# Aggregate, automate, assemble

Kroon, Peter

*DOI:*
10.33612/diss.132963667

# 5 - Summary

Molecular dynamics simulations have been around for about 60 years, yet still remain one of the most powerful tools available in a computation chemist's toolbox for investigating a wide range of systems in molecular detail, even though they are limited to classical mechanics. Advances in force fields and computer hardware have enabled the use of molecular dynamics in more complex systems, progressing from hard spheres — through soft spheres, different spheres, connected spheres, actual molecules, large molecules, and groups of molecules — to large groups of large molecules. However, there is an ever growing need to study systems of ever greater complexity. Rather than the hardware or molecular dynamics algorithms, the current generation of tools for *preparing* simulations cannot deal with this complexity and are becoming a limiting factor. Every molecule in a molecular dynamics simulation requires appropriate parameters for it to show physical behaviour. This can be further specified into two cases: 1) the molecules are built up of separate building blocks, *e.g.* polymers, for which parameters are available; and 2) the molecules have not been simulated before: since there are infinitely many different molecules, often new molecules need to be parametrized, which is a relatively tedious process. For atomistic force fields new molecules can be readily parametrized using automatic topology builders. Unfortunately, for coarse-grained force fields such as the popular Martini model these are not commonplace.

In **Chapter 2** we present the VERsatile MOdular Universal Transformation Helper (VerMoUTH), a general library capable of setting up more complex systems for simulation. On top of VerMoUTH we have built martinize2, as successor of the program martinize. These tools aim to address the first case, and are designed to generated molecular dynamics topologies for arbitrary polymers. Together these tools allow researchers to generate input topology files for molecular dynamics simulations for almost any system, using almost any force field, at any resolution. These tools enable facile and rapid exploration of simulation parameters, and enable studies of complex systems that until now were out of reach, such as branched or cyclic polymers and proteins with post-translational modifications. In addition, the tools are designed such that the separate elements of the process are as independent as possible making it relatively easy to extend them to systems and setups currently not supported without requiring intimate knowledge of the complete code base.

In **Chapter 3** we approach the second case — the problem of generating mappings for coarse-grained force fields — as a graph clustering problem, and rephrase and formalize the process accordingly. We employ this formalization to take the initial steps in creating an automatic topology builder for the Martini

coarse-grained force field, using machine learning to derive appropriate input parameters for the clustering algorithm. We find that clustering atoms into soft clusters using robust Perron cluster analysis is not a suitable approach to solve this problem. However, using the spectral graph partitioning algorithm, a hard clustering algorithm, shows the potential of our work. A lot of work remains to be done to arrive at a complete automatic topology builder for the Martini force field, the first of which is to identify and implement an appropriate soft clustering algorithm.

In **Chapter 4** we demonstrate that molecular dynamics, in combination with advanced experimental techniques, has the potential to elucidate complex (supra)molecular mechanisms. Recently an artificial peptide-based system was discovered that reliably shows self-replication, and can thus serve as a model system to study self-replication in the context of the origin of (artificial) life. The system consists of GLKFK pentapeptides, which are N-terminally capped by a benzene dithiol group. The thiols in this group can be oxidized to form disulphide bridges. This causes the reversible formation of ring structures of varying sizes. Of these various rings the hexamer can nucleate and form fibres. These fibres in turn catalyse the formation of more hexamers. The hexamer concentration shows a strong exponential growth to a plateau after an initial lag phase, and it is proven to be caused by self-replication *via* seeding experiments. To provide a more in-depth understanding of the growth process, we combined several experimental methods with molecular modelling. High speed atomic force microscopy movies reveal that for fibres to grow, a reservoir of material, presumably monomers, trimers and tetramers, has to be present near the end of the fibre. We show that coarse-grained molecular dynamics can provide molecular insight in this complex supramolecular process: the fibre morphology matches experimental and united atom molecular dynamics results, trimers do not unbind from hexamer fibres in 500 ns, and trimers readily diffuse along the fibre until they reach the end of the fibre, where they are stable. The results from the simulations are consistent with the experimental evidence, and point to a mechanism that hinges on reservoir assisted self-replication, where single molecules efficiently diffuse from the reservoir to the end of the fibre.