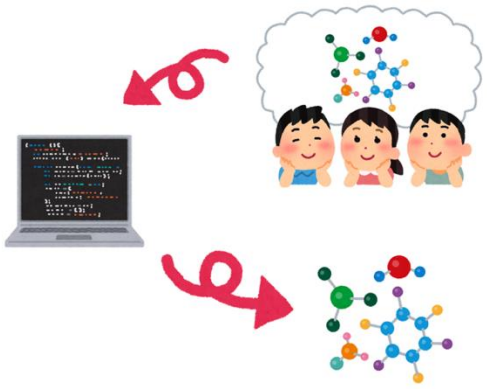
Automatic creation of new materials

Akinori Matsumoto

What would you do if you had a machine which could create new material with the properties you desire? You may think such a convenient machine is impossible, but something similar may become possible in the near future. Recently a group of researchers has succeeded in creating unreported molecules emitting fluorescence, just as they were intended to [1].

Fluorescent molecules have a variety of applications in our daily lives. For example, fluorescent molecules are useful in the medical field. By connecting fluorescent molecules to living cells, cellular activity can be directly viewed under a microscope. This makes it easier to observe the effects of new disease treatments in experiments on mice [2].



In addition, fluorescent molecules are beginning to be used as a new material for displays to replace liquid crystals. The displays made from fluorescent molecules not only have the advantage of emitting light without heat, similar to liquid crystals but also have other advantages such as flexibility and lower power consumption [3].

Although many types of fluorescent molecules have already been utilized, they often have drawbacks in terms of function, sustainability, and cost. To overcome such shortcomings, new fluorescent molecules have always been sought after. Then, how can we create new fluorescent molecules?

To answer this, let's first see how fluorescent molecules emit light. A molecule is a combination of many atoms that exhibit a single property. Based on the combination, several 'states', each of which corresponds to a certain energy range exist. A molecule at a state has energy in the corresponding range. When a photon strikes a fluorescent molecule, the molecule moves to a state with a range of higher energy. Then the molecule loses energy to the minimum value allowed in the new state. Eventually, the molecule returns to its original state with lower energy. On the transition from the 'new' state to the original state, a photon with the same energy as the energy difference between the states, i.e., fluorescence, is emitted. What we have to do is to find a new stracture that this fluorescence mechanism works.

However, this is not an easy task at all because we cannot determine whether the fluorescence emission occurs from the molecular structure alone. To confirm that a candidate of a fluorescent molecule really emits fluoresce requires extensive simulations. Therefore, the conventional approach has been to create a new fluorescent molecule based on the structure of a known fluorescent molecule. However, a method has recently been found that creates perfectly new fluorescent molecules almost automatically.

How on earth can we do that? The first step is to represent structures of fluorescent molecules as a single string by a method called SMILES (simplified molecular-input line-entry system). With the SMILES strings, we can represent any structures of molecules, including the types of bonds between atoms, as a one-dimensional sequence

of symbols. For example, “r, [, C, @, H,], (, C, l,), F” in the SMILES represents bromochlorofluoromethane, a molecule with a three-dimensional structure. Though fluorescent molecules generally have a very complex three-dimensional structure, they can be expressed with only a one-dimensional sequence of symbols. In this manner, we obtain a series of one-dimensional sequences of symbols for a huge number of known fluorescent molecules.

The second step is to build a model which predicts a SMILES string for an unknown fluorescent molecule from the data of the SMILES strings of the known fluorescent molecule. This can be done with an algorithm called a recurrent neural network (RNN). The RNN is a class of artificial neural networks, which handles an ordered sequence of data. Thanks to the SMILES, we can treat structures of molecules as machine-readable information and the RNN is well suited to handle this one-dimensional information.

Now that we are finally ready to create a string in SMILES of fluorescent molecules. Using a computer, we arrange symbols in SMILES almost randomly starting from the first letter checking if they are likely to be fluorescent molecules. Here I say "almost randomly" because the computer does not create strings completely at random, but rather it determines and focuses on sequences of symbols that are likely to be fluorescent molecules. This is based on the method used by computers to find out the best moves in chess. In the case of chess, a computer examines “almost randomly” possible moves and finds the next move that is likely to win the game. In the case of creating fluorescent molecules, the next move of chess corresponds to the next symbol in SMILES.

However, the options for arranging symbols are so numerous that it is not possible to arrange them in this way down to the last symbol. So, after arranging the symbols to some extent, we apply the trained RNN we have just prepared to the string. Thanks to training the RNN, it refers to the first half of the string and completes the rest of the string to be a fluorescent molecule. Finally, we have a candidate for the fluorescent molecule.

By this method, the group of researchers prepared 8 candidates of fluorescent molecules which can actually be synthesized. The researchers synthesized these molecules and found that six of the eight molecules were actually fluorescent molecules.

This success is very valuable. First, the newly synthesized fluorescent molecules themselves may have medical and industrial applications as I mentioned above. Another important thing is that the molecules with the desired property were created almost automatically. Creating such molecules has been difficult. Though the researchers do not discuss in their paper whether the new fluorescent molecules they have synthesized have an advantage compared with those that have already been utilized, we will be able to create more new fluorescent molecules, among which such useful molecules may exist. Furthermore, it will be possible to create new molecules with useful properties other than fluorescent molecules. In the future, a world may come true in which we can freely create materials with any properties we like.

References

[1] Masato Sumita, Kei Terayama, Naoya Suzuki, Shinsuke Ishihara, Ryo Tamura, Mandeep K. Chahal, Daniel T. Payne, Kazuki Yoshizoe, Koji Tsuda, "De novo creation of a naked eye–detectable fluorescent molecule based on quantum chemical computation and machine learning", Science Advances, 10.1126/sciadv.abj3906

[2] Mahadeva M. M. Swamy, Yuta Murai, Kenji Monde, Setsuko Tsuboi, and Takashi Jin, "Shortwave-Infrared Fluorescent Molecular Imaging Probes Based on π-Conjugation Extended Indocyanine Green", Bioconjugate Chemistry, 10.1021/acs.bioconjchem.1c00253

[3]キヤノングローバル. “有機 EL のしくみと特徴”. キヤノンサイエンスラボ・キッズ. https://global.canon/ja/technology/kids/mystery/m_04_12.html, (accessed 2022-06-07)