Sensors & Actuators: B. Chemical 329 (2021) 129248

Contents lists available at ScienceDirect

Sensors and Actuators: B. Chemical

journal homepage: www.elsevier.com/locate/snb

Kaleidoscopic fluorescent arrays for machine-learning-based point-of-care chemical sensing

Hyungi Kim [a,1], Sang-Kee Choi [a,1], Jungmo Ahn [b,1], Hojeong Yu [c], Kyoungha Min [c,d], Changgi Hong [e], Ik-Soo Shin [f], Sanghee Lee [g], Hakho Lee [c,h], Hyungsoon Im [c,h], JeongGil Ko [i], Eunha Kim [a,e,*]

[a] Department of Molecular Science and Technology, Ajou University, Suwon, 16499, South Korea
[b] Department of Computer Engineering, Ajou University, Suwon, 16499, South Korea
[c] Center for Systems Biology, Massachusetts General Hospital, Boston, MA, 02114, USA
[d] Department of Electrical Engineering and Computer Science, University of California Berkeley, Berkeley, CA, 94720, USA
[e] Department of Applied Chemistry and Biological Engineering, Ajou University, Suwon, 16499, South Korea
[f] Department of Chemistry, Soongsil University, Seoul, 07027, South Korea
[g] Center for Neuro-Medicine, Brain Science Institute, Korea Institute of Science and Technology, Seoul, 02792, South Korea
[h] Department of Radiology, Massachusetts General Hospital, Boston, MA, 02114, USA
[i] School of Integrated Technology, Yonsei University, Incheon, 21983, South Korea

ARTICLE INFO

Keywords:
Indolizine
Fluorescent compound array
Pattern recognition
Machine learning
Multiplexing

ABSTRACT

Multiplexed analysis allows simultaneous measurements of multiple targets, improving the detection sensitivity and accuracy. However, highly multiplexed analysis has been challenging for point-of-care (POC) sensing, which requires a simple, portable, robust, and affordable detection system. In this work, we developed paper-based POC sensing arrays consisting of kaleidoscopic fluorescent compounds. Using an indolizine structure as a fluorescent core skeleton, named Kaleidolizine (KIz), a library of 75 different fluorescent KIz derivatives were designed and synthesized. These KIz derivatives are simultaneously excited by a single ultraviolet (UV) light source and emit diverse fluorescence colors and intensities. For multiplexed POC sensing system, fluorescent compounds array on cellulose paper was prepared and the pattern of fluorescence changes of KIz on array were specific to target chemicals adsorbed on that paper. Furthermore, we developed a machine-learning algorithm for automated, rapid analysis of color and intensity changes of individual sensing arrays. We showed that the paper sensor arrays could differentiate 35 different volatile organic compounds using a smartphone-based handheld detection system. Powered by the custom-developed machine-learning algorithm, we achieved the detection accuracy of 97 % in the VOC detection. The highly multiplexed paper sensor could have favorable applications for monitoring a broad-range of environmental toxins, heavy metals, explosives, pathogens.

1. Introduction

There is a strong demand to develop a rapid, sensitive, portable, and affordable detection system for highly-multiplexed point-of-care (POC) chemical sensing [1]. A POC system that can detect subtle changes in a chemical environment will have broad applications in pathogen detection [2], volatile organic compound classification [3], explosive screening [4], and disease diagnosis [5,6]. However, it is currently challenging to detect multiple targets and perform multiplexed analysis, especially in the field settings where a large amount of non-target molecules could generate false-signals.

One of the most robust systems for highly multiplexed detection in the field is human sensory systems. For example, a human nose has nearly 400 active olfactory receptors [7]. The array of receptors detects chemical, molecular, and environmental changes [8]. Using the collected data, the brain could recognize, and discriminate thousands of odorants present in the environment. Key components underlying the robust multiplexed detection are the use of multiple olfactory receptors

* Corresponding author at: School of Integrated Technology, Yonsei University, Incheon, 21983, South Korea; Center for Systems Biology, Massachusetts General Hospital, Boston, MA, 02114, USA.
E-mail address: ehkim01@ajou.ac.kr (E. Kim).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.snb.2020.129248
Received 3 September 2020; Received in revised form 15 November 2020; Accepted 19 November 2020
Available online 1 December 2020
0925-4005/© 2020 Elsevier B.V. All rights reserved.

that generate complex signals from the sensor arrays and the human brain that analyzes and recognizes subtle changes from the sensory data [9]. Learning from nature, we aim to design sensing system composed of sensor arrays that independently interact and capture different molecular characteristics and machine-learning-based artificial intelligence for accurate discrimination of various chemicals.

Among various sensing systems, fluorometric sensors have been widely used in various chemical sensing applications because of their high sensitivity and broad applicability in different settings [10–14]. The fundamental understanding of fluorophores photophysical properties related to their structures [15–19] and the development of supramolecular chemistry synergistically have led to the discovery of various sensitive fluorescent sensors [12]. The fluorescent sensors are usually generated by incorporating a chelator or functional moiety into fluorophores. Intermolecular interactions between target molecules and fluorophores alter photophysical properties of fluorophore sensors; the changes are used molecular sensing [13,20]. Despite the substantial progress in the field, it is still challenging to develop highly specific and sensitive fluorescent sensor arrays with highly multiplexed sensing capabilities for multiple targets. For highly multiplexed analysis in broad applications, it is important to develop vast fluorometric sensor arrays that interact with multiple target analytes [21–23]. The fact that one molecular sensor is not specific only to a specific analyte imposes that a combination of signals from several molecular sensors can lead to unique differences for each sample allowing to identity them. Recent studies demonstrated that molecular array system could be applied to discriminate the types of bacteria [24], amino acids [25], mammalian cancer cells from normal cells and metastatic cells [26] in controlled conditions. In addition, more importantly, pioneering recent studies [1, 6] demonstrate that the paper-based POC testing systems powered by machine learning for providing significantly improved accuracy and robustness. Therefore, the construction of a fluorescent array system could contribute to the evolution of paper-based POC testing as a cost-effective multiplexed assay system along with good accuracy and precision.

Here, we report a new discovery of an indolizine chemical structure, named Kaleidolizine (KIz), and a library of 75 fluorescence compounds derived from the tunable KIz core skeleton. The 75 KIz-derived compounds are excited by single UV excitation and generate emission lights in the full spectra of visible wavelengths. Each compound interacts with multiple analytes, resulting in different photophysical properties upon binding (e.g. intensity increase/decrease and hypsochromic/bathochromic shifts). As a result, the combined fluorescent changes for each individual component can represent a unique ID for different analytes. We confirmed that exposure of the array with 35 different VOCs result unique fluorescent change patterns of the array according to the VOCs. By implementing machine-learning algorithms as “the brain” of our system, we demonstrate accurate and reproducible discrimination of 5 different volatile organic compounds (VOC) using a smartphone-based detection system with fluorescent compound arrays on disposable cellulose paper. The method is cost-effective, robust, easy-to-use, and readily applicable to highly multiplexed sensing applications.

2. Experimental

2.1. Compound characterization

^{1}H NMR and ^{13}C NMR data were recorded on Varian Mercury plus (400 MHz) spectrometer, and the chemical shifts (δ) were reported in parts per million (ppm) and calibrated using internal tetramethylsilane (TMS) standard. Multiplicity was indicated as follows: s (singlet); d (doublet); t (triplet); q (quartet); m (multiplet); dd (doublet of doublets); dt (doublet of triplets); br (broad) etc. Coupling constants (J) were reported in hertz (Hz). Low resolution mass spectrometry (LRMS) was obtained by LCMS-2020 [Shimadzu] and compact mass spectrometer [Advion].

2.2. Materials

All chemicals were purchased from Sigma-Aldrich, Tokyo Chemical Industry Co., Ltd, ThermoFisher Scientific, and Alfa-Aesar, and used without further purification unless otherwise specified. The progress of reaction was monitored using thin-layer chromatography (TLC) (silica gel 60, F254 0.25 mm), and components were visualized by observation under UV light (254 and 365 nm) or by treating the TLC plates either with p-anisaldehyde, $KMnO_4$, Phosphomolybdic (PMA) or ninhydrin followed by heating. Solvents were purchased from commercial venders and used without further purification.

2.3. Absorption and fluorescence related properties

Solution state photophysical property measurement: Fluorescence emission spectra and UV absorption spectra of KIz solutions (10 μM) in dichloromethane (DCM with 0.1 % DMSO) were recorded on JASCO FP-8200 spectrofluorometer and on JASCO V-670 spectrophotometer. Absolute quantum yield was measured by QE-2000 (Otsuka Electronics)

2.4. Solid state photophysical property measurement

The glass substrates were cleaned with deionized water, acetone, and isopropanol for 10 min each in an ultra sonicator. After drying, the substrates were treated with UV light(365 nm) for 20 min. The KIz solution in chloroform (200.0 μL, 20.0 mg/mL) was treated to the glass substrates and spin-coated at 1500 rpm for 60 s. JASCO FP-8200 spectrophotometer and a JASCO V-670 UV–vis spectrometer were used to perform the spectral measurement. UV–vis absorption and fluorescence emission were measured before and after the exposure of film with 3 different VOCs (acetic acid, ethyl acetate, ethyl amine) for 10 min.

2.5. Quantum mechanical calculations

All quantum mechanical calculations were performed in Gaussin09W. The ground state structures of Kaleidolizine were optimized using density functional theory (DFT) at the B3LYP/6−31G* level. Energy levels of the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) were calculated with the optimized ground state molecular geometry.

2.6. Experimental Procedure for sensor array preparation by wax printing

We designed KIz sensor array (23 mm × 23 mm) by printing patterns of hydrophobic barriers as black lines on a white background filter paper using word processer program (Hangul 2018). The black lines were printed on filter paper (HYUNDAI MICRO No. 22) employing a wax printer (Xerox color cube 8870). The printed filter paper was then heated at 150 ℃ for 5 min by placing on a hotplate, which melted the wax applied on the filter paper toward back side of the paper. After preparing the patterned filter paper, KIz stock solution in DMSO (10 mM) was spotted on the paper using glass capillaries (around 0.15 μL). Finally resulting paper was dried in dry oven at 80 ℃ for 2 h.

2.7. Smartphone-based detection system

We designed a 3D-printed plastic cradle housing for the iPhone 6 smartphone (Apple Inc., CA, USA). The smartphone was held firmly by the supporting structure fitted to the phone body. The rear-facing camera (8 megapixels, 1/3-inch sensor with a 1.5 μm pixel size) of the smartphone was aligned with a camera hole (11 mm in a diameter), a macro lens (LB1844-A, f = 50.0 mm), and a long pass filter (cut-off wavelength: 400 nm). A UV LED (M365D2, 365 nm ± 9 nm) was installed next to the camera hole to illuminate the KIz arrays. All the optical components were purchased from Thorlabs Inc (NJ, USA). The macro lens was placed in front of the camera to close up the image of the

array with inconsequential barrel distortion. The long pass filter was used to reduce the background noise occurred from the tail end of the UV signals reflected and scattered from the array while passing fluorescence signals from the KIz compounds. The LED is mounted on a metal-core printed circuit board and illuminate active area of the arrays with <15 % variation in the illumination intensity. The LED light source was powered by two AAA batteries and was activated by an on/off switch installed on the instrument body. All optical and electrical components, except the smartphone, were integrated and enclosed in the cradle housing. The inside of the cradle was lightproof. The overall dimensions of the instrument are about 105 mm (w) × 80 mm (h) × 115 mm (l) (Fig. S4).

2.8. Automated color information extraction from captured fluorescent array images

In order to extract the color information from fluorescent array images, before and after the exposure with VOC, we design an automated color information extraction application from the captured fluorescent array images. The application is implemented in Python 3.7 and uses the OpenCV package, containing various computer vision algorithms. The application starts by identifying the four vertices of the sensor array and extracts the sensor array region from a rectangle of these vertices. Due to artifacts such as optical distortion and camera shooting angle variation, the sensor array image may show an irregular rectangle shape, despite the sensor itself being a perfect square. In order to mitigate negative effects from such artifacts, the application applies affine transform to rectify the shape of the sensor array from an irregular rectangle image to a square image. We firstly create a mask that represents a region of interest (ROI, in our case the area of the sensor array) in the captured image. We convert the captured image to a binary image using the OTSU threshold algorithm. We then apply morphological transformation (i.e., five times of opening and dilating) to the binary images to remove noise from binary masking. Since the fluorescent compounds are spotted at uniform distances in matrix shape on the cellulose paper, the application identifies the locations of each fluorescent compound by extracting the red points on the vertical and horizontal reference lines (lines are virtual lines to assist readers in understanding the procedure) in Fig S5. Nevertheless, since a single pixel's color may not be a representative color due to reasons such as spotting errors or camera noise, we extract the average color of the nine adjacent pixels from this center pixel to extract a stable representative color for each fluorescent compound.

2.9. VOC classification with machine learning model

We exploit the random forest machine-learning algorithm to classify 500 collected sample data using Python 3.7 and scikit-learn package. We randomly separate the 500 collected data to 70 % (350 samples) of training data and 30 % (150 samples) of testing data for each VOC compound. In the inference model training phase, the testing data is completely excluded. We set the number of trees of the random forest model to 20.

3. Results and discussion

3.1. Design of fluorescent compounds and detection systems

For the molecular array system, exploiting the diversity of chemical structures and the electronic state of the fluorescence compounds are crucial for two main reasons. First, we can maximize the molecular interaction possibilities with various analytes, and second, it opens the possibility for extracting unique fluorescent patterns for different analytes. In addition, it is desirable that all the compounds are excited with single excitation wavelength so that the optics configurations of the system can be simplified. Taking this into account, we chose indolizine as a chemical core skeleton because previous studies [18,19] implied that indolizine chemical structure could be used as versatile environment sensitive molecular platform to synthesize various fluorescent compounds (See Table S1 for comparison of KIz system with other indolizine based fluorophores) [17–19]. We reasoned that the ideal positions for substituents are the carbon positions that show the greatest differences between HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital) [18,27]. Modification of the positions with substituents having divergent electronic densities would maximize the perturbation of electronic state of the indolizine core and produce sensitive fluorescent compounds exhibiting varying photophysical properties [18,28]. Using the density-functional theory, we first calculated atomic coefficients of indolizine core skeleton in HOMO and LUMO lobes. We finalized the KIz (Kaleidoscopic Indolizine) structure that modifies C-1, C-3, and C-7 positions considering the molecular orbitals (high differences between HOMO and LUMO lobes), as well as the synthetic feasibility (Fig. 1a). Next, we selected different functional groups on each diversification position (R_1, R_2 and R_3) to efficiently diversify the electronic state of fluorescent compounds and make them sensitive to environmental chemical changes (See Supporting Scheme). We used the Hammet constant (σ_p) [29] as a guideline to induce intramolecular charge transfer (ICT) phenomena [30,31]. Overall, a combination of 5 × 5 × 3 different substituents on the R_1, R_2 and R_3 positions produced 75 different KIz compounds that are simultaneously existed by single UV excitation and generate different emission spectra (Fig. 1b). To facilitate POC sensing applications (Fig. 1c), we developed a smartphone-based detection system for the KIz sensing arrays (Fig. 1d). The system was equipped with a UV light-emitting diode (LED) for fluorescence excitation and used a smartphone camera for imaging. A microlens was located in front of the camera to reduce the focal length and a long pass filter was used to block excitation light. We also designed a custom application (Fig. 1e) to constantly control imaging parameters in the camera (e.g., exposure time, ISO, focus, zoom factor). Finally, we developed and applied machine-learning algorithms to automatically analyze the color change patterns of sensing arrays and discriminate detected chemicals (Fig. 1f).

3.2. Photophysical properties of KIz compounds

After construction of fluorescent compounds library (See supporting information for detailed synthetic procedure and chemical characterization), we next measured photophysical properties of individual KIz compounds (Table S1). Modification of three diversification carbon positions with electronically divergent substituents (five, five and three different groups for R_1, R_2, and R_3, respectively) generated 75 different KIz compounds exhibiting diverse fluorescence both in solution and solid states (Fig. 2a) Photophysical properties (Tabls S2) and full spectrums (from Fig. S45 to Fig. S119) of all the compounds in both states are presented in supporting information. The KIz compounds in both states have absorption wavelengths (λ_{ab}) ranging from 356 to 464 nm and emission wavelengths (λ_{em}) ranging from 416 to 620 nm (Table S1). A more interesting feature of the KIz fluorescent core skeleton was that the systematic tuning of the photophysical property is possible in solid state [32]. It should be emphasized that, systematic tunability of photophysical properties could maximized potential of the array by making fluorescent compound with desirable properties on demand. Although many strategies have been empirically pursued for the discovery of novel fluorescence probes, however, complexity of the underlying photophysical phenomena makes the rational design of fluorescent probes in systematic way still difficult [15,17,32–38]. On the other hand, we can efficiently tune λ_{em} of KIz compounds simply by using electronically different substituents (Fig. 2b and c). For instance, we observed changing R_1 substituent from CF_3Ph (blue dots) to AcPh (green dots), Ph (yellow dots), MeOPh (orange dots) and to DMAPh (red dots) generally induced bathochromic shift of λ_{em} in solid state (Fig. 2c, right). In addition, compounds with larger σ_p (larger size dots) of R_2 exhibited longer λ_{em} than that of compounds having smaller σ_p (smaller size dots

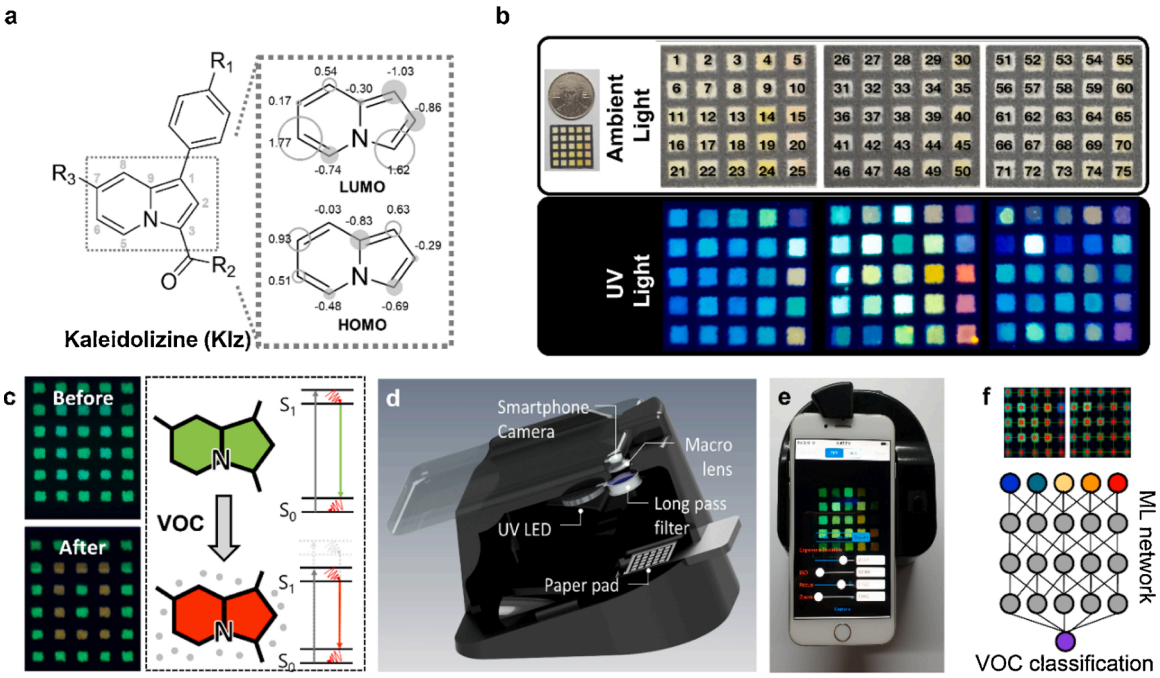


Fig. 1. Fluorescent compound array for portable classification of chemicals assisted by machine learning algorithms. a) Chemical structure of the Kaleidolizine (left) and schematic figure of the atomic coefficients of the HOMO and LUMO of indolizine core skeleton (right). The sizes and colors of the circles indicate the π-electron density and phase difference of the orbitals, respectively. b) Photograph of fluorescent compound array made with 75 different KIz compound on wax-printed cellulose paper under irradiation with ambient light (up) and 365 nm UV light (down). Inserted image represent comparable size of single array with the coin (korean hundred won). c) Photograph of fluorescent compound array for proof of concept study (made with KIz 42, and KIz 43) under 365 nm handheld UV before (left up) and after (left right) the exposure with volatile organic compound (3-Bromopropionic acid). Schematic representation for the emission wavelength changes induced by ICT process in KIz after VOC exposure (Right). d) 3D illustration of the cradle for portable application of the system. e) photograph for actual application (right). f) Schematic representation of automated analysis of fluorescent pattern changes assisted by machine learning algorithm.

in Fig. 2c, right). Therefore, strengthening the EDG (electron donating group) on R_1 and EWG (electron withdrawing group) on R_2 position of KIz induced bathochromic shift of λ_{em} in solid state. The electron density of substituent on KIz core skeleton have similar systematic tunability on solution state fluorescence (Fig. 2c, left). For the R_3 position, strengthening the EWG character generally induced a bathochromic shift of λ_{em} (Fig. 2b) in solid state (KIz compounds having Ac or CF_3 as R_3 substituent have longer λ_{em} than the KIz having H as R_3 substituent). Another point to note is that we observed predictable photophysical property of the KIz fluorescent core skeleton, which was strongly reminiscent of other indolizine based fluorescent core skeleton [18]. For example, simple linear correlation between the experimental λ_{em} and the calculated energy gap (E_{gap}) between and energy level of HOMO (E_{HOMO}) and that of LUMO (E_{LUMO}) was observed (Fig. 2b). Furthermore, strengthening EDG and EWG on R_1 and R_2 positions (inducing bathochromic shift of λ_{em}) generally reduced the theoretical E_{gap} between E_{HOMO} and E_{LUMO} in solid state rather than in solution state (Fig. 2c). In other words, we could predict the λ_{em} of fluorescent compounds based on the KIz fluorescent core skeleton simply by calculating the E_{gap} between E_{HOMO} and E_{LUMO}. Collectively, σ_p of the substituents on the R_1, R_2 and R_3 positions could provide molecular design guideline for systematic controlling the solid state photophysical property of the KIz compound, which can be predicted by calculating the E_{gap} between E_{HOMO} and E_{LUMO} values of the compound. To measure fluorescent quantum yield of the KIz compounds, solution of KIz in DMSO solvent (10 μM) was measured. Interestingly, the synthesized fluorophores have broad range of quantum yield. For instance, KIz 70 have negligible quantum yield, but KIz 31 have considerably high quantum yield (96 %). This is probably due to non-radiative decay of the energy is very different from compound to compound.

3.3. Environmental chemical sensing with KIz compounds

First of all, we initially designed the KIz library with the substituent having EDG and EWG characteristic to induced push-pi-pull structure for ICT characteristics. Computational calculation of molecular orbital of all 75 different KIz derivatives showed most of the compounds have drastic distribution differences between HOMO and LUMO lobes (See Supporting Information, Section 3), demonstrating molecular design and substituents selection induce ICT process in the molecule for molecular environment sensitive photophysical properties [30,31]. As like the solvatochromism of the molecule having ICT characteristics, we anticipated that molecular environmental changes of the KIz compound in solid state affect stability (depending on the KIz compounds and on the exposed molecule) of the excited state of the molecule, triggering photophysical changes of the molecule in solid state. To confirm environment sensitive photophysical property of the compounds, we exposed three different KIz compounds (KIz 15, KIz 40 and KIz 65), which was spin-coated on glass slides, with three different organic vapors (acetic acid, ethyl acetate, ethyl amine) and their emission spectra were obtained before and after the exposure. As shown in Fig. 3, this experiment confirmed that exposing KIz with organic vapor not only induce photophysical property changes, but also result very different response pattern especially in solid state (See Fig. 3). Therefore, this result showed that KIz compounds have different and unique photophysical response characteristics for each chemical, which indicates combinational analysis of photophysical pattern changes of KIz compound arrays could be used as a unique ID for detecting different chemicals.

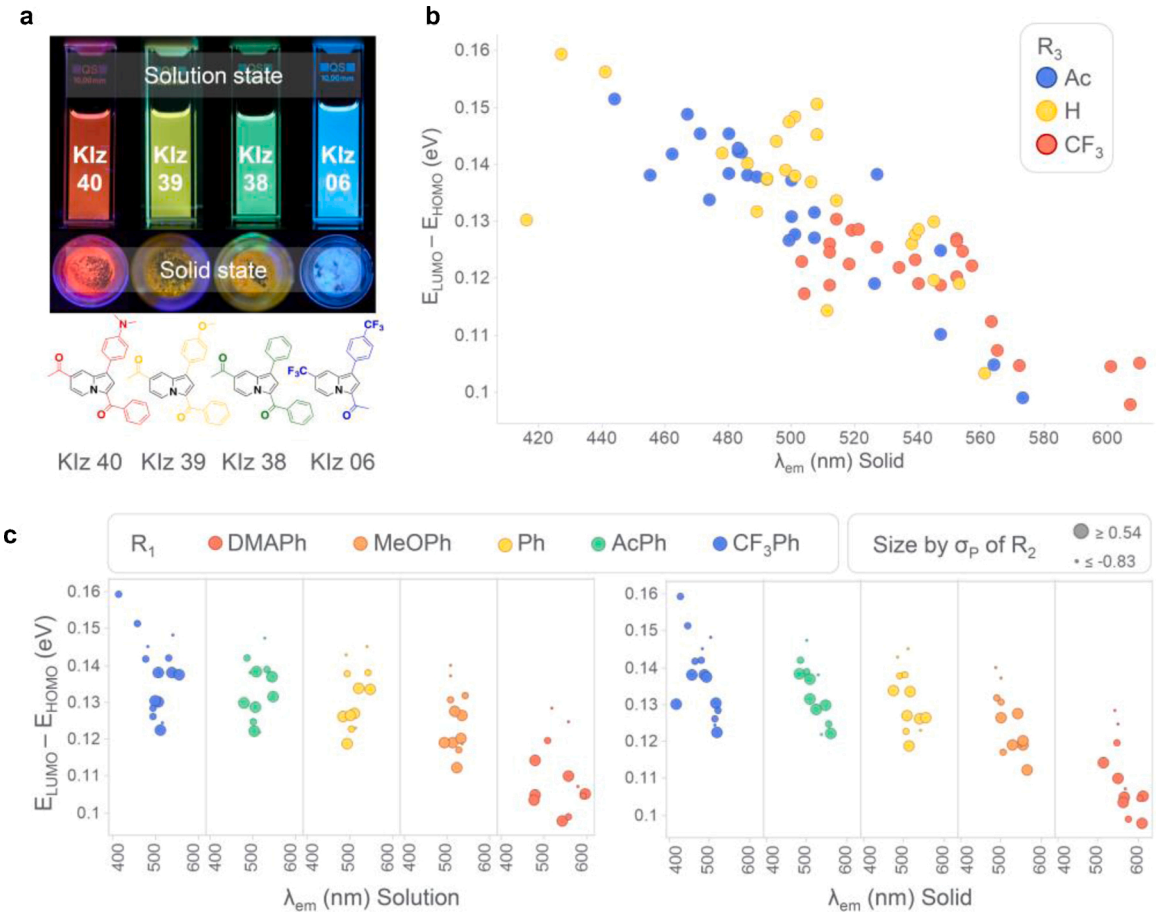


Fig. 2. The photophysical property of Kaleidolizine library. a) Photograph of KIz 40, KIz 39 KIz 38 and KIz 06 under UV irradiation (365 nm) in solution and solid state. b) Scatter plot of KIz derivatives, color coded according to the R_3 substituent (blue for acetyl, yellow for hydrogen and red for trifluoromethyl). c) Scatter plot of KIz derivatives, grouped according to R_1 substituent in solution (left) and in solid (right) state. Color of the dot represent the R_1 substituent (red, orange, yellow, green and blue for dimethylamino, methoxy, hydrogen, acetyl and trifluoromethyl, respectively). Size of the dot represent Hammet constant of R_2 substituent.

3.4. VOC classification with KIz compound arrays

For the high-dimensional multiplexed sensing, we constructed arrays of 75 KIz compounds on wax printed cellulose paper [39] (Fig. S1). Fluorescence images of the KIz compound array before and after the exposure to VOCs (Fig. S2) were taken using the hand-held smartphone system (Fig. S3 and Fig. S4) with a single UV LED excitation ($\lambda_{ex} = 365$ nm). Fluorescent color changes in individual array were extracted as Hue differences (Fig. S5). To confirm fluorescence pattern changes of the array enabling VOC classification, HUE value differences of the sensor array, exposed with 35 different VOC for 5 times, were analyzed. For the systematic investigation, we marked characteristic spot within array based on HUE difference analysis - 20 fold difference with three reproducible result among five different trial against individual VOC (Fig. 4a and Fig. S8). We confirmed that KIz fluorescence compound array exhibiting unique fluorescence changes pattern against each VOC. For instance, we confirmed each KIz compound differently respond with different organic acids (Fig. 4a) and marking of the KIz spot with higher than cut off value revealed unique fluorescence change pattern of the array for individual organic acids (Fig. 4b). This result demonstrates the power of KIz array system for VOC classification (see Figs. S7 and S9-S43 for pattern changes against all the chemicals tested).

3.5. Machine learning-based data analysis

To confirm whether increase the data collection could improve the accuracy and robustness of the system, we exposed the sensor arrays with 5 different VOC for 100 times (total 500 array sets were exposed with 5 different VOC; 100 sets for each VOC, Fig S43). In the data analysis, we first tested the principal components analysis (PCA) of 75 KIz arrays to discriminate 5 different VOC chemicals (Acetic acid, Acetone, Ethylenediamine, Phenol, and Toluene; Fig. 5a). Even with 100 measurement data for each chemical, however, the PCA approach showed relatively low accuracy as shown by the large overlaps between different groups (Fig. 5a). To improve the discriminate accuracy, we applied machine learning approaches using a random forest algorithm [40,43], one of the most popular machine learning algorithms for both classification and probability estimation. For training, we randomly select 70 % of the data as the training set and the remaining 30 % for testing. The training dataset is used by the inference model to learn the patterns of each VOC. We note that the test dataset was completely excluded from the learning process for testing and evaluations (Fig. 5b). For machine learning-based VOC classification, we converted the colors to RGB, HSV, and CIELAB color spaces [41,42] to compare the performance of each sample features on data analytic perspective. We note that we have used the Hue Difference information only from the HSV values. The main reason behind is that Hue Difference carried sufficient information to classify the VOC materials. This leads to a more

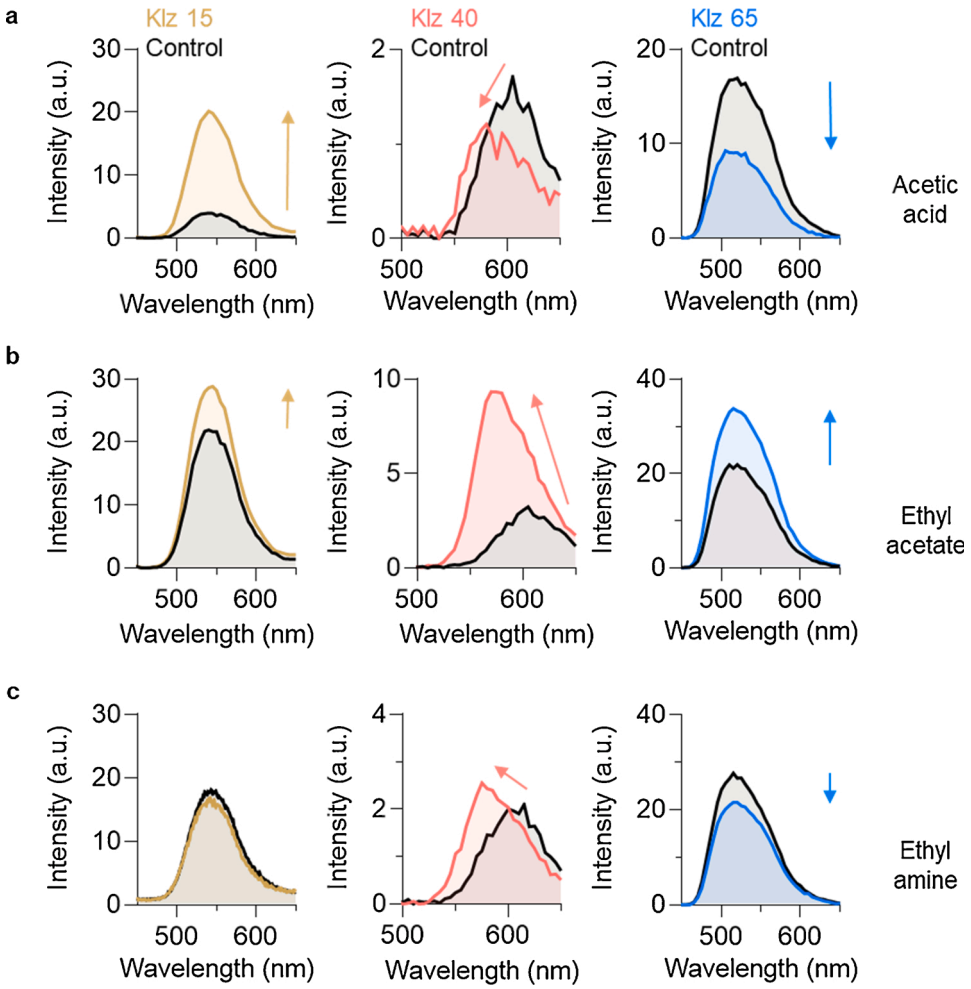


Fig. 3. Solid state photophysical property changes of KIz derivatives exposed with VOC. Photophysical property of KIz 15 (left column), KIz 40 (middle column), and KIz 65 (right column), which were spin-coated on glass, was measured before (gray spectrum) and after (colored spectrum) exposure with acetic acid (upper row), ethyl acetate (middle row), and ethylamine (downer row).

lightweight model that can be trained with smaller quantities of data. Given that we wanted to design a system that could be used in real-world use cases, we found the model size reduction to be an important aspect. Fig. 5c-e show confusion matrices of three random forest models trained with different color spaces. The row represents the ground truth and the column is the predicted label from the random forest model. The darkness of the blue cells in the confusion matrices show the frequency of inference. Specifically, the darker the blue color, the more classification results indicate a specific VOC type. With a perfect model, we would expect the diagonal cells to be dark and all the other cells to be white. Overall, on average, the three random forest models (one for each color space) show classification accuracies of higher than 95 %. For the models trained using the RGB Euclidian distance (Fig. 5c) and CIEDE2000 (Fig. 5e) shows accuracy results higher than 97 %. On a per-material perspective, when using the RGB Euclidian distance or CIEDE2000 color difference metric, the classification result for Acetic acid (C1) and Acetone (C2) shows 100 % accuracy, and all other materials show an accuracy of higher than 93 %. We conjecture that RGB Euclidean distance and CIEDE2000 color difference metrics show better results compared to the Hue difference-based model (Fig. 5d) given that the two metrics include both color difference and brightness differences. Although we have tried to analyze the data with PCA or Linear discriminant analysis (LDA) (Fig S6, Fig S7), we noticed that compared to the 90+% accuracy reported using the Random Forest model, conventional methods such as PCA or LDA was not sufficient to effectively exploit the features provided from the sensor arrays (Table S3).

4. Conclusion

For constructing fluorescence microarray system for multiplexed POC sensing, it is crucial to have molecular diversity of fluorescence compound library to maximize the molecular interaction possibilities with various analytes. In addition, dynamic molecular environment responding photophysical property of the compounds is highly desirable to make sensing array exhibit kaleidoscopic fluorescent pattern changes for increased discrimination power. Furthermore, it is necessary that all the compounds are excited with single excitation wavelength so that the optics configurations of the system can be simplified. Despite many enthusiastic strategies, it is still challenging to develop fluorescence compound library having synthetic flexibility and photophysical property diversity with single excitation light source because of complexity of the underlying photophysical phenomena. To address the bottleneck, we developed a library of KIz fluorescent compounds based on a single core skeleton. In this study, to maximize the possible fluorescent patterns of the array, chemical microenvironment sensitive new 75 KIz (Kaleidolizine) fluorophores were developed in combinatorial fashion. Structure photophysical property relationship study revealed that KIz system have systematic tunable and predictable photophysical properties especially in solid state. After confirming their chemical microenvironment sensitive photophysical property in solid state, we developed

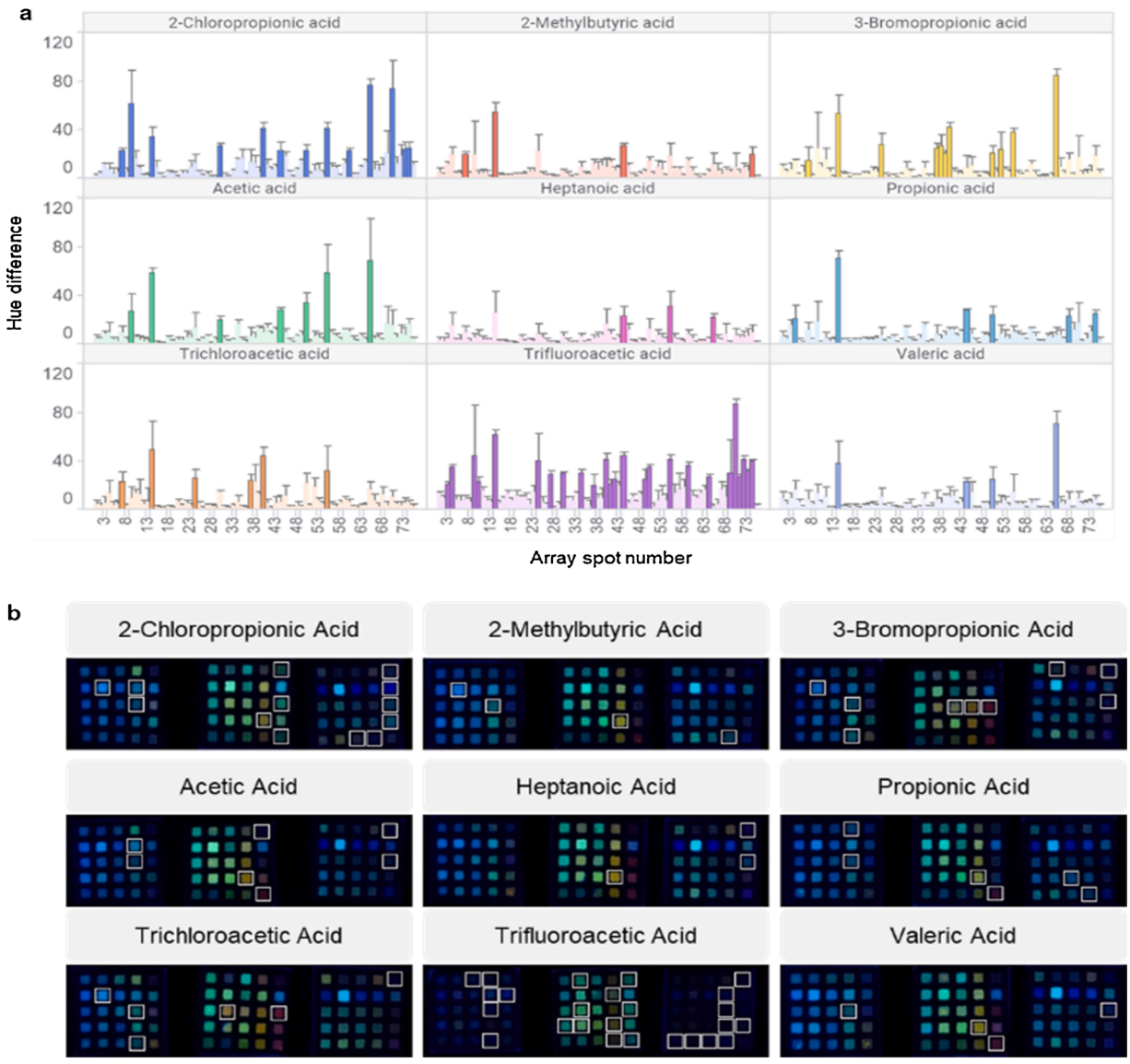


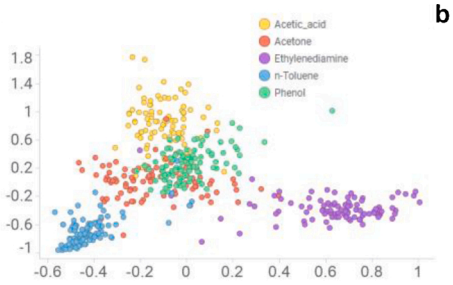
Fig. 4. Fluorescence pattern changes of KIz sensor array exposed with VOC. a) The representative bar chart analysis showing the hue difference (y axis) of individual KIz compounds on array (x axis) exposed with 9 different acids (5 replicates). Only the data from KIz compound exhibit the hue difference above the threshold (more than 20 hue difference) more than 3 times were marked in darker color than others. b) Representative photograph of the KIz sensor array under irradiation with 365 nm hand held UV lamp. White box indicate the array spot exhibit hue difference above the threshold more than 3 times. Please see supporting information for the complete data set (from Fig. S9 to Fig. S43).

the fluorescent compound array on wax printed cellulose paper. Fluorescent pattern changes of the array, induced by exposure of the array with VOC, were precisely captured by custom-designed detection instrument consists of a smartphone. The analysis of high-dimensional data from the 75 KIz compound arrays was powered by a machine-learning algorithm. A random forest is an ensemble-learning method which trains multiple inference models and measures the probability classification to a specific class with multiple models voting results. This means that a single random forest model is composed of multiple decision trees which are trained separately. Given the current size of the dataset, the random forest is a reasonable choice because multiple decision trees can make the inference model more robust to a dataset with reasonable size. We envision that the accuracy of the model will be further improved with a larger amount of training dataset that can be accumulated over time, and the next-generation algorithm could be applied to discriminate multiple targets in a background of non-target molecules. Finally, we confirmed machine learning algorithms allowed accurate classification of VOC, higher than 97 %.

In summary, KIz core skeleton provides infinite molecular platform to generate diverse fluorescent compounds. Combinatorial synthetic scheme allowed synthetic accessibility for expanding molecular diversity of fluorescent compound. In addition, systematic tunability and predictability of photophysical property is available both in solution and in solid state. Chemical environment sensitive photophysical property of the KIz system allowed unique fluorescence pattern generating fluorescent compound array on cellulose paper. The sensing array composed with KIz compounds could detect and classify various VOCs using a smartphone-based detection system powered by machine learning algorithms. Based on this result, we are currently trying to characterize VOC mixtures and the result of the study will be reported in due course.

We applied the sensor array for VOC classification in this study, however, we wanted to demonstrate that 75 sensing arrays can be readily made using the core skeleton and thereby to confirm that KIz core skeleton could provide molecular platform to generate diverse

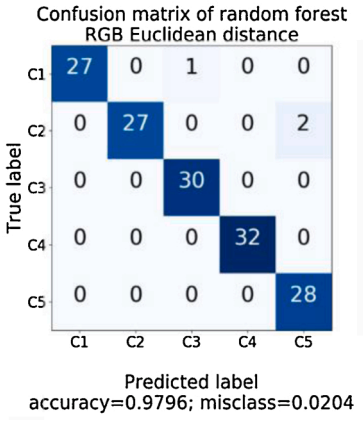
a



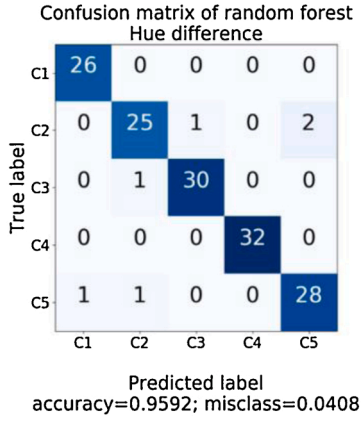
b

Species	Abb.	# of training data	# of test data	# of total data
Acetic acid	C1	67 (71.3%)	27 (28.7%)	94
Acetone	C2	69 (72.0%)	27 (28.1%)	96
Ethylenediamine	C3	67 (68.4%)	31 (31.6%)	98
Phenol	C4	68 (68.0%)	32 (32.0%)	100
Toluene	C5	70 (70.0%)	30 (30.0%)	100

c



d



e

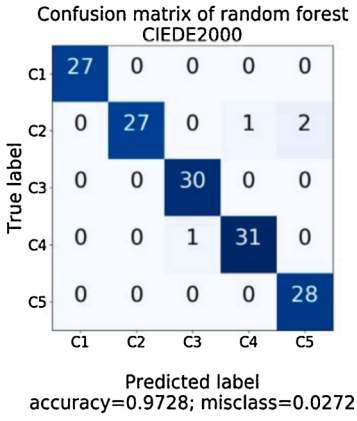


Fig. 5. Machine learning algorithm assisted VOC classification a) PCA analysis of data set acquired with pattern changes of fluorescent array, exposed with 5 different VOC (acetic acid (yellow), acetone (red), ethylenediamine (violet), phenol (green) and toluene (blue)). For each VOC, data from 100 replicates were analyzed. b) Data description and summary of classification accuracy for each VOC type. c-e) Confusion matrices of 5-folds cross-validation Random forest models with RGB Euclidean distance (c), Hue difference (d) and CIEDE2000 dataset (e).

fluorescent compounds, especially for machine-learning-based point of care chemical sensing. The developed fluorescent core skeleton sensitive to chemical microenvironment change could pave new way for molecular recognition by generating fluorescent compounds array working synergistically with machine learning algorithms. We envision that many of well-established strategies for the development of fluorescent sensors could be efficiently integrated into the KIz system. Since KIz compounds with various R_1, R_2 and R_3 substituents can be synthesized using straightforward combinatorial reactions, we could improve the specificity and sensitivity of the array against target of interest by construction of more focused fluorescent compound library via incorporation of the more specific functional group on the verification position. With the unique systematic tunability and predictability of the KIz system, we could design new fluorescent sensors with improved sensitivity and specificity for certain charges molecules using the same core skeleton. We expect that the KIz system will be highly beneficial in POC environmental monitoring; not only for VOC classification, but also for detection of pollutions, pathogens and even for biological markers, suited for health care system in point-of-care settings. Following progress will be reported in due course.

CRediT author statement

Hyungi Kim, Sang-Kee Choi, and Eunha Kim designed the study.

Hyungi Kim, Sang-Kee Choi, Jungmo Ahn, Hojeong Yu, Kyoungha Min and Changgi Hong collected the data.

Hyungi Kim, Sang-Kee Choi, Ik-Soo Shin, Sanghee Lee, Hyungsoon Im, JeongGil Ko and Eunha Km analyzed the data.

Hyungi Kim, Sang-Kee Choi, Jungmo Ahn, Sanghee Lee, Hakho Lee, Hyungsoon Im, JeongGil Ko and Eunha Kim interpreted the data and wrote the manuscripts.

All authors revised, reviewed, and approved of the final version of the manuscript.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgment

This study was supported in part by Ajou University research fund, National Institutes of Health (R00CA201248), Creative Materials Discovery Program through the National Research Foundation (2019M3D1A1078941), Technology Innovation Program (10077599) funded by the Ministry of Trade, Industry & Energy, the KRIBB Research Initiative Program [KGM9952011], and by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (NRF-2015R1A5A1037668, NRF-2020R1C1C1010044) (NRF-2019R1A6A1A11051471) and the ITRC Support Program supervised by the IITP (#IITP-2020-2020-0-01461).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.snb.2020.129248.

References

[1] Z.S. Ballard, H.-A. Joung, A. Goncharov, J. Liang, K. Nugroho, D.D. Carlo, O. B. Garner, A. Ozcan, Deep learning-enabled point-of-care sensing using multiplexed paper-based sensors, NPJ Digit. Med. 3 (2020) 66.

[2] N.P. Pai, C. Vadnais, C. Denkinger, N. Engel, M. Pai, Point-of-Care testing for infectious diseases: diversity, complexity, and barriers in low- and middle-income countries, PLoS Med. 9 (2012), e1001306.
[3] A.H. Jalal, F. Alam, S. Roychoudhury, Y. Umasankar, N. Pala, S. Bhansali, Prospects and challenges of volatile organic compound sensors in human healthcare, ACS Sens. 3 (2018) 1246–1263.
[4] J.-W. Oh, W.-J. Chung, K. Heo, H.-E. Jin, B.Y. Lee, E. Wang, C. Zueger, W. Wong, J. Meyer, C. Kim, S.-Y. Lee, W.-G. Kim, M. Zemla, M. Auer, A. Hexemer, S.-W. Lee, Biomimetic virus-based colourimetric sensors, Nat. Commun. 5 (2014) 3043.
[5] S. Shrivastava, T.Q. Trung, N.-E. Lee, Recent progress, challenges, and prospects of fully integrated mobile and wearable point-of-care testing systems for self-testing, Chem. Soc. Rev. 49 (2020) 1812–1866.
[6] H.-A. Joung, Z.S. Ballard, J. Wu, D.K. Tseng, H. Teshome, L. Zhang, E.J. Horn, P. M. Arnaboldi, R.J. Dattwyler, O.B. Garner, D.D. Carlo, A. Ozcan, Point-of-Care serodiagnostic test for early-stage lyme disease using a multiplexed paper-based immunoassay and machine learning, ACS Nano 14 (2020) 229–240.
[7] J.P. McGann, Poor human olfaction is a 19th-century myth, Science 356 (2017) eaam7263.
[8] B. Malnic, J. Hirono, T. Sato, L.B. Buck, Combinatorial receptor codes for odors, Cell 96 (1999) 713–723.
[9] L.B. Buck, Unraveling the sense of smell (Nobel lecture), Angew. Chem. Int. Ed. 44 (2005) 6128–6140.
[10] D. Wu, A.C. Sedgwick, T. Gunnlaugsson, E.U. Akkaya, J. Yoon, T.D. James, Fluorescent chemosensors: the past, present and future, Chem. Soc. Rev. 46 (2017) 7105–7123.
[11] J.L. Kolanowski, F. Liu, E.J. New, Fluorescent probes for the simultaneous detection of multiple analytes in biology, Chem. Soc. Rev. 47 (2017) 195–208.
[12] K.P. Carter, A.M. Young, A.E. Palmer, Fluorescent sensors for measuring metal ions in living systems, Chem. Rev. 114 (2014) 4564–4601.
[13] A.P. de Silva, H.Q.N. Gunaratne, T. Gunnlaugsson, A.J.M. Huxley, C.P. McCoy, J. T. Rademacher, T.E. Rice, Signaling recognition events with fluorescent sensors and switches, Chem. Rev. 97 (1997) 1515–1566.
[14] T.L. Mako, J.M. Racicot, M. Levine, Supramolecular luminescent sensors, Chem. Rev. 119 (2018) 322–477.
[15] L.D. Lavis, R.T. Raines, Bright building blocks for chemical biology, ACS Chem. Biol. 9 (2014) 855–866.
[16] L.D. Lavis, R.T. Raines, Bright ideas for chemical biology, ACS Chem. Biol. 3 (2008) 142–155.
[17] Y. Lee, Systematic Exploration of Indolizine-Based Small Fluorescent Molecules, Synthesis, Analysis and Application, Springer Theses; Springer Singapore, Singapore, 2018.
[18] E. Kim, Y. Lee, S. Lee, S.B. Park, Discovery, understanding, and bioapplication of organic fluorophore: a case study with an indolizine-based novel fluorophore, seoul-fluor, Acc. Chem. Res. 48 (2015) 538–547.
[19] Y. Ge, A. Liu, J. Dong, G. Duan, X. Cao, F. Li, A simple pH fluorescent probe based on new fluorophore indolizine for imaging of living cells, Sens. Actuators B Chem. 247 (2017) 46–52.
[20] A.P. de Silva, T.S. Moody, G.D. Wright, Fluorescent PET (Photoinduced Electron Transfer) sensors as potent analytical tools, Analyst 134 (2009) 2385–2393.
[21] N.A. Rakow, K.S. Suslick, A colorimetric sensor array for odour visualization, Nature 406 (2000) 710–713.
[22] J.R. Askim, M. Mahmoudi, K.S. Suslick, Optical sensor arrays for chemical sensing: the optoelectronic nose, Chem. Soc. Rev. 42 (2013) 8649–8682.
[23] Z. Li, J.R. Askim, K.S. Suslick, The optoelectronic nose: colorimetric and fluorometric sensor arrays, Chem. Rev. 119 (2019) 231–292.
[24] D. Svechkarev, M.R. Sadykov, K.W. Bayles, A.M. Mohs, Ratiometric fluorescent sensor array as a versatile tool for bacterial pathogen identification and analysis, ACS Sens. 3 (2018) 700–708.
[25] B. Wang, J. Han, N.M. Bojanowski, M. Bender, C. Ma, K. Seehafer, A. Herrmann, U. H.F. Bunz, An optimized sensor array identifies all natural amino acids, ACS Sens. 3 (2018) 1562–1568.
[26] S. Rana, S.G. Elci, R. Mout, A.K. Singla, M. Yazdani, M. Bender, A. Bajaj, K. Saha, U. H.F. Bunz, F.R. Jirik, V.M. Rotello, Ratiometric array of conjugated polymers–Fluorescent protein provides a robust mammalian cell sensor, J. Am. Chem. Soc. 138 (2016) 4522–4529.
[27] K. Higashiguchi, K. Matsuda, Y. Asano, A. Murakami, S. Nakamura, M. Irie, Photochromism of dithienylethenes containing fluorinated thiophene rings, Eur. J. Org. Chem. 2005 (2005) 91–97.
[28] E. Kim, M. Koh, J. Ryu, S.B. Park, Combinatorial discovery of full-color-Tunable emissive fluorescent probes using a single core skeleton, 1,2-Dihydropyrrolo[3,4-β] indolizin-3-one, J. Am. Chem. Soc. 130 (2008) 12206–12207.
[29] C. Hansch, A. Leo, R.W. Taft, A survey of Hammett substituent constants and resonance and field parameters, Chem. Rev. 91 (1991) 165–195.
[30] C. Reichardt, Solvatochromic dyes as solvent polarity indicators, Chem. Rev. 94 (1994) 2319–2358.
[31] Z.R. Grabowski, K. Rotkiewicz, W. Rettig, Structural changes accompanying intramolecular Electron transfer: focus on twisted intramolecular charge-transfer states and structures, Chem. Rev. 103 (2003) 3899–4032.
[32] X. Hou, C. Ke, C.J. Bruns, P.R. McGonigal, R.B. Pettman, J.F. Stoddart, Tunable solid-state fluorescent materials for supramolecular encryption, Nat. Commun. 6 (2015) 6884.
[33] J.B. Grimm, A.K. Muthusamy, Y. Liang, T.A. Brown, W.C. Lemon, R. Patel, R. Lu, J. J. Macklin, P.J. Keller, N. Ji, L.D. Lavis, A general method to fine-tune fluorophores for live-cell and in vivo imaging, Nat. Methods 14 (2017) 987–994.
[34] G.R. Rosania, J.W. Lee, L. Ding, H.-S. Yoon, Y.-T. Chang, Combinatorial approach to organelle-targeted fluorescent library based on the styryl scaffold, J. Am. Chem. Soc. 125 (2003) 1130–1131.
[35] T. Miura, Y. Urano, K. Tanaka, T. Nagano, K. Ohkubo, S. Fukuzumi, Rational design principle for modulating fluorescence properties of fluorescein-based probes by photoinduced Electron transfer, J. Am. Chem. Soc. 125 (2003) 8666–8671.
[36] Y. Lee, A. Jo, S.B. Park, Rational improvement of molar absorptivity guided by oscillator strength: a case study with furoindolizine-based core skeleton, Angew. Chem. Int. Ed. 127 (2015) 15915–15919.
[37] E. Kim, S.B. Park, in: A.P. Demchenko (Ed.), Discovery of New Fluorescent Dyes: Targeted Synthesis or Combinatorial Approach? In Advanced Fluorescence Reporters in Chemistry and Biology, 8, Springer Series on Fluorescence; Springer-Verlag Berlin Heidelberg, New York, 2010, pp. 149–186.
[38] K. Umezawa, A. Matsui, Y. Nakamura, D. Citterio, K. Suzuki, Bright, color-tunable fluorescent dyes in the Vis/NIR region: establishment of new "Tailor-Made" multicolor fluorophores based on borondipyrromethene, Chem. Eur. J. 15 (2009) 1096–1106.
[39] M. Sher, R. Zhuang, U. Demirci, W. Asghar, Paper-based analytical devices for clinical diagnosis: recent advances in the fabrication techniques and sensing mechanisms, Expert Rev. Mol. Diagn. 17 (2017) 351–366.
[40] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
[41] G. Sharma, W. Wu, E.N. Dalal, The CIEDE2000 color-difference formula: implementation notes, supplementary test data, and mathematical observations, Color Res. Appl. 30 (2004) 21–30.
[42] M.R. Luo, G. Cui, B. Rigg, The development of the CIE 2000 colour-difference formula: CIEDE2000, Color Res. Appl. 26 (2001) 340–350.
[43] Jungmo Ahn, Hyungi Kim, Eunha Kim, Jeonggil Ko, VOCkit: A Low-Cost IoT Sensing Platform for Volatile Organic Compound Classification, Ad Hoc Networks (2021). In press.

Hyungi Kim received his B.S. degree from Department of applied chemistry and biological technology (2015). He is currently a Doctor's candidate in the Department of Molecular Science and Technology of Ajou University. His research interests are focused on the development of fluorescent probes capable of sensing the toxic or bioactive species for variety of applications.

Sang-Kee Choi received his B.S. degree in 2016 from Department of applied chemistry, Konkuk University, Korea. He is currently a Doctor's candidate in the Department of Molecular Science and Technology of Ajou University, Korea. His research interests are focused on design and application of fluorescent probes based on aggregation-induced emission.

Jungmo Ahn received his B.S. degree in Software and Computer Engineering from Ajou University and is currently a Ph.D. candidate in the Department of Computer Engineering at Ajou University. His research interests are in designing intelligent embedded and mobile computing systems.

Kyoungha Min received his B.S. degree in Electrical Engineering and Computer Science from the University of California, Berkeley. He was previously working at the Center for Systems Biology at the Massachusetts General Hospital as a research assistant where he developed iOS apps for camera and image processing. He is currently a Software Engineer at Moloco, Inc, in Palo Alto, CA, USA.

Hojeong Yu is currently a Postdoc Research Fellow at the Center for Systems Biology at the Massachusetts General Hospital. He received his Ph.D. in Electrical Engineering from the University of Illinois at Urbana-Champaign. He developed a smartphone-based mobile imaging system.

Changgi Hong received his B.S degree from Department of applied chemistry and biological engineering at Ajou University. He is currently a Doctor's candidate in the Department of applied bio and technology at Seoul National University.

Ik-Soo Shin received his Ph.D. degree from Department of Chemistry at Seoul National University in 2007. After postdoctoral work at University of Texas at Austin under guidance with Richard M. Corooks. From 2008 to 2009, Ik-Soo Shin was a senior researcher at the Samsung Total and he was an research professor at the Department of chemistry at Seoul National University from 2009 to 2012. He is now associate professor at the department of chemistry at Soongsil University. His research interests are focused on the development of electrochemical materials for sensing the toxic or bioactive species.

Sanghee Lee received her Ph.D. from the Department of Chemistry at Seoul National University in 2014. After postdoctoral work at Novartis along with Harvard Medical School, she began her independent researcher career in 2017 in the Center for Neuro-Medicine at Korea Institute of Science and Technology in Seoul. Her research interests are focused on the development of novel fluorescence-based smart molecules for monitoring biological phenomena and development of novel therapeutic agents.

Hakho Lee is Associate Professor in Radiology at Harvard Medical School, Director of the Biomedical Engineering Program at the Center for Systems Biology (CSB), Massachusetts General Hospital (MGH), and Hostetter MGH Research Scholar. He received his Ph.D. in Physics from Harvard University and completed his postdoctoral training at MGH. Dr. Lee has extensive experience in exosomes, nanomaterials, biophysics, and electrical

engineering. His research focuses on developing novel biomedical sensors for clinical applications. Prof. Lee has been leading the Biomedical Engineering Program at CSB. He is also a global faculty at IBS Center for NanoMedicine in Korea.

Hyungsoon Im is Assistant Professor in Radiology at Harvard Medical School and a Principal Investigator at the Center for Systems Biology at the Massachusetts General Hospital. He received his Ph.D. in Electrical Engineering from the University of Minnesota, Twin-Cities. His laboratory focuses on developing next generation sensing technologies to better understand the make-up of human diseases and changes associated with disease progression and therapy.

JeongGil Ko received his Ph.D. degree in Computer Science from the Johns Hopkins University in 2012. From June 2012 to August 2015, JeongGil Ko was a senior researcher at the Electronics and Telecommunications Research Institute (ETRI) and he was an assistant professor at the Department of Software and Computer Engineering at Ajou University from September 2015 to August 2019. He is now assistant professor at the School of Integrated Technology at Yonsei University. He is a senior member of the IEEE since 2017. His research interests are in the general area of developing embedded and mobile computing systems with ambient intelligence.

Eunha Kim received his Ph.D. from the Department of Chemistry at Seoul National University in 2011. After postdoctoral work at Harvard Medical School, he began his academic career in 2015 in the Department of Molecular Science and Technology at Ajou University in Suwon. His research interests are focused on the development of novel fluorescence-based smart molecules for applications in the sensing and imaging.