

From volcanoes to the bench: Advantages of novel hyperthermoacidic archaeal proteases for proteomics workflows

Maxwell C. McCabe^a, Varun Gejji^{b,1}, Adam Barnebey^{b,1}, Gary Siuzdak^c, Linh Truc Hoang^c, Truc Pham^a, Keira Y. Larson^a, Anthony J. Saviola^a, Steven M. Yannone^{b,*,1}, Kirk C. Hansen^{a,*}

^a Department of Biochemistry and Molecular Genetics, University of Colorado Denver, Aurora, CO 80045, USA

^b Cinder Biological, Inc., 1933 Davis Street, STE 208, San Leandro, CA 94577, USA

^c Departments of Chemistry, Molecular, and Computational Biology, Scripps Research, La Jolla, CA 92037, USA

ARTICLE INFO

Keywords:

Archaea
Hyperthermophile
Ultrastable enzymes
Heat
Acid
Protease
Digestion
Sample preparation
Thermopsin
HTA-Protease

ABSTRACT

Here we introduce hyperthermoacidic archaeal proteases (HTA-Proteases©) isolated from organisms that thrive in nearly boiling acidic volcanic springs and investigate their use for bottom-up proteomic experiments. We find that HTA-Proteases have novel cleavage specificities, show no autolysis, function in dilute formic acid, and store at ambient temperature for years. HTA-Proteases function optimally at 70–90 °C and pH of 2–4 with rapid digestion kinetics. The extreme HTA-Protease reaction conditions actively denature sample proteins, obviate the use of chaotropes, are largely independent of reduction and alkylation, and allow for a one-step/five-minute sample preparation protocol without sample manipulation, dilution, or additional cleanup. We find that brief one-step HTA-Protease protocols significantly increase proteome and protein sequence coverage with datasets orthogonal to trypsin. Importantly, HTA-Protease digests markedly increase coverage and identifications for ribonucleoproteins, histones, and mitochondrial membrane proteins as compared to tryptic digests alone. In addition to increased coverage in these classes, HTA-Proteases and simplified one-step protocols are expected to reduce technical variability and advance the fields of clinical and high-throughput proteomics. This work reveals significant utility of heretofore unavailable HTA-Proteases for proteomic workflows. We discuss some of the potential for these remarkable enzymes to empower new proteomics methods, approaches, and biological insights.

Significance: Here we introduce new capabilities for bottom-up proteomics applications with hyperthermoacidic archaeal proteases (HTA-Proteases©). HTA-Proteases have novel cleavage specificity, require no chaotropes, and allow simple one-step/five-minute sample preparations that promise to reduce variability between samples and laboratories. HTA-Proteases generate unique sets of observable peptides that are non-overlapping with tryptic peptides and significantly increase sequence coverage and available peptide targets relative to trypsin alone. HTA-Proteases show some bias for the detection and coverage of nucleic acid-binding proteins and membrane proteins relative to trypsin. These new ultra-stable enzymes function optimally in nearly boiling acidic conditions, show no autolysis, and do not require aliquoting as they are stable for years at ambient temperatures. Used independently or in conjunction with tryptic digests, HTA-Proteases offer increased proteome coverage, unique peptide targets, and brief one-step protocols amenable to automation, rapid turnaround, and high-throughput approaches.

1. Introduction

The application of proteomics has transformed the landscape of modern biological research over the past decades. Since the development and popularization of mass spectrometry-based proteomics in the

late 20th century, advances in many aspects of sample preparation, sample size reduction towards nano-proteomics [1], and front-end chromatography technologies [2,3] have all advanced at an extraordinary pace. These advances led to the publication of the first draft of the human proteome in 2014 [4,5] and an updated high-stringency version

* Corresponding authors.

E-mail addresses: SteveYannone@CinderBio.com (S.M. Yannone), kirk.hansen@cuanenschutz.edu (K.C. Hansen).

¹ Authors contributed as part of employment at Cinder Biological, Inc., a for-profit business

<https://doi.org/10.1016/j.jprot.2023.104992>

Received 2 December 2022; Received in revised form 26 July 2023; Accepted 14 August 2023

Available online 25 August 2023

1874-3919/© 2023 CinderBiological, Inc. (dba CinderBio). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

in 2020 [6], signaling that proteomics is rapidly transitioning from research and development to applied fields including diagnostics.

Peptide-centric proteomic workflows, i.e., bottom-up proteomics, rely on the enzymatic digestion of proteins into peptides prior to analysis by tandem mass spectrometry [7]. Frequently, digestion is performed with trypsin due to its remarkable cleavage specificity, reliability, and widespread availability. However, several limitations can hamper the utility of trypsin for the comprehensive analysis of a proteome. For example, proteins lacking or with heavily modified lysine and arginine residues can be excluded from proteomic datasets [8,9] and its autolysis at basic pH and relatively short shelf-life can introduce experimental variability [10,11]. In addition, trypsin digestion is often the most time-consuming step of a protocol and is conventionally performed overnight. Therefore, identifying new proteases with complementary activities to trypsin that have simple, rapid, and reliable protocols can result in improvements to proteomic analyses.

The discovery of hyperthermoacidic life forms in the 1970's [12] and the ensuing work to understand these remarkable microbes has revealed enzymes with thermal and pH optima that would have previously been considered impossible (Fig. 1). Advantages of thermal denaturation of target proteins for proteolytic cleavage are reported as early as the 1980's and are shown to increase target protein digestibility and reduce digestion times [13]. Additionally, heat pre-treatments prior to tryptic digestions are reported to increase coverage from 15 to 86% for myoglobin and from 0 to 43% for ovalbumin [14]. These and other studies demonstrate advantages of thermal-denaturation treatments including ease of use, no sample dilution, no need for post-digestion removal of chaotropes, enhanced proteolysis, and reduced digestion times. Other works reveal additional practical advantages of thermostable proteases for challenging proteomic targets like the tightly folded globular capsid proteins of viral particles [15]. Furthermore, the effects of severe pH conditions on enzymatic digestions have also shown the potential for practical advantages for proteomics [16].

Recently developed technologies enable the production of many classes of hyperthermoacidic archaeal enzymes (HTA-enzymes) that function optimally at temperatures near the boiling point of water and at highly acidic pH [17]. The inherent hyperstable nature of these enzymes also manifests as shelf-lives on the order of years in solution at ambient temperatures with no detectable autolysis. Here we evaluate the utility of two such HTA-Proteases for proteomics applications. The experimental scope presented here is not intended to be comprehensive, but rather to introduce the robust nature of HTA-Proteases and to demonstrate the potential for advancing proteomic capabilities. This new class of protease can simplify sample preparation, reduce digestion times, and offers new cleavage specificity among several other characteristics that

can further empower the fields of proteomics.

2. Materials and methods

2.1. Enzyme and substrate preparation

2.1.1. Production and characterization of HTA-Proteases

CB14057 and CB23726 are both computationally classified as thermopsins (EC 3.4.23.42) based on domain sequence homology and source organisms [18]. Both enzymes were recombinantly expressed in a proprietary Archaeal host platform [19] with the source organisms for both genes coming from members of the lineage; Archaea; Crenarchaeota; Thermoprotei; Sulfolobales. Both proteases were purified to near homogeneity using standard biochemical techniques and chromatography [20]. Purified enzymes were characterized extensively for pH and thermal optima, as well as shelf-life using standard protease assays [21] on a series of substrates including BSA, IgG, casein, and those presented in this work (Fig. 1). The purity of protease activity in all preparations was validated by zymography showing at least 99% of detectable protease activity is ascribed to bands of the correct mobility for the enzyme of interest. The cleavage specificity for these enzymes were determined through a series of proteomics experiments (including this work). The ratios or dosages of enzyme to substrate throughout these studies are expressed as mass enzyme to mass substrate or defined activity units where determined and appropriate.

2.1.2. BSA Preparation and Pre-Treatment (Chaotrope versus heat/acid denaturation)

Two 10 µg/µL stock solutions of bovine serum albumin (BSA) were prepared in either 50 mM ammonium bicarbonate (ABC) pH 7.8 or 20 mM phosphate citric acid buffer (pH 3). One aliquot of each stock was reduced by addition of dithiothreitol (DTT) to a final concentration of 10 mM and incubated at 37 °C for 60 min. The reduced samples were then alkylated by adding iodoacetamide (IAA) at a final concentration of 32 mM and incubated at room temperature for 60 min in the dark. Urea was added to the DTT/IAA treated aliquots to 6 M final concentration and both aliquots were incubated for 1 h at 37 °C.

2.1.3. Soluble and insoluble *E. coli* extract preparation

Top10 *E. coli* was purchased from Invitrogen (cat # C404010) and grown in glass culture flasks at 37 °C with shaking at 300 rpm in 2× YT media with a total volume of 1 L. As the culture reached stationary phase (OD ~1.1), cells were harvested by centrifuge at 3000 rpm (1711 xg) at 4 °C. Pelleted cells were washed twice in 1× TBS buffer. After the final wash in TBS, cells were split equally into two tubes and stored at -20 °C.

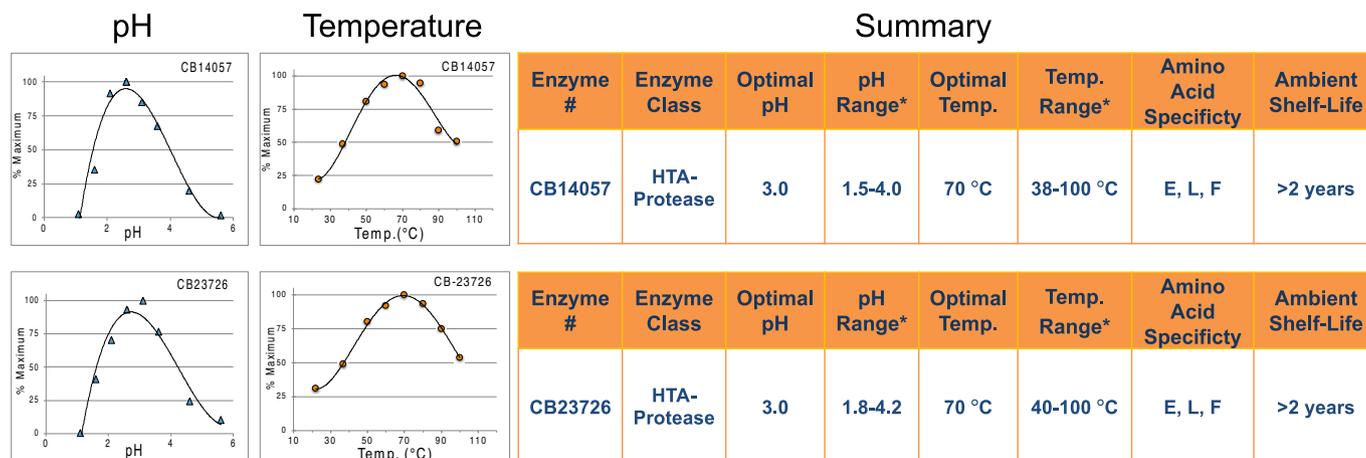


Fig. 1. Basic physical and biochemical characteristics of HTA-Proteases investigated in this study. Left: relative enzyme activity as a function of digest pH and temperature. Right: summary of relevant enzyme characteristics.

One half of the washed, pelleted, and frozen cells (500 mL of culture) was thawed and the cells resuspended in 15 mL of 20 mM NH_4HCO_3 pH 7 and sonicated 4 \times on ice for 3 min/cycle, constant duty, ~30% output with 5-min ice cooling between cycles using a Branson 250 Sonifier with a large probe. The sonicate was cleared by centrifugations at 14,000 rpm (15,339 \times g) for 10 min. The cleared lysate (soluble extract) was determined to contain 50 mg/mL protein using a Lowry assay with BSA as a standard and decanted and stored in 1 mL aliquots at -80°C until use. The insoluble cell portion (pellets) were resuspended in 5 mL of 20 mM NH_4HCO_3 pH 7 and sonicated as above an additional three times. This back-extraction was repeated with an additional 5 mL of buffer and both cleared by centrifugation and supernatants were combined. Lowry protein determinations indicated a final protein concentration of 27 mg/mL for these fractions (insoluble extract) which were aliquoted into 1 mL fractions and stored at -80°C until use.

2.1.4. Whole mouse powder (WMP) sample preparation

A whole male C57BL/6 J mouse was frozen in liquid nitrogen and fractured into approximately 1 cm^3 pieces. Fur and blood were not removed from the mouse before milling. The resulting pieces were kept frozen in liquid nitrogen before milling to a fine powder under liquid nitrogen using a SPEX 6870 freezer/mill. The milled powder was kept frozen and lyophilized for 24 h. Approximately 100 mg aliquots of lyophilized powder were then delipidated by 4 successive extractions with 2 mL 100% ice-cold acetone and briefly dried at room temperature in a fume hood and stored cryogenically until used.

2.1.5. Bovine neck ligament (BNL), bovine tendon (BT), and human blood clot (HBC) sample preparation

HBC samples were decellularized using successive washes of 6 M guanidine hydrochloride (Gnd-HCl) until clots were pale and translucent before being washed 2 \times with ddH $_2$ O to remove residual Gnd-HCl. Decellularized-HBC, BT and BNL tissues were then cryo-milled to a fine powder using liquid nitrogen and a mortar and pestle. Milled tissue samples were then reduced and alkylated by incubating in 10 mM DTT at 37°C for 30 mins followed by 50 mM IAA at room temp for 15 min. Samples were then frozen and lyophilized overnight and stored cryogenically until use.

2.2. Sample digestion

2.2.1. Trypsin digests of BSA (Chaotrope versus heat/acid denaturation)

Trypsin (NEB #P8101S) was prepared in 50 mM ABC and 25 μL of

serially diluted trypsin was added to 25 μL of pre-treated and untreated BSA diluted to 1 $\mu\text{g}/\mu\text{L}$ BSA with ABC pH 7.8 in PCR tubes. Dosages of 50, 17, 6, 2, 0.6, 0.07, 0.0 μg trypsin were added to digest 25 μg of BSA. Tryptic reactions were incubated for 5 min at 37°C in a thermocycler and 50 μL of 2 \times Laemmli buffer was added to 50 μL of each reaction mix, incubated at 95°C for 10 min to quench trypsin activity and reaction products visualized on 10% SDS-PAGE gel (Fig. 2).

2.2.2. HTA-Protease Digests of BSA (Chaotrope versus heat/acid denaturation)

The pretreated and untreated 10 mg/mL BSA solutions in 20 mM phosphate citric acid buffer (pH 3) were diluted to 1 $\mu\text{g}/\mu\text{L}$ in the same buffer. Twenty-five μL of serially diluted HTA-Protease was added to 25 μL of prewarmed treated and untreated 1 $\mu\text{g}/\mu\text{L}$ BSA in PCR tubes and incubated for 5 min at 80°C in a thermocycler. Dosages of 0.5, 0.166, 0.056, 0.018, 0.006, and 0.002 μg of CB23726 and 0.78, 0.26, 0.087, 0.028, 0.0093, and 0.0031 μg of CB14057 was added to digest 25 μg of BSA. Fifty μL of 2 \times Laemmli buffer was added to each reaction mix, incubated at 95°C for 10 min and visualized on 10% SDS-PAGE gels stained with Coomassie brilliant blue (Fig. 2).

2.2.3. Trypsin digests of BSA (Relative proteolytic rates)

BSA samples were prepared to 2 mg/mL final concentration in 50 mM ammonium bicarbonate pH 7.8. Two tubes containing 500 μL of 2 mg/mL BSA/pH 7.8 were prepared. Next, 1 μg of trypsin (NEB #P8101S) was added to 1 mg of BSA in the pH 7.8500 μL (1:1000 E/S) reaction tube and incubated at 37°C . Timepoint samples of 25 μL were taken at the indicated times and analyzed for by tyrosine equivalents of liberated soluble protease products using standard methodologies [22] (Fig. 3).

2.2.4. HTA-Protease digests of BSA (Relative proteolytic rates)

BSA samples were prepared to 2 mg/mL final concentration in 20 mM phosphate/citrate pH 3 buffer. Two tubes containing 2 mg/mL BSA/pH 3.0 were prepared. Next, 0.1 μg of either CB14057 or CB23726 was added to 500 μL reaction tubes containing 1 mg of BSA at pH 3.0 (1:10,000 E/S) and incubated at 80°C . Timepoint samples of 25 μL were taken at the indicated times and analyzed for by tyrosine equivalents of liberated soluble protease products using standard methodologies [22] (Fig. 3).

2.2.5. HTA-Protease digests in citrate buffer (GFAP and histones)

Glial fibrillary acidic protein (GFAP, Cat. #AG230), and core

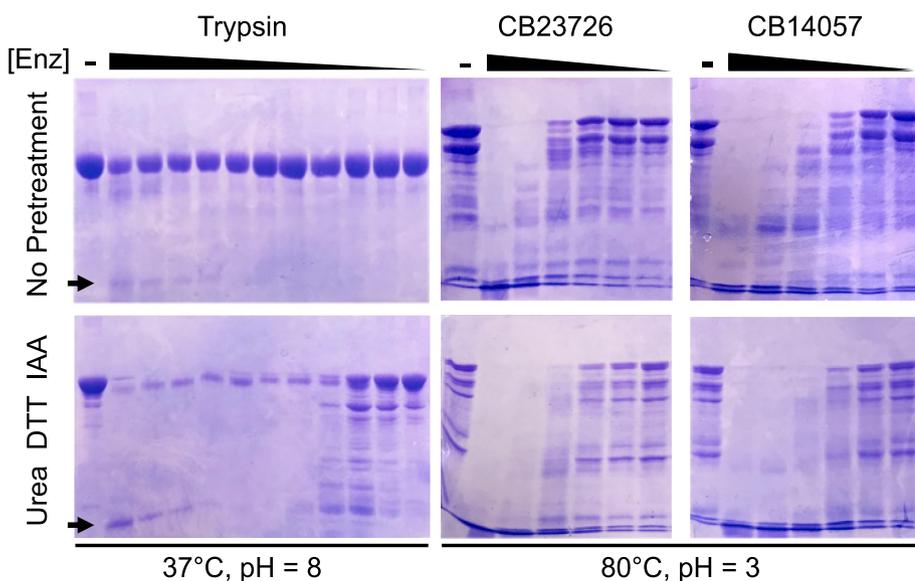


Fig. 2. Relative effects of Urea/DTT/IAA treatment on protease activity on BSA across serial enzyme dilutions (lower left, arrows indicate trypsin). Top row: five-minute trypsin and HTA-Protease digests performed without urea/DTT/IAA treatment. Bottom row: five-minute enzymatic digests performed after treatment with 6 M urea/10 mM DTT/32 mM IAA. Enzyme concentrations are diluted 3-fold with each dilution step. Left: tryptic digests of BSA (maximum dose 50 μg). Center: CB23726 digests of BSA (maximum dose 0.5 μg). Right: CB14057 digests of BSA (maximum dose 0.78 μg).

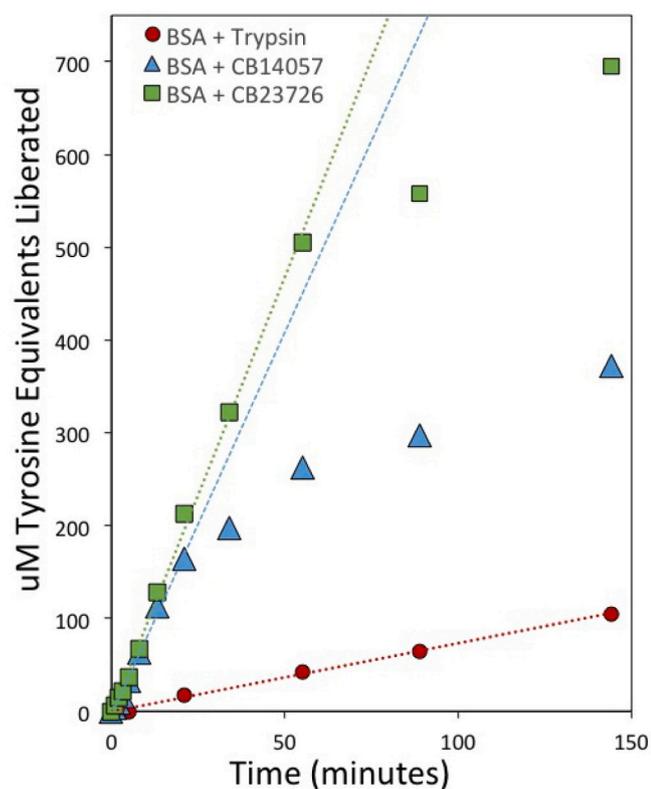


Fig. 3. Relative proteolytic rates showing CinderBio enzymes have significantly higher digestion rates (~10×) than trypsin with equal enzyme dose under respective optimal conditions. Plot displays μM tyrosine equivalents liberated over time for CB23726 (green), CB14057 (blue), and trypsin (red) with BSA substrate. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

histones (histones, Cat. #13–107) were purchased from Millipore Sigma. GFAP, and histones were reconstituted in 20 mM phosphate citrate buffer pH 3 at a concentration of 3 $\mu\text{g}/\mu\text{L}$. Seventy-five μL of substrate was distributed into PCR tubes and HTA-Proteases CB23726 and CB14057 were added at a concentration of 0.00032 and 0.00050 μg enzyme/ μg substrate respectively. Samples were incubated at 80 °C for 5 min and quenched by placing reactions on ice (Fig. 4).

Sequence Coverage (%)			
Substrate	Trypsin	CB14057	CB23726
H1	23	16	25
H2A-IV	63	52	49
H2B	58	61	62
H3.2	67	43	49
H4	70	62	40
Isoform 1	70	71	69
Isoform 2	71	73	72
Isoform 3	68	70	65
Median	67.5	61.5	55.5
No DTT/IAA			

Fig. 4. Sequence coverage values for one-step 5-min HTA-Protease protocol vs standard trypsin/urea protocol on purified chicken histones, and porcine GFAP. Trypsin-digested samples were reduced and alkylated before digestion and HTA-Protease reactions were not treated with DTT or IAA.

2.2.6. HTA-Protease digests of BSA (Formic acid vs. phosphate/citrate buffer)

BSA was reconstituted to a concentration of 1 $\mu\text{g}/\mu\text{L}$ in 200 μL of either 10 mM formic acid (FA) pH 3, or 20 mM phosphate citrate buffer at pH 3. 100 μL of each BSA solution was portioned into 2 PCR tubes, and ten Units (U) of each HTA-Protease CB23726 or CB1405 was added to each BSA sample. One Unit of activity was defined as the amount of enzyme that will release 1 micro-mole of tyrosine equivalents at 80 °C, pH 3, in 15-min with a hemoglobin substrate. Reaction mixes were incubated at 80 °C for 5 min and terminated by placing the PCR tubes on ice prior to MS/MS analysis (Fig. 6).

2.2.7. Trypsin and Chymotrypsin digests of *E. coli* lysate

For trypsin (NEB #P8101S) and chymotrypsin (Promega #V106A) digests, 50 $\mu\text{g}/\mu\text{L}$ stock solution of soluble *E. coli* extract or 27 mg/mL insoluble *E. coli* extract were thawed (see extract preparation above) and prepared by adding urea to a final of 8 M and incubated for 1 h incubation at 37 °C. The protein suspension for each enzyme was dialyzed overnight against their respective reaction buffers, then reduced by addition of DTT at a final concentration of 4 mM and incubated at 37 °C for 60 min. The reduced samples were then alkylated by adding IAA at a final concentration of 8 mM and incubated at room temperature for 60 min in the dark. The final concentration of soluble *E. coli* proteins was measured at an absorbance of 280 nm using Epoch take3 plate reader and adjusted to 10 $\mu\text{g}/\mu\text{L}$. Ten μL of protein suspension was split in two vials. Proteolytic digestion using trypsin was performed by adding trypsin at an enzyme/substrate ratio of 1:75 and incubated at 37 °C for 16 h. For enzymatic digestion using chymotrypsin, enzyme was also added at an enzyme/substrate ratio of 1:75 and incubating at 25 °C for 16 h (Fig. 7).

2.2.8. HTA-Protease digests of *E. coli* lysate

Soluble and insoluble *E. coli* extracts were dialyzed overnight against 20 mM phosphate citrate buffer, pH 3. The final concentration of proteins was measured at an absorbance of 280 nm using Epoch take3 plate reader and adjusted to 10 $\mu\text{g}/\mu\text{L}$ in 20 mM phosphate citrate buffer, pH 3. Ten μL of protein suspension was added to 2 PCR tubes and HTA-Protease, CB23726 and CB14057, were added at enzyme/substrate ratios of 1:13 and 1:50, respectively, and incubated at 80 °C for 16 h (Fig. 7).

2.2.9. HTA-Protease digests of WMP, BNL, BT, and HBC (Time-course analysis)

Milled, dry tissues were weighed into 1.5 mL safe-lock tube (Eppendorf #022363204) and resuspended in 20 mM phosphate citrate (PC) buffer, pH 3.0, at 5 mg tissue/mL buffer. HTA-Proteases were added at 64 U/mg starting material and samples were incubated at 80 °C with shaking. Aliquots of each digest were taken at 2 h and 24 h, immediately frozen, and stored at -80 °C until analysis. Before analysis, samples were spun at 18,000 rcf to remove insoluble debris and the supernatant was reserved (Fig. 8).

2.2.10. Tryptic digests of K562 human cell lysate

K562 tryptic digests were prepared by aliquoting 20 μg of K562 cell lysate (Promega # V6941) into a 1.5 mL safe-lock tube (Eppendorf #022363204) and bringing to 25 μL with 8 M urea, 50 mM Tris, pH 8. DTT was added to 10 mM and samples were incubated at 37 °C for 30 mins. IAA (50 mM) in 50 mM Tris, pH 8, was added to a 2.2 M excess over DTT and samples were incubated for 15 min in the dark. Samples were diluted to a final volume of 300 μL and trypsin (0.3 μg) was added. Samples were incubated at 37 °C for 16 h before being acidified to 1% FA and stored at -80 °C (Figs. 9 & 10).

2.2.11. HTA-Protease digests of K562 human cell lysate

CB14057 and CB23726 digests for trypsin comparison: HTA-Protease K562 digests were prepared by aliquoting 20 μg of K562 cell lysate

(Promega # V6941) into a 1.5 mL safe-lock tube (Eppendorf #022363204) and bringing to 300 μ L with 100 mM phosphate citrate buffer, pH 3.0, supplemented with 10 mM DTT and 0.2 U/ μ g of either CB14057 or CB23726. Samples were incubated in a heat block at 80 °C for 90 mins with vortexing every 20 mins. After digestion, samples were transferred to ice and IAA was added to 20 mM for reaction quenching and alkylation (Figs. 9 & 10).

2.2.12. Sample Quantification and Desalting

Unless otherwise indicated, samples from all methods were quantified after digestion using the Pierce™ Quantitative Colorimetric Peptide Assay kit (Thermo Scientific #23275) according to the manufacturer's protocol. Aliquots of digested samples containing 10 μ g of protein were desalted using Pierce™ C18 Spin Tips (Thermo Scientific #84850) according to the manufacturer's protocol.

2.3. MS/MS acquisition

2.3.1. LTQ XL (histones and GFAP)

Global untargeted proteomics for histone and GFAP analysis was carried out on an LTQ XL Linear Ion Trap mass spectrometer (Thermo Fisher Scientific) coupled to an Agilent 1100 capillary LC system through a nanoelectrospray LC – MS interface. Eight μ L of sample containing 4 μ g of digested peptides were separated by reverse-phase chromatography (RP-HPLC) using nanoelectrospray capillary columns packed in-house with Agilent Zorbax SB-C18 stationary phase (14 cm \times 75 μ m id, 5 μ m). RP-HPLC separation was performed over 60 mins using 0.1% FA in ddH₂O and 0.1% FA in ACN as the mobile phases. The gradient started at 5% ACN for 5 min, then ramped up to 55% ACN for 40 min, 95% ACN for 5 min, and re-equilibrated with 5% ACN for 10 min. The flow rate was set at 500 nL/min. Data acquisition was performed using Xcalibur™ (version 2.0 SR2). The mass spectrometer was operated in the positive ion mode. Each survey scan of m/z 400–2000 was followed by collision-induced dissociation (CID) MS/MS of the 3 most intense precursor ions with an isolation width of 2.00 m/z . Dynamic exclusion was performed after fragmenting a precursor 2 times within 30 s for a duration of 60 s with a 1.5 ppm mass tolerance. Singly charged ions were excluded from CID selection. Normalized collision energies of 35 eV were employed. MS1 and MS2 scans were performed at a resolution of 0.7 FWHM.

2.3.2. LTQ XL (*E. coli* lysate, BSA)

Global untargeted proteomics for *E. coli* lysate and BSA analysis was performed on the same column and LC-MS system mentioned in the previous section. Eight μ L of each sample containing 4 μ g of protein was separated over a 170 min gradient starting at 5% acetonitrile for 10 min, then ramped up to 8% ACN for 5 min, 35% ACN for 113 min, 55% ACN for 12 min, held at 95% ACN for 15 min and re-equilibrated with 5% ACN for 15 min. The flow rate was set at 300 nL/min. Data acquisition was performed using Xcalibur™ (version 2.0 SR2). The mass spectrometer was operated in the positive ion mode. Each survey scan of m/z 200–2000 was followed by collision-induced dissociation (CID) MS/MS of the 3 most intense precursor ions with an isolation width of 2.00 m/z . Dynamic exclusion was performed after fragmenting a precursor 2 times within 30 s for a duration of 60 s with a 1.5 ppm mass tolerance. Singly charged ions were excluded from CID selection. Normalized collision energies of 35 eV were employed. MS1 and MS2 scans were performed at a resolution of 0.7 FWHM.

2.3.3. Orbitrap Velos (WMP, BNL, BT, HBC)

Global untargeted proteomics for WMP, BNL, BT, and HBC digest time-course analysis was carried out on an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) coupled to an Eksigent nanoLC-2D system through a nanoelectrospray LC – MS interface. Four μ L of each sample containing 5 μ g of protein was injected into a 20 μ L loop using the autosampler. The analytical column was then switched on-line

at 600 nL/min over an in-house-made 100 μ m i.d. \times 150 mm fused silica capillary packed with 2.7 μ m CORTECS C18 resin (Waters; Milford, MA). After 10 min of sample loading at 600 nL/min, each sample was separated on a 90-min gradient. Flow was maintained at 600 nL/min of 2% ACN with 0.1% FA from minutes 0 to 3, followed by a linear gradient from 2 to 6% ACN from 3 min to 4 min and a subsequent gradient from 6 to 12% ACN from 4 to 20 min. From minutes 20 to 20.5, flow rate was lowered to 350 nL/min while maintaining 12% ACN, remaining at 350 nL/min for the rest of the run. Flow rate adjustment was followed by linear gradients from 12 to 30% ACN from 20.5 to 67 min and further to 45% ACN from 67 to 72 min. Gradient elution was followed by a linear increase to 80% ACN from 72 to 75 min and further to 90% ACN from 75 to 81 min to elute the remaining peptides. The column was then re-equilibrated with 2% ACN in 0.1% FA from minutes 81 to 90. LC mobile phase solvents consisted of 0.1% formic acid in water (Buffer A) and 0.1% formic acid in acetonitrile (Buffer B, Optima™ LC/MS, Fisher Scientific, Pittsburgh, PA). Data acquisition was performed using Xcalibur™ (version 4.1) software. The mass spectrometer was operated in the positive ion mode. Each orbitrap survey scan of m/z 300–1800 was followed by ion trap collision-induced dissociation (CID) MS/MS of the 20 most intense precursor ions with an isolation width of 2.5 m/z . Dynamic exclusion was performed after fragmenting a precursor 2 times within 15 s for a duration of 30 s with a 1.5 ppm mass tolerance. Singly charged ions were excluded from CID selection. The orbitrap was used for MS1 detection at a resolution of 60,000 and MS2 detection was performed in the ion trap. Normalized collision energies of 35 eV were employed using helium as the collision gas.

2.3.4. Fusion Lumos (K562 lysate)

Global untargeted proteomic comparisons of trypsin to CB14057 and CB23726 using K562 lysate (Promega # V6941) were carried out on a Fusion Lumos Tribid mass spectrometer (Thermo Fisher Scientific) coupled to an EASY-nLC 1200 (Thermo Fisher Scientific) through a nanoelectrospray LC – MS interface. Seven μ L of each sample containing 3 μ g of protein was injected into a 20 μ L loop using the autosampler. The analytical column was then switched on-line at 400 nL/min over an in-house-made 100 μ m i.d. \times 150 mm fused silica capillary packed with 2.7 μ m CORTECS C18 resin (Waters; Milford, MA). LC mobile phase solvents consisted of 0.1% FA in water (Buffer A) and 0.1% FA in 100% ACN (Buffer B, Optima™ LC/MS, Fisher Scientific, Pittsburgh, PA). After 22 μ L of sample loading at a maximum column pressure of 700 bar, each sample was separated on a 120-min gradient at a constant flow rate of 400 nL/min. The separation gradient for cell fractions consisted of 4% buffer B from 0 to 3 min, followed by a linear gradient from 4 to 32% buffer B from 3 min to 105 min. Gradient elution was followed by a linear increase to 55% buffer B from 105 to 110 min and further to 95% buffer B from 110 to 111 min. Flow at 95% buffer B was maintained from 111 min to 120 min to remove remaining peptides. Data acquisition was performed using Xcalibur™ (version 4.5). The mass spectrometer was operated in the positive ion mode. Each survey scan of m/z 300–1800 was followed by higher energy collisional dissociation (HCD) MS/MS (30% collision energy) with an isolation width of 1.6 m/z . The orbitrap was used for MS1 and MS2 detection at resolutions of 120,000 and 50,000, respectively. Dynamic exclusion was performed after fragmenting a precursor 1 time for a duration of 20 s. Singly charged ions were excluded from HCD selection.

2.4. Data analysis

2.4.1. LTQ XL Data Searching (PEAKS)

Raw data was searched using PEAKS X Pro version 10.6 (Bioinformatics Solutions, Inc.). Precursor tolerance was set to \pm 0.8 Da and fragment tolerance was set to \pm 0.7 Da. All searches were performed with fixed modifications set to carbamidomethyl (C) and no enzyme specificity. All databases additionally contained common contaminants derived from version 1.1 of the CRAPome [23]. Databases and

modification were chosen by analyzed sample as follows: BSA data was searched against a database containing the UniProt entry for BSA with variable modifications set to oxidation (M). *E. coli* lysate data was searched against the SwissProt database (4140 sequences, downloaded 2/12/2018) restricted to *E. coli* using oxidation (M) as the variable modification. Chicken histone data (Millipore 13–107) was searched against a database containing UniProt sequences for all core chicken histone proteins including H1 and H2 using oxidation (M), acetyl (CKS and protein N-terminal), methylation (KR), dimethylation (KR), and trimethylation (K) as variable modifications. Purified porcine GFAP data (Millipore AG230) was searched against a database containing all UniProt porcine GFAP sequences with variable modifications set to oxidation (M). Results were filtered to 1% false discovery rate (FDR) at the peptide and protein level.

2.4.2. Orbitrap Data Searching (Mascot)

Raw data was searched using an in-house Mascot server (Version 2.5, Matrix Science). Precursor tolerance was set to ± 15 ppm and fragment tolerance was set to ± 0.6 Da. WMP data was searched against SwissProt (17,016 sequences, downloaded 2/12/2018) restricted to *Mus musculus*. Data for BNL and BT was searched against UniProt (32,231 sequences, downloaded 2/12/2018) restricted to *Bos taurus*. All databases additionally contained common contaminants derived from version 1.1 of the CRAPome [23]. All searches were performed without cleavage limitations. Fixed modifications were set as carbamidomethyl (C). Variable modifications were set as oxidation (M), oxidation (P) (hydroxyproline), Gln- > pyro-Glu (N-term Q), and acetyl (Peptide N-term). Results were filtered to 1.3% FDR at the peptide level and 0.7% FDR at the protein level using Scaffold 4.8.9.

2.4.3. Fusion Lumos Data Searching (PEAKS)

K562 lysate data generated using the Fusion Lumos MS system were searched using PEAKS X Pro version 10.6 (Bioinformatics Solutions, Inc.). Precursor tolerance was set to ± 15 ppm and fragment tolerance was set to ± 0.08 Da. Data was searched against SwissProt (20,379 sequences, downloaded 3/15/2021) restricted to *Homo sapiens* with additional common contaminants derived from version 1.1 of the CRAPome [23]. All searches were carried out without cleavage limitations. Variable modifications were set as carbamidomethyl (C), oxidation (M), oxidation (P) (hydroxyproline), Gln- > pyro-Glu (N-term Q), and acetyl (Peptide N-term). Searches were performed with all replicates from each sample condition grouped to obtain protein coverage across all analyzed replicates. Results were filtered to 1% FDR at the peptide and protein level.

2.4.4. Coefficients of Variation (CVs)

Spectral count values were calculated within PEAKS X Pro and exported for further analysis. For each protein identified in all analyzed K562 lysate samples, the mean abundance value (μ) and the standard deviation (σ) of spectral count measurements across triplicate analyses were calculated. The coefficient of variation (CV) for each protein was then determined using the formula: $CV = (\sigma / \mu) \times 100$. The resulting CV values were expressed as percentages.

2.4.5. iceLogo Generation

Cleavage specificity for CB14057, and CB23726 was visualized using iceLogo [24]. In short, identified peptides were mapped onto the entire protein sequence and cleavage sites were identified in the context of the protein amino acid sequence. The abundance of amino acids at the P1 position was calculated and represented in iceLogo format.

3. Results and discussion

3.1. HyperThermoacidic Archaeal Proteases (HTA-Proteases®)

Thermophilic enzymes have revolutionized the laboratory practices

of modern biology with the most noted impacts arguably coming from the invention of PCR and the acceleration of crystallographic protein structural studies. The study of organisms thriving in hot and acidic volcanic springs has led to technological developments in the commercial production of hyperthermoacidic archaeal enzymes (HTA-enzymes), the most thermally and acid stable enzymes yet available [17,25]. These newly available HTA-enzymes include several highly active and specific proteases with previously unavailable characteristics that hold great promise to accelerate advances in several areas of proteomic research. The two HTA-Proteases under study here have extreme thermal and pH optima that enable denaturing proteolytic sample reaction conditions at approximately 80 °C and pH 3 (Fig. 1). Here we apply two of these proteases to several varied proteomics analyses to investigate their utility and unique capabilities for proteomics.

3.2. Heat and Acid Denaturation in HTA-Protease Digests Obviates Addition of Chaotropes

To visualize the effects of thermal and pH sample denaturation on protease activity, we compared proteolytic digestion of BSA under both native and denaturing conditions over a range of protease concentrations (Fig. 2). Immediately apparent is the ineffective tryptic digestion of untreated BSA in the absence of chaotropes and reduction/alkylation under brief but otherwise standard digestion conditions. Notably, even trypsin doses 100 times greater than standard concentrations, and visible by Coomassie blue staining, show only marginal protease activity on BSA without pretreatments (Fig. 2, top). Unsurprisingly, BSA denatured with urea, reduced, and alkylated in a multi-step treatment process is effectively digested by trypsin. In stark contrast, both HTA-Proteases show significant proteolytic activity on the untreated BSA and only relatively small gains in activity are observed after the multi-step chaotrope denaturing protocol (Fig. 2, right). Interestingly, even the highest dose of trypsin shows incomplete digestion of full-length BSA while no full-length BSA is visible at the maximal HTA-Protease dose, approximately 100 times less (mass/mass) than trypsin. Importantly, DTT is not an optimally effective reductant at pH 3, and if sample pH is set to 3 before reduction, an acid-tolerant reductant such as TCEP should be used if complete sample reduction is a priority. However, HTA-Proteases show effective digestion of BSA both under non-optimal reduction conditions and in the absence of reductant (Fig. 2). These experiments indicate that optimal reaction conditions for HTA-Proteases effectively denature target proteins and HTA-Protease digests do not require chaotropic chemicals (e.g. urea, guanidine), reduction, alkylation, or the associated sample manipulations, dilutions, and cleanup steps.

3.3. HTA-Protease Digestion Rates are Faster than Trypsin

To assess the relative reaction rates of HTA-Proteases as compared to tryptic digestion rates, parallel protease reactions digesting BSA were carried out at 37 °C, pH 7.8 and 80 °C, pH 3.0 to match the most commonly used conditions and determined optima for trypsin and HTA-Proteases, respectively (Fig. 3). Protease reactions were run at the respective pH and thermal optima and sampled over time (Fig. 3). This study includes only tryptic digests run at 37 °C but we acknowledge that higher temperature tryptic digest protocols are available. Protease activity was quantified at the noted time points by tyrosine equivalents of liberated soluble protease products using standard methodologies [22]. We found that the initial reaction rates for the HTA-Proteases were approximately 10-fold faster than the rate of trypsin on BSA under normalized reaction conditions. Given the greater kinetic energy available at 80 °C relative to the 37 °C trypsin reactions, it is not surprising that the HTA-enzyme proteolytic reaction rates are significantly faster. Importantly, in addition to greater kinetic energies at 80 °C relative to 37 °C, substrates in the hot/acid reactions are also retained in a denatured state thereby likely increasing protease access to peptide bond

cleavage sites (Figs. 1 & 2). The HTA-Protease reactions consume enough substrate within one hour to deviate from linearity while the trypsin reaction remains linear for over two hours. Irrespective of the precise mechanism(s), the initial hyperthermoacidic reaction rates are approximately 10-fold more rapid than trypsin (Fig. 3).

3.4. Competitive Protein Sequence Coverage is Achieved with a Simple Five-Minute Protocol

Initial tests for functional utility of HTA-Proteases for proteomic applications involved one-step reactions where HTA-Proteases were added to samples in acidic buffer and incubated at 80 °C and pH 3 for five minutes, quenched on ice, and analyzed via LC-MS/MS. In this set of experiments commercially purified target proteins were compared, 1) the five component proteins of core histones, and 2) the three isoforms of glial fibrillary acidic protein (GFAP), chosen to represent a variety of charge, mass, globularity, and modifications. The comparative tryptic digests were carried out with standard reduction, alkylation, urea denaturation, dilution, and over-night digestion. All samples were analyzed with the same proteomic computational analysis protocol without limitations on enzyme cleavage or any alteration to identification thresholds, columns, buffers, or MS settings between enzymes. These results reveal that even without reduction, alkylation, or optimizing MS protocols, the sequence coverage achieved with a one-step/five-minute protocol was competitive with the nine-step 18-h trypsin/urea results (Fig. 4).

3.5. Simplified Proteomics Sample Preparation Protocols Enabled by HTA-Proteases

The characteristics of the HTA-Protease described in part above, have led to a simplified proteomics sample preparation protocol used for the following studies (Fig. 5). When using HTA-Proteases, samples are resuspended in simple acidic buffers containing low concentrations of reductant. As stated above, DTT does not function optimally at pH 3, yet HTA-Proteases display effective digestion of disulfide-containing proteins under these conditions. A very small quantity of HTA-Proteases is then added (1/625–1/2000 of the sample mass) and the reaction heated to 80 °C for five minutes. Maximum optimal reaction times were later determined to be approximately 2 h as described in following sections. Samples were then treated with IAA to match conditions for tryptic digests and database searching.

Reduction and alkylation treatments are intended to break disulfide bonds and prevent reformation, respectively. In stark contrast to the strict dependence of trypsin on these treatments, HTA protease reactions

appear to be largely independent of reduction and alkylation treatments (Figs. 2 & 4). The use of IAA is likely not needed in HTA-Protease acidic reaction buffers as the pH of 3.0 should maintain protonated cysteines, thereby preventing disulfide reformation [26]. To further simplify HTA-Protease protocols, we next tested HTA-Protease activity directly in dilute formic acid (FA) due to the ubiquitous use of FA as a mobile phase for LC-MS/MS chromatography. Simple HTA-Protease digests of BSA were run and analyzed by MS/MS without reduction, alkylation, or chaotrope treatments (Fig. 6). These data reveal only modest changes in peptide identifications and sequence coverage of BSA in response to both tested buffering systems. Importantly, the absolute and relative differences observed between buffering systems were consistent and modest on this simple substrate. These data indicate that HTA-Proteases are functional in various buffering systems, and importantly, in dilute (10 mM) formic acid with no additives. Taken together with the data in Figs. 2 and 4, in at least some cases, the prototypical reduction/alkylation reactions, manipulations, and cleanup steps can be omitted. These HTA-Protease capabilities can further increase the simplicity and speed of sample preparation. Moreover, omitting reduction/alkylation may have utility where such treatments are undesirable such as when native cysteine status is of interest or off-target IAA chemistry modifies protein functional groups of interest. Direct sample digestion in FA without the addition of reductant or alkylating compounds allows samples to be analyzed without need for desalting or buffer exchange steps that can increase cost and variability while also reducing the risk of MS contamination. These new capabilities may have significant utility in automated sample preparation and high-throughput proteomic applications, including clinical proteomics.

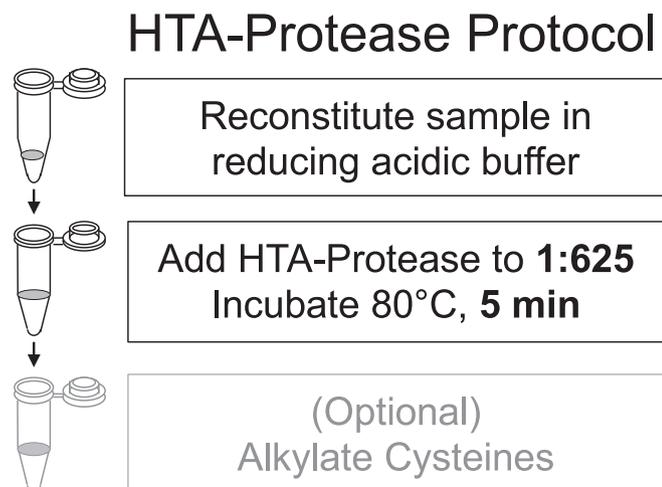


Fig. 5. Simple HTA-Protease proteomics sample preparation protocol.

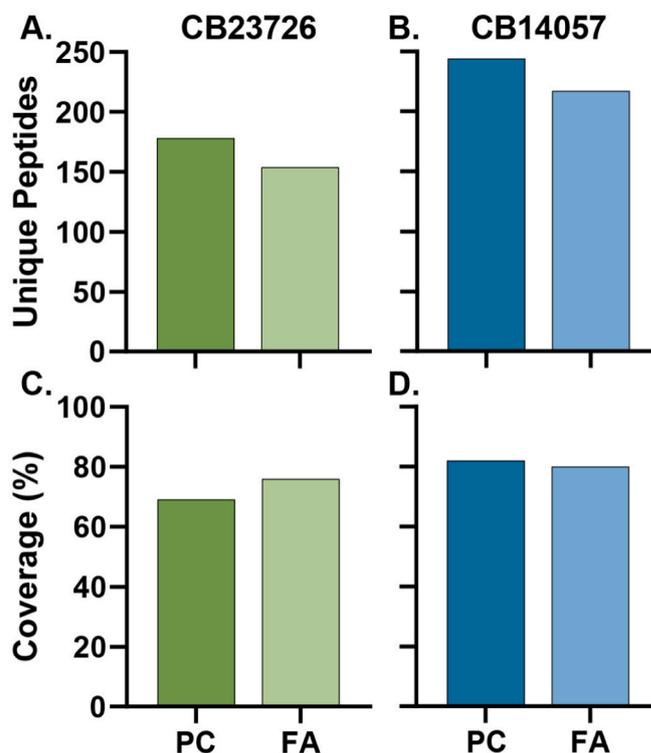


Fig. 6. Comparison of phosphate/citrate (PC) and formic acid (FA) buffers on HTA-Protease activity on BSA. Unique peptide identifications for CB23726 (A) and CB14057 (B) digests in either PC or FA. Average sequence coverage of BSA resulting from CB23726 (C) and CB14057 (D) digests using either PC or FA reaction buffers.

3.6. HTA-Protease Specificity and Cleavage Entropy is Akin to Chymotrypsin

Initial proteomic analysis of purified proteins and *E. coli* extracts gave preliminary indication of the cleavage specificity of the HTA-Proteases studied here (**data not shown**). To assess the relative cleavage specificity of HTA-Proteases more accurately, we used digests of the complex mixture of proteins in *E. coli* crude extracts and WMP. The resulting peptide data were used to generate both iceLogos [24] and to quantitatively measure protease specificity using cleavage entropy [27]. These studies compared the commercial proteomics-grade modified trypsin (NEB #P8101S) and chymotrypsin (Promega #V106A) with CinderBio proteomics-grade HTA-Protease preparations. Trypsin and chymotrypsin digests were run according to manufacturer recommended procedures (~17-h digests), while HTA-Protease digests were carried out according to the simple HTA-Protease protocol (1h). These data included peptide identifications from WMP and crude *E. coli* extracts to avoid substrate bias and to improve the statistical significance. The cleavage entropy study compared standard overnight urea-denatured digests for the commercial enzymes to 1-h HTA-Protease digestion protocols without urea, all digesting soluble *E. coli* extracts [27]. These data reveal that HTA-Proteases have a cleavage entropy akin to proteomic grade chymotrypsin (Fig. 7). As noted previously and apparent in Fig. 7, a major advantage of trypsin is the remarkable specificity for lysine and arginine residues. While commonly used proteases are significantly less specific than trypsin, it is clear from our studies that the specificity inherent to HTA-Proteases is sufficient for practical proteomic studies. Importantly, this together with the E, L, F cleavage and the speed and simplicity of sample preparation indicates the utility of these new enzymes in routine and specialized proteomics applications.

3.7. HTA-Protease Reactions Show Time-Dependent Cleavage at Aspartate

Heated acidic conditions are well known to hydrolyze peptide bonds in protein samples with cleavage efficiencies often depending on the specific amino acid and hydrolysis conditions [16,28,29]. We had noted in earlier experiments that digestion for over 10 h showed increasing cleavage at aspartate (D). In addition, studies with synthetic peptide libraries revealed elevated cleavage at aspartic acid residues in the absence of enzyme (**data not shown**). Because HTA-Proteases function optimally in heated acid reactions (80 °C and pH 3), we next investigated whether acid hydrolysis interfered with HTA-Protease proteomic applications.

To evaluate whether chemical cleavage at aspartate or other amino acids would interfere with HTA-Protease proteomic analyses, we digested complex samples of bovine blood clots and WMP at 80 °C and pH 3 for either 2 h or 24 h and analyzed the digests by LC-MS (Fig. 8). The resulting peptide identifications were used to generate iceLogos to visualize the relative abundance of cleavage sites (Fig. 8). The number of aspartate cleavages increased significantly in 24-h digests relative to 2-h digests. For practical purposes, significant aspartate cleavage can be largely avoided with HTA-Protease digestions that are no longer than 2 h. Cleavage at aspartate was by far the most significant additional cleavage site showing increased abundance at longer digestion times. Not surprisingly, the HTA-catalyzed cleavage of proteins appears to be far more rapid than the heated acid chemical cleavage of peptide bonds, and therefore can be avoided in these reactions by adjusting HTA-Protease dose and/or reaction times. Importantly, no increase in unique peptides or sequence coverage was observed in digestions over 2 h, in fact, both parameters were decreased with longer digestion times (**data not shown**). Therefore, the optimal HTA-Protease digestion times are between five minutes and two hours and interference from the slower chemical cleavage is largely negligible.

3.8. HTA-Proteases Increase Proteome Coverage with Data Orthogonal to Tryptic Digests

We next investigated the potential to increase proteome coverage by analyzing the orthogonal HTA-Protease datasets in combination with data generated using standard tryptic digests. Proteolytic digests of commercially available K562 human cell lysates were carried out with trypsin and both HTA-Proteases and reduction with DTT and alkylation with IAA were carried out on all samples in their respective digest buffers to maintain chemical parity between the reactions. Notably, these reactions were carried out with HTA-Protease aliquots that were stored in aqueous solutions at room temperature in the laboratory for over four-years prior to these experiments. Non-specific enzyme searches revealed significant populations of peptides uniquely identified in each enzyme group and of the 43,774 peptides identified across the 3 digests, only 4 peptides were shared between trypsin and HTA-Proteases (Fig. 9). This not only confirms the non-overlapping cleavage specificity of HTA-Proteases with trypsin but also reaffirms the specificity of these HTA-Proteases. Moreover, HTA-Protease digests provided an additional 19,619 peptide identifications over the trypsin digests alone, representing a 45% increase in unique peptide sequences. Of these additional peptides, 9157 were identified with CB14057 while 12,685 were identified with CB23726. Despite similar cleavage preferences, CB23726 provides 1.4-fold more uniquely identified peptides than CB14057 in these experiments. However, digestion with CB14057 provides an additional 6934 uniquely identified peptide sequences, demonstrating significant orthogonal benefit to using all three proteases when optimal protein coverage is a priority. Comparison of modification levels between trypsin and HTA-Protease peptide datasets revealed methionine oxidation was present on <1% of identified peptide spectral matches (PSMs) from all samples. Trypsin-digested samples displayed slightly greater levels of oxidation (0.85% of PSMs) than HTA-Protease digests (0.37% of PSMs). Despite the potentially inefficient reduction and alkylation of HTA-Protease digests with DTT and IAA at pH 3, we find that 19.6% of tryptic peptides, 15.3% of CB14057 peptides, and 12% of CB23726 peptides contain cysteines. In contrast, 19.6% of tryptic peptides showed alkylation (carbamidomethylation) while HTA-Protease peptides showed <0.1% in these experiments, a likely function of IAA chemistry at pH of 3. Coefficients of variation (CV) for spectral counts of proteins identified in all samples were calculated for trypsin (12.8%), CB23726 (16.5%), and CB14057 (24.2%) revealing reasonable reproducibility for the nascent HTA-Protease protocols.

In contrast to the effectively non-overlapping peptide identifications, 1083 of the 2909 proteins identified are common to tryptic and HTA-Protease digests, a likely reflection of protein abundance in the samples. Regardless, the simple and nascent HTA-Protease protocols produced approximately 60% of the protein identifications achieved with the tryptic protocol and contributed 91 unique protein identifications. Of the 91 unique protein identifications, 64 are attributed to CB14057 and 35 from CB23726. Despite the greater number of peptide sequences identified by CB23726, CB14057 provides more unique protein identifications, further highlighting slight variations in HTA-Protease behavior and utility. Notably, these two HTA-Proteases are compatible with simultaneous co-digestion in two-enzyme reactions, akin to the Lys-C and trypsin sequential digests reported to improve coverage [30,31]. The identifications unique to HTA-Proteases include five histone H2 and H3 subunit variants, Ubiquitin-60S ribosomal protein L40 [32] and several other proteins implicated in known disease. Curiously, nucleoproteins and mitochondrial proteins, many of which are membrane proteins, appear to be highly represented among these unique identifications (**Supplementary table 1**).

3.9. HTA-Proteases Markedly Increase Nucleoprotein and Membrane Protein Coverage

Given the apparent bias of HTA-Protease coverage and identification

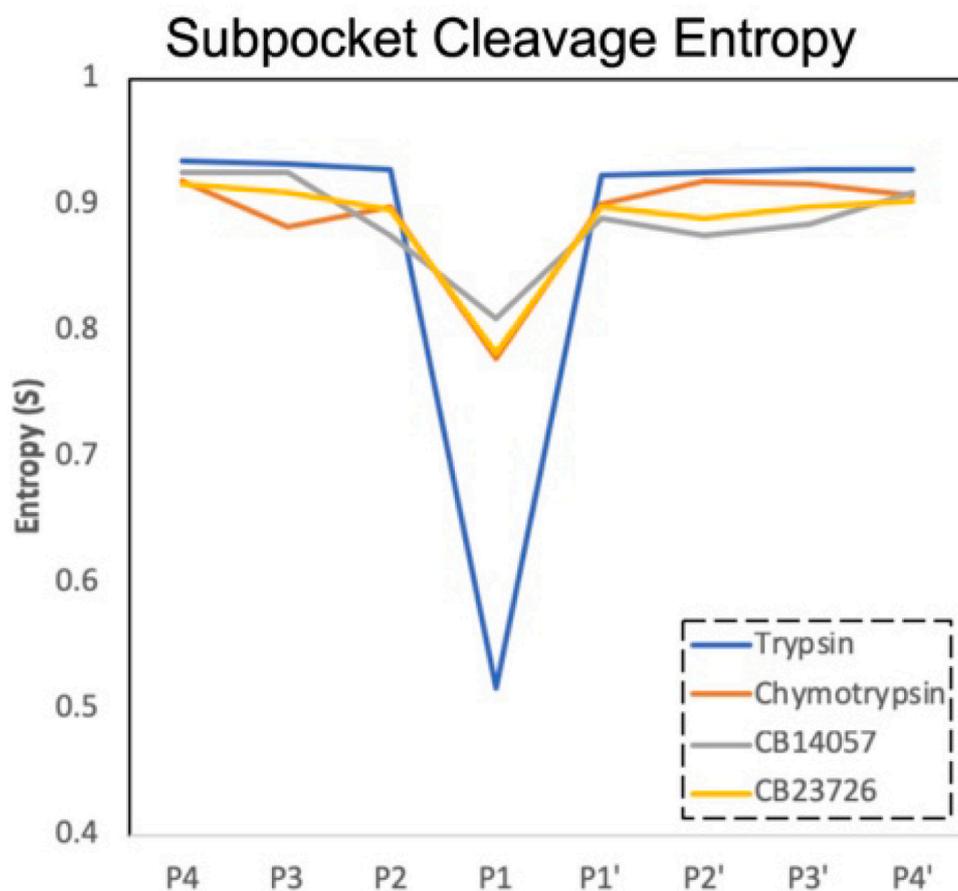
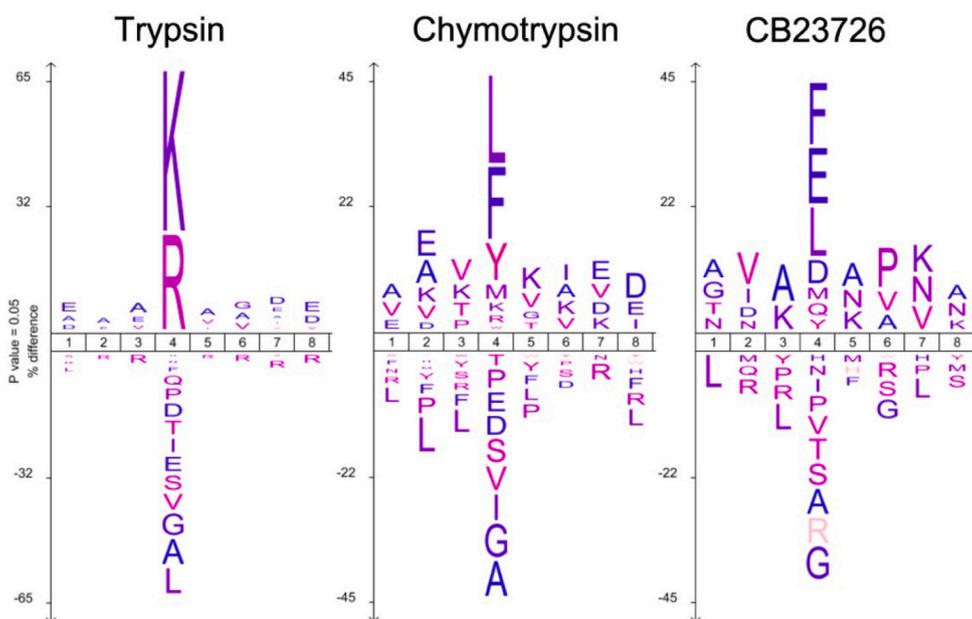


Fig. 7. Comparison of cleavage specificity (top) and cleavage entropy (bottom) between trypsin, chymotrypsin, and CB23726.

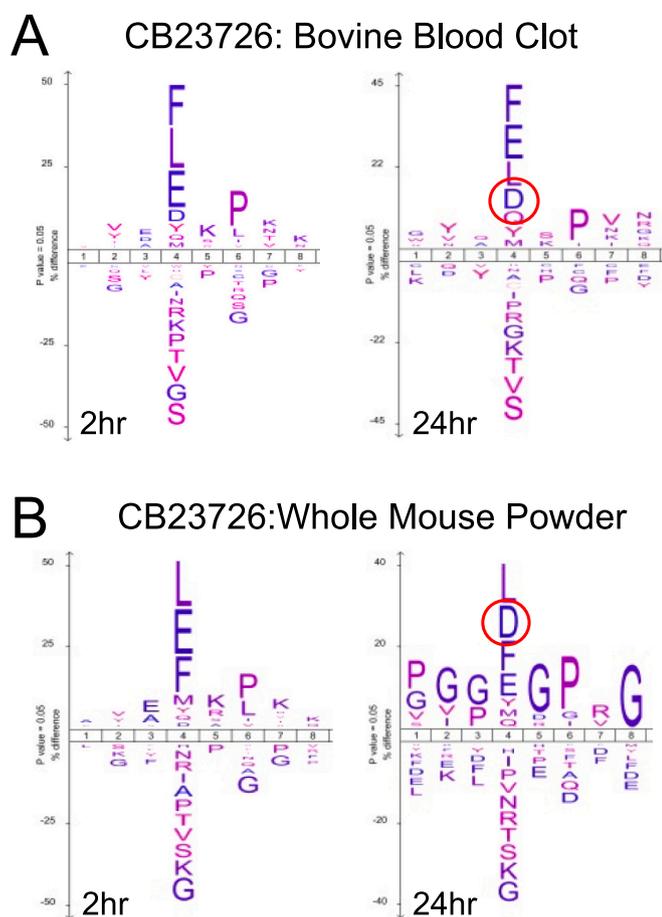


Fig. 8. Protein cleavage at aspartic acid is evident in extended HTA-Protease digestions. (A) CB23726 digests of bovine blood clot performed for 2 h (left) and 24 h (right). (B) CB23726 digests of whole mouse powder performed for 2 h (left) and 24 h (right). Abundance of peptides with terminal aspartate in longer digests is noted with a red circle. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

gains towards different protein classes, we next scrutinized the identity and nature of proteins that were either uniquely identified by HTA-Proteases or showed large increases in sequence coverage. For these analyses we first categorized the 91 protein identifications that were unique to HTA-Protease data sets. Interestingly, approximately 40% of the unique identifications were membrane-associated proteins and 20% were nuclear and nucleic acid interacting proteins (including histones). To further query these data for a protein-class bias, we plotted all 2809 proteins identified by trypsin according to sequence coverage percentages with and without the addition of HTA-Protease digest data. Importantly, 53% of the identified proteins showed improved coverage with inclusion of HTA-Protease data. (Fig. 10A). We next evaluated 285 proteins where the sequence coverage was increased by HTA-Protease data by 20% or more relative to trypsin alone (Fig. 10A, **dashed line**). Strikingly, 148 of the 285 queried proteins (52%) were nucleic acid-binding proteins, including a large representation of ribonucleo-proteins and histones (**Supplementary table 2**). Of the same 285 proteins, 40 were identified as mitochondrial proteins (14%), many of which are membrane proteins (Fig. 10A). An additional 13% of the proteins with significantly improved coverage were annotated as either chaperones or nuclear proteins. Taken together, a remarkable 79% of the proteins with 20% or more increased coverage classified into these four categories. To illustrate the significance of the gains in sequence coverage, we selected 11 representative proteins with the greatest coverage improvement from these classes and show the relative

coverage contributions of the respective enzymes (Fig. 10B). Overall, inclusion of HTA-Protease data in this analysis identified an additional 91 proteins, increased peptide identification by 45%, and increased sequence coverage for over half of the identified proteins despite inefficient reduction of proteins with DTT at pH 3. Possibly most striking, 79% of the proteins with the largest increases in coverage (>20%) classified into four discrete categories with 52% of those being nucleic acid binding proteins (Fig. 10).

4. Concluding remarks

Technological advances have driven the maturation and growth of modern proteomics and enabled the use of proteomic technologies in nearly all aspects of modern biological research. Trypsin was discovered over one hundred and fifty years ago [33] and is the predominant enzyme used in peptide-centric proteomic workflows exhibiting remarkable specificity and generating high-quality fragmentation spectra and reliable identifications. In fact, trypsin is so ubiquitous that nearly all facets of modern proteomics are optimized for trypsin by design or default, and tryptic digests account for 96% of the deposited data sets in the Global Proteome Machine Database [34]. However, trypsin is known to be inherently fragile, subject to autolysis, and requires a time-consuming multi-step protocol using chaotropes, dilutions, manipulations, and cleanup steps. Consequently, recognizing the value of alternative enzymes capable of expanding and diversifying the range of proteomic approaches is of interest to many researchers [9,35–37]. Here we introduce two novel HTA-Proteases and methods to offer entirely new capabilities for proteomic approaches. The new capabilities of HTA-Proteases promise to extend the reach of proteomics beyond what is possible with trypsin alone to new and powerful areas of clinical and fundamental proteomics applications.

The current study presents experimentation designed to test the utility of HTA-Proteases for bottom-up proteomic applications and to identify some of the key advantages over traditional tryptic digests. We first present the advantages of sample digestion in low pH and at high temperatures with HTA-Proteases which act faster and are optimally active in denaturing conditions (Figs. 1 & 2). We show that HTA-Protease reaction conditions efficiently denature mesophilic sample proteins without addition of chaotropes, sample manipulation, or reduction and alkylation (Fig. 2). Furthermore, the utility of heated/acid reactions include a drastically simplified and abbreviated sample preparation protocols (Fig. 5). The typically employed nine-step overnight tryptic protocol can be reduced to 1-step and 5 min without dilution and effectively no sample manipulation or associated losses. Moreover, a one-step protocol produces peptides useful for proteomic analysis and show significant protein coverage across a variety of different samples (Figs. 4, 6-10). In recent years there have been many efforts to increase the throughput and turnaround time of proteomic analyses for use in large-scale biomarker studies and other applications [38], further highlighting the utility of HTA-Proteases with high rates of digestion and one-step protocols.

The identified E, L, F, cleavage specificity is consistent and reproducible irrespective of sample complexity and across different mass spectrometers, database search tools, and disparate sample types in various proteomics laboratories (Figs. 4, 7-10). The ultra-stable nature of HTA-Proteases permits ambient storage in aqueous buffers for over two years, mass/mass dosages approximately 1/100 of that required for trypsin, and no detection of interfering autolysis products. We further demonstrate that HTA-Proteases function in dilute formic acid with no added salts or reductants allowing such reactions to be loaded directly on mass spectrometers. Furthermore, the acidic nature of the HTA-Protease reactions inhibits disulfide reformation and obviates the need for addition of IAA (Figs. 4 & 6). Importantly, simplified and abbreviated protocols with no sample manipulation, cleanup, or losses have significant potential to reduce sample-to-sample and lab-to-lab variation and facilitate data sharing via standardized protocols. Such improved

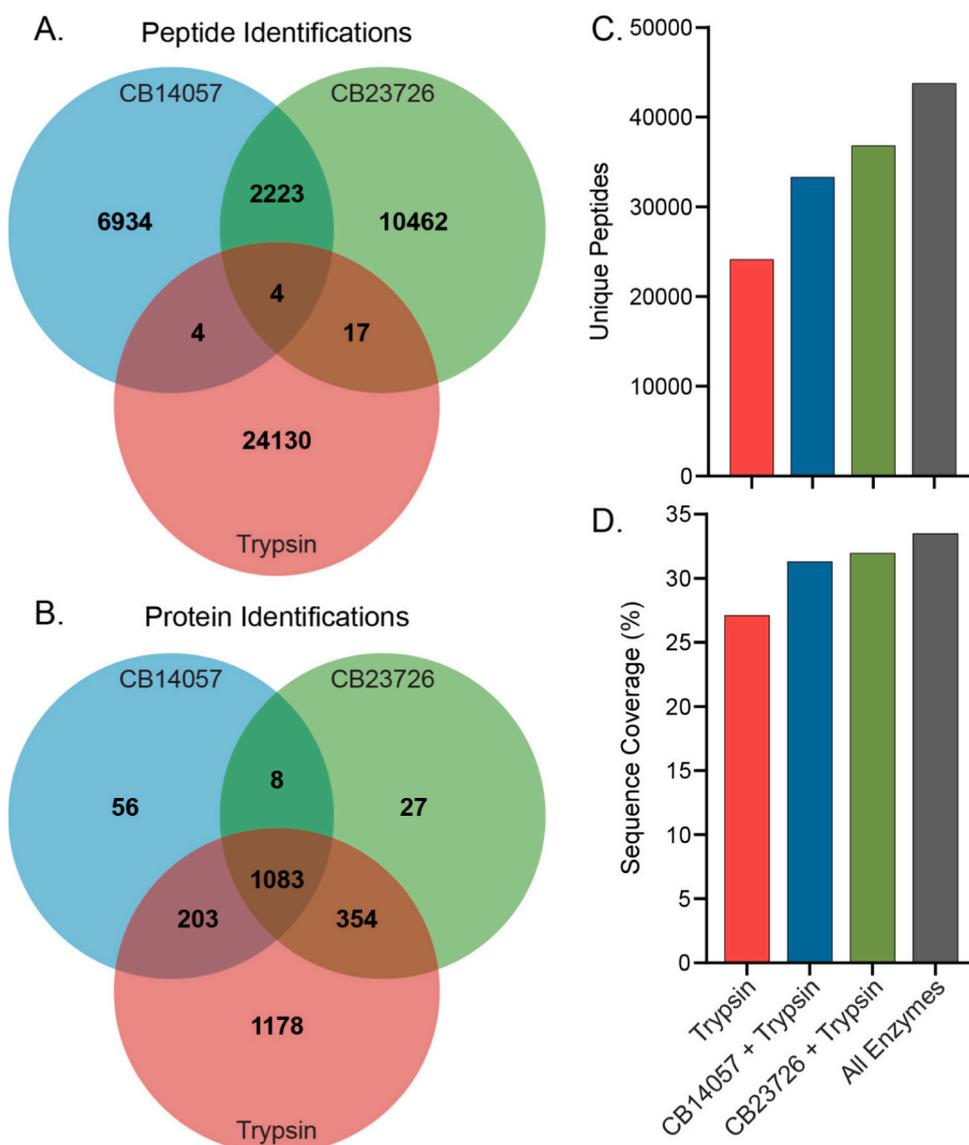


Fig. 9. Effect of HTA-Proteases used orthogonally with trypsin to improve coverage of K562 cell lysate proteome. Unique peptide (A) and protein (B) identifications using CB14057, CB23726, and trypsin alone. Unique peptide IDs (C) and average sequence coverage (D) using trypsin, trypsin and CB14057 combined, trypsin and CB23726 combined, and all enzymes combined. Coefficients of variation (CV) of spectral counts for proteins identified in all samples were 12.8%, 24.2%, and 16.5% for trypsin, CB14057, and CB23726, respectively.

reproducibility with no sample losses can be a major advance in facilitating clinical, diagnostic, and nano- proteomics methods. Likewise, the HTA-Protease protocols lend themselves to automated front-end sample-prep and high-throughput and rapid-turnaround applications such as those employed in clinical laboratories. These and other topics are the subjects of ongoing studies.

Enzyme specificity is a rational concern for proteomics studies and trypsin is the most 'specific' protease commonly used for proteomic applications. However, chymotrypsin and several other less-specific enzymes are routinely used in proteomic analyses. Like all common proteomics enzymes explored to date, HTA-Protease specificity is quantifiably lesser than trypsin but comparable to chymotrypsin and demonstrably useful for proteomic applications (Figs. 7-10). Historically, computational power limited the practicality of complex search algorithms and made limitation of search space through specific enzyme definitions a high priority. Computational advances have made this less of a concern and no-enzyme unbounded cleavage searches are practical and advantageous in some cases. In addition to the unmatched specificity of trypsin, proteomics investigators have found that having basic amino acids at peptide termini could be advantageous. This specificity for basic amino acids has manifested as a nearly exclusive use of positive-mode settings for proteomics experiments. The E, L, F

specificity of these first HTA-Proteases may permit further gains in coverage by combining both positive and negative mode data from the same sample and is also the topic of ongoing studies.

Importantly, the HTA-Protease experiments presented here use instrument settings and methods that were optimized for trypsin. Here we present a set of introductory experiments intended to exhibit the robust nature of HTA-Proteases and establish their basic suitability for proteomics. HTA-Proteases and protocols are very dissimilar to currently used enzymes and protocols and will therefore require continued study and optimization to take full advantage of these entirely new enzymes for proteomics. While the primary objective of this study was to demonstrate the capabilities of HTA-Proteases to improve proteomic identifications, studies are ongoing to further investigate the accuracy of protein quantification provided by these enzymes, as well as to determine potential improvements with more effective reduction at pH 3 using reductants like TCEP and to optimize front end chromatography and instrument settings, among others. Even without HTA-Protease protocol optimization, the coefficients of variation (CV's) observed in this study (trypsin = 12.8%, CB23726 = 16.5%, CB14057 = 24.2%) are sufficient for many studies but encourage method optimization to further reduce variability. Due to the significant divergence from traditional methods, the scope of work presented here is necessarily

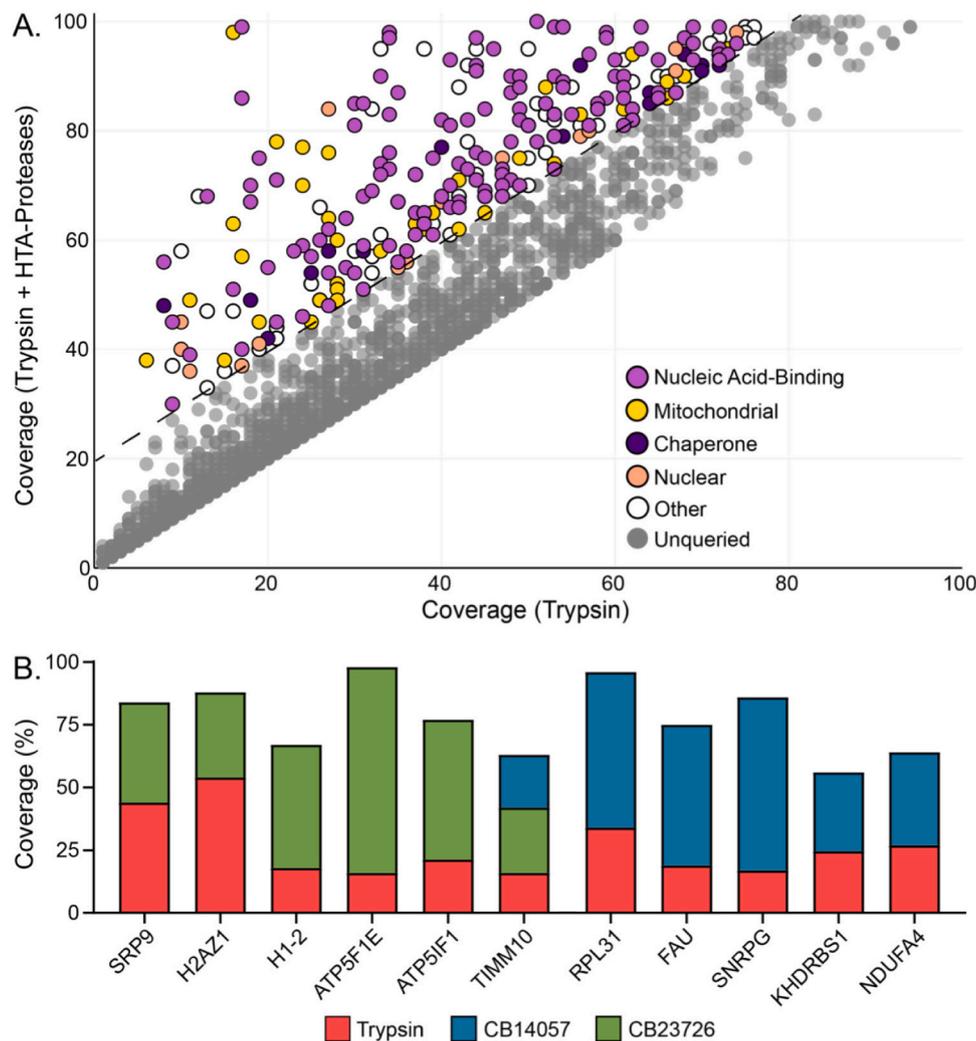


Fig. 10. HTA-Proteases provide improved coverage for specific protein classes when combined with trypsin digest data. (A) Scatter plot of protein coverage with trypsin alone vs. trypsin in combination with HTA-Protease data. Proteins with <20% coverage improvement are plotted in grey, proteins with >20% coverage improvement are colored by class: nucleic acid-binding (magenta), mitochondrial (yellow), chaperone (purple), nuclear (orange), and other (white). The dashed line indicates the 20% improvement boundary. (B) Bar plots showing coverage contributions by trypsin (red), CB14057 (blue), and CB23726 (green) for representative proteins from each listed class. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

narrow and addresses only a subset of the potential advantages of HTA-Proteases for proteomics. It is not unreasonable to expect as HTA-Protease methods mature, further gains will be realized with this new technology as has been the case with tryptic digest optimization over the past 50 years.

Currently, HTA-Proteases have several immediately useful ‘drop-in’ applications for proteomic studies. First, HTA-Proteases can be of immediate use in cases where a region of interest has too few tryptic cleavage sites and requires alternative approaches [39,40]. Notably, one such case has been validated during the drafting of this manuscript where a C-terminal region of murine fibrinogen was not identified after digestion with a variety of proteases, including trypsin, chymotrypsin, Arg—C, and limited pepsin proteolysis. This region was reliably detected in HTA-protease digested samples, advancing the research, and uncovering biologically relevant differences in *in vivo* proteolytic processing between sample groups (**Manuscript in preparation**). Also shown here is that the addition of HTA-Protease datasets to tryptic datasets can significantly increase proteome and protein coverage and produce nearly completely non-overlapping peptide data sets (Fig. 9). In comparing the protein coverage between trypsin alone and trypsin with HTA-Proteases we find a significant increase in coverage with the combined data relative to the tryptic digests alone. Trypsin alone gives an average protein coverage of 27.1% in these studies. The addition of CB14057 and CB23726 digest data increased the average coverage to 31.3% and 32.0%, respectively, while all three enzymes together

increased average coverage to 33.5%. The utility of additional coverage is self-evident in some cases and may permit the identification of specific target protein regions of interest and markers of disease that would not be possible in the absence of the HTA-Protease digests.

The unexpected bias for increased detection and coverage of ribonucleoprotein, DNA-binding, and mitochondrial/membrane proteins may be of the most immediate impact for proteomic research groups (Fig. 10). Upon considering these results, we reasoned that much like the chemical hydrolysis at aspartate discussed above (Fig. 8), nucleic acids are largely if not completely hydrolyzed in heated acid [41]. The preponderance of gains in DNA and RNA binding protein coverage and identifications may be a function of the hydrolysis of associated nucleic acids in HTA-Protease reactions. Alternatively, the loss of nucleic acid components under persistent hot/acid denaturing conditions may be responsible for the observed bias. Likewise, the harsh digestion conditions may also more effectively disrupt mitochondrial membrane complexes and proteins, thereby accounting for the increase in mitochondrial/membrane protein coverage. Irrespective of the precise mechanism(s), HTA-Proteases offer new capabilities for researchers studying the biologically critical functions of nucleoproteins, chromatin, epigenetics, mitochondrial, and membrane proteins.

Of the available alternative proteases used for proteomics [9], the inherent stability, rapid and simple protocols, denaturing reaction conditions, and bias towards biologically critical classes of proteins are all unique to the HTA-Proteases described herein. The availability of

HTA-Proteases and protocols and their unprecedented characteristics promise to deliver new options, capabilities, and approaches to further empower the growing fields of proteomics.

Declaration of Competing Interest

Procurement, inquiries, and information regarding HTA-Proteases can be found at <https://www.cinderbio.com>.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD041226.

Acknowledgements

Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number 5R44GM128540 (SMY) and the University of Colorado Cancer Center Support Grant (P30CA046934). We would like to thank Max Hopkins, Paula Matheus Carnevali, Kirti Hargunani, and Allison Michele Narlock-Brand for sharing their insights.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jprot.2023.104992>.

References

- [1] L. Yi, P.D. Pichowski, T. Shi, R.D. Smith, W.J. Qian, Advances in microscale separations towards nanoproteomics applications, *J. Chromatogr. A* 1523 (2017) 40–48.
- [2] K. Wang, C. Huang, E. Nice, Recent advances in proteomics: towards the human proteome, *Biomed. Chromatogr.* 28 (6) (2014) 848–857.
- [3] Y. Zhou, X. Zhang, L. Fornelli, P.D. Compton, N. Kelleher, M.J. Wirth, Chromatographic efficiency and selectivity in top-down proteomics of histones, *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* 1044–1045 (2017) 47–53.
- [4] M.S. Kim, S.M. Pinto, D. Getnet, R.S. Nirujogi, S.S. Manda, R. Chaerkady, A. K. Madugundu, D.S. Kelkar, R. Isserlin, S. Jain, J.K. Thomas, B. Muthusamy, P. Leal-Rojas, P. Kumar, N.A. Sahasrabudhe, L. Balakrishnan, J. Advani, B. George, S. Renuse, L.D. Selvan, A.H. Patil, V. Nanjappa, A. Radhakrishnan, S. Prasad, T. Subbannayya, R. Raju, M. Kumar, S.K. Sreenivasamurthy, A. Marimuthu, G.J. Sathé, S. Chavan, K.K. Datta, Y. Subbannayya, A. Sahu, S. D. Yelamanchi, S. Jayaram, P. Rajagopalani, J. Sharma, K.R. Murthy, N. Syed, R. Goel, A.A. Khan, S. Ahmad, G. Dey, K. Mudgal, A. Chatterjee, T.C. Huang, J. Zhong, X. Wu, P.G. Shaw, D. Freed, M.S. Zahari, K.K. Mukherjee, S. Shankar, A. Mahadevan, H. Lam, C.J. Mitchell, S.K. Shankar, P. Satishchandra, J. T. Schroeder, R. Sirdeshmukh, A. Maitra, S.D. Leach, C.G. Drake, M.K. Halushka, T. S. Prasad, R.H. Hruban, C.L. Kerr, G.D. Bader, C.A. Iacobuzio-Donahue, H. Gowda, A. Pandey, A draft map of the human proteome, *Nature* 509 (7502) (2014) 575–581.
- [5] M. Wilhelm, J. Schlegl, H. Hahne, A.M. Gholami, M. Lieberenz, M.M. Savitski, E. Ziegler, L. Butzmann, S. Gessulat, H. Marx, T. Mathieson, S. Lemeur, K. Schnatbaum, U. Reimer, H. Wenschuh, M. Mollenhauer, J. Slotta-Huspenina, J. H. Boese, M. Bantscheff, A. Gerstmaier, F. Faerber, B. Kuster, Mass-spectrometry-based draft of the human proteome, *Nature* 509 (7502) (2014) 582–587.
- [6] S. Adhikari, E.C. Nice, E.W. Deutsch, L. Lane, G.S. Omenn, S.R. Pennington, Y. K. Paik, C.M. Overall, F.J. Corrales, I.M. Cristea, J.E. Van Eyk, M. Uhlen, C. Lindskog, D.W. Chan, A. Bairoch, J.C. Waddington, J.L. Justice, J. LaBaer, H. Rodriguez, F. He, M. Kostrzewa, P. Ping, R.L. Gundry, P. Stewart, S. Srivastava, S. Srivastava, F.C.S. Nogueira, G.B. Domont, Y. Vandenbrouck, M.P.Y. Lam, S. Wennersten, J.A. Vizcaino, M. Wilkins, J.M. Schwenk, E. Lundberg, N. Bandeira, G. Marko-Varga, S.T. Weintraub, C. Pineau, U. Kusebauch, R.L. Moritz, S.B. Ahn, M. Palmlad, M.P. Snyder, R. Aebersold, M.S. Baker, A high-stringency blueprint of the human proteome, *Nat. Commun.* 11 (1) (2020) 5301.
- [7] E.J. Dupree, M. Jayathirtha, H. Yorkey, M. Mihasan, B.A. Petre, C.C. Darie, A critical review of bottom-up proteomics: the good, the bad, and the future of this field, *Proteomes* 8 (3) (2020).
- [8] E. Vandermarliere, M. Mueller, L. Martens, Getting intimate with trypsin, the leading protease in proteomics, *Mass Spectrom. Rev.* 32 (6) (2013) 453–465.
- [9] P. Giansanti, L. Tsiatsiani, T.Y. Low, A.J. Heck, Six alternative proteases for mass spectrometry-based proteomics beyond trypsin, *Nat. Protoc.* 11 (5) (2016) 993–1006.
- [10] T. Vajda, A. Garai, Comparison of the effect of calcium(II) and manganese(II) ions on trypsin autolysis, *J. Inorg. Biochem.* 15 (4) (1981) 307–315.
- [11] M. Sebel, T. Stosova, J. Havlis, N. Wielsch, H. Thomas, Z. Zdrahal, A. Shevchenko, Thermostable trypsin conjugates for high-throughput proteomics: synthesis and performance evaluation, *Proteomics* 6 (10) (2006) 2959–2963.
- [12] T.D. Brock, K.M. Brock, R.T. Belly, R.L. Weiss, *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature, *Arch. Mikrobiol.* 84 (1) (1972) 54–68.
- [13] U.M. Marquez, F.M. Lajolo, Composition and digestibility of albumin, globulins, and glutelins from *Phaseolus vulgaris*, *J. Agric. Food Chem.* 29 (5) (1981) 1068–1074.
- [14] Z.Y. Park, D.H. Russell, Thermal denaturation: a useful technique in peptide mass mapping, *Anal. Chem.* 72 (11) (2000) 2667–2670.
- [15] S.J. Bark, N. Muster, J.R. Yates 3rd, G. Siuzdak, High-temperature protein mass mapping using a thermophilic protease, *J. Am. Chem. Soc.* 123 (8) (2001) 1774–1775.
- [16] J.H. Northrop, Comparative hydrolysis of gelatin by pepsin, trypsin, acid, and alkali, *J. Gen Physiol* 4 (1) (1921) 57–71.
- [17] A.B. Steven, M. Yannone, Nucleic acids useful for integrating into and gene expression in hyperthermophilic acidophilic archaea, in: USPTO (Ed.) University of California, WO EP US BR CA, 2012.
- [18] X. Lin, J. Tang, Thermopsin, *Methods Enzymol.* 248 (1995) 156–168.
- [19] S.M. Yannone, A. Barnebey, in: USPTO (Ed.) (Ed.), Nucleic Acids Useful for Integrating into and Gene Expression in Hyperthermophilic Acidophilic archaea, University of California US, 2012.
- [20] X. Lin, J. Tang, Purification, characterization, and gene cloning of thermopsin, a thermostable acid protease from *Sulfolobus acidocaldarius*, *J. Biol. Chem.* 265 (3) (1990) 1490–1495.
- [21] C. Cupp-Enyard, Sigma's non-specific protease activity assay - casein as a substrate, *J. Vis. Exp.* 19 (2008).
- [22] M.L. Anson, The estimation of pepsin, trypsin, papain, and cathepsin with hemoglobin, *J. Gen Physiol* 22 (1) (1938) 79–89.
- [23] D. Mellacheruvu, Z. Wright, A.L. Couzens, J.P. Lambert, N.A. St-Denis, T. Li, Y. V. Miteva, S. Hauri, M.E. Sardi, T.Y. Low, V.A. Halim, R.D. Bagshaw, N.C. Hubner, A. Al-Hakim, A. Bouchard, D. Faubert, D. Fermin, W.H. Dunham, M. Goudreau, Z. Y. Lin, B.G. Badillo, T. Pawson, D. Durocher, B. Coulombe, R. Aebersold, G. Superti-Furga, J. Colinge, A.J. Heck, H. Choi, M. Gstaiger, S. Mohammed, I. M. Cristea, K.L. Bennett, M.P. Washburn, B. Raught, R.M. Ewing, A.C. Gingras, A. I. Nesvizhskii, The CRAPome: a contaminant repository for affinity purification-mass spectrometry data, *Nat. Methods* 10 (8) (2013) 730–736.
- [24] N. Colaert, K. Helsen, L. Martens, J. Vandekerckhove, K. Gevaert, Improved visualization of protein consensus sequences by iceLogo, *Nat. Methods* 6 (11) (2009) 786–787.
- [25] S.M. Yannone, J.O. Fuss, A. Barnebey, in: USPTO (Ed.) (Ed.), Methods for rapidly digesting biopolymers with ultrastable enzymes for mass spectrometry-based analyses, Cinder Biological, Inc., US, 2017.
- [26] J.R. Winther, C. Thorpe, Quantification of thiols and disulfides, *Biochim. Biophys. Acta* 1840 (2) (2014) 838–846.
- [27] J.E. Fuchs, S. von Grafenstein, R.G. Huber, M.A. Margreiter, G.M. Spitzer, H. G. Wallnoefer, K.R. Liedl, Cleavage entropy as quantitative measure of protease specificity, *PLoS Comput. Biol.* 9 (4) (2013), e1003007.
- [28] R.W. Zumwalt, J.S. Absheer, F.E. Kaiser, C.W. Gehrke, Acid hydrolysis of proteins for chromatographic analysis of amino acids, *J. Assoc. Off. Anal. Chem.* 70 (1) (1987) 147–151.
- [29] A.J. Darragh, P.J. Moughan, The effect of hydrolysis time on amino acid analysis, *J. AOAC Int.* 88 (3) (2005) 888–893.
- [30] P.D. Gershon, Cleaved and missed sites for trypsin, lys-C, and lys-N can be predicted with high confidence on the basis of sequence context, *J. Proteome Res.* 13 (2) (2014) 702–709.
- [31] A. Schniers, Y. Pasing, T. Hansen, Lys-C/trypsin tandem-digestion protocol for gel-free proteomic analysis of colon biopsies, *Methods Mol. Biol.* 2019 (1959) 113–122.
- [32] M. Kobayashi, S. Oshima, C. Maeyashiki, Y. Nibe, K. Otsubo, Y. Matsuzawa, Y. Nemoto, T. Nagaishi, R. Okamoto, K. Tsuchiya, T. Nakamura, M. Watanabe, The ubiquitin hybrid gene UBA52 regulates ubiquitination of ribosome and sustains embryonic development, *Sci. Rep.* 6 (2016) 36780.
- [33] W. Kühne, Verhandlungen des Naturhistorisch-medizinischen Vereins zu Heidelberg, Heidelberg, 1857, pp. 194–198.
- [34] D. Fenyo, R.C. Beavis, The GPMDB REST interface, *Bioinformatics* 31 (12) (2015) 2056–2058.
- [35] L. Switzar, M. Giera, W.M. Niessen, Protein digestion: an overview of the available techniques and recent developments, *J. Proteome Res.* 12 (3) (2013) 1067–1077.
- [36] L. Tsiatsiani, A.J. Heck, Proteomics beyond trypsin, *FEBS J.* 282 (14) (2015) 2612–2626.
- [37] Z. Ning, D. Seebun, B. Hawley, C.K. Chiang, D. Figeys, From cells to peptides: “one-stop” integrated proteomic processing using amphipols, *J. Proteome Res.* 12 (3) (2013) 1512–1519.
- [38] N. Slavov, Increasing proteomics throughput, *Nat. Biotechnol.* 39 (7) (2021) 809–810.
- [39] M. Vermachova, Z. Purkrtova, J. Santrucek, P. Jolivet, T. Chardot, M. Kodicek, Combining chymotrypsin/trypsin digestion to identify hydrophobic proteins from oil bodies, *Methods Mol. Biol.* 1072 (2014) 185–198.
- [40] J. Salplachta, M. Marchetti, J. Chmelik, G. Allmaier, A new approach in proteomics of wheat gluten: combining chymotrypsin cleavage and matrix-assisted laser desorption/ionization quadrupole ion trap reflectron tandem mass spectrometry, *Rapid Commun. Mass Spectrom.* 19 (18) (2005) 2725–2728.
- [41] J.D. Smith, R. Markham, Chromatographic studies on nucleic acids; the quantitative analysis of ribonucleic acids, *Biochem. J.* 46 (5) (1950) 509–513.