Journal of proteome research

Technical Note

pubs.acs.org/jpr

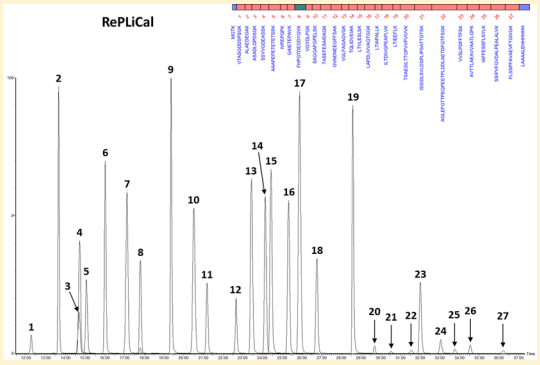
RePLiCal: A QconCAT Protein for Retention Time Standardization in Proteomics Studies

Stephen W. Holman, Lynn McLean, and Claire E. Eyers*

Centre for Proteome Research, Department of Biochemistry, Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool L69 7ZB, United Kingdom

Supporting Information

ABSTRACT: This study introduces a new reversed-phase liquid chromatography retention time (RT) standard, RePLiCal (**Re**versed-**p**hase **li**quid chromatography **cal**ibrant), produced using QconCAT technology. The synthetic protein contains 27 lysine-terminating calibrant peptides, meaning that the same complement of standards can be generated using either Lys-C or trypsin-based digestion protocols. RePLiCal was designed such that each constituent peptide is unique with respect to all eukaryotic proteomes, thereby enabling integration into a wide range of proteomic analyses. RePLiCal has been benchmarked against three commercially available peptide RT standard kits and outperforms all in terms of LC gradient coverage. RePLiCal also provides a higher number of calibrant points for chromatographic retention time standardization and normalization. The standard provides stable RTs over long analysis times and can be readily transferred between different LC gradients and nUHPLC instruments. Moreover, RePLiCal can be used to predict RTs for other peptides in a timely manner. Furthermore, it is shown that RePLiCal can be used effectively to evaluate trapping column performance for nUHPLC instruments using trap-elute configurations, to optimize gradients to maximize peptide and protein identification rates, and to recalibrate the m/z scale of mass spectrometry data post-acquisition.



KEYWORDS: *retention time, LC, standardization, SRM, proteomics, QconCAT, mass spectrometry, calibration, quality control*

INTRODUCTION

The method of choice for complex "bottom-up" proteomic sample analysis is low pH reversed phase−high-performance liquid chromatography−electrospray ionization-tandem mass spectrometry (RP−HPLC−ESI−MS/MS), with or without peptide prefractionation. Optimal LC−MS instrument performance is paramount to ensure that the highest quality data are acquired and maximum information is garnered from every experiment. Typical practice is to regularly analyze a reliable, well-characterized standard sample that permits assessment of instrumental performance with respect to relevant parameters of interest, for example, chromatographic peak width, mass spectrometric signal response, protein sequence coverage, and so on[1−3] Quality-control standards enable instrument performance to be monitored longitudinally,[4] expedite troubleshooting to maximize uptime,[5] and allow comparisons to be made between experimental conditions, instruments, analysts, and laboratories.[6−8]

One key parameter in the assessment of instrument performance in bottom-up proteomics is peptide retention time (RT). Repeatable RT is critical in many types of proteomics experiments, such as label-free quantification to allow accurate mass-retention time (AMRT) realignment of multiple data acquisitions[9,10] and for the scheduling of transitions in selected reaction monitoring (SRM) studies.[11] Peptide RT has also been used in many studies (and arguably could be used to even greater effect) as an orthogonal identification criterion alongside product ion information to filter false-positive identifications and improve peptide and protein identification rates, typically through the use of predictive models developed by machine learning algorithms.[12−16] Predictive programs based on fundamental physicochemical phenomena related to peptide chromatographic behavior have also been reported;[17−19] however, these models have been demonstrated to generally only perform well under the specific conditions that they were developed.[20] Such models cannot always be applied to an individual laboratory's LC−MS setup, and empirical determination of RT is often preferred. Several laboratories use enzymatic digests of one or more readily available proteins as RT standards for benchmarking performance.[21] While this can be efficient and cost-effective, the resultant set of peptides may not be suitable to fully test the instrument should they fail to elute across the whole LC gradient or be overly complicated as to hamper straightforward interrogation, for example, as with an *E. coli* digest. To address this, a number of groups have reported the development of dedicated standards and workflows for

Received: October 23, 2015
Published: January 18, 2016

DOI: 10.1021/acs.jproteome.5b00988
J. Proteome Res. 2016, 15, 1090−1102

benchmarking liquid chromatograph performance.[22−27] Of note is the cross-platform iRT standardization method reported by Escher and coworkers, underpinned by a set of 11 peptides for each of which a dimensionless elution parameter (iRT value) is attributed.[25] The iRT concept involves attributing an analyte-specific dimensionless value to each peptide in a scheduled SRM experiment, which is fixed relative to the standard peptides. By measuring the RT of the standard peptides on a different column, LC instrument, and so on, the RT of other peptides can be determined given their iRT value. Transfer of scheduled SRM methods between conditions thus only requires a single recalibration using the iRT standard peptides. The authors demonstrated high accuracy in RT prediction, allowing 4-fold narrower scheduling time windows than in silico prediction using SSRCalc.[17,19] Consequently, improved precision was achieved due to the increase in dwell time per transition. A similar strategy was described by Gallien and coworkers, although they achieved real-time rescheduling of time windows for SRM analysis.[26] If a change in expected RT was observed for one of nine standard peptides, the time window for targeted peptides was altered by the control software. A concomitant reduction was observed in the number of targets for which data was lacking due to peptides eluting outside of the originally scheduled time period, obviating the need for reanalysis; however, both of these RT standards are limited in their content of extremely hydrophilic and hydrophobic peptides and thus their ability to normalize over the early and late phases of the chromatographic gradient. This is particularly problematic for early eluting peptides, which demonstrate more variability in their RTs and thus the ability to calibrate for real-time RT scheduling at the beginning of an LC gradient.[26]

Here a new peptide RT standard called RePLiCal (**Re**versed-**p**hase **li**quid chromatography **cal**ibrant) is described. The standard is generated using QconCAT technology and is a designer protein containing 27 lysine-terminating peptides.[28,29] Proteolytic digestion of the protein using either Lys-C or trypsin repeatably generates the same set of peptides, demonstrated to elute over a wider time range than three commercially available peptide mixtures designed for RT calibration. The utility and benefits of RePLiCal will be discussed.

EXPERIMENTAL SECTION

Peptide Selection

Candidate peptides were selected using in-house RP−nano-ultra-high-performance liquid chromatography−nano-ESI−mass spectrometry with elevated energy (RP−nUHPLC−nESI−MS^E)[30] data from tryptic digests of *E. coli* and *S. cerevisiae*. Peptides containing Met, Trp, Cys, and *N*-terminal Gln residues and Asp-Pro, Asn-Pro, and Asn-Gly motifs were removed from the data set. Arginine-terminating peptides were excluded from consideration to ensure that the selected sequences would conform to both Lys-C and trypsin cleavage specificity once assembled into the QconCAT protein. Peptides with good elution profiles (minimal tailing, narrow peak width at FWHM, and 10% height) were chosen for further consideration. Candidates from throughout the gradient were selected. For peptides originating from *S. cerevisiae*, conservative mutations were made to the amino acid sequences that have been reported to have minimal effect on retention times, for example, Gly to Ala, Glu to Asp.[31] In addition, peptides with Asp and Glu in positions P4, P3, P2, P1′, P2′, and P3′ (Schechter & Berger nomenclature)[32] were permutated to move the acidic side chains further from the lysine residue and reduce the likelihood of missed cleavages in the final QconCAT protein.[33−36] Candidate peptides were BLAST[37] searched against the nonredundant protein sequences (nr) database (searched 12/04/2012) with prokaryotic organisms excluded to verify uniqueness as Lys-C or tryptic peptides, taking into consideration the inability of low-energy CID to differentiate Leu and Ile and also low resolving power instrumentation to discriminate between Gln and Lys. Sequences that matched a Lys-C or tryptic peptide in any eukaryotic organism were discarded. The remaining 61 peptides were prepared by SPOT synthesis[38,39] (JPT Peptide Technologies, Berlin, Germany) to enable evaluation of the novel sequences generated by conservative mutation and sequence permutation. Each peptide was resolubilized in 100 mM ammonium bicarbonate (AmBic)/MeCN [80:20 v/v] according to the manufacturer's instructions to a final concentration of 1 nmol μL^{-1}. One μL was taken from each stock, pooled, and dried to completeness using vacuum centrifugation before resolubilization in 100 mM AmBic/MeCN [80:20 v/v] to a final concentration of 1 pmol μL^{-1}, followed by vortexing and sonication for 5 min. Peptides were diluted 1 in 10 using 0.1% formic acid (FA) in H_2O/MeCN [97:3 v/v] and 1 μL analyzed in triplicate by RP−nUHPLC−MS^E using a nanoACQUITY LC instrument coupled to a Synapt HDMS Q-ToF mass spectrometer (Waters, Elstree, U.K.) (see Supporting Information). The data were evaluated and 27 consistently observed peptides eluting at regular time points throughout the gradient were selected for inclusion in the QconCAT protein.

Recombinant Expression of RePLiCal and Initial Characterization

RePLiCal was prepared by heterologous expression in *E. coli* using isopropyl β-D-1-thiogalactopyranoside (IPTG) induction, purified by affinity chromatography by virtue of a 6-His-tag, and quantified by SDS-PAGE with reference to a bovine serum albumin standard curve. (Full details can be found in the Supporting Information.) The purified material was solubilized in 25 mM AmBic and digested with either trypsin (Sigma-Aldrich, Dorset, U.K.) or Lys-C (Roche Diagnostics, West Sussex, U.K.) at an enzyme/substrate ratio of 1:50 [w/w] overnight at 37 °C. Initially, the sample was analyzed by RP−nUHPLC−MS^E as previously described to check that all of the RePLiCal peptides could be detected and to select the four most intense product ions above m/z 400 for each peptide to monitor in SRM assays.

Digestion of Yeast Lysate

Yeast (*S. cerevisiae*) was prepared as previously described.[40,41] The whole cell lysate (100 μg) was solubilized in 0.1% RapiGest *SF*,[42] reduced with dithiothreitol (final concentration 3 mM), alkylated with iodoacetamide (final concentration 9 mM), and digested with trypsin at an enzyme:substrate ratio of 1:50 [w/w] overnight at 37 °C. The addition of trifluoroacetic acid (TFA) to 0.5% [v/v] terminated the enzymatic reaction and degraded the RapiGest *SF* following incubation at 37 °C for 2 h. The sample was centrifuged at 13 000*g*, 4 °C for 15 min and the cleared supernatant fraction retained for analysis.

LC−MS Analyses

All LC−MS analyses were performed using a nanoACQUITY LC instrument (Waters), except for the comparison experiment

DOI: 10.1021/acs.jproteome.5b00988
J. Proteome Res. 2016, 15, 1090−1102

with the Ultimate 3000 RSLC LC instrument (ThermoFisher Scientific, Hemel Hempstead, U.K.). SRM-MS analyses were performed using a Xevo TQMS tandem quadrupole mass spectrometer (Waters) and nontargeted MS analyses were done on a variety of Orbitrap-based platforms (ThermoFisher Scientific). Full details of each experiment can be found in the Supporting Information.

RESULTS AND DISCUSSION

Design of RePLiCal

The QconCAT methodology[28,29] describes the synthesis of artificial designer proteins that, upon proteolysis, generate an ensemble of peptides. This technology was exploited here to generate a collection of peptides suitable for the testing and standardization of HPLC instrumentation used for bottom-up proteomics. The resultant artificial protein, RePLiCal, is a 379 amino acid residue, 39 193.3 Da protein containing 27 peptides designed for the calibration and standardization of HPLC instruments and a C-terminal His-tag for purification (Figure 1). RePLiCal was specifically designed to contain only Lys-terminating peptides, such that the same complement of proteolytic fragments could be generated upon enzymatic digestion using either Lys-C or trypsin, the two most commonly employed proteases in proteomics studies.[43−45]

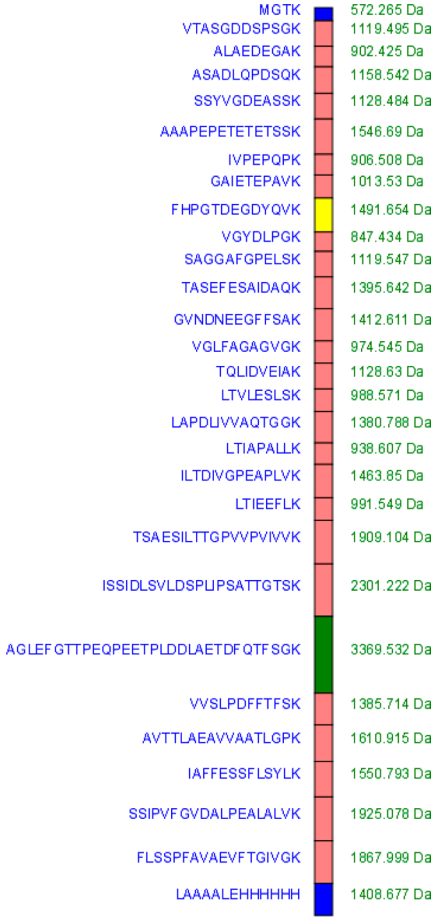


Figure 1. Structure of RePLiCal with cleavage sites for Lys-C and trypsin denoted by vertical black lines.[73] Pink indicates those peptides observed as $[M + 2H]^{2+}$ species, green block as $[M + 3H]^{3+}$ species, and yellow as both $[M + 2H]^{2+}$ and $[M + 3H]^{3+}$ species. Blue indicates peptides not used for RT standardization.

The generated peptide sequences are unique with respect to all eukaryotic proteomes. Therefore, RePLiCal can be implemented as a standard for a wide range of proteomics analyses without interfering with the sample under consideration. Key chemically reactive residues and motifs were avoided to aid stability of the protein during storage and sample preparation.[46] Crucially, the concatenation of the calibrant peptides into an artificial protein presents an ideal storage environment, limiting the selective loss of certain peptides. Hydrophobic peptides, in particular, are known to adhere to surfaces over time, in some cases irreversibly.[47−49] Therefore, RT standards stored as peptide mixtures can experience a loss of calibration points due adherence of one or more of the analytes to a surface over time. Because the peptides in RePLiCal are stored at the protein level there is no opportunity for differential adsorption to take place: While loss of the protein could potentially occur, the peptides will remain in a 1:1 stoichiometry prior to digestion, meaning that all calibration points will be available even when the protein standard is stored for an extended time period.

Comparison of RePLiCal with Commercially Available Peptide RT Standards

RePLiCal was directly compared with three commercially available peptide RT standards: iRT-Kit (Biognosys AG, Zurich, Switzerland), Peptide Retention Time Calibration Mixture (Pierce, Rockford, IL), and MS RT Calibration Mix (Sigma-Aldrich, Poole, U.K.). SRM assays were designed for each RePLiCal peptide by selecting the four most intense product ions with m/z values greater than 400 from LC−MS^E data. Three or four transitions per peptide were monitored for the commercial RT standards as recommended by the manufacturer. (SRM transitions can be found in the Supporting Information, Tables S5−S8.) Figure 2 shows the chromatograms for 10 fmol of each of the RT standards acquired under the same LC conditions (columns, mobile phases, gradient, etc.). It is evident that the RePLiCal peptides elute over a RT wider range using the 30 min LC gradient program and a trap-elute LC configuration than the peptides from the three commercially available kits. (Equivalent chromatograms for 10, 60, and 90 min LC gradient programs are shown in the Supporting Information Figures S4−S6, which demonstrate the same chromatographic behavior. All retention time data used to construct the figures are also included in the Supporting Information.) Crucially, the first two peptides from RePLiCal elute before the first peptide observed from any of the other standards assessed; indeed, RePLiCal peptide **1**, VTASGDDSPSGK, elutes 2.4 min (2.96% B over the 30 min gradient) before the first peptide from any of the commercially available kits. The Sigma standard contains two peptides, RGDSPASSPK and GLVK, which were not observed using a trap-elute LC configuration. A direct injection configuration, that is, no trapping column in the flow path, was tested to compare the performance of the early eluting peptide in RePLiCal with those in the Sigma standard. While the Sigma peptides eluted before the RePLiCal peptides on the nanoACQUITY LC system (Figure S7), the elution profiles were extremely poor. Furthermore, the Sigma peptides fail to describe the linear gradients under direct injection conditions (Figure S8); the trendlines deviate from linearity, and the relationship between gradient length and trendline gradient no longer correlates as expected: The slope of the 60 min gradient trendline should be twice that of the 30 min gradient trendline, which is not observed. Conversely, the early eluting RePLiCal peptides

DOI: 10.1021/acs.jproteome.5b00988
J. Proteome Res. 2016, 15, 1090−1102

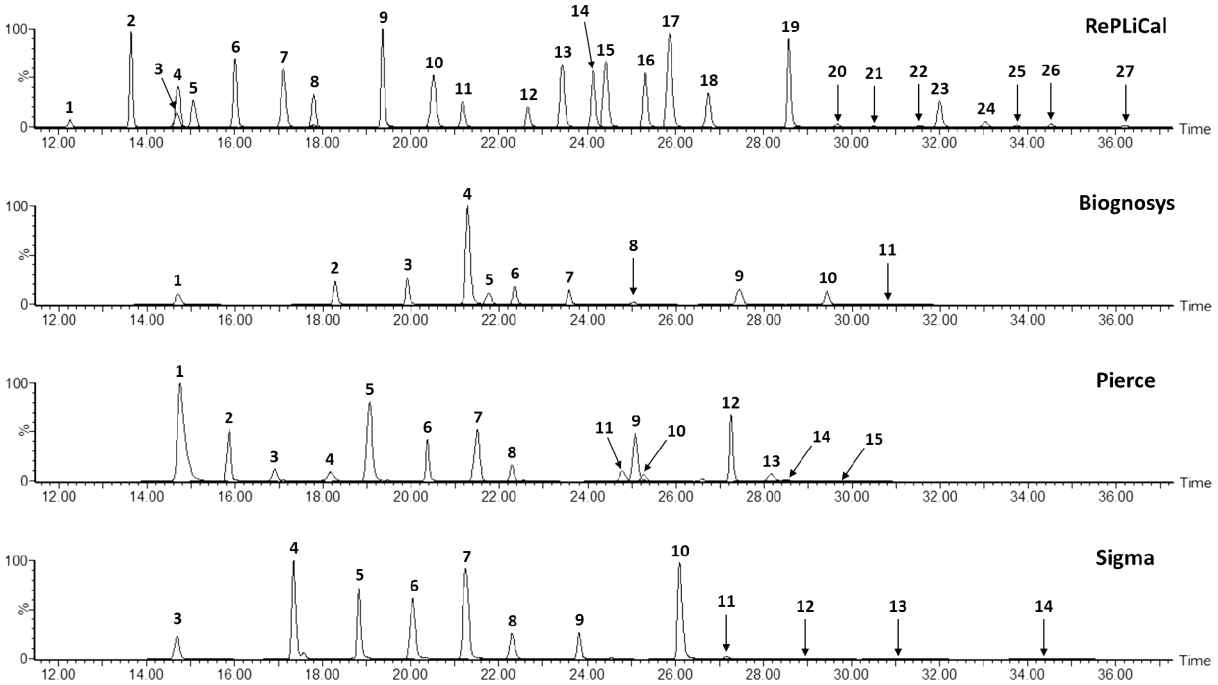


Figure 2. Comparison of chromatograms on a 30 min LC gradient (3−40% 0.1% FA in MeCN) with nESI−SRM−MS data acquisition for RePLiCal and three commercially available retention time standards; iRT-Kit (Biognosys), Peptide Retention Time Calibration Mixture (Pierce), and MS RT Calibration Mix (Sigma). The numerical annotations represent the elution order as provided by the manufacturer. Peptides 1 and 2 from the MS RT Calibration Mix were not observed. These peptides, RGDSPASSPK and GLVK, are very hydrophilic and are not trapped efficiently.

continue to chromatograph with good peak shape (Figure S7), and all four linear gradients evaluated are well described by the standard (Figure S8). Subsequently, the peptides were separated using a direct injection configuration on an RSLC LC instrument with a 30 min gradient, upon which changes in selectivity for the peptides were observed (Figure S9). VTASGDDSPSGK from RePLiCal now eluted earlier than both RGDSPASSPK and GVLK from the Sigma standard. VTASGDDSPSGK (and ALAEDEGAK) gave good chromatographic performance, whereas both RGDSPASSPK and GVLK produced poor peak shape. The data therefore indicate that RePLiCal outperforms all three standards evaluated in terms of characterizing the early part of a LC gradient. Additionally, RePLiCal provides between one and seven additional peptides that elute after the longest retained peptides in the commercially available kits. A significant number of additional data points for standardization are therefore obtained during the earlier and latter parts of the gradient. A whole cell yeast lysate tryptic digest was separated over a 90 min LC gradient to evaluate the relationship between these additional standardization points and a complex proteome sample. Of the 2696 identified peptides, only 64 (2.37%) eluted outside of the retention times covered by RePLiCal (Supporting Information Table S1). This compares with 307 (11.39%), 609 (22.59%), and 774 (28.71%) peptides for the Sigma, Biognosys and Pierce peptides, respectively. Therefore, RePLiCal significantly enhances the fraction of peptides in a complex proteome sample that can be subjected to standardization. The additional data points will also be advantageous for the iRT concept,[25] which as previously noted cannot be used for peptides with RTs before the first or after the last eluting reference peptide.[50] The reference peptides encoded within RePLiCal will thus help to extend the iRT concept by providing a greater number of calibration points over a wider range of elution times. RePLiCal also has a greater density of calibration points across the chromatographic gradient than the other standards. In concert with the additional early and late eluting peptides, the greater density of potential calibration points provided by RePLiCal is undoubtedly advantageous in terms of instrument standardization and realignment of data in label-free quantification studies. In addition, the high number of reference peptides provided by RePLiCal, which are interspersed consistently throughout the gradient, should allow more effective implementation of the dynamic scheduled SRM analysis described by Gallien and coworkers.[26] Furthermore, the inclusion of standard peptides that elute significantly earlier in the gradient will allow faster reaction to RT changes between runs and thus more efficient real-time correction of time-scheduled SRM experiments, where two recently detected reference peptides are required to adjust the scheduled time window for subsequent peptide analysis. These earliest eluting peptides from RePLiCal are particularly important given that RTs early in the gradient are known to be more variable.[26]

Comparison of Empirical and Predicted RTs of RePLiCal Peptides

A number of algorithms have been described that predict peptide RTs on C18 columns under low-pH RP conditions;[14,17−19,51,52] however, none of the models accounts for all phenomena resulting from the interaction of peptides with C18 stationary phases, such as the stabilization of helical structures recently demonstrated to be a key contributing factor to the elution of peptides from RP material; such RT prediction algorithms are thus prone to error.[53] Failure to predict RT accurately would detrimentally affect scheduled SRM analyses or the normalization for AMRT of peptides between experiments or experimental systems. The BioLCCC model was used to predict the behavior of RePLiCal peptides to assess the hypothesis that prediction of the elution order and RTs of the

DOI: 10.1021/acs.jproteome.5b00988
J. Proteome Res. 2016, 15, 1090−1102

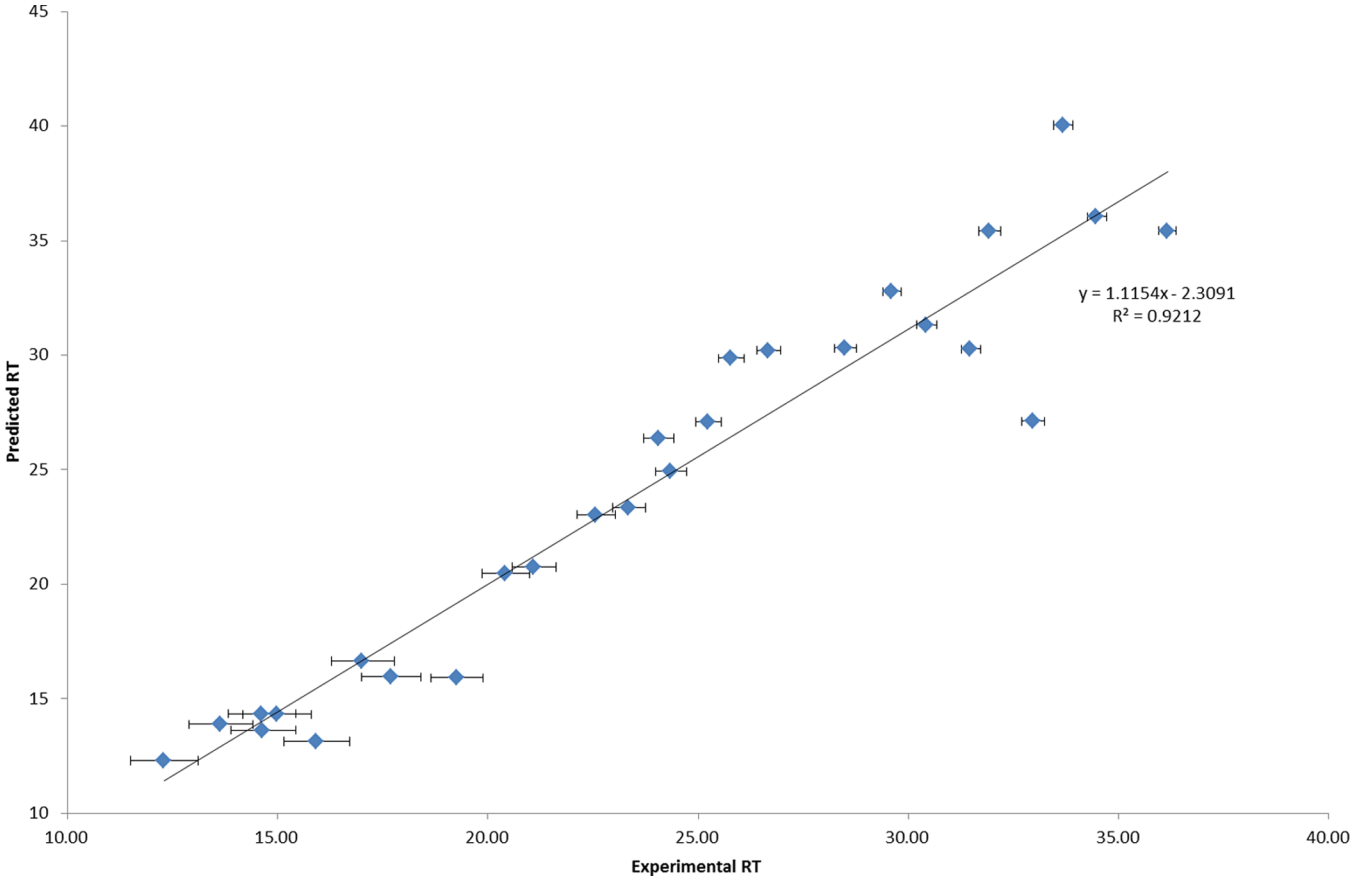


Figure 3. Comparison of predicted RTs of RePLiCal peptides using BioLCCC and the average experimentally determined values (n = 3) over a 30 min gradient (3−40% 0.1% FA in MeCN).

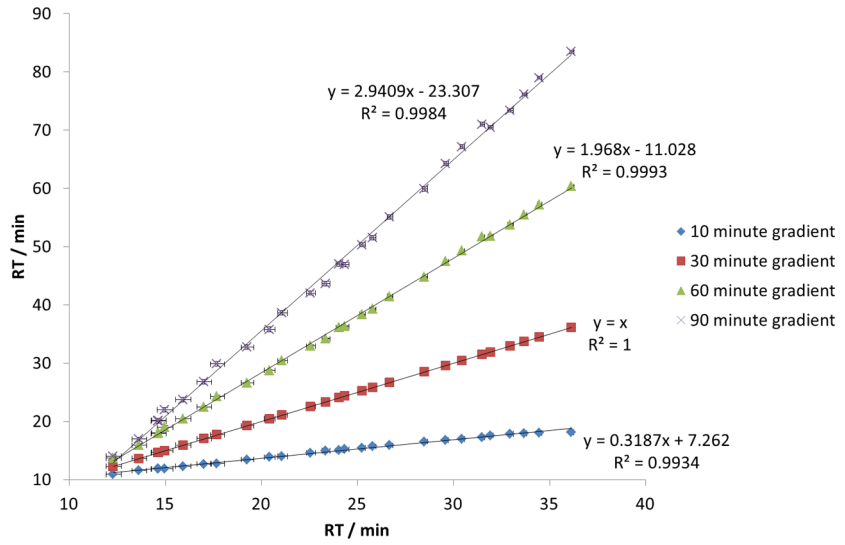


Figure 4. Comparison of the retention times of RePLiCal peptides on different length LC gradients (3−40% 0.1% FA in MeCN) using the 30 min gradient as a reference (n = 3). Error bars represent ±2 standard deviations.

RePLiCal peptides would differ significantly from that observed experimentally. BioLCCC was chosen for comparison due to the ability of the in silico predictor to be programmed to match the experimental conditions under which RePLiCal was analyzed.[18] Figure 3 shows that correlation between predicted and experimental RTs is reasonably strong (R^2 = 0.9212); however, for accurate prediction of unknown RTs, the equation for the regression line should be $y = x$. This was not observed (y = 1.1154x − 2.3091) and poor prediction of RTs was achieved using the BioLCCC model, with an average of a 1.81 min error in the predicted RT. The scatter of data points around the regression line also demonstrates that prediction of elution order is incorrect; for example, peptide x is predicted to elute in position y, thus suggesting that selectivity of separation is also poorly predicted. The observed discrepancies between the predicted and empirical data emphasize the need for reliable

DOI: 10.1021/acs.jproteome.5b00988
J. Proteome Res. 2016, 15, 1090−1102

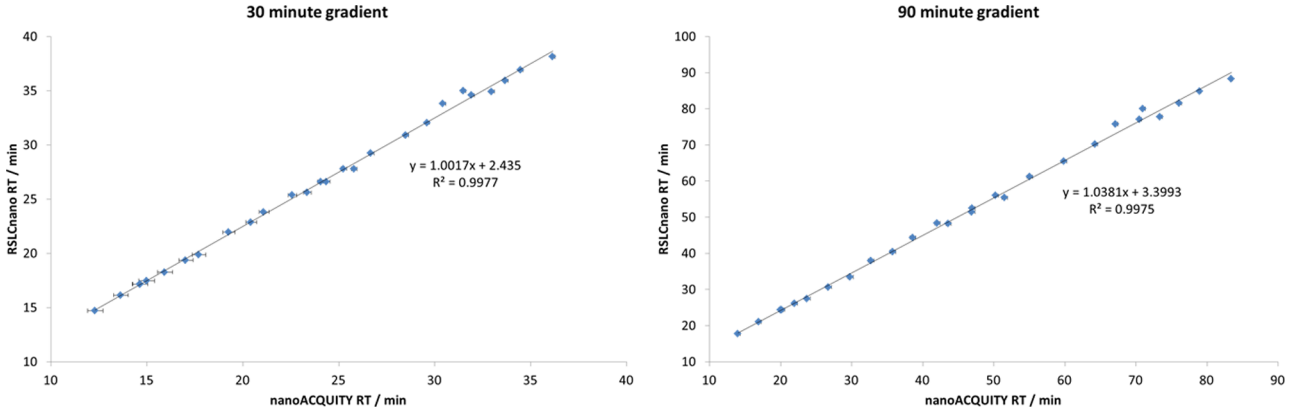


Figure 5. Comparison of the retention times of RePLiCal peptides on two different nano LC instruments (Thermo RSLC and Waters nanoACQUITY) using two different LC gradient lengths (3−40% 0.1% FA in MeCN) (n = 3). Error bars represent ±2 standard deviations.

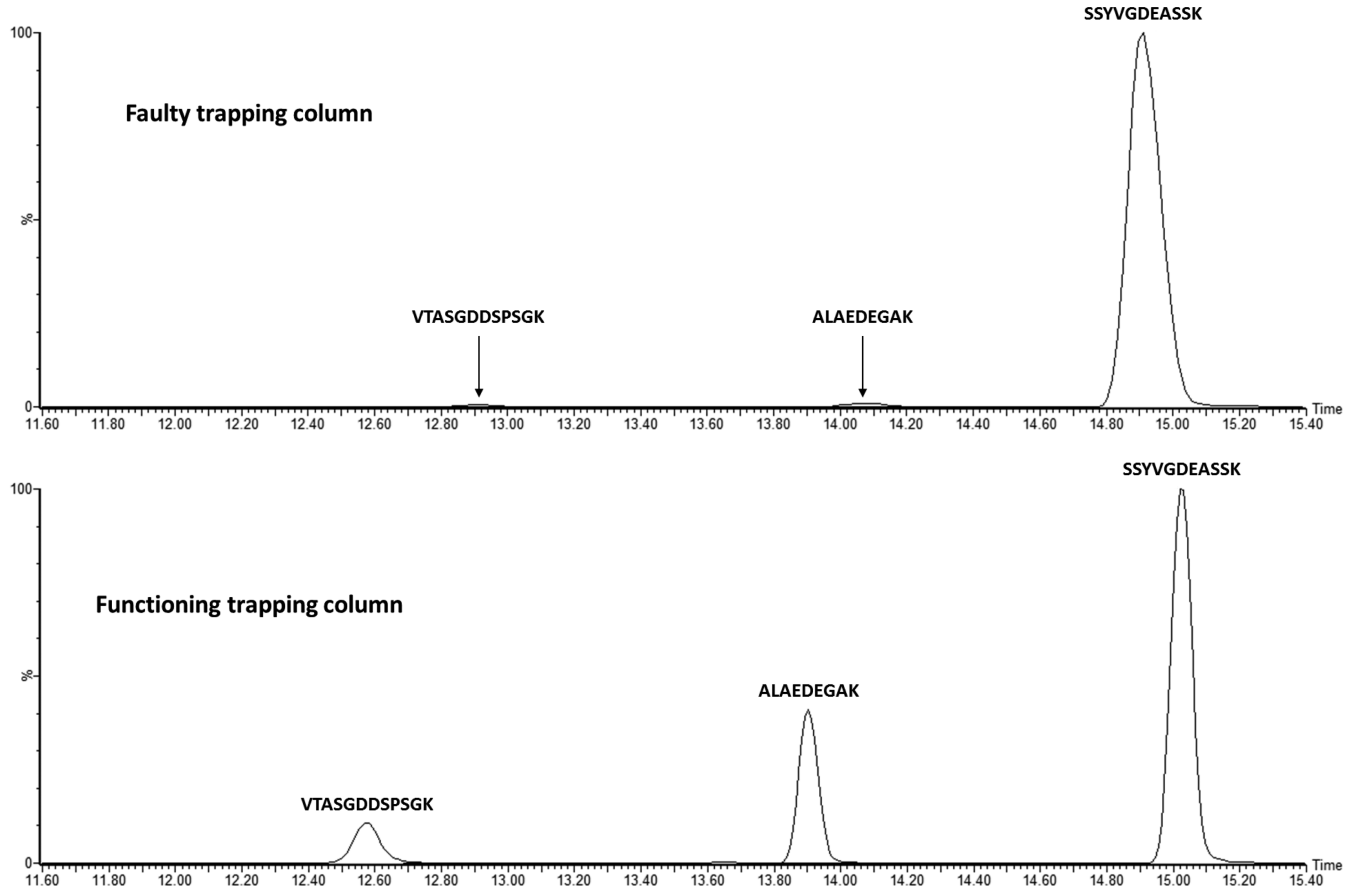


Figure 6. Comparison of trapping column performance using the intensities of the three earliest eluting peptides in RePLiCal.

standards to assess LC conditions and performance experimentally across multiple platforms.

Transferability of RePLiCal Across Gradients and LC Instrumentation

Figure 4 shows the retention times of the 27 RePLiCal peptides using four different length gradients (10, 30, 60, and 90 min, all 3−40% 0.1% FA in MeCN). Using the 30 min gradient as the reference, a high degree of proportionality in the transfer of RTs across gradients was observed. The linear relationship in RTs for the RePLiCal peptides for each individual gradient shows that the peptides are suited to characterizing the gradient, which itself was linear. The linearity observed for each of the gradients assessed means that, in principle, it should be possible to predict RTs of peptides on a new gradient simply by knowing their RT relative to the RePLiCal peptides on a reference gradient. By establishing the linear relationship between the reference and new gradient by a single analysis of RePLiCal, large-scale transfer of RTs of other peptides can be performed. Such an approach is particularly advantageous for adjusting scheduled SRM assays for different LC conditions on the same instrument (vide infra) or transporting assays between different LC instruments.[11] The ability of RePLiCal to permit the transfer of assays between different LC instruments was confirmed by analysis on platforms from two different manufacturers (Figure 5). Again, good linearity was observed,

DOI: 10.1021/acs.jproteome.5b00988
J. Proteome Res. 2016, 15, 1090−1102

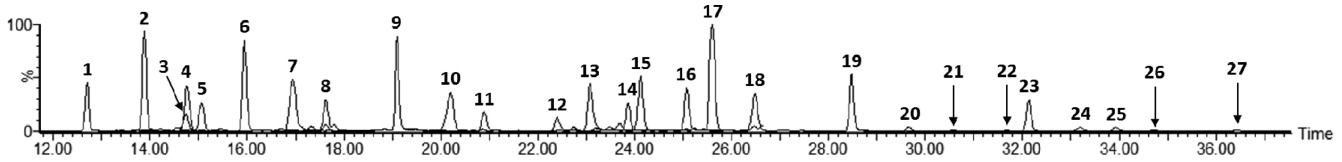


Figure 7. nLC−nESI−SRM−MS chromatogram of 5 fmol of RePLiCal spiked into 1 μg of a whole cell yeast lysate tryptic digest separated over a 30 min gradient (3−40% 0.1% FA in MeCN).

Table 1. Average RT, Peak Widths and Associated RSDs from 60 Injections of RePLiCal on 30 min LC Gradient (3−40% 0.1% FA in MeCN)

peptide no.	peptide sequence	average RT/min	RSD RT/min	average peak width (FWHM)/s	RSD peak width (FWHM)/s
1	VTASGDDSPSGK	12.55	1.14	5.17	3.00
2	ALAEDEGAK	13.75	0.93	5.41	8.98
3	ASADLQPDSQK	14.65	0.83	8.00	1.96
4	SSYVGDEASSK	14.67	0.82	7.72	2.63
5	AAAPEPETETETSSK	14.97	0.80	7.61	1.82
6	IVPEPQPK	15.89	0.71	6.13	13.03
7	GAIETEPAVK	16.90	0.69	7.11	3.63
8	FHPGTDEGDYQVK	17.60	0.66	6.35	5.51
9	VGYDLPGK	19.08	0.59	5.50	15.10
10	SAGGAFGPELSK	20.22	0.57	9.76	11.39
11	TASEFESAIDAQK	20.88	0.53	5.22	10.20
12	GVNDNEEGFFSAK	22.42	0.51	6.25	15.30
13	VGLFAGAGVGK	23.12	0.49	7.39	4.82
14	TQLIDVEIAK	23.90	0.44	6.34	1.51
15	LTVLESLSK	24.17	0.45	8.19	4.02
16	LAPDLIVVAQTGGK	25.10	0.40	6.20	12.61
17	LTIAPALLK	25.64	0.41	7.44	8.68
18	ILTDIVGPEAPLVK	26.51	0.39	6.25	6.18
19	LTIEEFLK	28.56	0.47	7.26	8.55
20	TSAESILTTGPVVPVIVVK	29.62	0.29	6.92	23.49
21	ISSIDLSVLDSPLIPSATTGTSK	30.55	0.31	7.61	2.76
22	AGLEFGTTPEQPEETPLDDLAETDFQTFSGK	31.72	0.32	8.34	5.27
23	VVSLPDFFTFSK	32.23	0.39	8.10	5.61
24	AVTTLAEAVVAATLGPK	33.34	0.49	8.53	13.53
25	IAFFESSFLSYLK	34.13	0.55	8.66	24.25
26	SSIPVFGVDALPEALALVK	34.82	0.30	6.37	12.74
27	FLSSPFAVAEVFTGIVGK	36.66	0.42	8.72	5.63

although slight changes in selectivity were evident, particularly during the latter part of the gradient; however, the changes in RT due to differential selectivity are minor and do not cause a significant deviation from perfect linearity, suggesting a negligible effect on transferring RTs from one instrument to another. Interestingly, changing the length of the chromatographic gradient results in a slight difference in selectivity of retention, as evidenced by changes in elution order of the RePLiCal peptides (Figure 4). For two pairs of peptides, AGLEFGTTPEQPEETPLDDLAETDFQTFSGK (**22**) and VVSLPDFFTFSK (**23**), and TQLIDVEIAK (**14**) and LTVLESLSK (**15**), a change in elution order was observed for the 90 min gradient. Rationalization of this observation can be provided by the linear solvent strength (LSS) theory (eq 1)[54−56]

$$\log k = \log k_0 - S_\phi \quad (1)$$

where k is the retention factor at a given organic solvent volume fraction ϕ (in this study, % MeCN/100) in the mobile phase and k_0 is the retention factor with a pure aqueous mobile phase. The S parameter is a constant for a particular peptide at a given ϕ value and essentially measures the rate of change of the retention factor as a function of changing organic solvent composition of the mobile phase.[57] This model can therefore been used to explain differences in selectivity for peptide separations as a function of gradient slope, that is, the rate of change of organic solvent, with peptides whose S curves intersect, showing reversals in retentivity either side of a critical concentration of organic solvent. The rate at which this critical concentration is reached, that is, the slope of the gradient, determines which of the two peptides is the first to elute from the column.[55] Therefore, it can be concluded that the peptides TQLIDVEIAK (**14**) and LTVLESLSK (**15**) and AGLEFGTTPEQPEETPLDDLAETDFQTFSGK (**22**) and VVSLPDFFTFSK (**23**) have S curves that intersect and thus demonstrate differential selectivity as a function of gradient slope. Practically, this is unlikely to be a problem as the observed deviation from the linear relationship between different gradients was minor. Good prediction of retention times on different gradients is thus eminently possible even when these two pairs of RePLiCal peptides are employed as reference points.

DOI: 10.1021/acs.jproteome.5b00988
J. Proteome Res. 2016, 15, 1090−1102

Testing of trapping column performance using RePLiCal

Many nano LC instruments used in proteomics experiments are operated in a trap-elute configuration, whereby peptides are loaded onto a short, larger internal diameter trapping column and then eluted onto a narrower, longer analytical column. The trapping column allows peptides to be loaded from the sample loop at higher flow rates than would be possible using the analytical column, helping to minimize band broadening. It can also serve to desalt samples, binding peptides while unwanted salts (and other nonbinding constituents present in the mixture) end up in the column effluent and are directed to waste. For this configuration to be effective peptides must partition into the stationary phase so that they are not eluted to waste; this is of particular concern for very hydrophilic peptides that are less likely to be retained. The inclusion of two very hydrophilic peptides in RePLiCal, VTASGDDSPSGK (**1**) and ALAEDEGAK (**2**), can allow trapping column performance to be tested. Figure 6 shows the signal intensities and RTs of these two peptides relative to the next eluting peptide, SSYVGDEASSK (**3**), on the same LC system with both a faulty trapping column and one that is functioning optimally. It is evident that the signal intensities for VTASGDDSPSGK and ALAEDEGAK are significantly decreased relative to SSYVGDEASSK (both <1% relative intensity) compared with when the trapping column is working effectively, under which conditions the relative intensities of VTASGDDSPSGK and ALAEDEGAK are ∼10 and 40%, respectively, compared with SSYVGDEASSK. Suboptimal peak widths for all three peptides are also observed (as would be expected). Chromatographic peak width in addition to relative signal intensity can thus be monitored as a measure of optimal trapping performance, providing multiple metrics by which trapping column efficiency can be assessed and the necessity for replacement defined.

Detection of RePLiCal in a Complex Matrix

RePLiCal peptides were spiked into a tryptic digest of a whole cell yeast lysate and analyzed by nLC−nESI−SRM−MS to test the utility of this novel standard in proteomics experiments (Figure 7). All 27 RePLiCal peptides were detectable with good signal-to-background ratio (≥29:1) when spiked at 5 fmol in a background of 1 μg of yeast tryptic peptides. The fact that RePLiCal is readily detectable in a complex matrix when added to the sample in minimal amounts (5 fmol is equivalent to ∼200 pg and represents a 0.02% increase in protein load on column) means that it can be used without itself increasing the complexity of the sample significantly. A greater quantity of RePLiCal would be required to ensure detection in nontargeted data-dependent acquisition (DDA) and data-independent acquisition (DIA) experiments, but sufficient amounts are still unlikely to increase the complexity of the sample loaded on column by >0.5%.

Repeatability of RePLiCal Performance over Extended Analysis Times

60 consecutive injections of 50 fmol RePLiCal in 1 μg of yeast tryptic peptides were loaded onto the LC column and analyzed using a 30 min LC gradient (>2 days instrument time with trapping, washing, and re-equilibration steps) to assess the repeatability of RePLiCal performance. Table 1 shows that the relative standard deviations (RSDs) of the RTs were <1.2% for all peptides, demonstrating that even over prolonged analysis times high repeatability in RT measurement is obtained. Furthermore, all but two of the peptides showed a <6% RSD in the peak width (FWHM), showing that the vast majority of the standards gave very consistent peak shapes. The two peptides that gave values exceeding 20%, TSAESILTTGPVVPVIVVK (**20**) and IAFFESSFLSYLK (**25**), were longer retained species, thus giving more scope for band broadening, which will, in part, account for the more variable peak widths. Nonetheless, all peptides retained good peak shapes with the variation in their widths in absolute terms being on the order of a few seconds. The repeatability of these experiments thus provides confidence that RePLiCal can be used to monitor LC performance longitudinally and that deviation from precise measurements in terms of RTs and peak widths will give rapid feedback on problems with LC instrumentation. The peak areas for the majority of the RePLiCal peptides also showed good precision, with 23 having RSDs of <20% (median of 7.91%). The four peptides with greater RSDs were those with the lowest average signal intensities (Supporting Information Table S2), and therefore would be expected to demonstrate greater injection-to-injection variability. These four peptides are also among the seven latest eluting species and thus represent the more hydrophobic peptides in RePLiCal. Several groups have shown that hydrophobic peptides are prone to adsorption to surfaces during storage;[47−49,58−60] therefore some of the variability observed in the peak areas is likely to be due to loss over the duration of the experiment via this mechanism.

Comparison of Formic Acid and Acetic Acids As Ion-Pairing Agents

Typically, low-pH RP−LC−MS is performed using FA as the ion-pairing agent due to its volatility and hence compatibility with ESI; however, acetic acid is used by several groups as an alternative.[61−63] For comparison, proteolyzed RePLiCal was analyzed on 10, 30, 60, and 90 min gradients using either 0.1% FA or 0.5% acetic acid to compare performance using the different ion-pairing agents. Identical elution orders were seen with both acids, indicating that no ion-pairing agent-dependent changes in selectivity occurred (Supporting Information Figure S10). Comparison of the peak widths (FWHM) showed no statistically significant differences at the 1% confidence level using the Mann−Whitney *U* test. These observations demonstrate that RePLiCal can be used as a standard irrespective of whether FA or acetic acid is used as the ion-pairing agent.

Prediction of RTs Using RePLiCal

As previously discussed, RePLiCal can, in theory, be implemented for RT prediction of a target peptide with reference to the known RT on the same or potentially different gradient characterized by RePLiCal by calculating the regression fit. In a practical sense, one might envisage RT measurement of a set of peptides on a long discovery-type LC gradient and then analysis of these peptides in a targeted manner by scheduled SRM, which typically employs shorter LC gradients and often requires a different LC−MS system. Determination of the RTs for these peptides for implementation of scheduled SRM would thus be required, a time-consuming step, particularly when methods for large numbers of peptides are being developed.[11] The elution times of RePLiCal and 100 peptides from moderate-to-high abundance proteins in yeast[40,41,64] targeted using transitions selected from SRMAtlas[65] (Supporting Information Table S9) were determined on a 30 min LC gradient to demonstrate the utility of RePLiCal in predicting RTs on different gradients. Subsequently, the RTs of the RePLiCal peptides were recorded on a 60 min LC gradient (using the same LC instrument and

DOI: 10.1021/acs.jproteome.5b00988
J. Proteome Res. 2016, 15, 1090−1102

columns) and a linear regression linking the two gradients was calculated. Using this linear regression, the RTs of the 100 yeast peptides were predicted and scheduled SRM methods using time windows of 1.5, 2, 2.5, and 3 min were created for the 60 min LC gradient. Each method was run in hexaplicate with detection of a peptide being regarded as the full elution profile of the peak being within the time window in all six analyses. Using a very narrow 1.5 min (0.93% B) time window, this criterion was only satisfied for 64% of the peptides; however, lengthening the time window to 2 min led to successful detection of 94% of the peptides. Further widening of the time windows to 2.5 and 3 min allowed successful detection of 99 and 100% of the peptides, respectively. Normalization of two different LC gradients using RePLiCal thus allows successful prediction and scheduling of RT windows for a large number of peptides whose RTs have been determined using a different LC gradient or chromatographic system, thereby allowing efficient transfer between discovery and targeted proteomics experiments.

Optimization of LC Gradient to Maximize Peptide and Protein Identification Rates

Despite the fast acquisition speeds of state-of-the-art mass spectrometers used for discovery proteomics,[66,67] many peptides in complex proteolytic digests are not selected for tandem mass spectrometry in typical top *n* DDA experiments (where *n* = the number of MS/MS events before the instrument returns to acquiring a full scan mass spectrum) due to high degrees of coelution and bias against lower intensity signals.[68] To address this problem, Moruz and coworkers proposed the implementation of nonlinear LC gradients to more evenly distribute the elution times of peptides. The likelihood of peptide coelution is thus reduced and the percentage of ions in a given MS scan selected for MS/MS increased, improving the chances of lower intensity peptides being selected during DDA and hence identified.[69] Having first acquired data for the sample of interest using a linear gradient, a nonlinear slope was optimized by determining even distributions of either MS1 features or predicted RTs for the proteome of interest using the program ELUDE.[14] The development of such optimized nonlinear gradients led to increases in peptide identifications by between, on average, 2 and 10% under different chromatography conditions; however, this approach necessitates that samples are analyzed twice, that is, with linear and nonlinear gradients. This may not be possible in sample-limited situations and puts an additional burden on instrument usage time. It was hypothesized that RePLiCal could be used to optimize nonlinear gradients as effectively as a whole proteome sample, given that the peptides fully describe a linear LC gradient. The analysis could be performed once, and the optimized nonlinear gradients applied for all subsequent experiments, only requiring reoptimization when a change to the LC system takes place, for example, when a new column is fitted. RePLiCal was therefore analyzed on three linear gradients (3−40% B) over 30, 60, and 90 min and used to generate "in-silico-optimized" and "custom distribution" nonlinear gradients (the latter being recently introduced by Moruz and Käll and involving the generation of an even distribution of inputted retention times).[70] Subsequent analysis of a whole cell yeast lysate tryptic digest using each of these two nonlinear gradients demonstrated increased identifications at both the peptide and protein levels compared with a comparable standard linear gradient (Table 2). As would be expected for a DDA experiment, there is an inverse relationship between gradient length and percentage increase in identifications: As the gradient length increases there is less coelution and hence a greater percentage of the peptide ions recorded in a given full scan mass spectrum will be selected for MS/MS. Nevertheless, even for the longest gradient considered, the nonlinear gradients provided a noticeable increase of over 10% in peptide identifications and nearly 5% in protein identifications, with the "in-silico-optimized" gradient consistently outperforming the "custom distribution". This difference in performance can be attributed to the former using chemical information, that is, the peptide sequence, as part of the optimization rather than simply evenly distributing a series of retention times, therefore providing a more refined nonlinear gradient to be used for a whole proteome sample.

Table 2. Average Numbers of Peptide and Protein Identifications (at a 1% FDR and requiring identification in all three technical replicates) from 1 μg of a Whole Cell Yeast Lysate Tryptic Digest Using a Linear and Two Nonlinear LC Gradients (3−40% 0.1% FA in MeCN)[a]

gradient					
type	time/min	average peptide identifications	% increase in peptides	average protein identifications	% increase in proteins
linear	30	2573		583	
custom distribution	30	2994	16.4	661	13.4
in silico optimized	30	3228	25.5	674	15.6
linear	60	4481		910	
custom distribution	60	4960	10.7	972	6.8
in silico optimized	60	5219	16.5	998	9.7
linear	90	6066		1166	
custom distribution	90	6676	10.1	1223	4.9
in silico optimized	90	6948	14.5	1237	6.1

[a]Percentage increases in identifications are relative to the linear gradient.

The closest like-for-like comparison with the work of Moruz and coworkers is for the "in-silico-optimized" gradient. In this study, a 14.5% increase in peptide identifications was observed for a 90 min gradient, whereas Moruz and coworkers saw only a 5.3% increase when using a 120 min gradient (albeit with a HeLa cell tryptic digest). This suggests that RePLiCal is more effective at optimizing nonlinear LC gradients than using the sample of interest itself, which is particularly advantegeous in sample-limited circumstances. Consistent with the observation of Moruz and coworkers, different populations of peptides were identified in this study using the three different gradients (Figure 8a). The data also showed a similar trend at the protein level, although the overlap between conditions was greater than that observed at the peptide level as the differentially identified peptides generally lead to the identification of the same proteins (Figure 8b).[69] The data thus support the proposal that the combination of linear and nonlinear gradients allows more comprehensive proteome coverage in shotgun proteomics experiments,[70] and preference should be given to nonlinear "in-silico-optimized" LC gradients for maximal protein (and peptide) identification rates over a single run.

DOI: 10.1021/acs.jproteome.5b00988
J. Proteome Res. 2016, 15, 1090−1102

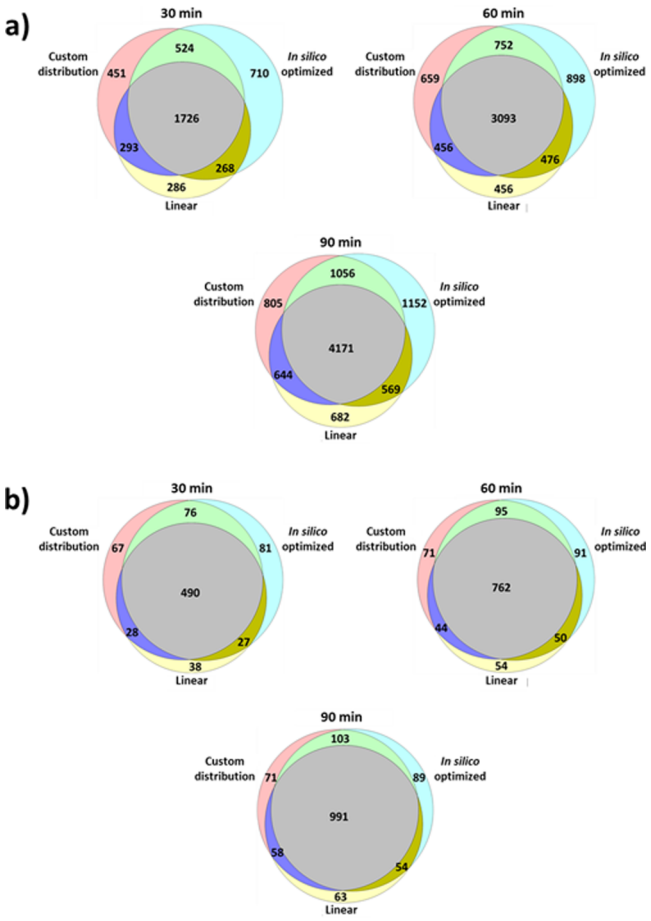


Figure 8. Venn diagrams showing the numbers and overlap for the (a) peptides and (b) proteins identified using linear and nonlinear ("custom distribution" and "in silico optimized") gradients.

Recalibration of *m/z* Scale Using RePLiCal

Consistent elution of RePLiCal peptides throughout the LC gradient prompted an investigation as to whether these peptides can also be used as lock masses to correct the *m/z* scale of high-resolution, accurate mass MS data. This approach has been previously demonstrated to improve total protein identifications in shotgun proteomics experiments.[23] RePLiCal was spiked into a whole cell yeast lysate tryptic digest and analyzed using a calibrated linear ion trap-Orbitrap mass spectrometer on a 270 min gradient. Searching the raw data enabled the identification of 8903 peptides at a 1% FDR, which were attributed to a total of 1083 proteins. The raw data were then miscalibrated by 0.125 *m/z* units using the middle eluting peptide from RePLiCal, TQLIDVEIAK, as the reference point. The miscalibrated data were then split into 27 sections, each containing a single RePLiCal peptide at the midpoint of the section in the time dimension. The exact *m/z* of the RePLiCal peptide was used to lock-mass-correct the data in the truncated section of the LC gradient, following which the 27 sections of the chromatogram were recombined and searched using the same database search parameters as for the raw data. At a 1% FDR, 8903 peptides were identified, leading to the identification of 1083 proteins, with high overlap with the data premiscalibration (1082 common proteins, 8810 common peptides). While an improvement in protein identifications was not observed as described by Mirzaei and coworkers (this may be a function of the quality of the initial instrument *m/z* scale calibration), the use of RePLiCal essentially allowed the rebuilding of the acquired raw data following miscalibration. This orchestrated scenario replicates the drifting or complete loss of calibration during an analysis and demonstrates the utility of RePLiCal to prevent data loss. This enables more efficient use of instrument time as reanalysis is not required and prevents complete loss of data in sample-limited situations. The addition of a standard such as RePLiCal in this situation is particularly advantageous for poorly characterized samples where knowledge of the expected endogeneous peptides, which could be used for *m/z* scale recalibration, is not available.

CONCLUSIONS

A retention time standard, RePLiCal, generated using QconCAT technology and complementing existing standards for mass spectrometry[71] and ion mobility[72] instrumentation, has been presented. It was demonstrated that RePLiCal can more effectively standardize nUHPLC instrumentation than three commercially available standards through greater coverage of the LC gradient, particularly at the start and end. Furthermore, RePLiCal can be used to identify poor trapping column performance due to the presence of two very hydrophilic peptides that are only trapped effectively when instrument performance is optimal. The standard has been analyzed on two different nUHPLC instruments and on a variety of gradient lengths and has performed stably and in a predictable manner, permitting transfer of peptide RTs between LC systems. RePLiCal has also been used effectively to generate nonlinear gradients to maximize peptide and protein identifications in nontargeted proteomics experiments, and the individual peptides have proven useful as evenly distributed reference points throughout the LC gradient to recalibrate the *m/z* scale postacquisition. It is envisaged that this standard could be used to benchmark RP LC instrument performance across laboratories, particularly given that the same peptides are generated by both Lys-C and trypsin digestion. Finally, given that the RePLiCal peptide sequences were designed so that they are not naturally occurring in any eukaryotic organism, this standard can be introduced into the vast majority of proteomics samples, permitting standardization across almost all proteomics experiments.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.5b00988.

Supplementary experimental details and Supplementary Figures S1−S7. (PDF)

Supplementary Tables S1−S9. (XLS)

AUTHOR INFORMATION

Corresponding Author

*E-mail: claire.eyers@liverpool.ac.uk. Tel: +44 151 795 4424.

Notes

The authors declare the following competing financial interest(s): RePLiCal is now being marketed and sold as a retention time standard for HPLC instruments by PolyQuant GMBH (http://www.polyquant.com/calibration_standards_for_proteomics/), for which S.W.H. and C.E.E. hold a joint patent application.

DOI: 10.1021/acs.jproteome.5b00988
J. Proteome Res. 2016, 15, 1090−1102

■ ACKNOWLEDGMENTS

This work was financially supported by the U.K. Biotechnology and Biological Sciences Research Council (BBSRC) under grant BB/G009058/1. We thank Dr. Markus Fischer (InnVentis) for assistance with the BLAST searching, Dr. Karin Lanthaler (University of Manchester) for preparation of the yeast sample, and Dr. Philip Brownridge (University of Liverpool) for technical assistance relating to the experiments performed on the Ultimate 3000 RSLC nUHPLC instrument.

■ REFERENCES

(1) Rudnick, P. A.; Clauser, K. R.; Kilpatrick, L. E.; Tchekhovskoi, D. V.; Neta, P.; Blonder, N.; Billheimer, D. D.; Blackman, R. K.; Bunk, D. M.; Cardasis, H. L.; Ham, A.-J. L.; Jaffe, J. D.; Kinsinger, C. R.; Mesri, M.; Neubert, T. A.; Schilling, B.; Tabb, D. L.; Tegeler, T. J.; Vega-Montoto, L.; Variyath, A. M.; Wang, M.; Wang, P.; Whiteaker, J. R.; Zimmerman, L. J.; Carr, S. A.; Fisher, S. J.; Gibson, B. W.; Paulovich, A. G.; Regnier, F. E.; Rodriguez, H.; Spiegelman, C.; Tempst, P.; Liebler, D. C.; Stein, S. E. Performance metrics for liquid chromatography-tandem mass spectrometry systems in proteomics analyses. *Mol. Cell. Proteomics* **2010**, *9*, 225−241.

(2) Beasley-Green, A.; Bunk, D.; Rudnick, P.; Kilpatrick, L.; Phinney, K. A proteomics performance standard to support measurement quality in proteomics. *Proteomics* **2012**, *12*, 923−931.

(3) Bereman, M. S. Tools for monitoring system suitability in LC MS/MS centric proteomic experiments. *Proteomics* **2015**, *15*, 891−902.

(4) Pichler, P.; Mazanek, M.; Dusberger, F.; Weilnböck, L.; Huber, C. G.; Stingl, C.; Luider, T. M.; Straube, W. L.; Köcher, T.; Mechtler, K. SIMPATIQCO: A server-based software suite which facilitates monitoring the time course of LC-MS performance metrics on Orbitrap instruments. *J. Proteome Res.* **2012**, *11*, 5540−5547.

(5) Scheltema, R. A.; Mann, M. SprayQC: A real-time LC-MS/MS quality monitoring system to maximize uptime using off the shelf components. *J. Proteome Res.* **2012**, *11*, 3458−3466.

(6) Paulovich, A. G.; Billheimer, D.; Ham, A.-J. L.; Vega-Montoto, L.; Rudnick, P. A.; Tabb, D. L.; Wang, P.; Blackman, R. K.; Bunk, D. M.; Cardasis, H. L.; Clauser, K. R.; Kinsinger, C. R.; Schilling, B.; Tegeler, T. J.; Variyath, A. M.; Wang, M.; Whiteaker, J. R.; Zimmerman, L. J.; Fenyo, D.; Carr, S. A.; Fisher, S. J.; Gibson, B. W.; Mesri, M.; Neubert, T. A.; Regnier, F. E.; Rodriguez, H.; Spiegelman, C.; Stein, S. E.; Tempst, P.; Liebler, D. C. Interlaboratory study characterizing a yeast performance standard for benchmarking LC-MS platform performance. *Mol. Cell. Proteomics* **2010**, *9*, 242−254.

(7) Ivanov, A. R.; Colangelo, C. M.; Dufresne, C. P.; Friedman, D. B.; Lilley, K. S.; Mechtler, K.; Phinney, B. S.; Rose, K. L.; Rudnick, P. A.; Searle, B. C.; Shaffer, S. A.; Weintraub, S. T. Interlaboratory studies and initiatives developing standards for proteomics. *Proteomics* **2013**, *13*, 904−909.

(8) Gallien, S.; Bourmaud, A.; Domon, B. A simple protocol to routinely assess the uniformity of proteomics analyses. *J. Proteome Res.* **2014**, *13*, 2688−2695.

(9) Lai, X.; Wang, L.; Tang, H.; Witzmann, F. A. A novel alignment method and multiple filters for exclusion of unqualified peptides to enhance label-free quantification using peptide intensity in LC-MS/MS. *J. Proteome Res.* **2011**, *10*, 4799−4812.

(10) Zhang, W.; Zhang, J.; Xu, C.; Li, N.; Liu, H.; Ma, J.; Zhu, Y.; Xie, H. LFQuant: A label-free fast quantitative analysis tool for high-resolution LC-MS/MS proteomics data. *Proteomics* **2012**, *12*, 3475−3484.

(11) Holman, S. W.; Sims, P. F. G.; Eyers, C. E. The use of selected reaction monitoring in quantitative proteomics. *Bioanalysis* **2012**, *4*, 1763−1786.

(12) Klammer, A. A.; Yi, X.; MacCoss, M. J.; Noble, W. S. Improving tandem mass spectrum identification using peptide retention time prediction across diverse chromatography conditions. *Anal. Chem.* **2007**, *79*, 6111−6118.

(13) Pfeifer, N.; Leinenbach, A.; Huber, C. G.; Kohlbacher, O. Statistical learning of peptide retention behaviour in chromatographic separations: a new kernel-based approach for computational proteomics. *BMC Bioinf.* **2007**, *8*, 468.

(14) Moruz, L.; Tomazela, D.; Käll, L. Training, selection, and robust calibration of retention time models for targeted proteomics. *J. Proteome Res.* **2010**, *9*, 5209−5216.

(15) Cao, W.; Ma, D.; Kapur, A.; Patankar, M. S.; Ma, Y.; Li, L. RT-SVR+q: A strategy for post-mascot analysis using retention time and q value metric to improve peptide and protein identifications. *J. Proteomics* **2011**, *75*, 480−490.

(16) Moruz, L.; Staes, A.; Foster, J. M.; Hatzou, M.; Timmerman, E.; Martens, L.; Käll, L. Chromatographic retention time prediction for posttranslationally modified peptides. *Proteomics* **2012**, *12*, 1151−1159.

(17) Krokhin, O. V.; Craig, R.; Spicer, V.; Ens, W.; Standing, K. G.; Beavis, R. C.; Wilkins, J. A. An improved model for prediction of retention times of tryptic peptides in ion pair reversed-phase HPLC: Its application to protein peptide mapping by off-line HPLC-MALDI MS. *Mol. Cell. Proteomics* **2004**, *3*, 908−919.

(18) Gorshkov, A. V.; Tarasova, I. A.; Evreinov, V. V.; Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A.; Gorshkov, M. V. Liquid chromatography at critical conditions: Comprehensive approach to sequence-dependent retention time prediction. *Anal. Chem.* **2006**, *78*, 7770−7777.

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

(20) Spicer, V.; Yamchuk, A.; Cortens, J.; Sousa, S.; Ens, W.; Standing, K. G.; Wilkins, J. A.; Krokhin, O. V. Sequence-specific retention calculator. A family of peptide retention time prediction algorithsm in reversed-phase HPLC: Applicability to various chromatographic conditions and columns. *Anal. Chem.* **2007**, *79*, 8762−8768.

(21) Tarasova, I. A.; Guryča, V.; Pridatchenko, M. L.; Gorshkov, A. V.; Kieffer-Jaquinod, S.; Evreinov, V. V.; Masselon, C. D.; Gorshkov, M. V. Gorshkov Standardization of retention time data for AMT tag proteomics database generation. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2009**, *877*, 433−440.

(22) Burkhart, J. M.; Premsler, T.; Sickmann, A. Quality control of nano-LC-MS systems using stable isotope-coded peptides. *Proteomics* **2011**, *11*, 1049−1057.

(23) Mirzaei, H.; Brusniak, M.-Y.; Mueller, L. N.; Letarte, S.; Watts, J. D.; Aebersold, R. Halogenated peptides as internal standards (H-PINS): Introduction of an MS-based internal standard set for liquid chromatography-mass spectrometry. *Mol. Cell. Proteomics* **2009**, *8*, 1934−1946.

(24) Krokhin, O. V.; Spicer, V. Peptide retention standards and hydrophobicity indexes in reversed-phase high-performance liquid chromatography of peptides. *Anal. Chem.* **2009**, *81*, 9522−9530.

(25) Escher, C.; Reiter, L.; MacLean, B.; Ossola, R.; Herzog, F.; Chilton, J.; MacCoss, M. J.; Rinner, O. Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* **2012**, *12*, 1111−1121.

(26) Gallien, S.; Peterman, S.; Kiyonami, R.; Souady, J.; Duriez, E.; Schoen, A.; Domon, B. Highly multiplexed targeted proteomics using precise control of peptide retention time. *Proteomics* **2012**, *12*, 1122−1133.

(27) Parker, S. J.; Rost, H.; Rosenberger, G.; Collins, B. C.; Malmström, L.; Amodei, D.; Venkatraman, V.; Raedschelders, K.; Van Eyk, J. E.; Aebersold, R. Identification of a set of conserved eukaryotic internal retention time standards for data-independent acquisition mass spectrometry. *Mol. Cell. Proteomics* **2015**, *14*, 2800.

(28) Beynon, R. J.; Doherty, M. K.; Pratt, J. M.; Gaskell, S. J. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nat. Methods* **2005**, *2*, 587−589.

DOI: 10.1021/acs.jproteome.5b00988
J. Proteome Res. 2016, 15, 1090−1102

(29) Pratt, J. M.; Simpson, D. M.; Doherty, M. K.; Rivers, J.; Gaskell, S. J.; Beynon, R. J. Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat. Protoc.* **2006**, *1*, 1029−1043.
(30) Bateman, R. H.; Carruthers, R.; Hoyes, J. B.; Jones, C.; Langridge, J. I.; Millar, A.; Vissers, J. P. C. A novel precursor ion discovery method on a hybrid quadrupole orthogonal acceleration time-of-flight (Q-TOF) mass spectrometer for studying protein phosphorylation. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 792−803.
(31) Remily-Wood, E. R.; Koomen, J. M. Evaluation of protein quantification using standard peptides containing single conservative amino acid replacements. *J. Mass Spectrom.* **2012**, *47*, 188−194.
(32) Schechter, I.; Berger, A. On the size of the active site in proteases. I. Papain. *Biochem. Biophys. Res. Commun.* **1967**, *27*, 157−162.
(33) Winter, D.; Pipkorn, R.; Lehmann, W. D. Separation of peptide isomers and conformers by ultra performance liquid chromatography. *J. Sep. Sci.* **2009**, *32*, 1111−1119.
(34) Winter, D.; Seidler, J.; Kugelstadt, D.; Derrer, B.; Kappes, B.; Lehmann, W. D. Minimally permutated peptide analogs as internal standards for relative and absolute quantification of peptides and proteins. *Proteomics* **2010**, *10*, 1510−1514.
(35) Lawless, C.; Hubbard, S. J. Prediction of missed proteolytic cleavages for the selection of surrogate peptide for quantitative proteomics. *OMICS* **2012**, *16*, 449−456.
(36) Fannes, T.; Vandermarliere, E.; Schietgat, L.; Degroeve, S.; Martens, L.; Ramon, J. Predicting tryptic cleavage from proteomics data using decision tree ensembles. *J. Proteome Res.* **2013**, *12*, 2253−2259.
(37) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403−410.
(38) Frank, R. Spot-synthesis: An easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. *Tetrahedron* **1992**, *48*, 9217−9232.
(39) Frank, R. The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports - principles and applications. *J. Immunol. Methods* **2002**, *267*, 13−26.
(40) Brownridge, P.; Lawless, C.; Payapilly, A. B.; Lanthaler, K.; Holman, S. W.; Harman, V. M.; Grant, C. M.; Beynon, R. J.; Hubbard, S. J. Quantitative analysis of chaperone network throughput in budding yeast. *Proteomics* **2013**, *13*, 1276−1291.
(41) Lawless, C.; Holman, S. W.; Brownridge, P.; Lanthaler, K.; Harman, V. M.; Watkins, R.; Hammond, D. E.; Miller, R. L.; Sims, P. F. G.; Grant, C. M.; Eyers, C. E.; Beynon, R. J.; Hubbard, S. J. Absolute quantification of over 1800 yeast proteins provides insights into translational control and protein dynamics. *Mol. Cell. Proteomics* **2016**, mcp.M115.054288.
(42) Yu, Y.-Q.; Gilar, M.; Lee, P. J.; Bouvier, E. S. P.; Gebler, J. C. Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins. *Anal. Chem.* **2003**, *75*, 6023−6028.
(43) Glatter, T.; Ludwig, C.; Ahrné, E.; Aebersold, R.; Heck, A. J. R.; Schmidt, A. Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *J. Proteome Res.* **2012**, *11*, 5145−5156.
(44) Achour, B.; Barber, J. The activities of *Achromobacter* lysyl endopeptidase and *Lysobacter* lysyl endoproteinase as digestive enzymes for quantitative proteomics. *Rapid Commun. Mass Spectrom.* **2013**, *27*, 1669−1672.
(45) Vandermarliere, E.; Mueller, M.; Martens, L. Getting intimate with trypsin, the leading protease in proteomics. *Mass Spectrom. Rev.* **2013**, *32*, 453−465.
(46) Brownridge, P.; Holman, S. W.; Gaskell, S. J.; Grant, C. M.; Harman, V. M.; Hubbard, S. J.; Lanthaler, K.; Lawless, C.; O'cualain, R.; Sims, P.; Watkins, R.; Beynon, R. J. Global absolute quantification of a proteome: Challenges in the deployment of a QconCAT strategy. *Proteomics* **2011**, *11*, 2957−2970.
(47) Pezeshki, A.; Vergote, V.; Van Dorpe, S.; Baert, B.; Burvenich, C.; Popkov, A.; De Spiegeleer, B. Adsorption of peptides at the sample drying step: Influence of solvent evaporation technique, vial material and solution additive. *J. Pharm. Biomed. Anal.* **2009**, *49*, 607−612.
(48) Kraut, A.; Marcellin, M.; Adrait, A.; Kuhn, L.; Louwagie, M.; Kieffer-Jaquinod, S.; Lebert, D.; Masselon, C. D.; Dupuis, A.; Bruley, C.; Jaquinod, M.; Garin, J.; Gallagher-Gambarelli, M. Peptide storage: Are you getting the best return on your investment? Defining optimal storage conditions for proteomics samples. *J. Proteome Res.* **2009**, *8*, 3778−3785.
(49) Kawashima, Y.; Takahashi, N.; Satoh, M.; Saito, T.; Kado, S.; Nomura, F.; Matsumoto, H.; Kodera, Y. Enhanced recovery of lyophilized peptides in shotgun proteomics by using an LC-ESI-MS compatible surfactant. *Proteomics* **2013**, *13*, 751−755.
(50) Malmström, L.; Malmström, J.; Selevsek, N.; Rosenberger, G.; Aebersold, R. Automated workflow for large-scale selected reaction monitoring experiments. *J. Proteome Res.* **2012**, *11*, 1644−1653.
(51) Petritis, K.; Kangas, L. J.; Yan, B.; Monroe, M. E.; Strittmatter, W.-J. Q.; Adkins, J. N.; Qian, W.-J.; Moore, R. J.; Xu, Y.; Lipton, M. S.; Camp, D. G., II; Smith, R. D. Improved peptide elution time prediction for reversed-phase liquid chromatography-MS by incorporating peptide sequence information. *Anal. Chem.* **2006**, *78*, 5026−5039.
(52) Gilar, M.; Jaworski, A.; Olivova, P.; Gebler, J. C. Peptide retention prediction applied to proteomic data analysis. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 2813−2821.
(53) Reimer, J.; Spicer, V.; Krokhin, O. V. Application of modern reversed-phase peptide retention prediction algorithms to the Houghten and DeGraw dataset: Peptide helicity and its effect on prediction accuracy. *J. Chromatogr. A* **2012**, *1256*, 160−168.
(54) Stadalius, M. A.; Gold, H. S.; Snyder, L. R. Optimization model for the gradient elution separation of peptide mixtures by reversed-phase high-performance liquid chromatography. *J. Chromatogr.* **1984**, *296*, 31−59.
(55) Spicer, V.; Grigoryan, M.; Gotfrid, A.; Standing, K. G.; Krokhin, O. V. Predicting retention time shifts associated with variation of the gradient slope in peptide RP-HPLC. *Anal. Chem.* **2010**, *82*, 9678−9685.
(56) Tarasova, I. A.; Perlova, T. Y.; Pridatchenko, M. L.; Goloborod'ko, A. A.; Levitsky, L. I.; Evreinov, V. V.; Guryca, V.; Masselon, C. D.; Gorshkov, A. V.; Gorshkov, M. V. Inversion of chromatographic elution orders of peptides and its importance for proteomics. *J. Anal. Chem.* **2012**, *67*, 1014−1025.
(57) Glajch, J. L.; Quarry, M. A.; Vasta, J. F.; Snyder, L. R. Separation of peptide mixtures by reversed-phase gradient elution. Use of flow rate changes for controlling band spacing and improving resolution. *Anal. Chem.* **1986**, *58*, 280−285.
(58) Stejskal, K.; Potěšil, D.; Zdráhal, Z. *J. Proteome Res.* **2013**, *12* (Suppression of peptidesample losses in autosampler vials), 3057−3062.
(59) Warwood, S.; Byron, A.; Humphries, M. J.; Knight, D. The effect of peptide adsorption on signal intensity and a simple approach to improve reliability of quantification. *J. Proteomics* **2013**, *85*, 160−164.
(60) Magdeldin, S.; Moresco, J. J.; Yamamoto, T.; Yates, J. R., III Off-line multidimensional liquid chromatography and auto sampling result in sample loss in LC/LC-MS/MS. *J. Proteome Res.* **2014**, *13*, 3826−3836.
(61) Michalski, A.; Neuhauser, N.; Cox, J.; Mann, M. A systematic investigation into the nature of tryptic HCD spectra. *J. Proteome Res.* **2012**, *11*, 5479−5491.
(62) Subramonian, D.; Raghunayakula, S.; Olsen, J. V.; Beningo, K. A.; Paschen, W.; Zhang, X.-D. Analysis of changes in SUMO-2/3 modification during breast cancer progression and metastasis. *J. Proteome Res.* **2014**, *13*, 3905−3918.
(63) Matic, K.; Eninger, T.; Bardoni, B.; Davidovic, L.; Macek, B. Quantitative phosphoproteomics of murine *Fmr1*-KO cell lines provides new insights into FMRP-dependent signal transduction mechanisms. *J. Proteome Res.* **2014**, *13*, 4388−4397.

DOI: 10.1021/acs.jproteome.5b00988
J. Proteome Res. 2016, 15, 1090−1102

(64) Wang, M.; Weiss, M.; Simonovic, M.; Haertinger, G.; Schrimpf, S. P.; Hengartner, M. O.; von Mering, C. PaxDb, a database of protein abundance averages across all three domains of life. *Mol. Cell. Proteomics* **2012**, *11*, 492−500.
(65) Picotti, P.; Clément-Ziza, M.; Lam, H.; Campbell, D. S.; Schmidt, A.; Deutsch, E. W.; Röst, H.; Sun, Z.; Rinner, O.; Reiter, L.; Shen, Q.; Michaelson, J. J.; Frei, A.; Alberti, S.; Kusebauch, U.; Wollscheid, B.; Moritz, R. L.; Beyer, A.; Aebersold, R. A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* **2013**, *494*, 266−270.
(66) Hebert, A. S.; Richards, A. L.; Bailey, D. J.; Ulbrich, A.; Coughlin, E. E.; Westphall, M. S.; Coon, J. J. The one hour yeast proteome. *Mol. Cell. Proteomics* **2014**, *13*, 339−347.
(67) Kelstrup, C. D.; Jersie-Christensen, R. R.; Batth, T. S.; Arrey, T. N.; Kuehn, A.; Kellmann, M.; Olsen, J. V. Rapid and deep proteomes by faster sequencing on a benchtop quadrupole ultra-high-field Orbitrap mass spectrometer. *J. Proteome Res.* **2014**, *13*, 6187−6195.
(68) Michalski, A.; Cox, J.; Mann, M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J. Proteome Res.* **2011**, *10*, 1785−1793.
(69) Moruz, L.; Pichler, P.; Stranzl, T.; Mechtler, K.; Käll, L. Optimized nonlinear gradients for reversed-phase liquid chromatography in shotgun proteomics. *Anal. Chem.* **2013**, *85*, 7777−7785.
(70) Moruz, L.; Käll, L. GradientOptimizer: An open-source graphical environment for calculating optimized gradients in reversed-phase liquid chromatography. *Proteomics* **2014**, *14*, 1464−1466.
(71) Eyers, C. E.; Simpson, D. M.; Wong, S. C. C.; Beynon, R. J.; Gaskell, S. J. QCAL - a novel standard for assessing instrument conditions for proteome analysis. *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1275−1280.
(72) Chawner, R.; McCullough, B.; Giles, K.; Barran, P. E.; Gaskell, S. J.; Eyers, C. E. QconCAT standard for calibration of ion mobility-mass spectrometry systems. *J. Proteome Res.* **2012**, *11*, 5564−5572.
(73) Beynon, R. J. A simple tool for drawing proteolytic peptide maps. *Bioinformatics* **2005**, *21*, 674−675.

DOI: 10.1021/acs.jproteome.5b00988
J. Proteome Res. 2016, 15, 1090−1102