Evolution of a mass spectrometry-grade protease with PTM-directed specificity

Duc T. Tran[a], Valerie J. Cavett[a], Vuong Q. Dang[a], Héctor L. Torres[a], and Brian M. Paegel[a,1]

[a]Department of Chemistry, The Scripps Research Institute, Jupiter, FL 33458

Edited by David Baker, University of Washington, Seattle, WA, and approved November 8, 2016 (received for review July 7, 2016)

Mapping posttranslational modifications (PTMs), which diversely modulate biological functions, represents a significant analytical challenge. The centerpiece technology for PTM site identification, mass spectrometry (MS), requires proteolytic cleavage in the vicinity of a PTM to yield peptides for sequencing. This requirement catalyzed our efforts to evolve MS-grade mutant PTM-directed proteases. Citrulline, a PTM implicated in epigenetic and immunological function, made an ideal first target, because citrullination eliminates arginyl tryptic sites. Bead-displayed trypsin mutant genes were translated in droplets, the mutant proteases were challenged to cleave bead-bound fluorogenic probes of citrulline-dependent proteolysis, and the resultant beads (1.3 million) were screened. The most promising mutant efficiently catalyzed citrulline-dependent peptide bond cleavage (k_{cat}/K_M = 6.9 × 10^5 $M^{-1}·s^{-1}$). The resulting C-terminally citrullinated peptides generated characteristic isotopic patterns in MALDI-TOF MS, and both a fragmentation product y_1 ion corresponding to citrulline (176.1030 *m/z*) and diagnostic peak pairs in the extracted ion chromatograms of LC-MS/MS analysis. Using these signatures, we identified citrullination sites in protein arginine deiminase 4 (12 sites) and in fibrinogen (25 sites, two previously unknown). The unique mass spectral features of PTM-dependent proteolytic digest products promise a generalized PTM site-mapping strategy based on a toolbox of such mutant proteases, which are now accessible by laboratory evolution.

posttranslational modification | proteomics | directed evolution | protease | in vitro compartmentalization

The identification and mapping of protein posttranslational modifications (PTMs), which underpin the regulation of virtually every cellular function, represents one of the most significant challenges of modern-day analytical science. Accessing this biochemical information requires direct observation of the modified site, often in low abundance within the complex cellular milieu. The ability to address this sample complexity is precisely why tandem mass spectrometry (MS/MS) now dominates proteomic analysis. LC-MS/MS instruments can analyze mixtures of thousands of peptides, acquiring the mass-to-charge ratio (*m/z*) of each peptide precursor ion (MS1) and fragmenting the most abundant precursors for a second dimension of MS analysis (MS2). Database searching of MS/MS "bottom-up" proteomic data entails matching observed precursor and fragmentation product ion masses with those of a theoretical protein digest, returning the amino acid sequence of each ion (including any PTMs) and statistical confidence of the sequence match (1, 2).

Bottom-up MS/MS workflows generate peptides for analysis almost ubiquitously by digestion using trypsin, which cleaves the C-terminal peptide bonds of Arg and Lys with high catalytic efficiency and selectivity (3). Regions of a protein too densely or too sparsely populated with tryptic sites result in sequencing coverage gaps, which can conceal PTM sites during mapping experiments. Digestion with alternative proteases (4, 5), which reinforce tryptic cleavage efficiency or introduce cleavages at nontryptic sites, enhance coverage by increasing the probability that more regions of the protein will be represented in peptides appropriate for MS analysis. However, only a handful of naturally occurring enzymes exists, and they suffer myriad drawbacks, such as extreme cleavage site promiscuity or low catalytic efficiency (6).

The relationship between protein cleavage and MS sequence coverage has spurred the exploration of methods to induce alternate cleavages, especially at PTM sites. Examples include chemoselective pSer side chain modification to establish pSer-dependent tryptic cleavage sites (7, 8), and directed evolution of novel proteolytic cleavage activity, yielding a pTyr-dependent subtilisin mutant (9) and a suite of OmpT mutants that cleaved the Ala–Arg (10), sTyr–Arg (11), or nTyr–Arg (12) P1–P1′ peptide bonds; P1–P1′ cleavage junction requirements precluded further implementation. Proteolytic specificity has also been successfully evolved in other enzymes (13, 14), but the MS workhorse enzyme, trypsin, has remained unexplored. Although the trypsin family is the textbook example of divergent evolution (the highly homologous family members possess diverse cleavage specificities), its network of specificity-determining residues—a constellation of approximately four point mutations (15–17) and two 4-residue surface-exposed loops (18)—is daunting. Evolution would require large libraries and either a high-throughput mutant screening or selection approach (19) to discover new cleavage specificity. We sought to develop such an approach in the light of trypsin's compelling phylogeny and the promise of PTM-directed proteases for proteomics, and pursued the evolution of proteolytic specificity targeting the PTM citrulline.

Citrullination, the product of protein arginine deiminase (PAD)-mediated deimination of key Arg side chains, is important in an array of epigenetic (20) and immunological functions (21), but detecting citrullination by MS is problematic. Citrullination eliminates tryptic sites, resulting in larger peptides that are harder to detect and sequence. Further, deamidation of Gln or Asn produces mass shifts identical to citrullination. Several chemical proteomic strategies aid in the identification of citrullinated peptides (22–25), but citrulline site identifications particularly (and PTM site identifications generally) require observation of an ion series containing

Significance

Protein posttranslational modifications (PTMs) regulate virtually every aspect of cellular function. Mapping PTMs is a key application of protein mass spectrometry (MS), which analyzes thousands of peptides generated in enzymatic protein digests. There are currently only about seven naturally occurring proteases for generating these digests, and with the exception of trypsin, most are poor or promiscuous enzymes that are not ideal for MS. This protease "toolbox" has remained essentially unchanged for decades. In this study, the directed evolution of trypsin's proteolytic specificity resulted in the discovery of a mutant protease that efficiently cleaves at citrullinated arginines, affording a powerful PTM site identification tool and promising a Darwinian wellspring of new, custom proteases.

Author contributions: D.T.T., V.J.C., and B.M.P. designed research; D.T.T., V.J.C., V.Q.D., H.L.T., and B.M.P. performed research; D.T.T., V.J.C., V.Q.D., H.L.T., and B.M.P. analyzed data; and D.T.T., V.J.C., and B.M.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: briandna@gmail.com.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1609925113/-/DCSupplemental.

the PTM-added mass in the MS2 spectrum. For example, if the fourth residue of a 10-residue peptide is modified, all fragmentation ions should reflect the PTM's added mass in the *b* series b_4–b_9 and *y* series y_7–y_9 (the b_i ion is the N-terminal product of fragmentation at the i^{th} amide bond counted from the N terminus; the y_i ion is the C-terminal product of fragmentation at the i^{th} amide bond counted from the C terminus) (26). A citrulline-dependent (or, generally, PTM-dependent) protease, however, would generate peptides with a C-terminally citrullinated (modified) residue, and MS2 spectra of all such citrulline-terminated peptides should feature a y_1 ion corresponding to the mass of citrulline.

Results

Compartmentalized Evolution of a Citrulline-Dependent Protease. We devised a compartmentalized microbead display (27, 28) strategy for high-throughput mutant protease expression and activity-based screening. Activity assay "probe–primer" beads (Fig. 1, *Upper*) display an oligonucleotide PCR primer and a bisamide rhodamine 110 (R110) protease activity assay probe (Figs. S1–S3) (29). Probe–primer beads and a trypsinogen mutant gene library were used in emulsion PCR (emPCR) to generate beads clonally populated with ~10,000 copies of a gene library member (probe–gene beads) (30) followed by emulsion in vitro transcription/translation (emIVTT) for highly parallel, low-volume protein expression and activity assay. FACS analysis isolated single beads encoding active proteases that dequenched bead-bound R110 probes (Fig. 1, *Lower*).

Control tryptic probe–primer beads displaying $(acGPR)_2R110$ that were digested with MS-grade or translated and enterokinase (EK)-activated trypsin and analyzed by flow cytometry exhibited good separation from untreated and unfunctionalized beads (Fig. 2*A*). With feasibility established, citrulline-dependent activity assay probe–primer beads displaying $(acIPcit)_2R110$ were elaborated with an EGGK(185–188)XXXX/D189S trypsinogen mutant library. This library explored the active site-proximal specificity-determining surface-exposed loop 1 (18) in the context of a fixed D189S (31), another key residue in the base of the active site that we also found to confer basal citrulline-dependent cleavage activity (Fig. S4). emIVTT and FACS screening of 1.3×10^6 $(acIPcit)_2R110$ probe-gene library beads yielded 53 events (Fig. 2*B*), 21 successful sorts to individual quantitative PCRs (qPCRs), and 15 clean amplifications. One sample exhibited enhanced citrulline-dependent proteolysis in a "real-time" IVTT activity assay (Fig. 2*C*) (32). The screening lead contained the citrulline-active loop 1 mutant EGGK(185–188)KGGA and one additional serendipitous mutation, L7P, in the posttranslationally cleaved N-terminal propeptide. The L7P mutation conferred no detectable citrulline-dependent cleavage activity to trypsin by real-time IVTT assay, although L7P increased real-time IVTT reaction progress of trypsin using a tryptic substrate. Screening of either loop 2 QKNK(221–224)XXXX or rescreening loop 1 with fixed L7P did not yield more compelling leads, though L7P modestly increased the hit rate (Fig. S4). The confirmed mutant L7P/EGGK(185–188)KGGA/D189S was declared trypsin^{+cit} to reflect its trypsin ancestry and citrulline-dependent activity.

Expression and Characterization of Trypsin^{+cit}. Overexpression of trypsin^{+cit} in *Escherichia coli* BL21 was readily observed in the insoluble fraction by SDS/PAGE analysis. Solubilized, purified, and refolded trypsin^{+cit} was observed as the singly charged molecular ion $[M + H]^+ = 26{,}888$ and doubly charged molecular ion $[M + 2H]^{2+} = 13{,}432$ in MALDI-TOF MS analysis before EK activation (full-length zymogen), and as $[M + H]^+ = 23{,}337$ and $[M + 2H]^{2+} = 11{,}658$ after activation (active enzyme). Overexpression of wild-type trypsin was unsuccessful under otherwise identical conditions. Installation of L7P alone onto trypsin was sufficient to generate overexpression product by SDS/PAGE, and the P7L trypsin^{+cit} revertant abolished overexpression product (Fig. S4).

Digestion of fluorogenic peptide probes (Fig. 3), black hole quencher (BHQ)-IPcitAA-fluorescein (FAM) and BHQ-IPRAA-tetramethylrhodamine (TMR) (Fig. S5), with trypsin, trypsin^{+cit}, or an activator control (EK) and fluorescence monitoring confirmed specific enzyme activities on more conventional substrates: after

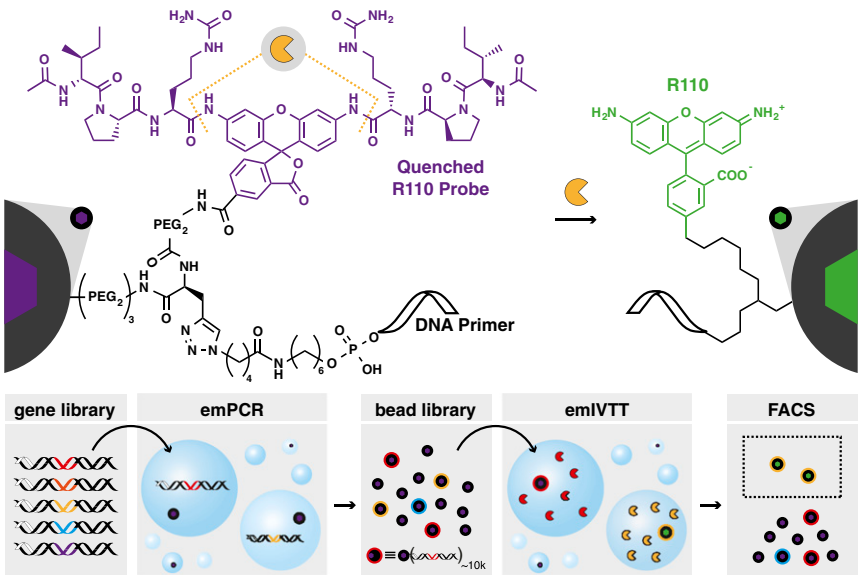


Fig. 1. Protease activity assay probe–primer bead and compartmentalized protease evolution. Quenched activity assay probe–primer beads (purple hexagon) display a PCR primer (white) and a quenched bisamide R110 probe (purple). Cleavage at the P1 amide bond (yellow dashed line) catalyzed by an active protease (yellow Pac-man) dequenches the R110 fluorophore to yield a fluorescent bead (green hexagon). emPCR of a limiting dilution of protease mutant gene library in the presence of the protease activity assay probe–primer beads yields a probe–gene bead library. Each probe–gene bead clonally displays ~10,000 copies of a protease mutant library member. The probe–gene bead library is washed and reemulsified with in vitro transcription/translation reaction mixture (emIVTT). Droplets with beads that display an inactive protease-encoding gene (red) contain inactive protease (red Pac-man), which does not dequench the R110 probe. Droplets with beads that display an active protease-encoding gene (yellow) contain active protease, which cleaves the bead-bound quenched R110 probes to generate fluorescent R110-labeled beads. Fluorescent beads are collected by FACS for single-bead PCR and activity assay.

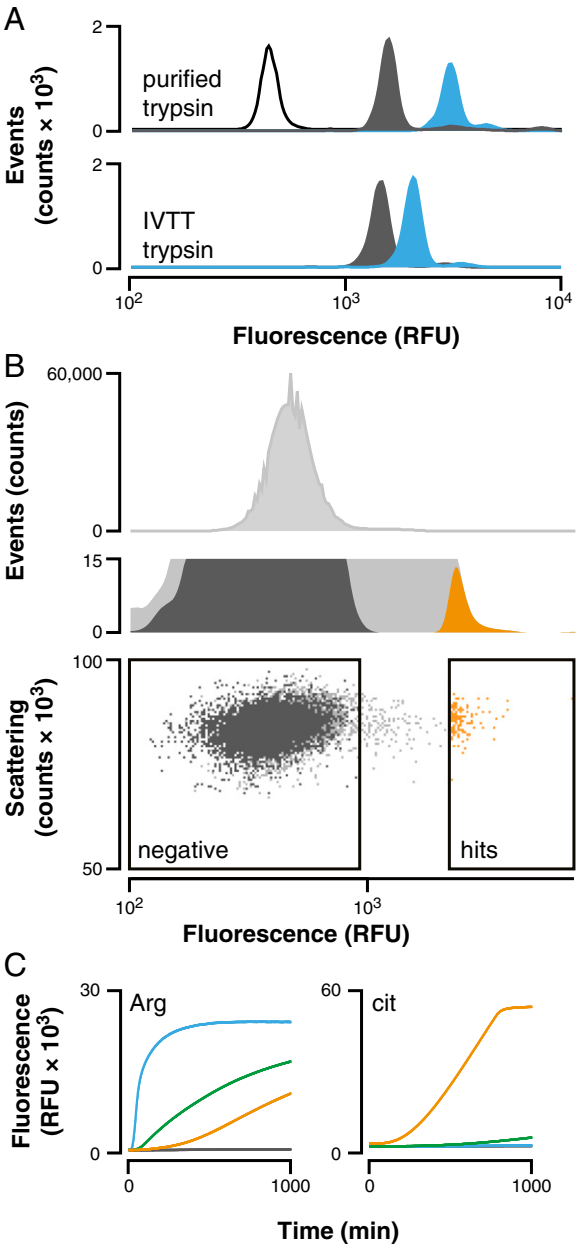


Fig. 2. Identification and analysis of probe–gene beads encoding active proteases. (*A*) Control tryptic activity assay probe–primer beads displaying $(acGPR)_2R110$ were incubated with sequencing grade purified trypsin and analyzed by flow cytometry (*Upper*). Blank beads (white) were significantly less fluorescent than either undigested tryptic probe–primer beads (gray, 1,500 ± 400) or trypsin-digested beads (cyan, 3,100 ± 600). Incubation of the tryptic probe–primer beads displaying trypsinogen-encoding DNA templates and EK in IVTT mixture (*Lower*) resulted in increased bead fluorescence of IVTT trypsin-digested beads (cyan, ~40 solution-phase trypsinogen template molecules/bead yielding ~3,000 protease molecules/bead, signal shifted to 2,100 ± 500) relative to probe–primer beads exposed to IVTT alone (gray). (*B*) emIVTT and FACS analysis of a trypsinogen EGGK(185–188)XXXX/D189S mutant gene library amplified on $(aclPcit)_2R110$ probe–primer beads yielded a hit pool of 53 beads (orange) that exhibited ~eightfold enhanced fluorescence over negative control beads (dark gray), cleanly shifted from the uncollected 1.3×10^6 events (light gray). (*C*) Individual hit beads from FACS were amplified and the PCR products analyzed in real-time IVTT assays containing either tryptic probe $(cbzIPR)_2R110$ (Arg) or citrulline-specific screening probe $(aclPcit)_2R110$ (cit). Activity traces include the wild-type trypsinogen gene (cyan), gene library (green), most promising screening hit (orange), and negative control (gray).

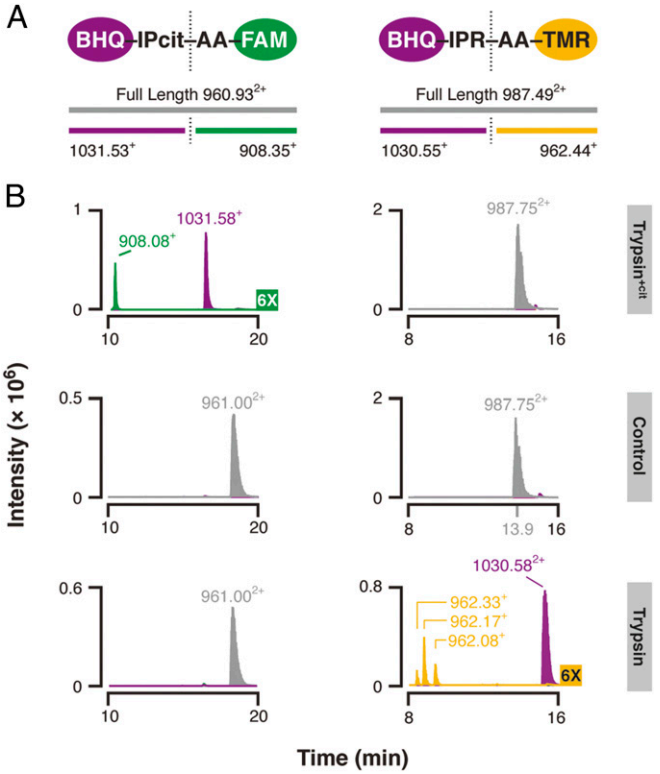


Fig. 3. Fluorogenic peptide digestion and LC-MS analysis. (*A*) Both fluorogenic substrates contained a BHQ quencher (purple). The cit-containing probe was C-terminally labeled with FAM (green) and the Arg-containing probe was C-terminally labeled with TMR (yellow). The digest schematics display the doubly charged full-length peptide ion and singly charged enzymatic digest product theoretical exact masses. (*B*) Digests of both peptides were performed using trypsin^{+cit} (*Upper*), EK control (*Middle*), or trypsin (*Lower*). XIC retention times differ for the full-length (gray), BHQ-labeled N-terminal fragment (purple), FAM-labeled citrulline-dependent C-terminal cleavage product (green), and TMR-labeled tryptic C-terminal cleavage products (yellow). Observed *m/z* and ion charge state appear above each XIC peak. The intensity of each C-terminal digest product XIC is magnified 6×.

25 min incubation, only the trypsin^{+cit} digest of BHQ-IPcitAA-FAM produced FAM signal; only the tryptic digest of BHQ-IPRAA-TMR produced TMR signal; and the EK control produced no signal. LC-MS/MS analysis confirmed each enzyme's site-specific cleavage activity. Only the trypsin^{+cit} digest of BHQ-IPcitAA-FAM contained extracted ion chromatogram (XIC) features corresponding to new N- and C-terminal products of cleavage at citrulline ($1{,}031.58^+$ and 908.08^+ *m/z*, respectively). Tryptic and control digest XIC only contained the full-length probe (961.11^{2+} *m/z*). Trypsin^{+cit} and EK control digests of the BHQ-IPRAA-TMR probe yielded only full-length probe (987.75^{2+} *m/z*) in the respective XICs, whereas the tryptic digest XIC contained new N- and C-terminal products of cleavage at Arg ($1{,}030.58^+$ and 962.33^+ *m/z*, respectively); the C-terminal fragment exists as three regioisomeric TMR coupling products, which are chromatographically distinct but otherwise indistinguishable by MS (Fig. S6). Trypsin^{+cit} exhibited greatly enhanced citrulline-dependent activity in steady-state kinetic analyses. The catalytic efficiency (k_{cat}/K_M) of BHQ-IPcitAA-FAM cleavage by trypsin^{+cit} was 6.9×10^5 $M^{-1}{\cdot}s^{-1}$, and that of BHQ-IPRAA-TMR cleavage was 6.3×10^4 $M^{-1}{\cdot}s^{-1}$, ~12-fold substrate selectivity. The catalytic efficiency of BHQ-IPRAA-TMR cleavage by trypsin was 3.9×10^7 $M^{-1}{\cdot}s^{-1}$. Tryptic cleavage of BHQ-IPcitAA-FAM was undetectable (Fig. S7).

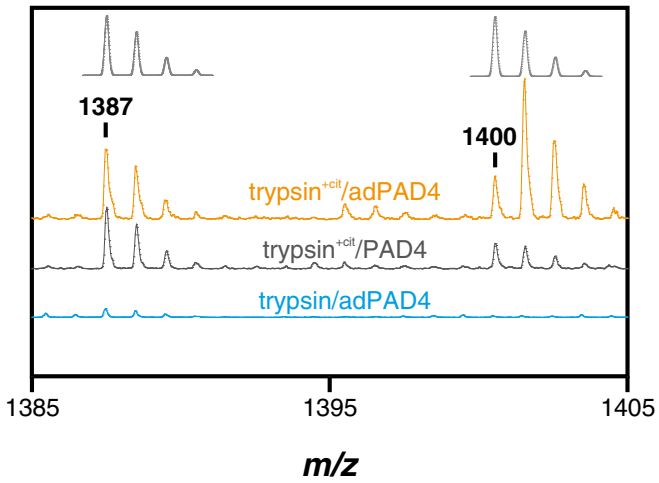
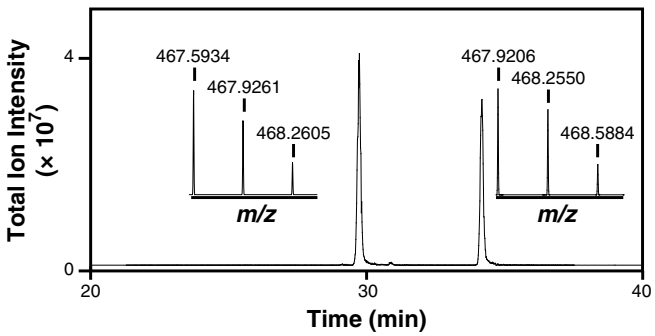


Fig. 4. MALDI-TOF MS analysis of PAD4. The MALDI-TOF mass spectrum (magnified) of the adPAD4 tryptic digest (cyan), untreated PAD4 trypsin^{+cit} digest (black), and adPAD4 trypsin^{+cit} digest (orange) displays two example novel features. One peptide is citrullinated (1,400 *m/z*) and the other is not (1,387 *m/z*) based on deviations from an expected theoretical isotopic envelope (top, light gray). The adPAD4 tryptic digest spectrum contains no features in this *m/z* range.

Trypsin^{+cit} Digestion Introduces Diagnostic Mass Signatures of Citrullination for PTM Site Mapping. Trypsin^{+cit} digestion of protein arginine deiminase 4 (PAD4) in its native and autodeiminated state (adPAD4, Ca^{2+}-treated PAD4) (33) produced notably different MALDI-TOF MS spectra, which were also dramatically different from feature-poor spectra of tryptic digests. Qualitatively, the tryptic digest spectra of PAD4 and adPAD4 are nearly identical (Fig. S8). However, trypsin^{+cit} digestion of adPAD4 produced numerous new features (428 vs. 173 with S/N > 3.5, intensity > 5% base peak). A magnified view displays two example features unique to the PAD4 and adPAD4 trypsin^{+cit} digest (Fig. 4). The 1,387 *m/z* ion (534-TLREHNSFVER-544) was prevalent in both PAD4 and adPAD4 trypsin^{+cit} digests, yet undetectable in the tryptic digest. The peptide's sequence yielded its molecular formula and thereby the theoretical abundance of each isotopic peak based on the binomial distribution and natural abundances of each element. The theoretical and observed isotopic envelopes agreed for both PAD4 and adPAD4 versions of the 1,387 *m/z* ion. The isotopic envelope of the 1,400 *m/z* ion (363-TLPVVFDSPRNR-374), another new trypsin^{+cit} feature, agreed with the expected theoretical envelope. The same feature in the adPAD4 spectrum contained a low-intensity molecular ion and high-intensity first isotope, inconsistent with the expected theoretical envelope for a peptide of this mass where the highest intensity peak should be the molecular ion; instead, it agreed more closely with the overlapping envelopes of two peptides whose molecular ions are shifted by ~1 Da, suggesting the presence of an unmodified and citrullinated (Δm = 0.984 Da) form of the same peptide.

LC-MS/MS investigation of the adPAD4 trypsin^{+cit} digest confirmed citrullination of the 1,400 *m/z* peptide and revealed other diagnostic signatures of citrullination. Database matching (34) returned 53 unique peptides from the tryptic adPAD4 digest (24 peptides contained C-terminal Arg) and 91 unique peptides from the trypsin^{+cit} digest (34 peptides contained C-terminal Arg). The 1387 *m/z* ion (534-TLREHNSFVER-544) appeared in both tryptic and trypsin^{+cit} digest analyses, the XIC contained only one chromatographic peak (463.5687^{3+} *m/z*), and the MS1 isotopic envelope agreed with the unmodified mass. In contrast, the XIC of the 1,400 *m/z* ion (Fig. 5, *Upper*) contained two chromatographic peaks. The two MS1 ions (467.5934^{3+} and 467.9206^{3+} *m/z*) differed by 0.9816 Da, in close agreement with the mass shift of citrullination. Another peptide ion unique to the adPAD4 trypsin^{+cit} digest (211-VRVFQATR-218) also yielded two distinct XIC peaks (488.7870^{2+} and 489.2799^{2+}), the second peak's MS1 again reflecting the mass shift of citrullination. The MS2 spectrum of this peptide contained 176.1030 *m/z*, the y_1 citrulline fragment, y_1(cit), and its ammonia neutral loss, 159.0764 *m/z*. A search for all MS1 precursors generating the 176.1030/159.0765 MS2 citrullination signatures yielded 66 unique MS1. Of these 66, 28 produced two-peak XIC, the earlier eluting peak corresponding to the unmodified peptide and the later peak corresponding to the C-terminal citrullinated peptide. The remaining 38 peptides contained no arginine pair peaks, generated uninterpretable XIC, or yielded no sequence data. Of the 28 citrullinated species, 22 were unique peptides corresponding to 12 sites of modification: 205, 212, 218, 372, 374, 383, 394, 427, 441, 488, 495, and 639 (Fig. 5, *Lower*). The XIC for R639cit contained no detectable peak corresponding to the unmodified peptide.



Arg	Peptide	XIC	MS1^{R}	MS1^{cit}
205	TNHTLVLHVAR		420.9104	421.2383
212 & 218	VRVFQATR		489.2805	489.7720
218	VRVFQATR		488.7870	489.2791
372	TLPVVFDSPR		565.8138	566.3055
374	TLPVVFDSPRNR		467.5934	467.9206
383	EFPIKR		395.2349	395.7259
394	VM$_{Ox}$GPDFGYVTR		629.2996	629.7914
427	GKEYPLGR		460.2537	460.7458
441	ILFGDSCYPSNDSR		815.8622	816.3535
488 & 495	RLLLASPR		463.7924	464.2842
495	LLLASPR		385.2497	385.7424
639	INDFFTYHIR		N.D.	663.8274

0 60
Time (min)

Fig. 5. Citrulline-dependent XIC and PTM site identification in adPAD4. All MS2 spectra containing the 176.1030 *m/z* fragmentation product ion y_1(cit) were identified. XIC analysis of each precursor MS1 ion and its corresponding unmodified ion yielded two distinct peaks, each having the expected theoretical isotopic envelope (*Upper*, example peptide 363-TLPVVFDSPRNR-374). The trypsin^{+cit} digest analysis of adPAD4 identified 12 citrullination sites. The Arg column enumerates sites based on translation start. An orange R indicates the modification site in the peptide sequence. Unmodified (cyan) and citrullinated (orange) MS1 peptide masses agree within 2 ppm of the predicted exact mass.

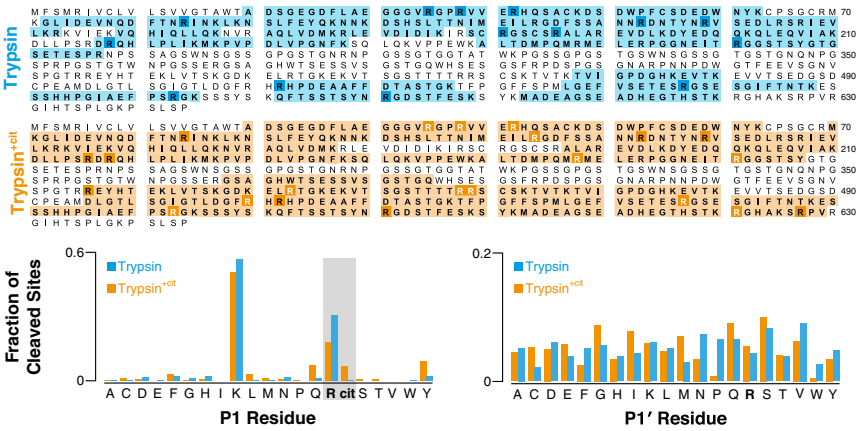


Fig. 6. Comparison of coverage, Arg deimination site mapping, and cleavage site census. cFGA sequence coverage is sparser in the tryptic (light cyan, 52%) compared with the trypsin^{+cit} (light orange, 68%) digest. Deiminated Arg sites in the tryptic digest (dark cyan R) are the result of database matching only. Deiminated Arg sites in the trypsin^{+cit} digest are the result of either y_1(cit) ion detection and database matching (dark orange, white R) or database matching alone (dark orange, black R). The fraction of identified cleavage junctions in the tryptic (cyan) and trypsin^{+cit} (orange) digests of cFGA, cFGB, and cFGG is plotted by P1 residue. Trypsin^{+cit} cleavages at cit, Gln, and Tyr are each >15% total cleavages. The majority of trypsin^{+cit} cleavages derive from its ancestral Arg- and Lys-directed activity. A similar analysis of cleavage junction P1′ residues revealed little preference for either enzyme.

Trypsin^{+cit} Digestion Identifies Citrullination in PAD4-Treated Fibrinogen and Increases Sequence Coverage. Trypsin^{+cit} digestion of native fibrinogen alpha, beta, and gamma chains (FGA, FGB, FGG) and PAD4-citrullinated fibrinogen (cFGA, cFGB, cFGG) produced analogous citrullination-dependent mass spectral signatures. The novel trypsin^{+cit} cleavages increased sequence coverage compared with the tryptic digest for the heavily modified cFGA chain (68% vs. 52% in the tryptic digest), as well as for cFGB and cFGG chains (88% vs. 77% and 78% vs. 75%, respectively; Fig. S9). Mascot (34) and X!Tandem (35) database searches of the two LC-MS/MS analyses (trypsin or trypsin^{+cit} digestion of cFG) generated two lists of peptides for filtering via direct assignment of Arg modification, y_1(cit) detection, or both (Fig. 6, *Upper*). The MS2 fragmentation product ion y_1(cit) is a unique feature of all C-terminally citrullinated peptides, though not all sites of citrullination were assigned as a result of a C-terminal citrulline; some sites were internal in the trypsin^{+cit} sample. All citrullinated sites were observed internal in the tryptic digest. The larger sequence space of cFG also allowed an analysis of trypsin^{+cit} cleavage junction P1 and P1′ residues. In addition to citrulline, Arg and Lys were the most common P1 residues. Cleavage products also frequently contained C-terminal Tyr and Gln (Fig. 6, *Lower*). The P1′ residue profile of trypsin^{+cit} and trypsin are fairly similar. P1′ Pro and Trp appeared to be particularly disfavored.

XIC analysis of the cFG trypsin^{+cit} digest peptides (modified and unmodified forms) that appeared in both database searches (Mascot and X!Tandem) as citrullinated usually resulted in only a single chromatographic peak corresponding to the modified form (Fig. S10). XIC analysis of FG trypsin^{+cit} digest peptides yielded peaks corresponding to the unmodified form of the peptide for most sites. The cFG trypsin^{+cit} digest yielded 49 unique peptides that cross-validated between two database searches as containing a deiminated Arg. Among the 18 FGA sites, 12 contained y1(cit) and 1 was previously unreported (R621). Among the 6 FGB sites, 1 contained y1(cit) and 1 was previously unreported (R376). For the 1 FGG site, y1(cit) was not detected. Previously identified citrullination sites not sequenced by trypsin^{+cit} included seven in FGA (287, 308, 353, 367, 394, 404, 425), four in FGB (121, 124, 285, 410), and two in FGG (40, 282). The cFG tryptic digest yielded 35 unique peptides (none containing C-terminal citrulline) mapping to 11 FGA sites, 6 FGB sites, and 1 FGG site. Previously identified citrullination sites not sequenced by trypsin included eight FGA sites (308, 353, 367, 394, 404, 425, 426, 510) and four FGB sites (121, 124, 285, 410) (36).

Discussion

Trypsin's high activity, solubility, and ancestry of divergent proteolytic specificity make it a promising scaffold for evolution. However, trypsin possesses six nonconsecutive disulfide bonds and it is translated as a zymogen that requires posttranslational proteolytic cleavage of an N-terminal propeptide to produce the mature enzyme, presenting significant challenges for conventional protein evolution involving recombinant expression. Supplementing IVTT to promote S–S bond formation, folding, and propeptide proteolysis sidestepped the throughput limitations of recombinant expression, but the cost of IVTT reagent and complexity of libraries containing all possible loop sequences required the much higher throughput and miniaturization of compartmentalized evolution (27).

Screening for citrulline-dependent proteolytic activity yielded trypsin^{+cit}, an efficient citrulline-dependent mutant. In addition to the expected reconfiguration of loop 1, another mutation, L7P, was observed in the posttranslationally cleaved N-terminal propeptide signal sequence. That L7P alone was required for bacterial overexpression of both trypsin and trypsin^{+cit} cast doubt on its role in the evolution of novel proteolytic specificity. Exogenous proteins possessing signal peptides are known to enter the secretion pathway, but are rapidly degraded (37); a single Leu-to-Pro mutation of the signal peptide hydrophobic region completely disrupts signal peptide function, restores cytosolic expression, and increases protein synthesis in vitro (38–40). These findings inform future library design, because L7P will be both advantageous for IVTT yield and convenient for moving screening hits directly into recombinant expression.

Deployment of trypsin^{+cit} in standard MS proteomics analyses of PAD4 and fibrinogen as case studies revealed unique signatures of modification. Trypsin^{+cit} analysis of PAD4, which catalyzes the deimination of other proteins and itself (33, 41), definitively mapped 12 sites of citrullination based on detection of y_1(cit), a signature of fragmenting C-terminally citrullinated peptides at the C-terminal amide bond. XIC analysis detected 11 of the 12 peptides as pairs of both modified and unmodified versions, providing additional site discovery confirmation. Fibrinogen is a target of PAD4 activity that exists as three chains, one of which (FGA) is extensively modified. Trypsin^{+cit} also identified the majority of known citrullination sites in cFG. Both case studies successfully demonstrated y_1(PTM) detection, setting the stage for a generalized approach to PTM discovery. PTM-dependent enzymatic cleavage and the resulting y_1(PTM)

could enable selected reaction monitoring (SRM) beyond MS-labile modifications, such as phosphorylation (42), with transition detection unambiguously assigning the modification site to the terminus. This study predicts such y_1(PTM) SRM experiments will not require an enzyme with high cleavage selectivity for the PTM over other residues. The enzyme must simply generate the cleavages efficiently and with appropriate frequency, and the PTM must generate a detectable y_1 fragment. Of course, these experiments are incumbent upon using the present approach to evolve other proteases, a prospect proposed more than a decade ago (8).

Methods

Emulsification. PCRs contain PCR buffer (1×), dNTPs (200 μM each), PEG-8000 (4.8 mM), butanediol (5% [vol/vol]), KF-6012 (0.02% [vol/vol]; ShinEtsu), Taq (0.3 U/μL), oligonucleotide primer 5′-TGCGTCCGGCGTAGAGGATC-3′, EGGK(185–188)XXXX/D189S library (800 pg), and $(acIPcit)_2$R110 probe–primer beads (5×10^6). IVTT reactions (New England BioLabs) contained beads from emPCR, solution A (74.7 μL), solution B (56 μL), disulfide enhancers 1 and 2 (14.9 μL each), and EK (1/200 dilution, 4.1 μL). The reaction (150 μL) was combined with ice-cold oil (76% [wt/vol] DMF-A-6cs; 20% [wt/vol] mineral oil; 4% [wt/vol] KF-6038; 600 μL) and a stainless steel bead (6 mm diameter). Sample was loaded into a homogenizer (TissueLyser; QIAGEN), emulsified (10 s, 15 Hz; 10 s, 17 Hz), and placed on ice. Aliquots of emPCR (50 μL) were transferred to PCR tubes for thermal cycling ([95 °C, 20 s; 68 °C, 90 s] × 30 cycles; 68 °C, 600 s). The emIVTT samples were incubated (67 h, 37 °C) protected from light.

Real-Time IVTT Assay. Purified PCR product from single-bead amplifications (5 ng) or plasmid template (22 ng) was added to IVTT reactions (5 μL, above) and either $(acIPcit)_2$R110 (3 μM) or $(cbzIPR)_2$R110 (3 μM). Reactions were incubated (37 °C, 1,000 min) with fluorescence monitoring (channel 1, CFX96; Bio-Rad).

Detailed methods are available in *SI Methods*.

ACKNOWLEDGMENTS. Prof. Paul R. Thompson (University of Massachusetts Medical School) provided the kind gift of purified PAD4 protein for analysis. Support for this work was provided by an NIH Director's New Innovator Award OD008535 (to B.M.P.) and a National Science Foundation Research Experiences for Undergraduates Grant 1359369 (to H.L.T.).

1. Altelaar AFM, Munoz J, Heck AJR (2013) Next-generation proteomics: Towards an integrative view of proteome dynamics. *Nat Rev Genet* 14(1):35–48.
2. Mayne J, et al. (2016) Bottom-up proteomics (2013–2015): Keeping up in the era of systems biology. *Anal Chem* 88(1):95–121.
3. Rodriguez J, Gupta N, Smith RD, Pevzner PA (2008) Does trypsin cut before proline? *J Proteome Res* 7(1):300–305.
4. Choudhary G, Wu S-L, Shieh P, Hancock WS (2003) Multiple enzymatic digestion for enhanced sequence coverage of proteins in complex proteomic mixtures using capillary LC with ion trap MS/MS. *J Proteome Res* 2(1):59–67.
5. Swaney DL, Wenger CD, Coon JJ (2010) Value of using multiple proteases for large-scale mass spectrometry-based proteomics. *J Proteome Res* 9(3):1323–1329.
6. Giansanti P, Tsiatsiani L, Low TY, Heck AJR (2016) Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. *Nat Protoc* 11(5):993–1006.
7. Rusnak F, Zhou J, Hathaway GM (2002) Identification of phosphorylated and glycosylated sites in peptides by chemically targeted proteolysis. *J Biomol Tech* 13(4):228–237.
8. Knight ZA, et al. (2003) Phosphospecific proteolysis for mapping sites of protein phosphorylation. *Nat Biotechnol* 21(9):1047–1054.
9. Knight ZA, Garrison JL, Chan K, King DS, Shokat KM (2007) A remodelled protease that cleaves phosphotyrosine substrates. *J Am Chem Soc* 129(38):11672–11673.
10. Varadarajan N, Gam J, Olsen MJ, Georgiou G, Iverson BL (2005) Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity. *Proc Natl Acad Sci USA* 102(19):6855–6860.
11. Varadarajan N, Georgiou G, Iverson BL (2008) An engineered protease that cleaves specifically after sulfated tyrosine. *Angew Chem Int Ed Engl* 47(41):7861–7863.
12. Varadarajan N, Pogson M, Georgiou G, Iverson BL (2009) Proteases that can distinguish among different post-translational forms of tyrosine engineered using multicolor flow cytometry. *J Am Chem Soc* 131(50):18186–18190.
13. Meyer JG, et al. (2014) Expanding proteome coverage with orthogonal-specificity α-lytic proteases. *Mol Cell Proteomics* 13(3):823–835.
14. Guerrero JL, O'Malley MA, Daugherty PS (2016) Intracellular FRET-based screen for redesigning the specificity of secreted proteases. *ACS Chem Biol* 11(4):961–970.
15. Craik CS, et al. (1985) Redesigning trypsin: Alteration of substrate specificity. *Science* 228(4697):291–297.
16. Hedstrom L, Perona JJ, Rutter WJ (1994) Converting trypsin to chymotrypsin: Residue 172 is a substrate specificity determinant. *Biochemistry* 33(29):8757–8763.
17. Perona JJ, Hedstrom L, Rutter WJ, Fletterick RJ (1995) Structural origins of substrate discrimination in trypsin and chymotrypsin. *Biochemistry* 34(5):1489–1499.
18. Hedstrom L, Szilagyi L, Rutter WJ (1992) Converting trypsin to chymotrypsin: The role of surface loops. *Science* 255(5049):1249–1253.
19. Evnin LB, Vásquez JR, Craik CS (1990) Substrate specificity of trypsin investigated by using a genetic selection. *Proc Natl Acad Sci USA* 87(17):6659–6663.
20. Christophorou MA, et al. (2014) Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature* 507(7490):104–108.
21. Lewis HD, et al. (2015) Inhibition of PAD4 activity is sufficient to disrupt mouse and human NET formation. *Nat Chem Biol* 11(3):189–191.
22. Kubota K, Yoneyama-Takazawa T, Ichikawa K (2005) Determination of sites citrullinated by peptidylarginine deiminase using 18O stable isotope labeling and mass spectrometry. *Rapid Commun Mass Spectrom* 19(5):683–688.
23. Holm A, et al. (2006) Specific modification of peptide-bound citrulline residues. *Anal Biochem* 352(1):68–76.
24. Tutturen AEV, et al. (2010) A technique for the specific enrichment of citrulline-containing peptides. *Anal Biochem* 403(1-2):43–51.
25. Bicker KL, Subramanian V, Chumanevich AA, Hofseth LJ, Thompson PR (2012) Seeing citrulline: Development of a phenylglyoxal-based probe to visualize protein citrullination. *J Am Chem Soc* 134(41):17015–17018.
26. Steen H, Mann M (2004) The ABC's (and XYZ's) of peptide sequencing. *Nat Rev Mol Cell Biol* 5(9):699–711.
27. Tawfik DS, Griffiths AD (1998) Man-made cell-like compartments for molecular evolution. *Nat Biotechnol* 16(7):652–656.
28. Sepp A, Tawfik DS, Griffiths AD (2002) Microbead display by in vitro compartmentalisation: Selection for binding using flow cytometry. *FEBS Lett* 532(3):455–458.
29. Leytus SP, Patterson WL, Mangel WF (1983) New class of sensitive and selective fluorogenic substrates for serine proteinases. Amino acid and dipeptide derivatives of rhodamine. *Biochem J* 215(2):253–260.
30. Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B (2003) Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. *Proc Natl Acad Sci USA* 100(15):8817–8822.
31. Gráf L, et al. (1988) Electrostatic complementarity within the substrate-binding pocket of trypsin. *Proc Natl Acad Sci USA* 85(14):4961–4965.
32. Capece MC, Kornberg GL, Petrov A, Puglisi JD (2015) A simple real-time assay for in vitro translation. *RNA* 21(2):296–305.
33. Slack JL, Jones LE, Jr, Bhatia MM, Thompson PR (2011) Autodeimination of protein arginine deiminase 4 alters protein-protein interactions but not activity. *Biochemistry* 50(19):3997–4010.
34. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20(18):3551–3567.
35. Craig R, Beavis RC (2004) TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics* 20(9):1466–1467.
36. van Beers JJ, et al. (2010) Mapping of citrullinated fibrinogen B-cell epitopes in rheumatoid arthritis by imaging surface plasmon resonance. *Arthritis Res Ther* 12(6):R219.
37. Gentz R, et al. (1988) Association of degradation and secretion of three chimeric polypeptides in *Escherichia coli*. *J Bacteriol* 170(5):2212–2220.
38. Bedouelle H, et al. (1980) Mutations which alter the function of the signal sequence of the maltose binding protein of *Escherichia coli*. *Nature* 285(5760):78–81.
39. Emr SD, Silhavy TJ (1983) Importance of secondary structure in the signal sequence for protein secretion. *Proc Natl Acad Sci USA* 80(15):4599–4603.
40. Ibrahimi I, Gentz R (1987) A functional interaction between the signal peptide and the translation apparatus is detected by the use of a single point mutation which blocks translocation across mammalian endoplasmic reticulum. *J Biol Chem* 262(21):10189–10194.
41. Andrade F, et al. (2010) Autocitrullination of human peptidyl arginine deiminase type 4 regulates protein citrullination during cell activation. *Arthritis Rheum* 62(6):1630–1640.
42. Unwin RD, et al. (2005) Multiple reaction monitoring to identify sites of protein phosphorylation with high sensitivity. *Mol Cell Proteomics* 4(8):1134–1144.
43. Chan TR, Hilgraf R, Sharpless KB, Fokin VV (2004) Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Org Lett* 6(17):2853–2855.
44. Wiśniewski JR, Zougman A, Nagaraj N, Mann M (2009) Universal sample preparation method for proteome analysis. *Nat Methods* 6(5):359–362.