Food Chemistry 173 (2015) 847–854

Contents lists available at ScienceDirect

Food Chemistry

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

Improved short peptide identification using HILIC–MS/MS: Retention time prediction model based on the impact of amino acid position in the peptide sequence

Solène Le Maux, Alice B. Nongonierma, Richard J. FitzGerald *

Department of Life Sciences and Food for Health Ireland (FHI), University of Limerick, Limerick, Ireland

ARTICLE INFO

Article history:
Received 11 June 2014
Received in revised form 4 October 2014
Accepted 18 October 2014
Available online 27 October 2014

Keywords:
UPLC–HILIC
Di-, tri- and tetrapeptides
Retention time prediction
Amino acid coefficients
N-, C-terminal

ABSTRACT

Short peptides can have interesting beneficial effects but they are difficult to identify in complex mixtures. We developed a method to improve short peptide identification based on HILIC–MS/MS. The apparent hydrophilicity of peptides was determined as a function of amino acid position in the sequence. This allowed the differentiation of peptides with the same amino acid composition but with a different sequence (homologous peptides). A retention time prediction model was established using the hydrophilicity and peptide length of 153 di- to tetrapeptides. This model was proven to be reliable (R^2 = 0.992), it was validated using statistical methods and a mixture of 14 synthetic peptides. A whey protein hydrolysate was analysed to assess the ability of the model to identify unknown peptides. In parallel to milk protein database and *de novo* searches, the retention time prediction model permitted reduction and ranking of potential short peptides, including homologous peptides, present in the hydrolysate.

1. Introduction

Many studies have reported the health benefits of short food protein-derived peptides, these include peptides with antioxidant, antidiabetic and antihypertensive properties (Kim et al., 2013; Matsui et al., 2002; Memarpoor-Yazdi, Mahaki, & Zare-Zardini, 2013; Power, Jakeman, & FitzGerald, 2013). One of the most successful strategies for the production of bioactive peptides has been via the hydrolysis of proteins. However, generating peptides via proteolysis/peptidolysis results in complex combinations of peptides having different masses and physicochemical properties. Therefore, identification of bioactive peptides within complex peptide mixtures requires their separation prior to sequence characterisation.

Classical chemical identification methods such as the Edman approach offer a high level of precision but require isolation of peptides prior to analysis. Strategies commonly used to achieve isolation and subsequent characterisation employ chromatographic techniques coupled to mass spectrometry (MS), and in particular liquid chromatography (LC) coupled to tandem MS (MS/MS). LC–MS/MS allows peptide isolation, determination of their molecular mass (M_w) and fragmentation, providing the data necessary for peptide sequencing. However, while this strategy has been successfully used for longer peptides, short peptides may be difficult to identify due to the challenge in matching a unique peptide sequence to the targeted M_w. This is exemplified by the fact that one M_w can correspond to several peptides with different sequences as well as peptides with the same amino acid composition but in a different order (homologous composition peptides). Moreover, because of the low amount of fragment peaks generated by MS/MS, the use of protein database and *de novo* searches to determine short peptide sequences (less than 5 amino acids) can be problematic as a single short peptide may correspond to a number of potential peptide sequences.

Chromatographic separation of peptides provides a useful source of information on each peptide; it can consequently be used to improve peptide identification. For instance, the retention time on different chromatographic matrices is dependent on peptide and column properties, and the chromatographic conditions used during separation (Aguilar & Hearn, 1996; Zou, Zhang, Hong, & Lu, 1992). The retention time observed in reversed phase high pressure liquid chromatography (RP-HPLC) is directly related to the apparent hydrophobic character of the analysed molecule. Consequently, some algorithms have been developed to link the apparent hydrophobicity of peptides to their retention time. Regression curves to predict the retention time based on the amino acid sequence have been described in the literature (Meek, 1980; Schweizer et al., 2007). Some authors have highlighted the

* Corresponding author. Tel.: +353 061 202598; fax: +353 061 331490.
E-mail addresses: Solene.LeMaux@ul.ie (S. Le Maux), Alice.Nongonierma@ul.ie (A.B. Nongonierma), dick.fitzgerald@ul.ie (R.J. FitzGerald).

http://dx.doi.org/10.1016/j.foodchem.2014.10.104
0308-8146/

importance of peptide length as well as amino acid sequence in their prediction models (Chabanet & Yvon, 1992; Harscoat-Schiavo et al., 2012). Additionally, the position of an amino acid within the sequence (i.e., at the N- or C-terminal) and its adjacent residues may also influence the peptides retention time (Babushok & Zenkevich, 2010; Kovacs, Mant, Kwok, Osguthorpe, & Hodges, 2006). The majority of the retention time models available utilise RP-HPLC, with only a few employing ion exchange or hydrophilic interaction liquid chromatography (HILIC) (Bouhallab, Henry, & Boschetti, 1996; Harscoat-Schiavo, Raminosoa, Ronat-Heit, Vanderesse, & Marc, 2010; Harscoat-Schiavo et al., 2012; Krokhin & Spicer, 2010, Chap. 13; Petritis et al., 2006; Schlichtherle-Cerny, Affolter, & Cerny, 2003). Despite its common use, RP-HPLC is limited by the lack of retention of polar molecules. Therefore, HILIC can be used as compounds elute in order of decreasing hydrophobicity or increasing polarity. Moreover, utilisation of HILIC can enhance MS sensitivity due to use of highly organic mobile phases which are easily desolvated, resulting in an improved ionisation efficiency and MS response (Shou & Naidong, 2005).

The objective of the present study was to optimise an HILIC separation method coupled to MS/MS, which could not only be utilised to separate, but also to identify short peptides and to differentiate peptides with homologous sequences. To achieve this, determination of the apparent hydrophilicity of each peptide was a key factor for the development of a retention time prediction model. This retention time prediction algorithm was generated using synthetic peptides and was subsequently validated using a known standard peptide mixture. Further confirmation was achieved by analysis of the peptide sequences present in a whey protein hydrolysate fraction using the retention time prediction method.

2. Materials and methods

The definition of a short peptide can differ from one research group to another. Some authors define a short peptide as having a molar mass <1000 Da, whereas others described short peptides as being <15–20 residues (Petritis et al., 2003; Schweizer et al., 2007; Harscoat-Schiavo et al., 2010, 2012). As the identification of peptides by LC–MS/MS is more complex for sequences <5 amino acids, short peptides are considered as being <5 amino acids for the purposes of this manuscript.

2.1. Chemicals, reagents and samples

Acetonitrile (MeCN), trifluoroacetic acid (TFA) and water were all MS grade and purchased from Sigma–Aldrich (St. Louis, MO). The twenty main amino acids were also purchased from Sigma–Aldrich. Amino acids and peptide sequences are described using the one-letter amino acid abbreviation code. Standard peptides (purity $\geqslant$ 95% *w*/*w*) described in Tables 1 and 2 were purchased from Bachem (Bubendorf, Switzerland), Thermo Fisher Scientific (Waltham, MA) or GenScript (Piscataway, NJ). These peptides were selected based on their range of hydrophilicity, peptide size and sequence. Peptides were dissolved at a concentration of 1 g L^{-1} in MeCN:Water:TFA (80:19:1 *v*/*v*/*v*) before being diluted in 100% MeCN to give a working concentration of 0.025 g L^{-1}. A fraction derived from a whey protein hydrolysate was prepared by solid-phase extraction (SPE) as previously described by Nongonierma et al. (2013). Briefly, whey proteins (88.3% *w*/*w* protein; 1% *w*/*w* fat; 2.4 *w*/*w* ash) were hydrolysed using a pancreatic enzyme preparation prior to SPE fractionation using an hydrophobic matrix (StrataX, Phenomenex, Cheshire, UK). The unbound SPE fraction

Table 1
Training set of standard peptides used for the HILIC retention time model: sequence, peptide length, M_w, observed and predicted retention time, difference between the observed and predicted retention time.

Peptide	*N*	$M_w + H$ (Da)	RT_{obs} (min)	RT_{pred} (min)	Diff (min)
GG	2	133.061	19.072 ± 0.165	20.435	1.363
GP	2	173.092	13.356 ± 0.174	15.148	1.792
AP	2	187.108	11.684 ± 0.111	10.790	0.894
GL*	2	189.123	11.489 ± 0.218	12.508	1.019
VA*	2	189.123	9.019 ± 0.159	8.803	0.216
AL*	2	203.139	8.961 ± 0.157	8.150	0.810
GQ*	2	204.098	38.975 ± 0.150	36.065	2.909
QG	2	204.098	27.586 ± 0.141	30.458	2.872
SL*	2	219.134	17.390 ± 0.167	19.461	2.071
DS	2	221.077	44.730 ± 0.318	44.896	0.166
PL*	2	229.155	5.058 ± 0.030	0.983	4.075
LP*	2	229.155	8.296 ± 0.076	8.133	0.164
IP*	2	229.155	8.010 ± 0.086	5.988	2.022
PI*	2	229.155	4.859 ± 0.045	0.170	4.689
NP	2	230.114	24.731 ± 0.600	25.942	1.211
NV	2	232.129	20.243 ± 0.183	23.260	3.017
VN	2	232.129	25.303 ± 0.127	27.669	2.365
FA	2	237.123	9.451 ± 0.052	11.397	1.946
YG*	2	239.103	22.113 ± 0.324	22.508	0.395
SH	2	243.109	48.986 ± 0.206	48.669	0.317
HS	2	243.109	51.906 ± 0.337	47.196	4.710
QP	2	244.129	18.047 ± 0.287	25.171	7.124
LL*	2	245.186	6.164 ± 0.117	5.493	0.671
AR	2	246.156	32.864 ± 0.148	31.093	1.771
RA	2	246.156	34.016 ± 0.192	34.565	0.549
PM	2	247.111	5.971 ± 0.102	4.447	1.524
MP	2	247.111	10.194 ± 0.027	7.077	3.117
SF*	2	253.118	19.413 ± 0.769	20.267	0.854
HP	2	253.130	30.760 ± 0.305	31.264	0.503
GW*	2	262.118	16.448 ± 0.435	16.753	0.305
WG*	2	262.118	16.447 ± 0.462	15.165	1.283
NK*	2	262.140	50.963 ± 0.300	48.301	2.662
FP	2	263.139	9.450 ± 0.060	8.133	1.317
VF	2	265.155	7.058 ± 0.028	3.704	3.353
YS	2	269.113	31.025 ± 0.471	33.154	2.129
HL*	2	269.161	25.037 ± 0.634	28.624	3.587
VR	2	274.187	23.526 ± 0.321	25.841	2.315
QQ	2	275.135	46.131 ± 0.322	46.088	0.043
QK	2	275.171	51.960 ± 0.648	47.530	4.431
KK	2	275.208	59.598 ± 0.449	55.725	3.872
QE	2	276.119	39.150 ± 0.619	39.874	0.724
AW*	2	276.134	13.032 ± 0.153	12.396	0.637
EK*	2	276.155	48.159 ± 0.149	47.870	0.290
KE	2	276.155	47.500 ± 0.547	48.069	0.570
PY	2	279.133	8.264 ± 0.042	10.888	2.624
YP	2	279.134	17.800 ± 0.061	17.221	0.579
FL*	2	279.170	6.914 ± 0.199	5.493	1.421
VY	2	281.150	12.545 ± 0.072	12.804	0.258
TY*	2	283.129	24.309 ± 0.089	30.188	5.880
YT	2	283.129	26.627 ± 0.597	26.937	0.310
KH	2	284.172	58.220 ± 0.032	59.935	1.715
WS*	2	292.129	24.101 ± 0.465	25.811	1.710
PW*	2	302.150	6.386 ± 0.089	5.228	1.158
WP	2	302.150	12.930 ± 0.141	9.878	3.052
WV*	2	304.166	8.948 ± 0.051	7.196	1.751
VW*	2	304.166	9.346 ± 0.092	7.144	2.202
WT*	2	306.145	20.126 ± 0.890	19.594	0.532
MR	2	306.159	27.668 ± 0.910	27.380	0.288
WC*	2	308.106	11.921 ± 0.169	13.041	1.120
CW*	2	308.106	13.768 ± 0.230	15.103	1.334
IW*	2	318.181	8.559 ± 0.067	7.594	0.965
WI	2	318.181	8.596 ± 0.206	6.426	2.170
WL*	2	318.181	8.630 ± 0.119	7.238	1.391
LW*	2	318.181	8.752 ± 0.215	9.738	0.986
WN*	2	319.140	28.813 ± 0.325	32.008	3.196
WD*	2	320.124	27.669 ± 0.582	29.275	1.606
DW	2	320.124	28.657 ± 0.590	30.569	1.912
RF	2	322.187	25.697 ± 0.332	29.467	3.770
WQ*	2	333.156	26.754 ± 0.556	30.795	4.042
WK*	2	333.192	27.787 ± 0.676	32.237	4.450
KW	2	333.192	32.115 ± 0.345	34.972	2.857
EW*	2	334.140	26.126 ± 0.573	27.116	0.990

Table 1 (*continued*)

Peptide	*N*	$M_w + H$ (Da)	RT_{obs} (min)	RT_{pred} (min)	Diff (min)
WE*	2	334.140	25.285 ± 0.568	24.581	0.705
MW	2	336.138	10.111 ± 0.146	8.683	1.429
WM	2	336.138	9.915 ± 0.144	10.703	0.787
YY*	2	345.144	23.656 ± 0.362	24.487	0.831
FW	2	352.166	9.804 ± 0.241	9.738	0.066
WF	2	352.166	9.855 ± 0.218	8.044	1.811
RW*	2	361.198	31.584 ± 0.625	32.906	1.322
WR	2	361.198	28.760 ± 0.542	30.181	1.421
WY*	2	368.160	18.395 ± 0.226	17.144	1.252
YW	2	368.160	16.459 ± 0.576	18.827	2.368
WW*	2	391.176	13.165 ± 0.248	11.483	1.682
GAD	3	262.103	41.737 ± 0.116	39.958	1.780
GGH	3	270.120	50.461 ± 0.841	48.398	2.063
VLG	3	288.192	14.481 ± 0.231	10.038	4.443
YGG	3	296.124	32.952 ± 0.127	33.449	0.497
ALP	3	300.192	11.691 ± 0.363	10.011	1.679
AIV	3	302.207	9.258 ± 0.068	9.978	0.720
MAS	3	308.127	29.497 ± 0.310	30.980	1.483
VPP	3	312.192	9.065 ± 0.045	8.830	0.236
LGE	3	318.166	30.627 ± 0.342	33.703	3.077
CQA	3	321.123	38.967 ± 0.450	38.038	0.929
IPP	3	326.207	7.807 ± 0.025	9.180	1.373
LPP	3	326.207	8.422 ± 0.190	10.849	2.427
PVD	3	330.166	19.222 ± 0.297	20.340	1.118
ITP	3	330.202	19.447 ± 0.521	19.240	0.207
IQA	3	331.198	30.622 ± 0.469	32.194	1.572
ALK	3	331.234	24.446 ± 0.173	27.414	2.968
IPI	3	342.239	7.865 ± 0.133	6.494	1.371
LPL	3	342.239	7.894 ± 0.171	8.795	0.900
VEP	3	344.181	28.069 ± 0.190	27.995	0.074
IIV	3	344.254	7.178 ± 0.093	6.241	0.937
SHC	3	346.118	46.541 ± 0.336	46.114	0.427
QKA	3	346.208	51.603 ± 0.102	48.606	2.997
ECT	3	352.117	41.087 ± 0.027	41.105	0.018
LQP	3	357.213	28.592 ± 0.064	31.322	2.730
IQP	3	357.213	28.419 ± 1.025	29.653	1.234
RGE	3	361.183	55.956 ± 0.227	51.735	4.221
DNQ	3	376.146	66.615 ± 0.094	66.496	0.119
VYV	3	380.218	17.649 ± 0.040	16.099	1.550
HQP	3	381.188	49.594 ± 0.217	49.325	0.269
KHP	3	381.224	52.295 ± 0.385	52.421	0.126
TRN	3	390.210	63.279 ± 0.114	61.824	1.455
LLF	3	392.254	7.301 ± 0.010	6.516	0.785
FLL	3	392.254	7.364 ± 0.020	5.888	1.475
IVY	3	394.234	13.508 ± 0.029	12.740	0.768
HDK	3	399.199	67.003 ± 0.185	68.206	1.203
KKK	3	403.303	71.258 ± 0.144	69.847	1.411
KYP	3	407.229	39.267 ± 0.210	39.845	0.579
KHK	3	412.267	70.164 ± 0.642	69.823	0.341
PWI	3	415.234	7.234 ± 0.021	9.136	1.902
RSR	3	418.252	66.498 ± 0.260	65.723	0.775
IQY	3	423.224	35.889 ± 0.320	35.308	0.581
KYI	3	423.260	34.179 ± 1.167	37.159	2.980
NRH	3	426.221	67.904 ± 0.180	67.628	0.276
LIW	3	431.265	9.575 ± 0.006	11.246	1.672
KYK	3	438.271	57.946 ± 0.402	57.248	0.698
YPY	3	442.197	27.907 ± 0.192	23.578	4.329
MRF	3	453.228	29.918 ± 0.204	30.839	0.921
RRK	3	459.315	68.713 ± 0.210	68.522	0.191
RFF	3	469.256	25.107 ± 1.015	25.019	0.087
WWW	3	577.256	21.787 ± 1.005	17.941	3.846
VLGP	4	385.245	15.340 ± 0.143	18.884	3.544
TSTP	4	405.198	49.774 ± 0.089	50.099	0.325
VRGP	4	428.262	39.426 ± 0.128	40.611	1.185
NSLP	4	430.230	41.028 ± 0.254	40.926	0.101
GDLE	4	433.193	42.509 ± 0.242	44.343	1.834
RGDS	4	434.199	69.342 ± 0.558	68.404	0.938
IPPL	4	439.291	8.739 ± 0.036	10.062	1.323
IPSK	4	444.282	41.826 ± 0.148	45.055	3.229
KVLP	4	456.318	26.341 ± 0.243	25.929	0.412
RANK	4	488.294	70.902 ± 0.790	71.294	0.391
FLQP	4	504.282	27.905 ± 0.970	28.459	0.554
RDMP	4	518.239	48.418 ± 0.096	46.869	1.549
IPQY	4	520.277	36.635 ± 0.516	34.415	2.220
HHMP	4	521.229	46.858 ± 0.224	46.827	0.031
WIQP	4	543.293	31.895 ± 0.336	31.407	0.488
RHKK	4	568.368	80.624 ± 0.121	81.820	1.196
RKKH	4	568.368	82.966 ± 0.297	84.672	1.706
RRKE	4	588.358	76.630 ± 0.567	76.937	0.307
YPYY	4	605.261	32.972 ± 0.956	32.363	0.609
LWMR	4	605.323	27.049 ± 0.400	29.849	2.801

Peptide sequences in bold are part of a homologous peptide pair.
N: peptide length; RT_{obs}: observed retention time average on three different runs ($n = 3$) ± standard deviation; RT_{pred}: predicted retention time; Diff: absolute value of the difference between the observed and predicted retention times.
* Dipeptides injected simultaneously onto the HILIC–MS/MS to highlight the absence of interaction between peptides during the analysis run (see Fig. 1).

was used to validate the retention time prediction model. Each sample was filtered through a 0.20 μm regenerated cellulose filter (VWR International LLC, Radnor, PA) prior to analysis.

2.2. LC–MS/MS analysis

Samples were analysed in triplicate on LC–MS/MS, using a Waters Acquity ultra performance liquid chromatography (UPLC) system (Waters, Milford, MA) coupled to a MicrOTOF-Q II (Quadrupole, Time-of-Flight) mass spectrometer (Bruker Daltonics GmbH, Bremen, Germany). The UPLC system was equipped with a tunable UV detector set at 214 and 280 nm. The mass spectrometer was fitted with an electrospray ionisation (ESI) source used in the positive ion mode. Hystar™ software (Bruker Daltonics GmbH) was used to control both the UPLC and MS systems.

The peptides analysed were eluted by injecting 2 μL of each sample onto an Acquity BEH Amide column (2.1 × 150 mm, 1.7 μm, 130 Å) equipped with an Acquity BEH Amide 1.7 μm vanguard pre-column, both from Waters. The column temperature was maintained at 40 °C. Mobile phase A was MeCN/H_2O/TFA (97:3:0.1, *v*/*v*/*v*) whereas mobile phase B was MeCN/H_2O/TFA (40:60:0.1, *v*/*v*/*v*). A linear gradient from 100% solvent A to 50% solvent A was applied for 120 min at a flow rate of 0.1 mL min^{-1}. The UPLC conditions employed were optimised initially to obtain the optimal separation of a complex milk protein hydrolysate sample.

MS measurements were performed over a 21–600 or 21–700 *m*/*z* acquisition range. Tandem MS was carried out using five automatically selected precursor ions present in the MS scan. MS data were processed on Compass DataAnalysis 4.0 SP5 (Bruker Daltonics).

2.3. Retention time prediction model

Matlab R2013b (The MathWorks Inc., Natick, MA) was used to generate the hydrophilic coefficient of each amino acid and subsequently the retention time prediction models. The training set of 153 standard peptides which was used in the models is described in Table 1. The amino acid hydrophilic coefficients were determined using a genetic evolutionary algorithm modified from Meek (1980). Briefly, each amino acid retention time was used as their initial coefficient. The hydrophilic coefficients of each peptide from the training set was calculated as being the sum of each amino acid coefficients depending on their position in its sequence. A linear regression was performed to predict retention time using both the peptides hydrophilic coefficients or the peptides hydrophilic coefficients and peptide length. One by one, amino acid coefficients were readjusted by iterations of 0.01 in order to obtain the highest correlation coefficient (R^2) of the linear regression.

Table 2
Synthetic peptides used for validation of the HILIC retention time model. The number of possible peptides was based on the M_w determined experimentally by LC–MS/MS with an error of 0.1 Da. Potential peptides having a maximum of 12 min difference between the observed retention time were considered in the model.

Peptide	$M_w + H$ (Da)	RT_{obs} (min)	RT_{pred} (min)	Diff (min)	No. of peptides with the targeted M_w	No. of potential peptides after the model	Rank after the model	No. of potential peptides in milk	No. of potential peptides in milk after the model	Rank in potential milk peptides after the model
VP	215.1	8.1	5.5	2.56	2	2	1	2	2	1
AM	221.1	9.6	11.6	1.97	7	4	4	7	4	4
IL	245.2	6.4	3.3	3.03	6	5	3	6	5	3
VE	247.1	20.1	20.2	0.12	13	7	1	8	7	1
DE	263.1	42.0	43.7	1.68	8	2	1	6	2	1
RP	272.2	30.1	31.3	1.22	8	4	1	5	3	1
LKP	357.3	24.7	32.8	8.15	132	94	89	22	21	11
IPM	360.2	8.9	9.8	0.88	185	22	4	35	18	1
VKE	375.2	45.1	42.2	2.88	462	261	67	50	43	19
KVKE	503.3	60.5	57.3	3.25	1176	615	235	29	23	14
MGG*	264.1	23.3	25.6	2.22	17	9	3	–	–	–
NW*	319.1	26.3	27.5	1.30	99	20	1	–	–	–
VHSP*	439.2	41.3	48.7	7.40	766	560	349	–	–	–
ILDL*	473.3	22.1	26.5	4.40	978	161	74	–	–	–

Diff: absolute value of the difference between the observed and theoretical retention times; RT_{obs}: observed retention time average on three different runs ($n = 3$); RT_{pred}: predicted retention time.
* Peptides not present in the major milk proteins.

Statistical validations of the model were performed to show the accuracy of the fit, its predictive ability, as well as its robustness. The R^2, the F-statistic, the significance (p-value) and the Durbin–Watson test statistical value (DW) were assessed. The model was also evaluated using diagnostic plots such as the Cook's distance as well as residuals analysis plots (histogram, normal probability plot, symmetry plot, residuals versus lagged residuals and residuals versus fitted values). Prediction methods were compared for efficiency using the root mean squared error (RMSE) of the model as well as the root mean squared errors of two cross-validations, leave-one-out and tenfold cross-validations (RMSECV1 and RMSECV10, respectively).

Where appropriate, data and models were compared using Matlab and a one-way analysis of variance (ANOVA) or a covariance analysis (ANCOVA), both with a Tukey's honestly significant difference comparison. A p-value <0.05 was deemed to be statistically significant.

2.4. Identification of short peptides by LC–MS

A set of 14 standard short peptides described in Table 2 and a whey protein hydrolysate fraction were used to validate the retention time prediction model. A Matlab script was used to generate a list of all possible peptide sequences which corresponded to the targeted M_w within an error of 0.1 Da. The predicted retention time of these potential peptides was then calculated and compared to the observed retention time. All peptides with a time difference >12 min between the predicted and observed retention times were excluded. The presence of the potential peptides in milk was searched using an in-house bovine milk protein database in order to discard non relevant peptides. The database was built from the main bovine milk proteins (β-lactoglobulin, α-lactalbumin, bovine serum albumin, lactoferrin, α_{s1}-, α_{s2}-, β-and κ-casein) using all the available genetic variants given in PubMed. In parallel, a *de novo* search was performed using Peaks Studio 6.0 (Bioinformatics Solutions Inc., Waterloo, Canada). Mass tolerances were set at 0.1 Da for both MS and MS/MS. Variable modifications were fixed at sodium adduct, oxidation of Methionine and phosphorylation of Serine, Threonine and Tyrosine. All potential peptides obtained via the Matlab and Peaks searches were combined and their MS/MS spectra were corroborated using Compass DataAnalysis and Biotools 3.2 (Bruker Daltonics).

3. Results and discussion

3.1. Optimisation of the HILIC conditions

Stock solutions of the synthetic peptides and the protein hydrolysate were dissolved at 1 g L^{-1} in 1% TFA and 80% MeCN. TFA was used due to its properties as a strong acid and an excellent solvent for the majority of peptides (Jao, Ma, Talafous, Orlando, & Zagorski, 1997). The stock solutions were then diluted to their working concentration in 100% MeCN. As the final peptide concentration was in the order of microgram per millilitre, the final concentration of water in the working solution was deemed negligible. Preliminary experiments had shown that a high water concentration in the working solution did not provide acceptable peak shapes (data not shown), being both too wide and asymmetrical as also shown by Ruta, Rudaz, McCalley, Veuthey, and Guillarme (2010) and Chauve, Guillarme, Cléon, and Veuthey (2010). It is therefore important to remain as close as possible to the initial solvent concentration (mobile phase A) in the working solution to provide the best peptide separation.

While formic acid is recommended when performing MS, the peak shapes observed in this study were not sufficiently well defined when this ion pairing agent was used (data not shown). As a result, TFA was employed in all solvents and mobile phases as it offered a better chromatographic resolution; this observation was also made by others (Ruta et al., 2010; Shou & Naidong, 2005; Yoshida, 2004). However, TFA is known to suppress the ESI signals of analytes and reduces their sensitivity (Shou & Naidong, 2005). This is due to the fact that being a strong acid, TFA increases the conductivity and surface tension of water, creating instability during ESI, which decreases the MS signal (Eshraghi & Chowdhury, 1993). Nevertheless, a number of authors have used TFA as an ion paring agent under HILIC conditions (Harscoat-Schiavo et al., 2012; Liu, Tweed, & Wujcik, 2009; Martens-Lobenhoffer, Postel, Tröger, & Bode-Böger, 2007). The decrease in MS signal caused by the presence of TFA can be compensated by the fact that the MS intensity in HILIC is higher than in RP-HPLC due to the high content of organic solvent in the mobile phase (Shou & Naidong, 2005; Simon, Enjalbert, Biarc, Lemoine, & Salvador, 2012). When Shou and Naidong (2005) employed HILIC separation with 0.025% TFA, the signal intensity was approximately five times higher than under RP-HPLC conditions using 0.2% formic acid. Following assessment of different concentrations of TFA in the solvents (data

not shown), it was found that the use of 0.1% TFA in the mobile phase gave acceptable MS signals combined with good peak resolution.

3.2. Development of the retention time model

One hundred and fifty-three synthetic di-, tri- and tetrapeptides were selected based on their range of potential retention times and as representatives of sequences containing all the 20 amino acids in N- and C-terminal positions (Table 1). Twenty-five pairs of homologous peptides, mainly dipeptide pairs, were selected in order to observe if the position of the amino acids in the peptide sequence would impact the retention time. Ten of the peptide pairs had retention times which were not significantly different from each other (p-value >0.05). For instance, FW and WF had retention times of 9.804 ± 0.241 min and 9.855 ± 0.218 min, respectively (Table 1). However, it was observed that even if some pairs of peptides did not have significantly different retention times, two separate peaks could be distinguished on a chromatogram. For example, EW and WE with retention times of 26.126 ± 0.573 min and 25.285 ± 0.568 min, respectively (Table 1), showed two different peaks when eluted from the same sample (Fig. 1). Fifteen of the homologous peptide pairs had significantly different retention times (p-value <0.05), the maximum difference being for GQ and QG with a retention time of 38.975 ± 0.150 min and 27.586 ± 0.141 min, respectively (Table 1). Sequence-specific retention differences for dipeptides under HILIC conditions have previously been observed by Alpert (1990) and Schlichtherle-Cerny et al. (2003). This phenomena was explained by the fact that reversing the sequence of peptides could bring the charged area closer, creating an ion pair (Alpert, 1990). This would lead to a decrease of charge and thus hydrophilicity, thus affecting the retention. Some retention time prediction models based on RP-HPLC have also highlighted the importance of the terminal amino acid groups on the retention time (Krokhin, 2006; Tripet et al., 2007). However, Harscoat-Schiavo et al. (2012) did not observe differences in the retention time of their homologous peptide pairs when studied using HILIC–MS. The only common peptide pair between their study and this manuscript, GW/WG, was also observed as coeluted herein. Alpert (1990) indicated that the lack of retention time differences for some homologous peptides may be due to the non-disruption of charge when the sequence was switched.

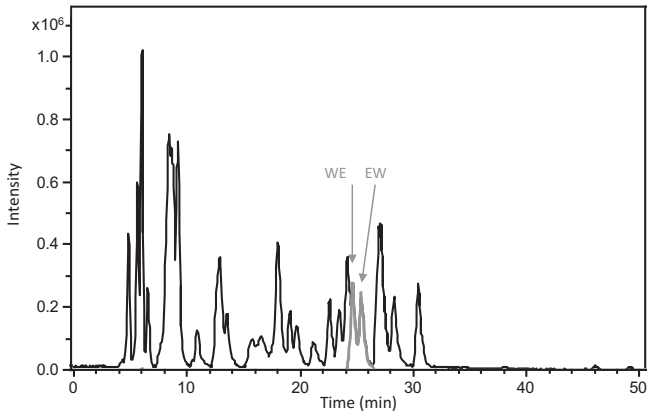

Fig. 1. HILIC–MS/MS base peak chromatogram of a mixture of forty dipeptides (black line). No significant difference in retention time was observed when the dipeptides were injected individually or simultaneously. The dipeptides sequences and retention time are listed in Table 1. The grey line represents the extracted ion chromatogram at 334.140 m/z. The peaks corresponded to WE and EW, in the order of elution, highlighting that two peaks could be distinguished even if these homologous peptides did not have significantly different retention times.

Because differences in retention time could be observed when an amino acid sequence was in a different order, a coefficient was calculated for each amino acid depending on whether it was located at the C-terminal, N-terminal or inside the peptide chain (centre of the peptide sequence). The centre coefficient corresponded to the N-2 residue of tripeptide and to the N-2 and N-3 residues of tetrapeptides. The coefficient of hydrophilicity (H) of each peptide was defined as the sum of each amino acid coefficient of the peptide sequence as previously described (Gilar & Jaworski, 2011; Harscoat-Schiavo et al., 2012; Meek, 1980). The first retention time prediction model developed herein defined the predicted retention time (RT_{pred}) as a linear function of H. Even if this model had an R^2 of 0.987, separate trendlines for dipeptides, tripeptides and tetrapeptides showed significant deviations (p-value <0.05) as highlighted by ANCOVA (see the trendlines on Fig. 2A). The model was therefore modified by inclusion of peptide length (N) to the prediction equation. Several papers have previously used the Napierian logarithm of peptide length ($\ln N$) in their model equation to rectify the deviation observed with different peptide sizes (Gilar & Jaworski, 2011; Harscoat-Schiavo et al., 2012). When $\ln N$ was used in the new model equation herein, no significant deviation was observed for the different peptide lengths (p-value >0.05), which highlighted a more appropriate algorithm (Fig. 2B).

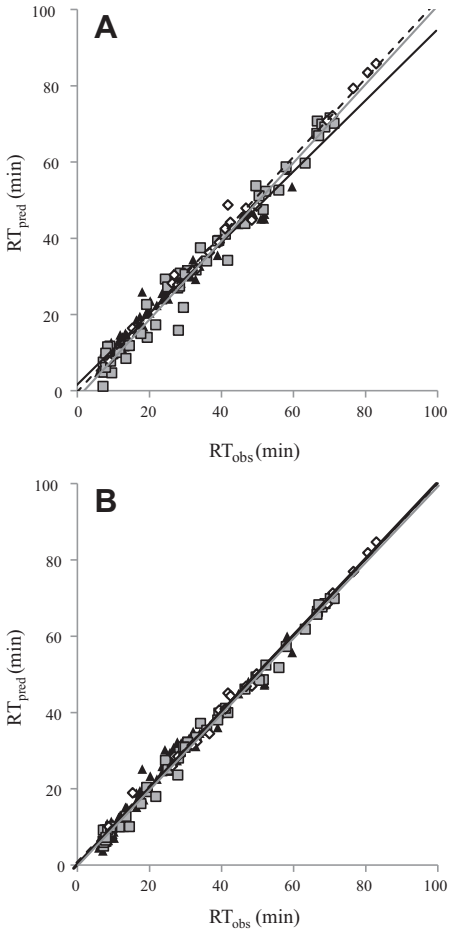

Fig. 2. Plots of predicted versus observed retention times of the 154 training peptides. The observed retention times (RT_{obs}) were the average of three replicates (n = 3). (A) The predicted retention times (RT_{pred}) were a function of the coefficient of hydrophilicity (H) (R^2 = 0.987). (B) The predicted retention times (RT_{pred}) were a function of H and of the Napierian logarithm of the peptide length ($\ln N$) (R^2 = 0.992). Dipeptides, filled triangle; tripeptides, grey box; tetrapeptides, white diamond; dipeptide trendline, black solid line; tripeptide trendline, grey solid line; tetrapeptide trendline, black dashed line.

The model herein was statistically validated as it explained about 99.24% of the variability in the response (R^2). The F-statistic on the regression model was 3.01×10^4 with an extremely low p-value ($p < 5 \times 10^{-324}$), which indicated that the model was significant at the 5% significance level. Moreover, the hypothesis that the residues were not autocorrelated was not rejected as DW had a value of 1.907 which was considered as acceptable according to the Savin and White tables. The Cook's distance and the residual plots of this prediction model were considered as acceptable as no outliers could be defined (data not shown). The RMSE, RMSECV1 and RMSECV10 were 2.207, 5.115 and 5.121, respectively, which revealed the robustness of the model. Consequently, the established amino acid coefficients were defined in Table 3 and the equation for the retention time prediction (RT_{pred}) model was:

$$RT_{pred} = (H - c)/(a + b \times \ln N) \quad (1)$$

The constants a, b and c were optimised through algorithm iterations and were established in the conditions employed herein as 0.488 ± 0.015, 0.668 ± 0.012 and 12.065 ± 0.198, respectively. The prediction of retention time had intervals of 11.6 min and 14.4 min for confidence levels of 95% and 99%, respectively. Therefore, a difference of 12 min between the observed retention time and the predicted retention time was considered as acceptable throughout the rest of the study. Gilar and Jaworski (2011) and Harscoat-Schiavo et al. (2012) considered a maximum uncertainty rate of 20% between the observed and the predicted retention time (standard deviation). Therefore, they tolerated a difference between observed and predicted retention time of a few seconds for poorly retained peptides (very hydrophobic peptides) and of 16 min for peptides eluting after 80 min (very hydrophilic peptides).

The amino acid coefficients of the C-, N-terminal or in the centre, that allowed the determination of H for every peptide, were calculated using the training set of peptides from Table 1 and Eq. (1) (Table 3). Therefore, the H of a peptide corresponded to the contribution of each hydrophilic coefficient of its constitutive amino acid residues, depending on their position in the peptidic chain. Regardless of the residue positions, the basic amino acids had the highest impact on retention time, followed by the acidic amino acids, then the hydrophilic polar uncharged amino acids (amide residues > hydroxyl residues > phenyl residue), prior to the hydrophobic amino acids. This is relatively similar to the order of contribution of each amino acid on peptide apparent hydrophilicity established by Yoshida, Okada, Hobo, and Chiba (2000), Gilar and Jaworski (2011) and Harscoat-Schiavo et al. (2012) when separated on amide columns. However, the results of Yoshida et al. (2000) and Gilar and Jaworski (2011) indicated that the contribution of Proline was more important and the contribution of Tyrosine was less important compared to our study and to that of Harscoat-Schiavo et al. (2012). Using polyhydroxylethyl A and polysulfoethyl A columns, Alpert (1990) studied the contribution of 16 amino acids to the retention of peptides. The basic amino acids were not represented but the order of contribution was similar to that found herein except for a lower contribution of the acidic amino acids. A limited number of studies have calculated the impact of amino acid residue position in the sequence on peptide retention time. Using a reverse-phase C18 column, Krokhin (2006) developed algorithms to define the apparent hydrophobicity of peptides using multiple parameters including amino acid composition, position within the peptide (N- and C-terminal), peptide size and the effect of adjacent residues. For hydrophobic amino acids in peptides of less than nine residues, Krokhin (2006) showed that the amino acid hydrophobic coefficients in the position N-1 were relatively similar to the coefficients in the centre of the peptide chain. Tripet et al. (2007) also defined amino acid hydrophobic coefficients depending on their position in a decapeptide model sequence, using a C18 column. Their results indicated similar coefficients for hydrophobic amino acids located in the C-terminal and in the centre of the chain. In the HILIC study herein, the most hydrophobic residues had a higher hydrophobic impact when positioned in the centre of the sequence than when at N- or C-terminus. This highlighted the fact that hydrophobic residues display large differences in interaction with LC column, based on their position in the peptide sequence and the LC conditions used.

When the retention time model herein was built on Eq. (1), but being only a function of peptide composition without differentiating the position of the residues, it explained about 97.23% of the variability in the response (R^2). The final retention time prediction model developed herein was improved ($R^2 = 0.9924$), by distinguishing the position of the amino acids in the peptide sequences compared to the previously developed HILIC retention time prediction models. Moreover, because it was focussed on di-, tri- and tetrapeptides, this new HILIC model may enhance the identification of short peptides and help the differentiation of peptides with homologous composition.

Table 3
Hydrophilic coefficients of amino acid residue determined when the amino is present at the N-, C-terminal and in the centre of a peptide sequence.

Amino acid	N-t	C-t	Centre
A	12.15	13.27	15.96
R	31.65	29.47	32.68
N	26.55	31.20	33.79
D	29.43	28.61	32.33
C	14.72	13.17	16.72
E	26.14	24.14	28.92
Q	25.82	30.05	30.52
G	16.29	15.19	19.45
H	31.61	35.42	32.30
I	7.59	6.89	4.48
L	7.72	7.66	1.97
K	33.61	31.42	32.33
M	8.62	10.95	4.16
F	9.63	8.43	2.55
P	5.34	10.17	5.52
S	22.90	25.31	31.21
T	23.68	19.40	17.80
W	11.29	11.70	11.00
Y	18.27	17.08	16.95
V	7.16	7.62	2.96

N-t: coefficient of amino acid in N-terminal position; C-t: coefficient of amino acid in C-terminal position; Centre: coefficient of amino acid inside the peptide sequence.

3.3. Validation of the model using standard peptides

A first validation of the model was made by checking if the retention of a peptide could be altered by the presence of several other peptides in the injected sample. A mixture of 40 dipeptides was therefore chosen randomly from the training peptide list and these were injected simultaneously onto the HILIC–MS/MS. None of the retention times were affected by the presence of other peptides, showing that no interaction occurred between peptides during the analysis run (Fig. 1). Because a mixture of peptides did not induce a retention time shift, it was concluded that the prediction model could be used for samples containing a complex mixture of peptides.

Fourteen known synthetic peptides were randomly selected and injected simultaneously onto the HILIC–MS/MS to validate the algorithm developed; none of which were previously used in the training model. Table 2 describes the analysis of this peptide mixture and the retention time prediction model. The number of

possible sequences corresponding to the observed M_w was assessed, followed by the application of the retention time prediction model. Furthermore, the potential peptides that matched the observed M_w and that had a predicted retention time within 12 min of the observed retention time were ranked as a function of their retention time difference (predicted versus observed retention times). The number of potential peptides (including the homologous peptide possibilities) identified using the M_w obtained by MS was low for peptides <300 g mol^{-1}, but was much higher for larger peptides. For example, over one thousand possibilities appeared for the tetrapeptide KVKE (Table 2). The number of potential peptides was considerably lower for peptides with a M_w > 300 g mol^{-1}. When the list of potential sequences was ranked based on their retention time differences, the synthetic peptides injected on HILIC–MS/MS were identified 5 times out of 14 as having the lowest retention time difference (ranked in first position). This highlighted the applicability of the developed retention time prediction model. As ten of the synthetic peptides tested correspond to sequences in bovine milk proteins, a milk peptide database was further used to reduce the number of potential peptides (Table 2). For instance, the M_w of peptide IPM ($M_w + H$ of 360.2 Da) corresponds to 185 possible peptides, but only 22 of these had a predicted retention time within 12 min of the observed retention time. However, the M_w of IPM corresponds to 35 peptide sequences in the milk proteome, of which 18 had a retention time within 12 min of the observed retention time (Table 2). Finally, IPM had the lowest difference between the observed retention time and the predicted retention time (ranked first as seen in Table 2), showing the relevance of the developed retention time prediction model. In some cases, such as for the synthetic peptide VP, the homologous peptide PV was one of the potential peptides that is also present in bovine milk. As its retention time was close to that of VP, it was not possible to rule out PV. One way to distinguish a specific peptide from its homologous sequence(s) may be by using MS/MS fragmentation data as outlined by Schlichtherle-Cerny et al. (2003). These authors could differentiate pairs of homologous dipeptides on the basis of their characteristic fragmentation spectra using HILIC–MS^n. In this study, the twenty-five pairs of homologous peptides had two different MS fragmentation spectra for each peptide (data not shown). This highlights the importance of MS/MS data in the identification of short peptides. To conclude, coupled to the source of the proteins and MS/MS data searches, the prediction model developed herein using HILIC retention time could support peptide identification and help in the differentiation of homologous peptides.

3.4. Validation of the model using a whey protein hydrolysate fraction

A whey protein hydrolysate SPE fraction was analysed with an MS method for short peptide detection (m/z range of 21–600) as only short peptides were used to validate the model. Forty five short peptides were detected and then analysed using their M_w, retention time and MS/MS spectrum. Twenty peptides taken randomly were submitted to data analysis using the HILIC retention time model associated to the protein source and to *de novo* data searches. Table 4 summarises the number of potential sequences that corresponded to the observed M_w. These were then reduced using a milk protein database and the prediction model developed in combination with MS fragmentation analysis. For instance, 462 peptides corresponded to the M_w of the peptide observed at a retention time of 40.3 min. Only 44 of these peptides were present in milk proteins, while only 4 of these matched the criteria of having a predicted retention time with a maximum of 12 min difference from the observed retention time, and an MS/MS spectrum corresponding to the targeted peptide. Of these 4 potential peptides, the peptide with the highest probability based on the fragmentation spectra was EVK.

In conclusion, it was possible to improve short peptide identification in complex hydrolysates, when separated with UPLC–HILIC and using the retention time prediction model established in this study. In some cases, the retention time prediction tolerance of 12 min was too large to allow ruling out certain homologous peptides. However, this approach allows a reduction in the number of possible peptides along with ranking the potential sequences, including homologous peptides. Consequently, it reduces the extent of manual analysis that may be required subsequently to analyse the MS spectra.

Table 4
Peptide identification applied to a whey protein hydrolysate fraction using the HILIC prediction retention time model developed herein. The number of possible peptides was based on the M_w determined experimentally by LC–MS/MS with an error of 0.1 Da. The time difference between the observed and predicted retention times was set at a maximum of 12 min.

RT_{obs} (min)	$M_w + H$ (Da)	No. of peptides corresponding to the M_w	No. of potential peptides in milk	No. of potential peptides after model and *de novo* analysis	Potential peptide sequences (ranked in potential peptides order)
7.2	447.2	1262	4	2	CIVL, LCVL
8.2	203.1	6	6	2	AL, LA
9.9	300.2	21	5	1	IPA
11.6	205.0	7	2	1	W
13.9	260.1	29	9	2	QI, QL
16.2	203.1	6	6	6	PS, LA, SP, IA, AL, AI
21.1	260.1	29	14	4	QI, QL, KI, LK
22.0	173.1	2	2	1	GP
22.2	366.2	90	22	6	LAY, HLP, HPI, PIH, CPF, LSF
25.6	274.2	23	8	1	VR
27.3	428.3	279	9	1	KIPA
28.8	318.2	105	24	4	DAI, IDA, LDA, DLA
30.2	525.3	518	5	1	AIPPK
32.2	329.2	30	4	1	VPN
32.1	347.2	683	31	2	QVT, KVT
34.5	391.2	592	17	1	TVAT
36.8	377.2	439	11	1	STAV
39.4	573.3	296	10	5	IIAEK
40.3	375.2	462	44	4	EVK, EIN, DIQ, VEQ
49.1	421.2	820	7	1	SSLD

RT_{obs}: observed retention time average on three different runs ($n = 3$).

4. Conclusion

UPLC–HILIC conditions were first optimised to allow good separation of peptides within a complex mixture of peptides. A new method to determine the apparent hydrophilicity of short peptides was developed. The originality of this model was to distinguish the impact of each residue on peptide hydrophilicity when the residue was in N-, C-terminal or in the centre of the peptide sequence, allowing for differentiation between homologous peptides. Moreover, the focus of this model was on short peptides by using a training set of short peptides, from di- to tetrapeptides. The algorithm developed herein links short peptide retention time to the apparent hydrophilicity coefficient and the size of the peptide sequence. This retention time prediction model was first validated statistically (cross-validation tests), then using a mixture of known synthetic peptides and finally applied to a whey protein hydrolysate fraction. Consequently, this method allows a more accurate prediction of the amino acid sequence of unknown short peptides, especially by being able to focus on di-, tri- and tetrapeptides, but also by differentiating retention times of homologous peptides. The improvement of knowledge on the retention time prediction of short peptides under the UPLC–HILIC conditions used herein may support the identification of short bioactive peptides in complex mixtures. Combination of retention time models using different separation modes such as HILIC, RP-HPLC, ion-exchange LC and capillary electrophoresis, may further enhance the accuracy of peptide identification (Krokhin, 2006; Harscoat-Schiavo et al., 2010, 2012). Furthermore, the different chromatographic conditions will result in differences in retention times, peak resolutions and MS signal intensities (Shou & Naidong, 2005; Simon et al., 2012). Finally, the analysis of hydrolysates containing larger peptides can also be achieved using separation methods, such as HILIC and RP-HPLC, coupled to MS methods detecting broader m/z ranges.

Acknowledgement

The work described herein was supported by Enterprise Ireland under Grant Number TC2013-0001.

References

Aguilar, M. I., & Hearn, M. T. W. (1996). High-resolution reversed-phase high-performance liquid chromatography of peptides and proteins. *Methods in Enzymology, 270*, 3–26.

Alpert, A. J. (1990). Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds. *Journal of Chromatography A, 499*, 177–196.

Babushok, V. I., & Zenkevich, I. G. (2010). Retention characteristics of peptides in RP-LC: Peptide retention prediction. *Chromatographia, 72*(9–10), 781–797.

Bouhallab, S., Henry, G., & Boschetti, E. (1996). Separation of small cationic bioactive peptides by strong ion-exchange chromatography. *Journal of Chromatography A, 724*(1–2), 137–145.

Chabanet, C., & Yvon, M. (1992). Prediction of peptide retention time in reversed-phase high-performance liquid chromatography. *Journal of Chromatography A, 599*(1–2), 211–225.

Chauve, B., Guillarme, D., Cléon, P., & Veuthey, J.-L. (2010). Evaluation of various HILIC materials for the fast separation of polar compounds. *Journal of Separation Science, 33*(6–7), 752–764.

Eshraghi, J., & Chowdhury, S. K. (1993). Factors affecting electrospray ionization of effluents containing trifluoroacetic acid for high-performance liquid chromatography/mass spectrometry. *Analytical Chemistry, 65*(23), 3528–3533.

Gilar, M., & Jaworski, A. (2011). Retention behavior of peptides in hydrophilic-interaction chromatography. *Journal of Chromatography A, 1218*(49), 8890–8896.

Harscoat-Schiavo, C., Nioi, C., Ronat-Heit, E., Paris, C., Vanderesse, R., Fournier, F., et al. (2012). Hydrophilic properties as a new contribution for computer-aided identification of short peptides in complex mixtures. *Analytical and Bioanalytical Chemistry, 403*(7), 1939–1949.

Harscoat-Schiavo, C., Raminosoa, F., Ronat-Heit, E., Vanderesse, R., & Marc, I. (2010). Modeling the separation of small peptides by cation-exchange chromatography. *Journal of Separation Science, 33*(16), 2447–2457.

Jao, S.-C., Ma, K., Talafous, J., Orlando, R., & Zagorski, M. G. (1997). Trifluoroacetic acid pretreatment reproducibly disaggregates the amyloid β-peptide. *Amyloid, 4*(4), 240–252.

Kim, E. K., Oh, H. J., Kim, Y. S., Hwang, J. W., Ahn, C. B., Lee, J. S., et al. (2013). Purification of a novel peptide derived from *Mytilus coruscus* and in vitro/in vivo evaluation of its bioactive properties. *Fish & Shellfish Immunology, 34*(5), 1078–1084.

Kovacs, J. A., Mant, C. T., Kwok, S. C., Osguthorpe, D. J., & Hodges, R. S. (2006). Quantitation of the nearest-neighbour effects of amino acid side-chains that restrict conformational freedom of the polypeptide chain using reversed-phase liquid chromatography of synthetic model using reversed-phase liquid chromatography of synthetic model peptides with L- and D-amino acid substitutions. *Journal of Chromatography A, 1123*(2), 212–224.

Krokhin, O. V. (2006). Sequence-specific retention calculator. Algorithm for peptide retention prediction in ion-pair RP-HPLC: Application to 300- and 100-Å pore size C18 sorbents. *Analytical Chemistry, 78*(22), 7785–7795.

Krokhin, O. V., & Spicer, V. (2010). Predicting peptide retention times for proteomics. *Current Protocols in Bioinformatics*. Unit 13 14.

Liu, A., Tweed, J., & Wujcik, C. E. (2009). Investigation of an on-line two-dimensional chromatographic approach for peptide analysis in plasma by LC–MS–MS. *Journal of Chromatography B, 877*(20–21), 1873–1881.

Martens-Lobenhoffer, J., Postel, S., Tröger, U., & Bode-Böger, S. M. (2007). Determination of ornithine in human plasma by hydrophilic interaction chromatography–tandem mass spectrometry. *Journal of Chromatography B, 855*(2), 271–275.

Matsui, T., Yukiyoshi, A., Doi, S., Sugimoto, H., Yamada, H., & Matsumoto, K. (2002). Gastrointestinal enzyme production of bioactive peptides from royal jelly protein and their antihypertensive ability in SHR. *The Journal of Nutritional Biochemistry, 13*(2), 80–86.

Meek, J. L. (1980). Prediction of peptide retention times in high-pressure liquid chromatography on the basis of amino acid composition. *Proceedings of the National Academy of Sciences of the United States of America, 77*(3), 1632–1636.

Memarpoor-Yazdi, M., Mahaki, H., & Zare-Zardini, H. (2013). Antioxidant activity of protein hydrolysates and purified peptides from *Zizyphus jujuba* fruits. *Journal of Functional Foods, 5*(1), 62–70.

Nongonierma, A. B., Gaudel, C., Murray, B. A., Flynn, S., Kelly, P. M., Newsholme, P., et al. (2013). Insulinotropic properties of whey protein hydrolysates and impact of peptide fractionation on insulinotropic response. *International Dairy Journal, 32*(2), 163–168.

Petritis, K., Kangas, L. J., Ferguson, P. L., Anderson, G. A., Pasa-Tolic, L., Lipton, M. S., et al. (2003). Use of artificial neural networks for the accurate prediction of peptide liquid chromatography elution times in proteome analyses. *Analytical Chemistry, 75*(5), 1039–1048.

Petritis, K., Kangas, L. J., Yan, B., Monroe, M. E., Strittmatter, E. F., Qian, W.-J., et al. (2006). Improved peptide elution time prediction for reversed-phase liquid chromatography–MS by incorporating peptide sequence information. *Analytical Chemistry, 78*(14), 5026–5039.

Power, O., Jakeman, P., & FitzGerald, R. J. (2013). Antioxidative peptides: enzymatic production, in vitro and in vivo antioxidant activity and potential applications of milk-derived antioxidative peptides. *Amino Acids, 44*(3), 797–820.

Ruta, J., Rudaz, S., McCalley, D. V., Veuthey, J.-L., & Guillarme, D. (2010). A systematic investigation of the effect of sample diluent on peak shape in hydrophilic interaction liquid chromatography. *Journal of Chromatography A, 1217*(52), 8230–8240.

Schlichtherle-Cerny, H., Affolter, M., & Cerny, C. (2003). Hydrophilic interaction liquid chromatography coupled to electrospray mass spectrometry of small polar compounds in food analysis. *Analytical Chemistry, 75*(10), 2349–2354.

Schweizer, M., Chevalot, I., Blanchard, F., Fournier, F., Harscoat-Schiavo, C., Vanderesse, R., et al. (2007). Prediction of short peptides composition by RP-HPLC coupled to ESI mass spectrometry. *Food Chemistry, 105*(4), 1606–1613.

Shou, W. Z., & Naidong, W. (2005). Simple means to alleviate sensitivity loss by trifluoroacetic acid (TFA) mobile phases in the hydrophilic interaction chromatography–electrospray tandem mass spectrometric (HILIC–ESI/MS/MS) bioanalysis of basic compounds. *Journal of Chromatography B, 825*(2), 186–192.

Simon, R., Enjalbert, Q., Biarc, J., Lemoine, J., & Salvador, A. (2012). Evaluation of hydrophilic interaction chromatography (HILIC) versus C-18 reversed-phase chromatography for targeted quantification of peptides by mass spectrometry. *Journal of Chromatography A, 1264*, 31–39.

Tripet, B., Cepeniene, D., Kovacs, J. M., Mant, C. T., Krokhin, O. V., & Hodges, R. S. (2007). Requirements for prediction of peptide retention time in reversed-phase high-performance liquid chromatography: Hydrophilicity/hydrophobicity of side-chains at the N- and C-termini of peptides are dramatically affected by the end-groups and location. *Journal of Chromatography A, 1141*(2), 212–225.

Yoshida, T. (2004). Peptide separation by hydrophilic-interaction chromatography: A review. *Journal of Biochemical and Biophysical Methods, 60*(3), 265–280.

Yoshida, T., Okada, T., Hobo, T., & Chiba, R. (2000). Calculation of amino acid hydrophilicity indices for retention of peptides on amide, diol and silica columns in normal-phase liquid chromatography. *Chromatographia, 52*(7–8), 418–424.

Zou, H., Zhang, Y., Hong, M., & Lu, P. (1992). Effect of column temperature on retention of dipeptide isomers in reversed-phase high performance liquid chromatography. *Journal of Liquid Chromatography, 15*(13), 2289–2296.