



**Exception Search in Databases for Polymers with Practically  
Contradictory Properties of Heat Resistance and  
Transparency**

Journal:	<i>Polymer Chemistry</i>
Manuscript ID	PY-ART-05-2023-000565.R1
Article Type:	Paper
Date Submitted by the Author:	27-Jun-2023
Complete List of Authors:	Nakamura, Yasuyuki; National Institute for Materials Science Gros, Alice; National Institute for Materials Science Zhang, Wenhao; The University of Tokyo Graduate School of Frontier Sciences Sodeyama, Keitaro; National Institute for Materials Science Naito, Masanobu; National Institute of Materials Science; The University of Tokyo Graduate School of Frontier Sciences

## COMMUNICATION

## Exception Search in Databases for Polymers with Practically Contradictory Properties of Heat Resistance and Transparency

Yasuyuki Nakamura,<sup>\*†‡a</sup> Alice Gros,<sup>‡b</sup> Wenhao Zhang,<sup>c</sup> Keitaro Sodeyama,<sup>b</sup> and Masanobu Naito<sup>\*a,c</sup>

Received 00th January 20xx,  
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Although the common data-driven studies for material development use property values present in the database, the effectiveness of analysing and exploring the areas without property values in the database has hardly been clarified. The necessity to analyse such area is evident for the fact that only 44% of the polymer chemical structures entries with property values are used for studies, while the remaining 56% are not in the analysis of glass transition temperature using PoLyInfo database. In this study, a method to discover polymer materials with undiscovered properties was demonstrated and experimentally verified. The study utilised a comprehensive database of polymer materials and properties with an exceptional searching procedure that defines unexplored area in the database. In addition, a filter based on the physicochemical mechanism of polymers was used to remove the common molecular structures to reveal polymers with contradictory properties. Similarly, a mechanism-based filter was also used to narrow down the candidates efficiently. We investigated a practically challenging heat resistant transparent polymer material with this procedure, and experimentally verified the screening candidate obtained by molecular dynamic simulations and machine-learning predictions. Consequently, potential polymer materials were discovered with thermal degradation temperature,  $T_{deg}$ , higher than 300 °C without glass transition in the entire temperature range, and high transparency (more than 80% transmittance) in the visible and UV region.

Databases on structure and property correlation of compounds and materials are indispensable for data-driven the development of polymer material.<sup>1–3</sup> One of a common use of these databases is the extraction of the chemical structure of monomers or polymers with target properties, which are used as training data for supervised learnings (Figure 1a). Molecules or materials design by such supervised learning is extensively studied recently,<sup>4,5</sup> however, it is not always practical for polymer material development, because of the difficulty and the time cost of an experimental verification of proposed candidate.<sup>6–10</sup>

Alternative interesting usage of a database is finding a potential material by the exploration of unidentified properties or functionalities of the known materials included in the database. With a comprehensive database such finding should be possible, because the database includes unidentified chemical space which are not recognized by human's insight. In

this exploration, the presence of information regarding synthetic accessibility is a significant advantage for experimental verification. However, only few studies of polymer materials focus on utilising the databases by screening unexplored space rather than simply using them as a training data for supervised learnings.<sup>11</sup>

To explore and define the unexplored areas in the database, it is essential that: first, the database is comprehensive and includes as many entries and properties as possible to ensure a good representation of existing polymers. Second, the database should be retrievable with the conditions for the specific purpose. There are several databases of polymers and polymer materials including PoLyInfo<sup>12</sup>, Polymer Genome<sup>13</sup>, and Polymer Property Predictor<sup>14</sup>. Among them, PoLyInfo is the database containing the most experimental polymer material property entries along with structural information (in simplified molecular input line entry system, SMILES).<sup>2</sup> In addition, the database allows filtering in a machine-readable structure based on polymer class, structural similarity, property category, and other factors, in a machine-readable format, as well as creating objective-oriented datasets. Thus, PoLyInfo meets the two above-stated criteria.

Thermal properties of polymers are well studied, and numerous values are present in the database, as they determine essential factors in material applications. For heat-resistant polymer materials, the glass transition temperature ( $T_g$ ) is used to determine the usage temperature range because various properties, such as viscoelasticity and linear expansion,

<sup>a</sup> Data-driven Polymer Design Group, Research Center for Macromolecules and Biomaterials, National Institute for Materials Science (NIMS), Tsukuba, Ibaraki, Japan. E-mail: NAKAMURA.Yasuyuki@nims.go.jp; NAITO.Masanobu@nims.go.jp

<sup>b</sup> Data-driven Materials Design Group, Center for Basic Research on Materials, and Energy Materials Design Group, Research and Services Division of Materials Data and Integrated System (MaDIS), National Institute for Materials Science (NIMS), Tsukuba, Ibaraki, Japan.

<sup>c</sup> Department of Advanced Material Science, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Japan.

† Footnotes relating to the title and/or authors should appear here.

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

‡ These authors contributed equally to this work.

vary at this point. Another important parameter is the thermal degradation temperature ( $T_{deg}$ ), which is the point of material breakdown. Heat-resistant polymer materials are required to have not only high  $T_g$  and  $T_{deg}$  but also other properties, such as transparency, which is desired, for example, for bonding glass and film, LED encapsulant, and photomasks.

The incorporation of aromatic rings (or  $\pi$ -conjugated structures) into the polymer main chain is an effective method for increasing thermal stability owing to their rigid and chemically stable characteristics. However, a disadvantage of the aromatic ring is the colouring and low optical transmission, as illustrated by the deep brown colour of aromatic polyimides and the opacity of polyether ether ketone (PEEK). However, representative transparent polymers with aliphatic chain structures, such as poly(methacrylate)s, poly(carbonate)s, and poly(cycloolefin)s, have low  $T_g$  values (mostly under 150 °C), which limit their applications. Therefore, heat resistance and transparency are contradictory properties in polymer materials in practice, and the identification of polymers that satisfy high heat resistance requirements without aromatic main chains remains challenging.

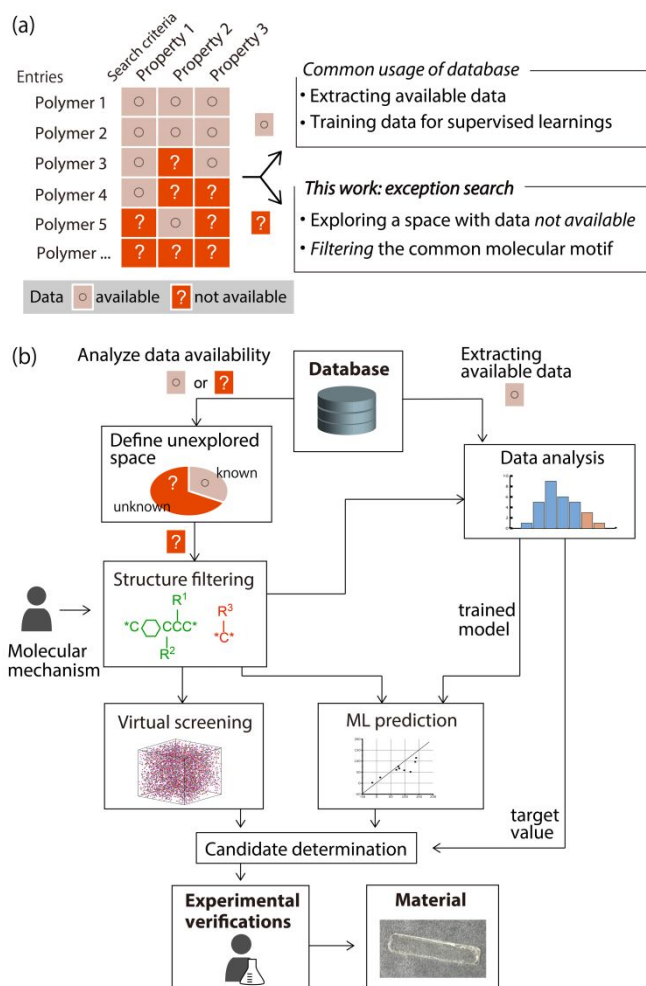


Figure 1. (a) Concept of exception search in this study, which illustrates the data availability of properties as search criteria for each polymer entries in a database. (b) Workflow of this study.

In this study, we focus on the unexplored space inside the database for the quick development of materials. Here we demonstrated this method by finding heat-resistant and transparent polymer materials verified by synthetic experiments with small burden. Figure 1 shows our workflow. First, the PoLyInfo database was analysed to define the unexplored property region in terms of  $T_g$ . And then, the defined region was further filtered through the exclusion of the aromatic main chain polymers. Second, to narrow down the exploration space efficiently, the filter of polymer molecular structure based on the physicochemical mechanism regarding  $T_g$  was adopted. A small space for candidates was shown to benefit them in virtual screening by molecular dynamic (MD) simulations and machine learning prediction for lower computational cost. The screening effectively outputted candidates for experimental verification. Finally, the promising candidates, a series of poly(metal fumarate)s, were synthesised, and the relevant heat-resistant and colourless transparent properties were identified, which could pass visible and UV light.

## Results and Discussion

First the  $T_g$  and  $T_{deg}$  values available in PoLyInfo were extracted and the distribution of these values with the classification of the main chain with or without aromatic rings were plotted in Figure 2. The plot for  $T_g$  show that the aliphatic polymers dominate at lower temperatures, while aromatic polymers become dominant at higher temperatures (Figure 2a). The population of aliphatic polymers was significantly smaller as less than 10% in each bin above 200 °C, and the aliphatic polymers for which  $T_g$  was >200 °C were the target of interest.

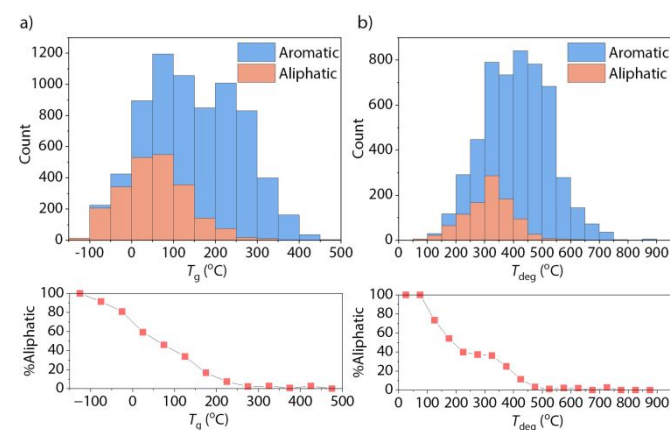


Figure 2. Histogram of  $T_g$  (a) or  $T_{deg}$  (b) of polymers in PoLyInfo (top) and the ratio of aliphatic polymers in each bin (bottom). The colour of bars indicates the main chain structure of polymers.

The same tendency was observed for  $T_{deg}$  for the population distribution of polymers with/without aromatic rings in the main chain (Figure 2b). The  $T_g$  and  $T_{deg}$  values have a positive correlation (Figure S1), regardless of the different physicochemical factors or mechanism of them.  $T_g$  is related to the dynamic flexibility of polymer chains, while  $T_{deg}$  is related to the chemical structure which is susceptible to thermal cleavage

or transformations. Therefore, searching for a polymer with a high  $T_g$  would give polymers with both high  $T_g$  and  $T_{deg}$  values.

Figure 3 shows the number of whole homopolymer entries in PoLyInfo categorized by the availability of  $T_g$  or  $T_{deg}$  and the main chain structure. The ratio of polymer entries without  $T_g$  data is 56%, and there are 3866  $T_g$  not-reported aliphatic polymers in the PoLyInfo database (Figure 3), which are interested in this study. This is a much lower number compared to those in common library generative machine learning studies, though it remains too large for the next virtual screening step.

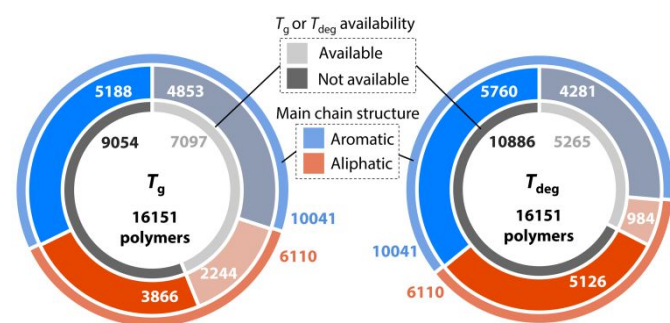


Figure 3. Number of polymers in PoLyInfo database categorised by the availability of  $T_g$  or  $T_{deg}$ , and/or main chain molecular structure.

The mechanism of glass transition is considered to involve the flexibility of the main chain.<sup>15,16</sup> Most main-chain aliphatic polymers have methylene units as seen in vinyl polymers. The presence of the methylene unit increases the flexibility of the chain and causes the  $T_g$  value to decrease. Therefore, the chemical space in this study was narrowed down according to a molecular mechanism-based assumption that all-carbon substituted polymers with C1-repeating chain structures have higher  $T_g$  values compared to the corresponding C2 or vinyl polymers. Substituted methylene polymers,  $(CR^1R^2)_n$  ( $R^1$ ,  $R^2$  = substituent group or H, denoted as “C1 polymers”), have substituents on all of the repeating carbon units, which is in contrast to common vinyl polymers,  $(CH_2CR^1R^2)_n$  (denoted as “C2 polymers”).<sup>17,18</sup> In particular, the densely substituted structures are expected to have a significant effect on suppressing the flexibility of the main chain, which results in high  $T_g$  values. However, there has been little knowledge on the potentially useful properties of these polymers.

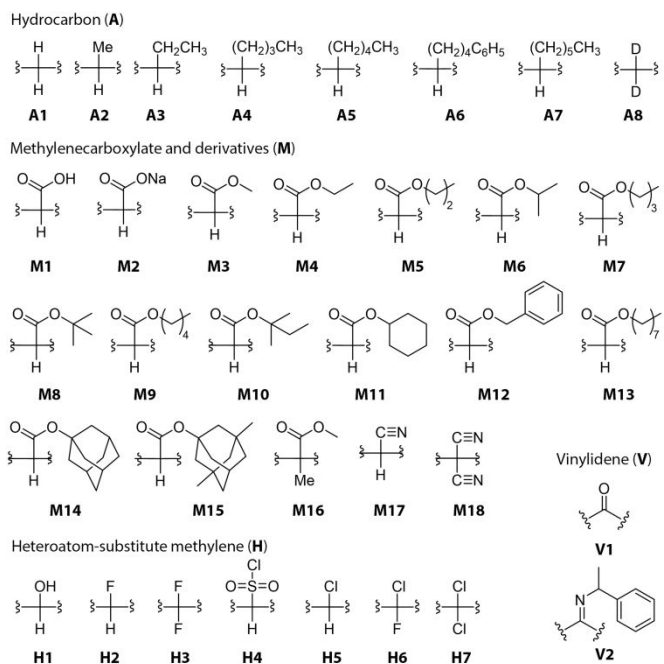


Chart 1. Extraction from PoLyInfo: repeating unit structures of polymers with substituents on all the carbon atom in the main chain.

The filtration based on C1-polymer structure was applied to the listed 3866 polymers and gave 35 candidates. These polymers were categorised into four groups by structural similarity, as shown in Chart 1. In searching for a high  $T_g$  polymer, group A was excluded due to the expected highly flexible structure and low  $T_g$ . The comparison of known  $T_g$  of C1 and the C2 polymers showed a clear tendency for C1 polymers to have higher  $T_g$  values than the corresponding C2 polymers (Figure S2). The difference roughly reflects the steric bulkiness, especially in the space close to the carbon main chain. This tendency agrees with the known mechanisms for the  $T_g$  values. A similar analysis was also attempted for  $T_{deg}$ , however, too few values were available.

The estimation of  $T_g$  values through MD simulations provided useful information for identifying synthetic targets. The  $T_g$  values of the C1 polymers and structurally corresponding C2 polymers were examined to compensate for the small number of experimental  $T_g$  values of C1 polymers under consideration and to examine the reliability of the calculations (Figure 4). The simulation was carried out with small cells, a general force field, and simulation parameters for versatility and low computational cost, which are acceptable for the selection of synthetic target (for the detail of the simulation conditions, see Supplementary Material). The deviation between known experimental and calculated  $T_g$  values ranged from -23 to 58 °C in each polymer, which indicated even small simulation cells gives reasonable information about the tendency of  $T_g$  for the present purpose of screening the candidate polymers. Accounting for this deviation, **M2**, **M8**, **M10**, **M14**, and **M18** were suggested to have  $T_g$  values above 200 °C among the listed polymers without reported experimental  $T_g$ . **M8** and **M10** are known to thermally

decompose at 180 °C ( $T_{\text{deg}}$ ), as experimentally examined later. **M14** has been recently

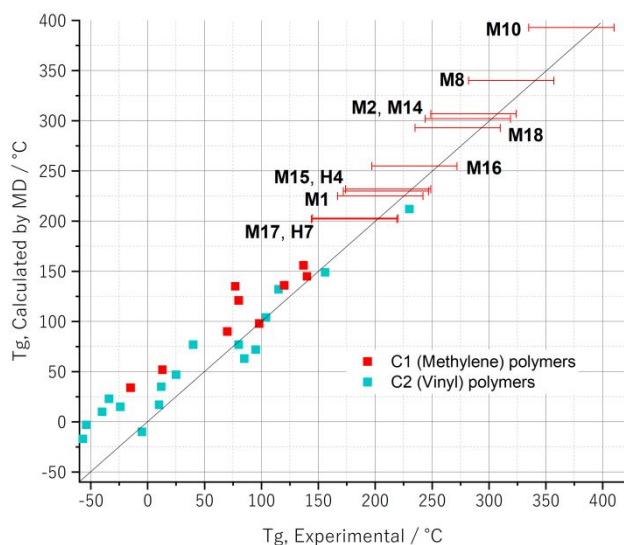


Figure 4. Verification of calculated glass transition temperature of substituted methylene polymers (C1 polymers) listed in PoLyInfo. Error bars shown are based on the estimated deviation of calculated values from the experimental values (-23 to 58 °C, see the text) for polymers with reported  $T_g$ . Group A polymers were excluded.

reported as an excellent heat-resistant polymer with a high  $T_{\text{deg}}$  value of 384 °C.<sup>19,20</sup> **M18** has been reported to absorb visible light, and the structure is unconfirmed.<sup>21</sup> On the other hand, **M2** is synthetically accessible, and the time-dependent density functional theory (TD-DFT) calculations indicated no absorption in visible, UV-A and UV-B regions, along with the rigid structure by the interaction between carboxylates and  $\text{Na}^+$  (Figure S3). Thus, **M2** is the optimal candidate for synthetic verification.

The  $T_g$  predictions were performed on the C1 polymers list, using the polymer properties prediction tool SMILES-X<sup>22</sup> using  $T_g$  data in PoLyInfo (see Figure 1b for the workflow, right part). The prediction model was constructed with the training dataset curated from PoLyInfo for all the available  $T_g$  values, and the model had a fair score of root-mean-square error (RMSE) value of  $36.3 \pm 3$  °C. It showed good agreement with the 9 experimentally known values among the C1 polymers, as expected as they are present in the training dataset. Punctual potential deviations (**M1** and **M8** for examples, for which no experimental  $T_g$  under 180 °C has been reported), attributed to a lack of available training data for the C1 polymers, are observed, and suggest an underestimation of high value of  $T_g$ . Importantly, it is still worth noting that **M2** is also confirmed to be a promising candidate as with MD simulations, given that it had the highest predicted  $T_g$  value among the C1 polymer list.

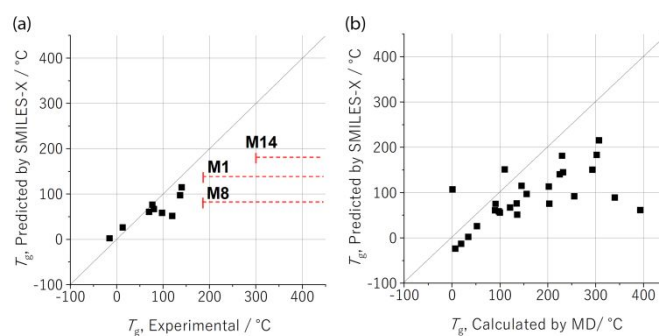
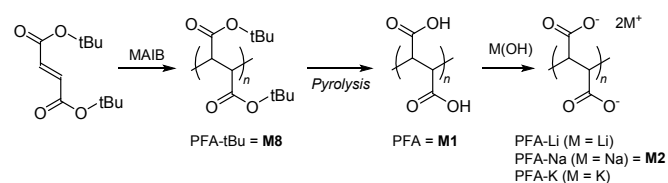


Figure 5. Comparison of  $T_g$  predicted by (a) SMILES-X and experimental and (b) MD-calculated values. Dashed red lines in (a) indicate the expected range of the experimental  $T_g$  values judging from the reported information. Group A polymers were excluded.

The carboxylate-substituted methylene polymers in **M** group have been reported to be synthesised by the polymerisation of dialkyl fumarates<sup>23,24</sup> or catalytic polymerisation of diazo compounds.<sup>25–27</sup> Therefore, the polymers are also known as polyfumarates. The polymers are the methine analogues of polyacrylates and are used in such as coatings and fibres.<sup>28,29</sup> Although there are several reports of applications of **M2** polymer and analogues (poly(metal fumarate)s) as electrolytes for use in film application<sup>30–32</sup> or as a liquid crystalline polymer,<sup>33–35</sup> fundamental properties of poly(metal fumarate)s have hardly been studied and there are no detailed reports on their thermal properties.

In the synthetic experimental verification, first the precursor poly(fumaric acid) (PFA) was synthesised according to the reported method (Scheme 1),<sup>23,24</sup> and the synthesis of poly(*di-tert*-butyl fumarate) (PFA-*t*Bu) is summarised in Table S1. The following pyrolysis was carried out as reported to yield PFA at approximately 85% yield. The resulting polymer was slightly brown, which is likely due to the unexpected decomposition of the polymer during the pyrolysis.<sup>36</sup>



Scheme 1. Synthesis of poly(metal fumarate)s.

And then, the poly(metal fumarate)s (PFA-M,  $\text{M}^+ = \text{Li}^+, \text{Na}^+, \text{K}^+$ ) were synthesised by the treatment of PFA ( $M_n = 4.18 \times 10^4$ ,  $\mathcal{D} = 3.70$ ) with metal hydroxides. The target polymer PFA-Na (= **M3**) was readily prepared, and in addition, the synthetic experiments expanded the variation of polymers with PFA-Li and PFA-K, which are not listed in PoLyInfo. These results are a benefit of the present exploration method with experimentations. The solubility of PFA-M is highly dependent on the counteraction,  $\text{M}^+$ . Although PFA-K and PFA-Li were readily soluble in water, the low solubility of PFA-Na in water and other solvents made it difficult to prepare its thick film (Figure 6).

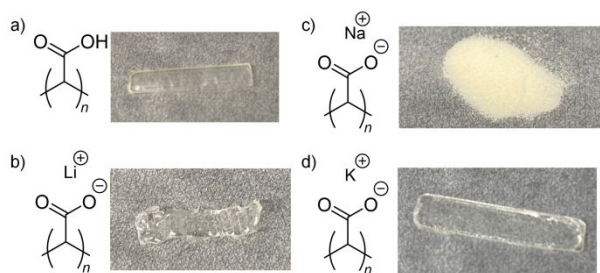


Figure 6. Pictures of polymers. PFA-Li (b) was easy to shrink during preparation. PFA-Na was difficult to prepare a thick film due to its low solubility in water.

For all polymers, thermal analyses showed remarkable heat-resistant properties. In differential scanning calorimetry (DSC) analysis, no signal of glass transition was observed up to approximately 50 °C below the thermal decomposition temperature (Figure S4). High  $T_g$  or the nature of no glass transition can be explained by the non-flexible chain structure of polyfumarates.<sup>17</sup>

The thermal degradation analysis of poly(metal fumarate)s is shown in Figure 7 and summarised in Table 1. The degradation of the polymer was carried out in an atmosphere consisting of a  $N_2+O_2$  mixture at a pressure ratio of 80:20 to simulate air, and therefore the residual weight of the PFA metal salts is ascribed to the carbonate of the metals. PFA-M exhibited higher thermal stability than PFA. The weight loss at approximately 300 °C for PFA could be assigned to the degradation derived from  $CO_2H$  group, and this indicates that the higher heat resistance of PFA-M is due to the formation of thermally stable metal carboxylate. Indeed, the measurement for a series of PFA-Na prepared with different amounts of sodium hydroxide revealed the disappearance of the derivative thermogravimetry (DTG) peak at approximately 300 °C, alongside an increase in the equivalence of NaOH to  $CO_2H$  ( $N_{NaOH}/N_{CO_2H}$ ) (Figure S5). The plot of the equivalence of NaOH, the residual weight, and DTG peak intensity at 285 °C clearly shows a linear correlation. Although the conversion of  $CO_2H$  to  $CO_2M$  is the primary factor in thermal stabilisation, intra- and intermolecular interactions of polymer chain(s) and Na cations will also affect it.

The thermogravimetry analysis (TGA) profiles of PFA-M depended on the metals, and the order of 5% weight loss temperature ( $T_{deg}$ ) and the main DTG peak temperature was as follows:  $K < Li < Na$ . This order does not reflect the physical properties of metal cations but indicates that the thermal stabilisation effect of salt is related to the intra- and/or interaction of the polymer chains.

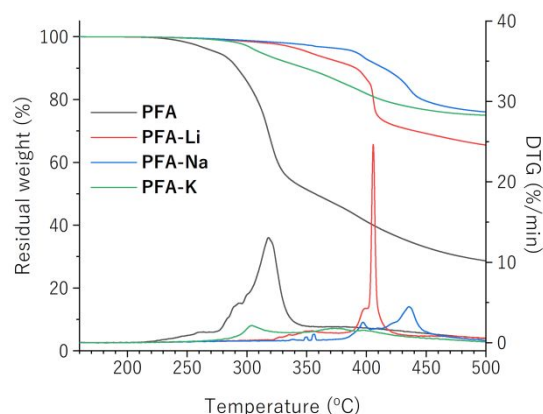


Figure 7. Thermogravimetry analysis profiles of poly(metal fumarate)s.

Table 1. Thermogravimetry analysis of poly(metal fumarate)s.  $T_{deg}$  corresponds to the temperature at 5% weight loss.

	PFA	PFA-Li	PFA-Na	PFA-K
$T_{deg}$ (°C)	285	352	391	308
DTG peak (°C)	318	405	397, 436	304, 375

The transparency of the polymers were evaluated by UV-vis spectroscopy and haze measurement for the polymer films prepared on a quartz glass. All the PFA-M films have almost no absorption band in the visible to UV region. The polymer films of PFA-Li and PFA-K showed high transparency (%transmittance, %T) over 85% throughout the visible region (Figure 8). However, PFA-Na showed lower transparency. Remarkably, the 50% cut-off wavelength of these films was approximately 240 nm, and the films exhibited high transmittance in both the UV-B (320–280 nm) and UV-C (< 280 nm) regions. This is due to the aliphatic polymer chain structure of PFA-salts without aromatic groups and is in sharp contrast to a common transparent heat-resistant film such as polyimide having 50% cut-off at  $\lambda > 350$  nm.

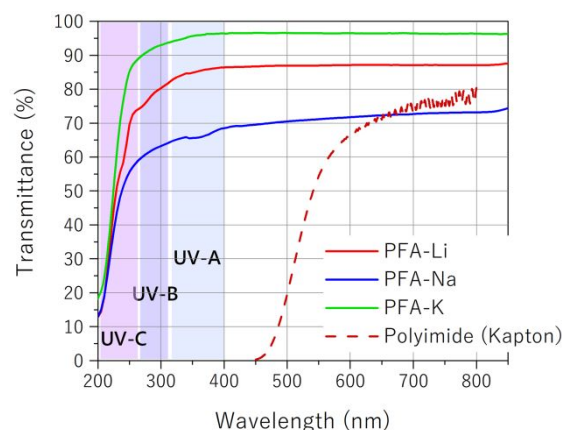


Figure 8. Wavelength-dependent transmittance of poly(metal fumarate)s films analysed by UV-vis spectrometer

Haze measurements of the films confirmed that films of PFA, PFA-Li, and PFA-K have low haze values below 1% (Table S3). The low transmittance of PFA-Li films in UV-vis measurement is due to a distortion of the film surface. Although the total



- 34 H. Noguchi, T. Michinobu, N. Fujii, M. Funahashi, M. Tokita, J. Watanabe and K. Shigehara, *J. Polym. Sci. Part Polym. Chem.*, 2008, **46**, 5101–5114.
- 35 S. Kim, W. Inoue, S. Hirano, R. Yagi, Y. Kuwahara, T. Ogata and S. Kurihara, *Polymer*, 2014, **55**, 871–877.
- 36 T. Kitano, A. Ishigaki, G.-I. Uematsu, S. Kawaguchi and K. Ito, *J. Polym. Sci. Part Polym. Chem.*, 1987, **25**, 979–986.