Definitive Protein Sequencing by Top-Down Mass Spectrometry Using timsOmni and OmniScape

OmniScape enables *de novo* protein sequence reconstruction from top-down data and full proteoform validation using multiple fragmentation methods.

Abstract

Peptide centric bottom-up proteomics can obscure proteoform level information due to reliance on protein inference. Here, a bottom-up-derived protein sequence was re evaluated using top-down mass spectrometry on a timsOmni™ platform. OmniScape™ workflows enabled systematic comparison of large numbers of candidate sequence variants, leading to sequence correction and validation with 100% sequence coverage.

Keywords:
Top-down, timsOmni, OmniScape, *de novo* sequencing, protein sequencing, proteoform analysis, ECD, clipping analysis, PTM screening, sequence validation

Introduction

Bottom-up (BU) proteomics remains the dominant approach for protein identification. However, its peptide centric measurements rely on protein inference, which can lead to incomplete or incorrect information in the proteoform context. In this study, a protein sequence originally inferred from bottom-up data was re examined using top-down (TD) mass spectrometry on a timsOmni mass spectrometer. Comprehensive analysis was carried out with the OmniScape software workflows including *de novo* sequencing, clipping analysis and targeted substitution screening. This resulted in the correction of the originally proposed sequence and 100% sequence coverage confirmed with multiple fragmentation methods.

Dalila Bensaddek[1], Miriam Escarlet Diaz Galicia[2], Magnus Rueping[2], Dominik Renn[2], Mariangela Kosmopoulou[3], Athanasios Smyrnakis[3], George Alevizos[3], Dimitris Papanastasiou[3], Maan Amad[1]; [1]Analytical Chemistry Core Lab, KAUST, Thuwal, Saudi Arabia; [2]KAUST Catalysis Center (KCC), Division of Physical Sciences & Engineering, KAUST, Thuwal, Saudi Arabia; [3]Fasmatech Science & Technology, Chalandri, Greece.

Experimental

Sample preparation and data acquisition

A ~9 kDa protein sample was dialyzed against water, lyophilized and reconstituted at 4 µM in 50:49.5:0.5 H_2O:ACN:FA. The sample was analyzed offline on a timsOmni system using an Apollo II source operated at 3 µL/min. MS² experiments were performed on the isolated 8+ charge state using multiple fragmentation modes, including isCID, $_C$CID, $_R$CID, ECD, ECciD and EID.

OmniScape workflows

Data analysis followed a stepwise approach, progressing from confirmation of the originally proposed BU derived sequence through iterative sequence revision:

1. Initial confirmation of the BU derived sequence using MS² ECD data
2. *De novo* sequencing to identify novel sequence elements
3. Clipping analysis to iteratively correct the C-terminal region
4. Targeted substitution and PTM screening to validate ambiguous regions

This structured approach enabled systematic refinement of the protein sequence and progressive increases in sequence coverage.

Results

1 Discrepancy Between BU and TD Measurements

The TD MS measured mass was higher than the BU predicted mass, indicating a sequence inconsistency. Application of the confirmation workflow to MS² ECD data resulted only in random fragment matches when mapped to the BU derived sequence (Figure 1), supporting the hypothesis that the original N- and C-terminal parts of the sequence were incorrect.

2 *De Novo* Sequencing

De novo analysis of MS² ECD data produced a top scored sequence tag that included a His tag connected to the N-terminal region of the protein (Figure 2). The first confidently identified fragment exhibited a mass shift corresponding to MHH amino acids. Confirmation of this first sequence revision is shown in Figure 3. This revision provided full N-terminal sequence coverage up to residue O45, with limited coverage at the C-terminus.

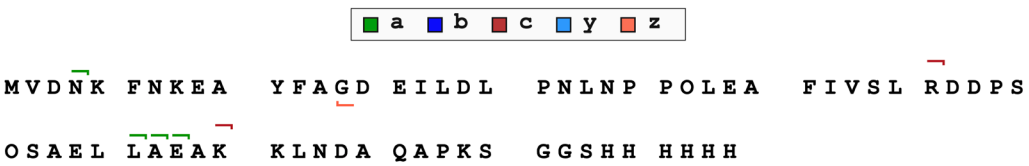

Figure 1. MS² ECD sequence map matched to the bottom-up derived sequence.

150	Sequence Tag	Terminus	Score
1	HHHHSSGREN[I/L]YFQGSVDNKFNKEAYFAGDE[I/L][I/L]D	N	29.85

Figure 2. Top-scoring *de novo* sequence tag including a His tag (red) connected to the protein N-terminus (green) linked by unexpected residues.

Figure 3. MS² ECD sequence map after N-terminal correction.

3 Clipping Analysis and C Terminal Correction

Despite improved coverage, the theoretical neutral mass of the first revision remained higher than the TD MS measured mass. *De novo* analysis of isCID data identified sequence tags near the C terminus consistent with the BU derived sequence, but systematically lower y ion masses indicated a shorter sequence than the proposed C-terminal region. Iterative clipping analysis was therefore applied to trim the C-terminal region for optimal coverage. The second sequence revision yielded strong coverage near both termini (Figure 4), although a coverage gap between residues O45 and O59 remained.

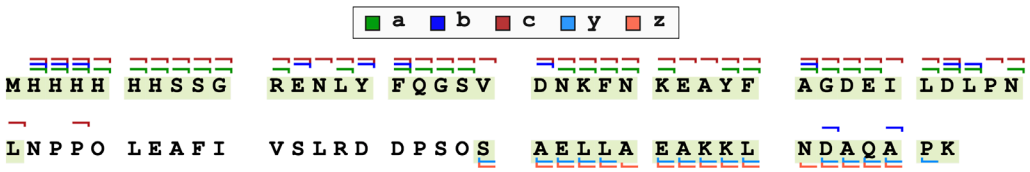

Figure 4. MS² ECD sequence map after C-Terminal correction.

4 Targeted Substitution and Final Confirmation

Targeted amino acid substitution analysis based on MS² ECD data was performed for residues O45 and O59, which corresponded to the region of missing sequence coverage after the prior corrections. This analysis revealed substitution of both residues with Q (Figures 5 and 6). The resulting third sequence revision achieved near-complete coverage. Further validation of the revised N-terminal region was conducted using single-site substitution screening across residues M1–V20 via the PTM Screening workflow, where the *de novo*–derived amino acid series provided the highest sequence and intensity coverage for MS² ECD data. Final confirmation using MS² CID, EID, and ECciD data further validated the revised sequence, with MS² ECciD achieving 100% sequence coverage (Figure 7).

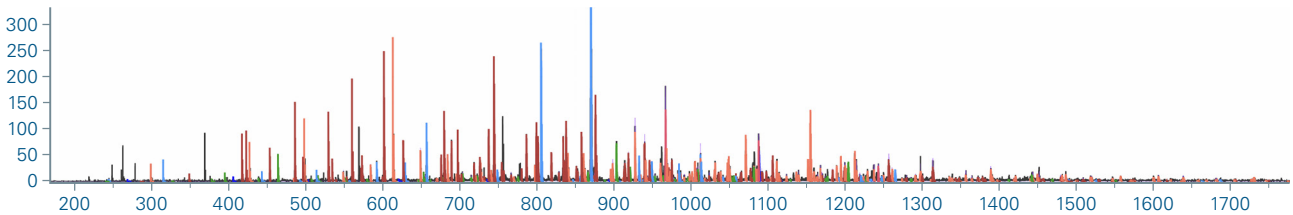

Figure 5. Fully annotated MS² ECD spectrum.

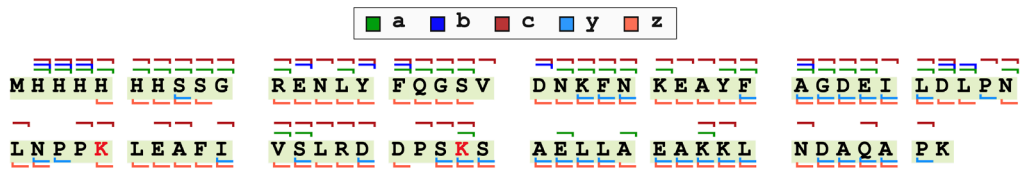

Figure 6. MS² ECD sequence map after amino acid substitution analysis.

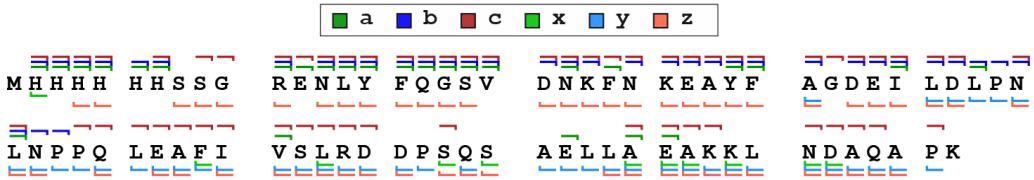

Figure 7. MS² ECciD sequence map on sequence revision 3.

Conclusion

- Definitive proteoform identification and 100% sequence coverage were achieved using timsOmni top-down data.
- OmniScape *de novo*-sequencing workflow was essential to characterize the N-terminal with unexpected His-tag and 12 amino acid sequence extension.
- OmniScape clipping analysis workflow successfully identified the correct C-terminus.
- The proteoform screening workflow enabled the rapid and systematic comparison of large numbers of single amino acid substitutions to refine and validate the reconstructed protein sequence.

Bruker Daltonics is continually improving its products and reserves the right to change specifications without notice. © Bruker Daltonics 06-2026, TN-66, 1929347

For Research Use Only. Not for use in clinical diagnostic procedures.

Bruker Switzerland AG

Fällanden · Switzerland
Phone +41 44 825 91 11

Bruker Scientific LLC

Billerica, MA · USA
Phone +1 (978) 663-3660

info.ms@bruker.com – www.bruker.com

Learn more at **www.bruker.com/timsomni**