Short thesis for the degree of doctor of philosophy (PhD)

Application of algorithms and neural networks to analyze the composition of copolymers

by Gergő Róth

Supervisor: Dr. Tibor Nagy

UNIVERSITY OF DEBRECEN

Doctoral School of Chemistry

Debrecen, 2024

1. Introduction and objectives

The use of copolymers has become almost essential in everyday life, and the production and application of purpose-designed copolymers to solve new problems is gaining ground. This demand also puts great pressure on the development of analytical methods for copolymers.

During my doctoral research, my goal was to develop a new algorithm for analyzing the mass spectrum of copolymers, that would allow this very complicated and time-consuming operation to be automated. For the development of the method, I used the [poly(ethylene oxide)]-[poly(propylene oxide)] copolymer family as sample compounds, since their use is very broad. One of my goals was to investigate whether a small change in the copolymer composition affects the polymer's physical properties. In addition, I wanted to visualize the extracted information, for which I wanted to apply the composition drift diagram.

In the course of the research, my goal was to implement the previously mentioned algorithm in a simpler way, so that it could also be implemented in simple spreadsheet software. To achieve this, I wanted to use the mass remainder analysis (MARA) discovered by our research group. In this case, I also used the aforementioned [poly(ethylene oxide)]-[poly(propylene oxide)] copolymer family as a sample compound family.

In addition, I wanted to find out whether the results of the mass spectrometry measurements could also be obtained using other methods. For this, I aimed to create an empirical model between the data from the gel permeation chromatography method and mass spectrometry. To solve the problem, I planned to use neural networks, since they can be universally used as empirical models. The great advantage of such a combination is that it would be possible to determine details on copolymers from inexpensive routine measurements.

2. Reaction conditions, applied equipment and methods

Mass spectrometric measurements were performed with an AutofMALDI-TOF-MS device. I used both reflectron and linear modes. In the case of reflectron measurements, the first voltage of the ion source was 19 kV, the voltage of the second was 16.65 kV, the first voltage of the reflector was 21 kV, while the second was 9.55 kV. In linear mode, the first voltage on the ion source was 19.5 kV, while the second voltage was 18.3 kV. The mass spectrometer was equipped with a solid-state laser (355 nm). During the research, I initially used 2,5-dihydroxybenzoic acid (DHB) and later trans-2-[3-(4-tert-butylphenyl)-2-methyl-2-propenylidene]malononitrile (DCTB) as a matrix, the ionizing agent in each case was sodium trifluoroacetate (NaTFA).

I used a Waters Alliance e2695 device for the gel permeation chromatography measurements. I used a Waters 2414 refraction index detector for detection. The measurements were performed in three different solvent mixtures or in pure solvent, in each case a different method was used. In the first case, the eluent was DMF, the columns were Shodex KF-603 and Shodex KF-602.5, the flow rate was 0.3 mL/min, the temperature was 42 °C, and the running time was 28 minutes. In the second case, the eluent was a mixture of THF:DMSO:piperidine (94:5:1 V/V%), the columns were Styragel HR 0.5, 1, 2 and 4, the flow rate was 0.5 mL/min, temperature 40 °C, running time was 35 minutes. In the third case, the eluent was a mixture of ACN:H2O (60:40 V/V%), the columns were PolySep-GFC-P 3000, the flow rate was 0.48 mL/min, the temperature was 40 °C, and the running time was 48 minutes.

The ^{1}H-NMR measurements were performed with a Bruker Avance II 500 MHz spectrometer. The measurement conditions of quantitative 1D 1H-NMR are as follows: 300 K, 90° ^{1}H excitation pulse, 15 s relaxation time, spectrum width 14.97 ppm, number of data points 32768, number of scans 1.

The dynamic light scattering measurement instrument was a Zetasizer Nano ZS to determine the critical micelle formation concentration and micelle size. The device was equipped with a He-Ne laser (633 nm). The measurements were at 37°C, the detecting angle was 173°.

3. New scientific results

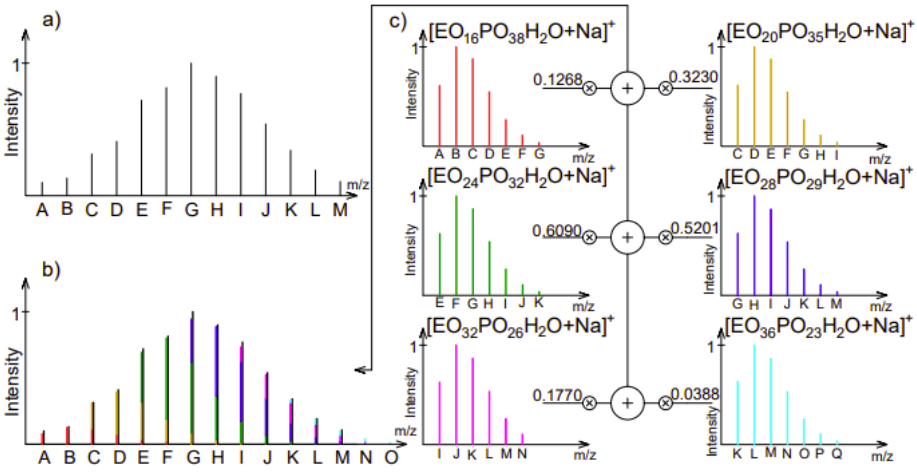
3.1. Detailed analysis of mass spectra of copolymers

3.1.1. I developed a mass spectra processing algorithm for the analysis of copolymers in details and tested it on the [poly(ethylene oxide)]-[poly-(propylene oxide)] (PEO)-(PPO) copolymer family

During my thesis I developed an algorithm for analyzing the mass spectra of copolymers, which can determine the parameters that characterize copolymers. The steps of the algorithm are as follows:

1. Import and normalization of the calibrated mass spectrum.
2. Calculation of possible components.
3. Calculation of the isotopes of the possible components and their distribution.
4. Checking the presence of the isotopes calculated for the components in the mass spectrum, thereby investigating the possibility of the component as a true element.
5. Determining the intensity fraction of all components by fitting.

Figure 1. illustrates the operation of the fifth step.



	m/z		m/z		m/z		m/z		m/z		m/z
A	2951.02	D	2954.02	G	2956.99	J	2959.96	M	2962.96	P	2965.92
B	2952.03	E	2955.00	H	2957.98	K	2960.96	N	2963.94	Q	2966.92
C	2953.02	F	2955.99	I	2958.97	L	2961.95	O	2964.94		

Figure 1: Identification of the components that make up the peak package in the mass spectrum of the copolymer and their fitting

The result of the algorithm is that the proportion of each component that makes up the copolymer can be determined and thus the polymer characteristics, such as the number average molecular weight (M_n), weight average molecular weight (M_w), and polydispersity index (PDI) can be calculated. In addition, it is possible to determine the number of monomers for each chain, and the averages for these can be calculated, such as the number average number of units (n_n^A) and number weighted average number of units (n_w^A). In addition, information on the distribution of monomers is obtained from these results (PDI^A). I validated the algorithm with data provided by ^{1}H-NMR.

3.1.2. I introduced a new copolymer parameter (polydispersity ratio, PDR), which explain the shape of composition drift diagram

I used the composition drift diagram to visualize the composition of the polymers making it easy to compare them. Furthermore, to characterize the drift slope, I introduced a new copolymer parameter, the polydispersity ratio (PDR), which can be calculated as the ratio of the polydispersity indexes of the monomers (PDI^A/PDI^B). This is represented in Figure 2.

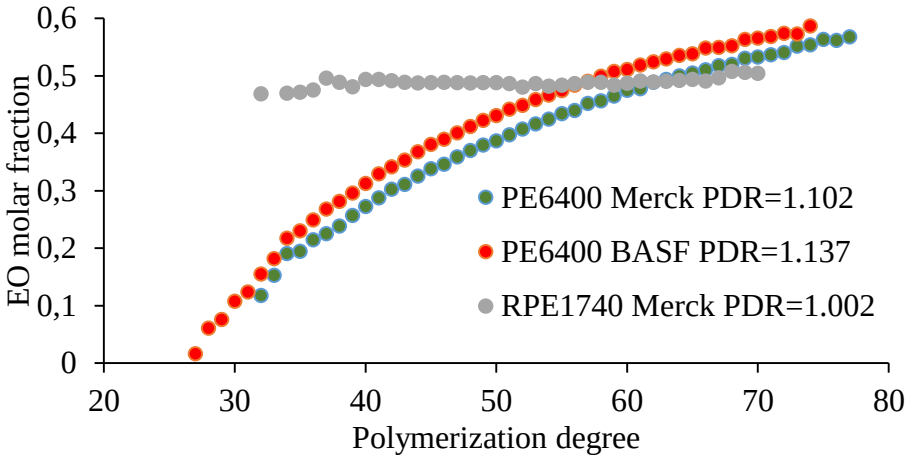


Figure 2: Composition drift diagram of PE6400 BASF (sample 1), PE6400 Merck (sample 2) and RPE1740 Merck (sample 3 (17R4)) copolymers

In the case of PEO-PPO-PEO triblock copolymers, the PDR value was always above 1, so the polydispersity index of EO is higher than the polydispersity index of PO. This is a consequence of the fact that the distribution of EO is wider. Therefore, I checked whether the wide distribution was due to the fact that only one PEO block was formed on the side of PPO chain during the polymerization, i.e. is it possible that

the diblock is present as a "contaminant". Based on the ^{1}H-NMR measurements, the percentage of primary hydroxyl endgroups is not 100%, in the case of PE6400 BASF (sample 1) it is 85.4% and in the case of PE6400 Merck (sample 2) 84.9%, so diblock copolymer is also present in the copolymers.

3.1.3. I proved that the differences revealed by the detailed characterization also appear in the self-assembly properties of the polymers, like critical micelle formation concentration and the hydrodynamic diameter of micelles

During the method development, I also analyzed analogue copolymers (PE6400 BASF (sample 1) and PE6400 Merck (sample 2)) according to the manufacturer's parameters and found differences in their characteristic parameters. Therefore, I measured the size of micelles formed from copolymers in aqueous solution, as well as the critical micelle formation concentration. In the case of PE6400 BASF (sample 1), the z-average hydrodynamic diameter of the micelle is 65.3 nm and the critical micelle formation concentration is 2.2 mmol/mL. In the case of PE6400 Merck (sample 2), the z-average hydrodynamic diameter was 18.2 nm, while the critical micelle formation concentration was 1.2 mmol/mL. In the case of both parameters, there is a significant difference between the two samples.

3.2. Data reduction analysis of mass spectra of copolymers

3.2.1. I proved that the composition of copolymers can also be determined from a reduced data set with a proper intensity correction procedure

The data reduction algorithm consists of the following main steps:

1. Selection of the most intense peak
2. Application of the MARA method to identify series and select peaks
3. Intensity correction
4. Calculation of polymer parameters

In the case of mass remainder analysis, the selected dividing factor was the molecular weight of the PO monomer, which is 58.042 g/mol. I used this, as well as a decryption code table, to determine the compositions. In addition, I developed an intensity correction procedure for the deconvolution of the overlapping peaks, which calculates the intensities from the monoisotopic peaks. This makes it possible to calculate the main parameters characterizing the copolymers in details. In addition, I plotted the results on the composition drift diagram and compared the results from the detailed analysis curve with the new results which can be seen in Figure 3.

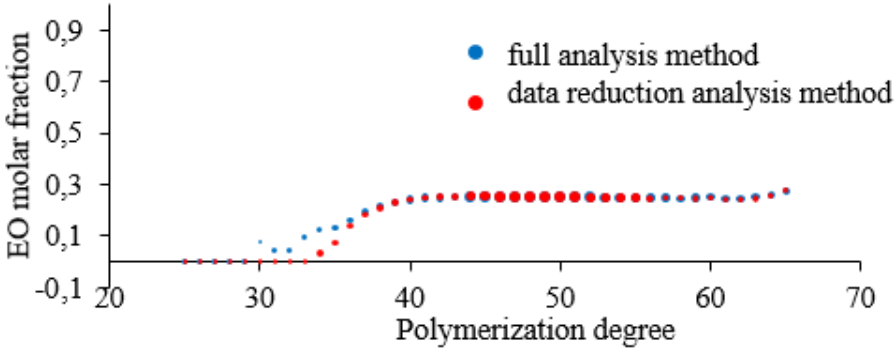


Figure 3: RPE2520 composition drift diagram with the complete analysis algorithm and the data reduction method

3.3. Artificial Intelligence Assisted Gel Permeation Chromatography (GPC) Copolymer Analysis

3.3.1. I developed a neural network based method that can calculate the exact number average molecular weight and monomers ratio of the commercially available [poly(ethylene oxide)]-[poly(propylene oxide)] copolymers from gel permeation chromatography (GPC) data

The input parameters of the neural network were the M_p values determined by gel permeation chromatography (GPC), while the output parameters were the M_n values measured by MALDI-TOF-MS and the EO n/n% content determined by the ^{1}H-NMR measurement method. The network contained one hidden feedforward layer with three neurons and the transfer functions were sigmoid functions. The calculation performance of the network is presented in Table 1.

Table 1: Correlation analysis of calculated and measured results

	M_n	EO%
r	0.988	0.990
accuracy %	4.9	2.3
p	0.95	0.51

3.3.2. I trained a neural network that can be applied for [poly-(ethylene oxide)]-[poly-(propylene oxide)] copolymers (Mn<5000 g/mol) for detailed analysis using GPC measurements

In this case, the input parameters of the second neural network were also the M_p values provided by GPC and the block sequence. The output

parameters were the number average number of units (n_n^{EO}, n_n^{PO}) and the number weighted average number of units (n_w^{EO}, n_w^{PO}) calculated for both monomers. Regarding the topology of the network, it contained one hidden feedforward layer, with five neurons in the layer, and the neurons with sigmoid transfer functions. Table 2. shows the precision of the network.

Table 2: Correlation analysis of calculated and measured results

	n_n^{EO}	n_w^{EO}	n_n^{PO}	n_w^{PO}
r	0.996	0.995	0.983	0.983
accuracy %	1.6	1.6	2.2	2.2
p	0.09	0.08	0.34	0.35

3.3.3. I showed correlation between the PDR and the ratio of the primary hydroxyl end groups, in the case of a normal block sequence for [poly-(ethylene oxide)]-[poly-(propylene oxide)] copolymers

During the analysis of the data, I found a correlation between the percentage of the primary hydroxyl end group measured by 1H-NMR and the PDR, which, as presented in Chapter 3.1.2, supports the presence of diblock as a "contaminant".

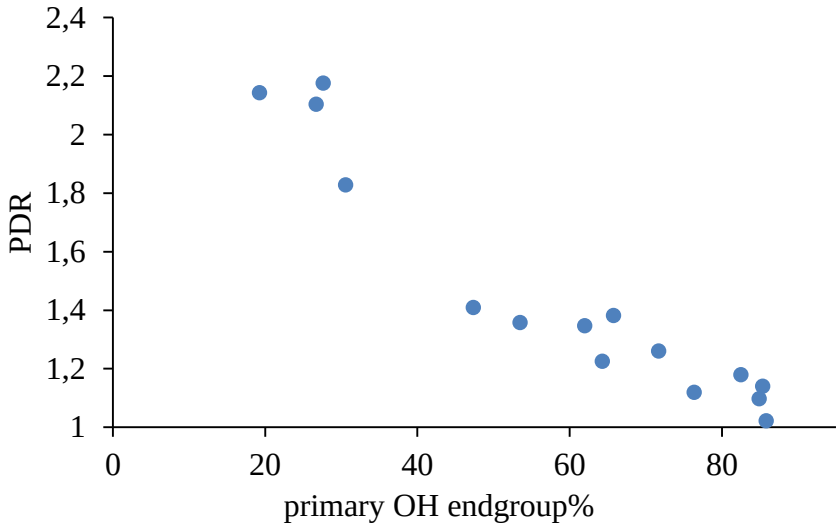


Figure 4: Polydispersity ratio as a function of the percentage of primary hydroxyl groups

4. Possible applications of the results

In today's world, one of the essential tools for precision production is the application of appropriate analytical procedures. In my research, I further developed polymer chemistry analytical procedures. I was the first to create a copolymer mass spectrum analysis algorithm, with which the evaluation of the mass spectrum can be made fast and automatic and the characteristic parameters (M_n, M_w, PDI, n_n^A, n_w^A, PDI^A) can be determined. Furthermore, I introduced a new copolymer specific parameter, the polydispersity ratio (PDR), which helps to understand the slope of the composition drift diagram. In addition, in the case of [poly-(ethylene oxide)]- [poly-(propylene oxide)]- [poly-(ethylene oxide)] block copolymers, the presence of diblock "contaminant" in the copolymer can be detected with this parameter. Furthermore, I developed a simpler, non-programing algorithm to process mass specra that can be easily implemented in spreadsheet software and produces results that are almost as accurate as those produced by the first algorithm. Thus simplifying the analysis, the procedure can also be easily implemented for other copolymers. Moreover, I expanded the set of information that can be extracted from the gel permeation chromatography (GPC) method with simple neural networks. This makes it possible to obtain results with a simple and cheap GPC method, which until now could only be determined by expensive and highly specialized methods. It can be stated that the field of application of the GPC method routinely used in industry has thus become much wider.

Acknowledgment

For the preparation of my thesis, I would like to thank the grants and projects GINOP-2.3.3-15-2016-00021, TKP2020-NKA-04, GINOP-2.3.3-15-2016-00004 and NKFI FK-132385, which are supported by the European Union and were realized with the co-financing of the European Regional Development Fund.

UNIVERSITY of DEBRECEN

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

Registry number: DEENK/8/2024.PL
Subject: PhD Publication List

Candidate: Gergő Róth
Doctoral School: Doctoral School of Chemistry
MTMT ID: 10078039

List of publications related to the dissertation

Foreign language scientific articles in international journals (3)

1. Nagy, T., **Róth, G.**, Benedek, M., Kuki, Á., Timári, I., Zsuga, M., Kéki, S.: Enhanced Copolymer Characterization for Polyethers Using Gel Permeation Chromatography Combined with Artificial Neural Networks.
   Anal. Chem. 95 (28), 10504-10511, 2023. ISSN: 1520-6882.
   DOI: http://dx.doi.org/10.1021/acs.analchem.2c02913
   IF: 7.4 (2022)

2. Kuki, Á., **Róth, G.**, Nagy, A., Zsuga, M., Kéki, S., Nagy, T.: A Short-Cut Data Mining Method for the Mass Spectrometric Characterization of Block Copolymers.
   Processes. 10 (1), 42-50, 2021. EISSN: 2227-9717.
   DOI: http://dx.doi.org/10.3390/pr10010042
   IF: 3.352

3. **Róth, G.**, Nagy, T., Kuki, Á., Hashimov, M., Vonza, Z., Timári, I., Zsuga, M., Kéki, S.: Polydispersity Ratio and Its Application for the Characterization of Poloxamers.
   Macromolecules. 54 (21), 9984-9991, 2021. ISSN: 0024-9297.
   DOI: http://dx.doi.org/10.1021/acs.macromol.1c01552
   IF: 6.057

List of other publications

Foreign language scientific articles in international journals (3)

4. Nagy, T., Kuki, Á., **Róth, G.**, Kyzy, Z. K., Zatalini, A. B., Nyul, D., Zsuga, M., Kéki, S.: Characterization of polyacrylamide diblock copolymers by mass spectrometry combined with Mass-remainder analysis (MARA).
   Polym. Test. 117, 1-7, 2023. ISSN: 0142-9418.
   DOI: http://dx.doi.org/10.1016/j.polymertesting.2022.107834
   IF: 5.1 (2022)

5. Pardi-Tóth, V. C., Kuki, Á., Kordován, M. Á., **Róth, G.**, Nagy, L., Zsuga, M., Nagy, T., Kéki, S.: Molecular data storage using direct analysis in real time (DART) ionization mass spectrometry for decoding.
Sci. Rep. 13 (1), 1-7, 2023. EISSN: 2045-2322.
DOI: http://dx.doi.org/10.1038/s41598-023-43658-x
IF: 4.6 (2022)

6. Nagy, T., **Róth, G.**, Kuki, Á., Zsuga, M., Kéki, S.: Mass Spectral Filtering by Mass-Remainder Analysis (MARA) at High Resolution and Its Application to Metabolite Profiling of Flavonoids.
Int. J. Mol. Sci. 22 (2), 1-12, 2021. ISSN: 1661-6596.
DOI: http://dx.doi.org/10.3390/ijms22020864
IF: 6.208

Hungarian abstracts (13)

7. Pardi-Tóth, V. C., Nagy, T., Kordován, M. Á., **Róth, G.**, Nagy, L., Kuki, Á., Zsuga, M., Kéki, S.: Anyagba integrált adatok dekódolása DART-MS ionforrással = Decoding material-integrated data with DART-MS ion source.
In: XXIX. Nemzetközi Vegyészkonferencia = 29th International Conference on Chemistry. Ed.: Majdik Kornélia, Erdélyi Magyar Műszaki Tudományos Társaság, Kolozsvár, 58, 2023, (ISSN 2734-7109)

8. Kaldybek, K. Z., Nagy, T., **Róth, G.**, Kuki, Á., Zsuga, M., Kéki, S.: Blokk kopolimerek előállítása és karakterizálása = Synthesis and characterization of block copolymer.
In: XXIX. Nemzetközi Vegyészkonferencia = 29th International Conference on Chemistry. Ed.: Majdik Kornélia, Erdélyi Magyar Műszaki Tudományos Társaság, Kolozsvár, 48, 2023, (ISSN 2734-7109)

9. Kordován, M. Á., Nagy, T., Nagy, L., Kuki, Á., **Róth, G.**, Zsuga, M., Kéki, S.: Fém-levegő akkumulátorok formatervezése és 3D nyomtatása = Design and 3D printing of metal-air secondary batteries.
In: XXIX. Nemzetközi Vegyészkonferencia = 29th International Conference on Chemistry. Ed.: Majdik Kornélia, Erdélyi Magyar Műszaki Tudományos Társaság, Kolozsvár, 49, 2023, (ISSN 2734-7109)

10. Benedek, M., Kuki, Á., **Róth, G.**, Nagy, T., Zsuga, M., Kéki, S.: Kopolimerek elemzése gél permeációs kromatográfiával = Kopolimerek elemzése gél permeációs kromatográfiával Analysis of copolymers by gel permeation chromatography.
In: XXIX. Nemzetközi Vegyészkonferencia = 29th International Conference on Chemistry. Ed.: Majdik Kornélia, Erdélyi Magyar Műszaki Tudományos Társaság, Kolozsvár, 45, 2023, (ISSN 2734-7109)

UNIVERSITY AND NATIONAL LIBRARY
UNIVERSITY OF DEBRECEN
H-4002 Egyetem tér 1, Debrecen
Phone: +3652/410-443, email: publikaciok@lib.unideb.hu

11. **Róth, G.**, Nagy, T., Kuki, Á., Pardi-Tóth, V. C., Nyul, D., Kaldybek, K. Z., Palacios, I. A. I., Benedek, M., Zsuga, M., Kéki, S.: Kopolimer MALDI-TOF-MS spektrumok elemzése = Analysis of copolymer MALDI-TOF-MS spectra.
In: XXIX. Nemzetközi Vegyészkonferencia = 29th International Conference on Chemistry. Ed.: Majdik Kornélia, Erdélyi Magyar Műszaki Tudományos Társaság, Kolozsvár, 59, 2023, (ISSN 2734-7109)

12. Nagy, T., Kuki, Á., **Róth, G.**, Zsuga, M., Kéki, S.: Tömegspektrometriával a kopolimerek nyomában = Chasing copolymers by mass spectrometry.
In: XXIX. Nemzetközi Vegyészkonferencia = 29th International Conference on Chemistry. Ed.: Majdik Kornélia, Erdélyi Magyar Műszaki Tudományos Társaság, Kolozsvár, 4, 2023, (ISSN 2734-7109)

13. Nagy, T., Kuki, Á., **Róth, G.**, Kaldybek, K. Z., Zatalini, A. B., Nyul, D., Zsuga, M., Kéki, S.: 4-akrioil-morfolin és N-izopropil-akrilamid blokk kopolimerenek előállítása és karakterizálása.
In: XXVIII. Nemzetközi Vegyészkonferencia = 28th International Conference on Chemistry. Szerk.: Majdik Kornélia, Erdélyi Magyar Műszaki Tudományos Társaság (EMT), Kolozsvár, 35, 2022, (ISSN 2734-7109)

14. **Róth, G.**, Nagy, T., Benedek, M., Kuki, Á., Timári, I., Zsuga, M., Kéki, S.: Neurális hálóval támogatott gélpermeációs kromatográfiás kopolimer analízis.
In: XXVIII. Nemzetközi Vegyészkonferencia = 28th International Conference on Chemistry. Szerk.: Majdik Kornélia, Erdélyi Magyar Műszaki Tudományos Társaság (EMT), Kolozsvár, 40, 2022, (ISSN 2734-7109)

15. Kuki, Á., Nagy, T., Nagy, L., Deák, G., Kordován, M. Á., **Róth, G.**, Zsuga, M., Kéki, S.: Új irányok a környezetbarát fém-levegő akkumulátorok fejlesztésében = Development of novel environmentally friendly metal-air batteries.
In: XXVIII. Nemzetközi Vegyészkonferencia = 28th International Conference on Chemistry. Szerk.: Majdik Kornélia, Erdélyi Magyar Műszaki Tudományos Társaság (EMT), Kolozsvár, 30, 2022, (ISSN 2734-7109)

16. **Róth, G.**, Nagy, T., Kuki, Á., Zsuga, M., Kéki, S.: Biológiai eredetű minták flavonoid tartalmának profilozása tömegmaradék analízissel.
In: XXVII. Nemzetközi Vegyészkonferencia = 27th International Conference on Chemistry. Ed.: Majdik Kornélia, Erdélyi Magyar Műszaki Tudományos Társaság (EMT), Kolozsvár, 43, 2021, (ISSN 2734-7109)

17. Nagy, T., **Róth, G.**, Kuki, Á., Hashimov, M., Timári, I., Vonza, Z., Zsuga, M., Kéki, S.: Polidiszperzitás arány alkalmazása kopolimerek elemzésére.
In: XXVII. Nemzetközi Vegyészkonferencia = 27th International Conference on Chemistry. Ed.: Majdik Kornélia, Erdélyi Magyar Műszaki Tudományos Társaság (EMT), online [Kolozsvár], 47, 2021, (ISSN 2734-7109)

18. **Róth, G.**, Nagy, T., Kuki, Á., Zsuga, M., Kéki, S.: Tömegmaradék analízis (MARA) alkalmazása flavonoidok profilozásra nagy felbontású tömegspektrumok esetén.
In: II. FKF Szimpózium : Absztraktfüzet. Szerk.: Ádám Anna Adél, Ziegenheim Szilveszter, Magyar Kémikusok Egyesülete, Budapest, 15, 2021.

19. **Róth, G.**, Nagy, T., Kuki, Á., Zsuga, M., Kéki, S.: Kopolimer tömegspektrumok automatikus kiértékelése.
In: XXVI. Nemzetközi Vegyészkonferencia = 26th International Conference on Chemistry. Ed.: Majdik Kornélia, Erdélyi Magyar Műszaki Tudományos Társaság (EMT) Kémia Szakosztálya, Kolozsvár, 28, 2020, (ISSN 2734-7109)

Total IF of journals (all publications): 32,717
Total IF of journals (publications related to the dissertation): 16,809

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

11 January, 2024