Analyst

Online Protein Digestion in Membranes Between Capillary Electrophoresis and Mass Spectrometry

Journal:	*Analyst*
Manuscript ID	AN-ART-01-2023-000106.R1
Article Type:	Paper
Date Submitted by the Author:	27-Feb-2023
Complete List of Authors:	Ryan, Kendall; University of Notre Dame, Department of Chemistry Bruening, Merlin; University of Notre Dame, Chemical and Biomolecular Engineering and Chemistry;

SCHOLARONE™
Manuscripts

Please do not adjust margins

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

ARTICLE

Online Protein Digestion in Membranes Between Capillary Electrophoresis and Mass Spectrometry

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Kendall A. Ryan[1] and Merlin L. Bruening[1,2]

This research employs pepsin-containing membranes to digest proteins online after a capillary electrophoresis (CE) separation and prior to tandem mass spectrometry. Proteolysis after the separation allows the peptides from a given protein to enter the mass spectrometer in a single plug. Thus, migration time can serve as an additional criterion for confirming the identification of a peptide. The membrane resides in a sheath-flow electrospray ionization (ESI) source to enable digestion immediately before spray into the mass spectrometer, thus limiting separation of the digested peptides. Using the same membrane, digestion occurred reproducibly during 20 consecutive CE analyses performed over a 10 h period. Additionally, after separating a mixture of six unreduced proteins with CE, online digestion facilitated protein identification with at least 2 identifiable peptides for all the proteins. Sequence coverages were >75% for myoglobin and carbonic anhydrase II but much lower for proteins containing disulfide bonds. Development of methods for efficient separation of reduced proteins or identification of cross-linked peptides should enhance sequence coverages for proteins with disulfide bonds. Migration times for the peptides identified from a specific protein differed by <~30 s, which allows for rejection of some spurious peptide identifications.

Introduction

This paper describes placement of a pepsin-containing membrane inline after a capillary electrophoresis (CE) protein separation to digest proteins just prior to electrospray ionization (ESI). Proteolysis immediately before mass spectrometry (MS) retains some of the benefits of bottom-up protein analysis, which employs digestion prior to separation of proteolytic peptides and tandem mass spectrometry (MS/MS).[1,2] Analysis of proteolytic peptides is attractive because MS/MS allows extensive peptide sequencing. In standard bottom-up analysis, peptides from all proteins emerge together during proteolysis. After CE or liquid chromatography, the peptides from a given protein appear over a wide range of separation times. Online digestion after separation of intact proteins should still afford the benefit of extensive peptide sequencing, but all the peptides from a given protein will appear

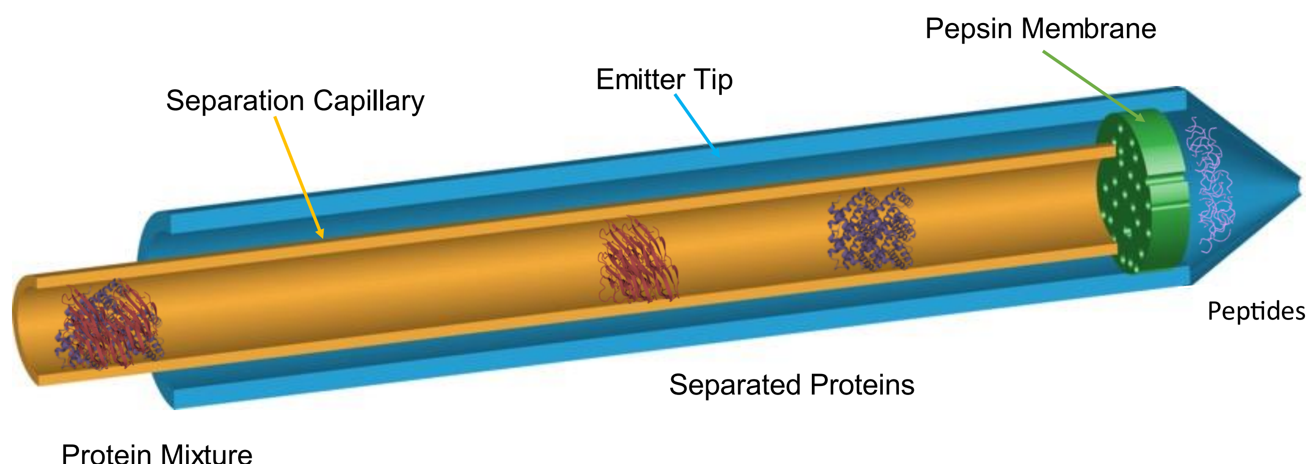

Fig. 1 Cross-section of a separation capillary inside an emitter tip showing placement of a membrane online after protein separation so the peptides from a given protein will appear in a narrow plug.

[1]*Department of Chemistry and Biochemistry*
[2]*Department of Chemical and Biomolecular Engineering*
University of Notre Dame
Notre Dame, IN 46556
Electronic Supplementary Information (ESI) available: Tables of identified proteins; electropherograms, and selected mass spectra. See DOI: 10.1039/x0xx00000x.

Please do not adjust margins

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

in a narrow plug (Fig. 1). Thus, migration time could provide an additional criterion for peptide identification.

Online proteolysis after a separation requires an Immobilized-Enzyme Reactor (IMER). Many studies show that IMERS can streamline proteolysis.[3–5] Compared to in-solution digestion conditions, IMERs have a higher protease density and may show greater resistance to enzyme degradation, more reproducible catalytic activity, and shorter reaction times.[1,6–8] Many IMERs also exhibit limited protease autolysis because immobilized enzymes do not interact with each other. Placement of an IMER in the emitter tip of a CE-MS apparatus enables the separation, digestion, and identification of a mixture of proteins without additional sample preparation related to digestion.

This research uses enzyme-containing microfiltration membranes for digestion after CE. Porous membranes are attractive substrates for IMERs because their limited thickness (100 µm) provides low resistance to flow. More importantly, small diffusion distances in µm-diameter membrane pores facilitate rapid reaction. Immobilization of proteases in a membrane support can provide a high enzyme density within the membrane (up to 60 mg of enzyme per mL of membrane).[9] This high enzyme density enables proteolysis during millisecond residence times in the membrane. In contrast digestion in monoliths requires second or minute residence times.[10,11] Online digestion can occur before separation for a bottom-up analysis,[12–15] but we are particularly interested in digestion after a separation. A few studies used IMERS for online digestion after a separation but included added complications such as online buffer exchange, T-junctions for pH adjustment, and online dilution.[16–19] Our proposed system integrates an IMER into existing equipment with no added complexity and uses a protease that is compatible with native CE separation buffers.

CE can separate proteins with over 100,00 theoretical plates due to the absence of pressure-driven flow.[20] Proteins separate according to their electrophoretic mobility, which depends on their size and charge. At acidic pH, most proteins are positively charged and migrate towards the cathode (the mass spectrometer inlet). Combining separations of proteins in an acidic background electrolyte (BGE) with online digestion requires a protease that is active at low pH, and pepsin is most active around pH 2.[21] Thus, pepsin is ideally suited for digestion after CE separations. Although pepsin is not highly specific, it is useful for proteolysis prior to analysis of modestly complex protein mixtures.[22–24]

Digestion of proteins after CE separations may enable the use of protein migration time as a criterion for peptide matching. A few studies attempted to use peptide elution or migration times as a criterion for their identification,[25–27] and CE outperformed LC with regard to predicting peptide migration or elution times for bottom-up analysis.[28–36] These approaches rely on mathematical prediction of peptide separations, which is particularly difficult for peptides with post-translational modifications.[37] Moreover, separations vary among instruments, especially for CE. Therefore, the addition of elution or migration time as a criterion for peptide identification is not common.

With digestion just prior to ESI, all the peptides from a given protein should enter the mass spectrometer in a narrow band. Therefore, the additional criterion for peptide matching would compare the migration times of peptides from a given protein to each other. This strategy does not depend on the absolute migration time. Using a simple mixture of proteins, this paper executes this strategy, although further work is needed to overcome challenges in either digestion without reduction or separation of reduced proteins.

Experimental

Materials and Reagents

Nylon membranes with a nominal pore size of 1.2 µm and a thickness of 110 µm were acquired from Pall (Port Washington, NY). Poly(sodium 4-styrenesulfonate) (M_W ~ 70 000) (PSS), sodium chloride, hydrochloric acid, acetonitrile, ammonium bicarbonate (ABC), equine myoglobin, bovine serum albumin (BSA), carbonic anhydrase II, α-lactalbumin, bovine insulin, ribonuclease B, LC-MS grade formic acid (FA), tris(2-carboxyethyl)phosphine hydrochloride (TCEP), sodium hydroxide, and LC-MS grade methanol were purchased from Sigma Aldrich (St Louis, MO). ACS grade acetic acid was obtained from VWR (Radnor, PA). Bovine myoglobin was acquired from Innovative Research (Novi, MI), and water was purified using a Millipore Milli-Q Reference system (18.2 MΩ cm). Borosilicate glass tubing with a 1 mm outer diameter (od) and a 0.75 mm inner diameter (id) was purchased from Sutter Instruments (Novato, CA). Fused silica capillaries were obtained from Polymicro (Phoenix, AZ) and have a 50 µm id and a 360 µm od. A mini microfilter assembly and capsules were purchased from IDEX Health and Science (Oak Harbor, WA). All solvents were filtered with a 0.22 µm filter and degassed via sonication before use.

Immobilization of Pepsin

Immobilization of pepsin was performed as previously described.[9,10] A nylon membrane with a 1.2 µm pore size was UV/O_3 cleaned for 10 min prior to modification, and a peristaltic pump circulated solutions through the membrane at a rate of 1 mL/min. The membrane was washed with 10 mL of water, then 10 mL of 0.18 mg/mL PSS in 0.5 M NaCl (adjusted to pH 2.3) was circulated through the membrane for 20 min. The membrane was then washed with 20 mL of 5% FA before circulating 2 mL of 2 mg/mL pepsin in 5% FA through the membrane for 2 hours. Finally, 10 mL of 5% FA was passed through the membrane before drying it under nitrogen and storing it at 4 °C until use. Fluorescence analysis of the pepsin loading solution before and after circulation suggests an immobilized pepsin concentration between 30 and 35 mg per mL of membrane.

CE-MS

A Next 840 CE power supply and autosampler from Prince Technologies (Emmen, Netherlands) provided the electric field and sample injection for intact protein separations. To reduce the electroosmotic flow (EOF), capillaries were coated with linear polyacrylamide (in-house coating) to reduce electro-osmotic flow and protein-wall interactions.[38,39] After coating,

This journal is © The Royal Society of Chemistry 20xx

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

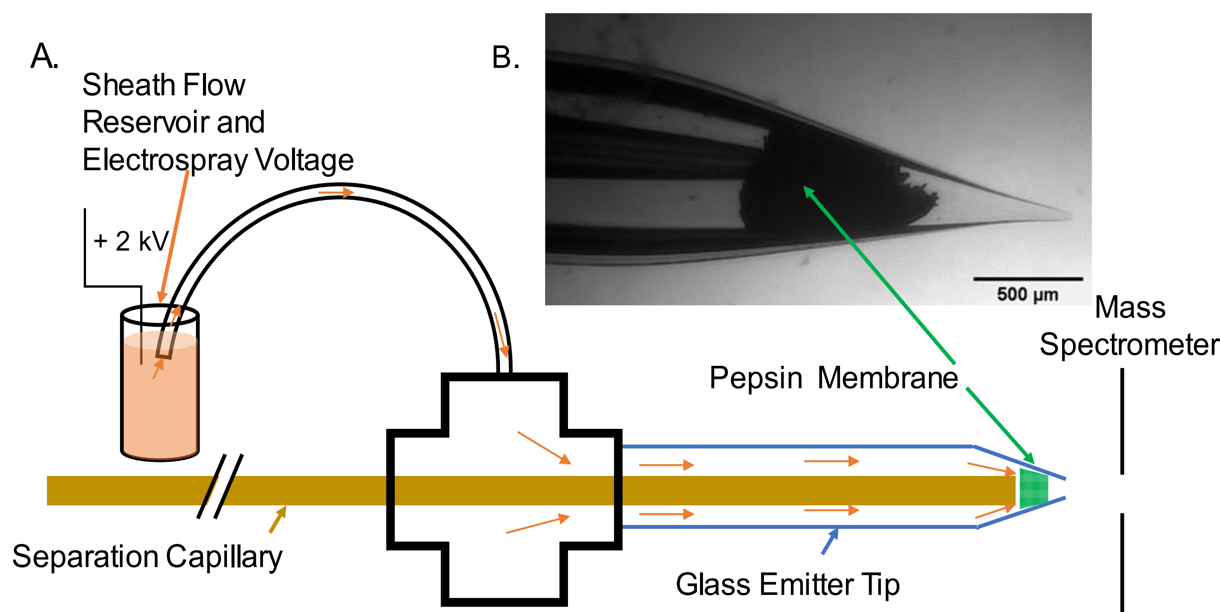

Fig. 2 A. Scheme of the sheath-flow apparatus with a pepsin membrane packed into the pulled glass emitter tip. The sheath-flow interface provides electrokinetically-pumped flow to maintain stable ESI.[41-45] B. Photograph of a membrane packed into the emitter tip. The scale bar was generated using ImageJ based on the capillary inner diameter.[47]

the capillary outlet was sharpened to a 5° point using a 3-D printed grinding apparatus.[40]

The CE was coupled to a Thermo Fisher (San Jose, CA) Linear Trap Quadrupole Velos-Orbitrap Mass spectrometer through a sheath flow electrospray ionization (ESI) interface developed by the Dovichi group.[41–45] The sheath fluid, 10% methanol and 0.5% FA in water, maintains stable ESI. The mass spectrometer was operated in MS/MS mode using higher-energy c-trap dissociation (HCD) with a normalized collision energy of 35%[46] running a top-10 data-dependent method, where a single mass spectrum at a resolution of 60,000 was acquired, and the top 10 precursors were selected for fragmentation. Xcalibur software used a 3-point boxcar smoothing for displaying total ion-current electropherograms (TICs) and extracted ion electropherograms (XIEs).

Protein Separation with CE

Capillaries were cut to 100 cm in length after the tip was ground. The separation employed a 3% acetic acid BGE, and a potential of 30 kV was applied between the BGE vial and the ground of the mass spectrometer. Additionally, a potential of ~2 kV was applied between the sheath-flow reservoir and the mass spectrometer (Fig. 2A). Separations employed a solution containing 0.1 mg/mL (each) of six proteins: albumin, α-lactalbumin, insulin, carbonic anhydrase II, myoglobin, and ribonuclease B in 10 mM ABC. All proteins were bovine. A 20 s hydrodynamic injection (120 nL) was performed at 5 psi. PEAKS online (Bioinformatics Solutions Inc. Waterloo, ON) was used for peptide identification with the following search parameters: precursor mass tolerance of 20 ppm; fragment mass error tolerance of 0.5 Da; peptide length between 6 and 45 amino acids; variable modifications of N-terminal acetylation (Acetyl Protein N-term), methionine oxidation (M), deamidation (NQ), Gln → pyro-Glu, and Glu → pyro-Glu; and non-specific as the enzyme. MS/MS spectra were searched against the *Bos taurus* (Bovine) proteome (proteome ID, UP000009136, 37,510 proteins) with the addition of porcine pepsin.

Emitter Tips

Emitter tips are pulled from glass tubing blanks on a Flaming/ Brown Micropipette Puller Model P-1000. A 10-step heating was used with a final pull and velocity setting of 33 to create a tip size of 30-35 µm. The tip size was chosen based on testing described in Table S1. A round 1.8 mm hole punch was used to section small membranes from a larger coupon, ensuring each tip contained the same amount of pepsin-containing membrane. This section was tightly packed into the emitter tip (Fig. 2B) using gentle pressure from a capillary. Membrane-containing emitter tips were stored at 4 °C until use.

Prior to protein separations, new membrane-containing emitter tips were flushed online by applying the electric field for electrokinetically pumped sheath flow and 5 psi for pressure-driven flow of BGE in the separation capillary for 90 minutes, followed by flushing the emitter tip with sheath fluid using a syringe.[42] After each protein separation, under 5 psi the capillary is flushed with methanol for 1 min and BGE for 5 minutes to remove protein bound to the wall.

Comigrating Proteins Digested Offline

Mixtures of BSA and Herceptin (HER) were prepared at equal concentrations of 0.1 mg/mL or with an excess of BSA at 0.5 mg/mL. Samples were digested offline by passing the solutions through a pepsin membrane using a syringe pump at a flow rate of 100 µL/hr.[47] The resulting peptide mixtures were dried in a speedvac and reconstituted in 10% methanol and 0.1% FA in water for direct infusion MS analysis. MS spectra were matched using Protein Prospector MS-Fit version 6.4.2 (San Francisco, CA) with a mass tolerance of 15 ppm.

Results and Discussion

This study aims to incorporate online peptic digestion after a CE protein separation and immediately prior to ESI. Our method for pepsinolysis includes immobilizing pepsin in a membrane and passing solution through the modified membrane pores. The sections below first describe in-membrane digestion and integration of the membrane into the emitter tip. Subsequently, we investigate the separation and digestion of a mixture of six proteins, along with identification of peptides using MS/MS. Finally, this work examines whether the comparison of migration times of peptides from a given protein can exclude some false-positive peptide identifications.

Ion Suppression with Comigrating Proteins

Because all peptides from a single protein will spray simultaneously after online digestion, ion suppression in ESI may prevent ionization and detection of some peptides. In addition, protein separations are never perfect, so comigrating proteins will send even more peptides through the emitter tip at the same time. To investigate possible ion suppression, we employed direct infusion to spray BSA and HER digests simultaneously to simulate large numbers of peptides from multiple proteins reaching the mass spectrometer at the same time. Equal concentrations of BSA and HER digests were compared to a 5-fold mass excess of BSA.

Table 1 Sequence coverages and numbers of identified peptides for proteins digested and directly infused together into the mass spectrometer to investigate ion suppression.

	1:1 BSA:HER		5:1 BSA:HER	
	Sequence Coverage	# Peptides	Sequence Coverage	# Peptides
BSA	99%	62	82%	58
HER(Heavy Chain)	94%	36	93%	40
HER(Light Chain)	96%	20	96%	21

Table 1 shows the sequence coverages that result from peptide mass fingerprinting. (All other results in this paper employ CE-MS/MS to identify peptides.) HER sequence coverages are high even in a 5-fold excess of BSA, suggesting that ion suppression will not be extensive when several proteins migrate together and undergo simultaneous online digestion. However, greater suppression will occur at higher protein concentrations.[48,49]

Membrane Integration with CE-MS

CE protein separations typically occur at low pH where most proteins carry a net positive charge. Pepsin digests proteins over a pH range of 2-5, with maximum activity between pH 1.5 and 2.5, so it is well suited for digestion under acidic CE conditions.[21] Initially we tried to integrate a pepsin-containing membrane in the separation capillary. A mini microfilter assembly held a membrane in-line between two pieces of capillary (see Fig. S1).[50] Offline proteolysis was effective during pressure-driven flow of proteins through the membrane at velocities typical of those experienced during CE separations. However, during CE the junction produced constant bubbles, which prevented the flow of current. A glass capillary union, shown in Fig. S2, also led to some bubble formation as well as a very high EOF. To overcome the problems associated with

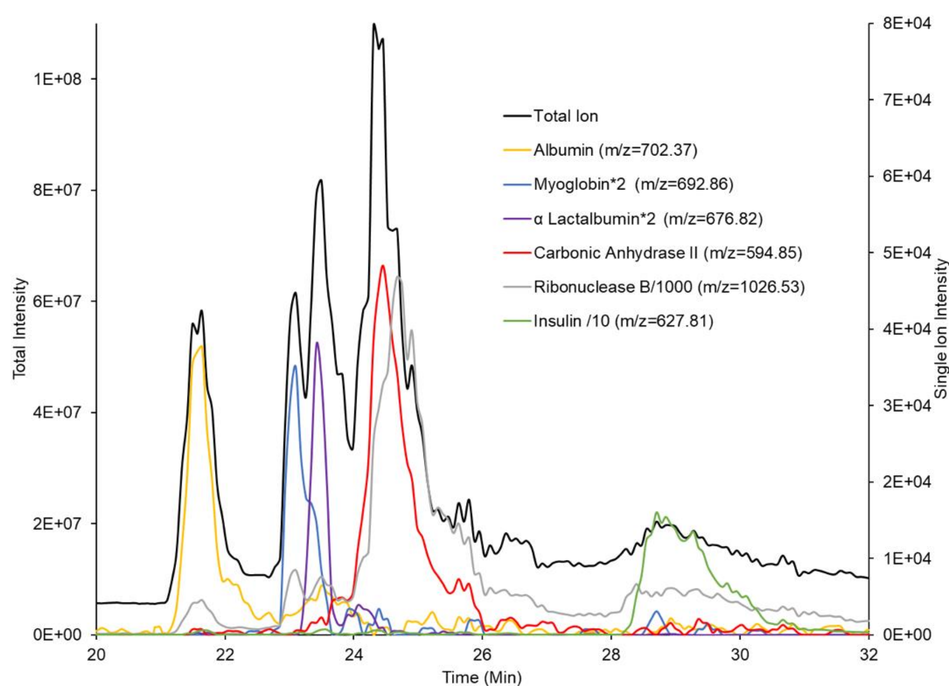

Fig. 3 Electropherogram of the separation and online digestion of six bovine proteins. The TIC is shown in black. m/z values in XIEs correspond to m/z values of individual peptides matched to each protein. Some of the XIEs are scaled as noted to make them visible in the window.

This journal is © The Royal Society of Chemistry 20xx

cutting the CE separation capillary, we chose to insert the membrane into the emitter tip (see Fig. 2).

Inserting the membrane in the emitter tip has two advantages over placing the membrane in the separation capillary: (1) digested peptides have little distance to travel and potentially separate before ESI into the mass spectrometer, and (2) we can easily replace membrane-containing emitter tips between analyses. To test the stability of the pepsin membrane in the sheath flow, we separated and digested intact equine myoglobin 20 times overnight with no loss in digestion performance (Table S2). The first 3 electropherograms showed significant pepsin peptides leaching from the membrane, but fewer pepsin peptides appeared afterward. All 20 experiments provided consistent myoglobin sequence coverage with at least 18 unique peptide identifications for runs 5 through 20. Injection of blank sample buffer after the 20th separation gave rise to no identified myoglobin peptides, suggesting that any peptides adsorbed during prior digestions did not elute from the membrane. Because leaching of pepsin peptides occurred most extensively during the first 3 separations, all subsequent

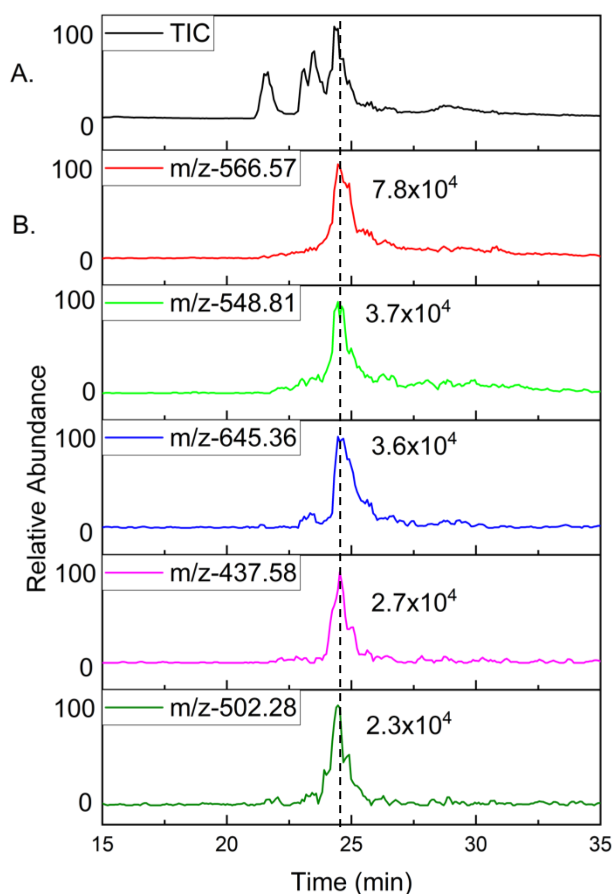

Fig. 4 A. TIC electropherogram for the six-protein CE separation. B. XIEs for the 5 most intense peptides that matched to carbonic anhydrase II. Each XIE is labelled with the m/z value. The maximum intensity is shown to the right of each XIE peak. The dotted black line highlights the consistency in the migration times of the peptides that stem from a given parent protein.

pepsin emitter tips were flushed with BGE for 90 minutes before separating proteins. The addition of a membrane into the emitter tip results in limited peak broadening as Fig. S3 shows. The full width at half maximum for BSA was 18 s wider with the membrane in the emitter tip. The peak broadening is partially due to the membrane and partially due to the increased distance between the end of the separation capillary and the tip of the emitter when the membrane is present.

Separation and Digestion of a Mixture of Six Proteins

We separated a mixture of the six proteins listed in Table 2 and digested them online to examine proteolysis using the pepsin membrane in the emitter tip. These proteins vary in size from insulin at 11 kDa to albumin at 66 kDa. Myoglobin and carbonic anhydrase II do not have disulfide bonds,[51] whereas albumin has 17 of them. These are all bovine proteins, which enables searching of MS/MS spectra against the Uniprot bovine proteome plus porcine pepsin. Because pepsin cleavage occurs with limited selectivity, searching against an entire database, rather than a selection of proteins, is important for achieving confident identifications.

Fig. 3 shows the TIC electropherogram and XIEs from the separation and online digestion of the six-protein mixture. The TIC (black) electropherogram shows partial separation of the proteins. The XIEs correspond to m/z values for peptides that matched to each of the various parent proteins and further show the partial separation of the proteins. Importantly, the peptide XIEs contain peaks that correspond to a unique peak in the TIC electropherogram, indicating that peptides from the separated proteins remain in a narrow plug through digestion

Table 2 Molecular weights, sequence coverages, numbers of disulfide bonds, and numbers of identified peptides for each of six proteins separated and digested online in CE. Sequence coverages are average values for four tips tested in triplicate (n=12), and uncertainties are standard deviations. The number of identified peptides is the median value, rounded to a whole number, from four tips in triplicate. Data are taken directly from Peaks, with no corrections based on migration times.

Protein	Molecular Weight (Da)	Disulfide Bonds	# Peptides	Sequence Coverage (%)
Myoglobin	17,078	0	18	89 ± 15
Carbonic Anhydrase II	29,114	0	24	75 ± 12
Ribonuclease B	14,700	4	3	20 ± 4
Insulin	11,393	2	2	32 ± 22
α-Lactalbumin	16,247	4	7	39 ± 12
Albumin	69,324	17	7	18 ± 10

This journal is © The Royal Society of Chemistry 20xx

and ESI. Despite the small differences in migration times for myoglobin and α-lactalbumin or ribonuclease B and carbonic anhydrase II, the XIEs show distinct migration times for peptides from each protein. For clarity, Fig. 3 gives XIEs of a single peptide for each protein in the mixture, however most matching peptides from a given protein have similar migration times.

Fig. 4 shows the TIC electropherogram and XIEs for the five carbonic anhydrase II peptides with the highest intensities. The migration times of the five peptides are within 10 s of the mean peptide migration time, confirming that most peptides from a given protein appear in a narrow plug. Moreover, the migration times for identified myoglobin and carbonic anhydrase II peptides were distinct and on average differed by 1 min.

Four different membrane-packed emitter tips were used to digest the six-protein mixture in triplicate to test reproducibility. Table 2 lists the molecular weights, numbers of unique identified peptides, sequence coverages, and numbers of disulfide bonds for each of the six proteins. Table S3 gives the numbers of peptides and sequence coverages for each individual replicate as well as cumulative sequence coverages. Although only two peptides are required to identify each protein,[52] the sequence coverages for proteins other than myoglobin and carbonic anhydrase II are <40%.

The low sequence coverages are likely due to disulfide bonds between cysteine residues. These bonds cross-link the protein and make digestion difficult.[53–55] Further, software packages are not designed to identify disulfide-bonded peptides after relatively nonspecific peptic digestion. The mass spectrum of α-lactalbumin in Fig. S4 indicates that the digestion of this protein is incomplete, as signals for both intact protein and peptides appear. Ribonuclease B, insulin, and BSA also show signals from intact protein in mass spectra. In contrast, the MS spectrum from myoglobin, which does not contain disulfide bonds, shows no obvious intact protein (Fig. S5). Like myoglobin, carbonic anhydrase II does not contain disulfide bonds and shows no intact protein along with a relatively high sequence coverage and many identified peptides. In principle, the intact protein signals could aid in protein identifications.[56–58] However, intact protein decreases the concentration of peptides available for analysis and may also cause significant ion suppression.[59] The average peptide length is 21 amino acids, suggesting that digestion occurs with some missed cleavage sites.

We attempted to enhance digestion by reducing proteins before injection in the capillary or adding TCEP to the sheath fluid (see Table S4). However, the electropherogram of reduced proteins (see Fig. S7) showed a loss of resolution, and TCEP in the sheath fluid did not increase sequence coverage. In the case of extensively digested, unreduced proteins with disulfide bonds, many peptides will be cross-linked and not identifiable. The Peaks software does not search for disulfide-bonded peptides. The number of MS/MS spectra that gave peptide-spectrum matches (PSMs) was less than 20% for most of the analyses, and some of these unmatched spectra may correspond to disulfide-bonded peptides. Enabling disulfide-bonded peptide searching would likely increase PSMs and sequence coverages for proteins with disulfide bonds.

Use of Migration Times as a Criterion for Peptide Matching

As Fig. 4 demonstrates, most of the peptides from a single protein reside in the same plug when using sheath-flow membrane digestion. Therefore, migration time could serve as an additional criterion for peptide matching. Table 3 shows all myoglobin peptide matches, their sequences, matching scores, and migration times for a single electropherogram. Peptides with an inconsistent migration time (more than 30 s away from the median migration time of all identified peptides) are highlighted in red. Fig. S8 shows XIEs for each peptide.

Because peptides should not have a migration time shorter than the parent protein, the peptides with migration times <23 min are most likely contamination so we should reject them. In addition, the XIEs for these peptides have no well-defined peak

Table 3 Peptides identified from myoglobin along with their matching scores (-10 log P), m/z values, and migration times (MT). Peptides with an inconsistent migration time are highlighted in red.

Peptide	Score	m/z	MT
L.FTGHPET.L	26.21	394.68	0.17
L.FTGHPETL.E	54.04	451.22	0.17
F.TGHPETL.E	16.09	754.38	0.17
A.DVAGHGQEVL.I	63.95	512.76	0.17
A.AQYKVLGFHG	87.38	560.30	23.10
M.AAQYKVLGFHG	94.13	595.82	23.10
Q.AAMSKALELFRNDMAAQYKVLGFHG	111.87	692.86	23.10
M.GLSDGEWQLVLNAWGKVEADVAGHGQEVL.I	120.67	1026.52	23.10
F.ISDAIIHVLHAKHPSDFGADAQ(+0.98)AAMSKALELFRNDMAAQYKVLGFHG	135.45	849.61	23.10
F.ISDAIIHVLHAKHPSDFGADAQAAMSKALELFRNDMAAQYKVLGFHG	73.62	728.24	23.23
Q.YKVLGFHG	66.41	460.76	23.26
D.AIIHVLHAKHPSDFGADAQAAMSKALELFRNDMAAQYKVLGFHG	99.01	683.22	23.26
L.FRNDMAAQYKVLGFHG	128.24	927.46	23.27
L.NAWGKVEADVAGHGQEVL.I	109.35	940.47	23.42
L.IRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGN(+0.98)TVL.T	38.83	778.76	23.42
L.IRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGNTVL.T	67.16	778.59	23.55
L.NAWGKVEA.D	38.17	437.73	24.61
Y.KVLGFHG	53.36	379.22	26.20

and typically show their highest signals near time zero. Notably, in a bottom-up method one could not determine that these peptides stem from contamination.

Identified peptides with longer migration times than the median might interact with the membrane and slowly desorb to increase the observed migration time, or they could come from

This journal is © The Royal Society of Chemistry 20xx

a different protein with a longer migration time. If the peptide slowly desorbs, we would expect a broad peak with tailing. In contrast, a peptide from another protein should exhibit a relatively sharp peak. The XIE for m/z= 437.73 (see Fig. S8) shows a distinct peak with a migration time (24.6 min) that matches to carbonic anhydrase II, so we would reject it as myoglobin peptide. (The median migration time for myoglobin peptides is 23.2 min.) Further, peptide mass fingerprinting shows that this signal (m/z=437.74, z=2) matches to the peptide QSPVDIDT from carbonic anhydrase II at < 20ppm. Manual interpretation of the MS/MS spectrum was inconclusive (see Fig. S10). The migration time, peak shape, and mass matching to a peptide of carbonic anhydrase II all suggest that this peptide was incorrectly assigned to myoglobin instead of carbonic anhydrase II. The XIE for m/z=379.22 shows a sharp peak at 26.2 min, which is 3 min outside the window where we would expect myoglobin peptides so we would reject it as well. Importantly, for peptides without a post-translational modification, the matching scores are higher for the peptides whose migration times are close to the median value.

The supporting information (Table S5 and Fig. S9) shows similar data for carbonic anhydrase II. Three peptides appear at short migration times, so we can reject them. Again, in a traditional bottom-up method it would not be possible to reject these identifications. Peptides with peak migration times significantly longer than the median showed signals that started to increase at the median migration time, although their maximum intensity appeared later due to what we think is tailing. Thus, in that case we would not reject peptides identified at longer migration times. If we reject all peptides that have an errant migration time in a single electropherogram, the sequence coverage for myoglobin decreases from 75% with 15 peptides to 73% with 11 peptides. For carbonic anhydrase II the sequence coverage decreases from 60% with 25 peptides to 55% with 14 peptides. Thus, rejection of peptides based on inconsistent migration time does not greatly decrease sequence coverage.

Challenges and Possible Solutions

Proteins with disulfide bonds give particularly low sequence coverages when using sheath-flow digestion. The low coverage stems from both incomplete digestion and the challenge of identifying cross-linked peptides. Enhanced digestion will require protein reduction, but this will necessitate changes to the CE method. Zhao et al. separated a reduced antibody heavy and light chain with CE, and Staub et al. proposed several improvements that can limit protein adsorption to the capillary wall.[60,61] When separating reduced proteins it is important to retain the resolution in the separation so that migration times are still distinct for each protein. Höcker et al. developed pressure-driven sheath flow with an additional electrode to provide the voltage needed for ESI. This system may allow longer digestion times to enhance proteolysis in the emitter tip.[62]

Alternatively, identification of cross-linked peptides could enhance sequence coverage. Specialized software can search a whole proteome for cross-linked peptides, but this remains computationally intensive.[63,64] This is especially true for the relatively nonspecific peptic digestion.

Conclusions

This work shows that a pepsin membrane integrated into the sheath flow after a CE separation can catalyze online digestion of separated proteins. Pepsin membranes enabled 20 consecutive digestions of myoglobin with little carryover of myoglobin peptides in a buffer blank. Although the peptides from a single protein enter the mass spectrometer simultaneously, infusion MS suggests that peptide ion suppression should not greatly decrease sequence coverages at low protein concentrations. However, incomplete digestion of proteins with disulfide bonds significantly reduces the number of identified peptides and sequence coverage. Despite incomplete digestion for many proteins, peptide migration times match those of the parent proteins. Thus, peptide migration time can serve as an additional matching criterion to exclude false positive peptide matches. Overall, CE coupled to an online pepsin membrane and MS/MS identification provide a quick and robust separation and digestion tool for protein analysis. Future work should address protein reduction for more complete digestion and identification of peptides.

Author contributions

Conceptualization: K.A.R, M.L.B.; methodology: K.A.R.; validation: K.A.R.; formal analysis K.A.R.; investigation K.A.R.; writing- original draft K.A.R.; writing – review and editing K.A.R., M.L.B.; visualization: K.A.R.; supervision: M.L.B.; funding acquisition: M.L.B.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We acknowledge Dr. Bill Boggess in the Mass Spectrometry and Proteomics core facility at Notre Dame for his expertise and assistance on this project. We also acknowledge Joseph Dragonjac, an undergraduate student who participated in this work. Finally, we gratefully acknowledge support from the National Science Foundation (CHE-1903967).

References

1 A. De Simone, M. Naldi, M. Bartolini, L. Davani and V. Andrisano, *Chromatographia*, 2019, **82**, 425–441.

2 M. Sielaff, J. Kuharev, T. Bohn, J. Hahlbrock, T. Bopp, S. Tenzer and U. Distler, *J. Proteome Res.*, 2017, **16**, 4060–4072.
3 A. M. Girelli and E. Mattei, *Journal of Chromatography B*, 2005, **819**, 3–16.
4 J. Ma, L. Zhang, Z. Liang, W. Zhang and Y. Zhang, *Analytica Chimica Acta*, 2009, **632**, 1–8.
5 J. Ma, L. Zhang, Z. Liang, Y. Shan and Y. Zhang, *TrAC Trends in Analytical Chemistry*, 2011, **30**, 691–702.
6 M. Naldi, A. Tramarin and M. Bartolini, *Journal of Pharmaceutical and Biomedical Analysis*, 2018, **160**, 222–237.
7 A. A. Homaei, R. Sariri, F. Vianello and R. Stevanato, *J Chem. Biol.*, 2013, **6**, 185–205.
8 B. Wouters, S. A. Currivan, N. Abdulhussain, T. Hankemeier and P. J. Schoenmakers, *TrAC Trends in Analytical Chemistry*, 2021, **144**, 116419.
9 Y. Pang, W.-H. Wang, G. E. Reid, D. F. Hunt and M. L. Bruening, *Anal. Chem.*, 2015, **87**, 10942–10949.
10 Y.-J. Tan, W.-H. Wang, Y. Zheng, J. Dong, G. Stefano, F. Brandizzi, R. M. Garavito, G. E. Reid and M. L. Bruening, *Anal. Chem.*, 2012, **84**, 8357–8363.
11 J. Sproß and A. Sinz, *Anal. Chem.*, 2010, **82**, 1434–1443.
12 H. Yuan, S. Zhang, B. Zhao, Y. Weng, X. Zhu, S. Li, L. Zhang and Y. Zhang, *Anal. Chem.*, 2017, **89**, 6324–6329.
13 Z. Kuklenyik, J. I. Jones, C. A. Toth, M. S. Gardner, J. L. Pirkle and J. R. Barr, *Instrumentation Science & Technology*, 2018, **46**, 102–114.
14 M. Bonichon, A. Combès, C. Desoubries, A. Bossée and V. Pichon, *Journal of Chromatography A*, 2016, **1461**, 84–91.
15 C. A. Toth, Z. Kuklenyik, J. I. Jones, B. A. Parks, M. S. Gardner, D. M. Schieltz, J. C. Rees, M. L. Andrews, L. G. McWilliams, J. L. Pirkle and J. R. Barr, *Journal of Proteomics*, 2017, **150**, 258–267.
16 G. W. Slysz, D. F. Lewis and D. C. Schriemer, *J. Proteome Res.*, 2006, **5**, 1959–1966.
17 L. Sun, J. Ma, X. Qiao, Y. Liang, G. Zhu, Y. Shan, Z. Liang, L. Zhang and Y. Zhang, *Anal. Chem.*, 2010, **82**, 2574–2579.
18 G. W. Slysz and D. C. Schriemer, *Anal. Chem.*, 2005, **77**, 1572–1579.
19 Y. Li, R. Wojcik and N. J. Dovichi, *Journal of Chromatography A*, 2011, **1218**, 2007–2011.
20 G. Zhu, L. Sun, X. Yan and N. J. Dovichi, *Anal. Chem.*, 2014, **86**, 6331–6336.
21 D. W. Piper and B. H. Fenton, *Gut*, 1965, **6**, 506–508.
22 D. Trevisan-Silva, A. V. Bednaski, J. S. G. Fischer, S. S. Veiga, N. Bandeira, A. Guthals, F. K. Marchini, F. V. Leprevost, V. C. Barbosa, A. Senff-Ribeiro and P. C. Carvalho, *Sci. Data*, 2017, **4**, 170090.
23 H. M. Kayili and B. Salih, *Talanta*, 2016, **155**, 78–86.
24 J. Ahn, M. C. Jung, K. Wyndham, Y. Q. Yu and J. R. Engen, *Anal. Chem.*, 2012, **84**, 7256–7262.
25 L. Paša-Tolić, C. Masselon, R. C. Barry, Y. Shen and R. D. Smith, *BioTechniques*, 2004, **37**, 621–639.
26 J. S. D. Zimmer, M. E. Monroe, W.-J. Qian and R. D. Smith, *Mass Spectrom. Rev.*, 2006, **25**, 450–482.
27 A. D. Norbeck, M. E. Monroe, J. N. Adkins, K. K. Anderson, D. S. Daly and R. D. Smith, *J. Am. Soc. Mass Spectrom.*, 2005, **16**, 1239–1249.
28 K. Petritis, L. J. Kangas, P. L. Ferguson, G. A. Anderson, L. Paša-Tolić, M. S. Lipton, K. J. Auberry, E. F. Strittmatter, Y. Shen, R. Zhao and R. D. Smith, *Anal. Chem.*, 2003, **75**, 1039–1048.
29 A. Barroso, E. Giménez, F. Benavente, J. Barbosa and V. Sanz-Nebot, *Analytica Chimica Acta*, 2015, **854**, 169–177.
30 F. Benavente, B. Andón, E. Giménez, J. Barbosa and V. Sanz-Nebot, *Electrophoresis*, 2008, NA-NA.
31 O. V. Krokhin and V. Spicer, *Curr. Protoc. Bioinform.*, 2010, **31**, 13.14.1-13.15.15
32 O. V. Krokhin, S. Ying, J. P. Cortens, D. Ghosh, V. Spicer, W. Ens, K. G. Standing, R. C. Beavis and J. A. Wilkins, *Anal. Chem.*, 2006, **78**, 6265–6269.
33 O. V. Krokhin, R. Craig, V. Spicer, W. Ens, K. G. Standing, R. C. Beavis and J. A. Wilkins, *Molecular & Cellular Proteomics*, 2004, **3**, 908–919.
34 V. Spicer, A. Yamchuk, J. Cortens, S. Sousa, W. Ens, K. G. Standing, J. A. Wilkins and O. V. Krokhin, *Anal. Chem.*, 2007, **79**, 8762–8768.
35 N. Klaassen, V. Spicer and O. V. Krokhin, *Journal of Chromatography A*, 2019, **1588**, 163–168.
36 O. V. Krokhin, G. Anderson, V. Spicer, L. Sun and N. J. Dovichi, *Anal. Chem.*, 2017, **89**, 2000–2008.
37 D. Chen, R. A. Lubeckyj, Z. Yang, E. N. McCool, X. Shen, Q. Wang, T. Xu and L. Sun, *Anal. Chem.*, 2020, **92**, 3503–3507.
38 G. Zhu, L. Sun and N. J. Dovichi, *Talanta*, 2016, **146**, 839–843.
39 B. Yu, P. Liu, H. Cong, J. Tang and L. Zhang, *Electrophoresis*, 2012, **33**, 3066–3072.
40 A. Tycova, J. Prikryl and F. Foret, *Electrophoresis*, 2016, **37**, 924–930.
41 Z. Zhang, A. S. Hebert, M. S. Westphall, Y. Qu, J. J. Coon and N. J. Dovichi, *Anal. Chem.*, 2018, **90**, 12090–12093.
42 R. Wojcik, O. O. Dada, M. Sadilek and N. J. Dovichi, *Rapid Commun. Mass Spectrom.*, 2010, **24**, 2554–2560.

This journal is © The Royal Society of Chemistry 20xx

43 Y. Zhao, N. M. Riley, L. Sun, A. S. Hebert, X. Yan, M. S. Westphall, M. J. P. Rush, G. Zhu, M. M. Champion, F. M. Medie, P. A. D. Champion, J. J. Coon and N. J. Dovichi, *Anal. Chem.*, 2015, **87**, 5422–5429.
44 L. Sun, G. Zhu, Z. Zhang, S. Mou and N. J. Dovichi, *J. Proteome Res.*, 2015, **14**, 2312–2321.
45 L. Sun, G. Zhu, Y. Zhao, X. Yan, S. Mou and N. J. Dovichi, *Angew. Chem. Int. Ed.*, 2013, **52**, 13661–13664.
46 J. V. Olsen, B. Macek, O. Lange, A. Makarov, S. Horning and M. Mann, *Nat. Methods*, 2007, **4**, 709–712.
47 F. Xu, W.-H. Wang, Y.-J. Tan and M. L. Bruening, *Anal. Chem.*, 2010, **82**, 10045–10051.
48 A. Furey, M. Moriarty, V. Bane, B. Kinsella and M. Lehane, *Talanta*, 2013, **115**, 104–122.
49 J. L. Sterner, M. V. Johnston, G. R. Nicol and D. P. Ridge, *J. Mass Spectrom.*, 2000, **35**, 385–391.
50 IDEX Health and Science. Microfluidic Connections [brocure]. IDEX Health and Science; 2014. [cited 9 Nov 2022]. Available from: https://www.genetec.se/media//documents/IDEX_-_Microfluidic_Connections.pdf
51 The UniProt Consortium, *Nucleic Acids Res.*, 2017, **45**, D158–D169.
52 J. B. Klein and M. A. Knepper, *JASN*, 2018, **29**, 1585–1587.
53 X. Guan, L. Zhang and J. Wypych, *mAbs*, 2018, **10**, 572–582.
54 L. Switzar, S. Nicolardi, J. W. Rutten, S. A. J. L. Oberstein, A. Aartsma-Rus and Y. E. M. van der Burgt, *J. Am. Soc. Mass Spectrom.*, 2016, **27**, 50–58.
55 J. Wiesner, A. Resemann, C. Evans, D. Suckau and W. Jabs, *Expert Review of Proteomics*, 2015, **12**, 115–123.
56 S. Wu, N. M. Lourette, N. Tolić, R. Zhao, E. W. Robinson, A. V. Tolmachev, R. D. Smith and L. Paša-Tolić, *J. Proteome Res.*, 2009, **8**, 1347–1357.
57 L. Cassidy, P. T. Kaulich, S. Maaß, J. Bartel, D. Becher and A. Tholey, *Proteomics*, 2021, **21**, 2100008.
58 D. López-Ferrer, K. Petritis, E. W. Robinson, K. K. Hixson, Z. Tian, J. H. Lee, S.-W. Lee, N. Tolić, K. K. Weitz, M. E. Belov, R. D. Smith and L. Paša-Tolić, *Molecular & Cellular Proteomics*, 2011, **10**, S1–S11.
59 W. Sun, S. Wu, X. Wang, D. Zheng and Y. Gao, *Eur. J. Mass Spectrom.*, 2005, **11**, 575–580.
60 Y. Zhao, L. Sun, M. D. Knierman and N. J. Dovichi, *Talanta*, 2016, **148**, 529–533.
61 A. Staub, S. Comte, S. Rudaz, J.-L. Veuthey and J. Schappler, *Electrophoresis*, 2010, **31**, 3326–3333.
62 O. Höcker, M. Knierman, J. Meixner and C. Neusüß, *Electrophoresis*, 2021, **42**, 369–373.
63 O. Rinner, J. Seebacher, T. Walzthoeni, L. N. Mueller, M. Beck, A. Schmidt, M. Mueller and R. Aebersold, *Nat. Methods*, 2008, **5**, 315–318.
64 S. McIlwain, P. Draghicescu, P. Singh, D. R. Goodlett and W. S. Noble, *J. Proteome Res.*, 2010, **9**, 2488–2495.

This journal is © The Royal Society of Chemistry 20xx