Polymer Chemistry

ROYAL SOCIETY OF CHEMISTRY

TUTORIAL REVIEW

Published on 05 October 2026
Licensed under CC-BY 4.0

Check for updates

Cite this: *Polym. Chem.*, 2026, **17**, 4527

Pyrolysis mass spectra as polymer descriptors: reference-free quantification of polymer sequence, degradation, and performance

Yusuke Hibi

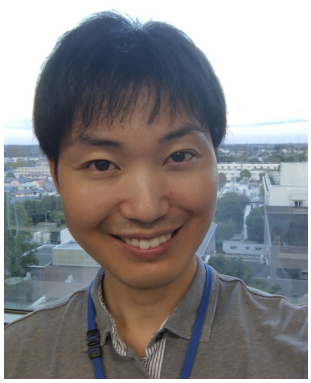
Pyrolysis mass spectrometry (Py-MS) holds unique potential for polymer characterisation because the masses and generation temperatures of pyrolysis fragments reflect the multi-scale structural distributions inherent in polymeric materials. Despite this promise, broader adoption of Py-MS has been relatively limited, owing to the complexity of pyrolysis and, more fundamentally, to the non-quantitative nature of mass spectrometry arising from fragment-specific ionisation efficiencies. This situation is now changing with the emergence of reference-free quantitative mass spectrometry (RQMS), which represents the structural distribution of interest as mixing ratios among monodisperse reference states defined according to the analytical purpose: sequence-defined polymers for sequence analysis; pristine and completely degraded polymers for degradation analysis. If reference spectra of these well-defined materials were available, their mixing ratios could be obtained by conventional spectral decomposition; however, such materials are often impossible to synthesise or isolate. RQMS instead infers the missing reference spectra from mixed or intermediate-state samples, thereby enabling previously inaccessible features to be analysed within a unified framework. In this Tutorial Review, we show that Py-MS/RQMS can serve as a universal polymer descriptor by mapping diverse polymer materials into a common mass–temperature data space, from which structural distributions and their relationships to material properties can be extracted.

Received 1st June 2026,
Accepted 3rd September 2026

DOI: 10.1039/d6py00532b

rsc.li/polymers

Data-driven Polymer Design Group, Research Center for Macromolecules and Biomaterials, National Institute for Materials Science (NIMS), 1-2-1, Sengen, Tsukuba, Ibaraki 305-0047, Japan. E-mail: hibi.yusuke@nims.go.jp



Yusuke Hibi

Yusuke Hibi is a Senior Researcher at the National Institute for Materials Science (NIMS), Japan. His research has spanned sequence-controlled polymerisation, polymer self-assembly, materials processing and data-driven polymer analysis. Since joining NIMS in 2020, he has developed the reference-free quantitative mass spectrometry (RQMS) framework for analysing complex structural distributions in polymers. His current research focuses on autonomous polymer analysis and control.

1. Introduction

Polymers do not exist as single molecular structures, but as ensembles of molecular chains with structural distributions, and these distributions govern their properties. In addition to distributions in primary structure, such as molecular weight, copolymer composition and monomer sequence, each sample also possesses distributions in chain packing, higher-order assembly, crosslinking, degradation state and filler incorporation. These distributions ultimately appear as differences in material performance.[1–3] To visualise the relationship between such structural distributions and material performance, and to feed this information back into rational polymer design, experimental descriptors capable of quantitatively representing sample-specific distributions that cannot be captured by chemical structures alone are required.[4–8]

Pyrolysis mass spectrometry (Py-MS) has considerable potential as such an experimental descriptor.[9–11] Because it records the mass patterns of volatile fragments generated by thermal decomposition, Py-MS can be directly applied to insoluble polymers, crosslinked materials and filler-containing composites. Chemical information is encoded along the m/z axis as fragment mass patterns, whereas physical and microenvironmental information is encoded along the temperature

axis through the thermal stability of the structures from which those fragments are generated. Thus, Py-MS provides a two-dimensional experimental descriptor for reading the multi-scale structural distributions inherent in polymeric materials.

Despite its promise as a polymer descriptor, this information has not yet been fully exploited. Pyrolysis is a complex process involving multiple pathways such as main-chain scission, side-chain decomposition, and radical reactions, so even a single homopolymer can generate a wide variety of fragments. In practical materials, spectra become more complex because copolymers, additives, fillers and degradation products coexist. Moreover, because ionisation efficiencies differ among fragments in mass spectrometry, peak-intensity ratios cannot be read directly as abundance ratios. Owing to this combination of spectral complexity and limited quantitative interpretability, Py-MS has often been used as a qualitative fingerprinting technique, despite the richness of the information it contains.

High-resolution MS and Kendrick mass defect (KMD) analysis have greatly improved the interpretation of complex polymer mass spectra. In KMD analysis, the mass scale is transformed so that peaks differing only in the number of a selected repeat unit have the same small difference from the corresponding integer mass, termed the Kendrick mass defect. When plotted against the integer mass, such peaks align horizontally, whereas systematic changes in other structural units can appear as diagonal series. KMD therefore groups and visualises peaks according to their mass relationships, making repetitive patterns in complex polymer spectra easier to recognise.[12–14] Complementary chemometric approaches have been used to resolve spectral complexity on the basis of intensity co-variation. In particular, principal component analysis (PCA)-based factor analysis has long been applied to separate complicated Py-MS into latent component spectra and their generation profiles.[15] However, PCA axes are defined by statistical criteria—maximising variance under an orthogonality constraint—that do not necessarily correspond to chemically meaningful Py-MS component spectra. Factor rotation can improve chemical interpretability, but the selection of component axes often requires substantial judgement by the analyst.[15] Overall, these approaches greatly improved the interpretability of Py-MS data, but quantitative analysis of the underlying polymer structures remained challenging.

To enable such quantitative interpretation, the intensities of signals assigned to pyrolysis products have been calibrated against structural distributions determined by external nuclear magnetic resonance (NMR) measurements, often in combination with chemometric models.[16,17] These studies demonstrate the quantitative potential of Py-MS, but the accessible structural resolution is naturally linked to that of the NMR measurements used for calibration.

Reference-free quantitative mass spectrometry (RQMS) enables structural distributions encoded in complex Py-MS data to be extracted quantitatively and in an interpretable form without external structural calibration. Rather than assigning and quantifying individual peaks, RQMS represents the entire spectrum as a superposition of reference spectra, each corresponding to a monodisperse state of the structural distribution of interest, and expresses the distribution in terms of their mixing ratios (Fig. 1). For example, the sequence distribution of a random copolymer can be represented as a mixing ratio of sequence-defined polymers, each composed of a single repeating sequence motif.[18] Similarly, the degree of degradation of a weathered plastic, which is difficult to define uniquely because many degradation modes coexist, can be defined as a mixing ratio between pristine and completely degraded states.[19] In this way, structural distributions that are otherwise difficult to

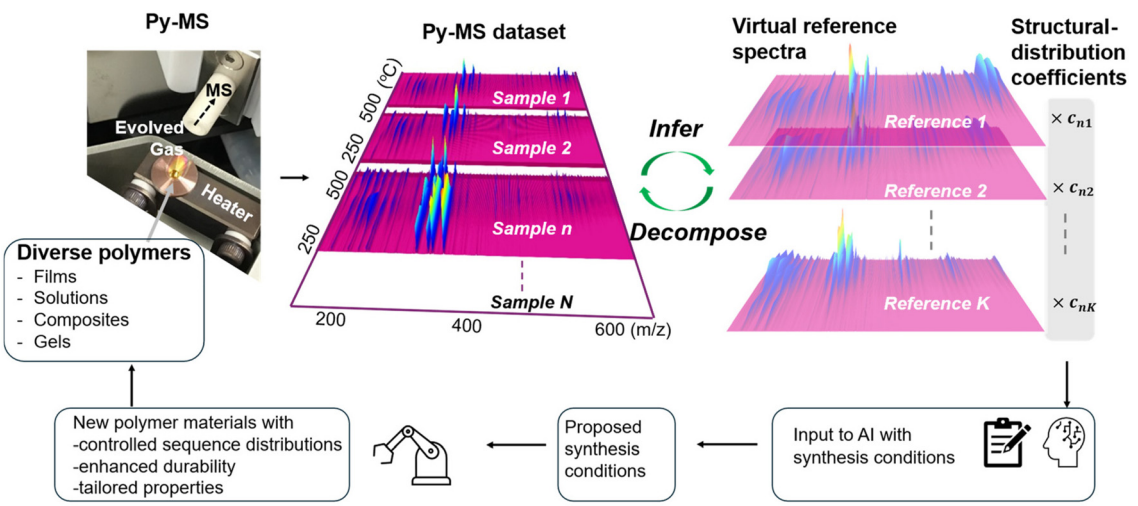


Fig. 1 Overview of Py-MS as a universal polymer descriptor for data-driven polymer development. Diverse polymer materials, regardless of their physical form, are described by the masses and generation temperatures of their pyrolysis fragments. From a dataset of N samples, RQMS infers a much smaller set of K virtual reference spectra ($N \gg K$) such that each sample spectrum can be reconstructed as their weighted superposition. Each virtual reference spectrum corresponds to a monodisperse state of the structural distribution of interest, so that their mixing coefficients represent the structural distribution of each sample. These coefficients can be combined with synthesis conditions as inputs for closed-loop optimisation.

 This journal is © The Royal Society of Chemistry 2026

measure or even define can be quantified within a unified compositional analysis framework.

The immediate difficulty, however, is whether materials corresponding to these monodisperse states can actually be prepared as references. In conventional quantitative MS, the spectra of pure components are measured in advance, and an analyte spectrum is decomposed into these reference spectra to calculate their contributions.[20,21] In the distributional problems targeted by RQMS, by contrast, materials corresponding to the monodisperse states that define the distribution are often experimentally inaccessible. Sequence-defined polymers and completely degraded polymers are typical examples. RQMS circumvents the need to prepare reference materials experimentally by inferring the corresponding virtual reference spectra from datasets composed of accessible mixed or intermediate-state samples (Fig. 1). An analyte spectrum is then decomposed into these inferred reference spectra, and the fitted coefficients provide quantitative estimates of its distribution among the analyst-defined references. At first sight, finer structural resolution could be achieved simply by defining a larger number K of monodisperse reference states. In practice, however, the corresponding reference spectra must themselves be inferred from N mixed samples, and K therefore cannot be increased arbitrarily without loss of identifiability and overfitting. The achievable structural resolution is thus intrinsically data-driven, depending on both the size and the structural diversity of the dataset.

However, representing such structural distributions as compositional mixtures over analyst-defined reference states is not always straightforward. This is particularly true when relating the multi-scale structural distributions encoded in Py-MS to material performance, because multiple structural factors may act simultaneously and the appropriate references are not necessarily obvious. Composite materials provide a representative example, in which polymer–filler interactions, filler dispersion and the crosslink density of the polymer matrix are coupled.[3] In such cases, Py-MS can instead be used directly as a material descriptor. Although this approach does not explicitly resolve the underlying structural distributions, spectral features associated with material performance can still be identified and correlated with independently measured properties. Once this relationship has been established using a training dataset, the properties of subsequently measured samples can be predicted rapidly from their Py-MS, enabling the inverse design of experimental conditions for materials with tailored properties.[22]

In this Tutorial Review, we first outline the measurement principles and instrumental configurations of Py-MS, and then introduce the basic concept of RQMS. We subsequently discuss representative applications, including monomer-sequence distribution analysis, sequence analysis of resist copolymers using temperature-domain information, quantification of the degree of degradation in plastics and stabiliser leaching, and performance prediction for composite materials. Through these examples, we aim to position Py-MS as an experimental descriptor of sample-specific structural distributions and RQMS as a framework for extracting purpose-specific quantitative distributions from it. Experiment-derived descriptors are becoming increasingly important in data-driven polymer science: automated clustering of small-angle X-ray and neutron scattering patterns guides active-learning-based exploration of formulation space;[23] nuclear magnetic resonance (NMR), gel permeation chromatography and dynamic light scattering measurements provide feedback for autonomous polymer synthesis;[8] chemometric decomposition of Raman spectra has resolved component-specific crystallisation behaviour;[24] and melt-viscosity monitoring enables real-time inspection of recycled plastics.[25] A distinctive feature of Py-MS is that its destructive measurement principle allows diverse polymer samples to be analysed largely independently of their form and solubility and mapped onto a common mass–temperature space in which structural information is encoded. RQMS can then extract purpose-specific quantitative structural distributions from this common space. In this sense, Py-MS/RQMS provides a universal descriptor and a common analytical basis for data-driven polymer science. To further assist readers, Table 1 provides a glossary of key terms used throughout this Review.

Table 1 Glossary of key terms used in this review

Term	Working definition
Spectral space	A coordinate space in which each m/z channel defines an axis and each spectrum is represented as a vector whose components are the signal intensities in those channels.
Composition space	A coordinate space in which each sample is represented by the mixing fractions of the reference states.
Reference spectrum	A spectrum corresponding to a reference state. A reference state may be either a pure component of the system or a monodisperse state used to define the structural distribution of interest.
Simplex	A convex hull whose vertices are the reference spectra in spectral space. A simplex defined by K reference spectra is $(K-1)$-dimensional.
Reference-free quantification	Quantification in which unavailable reference spectra are inferred from mixture data together with their coefficients, without requiring all reference spectra to be measured experimentally.
Data matrix	A matrix constructed by arranging multiple observed spectra as rows.
Matrix factorisation	A method that decomposes a data matrix into a small number of basis patterns and their corresponding coefficients. When both the basis patterns and their coefficients are constrained to be non-negative, the method is termed non-negative matrix factorisation (NMF).
Fragment spectrum	A basis spectrum extracted by NMF from a data matrix, comprising several coherently varying peaks. The corresponding NMF coefficient for each observed spectrum is termed the fragment abundance.
Minimum-volume NMF	An NMF approach that infers unknown reference spectra by selecting the smallest simplex that encloses the observed data.

2. Instrumental setup of Py-MS

Py-MS consists of three principal components: a pyrolysis unit, an evolved-gas transfer section and an analytical section. When a polymer sample is heated in the pyrolysis unit, volatile fragments are generated through processes such as main-chain scission and side-chain elimination. These fragments are transported to the analytical section, where mass spectra are continuously acquired. In constant-heating-rate Py-MS, the sample is typically heated at approximately 50 °C min^{-1} while mass spectra are acquired at 1–10 scans per s, providing a temperature-resolved mass-spectral dataset. Each fragment is therefore characterised by its mass and the temperature range in which it is generated.

This temperature-resolved nature is central to the use of Py-MS as a polymer descriptor. Fragment mass patterns mainly represent chemical information, whereas generation-temperature profiles provide an additional separation axis based on thermal stability. This Review therefore focuses primarily on constant-heating-rate Py-MS. Flash-heating Py-MS, most commonly Py-GC/MS, is highly effective for identifying individual pyrolysis products.[16,17,26] The combination of rapid pyrolysis, an enclosed transfer path, chromatographic separation and standardised electron ionisation (EI) provides reproducible pyrograms and mass spectra, facilitating the construction of extensive reference libraries.[27] However, flash heating compresses the thermal history into a single pyrolysis event, thereby losing temperature-resolved information on fragment generation and thermal stability. In addition, low-volatility oligomers may be lost through condensation or adsorption during passage through the GC column. By contrast, constant-heating-rate Py-MS allocates measurement time to resolving the thermal evolution of polymer pyrolysis rather than to chromatographic separation. Numerous fragments are generated within overlapping temperature ranges, resulting in highly congested mass spectra. However, the resulting mass–temperature dataset retains systematic co-variation among fragment peaks, allowing them to be computationally grouped and organised according to their shared structural origins, as described below.

Fig. 2 shows three representative instrumental implementations of constant-heating-rate Py-MS, grouped into two broader categories. The first comprises ambient-pressure configurations, in which the sample is heated near the MS inlet and the evolved gases are ionised under atmospheric conditions. Representative examples include direct analysis in real time (DART)[28] and proximal corona-discharge ionisation.[29] In both configurations, the short transfer path from the pyrolysis region to the MS inlet suppresses condensation and secondary decomposition. The principal difference between DART and proximal corona-discharge ionisation lies in the atmospheric flight distance travelled by ionised fragments before entering the MS and, consequently, in the diversity of adduct ions detected. In DART, ions travel a comparatively long distance through ambient air, during which gas-phase ion–molecule reactions and adduct exchange can occur; protonated and ammoniated species are therefore commonly observed (Fig. 2A). By contrast, proximal corona-discharge ionisation occurs in much closer proximity to the MS inlet, greatly shortening the atmospheric flight path of the ions (Fig. 2B). The short atmospheric flight path suppresses such reactions and favours the selective detection of protonated species. This simplifies spectral interpretation and reduces undesired variations in peak patterns caused by competing adduct formation. However, corona discharge can become unstable owing to variations in air dryness and deterioration or contamination of the corona needle during prolonged operation, and therefore requires frequent maintenance. DART is therefore preferable when operationally robust routine measurements are required. Both methods provide relatively soft ionisation and can detect high-m/z oligomeric fragments, in some cases above m/z 1000. Ambient-pressure soft-ionisation Py-MS is therefore particularly useful for sequence analysis, where oligomeric fragments retaining short sequence motifs are informative. A common limitation of these ambient-pressure methods is sensitivity to humidity and atmospheric composition, which may complicate long-term dataset compatibility.

The second category performs pyrolysis under a controlled atmosphere and then transports the evolved gases into the MS chamber. A representative example is thermogravimetry-coupled MS (TG-MS),[30] in which TG is used as the pyrolysis unit (Fig. 2C). Pyrolysis is conducted under an inert gas atmosphere, and the evolved gases are introduced into the MS, where ionisation is performed under high vacuum. This closed and controlled configuration is less affected by atmospheric fluctuations and generally provides long-term reproducibility. More importantly, the MS signal can be directly correlated with instantaneous mass loss, which enables correction for fragment-dependent ionisation efficiencies and improves weight-based quantification.[31] This is particularly valuable for trace-additive quantification and degradation analysis.[19,32]

TG-MS also has limitations. The transfer line from TG to the MS is often relatively long, so condensation and secondary reactions can limit the detection of high-m/z fragments. In addition, electron ionisation, which is widely used in TG-MS, induces further fragmentation and tends to bias the observed signals towards the low-m/z region.[33] Softer ionisation methods such as field ionisation have also been used to improve the detection of polymer-derived fragments, although the resulting spectral patterns can suffer from limited long-term reproducibility. Ion attachment (IA) ionisation is also attracting attention as a minimal-fragmentation ionisation method that offers high spectral reproducibility.[34,35]

Sequence analysis and material performance evaluation can be initiated relatively easily using ambient-pressure soft-ionisation Py-MS. Because DART is a well-established commercial ion source that is already available in many mass-spectrometry laboratories, a practical entry point is simply to add a compact programmable heater block to an existing DART-MS system. By contrast, TG-MS is preferable when the degradation of practical materials must be analysed together with the trace-level loss of additives, because the MS signals can be correlated

 This journal is © The Royal Society of Chemistry 2026

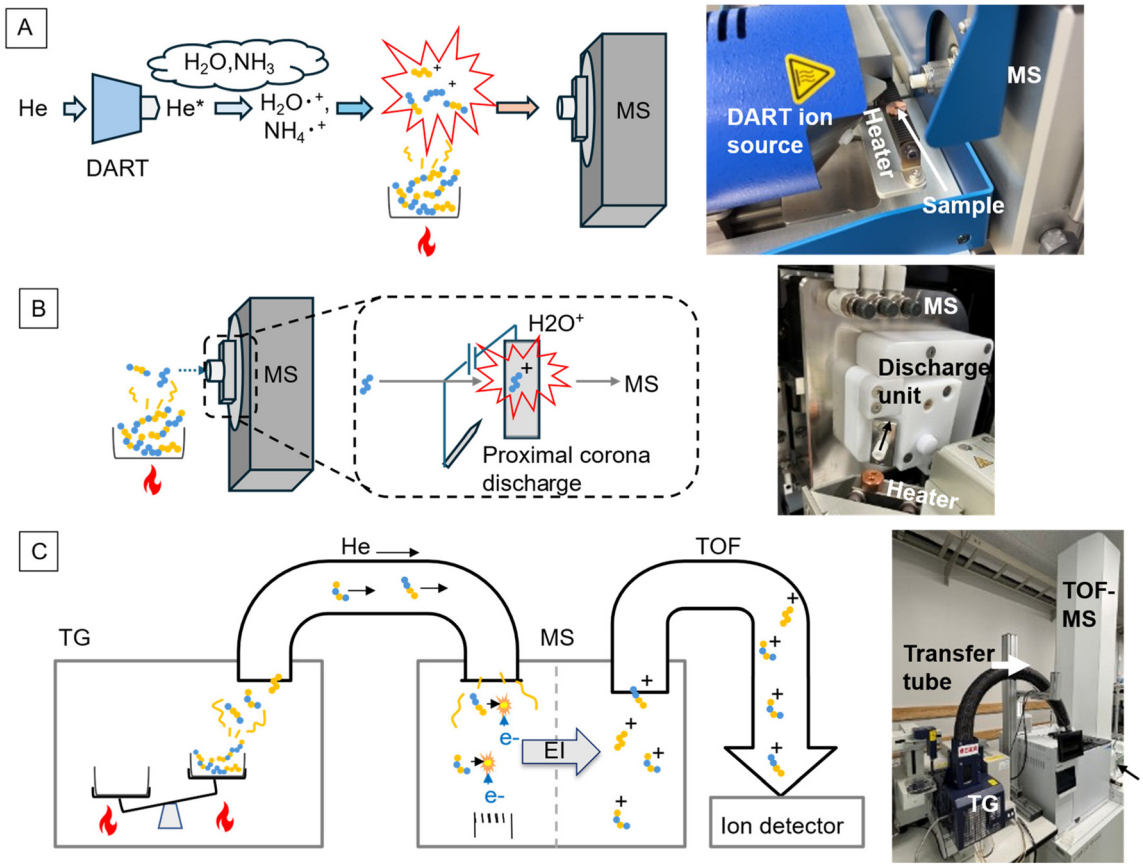


Fig. 2 Representative instrumental configurations of Py-MS. (A) Ambient DART-MS, (B) ambient proximal corona-discharge ionisation, and (C) He-atmosphere TG-MS. (A) In DART, metastable species first ionise ambient water and ammonia, which then ionise pyrolysis fragments through secondary reactions under relatively soft conditions. (B) In proximal corona-discharge ionisation, polymer fragments are ionised immediately in front of the MS inlet, minimising their atmospheric flight distance before transfer into the vacuum region of the spectrometer. (C) In TG-MS, pyrolysis is conducted under a controlled atmosphere, and the evolved gases are transferred into the MS while the accompanying mass loss is monitored simultaneously. This closed configuration provides stable datasets and supports mass-loss-correlated quantification.

with the accompanying mass loss to improve weight-based quantification. Constructing a reliable TG-MS system requires careful design of a continuously heated transfer path between the TG and MS to minimise condensation and cold spots; a dedicated integrated system is therefore generally advisable.

3. Principle of RQMS: Py-MS as reference-free compositional analysis

To clarify how RQMS is implemented in Py-MS, we first consider the reference-free quantification problem in its simplest, general form. Section 3.1 describes the linear mixing model[36,37] and its geometric interpretation in spectral space, including how unknown reference spectra can be inferred from mixed samples. Section 3.2 then extends this framework to temperature-resolved Py-MS using the two-stage NMF procedure, and section 3.3 describes the practical data processing required for its implementation.

3.1 General framework for reference-free quantification

Conventional reference-based analysis begins with experimentally measured pure-component spectra. For a three-component system, each spectrum can be represented as a raw vector in spectral space by plotting its signal intensities along the measurement channels (Fig. 3A, left). The references, $\boldsymbol{p}_1$, $\boldsymbol{p}_2$ and $\boldsymbol{p}_3$, define the vertices of a triangle, and, in the absence of noise, any mixture must lie inside it (Fig. 3A, middle). For example, a 0.5 : 0.5 : 0 mixture lies midway between $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$. A vector outside the plane violates the sum-to-one constraint of coefficients, whereas a point within the plane but outside the triangle requires a negative coefficient. Although Fig. 3 uses only three channels for visualisation, the same geometry applies to a general D-channel spectrum represented in a D-dimensional spectral space.

More generally, an observed spectrum in a system containing K components can be written as

$$x = \sum_{k=1}^{K} c_k \boldsymbol{p}_k + \text{noise} \quad (1)$$

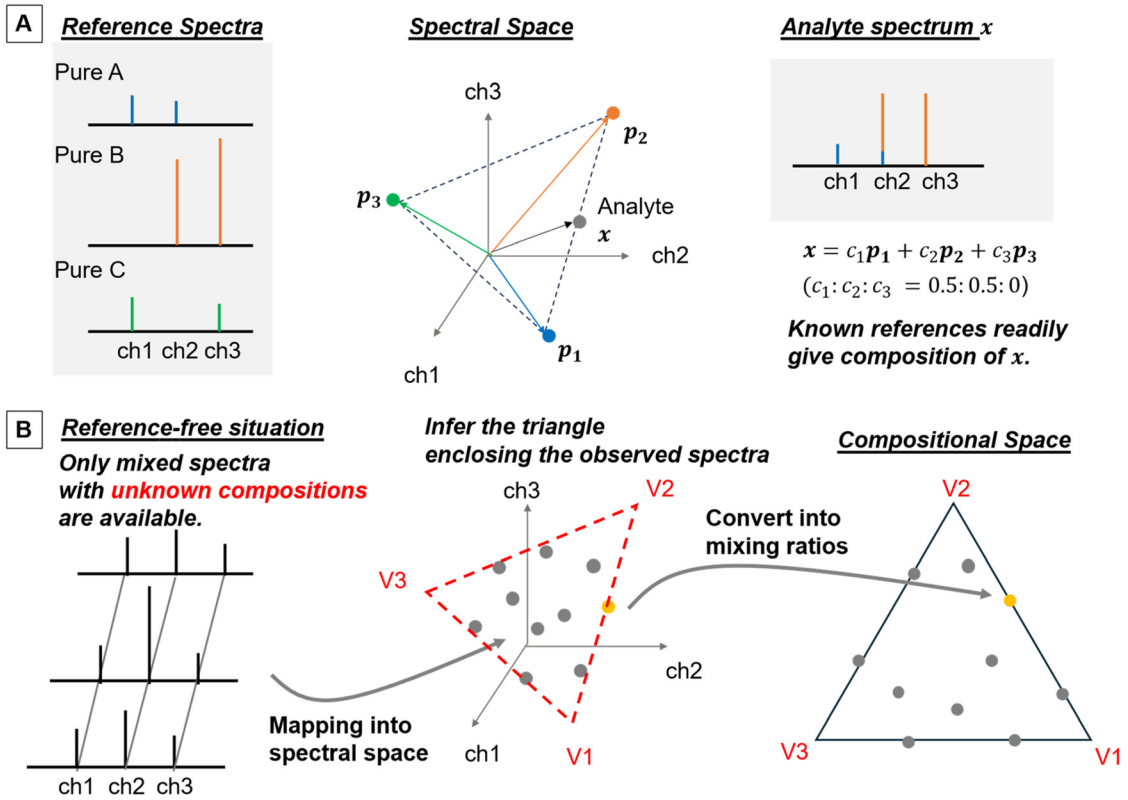


Fig. 3 Conventional reference-based spectral decomposition and reference-free inference of virtual reference spectra. (A) A spectrum is represented as a vector in spectral space by plotting its channel intensities. In a three-component system, any mixture spectrum lies inside the triangle spanned by the three measured pure reference spectra, and its composition is obtained from the corresponding linear-combination coefficients. (B) When the pure reference spectra cannot be measured, multiple observed mixture spectra are mapped into the same spectral space, and virtual reference spectra are inferred as the vertices of a triangle enclosing the observations. The resulting coefficients then provide the composition of each sample. Although a three-channel system is shown for visualisation, the same geometry applies to actual MS spectra with thousands of channels: mixtures of three components remain confined to the two-dimensional triangle spanned by the three reference spectra in the high-dimensional spectral space.

where c_k is the non-negative fraction of the k-th component, with all fractions summing to unity.

Collecting the reference vectors as the rows of a matrix $\boldsymbol{P}$ and their fractions in a vector $\boldsymbol{c}$, eqn (1) can be written as

$$\boldsymbol{x} = \boldsymbol{cP} + \text{noise}. \tag{2}$$

When $\boldsymbol{P}$ is known, $\boldsymbol{c}$ can be estimated by fitting the observed spectrum as a linear combination of the reference spectra. This is conventional reference-based analysis.

Reference-free analysis addresses the more challenging situation in which only mixture spectra are measurable, while both their mixing ratios and the pure reference spectra are unknown (Fig. 3B). A single mixture spectrum is insufficient to determine both its composition and the unknown references. However, when many mixtures with different compositions are measured, they must all lie inside the same triangle, increasingly constraining the positions of its vertices. If N observed spectra are collected as the rows of a data matrix $\boldsymbol{X}$, the complete dataset can be written as

$$\boldsymbol{X} \approx \boldsymbol{CP}. \tag{3}$$

Here, $\boldsymbol{C}$ is an $N \times K$ composition matrix, with each row giving the fractions of the K components in one sample. Its entries are non-negative, and each row sums to unity. The fractions in each row of $\boldsymbol{C}$ also specify the position of the corresponding sample in composition space, as shown in the ternary diagram in Fig. 3B, right. $\boldsymbol{P}$ is a $K \times D$ matrix whose rows are the non-negative reference spectra. Therefore, estimating $\boldsymbol{C}$ and $\boldsymbol{P}$ simultaneously from $\boldsymbol{X}$ can be formulated as non-negative matrix factorisation (NMF).[38] As more samples with diverse compositions are measured, the possible reference spectra become increasingly restricted. This is the basic principle of data-driven reference inference.

However, non-negativity and sum-to-one constraints alone do not uniquely determine $\boldsymbol{C}$ and $\boldsymbol{P}$.[39] This non-uniqueness is intuitive when the problem is viewed geometrically in spectral space. For three components, reference inference corresponds to determining the three vertices of a triangle that encloses all observed mixture spectra. However, many such triangles can be constructed by changing their size and orientation, with each triangle corresponding to a different set of candidate

 This journal is © The Royal Society of Chemistry 2026

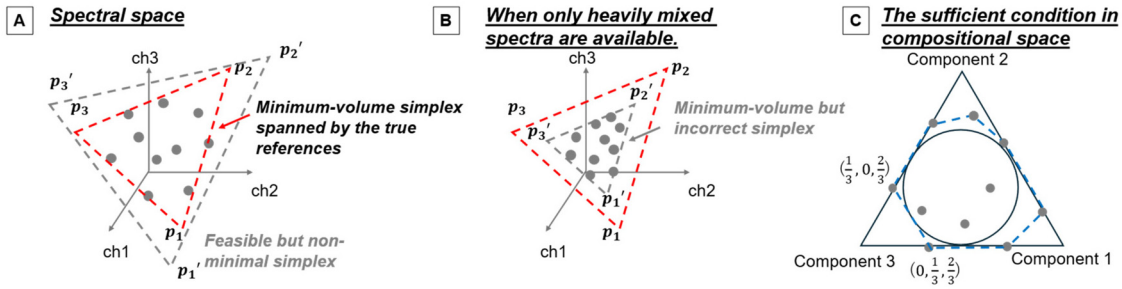


Fig. 4 Geometric intuition behind minimum-volume reference inference. (A) Many simplices can enclose the observed spectra (grey circles), and the minimum-volume constraint selects the smallest one. (B) If only heavily mixed spectra with compositions close to 1 : 1 : 1 are available, the inferred minimum-volume simplex can be incorrect. (C) The minimum-volume constraint is guaranteed to recover the true references when the convex hull of the sample compositions (blue lines) contains the incircle of the ternary composition diagram.[37] This gives a practical dataset-design guideline: each component should appear in both component-poor samples and samples containing approximately 67% or more of that component.

reference spectra (Fig. 4A). Minimum-volume NMF resolves this ambiguity by selecting the enclosing triangle with the smallest area.[37,40] More generally, for K components, the triangle is replaced by a $(K-1)$-dimensional simplex, and minimum-volume NMF selects the minimum-volume simplex that encloses the observed data.

The minimum-volume solution, however, corresponds to the true reference spectra only when the observed mixtures sufficiently constrain all reference vertices. If they are concentrated near the centre, the minimum-volume simplex shrinks inward and gives incorrect references (Fig. 4B). As this example illustrates, successful reference inference depends critically on the compositional coverage of the dataset. A sufficient condition for reliable reference recovery can be formulated in composition space. For $K = 3$, the sufficiently scattered condition is satisfied when the convex hull of the sample compositions—the blue polygon in Fig. 4C—contains the incircle of the full ternary composition diagram. More generally, for K components, the convex hull of the sample compositions must contain the inscribed sphere of the full composition simplex.[37] Chemically, each component should appear in both component-poor samples and high-fraction samples. The high-fraction guideline is approximately $2/K$ for each component; thus, it is about 67% in a three-component system and 50% in a four-component system. This is somewhat counterintuitive: as K increases, the required high-fraction level for each component decreases.[37] In this sense, the enrichment requirement for satisfying the sufficient condition becomes less stringent in higher-component systems, which can improve the reliability of reference inference when the dataset broadly covers the composition simplex.

However, because the true compositions of the observed mixtures are unknown, whether a given dataset satisfies this sufficient condition cannot generally be established in advance, and correct reference inference therefore cannot be guaranteed from the dataset alone. Importantly, the sufficiently scattered condition is sufficient rather than necessary; correct reference spectra may still be recovered even when this condition is not formally satisfied. The requirement is also expected to become less restrictive when some reference spectra are experimentally available, because these known references constrain part of the simplex directly. In practice, therefore, any experimentally accessible reference materials should be included in the dataset, together with samples that approach each remaining reference state as closely as experimental constraints permit. The inferred reference spectra and sample compositions should then be evaluated carefully against chemical expectations. An inferred reference spectrum containing chemically incompatible peaks or lacking expected peaks indicates that additional samples enriched towards that state are needed. Reference-free analysis is thus not a black-box computational procedure, but an iterative experimental–computational process that requires careful chemical judgement in both interpretation and dataset design.

3.2 Application to Py-MS: two-stage NMF

For Py-MS, applying eqn (3) directly to the observed spectra is too coarse. Py-MS gives a series of temperature-resolved spectra for each sample. If T spectra are obtained for each of N samples, the observed data are arranged as an $NT \times D$ matrix $\boldsymbol{X}$, where D is the number of m/z channels. A key difference from ordinary spectra is that Py-MS observes fragments, not polymers themselves. Even a single polymer can generate many fragments through various pyrolysis pathways. Thus, the number of independently varying fragment patterns, M, is much larger than the number of references, K. Directly decomposing $\boldsymbol{X}$ into only K reference spectra cannot capture these fragment-level variations.[18] RQMS therefore uses a two-stage decomposition. In the first stage, the observed Py-MS matrix $\boldsymbol{X}$ is decomposed into fragment abundances and fragment spectra:

$$\boldsymbol{X} \approx \boldsymbol{AS} \tag{4}$$

Here, $\boldsymbol{A}$ is an $NT \times M$ fragment-abundance matrix and $\boldsymbol{S}$ is an $M \times D$ fragment-spectral matrix. Each row of $\boldsymbol{S}$ represents a

fragment spectrum, that is, a set of m/z peaks that varies coherently across the dataset rather than a single chemical species. Fig. 5 illustrates how such fragment spectra are extracted. For homopolymers, isotope and adduct peaks from the same fragment, together with homologous oligomer peaks separated by the monomer-unit mass, are grouped when they vary coherently across the dataset (Fig. 5A). Because the relative distribution of oligomer lengths changes with generation temperature, the data are decomposed into several representative oligomer peak patterns. As evident from the raw 3D spectra, longer oligomers become more prominent at higher temperatures, allowing different oligomer peak patterns to be distinguished even when they share some m/z peaks (Fig. 5A, top and middle spectra). End-group-derived fragments, which tend to appear at lower temperatures, can likewise form a separate pattern.[41] Polymer pyrolysis proceeds through multiple pathways, including main-chain and side-chain scission, radical transfer or rearrangement, elimination and oxidation. For example, polystyrene can generate hydrocarbon, carbonyl- and aldehyde-containing fragments.[42] Under fixed pyrolysis and ionisation conditions, products derived from the same local structure tend to show approximately reproducible relative intensities and are therefore grouped into a single fragment spectrum (Fig. 5A, right). This grouping is also demonstrated experimentally in Fig. 6, where complex styrene-derived peak patterns are extracted as distinct fragment spectra shown in green. For a copolymer, sequence motifs that occur frequently generate a corresponding oligomer peak together with a characteristic set of shorter oligomer peaks in consistent relative proportions. Because these peaks vary together across the dataset, the motif-associated pattern can be extracted as a fragment spectrum (Fig. 5B). Sequence isomers overlap at the same m/z for the corresponding oligomer, but differ in the pattern or relative intensities of their shorter oligomer peaks. These differences allow their overlapping contributions to be separated into different fragment spectra.

The corresponding column a_m of $\boldsymbol{A}$ gives the abundance distribution of fragment spectrum s_m across samples and temperatures (Fig. 5). For standard compositional analysis, $\boldsymbol{A}$ is integrated along the temperature direction for each sample to give an $N \times M$ matrix $\boldsymbol{A}'$, because this analysis mainly uses the total abundance of each fragment spectrum rather than its temperature distribution. When the relevant information is encoded in temperature profiles, however, this integration step can be

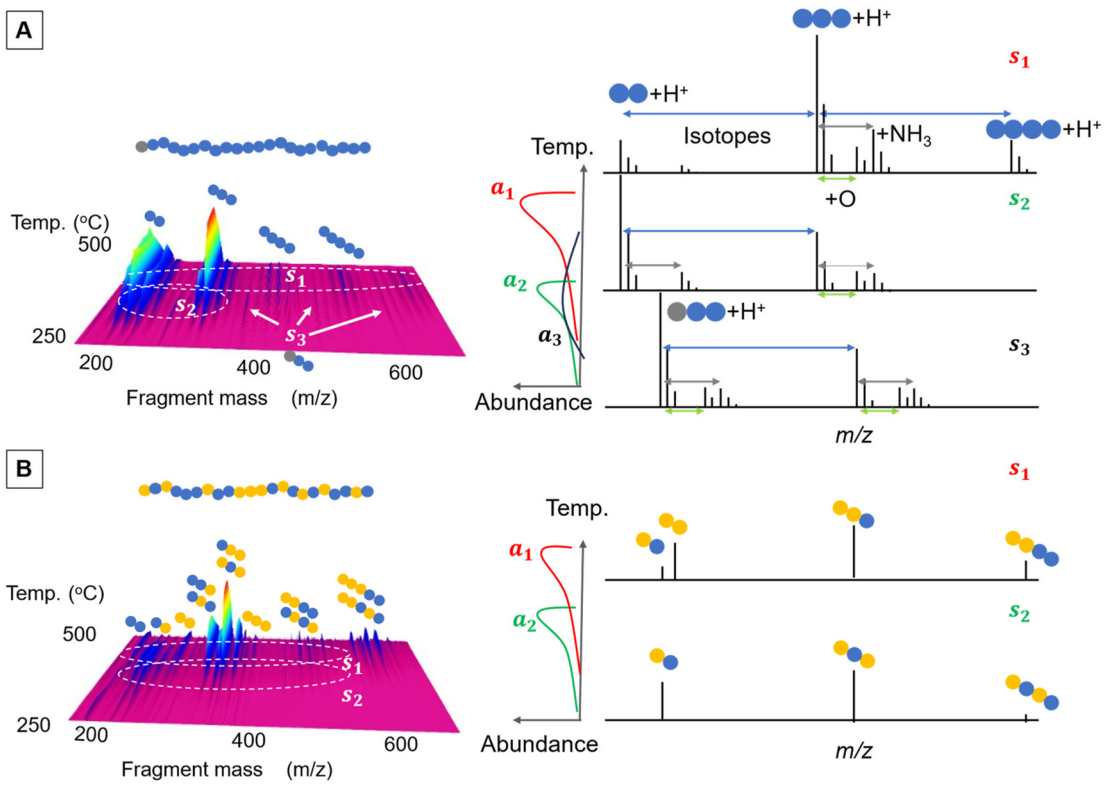


Fig. 5 Schematic extraction of fragment spectra s_m and their temperature-resolved abundance distributions a_m from Py-MS data for (A) homopolymers and (B) copolymers. (A) For a homopolymer, coherently varying peaks, including isotope and adduct peaks, homologous oligomer series with temperature-specific oligomer-length distributions, and end-group-derived peaks, are extracted as fragment spectra. (B) For a copolymer, recurring local sequence motifs, illustrated here by XXY and XYX, give characteristic combinations of a motif-containing oligomer peak and shorter oligomer peaks, which are extracted as fragment spectra. Isotope and adduct peaks are omitted in (B) for clarity. The 3D spectral maps show experimental Py-MS data, whereas the extracted fragment spectra and abundance profiles are schematic illustrations. Their decomposition is determined by coherent peak variations across the complete dataset, rather than from the displayed spectrum alone.

 This journal is © The Royal Society of Chemistry 2026

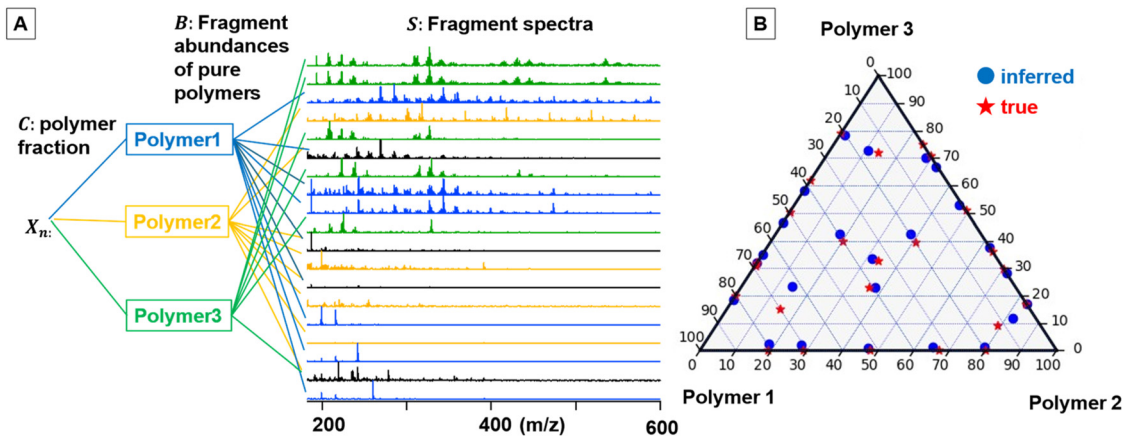


Fig. 6 Benchmark reference-free compositional analysis by RQMS. (A) Hierarchical Py-MS/RQMS model. Temperature-resolved Py-MS data are first decomposed into fragment spectra—dataset-wide patterns of m/z peaks whose intensities vary coherently across samples and temperature windows—and their temperature-resolved abundance profiles. The abundance profiles are then integrated over temperature to obtain the total abundance of each fragment spectrum for compositional analysis. (B) Inferred and ground-truth compositions of ternary polymer films. Reproduced from ref. 18 with permission from the Royal Society of Chemistry, copyright 2023.

omitted and the temperature-resolved fragment-abundance patterns can be analysed directly.

Minimum-volume NMF is then applied to $\boldsymbol{A}'$:

$$\boldsymbol{A}' \approx \boldsymbol{CB} \tag{5}$$

Here, $\boldsymbol{C}$ is an $N \times K$ composition matrix representing the contribution of each component to each sample. The matrix $\boldsymbol{B}$ is a $K \times M$ matrix whose elements represent how much of each fragment spectrum is generated by each reference. Although the first-stage decomposition is performed on the original temperature-resolved matrix $\boldsymbol{X}$, the corresponding temperature-integrated spectral matrix $\boldsymbol{X}'$ can be reconstructed from $\boldsymbol{A}'$ and $\boldsymbol{S}$. Combining the two stages therefore gives

$$\boldsymbol{X}' = \boldsymbol{A}'\boldsymbol{S} \approx (\boldsymbol{CB})\boldsymbol{S} = \boldsymbol{CP} \tag{6}$$

where $\boldsymbol{P} = \boldsymbol{BS}$ corresponds to the virtual reference spectra inferred by RQMS. In other words, RQMS first maps Py-MS data onto a fragment space and then performs reference-free compositional analysis in that space.

Here, K and M have different roles. The number of references K is defined by the analyst according to the structural distribution to be quantified. In contrast, the number of fragment-level components, M, is not known in advance and must be estimated from the Py-MS dataset. In RQMS, the first-stage decomposition is therefore started with a sufficiently large number of candidate fragment components, and redundant components are suppressed during fragment-spectrum extraction by automatic relevance determination (ARD), a sparse-modelling approach, to determine the effective value of M.[43] For initialisation, M spectra, typically 20–40, are randomly selected from the full dataset of NT spectra and used as the initial rows of $\boldsymbol{S}$. The matrices $\boldsymbol{A}$ and $\boldsymbol{S}$ are then alternately updated so as to minimise the squared error between $\boldsymbol{X}$ and $\boldsymbol{AS}$. Details of the algorithmic implementation are provided in the original RQMS reports.[18]

The two-stage NMF reflects the hierarchy latent in Py-MS (Fig. 6A): each mixed sample has mixing ratios of references; each reference has characteristic abundances of fragment spectra; and each fragment spectrum consists of a coherent set of peaks. This procedure has been validated using benchmark mixed-polymer datasets; even when the dataset contained only mixed samples with maximum component fractions of 80%, the estimated compositions agreed well with the true compositions (Fig. 6B). Taken together, this benchmark provides proof of principle for accurate composition recovery in a fully reference-free ternary system. As noted in section 3.1, higher-component systems require more extensive datasets, but less stringent enrichment towards each reference vertex, while any experimentally accessible references can further constrain the inference. Importantly, although the temperature axis is essential for extracting fragment spectra in the first-stage NMF, retaining it in the second-stage analysis does not necessarily improve interpretation. Because generation-temperature profiles may contain confounding effects, such as entangled sequence and tacticity contributions (section 4), polymer chemists should decide after fragment-spectrum extraction whether to integrate fragment abundances over temperature or retain their temperature profiles according to the analytical question.

3.3 Practical data processing

As discussed in section 2, signal intensities obtained by ambient-pressure MS are intrinsically sensitive to humidity. Even in closed TG-MS configurations, day-to-day changes in ion-source contamination and fragment-transfer efficiency can alter the overall sensitivity.[31] Careful intensity correction is therefore essential when constructing a comparative Py-MS dataset. This is particularly important for quantitative decomposition using either experimentally measured or inferred reference spectra. As illustrated by the geometry in

Fig. 3, all spectra must be placed on a common quantitative scale. Normalising each spectrum independently to unit total intensity should generally be avoided; instead, the signals should be corrected for both sample amount and instrumental sensitivity. Otherwise, the observed mixture spectra may no longer lie consistently within the simplex defined by the true reference spectra. The factorisation may then distort the inferred reference spectra, or introduce additional spurious components, in order to reproduce these non-structural intensity variations, leading to artefactual decomposition and overfitting. DART-MS can exhibit changes in spectral shape because ambient conditions alter adduct-ion ratios; overall intensity correction alone is therefore insufficient, and measurements should preferably be performed under controlled-humidity conditions. Proximal corona-discharge ionisation is less susceptible to adduct-ratio variations because it predominantly generates protonated species.

Accurate weighing of both the polymer sample and an internal standard provides a practical basis for this correction. Dividing the measured spectrum by the polymer weight W_p expresses the signal on a unit-polymer-mass basis. Day-to-day changes in sensitivity can then be corrected using an internal standard of known weight W_s and measured intensity I_s. The measured polymer spectrum is therefore multiplied by the correction factor $W_s/(W_pI_s)$. The internal standard should not react or interact strongly with the polymer, should remain stable during sample handling, and should produce signals at lower temperature than polymers to avoid peak overlapping. Biphenyl has been used for ambient-pressure measurements, whereas the less volatile 4,4′-di-*tert*-butylbiphenyl is useful when sublimation during TG-MS sample handling is a concern.

By measuring the internal standard alone and including the resulting spectrum in the dataset, signals from the standard and reproducible background can be extracted as fragment spectra by the first-stage NMF because they appear consistently across measurements. This treatment is particularly important for ambient-pressure MS, in which siloxane oligomer peaks derived from vacuum-pump oil and environmental plasticiser-derived peaks may persist as systematic background. By excluding the corresponding fragment abundances from subsequent analysis, these systematic backgrounds can be appropriately removed. Further procedural details are provided in the section "Derivation of the canonical correlation analysis filter" in the original report.[18]

In addition to intensity normalisation, accurate alignment of the m/z axis is essential for constructing the data matrix $\boldsymbol{X}$. Even a small mass shift can cause the same ion to occupy different channels across measurements and consequently be interpreted as artificial spectral variation during factorisation. The appropriate representation of the m/z axis—raw profile or centroided data—should therefore be selected according to the resolving power of the MS. Low-resolution quadrupole MS produces relatively broad peaks, typically with full widths at half maximum of approximately 0.5–0.8 m/z, while the apparent peak position can fluctuate by up to ±0.2 m/z between individual scans in our measurements. In the present workflow, the scans falling within each of the T temperature windows are summed before construction of the data matrix, thereby reducing random scan-to-scan variation in peak position. The resulting spectra should not be rounded to integer m/z values, because ions with the same nominal mass can differ in fractional mass and affect the position and shape of the integrated profile. For example, $C_{40}H_{40}$ corresponding to styrene pentamer, and the nominally isobaric saturated hydrocarbon $C_{37}H_{76}$ have monoisotopic masses of 520.313 and 520.595 Da, respectively, differing by 0.28 Da. Although these ions cannot be resolved as separate peaks by unit-resolution MS, changes in their relative contributions can alter the position, width, or asymmetry of the overlapping profile and thereby contribute to multivariate spectral decomposition. Low-resolution quadrupole data should therefore be retained as temperature-window-integrated profile spectra rather than independently centroided.

For high-resolution TOF-MS, by contrast, profile data may contain hundreds of thousands of m/z channels, making direct factorisation computationally inefficient. Centroided data are therefore more practical. In our measurements, mass drift is largely suppressed by correcting the m/z axis against the internal-standard peak during each measurement, leaving residual shifts within ±0.003 m/z. Peaks lying within this range are assigned to a common analysis channel represented by the modal m/z value across the dataset, making the remaining mass-axis variation negligible for subsequent analysis. Overall, raw profile data are generally preferable for unit-resolution quadrupole MS, whereas drift-corrected centroided data are more suitable for high-resolution TOF-MS.

The processed data can be exported from the instrument software in CDF format and imported into Python using the netCDF4 library. The subsequent analysis described in section 3.2 can be implemented following the algorithms presented in the original RQMS reports.[18]

4. Polymer sequencer for monomer-sequence distributions

The sophisticated properties and functions of biological macromolecules arise from their precisely defined monomer sequences. Sequence-defined synthetic polymers have therefore attracted considerable interest as a route to advanced functions. One prominent demonstration is molecular information storage, in which polymers constructed through iterative coupling reactions encode digital information in prescribed monomer sequences. This information can subsequently be decoded by tandem MS, in which a mass-selected precursor ion is fragmented and the monomer order is reconstructed from the mass differences among the resulting product ions.[44–49]

Such monomer-by-monomer sequence definition is substantially more difficult in chain-growth polymerisation, because successive additions at the active chain end are gov-

 This journal is © The Royal Society of Chemistry 2026

erned by propagation kinetics rather than individually programmed coupling steps. Although several ingenious approaches to sequence control in radical polymerisation have been demonstrated,[50–53] broadly applicable sequence definition has not yet been achieved. Nevertheless, recent studies have shown that short sequence motifs generated statistically during copolymerisation can govern specific polymer properties and functions.[1,54,55] These findings highlight the need to quantify sequence distributions in terms of the occurrence frequencies of sequence motifs such as triads and pentads.

This analytical target is fundamentally different from that of sequence-defined polymers. For a sequence-defined polymer, a single monomer order is prescribed synthetically and can be reconstructed from the mass pattern of an MS/MS fragment series; quantitative interpretation of fragment intensities as motif abundances is unnecessary. In a statistical copolymer, by contrast, multiple sequences coexist within the same sample. Analysis therefore requires not only identification of motif-related fragments but also quantitative determination of the abundance of each motif from the overall fragment pattern. The RQMS-based polymer sequencer is designed for this latter task: it quantifies the distribution of short sequence motifs across the chain ensemble rather than reconstructing the monomer order of an individual chain.

Py-MS appears well suited to implement such a sequencer. Thermal decomposition of a copolymer can generate oligomeric fragments that retain local monomer connectivity, and their masses reflect the monomer types and numbers contained in each fragment. However, peak-wise interpretation is insufficient for quantitative sequence analysis. Peak intensities are affected by fragment-dependent ionisation efficiencies and pyrolysis yields. Sequence isomers can share the same mass. In addition, pyrolysis does not selectively generate fragments of the desired sequence length; monomers, dimers, longer oligomers, and secondary decomposition and oxidation products are formed together. Under the ambient-pressure soft-ionisation conditions used in the studies reviewed here, monomeric to trimeric fragments generally give the most intense signals. This likely reflects the fragmentation pathways as well as the higher volatility and transfer efficiency of smaller products.

The RQMS-based polymer sequencer bypasses peak-wise assignment by comparing whole fragment patterns: an analyte copolymer spectrum is interpreted as a virtual mixture of spectra from sequence-defined polymers (Fig. 7). For an X/Y binary copolymer, triad analysis can be based on sequence-defined polymers such as $(XXX)_l$, $(XXY)_l$, $(XY)_l$, $(YYX)_l$ and $(YYY)_l$. Each reference polymer represents not a single triad fragment, but the full fragment pattern generated by pyrolysis of that sequence-defined polymer. This formulation absorbs several otherwise intractable factors into the virtual reference spectra: regioselective main-chain cleavage, mass overlap between sequence isomers, secondary reactions during pyrolysis and fragment-specific ionisation efficiencies. The analyte spectrum is then reconstructed as a linear combination of these virtual reference spectra, and the mixing coefficients are interpreted as the distribution of the corresponding sequence motifs. When this reconstruction is performed using a limited number of reference spectra, the fit is dominated by the relatively intense low-mass fragment peaks, mainly those of monomers, dimers and trimers. Dataset-wide variations in weaker higher-mass oligomer peaks are therefore not decomposed into independent components, and the sequence information encoded in these peaks is not quantified accurately. Resolving these weaker variations requires more references: K is 5 for binary triads, 9 for binary pentads and 13 for ternary triads.[18] Although the corresponding sequence-defined polymers generally cannot be synthesised, RQMS can infer their virtual reference spectra from experimentally accessible random-copoly-

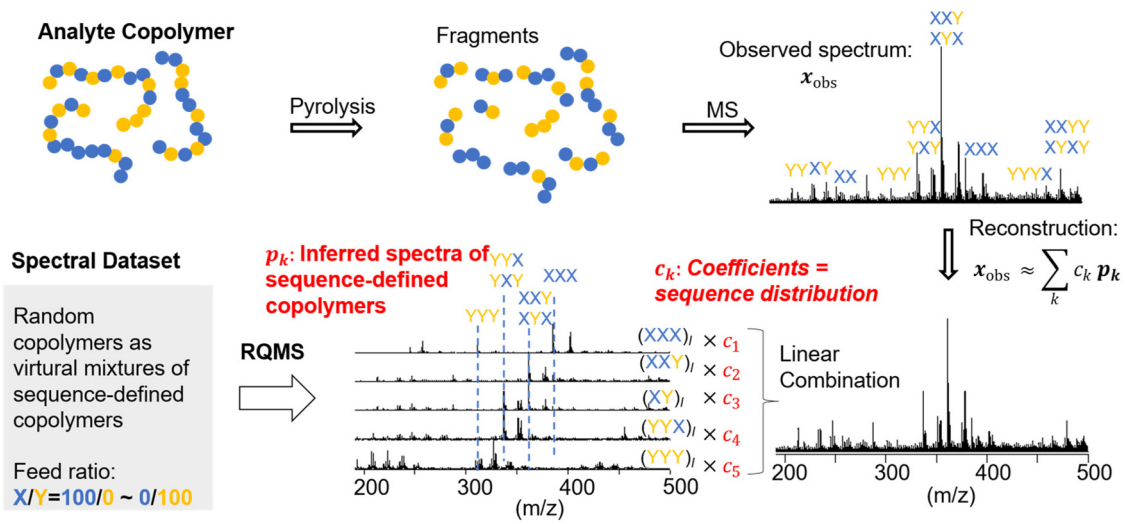


Fig. 7 Principle of RQMS polymer sequencing. The analyte spectrum is reconstructed as a virtual mixture of inferred temperature-integrated spectra corresponding to sequence-defined polymers. Reproduced from ref. 18 with permission from the Royal Society of Chemistry, copyright 2023.

mer datasets prepared with different feed ratios. K cannot be increased arbitrarily, because the virtual references must be sufficiently constrained by the available dataset. As discussed in section 3.1, this requires some samples enriched in the corresponding target sequence motifs. The polymer sequencer is therefore intrinsically data-driven: the complexity of sequence information that can be resolved increases with the size and sequence diversity of the dataset. This may provide access to deeper sequence information across a broader range of monomer systems than conventional NMR-based analysis.

As an example, Fig. 8 shows pentad analysis of styrene (S)/*n*-butyl acrylate (B) copolymers. Nine sequence-defined polymers were defined as monodisperse references to describe sequence distribution up to pentads. A dataset was constructed from 80 random copolymers prepared with different feed ratios and polymerisation temperatures, and RQMS was used to infer the corresponding virtual reference spectra. These reference spectra were then applied to S/B 1 : 1 copolymers sampled during reversible addition–fragmentation chain transfer (RAFT) copolymerisation.[56] At the early stage of polymerisation, S-rich pentads were dominant. As polymerisation progressed and S was consumed, B-rich pentads became more abundant. To validate the result, the RQMS-derived pentad distribution was converted into B-centred triad fractions and compared with ^{1}H NMR analysis and theoretical prediction based on the Alfrey–Mayo equation.[57] The time evolution of the triad distribution agreed well with both NMR and kinetic prediction, supporting the validity of the RQMS sequencer.

NMR sequence analysis is limited to monomer combinations in which neighbouring units induce resolvable sequence-sensitive peak shifts. The RQMS-based sequencer is not subject to this requirement in principle. This advantage is particularly important for copolymers composed of monomers with the same main-chain backbone, where NMR-based sequence analysis is extremely difficult. Methacrylate-based resist copolymers provide a representative example. However, these resist copolymers do not readily generate sequence-informative fragments in Py-MS. In ArF resist model copolymers composed of 2-methyl-2-adamantyl methacrylate (M) and γ-butyrolactone methacrylate (G), the major signals arise from M-side-chain fragments, depolymerised G monomer and G-side-chain fragments, while oligomeric peaks retaining short-sequence information are barely observed (Fig. 9A).[58]

Nevertheless, the loss of sequence-informative oligomer peaks does not mean that sequence information is completely lost. As the G fraction increases, the temperature profile of the M-side-chain fragment shifts to lower temperatures and becomes bimodal, indicating that neighbouring G units destabilise the M side chain (Fig. 9B). Thus, even when the m/z axis no longer retains sequence-informative oligomer peaks, sequence-dependent thermal stability can remain encoded in the temperature axis. Therefore, the temperature dimension is retained instead of being integrated before the second-stage NMF (see section 3.2). The temperature-dependent fragment abundances extracted from M/G random copolymers are decomposed into virtual sequence-defined polymer bases, such as $(MMM)_l$, $(MMG)_l$, $(MG)_l$, $(MGG)_l$ and $(GGG)_l$. The mixing coefficients give the triad-level sequence distribution, and the inferred reference profiles reveal sequence-dependent fragment thermal stability.

The RQMS-unmixed temperature profiles showed that both local sequence and tacticity affect the thermal stability of the M side chain. The inferred profiles, supported by density functional theory calculations, indicated deprotection sensitivities in the order MMM-*rr* > MMG > GMG > MMM-*rm*/*mm*, corresponding to an energy difference of approximately 10–15 kJ mol^{-1}. The $(MMM)_l$ profile was bimodal, whereas the inferred profiles of the G-containing virtual sequence-defined polymers

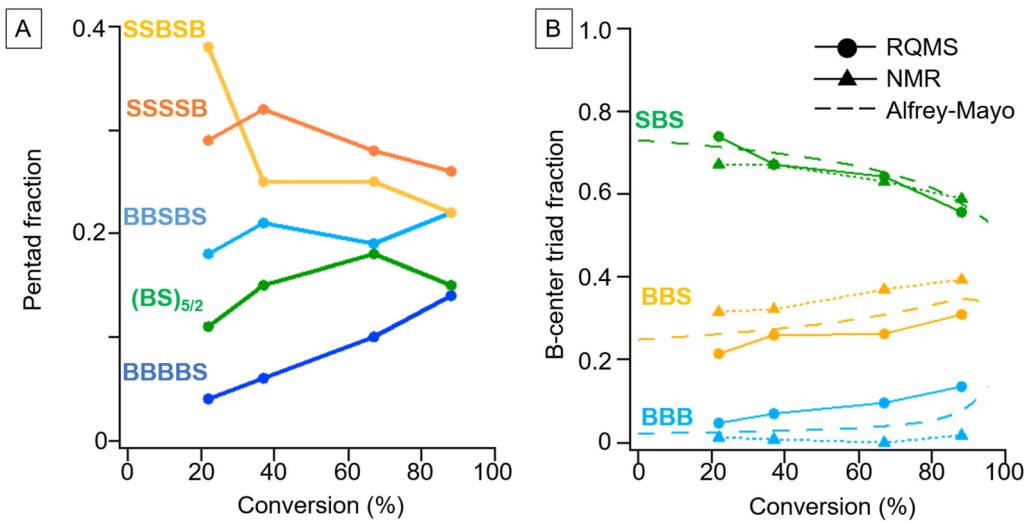


Fig. 8 RQMS sequencing of styrene (S)/*n*-butyl acrylate (B) copolymers during 1 : 1 RAFT copolymerisation. (A) Pentad sequence distributions during polymerisation. Negligible blocky motifs are omitted for clarity. (B) Comparison of B-centred triad fractions with ^{1}H NMR and Alfrey–Mayo prediction. Reproduced from ref. 18 with permission from the Royal Society of Chemistry, copyright 2023.

This journal is © The Royal Society of Chemistry 2026

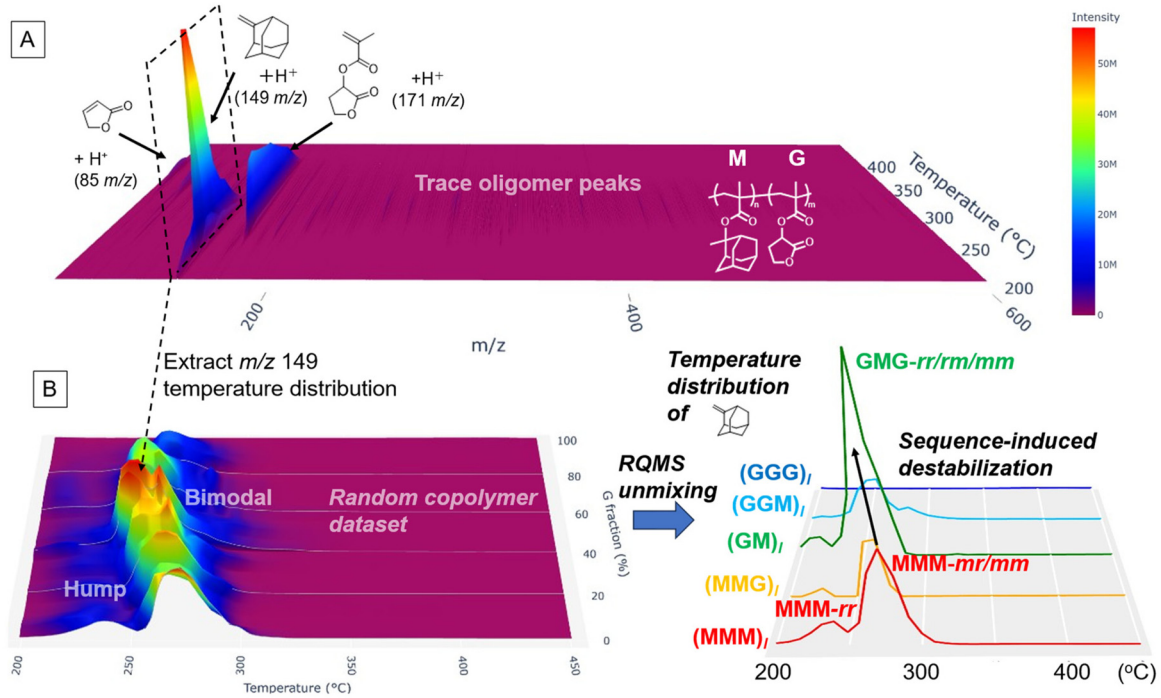


Fig. 9 RQMS sequencing of resist copolymers using temperature profiles. (A) Py-MS spectrum of an M/G 1 : 1 copolymer. (B) Temperature profiles of the M-side-chain fragment in copolymers prepared with various feed ratios and RQMS-unmixed temperature profiles of virtual sequence-defined copolymers. Reproduced from ref. 58 under the terms of the CC BY 4.0 license.

were relatively sharp and unimodal. This contrast suggests that tacticity has a smaller influence on M-side-chain decomposition when an M unit is adjacent to G. Although the detailed origin of the G-induced destabilisation remains unclear, the strongly electron-withdrawing substituent of the neighbouring G unit may electronically destabilise the acid-labile tertiary ester of M. These large sequence- and tacticity-dependent differences in deprotection sensitivity have direct implications for resist performance. Under comparable local acid exposure during post-exposure baking, polymer segments with different sequences and tacticities may undergo deprotection at different rates and to different extents. Because the dissolution rate of a chemically amplified resist depends strongly on the degree of deprotection, such molecular-scale heterogeneity can produce a spatially nonuniform dissolution front at the pattern boundary and thereby increase line-edge roughness.

As discussed in section 3.1, successful recovery of the reference spectra depends on how broadly the dataset spans the compositional space (Fig. 4). For sequence analysis, the dataset should include both homopolymers and copolymers prepared over a wide range of feed compositions. Importantly, copolymerisation should be stopped at low conversion for each feed composition. Because the monomers generally have different reactivities, their relative concentrations in the reaction mixture change as polymerisation proceeds. Polymers formed at early and late stages therefore have different sequence distributions, and their accumulation broadens the overall sequence distribution of the recovered sample. Limiting conversion suppresses this broadening and preserves a sequence distribution that more closely reflects the monomer composition at the beginning of polymerisation. At the same time, the degree of polymerisation should be sufficiently high, preferably 100 or higher, to minimise contributions from end-group-derived fragments. Conventional free-radical polymerisation may therefore be advantageous for constructing such datasets, because high-degree-of-polymerisation copolymers can be prepared rapidly even at low conversion.

Careful dataset design improves reference identifiability but does not guarantee correct recovery; the inferred sequence distributions must therefore be independently validated. When available, NMR and kinetic predictions provide direct benchmarks, as demonstrated in Fig. 8B. A simpler and more broadly applicable validation is to compare the overall monomer composition calculated from the inferred sequence distribution with that independently measured by NMR.[58] Although monomer composition does not uniquely determine a sequence distribution, a valid sequence distribution must reproduce the measured composition throughout the dataset. The inferred reference spectra should also be assessed for chemical consistency. For example, a virtual reference based on the repeating motif $(XXYXY)_I$ should not contain a peak attributable to an XXX fragment, whereas the reference for $(XXXYY)_I$ should contain such a peak, even though both pentad motifs have the same overall composition, X_3Y_2, and therefore the same mass.[18]

The effects of sequence and tacticity may also be entangled. Even for the same sequence, differences in stereochemical configuration can alter fragment selectivity and generation-temperature profiles. The M/G resist copolymers provide a clear example: the generation-temperature profiles of the M-side-chain fragment differed substantially between MMM-*rr* and MMM-*rm*/*mm* (Fig. 9B right). Strictly, these tacticity states should therefore be represented as separate references rather than as a single MMM vertex. In the present sequence-analysis framework, however, inclusion of the M homopolymer strongly anchors the MMM reference to the weighted average of the tacticity states present in that homopolymer. This should not cause substantial difficulty if MMM motifs embedded in the copolymers have approximately the same tacticity distribution as those in the homopolymer, because all MMM contributions can then be represented by a common reference spectrum. Although this approximation may hold for many radical copolymerisations, it is not guaranteed for sterically demanding monomers such as M. The tacticity distribution associated with an MMM motif may depend on the surrounding monomer sequence beyond the triad, so that MMM motifs in an M/G copolymer need not have the same *rr*/*rm*/*mm* distribution as those in the M homopolymer.

In the M-rich region of the dataset, the monomer composition back-calculated from the RQMS-derived sequence distribution was reported to deviate from that determined by NMR, indicating underestimation of the MMM fraction.[58] This discrepancy may partly reflect the tacticity-dependent reference mismatch described above. Conversely, independently measured monomer compositions could be used to constrain the allowable sequence fractions and thereby reduce the influence of tacticity-dependent reference mismatch on sequence quantification. When sufficient sequence information is retained in the mass-domain fragment patterns, a more straightforward approach is to integrate the first-stage fragment abundances over the temperature axis before the second-stage NMF (section 3.2). Because tacticity-dependent differences are expressed mainly in the generation-temperature profiles, this procedure may substantially reduce their influence on sequence quantification.

NMR analysis of copolymers is often complicated because sequence and tacticity simultaneously contribute to peak splitting. However, even partial prior knowledge of the sequence distribution can greatly simplify peak assignment.[59] Accordingly, RQMS could first provide a sequence distribution averaged over the tacticity states. Combining this information with NMR analysis may then enable more complete and separate quantification of sequence and tacticity distributions.

Unlike kinetic approaches based on monomer reactivity ratios,[57] the polymer sequencer directly quantifies sequence distributions in the generated polymer and therefore does not require prior knowledge of growth-end selectivity. This direct readout could ultimately be fed back during polymerisation to adjust the reaction conditions towards a target sequence distribution. Statistical sequences would then become experimentally observable and controllable design variables, defining a middle ground between fully sequence-defined synthesis and conventional random copolymerisation.[60–62]

5. Extensions of RQMS: TG-coupled quantification and nonlinear systems

Practical polymer materials often contain minor additives that strongly affect performance. Weathering stabilisers are typically added at concentrations of only a few wt% or less, yet their concentration and loss during use can determine material lifetime. RQMS for practical materials must therefore quantify not only the major polymer components, but also trace components at the 0.1 wt% level. At the same time, additives can interact with polymers and alter the pyrolysis response itself. Two extensions are therefore important: thermogravimetry-synchronised RQMS (TG-RQMS) for accurate trace quantification and nonlinear RQMS (NL-RQMS) for systems with interaction-induced spectral distortion.[31,63]

TG-RQMS improves quantitative accuracy by coupling thermogravimetry with MS. MS peak intensities do not directly represent weight fractions because ionisation efficiencies differ among fragments. A minor component with high ionisation efficiency can produce a disproportionately large peak, while a major component with low ionisation efficiency can appear weak. TG-RQMS synchronises the instantaneous mass loss measured by TG with the MS spectra, allowing MS intensities to be converted into weight-based fragment abundances. This corrects "visibility" differences in MS and improves RQMS quantification. In a ternary polymer system composed of styrene (S), methyl methacrylate (M) and ethyl methacrylate (E), the composition error decreased from 4.3 wt% to 1.3 wt% (Fig. 10A). Trace components at the 0.1 wt% level were also quantified without reference spectra (Fig. 10B).

A second problem arises when components react or interact before or during pyrolysis. In such cases, a mixed spectrum is no longer a simple linear combination of pure-component spectra. New fragments can appear upon mixing, and fragments derived from the original components can decrease. If such a spectrum is forced into a linear model, newly generated peaks are incorrectly assigned to one of the pure components, leading to biased composition estimates.

NL-RQMS addresses this problem by adding cross-component terms that represent reactions or interactions between components.[64] This separates fragments that decrease upon mixing from fragments that increase through reaction or interaction. In a reactive epoxy/amine ternary mixture, the composition error decreased from approximately 6 wt% with linear RQMS to approximately 3 wt% with NL-RQMS (Fig. 10C). The method also extracted, without reference spectra, epoxy-derived fragments consumed by reaction, tertiary-amine-related fragments generated by the amine–epoxy reaction, and cyclic fragments enhanced through interactions with hydroxyl groups (Fig. 10D). Together, TG-RQMS and NL-RQMS extend reference-free compositional analysis beyond the ideal linear case: TG coupling corrects fragment-dependent differences in

 This journal is © The Royal Society of Chemistry 2026

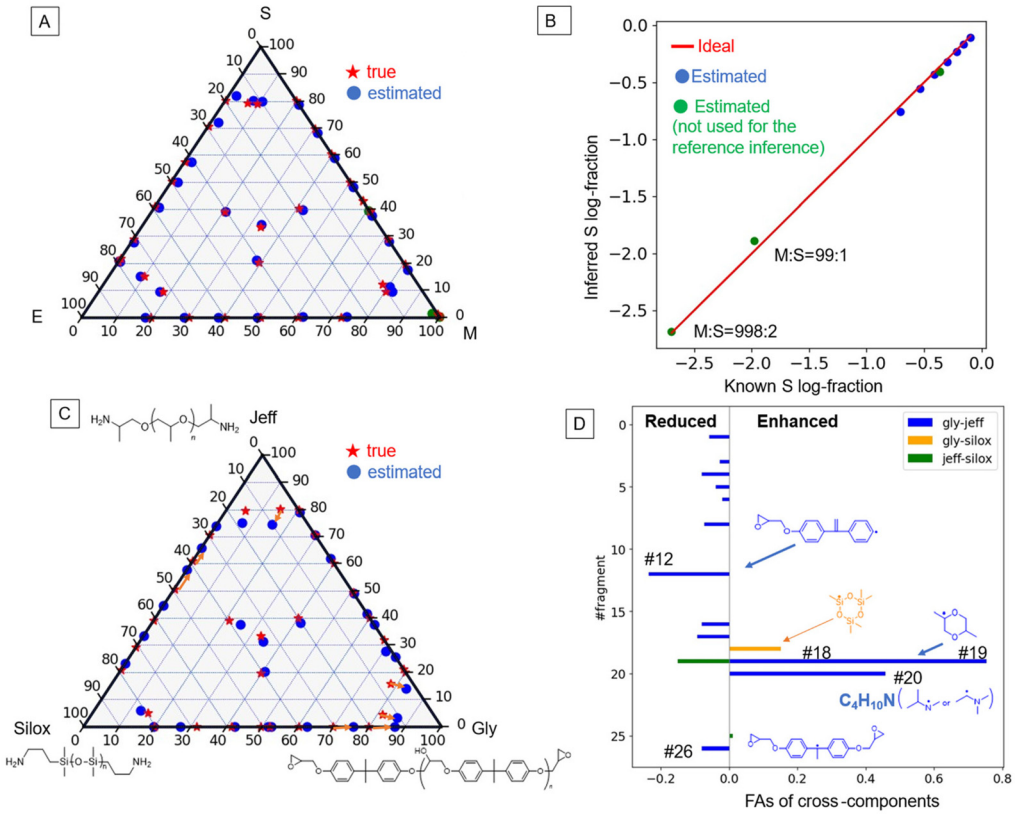


Fig. 10 RQMS extensions for practical polymer analysis. (A and B) TG-RQMS improves quantitative accuracy and enables trace quantification; the compositional error decreased from 4.3 wt% to 1.3 wt%, and 0.1 wt% components were quantified without reference spectra. (C and D) NL-RQMS analyses reactive mixtures with nonlinear spectral distortion; in an epoxy/amine ternary mixture, the error decreased from approximately 6 wt% to 3 wt%. Cross-component terms identify fragments enhanced or reduced by reactions and interactions. Adapted from ref. 31 and 63 with permission from The Royal Society of Chemistry, copyright 2024.

MS visibility on a weight basis, whereas nonlinear terms account for spectral changes caused by reactions and interactions between components. These extensions enable RQMS to address trace components and chemically interacting mixtures encountered in practical polymer materials.

6. Defining polymer degradation degree by RQMS

Plastic degradation proceeds through multiple overlapping processes, including oxidation, main-chain scission, cross-linking, side-chain modification, small-molecule elimination and additive leaching (Fig. 11A).[65,66] The dominant pathway depends on the service environment. The "degree of degradation" is therefore difficult to define from pathway-specific chemical indices, because such indices are neither directly comparable nor additive across different degradation processes. Mechanical properties are also insufficient as early degradation indicators. Hardness, modulus and strength often remain almost unchanged while chemical damage accumulates inside the material. In structural plastics, properties may collapse only after chemical degradation exceeds a critical level. Relying only on mechanical properties can therefore miss the early stages preceding sudden failure.

A widely used chemical index is the carbonyl index obtained by infrared (IR) spectroscopy, which detects oxidation-induced carbonyl formation. However, it cannot cover diverse degradation pathways, is difficult to apply to polymers that already contain carbonyl groups, and lacks sufficient sensitivity for tracking additives present at a few wt% or below.[67]

RQMS approaches this problem differently. Instead of tracking the time evolution of individual Py-MS peaks for each degradation mode, it defines degradation as a compositional transition between two polymer states: the pristine polymer and the completely degraded polymer (CDP) expected after long-term exposure under the target service environment (Fig. 11B). An actual degraded sample is then represented as an intermediate mixture between these two states. The CDP should therefore not be regarded as a single molecularly defined species, but as a condition-specific analytical endpoint representing the ensemble of structural changes generated under the target environment. Accordingly, samples exposed under different service environments should not be combined in a single pristine–CDP RQMS model, because each environment may generate a different balance of competing degra-

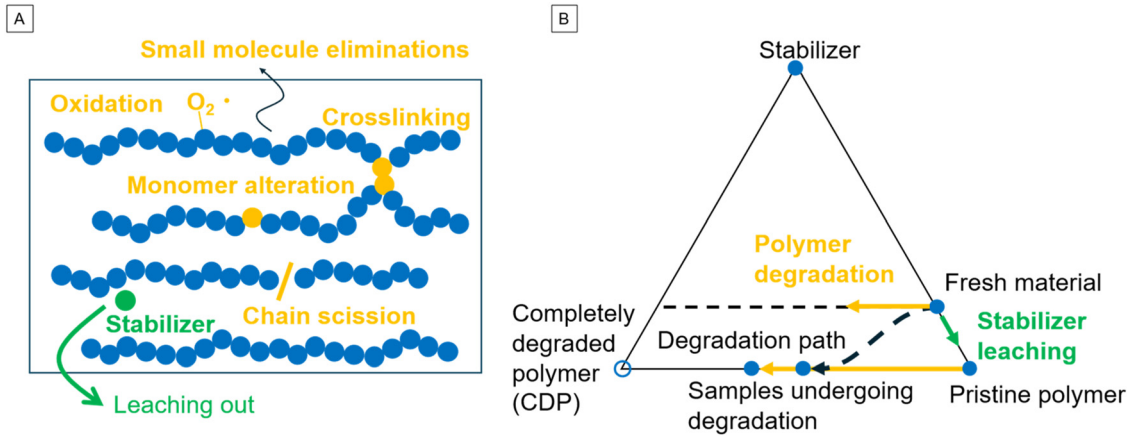


Fig. 11 RQMS definition of polymer degradation degree. (A) Multiple degradation pathways in plastics. (B) Ternary representation of chemical degradation and stabiliser leaching. Adapted from ref. 19 under the terms of the CC BY 4.0 license.

dation pathways and therefore define a distinct CDP endpoint. Although this definition integrates competing degradation pathways into a single quantitative index, comparison of the inferred CDP and pristine reference spectra may identify fragment groups that increase or decrease during degradation, providing clues to the underlying pathways. Interpreting these spectral differences is important for understanding degradation mechanisms and deriving molecular-design guidelines for improved durability. However, even with accurate-mass information, RQMS alone cannot generally identify individual degradation pathways uniquely, and mechanistic interpretation requires prior knowledge of degradation chemistry and, where necessary, complementary analyses. This formulation also allows additive loss to be analysed simultaneously: degradation behaviour is captured as a trajectory in a ternary composition diagram spanned by pristine polymer, CDP and hindered amine light stabiliser (HALS). Movement from pristine polymer towards CDP represents chemical degradation, whereas movement away from the HALS vertex represents stabiliser leaching (Fig. 11B).

Because minor additives can affect both pyrolysis behaviour and degradation reactions, the analysis was performed using TG-NL-RQMS. In accelerated degradation tests of crosslinked polyurethane films, the samples moved towards the CDP vertex in the ternary composition diagram, even though degradation time was not used as an input variable (Fig. 12A). This movement was strongly suppressed in HALS-containing films, quantifying the stabilising effect of HALS as a lower chemical degradation degree. At the same time, movement away from the HALS vertex visualised stabiliser leaching during degradation.

The RQMS-defined chemical degradation degree also acted as a leading indicator of mechanical failure. In polyurethane films, indentation depth and indentation modulus changed only slightly until the CDP fraction reached approximately 50%. Beyond this threshold, indentation depth decreased abruptly and modulus increased (Fig. 12B), indicating a tran-

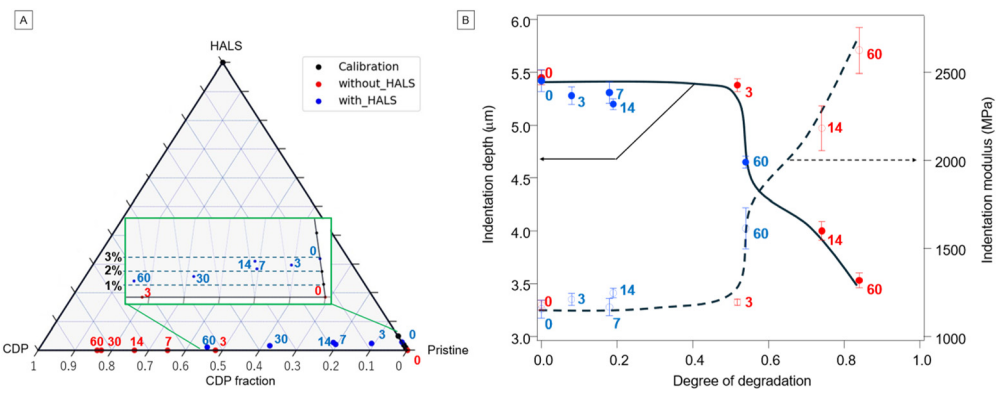


Fig. 12 Degradation behaviour of polyurethane films visualised by RQMS. (A) Ternary plot of degraded polyurethane films represented by pristine polymer, completely degraded polymer (CDP) and HALS stabiliser components. Numbers indicate degradation duration in days for films without additive (red) and with HALS (blue). (B) Relationship between RQMS-defined degradation degree and mechanical properties, showing abrupt changes in indentation depth and indentation modulus beyond a critical degradation level. Reproduced from ref. 19 under the terms of the CC BY 4.0 license.

 This journal is © The Royal Society of Chemistry 2026

sition to a harder and more brittle state. Thus, mechanical-property changes are delayed outputs of chemical degradation, whereas the RQMS-defined CDP fraction captures the preceding chemical changes.

7. Predicting composite-material properties using Py-MS descriptors

The RQMS applications discussed so far show that complex phenomena can often be clarified by recasting them as compositional shifts among analyst-defined reference states. Material properties, however, cannot always be reduced to explicit compositional fractions. This is particularly true for composite materials, where performance arises from coupled factors such as filler–polymer interactions, filler dispersion and the state of the matrix polymer.[3] In such cases, Py-MS can be used directly as a high-dimensional material descriptor.[22]

A representative example is an alumina/poly(dimethylsiloxane) (PDMS) heat dissipator (Fig. 13A). Although the tested commercial materials had nearly identical nominal compositions, consisting of approximately 90 wt% alumina filler and 10 wt% PDMS-based siloxane phase, their thermal conductivity and adhesion varied substantially. This indicates that the relevant information is not simply the bulk composition, but the state of the minor polymer phase and its interaction with the filler. Py-MS directly observes only this minor polymer phase; the inorganic filler itself is not detected. Nevertheless, the pyrolysis behaviour of PDMS chains is affected by crosslinking, interfacial immobilisation and the local thermal environment created by filler dispersion. The PDMS-derived fragment pattern can therefore serve as an indirect probe of the microenvironment of the whole composite.

The first-stage NMF used in RQMS was repurposed to extract peak groups that evolve together during heating, and their temperature-resolved abundances were used as material descriptors (Fig. 13B). These bases do not simply reflect end-group differences, whose contribution should be small for high-molecular-weight PDMS. Model hydrosilylation experiments instead showed that base 0 increased and base 1 decreased with crosslinking, allowing base 0 to be assigned to crosslinked PDMS domains and base 1 to uncrosslinked or weakly crosslinked siloxane domains (Fig. 13D). Base 2 showed little dependence on reaction time, suggesting an interface-related or immobilised PDMS domain.

The temperature-resolved base abundances were then correlated with thermal conductivity and adhesion using Fisher discriminant analysis (FDA) (Fig. 14).[68] FDA projects the descriptor space onto axes that best separate predefined property classes, such as high, medium and low performance. The projection coefficients indicate which NMF base and temperature region contribute to each property classification.

For thermal conductivity, high-performance samples were characterised by a high contribution of base 0 and a low contribution of base 1 over a broad temperature range. This

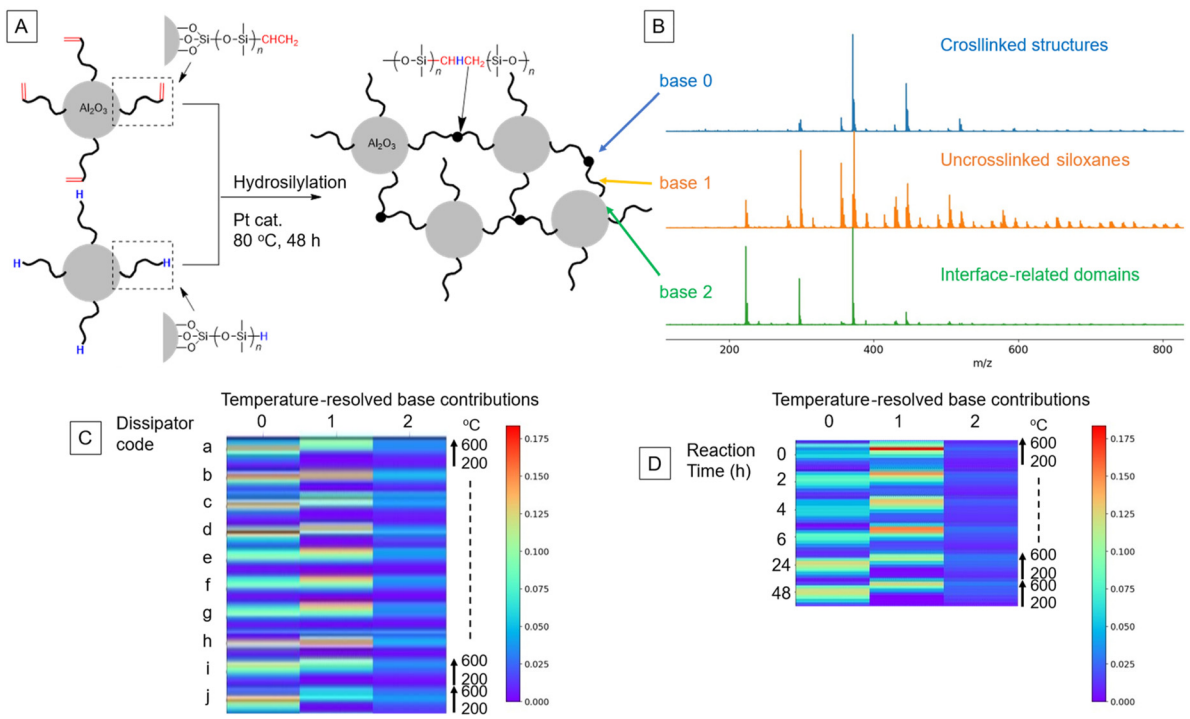


Fig. 13 Py-MS analysis of alumina-siloxane heat dissipators. (A) Hydrosilylation-based network formation. (B) NMF-extracted fragment spectra. (C) Temperature distribution of fragment abundances in commercial samples. (D) Evolution of the temperature-resolved fragment-abundance profiles during crosslinking at 30 °C for 0–48 h. Reproduced from ref. 22 under the terms of the CC BY 4.0 license.

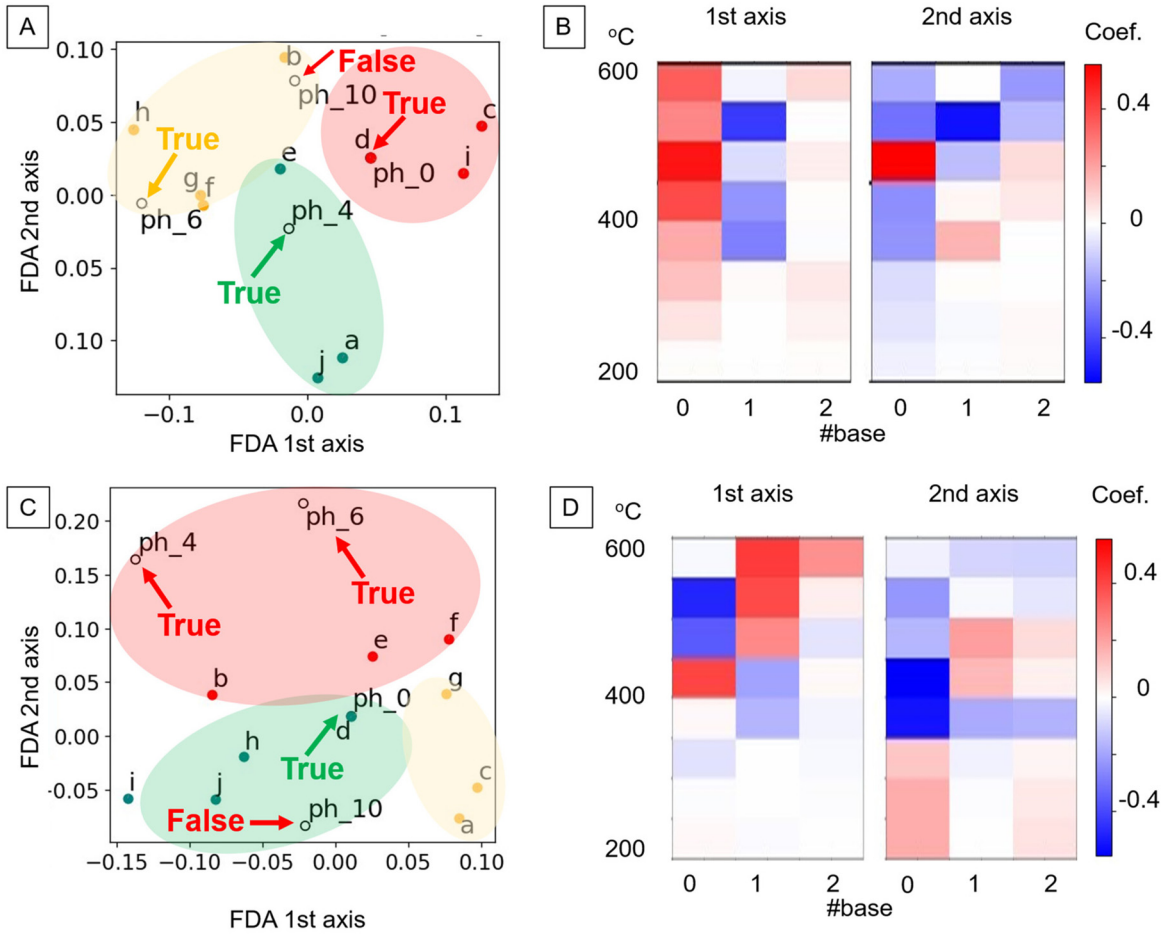


Fig. 14 Prediction of heat dissipator performance based on Py-MS descriptors. (A and C) Low-dimensional FDA projections constructed from temperature-resolved NMF-base abundances for thermal conductivity and adhesion, respectively. Background regions indicate the property classes learned from commercial heat dissipator samples. Labels "True" and "False" indicate whether each phenyl-substituted test sample was assigned to the experimentally corresponding class. (B and D) Weight maps of the FDA projection axes, showing the temperature ranges and NMF bases that contributed to each classification. Adapted from ref. 22 under the terms of the CC BY 4.0 license.

suggests that a more crosslinked matrix polymer favours thermal conduction. In filler-free polymers, higher crosslink density is known to improve thermal conductivity, but in highly filled composites the matrix state is difficult to evaluate by conventional viscoelastic measurements. Py-MS selectively observes the minor polymer phase and extracts the state of the matrix polymer, which can act as a bottleneck in the thermal conduction path.

For adhesion, the relationship could not be reduced to crosslink density alone. The FDA axis depended not only on base abundance, but also strongly on its temperature distribution. This indicates that adhesion is governed not merely by the amount of crosslinked PDMS, but also by local microenvironments that modulate the thermal stability of the polymer phase. These temperature-dependent Py-MS descriptors may therefore capture interfacial confinement and filler–polymer interactions that are difficult to evaluate from composition or crosslink density alone. Once calibrated against conventional property measurements, Py-MS descriptors could serve as rapid surrogate tests for material performance. When integrated with continuously monitored process variables,[25] they could provide at-line feedback to support process monitoring and industrial quality control.

8. Current limitations and perspectives

Py-MS can encode sample-specific structural distributions in the masses and generation temperatures of the numerous fragments produced by polymer pyrolysis. However, substantial information is also lost during this process, and Py-MS/RQMS should therefore be used in combination with conventional polymer-characterisation methods. The clearest example is the molecular-weight distribution. Once a polymer chain has been converted into small fragments, reconstructing its original chain-length distribution is inherently difficult. Molecular weight and molecular-weight distribution should therefore be

 This journal is © The Royal Society of Chemistry 2026

evaluated independently using techniques such as gel permeation chromatography. Likewise, crystalline and phase-separated structures may change or disappear during heating before pyrolysis. Their evaluation therefore requires complementary X-ray scattering and diffraction techniques. By contrast, structural factors that remain influential during pyrolysis, such as inorganic fillers and cross-linked networks, may still be detectable through their effects on polymer decomposition, allowing differences in the local polymer environment to be probed indirectly. Relatedly, components that remain non-volatile, for example through char formation, are not detected by Py-MS. If the fraction converted into detectable volatiles differs among samples, normalisation by the initial sample mass may not correctly compensate for sample loading and may introduce systematic bias.

Beyond these analytical boundaries, the coefficients obtained by RQMS should first be regarded as model-derived quantities. Their interpretation as actual structural fractions is an approximation whose accuracy depends on two distinct issues. First, the chosen reference states must adequately represent the structural distribution to be quantified. For example, in binary-triad sequence analysis using only five references, spectral reconstruction is dominated by the relatively intense low-mass fragment patterns, mainly those of monomers, dimers and trimers. Dataset-wide variations in the weaker higher-mass oligomer peaks are not decomposed into independent components, and the sequence information encoded in those peaks is therefore not quantified accurately. Tacticity-dependent differences in pyrolysis behaviour are likewise not explicitly resolved, so the inferred sequence distribution is averaged over the underlying tacticity states (section 4). Second, accurate reference inference depends on dataset coverage rather than simply on dataset size. Under the assumptions discussed in section 3.1, the sufficiently scattered condition guarantees recovery of the true references, but whether an experimental dataset satisfies this condition cannot be verified because the true compositions are unknown. Because this condition is sufficient rather than necessary, correct references may still be recovered from less broadly distributed datasets; however, any deviation from the true references cannot be quantified from the RQMS dataset alone. The inferred coefficients must therefore be validated against independent chemical evidence, such as NMR-derived compositions, copolymerisation kinetics, experimentally accessible references and chemically reasonable fragment patterns. If inconsistencies are found, samples should be added in the underrepresented composition region and the analysis repeated.

At present, the broader application of Py-MS/RQMS is also constrained by the relatively small datasets constructed independently for individual polymer systems. In sequence analysis, for example, a new dataset must be generated whenever the target monomer combination is changed. A long-term goal would therefore be to assemble a large-scale dataset linking Py-MS measurements with monomer structures, monomer reactivity ratios, compositions and polymerisation conditions, and to use this dataset to train a transferable sequence-analysis model. Such a model could then be adapted to a target monomer combination using only a limited number of additional samples, rather than requiring the analytical framework to be reconstructed from scratch for every new monomer combination. The same transferability may also provide access to more complex sequence motifs. To date, the most complex sequence distributions demonstrated are pentads for binary copolymers and triads for ternary copolymers. Because the number of possible sequence states increases rapidly with motif length and the number of monomer species, further extension is currently limited mainly by the amount of system-specific data required to constrain their reference spectra. Reusing fragment–sequence relationships learned across polymer systems may therefore enable pentad analysis of multicomponent copolymers and motifs longer than pentads in binary copolymers. Ion-mobility MS and tandem MS are also emerging as tools for distinguishing sequence isomers of relatively long oligomers and obtaining sequence information from diagnostic fragmentation patterns. Complementary advances across these mass-spectrometric approaches may ultimately enable the determination and quantification of sequence motifs far beyond the short-range sequence information typically accessible by conventional NMR analysis.[69,70]

The principal obstacle, however, is not simply the amount of available data, but how chemically diverse copolymers can be represented in a unified and machine-readable form. As of 29 May 2025, the large-scale polymer database PoLyInfo contained 19 227 homopolymers but only 8321 copolymers.[71] Copolymers are widely used in practical materials, and their chemical space is combinatorially larger than that of homopolymers because it expands not only with the combinations of constituent monomers but also with composition and sequence distribution. The relatively limited number of registered copolymers therefore likely reflects a fundamental representation problem: the chemical structures of the constituent monomers and their average composition alone cannot adequately describe the structural distributions that govern copolymer properties.

Py-MS may provide a means of overcoming this bottleneck. A major advantage is that diverse polymer materials can be directly analysed using essentially the same measurement procedure and mapped onto a common mass–temperature data space in which sample-specific structural distributions and property-related signatures are encoded. RQMS can then extract purpose-specific quantitative structural distributions from this shared representation, while Py-MS descriptors can also be related directly to material properties when an explicit compositional representation is not appropriate. In this sense, Py-MS/RQMS provides a universal analytical framework for polymer analysis and may serve as an analytical foundation for data-driven polymer science.

Conflicts of interest

The author declares no conflicts of interest.

Data availability

No original experimental data are reported in this Tutorial Review. The data discussed in this review are available in the cited original publications.

Acknowledgements

This Tutorial Review brings together and systematises a series of polymer-analysis studies based on the Py-MS/RQMS framework developed at the National Institute for Materials Science (NIMS) since 2021. I am deeply grateful to all academic and industrial collaborators involved in these studies. I sincerely thank Dr Masanobu Naito for his generous support through the JST CREST programme (JPMJCR19J3). I also thank Dr Yasuyuki Nakamura and Dr Kimiyoshi Naito for their valuable support and discussions during the development of this research. This work was also supported by JSPS KAKENHI grant number JP24K08520.

References

1 T. Jiang, A. Hall, M. Eres, Z. Hemmatian, B. Qiao, Y. Zhou, Z. Ruan, A. D. Couse, W. T. Heller, H. Huang, M. O. de la Cruz, M. Rolandi and T. Xu, *Nature*, 2020, **577**, 216–220.
2 J. Shen, X. Lin, J. Liu and X. Li, *Macromolecules*, 2019, **52**, 121–134.
3 N. Burger, A. Laachachi, M. Ferriol, M. Lutz, V. Toniazzo and D. Ruch, *Prog. Polym. Sci.*, 2016, **61**, 1–28.
4 R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi and C. Kim, *npj Comput. Mater.*, 2017, **3**, 54.
5 C. Kim, A. Chandrasekaran, T. D. Huan, D. Das and R. Ramprasad, *J. Phys. Chem. C*, 2018, **122**, 17575–17585.
6 P. A. Beaucage, D. R. Sutherland and T. B. Martin, *Macromolecules*, 2024, **57**, 8661–8670.
7 T. B. Martin and D. J. Audus, *ACS Polym. Au*, 2023, **3**, 239–258.
8 S. T. Knox, K. E. Wu, N. Islam, R. O'Connell, P. M. Pittaway, K. E. Chingono, J. Oyekan, G. Panoutsos, T. W. Chamberlain, R. A. Bourne and N. J. Warren, *Polym. Chem.*, 2025, **16**, 1355–1364.
9 K. Qian, W. E. Killinger, M. Casey and G. R. Nicol, *Anal. Chem.*, 1996, **68**, 1019–1027.
10 R. P. Lattimer, *J. Anal. Appl. Pyrolysis*, 2003, **68–69**, 3–14.
11 R. B. Cody, T. N. J. Fouquet and C. Takei, *Rapid Commun. Mass Spectrom.*, 2020, **34**, e8687.
12 H. Sato, S. Nakamura, K. Teramoto and T. Sato, *J. Am. Soc. Mass Spectrom.*, 2014, **25**, 1346–1355.
13 R. Watanabe, T. Ozawa, S. Nakamura and H. Watanabe, *Anal. Chem.*, 2026, **98**, 2815–2824.
14 T. N. J. Fouquet, *J. Mass Spectrom.*, 2019, **54**, 933–947.
15 W. Windig, E. Jakab, J. M. Richards and H. L. C. Meuzelaar, *Anal. Chem.*, 1987, **59**, 317–323.
16 W. C. Knol, S. Vos, T. Gruendling, B. W. J. Pirok and R. A. H. Peters, *Anal. Chim. Acta*, 2023, **1238**, 340635.
17 S. R. A. Molenaar, W. C. Knol, B. W. J. Pirok and R. A. H. Peters, *J. Anal. Appl. Pyrolysis*, 2023, **171**, 105978.
18 Y. Hibi, S. Uesaka and M. Naito, *Chem. Sci.*, 2023, **14**, 5619–5626.
19 Y. Hibi, T. Matsumoto, M. Midorikawa, S. Uesaka, M. Mizoshiri, T. Tanimura, M. Okada, K. Naito and M. Naito, *Polym. Degrad. Stab.*, 2025, **232**, 111128.
20 W. Zhu, J. W. Smith and C. M. Huang, *J. Biomed. Biotechnol.*, 2010, **2010**, 840518.
21 D. C. Heinz and C.-I. Chang, *IEEE Trans. Geosci. Remote Sens.*, 2001, **39**, 529–545.
22 Y. Hibi, Y. Tsuyuki, S. Ishii, E. Ide and M. Naito, *Sci. Technol. Adv. Mater.*, 2024, **25**, 2362125.
23 T. B. Martin, D. R. Sutherland, A. McDannald, A. G. Kusne and P. A. Beaucage, *Chem. Mater.*, 2025, **37**, 4272–4281.
24 D. E. Huang, A. P. Kotula, C. R. Snyder and K. B. Migler, *Macromolecules*, 2022, **55**, 10921–10932.
25 Y. Hibi, S. Uesaka, K. Hitomi, K. Niihara, A. Imai, S. Samitsu and M. Naito, *ACS Sustainable Resour. Manage.*, 2025, **2**, 673–680.
26 H. Sato, S. Nakamura and H. Shinzawa, *J. Anal. Appl. Pyrolysis*, 2025, **189**, 107056.
27 E. P. Erisman, Y. Simón-Manso and W. E. Wallace, *Environ. Sci. Technol.*, 2026, **60**, 19336–19346.
28 J. H. Gross, *Anal. Bioanal. Chem.*, 2014, **406**, 63–80.
29 D. T. Usmanov, K. Hiraoka, H. Wada, M. Matsumura, S. Sanada-Morimura, H. Nonami and S. Yamabe, *Anal. Chim. Acta*, 2017, **973**, 59–67.
30 F. Zitomer, *Anal. Chem.*, 1968, **40**, 1091–1095.
31 Y. Hibi, S. Uesaka and M. Naito, *Analyst*, 2024, **149**, 4388–4394.
32 T. Ozawa, S. Nakamura, H. Sato, H. Shinzawa, H. Hagihara and R. Watanabe, *Anal. Chem.*, 2025, **97**, 1665–1672.
33 J. H. Gross, *Eur. J. Mass Spectrom.*, 2020, **26**, 241–273.
34 T. Tsugoshi, T. Nagaoka, M. Nakamura, Y. Shiokawa and K. Watari, *Anal. Chem.*, 2006, **78**, 2366–2369.
35 Y. Kitahara, S. Takahashi, N. Kuramoto, M. Šala, T. Tsugoshi, M. Sablier and T. Fujii, *Anal. Chem.*, 2009, **81**, 3155–3158.
36 M. D. Craig, *IEEE Trans. Geosci. Remote Sens.*, 1994, **32**, 542–552.
37 X. Fu, K. Huang, B. Yang, W.-K. Ma and N. D. Sidiropoulos, *IEEE Trans. Signal Process.*, 2016, **64**, 6254–6268.
38 D. D. Lee and H. S. Seung, *Nature*, 1999, **401**, 788–791.
39 H. Laurberg, M. G. Christensen, M. D. Plumbley, L. K. Hansen and S. H. Jensen, *Comput. Intell. Neurosci.*, 2008, **2008**, 764206.
40 L. Miao and H. Qi, *IEEE Trans. Geosci. Remote Sens.*, 2007, **45**, 765–777.
41 S. Yamane, S. Nakamura, R. Inoue, T. N. J. Fouquet, T. Satoh, K. Kinoshita and H. Sato, *Macromolecules*, 2021, **54**, 10388–10394.
42 I. C. McNeill, *Angew. Makromol. Chem.*, 1997, **247**, 179–195.

 This journal is © The Royal Society of Chemistry 2026

43 M. Shiga, K. Tatsumi, S. Muto, K. Tsuda, Y. Yamamoto, T. Mori and T. Tanji, *Ultramicroscopy*, 2016, **170**, 43–59.
44 R. K. Roy, A. Meszynska, C. Laure, L. Charles, C. Verchin and J.-F. Lutz, *Nat. Commun.*, 2015, **6**, 7237.
45 N. Zydziak, W. Konrad, F. Feist, S. Afonin, S. Weidner and C. Barner-Kowollik, *Nat. Commun.*, 2016, **7**, 13672.
46 A. Al Ouahabi, J. A. Amalian, L. Charles and J.-F. Lutz, *Nat. Commun.*, 2017, **8**, 967.
47 J. M. Lee, M. B. Koo, S. W. Lee, H. Lee, J. Kwon, Y. H. Shim, S. Y. Kim and K. T. Kim, *Nat. Commun.*, 2020, **11**, 56.
48 J. M. Lee, J. Kwon, S. J. Lee, H. Jang, D. G. Kim, J. Song and K. T. Kim, *Sci. Adv.*, 2022, **8**, eabl8614.
49 H. Jang, H. Chu, H. Noh and K. T. Kim, *Angew. Chem., Int. Ed.*, 2025, **64**, e202415124.
50 K. Satoh, S. Ozawa, M. Mizutani, K. Nagai and M. Kamigaito, *Nat. Commun.*, 2010, **1**, 6.
51 X. Tong, B.-H. Guo and Y. Huang, *Chem. Commun.*, 2011, **47**, 1455–1457.
52 Y. Hibi, M. Ouchi and M. Sawamoto, *Nat. Commun.*, 2016, **7**, 11064.
53 C. Boyer, M. Kamigaito, K. Satoh and G. Moad, *Prog. Polym. Sci.*, 2023, **138**, 101648.
54 M. W. Urban, D. Davydovich, Y. Yang, T. Demir, Y. Zhang and L. Casabianca, *Science*, 2018, **362**, 220–225.
55 H. Liao, S. Hu, H. Yang, L. Wang, S. Tanaka, I. Takigawa, W. Li, H. Fan and J. P. Gong, *Nature*, 2025, **644**, 89–95.
56 G. Moad, E. Rizzardo and S. H. Thang, *Aust. J. Chem.*, 2012, **65**, 985–1076.
57 F. R. Mayo and F. M. Lewis, *J. Am. Chem. Soc.*, 1944, **66**, 1594–1601.
58 Y. Hibi, Y. Nakamura, S. Uesaka and M. Naito, *Macromolecules*, 2026, **59**, 1640–1649.
59 Y. Hibi, S. Tokuoka, T. Terashima, M. Ouchi and M. Sawamoto, *Polym. Chem.*, 2011, **2**, 341–347.
60 M. Rubens, J. H. Vrijsen, J. Laun and T. Junkers, *Angew. Chem., Int. Ed.*, 2019, **58**, 3183–3187.
61 S. T. Knox, S. J. Parkinson, C. Y. P. Wilding, R. A. Bourne and N. J. Warren, *Polym. Chem.*, 2022, **13**, 1576–1585.
62 C. Barner-Kowollik and B. T. Tuten, *Chem. Sci.*, 2023, **14**, 5825–5826.
63 Y. Hibi, *Analyst*, 2024, **149**, 5320–5328.
64 N. Dobigeon, J.-Y. Tourneret, C. Richard, J. C. M. Bermudez, S. McLaughlin and A. O. Hero, *IEEE Signal Process. Mag.*, 2014, **31**, 82–94.
65 S. S. Ali, T. Elsamahy, E. Koutra, M. Kornaros, M. El-Sheekh, E. A. Abdelkarim, D. Zhu and J. Sun, *Sci. Total Environ.*, 2021, **771**, 144719.
66 A. Chamas, H. Moon, J. Zheng, Y. Qiu, T. Tabassum, J. H. Jang, M. Abu-Omar, S. L. Scott and S. Suh, *ACS Sustainable Chem. Eng.*, 2020, **8**, 3494–3511.
67 E. Syranidou, K. Karkanorachaki, D. Barouta, E. Papadaki, D. Moschovas, A. Avgeropoulos and N. Kalogerakis, *Environ. Sci. Technol.*, 2023, **57**, 8130–8138.
68 W. Deng, J. Hu, J. Guo and H. Zhang, in *Fourth International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 2007)*, IEEE, 2007, vol. 3, pp. 248–252.
69 X. Liu, L. R. Cool, K. Lin, A. M. Kasko and C. Wesdemiotis, *Analyst*, 2015, **140**, 1182–1191.
70 S. Naskar, A. Izuagbe, V. Lemaur, Q. Duez, A. Minoia, J. De Winter, S. J. Blanksby, J. Cornil, C. Barner-Kowollik and P. Gerbaux, *Polym. Chem.*, 2025, **16**, 2143–2153.
71 PoLyInfo, NIMS, https://polymer.nims.go.jp, accessed 5 August 2026.