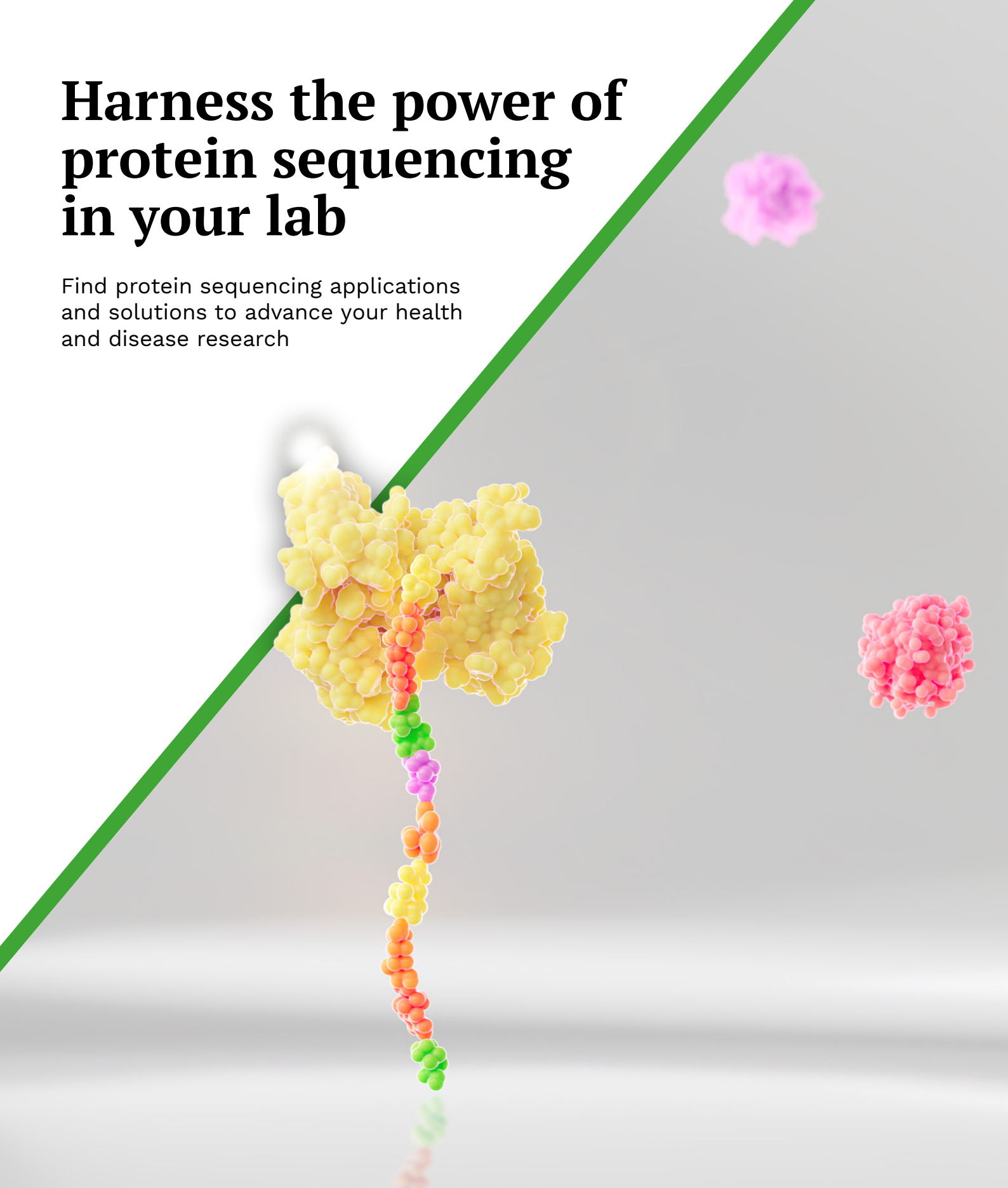
Harness the power of protein sequencing in your lab

Find protein sequencing applications and solutions to advance your health and disease research



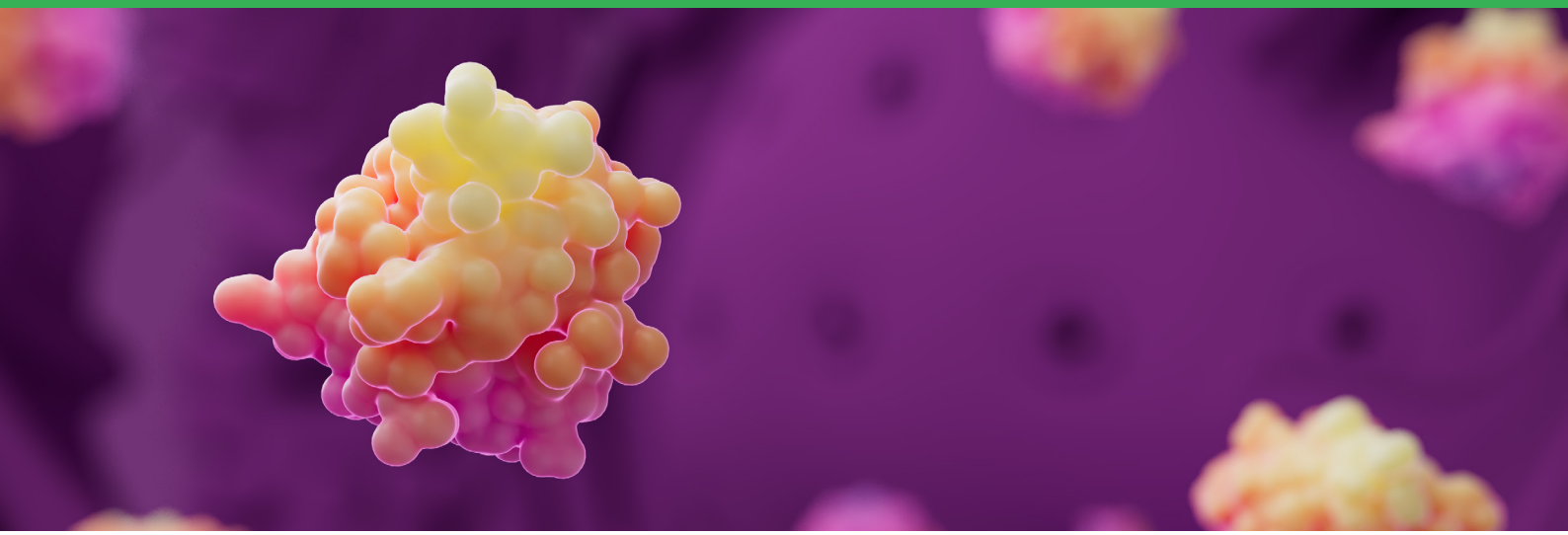
THE QUEEN'S AWARDS FOR ENTERPRISE: INTERNATIONAL TRADE 2021

BROUGHT TO YOU BY INDEPENDENT SCIENCE PUBLISHER

SelectScience®

IN PARTNERSHIP WITH

QUANTUM SI™



Introduction

Next-Generation Protein Sequencing™ technologies mark the dawn of a new era of scientific discovery in biology and medicine, just as the emergence of DNA sequencing tools revolutionized our understanding of human health.

Proteins are the vital engines of biological systems, playing crucial roles in health, and disease. Understanding their sequence, structure, and interactions is essential for biological discovery, developing new therapies and applying functional biology across industry. By studying proteins, researchers can gain deeper insights into biological processes, uncover disease mechanisms, and identify potential drug targets, leading to groundbreaking advancements in medical science.

Protein sequencing offers a unique means to investigate proteins at a level of resolution inaccessible with other technologies. By deciphering the amino acid sequence of proteins, scientists can predict their functions and interactions, leading to breakthroughs in understanding disease mechanisms, identifying new drug targets, and developing novel therapies. Additionally, protein sequencing allows for the detailed characterization of protein isoforms, providing deeper insights into the regulation and diversity of protein functions.

Moreover, with advanced single-molecule protein sequencing technologies like Quantum-Si's Platinum® Next-Generation Protein Sequencer™, researchers can now reliably identify changes to proteins at the amino acid level.

Contents

In this eBook, explore a curated selection of articles, methods, and application notes that demonstrate the power of protein sequencing for advancing our understanding of human health and disease. Read on to learn more about:

- Next-generation protein sequencing: How it works and key benefits
- How protein sequencing complements conventional proteomic techniques
- Applications that can be performed with the Platinum® instrument
- The impact of protein sequencing across various fields, from basic research to clinical

The power of protein sequencing

Almost all biological processes, whether in health or disease, manifest at the protein level. While genomic and transcriptomic analyses provide valuable information about gene expression, they only offer indirect glimpses into the resulting proteome. Alternative splicing, mRNA stability, translational regulation, and post-translational modifications can all expand the functional diversity of proteins encoded by a particular DNA sequence. Final protein sequence, isoform, structure, and abundance can only be revealed by direct protein-level analysis.

Protein sequencing facilitates the detailed characterization of protein isoforms, which offers deeper insights into the regulation and diversity of protein functions. This comprehensive analysis is crucial for studying protein variants and their impact on disease outcomes, understanding protein-protein interactions for drug identification and characterization, and ensuring the quality control of biotherapeutics like mRNA vaccines and antibodies.

Mass spectrometry and affinity-based technologies offer valuable and high-throughput approaches to proteomics, but when deeper characterization of a protein is needed, they lack the resolution to obtain information about a specific protein's amino acid composition. This is where next-generation protein sequencing (NGPS) on Quantum-Si's Platinum® instrument provides a highly complementary approach. NGPS excels in distinguishing between similar protein variants and provides superior data quality for applications requiring increased sensitivity or throughput on defined subsets of the proteome. These features of NGPS also make it highly suitable for protein barcoding applications, as it offers a means to directly read and identify protein barcode sequences with single-molecule resolution. This in turn creates opportunities for multiplexed protein characterization, variant screening, and many other applications.

The Platinum® instrument brings a novel approach to protein sequencing that relies upon carefully curated molecular biochemistry, photonics, and proprietary semiconductor technology to determine the precise amino acid sequence of hundreds of peptide strands simultaneously. Cloud-based software automatically analyzes the sequences, mapping them to their parent proteins and providing detailed information on proteoforms, post-translational modifications, and peptide abundance. All this is achieved in an instrument the size of a microwave, without the need for bioinformatics expertise, and at a much lower price point than legacy platforms.

By providing deeper insights into protein sequence variations and the dynamics of PTM regulation, instruments such as the Platinum® can drive major advances in the understanding of human biology, help identify novel drug targets, and accelerate the discovery of biomarkers for early diagnosis and personalized treatment. The remainder of this eBook covers some of the current applications that can be performed with the Platinum® instrument.

Protein identification

The separation, enrichment, and identification of proteins from complex mixtures is a key component of proteomic workflows. Western blot is one of the most widely used methods for identifying proteins following their separation. However, achieving reliable and reproducible results depends on high-quality antibodies with validated specificity for the target protein. Furthermore, differences in amino acid sequences and isoforms cannot be easily differentiated by western blot if the size of the protein is similar.

NGPS can overcome this limitation. In this application note, explore an in-gel protein digestion and peptide extraction procedure that can be used to separate and enrich single proteins from complex mixtures for subsequent identification on Quantum-Si's Platinum® instrument.

Protein screening

Protein screening with barcodes can enhance the ability of next-generation protein sequencing (NGPS) to detect protein variants.

Protein barcodes – short stretches of amino acids that can be genetically encoded within proteins to house extensive information – hold immense promise across fields such

as proteomics, synthetic biology, protein engineering, and drug discovery. Published studies employing protein barcoding have previously relied on mass spectrometry for decoding, but this technology cannot always differentiate among peptides of similar length and sequence.

The Platinum® instrument is the first to enable protein barcode sequences to be read directly and identified with single-molecule resolution. It relies on unique kinetic signatures to differentiate and identify amino acid residues, enabling the quantification of peptide sequences with high confidence. In doing so, the instrument introduces a convenient and user-friendly method for decoding peptide barcodes, opening avenues for multiplexed protein characterization, variant screening, and various other applications. Read the full application note to explore criteria for protein barcode design, methods for generating protein barcode libraries, and example applications that illustrate how protein barcodes can be used for screening proteins with desired properties.

Protein variant distinction

A key strength of NGPS lies in its ability to detect differences among protein variants, such as single amino acid differences in peptide sequences and the presence or absence of specific peptides. This capability is especially relevant for disease research and virology, where understanding subtle differences can have profound implications.

For instance, virus variant detection is crucial for understanding and responding to the evolving landscape of viral infections, as it enables the identification of specific changes in viral proteins that may impact infectivity, vaccine efficacy, and disease severity. In this application note, see how Quantum-Si's Platinum® was used to directly distinguish variants of the SARS-CoV-2 virus based on differences in the amino acid sequence of their spike proteins. Not only does variant detection support tracking and surveillance efforts, but detecting the sequence of residual viral proteins in long COVID patients could also lead to new understandings and treatment of the disease.

Detecting post-translational modifications

Post-translational modifications of amino acid side chains significantly influence protein function and cellular events, making their measurement crucial for understanding health and disease. While most proteomic techniques are unable to distinguish between the various proteoforms in the human proteome, the Platinum® instrument possesses the sensitivity to detect PTMs. In this application note, see how Platinum® technology can detect arginine methylation and citrullination in p38MAPKα and Vimentin proteins with single-molecule resolution. These modifications are biomedically significant, playing key roles in diseases like cardiovascular disease, autoimmune disease, and cancer.

Conclusion

Quantum SI's benchtop sequencer revolutionizes protein analysis with its unique kinetic signature technology. A kinetic signature comprises the measurable characteristics of the series of dynamic recognizer events that uniquely identifies a peptide. These signatures distinguish alterations in amino acids, and post translational modification events, providing a highly specific and accurate picture of peptides and their respective proteins. As Next-Generation Protein Sequencing™ becomes a standard for understanding biology at the amino-acid level, kinetic signatures will be crucial for identifying and exploring variation in a way never before possible. Quantum SI's technology is key to advancing our understanding of the proteome, offering unmatched sensitivity to protein variants and enabling deep interrogation of proteins to deliver breakthrough insights.

Protein Identification Using Next-Generation Protein Sequencing of In-Gel Digested Proteins

SUMMARY

Western blot is one of the most widely used methods for protein identification but achieving reliable and reproducible results depends on quality antibodies with validated specificity for the target protein. Furthermore, differences in amino acid sequences and isoforms cannot be easily differentiated by western blot if the size of the protein is similar. To overcome these limitations, non-antibody-based approaches are highly desired. In this technical note, we describe an in-gel protein digestion and peptide extraction procedure compatible with Quantum-Si's library preparation and protein sequencing workflows. This method can be used for the separation and enrichment of single proteins from complex mixtures using SDS-PAGE, followed by their identification on Quantum-Si's Platinum® next-generation protein sequencing instrument.

Q-SI TECHNOLOGY

Quantum-Si's benchtop Platinum® instrument enables protein sequencing from biological samples in a simple user-friendly workflow. Our technology utilizes dye-tagged N-terminal amino acid recognizers and semiconductor chip technology to detect the binding characteristics and binding order of N-terminal amino acids, resulting in unique kinetic signatures that can be used to differentiate and identify amino acid residues and PTMs. A more detailed overview of the workflow and technology can be found in our Science Paper.

INTRODUCTION

The separation and enrichment of complex samples is a key component of proteomic workflows. A relatively straightforward method for sample enrichment is gel electrophoresis. Among the different gel electrophoresis techniques, SDS-PAGE (sodium dodecyl-sulfate polyacrylamide gel electrophoresis) is widely used for protein mixtures. During this process, an electric field is applied to denatured proteins bound to the anionic detergent sodium-dodecyl sulfate (SDS), resulting in their migration through a crosslinked polyacrylamide gel matrix. This migration separates

the proteins by their length and mass-to-charge ratio, resulting in individual protein bands that are resolved in the gel upon staining. While SDS-PAGE enables the separation and enrichment of complex samples, additional downstream procedures are required to identify the separated components.

Western blot is one of the most common detection methods to identify specific proteins following their separation by SDS-PAGE. In this technique, proteins are transferred from the gel to a nitrocellulose membrane, followed by blocking and then by incubation with a primary antibody specific to the protein of interest. A secondary enzyme-conjugated antibody is then used to bind to the primary antibody, resulting in a signal that can be visualized on a western blot imager for protein identification. Despite being the most frequently used antibody-based method, the availability of properly validated antibodies without off-target binding remains an issue, leading to wasted reagents, time, and money as well as erroneous data.[1] Therefore, finding alternative antibody-free methods for sensitive protein identification can enhance research productivity and reduce waste.

One promising approach for identifying pre-separated components without the use of antibodies is through next-generation protein sequencing using Quantum-Si's Platinum workflow.[2] This innovative method entails cutting proteins from the SDS-PAGE gel, digesting proteins into peptide fragments, and linking them to macromolecular linkers at the C-terminus. The peptides are then immobilized on Quantum-Si's semiconductor chip and then probed with dye-labeled N-terminal amino acid (NAA) recognizers, resulting in distinctive pulsing patterns with characteristic kinetic properties. Individual NAAs are removed sequentially by aminopeptidases in solution, uncovering subsequent amino acids for recognition. Fluorescence lifetime, intensity, and kinetic data are collected in real time and analyzed with Cloud-based software to identify the peptide sequence and corresponding protein.

To combine the benefits of separating complex mixtures by gel electrophoresis with the protein identification capabilities by Platinum, we have developed an in-gel digestion procedure that is compatible with Quantum-Si's downstream library preparation and sequencing workflows. The in-gel digestion protocol described here is similar to procedures used in other proteomic workflows.[3] To demonstrate the utility of this procedure, we prepared in-gel digested peptide libraries using the protein CDNF

REFERENCES

1. Baker M. Reproducibility crisis: Blame it on the antibodies. Nature. 2015;521(7552):274-276. doi:10.1038/521274a
2. Reed BD, Meyer MJ, Abramzon V, et al. Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device. Science. 2022;378(6616):186-192. doi:10.1126/science.abo7651
3. Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M. In-gel digestion for mass spectrometric characterization of proteins and proteomes. Nat Protoc. 2006;1(6):2856-2860. doi:10.1038/nprot.2006.468

(Cerebral Dopamine Neurotrophic Factor) as a model protein. Sequencing of the in-gel digested peptide libraries produced single molecule peptide traces that aligned with peptides from CDNF, enabling the successful identification of the protein. These results demonstrate the compatibility of the in-gel digestion procedure with Quantum-Si's library preparation and sequencing workflow and suggest its potential in separating and enriching more complex sample mixtures to resolve differences in amino acid sequence.

METHODS

We began by resuspending recombinant CDNF protein (Bio-Techne Cat # 5097-CD-050) as a 1 mg/mL solution in 125 mM HEPES, pH 8.0. Samples containing either 0.1 µg, 0.5 µg, 1 µg, 5 µg, or 10 µg of CDNF were prepared in SDS loading buffer with reductant, heated at 95°C for 5 min, and loaded onto a 4–20% Novex TGX gel (Thermo Fisher Scientific Cat # XPO4200BOX). The gel was run in Novex TGX SDS running buffer (Thermo Fisher Scientific Cat # LC2675) at 225 V for 35 min. To visualize gel bands, the staining and destaining procedure was performed with an automated eStain Coomassie staining system (Genscript Cat # L00657).

Next, we excised each band from the gel with a new razor blade and placed the gel bands on a clean surface. Each gel band was then diced into 1-mm cubes and placed into a clean 1.5-mL tube. After dicing the gel bands, we used an in-gel digestion kit (Thermo Fisher Scientific Cat # 89871) to digest the protein and extract the peptides from the gel pieces. Briefly, the gel pieces were destained, reduced, alkylated, and then washed as per the standard protocol. Next, we utilized the LysC endoprotease from the Quantum-Si Library Preparation Kit instead of trypsin in the digestion process to ensure compatibility with Quantum-Si library preparation workflow. The digestion solution was prepared by adding 1 µg of LysC endoprotease into 50 µL of the digestion buffer. We then added the full 50 µL LysC digestion solution to the dehydrated gel pieces, wrapped the tubes in parafilm, and incubated the tubes overnight (16 hrs) at 37°C with light shaking on a thermomixer. On the following day, we collected the supernatants by aspirating the solution containing the CDNF digested peptides and placed solution from each sample in a new collection tube.

To maximize recovery of the digested peptide fragments for each sample, we performed two rounds of peptide extraction on the remaining gel pieces. In the first round, 50 µL of extraction solution 1 (50% acetonitrile, 0.02% trifluoroacetic acid) was added to the tube containing the gel pieces and incubated at room temperature for 45 min. The tubes containing the gel pieces were subsequently sonicated for 5 min in a water bath, and the extraction solution was removed and placed in the tube containing the digested peptide supernatant. This process was repeated a second time but with extraction solution B (90% acetonitrile, 0.02% trifluoroacetic acid). After all gel extractions were completed, the combined peptide supernatant samples were evaporated to dryness by SpeedVac at 30°C for 60-90 minutes to remove the volatile ammonium bicarbonate from the samples and facilitate Quantum-Si's downstream library preparation workflow.

To generate the peptide library for sequencing, the digested peptides were prepared using Quantum-Si's Library Preparation Kit and Protocol with the following modification. As the peptide concentration was unknown, we resuspended the 1 µg, 5 µg, and 10 µg dried peptide samples in 50 µL sample buffer. Due to the lower input concentrations, the 0.1 µg and 0.5 µg were resuspended in 25 µL sample buffer. After preparing the digested peptide solution, we began the protocol starting at Day 2 according to the Quantum-Si Library Preparation Protocol (Document # 950-10007-00) and followed the remaining steps, except that the volume of Activators E, F, and G was reduced by a fourth or half to account for the initial 50 µL or 25 µL sample volume. To ensure an excess of peptides to Solution K during the library reaction step, all reactions contained 1 µL of Additive N and 1 µL of Solution K, however 48 µL of the derivatized peptide solution was used for the 1 µg, 5 µg, and 10 µg input samples and 23 µL of the derivatized peptide solution was used for the 0.1 µg and 0.5 µg input samples. All reactions were wrapped in parafilm and incubated overnight (16 hrs) at 37°C. On the next day, samples were removed from heat and stored on ice until sequencing on Quantum-Si's Platinum instrument using the Protein Sequencing Kit and Protocol (Document # 950-10006-00).

RESULTS AND DISCUSSION

To explore the applicable protein input compatible with in-gel digestion in tandem with Quantum-Si's library preparation workflow, a titration series of CDNF from 0.1 µg to 10 µg was prepared. Each of the 5 samples were run side by side on a gradient gel (Figure 1), followed by excision and processing of the corresponding bands as described in the methods section (Band A). In addition to the bands at the expected CDNF molecular weight, we observed a slower migrating protein on the gel in both the 5 µg and 10 µg protein lanes (Band B). To determine the identity of this band, it was also excised from the gel of the 10 µg sample and subjected to the in-gel digestion and library preparation workflow.

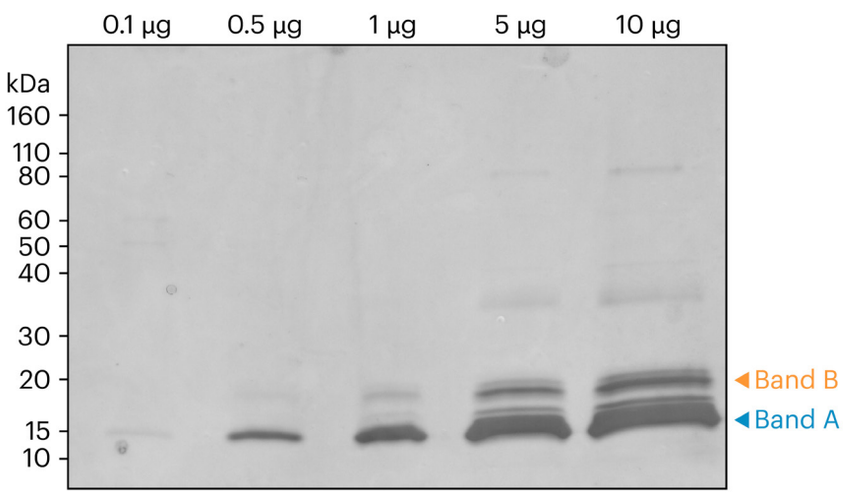


FIGURE 1

SDS-PAGE Gel Image of CDNF Samples Prior to Band Excision. Orange and blue arrows indicate the approximate location of the bands removed for in-gel digestion and subsequent sequencing.

We next sequenced the protein libraries on the Platinum instrument and generated single molecule sequencing traces, resulting in the identification of the input sample as CDNF. A total of 5 CDNF peptides were identified from the isolated 5 µg and 10 µg gel slices. Representative single molecule traces from 5 peptides are shown in Figure 2. While both the 5 µg and 10 µg libraries produced sufficient alignments to identify CDNF, the peptide libraries prepared from the 0.1 µg, 0.5 µg and 1 µg gel slices produced very few CDNF alignments (Figure 3). This outcome was expected given that the total protein starting input was outside of the recommended range of Quantum-Si's library preparation

protocol, considering the potential losses from the in-gel protein digestion and peptide recovery process. This issue could potentially be mitigated in future experiments by adding a digested carrier protein to the in-gel digested peptides to ensure that the minimum protein concentration requirements are met for the library reaction.

We also performed sequencing on the slower migrating band B in the 10 µg sample. Interestingly, this library also produced a significant number of CDNF peptide alignments, although fewer than both the 5 µg and 10 µg libraries (Figure 3). Single molecule traces representing all CDNF peptides, except ELISFCLTDK, were identified in the slower migrating 10 µg band. While the cause for the difference in the migration speed within the sample is unclear, the sequencing data clearly indicated that the slower migrating band was also CDNF.

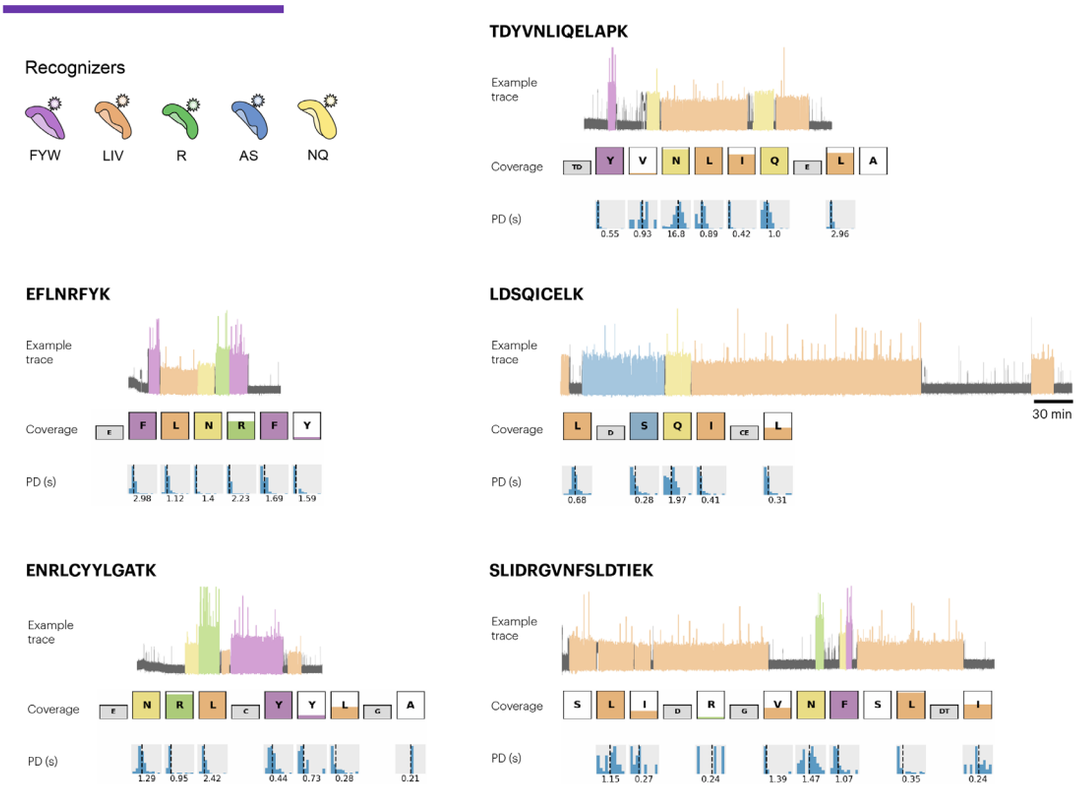


FIGURE 2

Representative Traces, Coverage, and Pulse Duration Data for the Peptides Identified in the 10 µg Sample of CDNF. Five recognizers were used to identify 12 amino acids (F, Y, W, L, I, V, A, S, N, Q, R, and K), enabling identification of 5 peptides as shown in the figure.

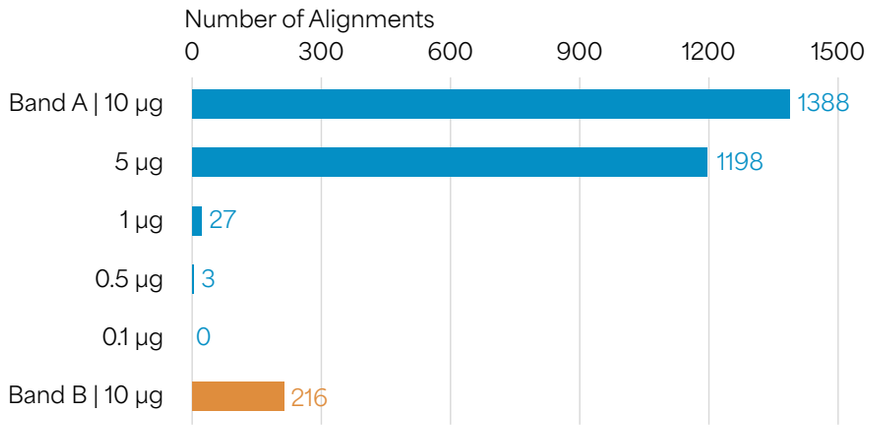


FIGURE 3

Bar Graph Showing the Relative Number of Peptide Alignments for Each In-Gel Digested Sample. For Band A, both the 5 µg and 10 µg samples produced more aligned CDNF peptides compared to the 0.1 µg, 0.5 µg, and 1 µg samples. Additionally, the 10 µg Band B sample also produced a significant number of CDNF aligned peptides. All alignments normalized to the total alignments from the 10 µg library.

CONCLUSION

In this technical note, we presented an in-gel digestion protocol compatible with Quantum-Si's next-generation protein sequencing workflow on Platinum. Using this procedure, we demonstrated the acceptable range of sample input for libraries prepared from in-gel digested samples that enabled the successful sequencing and identification of CDNF. This procedure has the potential to facilitate the pre-separation and identification of multiple proteins from complex mixtures, offering an alternative method to antibody-based western blot methods for protein identification from SDS-PAGE gels.

QUANTUM SI™

29 Business Park Drive, Branford, CT 06405

www.quantum-si.com | 866.688.7374

Protein Barcodes for Next-Generation Protein Sequencing™

January 2024

INTRODUCTION

Barcoding technologies that take advantage of the high information content that can be encoded in oligonucleotide sequences have enabled a wide array of applications in biotechnology, particularly when coupled with next-generation DNA sequencing (NGS) to decode this information in a high-throughput, cost-effective manner. For example, DNA barcoding combined with NGS readout is employed to track sample identity in multiplexed libraries, provide single-cell or spatial resolution in transcriptomic studies, and to track the enrichment of genotypes in directed evolution methods for protein engineering.[1-3] Protein barcodes, with their capacity to encode extensive information in short sequences, easy genetic encoding, and chemical versatility, present cutting-edge opportunities. Applications of peptide barcoding that use mass spectrometry for decoding, such as nanobody screening based on flycodes,[4] have been developed. However, there remains a critical need for accessible methods to directly read protein barcode sequences and to identify protein barcodes with single-molecule resolution. Next-Generation Protein Sequencing™ (NGPS) on Quantum-Si's Platinum® instrument offers researchers the ability to directly sequence protein barcodes with single-molecule resolution for the first time. This user-friendly benchtop platform combines the ease-of-access of DNA-based methods with the exciting, innovative capabilities of peptide-based barcoding approaches.

Quantum-Si's workflow for real-time, single-molecule protein sequencing begins with attachment of peptides to macromolecular linkers at the C-terminus for immobilization on a semiconductor

Q-SI TECHNOLOGY

Quantum-Si's benchtop Platinum® instrument enables protein sequencing from biological samples in a simple user-friendly workflow. Our technology utilizes dye-tagged N-terminal amino acid recognizers and semiconductor chip technology to detect the binding characteristics and binding order of N-terminal amino acids, resulting in unique kinetic signatures that can be used to differentiate and identify amino acid residues and PTMs. A more detailed overview of the workflow and technology can be found in our Science Paper.

chip. Following peptide immobilization, dye-labeled N-terminal amino acid (NAA) recognizers and aminopeptidases are added to the chip to initiate the sequencing run. The recognizers repetitively bind and unbind the immobilized peptides when their cognate NAAs are exposed at the N-terminus. This activity generates a distinct series of pulses, termed a recognition segment (RS), for each recognized NAA with characteristic fluorescence and kinetic properties. Aminopeptidases in solution sequentially remove individual NAAs, exposing subsequent residues for detection. This dynamic process repeats until the peptide has been completely sequenced. The temporal order of NAA recognition and associated kinetic properties over the time course of sequencing are highly characteristic for a given peptide and are termed its kinetic signature. Kinetic signatures are analyzed with Platinum Analysis Software to provide high confidence alignments to individual peptide sequences.[5]

Here, we present a novel protein barcoding approach using NGPS on Platinum. We designed and sequenced a set of synthetic peptides to demonstrate that they can be used as distinguishable barcodes. Statistical analysis of these sequences allowed us to establish criteria for protein barcode design and scalability estimation. We then developed a method for enzymatic library preparation of recombinantly expressed protein barcodes and demonstrated the relative quantitation of a mixture of protein barcodes on Platinum. Finally, we demonstrated the use of protein barcodes to select for proteins with characteristic properties. In this experiment, we observed a 300-fold enrichment in the relative abundance of an anti-GFP nanobody after positive selection using protein barcodes enzymatically cleaved from the enriched nanobody. This application note provides criteria for protein barcode design, methods for generating protein barcode libraries, and example applications that illustrate how protein barcodes can be used for screening proteins with desired properties.

METHODS

SYNTHETIC PEPTIDE PRODUCTION

Protein barcodes were synthesized by Innopep using solid-phase peptide synthesis. All peptides were prepared with an azido-lysine

REFERENCES

1. Wang, Y. & Navin, N. E. Advances and Applications of Single-Cell Sequencing Technologies. Mol. Cell 58, 598-609 (2015).
2. Whitmann, B. J., Johnston, K. E., Almhjell, P. J. & Arnold, F. H. evSeq: Cost-Effective Amplicon Sequencing of Every Variant in a Protein Library. ACS Synth Biol 11, 1313-1324 (2022).
3. Baysoy, A., Bai, Z., Satija, R. & Fan, R. The technological landscape and applications of single-cell multi-omics. Nat. Rev. Mol. Cell Biol. 24, 695-713 (2023).
4. Egloff, P. et al. Engineered Peptide Barcodes for In-Depth Analyses of Binding Protein Ensembles. Nat. Methods 16, 421-428 (2019).
5. Reed, B. D. et al. Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device. Science 378, 186-192 (2022).
6. Fridy, P. C. et al. A robust pipeline for rapid production of versatile nanobody repertoires. Nat. Methods 11, 1253-1260 (2014).
7. Zimmermann, I. et al. Synthetic single domain antibodies for the conformational trapping of membrane proteins. eLife 7, e34317 (2018).
8. Levenshtein, V. I. Binary codes capable of correcting deletions, insertions, and reversals. Dokl. Phys. 163, 845-848 (1965).
9. Mazmanian, S. K., Ton-That, H. & Schneewind, O. Sortase-catalysed anchoring of surface proteins to the cell wall of Staphylococcus aureus. Mol. Microbiol. 40, 1049-1057 (2001).
10. Jovčevska, I. & Muyldermans, S. The Therapeutic Potential of Nanobodies. BioDrugs 34, 11-26 (2020).

modification at the C-terminus. Peptides were confirmed to be 95% pure by HPLC and mass spectrometry analysis.

BARCODE EXPRESSION AND PURIFICATION

Each protein barcode was cloned into a plasmid containing an N-terminal HaloTag, a TEV protease site, a SUMO Tag, the protein barcode sequence, a Sortase A recognition motif (LPETGG), and a C-terminal 6x-Histidine tag. Following cloning, the barcode plasmids were transformed into SHuffle T7 express competent *E. coli* (New England Biolabs, Catalog No. C3029J) and cultivated overnight on the appropriate antibiotic selection media. The sequence of the barcode plasmids was confirmed by Sanger sequencing. Next, 10 mL of Terrific Broth was inoculated with bacteria overnight with shaking at 37°C. Barcode expression was induced with the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) at OD 0.6-0.8. The bacteria were harvested by centrifugation, and the resulting pellet was washed with 50 mM HEPES pH 7.3, 150 mM NaCl, and stored at -80°C until ready for purification.

For purification of the protein barcodes, the thawed cell pellets were resuspended in 0.2 mL of 50 mM HEPES, pH 7.3, 150 mM NaCl. Cells were lysed by adding 1 mL of NEB Express lysis buffer and mixing for 30 min at room temperature. The bacterial lysate was then centrifuged at 10,000 x g for 10 min to pellet cell debris, and the supernatant was collected for purification. Next, 200 µL of Ni-NTA resin was pre-equilibrated by washing 3x with 50 mM HEPES, pH 7.3, 150 mM NaCl, 10 mM imidazole buffer. The bacterial supernatant was diluted with 1 volume of 50 mM HEPES, pH 7.3, 150 mM NaCl, 10 mM imidazole buffer, and the supernatant solution was applied to the pre-equilibrated Ni-NTA resin and incubated for 30 min at room temperature with mixing. The resin was then pelleted by centrifugation at 1,000 x g for 2 min, and the supernatant was collected. Next, the resin was washed 3x with 400 µL of 50 mM HEPES, pH 7.3, 150 mM NaCl, 10 mM imidazole buffer. Elution was performed by adding 500 µL 50 mM HEPES, pH 7.3, 150 mM NaCl, and 300 mM imidazole buffer to the resin, followed by 5 min of incubation at room temperature. The resin was then pelleted by centrifugation at 1,000 x g for 2 min.

SORTASE A REACTION

For the Sortase labeling reaction, 10 µg of purified barcode peptide was introduced into a solution containing 1 µM of Sortase A pentamutant (BPS Biosciences), 1 mM of a synthetic tri-glycine-azide (GGG-azide) peptide (Click Chemistry Tools), and 1X Sortase Buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 10 mM $CaCl_2$). The final reaction volume of 50 µL was incubated for 1 hr at 37°C to facilitate the incorporation of the GGG-azide motif at the C-terminus of the barcode peptide.

For the purification of the azide-labeled protein barcode product, 20 µL of Magne HaloTag beads (Promega) were used per protein barcode. The HaloTag beads were first washed 4x for 5 min with 200 µL of HEB Buffer (HEPES 50 mM pH 7.3, 0.005% IGEPAL CA-630, NaCl 150 mM) using a magnetic stand for bead collection during the wash steps. Next, 50 µL of HEB buffer was added to the 50 µL Sortase reaction (1:1), and the combined solution was added to the beads. This mixture was incubated with end-over-end mixing for 1 hr at room temperature. After incubation, the beads were magnetically separated, and the flow-through was collected for SDS-PAGE analysis. Next, the beads were washed 3x for 5 min with end-over-end mixing using HEB buffer. The beads were then stored in the final wash buffer until ready to proceed to the click chemistry reaction.

CLICK CHEMISTRY REACTION

The click reaction was performed with the azide-labeled protein barcodes still attached to the beads. After the supernatant solution was removed from the beads, 22 µL of HEB buffer was introduced. Next, 0.5 µL of Cetyltrimethylammonium bromide (CTAB) and 1 µL of K-Linker from the Library Preparation Kit - Lys-C (Quantum-Si, Cat. No. 910-00012-00) were added to the mixture. The reaction was allowed to proceed overnight at 37°C in a Thermomixer (Eppendorf) at 1,400 rpm.

The next day, the beads were washed 3x with HEB buffer, as previously described. To elute the protein barcodes from the HaloTag, 25 µL of HEB Buffer supplemented with 1 mM DTT, and 1 U of Sumo Protease (ThermoFisher) was added to each protein barcode. The reaction was incubated for 1 hr at 37°C at 1,400 rpm in a Thermomixer. The beads were then collected using the magnetic stand, and the supernatant containing the eluted protein barcode

conjugated to the K-Linker was retrieved. The concentration of the eluted protein barcodes was determined by UV-Vis absorbance and gel densitometry.

NANOBODY SELECTION

The nanobodies targeting MBP (Sb_MBP#1) and GFP (LaG-16) have been previously characterized.[6,7] These constructs were inserted between the TEV protease site and the SUMO Tag in the peptide barcoding plasmid described above. Specifically, the MBP nanobody was fused to Barcode A (RLIFAA), and the GFP nanobody was fused to Barcode B (FLRAA). The barcoded nanobodies were then expressed at 22°C in *E. coli*, purified, functionalized with Sortase A, and conjugated via GGG-azide moiety with the macromolecular K-Linker for peptide immobilization on chip after proteolysis. A pre-selection sample fraction of this library was stored for sequencing on Platinum.

The barcoded model nanobody library underwent affinity purification selection using commercially purchased GFP (Sigma-Aldrich) immobilized on M-280 tosylactivated Dynabeads (ThermoFisher), adhering to manufacturer's instructions. Briefly, the barcoded nanobody library (10 nM) was incubated with the GFP-coated Dynabeads for 5 min at room temperature in 50 mM Tris pH 7.5, 0.5 % Tween-20. Subsequently, the beads were magnetically separated and washed 5x with buffer. Barcodes from the post-selection sample associated with the nanobodies that remained on the beads after washing were then harvested by proteolysis, resulting in a nanomolar barcode solution ready for NGPS.

NEXT-GENERATION PROTEIN SEQUENCING ON PLATINUM

Single-molecule sequencing of protein barcodes was performed according to Quantum-Si's user protocols. Briefly, conjugated peptide libraries were loaded onto Quantum-Si's semiconductor chip (Catalog No. 910-00011-00) via 15 min incubation. A solution containing dye-labeled recognizers was prepared using Quantum-Si's Sequencing Kit (Catalog No. 910-00011-00) and added to the chip. The chip was then installed in the Platinum instrument (Catalog No. 910-10904-00) and data was collected for 15 min. Next, a solution containing aminopeptidases from Quantum-Si's

Sequencing Kit was added to the chip with mixing and data was collected for 10 hours. Following completion of sequencing runs, data was analyzed using Quantum-Si's automated Platinum Software Analysis pipeline.

RESULTS & DISCUSSION

SHORT PEPTIDE SEQUENCES CAN BE USED AS BARCODES ON PLATINUM AND ALLOW BUILDING DIVERSE BARCODE SETS

Data from previously sequenced proteins was evaluated, and the following peptide sequences were identified as candidate protein barcodes: EFLNRFY, VRFLEQQN, DQFRLAGG, ARLAFAYPDDD, FQRIALNFA, RLAIQFAYPDDD, FAQLQARFAADDD, and ENRLCYYLGAT. To test whether this group of peptides had the required kinetic and sequence properties to enable use as a barcode set, we generated synthetic peptides and analyzed the sequencing performance of these synthetic peptides individually. Kinetic signature plots for the 8 different peptides are shown in **Figure 1**. The signals from each of these runs were also aligned to the entire set of 8 peptides to measure the false discovery rate (FDR), defined as the fraction of off-target alignments for a given peptide. The maximum FDR of all 8 peptides was 0.2%. These results highlight the strength of Platinum's NGPS in generating distinct patterns of RSs with characteristic kinetic properties that enables the detection of each protein barcode with high-confidence using Quantum-Si's NAA recognizers.

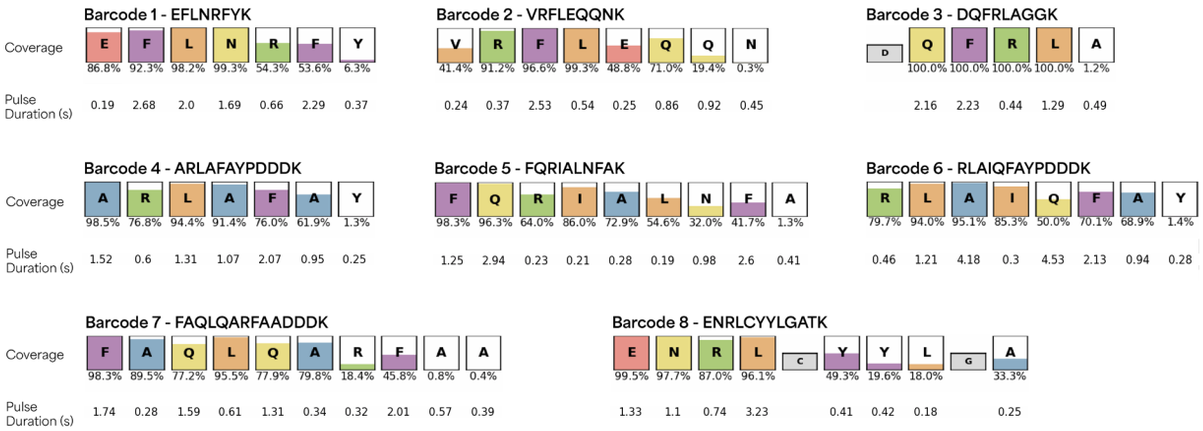


FIGURE 1

A set of protein barcodes displaying unique patterns of RSs and on-chip pulse durations, enabling accurate identification with Platinum. The color fill in each box represents the total percent coverage of observed amino acids in each peptide sequence with the numerical value reported under each box. The mean pulse duration for each observed residue, measured in seconds, is given under each peptide sequence coverage box.

Having characterized each peptide individually on Platinum, we next sought to establish that these peptides are distinguishable from one another when sequenced together in a mixture. The Levenshtein distance (**L**) is a common measure of sequence similarity in bioinformatics. It is defined as the minimum number of edit operations (insertions, deletions, and substitutions) needed to transform one sequence into another.[8] For our set of eight sequences, we computed the Levenshtein distance for every pair of sequences and observed **L** ≥ 3 for all pairs, with a mean **L** = 8.7 (**Figure 2**). This analysis, combined with the distinct kinetic signatures and low FDRs observed in NGPS, suggested that **L** ≥ 3 in all pairwise comparisons is a suitable threshold to generate barcode sets with highly distinguishable sequences. We used this threshold next to estimate how the sizes of protein barcode sets could be increased.

A

Sequence 1 - EFLNRFY
Sequence 2 - VRFLEQQN
Sequence 3 - DQFRLAGG
Sequence 4 - ARLAFAYPDDD
Sequence 5 - FQRIALNFA
Sequence 6 - RLAIQFAYPDDD
Sequence 7 - FAQLQARFAADDD
Sequence 8 - ENRLCYYLGAT

B

Sequence	1	2	3	4	5	6	7
2	6						
3	7	7					
4	9	9	10				
5	7	8	7	10			
6	10	11	10	3	10		
7	10	11	10	6	9	8	
8	9	10	9	9	7	11	11

C

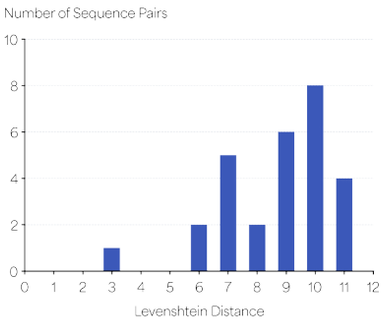


FIGURE 2

Levenshtein distance analysis of peptide sequences. (A) Sequences of peptides used in the Levenshtein distance analysis. (B) Levenshtein distance values for all pairs of sequences, showing **L** ≥ 3 for all pairs. (C) Histogram of orthogonal pair-wise Levenshtein distances between all members of the barcode set.

Many applications require extension of a barcode set to hundreds or even thousands of sequences. To assess the scalability of protein barcodes on Quantum-Si's Platinum, we constructed barcode sets in silico with **L** ≥ 3 for 6-amino-acid-character sets (L, F, R, N, A, and E), ranging in length from 3 to 8 residues, while restricting repetitive sequences to ensure an amino acid could only repeat every 5 residues. We successfully constructed barcode sets ranging from 6-member sets for 3-residue barcodes to 1,600-member sets for 8-residue barcodes (**Figure 3**). Depending on the specific application, one could consider other design parameters to align with the desired throughput, sensitivity, and accuracy. These findings underscore the scalability of peptide design needed for applications requiring thousands of unique barcodes.

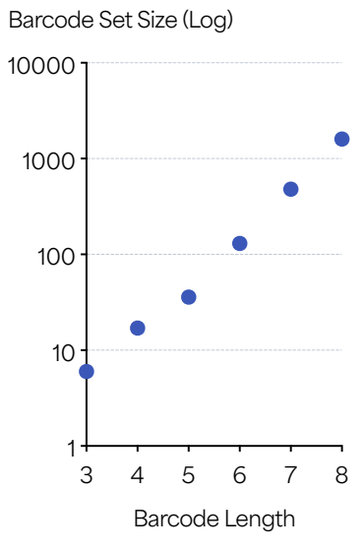


FIGURE 3

Predicted barcode set sizes at L ≥ 3 for barcode length of 3–8.

PROTEIN BARCODES CAN BE FUSED TO FUNCTIONAL ELEMENTS, RECOMBINANTLY EXPRESSED, AND SEQUENCED IN MIXTURES

The previous section discussed sequencing libraries based on C-terminal azido-lysine modified synthetic peptides. To assess their performance when translated, we prepared libraries from barcode peptides expressed recombinantly in *E. coli*. The standard Quantum-Si library preparation workflow involves digesting samples with endoproteinase LysC, producing peptides containing C-terminal lysine residues for conjugation to the K-linker. While this digestion process is necessary for proteomics workflows in which the production of multiple peptides from within the same protein is typically necessary for high-confidence identification of the target, barcoding applications require only the recovery of a single barcoded peptide sequence to identify the associated protein of interest. As such, digestion of samples that contain multiple proteins or protein variants may result in the production of an excess of non-barcoded peptides to barcoded ones. If both barcode and non-barcode peptides are subsequently sequenced on Platinum, this will significantly reduce the detection sensitivity of the assay as the barcode peptides will represent only a minor portion of the available peptides for immobilization into the reaction chambers of the semiconductor chip. We therefore developed an innovative solution to enrich for and specifically conjugate the K-linker to the barcode peptide from a complex sample.

Our novel library preparation approach utilized a Sortase A enzyme mediated transpeptidation reaction[9] (**Figure 4A**). This reaction employs a two-step mechanism in which Sortase A initially catalyzes the cleavage of a substrate, a short C-terminal LPETGG motif, between the threonine and glycine residues via a thioacyl intermediate. The intermediate is then resolved via nucleophilic attack in trans from a glycine containing substrate, resulting in ligation via amide bond formation. To apply this approach in our workflow, we appended the LPETGG motif immediately C-terminal to the protein barcode sequence. In addition, we used a Picolyl-Azide-Gly-Gly-Gly tripeptide to introduce the required nucleophilic attack. This substrate also serves a dual role in the reaction by providing the azide functional group for the subsequent click chemistry reaction to the K-linker.

With the Sortase A mediated transpeptidation enabling the digestion-free conjugation of barcode peptides, our next task was to develop a method to specifically enrich and elute the protein barcodes from a co-expressed protein of interest, as digestion is no longer part of the workflow. We addressed this by incorporating a SUMO tag at the N-terminus of the protein barcode sequence. After the sample is enriched via an affinity tag or a selection method, the SUMO tag enables precise cleavage of the barcode sequence from the protein coding sequence using the SUMO protease ULP1. The enzymatic cleavage results in a free protein barcode starting at the desired NAA of the peptide sequence. When combined with the Sortase A transpeptidation approach, this method establishes an enzymatic library preparation workflow that specifically elutes K-linker conjugated protein barcodes from enriched or selected samples, making them ready for sequencing on Platinum.

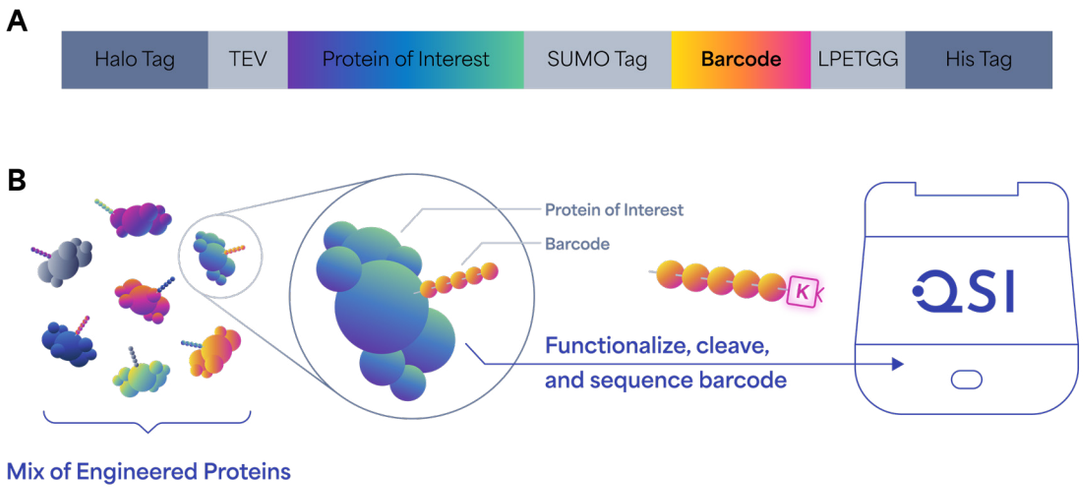


FIGURE 4

Design, generation, and sequencing workflow of barcoded protein libraries. (A) Schematic of barcode construct design, which includes a HaloTag for capturing on magnetic beads, TEV protease site to cleave the protein of interest with the barcode, a SUMO tag to specifically recover barcodes, a Sortase A recognition motif (LPETGG) for the enzymatic ligation of GGG-N3 peptide onto barcodes, and the His tag used for Ni-NTA/Talon purification of barcoded protein with all the tags. (B) Schematic representation of the generation of barcoded protein libraries. Barcoded protein libraries are expressed in vitro or in vivo and subject to screening or selection. Barcodes are then cleaved and conjugated to K-linker before sequencing on Platinum.

We next tested this fully enzymatic library preparation workflow on barcodes recombinantly expressed in *E. coli*, with the goal to demonstrate that these barcodes can be identified with minimal False Discovery Rate (FDR) in a controlled mixture of protein barcodes with different ratios. We prepared and sequenced an 8-protein barcode mixture library in which each protein barcode was added to the mixture at the indicated relative amounts shown in **Table 1**. As seen in **Figure 5**, each of the protein barcodes was identified in the mixture using NGPS, with the number of alignments decreasing with relative abundance. 7 out of 8 peptides in the mix yielded FDR values of < 10%. Of these 7 peptides, 6 yielded FDR values of < 1%, and one with ~5% FDR. This result demonstrates the capability of protein barcodes to undergo recombinant expression, purification, and sequencing on Platinum.

	Relative Abundance
Barcode 6	1x
Barcode 1	0.67x
Barcode 7	0.67x
Barcode 2	0.67x
Barcode 5	0.5x
Barcode 4	0.5x
Barcode 3	0.5x
Barcode 8	0.1x

TABLE 1

Relative abundance of barcodes.

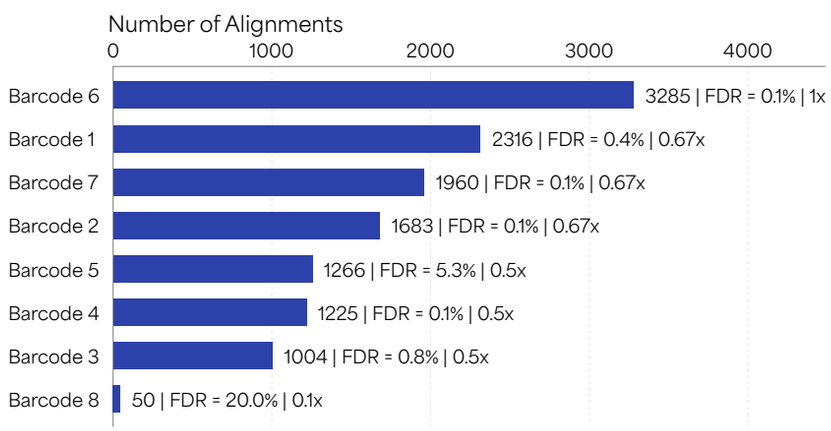


FIGURE 5

Sequencing results on Platinum of an 8-protein-barcode mixture library. Each protein barcode was added to the mixture at controlled relative amounts. The metrics at the end of the bars represent Number of Alignments | FDR | Relative Abundance. For example, Barcode 1 was added to the mixture at a relative abundance of 0.67x, yielding 2,316 alignments with an FDR of 0.4%.

NANOBODY ENRICHMENT SCREENING BY SINGLE-MOLECULE PEPTIDE SEQUENCING

Next, we sought to demonstrate the application of Quantum-Si's protein barcodes in a differential enrichment assay. We focused on nanobodies: single-domain antibodies from camelids that have

gained prominence in various applications, such as targeted drug therapy.[10] Nanobodies are often selected from large synthetic libraries via ultra-high-throughput methods such ribosome or phage display. However, these methods have limitations due to the requirement to maintain the phenotype-genotype linkage. By removing this constraint, protein barcodes offer a streamlined solution for the screening of nanobody variants.

In our experiment, we used a model library with two nanobodies, an anti-MBP Nanobody (targeting maltose binding protein, MBP) and an anti-GFP nanobody (targeting green fluorescent protein, GFP).[6,7] At the C-terminus, the nanobodies were fused with MBP protein barcode and GFP protein barcode, respectively (**Figure 6**). The barcoded nanobodies were then prepared for sequencing based on the targeted sequencing approach described in the previous section. During the affinity selection process, the barcoded nanobody library was added to a solution with GFP-coated magnetic beads. Subsequently, the protein barcodes were proteolytically eluted from the nanobodies bound to the beads. The post-selection library was then sequenced on Platinum alongside a pre-selection control library, and we compared the relative number of each barcode alignment pre- and post-selection.

As seen in the graph in **Figure 6**, the ratio of GFP:MBP protein barcode alignments was approximately 3.18:1 in the pre-selection control library. In the post-selection library, this ratio increased to 1079:1, indicating approximately 300-fold enrichment of the anti-GFP nanobody over the MBP nanobody. This result demonstrates the ability to measure the differential enrichment of nanobody clones in a library through direct single-molecule sequencing of protein-associated barcodes. This experiment also demonstrates the advantage of targeted sequencing in specific applications. Shotgun sequencing of these two nanobodies would have produced 11 individual digested peptides (9 from the two nanobodies + the two barcodes). In contrast, our targeted sequencing approach, which increases barcode sequencing capacity approximately 5-fold, was applied to two small proteins (15 kDa each). This effect would be even more pronounced in applications targeting larger proteins of interest, in which shotgun sequencing would produce substantially more total peptides.

Protein Barcodes for Next-Generation Protein Sequencing™

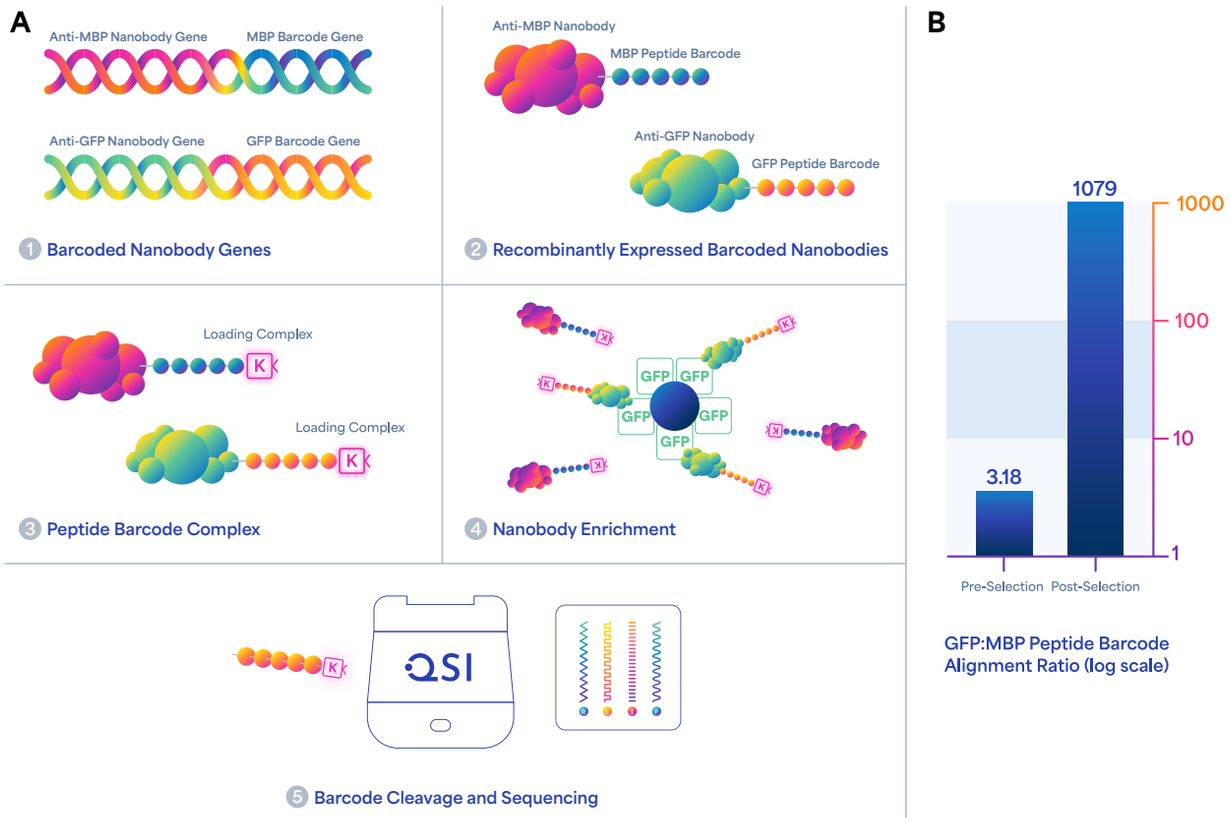


FIGURE 6

Workflow for enrichment and sequencing of barcoded nanobodies. (A) Schematic representation of the differential enrichment of an anti-GFP nanobody by peptide barcoding. Model anti-GFP and anti-MBP nanobody genes were encoded with unique protein barcodes and expressed recombinantly. The proteins were purified, labeled with an azide tag, and conjugated to macromolecular linkers. The nanobodies were enriched using GFP immobilized on magnetic beads. After selection, the barcodes were eluted by proteolysis and sequenced on Platinum. (B) Bar graph quantification of pre- and post-selection barcoded nanobody libraries indicating a > 300-fold enrichment for the anti-GFP nanobody post-selection based on the ratio of GFP to MBP protein barcode alignments.

CONCLUSION

Protein barcodes hold immense promise across various fields such as proteomics, synthetic biology, protein engineering, and drug discovery. In this application note, we demonstrated the development and utilization of protein barcodes with NGPS on Platinum. Our work reveals that short peptide sequences yield information-rich and distinctive kinetic signatures that can be

accurately recognized. Moreover, our optimized protein barcode designs ensure reliable relative quantification. The introduction of an enzymatic approach to targeted sequencing allows barcodes to be easily and specifically retrieved from complex matrices such as an *E. coli* lysate, a crucial capability for future applications. To explore the applicability of protein barcodes to nanobody engineering, we used protein barcodes to reliably monitor the enrichment of an anti-GFP nanobody from a model selection. By offering accessible and user-friendly instrumentation for decoding protein barcodes, Quantum-Si's Platinum stands ready to facilitate widespread adoption and maximize the potential impact of protein barcodes, paving the way for a new era of innovation in proteomics and beyond.

QUANTUM SI™

29 Business Park Drive, Branford, CT 06405

www.quantum-si.com | 866.688.7374

Quantum-Si's Next Generation Protein Sequencing™ Technology Enables Rapid and Accurate Distinction of Variants of the SARS-CoV-2 Virus

May 5, 2023

SUMMARY

Accurate and rapid identification of SARS-CoV-2 variants is crucial for effective surveillance and disease monitoring. Furthermore, patients with long COVID may still have residual spike protein in their blood. Distinguishing variants in long COVID patients via protein sequencing could lead to new understandings and treatment of the disease.[1] In this application note, we present a method utilizing Quantum-Si's protein sequencing technology on the Platinum® instrument to distinguish the Alpha, Delta, and Omicron variants based on differences in the amino acid sequence of their spike proteins.

INTRODUCTION

Since its emergence in December 2019, the SARS-CoV-2 virus has spread rapidly across the globe, leading to millions of infections and deaths. The emergence of new variants has raised concerns about the effectiveness of current countermeasures.[2,3] Currently, three dominant variants of concern, Alpha, Delta, and Omicron, have been identified by the World Health Organization (WHO) due to their increased transmission rates and potential to cause more

Q-Si Technology

Quantum-Si's benchtop Platinum® instrument enables protein sequencing from biological samples in a simple user-friendly workflow. Our technology utilizes dye-tagged N-terminal amino acid recognizers and semiconductor chip technology to detect the binding characteristics and binding order of N-terminal amino acids, resulting in unique kinetic signatures that can be used to differentiate and identify amino acid residues and post-translational modifications. A more detailed overview of the workflow and technology can be found in our Science Paper.

severe disease. These variants have different mutations in their genetic sequence, leading to differences in their characteristics and behavior.

Quantum-Si provides an accessible protein sequencing technology to identify the unique protein sequences of a virus and distinguish between different variants. The technology relies on recognition of single N-terminal amino acids (NAAs) with different NAA recognizers and sequential cleavage of NAAs with aminopeptidases. The information-rich output from protein sequencing enables the detection of differences among protein variants, such as single amino acid differences in peptide sequences and the presence or absence of specific peptides during infection and in long COVID patients with residual viral proteins. This feature makes the platform highly effective in detecting the various changes at the protein level that result from genetic differences among variant viral strains.

METHODOLOGY & WORKFLOW

The receptor-binding domain (RBD) regions of the spike proteins in the SARS-CoV-2 variants are sequenced on Quantum-Si's Platinum instrument using our library preparation and real-time sequencing workflow as previously described.[4] The method entails digesting proteins into peptide fragments then linking them to macromolecular linkers at the C-terminus. The peptides are then immobilized on Quantum-Si's semiconductor chip, exposing the N-termini for sequencing. Dye-labeled recognizers bind on and off to NAAs, resulting in pulsing patterns with characteristic kinetic properties. Individual NAAs are removed sequentially by aminopeptidases in solution, uncovering subsequent amino acids for recognition. Fluorescence lifetime, intensity, and kinetic data are collected in real time and analyzed with Cloud-based software to identify the peptide sequence and corresponding protein.

The trace-level output from protein sequencing on Quantum-Si's Platinum instrument consists of distinct pulsing regions called recognition segments (RSs). Each RS corresponds to a time interval between aminopeptidase cleavage events during which a recognizer binds on and off to the exposed NAA. Point substitutions to a target NAA can lead to binding events from a different recognizer or other changes in recognition pattern, resulting in a characteristic change in the binding kinetics observed as pulse durations

REFERENCES

1. Cheung CCL, Goh D, Lim X, et al. Residual SARS-CoV-2 viral antigens detected in GI and hepatic tissues from five recovered patients with COVID-19. *Gut*. 2022;71(1):226-229. doi:10.1136/gutjnl-2021-324280
2. Bernal JL, Andrews N, Gower C, et al. Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. *New England Journal of Medicine*. 2021;385(7). doi:10.1056/nejmoa2108891
3. Andrews N, Stowe J, Kirsebom F, et al. Covid-19 Vaccine Effectiveness against the Omicron (B.1.1.529) Variant. *New England Journal of Medicine*. 2022;386(16). doi:10.1056/nejmoa2119451
4. Reed BD, Meyer MJ, Abramzon V, et al. Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device. *Science*. 2022;378(6616):186-192. doi:10.1126/science.abo7651

(PD). Furthermore, substitutions at lysine positions can change the digestion pattern of the protein during library preparation by removing LysC cleavage sites, leading to the absence of peptides that would otherwise be present in the library and thus detected during sequencing.

RESULTS & DISCUSSION

ALPHA VARIANT CAN BE DIFFERENTIATED BY THE L452R MUTATION

We first sought to demonstrate the differentiation of the Alpha variant from the Delta and Omicron variants by single-molecule protein sequencing. We focused on a key mutation L452R that occurs in the sequences of Delta and Omicron but not Alpha.

We obtained the RBDs of the spike proteins of the SARS-CoV-2 variants Alpha (Sino Biological, Beijing, China; #40592-V02H1), Delta (Sino Biological, Beijing, China; #40592-V08H90), and Omicron (Sino Biological, Beijing, China; #40592-V08H130).

We sequenced each protein on Quantum-Si's Platinum instrument using five recognizers—PS610 (F, Y, W), PS1165 (A, S), PS1220 (R), PS1223 (L, I, V), and PS1259 (N, Q)—and analyzed data to identify RSs, determine the mean PD of each RS, and characterize the kinetic signature of the peptides of interest. For the Alpha variant, we focused on the peptide VGGNYNYLYRLFRK, while for Delta and Omicron, we investigated the peptide VGGNYNYRYRLFRK.

The two peptides displayed distinguishable patterns due to the distinct kinetic influences of the L452R mutation on recognizer binding (see example traces in Fig 1). The leucine in the Alpha variant was observed via the long PD recognition event upon binding with recognizer PS1223, whereas the arginine in the Delta and Omicron variants was recognized by PS1220 and observed as a shorter PD recognition event. This clear kinetic difference demonstrates that the L452R mutation can be detected with Quantum-Si's Platinum instrument, effectively differentiating the Alpha variant from Delta and Omicron.

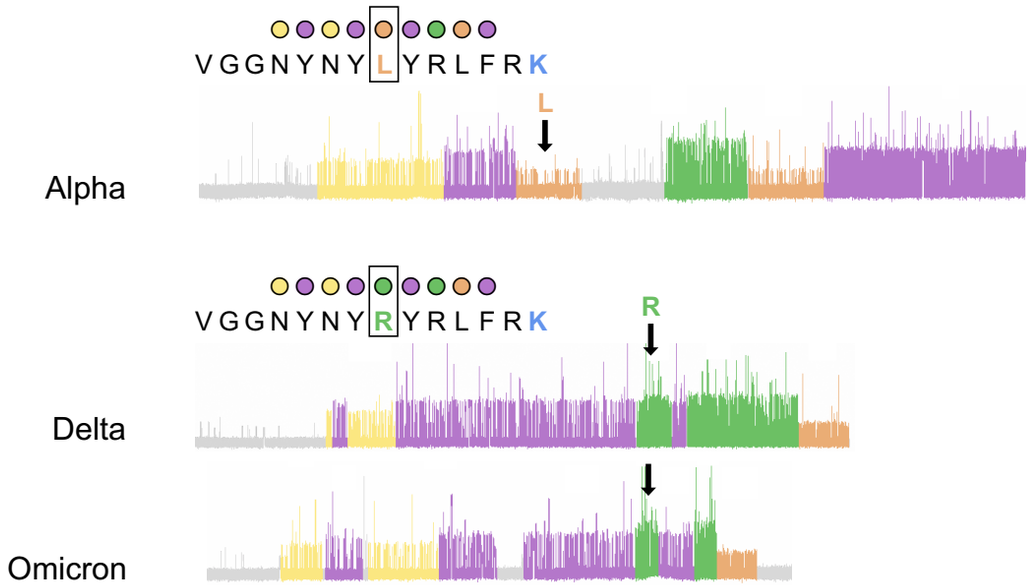


FIGURE 1

Example protein sequencing traces for the peptide where the L452R mutation occurs.

The two peptides displayed distinguishable patterns due to the distinct kinetic influences of the L452R mutation on recognizer binding (see example traces in Fig 1). The leucine in the Alpha variant exhibited was observed via the long PD recognition event upon binding with recognizer PS1223, whereas the arginine in the Delta and Omicron variants was recognized by PS1220 and observed as a shorter PD recognition event. This clear kinetic difference demonstrates that the L452R mutation can be detected with Quantum-Si's Platinum instrument, effectively differentiating the Alpha variant from Delta and Omicron.

OMICRON VARIANT CAN BE DIFFERENTIATED BY PEPTIDE I418-K424 DUE TO THE K417N MUTATION

Next, we sought to demonstrate the differentiation of the Omicron variant from the Alpha and Delta variants by focusing on the detection of peptide I418-K424 (IADYNYK). In the Alpha and Delta variants, this peptide is present due to LysC cleavage that occurs at residue K417 during library preparation. In the Omicron variant, the K417N mutation results in the loss of cleavage, leading to the absence of this peptide (Fig 2A).

We sequenced the recombinant spike RBD for each variant on Quantum-Si's Platinum instrument and analyzed the sequencing data to detect the presence of peptide I418-K424. For the Alpha and Delta variants, this peptide was readily detected and displayed a characteristic pulsing pattern (see example trace in Fig 2B). This peptide was not detected in the sequencing output from the Omicron variant. The absence of peptide I418-K424 distinguishes the Omicron variant from Alpha and Delta.

2A

2B

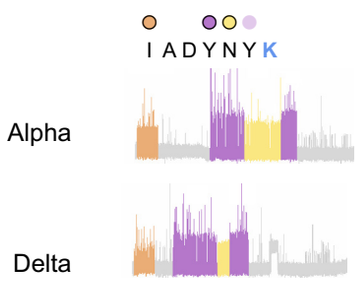


FIGURE 2A

The K417N mutation in the Omicron variant results in the loss of LysC cleavage and absense of the peptide I418-K424 (IADYNYK).

FIGURE 2B

Example protein sequencing trace for peptide I418-K424 in the Alpha and Delta variants. This peptide was not detected in the Omicron variant.

These results indicate that the three variants Alpha, Delta, and Omicron can be distinguished using Quantum-Si's Platinum instrument. For the Alpha variant, we detected a leucine at position 452 and the presence of peptide I418-K424. For Delta, we detected the L452R mutation and peptide I418-K424. For Omicron, we detected the L452R mutation but not peptide I418-K424. These results are summarized in the table below.

	Position 452	Peptide I418-K424
Alpha	L	Detected
Delta	R	Detected
Omicron	R	Not detected

CONCLUSION

In this application note, we accurately distinguished three SARS-CoV-2 variants, Alpha, Delta, and Omicron using Quantum-Si's Platinum instrument. As the virus continues to mutate and new variants emerge, it is critical that COVID-19 variants be identified and tracked rapidly so that scientists can understand their potential impact on public health and develop effective treatments and vaccines. Furthermore, detecting the sequence of residual viral proteins in long COVID patients will aid in understanding the long-term effects of the disease. The ability to directly distinguish the variants of SARS-CoV-2 virus via Quantum-Si's next-generation protein sequencing platform can potentially accelerate the diagnosis and treatment of COVID-19 patients and support surveillance efforts to contain the spread of the virus.

Detecting Arginine Post-Translational Modifications in p38MAPKα and Vimentin Proteins Using Quantum-Si's Next-Generation Protein Sequencing Technology

SUMMARY

Proteins undergo a diverse array of post-translational modifications (PTMs) to their amino acid side chains, which strongly affect protein function and mediate intricate cellular events. Measuring the diversity, dynamics, and functional consequences of PTM states of proteins across the proteome is essential to understanding the role of proteins in health and disease. However, the diversity of proteoforms in the human proteome remains largely unmapped, and the discovery and detection of PTMs is highly challenging.[1] New, more sensitive methods for PTM detection will greatly aid biomarker discovery, drug discovery, and the development of precise and personalized approaches to medicine.

Modifications of the arginine side chain are of biomedical interest. Methylation and citrullination of arginine residues in human proteins have been shown to play key roles in disease states such as cardiovascular disease, autoimmune disease, and cancer.[2-6] Here, we demonstrate the application of Platinum®, Quantum-Si's next-generation protein sequencing technology,[7] to the detection of arginine methylation and citrullination in p38MAPKα and Vimentin proteins with single-molecule resolution and sensitivity.

Quantum-Si's protein sequencing technology offers a sensitive platform for PTM detection and discovery with the power to address the critical need for accessible methods to study the role of PTMs in human health and disease.

Q-SI TECHNOLOGY

Quantum-Si's benchtop Platinum® instrument enables protein sequencing from biological samples in a simple user-friendly workflow. Our technology utilizes dye-tagged N-terminal amino acid recognizers and semiconductor chip technology to detect the binding characteristics and binding order of N-terminal amino acids, resulting in unique kinetic signatures that can be used to differentiate and identify amino acid residues and PTMs. A more detailed overview of the workflow and technology can be found in our Science Paper.

p38MAPKα: p38 mitogen-activated protein kinase α

For Research Use Only. Not for use in diagnostic procedures.

INTRODUCTION

Arginine plays a critical role in protein structure and function due to the unique properties of the guanidinium group that forms the terminus of its side chain. This group is both positively charged and capable of forming extended hydrogen bond networks and cation-π interactions with other amino acids and with nucleic acids. Arginine, therefore, often mediates key interactions between protein binding partners or between proteins and DNA.

The two most common arginine PTMs, dimethylation and citrullination, alter the arginine side chain and change its properties (Figure 1), potentially resulting in important downstream effects on cellular processes. Dimethylation retains arginine's positive charge but increases its size and hydrophobicity and blocks hydrogen bond formation. Citrullination eliminates arginine's positive charge, resulting in a neutral side chain, greatly impacting protein conformation and function. Dimethylation and citrullination of arginine are carried out by enzymes and may be part of the normal regulation of cellular processes or involved in disease states. Arginine dimethylation is catalyzed by protein arginine methyltransferases (PRMTs). PRMTs transfer two methyl groups either asymmetrically onto the same nitrogen atom, resulting in asymmetric dimethyl arginine (ADMA), or symmetrically onto opposite nitrogen atoms, resulting in symmetric dimethyl arginine (SDMA). Arginine citrullination is catalyzed by protein arginine deiminases (PADs). PADs carry out the hydrolysis of arginine's positively-charged guanidinium group, resulting in a neutral ureido group.

Arginine PTMs have emerged as important targets in biomedical research. Methylated arginine residues and their respective PRMTs have been implicated in diseases such as cardiovascular disease and cancers.[2,3] Research has demonstrated that arginine citrullination is critically involved in immune system function, skin keratinization, myelination, and gene expression regulation.[4-6] Notably, the removal of arginine's positive charge in some cases can cause proteins to activate the immune system, contributing to autoimmune diseases.[5]

REFERENCES

1. Aebersold R, Agar JN, Amster IJ, et al. How many human proteoforms are there?. Nat Chem Biol. 2018;14(3):206-214. doi:10.1038/nchembio.2576
2. Couto E Silva A, Wu CY, Citadin CT, et al. Protein Arginine Methyltransferases in Cardiovascular and Neuronal Function. Mol Neurobiol. 2020;57(3):1716-1732. doi:10.1007/s12035-019-01850-z
3. Chen L, Liu S, Tao Y. Regulating tumor suppressor genes: post-translational modifications. Signal Transduct Target Ther. 2020;5(1):90. Published 2020 Jun 10. doi:10.1038/s41392-020-0196-9
4. James EA, Pietropaolo M, Mamula MJ. Immune Recognition of α-Cells: Neoepitopes as Key Players in the Loss of Tolerance. Diabetes. 2018;67(6):1035-1042. doi:10.2337/dbi17-0030
5. Nguyen H, James EA. Immune recognition of citrullinated epitopes. Immunology. 2016;149(2):131-138. doi:10.1111/imm.12640
6. Wegner N, Lundberg K, Kinloch A, et al. Autoimmunity to specific citrullinated proteins gives the first clues to the etiology of rheumatoid arthritis. Immunol Rev. 2010;233(1):34-54. doi:10.1111/j.0105-2896.2009.00850.x
7. Reed BD, Meyer MJ, Abramzon V, et al. Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device. Science. 2022;378(6616):186-192. doi:10.1126/science.abo7651
8. Witze ES, Old WM, Resing KA, Ahn NG. Mapping protein post-translational modifications with mass spectrometry. Nat Methods. 2007;4(10):798-806. doi:10.1038/nmeth1100

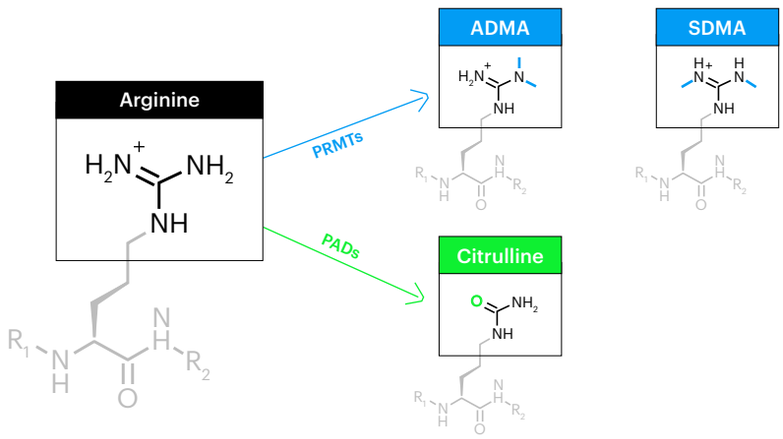


FIGURE 1

Post-translational modifications of arginine. The arginine side chain consists of a positively-charged guanidinium group at the terminus of a flexible aliphatic chain. In arginine dimethylation, protein arginine methyltransferases (PRMTs) transfer two methyl groups—either asymmetrically to the same nitrogen atom (ADMA) or symmetrically onto opposite nitrogen atoms (SDMA). These modifications increased size and hydrophobicity and block hydrogen bonding. In arginine citrullination, protein arginine deiminases (PADs) carry out the hydrolysis of arginine's positively-charged guanidinium group, resulting in a neutral ureido group (a transformation referred to as deimination). This transformation results in a negligible mass increase of 0.9840 Da, though the loss of positive charge can dramatically alter protein conformation and function.

CHALLENGES FOR THE DETECTION OF ARGININE PTMS

Research into these arginine PTMs has been particularly challenging because they are difficult to detect and differentiate with current proteomic methods.[8] Mass spectrometry (MS) is the most frequently utilized tool for detecting protein PTMs. However, MS cannot easily distinguish between ADMA and SDMA because they are constitutional isomers with identical masses.[9] Likewise, deimination of arginine to citrulline results in a negligible mass increase of 0.9840 Da. This mass difference can easily be confused with a ^{13}C isotope or misinterpreted as deamidation of nearby asparagine or glutamine residues.[10] In addition, MS techniques for arginine PTM detection, specifically arginine modifications, require highly specialized knowledge and training and advanced analysis methods.

Enzyme-linked immunosorbent assay (ELISA), another common method for PTM detection, uses antibodies specifically generated to detect a modified protein of interest.

REFERENCES

9. Afjehi-Sadat L, Garcia BA. Comprehending dynamic protein methylation with mass spectrometry. Curr Opin Chem Biol. 2013;17(1):12-19. doi:10.1016/j.cbpa.2012.12.023

10. Vitorino R, Guedes S, Vitorino C, Ferreira R, Amado F, Van Eyk JE. Elucidating Citrullination by Mass Spectrometry and Its Role in Disease Pathogenesis. J Proteome Res. 2021;20(1):38-48. doi:10.1021/acs.jproteome.0c00474

11. Larsen SC, Sylvestersen KB, Mund A, et al. Proteome-wide analysis of arginine monomethylation reveals widespread occurrence in human cells. Sci Signal. 2016;9(443):rs9. Published 2016 Aug 30. doi:10.1126/scisignal.aaf7329

12. Egelhofer TA, Minoda A, Klugman S, et al. An assessment of histone-modification antibody quality. Nat Struct Mol Biol. 2011;18(1):91-93. doi:10.1038/nsmb.1972

13. Jeong HJ, Lee SJ, Lee HJ, et al. Prmt7 promotes myoblast differentiation via methylation of p38MAPK on arginine residue 70. Cell Death Differ. 2020;27(2):573-586. doi:10.1038/s41418-019-0373-y

14. Van Steendam K, Tilleman K, De Ceuleneer M, De Keyser F, Elewaut D, Deforce D. Citrullinated vimentin as an important antigen in immune complexes from synovial fluid of rheumatoid arthritis patients with antibodies against citrullinated proteins. Arthritis Res Ther. 2010;12(4):R132. doi:10.1186/ar3070

15. Li FJ, Surolia R, Li H, et al. Citrullinated vimentin mediates development and progression of lung fibrosis. Sci Transl Med. 2021;13(585):eaba2927. doi:10.1126/scitranslmed.aba2927

Although arginine PTMs are estimated to be widespread in human cells,[11] commercially available antibodies against arginine PTMs are limited to specific sites on a few highly studied proteins. The requirement to generate new antibodies, along with complex workflows, high costs, limited antibody reproducibility,[12] and other challenges associated with ELISA assay development, will likely hinder discovery and further study of novel arginine PTM sites.

Continued development toward novel methods is needed to failitate direct detection of arginine PTMs in proteins. Next-generation protein sequencing on Quantum-Si's Platinum® offers a single-molecule sequencing solution to the detection of ADMA, SDMA, and citrulline that is not based on mass-to-charge ratio or antibody specificity, but rather on the kinetic signature of binding between recognizers and N-terminal amino acids (NAAs).

Quantum-Si's Platinum instrument harnesses the power of single-molecule protein sequencing to gain insights into these PTMs with single molecule resolution, overcoming current technological gaps, and providing direct detection of arginine PTMs.

METHODOLOGY & WORKFLOW

PTM detection on Platinum consists of isolating peptides and sujecting them to a single-molecule protein sequencing reaction.[7] Proteins are first digested into peptide fragments and conjugated C-terminally to macromolecular linkers. The peptide complexes are immobilized on Quantum-Si's semiconductor chip, resulting in single peptide molecules with exposed N-termini ready for sequencing. During the sequencing reaction, the surface-immobilized peptides are exposed to a solution containing dye-labeled NAA recognizers that bind on and off to their cognate NAAs with characteristic kinetic properties. Aminopeptidases in solution sequentially remove individual NAAs to expose subsequent amino acids for recognition. Fluorescence lifetime, intensity, and kinetic data are collected in real time and analyzed to determine amino acid sequence and PTM content.

Sequencing profiles of peptides are visualized as kinetic signature plots—simplified trace-like representations of the time course of complete peptide sequencing containing the median pulse duration (PD) for each recognition segment (RS) and the average

duration of each RS and non-recognition segment (NRS). When recognizers bind to NAAs, they also make important contacts with nearby downstream residues in the sequence, influencing the average PD of binding events between recognizers and target peptides. This kinetic sensitivity to nearby downstream residues provides a wealth of information on peptide sequence composition and is extremely beneficial for mapping traces to their proteins of origin. A summary of the workflow for sequencing of peptides and detection of PTMs is presented in Figure 2.

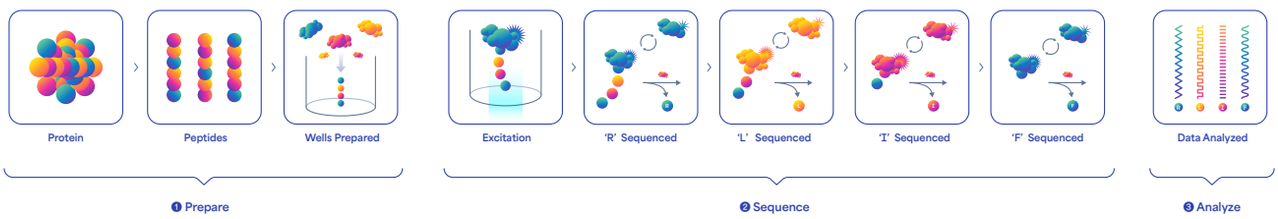


FIGURE 2

The Platinum Protein Sequencing Workflow. Proteins are digested and functionalized into individual peptides that are immobilized on a semiconductor chip with N-terminal amino acid (NAA) recognizers and aminopeptidases. Binding kinetics and order of binding are then used to determine the peptide sequence. Analysis software is used to identify proteins using the sequencing information.

RESULTS & DISCUSSION

DETECTION OF ARGININE DIMETHYLATION

We seek to demonstrate the detection and differentiation of arginine, ADMA, and SDMA using next-generation protein sequencing with single-molecule resolution. We focused on a key segment of the signaling protein p38MAPKα . Dimethylation of arginine residue 70 of p38MAPKα in myoblast cells by PRMT7 is a critical regulatory step in the activation of myoblast differentiation in humans.[13]

We generated synthetic peptides corresponding to residues 69 to 76 of p38MAPKα in three versions containing either arginine, ADMA, or SDMA at position 2: YRELRLLK, YR_{ADMA} ELRLLK,and YR_{SDMA} ELRLLK. We sequenced each peptide on the Quantum-Si Platinum® instrument using three recognizers with specificity for 7 NAAs (F, Y, W, L, I, V, and R) and analyzed data to identify RSs, determine the mean PD of each RS, and characterize the kinetic signature of each peptide. Each peptide displayed a distinguishable pattern due to the distinct kinetic influences of arginine,

ADMA, and SDMA on recognizer binding (see example traces in Figure 3).

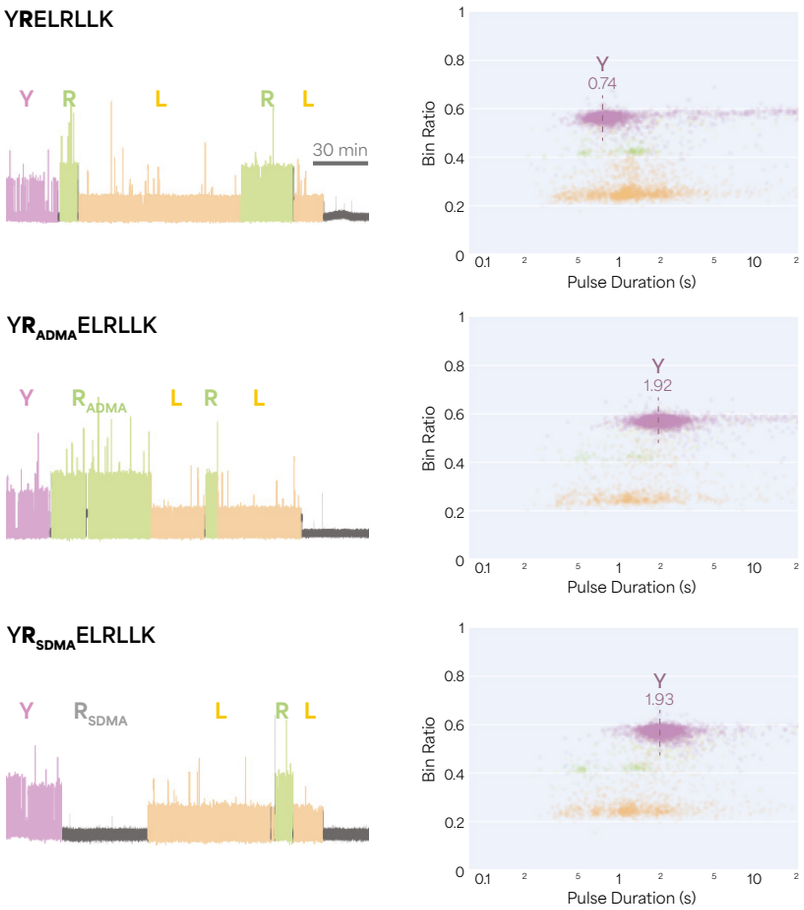


FIGURE 3

Kinetic signatures distinguish three synthetic p38MAPKα-derived peptides containing arginine (top), ADMA (middle), and SDMA (bottom) at position 2. Full length peptide sequences are indicated. Left: Example protein sequencing traces. Arginine and ADMA residues exhibited binding with the R recognizer, whereas SDMA exhibited no binding. Right: Kinetic profile from sequencing experiments of the synthetic peptides. The mean PD of tyrosine recognition increased for ADMA and SDMA.

Arginine and ADMA residues exhibited binding with the R recognizer, whereas SDMA exhibited no binding (Figure 3). This result indicates that symmetric dimethylation of arginine—in contrast to asymmetric dimethylation—reduces the affinity of the R recognizer for N-terminal arginine, providing a clear kinetic difference between these isomeric arginine PTMs. Quantum-Si's NAA recognizers contact residues at position 2 and 3 from the N-terminus when they bind to their target NAAs; therefore, modification of

these downstream residues can influence recognizer binding affinity.[7] We observed a strong influence of arginine dimethylation on recognition of the upstream tyrosine residue in these peptides by the FYW recognizer. The median PD of tyrosine recognition increased from 0.74 s for YRE to 1.92 s and 1.93 s for $YR_{ADMA}E$ and $YR_{SDMA}E$, respectively (Figure 3). The influence that these dimethylated arginine residues have on the recognition of preceding NAAs serves as a powerful feature of protein sequencing with single-molecule sensitivity and precision. These results demonstrate the capacity for unprecedented sensitivity in detection of arginine dimethylation using Platinum's next-generation protein sequencing technology.

DETECTION OF ARGININE CITRULLINATION

We next sought to demonstrate that differential binding kinetics can be used to rapidly differentiate citrullinated arginine residues from native arginine residues. We generated synthetic peptides corresponding to residues 47 to 62 of MLRV in two versions containing either arginine or citrulline at position 2: NDLRDTFAALGRVNVK and NDLCitDTFAALGRVNVK. We sequenced each peptide on the Quantum-Si Platinum instrument using five recognizers with specificity for 11 NAAs (F, Y, W, L, I, V, R, A, S, N, and Q). Each peptide displayed a highly distinguishable kinetic signature due to the influence of the different arginine and citrulline side chains on recognition (Figure 4). Citrullination eliminated N-terminal arginine recognition by the R recognizer, as demonstrated by the example traces (Figure 4) and the level of coverage (Figure 4, green box). Citrullination at position 4 also resulted in a large increase in the median PD of recognition of the N-terminal leucine located at the preceding position by the LIV recognizer. Median PD increased from 0.44 s for LRF to 0.72 s for LCitF (Figure 4, orange box). These results demonstrate the capability of Quantum-Si next-generation protein sequencing technology in detection and digital quantification of arginine citrullination with the capacity to overcome the challenges facing conventional methods.

DETECTION OF ARGININE CITRULLINATION IN VIMENTIN

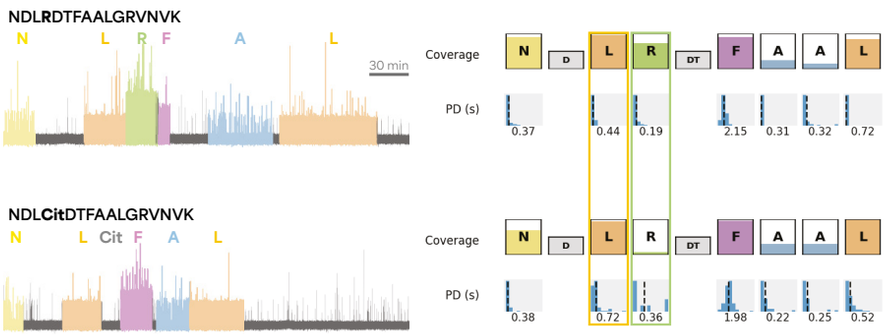


FIGURE 4

Kinetic signatures distinguish two synthetic MLRV-derived peptides containing arginine (top) and citrulline (bottom) at position 4. Full length peptide sequences are indicated. Left: Example protein sequencing traces. Arginine exhibited binding with the R recognizer, whereas Cit exhibited no binding. Right: Coverage and mean PD of the recognized residues. The mean PD of leucine recognition increased for Cit (orange box), while the coverage of the arginine was eliminated (green box).

We next sought to demonstrate the detection of arginine citrullination in vimentin, a type III intermediate filament protein that provides structural support to cells and is involved in maintaining cell shape and integrity. Citrullinated vimentin has been implicated in various autoimmune diseases, such as rheumatoid arthritis, as it can trigger an immune response leading to inflammation and tissue damage.[15,16] This protein when digested with LysC produces the peptide VRFLEQQNK in which citrullination happens at position 2.

We generated two synthetic peptide sequences containing either arginine or citrulline at this position–VRFLEQQNK and VCitFLEQQNK–and sequenced each peptide on the Quantum-Si Platinum® instrument using five recognizers with specificity for 11 NAAs (F, Y, W, L, I, V, R, A, S, N, and Q). Each peptide displayed a highly distinguishable kinetic signature as a result of the different binding kinetics of arginine and citrulline side chains (Figure 5). For the VRFLEQQNK peptide, the R residue was detected, whereas the preceding V residue was not detected due to the V residue's mean PD being too short for recognition. For the VCitFLEQQNK peptide, citrullination eliminated the recognition of the R residue and resulted in the recognition of the preceding V residue due to an increase in the mean PD.

Next, we digested and sequenced the full vimentin protein with the synthetic peptide VCitFLEQQNK spiked in at equal peptide concentration. The spike-in citrullinated peptide can be clearly distinguished from the non-citrullinated peptide of the vimentin protein (VRFLEQQNK), as each peptide displayed a distinct kinetic signature (Figure 5). Example sequencing traces demonstrated that for peptide VRFLEQQNK, the first V residue was not detected, while the second R residue was clearly detected. In contrast, for the spike-in peptide VCitFLEQQNK, citrullination led to the recognition of the first V residue, while the recognition of the second R residue was eliminated. These results demonstrate that Quantum-Si Platinum® sequencing instrument can detect arginine citrullination in vimentin, an important protein in autoimmune diseases.

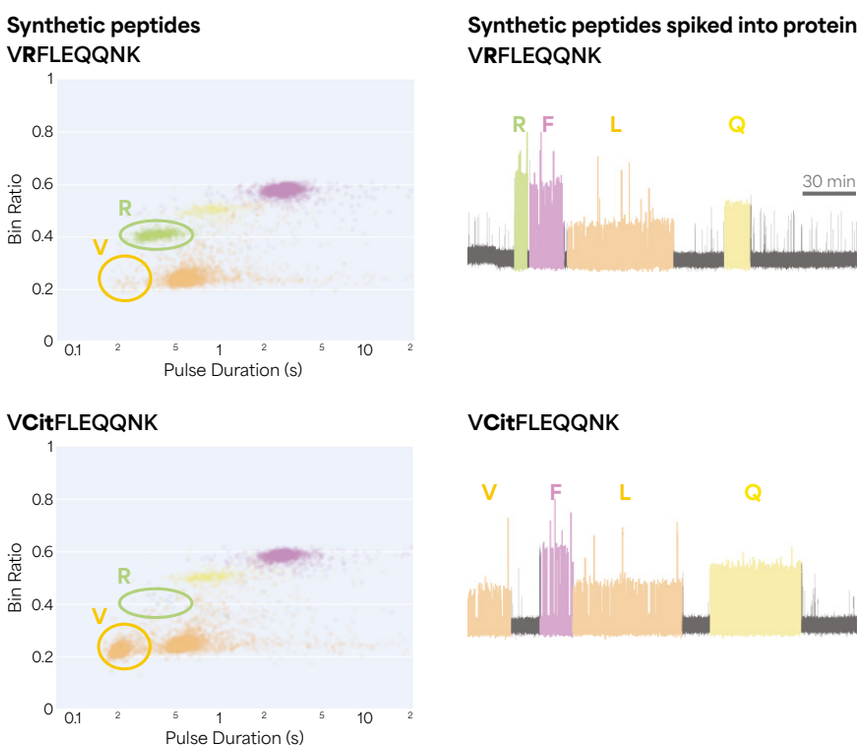


FIGURE 5

Kinetic signatures distinguish peptides of vimentin protein containing arginine and citrulline at position 2. Full length peptide sequences are indicated. Left: Kinetic profile from sequencing experiments for two synthetic peptides of vimentin. Right: Example protein sequencing traces of the non-citrullinated peptide of vimentin (top) and the spike-in synthetic citrullinated peptide (bottom).

CONCLUSION

In this application note, we directly detect arginine PTMs using Quantum-Si's next-generation protein sequencing platform. Arginine PTMs play important roles in human health and disease but have been challenging to study. Current proteomic methods such as mass spectrometry and ELISA have been capable of indirect identification of these arginine PTMs using highly specialized techniques or limited to a small set of specific proteins on the basis of antibody availability and other challenges. The ability to directly detect PTMs via next-generation protein sequencing demonstrated here offers great potential for accelerated biomedical research and for a wide range of commercial applications in drug discovery and biomarker development.

Featured products

Platinum® Next-Generation Protein Sequencer™ by Quantum-Si

Easy-to-use, accessible benchtop single-molecule protein sequencing platform that empowers any lab, anywhere to obtain critical proteomic insights quickly and efficiently from the convenience of your lab.

- **Convenient** — the benchtop instrument and collaboration-friendly cloud-based analysis software easily fit into the standard laboratory environment.
- **Accessible** — with simple, straightforward workflows and automated cloud-based analysis software, any lab can generate protein sequence and evaluate variation.
- **Insight-generating** — single amino acid variation and PTMs can be detected directly by sequencing without the need to acquire or generate highly specific affinity reagents directed against these variants and modifications.