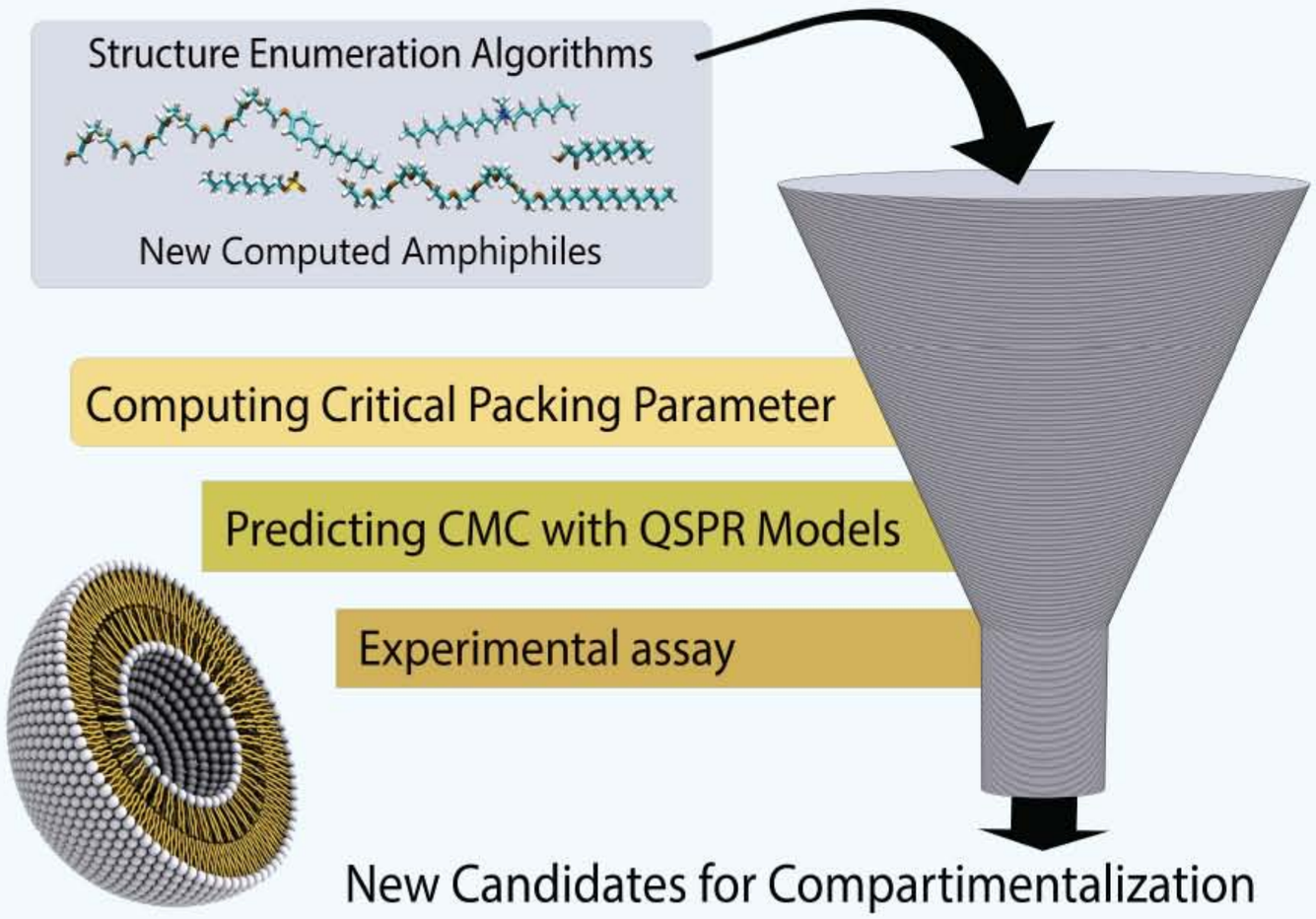
Computational Exploration of Lipid Chemical Space : Predicting Assembly using QSPR Models

Sélène Forget[1], H. James Cleaves[1,2], Tony Z. Jia[1], Richard J. Gillams[3], Markus Meringer[4]



Introduction

Compartmentalization is likely to have been essential for the emergence of life as it allows for the creation of unique chemical conditions that can be maintained out of equilibrium with the environment and the exclusion of parasites. Confining organic molecules also helps limit diffusion, increases concentration and can thus influence both the thermodynamics and kinetics of prebiotic reactions. Biology currently predominantly uses phospholipids to construct cell membranes. However, there are many other types of organic compounds that can form stable compartments in water, and many of these may have been abundant in the prebiotic environment.

Objectives

1. Explore the lipid chemical space by using structure enumeration algorithms and compute an exhaustive combinatorial library of surfactant molecules.
2. Predict the propensity of these compounds to self-assemble into membranes.
3. Identify novel molecule types which can be experimentally assayed as candidates for the emergence of protocells.

1. Earth-Life Sciences Institute, Tokyo Institute of Technology **2.** Institute for Advanced Studies, Princeton **3.**School of Electronics and Computer Science, University of Southampton, Southampton, UK **4.**Deutsches Zentrum für Luft- und Raumfahrt (DLR), Oberpfaffenhofen

Computational Exploration of Lipid Chemical Space : Predicting Assembly using QSPR Models

Structure Enumeration Algorithms

MOLGEN software [1] was used to generate an exhaustive library of 12662 plausible alkane and alkene chains with 2 to 12 carbons [2]. A diverse set of hydrophilic “heads” were selected in order to computationally generate the corresponding amphiphiles.

Heads
(hydrophilic)

SO_3-

COO-

$N+(CH_3)_3$

$N(CH_3)_3CH_2CH_2OPO_3$-
Phosphocholine

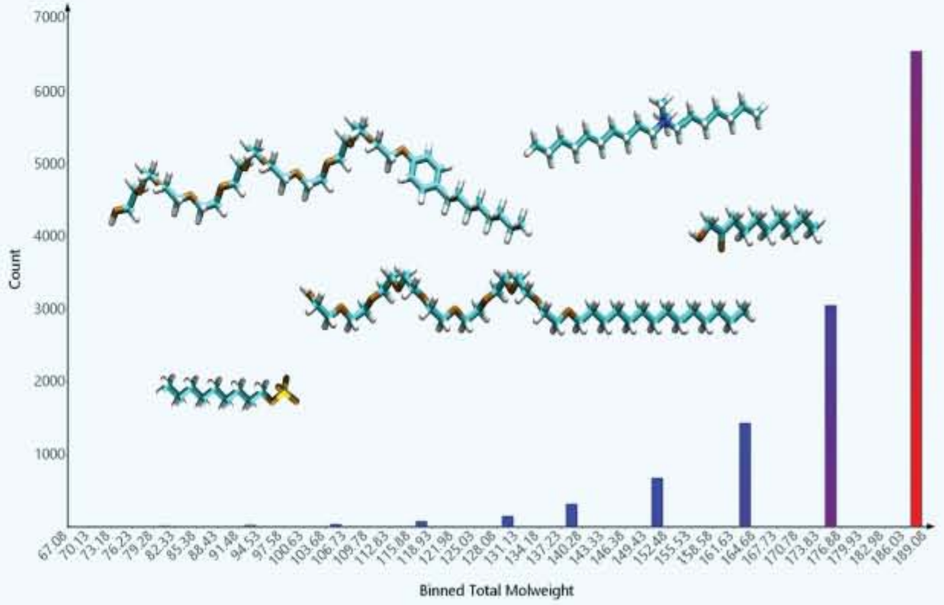
\+

Carbon Tails
(hydrophobic)

→

New Computed Amphiphiles

for fatty acids generated with up to 12 carbon.

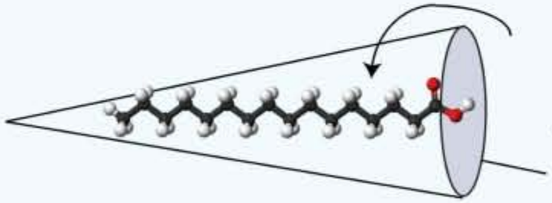


[1] Gugisch R et al., 2016, 113-138 MOLGEN 5.0, a molecular structure generator. In Advances in mathematical chemistry and applications.
[2] Cleaves H.J., Meringer M., Phil. Trans. R. Soc. A, 2017, 375, 7 Exploring astrobiology using in silico molecular structure generation.

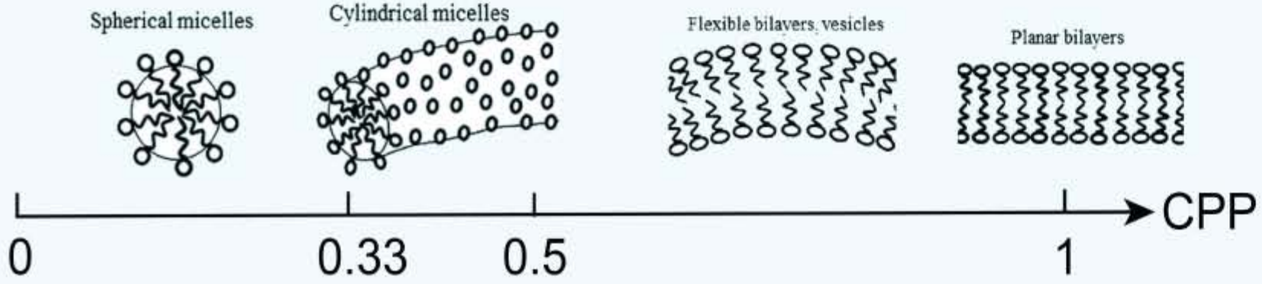
Computing Critical Packing Parameter

CPP geometrically quantifies the capacity of the molecules to aggregate and form a certain type of structure [3] :

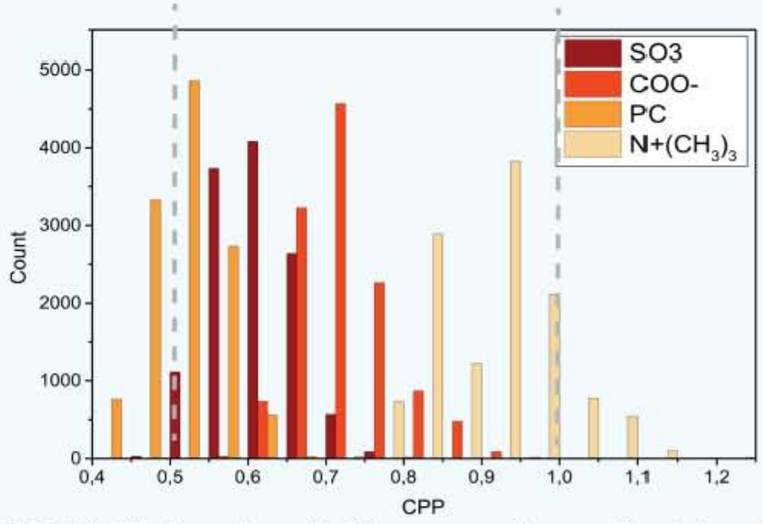
$$CPP = \frac{V}{a_e \times l}$$



V tail's volume
l length of the tail
a_e surface area of the head



We computed CPP for the computed amphiphiles. Only the compounds having a CPP between 0.5 and 1 are kept.



CPP Distribution of amphiphiles, grouped according to head type

[3] R. Nagarajan, Langmuir 2002, 18, 31-38, Molecular Packing Parameter and Surfactant Self-Assembly: The Neglected Role of the Surfactant Tail

Predicting CMC with Quantitative Structure-Property Relationships:

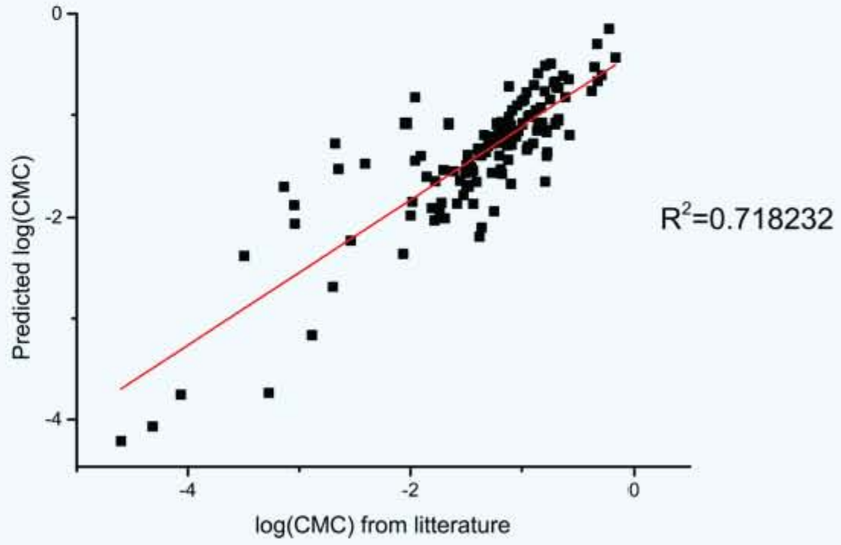
The Critical Micelle Concentration (CMC) is the minimum concentration to have the aggregation of a compound in solution. It is dependent of the temperature, the solvent, as well as the presence of other molecules such as ions. The CMC is an emperical value which can be predicted using QSPRs :

A QSPR is a mathematical model which relates molecular descriptors to a known variable (physico-chemical property) using supervised machine learning:

$$CMC = \sum a_i x_i + u$$

a_i coefficients
x_i molecular descriptors

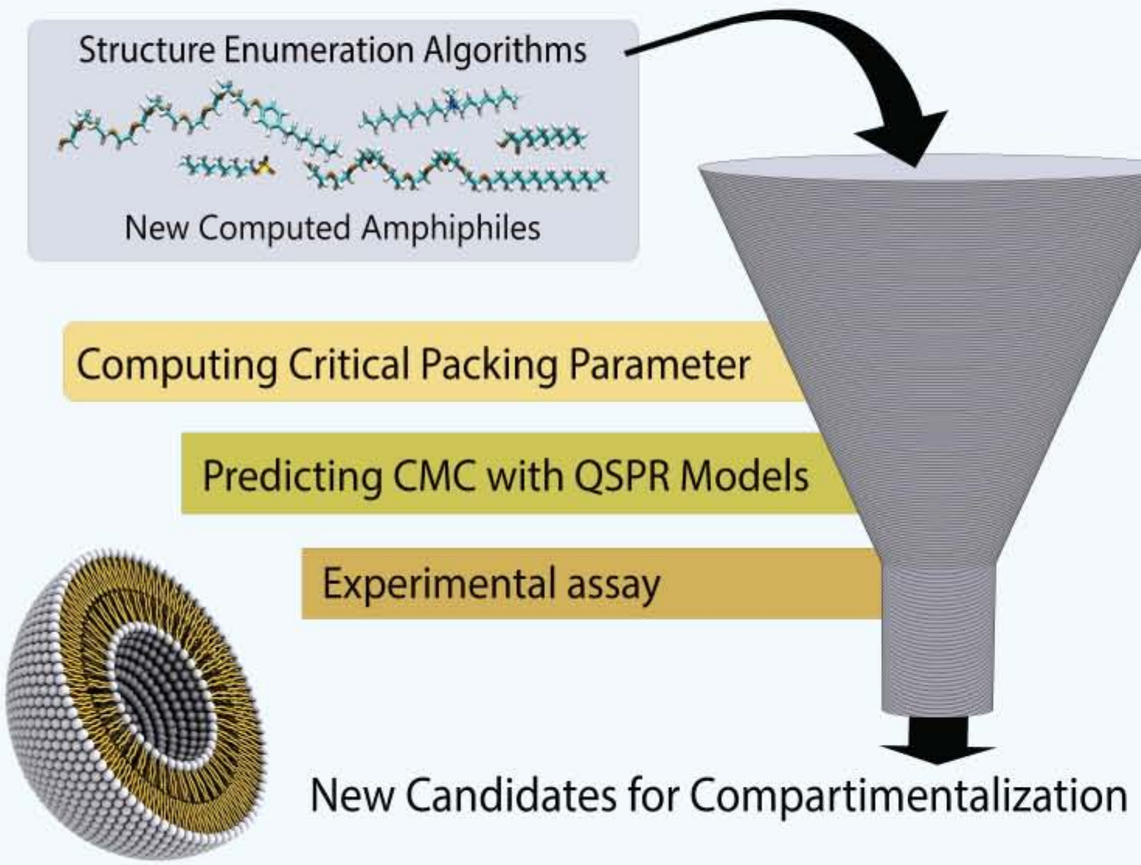
Using experimentally measured CMC values [4] (25 °C in water) for 786 compounds, we are able to predict the CMC of our computationally generated amphiphiles.



Predicted log (CMC) values of the training set using the QSPR model over the literature CMC data for SO3- amphiphiles

[4] Katritzky, Al. et al. Karelson, Ind. Eng. Chem. Res. 2008, 47, 9687–9695QSPR study of critical micelle concentrations of nonionic surfactants

Computational Exploration of Lipid Chemical Space : Predicting Assembly using QSPR Models



Future work :

- Synthesis and experimental assay to form diverse promising candidates

- Measurement of their ability of novel compartment systems to sequester RNA and other macromolecules

EARTH - LIFE SCIENCE INSTITUTE

東京工業大学
Tokyo Institute of Technology