0.64	0.80	1.02	0.07	0.51	1.08
T	–	0.57	0.78	1.26	0.47	0.95	0.48
D	–	0.26	0.26	1.61	0.57	0.61	−0.22
Q	Q^* 3.65	−0.39	−0.17	0.33	−0.16	0.12	−0.53
S	–	−0.24	−0.04	1.10	−0.09	0.19	−0.33
G	–	0.02	0.27	0.77	0.32	0.66	−0.35
R	–	−1.92	−3.13	−0.96	−0.77	−3.28	−2.58
N	–	−0.94	−1.10	0.87	−0.24	−0.72	−1.44
H	–	−2.43	−2.97	−1.61	−1.16	−2.63	−3.04
K	–	−1.29	−3.16	−0.2	−2.48	−2.57	−3.53

Q^*, C^* – retention coefficients for N-terminal pyroglutamic acid and pyro-cmC.

which takes into account the new residues – N-terminal pyroglutamic and pyro-cmC acids in our case.

We used the second scenario to modify our SSRCalc models. The SSRCalc algorithm is driven by a number of conditionally performed strings and numeric operations. When modified peptides are submitted for calculations, the program performs the following steps:

(1) It uses a separate retention coefficient value, R_{c1}, for the modified residue (denoted as Q^* and C^* in Table 2). These values are assigned only for position 1 as the location of these residues is pre-determined. The values of individual R_{c1} coefficients for Q^* and C^* are higher compared to R_{c1} of Gln and IAM-Cys in the original SSRCalc-TFA model reflecting an overall increase in retention.
(2) SSRCalc uses separate sets of retention coefficients for all residues in positions 2 and 3: R_{c2} and R_{c3} in Table 2. Since the N-terminal ion-pairing effect is removed for modified species, the retention coefficients of hydrophobic residues in positions 2 and 3 are increased compared to non-modified peptides. The observed effect for hydrophilic residues or amino acids of intermediate hydrophobicity is not that profound (Table 2). It is interesting to note that optimized values of R_{c2} and R_{c3} are higher than R_c – retention coefficients for internal positions. This highlights once again the uniqueness of terminal positions within the peptide; they are readily exposed to interact with the C18 phase compared to the internal hydrophobic residues.

Fig. 4d–f shows retention time prediction using a corrected version of SSRCalc models for all three retention data sets. To highlight the impact of the correction introduced, the correlation data are still provided separately for non-modified and modified peptides. Improvements in R^2-value correlation for modified species clearly indicate the applicability of our approach and ability to take into account sequence specific features of peptide retention. The effect of sequence specific corrections at pH 10 retention conditions was found to be less profound. The influence of this modification on the stabilization of amphipathic helical structures has yet to be implemented in the model, and likely will provide further improvements.

3.5. N-terminal cyclization of Glu

Similar to the loss of ammonia in the reaction shown in Fig. 1a, N-terminal Glu can lose water yielding pyroglutamic acid. It is conceivable that the retention time prediction for modified peptides will be identical, independent of the original residue: Gln or Glu. It was shown that cyclization of N-terminal Glu is a much slower reaction [24]. One should not expect to observe this modification in freshly prepared digests, unless this was a N-terminal residue in an intact protein sequence. It was of interest to verify this using our LC–ESI-MS data. Forty Glu-terminated peptides were identified in those two mixtures, and none of them exhibited significant amounts of −18.01 Da products with higher retention. Contrary to that, the modified peptides were often observed co-eluting with the parent ion, and identified using standard MS/MS search engines like X!Tandem [25] or Protein Pilot [26]. These are correct identifications showing N-terminal loss of water. The parent −18.01 Da ion, however, is formed during the ESI ionization process rather than before separation due to chemical reaction in solution. To confirm this further we performed LC–ESI-MS analysis on a GluFib (EGVNDNEEGFFSAR) peptide solution after 1 week storage at room temperature. Two doubly charged ions with m/z 776.83 are observed in addition to the original one at 785.84 (results not shown). One of them corresponds to the "in-source" loss of water and "co-elutes" with non-modified GluFib, while another 776.83 ion is "in-solution" formed peptide terminated by pyroglutamic acid with higher RP-HPLC retention. Overall, N-terminal cyclization on Glu residues will not be an issue for LC–MS analysis of freshly prepared digests.

4. Conclusions

We studied in detail RP chromatographic behavior of large sets of tryptic peptides carrying N-terminal pyroglutamic acid and pyro-cmC. Analysis of retention shifts between modified and non-modified species have allowed us to establish the sequence specific features responsible for the shift variations, and develop a peptide retention prediction model to accommodate these changes. It is interesting to note that the detailed study of the same modification has illuminated a variety of sequence specific factors, which effect peptide retention in reversed-phase systems. Development of accurate peptide retention prediction models that can accommodate massive proteomics-derived data became possible following discovery of these effects. Knowing the chromatographic properties of post-translationally/chemically modified peptides will facilitate the majority of applied proteomic workflows. The ability to predict not only mass alteration, but the retention shift for a particular modification will be extremely beneficial for detailed fingerprinting of protein sequences and directed "hypothesis-driven" proteomic studies. Additional studies encompassing further abundant protein/peptide modifications are in progress.

Acknowledgements

This work was supported in part by Grants from the Technology Transfer Office at the University of Manitoba and the Natural Sciences and Engineering Research Council of Canada (O.V.K.).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.chroma.2011.05.079.

References

[1] M.T.W. Hearn (Ed.), HPLC of Proteins, Peptides and Polynucleotides. Contemporary Topics and Applications, Wiley, New York, 1991.
[2] C.T. Mant, R.S. Hodges, HPLC of Biological Macromolecules, Marcel Dekker, New York, 2002.
[3] R.A. Houghten, S.T. DeGraw, J. Chromatogr. 386 (1987) 223.
[4] O.V. Krokhin, R. Craig, V. Spicer, W.E. Ens, K.G. Standing, R.C. Beavis, J.A. Wilkins, Mol. Cell. Proteomics 3 (2004) 908.
[5] D.C. Guo, C.T. Mant, R.S. Hodges, J. Chromatogr. 386 (1987) 205.
[6] T.J. Sereda, C.T. Mant, R.S. Hodges, J. Chromatogr. A 695 (1995) 205.
[7] K. Gevaert, J. Van Damme, M. Goethals, G.R. Thomas, B. Hoorelbeke, H. Demol, L. Martens, M. Puype, A. Staes, J. Vandekerckhove, Mol. Cell. Proteomics 1 (2002) 896.
[8] D. Guo, C.T. Mant, A.K. Taneja, J.M.R. Parker, R.S. Hodges, J. Chromatogr. 359 (1986) 499.
[9] O.V. Krokhin, W. Ens, K.G. Standing, Rapid Commun. Mass Spectrom. 17 (2003) 2528.
[10] O.V. Krokhin, M. Antonovici, W. Ens, J.A. Wilkins, K.G. Standing, Anal. Chem. 78 (2006) 6645.
[11] P. Liu, C.L. Feasley, F.E. Regnier, J. Chromatogr. A 1047 (2004) 221.
[12] J. Kim, K. Petritis, Y. Shen, D.G. Camp 2nd., R.J. Moore, R.D. Smith, J. Chromatogr. A 1172 (2007) 9.
[13] K. Gevaert, P. Van Damme, L. Martens, J. Vandekerckhove, Anal. Biochem. 345 (2005) 18.
[14] K. Gevaert, B. Ghesquiere, A. Staes, L. Martens, J. Van Damme, G.R. Thomas, J. Vandekerckhove, Proteomics 4 (2004) 897.
[15] K. Gevaert, M. Goethals, L. Martens, J. Van Damme, A. Staes, G.R. Thomas, J. Vandekerckhove, Nat. Biotechnol. 21 (2003) 566.
[16] K.F. Geoghegan, L.R. Hoth, D.H. Tan, K.A. Borzilleri, J.M. Withka, J.G. Boyd, J. Proteome Res. 1 (2002) 181.
[17] K.M. Khandke, T. Fairwell, B.T. Chait, B.N. Manjula, Int. J. Pept. Protein Res. 34 (1989) 118.
[18] O.V. Krokhin, Anal. Chem. 78 (2006) 7785.
[19] O.V. Krokhin, V. Spicer, Anal. Chem. 81 (2009) 9522.
[20] V. Spicer, M. Grigoryan, A. Gotfrid, K.G. Standing, O.V. Krokhin, Anal. Chem. 82 (2010) 9678.
[21] A.V. Loboda, A.N. Krutchinsky, M. Bromirski, W. Ens, K.G. Standing, Rapid Commun. Mass Spectrom. 14 (2000) 1047.
[22] O.V. Krokhin, V. Spicer, K.G. Standing, J.A. Wilkins, W. Ens, 54th ASMS Conference on Mass Spectrometry and Allied Topics, Seattle, USA, Poster, 2006.
[23] K. Petritis, L.J. Kangas, B. Yan, M.E. Monroe, E.F. Strittmatter, W.J. Qian, J.N. Adkins, R.J. Moore, Y. Xu, M.S. Lipton, D.G. Camp 2nd., R.D. Smith, Anal. Chem. 78 (2006) 5026.
[24] D. Chelius, K. Jing, A. Lueras, D.S. Rehder, T.M. Dillon, A. Vizel, R.S. Rajan, T. Li, M.J. Treuheit, P.V. Bondarenko, Anal. Chem. 78 (2006) 2370.
[25] R. Craig, R.C. Beavis, Bioinformatics 20 (2004) 1466.
[26] I.V. Shilov, S.L. Seymour, A.A. Patel, A. Loboda, W.H. Tang, S.P. Keating, C.L. Hunter, L.M. Nuwaysir, D.A. Schaeffer, Mol. Cell. Proteomics 6 (2007) 1638.