nature biotechnology

Article

https://doi.org/10.1038/s41587-026-03061-z

Single-molecule peptide sequencing through reverse translation of peptides into DNA

Received: 13 October 2025

Accepted: 18 February 2026

Published online: 18 March 2026

Check for updates

Liwei Zheng[1] ✉, Yujia Sun[1], Linus A. Hein[2], Michael Eisenstein[1,2] & Hyongsok Tom Soh[1,2,3] ✉

Despite advances in mass spectrometry and emerging single-molecule approaches, sequencing peptides at the single-molecule level remains a central challenge in proteomics. Here we present a 'reverse translation' strategy that enables single-molecule peptide sequencing with single-amino-acid resolution. In this approach, peptides undergo a modified Edman degradation that iteratively releases N-terminal amino acids tagged with peptide-specific DNA barcodes. Antibody-mediated proximity extension assays identify these barcoded amino acids and generate PCR-amplifiable DNA reporters that record the identity, position and originating peptide of each amino acid. The resulting DNA library is directly read by high-throughput sequencing, converting peptide sequences into digital DNA outputs. Using this approach, we demonstrate true single-molecule peptide sequencing, achieving full sequence coverage in millions of reads and accurate differentiation of both native and post-translationally modified peptides. These results establish a framework that redefines protein sequencing as a DNA sequencing problem and lays the foundation for high-throughput, de novo single-molecule protein sequencing.

The ability to sequence DNA and RNA with single-molecule sensitivity and single-nucleotide resolution has fueled progress in biological research in the past two decades. Enabled by high-throughput sequencing technologies, single-cell transcriptomics has matured into a powerful method for directly quantifying gene expression in individual cells[1–3]. Although transcriptomes are a highly informative indicator of protein expression activity, the ultimate consequences of changes in gene expression can only be fully understood by assessing a cell's protein content. This is because the final outcome of translation is heavily influenced by the effects of translational regulation and kinetics, post-translational modification (PTM) and other biochemical processes[4,5]. Despite ongoing advances in mass spectrometry (MS)-based proteomics[6–9], this technology still lacks the dynamic range and sensitivity to document the full breadth of the cellular proteome[10–12]. This has sparked considerable interest in the development of technologies that enable the single-molecule identification and quantification of proteins.

The past several years have witnessed considerable progress in this domain, including several early-stage commercial platforms (Supplementary Table 1)[13]. Nanopore-based methods have shown potential for profiling intact proteins[14–18] and distinguishing individual amino acids[19–21] but protein sequencing at the single-amino-acid level has not yet been demonstrated[16,18]. Other technologies are designed to sequence shorter peptides at single-amino-acid resolution. Fluorosequencing[22], which combines selective amino acid labeling with iterative Edman degradation, enables parallel fingerprinting of peptides but is constrained by the range of reagents available for selective amino acid labeling[23,24]. Another recently developed approach exploits N-terminal amino acid recognizers derived from ClpS family proteins to identify exposed N-terminal residues after digestion[13]. While these recognizers can discriminate subsets of amino acids, their affinity and binding kinetics are strongly influenced by neighboring residues, which limits their accuracy and generalizability[25,26]. As such, although considerable progress has been made in terms of

[1]Department of Radiology, Stanford University, Stanford, CA, USA. [2]Department of Electrical Engineering, Stanford University, Stanford, CA, USA. [3]Department of Bioengineering, Stanford University, Stanford, CA, USA. ✉e-mail: lwzheng@stanford.edu; tsoh@stanford.edu

https://doi.org/10.1038/s41587-026-03061-z

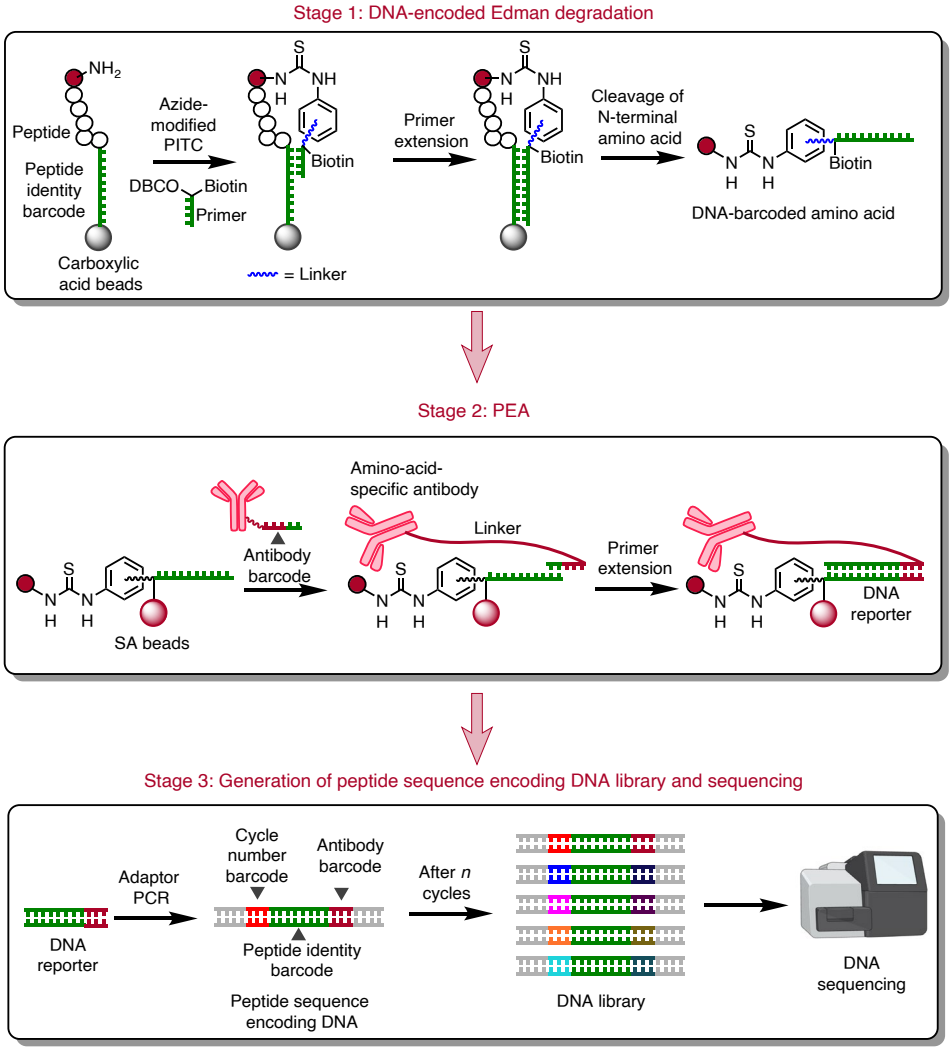


Fig. 1 | Overview of peptide sequencing through reverse translation. In stage 1, peptides are conjugated with peptide identity barcode sequences immobilized on magnetic beads. A modified Edman degradation reaction is performed using a PITC molecule modified with an azide group, enabling click coupling to a DBCO-modified, biotinylated primer that hybridizes directly to the peptide identity barcode sequence. This primer is extended to record the peptide identity barcode before N-terminal amino acid cleavage. In stage 2, the DNA-barcoded amino acids released after cleavage are pulled down by SA beads through their biotin moiety. These immobilized amino acids are recognized by DNA-barcoded amino-acid-specific antibodies and the DNA-barcoded amino acids are subsequently converted into DNA reporters using a PEA. In stage 3, the DNA reporters from each cycle are barcoded with cycle number barcodes during adaptor PCR, yielding a DNA sequence that encodes the identity, position and originating peptide of each amino acid. The final DNA library produced by this process can then be sequenced in a single run, enabling the reconstruction of peptide sequences in the sample.

single-molecule protein profiling and quantification, there remains an unmet need for methods that can accurately identify a wide range of proteins—including PTM variants—with single-molecule sensitivity, single-amino-acid resolution and high throughput.

To address these challenges, we developed a strategy that reframes protein sequencing as a DNA sequencing problem. Rather than directly detecting amino acids within peptide contexts, our approach iteratively releases amino acids from their parent peptides and encodes them into DNA, which are amplified and read with the sensitivity and scalability of next-generation sequencing. We demonstrate that peptide sequence information can be transduced into DNA readouts, enabling high-throughput single-molecule peptide sequencing. This molecular analog of PCR for peptide sequence information is referred to as 'reverse translation'. In our design, peptides undergo a modified Edman degradation that sequentially removes one amino acid from the N terminus of an immobilized peptide[27]. Before cleavage, each amino acid is tagged with a DNA barcode that traces it back to the source peptide. The barcoded amino acids are subsequently identified using proximity extension assays (PEA) with amino-acid-specific or PTM-specific antibodies, generating DNA reporters that encode the identity, position and originating peptide of each residue (Fig. 1). We establish the method with a single seven-residue model peptide and then validate it by carrying out multiplexed sequencing of multiple peptide species simultaneously. We also establish the feasibility of PTM mapping by using antibodies that selectively recognize modifications such as phosphotyrosine. By incorporating unique molecular identifiers (UMIs), we show that our platform is capable of true single-molecule peptide sequencing. Together, these results lay a foundation for high-throughput de novo protein sequencing with single-molecule sensitivity and single-amino-acid resolution.

Design of peptide sequencing through reverse translation

We designed a three-stage process to reverse-translate peptide sequence information into DNA. The first stage entails the conversion of peptides into DNA-barcoded amino acids (Fig. 1, stage 1). To begin,

the C terminus of the peptides being sequenced is conjugated to beads modified with a peptide identity barcode, which is used to differentiate each subpopulation of peptides or individual peptide molecules. The immobilized DNA-barcoded peptides are then subjected to a modified version of the Edman degradation reaction. This reaction is a well-established tool for protein sequencing, where the N-terminal amino acid is conjugated to phenylisothiocyanate (PITC) and cleaved from the peptide under acidic conditions. The resulting degradation fragment is converted to a more stable form and identified, after which the process is repeated to identify the next amino acid. Reaction efficiency is generally not limited by amino acid identity or peptide charge states. Thus, Edman degradation accommodates a wide range of peptide sequences[22]. In our implementation, an azide-modified PITC is used for N-terminal conjugation, which allows the click coupling of a dibenzocyclooctyne (DBCO)-modified biotinylated primer that hybridizes with the peptide identity barcode. The primer is then immediately extended by DNA polymerase, after which the Edman reaction is performed to cleave the N-terminal amino acid. This yields a degradation product that is covalently tagged with the complementary sequence of the peptide identity barcode.

In the second stage, the DNA-barcoded degradation products are converted to DNA reporters through a PEA (Fig. 1, stage 2)[28,29]. The DNA-barcoded degradation products are enriched by streptavidin (SA) pulldown and subsequently recognized by DNA-barcoded amino-acid-specific antibodies. The antibody-based identification of cleaved amino acids is more reliable than methods that attempt to detect N-terminal amino acids within the peptide context, as antibody binding is not affected by neighboring amino acids. Binding events are recorded by PEA, where the resulting DNA reporters contain both the initial peptide-specific barcodes and the antibody-specific barcodes.

In the final stage, the DNA reporters generated by PEA from each cycle are tagged with cycle number barcodes during adaptor PCR before sequencing (Fig. 1, stage 3). This yields a peptide sequence encoding DNA that comprises a set of barcodes encoding the identity, position and originating peptide of each amino acid. The three-stage process is performed iteratively on the immobilized peptide and the resulting DNA library is then sequenced, enabling the reconstruction of the amino acid sequence for each peptide present in the sample.

DNA-encoded Edman degradation

Preserving DNA integrity during Edman degradation is critical to our reverse translation strategy. However, trifluoroacetic acid (TFA), used during the cleavage and conversion reaction of conventional Edman degradation, can cause proton-catalyzed deglycosylation (Fig. 2a)[30,31]. We, therefore, developed reaction conditions that are compatible with DNA. BF_3 etherate-induced Edman degradation in acetonitrile reduces proton-catalyzed racemization, while completely cleaving PITC-modified N-terminal amino acids[32,33]. Moreover, the BF_3–methanol complex has been used as a nondepurinating detritylation reagent in DNA synthesis[34]. Encouraged by these reports, we hypothesized that DNA stability would be enhanced in a cleavage reaction mediated by BF_3 etherate in acetonitrile. However, we observed depurination when native DNA was subjected to treatment with 40 mM BF_3 etherate in anhydrous acetonitrile (Supplementary Fig. 1). We, therefore, explored the use of chemically modified 7-deazapurine deoxynucleotides (c^7dA and c^7dG), which have been reported to be resistant to acid-catalyzed depurination[35,36]. We synthesized **ODN-1** (sequences in Supplementary Table 2; structures of nonnucleotide modifications in Supplementary Table 3), in which dA and dG were substituted with c^7dA and c^7dG, respectively. This modified oligonucleotide was unaffected by BF_3 etherate treatment for 4 h (Fig. 2b and Supplementary Fig. 2). Given that N-terminal amino acid cleavage in the presence of 40 mM BF_3 etherate occurs within 5 min, the stability of 7-deazapurine-modified DNA is sufficient for multiple cycles of the cleavage reaction. Notably, 7-deazapurine deoxynucleotide triphosphates are accepted by several commonly used DNA polymerases (for example, Klenow fragment (exo−), Sequenase version 2.0 and *Bst* 3.0) and primer extension reactions using c^7dATP and c^7dGTP proceeded with efficiency comparable to that of dATP and dGTP (Supplementary Fig. 3).

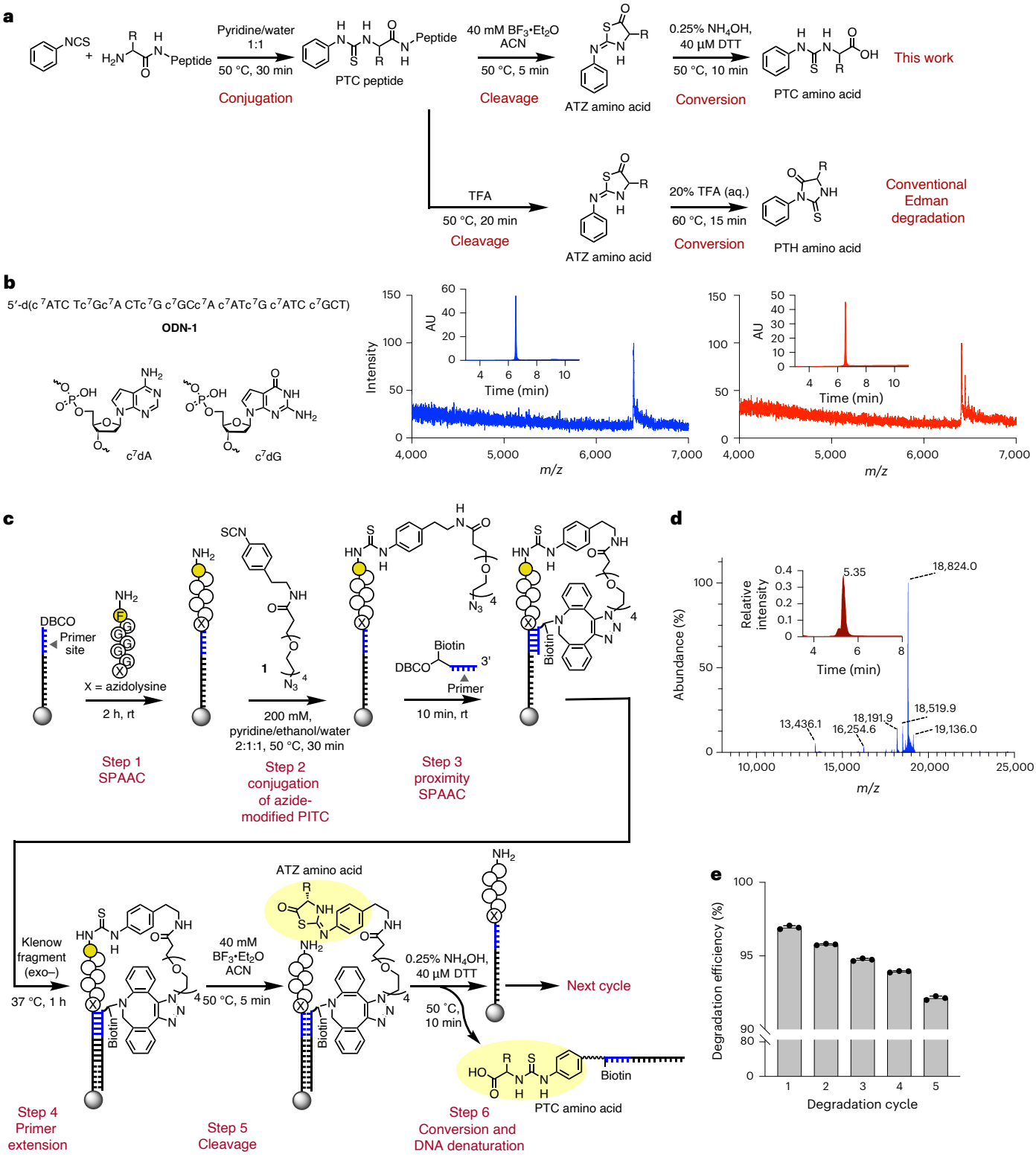
Anilinothiazolinone (ATZ) amino acids generated by BF_3 etherate-mediated cleavage reaction are unstable and, thus, not suitable for antibody-based detection[32]. We substituted the acidic conversion reaction used in conventional Edman degradation with a DNA-compatible conversion reaction under alkaline reducing conditions (Fig. 2a)[37,38]. This reaction converts the ATZ amino acids into stable phenylthiocarbamyl (PTC) amino acids and dithiothreitol (DTT) was added to inhibit oxidative degradation of PTC amino acids[39]. PTC amino acids are readily prepared by reacting PITC or its derivatives with amino acids, greatly simplifying the generation of antibodies against these modified amino acids.

We next developed the DNA-encoded Edman degradation process using our modified Edman degradation reaction to iteratively barcode the generated PTC amino acids (Fig. 2c). The degradation reaction was performed on model peptides bearing an azidolysine immobilized onto magnetic beads, which simplifies solvent exchange during the multistep chemical process while also increasing reagent accessibility to DNA–peptide conjugates when reactions are conducted in neat organic solvents[40–42]. To immobilize the azide-modified peptides, we prepared carboxylic acid beads conjugated with DBCO-modified barcode DNA by 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) coupling and subsequent DNA ligation (Methods). Subsequently, we conjugated a 7-aa peptide (FGGGGGX, where X is azidolysine) to the DNA by a strain-promoted alkyne–azide cycloaddition (SPAAC) reaction (Fig. 2c, step 1) and reacted azide-modified PITC (**1**) with the N terminus of the peptide (Fig. 2c, step 2). For the barcode transfer step, we first conjugated a DBCO-modified primer (**ODN-S13**) to the azide

Fig. 2 | Developing DNA-encoded Edman degradation. a, Reaction scheme of conventional Edman degradation (bottom) and the BF_3 etherate-mediated version presented in this work (top). The TFA used in conventional Edman degradation is incompatible with DNA. We instead use BF_3 etherate, which generates ATZ amino acids that are then converted to more synthetically accessible PTC amino acids under reducing, basic conditions. **b**, MS spectra of oligonucleotide **ODN-1** before (left) and after (right) treatment with 40 mM BF_3 etherate in anhydrous acetonitrile for 4 h. Inset, HPLC chromatograms of **ODN-1** before and after treatment. **c**, Reaction scheme for barcoding PTC amino acids through DNA-encoded Edman degradation on magnetic beads. The azidolysine-modified peptide FGGGGGX is attached to DBCO-modified barcode DNA by click reaction (step 1), after which the peptide's N terminus is reacted with azide-modified PITC (**1**) (step 2). The DBCO-modified primer (**ODN-S13**) is then coupled by proximity SPAAC reaction (step 3). Next, primer extension generates a complementary sequence to the peptide identity barcode (step 4). Cleavage by BF_3 etherate-mediated Edman degradation yields a DNA-barcoded ATZ amino acid, which remains hybridized on the bead because of the insolubility of DNA in acetonitrile (step 5). Lastly, hydrolysis under reducing basic conditions converts the ATZ amino acid to a PTC amino acid and denatures the DNA duplex, releasing the DNA-barcoded PTC amino acid from the solid support (step 6). **d**, MS spectrum of DNA-barcoded PTC-F generated by the first cycle of DNA-encoded Edman degradation on peptide FGGGGGX (calculated molecular weight, 18,824.2; observed molecular weight, 18,824.0). Peaks at 13,436.1, 16,254.6, 18,191.9 and 18,519.9 correspond to truncated products resulting from incomplete primer extension. The peak at 19,136.0 corresponds to the product from A-tailing. Inset, HPLC chromatograms of DNA-barcoded PTC-F. **e**, Efficiency of DNA-encoded Edman degradation over five cycles as measured by flow cytometry. The degradation efficiency was determined by (1 − [mean PE-A after degradation]/[mean PE-A before degradation]) × 100% (Methods). Data were obtained from three independent experiments and are presented as the mean ± s.d. (n = 3). Stepwise yields are 96.9% ± 0.1%, 95.8% ± 0.1%, 94.75% ± 0.08%, 93.94% ± 0.05% and 92.1% ± 0.1%. aq., aqueous solution; rt, room temperature.

https://doi.org/10.1038/s41587-026-03061-z

group on the PTC peptide by SPAAC (Fig. 2c, step 3). The hybridization between the template and primer accelerated the SPAAC reaction by increasing the effective local concentration of reactants[43] and we found that the primer conjugation reaction reached completion with peptide FGGGGGX within 10 min (Supplementary Fig. 4). In contrast, when the reaction was carried out using a noncomplementary poly(dT) sequence (**ODN-S14**), no detectable conjugation was observed (Supplementary Fig. 4), which is consistent with the relatively slow reaction rate of SPAAC (k = 0.2–0.5 $M^{-1} s^{-1}$). The rate acceleration of the proximity SPAAC reaction was general and the click reaction reached completion within 1 h for various peptides ranging from 10 to 30 aa in length and of varying rigidities (Supplementary Fig. 5). For the 10-aa peptide, SPAAC was accelerated 48.3-fold in the hybridization-assisted reaction compared with the reaction in the absence of hybridization. To evaluate the yield of each step, we used a Cy3-labeled anchor DNA bearing an internal disulfide linkage (**ODN-S15**), which enables the release of intermediates for analysis by gel electrophoresis. Steps 1–3 all proceeded with near-quantitative yields (Supplementary Fig. 4).



The primer was subsequently extended using Klenow fragment (exo−; Fig. 2c, step 4), after which the beads were subjected to Edman degradation with BF_3 etherate in acetonitrile (Fig. 2c, step 5). During this step, we took advantage of the insolubility of DNA in acetonitrile[44], which allows the DNA-barcoded ATZ amino acids to remain hybridized with the template DNA following the cleavage reaction. The subsequent basic conversion reaction converts ATZ amino acids into stable PTC amino acids while also denaturing the duplex DNA, releasing the barcoded amino acids (Fig. 2c, step 6)[37]. The structure of the DNA-barcoded PTC amino acid was confirmed by liquid chromatography (LC)–MS (Fig. 2d). The yield of the cleavage reaction was measured by flow cytometry (Fig. 2e) and an average cleavage yield of 95% ± 2% was observed for the five iterative cycles of cleavage with peptide FGGGGGX. We observed a slight decrease in available peptides after each cycle, as indicated by a reduction of fluorescence signal before degradation (Supplementary Fig. 6). It is unlikely that this is because of the instability of the DBCO–azide linkage, as no decomposition of this linkage was observed by MS (Fig. 2d). We hypothesize that the decrease in available peptides is caused by processes that irreversibly block the N terminus, such as oxidative desulfurization of the PTC moiety[45,46]. Lastly, we also validated that peptides bearing N-terminal tryptophan, tyrosine, aspartic acid, arginine or phosphotyrosine are compatible with this process. Correct DNA-barcoded PTC amino acids were observed in all cases, with degradation yields comparable to that of N-terminal phenylalanine (Supplementary Figs. 7 and 8).

Conversion of barcoded PTC amino acids to DNA reporters by proximity extension

The DNA-barcoded PTC amino acids were subsequently converted into DNA reporters by an antibody-binding-mediated PEA. For PTC amino acids with no existing antibody, we generated monoclonal antibodies by immunizing mice with BSA conjugated to various PTC amino acids (Supplementary Fig. 9). The binding performance of the antibodies was tested against PTC amino acids conjugated to a biotinylated DNA (**ODN-S19**) on biolayer interferometry (BLI; Supplementary Fig. 10). We obtained antibodies that achieved high specificity and affinity for PTC-modified phenylalanine (PTC-F; Fig. 3a), tryptophan (PTC-W; Fig. 3b), arginine (PTC-R; Fig. 3c) and aspartic acid (PTC-D; Fig. 3d), as measured by BLI. In addition, antibodies to PTC glutamine (PTC-Q; Supplementary Fig. 11a) and PTC tyrosine (PTC-Y; Supplementary Fig. 11b) were also obtained with modest selectivity against structurally similar PTC amino acids. On the other hand, antibodies targeting PTMs primarily recognize the modified amino acid side chains. We hypothesize that these antibodies will maintain their affinity against the corresponding PTC amino acids bearing PTMs. Consequently, we experimentally confirmed that commercially available antibodies to phosphotyrosine (^{P}Y), asymmetric dimethylarginine, phosphoserine (^{P}S) and acetylated lysine (^{Ac}K) all retained their high affinity for corresponding PTC amino acids (Fig. 3e and Supplementary Fig. 12).

To perform PEA, we first functionalized the Fc domain of anti-PTC amino acid antibodies with tetrazine (Tz; Fig. 3f and Supplementary Fig. 13a,b), then *trans*-cyclooctene (TCO)-modified DNA sequences consisting of the PCR reverse primer and the PEA primer were conjugated to antibodies by click chemistry (**ODN-2**, **ODN-3** or **ODN-4**; Fig. 3f and Supplementary Fig. 13c). Independently synthesized PTC amino acids were conjugated with a model barcode DNA (**ODN-5**) containing the PCR forward primer and the PEA primer-binding site by SPAAC and pulled down on SA beads (Fig. 3f). The immobilized DNA-barcoded PTC amino acids were recognized by their corresponding primer-functionalized antibodies and this binding event stabilized the primer–template complex and enabled primer extension.

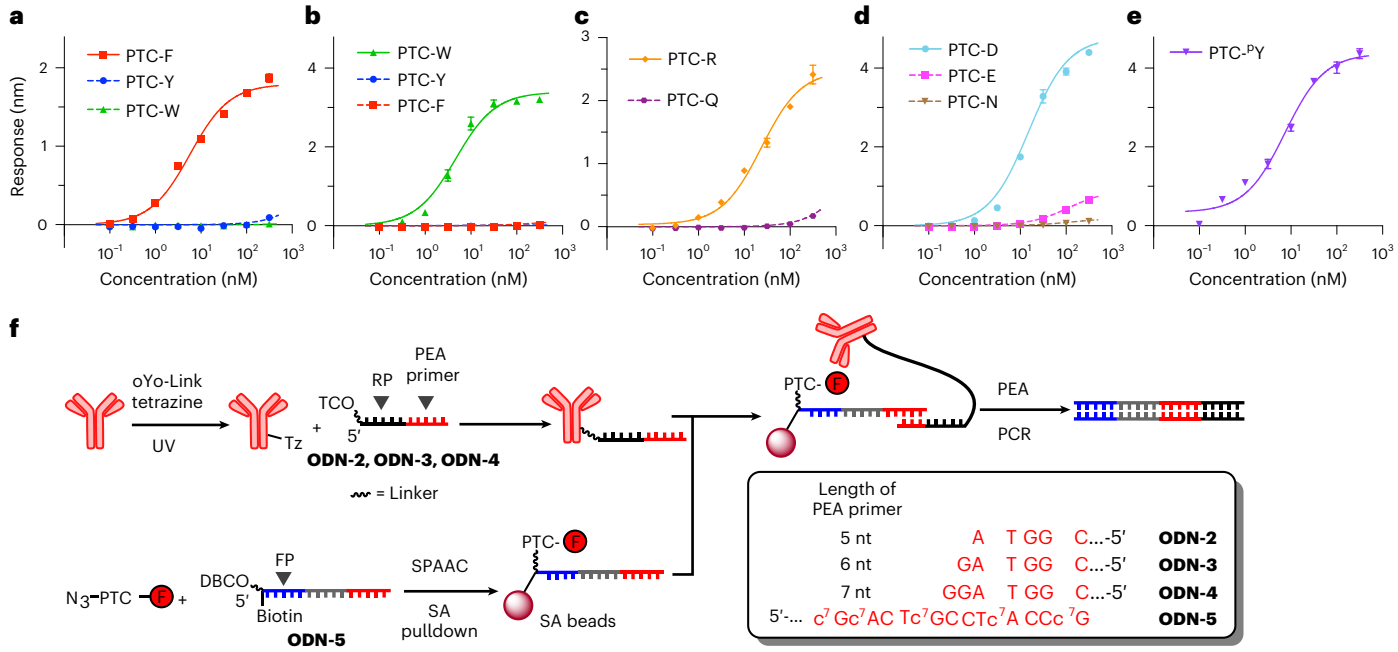
In conventional PEA, a quantitative DNA output is not required for the quantification of the molecule of interest[47]. However, PEA efficiency is critical for our method to achieve high sequencing coverage and single-molecule sensitivity. Specificity is also essential for achieving high-fidelity reads. There are two competing reactions during proximity extension: specific primer extension that only occurs with help from antibody–antigen binding and nonspecific primer extension initiated in the absence of antibody–antigen recognition. The efficiency of specific primer extension is affected by the affinity of the antibodies and the stability of the primer–template complex, while the extent of nonspecific primer extension is only affected by the latter. Thus, tuning primer–template complex stability could increase the yield of specific primer extension and suppress the nonspecific one. To test this, we assessed the specificity and efficiency of proximity extension when barcoded PTC-F, PTC-W or PTC-Y were incubated with anti-PTC-F antibodies modified with PEA primer sequences with lengths of 5–7 nt (**ODN-2**, **ODN-3** or **ODN-4**, respectively). Considerable nonspecific primer extension was observed with **ODN-4**, whereas the shorter primers favored specific primer extension (Fig. 3g). Next, we evaluated the efficiency of proximity extension by qPCR (Supplementary Fig. 14). At a concentration of 100 nM, anti-PTC-F antibody modified with **ODN-2** only converted 24% of PTC-F molecules to a DNA output (Supplementary Fig. 14b). On the other hand, **ODN-3**-modified anti-PTC-F yielded complete primer extension at the same concentration (Supplementary Fig. 14c). In addition to primer length, the stability of primer–template complexes is

Fig. 3 | PEA-mediated conversion of DNA-barcoded PTC amino acids into DNA reporters. a–**e**, Binding curves for anti-PTC-F (**a**), anti-PTC-W (**b**), anti-PTC-R (**c**), anti-PTC-D (**d**) and PY20 antibody (**e**) against target and nontarget DNA-tethered PTC-modified amino acids, as measured by BLI. Calculated K_d values for the primary antibody target were 5.7 nM (anti-PTC-F), 4.9 nM (anti-PTC-W), 23.7 nM (anti-PTC-R) and 15.2 nM (anti-PTC-D) and 7.4 nM (PY20). Data were obtained from three independent measurements and are presented as the mean ± s.d. ($n = 3$). **f**, Scheme for converting DNA-barcoded PTC amino acids to DNA reporters. First, anti-PTC-F antibody is modified with photoreactive antibody-binding domains (that is, oYo-Link) to introduce a Tz functional group onto the Fc region. Next, a TCO-modified DNA strand featuring a PEA primer sequence of 5–7 nt (**ODN-2**, **ODN-3** or **ODN-4**) is conjugated with the Tz group to form a primer–antibody conjugate. PTC-F is tagged with biotinylated template DNA (**ODN-5**) and immobilized onto SA beads. The beads are incubated with primer-modified anti-PTC-F antibody and then Klenow fragment (exo−) is added to perform PEA. The resulting product is PCR amplified for further analysis. **ODN-2**, **ODN-3** and **ODN-4** contain TCO modification. DNA segments relevant to PEA are shown for clarity. The linker consists of three hexaethylene glycol (that is, spacer18) modifications. **g**, Native PAGE of products generated by PEA with primers of varying lengths. The use of 5-nt and 6-nt PEA primers (**ODN-2** and **ODN-3**) can specifically distinguish antibody–antigen interactions, whereas the longer 7-nt primer (**ODN-4**) produces a nonspecific signal. Experiments were repeated twice independently, yielding similar results. **h**, Quantification of DNA output of PEA by qPCR. PEA was carried out with 30 nM **ODN-3**-tagged anti-PTC-F antibody and **ODN-5**–PTC-F immobilized on SA beads. For the positive control, primer extension was performed using a 12-nt primer (**ODN-S20**, 30 nM) and **ODN-5** immobilized on SA beads without PTC-F conjugation. Nonspecific PEA was performed as a negative control, with **ODN-5** immobilized on SA beads and 30 nM **ODN-3**-tagged anti-PTC-F antibody. **i**, PEA efficiency of other anti-PTC amino acid antibodies to their corresponding PTC-modified amino acids. All antibodies were modified with **ODN-3** and PEA was carried out with 100 nM primer–antibody conjugate. Data in **h** and **i** were obtained from three independent experiments and are presented as the mean ± s.d. ($n = 3$). **j**, Relative PEA efficiency of four custom anti-PTC amino acid antibodies to 18 PTC-modified amino acids. All antibodies were modified with **ODN-3** and PEA was carried out with 30 nM primer–antibody conjugate. The relative efficiency was determined by comparison to the PEA efficiency of the specific primer extension achieved by each antibody in the presence of its target PTC amino acid, which was set as 100%. Extension efficiencies are calculated from three independent experiments and the means are plotted. RP, reverse primer; FP, forward primer.

https://doi.org/10.1038/s41587-026-03061-z

also affected by the primer concentration. Because nonspecific primer extension solely originates from primer–template hybridization, it can be greatly inhibited by reducing the primer concentration. With 30 nM **ODN-3**-conjugated anti-PTC-F, the extent of nonspecific primer extension was just 1% of that of specific primer extension and the reduced concentration had minimal impact on the yield of specific primer extension (Fig. 3h). In contrast, increasing the antibody concentration to 300 nM only increased the extent of nonspecific primer extension (Supplementary Fig. 14c).

PEA efficiencies were determined for other antibodies conjugated with **ODN-3**. Specific primer extension reactions were observed for all antibodies, with a positive correlation between the yield of the proximity extension reaction and the affinity of the antibody (Fig. 3i). We evaluated the degree of cross-reactivity during proximity extension for all four custom antibodies that showed high specificity on BLI against 18 different PTC amino acids using qPCR (Fig. 3j). We excluded only cysteine and lysine from this assay, as their PTC amino acid counterparts cannot be independently synthesized for testing because of modification of the side chain by azide-modified PITC (**1**). All four antibodies demonstrated high specificity for their corresponding PTC-modified amino acids, with the majority exhibiting <2% cross-reactivity against nontarget amino acids.



https://doi.org/10.1038/s41587-026-03061-z

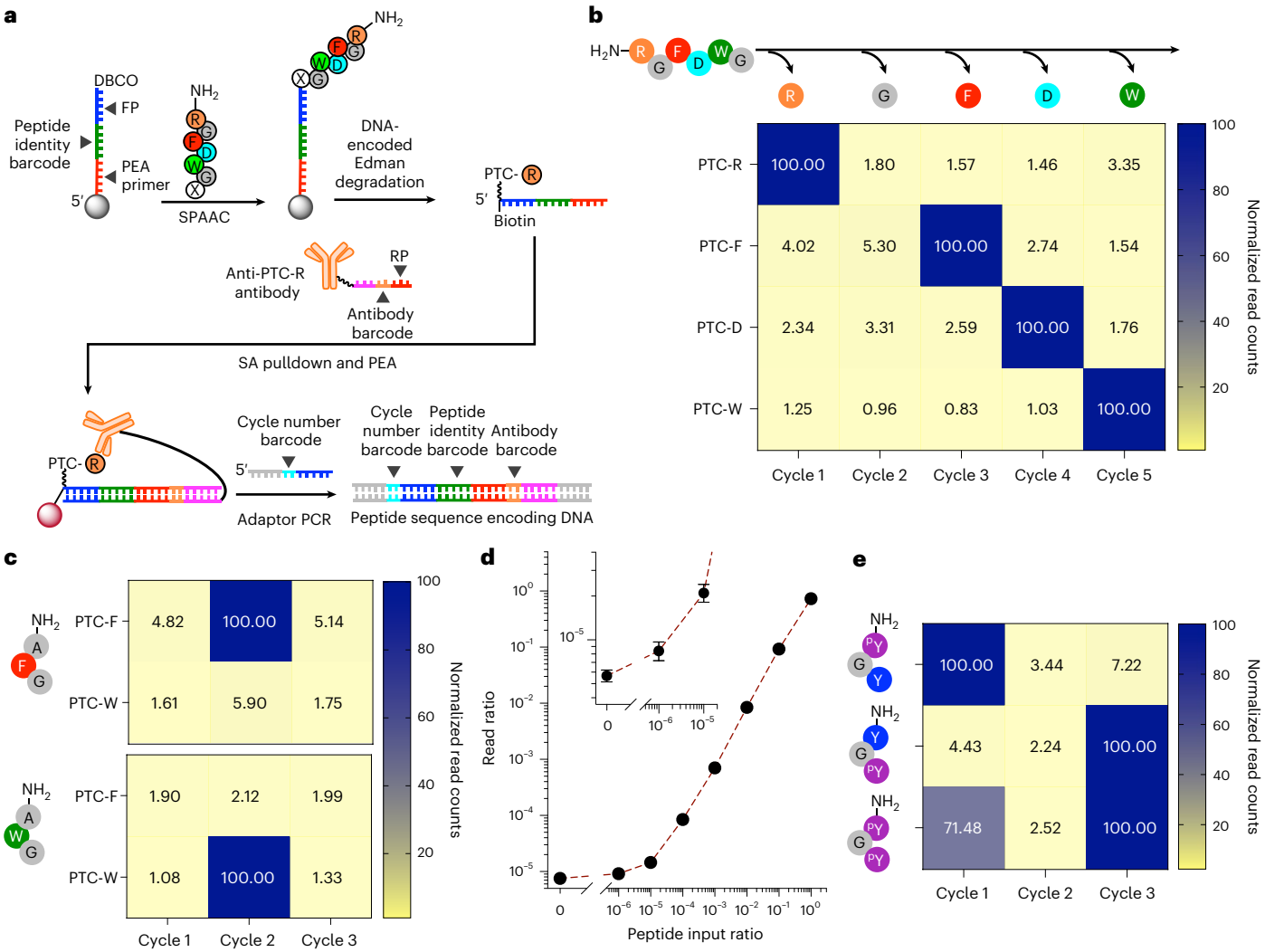


Fig. 4 | Demonstrations of ensemble peptide sequencing. a, Overview of the complete peptide reverse translation process. Carboxylic acid beads with a peptide identity barcode strand featuring a 2-nt peptide-barcoding region were prepared using EDC coupling and subsequent DNA ligation. The model peptide RGFDWGX is then linked to the DNA strand by SPAAC and subjected to our modified Edman degradation process, which yields a DNA-barcoded PTC amino acid. This is recognized by an antibody barcoded with a DNA consisting of the RP sequence, a 3-nt antibody barcode and a 6-nt PEA primer (for example, **ODN-S30** for PTC-R). The resulting PEA product is amplified by adaptor PCR, during which a 2-nt cycle number barcode is introduced through barcoded forward adaptor primers (for example, **ODN-S34** for cycle 1). The final DNA reporter sequence contains barcodes for peptide, antibody (that is, amino acid identity) and cycle number (that is, amino acid position). **b**, Heat map of normalized barcode read counts from the first five residues of RGFDWGX. The raw counts of reads bearing the same antibody barcode were normalized by setting the highest read count to 100. The correct residue at each position is shown at the top. **c**, Heat maps of normalized read counts for peptide sequences AFG (top) and AWG (bottom) after sequencing an equimolar mixture of the two peptides. The raw counts of reads bearing the same peptide identity barcode were normalized by setting the highest read count to 100. **d**, Detection of varying amounts of peptide AFG in the presence of peptide AWG. The input ratio on the *x* axis describes the proportion of peptide AFG relative to peptide AWG used during DNA-encoded Edman degradation. The output ratio on the *y* axis represents the read counts for peptide AFG relative to peptide AWG in the second sequencing cycle. At a 1:10^6 ratio, we could detect the equivalent of 0.5 zmol of DNA-barcoded PTC amino acid. This represents our estimated limit of detection in a bulk assay, calculated from the mean value of the blank sample plus 3 s.d. Inset, magnified region of the plot with input ratios ranging from 0 to 1:10^5. Data were from three independent biological replicates ($n = 3$) and are presented as the mean ± s.d. **e**, Heat maps of normalized read counts for peptide sequences ^{P}YGY, YG^{P}Y and ^{P}YG^{P}Y after analysis with PY20 antibody. Normalization for **e** was applied as in **c**.

Ensemble peptide sequencing by reverse translation

Having optimized the various steps of our procedure, we carried out ensemble sequencing on a model peptide (Fig. 4a). We prepared carboxylic acid beads with a peptide identity barcode strand featuring a 2-nt peptide-barcoding region as described above. We then immobilized the 7-aa peptide RGFDWGX (where X is azidolysine), which consists of four amino acids that can be detected by previously mentioned anti-PTC amino acid antibodies (anti-PTC-R, anti-PTC-F, anti-PTC-D and anti-PTC-W). The glycine residue serves as an exemplar of an amino acid that cannot be detected currently, which allows us to assess the true background in the absence of a specific antibody recognition event during the sequencing process. We performed five cycles of sequencing on ~100 pmol of peptide. To introduce cycle-number-specific barcodes during the adaptor PCR step, the PTC amino acids from each cycle were collected and processed separately. The DNA-barcoded PTC amino acids were enriched by SA bead pulldown, yielding a final loading density of ~10 pmol of biotinylated DNA-barcoded PTC amino acids per mg of beads. Then, PEA was carried out with a mixture of DNA-modified anti-PTC-R, anti-PTC-F, anti-PTC-D and anti-PTC-W antibodies, where each DNA strand comprised the reverse primer sequence for PCR amplification, a 3-nt antibody barcode

and a 6-nt PEA primer sequence (for example, **ODN-S30** for PTC-R). After PEA, sequencing adaptors were incorporated by PCR, during which 2-nt cycle number barcodes were introduced with forward adaptor primers (for example, **ODN-S34** for cycle 1). The PCR products were then combined for indexing and sequencing. After sequencing, barcodes encoding cycle number and antibody identity were extracted and counted and the raw counts of reads bearing the same antibody barcode were normalized by setting the highest read count to 100 (Fig. 4b). Overall, the correct amino acid assignment for each position consistently produced the highest normalized read count, producing signals that were 20–100 times above the background. While sequencing the undetectable glycine in the second cycle, we observed a normalized read count comparable to that of the background in other cycles. This demonstrates that undetectable amino acids at a given position can be clearly identified rather than being incorrectly assigned. The raw read counts were generally consistent with degradation yield and proximity extension efficiencies established previously (Supplementary Fig. 15a).

Our procedure is also capable of sensitively discriminating subtle sequence differences in mixtures of multiple peptides. Simultaneous degradation of multiple peptide species will yield a mixture of DNA-barcoded PTC amino acids that, when immobilized in close proximity, can lead to barcode crosstalk during PEA (Supplementary Fig. 16). We addressed this by reducing the loading density of DNA-barcoded PTC amino acids on the beads. At 0.5 pmol mg^{-1} loading, less than 3% of off-target proximity extension was observed (Supplementary Fig. 16c). We then set out to distinguish two peptides that differ by a single amino acid (AFGGGX and AWGGGX) by sequencing the first three amino acids in parallel. To sequence these two peptides simultaneously, we modified the initial stages of our method. Each of the two peptides was first conjugated with a unique peptide identity barcode. Afterward, these DNA-barcoded peptides were mixed and immobilized onto anchor strand-modified magnetic beads by ligation. Subsequent degradation and sample processing steps were performed similarly to the single-peptide sequencing process described above. The DNA-barcoded PTC amino acids were detected with an equimolar mixture of primer-conjugated anti-PTC-F and anti-PTC-W antibodies (30 nM each). The resulting primer extension products were then barcoded with cycle number barcodes and sequenced. After separating our analyses on the basis of peptide identity barcodes, we confirmed that the normalized read counts accurately reflected each peptide sequence, allowing for clear discrimination of the two (Fig. 4c and Supplementary Fig. 15b,c). Furthermore, our sequencing method enables the sensitive detection of peptides even when present in mixtures containing far higher concentrations of competing peptides (Fig. 4d). We sequenced a series of peptide mixtures containing different ratios of AFGGGX and AWGGGX and achieved a dynamic range of six orders of magnitude when the DNA output was sequenced on a MiSeq. At the limit of detection, we distinguished peptides at a $1{:}10^6$ ratio, equivalent to the detection of 0.5 zmol of DNA-barcoded amino acid, highlighting the potential to detect rare protein variants.

Lastly, we sequenced three peptides with different phosphorylation patterns in parallel: ^{p}YGYGGX, YG^{p}YGGX and ^{p}YG^{p}YGGX. Existing methods such as ELISA and MS often struggle to identify the correct position of a PTM given multiple candidate sites. Because our method allows sequential analysis of individual amino acids, phosphorylation sites are accurately assigned when sequenced with anti-^{p}Y (PY20, 30 nM; Fig. 4e and Supplementary Fig. 15d–f). This result highlights our platform's capacity to accurately identify both native and post-translationally modified versions of the same amino acid within a peptide sequence.

Single-molecule peptide sequencing through reverse translation

Our sequencing method establishes a framework for single-molecule peptide sequencing. Unlike methods relying on direct physical observation[13–18,22], we achieve single-molecule resolution through unique molecular barcoding. By conjugating each peptide of interest with a peptide identity barcode DNA containing a UMI, every original peptide is uniquely indexed and addressed at the single-molecule level. After DNA sequencing, reads sharing the same UMI are grouped and ordered by cycle number and antibody barcodes to reconstruct peptide sequences. This strategy eliminates the reliance on predefined identity barcodes used during ensemble peptide sequencing, leveraging the vast UMI sequence space to ensure each peptide is uniquely barcoded.

To obtain high sequence coverage, it is necessary to detect DNA reporters generated from the same peptide across multiple degradation cycles, which only becomes probable when the majority of DNA reporters are sequenced. However, the number of DNA reporters generated by the current DNA-encoded Edman degradation procedure far exceeds the throughput of current DNA sequencers. To mitigate this mismatch, we adopted a strategy inspired by the carrier proteome approach commonly used in MS-based single-cell proteomics (Fig. 5a)[48]. We first synthesized beads bearing a peptide of interest (FWFWGGGX) barcoded with a peptide identity barcode DNA containing a 30-nt randomized region. We then spiked beads bearing ~0.5 fmol of this peptide into beads bearing ~100 pmol of carrier peptide (GGGGGGX) conjugated to a DNA sequence that will not be amplified or sequenced. The addition of this carrier peptide allows the sequencing of small amounts of peptides with the current DNA-encoded Edman degradation procedure and minimizes the loss of DNA-barcoded amino acids generated from the peptides of interest.

The resulting DNA output was sequenced on a NovaSeq X Plus, which has sufficient throughput to sequence all DNA output at an average sequencing depth of 20. The reads were grouped by UMI to yield single-molecule peptide sequencing results. In total, 2.94×10^8 UMIs were detected with an average read depth of 23.8 during the first sequencing cycle (Supplementary Fig. 17), which is consistent with the amount of peptide input. As we would predict from the specificity of our PEA reaction, the average accuracy of amino acid assignment was 98% (Fig. 5b); 91.9% of peptide sequencing reads contained no misassigned amino acids (Fig. 5c). Of all peptide sequencing reads, 4.8×10^7 (17.9%) achieved full sequence coverage and 66.5% contained reads of multiple amino acids within the same peptide (Fig. 5d).

The UMI-resolved sequencing results (Supplementary Table 4) allowed us to track each individual molecule and revealed two types of mechanisms that can contribute to incomplete sequence coverage: truncations and deletions (Fig. 5e). Truncations are the premature termination of sequencing reactions caused by either irreversible capping of the N terminus or loss of beads or peptides during degradation, resulting in the complete loss of UMIs for subsequent sequencing cycles. Given that the fourth amino acid was detected for 44.4% of UMIs (Fig. 5f) and the average stepwise degradation yield was 95.4%, we estimated an average stepwise peptide recovery rate of 80.1% ± 4.8%. As we did not observe cleavage of DNA or the triazole linkage during degradation, loss of beads because of adsorption to surfaces and aggregation is likely to be the main contributor to peptide loss during sequencing, in addition to irreversible capping of the N terminus by oxidative desulfurization[45,46]. On the other hand, deletions occur when the conversion from DNA-barcoded amino acids to DNA reporters is incomplete, which do not lead to sequencing termination. This can be caused by the loss of DNA-barcoded amino acids during and before SA pulldown or by inefficient proximity extension. We observed deletions at a steady frequency in every degradation cycle. To calculate the average rate of deletion, we selected a subset of sequences in which the first and fourth amino acids were detected, thereby disregarding the impact of truncations. We determined an average deletion rate of 22.8% per cycle (Fig. 5g). Overall, these results highlight the capacity of our platform to sequence peptides with single-molecule sensitivity.

https://doi.org/10.1038/s41587-026-03061-z

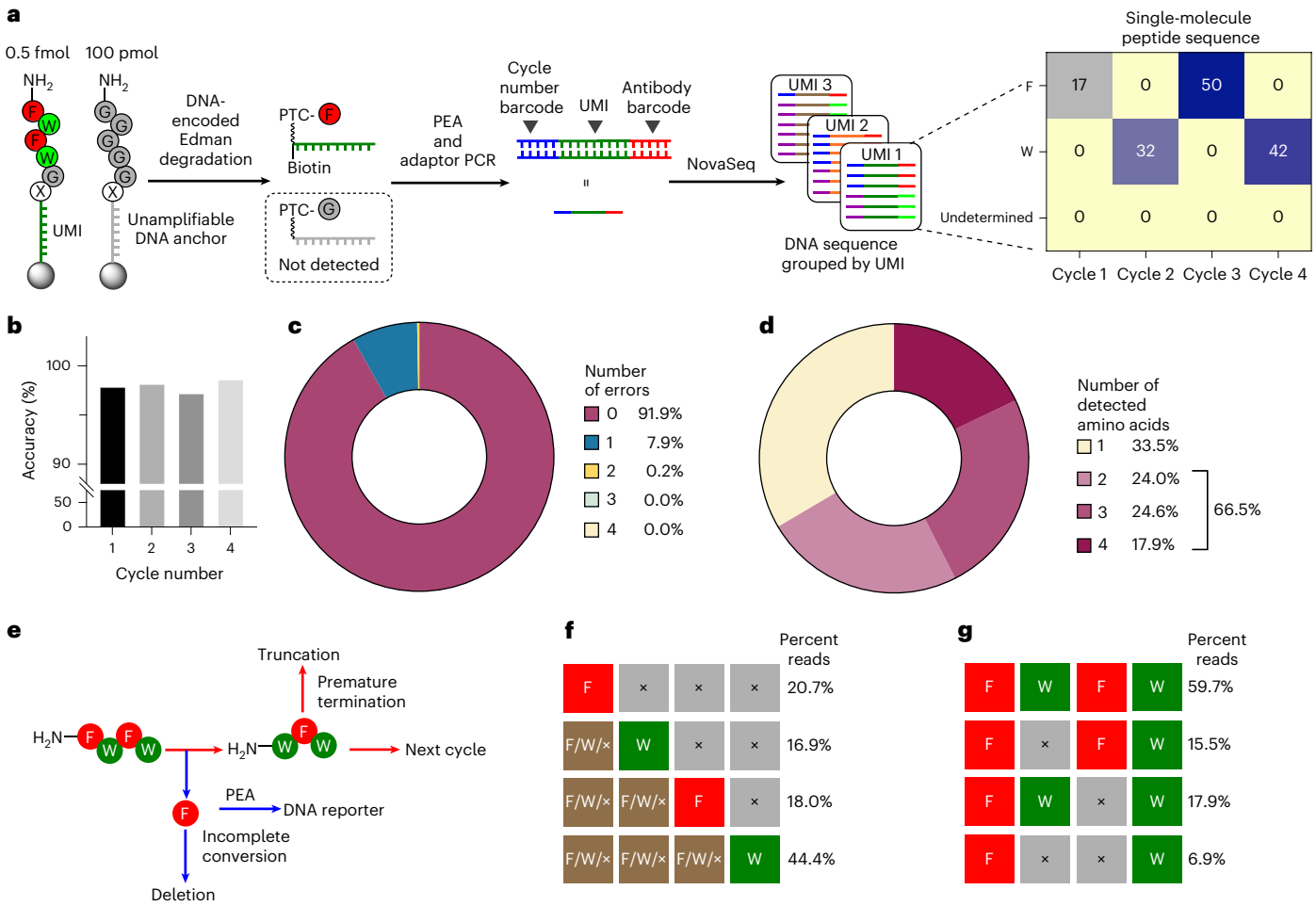


Fig. 5 | Single-molecule peptide sequencing. a, Workflow of single-molecule peptide sequencing through reverse translation. Beads bearing the peptide of interest (0.5 fmol) were mixed with beads bearing a carrier peptide (100 pmol). The carrier peptide was introduced to minimize the loss of the peptide of interest during DNA-encoded Edman degradation. The DNA-barcoded amino acids generated from the carrier peptide lacked biotin and primer-binding sites required for primer extension and were, therefore, not detectable by PEA. After amplification and sequencing on NovaSeq, the sequencing reads were grouped by UMI to yield single-molecule peptide sequencing results. Representative single-molecule sequencing results from the peptide bearing the UMI GGTGTACGTAGCTGTCCCGGATATCAGTGT are shown, where the raw read numbers are plotted in a heat map. Sequencing reads containing nonpredefined amino acid barcodes are categorized as 'undetermined'. Primer sites are omitted for clarity. **b**, Sequencing accuracy was maintained throughout the sequencing process. The accuracy for cycles 1–4 was 97.87%, 98.16%, 97.21% and 98.62%, respectively. **c**, Distribution of the number of errors in single-molecule peptide sequencing reads. No error was detected in 91.9% of reads. **d**, Distribution of the number of amino acids detected in single-molecule peptide sequencing reads. Multiple amino acids were detected in 66.5% of peptide sequences. **e**, Incomplete sequencing reads are caused by two distinct mechanisms: truncation and deletion. Truncations result from premature termination of the DNA-encoded Edman degradation reaction, leading to the complete loss of peptides for later cycles. Deletions are caused by incomplete conversion or loss of DNA reporters generated by PEA, which occurs at a constant rate during each sequencing cycle. **f**, Distribution of truncation positions. Undetected amino acids are denoted by '×'. The truncation that occurred in the first cycle was not detectable by our method. The rate of peptide recovery for cycles 2–4 was 83.1%, 82.5% and 74.6%, respectively. **g**, Deletions in sequencing reads beginning with phenylalanine at the first position and ending with tryptophan at the fourth position. The reads were sorted by whether amino acids were detected at the second and third positions. The average rate of deletion was estimated to be 22.8% per cycle.

Discussion

Our reverse translation strategy for peptide sequencing offers several advantages over other methods described to date. The Edman degradation-based method used here excises one amino acid at a time, enabling each to be analyzed independently outside the larger context of the peptide. This enables peptide sequencing at single-amino-acid resolution and eliminates the potential for interference from adjacent amino acids or PTMs. The released amino acids are subsequently detected by antibodies, which can, in principle, be directly generated against the full spectrum of proteinogenic amino acids. In this work, we demonstrated such capabilities with four antibodies that exhibit good affinity and specificity for their natural amino acid targets. Furthermore, we demonstrated that we can use PTM-specific antibodies to accurately identify both the presence and the position of their corresponding PTMs. Our method transforms peptide sequences into DNA output, allowing the amplification of peptide sequence information from peptides. This capability, when combined with the use of UMIs, enables our method to sequence peptides with single-molecule sensitivity. Furthermore, by simplifying peptide sequencing to a DNA sequencing problem, our approach is highly scalable and can leverage well-established infrastructure for high-throughput DNA sequencing platforms. With recent advancements in DNA sequencing technologies, such as sequencing by expansion, the throughput of our method can be further increased (Supplementary Information)[49].

Several areas of improvement are needed to sequence native peptides from biological samples. First, our method requires the conjugation of DNA to the C terminus of peptides. While several

strategies for selective labeling of the C-terminal carboxylic acid have been reported, their efficiency and substrate scope are still somewhat limited[13,50–54]. The development of more efficient and generalizable C-terminal-specific modification reactions will greatly expand the scope of peptides that can be sequenced by our method.

Additionally, while our method is designed to operate in a manner analogous to bottom-up proteomics, further improvements in the efficiency of sequencing process are needed for the sequencing of proteotypic peptides generated from digestion of native proteins with higher sequence coverage. The sequencing length of conventional Edman degradation is limited by incomplete reactions, which leads to the loss of synchronization[55]. Our approach can overcome this limitation by performing UMI-enabled single-molecule peptide sequencing in conjunction with the incorporation of cycle number barcodes during DNA-encoded Edman degradation. In addition, truncations arising from bead loss and deletions resulting from the loss of DNA-barcoded PTC amino acids currently limit sequencing length and coverage. To mitigate these effects, we anticipate that the present bead-based sequencing pipeline can be translated into an automated, miniaturized flow-cell-based sequencing platform. Lastly, further optimization of Edman degradation efficiency and primer extension will reduce truncation and deletion rates, respectively, and will be essential for achieving robust and efficient sequencing.

Lastly, the scope, sensitivity and accuracy of our sequencing method are determined by the performance of our PTC amino-acid-specific antibodies. In some applications, such as protein identification, detecting all amino acids is unnecessary and matching a partial sequence to a reference protein database is sufficient[13,22,56]. For example, using the antibodies presented in the present work, one could identify 74% of the human proteome by sequencing the first ten amino acids of peptides generated from Lys-C digestion (Supplementary Fig. 18). Beyond protein identification, this platform offers an advantage for profiling PTMs. By using commercially available anti-PTM antibodies, it allows for the characterization of proteins with complex, multisite PTM patterns that are challenging to resolve using traditional methods. Looking ahead, achieving comprehensive de novo protein sequencing with single-molecule sensitivity will ultimately require a complete set of amino-acid-specific recognizers. We anticipate that the repertoire of recognizers can be systematically expanded through a combination of antibody generation through immunization and alternative binder discovery through aptamer selection[57], protein directed evolution[58,59] and computational design[60,61]. Lastly, as multiple amino acid recognition modalities are introduced, careful optimization of PEA would be required to minimize sequence-dependent biases.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-026-03061-z.

References

1. Pollen, A. A. et al. Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat. Biotechnol.* **32**, 1053–1058 (2014).
2. Wang, N. et al. Single-cell microRNA-mRNA co-sequencing reveals non-genetic heterogeneity and mechanisms of microRNA regulation. *Nat. Commun.* **10**, 95 (2019).
3. Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610–620 (2015).
4. Sonneveld, S., Verhagen, B. M. P. & Tanenbaum, M. E. Heterogeneity in mRNA Translation. *Trends Cell Biol.* **30**, 606–618 (2020).
5. Smith, L. M. et al. The Human Proteoform Project: defining the human proteome. *Sci. Adv.* **7**, eabk0734 (2021).
6. Tajik, M., Baharfar, M. & Donald, W. A. Single-cell mass spectrometry. *Trends Biotechnol.* **40**, 1374–1392 (2022).
7. Ye, Z. et al. One-Tip enables comprehensive proteome coverage in minimal cells and single zygotes. *Nat. Commun.* **15**, 2474 (2024).
8. Guzman, U. H. et al. Ultra-fast label-free quantification and comprehensive proteome coverage with narrow-window data-independent acquisition. *Nat. Biotechnol.* 1–12 https://doi.org/10.1038/s41587-023-02099-7 (2024).
9. Huffman, R. G. et al. Prioritized mass spectrometry increases the depth, sensitivity and data completeness of single-cell proteomics. *Nat. Methods* **20**, 714–722 (2023).
10. MacCoss, M. J. et al. Sampling the proteome by emerging single-molecule and mass spectrometry methods. *Nat. Methods* **20**, 339–346 (2023).
11. Alfaro, J. A. et al. The emerging landscape of single-molecule protein sequencing technologies. *Nat. Methods* **18**, 604–617 (2021).
12. Restrepo-Pérez, L., Joo, C. & Dekker, C. Paving the way to single-molecule protein sequencing. *Nat. Nanotechnol.* **13**, 786–796 (2018).
13. Reed, B. D. et al. Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device. *Science* **378**, 186–192 (2022).
14. Martin-Baniandres, P. et al. Enzyme-less nanopore detection of post-translational modifications within long polypeptides. *Nat. Nanotechnol.* **18**, 1335–1340 (2023).
15. Yu, L. et al. Unidirectional single-file transport of full-length proteins through a nanopore. *Nat. Biotechnol.* **41**, 1130–1139 (2023).
16. Brinkerhoff, H., Kang, A. S. W., Liu, J., Aksimentiev, A. & Dekker, C. Multiple rereads of single proteins at single-amino acid resolution using nanopores. *Science* **374**, 1509–1513 (2021).
17. Nova, I. C. et al. Detection of phosphorylation post-translational modifications along single peptides with nanopores. *Nat. Biotechnol.* **42**, 710–714 (2023).
18. Motone, K. et al. Multi-pass, single-molecule nanopore reading of long protein strands. *Nature* **633**, 662–669 (2024).
19. Ouldali, H. et al. Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. *Nat. Biotechnol.* **38**, 176–181 (2020).
20. Wang, K. et al. Unambiguous discrimination of all 20 proteinogenic amino acids and their modifications by nanopore. *Nat. Methods* **21**, 92–101 (2024).
21. Zhang, M. et al. Real-time detection of 20 amino acids and discrimination of pathologically relevant peptides with functionalized nanopore. *Nat. Methods* **21**, 609–618 (2024).
22. Swaminathan, J. et al. Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures. *Nat. Biotechnol.* **36**, 1076–1082 (2018).
23. Boutureira, O. & Bernardes, G. J. L. Advances in chemical protein modification. *Chem. Rev.* **115**, 2174–2195 (2015).
24. Szijj, P. A., Kostadinova, K. A., Spears, R. J. & Chudasama, V. Tyrosine bioconjugation—an emergent alternative. *Org. Biomol. Chem.* **18**, 9018–9028 (2020).
25. Choi, W. S. et al. Structural basis for the recognition of N-end rule substrates by the UBR box of ubiquitin ligases. *Nat. Struct. Mol. Biol.* **17**, 1175–1181 (2010).
26. 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).

27. Edman, P., Högfeldt, E., Sillén, L. G. & Kinell, P.-O. Method for determination of the amino acid sequence in peptides. *Acta Chem. Scand.* **4**, 283–293 (1950).
28. McGregor, L. M., Gorin, D. J., Dumelin, C. E. & Liu, D. R. Interaction-dependent PCR: identification of ligand–target pairs from libraries of ligands and libraries of targets in a single solution-phase experiment. *J. Am. Chem. Soc.* **132**, 15522–15524 (2010).
29. Zhang, P. et al. Highly sensitive serum protein analysis using magnetic bead-based proximity extension assay. *Anal. Chem.* **94**, 12481–12489 (2022).
30. Zoltewicz, J. A., Clark, D. F., Sharpless, T. W. & Grahe, G. Kinetics and mechanism of the acid-catalyzed hydrolysis of some purine nucleosides. *J. Am. Chem. Soc.* **92**, 1741–1750 (1970).
31. Hong, J. M. et al. ProtSeq: toward high-throughput, single-molecule protein sequencing via amino acid conversion into DNA barcodes. *iScience* **25**, 103586 (2022).
32. 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).
33. 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).
34. Mitchell, M. J., Hirschowitz, W., Rastinejad, F. & Lu, P. Boron trifluoride–methanol complex as a non-depurinating detritylating agent in DNA synthesis. *Nucleic Acids Res.* **18**, 5321 (1990).
35. Seela, F., Menkhoff, S. & Behrendt, S. Furanoside–pyranoside isomerization of tubercidin and its 2′-deoxy derivatives: influence of nucleobase and sugar structure on the proton-catalysed reaction. *J. Chem. Soc. Perkin Trans.* **2**, 525–530 (1986).
36. Potowski, M. et al. Chemically stabilized DNA barcodes for DNA-encoded chemistry. *Angew. Chem. Int. Ed.* **60**, 19744–19749 (2021).
37. 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 (1993).
38. Matsudaira, P. (ed.) *A Practical Guide to Protein and Peptide Purification for Microsequencing* (Academic Press, 1993).
39. Rydberg, P., Lüning, B., Wachtmeister, C. A., Eriksson, L. & Törnqvist, M. Applicability of a modified Edman procedure for measurement of protein adducts: mechanisms of formation and degradation of phenylthiohydantoins. *Chem. Res. Toxicol.* **15**, 570–581 (2002).
40. Flood, D. T. et al. Expanding reactivity in DNA-encoded library synthesis via reversible binding of DNA to an inert quaternary ammonium support. *J. Am. Chem. Soc.* **141**, 9998–10006 (2019).
41. MacConnell, A. B., McEnaney, P. J., Cavett, V. J. & Paegel, B. M. DNA-encoded solid-phase synthesis: encoding language design and complex oligomer library synthesis. *ACS Comb. Sci.* **17**, 518–534 (2015).
42. Halpin, D. R., Lee, J. A., Wrenn, S. J. & Harbury, P. B. DNA display III. Solid-phase organic synthesis on unprotected DNA. *PLoS Biol.* **2**, e175 (2004).
43. Wang, Z., Li, D., Tian, X. & Zhang, C. A copper-free and enzyme-free click chemistry-mediated single quantum dot nanosensor for accurate detection of microRNAs in cancer cells and tissues. *Chem. Sci.* **12**, 10426–10435 (2021).
44. Nakano, S. & Sugimoto, N. The structural stability and catalytic activity of DNA and RNA oligonucleotides in the presence of organic solvents. *Biophys. Rev.* **8**, 11–23 (2016).
45. Niall, H. D. Automated Edman degradation: the protein sequenator. *Methods Enzymol.* **27**, 942–1010 (1973).
46. Laursen, R. A. Solid-phase Edman degradation. *Eur. J. Biochem.* **20**, 89–102 (1971).
47. Fredriksson, S. et al. Protein detection using proximity-dependent DNA ligation assays. *Nat. Biotechnol.* **20**, 473–477 (2002).
48. Budnik, B., Levy, E., Harmange, G. & Slavov, N. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **19**, 161 (2018).
49. Kokoris, M. et al. Sequencing by expansion (SBX)—a novel, high-throughput single-molecule sequencing technology. Preprint at *bioRxiv* https://doi.org/10.1101/2025.02.19.639056 (2025).
50. Bloom, S. et al. Decarboxylative alkylation for site-selective bioconjugation of native proteins via oxidation potentials. *Nat. Chem.* **10**, 205–211 (2018).
51. Zhang, L. et al. Photoredox-catalyzed decarboxylative C-terminal differentiation for bulk- and single-molecule proteomics. *ACS Chem. Biol.* **16**, 2595–2603 (2021).
52. Rehm, F. B. H. et al. Enzymatic C-terminal protein engineering with amines. *J. Am. Chem. Soc.* **143**, 19498–19504 (2021).
53. Antos, J. M. et al. Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J. Am. Chem. Soc.* **131**, 10800–10801 (2009).
54. Boga, S. B. et al. Site-selective synthesis of insulin azides and bioconjugates. *Bioconjug. Chem.* **30**, 1127–1132 (2019).
55. Edman, P. & Begg, G. A protein sequenator. *Eur. J. Biochem.* **1**, 80–91 (1967).
56. Swaminathan, J., Boulgakov, A. A. & Marcotte, E. M. A theoretical justification for single molecule peptide sequencing. *PLoS Comput. Biol.* **11**, e1004080 (2015).
57. Yang, K.-A. et al. Recognition and sensing of low-epitope targets via ternary complexes with oligonucleotides and synthetic receptors. *Nat. Chem.* **6**, 1003–1008 (2014).
58. Boder, E. T., Midelfort, K. S. & Wittrup, K. D. Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc. Natl Acad. Sci. USA* **97**, 10701–10705 (2000).
59. Javanpour, A. A. & Liu, C. C. Evolving small-molecule biosensors with improved performance and reprogrammed ligand preference using OrthoRep. *ACS Synth. Biol.* **10**, 2705–2714 (2021).
60. Lee, G. R. et al. Small-molecule binding and sensing with a designed protein family. Preprint at *bioRxiv* https://doi.org/10.1101/2023.11.01.565201 (2023).
61. An, L. et al. Binding and sensing diverse small molecules using shape-complementary pseudocycles. *Science* **385**, 276–282 (2024).

Methods

Solid-phase synthesis of oligonucleotides

Oligonucleotides were synthesized on an Applied Biosystems Expedite 8909 nucleic acid synthesis system using the procedures recommended by the manufacturer and standard β-cyanoethyl phosphoramidites. Detailed procedures for deprotection, desalting and purification are provided in the Supplementary Information.

Immobilization of oligonucleotides on magnetic carboxylic acid beads

Dynabeads MyOne carboxylic acid (5 mg, 500 µl) was washed once with 500 µl of 10 mM NaOH and five times with 500 µl of nuclease-free water. The washed beads were resuspended in a 150-µl reaction mixture containing 200 mM NaCl, 200 µM 5′-amino-modified oligonucleotide, 1 mM imidazole, 50% v/v DMSO and 250 mM EDC. Beads were mixed well with reagents, vortexed, sonicated and incubated overnight on a rotator at room temperature. After coupling the oligonucleotides, the beads were washed three times with cold 100 mM MES buffer (pH 4.7). The unconjugated carboxylic acid groups on the beads were activated with 250 mM EDC and 100 mM NHS in MES buffer at room temperature for 30 min. Activated beads were washed three times with cold PBS buffer and passivated with 20 mM methyl-PEG12-amine in PBS buffer for 2 h at room temperature. The passivated beads were washed three times with 500 µl of Tris–Tween buffer (100 mM Tris-HCl and 0.01% v/v Tween-20, pH 7.5) and stored in 500 µL of PBS–Tween (PBST) buffer (1× PBS pH 7.4 and 0.01% v/v Tween-20).

Investigation of DNA stability during BF_3 etherate treatment

Oligonucleotide **ODN-1** or **ODN-S1** in PBST buffer (10 µM, 100 µl) was annealed to Dynabeads MyOne carboxylic acid beads functionalized with amino-modified oligonucleotide **ODN-S2** for 10 min at room temperature. After hybridization, the beads were washed three times with 100 µl of PBST buffer to remove excess DNA and then washed five times with 100 µl of PBST plus acetonitrile (50% v/v) and five times with anhydrous acetonitrile. The beads were dried under vacuum in a desiccator for 30 min. The dried beads were incubated in 100 µl of 40 mM BF_3 etherate in anhydrous acetonitrile at 50 °C for up to 4 h. After incubation, the beads were washed five times with 100 µl of anhydrous acetonitrile and once with 50 µl of 20 mM NaOH to neutralize excess BF_3 etherate. The washed beads were then incubated with 100 µl of 20 mM NaOH at room temperature for 10 min to release the oligonucleotides. The procedure is illustrated in Supplementary Fig. 1a. The oligonucleotides were then purified by ethanol precipitation (0.3 M sodium acetate in 70% ethanol), resuspended in nuclease-free water and analyzed on an Agilent 1260 Infinity II high-performance LC (HPLC) instrument equipped with a Waters XBridge oligonucleotide BEH C18 column (130 Å, 2.5 µm, 4.6 mm × 50 mm). The mobile phases consisted of 100 mM triethylammonium acetate (A) and acetonitrile (B). The gradient began with 5% B from $t = 0$ to 2 min, ramped to 30% B linearly over 13 min, ramped to 97% B over 5 min and was then held at 97% B from $t = 20$ to 24 min. Mobile phase B was reduced to 5% linearly over 2 min and kept at 5% for 8 min. The flow rate was 1 ml min^{-1} and the column was maintained at ambient temperature. The oligonucleotides were monitored by the absorbance at 260 nm and 284 nm. A similar procedure was performed with oligonucleotides **ODN-S3** and **ODN-S4** hybridized to beads modified with amino-modified oligonucleotide **ODN-S5**.

Primer extension with c^7dATP and c^7dGTP

A typical 20-µl primer extension reaction contained 1.5 µM template (**ODN-S6**) and 1 µM Cy3-labeled primer (**ODN-S7** or **ODN-S8**). We tested different polymerases using the following conditions (conditions are shown for **ODN-S7**; for **ODN-S8** reactions, c^7dATP and c^7dGTP were replaced with dATP and dGTP):

Sequenase version 2.0 DNA polymerase. A mixture of 0.5 U of polymerase, 300 µM c^7dATP, 300 µM dTTP, 300 µM c^7dGTP, 300 µM dCTP, 40 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$ and 5 mM DTT was incubated for 15 or 30 min at 37 °C.

Klenow fragment (exo–). A mixture of 1 U of polymerase, 300 µM c^7dATP, 300 µM dTTP, 300 µM c^7dGTP, 300 µM dCTP, 10 mM Tris-HCl pH 7.9, 50 mM NaCl, 10 mM $MgCl_2$ and 1 mM DTT was incubated for 15 or 30 min at 37 °C.

***Bst* 3.0 DNA polymerase.** A mixture of 1 U of polymerase, 300 µM c^7dATP, 300 µM dTTP, 300 µM c^7dGTP, 300 µM dCTP, 20 mM Tris-HCl pH 8.8, 10 mM $(NH_4)_2SO_4$, 150 mM KCl, 2 mM $MgSO_4$, and 0.1% Tween-20 was incubated for 15 or 30 min at 55 °C.

After primer extension, the DNA was cleaned by ethanol precipitation (0.3 M sodium acetate, 70% ethanol) and analyzed by denaturing PAGE. The gel was imaged on an Amersham Typhoon RGB imager.

DNA-encoded Edman degradation

We began by preparing oligonucleotide-modified magnetic beads. The amino-modified oligonucleotide **ODN-S10** was immobilized onto 100 µl (1 mg) of Dynabeads MyOne carboxylic acid beads. We then ligated 10 µM 5′-phosphorylated, DBCO-modified DNA **ODN-S11** in a 100-µl reaction containing 10 µM splint DNA **ODN-S12**, T4 ligase (2,000 U), 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM ATP and 0.01% Tween-20 at room temperature for 2 h. After ligation, splint DNA was removed by incubation with 100 µl of 20 mM NaOH for 10 min, after which the beads were washed three times with PBST buffer.

Next, the beads were incubated with 10 mM peptide FGGGGGX (where X is azidolysine) dissolved in 100 µl of PBST at room temperature for 2 h. After the completion of the SPAAC reaction, the beads were washed with PBST buffer three times to remove unconjugated peptide and then washed with 100 µl of PITC conjugation solution (pyridine, ethanol and water, 2:1:1 v/v). Azide-modified PITC (**1**) (200 mM, 100 µL) in conjugation solution was added to the beads and the peptides were allowed to react for 30 min at 50 °C. The beads were then washed three times with 100 µl of conjugation solution and three times with 100 µl of PBST-Mg buffer (PBST plus 2 mM $MgCl_2$, pH 7.4). Next, proximity SPAAC was carried out by incubating the washed beads with 20 µM primer **ODN-S13** in 100 µl of PBST-Mg buffer for 10 min at room temperature.

For the primer extension step, the beads were washed three times with 100 µl of primer extension buffer (50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$ and 0.01% v/v Tween-20, pH 7.5). We then added 60 µl of primer extension mixture to the beads, which included Klenow fragment (exo−; 8 U), 300 µM c^7dATP, 300 µM dTTP, 300 µM c^7dGTP, 300 µM dCTP, 50 mM NaCl, 10 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$ and 0.01% v/v Tween-20, with incubation for 1 h at 37 °C. The beads were then washed three times with 100 µl of PBST buffer to remove excess reagents.

For the cleavage step, the beads were washed five times with 100 µl of 50% acetonitrile (v/v) in PBST and five times with anhydrous acetonitrile. The beads were dried under vacuum in a desiccator for 30 min and then incubated in 100 µL of 40 mM BF_3 etherate in anhydrous acetonitrile at 50 °C for 5 min. The beads were then washed five times with 100 µl of anhydrous acetonitrile and once with 50 µl of basic conversion solution (0.25% v/v NH_4OH, 200 µM DTT) to neutralize excess BF_3 etherate. The washed beads were then incubated with 100 µl of basic conversion solution at 50 °C for 10 min to release the DNA-barcoded PTC amino acid. The DNA-barcoded PTC amino acid was cleaned by ethanol precipitation (0.3 M sodium acetate in 70% ethanol). The DNA-barcoded PTC amino acid was resuspended in HPLC running buffer (10 mM ammonium acetate in 80% acetonitrile) and analyzed by HPLC–MS, as described below. The beads were washed three times with PBST buffer and stored in PBST buffer.

Isolation and analysis of intermediates of DNA-encoded Edman degradation

Amino-modified anchor DNA containing a disulfide linker and a Cy3 modification (**ODN-S15**) was immobilized onto 500 µl (5 mg) of Dynabeads MyOne carboxylic acid beads as described above. A 50-µl aliquot of these beads was treated with 50 µl of 10 mM TCEP in PBST buffer at room temperature for 30 min to cleave the oligonucleotide off the bead surface. The released oligonucleotide was purified by ethanol precipitation (0.3 M sodium acetate in 70% ethanol). The rest of the beads were subjected to the procedure described above, with 50-µl aliquots taken after DNA ligation, conjugation of peptide, conjugation of azide-modified PITC and conjugation of DBCO-modified primer. All aliquots were reduced with 50 µl of 10 mM TCEP and the released intermediates were cleaned by ethanol precipitation (0.3 M sodium acetate in 70% ethanol). All intermediates were analyzed by denaturing PAGE and the gel was imaged by an Amersham Typhoon RGB imager.

Time-course experiment of proximity SPAAC for peptides of varying length and rigidity

Dynabeads MyOne carboxylic acid beads (5 mg, 500 µl) were modified with oligonucleotides as described in the DNA-encoded Edman degradation procedure. Beads were divided into 100-µl aliquots and conjugated to peptides ranging from 10–30 aa with varying rigidity (Supplementary Table 5) as described previously, after which the peptides were modified with azide-modified PITC and subjected to proximity SPAAC with **ODN-S16**. As a no-proximity negative control, we performed a reaction using beads modified with the shortest peptide GGGGSGGGGSX and **ODN-S17**. The reactions were incubated at room temperature for 2 h, with 1-µl aliquots collected at 10 min, 30 min, 1 h and 2 h. These aliquots were washed with 100 µl of 20 mM NaOH for 10 min to remove unreacted DBCO-modified DNA, then resuspended in PBST and analyzed on a NovoCyte Quanteon flow cytometer.

Quantification of degradation yield

After completion of the primer conjugation by proximity SPAAC during the Edman degradation procedure above, we collected a 0.2-µl aliquot from the reaction mixture and incubated it with 20 µl of 10 µM Cy3-labeled complementary strand of the primer (**ODN-S18**) for 10 min at room temperature. The DBCO-modified primer (**ODN-S13**) contains a dT5 region that facilitates the hybridization of **ODN-S12** through toehold-mediated strand displacement. The beads were washed three times with PBST buffer to remove excess Cy3-labeled strand and then analyzed by a NovoCyte Quanteon flow cytometer. The mean fluorescence signal in PE channel was recorded as the mean PE-A concentration before degradation. Similarly, after the release of the DNA-tagged PTC amino acid, we removed a 0.2-µl aliquot from the reaction mixture and incubated with **ODN-S18**. The mean fluorescence signal in the PE channel was designated as the mean PE-A concentration after degradation. The percentage yield of the cleavage reaction was determined using Eq. (1). The procedure is illustrated in Supplementary Fig. 6a.

$$\text{Degradation yield} = 100\% - \frac{[\text{Mean PE-A after degradation}]}{[\text{Mean PE-A before degradation}]} \times 100\%. \quad (1)$$

LC–MS analysis of DNA-barcoded PTC amino acids

The DNA-barcoded PTC amino acids (10 µM, 5 µl) were analyzed by LC–electrospray ionization MS on a Waters Acquity ultra-HPLC (UPLC) instrument and a Thermo Exploris 240 BioPharma Orbitrap MS instrument. The UPLC was equipped with a Waters Acquity UPLC BEH Amide column (130 Å, 1.7 µm, 2.1 mm × 100 mm). The mobile phase consisted of 10 mM ammonium acetate in 80% acetonitrile (A) and 10 mM ammonium acetate in 25% acetonitrile (B). The gradient began with 10% B from $t = 0$ to 1 min, ramped to 50% B linearly over 2 min, ramped to 90% B over 7 min and was then held at 90% B from $t = 10$ to 14 min. The flow rate was 0.3 ml min^{-1} and the column was maintained at 60 °C. Full-scan MS spectra were acquired over the mass range 570–4,000 m/z with 120,000 resolution. The molecular weight of the DNA-barcoded PTC amino acids was obtained after deconvolution using Protein Metrics Suite software.

Preparation of PTC amino acid-conjugated BSA

BSA (1 mg) was dissolved in 270 µl of 100 mM sodium bicarbonate buffer (pH 8.2). TFP ester-PEG4-DBCO (70 mM, 30 µL) in anhydrous DMSO was added to the BSA solution and the reaction was incubated for 2 h on a rotator at room temperature. After conjugation, DBCO-modified BSA was purified by buffer exchange to 1× PBS using a Zeba spin desalting column (7-kDa molecular weight cutoff (MWCO)), yielding a final concentration of ~3 mg ml^{-1}. The degree of labeling was determined by SDS–PAGE and ultraviolet–visible light (UV–Vis) spectroscopy (Thermo Fisher Scientific Nanodrop 2000). PTC-modified amino acid azide (100 mM, 15 µl) in DMSO was added to 300 µL (~1 mg) of DBCO-modified BSA in 1× PBS and incubated for 1 h on a rotator at room temperature. Another 15-µl aliquot of 100 mM PTC-modified amino acid azide in DMSO was then added and the reaction was incubated for an additional hour. After conjugation, BSA conjugated to PTC amino acid was purified by buffer exchange to 1× PBS using a Zeba spin desalting column (7-kDa MWCO), yielding a final concentration of ~3 mg ml^{-1}. The degree of labeling was again determined as described above. Completion of the click reaction was marked by the disappearance of DBCO absorbance at 310 nm.

Immunization, hybridoma generation and antibody purification

Immunization and hybridoma generation was carried out by Antibody Solutions. All animal experiments were approved by the Institutional Animal Care and Use Committee of Antibody Solutions (protocol IP-01). Antibody Solutions holds an National Institutes of Health Office of Laboratory Animal Welfare Public Health Service assurance (D18-01024). Each BSA-conjugated PTC amino acid was administered to a cohort of three female CD1 mice as twice-weekly subcutaneous injections to a single hind footpad for 4 weeks. All injections consisted of 10 µL (10 µg) of BSA-conjugated PTC amino acid in 1× PBS buffer mixed with an equal volume of Sigma adjuvant system. On day 21, serum from each mouse was evaluated by ELISA for binding to its corresponding target. In short, ELISA plates were coated with each respective target, blocked and then incubated with serial dilutions of each serum sample. Plates were washed and then incubated with horseradish-peroxidase-conjugated mouse IgG-specific secondary antibody, followed by another wash step, development with TMB substrate and finally absorbance measurement at 450 nm. The specificity of each serum sample was evaluated in parallel against duplicate wells coated with the respective target antigen. Before the serum samples were added to the ELISA plates, each sample was preincubated with an excess amount of other BSA-conjugated PTC amino acids to block antibodies to the BSA carrier or conserved moieties (for example, the PEG_4 linker in azide-modified PITC and DBCO). On day 28, selected mice for each PTC amino acid received a final boost injection. Then, 3 days after the final injection, popliteal and inguinal lymph nodes were isolated from selected mice within each cohort and teased to isolate lymphocytes. These were then mixed with P3X63Ag8.653 myeloma cells lacking hypoxanthine–guanine phosphoribosyltransferase. Cell fusions were performed using the traditional PEG method by Kohler and Milstein. The cells were then cultured for 7 days in hypoxanthine–aminopterin–thymidine medium to select for hybridoma cells capable of growth in the presence of aminopterin while eliminating unfused lymphocytes and myeloma cells. On day 7, the cells were transitioned to hypoxanthine–thymidine (HT) medium and cultured for four more days. For each PTC amino acid target, the resulting population of hybridomas was split into six aliquots and cryopreserved to generate a hybridoma

library and the hybridoma-conditioned HT medium was harvested to evaluate the binding performance of the secreted antibodies by ELISA as described above.

To isolate monoclonal hybridomas, one vial of each hybridoma library was thawed and single cells were sorted by flow cytometry into ten 96-well culture plates. After 2 weeks in culture, supernatants from all ten plates were evaluated by ELISA for binding to the respective target antigen. Positive clones were identified and further characterized by ELISA against a panel of other PTC-modified amino acid targets to confirm specificity. Clones selective for each target were expanded from 96-well plates into T-25 culture flasks with HT medium. Once cells reached confluence, each clonal line was cryopreserved in two aliquots, as was the conditioned hybridoma supernatant from each line. For selected monoclonal hybridomas, the secreted antibody was purified by protein A chromatography.

Measurement of antibody affinity

The biotinylated DBCO-modified oligonucleotide (50 nmol, **ODN-S19**) was dissolved in 100 µl of 1× PBS buffer. Then, 10 µl of 100 mM PTC amino acid in DMSO was added to the oligonucleotide and the reaction was incubated for 2 h on a rotator at room temperature. After conjugation, the conjugated oligonucleotide was purified by ethanol precipitation (0.3 M sodium acetate in 70% ethanol). BLI experiments were carried out on an Octet RED384 instrument. Briefly, SA biosensors were equilibrated in PBST buffer for 10 min. The biotinylated DNA-tagged PTC amino acid (100 nM) was loaded onto the SA biosensors for 150 s. The reference biosensors were incubated with PBST buffer alone. All biosensors were washed with PBST and then incubated with varying concentrations of anti-PTC amino acid antibodies in PBST buffer for 800 s. Finally, the associated antibodies were allowed to dissociate in PBST buffer for 150 s. The responses after reference subtraction were plotted against concentration and fitted to a Langmuir binding isotherm assuming 1:1 interaction using GraphPad Prism 10.0 to determine the antibody dissociation constant (K_d). It should be noted that, because of the orientation of the assay and the two-site binding of IgGs, the binding curves may not fit exactly into a 1:1 binding model.

Preparation of primer-modified anti-PTC amino acid antibodies

oYo-Link Tz was received as dry pellet and resuspended in 100 µl of nuclease-free water. This solution was mixed with 100 µl of 1 mg ml^{-1} antibody in 1× PBS and then irradiated by UV (365 nm) on ice for 2 h. After UV irradiation, excess oYo-Link Tz was removed by filtering with an Amicon centrifugal filter (50-kDa MWCO). The Tz-modified antibody (~1 mg ml^{-1}, 100 µl) was mixed with 10 µl of 300 µM TCO-modified primer **ODN-2**, **ODN-3** or **ODN-4** and the reaction was incubated for 2 h at room temperature. After conjugation, excess primer was removed by filtering with an Amicon centrifugal filter (50-kDa MWCO). The concentrations of primer–antibody conjugate were determined by UV–Vis on a NanoDrop spectrophotometer and the labeling was verified by SDS–PAGE and matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF; AB Sciex 5800 system; Supplementary Table 6).

PEA

The model PEA template (50 nmol, **ODN-5**) was dissolved in 100 µl of 1× PBS buffer. Next, 10 µL of 100 mM PTC amino acid in DMSO was added to the oligonucleotide and the mixture was incubated for 2 h on a rotator at room temperature. After conjugation, the conjugated oligonucleotide was purified by ethanol precipitation (0.3 M sodium acetate in 70% ethanol). Dynabeads MyOne SA C1 beads (1 mg) were washed twice with 1× binding and washing (B&W) buffer (5 mM Tris-HCl, 0.5 mM EDTA and 1 M NaCl, pH 7.5). Then, 100 µL of 100 nM biotinylated DNA-tagged PTC amino acids in 1× B&W buffer were incubated with the washed beads on a rotator for 2 h at room temperature and then washed three times with PBST buffer. DNA-loaded beads (0.01 mg) were incubated with 10 µl of 200 nM primer-modified anti-PTC amino acid antibody in 50 mM NaCl, 10 mM Tris-HCl, 10 mM $MgCl_2$ and 0.01% v/v Tween-20 (pH 7.5) for 30 min at room temperature. We next added 10 µl of mixture containing Klenow fragment (exo−; 1 U), 600 µM dATP, 600 µM dTTP, 600 µM dGTP, 600 µM dCTP, 50 mM NaCl, 10 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, and 0.01% v/v Tween-20 and performed primer extension for 30 min at 37 °C. After extension, beads were washed three times with PBST and resuspended in 20 µl of PBST.

Quantification of DNA output of PEA by qPCR

A 50-µl PCR reaction was prepared with 1 µl (0.5 µg) of beads after the PEA step and 2× GoTaq qPCR master mix. Final concentrations of forward and reverse primers (**ODN-S22** and **ODN-S23**) were 1 µM. Reactions were run on a Bio-Rad CFX96 real-time PCR detection system using the following program: 3 min at 95 °C, followed by 40 cycles of 10 s at 95 °C and 1 min at 55 °C. The C_q values were determined using Bio-Rad CFX Manager software. Negative control reactions were performed without beads.

Sequencing of a single peptide by ensemble peptide sequencing

We generally followed the Edman procedure described above, with some adjustments. To begin, 1 mg of Dynabeads MyOne carboxylic acid beads were modified with anchor DNA (**ODN-S24**) and DBCO-modified DNA containing peptide identity barcode (**ODN-S25**) was then conjugated by DNA ligation. Peptide RGFDWGX was subsequently conjugated by SPAAC. The subsequent proximity SPAAC reaction was carried out with DBCO-modified primer (**ODN-S29**). After each cycle, the PTC amino acid was precipitated and processed separately. PTC amino acids were resuspended in 1 ml of 1× B&W buffer and a 100-µl aliquot was incubated with 1 mg of Dynabeads MyOne SA C1 beads on a rotator for 2 h at room temperature. Subsequently, PEA was carried out with a mixture of DNA-barcoded (Supplementary Table 7) anti-PTC-R, anti-PTC-F, anti-PTC-D and anti-PTC-W antibodies (100 nM each) using the procedure described above.

For each individual PEA reaction, a 100-µl adaptor PCR reaction was prepared with 2 µl (1 µg) of the post-PEA beads and 2× colorless GoTaq G2 hot start master mix. The final concentration of forward and reverse adaptor primers was 1 µM each, where the forward primer contained the cycle number barcode. Reactions were run on an Eppendorf Mastercycler X50 PCR thermocycler using the following program: 2 min at 95 °C, followed by eight cycles of 15 s at 95 °C, 15 s at 54 °C and 30 s at 72 °C. The adaptor PCR reaction was cleaned up using an Axygen AxyPrep MAG PCR cleanup kit. PCR products from all peptide sequencing cycles were combined and indexed with the Nextera XT index kit v2. The indexed library was sequenced on an Illumina MiSeq sequencing system. The sequencing data were processed with Galaxy. Briefly, barcodes for cycle number, antibody and peptide were extracted using the 'trim sequences' tool. The extracted barcodes were joined to give a single barcode using the 'FASTQ joiner' tool and the counts of each barcode were counted with 'collapse' tool.

Probing barcode crosstalk during PEA

A 1:1 mixture of the 42-nt template coupled to PTC-F **ODN-5-PTC-F** and the 60-nt template **ODN-S40** (total DNA concentration = 1 µM) was prepared and immobilized onto SA beads at densities ranging 0.5–100 pmol per mg of beads. PEA reactions were carried out with 0.01 mg of modified SA beads using anti-PTC-F antibody conjugated with the 6-nt PEA primer **ODN-3** as described previously. After PEA, beads were resuspended in 20 µl of PBST. A 50-µl PCR reaction was prepared with 1 µl (0.5 µg) of beads and 2× GoTaq master mix plus a 1 µM final concentration of forward and reverse primers **ODN-S22** and **ODN-S23**. Reactions were run on an Eppendorf Mastercycler X50 PCR thermocycler using the following program: 2 min at 95 °C, followed by eight cycles of 15 s at 95 °C, 15 s at 54 °C and 30 s at 72 °C. The products

were analyzed by native PAGE. The gel was stained with GelStar nucleic acid gel stain and imaged by a Bio-Rad Gel Doc XR+ gel imaging system.

Parallel sequencing of multiple peptide species by ensemble peptide sequencing

To sequence multiple peptides simultaneously, we made the following modifications to the procedure for peptide immobilization. First, Dynabeads MyOne carboxylic acid (1 mg) were modified with anchor DNA (**ODN-S24**). Two azide-modified peptides, AFGGGX and AWGGGX, were clicked with DBCO-modified peptide-barcoding strand **ODN-S25** and **-S26**, respectively. For Fig. 5e, the azide-modified peptides ^{p}YGYGGX, YG^{p}YGGX and ^{p}YG^{p}YGGX were clicked with DBCO-modified peptide-barcoding strand **ODN-S25**, **ODN-S26** and **ODN-S27**, respectively. The peptide–DNA conjugates were mixed and immobilized on beads by DNA ligation. After each cycle of Edman degradation, a pull-down was carried out with a 10-µl aliquot of DNA-barcoded PTC amino acids and 2 mg of Dynabeads MyOne SA C1 beads to minimize barcode crosstalk during PEA. For Fig. 5c, PEA was carried out with a mixture of DNA-barcoded anti-PTC-F and anti-PTC-W antibodies (30 nM each). For Fig. 5e, two separate PEA reactions were performed: one with PY20 (30 nM) and the other with anti-PTC-Y antibody (100 nM). The rest of the procedure was identical to that used to sequence a single peptide species.

Single-molecule peptide sequencing

We generally followed the sequencing procedure described above with some adjustments. To begin, 1 mg of Dynabeads MyOne carboxylic acid beads were modified with anchor DNA (**ODN-S24**) and DBCO-modified DNA containing 30-nt UMI (**ODN-S41**) was then conjugated by DNA ligation. The peptide of interest, FWFWGGGX, was then immobilized by SPAAC. This procedure ensured that every peptide molecule was statistically tethered to a unique DNA barcode. Similarly, Dynabeads bearing the carrier peptide were prepared using another 5′-amino-modified anchor DNA (**ODN-S42**) and DBCO-modified DNA (**ODN-S43**). With this procedure, we estimated that approximately 100 pmol of peptides were immobilized on 1 mg of beads. Through serial dilution, ~5 ng of beads bearing the peptide of interest were mixed with 1 mg of beads bearing the carrier peptide. The mixture was subjected to DNA-encoded Edman degradation, where the proximity SPAAC reaction was carried out with a mixture of **ODN-S29** and **ODN-S45** (20 µM each). After DNA-encoded Edman degradation, the DNA-barcoded amino acids were cleaned up by ethanol precipitation and resuspended in 5 µl of 1× B&W buffer. Biotinylated DNA-barcoded amino acids were pulled down with 1 µg of Dynabeads MyOne SA C1 beads on a rotator for 16 h at room temperature. Subsequently, PEA was carried out with a mixture of DNA-barcoded anti-PTC-F and anti-PTC-W antibodies (30 nM each; Supplementary Table 7) using the procedure described above. A set of adaptor primers bearing a randomized stagger sequence were installed on the PEA product by PCR. The stagger sequence provides the necessary base diversity for sequencing on NovaSeq X plus. The PCR (100 µl) reactions were prepared with 1 µg of beads, Q5 High-Fidelity 2× master mix and forward and reverse primers (0.5 µM each). Reactions were run on an Eppendorf Mastercycler X50 PCR thermocycler using the following program: 2 min at 95 °C, 15 cycles of 15 s at 95 °C, 15 s at 64 °C and 30 s at 72 °C. The indexed libraries were pooled and sequenced on a NovaSeq X plus sequencer using a 25 B flow cell. The sequencing is carried out by Novogene.

Data processing of single-molecule peptide sequencing

The DNA sequencing data generated by NovaSeq is provided in compressed FASTQ.GZ format. Because of the large file sizes and the high number of UMIs, a customized C++ data-processing pipeline was developed to efficiently handle this data, using multithreading and techniques to reduce RAM usage. The source code is available from GitHub (https://github.com/whulwzheng-source/smPeptideSeq) and the process is described briefly below.

We used streaming decompression to avoid instantiating the full-size sequencing data. From this stream of decompressed data, we extracted the raw DNA sequences from each sequencing record. For each sequence, we identified the location and length of the UMIs and the amino acid barcode by searching for the conserved regions flanking them. A search for the cycle number was not required for this dataset because the sequencing result was already presorted as different sequencing cycles into separate files on the basis of sequencing indices. We disregarded any UMIs that were longer than 32 bp as they were unlikely to be valid sequences. The UMI was then encoded in a more efficient format to minimize RAM usage. For each UMI, we increased the running counter for each identified amino acid, bucketing them into three categories: ATA, TAT and undetermined. For each sequencing cycle, the resulting tallies were stored as a CSV file, with the columns UMI, ATA counts, TAT counts and undetermined counts, sorted by UMI. Next, the files for the different sequencing cycles were merged, exploiting the sortedness of the CSV files for memory efficiency, resulting in another CSV file with the columns UMI, R1_ATA, R1_TAT, R1_undetermined, …, R4_TAT and R4_ undetermined.

Subsequent data processing was performed in Python with the data-processing library Polars. To generate summarizing statistics, the CSV file was compiled into Boolean flags for each row, which are compactly stored as binary files. By using simple logic operations (NOT, AND, OR) and counting the number of true values, the statistics of sequencing results were obtained. Here, UMIs whose count never exceeded 3 in any of the four sequencing cycles were filtered, as they were more likely to be caused by errors during PCR and sequencing. This resulted in 2.94×10^8 UMIs passing the filter, the statistics of which were subsequently calculated. The results are summarized in Supplementary Table 4.

Simulation of proteome coverage by reverse translation with a selected set of amino acids

To predict the potential proteome coverage using the current set of antibodies with increasing sequencing length, a simulation was carried out following the method previously reported with UniProtKB/Swiss-Prot complete *Homo sapiens* proteome (UP000005640), comprising 20,405 proteins[56]. Similar to the reported procedure, we first loaded all protein sequences and then simulated a digestion by splitting the amino acid sequence strings after certain amino acids. Next, for the purposes of this simplistic simulation, efficiency and accuracy of sequencing were not considered; the only variables under consideration were the choice of protease, antibodies for amino acids and the number of sequencing cycles. To adapt the code to our method, for each of the peptides generated by simulated digestion, we masked all unidentifiable amino acids (that is, amino acids without an antibody) by replacing the letter with an underscore in the string and trimming its length to equal the number of sequencing cycles (or adding '_' in the C-terminal direction, when the length of peptides is shorter than the number of read cycles). Subsequently, we checked how many of the proteins had at least one peptide that was unique compared to all other proteins in the dataset and calculated the resultant fraction of uniquely identifiable proteins.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data supporting the findings of this study are available within the article and its Supplementary Information. The next-generation sequencing reads for ensemble and single-molecule peptide sequencing were deposited to the National Center for Biotechnology Information Sequence Read Archive under BioProjects PRJNA1420480 and PRJNA1423337, respectively. Source data are provided with this paper.

Code availability

The C++ and Python scripts for data processing and visualization of single-molecule peptide sequencing are available from GitHub (https://github.com/whulwzheng-source/smPeptideSeq).

Acknowledgements

This work was supported by the Helmsley Trust, Wellcome LEAP SAVE program. We thank A. Hugenmatter at Stanford Innovative Medicines Accelerator for the helpful discussion. We thank J. Lowitz at Antibody Solutions for his assistance with custom antibody generation. We thank T. McLaughlin at Stanford University MS for her assistance with developing LC–MS methods for oligonucleotides. This work was supported by the Vincent Coates Foundation MS Laboratory, Stanford University MS (RRID:SCR_017801) using the Bruker Microflex MALDI-TOF MS instrument (RRID:SCR_018696) and Thermo Exploris 240 LC–MS system (RRID:SCR_022216) that was purchased with funding from Stanford C-ShaRP (RRID:SCR_022986).

Author contributions

L.Z. and H.T.S. conceptualized the study. L.Z. and Y.S. performed the experiments and analyzed the experimental data. L.Z. and L.A.H. analyzed the single-molecule peptide sequencing data. H.T.S. supervised the research. L.Z. and H.T.S. wrote the original draft. L.Z., Y.S., M.E. and H.T.S. reviewed and edited the paper.

Competing interests

L.Z., Y.S. and H.T.S. are listed as coinventors on a pending patent application related to this work filed at the US Patent and Trademark Office (no. PCT/US2024/017167). L.A.H. and M.E. declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-026-03061-z.

Correspondence and requests for materials should be addressed to Liwei Zheng or Hyongsok Tom Soh.

Peer review information *Nature Biotechnology* thanks Chirlmin Joo, Stefan Howorka and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

nature portfolio

Corresponding author(s): *Double-anonymous peer review submissions: write DAPR and your manuscript number here instead of author names.*

Last updated by author(s): *YYYY-MM-DD*

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	MassLynx 4.2, flexControl 3.4, Xcalibur 4.7, VnmrJ 4.2, AB SCIEX Data Explorer® Software 3.7, CytExpert 2.4, OpenLab CDS 2.8, Image Lab 6.1, CFX manager 3.1, Fortebio 10.
Data analysis	Prism 10, FlowJo 10, Image Lab 6.1, CFX manager 3.1,Fortebio DA 10, Mnova 15, Protein Metrics Suite 5.6, flexAnalysis 3.4, ImageJ 1.54, Galaxy Server 25.0.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The main data supporting the results in this study are available within the paper and its supplementary information. The sequencing data for ensemble peptide

sequencing are available for download at https://doi.org/10.6084/m9.figshare.25901986.v1. The sequencing data for single-molecule peptide sequencing are too large to be publicly shared, but they are available from the corresponding author upon reasonable request. The C++ and Python scripts for data processing and visualization can be accessed from GitHub at https://github.com/whulwzheng-source/single-molecule-reverse-translation-sequencing.git.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	No human participants used
Reporting on race, ethnicity, or other socially relevant groupings	No human participants used
Population characteristics	No human participants used
Recruitment	No human participants used
Ethics oversight	No human participants used

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	*Describe how sample size was determined, detailing any statistical methods used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data exclusions	*Describe any data exclusions. If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Replication	*Describe the measures taken to verify the reproducibility of the experimental findings. If all attempts at replication were successful, confirm this OR if there are any findings that were not replicated or cannot be reproduced, note this and describe why.*
Randomization	*Describe how samples/organisms/participants were allocated into experimental groups. If allocation was not random, describe how covariates were controlled OR if this is not relevant to your study, explain why.*
Blinding	*Describe whether the investigators were blinded to group allocation during data collection and/or analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and*

	whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used: Custom generated antibodies are described in the manuscript. Phosphotyrosine antibody (PY20, Catalog No.ab10321) was purchased from Abcam. Phosphoserine antibody (3C171, Catalog No. MA1-91608), asymmetric dimethyl arginine antibody (21C7, Catalog No. 90005-100UG), and acetylated lysine antibody (RM101, Catalog No. MA5-33031) were purchased from Thermo Fisher Scientific

Validation: The affinity and specificity of antibodies used in the study are measured by Bio-layer interferometry (BLI).

Plants

Seed stocks: N/A

Novel plant genotypes: N/A

Authentication: N/A

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☐ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☐ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation: No cells or biological samples are used. Beads functionalized DNA barcoded peptides were prepared as outlined in the methods section. The amount of immobilized DNA is quantified by annealing with fluorescently labeled complementary DNA strands.

Instrument: Cytoflex (Beckman Coulter)

Software: FlowJo

Cell population abundance: This is not relevant because no cells were used in the flow cytometry experiments.

Gating strategy: Beads were gated by FSC/SSC and singlets to used for quantification.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.