Self-assembly is a process in which multiple subcomponents spontaneously assemble into larger molecular structures. This approach to creating large structures is advantageous in terms of the economy of design information, i.e., the resulting self-assembled structures are determined by the information encoded in their components. While the scientific community has long recognized the practical utility of such a design strategy, scientists, in comparison to Nature, are still in the initial developing stages with respect to understanding mechanisms of self-assembly. Although we have been successful at self-assembling structures from a relatively small number of components, increasing the number of subcomponents to 100 or more has been extraordinarily challenging. The intrinsic difficulty derives from entropy, in that the ratio of microstates resulting in the target structure in comparison to the vast number of microstates that do not, decreases dramatically as the number of components required for the target increases. Such is the seemingly inevitable entropic barrier we face, and so the design of self-assembled structures composed of very large numbers of highly regulated and repeating components has yet to be realized in the lab. Nonetheless, Nature can somehow overcome this issue. How can this be? In addressing this fundamental question, I focused on the fact that most virus cages have icosahedral symmetry and hypothesized that this polyhedral symmetry is one of the most crucial design features to consider. Through the PRESTO study, I propose a new design concept for the self-assembly of structures with a large number of components based on the mathematics of graph theory, wherein graphs are equivalent to different polyhedron shapes. The core results have been published in Nature (attached); the applicant was the first and co-corresponding author therein.

More Details: In the last decade or two, chemists have constructed several artificial analogs of natural self-assembled structures using reversible metal-ligand coordination bonds. We have been developing a self-assembly motif in which n Pd(II) ions (M) and $2n$ bent bridging ligands (L), that possess two binding sites for the metal ions, self-assemble into a family of discrete M_nL_{2n} polyhedra, and to date those with n=6,12,24 have been synthesized. Recently, I serendipitously observed the self-assembly of a new polyhedral solid $M_{30}L_{60}$. This large metal complex consists of a simple combination of eight triangles and 24 squares, seemingly like an Archimedean solid, but does not belong to this family of shapes. Combining this new finding and the reported fact of the virus capsid shapes, I succeed to build a clear-cut theory that generalize polyhedral molecular self-assembly based on mathematical graph theory. As a demonstration, we predicted the self-assembly of higher tetravalent Goldberg polyhedra — a topology which has been abstracted in mathematical graph theory but has never been reported at the molecular level either in nature or the laboratory — and, after many synthetic attempts, I discovered the self-assembly of the predicted topology with $M_{48}L_{96}$ composition. The extended Goldberg series of polyhedra provides an indefinite number of ever-greater structures, so in principle there is no intrinsic mathematical limit to size. There are two important scientific significations in this achievement: 1) Mathematics can be a powerful tool that guides self-assembly system design; 2) Molecular self-assembly is theoretically scalable to a very large number of components. The M_nL_{2n} family is the first class of synthetic assemblies that are capable of encapsulating whole protein molecules. The project is now expanding as a chemical-biological hybrid as well.