# Development of a chemometric approach to improve the accuracy of copolymer sequence information obtained from pyrolysis gas-chromatography mass-spectrometry data

Molenaar, S.R.A.; Knol, W.C.; Pirok, B.W.J.; Peters, R.A.H.

**DOI**
10.1016/j.jaap.2023.105978

**Publication date**
2023

**Document Version**
Final published version

**Published in**
Journal of analytical and applied pyrolysis

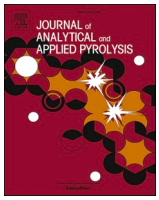
**License**
CC BY

Journal of Analytical and Applied Pyrolysis 171 (2023) 105978

Contents lists available at ScienceDirect

Journal of Analytical and Applied Pyrolysis

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



# Development of a chemometric approach to improve the accuracy of copolymer sequence information obtained from pyrolysis gas-chromatography mass-spectrometry data

Stef R.A. Molenaar [a,b,*,1], Wouter C. Knol [a,b,1], Bob W.J. Pirok [a,b], Ron A.H. Peters [a,b,c]

[a] van 't Hoff Institute for Molecular Sciences, Analytical Chemistry Group, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, the Netherlands
[b] Centre for Analytical Sciences Amsterdam (CASA), the Netherlands
[c] Covestro, Group Innovation, Physics and Material Science, Waalwijk, the Netherlands

ARTICLE INFO

*Keywords:*
Polymer
Sequence length
Pyrolysis-gas chromatography
Chemometrics

ABSTRACT

Number-average copolymer sequence length information can be obtained by pyrolysis-gas chromatography (Py-GC) by comparing the ratios of formed oligomers (*i.e.* dimers and trimers). The formation constants of the oligomers and their detection efficiency are not constant for all fragments, however. This can lead to unrepresentative peak ratios in the chromatogram. In these cases, calibration with an external method (*e.g.* NMR) is required. In this work, we introduce an algorithm that improves the copolymer sequence accuracy yielded from chromatograms with unrepresentative peak areas. The algorithm even functions in cases where oligomer data is missing as the rate of formation of certain oligomers is too low to detect them. One Py-GC measurement and one NMR measurement are required to train the developed algorithm for the determination of average monomer reactivity ratios and relative pyrolysis constants. Afterwards, Py-GC measurements of copolymers containing the same monomers, albeit with different compositions, can be corrected using the previously estimated constants. The algorithm was tested on various styrene-acrylate copolymers, yielding more accurate sequence information, even when limited oligomer information was available.

## 1. Introduction

Pyrolysis-gas chromatography (Py-GC) has been an instrumental tool in copolymer analysis for decades [1,2]. Py-GC is mostly used for the chemical identification of copolymers and their additives, although quantitative insights in various copolymer properties, such as end-groups and chemical composition can also be obtained. Besides these commonly studied copolymer characteristics, Py-GC can also be used to study the copolymer sequence [3]. The copolymer sequence describes the order of different monomers in the copolymer backbone. The sequence of a copolymer can vary from perfectly alternating to random to block. It is of interest since it affects a variety of copolymer properties such as solubility, glass transition temperature and inter particle interactions [4–7]. Tsuge et al. and later Wang et al. performed pioneering work in the field of copolymer sequence determination by Py-GC. They have published oligomer assignments for various copolymers, *e.g.* methyl methacrylate-styrene (MMA-St), butyl acrylate-styrene (BA-St) and methyl acrylate-styrene (MA-St), relating the found oligomer ratios to copolymerization models [3,8–14]. When Py-GC is applied to study the sequence of copolymers, oligomers (*i.e.* dimers or trimers) in the chromatogram are related to the same subunits in the intact copolymer. This is not trivial as both the detection efficiencies of the oligomers and their formation constants in pyrolysis affect the determined sequence value. Therefore, calibration with other methods capable of determining the sequence is required. In practice NMR (often $^{13}$C NMR) is typically the only method capable of providing such a reference value [15,16]. While $^{13}$C NMR might seem like a competitive method to Py-GC, both methods are rather complementary. Where Py-GC lacks in quantitative accuracy, it offers improved sensitivity [17]. Moreover, the chromatographic resolution between oligomer signals is often greater than the spectral resolution yielded by NMR [18]. The poor resolution in NMR often results from additional signal-splitting caused by tacticity effects, which can render spectra too complex to interpret [18].

* Correspondence to: Science Park 904, 1098 XH Amsterdam, the Netherlands.
*E-mail address:* S.R.A.Molenaar@uva.nl (S.R.A. Molenaar).
[1] Equal contribution

https://doi.org/10.1016/j.jaap.2023.105978
Received 18 November 2022; Received in revised form 20 March 2023; Accepted 17 April 2023
Available online 18 April 2023

As mentioned before, for copolymers the pyrolysis rate (*i.e.* fragmentation) and ionization efficiency (in the case of mass spectrometry (MS)) might differ for each formed oligomer. Moreover, the stability of all formed dimers and trimers is not equal. Therefore, fragments are sometimes detected in low quantities or not detected at all. For example, in MMA-St copolymers, MMA-MMA bonds tend to depolymerize completely, yielding a chromatogram where only trimers containing St-St and MMA-St bonds are observed [9,17]. This leads to a chromatogram wherein the observed oligomer peak areas are not representative of the oligomer fractions in the intact polymer. This issue was described fundamentally by Shibasaki and is referred to as the boundary effect [19].

While various authors studied MMA-St copolymers, none have attempted to utilize all the oligomer data present in the chromatogram. Instead, only the ratio between the MMA-St and St-St dimers was used. This means that all other sequence information in the chromatogram is not used. In general, most developments of computer-aided tools to interpret Py-GC data focus on identification and feature selection [20,21]. To the authors knowledge, no computational study has focused on the extraction and correction of sequence information.

Here we introduce an algorithm to improve the accuracy of the yielded sequence and composition information by statistical means. Utilizing a single $^{13}$C NMR measurement of a reference sample, the algorithm estimates average monomer reactivity ratios (MRR), which describe the reference copolymer. The rates that match the NMR results the best are selected. Using the determined MRRs, similar copolymers are generated and pyrolyzed using varying relative pyrolysis constants (RPC) *in silico*. Consequently, the algorithm is able to determine the most likely average RPCs and predict the formation of oligomers to correspond to the measured Py-GC data of the reference sample. Pyrolysis data of other samples of interest can then be translated to a more accurate chemical characterization, using the estimated formation and degradation constants. Even when some dimer or trimer fragments are not detected during pyrolysis, the algorithm is still able to improve the accuracy of the yielded sequence and monomer ratio information. The formation and degradation ratios are specific enough to estimate the sequence for the intact polymer. This tool should render Py-GC a more reliable and easy-to-use alternative in copolymer sequence studies. Hopefully it will serve as a basis for future developments in the field.

## 2. Experimental

### 2.1. Experimental conditions

#### 2.1.1. Samples

Samples used in this work consist of MA-St, ethyl acrylate-styrene (EA-St), BA-St and MMA-St copolymers with various chemical compositions. Table 1 shows the measured molar fraction of styrene by $^{1}$H NMR and the average molecular weight and the molecular weight dispersity measured by size-exclusion chromatography. Sample preparation and analysis methods are described elsewhere [17,22].

#### 2.1.2. Py-GC-MS

Py-GC-MS measurements of the MMA-St were described earlier [17]. In short, a 2010 Shimadzu two-dimensional GC–MS instrument (ˈs-Hertogenbosch, The Netherlands) operated in 1D mode was used for the GC separation. An OPTIC-4 PTV (programmed temperature vaporizer; GL Sciences, Eindhoven, The Netherlands) was used for pyrolysis of the samples. The separation was performed on a DB-5 stationary phase column (30 m × 0.25 mm i.d., 0.25 μm film thickness; Agilent, Waldbronn, Germany). The MA-St, EA-St and BA-St samples were analyzed as described in ref. [22]. In short, a Shimadzu 2010 GC-MS coupled to an OPTIC-3 PTV was used for Py-GC analysis of the synthesized samples. The pyrolysate was separated on a RESTEK (Bellefonte, PA, USA) Rxi-5 ms column (30 m x 0.25 mm ID, 0.25 μm film thickness). Data of both analysis methods was processed using (Shimadzu) Labsolutions version 2.71.

**Table 1**
Chemical properties of studied copolymers, featuring a number, comonomers, styrene fraction, average molecular weight ($M_w$), and molecular weight dispersity ($Đ_M$) for each sample.

| Sample Number | Comonomers | $^{1}$H NMR fraction styrene | $M_w$ (kDa) | $Đ_M$ |
|---|---|---|---|---|
| **1** | MA-St | 0.81 | 73 | 2.1 |
| **2** | MA-St | 0.72 | 74 | 2.2 |
| **3** | MA-St | 0.63 | 84 | 2.2 |
| **4** | MA-St | 0.54 | 86 | 2.3 |
| **5** | MA-St | 0.41 | 81 | 2.3 |
| **6** | EA-St | 0.80 | 58 | 2.2 |
| **7** | EA-St | 0.73 | 70 | 2.3 |
| **8** | EA-St | 0.61 | 73 | 2.3 |
| **9** | EA-St | 0.53 | 78 | 2.3 |
| **10** | EA-St | 0.40 | 95 | 2.4 |
| **11** | BA-St | 0.85 | 60 | 2.2 |
| **12** | BA-St | 0.77 | 70 | 2.3 |
| **13** | BA-St | 0.62 | 76 | 2.1 |
| **14** | BA-St | 0.50 | 82 | 2.3 |
| **15** | BA-St | 0.42 | 90 | 2.4 |
| **16** | MMA-St | 0.26 | 39 | 1.6 |
| **17** | MMA-St | 0.46 | 34 | 1.6 |
| **18** | MMA-St | 0.67 | 27 | 1.7 |

#### 2.1.3. NMR

NMR measurements were performed as described in ref. [17]. In short, proton decoupled $^{13}$C NMR spectra were recorded in $CDCl_3$ (Sigma Aldrich, Darmstadt, Germany) Bruker (Rheinstetten, Germany) Avance NEO 500 MHz NMR spectrometer at 298 K. 8192 Scans were recorded with a pulse delay of 4 s $^{1}$H NMR were recorded on the same instrument at the same conditions instead 32 scans were recorded with a pulse delay of 5 s. NMR data was analyzed using TopSpin version 4.1.1 (Bruker).

### 2.2. Data processing

The entire algorithm was written using MATLAB 2020b (Mathworks, Natick, MA, USA) for the in-house MOREDISTRIBUTIONS program [23,24]. Core functions to fit parameters make use of the build-in functions *ga* and *fmincon* within MATLAB.

## 3. Results and discussion

### 3.1. Creating a copolymer distribution

Considering that simulating copolymer degradation *in silico* requires an accurate representation of the chemical distribution, an understanding of the formation of the copolymer is needed on which further steps can be build. An accurate representation of the sequence of a copolymer must therefore be determined. Consequently, the first step within the workflow was fitting the MRR for each monomer to the sequence information obtained from $^{13}$C NMR data.

The fraction of monomer A ($F_A$) in the copolymer can be calculated from the $^{13}$C NMR results by dividing the number-average sequence length of monomer A ($\overline{n}_A$) by the sum of $\overline{n}_A$ and $\overline{n}_B$ (Eq. 1) [9]. Eqs. 2 and 3, can be used to describe the number-average sequence length, where AAA, AAB, BAB, BBB, BBA and ABA represent the peak areas of the corresponding trimers.

$$F_A = \frac{\overline{n}_A}{\overline{n}_A + \overline{n}_B} \quad (1)$$

$$\overline{n}_A = \frac{AAA + AAB + BAB}{0.5 * AAB + BAB} \quad (2)$$

$$\overline{n}_B = \frac{BBB + BBA + ABA}{0.5 * BBA + ABA} \quad (3)$$

While the average ratios, or the likelihood, of the addition of monomer A or B can be calculated indirectly post-synthesis, it is impossible to determine the exact reaction rates in terms of reaction speed post-synthesis. Utilizing Markov equations [25–27] the probability of adding monomer A or B to the current chain can be calculated. To accomplish this, a number of copolymer chains were initiated with two monomers. The copolymer chain then was grown by the addition of monomers until the maximum allowed chain length (*e.g.* 1000 monomers) was achieved. As long as the *in silico* polymerization is not terminated, the probability of the addition of monomer A or B can be described by Eq. 4 & 5, and the probability of adding any monomer (*i.e.* monomer A or B) to the chain is equal to 1 (Eq. 6).

$$P_{YZA} = \frac{N_A}{N_A + N_B * r_{YZB}} \quad (4)$$

$$P_{YZB} = \frac{N_B}{N_A * r_{YZA} + N_B} \quad (5)$$

$$P_{YZA} + P_{YZB} = 1 \quad (6)$$

Here $P_{YZA}$ and $P_{YZB}$ are the probabilities of adding monomer A and B to a YZ terminated chain, respectively. Y and Z can both represent either monomer A or monomer B. $N_A$ and $N_B$ are the number of monomers of A and B and $r_{YZB}$ is the MRR of adding monomer B over monomer A to a YZ terminated chain, defined as $r_{YZB} = \frac{k_{YZB}}{k_{YZA}}$. Likewise, $r_{YZA}$ is the MRR of adding monomer A over monomer B to a YZ terminated copolymer chain, defined as $r_{YZA} = \frac{k_{YZA}}{k_{YZB}}$ with $k_{YZA}$ and $k_{YZB}$ being the reaction rates of adding the corresponding monomer to a YZ terminated chain. Note that $P_{YZA}$ and $P_{YZB}$ are related by Eq. 6 and therefore, only one probability needs to be calculated to know the other. Furthermore, $r_{YZA}$ and $r_{YZB}$ are related to each other by $r_{YZA} = \frac{1}{r_{YZB}}$.

The MRRs depend on the monomers at the reactive site of the chain. The monomers closest to the reactive site have the largest impact on the MRR. Therefore, the last two monomers (previously indicated by YZ) in the chain were taken into account, this is known as the penultimate model [27]. This means that in total four different MRRs, one for each unique penultimate unit, were used within the algorithm: $r_{AAB}$, $r_{ABB}$, $r_{BAB}$, $r_{BBB}$. Where $r_{AAB}$ (defined as $r_{AAB} = \frac{k_{AAB}}{k_{AAA}}$) is the MRR of adding monomer B over monomer A when AA in the penultimate unit, $r_{ABB}$ is the MRR of adding monomer B over monomer A when AB is the penultimate unit, and so forth. The determined MRRs provide an average value for the formation of the copolymer. Composition drift is therefore not taken into account.

The resulting simulated copolymer distributions need to be statistically viable. A distribution of a 1000 copolymer chains, with an average length of 1000 monomers and a standard deviation of 100 monomers were thus created, to yield an adequate sample size. Initial tests proved no gain in accuracy when generating additional or longer chains, as an average was calculated. When the fraction of monomer A is determined utilizing the NMR results (Eq. 4) and the length of each copolymer chain is determined ($N_{tot}$), the number of required A and B monomers can be calculated by $N_A = \lfloor F_A * N_{tot} \rceil$ and $N_B = N_{tot} - N_A$. Due to a lower sensitivity and poorly resolved trimer signals [28], $^{13}$C NMR often results in less accurate composition information based on the determined number-average sequence length (Eqs. 1–3) compared to $^{1}$H NMR. Therefore, if a more accurate monomer fraction is known, the user is free to input this value when calculating the needed number of monomers. Now that the total amount of monomers A and B were known, each polymer chain was initiated by two monomers based on their relative abundance (*i.e.* $P_A = \frac{N_A}{N_A + N_B}$) and the number of monomers was adjusted accordingly by subtracting the used monomers from the totals. Afterwards, the algorithm continued to determine the last two monomers (YZ) and generate random values between 0 and 1. Then, the generated value was compared to the probability of adding monomer A to the YZ terminated chain according to Eq. 4. If the random value was lower than the probability, monomer A was added to chain. And if the random value was higher than the probability of adding monomer A, monomer B was added to the copolymer chain. After the addition of a monomer, the number of monomers were again adjusted accordingly by subtracting the used monomer from the totals. Each polymer chain was grown at the same time, in contrast to making each polymer chain individually before creating the next one. This ensured each created polymer chain consisted of a representative copolymer even in the unlikely case where the initial concentrations are heavily skewed.

Using the copolymer formation as described above, a genetic algorithm was utilized to estimate the MRRs. A genetic algorithm was chosen because genetic algorithms can handle variation robustly [29]. While the genetic algorithm is being trained, random parameters are selected and evaluated. The repeated analysis of the same data might thus produce very minor differences, which may cause linear optimization algorithms to search in a wrong direction. Furthermore, genetic algorithms are less likely to converge to local minima and can be performed in parallel. By varying the MRRs, the genetic algorithm minimized the difference between the created copolymers and the $^{13}$C NMR results. This was done by comparing the ratios between the A-centered trimers (AAA, AAB/BAA, BAB) and the B-centered trimers (BBB, BBA/ABB, ABA). This originates from $^{13}$C NMR where the splitting of the signal of different carbon atoms is examined. The response of these observed carbon-atoms cannot be assumed to be equal. Therefore, pooling all trimers data gives rise to less accurate results. The genetic algorithm was set up with an initial population of 20 of randomly selected values between 0 and 10, a lower boundary of 0 and no upper boundary.

### 3.2. Simulated pyrolysis of copolymer distributions

To correct Py-GC data to more representative sequence information, it is of high importance to understand how the thermal degradation of a copolymer takes place. Not all oligomers are formed at the same frequency, due to the stability of intermediates during radical transfer and the stability of the created fragments. In literature, thermal degradation mechanisms of (co)polymers have been described frequently [30–32]. It is presumed that the process is initiated with a random chain scission. This generates two different fragments which both contain a radical, one of the fragments contains a methylene radical the other is a substituted radical (See Fig. 1). The random scission is then followed by three dominant degradation pathways after the initiation. The fragments can undergo depolymerization (*i.e.* forming monomers) by a 2,1-radical transfer, which is the most dominant process in the studied systems. Or the fragments could undergo back biting by a 1,5-radical transfer, forming a trimer. In older literature, the possibility of β-scission is described where a 1,5-radical transfer leads to the formation of a monomer and a monomer radical [30]. This monomer radical can then interact with a polymer chain, forming a dimer and transferring the radical back to the polymer chain. This pathway dilutes the original sequence within the copolymer, as randomization is introduced. A more recent insight is that dimers are more likely formed by a 1,7-radical transfer, followed by a 7,3-radical transfer [32], although the relative rates are temperature dependent. A preliminary study indicated that the monomer radical pathway is not significant for the used statistical copolymers, as is shown in Supplementary Materials Section S-1. Therefore, only the 1,7-radical transfer pathway was considered in this work when pyrolyzing the created distributions. When block-copolymer or polymer blends are used, the monomer-radical pathway might need to be considered, however.

It is hypothesized that the formation of monomers (depolymerization) is mostly dependent on the last two monomers, whereas the backbiting constant for the 1,5-radical transfer (*i.e.* trimer formation) is

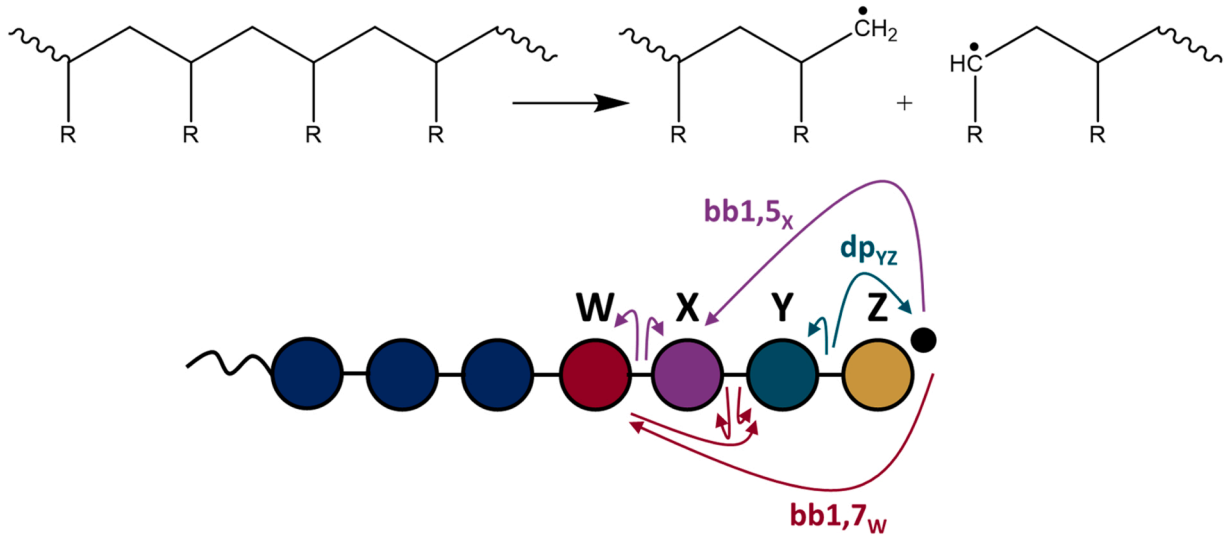



**Fig. 1.** Top) Random scission creating two different radical fragments. Bottom) Indication of monomer locations and the three most dominant degradation pathways with respect to the radical (Black dot on the right). With Z the location of the monomer containing the radical, Y the second position, X the third position and W the fourth position.

mostly dependent on the monomer on the third position and the backbiting constant for dimer formation, or the 1,7-radical transfer, is mostly dependent on the monomer on the fourth position. Which means that for the degradation mechanism $dp_{AA}$, $dp_{AB}$, $dp_{BA}$, $dp_{BB}$, $bb1,5_A$, $bb1,5_B$, $bb1,7_A$ and $bb1,7_B$ are the needed pyrolysis constants. Here $dp_{YZ}$ is the depolymerization constant for any combination of two monomers at position Z (first) and Y (second), $bb1,5_X$ is the 1,5-radical transfer constant and $bb1,7_W$ is the constant for a 1,7-radical transfer with the monomer present at position X (third) and W (fourth) respectively. Where W, X, Y, Z indicate the monomers on the corresponding positions with position Z being the location of the radical, as shown in Fig. 1.

Furthermore, fragments formed by pyrolysis don't have the same ionization efficiencies in MS detection, thus the following equations predict, in essence, the probability of forming monomers and oligomers multiplied by their relative detector response. The probability of detecting each variation is described by Eqs. 7 to 9.

$$P_{mono} = \frac{dp_{YZ}}{bb1,7_W + bb1,5_X + dp_{YZ}} \tag{7}$$

$$P_{di} = \frac{bb1,7_W}{bb1,7_W + bb1,5_X + dp_{YZ}} \tag{8}$$

$$P_{tri} = \frac{bb1,5_X}{bb1,7_W + bb1,5_X + dp_{YZ}} \tag{9}$$

Where $P_{mono}$, $P_{di}$ and $P_{tri}$ are the probabilities of forming a monomer, dimer or trimer respectively and $P_{mono} + P_{di} + P_{tri} = 1$.

Since the MRRs were estimated in the previous step, the algorithm was able to simulate realistic polymer distributions similar to the NMR results. These polymer distributions could then undergo simulated pyrolysis. By utilizing another genetic algorithm, the average RPCs were estimated. This was done by creating polymer distributions using the previous estimated MRRs. Subsequently, these distribution are pyrolyzed with varied pyrolysis constants until the differences between the ratios of dimers, MA-centered trimers and St-centered trimers of the calculated fragments and the pyrolysis data were minimized.

Given that these pyrolysis constants are effectively ratios of one another, $dp_{AA}$ was set to 1 by default. Meaning that seven likelihoods of forming other fragments compared to the A monomer formation were fitted for the two different radical fragments (14 in total). The genetic algorithm was set up with an initial population of 200 of randomly selected values between 0 and 10, a lower boundary of 0 and no upper boundary. Furthermore, to evaluate if both different radical fragments need to have their own set of RPCs, a second genetic algorithm was set up. This optimization used only one set of RPCs and thus required a smaller initial population of 20.

Fig. 2 shows the results of a MA-St copolymer for both cases. The adjusted ratios between dimers, St-centered trimers and MA-centered trimers after the genetic algorithm estimated the RPCs *versus* the case where each fragment was as likely to form (*i.e.* all MRRs and RPCs equal 1). A clear improvement in the capability of predicting the trimer ratios is shown compared to all MRRs and RPCs equal to 1. The pyrolysis and $^{13}C$ NMR results show that the copolymer has an alternating tendency. However, the fragments predicted when each fragment is equally likely to form overestimate the homo-trimers and underestimates the alternating trimers. Furthermore, the St-St-MA trimer fraction is overestimated by the pyrolysis results (and thus St-St-St and MA-St-MA are underestimated), compared to the $^{13}C$ NMR results. Note that the MA-MA dimer is not detected as it degrades completely [33], it is therefore not shown. When using two different sets of RPCs the St-centered trimer ratios is predicted more accurate than when using only one set of RPCs. Otherwise, there were no other significant difference seen between using two sets of RPCs and only using one RPCs set.

### 3.3. Fitting composition and randomness of a new sample

When the formation and thermal degradation of a copolymer is understood, and constants such as the MRR and the RPC are calculated, the gained understanding can be utilized on copolymer systems with identical monomers. This will enable accurate sequence determination by Py-GC, rendering sequence determination more routinely applicable as it is much faster and less expensive than NMR.

The found Py-GC areas of each detected dimer and trimer of a new sample was used as input in the following step. By varying the fraction of monomer A using the *fmincon* function within MATLAB, the algorithm generated a copolymer distribution and simulated pyrolysis of this distribution according to the previously calculated constants. The *fmincon* function then minimized the differences between the calculated fragments and the new input data. When a minimum was found, the algorithm returned the found monomer fractions, the randomness of the copolymer and the number-average sequence length of monomer A and B.

Note that the fitting of MRR and RPC has to be performed only once.

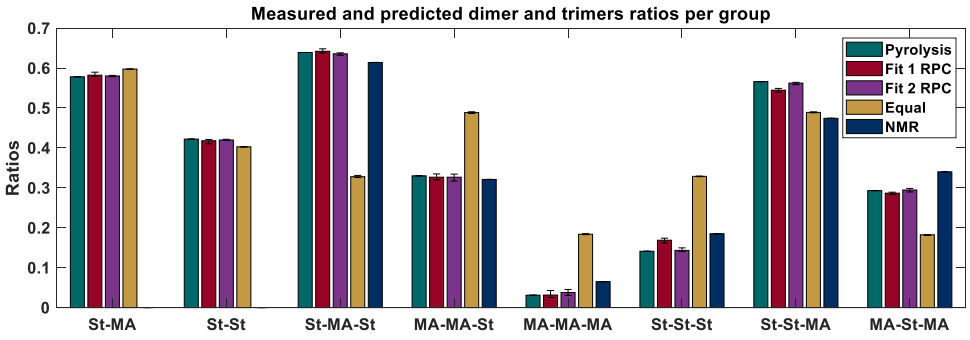



**Fig. 2.** Bar plots showing the ratios between dimers, methyl acrylate-centered trimers and styrene-centered trimers as measured (Green), predicted by the algorithm after the calculation of one set of RPCs (Red), predicted by the algorithm after the calculation of two sets of RPCs (Purple), when all MRR and RPC are assumed equal (Yellow) and as measured by $^{13}$C NMR (Blue). Error bars are shown over five different calculations. MA-MA dimers are not detected and therefore not shown.

The obtained constants can subsequently improve the sequence determination of any number of new samples, given that the pyrolysis is performed under the same conditions. By creating an in-house database for each combination of monomers of interest, the first two steps of the algorithm are only needed once. Here it assumed the pyrolysis process remains constant to a certain degree. If the experimental conditions significantly change, the RPC might need to be updated.

The overall designed workflow is shown in Fig. 3. Supplementary Material Section S-2 shows detailed figures explaining each, previously explained, individual step. In brief, using NMR and pyrolysis data from one copolymer sample, MRR and RPC are calculated. Subsequently, the adjusted composition of a copolymer, consisting of the same monomers, can be calculated utilizing the formerly calculated MRR, RPC, and the dimer and trimer information of new pyrolysis data. In essence, calibration parameters are established on a single sample of which sequence data (from $^{13}$C NMR) is available, which can then be applied to interpret Py-GC data of any sample consisting of identical monomers.

### 3.4. Robustness of the algorithm

Firstly, copolymer Sample 3 (Table 1) was used to fit MRR and RPC, using two sets of RPC and using one set of RPC five times. Then, four other MA-St copolymers (Samples 1, 2, 4 and 5, Table 1) underwent pyrolysis, and the data was used as input for the last part of the algorithm. The four samples were made with different synthesis recipes. In all samples the MA-MA dimer was not detected. For Sample 1, the MA-trimer was also not detected. This was to be expected the sample contains little MA. No accurate composition could be estimated for this sample, using conventional methods, due to the missing MA trimer signal. Additional $^{13}$C NMR measurements were performed on Samples 2 and 4 as an external validation to compare calculated trimer fractions with the experimental data.

When using two different sets of RPCs, the algorithm predicted a styrene fraction of $0.75 \pm 0.02$, $0.66 \pm 0.01$, $0.47 \pm 0.01$ and $0.36 \pm 0.01$ for sample 1, 2, 4 and 5 respectively. When fitting only one set of RPCs, the styrene fractions were predicted as $0.72 \pm 0.01$, $0.65 \pm 0.01$, $0.49 \pm 0.01$ and $0.36 \pm 0.01$. When comparing these values to the styrene fractions determined with $^{1}$H NMR of 0.81, 0.72, 0.54 and 0.41, there was no significant increase of accuracy when using two different sets of RPCs. Therefore, it was chosen to continue with only one set of RPCs, since the genetic algorithm needs less computational resources when less variables are determined.

As most functions used within the algorithm were probability based, the robustness of the developed algorithm was of high importance. Therefore, the full algorithm was initiated 50 times with only the $^{13}$C NMR data and an additional 50 times with an addition of the monomer fraction as determined by $^{1}$H NMR. Styrene fractions as determined by NMR, pyrolysis and as estimated by the algorithm are shown in Table 2. When the initial fractions by $^{1}$H NMR are included, the comparison of calculated monomer fractions to the fraction determined by $^{1}$H NMR improved, as expected. However, the estimation of the number-average sequence length, and thus the sequence information, worsened. Depending on the needs of the user, the monomer fraction obtained by $^{1}$H NMR may or may not be included for a fair comparison between samples.

Even though the number-average sequence lengths of both monomers were in good agreement with the $^{13}$C NMR results, the trimer ratios provided a different picture. Comparing the Py-GC-MS trimer fractions to the fractions obtained by $^{13}$C NMR, the trimer ratios differ by factors between 0.56 and 2.06 for Sample 2 and between 0.67 and 1.89 for Sample 4. The trimer ratios of the predicted results however, only show difference factors between 0.83 and 1.11 for Sample 2 and between 0.70 and 1.21 for Sample 4. The calculated and measured trimer fractions of Sample 2 are shown in Fig. 4. Here a clear improvement in trimer

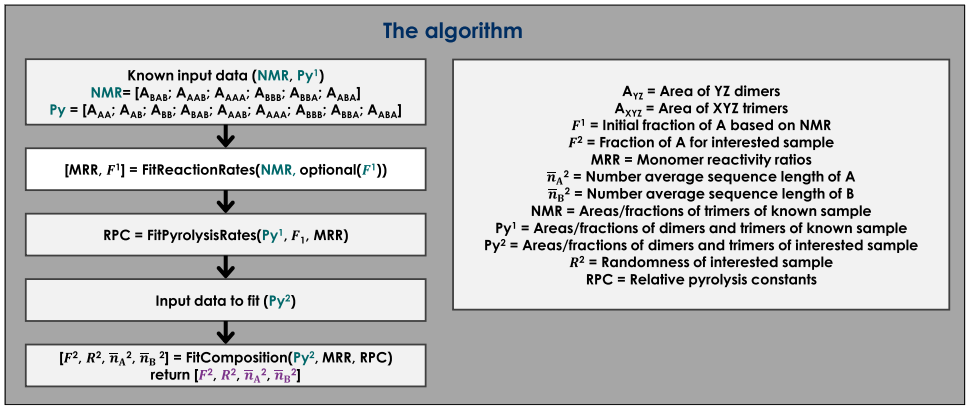



**Fig. 3.** Simplified flowchart of the algorithm. Input variables are indicated in green and output variables are indicated in purple.

**Table 2**
Styrene fractions within the copolymer samples by $^{13}$C- and $^{1}$H NMR, pyrolysis and the mean estimate by the algorithm. Standard deviations for the algorithm's fraction estimate over 50 calculations are also given. The numbers followed by an asterix (*) show a value that is calculated with missing values and are therefore not accurate.

| | Sample 1 | Sample 2 | Sample 3 | Sample 4 | Sample 5 |
|---|---|---|---|---|---|
| **$^{13}$C NMR: $F_{St}$** | — | 0.65 | 0.57 | 0.48 | — |
| **$^{13}$C NMR: $\overline{n}_{St}$ \| $\overline{n}_{MA}$** | — | 2.22 \| 1.19 | 1.73 \| 1.29 | 1.39 \| 1.49 | — |
| **$^{1}$H NMR: $F_{St}$** | 0.81 | 0.72 | 0.63 | 0.54 | 0.41. |
| **Pyrolysis: $F_{St}$** | 0.74 * | 0.66 | 0.58 | 0.50 | 0.32 |
| **Pyrolysis: $\overline{n}_{St}$ \| $\overline{n}_{MA}$** | 3.09 \| 1.06 * | 2.19 \| 1.13 | 1.74 \| 1.24 | 1.46 \| 1.43 | 1.28 \| 1.87 |
| **Algorithm $^{13}$C NMR only: $F_{St}$** | 0.73 ± 0.02 | 0.66 ± 0.01 | Input | 0.49 ± 0.01 | 0.38 ± 0.02 |
| **Algorithm $^{13}$C NMR only: $\overline{n}_{St}$ \| $\overline{n}_{MA}$** | 3.01 \| 1.09 | 2.22 \| 1.15 | Input | 1.49 \| 1.52 | 1.28 \| 2.06 |
| **Algorithm $^{1}$H NMR added: $F_{St}$** | 0.77 ± 0.04 | 0.72 ± 0.04 | Input | 0.55 ± 0.08 | 0.40 ± 0.10 |
| **Algorithm $^{1}$H NMR added: $\overline{n}_{St}$ \| $\overline{n}_{MA}$** | 3.90 \| 1.14 | 3.07 \| 1.20 | Input | 2.11 \| 1.72 | 1.75 \| 2.70 |

fraction can be seen compared to the pyrolysis results. Exact values for both samples and a calibration figure for Sample 4 can be found in Supplementary Material Section S-3.

Fig. 4 shows one million different fraction possibilities *versus* the randomness ($R$, see Eq. 10) for a 65% styrene MA-St copolymer (*i.e.* Sample 2). The found fraction by pyrolysis, the algorithm and $^{13}$C NMR for validation are also shown on this figure. Here it is clear that some found values are statistically impossible within the pyrolysis data. For example, the St-MA-St trimer fraction is not within the calibration range. The $^{13}$C NMR results of the same sample show a completely different fraction composition, that does fall within the calibration areas and where the order of relative abundance of some trimers (*e.g.* MA-MA-MA *versus* St-St-MA or MA-MA-St *versus* MA-St-MA) are exchanged. The algorithm was able to predict trimer fractions for Sample 2 in close agreement with the $^{13}$C NMR data used for external validation.

$$R = \frac{1}{\overline{n}_A} + \frac{1}{\overline{n}_B} \tag{10}$$

### 3.5. Algorithms performance with missing values

After the precision of the algorithm had been estimated, the algorithm performance with missing values was tested. In the previous calculations the MA-MA dimer was missing within all pyrolysis data and, in the case of Sample 1, the MA-MA-MA trimer was also missing. Now the algorithm was tested when other variables were missing as well. Fig. 5 shows the predicted values compared to $^{1}$H NMR of Samples 1, 2, 4 and 5, again using Sample 3 as input. A table with means and standard deviation can be found in Supplementary Material Section S-4. I1 means that one additional randomly selected input variable was missing from the data when the RPCs were being fitted. I2, I3 and I4 mean that two, three and four additional input variables were left out. The rows coded C1 to C4 are missing one to four additional datapoints for the compounds pyrolysis data that needed to be adjusted. I1C1 to I4C4 mean that both the input sample and the tested sample were missing 1–4 additional datapoints.

No clear trend can be seen in the prediction errors, indicating that the number of missing values has no distinct impact in the algorithm's

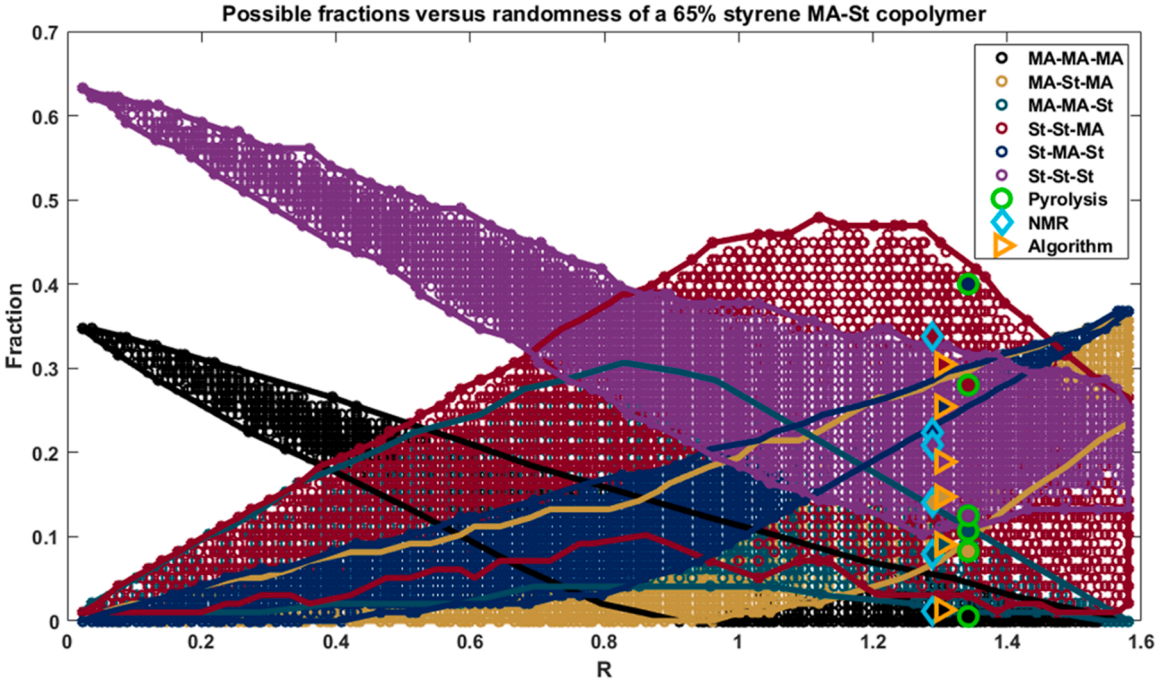



**Fig. 4.** One million different fraction possibilities of a 65% styrene MA-St copolymer (Sample 2) *versus* the randomness. Contours of each trimer fractions are provided. Green circles represent the found pyrolysis fractions, blue diamonds represent the found $^{13}$C NMR fractions for external validation and orange right-pointing triangle represent the fractions as predicted by the algorithm based on the pyrolysis data. Pyrolysis, $^{13}$C NMR and predicted fractions are filled in with the corresponding color matching the trimers.

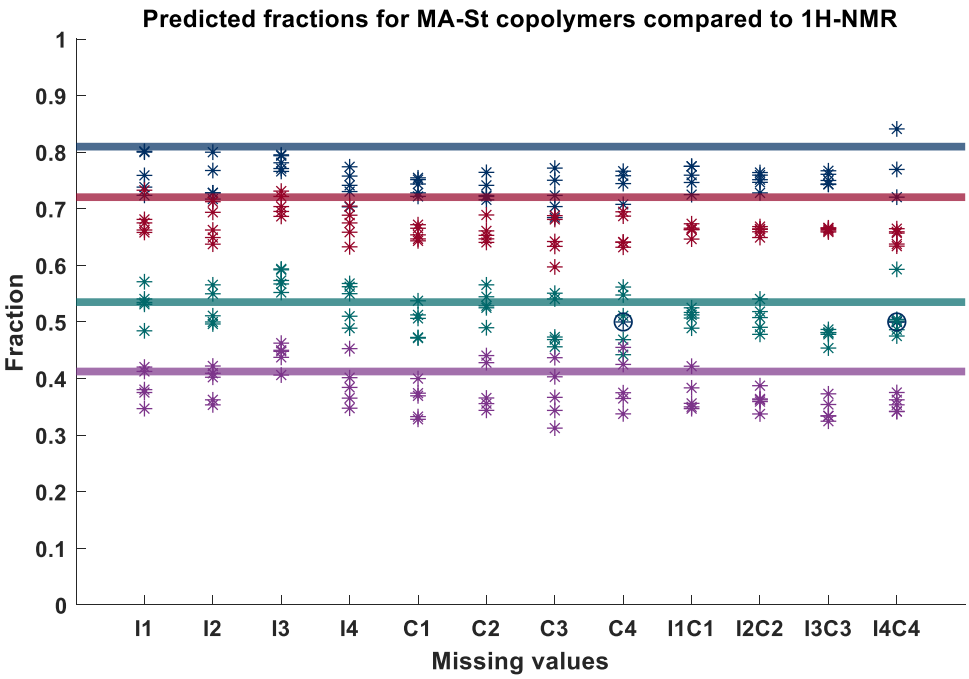



**Fig. 5.** Predicted styrene fractions when additional datapoint are left out of the data set. Each asterix is a predicted value. Lines indicate the measured styrene fractions by $^{1}$H NMR. Colors indicate different samples: blue indicates Sample 1, red indicates Sample 2, green indicates Sample 4 and purple indicates Sample 5. IX means that X input values from the fit were left out, CX means that X values from the compound of interest were left out. The number indicated how many values were randomly selected. Two blue circles are placed at 'C4, 0.5′ and 'I4C4, 0.5′. These values were unable to be calculated as too many datapoints were missing.

performance, however Supplementary Material Section S-4 shows the exact values that were left out for each calculation and the predicted values. When looking at individual calculations, it becomes clear that the algorithm calculates wrong values when multiple values from the same centered monomer are missing (*e.g.* all A-centered trimers: AAA, AAB and BAB). The algorithm normalizes in three groups; dimers, A-centered trimers and B-centered trimers. When multiple values from the same group were missing, the accuracy of the algorithm went down. In groups C4 and I4C4 for Sample 1, there were too many values missing as with this sample the MA-MA-MA trimer was not detected. Meaning that sometimes effectively six values were missing from the fit and if these missing values were spread-out over each group, the algorithm could not calculate a value. If only one value in each group was present, no ratios could be calculated and the algorithm predicted a fraction of 0.5, which is the starting point from the *fmincon* function. To illustrate, dimer ratios are calculated using AA, AB and BB. A-centered trimer ratios are calculated using AAA, AAB and BAB and the B-centered trimer ratios are calculated using BBB, BBA and ABA. If from each group only one value can be determined, for example AB, BAB and BBB, no ratios can be calculated per group and the algorithm fails.

### 3.6. Algorithms performance with ethyl acrylate-styrene and butyl acrylate-styrene

The algorithm was further tested on EA-St and BA-St copolymers. Within these samples the discrepancy between pyrolysis, $^{13}$C NMR and $^{1}$H NMR were larger than with the MA-St samples. Table 3 below shows the input values for the two samples (Sample 8 for EA-St and Sample 13 for BA-St). Here it can be seen that $F_{St}$ as obtained by $^{1}$H- and $^{13}$C NMR vary 0.09 and 0.08. When the algorithm is initiated without an initial fraction by $^{1}$H NMR, the discrepancy between the calculated fraction and the fraction by $^{1}$H NMR, also varied on average 0.076, as can be seen in Table 4. When the algorithm was initiated with an initial fraction by $^{1}$H NMR however, the error between calculated and experimental fraction was reduced to an average of 0.031. The first EA-St sample (Sample 6) is accountable for the largest error of 0.08. This is the only sample where the EA trimer wasn't detected, which could explain the large discrepancy.

**Table 3**
Fractions, randomness and number-average sequence lengths provided by pyrolysis and NMR for the input Samples 8 (top) and 13 (bottom).

| | Pyrolysis | | | | $^{13}$C NMR | | | | $^{1}$H NMR |
|---|---|---|---|---|---|---|---|---|---|
| # | $F_{St}$ | $R$ | $\overline{n}_{St}$ | $\overline{n}_{EA}$ | $F_{St}$ | $R$ | $\overline{n}_{St}$ | $\overline{n}_{EA}$ | $F_{St}$ |
| **8** | 0.63 | 1.36 | 1.98 | 1.16 | 0.52 | 1.47 | 1.42 | 1.31 | 0.61 |
| # | $F_{St}$ | $R$ | $\overline{n}_{St}$ | $\overline{n}_{BA}$ | $F_{St}$ | $R$ | $\overline{n}_{St}$ | $\overline{n}_{BA}$ | $F_{St}$ |
| **13** | 0.66 | 1.22 | 2.43 | 1.24 | 0.54 | 1.48 | 1.47 | 1.24 | 0.62 |

### 3.7. Algorithms performance with methyl methacrylate-styrene

Lastly, the algorithm was initiated with data from a MMA-St sample. Any dimer or trimer containing two or more MMA monomers (MM, MMS/SMM, MSM & MMM, in which S stands for St and M stands for MMA) was not stable within the Py-GC-MS set-up and therefore were not detectable. This created extremely flawed results of the MMA-St samples of interest, as can be seen in Table 5. The algorithm could only compare two values with each other. There was only one M-centered trimer available and thus this ratio could not be used. The two remaining values, the dimer ratios and the St-centered trimer ratios were skewed since they both missed a value (MM and MSM respectively). Nevertheless, even though only five out of nine values were provided in the pyrolysis data (MS, SS, SSM/MSS, SMS & SSS) and used as input, the algorithm still provided values very similar to $^{13}$C NMR.

A preliminary study indicates that the margin of error increased when using the fitted RPCs on data measured on a different instrument. The results of this study are shown in Supplementary Material Section S-5. The algorithm was initiated to predict the $F_{St}$ for Samples 16 and 18. Both the predicted $F_{St}$ differ 0.06 when comparing to the fraction given by $^{13}$C NMR. The results in Table 5 however, differ only 0.04 and or don't even have a (significant) difference. First fitting the RPCs for the new system and then predicting the $F_{St}$ reduced the errors from 0.06 to 0.04 and 0.03. Therefore, it is recommended that RPCs are calculated for each system and that the experimental settings are kept constant. Moreover, the system should be well-maintained and/or recalibrated regularly. The best results would most likely be obtained when one known sample is measured within a sequence, to make sure the RPCs for the developed chemometric approach remain up-to-date.

**Table 4**
Fractions, randomness and number-average sequence lengths provided by pyrolysis, $^{1}$H NMR and the calculated values by the algorithm for the tested EA-St and BA-St samples. Numbers followed by an asterix (*) show values that are calculated with missing values and are therefore not accurate.

| | Pyrolysis | | | | Algorithm with $^{1}$H | | | | $^{1}$H NMR | Algorithm without $^{1}$H | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # | $F_{St}$ | $R$ | $\overline{n}_{St}$ | $\overline{n}_{EA}$ | $F_{St}$ | $R$ | $\overline{n}_{St}$ | $\overline{n}_{EA}$ | $F_{St}$ | $F_{St}$ | $R$ | $\overline{n}_{St}$ | $\overline{n}_{EA}$ |
| **6** | 0.83 * | 1.16 * | 5.20 | 1.04 * | 0.88 | 1.09 | 7.31 | 1.05 | 0.80 | 0.75 | 1.25 | 3.20 | 1.07 |
| **7** | 0.72 | 1.26 | 2.87 | 1.10 | 0.71 | 1.26 | 2.74 | 1.12 | 0.73 | 0.62 | 1.41 | 1.89 | 1.14 |
| **9** | 0.54 | 1.41 | 1.55 | 1.31 | 0.52 | 1.50 | 1.38 | 1.29 | 0.53 | 0.46 | 1.45 | 1.28 | 1.51 |
| **10** | 0.44 | 1.34 | 1.33 | 1.71 | 0.43 | 1.49 | 1.17 | 1.58 | 0.40 | 0.37 | 1.38 | 1.16 | 1.94 |
| # | $F_{St}$ | $R$ | $\overline{n}_{St}$ | $\overline{n}_{BA}$ | $F_{St}$ | $R$ | $\overline{n}_{St}$ | $\overline{n}_{BA}$ | $F_{St}$ | $F_{St}$ | $R$ | $\overline{n}_{St}$ | $\overline{n}_{BA}$ |
| **11** | 0.86 | 1.01 | 6.87 | 1.16 | 0.85 | 1.13 | 5.79 | 1.05 | 0.85 | 0.80 | 1.20 | 4.22 | 1.04 |
| **12** | 0.75 | 1.20 | 3.39 | 1.11 | 0.71 | 1.27 | 2.72 | 1.10 | 0.77 | 0.62 | 1.44 | 1.81 | 1.13 |
| **14** | 0.54 | 1.41 | 1.54 | 1.32 | 0.52 | 1.45 | 1.43 | 1.33 | 0.50 | 0.45 | 1.44 | 1.26 | 1.56 |
| **15** | 0.42 | 1.44 | 1.21 | 1.64 | 0.45 | 1.43 | 1.28 | 1.54 | 0.42 | 0.32 | 1.32 | 1.12 | 2.36 |

**Table 5**
Fractions, randomness and number-average sequence lengths provided by pyrolysis, $^{13}$C NMR and the calculated values by the algorithm for the tested MMA-St samples. Numbers followed by an asterix (*) show values that are calculated with missing values and are therefore not accurate.

| | Pyrolysis | | | | $^{13}$C NMR | | | | Algorithm | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # | $F_{St}$ | $R$ | $\overline{n}_{St}$ | $\overline{n}_{MMA}$ | $F_{St}$ | $R$ | $\overline{n}_{St}$ | $\overline{n}_{MMA}$ | $F_{St}$ | $R$ | $\overline{n}_{St}$ | $\overline{n}_{MMA}$ |
| **16** | 0.70 * | 1.43 * | 2.33 * | 1.00 * | 0.29 | 1.20 | 1.18 | 2.87 | 0.29 | 1.21 | 1.16 | 2.88 |
| **17** | 0.74 * | 1.32 * | 3.17 * | 1.00 * | 0.45 | 1.29 | 1.41 | 1.72 | Input | Input | Input | Input |
| **18** | 0.84 * | 1.21 * | 4.81 * | 1.00 * | 0.61 | 1.08 | 2.37 | 1.52 | 0.57 | 1.26 | 1.86 | 1.39 |

## 4. Conclusion

Although NMR can be seen as the "golden standard" in sequence and monomer fraction information for copolymers, NMR is not always a viable tool. Measurements are time consuming and the resolution between signals relating to oligomers isn't always sufficient. This renders NMR challenging to apply, especially to larger sample sizes. Py-GC-MS has the potential to overcome this issue. Data interpretation is not straightforward however, as observed trimer fractions cannot be directly related to the original intact polymer.

The developed algorithm was able to improve the average sequence accuracy obtained from experimental pyrolysis data using NMR and Py-GC-MS data of a single reference sample. After fitting MRR and RPC on the reference, the algorithm is able to predict more accurate sequence information when fed pyrolysis data from a new (chemically similar) sample. Although the initial training of the genetic algorithms can take up to three hours, the MRR and RPC can be saved in a library after which they can be readily applied to subsequent samples. Moreover, when the monomer reaction rates of specific samples are known, the first step in the developed algorithm might be omitted. This is the case for many polymerization reactions (*e.g.* [34,35]) and calibration with NMR might be avoidable.

The algorithm can be especially useful when analyzing copolymer samples that produce unstable dimers and trimers. Almost no useful information could be found when analyzing the Py-GC-MS results of the methyl methacrylate-styrene copolymer. All dimers and trimers which included two or more MMA monomers completely degraded and were therefore not detectable. However, even with the limited datapoints available, the algorithm was able to provide accurate information on monomer fractions. Furthermore, next versions of the algorithm could be useful on specific cases where the sensitivity of NMR might not be sufficient. For example, in recent work SEC-Py-GC-MS has been performed to gain insight in sequence information over the molar mass distribution [17]. However, NMR is not feasible as the concentration in the collected fractions is limited.

We envision that the reported work will make the use of Py-GC as a tool for copolymer sequence determination more widespread and reliable. Furthermore, we hope that it will provide a basis for additional chemometric developments in Py-GC data interpretation. It could be highly interesting to apply the workflow to copolymers containing more than two different monomers to increase the understanding of their formation and degradation. Also, the obtained RPCs might be of interest to improve number-average sequence determination (*i.e.* block length) in block copolymers. As with statistical copolymers, the obtained Py-GC results might also contain unrepresentative peak areas. This however, needs additional considerations as the algorithm in current form treats all Py-GC-MS data as if obtained from a statistical copolymer.

## CRediT authorship contribution statement

**Stef R.A. Molenaar:** Conceptualization, Methodology, Computational, Software, Validation, Formal analysis, Data curation, Writing – original draft, Visualization. **Wouter C. Knol:** Conceptualization, Methodology, Experimental, Investigation, Validation, Data curation, Writing – original draft. **Bob W.J. Pirok:** Writing – review & editing, Supervision, Project administration. **Ron A.H. Peters:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data Availability

Data will be made available on request.

## Acknowledgements

SM and WK acknowledge the UNMATCHED project, which is supported by BASF, DSM and Nouryon, and receives funding from the Dutch Research Council (NWO) in the framework of the Innovation Fund for Chemistry (CHIPP Project 731.017.303) and from the Ministry of Economic Affairs in the framework of the "PPS-toeslagregeling".

This work was performed in the context of the Chemometrics and Advanced Separations Team (CAST) within the Centre for Analytical

Sciences Amsterdam (CASA). The valuable contributions of the CAST members are gratefully acknowledged.

## Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jaap.2023.105978.

## References


[1] F.C.Y. Wang, Polymer analysis by pyrolysis gas chromatography, J. Chromatogr. A. 843 (1999) 413–423, https://doi.org/10.1016/S0021-9673(98)01051-6.

[2] R. Rial-Otero, M. Galesio, J.-L. Capelo, J. Simal-Gándara, A review of synthetic polymer characterization by pyrolysis–GC–MS, Chromatographia 70 (2009) 339–348, https://doi.org/10.1365/s10337-009-1254-1.

[3] F.C.-Y. Wang, B.B. Gerhart, P.B. Smith, Structure determination of polymeric materials by pyrolysis gas chromatography, Anal. Chem. 67 (1995) 3536–3540, https://doi.org/10.1021/ac00115a024.

[4] A. Staubli, E. Mathiowitz, R. Langer, Sequence distribution and its effect on glass transition temperatures of poly(anhydride-co-imides) containing asymmetric monomers, Macromolecules 24 (1991) 2291–2298, https://doi.org/10.1021/ma00009a026.

[5] C.-L. Lin, W.-C. Chen, C.-S. Liao, Y.-C. Su, C.-F. Huang, S.-W. Kuo, F.-C. Chang, Sequence distribution and polydispersity index affect the hydrogen-bonding strength of poly(vinylphenol- co -methyl methacrylate) copolymers, Macromolecules 38 (2005) 6435–6444, https://doi.org/10.1021/ma050639t.

[6] I. Dimitrov, B. Trzebicka, A.H.E. Müller, A. Dworak, B. Tsvetanov, Christo, Thermosensitive water-soluble copolymers with doubly responsive reversibly interacting entities, Prog. Polym. Sci. 32 (2007) 1275–1343, https://doi.org/10.1016/j.progpolymsci.2007.07.001.

[7] A.J. Trazkovich, M.F. Wendt, L.M. Hall, Effect of copolymer sequence on structure and relaxation times near a nanoparticle surface, Soft Matter 14 (2018) 5913–5921, https://doi.org/10.1039/c8sm00976g.

[8] S. Tsuge, T. Kobayashi, Y. Sugimura, T. Nagaya, T. Takeuchi, Pyrolysis gas chromatographic characterization of highly alternating copolymers containing styrene and tetracyanoquinodimethan, methyl acrylate, acrylonitrile, or methyl methacrylate units, Macromolecules 12 (1979) 988–992, https://doi.org/10.1021/ma60071a041.

[9] F.C.-Y. Wang, P.B. Smith, Quantitative analysis and structure determination of styrene/methyl methacrylate copolymers by pyrolysis gas chromatography, Anal. Chem. 68 (1996) 3033–3037, https://doi.org/10.1021/ac960196e.

[10] S. Tsuge, S. Hiramitsu, T. Horibe, M. Yamaoka, T. Takeuchi, Characterization of sequence distributions in methyl acrylate-styrene copolymers to high conversion by pyrolysis gas chromatography, Macromolecules 8 (1975) 721–725, https://doi.org/10.1021/ma60048a010.

[11] T. Nagaya, Y. Sugimura, S. Tsuge, Studies on sequence distributions in acrylonitrile-styrene copolymers by pyrolysis-glass capillary gas chromatography, Macromolecules 13 (1980) 353–357, https://doi.org/10.1021/ma60074a027.

[12] T. Okumoto, T. Takeuchi, S. Tsuge, Pyrolysis gas chromatographic study on sequence distribution of dyads in styrene-m-chlorostyrene and styrene-p-chlorostyrene copolymers, Macromolecules 6 (1973) 922–924, https://doi.org/10.1021/ma60036a026.

[13] T. Okumoto, S. Tsuge, Y. Yamamoto, T. Takeuchi, Pyrolysis gas chromatographic evaluation of sequence distribution of dyads in vinyl-type copolymers: acrylonitrile-m-chlorostyrene and acrylonitrile-p-chlorostyrene copolymers, Macromolecules 7 (1974) 376–380, https://doi.org/10.1021/ma60039a021.

[14] F.C.Y. Wang, The microstructure exploration of thermoplastic copolymers by pyrolysis-gas chromatography, J. Anal. Appl. Pyrolysis 71 (2004) 83–106, https://doi.org/10.1016/S0165-2370(03)00100-1.

[15] J.C. Randall, Polymer Sequence Determination Carbon-13 NMR Method, Academic Press, New York, 1977.

[16] A.M. Aerdts, J.W. de Haan, A.L. German, Proton and carbon NMR spectra of alternating and statistical styrene-methyl methacrylate copolymers revisited, Macromolecules 26 (1993) 1965–1971, https://doi.org/10.1021/ma00060a025.

[17] W.C. Knol, T. Gruendling, P.J. Schoenmakers, B.W.J. Pirok, R.A.H. Peters, Co-Polymer sequence determination over the molar mass distribution by size-exclusion chromatography combined with pyrolysis - gas chromatography, J. Chromatogr. A. 1670 (2022), 462973, https://doi.org/10.1016/j.chroma.2022.462973.

[18] W.C. Knol, S. Vos, T. Gruendling, B.W.J. Pirok, R.A.H. Peters, Expansion of the application range of pyrolysis-gas chromatography to copolymer sequence determination: acrylate copolymers, J. Anal. Appl. Pyrolysis 165 (2022), 105578, https://doi.org/10.1016/j.jaap.2022.105578.

[19] Y. Shibasaki, Bound. Eff. Therm. Degrad. Copolym. 5 (1967) 21–34, https://doi.org/10.1002/pol.1967.150050103.

[20] B.K. Alsberg, R. Goodacre, J.J. Rowland, D.B. Kell, Classification of pyrolysis mass spectra by fuzzy multivariate rule induction-comparison with regression, K-nearest neighbour, neural and decision-tree methods, Anal. Chim. Acta 348 (1997) 389–407, https://doi.org/10.1016/S0003-2670(97)00064-0.

[21] D. Broadhurst, R. Goodacre, A. Jones, J.J. Rowland, D.B. Kell, Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression, with applications to pyrolysis mass spectrometry, Anal. Chim. Acta 348 (1997) 71–86, https://doi.org/10.1016/S0003-2670(97)00065-2.

[22] W.C. Knol, S. Vos, T. Gruendling, B.W.J. Pirok, R.A.H. Peters, Improving the accuracy of copolymer sequence length determination by pyrolysis gas chromatography: a comprehensive study, Anal. Chim. Acta 1238 (2023), 340635, https://doi.org/10.1016/j.aca.2022.340635.

[23] S.R.A. Molenaar, B. van de Put, B.W.J. Pirok, Multivar. Other Rapid Effic. Determ. Identif. Softw. Thorough Represent. Interpret. Unveiling Traits Inf. Nov. Synth. (MOREDISTRIBUTIONS) (2021), https://doi.org/10.5281/zenodo.5710530.

[24] S.R.A. Molenaar, B. Van De Put, J.S. Desport, S. Samanipour, R.A.H. Peters, B.W. J. Pirok, Automated feature mining for two-dimensional liquid chromatography applied to polymers enabled by mass remainder analysis, Anal. Chem. 94 (2022) 5599–5607, https://doi.org/10.1021/acs.analchem.1c05336.

[25] F.R. Mayo, C. Walling, Copolymerization, Chem. Rev. 46 (1950) 191–287, https://doi.org/10.1021/cr60144a001.

[26] I.R. Herbert, Statistical analysis of copolymer sequence distribution, in: R.N. Ibbet (Ed.), NMR Spectrosc. Polym., Springer, Netherlands, Dordrecht, 1993, pp. 50–79, https://doi.org/10.1007/978-94-011-2150-7_2.

[27] E. Merz, T. Alfrey, G. Goldfinger, Intramolecular reactions in vinyl polymers as a means of investigation of the propagation step, J. Polym. Sci. 1 (1946) 75–82, https://doi.org/10.1002/pol.1946.120010202.

[28] P. Giraudeau, J.L. Wang, É. Baguet, Improvement of the inverse-gated-decoupling sequence for a faster quantitative analysis by $^{13}C$ NMR, Comptes Rendus Chim. 9 (2006) 525–529, https://doi.org/10.1016/j.crci.2005.06.030.

[29] S. Katoch, S.S. Chauhan, V. Kumar, A review on genetic algorithm: past, present, and future, Multimed. Tools Appl. (2021), https://doi.org/10.1007/s11042-020-10139-6.

[30] M. Yang, Y. Shibasaki, Mechanisms of thermal degradation of polystyrene, polymethacrylonitrile, and their copolymers on flash pyrolysis, J. Polym. Sci. Part A Polym. Chem. 36 (1998) 2315–2330, https://doi.org/10.1002/(SICI)1099-0518(19980930)36:13<2315::AID-POLA19>3.0.CO;2-D.

[31] T. Brooijmans, R. Okhuijsen, I. Oerlemans, P.J. Schoenmakers, R.A.H. Peters, Acid monomer analysis in waterborne polymer systems by targeted labeling of carboxylic acid functionality, followed by pyrolysis – gas chromatography, J. Chromatogr. A. 1560 (2018) 63–70, https://doi.org/10.1016/j.chroma.2018.05.024.

[32] S.E. Levine, L.J. Broadbelt, Reaction pathways to dimer in polystyrene pyrolysis: a mechanistic modeling study, Polym. Degrad. Stab. 93 (2008) 941–951, https://doi.org/10.1016/j.polymdegradstab.2008.01.029.

[33] S. Tsuge, H. Ohtani, C. Watanabe, Pyrolysis–GC/MS Data Book of Synthetic Polymers: Pyrograms. Thermograms and MS of Pyrolyzates, Elsevier, 2012.

[34] A. Issa, A. Luyt, Kinetics of alkoxysilanes and organoalkoxysilanes polymerization: a review, Polym. (Basel) 11 (2019) 537, https://doi.org/10.3390/polym11030537.

[35] L.C. Kröger, W.A. Kopp, K. Leonhard, Prediction of chain propagation rate constants of polymerization reactions in aqueous NIPAM/BIS and VCL/BIS systems, J. Phys. Chem. B. 121 (2017) 2887–2895, https://doi.org/10.1021/ACS.JPCB.6B09147/ASSET/IMAGES/LARGE/JP-2016-09147V_0004.JPEG.