29 January 2026

Optimization of Quantum-Si Platinum single-molecule protein sequencing platform towards improved complex-matrix protein identification

Tomasz A Leski[1], Sean M Brown[2], Zachary T Johnson[1], Scott N Dean[1], Ellen R Goldman[1], David A Stenger[1]

1. Center for Biomolecular Science and Engineering, US Naval Research Laboratory

2. National Research Council Research Associate at U.S. Naval Research Laboratory

Abstract

Proteins are a class of macromolecules with essential roles in processes and structures associated with life. Protein sequencing technologies are therefore fundamental for understanding cell metabolic pathways, disease mechanisms, and how pathogenic agents and toxins function. Emerging next generation protein sequencing (NGPS) technologies promise a dramatic improvement of proteomics methods enabling identification of pathogens and toxins with unparalleled sensitivity and precision. The Quantum-Si (QSi) Platinum Sequencer is a novel single molecule protein sequencing technology capable of single amino acid resolution. In this work, we conducted significant optimization of the QSi protein library preparation protocol, reducing sample preparation time from 32 to 10 hours without sacrificing substantial sequencing quality, allowing for a sample-to-answer timeline in less than 24 hours. The modified protocol was applied for analyzing a set of proteins including sixteen single-domain antibodies with diverse sequences and a nontoxic derivative of staphylococcal enterotoxin B. We were further able to determine the library dilution threshold: losing the ability to sequence beyond 100x dilution. Finally, we were able to successfully obtain protein sequences within a crude lysate background, demonstrating the effectiveness of sequencing within complex protein mixtures. Improvements in sequencing chemistry and data processing may soon lessen or eliminate the dependence on reference sequences: a current obstacle for efficiently characterizing unknown proteins. By further condensing and optimizing library preparation, this technique presents a potential application for proteomics that require rapid characterization of highly complex biological systems, significantly improving protein-based diagnostic technologies.

Optimization of Quantum-Si Platinum single-molecule protein sequencing platform towards improved complex-matrix protein identification

Tomasz A Leski [1]*, Sean M. Brown[2], Zachary T. Johnson[1], Scott N. Dean[1], Ellen R. Goldman[1], David A. Stenger[1]

[1]Center for Biomolecular Science and Engineering, US Naval Research Laboratory, Washington, DC.

[2]National Research Council Research Associate at U.S. Naval Research Laboratory, Washington, DC.

*Corresponding author
Tomasz A. Leski
email: tomasz.a.leski.civ@us.navy.mil

Abstract

Proteins are a class of macromolecules with essential roles in processes and structures associated with life. Protein sequencing technologies are therefore fundamental for understanding cell metabolic pathways, disease mechanisms, and how pathogenic agents and toxins function. Emerging next generation protein sequencing (NGPS) technologies promise a dramatic improvement of proteomics methods enabling identification of pathogens and toxins with unparalleled sensitivity and precision. The Quantum-Si (QSi) Platinum Sequencer is a novel single molecule protein sequencing technology capable of single amino acid resolution. In this work, we conducted significant optimization of the QSi protein library preparation protocol, reducing sample preparation time from 32 to 10 hours without sacrificing substantial sequencing quality, allowing for a sample-to-answer timeline in less than 24 hours. The modified protocol was applied for analyzing a set of proteins including sixteen single-domain antibodies with diverse sequences and a nontoxic derivative of staphylococcal enterotoxin B. We were further able to determine the library dilution threshold: losing the ability to sequence beyond 100x dilution. Finally, we were able to successfully obtain protein sequences within a crude lysate background, demonstrating the effectiveness of sequencing within complex protein mixtures. Improvements in sequencing chemistry and data processing may soon lessen or eliminate the dependence on reference sequences: a current obstacle for efficiently characterizing unknown proteins. By further condensing and optimizing library preparation, this technique presents a potential application for proteomics that require rapid characterization of highly complex biological systems, significantly improving protein-based diagnostic technologies.

Introduction

Protein and peptide sequencing techniques are fundamental tools in molecular biology and biotechnology, providing insights into the primary structure of proteins, which allows us to explore their functional roles in living systems.[1] This information may be critical for understanding protein function, disease mechanisms, and provide indispensable data for development of a broad range of drugs and therapies [2]. Over the decades, sequencing approaches have evolved from foundational methods, such as Edman degradation,[3] to advanced techniques utilizing mass spectrometry [4] and fluorosequencing,[5] offering remarkable speed and resolution. Today, innovations such as single-molecule sequencing,[6] top-down proteomics,[7] and machine learning[8, 9] are redefining the field by expanding its applications in research, personalized medicine, and diagnostics.

Single-molecule protein sequencing is emerging as a transformative tool for identifying proteins and peptides with unparalleled sensitivity and precision. Unlike traditional methods that require bulk, homogeneous protein samples, single-molecule techniques analyze individual protein molecules. Example applications of this technology include detection of rare or low-abundance peptides, often in complex protein mixtures, which may prove useful for early diagnosis of some diseases or monitoring of industrial biomanufacturing processes. This approach is particularly valuable in pathogen and toxin identification, where rapid and accurate detection can for example, prevent the spread of infectious disease. Proteomic technologies, such as nanopore-based sequencing and fluorescence-based methods, can resolve amino acid sequences and post-translational modifications directly from small sample volumes. These advancements hold immense potential for real-time diagnostics, offering a faster, more specific alternative to culture-, immunoassay-, or genomic-based approaches while uncovering new biomarkers and virulence factors in complex biological samples.

Quantum Si (QSi) is a novel technology, currently available in benchtop format, that achieves "*single-molecule protein sequencing using a dynamic approach in which single peptides are probed in real time by a mixture of dye-labeled N-terminal amino acid recognizers and simultaneously cleaved by*

aminopeptidases." [10] To do so, proteins are first cleaved into peptides with the endoproteinase LysC, cleaving peptide bonds at the carboxyl sides of lysines. These peptides are then immobilized onto wells within a semiconductor chip before iterative cycles of N-terminal florescence recognition and cleavage by aminopeptidases. More specifically, fluorescence intensity and time from amino acid recognizers (which bind to one or multiple types of amino acids) is measured after excitation with a 532-nm laser. From this data, kinetic signatures, or measurable properties of recognizer events in real-time, provide data for which to align peptides to reference sequences of a specified target(s).

Altogether, the system offers some notable advantages and drawbacks when compared to comparable protein sequencing technologies (for a comprehensive review on contemporary protein sequencing techniques see Alfaro et al. [1]). Edman degradation [11], the first developed method to sequence a pure peptide, while transformative in the 1950s when it was first introduced, is arduous and requires chemically labeled peptides for multiplexed sequencing peptides of high purity, known as flourosequencing [12]. Mass spectrometry (MS) is another method currently used for proteome characterization and capable of sensitive identification of proteins in complex mixtures. De-novo sequencing by MS however relies on existing signature databases where sequencing unknown proteins remains an unsolved challenge [13].

In contrast, rapid advancements in applying the same fundamental technology which uses biological and solid-state nanopores for nucleic acid sequencing shows promise as a peptide sequencing methodology [14-16]. At current however, research has not yet matured enough, thus limiting these approaches to nucleotides. Interestingly, an approach similar to nanopore sequencing, known as quantum field-effect nanogap sensing[17], has begun to gain momentum as it uses the quantum transport method coupled with machine learning to allow for high precision nucleotide and potential protein sequencing [18]. In brief, as molecules pass by a nano-sized junction, the electron tunneling current of each single molecule is measured and then elucidated through machine learning approaches. This technique has already been validated for DNA sequencing and has been speculated to work with protein sequencing [17]. It however

remains unknown to what extent it will address the protein sequencing problem. Regardless, work here is still maturing as no commercialized device exists yet for either nanopore or nano gap peptide sequencing.

QSi therefore currently presents the only benchtop, commercialized, single-molecule sequencing approach that eliminates arduous procedures typically associated with traditional sequencing methods. The QSi technology uses semiconductor chips with nanoscale reaction chambers that circumvent conventional stepwise chemistries and complex fluidics, which is combined with a laser-based optical readout providing nanosecond precision and ordered recognition based upon on-off binding kinetics. The actual sequencing process relies on attachment of individual peptides in the microscopic reaction chambers at their C-terminus and detection of N-terminal amino acids using fluorescently labelled recognizers. The N-terminal amino acid is subsequently cleaved, and the next one is exposed, recognized and detected. This process is repeated until the whole peptide is fully digested. For the current generation of the technology, not all genetically encoded amino acids have a corresponding recognizer, leaving them invisible to sequencing. Additionally, some recognizers bind to multiple amino acids which leaves ambiguity in calling one amino acid over another. The output of the sequencing process is therefore a sequence of recognizers. This sequence of recognizers is then aligned to either a target sequence in the case of the QSi "alignment" script or a proteome in the case of the QSi "*protein inference*" script based on sequencable peptides.[6]

In this work, single-molecule sequencing was evaluated using a QSi Platinum Protein Sequencer with version 3 (V3) sequencing chemistry. We explored QSi technology and protocols with the following goals: (1) modify the existing QSi protocol to demonstrate library preparation and sequencing data within 24 hours without substantial loss of performance, (2) determine the performance of sequencing for proteins with diverse sequences and sizes: from single-domain antibodies (sdAbs – Table 1) to toxoids (SEBv) and (3) determine the performance of sequencing of target proteins in mixtures of varying complexity.

Table 1. Proteins used in the study.

Name	*Type*	*Length (aa)*	*MW (kD)*	*Reference*
V2B3	sdAb	135	14.3	[27]
V2C3	sdAb	126	13.7	[27]
V3A8f	sdAb	138	15.0	[27]
WD11f	sdAb	136	14.6	[28]
WE11f	sdAb	139	15.1	[28]
WB9	sdAb	127	13.9	[28]
WF4	sdAb	126	13.7	[28]
WC10	sdAb	127	13.9	[28]
WH11	sdAb	129	14.2	[28]
WE10	sdAb	129	14.2	[28]
a16	sdAb	136	15.0	[29]
a18	sdAb	141	15.3	[29]
a19	sdAb	133	14.5	[29]
a86	sdAb	134	14.6	[29]
a155	sdAb	134	14.5	[29]
ACVE	sdAb	153	15.9	[30]
SEBv	toxoid	248	29.2	[24, 31]

Results and Discussion

1. Determination of run performance criteria

We used the QSi Platinum protein sequencing system in offline configuration for this study (see methods), which does not explicitly provide assessment of the quality of each sequencing run. Several values, however, including chip loading (CL) and number of high-quality reads (HQR) are available after performing primary analysis of sequencing data. CL expressed as a percentage reflects the proportion of sequencing chip wells (apertures) occupied by peptides after conducting a library loading step. The CL value is calculated based on the count of wells occupied by one or more peptides. Although only wells containing a single peptide provide reliable sequencing results [6]. For the 79 runs analyzed here, the observed CL values ranged between 0.2% to 80.4%, and HQR ranged from 24 to 125,151. Chip loading, and to a certain degree HQR, ranges appear dependent on the protein analyzed (Figure SI.1G-H). Another metric considered for run quality assessment was the number of alignments for sequencing reads of an analyzed protein to its own sequence. For the purposes of this work, we considered two alignment values:

total alignments (TA) and high-quality alignments (HQA), which only included alignments with false detection rates (FDR) less than 0.05 (see methods).

To determine the impact of chip loading on run quality and the relationship between high-quality read and alignment numbers, we calculated coefficients of determination (R^2): (1) between loading (CL) versus HQR, TA, and HQA, (2) between reads (HQR) versus alignment values (TA and HQA), and (3) between total and high-quality alignments (TA and HQA) shown in Figure SI.1. We observe a weak linear correlation (R^2=0.058) between CL and HQR (Figure SI.1A). We observed even weaker linear correlation (R^2=0.008) between CL and both TA and HQA (Figure SI.1B-C) in both cases. Since both TA and HQA strongly correlate with HQR (Figure SI.1D-F) and with each other (Figure SI.1E) we treat HQR as the main metric representing sequencing output and run quality.

The lack of strong correlation between CL and HQR is especially noticeable for loading values higher than 30%. Analysis of runs with CL at or lower than 30% was performed using only SEBv sequencing runs, as the only sample type for which sufficient number of runs was performed in the range of CL values between 0 and 30% (Figure SI.1A inset). On one hand, the analysis shows that for CL values < 30%, HQRs are strongly correlated with CL (R^2=0.8) where decreasing loading is associated with a significant decrease in high-quality reads. On the other, changes of CL above approximately 5 to 10% have minimal impact on HQRs thus consequently on sequencing results. We therefore deemed runs with CL above 5% successful while also considering HQR and HQL numbers. Exceptions here are experiments with deliberate use of lower peptide library concentrations to test the sequencer's limit of detection.

2. Sample-to-Answer Protocol Optimization

Results from all sequencing rounds discussed below can be found in Table SI.1. We tested various method optimization steps (Figure 1 and Figure SI.2) in attempt to either increase quality and/or quantity of the protein sequence output or shorten the sample-to-answer protocol originally developed by QSi [19, 20] without significantly compromising the sequencing results. The following modifications

were tested: (a) use of an alternative protein digestion enzyme, (b) shortening the library preparation time by reduction of protein digestion and K-linker conjugation incubations and, (c) shortening the sequencing run times.

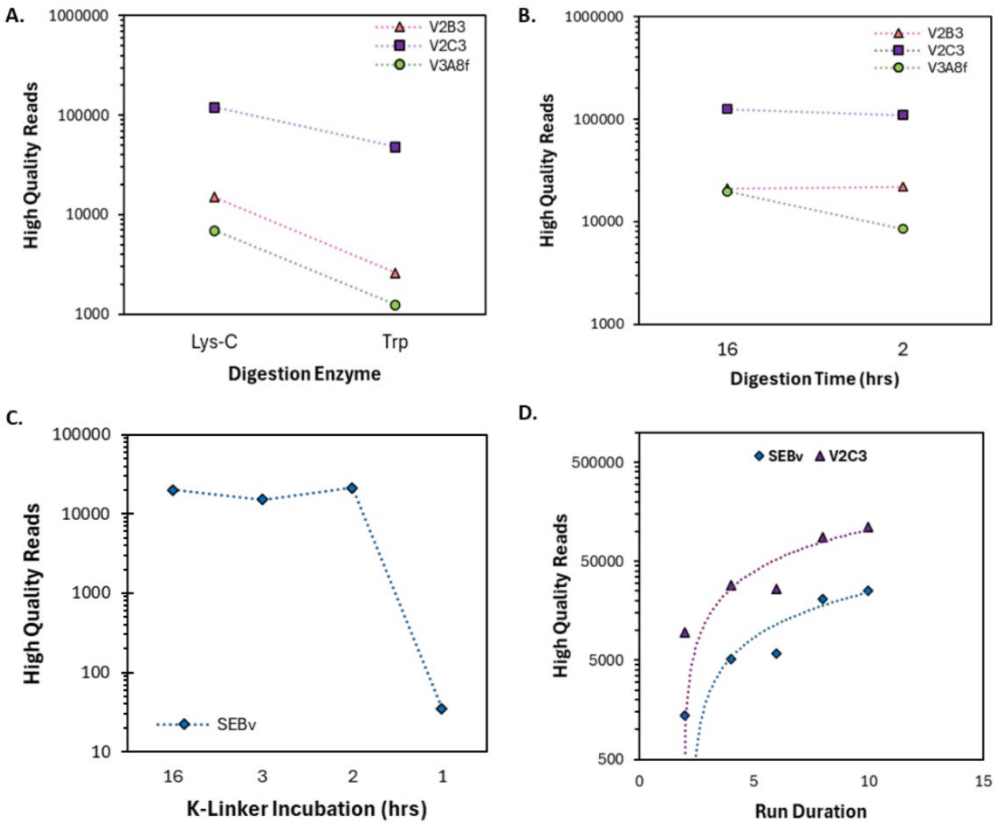

Figure 1 | Development of library preparation and sequencing protocol. A. High quality read (HQR) counts for V2B3 (pink triangles), V2C3 (purple squares), and V3A8f (lime circles) sdAbs decrease when digested with Tryp instead of Lys-C during library preparation. Library preparation conditions: 16 hr protease digestion time (DT), 16 hr K-linker incubation time (KT). Sequencing conditions: 0.4 nM loading concentration (LC), 10 hr sequencing run time (RT). **B.** HQR counts for V2B3 (pink triangles), V2C3 (purple squares), and V3A8f (lime circles) sdAbs remain relatively constant when decreasing Lys-C digestion time from sixteen to two hours (KT:16hr, LC: 0.4 nM, RT: 10hr) **C.** HQR counts for SEBv (blue diamonds) decrease for K-linker incubation shorter than 3 hr (DT: 2hr, LC: 0.2 nM, RT: 10hr). **D.** HQR counts for SEBv (blue diamonds) and V2C3 (purple triangles) decrease with lower sequencing run times (for both proteins - DT: 2hr, LC: 0.2 nM; V2C3 - KT: 16hr, SEBv - KT: 3hr). Linear correlation coefficient values (R^2) equal 0.89 and 0.88 for SEBv and V2C3 respectively.

(a) Use of Trypsin for protein digestion. To expand the possible search space of lysed peptides, we tested the use of Trypsin (Tryp) instead of Lys-C used in the standard QSi protocol. Lys-C cleaves peptide bonds at the carboxyl side of lysine residues while Tryp cleaves at the carboxyl side of both lysine and arginine residues. As a result, digestion with Tryp produces a higher number of shorter peptides. The main benefit of using Tryp instead of Lys-C is the introduction

of additional sequencing initiation points in peptides generated by cutting at arginine. While Tryp increases the number of sequencing initiation points, due to the chemistry used for peptide attachment to the chip surface by the original QSi library preparation protocol [19], the molecules needed to link peptides to the chip (K-linkers) cannot be conjugated with peptides ending with arginines. As a result, these peptides are unable to attach to the chip during the loading step and therefore are not sequenced. Another complication when using Tryp for peptide generation is that QSi sequencing analysis software is designed around Lys-C digestion, thus does not recognize arginine cutting sites when aligning reads. To bypass this limitation, arginines were replaced with lysines in the reference sequences used to generate alignments for libraries prepared using Tryp. To test the impact of Tryp substitution, we prepared libraries of three single domain antibodies (sdAbs; ~15 kDa binding domains): V2B3, V2C3, and V3A8f. For each of these sdAbs, two libraries were prepared: (i) the QSi standard 16 hr Lys-C digestion protocol and (ii) 16 hr Tryp digestion. Since Lys-C libraries had lower than expected concentration after the K-linker attachment step, they were loaded at 0.4 nM for sequencing (while the standard 0.2 nM concentrations of Tryp libraries were used). All six libraries were sequenced using the standard QSi protocol. The CL values and comparison of HQR, TA, and HQA between these two library types are shown in Figure 1 (Figure 1A), Figure SI.2A and Table 2. All the CL values were found to be above 20%. We observed CL for Lys-C libraries between 63.1-75.9% and between 23.9-38.3% for Tryp libraries. To account for difference between loading concentrations of Lys-C and Tryp libraries, we adjusted the HQR numbers using a factor of 2. This adjustment factor was calculated based on the finding that HQR number is directly correlated with library concentration (Table SI.1). For all libraries, HQR (both before and after the concentration-related adjustment) are much lower for Tryp than for Lys-C. On the other hand, the alignment numbers (both TA and HQA) are higher for V2B3 and V2C3 Tryp libraries even when not adjusted for library concentration. This may be explained by increased proportion of highly sequencable peptides in V2B3 and V2C3 Tryp compared with the Lys-C libraries. This phenomenon was not observed in

case of V3A8f; we observed negligible TA and zero HQA for this sample suggesting it may be highly dependent on the composition of the protein. All the findings discussed above warrant further investigation of library preparation methods using Tryp, offering a possible avenue of improvement by increasing the sequencing efficiency.

Table 2. Use of Trypsin for protein digestion.

Sample	***Enzyme***	***Library concentration (nM)***	***CL (%)***	***HQR***	***TA***	***HQA***	***HQR***	***TA***	***HQA***
				actual			**normalized**		
V2B3	Lys-C	0.4	67.7	15041	91	64	15041	91	64
V2B3	Trp	0.2	31.1	2583	221	198	5166	442	396
V2C3	Lys-C	0.4	63.1	120680	13017	12920	120680	13017	12920
V2C3	Trp	0.2	38.3	47936	15884	15827	95872	31768	31654
V3A8f	Lys-C	0.4	75.9	6924	16	0	6924	16	0
V3A8f	Trp	0.2	23.9	1242	18	0	2484	36	0

(b) Reduction of sample-to-answer time. To assess the viability of shortening the library preparation time, we tested variations in: (i) protease digestion time, (ii) k-linker incubation time

i. **Protease digestion time.** To test the impact of Lys-C digestion time we prepared libraries of three sdAbs: V2B3, V2C3, and V3A8f. Two libraries were prepared for each protein with either the standard QSi 16 hr Lys-C digestion protocol or 2 hr Lys-C digestion. All six libraries were sequenced using the standard QSi protocol with reported CL values between 34.8% and 74.7%. The CL values and comparison of HQR, TA and HQA numbers between these two library types are shown in Figure 1B, Figure SI.2B and Table 3. When comparing libraries incubated for 16 hr and 2 hr, we observed relatively minor variation in HQR for V2B3 and V2C3. We however observed a 2-fold decrease in HQRs for 2 hr incubation in case of V3A8f. The alignment numbers (both TA and HQA) regardless were significantly higher for all libraries using 2 hr incubation time. For V2B3 and V3A8f, the HQA numbers increased from 0 to 428 and 470 respectively, while the increase was approximately 3-fold (11024 vs. 32307) for V2C3. The data show that

shortening the Lys-C protease digestion time for the proteins tested did not negatively impact the quality of the sequencing run, rather we observe the opposite effect indicating improved quality of obtained sequences.

Table 3. Protein digestion time.

Sample	***Digestion time***	***CL (%)***	***HQR***	***TA***	***HQA***
V2B3	16	74.7	21048	123	0
V2B3	2	62.1	21962	976	961
V2C3	16	61.8	125151	11087	11024
V2C3	2	68.7	110034	32359	32307
V3A8f	16	62.6	19779	46	0
V3A8f	2	34.8	8521	486	470

ii. **K-linker incubation time.** The K-linker incubation was another time-consuming step of the standard QSi library preparation protocol targeted for optimization. To explore the impact of the K-linker incubation time on sequencing quality, we prepared four libraries of an inactivated staphylococcal enterotoxin B derivative (SEBv, see methods). The libraries were prepared using a modified QSi protocol with short, 2 hr Lys-C incubation, with four K-linker incubation times at 16 hr, 3 hr, 2 hr, or 1 hr All four libraries were sequenced using the standard QSi protocol with reported CL values between 26.5% (3 hr incubation) and 0.2% (1 hr incubation), shown in Figure 1D, Figure SI.2D, and Table 4.

Table 4. K-linker conjugation time.

Sample	***Conjugation time***	***CL (%)***	***HQR***	***TA***	***HQA***
SEBv	16	19.6	20178	2355	2054
SEBv	3	26.5	25159	2962	2617
SEBv	2	13.9	21388	2079	1901
SEBv	1	0.3	35	0	0

The CL values for the libraries with K-linker incubation of 2 hrs or more fluctuated around 20% with the highest value (26.5%) for the 3 hr incubation and lowest for the 2 hr incubation (13.9%). Despite the low loading (as compared to previously analyzed sdAb

proteins), we observed over 20,000 HQRs for these three libraries with relatively high (~2000) TA and HQA numbers. We noticed, however, that the 2 hr K-linker incubated library compared to the 3 hr incubated one shows a decrease in HQA number by 30% while the HQR is lower by only 15% suggesting a significant deterioration of the read quality. For the 1 hr K-linker incubated library, the CL value dropped below 1%, HQR below 100, and zero alignments to the reference sequence were found. This indicates a complete failure of the library as these values are typically observed in for runs without library loaded onto the chip (Table SI.2). These results therefore indicate the minimum K-linker incubation time for SEBv without significant deterioration is 2-3 hours.

Shortened library protocol. Based on the results presented above, we designed a modified library preparation protocol for use in subsequent parts of this study. The protocol included use of Lys-C for protein digestion, 2 hr Lys-C digestion and 3 hr K-linker incubation. The elimination of overnight incubations means that this modified protocol reduces the total time to generate a high-quality sequencing result from 37 hrs. to merely 10 hrs.

(c) Sequencing run duration. To test the impact of run duration on sequencing quality, we conducted a series of tests with two proteins: SEBv and V2C3. Five sets of runs were conducted with run times decreasing from the standard 10 hrs to 2 hrs CL, HQR, TA and HQA obtained for these runs are shown in Table 5, Figure 1C, and Figure SI.2C. CL values for SEBv ranged from 24.1% to 30.6% and from 21% to 68.7% for V2C3. In both cases, the decrease of runtime resulted in reduced HQRs, which were generally mirrored by the decrease of alignments (both TA and HQA). The decrease of the runtime to 8 hrs. resulted in approximately 20% decrease in HQRs. Both the 6 hr and 4 hr sequencing runs resulted in approximately 80% reduced HQR numbers compared to the 10 hr run. For the 2 hr runs, HQRs were reduced by more than 90% for both samples. These results indicate that, despite a high reduction of obtained reads, 2 hr runs may be still capable of producing enough reads to allow for protein identification. This outcome

however will be strongly dependent on the protein's sequenceability and its concentration in the analyzed sample.

Table 5. Sequencing run time.

Sample	*Run time*	*CL (%)*	*HQR*	*TA*	*HQA*
SEBv	10	26.5	25159	2962	2617
	8	28.1	20651	2066	1787
	6	24.1	5804	670	607
	4	30.6	5124	689	637
	2	27	1387	161	128
V2C3	10	68.7	110034	32359	32307
	8	47.8	87597	35244	35155
	6	21	26326	11743	11720
	4	35.8	28909	13417	13407
	2	32.7	9602	5202	5202

3. Testing optimized protocol for samples of varying composition

(a) sdAbs with diverse sequences

Sixteen different sdAbs were sequenced to test the sequencer's performance for samples with diverse amino acid sequences. See Table 1 for the detailed information about the sdAbs used. The sequencing libraries were prepared using the shortened protocol described above and loaded at 0.2 nM except for three samples loaded at 0.4 nM (V2B3, V2C3 and, V3A8f) for which an earlier version of the library protocol was used (with 2 hr, Lys-C digestions and 16 hr K-linker incubation). The results of the sequencing including CL values and HQR, TA and HQA numbers are shown in Table 6. While multiple samples were sequenced only once, many selected samples were run multiple times (from 2 to 19 replicates) to explore the reproducibility of the sequencing system. We analyzed run-to-run reproducibility of three samples (two sdAbs: a16, WE11f and SEBv) by analyzing variability of CL, HQRs, and both types of alignments (Figure 2, Table 9, and Table SI.3). The variability appears to be relatively high and specific to sample type (higher for sdAbs vs. SEBv). This result most likely reflects a change of library quality as they were subjected to freeze/thaw cycles and impact of environmental factors such as changes in ambient temperatures in which sequencing runs were conducted.

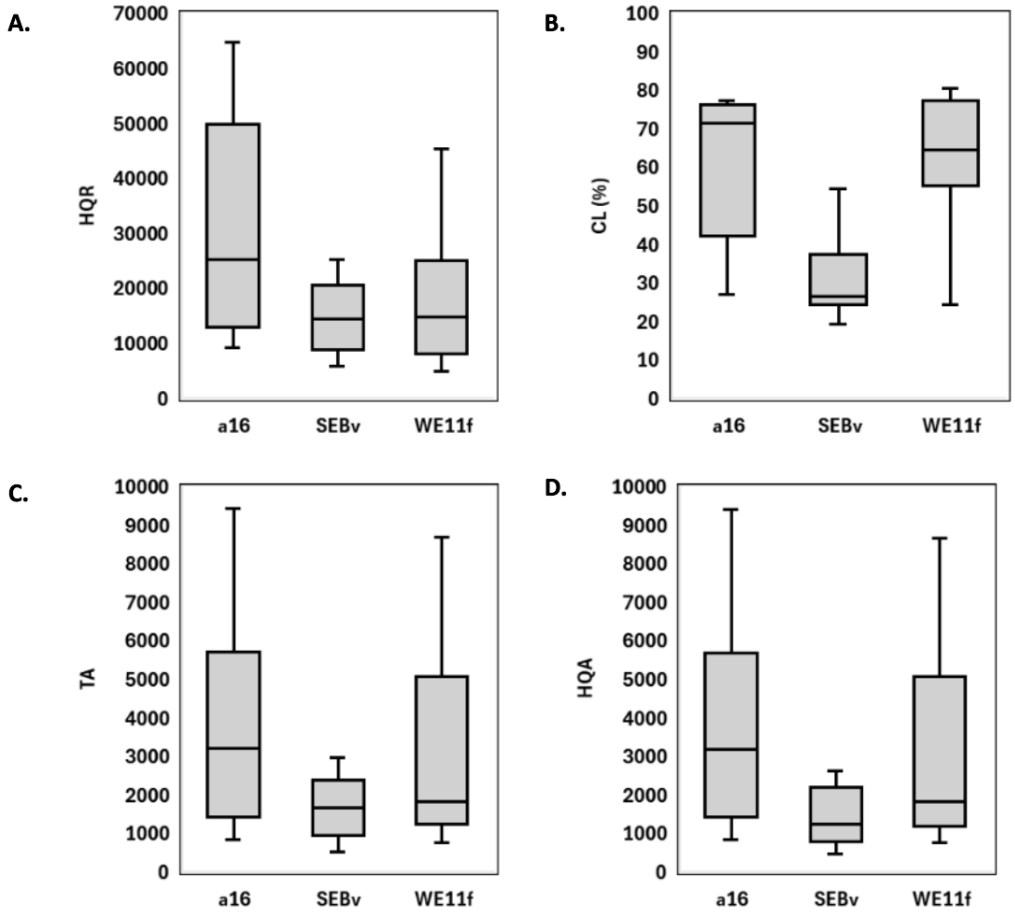

Figure 2 | Variability among QSI sequencing runs across three proteins: a16, SEBv, and WE11f. A. Box and whisker plots of high-quality reads for a16 (N = 8, $\bar{x}$ = 30356, std = 20289), SEBv (N = 9, $\bar{x}$ = 15055, std = 6567), and WE11f (N = 19, $\bar{x}$ = 18237, std = 12735). B. Box and whisker plots of chip loading percentage for a16 ($\bar{x}$ = 61, std = 20), SEBv ($\bar{x}$ = 31, std = 31), WE11f ($\bar{x}$ = 64, std = 14). C. Box and whisker plots of total alignments to each respective protein for a16 ($\bar{x}$ = 3794, std = 2991), SEBv ($\bar{x}$ = 1655, std = 809), and WE11f ($\bar{x}$ = 3213, std = 2494). D. Box and whisker plots of high-quality alignments to each respective protein for a16 ($\bar{x}$ = 3777, std = 2984), SEBv ($\bar{x}$ = 1447, std = 745), and WE11f ($\bar{x}$ = 3190, std = 2485).

CL for all analyzed sdAb runs varied between 34.8% and 75.8% with the majority in the 50%-70% range. HQRs ranged from 1400 (WB9) to 110034 (V2C3) with a median value of 10783. The alignment numbers in general correlate with HQR apart from several exceptions. Despite a significant number of HQRs (ranging from 3632 to 10639) the alignment numbers for WD11f, WF4, a18 and a86 were within levels associated with experimental noise. TA for all four samples were below 40, and there were no HQAs for three of them. Multiple runs of some of these samples have shown similar low alignment numbers making a technical run failure an unlikely explanation of these results. More probably,

an explanation of the low number of alignments is therefore the QSi Platinum sequencing system's inability to sequence peptides comprising certain amino acid sequences (e.g. proline-rich peptides).

Table 6. Sequencing of diverse sdAbs.

Sample*	***Replicates***	***CL***		***HQR***		***TA***		***HQA***	
		%	std dev	n	std dev	n	std dev	n	std dev
V2B3	1	62.1	n/a	21962	n/a	649	n/a	428	n/a
V2C3	1	68.7	n/a	110034	n/a	32331	n/a	32307	n/a
V3A8f	1	34.8	n/a	8521	n/a	478	n/a	470	n/a
WD11f	*2*	*71.7*	*0.2*	*4041*	*1159*	*19*	*1*	*0*	*0*
WE11f	19	64.3	14	18237	12735	3213	2494	3190	2485
WB9	1	52.5	n/a	1400	n/a	170	n/a	168	n/a
WF4	*1*	*46.5*	*n/a*	*10639*	*n/a*	*33*	*n/a*	*0*	*n/a*
WC10	1	75.8	n/a	10928	n/a	1235	n/a	1222	n/a
WH11	1	75.6	n/a	26645	n/a	2181	n/a	2163	n/a
WE10	1	71.3	n/a	15879	n/a	1594	n/a	1540	n/a
a16	9	61.7	17	30388	17894	3884	2650	3864	2643
a18	*2*	*63.4*	*10.2*	*3632*	*231*	*35.0*	*6*	*14.5*	*21*
a19	1	46.7	n/a	5031	n/a	143	n/a	121	n/a
a86	*1*	*52.9*	*n/a*	*4287*	*n/a*	*20*	*n/a*	*0*	*n/a*
a155	1	65.1	n/a	6578	n/a	1038	n/a	1031	n/a
ACVE	3	62.1	2.4	30263	1445	5225	2096	5215	2095

*Library preparations of all sdAbs were prepared using the optimized/shortened protocol including: 2 hr Lys-C digestion, 3 hr K-linker incubation and were loaded at concentration of 0.2 nM except for V2B3, V2C3 and V3A8f for which 16 hr K-linker incubation was performed and they were loaded at 0.4 nM. All libraries were sequenced using standard 10 hr Qsi protocol. Italic font was used for four samples (WD11f, WF4, a18 and, a86) with no or very low HQA numbers indicating low efficiency of sequencing,

(b) Diluted samples.

To explore the QSi Platinum system's ability to sequence samples containing lower initial concentrations of target proteins, we sequenced a series of SEBv library dilutions (Table 7). For the 10x dilution library, both the chip loading and read numbers decreased significantly with CL dropping below 5% and HQRs and alignments to approximately 10% of what is typical in a 0.2 nM library. For the 100x diluted library, both CL, HQRs and alignments dropped to levels observed for runs without samples loaded on the chip.

Table 7. Sequencing of diluted libraries.

Sample	*Library concentration (nM)*	*CL (%)*	*HQR*	*TA*	*HQA*
SEBv	0.2	29.4	22871	2496	2231
	0.02	4.5	2460	264	217
	0.002	0.4	77	9	0

(c) Mixed proteins and complex backgrounds.

To explore the performance of the QSi Platinum sequencing system's ability to sequence both mixed proteins samples and those within complex backgrounds, we conducted two sequencing experiments: i) a mixture of libraries containing SEBv and V2C3 and ii) libraries prepared using crude lysates of *E. coli* cells overexpressing two different proteins (SEBv and ACVE) compared to libraries prepared from partially purified and pure preparations SEBv and ACVE to libraries prepared from the same strain of *E. coli* cells without protein expression plasmids. The results of these experiments are summarized in Table 8, where CL ranged from 22.7% to 62.3%.

Table 8. Sequencing of mixed libraries and samples in complex matrices.

Sample	***Replicates***	***CL***		***HQR***		***SEBv***				***V2C3***			
						TA		**HQA**		**TA**		**HQA**	
		n	std dev	n	std dev	n	std dev	n	std dev	n	std dev	n	std dev
SEBv+V2C3	1	30.0		29277		916		811		10601		10574	
						SEBv				**ACVE**			
SEBv-1	1	32.2	n/a	6541	n/a	381	n/a	179	n/a	13	n/a	0	n/a
SEBv-2	1	22.7	n/a	6577	n/a	458	n/a	390	n/a	4	n/a	0	n/a
SEBv-3	1	47.1	n/a	9796	n/a	764	n/a	683	n/a	6	n/a	0	n/a
ACVE-1	2	34.9	17.3	8018	4723	26	15	0	0	435	243	427	233
ACVE-2	2	49.7	1.7	20505	12262	35	8	0	0	2295	1583	2285	1579
ACVE-3	2	61.0	1.8	29633	1337	29	1	0	0	4019	221	4009	214
E. coli	2	62.3	5.8	11264	5093	88	31	0	0	36	6	0	0

(i) Mixed library. Sequencing the library created by mixing previously prepared libraries of SEBv and V2C3 in 1:1 proportion produced a high number of HQR (n=29277) and alignments with SEBv and V2C3 reference sequences resulting in high numbers of TA and HQA for both references. **N**umbers of alignments for each reference were reduced by approximately 60% when compared to running each library individually: for HQA, the alignments dropped from 2617 to 811 in case of SEBv and from 32307 to 10574 in case of V2C3. This reduction in HQA **is** plausibly a reflection of lower concentrations of peptides from each protein in the library. The results though demonstrate that both proteins may be easily detected within a mixed sample.

(ii) Libraries with complex backgrounds. To investigate the ability of the QSi Platinum system to obtain sequences of proteins present in a background of complex mixtures, we prepared libraries using the following samples: (i) three libraries of crude *E. coli* lysates expressing SEBv, ACVE, or the same *E. coli* cells not containing any expression vectors as a negative control (ii) two libraries of enriched SEBv and ACVE preparations; the first step of protein purification by nickel affinity enrichment (iii) two libraries of SEBv and ACVE preparations fully purified from the aforementioned lysates. All libraries were sequenced using the standard 10 hr sequencing protocol. Libraries containing SEBv were sequenced once while the ACVE and non-expressing *E. coli* libraries were prepared and run in duplicate. For each of the sequencing runs, alignment analysis was performed using SEBv and ACVE reference sequences. These results are summarized in Table 8 and Figure 3. We observed high HQRs for all libraries with generally lower HQR (6541-9796) for SEBv libraries and higher HQR (8017-29632) for libraries containing ACVE and the controls not expressing any extrinsic proteins.

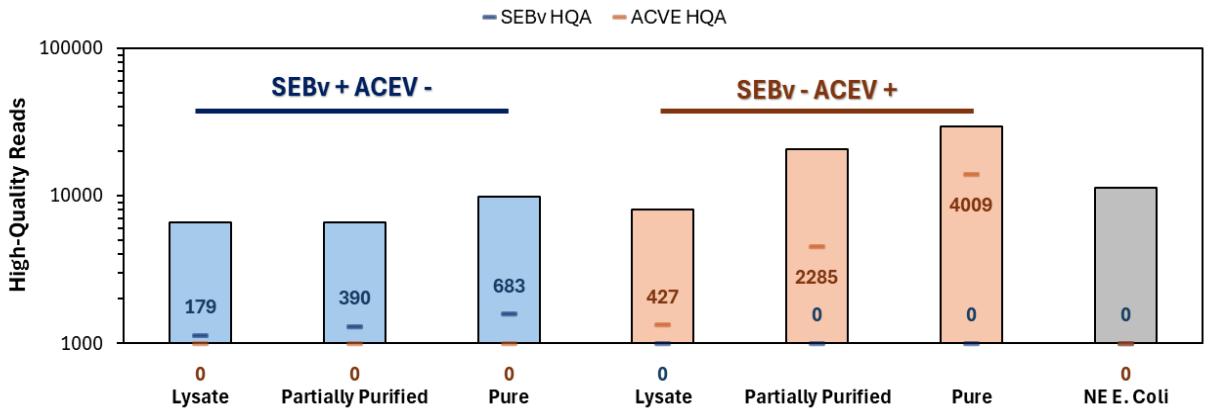

Figure 3 | Sequencing SEBv on QSi Instrument after various stages of protein purification. SEBv high-quality reads (SEBv + ACEV - ; blue bars) are plotted at multiple purification stages after expression in *E. coli*: (i) *E. coli* lysate, (ii) Nickel -purification, (iii) FPLC-purification (see methods). *E. coli* expressing ACVE (SEBv – ACVE +; orange bars) are plotted at multiple purification stages after expression in *E. coli*: (i) *E. coli* lysate, (ii) Nickel-purification, (iii) FPLC-purification. Non-expressing *E. coli* lysate number of high-quality reads (grey bar) shown as negative control. Blue horizontal lines and values represent the number of high-quality alignments to enterotoxin SEBv, whereas orange horizontal lines and values equal high-quality alignments to sdAb ACEV.

Table 9. Run-to-run variability – summary.

Sample	***Replicates (n)***	***CL***		***HQR***		***TA***		***HQA***	
		%	std dev	n	std dev	n	std dev	n	std dev
a16	8	61	20	30356	20289	3794	2991	3777	2984
WE11f	19	64	14	18237	12735	3213	2494	3190	2485
SEBv	9	31	11	15055	6567	1655	809	1447	745

When comparing runs from raw lysate through partially purified to pure protein, we observed an increase in HQRs with increasing purity and decreasing complexity of the library. For both SEBv and ACVE containing lysates, we observed alignment numbers (both TA and HQA) significantly above the background for each specific reference sequence (e.g. SEBv for SEBv containing libraries) while only background level numbers of alignments for the other reference sequence (e.g. ACVE for SEBv containing libraries). The number of alignments increased with increasing purity of the protein in the library with numbers of alignments for pure proteins around an order of magnitude higher than the number for crude cell lysates. We observed background noise level numbers of alignments for both reference sequences for the libraires containing controls not expressing any extrinsic proteins. Notably these runs, especially libraries prepared from crude *E. coli* lysates expressing or not expressing extrinsic

proteins, are complex mixtures containing hundreds of different proteins. Results here in context of the original use case for QSi sequencing, which is recommended for mixtures of no more than 10 proteins, [19] demonstrate that even low HQAs is remarkable as this potentially indicates the ability to detect proteins in complex mixtures, at least for the proteins tested in this study.

Conclusions

Our revised sequencing protocol demonstrated a significant reduction in the sample-to-answer timeline, enabling a complete sequencing workflow in less than 24 hours. Results presented here furthermore suggest that this protocol could still be shortened more by optimizing sequencing runtime without substantial loss in sequencing quality. We additionally confirmed the applicability of this approach for a range of proteins with diverse sequences including single domain antibodies and a staphylococcal enterotoxin. However, protiens with sizes and compositions different to those tested here could perform differently. Regardless, this method will significantly benefit from the improved sequencing chemistry offered in the V4 QSi sequencing kit which allows for improved detection of alanine and serine, permits for the detection of glycine, and expands the sequencable range of proline-containing peptides [21].

Beyond the optimized sequencing protocol presented here, we have shown that it is possible to obtain sequences for diluted target libraries and target proteins in the presence of more complex backgrounds than currently indicated by the manufacturer. While the current QSi protocol does not recommend sequencing of mixtures of more than 10 proteins [19], we were successful in sequencing and detecting targets in matrices of significantly higher complexity. This should also improve with the introduction of the V4 sequencing kit.

However, while these results certainly advance the field of single molecule protein sequencing, especially within time-sensitive applications, a clear limitation lies at the reliance of reference sequences during the analysis of primary data from the instrument. In other words, due to the nature of the current (V3) sequencing chemistry, obtaining *de-novo* sequences for unknown targets is very limited.

To advance the QSi technology on its progress towards truly *de-novo* protein sequencing, significantly improved chemistry is needed. In addition, the analysis of the raw sequencing output will certainly benefit from advancements in data processing technologies taking advantage of rapidly improving machine learning tools. Even with progress here however, significant further advancements are required to sequence modified amino acids and non-proteinogenic amino acids at the heart of rapidly advancing contemporary biotechnology. [22, 23] Overall, the QSi Platinum technology has the potential for application in many areas of proteomics which require rapid high-resolution characterization of complex biological samples to significantly improve protein based diagnostic technologies.

Methods

Protein samples. The protein preparations used in this study included inactivated staphylococcal enterotoxin B (SEBv) and sixteen single domain antibodies (sdAbs). See Table 1 for detailed protein information. All the protein samples were synthesized and purified *in-house* as described below.

SEBv. The coding sequence for SEBv was PCR amplified from the previously described pET15-based expression vector to introduce flanking Nco I and Not I restriction sites [24]. The PCR product was purified using a QIAquick PCR purification kit (Qiagen) prior to digesting with Nco I and Not I. The digested fragment was ligated into pET22b that had been similarly digested and treated with CIP. The SEBv expression vector was transformed into the Tuner (DE3) strain of *E. coli* and a fresh colony was used to start a 50 mL overnight culture in Terrific Broth supplemented with ampicillin at 100 µg/mL (TB-amp). All growth steps in TB-amp, including the overnight, were conducted at 25 °C. The overnight culture was used to inoculate 450 mL TB-amp in a 2L baffled flask and grown for a further ~8 hours before induction with IPTG (0.5 mM final concentration). Induced cultures were grown overnight. The next morning the cells were centrifuged, and cell pellets were subjected to an osmotic shock protocol; cells were kept on ice throughout the osmotic shock procedure. Briefly, the pellets were resuspended in 14 mL of cold Tris-Sucrose buffer (100 mM Tris, 0.75M sucrose pH 7.5) and 1 mL of lysozyme (1 mg/mL made up in the Tris-Sucrose buffer) was added to the homogenized cells. As the

cells were shaking on a rotating platform, 28 mL of 1 mM EDTA was added dropwise. After fifteen minutes, 1 mL of 0.5 M $MgCl_2$ was added and the mix incubated for a further 15 minutes, prior to pelleting the spheroplasts. The supernatant (termed the lysate) was poured into a 50 mL conical tube that contained 5 mL of 10× IMAC buffer (0.2 M Na_2HPO_4, 4 M NaCl, 0.2 M imidazole, pH 7.5) and 0.5 mL of Ni Sepharose (GE Healthcare) that had been equilibrated with 1X IMAC buffer. The sample was tumbled at between 1 and 2 hours at 4 °C on a rotisserie. Afterwards, resin was washed twice in batch with 25 mL 1× IMAC buffer. The resin was poured into a small column, washed with a further ~10 mL 1× IMAC buffer and eluted with 1 mL of 1× IMAC buffer containing 250 mM imidazole. Protein was then further purified into PBS by size exclusion chromatography using a Bio-Rad Enrich SEC70 10 300 and a Bio-Rad Duo-Flow System.

sdAbs. The sdAbs used in this study were purified as described previously [25]. Briefly, expression plasmid was transformed into Tuner (DE3) and a single colony was used to start an overnight culture as described above. The overnight was used to inoculate 450 mL TB-amp, grown for 2 hours followed by induction with IPTG (final concentration 0.5 mM) and grown for a further 2 hours. Cells were pelleted and processed using our standard periplasmic production protocol and purified through a combination of immobilized metal affinity chromatography (IMAC), followed by size exclusion chromatography as described for SEBv above. All protein preparations were aliquoted and stored frozen at -80°C until use.

Library preparation. The purified protein preparations were quantified using Qubit Protein Assay Kit or Qubit Protein BR Assay Kit and Qubit 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA). Then they were processed to prepare peptide libraries ready for loading on sequencing chips according to the Quantum Si Inc. (QSi) library preparation protocol using the Quantum-Si Library Preparation Kit – Lys-C [19].

Briefly, buffer exchange was performed to remove buffer components incompatible with library preparation or sequencing. The buffer replacement was accomplished using Amicon Ultra-0.5 3kDa Centrifugal Filter Units (Sigma-Aldrich, Inc., St. Louis, MO) following the QSi library preparation

protocol. Sample buffer provided in the QSi library preparation reagent kit was used as the replacement buffer. Subsequently, the concentration of the samples was measured again as described earlier and diluted to obtain 100 µL at 5 µM concentration. Diluted sample was subjected to cysteine reduction by 30 min incubation with TCEP (Tris(2-carboxyethyl) phosphine hydrochloride) at 37°C to disrupt disulfide bridges. Next, exposed thiol groups on cysteine side chains were alkylated by 30 min incubation with CAA (Chloroacetamide) at room temperature (RT) to prevent the disulfide bridges from reforming. The reduced proteins were then digested using Lys-C endopeptidase (or trypsin). The endopeptidase digestion was conducted at 37°C overnight/16 hr (standard protocol) or for 2 hr (shortened protocol). As a result of endopeptidase digestion, variable length peptides (with a lysine (K) residue at the C-terminus in case of Lys-C and with a lysine (K) or arginine (R) residue at the C-terminus in case of trypsin) were obtained. In the next step the pH of the peptide mixture was adjusted by adding potassium carbonate (K_2CO_3), and lysines are converted to azido-terminated lysines by 90-minute incubation with ISA (imidazole-1-sulfonyl azide) and copper (II) sulfate ($CuSO_4$) at RT. Then, the obtained mixture was incubated with polyamide beads with shaking for 30 minutes at RT to quench the excess of ISA. Next, the polyamide beads were removed by filtration and the pH of the derivatized peptide mixture was adjusted by addition of acetic acid. In the next step, CTAB and K-linker were added to the peptide preparation and K-linker is conjugated to the peptides during incubation at 37°C overnight/16 hr (standard protocol) or for shorter period (as short as 2hrs). Finally, quantification of the K-linked peptide library was conducted using Qubit 4 Fluorometer using the fluorometer function of the instrument. Measurements were performed using the Blue (470 nm) excitation setting and RFU values from the Green (510-580 nm) emission were used for standard curve generation and sample concentration measurements. The libraries were stored at -80°C until use. The libraries were diluted to 0.2 nM (or in some cases to 0.4 nM) before loading on the sequencing chips.

Protein sequencing. All protein sequencing in this work followed the QSi Platinum Sequencing Kit V3 Protocol using sequencing V3 chips and sequencing chemistry [20]. We optimized this workflow, however, for shortened runs (see Figure 1 and *Supplementary Information*) as compared to the original

QSi method. The QSi chip was first washed three times by pipetting 50µL QSi Wash Buffer into the upper flow cell ports and pipette-mixed ten times. The chip was then loaded into the QSi sequencer for a 'chip-check' with 50µL of QSi Wash Buffer loaded. Following a successful chip-check, the chip was then loaded with 147µL of loading buffer comprising QSi Sequencing Buffer, nuclease-free water, and 3uL of 0.2nM solution of protein library prepared in loading buffer, again by pipette mixing ten times on each side. Following a 15-minute incubation at room-temperature, 30µL of imaging solution (nuclease-free water and kit components: Seq Buffer 3, Quencher, GOx, and Trolox) were loaded to the upper flow cell ports by pipette mixing ten times and the chip was then loaded into the sequencer for a loading check. After a successful loading check, 27µL of Recognizer Solution (nuclease-free water and kit components: Seq Buffer 3, Quencher, GOx, Trolox, COT, and Recognizer Mix 3) was loaded into the upper flow cell ports by pipette mixing ten times and the chip is reinserted into the QSi instrument for sequencing. Following 15 minutes of sequencing, 3uL of aminopeptidase mix (AP mix 3) was added to the chip's upper reservoir, mixed three times with the existing Recognizer Solution previously loaded onto the chip with a 12 µL pipette, and then loaded into the flow cell. The aminopeptidases are mixed by pipetting ten times in both the upper and lower flow cell ports before reloading the chip onto the QSi instrument after the addition of a chip plug for a total of a remaining 9.75-hour sequencing run, 10 hours in total. The run length was adjusted to the total sequencing time to shorter times for length of run experiments. For run-to-run variability see Table 9 and Table SI.3.

Sequencer configuration and sequencing data analysis. Quantum Si Platinum sequencer can be set up in two different configurations which differ in the way the data analysis is performed. In the online configuration the data is sent to a remote server where it is processed and analyzed using a web-based QSi Platinum Analysis Software. However, for this study, we used an offline configuration in which the data is transferred from the sequencing unit to a local server. In the offline system no internet access is needed for data analysis which is performed locally. The data analysis was performed using a web browser-based user interface and involved running analysis scripts provided by QSi. Upon completion of sequencing, all runs described in this study were analyzed by executing

"peptide_alignment_v2.5.2" script which performed both primary data analysis and alignment to a reference sequence.

Primary data analysis provided high level run metrics including chip loading (CL) and number of high-quality reads (HQR) and generated sequencing reads subsequently used for alignment. CL is defined as the percentage of analyzed apertures (chip wells) that were loaded including single-loaded and multi-loaded apertures and HQRs are defined as reads with recognizer read lengths ≥ 4 and unique active recognizers ≥ 3.[26]

Alignment analysis required a reference sequence, which for the most analyses in this study, was the sequence of the analyzed protein. Alignment analysis produced a list of reference sequence-derived peptides with the numbers of sequencing reads aligning with each of these peptides and false discovery rate (FDR) for each of these alignment sets. FDR is an alignment quality metric developed by QSi using a decoy generation method adapted from methods used in peptide identification by mass spectrometry [26]. The alignments with FDR < 0.05 were considered high-quality alignments (HQA) and were a subset of total alignment (TA) number which included all observed alignments.

Statistical and graphics software. Figures and statistical calculations were made using Microsoft Excel (v. 2502) and Python (v. 3.11.11, https://www.python.org) with the following packages: seaborn (v. 0.13.2, https://seaborn.pydata.org), pandas (v. 2.2.3, https://pandas.pydata.org), matplotlib (v. 3.10.0, https://matplotlib.org), and numpy (v. 2.2.5, https://numpy.org).

Acknowledgments

This work was supported by the Defense Threat Reduction Agency and the National Academies of Sciences Engineering and Medicine's NRC Research Associateship Program. The opinions and assertions contained herein are those of the authors and are not to be construed as those of the U.S. Navy, military service at large or U.S. Government. T.A.L., Z.T.J., S.N.D., E.R.G., and D.A.S. are employees of the US Government. DISTRIBUTION STATEMENT A. Approved for public release: distribution is unlimited.

Conflicts of Interest

The authors completed this work while in collaborative efforts with QSi. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Alfaro, J.A., et al., *The emerging landscape of single-molecule protein sequencing technologies.* Nature Methods, 2021. **18**: p. 604.
2. Restrepo-Pérez, L., C. Joo, and C. Dekker, *Paving the way to single-molecule protein sequencing.* Nature Nanotechnology, 2018. **13**: p. 786.
3. Edman, P., et al., *Method for Determination of the Amino Acid Sequence in Peptides.* Acta Chemica Scandinavica, 1950. **4**(7): p. 283.
4. Miyashita, M., et al., *Attomole level protein sequencing by Edman degradation coupled with accelerator mass spectrometry.* Proceedings of the National Academy of Sciences, 2001. **98**(8): p. 4403.
5. Deol, H., et al., *After 75 Years, an Alternative to Edman Degradation: A Mechanistic and Efficiency Study of a Base-Induced Method for N-Terminal Peptide Sequencing.* Journal of the American Chemical Society, 2025. **147**(16): p. 13973.
6. Reed, B.D., et al., *Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device.* Science, 2022. **378**(6616): p. 186.
7. Roberts, D.S., et al., *Top-down proteomics.* Nature Reviews Methods Primers, 2024. **4**(38).
8. Bordin, N., et al., *Novel machine learning approaches revolutionize protein knowledge.* Trends in Biochemical Sciences, 2023. **48**(4): p. 345.
9. Gallo, E., *Revolutionizing Synthetic Antibody Design: Harnessing Artificial Intelligence and Deep Sequencing Big Data for Unprecedented Advances.* Molecular Biotechnology, 2025. **67**: p. 410.
10. Reed, B.D., et al., *Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device.* Science, 2022. **378**(6616): p. 186-192.
11. Edman, P., *A method for the determination of amino acid sequence in peptides.* Arch Biochem, 1949. **22**(3): p. 475.
12. Swaminathan, J., et al., *Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures.* Nat Biotechnol, 2018.
13. O'Bryon, I., S.C. Jenson, and E.D. Merkley, *Flying blind, or just flying under the radar? The underappreciated power of de novo methods of mass spectrometric peptide identification.* Protein Sci, 2020. **29**(9): p. 1864-1878.
14. Cao, C., et al., *Single-molecule sensing of peptides and nucleic acids by engineered aerolysin nanopores.* Nat Commun, 2019. **10**(1): p. 4918.
15. Wilson, J., et al., *Rapid and Accurate Determination of Nanopore Ionic Current Using a Steric Exclusion Model.* ACS Sens, 2019. **4**(3): p. 634-644.
16. Di Muccio, G., et al., *Insights into protein sequencing with an alpha-Hemolysin nanopore by atomistic simulations.* Sci Rep, 2019. **9**(1): p. 6440.
17. Ryu, J., et al., *Direct biomolecule discrimination in mixed samples using nanogap-based single-molecule electrical measurement.* Sci Rep, 2023. **13**(1): p. 9103.

18. Taniguchi, M., *Combination of Single-Molecule Electrical Measurements and Machine Learning for the Identification of Single Biomolecules.* ACS Omega, 2020. **5**(2): p. 959-964.
19. Quantum Si Inc, *Library Preparation Kit - Protein Protocol*. 2024, Quantum Si Inc.: Branford, CT.
20. Quantum Si Inc, *Platinum Sequencing Kit V3 Protocol*. 2024, Quantum Si Inc,: Branford, CT.
21. Quantum Si Inc. *Unlock Proteome Depth with the Platinum® Pro Instrument and Sequencing Kit V4*. [PDF] 2025 9/8/2025 [cited 2025 10/14/2025]; Product Data Sheet]. Available from: https://www.quantum-si.com/wp-content/uploads/Q423-DS-Sequencing-Kit-V4-090825.pdf.
22. Brown, S.M., C. Mayer-Bacon, and S. Freeland, *Xeno Amino Acids: A Look into Biochemistry as We Do Not Know It.* Life (Basel), 2023. **13**(12).
23. Brown, S.M., E. Allgair, and R. Krystufek, *Mapping the Edges of Mass Spectral Prediction: Evaluation of Machine Learning EIMS Prediction for Xeno Amino Acids.* Anal Chem, 2025. **97**(19): p. 10282-10288.
24. Anderson, G.P., et al., *Comparison of immunoreactivity of staphylococcal enterotoxin B mutants for use as toxin surrogates.* Anal Chem, 2012. **84**(12): p. 5198-203.
25. Goldman, E.R., et al., *A single domain antibody-based Luminex assay for the detection of SARS-CoV-2 in clinical samples.* Front Immunol, 2024. **15**: p. 1446095.
26. Quantum Si Inc, *Platinum® Next-Generation Protein Sequencing™ Advanced Data Analysis Technical Note*. Quantum Si Inc,.
27. Liu, J.L., et al., *Bivalent single domain antibody constructs for effective neutralization of Venezuelan equine encephalitis.* Sci Rep, 2022. **12**(1): p. 700.
28. Liu, J.L., et al., *Selection of Single-Domain Antibodies towards Western Equine Encephalitis Virus.* Antibodies (Basel), 2018. **7**(4).
29. Liu, J.L., et al., *Generative Deep Learning Design of Single-Domain Antibodies Against Venezuelan Equine Encephalitis Virus.* Antibodies (Basel), 2025. **14**(2).
30. Anderson, G.P., et al., *Oriented Immobilization of Single-Domain Antibodies Using SpyTag/SpyCatcher Yields Improved Limits of Detection.* Anal Chem, 2019. **91**(15): p. 9424-9429.
31. Ulrich, R.G., M.A. Olson, and S. Bavari, *Development of engineered vaccines effective against structurally related bacterial superantigens.* Vaccine, 1998. **16**(19): p. 1857-1864.