

New Results

[Follow this preprint](#) [Previous](#)

[Next](#)

Next-Generation Protein Sequencing and individual ion mass spectrometry enable complementary analysis of interleukin-6

Posted February 08, 2025.

[Kenneth A Skinner PhD](#), [Troy D Fisher](#), [Andrew Lee PhD](#), [Taojunfeng Su](#), [Eleonora Forte PhD](#), [Aniel Sanchez PhD](#), [Michael A Caldwell PhD](#), [Neil L Kelleher PhD](#)

doi: <https://doi.org/10.1101/2025.02.07.637157>

This article is a preprint and has not been certified by peer review [what does this mean?].

[Download PDF](#)
[Print/Save Options](#)

[Email](#)
[Share](#)
[Citation Tools](#)
[Get QR code](#)

Abstract | [Info/History](#) | [Metrics](#) | [Preview PDF](#)

Subject Area

Biochemistry

Abstract

The vast complexity of the proteome currently overwhelms any single analytical technology in capturing the full spectrum of proteoform diversity. In this study, we evaluated the complementarity of two cutting-edge proteomic technologies—single-molecule protein sequencing and individual ion mass spectrometry—for analyzing recombinant human IL-6 (rhIL-6) at the amino acid, peptide, and intact proteoform levels. For single-molecule protein sequencing, we employ the recently released Platinum® instrument. Next-Generation Protein Sequencing™ (NGPS™) on Platinum utilizes cycles of N-terminal amino acid recognizer binding and aminopeptidase cleavage to enable parallelized sequencing of single peptide molecules. We found that NGPS produces single amino acid coverage of multiple key regions of IL-6, including two peptides within helices A and C which harbor residues that reportedly impact IL-6 function. For top-down proteoform evaluation, we use individual ion mass spectrometry (I2MS), a highly parallelized orbitrap-based charge detection MS platform. Single ion detection of gas-phase fragmentation products (I2MS2) gives significant sequence coverage in key regions in IL-6, including two regions within helices B and D that are involved in IL-6 signaling. Together, these complementary technologies deliver a combined 52% sequence coverage, offering a more complete view of IL-6 structural and functional diversity than either technology alone. This study highlights the synergy of complementary protein detection methods to more comprehensively cover protein segments relevant to biological interactions.

Competing Interest Statement

KAS is a shareholder and former employee of Quantum-Si. NLK is engaged in various entrepreneurial activities related to TD-MS.

Copyright The copyright holder for this preprint is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](#).

bioRxiv and medRxiv thank the following for their generous financial support:

The Chan Zuckerberg Initiative, Cold Spring Harbor Laboratory, the Sergey Brin Family Foundation, California Institute of Technology, Centre National de la Recherche Scientifique, Fred Hutchinson Cancer Center, Imperial College London, Massachusetts Institute of Technology, Stanford University, University of Washington, and Vrije Universiteit Amsterdam.

[Back to top](#)

Reviews and Context

- [0 Comment](#)
- [0 TRIP Peer Reviews](#)
- [0 Community Reviews](#)
- [1 Automated Services](#)
- [0 Blogs/Media](#)
- [0 Author Videos](#)

Subject Areas

All Articles

- Animal Behavior and Cognition
- Biochemistry
- Bioengineering
- Bioinformatics
- Biophysics
- Cancer Biology
- Cell Biology
- Clinical Trials*
- Developmental Biology
- Ecology
- Epidemiology*
- Evolutionary Biology
- Genetics
- Genomics
- Immunology
- Microbiology
- Molecular Biology
- Neuroscience
- Paleontology
- Pathology
- Pharmacology and Toxicology
- Physiology
- Plant Biology
- Scientific Communication and Education
- Synthetic Biology
- Systems Biology
- Zoology

* The Clinical Trials and Epidemiology subject categories are now closed to new submissions following the completion of bioRxiv's clinical research pilot project and launch of the dedicated health sciences server medRxiv (submit.medrxiv.org). New papers that report results of Clinical Trials must now be submitted to medRxiv. Most new Epidemiology papers also should be submitted to medRxiv, but if a paper contains no health-related information, authors may choose to submit it to another bioRxiv subject category (e.g., Genetics or Microbiology).

Next-Generation Protein Sequencing and individual ion mass spectrometry enable complementary analysis of interleukin-6

Kenneth A. Skinner

Quantum-Si, Incorporated, Branford, Connecticut, United States

ORCID ID: 0000-0001-6407-5034

Troy D. Fisher

Proteomics Center of Excellence, Northwestern University, Evanston, Illinois, United States

ORCID ID: 0009-0006-1306-0397

Andrew Lee

Departments of Molecular Biosciences, Chemistry and Chemical and Biological Engineering,
Northwestern University, Evanston, IL, United States

ORCID ID: 0000-0002-7665-1999

Taojunfeng Su

Departments of Molecular Biosciences, Chemistry and Chemical and Biological Engineering,
Northwestern University, Evanston, IL, United States

ORCID ID: 0000-0002-7111-1248

Eleonora Forte

Proteomics Center of Excellence, Northwestern University, Evanston, Illinois, United States
Department of Surgery, Feinberg School of Medicine, Comprehensive Transplant Center,
Northwestern University, Chicago, Illinois, United States

ORCID ID: 0000-0002-6095-2797

Aniel Sanchez

Proteomics Center of Excellence, Northwestern University, Evanston, Illinois, United States

ORCID ID: 0000-0002-0278-8802

Michael A. Caldwell

Proteomics Center of Excellence, Northwestern University, Evanston, Illinois, United States
Department of Medicine, Division of Hematology Oncology, Feinberg School of Medicine,
Northwestern University, Chicago, IL, United States

ORCID ID: 0000-0002-8636-0706

Neil L. Kelleher[#]

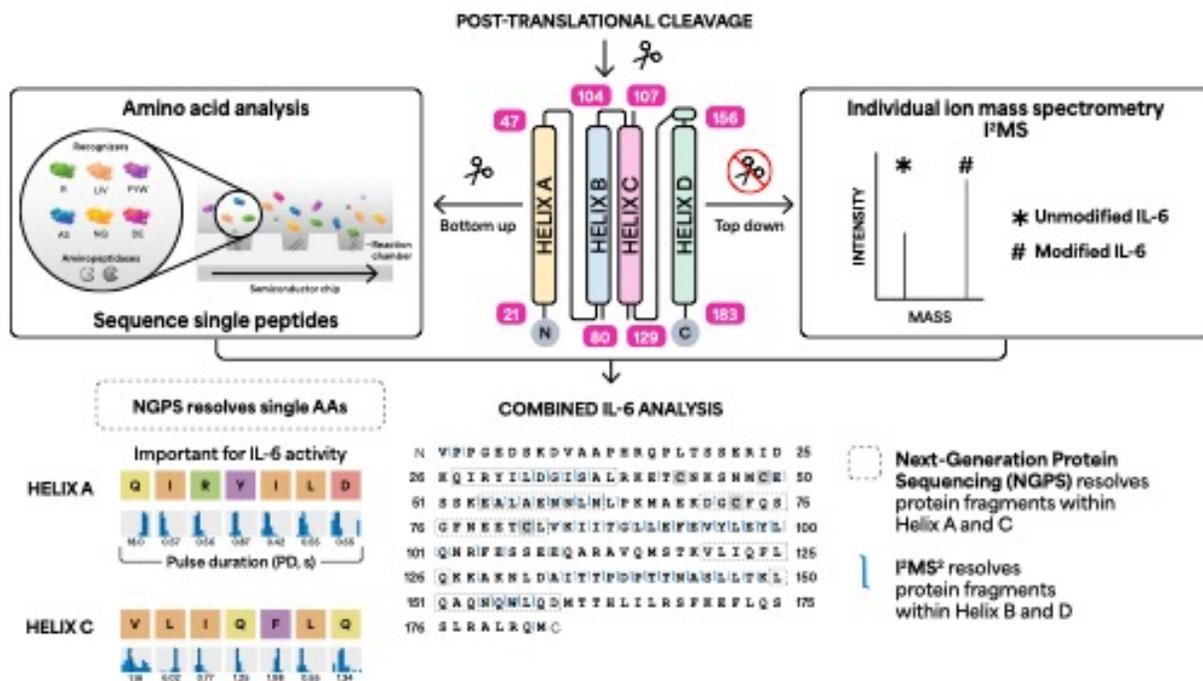
Proteomics Center of Excellence, Northwestern University, Evanston, Illinois, United States
Departments of Molecular Biosciences, Chemistry and Chemical and Biological Engineering,
Northwestern University, Evanston, IL, United States
Department of Biochemistry and Molecular Genetics, Feinberg School of Medicine,
Northwestern University, Chicago, IL, United States

ORCID ID: 0000-0002-8815-3372

[#]Corresponding author email: n-kelleher@northwestern.edu

Abstract

The vast complexity of the proteome currently overwhelms any single analytical technology in capturing the full spectrum of proteoform diversity. In this study, we evaluated the complementarity of two cutting-edge proteomic technologies—single-molecule protein sequencing and individual ion mass spectrometry—for analyzing recombinant human IL-6 (rhIL-6) at the amino acid, peptide, and intact proteoform levels. For single-molecule protein sequencing, we employ the recently released Platinum[®] instrument. Next-Generation Protein Sequencing[™] (NGPS[™]) on Platinum utilizes cycles of N-terminal amino acid recognizer binding and aminopeptidase cleavage to enable parallelized sequencing of single peptide molecules. We found that NGPS produces single amino acid coverage of multiple key regions of IL-6, including two peptides within helices A and C which harbor residues that reportedly impact IL-6 function. For top-down proteoform evaluation, we use individual ion mass spectrometry (I²MS), a highly parallelized orbitrap-based charge detection MS platform. Single ion detection of gas-phase fragmentation products (I²MS²) gives significant sequence coverage in key regions in IL-6, including two regions within helices B and D that are involved in IL-6 signaling. Together, these complementary technologies deliver a combined 52% sequence coverage, offering a more complete view of IL-6 structural and functional diversity than either technology alone. This study highlights the synergy of complementary protein detection methods to more comprehensively cover protein segments relevant to biological interactions.



Keywords

Single-molecule protein sequencing, Next-Generation Protein Sequencing, Platinum, cytokine, top-down mass spectrometry, individual ion mass spectrometry, proteoform

Statements and Declarations

Competing Interests

KAS is a shareholder and former employee of Quantum-Si. NLK is engaged in various entrepreneurial activities related to TD-MS.

Introduction

Proteins are fundamental to nearly all biological processes, serving as key cellular components and prime drug targets [1, 2]. Protein-based therapies are among the top-selling pharmaceutical products [3], and recombinant cytokines and biologic inhibitors of cytokine-mediated signaling are used clinically [4, 5]. The structure and function of proteins are determined by their primary amino acid sequence and post-translational modifications (PTMs), making the analysis of these features crucial for drug development and biopharmaceuticals [6, 7].

Rational design and directed evolution are frequently employed to engineer proteins with desirable properties [8, 9]. However, amino acid substitutions, errors in protein synthesis, and variations in expression systems can lead to subtle differences in protein structure and function, necessitating highly sensitive methods for detection [10, 11]. Furthermore, proteins often exist in multiple forms, or proteoforms [12], generated not only by genetic variation but also by alternative splicing and PTMs. Current proteomic technologies struggle to fully capture the complexity and heterogeneity of proteoforms, underscoring the need for innovative solutions [1, 13-16].

Several new technologies offer promising approaches to overcoming these challenges. Next-Generation Protein Sequencing (NGPS) uses engineered protein-based recognizers to sequence peptides with single amino acid resolution, while individual ion mass spectrometry (I²MS) is a mass spectrometry technique that measures the intact mass of proteins. Both NGPS and I²MS are highly parallelized, single-molecule approaches to analyze protein sequences and PTMs.

Proof-of-concept studies support the enormous potential of single-molecule protein sequencing and analysis [17-19]. The first of these technologies to be commercialized is NGPS on the Platinum and Platinum Pro instruments, which employ fluorophore-labeled N-terminal amino acid (NAA) recognizers that reversibly bind cognate NAAs [20]. These sensitive biophysical interactions produce kinetic signatures that summarize the average sequencing behavior of an ensemble of single peptide molecules, including their primary sequence and PTMs.

The application of individual ion technology to top-down mass spectrometry (TD-MS) provides a >500x improvement in analytical sensitivity, >10x increase in mass range, and >10x higher resolution for the characterization of proteoforms than traditional TD-MS approaches [21, 22]. I²MS uses multiplexed orbitrap-based charge detection mass spectrometry (CDMS) to directly measure the charge of individual proteoform ions, producing significantly simplified mass-domain spectra for mixtures of proteoforms without the challenges of deconvolution [23]. Mass differences between the measured protein molecular weight and that of its DNA-predicted sequence may reveal sequence variations and/or PTMs [24]. Further, tandem-MS with the detection of individual ions (I²MS²) resolves fragment ions from the measured molecular ion to provide sequence analysis across the whole protein, confident protein identification and localization of PTMs [21].

NGPS and I²MS utilize distinct fluorescence- or mass spectrometry-based techniques to characterize proteins and their proteoforms. In the context of biopharmaceutical quality control, there is an increasing demand for orthogonal methods—those that employ different physical principles to measure the same property—and for complementary approaches that provide a more comprehensive assessment of protein characteristics [25]. Both I²MS and NGPS represent promising single-molecule technologies that meet these needs. Moreover, when used together, NGPS and I²MS can provide complementary insights, covering overlapping and distinct protein regions, thus enhancing overall sequence coverage and depth.

Cytokines, such as interleukin-6 (IL-6), play a vital role in immune regulation and are widely used in therapeutic applications [5, 9, 26-28]. The functional versatility and adjustable properties of cytokines have driven the development of engineered variants with optimized pharmacological profiles and enhanced biological activity [9, 29-32]. Beyond genetic modifications, cytokines can undergo alternative splicing [33], post-translational proteolysis [34], and various PTMs [35], all of which can affect their stability and interactions with biological targets [36, 37]. These variations may also differ depending on the expression system used [38, 39]. Understanding and controlling these factors is essential for refining recombinant IL-6 formulations and optimizing their clinical efficacy.

In this study, we use two independent technologies to analyze a recombinant human IL-6 (rhIL-6) sample: NGPS for single-molecule protein sequencing [20] and I²MS² for intact mass profiling and readout of top-down fragmentation spectra [22]. By integrating these orthogonal methods for the first time, we show how they complement each other to provide more comprehensive coverage of key regions crucial for IL-6 function and therapeutic potential.

Materials and Methods

Peptide sequencing on Platinum

Experiments were conducted in accordance with [Library Preparation Kit -Lys-C Data Sheet](#) and [Platinum Instrument and Sequencing Kit V3 Data Sheet](#) (February 27, 2024).

Platinum uses bottom-up processing to sequence single peptides and identify proteins. Briefly, the disulfide bonds of rhIL-6 sourced from AcroBiosystems (IL6-H4218) were reduced, enabling alkylation of cysteine residues. Digestion of rhIL-6 with Lys-C endoproteinase [40] generated peptides with C-terminal lysine residues. These peptides underwent diazotransfer [41] and bioconjugation to a macromolecular linker [42, 43]. Conjugated peptides were immobilized in nanoscale reaction chambers on a semiconductor chip for exposure to a mixture of freely diffusing NAA recognizers and aminopeptidases [20]. The sequencing mixture consisted of six NAA recognizers that target 13 NAAs. During on-chip sequencing, fluorophore-labeled NAA recognizers reversibly bind cognate NAAs, producing recognition segments (RSs) and fluorescence properties that are captured by the semiconductor chip. To carry out the sequencing process, aminopeptidases cleave the peptide bond and expose the subsequent NAA for recognition.

Analysis of sequencing data; analysis versions

Primary Analysis v2.5.1, Peptide Alignment v2.3.0, and Protein Inference v2.5.2 were used for cloud-based analysis of sequencing data. Details can be found in the [Platinum Analysis Software Data Sheet](#) (February 2, 2024). The Primary Analysis workflow is the first step in processing data which characterizes the apertures across the chip based on peptide loading, recognizer activity, recognizer reads, and recognizer read lengths.

Peptide Alignment v2.3.0 workflow

For the Peptide Alignment workflow, a reference sequence is required to call amino acids from the recognizer reads. Reads from the sequencing data were aligned to the FASTA reference of mature rhIL-6. Peptides were aligned based on the correspondence of observed recognition segments to the expected reference profile, using recognizer identity.

Peptide Alignment workflow also computes a false discovery rate (FDR) for each aligned peptide. This calculation is adapted from methods used in peptide identification by mass spectrometry [44], based on decoy peptide matching. Thus, FDR represents the relative number of alignments to the reference peptide sequence versus the total number of off-target alignments, such as scrambled sequences with the same length as the target peptide.

Protein Inference workflow v2.5.2

A pre-defined reference set of 8,076 human proteins was used to infer proteins from unknown samples and confirm the identity of samples. The proteins in this reference panel

span 10–70 kDa and contain at least three in silico LysC-digested peptides with three unique, visible residues. Inferred proteins are ranked by their respective Inference Score. The Inference Score is a natural log calculation of the FDR associated with the inferred protein.

Sample preparation and data collection for I²MS and I²MS² analysis

RhIL-6 (AcroBiosystems, IL6-H4218) was reconstituted in Optima™ LC/MS grade water (Fisher Scientific, W64), aliquoted, and stored at -80°C until used following the manufacturer's recommendations. Samples were analyzed with the SampleStream Platform (Integrated Protein Technologies, Evanston, IL) [45] coupled to a modified Q-Exactive HF (Thermo Fisher Scientific; Bremen, Germany) mass spectrometer. Briefly, rhIL-6 was diluted to ~0.8 μM with water and transferred to a low-retention autosampler vial (Waters, 186009186). For each injection, 10 μL was buffer exchanged with SampleStream into 80 μL denaturing buffer (70:30 water:acetonitrile with 0.2% formic acid), deposited into a clean vial to mix, and then 75 μL aspirated to infuse at ~0.1 μM. SampleStream parameters include a 125 μL push volume, 125 μL focus volume, 225 μL focus flow rate, 60°C flow cell temperature, and a 5 kDa molecular weight cut-off membrane. Source conditions include a custom nano-electrospray emitter (CoAnn Tech, TIP36007540-10), 1.8 kV spray voltage, 1.0 μL/min flow rate, and 320°C inlet capillary temperature. Instrument parameters include RF 50%, 5e6 AGC target, 120,000 resolution, 1 μscan, -1 kV central electrode voltage, 0.3 (arb) trapping gas pressure setting, 600-2500 *m/z* scan range, and a 78-minute acquisition length (4583 scans). Injection time was determined via automated ion control (AIC) to remain on the individual ion level [46].

For I²MS² experiments, a charge state for each precursor was selected using The Fisher, a tool developed internally by the Kelleher group [47]. Briefly, ions within a 0.8 *m/z* isolation window width and 20 Da mass window width centered on a precursor mass were counted as ions corresponding to either the desired precursor or other species. The charge state which maximized the number of desired precursor ions and minimized the number of ions from other species was selected for isolation. Precursors were isolated and fragmented via higher-energy collisional dissociation (HCD) normalized to charge state. Fragment ions were measured within a 150-2500 *m/z* scan range. Normalized collisional energy (NCE) and injection time were set manually to optimize fragmentation and ion counts.

I²MS and I²MS² data processing and analysis

The I²MS charge assignment and ion mass determination process has been previously described [22, 48]. The *m/z* of individual ions were determined through normal Fourier-transform Mass Spectrometry (FTMS) analysis, while the charge of the ion was determined using the summation of its ion signal as a function of time (STORI plot) [48]. Once the quantized charge and *m/z* values of the ion were assigned, the mass of the species was calculated and utilized to create a true mass spectrum.

A fragment search against the precursor amino acid sequence and expected PTMs was carried out using TDValidator (Proteinaceous, Inc., Evanston, IL). Fragment ions were identified by matching their isotopic distributions to theoretical isotopic distributions generated using an averagine model [49] and the Mercury7 [50] algorithm. To make ions searchable in TDValidator, neutral mass I^2MS spectra were transformed into theoretical +1 (M+H) distributions. All fragment ions were identified within a ± 10 ppm tolerance of their theoretical values for the isotopic distribution (Max PPM tolerance) and ± 5 ppm tolerance for isotopologues within the same distribution (Sub PPM tolerance). Other search metrics included a +1 charge state, 1.5 S/N cutoff, and 0.01 score cutoff. Spectra were manually curated to remove poor fragment ion matches. To calculate P-scores, fragment monoisotopic masses were generated with the THRASH algorithm in TDValidator using a 1.5 S/N cutoff, charge +1, 30,000 Da maximum mass, and 0.9 minimum RL value and uploaded to ProSight Lite (<http://prosightlite.northwestern.edu/>) with 10 ppm tolerances.

Results and Discussion

To evaluate the ability of each proteomic technology to detect key regions of rhIL-6, we first performed NGPS of rhIL-6 using the Platinum instrument (**Figure 1A**). The Platinum platform includes three components: kits for bottom-up protein processing and sequencing; a benchtop Platinum instrument that accommodates semiconductor chips for sequencing of polypeptides (**Figure 1B and C**); and cloud-based software for analysis of sequencing data (**Figure 1D and E**).

For on-chip sequencing, surface immobilized peptides are exposed to a mixture of freely diffusing NAA recognizers and aminopeptidases (**Figure 1B**) [20]. Six NAA recognizers, labeled with different fluorophores, reversibly bind 13 target NAAs (**Figure 1B**) and elicit characteristic pulsing patterns upon binding to each NAA (**Figure 1C**). The semiconductor chip converts fluorescence signal into digital readouts, enabling real-time sequencing of single peptide molecules in parallel [20]. Cycles of binding and cleavage proceed to sequentially reveal the order of NAAs and enable identification of the peptide sequence [20].

For analysis of sequencing data, only peptides that meet specific thresholds (see **Methods** section) are eligible for high-confidence alignment to the reference rhIL-6 sequence (**red asterisks, Figure 1D**). Based on these criteria, three rhIL-6 peptides ($V_{1PPGEDSK_8}$, $E_{41TCNK_{45}}$, and $M_{66AEK_{69}}$) are ineligible for alignment. In addition, the C-terminal segment 171 – 183 ($E_{171FLQSSLRALRQM_{183}}$) does not contain K and thus is not amenable to conjugation and on-chip sequencing. NGPS sequences five rhIL-6 peptides (**Figure 1E**) and determines the identity of 46/183 single amino acids, indicating ~25% amino acid level coverage within rhIL-6 (**Figure 1E**).

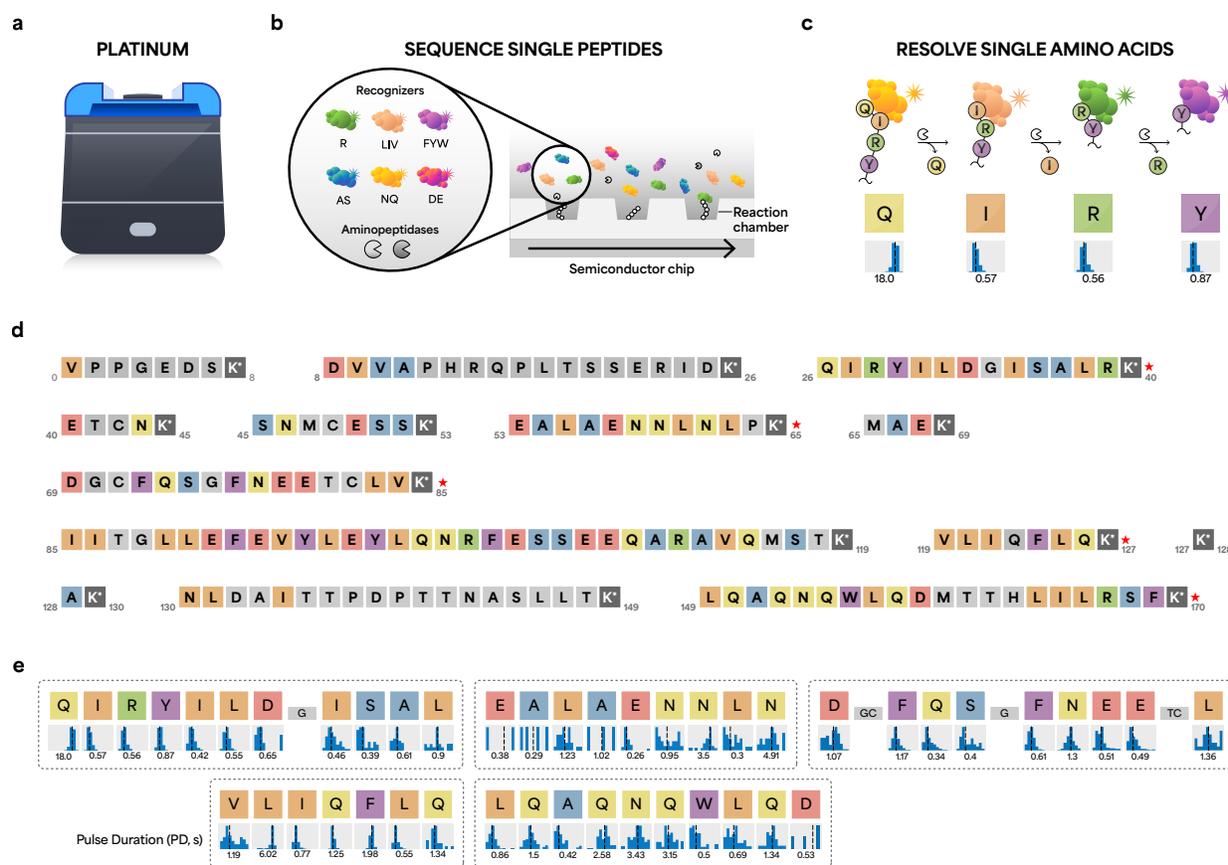


Figure 1. Next-Generation Protein Sequencing (NGPS) of IL-6 with the Platinum instrument.

(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. (d) Results of *in silico* Lys-C digestion of mature IL-6. Missing C-terminal residues 171-183 lack a lysine residue (K) and thus are not amenable for peptide capture and sequencing on Platinum. Colored boxes indicate potential recognition events. Gray boxes indicate amino acids not amenable to recognition. Red asterisks indicate automated alignments of sequencing data to reference sequence. (e) NGPS of rhIL-6 on Platinum detects five IL-6 peptides (dotted rectangles). NAA recognizers elicit kinetic signatures such as pulse duration (PD), which reflects the affinity between specific recognizers and NAAs. PD histograms represents the statistical distribution of kinetic data for all pulses associated with a specific residue and supports inference of the corresponding amino acid. Values are reported as the median of the mean across reaction chambers.

To support high-confidence identification of IL-6, we also examined the false discovery rate (FDR) for each peptide and used Protein Inference analysis (see **Methods section**) to determine the specificity of IL-6 mapping relative to a protein panel. Platinum analysis outputs FDR for each sequenced peptide using a target decoy approach that is analogous to methodologies employed in MS [44]. FDR scores for each peptide are Q₂₇IRYILDGISALRK₄₀ (FDR: 0.0), V₁₂₀LIQFLQK₁₂₇ (FDR: 0.0),

D₇₀GCFQSGFNEETCLVK₈₅ (FDR: 0.02), E₅₄ALAENNLNLPK₆₅ (FDR: 0.07), and L₁₅₀QAQNQLQDMTTHLILRSFK₁₇₀ (FDR: 0.09). Interestingly, an FDR score of 0.0 was computed for peptide V₁₂₀LIQFLQK₁₂₇, despite contiguous isobaric residues leucine (L) in position 2 and isoleucine (I) in position 3 (**Figure 1E**). The recognizer for branched chain NAAs (LIV) exhibits differential pulse duration (PD) profiles for L (6.02 s) and I (0.77 s), discerning the order of L and I residues with the same mass. This example demonstrates that a single NAA recognizer can differentiate NAAs with similar physiochemical properties on the basis of differences in PD. In addition to FDR, we also used the Protein Inference analysis software (see Methods), which screens the sequencing data against a reference panel of 8,076 proteins. IL-6 was identified as the top protein hit with 99.99% confidence (**Supplementary Figure 1**). These results demonstrate NGPS confidently identifies rhIL-6 with single amino acid resolution.

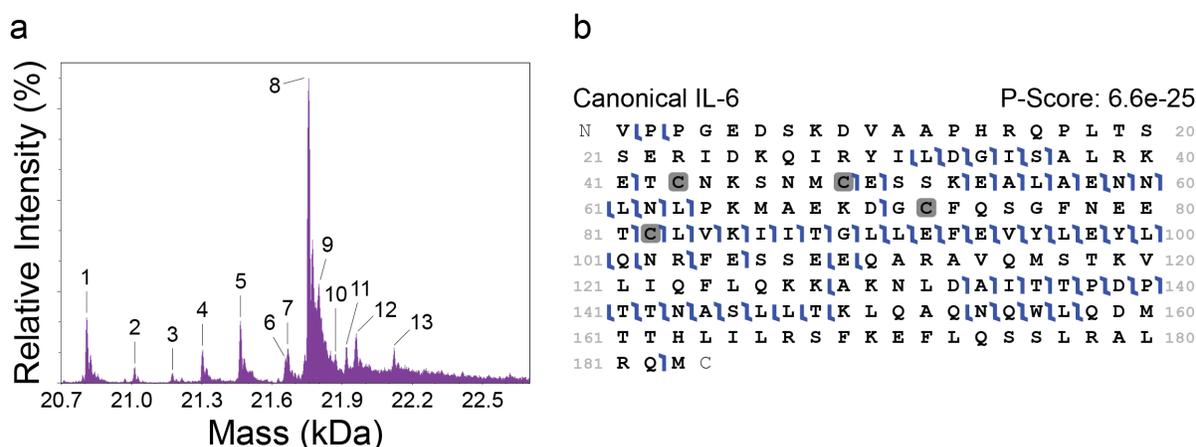


Figure 2. I²MS detects intact rhIL-6 proteoforms. I²MS² covers broad regions of rhIL-6 primary structure. **(a)** The application of I²MS to TD-MS enables measurement of the intact mass of intact protein ions of canonical rhIL-6 (~20.8 kDa). Higher mass proteoforms indicate the addition of PTMs. **(b)** TD-MS provides broad sequence coverage of canonical rhIL-6 (PFR 1) by higher energy collision dissociation (HCD), achieving 39% sequence coverage.

Next, we deployed the same rhIL-6 sample for intact mass measurement via I²MS. I²MS accurately identifies unmodified rhIL-6 (~20.8 kDa) and several higher molecular weight proteoforms up to ~22 kDa (**Figure 2A**). Human IL-6 has been shown to undergo a number of PTMs including O- and N-linked glycosylation and phosphorylation [35]. O-glycans have been reported in species under 25 kDa, while heavier species up to 29 kDa have been reported to contain N-glycans, though few reports detail the composition or localization of IL-6 O-glycans [51]. Using the intact masses of these proteoforms, the composition of the putative O-glycans was determined by searching the GlyGen database [52] for those glycans corresponding to the observed intact mass shift, and their chemical formula and mass were calculated using the publicly available NIST Glyco Mass Calculator [53] (**Supplementary Table 1**). Thus, I²MS enables the observation of intact proteins and proteoforms, which is distinct from the data produced by NGPS.

We then isolated and fragmented discrete rhIL-6 proteoforms via I²MS². Using higher energy collision dissociation (HCD), I²MS² achieves 39% fragmentation coverage of the unmodified form of rhIL-6, with shared and distinct regions identified by NGPS (**Figures 2B and 1E**). Of the 12 specified higher mass species, 7 more abundant species showed satisfactory ion counts of the desired precursor with minimal co-isolation of other species (PFRs 4, 5, 8, 9, 11, 12, 13) using The Fisher, an in-house tool for determining I²MS² isolation windows, on the intact I²MS² spectrum. I²MS² results of these selected higher mass species are consistent with putative O-glycans with compositions of N-acetyl hexosamine, hexose, deoxyhexose, and sialic acid variably located towards the C-terminus in Helix D and the adjacent C-D loop. Localization of the O-glycans at specific serine or threonine residues along the base rhIL-6 sequence was predicted using the OGP repository [54]. Sites T6, T48, S49, T117, T147, T166, T170, T171, and T177 generated a probability for O-glycosylation greater than 0.5 and thus were used as potential O-glycan locations. Canonical rhIL-6 and select higher mass species with their respective O-glycan composition and localization are supported with fragment map P-Scores ranging from 6.6E⁻²⁵ to 2.3E⁻⁰⁴ (**Supplementary Figures 2-9**). The O-glycan localization (T147, T166, and T177 across all glycoforms) which generated the smallest P-Score for each proteoform is reported. O-glycosylation located towards the C-terminus of human IL-6 is consistent with that reported in lung adenocarcinoma cells isolated from malignant pleural effusion [51].

Both NGPS and I²MS² detect the E₅₄ALAEAENNLNL₆₃ sequence region (**Figures 1E and 2B**). Also, while NGPS provides single amino acid resolution of the D₇₀GCFQSGFNE₇₉ fragment, only a single residue cleavage was observed within this peptide by I²MS² (**Figures 1E, 2B**). Similarly, NGPS reports the sequence of the V₁₂₀LIQFLQK₁₂₇ peptide (**Figure 1E**), which also contains no residue cleavages by I²MS² (**Figure 2B**).

For the region encompassing Q₂₇IRYILDGISAL₃₈ (**Figures 1E, 2B**), NGPS and I²MS² provide complementary coverage. While NGPS enables single amino acid resolution of 11/12 amino acids (**Figure 1E**), I²MS² detects fragment ions on either side of the glycine (G) residue, which is not detected by NGPS due to the lack of a G recognizer (**Figure 1B**). While NGPS does not currently provide a C recognizer, I²MS² detects C-containing fragments C₄₃NKSNMCESSK₅₃ and C₈₂LVKIITGLLEFEVYLEYLQNR₁₀₃ (**Figure 2B**). In addition to this 22 amino acid-long internal fragment, I²MS² also provides near complete amino acid coverage across N₁₃₁LDAIITTPDPTTNASLLTK₁₅₀ (**Figure 2B**). Both NGPS and I²MS² sequence the C-terminal region of rhIL-6 encompassing L₁₅₀QAQNQW₁₅₆ (**Figures 1E, 2B**), which contains the only tryptophan (W) in IL-6 [55, 56]. Interestingly, amino acids within this fragment have been implicated in IL-6 binding interactions with receptors.

Overall, our results demonstrate NGPS and TD-MS provide sequence information for overlapping and distinct regions within rhIL-6. To place these sequencing results in the context of IL-6 tertiary structure and function, we mapped the amino acid regions detected by NGPS (**Figure 1E**) and I²MS² (**Figure 2B**) using an IL-6 crystal structure as a reference [57] (**Figure 3A**).

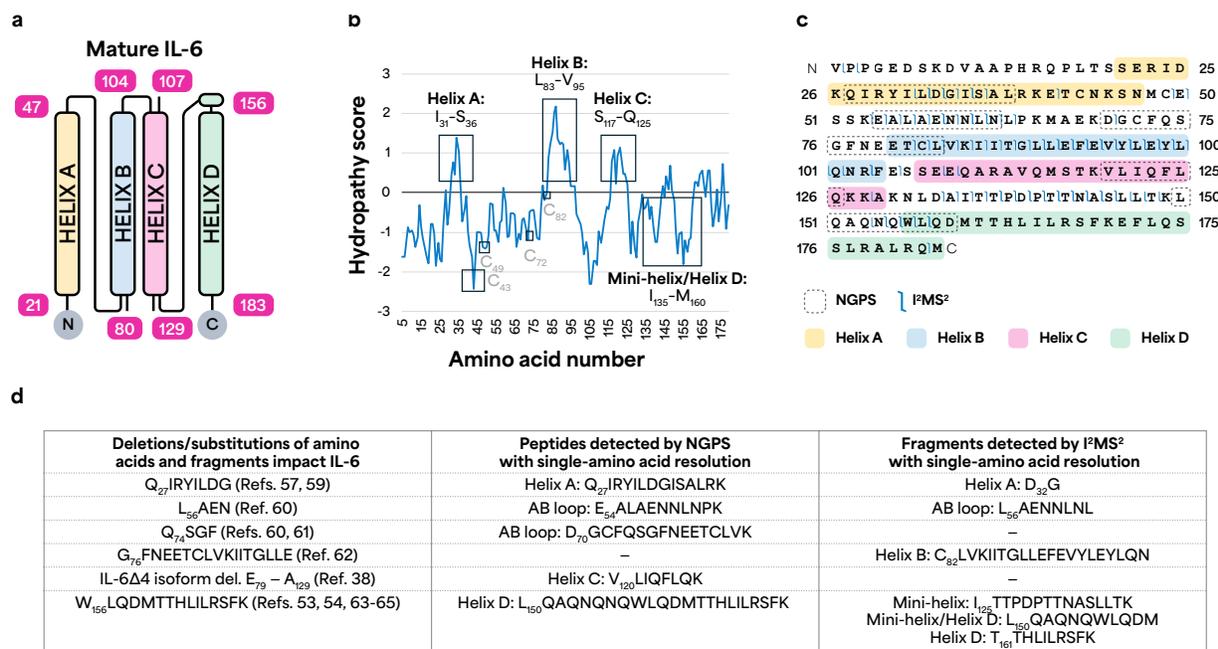


Figure 3. NGPS and I²MS/I²MS² cover hydrophobic and hydrophilic regions of IL-6 important for IL-6 interactions. (a) Mature human IL-6 schematic. Numbering is based on a reported three-dimensional structure of human IL-6. (b) Kyte-Doolittle hydropathy plot reports surface-exposed regions. Positive scores correspond to hydrophobic regions. Negative scores correspond to hydrophilic regions. Amino acid sequence position is on the x axis. Average hydropathy score is calculated for windows of 9 amino acids. (c) NGPS (Platinum) and I²MS provide broad coverage of IL-6, combining to resolve 52% of single amino acids. (d) NGPS (Platinum) and I²MS detect single amino acids and protein fragments (denoted with numbers in columns 2 and 3) reported to impact IL-6 interactions (denoted with numbers in column 1).

IL-6, a four-helix bundle cytokine that contains 183 amino acids after signal peptide cleavage [58-60], is a multi-functional protein that transmits cellular signaling via IL-6 receptor alpha (IL-6R) and beta (gp130) [61]. The molecular information for IL-6 binding and activity is enabled via adoption of a four-helix fold, with a mini-helix before Helix D [57]. Each alpha helix contains 20-30 amino acids, with long AB and CD loops that accommodate an up-up-down-down helical orientation (**Figure 3A**). Previous studies indicate that multiple segments within IL-6 topology are necessary for biological interactions [62]. Hence, broad coverage of IL-6 sequence is key for elucidating residues important for IL-6 interactions, which occur via hydrophobic and hydrophilic interactions spread across different domains.

To determine if NGPS and I²MS provide broad coverage of the helices and loop regions, we generated a Kyte-Doolittle hydropathy plot [63] (**Figure 3B**). Interestingly, NGPS and I²MS analyze broad regions of IL-6 that are hydrophobic (positive scores) and hydrophilic (negative scores) (**Figure 3B**). NGPS resolves single amino acids within peptide Q₂₇IRYILDGISAL₃₈ (**Figure 3C**), located within Helix A, which is one of the most hydrophobic regions in IL-6 (**Figure 3B**). Further, NGPS also detects V₁₂₀LIQFLQK₁₂₇ located within a hydrophobic portion of Helix C (**Figure 3B**).

The most hydrophobic portion of IL-6 encompasses C₈₂LVKIITGLLEFEVYLEYLQNR₁₀₃ within Helix B (**Figure 3B**), an internal fragment that I²MS detects (**Figure 3C**). While this fragment is nonpolar, other cysteine-containing segments lie within hydrophilic regions (**Figure 3B**). I²MS also discerns N₁₃₁LDAIITTPDPTTNASLLTK₁₅₀ located within an amphipathic region of the CD loop and the mini helix that precedes Helix D (**Figures 3A, C**). Both I²MS and NGPS cover amino acids within Helix D (**Figure 3C**), a C-terminal region that contains a stretch of amphipathic residues (**Figure 3B**). Overall, NGPS and I²MS sequence various regions of IL-6 with different amino acid compositions and properties, highlighting the complementarity of these orthogonal analytical methods (**Figure 3C**).

For sequence structure-function analysis, we performed a literature survey to determine which sequenced regions are relevant to IL-6 interactions. Removal of amino acids within the Q₂₇IRYILD₃₃ segment of Helix A reduces or abolishes IL-6 activity [62, 64] (**Figure 3D, Row 1**). Within the AB loop, E₅₄ALAEAENNLNL₆₅ contains residues important for IL-6 binding to IL-6R [65] (**Figure 3D, Row 2**). Similarly, D₇₀GCFQSGFNE₇₉ harbors residues such as phenylalanine (F), both of which are relevant for IL-6 binding interactions (**Figure 3D, Row 3**) [65-67]. We demonstrate that NGPS resolves both F residues within D₇₀GCFQSGFNE₇₉ (**Figure 1E, Figure 3C**). NGPS also detects peptide V₁₂₀LIQFLQK₁₂₇ within Helix C (**Figure 1E, Figure 3C**). Notably, a segment including V₁₂₀LIQFLQK₁₂₇ is truncated in an IL-6 splice variant with altered signaling (**Figure 3D, Row 4**) [33]. The mini-helix and amphipathic D helix contain a stretch of leucine and polar Q/N residues that influence IL-6 binding [55, 56, 68-70]. Therefore, both NGPS and I²MS detect regions of IL-6 relevant to its biological function and interactions.

Limitations of this study include the use of recombinant IL-6 and, as a proof-of-concept study, its small sample size. Future studies will aim to expand the scope of analysis for combining NGPS and I²MS² across a broader range of target proteins and proteoforms.

Conclusion

Taken together, our data demonstrate complementarity of single-molecule protein sequencing (NGPS on Platinum) and I²MS to cover key regions of IL-6. While NGPS provided single amino acid resolution for fragments like V₁₂₀LIQFLQK₁₂₇, which are

essential for IL-6 receptor binding, I²MS enabled detection of larger proteoforms, providing critical information on PTMs, such as glycosylation, that affect IL-6's stability and bioactivity. When combined, NGPS and I²MS cover many key regions of IL-6, achieving 52% combined sequence coverage of single amino acids. This sets up improvements in technology and informs future studies of how polymorphisms or mutations affect PTMs on endogenous IL-6. These complementary datasets underscore the potential of combining single-molecule protein sequencing and mass spectrometry to obtain a more comprehensive picture of IL-6 structural and functional diversity, which is vital for understanding its therapeutic potential.

Author Contributions

Conception and project design: KAS, MAC, NLK

Data acquisition: KAS, TDF, AL, TS, EF, AS

Data analysis and interpretation of results: KAS, TDF, MW, HH, TS, EF, AS, MAC, NLK

Original draft preparation: KAS and MAC

All authors reviewed the results and approved the final version of the manuscript.

Acknowledgments

We thank researchers at Quantum-Si, particularly Brian Reed and Meredith Carpenter, and members of the Northwestern Proteomics Center of Excellence (PCE) for technical assistance. Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number P41GM108569. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

1. Patterson, S.D. and R.H. Aebersold, *Proteomics: the first decade and beyond*. Nature Genetics, 2003. **33**(S3): p. 311-323.
2. Cohen, P., *Protein kinases — the major drug targets of the twenty-first century?* Nature Reviews Drug Discovery, 2002. **1**(4): p. 309-315.
3. Chen, F., et al., *A Comprehensive Analysis of Biopharmaceutical Products Listed in the FDA's Purple Book*. AAPS PharmSciTech, 2024. **25**(5).
4. Beck, L.A., et al., *Dupilumab Treatment in Adults with Moderate-to-Severe Atopic Dermatitis*. New England Journal of Medicine, 2014. **371**(2): p. 130-139.
5. Gutterman, J.U., *Cytokine therapeutics: lessons from interferon alpha*. Proceedings of the National Academy of Sciences of the United States of America, 1994. **91**(4): p. 1198-1205.
6. Anfinsen, C.B., *Principles that Govern the Folding of Protein Chains*. Science, 1973. **181**(4096): p. 223-230.
7. Dobson, C.M., *Protein folding and misfolding*. Nature, 2003. **426**(6968): p. 884-890.
8. Marshall, S.A., et al., *Rational design and engineering of therapeutic proteins*. Drug Discovery Today, 2003. **8**(5): p. 212-221.
9. Saxton, R.A., C.R. Glassman, and K.C. Garcia, *Emerging principles of cytokine pharmacology and therapeutics*. Nature reviews. Drug discovery, 2023. **22**(1): p. 21-37.
10. Drummond, D.A. and C.O. Wilke, *The evolutionary consequences of erroneous protein synthesis*. Nature reviews. Genetics, 2009. **10**(10): p. 715-724.
11. Wong, H.E., C. Huang, Jr., and Z. Zhang, *Amino acid misincorporation in recombinant proteins*. Biotechnology Advances, 2018. **36**(1): p. 168-181.
12. Smith, L.M., N.L. Kelleher, and P. Consortium for Top Down, *Proteoform: a single term describing protein complexity*. Nat Methods, 2013. **10**(3): p. 186-7.
13. Zeck, A., et al., *Low level sequence variant analysis of recombinant proteins: an optimized approach*. PloS one, 2012. **7**(7): p. e40328-e40328.
14. Gooding, J.J. and K. Gaus, *Single-Molecule Sensors: Challenges and Opportunities for Quantitative Analysis*. Angewandte Chemie International Edition, 2016. **55**(38): p. 11354-11366.
15. Smith, L.M. and N.L. Kelleher, *Proteoforms as the next proteomics currency*. Science (New York, N.Y.), 2018. **359**(6380): p. 1106-1107.
16. Carbonara, K., M. Andonovski, and J.R. Coorsen, *Proteomes Are of Proteoforms: Embracing the Complexity*. Proteomes, 2021. **9**(3): p. 38.
17. Alfaro, J.A., et al., *The emerging landscape of single-molecule protein sequencing technologies*. Nature methods, 2021. **18**(6): p. 604-617.
18. Martin-Baniandres, P., et al., *Enzyme-less nanopore detection of post-translational modifications within long polypeptides*. Nature nanotechnology, 2023. **18**(11): p. 1335-1340.
19. Swaminathan, J., et al., *Highly parallel single-molecule identification of proteins in zeptomole-scale mixtures*. Nature biotechnology, 2018: p. 10.1038/nbt.4278.

20. Reed, B.D., et al., *Real-time dynamic single-molecule protein sequencing on an integrated semiconductor device*. *Science*, 2022. **378**(6616): p. 186-192.
21. Kafader, J.O., et al., *Individual Ion Mass Spectrometry Enhances the Sensitivity and Sequence Coverage of Top-Down Mass Spectrometry*. *Journal of proteome research*, 2020. **19**(3): p. 1346-1350.
22. Kafader, J.O., et al., *Multiplexed mass spectrometry of individual ions improves measurement of proteoforms and their complexes*. *Nature methods*, 2020. **17**(4): p. 391-394.
23. Jarrold, M.F., *Single-Ion Mass Spectrometry for Heterogeneous and High Molecular Weight Samples*. *Journal of the American Chemical Society*, 2024. **146**(9): p. 5749-5758.
24. Sze, S.K., et al., *Top-down mass spectrometry of a 29-kDa protein for characterization of any posttranslational modification to within one residue*. *Proceedings of the National Academy of Sciences of the United States of America*, 2002. **99**(4): p. 1774-1779.
25. Lin, T.J., et al., *Evolution of a comprehensive, orthogonal approach to sequence variant analysis for biotherapeutics*. *mAbs*, 2019. **11**(1): p. 1-12.
26. Berraondo, P., et al., *Cytokines in clinical cancer immunotherapy*. *British journal of cancer*, 2019. **120**(1): p. 6-15.
27. Abbas, A.K., et al., *Revisiting IL-2: Biology and therapeutic prospects*. *Science Immunology*, 2018. **3**(25).
28. Vilček, J. and M. Feldmann, *Historical review: Cytokines as therapeutics and targets of therapeutics*. *Trends in Pharmacological Sciences*, 2004. **25**(4): p. 201-209.
29. Scheller, J., et al., *Exploring the landscape of synthetic <scp>IL</scp>-6-type cytokines*. *The FEBS Journal*, 2023. **291**(10): p. 2030-2050.
30. Zheng, X., et al., *The use of supercytokines, immunocytokines, engager cytokines, and other synthetic cytokines in immunotherapy*. *Cellular & molecular immunology*, 2022. **19**(2): p. 192-209.
31. Lipiäinen, T., et al., *Formulation and Stability of Cytokine Therapeutics*. *Journal of Pharmaceutical Sciences*, 2015. **104**(2): p. 307-326.
32. Savino, R., et al., *Rational design of a receptor super-antagonist of human interleukin-6*. *The EMBO journal*, 1994. **13**(24): p. 5863-5870.
33. Bihl, M.P., et al., *Identification of a Novel IL-6 Isoform Binding to the Endogenous IL-6 Receptor*. *American Journal of Respiratory Cell and Molecular Biology*, 2002. **27**(1): p. 48-56.
34. Afonina, Inna S., et al., *Proteolytic Processing of Interleukin-1 Family Cytokines: Variations on a Common Theme*. *Immunity*, 2015. **42**(6): p. 991-1004.
35. Santhanam, U., et al., *Post-translational modifications of human interleukin-6*. *Archives of Biochemistry and Biophysics*, 1989. **274**(1): p. 161-170.
36. Vanheule, V., et al., *How post-translational modifications influence the biological activity of chemokines*. *Cytokine*, 2018. **109**: p. 29-51.

37. Höglund, M., *Glycosylated and non-glycosylated recombinant human granulocyte colony-stimulating factor (rhG-CSF)—what is the difference?* Medical Oncology, 1998. **15**(4): p. 229-233.
38. Ebrahimi, S.B. and D. Samanta, *Engineering protein-based therapeutics through structural and chemical design*. Nature communications, 2023. **14**(1): p. 2411-2411.
39. Assenberg, R., et al., *Advances in recombinant protein expression for use in pharmaceutical research*. Current Opinion in Structural Biology, 2013. **23**(3): p. 393-402.
40. Jekel, P.A., W.J. Weijer, and J.J. Beintema, *Use of endoproteinase Lys-C from Lysobacter enzymogenes in protein sequence analysis*. Analytical Biochemistry, 1983. **134**(2): p. 347-354.
41. Gironde-Martínez, A., et al., *DNA-Compatible Diazo-Transfer Reaction in Aqueous Media Suitable for DNA-Encoded Chemical Library Synthesis*. Organic Letters, 2019. **21**(23): p. 9555-9558.
42. Gordon, C.G., et al., *Reactivity of biarylazacyclooctynones in copper-free click chemistry*. Journal of the American Chemical Society, 2012. **134**(22): p. 9199-9208.
43. Sittipongpittaya, N., et al., *Protein sequencing with single amino acid resolution discerns peptides that discriminate tropomyosin proteoforms*. 2024, Cold Spring Harbor Laboratory.
44. Käll, L., et al., *Assigning Significance to Peptides Identified by Tandem Mass Spectrometry Using Decoy Databases*. Journal of Proteome Research, 2007. **7**(1): p. 29-34.
45. Park, H.M., et al., *Novel Interface for High-Throughput Analysis of Biotherapeutics by Electrospray Mass Spectrometry*. Anal Chem, 2020. **92**(2): p. 2186-2193.
46. McGee, J.P., et al., *Automated Control of Injection Times for Unattended Acquisition of Multiplexed Individual Ion Mass Spectra*. Anal Chem, 2022. **94**(48): p. 16543-16548.
47. McGee, J.P., et al., *Automated imaging and identification of proteoforms directly from ovarian cancer tissue*. Nat Commun, 2023. **14**(1): p. 6478.
48. Kafader, J.O., et al., *STORI Plots Enable Accurate Tracking of Individual Ion Signals*. Journal of the American Society for Mass Spectrometry, 2019. **30**(11): p. 2200-2203.
49. Senko, M.W., S.C. Beu, and F.W. McLafferty, *Determination of monoisotopic masses and ion populations for large biomolecules from resolved isotopic distributions*. J Am Soc Mass Spectrom, 1995. **6**(4): p. 229-33.
50. Rockwood, A.L. and P. Haimi, *Efficient calculation of accurate masses of isotopic peaks*. J Am Soc Mass Spectrom, 2006. **17**(3): p. 415-9.
51. Hung, C.H., et al., *Defective N-glycosylation of IL6 induces metastasis and tyrosine kinase inhibitor resistance in lung cancer*. Nat Commun, 2024. **15**(1): p. 7885.
52. York, W.S., et al., *GlyGen: Computational and Informatics Resources for Glycoscience*. Glycobiology, 2020. **30**(2): p. 72-73.
53. De Leoz, M.L. and S. Stein. *NIST Glyco Mass Calculator*. Available from: <https://www.nist.gov/static/glyco-mass-calc/#/>.

54. Huang, J., et al., *OGP: A Repository of Experimentally Characterized O-glycoproteins to Facilitate Studies on O-glycosylation*. Genomics Proteomics Bioinformatics, 2021. **19**(4): p. 611-618.
55. Nishimura, C., et al., *Chemical modification and ¹H-NMR studies on the receptor-binding region of human interleukin 6*. European Journal of Biochemistry, 1991. **196**(2): p. 377-384.
56. Simpson, R.J., et al., *Interleukin-6: structure-function relationships*. Protein science : a publication of the Protein Society, 1997. **6**(5): p. 929-955.
57. Somers, W., M. Stahl, and J.S. Seehra, *1.9 A crystal structure of interleukin 6: implications for a novel mode of receptor dimerization and signaling*. The EMBO journal, 1997. **16**(5): p. 989-997.
58. Widjaja, A.A., S.P. Chothani, and S.A. Cook, *Different roles of interleukin 6 and interleukin 11 in the liver: implications for therapy*. Human vaccines & immunotherapeutics, 2020. **16**(10): p. 2357-2362.
59. Schumertl, T., et al., *Function and proteolytic generation of the soluble interleukin-6 receptor in health and disease*. Biochim Biophys Acta Mol Cell Res, 2022. **1869**(1): p. 119143.
60. Widjaja, A.A., S.P. Chothani, and S.A. Cook, *Different roles of interleukin 6 and interleukin 11 in the liver: implications for therapy*. Hum Vaccin Immunother, 2020. **16**(10): p. 2357-2362.
61. Taga, T. and T. Kishimoto, *gp130 AND THE INTERLEUKIN-6 FAMILY OF CYTOKINES*. Annual Review of Immunology, 1997. **15**(1): p. 797-819.
62. Snouwaert, J.N., K. Kariya, and D.M. Fowlkes, *Effects of site-specific mutations on biologic activities of recombinant human IL-6*. The Journal of Immunology, 1991. **146**(2): p. 585-591.
63. Kyte, J. and R.F. Doolittle, *A simple method for displaying the hydropathic character of a protein*. Journal of Molecular Biology, 1982. **157**(1): p. 105-132.
64. Brakenhoff, J.P., M. Hart, and L.A. Aarden, *Analysis of human IL-6 mutants expressed in Escherichia coli. Biologic activities are not affected by deletion of amino acids 1-28*. The Journal of Immunology, 1989. **143**(4): p. 1175-1182.
65. Toniatti, C., et al., *Engineering human interleukin-6 to obtain variants with strongly enhanced bioactivity*. The EMBO Journal, 1996. **15**(11): p. 2726-2737.
66. Ehlers, M., et al., *Identification of Single Amino Acid Residues of Human IL-6 Involved in Receptor Binding and Signal Initiation*. Journal of Interferon & Cytokine Research, 1996. **16**(8): p. 569-576.
67. Kalai, M., et al., *Analysis of the Human Interleukin-6/Human Interleukin-6 Receptor Binding Interface at the Amino Acid Level: Proposed Mechanism of Interaction*. Blood, 1997. **89**(4): p. 1319-1333.
68. Nishimura, C., et al., *Role of leucine residues in the C-terminal region of human interleukin-6 in the biological activity*. FEBS Letters, 1992. **311**(3): p. 271-275.
69. Nishimura, C., et al., *Site-specific mutagenesis of human interleukin-6 and its biological activity*. FEBS Letters, 1991. **281**(1-2): p. 167-169.

70. Robins, L.I., et al., *Modifications of IL-6 by Hypochlorous Acids: Effects on Receptor Binding*. ACS omega, 2021. **6**(51): p. 35593-35599.

Supplementary Information

Next-Generation Protein Sequencing and individual ion mass spectrometry enable complementary analysis of interleukin-6

Kenneth A. Skinner¹; Troy D. Fisher²; Andrew Lee³; Taojunfeng Su³; Eleonora Forte^{2,4}; Aniel Sanchez²; Michael A. Caldwell^{2,5}; Neil L. Kelleher^{2,3,6#}

¹Quantum-Si, Incorporated, Branford, Connecticut 06405

²Proteomics Center of Excellence, Northwestern University, Evanston, Illinois 60208

³Departments of Molecular Biosciences, Chemistry and Chemical and Biological Engineering, Northwestern University, Evanston, IL, USA

⁴Department of Surgery, Feinberg School of Medicine, Comprehensive Transplant Center, Northwestern University, Chicago, Illinois 60611, United States

⁵Department of Medicine, Division of Hematology Oncology, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

⁶Department of Biochemistry and Molecular Genetics, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA

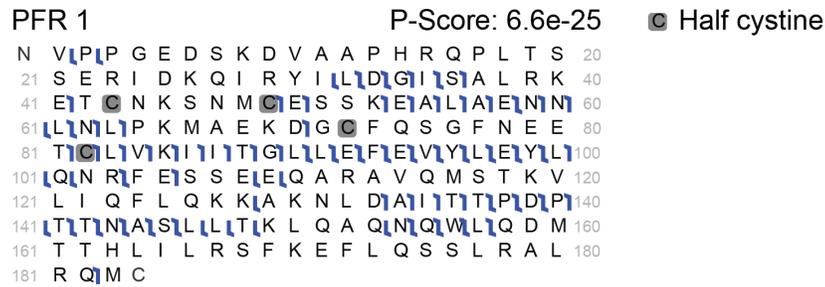
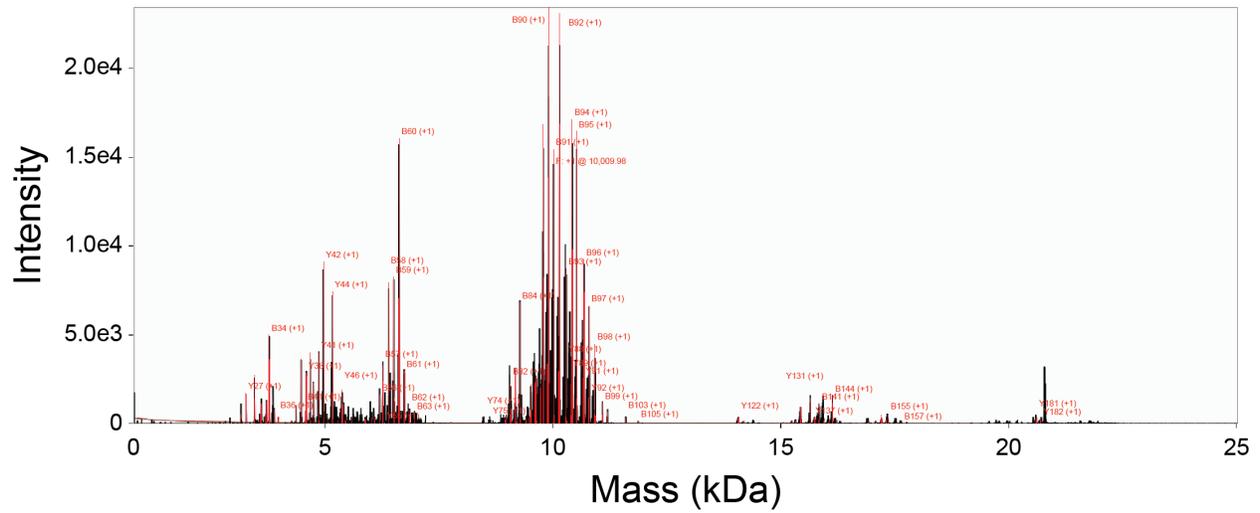
#Corresponding author email: n-kelleher@northwestern.edu

Rank	Inferred Protein	Score	Likelihood	Mass (kDa)	Length	Digested Peptides	Inferred Peptides
1	spIP05231IIL6_HUMAN	9.222594	99.99%	24	212	5	5
2	spIQ6NXT1IANR54_HUMAN	0.257477	22.7%	32	300	1	1

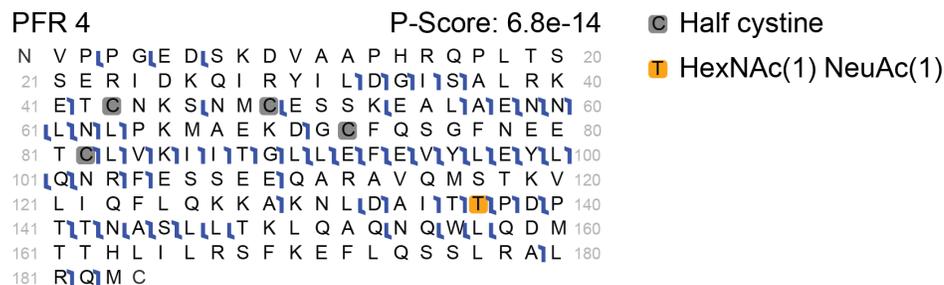
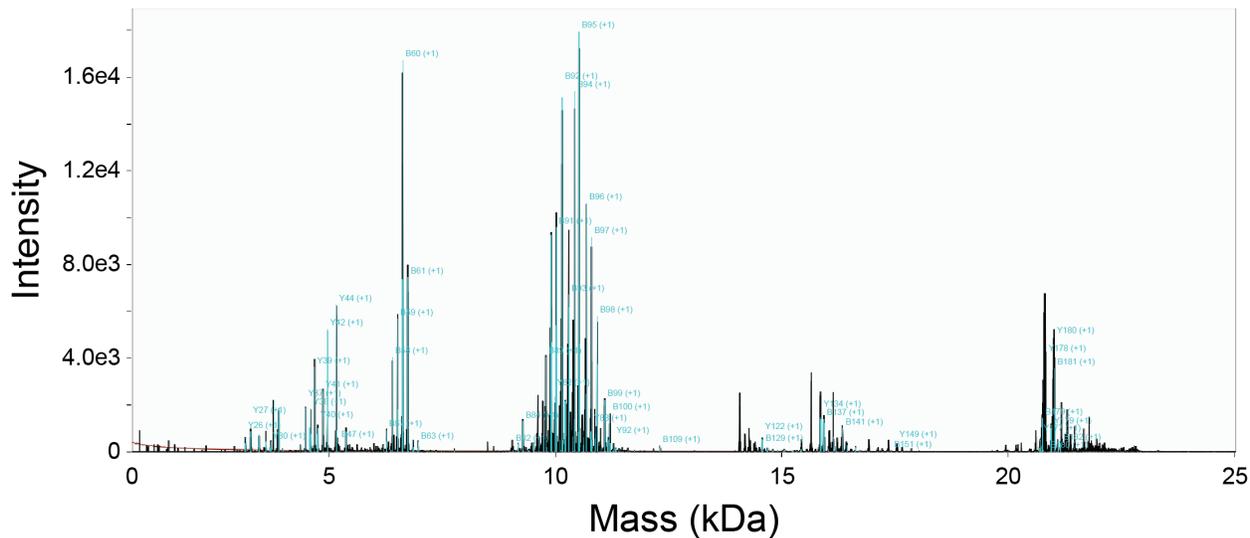
Supplementary Figure 1. Protein inference v2.5.2 aligns observed kinetic signatures from sequencing data of rhIL-6 to a database of predicted kinetic signatures. Protein Score is an estimate of the likelihood of protein identity. While ankyrin repeat domain-containing protein 54 (UniProt ID: Q6NXT1) contains peptide QIIHMLREYLERLGQHEQRERLDDLCTRLQMTSTK, which is similar in sequence to QIRYILDGISALRK in rhIL-6, Platinum unambiguously identifies rhIL-6.

Supplementary Table 1. Intact Proteoforms of rhIL-6 detected with I²MS. 13 proteoforms of rhIL-6 were detected by I²MS and abundant species were targeted for tandem-MS with I²MS². Based on the intact masses, putative glycan compositions were proposed and supported with I²MS² data when available.

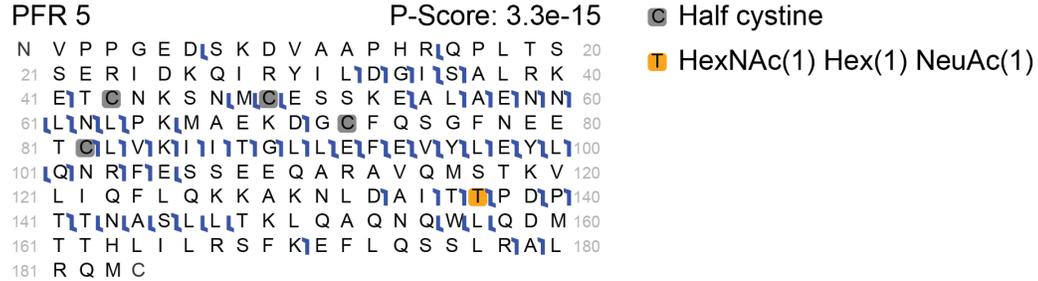
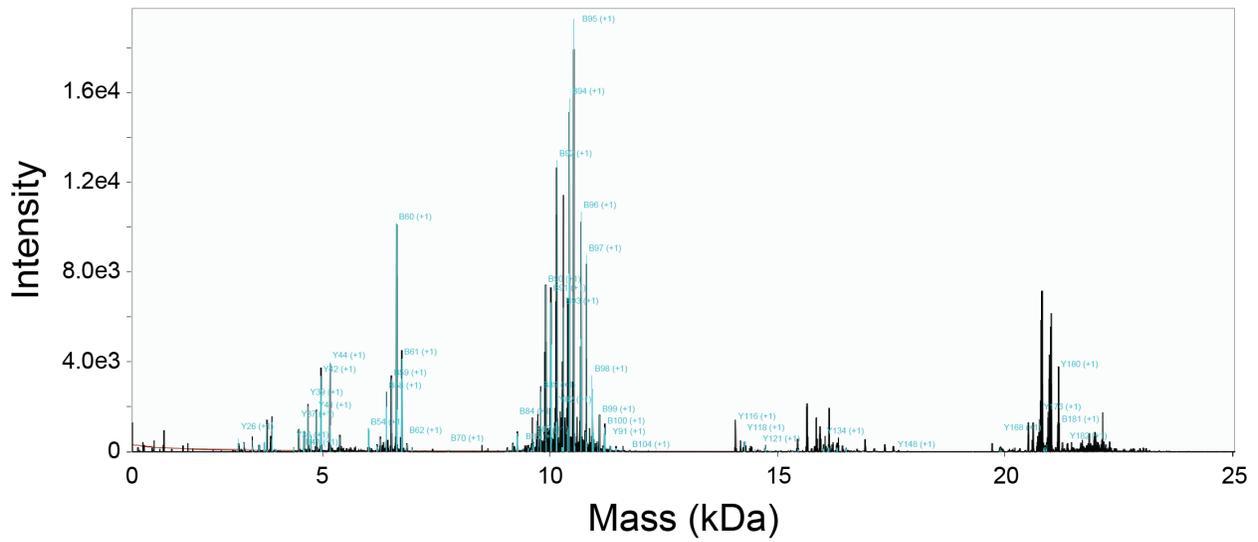
PFR	Theoretical Monoisotopic Mass (Da)	Observed Monoisotopic Mass (Da)	Mass Error (ppm)	Half Cystine Sites	Putative O-Glycan Composition	Glycan Chemical Formula (unbound)	Monoisotopic Glycan Mass (Da)	I ² MS ² ?
1	20795.58	20795.55	-1.55	C72, C78, C101, C111	N/A	N/A	N/A	Yes
2	20998.66	20998.64	-1.03	C72, C78, C101, C111	HexNAc(1)	C8H13NO5	+221.08994	No
3	21160.71	21160.69	-1.16	C72, C78, C101, C111	HexNAc(1) Hex(1)	C14H23NO10	+383.14276	No
4	21289.76	21289.72	-1.74	C72, C78, C101, C111	HexNAc(1) NeuAc(1)	C19H30N2O13	+512.18535	Yes
5	21452.17	21451.79	-17.71	C72, C78, C101, C111	HexNAc(1) Hex(1) NeuAc(1)	C25H40N2O18	+674.23818	Yes
6	21646.87	21646.84	-1.52	C72, C78, C101, C111	HexNAc(1) Hex(4)	C32H53NO25	+869.30123	No
7	21654.89	21654.86	-1.35	C72, C78, C101, C111	HexNAc(2) Hex(1) NeuAc(1)	C33H53N3O23	+877.31755	No
8	21742.91	21742.90	-0.25	C72, C78, C101, C111	HexNAc(1) Hex(1) NeuAc(2)	C36H57N3O26	+965.33359	Yes
9	21785.95	21785.89	-2.64	C72, C78, C101, C111	HexNAc(4) Hex(1)	C38H62N4O25	+992.38088	Yes
10	21858.99	21858.91	-3.62	C72, C78, C101, C111	HexNAc(3) dHex(2) Hex(1)	C42H69N3O28	+1081.41732	No
11	21905.98	21905.95	-1.30	C72, C78, C101, C111	HexNAc(1) dHex(2) Hex(2) NeuAc(1)	C43H70N2O31	+1128.40682	Yes
12	21947.01	21946.97	-1.60	C72, C78, C101, C111	HexNAc(2) dHex(2) Hex(1) NeuAc(1)	C45H73N3O31	+1169.43337	Yes
13	22109.06	22109.03	-1.26	C72, C78, C101, C111	HexNAc(2) dHex(2) Hex(2) NeuAc(1)	C51H83N3O36	+1331.48619	Yes



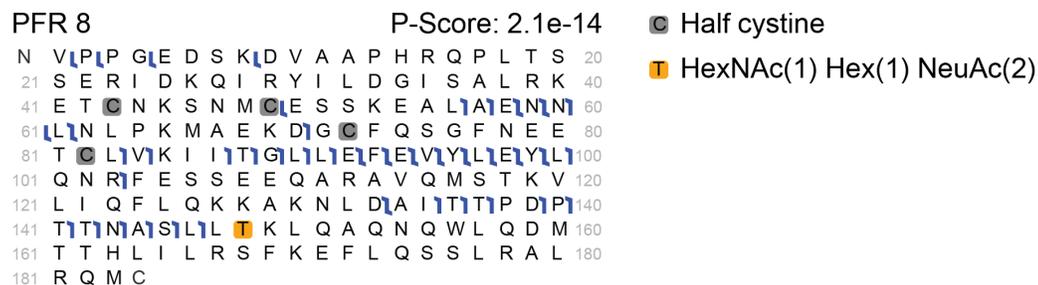
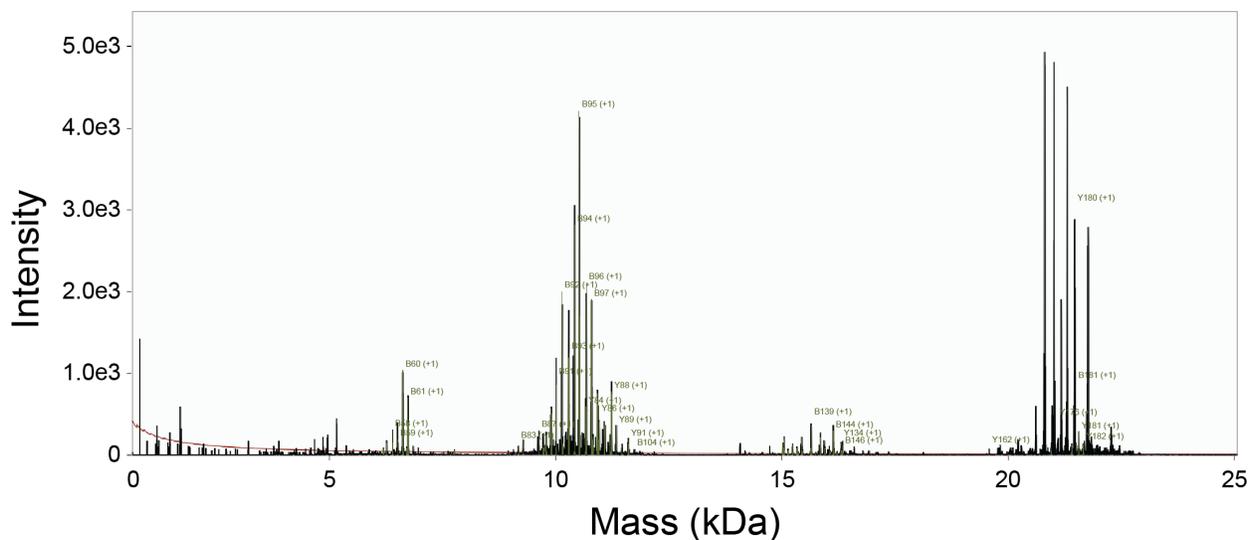
Supplementary Figure 2. Fragmentation spectrum and graphical fragment map for rhIL-6 proteoform 1.



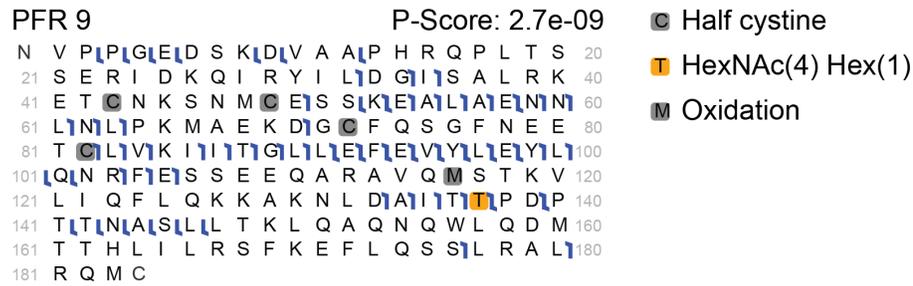
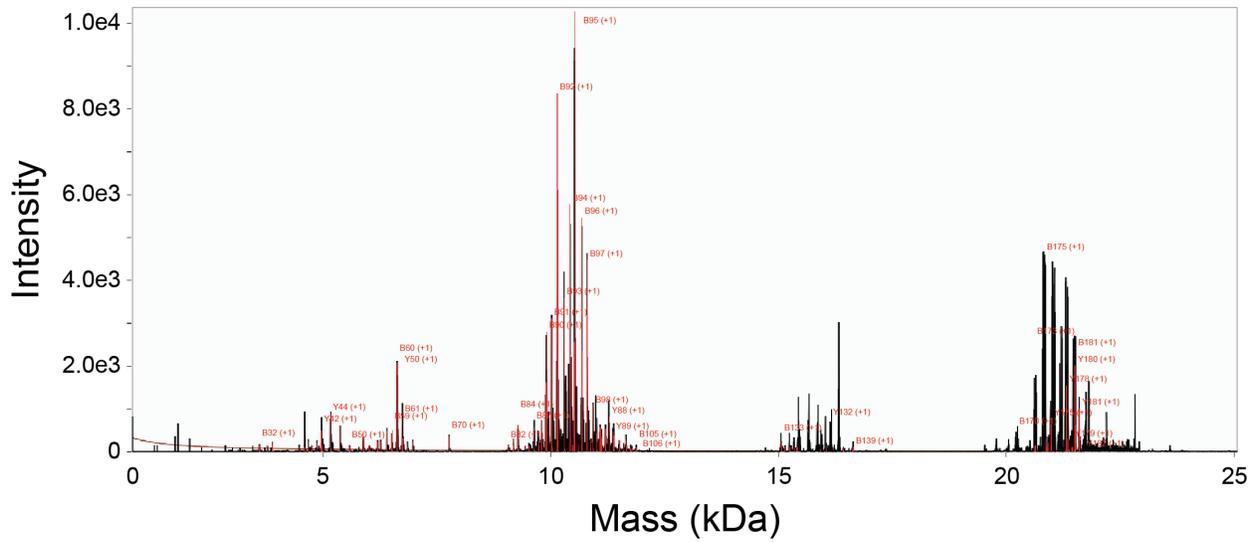
Supplementary Figure 3. Fragmentation spectrum and graphical fragment map for rhIL-6 proteoform 4.



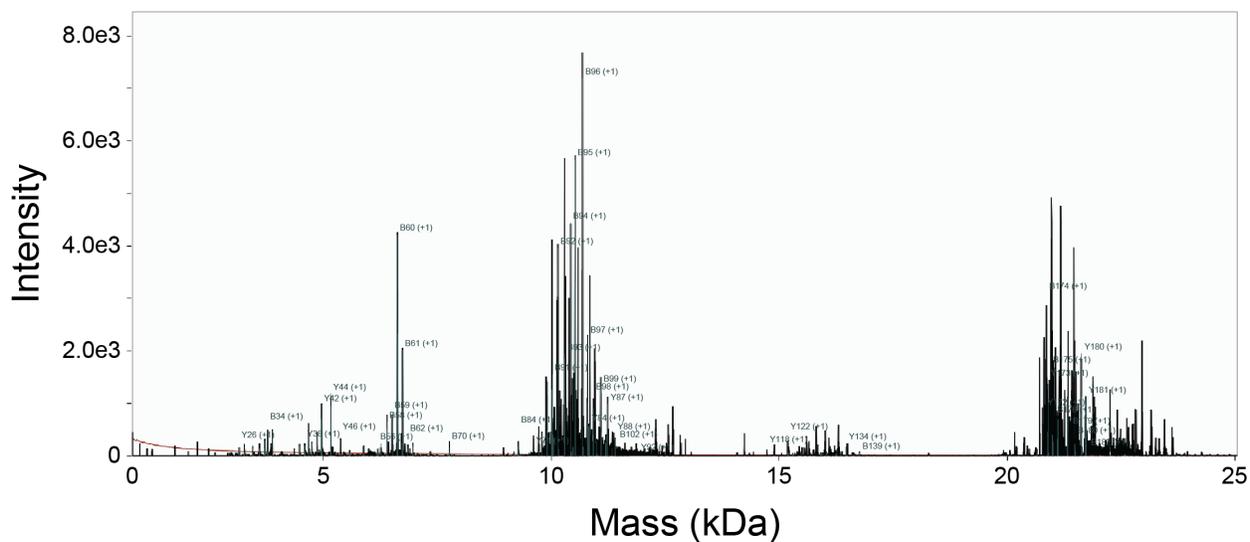
Supplementary Figure 4. Fragmentation spectrum and graphical fragment map for rhIL-6 proteoform 5.



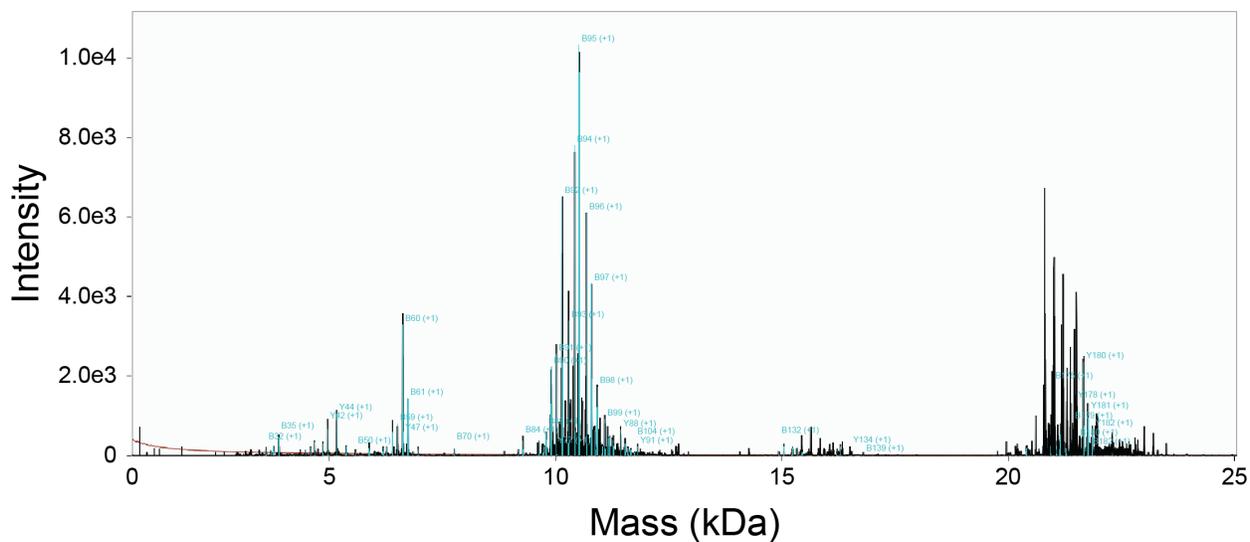
Supplementary Figure 5. Fragmentation spectrum and graphical fragment map for rhIL-6 proteoform 8.



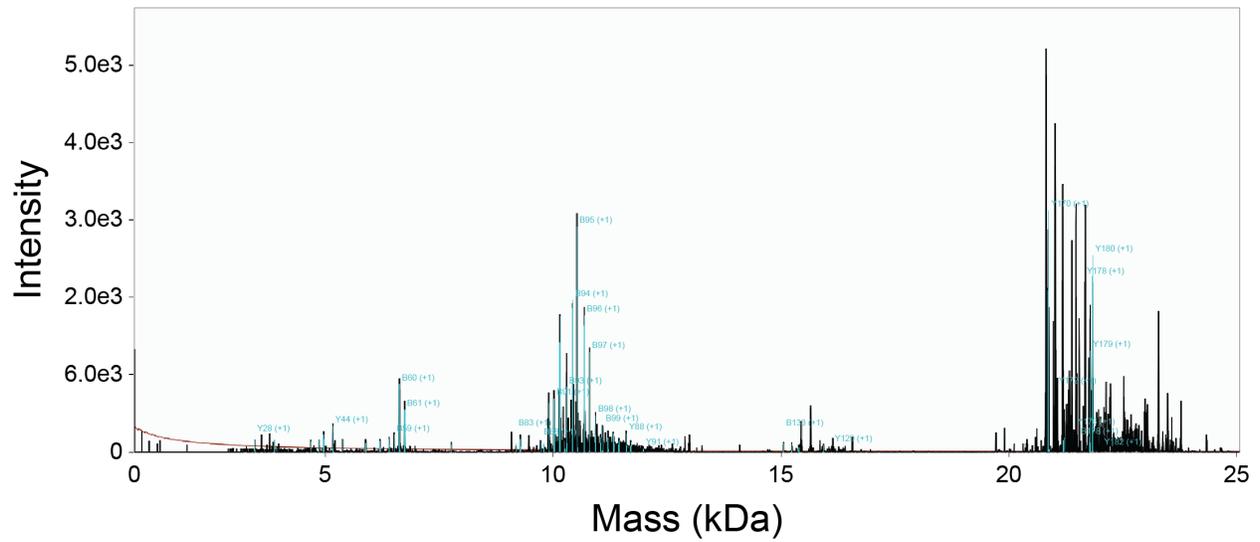
Supplementary Figure 6. Fragmentation spectrum and graphical fragment map for rhIL-6 proteoform 9.



Supplementary Figure 7. Fragmentation spectrum and graphical fragment map for rhIL-6 proteoform 11.



Supplementary Figure 8. Fragmentation spectrum and graphical fragment map for rhIL-6 proteoform 12.



Supplementary Figure 9. Fragmentation spectrum and graphical fragment map for rhIL-6 proteoform 13.