QUANTUM SI

Single-molecule protein sequencing for the characterization of proteins, purification impurities, and binding kinetics

David Sloan,[1] Sebastian Hutchinson,[1] Marco Ribezzi-Crivellari,[1] Anthony Oppedisano,[2] Nigel O'Neil,[2] Ajay Vashisht,[1] John Vieceli[1]

1. Quantum-Si Incorporated, 29 Business Park Drive Branford, CT 06405, USA
2. Arrowsmith Genetics, Vancouver, BC, Canada

Abstract

Introduction/methods

Accurate and efficient protein identification remains a central challenge in modern biochemistry and molecular biology. Traditional methods often depend on antibody-based detection, which can be limited by antibody availability and specificity, or on complex and costly instrumentation that restricts accessibility for many laboratories. Consequently, there is a growing need for innovative approaches that enable direct and unbiased protein analysis without these constraints.

Recent advances in single-molecule protein sequencing have opened new avenues for proteomic analysis, allowing for precise identification of proteins and characterization of protein assemblies. In this study, we performed next-gen protein sequencing (NGPS) on the Platinum®/Platinum® Pro benchtop instruments. These single-molecule protein sequencers utilize fluorescently labeled recognizer proteins that bind to and sequentially identify amino acids, generating sequences for individual peptides attached to a semiconductor chip.

We applied NGPS to several proteomic analyses. We first used NGPS to sequence the variable regions of a recombinant therapeutic antibody (trastuzumab) as well as PD-L1. We also sequenced AAV8 and AAV9 with the goal of distinguishing between these two highly similar AAVs in mixed populations. Next, we used NGPS to identify contaminating proteins within a purified protein product. Lastly, we used the single-molecule detection capabilities of Platinum Pro to characterize and identify anti-GFP nanobodies across a broad range of binding affinities.

Results/conclusions

NGPS successfully sequenced a large percentage of PD-L1 and the variable domains of trastuzumab on both the heavy and light chains of the antibody. We also successfully differentiated between AAV8 and AAV9 and were able to detect as low as a 10% contamination of AAV8 within an AAV9 preparation. Lastly, we simultaneously characterized the binding kinetics for 20 different anti-GFP nanobodies across a 1,000-fold difference in affinity. These findings establish NGPS on Platinum Pro as a versatile tool for comprehensive and accessible proteomic analysis in biologics development.

Amino acid-level protein characterization

PD-L1 sequencing

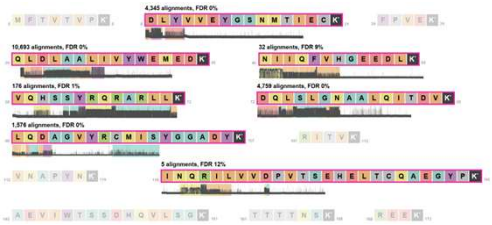
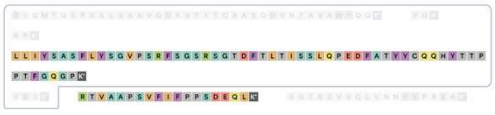


Figure 1. Sequence coverage of *programmed death ligand 1*. Sequence coverage for 7 of 14 peptides.

Trastuzumab sequencing

Light chain



Heavy chain

Figure 2. Sequence coverage of trastuzumab. Sequence coverage for four of ten peptides in the antibody variable regions.

References

1. NGPS enables accurate AAV serotype profiling and mixture resolution for gene therapy. Quantum-Si application note. November 3, 2025.
2. Brian D. Reed et al, *Science* 2022, 378 (6166) 186–192.
3. Hutchinson et al, *Biorxiv*, 2025, 10.1101/2025.06.05.658126.

Next-generation protein sequencing (NGPS)

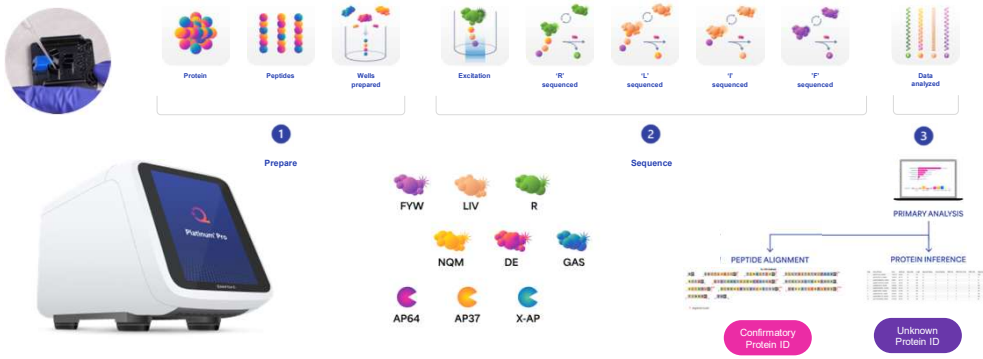


Figure 3. NGPS sample chip and schematic workflow for bottom-up proteomic analysis and sequencing via fluorescent recognizer proteins and a cocktail of aminopeptidases. Image of Platinum Pro sequencer and representation of fluorescent recognizers and amino acids detected. Schematic of two different workflows, peptide alignment for known samples and protein inference for identification of unknown proteins from a reference proteome.

Definitive discrimination between AAV8 and AAV9

Four amino acids on three peptides clearly distinguish AAV8 and AAV9.

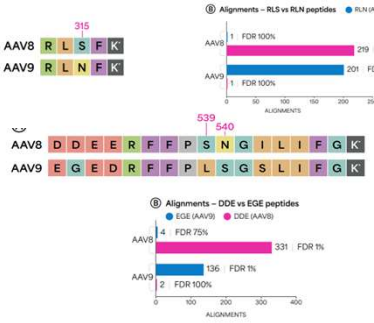


Figure 4. Two peptides with three AA substitutions which differentiate between AAV8 and AAV9. Bar graphs show the number of alignments for all sequences against AAV8 or AAV9 reference sequences and false discovery rates (FDR).

In mixtures of AAV8/AAV9, NGPS was able to clearly detect a 10% contamination of either AAV8 or AAV9 within a background of the other serotype.

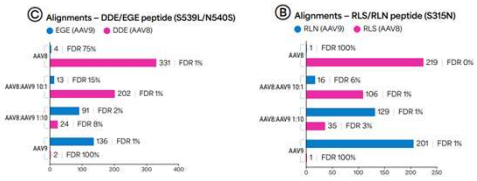
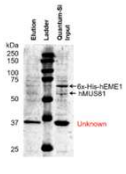


Figure 5. Bar graphs show the number of alignments for all sequences against AAV8 and AAV9 reference sequences and FDR. Samples are either pure AAV8, pure AAV9, or 90/10% mixtures of both AAV8 and AAV9.

Using NGPS as quality control of a purified protein sample

- hMUS81/hEME1-6xHis expressed in yeast and purified on nickel column.
- SDS-PAGE protein gel revealed bands from multiple unknown contaminants.
- Eluted sample analyzed via NGPS to identify contaminants.



NGPS inference results

Rank	Inferred protein	Score (rounded)	Likelihood (rounded)	Mass (kDa)	Length (aa)
1	DPH1	1	66%	48	425
2	TDH1	0.4	33%	36	332
3	ICP55	0.4	33%	58	511
4	RKM2	0.2	20%	56	479
5	DAN1	0.2	19%	30	298
6	Dubious ORF	0.2	19%	13	109
7	ACH1	0.2	16%	59	526

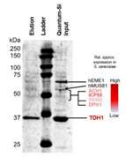
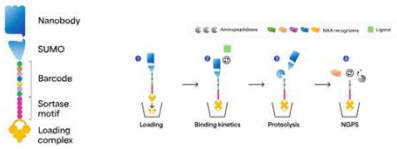


Figure 6. Top: gel elution profile of hMUS81/hEME-1 from nickel affinity column showing proteins of interest as well as contaminating bands. Bottom: inference results of eluted proteins, from peptide alignments against reference proteome.

Single-molecule binding kinetics



Parallel single-molecule binding kinetics and protein barcode sequencing workflow

1. **Loading:** barcoded nanobodies are immobilized via biotin-streptavidin interaction at the bottom of reaction chambers.
2. **Binding kinetics:** labelled ligand is flowed into the reaction chamber, and bimolecular interactions are recorded.
3. **Proteolysis:** the N-terminus of the barcode is exposed by proteolysis of SUMO by Ulp1.
4. **NGPS:** the peptide barcode is sequenced.

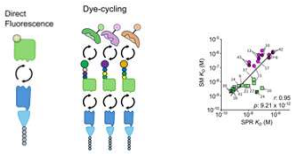


Figure 7. Left: detection of ligand binding either through direct fluorescence or dye-cycling to avoid fluorescence bleaching for high-affinity interactions. Right: comparison of K_D measurements between single-molecule binding assay and SPR.

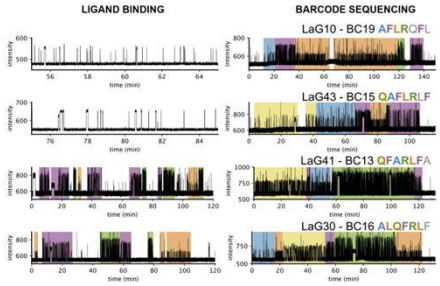


Figure 8. Left: ligand binding results for four nanobodies detected either with direct fluorescence (top) or with dye cycling (bottom). Right: sequencing results for the barcodes, identifying which nanobody's binding affinity was being measured in that microwell.

Conclusions

Next-gen protein sequencing is an accessible, benchtop workflow that can be used for multiple applications in biologics development, including:

- QC to verify protein sequences and detect low-level contaminants within final products
- Sequencing variable regions of antibodies
- Distinguishing between highly similar proteins/AAVs in mixtures
- Characterizing binding affinities/specificities for antibodies or engineered proteins in mixtures by utilizing protein barcodes for coupling of phenotype and protein identity.