Team COUNCIL OF RICKS submission for EUPA YPIC 2017

Council of Ricks (Pino, Lindsay K; Searle, Brian C; Yang, Han-Yin)
Department of Genome Sciences
University of Washington, Seattle

ABSTRACT

Using primarily LC-MS/MS *de novo* sequencing techniques, we concluded that the peptides form the following sentence from Sir JJ Thomson's preface to Rays of Positive Electricity and Their Application to Chemical Analyses:

> *"I feel sure that there are many problems in chemistry which could be solved with far greater ease by this than by any other method. The method is surprisingly sensitive, more so even than that of spectrum analysis, requires an infinitesimal amount of material, and does not require this to be specially purified."*

Here, we detail our process for deciphering the peptide mixture composition and finding the sentence they form and from what book the sentence comes from.

RESULTS AND DISCUSSION

Our final approach reflected our collective prior experience in liquid chromatography mass spectrometry (LC-MS), especially focusing on the interpretation and analysis of data dependent acquisition (DDA). We first acquired MS1 and MS2 spectra of a diluted sample of the peptide mixture using a DDA method. Several distinct peaks were visually apparent in the Total Ion Current (TIC), suggesting high abundance of some ion species, and so we manually *de novo* sequenced MS/MS spectra of the most intense peaks. Manual sequencing produced several English-word "tags", including SENSITIVE (Figure 1), EVENT, ATTHE, and OTHER.

To produce more English-word "tags", we then used an automated *de novo* sequencing algorithm, Novor. The Novor results from digested and undigested samples of the peptide mixture were concatenated and the data cleaned using Excel ("K(Acetyl)" replaced with "O", "S(Phosphorylated)" replaced with B; duplicate results dropped). The results were sorted by score, revealing the additional candidate sequences ANYOTHER, ANALYSIS, THEMETHOD, and SPECTR.

We combined these confident sequences from manual and automated *de novo* sequencing into a single search term (sensitive+spectrum+"any other method"+analysis+more+sensitive+event+"the method") which we queried across Google Book (https://books.google.com/, accessed 09/08/2017). Out of 1,550 results, the second hit for this search term was Rays of Positive Electricity and Their Application to Chemical Analyses by JJ

Thomson (Figure 2). We used the two sentences from which our confident sequences arose as a primary candidate, and searched for additional evidence supporting the candidate.

The 92,804 unique Novor result values were used to create a pseudo-FASTA along with the candidate sentence. The candidate sentence was manipulated so that all instances of “U” were replaced with “O”, as Novor was unable to distinguish methylated arginine (“U”) and therefore all Novor results included O in place of U.

Additional attempts not detailed here include automated *de novo* sequencing softwares LutefiskXP (v1.0.7) and DeepNovo; and a narrow isolation window data-independent acquisition (DIA) strategy combined with DIA-Umpire to produce pseudo-spectra for *de novo* sequencing.

CONCLUSIONS

From the LC-MS/MS evidence, we conclude that the peptide mixture comes from Sir JJ Thomson’s preface to Rays of Positive Electricity and Their Application to Chemical Analyses:

> *“I feel sure that there are many problems in chemistry which could be solved with far greater ease by this than by any other method. The method is surprisingly sensitive, more so even than that of spectrum analysis, requires an infinitesimal amount of material, and does not require this to be specially purified.”*

MATERIALS AND METHODS

Sample preparation. Sample was divided into four equal-volume aliquots of approximately 10ul each. One 10ul aliquot was serially diluted with 2% MeCN, 0.1% formic acid in water to an estimated 5,000 fmol/ul sample, a 500 fmol/ul sample, and a 50 fmol/ul sample. Additionally, 18ul of the 5,000 fmol/ul diluted sample were enzymatically digested for 4 hours at 37C with 0.4ug trypsin.

Liquid Chromatography Mass Spectrometry. Peptides were separated with a Waters NanoAcquity UPLC and emitted into a Thermo Q-Exactive HF tandem mass spectrometer. Pulled tip columns were created from 75 µm inner diameter fused silica capillary in-house using a laser pulling device and packed with 2.1 µm C18 beads (Dr. Maisch GmbH) to 300 mm. Trap columns were created from 150 µm inner diameter fused silica capillary fritted with Kasil on one end and packed with the same C18 beads to 25 mm. Buffer A was water and 0.1% formic acid, while buffer B was 98% acetonitrile and 0.1% formic acid. For each injection, 3 µl of each sample was loaded with 5 µL 2% B and eluted using the following program: 0-90 minutes 2%-35% B, 90-100 minutes 35%-60% B, followed by a 35 minute washing gradient. Data were acquired using data-dependent acquisition (DDA).

DDA Acquisition and Processing. The Thermo Q-Exactive HF was set to positive mode in a top-20 configuration. Precursor scans (300-2000 M/Z) were collected at 60,000 resolution to hit

an AGC target of 3e6. The maximum inject time was set to 100 ms. Fragment scans were collected at 30,000 resolution to hit an AGC target of 1e5 with a maximum inject time of 55 ms. The isolation width was set to 1.6 M/Z with a normalized collision energy of 27, 30, or 33. Precursors with charge up to +6 that achieved a minimum AGC of 5e3 were acquired. Dynamic exclusion was disabled. Thermo RAW files were converted to ms2 format using MSConvert and searched using Novor (version 1.6.634), allowing for variable acetylated lysine, phosphorylated serine and threonine. Cysteines were assumed to not be carbamidomethylated. Searches were performed using a 15 ppm precursor error tolerance and a 0.1 Da fragment tolerance and enzymatic digestion by trypsin. XCalibur was used to visualize the raw LC-MS/MS data.

FIGURES

Figure 1. Manual de novo sequencing of “SENSITIVEM--”

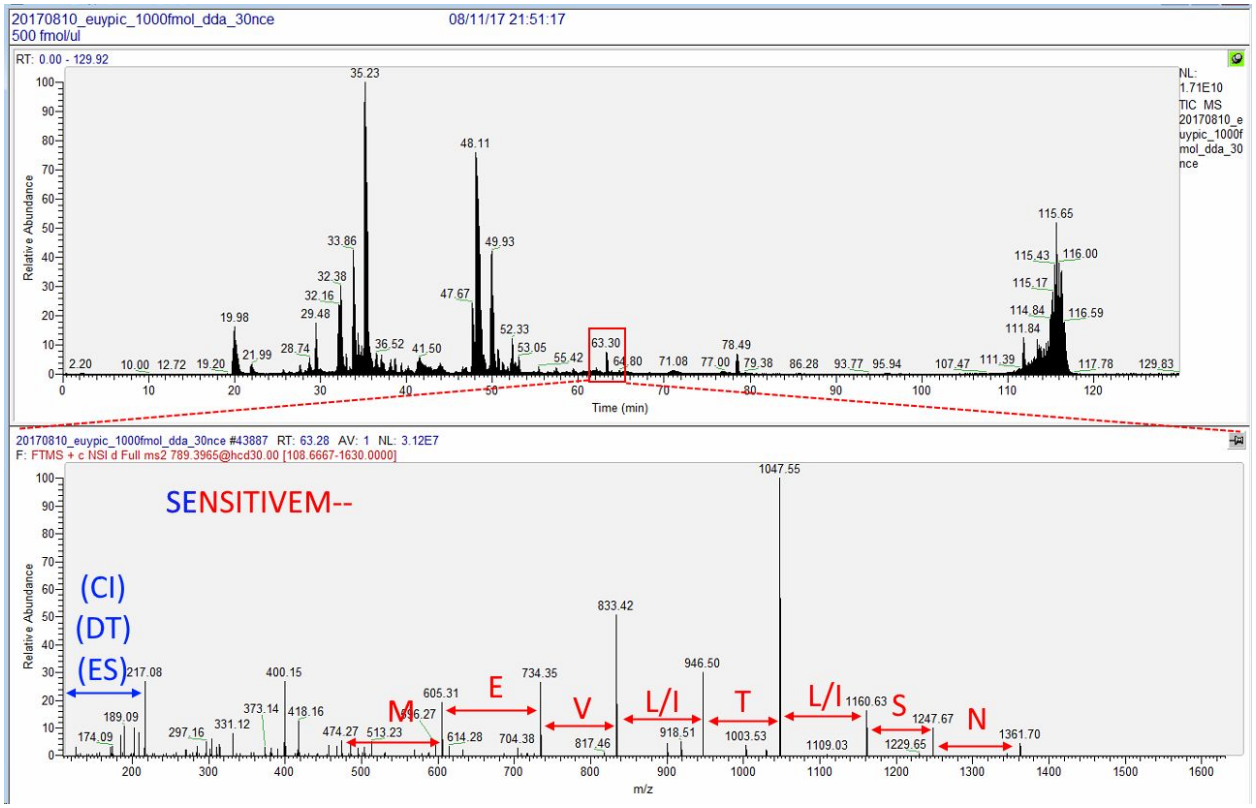

Figure 2. Screenshot of Google Books search results