# Definitive Protein Sequencing by Top-Down MS with timsOmni and OmniScape

Learn more about TD Proteomics

**Dalila Bensaddek[1]; Miriam Escarlet Diaz Galicia[2]; Magnus Rueping[2]; Mariangela Kosmopoulou[3]; Athanasios Smyrnakis[3]; George Alevizos[3]; Dimitris Papanastasiou[3]; Maan Amad[1]**

[1]Analytical Chemistry Core Lab, King Abdullah University of Science and Technology, KAUST, Thuwal, Saudi Arabia; [2]KAUST Catalysis Center (KCC), Division of Physical Sciences & Engineering, King Abdullah University of Science and Technology, KAUST, Thuwal, Saudi Arabia; [3]Fasmatech Science & Technology, Chalandri, Greece

## Introduction

- Bottom-up (BU) proteomics is the dominant approach for protein identification and characterization; however, its reliance on peptide-centric measurements and protein inference can lead to incorrect or incomplete sequence assignments, particularly in the presence of sequence scrambling or proteolytically inaccessible regions.
- In contrast, top-down (TD) mass spectrometry directly analyzes intact proteoforms, preserving sequence connectivity that may be lost in bottom-up workflows. Despite this advantage, *de novo* sequence analysis by top-down MS remains challenging due to limited availability of dedicated software.
- Here, a protein sequence originally inferred from bottom-up data was re-examined using top-down experiments on a timsOmni™ platform with multiple fragmentation modes. Comprehensive OmniScape™ analysis enabled reconstruction of a corrected protein sequence.

## Methods

- The ~9 kDa protein was initially dialyzed against water and lyophilized. Prior to analysis, the sample was reconstituted at 4 µM in 50:49.5:0.5 $H_2O$:ACN:FA and analyzed offline on a timsOmni platform using an Apollo II source at 3 µL/min. MS² experiments were performed on the 8+ precursor charge state using multiple fragmentation methods, including $_C$CID, $_R$CID, ECD, ECciD and EID.
- Data processing was performed in OmniScape using multiple workflows. Data analysis followed a stepwise approach (Results 1-4), starting from zero sequence coverage of the initially proposed, BU-derived protein sequence, progressing through *de novo* sequence reconstruction and clipping analysis, and concluding with mutation screening to confirm the revised sequence, ultimately achieving 100% sequence coverage (SC).

## Results

### 1

**Observation**

TD-MS measured mass is higher than the BU-predicted mass, indicating sequence inconsistency.

Additionally, the confirmation workflow on MS2 ECD data resulted only in random fragment matches (Fig. 1).


Fig. 1 MS² ECD sequence map on the BU-derived sequence.

**Method**

*De novo* on MS² ECD data: the top-scored sequence (Fig. 2) tag includes a His tag connected to the N-terminal part of the protein. Additionally, the first fragment identified has a mass shift corresponding to MHH amino acids.

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

Fig. 2 top-scored *de novo* sequence tag including a His tag (red) connected to the protein N-terminus (green).

**Revised Sequence**

Confirmation result of the first sequence revision, based on the *de novo* analysis, is shown in Fig. 3.


Fig. 3 MS² ECD sequence map on sequence revision 1.

### 2

The first sequence revision provided full N-term SC until amino acid O45 and minor coverage on the protein C-terminus.

Additionally, the theoretical neutral mass is still higher than the TD-MS measured mass.

*De novo* on isCID data: another sequence tag was consistent with the BU-derived sequence near C-terminus; however, the systematically lower y-ion masses further underlined that the BU-proposed sequence is erroneously longer.

Clipping analysis was subsequently applied to iteratively trim the C-terminal region of the first sequence revision.

Confirmation result (Fig. 4) of the second sequence revision provided great coverage near both protein termini.

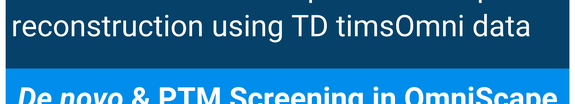


Fig. 4 MS² ECD sequence map on sequence revision 2.

### 3

Coverage gap is observed between O45-O59 for the second sequence revision, while some high-abundance peaks remain unassigned in the MS² ECD.

Targeted amino acid substitution analysis for O45 and O59, based on the MS² ECD data, revealed the substitution of both O residues with Q (Fig. 5, 6).

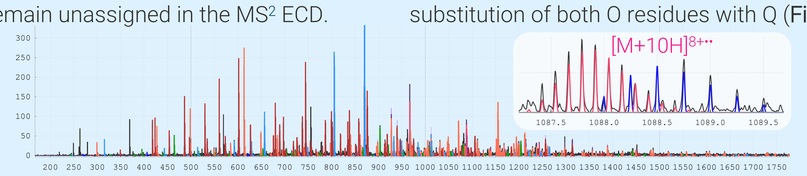


Fig. 5 Fully annotated MS² ECD spectrum and perfectly matched precursor ion isotopic distribution.

Confirmation result (Fig. 6) of the third sequence revision provided full sequence coverage.


Fig. 6 MS² ECD sequence map on sequence revision 3 (K45Q - K59Q best matching proteoform).

### 4

Despite the almost full coverage of the third sequence revision, further validation of the heavily revised N-terminal region was considered necessary.

Single-site Substitution Screening across residues M1–V20 via PTM Screening workflow, further validated the revised N-terminal region, with the *de novo*–derived amino acid series yielding the highest sequence and intensity coverage for the MS² ECD data.

Finally, application of the confirmation workflow on the MS² CID, MS² EID, and MS² ECciD data further confirmed the sequence identified in OmniScape, with MS² ECciD achieving SC=100% (Fig. 7).


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

## Summary

**Definitive sequence identification**

- ✓ His-tag confirmed at N-terminus
- ✓ A new 12 amino acid sequence was identified *de novo* as part of the N-terminal region of the protein
- ✓ C-terminal over-length was identified *de novo* and corrected by Clipping Analysis
- ✓ O45Q and O59Q substitutions identified via PTM Screening
- ✓ Revised sequence was validated by Single-site Substitution Screening and further confirmed by multiple fragmentation methods with SC=100%

## Conclusion

*De novo* sequencing & PTM Screening workflows enable protein sequence reconstruction using TD timsOmni data

***De novo* & PTM Screening in OmniScape**