

## CHEMISTRY

## 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>9</sup>, Daniel T. Payne<sup>2,10</sup>, Kazuki Yoshizoe<sup>1,11</sup>, Koji Tsuda<sup>1,7,8\*</sup>

Designing fluorescent molecules requires considering multiple interrelated molecular properties, as opposed to properties that straightforwardly correlated with molecular structure, such as light absorption of molecules. In this study, we have used a de novo molecule generator (DNMG) coupled with quantum chemical computation (QC) to develop fluorescent molecules, which are garnering significant attention in various disciplines. Using massive parallel computation (1024 cores, 5 days), the DNMG has produced 3643 candidate molecules. We have selected an unreported molecule and seven reported molecules and synthesized them. Photoluminescence spectrum measurements demonstrated that the DNMG can successfully design fluorescent molecules with 75% accuracy ( $n = 6/8$ ) and create an unreported molecule that emits fluorescence detectable by the naked eye.

## INTRODUCTION

Fluorescent compounds are important as visible photoemitters in applications across several disciplines, including organic light-emitting diodes (1–4), sensors (5–8), and bioimaging (9–11). Although numerous fluorescent molecules have been developed for these and other applications, new ones are in constant demand to address the shortcomings of current materials in terms of function, sustainability, and low cost. Even subtle changes in chemical structures might lead to major improvements. Fluorescence is a photochemical property governed by quantum mechanics. However, despite the long history of the study of fluorescence, there are no clear guidelines for creating the fluorescent molecules as there are for light-absorbing ones (1, 5, 9).

A simplified physicochemical mechanism of fluorescence emission from a molecule is illustrated in Fig. 1. Initially, we consider the molecule to be in its singlet ground ( $S_0$ ) state; at the  $S_0$  minimum, it absorbs light and transits 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, emitting the energy difference between the  $S_1$  and  $S_0$  states as light (fluorescence). The excited molecule should travel to the minimum in the  $S_1$  state for emitting light without deactivation. Several factors, including reactions with oxygen molecules, molecular

collisions, intra/intermolecular electron transfer, and aggregation, may deactivate the molecule as it travels in the excited state (12). This makes it difficult to correlate fluorescence with molecular structure. Hence, automatizing fluorescent molecule design would be helpful.

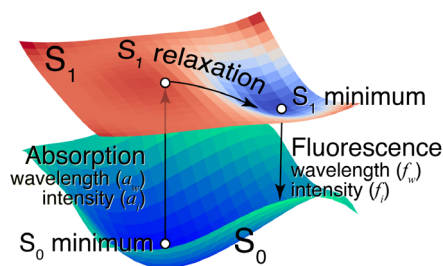
Recently, de novo molecule generators (DNMGs) based on machine learning (ML) (13–16) have been developed for designing molecules with simple and predictable values such as the logarithm of the partition coefficient ( $\log P$ ), which can be estimated from the constitute parts of a molecule (17–19). Combining the DNMGs with classical simulations has successfully generated molecules with improved versatility and practicality. For instance, the combination of a DNMG and docking simulation can be used to design biologically active molecules; this was investigated by organic syntheses (20). In combination with molecular dynamics or predictive models, DNMGs can also direct the synthesis of functional polymers (21, 22).

In previous study, we have combined quantum chemical computations (QCs) with our DNMG, called ChemTS (23), which (in principle) can design functional molecules characterized by their quantum mechanical (QM) properties de novo (24). Thus, ChemTS coupled with QC was applied to design molecules that can absorb light with a desired wavelength (24). Among the 86 designed and generated molecules, 6 known molecules that were not included in the training dataset were selected for the ultraviolet-visible (UV-vis) absorption measurement. The results were consistent with the target wavelength of the generator. In addition, DNMGs increase the possibility of discovering new molecules because the search area of a DNMG is not limited in the dataset in contrast to the traditional high-throughput QM and screening with ML models. We also performed functional group enrichment analysis of the molecules produced by ChemTS with QC to maximize the electron gain energy and found important functional groups that are not included in the electret literature (25).

Although relatively simple properties, such as light absorption and electron gain energy, can be straightforwardly correlated with molecular structure, complex phenomena such as fluorescence that are exhibited only by specific molecules pose vastly greater difficulty. In the case of fluorescence, as shown in Fig. 1, it is necessary to consider multiple properties, which are intricately intertwined. This

<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, Kanagawa 230-0045, Japan. <sup>4</sup>Graduate School of Medicine, Kyoto University, 53 Shogoin-Kawaharacho, Sakyo-ku, Kyoto 606-8507, 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, 1-1 Namiki, Tsukuba, Ibaraki 305-0044, 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>Department of Chemistry, University of Southampton, University Road, Highfield, Southampton SO17 1BJ, UK. <sup>10</sup>International Center for Young Scientists (ICYS), National Institute for Materials Science, 1-1 Namiki, Tsukuba, Ibaraki 305-0044, Japan. <sup>11</sup>Research Institute for Information Technology (RIIT), Kyushu University, 744 Motooka, Nishi-ku, Fukuoka City, Fukuoka 819-0395, Japan.

\*Corresponding author. Email: masato.sumita@riken.jp (M.S.); tsuda@k.u-tokyo.ac.jp (K.T.)



**Fig. 1. Schematic of the PESs of the singlet ground ( $S_0$ ) state and the singlet first excited ( $S_1$ ) state of a fluorescent molecule.** At the  $S_0$  minimum, the molecule is excited vertically to the  $S_1$  excited state, absorbing light of wavelength  $\lambda_a$  with the intensity  $i_a$ . Relaxing in the  $S_1$  state, the molecule reaches the  $S_1$  minimum and fluoresces with wavelength  $\lambda_f$  and intensity  $i_f$ .

makes it difficult to establish intuitive guidelines for the design of molecular structures. To design practically useful compounds, the complex mechanisms governing the target molecular property must be appropriately digitized for utilization in DNMGs. Moreover, the growth of computational cost with increasing the complexity of the mechanism to explore chemical space must be considered.

In this study, we designed fluorescent molecules with a massively parallelized version of ChemTS. The program package used QC to digitize the minimum requirements of the fluorescence mechanism (Fig. 1). There are several user-friendly software packages (26–28) based on electronic structure theories that can be used for QC on molecules and materials. Balancing reliability against computational costs, we used density functional theory (DFT) (29) to evaluate the potential energy surfaces (PESs) shown in Fig. 1. To solve the computational costs of extensively exploring the chemical space, we parallelized ChemTS massively to use 1024 cores based on the concept of virtual loss and generated 3643 molecules. For validation, we synthesized an unreported compound and several reported ones from those generated. Six compounds, including an unreported one, fluoresced as expected. The unreported molecule, which is synthesizable through coupling between commercially available reagents, has an unexpected property, although it is composed of common fragments [coumarin (5), pyridine (30), and pyrazolopyrimidine (31)]. This suggests that the massive parallel DNMG has the potential to induce a paradigm shift in the molecular design. Our implementations of the digitalized mechanism of fluorescence and the parallelized DNMG are available at <https://github.com/tsudalab/GaussianRunPack> and [https://github.com/tsudalab/FL\\_ChemTS](https://github.com/tsudalab/FL_ChemTS).

## RESULTS

### Molecule generation

Assuming that fluorescent molecules could be used for bioimaging (10), we set the upper limits of the absorption and fluorescence wavelengths at 700 and 1200 nm, respectively, and the lower limits of the oscillator strengths (OSs) at 0.01. ChemTS designed 3643 candidate fluorescent molecules using 1024 cores for 5 days. The distribution profiles of the absorption and fluorescence wavelengths and OSs of the molecules evaluated by DFT are shown in Fig. 2 with the structural characteristics, i.e., the number of aromatic rings and the conjugate length. Their theoretical  $S_1$  absorption wavelengths were distributed from 120 to 1200 nm; the fluorescence wavelengths from their  $S_1$ -minimum states were all >180 nm. The molecules

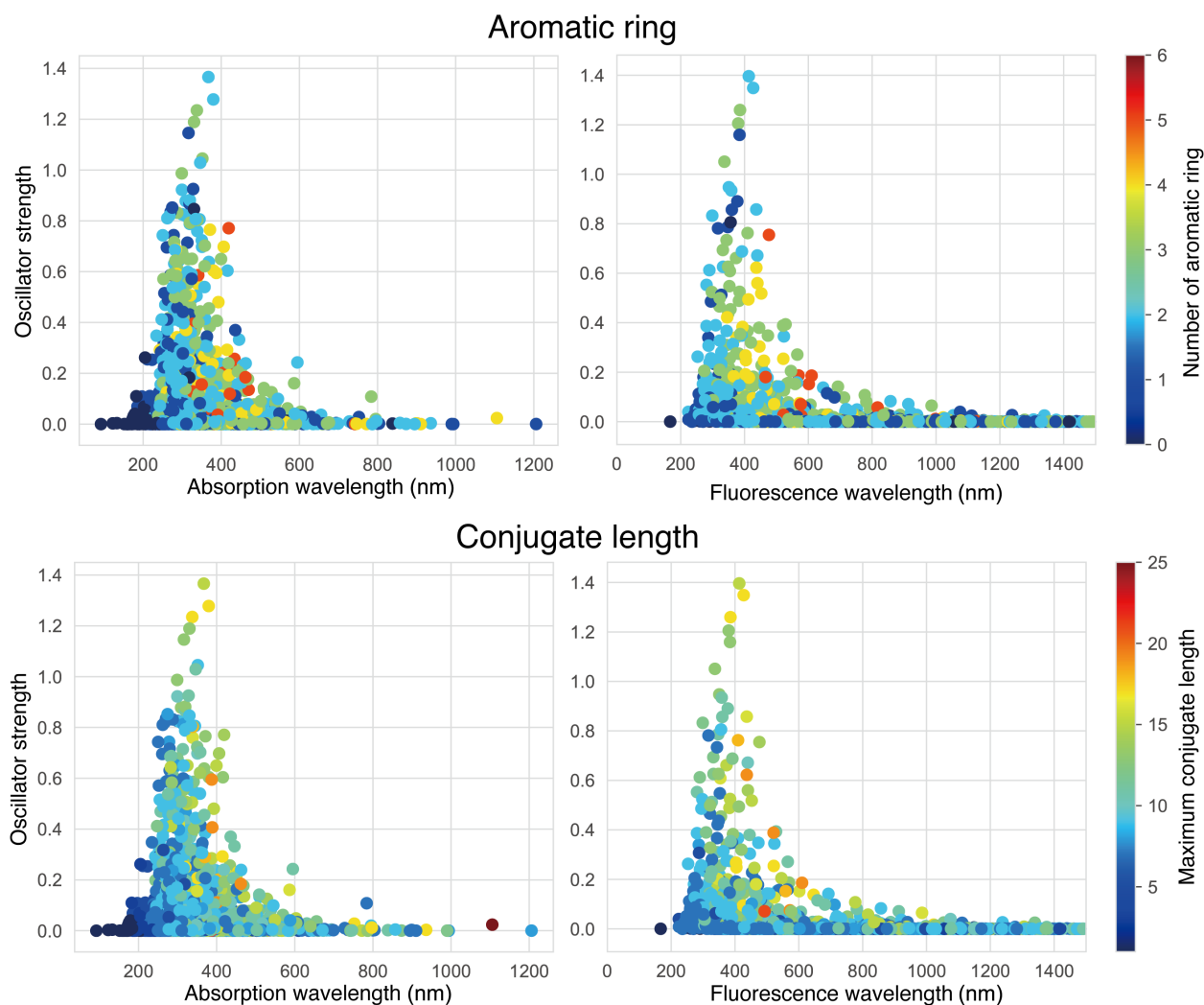
with high OS for adsorption and fluorescence were distributed from approximately 200 to 600 nm.

Typical chromophores often feature the conjugation of multibonds and aromatic rings in organic molecules (1). To determine whether this holds for the candidate molecules produced by the generator, we have analyzed the number of aromatic rings and the conjugate lengths. In Fig. 2, the colored points representing various numbers of aromatic rings are scattered over the entire distribution. However, the brightly colored points representing the conjugate length seem to be concentrated in the high-OS areas. This implies that the conjugate length might correlate with the absorption and fluorescence wavelengths, along with the corresponding intensities, but the number of aromatic rings might not. To obtain more quantitative data, we computed the Pearson's correlation coefficients listed in Table 1. The highest correlation (0.49) was between the conjugate length and the absorption wavelength and its coefficient, which supports the traditional prescription that designing molecules that absorb light with long wavelengths implies the elongation of the conjugate length (32).

To investigate the relationship between molecular features and the properties of absorption and fluorescence more fully, we developed prediction models on the basis of random forest (33) using the Mordred descriptors (34), as shown in table S1 and fig. S3; these models have been recently used in the field of cheminformatics. The absorption wavelength was found to be predictable to some extent ( $R = 0.73$  whose SD was 0.028, evaluated by using fivefold cross validation), and features related to conjugation length appeared at the top as important features. On the other hand, it was difficult to predict the OS of absorption or the wavelength and OS of fluorescence, although a relationship between molecular structure and fluorescence in derivatives that include a common scaffold has been reported (35). All these analyses were only preliminary and did not exclude other molecular properties associated with fluorescence. Nevertheless, the fact that we could not find any correlation between fluorescence (or its OS) and molecular features, such as conjugate length and number of aromatic rings in the wide variety of molecules produced by ChemTS, illustrates the difficulty of designing fluorescent molecules by optimizing the molecular structure.

To select the candidate molecules for the synthesis, we imposed the following three conditions: (i) The OS of the molecule should be greater than 0.1 for  $S_1$  states from the  $S_0$  minimum. (ii) The molecule should emit fluorescence with a wavelength of over 400 nm and the OS of over 0.01. (iii) The difference between the wavelength of absorption and fluorescence in the  $S_1$  states should be greater than 100 nm to ensure distinguishability. Imposing (i) to (iii), we successfully filtered 87 molecules (table S2).

For comparison, we randomly selected 1000 molecules from the training dataset and evaluated their PESs to see whether they were computationally predicted to fluoresce. We found 661 computationally fluorescent molecules, four of which (constituting 0.61% of the training dataset) fluoresced with intensities detectable to the naked eye. By contrast, 87 of molecules generated by ChemTS (2.39% of all candidates) exhibited naked eye-detectable fluorescence. These molecules were relatively discovered in the second half of the run; the median order of generation of the 87 selected molecules was estimated to be 2035 (of 3643). Among them, we found seven known molecules or tautomers (see table S2) included in the chemical database SciFinder (36). We have experimentally confirmed that five of these seven molecules emit the visible fluorescence (see figs. S14 to S17



**Fig. 2.** Distribution profiles of OSs 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 indicate the number of aromatic rings; in the (bottom) images, they indicate the conjugate length.

for their emission spectra and images and table S5 for the quantum yields). The other 80 molecules had not been reported yet; they include four fluorescent molecules that are predicted to emit light in the near-infrared region (around 700 nm) with moderate OS (Fig. 3) and might be useful for bioimaging molecules (9–11). None of these molecules have the common skeletons of known fluorescent molecules. This fact indicates the strength of ML; it can choose skeletons that synthetic chemists would not.

### Photochemistry of PC

To demonstrate the generator's ability to discover novel compounds, we selected an unreported molecule that was not included in SciFinder (36). Because the molecules with novel  $\pi$ -aromatic frameworks are often difficult to synthesize, we chose our test case from the unreported coumarin derivatives, widely known as fluorescent molecules (5). The molecule that we selected, shown in Fig. 3, was 3-[3-(pyridin-3-yl)pyrazolo[1,5-a]pyrimidin-6-yl]-2H-chromen-2-one (PC).

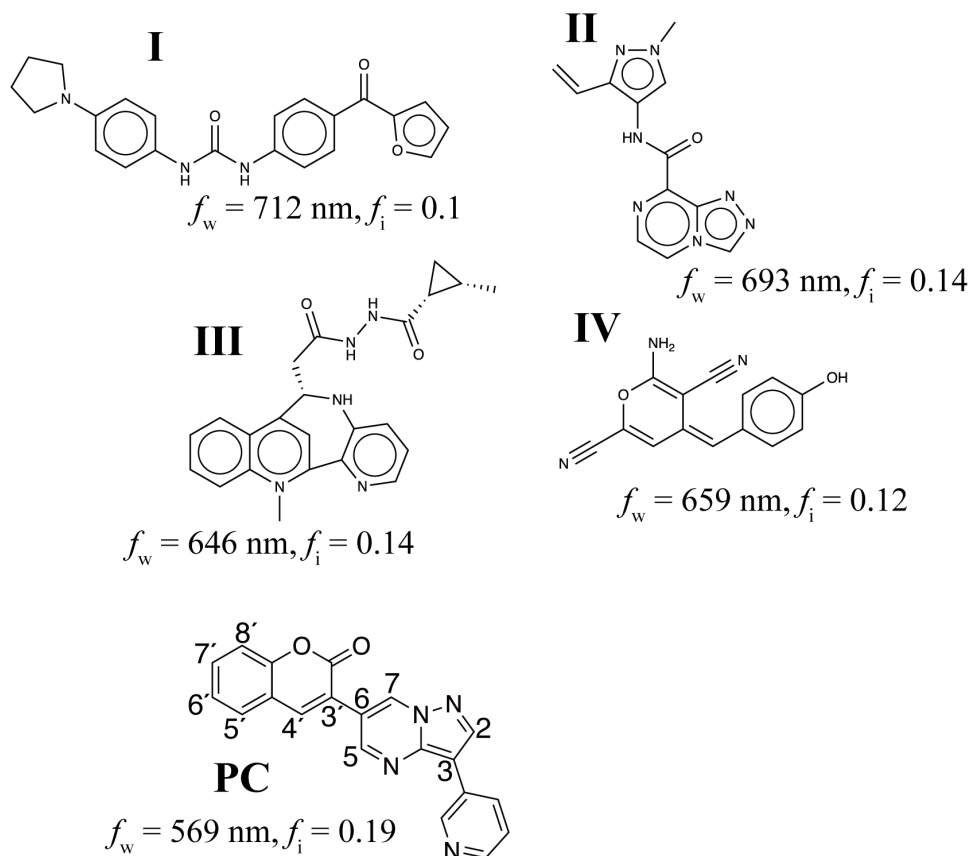
This design principle of PC is a little bit strange from the professional viewpoint. Pyrazolopyrimidine, one of the groups in PC, has

been used as a bioactive component (31) and has recently attracted as a fluorophore (37). Nevertheless, because of the way that the heterocycles are connected, PC is not a compound that would ordinarily attract the interest of chemists interested in synthesizing fluorescent materials. Although the substitution on the 3' position of coumarin has been used to expand the  $\pi$  conjugation (5), a direct connection between two six-membered rings tends to induce a twisted structure (as in biphenyl derivatives) because of steric repulsion; this deformation breaks  $\pi$  conjugation and shortens the absorption wavelength (32). Hence, the experts intuitively avoid the chromophore designs in which the six-membered ring of the pyrazolopyrimidine group directly connects to the 3' position of coumarin (5). Unfavorable deformation can also be induced by the pyridinyl group connecting to the pyrazolopyrimidine group via its meta-position. We will return to the design principles of PC in a later subsection.

Figure 4A shows the optimized structure of PC including its conformations (38) at the B3LYP/6-31G\* level. In agreement with the arguments in the previous paragraph, PC had a slightly twisted structure. Its torsion angle, however, was no larger than we expected: The torsion angle between pyrazolopyrimidine and coumarin was

**Table 1. Pearson's correlation coefficients ( $R$  values) of wavelengths and their OS 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 Materials for their correlation graphs (fig. S1).**

	$S_1$ absorption		$S_1$ fluorescence	
	Wavelength	OS	Wavelength	OS
Number of aromatic rings	0.35	0.18	-0.04	0.13
Conjugate length	0.49	0.26	-0.03	0.24

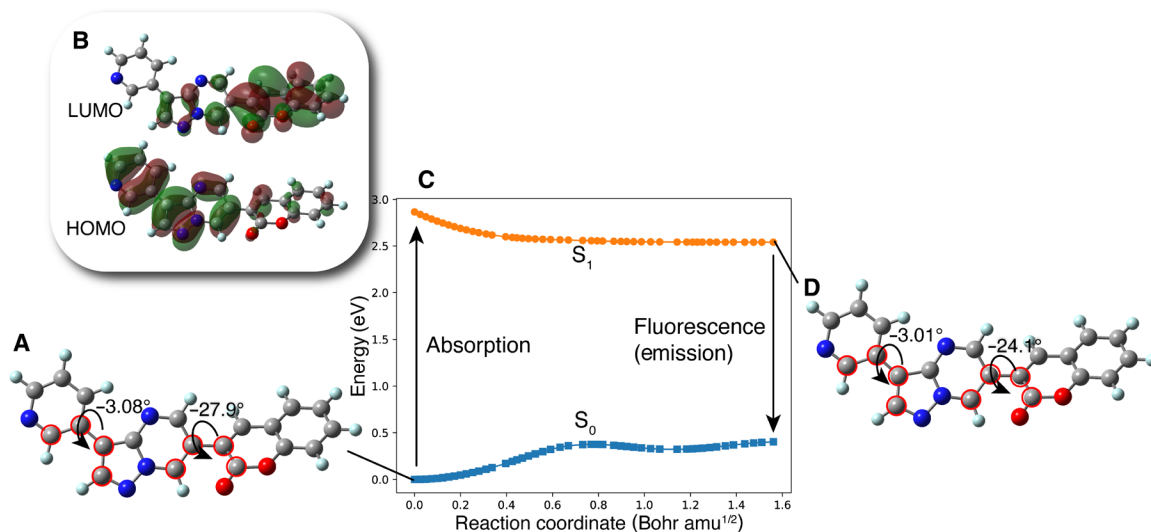


**Fig. 3. Unreported fluorescent molecules designed using ChemTS.** Molecules I to IV are expected to emit near-infrared light. Molecule **PC** is 3-[3-(Pyridin-3-yl)pyrazolo[1,5-a]pyrimidin-6-yl]-2H-chromen-2-one, which was synthesized to validate in this work. The fluorescent light wavelength ( $f_w$ ) and corresponding OS ( $f_i$ ) were computed at the B3LYP/3-21G\* level.

estimated to be  $27.9^\circ$ . This is relatively small compared to that of 3-phenyl-coumarin, which has a torsion angle of  $35.3^\circ$ , according to our preliminary calculation at the B3LYP/6-31G\* level; the Cartesian coordinate is shown in the Supplementary Materials. **PC**'s relatively small torsion angle is probably due to the weakness of the repulsion between  $\pi$  orbitals of the rings. The orbitals of nitrogen-rich ring, pyrazolopyrimidine, have higher energies than that of hydrocarbon rings. The highest occupied molecular orbital (HOMO) of **PC** attributes to  $\pi$  orbital of the pyridinyl-pyrazolopyrimidine moiety, as shown in Fig. 4B. In addition, the introduction of the nitrogen-rich ring results in charge transfer (CT) excitation. On the other hand, the torsion angle between pyridinyl and pyrazolopyrimidine is only  $3.08^\circ$ , reflecting the small steric repulsion induced by the pyridinyl

group linked to the five-member ring of pyrazolopyrimidine at the meta-position.

According to the time-dependent (TD)-DFT computation, the vertical excitation to the  $S_1$  state from the  $S_0$  state of compound **PC** is due to the transition of the one electron from HOMO to the lowest unoccupied molecular orbital (LUMO) ( $\pi$ - $\pi^*$  excitation). As shown in Fig. 4B, the excitation to the  $S_1$  state induces a CT from the pyrazolopyrimidine moiety to the coumarin moiety. The  $S_1$  excited molecule reaches a stationary point in the  $S_1$  state with little increasing structural flatness, and this is enough to stabilize the  $S_1$  excited **PC** and over 0.33 eV from the  $S_1$  vertically excited point. We confirmed by the minimum energy path calculation that this minimum in the  $S_1$  state is accessible from the  $S_1$  vertically excited point without a



**Fig. 4. Photoinduced process of PC.** (A)  $S_0$  minimum structure of PC. (B) The HOMO and the LUMO of optimized PC in the  $S_0$  state at the B3LYP/6-31G\* level. The  $S_1$  state of PC results from the transition of one electron from HOMO to LUMO ( $\pi$ - $\pi^*$  excitation); the excitation (absorption) energy was computed to be 2.87 eV (433 nm) with OS of 0.192. The energy of fluorescence was computed to be 2.14 eV (580 nm) with OS of 0.128. (C)  $S_1$  minimum-energy path up to the  $S_1$  minimum structure. Energy is relative to  $S_0$  minimum. The nature of orbitals involved in the transition was not changed along the path. (D)  $S_1$  minimum structure of PC. The values of dihedral angles in (A) and (D) are measured from the carbon atoms marked by red circles.

barrier, as shown in Fig. 4C. Hence, PC is fluorescent, as was designed to be by ChemTS. Molecules in the excited state with the CT characteristic tend to show energetically large relaxations in their excited state, thereby causing moderate differences between the absorption and fluorescence; in the present case, the theoretically estimated difference between the two was 147 nm ( $S_1$  absorption wavelength, 433 nm;  $S_1$  fluorescence wavelength, 580 nm at the B3LYP/6-31G\* level).

### Experimental measurements

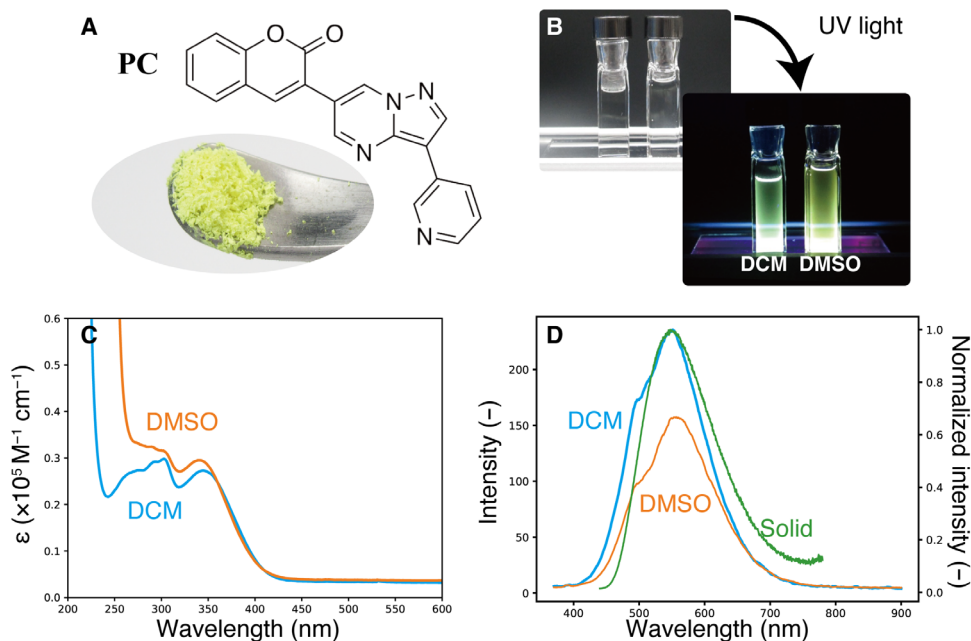
Next, we describe validating the generator, synthesizing PC, and measuring its photoluminescence (PL) spectra. The synthesis route to PC and its characterization are described in the Supplementary Materials. Figure 5A shows that PC is a yellow solid material at room temperature. Figure 5B shows PC at a concentration of 0.1 mM in dichloromethane (DCM) and dimethyl sulfoxide (DMSO) solvents, first under room light and then upon the irradiation of UV light (365 nm). The fluorescence (yellow or green depending on the solvent) can be observed with the naked eye. The UV-vis absorption and PL spectra of PC in DCM and DMSO are shown in Fig. 5 (C and D, respectively). The PL spectrum of PC in the solid state resembled those of PC in the solutions, as shown in Fig. 5D. For more information on the concentration dependence of their spectra, see the Supplementary Materials. 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 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 Materials).

The experimental difference between the absorption and fluorescence spectra was evaluated to be approximately 125 nm, which

is comparable to the theoretically predicted shift (147 nm). Although static theoretical computation succeeded in reproducing the experimental results, molecular dynamics did not. If the real crossing sites between the  $S_1$  and  $S_0$  PESs (conical intersections) were accessible by the molecular vibration in the  $S_1$  state (39, 40), then  $S_1$  excited molecules could go back to their  $S_0$  states nonradiatively and the fluorescence yield would clearly decrease. This photoinduced dynamical behavior of the molecule in the  $S_1$  excited state was reflected in the value of the quantum yield ( $\Phi$ ); the  $\Phi$  of PC measured by an integrating sphere in DCM was 0.007. This small  $\Phi$  indicates that there were some nonradiative paths to the ground states (39, 40). 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. Because a nonadiabatic decay route to the  $S_0$  state after  $\pi$ - $\pi^*$  excitations of coumarin itself has been reported (41), there might be a nonradiative way to the  $S_0$  state after the internal conversion from the  $S_1$  state to the  $S_2$  state.

### Meaning of steric structure of PC

Although we succeeded in validating our fluorescent molecule generation system by synthesizing PC and measuring its fluorescence, we could not yet understand the design principle of PC. We therefore prepared isomers of PC and a coumarin derivative (V) that from standpoint of conventional design seemed likely to fluoresce on the basis of the computational results at the B3LYP/6-31G\* level summarized in Table 2. We designed three PC isomers, differing in the connecting positions of pyridine to pyrazolopyrimidine and of pyrazolopyrimidine to coumarin. The steric repulsion was expected to be lower when coumarin was added at the 5-position of pyrazolopyrimidine (PC2 and PC3) than when it was added at the 6-position (PC and PC1). In the case of PC1 and PC2, the pyridine groups were linked with the pyrazolopyrimidine groups via their ortho-position to decrease the steric repulsion between them. Last, molecule V is a chromophore with the coumarin scaffold designed from the viewpoint of the structural planarity with a conjugated length similar to that of the PC species.



**Fig. 5. Photochemical properties of PC.** (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 with 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) 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).

Comparing PC itself with the other PC species, we can understand why PC has connections that increase the steric repulsion among coumarin, pyrazolopyrimidine, and pyridine groups. In the case of PC2 and PC3, where the steric repulsion between pyrazolopyrimidine and coumarin is reduced [as reflected in the small torsion ( $\theta_1$ ) angles between them], the wavelengths of fluorescence from the  $S_1$  state show a drastic red shift, while the OSs of absorption and fluorescence are decreased. For molecules that show thermally activated delayed fluorescence, increasing the steric repulsion around a bond is known to split the electronic distribution between HOMO and LUMO but to result in low OS (4). In the PC species, however, increasing the steric repulsion between pyrazolopyrimidine and coumarin increases their OS. Coumarin derivatives attached to the 7-position of pyrazolopyrimidine have already been reported as fluorophores (37) with large deformation from the flat plate ( $-37.4^\circ$ ). Although their intensity is high, their absorption (331 to 440 nm) and fluorescence (470 to 538 nm) show a slight blue shift in comparison with those of PC molecules (absorption, 425 nm; fluorescence, 550 nm). Therefore, PC is the result of ChemTS considering the trade-off between OS and wavelength.

On the other hand, the molecule V, designed from the professional viewpoint to fluoresce, certainly has a planar structure. The  $S_1$  excitation of V is a typical  $\pi$ - $\pi^*$  excitation with high OS that induces the bond alternation. However, the fluorescence wavelength is not very long (493 nm), and the difference between it and the absorption wavelength (53.0 nm) is also small. These facts illustrate the difficulty of designing chromophores with the intention of absorbing and emitting target wavelength of lights.

## DISCUSSION

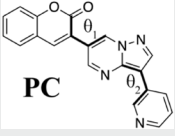
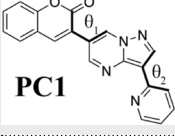
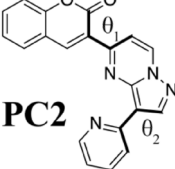
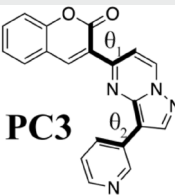
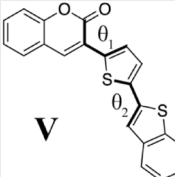
For several decades, QC has played an important role in chemistry and materials science. During this time, computer-aided molecular

design has been used in drug discovery (42). However, QC has solely focused on the analysis and speculation of the experimental results; the creative work such as the prediction of various phenomena and designing materials (6, 43–47) has rarely been considered. The recent applications of ML algorithms to chemistry and materials science (48, 49) represent a positive turning point for computer-aided chemistry and materials science. To design molecules useful for organic electronics automatically, combining QC with a DNMG is crucial because, in such applications, quantum mechanics cannot be ignored. Nevertheless, the value of a DNMG based on QC must be proven before it can be adopted in practice.

In this study, we used a DNMG to create molecules with a property that cannot currently be easily predicted: fluorescence. We designed fluorescent compounds using DFT, an inherently quantum mechanical method. Although it is known that molecules are governed by the rules of quantum mechanics, it is difficult to create a new molecule de novo with only QC. Although fluorescent molecules have simple PESs, they are hard to design from first principles because their diversity makes it very difficult to correlate fluorescence with molecular structure.

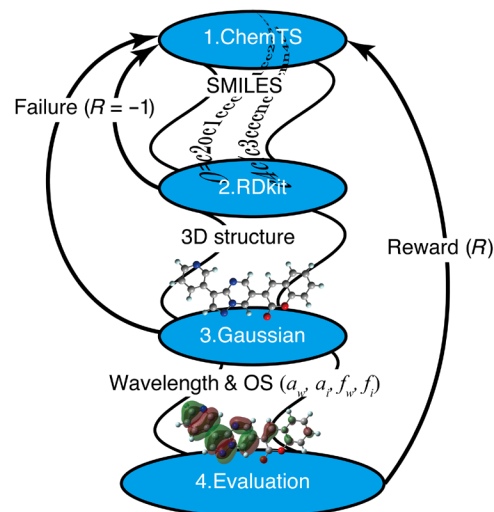
However, the generator dealt with this diversity and succeeded in designing fluorescent molecules from scratch. The substantial ab initio computations based on QC required massive parallel computations (1024 cores, 5 days); nevertheless, the generator succeeded in producing 3643 candidate molecules.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

**Table 2. PC species and a coumarin derivative V (a conventionally designed chromophore).** Here,  $a_w$  and  $a_i$  are absorption wavelength and corresponding OS, and  $f_w$  and  $f_i$  are fluorescence wavelength and its OS, computed by the TD-DFT computation at the B3LYP/6-31G\* level;  $\theta_1$  and  $\theta_2$  are torsion angles of each  $S_0$  optimized structure measured from atoms linked by bold type bonds.

Species	$a_w$ (nm)	$a_i$	$f_w$ (nm)	$f_i$	$\theta_1$	$\theta_2$
 PC	433	0.192	577	0.128	-27.9	-3.08
 PC1	430	0.214	565	0.141	-28.2	-0.36
 PC2	486.8	0.096	718	0.063	13.7	4.20
 PC3	475	0.110	702	0.070	6.40	20.9
 V	440	0.958	493	0.990	0.189	-0.802

We selected seven known compounds for validation and one candidate compound for further study on the basis of criteria of synthesizability and visible fluorescence. Experimental validation exhibited that the DNMG successfully designed fluorescent compounds with 75% (six of eight). The innovation potential of the DNMG was proven by the fluorescence (visible to the naked eye) of PC. In designing PC, the generator introduced an unfamiliar group, pyrazolopyrimidine, to coumarin; the connection induced a high steric repulsion but still resulted in an increase in OS. It is difficult for common chemists to come up with a way to enhance fluorescence and its intensity by increasing the steric repulsion between fragments. This shows that the generator is a tool that can surpass professional knowledge or intuition. DNMGs have the potential to induce a paradigm shift in molecular design.

Although the fluorescence of the molecule synthesized in this work was intended to be detected by the naked eye, more interesting molecules would be produced by removing the restrictions on the types of atoms and extending the design time. In addition, better molecular



**Fig. 6. Workflow of fluorescent molecule generation.** ChemTS generates the molecules represented in the SMILES strings (56), which are translated to a three-dimensional (3D) structure via RDKit (58). The absorption and fluorescence wavelengths and corresponding OSs obtained by the Gaussian package (26) are evaluated using the reward function (Eq. 1). ChemTS produces the molecules based on the electronic structure theory by iterating this process.

design would be possible by including the photoinduced dynamics of molecules. By further developing the QC, more complex functional molecules could be designed. Thus, generators with massively parallel computation would be able to create complex molecules with diverse and intriguing functions, possibly resulting in very complex synthesis routes that would increase the requirements of the recently developed planning retrosynthesis route based on ML (50–52).

## METHODS

### Molecular design

A simple fluorescent molecule has a minimum in the excited state with the same spin multiplicity as its ground state. Figure 1 shows an idealized PES of a molecule that emits fluorescence from its  $S_1$  state. Ignoring several factors that quench fluorescence, we designed the molecules with PESs similar to that shown in Fig. 1. Even in this simple framework, no qualitative guidelines are available for designing fluorescent molecules, owing to their diversity. A DNMG can address this. We used the molecule generator, ChemTS (23), which uses Monte Carlo tree search (MCTS) (53) and a recurrent neural network (RNN) (54, 55) to generate molecules. Initially, the RNN was trained with a set of SMILES (simplified molecular-input line-entry system) strings (56). In our case, 153,253 molecules containing H, C, N, and O, obtained from the ZINC database (57), were used.

The PESs of the generated molecules were evaluated at the B3LYP/3-21G\* level, implemented in the Gaussian 16 package (26). The excited energy and the fluorescence wavelength were computed by 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. A flow-chart illustrating the design of fluorescent molecules with ChemTS is shown in Fig. 6. To evaluate the absorption and fluorescence wavelengths and intensity simultaneously, we used the reward function

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

Here,  $W_{a_w, a_i, f_w, f_i}$ , respectively, denote the weights of the reward functions  $R_{a_w, a_i, f_w, f_i}$

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

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

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

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

$T_{a_w}$  and  $T_{a_i}$  denote the upper absorption wavelength and its OS, respectively. Similarly,  $T_{f_w}$  and  $T_{f_i}$  are the upper fluorescence wavelength and its OS, respectively. In this work, we set  $T_{a_w}$  and  $T_{f_w}$  to 700 and 1200 nm, respectively, and their OS of  $T_{a_i}$  and  $T_{f_i}$  to 0.01. The values of  $W_{a_w}$ ,  $W_{a_i}$ ,  $W_{f_w}$ , and  $W_{f_i}$  were set to 0.4, 0.1, 0.4, and 0.1, respectively. We fixed values of  $\sigma_a$ ,  $\sigma_f$  and  $\epsilon$  to 150, 150, and  $10^{-8}$ , respectively. The absorption and fluorescence wavelengths of the generated molecules are denoted by  $a_w$  [with corresponding intensity ( $a_i$ )] and  $f_w$  [with corresponding intensity ( $f_i$ )], respectively (see Fig. 1).

In this study, to accelerate the MCTS, we used a parallelization strategy based on the concept of virtual loss (53). When MCTS is parallelized, concentration of computation on a particular leaf node becomes a problem: If a subprocess on a leaf node has not finished its computation, then the same leaf node is selected by other subprocesses in the selection step. To solve this problem, parallelization with virtual loss explicitly treats whether a subprocess is computing on a node and uses the following score for each child node  $i$  in selection step

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

If a subprocess is computing on a node, then the denominators increase, and that node will be less likely to be selected, resulting in distributed computation. 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 the conjugate length. We counted aromatic rings using an RDKit function (58). For the conjugate length, we counted a unity that is the sequence of single bond–double bond–single bond (e.g.,  $-\text{C}=\text{C}-$ ).

### Molecule synthesis and measurement

We selected one compound, **PC**, from those generated by ChemTS for chemical synthesis using ease of synthesis as a criterion. Because the computational level used for molecular design was B3LYP/3-21G\*, which is somewhat unreliable, the PES of **PC** was checked at the B3LYP/6-31G\* level before synthesis. In addition to **PC**, we synthesized seven molecules reported in SciFinder (36) for analyzing absorption and emission spectra. Compound **PC** and the seven other compounds were obtained from Tokyo Chemical Industry Co. Ltd. through custom synthesis (refer to the Supplementary Materials for the detailed synthesis process and its characterization of **PC**).

The electronic absorption spectra were measured at 20°C using a Shimadzu UV-3600 UV–vis–near-infrared spectrophotometer. The emission spectra in solution and solid state were measured using FP-8300 (JASCO, Japan). A quartz cell with an optical length of 1.0 mm or 1.0 cm was used. Spectroscopic-grade solvents (DCM and DMSO) 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.

### SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <https://science.org/doi/10.1126/sciadv.abj3906>

### REFERENCES AND NOTES

1. Y. Qin, G. Li, T. Qi, H. Huang, Aromatic imide/amide-based organic small-molecule emitters for organic light-emitting diodes. *Mater. Chem. Front.* **4**, 1554–1568 (2020).
2. M. Mamada, K. Inada, T. Komino, W. J. Potscavage Jr., H. Nakanotani, C. Adachi, Highly efficient thermally activated delayed fluorescence from an excited-state intramolecular proton transfer system. *ACS Cent. Sci.* **3**, 769–777 (2017).
3. H. Uyama, K. Goushi, K. Shizu, H. Nomura, C. Adachi, Highly efficient organic light-emitting diodes from delayed fluorescence. *Nature* **492**, 234–238 (2012).
4. H. Kaji, H. Suzuki, T. Fukushima, K. Shizu, K. Suzuki, S. Kubo, T. Komino, H. Oiwai, F. Suzuki, A. Wakamiya, Y. Murata, C. Adachi, Purely organic electroluminescent material realizing 100% conversion from electricity to light. *Nat. Commun.* **6**, 8476 (2015).
5. D. Cao, Z. Liu, P. Verwilst, S. Koo, P. Jangjili, J. S. Kim, W. Lin, Coumarin-based small-molecule fluorescent chemosensors. *Chem. Rev.* **119**, 10403–10519 (2019).
6. H. Izawa, S. Nishino, M. Sumita, M. Akamatsu, K. Morihashi, S. Ifuku, M. Morimoto, H. Saimoto, A novel 1,8-naphthalimide derivative with an open space for an anion: Unique fluorescence behaviour depending on the binding anion's electrophilic properties. *Chem. Commun.* **51**, 8596–8599 (2015).
7. M. C.-L. Yeung, V. W.-W. Yam, Luminescent cation sensors: From host–guest chemistry, supramolecular chemistry to reaction-based mechanisms. *Chem. Soc. Rev.* **44**, 4192–4202 (2015).
8. P. A. Gale, C. Caltagirone, Anion sensing by small molecules and molecular ensembles. *Chem. Soc. Rev.* **44**, 4212–4227 (2015).
9. S. Banerjee, E. B. Veale, C. M. Phelan, S. A. Murphy, G. M. Tocci, L. J. Gillespie, D. O. Frimannsson, J. M. Kelly, T. Gunnlaugsson, 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).
10. J. A. Carr, D. Franke, J. R. Caram, C. F. Perkinson, M. Saif, V. Askoxylyakis, M. Datta, D. Fukumura, R. K. Jain, M. G. Bawendi, O. T. Bruns, Shortwave infrared fluorescence imaging with the clinically approved near-infrared dye indocyanine green. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4465–4470 (2018).
11. R. Chouket, A. Pellissier-Tanon, A. Lemarchand, A. Espagne, T. Saux, L. Jullien, Dynamic contrast with reversibly photoswitchable fluorescent labels for imaging living cells. *Chem. Sci.* **11**, 2882–2887 (2020).
12. B. Valeur, M. N. Berberan-Santos, in *Molecular Fluorescence: Principles and Applications* (Wiley, ed. 2, 2012), pp. 141–179.
13. D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton, Y. Chen, T. Lillicrap, F. Hui, L. Sifre, G. van der Drriessche, T. Graepel, D. Hassabis, Mastering the game of Go without human knowledge. *Nature* **550**, 354–359 (2017).
14. I. Goodfellow, J. Pouget-Abadie, M. Miza, B. Xu, D. Warda-Farley, S. Ozair, A. Couville, Y. Bengio, Generative adversarial nets. *Adv. Neural Inform. Process. Sys.* **2**, 2672–2680 (2014).
15. J. -P. Briot, G. Hadjeres, F. Pachet, Deep learning techniques for music generation—A survey. arXiv:1709.01620 [cs.LG] (2017).
16. J. Devlin, M. -W. Chang, K. Lee, K. Toutanova, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (Long and Short Papers, 2017), pp. 4171–4186, vol. 1.
17. R. Gómez-Bombarelli, J. N. Wei, D. Duvenaud, J. M. Hernández-Lobato, B. Sánchez-Lengeling, D. Sheberla, J. Aguilera-Iparraguirre, T. D. Hirzel, R. P. Adams, A. Aspuru-Guzik, Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **4**, 268–276 (2018).
18. B. Sanchez-Lengeling, A. Aspuru-Guzik, Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **361**, 360–365 (2018).
19. D. C. Elton, Z. Boukouvalas, M. D. Fuge, P. W. Chung, Deep learning for molecular design—A review of the state of the art. *Mol. Syst. Des. Eng.* **4**, 828–849 (2019).
20. D. Merk, F. Grisoni, J. Friedrich, G. Schneider, Tuning artificial intelligence on the *de novo* design of natural-product-inspired retinoid X receptor modulators. *Commun. Chem.* **1**, 68 (2018).



21. S. Wu, Y. Kondo, M. Kakimoto, B. Yang, H. Yamada, I. Kuwajima, G. Lombard, K. Hongo, Y. Xu, J. Shimoi, C. Schick, J. Morikawa, R. Yoshida, Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. *npj Comput. Mater.* **5**, 66 (2019).
22. S. Kajita, T. Kinjo, T. Nishi, Autonomous molecular design by Monte-Carlo tree search and rapid evaluations using molecular dynamics simulations. *Commun. Phys.* **3**, 77 (2020).
23. X. Yang, J. Zhang, K. Yoshizoe, K. Terayama, K. Tsuda, ChemTS: An efficient Python library for *de novo* molecular generation. *Sci. Technol. Adv. Mater.* **18**, 972–976 (2017).
24. M. Sumita, X. Yang, S. Ishihara, R. Tamura, K. Tsuda, Hunting for organic molecules with artificial intelligence: Molecules optimized for desired excitation energies. *ACS Cent. Sci.* **4**, 1126–1133 (2018).
25. Y. Zhang, J. Zhang, K. Suzuki, M. Sumita, K. Terayama, J. Li, Z. Mao, K. Tsuda, Y. Suzuki, Discovery of polymer electret material via *de novo* molecule generation and functional group enrichment analysis. *App. Phys. Lett.* **118**, 223904 (2021).
26. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, J. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, Gaussian 16, Revision C.01, (Gaussian Inc., 2016).
27. M. W. Schmidt, K. K. Baldrige, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery Jr., General atomic and molecular electronic structure system. *J. Comput. Chem.* **14**, 1347–1363 (1993).
28. E. Aprà, E. J. Bylaska, W. A. de Jong, N. Govind, K. Kowalski, T. P. Straatsma, M. Valiev, H. J. J. van Dam, Y. Alexeev, J. Anchell, V. Anisimov, F. W. Aquino, R. Atta-Fynn, J. Autschbach, N. P. Bauman, J. C. Becca, D. E. Bernholdt, K. Bhaskaran-Nair, S. Bogatko, P. Borowski, J. Boschen, J. Brabec, A. Bruner, E. Cauët, Y. Chen, G. N. Chuev, C. J. Cramer, J. Daily, M. J. O. Deegan, T. H. Dunning Jr., M. Dupuis, K. G. Dyall, G. I. Fann, S. A. Fischer, A. Fonari, H. Früchtl, L. Gagliardi, J. Garza, N. Gawande, S. Ghosh, K. Glaesemann, A. W. Götz, J. Hammond, V. Helms, E. D. Hermes, K. Hirao, S. Hirata, M. Jacquelin, L. Jensen, B. G. Johnson, H. Jónsson, R. A. Kendall, M. Klemm, R. Kobayashi, V. Konkov, S. Krishnamoorthy, M. Krishnan, Z. Lin, R. D. Lins, R. J. Littlefield, A. J. Logsdail, K. Lopata, W. Ma, A. V. Marenich, J. M. del Campo, D. Mejia-Rodriguez, J. E. Moore, J. M. Mullin, T. Nakajima, D. R. Nascimento, J. A. Nichols, P. J. Nichols, J. Nieplocha, A. Otero-de-la-Roza, B. Palmer, A. Panyala, T. Pirojsirikul, B. Peng, R. Peverati, J. Pittner, L. Pollack, R. M. Richard, P. Sadayappan, G. C. Schatz, W. A. Shelton, D. W. Silverstein, D. M. A. Smith, T. A. Soares, D. Song, M. Swart, H. L. Taylor, G. S. Thomas, V. Tipparaju, D. G. Truhlar, K. Tsemekhman, T. Van Voorhis, A. Vázquez-Mayagoitia, P. Verma, O. Villa, A. Vishnu, K. D. Vogiatzis, D. Wang, J. H. Weare, M. J. Williamson, T. L. Windus, K. Woliński, A. T. Wong, Q. Wu, C. Yang, Q. Yu, M. Zacharias, Z. Zhang, Y. Zhao, R. J. Harrison, NWChem: Past, present, and future. *J. Chem. Phys.* **152**, 184102 (2020).
29. R. G. Parr, W. Yang, *Density-Functional Theory of Atoms and Molecules* (Oxford Univ. Press, 1989).
30. D. M. Smith, Compounds containing six-membered rings with one nitrogen atom; pyridine and its derivatives, in *Rod's Chemistry of Carbon Compounds*, S. Coffey, Ed. (Elsevier, 2008) vol. 4, chap. 24.
31. J.-C. Castillo, J. Portilla, Recent Advances in the synthesis of new pyrazole derivatives. *Targets Heterocycl. Syst.* **22**, 194–223 (2018).
32. J. Guay, P. Kasai, A. Diaz, R. Wu, J. M. Tour, L. H. Dao, Chain-length dependence of electrochemical and electronic properties of neutral and oxidized soluble  $\alpha,\alpha$ -coupled thiophene oligomers. *Chem. Mater.* **4**, 1097–1105 (1992).
33. L. Breiman, Random forests. *Mach. Learn.* **45**, 5–32 (2001).
34. H. Moriwaki, Y. S. Tian, N. Kawashita, T. Takagi, Mordred: A molecular descriptor calculator. *J. Cheminfo.* **10**, 4 (2018).
35. A. Schüller, G. B. Goh, H. Kim, J.-S. Lee, Y.-T. Chang, Quantitative structure-fluorescence property relationship analysis of a large BODIPY library. *Mol. Inform.* **29**, 717–729 (2010).
36. *SciFinder*; Chemical Abstracts Service: Columbus, OH; <https://scifinder.cas.org> [accessed 2020 Sep 13].
37. A. Tigres, S.-L. Aranzazu, N.-F. Bravo, J. Zapata-Rivera, J. Portilla, Pyrazolo[1,5-*a*] pyrimidines-based fluorophores: A comprehensive theoretical-experimental study. *RSC Adv.* **10**, 39542–39552 (2020).
38. K. Terayama, M. Sumita, M. Katouda, K. Tsuda, Y. Okuno, Efficient search for energetically favorable molecular conformations against metastable states via gray-box optimization. *J. Chem. Theory, Comput.* **17**, 5419–5427 (2021).
39. M. A. Robb, M. Garavelli, M. Olivucci, F. Bernardi, A computational strategy for organic photochemistry. *Rev. Comput. Chem.* **15**, 87–146 (2000).
40. M. Domcke, D. R. Yarkony, H. Köppel, *Conical Intersections: Electronic Structure, Dynamics & Spectroscopy* (World Scientific, 2004).
41. Y. Gan, L. Yue, X. Guo, C. Zhu, Z. Cao, Multi-state nonadiabatic deactivation mechanism of coumarin revealed by *ab initio* on-the-fly trajectory surface hopping dynamic simulation. *Phys. Chem. Chem. Phys.* **19**, 12094–12106 (2017).
42. G. R. Marshall, Computer-aided drug design. *Annu. Rev. Pharmacol. Toxicol.* **27**, 193–213 (1987).
43. H. Struebing, Z. Ganase, P. G. Karamertzanis, E. Siougkrou, P. Haycock, P. M. Piccione, A. Armstrong, A. Galindo, C. A. Adjiman, Computer-aided molecular design of solvents for accelerated reaction kinetics. *Nat. Chem.* **5**, 952–957 (2013).
44. L. P. Wang, A. Titov, R. MacGibbon, F. Liu, V. S. Pande, T. J. Martínez, Discovering chemistry with an *ab initio* nanoreactor. *Nat. Chem.* **6**, 1044–1048 (2014).
45. S. Zhang, J. Zhou, Q. Wang, X. Chen, Y. Kawazoe, P. Jena, Penta-Graphene: A new carbon allotrope. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 2372–2377 (2015).
46. M. Sumita, Y. Tanaka, M. Ikeda, T. Ohno, Theoretically designed  $\text{Li}_3\text{PO}_4$  (100)/ $\text{LiFePO}_4$  (010) coherent electrolyte/cathode interface for all solid-state Li ion secondary batteries. *J. Phys. Chem. C* **119**, 14–22 (2015).
47. M. Sumita, R. Tamura, K. Homma, C. Kaneta, K. Tsuda, Li-Ion Conductive  $\text{Li}_3\text{PO}_4$ - $\text{Li}_2\text{BO}_3$ - $\text{Li}_2\text{SO}_4$  mixture: Prevision through density functional molecular dynamics and machine learning. *Bull. Chem. Soc. Jpn.* **92**, 1100–1106 (2019).
48. R. Pollice, G. P. Gomes, M. Aldeghi, R. J. Hickman, M. Krenn, C. Lavigne, M. Lindner-D'Addario, A. Nigam, C. T. Ser, Z. Yao, A. Aspuru-Guzik, Data-driven strategies for accelerated materials design. *Acc. Chem. Res.* **54**, 849–860 (2021).
49. K. Terayama, M. Sumita, R. Tamura, K. Tsuda, Black-box optimization for automated discovery. *Acc. Chem. Res.* **54**, 1334–1346 (2021).
50. M. H. S. Segler, M. Preuss, P. Mark, Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **555**, 604–610 (2018).
51. C. Coley, R. Barzilay, T. Jaakkola, W. H. Green, K. F. Jensen, Prediction of organic reaction outcomes using machine learning. *ACS Cent. Sci.* **3**, 434–443 (2017).
52. R. Shibukawa, S. Ishida, K. Yoshizoe, K. Wasa, K. Takasu, Y. Okuno, K. Terayama, K. Tsuda, CompRet: A comprehensive recommendation framework for chemical synthesis planning with algorithmic enumeration. *J. Cheminfo.* **12**, 52 (2020).
53. C. B. Browne, E. Powley, D. Whitehouse, S. M. Lucas, P. I. Cowling, P. Rohlfshagen, S. Tavener, D. Perez, S. Smothrakis, S. Colton, A survey of Monte Carlo tree search methods. *IEEE Trans. Comput. Intell. AI Games* **4**, 1–43 (2012).
54. S. Hochreiter, J. Schmidhuber, Long short-term memory. *Neural Comput.* **9**, 1735–1780 (1997).
55. K. Cho, 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* (2014), pp 1724–1734.
56. D. Weininger, SMILES, A chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Model.* **28**, 31–36 (1988).
57. J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad, R. G. Coleman, ZINC: A free tool to discover chemistry for biology. *Chem. Inf. Model.* **52**, 1757–1768 (2012).
58. G. Landrum, RDKit: Open-source cheminformatics; [www.rdkit.org](http://www.rdkit.org) [accessed 2017 Nov 7].

#### Acknowledgments

**Funding:** This work was partially supported by a project subsidized by the New Energy and Industrial Technology Development Organization (NEDO) and MEXT as 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 RAIDEN. **Author contributions:** M.S., K.Ts., and R.T. planned and supervised the project. M.S. and K.Te. performed the computational experiments. K.Te. 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 authors contributed toward preparing the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. 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 the Supplementary Materials. Cartesian coordinates of **PC** species and molecules **V** optimized by B3LYP/6-31G\* level are also listed. The training dataset that we used in this work is located at [FL\\_ChemTS/data/](https://github.com/tsudalab/FL_ChemTS/) in our public GitHub site: [https://github.com/tsudalab/FL\\_ChemTS/](https://github.com/tsudalab/FL_ChemTS/). Our dataset and code were also archived at Zenodo (<https://doi.org/10.5281/zenodo.5716545>).

Submitted 11 May 2021

Accepted 19 January 2022

Published 9 March 2022

10.1126/sciadv.abj3906

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

Masato Sumita, Kei Terayama, Naoya Suzuki, Shinsuke Ishihara, Ryo Tamura, Mandeep K. Chahal, Daniel T. Payne, Kazuki Yoshizoe, and Koji Tsuda

*Sci. Adv.* **8** (10), eabj3906. DOI: 10.1126/sciadv.abj3906

### View the article online

<https://www.science.org/doi/10.1126/sciadv.abj3906>

### Permissions

<https://www.science.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of service](#)

---

*Science Advances* (ISSN 2375-2548) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.

Copyright © 2022 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).