Universität Basel
Swiss Nanoscience Institute

SNI SWISS NANOSCIENCE INSTITUTE
EINE INITIATIVE DER UNIVERSITÄT BASEL UND DES KANTONS AARGAU

Machine learning assisted design of heteromeric self-assembled molecular capsules

PI: Konrad Tiefenbacher, Department of Chemistry, University of Basel/DBSSE, ETH Zürich; konrad.tiefenbacher@unibas.ch
CoPI: Anatole von Lilienfeld, Department of Chemistry, University of Basel; anatole.vonlilienfeld@unibas.ch

1. Current state of research in the field:

Catalysis in hydrogen bond-based supramolecular capsules has attracted great interest over the last years.[1] Much like enzymes, they feature a nanometer-sized cavity where suitable reactions can be accelerated due to noncovalent interactions between the substrate and host. Exciting results concerning complex terpene cyclizations, which were inaccessible with man-made catalysts before, have been obtained in our laboratory.[2] Although these examples clearly demonstrated the potential of this approach, the field is still in its infancy and real world applications are lacking. This is mainly due to the selectivity issue. Due to the high symmetry of the nanometer-sized reaction compartment on the inside of the molecular capsule (depicted in blue in Fig. 1a), the conformation of the flexible acyclic terpene substrates cannot be controlled well. The floppy binding translates into a poor selectivity during terpene cyclizations (Fig. 1b).[2a] To achieve selective cyclizations for biologically relevant sesqui- and diterpenes, less symmetric cavities, able to constrain the conformationally freedom of the encapsulated the substrate, are required. Unfortunately, **all known self-assembled molecular capsule are of high symmetry**. Therefore, modification of the building block (BB) to reduce the symmetry of the cavity is required. Most molecular capsules known are homomeric assemblies; meaning they assemble from one single BB. Modification of this BB, produces smaller but still highly symmetric cavities (Fig. 1c, top). Less symmetric cavities are only formed if the modified BB assembles with other BBs (heteromeric assemblies, Fig 1c bottom). Heteromeric assemblies based on hydrogen bonding are extremely rare.[3] The few examples known are still of high symmetry and far too small for terpene cyclizations. More importantly, **there is no strategy available to predict the self-assembly of stable heteromeric capsules**.

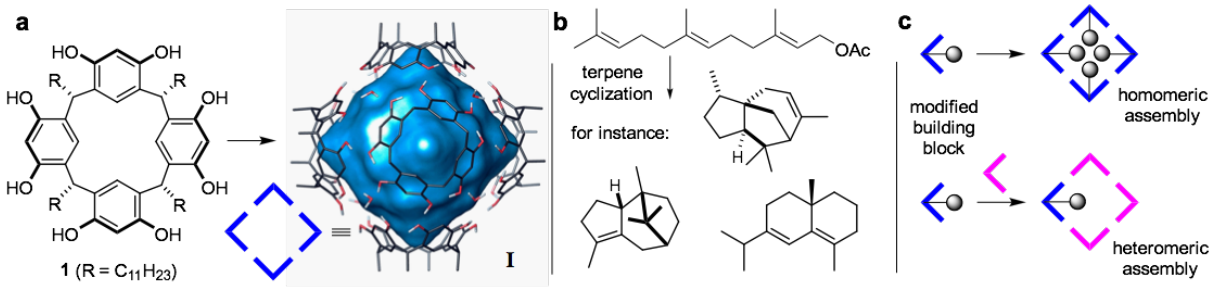

Figure 1. (a) The catalytically active supramolecular capsule I based on resorcinarene 1 as building block. The nanometer-sized cavity space, highlighted in blue, is large and of high symmetry. (b) Flexible acyclic terpenes like farnesyl acetate do not cyclize selectively inside this container due to their conformational freedom. (c) Modification of the building block of capsule I, leads to smaller but still highly symmetric cavities. Only in the case of heteromeric assemblies less symmetric cavities can be formed.

2. Research Proposal:

The overall aim of this project is the discovery of novel heteromeric self-assembled molecular capsules suitable for selective sesqui- and diterpene cyclizations. Since there is no rational approach available, we aim for a broad screening approach aided by machine learning algorithms. To achieve this ambitious goal, this study takes advantage of established expertise and experimental capabilities in both collaborating labs.

We identified a library of more than fifty concave and rigid building blocks suitably equipped with hydrogen bond motives to form self-assembled structures. This collection contains literature known compounds like **1**, but also unpublished molecules from the Tiefenbacher lab. Since no general guidelines for the search of heteromeric assemblies are available, we will start by manually screening binary combinations of this library (1'225 unique binary combinations available from 50 BB). Subsequently, ternary (19'600 unique combinations) and quaternary mixtures (230'300 unique combinations) will be screened. It is obvious that such numbers of combinations cannot be processed completely, at least manually, anymore. We will leverage modern quantum machine learning (QML) techniques[4] to navigate this combinatorial chemical space. Due to the convergence of learning curves in QML, we expect that, after providing sufficiently many experimental training examples, reliable models can be generated which enable directed screening and as a result dramatically reduce the number of additionally required experiments.

[1] a) Q. Zhang, L. Catti, K. Tiefenbacher, *Acc. Chem. Res.* **2018**, *51*, 2107-2114; b) M. Otte, *ACS Catal.* **2016**, *6*, 6491-6510.

[2] a) Q. Zhang, J. Rinkel, B. Goldfuss, J. S. Dickschat, K. Tiefenbacher, *Nat. Cat.* **2018**, *1*, 609-615; b) Q. Zhang, L. Catti, J. Pleiss, K. Tiefenbacher, *J. Am. Chem. Soc.* **2017**, *139*, 11482–11492; c) Q. Zhang, K. Tiefenbacher, *Nat. Chem.* **2015**, *7*, 197-202.

[3] D. Ajami, J.-L. Hou, T. J. Dale, E. Barrett, J. Rebek, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 10430-10434.

[4] O. A. von Lilienfeld, *Angew. Chem. Int. Ed.* **2018**, *57*, 4164.