Principles of *in situ* protein sequencing: expansion microscopy-adapted Edman degradation and amino acid recognition

Camille M. Mitchell[1,2], Sara Z. Tavana[1,2], Joanne Z. Peng[2], Hao Wang[2,13], Jiuhan Shi[2], Chi Zhang[2], Lilia Evgeniou[14], Masy Domecillo[11,12], Shiwei Wang[2,9,10], Daniel M. Estandian[1,2], Alexi G. Choueiri[1,2], Evelyn Wong[1], Sarah Dohadwala[1], Nicholas F. Polizzi[11,12], Laura L. Kiessling[4,9,10], Edward S. Boyden[1,2,3,4,5,6,7,8*]

[1]Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA
[2]McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, USA
[3]Media Arts and Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA
[4]Koch Institute, Massachusetts Institute of Technology, Cambridge, MA, USA
[5]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
[6]Center for Neurobiological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA
[7]K. Lisa Yang Center for Bionics and Yang Tan Collective, Massachusetts Institute of Technology, Cambridge, MA, USA
[8]Howard Hughes Medical Institute, Cambridge, MA, USA
[9]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA
[10]Broad Institute of MIT and Harvard, Cambridge, MA, USA
[11]Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA.
[12]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02215, USA.
[13]The Picower Institute for Learning and Memory, Massachusetts Institute of Technology, Cambridge, MA, USA
[14]Department of Systems Biology, Harvard Medical School, Boston, MA, USA
* correspondence to edboyden@mit.edu

Abstract

The ability to map protein identity, with resolution sufficient to infer interactions, would support analysis of how proteins work together, or malfunction, in biological processes and diseases. Although several emerging technologies aim towards single-molecule protein sequencing, they require proteins to be removed from the nanoscale spatial context of cells and tissues. Expansion microscopy (ExM) has facilitated a diversity of chemical analyses by isotropically separating molecules throughout a specimen after permeation via a charged hydrogel, followed by gel swelling. Here, we adapt key protein sequencing steps - Edman degradation and amino acid recognition - to the ExM gel context. Using testbed peptides in ExM gels, we show that N-terminal amino acids can be recognized over multiple cycles of in-gel Edman degradation. These results establish principles of *in situ* protein sequencing and provide a framework for future *in situ* protein sequencing developments, including the development of higher specificity and affinity amino acid binders.

Main

Proteins are central to life processes[1], and their interactions achieve metabolic or signaling outcomes with great efficacy and precision, in a fashion that depends on nanoscale context[2–4]. As just one example, synaptic NMDA receptor signaling can promote neuronal health, whereas extrasynaptic NMDA receptor signaling can result in neuronal death[5]. A strategy to map all protein identities and positions, within intact cellular contexts, ideally at single molecule resolution (e.g., 1 nm or better) so that interactions can be inferred, could enable the generation of novel hypotheses and reveal insights into how proteins interact to mediate essential and disease-related processes. However, existing spatial protein mapping techniques either do not scale to the entire proteome, and/or fall short of single molecule spatial resolution and sensitivity (see **Supplementary Table 1** for a list of spatial protein mapping strategies and their quantitative properties).

To map all protein identities, locations, and putative interactions, ideally one would be able to identify individual proteins, with single molecule resolution (e.g., ~1 nm or better). To this end, expansion microscopy (ExM) is a tool that involves the chemical anchoring of specific amino acids within a protein to a cell- or tissue-permeating swellable hydrogel; the biological specimen is then chemically softened (e.g., with heat and detergent, or enzymatic, treatment), and then immersed in water, which causes isotropic expansion of the specimen-hydrogel composite[6,7]. The net outcome is that a microscope's effective resolution is improved by an expansion factor (e.g., ~4.5-fold, in the original version). ExM has been used in hundreds of experimental studies throughout biology[8]. ExM can be iterated, as a sample can be expanded, then a second hydrogel chemically formed in the space opened up by the first expansion, and finally the specimen expanded again, with >50x expansion already having been demonstrated at proof of concept level (by iterating ~4x expansion three times); with novel ExM chemistries that achieve 20-fold expansion in a single step, expansion factors of 100-1000x, or even more, may be feasible[9–11]. The utility of such expansion factors might be justified by a recent study that has shown, by following a specific 10-fold expansion protocol by super-resolution radial fluctuations (SRRF) imaging, that ExM can be used to separate a protein's proteolyzed fragments from each other with submolecular (e.g., <1 nm) precision and in an isotropic fashion that significantly preserves protein shape[12]. We therefore postulated that ExM could facilitate *in situ* deployment of protein sequencing.

Many efforts are ongoing towards *ex situ* single-molecule protein sequencing. These include using fluorescent N-terminal amino acid binders, or covalent amino acid labels, in conjunction with iterative amino acid removal strategies (e.g. Edman degradation). These technologies operate upon proteins *ex situ*, extracted from the spatial context of intact biological cells and tissues[13–17]. If proteins throughout an expanded cell or tissue were fragmented into pieces, and the pieces separated from each other (**Fig. 1A**), then one could in principle walk down each peptide fragment, amino acid by amino acid using Edman degradation (**Fig. 1B**), applying N-terminal binders (or covalent amino acid labels) to such peptide fragments to identify each N-terminal amino acid in turn (**Fig. 1C**). Finally, computational methods could assemble these sequences into an overall protein identity or even sequence. ExM would support such an effort in at least two ways: it would provide the magnification needed to resolve individual peptide fragments (perhaps assisted by further super-resolution imaging), and would enable peptides to be surrounded by a well-defined chemical environment appropriate for protein sequencing (and to allow them to be more easily followed over the many rounds of imaging required for sequencing, as previously demonstrated for nucleic acids[18–20]).

Edman degradation removes amino acids one at a time using a two-step process. First, phenylisothiocyanate (PITC) reacts with the N-terminal amino acid, and then trifluoroacetic acid (TFA) results in cyclization followed by cleavage[21], at an efficiency of up to ~99% per round[22]. *Ex situ* single molecule protein sequencing using Edman degradation is an active area of research, e.g., through covalent fluorescent tagging of amino acids[15], or the application of N-terminal amino acid binders (NAABs)[14,16,23]. Here, we adapted Edman degradation chemistry to ExM to afford an in-gel Edman degradation protocol. We screened through solvents, Edman reagents, and reaction conditions to optimize isothiocyanate (ITC)–peptide conjugation in ExM gels. To optimize the chemical reactions that underpin the process, we analyzed testbed peptides with well-defined sequences in gels, analogous to the path taken by early *in situ* nucleic acid sequencing[24]. We demonstrated that we can

evaluate N-terminal amino acid binders as they bind to iteratively exposed N-terminal amino acids of testbed peptides. The creation of such binders is an active area of interest for *ex situ* protein sequencing[14,23,25–31], and further research in this space will be required for a full protein sequencing protocol to be developed. However, the principles revealed here are fundamental to realizing this objective.

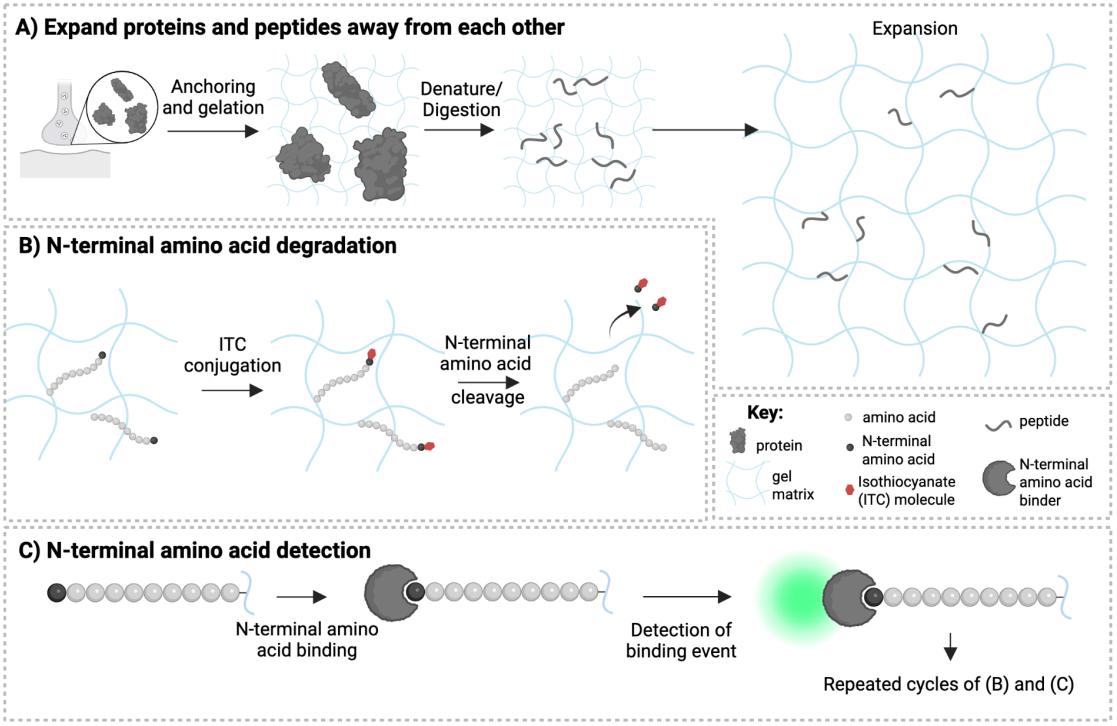


Figure 1 In-gel Edman degradation concept: a chemistry platform to support *in situ* protein sequencing technology development and application.

(A) Expansion of the gel network to separate proteins and protein fragments, prior to peptide sequencing, exploits expansion microscopy (ExM)-style processing of a biological sample (e.g., a cell), including covalent anchoring key amino acids (e.g., lysine side chain modified with 6-((acryloyl)amino)hexanoic acid (AcX), in the original proExM protocol[6,7]), gelation, denaturation and/or digestion of proteins, and expansion, followed by re-embedding of ExM gels in charge-neutral gels to stabilize them in the expanded state (not shown).

(B) Edman-in-gel chemistry can be conducted via conjugation of an isothiocyanate (ITC) derivative (e.g., phenylisothiocyanate (PITC)) to N-terminal amino acids of protein fragments or peptides, followed by cleavage of this N-terminal amino acid (e.g., with trifluoroacetic acid (TFA)).

(C) The N-terminal amino acid at the end of a protein fragment or peptide can be detected with an N-terminal amino acid binder (e.g., ClpS2 St-V1[25,26]). The binding event is monitored (e.g., with antibodies with or without signal amplification, depicted as a green fluorescent signal next to the binder), before washing away of the binder and a new round of N-terminal amino acid degradation (B). Steps (B) and (C) are performed iteratively to sequentially identify the amino acids of the protein fragment or peptide.

Results

Solvent considerations for in-gel Edman degradation

In-gel Edman degradation is a version of Edman degradation that requires anchoring of peptides via some functional groups (akin to solid phase sequencing), for instance through carboxyl groups as

performed in *ex situ* protein sequencing[15,32] or via the primary amine groups, as performed traditionally in ExM[7]. The sequencing of peptides could be performed on neo N-termini generated by selective peptide cleavage after denaturation/digestion, enabling sequencing of many peptides in parallel (see **Supplementary Note 10** and **Supplementary Figure 13c** for modeled exploration of the effect of anchoring and digestion efficiency on the percent of amino acids that are detected). To adapt Edman degradation to the ExM gel, we had to first discover a solvent that was compatible with gels remaining in the expanded state, while enabling PITC conjugation to peptides. The overall chemical reaction architecture for in-gel Edman chemistry is schematized in **Fig. 2a** (background information on Edman degradation, is in **Supplementary Note 1**, with reagents listed in **Supplementary Table 2**). We cast ExM gels by reacting sodium acrylate, acrylamide, and N,N'-methylenebisacrylamide (BIS, at a slightly high concentration to ensure gel robustness in subsequent steps, resulting in an expansion factor of 3.5x), in a fashion that would either incorporate gel-anchored peptides, or that would omit peptides, as a control. We then cast an acrylamide gel throughout the expanded gel (resulting in a re-embedded ExM gel, or ExMre for short), to stabilize it (at 2.5x expansion, increasing back to 3.5x when exposed to water) and reduce shrinkage under varying solvent and salt compositions (PBS causes ExMre gels to shrink to 2.7x, vs. 1.8x for native ExM gels; **Fig. 2b**; see **Supplementary Note 2** for information on the ionic strength of these solutions), as used previously for *in situ* RNA sequencing[19].

To understand how the reaction conditions might influence the gel, we quantified the flat/top side surface of the gels with a ruler, after placing PBS-washed gels in various PITC conjugation solvents for 30 min at 50 °C, followed by adding PITC (30 min at 50 °C). Pyridine and acetonitrile (ACN) are solvents commonly used for PITC conjugation that resulted in gel shrinkage, and in some cases opacity, which could prevent chemical access to the inside of the gel (**Fig. 2ci-iii**; see **Supplementary Figure 1** and **Supplementary Figure 2** for images and quantifications of gels cited throughout **Fig. 2**; see **Supplementary Table 4** for full data and statistics for **Fig. 2**), alone or supplemented with PITC (either at 1:9, or 1:1000, which we also found to suffice; **Supplementary Figure 4**). Many other solvents also turned the gel opaque and/or shrank gels (see **Supplementary Figure 1g** for images). We reasoned that solvents not previously used in Edman degradation, such as dimethylsulfoxide (DMSO) and formamide, might solubilize PITC while being compatible with hydrophilic ExM gels. Such solvents did not cause such severe gel shrinkage (**Fig. 2civ-v**), alone or with PITC. FITC, a nontraditional Edman reagent that is more water soluble, administered in DMSO with sodium bicarbonate led to no change in gel flat/top surface size (**Fig. 2cvi**), suggesting these conditions might be useful in the first Edman degradation step. Finally, TFA, needed in the cleavage step, did not shrink ExMre gels (**Fig. 2cvii**). Thus, we could find solvents compatible with Edman degradation conditions in expanded gel states.

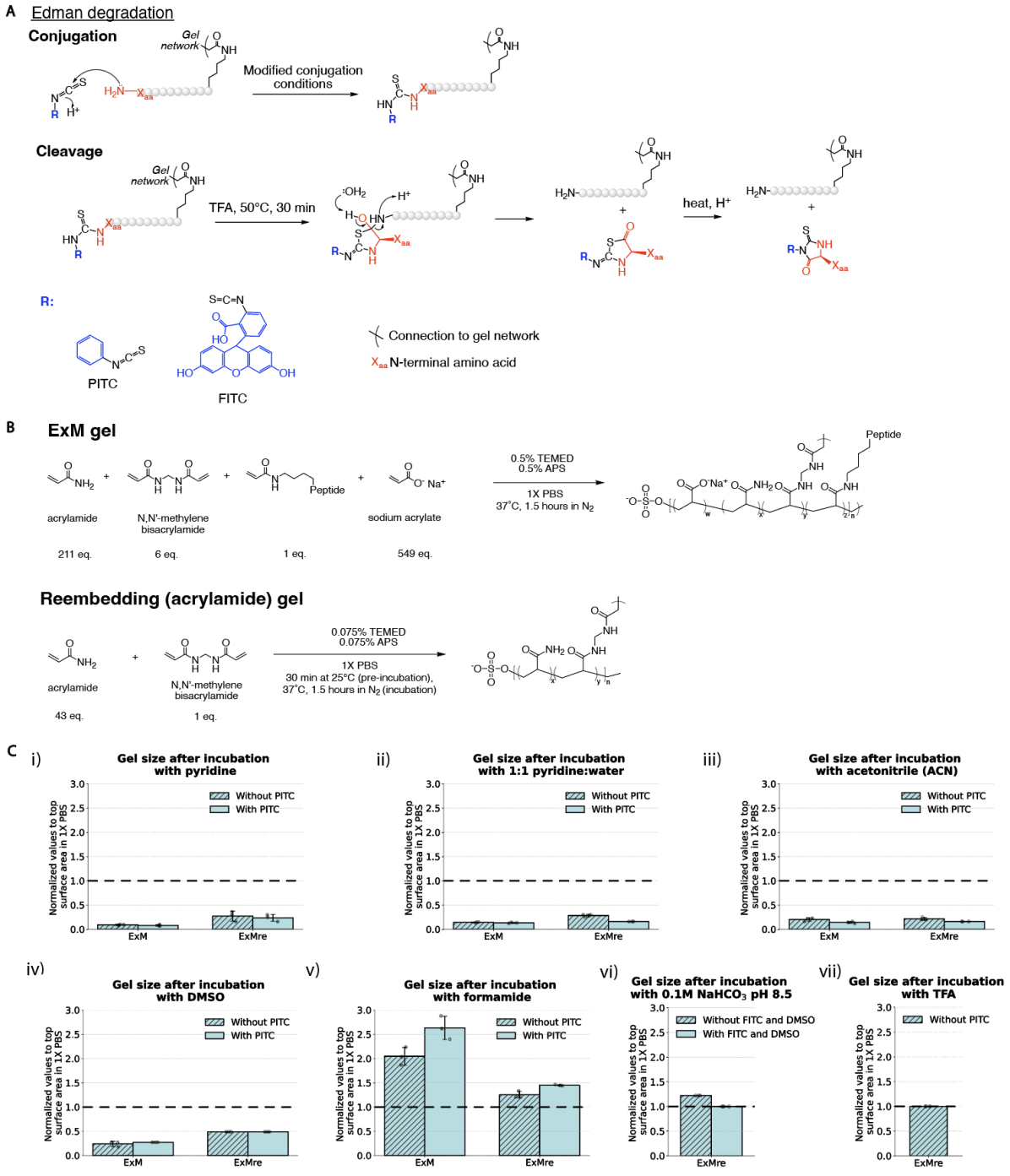


Figure 2 Edman degradation: chemical strategies for in-gel adaptation, and initial characterization of gel compatibility.

(A) Proposed reaction scheme of Edman degradation within an ExM gel. To develop conditions, specific peptides were anchored to the gel. The sequencing chemistry includes conjugation of an isothiocyanate (ITC) molecule (i.e., Edman reagent) to the N-terminal amino acid (top). Edman degradation can be performed with a variety of isothiocyanates (denoted "R", and depicted in blue). The conjugated N-terminal amino acid (denoted "Xaa", and depicted in red) is then cleaved with TFA, eliminating the thiohydantoin amino acid derivative.

(B) Reaction scheme for ExM, and re-embedding of ExM gels. Proteins will be anchored to the gel and then pulled apart from each other. We focus on using synthetic peptides with a C-terminal acryloyl functional group for covalent conjugation to the gel during free-radical polymerization. TEMED, tetramethylethylenediamine; APS, ammonium persulfate.

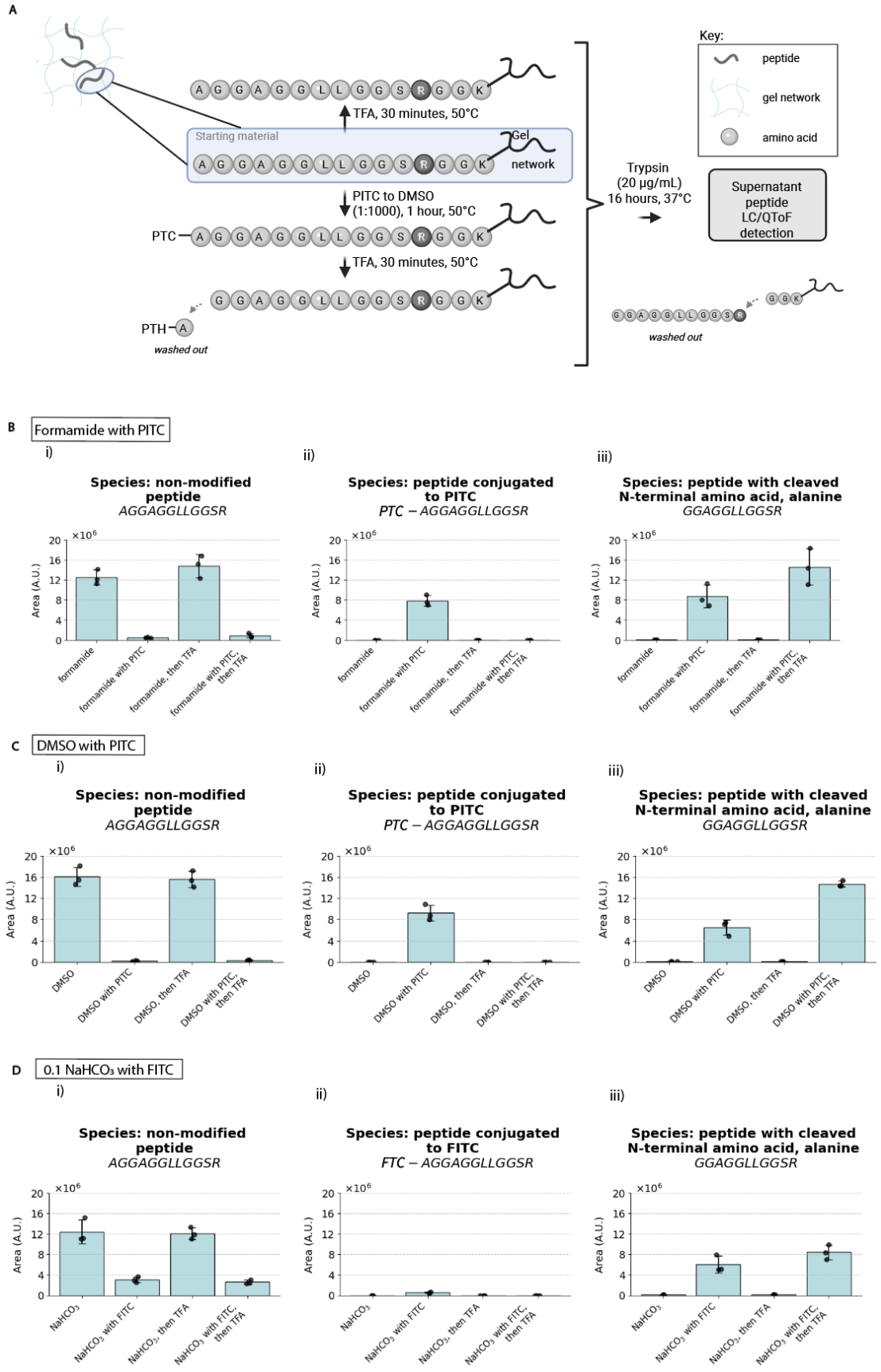
(C) Flat/top side surface size of ExM and ExM re-embedded (ExMre) gels when placed in solvents used in Edman degradation: (i) 100% pyridine, (ii) 1:1 pyridine:water (all ratios throughout are of volumes added, unless otherwise indicated), (iii) acetonitrile (ACN), (iv) dimethylsulfoxide (DMSO), (v) formamide, (vi) 0.1 M sodium bicarbonate ($NaHCO_3$). pH 8.5, (vii) trifluoroacetic acid (TFA), with and without Edman reagents (PITC 1:9 ratio PITC:solvent for (i), PITC 1:1000 PITC:solvent for (ii-v); FITC, 5.9 mM, and 23:77 DMSO:0.1 M sodium bicarbonate buffer pH 8.5, for (vi)). (Black dashed line: top/flat surface size of the gel in PBS; error bar: standard deviation; black dots, individual experiments; n=3 separate gelation solutions for (ii-vii); n=3 gels with same starting gelation solution for (i).)

Testbed peptides anchored throughout ExMre gels for validation of in-gel Edman degradation

Using testbed peptides, we devised a set of mass spectrometry and fluorescent measurements to characterize the efficacy of Edman reagent conjugation and subsequent cleavage. For mass spectrometry, we used a peptide that could be trypsinized from the gel to enable analysis via liquid chromatography coupled with electrospray ionization quadrupole time-of-flight mass spectrometry (or LC/QToF for short) (**Fig. 3a**; see **Supplementary Figure 12** for discussion of the control peptide). We note that the phenylthiocarbamyl (PTC)-peptide bond resulting from Edman reagent conjugation is labile under preparation conditions for mass spectrometry (see **Supplementary Note 3** for more detail about this limitation), and thus we focus our analysis on the efficacy of ITC conjugation. As expected, conventional Edman solvents that caused gel shrinkage or opacity did not result in good conjugation (see **Supplementary Note 4**). Formamide and DMSO, in contrast, enabled excellent conversion to the PTC-peptide with a conversion rate of ~96% for 1:1000 ratio PITC:formamide (**Fig. 3b**; see **Supplementary Table 5** for full data and statistics related to **Fig. 3**) and a conversion rate of >99% for 1:1000 PITC:DMSO (**Fig. 3c**). FITC in 0.1 M sodium bicarbonate resulted in a lower conversion rate of ~75% (**Fig. 3d**), as expected from the lower FITC conjugation rate reported in the literature[33]. All three protocols (**Fig. 3e**), could be conducted without substantial gel shrinkage.

To assess phenylthiohydantoin release, we devised a strain-promoted alkyne-azide cycloaddition (SPAAC) click chemistry fluorescence experiment (**Fig. 3f**). We used an N-terminal azidolysine peptide that was clicked to a dibenzocyclooctyne (DBCO) fluorophore after various chemical treatments. When the gel was exposed to our Edman degradation conditions, we observed the expected drop in fluorescence due to cleavage (**Fig. 3gi-ii**), about ~70% of the total (**Fig. 3giii**; trypsin-mediated positive control). We further explored whether the cleaved phenylthiohydantoin (PTH)-amino acid could be detected directly via LC/QToF (**Fig. 3hi**), and confirmed PTH-amino acid release under Edman conditions (**Fig. 3hii**; see **Supplementary Figure 6** and **Supplementary Note 5** for results regarding other amino acids).

Figure 3



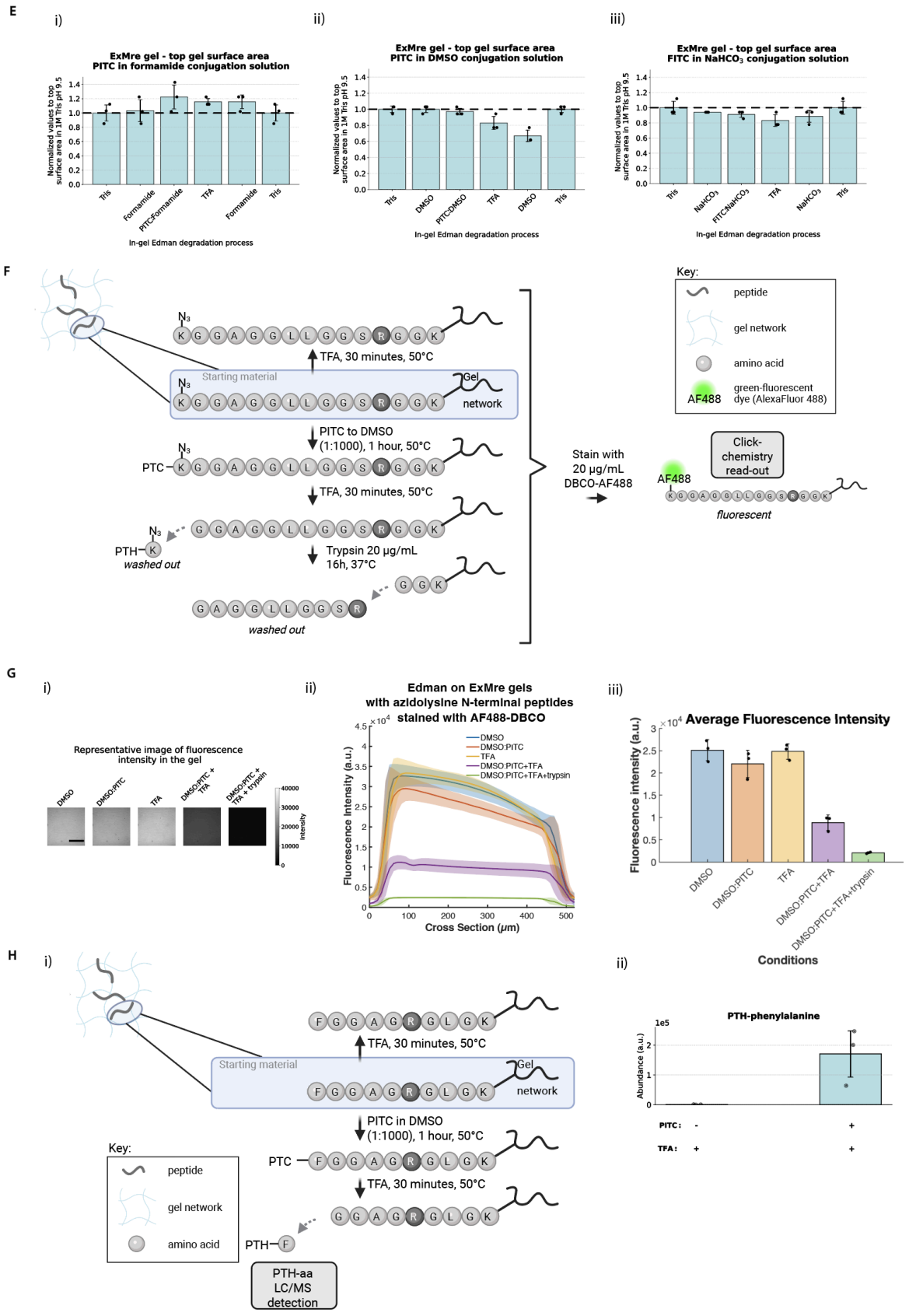


Figure 3 Assessing Edman degradation compatibility on peptides in ExMre gels using LC/QToF trypsinization assay, SPAAC fluorescence assay, and PTH-aa detection.

(A) Assay to assess efficiency of Edman reagent conjugation on synthetic peptides in ExMre gels. Synthetic peptide AGGAGGLLGGSRGGK{acr} (abbreviated as A15-peptide; K{acr} denotes an acryloyl functional group on the lysine side chain); amino acids of the peptide chain are depicted as light grey beads, with arginine in the dark grey bead, as site of trypsin cleavage. Peptides are embedded into ExMre gels during free-radical polymerization of the swellable gel, and separate gels are subjected to various in-gel Edman conditions. All ExMre gels are then immersed in trypsin for digestion overnight to acquire the fragment upstream to the trypsin cleavage site at arginine, for further analysis. Supernatant is analyzed with liquid chromatography coupled with electrospray ionization quadrupole time-of-flight mass spectrometry (LC-ESI-QToF MS, abbreviated LC/QToF).

(B) In-gel Edman chemistry with formamide as conjugation solvent and PITC as the Edman reagent. Bar graphs representing the relative abundance (arbitrary units, a.u., with all samples processed with spiked-in control compounds; see **Methods** and **Supplementary Figure 12**) of different peptide ion species detected on the LC/QToF. Bar graphs were obtained from measuring the area under the curve (AUC) of the chromatogram of various species (extracted based on the exact mass, see **Methods** for details, and **Source Data** for raw traces). The separate conditions were solvent only ("formamide"), PITC to solvent (1:1000 ratio PITC:solvent) for 1 hour at 50 °C ("formamide with PITC"), TFA for 30 min at 50 °C ("formamide, then TFA"), PITC to solvent (1:1000 ratio PITC:solvent) for 1 hour at 50 °C followed by TFA for 30 min at 50 °C ("formamide with PITC, then TFA"). The relative abundance of the ion species: (i) non-modified peptide (AGGAGGLLGGSR), (ii) peptide conjugated to PITC, phenylthiocarbamyl (PTC)-peptide (*PTC*-AGGAGGLLGGSR), and (iii) peptide with cleaved N-terminal amino acid (GGAGGLLGGSR), were reported throughout the in-gel Edman degradation process in the various conditions (dots, individual experiments; blue bar, mean; error bar, standard deviation, n=3 separate gelation solutions).

(C) In-gel Edman chemistry as in B, but with DMSO as solvent.

(D) In-gel Edman chemistry as in C, but with 23:77 of DMSO:0.1 M sodium bicarbonate pH 8.5 as conjugation solution with FITC (5.9 mM) as Edman reagent. The peptide conjugated to FITC is fluorescein-thiocarbamyl (FTC)-peptide (*FTC*-AGGAGGLLGGSR).

(E) Top/flat surface size of ExMre gels throughout the Edman degradation process. Normalized top/flat surface size at each step for ExMre gels with (i) formamide conjugation with PITC, (ii) DMSO conjugation with PITC, or (iii) 5.9 mM FITC in 23:77 DMSO:0.1 M $NaHCO_3$ pH 8.5 (thick dashed black line, top/flat surface size of the gel after washes in 1 M Tris pH 9.5 (1 mL x 3) before Edman degradation; error bar, standard deviation; black dots, individual experiments; n=3 separate gelation solutions).

(F) Assay to assess efficiency of Edman degradation on synthetic peptides in ExMre gels. Synthetic peptide: K{N_3}GGAGGLLGGSRGGK{acr} (abbreviated K{N_3}15-peptide, where K{N_3} is 6-azido-lysine), and amino acids of the peptide chain are depicted as light grey beads, with arginine, "R", with the dark grey bead, as site of trypsin cleavage. The peptides are embedded into the ExMre gels during free-radical polymerization of the first gel, and separate gels are subjected to various Edman conditions. Read-out is performed using strain-promoted alkyne-azide cycloaddition (SPAAC) with dibenzocyclooctyne AlexaFluor 488 (DBCO-AF488) on the embedded peptides followed by analysis of the fluorescence. Note: sloped intensity profiles were likely due to excitation light attenuation in deeper layers of highly fluorescent gels.

(G) ExM gels containing K{N_3}15mer-peptide at 1 mM were cast in a gelation chamber, expanded and re-embedded to reach ~2.7X expansion factor at a final peptide concentration of ~50 µM. The separate conditions were DMSO only ("DMSO"), PITC to DMSO (1:1000 ratio PITC:DMSO; "PITC:DMSO"), DMSO followed by TFA, PITC to DMSO (1:1000 ratio PITC:DMSO) followed by TFA, and the same condition but with 20 ug/mL trypsin ("PITC:DMSO+TFA+trypsin"). After SPAAC with 20 µg/mL DBCO-AF488 in PBS, bulk gel

fluorescence was imaged using a confocal microscope with a 10 µm Z-step. Analysis was performed on raw images. (i) A representative raw image of the fluorescence intensity is depicted for each gel condition, taking the 20th slice of the Z-stack for each (~200 µm deep into the gel). Scale bar is 500 µm. (ii) The fluorescence intensity of the ExMre gels in different conditions was compared throughout the gel thickness (~500 µm) (line, mean; shaded area, standard deviation; n=3 separate gelation solutions). (iii) Average fluorescence intensity of ExMre gels in the different conditions across the whole volume imaged (colored bar, mean; black dots, individual experiments; error bar, standard deviation, n=3 separate gelation solutions).

(H) (i) Independent assay for in-gel Edman degradation, with phenylthiohydantoin (PTH)-F detection of Edman degraded synthetic peptide: FGGAGRGLGK{acr} (abbreviated "F_1 peptide") embedded in ExMre gels, as in (A). Separate gels were subjected to various Edman conditions. The conditions included TFA only for 30 min at 50 °C, and PITC to DMSO (1:1000 ratio PITC:DMSO) for 1 hour at 50 °C followed by TFA for 30 min at 50 °C. Subsequently, TFA was removed from the gels and they were immersed in 50 µL of 1:1 acetonitrile to water and agitated. Read-out was then performed by injecting the supernatant into LC/QToF using **LC Method (see Methods)**. (ii) Results acquired as in (i). Analysis of PTH-F abundance was performed using PTH-F exact mass, 282.0827 ± 0.0056 Da (see **Methods for Edman degradation and PTH detection** for details). (blue bar, mean; error bar, standard deviation; black dots, individual experiments; n=3 separate gelation solutions).

Combining in-gel Edman degradation with bulk N-terminal amino acid detection

We set out to see if an existing N-terminal amino acid binder (NAAB) could detect N-terminal amino acids in the expansion gel. An ideal NAAB would have high affinity and specificity, and a slow rate of dissociation - slow enough to support amplification (e.g., by attaching many fluorophores through hybridization chain reaction (HCR) or enzymatic means[19,34], which takes minutes to hours) so that ordinary microscopes could be used to identify individual amino acids of individual peptides. Existing NAABs do not yet reach these criteria, with poor affinity, dwell time, and with affinity dependence on the downstream amino acids of the peptide. Making better NAABs is an active area of research[14,23,25–30], and will likely benefit from efforts in immune and display-based selection, structure based design, and/or AI-driven modeling and design[35].

We chose as our NAAB *Agrobacterium tumefaciens* ClpS2, engineered for higher specificity towards Phe (ClpS2 St-V1)[25] (see **Supplementary Note 6** for thoughts on affinity and sequence information). We measured bulk fluorescence throughout the gel upon applying ClpS2 St-V1 with a fused HA tag, followed by addition of a fluorescent anti-HA tag antibody, using ExMre gels equipped with peptides with N-terminal Phe, Trp, Tyr, and Ala (**Fig. 4a**). We discuss some assumptions and calculations to help interpret images amidst fast binder off kinetics, in **Supplementary Note 7**. We observed high fluorescence for the Phe-bearing peptides, and not the other three (**Fig. 4b-d**), for instance with an ~8 fold difference comparing Phe to Tyr (for all raw data for **Fig. 4** see **Supplementary Table 9**) comparable to our calculated estimate mentioned above (~7 fold difference comparing Phe to Tyr, see **Supplementary Note 8** for calculations). Having established that N-terminal binding was possible, we next sought to test the incorporation of N-terminal binding into a sequencing pipeline. We created gels with peptides bearing Phe on the N-terminus (denoted F_1 peptide) or in the second position (denoted G_1F_2 peptide), so that the NAAB would bind only at the appropriate stage in sequencing (**Fig. 4ei-ii**). We observed the NAAB to bind F_1 peptide only in cases when the N-terminal Phe was still remaining (e.g., the DMSO and TFA-only cases), and not when it was occluded (after PITC conjugation) or cleaved (PITC followed by TFA, or the trypsin control) (**Fig. 4f**). In contrast, for G_1F_2 peptide, we only saw NAAB binding when the Phe was exposed after Edman degradation (PITC followed by TFA), and not under conditions where it was occluded by the N-terminal amino acid (DMSO, TFA-only) or by that amino acid bearing PITC, or after trypsin elimination of the peptide (**Fig. 4g**). In summary, we were able to demonstrate rudimentary protein sequencing using NAABs, over multiple cycles of in-gel Edman degradation. Comparing gel

fluorescence for Phe in the 1st vs. 2nd position, we observed a ~50% drop in fluorescence, suggesting a yield comparable to what was estimated from the click chemistry experiment above.

As noted above, single molecule imaging with such a binder would not be straightforward on ordinary microscopes, one of our criteria for technological success. ClpS2 St-V1 exhibits weak affinity (K_d of ~1.1 μM) with a dwell time of ~10 seconds[14,23,26]. However, single molecule amplification methods used in ExMre gels for read-out with conventional confocal microscopy, such as hybridization chain reaction (HCR) or rolling circle amplification (RCA), require on the order of hours for signal amplification[19,34]. The binder used here would dissociate before any amplification product could be made. Therefore, this NAAB would not be suitable for single molecule detection in an *in situ* setting.

Oxidation analysis of amino acid side chains in ExM gels

As proteins being attached to polymerized ExM gels experience a free-radical filled environment, we examined the susceptibility of amino acid side chains to oxidation as a result of free-radical polymerization. The results are described in **Supplementary Note 9**. In summary, we observed a low number of oxidation products for N-terminal Tyr, Phe, Trp, His, Pro and Arg after ExM polymerization, but more substantial oxidation products for both Cys and Met residues as a result of the gelation process (other amino acids, not tested, do not have appropriately reactive side chains). There might be different ways to prevent, reverse, or compensate for such modifications amidst *in situ* protein sequencing; however, one might also be able to guess the identity of a protein even when only some of its amino acids are recognized - and thus a low number of oxidation sites might not hamper the development of a first working *in situ* protein sequencing technology (see **Supplementary Note 10** and **Supplementary Figure 13d** for a simulation of the fraction of correctly identified proteins using a subset of NAABs). Of course, ExM gels that do not involve free radicals are also possible[36,37].

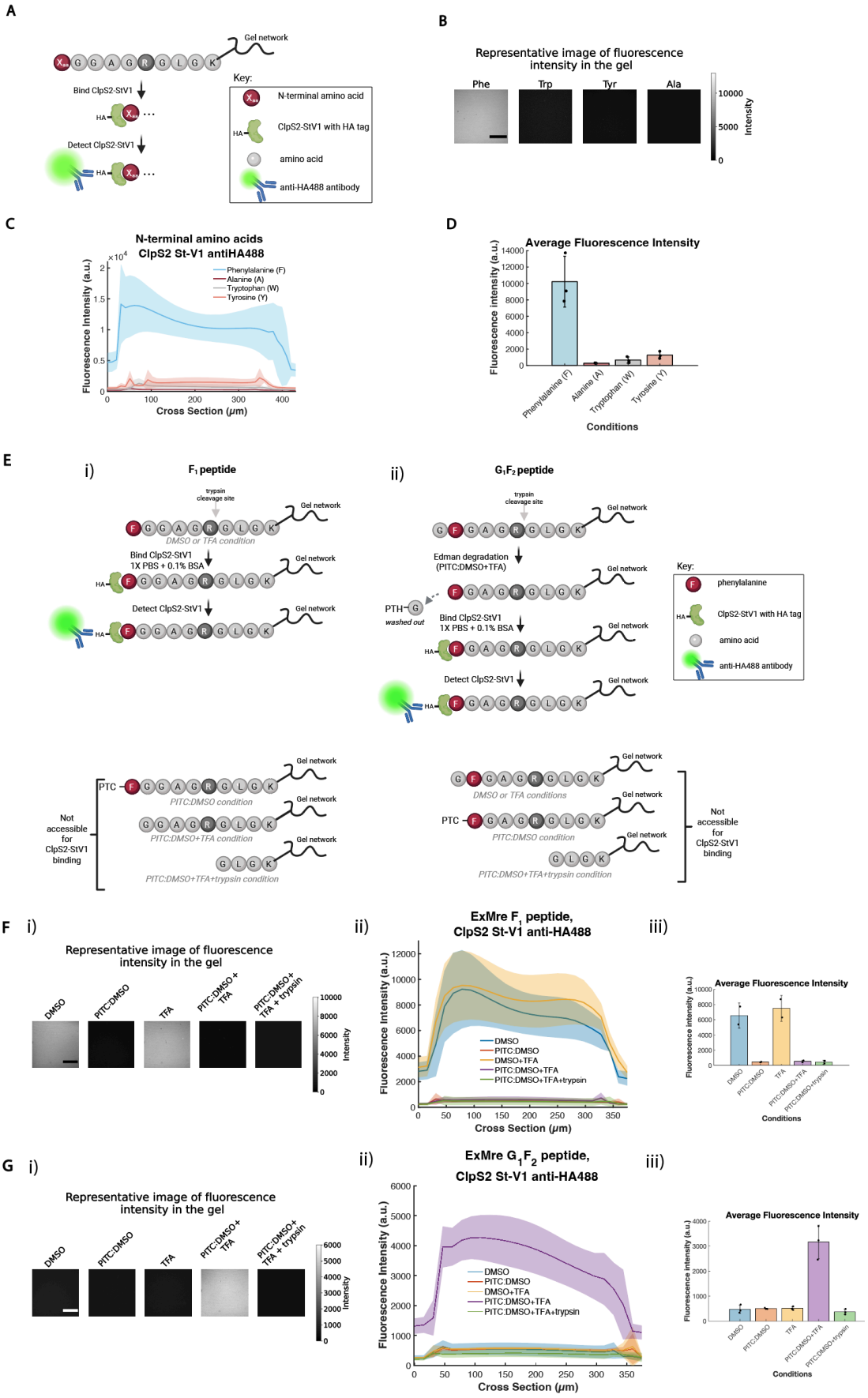


Figure 4 ClpS2 St-V1 serves as read-out for N-terminal phenylalanine on peptides in ExMre gels

(A) Schematic representing different N-terminal peptides with otherwise similar sequence (**Xaa**GGAGRGLGK{acr}, where Xaa can be phenylalanine, alanine, tryptophan, tyrosine). When Xaa is F, abbreviated F_1 peptide; A, abbreviated A_1 peptide; W, abbreviated W_1 peptide; Y, abbreviated Y_1 peptide. The peptide is depicted as a chain of grey beads, with arginine "R" as the dark grey bead, and the N-terminal amino acid as a red bead. First, 10 µM ClpS2-StV1 protein with an HA tag is placed in solution to bind peptides in ExMre gels. The protein is briefly washed out, and the gels are stained with 67 nM anti-HA tag antibody conjugated to Alexa Fluor 488. Fluorescence of the gels then reports the relative binding of ClpS2-StV1 to the various N-terminal peptides via a population read-out using confocal imaging of the ExMre gels.

(B) ExM gels containing F_1 peptide, A_1 peptide, W_1 peptide and Y_1 peptide are separately made at 1 mM concentration cast in a gelation chamber, expanded and re-embedded to reach ~2.7X expansion factor, reaching a final concentration of ~50 µM peptide (each gel with a volume of 1.4 µL). ExMre gels were incubated with 10 µM ClpS2-StV1, washed once, and placed in a solution of 67 nM anti-HA tag antibody 488 in 30 µL solution. The antibody solution is removed right before imaging on a confocal microscope, with a 10 µm step size covering the entirety of the gel. A representative raw image of the fluorescence of the gels is depicted for each ExMre gel containing a different N-terminal peptide, taking the 20th slice of the Z-stack for each (~200 µm deep into the gel). In black: scale bar, 100 µm.

(C) Cross-section of the imaging performed in (B) for the different N-terminal peptides in ExMre gels after staining with ClpS2 St-V1 and anti-HA tag antibody 488. Raw intensity of the gels across the Z-stack were compared (line, mean; shaded area, standard deviation; n=3, with three different peptide aliquots, 3 different gelation mixtures).

(D) Average fluorescence intensity of the images obtained in (B) for the different N-terminal peptides in ExMre gels after staining with ClpS2 St-V1 and anti-HA tag antibody 488. Fluorescence intensity of ExMre gels in the different conditions was averaged across the whole volume imaged (black dots, individual experiments; colored bar, mean; error bar, standard deviation, n=3, with three different peptide aliquots, 3 different gelation mixtures).

(E) (i) F_1 peptide, with the F amino acid depicted as a red bead, in ExMre gels. ClpS2-StV1 staining with anti-HA tag antibody 488 is performed as in (A). (ii) A peptide with sequence GFGAGRGLGK{acr}, abbreviated G_1F_2 peptide, where the F amino acid is depicted in red. 1 round of in-gel Edman degradation degrades N-terminal glycine (G) in ExMre gels. Then, ClpS2-StV1 and anti-HA tag 488 antibody staining is performed as in (A).

(F) (i) A representative image of the fluorescence of the gels is depicted for each condition of in-gel Edman degradation for the F_1 peptide, taking the 14th slice of the Z-stack for each (~210 µm deep into the gel, with 15 µm z steps). Imaging performed on a confocal microscope (~375 µm thick Z-stack). Scale bar, 100 µm. (ii) The gels in different conditions are compared in fluorescence intensity throughout their cross-section for the F_1 peptide (line, mean; shaded area, standard deviation; n=2, different gelation solutions). (iii) Average fluorescence intensity of the ExMre gels in the different conditions for the F_1 peptide by averaging across the whole volume imaged (black dots, individual experiments; colored bar, mean; error bar, standard deviation, n=2, different gelation solutions).

(G) (i) As in Fi, but with G_1F_2 peptide, taking the 13th slice of the Z-stack for each (~195 µm deep into the gel, with 15 µm z steps). (ii) As in Fii, but with G_1F_2 peptide (line, mean; shaded area, standard deviation; n=3, different gelation solutions). (iii) As in Fiii, but with G_1F_2 peptide (black dots, individual experiments; colored bar, mean; error bar, standard deviation, n=3, different gelation solutions).

In-gel Edman degradation of multiple rounds

We performed assessment of in-gel Edman degradation over 3 cycles, using the LC/QToF-trypsinization assay analysis of embedded peptides, described above, and recording the various peptide species (peptides with up to 3 cleaved N-terminal amino acids, and PTC-peptide from the conjugation step at each round of Edman). Importantly, performing multiple rounds of in-gel Edman degradation provides a way to confirm cleavage of PTC-peptide in the cleavage step of the in-gel Edman degradation reaction, overcoming the mass spectrometry caveat in quantitating the yield of the cleavage reaction in a single round, as noted in **Supplementary Note 3**. In short, formic acid addition, and other sample processing steps that compromise interpretation of PTC-peptide due to simulating the effects of TFA cleavage, occur only at the final step of the procedure, within the instrument itself. Thus such confounds do not compromise the interpretation of PTC-peptide from second and third round of in-gel Edman degradation, which require a fresh N-terminus from a successful cleavage reaction within the gel itself, followed by a successful conjugation, to detect these species. Gel size was relatively constant throughout (**Fig. 5a**; see **Supplementary Table 6** for all values for **Fig. 5**), and the products yielded by each step of Edman degradation largely matched what was expected from the 1- and 2-round experiments above (**Fig. 5b**). In summary, conversion rate of peptide upon reaction with PITC was high each round, with 98-99% conversion (**Fig. 5bi**, **Fig. 5biii**, **Fig. 5bv**). Small amounts of some side products were observed (see **Source Data** for **Fig. 5b** for all raw chromatogram traces and mass spectra used in **Fig. 5**), as well as some PTC-peptide from earlier rounds in later rounds of Edman degradation (**Fig. 5biv**, **Fig. 5bvi**). These could be the result of reactions involving oxidative desulfuration of the PTC moiety, as reported in the literature[32,38], or it is possible that some of the novel ingredients used here could result in byproducts (e.g., DMSO and TFA[39,40]), further evidenced by different effects of TFA on gel size after DMSO treatment (**Fig. 3eii**; **Fig. 5a**) vs. without DMSO (**Fig. 2vii**). Perhaps using a different acid (e.g. glacial acetic acid or hydrochloric acid; as previously reported in literature; see **Supplementary Table 2**), temperature, and/or incubation time could further optimize conditions for the ExMre solid-phase conditions. Recent results have also demonstrated the possibility for base-induced Edman degradation[41].

Figure 5

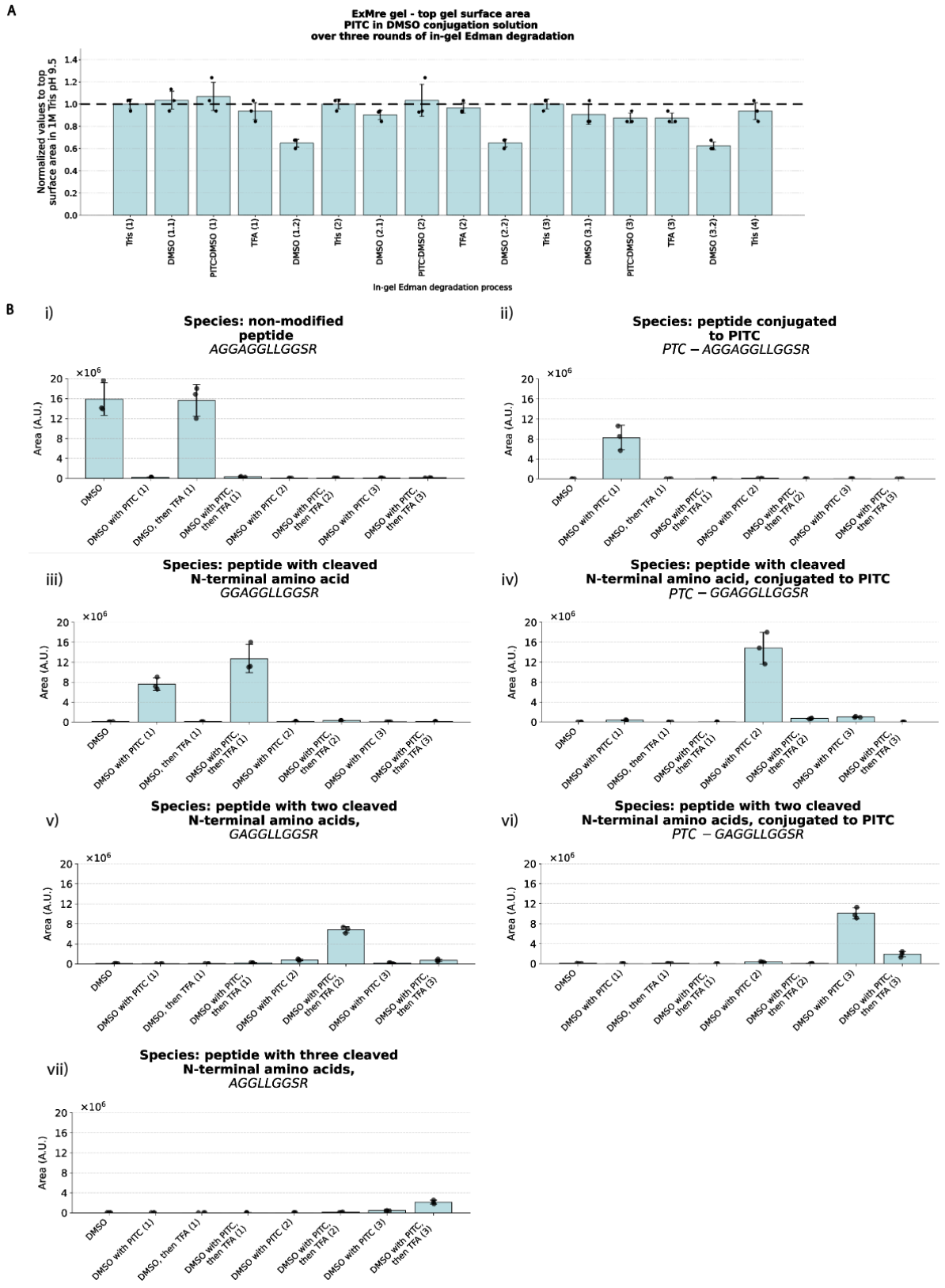


Figure 5 In-gel Edman degradation over multiple rounds using LC/QToF trypsinization assay

(A) Top/flat surface size of ExM and ExMre gels when placed in solvents used in Edman degradation. Normalized top/flat surface size to top/flat surface size in 1X PBS throughout 3 rounds of in-gel Edman degradation for ExMre gels with 1:1000 ratio PITC to DMSO. The round number is specified in parentheses (eg., "1"), and if solution is used at various steps of the same round, the first and second submersion of the gel is specified with an additional number (eg., "1.2"). (Thick dashed black line: top/flat surface size of the gel in 1M Tris pH 9.5 before Edman degradation, error bar: standard deviation, black dots, individual experiments, n=3, three separate gelation solutions.)

(B) Bar graphs representing the relative abundance (arbitrary units, a.u.) of different peptide ion species from the LC/QToF, obtained from the area under the curve (AUC) of the chromatogram based on the exact mass of the various species (Agilent MassHunter Qualitative Analysis software). The relative abundance of the ion species: (i) non-modified peptide (AGGAGGLLGGSR), (ii) peptide conjugated to PITC (*PTC*-AGGAGGLLGGSR), (iii) peptide with cleaved N-terminal amino acid (GGAGGLLGGSR), (iv) peptide with cleaved N-terminal amino acid conjugated to PITC (*PTC*-GGAGGLLGGSR), (v) peptide with two cleaved N-terminal amino acids (GAGGLLGGSR), (vi) peptide with two cleaved N-terminal amino acids conjugated to PITC (*PTC*-AGGLLGGSR), (vii) peptide with three cleaved N-terminal amino acids (AGGLLGGSR), are reported throughout the Edman degradation process in the gel in the various conditions. Conditions include: DMSO only, DMSO with PITC, TFA only, and sequential DMSO with PITC followed by TFA (black dots, individual experiments; blue bar, mean; error bar, standard deviation, n=3, three separate gelation solutions).

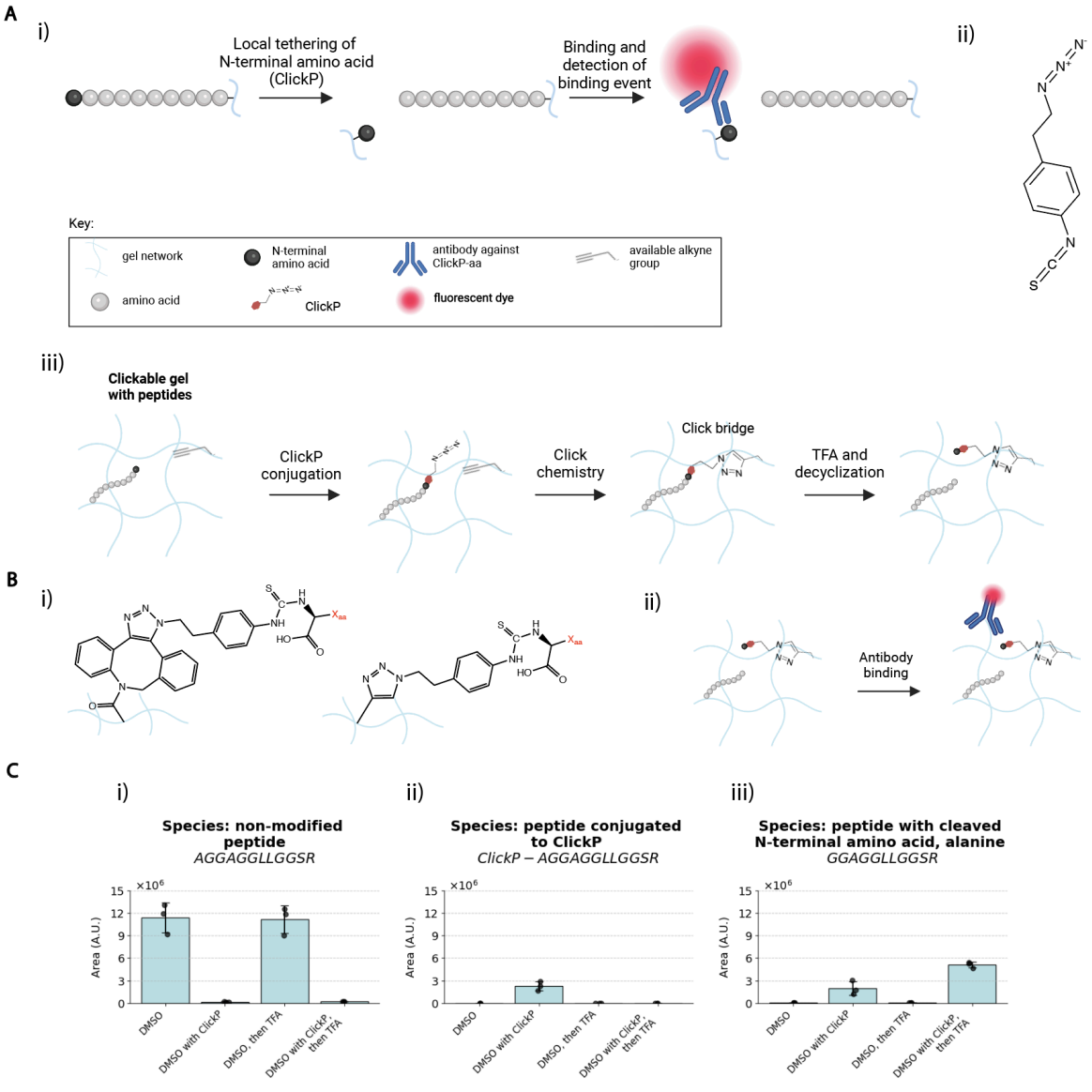
An Edman reagent that permits local tethering of an isolated N-terminal amino acid

One concern is that the binding of existing NAABs to an N-terminal amino acid of a peptide is modulated by the second amino acid in the chain (or other downstream amino acids), which could present both chemical and physical hindrance[14,25,26,42]. We thus, over the past several years[42–44], envisioned a complementary sequencing strategy, with a general workflow as depicted in **Figure 6ai**, wherein the N-terminal amino acid is first locally tethered to the polymer backbone before being removed from the peptide chain by Edman degradation, for subsequent read-out with an amino acid binder that was designed to bind an isolated amino acid. There would be no second amino acid to interfere with the binder interaction with the first amino acid, and thus perhaps an amino acid binder could have higher affinity and specificity than otherwise possible. As such, ClickP (as we named it), a bifunctional isothiocyanate derivative with an azide group (**Figure 6aii**), could in principle be used to isolate the N-terminal amino acid of a peptide for later isolated binder interrogation. There could be various ways of implementing this within the gel context, e.g., through initial conjugation of ClickP to the N-terminus of the peptide (i.e., the first step of Edman degradation), followed by local tethering of the N-terminal amino acid to the polymer backbone through click chemistry onto alkyne groups attached to the polymer backbone, followed by cleavage of the amide bond with TFA (**Figure 6aiii**). The amino acid would then remain bound to the polymer backbone, via click chemistry, in local proximity to its parent peptide. The antigen to be detected would simply be an isolated amino acid, attached to the ClickP group (see **Figure 6b**).

We here sought to validate whether a binder to such an amino acid could bind to its target in the gel; full realization of a ClickP-based *in situ* sequencing method would require significant further work. We designed binders to target amino acids that have been cleaved off the peptide chain after having been conjugated to ClickP. The ones here were created by Glyphic Biotechnologies (where they are commercially available) - a phenylalanine antibody (denoted F-antibody) and a valine antibody (denoted V-antibody), that target ClickP-tethered and decyclized amino acids (where the PTH-aa form is reverted back into PTC form, exposing the C-terminal carboxyl group, as done previously[45]). The K_d values for the F-antibody suggest a dissociation constant of ~75 pM and k_{off} of ~$4.5 * 10^{-5}$ s^{-1}, which corresponds to an average dwell time of ~6.2 hours (see **Supplementary Figure**

9**b**); the V-antibody was similar (see Supplementary Figure 9**a**). These dwell times would, unlike the previous NAABs discussed, support amplification strategies in an *in situ* context, which takes minutes to hours. ClickP showed good conjugation, with a conversion rate of ~99%, and more cleavage after TFA treatment than without, with the caveats acknowledged above regarding interpretation of PTC-amino acid cleavage with mass spec (Fig. 6**ci-iii;** see Supplementary Table 7 for all values from Fig. 6**c**).

ClickP-antigen for various amino acid side chains (comparing Gly, Val, Phe) carrying a biotin group were incubated with separate gels containing anchored streptavidin (Fig. 6**d;** chemical structures in Supplementary Figure 10). Subsequent antibody (F-antibody and V-antibody) binding to the antigens was monitored using bulk fluorescence read-out for both antibodies, as in Fig. 4. The V-antibody showed binding to the valine target over glycine and phenylalanine (Fig. 6**ei-iii**; see Supplementary Table 10 for all raw values for Fig. 6**e-f**). Similarly, the F-antibody showed binding to the phenylalanine target over glycine and valine (Fig. 6**fi-iii**). Thus, ClickP-amino acid antibodies can bind to their targets in the expansion gel.



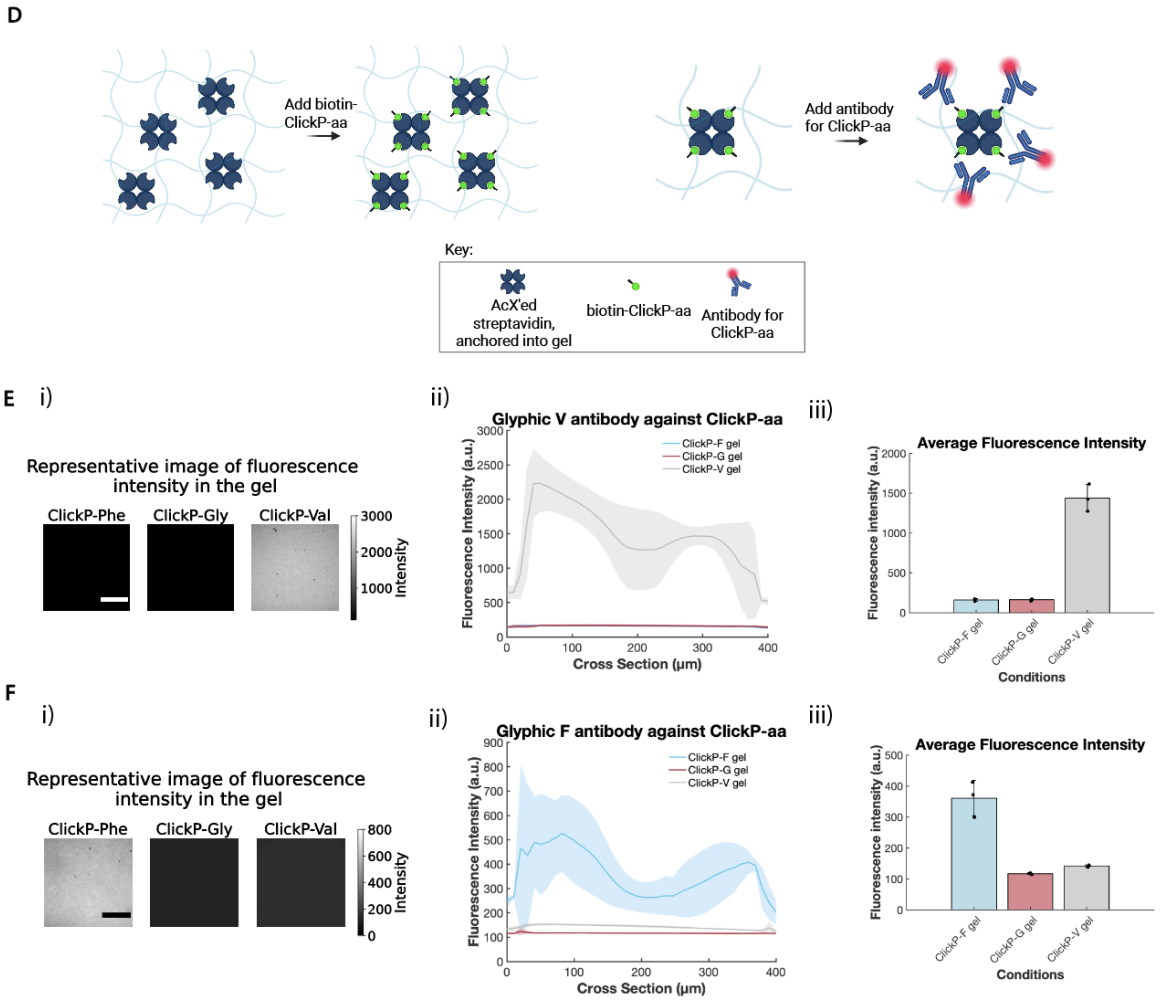


Figure 6 ClickP as Edman reagent, and ClickP-specific binders in ExMre gels.

(A) (i) 1-(2-azidoethyl)-4-isothiocyanatobenzene, abbreviated ClickP ITC derivative, is used for local tethering of the ClickP-conjugated N-terminal amino acid to the polymer network, followed by cleavage with TFA. Binders (eg., commercially available Glyphic antibodies) are used to detect the cleaved N-terminal amino acid bound to the polymer network. N-terminal amino acid local tethering followed by detection are performed iteratively to sequentially identify the amino acids of the chain. (ii) The structure of the ClickP ITC derivative. (iii) Peptides cast in an ExMre hydrogel conjugated with "clickable" groups (denoted: "Clickable gel", e.g. with alkyne groups for copper-catalyzed chemistry). The bifunctional molecule, with an azide and isothiocyanate group, first reacts to primary amines at the N-terminus of peptides via the isothiocyanate group. Then, the copper-catalyzed click chemistry forms a bridge between the ClickP-conjugated N-terminus (4-(2-azidoethyl)phenylthiocarbamoyl-peptide) via the azide group with the alkyne group in the gel. TFA is then used to cleave the N-terminal amino acid from the peptide, and the covalent click reaction locally tethering the amino acid to the surrounding hydrogel network, prevents its diffusion out of the gel. The N-terminal amino acid is decyclized from the thiohydantoin or thiozolinone intermediates, from post-cleavage, into the thiocarbamoyl derivative with covalently tethered N-terminus and an available C-terminus (carboxyl group).

(B) (i) The final antigen the Glyphic antibodies recognize. Either a reacted DBCO-azide click or a reacted alkyne-azide click group, with a given amino acid and its side chain exposed (amino acid side chain: "Xaa", in red text), including the free C-terminus of the amino acid. (ii)

Antibody binding can be performed after the N-terminal amino acid is tethered locally to the surrounding hydrogel network, cleaved from the peptide, and decyclized into the PTC form.

(C) Bar graphs representing the relative abundance (arbitrary units, a.u.) of different peptide ion species from the LC/QToF, obtained from the area under the curve (AUC) of the chromatogram based on the exact mass of the various species (Agilent MassHunter Qualitative Analysis software). The relative abundance of the ion species: (i) non-modified peptide (AGGAGGLLGGSR) , (ii) peptide conjugated to ClickP (4-(2-azidoethyl)phenylthiocarbamoyl-peptide, abbreviated: *ClickP*-AGGAGGLLGGSR), and (iii) peptide with cleaved N-terminal amino acid (GGAGGLLGGSR), is reported throughout the Edman degradation process in the gel in the various conditions (black dots, individual experiments; blue bar, mean; error bar, standard deviation, n=3, three separate gelation solutions).

(D) Schematic of the binding assay with biotin-ClickP-aa in ExMre gels with streptavidin. ExMre gels are cast with AcX-reacted streptavidin protein. Separate gels are made for each tested amino acid, by adding biotin-ClickP-amino acid (abbreviated: biotin-ClickP-aa), where the aa is either phenylalanine (Phe, or F), glycine (Gly or G), or valine (Val or V), to bind the streptavidin protein embedded in the gel (left). After the binding of biotin-ClickP-aa to the gels, which exposes the final antigen that Glyphic antibodies recognize, the GFP-conjugated Glyphic antibody against ClickP-V (abbreviated: Glyphic V) and RFP-conjugated Glyphic antibody against ClickP-F (abbreviated: Glyphic F), are added to the gels to assess specificity of binding towards the various ClickP-aa (right).

(E) Binding assay with biotin-ClickP-aa (aa: G, F, V) in ExMre gels with streptavidin using Glyphic V. First, streptavidin is embedded into ExM gels at a concentration of 0.1 mg/mL and after expansion and re-embedding reach a concentration of ~5 μg/mL. Biotin-ClickP-aa is added at a concentration of ~17 μg/mL. Then, Glyphic V is added at concentration of 5 μg/mL. (i) A representative raw image of the fluorescence of the gels is depicted for the gels in different conditions, comparing binding of Glyphic V against biotin-ClickP-F, biotin-ClickP-V and biotin-ClickP-G ExMre gels obtained from a confocal microscope, showing a slice 90 μm deep into the gel. Scale bar in white for the ClickP-Phe image (left): 100 μm. (ii) The ExMre gels in different conditions are compared in fluorescence intensity throughout the Z-stack (line, mean; shaded area, standard deviation; n=3 different gelation solutions). (iii) Average fluorescence intensity of ExMre gels in the different conditions by averaging across the whole volume imaged (black dots, individual experiments; colored bar, mean; error bar, standard deviation, n=3 different gelation solutions).

(F) As in E, but using Glyphic F (black dots, individual experiments; colored bar, mean; error bar, standard deviation, n=3 different gelation solutions).

A theoretical assessment of *in situ* protein sequencing

Having established core chemical principles for *in situ* protein sequencing, but with unknowns remaining - particularly when it comes to the affinity and specificity of binders to be invented in the future - we next set out to perform a theoretical assessment of *in situ* protein sequencing, that could guide future developments in the field (details on our strategy and results can be found in **Supplementary Note 10**). In short, the percentage of residues of a protein that remained accessible for read-out through in-gel Edman degradation and N-terminal binding depended on the extent and specificities of fixation, anchoring and digestion chemistries. However, the model suggests that even partial realization of an *in situ* protein sequencing platform could be extremely useful: under the assumptions of our model, with a subset of only 10 amino acid binders, with medium specificity (as defined by our model), 12 rounds of sequencing, with 10% conjugation error, and 30% cleavage error, could correctly identify ~90% of the proteins in the mycoplasma proteome. The human proteome, similarly, can be interrogated quite successfully even with a partially efficacious sequencing scheme.

Conclusion

We here derive principles of *in situ* protein sequencing. Building from the idea that expansion microscopy (ExM) could decrowd proteins from each other for sequencing, as previously shown for nucleic acids, we demonstrate several chemistries at the proof-of-concept level, including forms of Edman degradation compatible with the ExM hydrogel, and detections of N-terminal amino acids with NAABs. Currently, there are not enough high quality NAABs to support single molecule *in situ* protein sequencing in 3D specimens, but making better ones is an active area of research by many groups. In addition, commercialization efforts are often required to mature a sequencing chemistry (even DNA sequencing required pushes by companies like Illumina to mature to the point of everyday utility in biology). As a result, we focused on bulk measurements of Edman conjugation and degradation in the gel, e.g. with mass spec and bulk gel fluorescence, to evaluate the fundamental chemistries required for *in situ* protein sequencing to take place. Our current results offer a platform for facilitating further *in situ* protein sequencing technology development, e.g. for testing of future NAAB designs, or validating chemical probes for covalent amino acid sidechain labeling in the gel. We showed up to 3 rounds of Edman degradation on peptides in a gel, showing the iterative nature of Edman degradation could be implemented in expansion hydrogels. We also explored whether we could bind isolated amino acids, as might be possible with tethering of them to the gel before cleavage, to reduce steric effects from the downstream amino acids of the peptide.

Several chemical optimizations and validation were required to derive these principles of *in situ* protein sequencing. To enable N-terminal amino acid degradation from peptides embedded in ExMre gels, we screened solvents conventionally employed in Edman degradation chemistry, finding that traditional Edman conjugation solutions led to substantial gel shrinkage and opacity, whereas solvents such as DMSO did not. This preservation was essential for successful Edman degradation to take place. Further work will be required to optimize the yield beyond its current ~70% level, which we speculate to be the result of solvents and reagent interactions with the gel, or potentially interactions with the surrounding gel network during cyclization, unique challenges compared to other Edman degradation platforms.

Existing NAABs have a rate of dissociation that is too fast (~seconds) for fluorescence-based imaging of single molecules via signal amplification schemes (requiring hours), and the second amino acid of the peptide chain can influence binding to the first[25]. Prior work suggests that single amino acid side chains can be determinants of an epitope[46–50]. For instance, the armadillo modular protein, when using a constant peptide for K_d ~nM baseline affinity, can exhibit single side chain selectivity[51]. Nevertheless, the relatively small epitope size of a single amino acid side chain, compared to the 4-12 amino acids targeted by many antibodies[52], and the chemical and physical proximity of subsequent amino acid residues in the peptide chain to the N-terminal amino acid, may make NAAB design challenging. One could imagine using PITC or FITC N-terminally conjugated peptides to raise the

baseline affinity of the binders (or even an alternative ITC molecule that is both highly compatible with the hydrogel and provides additional groups for binder attachment), while maintaining specificity for the side chain. Binder design may be significantly enhanced by progress in computational protein design, particularly through AI-driven approaches for modeling and designing protein interactions[53–55].

We proposed an alternative strategy to overcome the limitations of NAABs, via our heterobifunctional ClickP molecule. This molecule features an isothiocyanate moiety that binds to the N-terminal amine of peptides, along with a click group that can locally tether the amino acid conjugated to the ClickP molecule, e.g. to the hydrogel. We showed this molecule to be compatible with in-gel Edman degradation. In addition, commercially available antibodies produced against these specific antigens had strong affinity, and we demonstrated their specificity to their expected target within the ExMre gels. Local tethering of an amino acid away from its parent peptide chain would prevent influence from subsequent amino acids in the peptide chain and, thus, potentially improve specificity. Future work could include development of the entire workflow combining in-gel Edman degradation with local tethering of N-terminal amino acid with ClickP for a complete cycle of degradation, tethering, and detection of the amino acid from a parent peptide within the polymer network. Other modifications of PITC, e.g. one that allows for DNA oligo conjugation[31], have been put forth as well.

The ExM protocol itself could be improved in terms of expansion factor, for better decrowding of proteins from each other. Indeed, ExM preserves key aspects of protein shape, enabling imaging at 1 nanometer resolution[12]; future expansion protocols that can expand 100x or 1000x or more, that preserve this high fidelity of expansion, would be particularly valuable. It would also be beneficial to show that such protocols can be applied to proteins in a densely packed, heterogeneous biological sample, such as in a cell or tissue.

Our theoretical assessment of *in situ* protein sequencing, under the assumptions of our model, reveals potential directions for improvement and optimization. Our modeling may guide experimental strategies, such as refining selection parameters and reaction efficiencies to improve peptide fragment retention for subsequent protein identification. Importantly, our model can be modularly modified to test alternative conditions and constraints, and the underlying assumptions can be updated as they become better constrained by empirical data.

Methods

Software for figure making

Figure 1**a-c;** Figure 3**a, f, hi;** Figure 4**a, e;** Figure 6**a, b, d;** Supplementary Figure 13**a-b** were made with BioRender. Figure 2**a, b;** Figure 6**a, b;** Supplementary Figure 8**ai-hi;** Supplementary Figure 10**;** Supplementary Figure 11 were made with ChemDraw.

Synthetic peptide designs

All peptides were ordered from AAPPTEC in aliquots with > 95% purity. 5 mM stock solutions of each peptide were made by diluting lyophilized peptides with UltraPure distilled water, DNase/RNase free (Invitrogen). See Supplementary Table 3 for exact mass and chemical formulae of all peptides used in the paper, and Supplementary Figure 11 for structures of all peptides used in the paper.

Edman degradation readouts with liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry (LC-ESI-QToF MS, abbreviated LC/QToF) (results from Figure 3**b-d**, Figure 5**b,** Figure 6**c**) were performed with a peptide that we denoted "A15-peptide" (which stands for AGGAGGLLGGSRGGK{acr}; K{acr} is an abbreviation for lysine modified with an acryloyl functional group). A peptide that we denoted "A9-peptide," AGGAGK{acr}GLR, was spiked into samples to account for any ionization efficiency fluctuations.

In-gel Edman degradation monitored through bulk fluorescence in Figure 3**g** was performed with a peptide that we denoted "K{N_3}15-peptide," K{N_3}GGAGGLLGGSRGGK{acr}, where K{N_3} is an abbreviation for lysine modified with an azide functional group.

Phenylthiohydantoin-phenylalanine (PTH-F) detection was performed with "F_1" peptide, FGGAGRGLGK{acr} (Figure 3**h**).

ClpS2 St-V1 bulk fluorescence read-outs comparing different N-terminal amino acids (Fig. 4**a**) in the ExMre gels were performed with "F_1" peptide, "W_1" peptide, WGGAGRGLGK{acr}, "Y_1" peptide, YGGAGRGLGK{acr}, and "A_1" peptide, AGGAGRGLGK{acr}. ClpS2 St-V1 bulk fluorescence read-outs combined with Edman degradation chemistry (Fig. 4**f-g**) were performed with "F_1" peptide and "G_1F_2" peptide, GFGAGRGLGK{acr}.

Hemagglutinin-tag (HA) peptide with a K{acr} group (or YPYDVPDYAK{acr}) was used in 9% acrylamide gels in Supplementary Figure 7**b**.

Amino acid oxidation analysis and results (Supplementary Figure 8**a-h**) in ExM gels were performed with XaaGGAGRGLGK{acr} peptides, where Xaa was one of: M, C, W, Y, F, P, H, R. Oxidation analysis, as a result of free-radical polymerization, was performed after trypsin cleavage of the peptides from the gel (or in solution with/without ammonium persulfate (APS; Sigma A3678-25G; we include the catalog number the first time we introduce a chemical in the Methods section) and tetramethylethylenediamine (TEMED; Sigma T7024-25ML)) followed by injection in LC/QToF. Post-translational oxidation modifications, exact mass and chemical formulae, are detailed in Supplementary Table 3.

Making acrylamide, expansion microscopy (ExM), and ExM re-embedded (ExMre) gels

All washes were performed at room temperature (RT), unless otherwise specified (for all sections).

Acrylamide gel: making the empty gel gelation mixture (i.e. no peptide/protein)

The empty gel gelation mixture (used for gel size change measurements in Supplementary Figure 2) was prepared to have a final concentration of 9% acrylamide (w/v) (A9099-25G), 10 mM N,N'-methylenebisacrylamide (abbreviated: BIS; Sigma M7279-25G), 1X PBS (from diluting

ThermoFisher 70011 10X PBS pH 7.4), in UltraPure water (Invitrogen 10977015; below, whenever water is mentioned, we use this UltraPure water unless otherwise indicated) to a total volume of 100 µL (see **Table 1: Acrylamide gelation solution)**. Subsequently, pre-chilled 10% (v/v) TEMED and 10% (w/v) APS stock solutions were added to the mix (5 µL each) on ice. Casting the gel was performed as described in section **Casting the gel**.

Acrylamide gelation solution	**Concentration**	**Final concentration**	**Add**
Acrylamide	50% (w/v)	9% (w/v)	20 µL
BIS	130 mM	9.1 mM	7.7 µL
PBS	10x	0.9x	10 µL
Water			62.3 µL
APS	10% (w/v)	0.45% (w/v)	5 µL
TEMED	10% (v/v)	0.45% (v/v)	5 µL
Total volume			110 µL

Table 1: Acrylamide gelation solution

Acrylamide gel: making the gelation mixture with peptide

The peptide-containing gelation mixture (used for **Supplementary Figure 7a-b**) was prepared to have a final concentration of 9% acrylamide (w/v), 10 mM BIS, 1X PBS, 20 µL of peptide (using a stock of 25 mM for **Supplementary Figure 7a**, and using a stock of 5 mM for **Supplementary Figure 7b**), and water to a total volume of 100 µL (final peptide concentration of 5 mM for **Supplementary Figure 7a**, and final concentration of 1 mM for **Supplementary Figure 7b**). Subsequently, pre-chilled 10% (v/v) TEMED and 10% (w/v) APS stock solutions were added to the mix (5 µL each) on ice. Casting the gel was performed as described in section **Casting the gel**.

ExM gel: making the empty gel gelation mixture

The empty gel gelation mixture (used for gel size change measurements in **Fig. 2c** and in **Supplementary Figure 2**) was prepared with 60 µL StockX monomer solution (see **Table 2: StockX monomer solution**), 10 µL 130 mM BIS, and water to a total volume of 90 µL. Subsequently, pre-chilled 10% TEMED (v/v) and 10% APS (w/v) stock solutions were added to the mix (5 µL each) on ice (see **Table 3: ExM empty gel and peptide-containing gelation solutions**). StockX monomer solution was made as follows as in the previously described protocol[7]. Aliquots of StockX monomer solution were kept at -20 °C. Casting the gel was performed as described in section **Casting the gel**.

StockX monomer solution	**Concentration**	**Amount (mL)**
Sodium acrylate (Combi-Blocks QC-1489)	38% w/v	2.25
Acrylamide	50% w/v	0.5
BIS	2% w/v	0.75
Sodium chloride (J60434.AK)	5 M	4
10X PBS	10X	1

Water		0.9
Total		**9.4**

Table 2: StockX monomer solution

ExM gel: making the gelation mixture with peptide

The peptide-containing gelation mix (used for **Fig. 3b-e,g,hii; 4b-d,f-g; 5a-b,f-g; 6c,e-f**) was prepared with 20 µL 5 mM peptide of the given sequence, 60 µL StockX, 10 µL 130 mM BIS, with water to a total volume of 90 µL. Subsequently, pre-chilled 10% TEMED v/v and 10% APS w/v stock solutions were added to the mix (5 µL each) on ice (see **Table 3: ExM empty gel and peptide-containing gelation solutions**). Casting the gel was performed as described in section **Casting the gel**.

Reagent	Concentration	Final concentration	Add
StockX monomer solution			60 µL
Additional BIS (not including concentration of BIS in the StockX solution)	130 mM (2% (w/v))	13 mM	10 µL
Peptide (replace with water for the empty gel gelation solution)	5 mM	1 mM	20 µL
APS	10% (w/v)	0.5% (w/v)	5 µL
TEMED	10% (v/v)	0.5% (v/v)	5 µL
Total volume			100 µL

Table 3: ExM empty gel and peptide-containing gelation solutions

Casting the gel

After gentle mixing, the mixture (taking ~70 µL of solution) was pipetted into a gelation chamber, created using a glass slide (VWR, #48300-026), a 22 x 22 mm #1.5 coverslip (VWR, #48366-227), and one or two layers of parafilm (Uline, S-25929) as spacers, as described previously [56]. **Fig. 2c; 3b-e,g; 5a-b**, **6c; Supplementary Figure 1a-h; Supplementary Figure 2a-m; Supplementary Figure 3a-b; Supplementary Figure 4a-c; Supplementary Figure 7a-b; Supplementary Figure 8a-h**, were made with 2 parafilm-thick (~260 µm) chambers ("thick" gel), whereas **Figure 3hii**; **4b-d,f-g; 6e-f; Supplementary Figure 6a-b** were made with 1 parafilm-thick (~130 µm) chambers ("thin" gel). By spacing the parafilm stacks 10 mm apart on the glass slide, the final gelation chamber was 10 mm long, 22 mm wide, and with height determined by the parafilm thickness. The gelation chamber was placed within a Tupperware box (with two holes drilled to insert a nozzle for nitrogen purging, and tape to close the holes shut after purging) with a damp paper towel, purged with nitrogen for 5 minutes to remove oxygen, and then incubated at 37 °C for 1.5 hr for gelation. Upon completion of the gelation process, the chamber was disassembled, and the gel was cut into 10 mm x 5 mm rectangles before being transferred to 1X PBS. The gels were washed 3 times in 1X PBS, 5 minutes each.

Expanding the ExM gels

ExM gels, cut to 10 mm x 5 mm rectangles after gelation, were placed on a glass slide inside a 4 well plate in 1X PBS. 1X PBS was removed and the gels were expanded at RT by washing 3 x 20 min with 5 mL water until they reached ~3.5x expansion. Expanded gels were then cut down to either 5 mm x 5

mm (solvent testing on ExM gels in **Fig. 2c**, **Supplementary Figure 2a-m**) or used for further re-embedding in **Fig. 2c; 3b-e,g,hii; 4b-d,f-g; 5a-b, 6c,e-f**.

Re-embedded ExM (ExMre) gels

Expanded ExM gels (10 mm x 5 mm before expansion, and reaching ~35 mm x ~17 mm in width and length after expansion) were re-embedded following a similar strategy as a previously published protocol [19]. Any remaining water was removed from the 4 well plate, and the gels were submerged in a 5 mL solution with 3% acrylamide (w/v), 0.15% BIS (w/v) obtained from a 1:13 dilution of acrylamide/BIS 19:1, 40% (w/v) solution (AM9024; Fisher Scientific), 5 mM Tris pH 8, 0.075% APS (w/v), 0.075% TEMED (v/v) for 30 min at RT on a shaker (50 rpm). The re-embedding solution was removed from the 4-well plate. A sandwich of 22 x 22 mm #1.5 coverslips (VWR, #48366-227) glued together were placed on either side of the gel. The sandwich consisted of four coverslips for gels with a height of ~700 µm (i.e., the same height as expanded ExM gels made with 2 parafilm thick chambers): in **Fig. 2c; 3b-e,g; 5a-b**; **6c; Supplementary Figure 1a-f,h; Supplementary Figure 2a-m; Supplementary Figure 3a-b; Supplementary Figure 4a-c**. The sandwich consisted of two coverslips with a height of ~350 µm (i.e., the same height as expanded ExM gels made with 1 parafilm thick chambers): in **Fig. 3hii**; **4b-d,f-g; 6e-f; Supplementary Figure 6a-b**. Then, for all gels, a 50 mm x 24 mm #2 coverslip (VWR, #48382-136) was gently placed on the top of the gel. Gel chambers were placed inside a closed Tupperware box with a damp paper towel (as in section **Casting the gel**), which was purged with nitrogen for 10 min. Gel chambers were then incubated for 1.5 hr at 37 °C. Expanded re-embedded (ExMre) gels were then cut down to 10 mm x 5 mm (for **Supplementary Figure 3b**) or 5 mm x 5 mm (for **Fig. 2c, 3b-e,g,hii, 5a-b, 6c; Supplementary Figure 1a-f,h; Supplementary Figure 2a-m; Supplementary Figure 3a; Supplementary Figure 4a-c; Supplementary Figure 6a-b**) or 2 mm x 2 mm size gels (for **Figure 4b-d,f-g, 6e-f; Supplementary Figure 7a-b**) for downstream steps.

Gel flat/top side surface size tests

General solvent tests on gel flat/top side surface size

Three separate batches of ExM gels (in expanded state) and ExMre gels, without embedded peptide, were made, cut into ~5 mm x 5 mm squares (~700 µm thick) for solvent testing in **Figure 2c**. Similarly, 9% acrylamide, ExM (in expanded state) and ExMre gels, without embedded peptide, were cut into ~5 mm x 5 mm squares (~700 µm thick) for solvent testing in **Supplementary Figure 1** (gel images) and **Supplementary Figure 2** (gel size changes). All gels were washed 3 times 10 min washes in 1X PBS. The exact dimensions of each gel were recorded after the 1X PBS washes by measuring the gel flat/top side surface size on a glass slide using a ruler. In cases where the gel tore, the edges were carefully realigned prior to measurement. When the gel folded into a solid form after exposure to the solvents described below and could not be unfolded, measurements were taken from the accessible edges by rotating the vial to obtain the most representative size.

Glass vials (Chemglass, CG-4912-01) were filled with 300 µL of the following organic solvents: trifluoroacetic acid (TFA) (Sigma, T6508-100ML), 1X PBS, water, phenylisothiocyanate (PITC; Sigma, 317861), acetonitrile (ACN) (Sigma Aldrich, 271004), 1:1 ratio pyridine to water (Millipore Sigma, 270407) (all ratios throughout are of volumes added, unless otherwise indicated), dimethylsulfoxide (DMSO) (Millipore Sigma, 276855), formamide (Fisher Scientific, AM9342), 1M Tris pH 8 (AM9856), 1M Tris pH 9.5 (J62084.K2), 0.1 M sodium bicarbonate pH 8.5 (J60408) (**Fig. 2cii-vii**, and **Supplementary Figure 1bi-ii, ci-ii, di-ii, ei-ii, f**, **h** and **Supplementary Figure 2a-d**). For other experiments, the glass vials were filled with 270 µL of the following organic solvents: ACN, DMSO, formamide, pyridine, 1:1 pyridine:water, triethylamine (Millipore Sigma, 471283), dimethylformamide (DMF) (Millipore Sigma, 270547), 1:1 ratio DMSO to water (**Supplementary Figure 1a, biii, ciii, diii, eiii**, and **Supplementary Figure 2e-l**). The glass vials were filled with 231 µL 0.1 M sodium bicarbonate pH 8.5 for **Supplementary Figure 2m**. Gels were submerged in the solvent or buffer on a heat block at 50 °C for 30 min. For **Fig. 2cii-vii**, and **Supplementary Figure**

1bii, cii, dii, eii, the solutions were then removed, and the flat/top side surface size of each gel was measured (from the bottom, through the glass vial). Then, PITC or ClickP (Enamine, EN300-37440925) at 1:1000 ratio PITC:solvent or ClickP:solvent, respectively, or fluorescein isothiocyanate (FITC) 'Isomer 1' (ThermoFisher, F1906) at a final concentration of 5.9 mM FITC (23:77 of DMSO:0.1 M sodium bicarbonate pH 8.5), was added to fresh solvent or buffer. This solution was added into each glass vial to submerge the gels at 50 °C for 30 min, followed by gel flat/top side surface size measurement again (from the bottom, through the glass vial). For **Supplementary Figure 2e-l**, PITC was added at 1:9 ratio PITC:solvent in the same original solvent (30 µL added to the same 270 µL solvent), and for **Supplementary Figure 2m** FITC in the same original buffer at a final concentration of 5.9 mM FITC (23:77 ratio DMSO:0.1 M sodium bicarbonate pH 8.5; 69 µL of 10 mg/mL FITC in DMSO added to 231 µL buffer) and incubated for 30 min at 50 °C. The gel size was measured again after the solution was removed (from the bottom, through the glass vial).

Surface flat/top side size during Edman degradation chemistry

ExMre gels, with embedded peptide, were cut into ~5 mm x 5 mm squares (~700 µm thick) after 3x 10 min 1 mL washes in 1M Tris pH 9.5 at RT in 1.5 mL Eppendorf tubes. The exact dimensions of each gel flat/top side surface size throughout the in-gel Edman degradation were recorded for **Fig. 3ei-iii** and **Fig. 5a**. After the 1M Tris pH 9.5 washes, gels were placed on a glass slide and the gel surface size was measured on the glass slide using a ruler. In all the next steps, the gels were submerged in 300 µL of solution and a heat block at 50 °C in glass vials. Gels were transferred to glass vials, washed 2 times with solvent (eg. DMSO or formamide) 5 min each, and gel flat/top side surface size was measured again after removal of the solution from the second wash (measured from the bottom of the glass vial). Then, PITC to solvent (1:1000 ratio, e.g., DMSO or formamide) was added to the gel for 1 hour before removal, followed by the next measurement. The gel was then washed 3 times with solvent (eg. DMSO or formamide) 5 min each before TFA was added for 30 min at 50 °C. After discarding the TFA, the gels were measured again (from the bottom of the glass vial). The gels were washed 2 times with solvent (e.g., DMSO or formamide) 5 min each, and the gel flat/top side surface size was measured again after removal of the last wash. Then, the gels were washed 3 times with Tris pH 8 (**Fig. 3ei-iii**), 10 min each at RT, measured, and transferred to 1.5 mL plastic Eppendorf tubes. For **Fig. 5a**, the gels were washed 3 times with 1M Tris buffer pH 9.5, 10 min each at RT, measured, before the next cycle of in-gel Edman degradation. During the two subsequent rounds of in-gel Edman degradation for **Fig. 5a** the flat/top side surface size was similarly measured.

In-gel Edman degradation

In-gel Edman degradation was performed as outlined in the schematic in **Fig. 3a**. First, a heating block in a chemical hood was preheated to 50 °C. Gels were washed in 1.5 mL Eppendorf tubes, 3 times 5 min with 1 mL of 1M Tris pH 9.5 at RT. The gel samples were carefully transferred into glass vials with solvent on the heat block. In all the next steps, the gels were submerged in 300 µL of solution in glass vials that were placed on a 50 °C heat block. Samples were washed with solvent or buffer (e.g., DMSO for **Fig. 3c**, formamide for **Fig. 3b**, 0.1 M sodium bicarbonate pH 8.5 for **Fig. 3d**) two times, 5 minutes each. Samples were then incubated with PITC to solvent (1:1000 ratio) for a duration of 1 hour (**Fig. 4f-g**, **Supplementary Figure 3b** performed with 1:9 of PITC:solvent; **Supplementary Figure 4a** with 1:100 ratio of PITC:solvent; **Supplementary Figure 4c** with 1:10,000 ratio of PITC:solvent; **Fig. 6c** performed with 1:1000 ratio ClickP:DMSO). For FITC conjugation (**Fig. 3d**), 5.9 mM FITC in 23:77 of DMSO:0.1 M sodium bicarbonate pH 8.5 was added to the samples for a duration of 2 hours, in the dark. Following this, samples were washed 3 times 5 minutes in solvent or buffer (except **Supplementary Figure 3b** which was performed with 1 solvent wash; **Fig. 4f-g** was performed with 2 solvent washes). After removal of the solvent, gels were submerged in TFA for 30 min. TFA was removed and samples were washed with solvent twice, 5 minutes each (this step was omitted for PTH detection in **Fig. 3h**, see **Edman degradation and PTH detection** methods section). Glass vials were discarded, and gels were placed back into Eppendorf

tubes at RT. If only 1 in-gel Edman degradation round was performed, as in **Fig. 3b-d, g**; **Fig. 4f-g**; **Fig. 6c,** the tubes were washed 3 times 10 min each, with 1 mL of 1M Tris buffer pH 8 at RT before trypsinization. If another round of in-gel Edman degradation was performed, as in **Fig. 5a-b**, the tubes were filled with 1ml of 1M Tris buffer (pH 9.5), and washed 3 times 10 min each at RT, and the pH was checked by immersing pH strips in the buffer solution after the last wash. (Other gels that served as controls with 1 or 2 rounds of in-gel Edman degradation during the multiround experiment were washed 3 times 10 min with 1 mL of 1M Tris buffer pH 8 and stored at 4 °C.) After the last round was performed, the tubes were washed 3 times 10 min with 1 mL of 1M Tris buffer pH 8 at RT before trypsinization.

LC/QToF analysis of synthetic peptides in the gel

Gel preparation

Edman degradation read-outs with LC/QToF were performed with A15-peptide in ExMre gels. Each gel size was 5 mm x 5 mm x ~700 µm after re-embedding (see section: **Methods: Re-embedded ExM (ExMre) gels**) .

Trypsinization of peptides from the gels

Peptides were trypsinized from the gels after in-gel Edman degradation (see **Fig. 3a**; right hand side). After the 3 times 10 min 1 mL 1M Tris buffer (pH 8) washes, they were submerged in 50 µL of 20 µg/mL trypsin (New England Biolabs, #P8101S) in 50 mM Tris-HCl pH 8 in the bottom of the 1.5 mL Eppendorf tube (ensuring they were completely submerged in trypsin solution). The gels were then incubated overnight (~16 hours) at 37 °C. The supernatant (~50 µL) was placed in a mass-spectrometry vial (ThermoFisher, 6PSV9-1P), with the glass insert (ThermoFisher, 6PME03C1SP) and screwed shut (ThermoFisher, 6PSC9ST101X).

The A9-peptide was spiked into all the samples, during trypsinization, at a 5 µM final concentration (the A9-peptide was added along with the 20 µg/mL trypsin for the overnight, ~16 hours, incubation at 37 °C) to monitor any fluctuations in ionization efficiency of the samples (the abundance of A9-peptide from experiments in **Fig. 3b-d**; **Fig. 5b**; **Fig. 6c** were plotted in **Supplementary Figure 12**).

Mass spectrometry (LC/QToF)

Samples were brought to the Department of Chemistry Instrumentation Facility (DCIF) at MIT for injection into the LC/QToF. This instrument was a high-resolution Agilent 6545 mass spectrometer coupled to an Agilent Infinity 1260 LC system running a Jet Stream ESI source (mass accuracy of 1-3 ppm using real-time calibration, with a mass resolving power of 45,000 (FWHM) at m/z of 2722; measurable m/z range from 50 to 10,000). The column used was a reversed-phase ZORBAX Eclipse AAA, 3.0 x 150 mm, 3.5 µm, C18 (Agilent part number: 961400-302). For **Supplementary Figure 3b**, a reverse-phase ZORBAX StableBond 300 C3, 2.1 x 150 mm, 5 µm column was used (Agilent part number: 883750-909). Before each run, the instrument was positively tuned, and 10 µL blank water washes were used to wash the column twice. 4 µL of the trypsinized supernatant was injected into the mass spectrometer for ExMre gels with 5 x 5 mm top/flat surface size in **Fig. 3b-d**; **Fig. 5b**; **Fig. 6c**; **Supplementary Figure 3a**; **Supplementary Figure 4a-c;** and for the control curve in **Supplementary Figure 5** (with an estimated concentration of ~50 µM peptide in ExMre gels, ~50 pmol (~50 ng) of trypsinized peptides were injected). 2 µL of the trypsinized supernatant was injected into the mass spectrometer for ExMre gels with 10 x 5 mm top/flat surface size in **Supplementary Figure 3b**. 1 µL of the trypsinized supernatant was injected into the mass spectrometer for non-expanded ExM gels with 10 x 5 mm top/flat surface size for oxidation analysis in **Supplementary Figure 8**. 10 µL of the supernatant was injected into the mass spectrometer for PTH-F detection in **Fig. 3hii**. The samples were injected using **LC Method**, which details the mobile phase during the ~23 min estimated run time.

LC Method: chromatographic separation method (LC/QToF):

Time (min)	Solvent A (%)	Solvent B (%)	Flow (mL/min)	Max Pressure Limit (bar)
0.00	99.5	0.5	0.300	400.00
1.50	99.5	0.5	0.300	400.00
15.00	40.0	60.0	0.300	400.00
18.00	5.0	95.0	0.300	400.00
19.00	5.0	95.0	0.300	400.00
20.00	99.0	1.0	0.300	400.00

From 0.00 to 1.50 min, the eluate was sent to waste and not to the Q-ToF. The Q-ToF was turned on at 1.50 min. Solvent A: 0.1% formic acid in water. Solvent B: 0.1% formic acid in ACN.

Analysis of LC/QToF data

Bar graphs representing the relative abundance (arbitrary units, a.u.) of different peptide ion species from the LC/QToF were obtained from the area under the curve (AUC) of the chromatogram based on the exact mass of the various species (Agilent MassHunter Qualitative Analysis software). The relative abundance of the ion species was reported throughout the Edman degradation process in the gel in the various conditions.

Specifically, to obtain the AUC in an automated fashion (**Fig. 3b-d**, **Fig. 5b**, **Fig. 6c**, **Supplementary Figure 3a-b**, **Supplementary Figure 4a-c**, **Supplementary Figure 5** and **Supplementary Figure 12**), Agilent MassHunter Qualitative Analysis software was opened with the files to analyze. Using the "Compound Analysis" tab, with "Find by Formula", the given formula and exact mass (see **Supplementary Table 3**) was used to extract the information about each species of interest from the total ion chromatograms (TICs). They were saved as an excel spreadsheet for downstream analysis and plotting. Plots were then made in python (see **Code**).
The results from this search were manually inspected to ensure correct peak identification. The raw TIC, and the corresponding mass spectra of the species of interest were exported as metafiles ('.emf') (available in **Source Data**, with an arrow pointing towards retention time and peak of interest).

For **Figure 3hii** and **Supplementary Figure 8a-h**, the AUC for the species of interest were obtained by manual integration by right clicking the chromatogram and selecting "Extract chromatogram". In the pop-up window, the mass of the $[M+H]^+$ species was inputted, and the detected peak(s) from this extraction were integrated into a single value (the integration considers an error of ± 20 ppm (Δm) : Δm = (ppm error × exact mass) / 1,000,000 for detection).

In-gel Edman degradation with PTH detection

F_1 peptide (3 separate vials received from AAPPTEC), were cast in ExM gels with a peptide concentration of 1mM in the gelation solution, and ~130 µm (i.e., 1 parafilm) thick gelation chamber (see section **Methods: ExM gel: making the gelation mixture with peptide** and **Re-embedded ExM (ExMre) gels**). After they were re-embedded into ExMre gels, they were washed 3x with 1X PBS 10 min, and cut into 5 x 5 mm squares on a glass slide. In-gel Edman degradation was performed with PITC to DMSO (1:1000 ratio PITC:DMSO), as described in section **In-gel Edman degradation**, with some modifications detailed here. Positive control included both PITC and TFA treatments, whereas negative control omitted the PITC. After the 30 min TFA incubation at 50 °C, the 300 µL of TFA was removed, and the gel was directly resuspended in 50 µL of 1:1 ACN to water. After shaking the glass vial 10 times, the 50 µL supernatant was pipetted up and down 20 times. The gel shrank and became slightly opaque in this process. The supernatant, containing the PTH-aa, was placed in a mass spec vial for downstream analysis, where 10 µL of the sample was injected into the LC/QToF (an

estimated 2 pmol was injected, if we assume a ~70% Edman efficiency). Analysis of PTH-F abundance was performed using PTH-F exact mass: 282.0827 ± 0.0056 Da (using an error of ± 20 ppm (Δm) : Δm = (ppm error × exact mass) / 1,000,000). Positive control of PTH-phenylalanine, in **Supplementary Figure 6b**, was purchased from TCI, P0367. A stock concentration of 12.5 mM in 1:1 ACN to water (volume ratio, used throughout unless otherwise indicated) with 0.1% TFA was made and stored at -20 °C in 100 µL aliquots. The stock solution was diluted to 100 µM in 1:1 ACN to water and 10 µL of this sample was injected into the LC/QToF (~1 nmol).

Control curve for calculating yield on LC/QToF

Peptide sequences were the fragments from A15 peptide expected after its trypsinization from the gel (fragments recovered in **Fig. 3b-d**): AGGAGGLLGGSR (non-modified peptide), and GGAGGLLGGSR (peptide with cleaved N-terminal amino acid). These sequences were ordered in 3 different 0.8 mg peptide aliquots of each peptide from AAPPTEC (> 95% purity). Peptides were prepared together in solution at various concentrations from 5 µM to 60 µM for the 3 separate replicates for each concentration ("C_product" in **Supplementary Figure 5**: 5, 10, 20, 30, 40, 50, 60 µM, and "I_product" was the abundance of the peptide). In addition, in each solution, a control peptide, namely the A9-peptide, was spiked at 5 µM final concentration in all triplicate samples for each concentration and was used as the standard ("C_standard" was 5 µM, and "I_standard" was the abundance of the A9-peptide). Peptide abundances were analyzed on Agilent MassHunter Qualitative Analysis software. The method for **Analysis of LC/QToF data** was performed for those species to obtain the AUC in an automated fashion.

Strain-promoted alkyne-azide cycloaddition (SPAAC) read-out of in-gel Edman degradation

Edman degradation followed by SPAAC for bulk fluorescence read-outs (**Fig. 3f-g**) were performed with K{N_3}15-peptide. The ExM gelation solution containing peptides at 1 mM concentration was cast in a gelation chamber with thickness ~260 µm (i.e., 2 parafilm thickness), expanded and re-embedded to reach ~2.7X expansion factor in 1X PBS. Thus, the final peptide concentration in the gels reached ~50 µM. ExMre gels were cut to 5 x 5 mm and reached a thickness of ~700 µm with expansion, with each gel containing ~900 pmol peptide.

The steps were the same as the protocol written in the section **In-gel Edman degradation**, but with K{N_3}15-peptide, and with modifications detailed here. As in the in-gel Edman degradation, PITC to DMSO (1:1000 ratio PITC:DMSO) was used for the conjugation step and after removing TFA, the gels were washed 3 times, 10 min each, with 1 mL of 1M Tris buffer pH 8 in 1.5 mL Eppendorf tubes. The gels with condition "PITC:DMSO+TFA+trypsin" were then submerged in 50 µL of 20 µg/mL trypsin in 50 mM Tris-HCl pH 8 in the bottom of the 1.5 mL Eppendorf tube (ensuring they were completely submerged in trypsin solution). The gels for other conditions were submerged in 50 µL of 50 mM Tris-HCl pH 8 (without trypsin). The gels were then incubated overnight (~16 hours) at 37 °C. Then, they were washed in 1X PBS 3 times,10 minutes each, at RT. 20 µg/mL of DBCO-AF488 (Lumiprobe, #218F0) was prepared, from a stock of 10 mg/mL fluorophore in DMF stored at -20 °C, in 1X PBS (500 µL per tube). Gels were submerged in 500 µL of the solution and checked to be well suspended within the Eppendorf tube (not stuck to the edge throughout the staining process). The click reaction was left to react overnight (~16 hours) at 4 °C, and washed 4 times the next day with 1X PBS, 10 min per wash prior to imaging.

The yield calculation based on the SPAAC assay was defined as:
% yield = ("1"-"4")/("1"-"5")
where the numbers represent the separate conditions as indicated in **Fig. 3g**: where (1) DMSO only, (2) PITC to DMSO (1:1000 ratio PITC:DMSO), (3) TFA only, (4) PITC to DMSO (1:1000 ratio PITC:DMSO) followed by TFA, and (5) same as (4) but with the additional trypsinization step.

Amino acid read-outs: ClpS2-StV1 N-terminal binding and Glyphic antibody

Binding ClpS2-StV1 and anti hemagglutinin-tag (HA) antibody to different N-terminal peptides in gel

For **Fig. 4a-d**, ExM gels were cast at a concentration of 1mM peptide in the gelation solution, for all peptides: F_1 peptide, W_1 peptide, Y_1 peptide, A_1 peptide. This was done with 3 different aliquots of 0.8 mg lyophilized peptide (from AAPPTEC, > 95% purity) reconstituted to a concentration of 5 mM in water. The gels were made using a ~130 µm thick gelation chamber (i.e., 1 parafilm thickness). The gels were then expanded and re-embedded, as detailed in **Methods: ExM gel: making the gelation mixture with peptide** and **Re-embedded ExM (ExMre) gels**. The gels were cut to 2 x 2 mm top/flat surface size and stored at 4°C in 1X PBS, reaching a thickness of ~350 µm after expansion, with each gel containing ~70 pmol peptide for **Fig. 4a-d, f-g**. For **Supplementary Figure 7a**, the same peptides were reconstituted to a concentration of 25 mM in water to make 9% acrylamide gels using ~130 µm thick gelation chamber (i.e., 2 parafilm thickness) to reach a thickness of ~310 µm, cut to 2 x 2 mm with each gel containing ~4 nmol peptide. After 3 times 1X PBS washes, 10 min each, these gels were then placed in plastic microcentrifuge tubes (VWR, 87003-292) with 500 µL water and washed 3 times, 5 minutes each, with 500 µL water.

A stock of ClpS2 St-V1 in a buffer of 50 mM sodium phosphate, 150 mM NaCl, pH 7.4 was made. The stock vial was spun down to remove any protein aggregates, and a concentration of ~116 µM was measured in the supernatant using NanoDrop (molecular weight of ClpS2 St-V1 is 13034 g/mol, and an extinction coefficient of 8940 $M^{-1}cm^{-1}$ at 280 nm). The protein sequence ClpS2 St-V1 (and for ClpS2, and ClpS2 V1) is in **Supplementary Note 6.2**. The protein was aliquoted and stored at -20 °C. Before use, the protein was diluted down to 10 µM in a buffer solution of 0.1% BSA in water, with ~4.3 mM sodium phosphate, 13 mM NaCl (for **Fig. 4a, f-g**), or 20 µM in a buffer solution of 0.1% BSA in water, with 8.6 mM sodium phosphate, 26 mM NaCl (for **Supplementary Figure 7a**). 30 µL of the solution was incubated with each gel for 6h (for **Fig. 4a, f-g**) or 1h (for **Supplementary Figure 7a**). For **Fig. 4a,f-g**, gels were washed with 0.1% BSA (Fisher Scientific, #NC1303417) in 1X PBS for ~1-3 min, and replaced with a 30 µL solution of 10 µg/mL (67 nM) (for **Fig. 4a**) or 20 µg/mL (133 nM) (for **Fig. 4f-g**) of anti-HA tag antibody 488 (ThermoFisher, Catalog # 26183-A488) incubated at 4 °C for 7 days (for **Fig. 4a**) or 3 days (for **Fig. 4f-g**). For **Fig. 4a**, after 5 days of antibody incubation, the gels were mixed with a pipette tip to make sure they were not stuck to one side of the plastic tube (which could disrupt equal diffusion of protein from either side of the gel). For **Fig. 4f-g**, after 1 day of antibody incubation, the gels were mixed with a pipette tip to make sure they were not stuck to one side of the plastic tube (which could disrupt equal diffusion of protein from either side of the gel). For **Supplementary Figure 7a**, gels were washed with 0.1% BSA in 1X PBS for ~1-3 min, and replaced with 100 µL solution of 200 µg/mL anti-HA tag antibody 647 (1.3 µM) (ThermoFisher, Catalog # 26183A647) incubated at 4 °C overnight (O/N).

Binding to HA-tagged peptides in acrylamide gels

For **Supplementary Figure 7b**, the HA-tagged peptides were reconstituted to a concentration of 5 mM in water to make 9% acrylamide gels using ~260 µm thick gel chambers (i.e., 2 parafilm thickness; as detailed in **Methods: Acrylamide gel: making the gelation mixture with peptide**). The gels were cut 2 x 2 mm top/flat surface size and stored at 4 °C in 1X PBS (reaching a thickness of ~310 µm in 1X PBS, with each gel containing ~750 pmol peptide). After 3 times 1X PBS washes, 10 min each, these gels were then placed in plastic microcentrifuge tubes with 500 µL water and washed 3 times, 5 minutes each, with 500 µL water. 30 µL of the 20 µM ClpS2St-V1 was incubated with each gel for 1h. The gels were washed with 0.1% BSA (Fisher Scientific, #NC1303417) in 1X PBS, and replaced with 30 µL solution of 20 µg/mL anti-HA tag antibody 647 (ThermoFisher, Catalog # 26183A647) incubated at 4 °C O/N.

In-gel Edman degradation for binding experiments with F_1 peptide and G_1F_2 peptide ExMre gels

ExM gels were cast at a concentration of 1mM peptide, for both peptides: F_1 peptide, and G_1F_2 peptide using 3 different aliquots of 0.8 mg lyophilized peptide (from AAPPTEC, > 95% purity) reconstituted

to a concentration of 5 mM in water. The gels were made using 1 parafilm thickness gel chamber. The gels were expanded and re-embedded (as detailed in **Methods:** **ExM gel: making the gelation mixture with peptide** and **Re-embedded ExM (ExMre) gels**). ExMre gels were cut 5 x 5 mm squares and stored at 4 °C in 1X PBS. In-gel Edman degradation was performed on the gels as previously described (see section **In-gel Edman degradation**), except for the two following modifications: 1:9 of PITC:DMSO by adding 30 µL PITC to 270 µL DMSO, and with only 2 washes with DMSO, 5 min each, at 50 °C, between conjugation (PITC) and cleavage (TFA) steps. After TFA was removed, the gels were washed 3 times with 1M Tris pH 8, 10 min each, at RT. The gels for "PITC:DMSO+TFA+trypsin" condition were then submerged in 50 µL of 20 µg/mL trypsin in 50 mM Tris-HCl pH 8 in the bottom of the 1.5 mL Eppendorf tube (ensuring they were completely submerged in trypsin solution). The gels for other conditions were submerged in 50 µL of 50 mM Tris-HCl pH 8. All gels were then incubated for 2 hours at 37 °C. Each ExMre gel was further cut into 2 x 2 mm squares on a glass slide.

Glyphic binder specificity in gel and Biolayer Interferometry (BLI) curves

As depicted in **Fig. 6d**, antigens from Glyphic Biotechnologies were tested by embedding streptavidin in ExMre gels and subsequently adding biotinylated ClickP-amino acid (abbreviated: biotin-ClickP-aa). For AcX modification of streptavidin, 1.5 µL of 10 mg/mL AcX, (ThermoFisher, A20770) stored at -20 °C in DMSO, was added to 25 µL of 10 mg/mL streptavidin (ThermoFisher, 21122) resuspended in water and stored at 4 °C, to a total volume of 50 µL in 0.1 M $NaHCO_3$ pH 8.5. This solution was left to incubate overnight at room temperature. Using the ~5 mg/mL AcX-modified streptavidin stock, ExM gels were cast, using ~130 µm thick gelation chamber (i.e., 1 parafilm thick), to a final concentration of ~0.1 mg/mL AcX-modified streptavidin. The gels were expanded and re-embedded (as in **Methods:** **ExM gel: making the gelation mixture with peptide** and **Re-embedded ExM (ExMre) gels**), to an approximate concentration of ~5 µg/mL (0.1 mg/mL / 2.7^3) and stored in 1X PBS at 4 °C. The ExMre gels were cut into 2 x 2 mm squares, placed in separate plastic microcentrifuge tubes, and washed 3 times with water. The gels were incubated with biotin-ClickP-aa (where aa: G, V, F) at a concentration of ~16.7 µg/mL, freshly made from a stock of 0.5 mg/mL (from Glyphic) in a total volume of 100 µL in 1X PBS with 0.1% BSA for 24 hours at room temperature. These gels were washed 3 times, 20 min each, with 0.1% BSA in 1X PBS, and then incubated for 2 days with 30 µL of 10 µg/mL RFP-conjugated Glyphic antibody against ClickP-F (abbreviated: Glyphic F) or 5 µg/mL GFP-conjugated Glyphic antibody against ClickP-V (abbreviated: Glyphic V) in 0.1% BSA in 1X PBS at 4 °C.

BLI results of the Glyphic antibodies in **Supplementary Figure 9** were provided by the company.

Imaging

All imaging was performed using a spinning disk confocal microscope (Andor Dragonfly Spinning Disk confocal).

For **Fig. 3g**, the gels were placed on a 24-well glass plate with 1 mL 1X PBS per well. Before imaging, the 1X PBS solution was removed, and the gel was imaged using a 10X magnification lens, acquiring a Z-stack covering the whole gel thickness, with a 10 µm Z-step size (520 µm total Z-stack, except condition "DMSO", replicate 3, with a 510 µm total Z-stack), 40% laser power, 100 ms exposure time, with 488 laser line and 525/50 emission filter.

For **Fig. 4b-d** and **Fig. 4f-g**, the supernatant antibody solution was first removed from the gels via a transfer pipette before imaging. They were placed on a droplet of 50 µL of 1X PBS with 0.1% BSA on a 24-well plate for the smooth transfer of the gel, and the droplet was then removed after the transfer. Imaging was performed with 488 laser and an emission filter of 525/50, with 100% laser power and 500 ms exposure time, using a 40X objective (water immersion) and 10 µm step size (**Fig. 4b-d**) or 15 µm step size (**Fig. 4f-g**) covering the entirety of the gel cross-section (for a final ~430 µm Z-stack for **Fig. 4b-d**; ~375 µm Z-stack for **Fig. 4f-g**).

For **Fig. 6e-f**, the gels were placed on a 24-well plate on a droplet of 50 µL 1X PBS 0.1% BSA in the center of the well to transfer the gel before imaging. Imaging was performed with a 40X objective (water immersion), and a Z-stack obtained from 10 µm step size across the depth of the gels (or a final ~400 µm Z-stack) using 500 ms exposure time, 100% laser power with excitation wavelength in 488 channel and an emission filter of 525/50 (Glyphic V antibody) or 561 and an emission filter of 582/15 (Glyphic F antibody).

For **Supplementary Figure 7a**, the imaging was performed after removing the antibody (anti-HA tag antibody 647) solution using a confocal microscope with 10X objective, 100% laser power in 637 laser with 676/37 emission filter, 400 ms exposure time, 10 µm Z-steps covering the entire gel cross-section.

For **Supplementary Figure 7b**, the imaging was performed after removing the antibody (anti-HA tag antibody 647) solution using a confocal microscope with 10X objective, 100% laser power in 637 laser with 676/37 emission filter, 100 ms exposure time, 10 µm Z-steps covering the entire gel cross-section.

Image analysis

For **Fig. 3g, Fig. 4b-d, Fig. 4f-g** and **Figure 6e-f,** images ('.ims' files) were opened in FIJI and gel Z-stacks were inspected by plotting the Z-axis profile. The final Z-stack thickness for all conditions was set to capture the entire thickness of the gel and was made identical for each replicate (see **Imaging** section). The images were then saved as '.tif' files.

The plots of gel fluorescence throughout the Z-stack (**Fig. 3gii, Fig. 4c, fii, gii, and Fig. 6eii, fii**) and bar plots of the average fluorescence throughout the gel volume (**Fig. 3giii, Fig. 4d, fiii, giii, and Fig. 6eiii, fiii**) were made in MATLAB (see GitHub **Code**). For the Z-stack plot showing fluorescence throughout the gel, the mean and standard deviation of fluorescence for each Z-slice (over the replicates of the same condition for the same Z-slice) was calculated and plotted in different colors for each condition. For the bar plots, the average fluorescence of the gels across the whole volume was computed (i.e., for each Z-slice of a given sample, the mean fluorescence intensity of all pixel values was computed. Then, for each sample, the average intensity across all its Z-slices was calculated, and plotted as a separate data point. Finally, the average intensity for each condition was obtained by computing the mean of these sample-wise average intensities).

The representative images of each gel (**Fig. 3gi, 4b, 4fi, 4gi, 6ei, 6fi**) was obtained from one gel replicate for a given condition. The same arbitrarily selected depth, reported in the figure captions, was taken for each condition by making a substack of a single Z-slice (e.g. to obtain the image for Z-slice # 20: Image → Stacks → Tools → Make Substack, input "20-20"). A scale bar was also added to every image representing the first condition of each experiment. These average intensity profiles for each condition were then plotted in Python using the *mpimg* Matplotlib image module, with an intensity scale for comparison (see GitHub **Code**).

Oxidation analysis in ExM gels

Amino acid oxidation analysis in ExM gels were performed with XaaGGAGRGLGK{acr} peptides, where Xaa was one of 8 amino acids: M, C, W, Y, F, P, H, R (as described in **Synthetic peptide designs** section). Post-translational oxidation modifications for each amino acid can be found in **Supplementary Figures 8a-h**.

Three conditions were prepared for each of the N-terminal acid peptides. Two in solution, where condition 1 was: 20 µL of 5 mM peptide with 90 µL water, and condition 2 was: 20 µL of 5 mM peptide with 80 µL water, 5 µL of 10% TEMED (v/v), and 5 µL 10% APS (w/v), incubated at 37 °C for 1h30m. The condition 3 was ExM gelation: 20 µL of 5 mM peptide with 60 µL of StockX, 10 µL

of 130 mM BIS, 5 µL of 10% TEMED (v/v), and 5uL 10% APS (w/v) gelled into 20 x 10 mm surface size chamber with ~260 µm chamber thickness (i.e., 2 parafilm thickness). Gelation was performed as described in section **ExM gel: making the gelation mixture with peptide**.

Condition 1 and condition 2 were further diluted by taking 15 µL of the sample and adding 50 µL 20 µg/mL trypsin in 1M Tris pH 8 for a total volume of 65 µL.Condition 3 gels were cut into 5 mm x 10 mm ExM gel using a razor blade (right after gelation; without expanding) and washed twice with 1X PBS, 5 min each, then once with 1M Tris pH 8 for 5 min. The wash solution was then replaced with 50 µL of 20 µg/mL trypsin in 1M Tris pH 8 on top of the gel. Condition 1, condition 2 and condition 3 were all incubated for 4 hours at 37 °C with the trypsin solution. After trypsinization, the samples were placed in the fridge until analysis on LC/QToF. 50 µL of the solutions, for each condition, were placed in separate mass spec vials for analysis. Condition 1, 2, and 3 samples were injected into the LC/QToF (see section on **Mass spectrometry**).

Analysis of the N-terminal amino acid oxidation was performed as detailed in **Analysis of LC/QToF data** using the information about post-translational oxidation modification of peptides documented in **Supplementary Table 3**. For each N-terminal amino acid, the abundance (a.u.) of the original fragment and the fragment with various post-translational oxidation modifications was extracted for the three different sample conditions and documented in a table (see **Supplementary Figures 8aii-hii**). Then, using these values, the relative abundance of each of the fragments were compared using a bar graph. For this, the abundance of each species was normalized with max absolute scaling for plotting (see **Supplementary Figures 8aiii-hiii**). The raw total ion chromatograms and mass spectra were exported as metafiles ('.emf'), and available in **Source Data**.

Theoretical assessment of *in situ* protein sequencing

All analyses were performed either on mycoplasma genitalium proteome (483 Swiss-Prot reviewed proteins; UniProt ID 243273: by searching "taxonomy_id:243273" in UniProt), or the human proteome (20,421 Swiss-Prot reviewed proteins, UniProt ID 9606: by searching "taxonomy_id:9606" in UniProt). All references throughout for mycoplasma proteome and the human proteome are for these two proteomes, respectively. All code can be found on GitHub, at: https://github.com/camimarie/insituprotein/. Amino acid residues are abbreviated using their single letter amino acid code. N-terminal is abbreviated "N-term".

Part 1: results of exploring the chemistries and their effects on preserving protein sequence coverage in the gel

Percent amino acid count

For **Supplementary Figure 13c** (mycoplasma proteome) heatmap, the protein database went through a three step chemical process that modified the primary sequence of amino acids. A gridsearch was performed with various combinations of P(fix), P(anchor), P(digest), where these represent the probability of these three chemical steps to modify a subset of amino acid side chains. In addition, the anchoring could be either AcX or epoxide, and digestion could be either endoprotease Lys-C (abbreviated Lys-C), Proteinase K (abbreviated ProK) or trypsin. The gridsearch values were: P(fixed): 0.00, 0.01, 0.05, 0.10, 0.50, 1.00; P(anchored): 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0; P(digestion): 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0. First, the proteome (mycoplasma or human) was selected from UniProt and loaded into the computational environment. Then, N-term, K, C, R, and Y residues were modified with probability P(fixed), where they were replaced with a "J" (arbitrarily, since "J" is not overlapping with any other amino acid single letter code) after fixation. Similarly, for anchoring, N-term and K residues for AcX, or N-term, K, C, H, Y, D and E for epoxide, were modified with probability of P(anchored), and replaced with a "Z" (since "Z" was not overlapping with any other letter). Lysine residues that were already converted to a J could not be

modified to a Z in this anchoring step). For digestion, residue K was digested at its C-terminus for Lys-C, residues R and K were digested at their C-terminus for trypsin, whereas for ProK, Y, W, E, L, V, A, I, F, T residues were digested at their C-terminus. For all digestion types (Lys-C, trypsin, ProK), digestion did not occur if they had already been converted to a J (fixed) or Z (anchored). A fragment was only retained if it had at least one anchored residue to the gel matrix from the P(anchor), otherwise it was considered lost. Each grid in the grid search represents a unique combination of those three parameters, where each unique combination was run 10 times. The "percent amino acid count" was calculated as the percentage of residues of a protein that are not modified (i.e., remains a canonical amino acid), and remain accessible for read-out over 15 rounds of in-gel Edman degradation after the chemical steps. This was computed by summing over all the amino acids of a fragment that are unmodified by any chemical process, stopping at the last anchor of a fragment (subsequent amino acids would be lost and not be readable by in-gel Edman degradation) and, if the final anchor is not within the first 15 amino acids of the fragment, stopping at the 15th amino acid. The N-term modifications in either the fixation or anchoring steps prevented read-out of the given fragment, for the mycoplasma proteome. The mean percent amino acid count was calculated by averaging over the 10 replicated computational experiments and also over the proteome and plotted on the heatmap in **Supplementary Figure 13c.**

For **Supplementary Figure 18** (human proteome), it was the same as described for **Supplementary Figure 13c** (mycoplasma proteome), but the N-terminal fragment was always considered inaccessible due to N-terminal acetylation. In addition the gridsearch values for P(fixed), P(anchored) and P(digestion) were the same values as above.

For **Supplementary Figure 14a** and **Supplementary Figure 14c**, the distribution of the percent amino acid count (defined above) was plotted as a distribution for all the proteins in the proteomes averaged over the 10 rounds of computational experiments for P(fixation)=0.05, P(anchoring)=0.80, P(digestion)=0.80, for both mycoplasma and human proteomes, where the error bar represented the standard deviation over the multiple computational repetitions. For **Supplementary Figure 14b, Supplementary Figure 14d**, the distribution of the percent amino acid count was plotted as a distribution for all the proteins in the proteomes over the 10 rounds of computational experiments for the highest percent amino acid count for each condition (best performing parameters given the condition) as calculated and plotted in **Supplementary Figure 13c** and **Supplementary Figure 18** (for the mycoplasma proteome and human proteome, respectively). The bin-size is 40.

Fragment lengths

For **Supplementary Figure 15**, the fragment length was defined as the number of amino acids for a given fragment retained in the gel, including amino acids that are fixed or anchored (unlike the percent amino acid count), but not considering amino acids that are located downstream, C-terminal, of the last anchoring amino acid of the fragment. We also do not cap at 15 amino acids, unlike the percent amino acid count. The results for these fragment lengths were plotted as distributions of protein count versus fragment length, where the mean and median values were recorded by plotting a vertical line along with their values. The y-axis was log-scale, and standard deviations were the result of 10 computational repetitions, with bin-size of 40. Results of the fragment lengths using a parametrization of P(fixation)=0.05, P(anchoring)=0.8, P(digestion)=0.8, were plotted for the mycoplasma proteome, human proteome (**Supplementary Figure 15a-b**, respectively), which were the same parameterizations used in **Part 2**.

Part 2: Simulating *in situ* protein sequencing with NAABs and in gel Edman degradation

Reference fragment dataset (T_s)

The reference fragment dataset represents the database of ground-truth sequences present in the gel, after fixation, anchoring, gelation, digestion and expansion, but prior to the *in situ* protein sequencing

read-out. To generate the dataset, first, similarly to **Part 1** methods, the mycoplasma proteome was selected from UniProt and loaded into the computational environment. Then, the primary sequence of amino acids was modified according to parameters for P(fixed), P(anchored), P(digestion), where these represent the probability of these three chemical steps to modify a subset of amino acid side chains. However, in **Part 2**, these values were fixed with P(fixed)=0.05, P(anchored)=0.8, P(digestion)=0.8 using paraformaldehyde, AcX and trypsin for fixation, anchoring and digestion in the model, respectively. The amino acid side chain specificities of the paraformaldehyde fixation, anchoring with AcX, and digestion with trypsin, remained the same as in **Part 1**. Only fragments with at least one anchored residue to the gel matrix from the P(anchoring) were retained. Subsequently, an idealistic in-gel Edman degradation read-out was conducted on the remaining fragments by considering all the sequencing chemistries, including the conjugation and cleavage of in-gel Edman degradation to be 100%. In addition, the idealistic read-out could detect all amino acids with 100% accuracy, except ones that have been fixed or anchored at previous steps (fixation and anchoring). In this way, the fragments generated were perfectly corresponding to the ground-truth sequence that would be expected when reading them out in idealistic conditions. In addition, if Edman leads to the loss of the only anchor, amino acids following are gone and will be substituted with an unknown read ("X") since there will be no further amino acids to readout. After, amino acid sidechains modified to J or Z are unrecognized by binders, and thus will return X (given that modifications like J and Z cannot be distinguished during read-out, and will also result in failure to bind) and stored as a fragment. This was performed 1,000 times, and the fragments from these computational experiments were then aggregated into the reference fragment dataset, also named $\boldsymbol{T_s}$ (peptides anchored in the expansion gel independent of read errors). For **Supplementary Figure 16**, the number of unique fragments generated were compared for 1, 10, 100, 1,000, and 10,000 simulations and plotted for mycoplasma.

Error-prone fragment dataset and assignment to proteins

The proteome (mycoplasma) was selected from UniProt and loaded into the computational environment. The first steps of the error-prone fragment dataset were the same as the reference fragment dataset generation up until the modification of the primary protein sequences into fragments through digestion (with values for P(fixed)=0.05, P(anchored)=0.8, P(digestion)=0.8 using paraformaldehyde, AcX and trypsin for fixation, anchoring and digestion in the model, respectively). Only fragments with at least one anchored residue to the gel matrix from the P(anchoring) were retained. After, amino acid sidechains modified to J or Z are unrecognized by binders, and thus will return X (given that modifications like J and Z cannot be distinguished during read-out). At this stage, instead of considering an idealistic in-gel Edman degradation read-out, the read-out was simulated with errors, including the sequencing chemistry and the binder kinetics.

Sequencing chemistry errors

Specifically, for the sequencing chemistry errors, the conjugation of PITC was considered to have a probability of success of 0.9, and the cleavage with TFA with a 0.7 success rate. Amino acids with a modified sidechain (by anchoring or fixation) were considered to be cleavable via in-gel Edman degradation, but when reaching the last anchor of the peptide, the rest of the fragment was lost and subsequent amino acids would be read as unknown (i.e. "X").

Binder read-out errors and error-prone fragment dataset ($\boldsymbol{F_s}$)

For simulation of the binding, after the PITC conjugation step, binders were bound probabilistically to PITC-conjugated N-terminal amino acids, and this was performed sequentially for every binder. Results were performed with several subsets of binders that recognized various PITC-conjugated N-terminal amino acids. The 20 amino acid binder subset with recognizers towards all canonical amino acid side-chains (but not the X sidechains, of course). The 15 amino acid binder subset contained the following recognizers: {A, N, D, E, Q, G, I, L, K, F, S, T, P, Y, V}, the 10 amino acid binder subset contained the following recognizers: {A, N, D, E, Q, I, L, K, F, V}, and the 5 amino acid binder subset contained the following recognizers: {N, E, L, F, V}. The binding was simulated in

a condition with excess binder, and where binding reaches equilibrium (using the Langmuir equation for binding), such that after the washing, the probability of being bound to on-target and off-target was defined as: $p_{correct} = \left(\frac{[C]}{[C] + K_d^{on-target}}\right) \cdot e^{-k_{off}^{on-target} t_{wash}}$ and $p_{incorrect} = \left(\frac{[C]}{[C] + K_d^{off-target}}\right) \cdot e^{-k_{off}^{off-target} t_{wash}}$, where $[C]$ is the concentration of binder, $K_d^{on-target}$ the equilibrium dissociation constant for the on-target, $K_d^{off-target}$ the equilibrium dissociation constant for the off-target, $k_{off}^{on-target}$ the dissociation rate for the on-target, and $k_{off}^{off-target}$ the dissociation rate for the off-target, and t_{wash} is the time elapsed during the dissociation phase (during the washing). **Supplementary Figure 17** was plotted with these equations (plotting $p_{correct} - p_{incorrect}$) by modifying the values of binder washing, t_{wash} from 0 to 200 min, and the concentration of binder, $[C]$, from 1 pM to 100 mM. These were plotted for a binder with different levels of specificity according to table **Supplementary Table 12**: very high, high, medium, low, and very low specificities, where the equilibrium dissociation constant and dissociation rates for on-target and off-target were modified in the equation for each of these specificity levels. Very high, high, and medium resulted in the same on-target and off-target binding profiles, and thus only one (medium) specificity was included in **Supplementary Figure 13d**. For the results in **Supplementary Figure 13d** (mycoplasma proteome), the t_{wash} was considered as 30 min, and the $[C]$ as 1 µM, and results were generated for different levels of binder specificities according to table **Supplementary Table 12**. For each in-gel Edman degradation round, independent reads were obtained for each amino acid (20, 15, 10, and 5 times) since the binding was performed sequentially for each binder. Thus, several read-outs were associated with the same position in a peptide fragment (this was taken into consideration for the downstream Hidden Markov Model (HMM)). After performing the binding, the N-terminal amino acid was cleaved with sequencing chemistry errors mentioned above, and the process was repeated from 5 to 15 rounds of in-gel Edman degradation and sequential binder read-out. This resulted in error-prone fragment sequences for each protein, which collectively became the error-prone fragment dataset (F_s).

Hidden Markov Model (HMM) based matching and fraction of proteome correctly identified

Pre-filtering uninformative fragments

Following the generation of the error-prone dataset, which included for every protein in the mycoplasma proteome at each condition (number of rounds × number of binder cases × number of specificity cases × 10 repetitions = 1,760 conditions), low quality fragments were filtered. First, uninformative "query fragments", originating from the error-prone dataset F_s, were defined as fragments that consisted of more than half of their total Edman degradation rounds contained only unknown reads ("X"), e.g. rounds where no binder sequentially bound at all. These fragments were removed. Second, uninformative "reference fragments", originating from the reference dataset T_s, were defined as fragments that consisted of more than half of the total reads as unknown reads, denoted by an "X". These fragments were also removed.

Algorithmic details for two-stage HMM-based matching:

Next, for each remaining fragment, a two-stage HMM-based matching was performed. The HMM models the joint probability $P(F|T, \theta)$ that a true fragment T generates an observed fragment F given the experimental error parameters θ, which include the following:

- Conjugation failures (c): probability that no conjugation occurs in an Edman round.
- Cleavage failures (d): probability that the N-terminal amino acid is not cleaved.

- Binder errors: per-binder probabilities describing on and off-target binding, where unbound is denoted as "X": $p_{bound,\,on-target}$, $p_{unbound,\,on-target}$, $p_{bound,\,off-target}$, $p_{unbound,\,off-target}$.

At each Edman round, each amino acid from the true fragment T emits m observations, one per binder in the binder sets described above. The model sums over all possible emission paths and state transitions, including insertion (failed cleavage) and stay (failed conjugation) transitions.

To map each observed fragment, subject to Edman conjugation failures, Edman cleavage failures, and binder errors, back to its source, we framed the problem as follows: for every "true" fragment T in the error-free trie, T_s (reference fragment dataset), what was the probability that T under set parameters θ would generate the observed, error-prone fragment F in the error-prone dataset (F_s). Thus, for each error-prone fragment T, the Viterbi log-likelihood, $\max_{\pi} \log P(F, \pi|T, \theta)$ was computed, which represents the probability of the most likely path $\pi*$ generating the observed fragment F, given each candidate sequence T. After computing these Viterbi scores for all candidates, we apply a pruning step to discard less likely candidates: $\log P(F, \pi*|T,\ \theta) \geq \max_{S'} \log P(F, \pi*|S',\ \theta) - \epsilon_{vit}$, where S' is the highest scoring fragment across all candidates. ϵ_{vit} was set as 5.0, which was chosen to be deliberately wide as there is increased variance with certain conditions: in log-likelihood space, any candidate that scores $1/e^5 \approx 1/150$ of the top score is retained. On the pruned candidate set, a forward dynamic programming algorithm is employed that computes the total likelihood $\log L(T \rightarrow F) = \log P(F|T,\ \theta) = \log \sum_{\pi} P(F, \pi|T,\ \theta)$, summing over almost all possible paths with a milder beam $\epsilon_{for} = 10.0$. Note that no candidates are ever removed here, as the forward epsilon only prunes unlikely paths within the possible paths.

Fraction of the proteome correctly identified

For each of the top 10 candidate fragments by forward log-likelihood $T \in T_S$ (if such fragments exist and have not been filtered out), we log-summed fragment likelihoods weighted by counts using the reference fragment dataset's protein occurrence counts for each fragment T. We then normalize and rank proteins by their total weight. The highest scoring protein is returned as the fragment's originating protein only if the weight is at least 1.5x the second-highest scoring protein. Otherwise, the fragment is labeled "uncertain". For each protein, we accumulate all the fragments' labels. If no valid predictions are produced, we return "no predictions". Otherwise, we count fragment-level votes per predicted protein: the protein is identified only if the top vote count is at least double the runner up, otherwise we return "uncertain". We report the counts of "correct", "false positive", "uncertain", and "no predictions". From these decisions, the fraction correctly identified over the whole proteome was a metric for how many of the proteins in the mycoplasma proteome were correctly identified from the simulation of the error-prone fragments (not including uncertain or false positive matches). For **Supplementary Figure 13d**, the whole simulation was performed 10 times and the average and standard deviation of the fraction of correctly identified proteins was reported on the graph of results. In **Supplementary Figure 19**, we plot the average and standard deviation of the fraction of false positively identified proteins.

References

1. Alberts, B. *et al.* Protein Function. *Molecular Biology of the Cell. 4th edition* (2002).
2. Haucke, V., Neher, E. & Sigrist, S. J. Protein scaffolds in the coupling of synaptic exocytosis and endocytosis. *Nat. Rev. Neurosci.* **12**, 127–138 (2011).
3. Good, M. C., Zalatan, J. G. & Lim, W. A. Scaffold proteins: hubs for controlling the flow of cellular information. *Science* **332**, 680–686 (2011).
4. DiRusso, C. J., Dashtiahangar, M. & Gilmore, T. D. Scaffold proteins as dynamic integrators of biological processes. *J. Biol. Chem.* **298**, 102628 (2022).
5. Hardingham, G. E. & Bading, H. Synaptic versus extrasynaptic NMDA receptor signalling: implications for neurodegenerative disorders. *Nat. Rev. Neurosci.* **11**, 682–696 (2010).
6. Chen, F., Tillberg, P. W. & Boyden, E. S. Expansion microscopy. *Science* **347**, 543–548 (2015).
7. Tillberg, P. W. *et al.* Protein-retention expansion microscopy of cells and tissues labeled using standard fluorescent proteins and antibodies. *Nat. Biotechnol.* **34**, 987–992 (2016).
8. Dolgin, E. 'Expansion microscopy' turns ten: how a tissue-swelling method brought super-resolution imaging to the masses. *Nature* **637**, 752–754 (2025).
9. Chang, J.-B. *et al.* Iterative expansion microscopy. *Nat. Methods* **14**, 593–599 (2017).
10. Sarkar, D. *et al.* Revealing nanostructures in brain tissue via protein decrowding by iterative expansion microscopy. *Nat. Biomed. Eng.* **6**, 1057–1073 (2022).
11. Wang, S. *et al.* Single-shot 20-fold expansion microscopy. *Nat. Methods* 1–7 (2024).
12. Shaib, A. H. *et al.* One-step nanoscale expansion microscopy reveals individual protein shapes. *Nat. Biotechnol.* 1–9 (2024).
13. Alfaro, J. A. *et al.* The emerging landscape of single-molecule protein sequencing technologies. *Nat. Methods* **18**, 604–617 (2021-6).
14. Reed, B. D. *et al.* Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device. *Science* (2022) doi:10.1126/science.abo7651.
15. Swaminathan, J. *et al.* Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. *Nat. Biotechnol.* 10.1038/nbt.4278 (2018).

16. Rodriques, S. G., Marblestone, A. H. & Boyden, E. S. A theoretical analysis of single molecule protein sequencing via weak binding spectra. *PLoS One* **14**, e0212868 (2019).

17. Ritmejeris, J., Chen, X. & Dekker, C. Single-molecule protein sequencing with nanopores. *Nat Rev Bioeng* **3**, 303–316 (2024).

18. Wang, G., Moffitt, J. R. & Zhuang, X. Multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy. *Sci. Rep.* **8**, 4847 (2018).

19. Alon, S. *et al.* Expansion sequencing: Spatially precise in situ transcriptomics in intact biological systems. *Science* **371**, eaax2656 (2021).

20. Labade, A. S. *et al.* Expansion in situ genome sequencing links nuclear abnormalities to hotspots of aberrant euchromatin repression. *bioRxivorg* 2024.09.24.614614 (2024) doi:10.1101/2024.09.24.614614.

21. Edman, P., Högfeldt, E., Sillén, L. G. & Kinell, P. O. Method for Determination of the Amino Acid Sequence in Peptides. *Acta Chemica Scandinavica* **4**, 283–293 (1950).

22. Edman, P. & Begg, G. A Protein Sequenator. *Eur. J. Biochem.* **1**, 80–91 (1967).

23. Tullman, J., Marino, J. P. & Kelman, Z. Leveraging nature's biomolecular designs in next-generation protein sequencing reagent development. *Appl. Microbiol. Biotechnol.* **104**, 7261–7271 (2020).

24. Mitra, R. D., Shendure, J., Olejnik, J., Edyta-Krzymanska-Olejnik & Church, G. M. Fluorescent in situ sequencing on polymerase colonies. *Anal. Biochem.* **320**, 55–65 (2003).

25. Tullman, J., Callahan, N., Ellington, B., Kelman, Z. & Marino, J. P. Engineering ClpS for selective and enhanced N-terminal amino acid binding. *Appl. Microbiol. Biotechnol.* **103**, 2621–2633 (2019).

26. Tullman, J., Christensen, M., Kelman, Z. & Marino, J. P. A ClpS-based N-terminal amino acid binding reagent with improved thermostability and selectivity. *Biochem. Eng. J.* **154**, 107438 (2020).

27. Callahan, N., Tullman, J., Kelman, Z. & Marino, J. Strategies for Development of a Next-Generation Protein Sequencing Platform. *Trends Biochem. Sci.* **45**, 76–89 (01/2020).

28. Havranek, J. J. & Borgo, B. Molecules and methods for iterative polypeptide analysis and processing. *US Patent* (2014).

29. Callahan, N., Siegall, W. B., Bergonzo, C., Marino, J. P. & Kelman, Z. Contributions from ClpS surface residues in modulating N-terminal peptide binding and their implications for NAAB development. *Protein Eng. Des. Sel.* **36**, gzad007 (2023).

30. Ikonomova, S. P. *et al.* Engineering GID4 for use as an N-terminal proline binder via directed evolution. *Biotechnol. Bioeng.* **122**, 179–188 (2025).

31. Zheng, L., Sun, Y., Eisenstein, M. & Soh, H. T. Peptide sequencing via reverse translation of peptides into DNA. *bioRxiv* (2024) doi:10.1101/2024.05.31.596913.

32. Laursen, R. A. Solid-phase Edman degradation. An automatic peptide sequencer. *Eur. J. Biochem.* **20**, 89–102 (1971).

33. Muramoto, K., Nokihara, K., Ueda, A. & Kamiya, H. Gas-phase microsequencing of peptides and proteins with a fluorescent Edman-type reagent, fluorescein isothiocyanate. *Biosci. Biotechnol. Biochem.* **58**, 300–304 (1994).

34. Chen, F. *et al.* Nanoscale Imaging of RNA with Expansion Microscopy. *Nat. Methods* **13**, 679–684 (2016-8).

35. Fry, B., Slaw, K. & Polizzi, N. F. Zero-shot design of drug-binding proteins via neural selection-expansion. *bioRxiv* 2025.04.22.649862 (2025) doi:10.1101/2025.04.22.649862.

36. Gao, R. *et al.* A highly homogeneous polymer composed of tetrahedron-like monomers for high-isotropy expansion microscopy. *Nat. Nanotechnol.* **16**, 698–707 (2021).

37. Lee, H., Yu, C.-C., Boyden, E. S., Zhuang, X. & Kosuri, P. Tetra-gel enables superior accuracy in combined super-resolution imaging and expansion microscopy. *Sci. Rep.* **11**, 16944 (2021).

38. Niall, H. D. Automated Edman degradation: the protein sequenator. *Methods Enzymol.* **27**, 942–1010 (1973).

39. Otaka, A., Koide, T., Shide, A. & Fujii, N. Application of dimethylsulphoxide(DMSO) / trifluoroacetic acid(TFA) oxidation to the synthesis of cystine-containing peptide. *Tetrahedron Lett.* **32**, 1223–1226 (1991).

40. Spetzler, J. C. & Meldal, M. Evaluation of strategies for ?one-pot? deprotection, cleavage and disulfide bond formation in the preparation of cystine-containing peptides. *Lett. Pept. Sci.* **3**, 327–332 (1997).

41. Deol, H. *et al.* After 75 years, an alternative to Edman degradation: A mechanistic and efficiency study of a base-induced method for N-terminal peptide sequencing. *J. Am. Chem. Soc.* **147**, 13973–13982 (2025).

42. Estandian, D. M. Enabling tools for de-novo single molecule protein sequencing. (Massachusetts Institute of Technology, 2021).

43. Estandian, D. M., Choueiri, A. G., Boyden, E. S. & Wassie, A. Single-molecule protein and peptide sequencing. Preprint at https://patents.google.com/patent/US20230305017A1/en (2023).

44. Choueiri, A. G. Single-Molecule Protein Sequencing (I) and Genetically Dominant mRNA Therapies to Combat Viral Evolution (II). (Massachusetts Institute of Technology, 2022).

45. Farnsworth, V. & Steinberg, K. The Generation of Phenylthiocarbamyl or Anilinothiazolinone Amino Acids from the Postcleavage Products of the Edman Degradation. *Anal. Biochem.* **215**, 200–210 (12/1993).

46. Braun, M. B. *et al.* Peptides in headlock – a novel high-affinity and versatile peptide-binding nanobody for proteomics and microscopy. *Sci. Rep.* **6**, 19211 (2016).

47. Hodges, R. S., Heaton, R. J., Parker, J. M., Molday, L. & Molday, R. S. Antigen-antibody interaction. Synthetic peptides define linear antigenic determinants recognized by monoclonal antibodies directed to the cytoplasmic carboxyl terminus of rhodopsin. *J. Biol. Chem.* **263**, 11768–11775 (1988).

48. Pinilla, C., Appel, J. R. & Houghten, R. A. Functional importance of amino acid residues making up peptide antigenic determinants. *Mol. Immunol.* **30**, 577–585 (1993).

49. Rodgers, J. R. & Rich, R. R. 6 - Antigens and antigen presentation. in *Clinical Immunology (Fourth Edition)* (eds. Rich, R. R. et al.) 77–89 (Elsevier, London, 2013).

50. Stark, S. E. & Caton, A. J. Antibodies that are specific for a single amino acid interchange in a protein epitope use structurally distinct variable regions. *J. Exp. Med.* **174**, 613–624 (1991).

51. Stark, Y. *et al.* Modular binder technology by NGS-aided, high-resolution selection in yeast of designed armadillo modules. *Proc. Natl. Acad. Sci. U. S. A.* **121**, e2318198121 (2024).

52. Buus, S. *et al.* High-resolution Mapping of Linear Antibody Epitopes Using Ultrahigh-density Peptide Microarrays. *Mol. Cell. Proteomics* **11**, 1790–1800 (2012).

53. Khakzad, H. *et al.* A new age in protein design empowered by deep learning. *Cell Syst.* **14**, 925–939 (2023).

54. Cao, L. *et al.* Design of protein-binding proteins from the target structure alone. *Nature* **605**, 551–560 (2022).

55. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

56. Asano, S. M. *et al.* Expansion Microscopy: Protocols for Imaging Proteins and RNA in Cells and Tissues. *Curr. Protoc. Cell Biol.* **80**, e56 (2018-9).

57. Hoshi, T. & Heinemann, S. H. Regulation of cell function by methionine oxidation and reduction. *J. Physiol.* **531**, 1–11 (2001).

58. Jortzik, E., Wang, L. & Becker, K. Thiol-Based Posttranslational Modifications in Parasites. *Antioxid. Redox Signal.* **17**, 657–673 (2011).

59. Giulivi, C., Traaseth, N. J. & Davies, K. J. A. Tyrosine oxidation products: analysis and biological relevance. *Amino Acids* **25**, 227–232 (2003).

60. Galano, A. & Cruz-Torres, A. OH radical reactions with phenylalanine in free and peptide forms. *Org. Biomol. Chem.* **6**, 732–738 (2008).

61. Todorovski, T., Fedorova, M., Hennig, L. & Hoffmann, R. Synthesis of peptides containing 5-hydroxytryptophan, oxindolylalanine, N-formylkynurenine and kynurenine. *J. Pept. Sci.* **17**, 256–262 (2011).

62. Uchida, K. Histidine and lysine as targets of oxidative modification. *Amino Acids* **25**, 249–257 (2003).

63. Xu, G. & Chance, M. R. Hydroxyl Radical-Mediated Modification of Proteins as Probes for Structural Proteomics. *Chem. Rev.* **107**, 3514–3543 (2007).

64. Phang, J. M., Liu, W. & Zabirnyk, O. Proline Metabolism and Microenvironmental Stress. *Annu. Rev. Nutr.* **30**, 441–463 (2010).

65. Petrov, D., Daura, X. & Zagrovic, B. Effect of Oxidative Damage on the Stability and Dimerization of Superoxide Dismutase 1. *Biophys. J.* **110**, 1499–1509 (2016).

66. Tayeh, M. A. & Marletta, M. A. Macrophage Oxidation of L-Arginine to Nitric Oxide, Nitrite, and Nitrate: Tetrahydrobiopterin is required as a cofactor*. *J. Biol. Chem.* **264**, 19654–19658 (1989).

67. Kostić, M. D., Mihajlović, K. & Divac, V. M. Kinetic Study of the Pyridine-Catalyzed Selenolactonization of 4-Pentenoic Acid. *Catal. Letters* **150**, 2076–2081 (2020).

68. Laursen, R. A. [27] Automatic solid-phase Edman degradation. in *Methods in Enzymology* vol. 25 344–359 (Academic Press, 1972).

69. Boehnert, M. & Schlesinger, D. H. Improved manual sequential analysis of peptides. *Anal. Biochem.* **96**, 469–473 (1979).

70. Eriksson, S. & Sjöquist, J. Quantitative determination of N-terminal amino acids in some serum proteins. *Biochim. Biophys. Acta* **45**, 290–296 (1960).

71. Sullivan, S. & Wong, T. W. A manual sequencing method for identification of phosphorylated amino acids in phosphopeptides. *Anal. Biochem.* **197**, 65–68 (1991).

72. Tarr, G. E. A general procedure for the manual sequencing of small quantities of peptides. *Anal. Biochem.* **63**, 361–370 (1975).

73. Zimmerman, C. L., Appella, E. & Pisano, J. J. Rapid analysis of amino acid phenylthiohydantoins by high-performance liquid chromatography. *Anal. Biochem.* **77**, 569–573 (1977).

74. Pramanik, B. C., Hinton, S. M., Millington, D. S., Dourdeville, T. A. & Slaughter, C. A. Analysis of phenylthiohydantoin amino acid mixtures for sequencing by thermospray liquid chromatography/mass spectrometry. *Anal. Biochem.* **175**, 305–318 (1988).

75. Matsudaira, P. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J. Biol. Chem.* **262**, 10035–10038 (1987).

76. Iida, T., Santa, T., Toriba, A. & Imai, K. Semi-automatic amino acid sequencing and D/L-configuration determination of peptides with detection of liberated N-terminal phenylthiocarbamoylamino acids. *Analyst* **123**, 2829–2834 (1998).

77. Matsunaga, H. *et al.* Proton: a major factor for the racemization and the dehydration at the cyclization/cleavage stage in the Edman sequencing method. *Anal. Chem.* **68**, 2850–2856 (1996).

78. Mustățea, G., Ungureanu, E. & Iorga, E. Protein acidic hydrolysis for amino acids analysis in food-progress over time : A short review. (2019).

79. Giera, M., Aisporna, A., Uritboonthai, W. & Siuzdak, G. The hidden impact of in-source fragmentation in metabolic and chemical mass spectrometry data interpretation. *Nat. Metab.* **6**, 1647–1648 (2024).

80. van der Rest, G., He, F., Emmett, M. R., Marshall, A. G. & Gaskell, S. J. Gas-phase cleavage of PTC-derivatized electrosprayed tryptic peptides in an FT-ICR trapped-ion cell: mass-based protein identification without liquid chromatographic separation. *J. Am. Soc. Mass Spectrom.* **12**, 288–295 (2001).

81. Stein, B. J., Grant, R. A., Sauer, R. T. & Baker, T. A. Structural Basis of an N-Degron Adaptor with More Stringent Specificity. *Structure* **24**, 232–242 (2016).

82. Sass, H.-J., Musco, G., Stahl, S. J., Wingfield, P. T. & Grzesiek, S. Solution NMR of proteins within polyacrylamide gels: Diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J. Biomol. NMR* **18**, 303–309 (2000).

83. Albe, K. R., Butler, M. H. & Wright, B. E. Cellular concentrations of enzymes and their substrates. *J. Theor. Biol.* **143**, 163–195 (1990).

84. Murray, E. *et al.* Simple, Scalable Proteomic Imaging for High-Dimensional Profiling of Intact Systems. *Cell* **163**, 1500–1514 (2015).

85. Crank, J. *The Mathematics of Diffusion. United Kingdom*. (Clarendon Press, 1964).

86. Fischer, H., Polikarpov, I. & Craievich, A. F. Average protein density is a molecular-weight-dependent function. *Protein Sci.* **13**, 2825–2828 (2004).

87. Trnková, L., Dršata, J. & Boušová, I. Oxidation as an important factor of protein damage:

Implications for Maillard reaction. *J. Biosci.* **40**, 419–439 (2015).

88. Heinonen, M., Gürbüz, G. & Ertbjerg, P. Chapter 3 - Oxidation of proteins. in *Chemical Changes During Processing and Storage of Foods* (eds. Rodriguez-Amaya, D. B. & Amaya-Farfan, J.) 85–123 (Academic Press, 2021).
89. Williams, B. J., Barlow, C. K., Kmiec, K. L., Russell, W. K. & Russell, D. H. Negative ion fragmentation of cysteic acid containing peptides: cysteic acid as a fixed negative charge. *J. Am. Soc. Mass Spectrom.* **22**, 1622–1630 (2011).
90. Yao, Y., Docter, M., van Ginkel, J., Ridder, D. de & Joo, C. Single-molecule protein sequencing through fingerprinting: computational assessment. *Phys. Biol.* **12**, 055003 (2015).
91. Swaminathan, J., Boulgakov, A. A. & Marcotte, E. M. A Theoretical Justification for Single Molecule Peptide Sequencing. *PLoS Comput. Biol.* **11**, e1004080 (2015).
92. Ohayon, S., Girsault, A., Nasser, M., Shen-Orr, S. & Meller, A. Simulation of single-protein nanopore sensing shows feasibility for whole-proteome identification. *PLoS Comput. Biol.* **15**, e1007067 (2019).
93. Smith, M. B., Simpson, Z. B. & Marcotte, E. M. Amino acid sequence assignment from single molecule peptide sequencing data using a two-stage classifier. *PLoS Comput. Biol.* **19**, e1011157 (2023).
94. Mapes, J. H. *et al.* Robust and scalable single-molecule protein sequencing with fluorosequencing. *bioRxivorg* (2023) doi:10.1101/2023.09.15.558007.
95. Smith, M. B. *et al.* Estimating error rates for single molecule protein sequencing experiments. *PLoS Comput. Biol.* **20**, e1012258 (2024).
96. Pham, T. L. H. *et al.* Peptide sequencing via protein language models. *arXiv [q-bio.BM]* (2024) doi:10.48550/arXiv.2408.00892.
97. Thavarajah, R., Mudimbaimannar, V. K., Elizabeth, J., Rao, U. K. & Ranganathan, K. Chemical and physical basics of routine formaldehyde fixation. *J. Oral Maxillofac. Pathol.* **16**, 400–405 (2012).
98. Gambarotto, D. *et al.* Imaging cellular ultrastructures using expansion microscopy (U-ExM).

Nat. Methods **16**, 71–74 (2019).

99. Truckenbrodt, S., Sommer, C., Rizzoli, S. O. & Danzl, J. G. A practical guide to optimization in X10 expansion microscopy. *Nat. Protoc.* **14**, 832–863 (2019).
100. Metz, B. *et al.* Identification of formaldehyde-induced modifications in proteins: reactions with model peptides. *J. Biol. Chem.* **279**, 6235–6243 (2004).
101. Bucur, O. *et al.* Nanoscale imaging of clinical specimens using conventional and rapid-expansion pathology. *Nat. Protoc.* **15**, 1649–1672 (2020).
102. Schnell, U., Dijk, F., Sjollema, K. A. & Giepmans, B. N. G. Immunolabeling artifacts and the need for live-cell imaging. *Nat. Methods* **9**, 152–158 (2012).
103. Chung, K. *et al.* Structural and molecular interrogation of intact biological systems. *Nature* **497**, 332–337 (2013).
104. Klimas, A. *et al.* Magnify is a universal molecular anchoring strategy for expansion microscopy. *Nat. Biotechnol.* **41**, 858–869 (2023).
105. Ku, T. *et al.* Multiplexed and scalable super-resolution imaging of three-dimensional protein localization in size-adjustable tissues. *Nat. Biotechnol.* **34**, 973–981 (2016-9).
106. Cui, Y. *et al.* Expansion microscopy using a single anchor molecule for high-yield multiplexed imaging of proteins and RNAs. *PLoS One* **18**, e0291506 (2023).
107. Keil, B. *Specificity of Proteolysis*. (Springer, Berlin, Germany, 2012).
108. Morihara, K. & Tsuzuki, H. Specificity of proteinase K from Tritirachium album limber for synthetic peptides. *Agricultural and biological chemistry* **39**, 1489–1492 (1975).
109. Saenger, W. Proteinase K. in *Handbook of Proteolytic Enzymes* 3240–3242 (Elsevier, 2013).
110. Yu, C.-C. J. *et al.* Expansion microscopy of C. elegans. *Elife* **9**, (2020).
111. Wang, U.-T. T. *et al.* Protein and lipid expansion microscopy with trypsin and tyramide signal amplification for 3D imaging. *Sci. Rep.* **13**, 21922 (2023).
112. Kang, J. *et al.* Multiplexed expansion revealing for imaging multiprotein nanostructures in healthy and diseased brain. *Nat. Commun.* **15**, 9722 (2024).
113. Valdes, P. A. *et al.* Improved immunostaining of nanostructures and cells in human brain

specimens through expansion-mediated protein decrowding. *Sci. Transl. Med.* **16**, eabo0049 (2024).

114. Olsen, J. V., Ong, S.-E. & Mann, M. Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Mol. Cell. Proteomics* **3**, 608–614 (2004).
115. Raijmakers, R., Neerincx, P., Mohammed, S. & Heck, A. J. R. Cleavage specificities of the brother and sister proteases Lys-C and Lys-N. *Chem. Commun. (Camb.)* **46**, 8827–8829 (2010).
116. Ebeling, W. *et al.* Proteinase K from Tritirachium album limber. *Eur. J. Biochem.* **47**, 91–97 (1974).
117. Bonissone, S., Gupta, N., Romine, M., Bradshaw, R. A. & Pevzner, P. A. N-terminal Protein Processing: A Comparative Proteogenomic Analysis. *Mol. Cell. Proteomics* **12**, 14–28 (2013-1).
118. Deng, S. & Marmorstein, R. Protein N-terminal acetylation: Structural basis, mechanism, versatility, and regulation. *Trends Biochem. Sci.* **46**, 15–27 (2021).
119. Saveliev, S. *et al.* Trypsin/Lys-C protease mix for enhanced protein mass spectrometry analysis. *Nat. Methods* **10**, i–ii (2013).
120. Kashina, A. S. & Yates, J. R., Iii. Analysis of arginylated peptides by subtractive Edman degradation. *Methods Mol. Biol.* **2620**, 153–155 (2023).
121. Khoury, G. A., Baliban, R. C. & Floudas, C. A. Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci. Rep.* **1**, 90 (2011).
122. Landry, J. P., Ke, Y., Yu, G.-L. & Zhu, X. D. Measuring affinity constants of 1450 monoclonal antibodies to peptide targets with a microarray-based label-free assay platform. *J. Immunol. Methods* **417**, 86–96 (2015).

Acknowledgements

We thank E. Anslyn, J. Choi, A. Sinha, J. Chang, J. Mou, R. Raines, M. Kumar, S. Babakhanova for thoughtful discussions. We are grateful to J. Chang, J. Yao for delivering the protein materials. For funding, E.S.B. acknowledges Ashar Aziz, Lisa Yang, John Doerr, HHMI, Jed McCaleb, James Fickel, the ERC Synergy Grant under the European Union's Horizon 2020 research and innovation programme (grant agreement No 835102), NIH R01AG087374, R01EB024261, 1R01AG070831, Tom Stocky and Avni Shah, Kathleen Octavio, Lore McGovern, Open Philanthropy and Good Ventures. N.F.P acknowledges funding from NIH (R00GM135519). C.M.M. acknowledges MIT Health and Life Sciences Collaborative Graduate Fellowship (MIT HEALS) 2025/2026, Friends of the McGovern Institute Student Fellowship 2024/2025, MIT Center of Neurobiological Engineering Training Program 2023/2024, Brain and Cognitive Sciences (BCS) Hilibrand Graduate Student Fellowship 2021-2024. S.Z.T. acknowledges support from the Brain and Cognitive Sciences (BCS) Singleton Fellowship (2024/2025), the BCS Hubert J. P. Schoemaker (1976) Fellowship (2024–2025), and the MIT Praecis Presidential Fellowship (2023–2024). S.W. was supported by MIT Presidential Graduate Fellowship and Weedon Alzheimer's Graduate Fellowship.

Author information

Authors and Affiliations

Contributions

C.M.M. and E.S.B. led the study, receiving early valuable insights from L.L.K, as well as early experimental input from C.Z., H.W., D.M.E., S.W.. C.Z. designed the initial experimental framework for in-gel Edman degradation. C.M.M., H.W., S.Z.T. designed the final experimental framework and the final protocol for in-gel Edman degradation. C.Z., C.M.M., H.W. contributed to the early conceptualization of the study, and performed early feasibility studies. Gel size changes were performed by C.M.M. and S.Z.T., and mass spectrometry was performed by C.M.M, S.Z.T., J.S. with experimental insights from C.Z., L.L.K.. Data analysis for gel size change and mass spectrometry was performed by C.M.M and S.Z.T.. Developing the in-gel Edman degradation protocol was done by C.M.M with help from H.W., S.Z.T., J.S. and J.P., and early input from C.Z., S.W., L.L.K.. ClpS binder protein and binder insight was provided by N.F.P and M.D.. Testing of ClpS binders was done by C.M.M., with early input from H.W. Amino acid oxidation experiments and data analysis was performed by C.M.M and L.E.. The creation and initial work on ClickP was performed by D.M.E. and A.G.C.. The development and testing of ClickP-AA binders was performed by D.M.E., A.G.C., E.W., and S.D.. Testing of Glyphic binders was done by C.M.M., with input from Glyphic. Imaging and imaging data analysis was performed by C.M.M.. Computational modeling was designed and developed by J.P. and C.M.M, performed by J.P., and written by C.M.M., J.P., E.S.B.. C.M.M and E.S.B wrote and edited other sections of the manuscript. E.S.B. supervised the project.

Corresponding authors

Correspondence to Ed Boyden.

Ethics declarations

ESB is a co-inventor on multiple patents related to expansion microscopy, and co-founder of a company seeking commercial interests in the space. CMM, CZ and ESB are co-inventors on a pending patent related to *in situ* single-molecule peptide sequencing.

Supplementary Materials

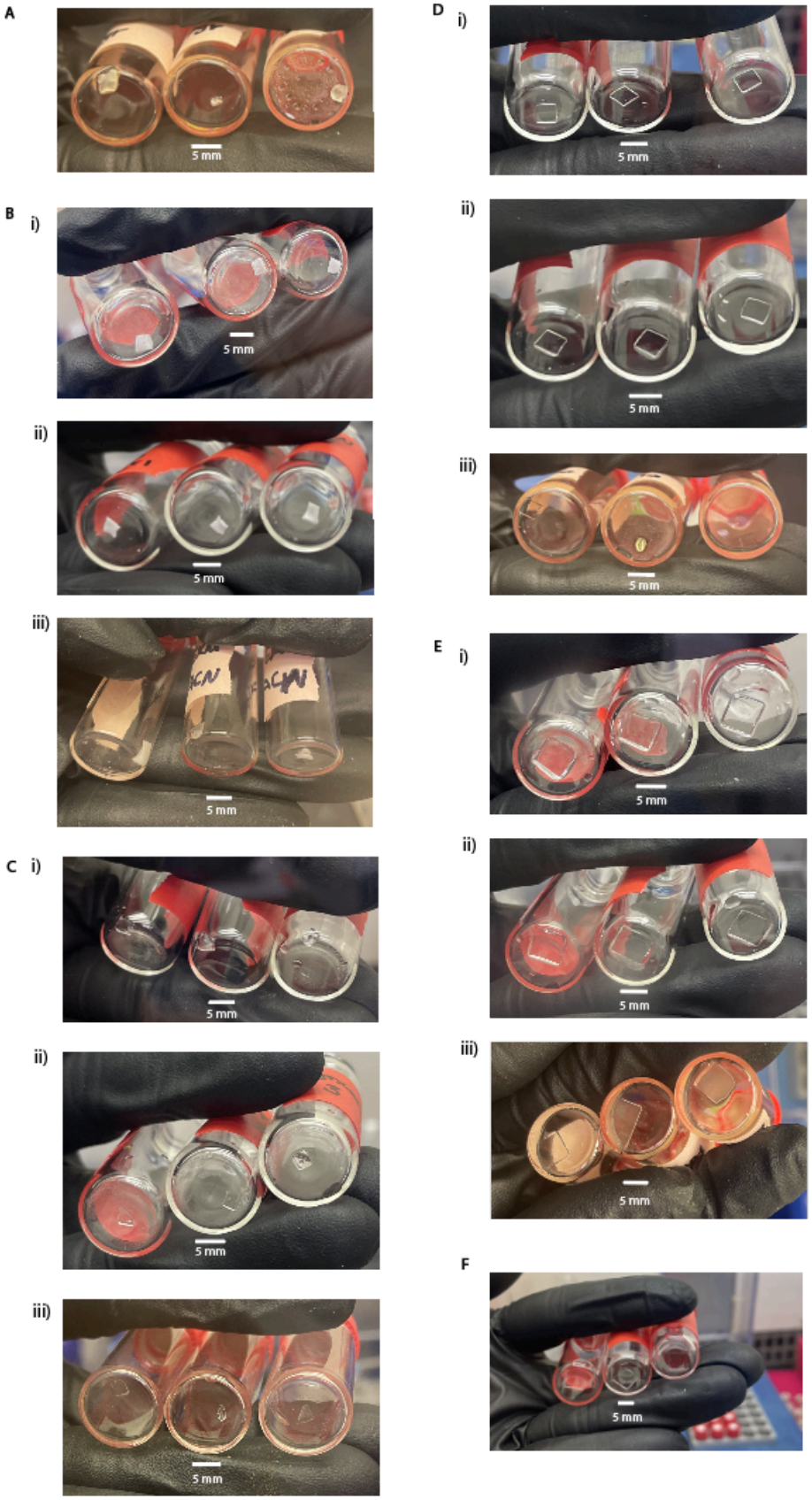
Principles of *in situ* protein sequencing: expansion microscopy-adapted Edman degradation and amino acid recognition

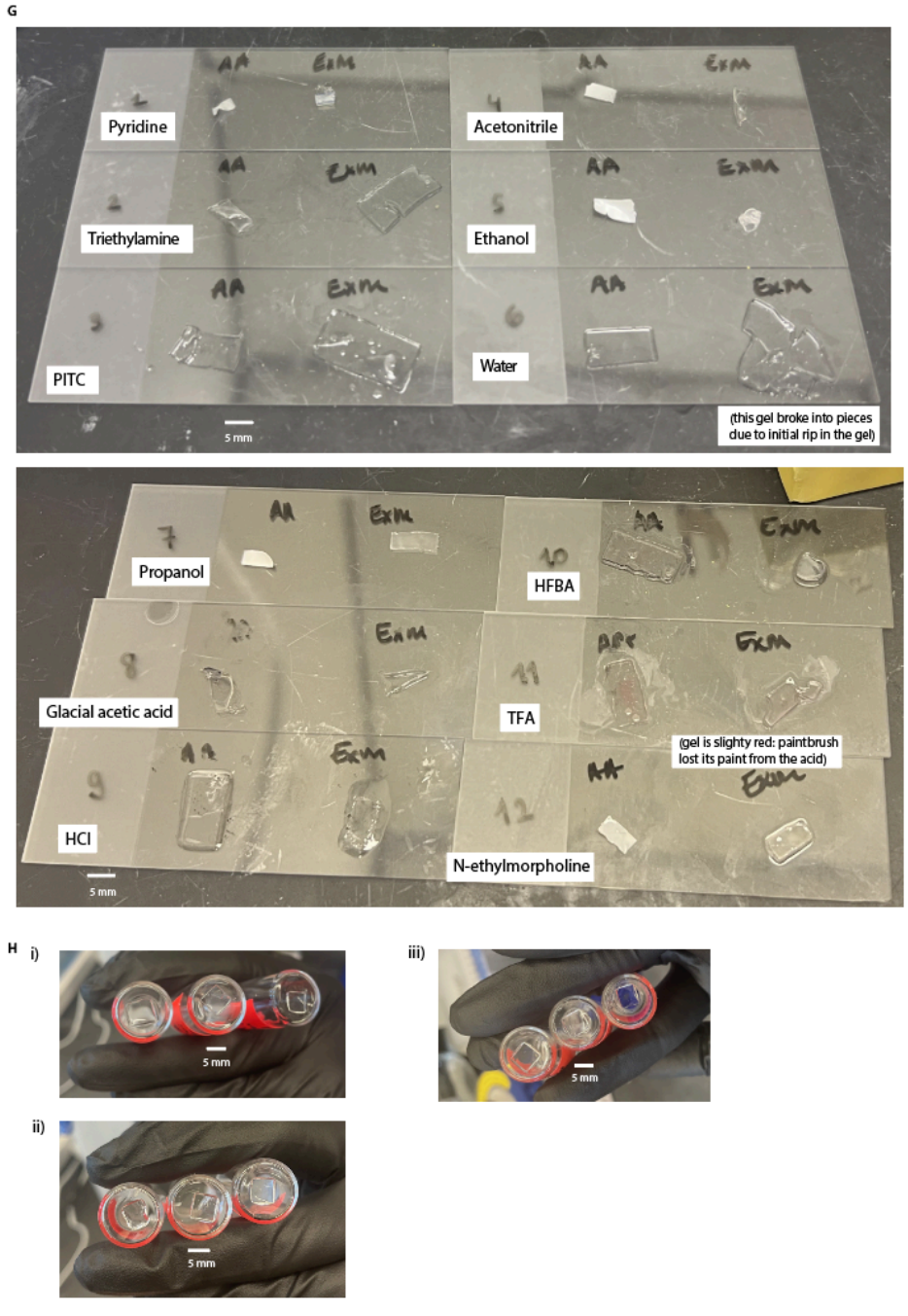
- Supplementary Materials
  - Supplementary Figures
    - Supplementary Figure 1
    - Supplementary Figure 2
    - Supplementary Figure 3
    - Supplementary Figure 4
    - Supplementary Figure 5
    - Supplementary Figure 6
    - Supplementary Figure 7
    - Supplementary Figure 8
    - Supplementary Figure 9
    - Supplementary Figure 10
    - Supplementary Figure 11
    - Supplementary Figure 12
    - Supplementary Figure 13
    - Supplementary Figure 14
    - Supplementary Figure 15
    - Supplementary Figure 16
    - Supplementary Figure 17
    - Supplementary Figure 18
    - Supplementary Figure 19
  - Supplementary Notes
    - Supplementary Note 1
    - Supplementary Note 2
    - Supplementary Note 3
    - Supplementary Note 4
    - Supplementary Note 5
    - Supplementary Note 6
    - Supplementary Note 7
    - Supplementary Note 8
    - Supplementary Note 9
    - Supplementary Note 10
  - Supplementary Tables
    - Supplementary Table 1
    - Supplementary Table 2

Supplementary Table 3
Supplementary Table 4
Supplementary Table 5
Supplementary Table 6
Supplementary Table 7
Supplementary Table 8
Supplementary Table 9
Supplementary Table 10
Supplementary Table 11
Supplementary Table 12

Supplementary Figures

Supplementary Figure 1





Supplementary Figure 1
Images of various gels (acrylamide, ExM, ExMre) after solvent with/without PITC. All gels were cut to 5 mm x 5 mm (the height depended on the gel type, ~0.3 mm for acrylamide, ~0.9 mm for ExM, ~0.7 mm for ExMre, i.e., a final gel volume of 7.8 µL, 22.8 µL and 17.5 µL, respectively) and washed 3 times with 1X PBS at room temperature (RT). ExM gels were made using 12:4:2:1:1 of StockX:water:2% BIS (w/v):10% APS (w/v):10% TEMED (v/v), as depicted in Fig. 2**b** (see Methods: ExM gel: making the empty gel gelation mixture for details), unless otherwise stated (all ratios throughout are of volumes added, unless otherwise indicated). For the gels treated with PITC to solvent (1:9 ratio PITC:solvent), they were submerged 30 min in 270 µL of solvent at 50 °C, and another 30 min at 50 °C with an additional 30 µL of PITC without removing the original solvent. All

solution was then removed, and gels were imaged in glass vials, from the bottom (**Supplementary Figure 1a, biii, ciii, diii, eiii**). For all the other gels, 300 µL was added to the gels at 50 °C for 30 min unless otherwise noted, and the gel flat/top side surface size was measured through the glass vial (**Supplementary Figure 1bi, ci, di, ei, f, hi-iii**); for the PITC ratio 1:1000 PITC:solvent case, the previous solution was then removed, and then 300 µL of a fresh solution was then added to the same gel, with PITC at a final ratio of 1:1000 ratio PITC:solvent (again, at 50 °C for 30 min). The solution was removed, and gels were similarly measured through the bottom of the glass vial after removal (**Supplementary Figure 1bii, cii, dii, eii**). In more detail, the gel flat/top side surface size was measured through the bottom of the glass vial after solution removal. In cases where the gel tore, the edges were carefully pushed back together prior to measurement. When the gel folded into a solid form and could not be unfolded, measurements were taken from the accessible edges by rotating the vial to obtain the most representative size and calculate the flat/top surface size as if it were unfolded (**Supplementary Figure 1a, ciii, diii)** . Flat/top surface sizes of gels are plotted in **Supplementary Figure 2**. Scale bar in white (using the central vial from the image as the reference for **Supplementary Figure 1a-f, h**; using the bottom leftmost microscope slide as reference for **Supplementary Figure 1g**; please note that given the 3-d nature of the vials and specimens, this means the scale bar is approximate for the other specimens): 5 mm.

(A) Gels submerged in pyridine followed by PITC to pyridine (1:9 ratio PITC:pyridine). Different gels are in three separate glass vials: a 9% acrylamide gel (left), ExM gel (middle) and ExMre gel (right).

(B) Gels submerged in acetonitrile (ACN) with/without PITC (1:9 and 1:1000 ratios PITC:ACN) (i) ExMre gels submerged in ACN without PITC. 3 different ExMre gels in separate glass vials, from different gelation solutions. (ii) Same gels as (i), after removal of solution and further submerging the same gels in fresh solution of PITC to ACN (1:1000 ratio PITC:ACN). (iii) 3 different gel types, 9% acrylamide gel (left), ExM gel (middle) and ExMre gel (right), treated with ACN followed by PITC to ACN(1:9 ratio PITC:ACN).

(C) As in B, but with 1:1 pyridine to water as solvent. For panel (iii) placement of the gels differs from (B): ExMre gel (left), ExM gel (middle) and a 9% acrylamide gel (right).

(D) As in B, but with DMSO as solvent. For panel (iii): a 9% acrylamide gel (left), ExM gel (middle) and ExMre gel (right).

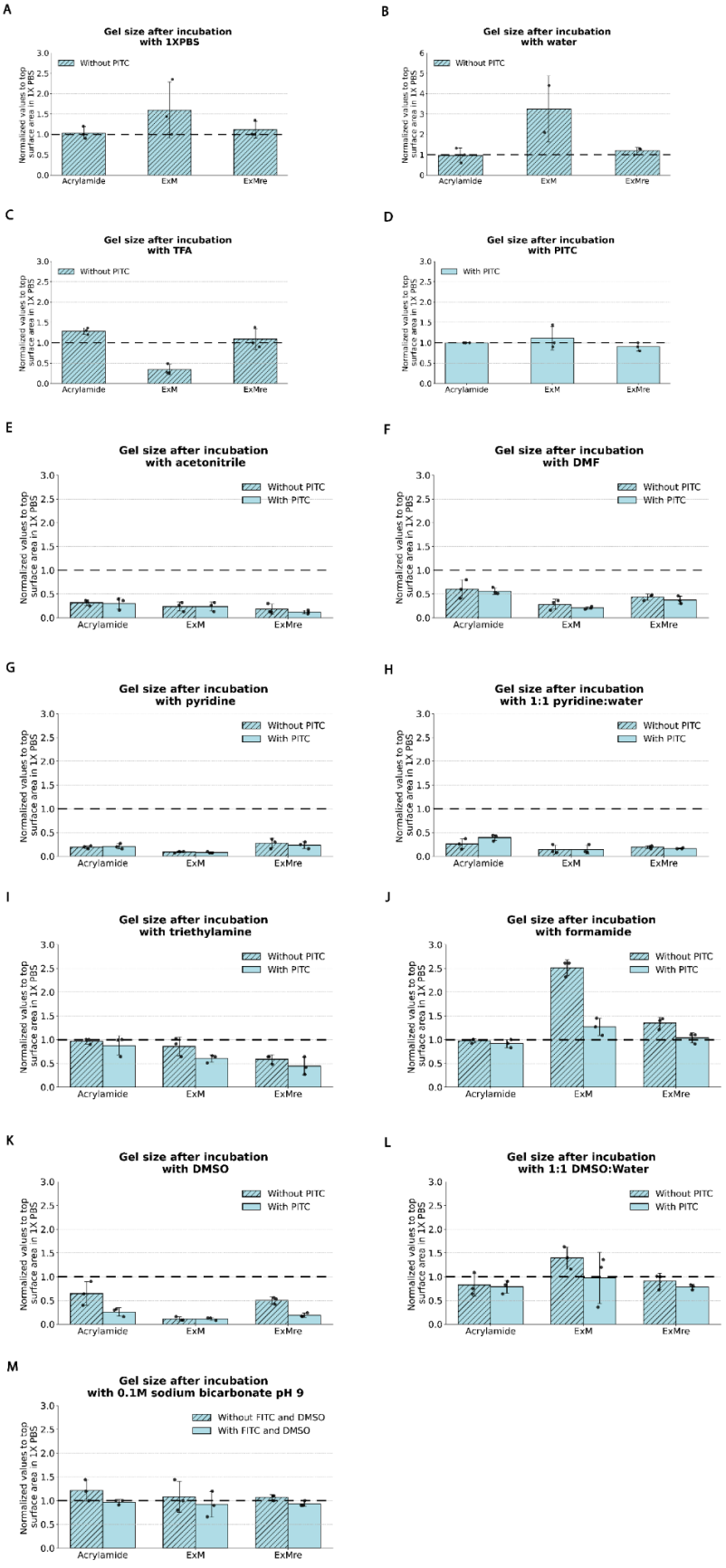
(E) As in B, but with formamide as solvent (note: the gels imaged for (i) and (ii) are not the same gels that were used for gel size change results in **Fig. 2civ**): a 9% acrylamide gel (left), ExM gel (middle) and ExMre gel (right).

(F) ExMre gels submerged in TFA. 3 different ExMre gels in separate glass vials from different gelation solutions.

(G) Other solvent testing, based on protocols listed in **Supplementary Table 2**, with various solvents used in conjugation and cleavage. 9% acrylamide gels and ExM gels (note: the gelation formula for ExM gels was different here, with 48:1:1 of StockX:10% TEMED (v/v):10% APS (w/v), see **Methods**: **ExM gel: making the empty gel gelation mixture** for StockX formula) tested in various solvents including pyridine, triethylamine, PITC, acetonitrile, ethanol, water, propanol, glacial acetic acid, hydrochloric acid (abbreviated: HCl), heptafluorobutyric acid (abbreviated: HFBA), trifluoroacetic acid (TFA), and N-ethylmorpholine. Each gel was washed twice with 1x phosphate buffered saline (PBS for 10 min) and then submerged in organic solvent for 10 min.

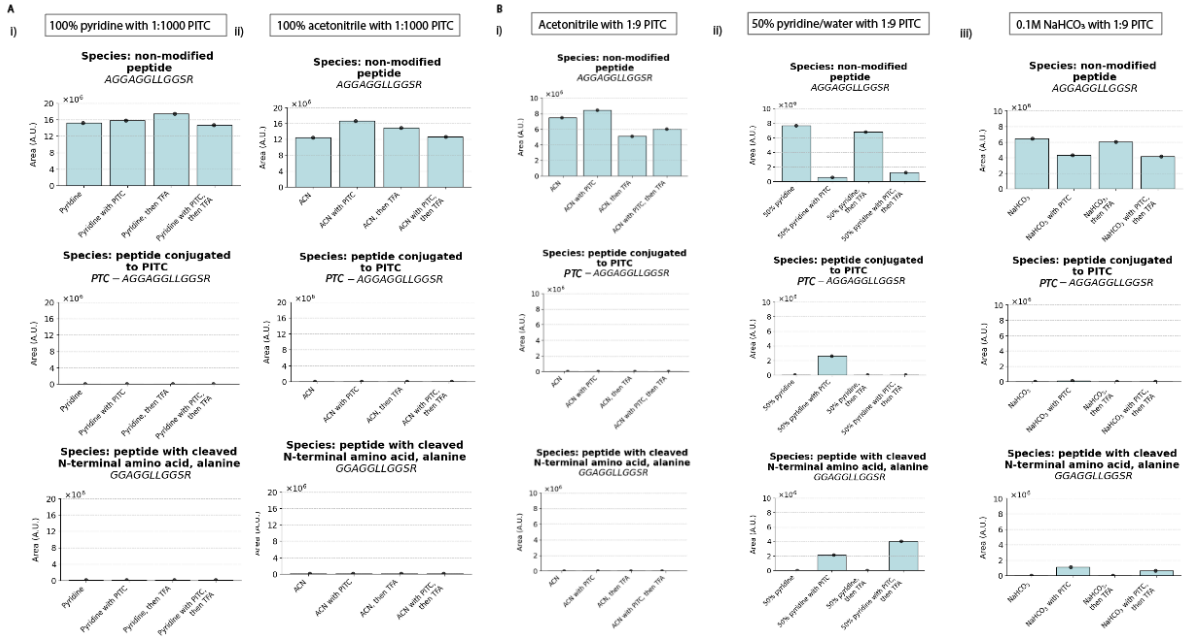
(H) ExMre gels submerged in various aqueous buffers. 3 different ExMre gels in separate glass vials from different gelation solutions. (i) 0.1 M sodium bicarbonate pH 8.5 at 50 °C. (ii) 1M Tris pH 8 at room temperature (RT). (iii) 1M Tris pH 9.5 at RT.

Supplementary Figure 2



Supplementary Figure 2: Surface flat/top side size of acrylamide, ExM, and ExMre gels in various solvents with/without PITC or FITC. Surface size of flat/top side surface size of acrylamide, ExM, and ExMre gels when placed in various solutions (some previously documented for Edman degradation, as in **Supplementary Table 2**), normalized to surface size (again, flat/top side surface size) of the gel in 1X PBS before various solution treatment. All gels were cut to 5 mm x 5 mm (the height depended on the gel type, ~0.3 mm for acrylamide, ~0.9 mm for ExM, ~0.7 mm for ExMre, i.e., a final gel volume of 7.8 µL, 22.8 µL and 17.5 µL, respectively) and washed 3 times with 1X PBS at room temperature (RT). For gels treated with 1X PBS, water, TFA, PITC, they were submerged 30 min with 300 µL of solution for 30 min at 50 °C (**Supplementary Figure 2a-d**). For the gels treated with PITC (1:9 ratio PITC:solvent), they were submerged 30 min in 270 µL of solvent at 50 °C, measured, then another 30 min at 50 °C with 30 µL of PITC in the original 270 µL solvent. All solution was then removed, and gels were measured again (**Supplementary Figure 2e-j**). For the gels treated with FITC, they were first submerged in 231 µL of 0.1 M sodium bicarbonate pH 8.5, at 50 °C, measured, then another 30 min at 50 °C with 69 µL of 10 mg/mL FITC in DMSO. (A) 1X PBS, (B) water, (C) TFA, (D) 100% PITC, (E) acetonitrile (ACN), (F) dimethylformamide (DMF), (G) 100% pyridine, (H) 1:1 pyridine to water, (I) 100% triethylamine, (J) 100% formamide, (K) dimethylsulfoxide (DMSO), (L) 1:1 DMSO to water, (M) 0.1 M sodium bicarbonate pH 8.5. "With PITC" is PITC added (1:9 ratio PITC:solution) in the given solution (unlike **Figure 2c**, where PITC is added at 1:1000 ratio PITC:solvent. (Dashed black line: surface size of the flat side of the gel in 1X PBS; error bar: standard deviation; black dots, individual experiments; n=3 gels from the same starting gelation solution).

Supplementary Figure 3

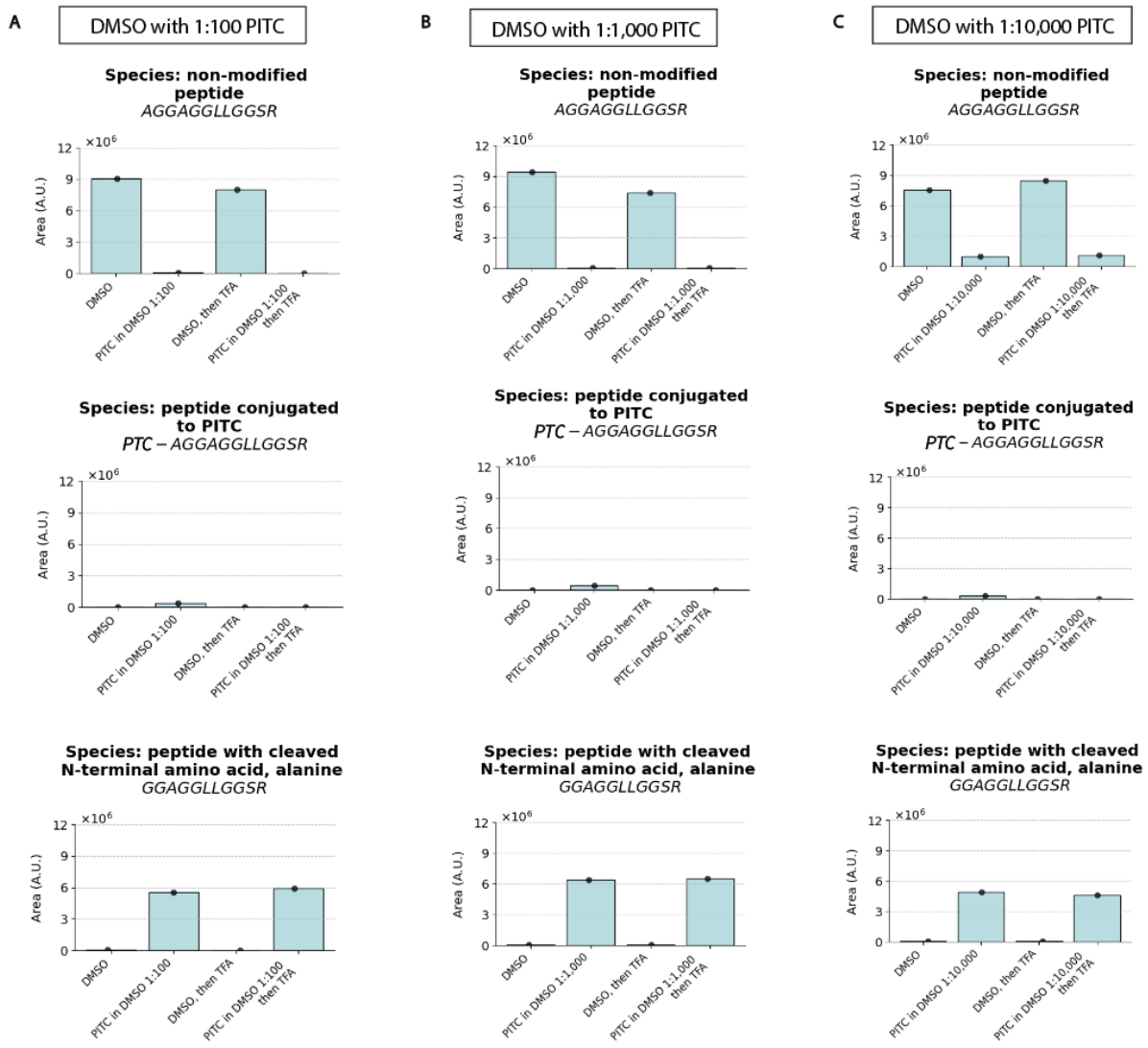


Supplementary Figure 3:

(A) In-gel Edman degradation with A15-peptide (as used in **Fig. 3a-d**) in ExMre gels. Here 100% pyridine or acetonitrile (ACN) were tested as conjugation solvent, PITC as Edman reagent (1:1000 ratio PITC:solvent). Bar graphs representing the relative abundance (arbitrary units, a.u.; all samples were processed with spiked-in controls; see **Methods**) of different peptide ion species detected on the LC/QToF. The bar graphs are obtained from comparing the area under the curve (AUC) of the chromatogram of various species (extracted based on the exact mass, see **Analysis of LC/QToF data**). The separate conditions compared include solvent only ("pyridine" or "ACN"), PITC to solvent (1:1000 ratio PITC:solvent) for 1 hour at 50 °C (e.g., "pyridine with PITC"), TFA for 30 min at 50 °C (e.g., "pyridine, then TFA"), PITC to solvent (1:1000 ratio PITC:solvent) for 1 hour at 50 °C followed by TFA for 30 min at 50 °C (e.g., "pyridine with PITC, then TFA"). The graphs in (i) were performed with 100% pyridine, and in (ii) were performed with ACN, as solvent for in-gel Edman conjugation. The relative abundance of the ion species: (top) non-modified peptide (AGGAGGLLGGSR), (middle) peptide conjugated to PITC, phenylthiocarbamoyl (PTC)-peptide (*PTC*-AGGAGGLLGGSR), and (bottom) peptide with cleaved N-terminal amino acid (GGAGGLLGGSR), were reported throughout the in-gel Edman degradation process in the various conditions (black dots, individual experiments; blue bar, mean; n=1 gelation solution).

(B) In-gel Edman degradation, as in (A), but with (i) ACN, (ii) 1:1 pyridine to water, and (iii) 0.1 M $NaHCO_3$ pH 8.5 as conjugation buffer, PITC as Edman reagent (PITC 1:9 ratio PITC:buffer). See **Methods** for further details on this protocol (black dots, individual experiments; blue bar, mean; error bar: standard deviation; n=1 gelation solution).

Supplementary Figure 4



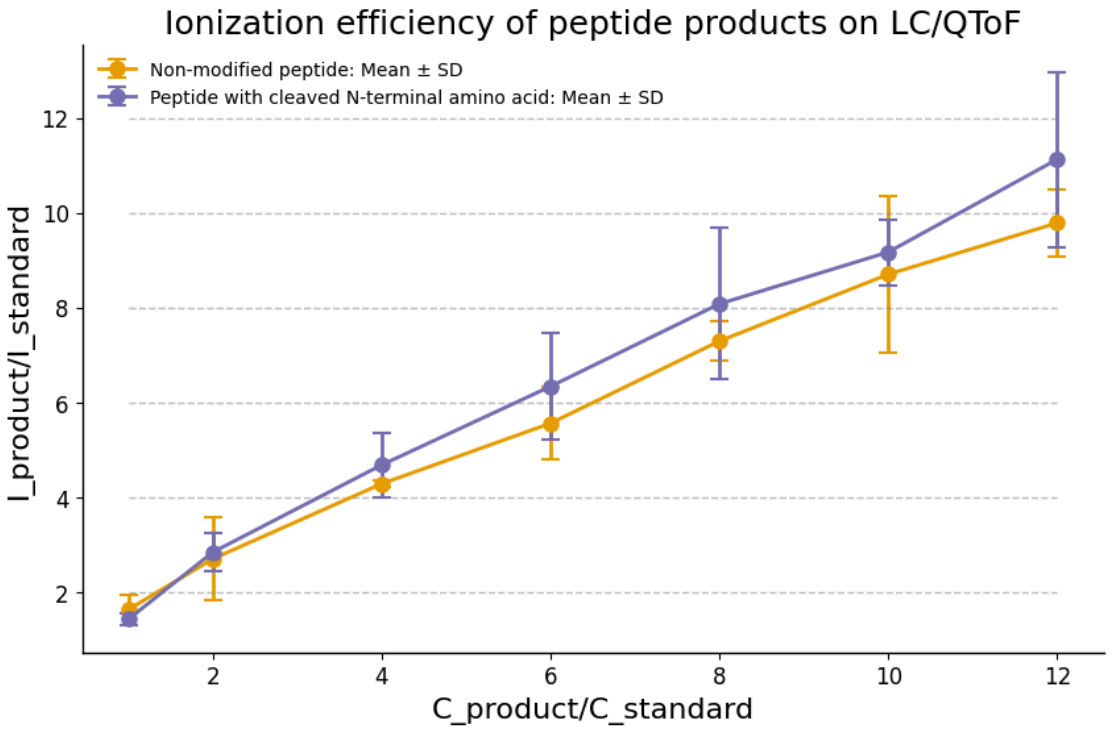
Supplementary Figure 4

(A) In-gel Edman degradation with A15 peptide in ExMre gels. Here dimethylsulfoxide (DMSO) is tested as a conjugation solvent, PITC as an Edman reagent at different ratios in the solvent. Bar graphs representing the relative abundance (arbitrary units, a.u.; all samples were processed with spiked-in controls; see **Methods**) of different peptide ion species detected on the LC/QToF. The bar graphs are obtained from comparing the area under the curve (AUC) of the chromatogram of various species (extracted based on the exact mass, see **Analysis of LC/QToF data**). The separate conditions compared include DMSO only ("DMSO"), PITC to solvent (1:100 ratio PITC:DMSO) for 1 hour at 50 °C ("PITC to DMSO 1:100"), TFA for 30 min at 50 °C (on the graph, denoted as "DMSO, then TFA"), (4) PITC to DMSO (1:100 ratio PITC:DMSO) for 1 hour at 50 °C followed by TFA for 30 min at 50 °C ("PITC to DMSO 1:100, then TFA"). The relative abundance of the ion species: (top) non-modified peptide (AGGAGGLLGGSR), (middle) peptide conjugated to PITC (*PTC*-AGGAGGLLGGSR), and (bottom) peptide with cleaved N-terminal amino acid (GGAGGLLGGSR), were reported throughout the in-gel Edman degradation process in the various conditions (black dots, individual experiments; blue bar, mean; error bar: standard deviation; n=1 gelation solution).

(B) As in A, but with PITC to DMSO (1:1,000 ratio PITC:DMSO) .

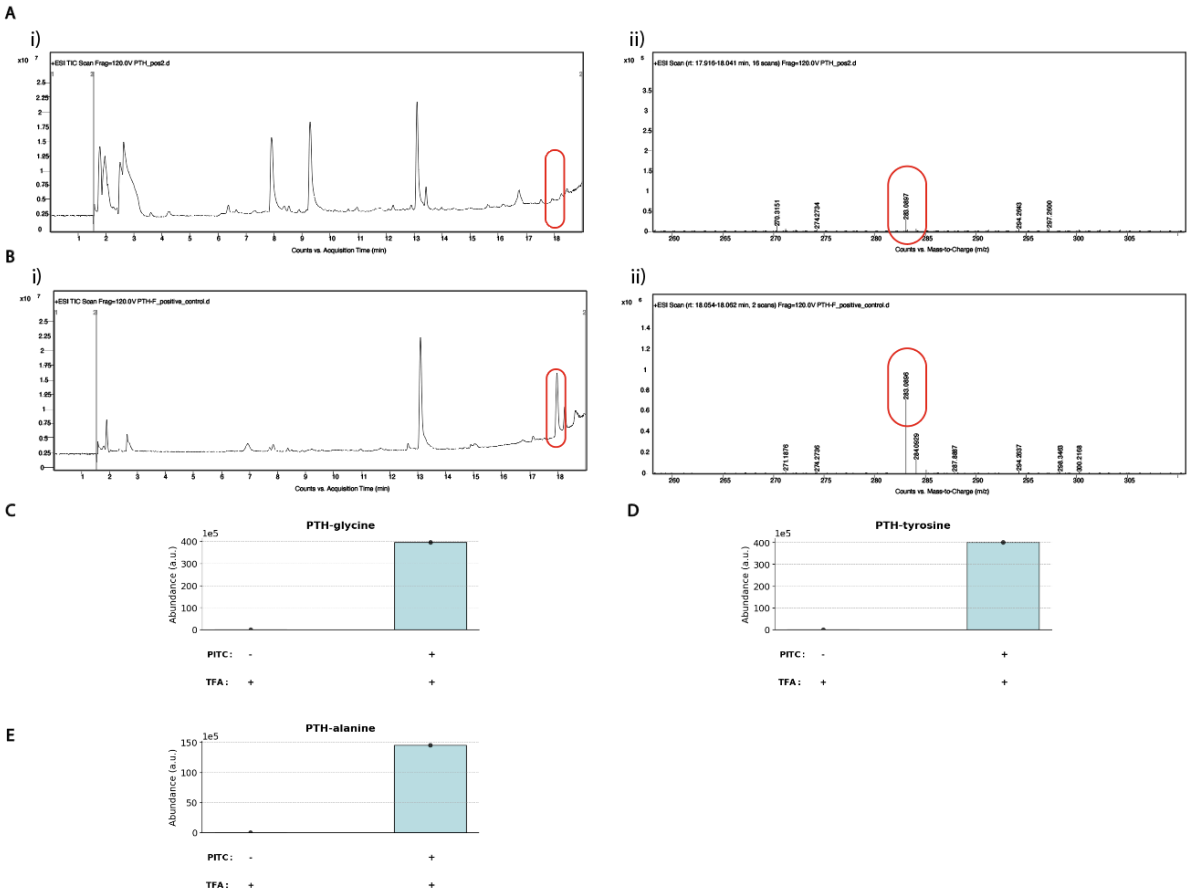
(C) As in A, but with PITC to DMSO (1:10,000 ratio PITC:DMSO) .

Supplementary Figure 5



Supplementary Figure 5: Control curve of the non-modified peptide and the peptide with cleaved N-terminal amino acid ionization efficiencies with increasing concentration. Peptide fragments derived from the peptide denoted A15 (AGGAGGLLGGSRGGK{acr}) expected after trypsinization from the gel (denoted "product"). Two products are shown as separate control curves: the first, "non-modified peptide" (AGGAGGLLGGSR; mean value across replicates: orange dots; orange vertical lines: standard deviation of given concentration, n=3 separate peptide vials) and "peptide with cleaved N-terminal amino acid" (GGAGGLLGGSR; mean value across replicates: purple dots; purple vertical lines: standard deviation of given concentration, n=3 separate peptide vials). They are prepared in solution at various "C_product" concentrations from 5 µM to 60 µM (5, 10, 20, 30, 40, 50, 60 µM) and the area under the curve (AUC) of the extracted ion chromatogram of the two species is recorded as "I_product". The A9-peptide is denoted "standard" (AGGAGK{acr}GLR) and was used as the standard with a constant "C_standard" of 5 µM and the AUC of the extracted ion chromatogram is recorded as "I_standard" (extracted based on the exact mass, see **Analysis of LC/QToF data**). The statistical significance of the two means were tested with two-sided Welch's t-test for n=3 for the two conditions (non-modified peptide and peptide with cleaved N-terminal amino acid) at each "C_product/C_standard" level. None of the conditions were statistically significant at the 95% confidence level ("C_product/C_standard": 1, t-statistic: 0.847, p-value: 0.470; "C_product/C_standard": 2, t-statistic: -0.208, p-value: 0.849; "C_product/C_standard": 4, t-statistic: -0.816, p-value: 0.499; "C_product/C_standard": 6, t-statistic: -0.808, p-value: 0.470; "C_product/C_standard": 8, t-statistic: -0.672, p-value: 0.563; "C_product/C_standard": 10, t-statistic: -0.372, p-value: 0.737; "C_product/C_standard": 12, t-statistics: -0.961, p-value: 0.418). After adjusting for multiple-comparison testing, using Holm-Bonferroni, p-values are corrected to 1.0 at each concentration level.

Supplementary Figure 6



Supplementary Figure 6: Phenylthiohydantoin (PTH) detection assay from **Fig. 3h** for F_1 peptide (FGGAGRGLGK{acr}) embedded in ExMre gels

(A) (i) Total ion chromatogram (TIC) of one of the three replicates for PTH-phenylalanine (PTH-F) detection from ExMre gels containing F_1 peptide. Peptides in ExMre gels cut 5 mm x 5 mm x 0.35 mm (8.8 µL) with ~50 µM peptide are Edman degraded, and resuspended in 1:1 acetonitrile to water before injection in the LC/QToF (y axis: abundance in arbitrary units, x axis: retention time in minutes, red oval at ~18 min when PTH-F elutes). Considering expected (from earlier experiments) ~70% cleavage of PTH-F from the peptide, ~2 pmol is injected into the LC/QToF. (ii) The associated mass spectrum of the TIC at ~18 min showing presence of PTH-F (exact mass: 282.0827, and $[M+H]^+$: 283.0905; y axis: abundance in arbitrary units, x axis: mass to charge ratio, red oval showing detection of this species).

(B) (i) TIC of the positive control with pure PTH-F in 1:1 acetonitrile to water before injection in the LC/QToF (y axis: abundance in arbitrary units, x axis: retention time in minutes, red oval at ~18 min when PTH-F elutes). About ~1 nmol is injected into the LC/QToF, compared to ~2 pmol in (a). (ii) The associated mass spectrum of the TIC at ~18 min showing presence of PTH-F (exact mass: 282.0827, and $[M+H]^+$: 283.0905; y axis: abundance in arbitrary units, x axis: mass to charge ratio).

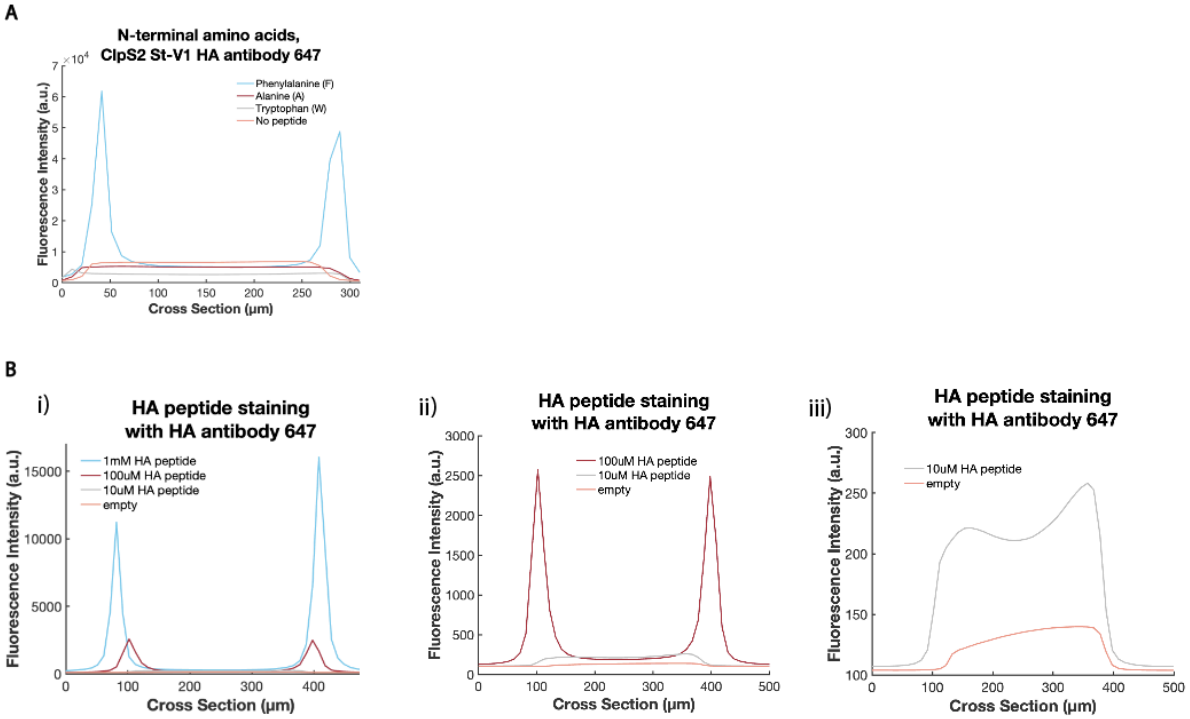
(C) PTH-glycine (PTH-G) detection of GGGAGRGLGK{acr} (abbreviated G-peptide) embedded in 9% acrylamide gels. Separate gels were subjected to various Edman conditions. The conditions included TFA only for 30 min at 50 °C, and PITC toDMSO (1:9 ratio PITC:DMSO) for 1 hour at 50 °C followed by TFA for 30 min at 50 °C. Subsequently, TFA was removed from the gels and they were immersed in 50 µL of 1:1 acetonitrile:water and agitated. Read-out was then performed by injecting the supernatant into LC/QToF using **LC Method (see Methods)**. Analysis of PTH-G abundance was performed using PTH-G exact

mass, 193.0435 ± 0.0039 Da (see **Methods for Edman degradation and PTH detection** for details). (blue bar, mean; black dots, individual experiments, n=1 gelation solution).

(D) As in (C), but with PTH-Y detection of YGGAGRGLGK{acr} (abbreviated Y-peptide) embedded in 9% acrylamide gels. Analysis of PTH-Y abundance was performed using PTH-Y exact mass, 299.0854 ± 0.0060 Da.

(E) As in (C), but with PTH-A detection of AK{N_3}GAGGLLGGSRGGK{acr} embedded in ExMre gels. Analysis of PTH-A abundance was performed using PTH-A exact mass, 207.0592 ± 0.0041 Da.

Supplementary Figure 7



Supplementary Figure 7:

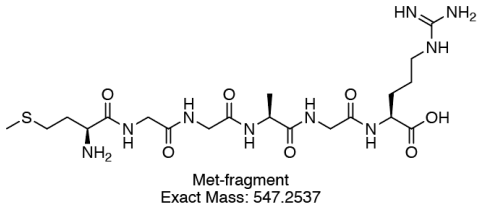
(A) 9% acrylamide gels were cast with different N-terminal peptides (XaaGGAGRGLGK{acr}, where Xaa was: F, A, W) at a concentration of ~5 mM or with no peptides embedded (labeled: "No peptide"). The gels (each with a volume of 1.2 µL) are compared for their fluorescence intensity across the thickness of the gel after staining with 20 µM ClpS2 St-V1 for 1 hour at RT, then washed for ~1 min, prior to incubating with 200 µg/mL anti hemagglutinin (HA) tag antibody 647 (HA-antibody 647) overnight (O/N) in 100 µL of solution at 4 °C in the dark. Imaging was performed using a confocal microscope with 10X objective, 100% laser power, 400 ms exposure time, 10 µm Z-steps.

(B) 9% acrylamide gels were cast with HA-tag peptide (YPYDVPDYAK{acr}) at a concentration of ~1 mM, ~100 µM and ~10 µM, and with no peptides embedded (labeled: "empty"). The gels are compared for their fluorescence intensity across the thickness of the gel after staining with 20 µM of ClpS2 St-V1 for 1 hour at RT and incubating with 20 µg/mL HA-antibody 647 O/N at 4 °C in the dark. Imaging was performed using a confocal microscope with 10X objective, 100% laser power, 100 ms exposure time, 10 µm Z-steps. (i) Shows all 4 conditions tested (1 mM, 100 µM, 10 µM, and empty). (ii) As in (i) but shows only 3 conditions (100 µM, 10 µM, and empty) with an adjusted y-axis. (iii) As in (ii) but shows only 2 conditions (10 µM and empty) with an adjusted y -axis.

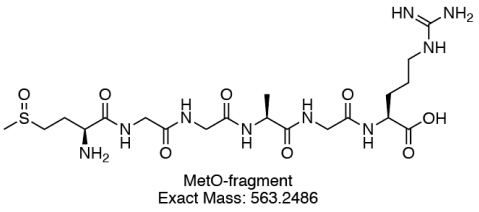
Supplementary Figure 8 Analysis of amino acid oxidation in ExM gels. Different N-termini peptides were embedded in ExM gels (XaaGGAGRGLGK{acr}, with Xaa: M, C, Y, W, F, H, P, R).

a)

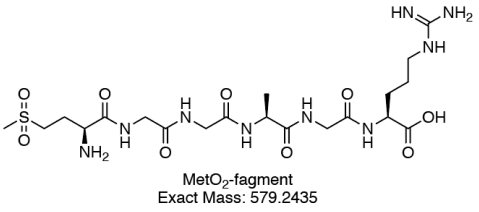
i) A)



B)



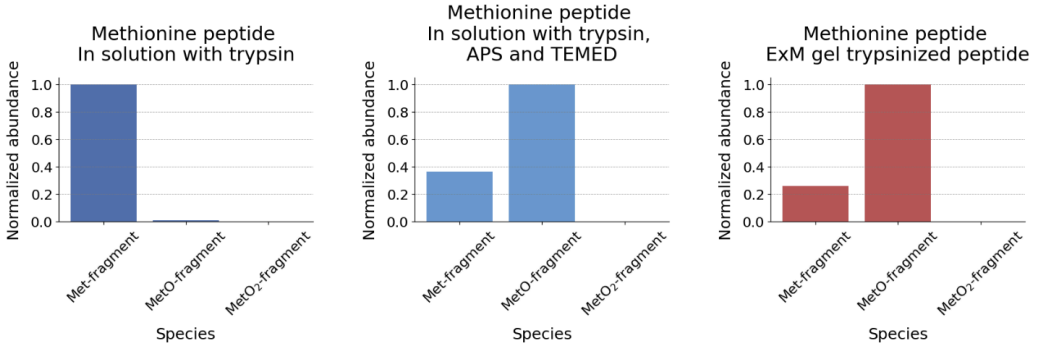
C)



ii)

Methionine	A) Met-fragment	B) MetO-fragment	C) $MetO_2$-fragment
Exact mass	*547.2537*	*563.2486*	*579.2435*
Condition 1: trypsin	19,007,365	183,009.5	366.81
Condition 2: TEMED+APS+trypsin	1,021,971	2,801,300	262.52
Condition 3: TEMED+APS+trypsin in ExM gel	3,628,434.7	13,817,063	5,788.13

iii)



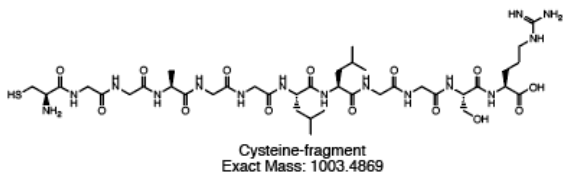
Supplementary Figure 8a Methionine and post-translational oxidation.

i) Structure of the various fragments resulting from enzymatic cleavage with trypsin from the parent peptide (sequence of parent peptide: MGGAGRGLGK{acr}), (A) "Methionine fragment" (MGGAGR, abbreviated: "Met-fragment") resulting from trypsin cleavage, (B) modification of the fragment in (a) with known methionine oxidation products, notably methionine sulfoxide, to form "Methionine sulfoxide fragment" (abbreviated: "MetO-fragment") (C), modification of the fragment in (a) with known methionine oxidation products, notably "Methionine sulfone fragment" (abbreviated: "$MetO_2$-fragment") [57].

ii) Table detailing the exact mass, and absolute abundance recorded for the different products (from A-C peptide fragments from **Supplementary Figure 8ai**) in the different conditions (Condition 1: in solution with trypsin; Condition 2: in solution with trypsin, ammonium persulfate (APS) and N,N,N',N'-Tetramethylethylenediamine (TEMED); Condition 3: ExM gel trypsinized peptide). AUC for each species in the extracted ion chromatogram was obtained as in **Methods: Analysis of LC/QToF data** and the raw total ion chromatograms (TIC) located in **Source Data**. The dash ('-') means the species was not detected when using search parameters (the extracted ion chromatogram was generated using the $[M+H]^+$ ion of the species, with a mass tolerance of ±20 ppm, with all detected peaks integrated into the final value).
iii) Normalized abundance of methionine N-terminal peptide and oxidation products in different conditions. Normalized abundance of the 3 different species (Met-fragment, MetO-fragment, $MetO_2$-fragment; as detailed in **Supplementary Figure 8ai**) in: condition 1, after trypsin cleavage in solution (left); condition 2, after addition of APS and TEMED with trypsin cleavage in solution (middle); and condition 3, after trypsin cleavage of peptides from ExM gels (condition 3; right). AUC for each species in the extracted ion chromatogram was obtained as in **Methods: Analysis of LC/QToF data** and the raw total ion chromatograms (TIC) located in **Source Data**. The abundance of each species was normalized with max absolute scaling for plotting on the bar graph (n=1 gelation solution for each condition).

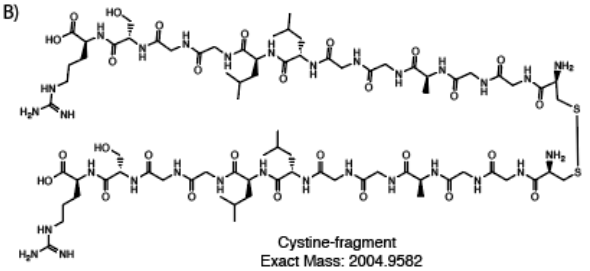
b)

i) A)



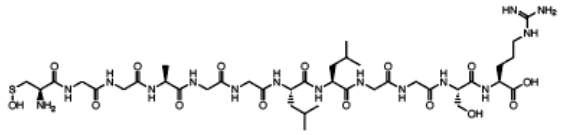
Cysteine-fragment
Exact Mass: 1003.4869

B)



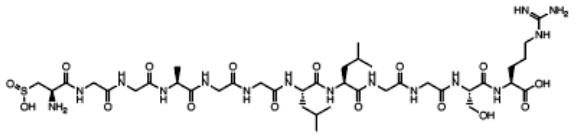
Cystine-fragment
Exact Mass: 2004.9582

C)



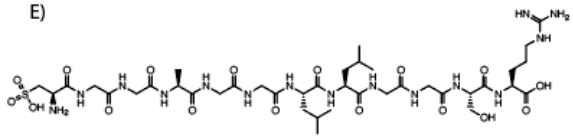
Sulfenic-fragment
Exact Mass: 1019.4818

D)



Sulfinate-fragment
Exact Mass: 1035.4767

E)

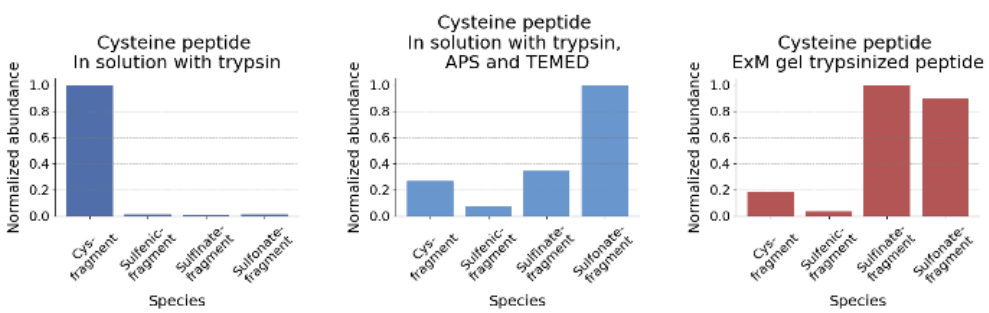


Sulfonate-fragment
Exact Mass: 1051.4717

ii)

Cysteine	A) Cys-fragment	C) Sulfenic-fragment	D) Sulfinate-fragment	E) Sulfonate-fragment
Exact mass	*1003.4869*	*1019.4818*	*1035.4767*	*1051.4717*
Condition 1: trypsin	11,412,844	162,653.3	96,467.32	132,344.9
Condition 2: TEMED+APS+trypsin	354,068.38	102,825.1	457,772.2	1,327,463
Condition 3: TEMED+APS+trypsin in ExM gel	54,427.54	10,445.3	299,666.7	268,362.1

iii)



Supplementary Figure 8b Cysteine and post-translational oxidation.

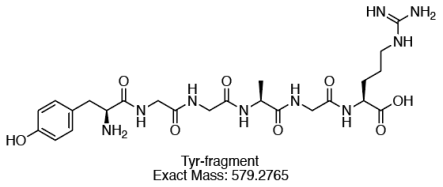
i) Structure of the various fragments resulting from enzymatic cleavage with trypsin from the parent peptide (sequence of parent peptide: CGGAGGLLGGSRGGK{acr}), (A) "Cysteine fragment" (CGGAGGLLGGSR, abbreviated: "Cys-fragment") resulting from trypsin cleavage, (B) modification of the fragment in (A) with known cysteine oxidation products, notably cystine, to form "Cystine fragment" (abbreviated: "Cystine-fragment"), (C) modification of the fragment in (A) with known cysteine oxidation products, notably "Sulfenic acid fragment" (abbreviated: "Sulfenic-fragment"), (D) modification of the fragment in (A) with known cysteine oxidation products, notably "Sulfinate fragment" (abbreviated: "Sulfinate-fragment"), (E) modification of the fragment in (a) with known cysteine oxidation products, notably "Sulfonate fragment" (abbreviated: "Sulfonate-fragment") [58].

ii) Table detailing the exact mass, and absolute abundance recorded for the different products (from A-E peptide fragments from **Supplementary Figure 8bi;** note that Cysteine-fragment is omitted due to $[M+2]^{2+}$ mass having nearly identical mass to Cys-fragment - only the mass of Cys-fragment is plotted but may include contributions from Cysteine-fragment with one carbon-13 atom) in the different conditions (Condition 1: in solution with trypsin; Condition 2: in solution with trypsin, APS and TEMED; Condition 3: ExM gel trypsinized peptide). AUC for each species in the extracted ion chromatogram was obtained as in **Methods: Analysis of LC/QToF data** and the raw total ion chromatograms (TIC) located in **Source Data**. The dash ('-') means the species was not detected when using search parameters (the extracted ion chromatogram was generated using the $[M+H]^{+}$ ion of the species, with a mass tolerance of ±20 ppm, with all detected peaks integrated into the final value).

iii) Normalized abundance of cysteine N-terminal peptide and oxidation products in different conditions. Normalized abundance of the 5 different species (Cys-fragment, Cystine-fragment, Sulfenic-fragment, Sulfinate-fragment, Sulfonate-fragment; as detailed in **Supplementary Figure 8bi**) in: condition 1, after trypsin cleavage in solution (left); condition 2, after addition of APS and TEMED with trypsin cleavage in solution (middle); and condition 3, after trypsin cleavage of peptides from ExM gels (condition 3; right). AUC for each species in the extracted ion chromatogram was obtained as in **Methods: Analysis of LC/QToF data** and the raw total ion chromatograms (TIC) located in **Source Data**. The abundance of each species was normalized with max absolute scaling for plotting on the bar graph (n=1 gelation solution for each condition).

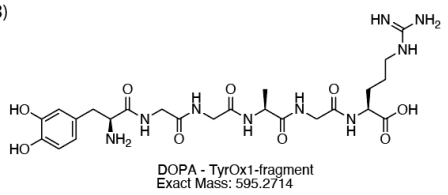
c)

i) A)

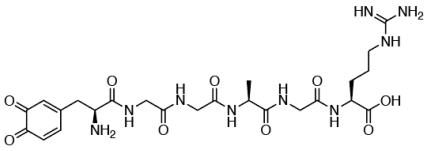


Tyr-fragment
Exact Mass: 579.2765

B)

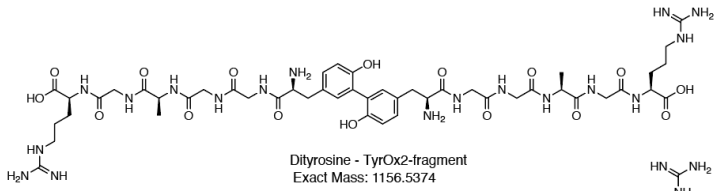


DOPA - TyrOx1-fragment
Exact Mass: 595.2714

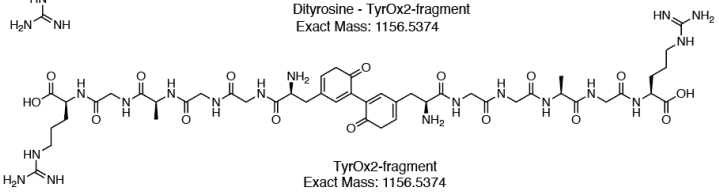


DOPA-quinone - TyrOx1-fragment
Exact Mass: 595.2714

C)



Dityrosine - TyrOx2-fragment
Exact Mass: 1156.5374

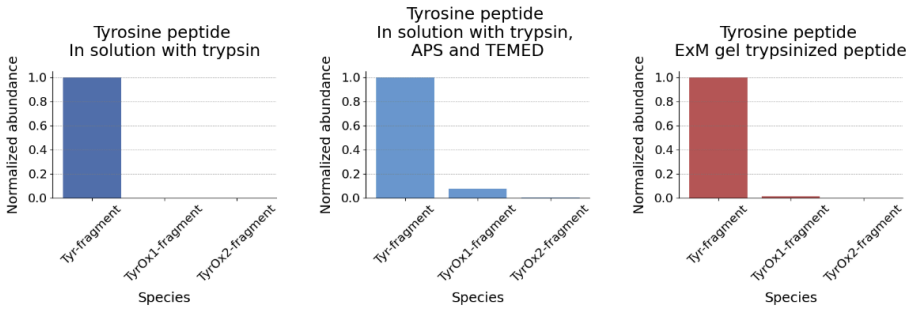


TyrOx2-fragment
Exact Mass: 1156.5374

ii)

Tyrosine	A) Tyr-fragment	B) TyrOx1-fragment	C) TyrOx2-fragment
Exact mass	*579.2765*	*595.2714*	*1156.5374*
Condition 1: trypsin	28,020,395	18,620.58	240.48
Condition 2: TEMED+APS+trypsin	13,289,453	1,054,274	71,283.85
Condition 3: TEMED+APS+trypsin in ExM gel	2,509,423.3	30,242.02	255.59

iii)



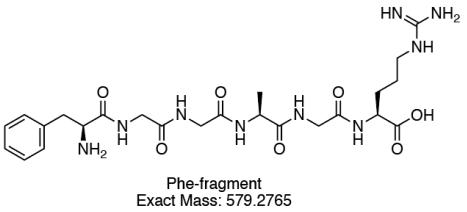
Supplementary Figure 8c Tyrosine and post-translational oxidation

i) Structure of the various fragments resulting from enzymatic cleavage with trypsin from the parent peptide (sequence of parent peptide: YGGAGRGLGK{acr}), (A) "Tyrosine fragment" (YGGAGR, abbreviated: "Tyr-fragment") resulting from trypsin cleavage, (B) modification of the fragment in (A) with known tyrosine oxidation products, notably "3,4-Dihydroxyphenylanine (DOPA) and/or dopamine quinone", to form "Tyrosine Oxidation products 1 fragment" (abbreviated: "TyrOx1-fragment"), (C) modification of the fragment in (A) with known tyrosine oxidation products, notably "Dityrosine fragment" (abbreviated: "TyrOx2-fragment") [59].

ii) Table detailing the exact mass, and absolute abundance recorded for the different products (from A-C peptide fragments from **Supplementary Figure 8ci**) in the different conditions (Condition 1: in solution with trypsin; Condition 2: in solution with trypsin, APS and TEMED; Condition 3: ExM gel trypsinized peptide). AUC for each species in the extracted ion chromatogram was obtained as in **Methods: Analysis of LC/QToF data** and the raw total ion chromatograms (TIC) located in **Source Data**. The dash ('-') means the species was not detected when using search parameters (the extracted ion chromatogram was generated using the $[M+H]^+$ ion of the species, with a mass tolerance of ±20 ppm, with all detected peaks integrated into the final value).
iii) Normalized abundance of tyrosine N-terminal peptide and oxidation products in different conditions. Normalized abundance of the 3 different species (Tyr-fragment, TyrOx1-fragment, TyrOx2-fragment; as detailed in **Supplementary Figure 8ci**) in: condition 1, after trypsin cleavage in solution (left); condition 2, after addition of APS and TEMED with trypsin cleavage in solution (middle); and condition 3, after trypsin cleavage of peptides from ExM gels (condition 3; right). AUC for each species in the extracted ion chromatogram was obtained as in **Methods: Analysis of LC/QToF data** and the raw total ion chromatograms (TIC) located in **Source Data**. The abundance of each species was normalized with max absolute scaling for plotting on the bar graph (n=1 gelation solution for each condition).

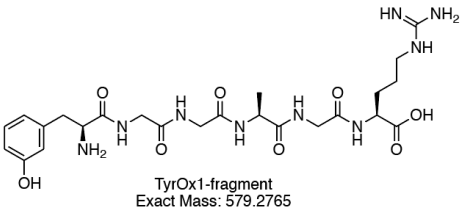
d)

i) A)



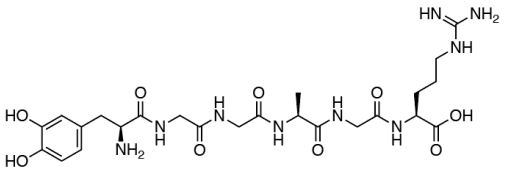
Phe-fragment
Exact Mass: 579.2765

B)

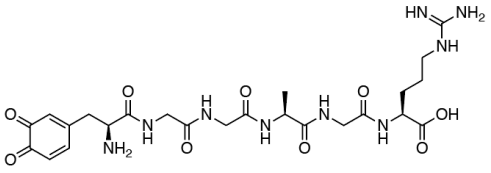


TyrOx1-fragment
Exact Mass: 579.2765

C)



DOPA - PheOx2-fragment
Exact Mass: 595.2714

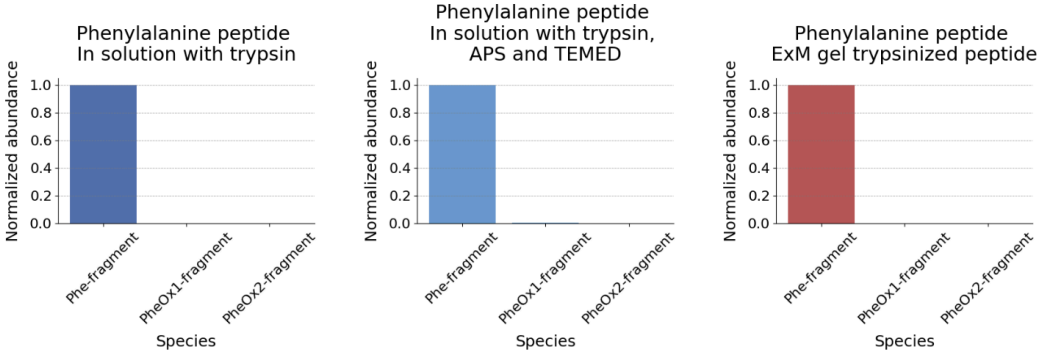


DOPA-quinone - PheOx2-fragment
Exact Mass: 595.2714

ii)

Phenylalanine	A) Phe-fragment	B) PheOx1-fragment	C) PheOx2-fragment
Exact mass	*563.2816*	*579.2765*	*595.2714*
Condition 1: trypsin	39,692,812	230.34	502.13
Condition 2: TEMED+APS+trypsin	27,486,226	128,252.7	2,134.02
Condition 3: TEMED+APS+trypsin in ExM gel	18,738,042	1,166.65	-

iii)



Supplementary Figure 8d Phenylalanine and post-translational oxidation

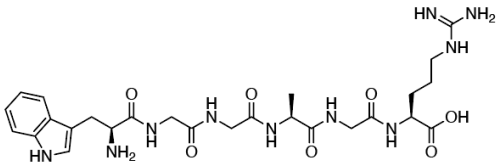
i) Structure of the various fragments resulting from enzymatic cleavage with trypsin from the parent peptide (sequence of parent peptide: FGGAGRGLGK{acr}), (A) "Phenylalanine fragment" (FGGAGR, abbreviated: "Phe-fragment") resulting from trypsin cleavage, (B) modification of the fragment in (A) with known phenylalanine oxidation products, notably "Meta- or ortho-tyrosine", to form "Oxidation products 1 fragment" (abbreviated: "PheOx1-fragment"), (C) modification of the fragment in (A) with known phenylalanine oxidation products, notably "Dihydroxyphenylalanine (DOPA) and/or dopamine quinone fragment" (abbreviated: "PheOx2-fragment") [60].

ii) Table detailing the exact mass, and absolute abundance recorded for the different products (from A-C peptide fragments from **Supplementary Figure 8di**) in the different conditions (Condition 1: in solution with trypsin; Condition 2: in solution with trypsin, APS and TEMED; Condition 3: ExM gel trypsinized peptide). AUC for each species in the extracted ion chromatogram was obtained as in **Methods: Analysis of LC/QToF data** and the raw total ion chromatograms (TIC) located in **Source Data**. The dash ('-') means the species was not detected when using search parameters (the extracted

ion chromatogram was generated using the $[M+H]^+$ ion of the species, with a mass tolerance of ±20 ppm, with all detected peaks integrated into the final value).
iii) Normalized abundance of phenylalanine N-terminal peptide and oxidation products in different conditions. Normalized abundance of the 3 different species (Phe-fragment, PheOx1-fragment, PheOx2-fragment; as detailed in **Supplementary Figure 8di**) in: condition 1, after trypsin cleavage in solution (left); condition 2, after addition of APS and TEMED with trypsin cleavage in solution (middle); and condition 3, after trypsin cleavage of peptides from ExM gels (condition 3; right). AUC for each species in the extracted ion chromatogram was obtained as in **Methods: Analysis of LC/QToF data** and the raw total ion chromatograms (TIC) located in **Source Data**. The abundance of each species was normalized with max absolute scaling for plotting on the bar graph (n=1 gelation solution for each condition).

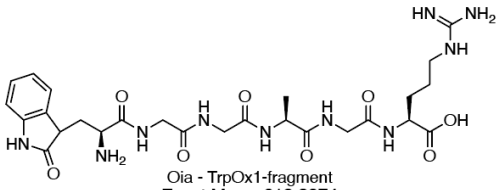
e)

i) A)

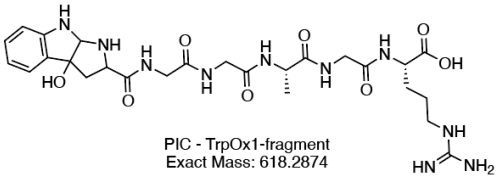


Trp-fragment
Exact Mass: 602.2925

B)

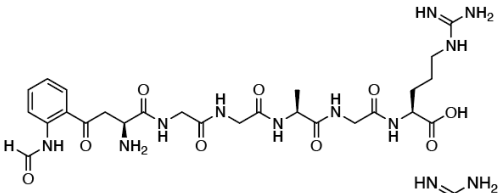


Oia - TrpOx1-fragment
Exact Mass: 618.2874

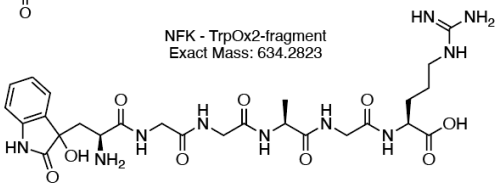


PIC - TrpOx1-fragment
Exact Mass: 618.2874

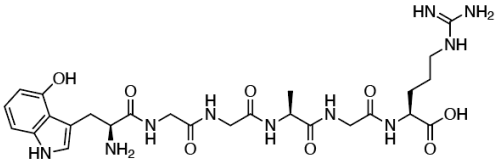
C)



NFK - TrpOx2-fragment
Exact Mass: 634.2823

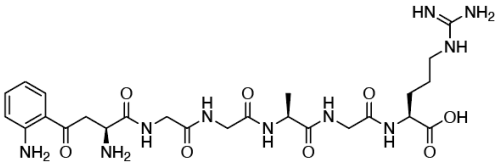


DiOia - TrpOx2-fragment
Exact Mass: 634.2823



Hydroxytryptophan - TrpOx1-fragment
Exact Mass: 618.2874

D)

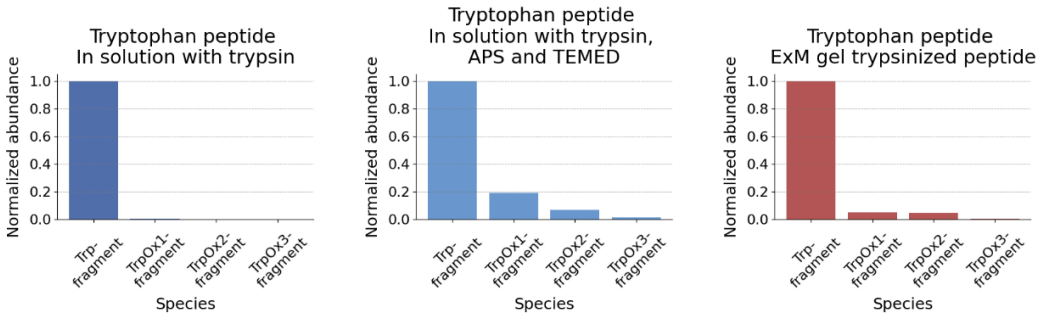


KN - TrpOx3-fragment
Exact Mass: 606.2874

ii)

Tryptophan	A) Trp-fragment	B) TrpOx1-fragment	C) TrpOx2-fragment	D) TrpOx3-fragment
Exact mass	*602.2925*	*618.2874*	*634.2823*	*606.2874*
Condition 1: trypsin	37,058,997	91,994.91	9,823.04	20,127.28
Condition 2: TEMED+APS+trypsin	8,511,649.4	1,601,600	582,627	107,597.4
Condition 3: TEMED+APS+trypsin in ExM gel	2,119,325.3	108,509.7	97,603.25	11,329.72

iii)



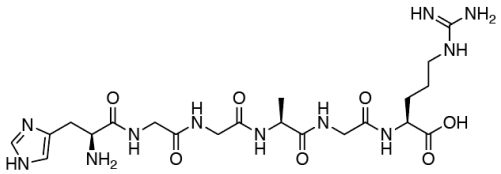
Supplementary Figure 8e Tryptophan and post-translational oxidation.

i) Structure of the various fragments resulting from enzymatic cleavage with trypsin from the parent peptide (sequence of parent peptide: WGGAGRGLGK{acr}), (A) "Tryptophan fragment"

(WGGAGR, abbreviated: “Trp-fragment”) resulting from trypsin cleavage, (B) modification of the fragment in (A) with known tryptophan oxidation products, notably “Oxindolylalanine (Oia), and/or 3-Hydroxypyrroloindole carboxylic acid (PIC), and/or hydroxytryptophan”, to form “Oxidation products 1 fragment” (abbreviated: “TrpOx1-fragment”), (C) modification of the fragment in (A) with known tryptophan oxidation products, notably “N-formylkynurenine (NFK) and/or dioxindolylalanine (DiOia)” (abbreviated: “TrpOx2-fragment”), (D) modification of the fragment in (A) with known tryptophan oxidation products, notably “Kynurenin (KN)” (abbreviated: “TrpOx3-fragment”) [61].
ii) Table detailing the exact mass, and absolute abundance recorded for the different products (from a-c peptide fragments from **Supplementary Figure 8ei**) in the different conditions (Condition 1: in solution with trypsin; Condition 2: in solution with trypsin, APS and TEMED; Condition 3: ExM gel trypsinized peptide). AUC for each species in the extracted ion chromatogram was obtained as in **Methods: Analysis of LC/QToF data** and the raw total ion chromatograms (TIC) located in **Source Data**. The dash (‘-’) means the species was not detected when using search parameters (the extracted ion chromatogram was generated using the $[M+H]^+$ ion of the species, with a mass tolerance of ±20 ppm, with all detected peaks integrated into the final value).
iii) Normalized abundance of tryptophan N-terminal peptide and oxidation products in different conditions. Normalized abundance of the 4 different species (Trp-fragment, TrpOx1-fragment, TrpOx2-fragment, TrpOx3-fragment; as detailed in **Supplementary Figure 8di**) in: condition 1, after trypsin cleavage in solution (left); condition 2, after addition of APS and TEMED with trypsin cleavage in solution (middle); and condition 3, after trypsin cleavage of peptides from ExM gels (condition 3; right). AUC for each species in the extracted ion chromatogram was obtained as in **Methods: Analysis of LC/QToF data** and the raw total ion chromatograms (TIC) located in **Source Data**. The abundance of each species was normalized with max absolute scaling for plotting on the bar graph (n=1 gelation solution for each condition).

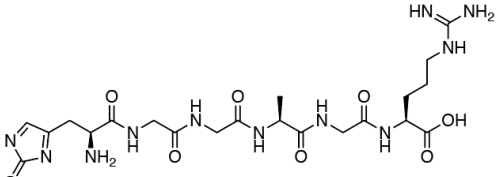
f)

i) A)



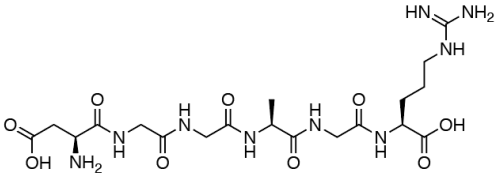
His-fragment
Exact Mass: 553.2721

B)



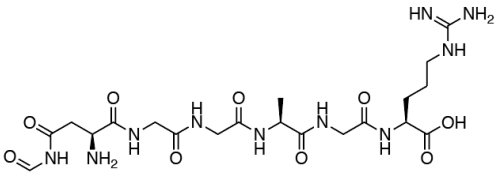
2-oxo-histidine-fragment
Exact Mass: 567.2514

C)



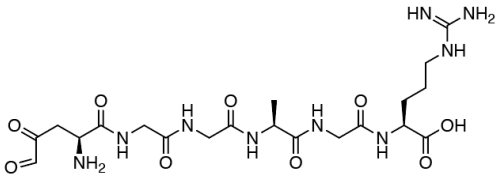
Aspartate-fragment
Exact Mass: 531.2401

D)



Formylasparagine-fragment
Exact Mass: 558.2510

E)

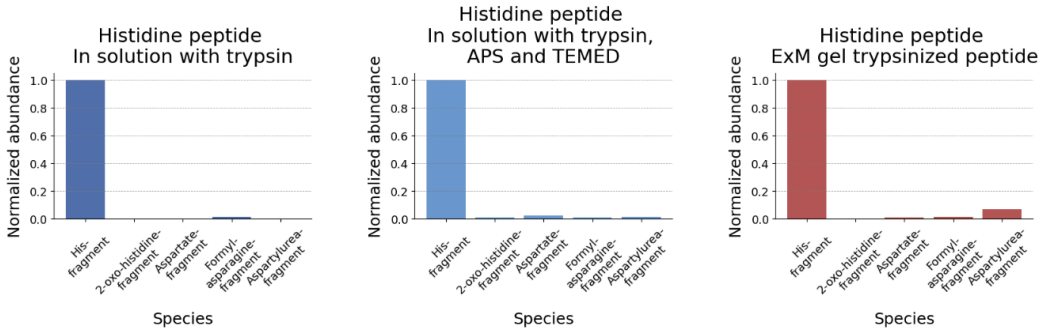


Aspartylurea-fragment
Exact Mass: 543.2401

ii)

Histidine	A) His-fragment	B) 2-oxo-histidine-fragment	C) Aspartate-fragment	D) Formylasparagine-fragment	E) Aspartylurea-fragment
Exact mass	*553.2721*	*567.2514*	*531.2401*	*558.251*	*543.2401*
Condition 1: trypsin	24,178,48.5	442.7	9,000.49	33,615.51	235.33
Condition 2: TEMED+APS+trypsin	1828127.9	22,679.29	44,921.53	19,516.84	28,212.79
Condition 3: TEMED+APS+trypsin in ExM gel	441,471.91	-	4,632.48	5,975.27	30,999.35

iii)



Supplementary Figure 8f Histidine and post-translational oxidation.

i) Structure of the various fragments resulting from enzymatic cleavage with trypsin from the parent peptide (sequence of parent peptide: HGGAGRGLGK{acr}), (A) "Histidine fragment" (HGGAGR, abbreviated: "His-fragment") resulting from trypsin cleavage, (B) modification of the fragment in (A) with known histidine oxidation products, notably 2-oxo-histidine, to form "2-oxo-histidine fragment" (abbreviated: "2-oxo-histidine-fragment"), (C) modification of the fragment in (A) with known histidine oxidation products, notably "Aspartate fragment" (abbreviated: "Aspartate-fragment"), (D) modification of the fragment in (A) with known histidine oxidation products, notably

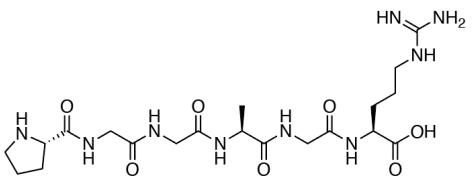
"Formylasparagine fragment" (abbreviated: "Formylasparagine-fragment"), (E) modification of the fragment in (A) with known histidine oxidation products, notably "Aspartylurea fragment" (abbreviated: "Aspartylurea-fragment") [62,63]**.**

ii) Table detailing the exact mass, and absolute abundance recorded for the different products (from A-E peptide fragments from **Supplementary Figure 8fi**) in the different conditions (Condition 1: in solution with trypsin; Condition 2: in solution with trypsin, APS and TEMED; Condition 3: ExM gel trypsinized peptide). AUC for each species in the extracted ion chromatogram was obtained as in **Methods: Analysis of LC/QToF data** and the raw total ion chromatograms (TIC) located in **Source Data**. The dash ('-') means the species was not detected when using search parameters (the extracted ion chromatogram was generated using the $[M+H]^+$ ion of the species, with a mass tolerance of ±20 ppm, with all detected peaks integrated into the final value).

iii) Normalized abundance of histidine N-terminal peptide and oxidation products in different conditions. Normalized abundance of the 5 different species (His-fragment, 2-oxo-histidine-fragment, Aspartate-fragment, Formylasparagine-fragment, Aspartylurea-fragment; as detailed in **Supplementary Figure 8fi**) in: condition 1, after trypsin cleavage in solution (left); condition 2, after addition of APS and TEMED with trypsin cleavage in solution (middle); and condition 3, after trypsin cleavage of peptides from ExM gels (condition 3; right). AUC for each species in the extracted ion chromatogram was obtained as in **Methods: Analysis of LC/QToF data** and the raw total ion chromatograms (TIC) located in **Source Data**. The abundance of each species was normalized with max absolute scaling for plotting on the bar graph (n=1 gelation solution for each condition).

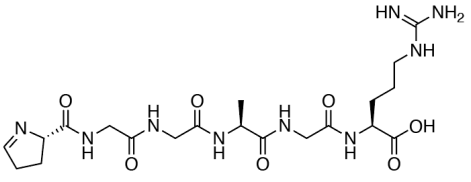
g)

i) A)



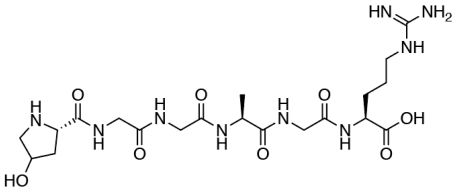
Pro-fragment
Exact Mass: 513.2659

B)



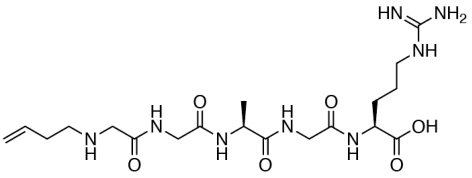
Pyrroline-5-carboxylate-fragment
Exact Mass: 511.2503

C)



Hydroxyproline-fragment
Exact Mass: 529.2609

D)

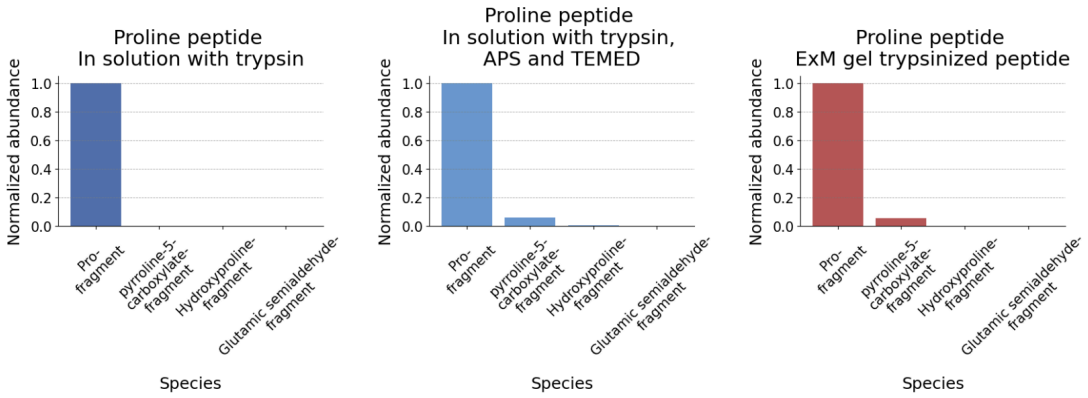


Glutamic semialdehyde-fragment
Exact Mass: 470.2601

ii)

Proline	A) Pro-fragment	B) Pyrroline-5-carboxylate-fragment	C) Hydroxyproline-fragment	D) Glutamic semialdehyde-fragment
Exact mass	*513.2659*	*511.2503*	*529.2609*	*527.2816*
Condition 1: trypsin	37,903,603	371.23	6,882.07	5,223.48
Condition 2: TEMED+APS+trypsin	9,000,359.9	544,009.1	23,130.87	4,143.66
Condition 3: TEMED+APS+trypsin in ExM gel	5,514,152.5	294,169.1	3,102.69	1,426.29

iii)



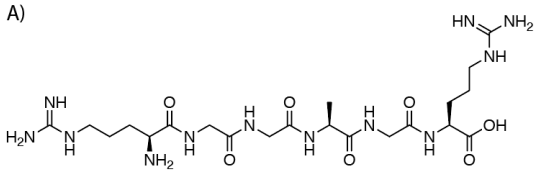
Supplementary Figure 8g Proline and post-translational oxidation.

i) Structure of the various fragments resulting from enzymatic cleavage with trypsin from the parent peptide (sequence of parent peptide: PGGAGRGLGK{acr}), (A) "Proline fragment" (sequence: PGGAGR, abbreviated: "Pro-fragment") resulting from trypsin cleavage, (B) modification of the fragment in (A) with known proline oxidation products, notably "Pyrroline-5-carboxylate", to form "Pyrroline-5-carboxylate fragment" (abbreviated: "Pyrroline-5-carboxylate-fragment"), (C) modification of the fragment in (A) with known proline oxidation products, notably "Hydroxyproline" (abbreviated: "Hydroxyproline-fragment"), (D) modification of the fragment in (A) with known proline oxidation products, notably "Glutamic semialdehyde" (abbreviated: "Glutamic semialdehyde-fragment") [64].

ii) Table detailing the exact mass, and absolute abundance recorded for the different products (from A-D peptide fragments from **Supplementary Figure 8gi**) in the different conditions (Condition 1: in solution with trypsin; Condition 2: in solution with trypsin, APS and TEMED; Condition 3: ExM gel trypsinized peptide). AUC for each species in the extracted ion chromatogram was obtained as in **Methods: Analysis of LC/QToF data** and the raw total ion chromatograms (TIC) located in **Source Data**. The dash ('-') means the species was not detected when using search parameters (the extracted ion chromatogram was generated using the $[M+H]^+$ ion of the species, with a mass tolerance of ±20 ppm, with all detected peaks integrated into the final value).
iii) Normalized abundance of proline N-terminal peptide and oxidation products in different conditions. Normalized abundance of the 4 different species (Pro-fragment, Pyrroline-5-carboxylate-fragment, Hydroxyproline-fragment, Glutamic semialdehyde-fragment; as detailed in **Supplementary Figure 8gi**) in: condition 1, after trypsin cleavage in solution (left); condition 2, after addition of APS and TEMED with trypsin cleavage in solution (middle); and condition 3, after trypsin cleavage of peptides from ExM gels (condition 3; right). AUC for each species in the extracted ion chromatogram was obtained as in **Methods: Analysis of LC/QToF data** and the raw total ion chromatograms (TIC) located in **Source Data**. The abundance of each species was normalized with max absolute scaling for plotting on the bar graph (n=1 gelation solution for each condition).

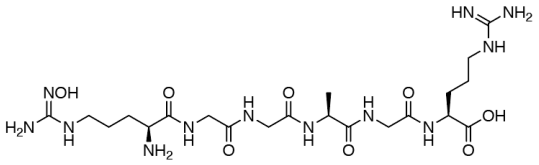
h)

i) A)



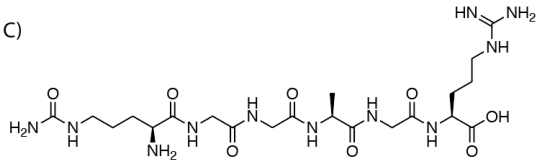
Arg-fragment
Exact Mass: 572.3143

B)



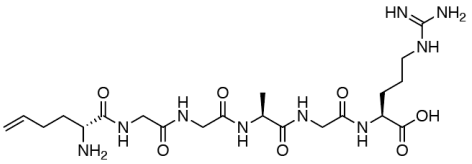
HydroxyArg-fragment
Exact Mass: 588.3092

C)



Oxoarginine
Exact Mass: 573.2983

D)

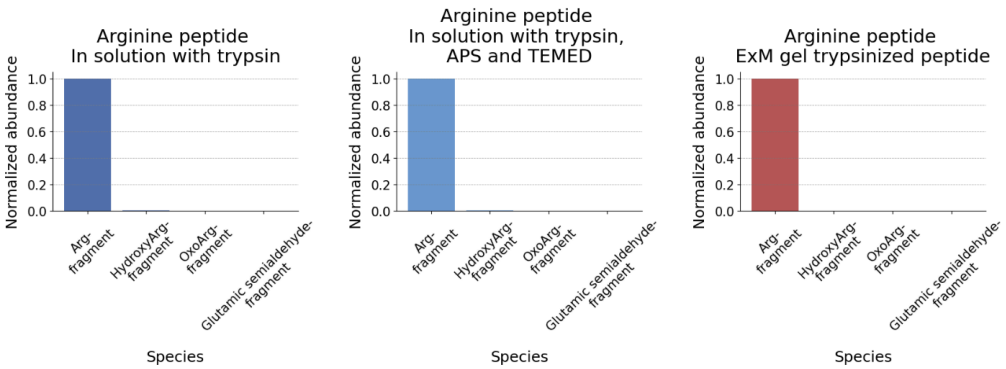


Glutamic semialdehyde-fragment
Exact Mass: 527.2816

ii)

Arginine	A) Arg-fragment	B) HydroxyArg-fragment	C) OxoArg-fragment	D) Glutamic semialdehyde-fragment
Exact mass	*572.3143*	*588.3092*	*573.2983*	*527.2816*
Condition 1: trypsin	17,883,221	94,152.55	569.49	821.92
Condition 2: TEMED+APS+trypsin	7,767,661.7	51,065.82	13,029.76	1,481.44
Condition 3: TEMED+APS+trypsin in ExM gel	4,906,487.6	6,285.23	2,591.88	1,105.52

iii)



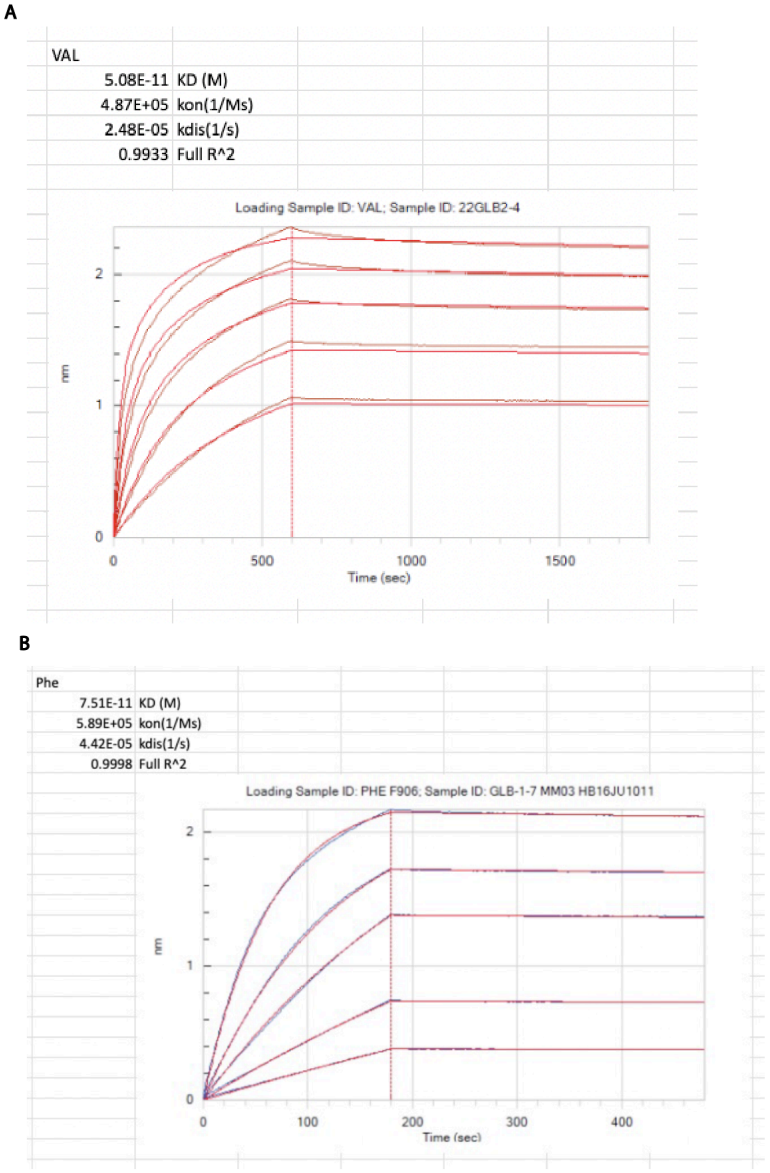
Supplementary Figure 8h Arginine and post-translational oxidation.

i) Structure of the various fragments resulting from enzymatic cleavage with trypsin from the parent peptide (sequence of parent peptide: RGGAGRGLGK{acr}), (A) "Arginine fragment" (RGGAGR, abbreviated: "Arg-fragment") resulting from trypsin cleavage, (B) modification of the fragment in (A) with known arginine oxidation products, notably "Hydroxyarginine", to form "Hydroxyarginine fragment" (abbreviated: "HydroxyArg-fragment"), C) modification of the fragment in (A) with known arginine oxidation products, notably "Oxoarginine" (abbreviated: "OxoArg-fragment"), (D) modification of the fragment in (A) with known arginine oxidation products, notably "Glutamic semialdehyde" (abbreviated: "Glutamic semialdehyde-fragment") [65,66].

ii) Table detailing the exact mass, and absolute abundance recorded for the different products (from A-D peptide fragments from **Supplementary Figure 3hi**) in the different conditions (Condition 1: in solution with trypsin; Condition 2: in solution with trypsin, APS and TEMED; Condition 3: ExM gel trypsinized peptide). AUC for each species in the extracted ion chromatogram was obtained as in **Methods: Analysis of LC/QToF data** and the raw total ion chromatograms (TIC) located in **Source Data**. The dash ('-') means the species was not detected when using search parameters (the extracted

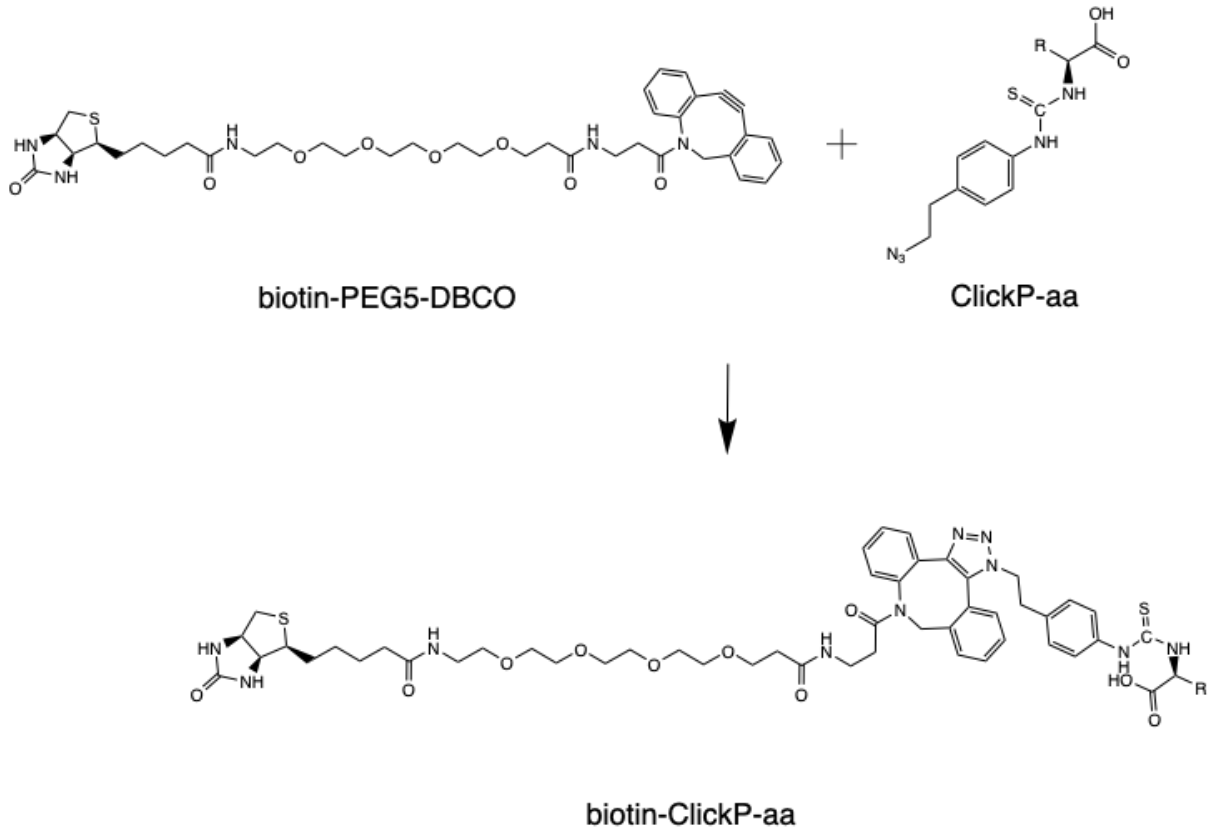
ion chromatogram was generated using the $[M+H]^+$ ion of the species, with a mass tolerance of ±20 ppm, with all detected peaks integrated into the final value).
iii) Normalized abundance of arginine N-terminal peptide and oxidation products in different conditions. Normalized abundance of the 4 different species (Arg-fragment, HydroxyArg-fragment, OxoArg-fragment, Glutamic semialdehyde-fragment; as detailed in **Supplementary Fig. 3hi**) in: condition 1, after trypsin cleavage in solution (left); condition 2, after addition of APS and TEMED with trypsin cleavage in solution (middle); and condition 3, after trypsin cleavage of peptides from ExM gels (condition 3; right). AUC for each species in the extracted ion chromatogram was obtained as in **Methods: Analysis of LC/QToF data** and the raw total ion chromatograms (TIC) located in **Source Data**. The abundance of each species was normalized with max absolute scaling for plotting on the bar graph (n=1 gelation solution for each condition).

Supplementary Figure 9



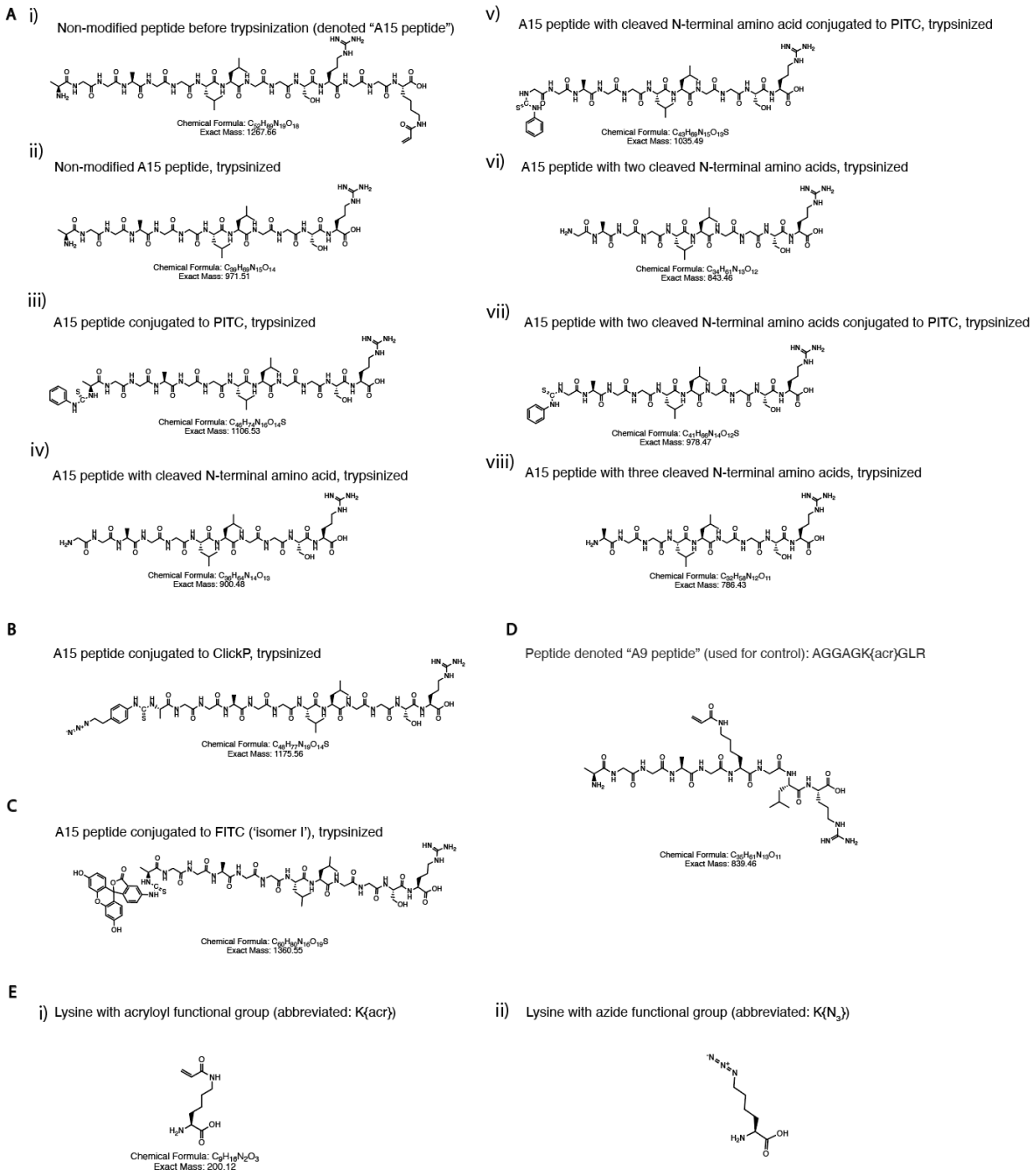
Supplementary Figure 9: Biolayer interferometry (BLI) curves for GFP-conjugated Glyphic antibody against ClickP-V (abbreviated: Glyphic V) and RFP-conjugated Glyphic antibody against ClickP-F (abbreviated: Glyphic F). (A) BLI curve obtained for Glyphic V against ClickP-valine antigen using five different antibody concentrations. The global fit was obtained using a bivalent fitting model and shows a dissociation constant of K_d = 50.8 pM, a k_{on} = 4.87 * 10^5 $M^{-1}s^{-1}$, a k_{dis} = 2.48 * 10^{-5} s^{-1}. (B) As in A, but for Glyphic F. K_d = 75.1 pM, a k_{on} = 5.89 * 10^5 $M^{-1}s^{-1}$, a k_{dis} = 4.42 * 10^{-5} s^{-1}.

Supplementary Figure 10



Supplementary Figure 10: Reaction of biotin-PEG5-DBCO (top left) with decyclized ClickP-amino acid (abbreviated: ClickP-aa) (top right) to form biotin-ClickP-aa (bottom). Biotin-ClickP-aa is the reagent used in **Figure 6d-f** for assessing specificity of binders in the ExMre gels. "R" in this case represents the variable group of the amino acid side chain.

Supplementary Figure 11



Supplementary Figure 11: Structure of various peptides (chemical formula and exact mass also detailed under each structure).

(A) (i) Structure of the "A15" peptide (i.e., AGGAGGLLGGSRGGK{acr}), (ii-viii) are structures of this peptide once it has been trypsinized from the gel (cleavage at arginine, R), (ii) the non-modified fragment trypsinized from the gel AGGAGGLLGGSR, (iii) the same fragment as (ii) but after conjugation with PITC, (i.e., *PTC*-AGGAGGLLGGSR), (iv) the same fragment as (ii) but with cleaved N-terminal amino acid (i.e., GGAGGLLGGSR), (v) the same fragment as (iv) but after conjugation with PITC (i.e., *PTC*-GGAGGLLGGSR), (vi) the same fragment as (ii) but with two cleaved N-terminal amino acids (i.e., GAGGLLGGSR),

(vii) the same fragment as (vi) but after conjugation with PITC (i.e., *PTC*-GAGGLLGGSR), (viii) the same fragment as (ii) but with three cleaved N-terminal amino acids (i.e., AGGLLGGSR).

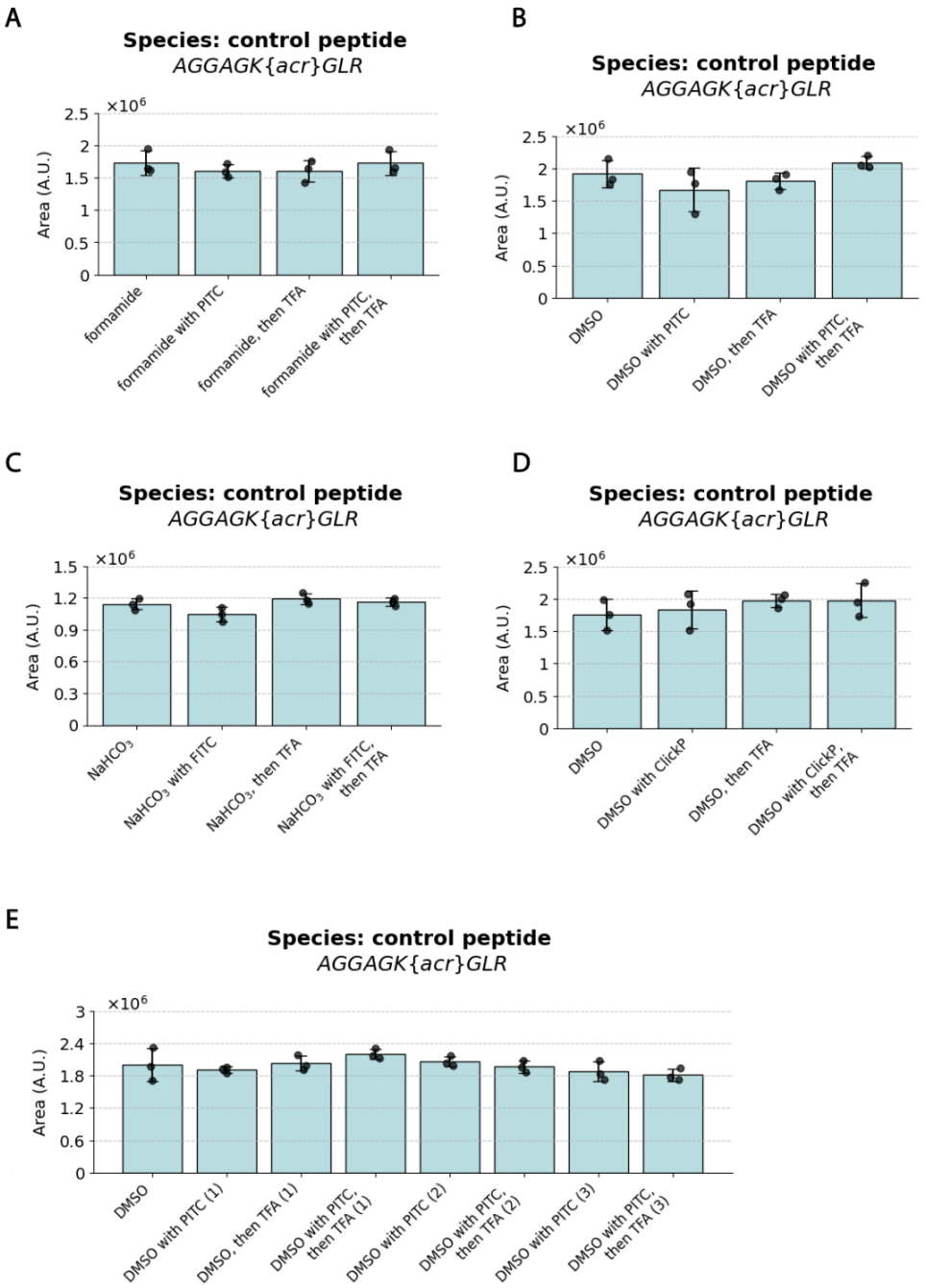
(B) Structure of the same fragment as (ii) but after conjugation with ClickP (i.e., *ClickP*-AGGAGGLLGGSR).

(C) Structure of the same fragment as (ii) but after conjugation with FITC ('Isomer I') (*FITC*-AGGAGGLLGGSR).

(D) Structure of peptide denoted A9 peptide, used as ionization efficiency standard, with sequence AGGAGK{acr}GLR, where the acryloyl group on the side chain of lysine is abbreviated K{acr}.

(E) (i) Structure of the lysine amino acid modified with acryloyl group (K{acr}), (ii) structure of the lysine amino acid modified with azide group (abbreviated: K{N_3}).

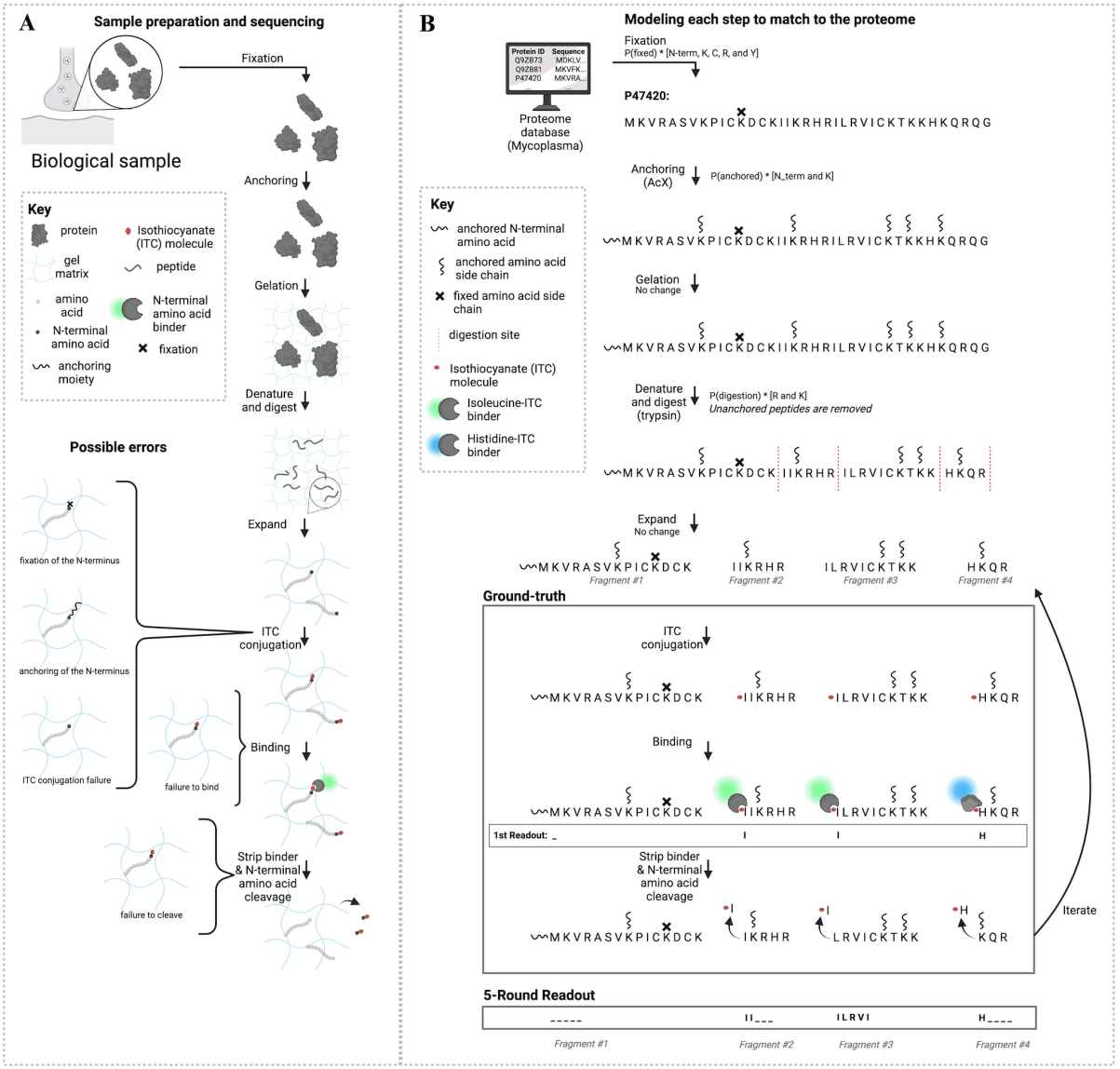
Supplementary Figure 12

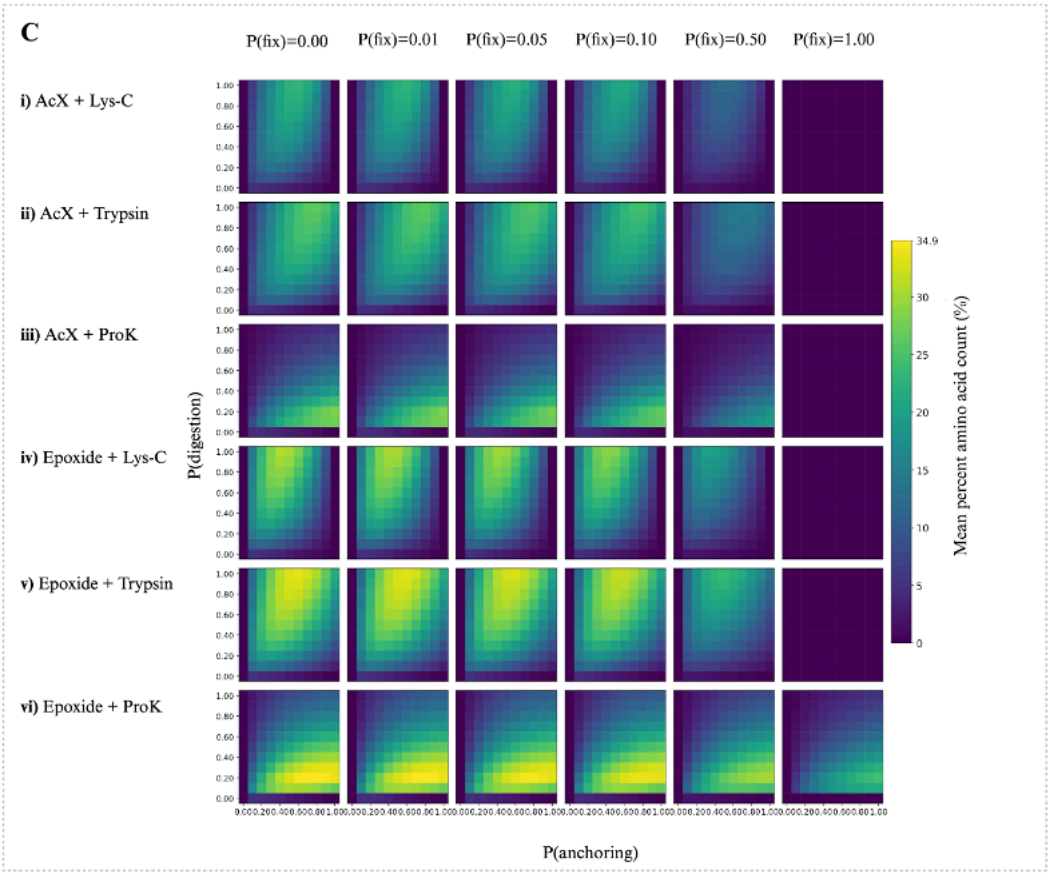


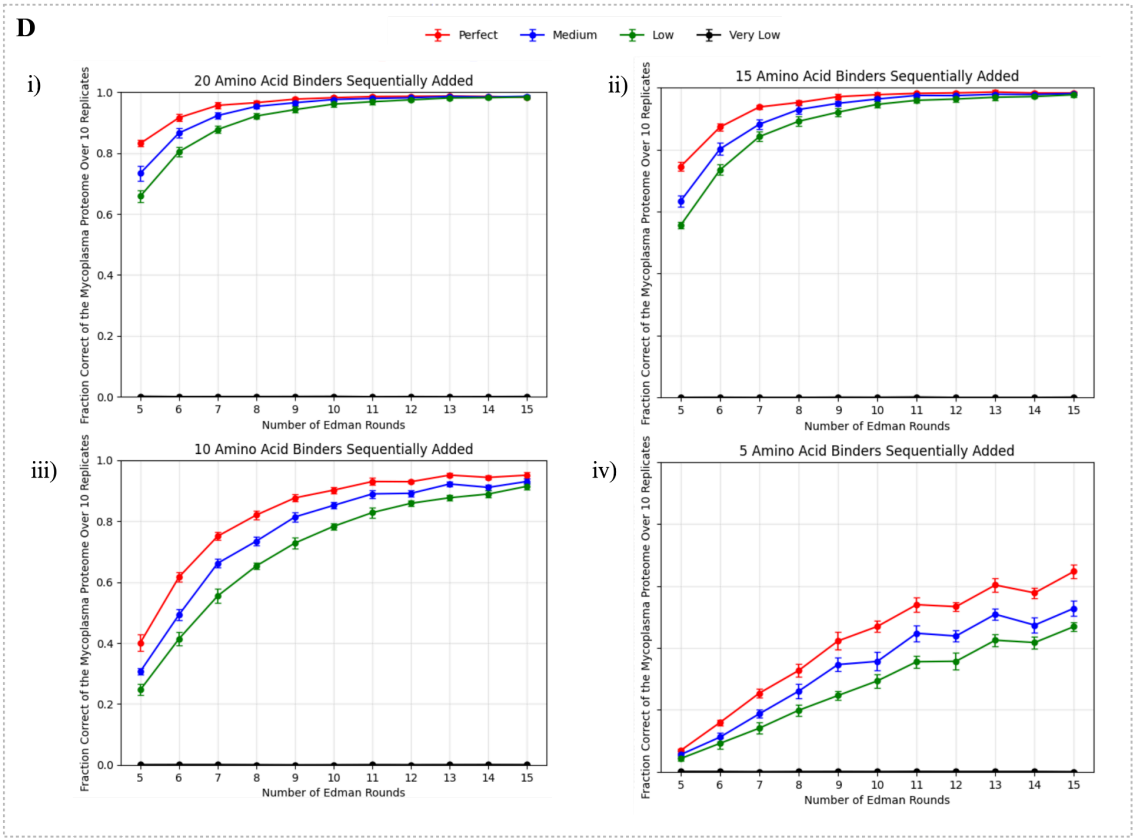
Supplementary Figure 12: Bar graphs representing the relative abundance (arbitrary units, a.u.) of A9-peptide (AGGAGK{acr}GLR) ion species detected on the LC/QToF. The bar graphs are obtained from comparing the area under the curve (AUC) of the chromatogram for the A9 peptide. It was spiked into the supernatant of the trypsinized gels, as an ionization efficiency standard, for downstream LC/QToF analysis in different in-gel Edman degradation conditions. A9 peptide abundance after the gel-anchored peptide (*not* the A9) was treated with: (A) PITC to formamide (1:1000 ratio PITC:formamide) conjugation solvent (as well as solvent only, TFA, and PITC followed by TFA; **Figure 3b**), (B) PITC to DMSO (1:1000 ratio PITC:DMSO) conjugation solvent (and related conditions, as in A; **Figure 3c**), (C) FITC at concentration 5.9 mM with 23:77 DMSO:0.1 M sodium bicarbonate ($NaHCO_3$) pH 8.5 (and related conditions, as in A; **Figure 3d**) (D) ClickP to DMSO

(1:1000 ratio ClickP:DMSO) as conjugation solvent (and related conditions, as in A; **Figure 6c**), (E) PITC to DMSO (1:1000 ratio PITC:DMSO) as conjugation solvent, over 3 rounds of in-gel Edman degradation (and related conditions, as in A; **Figure 5b**). Black dots, individual experiments; blue bar, mean; error bar: standard deviation; n=3 gelation solutions.

Supplementary Figure 13




C
P(fix)=0.00
P(fix)=0.01
P(fix)=0.05
P(fix)=0.10
P(fix)=0.50
P(fix)=1.00
i) AcX + Lys-C
ii) AcX + Trypsin
iii) AcX + ProK
iv) Epoxide + Lys-C
v) Epoxide + Trypsin
vi) Epoxide + ProK
P(digestion)
P(anchoring)
Mean percent amino acid count (%)


D
Perfect
Medium
Low
Very Low
i)
20 Amino Acid Binders Sequentially Added
ii)
15 Amino Acid Binders Sequentially Added
iii)
10 Amino Acid Binders Sequentially Added
iv)
5 Amino Acid Binders Sequentially Added
Fraction Correct of the Mycoplasma Proteome Over 10 Replicates
Number of Edman Rounds

Supplementary Figure 13: A theoretical assessment of in situ protein sequencing.

(A) Schematic of the experimental workflow being modeled. Starting from the biological sample, the chemical steps include fixation, anchoring, gelation, denature/digestion, expanding, and the sequencing chemistry. Major potential sequencing chemistry errors are depicted, including during fixation or anchoring of the N-terminal amino acid, ITC conjugation failure, failure for the NAAB to bind, and/or failure to cleave the N-terminal amino acid.

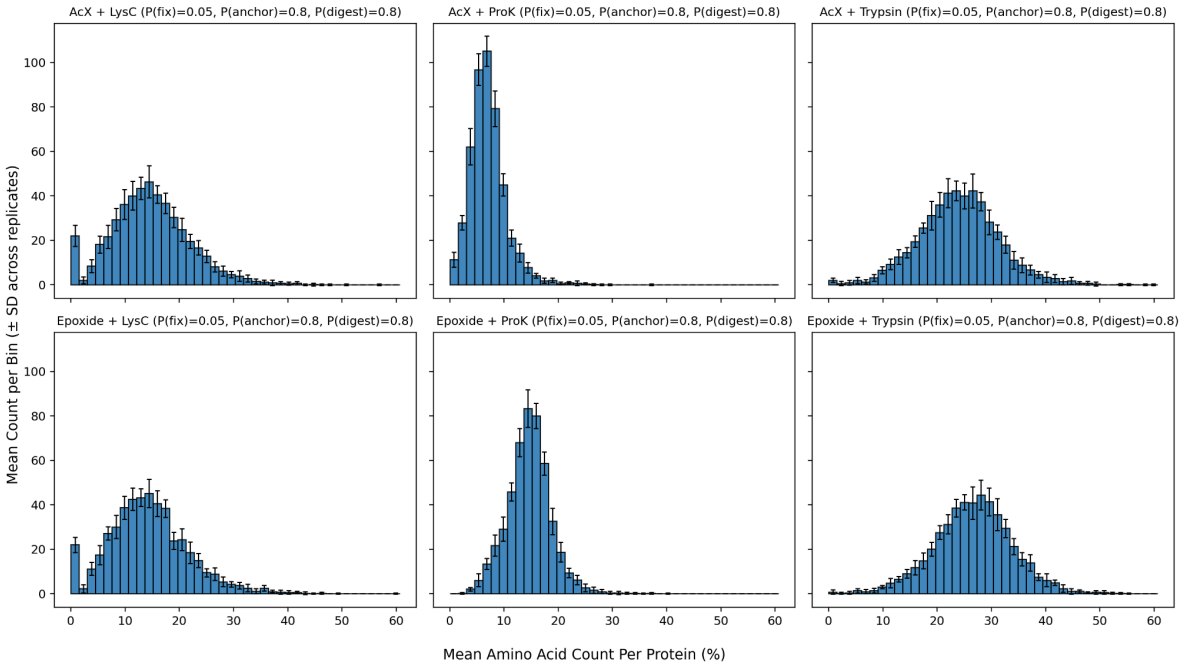
(B) Schematic of proteins in the pipeline of the computational model. Starting from modeling the upstream preparation steps, where fixation step modifies N-terminal (N-term), K, C, R, Y with probability P(fixed); the anchoring step that modifies N-term and K with probability P(anchored); gelation which does not lead to sequence modification; denature/digest step that cleaves peptides at R and K residues to expose new N-termini; and expansion which does not modify the sequences. Sequencing chemistry is performed with sequential addition of ITC to the N-termini of the peptide fragments, followed by NAAB binding and read-outs, and iteration of these two previous steps. The 5-round read-out shows the result of 5 iterations of those steps on 4 different fragments of the protein P47420 from the mycoplasma proteome, considering no errors (fragment #1 is anchored at the N-terminus and no reads are recorded; fragment #2 read-out is II for two isoleucine read-outs, abbreviated I, followed by loss of fragment from the gel after in-gel Edman degradation cleaves the fragment from the gel; fragment #3 read-out is ILRVI, where L is leucine, R is arginine, V is valine, assuming these binders are available; fragment #4 read-out is H for histidine followed by loss of fragment from the gel after in-gel Edman degradation cleaves the fragment from the gel).

(C) Heatmap of mean percent amino acid count in the *Mycoplasma genitalium* (referred to as "mycoplasma" hereafter) proteome across different types of anchoring conditions (AcX, epoxide) and types of digestion conditions (Lys-C, trypsin, Proteinase-K) at different probabilities of fixation (abbreviated: P(fix)), anchoring (abbreviated: P(anchoring)), and digestion (abbreviated: P(digestion)), with 15 in-gel Edman degradation rounds. Mean percent amino acid count is defined as the number of unmodified sequenceable amino acids (i.e. unfixed, unanchored, retained in gel after digestion or Edman degradation) divided by the length of the protein, averaged across all proteins in the proteome. We run this probabilistic process 10 times and average the resulting mean percentage amino acid count across trials for every combination of conditions.

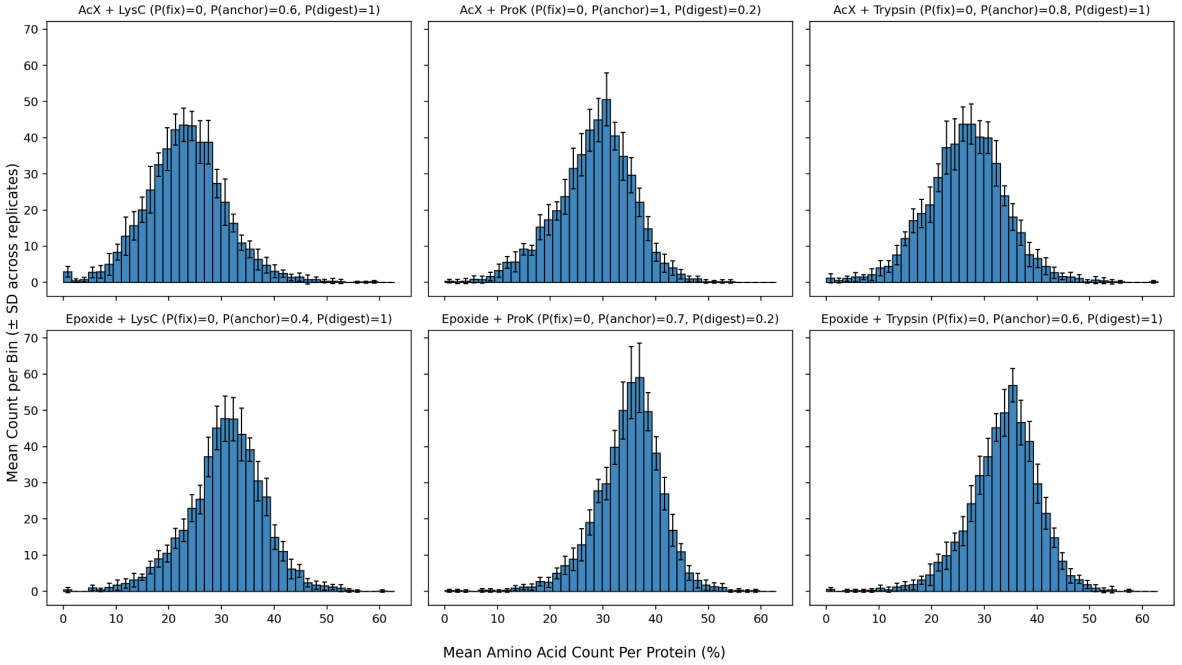
(D) Fraction correct of the identified proteins out the whole mycoplasma proteome for very low (black line on the plots), low (green line on the plots), medium (blue line on the plots) and perfect (red line on the plots) binder specificities (see Supplementary Table 12 for the kinetic values describing these specificity regimes) over increasing numbers of in-gel Edman degradation (from 5 to 15 rounds). Fraction correct is defined as the number of proteins identified correctly (i.e. not false positive and not uncertain; see Methods: Hidden Markov Model (HMM) based matching and fraction of proteome correctly identified for more details) divided by the size of the proteome. (i) assuming all 20 amino acid binders are sequentially added for recognition after each round, (ii) same as (i) but with 15 amino acid binders, against {A, N, D, E, Q, G, I, L, K, F, S, T, P, Y, V} (iii) 10 amino acid binders, against {A, N, D, E, Q, I, L, K, F, V} (iv) 5 amino acid binders, against {N, E, L, F, V} (dot: mean value of the fraction correct for that given condition; error bar: standard deviation for that given condition, n=10 simulations of the fixation, anchoring and digestion with 0.05 fixation, 0.8 AcX anchoring and 0.8 trypsin digestion, 0.1 Edman conjugation failure, 0.3 Edman cleaving failure).

Supplementary Figure 14

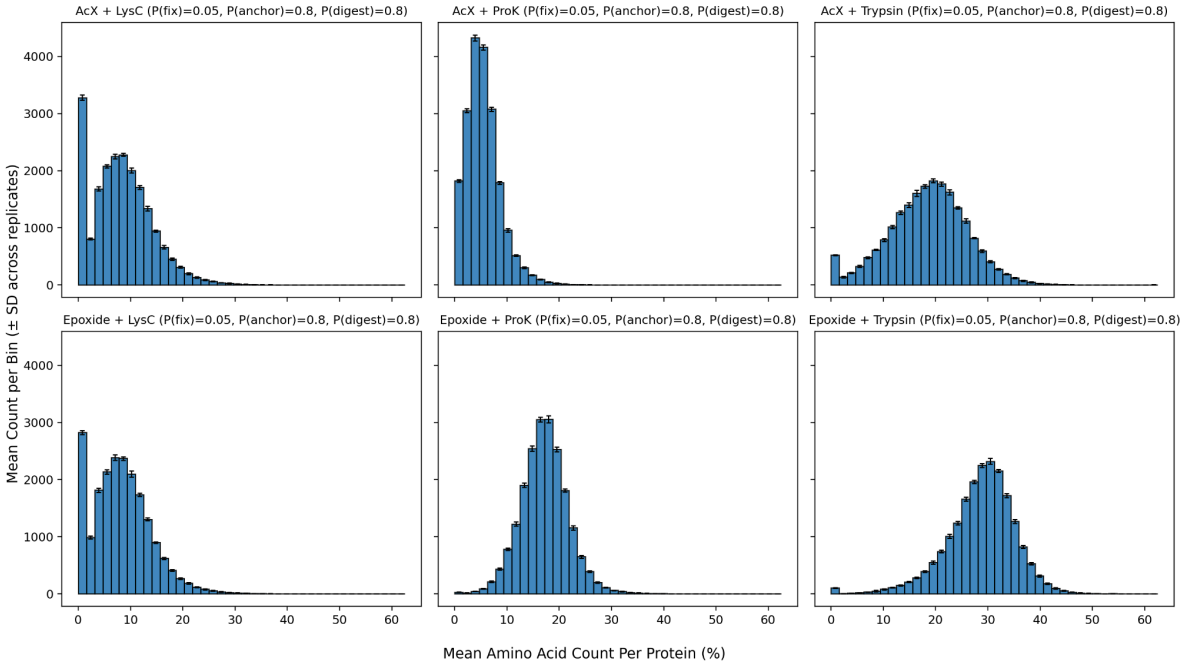
(A)



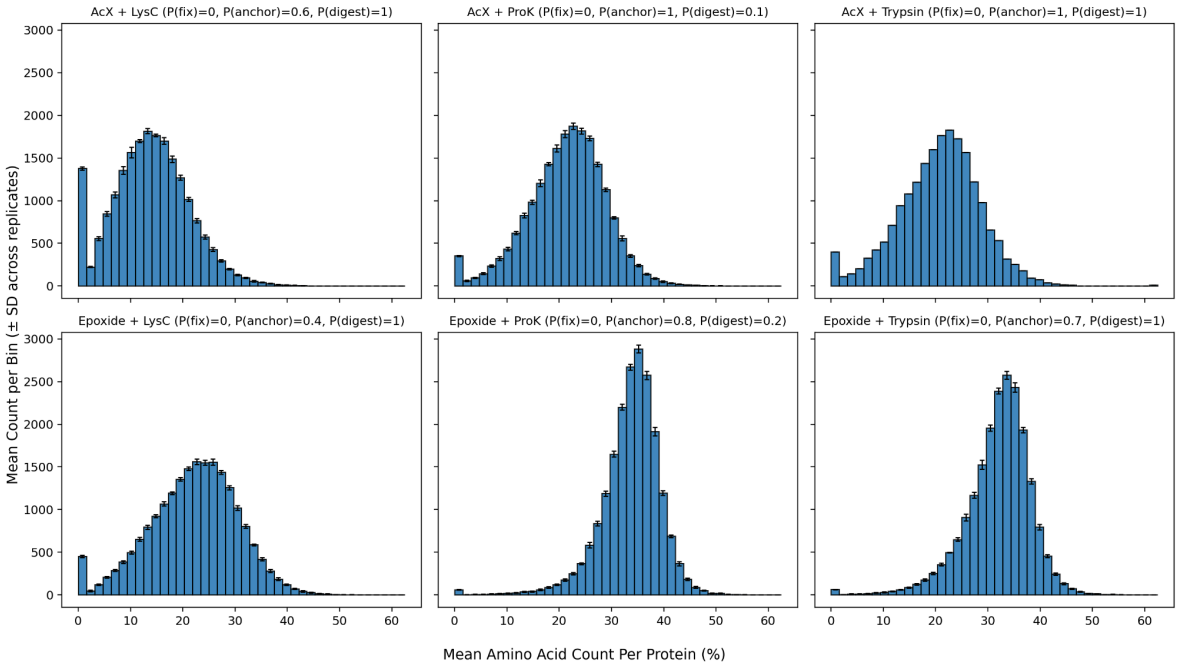
(B)



(C)



(D)

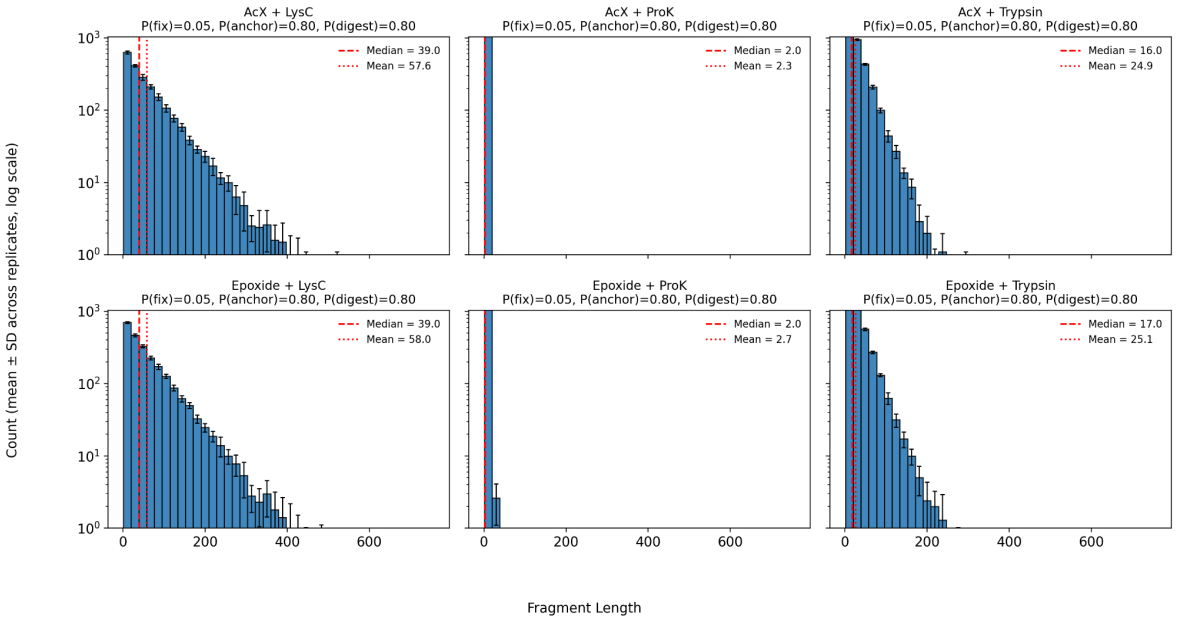


Supplementary Figure 14: Related to the theoretical assessment of in situ protein sequencing Supplementary Figure 13**c** (mycoplasma proteome) and Supplementary Figure 18 (human proteome) for the analysis of the distribution of the mean percent amino acid count for a certain set of parameters. Histogram distribution of the percent amino acid count for six different conditions of the chemistry: 6-((acryloyl)amino)hexanoic acid, succinimidyl ester (abbreviated: AcX) or epoxide anchoring in combination with endopeptidase Lys-C (abbreviated: Lys-C), proteinase K (abbreviated: ProK) or trypsin digestion (specific conditions are listed above each histogram). Mean protein count

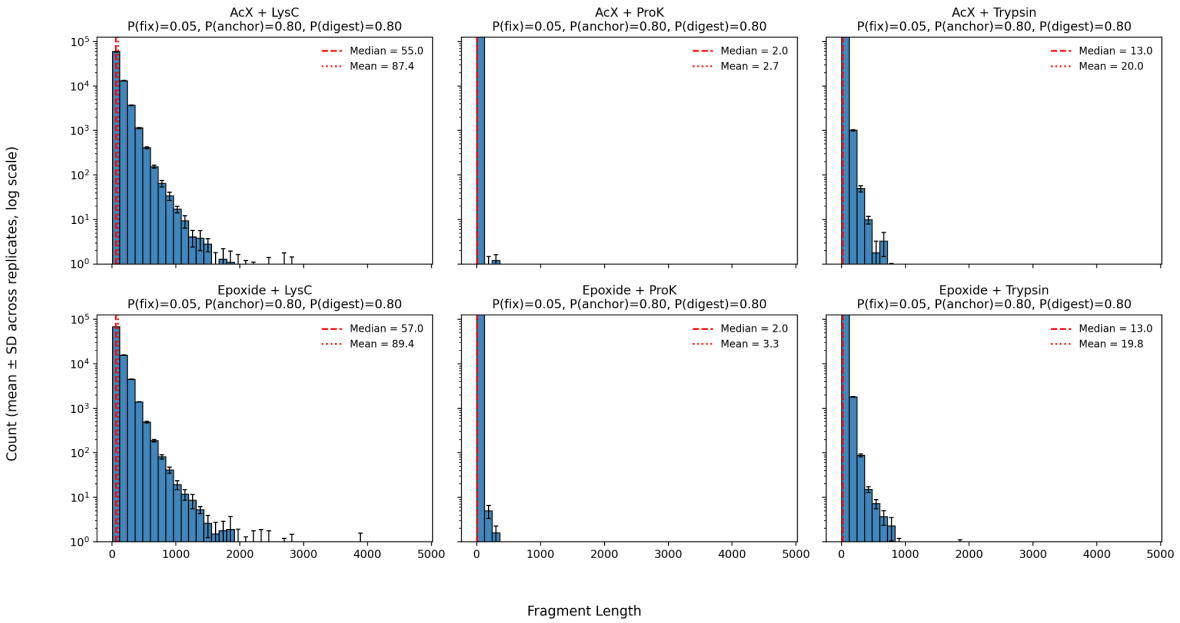
per bin is the frequency of proteins over the proteome (each bin has a range of 2.5% amino acid count). Percent amino acid count is defined as the percentage of residues of a collection of fragments making up a protein that are not modified and remain accessible for read-out over 15 rounds of in-gel Edman degradation after the chemical steps. The simulation is run 10 times for fixation, anchoring and digestion and the resulting fragments are assessed for their percent amino acid count in the gel.
(A) The histogram results showing the distribution of percent amino acid count for all the proteins in the mycoplasma proteome (UniProt ID 243273, total of 483 proteins) for each condition with a fixed parametrization P(fixation)=0.05, P(anchoring)=0.80, P(digestion)=0.80 (abbreviated P(fix), P(anchor), P(digest), respectively). Blue bars: number of proteins with the given percent amino acid count; black error bars: standard deviation across the 10 simulations, throughout this figure.
(B) The histogram results showing the distribution of percent amino acid count for all the proteins in the mycoplasma proteome for each condition with the best parametrization maximizing the averaged percent amino acid count (highest parametrization for the given condition in **Supplementary Figure 13c**), where the best parametrization is labeled at the top of each histogram.
(C) The histogram results showing the distribution of percent amino acid count for all the proteins in the human proteome (UniProt ID 9606, total of 20,421 proteins) for each condition with a fixed parametrization, as in (A).
(D) The histogram results showing the distribution of percent amino acid count for all the proteins in the human proteome for each condition with the best parametrization maximizing the averaged percent amino acid count (highest parametrization for the given condition in **Supplementary Figure 18**), where the best parametrization is labeled at the top of each histogram.

Supplementary Figure 15

(A)



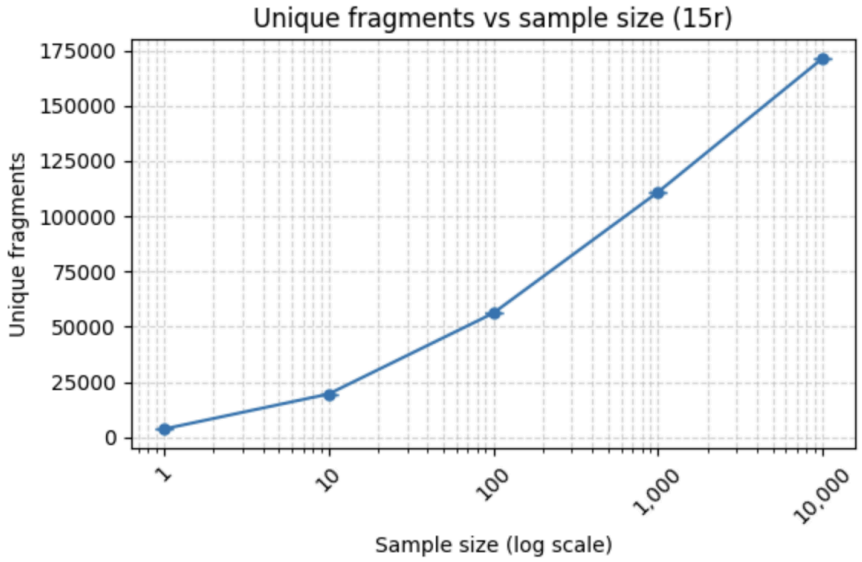
(B)



Supplementary Figure 15: Related to the theoretical assessment of *in situ* protein sequencing Supplementary Figure 13**c (**mycoplasma proteome) and Supplementary Figure 18 (human proteome), analyzing the fragment lengths under various conditions for fixation, anchoring and digestion. Fragment length is defined as the number of amino acids for a given fragment retained in the gel, but not considering amino acids that are located downstream (C-terminal) of the last

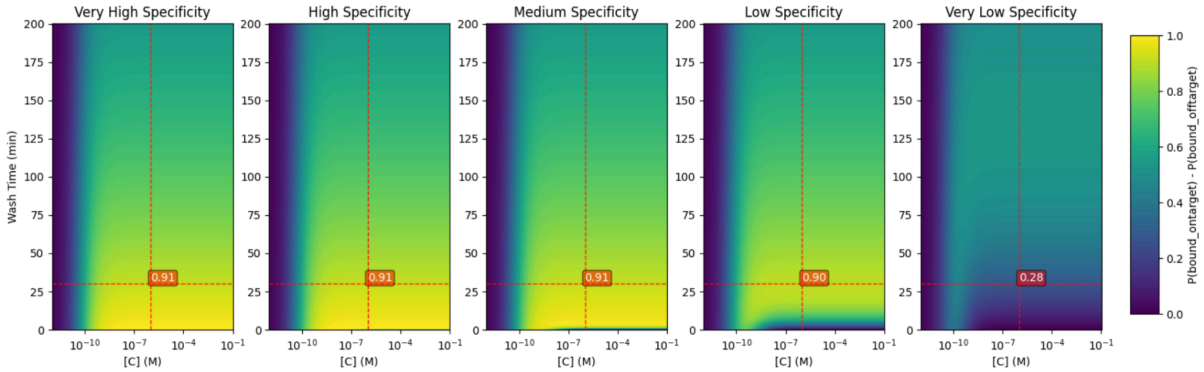
anchoring amino acid of the fragment (since this fragment would be lost after in-gel Edman degradation). Histogram of the distribution of fragment lengths for different parametrization of the chemistry: AcX or epoxide anchoring with Lys-C, ProK or trypsin digestion (specific condition listed above each histogram). The simulation is run 10 times for fixation, anchoring and digestion and the resulting fragments are assessed for their fragment length.
(A) The results for the mycoplasma proteome (UniProt ID 243273, total of 483 proteins). Varying the anchoring conditions (AcX, or epoxide digestion conditions) and the digestion condition (trypsin, Lys-C, or ProK digestion conditions), where P(fixation)=0.05, P(anchoring)=0.80, P(digestion)=0.80 for all conditions (abbreviated: P(fix)=0.05, P(anchor)=0.80, P(digest)=0.80). (blue bars: number of proteins with the given fragment length; error bars: standard deviation for each bin across 10 simulations; vertical red dashed line with long dash: median fragment length; vertical red dashed line with short dash: mean fragment length).
(B) As in (A), but the results for the fixed parametrization (P(fix)=0.05, P(anchor)=0.80, P(digest)=0.80) for each condition for the human proteome (UniProt ID 9606, total of 20,421 proteins).

Supplementary Figure 16



Supplementary Figure 16: Related to the theoretical assessment of *in situ* protein sequencing **Supplementary Figure 13b**, analyzing how increasing the number of simulations impacts the number of unique fragments stored in what we named the "reference fragment dataset" (defined as the database of the ground-truth sequences). Plot showing the number of unique fragments (over the first 15 amino acids from the N-terminus; relevant to 15 rounds of in-gel Edman degradation; abbreviated "15r") versus the number of simulations of fixation, anchoring and digestion chemistries (this is referred to as "Sample size" on the graph x-axis). The results for the mycoplasma proteome (UniProt 243273). For the true fragment space (i.e. no readout errors), we plot the number of unique fragments generated for 15 Edman rounds where the sample size is the number of times we perform the simulation of fixation, anchoring and digestion steps on the mycoplasma proteome.

Supplementary Figure 17

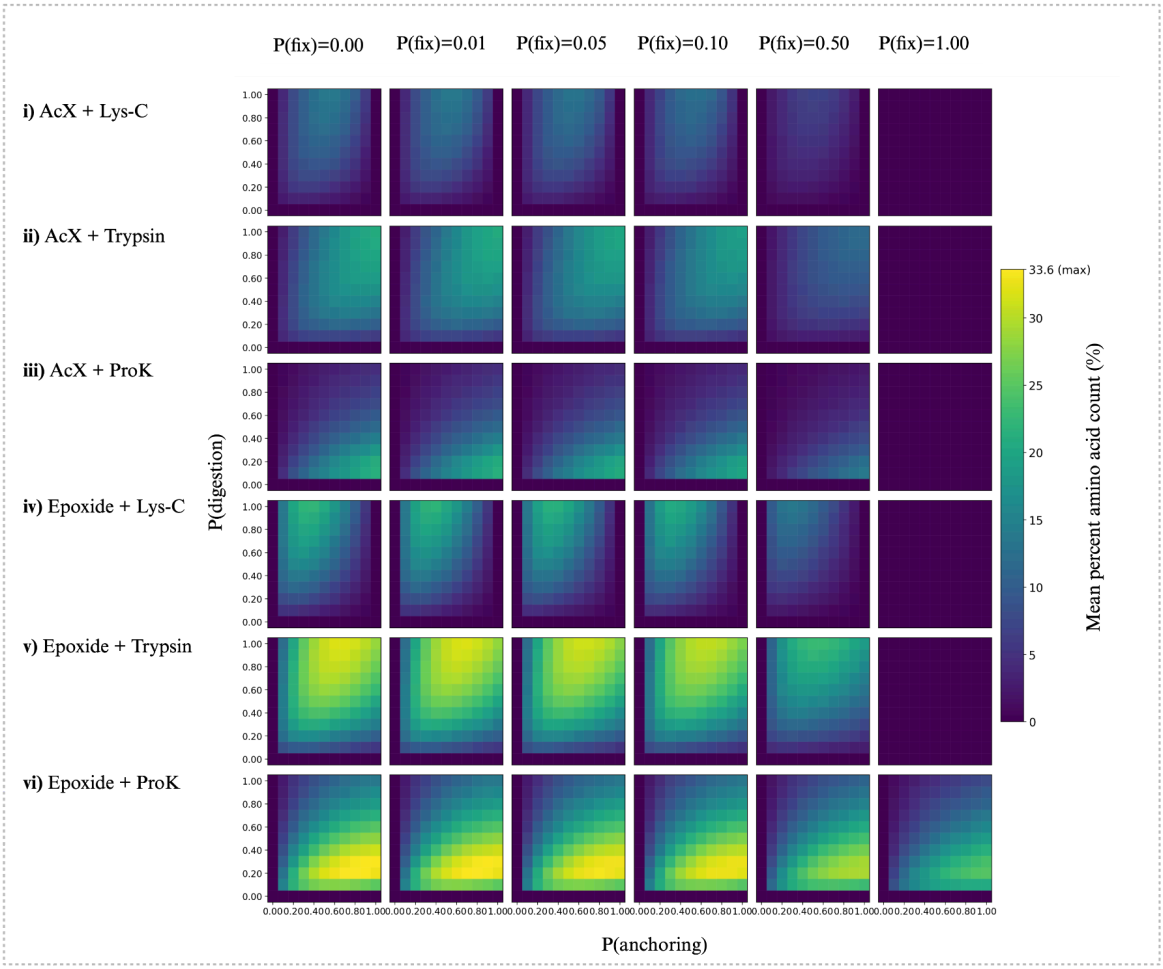


Supplementary Figure 17: Related to the theoretical assessment of *in situ* protein sequencing **Supplementary Figure 13d**, analyzing how wash time (min) and concentration of binder affects read-out. Heatmaps showing the difference between probability that a binder correctly binds its on-target and the probability that it incorrectly binds off-target sites when changing the wash time and the concentration of binder. This difference represents how much more likely an on-target site is to be occupied compared to an off-target site, but it does not reflect the overall binding probability across all possible targets, since it assumes equal site availability between on-target and off-target sites. These results are plotted across varying binder specificity regimes ("very high", "high", "medium", "low" and "very low"; where the kinetic values associated to these binder specificity regimes are detailed in **Supplementary Table 12)**. The results assume that we add excess binder, and that binding reaches equilibrium (using the Langmuir equation for binding) such that after the washing, the probability of being bound to on-target and off-target is defined as:

$p_{bound-ontarget} = \left(\frac{[C]}{[C] + K_d^{on-target}}\right) \cdot e^{-k_{off}^{on-target} t_{wash}}$ and

$p_{bound-offtarget} = \left(\frac{[C]}{[C] + K_d^{off-target}}\right) \cdot e^{-k_{off}^{off-target} t_{wash}}$, where we plot $p_{bound-ontarget} - p_{bound-offtarget}$. This does not consider any effects that the gel environment would have on the kinetics of binding, nor does it consider any non-specific background from binders bound non-specifically to the gel or not washed out. Here, $[C]$ is the concentration of binder, $K_d^{on-target}$ the equilibrium dissociation constant for the on-target, $K_d^{off-target}$ the equilibrium dissociation constant for the off-target, $k_{off}^{on-target}$ the dissociation rate for the on-target, and $k_{off}^{off-target}$ the dissociation rate for the off-target, and t_{wash} is the time elapsed during the dissociation phase (during the washing). For $[C] = 1\ \mu M$ and $t_{wash} = 30\ min$, the probability difference is noted at the intersection of two dashed red lines for each binder specificity regime. These probabilities, for medium, low, and very low specificity, are used in the simulation of the sequencing in **Supplementary Figure 13d**.

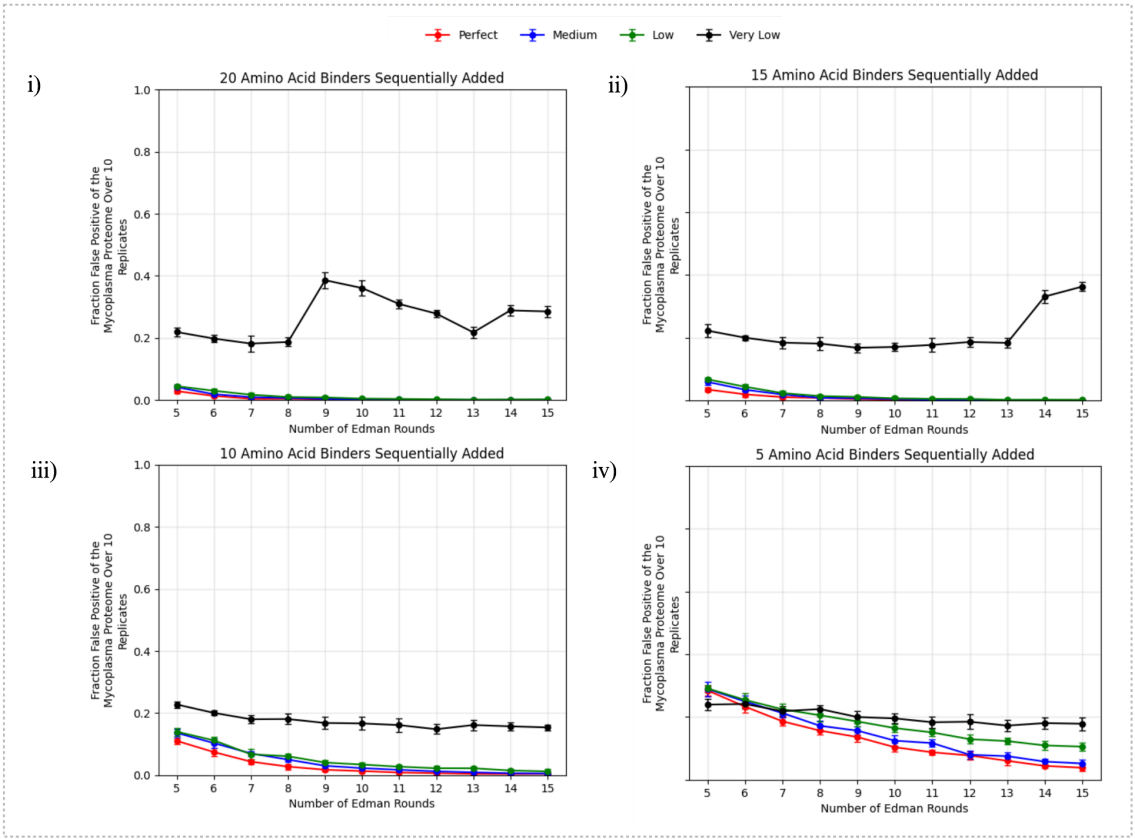
Supplementary Figure 18



Supplementary Figure 18:

(A) Related to the theoretical assessment of in situ protein sequencing **Supplementary Figure 13c**, but rather than plotting the mycoplasma proteome results, plotting the human proteome results. Heatmap of mean percent amino acid count amino acid count (%) in the human proteome (UniProt ID 9606) across different types of anchoring conditions (AcX, epoxide) and types of digestion conditions (Lys-C, trypsin, Proteinase-K) at different probabilities of fixation, anchoring, and digestion, with 15 Edman rounds. Percent amino acid count is defined as the percentage of residues of a collection of fragments making up a protein that are not modified and remain accessible for read-out over 15 rounds of in-gel Edman degradation after the chemical steps. We run this probabilistic process 10 times and average the resulting mean percentage amino acid count across trials for every combination of conditions.

Supplementary Figure 19



Supplementary Figure 19: Fraction false positive of the identified proteins out of the whole mycoplasma proteome for very low (black line on the plots), low (green line on the plots), medium (blue line on the plots) and perfect (red line on the plots) binder specificities (see **Supplementary Table 12** for the kinetic values describing these specificity regimes) over increasing numbers of in-gel Edman degradation (from 5 to 15 rounds). (i) assuming all 20 amino acid binders are sequentially added for recognition after each round, (ii) same as (i) but with 15 amino acid binders, for {A, N, D, E, Q, G, I, L, K, F, S, T, P, Y, V} (iii) 10 amino acid binders, for {A, N, D, E, Q, I, L, K, F, V} (iv) 5 amino acid binders, for {N, E, L, F, V} (dot: mean value of the fraction correct for that given condition; error bar: standard deviation for that given condition, n=10 simulations of the fixation, anchoring and digestion with 0.05 fixation, 0.8 AcX anchoring and 0.8 trypsin digestion, 0.1 Edman conjugation failure, 0.3 Edman cleaving failure).

Supplementary Notes

Supplementary Note 1: Information about previous solvents used in Edman degradation

This section is to add more background to the section **Solvent considerations for in-gel Edman degradation**. Edman degradation can determine the sequence of the first ~30 amino acids from the N-terminus (since over multiple rounds the cumulative yield decreases exponentially, with, for instance, over 30 rounds 0.99^{30} = 74% yield due to accumulated error) by detecting the cleaved phenylthiohydantoin amino acid (PTH-aa)[21,22]. Organic solvents solubilize PITC, a hydrophobic molecule, minimize its hydrolysis with water, and promote efficient isolation of the PTC-peptide product. Typically, pyridine and triethylamine maintain slightly alkaline conditions (e.g., pH 8-9), without adding additional buffers[67]. Acetonitrile (ACN), used in modern protocols[15,32,68], has advantages such as its higher polarity, volatility, reduced UV absorbance for chromatographic analysis, and reduced odor and handling concerns. Maintaining alkaline conditions in solvents such as ACN can be performed by adding triethylamine or pyridine[32,68–72]. After conjugation, PTC-peptides are recovered by evaporation, precipitation, or extraction, and are then cleaved with anhydrous acid (e.g., most commonly trifluoroacetic acid (TFA)). The differing solubilities of PTH-aa (hydrophobic) and peptide (hydrophilic) enable their separation by phase or solid-liquid partitioning. In the latter case, a non-polar solvent can precipitate the peptide, that is then separated from the soluble PTH-aa that is detected using paper chromatography[21]. Improvement in speed and sensitivity of the sequencing method include an automated version of this chemistry[22], peptide degradation on solid support[32], and detection of PTH-aa using HPLC[73], and/or coupled with mass spectrometry[74]. Despite these various modifications, all of these procedures rely on bulk PTH-aa extraction and detection[75].

In terms of the cleavage step, other solvents than TFA can enable cleavage (see **Supplementary Table 2** for a non-exhaustive list of Edman degradation protocol variations). Notably, boron trifluoride etherate in ACN[76,77] has been recently used for DNA-encoded Edman degradation. It also required chemically modified 7-deazapurine deoxynucleotides (c7dA and c7dG) for DNA resistance to depurination[31].

Supplementary Note 2: Information about ionic strength of different buffers and solutions used on ExM and ExMre gels

Different buffers have different ionic strengths, calculated below:

Ionic strength (I) = $\frac{1}{2}\sum_{i=1}^{n}(c_i * z_i^2)$, where c_i is the molar concentration of an ion and z_i is its charge, summed over all ions, n, present in the solution.

- For 1X PBS pH 7.4, three different types of salts: potassium phosphate monobasic (1.06 mM KH_2PO_4), sodium chloride (155.17 mM NaCl), sodium phosphate dibasic (2.97 mM Na_2HPO_4) contribute to ionic strength:

  $I = \frac{1}{2}\sum_{i=1}^{n}(c_i * z_i^2) = \frac{1}{2}\{(c_{K^+} * z_{K^+}^2) + (c_{H_2PO_4^-} * z_{H_2PO_4^-}^2) + (c_{Na^+} * z_{Na^+}^2) + (c_{Cl^-} * z_{Cl^-}^2) + (c_{HPO_4^{2-}} * z_{PO_4^{2-}}^2)\}$

  $I = \frac{1}{2}\sum_{i=1}^{n}(c_i * z_i^2) = \frac{1}{2}\{(1.06 * 1^2) + (1.06 * 1^2) + (161.11 * 1^2) + (155.17 * 1^2) + (2.97 * 2^2)\}$

  $I \simeq 165\, mM.$

- For re-embedding solution, two different types of salts: ammonium persulfate (0.075% APS i.e., 3.29 mM; $(NH_4)_2S_2O_8$) and Tris pH 8 (5 mM Tris-HCl) contribute to ionic strength:

  $I = \frac{1}{2}\sum_{i=1}^{n}(c_i * z_i^2) = \frac{1}{2}\{(c_{NH_4^+} * z_{NH_4^+}^2) + (c_{S_2O_8^{2-}} * z_{S_2O_8^{2-}}^2) + (c_{Tris-H^+} * z_{Tris-H^+}^2) + (c_{Cl^-} * z_{Cl^-}^2)\}$

  $I = \frac{1}{2}\sum_{i=1}^{n}(c_i * z_i^2) = \frac{1}{2}\{(6.57 * 1^2) + (3.29 * 2^2) + (5 * 1^2) + (5 * 1^2)\}$

  $I \simeq 15\, mM.$

The higher ionic strength of 1X PBS, compared to the re-embedding solution, contributes to why ExM gel shrinks more in 1X PBS than in the re-embedding solution.

After re-embedding is performed, the gel has a new internal structure. When placed in 1X PBS ExMre gels expand from hydration, and then can expand more in water.

Supplementary Note 3: Information about LC/QToF solvents and implications for Edman degradation read-out

The effect of formic acid and electrospray ionization (ESI) on read-out of PTC-peptides from in-gel Edman degradation

As mentioned in section **Testbed peptides anchored throughout ExMre gels for validation of in-gel Edman degradation**, peptides with a cleaved N-terminal amino acid can be detected using LC/QToF in the condition where only PITC has been added, without the second neat TFA cleavage step. This can be explained by the presence of 0.1% formic acid in the solutions used for liquid chromatography and ionization (see **Methods** section, under **LC/QToF analysis of synthetic peptides in the gel**, **LC Method**), as well as the use of ESI.

First, the acidic ~pH 2.7 of the aqueous solution could promote cleavage of the PTC-peptide. An aqueous acidic solution (i.e., hydrolytic solution) can also cleave peptide bonds within the peptide, but this requires high temperature, 100-160 °C, for 18-72 hours[78]. As such, peptide bond cleavage, beyond N-terminal amino acid cleavage via Edman degradation chemistry, is unlikely with the methods used here for LC/QToF detection of trypsinized fragments from the gel (0.1% formic acid for ~23 min at 30 °C). This would correspond to on-column cleavage where the peak of the peptide with cleaved N-terminal amino acid matches the retention time of the cleaved peptide in the PITC to DMSO (1:1000 ratio PITC:DMSO) followed by TFA condition (see **Source Data** for raw chromatograms and associated mass spectra).

Furthermore, we discovered that ESI can also promote the conversion of PTC-peptide to peptide with cleaved N-terminal amino acid from in-source fragmentation. Previously, ESI was shown to have an impact on interpretation of data[79]. In addition, previous reports show PTC-peptides were placed in a collision cell for cleavage[80]. Data here also demonstrates this occurs with PTC-peptides with ESI. In this case, unlike the on-column cleavage with formic acid, the peptide with cleaved N-terminal amino acid has a later retention time that matches the PTC-peptide retention time, since the cleavage happens after the elution, most likely inside the ESI source (see **Source Data** for the raw chromatograms and associated mass spectra).

Thus, both on-column cleavage and the in-source cleavage of the PTC-peptide could occur for samples in the mass spec instrument, either from the 0.1% formic acid in the solution for ionization and/or ESI.

Comparing the ionization efficiency of non-modified peptide (AGGAGGLLGGSR) and peptide with cleaved N-terminal amino acid (GGAGGLLGGSR)

The effect of formic acid and electrospray ionization on PTC-peptide precludes assessing the yield of the reaction by comparing the abundance of the non-modified peptide with the peptide with cleaved N-terminal amino acid. However, a control curve was still performed in order to assess the relative ionization efficiency of the non-modified peptide to the peptide with cleaved N-terminal amino acid. The ionization efficiencies were considered similar between the two species, as assessed with the control, which did not provide enough evidence to confidently say there's a difference in ionization efficiency between the two species (see **Supplementary Figure 5** for the control curve; in addition, since alanine is small and not ionizable, its loss after Edman degradation would suggest a similar peptide ionization efficiency to the parent peptide).

Supplementary Note 4: Conventional Edman degradation conditions in the gel

This Supplementary Note discusses conventional Edman degradation conditions that were tested using trypsinization-with-LC/QToF assay in the gel. We investigated two PITC-delivery solvents that we had shown to cause extreme gel shrinkage and opacity - pyridine and ACN - and found that PITC conjugation to gel-anchored peptide did not occur, even when PITC was administered at 1:9 in ACN (see **Supplementary Figure 3a** and **Supplementary Figure 3bi** for results of peptide species detected on LC/QToF using these conditions; see **Supplementary Table 8** for full raw data and descriptive statistics related to **Supplementary Figure 3**). 1:1 pyridine:water, which caused comparable shrinkage but did not result in the visible opacity caused by pyridine and ACN, did result in PITC conjugation after adding 1:9 PITC:solvent, although with still some (~20%) peptide unreacted (top graph in **Supplementary Figure 3bii**, n=1). In addition, to assess the possibility of performing PITC conjugation to the N-terminus of peptides in the gel using a purely aqueous solution, which is known to be compatible with ExMre gels, we tested 0.1 M sodium bicarbonate ($NaHCO_3$) pH 8.5 with PITC (1:9 PITC:buffer) for PTC-peptide formation. The results showed ~70% unreacted peptide (**Supplementary Figure 3biii**, n=1). This suggested a partial Edman conjugation reaction in the ExMre gels in these last two cases. Although 0.1 M $NaHCO_3$ pH 8.5 did not shrink the gel, and did not lead to gel opacity (see **Supplementary Figure 1hi** for representative gel images), the low yield of PITC conjugation in 0.1 M sodium bicarbonate pH 8.5 might be explained by the insoluble nature of PITC in aqueous buffer. In summary, Edman conjugation in the ExMre context correlates strongly with lack of shrinkage or opacity. In addition, it may also be affected by PITC solubility in the solvent at hand.

Supplementary Note 5: Phenylthiohydantoin amino acid (PTH-aa) detection

Calculation of the total dry weight of PTH-F recovered from the ExMre gels and injected into the LC/QToF:

Gel volume: 5*5*0.35 mm is ~9 µL

Several N-terminal amino acid peptides were tested for PTH-aa recovery after in-gel Edman degradation using 9% acrylamide gels (due to ease and speed of making and testing in these gels) and ExMre gels. We thought it useful to comment on these results, we note that our data should be regarded as preliminary (n=1 gelation solution for each condition).

PTH-tyrosine (PTH-Y), and PTH-glycine (PTH-G) were detected after PITC (1:9 ratio PITC:DMSO) conjugation reaction and TFA cleavage in 9% acrylamide gel (5 x 5 x 0.31 mm) containing N-terminal tyrosine or N-terminal glycine (**Xaa**GGAGRGLGK{acr}, where Xaa can be tyrosine or glycine), respectively. PTH-alanine was also detected after PITC (1:9 ratio PITC:DMSO) conjugation reaction and TFA cleavage in ExMre gel (5 x 5 x 0.7 mm) containing N-terminal alanine with the same sequence as above (i.e., AGGAGRGLGK{acr}). This included removal of TFA from the vial and resuspension in 1:1 acetonitrile to water, as in **Figure 3h**. These specific species, with exact mass (calculated with a mass accuracy within a 20 ppm window, see **Edman degradation and PTH detection** section in **Methods** for the relevant equation for calculating the mass accuracy window), matched the expected species. They sometimes exhibited different retention times to the PTH-aa positive control, which may be a result of the presence of an anilinothiozolinone (ATZ) intermediate in the experimental conditions. Indeed, PTH-aa and ATZ-aa have different chemical structures and properties, where treating with aqueous acid and heat can convert the ATZ isomer to PTH isomer. However, using this same protocol, PTH-tryptophan (PTH-W) was not detected, and may be due to differences in solubility in 1:1 acetonitrile to water, and/or differences in extraction of the PTH-aa from the gel after TFA due to distinct biochemical properties.

Supplementary Note 6: ClpS2-StV1 thoughts on affinity and sequence information

This note serves to add information about ClpS2-StV1 protein used in **Figure 4**.

1. Thoughts on affinity

The ClpS2 V1 binder has been previously used for non-in situ peptide sequencing[14] (sequence information for ClpS2 variants in **Supplementary Note 6.2**, below). ClpS2 V1 has a measured k_{off} of ~0.1 s^{-1} (or ~10 second dwell time) towards phenylalanine[25]; the variant ClpS2 St-V1 has improved thermostability while retaining the previously engineered higher affinity towards Phe[26]. However, with a dissociation rate on the order of seconds, this NAAB cannot be used for reliable in situ localization of single molecules when using downstream multi-step signal amplification strategies (requiring hours) employed in ExM protocols, which rely on enzymes or self-assembly cascades[19,34]. Here, to focus on the fundamental compatibility of protein sequencing technology components with the in situ milieu, we do not attempt single molecule imaging, and instead use bulk gel fluorescence to gauge whether NAAB binding can work on peptides in an ExM gel environment. As NAAB quality improves, an exciting next step will be to follow binding with signal amplification and then to do single molecule imaging - as has been done for other molecule types, e.g. RNA, in past expansion microscopy papers[19,34]. Our current experiments also allow an exploration of binder diffusion in the gel, important for any functioning in situ protein sequencing protocol.

2. Sequence information

Agrobacterium tumefaciens ClpS2 was first engineered for higher specificity towards Phe (ClpS2 V1), and then for higher thermostability, named ClpS2 St-V1[25,26]. The ClpS2 St-V1 protein used in **Figure 4** has a C-terminal hemagglutinin (HA) tag. The sequence is as follows:

SSDSPVDLKPKPKVKPKLERPKLYKVMLLNDDYTPMSFVTEVLKAVFNMSEDQGRRVMMTA
HRFGSAVVGVSTRDIAETKAKQATDLAREAGFPLMFTTEPEE-*GGSGGSYPYDVPDYA*

where red highlighting indicates the proposed substrate contacts with phenylalaninamide[25,81], underlined and bolded residues represent the mutations of *A. tumefaciens* ClpS2 wild-type to result in ClpS2 V1 (R35M and E36S), and yellow highlighting indicates amino acids that are different between ClpS2 V1 and ClpS2 St-V1[26]. The ClpS2 V1 sequence is as follows [14]:

MSDSPVDLKPKPKVKPKLERPKLYKVMLLNDDYTPMSFVTVVLKAVFRMSEDTGRRVMMT
AHRFGSAVVVVCERDIAETKAKEATDLGKEAGFPLMFTTEPEE

The first amino acid "S" in the ClpS2 St-V1 sequence remains from cleaving off the N-terminal TEV cleavage site: (ENLYFQ/S) with Tobacco Etch Virus (TEV) protease.

Supplementary Note 7

In the future, higher affinity NAABs will enable more precise measurements, and perhaps more quantitative characterizations. For our low affinity NAABs, we first conducted our exploration in a series of steps. We first assumed that if the binder is added in excess of peptide in the gel, and incubated long enough, it will reach equilibrium, with a binder-bound fraction (and by consequence, concentration) highest in a gel containing the peptide bearing the on-target N-terminal amino acid (see **Supplementary Note 8.2** for equations describing the kinetics of the binder in this condition against different targets). Then, washes to remove binder still in solution, on the order of a couple of minutes, would not have time to wash out all of the binder from the gel containing the on-target peptide, because of the nonzero time required for binder to diffuse out of the gel, even with a short dwell time, with an estimate of > ~70% binder remaining in the gel after a 1-3 min wash (see **Supplementary Note 8.3** for full calculations; this is framed as a lower bound because while we do take into account the slowing of diffusion from the gel itself[82], we do not take into account on-target binding, which would presumably keep more binder in the gel). Then, the gel, containing some fraction of binder within, which would be higher in the on-target gel, is immersed in a solution containing fluorescent antibody against HA-tag, keeping the overall volume small (i.e., a 1.4 µL gel is placed in 30 µL antibody solution), so that even as binder flows out from the gel into the solution, there is still a nonzero fraction of the binder in the volume of the gel labeled by the fluorescent antibody (see **Supplementary Note 8.4**). Finally, the external solution is removed, and the gel imaged. The net result is that some fraction of the NAAB will be retained within the gel, dependent on the initial high fraction of binder bound in the first step, despite the external solution exchanges, and that will result in a greater fluorescence in the gel, accordingly. The prediction is a final expected ~7 fold higher fluorescence intensity in gels with peptide bearing on-target Phe N-terminal amino acid vs. secondary substrates Tyr or Trp (presumably Ala would show no binding, which is why for the purposes of this calculation we compared to Tyr and Trp; see **Supplementary Note 8.4** for full calculations). As noted, the assumptions above are perhaps overestimates of loss of binders from the gel, and thus the fluorescences observed should be regarded as a lower bound on the actual amount; that said, the data should be regarded generally qualitatively, although of course comparisons of different conditions that use the same binder, should be possible.

Having established the assumptions, calculations, and bounds governing our experiment, we next sought to address practicalities - the actual properties of diffusion of binder and antibody in the gel, in relation to the peptide concentration in the gel, and other key experimental parameters like gel thickness. To this end, we tested a concentration of homogeneously distributed peptide of 5 mM, a concentration much higher than would be in an ultimate 1000x expanded cell (indeed, it approximates the density in a living cell, to order of magnitude[83]), in 9% acrylamide gels, with a thickness of ~310 µm and volume of ~1.2 µL, and a binder concentration of 20 µM (well above the K_d of 1.1 µM for Phe). The gel was incubated with 30 µL of the ClpS2 St-V1 binder for an hour at room temperature, followed by a short wash of ~1 min, then was incubated in 1.3 µM antibody overnight at 4 °C in 100 µL. Since wild-type ClpS2 binds N-terminal Phe with highest affinity, binds Trp and Tyr with reduced affinity, and does not bind N-terminal Ala[81], we chose to compare binding of this one binder to all four N-terminal amino acids in acrylamide gels. We cast gels with N-terminal Phe, Trp, Tyr and Ala peptides anchored to gels on their C-termini. The result of this experiment showed differential fluorescent intensity across N-termini peptides, confirming that this experimental design can observe population read-out differences in binding. However, the differential signal was only visible on the edges of the gel, and not in the middle (**Supplementary Figure 7a**): the edge-to-center ratio for Phe was ~12, but ~1 for Trp, Tyr and Ala. We hypothesized that, as in earlier studies of binder diffusion into dense samples[84]. there was a lack of diffusion of ClpS2 St-V1 or HA-tag antibody into ExMre gels, because the dense amount of Phe-bearing peptide target on the edge of the gel soaked up all available tags (being in excess), before they could get into the center (which of course was not a problem if there was no binding, e.g. for Trp, Tyr, and Ala). We reasoned that this depletion could be ameliorated by reducing the concentration of peptides in the gels (similar to expanding a biological sample). Further, longer incubations of binder and antibody might also support protein diffusion into

the core, as long as depletion was not an issue. Nevertheless, this experiment was still valuable since it was consistent with the NAAB binding to its target in an amino acid-specific way.

We next sought to explore more practical concentrations of peptide, e.g. such as might occur after expansion, and less than the binder concentration, to avoid binder depletion. The anti-HA tag antibody was the most likely to encounter a diffusion constraint, based on the Stokes-Einstein relation: larger molecules have larger diffusion coefficients, and slower diffusion flux (estimated ~2.5 fold higher diffusion coefficient for the binder compared to the antibody; see **Supplementary Note 8.1** for calculations). We tested HA-tag antibody diffusion on a set of acrylamide gels with varying HA tag concentrations, with fixed incubation time (O/N at 4 °C), and with an antibody concentration of 0.13 µM. Gels showed fluorescent intensity higher on the edges, than in the center, when the peptide concentration was >=100 µM, but at 10 µM concentration the signal was detected throughout (**Supplementary Figure 7b**). This suggests that the depletion of antibodies occurs on the edges of the gels due to an abundance of binding sites (peptides) when above ~10 µM. Based on this result, and considering the estimated difference in diffusion coefficient of ClpS2-StV1, a ~4 hour incubation at room temperature might be expected to enable similar diffusion for the binder throughout the gel (since $t_{diffusion} \propto \frac{1}{D}$, see **Supplementary Note 8.1** for calculations). We next determined a finalized experimental protocol for ExMre gels. The final thickness of the ExMre gels was ~350 µm, and the final concentration of peptide ~50 µM. As a result, we increased the ClpS2-StV1 incubation time to 6 hours at room temperature, and the antibody incubation to 7 days at 4 °C to let proteins diffuse throughout the gel. In addition, to save on protein binder material, the binder concentration was cut to 10 µM (still above the on-target K_d value), which we estimated to enable a high fraction bound after equilibrium (calculated ~90% for on-target, see **Supplementary Note 8.2**). This protocol was used to test ClpS2 St-V1 and HA-tag antibody binding in the gels, comparing different N-terminal amino acids as before (**Supplementary Figure 7a**), but with the above modifications to observe signal throughout the gel.

Supplementary Note 8

Adapting from "The mathematics of diffusion"[85], we calculate an estimate as follows:

1. Protein radii (assuming spherical shape) and diffusion coefficients (D) of ClpS2 St-V1 and anti-HA antibody:

Assuming a protein density of ~1.35 g/cm[386] and molecular weight of 13,034 g/mol for ClpS2-StV1 and ~150,000 g/mol for anti-HA antibody:

$$Volume_{ClpS2\ St-V1} = \frac{mass}{density} \simeq \frac{13{,}034\ g/mol \div 6.022*10^{23}\ molecules/mol}{1.35\ g/cm^3} \simeq 16.03\ nm^3$$

$$Radius_{ClpS2\ St-V1} \simeq 1.56\ nm \text{ (using the exact result from the volume calculation)}$$

$$Volume_{anti-HA\ antibody} = \frac{mass}{density} \simeq \frac{150{,}000\ g/mol \div 6.022*10^{23}\ molecules/mol}{1.35\ g/cm^3} \simeq 184.51\ nm^3$$

$$Radius_{anti-HA\ antibody} \simeq 3.53\ nm \text{ (using the exact result from the volume calculation)}$$

Abbreviations: diffusion coefficient (D in $m^2\ s^{-1}$), Boltzmann constant (k_B: $1.380649*10^{-23} m^2\ kg\ s^{-2}\ K^{-1}$), temperature ($T$: 298 K for 25 °C and 277 K for 4 °C), dynamic viscosity (η: $0.89*10^{-3}\ kg\ m^{-1}\ s^{-1}$ at 25 °C and $1.57*10^{-3}\ kg\ m^{-1}\ s^{-1}$ at 4 °C) and Stokes radius (r in m). We assume free diffusion in solution, in water:

$$D_{ClpS2\ St-V1} = \frac{k_B T}{6\pi\eta r} \simeq 1.57\ *\ 10^{-6}\ cm^2\ s^{-1}$$ (calculated at room temperature, 25 °C, assuming r = 1.56 * 10^{-9} m)

$$D_{anti-HA\ antibody} = \frac{k_B T}{6\pi\eta r} \simeq 3.66\ *\ 10^{-7}\ cm^2\ s^{-1}$$ (calculated at 4 °C, assuming r = 3.53 * 10^{-9} m)

$$Ratio_D = \frac{D_{ClpS2\ St-V1}}{D_{anti-HA\ antibody}} \simeq 4.3$$

Where $Ratio_D$ represents the ratio between the ClpS2 St-V1 and anti-HA antibody diffusion coefficients, which is directly proportional to their difference in rate of diffusion (for a fixed distance: $t \propto \frac{1}{D}$).

Thus, if O/N at 4 °C is needed for the antibody diffusion (~16 hours), then ~4 hours is needed for ClpS2 St-V1 binder diffusion at room temperature.

2. Kinetics of ClpS2 St-V1; fraction bound at equilibrium:

At equilibrium (i.e., when $t \rightarrow \infty$), the fraction of peptides bound ($f_{b,\ equilibrium}$) depends on the concentration of ClpS2 St-V1 (i.e., $[ClpS2\ St\ -\ V1\ binder]$) and its dissociation constant (i.e., K_d) towards the target:

$$f_{b,\ equilibrium} = \frac{[ClpS2\ St-V1\ binder]}{[ClpS2\ St-V1\ binder] + K_d}$$

This assumes that $[ClpS2\ St\ -\ V1\ binder]_{free} \simeq [ClpS2\ St\ -\ V1\ binder]_{total}$ since the binder is added in excess compared to the peptide (e.g., in **Fig. 4**, ~4.3x molar excess binder compared to peptide in the gel).

- For Phe – Assume K_d= 1.1μM (data for ClpS2 St-V1 against FRVECK-biotin) from [26]. Assume $k_{off} = 0.1\ s^{-1}$ (data for ClpS2 V1 against FGVECK-biotin) from [25].
- For Trp – Assume: K_d= 11.2 μM (data for ClpS2 St-V1 against WRVECK-biotin) from [26]. Assume $k_{off} = 0.76\ s^{-1}$ (data for ClpS2 V1 against WGVECK-biotin) from [25].
- For Tyr – Assume: K_d= 11.6 μM (data for ClpS2 V1 against YGVECK-biotin) from [25]. Assume $k_{off} = 0.5\ s^{-1}$ (data for ClpS2 V1 against YGVECK-biotin) from [25].

$$f_{b,\,equilibrium-Phe} = \frac{10\ \mu M}{10\ \mu M + 1.1\ \mu M} \simeq 90\%$$

$$f_{b,\,equilibrium-Trp} = \frac{10\ \mu M}{10\ \mu M + 11.2\ \mu M} \simeq 47\%$$

$$f_{b,\,equilibrium-Tyr} = \frac{10\ \mu M}{10\ \mu M + 11.6\ \mu M} \simeq 46\%$$

3. Diffusion of proteins out of the gel:

To calculate the percent of binders that diffuse out of the gel, we model the gel as a 3-d rectangular prism, where the binder is initially confined in the gel with dimensions 2 x 2 x 0.35 mm.

In a 1-d system, where the binder is confined in a region of $-h < x < +h$ and assuming a constant diffusion coefficient [85], this is described by the equation, from Fick's second law of diffusion:

$$C(x,t) = \frac{1}{2} C_0 \left\{ erf \frac{h-x}{2\sqrt{Dt}} + erf \frac{h+x}{2\sqrt{Dt}} \right\}$$

where erf is the error function defined as:

$$erf\ z = \frac{2}{\pi^{1/2}} \int_0^z exp(-\eta^2)\, d\eta$$

and where $C(x,t)$ is the concentration at point x at time t.

In the 1-d case, the initial total amount of binder present in the gel before diffusion ($n_{start-1d}$), is given by:

$$n_{start-1d} = C_0 * 2h$$

where C_0 is the initial concentration.

The total amount of binder present in the gel, after diffusion from the wash is $n_{wash-1d}$, and the fraction of binder starting material still in the gel (M_{1d}) after a defined time, t in seconds is then:

$$n_{wash-1d}(t) = \int_{-h}^{h} C(x,t)\, dx$$

$$M_{1d}(t) = \frac{n_{wash-1d}(t)}{n_{start-1d}}$$

Thus,

$$n_{wash-1d}(t) = \int_{-h}^{h} \frac{1}{2} C_0 \{erf\frac{h-x}{2\sqrt{Dt}} + erf\frac{h+x}{2\sqrt{Dt}}\} dx = 2C_0\sqrt{Dt}\left(\frac{h}{\sqrt{Dt}} erf(\frac{h}{\sqrt{Dt}}) + \frac{e^{-h^2/Dt}}{\sqrt{\pi}} - \frac{1}{\sqrt{\pi}}\right)$$

$$M_{1d}(t) = \frac{n_{wash}(t)}{n_{start}} = \frac{\sqrt{Dt}}{h}\left(\frac{h}{\sqrt{Dt}} erf(\frac{h}{\sqrt{Dt}}) + \frac{e^{-h^2/Dt}}{\sqrt{\pi}} - \frac{1}{\sqrt{\pi}}\right)$$

The solution in three dimensions, for the gel, is the product of the solutions for each dimension:

For x and y dimensions; $h = 1\,mm$ with $-h < x < +h$ and $-h < y < +h$; and in the z dimension, $w = 0.175\,mm$ with $-w < z < +w$

Then:

$$n_{start-3d} = C_0 * 2h * 2h * 2w$$

$$n_{wash-3d}(t) = \int_{volume} C(x,y,z,t)\,dx\,dy\,dz \text{ and } M_{3d}(t) = \frac{n_{wash-3d}(t)}{n_{start-3d}}$$

Assuming diffusion process can be separable in Cartesian coordinates, where diffusion along x, y, and z is independent:

$$M_{3d}(t) = \left(\frac{\sqrt{Dt}}{h}\left(\frac{h}{\sqrt{Dt}} erf(\frac{h}{\sqrt{Dt}}) + \frac{e^{-h^2/Dt}}{\sqrt{\pi}} - \frac{1}{\sqrt{\pi}}\right)\right)^2 * \frac{\sqrt{Dt}}{w}\left(\frac{w}{\sqrt{Dt}} erf(\frac{w}{\sqrt{Dt}}) + \frac{e^{-w^2/Dt}}{\sqrt{\pi}} - \frac{1}{\sqrt{\pi}}\right)$$

Assuming a wash of 60 seconds to 180 seconds, with a diffusion coefficient ⅕ of $D_{ClpS2\ St-V1}$ to account for diffusion in the gel compared to diffusion in free solution, based on existing literature [82].

Then, the estimate becomes:

$$M_{3d}(60\ seconds) \simeq 82\%$$

$$M_{3d}(180\ seconds) \simeq 69\%$$

4. <u>Calculations for ClpS2 St-V1 in the gel, after binding (a), after the wash (b), and after fluorescent antibody staining (c)</u>

a) After, the first binding of ClpS2 St-V1: $f_{b,\,equilibrium-Phe} \simeq 90\%$, $f_{b,\,equilibrium-Trp} \simeq 47\%$ and $f_{b,\,equilibrium-Tyr} \simeq 46\%$. This means that in the 1.4 µL gel

with 70 pmol peptide, there is then a number of moles ($n_{equilibrium}$) of ClpS2 St-V1 in the gel: $n_{equilibrium-Phe} \simeq 63\ pmol$, $n_{equilibrium-Trp} \simeq 33\ pmol$ and $n_{equilibrium-Tyr} \simeq 32\ pmol$

b) After the wash, assuming ~75% of binders remain in all of the gels (from calculation point 3, above): $n_{wash-Phe} \simeq 47\ pmol$, $n_{wash-Trp} \simeq 25\ pmol$ and $n_{wash-Tyr} \simeq 24\ pmol$

c) In the 30 µL incubation in anti-HA antibody, the concentration of binder (C_{binder}) in solution, for each gel, is then: $C_{binder-Phe} \simeq 1.5\ \mu M$, $C_{binder-Trp} \simeq 0.75\ \mu M$ and $C_{binder-Tyr} \simeq 0.73\ \mu M$. Then, the fraction bound before imaging is defined by the more complex binding equation (since we cannot work under the $[ClpS2\ St-V1\ binder]_{free} \simeq [ClpS2\ St-V1\ binder]_{total}$ assumption in this case), where $[peptide]_{total} \simeq 2.3\ \mu M$ (treating the gel-bound ligand as if it were free and uniformly distributed since the system is equilibrated):

$$f_b = \frac{([peptide]_{total}+[ClpS2\ St-V1\ binder]_{total}+K_d)-\sqrt{\left([peptide]_{total}+[ClpS2\ St-V1\ binder]_{total}+K_d\right)^2-4*[peptide]_{total}*[ClpS2\ St-V1\ binder]_{total}}}{2*[peptide]_{total}}$$

Then:

$f_{b,imaging-Phe} \simeq 37\%$, $f_{b,imaging-Trp} \simeq 5\%$, and $f_{b,imaging-Tyr} \simeq 5\%$.

So approximately ~7 fold higher intensity signal expected in the Phe gels compared to the Trp and Tyr gels.

Supplementary Note 9

This Supplementary Note serves to give more information on the discussion and analysis of oxidation of amino acids in the process of free-radical polymerization, as part of a broader discussion on N-terminal amino acid binders for **Fig. 4**.

As proteins being attached to polymerized ExM gels are experiencing a free-radical filled environment, we examined the susceptibility of amino acid side chains to oxidation as a result of free-radical polymerization. Methionine and cysteine, sulfur-containing side chains, are most prone to free radical oxidation; aromatic amino acid side chains (Tyr, Trp, Phe), and some others (His, Pro, Lys, Arg) can also be oxidized, although they are less prone[87,88]. Thus, the amino acid side chains of Met, Cys, Tyr, Phe, Trp, His, Pro, Arg, at the N-terminus of an otherwise identical peptide chain, were tested for known oxidation products after free-radical polymerization in ExM gels. We here aimed only to gauge the presence or absence of an oxidation product, and, thus, analyzed only one batch of gels and one peptide aliquot of each N-terminal peptide. Each peptide and its ratio to oxidation products was obtained by analyzing relative abundances via LC/QToF (see **Source Data** for raw traces and spectra). The known oxidation products of each amino acid side chain were determined and listed in **Supplementary Figure 8a-hi**. The raw area under the curve for each species (oxidized and non-oxidized) was documented for the following three conditions: (1) in solution with trypsin, a non-oxidative environment; (2) in solution with trypsin and ammonium persulfate (APS) and tetramethylethylenediamine (TEMED), which should capture the key environment that causes oxidation (since APS decomposes to sulfate radicals, and TEMED accelerates the decomposition of APS); (3) and ExM gel followed by trypsinization, which should similarly recapitulate the oxidative environment. Finally, we plotted the max-normalized area under the curve of the chromatogram, representing the relative abundance of each of the species. Note that the abundance of the various products cannot be compared directly, due to possible differences in ionization efficiency. However, the ratio of oxidized/unoxidized peptide (for a given oxidation product) can be estimated.

In summary: after ExM polymerization, the ratio of methionine sulfoxide/original methionine peptide went from ~0 (in non-free radical solution) to ~4 (in ExM gels) (**Supplementary Figure 8a**). The ratio of sulfinate/original cysteine peptide went from ~0 (in non-free-radical solution) to ~6 (in ExM gels) (**Supplementary Figure 8b**). On the other hand, all other amino acid side chains had a ratio of oxidation product/original peptide below ~0.07, for all oxidation products analyzed (**Supplementary Figure 8c-h**; see **Supplementary Table 11** for descriptive values). Since, in an *in situ* peptide sequencing workflow, amino acid oxidation might prevent recognition of binder reagents to the N-terminus amino acid (assuming the binders are specific to the non-modified side chain - of course, if an oxidation product is reliably produced in an ExM gel, one could simply make binders specific to the modified amino acid), side reactions, such as oxidation, might need to be regulated. Strategies might be adopted to prevent or recover oxidized residues. For instance, methionine sulfoxide can in principle be converted back to methionine by methionine sulfoxide reductase. For cysteine, reducing agents (i.e., TCEP, or DTT) can reduce disulfide bridges and thioredoxin can convert cysteine sulfenic acid (RSOH) back to a thiol group[87]. Nevertheless, some oxidation states of cysteine (i.e., sulfinate RSO_2, or sulfonate RSO_3) are likely irreversible[89]. Mitigation of side chain modifications from oxidation may benefit from the use of alternative gel polymerization strategies. Indeed, expansion microscopy has been demonstrated using a monomer that forms the ExM polymer network through click chemistry, which would be bioorthogonal to the amino acid detection questions at hand[36,37].

Supplementary Note 10 - A theoretical assessment of *in situ* protein sequencing

For our theoretical assessment of *in situ* protein sequencing, we decided to focus on the case of N-terminal binding to an amino acid while still attached to a peptide, as a conservative case for a binding event. Previous modeling approaches have explored how well a given *ex situ* single molecule protein sequencing or fingerprinting approach would perform in terms of proteomic coverage, using various proxies or metrics[90–96]; we here sought to develop such a framework for in situ protein sequencing.

In situ protein sequencing, using ExM chemistries and in-gel Edman degradation, involves stochastic chemical steps that need to be modeled as accurately as possible in order to assess the feasibility of mapping empirical protein sequences to the actual proteome (**Supplementary Figure 13a** for schematic of the whole experimental workflow) - or to be modeled pessimistically, to set a bound on future performance. Common alignment techniques, like BLASTP, are useful for homology detection between proteins, but are not suitable for *in situ* protein sequencing modeling. This is because the error profiles (from mismatches, gaps and deletions) of BLASTP do not consider the chemistries at hand (e.g., fixation, anchoring and digestion) and the read-out errors that are expected (i.e., in-gel Edman degradation reaction efficiencies, number of binders and their specificities), nor the complexity of sequencing multiple fragments from one protein, as expected in *in situ* protein sequencing. We thus set out to build our own model specifically for *in situ* protein sequencing (**Supplementary Figure 13b** for an outline of our modeling strategy).

We chose to perform these analyses on *Mycoplasma genitalium* and human proteomes. The former (referred to as "mycoplasma" hereafter) has the smallest proteome (483 Swiss-Prot reviewed proteins; UniProt ID 243273: by searching "taxonomy_id:243273" in UniProt), and thus could be a useful testbed for in situ protein sequencing. The human proteome (20,421 Swiss-Prot reviewed proteins, UniProt ID 9606: by searching "taxonomy_id:9606" in UniProt) is of course key to confronting human disease. We explored how varying the type and success rate of each chemical step (e.g., different anchoring chemistries, different digestion enzymes; % fixation, % anchoring, % digestion) affects the fidelity of the peptide fragments retained in the gel. In particular, we defined "percent amino acid count" as the percentage of residues of a protein that are not modified, and remain accessible for read-out, for 15 rounds of in-gel Edman degradation after chemical preprocessing (these results are contained in the section below, **Theoretical assessment Part 1: results regarding fixation, anchoring, gelation, and digestion**).

For a given parametrization of the chemical steps explored above (i.e., probabilities of 0.05 for fixation, 0.8 for anchoring with AcX, and 0.8 for digestion with trypsin), we subsequently simulated possible outcomes of the series of steps many times, and stored the results in a dataset containing the list of all such peptide fragments (which we call the "reference fragment dataset", representing the ground-truth sequences). Then, we simulated the rest of the workflow, modeling the errors expected with in-gel Edman chemistry, and molecular recognition with NAABs. This led to output reads for every fragment retained in gel, which we call the "error-prone fragment dataset". We assumed prior knowledge of which fragments originated from each parent protein, with no mixing or exchange occurring between neighboring proteins. (Note well, this assumption may not hold for protein complexes that are densely packed, e.g. many receptors that are multimeric, or the ribosome, or other structures - a full treatment of the problem will require additional work.) These assumptions were inspired by recent findings in the literature showing that it is possible to reach ~1 nm resolution, and to identify protein shape with ExM[12]. Fragment read-outs were then independently evaluated for their match to proteins, using the reference fragment dataset previously mentioned. A final assignment to a protein was performed if the number of fragments matching a protein was at least double that of the next best matching protein, a somewhat arbitrary choice (these results are contained in the section below, **Theoretical assessment Part 2: results for NAAB binding and Edman degradation**).

Theoretical assessment Part 1: considerations for fixation, anchoring, gelation, and digestion

The first step for a biological specimen involves fixation (**Supplementary Figure 13a**), which stabilizes the cellular architecture and the spatial relationships between proteins[97]. It is often

performed with formaldehyde, a cross-linking agent that forms chemical bonds between amino acids, in ExM[7,11,56,98,99]. Chemical fixation via formaldehyde involves primarily amine (e.g., N-termini (or N-term for short) and lysines) and thiol groups (e.g. cysteine) forming methylol groups, and a subsequent step involving crosslinking via methylene bridges which can include other amino acid side chains (e.g., arginine, tyrosine, and to a lesser extent asparagine, glutamine, histidine and tryptophan)[97,100]. As such, our model was constructed to modify residues most likely to be fixed (e.g. N-term, lysine, cysteine, arginine, tyrosine), under the assumption that they are modified equally (arbitrarily, since we had no ground-truth data to support our decision), at a given probability (**Supplementary Figure 13b**). Alternative fixation strategies exist that rely on organic solvents, such as methanol or ethanol, which do not form chemical bonds between amino acids on proteins. Such fixation has been demonstrated with ExM[101]. However, this comes with a likelihood of poorer soluble-protein retention[102], and thus higher protein loss. Protein loss is a concern for any fixation strategy, and various fixation strategies and their relative protein retention have been studied, with for instance a decrease from ~65% to ~8% protein loss via one form of hydrogel embedding[103]. These results, and others[104,105], suggest that hydrogel-tethering can retain the majority of proteins under the right conditions. For this reason, we chose not to account for protein loss in the chemical fixation step. This also allowed us to isolate the percent amino acid count (defined above) from fixation parameters that could affect protein loss. It is important to note that protein loss beyond the fixation step was accounted for in our model, in subsequent steps such as anchoring, digestion and in-gel Edman degradation.

Anchoring is the step after fixation (**Supplementary Figure 13a**), and included in ExM to create covalent linkages between proteins and the hydrogel network. For anchoring chemistries, we compared two strategies previously reported in ExM studies, using the bifunctional molecules AcX and acrylate epoxide (which we abbreviate “epoxide”). AcX harbors an NHS (N-hydroxy succinimide) ester moiety that reacts with primary amines (e.g., N-terminus and lysine side chains). We assumed for AcX anchoring that all primary amine groups were modified with equal probability (**Supplementary Figure 13b**). On the other hand, epoxide was assumed to react with multiple side chains (e.g., N-term, lysine, cysteine, histidine, tyrosine aspartic acid, and glutamic acid)[106], leading to different outcomes in fragment retention in the gel. For epoxide anchoring we assumed that all side chains, listed above, were modified with equal probability (not depicted on **Supplementary Figure 13b**). In general, we assumed that protein anchoring to the gel occurs independently for each molecule. We also assume, for the purposes of this model, that amino acids are anchored at the same probability regardless of where they are within the tertiary structure of the protein; of course, this may not be the case in real life. (Note well - in practice we can denature proteins, and/or iteratively anchor and expand, to compensate for such issues.)

After a sample is anchored, the gel is cast. During the gelation step we assumed no change in primary protein sequence. Finally, for digestion (**Supplementary Figure 13a**), we wanted to compare enzymatic cleavage using trypsin (targeting K and R sidechains as shown in **Supplementary Figure 13b**), endoproteinase Lys-C (Lys-C; targeting lysine sidechain), and proteinase K (ProK; targeting aliphatic, aromatic and hydrophobic side chains including Y, W, E, L, V, A, I, F, and T[107–109]). These digestion enzymes are commonly used in ExM to chemically soften the sample and allow for expansion to occur[7,11,12,110,111]. Alternatively, ExM can be performed with a denaturation step (e.g., using beta-mercaptoethanol, SDS and/or high temperature, i.e., 95 °C)[7,112,113]. We did not explore denaturation further in our modeling, since fixation and anchoring, without a downstream digestion step, does not lead to fresh N-termini for sequencing via in-gel Edman degradation. Digestion enzymes expose fresh N-terminal sites that can be targeted by in-gel Edman degradation, generating several fragments of a single protein that can be sequenced in parallel. Sequencing many fragments in parallel would lead to many shots on goal. This, of course, assumes that the peptide fragments derived from the digestion of a single protein are sufficiently separated to be distinguished as discrete fluorescent puncta. We assumed prior knowledge of which fragments originated from each parent protein, with no mixing or exchange occurring between neighboring proteins. During this evaluation, at a given probability of digestion, it was assumed that all amino acids targeted by the enzyme were equally accessible for digestion, unless the residue was anchored to the gel, in which case the enzyme

was modeled to be unable to cleave the downstream amide bond (due to specific side-chain recognition loss for the various enzymes[107–109,114–116]). All other amino acids were assumed available given that proteins can be denatured and linearized into peptide fragments prior to digestion and that we had no other ground-truth data to support our decision. Simulation of digestion is depicted in **Supplementary Figure 13b** for trypsin cleavage.

Importantly, we note that all of the chemical parameters for fixation, anchoring, and digestion can be independently modified (not just the probability values of the parameters, but also the reagent, thus affecting amino acid specificity at each step) within our code (**Methods: Theoretical assessment of in situ protein sequencing** section for code) to test alternatives.

Theoretical assessment Part 1: results regarding fixation, anchoring, gelation, and digestion

As mentioned above, we defined "percent amino acid count" as a metric to assess the number of amino acids that remain accessible for downstream sequencing for a given set of values in the chemical parameter space. We used this metric to determine the mean percent amino acid count, which is the percent amino acid count averaged over all proteins of the proteome, for several combinations of values in the parameter space. In this analysis, we assumed that for the human proteome, the first fragments of proteins were inaccessible for sequencing, since N-terminal acetylation affects ~70-80% of proteins[117], thus preventing the initiation of Edman degradation on original N-termini. However, this was not assumed for the mycoplasma proteome, since, although the extent of this post-translational modification (PTM) is not well characterized in *Mycoplasma genitalium*, N-terminal acetylation is significantly less common in bacterial and archaeal proteomes (10-29% N-terminal acetylation for *Mycobacterium tuberculosis* and *Pseudomonas aeruginosa PA14*)[118].

We conducted a systematic overview of how different fixation rates, anchoring rates for two anchoring methods (AcX, epoxide), and digestion rates for three digestion methods (Lys-C, trypsin, ProK), would affect the mean percent amino acid count in the mycoplasma and human proteomes. We varied the probability of reactivity towards amino acid targets from 0 to 1 (where 0 is no reaction, and 1 is complete reaction, assessed independently at each amino acid in the sequence). The results of this grid search are visualized in a heatmap (**Supplementary Figure 13c** and **Supplementary Figure 18**, for mycoplasma and human proteome results, respectively; raw data CSV files are available in our code on GitHub in **Methods: Theoretical assessment of in situ protein sequencing**). We also analyzed the variability in the percent amino acid count for certain parameters (discussed more below; with distributions plotted in **Supplementary Figure 14**).

Some patterns emerged: increased fixation generally reduced the number of residues available for anchoring to the gel network and led to decreased percent amino acid count in the gel, with the caveat, mentioned above, that we did not model protein loss at low fixation (**Supplementary Figure 13c**, going from left to right; **Supplementary Figure 18** for the same results in the human proteome). The results also showed differences in mean percent amino acid count when comparing AcX and epoxide (**Supplementary Figure 13ci-iii** for AcX, **Supplementary Figure 13civ-vi** for epoxide for the mycoplasma proteome; **Supplementary Figure 18i-iii** for AcX, **Supplementary Figure 18iv-vi** for epoxide for the human proteome). For AcX anchoring, fixation directly competed with anchoring since both processes target amines. Thus, as fixation probability increased, anchoring efficiency with AcX decreased, leading to greater fragment loss and less mean percent amino acid count (**Supplementary Figure 13ci-iii, Supplementary Figure 18i-iii**). For epoxide, the effect was not as pronounced because epoxide targets additional residues not modeled to be reactive to fixative (**Supplementary Figure 13civ-vi, Supplementary Figure 18iv-vi**).

Some anchoring was essential, as no anchoring resulted in complete fragment loss (**Supplementary Figure 13c** and **Supplementary Figure 18**). On the other hand, complete anchoring eliminated all Lys-C cleavage sites and prevented digestion entirely (**Supplementary Figure 13ci** and **Supplementary Figure 13civ** and **Supplementary Figure 18i** and **Supplementary Figure 18iv**). The best anchoring probability to maximize mean percent amino acid count from our model, represented a balance between these competing effects, with different ideal conditions (based on our assumptions) for AcX and epoxide due to their different amino acid selective reactivities. For

instance, for AcX with Lys-C digestion for P(fix)=0.1, the best option was P(anchor)=0.6. However, for epoxide with Lys-C the best condition was P(anchor)=0.4 (i.e, shifted to the left, towards lower anchoring probability). This can be explained since, at a similar reaction anchoring efficiency rate, more residues are modified with epoxide than AcX. Our results suggested that epoxide, on average, was able to retain more mean percent amino acid count in its best conditions than AcX in its best conditions, at least in this theoretical assessment.

We next compared the effects of the three proteases (trypsin, Lys-C and ProK) on the change in percent amino acid count of fragments, and overall fragment lengths. Fragment length was defined as the number of amino acids of a fragment retained in the gel, not considering amino acids that are located downstream (C-terminal) of the last anchoring amino acid of the fragment (since this fragment would be lost after in-gel Edman degradation), but unlike the percent amino acid metric, was not limited to the first 15 amino acids of the fragment. For ProK, a P(digest)=0.2 resulted in the highest mean percent amino acid count. Higher probabilities, above 0.2, resulted in higher fragment loss and digestion leading to an overall lower mean percent amino acid count (**Supplementary Figure 13ciii** and **Supplementary Figure 13cvi** for mycoplasma proteome, **Supplementary Figure 18iii** and **Supplementary Figure 18vi** for the human proteome). However, for both trypsin and Lys-C conditions, a maximal probability of digestion, P(digest)=1.0, enabled the highest mean percent amino acid count (**Supplementary Figure 13ci-ii** and **Supplementary Figure 13civ-v; Supplementary Figure 18i-ii** and **Supplementary Figure 18iv-v**). This is consistent with the fact that ProK has broad specificity against amino acids (9 targets;[107–109]), thus generating many more fragments with free N-termini at a given P(digest) compared to trypsin (2 targets) or Lys-C (1 target) (see fragment lengths in **Supplementary Figure 15a-b** for mycoplasma proteome and human proteome, respectively). Taken together, these results suggested ProK could reach its highest mean percent amino acid count by tuning its probability of digestion to P(digest)=0.2, but above this value would start to lead to fragment loss and very short fragments (e.g. ~2 amino acids in median length with P(digest)=0.8) leading to a reduction in mean percent amino acid count. For Lys-C, the median fragment length was above 15 amino acids with P(digest)=0.8, suggesting incomplete coverage of the retained fragments with only 15 rounds of in-gel Edman degradation. In addition, for Lys-C and trypsin, given that their best condition was at their maximal digestion capabilities (P(digest)=1.0), both of these enzymatic strategies may not be reaching ideal fragment lengths for maximal coverage over 15 rounds of in-gel Edman degradation.

Finally, we looked at the distribution of percent amino acid count in the gel for every protein for a given set of the chemical parameters, to have a sense of the variability of this metric. In the case of digestion with Lys-C with either AcX or epoxide anchoring (P(fix)=0.05, P(anchor)=0.8 with AcX or epoxide, and P(digest)=0.8 with Lys-C), there were 4% and 11% of the proteins in the mycoplasma and human proteomes, respectively, with <2.5% amino acid count in the gel (**Supplemental Figure 14a,** two leftmost panels, for the mycoplasma proteome and **Supplemental Figure 14c,** for the human proteome). This percentage of proteins was therefore mostly lost in the process of anchoring or digestion. On the other hand, with the same P(fix), P(anchor) and P(digest) values, epoxide with trypsin led to the highest median percent amino acid count in the gel, where close to all proteins showed a percent amino acid count >2.5% (**Supplemental Figure 14a,** bottom rightmost panel, where mycoplasma has ~1.2% of proteins and **Supplemental Figure 14c** human has 0.6% of proteins with <2.5% amino acid count in the gel). We also analyzed the distribution for the percent amino acid count for the best parametrization (i.e., the one that maximized the mean percent amino acid count) for each condition in **Supplemental Figure 14b** and **d** (mycoplasma, human). In this case, all conditions' distributions were shifted towards higher mean percent amino acid count and fewer proteins had <2.5% amino acid count. However, for the human proteome, unlike the mycoplasma proteome, Lys-C digestion with AcX or epoxide still led to >2% of proteins with <2.5% amino acid count (the mycoplasma proteome showed <1% proteins with <2.5 % amino acid count in the same conditions). This discrepancy arises from modeling human proteins with capped N-termini from N-terminal acetylation, unlike those in the mycoplasma proteome, combined with the limited ability of Lys-C digestion to generate sufficient new free N-termini (**Supplemental Figure 14b** for mycoplasma, and **Supplemental Figure 14d** for human).

Theoretical assessment Part 2: considerations for NAAB binding and Edman degradation

For simulating *in situ* protein sequencing with in-gel Edman degradation and NAAB sequence read-out, where we assume that NAABs bind PITC-modified N-terminal amino acids. We assumed this since this approach would yield a read-out similar to that of NAABs targeting native N-terminal amino acids, except that PITC conjugation at the N-terminus could enhance NAAB dwell times required for reliable imaging detection. We selected probability values for each of the fixation, anchoring and digestion steps explored above. Specifically, we selected 0.05 for fixation, AcX at 0.8 probability of anchoring and 0.8 probability of digestion with trypsin. We made the 0.05 fixation selection, since amino acid side chain modifications on the order of ~3-22% conversion have been observed when incubating peptides with 50 times excess formaldehyde, at pH 7.2 and 35 °C for 48 h[97,100]. (Again, we did not consider protein loss related to incomplete fixation.) AcX is the most commonly used anchoring chemistry used in ExM protocols; although the true percentage of amine groups modified by AcX is unknown, NHS esters react selectively and efficiently with primary amines. Due to potential limitations on reaction probability by tertiary protein structure in a cell, we considered P(anchor)=0.8 when modeling the expected probability of AcX reaction to amines. Finally, for digestion, although the true value of digestion efficiency is unknown in-gel, it has been shown that trypsin digestion efficiency is ~80% in samples from yeast total protein extract analyzed with mass spectrometry[119]. Thus, we considered P(digest)=0.8. Since epoxide and ProK, rather than AcX and trypsin, showed the strongest performance (these results are contained in **Theoretical assessment Part 1: results regarding fixation, anchoring, gelation, and digestion**), the model might be expected to perform less well with AcX and trypsin than with epoxide and ProK, when considering their respective best parametrization.

Using this selection, we constructed an indexing structure (trie) that captures the space of different fragments that could be generated and retained in the gel, which we also call our reference fragment dataset (**Supplementary Figure 13b**). We simulated the experiment 1,000 times (to be conservative; varying the number of runs from 1 to 10,000 did increase the number of unique fragments that were added to the reference fragment dataset; **Supplementary Figure 16**). Consequently, selecting only 1,000 simulations constrained the reference fragment dataset, likely underestimating our ability to map back to the proteome. Using this approach, we sought to determine whether it was still possible to map a significant portion of the proteome.

In terms of modeling errors with in-gel Edman chemistry, we modeled conjugation with 10% PITC failure, and cleavage of N-terminal amino acid with TFA as 30% failure (**Figure 3g** suggesting ~70% yield). In addition, we assumed that in-gel Edman degradation can proceed past fixed or anchored amino acids. Indeed, previous work has demonstrated Edman degradation proceeding through bulky side-chains, including covalent modification of side-chains with fluorophores[15], arginylation of the side chain of acidic amino acids[120] and a 30 amino-acid polyproline linker between the side chain and the fluorophore[94], suggesting that this may be a realistic assumption.

In terms of the read-out with the NAABs, we assumed that binders cannot bind to fixed, anchored, or unconjugated (i.e. not bound to PITC) amino acids. To model the binding of NAABs for read-out, we evaluated how increasing the number of NAABs (targeting 5, 10, 15 or 20 amino acids), and increasing the specificity of the NAABs, impacted protein identification. PTMs, present on certain amino acid side chains (most abundantly lysine, tyrosine, asparagine, threonine, and serine), were not taken into consideration. However, since ~0.2% of side chains carry a PTM (~300k experimentally validated modifications out of ~190M amino acids, as found for the Swiss-Prot database[121]), we reasoned that this assumption would not alter the main results of our model. Modifications to amino acid side-chains by free-radical polymerization were not considered since results suggest that oxidation may remain minimal on most sidechains, except cysteine and methionine (**Supplementary Figure 8a-h**), and also because alternative chemistries exist that avoid such side reactions, such as tetra-gel ExM, which does not rely on free-radical polymerization[36,37]. Binders were assumed to bind PITC-conjugated N-terminal amino acids (with each having a specificity towards one amino acid). To simulate a range of specificities and affinities, we constructed a binding affinity matrix shown in **Supplementary Table 12,** for “perfect”, “very high”, “high”,

"medium", "low" and "very low" binding profiles. Specifically, perfect binding was used as a theoretical limit of the read-out, modeled as a binary outcome for on-target versus off-target. The medium binding scenario was modeled as a binder with 500x longer dwell time against the on-target compared to the off-target, where the on-target binding affinity has K_d ~50 pM , k_{off} ~5 * 10^{-5} s^{-1} (as found for some high-affinity antibodies[122]), whereas off-target amino acid affinity was modeled as K_d ~25 nM, k_{off} ~2.5 * 10^{-2} s^{-1}. The low binding scenario was modeled with on-target kinetics as before, but only assumed a 50x longer dwell time for on-target compared to off-target binding (k_{off}~2.5 * 10^{-3} s^{-1}). Finally, for very low affinity, we similarly modeled on-target kinetics as before, but only assumed a 5x longer dwell time for on-target compared to off-target binding (k_{off}~2.5 * 10^{-4} s^{-1}). During the simulation, we performed sequential binder read-out (where the binders are not added simultaneously, but one at a time). We assumed that for each binder, the kinetics of binding reached equilibrium conditions prior to washing. In addition, we did not consider any specific effects that the gel environment would have on the kinetics of binding, nor any non-specific background from binders bound non-specifically to the gel or not washed out. To understand how binder concentration and wash time could impact read-out under these assumptions, we plotted heatmaps showing how different conditions influence difference in probability of on-target binding and off-target binding for very high, high, medium, low and very low specificity binders (**Supplementary Figure 17** for these heatmaps). Assuming an excess of binder compared to targets, and from these results in **Supplementary Figure 17**, we selected a 1 μM binder concentration because it fell in the range of values with a large difference between probability of on versus off-target binding (large differences could be noted for concentrations between 10 nM and 1 mM). Also since the binder is assumed to have high affinity, 1 μM is well above its K_d value of 50 pM. In addition, even if the concentration of target (i.e. an N-terminal amino acid) has a higher concentration than 1 μM in the gel, since the target is immobilized in the gel, the solution volume is assumed to be in large enough excess that the binder depletion is negligible for the Langmuir equation to hold (**Supplementary Figure 17** caption for the equations describing the binding kinetics). We selected a 30 min wash prior to imaging since this fell in the range of ~0.8 for the difference in probability for on versus off target binding and can be a realistic timing for a wash step (although diffusion may pose a constraint during a 30 min wash, we did not model this further). For the sequential binder read-out, we assumed that we can successfully wash out the binders from their targets by stripping them from the gel, and perform a second round of binding and imaging (as performed previously[112]).

Using the parameters described above, we generated the error-prone fragment dataset (in contrast to the reference fragment dataset that did not account for readout errors) through Monte Carlo simulation. Monte Carlo simulation captures the stochastic realization of fragments generated from imperfect chemical steps, and was a natural way to generate synthetic data for the error prone fragment dataset. Building on previous single-molecule sequencing studies that have used Hidden Markov Models (HMMs)[91,93], we mapped the resulting fragments back to the reference fragment dataset using a HMM. HMMs in this context predict the likelihood that a sequence of hidden states (the true amino acids of a protein) produce the observed data (noisy amino acid readout from correct plus incorrect binding). We first remove any "uninformative fragments", here defined as sequences with more than half unknown reads ("X"), from both the error-prone fragment and reference fragment database. Then, in the first stage of the HMM, we applied the Viterbi algorithm to each error-prone sequence to efficiently identify the single most likely peptide sequence given the observed binder readouts and modeled failure modes; this choice maximized computational speed and paralleled the pruning strategies used by Smith *et al.* for rapid path selection[93]. This dynamic programming approach traces the most likely sequence of hidden states, which we then use to remove candidate sequences whose log-likelihoods are less than about 1/150 as probable as the top sequence. In the second stage of the HMM, we used the forward algorithm to evaluate the top candidate alignments. Unlike Viterbi, which returned only the most likely path, the forward algorithm summed over almost all possible paths to compute the overall likelihood of the observed fragment from the reference fragments. This provided a probabilistic score that captured uncertainty across alternative fragment assignments and allowed us to rank candidate proteins. Unlike previous implementations, our HMM explicitly integrated binding-specificity parameters and chemically derived error rates at each degradation

cycle, relevant for the unique challenges of our *in situ* protein sequencing platform, which requires evaluation of how experimental variables influence the overall protein identification accuracy (i.e., including the chemical steps: fixation, anchoring and digestion and Edman degradation and readout inefficiencies). We chose to run the simulation with in-gel Edman degradation from 5 to 15 cycles, as the median length of a fragment is ~15 with the same parameters described above (**Supplementary Figure 15a-b**). The specific amino acid binders that were selected for smaller subsets (5, 10 and 15 binders) were chosen arbitrarily and the list of binders in those subsets are noted in the captions of **Supplementary Figure 13d**.

For the final assignment of fragments to a parent protein, a fragment was assigned only if the highest-scoring protein, computed as the sum of the weighted log-likelihoods across the top ten candidate fragments, had a total score at least 1.5x greater than that of the second-highest-scoring protein, otherwise it was "uncertain". At the protein level, a parent protein was considered identified if the total number of its assigned fragments was at least 1.5x the number of fragments supporting the next-highest-scoring protein, otherwise it was "uncertain". We recorded if each identification was "correct", "false positive", or "uncertain". See **Methods: Hidden Markov Model (HMM) based matching and fraction of proteome correctly identified** for more details.

Theoretical assessment Part 2: results for NAAB binding and Edman degradation

In **Supplementary Figure 13d**, we quantified the percent identified proteins from a proteome as the fraction of proteins for which our model's top candidate matched the true source protein. We plotted the "fraction correct" of the total mycoplasma proteome, defined as the number of proteins identified as correct (see above, i.e. not false positive and not uncertain; see **Methods: Hidden Markov Model (HMM) based matching and fraction of proteome correctly identified** for more details) divided by the size of the proteome, averaged across 10 repetitions. Each of the four panels shows how this fraction correct changes as we increase the number of Edman rounds, under four binder-specificity regimes (perfect, medium, low, and very low) and four different binder library sizes (20, 15, 10, and 5 amino acid binders). From **Supplementary Figure 17** and **Supplementary Table 12**, we note that high and very high specificity binders under chosen wash times of 30 minutes and concentrations of 1 μM resulted in similar on and off-target binding probabilities, and thus were not plotted and concluded to be comparable to medium specificity. With 20 amino acid binders, perfect, medium, and low specificity converged to >90% accuracy by round 8, and reached 98-99% accuracy by round 12, whereas very low specificity remained at almost zero. Dropping to 15 amino acid binders looked very similar, with convergence of perfect, medium, and low specificity at >90% accuracy by round 8, and reaching ~98% accuracy by round 12. At 10 amino acid binders, perfect and medium specificity binders required 12 rounds to hit ~90%, while low specificity was roughly ~85% at 12 rounds. Finally, with only 5 amino acid binders, even perfect specificity binders struggled to pass ~60% identified proteins after 15 rounds, and medium and low specificity remained below ~60%.

We observed a plateau for perfect binding even with all 20 amino acid binders, with ~99% rather than 100% correctly identified proteins. This plateau occurred since some proteins have a low number of lysine residues that can be anchored to the gel matrix, resulting in lost proteins and/or few retained fragments. An example is protein Uniprot accession P47377 in the mycoplasma proteome with only 2 lysines. Thus, increasing the number of amino acid binders may yield diminishing returns beyond 15 distinct amino acid targets, where > 90% of proteins can be correctly identified after 8 in-gel Edman rounds for both 15 and 20 amino acid binders across perfect, medium, and low specificity binders (**Supplementary Figure 13di-ii**). We reiterate that in this model, all on-target binding is considered to have a k_{off} of $5 * 10^{-5}$ s^{-1} irrespective of the specificity (**Supplementary Table 12**). Moreover, the magnitude of the k_{off} difference between on-target and off-target binding was critical: a 50-fold difference enabled correct protein identification, whereas a 5-fold difference did not, as reflected in the loss of identified proteins at very low specificity. There were diminishing returns in higher specificities, as noted by the similarity of curves in medium (500x difference between on-target and off-target in k_{off}) and low (50x difference) specificity (**Supplementary Figure 13di-iv**). Increasing the number of in-gel Edman rounds resulted in higher percentages of correctly identified

proteins. Finally, with the smaller binder library sizes (5 and 10 NAABs), increasing the number of Edman rounds to 15 did not fully rescue the percent proteins correctly identified, reaching ~90% for low specificity binders with 10 NAABs, and ~50% for medium specificity binders with 5 NAABs.

Together, these curves map out an estimate of how certain characteristics of *in situ* protein sequencing impact the percentage of correctly identified proteins, under the assumptions of our model. The results could inform experimental directions, such as which chemical steps require the most attention, in the future. The most critical development awaiting, of course, is the creation of new binders. As underlying chemistries are updated, to achieve more realistic simulations, we will need to update the assumptions of our model, to better constrain the parameters by empirical data. Once we do have sufficient empirical data, a more straightforward strategy would be to use standard machine learning techniques.

Supplementary Tables

Supplementary Table 1

Table documenting different in situ proteomic technologies compared based on spatial resolution, sensitivity, hardware and scale (‘omics’). The table categorizes the technologies based on general strategy (left-most column), spatial resolution, sensitivity, hardware, and scale.

Comparison of spatial proteomic technologies								
					Points for comparison			
	Author	Title	Year	Link	Spatial resolution (best)	Sensitivity	Hardware	Scale ('omics')
Proximity labeling techniques	Branon *et al.*	Efficient proximity labeling in living cells and organisms with TurboID	2018	https://www.nature.com/articles/nbt.4201	Proteins will be labeled around a few nanometers of the enzyme. Averaged proteomic landscape across all the locations where the enzyme was active in the sample.	Limited to mass spectrometry sensitivity, which is at best, attomole sensitivity (approximately ~ 10^6 molecules needed for detection).	LC-MS/MS.	In theory MS can detect any protein, but is limited by: (1) Protein not in the right dynamic range to be detected (eg. protein abundance is too low) (2) Ionization efficiency of the proteins (eg. in highly complex samples, peptides can compete for ionization), (3) sample preparation considerations, (eg. membrane proteins may not solubilize, and degrade).
Expansion Microscopy with Mass Spectrometry	Drelich *et al.*	Toward High Spatially Resolved Proteomics Using Expansion Microscopy	2021	https://pubs.acs.org/doi/10.1021/acs.analchem.0c05372	~ 330 um lateral resolution.		Manual microsection; LC-MS/MS.	
	Li *et al.*	Spatially resolved proteomics via tissue expansion	2022	https://www.nature.com/articles/s41467-022-34824-2	~ 160 um lateral resolution.		Manual microsection; LC-ESI/MS.	
	Chan *et al.*	Gel-assisted mass spectrometry imaging enables sub-micrometer spatial lipidomics	2024	https://www.nature.com/articles/s41467-024-49384-w	~ 1.3 um lateral resolution.		MALDI-MSI.	
	Dong *et al.*	Spatial proteomics of single cells and organelles on tissue slides using filter-aided expansion proteomics	2024	https://www.nature.com/articles/s41467-024-53683-7	~ 73 um lateral resolution.		LC-ESI/MS.	
	Zhang *et al.*	TEMI: tissue-expansion mass-spectrometry imaging	2025	https://www.nature.com/articles/s41592-025-02664-9	~ 20 um lateral resolution.		MALDI-MSI.	
	Wang *et al.*	iPEX enables micrometre-resolution deep spatial proteomics via tissue expansion	2025	https://www.nature.com/articles/s41586-025-09734-0	~ 1-5 um lateral resolution.		MALDI-MSI.	
Tissue clearing techniques / Expansion Microscopy with multiplexed antibody staining	Chung *et al.*	Structural and molecular interrogation of intact biological systems	2013	https://www.nature.com/articles/nature12107	No expansion of tissue. Diffraction limit of light resolution ~ 300-700 nm.	At this resolution, not single molecule. Antibody read-out is dependent on affinity, and specificity. Also depends on fluorophore brightness, signal amplification and detection strategies.	Confocal microscope, (Leica SP5), single-photon and two-photon imaging.	Three rounds of antibody staining (3 colors per round). Total of 9 protein targets.
	Murray *et al.*	Simple, Scalable Proteomic Imaging for High-Dimensional Profiling of Intact Systems	2015	https://www.cell.com/cell/comments/S0092-8674(15)01505-6	No expansion of tissue. Diffraction limit of light resolution ~ 300-700 nm.		Confocal microscope, and custom-built light-sheet microscope.	22 rounds of antibody staining (3 colors per round). Total of 66 targets. total
	Ku et al.	Multiplexed and scalable super-resolution imaging of threedimensional protein localization in size-adjustable tissues	2016	https://www.nature.com/articles/nbt.3641	Expansion factor of ~ 4 fold. States ~ 60 nm lateral resolution.		Single-photon confocal laser scanning imaging (Olympus).	Seven rounds of antibody staining (3 colors per round). Total of 21 targets.
	Park *et al.*	Protection of tissue physicochemical properties using polyfunctional crosslinkers	2018	https://www.nature.com/articles/nbt.4281	Expansion factor of ~ 3 fold. (resolution not stated, but perhaps approximately ~ 130 nm).		Confocal microscope (Olympus Confocal FV1000MPE 1, Leica TCS SP8), light sheet microsope (SmartSPIM, LifeCanvas).	Five rounds of antibody staining (3 colors per round). Total of 15 targets.
	Park *et al.*	Epitope-preserving magnified analysis of proteome (eMAP)	2021	https://www.science.org/doi/10.1126/sciadv.abf6589	Expansion factor of ~ 10 fold (resolution can be estimated as approximately ~ 40 nm).		Confocal microscope (Leica TCS SP8).	Three rounds of antibody staining.
	Bai *et al.*	Expanded vacuum-stable gels for multiplexed high-resolution spatial histopathology	2023	https://www.nature.com/articles/s41467-023-39616-w	Expansion factor of ~ 3.7 fold (resolution can be estimated as approximately ~ 100 nm).		Custom MIBI-TOF mass spectrometer equipped with an oxygen duoplasmatron ion gun (Alpha), a custom MIBI-TOF mass spectrometer (Betty) equipped with a xenon ion source (Hyperion, Oregon Physics), and a commercially available MIBIscope from Ionpath equipped with a xenon ion source (with MIBI software version v.1.7.0-0f60ffbc).	Detects > 40 targets total using isotope-conjugated antibodies.
	Park *et al.*	Integrated platform for multiscale molecular imaging and phenotyping of the human brain	2024	https://www.science.org/doi/10.1126/science.adh9979	Expansion factor of ~ 4.5 fold (resolution can be estimated as ~ 90 nm).		Confocal and MegaSPIM light-sheet microscope.	Seven rounds of antibody staining.
	Kang *et al.*	Multiplexed expansion revealing for imaging multiprotein nanostructures in healthy and diseased brain	2024	https://www.nature.com/articles/s41467-024-53729-w	Expansion factor of ~ 20 fold, and ~ 20 nm lateral resolution.		Nikon CSU-W1 or SORA confocal microscope.	Seven rounds of antibody staining.

Supplementary Table 2

A table documenting a non-exhaustive list of Edman degradation variations and optimizations over the years (from 1950 to 1996), including a wide range of conjugation solvents and cleavage conditions. All of the strategies rely on phenylthiohydantoin amino acid (PTH-aa) detection to separate and extract the PTH-aa.

Edman degradation protocols

Title	Author	Year	Conjugation condition	Cleavage condition	Link
Method for Determination of the Amino Acid Sequence in Peptides	Pehr Edman	1950	50% pyridine in water at 40°C with NaOH to about pH 9. 4.8% PITC (v/v) was added to that solution for 30 minutes. The pH was maintained throughout the reaction by adding small portions of NaOH.	Anhydrous nitromethane saturted with hydrogen chloride, 15 minutes.	http://actachemscand.org/pdf/acta_vol_04_p0283-0293.pdf
A technique for stepwise degradation of proteins from the amino-end	H. Fraenkel-Conrat	1954	20% PITC in peroxide-free dioxane incubation for 2-3h at 40 °C.	A mix of glacial acetic acid and 5.7N HCl incubation for 4-16h.	https://pubs.acs.org/doi/10.1021/ja01642a085
Quantitative determination of N-terminal amino acids in some serum proteins.	Sten Eriksson and John Sjöquist	1960	100:3:1 of pyridine:triethylamine:PITC incubation at 40°C for 1.5 hours.	Mix of 1 mL water and 2 mL of HCl saturated with acetic acid and kept for 2 hours at 40 °C.	https://www.sciencedirect.com/science/article/pii/0006300260914530?via%3Dihub
A protein sequenator	Pehr Edman and Geoffrey Begg	1967	5% v/v PITC in heptane was added after prior incubation with basic solvent (Quadrol). Incubated at 55°C for 30 minutes.	Anydrous HFBA was repeatedly added at 50°C.	https://febs.onlinelibrary.wiley.com/doi/full/10.1111/j.1432-1033.1967.tb00047.x
Solid-Phase Edman Degradation. An Automatic Peptide Sequencer.	Richard A. Laursen	1971	1.7 mL of buffer (3:2 mixture of pyridine and a N-methylmorpholiniumtrifluoroacetate buffer (pH 8.1)) and 0.7 mL 20% PITC in acetonitrile incubation for 20 minutes at 45°C.	TFA incubation for 30 minutes at 45 °C.	https://febs.onlinelibrary.wiley.com/doi/abs/10.1111/j.1432-1033.1971.tb01366.x
A manual method of sequential edman degradation followed by dansylation for the determination of protein sequences.	Maire E. Percy and Barbara M. Buchwald	1972	500 µL of 50% pyridine and 63 µL of N-ethylmorpholine prior incubation before 25 µL PITC addition and bubbling with nitrogen for 1 minute.. Incubated for 1 hour at 37°C or 45°C.	TFA incubation for 20 minutes at 37°C.	https://www.sciencedirect.com/science/article/pii/0003269772900073
Automatic solid-phase Edman degradation	Richard A. Laursen	1972	0.7 mL 5% PITC in acetonitrile in 1.7 mL of 3:2 pyridine and N-methylmorpholinium trifluoroacetate buffer (pH 8.1) at 45°C for 25 min.	TFA incubation for 30 minutes at 45°C.	https://www.sciencedirect.com/science/article/pii/S0076687972503037?via%3Dihub
Rapid manual sequencing of multiple peptide samples in a nitrogen chamber.	Richard B. Meagher	1975	7.5% v/v PITC in 50% pyridine with 10^{-5} M dithiothreitol (DTT) for 60-90 minutes at 45°C.	TFA incubation at 45°C for 15 minutes.	https://www.sciencedirect.com/science/article/pii/0003269775903127?via%3Dihub
A general procedure for the manual sequencing of small quantities of peptides.	George E. Tarr	1975	30% pyridine, 25% aqueous trimethylamine (0.1% in ethanthiol), 10% PITC in pyridine added at 1:1:4 ratio v/v for 25-30 minutes at 50 °C.	TFA purging for 6 minutes at 50 °C.	https://www.sciencedirect.com/science/article/pii/0003269775903589?via%3Dihub
Improved manual sequential analysis of peptides.	Mark Boehnert and David H. Schlesinger	1979	Under nitrogen, 0.4M triethylamine in propanol:water (3:2) solvent previously adjusted to pH 9.5 with TFA added prior to PITC. Performed at 54 °C for 30 minutes.	Concentrated HCl at 22°C for 5 minutes.	https://www.sciencedirect.com/science/article/pii/0003269779006080
A gas-liquid solid phase peptide and protein sequenator.	Hewick et al.	1981	15% PITC in n-heptane at 42°C for ~ 8 minutes twice.	TFA at 42°C and 0.01% dithiothreitol (DTT) twice once for ~ 5 minutes, another ~ 7 minutes.	https://www.sciencedirect.com/science/article/pii/S0021925818433777
Manual Edman Degradation Peptides. Methods in Molecular Biology. Humana Press	Per Klemm	1984	2.5% PITC 50% pyridine for 20 minutes at 50°C.	TFA incubation for 10 minutes at 45 °C.	https://link.springer.com/protocol/10.1385/0-89603-062-8:243
A manual sequencing method for identification of phosphorylated amino acids in phosphopeptides.	Sean Sullivan and Tai Wai Wong	1991	Methanol:water:triethylamine:PITC mix (7:1:1:1 v/v) incubation for 10 minutes at 50 °C.	TFA incubation at 50 °C for 6 minutes.	https://www.sciencedirect.com/science/article/pii/000326979190356X
Semi-automatic amino acid sequencing and D/L-configuration determination of peptides with detection of liberated *N*-terminal phenylthiocarbamoylamino acids	Lida et al.	1998	5% v/v PITC in heptane, and 12.5% m/v trimethylamine.	10-200 mol dm^{-3} boron trifluoride etherate in acetonitrile at 48°C for 5 minutes.	https://pubs.rsc.org/en/content/articlelanding/1998/an/a806109b
Proton: a major factor for the racemization and the dehydration at the cyclization/cleavage stage in the Edman sequencing method	Matsunaga et al.	1996	20 mM 7-[(N,N-dimethylamino)sulfonyl]-4-(2,1,3-benzoxadiazolyl) isothiocyanate (i.e., DBD-NCS) in 50% pyridine in water with 10 µL 10-100 µM dipeptide, mixed and heated at 50°C for 15 minutes.	1% v/v boron trifluoride and 0.1% v/v ethanethiol in dichloroethane at 50°C for 5 minutes.	https://pubs.acs.org/doi/10.1021/ac951253r

Supplementary Table 3

Table documenting the name of compounds, exact mass, and chemical formula. These values are used for searching in the Agilent MassHunter Software for automated AUC extraction of species from the LC/QToF data. Greyed out characters are not counted in formula and exact mass, and are fragments eliminated after trypsinization from the gel. Curly bracket after amino acid is to specify modifications that have occurred to that specific side chain (e.g., modified with acryloyl functional group "acr", specific post-translational modification from oxidation, such as P{hydroxyproline} etc.). Post-translational oxidation modifications for each amino acid are described in more detail in **Supplementary Figure 8** (abbreviations: MetO is methionine sulfoxide, $MetO_2$ is methionine sulfone, DOPA is 3,4-dihydroxyphenlylanine, Oia is oxindolylalanine, NFK is N-formylkynurenine, KN is kynurenin).

Peptide species	Monoisotopic mass	Formula
Peptide for LC/QToF trypsinization assay ("A15-peptide") with PITC		
AGGAGGLLGGSRGGK{acr}	971.5148	C39H69N15O14
PTC-AGGAGGLLGGSRGGK{acr}	1106.5291	C46H74N16O14S
GGAGGLLGGSRGGK{acr}	900.4777	C36H64N14O13
PTC-GGAGGLLGGSRGGK{acr}	1035.4920	C43H69N15O13S
GAGGLLGGSRGGK{acr}	843.4563	C34H61N13O12
PTC-AGGLLGGSRGGK{acr}	978.4705	C41H66N14O12S
GGLLGGSRGGK{acr}	786.4348	C32H58N12O11
Peptide for LC/QToF trypsinization assay ("A15-peptide") with ClickP		
ClickP-AGGAGGLLGGSRGGK{acr}	1175.5618	C48H77N19O14S
Peptide for SPAAC fluorescence assay ("K{N_3}15-peptide")		
K{N3}GGAGGLLGGSRGGK{acr}	1350.7116	C52H92N22O17
Peptide for LC/QToF trypsinization assay ("A15-peptide") with FITC		
FTC-AGGAGGLLGGSRGGK{acr}	1360.5506	C60H80N16O19S
Peptide for ionization efficiency standard ("A9-peptide")		
AGGAGK{acr}GLR	839.4613	C35H61N13O11
C-terminal, after arginine, of the N-terminal peptides		
XaaGGAGRGLGK{acr}	427.2431	C19H33N5O6
Methionine peptide ("M-peptide") and its oxidation products		
MGGAGRGLGK{acr}	547.2537	C20H37N9O7S
M{MetO}GGAGRGLGK{acr}	563.2486	C20H37N9O8S
M{$MetO_2$}GGAGRGLGK{acr}	579.2435	C20H37N9O9S
MGGAGRGLGK{acr}	956.4862	C39H68N14O12S
Cysteine peptide ("C-peptide") and its oxidation products		
CGGAGRGLGK{acr}	519.2224	C18H33N9O7S
K{acr}GLGRGAGGC{Cystine}GGAGRGLGK{acr}	1036.4291	C36H64N18O14S2
C{Sulfenic acid}GGAGRGLGK{acr}	535.2173	C18H33N9O8S
C{Sulfinate}GGAGRGLGK{acr}	551.2122	C18H33N9O9S
C{Sulfonate}GGAGRGLGK{acr}	567.2071	C18H33N9O10S
CGGAGRGLGK{acr}	928.4549	C37H64N14O12S
Tyrosine peptide ("Y-peptide") and its oxidation products		
YGGAGRGLGK{acr}	579.2765	C24H37N9O8
Y{DOPA}GGAGRGLGK{acr}	595.2714	C24H37N9O9
K{acr}GLGRGAGGY{dityrosine}YGGAGRGLGK{acr}	1156.5374	C48H72N18O16

YGGAGRGLGK{acr}	988.509	C43H68N14O13
Phenylalanine peptide ("F_1" peptide) and its oxidation products		
FGGAGRGLGK{acr}	563.2816	C24H37N9O7
F{Meta-tyrosine}GGAGRGLGK{acr}	579.2765	C24H37N9O8
F{DOPA}GGAGRGLGK{acr}	595.2714	C24H37N9O9
FGGAGRGLGK{acr}	972.5141	C43H68N14O12
Tryptophan peptide ("W-peptide") and its oxidation products		
WGGAGRGLGK{acr}	602.2925	C26H38N10O7
W{Oia}GGAGRGLGK{acr}	618.2874	C26H38N10O8
W{NFK}GGAGRGLGK{acr}	634.2823	C26H38N10O9
W{KN}GGAGRGLGK{acr}	606.2874	C25H38N10O8
WGGAGRGLGK{acr}	1011.525	C45H69N15O12
Histidine peptide ("H-peptide") and its oxidation products		
HGGAGRGLGK{acr}	553.2721	C21H35N11O7
H{2-oxo-histidine}GGAGRGLGK{acr}	567.2514	C21H33N11O8
H{aspartate}GGAGRGLGK{acr}	531.2401	C19H33N9O9
H{formylasparagine}GGAGRGLGK{acr}	558.251	C20H34N10O9
H{aspartylurea}GGAGRGLGK{acr}	543.2401	C20H33N9O9
HGGAGRGLGK{acr}	962.5046	C40H66N16O12
Proline peptide ("P-peptide") and its oxidation products		
PGGAGRGLGK{acr}	513.2659	C20H35N9O7
P{pyrroline-5-carboxylate} GGAGRGLGK{acr}	511.2503	C20H33N9O7
P{hydroxyproline}GGAGRGLGK{acr}	529.2609	C20H35N9O8
P{glutamic semialdehyde} GGAGRGLGK{acr}	527.2816	C21H37N9O7
PGGAGRGLGK{acr}	922.4985	C39H66N14O12
Arginine peptide ("R-peptide") and its oxidation products		
RGGAGRGLGK{acr}	572.3143	C21H40N12O7
R{hydroxyarginine} GGAGRGLGK{acr}	588.3092	C21H40N12O8
R{oxoarginine}GGAGRGLGK{acr}	573.2983	C21H39N11O8
R{glutamic semialdehyde} GGAGRGLGK{acr}	527.2816	C21H37N9O7
RGGAGRGLGK{acr}	981.5468	C40H71N17O12
PTH-amino acids		
PTH-phenylalanine	282.0827	C16H14N2OS
PTH-alanine	206.0514	C10H10N2OS
PTH-tyrosine	298.0776	C16H14N2O2S
PTH-glycine	192.0357	C9H8N2OS

Supplementary Table 4

Tables documenting the results for all ExM and ExMre gels placed in various solvents plotted in **Fig. 2ci-vii**. (i) Surface size (flat/top side of gel) of ExM and ExM re-embedded (ExMre) gels when placed in solvents used in Edman degradation, namely 1:1 pyridine to water, acetonitrile (ACN), dimethylsulfoxide (DMSO), formamide, 0.1 M sodium bicarbonate ($NaHCO_3$) pH 8.5, Trifluoroacetic acid (TFA), with and without Edman reagents (PITC, 1:1000; FITC, 5.9 mM in 23:77 DMSO:0.1 M sodium bicarbonate pH 8.5). (ii) Normalized values from (i) to flat/top surface size in 1X PBS. (iii) Mean percent shrinkage going from 1X PBS to various solvents based on values obtained in (ii).

Note: Some columns contain what appear to be repeated values - this is not an error. Instead, this often occurred because of the way we made and trimmed the gels, to a similar size of approximately ~0.5 x 0.5 cm and then measured exactly when washed 3x with 1X PBS 10 min at room temperature. Due to the limited values of length and width that could be measured with our ruler (gel length and width sizes between ~0.1 - 0.8 cm), we were only able to measure a small range of length and width values. This also limited the output values of the area when these two values were multiplied. As a result, since our gels had similar starting sizes prior to incubation in solvent or solvent or buffer with PITC or FITC, many gels ended up with similar or identical measured areas after treatment.

(i) Surface size of the flat/top side of ExM and ExMre gels, made from 3 separate gelation solutions, in various solvents in with/without PITC or FITC, except for pyridine which is 3 replicates from the same gelation solution.

	Area of the gel (cm^2)					
Gelation number	**ExM 1X PBS**	**ExMre 1X PBS**	**ExM pyridine**	**ExMre pyridine**	**ExM pyridine with PITC**	**ExMre pyridine with PITC**
1 (replicate 1)	0.300	0.250	0.030	0.040	0.020	0.040
1 (replicate 2)	0.200	0.300	0.020	0.090	0.020	0.090
1 (replicate 3)	0.275	0.250	0.020	0.090	0.020	0.063
Gelation number	**ExM 1X PBS**	**ExMre 1X PBS**	**ExM 1:1 pyridine to water**	**ExMre 1:1 pyridine to water**	**ExM 1:1 pyridine to water with PITC**	**ExMre 1:1 pyridine to water with PITC**
1	0.275	0.250	0.040	0.075	0.040	0.040
2	0.330	0.250	0.050	0.075	0.040	0.040
3	0.300	0.250	0.040	0.063	0.040	0.040
Gelation number	**ExM 1X PBS**	**ExMre 1X PBS**	**ExM ACN**	**ExMre ACN**	**ExM ACN with PITC**	**ExMre ACN with PITC**
1	0.330	0.250	0.075	0.063	0.040	0.040
2	0.275	0.250	0.060	0.050	0.040	0.040
3	0.250	0.250	0.040	0.050	0.040	0.040
Gelation number	**ExM 1X PBS**	**ExMre 1X PBS**	**ExM DMSO**	**ExMre DMSO**	**ExM DMSO with PITC**	**ExMre DMSO with PITC**
1	0.330	0.250	0.060	0.123	0.090	0.123
2	0.275	0.250	0.075	0.123	0.075	0.123
3	0.275	0.250	0.075	0.123	0.075	0.123
Gelation number	**ExM 1X PBS**	**ExMre 1X PBS**	**ExM formamide**	**ExMre formamide**	**ExM formamide with PITC**	**ExMre formamide with PITC**
1	0.300	0.225	0.560	0.275	0.720	0.330
2	0.275	0.225	0.560	0.300	0.720	0.325
3	0.250	0.250	0.560	0.303	0.720	0.360
Gelation	**ExMre 1X PBS**	**ExMre 0.1 M**	**ExMre 5.9 mM**			

number		NaHCO3 pH 8.5	FITC in 23:77 DMSO:0.1 M NaHCO3 pH 8.5			
1	0.225	0.275	0.225			
2	0.225	0.275	0.225			
3	0.225	0.275	0.225			
Gelation number	**ExMre 1X PBS**	**ExMre TFA**				
1	0.250	0.250				
2	0.250	0.250				
3	0.250	0.250				

(ii) Normalized values to top/flat surface size in 1X PBS.

	ExM top/flat surface size normalized to 1X PBS		ExMre top/flat surface size normalized to 1X PBS	
Gelation number	**pyridine**	**pyridine with PITC**	**pyridine**	**pyridine with PITC**
1(replicate 1)	0.100	0.067	0.160	0.160
1(replicate 2)	0.100	0.100	0.300	0.300
1(replicate 3)	0.072	0.072	0.360	0.250
Stdev	*0.016*	*0.018*	*0.103*	*0.071*
	ExM top/flat surface size normalized to 1X PBS		ExMre top/flat surface size normalized to 1X PBS	
Gelation number	**1:1 pyridine to water**	**1:1 pyridine to water with PITC**	**1:1 pyridine to water**	**1:1 pyridine to water with PITC**
1	0.145	0.145	0.300	0.160
2	0.152	0.121	0.300	0.160
3	0.133	0.133	0.250	0.160
Stdev	*0.009*	*0.012*	*0.029*	*0.000*
	ExM top/flat surface size normalized to 1X PBS		ExMre top/flat surface size normalized to 1X PBS	
Gelation number	**ACN**	**ACN with PITC**	**ACN**	**ACN with PITC**
1	0.227	0.121	0.250	0.160
2	0.218	0.145	0.200	0.160
3	0.160	0.160	0.200	0.160
Stdev	*0.036*	*0.020*	*0.029*	*0.000*
	ExM top/flat surface size normalized to 1X PBS		ExMre top/flat surface size normalized to 1X PBS	
Gelation number	**DMSO**	**DMSO with PITC**	**DMSO**	**DMSO with PITC**
1	0.182	0.273	0.490	0.490
2	0.273	0.273	0.490	0.490
3	0.273	0.273	0.490	0.490
Stdev	*0.052*	*0.000*	*0.000*	*0.000*
	ExM top/flat surface size normalized to 1X PBS		ExMre top/flat surface size normalized to 1X PBS	
Gelation number	**Formamide**	**Formamide with PITC**	**Formamide**	**Formamide with PITC**
1	1.867	2.400	1.222	1.467
2	2.036	2.618	1.333	1.444
3	2.240	2.880	1.210	1.440
Stdev	*0.187*	*0.240*	*0.068*	*0.014*
	ExM top/flat surface size normalized to 1X PBS			
Gelation number	**0.1 M NaHCO3 pH 8.5**	**5.9 mM FITC in DMSO:0.1 M NaHCO3 pH 8.5**		
1	1.222	1.000		

2	1.222	1.000	
3	1.222	1.000	
Stdev	*0.000*	*0.000*	
	ExMre top/flat surface size normalized to 1X PBS		
Gelation number	**TFA**		
1	1.000		
2	1.000		
3	1.000		
Stdev	*0.000*		

(iii) Percent shrinkage of ExM and ExMre gels when placed in different solvents with/without PITC or FITC.

pyridine	**% shrinkage from a to b (a vs. b)**
ExM in 1X PBS vs. pyridine without PITC	90.9
ExM in 1X PBS vs. pyridine ExM with PITC (1:9 ratio PITC:pyridine)	92.0
ExM in pyridine without PITC vs. ExM in pyridine with PITC (1:9 ratio PITC:pyridine)	1.1
ExMre in 1X PBS vs. pyridine without PITC	72.7
ExMre in 1X PBS vs. pyridine with PITC (1:9 ratio PITC:pyridine)	76.3
ExMre in pyridine without PITC vs. ExMre in pyridine with PITC (1:9 ratio PITC:pyridine)	3.7
1:1 pyridine to water	**% shrinkage from a to b (a vs. b)**
ExM in 1X PBS vs. 1:1 pyridine to water without PITC	85.7
ExM in 1X PBS vs. 1:1 pyridine to water ExM with PITC (1:1000 ratio PITC to 1:1 pyridine to water)	86.7
ExM in 1:1 pyridine to water without PITC vs. ExM in 1:1 pyridine to water with PITC (1:1000 ratio PITC to 1:1 pyridine to water)	1.0
ExMre in 1X PBS vs. 1:1 pyridine to water without PITC	71.7
ExMre in 1X PBS vs. 1:1 pyridine to water with PITC (1:1000 ratio PITC to 1:1 pyridine to water)	84.0
ExMre in 1:1 pyridine to water without PITC vs. ExMre in 1:1 pyridine to water with PITC (1:1000 ratio PITC to 1:1 pyridine to water)	12.3
Acetonitrile	**% shrinkage from a to b (a vs. b)**
ExM in 1X PBS vs. acetonitrile without PITC	79.8
ExM in 1X PBS vs. acetonitrile ExM with PITC (1:1000 ratio PITC:acetonitrile)	85.8
ExM in acetonitrile without PITC vs. ExM in acetonitrile with PITC (1:1000 ratio PITC:acetonitrile)	6.0
ExMre in 1X PBS vs. acetonitrile without PITC	78.3
ExMre in 1X PBS vs. acetonitrile with PITC (1:1000 ratio PITC:acetonitrile)	84.0
ExMre in acetonitrile without PITC vs. ExMre in acetonitrile with PITC (1:1000 ratio PITC:acetonitrile)	5.7
DMSO	**% shrinkage from a to b (a vs. b)**
ExM in 1X PBS vs. DMSO without PITC	75.8
ExM in 1X PBS vs. DMSO ExM with PITC (1:1000 ratio PITC:DMSO)	72.7
ExM in DMSO without PITC vs. ExM in DMSO with PITC	-3.0

(1:1000 ratio PITC:DMSO)	
ExMre in 1X PBS vs. DMSO without PITC	51.0
ExMre in 1X PBS vs. DMSO with PITC (1:1000 ratio PITC:DMSO)	51.0
ExMre in DMSO without PITC vs. ExMre in DMSO with PITC (1:1000 ratio PITC:DMSO)	0.0
Formamide	**% shrinkage from a to b (a vs. b)**
ExM in 1X PBS vs. formamide without PITC	-104.8
ExM in 1X PBS vs. formamide ExM with PITC (1:1000 ratio PITC:formamide)	-163.3
ExM in formamide without PITC vs. ExM in formamide with PITC (1:1000 ratio PITC:formamide)	-58.5
ExMre in 1X PBS vs. formamide without PITC	-25.5
ExMre in 1X PBS vs. formamide with PITC (1:1000 ratio PITC:formamide)	-45.0
ExMre in formamide without PITC vs. ExMre in formamide with PITC (1:1000 ratio PITC:formamide)	-19.5
0.1 M NaHCO3 pH 8.5	**% shrinkage from a to b (a vs. b)**
ExMre in 1X PBS vs. 0.1 M NaHCO3 pH 8.5 without FITC	-22.2
ExMre in 1X PBS vs. 0.1 M NaHCO3 pH 8.5 with FITC in DMSO (23:77 ratio DMSO:0.1 M sodium bicarbonate buffer pH 8.5)	0.0
ExMre in 0.1 M NaHCO3 pH 8.5 without FITC vs. ExMre in 0.1 M NaHCO3 pH 8.5 with FITC in DMSO (23:77 ratio DMSO:0.1 M sodium bicarbonate buffer pH 8.5)	22.2
TFA	**% shrinkage from a to b (a vs. b)**
ExMre in 1X PBS vs. in TFA	0.0

Supplementary Table 5

Values and analysis for the LC/QToF trypsinization assay from Fig. 3 and strain-promoted alkyne-azide cycloaddition (SPAAC) chemistry click chemistry fluorescence assay.

(A) Formamide with PITC

(i) Raw data values for the area under the curve (AUC) of the chromatogram of various species for PITC (1:1000 ratio PITC:formamide) conjugation(**Fig. 3b**) (extracted based on the exact mass, see **Methods** for more details, and **Source Data** for raw traces). Three tables for the three separate gelation solutions. The cells highlighted in grey are the ones used to calculate the conversion rate (solvent with PITC["A15-peptide"] / solvent ["A15-peptide"]). A15-peptide stands for AGGAGGLLGGSRGGK{acr}. Cells with '-' means that the species was not detected in that sample using the automatic extraction.

Conjugation with PITC to formamide (1:1000 ratio PITC:formamide)	Gelation	Formamide	Formamide with PITC	Formamide, then TFA	Formamide with PITC, then TFA	Conversion rate
A15-peptide	Gelation 1	12,117,368	394,015	15,178,608	680,108	96.75
PITC conjugated A15-peptide		-	7,493,956	-	-	
A15-peptide with cleaved N-terminal amino acid		54,993	6,919,759	87,591	14,344,037	
A9-peptide (control)		1,638,492	1,583,580	1,423,334	1,583,069	

Conjugation with PITC to formamide (1:1000 ratio PITC:formamide)	Gelation	Formamide	Formamide with PITC	Formamide, then TFA	Formamide with PITC, then TFA	Conversion rate
A15-peptide	Gelation 2	14,191,286	567,866	16,829,402	1,384,801	96.00
PITC conjugated A15-peptide		-	9,085,305	-	-	
A15-peptide with cleaved N-terminal amino acid		69,133	11,251,107	100,469	18,304,827	
A9-peptide (control)		1,610,108	1,518,859	1,642,476	1,658,903	

Conjugation with PITC to formamide (1:1000 ratio PITC:formamide)	Gelation	Formamide	Formamide with PITC	Formamide, then TFA	Formamide with PITC, then TFA	Conversion rate
A15-peptide	Gelation 3	11,261,653	472,423	12,264,577	555,385	95.81
PITC conjugated A15-peptide		-	6999122	-	-	
A15-peptide with cleaved N-terminal amino acid		50,898	8,048,591	77,366	11,057,572	
A9-peptide (control)		1,950,392	1,713,521	1,750,512	1,927,406	

(ii) Conversion rate of A15-peptide to PITC conjugated A15-peptide, and % yield for conjugation with PITC to formamide (1:1000 ratio PITC:formamide) (**Fig. 3b**).

Conjugation with PITC to formamide (1:1000 ratio PITC:formamide) **Average % yield (peptide with cleaved N-terminal amino acid / non-modified peptide * 100)**	98.72
Average conversion rate (%)	96.18

(iii) Statistics: One-way Analysis of Variance (ANOVA) and Tukey's post-hoc Honestly Significant Difference (HSD) test on the abundance of peptide with cleaved N-terminal amino acid using PITC to formamide (1:1000 ratio PITC:formamide) for conjugation (**Fig. 3biii**). In Tukey HSD results, conditions are abbreviated, A: formamide, B: formamide with PITC, C: formamide then TFA, D: formamide with PITC then TFA.

ANOVA Results					
Test	Comparison	sum_sq	df	F	PR(>F)
ANOVA	group	4.53E+14	3.00E+00	3.32E+01	7.31E-05
ANOVA	Residual	3.64E+13	8.00E+00		

Tukey HSD Results							
Test	group1	group2	meandiff	p-adj	lower	upper	reject
Tukey HSD	A	B	8.68E+06	4.73E-03	3.10E+06	1.43E+07	TRUE
Tukey HSD	A	C	3.01E+04	1.00E+00	-5.55E+06	5.61E+06	FALSE
Tukey HSD	A	D	1.45E+07	1.51E-04	8.93E+06	2.01E+07	TRUE
Tukey HSD	B	C	-8.65E+06	4.83E-03	-1.42E+07	-3.07E+06	TRUE
Tukey HSD	B	D	5.83E+06	4.09E-02	2.49E+05	1.14E+07	TRUE
Tukey HSD	C	D	1.45E+07	1.53E-04	8.90E+06	2.01E+07	TRUE

(iv) Gel size changes throughout in-gel Edman degradation with PITC to formamide (1:1000 ratio PITC:formamide) (**Fig. 3ei**).

Conjugation with PITC to formamide (1:1000 PITC:formamide) - Gel size changes in Edman steps	**Gelation 1, area (cm^2)**	**Gelation 2, area (cm^2)**	**Gelation 3, area (cm^2)**	**Mean (normalized to surface area in 1M Tris pH 9.5) (cm^2)**	**Standard deviation (cm^2)**
1. 3x wash with Tris 1M pH 9.5	0.330	0.303	0.250	1.000	0.000
2. 2x formamide wash	0.300	0.360	0.250	1.031	0.143
3. PITC conjugation	0.360	0.420	0.300	1.224	0.151
4. TFA	0.330	0.360	0.330	1.156	0.161
5. 2x formamide wash	0.360	0.360	0.300	1.156	0.060
6. 3x wash with Tris 1M pH 8	0.330	0.303	0.250	1.000	0.000

(B) DMSO with PITC

(i) Raw data values for the area under the curve (AUC) of the chromatogram of various species for PITC to DMSO (1:1000 ratio PITC:DMSO) for conjugation (**Fig. 3c**) (extracted based on the exact mass, see **Methods** for more details, and **Source Data** for raw traces). Three tables for the three separate gelation solutions. The cells highlighted in grey are the ones used to calculate the conversion rate (solvent with PITC["A15-peptide"] / solvent ["A15-peptide"]). A15-peptide stands for AGGAGGLLGGSRGGK{acr}. Cells with '-' means that the species was not detected in that sample using the automatic extraction.

Conjugation with PITC to DMSO (1:1000 ratio PITC:DMSO)	**Gelation**	**DMSO**	**DMSO with PITC**	**DMSO, then TFA**	**DMSO with PITC, then TFA**	**Conversion rate**

A15-peptide	Gelation 1	15,459,666	276,139	15,338,350	323,062	98.21
PITC conjugated A15-peptide			10,904,256		23,333	
A15-peptide with cleaved N-terminal amino acid		84,704	7,448,340	103,752	14,365,449	
A9-peptide (control)		1,835,948	1,773,090	1,663,375	2,047,203	

Conjugation with PITC to DMSO (1:1000 ratio PITC:DMSO)	**Gelation**	**DMSO**	**DMSO with PITC**	**DMSO, then TFA**	**DMSO with PITC, then TFA**	**Conversion rate**
A15-peptide	Gelation 2	18,095,571	230,579	17,231,541	344,879	98.73
PITC conjugated A15-peptide			8,054,863		29,359	
A15-peptide with cleaved N-terminal amino acid		106,966	7,135,934	89,852	15,418,384	
A9-peptide (control)		2,155,301	1,298,155	1,840,897	2,023,220	

Conjugation with PITC to DMSO (1:1000 ratio PITC:DMSO)	**Gelation**	**DMSO**	**DMSO with PITC**	**DMSO, then TFA**	**DMSO with PITC, then TFA**	**Conversion rate**
A15-peptide	Gelation 3	14,680,209	118,927	14,148,945	329,337	99.19
PITC conjugated A15-peptide			8,729,040		25,109	
A15-peptide with cleaved N-terminal amino acid		87,737	4,869,072	77,527	14,377,025	
A9-peptide (control)		1,763,566	1,950,219	1,915,314	2,201,666	

(ii) Conversion rate of A15-peptide to PITC conjugated A15-peptide, and % yield for PITC (1:1000 ratio PITC:DMSO) conjugation (**Fig. 3c**).

Conjugation with PITC to DMSO (1:1000 ratio PITC:DMSO)
Average % yield (peptide with cleaved N-terminal amino acid / non-modified peptide * 100) 94.52
Average conversion rate (%) 98.71

(iii) Statistics: One-way Analysis of Variance (ANOVA) and Tukey's post-hoc Honestly Significant Difference (HSD) test on the abundance of peptide with cleaved N-terminal amino acid using PITC to DMSO (1:1000 ratio PITC:DMSO) for conjugation (**Fig. 3ciii**). In Tukey HSD results, conditions are abbreviated, A: DMSO, B: DMSO with PITC, C: DMSO then TFA, D: DMSO with PITC then TFA.

ANOVA Results					
Test	Comparison	sum_sq	df	F	PR(>F)
ANOVA	group	4.33E+14	3.00E+00	2.46E+02	3.24E-08

ANOVA	Residual	4.70E+12	8.00E+00		

Tukey HSD Results							
Test	group1	group2	meandiff	p-adj	lower	upper	reject
Tukey HSD	A	B	6.39E+06	3.37E-05	4.39E+06	8.39E+06	TRUE
Tukey HSD	A	C	-2.76E+03	1.00E+00	-2.01E+06	2.00E+06	FALSE
Tukey HSD	A	D	1.46E+07	5.64E-08	1.26E+07	1.66E+07	TRUE
Tukey HSD	B	C	-6.39E+06	3.36E-05	-8.40E+06	-4.39E+06	TRUE
Tukey HSD	B	D	8.24E+06	4.96E-06	6.23E+06	1.02E+07	TRUE
Tukey HSD	C	D	1.46E+07	5.63E-08	1.26E+07	1.66E+07	TRUE

(iv) Gel size changes throughout in-gel Edman degradation with PITC (1:1000 ratio PITC:DMSO) conjugation (**Fig. 3eii**).

Conjugation with PITC to DMSO (1:1000 ratio PITC:DMSO) - Gel size changes in Edman steps	**Gelation 1, area (cm^2)**	**Gelation 2, area (cm^2)**	**Gelation 3, area (cm^2)**	**Mean (normalized to surface are in 1M Tris pH 9.5) (cm^2)**	**Standard deviation (cm^2)**
1. 3x wash with Tris 1M pH 9.5	0.300	0.300	0.275	1.000	0.000
2. 2x DMSO wash	0.300	0.300	0.275	1.000	0.000
3. PITC conjugation	0.300	0.275	0.275	0.971	0.048
4. TFA	0.225	0.275	0.225	0.829	0.084
5. 2x DMSO wash	0.180	0.225	0.180	0.669	0.076
6. 3x wash with Tris 1M pH 8	0.300	0.300	0.275	1.000	0.000

(C) $NaHCO_3$ with FITC

(i) Raw data values for the area under the curve (AUC) of the chromatogram of various species for 5.9 mM FITC conjugation in 23:77 DMSO:0.1 M sodium bicarbonate ($NaHCO_3$) pH 8.5 (**Fig. 3d**) (extracted based on the exact mass, see **Methods** for more details, and **Source Data** for raw traces). Three tables for the three separate gelation solutions. The cells highlighted in grey are the ones used to calculate the conversion rate (solvent with FITC["A15-peptide"] / solvent ["A15-peptide"]). A15-peptide stands for AGGAGGLLGGSRGGK{acr}. Cells with '-' means that the species was not detected in that sample using the automatic extraction.

Conjugation with FITC in 0.1 M $NaHCO_3$ pH 8.5 (5.9 mM FITC in 23:77 ratio DMSO:0.1 M $NaHCO_3$ pH 8.5)	**Gelation**	**0.1 M $NaHCO_3$**	**23:77 DMSO:0.1 M $NaHCO_3$ with FITC**	**0.1 M $NaHCO_3$, then TFA**	**23:77 DMSO:0.1 M $NaHCO_3$ with FITC, then TFA**	**Conversion rate**
A15-peptide		11,073,221	2,942,798	11,874,162	2,485,449	
FITC conjugated A15-peptide		-	495304	-	-	73.42
A15-peptide with cleaved N-terminal amino acid		130,625	5,100,100	131,053	8,326,486	
A9-peptide (control)	Gelation 1	1,192,016	1,113,044	1,246,017	1,198,247	

Conjugation with FITC in 0.1 M $NaHCO_3$ pH 8.5 (5.9 mM FITC in 23:77 ratio DMSO:0.1 M $NaHCO_3$ pH 8.5)	**Gelation**	**0.1 M $NaHCO_3$**	**23:77 DMSO:0.1 M $NaHCO_3$ with FITC**	**0.1 M $NaHCO_3$, then TFA**	**23:77 DMSO:0.1 M $NaHCO_3$ with FITC, then TFA**	**Conversion rate**
A15-peptide	Gelation 2	15,174,962	3,649,573	13,304,213	3,063,850	75.95

FITC conjugated A15-peptide		-	638249	-	-	
A15-peptide with cleaved N-terminal amino acid		141,670	7,944,263	137,796	9,897,684	
A9-peptide (control)		1,089,661	976,075	1,146,393	1,124,571	

Conjugation with FITC in 0.1 M $NaHCO_3$ pH 8.5 (5.9 mM FITC in 23:77 ratio DMSO:0.1 M $NaHCO_3$ pH 8.5)	**Gelation**	**0.1 M $NaHCO_3$**	**23:77 DMSO:0.1 M $NaHCO_3$ with FITC**	**0.1 M $NaHCO_3$, then TFA**	**23:77 DMSO:0.1 M $NaHCO_3$ with FITC, then TFA**	**Conversion rate**
A15-peptide		11,107,184	2,580,088	11,099,956	2,328,726	76.77
FITC conjugated A15-peptide		-	443245	-	-	
A15-peptide with cleaved N-terminal amino acid		98,419	5,065,557	126,352	7,022,611	
A9-peptide (control)	Gelation 3	1,142,427	1,047,343	1,180,805	1,163,367	

(ii) Conversion rate of A15-peptide to PITC conjugated A15-peptide, and % yield for 5.9 mM FITC conjugation in 23:77 DMSO:0.1 M sodium bicarbonate ($NaHCO_3$) pH 8.5 (**Fig. 3d**).

Conjugation with FITC in 0.1 M $NaHCO_3$ pH 8.5 (5.9 mM FITC in 23:77 ratio DMSO:0.1 M $NaHCO_3$ pH 8.5) **Average % yield (peptide with cleaved N-terminal amino acid / non-modified peptide * 100)**	69.59
Average conversion rate (%)	75.38

(iii) Statistics: One-way Analysis of Variance (ANOVA) and Tukey's post-hoc Honestly Significant Difference (HSD) test on the abundance of peptide with cleaved N-terminal amino acid using 5.9 mM FITC conjugation in 23:77 DMSO:0.1 M sodium bicarbonate ($NaHCO_3$) pH 8.5 (**Fig. 3diii**). In Tukey HSD results, conditions are abbreviated, A: $NaHCO_3$, B: $NaHCO_3$ with FITC, C: $NaHCO_3$ then TFA, D: $NaHCO_3$ with FITC then TFA.

ANOVA Results					
Test	Comparison	sum_sq	df	F	PR(>F)
ANOVA	group	1.60E+14	3.00E+00	4.43E+01	2.49E-05
ANOVA	Residual	9.61E+12	8.00E+00		

Tukey HSD Results							
Test	group1	group2	meandiff	p-adj	lower	upper	reject
Tukey HSD	A	B	5.91E+06	7.62E-04	3.05E+06	8.78E+06	TRUE
Tukey HSD	A	C	8.16E+03	1.00E+00	-2.86E+06	2.87E+06	FALSE
Tukey HSD	A	D	8.29E+06	6.93E-05	5.43E+06	1.12E+07	TRUE
Tukey HSD	B	C	-5.91E+06	7.69E-04	-8.77E+06	-3.04E+06	TRUE
Tukey HSD	B	D	2.38E+06	1.08E-01	-4.86E+05	5.24E+06	FALSE
Tukey HSD	C	D	8.28E+06	6.98E-05	5.42E+06	1.12E+07	TRUE

(iv) Gel size changes throughout in-gel Edman degradation with using 5.9 mM FITC conjugation in 23:77 DMSO:0.1 M sodium bicarbonate ($NaHCO_3$) pH 8.5 (**Fig. 3eiii**).

Conjugation with FITC in 23:77 DMSO:$NaHCO_3$ - Gel size changes in Edman steps	**Gelation 1, area (cm^2)**	**Gelation 2, area (cm^2)**	**Gelation 3, area (cm^2)**	**Mean (normalized to top/flat surface size in 1M Tris pH 9.5) (cm^2)**	**Standard deviation (cm^2)**
1. 3x wash with Tris 1M pH 9.5	0.303	0.360	0.303	1.000	0.000
2. 2x 0.1 M $NaHCO_3$wash	0.303	0.303	0.303	0.940	0.092
3. FITC conjugation	0.275	0.303	0.303	0.912	0.080
4. TFA	0.250	0.303	0.250	0.832	0.008
5. 2x 2x 0.1 M $NaHCO_3$ wash	0.303	0.303	0.250	0.886	0.096
6. 3x wash with Tris 1M pH 8	0.303	0.360	0.303	1.000	0.000

(D) Fluorescence assay with azidolysine peptide

(i) Raw average fluorescence intensity values for each gel, and average over the replicates.

Condition name and number	**Average intensity ExMre gel 1**	**Average intensity ExMre gel 2**	**Average intensity ExMre gel 3**	**Average intensity over replicates**	**Standard deviation over replicates**
(1) DMSO	25,068.05586	26,800.6509	22,135.4875	**24,668.06475**	2,358.16285
(2) PITC:DMSO	23,926.83914	22,876.96382	18,189.04312	**21,664.28203**	3,055.08161
(3) DMSO+TFA	24,913.99409	25,783.96689	22,445.86609	**24,381.27569**	1,731.63821
(4) PITC:DMSO+TFA	9,763.790698	9,629.696378	6,710.793017	**8,701.426698**	1,725.24264
(5) PITC:DMSO+TFA+trypsin	2,020.303932	1,890.965337	2,020.303932	**1,977.191067**	74.673673

(ii) % yield calculation based on the fluorescence azidolysine read-out. The "averageIntensities(x)" represents the values obtained in **Supplementary Table 5di** after averaging over the replicates of a given condition, where "x" represents the number associated with that condition.

% Yield
Formula: efficiencyAvg = (1 - ((averageIntensities(4) - averageIntensities(5))/(averageIntensities(1) - averageIntensities(5)))) * 100
70.3659025

(E) Phenylthiohydantoin-phenylalanine (PTH-F) detection

Analysis of PTH-F abundance was performed using PTH-F exact mass, 282.0827 ± 0.0056 Da (see **Methods: In-gel Edman degradation with PTH-aa detection** for details on extracting the abundance; see **Supplementary Figure 6a-b** for raw traces and mass spectra). The abundance extracted from the chromatograms are documented in the table below.

	Abundance (area) gel 1	**Abundance (area) gel 2**	**Abundance (area) gel 3**	**Average abundance over replicates**	**Standard deviation across replicates**
PTH-F negative control (DMSO+TFA)	787.54	1,196.55	235.46	739.85	482.31

PTH-F experiment (PITC:DMSO+TFA)	197,004.58	246,786.47	63,878.04	169,223.03	94,566.031
PTH-F positive control	-	-	-	35,445,968.85	-

Supplementary Table 6

Values and analysis for the LC/QToF trypsinization assay from Fig. 5**b**.

(i) Raw data values for the area under the curve (AUC) of the chromatogram of various species for PITCto DMSO (1:1000 ratio PITC:DMSO) for conjugation over multiple rounds (Fig. 5**b**) (extracted based on the exact mass, see Methods for more details, and Source Data for raw traces). Three tables for the three separate gelation solutions. The cells highlighted in grey are the ones used to calculate the conversion rate (solvent with PITC["A15-peptide"] / solvent ["A15-peptide"]). A15-peptide stands for AGGAGGLLGGSRGGK{acr}. Cells with '-' means that the species was not detected in that sample using the automatic extraction.

Gelation 1	**DMSO**	**DMSO with PITC**	**DMSO, then TFA**	**DMSO with PITC, then TFA**	**1 round, then DMSO with PITC**	**1 round, then DMSO with PITC, then TFA**	**2 rounds, then DMSO with PITC**	**2 rounds, then DMSO with PITC, then TFA**
A15-peptide	14,152,440	234,022	16,895,166	241,708	55,744	55,988	127,394	105,796
PITC conjugated A15-peptide	-	8,528,288	-	31,819	94,071	-	42,671	-
A15-peptide with cleaved N-terminal amino acid	76,571	7,159,311	101,167	11,042,843	110,995	302,012	69,181	103,014
PITC conjugated A15-peptide with cleaved N-terminal amino acid	-	354,635	-	23,086	11,642,582	605,347	937,425	-
A15-peptide with two cleaved N-terminal amino acids	73,330	21,890	62,400	187,758	624,556	6,975,236	214,828	611,031
PITC conjugated A15-peptide with two cleaved N-terminal amino acid	64,524	30,557	78,959	4,345	271,111	39,892	11,299,778	1,992,023
A15-peptide with three cleaved N-terminal amino acids	78,578	5,505	10,341	29,603	26,503	83,453	444,363	2,076,406
A9-peptide (control)	1,717,854	1,950,484	1,916,119	2,305,651	1,992,306	2,085,178	1,726,787	1,770,300

Gelation 2	**DMSO**	**DMSO with PITC**	**DMSO, then TFA**	**DMSO with PITC, then TFA**	**1 round, then DMSO with PITC**	**1 round, then DMSO with PITC, then TFA**	**2 rounds, then DMSO with PITC**	**2 rounds, then DMSO with PITC, then TFA**

A15-peptide	19,664,727	168,662	18,026,845	382,515	61,289	73,152	23,823	175,778
PITC conjugated A15-peptide	-	10,588,346	-	37,486	119,421	-	39,940	-
A15-peptide with cleaved N-terminal amino acid	114,996	9,038,666	101,342	16,031,614	186,135	323,530	67,534	159,065
PITC conjugated A15-peptide with cleaved N-terminal amino acid	-	464,449	-	28,326	18,012,585	679,072	939,101	-
A15-peptide with two cleaved N-terminal amino acids	97,726	74,181	72,097	270,900	973,920	7,341,481	225,736	952,458
PITC conjugated A15-peptide with two cleaved N-terminal amino acid	93,023	45,738	82,455	6,666	426,174	71,780	9,759,509	2,432,709
A15-peptide with three cleaved N-terminal amino acids	13,299	-	13,090	43,111	29,064	152,920	421,886	2,545,115
A9-peptide (control)	2,329,190	1,932,427	1,994,133	2,169,287	2,175,041	1,959,964	2,075,030	1,728,591

Gelation 3	**DMSO**	**DMSO with PITC**	**DMSO, then TFA**	**DMSO with PITC, then TFA**	**1 round, then DMSO with PITC**	**1 round, then DMSO with PITC, then TFA**	**2 rounds, then DMSO with PITC**	**2 rounds, then DMSO with PITC, then TFA**
A15-peptide	13,935,909	283,570	12,028,033	271,999	48,147	128,729	113,952	113,403
PITC conjugated A15-peptide	-	5,677,587	-	16,631	102,012	-	37,689	-
A15-peptide with cleaved N-terminal amino acid	77,417	6,630,107	72,205	11,187,479	141,646	344,558	82,374	114,200
PITC conjugated A15-peptide with cleaved N-terminal amino acid	-	318,844	-	18,965	14,851,035	786,992	1,140,599	-
A15-peptide with two cleaved N-terminal amino acids	68,099	20,399	44,486	207,266	816,241	6,151,386	49,806	561,106
PITC conjugated A15-peptide with two cleaved N-terminal amino acid	71,146	31,762	71,623	3,814	332,970	56,436	9,159,739	1,296,611
A15-peptide with three	6,102	-	6,064	35,149	23,037	80,978	392,714	1,781,781

cleaved N-terminal amino acids								
A9-peptide (control)	1,974,781	1,848,514	2,181,989	2,129,434	2,027,558	1,856,479	1,833,510	1,938,633

(ii) Conversion rate of A15-peptide to PITC conjugated A15-peptide, and % yield for conjugation with PITC to DMSO (1:1000 ratio PITC:DMSO) over multiple rounds (**Fig. 5b**).

First round average % yield (peptide with cleaved N-terminal amino acid / non-modified peptide * 100)	82.43
First round average conversion rate from A15-peptide to PITC conjugated to A15-peptide (%)	98.48
Second round average conversion rate from A15-peptide with cleaved N-terminal amino acid to PITC conjugated to A15-peptide with cleaved N-terminal amino acid (%)	98.86
Third round average conversion rate from A15-peptide with two cleaved N-terminal amino acids to PITC conjugated to A15-peptide with two cleaved N-terminal amino acids (%)	97.68

(iii) Gel size changes throughout in-gel Edman degradation for PITC (1:1000 ratio PITC:DMSO) conjugation over multiple rounds (**Fig. 4a**).

Conjugation with PITC to DMSO (1:1000 ratio PITC:DMSO)- Gel size changes in Edman steps	**Gelation 1, area (cm^2)**	**Gelation 2, area (cm^2)**	**Gelation 3, area (cm^2)**	**Mean (normalized to surface area in 1M Tris pH 9.5) (cm^2)**	**Standard deviation (cm^2)**
1. 3x wash with Tris 1M pH 9.5	0.250	0.275	0.275	1.000	0.000
2. 2x DMSO wash	0.250	0.303	0.275	1.034	0.058
3. PITC conjugation	0.250	0.330	0.275	1.069	0.115
4. TFA	0.250	0.275	0.225	0.938	0.105
5. 2x DMSO wash	0.160	0.180	0.180	0.650	0.008
6. 3x wash with Tris 1M pH 9.5	0.250	0.275	0.275	1.000	0.000
7. 2x DMSO wash	0.225	0.250	0.248	0.903	0.005
8. PITC conjugation	0.250	0.330	0.248	1.034	0.153
9. TFA	0.250	0.275	0.248	0.966	0.058
10. DMSO wash	0.160	0.180	0.180	0.650	0.008
11. 3x wash with Tris 1M pH 9.5	0.250	0.275	0.275	1.000	0.000
12. 2x DMSO wash	0.225	0.275	0.225	0.906	0.091
13. PITC conjugation	0.225	0.250	0.225	0.875	0.050
14. TFA	0.225	0.250	0.225	0.875	0.050
15. DMSO wash	0.160	0.180	0.160	0.625	0.038
16. 3x wash with Tris 1M pH 8	0.250	0.275	0.225	0.938	0.105

Supplementary Table 7

Values and analysis for the LC/QToF trypsinization assay from Fig. 6.

(i) Raw data values for the area under the curve (AUC) of the chromatogram of various species for ClickP (1:1000 ratio ClickP:DMSO) conjugation (Fig. 6c) (extracted based on the exact mass, see Methods for more details, and Source Data for raw traces). Three tables for the three separate gelation solutions. The cells highlighted in grey are the ones used to calculate the conversion rate (solvent with ClickP["A15-peptide"] / solvent ["A15-peptide"]). A15-peptide stands for AGGAGGLLGGSRGGK{acr}. Cells with '-' means that the species was not detected in that sample using the automatic extraction.

	Gelation	DMSO	DMSO with ClickP	DMSO, then TFA	DMSO with ClickP, then TFA	Conversion rate
A15-peptide	Gelation 1	13,053,858	208,055	11,854,227	214,082	98.41
PITC conjugated A15-peptide		-	2,892,880	-	-	
A15-peptide with cleaved N-terminal amino acid		86,430	3,028,163	84,602	5,280,782	
A9-peptide (control)		1,990,306	1,513,936	2,062,392	1,734,325	

	Gelation	DMSO	DMSO with ClickP	DMSO, then TFA	DMSO with ClickP, then TFA	Conjugation Efficiency
A15-peptide	Gelation 2	11,948,873	169,549	12,536,770	188,635	98.58
PITC conjugated A15-peptide		-	2,317,192	-	13,628	
A15-peptide with cleaved N-terminal amino acid		67,053	1,757,956	78,724	4,698,120	
A9-peptide (control)		1,512,368	2,079,931	1,997,604	2,253,932	

	Gelation	DMSO	DMSO with ClickP	DMSO, then TFA	DMSO with ClickP, then TFA	Conjugation Efficiency
A15-peptide	Gelation 3	9,167,841	108,656	9,047,914	197,729	98.81
PITC conjugated A15-peptide		-	1,667,014	-	10,077	
A15-peptide with cleaved N-terminal amino acid		61,582	1,218,684	63,182	5,427,226	
A9-peptide (control)		1,764,034	1,930,392	1,859,683	1,953,816	

(ii) Conversion rate of A15-peptide to ClickP conjugated A15-peptide, and % yield for conjugation

with ClickP (1:1000 ratio ClickP:DMSO) (**Fig. 6c**).

Average % yield (peptide with cleaved N-terminal amino acid / non-modified peptide * 100)	49.26
Average conversion rate (%)	98.60

(iii) Statistics: One-way Analysis of Variance (ANOVA) and Tukey's post-hoc Honestly Significant Difference (HSD) test on the abundance of peptide with cleaved N-terminal amino acid using ClickP (1:1000 ratio ClickP:DMSO) for conjugation (**Fig. 6ciii**). In Tukey HSD results, conditions are abbreviated, A: DMSO, B: DMSO with ClickP, C: DMSO then TFA, D: DMSO with ClickP then TFA.

ANOVA Results					
Test	Comparison	sum_sq	df	F	PR(>F)
ANOVA	group	5.14E+13	3.00E+00	6.77E+01	5.01E-06
ANOVA	Residual	2.02E+12	8.00E+00		

Tukey HSD Results							
Test	group1	group2	meandiff	p-adj	lower	upper	reject
Tukey HSD	A	B	1.93E+06	6.71E-03	6.15E+05	3.25E+06	TRUE
Tukey HSD	A	C	3.81E+03	1.00E+00	-1.31E+06	1.32E+06	FALSE
Tukey HSD	A	D	5.06E+06	8.19E-06	3.75E+06	6.38E+06	TRUE
Tukey HSD	B	C	-1.93E+06	6.79E-03	-3.24E+06	-6.11E+05	TRUE
Tukey HSD	B	D	3.13E+06	2.82E-04	1.82E+06	4.45E+06	TRUE
Tukey HSD	C	D	5.06E+06	8.24E-06	3.75E+06	6.38E+06	TRUE

Supplementary Table 8

Values and analysis for the LC/QToF trypsinization assay from **Supplementary Figures**.

(A & B) Raw data values for the area under the curve (AUC) of the chromatogram of various species for results from in-gel Edman degradation with 1:1000 ratio PITC:pyridine (**Supplementary Table 8Ai**), 1:1000 PITC ratio PITC:acetonitrile (**Supplementary Table 8Aii**) (see **Supplementary Figure 3ai and 3aii** for the plots of these results), 1:9 ratio PITC to acetonitrile) (**Supplementary Table 8Bi**), 1:9 ratio PITC to 1:1 pyridine and water (**Supplementary Table 8Bii**), 1:9 ratio PITC:0.1 M $NaHCO_3$ pH 8.5 (**Supplementary Table 8Biii**) (see **Supplementary Figure 3bi and 3bii and 3biii** for the plots of these results) (extracted based on the exact mass, see **Methods** for more details, and **Source Data** for raw traces). One table is produced for the single gelation solution for each solvent. The cells highlighted in grey are the ones used to calculate the % yield (solvent with PITC and then TFA["A15-peptide"] / solvent and then TFA["A15-peptide"]). A15-peptide stands for AGGAGGLLGGSRGGK{acr}. Cells with '-' means that the species was not detected in that sample using the automatic extraction.

Table 8 A i) Conjugation with using PITC to pyridine (1:1000 ratio PITC:pyridine)	**Gelation**	**pyridine**	**pyridine with PITC**	**pyridine, then TFA**	**pyridine with PITC, then TFA**	**% yield (peptide with cleaved N-terminal amino acid / non-modified peptide * 100)**
A15-peptide	Gelation 1	15,197,533	15,865,081	17,420,478	14,652,655	0.57
PITC conjugated A15-peptide		-	-	-	-	
A15-peptide with cleaved N-terminal amino acid		90,340	85,353	102,755	98,890	
A9-peptide (control)		3,669,196	3,818,575	3,538,687	3,724,464	

Table 8 A ii) Conjugation with using PITC to acetonitrile (1:1000 ratio PITC:acetonitrile)	**Gelation**	**Acetonitrile**	**acetonitrile with PITC**	**acetonitrile, then TFA**	**acetonitrile with PITC, then TFA**	**% yield (peptide with cleaved N-terminal amino acid / non-modified peptide * 100)**
A15-peptide	Gelation 1	12,415,553	16,650,344	14,832,564	12626395	0.46
PITC conjugated A15-peptide		-	-	-	-	
A15-peptide with cleaved N-terminal amino acid		60,749	99,444	69,906	67,941	
A9-peptide (control)		2,350,910	3,277,849	3,766,113	3,979,891	

Table 8 B i) Conjugation using PITC to acetonitrile (1:9 ratio PITC:acetonitrile)	**Gelation**	**Acetonitrile**	**acetonitrile with PITC**	**acetonitrile, then TFA**	**acetonitrile with PITC, then TFA**	**% yield (peptide with cleaved N-terminal amino acid / non-modified peptide * 100)**
A15-peptide	Gelation 1	7,493,808	8,463,312	5,112,020	6,006,358	0.27

PITC conjugated A15-peptide		-	-	-	-	
A15-peptide with cleaved N-terminal amino acid		21,128	21,568	11,002	13,558	
A9-peptide (control)		2,150,273	1,945,023	1,902,023	2,216,702	

Table 8 B ii) Conjugation using PITC in 1:1 pyridine to water (1:9 ratio PITC to 1:1 pyridine to water)	**Gelation**	**1:1 pyridine to water**	**1:1 pyridine to water with PITC**	**1:1 pyridine to water, then TFA**	**1:1 pyridine to water with PITC, then TFA**	**% yield (peptide with cleaved N-terminal amino acid / non-modified peptide * 100)**
A15-peptide	Gelation 1	7,681,288	565,768	6,802,453	1,242,845	59.34
PITC conjugated A15-peptide		5,431	2,594,118	-	-	
A15-peptide with cleaved N-terminal amino acid		12,188	2,144,764	10,933	4,036,719	
A9-peptide (control)		1,921,927	1,836,294	2,052,361	2,052,891	

Table 8 B iii) Conjugation using PITC 0.1 M $NaHCO_3$ pH 8.5 (1:9 ratio PITC:0.1 M $NaHCO_3$ pH 8.5)	**Gelation**	**0.1 M $NaHCO_3$ pH 8.5**	**0.1 M $NaHCO_3$ pH 8.5 with PITC**	**0.1 M $NaHCO_3$ pH 8.5, then TFA**	**0.1 M $NaHCO_3$ pH 8.5 with PITC, then TFA**	**% yield (peptide with cleaved N-terminal amino acid / non-modified peptide * 100)**
A15-peptide	Gelation 1	6,429,218	4,290,786	6,016,772	4,163,519	10.55
PITC conjugated A15-peptide		-	90,914	7,154	5,535	
A15-peptide with cleaved N-terminal amino acid		24,410	1,089,746	22,675	634,809	
A9-peptide (control)		1,940,911	2,125,041	1,879,287	2,120,299	

(C) Raw data values for the area under the curve (AUC) of the chromatogram of various species from in-gel Edman degradation with different ratios of PITC to DMSO (1:100, 1:1,000, 1:10,0000 ratios PITC:DMSO) (see **Supplementary Figure 4** for the plots of these results) (extracted based on the exact mass, see **Methods** for more details, and **Source Data** for raw traces). Three tables for the three separate ratios of PITC to DMSO (1:100, 1:1,000, 1:10,000 PITC:DMSO). The cells highlighted in grey are the ones used to calculate the conversion rate (solvent with PITC["A15-peptide"] / solvent ["A15-peptide"]). A15-peptide stands for AGGAGGLLGGSRGGK{acr}. Cells with '-' means that the species was not detected in that sample using the automatic extraction.

Conjugation with PITC to DMSO (1:100 ratio PITC:DMSO)	**Gelation**	**DMSO**	**DMSO with PITC**	**DMSO, then TFA**	**DMSO with PITC, then TFA**	**Conversion rate**
A15-peptide	Gelation 1	9,052,218	47,171	7,989,893	42,236	99.48

PITC conjugated A15-peptide		-	392,164	-	-	
A15-peptide with cleaved N-terminal amino acid		48,959	5,573,096	40,151	5,936,781	
A9-peptide (control)		4,742,905	4,541,611	4,878,650	5,148,177	

Conjugation with 1:1000 PITC to DMSO (1:1000 ratio PITC:DMSO)	**Gelation**	**DMSO**	**DMSO with PITC**	**DMSO, then TFA**	**DMSO with PITC, then TFA**	**Conversion rate**
A15-peptide		9,382,061	64,147	7,360,985	77,077	
PITC conjugated A15-peptide		-	438,161	-	-	
A15-peptide with cleaved N-terminal amino acid	Gelation 1	48,913	6,422,687	35,693	6,522,567	99.32
A9-peptide (control)		4,920,745	4,448,116	5,565,431	4,697,838	

Conjugation with 1:10,000 PITC to DMSO (1:10,000 ratio PITC:DMSO)	**Gelation**	**DMSO**	**DMSO with PITC**	**DMSO, then TFA**	**DMSO with PITC, then TFA**	**Conversion rate**
A15-peptide		7,542,918	959,140	8,492,660	1,109,025	
PITC conjugated A15-peptide		-	297842	-	-	
A15-peptide with cleaved N-terminal amino acid	Gelation 1	37,026	4,876,060	44,617	4,574,983	87.28
A9-peptide (control)		4,616,757	4,847,843	5,146,408	4,922,362	

(D) Raw data values for **Supplementary Figure 5** representing the area under the curve (AUC) of the chromatogram of various peptide fragments derived from the peptide denoted A15 (AGGAGGLLGGSRGGK{acr}) expected after trypsinization from the gel (denoted "I_product"). Two products are shown in different tables: the first, "non-modified A15-peptide" (AGGAGGLLGGSR) and "A15-peptide with cleaved N-terminal amino acid" (GGAGGLLGGSR). The A9-peptide is denoted "standard" (AGGAGK{acr}GLR) and was used in all samples as the standard with a constant "C_standard" of 5 µM. TheAUC of the extracted ion chromatogram for the A9-peptide is recorded as "I_standard" (extracted based on the exact mass, see **Analysis of LC/QToF data** for details on method of extraction). The AUC of non-modified A15-peptide and A15-peptide with cleaved N-terminal amino acid were measured from 5 µM to 60 µM (in 5 µM increments; denoted: "C_product") with A9-peptide concentration always at 5 µM.

Replicate 1	**5 µM peptide "C_product"**	**10 µM peptide "C_product"**	**20 µM peptide "C_product"**	**30 µM peptide "C_product"**	**40 µM peptide "C_product"**	**50 µM peptide "C_product"**	**60 µM peptide "C_product"**

Non-modified A15-peptide "I_product"	3,799,448	6,085,960	11,319,222	14,253,684	19,809,211	23,947,646	25,596,532
A15-peptide with cleaved N-terminal amino acid "I_product"	2,525,298	5,020,612	10,220,593	13,724,111	17,279,065	19,271,410	23,143,493
A-9 control peptide "I_standard"	1,975,582	1,551,193	2,617,067	2,297,308	2,646,120	2,173,907	2,440,661
Non-modified A15-peptide "I_product/I_standard"	1.92	3.92	4.33	6.20	7.49	11.02	10.49
A15-peptide with cleaved N-terminal amino acid "I_product/I_standard"	1.28	3.24	3.91	5.97	6.53	8.86	9.48

Replicate 2	**5 µM peptide "C_product"**	**10 µM peptide "C_product"**	**20 µM peptide "C_product"**	**30 µM peptide "C_product"**	**40 µM peptide "C_product"**	**50 µM peptide "C_product"**	**60 µM peptide "C_product"**
Non-modified A15-peptide "I_product"	2,744,052	4,939,161	8,944,190	12,276,463	15,692,870	17,966,806	20,359,400
A15-peptide with cleaved N-terminal amino acid "I_product"	3,446,088	6,532,592	11,438,018	16,089,772	20,973,147	23,719,649	27,786,734
A-9 control peptide "I_standard"	2,302,553	2,156,882	2,052,737	2,045,822	2,041,539	2,341,466	2,028,886
Non-modified A15-peptide "I_product/I_standard"	1.19	2.29	4.36	6.00	7.69	7.67	10.03
A15-peptide with cleaved N-terminal amino acid "I_product/I_standard"	1.50	3.03	5.57	7.86	10.27	10.13	13.70

Replicate 3	**5 µM peptide "C_product"**	**10 µM peptide "C_product"**	**20 µM peptide "C_product"**	**30 µM peptide "C_product"**	**40 µM peptide "C_product"**	**50 µM peptide "C_product"**	**60 µM peptide "C_product"**
Non-modified A15-peptide "I_product"	4,068,182	4,683,734	8,511,458	11,651,232	15,479,768	17,802,635	19,996,525
A15-peptide with cleaved N-terminal amino acid "I_product"	3,455,652	5,597,608	9,316,407	13,460,858	17,144,374	20,441,281	23,081,247
A-9 control peptide "I_standard"	2,249,002	2,450,888	2,032,303	2,595,972	2,306,414	2,400,407	2,264,781
Non-modified A15-peptide "I_product/I_standard"	1.81	1.91	4.19	4.49	6.71	7.42	8.83

A15-peptide with cleaved N-terminal amino acid "I_product/I_standard"	1.54	2.28	4.58	5.19	7.43	8.52	10.19

	5 µM peptide "C_product"	**10 µM peptide "C_product"**	**20 µM peptide "C_product"**	**30 µM peptide "C_product"**	**40 µM peptide "C_product"**	**50 µM peptide "C_product"**	**60 µM peptide "C_product"**
A-9 control peptide "C_standard"	5	5	5	5	5	5	5
Mean "C_product/C_standard"	1	2	4	6	8	10	12

Concentration of peptide (µM) "C_product"	**5 µM peptide "C_product"**	**10 µM peptide "C_product"**	**20 µM peptide "C_product"**	**30 µM peptide "C_product"**	**40 µM peptide "C_product"**	**50 µM peptide "C_product"**	**60 µM peptide "C_product"**
Mean non-modified peptide "I_product/I_standard"	1.63	2.55	4.29	5.50	7.29	8.63	9.79
Standard deviation non-modified peptide	0.39	1.07	0.09	0.94	0.51	2.01	0.86
Mean peptide with cleaved N-terminal amino acid "I_product/I_standard"	1.44	2.78	4.62	6.241	7.92	9.17	10.99
Standard deviation peptide with cleaved N-terminal amino acid	0.14	0.50	0.84	1.38	1.95	0.85	2.26
T-statistic (Welch's t-test, n=3) comparing the two means of "I_product/I_standard" for non-modified A15-peptide and A15-peptide with cleaved N-terminal amino acid	0.847	-0.208	-0.816	-0.808	-0.672	-0.372	-0.961
P-value (Welch's t-test, n=3) comparing the two means of "I_product/I_standard" for non-modified A15-peptide and peptide with A15-peptide with cleaved N-terminal amino acid	0.470	0.849	0.499	0.470	0.563	0.737	0.418

Supplementary Table 9

a) ClpS2-StV1 assay with anti-HA antibody 488 comparing different N-terminal amino acids (phenylalanine, alanine, tryptophan, tyrosine)

(i) Raw average fluorescence intensity values for each gel, average, and standard deviation, over the replicates for **Fig. 4d**.

Condition name	Average intensity ExMre gel 1	Average intensity ExMre gel 2	Average intensity ExMre gel 3	Average intensity over replicates	Standard deviation over replicates
Phenylalanine (F)	13,731	7,847	9,084	**10,221**	3102.31725
Alanine (A)	308	277	255	**280**	26.6270539
Tryptophan (W)	561	347	1084	**664**	379.142453
Tyrosine (Y)	1,741	873	1,179	**1,264**	440.2469

(ii) Fold difference between the average intensity read-out for **Fig. 4d** based on values documented in (i).

Fold change	
F/A	37
F/W	15
F/Y	8

b) ClpS2-StV1 assay with anti-HA antibody 488 on F_1 peptide combined with in-gel Edman degradation.

(i) Raw average fluorescence intensity values for each gel, average, and standard deviation, over the replicates for **Fig. 4fiii**.

Condition name	Average intensity ExMre gel 1	Average intensity ExMre gel 2	Average intensity over replicates	Standard deviation over replicates
(1) DMSO	5,381	7,731	**6,556**	1,662
(2) PITC:DMSO	474	411	**443**	45
(3) DMSO+TFA	6,290	8,694	**7,492**	1,700
(4) PITC:DMSO+TFA	400	640	**520**	170
(5) PITC:DMSO+TFA+trypsin	250	606	**428**	252

(ii) Fold difference between the average intensity read-out for **Fig. 4fiii** based on values documented in (i) (e.g., condition (1) is DMSO and (3) is TFA, as number in (i))

Fold change in average fluorescence intensity read-out	Ratio
(1)/(2)	14.82
(1)/(3)	0.88
(1)/(4)	12.61
(1)/(5)	15.32

Fold change in average fluorescence intensity read-out	Ratio
(3)/(1)	1.14
(3)/(2)	16.93
(3)/(4)	14.41
(3)/(5)	17.50

c) ClpS2-StV1 assay with anti-HA antibody 488 on G_1F_2 peptide combined with in-gel Edman degradation.

(i) Raw average fluorescence intensity values for each gel, average, and standard deviation, over the replicates for **Fig. 4giii**.

Condition name	**Average intensity ExMre gel 1**	**Average intensity ExMre gel 2**	**Average intensity ExMre gel 3**	**Average intensity over replicates**	**Standard deviation over replicates**
(1) DMSO	658	418	339	**472**	166
(2) PITC:DMSO	485	526	491	**501**	22
(3) DMSO+TFA	499	589	450	**513**	71
(4) PITC:DMSO+TFA	2,463	3,802	3,235	**3167**	672
(5) PITC:DMSO+TFA +trypsin	266	357	484	**369**	109

(ii) Fold difference between the average intensity read-out for **Fig. 4giii** based on values documented in (i) (e.g., condition (4) is PITC:DMSO+TFA, as number in (i))

Fold change in average	Ratio

fluorescence intensity read-out	
(4)/(1)	6.71
(4)/(2)	6.32
(4)/(3)	6.18
(4)/(5)	8.58

Supplementary Table 10

a) Raw fluorescence intensity values for the binding assay with biotin-ClickP-aa in ExMre gels with streptavidin using Glyphic-V (**Fig. 6e**), including the average intensity for each gel, the average and standard deviation over the replicates.

Condition name	Average intensity ExMre gel 1	Average intensity ExMre gel 2	Average intensity ExMre gel 3	Average intensity over replicates	Standard deviation over replicates
Phenylalanine	166	171	141	159	16
Glycine	169	170	148	162	12
Valine	1,615	1,425	1,275	1,438	170

Fold change	
Val/Phe	9.03
Val/Gly	8.86

b) Raw fluorescence intensity values for the binding assay with biotin-ClickP-aa in ExMre gels with streptavidin using Glyphic-F (**Fig. 6f**), including the average intensity for each gel, the average and standard deviation over the replicates.

Condition name	Average intensity ExMre gel 1	Average intensity ExMre gel 2	Average intensity ExMre gel 3	Average intensity over replicates	Standard deviation over replicates
Phenylalanine	371	299	413	361	58
Glycine	114	117	119	117	3
Valine	138	145	142	142	4

Fold change	
Phe/Gly	3.09
Phe/Val	2.55

Supplementary Table 11

Table recording the highest ratio of known oxidation peptide species to original peptide in the ExM gels (descriptive values of **Supplementary Figure 8**).

	Highest ratio of oxidized species to original peptide (oxidized/non-oxidized)	
Methionine	13817063/3628434.7	3.80800
Cysteine	299666.7/54427.54	5.50579
Tyrosine	30242.02/2509423.3	0.01205
Phenylalanine	1166.65/18738042	0.00006
Tryptophan	108509.7/2119325.3	0.05120
Histidine	30999.35/441471.91	0.07022
Proline	294169.1/5514152.5	0.05335
Arginine	6285.23/4906487.6	0.00128

Supplementary Table 12

This table relates to **Supplementary Figure 17** and the kinetic parameters of the binders used in the simulation of *in situ* protein sequencing. The following table shows the binding probabilities for different specificity regimes at a 1 μM binder concentration reaching equilibrium and a subsequent wash time of 30 min. The ratio α represents the ratio between k_{off} of the on-target versus the off-target.

The bound probability for the on-target (correct) is $p_{correct} = (\frac{[C]}{[C] + K_d^{on-target}}) \cdot e^{-k_{off}^{on-target} t_{wash}}$, the unbound probability for the on-target is $p_{correct-unbound} = 1 - p_{correct}$, whereas the bound probability for off-target (incorrect) is $p_{incorrect} = (\frac{[C]}{[C] + K_d^{off-target}}) \cdot e^{-k_{off}^{off-target} t_{wash}}$, and the unbound probability for the off-target is $p_{incorrect-unbound} = 1 - p_{incorrect}$. The k_{off} (s^{-1}) is the inverse of the dwell time, and t_{wash} (min) is the wash time. The k_{on} value is assumed constant across all binders with a value of 10^6.

Specificity	Ratio (α)	Bound probability	Unbound probability	Concentration of binder (μM)	Wash Time (min)	k_on (M^{-1} s^{-1})	k_off (s^{-1})	k_D (M)
Perfect	∞	1	0	-	-	-	-	-
Perfect (off-target)	∞	0	1	-	-	-	-	-
Very High	50,000	0.9139	0.0861	1	30	10^6	$5*10^{-5}$	$5*10^{-11}$
Very High (off-target)	50,000	~0	~1	1	30	10^6	2.5	$2.5*10^{-6}$
High	5000	0.9139	0.0861	1	30	10^6	$5*10^{-5}$	$5*10^{-11}$
High (off-target)	5000	~0	~1	1	30	10^6	0.25	$2.5*10^{-7}$
Medium	500	0.9139	0.0861	1	30	10^6	$5*10^{-5}$	$5*10^{-11}$
Medium (off-target)	500	~0	~1	1	30	10^6	$2.5*10^{-2}$	$2.5*10^{-8}$
Low	50	0.9139	0.0861	1	30	10^6	$5*10^{-5}$	$5*10^{-11}$
Low (off-target)	50	0.0111	0.9890	1	30	10^6	$2.5*10^{-3}$	$2.5*10^{-9}$
Very Low	5	0.9139	0.0861	1	30	10^6	$5*10^{-5}$	$5*10^{-11}$
Very Low (off-target)	5	0.6375	0.3625	1	30	10^6	$2.5*10^{-4}$	$2.5*10^{-10}$