Filling the gaps: multi-experiment *de novo* tag integration for confident top-down protein characterization in OmniScape

George Alevizos[1]; Mariangela Kosmopoulou[1]; Georgia Orfanoudaki[1]; Detlev Suckau[2]; Athanasios Smyrnakis[1]; Liliane Soares[3]; Dimitris Papanastasiou[1]

[1]Fasmatech Science & Technology, Chalandri, Greece; [2]Bruker Daltonics GmbH & Co KG, Bremen, Germany; [3]Bruker, San Diego, CA

Learn more

Introduction

- Top-down mass spectrometry (TD-MS) enables direct analysis of intact proteoforms; however, *de novo* interpretation of tandem mass spectra remains challenging due to spectral complexity, diverse fragmentation behaviour, incomplete sequence coverage, and instrument-dependent factors.
- Here, we present a novel *de novo* sequencing algorithm developed in the OmniScape™ software that produces high-confidence sequence tags by leveraging multimodal fragmentation and multiple information-rich MS^n spectra, generated using the timsOmni™ mass spectrometer.
- This approach allows generating sequence tags by combining MS^n spectra enhancing downstream local alignment searches (LYRA) resulting in improved protein identification.

Methods

- The advanced *de novo* workflow was implemented in C++ incorporating the OmniWave™ deconvolution algorithm. Averagine-based isotopic distributions were identified and used to construct a spectral graph. Sequence tags were evaluated using composite scoring metrics that account for both the quality and significance of fragmentation features.
- Algorithmic performance was evaluated using annotated MS/MS spectra of bovine carbonic anhydrase II (CA II) and histones acquired by direct infusion and multimodal fragmentation ($_C$CID, $_R$CID, ECD, ECciD, and EID) on the timsOmni platform.
- The sequence tags generated by the combined spectral graphs were submitted to local alignment searches using LYRA in an integrated workflow within OmniScape.

Results

CA II

MS^2 ECD data were acquired for CA II precursor charge states z = 31+, 34+, and 37+. While individual charge states yielded incomplete and only partial overlapping sequence tags, combining the experiments generated longer composite sequence tags. This integration substantially improved the total sequence coverage (SC) increasing it from approximately 29% to 45%.

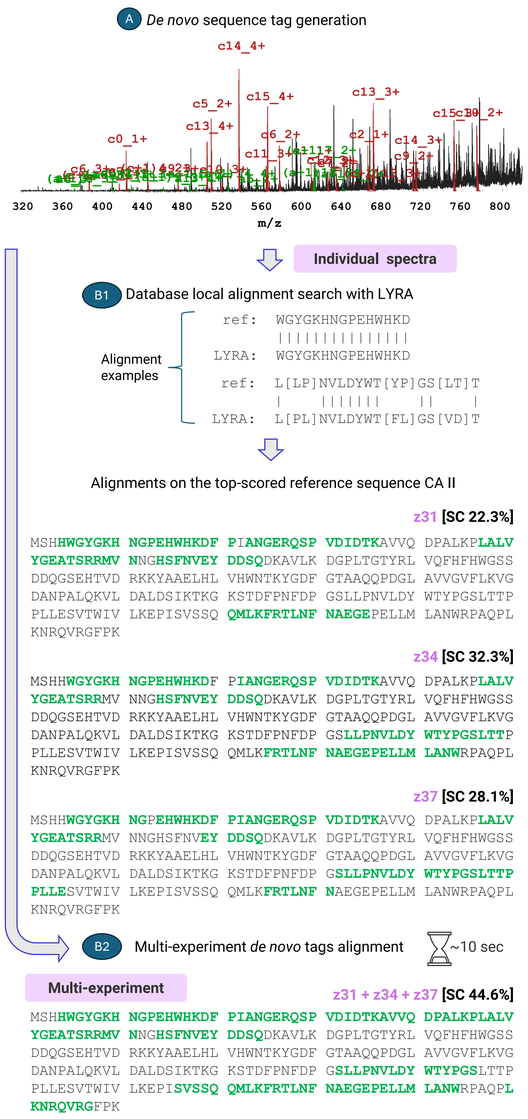

Fig. 1 A) *De novo* sequence tag generation and B1) LYRA alignment results for individual MS^2 ECD data *vs* B2) Multi-experiment *de novo* tags alignment on CA II sequence.

Figure 2 illustrates the improvement in top-down homology searches achieved by generating sequence tags from multiple experiments. For this sequence region, the integrated analysis increased the alignment bit score by ~57%, from ~35 to 55.

```
                                                    z31
ref:    220 Q[MLK]FRTLNFNAEGE 234
                  ||||||||  |
LYRA:       E[LRC]FRTLNFNANAE

                                                    z34
ref:    224 FRTLNFNA[EGEP][EL]LMLA[NW] 243
            |||||  |       ||||
LYRA:       FRTLNMEA[NANL][NK]LMLA[HY]

                                                    z37
ref:    224 FRTLNFN 230
            |||||||
LYRA:       FRTLNFN

                                                    z31 + z34 + z37
ref:    215 [SVSSQ][QMLK]FRTLNFNA[EGEP]ELLMLA[NW] 243
                        |||| |||      ||||||
LYRA:       [FLDL ][LKRC]FRTLLFNA[NANL]ELLMLA[HY]
```

Fig. 2 Comparison of sequence tag alignments highlighting the improvement of homology search performance.

MS-BLAST bit scores quantify alignment significance using a database-independent, log-scaled measure, with higher scores indicating matches that are less likely to occur by chance. Because standard MS-BLAST scoring treats isobaric sequence segments as gaps or mismatches, alignment scores are reduced. LYRA incorporates isobaric segments as valid matches, improving alignment sensitivity (**Fig. 3**).

```
ref:        L[LP]NVLDYWT[YP]GS[LT]T
            |    |||||||    ||    |
LYRA:       L[PL]NVLDYWT[FL]GS[VD]T      48.42 bits
                 ||||||| -- || -- |
MS-BLAST:        NVLDYWT YP GS LT T      34.30 bits


ref:        FRTLNFNA[EGEP][EL]LMLA[NW]
            |||||  |       ||||
LYRA:       FRTLNMEA[NANL][NK]LMLA[HY]   42.04 bits
            |||||--| ----  -- ||||
MS-BLAST:   FRTLNFNA EGEP  EL LMLA       22.00 bits
```

Fig. 3 Comparison of sequence tag alignments highlighting the LYRA isobaric segment identification (in green) leading to higher-length and better-scored matches.

Histone 3.1

In addition to combining data from different charge states, complementary fragmentation methods were applied to histone 3.1 using the timsOmni platform. **Figure 4** highlights a multimodal *de novo* sequencing workflow for H3.1 that leverages these complementary datasets.

```
                                                    H3.1 EID [SC 39.7%]
MARTKQTARK STGGKAPRKQ LATKAARKSA PATGGVKKPH RYRPGTVALR
EIRRYQKSTE LLIRKLPFQR LVREIAQDFK TDLRFQSSAV MALQEACEAY
LVGLFEDTNL CAIHAKRVTI MPKDIQLARR IRGERA

                                                    H3.1 ECD [SC 19.9%]
MARTKQTARK STGGKAPRKQ LATKAARKSA PATGGVKKPH RYRPGTVALR
EIRRYQKSTE LLIRKLPFQR LVREIAQDFK TDLRFQSSAV MALQEACEAY
LVGLFEDTNL CAIHAKRVTI MPKDIQLARR IRGERA

                                                    H3.1 CID [SC 14%]
MARTKQTARK STGGKAPRKQ LATKAARKSA PATGGVKKPH RYRPGTVALR
EIRRYQKSTE LLIRKLPFQR LVREIAQDFK TDLRFQSSAV MALQEACEAY
LVGLFEDTNL CAIHAKRVTI MPKDIQLARR IRGERA

                                                    H3.1 EID + ECD + CID [SC 66.2%]
MARTKQTARK STGGKAPRKQ LATKAARKSA PATGGVKKPH RYRPGTVALR
EIRRYQKSTE LLIRKLPFQR LVREIAQDFK TDLRFQSSAV MALQEACEAY
LVGLFEDTNL CAIHAKRVTI MPKDIQLARR IRGERA
```

Fig. 4 Comparison of sequence tag alignments highlighting the benefit of combining multiple experiments leading to higher-length and better-scored matches.

Summary

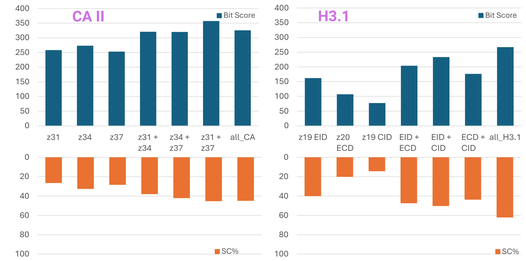

Fig. 5 Bit Score and SC% overview for CAII and H3.1.

Conclusions

- Multimodal fragmentation and multi-experiment analysis generating longer sequence tags, filling sequence gaps, and improving *de novo* sequencing performance.
- LYRA improving top-down homology searches by accounting for isobaric regions.

***De novo* sequencing in OmniScape**

© 2026 Bruker

Conflict of Interests
Authors from Fasmatech - Bruker may hold financial interests in the company.

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

Conflict of Interests
GA, MK, GO, DS, AS, LS and DP are employees of Bruker, which develops and manufactures advanced mass spectrometry platforms for life science applications. This presentation includes data and findings related to Bruker's OmniScape product.

Innovation with Integrity