

# **INTERNSHIP PROPOSAL**

*(One page maximum)*

Laboratory name: Gulliver lab, ESPCI Paris

CNRS identification code: UMR 7083

Internship director's surname: Z. Zeravcic, R. van Mastrigt. J. Brujic

e-mail: zorana.zeravcic@espci.fr

Phone number:

Web page:

Internship location: ESPCI Paris

Thesis possibility after internship: YES

Funding: NO

If YES, which type of funding:

## **Machine learning for designability**

Self-assembly turns building blocks into a structure without outside guidance. A versatile programmable building block is the DNA-coated colloid: a micrometer-sized particle covered with short single strands of DNA. Particles stick when their strands are complementary, so the DNA sequences set which particles bind and how strongly. Particles carrying the same DNA form one type, or “color”, and the interactions between colors are programmed by choosing the sequences. This raises a design question: which interactions make a given target cluster, and not another, assemble with high yield?

Some targets are far easier to design than others. The designability of a cluster geometry is the number of interaction rules that assemble it, and in small colloidal clusters it varies by two to three orders of magnitude. A few “magic” geometries can be reached with very many designs, while others can hardly be reached at all. Designability also depends on how assembly proceeds, for example from a free gas or from a chain, and on the number of colors. Biology shows the same phenomenon: in protein folding, a few structures are encoded by far more sequences than others, which may explain why proteins use a limited repertoire of folds. These results come from exhaustively enumerating geometries and interaction rules, which quickly becomes infeasible: twelve particles already form thousands of distinct cluster geometries.

In this project you will use machine learning on the existing designability dataset to find what makes a geometry designable. You will identify structural features that predict designability, test whether they hold for systems too large to enumerate, and compare the rules across assembly methods. These results will help guide particle and sequence design in experiments.

[1] N. Blot, PhD thesis 2025

Please, indicate which speciality(ies) seem(s) to be more adapted to the subject:

Condensed Matter Physics: NO

Soft Matter and Biological Physics: YES

Quantum Physics: NO

Theoretical Physics: YES