

Mathematical design of molecular self-assembly

Molecular self-assembly is a phenomenon in which multiple molecular subunits autonomously assemble into an ordered structural form or pattern without any external direction. Self-assembled monolayer (SAM) films and mesoporous materials are the well-established examples that harness the mechanism of molecular self-assembly. Molecular machines, which were the topic of the 2016 Nobel Prize in Chemistry, are also closely related to molecular self-assembly. Accordingly, many people are increasingly likely to recognize the word “self-assembly” as a topic in chemistry. The concept of self-assembly has, however, never been limited to chemistry. Indeed many aspects of biological activity inside a cell can be appreciated as the ultimate form of molecular self-assembly. We have a lot to learn from such biological systems to realize more advanced technology. Unfortunately, it is not yet common to discuss the relationship between chemical and biological self-assembly. This is presumably because the number of subcomponents participating in self-assembly substantially differs in chemical and biological self-assembly. More specifically, the maximum number of subcomponents that can be handled in a chemically designed system is typically about ten, with the maximum number of components reported in the literature only being on the order of 10. In nature, on the other hand, there are many self-assembly systems that assemble hundreds of subcomponents even as discrete molecules. For instance, the capsid structure of Bluetongue virus shown in Fig. 1 consists of 900 protein subunits. Such a level of self-assembly has been a daunting challenge and beyond the reach of chemists. What we have achieved in our recent work is a proposal and demonstration of a geometric design strategy that enables self-assembly with a large number of components, which had previously been unachievable. It is not common to exploit the knowledge of mathematics in synthetic chemistry, but such a new

approach was found to be effective. A mathematical description based on a geometric theorem clearly explained a previously reported self-assembly. Moreover, we succeeded in actually synthesizing a structure that was predicted by the theory. This is a structure self-assembled from 144 components, which is the largest self-assembled supramolecular cage ever reported. The methodology and the synthesis were described in Ref. 1. The following paragraphs summarize the results of the research.

The greatest challenge in a designing multi-component self-assembly is the “dispersion” of the number of components participating in the assembly. Similarly to micelles, vesicles, and polymers, upon assembling subunits, the dispersion of the number of subunits, and accordingly the dispersion of the resultant structure, is generally unavoidable. Such dispersion is a major obstacle to designing and implementing a precise function. Nonetheless, this issue can somehow be overcome in nature to achieve an accurate 900-component assembly, as shown in Fig. 1. How can it be? We focused on the fact that most spherical virus capsids have icosahedral symmetry. In geometry, restrictions such as congruity of the shape of faces or vertices significantly limit the number of possible polyhedra, for example regular and semiregular polyhedra are limited to only 5 and 13 species, respectively. Our hypothesis was as follows. If we can embed the nature of polyhedra within a system of a molecular assembly, only discrete products with limited numbers, sizes, and shapes of components will be realizable, thus enabling a molecular assembly without structural dispersion. We tested this hypothesis using a self-assembly system, in which bent ditopic organic ligands (L) and divalent palladium ions with square planar coordination sites (M) assemble into discrete M_nL_{2n} metal complexes [2]. The bend angle of L is the key to obtaining closed spherical structures rather than infinite networked structures (Fig. 2(a)). If we view the positions of metal ions (M) as vertices and those of organic ligands (L) as edges, each of the previously reported self-assembled structures can be recognized as semiregular polyhedra. In fact, when we fix the valency of vertices to four, which originates from the square planar coordination nature of palladium ions, only one of the regular polyhedra and four of the semiregular polyhedra fulfill the condition. Three of the five structures had been synthesized and the other two structures, the icosidodecahedron and rhombicosidodecahedron, were unprecedented. We

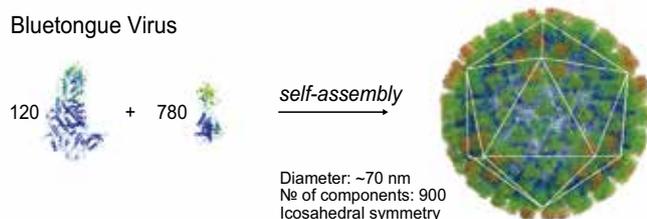


Fig. 1. Bluetongue virus. An example of a self-assembled molecule in nature.

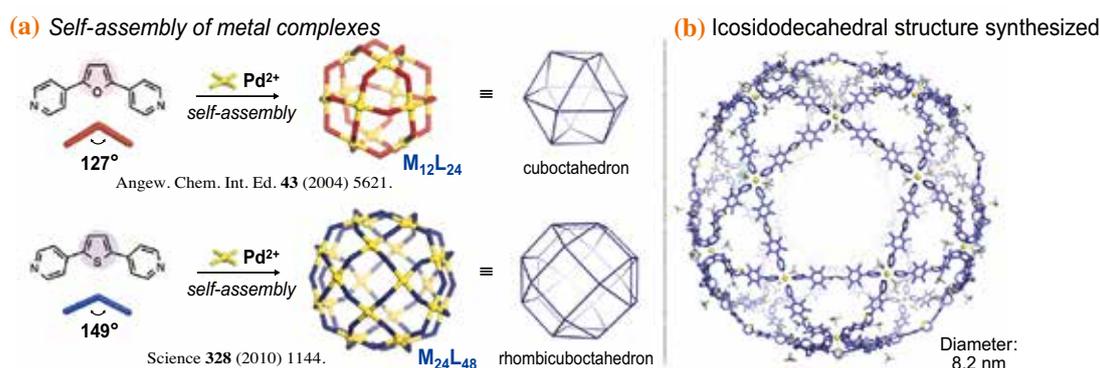


Fig. 2. Components and assembled structures of polyhedra.

then targeted the realization of the icosidodecahedron structure. If our hypothesis is correct, the fabrication of the icosidodecahedron structure will be challenging but theoretically feasible. We lack space to discuss the details here; however, by carefully tuning the structure of the organic ligand on the basis of the polyhedral design, we succeeded in synthesizing the icosidodecahedron structure at beamlines **BL38B1** and **BL41XU** in SPring-8 and at BL1A in Photon Factory (Fig. 2(b)) [3].

This approach based on the knowledge of polyhedra was further developed by combining it with more detailed mathematical discussion. The preceding discussion was limited to regular and semiregular polyhedra without a convincing theoretical justification. Also, the designable structures have an upper limit in terms of the number of subcomponents. We thus reconstructed the design theory based on the deduction from two chemical requirements: 1) all the vertices are tetravalent — a palladium ion has square-planar coordination sites, 2) all the edges are equilateral — the chemical structure of organic

ligands remains unchanged on assembly, and one symmetrical postulation. As a result, by defining the $F(h, k)$ index ($h, k = \text{natural numbers}$) and the Q value ($Q = h^2 + k^2$) as a representative value, we realized that it clearly described the reported products of the M_nL_{2n} self-assembly without any defect or overlap. (Mathematically, such a discussion corresponds to the tetravalent Goldberg-Coxeter construction in combination mathematics.) This discussion is not limited to the analysis of observed phenomena and can be exploited as a design tool for creation. As predicted from the Q value, we succeeded in synthesizing an $M_{48}L_{96}$ structure consisting of 144 subcomponents, which is the largest synthetic supramolecular structure ever reported (Fig. 3) [1].

The methodology we developed has opened up a new landscape in the field of molecular self-assembly, where no logical synthetic strategy previously existed. It revealed that a mathematical approach is beneficial for predicting or designing unknown self-assembled structures. We expect that this research will accelerate the development of the field of molecular self-assembly.

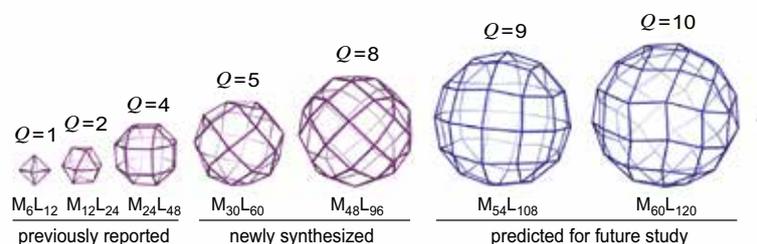


Fig. 3. The family of M_nL_{2n} polyhedra.

Daishi Fujita and Makoto Fujita*

Department of Applied Chemistry, The University of Tokyo

*Email: mfujita@appchem.t.u-tokyo.ac.jp

References

- [1] D. Fujita, Y. Ueda, S. Sato, N. Mizuno, T. Kumasaka, M. Fujita: *Nature* **540** (2016) 563.
- [2] K. Harris *et al.*: *Chem. Commun.* **49** (2013) 6703.
- [3] D. Fujita, Y. Ueda, S. Sato, H. Yokoyama, N. Mizuno, T. Kumasaka, M. Fujita: *Chem* **1** (2016) 91.