Rediscovering Armstrong's quinoid theory with machine learning using quantum chemistry

Takehiro Fujita[1,†], Kei Terayama[2,3,†], Masato Sumita[3,4,†,*], Ryo Tamura[3-6], Yasuyuki Nakamura[1], Masanobu Naito[1], Koji Tsuda[3,5,6,*]

1. *Data-driven Polymer Design Group, Research and Services Division of Materials Data and Integrated System (MaDIS), National Institute for Materials Science (NIMS), 1-2-1, Sengen, Tsukuba, Ibaraki 305–0047, Japan*

2. *Graduate School of Medical Life Science, Yokohama City University, Tsurumi-ku 230–0045, Japan*

3. *RIKEN Center for Advanced Intelligence Project, Tokyo 103–0027, Japan*

4. *International Center for Materials Nanoarchitectonics (WPI-MANA), National Institute for Materials Science (NIMS), Tsukuba 305–0044, Japan*

5. *Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science (NIMS), Tsukuba 305-0044, Jpan*

6. *Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa 277–8561, Japan*

[†]*These authors contributed equally.*

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

Throughout the history of chemistry, human efforts to design functional molecules have caused the discovery of numerous theories. Recently, artificial intelligence (AI)-enabled de novo molecular generators (DNMGs) have automated molecular design based on data-driven or simulation-based property estimates, eliminating the need for chemical-theory-based guidelines. However, it is unclear whether these DNMGs can discover theories that elucidate molecular design and chemistry. Herein, we demonstrate that an AI-enhanced DNMG can discover a chemical theory regarding the molecular structure. We attempted to elucidate the theory used by the DNMG to generate pure organic molecules (consisting of H, C, N, and O) for absorbing long-wavelength light by observing the functional group enrichment of molecules during the density-functional-theory-based generation. Although the DNMG initially targeted diketone and aniline derivatives, it later focused on quinone derivatives with a long absorption wavelength, as if it found a rule for chemical constitutions relevant to colour, previously known as Armstrong's quinonoid theory, which claimed that the colour originates from 1,4-quinon derivatives. Additionally, the DNMG shows the potential of 1,2-quinone derivatives as chromophores, as demonstrated by our experimental validation by synthesising one mimetic generated molecule. This study confirms that DNMGs have the potential to discover and expand

chemical theories.

Introduction

Since the beginning of modern chemistry, the relationship between colour and molecular constitution has been extensively investigated.[1-3] These investigations had led to the discover of several kind of chemical compounds that were considered as origins of colour, among which aniline, azo, and quinone derivatives were representative compounds. These investigations have eventually guided several fundamental and industrial studies, such as those related to aromatic[4], near infrared[5], and dye molecules[6]. From the viewpoint of the molecular electronic structure theory based on quantum mechanics[7-9], colour formation in molecules is mainly attributed to light absorption mediated by the transition among quantised electronic structures in molecules. Based on this approach, a technique for tuning the light absorption of a molecule by changing the gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO) is established. This technique has accelerated the molecular-level structural design of photoelectronic devices[10,11] through facilitating charge generation in copy/print machines[12], information storage in optical disks[13], and development of photosensitiser materials for photodynamic therapy[14]. Recently, dye-sensitised photovoltaic cells that can harvest light with long wavelengths in the near-infrared (NIR) region—a key to sustainable life[6,15]. Simultaneously, several theories for designing NIR molecules have

been developed, focusing on narrowing the HOMO/LUMO gap results in a red shift in the absorption. Although there are several theories for designing molecular structures to tune the HOMO/LUMO gap, such as by increasing conjugate length[11,16], through bond alternation[17] and charge transfer[18], their implementation causes an increase in the molecular size. The relationship between colour and molecular structure is not fully understood; for example, it is still challenging to explore the limitation of absorption wavelength due to moderate molecular size.[10,19]

Recently, automated molecular design approaches based on theoretical simulations and data-driven approaches, combined with the use of de novo molecular generators (DNMGs) and molecular property estimators, have been intensively developed; these approaches have led to the successful production of molecules as good as or better than those imaged by human under limited circumstances.[20-23] Traditionally, functional organic molecules have been extensively researched and developed using chemists' knowledge (of chemical theories) industrially and academically.[12-14, 24-26] These molecules are typically developed on the basis of scaffold molecules through experimentation in accordance with professional knowledge and experience. However, the synthesis of molecules through this strategy depends on professional knowledge, and their performance is dependent on that of scaffold molecules. To circumvent the

shortcomings of traditional molecular design, inverse molecular design is desired.[27]

DNMGs, which are actively developed in informatics, facilitate inverse molecular design.

Despite the limitation of the target property of the DNMGs developed in informatics, any

molecular properties can become the target of the DNMGs.[28-31] We applied our developed

DNMG, ChemTS[32], to develop several functional materials[20,21,32], combining it with

electronic structure simulation. However, the chemical insights into the DNMG beyond

its application for the proposal of novel compounds have not been sufficiently obtained.

In this study, we investigated whether ChemTS, a DNMG that combines reinforcement

learning with deep learning, can discover chemical theories during the process of

exploring chemical space to optimise a single chemical property. ChemTS employs the

Monte Carlo tree search (MCTS) algorithm[33], one of the reinforcement learning methods

used in AlphaGo[34], to efficiently search for a target property in chemical space. Through

this tool, we generated more than 40,000 molecules with long absorption wavelengths,

with the goal of elongating the absorption wavelength of light, as evaluated by density

functional theory (DFT) calculations[8,9]. The history of modern chemistry starts from

distilling molecular constitutions that characterise the molecular properties of available

chemical compounds. In line with this, we examined the chemical knowledge learned by

ChemTS through the analysis of the molecules generated by ChemTS via functional

group enrichment analysis. From this analysis, we showed that the DNMG learned the potential of quinonoids as a suitable chromophore absorbing long-wavelength light. In the past, quinonoids were empirically considered to induce colouration, according to Armstrong's quinonoid theory.[1,2] We can conclude that ChemTS rediscovered this theory from the world of pure organic molecules whose properties were described at the time-dependent DFT (B3LYP/3-21G*) level. According to Armstrong's quinonoid theory, 1,4-quinon derivatives are responsible for colour formation. Additionally, ChemTS revealed the potential of 1,2-quinon derivatives. Even in modern chemistry, the 1,4-quinon moiety is regarded as one of the most important chromophores.[35,36] However, 1,2-quinon derivatives have not received considerable attention for their application as chromophores.[37] In this study, we report the discovery of chemical theories and their extension by artificial intelligence (AI) using deep learning and quantum chemical simulation, similar to how *joseki* was learned by AlphaGo.[38]

Results and discussions

Molecule generation

ChemTS succeeded in generating 45,321 organic molecules, using 2048 cores over 120 h with the goal of maximising the absorption wavelength of the generated molecules.

Figure 1(a) shows the evolution of the absorption wavelength as a function of the number of generated molecules. After generating 10,000 molecules, the average value of the absorption wavelength began to elongate and reached more than 600 nm. The maximum absorption wavelength was over 1,200 nm after generating 40,000 molecules. With an increase in the wavelength, the average HOMO/LUMO gap of the generated molecules monotonically decreases, as shown in Figure 1(b). Hence, there is an inversely proportional correlation between the absorption wavelength and the HOMO/LUMO gap. ChemTS also clearly uses this correlation and design molecules whose HOMO/LUMO gap is greater than 1.0 eV. However, the oscillator strength (OS) of the absorption does not grow and saturates around 0.05, which is not different from the average OS of the molecules included in the training data (Figure 1(c)). After 40,000 molecules are generated, the average value of OS is degraded, in contrast to the elongation of the absorption wavelength. This trend agrees with the intuition and statistical results that the organic molecules that absorb the long-wavelength light with high intensity are rare.[39] Hence, the relationship among electronic structures is reasonable. However, the relationship between the electronic structure and the molecular structure is not.

The design principle of ChemTS for molecules that absorb long-wavelength light is not dependent on the expanding molecular size. As shown in Figure 1(d), the average

molecular weight gradually increased until 40,000 molecules were generated. However, from the generation of 40,000 molecules, the average molecular weight decreased. This tendency can be observed in the average conjugate length and the number of aromatic rings, as shown in Figure 1(e) and 1(f), respectively. Up to the generation of 40,000 molecules, the absorption wavelength increases with increasing conjugation length and the number of aromatic rings (Figure 1(a)). This means that the traditional strategy of longer conjugation lengths leading to longer absorption wavelengths is working well[16], and ChemTS supports the same strategy. However, a slight deviation has occurred since the generation of 40,000 molecules. This means that ChemTS takes another strategy to elongate the absorption wavelength.
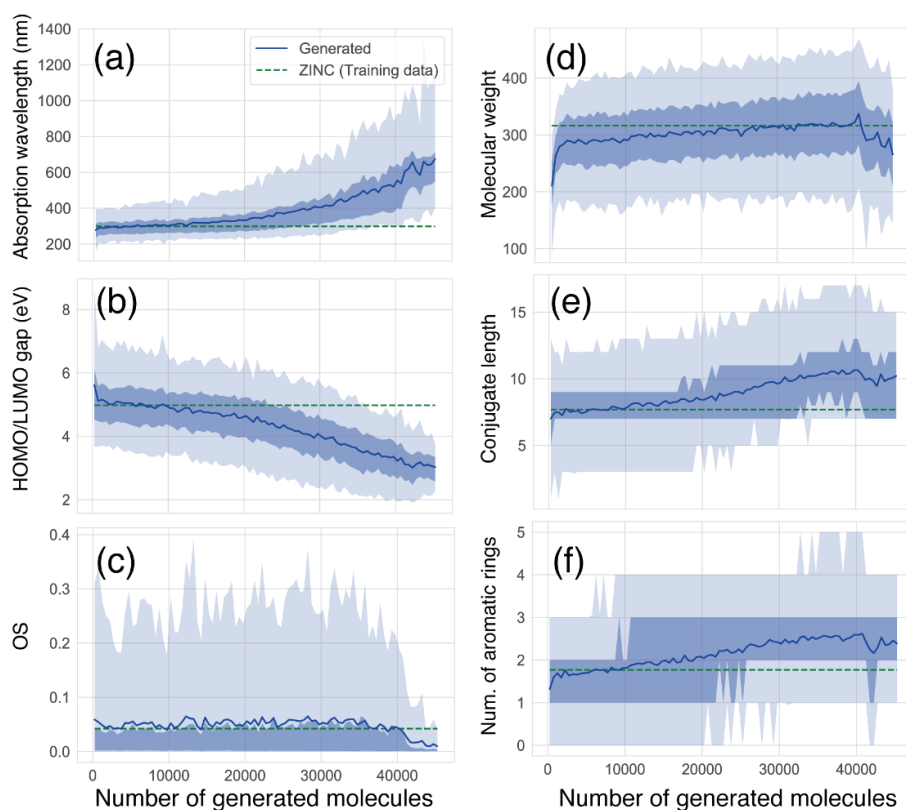
Figure 1. Evolution of several molecular properties as a function of the increment of generated molecules. (a) absorption wavelength (nm) to S_1 excited state, (b) HOMO/LUMO gap (eV) (c) absorption intensity (oscillator strength; OS), (d) molecular weight (g mol^{-1}), (e) conjugate length, (f) number of aromatic rings. Average values of ZINC and generated molecules at each step are depicted by green broken line and blue solid line, respectively. The shaded area depicts the distribution profiles of generated molecules for each property. A thin shade area represents 5%–95% of the total distribution, while a dense shade area represents 15%–75% of the total distribution in each number of generated molecules.
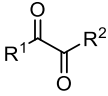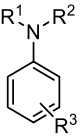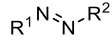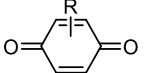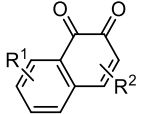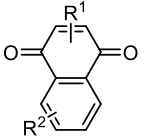
Functional group enrichment analysis

The percentages of several functional groups included in the whole generated molecules

and the training data and P_E values are listed in Table 1. Although ketone derivatives are

predominantly generated (~50%), their P_E value is not high (0.649). Similarly, the P_E of

traditional chromophore derivatives (azo, aniline) is not high. Hence, ChemTS does not

regard the azo and aniline derivatives as suitable molecules for absorbing long-

wavelength light. In contrast, 1,4-quinone shows a high P_E of > 30, despite the low

percentage of generated molecules. Among the generated molecules, 1,2-quinone shows

a relatively high odds ratio (Table 1). This result indicates that ChemTS predicted that

1,2-quinone is an important functional group for long-wavelength absorption. Dyes with

anthraquinone and 1,4-quinone structures are well known.[35,36] However, although 1,2-

quinone is considered a cofactor[41] and a building block in heterocyclic synthesis[40], it has

received little attention as a chromophore. Among the 1,2-quinone derivatives, the P_E of

1,2-naphthoquinone is the highest in Table 1, in contrast to 1,4-naphthoquinone, an

isomeric derivative of 1,2-naphthoquinone.

The evolution of P_E for the functional groups shown in Figure 2 indicates that ChemTS

focuses on the quinone derivatives as the chromophore for absorbing long-wavelength

light rather than ketones, diketones, and aniline. As shown in Figure 2(a)–2(c), ChemTS

insisted on ketone, diketone, and aniline derivatives from the initial stage to 20,000

molecule generation. From 20,000, however, P_E gradually decreased. Instead, the P_E of

azo [Figure 2 (d)] and quinone [Figure 2(e)–2(h)] derivatives gradually increased. In

particular, the P_E of azo, 1,4-quinone, and 1,2-quinone derivatives suddenly increased

after the generation of 40,000 molecules. This behaviour corresponds to the elongation of

the absorption wavelength, as shown in Figure 1(a). The P_E values of 1,4-quinone and

1,2-quinone were considerably higher than that of azo. This indicates that ChemTS

considers that 1,4-quinone is important for designing long-wavelength light absorption

chromophores. Compounds with 1,4-quinone and anthraquinone structures have been

known to serve as effective chromophores since B.C. After clarification of their molecular

composition, these compounds are known as the origin of colour according to

Armstrong's quinone theory. These compounds were industrially produced in the mid-

1800s. We concluded that ChemTS also found this theory by itself with the flamework of

DFT. Additionally, ChemTS has expanded Armstrong's quinone theory and predicted the

potential of 1,2-quinone derivatives after exploring the possibilities of ketone, diketone,

and aniline derivatives during the computation.

Table 1. Functional group enrichment analysis for various functional groups and their percentage of generated molecules and training data. Odds ratio is given as P_E.

Functional group	P_E	Generated mol. (%)	Training data (%)
Ketone	0.649	49.7	76.6
Diketone	0.375	0.847	2.26
Aniline	0.537	16.2	30.1
Azo	2.37	0.878	0.371
1,2-quinone	15.5	1.05	0.668
1,4-quinone	31.5	0.682	0.0217
1,2-naphthoquinone	78.5	0.0486	0.000619
1,4-naphthoquinone	0.938	0.0110	0.0118
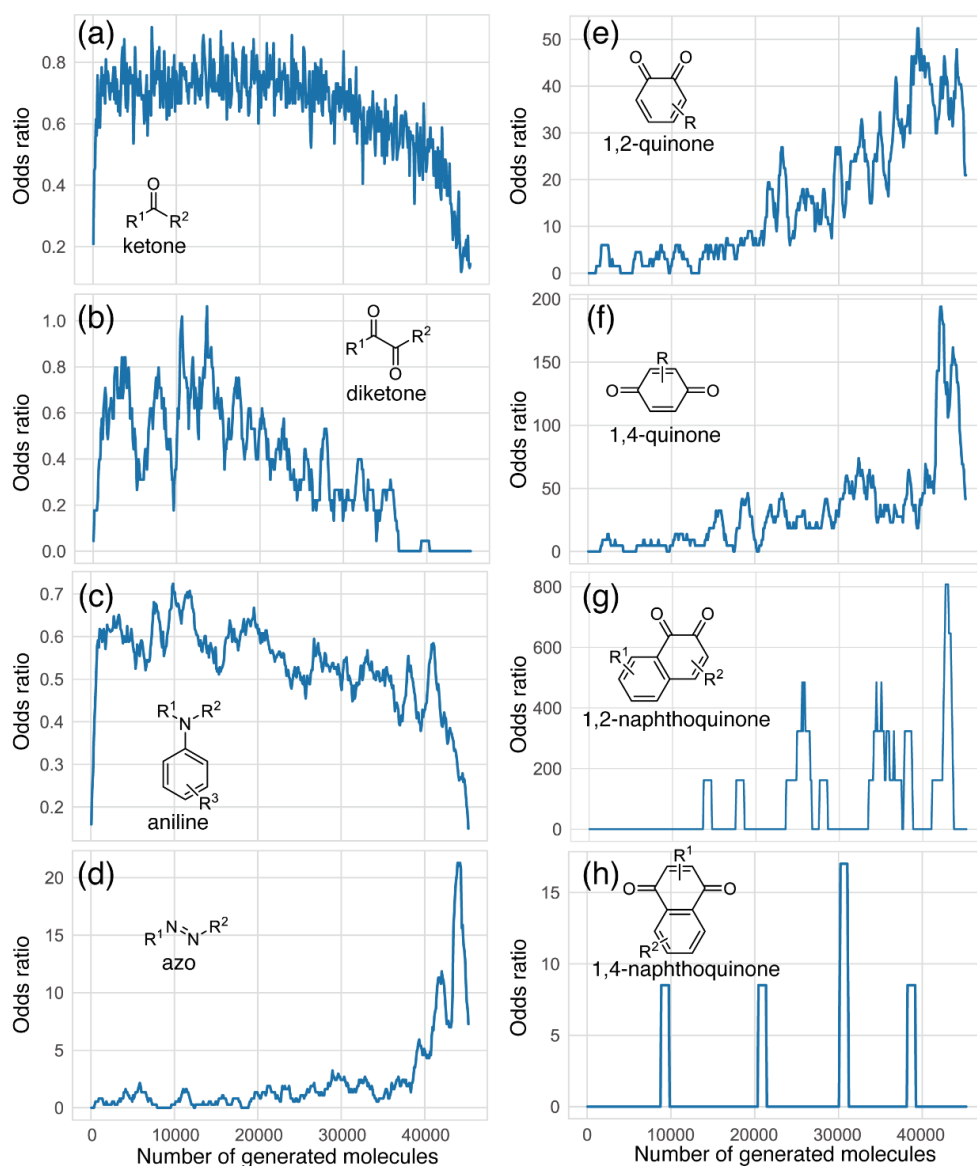
Figure 2. Odds ratio evolution of several functional groups shown in Table 1 as a function of the number of generated molecules. Odds ratios are computed for every 100 generated molecules.

Experimental validation

Among the 1,2-quinone derivatives, 1,2-naphthoquinone shows a high odds ratio of 78.5,

and its evolution behaviour [Figure 2(g)] is similar to that of 1,2-quinone, as shown in Figure 2(e). Hence, ChemTS learned the importance of 1,2-naphthoquinone compounds, despite the low percentage of training data. In total, 22 different 1,2-naphthoquinone structures were generated (see Figure S9 of the ESI). Among them, we focused on **1** in Figure 3(a), which showed the longest wavelength absorption of these molecules. Compound **1** has a structure consisting of an enol and a carbonyl skeleton attached to the 1,2-naphthoquinone skeleton; time-dependent DFT (TD-DFT) calculations at the APFD/6-31G* level (computational level is changed for conformation) predicted the appearance of the absorption at approximately 949 nm [Figure 3(a)]. For the actual synthesis, we simplified **1** by trimming the functional groups that would not be as important for their function as chromophores [Figure 3(a)]. First, **2** was obtained by replacing the moiety highlighted by red in **1** with hydrogen. According to our preliminary computation, TD-DFT calculations estimated that the absorption wavelength of **2** was approximately 820 nm, which is shorter than that of **1** by 130 nm. Second, **3** was obtained by replacing the triazole group in **2** with the phenyl group that also has an π conjugate structure because the triazole group is difficult to introduce. TD-DFT calculations estimated the absorption wavelength of **3** to be 757 nm. Although enol skeletons are thermodynamically unfavoured and isomerise to their keto forms, we tried synthesising

the enol but failed to isolate the desired compound. Accordingly, novel **4**, in which the enol skeleton of **3** was replaced by a simple olefin, was found as a synthesisable model of the target chromophore (**1**). The absorption wavelength of **4** to its first excited state owing to HOMO–LUMO single electron transition is estimated to be 575 nm, which undergoes a considerable blue shift from that of **1**; nonetheless, **4** is expected to be a chromophore. The nature of excitation to their first excited state is preserved (see Figure S10 of ESI) during the conversion of **1** to **4**.

The retrosynthetic analysis of **4** is shown in Figure 3(b). Compound **4** is prepared from 2-naphthol **5** by oxidation, which is synthesised from naphthol and an olefin through a cross-coupling reaction. As shown in Figure 4(a), using a commercially available compound (**6**) as the starting material, the Suzuki–Miyaura coupling reaction with vinylboronic acid pinacol ester gave **5**, which introduced the olefinic moiety into the naphthol skeleton.[42] Compound **5** was oxidised by 2-idoxybenzoic acid (IBX) to form the target compound (**4**) as a dark-purple solid with 95% yield.[43] Compound **4** was found to be stable in air and soluble in many common solvents such as $CHCl_3$, CH_2Cl_2, tetrahydrofuran, acetonitrile, and acetone. The product was characterised by NMR spectroscopy and mass spectrometry (see Figure S1-S4 of ESI). The solution of **4** in acetonitrile (1×10^{-5} mol L^{-1}) exhibited a red-purple colour; its UV-vis absorption

spectrum showed the first peak appearing at 550 nm, which agrees well with the predicted

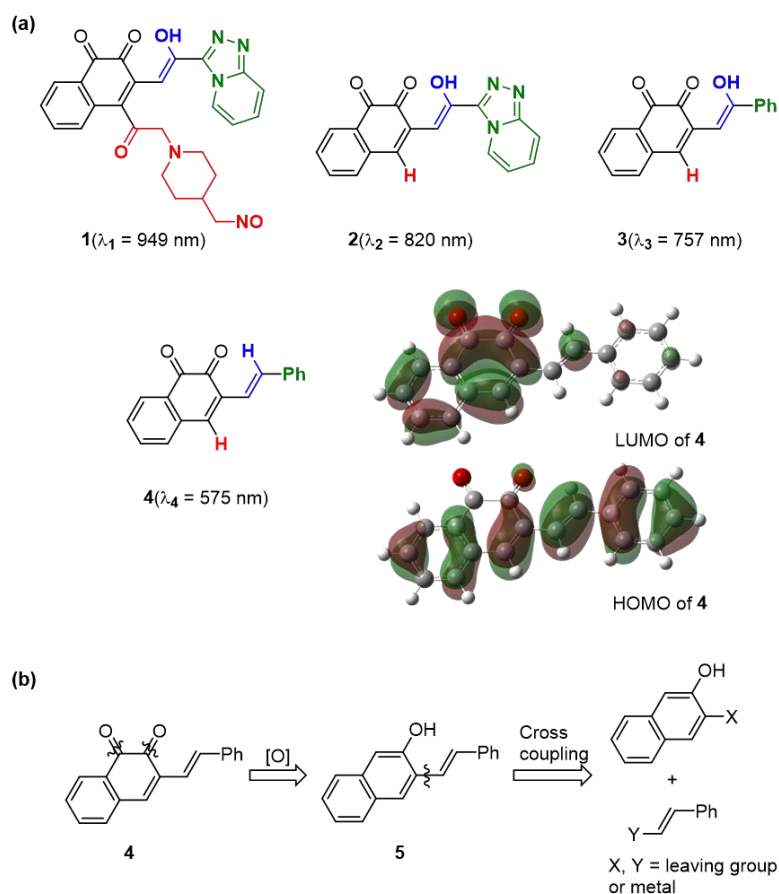result at the APFD/6-31G* level [Figure 4(b)].

Figure 3. Target molecule inspired by ChemTS. (a) The generated molecule by

ChemTS is **1**. **2**–**4** molecules are synthesis models of **1**. The absorption wavelength of

each molecule is estimated at the APFD/6-31G* level. Surfaces of HOMO–LUMO

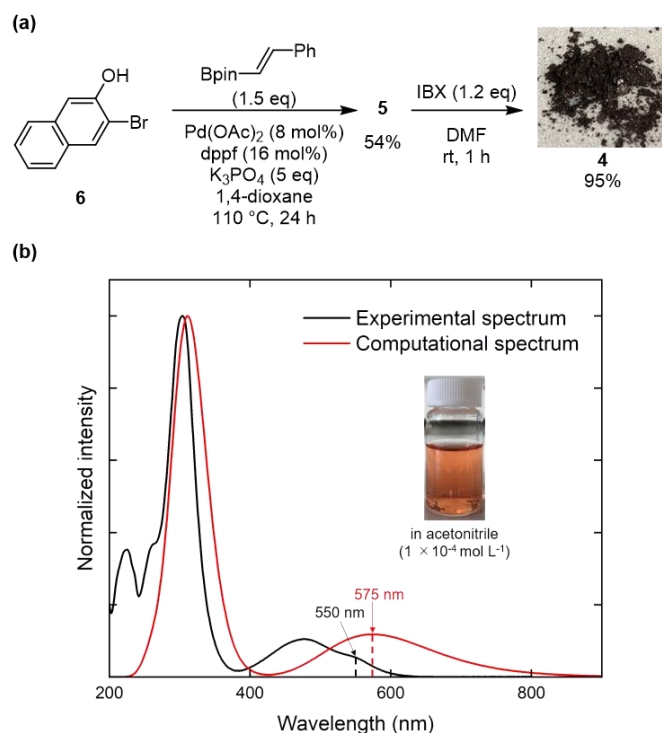orbitals of **4** are drawn at an isodensity value of 0.02. (b) Retro-synthesis of **4**.

Figure 4. Synthetic route and UV-vis spectrum of molecule 4. (a) Synthesis process of

4. (b) UV-vis absorption spectrum of the solution of acetonitrile **4** (1×10^{-5} mol L^{-1}) and

computational absorption spectrum of **4** obtained by TD-DFT calculation at the APFD/6-

31G* level. Photograph of a solution of **4** in acetonitrile (1×10^{-4} mol L^{-1}) under ambient

light is also shown.

Discussion

Many theories have been discovered in the history of chemistry through reproducibility

experiments or deductive methods based on physics (physicochemistry). As shown by

AlphaGo Zero[38], AI can discover *joseki* (tactics) without human knowledge and can be

trained at the superhuman level. This can also be applied in chemistry. We demonstrated

a novel approach to the theory discovery in chemistry by analysing molecules generated

by a deep-learning-based DNMG, ChemTS, coupled with DFT calculations. We used

ChemTS to generate pure organic molecules to maximise their absorption wavelength

under the constraints that atoms are limited to H, C, N, and O (pure organic molecules).

ChemTS succeeded in generating 45,321 molecules with absorption wavelengths of 1,200

nm at the B3LYP/3-21G* level. Functional group enrichment of molecules during

molecule generation exhibits the evolution of molecules from ketone and diketone

compounds to quinone and azo compounds with increasing absorption wavelength.

According to Armstrong's quinonoid theory, 1,4-quinones such as anthraquinone were

once considered as the origin of colour. ChemTS also focused on quinone derivatives by

elongating the absorption wavelength, in accordance with Armstrong's quinonoid theory.

Since the beginning of the research on colour formation and chemical constitution in 1856

by Perkin, the human being spent 32 years to achieve Armstrong's quinonoid theory.[1,2]

The deep-learning-based DNMG succeeded in simulating this evolution in just 120 h

through the aid of quantum chemistry. In addition to 1,4-quinone derivatives, functional

group enrichment analysis showed the potential of 1,2-naphthoquinone derivatives.

Moreover, we successfully verified the potential of the 1,2-naphthoquinone derivatives

by selecting one 1,2-naphthoquinone derivative generated by ChemTS and modelling it for the synthesis of a new red-purple compound whose UV absorption spectrum had the first peak at approximately 550 nm. Therefore, we can conclude that ChemTS not only rediscovered Armstrong's quinonoid theory but also simulated this theory in just 120 h and predicted that 1,2-naphthoquinone derivatives are suitable for long wavelength absorption molecules rather than 1,4-naphthoquinone derivatives, which known as representative chromophores.

Method

Molecule generator

To elongate the absorption wavelength of pure organic molecules consisting of C, N, O, and H atoms, we used our developed DNMG, ChemTS, which employs the MCTS[33] algorithm and the recurrent neural network (RNN)[44]. The RNN model was trained with a set of 153,253 SMILES[45] strings that only consisted of H, O, N, and C elements obtained from the ZINC database[46]. After translating SMILES to Cartesian coordinates using the RDKit package[47], the absorption wavelength of each generated molecule was computed using TD-DFT at the B3LYP/3-21G* level, implemented in the Gaussian 16 package[48]. The lowest 10 states for each molecule were calculated after geometry optimisation. We

used the following reward function, $r(I)$, of a generated molecule, I, in the MTCS-based

search process:

$$r(I) = F(I) * G(I), \quad (1)$$

$$F(I) = \frac{\tanh(0.003(\lambda_I - \theta))}{2}, \quad (2)$$

$$G(I) = \frac{-\tanh(SA_I - 4) + 1}{2}. \quad (3)$$

The reward function consists of the product of a term relating to the absorption

wavelength, $F(I)$, and a term relating to synthesizability, $G(I)$. $F(I)$ takes values

between 0 and 1; the longer the computational wavelength, λ_I, of molecule I, the larger

the value. θ is the comparative criterion, which is set to 400 nm in this study. For a

wavelength of 400 nm, the value of $F(I)$ is 0.5. $G(I)$ also takes values between 0 and 1

and is calculated from the synthetic accessibility score, SA_I, of molecule I, which predicts

the difficulty of the synthesis. To accelerate the MCTS search, we adopted the virtual loss

strategy to parallelise the computation. We used the following score (u_i) for each child

node i in the selection step:

$$u_i = \frac{tR_i}{v_i + vl_i} + CP_i \frac{\sqrt{v_p + vl_p}}{v_i + vl_i + 1}. \quad (4)$$

Here, tR_i is the total reward of node i, v_i is the total number of visits to node i, and

vl_i is the total number of virtual visits of i (virtual loss). C is the search parameter,

which controls the exploration–exploitation trade-off, and is set to 2 in this study. Without

depending on C, the discussion in the next section were almost valid (see the ESI for $C = 2$ and 4). P_i is the probability of node i, which is calculated using the RNN model. v_p and vl_p are the total number of visits and the total number of virtual visits of parent node p of child node i. See Ref. 1 for details of the virtual loss.

To clarify the design principle of ChemTS, we performed functional group enrichment analysis using odds ratio (P_E) for long-wavelength absorption. P_E of a fraction of molecule, f, is calculated as

$$P_E(f) = {P(f)} \big/ {P_t(f)}, \qquad (5)$$

where $P(f)$ is the fraction of molecules containing in the generated molecules, and $P_t(f)$ is the fraction of molecule containing in the training set. Hence, the higher value of P_E is the more important fraction for the DNMG.

Data availability

Chemical formulae of 1,2-naphthoquinone compounds generated by ChemTS, computational validation of the model molecules at a high level, and details of the chemical synthesis and characterisation of products are available in the Supplementary Information. The generated 45,321 molecules are listed in the ALW_ChemTS/generated_mols/result.csv file. The file contains following information:

generated molecules (SMILES), calculated absorption wavelengths and their oscillator strengths, and basic information such as molecular weight.

Code availability

Our implementation of parallel version of ChemTS for fluorescent molecules is available at https://github.com/tsudalab/ALW_ChemTS.

Acknowledgements

Author contributions

T. F., Y. N., and M. N. selected molecules from the generated molecules and planned the synthesis strategy. T. F. performed the chemical experiments. K. Terayama and M. S. performed the computational experiments. M.S., R.T., and K. Tsuda planned and supervised the project. All members contributed to the preparation of the manuscript.

Competing interests

The authors declare no competing interests.

References

1. Travis, A. S. Colour by design before 1890. *Text. Chem. Colorist* **22**, 23-27 (1990)

2. Maccoll, A. Colour and constitution. *Q. Rev. Chem. Soc.* **1**(1), 16-58 (1947). 10.1039/qr9470100016

3. Kanetkar, V. R. Colour: History and advancements. *Resonance* **15**(9), 794-803 (2010). 10.1007/s12045-010-0089-2

4.   Qin, Y., Li, G., Qi, T. & Huang, H. Aromatic imide/amide-based organic small-molecule emitters for organic light-emitting diodes. *Mater. Chem. Front.* **4**(6), 1554-1568 (2020). 10.1039/D0QM00084A

5.   Fabian, J., Nakazumi, H. & Matsuoka, M. Near-infrared absorbing dyes. *Chem. Rev.* **92**(6), 1197-1226 (1992). 10.1021/cr00014a003

6.   Hagfeldt, A., Boschloo, G., Sun, L., Kloo, L. & Pettersson, H. Dye-sensitized solar cells. *Chem. Rev.* **110**(11), 6595-6663 (2010). 10.1021/cr900356p, 20831177

7.   Atkins, P., De Paula, J. & Keeler, J. *Atkins, Physical Chemistry*. 11th ed (Oxford University Press, 2017)

8.   Parr, R. G. & Yang, W. *Density-Functional Theory of Atoms and Molecules* (Oxford University Press, 1989)

9.   Cohen, A. J., Mori-Sánchez, P. & Yang, W. Challenges for density functional theory. *Chem. Rev.* **112**(1), 289-320 (2012). 10.1021/cr200107z, 22191548

10.   Qian, G. & Wang, Z. Y. Near-infrared organic compounds and emerging applications. *Chem. Asian J.* **5**(5), 1006-1029 (2010). 10.1002/asia.200900596, 20352644

11. Cao, X., et al. An extremely narrow bandgap conjugated polymer for photovoltaic devices covering UV to near-infrared light. *Dyes Pigm.* **158**, 319-325 (2018). 10.1016/j.dyepig.2018.05.052

12. Law, K. Y. Organic photoconductive materials: Recent trends and developments. *Chem. Rev.* **93**(1), 449–486 (1993). 10.1021/cr00017a020

13. Mustroph, H., Stollenwerk, M. & Bressau, V. Current developments in optical data storage with organic dyes. *Angew. Chem. Int Ed Engl* **45**(13), 2016-2035 (2006). 10.1002/anie.200502820, 16518782

14. Swamy, P. C. A., et al. A. Near infrared (NIR) absorbing dyes as promising photosensitizers for photodynamic therapy. *Coord. Chem. Rev.* **411** (2020). 213233

15. Ferdowsi, P., et al. Molecular engineering of simple metal-free organic dyes derived from triphenylamine for dye-sensitized solar cell applications. *ChemSusChem* **13**(1), 212-220 (2020). 10.1002/cssc.201902245, 31592574

16. Guay, J., et al. Chain-length dependence of electrochemical and electronic properties of neutral and oxidized soluble .alpha.,.alpha.-coupled thiophene oligomers. *Chem. Mater.* **4**(5), 1097-1105 (1992). 10.1021/cm00023a031

17. Sezukuri, K., et al. A laterally expanded fluorone dye as an efficient near infrared fluorophore. *Chem. Commun.* **52**(27), 4872-4875 (2016). 10.1039/C6CC00237D

18. Kawano, S. I., et al. Near-infrared absorption by intramolecular charge-transfer transition in 5, 10,15,20-tetra(N-carbazolyl)porphyrin through protonation. *Chem. Commun. (Camb)* **55**(20), 2992-2995 (2019). 10.1039/c8cc09667h, 30785132

19. Bredas, J. L., Silbey, R., Boudreaux, D. S. & Chance, R. R. Chain-length dependence of electronic and electrochemical properties of conjugated systems: polyacetylene, polyphenylene, polythiophene, and polypyrrole. *J. Am. Chem. Soc.* **105**(22), 6555-6559 (1983). 10.1021/ja00360a004

20. 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**(9), 1126–1133 (2018). 10.1021/acscentsci.8b00213, 30276245

21. Zhang, Y., et al. Discovery of polymer electret material via de novo molecule generation and functional group enrichment analysis. *App. Phys. Lett.* **118** (2021). 223904

22. Zhavoronkov, A. et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* **37**(9), 1038–1040 (2019). 10.1038/s41587-019-0224-x, 31477924

23. Terayama, K. Sumita, M. Tamura, R., Tsuda, K. Black-Box Optimization for Automated Discovery, *Acc. Chem. Res.* **54**, 1334-1346 (2021). 10.1021/acs.accounts.0c00713

24. Nakanotani, H., Tsuchiya, Y. & Adachi, C. Thermally-activated delayed fluorescence for light-emitting devices. *Chem. Lett.* **50**(5)**,** 938-948 (2021). 10.1246/cl.200915

25. Leonhardt, E. J. & Jasti, R. Emerging applications of carbon nanohoops. *Nat. Rev. Chem.* **3**(12), 672-686 (2019). 10.1038/s41570-019-0140-0

26. Kang, B., Lee, W. H. & Cho, K. Recent advances in organic transistor printing processes. *ACS Appl. Mater. Interfaces* **5**(7), 2302-2315 (2013). 10.1021/am302796z, 23446358

27. Sanchez-Lengeling, B. & Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **361**(6400), 360–365 (2018). 10.1126/science.aat2663, 30049875

28. Merk, D., Grisoni, F., Friedrich, L. & Schneider, G. Tuning artificial intelligence on the de novo design of natural-product-inspired retinoid X receptor modulators. *Commun. Chem.* **1**(1), 68 (2018). 10.1038/s42004-018-0068-1

29. Wu, S. *et al.* Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. *npj Comput. Mater.* **5**(1), 66 (2019). 10.1038/s41524-019-0203-2

30. Kajita, S., Kinjo, T. & Nishi, T. Autonomous molecular design by Monte-Carlo tree search and rapid evaluations using molecular dynamics simulations. *Commun. Phys.* **3**(1), 77 (2020). 10.1038/s42005-020-0338-y

31. 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**(1), 972–976 (2017). 10.1080/14686996.2017.1401424, 29435094

32. Zhang, J. *et al*. NMR-TS: De novo molecule identification from NMR spectra. *Sci. Technol. Adv. Mater.* **21**(1), 552-561 (2020).

    10.1080/14686996.2020.1793382, 32939179

33. Browne, C. B. *et al.* A Survey of Monte Carlo tree search methods. *IEEE Trans. Comput. Intell. AI Games* **4**(1), 1–43 (2012). 10.1109/TCIAIG.2012.2186810

34. Silver, A. *et al.* Mastering the game of Go with deep neural networks and tree search. *Nature* **529**(7587), 484-489, (2016).

35. Griffiths, J. *Ullmann's Encyclopaedia of Industrial Chemistry* 349-354 (Wiley-VCH Verlag GmbH & Co. KGaA, 2012).

36. Dulo, B., Phan, K., Githaiga, J., Raes, K. & Meester, S. D. Natural quinone dyes: A review on structure, extraction techniques, analysis and application potential. *Waste Biomass Valorization* 1-36 (2021). 10.1007/s12649-021-01443-9

37. Ros-Lis, J. V., Martínez-Máñez, R., Benito, A. & Soto, J. Naphthoquinone derivatives as receptors for the chromogenic sensing of metal cations and anions. *Polyhedron* **25**(7), 1585-1591 (2006). 10.1016/j.poly.2005.10.034

38. Silver, D. *et al.* v. Silver, D. *et al.* Mastering the game of Go without human knowledge. *Nature* **550**(7676), 354-359 (2017). 10.1038/nature24270, 29052630

39. Terayama, K. *et al.* Pushing property limits in materials discovery via boundless objective-free exploration. *Chem. Sci.* **11**(23), 5959-5968 (2020). 10.1039/d0sc00982b, 32832058

40. Wendlandt, A. E. & Stahl, S. S. Quinone-catalyzed selective oxidation of organic molecules.

Angew. Chem. Int Ed Engl **54**(49), 14638-14658 (2015). 10.1002/anie.201505017,

26530485

41. Mishra, A. K., Mukhopadhyay, A. & Moorthy, J. N. One-pot multistep synthesis

of bipolar carbazolo-phenazines: Hydrogen bond control of Diels-Alder

cycloaddition and application for fluoride sensing. *Tetrahedron* **73**(16), 2210-2216

(2017). 10.1016/j.tet.2017.02.052

42. Parasram, M., Iaroshenko, V. O. & Gevorgyan, V. Endo-selective Pd-catalyzed

silyl methyl Heck reaction. *J. Am. Chem. Soc.* **136**(52), 17926-17929 (2014).

10.1021/ja5104525, 25494921

43. Magdziak, D., Rodriguez, A. A., Van De Water, R. W. W. D. & Pettus, T. R. R.

Regioselective oxidation of phenols to *o-quinones* with o-iodoxybenzoic acid

(IBX). *Org. Lett.* **4**(2), 285-288 (2002). 10.1021/ol017068j, 11796071

44. Cho, K. Learning phrase representations using RNN encoder-decoder for

statistical machine translation. Proceedings of the 2014 Conference on Empirical

Methods in Natural Language Processing. *EMNLP* **2104***, 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**(1), 31-

36 (1988). 10.1021/ci00057a005

46. Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S. & Coleman, R. G. Zinc: A free tool to discover chemistry for biology. *J. Chem. Inf. Model.* **52**(7), 1757-1768 (2012). 10.1021/ci3001277, 22587354

47. Landrum, G. RDKit: Open-source cheminformatics, Available from: http://www.rdkit.org

48. Frisch, M. J. *et al.*, Gaussian 16, Revision B.01, Gaussian, Inc., Wallingford CT, 2016.