

# **De novo creation of a naked-eye-detectable fluorescent molecule based on quantum-chemical computation and machine learning**

Masato Sumita,<sup>1,2\*</sup> Kei Terayama,<sup>1,3,4,5</sup> Naoya Suzuki,<sup>6</sup> Shinsuke Ishihara,<sup>2</sup> Ryo

Tamura,<sup>2,7,8</sup> Mandeep K. Chahal,<sup>2</sup> Daniel T. Payne,<sup>2,9</sup> Kazuki Yoshizoe,<sup>1</sup> Koji Tsuda<sup>1,7,8\*</sup>

<sup>1</sup>*RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan.*

<sup>2</sup>*International Center for Materials Nanoarchitectonics (WPI-MANA), National Institute for Materials Science, 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan.*

<sup>3</sup>*Graduate School of Medical Life Science, Yokohama City University, 1-7-29, Suehiro-cho, Tsurumi-ku, 230-0045 Kanagawa, Japan.*

<sup>4</sup>*Graduate School of Medicine, Kyoto University, 53 Shogoin-Kawaharacho, Sakyo-ku, 606-8507 Kyoto, Japan.*

<sup>5</sup>*Medical Sciences Innovation Hub Program, RIKEN Cluster for Science, Technology and Innovation Hub, Tsurumi-ku, Kanagawa 230-0045, Japan.*

<sup>6</sup>*Materials Science and Engineering, Osaka Prefecture University, 1-1 Gakuen-cho, Nakaku, Sakai, Osaka 599-8531, Japan.*

<sup>7</sup>*Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan.*

<sup>8</sup>*Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwa-no-ha, Kashiwa, Chiba 277-8561, Japan.*

<sup>9</sup>*International Center for Young Scientists (ICYS), National Institute for Materials Science, 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan.*

Correspondence and requests for materials should be addressed to M. S. (email: masato.sumita@riken.jp) or to K. Tsuda. (email: tsuda@k.u-tokyo.ac.jp)

## **Abstract**

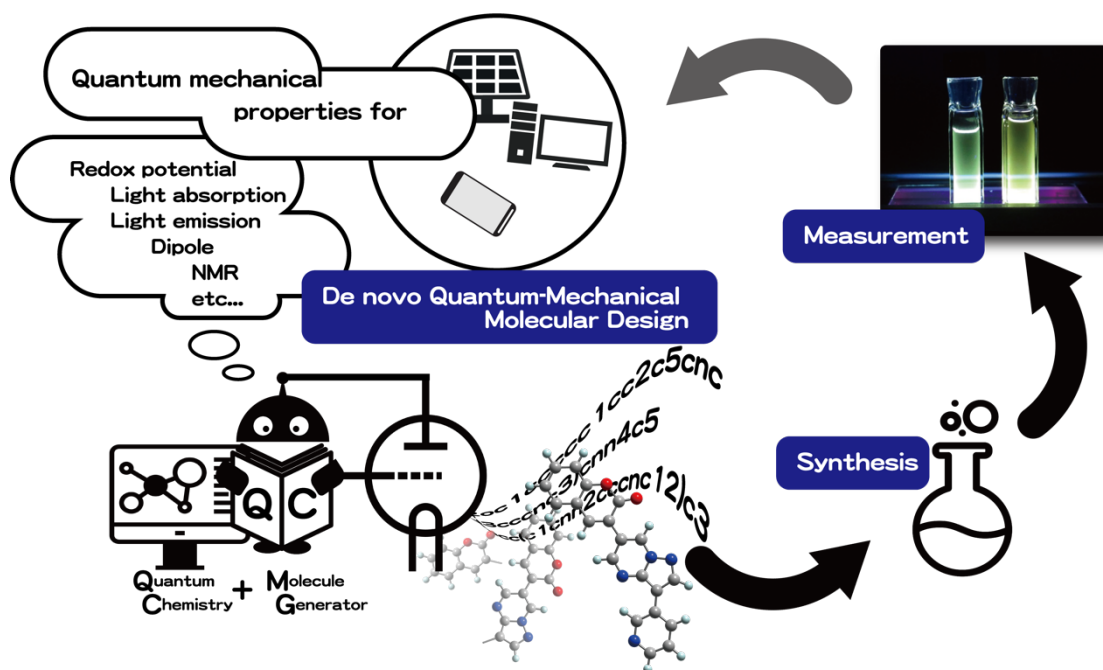
Correlations between molecular properties and structures, such as those between the absorption wavelength and conjugate length, are beneficial for designing materials and controlling their properties. However, determining the molecular structures that correlate with the target molecular properties (such as molecular fluorescence) is not an easy task. In this study, we have used a de novo molecule generator (DNMG) coupled with quantum-chemical computation (QC) to develop new fluorescent molecules, which are garnering significant attention in various disciplines. With massive parallel computation (1024 cores, 5 days), DNMG has produced 3,643 candidate molecules within the density functional theory (DFT; one of QC) framework. Among the generated molecules, we have selected an unreported molecule and synthesized it for photoluminescence spectrum measurement. Our experimental verification demonstrated that DNMG can successfully create a new molecule which emits fluorescence detectable by the naked eye, as predicted by the DFT.

## 1. Introduction

Professional knowledge or intuition is an effective tool for producing new functional molecules. Machine learning (ML) algorithms are lately being used as an alternative tool for professional knowledge or intuition. The application of ML<sup>1-4</sup> in games and artificial creativities, such as image, music, and text, to chemistry and materials science has produced significant results that have reduced the workload involved in discovering out-of-trend materials<sup>5</sup> and accelerated the development of new materials.<sup>6-8</sup> The synergy of robotics,<sup>9-11</sup> chemistry and materials science with ML can enable development of automated fabricating material systems (AFMSs). Among the components of an AFMS, the molecular design based on ML from scratch (de novo molecule generators; DNMG),<sup>12,13</sup> whose validation has just begun, plays an important role in deciding the target molecules to be synthesized. DNMGs can design new molecules in computers by setting the target properties (inverse molecular design), recognizing the valence of atoms through a few ML algorithms with the help of computational methods in quantum chemistry, computational materials science, and drug science. Therefore, the effective use of DNMGs can increase the possibility of discovering new functional molecules. Traditionally, new molecules are often developed based on known molecules by adjusting or replacing their functional groups.<sup>14-16</sup> However, even just screening the effects of

several functional groups to a known molecule is burdensome owing to the diversity of molecules; DNMGs can implicitly automate this process.

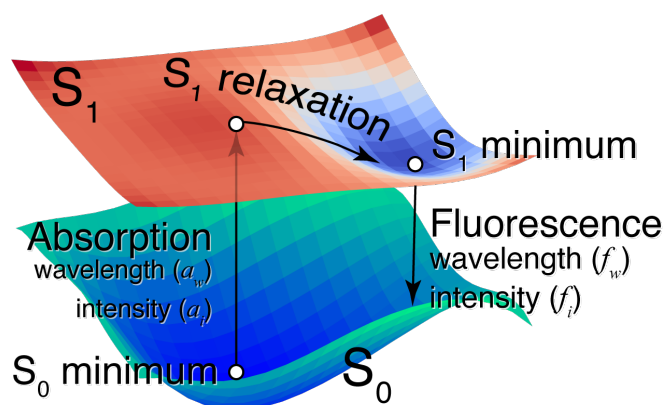
Various DNMGs based on ML have been developed for designing molecules that require simple values such as the logarithm of the partition coefficient (logP), which can be estimated from the fragments constituting a molecule.<sup>13</sup> Combining the DNMGs with several classical simulations improves the versatility and practicality of the generated molecule. For instance, DNMGs with docking simulation for proteins are investigated by synthesizing mimetic molecules.<sup>17</sup> These, coupled with classical molecular dynamics or prediction models, have directed the synthesis of functional polymers.<sup>18,19</sup> However, we have combined quantum-chemical computations (QCs) with our DNMG, called ChemTS,<sup>20</sup> for designing functional molecules characterized by their quantum mechanical properties (Figure 1).<sup>21</sup> In principle, this generator can design any molecule from scratch, by setting properties that are available in QC.



**Figure 1.** Molecule creation based on quantum-chemical computations (QCs) and machine learning (ML). The generator (ChemTS) creates molecules from scratch in accordance with QCs. Because the generator uses QCs as an evaluation tool, it is possible to design candidate molecules that can be applicable to materials for various electronic devices. In this work, using the generator, we created a new fluorescent molecule and measured its fluorescence.

In our previous work, ChemTS coupled with QC was applied to design molecules that can absorb light with the desired wavelength.<sup>21</sup> Among the 86 designed and generated molecules, six known molecules were selected that were not included in the training

dataset for the ultraviolet visible (UV) absorption measurement. Subsequently, it was confirmed that the measured results were consistent with the target wavelength of the generator. This result proved that the synthesizability and reliability of the molecules developed by ChemTS can be drastically improved by the combination of QCs because the QC simulation filtered chemically legal molecules from generated molecule due to the failure of simulation on illegal molecules. However, the “new” unreported molecules designed by ChemTS were not synthesized and thus, the innovation of ChemTS was not validated. Here, we employed ChemTS to design molecules with properties beyond the technical understanding of chemists, i.e., fluorescent molecules. In contrast to light absorption molecules, there are no clear guidelines for creating the fluorescent molecules without base molecules because there is no correlation between the fluorescence wavelength and molecular structure. In this study, we successfully synthesized a new fluorescent molecule, 3-[3-(pyridin-3-yl)pyrazolo[1,5-a]pyrimidin-6-yl]-2H-chromen-2-one, designed by ChemTS by evaluating the potential energy surfaces (PESs) shown in Figure 2. This work demonstrates the practical creation of a new molecule created by a DNMG.



**Figure 2.** Schematic of the potential energy surfaces (PESs) of the singlet ground ( $S_0$ ) state and the singlet first excited ( $S_1$ ) state of a fluorescent molecule. A fluorescent molecule at a  $S_0$  minimum is excited to its  $S_1$  excited state (vertical excitation), absorbing light of wavelength  $a_w$  with the intensity  $a_i$ . After relaxing in the  $S_1$  state, the  $S_1$  excited molecule reaches to  $S_1$  minimum and emits fluorescence whose wavelength and intensity are  $f_w$  and  $f_i$ , respectively.

Chemists and material scientists have been intrigued by fluorescent molecules owing to their noticeable visibility results in various application as photoemitters in several disciplines.<sup>14-16,22-29</sup> Representative application examples include organic light-emitting diodes,<sup>14,22-24</sup> sensors,<sup>15,25-27</sup> and bioimaging.<sup>16,28,29</sup> Although numerous fluorescent molecules had been developed for several applications, still new fluorescent molecules are constantly desired to meet the several needs. Fluorescence is a representative quantum-mechanical phenomenon; however, it does not have any correlation with



molecular structures although representative fluorescent molecules are known.<sup>14-16</sup> A simple understanding of the mechanism of fluorescence emission from a molecule is illustrated in Figure 2 from the physicochemical viewpoint. Let a molecule be in its singlet ground ( $S_0$ ) state; the molecule at the  $S_0$  minimum absorbs light to transit to the first excited singlet ( $S_1$ ) state. The  $S_1$  excited molecule relaxes to a minimum in the  $S_1$  state and back to the  $S_0$  state by emitting the energetic difference between the  $S_1$  and  $S_0$  states as light (fluorescence). Thus, by using ChemTS with our program to evaluate the PES shown in Figure 2, we controlled the fluorescence and designed the fluorescent molecules.

Several electronic structure theories for QC on molecules and materials are available through various user-friendly packages without theoretical knowledge.<sup>30</sup> We employed QC (density functional theory; DFT<sup>31</sup>) to evaluate the PESs shown in Figure 2, taking into account the balance between reliability and computational costs. However, DFT calculations require massive computational powers to extensively explore the chemical space by evaluating the generated molecules at the DFT level. Therefore, we parallelized our DNMG massively to utilize 1024 cores.

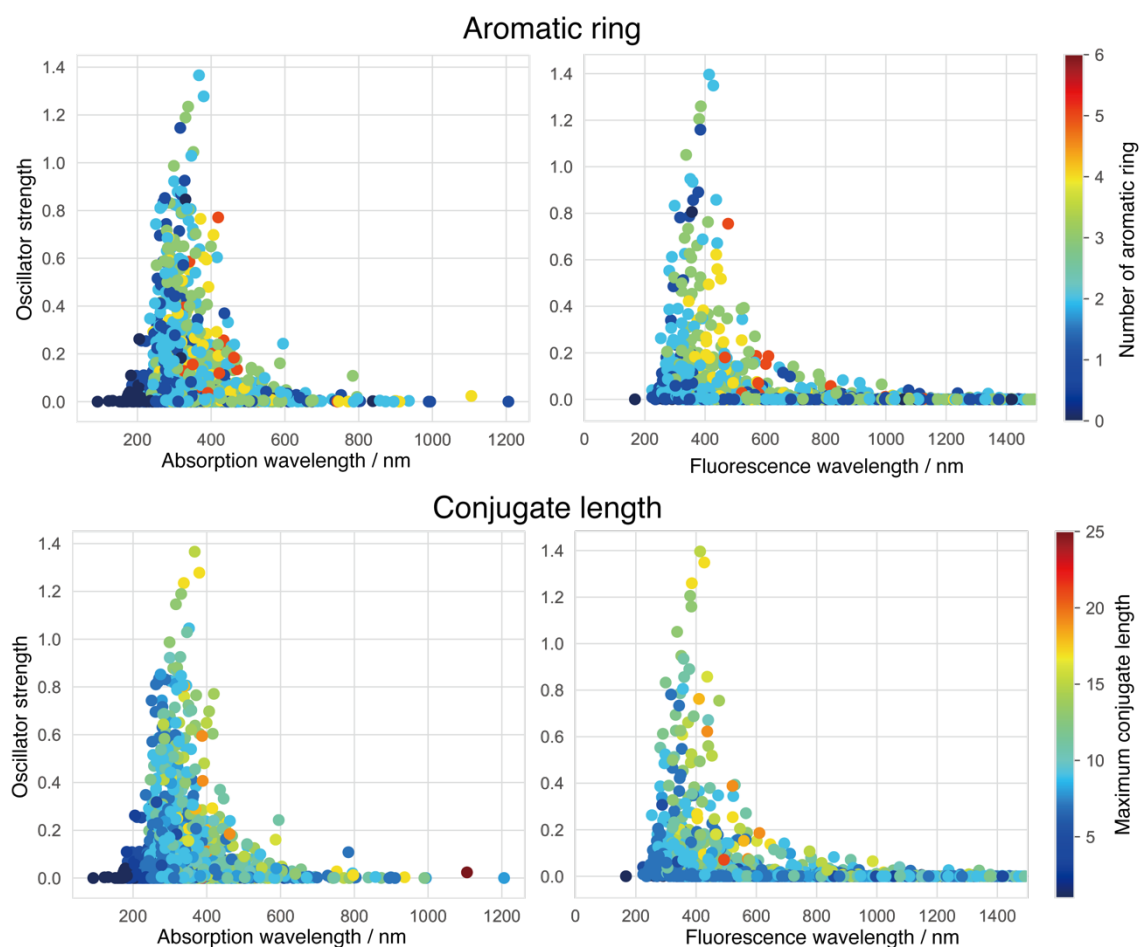
## 2. Results

## 2.1 Molecule generation

Using ChemTS, we designed 3,643 molecules as candidates of fluorescent molecules with 1024 cores for five days. The distribution profiles of the absorption and fluorescence wavelength and their oscillator strength (OS) of the molecules evaluated by DFT are shown in Figure 3 against the structural characteristics, i.e., the number of aromatic rings and the conjugate length. Their theoretical  $S_1$  excitation wavelengths are distributed from 120 to 1,200 nm. In contrast, the fluorescence wavelengths from their  $S_1$ -minimum states are distributed from 180 nm. The molecules with high OS for adsorption and fluorescence are distributed at around 200–600 nm.

The conjugation of multi-bonds and aromatic ring in organic molecules is often responsible for typical chromophores.<sup>14</sup> To confirm whether the generator can reproduce this traditional prescription or not, we have analyzed the number of aromatic rings and conjugate length. As shown in Figure 3, various color points for the number of aromatic rings are scattered on the entire distribution. Meanwhile, the bright color points for the conjugate length seems to be distributed in the high OS areas. This implies that the conjugate length might correlate with the absorption and fluorescence wavelengths, along with its intensity, instead of the number of aromatic rings. To obtain more quantitative data, we computed the correlation coefficients listed in Table 1. The highest correlation

can be observed between the conjugate length and the wavelength, and its coefficient is computed to be 0.49, which supports the traditional prescription (designing molecules that absorb light with long wavelengths implies the elongation of the conjugate length). However, this result indicates that there is no correlation between fluorescence/its OS and conjugate length/number of aromatic rings, thereby demonstrating the difficulties involved in designing fluorescent molecules with simultaneous optimization for the OS (intensity) of absorption and fluorescence by adjusting the conjugate length or the number of aromatic rings.



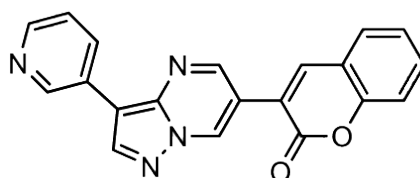
**Figure 3.** Distribution profiles for absorption to  $S_1$  states and fluorescence from  $S_1$  states of the generated molecules at the B3LYP/3-21G\* level. In the top images, the colors are discriminated based on the number of aromatic rings, whereas in the bottom images, the colors depend on the conjugate length.

**Table 1.** Correlation coefficients (R values) of wavelengths and their intensities of absorption to  $S_1$  states and fluorescence at the B3LYP/3-21G\* level from  $S_1$  states with structural parameters (number of aromatic rings and conjugate length) of the generated molecules. See the Supplementary Information for their correlation graphs.

	Wavelength	OS
$S_1$ absorption		
Number of aromatic rings	0.35	0.26
Conjugate length	0.49	0.26
$S_1$ fluorescence		
Number of aromatic rings	-0.03	0.24
Conjugate length	-0.03	0.24

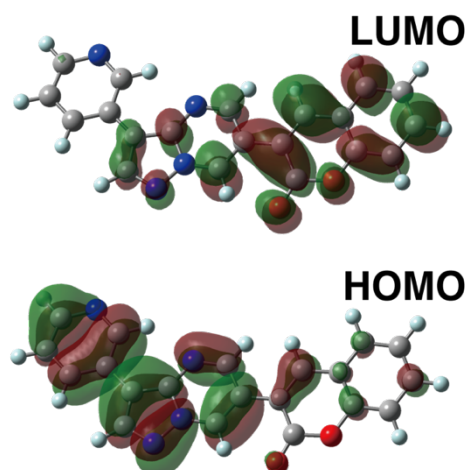
To select the candidate molecules for the synthesis, we filtered the generated molecules based on the visibility criterion. Regarding the intensity, the  $S_1$  state remains invisible until transition to the  $S_1$  state is allowed, we devised the condition that (i) the OS of a molecule should be greater than 0.1 for  $S_1$  states from the  $S_0$  minimum. For the wavelength, we imposed the condition that (ii) the molecule should emit fluorescence

with a wavelength of over 400 nm and OS of over 0.01. Finally, we constrained the difference between the wavelengths of the absorption and fluorescence because the small differences prevent the separation of fluorescence from the excitation light. Hence, we imposed the condition that (iii) the difference between the wavelength of absorption and fluorescence in the  $S_1$  states should be greater than 100 nm. Imposing (i)-(iii), we successfully filtered 87 molecules, which have been summarized in Table S1 of the Supplementary Information. Among them, we found seven known molecules or tautomers included in the chemical database (SciFinder<sup>32</sup>). To demonstrate the innovation of the generator, we selected a molecule that was not included in SciFinder<sup>32</sup>. Because it may be difficult to synthesize molecules with a novel  $\pi$ -aromatic framework, we selected a new coumarin derivatives, widely known as fluorescent molecules.<sup>15</sup> The molecule that we selected for organic synthesis in this study is shown in Scheme 1, 3-[3-(pyridin-3-yl)pyrazolo[1,5-a]pyrimidin-6-yl]-2H-chromen-2-one (**PC**).



**Scheme 1.** 3-[3-(Pyridin-3-yl)pyrazolo[1,5-a]pyrimidin-6-yl]-2H-chromen-2-one (**PC**)

The vertical excitation to the  $S_1$  state from the  $S_0$  state of compound **PC** is attributed to the transition of the one electron from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO) ( $\pi$ - $\pi^*$  excitation). As shown in Figure 4, the excitation to the  $S_1$  state induces a charge transfer (CT) from the pyrazolopyrimidines moiety to the coumarin moiety. The  $S_1$  excited molecule reaches a stationary point in the  $S_1$  state without large structural deformation, which is enough to stabilize the  $S_1$  excited **PC** to over 0.34 eV from the  $S_1$  vertical excited point. Molecules in the excited state with the CT characteristic tend to show energetically large relaxation in their excited state, thereby causing moderate differences between the absorption and fluorescence; in this molecule, the theoretically estimated difference between the two was 135.48 nm ( $S_1$  absorption wavelength: 433.75 nm,  $S_1$  fluorescence wavelength: 569.23 nm at the B3LYP/3-21G\*). Here, we validate the generator, synthesizing this new molecule, **PC**, and measure its photoluminescence (PL) spectra. The synthesis route to **PC** and its characterization are described in the Supplementary Information.



**Figure 4.** The highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) of optimized **PC** at the B3LYP/3-21G\* level. The  $S_1$  state of **PC** represents the transition of one electron from HOMO to LUMO ( $\pi$ - $\pi^*$  excitation) whose excitation (absorption) energy was computed to be 433.75 nm with an OS of 0.2567. On the other hand, the fluorescence was computed to be 569.23 nm with an OS of 0.1869, the nature of orbitals involved in the transition was not changed from that in the case of the absorption.

## 2.2 Experimental measurements

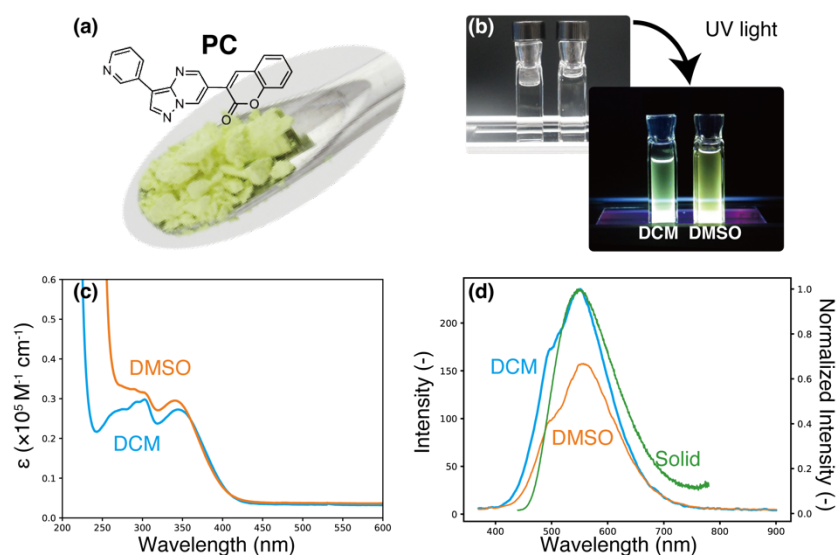
Figure 5(a) shows the image of compound **PC**, that is obtained as a yellow solid material at room temperature. The images of **PC** at a concentration of 0.1-mM in dichloromethane (DCM) and dimethyl sulfoxide (DMSO) solvents under room light and upon the irradiation of UV light (365 nm) are shown in Figure 5(b). The fluorescence (yellow or

green depending on the solvent) of **PC** can be observed by the naked eye. The UV-vis absorption and PL spectra of **PC** in DCM and DMSO are shown in Figures 5 (c) and (d) respectively. We obtained a similar PL spectrum of **PC** in the solid state as those of **PC** in the solutions, as shown in Figure 5 (d). For more information on the dependence of concentration of their spectra, see the Supplementary Information. The UV-vis absorption spectrum of **PC** exhibits peaks with edges beginning at 425 nm, which is consistent with the theoretically predicted wavelength of the excitation energy to the  $S_1$  state. On the other hand, the first peak of the PL spectrum lies at around 550 nm, which is significantly consistent with the theoretically predicted fluorescence from the  $S_1$  state. We could not find any increments in the intensity of PL when bubbled with dry  $N_2$  for 20 min; therefore, it can be concluded that phosphorescence does not contribute to the PL spectrum at room temperature (see the Supplementary Information).

The experimental difference between the absorption and fluorescence spectra was evaluated to be approximately 200 nm, which is comparable to the theoretically predicted shift (135.48 nm). Although the values estimated by static theoretical computation succeeded in reproducing the experimental results, the values involved in molecular dynamics are not the case. The dynamical behavior of the photoinduced molecule is reflected in the value of the quantum yield ( $\Phi$ ); the  $\Phi$  of **PC** measured by an integrating



sphere in DCM was  $0.007 \pm 0.03$ . This small  $\Phi$  indicates that there would be some non-radiative paths to the ground states.<sup>33</sup> In the case of **PC**, the  $S_2$  state is energetically close to the  $S_1$  state at the  $S_1$  minimum by 0.67 eV according to the DFT computation. Hence, there might be a non-radiative way to the  $S_0$  state after the internal conversion from the  $S_1$  state to the  $S_2$  state.



**Figure 5.** (a) Image of **PC** in powdery form. (b) Images of **PC** dissolved in DCM and DMSO (100  $\mu\text{M}$ ) under room light and upon irradiation of UV light (365 nm). (c) UV-vis absorption spectra of **PC** in DCM and DMSO (10  $\mu\text{M}$ , optical length = 1.0 mm, under air). (d) Photoluminescence (PL) spectra ( $\lambda_{\text{ex}} = 350 \text{ nm}$ ) of **PC** in DCM, DMSO (10  $\mu\text{M}$ , under air), and solid state. The intensity of the PL spectrum in solid state is normalized by its maximum value (right vertical axis).

### 3. Discussion

For the several decades, computer-aided chemistry has played a significant role in developing an understanding of several phenomena in chemistry and materials science. However, its work solely focuses on the analysis and speculation of the experimental results; the creative work such as the prediction of various phenomena and designing materials<sup>25,34-38</sup> has rarely been considered. The recent applications of ML algorithms to chemistry and materials science represent a good turning point for computer-aided chemistry and materials science. To automatically design the useful molecules for organic electronics, combining QC with a DNMG is crucial because the molecules that are applicable as materials for organic electronics feature characteristic quantum mechanics properties. However, the innovation of the DNMG based on QC should be proved before its practical use.

In this study, we used a DNMG for creating molecules with properties beyond the technical understanding, i.e., fluorescence. We designed fluorescent molecules under the framework of DFT, which is a type of QC. Although the PES of the fluorescent molecules is simple, we can not create a fluorescent molecule from scratch because the diversity in molecules prevents us from obtaining the correlation between the fluorescence and

molecular structures. However, the generator deals with this diversity and succeeds in designing fluorescent molecules from scratch. The substantial ab initio computations based on QC require massive parallel computations (1024 cores, 5 days); nevertheless, the generator succeeded in producing 3643 candidate molecules. The generator produced molecules that absorbed light with long wavelengths, controlling the conjugate length of molecules in a similar manner as the professional understanding; however, the generator could not find clear correlations between the fluorescence wavelength/intensity of molecules and the conjugate length/number of aromatic rings. This indicates that the difficulties involved in the de novo design of fluorescent molecules.

We selected one molecule considering the detection of its fluorescence from the naked eye and synthesizability. The innovation of the DNMG was proved by synthesizing 3-[3-(pyridin-3-yl)pyrazolo[1,5-a]pyrimidin-6-yl]-2H-chromen-2-one (**PC**) whose fluorescence was evident by the naked eye. Even though it is known that the world of molecules is governed by the rules of quantum mechanics, it is difficult to create a new molecule from scratch with only QC. However, this defect can be compensated by the generator.

This work shows that the generator can be an alternative tool to surpass professional knowledge or intuition. Although the fluorescence of the molecule synthesized in this

work could only be detected by naked eye, practical molecule design would be possible by including the dynamics of photo-induced molecules into the evaluation of the molecule generation. By further developing the QC, more complex functional molecules can be designed. Thus, generators with massive parallel computation would be able to create complex molecules with various intriguing functions, which may result in very complex synthesis routes that would increase the requirements of the recently developed planning retrosynthesis route based on ML.<sup>39-41</sup>

## **4. Method**

### **4.1. Molecular design**

A simple fluorescent molecule is considered as a molecule with a minimum in the excited state with the same spin multiplicity of its ground state. Figure 2 shows a simple PES of the molecule that emitted fluorescence from its  $S_1$  state. Ignoring several factors for quenching fluorescence, we designed the molecules whose PESs are shown in Figure 2. Despite the simple framework, no qualitative guidelines are available for designing fluorescent molecules owing to the diversity of molecules. To address this diversity, a DNMG can be employed. We used the molecule generator, ChemTS<sup>20</sup>, which uses Monte Carlo Tree Search (MCTS)<sup>42</sup> and recurrent neural network (RNN)<sup>43,44</sup> to generate

molecules. Initially, the RNN was trained with a set of SMILES strings.<sup>45</sup> In our case, 153,253 molecules that only contained H, O, N, and C elements, obtained from the ZINC database,<sup>46</sup> were used.

The PESs of the generated molecules were evaluated at the B3LYP/3-21G\* level, implemented in the Gaussian 16 package.<sup>30</sup> The excited energy and the fluorescence wavelength were computed by the time-dependent density functional theory (TD-DFT) at the same level. The lowest 10 states of each molecule were calculated for each minimum in the  $S_0$  and  $S_1$  states. The flowchart of designing fluorescent molecules with ChemTS is illustrated in Figure 6. We used the following reward function to evaluate the absorption and fluoresce with their intensity simultaneously:

$$R(a_w, a_i, f_w, f_s) = W_{a_w} R_{a_w}(a_w) + W_{a_i} R_{a_i}(a_i) + W_{f_w} R_{f_w}(f_w) + W_{f_s} R_{f_s}(f_s), \quad (1)$$

where,  $W_{a_w}, a_i, f_w, f_s$  denote the weights of the respective  $R_{a_w}, a_i, f_w, f_s$ , represented as following:

$$R_{a_w}(a_w) = \exp\left(-\frac{(a_w - T_{a_w})^2}{2\sigma_a^2}\right), \quad (2)$$

$$R_{a_i}(a_i) = \frac{\tanh(\log_{10}(a_i + \epsilon) - \log_{10} T_{a_i})}{2}, \quad (3)$$

$$R_{f_w}(f_w) = \exp\left(-\frac{(f_w - T_{f_w})^2}{2\sigma_w^2}\right), \quad (4)$$

$$R_{f_s}(f_s) = \frac{\tanh(\log_{10}(f_s + \epsilon) - \log_{10} T_{f_s})}{2}. \quad (5)$$

$T_{a_w}$  and  $T_{a_i}$  denote the upper absorption wavelength and its OS, respectively. Similarly,

$T_{fw}$  and  $T_{fi}$  are the upper fluorescence wavelength and its OS, respectively. In this work, we set  $T_{aw}$  and  $T_{fw}$  to 700 and 1200 nm, respectively, and their OS of  $T_{ai}$  and  $T_{fi}$  to 0.01. The absorption and fluorescence wavelengths of the generated molecules are denoted by  $a_w$  with its intensity ( $a_i$ ) and  $f_w$  with its intensity ( $f_i$ ), respectively (See Figure 2).  $\varepsilon$  is an infinitesimal constant to avoid the undefinition of logarithm. In this study, to accelerate the MCTS-based search, we employed a parallelization strategy based on the concept of virtual loss<sup>47</sup>. In the selection step of the MCTS algorithm, we computed the following score for each child node  $i$ :

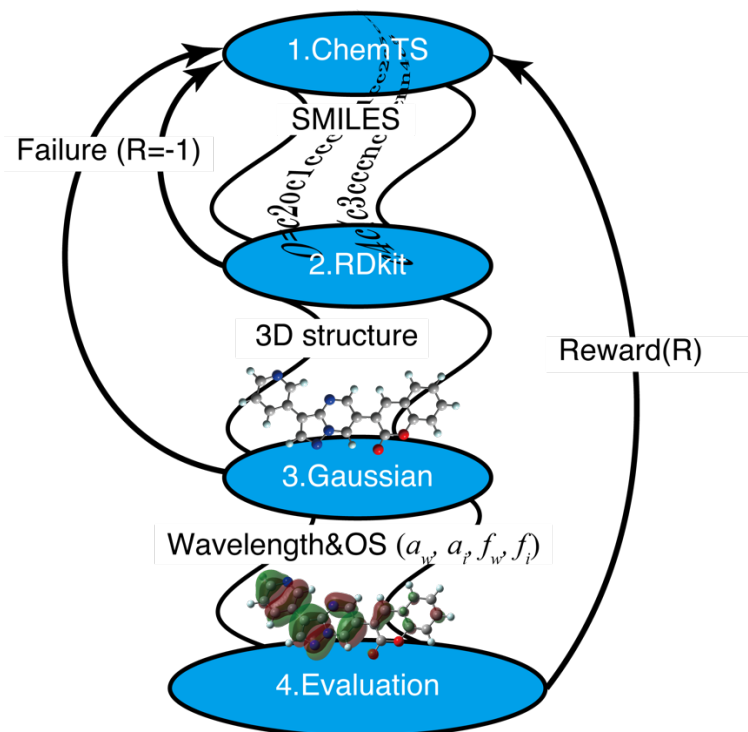
$$S_i = \frac{R_i}{v_i + w_i} + CP_i \sqrt{\frac{v_p + w_p}{1 + v_i + w_i}}. \quad (6)$$

Here,  $R_i$  is the cumulative sum of the above reward  $R(a_w, a_i, f_w, f_i)$  of node  $i$ ;  $v_i$  is the total visit number of node  $i$ ;  $w_i$  is the total virtual visit number (virtual loss) of node  $i$ ;  $C$  is the hyperparameter for search; and  $P_i$  is the selection probability of node  $i$ .  $v_p$  and  $w_p$  denote the total number of visit and total virtual visit numbers of the parent node  $p$ , respectively.

In this study, we set  $C$  to 4.

For geometric analysis, we selected the number of aromatic rings and conjugate length.

For the aromatic rings, we counted them using an Rdkit function.<sup>48</sup> For the conjugate length, we counted a unity that is the sequence of single bond-double bond-single bond (such as -C=C-).



**Figure 6.** Workflow of fluorescent-molecule generation. ChemTS generates the molecules represented in the SMILES string.<sup>45</sup> The molecules in the SMILES strings are translated to a 3D structure via RDkit<sup>48</sup>. The absorption, fluorescence, and molecular OS obtained by Gaussian package were evaluated by the reward function (eq. (1)). ChemTS produced the molecules based on the electronic structure theory iterating this process.

## 4.2. Molecule synthesis and measurement

We selected a new molecule, 3-[3-(pyridin-3-yl)pyrazolo[1,5-a]pyrimidin-6-yl]-2H-chromen-2-one (**PC**), from the molecules that generated by ChemTS for chemical

synthesis from the view point of its easiness of synthesis. Compound **PC** was obtained from Tokyo Chemical Industry Co., Ltd. through custom synthesis. Refer to the Supplementary Information for the detailed synthesis process and its characterization.

The electronic absorption spectra were measured using a Shimadzu UV-3600 UV–vis–NIR spectrophotometer at 20 °C. The emission spectra in solution and solid state were measured by FP-8300 (JASCO, Japan). A quartz cell with an optical length of 1.0 mm was used. Spectroscopic grade solvents (DCM (dichloromethane) and DMSO (dimethyl sulfoxide)) were obtained from Fujifilm Wako Pure Chemical Corporation. Absolute fluorescence quantum yields were determined with a Hamamatsu Photonics C-9920-02 calibrated integrating sphere system.

## **Acknowledgements**

This work is partially supported by a project subsidized by the New Energy and Industrial Technology Development Organization (NEDO) and MEXT as Priority Issue on Post-K Computer" (Building Innovative Drug Discovery Infrastructure through Functional Control of Biomolecular Systems) and Program for Promoting Researches on the Supercomputer Fugaku" (MD-driven Precision Medicine), AMED JP20nk0101111, SIP (Technologies for Smart Bio-industry and Agriculture) and JST ERATO JPMJER1903.



The computations in this work were performed in the supercomputer centers at NIMS and RAIDEN of AIP (RIKEN).

### **Author contributions**

M.S, K. Tsuda, and R.T planned and supervised the project. M.S and K. Terayama performed the computational experiments, K. Terayama and K. Y parallelized ChemTS. N.S selected a molecule from the generated molecules. N.S., S.I., M.K.C, and D.T.P performed the chemical measurements. All members contributed toward preparing the manuscript.

### **Competing interests**

The authors declare no competing financial interest.

### **Data availability**

Correlation graphs of molecular structures with  $S_1$  state relevant molecular properties, a table of selected generated molecules as detectable fluorescent molecules, the synthesis process of **PC** and its characterization by NMR and HRMS, concentration dependence of PL spectra, and PL spectra of **PC** under  $N_2$  are shown in Supplementary Information.

**Code availability**

Our implementation of parallel version of ChemTS for fluorescent molecules is available at [https://github.com/tsudalab/FL\\_ChemTS](https://github.com/tsudalab/FL_ChemTS).

## References

1. Silver, D. Mastering the Game of Go without Human Knowledge. *Nature*, **550**, 354–359 (2017).
2. Goodfellow, I. et al. Generative adversarial nets, Advances in neural information processing systems, 2672-2680 (2014).
3. Briot, J.-P., Hadjeres, G. and Pachet, F. Deep learning techniques for music generation--a survey, arXiv preprint arXiv:1709.01620 (2017).
4. Devlin, J., Chang, M.-W. Lee, K. Toutanova, K. Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 4171-4186 (2019).
5. Terayama, K. *et al.* Pushing property limits in materials discovery via boundless objective-free exploration. *Chem. Sci.* **11**, 5959–5968 (2020).
6. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559**, 547–555 (2018).
7. Cai, J., Chu, X., Xu, K., Li, H., and Wei, J. Machine Learning-driven new material discovery, *Nanoscale Adv.* **2**, 3115-3130 (2020).
8. Homma, K. *et al.* Optimization of a Heterogeneous Ternary  $\text{Li}_3\text{PO}_4\text{--Li}_3\text{BO}_3\text{--Li}_2\text{SO}_4$  Mixture for Li-Ion Conductivity by Machine Learning. *J. Phys. Chem. C* **124**, 12865–12870 (2020).
9. Henson, A. B., Gromski, P. S. & Cronin, L. Designing Algorithms to Aid Discovery by Chemical Robots. *ACS Cent. Sci.* **4**, 793–804 (2018).
10. Granda, J. M., Donina, L., Dragone, V., Long, D. L. & Cronin, L. Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature* **559**, 377–381 (2018).

11. Coley, C. W. *et al.* A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science*. **365**, eaax1566, (2019).
12. Sanchez-lengeling, B. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **365**, 360–365 (2018).
13. Elton, D. C., Boukouvalas, Z., Fuge, M. D. & Chung, P. W. Deep learning for molecular design - A review of the state of the art. *Mol. Syst. Des. Eng.* **4**, 828–849 (2019).
14. Qin, Y., Li, G., Qi T., and Huang, H. Aromatic imide/amide-based organic small-molecule emitters for organic light-emitting diodes. *Mater. Chem. Front.* **4**, 1554–1568 (2020).
15. Cao, D. *et al.* Coumarin-Based Small-Molecule Fluorescent Chemosensors. *Chem. Rev.* **119**, 10403–10519 (2019).
16. Banerjee, S. *et al.* Recent advances in the development of 1,8-naphthalimide based DNA targeting binders, anticancer and fluorescent cellular imaging agents. *Chem. Soc. Rev.* **42**, 1601–1618 (2013).
17. Merk, D., Grisoni, F., Friedrich, L., Shneider, G. Tuning artificial intelligence on the de novo design of natural-product-inspired retinoid X receptor modulators, *Commun. Chem.* **1.1**, 68 (2018).
18. Wu, S. *et al.* Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. *npj Comput. Mater.* **5**, (2019).
19. Kajita, S, Kinjo, T., Nishi, T. Autonomous molecular design by Monte-Carlo tree search and rapid evaluations using molecular dynamics simulations. *Commun. Phys.* **3**, 77 (2020).
20. Yang, X., Zhang, J., Yoshizoe, K., Terayama, K. & Tsuda, K. ChemTS: an efficient python library for de novo molecular generation. *Sci. Technol. Adv. Mater.* **18**, 972–976 (2017).

21. Sumita, M., Yang, X., Ishihara, S., Tamura, R. & Tsuda, K. Hunting for Organic Molecules with Artificial Intelligence: Molecules Optimized for Desired Excitation Energies. *ACS Cent. Sci.* **4**, 1126–1133 (2018).
22. Mamada, M. et al. Highly Efficient Thermally Activated Delayed Fluorescence from an Excited-State Intramolecular Proton Transfer System. *ACS Cent. Sci.* **3**, 769–777 (2017).
23. Uoyama, H., Goushi, K., Shizu, K., Nomura, H. & Adachi, C. Highly efficient organic light-emitting diodes from delayed fluorescence. *Nature* **492**, 234–238 (2012).
24. Kaji, H. et al. Purely organic electroluminescent material realizing 100% conversion from electricity to light. *Nat. Commun.* **6**, 8476 (2015).
25. Izawa, H. et al. Novel 1,8-Naphthalimide derivative with an open space for an anion: Unique fluorescence behaviour depending on the binding anion's electrophilic property. *Chem. Commun.* **51**, 8596–8599 (2015).
26. Yeung, M. C.-L. & Yam, V. W.-W. Luminescent cation sensors: from host–guest chemistry, supramolecular chemistry to reaction-based mechanisms. *Chem. Soc. Rev.* **44**, 4192–4202 (2015).
27. Gale, P. a & Caltagirone, C. Anion sensing by small molecules and molecular ensembles. *Chem. Soc. Rev.* **44**, 4212–4227 (2015).
28. Carr, J. A. et al. Shortwave infrared fluorescence imaging with the clinically approved near-infrared dye indocyanine green. *Proc Natl Acad Sci USA.* **115**, 4465–4470 (2018).
29. Chouket, R. et al. Dynamic contrast with reversibly photoswitchable fluorescent labels for imaging living cells. *Chem. Sci.* **11**, 2882–2887(2020).
30. Gaussian 16, Revision C.01, Frisch, M. J. et al. Gaussian, Inc., Wallingford CT, 2016.

31. Parr, R. G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, 1989
32. *SciFinder*; Chemical Abstracts Service: Columbus, OH (accessed Sep 13, 2020); <https://scifinder.cas.org>.
33. Domcke, M., Yarkony, D. R., Köppel, H., *Conical Intersections: Electronic Structure, Dynamics & Spectroscopy*, (World Scientific Publishing Co. Pte. Ltd. 2004).
34. Struebing, H. *et al.* Computer-aided molecular design of solvents for accelerated reaction kinetics. *Nat. Chem.* **5**, 952–957 (2013).
35. Wang, L. P. *et al.* Discovering chemistry with an ab initio nanoreactor. *Nat. Chem.* **6**, 1044–1048 (2014).
36. Zhang, S. *et al.* Penta-Graphene: A New Carbon Allotrope. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 2372–2377 (2015).
37. Sumita, M., Tanaka, Y., Ikeda, M. & Ohno, T. Theoretically designed Li<sub>3</sub>PO<sub>4</sub> (100)/LiFePO<sub>4</sub> (010) coherent electrolyte/cathode interface for all solid-state li ion secondary batteries. *J. Phys. Chem. C* **119**, 14–22 (2015).
38. Sumita, M., Tamura, R., Homma, K. & Tsuda, K. Li-Ion Conductive Li<sub>3</sub>PO<sub>4</sub>-Li<sub>3</sub>BO<sub>3</sub>-Li<sub>2</sub>SO<sub>4</sub> Mixture: Prevision through Density Functional Molecular Dynamics and Machine Learning. *Bull. Chem. Soc. Jpn.* **92**, 1100–1106 (2019).
39. Segler, M. H. S., Preuss, M. & Mark, P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature*, **555**, 604–610 (2018).
40. Coley, C. W., Barzilay, R., Jaakkola, T. S., Green, W. H. & Jensen, K. F. Prediction of Organic Reaction Outcomes Using Machine Learning. *ACS Cent. Sci.* **3**, 434–443 (2017).
41. Shibukawa, R., *et al.* CompRet: a comprehensive recommendation framework for chemical synthesis planning with algorithmic enumeration. *J. Cheminform.* **12**, 52

(2020)

42. Browne, C., Powley, E., and Whitehouse, D. A Survey of Monte Carlo Tree Search Methods. *IEEE Transactions on Computational Intelligence and AI in Games*, **4**, 1–43 (2012).
43. Hochreiter, S., Schmidhuber, J. Long Short-Term Memory. *Neural Computation*, **9**, 1735–1780 (1997).
44. Cho, K., Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014*, pp 1724–1734 (2014).
45. Weininger D. SMILES a chemical Language and information system 1. Introduction to methodology and encoding rules, *J. Chem. Inf. Model.* **28**, 31-36, 1988.
46. Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., and Coleman, R. G. ZINC: A Free Tool to Discover Chemistry for Biology *J. Chem. Inf. Model.* **52**, 1757–1768 (2012).
47. Browne CB, Powley E, Whitehouse D, et al. A survey of monte carlo tree search methods. *IEEE Trans Comput Intell AI Games.* **4**, 1–43, 2012.
48. Landrum, G. RDKit: Open-source cheminformatics. <http://www.rdkit.org> (accessed Nov 7, 2017).

## Supplementary Information

for

### De novo creation of a naked-eye-detectable fluorescent molecule based on quantum-chemical computation and machine learning

**Masato Sumita,<sup>1,2\*</sup> Kei Terayma,<sup>1,3,4,5</sup> Naoya Suzuki,<sup>6</sup> Shinsuke Ishihara,<sup>2</sup> Ryo Tamura,<sup>2,7,8</sup> Mandeep K. Chahal,<sup>2</sup> Daniel T. Payne,<sup>2,9</sup> Kazuki Yoshizoe,<sup>1</sup> Koji Tsuda<sup>1,7,8\*</sup>**

<sup>1</sup>*RIKEN Center for Advanced Intelligence Project, 1-4-1 Nihonbashi, Chuo-ku, Tokyo 103-0027, Japan.*

<sup>2</sup>*International Center for Materials Nanoarchitectonics (WPI-MANA), National Institute for Materials Science, 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan.*

<sup>3</sup>*Graduate School of Medical Life Science, Yokohama City University, 1-7-29, Suehiro-cho, Tsurumi-ku, 230-0045 Kanagawa, Japan*

<sup>4</sup>*Graduate School of Medicine, Kyoto University, 53 Shogoin-Kawaharacho, Sakyo-ku, 606-8507 Kyoto, Japan*

<sup>5</sup>*Medical Sciences Innovation Hub Program, RIKEN Cluster for Science, Technology and Innovation Hub, Tsurumi-ku, Kanagawa 230-0045, Japan.*

<sup>6</sup>*Materials Science and Engineering, Osaka Prefecture University, 1-1 Gakuen-cho, Nakaku, Sakai, Osaka 599-8531, Japan.s*

<sup>7</sup>*Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, 1-2-1 Sengen, Tsukuba, Ibaraki 305-0047, Japan.*

<sup>8</sup>*Graduate School of Frontier Sciences, The University of Tokyo, 5-1-5 Kashiwa-no-ha, Kashiwa, Chiba 277-8561, Japan.*

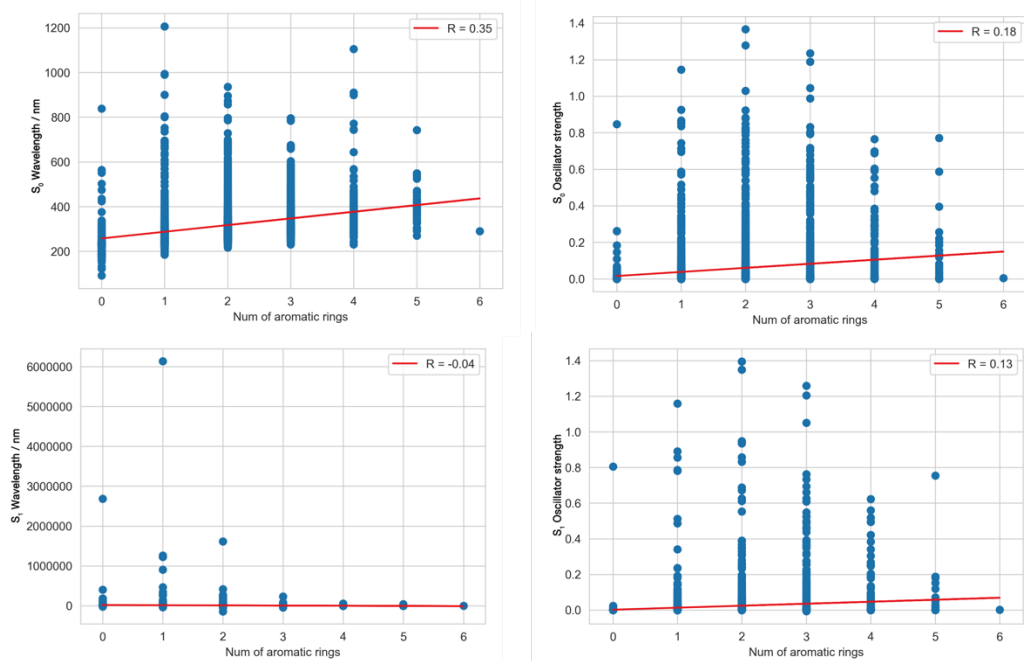
<sup>9</sup>*International Center for Young Scientists (ICYS), National Institute for Materials Science, 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan.*

Correspondence and requests for materials should be addressed to M. S. (email: masato.sumita@riken.jp) or to K. T. (email: tsuda@k.u-tokyo.ac.jp)



## 1. Correlation with number of aromatic rings

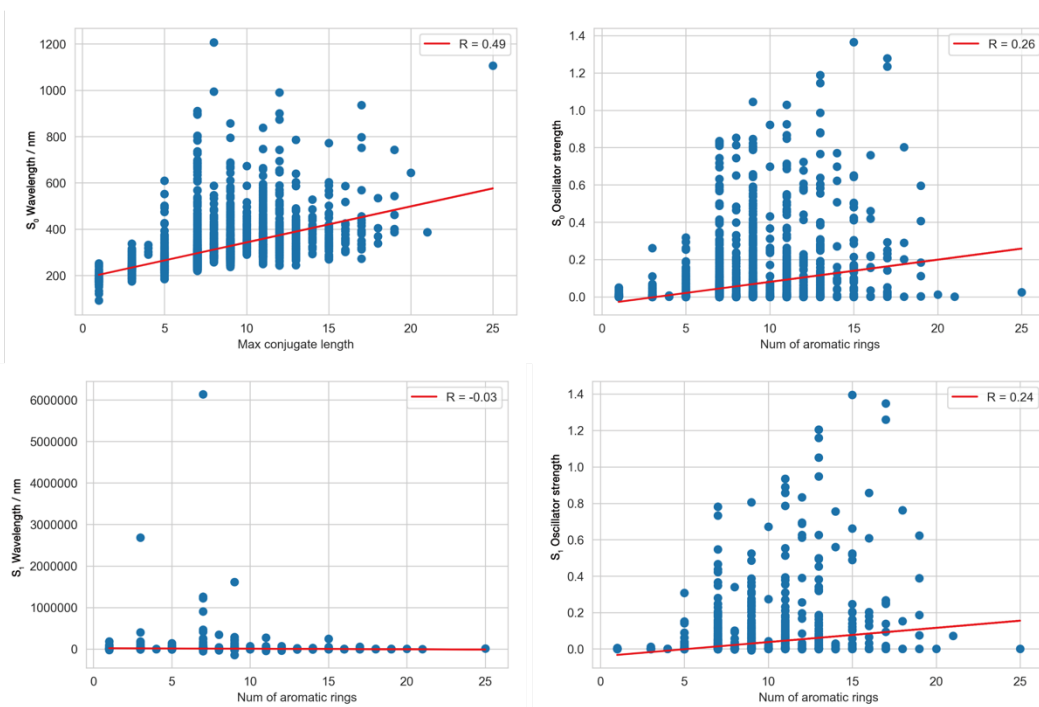
To see the correlations of absorption wavelength to  $S_1$  states/emission wavelength from  $S_1$  states and its oscillator strengths (OSs) with number of aromatic rings, the correlation graphs with them are shown in Figure S1. We used the tools of RDkit to count aromatic rings.



**Figure S1.** Correlation graphs of number of aromatic rings with  $S_0$  absorption wavelength, its Oscillator strength (upper two). Correlation graphs with  $S_1$  emission wavelength, its oscillator strength (bottom two).

## 2. Correlation with conjugate length

To see the correlations of absorption wavelength to  $S_1$  states/emission wavelength from  $S_1$  states and its oscillator strengths (OSs) with conjugate length, the correlation graphs with them are shown in Figure S1. We have count conjugate length whoes unity is defined as single-double-single bond sequence.

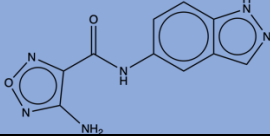
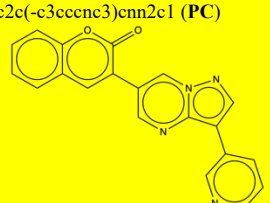


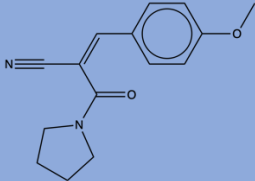
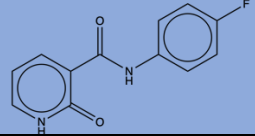
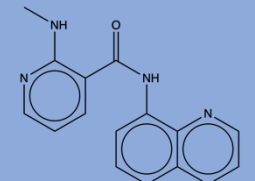
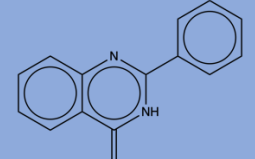
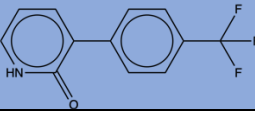
**Figure S2.** Correlation graphs of number of conjugate lengths with S<sub>0</sub> absorption wavelength, its Oscillator strength (upper two). Correlation graph with S<sub>1</sub> emission wavelength, its oscillator strength (bottom two).

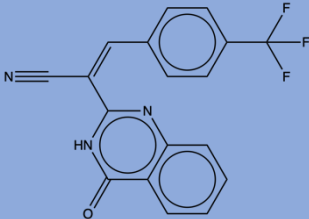
### 3. Selected molecules

From the viewpoint of detectability, we have selected 87 molecules in accordance with the condition mentioned in main text. In Table S1. Selected 87 molecules in SMILES string are summarized with their properties computed at the B3LYP/3-21G\* level. The molecules in blue shadow cells are known molecules or its tautomers that are found in Scifinder. The molecule we synthesize in this study are shown in yellow cells.

**Table S1. Selected 87 molecules and their properties from the viewpoint of detectability**

Species in SMILES	S <sub>0</sub> absorption /nm	S <sub>0</sub> OS	S <sub>1</sub> emission /nm	S <sub>1</sub> OS
<chem>Nc1nc(C(=O)Nc2ccc2[nH]nc2c1</chem> 	362.13	0.2032	533.04	0.1537
<chem>Cn1c(C(=O)Nc2ccc(F)c2F)nc2ccccc21</chem>	305.41	0.5429	445.63	0.3797
<chem>N/C=C/c1c(O)cccc1O</chem>	268.87	0.2485	516.14	0.0447
<chem>Nc1n[nH]nc1-c1ccccc1</chem>	293.59	0.1543	497.03	0.0128
<chem>Cn1cc(C(=O)Nc2nnc3ccccc32)c2ccccc12</chem>	338.71	0.5523	446.5	0.3039
<chem>Nc1c(-c2n[nH]c3ccccc23)cccc1C(F)(F)F</chem>	352.03	0.1669	459.16	0.0638
<chem>C/C(=C/C(=O)c1cnc(C(=O)Nc2ccc2nc(C)ccc2c1</chem>	409.79	0.1958	565.32	0.2712
<chem>O=c1n2nc(-c3ccccc3)oc2cc2cc(N3CCCC3)cc(C(F)(F)F)c12</chem>	409.37	0.1185	654.78	0.0283
<chem>O=c1[nH]cc(C(=O)Nc2ccc(O)nc2)c1CCCCC2</chem>	325.05	0.4937	440.26	0.0191
<chem>C/C(=C/C1ccc(C(C)C)cc1)O[C@@H](C)C1CC1</chem>	253.04	0.3179	1238.28	0.0265
<chem>CN1N=C(C(=O)Nc2ccc(NC(=O)c3ccccc3)c2)CC1=O</chem>	358.49	0.1338	600.06	0.0491
<chem>CNc1nc(-c2c(F)cccc2F)ccc1C(=O)N1CCCCC1</chem>	381.54	0.177	504.92	0.083
<chem>NN[C@H](CCCC1=C2ccccc2=NC1)C(=O)Nc1ccccc1F</chem>	377.65	0.1293	769.99	0.0167
<chem>O=C(c1ccc(NC(=O)Nc2ccc(N3CCCC3)c2)cc1)c1ccccc1</chem>	459.94	0.1315	711.83	0.0974
<chem>Cc1cc(C)c(C#N)c(=O)n1Cc1ncc(C)n1</chem>	307.09	0.1892	452.59	0.0233
<chem>O=c1ccc2c3ccccc3nc2cc1C</chem>	360.29	0.1657	549.85	0.0566
<chem>CCn1cc(C(=O)Nc2ccc(C)cc2)c2c(O)nn(C)c21</chem>	379.41	0.2324	643.49	0.028
<chem>O=C1N(CCC(F)(F)F)c2c(ccn2-c2ccc(F)cc2)/C1=C/C1ccc(F)c1</chem>	433.58	0.148	841.18	0.034
<chem>Fe1cn(CN2CCCC[C@@H](2)CO)nc1/N=C1/C=C(c2ccccc2)OC1</chem>	368.1	0.6382	3873.68	0.0197
<chem>O=c1oc2ccccc2cc1-c1cnc2c(-c3ccccc3)cn2c1 (PC)</chem> 	433.75	0.2567	569.23	0.1869
<chem>Fe1ccc(-c2ccccc2Nc2nnc2Cc2ccccc2)o1</chem>	308.57	0.1606	472.37	0.0571
<chem>C=Cc1nn(C)cc1NC(=O)c1cncn2cnn2c1</chem>	509.78	0.1333	692.66	0.1415
<chem>Fe1cc2cenn2c(-c2ccccc2)n1</chem>	346.52	0.1708	473.54	0.101
<chem>Cn1c2cc([C@H](CC(=O)NNC(=O)[C@@H](3C[C@@H](3C)nc3ccccc3)c2ccccc21</chem>	455.28	0.2315	646.3	0.1377
<chem>O=c1nc(NCc2ccccc(C(F)(F)F)c2)cn1-c1ccc(F)cc1</chem>	348.16	0.3058	537.98	0.1568
<chem>C#Cc1ncc2c3ccccc3n(CC(=O)N3CC[C@H](C)C[C@H](3c3ccccc3)cc12</chem>	440.25	0.1567	752.37	0.0492
<chem>CCn1c(C)nc2cc(NC(=O)c3cc(C)no3)ccc21</chem>	343.37	0.3214	473.29	0.3479
<chem>CCO[C@@H](1C[C@@H](NC(=O)c2ccc(-n3cc(C(=O)NC(C)C)ccc3=O)cc2)C12CCCCC2</chem>	321.15	0.2329	456.86	0.0258
<chem>C/C(=C/C1NC(=O)c1ccccc1)c1ccccc(C#N)c1</chem>	323.77	0.5903	523.36	0.3448
<chem>COc1nn(CC)cc1CNc1ccc(-c2ccccc2)cc1</chem>	289.02	0.6073	570.76	0.0124
<chem>Nc1n[nH]nc1-c1ccc(C(=O)NC2CC2)cc1</chem>	305.51	0.2863	455.27	0.0201
<chem>COc1ccc(C(=O)Nc2ccc(C(=O)N2CCCC2)cc1</chem>	324.84	0.2629	663.0	0.0134

				
<chem>Nc1cc(NC(=O)c2ccc3c(c2)OCCO3)nn1-c1ccnc1</chem>	309.23	0.2006	494.29	0.0925
<chem>N#C/C1=C/C(=C/c2ccc(O)cc2)C(C#N)=C(N)O1</chem>	426.86	0.2094	658.55	0.1152
<chem>CC[C@H](CCO)CNe1enn(-c2ccc(F)cc2)c(=O)c1C</chem>	318.35	0.2245	472.03	0.1775
<chem>CCc1c(/C=C/c2cccnc2)[C@H](C(=O)O)[C@H](C)N=C1N</chem>	383.35	0.1229	754.33	0.0231
<chem>Cc1nc(N2CCN(C[C@H]3CCCCN3CCC)CC2)c2nc[nH]c2c1</chem>	279.67	0.1966	598.48	0.0201
<chem>O=C(N/N=C/c1ccccc1)c1nc(C2CC2)n2ccccc12</chem>	347.02	0.2481	458.5	0.0884
<chem>O=C([C@H]1Cc2ccccc2CN1C(=O)C1=NN(c2ccccc2)C(=O)CC1)N1CCOCC1</chem>	322.73	0.1192	873.67	0.017
<chem>O=c1c(-c2nc(C3CC3)no2)cc(O)c2ccn(-c3ccc(F)cc3)nc12</chem>	594.94	0.2428	913.23	0.0711
<chem>C#CCN(CC#C)C(=O)c1ccccc1Nc1cccc(C)cc1</chem>	403.92	0.1074	715.75	0.0409
<chem>O=C(N/N=C/c1enn(-c2ccccc2)c1)c1ccccc1N</chem>	319.96	0.7927	792.72	0.0513
<chem>NC(=O)[C@H]1N=c2ccccc2=C1NC(=O)c1ccccc1</chem>	462.55	0.2392	921.19	0.0364
<chem>C/C(=N#Nc1ccccc2nnc(C)nc21)c1ccc(O)cc1</chem>	509.06	0.142	781.58	0.0807
<chem>Fe1cc2c(-c3ccccc3)nc3c(-c4ccc(F)cc4)enn3c2cc1F</chem>	425.74	0.2203	601.32	0.1523
<chem>Nc1c(-c2ccccc2)[nH]c2c(C(F)(F)F)ccnc12</chem>	420.08	0.1008	662.06	0.029
<chem>NC(=O)[C@H]1N=C(c2ccc(F)cc2)CO1</chem>	317.4	0.1103	544.05	0.0402
<chem>N#C/C1=C(c2ccccc2)Nc2nc3ccccc3n2C1</chem>	374.41	0.1079	500.62	0.1426
<chem>O=C(Cc1ccccc1)NCCCn1cnc2c(cnn2-c2ccccc2)c1=O</chem>	286.54	0.3421	402.98	0.1914
<chem>Nc1nccc(Nc2ccccc2C#N)c1[C@H]1CCCCO1</chem>	312.19	0.1478	431.48	0.0269
<chem>O=c1[nH]c(Oc2cccc(F)c2)c2ccccc2c1C(F)(F)F</chem>	409.86	0.1207	536.67	0.078
<chem>O=c1[nH]cnc(Oc2ccc(F)cc2)c1C(F)(F)F</chem>	312.5	0.1484	565.67	0.0505
<chem>O=c1c(-c2ccccc2)cc(NCC(=O)N(c2ccccc2)C2CCCC2)co1</chem>	388.56	0.1264	495.55	0.1068
<chem>CC[C@H]1c2ccccc2C(=O)N1Cc1nc(-c2ccc(N)cc2)no1</chem>	302.58	0.2274	506.8	0.0202
<chem>O=c1[nH]cccc1C(=O)Nc1cccc(F)cc1</chem>	386.74	0.1526	564.63	0.1585
				
<chem>CNc1ncccc1C(=O)Nc1cccc2ccnc12</chem>	376.23	0.1231	577.62	0.0457
				
<chem>O=c1ccc(-c2ccccc2)nc2c1-c1ccccc1-2</chem>	586.56	0.1608	986.1	0.0648
<chem>N#CCn1cc(Nc2nccc3c2enn3-c2ccccc2F)c(-c2ccccc2)c1</chem>	327.66	0.3947	473.03	0.118
<chem>O=c1[nH]c(-c2ccccc2)nc2ccccc12</chem>	303.03	0.4188	523.4	0.1267
				
<chem>O=c1[nH]c(O[C@H]2COCCO2)c2ccccc2c1C(F)(F)F</chem>	410.09	0.1083	541.93	0.0609
<chem>O=c1n(c2ccc3c(c2)OCCCO3)nc[nH]c1=O</chem>	347.48	0.1642	479.25	0.0762
<chem>O=C(/C=C/C1CCCC1)NNC(=O)c1cccc(=O)n(-c2ccccc2)n1</chem>	350.84	0.2809	464.08	0.1964
<chem>N/C(=N/Nc1cnc2ccc(F)cc2n1)c1cccc2ccccc2c1</chem>	419.6	0.1922	594.91	0.0382
<chem>Cc1nn2c(nnc3c(=O)n(Cc4ccc(F)cc4)ene32)c1-c1ccc(F)cc1</chem>	470.16	0.1353	816.86	0.0574
<chem>COc1ccc(-n2nc(C)c(CNC(=O)[C@H]3CCCN3C(=O)[C@H](C)NC)c2C)cc1</chem>	256.59	0.1149	481.11	0.119
<chem>O=c1[nH]cccc1-c1ccc(C(F)(F)F)cc1</chem>	322.04	0.3595	431.63	0.0627
				
<chem>O=C(/C=C/c1ccc(O)ccc1O)Nc1cccc2ccccc12</chem>	368.37	0.4556	528.58	0.3935
<chem>Fe1cc2nc3c(nn2c1)CCN(Cc1ccccc1)CN3</chem>	380.55	0.1032	553.59	0.0334
<chem>N#C/C(=C/c1ccc(C(F)(F)F)cc1)c1nc2ccccc2c(=O)[nH]1</chem>	369.71	0.2901	558.63	0.1524

				
Fe1c(/N=C/c2cc3ccccc3c2=O)ccc2c1C(=O)c1ccccc1-2	414.87	0.2923	521.43	0.2544
Ne1[nH]c(CN2CCCC2)c(C(=O)NC[C@@H]2CCCO2)c1N	280.25	0.1087	540.41	0.0308
Ne1ccc(-c2cc(C(=O)O)ncn2)cc1	361.57	0.2141	469.22	0.0841
O[C@H](CCC1=C2CCCC2=NC1)c1enn(-c2ccccc2)c1	376.64	0.1375	759.79	0.0184
Fe1cc2c3nc(CN4CCCC4)cc(-c4ccccc4)n3nc2cc1F	401.36	0.1115	571.98	0.0748
O=C(N/N=C/C1=C2CCCC2=NC1)N1CCN(C(=O)C2CCCCC2)CC1	435.92	0.3699	681.72	0.0982
O=c1c(O[C@@H]2C[C@H]3CC[C@@H]2C3)c(-c2ccccc2)[nH]1	435.51	0.111	1075.27	0.0396
COCc1ccccc1NC(=O)Nc1ccc(N(C)C(C)=O)cc1	284.57	0.2614	540.27	0.0141
O=c1[nH]cnc(N2C[C@H]3CCCC[C@H]3C2)c1-c1ccccc1	321.31	0.1969	463.4	0.1498
O=C(O)c1cccc(C(=O)Nc2ccnc2F)c1	297.84	0.1078	452.43	0.1015
Cc1ccc(/C=N/c2n(C)c(=O)c3c(F)c(F)cc(F)c3n2)c(F)c1	388.84	0.4068	520.85	0.3885
N/C(N/N=C/c1nc(-c2ccccc2)nc2cccn12)c1c[nH]c2ccccc12	422.06	0.1191	576.42	0.0704
O=c1[nH]c(/C=N/c2nnc(-c3ccccc3)o2)cn1C	399.51	0.6497	776.78	0.1075
Ne1nn(/NCc2ccccc2)c(=O)c2cn[nH]c12	342.04	0.1182	617.38	0.0277
O[C@H](CCC1=C2CCCC2=NC1)c1ccc(F)cc1F	375.84	0.1275	814.67	0.0156
COc1cc(-c2nc3c(C)ccccc3c(=O)[nH]2)c2c(C(F)F)cc(=O)n(Cc3ccccc3)c2c1	350.0	0.1558	466.03	0.1805
O=c1n2cnc(-c3ccccc3F)nc2cc(N2CCc3ccccc3C2)c1-c1ccc(F)cc1	461.86	0.1845	610.14	0.1861
O=C(c1c[nH]nc1-c1ccc2c(c1)OCO2)N1CCOC2(CCCCC2)C1	284.99	0.1557	440.55	0.074

## 4. Experimental Details

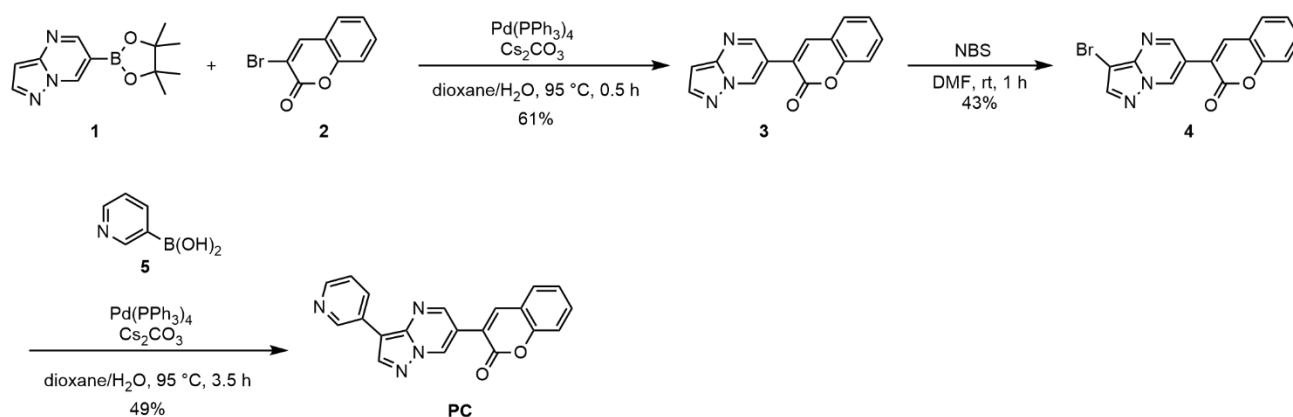
### 4.1. Materials

3-[3-(Pyridin-3-yl)pyrazolo[1,5-a]pyrimidin-6-yl]-2H-chromen-2-one (PC) was obtained from Tokyo Chemical Industry Co., Ltd. through custom synthesis. Spectroscopic grade solvents (CH<sub>2</sub>Cl<sub>2</sub> and DMSO) were obtained from Fujifilm Wako Pure Chemical Corporation.

### 4.2. General methods

ATR-FTIR spectra were obtained using a Thermo-Nicolet 760X FTIR spectrophotometer equipped with a SMART-iTX ATR accessory. <sup>1</sup>H-NMR spectra were obtained using a JEOL JNM-ECA400 spectrometer operating at 400 MHz and using tetramethylsilane (TMS) as an internal standard. Proton decoupled <sup>13</sup>C-NMR spectra were obtained using a JEOL JNM-ECA400 spectrometer operating at 101 MHz and using TMS as an internal standard. Data was processed using Delta version 5.0.5.1. <sup>1</sup>H NMR chemical shifts (δ) are reported in ppm relative to TMS in DMSO-*d*<sub>6</sub> (δ = 0.00). <sup>13</sup>C NMR chemical shifts (δ) are reported in ppm relative to the solvent reported. Coupling constants (*J*) are expressed in Hertz (Hz), shift multiplicities are reported as singlet (s), doublet (d), triplet (t), quartet (q), double doublet (dd), multiplet (m) and broad singlet (bs). High resolution ESI-MS mass spectra were measured using a Thermo Scientific Q-Exactive Plus instrument in methanol with 0.1% formic acid.

### 4.3. Synthesis



**Scheme 1.** Synthesis of 3-[3-(Pyridin-3-yl)pyrazolo[1,5-a]pyrimidin-6-yl]-2H-chromen-2-one (**PC**).

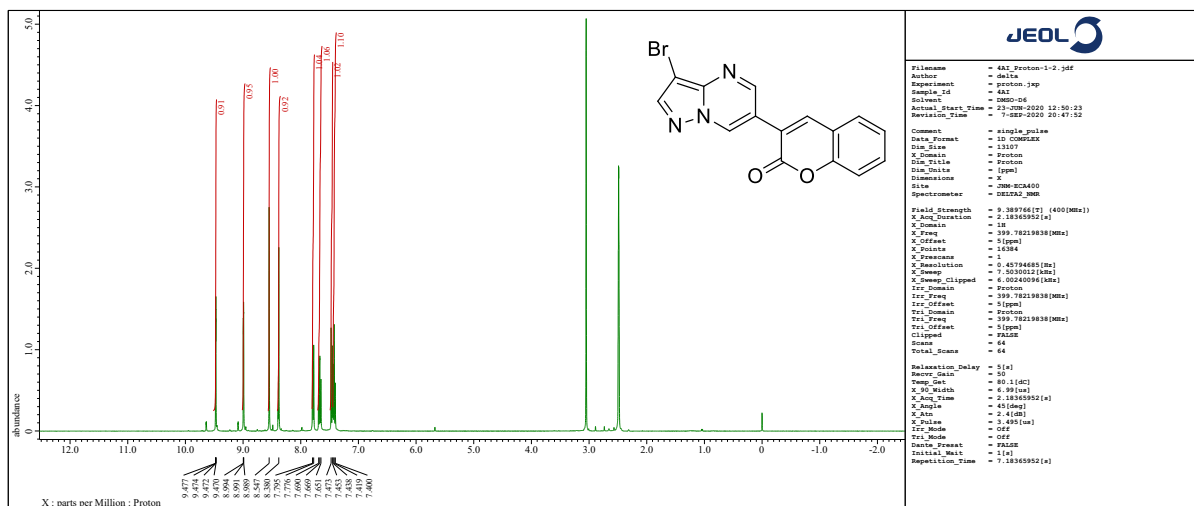
Synthesis of **3**. 6-(4,4,5,5-Tetramethyl-1,3,2-dioxaborolan-2-yl)pyrazolo[1,5-a]pyrimidine (**1**) (0.80 g, 3.3 mmol), 3-bromo-2-chromenone (**2**) (0.74 g, 3.3 mmol), tetrakis(triphenylphosphine)palladium(0) (188 mg, 0.163 mmol), and  $\text{Cs}_2\text{CO}_3$  (3.2 g, 9.8 mmol) were dispersed in a mixture of 1,4-dioxane (16 mL) and water (4 mL), and then stirred at 95 °C for 30 min under  $\text{N}_2$ . After cooling to room temperature, water (20 mL) was added. Product was extracted with ethyl acetate (20 mL) three times. The organic layer was washed with brine (10 mL), and dried over anhydrous  $\text{Na}_2\text{SO}_4$ . After filtration, solvents were removed under reduced pressure to yield a crude solid (1.1 g). The crude solid was washed with diisopropyl ether, and yellow solid **3** was obtained (0.52 g, 2.0 mmol, Yield 61%). Compound **3** was used in the next step without further purification.

Synthesis of **4**. *N*-Bromosuccinimide (0.35 g, 2.0 mmol) was added to compound **3** (0.52 g, 2.0 mmol) dissolved in *N,N*-dimethylformamide (33 mL) at 0 °C. The solution was allowed to warm to room temperature, and stirred for 1 hour. Precipitated solid was collected by filtration, and washed with diisopropyl ether. Thus, compound **4** was obtained as a yellow solid (0.29 g, 0.85 mmol, Yield 43%).  $^1\text{H}$ -NMR (400 MHz,  $\text{DMSO}-d_6$ ):  $\delta$  = 9.47 (m, 1H), 8.99 (m, 1H), 8.55 (s, 1H), 8.38 (s, 1H), 7.79 (d,  $J$  = 8.0 Hz, 1H), 7.67 (t,  $J$  = 7.7 Hz, 1H), 7.46 (d,  $J$  = 8.0 Hz, 1H), 7.42 (t,  $J$  = 7.7 Hz, 1H) ppm;  $^{13}\text{C}$  NMR (101 MHz,  $\text{DMSO}-d_6$ ):  $\delta$  = 160.0, 153.7, 151.4, 146.1, 144.4, 142.1, 135.9, 132.9, 129.4, 125.4, 120.9, 119.7, 117.5, 116.6, 84.0 ppm; FT-IR(ATR):  $\nu$  = 3128, 3060, 1697, 1679, 1606  $\text{cm}^{-1}$ ; HRMS (ESI-MS, methanol with 0.1% formic acid); calculated for  $[\text{C}_{15}\text{H}_9\text{N}_3\text{O}_2^{79}\text{Br}]$  = 341.9873, found: = 341.9884.

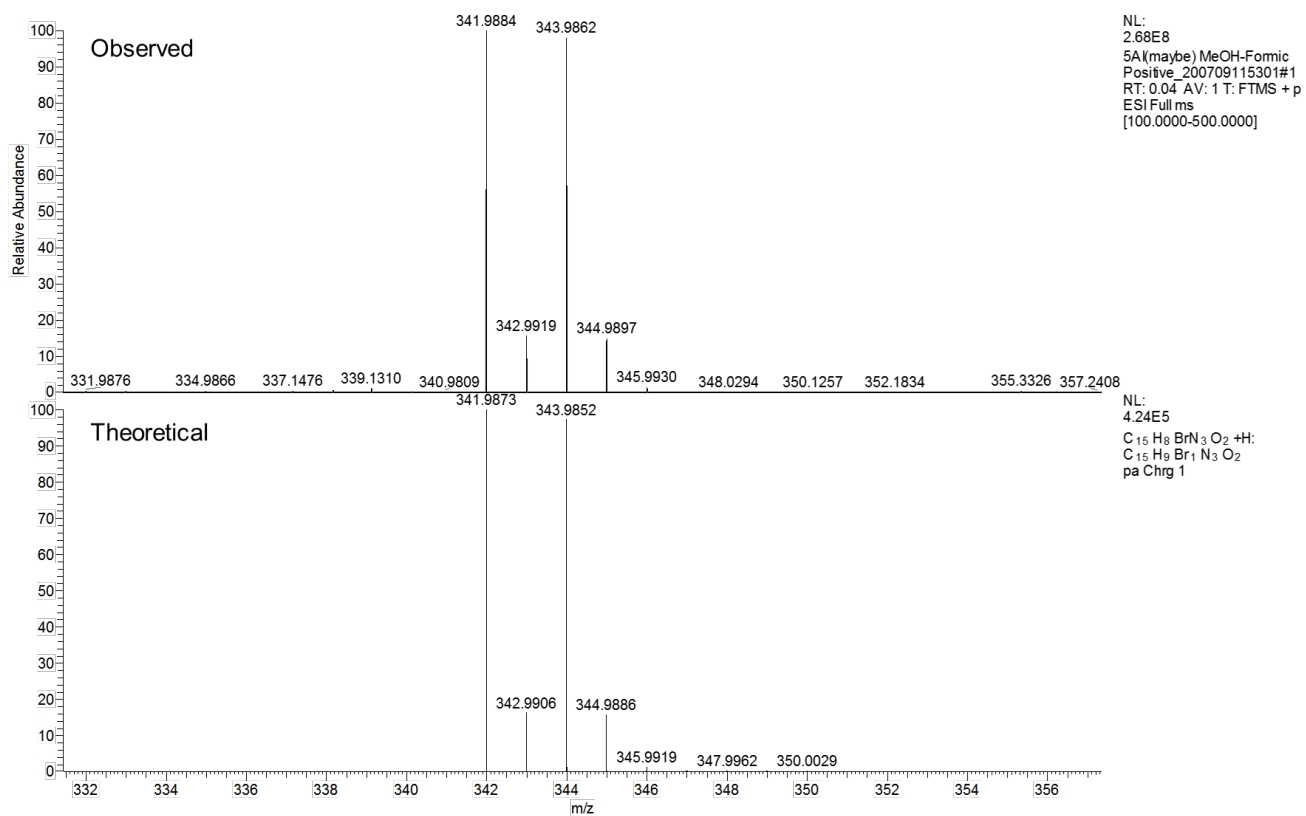
Synthesis of **PC**. Compound **4** (0.29 g, 0.85 mmol), 3-pyridylboronic acid (**5**) (0.40 g, 3.3 mmol), tetrakis(triphenylphosphine)palladium(0) (0.14 g, 0.12 mmol), and  $\text{Cs}_2\text{CO}_3$  (0.57 g, 1.7 mmol) were dispersed in a mixture of 1,4-dioxane (12 mL) and water (1.5 mL), and then stirred at 95 °C for 3.5 hour under  $\text{N}_2$ . After cooling, precipitated solid was collected by filtration, and then washed with ethyl acetate and methanol. The crude product was purified by medium pressure column chromatography (neutral silica gel 10 g,  $\text{CH}_2\text{Cl}_2/\text{Methanol}$  = 99/1 ~ 0/100), and the solid product was washed with methanol. Thus, compound **PC** was obtained as a yellow solid (0.14 g, 0.41 mmol, Yield 49%).  $^1\text{H}$ -NMR (400 MHz,  $\text{DMSO}-d_6$ ):  $\delta$  = 9.52 (d,  $J$  = 1.8 Hz, 1H), 9.32 (s, 1H), 9.06 (d,  $J$  = 2.3 Hz, 1H), 8.84 (s, 1H), 8.57 (s, 1H), 8.48-8.46 (m, 2H), 7.80 (dd,  $J$  = 7.8, 1.4 Hz, 1H), 7.70-7.66 (m, 1H), 7.48-7.41 (m, 3H) ppm;  $^{13}\text{C}$  NMR (101 MHz,  $\text{DMSO}-d_6$ ):  $\delta$  = 159.1, 152.8, 150.1, 146.6, 146.5, 143.3, 143.1, 140.9, 134.9, 132.3, 131.9, 128.4, 127.3, 124.4, 123.2, 120.1, 118.8, 116.2, 115.6, 106.2 ppm; FT-IR(ATR):  $\nu$  = 3037, 1714, 1625, 1606  $\text{cm}^{-1}$ ; HRMS (ESI-MS, methanol with 0.1% formic acid); calculated for  $[\text{C}_{20}\text{H}_{13}\text{N}_4\text{O}_2]$  = 341.1033, found: = 341.1043.

**Table S2.** Summary of HRMS results.

Compound	Molecular Formula	Observed Ion	Theoretical Mass	Observed Mass	Error (ppm)
<b>4</b>	C <sub>15</sub> H <sub>9</sub> N <sub>3</sub> O <sub>2</sub> <sup>79</sup> Br	[M+H] <sup>+</sup>	341.9873	341.9884	3.24
<b>PC</b>	C <sub>20</sub> H <sub>13</sub> N <sub>4</sub> O <sub>2</sub>	[M+H] <sup>+</sup>	341.1033	341.1043	2.86

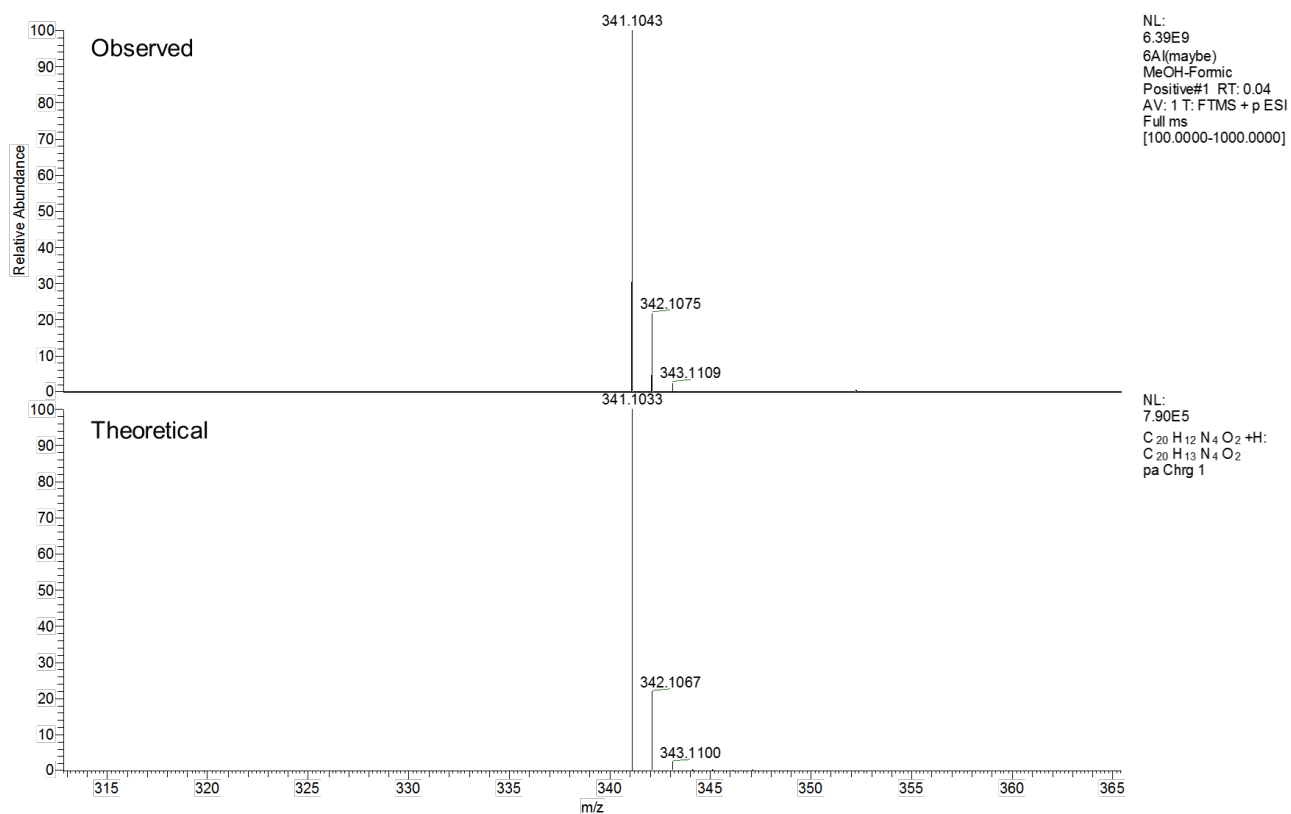






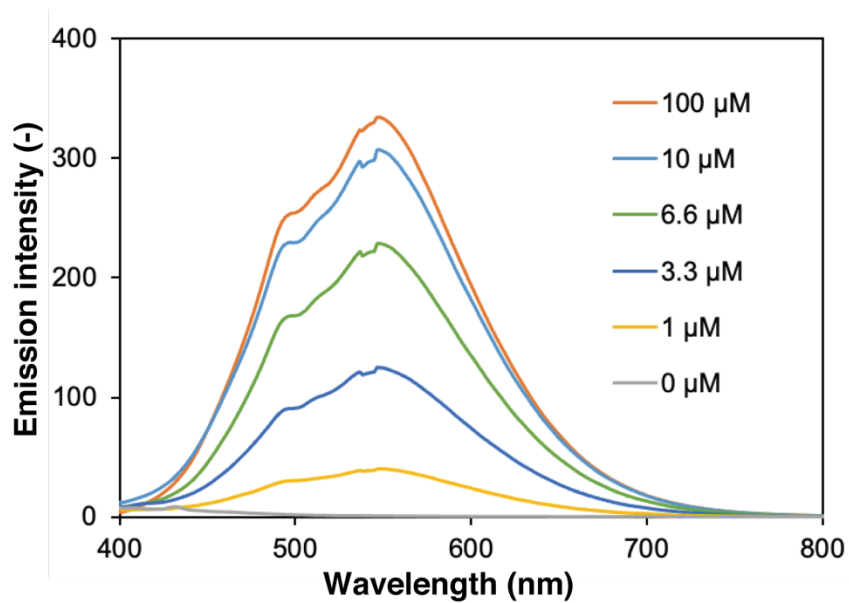
**Figure S5.** HRMS analysis of compound **4**.





**Figure S8.** HRMS analysis of compound **PC**.

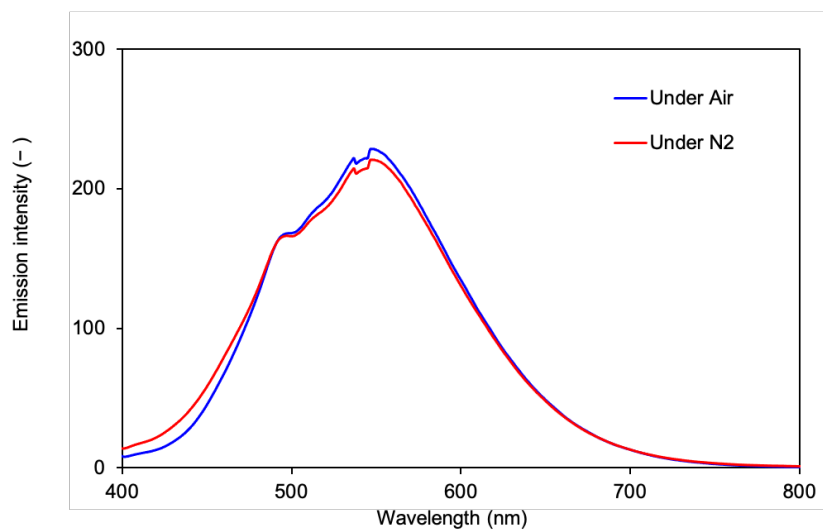
## 5. Concentration dependence of photoluminescence spectra



**Figure S9.** Photoluminescence (PL) spectra of several concentration of **PC** in DCM solvent. The wavelength of excitation light is 350 nm.

## 6. Photoluminescence (PL) spectra under N<sub>2</sub>

PL spectra spectra under air and N<sub>2</sub> (after N<sub>2</sub> bubble for 20 min) are measured at room temperature. Since there is no difference between them, phosphorescence should not be involved in PL of **PC**.



**Figure S10.** PL spectra of **PC** under air and N<sub>2</sub>.