

Machine learning analysis for photo-induced phenomena of 1-methyl-3-(*N*-(1,8-naphthalimidyl)ethyl)imidazolium salts

Takehiro Fujita,^{*1,2} Kei Terayama,^{3,4,5} Masato Sumita,^{*3} Hironori Izawa,⁶ Koji Tsuda,^{3,7,8} and Kenji Morihashi²

¹ Advanced Technology Research Laboratories, Analytical Technology Center, Idemitsu Kosan Co., Ltd., 1280 Kami-izumi, Sodegaura, Chiba, 299-0293, Japan

² Department of Chemistry, University of Tsukuba, 1-1-1 Tennoudai, Tsukuba, Ibaraki 305-8571, Japan

³ Advanced Intelligence Project, RIKEN, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan.

⁴ Medical Sciences Innovation Hub Program, RIKEN, 1-7-22 Suehiro-cho, Tsurumi-ku, Kanagawa 230-0045, Japan.

⁵ Graduate School of Medicine, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan.

⁶ Department of Chemistry and Biotechnology, Tottori University, 4-101 Koyama-Minami, Tottori, 680-8550, Japan.

⁷ Department of Computational Biology and Medical Sciences, University of Tokyo, 5-1-5, Kashiwa-no-ha, Kashiwa, Chiba 277-8561, Japan.

⁸ Materials Research by Information Integration Initiative, NIMS, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan.

Abstract

1-Methyl-3-(*N*-(1,8-naphthalimidyl)ethyl)imidazolium (MNEI) salts show various photochemical phenomena depending on the guest anion. We used machine learning techniques to extract decisive factors that dominate the fate of MNEI salts after UV irradiation and prepared a decision tree diagram to deduce their photochemical phenomena before experiments.

Keywords: anion sensor, decision tree, density functional theory

1. Introduction

1-Methyl-3-(*N*-(1,8-naphthalimidyl)ethyl)imidazolium¹ (MNEI; Fig. 1A), a 1,8-naphthalimide derivative, is designed as a versatile molecular sensor that exhibits a high affinity for anions due to the positive charge on the imidazolium group, allowing the detection of the electronegativity of a guest anion by fluorescence intensity.¹⁻³ While showing a potential as the fluorescent molecular sensor, MNEI shows yellow coloration with anionic polysaccharides upon irradiation.⁴ Although MNEI can interact with various anions, its photo-induced phenomena are different depending on the guest anions. Hence, to use MNEI as a molecular sensor, we need to elucidate the applicable limit of MNEI, that is, the factors that cause the MNEI–anion complex to show fluorescence or coloration.

Conventional experimental and computational approaches show that complex chemical phenomena, such as non-adiabatic transition and excimer formation, are involved in the photo-induced process of MNEI complexes.¹⁻⁴ Hence, much time is required to detect and identify the physical characteristics that allow MNEI to act as an anion sensor. Furthermore, because the number of experimentally available MNEI salt is only a degree of countable, predicting the properties of unidentified chemical compounds is impossible. Therefore, machine learning techniques that require considerable data may not be applicable.

Recently, the application of machine learning techniques has been attracting considerable attention in the field of chemistry. Representative research includes physical property

prediction, chemical reactivities, and design of de novo compounds⁵⁻⁹, which require a large amount of training data. On the other hands, one of authors reported a practical method to estimate oxygen-K edges of the ELNES/XANES spectra of oxide compounds with a small amount of training data¹⁰.

In this study, we employed machine learning techniques to extract the factors that dominate the photo-induced phenomena to MNEI complexes and made the estimation model to deduce the phenomena of MNEI complexes.

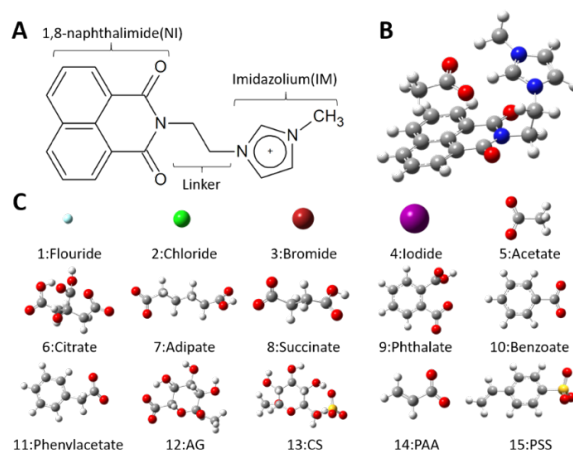


Figure 1. Chemical structure of MNEI (A), input model structures of MNEI–acetate (B) and anions (C). The anions have a net charge of -1. The difference in the ionic valence is considered as the feature value. AG (alginic acid), CS (chondroitin sulfate), PAA (polyacrylate), and PSS (polystyrene sulfonate) are the simplified unit structures of the anionic polymer. To distinguish these polymer species, we used a binary feature value.

2. Experimental and computational details

Fluorescence spectra of 1.50 mM MNEI-Cl aqueous solutions containing 0.5 wt% anionic molecules were recorded on an RF-6000 (Shimadzu, Kyoto, Japan). UV-Vis spectra were recorded on a Multiskan GO (Thermo Fischer Scientific, Waltham, MA). 1.50 mM MNEI aqueous solution containing 0.5 wt% anionic molecules placed in a quartz cell (path length: 1 mm) was UV irradiated by an SP-7 spot cure (Ushio, Tokyo) equipped with a deep UV lamp (main wavelength: 365 nm) at room temperature. The length between the quartz cell and the UV lamp was 19.2 cm. After irradiation, the irradiated solution was immediately subjected to UV-Vis measurement.

Computational data are prepared through the density functional theory¹¹ (DFT) by using GAUSSIAN 09¹². The effect of aqueous solution is included by IEFPCM¹³. The extended hybrid functional combined with Lee-Yang-Parr parameters (X3LYP¹⁴) were used with 6-31G* basis set except for Red-Ox potential calculations and iodine atom. The Red-Ox potentials were calculated at the X3LYP/6-31++G** level. For the iodine atom, DGDZVP¹⁵ basis set was adopted. In single point energy calculations of MNEI salt, the basis set superposition error (BSSE¹⁶) is corrected by counterpoise (CP¹⁷) method.

3. Results and Discussion

We employed applicable machine learning techniques even for a small amount data to extract the physical values that dominate the photo-induced phenomena of MNEI salts from computationally prepared data in contrast to recent applications in which a large amount of data was needed. Based on the extracted values, we prepared a decision tree diagram to predict the photo-induced phenomena of MNEI salts (and vice versa, i.e., to predict the physical properties of unidentified anions). To validate this diagram, we compared the deduced phenomena with the experimentally ones for two MNEI salts whose phenomena had not been observed.

We prepared 15 species (four halides, four anionic polymers, and seven carboxylates) whose experimental

phenomena were already known as a training data set. To prepare the computational data, we performed quantum chemistry calculations based on DFT. The structures of a representative MNEI salt and anions are shown in Fig. 1B and 1C, respectively.

We used physical values that were computed or obtained from a database¹⁸ as the feature values (explanatory variables). The integrated values of the experimental fluorescence (integration section: 370–550 nm) and UV-vis spectra (integration section: 390–500 nm) were used as scores (objective variables). All feature values and scores are tabulated in the Supporting information. Machine learning analyses were performed using these feature values and scores through the following procedures. (1) We performed the feature selection by lasso regression¹⁹ with “leave-one-out” cross-validation. The features that minimized the root mean squared error (RMSE) between the experimental scores and the regression results were selected. (2) We classified 15 anions into five clusters using the hierarchical cluster analysis²⁰ to divide the scores into tree-shaped clusters called a “dendrogram.” (3) We prepared a “decision tree²¹” diagram to estimate the photo-induced phenomena of unspecified anions based on the selected features and the generated clusters. We utilized the scikit-learn²² library to implement the machine learning algorithms.

The selected features and their contribution ratios through lasso regression are shown in Fig. 2. For the regression of the fluorescence intensity, the RMSE was 0.215 (Fig. 2A). The interaction energy between MNEI and the anion had the largest contribution ratio of 37% to the fluorescence intensity followed by pK_a , the ionic valence, and Mulliken charge of an anion (Fig. 2B). Thus, we can quantitatively guess the anion that strongly interacts with MNEI to induce fluorescence emission. This result also agrees with the suggestion that MNEI could be a sensor to detect anions whose electronegativities are high.³ For the regression of the coloration intensity, the RMSE was 0.283 (Fig. 2C). Although this error is slightly large, the differentiation between the colored group and the uncolored group is relatively clear. A conjugation extension had the largest contribution ratio of 35%. The extended π -conjugated system tends to quench the coloration reaction (see the Supporting information for more details). Additionally, the molecular weight and pK_a had the same contribution ratio of 28%. A heavy anion with a weak acidity induced the coloration. This means that molecular bulkiness induces coloration because the correlation coefficient between the molecular volume and the molecular weight was 0.940. Therefore, we guess the features reflect a non-electrostatic interaction, such as a CH- π interaction and London dispersion force. Consequently, the feature selection suggests that the magnitude of electrostatic correlation contributes to the fluorescence intensity and that of the non-electrostatic interaction contributes to the yellow coloration.

Figure 3 shows the dendrogram obtained based on the data from the experimental scores. The vertical axis represents the similarity between experimental photochemical behaviors. Anions that are clustered with lower values have higher similarity. Upon setting the depth to 0.50 as the classification

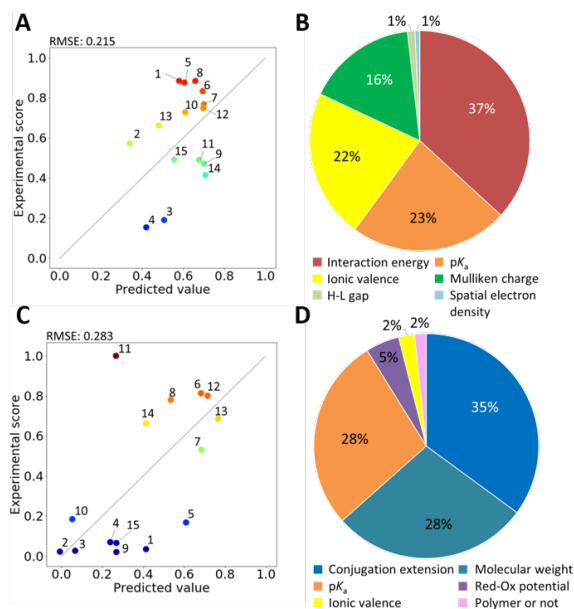


Figure 2. Results of lasso regression, selected features, and their contribution ratios for fluorescence (A, B) and coloration (C, D). The numbers in A and C correspond to Fig. 1C. The vertical axis of A and C is the experimental score, and the horizontal axis is the regression value. The experimental and regression values are equivalent on the dotted line.

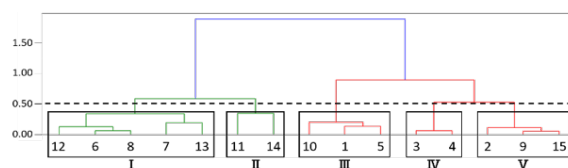


Figure 3. Dendrogram by hierarchical cluster analysis. The numbers correspond to Fig. 1C. Anions that are clustered with lower values having a higher similarity.

threshold, five clusters (shown in Table 1) were obtained. These clusters were used as labels in the decision tree diagram. Anions belonging to Cluster I are multivalent ions and bulky. Anions in Cluster II are also bulky but have small electrostatic interactions. Anions in Cluster III have a large interaction energy. Surprisingly, the fluorine anion was classified into Cluster III along with the carboxylate. This is due to its high fluorescence intensity caused by strong electrostatic interactions. Although both anions in Cluster IV and V are common from the viewpoint of their strong acidity and small interaction energy, they are classified into different clusters because of their low numerical similarity.

The decision tree diagram based on the selected features is shown in Fig. 4. This is a top-down type of estimation flow chart used to deduce the photochemical phenomena of the MNEI-anion complex. If a condition in a square is true, the target goes along the direction that the outlined arrow indicates. If the condition is false, the target goes along the direction that the black arrow indicates. Any features used in the model can be easily obtained through simple quantum chemical calculations and the literature database¹. Furthermore, when interpreting this diagram as a bottom-up type, we can evaluate the properties of unidentified anions by measuring the fluorescence/UV-vis spectra.

To ascertain the validity of the decision tree, two species of anions were prepared whose photo-induced phenomena as MNEI salts had not been observed, namely, trifluoroacetate (TFA; **A**) and laurate (**B**); see Fig. 5. TFA has a structure similar to acetate; however, its acidity and electrostatic properties are very different from those of acetate. Laurate is a kind of higher bulky carboxylate. We predicted the phenomena of MNEI with these anions using the decision tree. The feature values were calculated in the same way as the training data (see the Supporting information also). According to the decision tree and using the feature values in Table 2, TFA was classified into Cluster III, and laurate was classified into Cluster II. Therefore, it was predicted that MNEI-TFA emits an intense fluorescence without coloration and MNEI-laurate emits a weak fluorescence with coloration.

To confirm the estimation results, fluorescence/UV-vis spectra of both MNEI salts were measured. Figure 6 shows the fluorescence (**A**) and UV-vis (**B**) spectra of these MNEI salts. A comparison with experimental results of the estimated phenomena is summarized in Table 3. MNEI-TFA showed intense fluorescence, and there was no absorption peak at

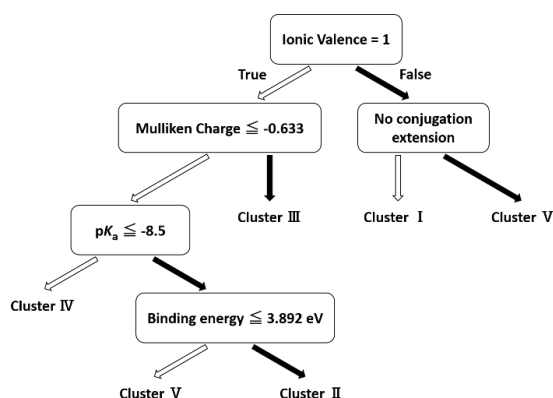


Figure 4. Decision tree diagram based on selected features. If the condition in the square is true, the target branches to the “True” side (white arrow). If the condition is false, the target branches to the “False” side (black arrow)

approximately 430 nm, resulting in a yellow coloration. Therefore, TFA was classified into Cluster III, which agrees with the estimation result. On the other hand, the coloration of MNEI-laurate was consistent with the estimated result, and the fluorescence intensity was moderate, which was weaker than TFA. Based on the experimental score, MNEI-laurate was numerically classified into Cluster I, which is different from the estimation in agreement with the decision tree diagram. This error occurred because all anions in Cluster I are multivalent ions. However, this can be improved by adding laurate and other higher bulky carboxylate into the training data. Likewise, we will investigate anions not included in the current data, such as polycyclic aromatic carboxylates and inorganic acids, and we will robustly improve the decision tree diagram.

Table 1. Characteristics of each cluster and anions belonging to the clusters

Cluster I :	Strong fluorescence, Yellow coloration
	Citrate, Adipate, Succinate, AG, CS
Cluster II :	Weak fluorescence, Yellow coloration
	Phenylacetate, PAA
Cluster III :	Strong fluorescence, No coloration
	Fluoride, Acetate, Benzoate
Cluster IV :	Faint fluorescence, No coloration
	Bromide, Iodide
Cluster V :	Weak fluorescence, No coloration
	Chloride, Phthalate, PSS

Table 2. Features of TFA and laurate used in the decision tree

Feature	TFA	Laurate
Ionic valence	1	1
Mulliken charge	-0.610	-0.647
Conjugate extension	0	0
pKa	0.23	5.30
Interaction energy (eV)	3.805	4.189

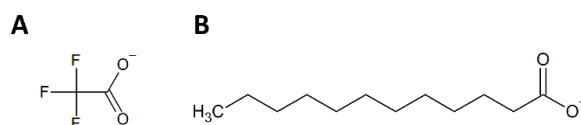


Figure 5. Chemical structures of trifluoroacetate (**A**) and laurate (**B**).

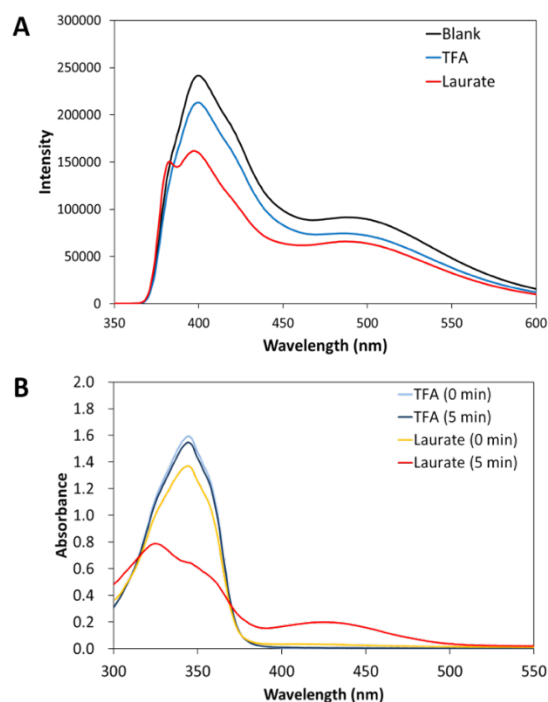


Figure 6. Fluorescence spectra excited at 380 nm in 1.5 mM MNEI-TFA and MNEI-laurate aqueous solutions (A). UV-vis spectra in 1.5 mM MNEI-TFA and MNEI-laurate aqueous solutions (B). The UV-vis spectra were measured immediately after (0 min) and five minutes after (5 min) UV irradiation.

Table 3. Characteristics deduced (Ded.) using the decision tree diagram based on the features tabulated in Table 2 and the experimental (Exp.) result

	Fluorescence		Coloration	
	Ded.	Exp.	Ded.	Exp.
TFA	Strong	Strong	No	No
Laurate	Weak	Strong	Yes	Yes

4. Conclusion

In this study, we successfully demonstrated that a machine learning approach can extract decisive factors that dominate the photo-induced phenomena of MNEI salts. The magnitude of interaction energy between MNEI and a guest anion contributes to the fluorescence emission. The magnitude of the non-electrostatic interaction, which is characterized by the molecular weight and the molecular volume, contributes to the yellow coloration. Therefore, MNEI and its derivatives have potential as an anion sensor that distinguishes electrostatic properties by fluorescence and non-electrostatic properties by coloration. We also prepared a decision tree diagram to deduce the photo-induced phenomena of the MNEI salts. The feature values used in the diagram can be obtained from simple theoretical calculations and literature data. Using the decision tree diagram, we can easily evaluate the properties of unidentified anions by measuring the fluorescence/UV-vis spectra. During the validation of the decision tree diagram, we succeeded in estimating the phenomena of MNEI-TFA. Although there was an error for the laurate system, its accuracy could be improved

with an increased amount of training data. In addition, the spectrum estimation method of Kiyohara *et al.*¹⁰ can be applied to interpret any type of spectrum. Therefore, we expect to increase the accuracy of estimation of the fluorescence/UV-vis spectra of the MNEI salt with increasing the amount of training data. Simultaneously, we will approach the cause of the photochemical phenomena and expand the application range of MNEI.

Acknowledgement

T. F. and M. S. thank the advice from Dr. Hou Zhufeng. This work was supported in part by KAKENHI 17H03034.

References

- H. Izawa, S. Nishino, M. Sumita, M. Akamatsu, K. Morihashi, S. Ifuku, M. Morimoto and H. Saimoto, *Chem. Commun.*, 2015, **51**, 8596.
- T. Otsuka, M. Sumita, H. Izawa and K. Morihashi, *Phys. Chem. Chem. Phys.*, 2016, **18**, 17795.
- T. Otsuka, M. Sumita, H. Izawa and K. Morihashi, *Phys. Chem. Chem. Phys.*, 2018, **20**, 3911.
- H. Izawa, M. Wada, S. Nishino, M. Sumita, T. Fujita, K. Morihashi, S. Ifuku, M. Morimoto and H. Saimoto, *Bull. Chem. Soc. Jpn.*, 2018, **91**, 1220.
- L. Ward, A. Agrawal, A. Choudhary and C. Wolverton, *npj Comput. Mater.*, 2016, **2**, 16028.
- M. Sumita, X. Yang, S. Ishihara, R. Tamura and K. Tsuda, *ACS Cent. Sci.*, 2018 **4**, 1126.
- M. H. S. Segler, T. Kogej, C. Tyrchan and M. P. Waller, *ACS Cent. Sci.*, 2018 **4**, 120.
- N. Mounet, M. Gibertini, P. Schwaller, D. Campi, A. Merkys, A. Marrazzo, T. Sohler, I. E. Castelli, A. Cepellotti, G. Pizzi and N. Marzari, *Nature Nanotech.*, 2018, **13**, 246.
- S. Steiner, *et al.*, *Science*, 2019, **363**, 11.
- S. Kiyohara, T. Miyata, K. Tsuda and T. Mizoguchi, *Sci. Rep.*, 2018, **8**, 13548.
- P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, B864.
- GAUSSIAN 09, Revision D. 01, M. J. Frisch, *et al.*, Gaussian Inc. Wallingford CT, 2009.
- B. Mennucci, *Adv. Rev.*, 2012, **2**, 386.
- X. Xu, W. A. Goddard, *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 2673.
- N. Godbout, D. R. Salahub, J. Andzelm and E. Wimmer, *Can. J. Chem.*, 1992, **70**, 560.
- S. F. Boys, F. Bernardi, *Mol. Phys.*, 1970, **19**, 553.
- S. Simon, M. Duran, *J. Chem. Phys.*, 1996, **105**, 11024.
- pKa Data compiled by R. Williams, https://www.chem.wisc.edu/areas/reich/pkatable/pKa_compilation-1-Williams.pdf, (accessed April 2018).
- R. Tibshirani, *J. R. Stat. Soc. Series B*, 1996, **58**, 267.
- S. C. Johnson, *Psychometrika*, 1967, **32**, 241.
- J. R. Quinlan, *Machine Learning*, 1986, **1**, 81.
- F. Pedregosa *et al.*, *J. Mach. Learn. Res.*, 2011, **12**, 28