Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Digital Discovery*, 2025, **4**, 3292

Unveiling the knowledge of a RAFT polymerization database obtained from an automated parallel synthesizer

Michael Ringleb, †[ab] Yannik Köster, †[ab] Stefan Zechel[ab] and Ulrich S. Schubert *[abcd]

An automated parallel synthesizer has been utilized to perform 539 polymerization kinetic experiments in a semi-automated high-throughput approach utilizing size exclusion chromatography and NMR spectroscopy for characterization. The polymerization parameters were systematically varied between 16 different monomers, seven RAFT agents and three solvents. More than 7200 data points were obtained and curated into 234 accurately described and successful reaction kinetics. The kinetic curves were fitted with appropriate negative growth functions, thereby enabling interpolation between the sampling times. The resulting knowledge is made available *via* a web interface to allow a quick search for the optimal reaction conditions for synthesizing a polymer with desired properties such as a defined molar mass. This work can be regarded both, as a demonstration of using a high-throughput approach to generate a reliable database, which is also in accordance with the FAIR principles, and further, as an explanatory data augmentation of such datasets, visualizing their inherent limitations and potentials in utilization.

Received 30th April 2025
Accepted 1st October 2025

DOI: 10.1039/d5dd00177c

rsc.li/digitaldiscovery

Introduction

Since its advent in the early 1970s, high-throughput experimentation (HTE) has become a valuable asset in pharmaceutical[1–4] and materials science.[5–10] In the field of polymer research, the concept was utilized first in the 2000s with, *e.g.*, works focusing on automated parallel synthesizers to carry out multiple sets of polymerizations.[11–14] Since then, a variety of polymerization techniques have been performed in an automated manner and carried out under high-throughput conditions, ranging from ionic to atom-transfer and nitroxide-mediated polymerizations.[9,15–21]

The latter two polymerization types belong to the group of reversible deactivation radical polymerizations (RDRP), which are particularly noteworthy in combination with automated syntheses. These polymerization types yield polymers with a low dispersity, which is an optimal basis for the production of materials with tailored properties.[22] Another interesting RDRP technique is the reversible-addition–fragmentation chain-transfer (RAFT) polymerization procedure.[23–25] It was invented by the Commonwealth Science and Industrial Research Organization of Australia and simultaneously by Rhodia as the "macromolecular design *via* the interchange of xanthates" (MADIX).[24,26,27] Since the technique was developed, numerous different RAFT polymerizations have been conducted over the past approximately 30 years, both manually and in an automated fashion.[24,28–44]

The RAFT process poses the opportunity to polymerize a wide range of monomers, from styrenes[45,46] to vinyl ethers[47] to (meth-)acrylates/(meth-)acrylamides.[24,25]

For the regulation of this polymerization, the selection of the initiator as well as the chain transfer agent (CTA), also termed RAFT agent, and solvent is of significant importance.[23,48–50]

To date, studies regarding the required CTA's substituents for specific groups of monomers have been conducted.[51–53]

Nevertheless, a systematic and conclusive review of the interplay between solvent and CTA for a broad range of monomers is, to the best of our knowledge, currently not available in literature. In this context, a comprehensive database for RAFT polymerizations is still missing. Addressing this requirement would facilitate the evaluation of the interactions between, *e.g.* monomer and CTA, by utilizing novel data science solutions.

Hence, this study employs an automated high-throughput approach for the execution and kinetic sampling of RAFT polymerizations to start addressing this gap in knowledge.

[a]*Laboratory of Organic and Macromolecular Chemistry (IOMC), Friedrich Schiller University Jena, 07743 Jena, Germany. E-mail: ulrich.schubert@uni-jena.de*

[b]*Jena Center for Soft Matter (JCSM), Friedrich Schiller University Jena, 07743 Jena, Germany*

[c]*Helmholtz Institute for Polymers in Energy Applications Jena (HIPOLE Jena), Lessingstraße 12-14, 07743 Jena, Germany*

[d]*Helmholtz-Zentrum für Materialien und Energie (HZB), Hahn-Meitner-Platz 1, 14109 Berlin, Germany*

† These authors have contributed equally.

Once such a database is established, one can follow a guide like the FAIR (findable, accessible, interoperable, and reproducible) principles.[54] These principles offer reasonable orientation for the publishing researcher to present their data in a scientific way. Nevertheless, their implementation proves to be challenging as often powerful search functionalities, *e.g.*, sketch search or similarity search for molecules are expected.[55,56] In particular, with the trend to apply machine-learning techniques onto such datasets and the general advice from many research groups to keep humans in the loop, it is important to convey expertise and research intent of the dataset.[57,58]

Hence, in addition to ensuring the FAIR principles, regarding the access and interoperation of the information in the resulting database, data science methods were utilized for the purposes of interpolation of the kinetics to augment and render the knowledge to an interactive dedicated online platform.

Because contemporary research often highlights machine learning approaches, it is important to note that the reaction curves generated in this work are extended by appropriately lightweight automatic adaptation of suitable graphs to them. Hence, the integration of this dataset by other groups as a part of a larger data set and consecutive machine learning is considered fruitful and suggested.

Results and discussion

Concept of the study

539 RAFT polymerizations utilizing 16 different monomers, seven different CTAs and three different solvents (see permutation Tables S4 to S6 in the SI and Fig. 1) were performed.[59]

After the conduction of the experiments, the prepared samples are investigated by means of nuclear magnetic resonance spectroscopy (NMR) and size exclusion chromatography (SEC). Information about the molar mass of the polymer and conversion at the time of sampling are gathered. The data are then curated and fed into a database to make it publicly available. Exemplary spectra and SEC curves for each polymer class are represented in the SI (Fig. S1 to S6). The constant parameters for the reactions were chosen based on previous work within our group.[60] For example, the monomer concentration was selected to avoid vitrification of the reaction solution and enable automated sampling using the setup presented in this manuscript. However, to test our setup and data curation process we also employed conditions which were likely to be unsuccessful (*e.g.*, polymerization of nonpolar monomers in polar solvents). In general, we do not aim to provide an optimized set of reaction conditions for the investigated monomers but rather a comparison of the polymerization between combinations of different educts (monomers, CTA, and solvent) under comparable reaction conditions. This enables the identification of influences of the reactants on the polymerization kinetics.

Automation of polymerization and sampling workflow

The first challenge of the study was to implement a robust protocol for the execution of the polymerizations and the sampling process. For this purpose, a previously reported protocol was adapted to the needs of conversion measurements *via* NMR spectroscopy instead of gas chromatography.[11]

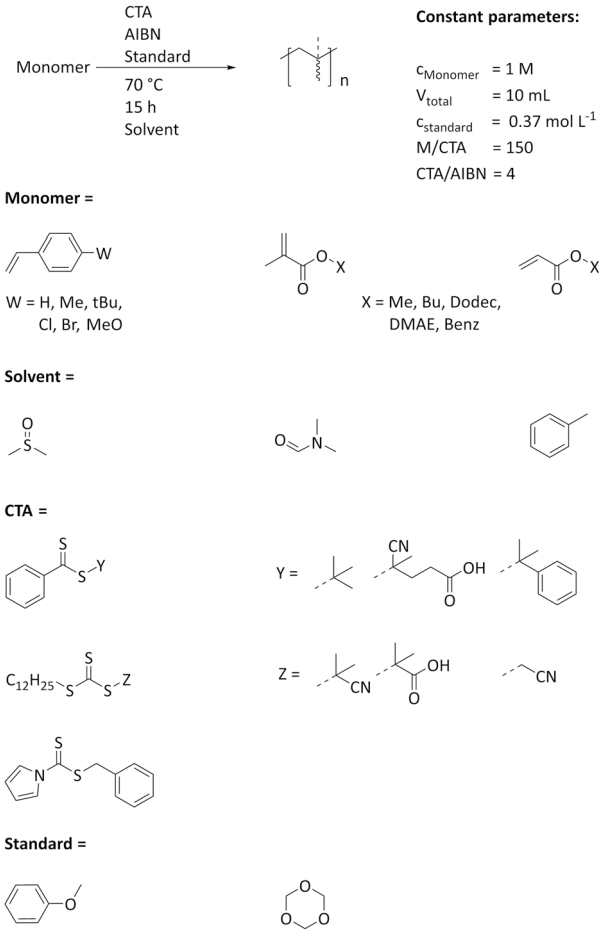

Fig. 1 Schematic representation of the performed reactions with constant parameters and variables (monomer, solvent, and CTA).

An automated parallel synthesizer was utilized to prepare the polymerization solutions by mixing solvent, monomer, standard and CTA and initiator solutions in different ratios in 15 reactors per batch. Each reactor contained a different combination of CTA, solvent and monomer (see Tables S1 to S6 in the SI for more details on the different combinations). At defined times after the start of the polymerization, samples were taken from each reactor, utilizing the 4-needle head of the synthesizer, and filled into respective vials or tubes depending on the sample's nature (SEC or NMR). The general polymerization time was 15 h.

As the preparation of the samples should be mainly automated, it was required to choose a setup, which omitted substantial human intervention (*e.g.*, in the form of closing NMR tubes with lids or transferring individual tubes to the NMR spectrometer). Hence, a high-throughput NMR autosampler was chosen (Bruker SampleJet), which offers a 96 well plate like footprint for NMR tubes. The 4″ tubes are arranged in a grid with size 8 × 12. The lids of the tubes have a small (<2 mm diameter) hole, which rendered the filling with the dispensing system of the robot nearly impossible. As presented in

a previous work, a funnel module in the same format was developed and manufactured to still enable the sampling.[10]

The second required technique was sampling for SEC samples. To reduce the amount of human intervention, a previously published sampling technique was utilized, involving vials with slitted septum caps, which ensured little to no evaporation of the sample between sampling and measurement.[59]

Evaluation procedure

After the samples had been measured by means of NMR spectroscopy and/or SEC chromatography, the data had to be evaluated to feed it into the initial database.

For the evaluation of the SEC data, the program WinGPC from PSS was utilized. The elugrams were visually inspected and the visible distributions were integrated to gain the values for the number average (M_n) and weight average (M_w) molar mass. The determination of the values was performed utilizing the internal calculation based on the previously conducted calibration. Poly(styrene) (PS) or poly(methyl methacrylate) (PMMA) calibration standards were utilized depending on the monomer – PS for styrenic polymers and PMMA for others. If the values for M_n were below 2000 g mol^{-1}, the reported value was rounded to the closest 10 g mol^{-1}. If it was between 2000 and 20 000 g mol^{-1}, the value was rounded to the closest 100 g mol^{-1} and if it was above 20 000 g mol^{-1}, the value was rounded to the nearest 1000 g mol^{-1}. The dispersity $Đ$ was then calculated by dividing the weight average molar mass by the number average molar mass (eqn (1)).

$$Đ = \frac{M_w}{M_n} \quad (1)$$

For the evaluation of the NMR data, the peaks of the standard substances (trioxane or anisole) were set to their expected proton count. For trioxane the six methylene-protons and for anisole, the aromatic protons were utilized as reference. This procedure was performed for each of the spectra in a time row of an individual polymerization. Afterwards, the integral of one of the vinyl methylene peaks of the monomer was evaluated. The value for the $t = 0$ h sample was then utilized as the marker for 0% conversion. The conversion of the other samples was calculated by subtracting the quotient of the methylene peak's integral for the specific time by the integral of the same peak at $t = 0$ h (eqn (2)).

$$\text{conversion} = 1 - \frac{I_{t=0\text{ h}}}{I_{t=x\text{ h}}} \quad (2)$$

The obtained values were inserted into an initial database sorted by the composition of the polymerization solution and the time of the individual sample.

Calculation of the specific sampling times

Inside the robot program, the individual sampling times (*e.g.*, sampling after 1 h, 2 h, *etc.*) are defined as the elapsed time since the heating was activated after the sampling for the initial set of samples ($t = 0$ h). However, only two reactors could be sampled at the same time due to practical limitations of the robot. Hence, deviations between the stated time in the initial dataset and correct sampling time for specific reactors occurred, which were up to *ca.* 20 to 30 min for the reactors sampled as last. For the generation of the published dataset and particularly the fitting, it was, however, important to correlate specific time points of sampling with the result of the measurement (*e.g.*, conversion value determined by NMR spectroscopy). A deviation in the utilized sampling time would lead to a divergence for the fitting of the polymerization kinetics. Due to this fact, the specific sampling times for each reactor had to be determined in more detail in comparison to the programming code of the machine, rendering the fitting more successful. The exact sampling times for each reactor are presented in the respective tab of the experimenter sheet xlsx file on the website.

Investigation of the reproducibility

The dataset was also checked for the reproducibility of the polymerizations under given reaction conditions. Hence, duplicates of specific reactions were performed to investigate whether the results are reproducible when carried out several times. From these experiments, it became obvious that syntheses with an overall conversion above *ca.* 50% after 7.5 h were easily comparable and reproducible. The same applies to the reactions with lower conversions (around or below 20% after 15 h) within the accuracy of the NMR evaluation. However, as the relative deviation is higher (because of the lower final conversion, see Fig. 2) this could lead to the misinterpretation that these reactions are more challenging to replicate. Nevertheless, this is not the case and it is visible that the automated process generally ensures reproducibility of the experimental results. For a comparison of the divergence of the conversion of all replicates of successful polymerizations see SI Fig. S9.

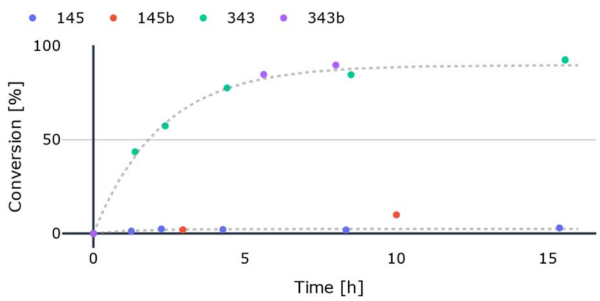

Fig. 2 Representative reproducing attempts ("b"/red and violet) of single kinetic time points at low conversion (experiment number 145/blue) and high conversion (experiment number 343/light green) and the respective fitting/prediction curves of the original kinetics (grey, dashed). Experiment 145 consisted of the polymerization of 4-methylstyrene with CTA 4-cyano-4-(phenylcarbonothioylthio)pentanoic acid. In experiment 343 the monomer benzyl acrylate was polymerized with the RAFT-agent cyanomethyl dodecyl trithiocarbonate. Both polymerizations were performed in dimethylformamide.

Fitting the experimental data with the theoretical course of the kinetics

In order to inter- and extrapolate beyond the sampling times of the kinetic studies performed, the conversion-over-time graphs had to be fitted with suitable growth functions. Theoretically, a linear function should be appropriate for describing the performed RDRPs (Fig. 3, red).[61] However, in practice, the ideal first-order kinetic plot may be subject to alterations. A first alteration can be attributed to a slow initiation of the polymerization (induction period). This is typically connected to the stability of the radicals generated by the RAFT-agent's leaving group, and the monomer reactivity, as both influence the rate of polymer chain propagation. In cases where these are not tuned to each other the discrepancy from the ideal can present itself as a low, then increasing conversion rate, the first part of a sigmoidal function. One such example within our dataset is the addition of tertiary cyanoalkyl radicals (derived from the thermal activation of AIBN) to acrylates. The rate of this reaction is substantially lower than the rate of polymerization for those monomers, leading to this phenomenon.[62] Furthermore, the decay rate of the utilized initiator plays an important role in this step.[28] However, as the initiator and the temperature for the polymerization were kept constant for all processes within our study, the influence of this parameter can be neglected.

Second and in a similar manner, depending on the combination of CTA, solvent and monomer, side-reactions can occur, leading to a decrease of the propagation rate. Example side reactions are, among others, the decomposition of the RAFT agent or disproportionation reactions. These can potentially also stop the reaction, presenting themselves as a decrease in the slope of the conversion-over-time plot, corresponding to the asymptotic part of a sigmoidal curve (Fig. 3, green).[63–66]

In consideration of the experimental setup, a third correction for the fitting of kinetic curves must be considered. The first sampling at $t = 1$ h could potentially occur (for combinations of parameters leading to high propagation rates) after the initialization step has been completed, skipping the increase in conversion rate. Taking these aspects into consideration, a negative growth function (violet) proved to be the most appropriate description in such cases. Consequently, the prediction of a conversion over time cannot be achieved in the initial stage of the reaction, which probably (depending on the propagation rates and reactivities of the CTAs and the monomers) only lasts for a few minutes. However, literature reports indicate that this phase can extend significantly longer compared to the overall polymerization time.[67] An in-depth analysis of the different fitting types is represented in the SI (Fig. S9 to S12).

Investigation of the fitting extrapolation

Some of the polymerizations (in particular for styrenic monomers) only reached low conversions (around 20%) over the time of the kinetic investigation study (last sampling time = 15 h). An explanation for this is that the reaction conditions are not optimized for the RAFT of styrenes, since in literature, often higher temperatures are applied for such polymerizations. As previously described, an ideal RAFT would exhibit a linear kinetic curve for their conversion, allowing for extrapolation beyond the sampled values. Hence, we tested the validity of the extrapolation of these kinetics. For this, a polymerization of styrene with 2-(dodecylthiocarbono-thioylthio)-2-methylpropionic acid in DMF was performed for an elongated period of time. As presented in Fig. 4 the assumption of linear growth over time could not be confirmed. Practical limitations of the setup such as oxygen creeping into the reaction, quenching it partially and the rate of decomposition of the initiator at the reaction temperature need to be considered. As presented in the example, RAFT reactions, in particular those with low conversion after 15 h (here 15% and 25%), were not necessarily following the course required for extrapolation. Hence, an estimation of the end of the reaction could only be performed when the asymptotic slowdown in the conversion rate occurred within the sampling time window. In conclusion, while interpolation *via* the fit is robust and

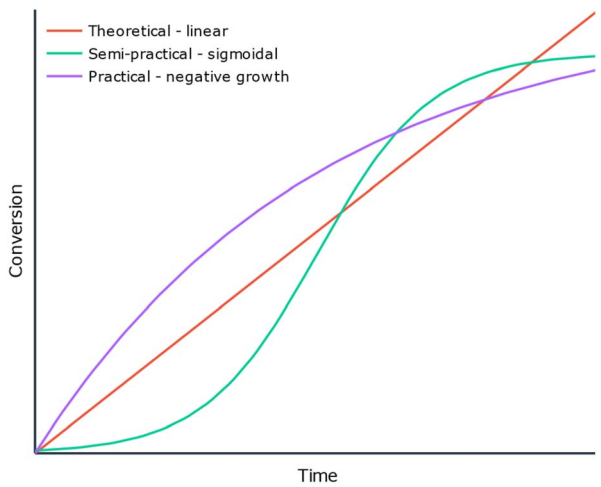

Fig. 3 Comparing three courses of fitting curves for RDRPs: theoretical, semi-practical and practical. The curves do not follow the same magnitude.

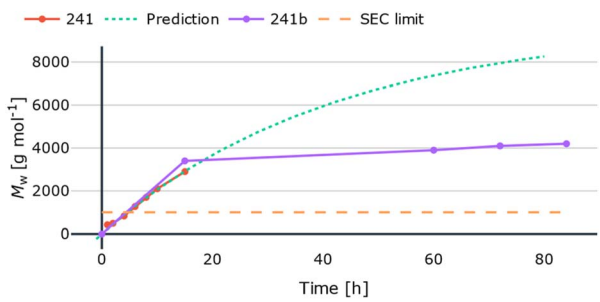

Fig. 4 Long-term estimation of a kinetic (red, green-dotted) and its actual course over 84 hours (violet). The orange-dashed line represents the lower SEC resolution limit at *ca.* 1000 g mol^{-1}. Experiment 241 represents the polymerization of styrene with CTA 2-(dodecylthiocarbonothioylthio)-2-methylpropionic acid in dimethylformamide. After the last sampling, the conversion was 15% and 25% for experiments 241 and 241b (repetition with longer run time), respectively.

straightforward, extrapolation, *e.g.*, to find out about the end time of the polymerization, was not possible.

Curation criteria of the experimental data

After the reactions for the initial dataset were performed, it became obvious that not all of the polymerizations were successful. Hence, a curation of the initial dataset was required to discard a fraction of the data before the polymerizations were fitted. The curation of the initial database was performed utilizing an automated rule-based approach. The rules for discarding individual datapoints or whole polymerization kinetics are as follows:

Individual datapoints were discarded:

(1) If the value of the weight average molar mass (M_w) was above 100 000 g mol^{-1} as these values were out of the calibration range of the size exclusion chromatography setup.

(2) If the calculated conversion for the datapoint was lower than −5% (conversions of less than 0% can be attributed to the precision of the NMR measurement and evaluation).

Whole kinetic sets of a polymerization were discarded:

(1) If experimental problems occurred during sampling, *e.g.*, dispensing failures of the robot.

(2) If the fluid level of the reactor had dropped by more than 20% after 15 h.

(3) If considerable gelation or precipitation, and hence, phase separation occurred inside the reactor (*e.g.*, for monomers with bulky alkyl chains in polar solvents).

(4) If, after discarding data points for a polymerization kinetic, less than four datapoints from the SEC samples (M_n, M_w, $Đ$) or the NMR samples (conversion) remained.

(5) If the conversion average over all data points of one kinetic was below 1%.

(6) If the value for M_n, M_w or the conversion was decreasing by more than 10% (corresponding to the method accuracy) over the course of the experiment.

From the aforementioned criteria, the low conversion (while a homogeneous solution remained) can be attributed to non-polymerization due to the combination of educts. Furthermore, decreasing molar mass values with increasing conversions could be attributed to a reaction course which is closer to a free radical polymerization rather than a RDRP. Hence, these reactions were attributed as failed. All other criteria were summarized under the category discarded. A closer discussion on the criteria and the differentiation process is presented in the SI.

Final dataset description

In total, 7221 datapoints were recorded when 539 kinetic experiments were conducted. All runs were sorted by the aforementioned criteria into utilizable (for evaluation) and discarded (potential experimentation errors). In this process, 257 kinetics (including 89 repetitions) were marked as discarded to enforce that the setup would not influence our subsequent decision if combinations of educts were delivering successful or failed RAFT polymerizations under the given conditions. This subset also includes combinations of educts which lead to polymerization, however, the polymerization itself cannot be evaluated utilizing our setup. For example, the polymerizations of lauryl methacrylate in DMSO fell into this category. These polymerizations lead to the formation of a polymer gel/solid inside the reactor which hinders the sampling and, hence, renders the evaluation of the conversion and molar mass difficult to impossible.

The utilizable experiments were then divided into successful and failed experiments according to the previously established criteria. In terms of publishing our experimental data in a FAIR way, all three sets are downloadable with their respective metadata (see Data availability).

In more detail, 234 experiments can be described as successful polymerizations, based on the criteria defined above. Fig. 5, a sunburst chart, sets the numbers of successful, discarded and failed polymerizations in context to each other. It also presents the number of replicated polymerizations. The number of replicates for the successful polymerizations is 48, meaning that *ca.* 21% of those reactions were reperformed to confirm their outcome. 305 experiments can be attributed as unsuccessful (257 discarded and 48 failed). Of these, 110 were replicated which corresponds to *ca.* 35%, nearly 1.7 times the ratio of replicated reactions of the successful polymerizations. This shows our commitment to verify the results of the unsuccessful polymerizations in an attempt to uncover the reasons leading to the missing success and to publish useable negative results on RAFT polymerizations.

Although over half of the unique experiments were successful and may represent a promising choice of reaction

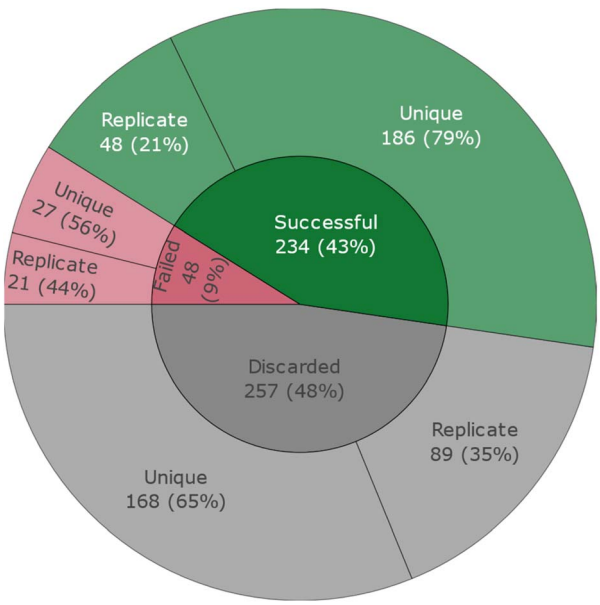

Fig. 5 Categorization of all experiments in the original dataset. A polymerization was marked as successful when all curation criteria mentioned above were met. While 35% of the failed/discarded experiments were retried (110 in total), 21% of the successful reactions were repeated to ensure reproducibility (48 in total).

conditions, the remaining failed experiments are also of value. Beyond the failed cases, the discarded polymerizations could as well add to the generation of knowledge if they are properly categorized. For example, based on our dataset it is clearly visible that the polymerization of lauryl (meth)acrylate in DMSO is occurrent, but not feasible in the presented setup due to gelation. In most cases, however, the results of these tested but inefficient reactions are discarded and not published, which in turn leads to unnecessary re-creation of the same outcome by other researchers. In essence, as Strieth-Kalthoff *et al.* have posited, the lack of failed experiments leads to creeping bias towards the positive experiments in following research and machine learning.[57,68] To counteract this bias, we decided to publish the unsuccessful reaction outcomes together with the successful ones.

Comparison to other dataset efforts

To the best of our knowledge, no other RAFT dataset exists which matches the complexity of this work – providing kinetic data dependent on CTA and solvents for a wide range of monomers, also including "negative" results on polymerizations. Nevertheless, our dataset can be contextualized alongside other recent automated parallel synthesizer generated efforts such as the RAFT emulsion copolymer series of Clothier *et al.*, who generated RAFT multiblock copolymers utilizing monomers from four different families.[69] Like many studies in the HTE domain, the authors focused on a curated, small dataset to highlight the performance and capabilities of their experimental setup. Conversely, we deliberately explored a broader range of parameters, including suboptimal ones. It should be noted that the underlying goal was not to optimize the polymerization of each monomer but to illuminate the complex interplay between the reactants.

In contrast to other major data collection efforts in polymer science, such as the Polymer Handbook or PoLyInfo, our database does not focus on reporting the physical properties or solely the properties of the final polymers.[70–74] Instead, we place emphasis on documenting the kinetics and temporal evolution of the RAFT polymerizations for a diverse set of parameters to facilitate the estimation of the right set of parameters (solvent, CTA, and reaction time) for other researchers.

Relevance of the dataset to machine learning

In times of great experimental data scarcity for machine learning algorithms, we wanted to set an example for how "negative" results can also be made public to other researchers. Machine learning models, especially when tasked with generating innovative polymers, synthesis routes, *etc.*, are vulnerable to hallucinate results far off reality, when exclusively trained on optimized positive data. The lack of failure data impedes the model's ability to learn the boundaries of chemical feasibility in the specific field.[68,75] For this reason, this publication includes well-described experimental limitations and does not hide negative outcomes.

Even though the training on this modest dataset exceeds the scope of this article by far we laid the foundation for subsequent machine learning. Molecular machine-readable representations for each monomer and CTA are presented in the "Legend for Abbreviations" tab of the experimenter sheet xlsx file (accessible *via* GitHub or directly on the website). These include SMILES of the monomers as educts. Additionally, for the repeating units (monomers in polymer), PSMILES from the Ramprasad research group[72] and plain explicit SMILES[76] with radicals where repetition units connect/repeat are displayed there, too.

Buildup of the web application

The initial concept behind this dataset was to investigate the potential of RAFT polymerizations. To make this knowledge accessible, it was required to make the entire database information accessible. During the time of data culmination an Excel spreadsheet with each line representing an experiment and separate columns for the conversions and molar masses at the specific timepoints was the interface to this data. As this spreadsheet was only meant to be easily filled by the experimenter and inadequate for visualization purposes it was required to create a more recapitulating interface, which is focused on conveying an overview instead. Hence, a search logic was incorporated into a website allowing easy access for anyone with the bare minimum of having a common internet-browser on their technological device of choice and a connection to the world wide web.

Intuition score

As Boeckhout *et al.* state, the FAIR principles do not provide a complete set of guiding principles for improving data driven science.[77] We believe that it is of utmost importance to not just follow the FAIR principles, when publishing a database but also to transport the intention to create the database to the reader, who does not directly know what parameters to look out for. This is why we introduced an intuition score to our interactive website. The score includes estimations of what the authors consider to be a good kinetic/reaction, the user might want to see first when choosing a specific set of choices of monomers and/or RAFT-agents and/or solvents (Fig. 6).

The intuition score consists of three sub-scores that have been normalized and weighted by their importance (eqn (3)). The importance given to each aspect is chosen carefully from

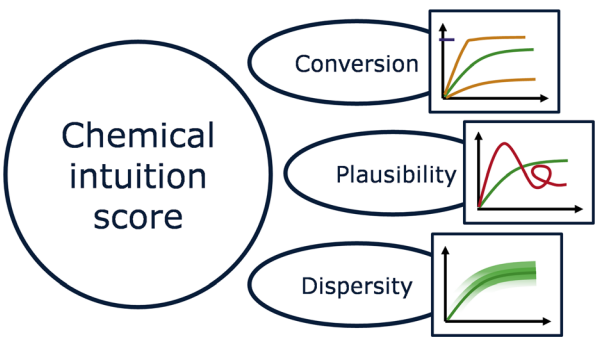

Fig. 6 Contextualization of the kinetic experiments regarding feasibility.

our experience to support practical applications, *e.g.*, industrial ones where a predictable and fast rise in conversion is given more importance than a narrow polymer dispersity.

$$\text{Score} = S_{\text{conversion}} + S_{\text{error,fit}} + S_{D,\text{mean}} \quad (3)$$

Firstly, the calculation of the intuition score includes the maximum conversion achieved by the kinetics. The contribution to a high and, therefore, good score is bigger the closer the value of the conversion at the end of the reaction period (in this study typically 15 h) is to 80% (eqn (4)). An even higher conversion usually indicates more unwanted side reactions, increasing $Đ$ due to an increasing amount of dead polymer chains.[78] On the other hand, a lower one implies a waste of resources as a lot of the monomer remains unpolymerized. Furthermore, it suggests a generally slow or inadequate polymerization process under the chosen conditions. Hence this parameter has already a strong significance for the description of a desired reaction.

Secondly, based on the disparity between experimental results and the fitting an assessment about the credibility of the experimentation run was given. Despite great effort, it has not been possible to give a perfect measure for the extent to which reaction data may deviate from the trend of an ideal process. It should be noted that the fitting curve function was merely an approximation *a priori* and, therefore, unsuitable for some reactions. Hence, to sort reactions similar in conversion and time, yet more by slight distinction in their credibility, the fitting error was ascribed with only 0.5 weight (eqn (5.1)). Moreover, for the calculation of this fitting error ($\text{error}_{\text{fit}}$), the squared error, variance of the two fitting parameters and the covariance between those parameters were sorted and divided into quartiles. Subsequently, every kinetic fit was assigned a score ($\text{score}_{\text{error type}}$) of how good it follows the negative growth fit by scoring points for each quartile further away from the one with the highest error (0 to 3; maximum 12 points in total for the four error types). In this manner, all kinetic fits are ordered according to the aforementioned errors (eqn (5.2)).

Finally, the importance of polymer dispersity (S_D) is recognized with the reciprocal eqn (6). The dispersity represents an important parameter as polymers synthesized by controlled polymerizations and RDRPs should generally exhibit low values. Hence, it is important to consider it in our score. However, the value is also dependent on other circumstances besides the actual polymerization. For example, the accuracy of the molar mass determination by SEC is vital for its determination which, in turn, depends on different parameters such as similarity between the investigated polymer and the polymer standard utilized for calibration.[79] Furthermore, in connection with intermittent sampling, it could not be completely ruled out that small amounts of oxygen entered the reactors, potentially leading to an increase in dispersities compared to literature reactions where no sampling took place. Consequently, a weighting of 0.3 is assigned to this parameter. The arithmetic mean of the dispersities ($Đ_{\text{mean}}$) of the last three sampling times was employed for each kinetic. The intuition score was normalized between 0 and 1.

In conclusion, this logic can effectively categorize the kinetics according to the three mentioned aspects. A high intuition score for a reaction with a specific monomer indicates that the respective combination of reactants and reaction conditions were suitable. A low score, however, does not necessarily mean that this monomer performs poorly in RAFT polymerizations in general. One must keep in mind that the reaction conditions presented here were not optimized for the different monomers, and, hence, different conditions may exist under which the intuition score of the monomer's polymerization would be higher.

$$S_{\text{conversion}} = \frac{0.8 - \text{abs}(\text{conv}_{\text{theo,max}} - 80\%)}{80\%} \quad (4)$$

$$S_{\text{error,fit}} = 0.5 \cdot \frac{12 - \text{error}_{\text{fit}}}{12} \quad (5.1)$$

$$\text{error}_{\text{fit}} = \sum_{\text{error type}=1}^{4} \text{score}_{\text{error type}}(\text{fit}) \quad (5.2)$$

$$S_{D,\text{mean}} = 0.3 \cdot \left(\frac{1}{2 \cdot Đ_{\text{mean}} - 1}\right) \quad (6)$$

An overview of the underlying software architecture can be found in the software documentation in the SI (see Fig. S7).

Our platform, with all its features, aims to be an inspiring service for the actual chemist interested in RAFT polymerizations in general, or a set of specific reactions. We expect the users to have at least one demand for a polymer they want to synthesize in mind (*e.g.*, a fast reaction). From this point, all relevant information will be presented to the users in an interactive tabular form, with the additional possibility to view graph visualization and more detail to examine it. Giving just some examples, the experimentalist can search for the reactants most compatible with a specific RAFT agent which needs to be used (search for positives), or could define all available RAFT agents as well as monomers and see which are not sensible to try out (search for negatives, Fig. 7). The web application is available at **http://www.raft-knowledgebase.de**. An example overview of the website is presented in Fig. S13.

Further accessible knowledge

Based on the data reported in the database, different chemical knowledge is extractable after making it accessible in this way.

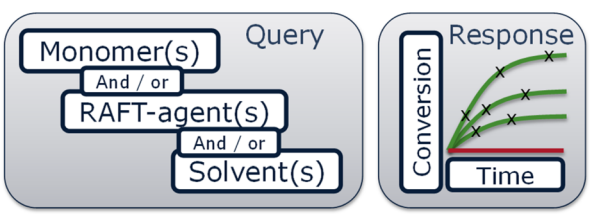

Fig. 7 Example use case of the site, where the user prompts the site through dropdown menus with monomer, CTA and solvent to receive a view of multiple kinetics and information about failed ones at once.

As a specific example, we could find that the conversion for all reactions using styrenic monomers was typically rather low (around 20%) in the investigated time frame. This is most probably due to the chosen reaction conditions used for the generation of our dataset and indicates that, *e.g.*, higher reaction temperatures should be utilized for these monomers in future reactions. Besides that, we could find that the success rate of the polymerization of lauryl acrylate and lauryl methacrylate was underwhelming for most of the polymerizations. Only seven out of 22 lauryl acrylate and five out of 23 lauryl methacrylate experiments yielded adequate conversions. As visible from the entries in the database, only the polymerizations in toluene were successful. This can potentially be attributed to the better solubility of the growing polymer chain inside the less polar toluene compared to dimethyl sulfoxide (DMSO) or dimethylformamide (DMF). The poly(lauryl acrylate) and its methacrylate analog become more nonpolar with each added repetition unit. Hence, at a certain point they precipitate from the reaction solution, forming a gel and the polymerization is discarded as conversion and molar mass values cannot be calculated with confidence.

Outlook for further features

Further features, such as similarity prediction based on simple parameters (*e.g.*, the polarity of solvents), or substructure comparison of the chemical structures of monomer and RAFT agent *via* fingerprints, were considered but ultimately excluded from the platform's initial implementation. We have opted to forego the implementation of such features, which could be perceived as an example of excessive development (feature creep), as it would divert from the idea of this platform representing an explanatory FAIR system.

Experimental

Materials

All chemicals were utilized as received from TCI, Sigma Aldrich, VWR, Roth and Thermo Fisher Scientific if not stated otherwise. The utilized monomers (Table 1) were destabilized *via* a short column of aluminum oxide (neutral aluminum oxide, obtained from Molekula).

Chain transfer agents were purchased from Sigma Aldrich or TCI and used as received (Table 2).

DMF (99.9% purity) was utilized after purification inside a solvent purification system (MBraun, SPS 800). Toluene (Tol) was purchased from VWR in 99.5% purity. DMSO (99.9% purity) was purchased from Sigma Aldrich. 2,2′-Azobis(isobutyronitrile) (AIBN) was purchased from Sigma Aldrich in 98% purity.

The anisole (99% purity) standard is utilized as obtained from Sigma Aldrich. Trioxane (>99% purity) was also purchased from Sigma Aldrich and subsequently dissolved in the respective solvent (DMF, DMSO or toluene) to yield a solution with a concentration of $c(\text{trioxane}) = 1.465\ \text{mol L}^{-1}$. As NMR solvent, deuterated chloroform ($CDCl_3$) (99.8% D) was purchased from Eurisotop. As eluent for the SEC measurements, a mixture of chloroform (>99.8% purity, VWR) (94 vol%), triethylamine (>99.6% purity, VWR) (4 vol%) and iso-propanol (>99.8% purity, VWR) (2 vol%) was utilized.

Table 1 Overview of utilized monomers with their respective abbreviations, purities and suppliers

Monomer	Purity [%]	Supplier
Styrene	99.9	Sigma Aldrich
4-Chlorostyrene	>98	Sigma Aldrich
4-Bromostyrene	97	Sigma Aldrich
4-Methylstyrene	96	Sigma Aldrich
4-Methoxystyrene	>97.5	Sigma Aldrich
4-*tert*-Butylstyrene	93	Sigma Aldrich
Methyl methacrylate	99	Sigma Aldrich
Butyl methacrylate	99	Sigma Aldrich
Lauryl methacrylate	>97	TCI
2-(Dimethylamino)ethyl methacrylate	>98.5	TCI
Benzyl methacrylate	>98	TCI
Methyl acrylate	99	Sigma Aldrich
Butyl acrylate	>99	Sigma Aldrich
Lauryl acrylate	>98	TCI
2-(Dimethylamino)ethyl acrylate	>98	TCI
Benzyl acrylate	>97	TCI

Instruments

^{1}H-NMR spectra were recorded on a Bruker Avance Neo Nanobay (300 MHz) spectrometer equipped with a ^{1}H, ^{13}C, ^{19}F, and ^{31}P-BBO probe and a SampleJet sample changer at room temperature. The chemical shifts are given in parts per million (ppm on δ scale) related to the deuterated solvent.

SEC measurements were performed on the following setup: Agilent 1260 Infinity II series, PSS (degasser), G7110B (pump), G7129A (autosampler), TCC6500 (oven), G7162A (RI detector), and PSS SDV guard column and PSS LinearS (column set). A mixture of 94/4/2 vol% of chloroform, triethylamine, isopropanol was utilized as eluent at 1 mL min^{-1} at 30 °C. Poly(methyl methacrylate) or poly(styrene) was utilized as the standard depending on the polymers (PMMA for poly(methacrylates) and (acrylates); PS for poly(styrenes)).

Description of the setup

The automated polymerizations were performed utilizing a Chemspeed Accelerator SLT100 automated parallel synthesizer robot, equipped with a tool exchange interface, ambient pressure pumps and a 4-needle head (4-NH) with septa piercing needles (which were rounded at the tip) for liquid handling. The 4-NH was connected to two 1-L flasks of chloroform, which was utilized as rinsing solvent between the individual liquid transfers. Furthermore, peripheral heating (Huber Unistat T326) and cooling (LAUDA microcool MC 600) devices were utilized. The robot was equipped with three stock solution racks and a block of 16 individual glass reactors with thermal jackets connected in series through the reactor block. Each individual reactor had a volume of 13 mL. Additionally, all reactors were equipped with reflux condensers (coolant temperature: 5 °C). Homogenization of the polymerization solutions was achieved through vortex

Table 2 Overview of CTAs with their respective abbreviations, purities and suppliers

CTA	Abbreviation	Purity [%]	Supplier
2-Cyano-2-propyl benzodithioate	CPDB	>97	Sigma Aldrich
4-Cyano-4-(phenylcarbonothi–ylthio)pentanoic acid	CPCTTPA	88–112	Sigma Aldrich
2-Phenyl-2-propyl benzodithioate	PPBDT	>98	TCI
2-Cyano-2-propyl dodecyl trithiocarbonate	CPDTTC	>96.5	Sigma Aldrich
2-(Dodecylthiocarbonothioylthio)-2-methylpropionic acid	DTCTTMPA	>98	TCI
Cyanomethyl dodecyl trithiocarbonate	CMDTTC	98	Sigma Aldrich
Benzyl 1*H*-pyrrole-1-carbodithioate	B1HPCDT	97	Sigma Aldrich

agitation at 400 rpm. Moreover, the reactors were connected to an external nitrogen supply and nitrogen was flushed through them from the start of the addition of solvents until the end of the reaction after 15 hours at 70 °C. The previously sealed sampling vials (with slit septa lids) were placed inside two individual sampling vial racks, each with a capacity of 80 vials, for SEC sampling. For NMR-sampling a special NMR-rack for the SampleJet 96 well plate format was manufactured and utilized to automatically prepare the NMR samples without any additional steps.[10] The setup for the polymerization execution is presented in Fig. 8.

Polymerization and sampling workflow

For the kinetic investigation of the polymerizations over 15 h, in the beginning the reactors are flushed with nitrogen. Then, the three different solvents are filled into the respective reactors so that after all subsequent additions (of monomers, standard solutions and RAFT/AIBN solution), the total volume of the solutions in a reactor is 10 mL. Subsequently, the analytical solvents (SEC eluent and deuterated chloroform) are filled in the vials and tubes for the first samples. Thereafter, the monomers (10 mmol), anisole (0.4 mL, 3.66 mmol) or trioxane in the respective solvent (c(trioxane) = 1.465 mol L^{-1}, 2.5 mL, 3.66 mmol) as a standard and finally 1.99 mL of the solution of RAFT agent (c(RAFT) = 33.5 mmol L^{-1}) and AIBN (c(AIBN) = 8.75 mmol L^{-1}) in the respective solvent are added to the reactors. Afterwards, the solutions are deoxygenated for 15 to 30 min by sparging with nitrogen gas. Following this, 0.4 mL solution from each reactor are aspirated with the needle of the robot and 0.2 mL dispensed to an NMR tube prefilled with deuterated chloroform. The remaining volume is dispensed to a SEC vial, prefilled with SEC eluent. These samples are assigned as $t = 0$ h samples with 0% conversion.

After the $t = 0$ h sample from the last reactor is taken, the timestamp for the further samplings is set and the reactors are heated to 70 °C. Afterwards, samples are taken at specified timepoints.

The batch sampling process for SEC is performed at the following timepoints: 0 h, 1 h, 2 h, 4 h, 6 h, 8 h, 10 h and 15 h. The sampling process always follows the same course: before the sampling from the reactor block, the vials are filled with 1 mL of SEC eluent to serve as quenching agent for the radical polymerization. Afterwards, the sample is dispensed to the vial and stands there until the end of the kinetic sampling ($t = 15$ h). Then, the SEC samples are taken out of the synthesizer and filtered through a 0.45 μm PTFE filter into a fresh vial prior to measurement.

The NMR sampling process is performed at 0 h, 1 h, 2 h, 4 h, 8 h and 15 h. The steps for the preparation are as follows: first 0.35 mL of deuterated chloroform is added to the 4″ NMR tubes (prior to sampling), followed by 0.2 mL of sample. Afterwards, the tubes stand until the end of the kinetic sampling. Then, the NMR funnel module is changed and once more 0.4 mL of NMR solvent is added for filling of the tubes. Afterwards, the rack is tilted and shaken manually to homogenize the solutions and then placed inside the SampleJet autosampler for measurement. For a more detailed overview of the program course, see Table S7 in the SI.

For reproducibility and longtime reaction tests, the same workflow was utilized with different sampling and polymerization times of up to 84 h.

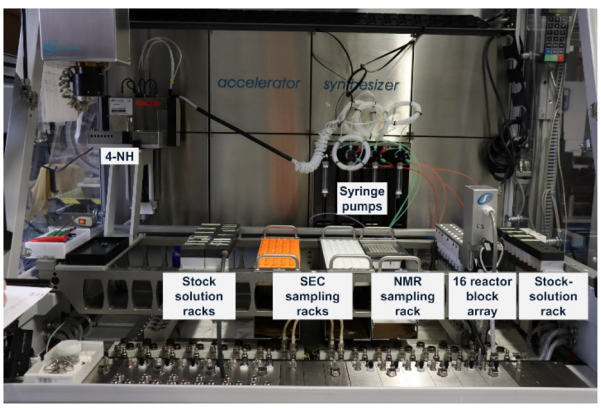

Fig. 8 Overview of the setup inside the parallel synthesizer platform with a reactor block, two SEC sampling racks, NMR sampling rack and three stock solution racks.

Sampling strategy

For the reproducibility of the experiments, it is important that the reactions are always conducted under the same conditions (*e.g.*, chosen reactor array). Hence, the monomers were clustered into three groups (methacrylates, acrylates, and styrenes). A RAFT agent was chosen and one monomer group after the other was sampled. Normally, this took one day per group. As the utilization of the same reactor array for all polymerizations

was not possible due to practical restrictions it was decided to utilize one reactor array per group of monomers.

Conclusions

In conclusion, we present a workflow to generate a new dataset of RAFT homopolymerization kinetics initiated with AIBN at 70 °C as well as the database itself. In detail, we utilized the workflow to generate a total of more than 7200 samples divided into conversion and molar mass data. A web interface was programmed, which allows access to the kinetic data from all over the world. We showed how this data can be utilized to guide polymer chemists, concerning which combination of RAFT, monomer and solvent to choose to yield a specific polymer material. Furthermore, the system is agnostic to the data-amount as long as it contains the same reaction parameters and, thus, can be easily made more comprehensive by expanding the database with additional reactions. The intuition score provides a rapid and practical, yet inherently subjective, overview of the dataset. It offers users insight into the authors' expertise while alerting about underlying biases common in all datasets. Ultimately, this work provides a tool to do (possibly big) data analysis with one's own neural network and assistance from the chemists behind it.

In future work, we aim to widen the scope of the dataset by implementing new variables to the reaction space. Targeted variables are the utilization of further initiators, polymerization temperatures, monomer concentrations, initiator-to-CTA ratios and further solvents (*e.g.*, tetrahydrofuran) as well as CTAs.

Author contributions

Conceptualization: MR and SZ; setup of the automated platform: MR; investigation & data curation: MR and YK; kinetic fitting: YK; programming of website: YK; writing – original draft: MR and YK; writing – review & editing: MR, YK, SZ and USS; supervision: SZ and USS; resources: USS; project administration: USS; funding acquisition: USS. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

There are no conflicts to declare.

Data availability

The primary data in an Excel spreadsheet as well as the curated database and software for curation and running the webserver are available at: **https://github.com/Bizbalt/RAFT-knowledgebase** and **https://doi.org/10.5281/zenodo.15230215** (5.1 MB). Furthermore, the raw data are available in a separate Zenodo archive at **https://doi.org/10.5281/zenodo.14018513** (3.3 GB, compressed).

The website is available at the domain **https://www.raft-knowledgebase.de**.

Supplementary information: example SEC curves and NMR spectra, overview of performed polymerizations, table of robot program workflow, program diagram of the software, information on classification of the intial gathered data. See DOI: **https://doi.org/10.1039/d5dd00177c**.

Acknowledgements

The authors thank the "Deutsche Forschungsgemeinschaft" for funding under the regime of the priority program SPP 2363 "Utilization and Development of Machine Learning for Molecular Applications – Molecular Machine Learning" (SCHU 1229/63-1; project number 497115849). Additionally, we thank the Thüringer Aufbaubank (VFE 1000241) for financial support. Furthermore, we thank the Free State of Thuringia under no. 2016 IZN 0009 for financial support, cofinanced by funds from the European Union within the framework of the European Regional Development Fund (EFRE). Also, Ludwig Büttner and Jacob Meyer are acknowledged for help with the experimental execution. We would also like to thank Dr Grit Festag for help with the SEC equipment as well as Leon Lange for help with the SampleJet NMR.

References

1 K. H. Bleicher, H.-J. Böhm, K. Müller and A. I. Alanine, *Nat. Rev. Drug Discovery*, 2003, **2**, 369–378.
2 S. W. Krska, D. A. DiRocco, S. D. Dreher and M. Shevlin, *Acc. Chem. Res.*, 2017, **50**, 2976–2985.
3 S. M. Mennen, C. Alhambra, C. L. Allen, M. Barberis, S. Berritt, T. A. Brandt, A. D. Campbell, J. Castañón, A. H. Cherney, M. Christensen, D. B. Damon, J. Eugenio de Diego, S. García-Cerrada, P. García-Losada, R. Haro, J. Janey, D. C. Leitch, L. Li, F. Liu, P. C. Lobben, D. W. C. MacMillan, J. Magano, E. McInturff, S. Monfette, R. J. Post, D. Schultz, B. J. Sitter, J. M. Stevens, I. I. Strambeanu, J. Twilton, K. Wang and M. A. Zajac, *Org. Process Res. Dev.*, 2019, **23**, 1213–1242.
4 M. Shevlin, *ACS Med. Chem. Lett.*, 2017, **8**, 601–607.
5 J. N. Cawse, *Acc. Chem. Res.*, 2001, **34**, 213–221.
6 J. G. deVries and A. H. deVries, *Eur. J. Org Chem.*, 2003, **2003**, 799–811.
7 T. N. Nguyen, T. T. P. Nhat, K. Takimoto, A. Thakur, S. Nishimura, J. Ohyama, I. Miyazato, L. Takahashi, J. Fujima, K. Takahashi and T. Taniike, *ACS Catal.*, 2020, **10**, 921–932.
8 W. F. Maier, *ACS Comb. Sci.*, 2019, **21**, 437–444.
9 S. Oliver, L. Zhao, A. J. Gormley, R. Chapman and C. Boyer, *Macromolecules*, 2019, **52**, 3–23.
10 M. Ringleb, L. Klähn, S. Zechel and U. S. Schubert, *ChemPlusChem*, 2025, **90**, e202400686.
11 C. R. Becer, A. M. Groth, R. Hoogenboom, R. M. Paulus and U. S. Schubert, *QSAR Comb. Sci.*, 2008, **27**, 977–983.
12 H. Zhang, V. Marin, M. W. M. Fijten and U. S. Schubert, *J. Polym. Sci., Part A: Polym. Chem.*, 2004, **42**, 1876–1885.
13 R. Hoogenboom and U. S. Schubert, *Rev. Sci. Instrum.*, 2005, **76**, 1–7.

14 L. I. Majoros, B. Dekeyser, R. Hoogenboom, M. W. M. Fijten, J. Geeraert, N. Haucourt and U. S. Schubert, *J. Polym. Sci., Part A: Polym. Chem.*, 2010, **48**, 570–580.
15 N. Zaquen, M. Rubens, N. Corrigan, J. Xu, P. B. Zetterlund, C. Boyer and T. Junkers, *Prog. Polym. Sci.*, 2020, **107**, 101256.
16 C. Rosenfeld, C. Serra, S. O'Donohue and G. Hadziioannou, *Macromol. React. Eng.*, 2007, **1**, 547–552.
17 C. R. Becer, R. M. Paulus, R. Hoogenboom and U. S. Schubert, *J. Polym. Sci., Part A: Polym. Chem.*, 2006, **44**, 6202–6213.
18 C. Guerrero-Sanchez, C. Abeln and U. S. Schubert, *J. Polym. Sci., Part A: Polym. Chem.*, 2005, **43**, 4151–4160.
19 C. Guerrero-Sanchez, R. M. Paulus, M. W. Fijten, M. J. de La Mar, R. Hoogenboom and U. S. Schubert, *Appl. Surf. Sci.*, 2006, **252**, 2555–2561.
20 A. W. Bosman, A. Heumann, G. Klaerner, D. Benoit, J. M. Fréchet and C. J. Hawker, *J. Am. Chem. Soc.*, 2001, **123**, 6461–6462.
21 M. J. Nasrullah and D. C. Webster, *Macromol. Chem. Phys.*, 2009, **210**, 640–650.
22 X. Liu, C.-G. Wang and A. Goto, *Angew. Chem., Int. Ed.*, 2019, **58**, 5598–5603.
23 S. Perrier, *Macromolecules*, 2017, **50**, 7433–7447.
24 J. Chiefari, Y. K. Chong, F. Ercole, J. Krstina, J. Jeffery, T. P. T. Le, R. T. A. Mayadunne, G. F. Meijs, C. L. Moad, G. Moad, E. Rizzardo and S. H. Thang, *Macromolecules*, 1998, **31**, 5559–5562.
25 R. T. A. Mayadunne, E. Rizzardo, J. Chiefari, Y. K. Chong, G. Moad and S. H. Thang, *Macromolecules*, 1999, **32**, 6977–6980.
26 P. Corpart, D. Charmot, S. Z. Zard, T. Bladatti and D. Michelet, *US Pat.*, US6153705 (A), 1999.
27 D. Charmot, P. Corpart, H. Adam, S. Z. Zard, T. Biadatti and G. Bouhadir, *Macromol. Symp.*, 2000, **150**, 23–32.
28 C. Barner-Kowollik, *Handbook of RAFT polymerization*, Wiley-VCH, Weinheim, 2008.
29 G. Moad, E. Rizzardo and S. H. Thang, *Aust. J. Chem.*, 2005, **58**, 379.
30 M. H. Stenzel, L. Cummins, G. E. Roberts, T. P. Davis, P. Vana and C. Barner-Kowollik, *Macromol. Chem. Phys.*, 2003, **204**, 1160–1168.
31 A. J. Gormley, J. Yeow, G. Ng, Ó. Conway, C. Boyer and R. Chapman, *Angew. Chem., Int. Ed.*, 2018, **57**, 1557–1562.
32 M. W. M. Fijten, R. M. Paulus and U. S. Schubert, *J. Polym. Sci., Part A: Polym. Chem.*, 2005, **43**, 3831–3839.
33 J. van Herck, I. Abeysekera, A.-L. Buckinx, K. Cai, J. Hooker, K. Thakur, E. van de Reydt, P.-J. Voorter, D. Wyers and T. Junkers, *Digital Discovery*, 2022, **1**, 519–526.
34 B. Zhang, A. Mathoor and T. Junkers, *Angew. Chem., Int. Ed.*, 2023, **62**, e202308838.
35 M. Rubens, J. van Herck and T. Junkers, *ACS Macro Lett.*, 2019, **8**, 1437–1441.
36 M. Rubens, J. H. Vrijsen, J. Laun and T. Junkers, *Angew. Chem., Int. Ed.*, 2019, **58**, 3183–3187.
37 M. W. M. Fijten, M. A. R. Meier, R. Hoogenboom and U. S. Schubert, *J. Polym. Sci., Part A: Polym. Chem.*, 2004, **42**, 5775–5783.
38 J. de Breuck, M. Streiber, M. Ringleb, D. Schröder, N. Herzog, U. S. Schubert, S. Zechel, A. Traeger and M. N. Leiske, *ACS Polym. Au*, 2024, **4**, 222–234.
39 M. Semsarilar and S. Perrier, *Nat. Chem.*, 2010, **2**, 811–820.
40 G. K. K. Clothier, T. R. Guimarães, S. W. Thompson, J. Y. Rho, S. Perrier, G. Moad and P. B. Zetterlund, *Chem. Soc. Rev.*, 2023, **52**, 3438–3469.
41 J. Wan, B. Fan and S. H. Thang, *Chem. Sci.*, 2022, **13**, 4192–4224.
42 Y. A. Vasilieva, C. W. Scales, D. B. Thomas, R. G. Ezell, A. B. Lowe, N. Ayres and C. L. McCormick, *J. Polym. Sci., Part A: Polym. Chem.*, 2005, **43**, 3141–3152.
43 A. P. Grimm, S. T. Knox, C. Y. P. Wilding, H. A. Jones, B. Schmidt, O. Piskljonow, D. Voll, C. W. Schmitt, N. J. Warren and P. Théato, *Macromol. Rapid Commun.*, 2025, **46**, e2500264.
44 V. Klimkevicius, M. Steponaviciute and R. Makuska, *Eur. Polym. J.*, 2020, **122**, 109356.
45 H. Zhang, J. Deng, L. Lu and Y. Cai, *Macromolecules*, 2007, **40**, 9252–9261.
46 K. Ponnusamy, R. P. Babu and R. Dhamodharan, *J. Polym. Sci., Part A: Polym. Chem.*, 2013, **51**, 1066–1078.
47 G. Puts, V. Venner, B. Améduri and P. Crouse, *Macromolecules*, 2018, **51**, 6724–6739.
48 X. Pan, F. Zhang, B. Choi, Y. Luo, X. Guo, A. Feng and S. H. Thang, *Eur. Polym. J.*, 2019, **115**, 166–172.
49 A. R. Santha Kumar, M. Roy and N. K. Singha, *Eur. Polym. J.*, 2018, **107**, 294–302.
50 I. Fraga Domínguez, J. Kolomanska, P. Johnston, A. Rivaton and P. D. Topham, *Polym. Int.*, 2015, **64**, 621–630.
51 J. Chiefari, T. A. M. Roshan, C. L. Moad, G. Moad, E. Rizzardo, A. Postma, M. A. Skidmore and S. H. Thang, *Macromolecules*, 2002, **36**, 2273–2283.
52 Y. K. Chong, J. Krstina, T. P. T. Le, G. Moad, A. Postma, E. Rizzardo and S. H. Thang, *Macromolecules*, 2003, **36**, 2256–2272.
53 D. J. Keddie, G. Moad, E. Rizzardo and S. H. Thang, *Macromolecules*, 2012, **45**, 5321–5342.
54 M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. Da Silva Santos, P. E. Bourne, J. Bouwman, A. J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S. J. Lusher, M. E. Martone, A. Mons, A. L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M. A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao and B. Mons, *Sci. Data*, 2016, **3**, 160018.
55 A. Dunning, M. de Smaele and J. Böhmer, *Int. J. Digit. Curation*, 2017, **12**, 177–195.
56 R. Mercado, S. M. Kearnes and C. W. Coley, *J. Chem. Inf. Model.*, 2023, **63**, 4253–4265.

57 F. Strieth-Kalthoff, S. Szymkuć, K. Molga, A. Aspuru-Guzik, F. Glorius and B. A. Grzybowski, *J. Am. Chem. Soc.*, 2024, **146**, 11005–11017.
58 E. Mosqueira-Rey, E. Hernández-Pereira, D. Alonso-Ríos, J. Bobes-Bascarán and Á. Fernández-Leal, *Artif. Intell. Rev.*, 2023, **56**, 3005–3054.
59 M. Ringleb, T. Schuett, S. Zechel and U. S. Schubert, *Digital Discovery*, 2023, **2**, 1883–1893.
60 T. Schuett, J. Kimmig, S. Zechel and U. S. Schubert, *Polymers*, 2022, **14**, 292.
61 J. Qiu, B. Charleux and K. Matyjaszewski, *Polimery*, 2001, **46**, 454–460.
62 H. Fischer and L. Radom, *Angew. Chem., Int. Ed.*, 2001, **40**, 1340–1371.
63 T. G. McKenzie, Q. Fu, E. H. H. Wong, D. E. Dunstan and G. G. Qiao, *Macromolecules*, 2015, **48**, 3864–3872.
64 M. J. Monteiro and H. de Brouwer, *Macromolecules*, 2001, **34**, 349–352.
65 J. Qiu, B. Charleux and K. Matyjaszewski, *Prog. Polym. Sci.*, 2001, **26**, 2083–2134.
66 X. Zhang, O. Giani, S. Monge and J.-J. Robin, *Polymer*, 2010, **51**, 2947–2953.
67 P. Vana, T. P. Davis and C. Barner-Kowollik, *Macromol. Theory Simul.*, 2002, **11**, 823–835.
68 F. Strieth-Kalthoff, F. Sandfort, M. Kühnemund, F. R. Schäfer, H. Kuchen and F. Glorius, *Angew. Chem., Int. Ed.*, 2022, **61**, e202204647.
69 G. K. K. Clothier, T. R. Guimarães, S. W. Thompson, S. C. Howard, B. W. Muir, G. Moad and P. B. Zetterlund, *Angew. Chem., Int. Ed.*, 2024, **136**, e202320154.
70 G. Wypich, *Handbook of Polymers*, ChemTec Publishing, Toronto, Canada, 1st edn, 2012.
71 *Polymer Handbook*, ed. D. R. Bloch, Wiley, Hoboken, N. J., 4th edn, 1999.
72 C. Kuenneth and R. Ramprasad, *Nat. Commun.*, 2023, **14**, 4099.
73 M. Ishii, T. Ito, H. Sado and I. Kuwajima, *Sci. Technol. Adv. Mater.:Methods*, 2024, **4**, 2354649.
74 T. D. Huan, A. Mannodi-Kanakkithodi, C. Kim, V. Sharma, G. Pilania and R. Ramprasad, *Sci. Data*, 2016, **3**, 160012.
75 R.-R. Griffiths, P. Schwaller and A. A. Lee, Dataset Bias in the Natural Sciences: A Case Study in Chemical Reaction Prediction and Synthesis Design, *arXiv*, 2021, preprint, arXiv:2105.02637, DOI: **10.48550/arXiv.2105.02637**, **https://arxiv.org/pdf/2105.02637**.
76 D. Weininger, *J. Chem. Inf. Comput. Sci.*, 1988, **28**, 31–36.
77 M. Boeckhout, G. A. Zielhuis and A. L. Bredenoord, *Eur. J. Hum. Genet.*, 2018, **26**, 931–936.
78 J. Vandenbergh, T. de Moraes Ogawa and T. Junkers, *J. Polym. Sci., Part A: Polym. Chem.*, 2013, **51**, 2366–2374.
79 T. Shimizu, R. Whitfield, G. R. Jones, I. O. Raji, D. Konkolewicz, N. P. Truong and A. Anastasaki, *Chem. Sci.*, 2023, **14**, 13419–13428.