EuPA Open Proteomics 22–23 (2019) 25–27

Contents lists available at ScienceDirect

EuPA Open Proteomics

journal homepage: www.elsevier.com/locate/euprot

Math, science, history, unraveling the mystery—That all started with *de novo*!

Ekaterina Ilgisonis, Olga Kiseleva, Ksenia Kuznetsova (Dr.)*

Institute of Biomedical Chemistry, 10/8 Pogodinskaya, Moscow, Russia

ARTICLE INFO

Keywords:
YPIC
EuPA
Challenge
Peptides
*De novo*sequencing

ABSTRACT

This work on solving the mystery of words encoded by amino acids in peptides was derived by the YPIC-EuPA Challenge. We received a dry synthetic peptide sample and performed a mass spectrometric analysis followed by *de novo* peptide sequencing. As a result, a part of “Rays of positive electricity and their application to chemical analyses” by J.J.Tomson was found to be encoded in the peptides of the sample. The words were first revealed from the peptides, that matched by Google search to find the answer. After that, the answer was validated using a standard proteomic search against a database constructed from the quotation found.

1. Introduction

Since, by now, the variety and complexity of the human proteome has not been studied completely, proteome investigators often face the necessity to analyze mixtures with unknown contents. To our mind, the most efficient approach to this is *de novo* sequencing from high- resolution LC–MS/MS data. The YPIC-EuPA Challenge turned out to be a great chance not only to improve our practical deciphering skills, but also to refresh the history of mass spectrometry (Fig. 1).

2. Materials and methods

The mixture consisting of 19 synthetic peptides obtained from the Challenge organizing team was analyzed using Dionex Ultimate 3000 (Thermo Fisher Scientific) connected to a Hybrid Ion Trap-Orbitrap Elite mass spectrometer (Thermo Fisher Scientific), equipped with a nanoelectrospray ion source (Thermo Scientific). Peptides were loaded onto the trap column Zorbax 300SB-C18 (C18 5 µm 0.3 mm inner diameter and 5 mm length, Agilent Technologies, USA) and washed for 5 min at a flow rate of 10 µl/min. Peptide separation was performed on a RP-HPLC Zorbax 300SB-C18 column (C18 3.5 µm 75 µm inner diameter and 150 mm length, Agilent Technologies, USA) using a linear gradient from 5% to 60% solvent B (0.1% formic acid, 80% acetonitrile) over 30 min at a flow rate of 0.4 µl/min.

CID has been used as a fragmentation method. Both MS and MS/MS spectra have been obtained in an orbitrap analyzer. Resolution was set at 60,000 (m/z400) for MS and 15,000 (m/z400) for MS/MS scans.

The mass spectra have been analyzed using the trial version of PEAKS (Bioinformatics solutions Inc.) [1] and SearchGUI [2] with the parameters described in the next section.

3. Results

For *de novo* sequencing we used the trial version of PEAKS, all the results were exported to a CSV-file and transferred into an MS Excel table (Supplementary 1). We changed all the identified PTMs, mentioned in the description of the Challenge:

Methylation of R (R(+14.02)) → U Acetylation of K (K(+42.01)) → O Phosphorylation of S (S(+79.97)) → B

After that we changed all Ls to Is, because it is hard to distinguish them using mass- spectrometry [3]. All obtained results were processed manually, because the results of *de novo* sequencing may be jumbled [4] and the human brain has an ability to read jumbled words [5]. Thus, we tried to identify real words in the variety of sequences of the identified peptides. All the decoded words are highlighted in the Supplementary 1 with green color. The discovered words were non specific, but google search revealed a J.J. Thomson citation (see Fig. 2) from the book “Rays of positive electricity and their application to chemical analyses”.

4. Validation

For validation of the found citation we have created a. fasta file (Supplementary 2), containing only 1 protein with sequence, equivalent to the phrase. We used SearchGUI for the search [2]. Using SearchGUI we have identified the most part of the text fragments. Nevertheless words “small amount”, “infinitesimal amount” have not been identified.

* Corresponding author.
E-mail address: kuznetsova.ks@gmail.com (K. Kuznetsova).

https://doi.org/10.1016/j.euprot.2019.07.011
Received 9 July 2019; Accepted 17 July 2019
Available online 17 August 2019

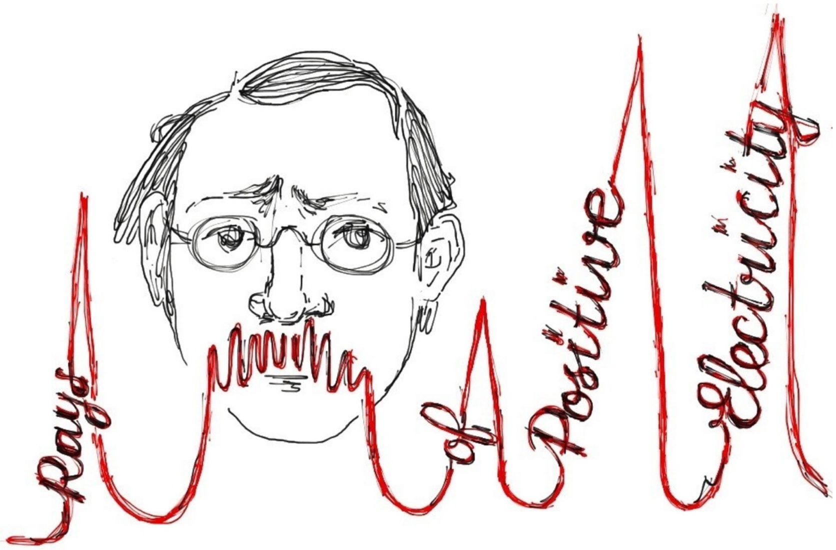
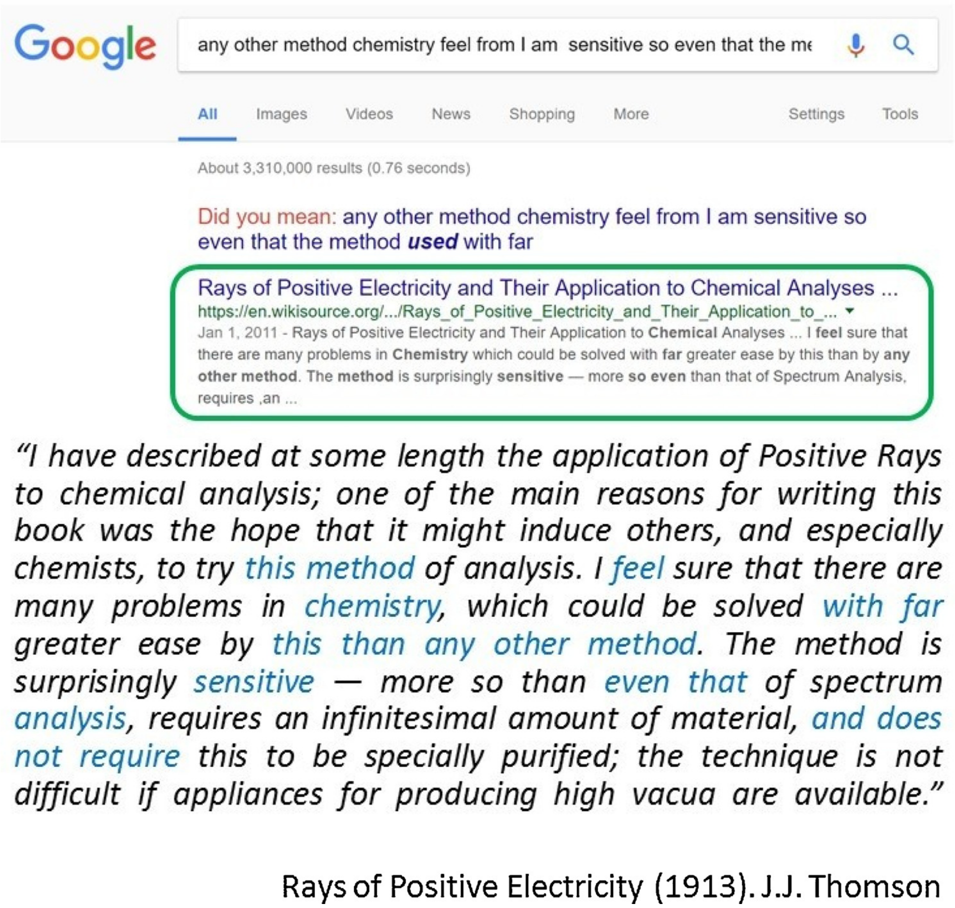


Fig. 1. Abstract graph.



any other method chemistry feel from I am sensitive so even that the me

About 3,310,000 results (0.76 seconds)

Did you mean: any other method chemistry feel from I am sensitive so even that the method ***used*** with far

Rays of Positive Electricity and Their Application to Chemical Analyses ...
https://en.wikisource.org/.../Rays_of_Positive_Electricity_and_Their_Application_to_... ▾
Jan 1, 2011 - 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 ...

“I have described at some length the application of Positive Rays to chemical analysis; one of the main reasons for writing this book was the hope that it might induce others, and especially chemists, to try this method of analysis. I feel sure that there are many problems in chemistry, which could be solved with far greater ease by this than any other method. The method is surprisingly sensitive — more so than even that of spectrum analysis, requires an infinitesimal amount of material, and does not require this to be specially purified; the technique is not difficult if appliances for producing high vacua are available.”

Rays of Positive Electricity (1913). J.J. Thomson

Fig. 2. The print screen of a Google search result revealing the quotation from J.J. Tomson. All the words identified by PEAKS are highlighted with blue color.

5. Instead of conclusion

We are grateful to the Challenge organizers for the chance to participate in such scientific riddle. And though we were given quite a lot of hints, like modifications and linguistic meaning of amino acids, this was not a piece of cake. Terrific task of highly complicated human proteome exploration remains a real challenge for lion-hearted.

Data

All the experimental data, *de novo* sequencing results and the fasta file used are available here at the Mendeley public repository under the title of this article.

Acknowledgments

We also thank our colleague Svetlana Novikova for the help during the mass spectrometry analysis and Alexey Chernobrovkin for valuable advice reguarding the bioinformatics part.

References

[1] B. Ma, K. Zhang, C. Hendrie, C. Liang, M. Li, A. Doherty-Kirby, G. Lajoie, PAEKS: Powerful Software for Peptide De Novo Sequencing by Tandem Mass Spectrometry, (2003), https://doi.org/10.1002/rcm.1196.

[2] M. Vaudel, H. Barsnes, F.S. Berven, A. Sickmann, L. Martens, SearchGUI: An Open-source Graphical User Interface for Simultaneous OMSSA and X!TAndem Searches, (2011), https://doi.org/10.1002/pmic.201000595.

[3] Y. Xiao, M.M. Vecchi, D. Wen, Distinguishing Between Leucine and Isoleucine by Integrated LC-MS Analysis Using an Orbitrap Fusion Mass Spectrometer Distinguishing Between Leucine and Isoleucine by Integrated LC- MS Analysis Using an Orbitrap Fusion Mass Spectrometer, (2016), https://doi.org/10.1021/acs.analchem.6b03409.

[4] J. Zhang, L. Xin, B. Shan, W. Chen, M. Xie, D. Yuen, et al., PEAKS DB : De Novo Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification, (2012), pp. 1–8, https://doi.org/10.1074/mcp.M111.010587.

[5] M. Perea, M. Jiménez, M. Martín-Suesta, P. Gómez, Letter Position Coding Across Modalities : Braille and Sighted Reading of Sentences With Jumbled Words, (2014), https://doi.org/10.3758/s13423-014-0680-8.