Accurate Measurement of Peptide Variant Mixtures with Next-Generation Protein Sequencing™ (NGPS™)

Poster P01.29
Presenter: Deepak Mistry
Mathivanan Chinnaraj, Jianan Lin, Kristin Blacklock, Eric Hermes, Michael Meyer, Douglas Pike, John Vieceli, Ilya Chorny

Quantum-Si Incorporated, 29 Business Park Drive
Branford, CT 06405, USA

ABSTRACT

ProteoVue™ is a comprehensive bioinformatics pipeline for single-amino acid variant (SAAV) detection and quantification using the Quantum-Si Platinum® and Platinum® Pro NGPS instruments. ProteoVue integrates pulse calling, recognition segment detection, fluorescence dye classification, and a neural network-driven kinetic signature database to enable variant calling with high sensitivity and specificity. These components feed into a scoring-based alignment and clustering framework, allowing accurate differentiation of binary peptide mixtures and proteoforms with subtle sequence variations. We demonstrate that ProteoVue successfully detects SAAVs across a range of substitution types, including variants lacking direct recognizer interactions, by leveraging kinetic features and sequence context. While some substitutions present challenges, the pipeline consistently captures key kinetic signatures for variant discrimination, underscoring its potential as a versatile tool for proteomic research. As NGPS technology matures and recognizer libraries expand, ProteoVue will enable more refined variant analysis, supporting applications in basic research, biomarker discovery, and clinical proteomics.

METHODS

Peptide Design and Synthesis

- Seven peptide variants (F, W, R, M, N, A, C) were designed based on the RFNELXFDISRYLANK(N3) sequence, with substitutions at the sixth position
- Peptides were synthesized with C-terminal azido-lysine modifications and prepared for sequencing with Quantum-Si's Library Preparation Kit-LysC kit
- Preceding and succeeding positions impact kinetic variations, crucial for accurate variant calling

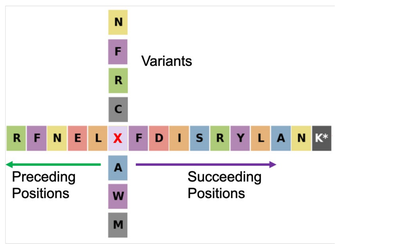


Fig. 1. Design of peptide sequence with all 7 variants at sixth position

Next-Generation Protein Sequencing (NGPS) Workflow

- Peptides were loaded onto the Quantum-Si Platinum instrument and analyzed using the V3 Sequencing Kit. Kit includes six dye-labeled N-terminal amino acid (NAA) recognizers targeting 13 of the 20 canonical amino acids, plus aminopeptidases that cleave each NAA sequentially.
- Recognizer binding and dissociation events were captured as fluorescent pulses, allowing for single-molecule sequencing and kinetic profiling.
- The sequencing run lasted ~10 hours, capturing real-time single-molecule kinetic data.

RESULTS

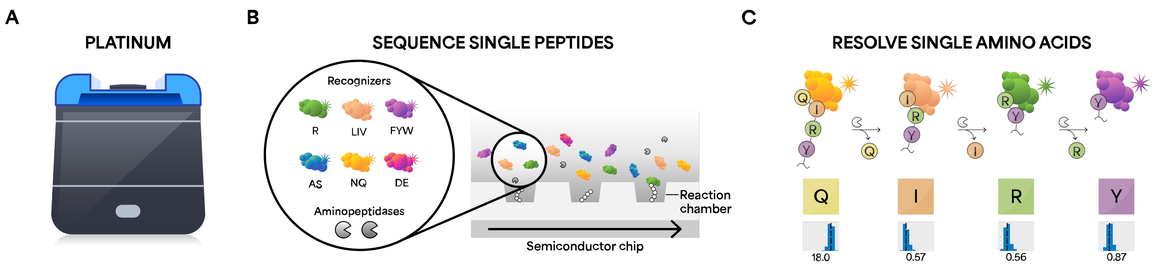


Fig. 1(A): The Platinum instrument sequences single peptide molecules with single amino acid resolution. (B) Sequencing kits include semiconductor chips, aminopeptidases, and six dye-labeled NAA recognizers that reversibly bind 13 target NAAs. (C) Binding of dye-labeled NAA recognizers generates kinetic information indicating which amino acid is being detected.

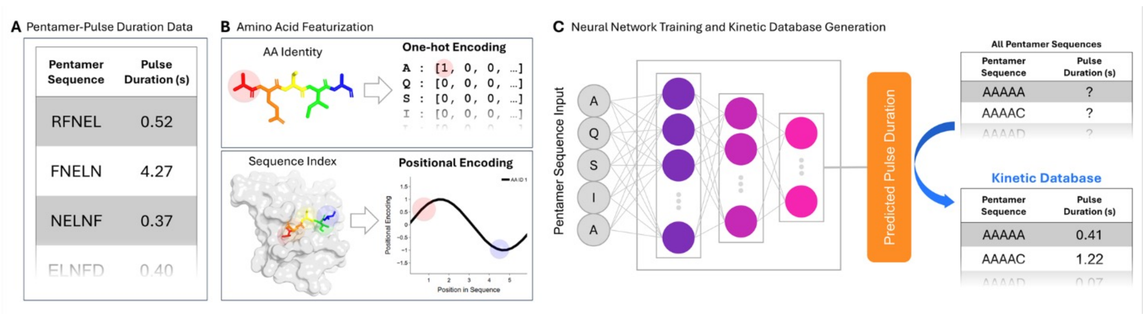
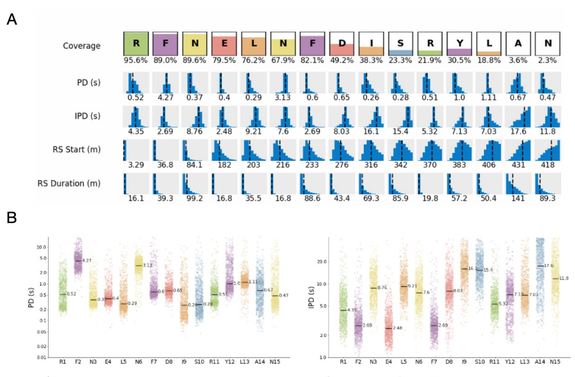


Fig. 7. Architectural overview of the neural network used to generate the kinetic database.

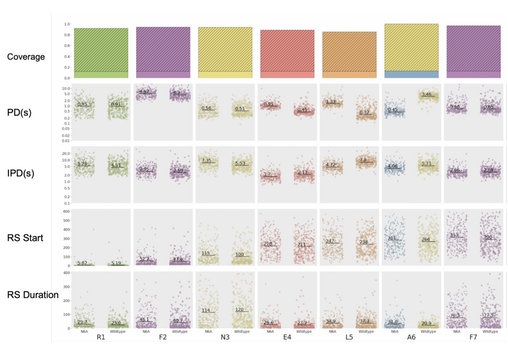
- **ProteoVue Workflow** uses peptide sequences and reference profiles to estimate SAAV ratios by integrating kinetic data and one-hot encoding.
- **Neural Network and Kinetic Database** predicts pulse durations for pentamer sequences, creating a kinetic database for accurate SAAV detection.
- **Training and Validation:** A balanced dataset of 30 proteins and 32 synthetic peptides ensures robust training and reliable performance.

Kinetic Properties of the Pure N6 Peptide



- The study demonstrates **SAAV detection** through the sequencing of seven variant peptides and their kinetic characterization using **ProteoVue**.
- Key metrics include mean **Pulse Duration** (PD), mean **Inter-Pulse Duration** (IPD), **Reference Start** (RS) times, and **RS Duration** across 177,807 reads for the N6 variant. Coverage data shows decrease in alignment efficiency towards the peptide's end.
- Evaluated across four titration series (N6A, F6W, R6M, C6M), covering a **100-fold dynamic range**. Mean Absolute Error (MAE): 0.93–1.39 (log scale), reflecting variant quantification accuracy.

Kinetics Profile and SAAV Detection



- **Variant Differentiation:** ProteoVue effectively distinguishes between kinetic profiles of N6A peptide variants at a 1:1 ratio, showing clear separation in pulse duration values at variant and upstream positions.
- **Population Quantification:** The workflow accurately estimates relative SAAV abundances to within a factor of 10, enabling precise population differentiation. Limit of detection: (SAAVs detected down to 2 pM on chip, confirming high sensitivity).
- **Demonstrates the ability of NGPS** to perform robust SAAV detection and characterization, with consistent trends across diverse variant types.

CONCLUSION

The study demonstrates the capability of Next-Generation Protein Sequencing (NGPS) with ProteoVue in detecting Single Amino Acid Variants (SAAVs) with high sensitivity and accuracy. By leveraging N-terminal amino acid recognizers and real-time kinetic profiling, **NGPS enables quantitative and qualitative detection of SAAVs, including substitutions with no direct recognizer interaction.**

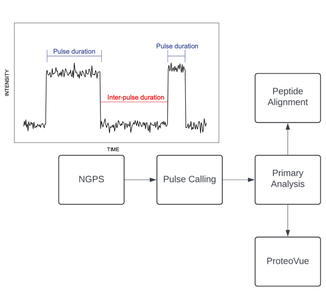
- The ProteoVue workflow integrates pulse-calling, recognition segment detection, and a neural network-driven kinetic database, allowing for precise variant identification in binary peptide mixtures.
- ProteoVue successfully recovers expected variant ratios, demonstrating its reliability in protein variant screening.
- The ability to distinguish isobaric amino acids and detect subtle proteoform changes further highlights NGPS as a valuable tool in proteomic analysis.
- Expanding the recognizer library, refining bioinformatics pipelines, and improving kinetic modeling will further enhance SAAV detection.

As NGPS technology advances, it is expected to become a powerful tool for precision proteomics, complementing mass spectrometry and other analytical techniques in detecting protein modifications and disease-associated variants.

REFERENCES

1. Reed, B.D., et al., Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device. *Science*, 2022. 378(6616): p. 186-192.
2. Mathivanan Chinnaraj, et al. Detecting Amino Acid Variants Using Next-Generation Protein Sequencing (NGPS). bioRxiv 2024.12.17.629036; doi: https://doi.org/10.1101/2024.12.17.629036

Key Terminology for ProteoVue™: Bioinformatics Tool for Detecting and Quantifying Single-amino Acid Variants



Kinetic Signatures: Unique amino acid binding patterns enable precise detection of protein variants, revealing subtle proteoform changes missed by traditional methods

Pulse Calling: Real-time fluorescence signal capture achieves single-molecule accuracy and high-throughput resolution for proteomic analysis.

Recognition Segment Detection: High-precision detection of amino acid sequences improves accuracy in identifying substitutions and proteoform isoforms.

Proteoforms: Variations in protein structures, including PTMs and isoforms, provide critical insights into disease mechanisms and drug discovery.

Variant Ratios: Quantitative measures of variant occurrences enhance reliability in population-scale proteomics and biomarker identification.

Neural Network-driven Kinetic Database: AI-powered prediction of kinetic signatures supports scalable, high-throughput workflows for complex proteomics analysis.