



The triumvirate of effective and rapid synthesis, analysis, and artificial intelligence to explore the structure-property relationship of copolymers

Full-length article

Tibor Nagy^a, Gergő Róth^{a,b}, Ákos Kuki^a, Veronika Pardi-Tóth^{a,b}, Dávid Nyul^{a,b}, Zuura Kaldybek Kyzy^a, Isaac Alexander Iglesias Palacios^a, Máté Benedek^a, Lajos Nagy^a, Miklós Zsuga^a, Sándor Kéki^{a,*}

^a Department of Applied Chemistry, Faculty of Science and Technology, University of Debrecen, Egyetem tér 1, H-4032 Debrecen, Hungary

^b Doctoral School of Chemistry, University of Debrecen, Egyetem tér 1, H-4032 Debrecen, Hungary

Keywords: Copolymers, Structure-property relation, MALDI-TOF MS, artificial intelligence

Understanding the structure-property relationship is of paramount importance for tailoring copolymers for specific applications. Poly(N-acryloylmorpholine)-block-poly(N-isopropylacrylamide) (PNAM-b-PNIPAM) diblock copolymers were synthesized by reversible addition-fragmentation chain transfer (RAFT) polymerization with varying M_n and composition, providing the basis for deducing structure-property relationships. The chemical structure of the copolymers was analyzed by mass spectrometry (MS). A novel and efficient mass spectrum processing methodology was developed for the detailed analysis of polymers/copolymers that greatly expands the upper mass limit of the time-of-flight (TOF) analyzers in the linear mode up to 20,000 Da. Our method “makes visible” the mass peaks of the individual copolymer species and their isotopologues providing effective and fast automatized analysis. The self-assembly property of the thermoresponsive PNAM-b-PNIPAM diblocks in aqueous solutions was investigated by dynamic light scattering (DLS) experiments, and quantified by determining the incipient temperature of the phase transition. For rapid evaluation, an artificial neural network (ANN) was created to explore the hidden relationships between the structural information obtained by our novel mass analysis method and the properties as well as to predict the self-assembly behavior of the copolymers.

Introduction

Block copolymers are a major group of smart, stimulus-responsive materials. The self-assembly of block copolymers has become increasingly important over the past several decades because it allows the creation of a variety of nanostructures [1–5]. The

* Corresponding author.

E-mail address: keki.sandor@science.unideb.hu (S. Kéki).

Received 20 December 2023; Received in revised form 21 February 2024; Accepted 22 February 2024

properties of the block copolymers can be effectively adjusted and tuned by their chemical composition, sequence distribution of the co-monomers and by the molecular weight of the macromolecules formed. Consequently, an arbitrary number of copolymers with various properties can be produced even in the low molecular weight range [6]. Understanding their structure-property relationship is essential for tailoring copolymers for specific applications. In general, copolymers with a narrow molecular weight distribution can be used to map the structure-property relationship. These copolymers can be prepared from a wide range of monomers by living/controlled polymerization that allows a high degree of control over the molecular weight and composition. Such methods include atom transfer radical polymerization (ATRP) [7,8], stable free-radical polymerization (SFRP) [9], and reversible addition-fragmentation chain transfer (RAFT) polymerization [10]. An additional advantage of RAFT polymerization is that it offers fast and simple methods within a short reaction time (a few minutes to an hour) [11,12]. This is important because the inclusion of a large number of copolymer samples gives a more detailed insight into the structure-property relationship. In parallel, there is a growing need for characterization methods that are fast and provide detailed information about the copolymer in addition to the molecular weight and composition [6]. The comprehensive characterization of the block copolymer systems is challenging as it is not enough just to determine the average molecular weights. In addition, we also need the details of the chain microstructure regarding the average length of the blocks and block length distributions as these have significant effects on the self-assembly behavior of the amphiphilic block copolymers [13–15]. Mass spectrometry (MS) is a powerful technique for copolymer analysis, but evaluation of the large number of peaks in the mass spectra requires efficient data processing methods [16–18]. Recently, we have developed a mass spectrometry based method called Mass-Remainder Analysis (MARA) and related techniques, which can even be used to determine the block length distribution in the block copolymers [19–21]. These data mining methods extend the role and applicability of mass spectrometry in polymer chemistry for the evaluation of polymers and copolymers with molecular weights up to 6000 g/mol. The limited mass range is a general problem in the mass spectrometric analysis of polymeric systems and the development of mass spectrometric methods is of paramount importance for a more thorough understanding of the structure-property relationships [6,22,23]. A method that could exploit the advantages of mass spectrometry for copolymers up to a molecular weight of 20,000 g/mol would open a new perspective in the elucidation of structure-property relationships of copolymers. The most commonly used ionization method for the MS analysis of copolymers is the matrix-assisted laser desorption/ionization (MALDI) [24,25]. MALDI typically enables a fast measurement, making it well-suited to rapid polymerization processes. In this way, when exploring the structure-property relationships of polymer and copolymer systems, sufficiently large data sets are generated that can efficiently be evaluated using machine learning (ML) methods.

Machine learning, as a method of artificial intelligence (AI) can uncover the hidden trends and relationships in large data sets

to solve complex problems [26]. ML methods have recently been used to predict and understand the self-assembly behavior of block copolymers and multi-component nanocomposites [27,28]. One of the bottlenecks of the AI-based methods can be that it requires a large amount of data sets. Therefore, simulations are often used to generate sufficient data [29], or intensive data collection is performed from published papers or publicly accessible databases [30,31]. The data sets from the experimental work are limited in size and can only be analyzed using simpler ML models. Usually, rules of thumb (e.g. “rule of 10” as a rough guideline, having at least 10 times more samples than number of features) are used to define the requirements for the size of the data set [32]. It is worth noting that by improving the methodology, the data requirements of ML can be reduced [33]. One of the most widely used AI methods is the artificial neural network (ANN), which models the cognitive processes of the human brain [34–38]. ANN is capable of building a regression model using the supervised learning method, i.e., the system is trained with known input and output data and the resulting model is able to predict the output variables based on the input parameters. Very recently, we have shown that an ANN method can be developed for the analysis of copolymers by gel permeation chromatography (GPC) giving results that can only be obtained experimentally by MALDI-MS or nuclear magnetic resonance (NMR) spectroscopy methods [39].

In this paper, we study the self-assembly behavior of the acrylamide-based homo- and copolymers. One of the most important members of this family is the poly(*N*-isopropylacrylamide) (PNIPAM), which undergoes a reversible hydrophilic-hydrophobic phase transition with increasing the temperature above the lower critical solution temperature (LCST) [40,41]. By varying the chain length and the end groups, it is possible to tune the properties of PNIPAM in the lower molecular weight range [42,43]. Moreover, addition of new monomers and their blocks can further modify the smart and self-assembly behavior of the formed copolymers [41]. These stimuli-responsive copolymers are in the focus of scientific interest, their phase transition and micelle formation have been studied in the form of various microstructures, such as block, random, and triblock copolymers [1,44,45].

This article aims to achieve three objectives. First, we report the synthesis of various PNIPAM homopolymers and poly(*N*-acryloylmorpholine)-block-poly(*N*-isopropylacrylamide) diblocks (PNAM-*b*-PNIPAM). Secondly, we characterize the chemical structure and the aggregation properties of the copolymer. As a major innovation, we have developed a mass spectrum processing method for the very detailed analysis of copolymers by MALDI-TOF MS in the linear mode up to the mass of approximately 20,000 Da. Finally, we have designed and implemented an ANN that takes the detailed characterization data as input and predicts the aggregation behavior in an aqueous solution of the copolymers. Our process flow resulted in the rapid exploration of the structure-property relation of copolymers involving detailed structural data.

Experimental

Chemicals

N-isopropylacrylamide (NIPA), *N*-acryloylmorpholine (NAM), 2,2'-Azobis(2-methylpropionamide) dihydrochloride (initiator)

and dioxane were purchased from Merck (Darmstadt, Germany). 2-(Butylthiocarbonothioylthio)propanoic acid (CTA) was received from ABCR GmbH (Karlsruhe, Germany) and used as received. Tetrahydrofuran and dimethyl sulfoxide were obtained from VWR International (Leuven, Belgium). Water was purified by a Direct-Q water system (Millipore, Molsheim, France).

Matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS)

The mass spectrometric measurements were carried out by an Autoflex Speed MALDI-TOF MS instrument (Bruker Daltonik, Bremen, Germany). Linear mode was applied where the ion-source voltage 1 and ion-source voltage 2 were 19.5 kV and 18.3 kV, respectively. The instrument was equipped with a solid phase (355 nm) laser. All mass spectra were externally calibrated with PNAM homopolymers. The matrix and ionizing agent was *Trans*-2-[3-(4-*tert*-Butylphenyl)-2-methyl-2-propenylidene] malononitrile (DCTB) and CsCl, respectively. The solvent for the matrix was THF, while for CsCl and PNAM homopolymers, water and THF:water 4:1 mixture was used, respectively. The concentration of the matrix, ionizing agent and samples were 20 mg/mL, 2.5 mg/mL and 10 mg/mL. The mixing ratio was 20:4:1, respectively.

Nuclear magnetic resonance spectroscopy (NMR)

A Bruker Avance II 400 MHz spectrometer equipped with a 5 mm z-gradient BBI probe was applied for the ^1H -NMR measurements. 1D ^1H -NMR experiments were performed at 297 K with a 45-degree ^1H excitation pulse, a relaxation delay of 5 s, spectral width of 13.97 ppm and 32k total data points. Bruker TopSpin 4.0.9 and Mestrelab MestreNova 14.1 software were used for NMR spectral processing and integration. For all measurements, 6 mg of the sample was dissolved in 600 μL D_2O .

Gel permeation chromatography (GPC)

The gel permeation chromatographic measurements were carried out with a Waters Alliance e2695 separation module. For detection, a Waters 2414 refractive index detector was applied (Waters Corp., Milford, MA, USA). THF/DMSO (95/5 V/V%) eluent was used (flow rate: 0.5 mL/min, T: 40°C, run: 35 min), columns: Styragel HR 0.5, 1,2 and 4 (300 \times 4.6 mm, Waters, Milford, MA, USA). The GPC methods were calibrated with polystyrene standard. Samples were dissolved in the eluents at a concentration of approximately 10 mg/mL and filtered by a syringe filter (0.45 μm) prior to analysis.

Dynamic light scattering (DLS)

A Zetasizer Nano ZS instrument (Malvern Instruments, Malvern, UK) equipped with a He-Ne laser (633 nm) with a detector position of 173° was used for dynamic light scattering experiments. The derived count rate and the hydrodynamic particle diameter (d , Z-average) were recorded. The phase transition was monitored with 0.2°C temperature steps, and the heating rate was lower than 0.025 K min^{-1} , as it was suggested by previous publications [41,46]. The Aggregation points were identified as the incipient temperature of the phase transition. The Z-average size was calculated by cumulative analysis.

Artificial neural network (ANN)

We developed an ANN model based on supervised learning, which is the most common form of machine learning [26,34]. The ANN modeling was performed using the Neural Network Toolbox of MATLAB 2020b (The MathWorks, Inc., Natick, MA, USA). A feed-forward network with sigmoid hidden neurons and sigmoid output neurons was constructed. The architecture of the ANN will be detailed later. The network was trained using the Bayesian regularization algorithm. The input data were randomly divided into training (90%) and test (10%) sets. The performance of the network was evaluated by the mean square error (MSE) and Pearson's correlation coefficient (r). The hyperparameters of the ANN were optimized by varying the number of hidden layers (1, 2), number of hidden neurons (1-4), and varying the type of the transfer functions (*relu*, *radial basis*, *tansig*, and *purelin*). When the increase of the hidden layers or the hidden neurons did not yield better performance, the network with the smallest number of layers or neurons at the maximum performance was chosen.

The development and evaluation of the ANN were based on the entire data set (the training set plus test data). The accuracy (%) of the ANN prediction was calculated as the average of $abs(1 - \text{predicted value}/\text{experimental target value}) \times 100$ values for the entire (the training set plus test data) and test data set [47].

Results and discussion

Synthesis of copolymers

The PNIPAM homopolymers and PNAM-b-PNIPAM block copolymers were prepared using a RAFT polymerization process based on the rapid method developed by Gody et al. [11,12], detailed procedure is in the Supporting Information. To synthesize blocks in 20 minutes, we used lower monomer concentrations and increased reaction times. This resulted in a more controllable and less viscous reaction mixture. Altogether 58 polymers were synthesized with varying M_n and comonomer (NIPAM) mole fractions (see Supporting Information Table S1). The M_n and mole fraction values are approximately evenly distributed in the 3,000 – 15,000 Da and 0 – 1 range, respectively. In addition to the detailed mass spectrometric characterization, which will be discussed in the next chapter, the average characteristics of the polymers were determined by ^1H -NMR spectroscopy [48] and GPC (see Supporting Information Table S1). Based on the GPC results, the distribution of the prepared polymers is narrow. The ^1H -NMR and GPC results are in good agreement with those obtained by MALDI-TOF analysis.

Structure analysis

MALDI-TOF MS in the linear mode can be used for the molecular weight determination of polymers over a wide mass range, typically up to 200,000 Da but with limited resolving power. The separation and accurate mass determination of the isobaric and isotopic ions cannot be achieved. A remarkable feature of the mass spectra of the copolymers is that the peaks corresponding to the individual chains are grouped into clusters [49]. Our NAM-NIPAM model system is a particularly appropriate example of this clustering, as replacing 4 NAM with 5 NIPAM units results only in a 1 Da difference (see Fig. 1b). Moreover, in the higher molecular weight range, the number of possible overlaps rises sharply, partly

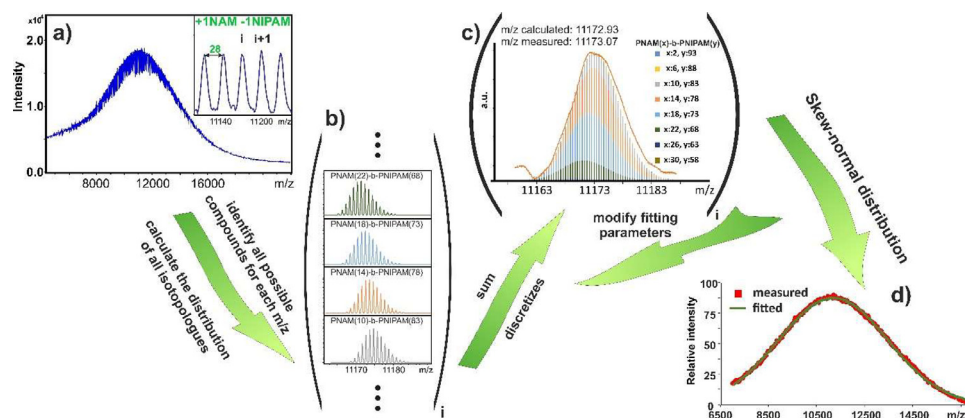


Fig. 1

Mass spectrometric analysis of the poly(N-acryloylmorpholine)-block-poly(N-isopropylacrylamide) copolymer **17** (NAM₁₆NIPAM₇₇) with the designed composition NAM₂₀NIPAM₈₀. a) Mass spectrum recorded by MALDI TOF in the linear mode; b) isotope distributions of some PNAM(x)-b-PNIPAM(y) species that can contribute to the m/z peak at 11,173; c) measured (solid curve) and calculated (histogram) composite mass peak at m/z 11,173; d) envelope curves of the measured and calculated (fitted) mass spectra.

due to the relatively low resolution, thus, these peak clusters appear as broad composite peaks (see Fig. 1a inset).

As a representative example, Fig. 1a shows the mass spectrum of the poly(N-acryloylmorpholine)-block-poly(N-isopropylacrylamide) copolymer **17** (with the designed and determined composition NAM₂₀NIPAM₈₀ and NAM₁₆NIPAM₇₇) recorded by MALDI time-of-flight (TOF) MS in the linear mode. As seen in the inset of Fig. 1a, the envelope curves of the peak clusters can be observed as individual broad merged peaks in the mass spectrum. (The characteristic difference of 28 Da between the adjacent peaks corresponds to the mass difference between the NAM and NIPAM repeat units.) A single broad composite peak is composed of numerous copolymer molecules with their isotopes (see Fig. 1b and c). Therefore, assigning the m/z value of a composite peak to a single component will, of course, lead to the misinterpretation of the spectrum. However, by taking all the measured composite peaks in parallel and considering their all possible copolymer molecules with their isotopologues as constituents it is possible to identify the individual copolymer chains using a suitable fitting algorithm. Furthermore, despite the overlapping peaks, it is also feasible to determine the corresponding abundance of each individual species, thus analyzing the copolymer sample in detail. Our data processing method generates a mass spectrum, taking into account the possible individual copolymer chains and the resolution of the instrument and iteratively fits it to the measured spectrum. The detailed description of our method and the source code of the implementation is presented in the Supporting Information. We highlight here only some essential elements of the process. Of course, considering several thousand m/z peaks as fitting variables is not feasible, so a key to our algorithm is to mathematically describe the block length distribution in the copolymer sample and use the parameters of these distributions as fitting variables. In the case of controlled polymerization, the use of simple distributions is reasonable, thus, we have chosen the skewed normal distribution to account for possible asymmetries. The equation of the skewed normal distribution can be found in the

supporting information while its parameters are the following: SD_{monomer} : standard deviation of a monomer unit, N_{avg} : average number of a monomer unit and α stands for the skew parameter. A very important step in our data processing algorithm is the construction of the measured composite peaks (see Fig. 1c). For example, the peak at m/z 11,173 in Fig. 1c is composed of 8 different copolymer molecules with their isotopologues (25 for each of them) and 200 peaks in total. (In addition, when the mass spectrum is constructed, all peaks are taken as a Gaussian curve with approximately 60 discrete intensity vs m/z points, in accordance with the limited m/z resolution of the TOF linear mode. The fitting algorithm modifies the parameters of the block length distributions (6 parameters in the case of diblocks) and hence the abundances of the individual molecules in the population, including the isotopologues. Thus, for example, the intensities of the 200 theoretical peaks mentioned above that form the peak at m/z 11,173 can be determined, and in this way the fitting algorithm approximates the calculated composite peaks to the measured ones (see the histogram and solid line in Fig. 1c). The residual of the fit has two components, the differences between the m/z and between the intensity values of the maximum of the calculated composite peaks and the corresponding observed peaks. As seen in Fig. 1c, for an arbitrarily selected peak, the shapes of the calculated and observed peaks are in good agreement. A similar good fit can be observed when comparing the calculated and measured envelope curves of the full spectrum (see Fig. 1d). The results of the detailed MALDI-TOF MS analysis are shown in Table S1 for all polymers. The values are in good agreement with those obtained by ¹H-NMR and GPC. Additional details are presented about our data processing method in the Supporting Information; the residual plots of the regression are reported in Fig. S1, a reproducibility analysis is given in Table S2 and the main copolymer parameters are depicted in Fig. S2. The Pearson correlation coefficient between the mole fraction of NIPAM measured by ¹H-NMR and MALDI-TOF MS is high, $r = 0.951$ (only copolymers were used for calculation, sample 1 to sample 48).

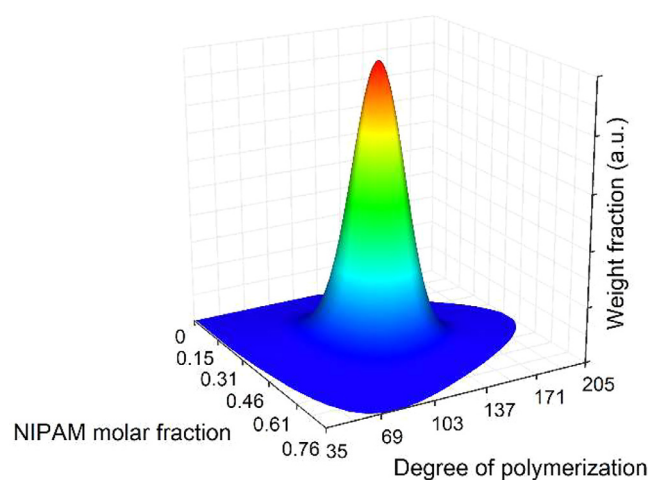


Fig. 2

Bivariate distribution of the poly(N-acryloylmorpholine)-block-poly(N-isopropylacrylamide) copolymer **43** (M_n :15800 Da, $NAM_{77}NIPAM_{42}$).

As a result of our data processing method, we have revealed the "invisible peaks" of the mass spectrum, and we have compiled the list of the individual copolymer chains (including their isotopes) and estimated their abundances applying the fitted $SD_{monomer}$ and N_{avg} parameters for both monomer units. This huge data set enables the detailed characterization of the copolymer sample, namely the calculation of the number-average molecular weight (M_n), weight-average molecular weight (M_w), dispersity $D=M_w/M_n$, the average molar fraction (c_A) and weight fraction (w_A) of unit *A* in the copolymer, the number average number of units *A* and *B* (n_n^A , n_n^B), weight average number of units *A* and *B* (n_w^A , n_w^B), and the dispersity for the monomers ($D_A=n_w^A/n_n^A$, $D_B=n_w^B/n_n^B$) [50]. In addition, unambiguous information about the structure of the copolymer can be obtained by determining the bivariate compositional distribution, where one variable is the degree of polymerization and the other is the composition (see Fig. 2) [51,52].

As seen in Fig. 2, the bivariate distribution provides an excellent visual representation of the copolymer sample and allows scientists and engineers to quickly assess the composition of copolymers during copolymerization process optimization, material design or quality control.

Property analysis

For applications such as drug delivery systems, the most important property of poly(N-acryloylmorpholine)-block-poly(N-isopropylacrylamide) copolymers is their LCST-type thermoresponsive aggregation behavior. We quantified this property by determining the incipient temperature of the phase transition (T_i , aggregation point) [53], above which the PNAM-b-PNIPAM copolymers aggregate. Dynamic light scattering (DLS) measurements were performed on aqueous solutions of PNAM-b-PNIPAM and the increase in the derived count rate (DCR, which represents the light scattering intensity) indicated the onset of the phase transition, i.e., the appearance of the copolymer aggregates.

Fig. 3a shows the DCR and Z-average hydrodynamic diameter *versus* temperature plots of the copolymer **16** ($NAM_{10}NIPAM_{54}$).

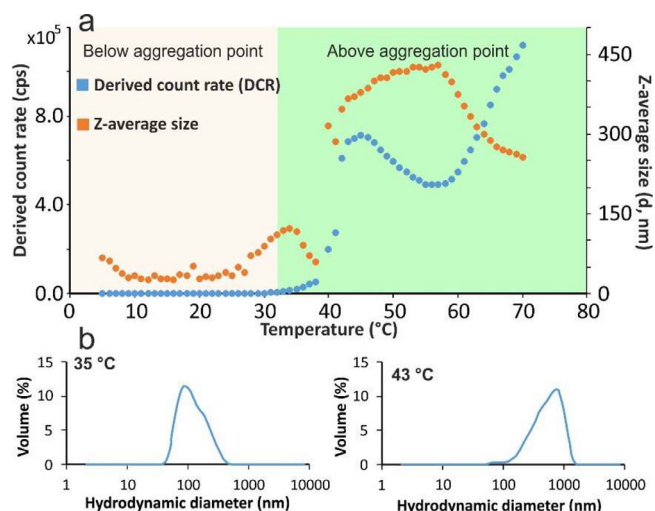


Fig. 3

Dynamic light scattering (DLS) analysis of the poly(N-acryloylmorpholine)-block-poly(N-isopropylacrylamide) copolymer **16** ($NAM_{10}NIPAM_{54}$) in 1 m/m% aqueous solution. a) Derived count rate (DCR) and Z-average hydrodynamic diameter as a function of temperature. b) Volume-weighted size distribution of the copolymer aggregates at various temperatures.

As can be seen in the figure, the change in z-average particle size is consistent with the change in DCR. It should be noted, however, that both curves (DCR and z-average *versus* temperature) have more complex shapes compared to the shape of the decrease in transparency (increase in turbidity) measured with UV-VIS spectrometers, which is the most common method used to study the LCST behavior of copolymers. The drop in the DCR curve after the steep rise is probably due to the appearance of multiple scattering, but the subsequent increase indicates a change in the size of the copolymer aggregates. The volume distributions shown in Fig. 3b also support this assumption about the complex nature of aggregation behavior, but the discussion of this finding is beyond the scope of this paper. In Fig. 3a, the lower bound of the green-colored background indicates T_i , which will be used in the following as a measure of aggregation behavior when examining the structure-property relationship. The T_i values together with the Z-average diameters recorded at T_i for all the copolymers synthesized are compiled in the Supporting Information Table S3, while Fig. 4 plots these as a bivariate function of NIPAM mole fraction and M_n .

As seen in Fig. 4, the increasing length of the NAM block (reflected in the mole fraction of NIPAM) modifies the self-assembly behavior of PNIPAM, so that T_i is shifted to higher temperatures. An additional figure is given in the Supporting Information (Fig. S3) demonstrating the effect of NAM blocks on the Z-average diameters of the aggregates formed at T_i . However, Fig. 4 also shows that T_i values can differ for nearly identical comonomer mole fractions and M_n . This underlines the importance of detailed copolymer characterization for the elucidation of structure-property relationships. Our novel mass spectrum processing method, discussed in the previous chapter, provides the chain length distributions of the comonomers (not just their average fractions), and the parameters of these

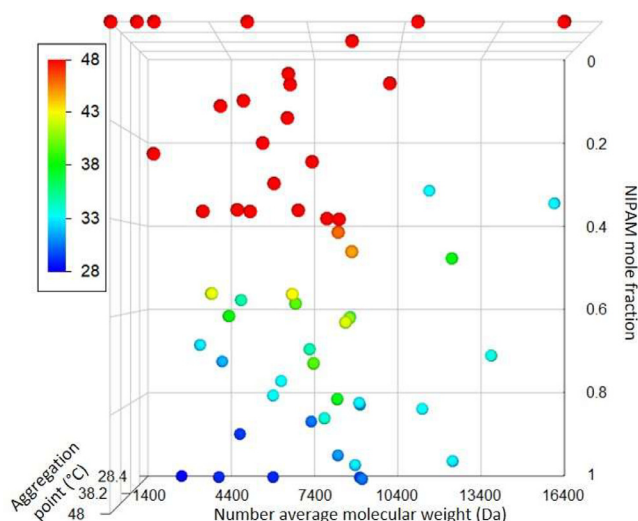


Fig. 4

The incipient temperature of the phase transition (T_i) of the synthesized copolymers determined by Dynamic light scattering (DLS) measurements as a bivariate function of NIPAM mole fraction and M_n .

distributions serve as inputs to the AI-based method, which is presented in the following.

Study of structure-property relationships by artificial neural network (ANN)

The block length distributions obtained as the output results of our mass spectrum processing method describe the copolymer systems in details but are not explicitly related to T_i , as a measure of aggregation behavior. However, the number of copolymer samples synthesized and characterized allows the discovery of hidden relationships by means of ANNs.

A two-layer feed-forward ANN with three sigmoid hidden neurons and one sigmoid output neuron was constructed (Fig. 5). The network has four input variables, such as the mean and standard deviation of the block length distribution of NAM and NIPAM (SD_{NAM} , SD_{NIPAM} , $N_{avg, NAM}$ and $N_{avg, NIPAM}$). (As described earlier, skewed normal distributions were generally used to evaluate the mass spectra, but for the copolymers studied, the skewness was found to be zero, so we use only four input parameters instead of six. All the used data are in the Table S3.) The output of the ANN is the predicted T_i . The number of hidden neurons was determined as detailed in the Experimental Section. The target T_i values of the ANN were determined by DLS, as detailed in the previous chapter. Supplementary Fig. S4 shows the parity plot comparing the experimental onset temperature of the phase transition (T_i) with those predicted by the ANN. The correlation between the experimental and the predicted values is very good, the Pearson's correlation coefficient is $r = 0.993$. Furthermore, the predicted T_i values are quite close to the diagonal line, indicating an accurate prediction. The accuracy of the ANN prediction was quantified as detailed in the Experimental Section and found to be 1.75% and 1.79 % for the full and the test dataset, respectively. The paired T-test shows no significant differences between the experimental and predicted T_i values ($p = 0.72$, $df:57$). The accuracy of the model is high and adequate to give the chemical composition and molecular weight of a copolymer with a specific aggregation point depending on its use.

Our ANN, developed by analyzing nearly 60 synthesized polymers, can also be used to investigate theoretically assembled copolymers (M_n : 3500-12500 Da, step:1000, and NIPAM mole fraction: 0-1, step:0.1). As seen in Fig. 6a, as expected, the mole fraction of the comonomers has a significant effect on the incipient aggregation temperature (T_i). However, our novel mass spectrum processing method can even determine the

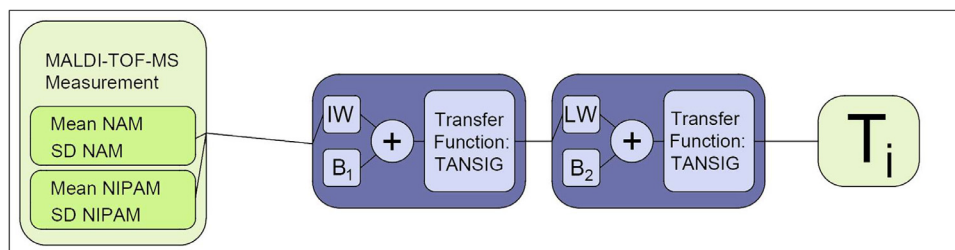


Fig. 5

The ANN with a 4-3-1 architecture used for the determination of the incipient temperature of the phase transition (T_i). (SD stands for standard deviation.)

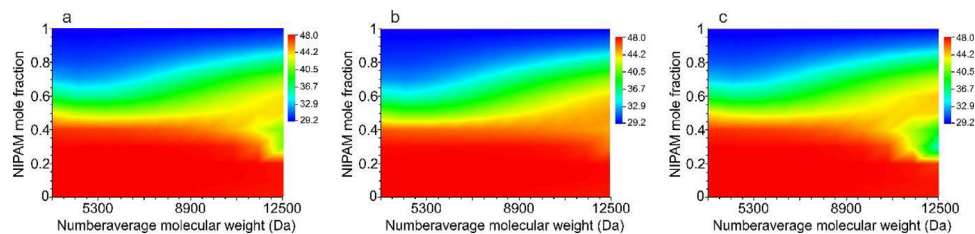


Fig. 6

The application of ANN for the study of structure-property relationships using theoretical input data. a) The standard deviations of the two block length distributions are identical ($SD_{NAM} \sim SD_{NIPAM}$); b) $SD_{NAM} < SD_{NIPAM}$ c) $SD_{NAM} > SD_{NIPAM}$.

block length distributions, which also influence the aggregation behavior as shown in Fig. 6. The very detailed structure-property relationship analysis performed by ANN can be a useful tool for designing copolymer molecules, developing synthesis methods, and controlling quality.

Conclusions

In this paper, we reported the synthesis of different poly(N-acryloylmorpholine)-block-poly(N-isopropylacrylamide) diblock copolymers. Their detailed chemical structure was analyzed by MALDI-TOF MS in the linear mode. Our novel mass spectrum processing methodology for the detailed characterization of copolymers allowed the detection and intensity assignment of the overlapping mass peaks up to 20,000 Da, thus, greatly extending the mass range of MALDI-TOF MS method. By resolving the individual copolymer molecules and their isotopologues, not only the M_n and average comonomer fractions but the length distribution of the blocks representing the detailed chemical composition of the copolymer was also determined. Due to the automation, the large dataset was evaluated in a relatively short time. The large number of copolymers synthesized and their detailed mass spectrometric characterization provided the basis for the discovery of the structure-property relationship using artificial intelligence. We constructed an ANN that takes characteristic copolymer parameters of the thermoresponsive copolymer (PNAM-*b*-PNIPAM specific copolymers) as input and predicts its aggregation temperature. Comparing the experimental target values to the corresponding predicted results (the r coefficient is above 0.99) it was found that the ANN correlates them at a very high level. The accuracy of the prediction is also high, 1.75%. Our new mass spectrum processing method and ANN can be adopted for other copolymer systems even using a different target value to characterize the polymer property. Once built and trained, the ANN helps polymer chemists design new smart copolymers, develop novel synthesis methods and monitor production quality.

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

The data used in this article is shared as supplementary material.

CRediT authorship contribution statement

Tibor Nagy: Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Gergő Róth:** Writing – original draft, Software, Data curation, Conceptualization. **Ákos Kuki:** Writing – original draft, Methodology, Investigation, Conceptualization. **Veronika Pardi-Tóth:** Methodology, Investigation. **Dávid Nyul:** Methodology, Investigation. **Zuura Kaldybek Kyzy:** Methodology, Investigation. **Isaac Alexander Iglesias Palacios:**

Methodology, Investigation. **Máté Benedek:** Methodology, Investigation. **Lajos Nagy:** Methodology, Investigation. **Miklós Zsuga:** Writing – review & editing. **Sándor Kéki:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization.

Acknowledgments

The authors would like to thank for the financial support provided by the following grants: GINOP-2.3.3-15-2016-00021 projects Project no. RRF-2.3.1-21-2022-00009, titled National Laboratory for Renewable Energy has been implemented with the support provided by the Recovery and Resilience Facility of the European Union within the framework of Programme Széchenyi Plan Plus. Furthermore, this paper was also supported by Grant No. FK-132385 from National Research, Development and Innovation Office (NKFI). T.N. acknowledge the support of the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (BO/00212/20/7 (T.N.) and ÚNKP-22-05-DE-426 (T.N.) New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund. Thanks for the financial support provided by the Mol Group, Hungary.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.giant.2024.100248.

References

- [1] A. Prause, M. Hechenbichler, B. von Lospichl, A. Feoktystov, R. Schweins, N. Mahmoudi, A. Laschewsky, M. Gradzielski, Aggregation behavior of nonsymmetrically end-capped thermoresponsive block copolymers in aqueous solutions: between polymer coils and micellar states, *Macromolecules* 55 (2022) 5849–5863.
- [2] M. Karayianni, S. Pispas, Block copolymer solution self-assembly: Recent advances, emerging trends, and applications, *J. Polym. Sci.* 59 (2021) 1874–1898.
- [3] J. Zhang, B. Farias-Mancilla, I. Kulai, S. Hoepfner, B. Lonetti, S. Prévost, J. Ulbrich, M. Destarac, O. Colombani, U.S. Schubert, C. Guerrero-Sanchez, S. Harrison, Effect of hydrophilic monomer distribution on self-assembly of a pH-responsive copolymer: spheres, worms and vesicles from a single copolymer composition, *Angew. Chem. Int. Ed* 60 (2021) 4925–4930.
- [4] M. Sponchioni, U. Capasso Palmiero, D. Moscatelli, Thermo-responsive polymers: Applications of smart materials in drug delivery and tissue engineering, *Mater. Sci. Eng. C* 102 (2019) 589–605.
- [5] D. Roy, W.L.A. Brooks, B.S. Sumerlin, New directions in thermoresponsive polymers, *Chem. Soc. Rev.* 42 (2013) 7214–7243.
- [6] E.S. Bates, M.A. Hillmyer, T.P. Lodge, C.M. Bates, K.T. Delaney, G.H. Fredrickson, Multiblock polymers: panacea or Pandora's Box? *Science* 336 (2012) 434–440.
- [7] S. Perrier, 50th anniversary perspective: RAFT polymerization—a user guide, *Macromolecules* 50 (2017) 7433–7447.
- [8] G. Gody, R. Barbey, M. Danial, S. Perrier, Ultrafast RAFT polymerization: multiblock copolymers within minutes, *Polym. Chem.* 6 (2015) 1502–1511.
- [9] G. Gody, T. Maschmeyer, P.B. Zetterlund, S. Perrier, Rapid and quantitative one-pot synthesis of sequence-controlled polymers by radical polymerization, *Nat. Commun.* 4 (2013) 2505.
- [10] C. Zhang, M.W. Bates, Z. Geng, A.E. Levi, D. Vigil, S.M. Barbon, T. Loman, K.T. Delaney, G.H. Fredrickson, C.M. Bates, A.K. Whittaker, C.J. Hawker, Rapid generation of block copolymer libraries using automated chromatographic separation, *J. Am. Chem. Soc.* 142 (2020) 9843–9849.
- [11] M.-N. Antonopoulou, R. Whitfield, N.P. Truong, D. Wyers, S. Harrison, T. Junkers, A. Anastasaki, Concurrent control over sequence and dispersity in multiblock copolymers, *Nat. Chem.* 14 (2022) 304–312.
- [12] M.-N. Antonopoulou, R. Whitfield, N.P. Truong, A. Anastasaki, Controlling polymer dispersity using switchable RAFT agents: unravelling the effect of the organic content and degree of polymerization, *Eur. Polym. J.* 174 (2022) 111326.
- [13] K. Parkatzidis, N.P. Truong, R. Whitfield, C.E. Campi, B. Grimm-Lebsanft, S. Buchenau, M.A. Rübhausen, S. Harrison, D. Konkolewicz, S. Schindler, A. Anastasaki, Oxygen-enhanced atom transfer radical polymerization through the formation of a copper superoxido complex, *J. Am. Chem. Soc.* 145 (2023) 1906–1915.

- [14] K. Matyjaszewski, *Advanced materials by atom transfer radical polymerization*, *Adv. Mater.* 30 (2018) 1706441.
- [15] K.-A. Hansen, J.P. Blinco, Nitroxide radical polymers – a versatile material class for high-tech applications, *Polym. Chem.* 9 (2018) 1479–1516.
- [16] M.S. Engler, S. Crotty, M.J. Barthel, C. Pietsch, K. Knop, U.S. Schubert, S. Böcker, Coconut—an efficient tool for estimating copolymer compositions from mass spectra, *Anal. Chem.* 87 (2015) 5223–5231.
- [17] T. Fouquet, H. Sato, Extension of the Kendrick mass defect analysis of homopolymers to low resolution and high mass range mass spectra using fractional base units, *Anal. Chem.* 89 (2017) 2682–2686.
- [18] J.S. Town, Y. Gao, E. Hancox, E. Liarou, A. Shegiwal, C.J. Atkins, D. Haddleton, Automatic peak assignment and visualisation of copolymer mass spectrometry data using the ‘genetic algorithm’, *Rapid Commun. Mass Spectrom.* 34 (2020) e8654.
- [19] T. Nagy, Á. Kuki, M. Zsuga, S. Kéki, Mass-remainder analysis (MARA): a new data mining tool for copolymer characterization, *Anal. Chem.* 90 (2018) 3892–3897.
- [20] T. Nagy, Á. Kuki, M. Hashimov, M. Zsuga, S. Kéki, Multistep mass-remainder analysis and its application in copolymer blends, *Macromolecules* 53 (2020) 1199–1204.
- [21] G. Róth, T. Nagy, Á. Kuki, M. Hashimov, Z. Vonza, I. Timári, M. Zsuga, S. Kéki, Polydispersity ratio and its application for the characterization of poloxamers, *Macromolecules* 54 (2021) 9984–9991.
- [22] K. De Bruycker, A. Welle, S. Hirth, S.J. Blanksby, C. Barner-Kowollik, Mass spectrometry as a tool to advance polymer science, *Nat. Rev. Chem.* 4 (2020) 257–268.
- [23] C. Wesdemiotis, Multidimensional mass spectrometry of synthetic polymers and advanced materials, *Angew. Chem. Int. Ed.* 56 (2017) 1452–1464.
- [24] M. Karas, F. Hillenkamp, Laser desorption/ionization of proteins with molecular masses exceeding 10,000 daltons, *Anal. Chem.* 60 (1988) 2299–2301.
- [25] K. Tanaka, H. Waki, Y. Ido, S. Akita, Y. Yoshida, T. Yoshida, T. Matsuo, Protein and polymer analyses up to m/z 100 000 by laser ionization time-of-flight mass spectrometry, *Rapid Commun. Mass Spectrom.* 2 (1988) 151–153.
- [26] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (2015) 436–444.
- [27] K.-H. Tu, H. Huang, S. Lee, W. Lee, Z. Sun, A. Alexander-Katz, C.A. Ross, Machine learning predictions of block copolymer self-assembly, *Adv. Mater.* 32 (2020) 2005713.
- [28] E. Vargo, J.C. Dahl, K.M. Evans, T. Khan, P. Alivisatos, T. Xu, Using machine learning to predict and understand complex self-assembly behaviors of a multicomponent nanocomposite, *Adv. Mater.* 34 (2022) 2203168.
- [29] A.L. Nazarova, L. Yang, K. Liu, A. Mishra, R.K. Kalia, K.-i. Nomura, A. Nakano, P. Vashishta, P. Rajak, Dielectric polymer property prediction using recurrent neural networks with optimizations, *J. Chem. Inf. Model.* 61 (2021) 2175–2186.
- [30] G. Bradford, J. Lopez, J. Ruza, M.A. Stolberg, R. Osterude, J.A. Johnson, R. Gomez-Bombarelli, Y. Shao-Horn, Chemistry-informed machine learning for polymer electrolyte discovery, *ACS Cent. Sci.* 9 (2023) 206–216.
- [31] L. Tao, J. Byrnes, V. Varshney, Y. Li, Machine learning strategies for the structure-property relationship of copolymers, *iScience* 25 (2022) 104585.
- [32] A. Alwosheel, S. van Cranenburgh, C.G. Chorus, Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis, *J. Choice Model.* 28 (2018) 167–182.
- [33] M. Bataineh, T. Marler, Neural network for regression problems with reduced training sets, *Neur. Netw.* 95 (2017) 1–9.
- [34] L.B. Ayres, E.J.V. Gomez, J.R. Linton, M.F. Silva, C.D. Garcia, Taking the leap between analytical chemistry and artificial intelligence: a tutorial review, *Anal. Chim. Acta* 1161 (2021) 338403.
- [35] J. Gasteiger, Chemistry in times of artificial intelligence, *Chemphyschem* 21 (2020) 2233–2242.
- [36] J. Behler, First principles neural network potentials for reactive simulations of large molecular and condensed systems, *Angew. Chem. Int. Ed.* 56 (2017) 12828–12840.
- [37] D. Packwood, L.T.H. Nguyen, P. Cesana, G. Zhang, A. Staykov, Y. Fukumoto, D.H. Nguyen, Machine learning in materials chemistry: an invitation, *Mach. Learn. Appl.* 8 (2022) 100265.
- [38] Q. Zhou, S. Lu, Y. Wu, J. Wang, Property-oriented material design based on a data-driven machine learning technique, *J. Phys. Chem. Lett.* 11 (2020) 3920–3927.
- [39] T. Nagy, G. Róth, M. Benedek, Á. Kuki, I. Timári, M. Zsuga, S. Kéki, Enhanced copolymer characterization for polyethers using gel permeation chromatography combined with artificial neural networks, *Anal. Chem.* 95 (2023) 10504–10511.
- [40] G.-F. Luo, W.-H. Chen, X.-Z. Zhang, 100th anniversary of macromolecular science viewpoint: poly(N-isopropylacrylamide)-based thermally responsive micelles, *ACS Macro Lett.* 9 (2020) 872–881.
- [41] A. Halperin, M. Kröger, F.M. Winnik, Poly(N-isopropylacrylamide) phase diagrams: fifty years of research, *Angew. Chem. Int. Ed.* 54 (2015) 15342–15367.
- [42] R. Pamies, K. Zhu, A.-L. Kjøniksen, B. Nyström, Thermal response of low molecular weight poly(N-isopropylacrylamide) polymers in aqueous solution, *Polym. Bull.* 62 (2009) 487–502.
- [43] S. Furryk, Y. Zhang, D. Ortiz-Acosta, P.S. Cremer, D.E. Bergbreiter, Effects of end group polarity and molecular weight on the lower critical solution temperature of poly(N-isopropylacrylamide), *J. Polym. Sci., Part A: Polym. Chem.* 44 (2006) 1492–1501.
- [44] M. Sahn, T. Yildirim, M. Dirauf, C. Weber, P. Sungur, S. Hoepfner, U.S. Schubert, LCST behavior of symmetrical PNIPAm-b-PtEOx-b-PNIPAm triblock copolymers, *Macromolecules* 49 (2016) 7257–7267.
- [45] Y. Kotsuchibashi, M. Ebara, T. Aoyagi, R. Narain, Recent advances in dual temperature responsive block copolymers and their potential as biomedical applications, *Polymers* 8 (2016) 380.
- [46] Q. Zhang, C. Weber, U.S. Schubert, R. Hoogenboom, Thermoresponsive polymers with lower critical solution temperature: from fundamental aspects and measuring techniques to recommended turbidimetry conditions, *Mater. Horiz.* 4 (2017) 109–116.
- [47] A. Tropsha, Best practices for QSAR model development, validation, and exploitation, *Mol. Inform.* 29 (2010) 476–488.
- [48] T. Nagy, A. Kuki, G. Roth, Z.K. Kyzy, A.B. Zatalini, D. Nyul, M. Zsuga, S. Kéki, Characterization of polyacrylamide diblock copolymers by mass spectrometry combined with Mass-remainder analysis (MARA), *Polym. Test.* 117 (2023) 107834.
- [49] Á. Kuki, G. Róth, A. Nagy, M. Zsuga, S. Kéki, T. Nagy, A short-cut data mining method for the mass spectrometric characterization of block copolymers, *Processes* 10 (2022) 42.
- [50] G.J. van Rooij, M.C. Duursma, C.G. de Koster, R.M.A. Heeren, J.J. Boon, P.J.W. Schuyf, E.R.E. van der Hage, Determination of block length distributions of poly(oxypropylene) and poly(oxyethylene) block copolymers by MALDI-FTICR mass spectrometry, *Anal. Chem.* 70 (1998) 843–850.
- [51] M.S. Montaudo, G. Adamus, M. Kowalczyk, Bivariate distribution in copolymers: a new model, *J. Polym. Sci., Part A: Polym. Chem.* 40 (2002) 2442–2448.
- [52] O. Trhlíková, Z. Walterová, M. Janata, L. Kanizsová, J. Horský, Compositional distribution of binary living copolymers and their end sequences, *Macromol. Theory Simul.* 30 (2021) 2100002.
- [53] Q.-T. Pham, Z.-H. Yao, Y.-T. Chang, F.-M. Wang, C.-S. Chern, LCST phase transition kinetics of aqueous poly(N-isopropylacrylamide) solution, *J. Taiwan Inst. Chem. Eng.* 93 (2018) 63–69.