

COMMUNICATION



Cite this: *Chem. Commun.*, 2018, 54, 7955

Received 10th April 2018,
Accepted 25th June 2018

DOI: 10.1039/c8cc02850h

rsc.li/chemcomm

Prediction and understanding of AIE effect by quantum mechanics-aided machine-learning algorithm†

Jia Qiu,^{‡,abc} Kun Wang,^{‡,cd} Zhouyang Lian,^{‡,b} Xing Yang,^e Wenhui Huang,^{ab}
Anjun Qin,^{ibf} Qian Wang,^g Jie Tian,^{*cd} Benzong Tang^{*fn} and Shuixing Zhang^{ib*ab}

Significant effort has been devoted to the research of aggregation-induced emission (AIE); however, the discovery of new AIE materials is driven mainly by laborious trial-and-error. In this study, taking triphenylamine (TPA)-based luminophores as an example, we propose an efficient machine-learning scheme for predicting AIE-activity based on quantum mechanics.

Organic luminescent materials have been drawing the attention of researchers worldwide because of their unique chemical and optoelectronic properties.^{1,2} In most cases, luminescent materials are used in the solid state (or aggregate state) for technological applications; however, most of them have very low fluorescence quantum yields (QY) in the solid/aggregate state due to the aggregation-caused quenching ACQ effect.³ Over the past few decades, considerable research has been performed to mitigate this problem;^{4–6} for example, by using guest–host-doped emitter systems, blending with transparent polymers, or the introduction of bulky dendrons. However, these methods are tedious and always accompanied by severe side-effects. Fortunately,

aggregation-induced emission (AIE) provides a revolutionary solution to the fluorescence self-quenching problem at high concentrations or in the aggregate state.^{7,8} According to their research, AIE luminophores are almost non-emissive when dissolved in a good organic solvent, whereas intense fluorescence is generated in the aggregate state.^{3,8} Thus far, several possible qualitative mechanisms have been proposed for elucidating the AIE working principle, including restriction of intramolecular motion (RIM), twisted intramolecular charge transfer (TICT), and excited-state intramolecular proton transfer (ESIPT).^{8,9} Based on the existing studies, thousands of AIE luminophores have been synthesized for widespread applications;^{3,8} however, none of the existing theories can predict the AIE effect accurately prior to experimentation, except for the ones with known AIE elements. Therefore, AIE luminophores have been developed mainly by trial and error based on the researchers' experience.

Since the proposal of the Materials Genome Initiative (MGI), Machine Learning (ML) has been displaying powerful ability in computational simulation, modeling, and high-throughput computational screening of materials.^{10–13} On the other hand, traditional theoretical calculation methods like quantum mechanics can yield important molecular structural parameters; however, efficient prediction of the macro-properties of materials is difficult because computational cost must be high to obtain good performance.^{11,12} A combination of both methods may provide a rational solution to these problems. Herein, we attempted to use a ML scheme to predict the AIE activity of luminophores based on their quantum mechanics data, the critical step of which is to find the correlation between AIE activity and molecular structural parameters.

Although various AIE luminophores have been reported, we start with triphenylamine (TPA) derivatives because both AIE-active and AIE-inactive (including ACQ luminophores and those not having any distinct AIE or ACQ effect) TPA derivatives are available, which have a common TPA core that facilitates the following feature extraction. Besides, TPA-based luminophores possess excellent photoelectric properties and find extensive applications due to their special donor–acceptor (D–A) structure.^{14,15}

^a School of Medicine, South China University of Technology, Guangzhou Higher Education Mega Center, Guangzhou 510006, China

^b Department of Radiology, Guangdong General Hospital/Guangdong Academy of Medical Sciences, Guangzhou, 510080, China. E-mail: shui7515@126.com

^c CAS Key Laboratory of Molecular Imaging, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China. E-mail: tian@ieec.org

^d University of Chinese Academy of Sciences, Beijing 100049, China

^e Department of Nuclear Medicine, Peking University First Hospital, Beijing 100034, China

^f State Key Laboratory of Luminescent Materials and Devices, South China University of Technology, Guangzhou 510640, China

^g Department of Diagnostic Imaging, National Cancer Center/Cancer Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, No. 17, Panjiayuan, Chaoyang District, Beijing 100021, China

^h Department of Chemistry, Hong Kong Branch of Chinese National Engineering Research Center for Tissue Restoration and Reconstruction, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, China. E-mail: tangbenz@ust.hk

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c8cc02850h

‡ These authors contributed equally to this work.

Significant research effort has been devoted to TPA-based luminophores, making more data available to researchers. Of the various ML-based methods available, support vector machine (SVM),^{16,17} one of the most popular and powerful techniques for data classification, was employed herein with a radial basis function kernel to develop the ML model.^{18,19}

In past studies, TPA was considered a typical ACQ luminophore.²⁰ However, many TPA derivatives display remarkable AIE effects.^{14,15,21} According to previous research, ACQ components combined with AIE elements (such as tetraphenylethene and hexaphenylsilole) as substituent groups usually yield AIE products;^{3,22} therefore, our study focuses on those without traditional AIE substituent groups. In order to build a proper database for SVM classifiers, we must obtain information on both AIE-active and AIE-inactive compounds. Unfortunately, the AIE-inactive compounds have not been extensively reported due to their poor emission performances. 61 reported TPA-based luminophores including 41 AIE-active ones and 20 AIE-inactive ones (ESI†, Section S1) were obtained according to literature survey. The molecular geometries of all these compounds have been fully optimized using density functional theory (DFT) calculations based on B3LYP/6-31G(d). The procedure for ML model building and the operating principle are illustrated in Fig. 1.

Since TPA is an electron donor, the TICT theory, is a popular mechanism for explaining the AIE phenomenon of TPA derivatives; however, other relevant theories, including RIM and non-planar conformations, have also been reported.^{9,15,21,23,24} Inspired by this, the charge distributions in TPA, AIE-active TPA derivatives, and AIE-inactive TPA derivatives were calculated using the natural bond orbital (NBO) analysis (Table S1, ESI†).²⁵ To simplify the problem, we focus on the common TPA core of these compounds for feature extraction. Interestingly, the charge distribution on the TPA core of AIE-active derivatives

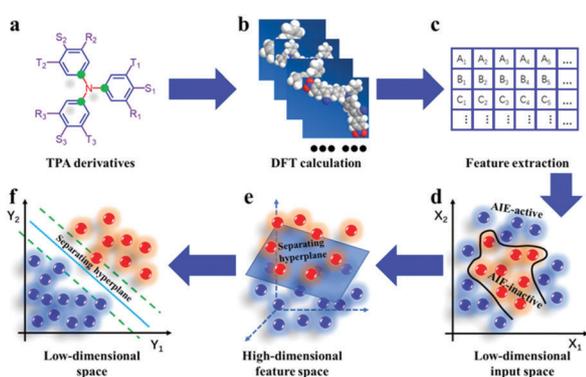


Fig. 1 The procedure for ML model building and the operating principle of the model. (a) General molecular formula of the TPA derivatives (including TPA) in our database, central nitrogen-connected carbons are marked by green circles and substituents are signified by purple characters. (b) Quantum mechanics calculation was performed using DFT. (c) Feature extraction from calculated parameters. (d and e) Distribution of feature points is complex and linearly non-separable in low-dimensional input space, while it is simple and tends to be linearly separable after mapping to high-dimensional feature space. (f) Projection of feature points displayed in (e) on a section perpendicular to the separating hyperplane.

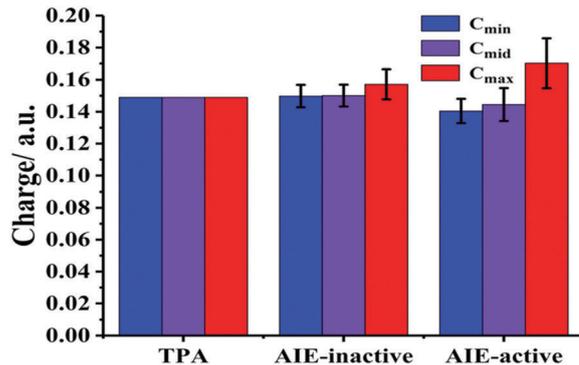


Fig. 2 The charges on the central nitrogen and the three adjacent carbons of TPA, AIE-active TPA derivatives, and AIE-inactive TPA derivatives. For the compounds in our database, the blue, violet, and red bars refer to the mean of minimum, medium, and maximum charges of the nitrogen-connected carbons, respectively.

was significantly different from those on TPA and the AIE-inactive derivatives. In other words, AIE-active derivatives have a more asymmetrical charge distribution, which is particularly obvious for the three carbons adjacent to the central nitrogen (Fig. 2 and Fig. S1, ESI†). Thereby, the charge values of these carbons were collected as the input parameters of the SVM classifier.

To build an SVM model, sufficient data are required. In this study, however, the 61 sets of data obtained are far from enough for conventional verification methods. Hence we applied the leave-one-out cross-validation (LOOCV) technique,²⁶ by which only one molecule was selected as the testing data set while the others were used as training data set each time, and the comprehensive performance of this model was evaluated until all the molecules have been used as testing data and training data by turns for the same model parameters. To ensure that the prediction results are not affected by the sequence of input parameters, the data obtained by exchanging the coordinate positions of the input vectors was also used for training. For example, if (a, b, c) belong to the training data set, then (a, c, b), (b, a, c), ..., (c, b, a) do too.

To evaluate the performance of the model, the AIE-active and AIE-inactive molecules were classified as “Positive class” and “Negative class”, respectively. Therefore, four types of prediction results, *viz.* TP (True Positive), FN (False Negative), FP (False Positive), and TN (True Negative) were obtained. The accuracy, sensitivity and specificity is defined by:

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (1)$$

$$\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN}) \quad (2)$$

$$\text{Specificity} = \text{TN} / (\text{FP} + \text{TN}) \quad (3)$$

respectively, where the TP, TN, FP, and FN represent the numbers of the corresponding prediction results, while sensitivity and specificity indicate the ability to recognize the AIE-active and the AIE-inactive luminophores from relevant classes, respectively. Since we not only need to identify the former from the positive class but also hope to avoid mistaking the

AIE-inactive molecules for AIE-active ones, both sensitivity and specificity are crucial terms, in addition to accuracy. Here, we defined a new parameter $MSS = \min \{\text{sensitivity, specificity}\}$ as an indicator of model performance. The best SVM classifier is selected based on the following criteria: (1) largest MSS value and (2) largest accuracy while there are more than one largest MSS, by which models with large values for accuracy, sensitivity, and specificity simultaneously will be picked out. Thereafter, two SVM parameters, C (trades off misclassification of training examples against simplicity of the decision surface) and γ (defines how much influence a single training example has) were varied to adjust the performance of the classifier.

The model performances for different values of parameters are shown in Fig. 3. The variation trends of sensitivity and specificity are opposite with the changes of C and γ . The optimum model was obtained when $C = 2^{-4}$ and $\gamma = 2^{-1}$ (marked by yellow star in Fig. 3), with accuracy, sensitivity, and specificity values of up to 0.84, 0.80, and 0.90, respectively (Fig. 4a). Such good performance is enough to demonstrate the close relationship between intramolecular charge distribution and AIE effect of the TPA-based luminophores. Relevant model performance was also evaluated using the above-mentioned training method considering the charge on the central nitrogen atom as an input descriptor (Fig. 4a). However, the addition of the nitrogen charge did not enhance the performance of the model. Our explanation is that there is a correlation between the charges on the central nitrogen and the adjacent carbon atoms because they are spatially adjacent and belong to the same conjugated system; this was supported by the linear regression between the charge on the nitrogen and the sum of charges on the three carbons (Fig. 4b). However, if only the charge on the central nitrogen was used as the input descriptor, the model performance will be significantly reduced because the input information is insufficient for finding the difference between the two types of molecules (Fig. 4a).

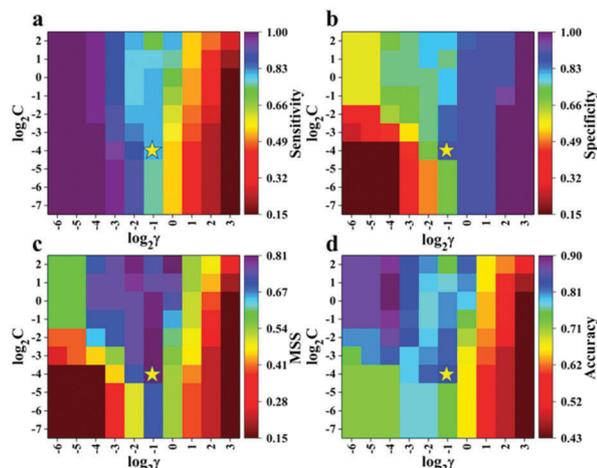


Fig. 3 Performance of the SVM classifier for different training parameters. (a–d) Sensitivity, specificity, MSS, and accuracy values of the models trained with different parameter. The parameters corresponding to best model are marked with yellow star.

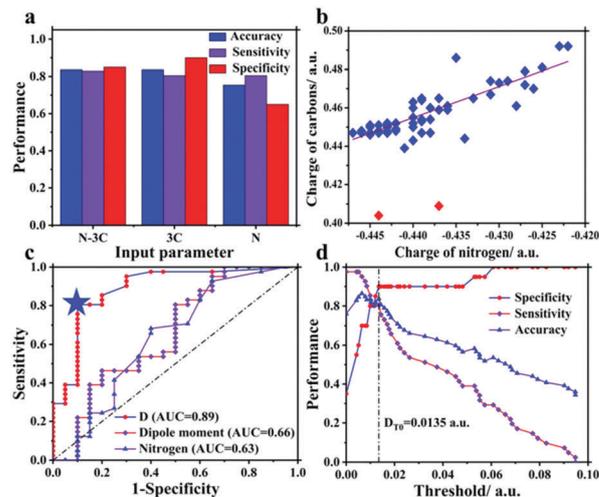


Fig. 4 Performance comparison between classifiers trained by different types of parameters, the ROC curves of different physical quantities and classification performances of D . (a) Classifiers trained by N-3C and 3C show similar performance, while the single N yields the worst result. (b) Linear regression analysis between the charge on the central nitrogen and the sum of the three adjacent carbons' charges. The red points were not included in the calculation due to their large deviation. (c) D yields the largest AUC, while the dipole moment of the entire molecule and the charge on the nitrogen show poor classification ability. The critical point corresponding to the best threshold for parameter D is marked by a blue star. (d) The classifying qualities of parameter D at different threshold levels; the dot-dash line refers to the best threshold.

According to the research on ML schemes, the good prediction ability has demonstrated the close correlation between charge distribution on the TPA core and the AIE effect, which is consistent with the TICT theory. The rough statistical analysis (Fig. 2) also indicated that AIE-active luminophores have a more asymmetrical charge distribution on the TPA core than the AIE-inactive ones. To further demonstrate the importance of the charge distribution on TPA core to the AIE effect of TPA-based luminophores, we have attempted to figure out the relationship between the asymmetry of charge distribution and the AIE activity by semi-quantitative analysis.

Since studying the entire molecule is difficult, we still focus on the three carbons adjacent to the central nitrogen, which have been demonstrated significant by the above research. Here, on the analogy of dipole moment, we deduced a parameter “ D ” with the same physical dimensions as atomic charge by the following equation:

$$D = \sqrt{(E_1 - E_2)^2 + (E_2 - E_3)^2 + (E_3 - E_1)^2} \quad (4)$$

where E_1 , E_2 , and E_3 are the charges on the three carbons adjacent to the central nitrogen of the TPA core. Detailed derivation can be found in ESI,† Section S2. For eqn (4), if the charge distribution on the TPA core was completely symmetric, *i.e.* $E_1 = E_2 = E_3$, then $D = 0$; otherwise, $D > 0$, which will increase with increasing asymmetry of charge distribution. Hence the parameter D can be used to describe the asymmetry of charge distribution, or the dipole moment, of the TPA core. Thereafter,

the parameter D was used to classify the two types of luminophores at a specific threshold. The receiver operating characteristic (ROC) curve,²⁷ was plotted to study the classification effect at different threshold D_T values (Fig. 4c).

Fig. 4c shows that the ROC curve obtained by parameter D has a large area under curve (AUC) value of 0.89, indicating its efficiency in evaluating the AIE activity. The point on the curve nearest to the upper-left corner (*i.e.* the (0,1) coordinate, indicating maximum sensitivity and specificity) is selected as critical point, with the corresponding threshold $D_{T0} = 0.0135$ a.u. serving as the best threshold (marked by blue star in Fig. 4c). The classifying qualities of the parameter D at different threshold values are shown in Fig. 4d. Hence, for a specific molecule, if $D > 0.0135$ a.u., it can be considered AIE-active, otherwise inactive. Incredibly, thus-obtained results is the same as that obtained by the SVM classifier (Table S1, ESI[†]), further confirming the veracity of this model and demonstrating that large enough dipole moment of the TPA core is the key to activate the AIE effect of TPA derivatives. In addition to the parameter D , the dipole moment of the entire molecule and the charge on the central nitrogen were similarly used for classification; however, the corresponding ROC curves have much smaller AUC values, indicating poor classification ability (Fig. 4c). Consequently, these two physical variables, especially the dipole moment of the entire molecule (representing molecular polarity), is not the key determining factor for AIE effect. An interesting phenomenon is that most symmetrically tri-substituted TPA derivatives with distinct D-A structures do not show AIE activity, while their mono-substituted and (or) disubstituted analogues display obvious AIE-effects.^{28–30} An unsymmetrical polysubstituted TPA-based luminophore with five strong electron-withdrawing substituents (including four hydroxyls and a cyano group), reported by Wu *et al.*, is also AIE-inactive,³¹ with D value (0.004 a.u.) much less than D_{T0} . Based on these phenomena, we give two interpretations here. (1) Large enough dipole moment of TPA core is the key to activate the AIE effect of TPA-based luminophores, and TICT is the essential mechanism. The non-zero D_{T0} value means that the AIE effect can be activated only when the dipole moment of the TPA core exceeds a certain value, sufficient enough to produce distinct TICT effect. (2) Symmetrical substitution results in symmetrical charge distribution therefore, AIE effect cannot be activated. Finally, using the SVM model, we predicted an AIE-active TPA derivative 4-diphenylaminobenzaldehyde (DPAB), and the AIE effect was verified by experiment (Section S3, ESI[†]).

In summary, we developed a ML scheme with excellent performance based on TPA-based luminophores, for predicting the AIE activity *via* a combination of quantum mechanics and SVM classifier. In this study, our models showed good prediction ability, explained the key factor required to activate the AIE effect of TPA-based luminophores, and showed that TICT is the operating mechanism. Remarkably, we used only the charge on the three carbons, though there are tens to hundreds of atoms in each complex molecule. This indicates a strong correlation among different parts of the entire molecule because of the large conjugated system.

We acknowledge financial support from the National Natural Science Foundation of China (81571664, 61671449); the National Key

Research and Development Program of China (2017YFA0205200, 2016YFA0201401); the Science and Technology Planning Project of Guangdong Province (2014A020212244, 2016A020216020); the Scientific Research General Project of Guangzhou Science Technology and Innovation Commission (201707010328); and the China Postdoctoral Science Foundation (2016M600145).

Conflicts of interest

There are no conflicts to declare.

Notes and references

- 1 J. Li, D. Yim, W. Jang and J. Yoon, *Chem. Soc. Rev.*, 2017, **46**, 2437–2458.
- 2 J. Liu, W. Bu and J. Shi, *Chem. Rev.*, 2017, **117**, 6160–6224.
- 3 J. Mei, N. L. C. Leung, R. T. K. Kwok, J. W. Y. Lam and B. Z. Tang, *Chem. Rev.*, 2015, **115**, 11718–11940.
- 4 S. Hecht and J. Frechet, *Angew. Chem., Int. Ed.*, 2001, **40**, 74–91.
- 5 M. T. Lee, H. H. Chen, C. H. Liao, C. H. Tsai and C. H. Chen, *Appl. Phys. Lett.*, 2004, **85**, 3301–3303.
- 6 S. F. Lim, R. H. Friend, I. D. Rees, J. Li, Y. G. Ma, K. Robinson, A. B. Holmes, E. Hennebicq, D. Beljonne and F. Cacialli, *Adv. Funct. Mater.*, 2005, **15**, 981–988.
- 7 J. D. Luo, Z. L. Xie, J. Lam, L. Cheng, H. Y. Chen, C. F. Qiu, H. S. Kwok, X. W. Zhan, Y. Q. Liu, D. B. Zhu and B. Z. Tang, *Chem. Commun.*, 2001, 1740–1741.
- 8 G. Feng, R. T. K. Kwok, B. Z. Tang and B. Liu, *Appl. Phys. Rev.*, 2017, **4**, 021307.
- 9 J. Mei, Y. Hong, J. W. Y. Lam, A. Qin, Y. Tang and B. Z. Tang, *Adv. Mater.*, 2014, **26**, 5429–5479.
- 10 P. Raccuglia, K. C. Elbert, P. D. F. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier and A. J. Norquist, *Nature*, 2016, **533**, 73–76.
- 11 M. Fernandez, P. G. Boyd, T. D. Daff, M. Z. Aghaji and T. K. Woo, *J. Phys. Chem. Lett.*, 2014, **5**, 3056–3060.
- 12 J. P. Janet and H. J. Kulik, *Chem. Sci.*, 2017, **8**, 5137–5152.
- 13 O. Isayev, C. Oses, C. Toher, E. Gossett, S. Curtarolo and A. Tropsha, *Nat. Commun.*, 2017, **8**, 15679.
- 14 X. Han, Q. Bai, L. Yao, H. Liu, Y. Gao, J. Li, L. Liu, Y. Liu, X. Li, P. Lu and B. Yang, *Adv. Funct. Mater.*, 2015, **25**, 7521–7529.
- 15 T. Liu, L. Zhu, C. Zhong, G. Xie, S. Gong, J. Fang, D. Ma and C. Yang, *Adv. Funct. Mater.*, 2017, **27**, 1606384.
- 16 C. Cortes and V. Vapnik, *Mach. Learn.*, 1995, **20**, 273–297.
- 17 Y. T. Sun, H. Y. Bai, M. Z. Li and W. H. Wang, *J. Phys. Chem. Lett.*, 2017, **8**, 3434–3439.
- 18 V. N. Hien and F. Porikli, *IEEE Trans. Pattern Anal.*, 2013, **35**, 970–982.
- 19 C. Chang and S. Chou, *Pattern Recogn.*, 2015, **48**, 3983–3992.
- 20 W. Z. Yuan, P. Lu, S. Chen, J. W. Y. Lam, Z. Wang, Y. Liu, H. S. Kwok, Y. Ma and B. Z. Tang, *Adv. Mater.*, 2010, **22**, 2159–2163.
- 21 M. Gao, H. Su, Y. Lin, X. Ling, S. Li, A. Qin and B. Z. Tang, *Chem. Sci.*, 2017, **8**, 1763–1768.
- 22 W. Qin, D. Ding, J. Liu, W. Z. Yuan, Y. Hu, B. Liu and B. Z. Tang, *Adv. Funct. Mater.*, 2012, **22**, 771–779.
- 23 Z. Ning, Z. Chen, Q. Zhang, Y. Yan, S. Qian, Y. Cao and H. Tian, *Adv. Funct. Mater.*, 2007, **17**, 3799–3807.
- 24 Y. Liu, M. Kong, Q. Zhang, Z. Zhang, H. Zhou, S. Zhang, S. Li, J. Wu and Y. Tian, *J. Mater. Chem. B*, 2014, **2**, 5430–5440.
- 25 B. G. Caulkins, R. P. Young, R. A. Kudla, C. Yang, T. J. Bittbauer, B. Bastin, E. Hilario, L. Fan, M. J. Marsella, M. F. Dunn and L. J. Mueller, *J. Am. Chem. Soc.*, 2016, **138**, 15214–15226.
- 26 J. Wan, G. Guo and S. Z. Li, *IEEE Trans. Pattern Anal.*, 2016, **38**, 1626–1639.
- 27 A. P. Bradley, *Pattern Recogn.*, 1997, **30**, 1145–1159.
- 28 X. Zhang, X. Gan, S. Yao, W. Zhu, J. Yu, Z. Wu, H. Zhou, Y. Tian and J. Wu, *RSC Adv.*, 2016, **6**, 60022–60028.
- 29 C. Wang, S. Yan, Y. Chen, Y. Zhou, C. Zhong, P. Guo, R. Huang, X. Weng and X. Zhou, *Chin. Chem. Lett.*, 2015, **26**, 323–328.
- 30 Z. Liang, X. Wang, G. Dai, C. Ye, Y. Zhou and X. Tao, *New J. Chem.*, 2015, **39**, 8874–8880.
- 31 J. Wu, W. Chen and G. Liou, *Polym. Chem.*, 2016, **7**, 1569–1576.