Next-generation protein sequencing integrates with top-down MS to distinguish native and artifactual proteoforms

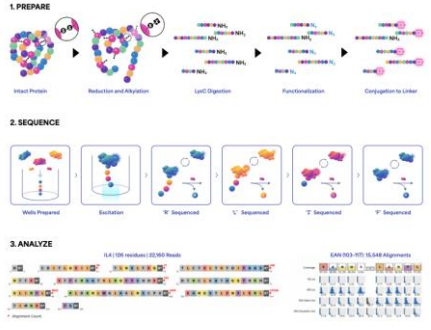
Andrew Lee[1*]; Kenneth Skinner[2*]; Taojunfeng Su[1]; Michael Warchol[2]; Aniel Sanchez[1]; Haidong Huang[2]; Troy D Fisher[1]; Eleonora Forte[1]; Michael A Caldwell[1]; Neil L Kelleher[1]
[1]Northwestern University, Evanston, IL; [2]Quantum-Si Inc., Branford, CT

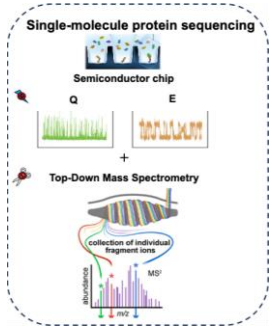
Introduction

Understanding proteoform dynamics is essential to interrogate and understand the roles of proteins in healthy and diseased states. The posttranslational modification pyroglutamate (pE) plays a role in proteolytic stability and bioactivity and is potentially involved in neurodegenerative processes. Exposed glutamine (Q) and glutamate (E) residues can undergo intramolecular cyclization to form pE under several conditions, including: 1) biological posttranslational modification; 2) spontaneous chemical reaction in solution (e.g., during preparation or storage of protein digests); 3) during ESI and CID in MS. Measurement of biologically generated pE thus faces the challenge posed by pE generated as artifacts during sample handling and MS experimentation. Top-down MS and Next-generation protein sequencing (NGPS) via Platinum can detect Q/E/pE-terminated peptides and provide an approach to distinguish biological pE modifications from analysis artifacts. Here, we show that experimental parameters of top-down MS can limit pE formation (NCE in MS2) in the analysis of intact proteins and that NGPS via Platinum can differentiate peptides by their N-terminal residues (Q, pE, -pE).

Novel Aspect

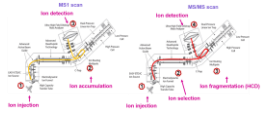
Our findings highlight that under high-energy collision conditions, the abundance of peptides with N-terminal Q can be underestimated using bottom-up MS/MS analysis, complicating protein quantitation. This problem is mitigated through top-down MS and the Platinum® protein sequencing workflow. Complementary NGPS via Platinum® and top-down MS accurately detect and quantify peptides and proteoforms that are otherwise underestimated due to unexpected artifacts and modifications.



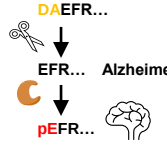


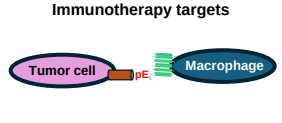
Platinum™ benchtop protein sequencer

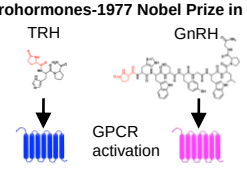
MS1 and MS2 with Orbitrap Fusion Lumos

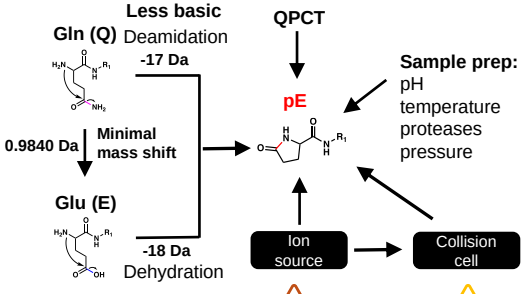


Pyroglutamate (pE): a PTM, drug target, neurohormone, and instrument-induced artifact







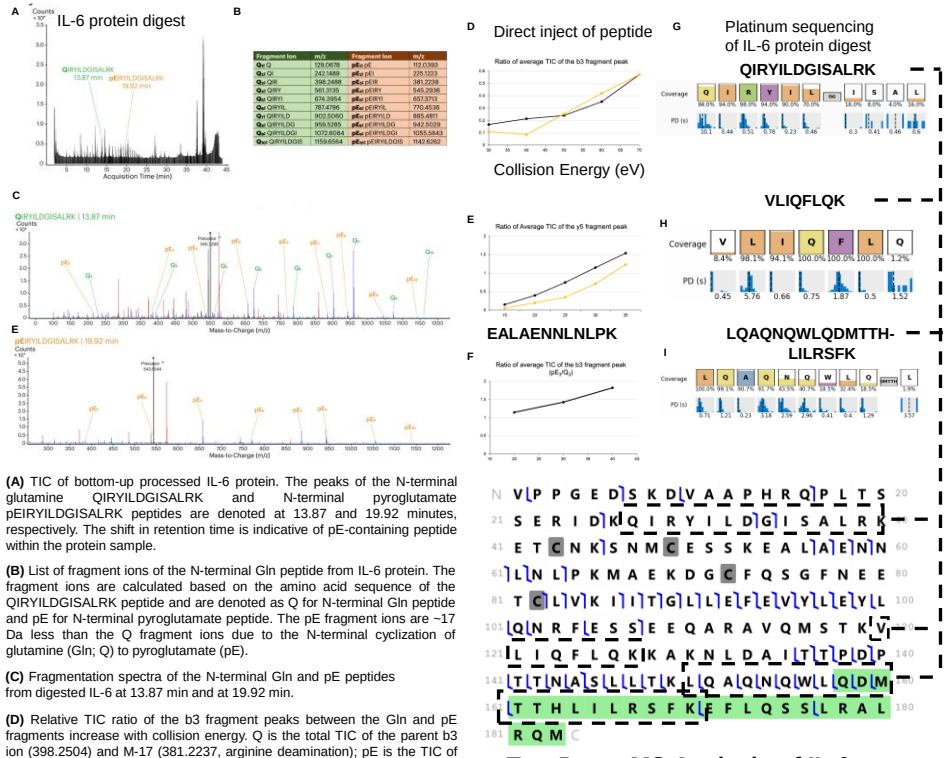


In situ formation of pE under MS conditions

Mobile proton model:

- Protons are initially localized on the most basic sites (N-terminus & K/R residues).
- After ion activation, ionizing protons can be transferred to various peptide linkages → heterogenous populations of protonated forms.
- Certain peptide ions don't fragment preferentially at the peptide bond linking the 2nd & 3rd residues from N-terminus (b_2/y_{n-2}).
- Loss of NH_3/H_2O from N-terminal Q/E → decreased basicity of N-terminus and unexpected fragmentation patterns (ie closer to center of peptide) that can hamper database matching.
- N-terminal conversion reduces proton affinity → 2nd proton on peptide ion is no longer localized at this site → more mobile and effect dissociation at various peptide bonds.

Amino acid recognizers detect Gln-containing IL-6 peptide that escapes LC-MS/MS database matching



(A) TIC of bottom-up processed IL-6 protein. The peaks of the N-terminal glutamine QIRYILDGISALRK and N-terminal pyroglutamate pEIRYILDGISALRK peptides are denoted at 13.87 and 19.92 minutes, respectively. The shift in retention time is indicative of pE-containing peptide within the protein sample.

(B) List of fragment ions of the N-terminal Gln peptide from IL-6 protein. The fragment ions are calculated based on the amino acid sequence of the QIRYILDGISALRK peptide and are denoted as Q for N-terminal Gln peptide and pE for N-terminal pyroglutamate peptide. The pE fragment ions are ~17 Da less than the Q fragment ions due to the N-terminal cyclization of glutamine (Gln; Q) to pyroglutamate (pE).

(C) Fragmentation spectra of the N-terminal Gln and pE peptides from digested IL-6 at 13.87 min and at 19.92 min.

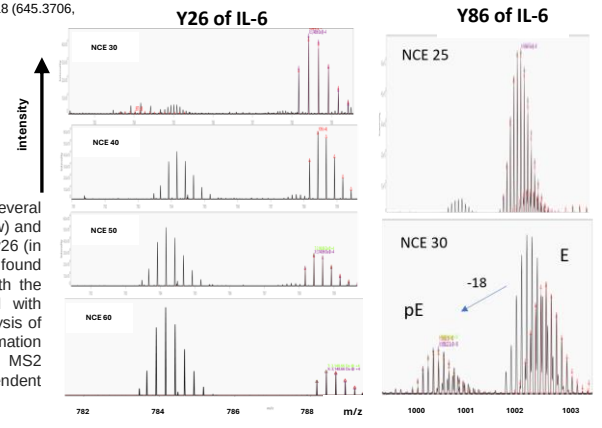
(D) Relative TIC ratio of the b3 fragment peaks between the Gln and pE fragments increase with collision energy. Q is the total TIC of the parent b3 ion (398.2504) and M-17 (381.2237, arginine deamination); pE is the TIC of the M-34 peak (364.1974, arginine deamination, pyroglutamate); dhQ is the TIC of the M-35 (363.2131, arginine deamination, glutamine dehydration).

(E) Q is the TIC of the parent y5 ion (663.3813); pE is the TIC of the M-17 peak (646.3540, pyroglutamate); dhQ is the TIC of the M-18 (645.3706, glutamine dehydration).

(F) Q is the TIC of the parent b3 ion (385.2076); pE is the TIC of the M-17 peak (367.1972, pyroglutamate).

(G-I) Sequencing traces of VLIQFLQK, QIRYILDGISALRK, and LQAQNQWLQDMTTHLILRSFK peptides of IL-6 protein digests on Platinum.

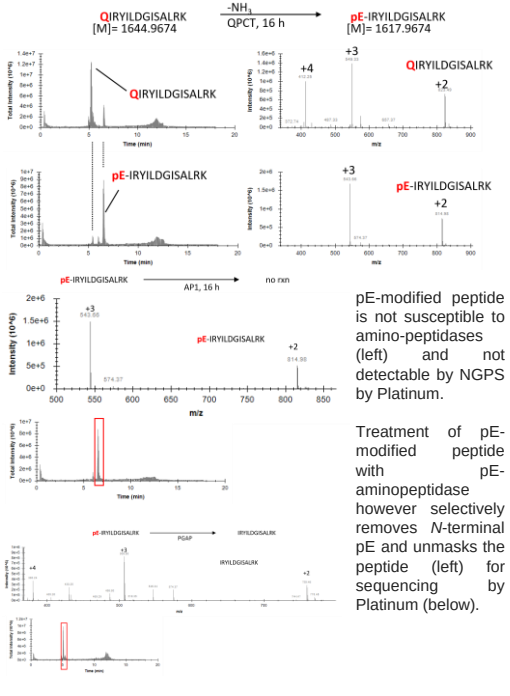
Top-Down MS Analysis of IL-6 Observes pE Formation in MS2



Right: The TD-MS analysis of IL-6 detects several daughter ions in MS2 fragmentation (shown below) and provides broad coverage of IL-6. Analysis of the y26 (in green), one of several ions with *N*-terminal Q, found NCE-dependent generation of daughter ions with the loss of NH_3 or H_2O. These ions correspond with formation of pE or pD residues (right). MS3 analysis of y26 furthermore confirmed the spontaneous formation of the pE-terminated daughter ion in MS2. MS2 analysis of y86 also shows the NCE-dependent cyclization of *N*-terminal E to pE.

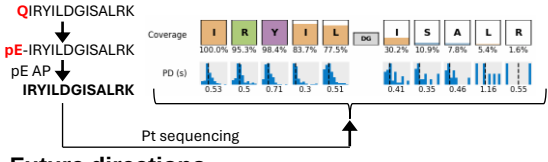
Assay development with pE-aminopeptidase for Platinum sequencing

Spontaneous and Enzymatic Conversion of QIRYILDGISALRK to pE-IRYILDGISALRK



pE-modified peptide is not susceptible to amino-peptidases (left) and not detectable by NGPS by Platinum.

Treatment of pE-modified peptide with pE-aminopeptidase however selectively removes *N*-terminal pE and unmasks the peptide (left) for sequencing by Platinum (below).



Future directions

Analysis of IL-6 from biological samples for detection of proteoforms
Analysis of pE proteoforms (e.g., β-amyloid)

References

Bayer, Thomas A. *Molecular Psychiatry* 27.4 (2022): 1880-1885.
Boyd, Robert, and Árpád Somogyi. *Journal of the American Society for Mass Spectrometry* 21.8 (2010): 1275-1278.
Godugu, Bhaskar, *et al. Journal of the American Society for Mass Spectrometry* 21 (2010): 1169-1176.
Kafader, Jared O., *et al. Journal of proteome research* 19.3 (2020): 1346-1350.
Liu, Zhihua, *et al. Journal of pharmaceutical sciences* 108.10 (2019): 3194-3200.
Neta, Pedatsur, *et al. Journal of the American Society for Mass Spectrometry* 18 (2007): 27-36.
Reed, Brian D., *et al. Science* 378.6616 (2022): 186-192.

Conflicts of Interest disclosure

K.S., M.W., H.H. have financial interest in Quantum-Si Inc.
N.L.K is involved in entrepreneurial activities in top-down proteomics and consults for Thermo Fisher Scientific.