QuantumSi

Comparison of Protein Sequencing Analysis of CDNF on Platinum™ and Mass Spectrometry

04 MAY 2023

INTRODUCTION

The rapid growth of the field of proteomics since the early 2000s has been closely tied to the technological advancements in mass spectrometry. Mass spectrometry analysis is used to identify proteins and is commonly performed as a service at core facilities due to the high cost, space, and expertise required. A typical workflow for protein identification via mass spectrometry involves digestion of proteins into peptides followed by LC-MS/MS on a high-resolution analyzer. In MS1, peptides are separated by reversed phase chromatography and the mass-to-charge ratio is detected as intact peptides elute over time. In MS2, peptides are fragmented by collision-induced dissociation, generating spectra that can be analyzed to determine peptide sequenc.[1] Complex software is used to map the spectra to a large database of simulated spectra from the proteome generated via in silico protein digestion to identify peptides and determine the protein of origin.

While mass spectrometry has become the gold standard for protein identification, several confounding factors can limit the unambiguous, proteome-wide mapping of peptides. For example, missed or unanticipated cleavages and post-translational modifications (PTMs) can lead to peptides that are not detected by the search algorithm. Amino acids of identical or similar masses also present challenges in standard mass spectrometry workflow. Examples include the differentiation of isoleucine and leucine, asymmetric dimethylarginine and symmetric dimethylarginine, and trimethyllysine and acetyllysine.

REFERENCES

1 Neagu AN, Jayathirtha M, Baxter E, Donnelly M, Petre BA, Darie CC. Applications of Tandem Mass Spectrometry (MS/MS) in Protein Analysis for Biomedical Research. Molecules. 2022 Apr 8;27(8):2411.

2 Reed BD et al. Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device. Science. 2022 Oct 14;378(6616):186-192.

In addition, the workflow is laborious and requires expensive capital equipment and advanced expertise in analysis to properly identify the protein sequence. Researchers often resort to simpler proteomic techniques—such as western blots—to detect protein changes, with the tradeoff of not uncovering deeper insights at the amino acid and peptide level of proteins.

Quantum-Si's single-molecule protein sequencing workflow on Platinum overcomes these challenges by offering an accessible and convenient benchtop instrument with Cloud based analysis software that enables researchers to discover deeper proteomic insights without the need for expensive equipment and expertise. Proteins are digested with a protease to generate peptide libraries and immobilized on a semiconductor chip in as little as 2-3 hours of hands-on time. Protein sequencing occurs on the Platinum™ instrument in 10 hours or less without the need for complex cyclical chemistry and fluidics.
Sequencing data is automatically and securely transferred to a Cloud-based software environment for analysis. Protein identification output is provided without the need for expert analyses, enabling researchers to quickly make decisions about their protein samples.

Protein sequencing on Platinum offers several advantages for protein identification relative to tandem mass spectrometry, including the convenience of a benchtop sequencer without the need to send samples to a core facility, the simplicity of the workflow without the need for advanced expertise, and the ability to gain deeper insights into the proteome thanks to the information-rich binding signatures of N-terminal amino acid (NAA) recognizers.

To compare protein sequencing to tandem mass spectrometry for the purpose of identification of an isolated protein, we sequenced the protein CDNF on Platinum and sent the same protein to a mass spectrometry core facility for analysis.

Methodology and Workflow

Proteins are sequenced on Quantum-Si's Platinum instrument using Library Preparation Kits, Protein Sequencing Kits, and our real-time dynamic sequencing workflow as previously described.[2] Briefly, proteins are digested into peptide fragments and conjugated to macromolecular linkers. The conjugated peptides are then immobilized on Quantum-Si's semiconductor chip with exposed N-termini for sequencing.

Dye-labeled recognizers bind on and off to NAAs, generating pulsing patterns with characteristic fluorescence and kinetic properties. Regions corresponding to NAA recognition are termed recognition segments (RSs). 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 distinguish amino acids of the peptide sequence.

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 RS and the average duration of each RS and non-recognition segment (NRS). As described previously, when recognizers bind to NAAs, they also make important contacts with nearby downstream residues that influence 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 from peptides to their proteins of origin.

In this study, the human protein cerebral dopamine neurotrophic factor (CDNF, 161 amino acids, R&D Systems, Cat# 5097-CD-050) was used as a model protein to demonstrate protein identification from sequencing data based on our kinetic model and proteome mapping software. We sequenced recombinant CDNF using a set of five NAA recognizers [PS610 recognizes N-terminal phenylalanine (F), tyrosine (Y), and tryptophan (W); PS961 recognizes leucine (L), isoleucine (I), and valine (V); PS691 recognizes arginine (R); PS1259 recognizes N-terminal glutamine (Q) and asparagine (N); and PS1165 recognizes N-terminal alanine (A) and serine (S)]. This set of five recognizers recognizes a total of 11 NAAs. We digested CDNF using the endopeptidase Lys-C and prepared a peptide library for sequencing. Data was collected and analyzed using our cloud-based software.

The same full-length CDNF protein was sent to a university mass spectrometry core facility for LysC digestion and LC-MS/MS analysis on the Q Exactive™ HF-X Hybrid Quadrupole-Orbitrap™ Mass Spectrometer equipped with dual pump Ultimate 3000 nanoLC (Thermo Fischer, San Jose, CA). The UniProt human protein database was used as a reference for peptide identification.

Results and Discussion

The results from both mass spectrometry and Quantum-Si's Platinum protein sequencing platform correctly identified the CDNF protein, despite different workflows and identification of different peptides (**Figure 1**).

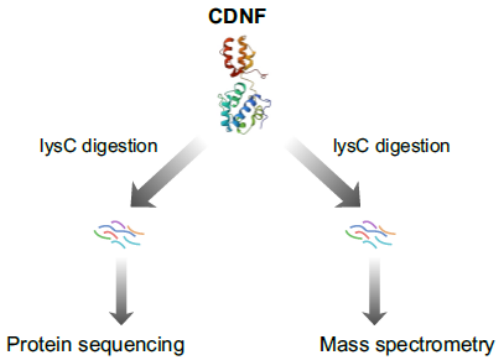

Figure 1: CDNF Protein Sequencing and Mass Spectrometry Workflow

The workflow for submitting samples to the mass spectrometry core was carried out by multiple people and required shipment of samples and analysis. Sample submission involved scheduling time with the core facility, preparing the samples, and shipping the samples to the core. Once received, the samples were processed, digested, and run on mass spectrometry equipment by highly trained staff. The data was then processed using a suite of software tools by the core facility and sent to Quantum-Si in the form of a spreadsheet summarizing the mapping of spectra to peptides. The total time from protein preparation for shipment to receiving results was around 5 days.

The workflow on Platinum—including sample preparation, sequencing, and data analysis—was completed by one scientist. Proteins were digested into peptides and prepared for sequencing in less than 3 days and required less than 3 hours of hands-on time. Protein sequencing on the Platinum was initiated immediately following sample preparation and data collected during the 10-hour sequencing run was automatically and securely transferred to the Cloud for analysis. Peptides were automatically identified by the analysis software based on the fluorescence and kinetic properties of dye-labeled NAA recognizers as described previously.[2] The total time from protein digestion to results interpretation was less than 4 days.

While different peptides were identified via mass spectrometry and Platinum, the CDNF protein was correctly identified in both approaches. Mass spectrometry analysis of CDNF resulted in the identification of 25 total peptides mapping to CDNF. This set of peptides included a number of peptides derived from incomplete LysC cleavage (**Figure 2a**).

In mass spec, the software commonly uses a search engine to sieve through acquired mass spectra to identify individual peptides by comparing the observed mass-to-charge ratios (m/z) with those expected for known peptides in pre-selected databases. The software also performs statistical analysis to estimate the probability that each peptide identified is correct. Subsequently, mass spectrometry software tools assign identified peptides to specific proteins that contain those peptides—here, an inference algorithm is used to group peptides together into proteins and assign a probability score for each protein identified.

Platinum analysis of CDNF resulted in unique identification of 6 peptides using kinetic binding signatures to correctly recognize amino acids and map them to the CDNF protein (**Figure 2b**). These peptides are sufficient to identify CDNF as the protein of origin from the human proteome with high confidence (Read the application note).

Both Platinum and mass spectrometry resulted in identification of CDNF using distinct methods to confidently call unique peptides.

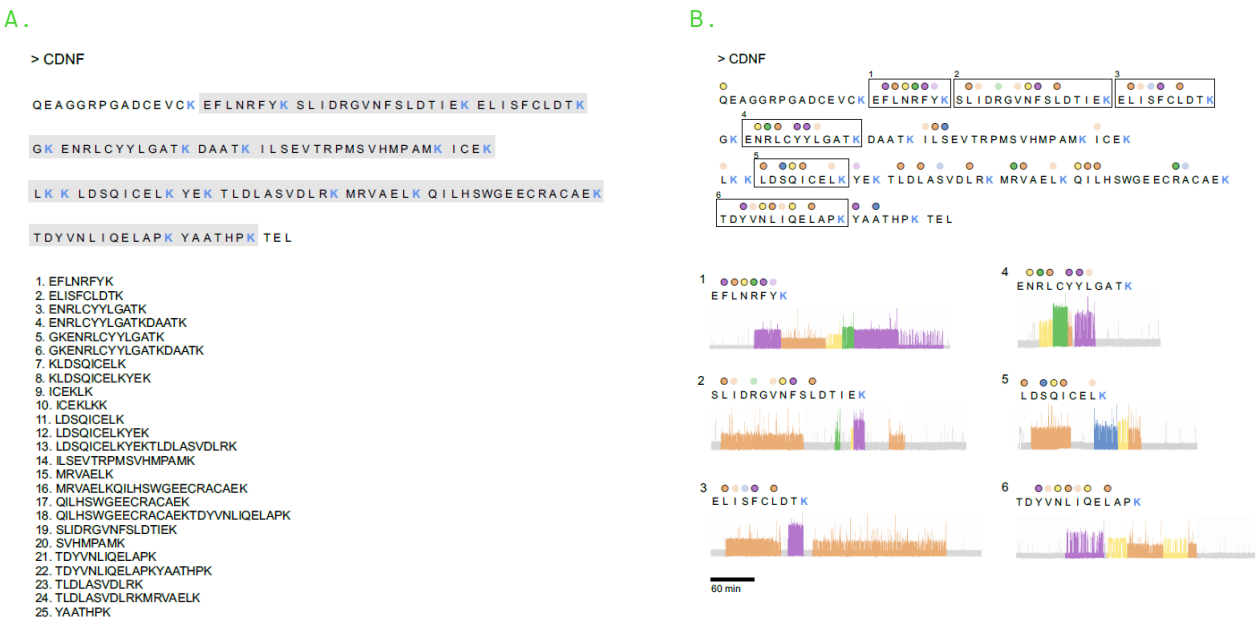

Figure 2: Comparison of Mass Spectrometry and Platinum Protein Analysis of CDNF

Conclusion

Quantum-Si's Platinum workflow offers a convenient solution for single-molecule protein identification without the need for expensive capital equipment and advanced expertise. Results are automatically provided in a Cloud-based software environment, enabling easy and straightforward interpretation of protein identification results. Single-molecule protein sequencing on Platinum offers a solution for deeper interrogation of proteoforms and PTMs through the use of kinetic signatures and enables researchers to make new discoveries about the proteome.