QuantumSi

Sequencing and Identification of Proteins from Five-protein Mixtures on Platinum™

05 MAY 2023

Introduction

Complex biological samples consist of multiple proteins. Understanding the presence and relative abundance of different proteins in a sample can provide important information about the biological processes underlying human health and disease. Some protein detection techniques identify proteins using antibodies or other affinity-based reagents with specificity for the proteins of interest. These methods require the generation of protein-specific antibodies or the use of commercially available or custom targeted antibody panels. The need to source or develop high-quality antibodies with the correct specificity for target proteins and panels limits the proteins that can be identified and typically places key information about proteins with variant peptide sequences out of reach.

Quantum-Si's Platinum™ instrument offers a solution to identifying proteins in a mixture by directly sequencing them, eliminating the need for targeted antibody-based detection assay. To demonstrate feasibility of detecting multiple proteins in a sample on Platinum, we combined five recombinant proteins (CDNF, FGF2, IL4, GMFB, and PDIA1) in a mixture and used protein sequencing to identify them **(Table 1)**. In this application note, we describe a method for optimizing sample preparation for a mixture of five proteins. Protein identification through next-generation protein sequencing can confirm the presence or absence of specific proteins in a mixture. The ability to interrogate multi-protein sample mixtures at reduced input concentrations compared to the single protein concentration recommendations offers increased

REFERENCES

1 Lin CQ, Chen LK. Cerebral dopamine neurotrophic factor promotes the proliferation and differentiation of neural stem cells in hypoxic environments. Neural Regen Res. 2020 Nov;15(11):2057-2062.

2 Ornitz DM, Itoh N. The Fibroblast Growth Factor signaling pathway. Wiley Interdiscip Rev Dev Biol. 2015 May-Jun;4(3):215-66.

3 Gadani SP, Cronk JC, Norris GT, Kipnis J. IL-4 in the brain: a cytokine to remember. J Immunol. 2012 Nov 1;189(9):4213-9.

4 Fan J, Fong T, Chen X, Chen C, Luo P, Xie H. Glia maturation factor-β: a potential therapeutic target in neurodegeneration and neuroinflammation. Neuropsychiatr Dis Treat. 2018 Feb 9;14:495-504.

5 Rahman NSA, Zahari S, Syafruddin SE, Firdaus-Raih M, Low TY, Mohtar MA. Functions and mechanisms of protein disulfide isomerase family in cancer emergence. Cell Biosci. 2022 Aug 14;12(1):129.

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

process efficiency and testing throughput. These results demonstrate the applicability of the Quantum-Si Platinum sequencing platform to interrogate complex protein samples via real-time dynamic sequencing.

TABLE 1

Protein	Function
CDNF	Cerebral dopamine neurotrophic factor (CDNF) plays a role in neuronal function[1]
FGF2	Fibroblast growth factor 2 (FGF2) is part of a family of cell signaling proteins involved in normal development in animal cells[2]
IL4	Interleukin-4 (IL4) is a cytokine that suppresses inflammation and plays a role in tissue repair[3]
GMFB	Glia maturation factor beta (GMFB) is a growth and differentiation factor for both glia and neurons and involved in neurodegeneration[4]
PDIA1	Protein disulfide isomerase (PDIA1) serves many functions in chaperoning and assisting with folding proteins[5]

Table 1: The 5 Proteins Sequenced and their Function

Methods

Quantum-Si's library preparation workflow uses the Library Preparation Kit (Catalog Number: 910-10011-00) to perform reduction, alkylation, and endopeptidase digestion on intact proteins. The resulting peptide fragments then undergo azide-functionalization on C-terminal lysines, followed by attachment to molecular linkers using click chemistry conjugation (as described in Reed *et al.*, 2022).[6] To demonstrate increased sample complexity by combining multiple individual proteins prior to library preparation, we prepared libraries using two different methods **(Figure 1)**.

In the first method, individual libraries were prepared separately from each of the five commercially available recombinant proteins CDNF, FGF2, IL4, GMFB, and PDIA1 following our standard library preparation protocol. Briefly, each protein was exchanged into sample buffer and diluted to a 5 µM final concentration. The proteins were then reduced, alkylated, and digested with the endopeptidase LysC. The resulting peptide fragments were subsequently functionalized and conjugated to macromolecular linkers as described above using the reagents provided in the Library Preparation kit. After library preparation, the individual libraries were combined into a single solution at equimolar peptide concentrations (100 nM final peptide concentration for each library). We term this sample the 'post-mix' sample.

FIG. 1

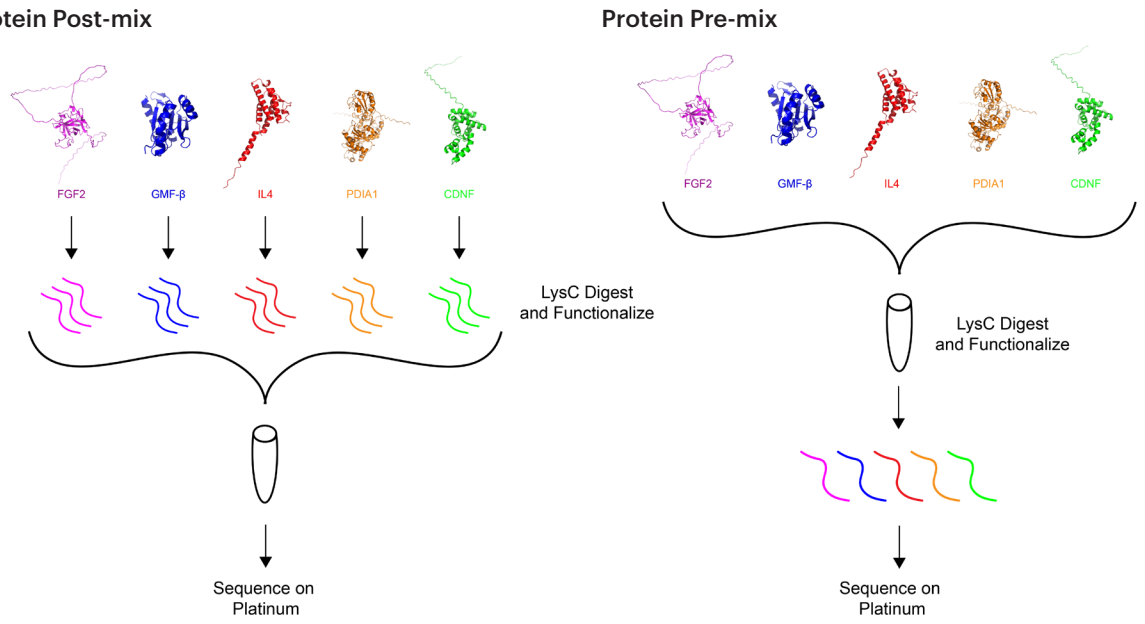

Figure 1: Schematic representation of library preparation workflows. In the Post-mix Workflow, 5 µM of individual CDNF, FGF2, GMFb, IL4, and PDIA1 proteins were digested with LysC and functionalized with linkers generating 5 peptide libraries which were then mixed together in equimolar ratios prior to sequencing. In the Pre-mix Workflow, 1 µM of each CDNF, FGF2, GMFb, IL4, and PDIA1 proteins were mixed together resulting in a 5 µM protein mixture. The proteins were then digested with LysC and functionalized with linkers generating a single library ready for sequencing.

In the second method, the five full-length proteins were combined at equimolar concentrations prior to library preparation. Here, a 5 µM solution containing 1 µM each of CDNF, FGF2, IL4, GMFB, and PDIA1 was first prepared. This solution was processed using the standard library preparation workflow described above. We term this sample the 'pre-mix' sample.

After library preparation, the post-mix and pre-mix protein libraries were sequenced on Quantum-Si's Platinum instrument with the commercially available Quantum-Si protein sequencing kits (PN 860-00002-00) using the previously described workflow for real-time dynamic protein sequencing.[6] Briefly, a 7 nM solution of the respective protein library was injected into a flow cell and the conjugated peptides with exposed N-termini were immobilized onto the semiconductor chip surface. After washing of each flow cell to remove unbound peptides, fluorescently labeled N-terminal amino acid (NAA) recognizers and aminopeptidases were added to the flow cell. The NAA recognizers bind to the NAAs of the immobilized peptides and the fluorescence signal from these binding events is recorded by the chip.

The aminopeptidases sequentially cleave amino acids from the N-terminus, exposing the next NAA for recognition, and the process repeats until the whole peptide is sequenced. Data is automatically and securely uploaded to the Cloud software environment for peptide identification and mapping to the correct protein.

Results and Discussion

To compare the performance of the library prep kits with a mixture of five proteins on Platinum, we first examined the peptide coverage for each protein in the pre-mix and post-mix samples. Given the five-fold lower sample input for each protein in the pre-mix sample, we sought to determine if the overall peptide coverage for each protein was reduced compared to the post-mix sample. As shown in **Figure 2**, analysis of the sequencing data resulted in the identification of an identical set of peptides aligning to each of the proteins present in both libraries.

FIG. 2

Figure 2: Comparison of peptide sequence coverage from the pre-mix and post-mix samples. Protein sequences were digested with LysC in silico and run through Quantum-Si's kinetic prediction software. Amino acid positions with high predicted average pulse duration (> 0.3 s) are indicated by circles with bold shading. Positions with low predicted average pulse duration (0.1-0.3 s) are indicated with light shading. Black boxes indicate peptides that were identified in the sequencing output by the software analysis pipeline. The same set of 9 total peptides aligned to PDIA1, 6 peptides aligned to CDNF, 5 peptides aligned to FGF2, and 3 peptides each aligned to GMFB and IL4 in both the pre-mix and post-mix samples.

The same set of 9 total peptides aligned to PDIA1, 6 peptides aligned to CDNF, 5 peptides aligned to FGF2, and 3 peptides each aligned to each of GMFB and IL4 in both the pre-mix and post-mix samples. This result demonstrates that multiple proteins can be identified from complex mixtures while producing equivalent peptide coverage from a reduced sample input. Example traces are displayed for each protein in the post-mix and *pre-mix* samples in **Figure 3**.

FIG. 3

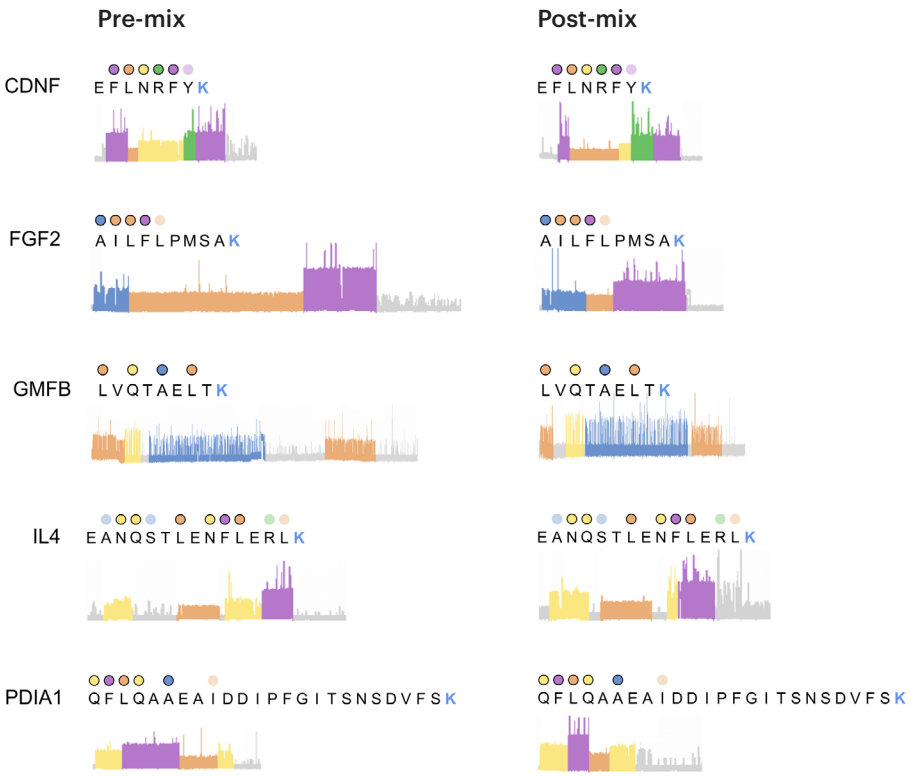

Figure 3: Comparison of sequencing output of peptides from each library workflow. An example trace from a single aligned peptide is shown for each protein. Peptides exhibited highly consistent kinetic behavior during sequencing in the pre-mix and post-mix workflows.

Conclusion

In this application note, we demonstrated that the library preparation kit was effective in the preparation of multi-protein mixtures for sequencing on Quantum-Si's Platinum instrument. We demonstrated that the multiple protein workflow results in equivalent peptide alignment and sequence coverage compared to preparation of a single library for each protein of interest. This new workflow will enable an increase in efficiency and throughput as multiple proteins can be positively identified within a single library preparation and sequencing run.