Impact of internal fragment ion identification on oligonucleotides scoring metric

Abhishek Roushan, Nelson Pang, Ignat Shilov, St John Skilton

Protein Metrics, Boston, MA, USA

Overview: Oligonucleotide analysis | Project inspection and probabilistic scores | MS2 and internal fragmentation | Discriminant analysis and correlation

Introduction

Using mass spectrometry for the analysis of complex oligomer molecules is a common practice nowadays. For such experiments, MS2 analysis of oligomers poses a particular challenge as oligonucleotides, unlike peptide molecules, produce very dense MS2 spectra with a lot of byproduct ions resulting from nucleotide base losses, ether linker fragmentations, and internal fragments. For larger oligonucleotides >30mer the MS2 spectra are particularly complex, requiring sophisticated isotope and charge analysis to reduce chances of random fragment matching.

In this work, we address the open question on whether the inclusion of internal fragments helps to improve discrimination power of the MS2 scoring scheme based on a probabilistic binomial model.

Methods: Oligo MS2 analysis with internal fragment scoring

In this workflow, we assess the value of inclusion of internal fragments to discriminate between true and scrambled/shuffled sequences. We vary scrambling to a different degree, from fully shuffled cases to a single residue pair flips, to to assess the discrimination power for various alternatives.

We also analyze the discrimination value of Protein Metrics's isotope envelope based confidence scoring for individual MS2 ions.

As an extension of the work, we also assessed scoring of common single base span internal fragment intensities with the frequency of the bases within oligonucleotide sequence.

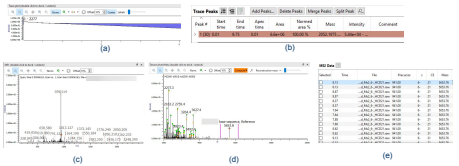


Figure 1: *Workflow (a) Sample traces are (b) segmented to extract (c) MS1 plots and to compute (d) Deconvolved masses, whose regions of interest in time are determined for extracting (e) MS2 data and thereby computing the oligo scores.*

Project inspection insights

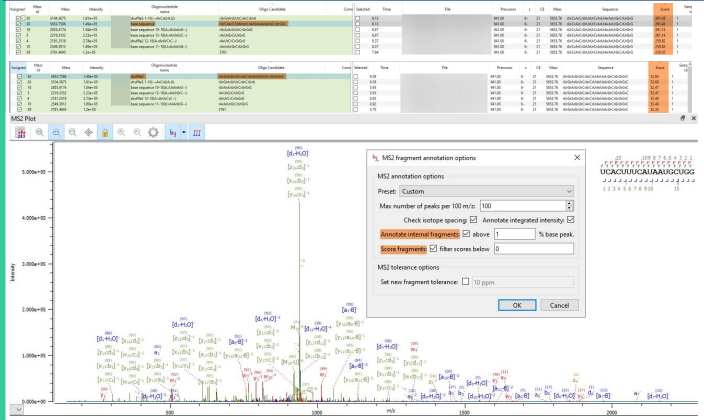


Figure 2: *Oligonucleotide MS2 analysis and binomial 'Score' and control to annotate, score internal fragments*

Internal fragment ions illustration and residue span visualization in MS2 Plot

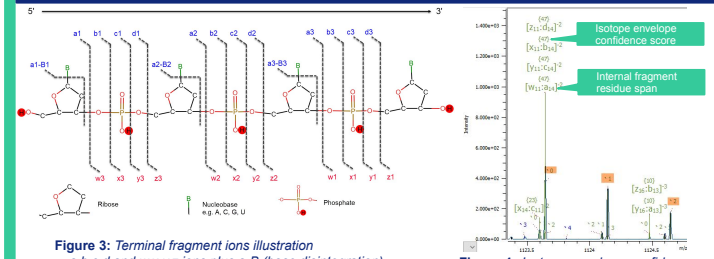
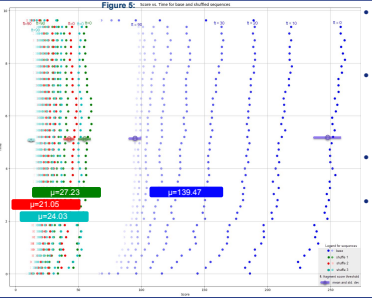


Figure 3: *Terminal fragment ions illustration*
a,b,c,d and w,x,y,z ions plus a-B (base disintegration)
Internal fragment ions with residue span across nucleobases e.g. [w2:d1] ion spans the 2nd nucleobase

Figure 4: *Isotope envelope confidence score {47} for internal fragment residue span e.g. w11 to a14*

Score discriminant analysis

Discrimination of base and shuffled sequences scores with fragment score threshold



- Applying different fragment score threshold (*ft*) filters {0,10,20,...,90} result in the formation of 'bands' of score-time points
- Std. deviation decreases with increasing *ft*, both within and across base, shuffled sequences distributions indicating a high discrimination power (inv. proportional to *ft*)
- Distribution of base scores separated from shuffled sequences by a min. 112 score
- Inclusion of internal fragments increases the gap, improves discrimination, and their exclusion weakens the discrimination of base and shuffled sequences by increasing the chances of 'bands' to overlap.

Correlation between nucleobase frequency and internal fragment ion intensity

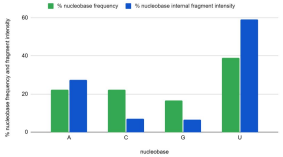


Figure 6: *Column chart of % frequency and internal fragment intensity for all nucleobases illustrating the correlation.*

We utilized consecutive span [w:d] ions' observed intensities for reporting intensity of nucleobases.

Nucleobase frequency is computed from no. of times the nucleobase occurs in the oligonucleotide sequence.

We find a positive correlation between nucleobase frequency and corresponding nucleobase span fragment ion intensities.

The helps us in predicting fragment ion intensities for base from oligonucleotide sequence nucleobase composition, and vice versa.

Conclusions and future work

- Fragment ion score thresholds are shown to assist the software and the user, in discriminating base sequence matches against shuffled ones in oligonucleotide analysis.
- Correlation between internal fragment ion intensities and nucleobase frequency is demonstrated. We expect them to be used in each other's prediction for future applications.
- We plan to incorporate oligonucleotide mutation into score discriminant & frequency analyses.
- We also plan to experiment with the behavior of other internal fragment ions like a-B and B (nucleobase disintegration) for their impact on on oligonucleotide MS2 scores