Journal of proteome research

Open Access

pubs.acs.org/jpr

Article

Enhanced Proteomics Analysis with a Novel Recombinant Chymotrypsin Analogue Engineered for High Cleavage Specificity

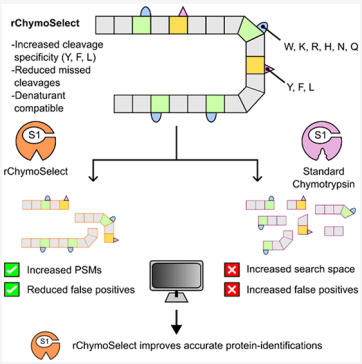
Kish R. Adoni, Jonathan E. Ditcham, Alba Katiria González Rivera, Georgina H. Charlton, Sergei V. Saveliev, Konstantinos Thalassinos,* and Riccardo Zenezini Chiozzi*

Cite This: *J. Proteome Res.* 2026, 25, 2511−2519

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Chymotrypsin is widely used in shotgun proteomics, owing to its orthogonal cleavage specificity relative to trypsin, which enhances sequence coverage of hydrophobic protein regions. However, commercial preparations often display variable cleavage specificity, trypsin contamination, and elevated missed cleavage rates, which can collectively reduce the proteome coverage and data reproducibility. To address these limitations, we present a novel recombinant chymotrypsin (*rChymoSelect*) engineered for improved cleavage specificity and robustness in proteomics workflows. Benchmarking against standard bovine chymotrypsin revealed 97% C-terminal cleavages after tyrosine (Y), phenylalanine (F), and leucine (L) for *rChymoSelect*, compared to 72% for the standard enzyme. This enhanced cleavage specificity reduced missed cleavages and increased peptide-spectrum matches across charge states. Across 3,638 identified proteins, *rChymoSelect* yielded 22.2% unique identifications compared with 8.2% for standard chymotrypsin, while maintaining similar peptide length, m/z, and hydrophobicity distributions. The enzyme remained active in up to 6 M urea and achieved near-maximal proteome coverage within 2 h (only a 2.4% gain after overnight digestion). These results establish *rChymoSelect* as an advanced tool with improved cleavage specificity that reduces analytical complexity and enhances the reliability of proteomic analysis, while expanding chymotryptic digestion to hydrophobic and high-denaturant proteomics applications.



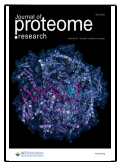
KEYWORDS: *chymotrypsin, proteomics, protein digestion, sample preparation, missed cleavages*

INTRODUCTION

Recent advancements in sample preparation methodologies, liquid chromatography, mass spectrometry, and data processing workflows have spearheaded a revolution in the field of proteomics. Together, these developments have propelled this technique to represent the gold standard for high-throughput protein identification.[1,2] During proteomics sample preparation, trypsin is commonly used as the enzyme of choice for the conversion of proteins into appropriately charged and sized peptides for subsequent proteomics mass spectrometry analysis. Since trypsin cleaves specifically at the C-terminus of lysine and arginine, there are some limitations to its utility for protein analysis, particularly in the context of hydrophobic proteins, membrane proteins, proteins with excessive lysine/arginine amino acids in the primary sequence, and proteins with excessive post-translational modifications at lysine/arginine residues.[3]

To this end, chymotrypsin provides a powerful orthogonal approach toward proteomics sample preparation.[4] While both proteases share the same serine-protease catalytic triad, via serine (S195), histidine (H57), and aspartic acid (D102), their differences are driven by the differential substrate binding pocket (S1), with the negatively charged aspartic acid of trypsin's S1 pocket facilitating hydrolysis at the C-terminus of lysine (K) and arginine (R). Conversely, the deeply hydrophobic nature of chymotrypsin's S1 binding pocket facilitates its preference for cleavage at larger, hydrophobic residues such as phenylalanine (F), tyrosine (Y), and tryptophan (W).[5] Consequently, chymotrypsin targets the C-terminus of residues: phenylalanine, tryptophan, tyrosine, methionine (M), and leucine (L). As such, these properties have been leveraged for more comprehensive hydrophobic protein digestion.[6,7] Common limitations of chymotrypsin in the context of proteomics include its promiscuity of cleavage specificity and the increased occurrence of missed cleavages.[8] Furthermore, chymotrypsin is traditionally derived from bovine pancreatic tissue, potentially leading to contamination by trypsin in commercial products. Several manufacturers note this explicitly, as some commercial formulations are treated with TLCK to inhibit trypsin activity.[9] The increased heterogeneity of this digested peptide population, from chymotrypsin digestion, can greatly hinder the data processing step of proteomics experiments, on account of increased

Received: December 19, 2025
Revised: February 24, 2026
Accepted: March 12, 2026
Published: April 10, 2026



https://doi.org/10.1021/acs.jproteome.5c01262
J. Proteome Res. 2026, 25, 2511−2519

computational search space requirements, leading to reduced and even false peptide identifications.[10]

To address these limitations, we evaluated rChymoSelect MS grade Protease (Promega), a commercially available recombinant chymotrypsin engineered by Promega for enhanced cleavage specificity at Y, F, and L and reduced nonspecific and missed cleavages relative to bovine chymotrypsin. In this work, we benchmarked rChymoSelect against standard chymotrypsin in the context of high-throughput proteomics. Our findings demonstrate that rChymoSelect's enhanced cleavage specificity boosts efficiency and reproducibility for modern proteomics workflows, leading to increased protein identifications and protein sequence coverage, relative to standard chymotrypsin, without affecting the physicochemical properties of the digested peptides.

MATERIALS AND METHODS

rChymoSelect Production

rChymoSelect MS grade Protease (Promega) is a commercially available recombinant chymotrypsin engineered for enhanced specificity toward Y, F, and L residues with reduced nonspecific cleavage relative to bovine chymotrypsin. Detailed sequence information and molecular features are proprietary to Promega. The enzyme was provided by Promega at a concentration of 1 μg/μL and stored at −80 °C. Prior to use, aliquots were thawed on ice and used without further preparation. Prior to use, aliquots were thawed on ice and used without further preparation.

Protein Preparation

For all experiments relating to the benchmarking of rChymoSelect against standard chymotrypsin (V1061, Promega), a standard human K562 intact extract (10 mg/mL of human cell K562 lysed in 6 M Urea, Promega V6941) was used. For comparison of rChymoSelect to trypsin (Trypsin Gold, V5280, Promega); human bone osteosarcoma epithelial (U2-OS) cells were lysed in 100 mM Tris-Cl (pH 8.5), 5% sodium dodecyl sulfate (SDS), 5 mM Tris(2-carboxyethyl)phosphine hydrochloride (TCEP), and 20 mM chloroacetamide. Samples were boiled for 10 min and then cooled to room temperature before being subjected to ultrasonic bath sonication for 10 min. Protein content was quantified via Bicinchoninic Acid (BCA) assay[11] (Thermo Fisher Scientific).

Optimization of in-Solution and SP3 Digestion

20 μg of K562 lysate was diluted to a final concentration of 2, 4, and 6 M urea for rChymoSelect, 2 and 0.6 M urea for standard chymotrypsin, and 0.6 M urea for trypsin. All digestions were performed at 25 °C for in-solution digestion at 2- or 16 h using a thermoshaker (Eppendorf) at 1500 rpm For SP3 digestion, the samples were prepared in a KingFisher APEX robot (Thermo Fisher Scientific) using a protocol from Koenig et al.,[12] with the following modifications: the 96-well comb was stored in plate 1, plate 2 contained reconstituted sample in 70% acetonitrile with magnetic MagReSyn Hydroxyl beads in a protein/bead ratio of 1:2. Washing solutions were in plates 3−5 (95% acetonitrile) and plates 6−7 (70% ethanol). Plate 8 contained 200 μL of digestion solution, which consisted of protease (rChymoSelect or standard chymotrypsin) in 50 mM ammonium bicarbonate (pH 8.5). For this optimization, different protease/protein ratios (1:100, 1:40, 1:20) and guanidinium hydrochloride (GND) concentrations (0.1, 0.2, and 0.4 M) were tested, with incubation for 2 h, at 37 °C. For both digestion protocols, protease activity was quenched by acidification with trifluoroacetic acid (TFA) to pH 2 before the resulting peptide-mixture was desalted on an OASIS HLB 96-well plate (Waters), before drying *in-vacuo* using a Savant DNA120 (Thermo Fisher Scientific).

Comparison of Trypsin and rChymoSelect with Deep Offline Fractionation

200 μg of proteins from U2-OS cells lysed in SDS buffer were prepared using the SP3 lysis method and digested with either trypsin overnight or rChymoSelect for 2 h at 37 °C. Digestion was quenched by TFA, and peptides were desalted using the OASIS HLB 96-well plate (Waters). Peptides were then fractionated with a Vanquish HPLC (Thermo Fisher Scientific) using an Acquity BEH C18 column (2.1 × 50 mm with 1.7 μm particles, Waters), mobile phase A: 10 mM ammonium formate, pH 10, mobile phase B: 80% acetonitrile, 10 mM ammonium formate pH10, with a flow rate of 1000 μL/min. Twenty-four fractions were collected across a 6 min gradient. Fractions were then dried and reconstituted in 0.5% TFA before undergoing liquid chromatography−tandem mass spectrometry (LC-MS/MS) analysis.

LC-MS/MS

Peptides were reconstituted in 0.5% TFA, and 500 ng were injected on an UltiMate 3000 RSLCnano liquid chromatography system (Thermo Fisher Scientific). A 5.5 cm μPAC Neo HPLC analytical column (COL-CAPHTNEOB) was connected to a Silica Tip emitter. Column temperature was set at 45 °C, and column flow rate was set to 1500 nL/min. Mobile phase A (0.1% formic acid) and mobile phase B (80% acetonitrile, 0.1% formic acid) were applied with an elution gradient from 1.0 to 35.0% mobile phase B over either 25 min (for the fractionations) or 51 min (protocol optimization). Peptides were ionized by using a spray voltage of 2.0 kV (275 °C). The mass spectrometer was set to acquire full-scan MS spectra (350 to 1400 m/z) for a maximum injection time set to Auto at a resolution of 120,000 and an automated gain control (AGC) target value of 250%. The instrument was set up in DDA with a Wide Window Acquisition (fourth window) for all the precursors with a charge state in the range of 2+ to 5+; most abundant ions were then fragmented in HCD (30% normalized collision energy) with MS2 acquisition at a resolution of 30,000, AGC: 400%, maximum injection time: dynamic, dynamic exclusion: 20s. For protocol optimization with Data-Independent Acquisition: 56 windows of 12 Da (m/z range 361−1033) were fragmented via HCD with normalized energy: 30%, AGC: 1000% and resolution: 15,000.

DATA ANALYSIS

MS data were searched against the Human SWISS-Prot protein database (August 2023) with SAGE proteomics software for downstream analysis[13] (for DDA). The increased capacity to handle a large search space (due to increased cleavage sites and missed cleavages), as well as its fast performance, made it an appropriate software choice for this work.[1,13] For cleavage specificity analysis, the following parameters were set: cleavage sites: F, Y, W, L, K, R, H, N, Q (except before P), missed cleavages: 6. For onward analysis, cleavage sites: FYWL (except before P), missed cleavages: 3. For all searches, variable modifications: [+15.9949, M]; static modifications: [+57.0215, C], max variable modifications: 3. PSMs with q-value <0.01 were selected for onward analysis, ambiguous proteins and contaminants were filtered out. Further, all validated proteins were identified in at least 2 of 3 biological replicates. Raw SAGE search outputs were processed with the PickedGroupFDR pipeline[14] to produce a combined_protein.tsv file that was used for protein quantification and downstream enrichment analyses. Peptide−protein rollup and protein group collapse were performed using the PickedGroupFDR implementation. Downstream statistics, plotting, and figure generation were performed in GraphPad Prism.[15] For Data-Independent Analysis (DIA), RAW data was processed with FragPipe v23.1,[16] with enzyme: chymotrypsin cleavage sites: F, L, W, Y, (except before P), missed cleavages: 1, variable modifications: [+15.9949, M], [+42.0106, Nt], fixed

https://doi.org/10.1021/acs.jproteome.5c01262
J. Proteome Res. 2026, 25, 2511−2519

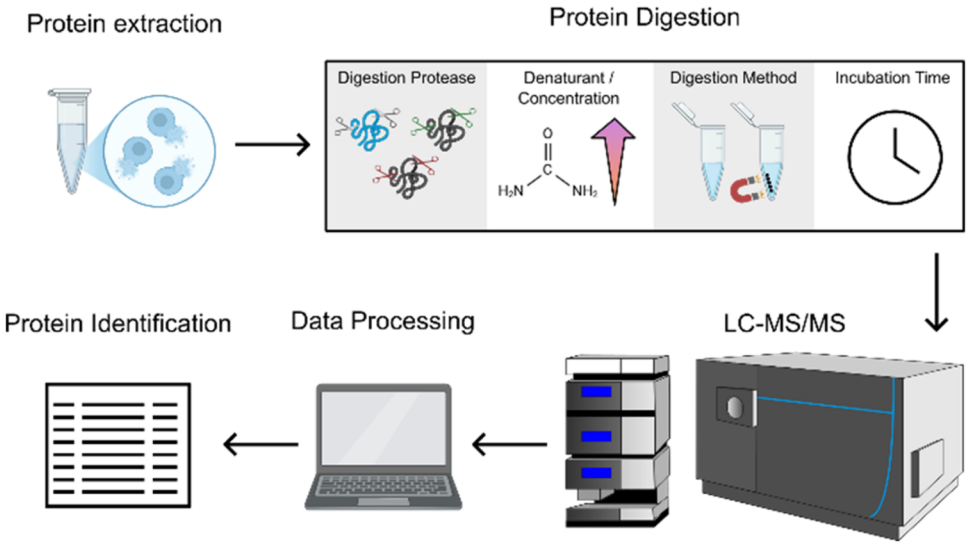


Figure 1. Benchmarking workflow for rChymoSelect vs standard chymotrypsin. Digestion enzyme (rChymoSelect vs chymotrypsin), buffer denaturant concentration (urea: 2, 4, 6 M, guanidinium hydrochloride: 0.1, 0.2, 0.4 M), digestion methodology (in-solution vs SP3), and incubation time (2 h vs 16 h) were all investigated against intact human protein lysate (Promega, V6941).

modifications: [C, +57.02146] and modifications per peptide: 3. For DIA library-based searches, the DDA RAW files from the corresponding runs were used to generate a spectral library, using DIA-NN[17] Unless otherwise stated, all samples were analyzed in experimental triplicate, and all mass spectrometry data is available via Proteome Xchange (PRIDE accession: PXD072165).

RESULTS AND DISCUSSION

We performed a comprehensive characterization of rChymoSelect against its commercially available analogue (defined here as standard chymotrypsin), as well as investigating the optimal digestion conditions in the context of buffer constitution, digestion methodology, and incubation time (Figure 1).

rChymoSelect Boosts Cleavage Specificity and Reduces Missed Cleavage Identifications

Standard chymotrypsin's promiscuity in the context of cleavage site specificity can impede downstream proteomics data processing pipelines. This occurs due to the increased search space that is associated with the inclusion of up to 9 amino acid cleavage sites as well as the increased number of missed cleavages that must be accounted for in the processing software. Consequently, the larger computational burden can drive slower data processing times, and more importantly, an increase in false positive identifications.[10] Two computational approaches can be used to curtail this cleavage site heterogeneity. First, one can include only the most probable cleavage sites (L, F, Y) and ignore all peptides that were generated from other "off-target" cleavages. However, this approach neglects these non-L, -F, and-Y peptides from the theoretical candidate peptide library for onward peptide spectral match (PSM) identification, consequently introducing missed annotations. Further, incorrect spectral assignments, in which an experimental spectrum is matched to an inappropriate theoretical spectrum, result in false positive identifications. In complex samples such as the complete human proteome, this effect is exacerbated and can substantially increase the rate of false positive peptide identifications. Alternatively, all 9 possible cleavage sites can be incorporated into the data processing workflow to ensure that all peptides can theoretically be identified. While this facilitates the identification of more PSMs, the pitfalls that are associated with this exponential increase in computation search space requirements are incorporated into the results.

To investigate whether cleavage specificity was improved with rChymoSelect, we initially performed a chymotrypsin proteomics experiment against commercially available intact protein lysate from human K562 cells (Promega; Figure 1). Comparison of both digestion enzymes at 2-h and 16-h incubation, with 2 M Urea, revealed a significant improvement in cleavage specificity for rChymoSelect vs standard chymotrypsin. Standard chymotrypsin analysis revealed that Leucine, Phenylalanine, and Tyrosine cleavage sites accounted for just 72 and 82%, for 2- and 16-h incubations, respectively. The remaining cleavage sites of lysine, arginine, tryptophan, histidine, asparagine, and glutamine accounted for 28 and 17% of cleavages, for 2- and 16-h incubation, respectively. In contrast, in the case of rChymoSelect, LFY site cleavage accounted for 97 and 95% for 2- and 16-h incubation, respectively (Figure 2a). The greatly improved cleavage specificity of rChymoSelect, whereby nearly all identified peptides were cleaved at L, F, or Y, enables the user to incorporate just three cleavage sites with minimal loss of peptides, reduced computational burden, and false positive identifications.

As well as characterizing the improved cleavage specificity of rChymoSelect, we probed the optimal urea concentration for sample preparation. Urea represents a useful buffer component for proteomics sample preparation due to its protein-denaturing capacity. As such, more protein residues are exposed to the enzyme for digestion. Our findings suggested minimal changes in peptide and protein identifications for rChymoSelect at 2-, 4-, and 6 M-urea (Figure S1a). Importantly, the application of standard chymotrypsin (and most proteases) is not amenable to high concentrations of chaotropic agents.[19] For example, urea concentration must be reduced to <1 M in most currently available digestion protocols. These data demonstrate the effective functionality of rChymoSelect well beyond this threshold, suggesting its

https://doi.org/10.1021/acs.jproteome.5c01262
J. Proteome Res. 2026, 25, 2511−2519

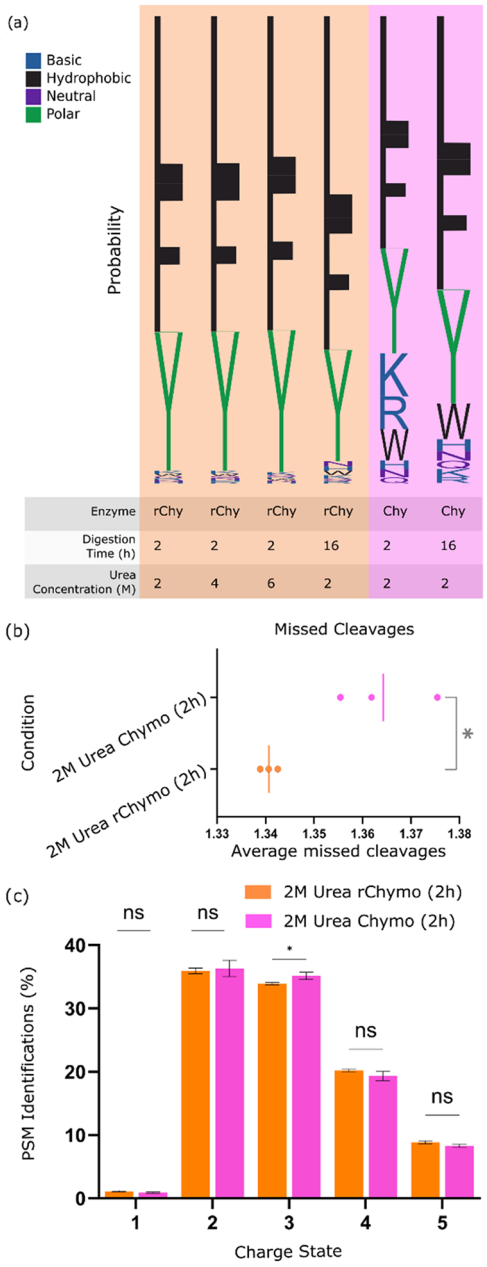


Figure 2. rChymoSelect enhanced cleavage specificity to reduce missed cleavage occurrence and increase peptide identifications vs standard chymotrypsin. (a) Cleavage site specificity analysis for rChymoSelect (rChymo) (digestion time: 2, 16 h; urea concentration: 2, 4, 6 M) vs standard chymotrypsin (Chymo) (digestion time: 2, 16 h; urea concentration: 2 M). Cleavage sequence analysis performed using Logomaker.[18] (b) Number of missed cleavages for three experimental replicates of Chymotrypsin vs rChymoSelect (2 M urea, 2 h digestion). Statistical analysis performed using a 2-sided unpaired *t* test, * represents $p < 0.05$. (c) Bar chart to represent percentage of PSMs for different charge states, rChymoSelect vs standard Chymotrypsin (2 M urea, 2 h incubation).

utility for digesting hydrophobic proteins that require aggressive denaturation conditions.

We next probed how the improved specificity of rChymoSelect could reduce the generation of missed cleavage peptides (Figure 2b). Not only did rChymoSelect reduce the average number of missed cleavages versus standard chymotrypsin, it also boosted the reproducibility across three experimental replicates. We reason that these improvements could result in more peptide spectral matches (PSMs) for a given peptide, with positive consequences for peptide confidence and quantitation statistics. Further, our analysis revealed that rChymoSelect generated a similar profile of PSM charge-state distributions, corresponding to 2+, 3+, 4+, and 5+ charges, relative to standard chymotrypsin. Importantly, more PSMs were identified across all charges (2+ − 5+) for rChymoSelect vs standard chymotrypsin (Figure S1b).

Ultimately, rChymoSelect demonstrated reduced heterogeneity of cleavage sites, reduced missed cleavages, and increased reproducibility and number of PSMs. These features combinatorially lead to greater peptide and, consequently, protein identifications.

rChymoSelect Does Not Alter the Chymotryptic Digested Peptide Physicochemical Properties, Relative to Standard Chymotrypsin

Having demonstrated the improved cleavage specificity and reproducibility of rChymoSelect vs standard chymotrypsin, we next investigated whether these changes would modify the properties of the resulting chymotryptic peptides. No significant changes in peptide length were identified from our comparison across 2 M urea and 2 h incubations (Figure 3a,b). Comparison of PSM m/z distribution revealed no significant increase, from rChymoSelect to standard chymotrypsin (Figure 3c). PSM hydrophobicity, as measured by the Grand average of hydrophobicity for a peptide or protein (GRAVY), revealed no significant changes in peptide hydrophobicity from rChymoSelect, compared to standard chymotrypsin when comparing all peptides identified from the two enzymes (Figure 3d). Interestingly, analysis of the unique peptide identifications of rChymoSelect and standard chymotrypsin revealed increased hydrophobicity of those peptides that were uniquely accessed upon rChymoSelect digestion, perhaps on account of its improved cleavage specificity at hydrophobic L, F, and Y residues (Figure S2). Together, these findings suggested that rChymoSelect enables improved digestion efficiency and reproducibility, leading to increased peptide and protein identifications without significantly modifying the physiochemistry of peptides that were generated during sample preparation.

rChymoSelect Significantly Boosts Protein Identifications, Relative to Standard Chymotrypsin

Our benchmarking had thus-far revealed that rChymoSelect demonstrated improved cleavage specificity and reduced missed cleavage occurrence without significant changes in peptide physiochemistry. These improvements led to a boost in PSM identifications across the standard proteomics precursor charge state distribution, relative to standard chymotrypsin (Figure S1b). We next sought to investigate how these technical advancements in chymotrypsin activity could deliver more informative proteomics data. For this, we investigated the overall protein sequence coverage distribution for rChymoSelect vs standard chymotrypsin, across the proteome of human K562 cells (Figure 4a). rChymoSelect demonstrated greater protein sequence coverage across the proteome, from the most abundant to the least abundant identified protein of the data set.

We next probed the differential proteomics data for rChymoSelect versus standard chymotrypsin. Of the 3638 proteins that were identified in our benchmarking experiment, 22.2% were exclusively identified from rChymoSelect,

https://doi.org/10.1021/acs.jproteome.5c01262
J. Proteome Res. 2026, 25, 2511−2519

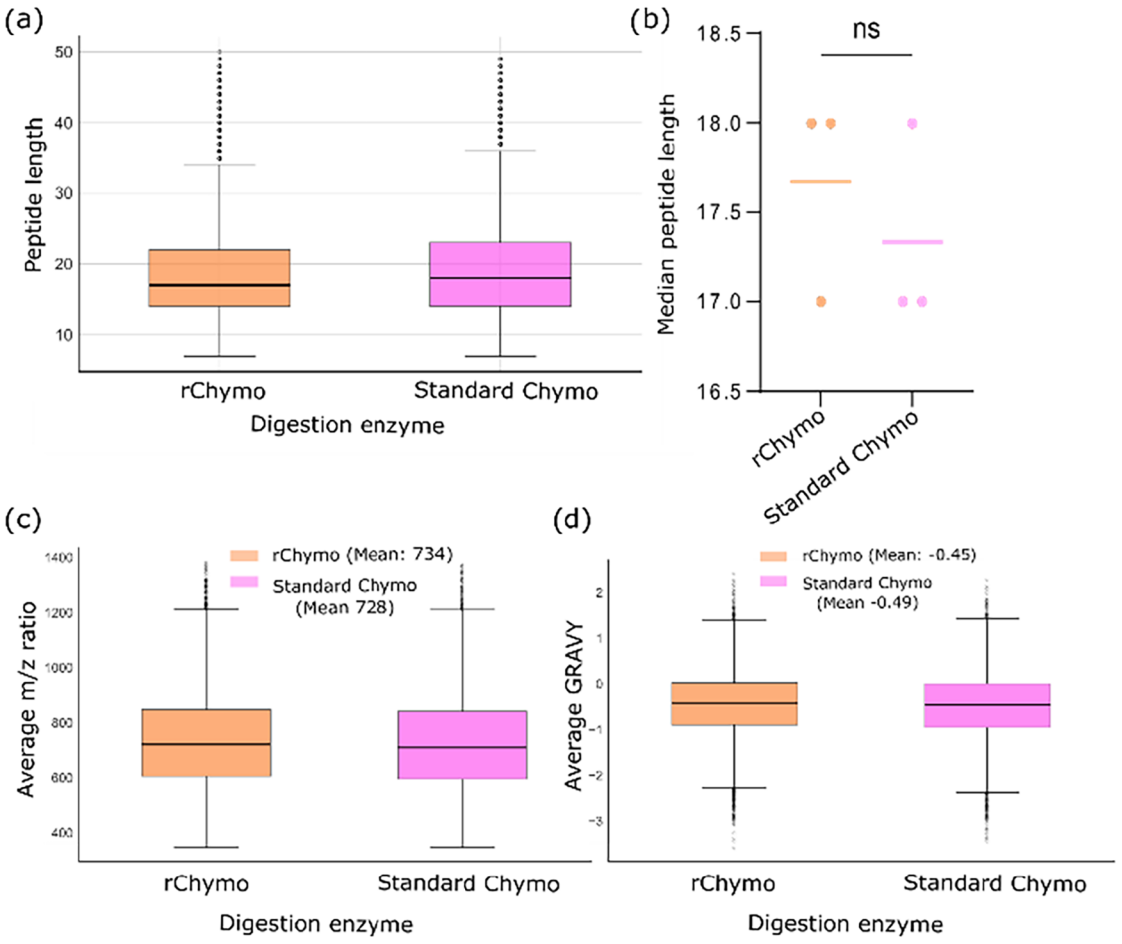


Figure 3. rChymoSelect does not alter peptide properties, vs standard chymotrypsin. Comparison of peptide length from both proteases, as plotted by (a) boxplot and (b) scatter plot. Statistical analysis performed using a 2-sided unpaired *t* test, ns represents $p > 0.05$. Boxplot comparison of (c) PSM *m*/*z* ratios and (d) Grand average of hydrophobicity for a peptide or protein (GRAVY) score distribution as a measure of PSM hydrophobicity for each condition.

compared to 8.2% for standard chymotrypsin. These findings demonstrated the improved performance of rChymoSelect in generating a comprehensive proteomics data set, vs standard chymotrypsin (Figure 4b). Investigation of the cellular compartment of proteins that were identified from each proteolytic enzyme revealed broadly similar profiles. The improved protein identifications that were associated with rChymoSelect were seen across proteins associated with cytoplasmic, nuclear, mitochondrial, and vesicular Gene Ontology Cellular Compartments (GO:CC), both in terms of FDR and total proteins (Figure 3c).

Following our benchmarking experiments for rChymoSelect versus standard chymotrypsin, we next investigated the optimal conditions across our proteomics sample preparation workflow to enhance rChymoSelect performance. Minimal differences in protein identifications were identified between in-solution vs SP3 digestion (2.15% more identifications for in-solution, Figure 4d). However, automated SP3 digestion, using the KingFisher robot, provides multiple advantages over in-solution-based digestion strategies. First, throughput is vastly boosted to enable the efficient processing of up to 96 samples in one step. Second, human-introduced variability is curtailed to deliver more robust data. Third, as protein is extracted from its lysis buffer before digestion, any detergent can be leveraged to boost protein denaturation predigestion, without compromising protease efficiency. As such, the comparable performance of in-solution versus SP3 digestion for rChymoSelect suggests that performance can be boosted by leveraging these denaturation strategies to optimize performance.

Varying the urea concentration demonstrated no significant differences in protein identifications or sequence coverage distribution (Figure 4e). However, manipulating guanidinium hydrochloride denaturant concentration and protease/protein ratio revealed more significant changes. An inverse relation between the guanidinium hydrochloride concentration and protein identifications was observed, such that 0.1 M boosted protein identifications by 5.6%, relative to 0.4 M (Figure 4f). Importantly, the functionality of rChymoSelect at 6 M urea and 0.4 M guanidinium hydrochloride demonstrates its applicability for proteins that require aggressive denaturation conditions. As expected, increasing the protease/protein ratio also boosted identifications, with 1:20 driving increased protein identifications by 7.2% versus 1:100 (Figure 4g). We next probed whether increasing the rChymoSelect digestion time would drive a significant increase in protein identifications. For this, we compared a 2-h incubation against a 16-h incubation (with in-solution digestion, 2 M urea, and 1:40 protease:protein ratio). No significant change in protein identifications was observed when comparing the two conditions, with just a 2.4% increase from 2 to 16 h (Figure

https://doi.org/10.1021/acs.jproteome.5c01262
J. Proteome Res. 2026, 25, 2511−2519

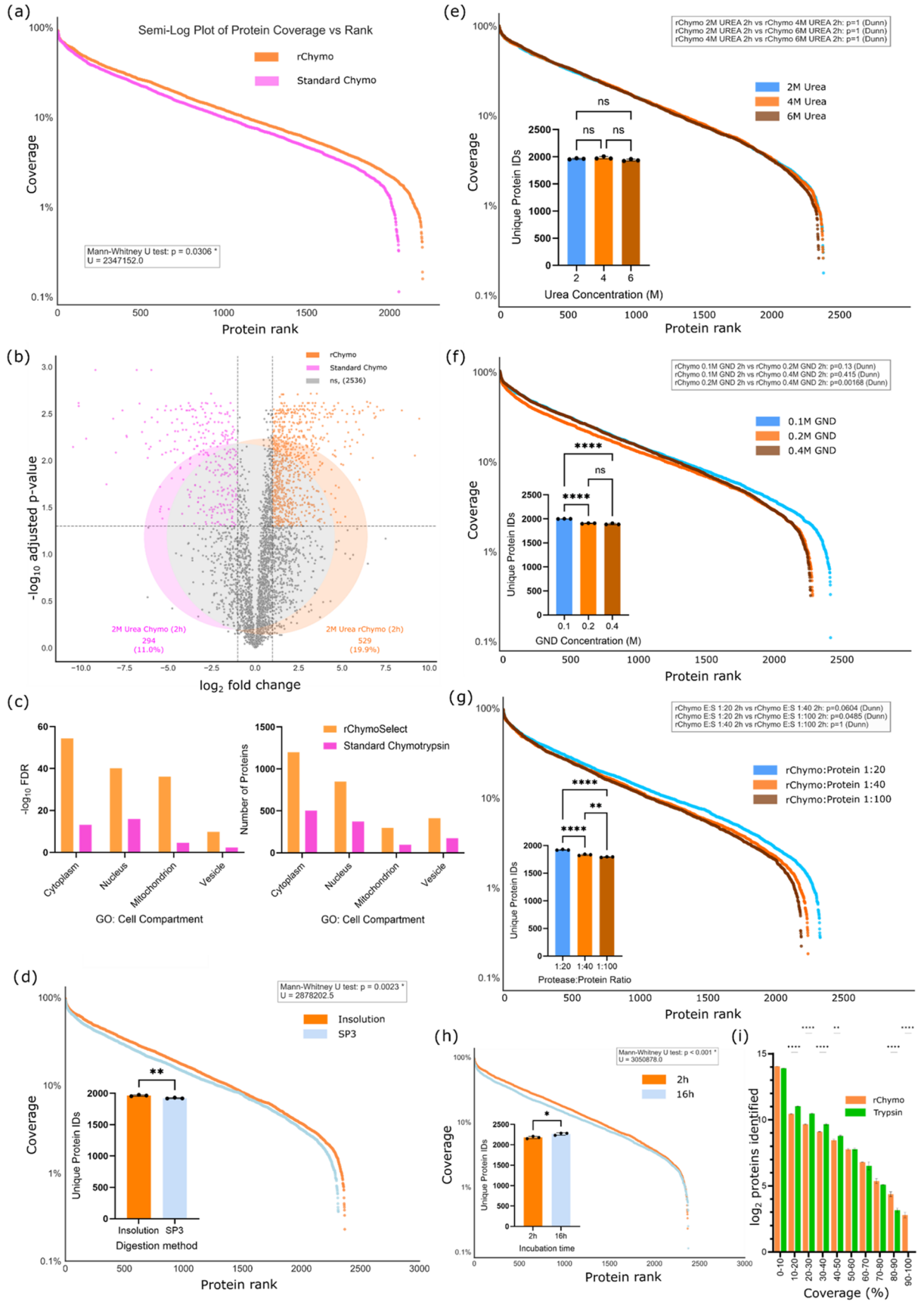


Figure 4. rChymoSelect significantly boosts protein identifications, relative to standard chymotrypsin, based on (a) protein sequence coverage distribution and (b) volcano plot and protein identifications Venn diagram. (c) Comparison of rChymoSelect vs standard chymotrypsin for identification of cytoplasmic, nuclear, mitochondrial, and vesicle proteins, using Gene Ontology: Cell Compartment annotations include $-\log_{10}$ False Discovery Rate (left) and number of proteins for each subcellular location (right). All experiments were performed with 2 M urea and 2-h

https://doi.org/10.1021/acs.jproteome.5c01262
J. Proteome Res. 2026, 25, 2511−2519

Figure 4. continued

incubation using in-solution digestion. Comparison of different sample preparation methodologies, including sequence coverage distribution and protein identifications bar chart for (d) in-solution vs SP3 digestion (2 M urea, 2 h digestion, 1:40 rChymoSelect/protein ratio), (e) urea concentration (in-solution digestion, 2-h incubation, 1:40 rChymoSelect/protein ratio), (f) guanidinium hydrochloride (GND) concentration (SP3 digestion, 2-h incubation, 1:20 rChymoSelect/protein ratio), (g) rChymoSelect to protein ratio (SP3 digestion, no denaturant, 2-h incubation), and (h) 2- vs 16-h incubation (in-solution digestion, 2 M urea, 1:40 rChymoSelect/protein ratio). (i) Protein identifications at different sequence coverages, across 10% sequence coverage bins, with comparison between rChymoSelect and trypsin using lysed U2-OS cells, with offline peptide fractionation. Protein sequence coverage distribution plot statistical analysis performed using 2-sided Mann−Whitney *U* rank test on two independent samples, where *U* represents the Mann−Whitney statistic. Volcano plot statistical analysis performed using 2-sided independent *t* test, with Benjamin-Hochberg adjusted *p*-values. Bar chart statistical analysis performed using a two-sided Kruskal−Wallis test with Dunn's post hoc test. rChymo vs trypsin bar chart statistical analysis performed using 2-way ANOVA, **represents *p*-value <0.01 and *** represents *p*-value <0.0001.

4h). These data suggest rChymoSelect enables faster sample preparation by avoiding the standardized overnight digestion, with minimal concomitant loss of identifications.

Following this, we investigated how this improved iteration of chymotrypsin would perform relative to trypsin, the gold standard of proteomics research. For this, we applied a deep proteomics pipeline, including offline peptide fractionation of lysed U2-OS cells before LC-MS/MS analysis. As expected, trypsin digestion outperformed rChymoSelect digestion for the identification of proteins with a sequence coverage lower than 50%. Intriguingly, for more "digestible" proteins, whose sequence coverage was greater than 60%, rChymoSelect demonstrated improved identifications, particularly in the case of >90% sequence coverage (Figure S3). Thus, while trypsin remains the enzyme of choice for global deep proteomics, rChymoSelect provides complementary coverage for proteins where near-complete sequence coverage is required

rChymoSelect Facilitates Improvements in Data-Independent Acquisition Proteomics Strategies

Recently, Data-Independent Acquisition (DIA) has become the premier strategy for high-throughput protein identification.[20] Briefly, while Data-Dependent Acquisition modes sequentially fragment the most abundant precursor ions, DIA performs fragmentation over a predefined window of m/z's, scanning m/z bins across the precursor mass range, for each eluted chromatographic peak. This enables greater protein identifications, as well as curtailing issues of missing values across replicates. From our comparison of DIA versus DDA for rChymoSelect-based proteomics, DIA facilitated a 45% increase in protein identifications, versus DDA, against standard K562 cell lysate (Figure 5a). This improvement is comparable to that of trypsin-based proteomics, where at least a 50% improvement in protein identifications is facilitated by DIA workflows, vs DDA.[21] Because of the heterogeneity of sequence cleavage specificity of standard chymotrypsin, DIA data processing methodologies struggle to manage the unpredictable spectra that are generated from off-target cleavages of standard chymotrypsin. Therefore, we reasoned that the superior cleavage specificity of rChymoSelect would alleviate such issues, improving DIA workflows with chymotrypsin-based proteomics. While a 3.6% increase in protein identifications was observed from a library-free DIA search with rChymoSelect vs standard chymotrypsin (Figure 5b), the use of a DDA-generated spectral library within the DIA pipeline generated a 16.6% increase in peptide-spectrum matches, leading to a 4.6% increase in protein identifications (with 2 M urea denaturation). Further, 4 M urea denaturation increased PSMs and protein identifications by 22.4 and 6.2% vs standard chymotrypsin, respectively (Figure 5c,d).

While these DIA experiments were performed against one experimental replicate per condition, the data are encouraging in the context of leveraging the advantages of rChymoSelect against the increased protein identifications and robustness across replicates that are associated with DIA strategies.

CONCLUSIONS

Chymotrypsin offers a complementary analytical capability for proteomic applications through its preferential cleavage at hydrophobic amino acid residues, a specificity that is orthogonal to that of trypsin. Nevertheless, its broader substrate specificity and tendency to produce missed cleavages can limit digestion efficiency and expand the computational search space required for peptide identification during data processing. Benchmarking of a novel recombinant chymotrypsin analogue, rChymoSelect (Promega), against conventional bovine chymotrypsin demonstrated marked improvements in cleavage specificity and proteome sequence coverage. Whereas standard chymotrypsin cleaved at the C-termini of nine distinct amino acid residues, the recombinant chymotrypsin variant exhibited vastly enhanced specificity, with up to 97% of peptide bond cleavages occurring after leucine, phenylalanine, or tyrosine residues. This increased cleavage selectivity led to fewer missed cleavages, reduced database search space, and resulted in a significant rise in peptide spectral matches, culminating in greater protein identifications and improved data reproducibility relative to the standard enzyme. Workflow optimization confirmed a comparable performance between in-solution and SP3 digestion, supporting its flexibility in diverse sample preparation formats. Further, denaturation with either 2 M urea or 0.1 M guanidinium hydrochloride provided sufficient protein unfolding for efficient proteolysis, although rChymoSelect's effective functionality at up to 6 M urea and −0.4 M guanidinium hydrochloride positions it as a highly attractive protease for the digestion of hydrophobic proteins that require harsh denaturation conditions. Increasing the incubation time from 2 h to overnight (16 h) did not yield a substantial improvement in proteome coverage, suggesting that rChymoSelect provides fast digestion efficiency and enables the entire sample preparation workflow to be completed in a matter of hours. Incorporation of Data-Independent Acquisition (DIA) strategies increased protein identifications by 45% relative to DDA, with reduced cleavage site heterogeneity, simplifying spectra and enhancing peptide assignments for rChymoSelect. Together, these findings establish rChymoSelect as a highly valuable complementary protease for proteomics analysis.

https://doi.org/10.1021/acs.jproteome.5c01262
J. Proteome Res. 2026, 25, 2511−2519

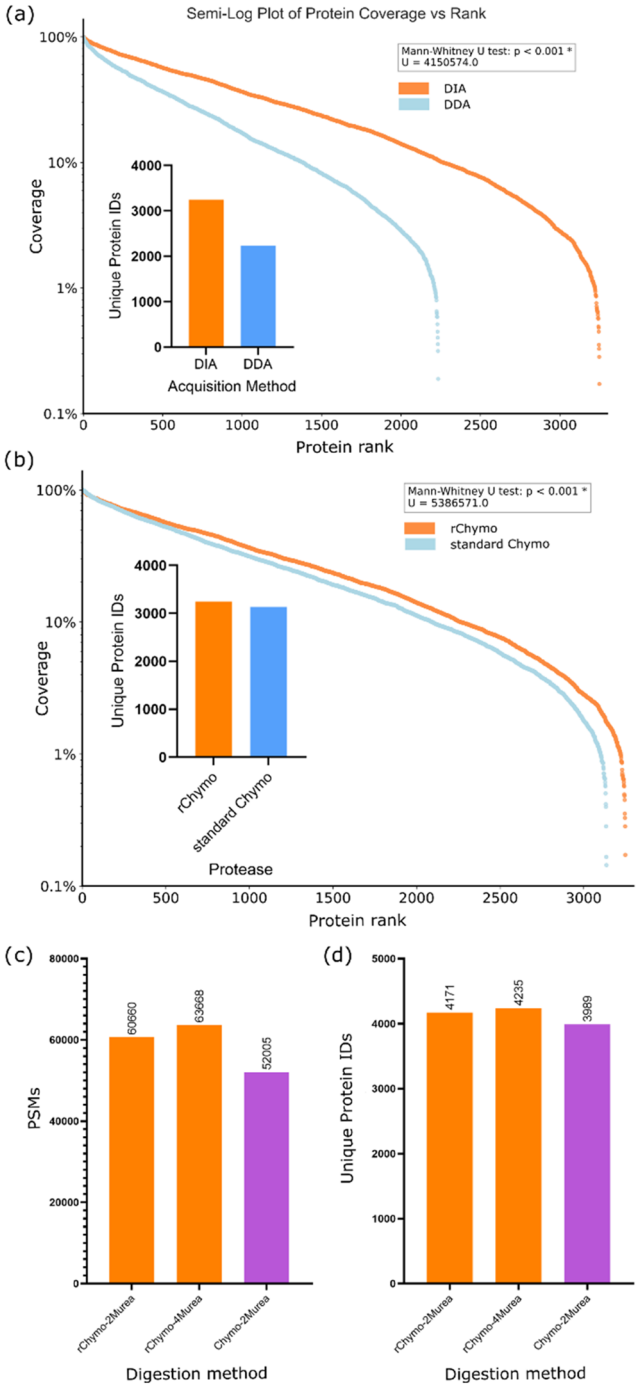


Figure 5. Incorporation of Data-Independent Acquisition (DIA) strategies into chymotrypsin proteomics. Protein sequence coverage distribution plot and unique protein identifications bar chart for (a) DIA vs DDA with rChymoSelect, and (b) rChymoSelect-DIA vs standard chymotrypsin-DIA, using standard K562 cell lysate. Samples were digested in 2 M urea, using in-solution digestion, with 2 h incubation at 25 °C (library-free search). Spectral library-based DIA searches are represented in terms of (c) PSM identifications and (d) unique protein identifications.

■ ASSOCIATED CONTENT

Data Availability Statement

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[22] partner repository with the data set identifier PXD072165.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.5c01262.

(a) Peptide and protein identifications for 2, 4, and 6 M urea concentration (2 h incubation) for rChymoSelect. (b) Number of PSMs for rChymoSelect vs standard chymotrypsin for 2+ − 5+ charge-state (Figure S1); comparison of GRAVY scores of uniquely identified peptides with rChymo (orange) and standard chymotrypsin (orange), visualized as a histogram (a) and box plot (b) (Figure S2); sequence coverage for rChymoSelect (orange) and Trypsin (green) for selected proteins with sequence coverage 90−100% with rChymoSelect (Figure S3) (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Konstantinos Thalassinos − *Institute of Structural and Molecular Biology, Division of Biosciences and University College London Mass Spectrometry Science Technology Platform, Division of Biosciences, University College London, London WC1E 6BT, U.K.; Institute of Structural and Molecular Biology, School of Natural Sciences, Birkbeck College, University of London, London WC1E 7HX, U.K.;* orcid.org/0000-0001-5072-8428; Email: k.thalassinos@ucl.ac.uk

Riccardo Zenezini Chiozzi − *Institute of Structural and Molecular Biology, Division of Biosciences and University College London Mass Spectrometry Science Technology Platform, Division of Biosciences, University College London, London WC1E 6BT, U.K.*; Email: r.chiozzi@ucl.ac.uk

Authors

Kish R. Adoni − *Institute of Structural and Molecular Biology, Division of Biosciences, University College London, London WC1E 6BT, U.K.;* orcid.org/0000-0001-7390-501X

Jonathan E. Ditcham − *Institute of Structural and Molecular Biology, Division of Biosciences, University College London, London WC1E 6BT, U.K.*

Alba Katiria González Rivera − *Promega Corporation, Madison, Wisconsin 53711, United States*

Georgina H. Charlton − *Institute of Structural and Molecular Biology, Division of Biosciences, University College London, London WC1E 6BT, U.K.;* orcid.org/0000-0002-3457-2694

Sergei V. Saveliev − *Promega Corporation, Madison, Wisconsin 53711, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.5c01262

Author Contributions

The manuscript was written by K.R.A. with contributions from all authors. All authors have given approval to the final version of the manuscript. Conceptualization of methodology: R.Z.C.: Software: J.D.: Validation: K.R.A. and J.D. Formal analysis: K.R.A. and J.D. Investigation: K.R.A., J.D., G.H.R. and Z.C. Writing—original draft: K.R.A. Writing—editing and reviewing: K.R.A. and K.T. Visualization: K.R.A. and J.D. Supervision: K.T. and R.Z.C. Project administration: K.T. and R.Z.C. Funding acquisition: K.T.

https://doi.org/10.1021/acs.jproteome.5c01262
J. Proteome Res. 2026, 25, 2511−2519

Funding

Wellcome Trust Multiuser Equipment grant 221521/Z/20/Z (K.T). Wellcome Trust Collaborative Award in Science 209250/Z/17/Z (K.T).

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

rChymo, recombinant chymotrypsin; Chymo, chymotrypsin; SP3, Single-Pot, Solid-Phase-enhanced Sample Preparation; SDS, sodium dodecyl sulfate; SDC, sodium deoxycholate; GND, guanidinium hydrochloride; LC, liquid chromatography; MS, mass spectrometry; PSM, peptide spectral match; GO:CC, gene ontology:cell compartment; DDA, Data-Dependent Acquisition; DIA, Data-Independent Acquisition

■ REFERENCES

(1) Peters-Clarke, T. M.; Coon, J. J.; Riley, N. M. Instrumentation at the Leading Edge of Proteomics. *Anal. Chem.* **2024**, *96* (20), 7976−8010.
(2) Guo, T.; Steen, J. A.; Mann, M. Mass-Spectrometry-Based Proteomics: From Single Cells to Clinical Applications. *Nature* **2025**, *638* (8052), 901−911.
(3) Tsiatsiani, L.; Heck, A. J. R. Proteomics beyond Trypsin. *FEBS J.* **2015**, *282* (14), 2612−2626.
(4) Nagy, K.; Sándor, P.; Vékey, K.; Drahos, L.; Révész, Á. The Enzyme Effect: Broadening the Horizon of MS Optimization to Nontryptic Digestion in Proteomics. *J. Am. Soc. Mass Spectrom.* **2025**, *36* (2), 299−308.
(5) Di Cera, E. Serine Proteases. *IUBMB Life* **2009**, *61* (5), 510−515.
(6) Min, L.; Choe, L. H.; Lee, K. H. Improved Protease Digestion Conditions for Membrane Protein Detection. *Electrophoresis* **2015**, *36* (15), 1690−1698.
(7) Vermachova, M.; Purkrtova, Z.; Santrucek, J.; Jolivet, P.; Chardot, T.; Kodicek, M. Combining Chymotrypsin/Trypsin Digestion to Identify Hydrophobic Proteins from Oil Bodies. In *Plant Proteomics: Methods and Protocols*; Jorrin-Novo, J. V.; Komatsu, S.; Weckwerth, W.; Wienkoop, S., Eds.; Humana Press: Totowa, NJ, 2014; pp 185−198 DOI: 10.1007/978-1-62703-631-3_14.
(8) Dau, T.; Bartolomucci, G.; Rappsilber, J. Proteomics Using Protease Alternatives to Trypsin Benefits from Sequential Digestion with Trypsin. *Anal. Chem.* **2020**, *92* (14), 9523−9527.
(9) Ovsyanko, S. L.; Chernyavskii, E. A.; Yurkshtovich, T. L.; Shkumatov, V. M.; Kaputskii, F. N. A Comparative Analysis of Various Commercial α-Chymotrypsin Preparations by HPLC Under Nondenaturing Conditions. *Pharm. Chem. J.* **2001**, *35* (10), 573−576.
(10) Savitski, M. M.; Wilhelm, M.; Hahne, H.; Kuster, B.; Bantscheff, M. A Scalable Approach for Protein False Discovery Rate Estimation in Large Proteomic Data Sets. *Mol. Cell. Proteomics* **2015**, *14* (9), 2394−2404.
(11) Walker, J. M. The Bicinchoninic Acid (BCA) Assay for Protein Quantitation. *Methods Mol. Biol.* **1994**, *32*, 5−8.
(12) Koenig, C.; Martinez-Val, A.; Naicker, P.; Stoychev, S.; Jordaan, J.; Olsen, J. V. Protocol for High-Throughput Semi-Automated Label-Free- or TMT-Based Phosphoproteome Profiling. *STAR Protoc.* **2023**, *4* (3), No. 102536.
(13) Lazear, M. R. Sage: An Open-Source Tool for Fast Proteomics Searching and Quantification at Scale. *J. Proteome Res.* **2023**, *22* (11), 3652−3659.
(14) The, M.; Samaras, P.; Kuster, B.; Wilhelm, M. Reanalysis of ProteomicsDB Using an Accurate, Sensitive, and Scalable False Discovery Rate Estimation Approach for Protein Groups. *Mol. Cell. Proteomics* **2022**, *21* (12), No. 100437.
(15) Swift, M. L. GraphPad Prism, Data Analysis, and Scientific Graphing. *J. Chem. Inf. Comput. Sci.* **1997**, *37* (2), 411−412.
(16) Kong, A. T.; Leprevost, F. V.; Avtonomov, D. M.; Mellacheruvu, D.; Nesvizhskii, A. I. MSFragger: Ultrafast and Comprehensive Peptide Identification in Mass Spectrometry−Based Proteomics. *Nat. Methods* **2017**, *14* (5), 513−520.
(17) Demichev, V.; Messner, C. B.; Vernardis, S. I.; Lilley, K. S.; Ralser, M. DIA-NN: Neural Networks and Interference Correction Enable Deep Proteome Coverage in High Throughput. *Nat. Methods* **2020**, *17* (1), 41−44.
(18) Tareen, A.; Kinney, J. B. Logomaker: Beautiful Sequence Logos in Python *bioRxiv* 2019635029.
(19) Martin, C. J.; Frazier, A. R. The Urea Denaturation of α-Chymotrypsin. THE EFFECT OF PH ON REVERSIBLE AND IRREVERSIBLE INACTIVATION. *J. Biol. Chem.* **1963**, *238*, 3268−3273.
(20) Fröhlich, K.; Fahrner, M.; Brombacher, E.; Seredynska, A.; Maldacker, M.; Kreutz, C.; Schmidt, A.; Schilling, O. Data-Independent Acquisition: A Milestone and Prospect in Clinical Mass Spectrometry-Based Proteomics. *Mol. Cell. Proteomics* **2024**, *23* (8), No. 100800.
(21) Zou, X.; Wang, L.; Chen, Y.; Fu, H.; Gao, Y.; Liu, B.; Tan, M.; Zhai, L. In-Depth Analysis of Data Characteristics and Comparative Evaluation of Dda and Dia Accuracy in Label-Free Quantitative Proteomics of Biological Samples. *Clin. Proteomics* **2026**, *23*, No. 2.
(22) Martens, L.; Hermjakob, H.; Jones, P.; Adamski, M.; Taylor, C.; States, D.; Gevaert, K.; Vandekerckhove, J.; Apweiler, R. PRIDE: The Proteomics Identifications Database. *Proteomics* **2005**, *5* (13), 3537−3545.

https://doi.org/10.1021/acs.jproteome.5c01262
J. Proteome Res. 2026, 25, 2511−2519