Aggregate, automate, assemble

Kroon, Peter

DOI:
10.33612/diss.132963667

1 - Introduction

Une intelligence qui, pour un instant donné, connaitrait toutes les forces dont la nature est animée et Ia situation respective des êtres qui la composent, [...], *embrasserait dans la même formule Ies mouvements des plus grands corps de l'univers et ceux du plus léger atome: rien ne serait incertain pour elle, et l'avenir, comme Ie passé, serait présent à ses yeux.*

— *Pierre Simon de Laplace* (*1749-1827*)

An intelligence which could, at any moment, comprehend all the forces by which nature is animated and the respective positions of the beings of which it is composed, [...], *it would encompass in that formula both the movements of the largest bodies in the universe and those of the lightest atom: to it nothing would be uncertain, and the future, as well as the past, would be present to its eyes.*

— *Pierre Simon de Laplace* (*1749-1827*)

1

The earliest computers were already used for chemical simulations [1–6]. Advances in hard- and software since those early years have allowed researchers to study larger and more complex systems, as well as slower processes, in greater detail than before [7–13]. This, effectively, has turned computers into computational microscopes [14, 15], which give the ability to study diverse systems at a level of spatio-temporal resolution unmatched by any experimental technique. Simulation techniques can help provide a molecular interpretation for experimental data [16–19]. An example of this is described in Chapter 4.

There are several scales of simulations, spanning from the very small (quantum chemical simulations) to the very large (planetary bodies, hydrodynamics simulations) [9]. Molecular dynamics (MD), the most popular technique to simulate the motions of molecules, can be applied to study a few molecules in high detail or large supramolecular assemblies in less detail [9]. At the small end of the spectrum electronic polarizabilities and possibly chemical reactions are included in the model, at the large end multiple atoms, or even complete molecules are described as single, coarse-grained (CG), interaction sites. In most cases however, the solvent is explicitly represented in the simulation [9]. This thesis spans MD simulations ranging from all atom to coarse-grained.

Molecular dynamics simulations work by numerically integrating Newton's second law over time, resulting in a trajectory describing the motion of all particles in the system [20]. For every time step forces are evaluated based on the current coordinates, and the resulting accelerations are integrated to arrive at a displacement. Since this involves a numerical integration the time step cannot be too large, or the result will be inaccurate. The advantage of limiting MD simulations to classical mechanics is that this makes the simulations relatively cheap, when compared to more accurate and detailed methods. However, due to the inherent inaccuracy of the method the forces that are found are of critical importance [21, 22]. The force field determines how, from coordinates, effective forces should be derived for all particles in the simulation. MD simulations are still expensive, and it often requires supercomputers to run appreciable systems [23].

Within a force field the parameters are often separated in those involved in 1) bonded forces: the forces due to *e.g.* bonds and angles, and 2) non-bonded forces: the forces due to *e.g.* coulombic and dispersion interactions. This means, that for every molecule in the simulation a topology has to be defined, which describes how the bonded and non-bonded forces should be derived. For small molecules this can still be done by hand (although automated tools also exist, Chapter 3), but for larger molecules, such as polymers, specialized tools are required. Due to the historical focus of MD on biochemical systems [4, 5, 24–28] the existing tools are limited to the assumptions that are true

for proteins and DNA. However, chemical space is much larger than linear biopolymers, and this thesis describes one of the first attempts to solve this in a general way (Chapter 2).

Coarse-graining is a technique for reducing the number of particles in an MD simulation: multiple atoms are combined to form a single interaction site (bead). The advantages are that there are fewer particles in the simulation, easing the computational load [10, 12, 13, 21]. In addition, the potential energy landscape is generally also smoother [29], allowing for a larger time step during the numerical integration. Lastly, in a coarse-grained molecule there are fewer interactions, and hence parameters, which makes generating a topology easier. The downside, however, is the loss of resolution. This is observable in the generated trajectory, which no longer describes the position for every single atom, but rather for every CG bead. In addition, there is also a loss of resolution in the effective interactions: for example, (most) CG force fields have no concept of directional hydrogen bonds.

Fundamentally there is a trade-off between accuracy and speed [9]. Combined with constraints on computational power, some systems are simply too large, and some processes too slow to simulate with great accuracy. From a more practical point of view there is the question of whether a topology for your molecule is already available, and how hard it would be to create one if it is not. This is related to the concepts of transferability, which somehow captures how hard it is to add new molecules, and generality, which captures how many classes of molecules can be simulated with a force field. In general, force fields which are specific to one class of molecules perform better (for that class), but are more difficult to extend to different classes. Because of this, different applications require different (CG) force fields [12].

One of the most popular, generally applicable, CG force field is the Martini force field [21]. It was originally developed for lipid membranes [29, 30], and later extended to proteins [31–33], DNA and RNA [34, 35], sugars [36], and several synthetic polymers [37, 38]. By design it maps roughly four non-hydrogen atoms to one bead, preferring to keep functional groups together within one bead. The hypothesis is that an *e.g.* ester group in one molecule behaves the same as an ester group in another molecule. This assumption makes the force field highly transferable and easy to use, while still being computationally much cheaper than atomistic simulations. However, since not every atom is explicitly represented in the simulation there are also no explicit hydrogen bonds. Mostly because of this, the secondary structure of *e.g.* proteins has to be restrained, for example using an elastic network [33, 39]. This means that Martini MD simulations cannot be used to study, for example, secondary structure changes in proteins.

1

This thesis will describe the development of new software and methods that enable the setup of complex (polymeric) molecules in Chapter 2, going beyond the simple linear cases that are currently possible. Chapter 3 will deal with the development of new software and methods that enable the automatic generation of topologies for Martini molecules. Chapter 4 will deal with the application of the Martini force field in studying the self-replication of supramolecular polymers for which detailed experimental data is available. Finally, a glossary is provided that will describe some of the used terms, with which not all readers may be familiar.

References

[1] Alder, B. J.; Wainwright, T. E. (1957). Phase Transition for a Hard Sphere System, *The Journal of Chemical Physics*, Vol. 27, No. 5, 1208–1209. doi:10.1063/1.1743957

[2] Rahman, A. (1964). Correlations in the Motion of Atoms in Liquid Argon, *Physical Review*, Vol. 136, No. 2A, A405–A411. doi:10.1103/PhysRev.136.A405

[3] Gibson, J. B.; Goland, A. N.; Milgram, M.; Vineyard, G. H. (1960). Dynamics of Radiation Damage, *Physical Review*, Vol. 120, No. 4, 1229–1253. doi:10.1103/PhysRev.120.1229

[4] Levitt, M.; Warshel, A. (1975). Computer simulation of protein folding, *Nature*, Vol. 253, No. 5494, 694–698. doi:10.1038/253694a0

[5] Warshel, A. (1976). Bicycle-pedal model for the first step in the vision process, *Nature*, Vol. 260, No. 5553, 679–683. doi:10.1038/260679a0

[6] Rahman, A.; Stillinger, F. H. (1971). Molecular Dynamics Study of Liquid Water, *The Journal of Chemical Physics*, Vol. 55, No. 7, 3336–3359. doi:10.1063/1.1676585

[7] Nerenberg, P. S.; Head-Gordon, T. (2018). New developments in force fields for biomolecular simulations, *Current Opinion in Structural Biology*, Vol. 49, 129–138. doi:10.1016/j.sbi.2018.02.002

[8] Marrink, S. J.; Corradi, V.; Souza, P. C. T.; Ingólfsson, H. I.; Tieleman, D. P.; Sansom, M. S. P. (2019). Computational Modeling of Realistic Cell Membranes, *Chemical Reviews*. doi:10.1021/acs.chemrev.8b00460

[9] Ingolfsson, H. I.; Arnarez, C.; Periole, X.; Marrink, S. J. (2016). Computational 'microscopy' of cellular membranes, *Journal of Cell Science*, Vol. 129, No. 2, 257–268. doi:10.1242/jcs.176040

[10] Maximiano, P.; Durães, L.; Simões, P. (2019). Overview of Multiscale Molecular Modeling and Simulation of Silica Aerogels, *Industrial & Engineering Chemistry Research*, Vol. 58, No. 41, 18905–18929. doi:10.1021/acs.iecr.9b03781

[11] Kumar, A.; Sharma, K.; Dixit, A. R. (2019). A review on the mechanical and thermal properties of graphene and graphene-based polymer nanocomposites: understanding of modelling and MD simulation, *Molecular Simulation*, 1–19. doi:10.1080/08927022.2019.1680844

[12] Singh, N.; Li, W. (2019). Recent Advances in Coarse-Grained Models for Biomolecules and Their Applications, *International Journal of Molecular Sciences*, Vol. 20, No. 15, 3774. doi:10.3390/ijms20153774

[13] Ingólfsson, H. I.; Lopez, C. A.; Uusitalo, J. J.; de Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S. J. (2014). The power of coarse-graining in biomolecular simulations, *Wiley Interdisciplinary Reviews: Computational Molecular Science*, Vol. 4, No. 3, 225–248. doi:10.1002/wcms.1169

[14] Lee, E. H.; Hsin, J.; Sotomayor, M.; Comellas, G.; Schulten, K. (2009). Discovery Through the Computational Microscope, *Structure*, Vol. 17, No. 10, 1295–1306. doi:10.1016/j.str.2009.09.001

[15] Dror, R. O.; Dirks, R. M.; Grossman, J. P.; Xu, H.; Shaw, D. E. (2012). Biomolecular Simulation: A Computational Microscope for Molecular Biology, *Annual Review of Biophysics*, Vol. 41, No. 1, 429–452. doi:10.1146/annurev-biophys-042910-155245

[16] Frederix, P. W. J. M.; Idé, J.; Altay, Y.; Schaeffer, G.; Surin, M.; Beljonne, D.; Bondarenko, A. S.; Jansen, T. L. C.; Otto, S.; Marrink, S. J. (2017). Structural and Spectroscopic Properties of Assemblies of Self-Replicating Peptide Macrocycles, *ACS Nano*, Vol. 11, No. 8, 7858–7868. doi:10.1021/acsnano.7b02211

[17] Frederix, P. W. J. M.; Patmanidis, I.; Marrink, S. J. (2018). Molecular simulations of self-assembling bio-inspired supramolecular systems and their connection to experiments, *Chemical Society Reviews*, Vol. 47, No. 10, 3470–3489. doi:10.1039/c8cs00040a

[18] Wong-ekkabut, J.; Karttunen, M. (2016). The good, the bad and the user in soft matter simulations, *Biochimica et Biophysica Acta* (*BBA*) - *Biomembranes*, No. JANUARY. doi:10.1016/j.bbamem.2016.02.004

[19] Piskorz, T. K.; Gobbo, C.; Marrink, S. J.; De Feyter, S.; de Vries, A. H.; van Esch, J. H. (2019). Nucleation Mechanisms of Self-Assembled Physisorbed Monolayers on Graphite, *The Journal of Physical Chemistry C*, Vol. 123, No. 28, 17510–17520, research-article. doi:10.1021/acs.jpcc.9b01234

[20] Frenkel, D.; Smit, B. (2002). *Understanding Molecular Simulation: From Algorithms to Applications* (2nd ed.), Academic Press, London

[21] Marrink, S. J.; Tieleman, D. P. (2013). Perspective on the Martini model., *Chemical Society Reviews*, Vol. 42, No. 16, 6801–22. doi:10.1039/c3cs60093a

[22] Monticelli, L.; Tieleman, D. P. (2013). Force Fields for Classical Molecular Dynamics, 197–213. doi:10.1007/978-1-62703-017-5_8

[23] Shaw, D. E.; Chao, J. C.; Eastwood, M. P.; Gagliardo, J.; Grossman, J. P.; Ho, C. R.; Lerardi, D. J.; Kolossváry, I.; Klepeis, J. L.; Layman, T.; McLeavey, C.; Deneroff, M. M.; Moraes, M. A.; Mueller, R.; Priest, E. C.; Shan, Y.; Spengler, J.; Theobald, M.; Towles, B.; Wang, S. C.; Dror, R. O.; Kuskin, J. S.; Larson, R. H.; Salmon, J. K.; Young, C.; Batson, B.; Bowers, K. J. (2008). Anton, a special-purpose machine for molecular dynamics simulation, *Communications of the ACM*, Vol. 51, No. 7, 91. doi:10.1145/1364782.1364802

[24] Chothia, C.; Lesk, A. M.; Tramontano, A.; Levitt, M.; Smith-Gill, S. J.; Air, G.; Sheriff, S.; Padlan, E. A.; Davies, D.; Tulip, W. R.; Colman, P. M.; Spinelli, S.; Alzari, P. M.; Poljak, R. J. (1989). Conformations of immunoglobulin hypervariable regions, *Nature*, Vol. 342, No. 6252, 877–883. doi:10.1038/342877a0

[25] Levitt, M.; Chothia, C. (1976). Structural patterns in globular proteins, *Nature*, Vol. 261, No. 5561, 552–558. doi:10.1038/261552a0

[26] Warshel, A.; Levitt, M. (1976). Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme, *Journal of Molecular Biology*, Vol. 103, No. 2, 227–249. doi:10.1016/0022-2836(76)90311-9

[27] Levitt, M. (1976). A simplified representation of protein conformations for rapid simulation of protein folding, *Journal of Molecular Biology*, Vol. 104, No. 1, 59–107. doi:10.1016/0022-2836(76)90004-8

[28] Hess, B.; Scheek, R. M. (2003). Orientation restraints in molecular dynamics simulations using time and ensemble averaging, *Journal of Magnetic Resonance*, Vol. 164, No. 1, 19–27. doi:10.1016/S1090-7807(03)00178-2

[29] Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. (2007). The MARTINI force field: coarse-grained model for biomolecular simulations., *The Journal of Physical Chemistry. B*, Vol. 111, No. 27, 7812–24. doi:10.1021/jp071097f

[30] Marrink, S. J.; de Vries, A. H.; Mark, A. E. (2004). Coarse-grained Model for Semiquantitative Lipid Simulations, *The Journal of Physical Chemistry B*, Vol. 108, No. 2, 750–760. doi:10.1021/jp036508g

[31] Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J. (2008). The MARTINI coarse-grained force field: Extension to proteins, *Journal of Chemical Theory and Computation*, Vol. 4, 819–834. doi:10.1021/ct700324x

[32] de Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. a; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. (2013). Improved Parameters for the Martini Coarse-Grained Protein Force Field, *Journal of Chemical Theory and Computation*, Vol. 9, No. 1, 687–697. doi:10.1021/ct300646g

[33] Periole, X.; Cavalli, M.; Marrink, S.-J.; Ceruso, M. A. (2009). Combining an Elastic Network With a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition, *Journal of Chemical Theory and Computation*, Vol. 5, No. 9, 2531–2543. doi:10.1021/ct9002114

[34] Uusitalo, J. J.; Ingólfsson, H. I.; Akhshi, P.; Tieleman, D. P.; Marrink, S. J. (2015). Martini Coarse-Grained Force Field: Extension to DNA., *Journal of Chemical Theory and Computation*, Vol. 11, No. 8, 3932–45. doi:10.1021/acs.jctc.5b00286

[35] Uusitalo, J. J.; Ingólfsson, H. I.; Marrink, S. J.; Faustino, I. (2017). Martini Coarse-Grained Force Field: Extension to RNA, *Biophysical Journal*, Vol. 113, No. 2, 246–256. doi:10.1016/j.bpj.2017.05.043

[36] López, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Hünenberger, P. H.; Marrink, S. J. (2009). Martini Coarse-Grained Force Field: Extension to Carbohydrates, *Journal of Chemical Theory and Computation*, Vol. 5, No. 12, 3195–3210. doi:10.1021/ct900313w

[37] Grunewald, F.; Rossi, G.; de Vries, A. H.; Marrink, S. J.; Monticelli, L. (2018). Transferable MARTINI Model of Poly(ethylene Oxide), *The Journal of Physical Chemistry B*, Vol. 122, No. 29, 7436–7449, research-article. doi:10.1021/acs.jpcb.8b04760

[38] Xue, M.; Cheng, L.; Faustino, I.; Guo, W.; Marrink, S. J. (2018). Molecular Mechanism of Lipid Nanodisk Formation by Styrene-Maleic Acid Copolymers, *Biophysical Journal*, Vol. 115, No. 3, 494–502. doi:10.1016/j.bpj.2018.06.018

[39] Alessandri, R.; Souza, P. C. T.; Thallmair, S.; Melo, M. N.; de Vries, A. H.; Marrink, S. J. (2019). Pitfalls of the Martini Model, *Journal of Chemical Theory and Computation*, Vol. 15, No. 10, 5448–5460. doi:10.1021/acs.jctc.9b00473

1