

## Manuscript Details

<b>Manuscript number</b>	EUPROT_2017_18
<b>Title</b>	Math, Science, History, Unraveling the mystery. That all started with the big bang de novo!
<b>Short title</b>	Math, Science, History, Unraveling the mystery
<b>Article type</b>	Short communication

### 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.

<b>Keywords</b>	YPIC; EuPA; challenge; peptides; de novo sequencing
<b>Corresponding Author</b>	Ksenia Kuznetsova
<b>Corresponding Author's Institution</b>	Institute of Biomedical Chemistry
<b>Order of Authors</b>	Ekaterina Ilgisonis, Olga Kiseleva, Ksenia Kuznetsova
<b>Suggested reviewers</b>	Dina Rešetar, Maarten Dhaenens

## Submission Files Included in this PDF

### File Name [File Type]

cover\_letter.docx [Cover Letter]

Ilgisonis\_ypic.docx [Manuscript File]

fig.1.jpg [Figure]

fig.2.jpg [Figure]

scan (1).pdf [Conflict of Interest]

To view all the submission files, including those not included in the PDF, click on the manuscript title on your EVISE Homepage, then click 'Download zip file'.

## Research Data Related to this Submission

**Data set** <https://data.mendeley.com/datasets/8gp9wnsjh3/draft?a=f0d298a9-63ef-4b11-b040-0ad78e0418dc>

Data for: Math, Science, History, Unraveling the mystery. That all started with the big bang de novo!

Mass spectra and other data for the YPIC-EuPA Challenge 2017 from the Russian team.

Dear Editor,

We are submitting this little article as a final manuscript for the YPIC-EuPA Challenge 2017. The manuscript is written according to the challenge's recommendations. All the data files have been stored in the Mendeley repository and linked to this submission.

Despite the exceptional educational approach of the work, we find it very useful and interesting. First, it may act as a demonstration of a *de novo* peptide sequencing technique. Second, it may be used as a sort of a benchmark for teams from different countries if the results are shared amongst the teams.

Finally, we are grateful to the organizers of the challenge for the opportunity to participate in such a captivating activity and happy to submit our manuscript with the results.

Sincerely,

Ekaterina Ilgisonis, Olga Kiseleva and Ksenia Kuznetsova

# Math, Science, History, Unraveling the mystery. That all started with the ~~big bang~~ *de novo*!

Ekaterina Ilgisonis<sup>1</sup>, Olga Kiseleva<sup>1</sup>, Ksenia Kuznetsova<sup>1\*</sup>

<sup>1</sup> Institute of Biomedical Chemistry, Moscow, Russia

e-mail: kuznetsova.ks@gmail.com

postal address: 10/8 Pogodinskaya, Moscow, Russia

## 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.

## 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.

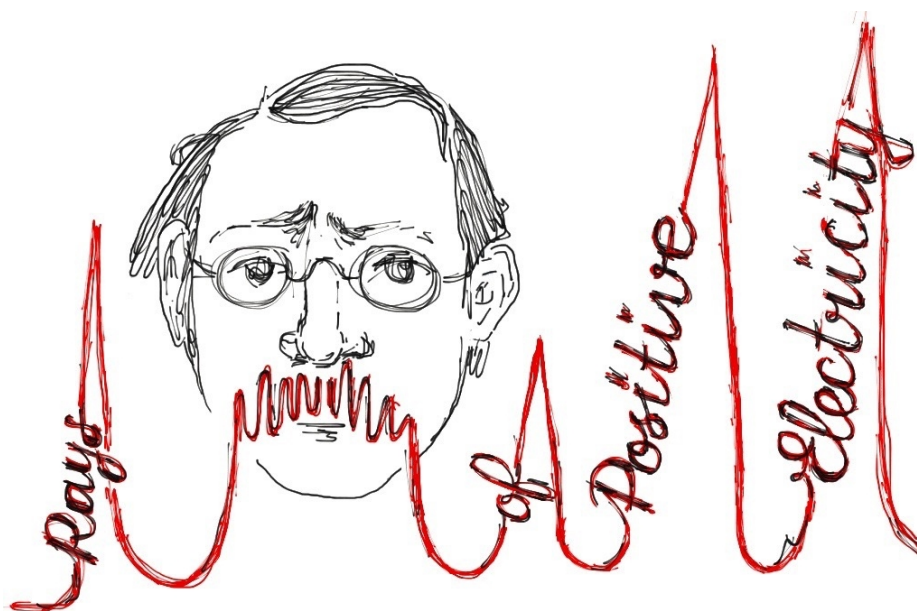


Figure 1. Abstract graph.

## 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  $\mu$ m 0.3 mm inner diameter and 5 mm length, Agilent Technologies, USA) and washed for 5 min at a flow rate of 10  $\mu$ l/min. Peptide separation was performed on a RP-HPLC Zorbax 300SB-C18 column (C18 3.5  $\mu$ m 75  $\mu$ m 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 $\mu$ l/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/z 400) for MS and 15,000 (m/z 400) for MS/MS scans.

The mass spectra have been analyzed using the trial version of PEAKS (Bioinformatics solutions Inc.) (Ma et al., 2003) and SearchGUI (Vaudel et al., 2011) with the parameters described in the next section.

## Results

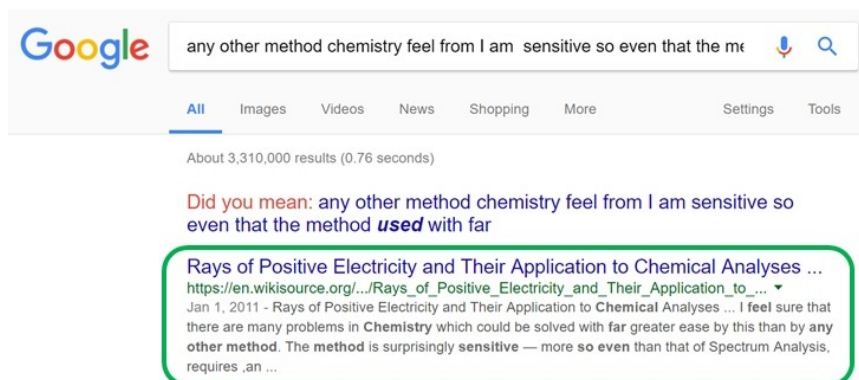
For *de novo* sequencing we used the trial version of PEAKS, all the results were exported to a CSV-file and transferred into an MS Excel table (Supplementary 1). We changed all the identified PTMs, mentioned in the description of the Challenge:

Methylation of R (R(+14.02))  $\rightarrow$  U

Acetylation of K (K(+42.01))  $\rightarrow$  O

Phosphorylation of S (S(+79.97))  $\rightarrow$  B

After that we changed all Ls to Is, because it is hard to distinguish them using mass-spectrometry (Xiao, Vecchi, & Wen, 2016). All obtained results were processed manually, because the results of *de novo* sequencing may be jumbled (Zhang et al., 2012) and the human brain has an ability to read jumbled words (Perea et al, 2014). 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 Figure 2) from the book “Rays of positive electricity and their application to chemical analyses”.



*"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

Figure 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.

### 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 (Vaudel et al., 2011). Using SearchGUI we have identified the most part of the text fragments. Nevertheless words "small amount", "infinitesimal amount" have not been identified.

### 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.

## Acknowledgments

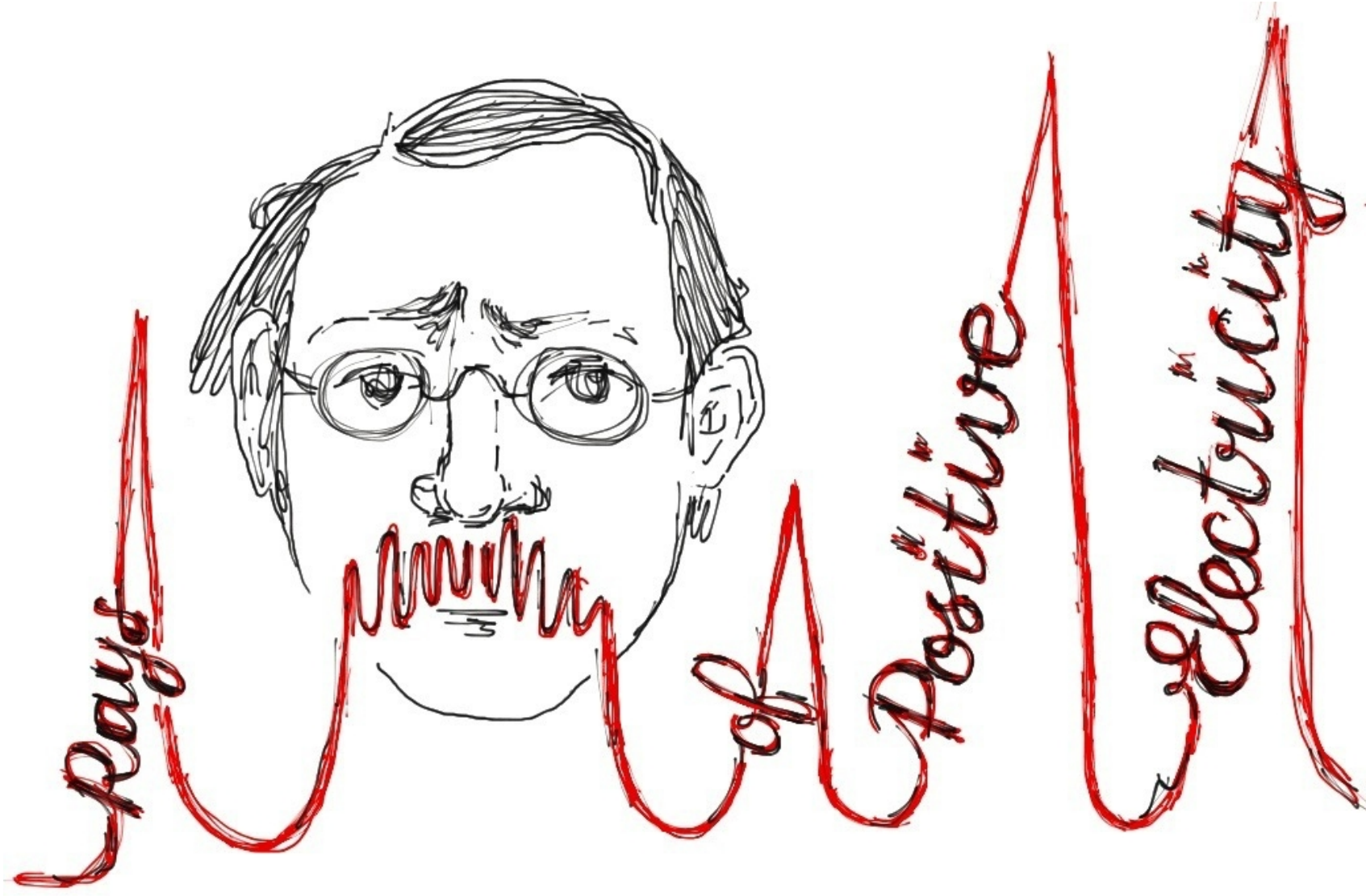
We also thank our colleague Svetlana Novikova for the help during the mass spectrometry analysis and Alexey Chernobrovkin for valuable advice regarding the bioinformatics part.

## 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.

## References:

- Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A., Lajoie, G. (2003). PAEKS: powerful software for peptide *de novo* sequencing by tandem mass spectrometry. <https://doi.org/10.1002/rcm.1196>
- Vaudel, M., Barsnes, H., Berven, F.S., Sickmann, A., & Martens, L. (2011). SearchGUI: An open-source graphical user interface for simultaneous OMSSA and X!Tandem searches. <https://doi.org/10.1002/pmic.201000595>
- Perea, M., Jiménez, M., Martín-Suesta, M., & Gómez, P. (2014). Letter position coding across modalities: Braille and sighted reading of sentences with jumbled words. <https://doi.org/10.3758/s13423-014-0680-8>
- Xiao, Y., Vecchi, M. M., & Wen, D. (2016). 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. <https://doi.org/10.1021/acs.analchem.6b03409>
- Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., ... Ma, B. (2012). PEAKS DB : De Novo Sequencing Assisted Database Search for Sensitive and Accurate Peptide Identification, 1–8. <https://doi.org/10.1074/mcp.M111.010587>







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



All

Images

Videos

News

Shopping

More

Settings

Tools

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/wiki/Rays\\_of\\_Positive\\_Electricity\\_and\\_Their\\_Application\\_to\\_Chemical\\_Analyses](https://en.wikisource.org/wiki/Rays_of_Positive_Electricity_and_Their_Application_to_Chemical_Analyses) ...

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



## Conflicts of Interest Statement

Manuscript title: Math, Science, History, Unraveling the mystery.  
That all started with the big bang de novo!

The authors whose names are listed immediately below certify that they have NO affiliations with or involvement in any organization or entity with any financial interest (such as honoraria; educational grants; participation in speakers' bureaus; membership, employment, consultancies, stock ownership, or other equity interest; and expert testimony or patent-licensing arrangements), or non-financial interest (such as personal or professional relationships, affiliations, knowledge or beliefs) in the subject matter or materials discussed in this manuscript.

Author names:

Ekaterina Ilgisonis

Olga Kiseleva

Ksenia Kuznetsova

The authors whose names are listed immediately below report the following details of affiliation or involvement in an organization or entity with a financial or non-financial interest in the subject matter or materials discussed in this manuscript. Please specify the nature of the conflict on a separate sheet of paper if the space below is inadequate.

Author names:

This statement is signed by all the authors to indicate agreement that the above information is true and correct (a photocopy of this form may be used if there are more than 10 authors):

Author's name (typed)

Author's signature

Date

Ekaterina Ilgisonis



09.09.2017

Olga Kiseleva



09.09.2017

Ksenia Kuznetsova



09.09.2017