UiO : University of Oslo

Sigbjørn Løland Bore

Advances in the Hybrid Particle-Field Approach: Towards Biological Systems

Thesis submitted for the degree of Philosophiae Doctor

Department of Chemistry
Faculty of Mathematics and Natural Sciences

Hylleraas Centre for Quantum Molecular Sciences

2020

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo

ISSN ISSN

Cover: Hanne Baadsgaard Utigard.
Print production: Reprosentralen, University of Oslo.

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree of *Philosophiae Doctor* at the University of Oslo. The research presented here is conducted under the supervision of Professor Michele Cascella.

Acknowledgments

I acknowledge the Department of Chemistry at the University of Oslo, the Hylleraas Centre for Quantum Molecular Sciences (Grant No. 262695) and the Norwegian Supercomputing Program (NOTUR) for financial support through a grant of computer time (Grant No. NN4654K).

In the process of writing this thesis, I have received constructive feedback in form of corrections and suggestions. Therefore, I thank the following people: Gaute Linga, Manuel Carer, Thomas K. Sæbøe, Erik Tellgren and David Balcells.

I am very grateful to have been a part two centers of excellence, the Centre for Theoretical and Computational Chemistry and the Hylleraas center for Quantum Molecular Sciences. This has given me the opportunity to meet and learn from a strong community of theoretical chemists and has provided me with excellent working conditions. Special thanks goes to Jan Ingar Johnsen and Trygve Ulf Helgaker, who have been very helpful in terms of everyday logistics.

The work presented in this thesis would not have been possible without my collaborators. Special thanks goes to my Italian and Japanese collaborators Antonio de Nicola, Antonio Pizzirusso, Toshiro Kawakatsu and Giuseppe Milano. They have been exceptionally helpful in sharing their knowledge on hybrid particle-field molecular dynamics and have assisted me in day-to-day operations. At the University of Oslo I would like to especially thank Hima Bindu Kolli, Victoria Ariel Bjørnestad and Morten Ledum. It has been very useful having you working on the same topic as me. I strongly believe that our research output has benefited from us standing together.

Last, but most of all people, I must thank my supervisor, Professor Michele Cascella. We have had many tough scientific discussions, but always with our eyes on the prize.

Oslo, December 2019
Sigbjørn Løland Bore

List of Papers

The thesis is comprised by results included in the following papers.

Paper I: *Hybrid Particle-Field Model for Conformational Dynamics of Peptide Chains*
Sigbjørn Løland Bore, Giuseppe Milano and Michele Cascella
Journal of Chemical Theory and Computation **14**, 1120–1130 (2018)

Paper II: *Hybrid Particle-Field Molecular Dynamics Simulations of Charged Amphiphiles in Aqueous Environment*
Hima Bindu Kolli, Antonio De Nicola, **Sigbjørn Løland Bore**, Ken Schäfer, Gregor Diezemann, Jürgen Gauss, Toshihiro Kawakatsu, Zhongyuan Lu, You-Liang Zhu, Giuseppe Milano and Michele Cascella
Journal of Chemical Theory and Computation **14**, 4928–4937 (2018)

Paper III: *Mesoscale Electrostatics Driving Particle Dynamics in Nonhomogeneous Dielectrics*
Sigbjørn Løland Bore, Hima Bindu Kolli, Toshihiro Kawakatsu, Giuseppe Milano and Michele Cascella
Journal of Chemical Theory and Computation **15**, 2033 (2019)

Paper IV: *Aggregation of Lipid A Variants: a Hybrid Particle-Field Model*
Antonio De Nicola, Thereza A. Soares, Denys E. S. Santos, **Sigbjørn Løland Bore**, G. J. Agur Sevink, Michele Cascella and Giuseppe Milano
Accepted for publication in BBA – General Subjects (2020)

Paper V: *Beyond the Molecular Packing Model: Understanding Morphological Transitions of Charged Surfactant Micelles*
Ken Schäfer, Hima Bindu Kolli, Mikkel Killingmoe Christensen, **Sigbjørn Løland Bore**, Gregor Diezemann, Jürgen Gauss, Giuseppe Milano, Reidar Lund and Michele Cascella
(In preparation)

Paper VI: *Hybrid Particle-Field Molecular Dynamics Under Constant Pressure*
Sigbjørn Løland Bore, Hima Bindu Kolli, Antonio De Nicola, Maksym Byshkin, Toshihiro Kawakatsu, Giuseppe Milano and Michele Cascella
(In preparation)

The following papers were published in peer reviewed journals during the course of the PhD, but are not included in the thesis.

- *Coupling Spin to Velocity: Collective Motion of Hamiltonian Polar Particles*
  Sigbjørn Løland Bore, Michael Schindler, Khanh-Dang Nguyen Thu Lam, Eric Bertin and Olivier Dauchot
  Journal of Statistical Mechanics: Theory and Experiment **3**, 033305 (2016)

- High-Resolution Large Time-Step Schemes for Inviscid Fluid Flow
  Sigbjørn Løland Bore and Tore Flåtten
  Applied Mathematical Modelling **81**, 263–278 (2020)

Contents

Chapter 1

Introduction

1.1 State of the art

Can we understand complex biological systems, such as bacterias, organelles and cells, by considering their constituents, the atoms? Such a bottom-up approach is attractive as it offers an understanding at the most fundamental level.

Already in 1929, Paul M. Dirac stated [1]:

> *The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble.*

Years of research have strengthened the validity of Dirac's powerful statement. Beyond a single hydrogen atom, the innate complexity of chemical systems makes investigation by pen and paper an intractable endeavor. Since the invention of the transistors and the ensuing informatics revolution in the fifties, major effort has been put into developing reliable computational methods that can deal with complicated chemical and biological systems. This effort has resulted in the establishment of *computational chemistry*. Within this field, the problem of investigating chemical systems by means of theoretical models is solved by algorithms performed by the brute force of computers.

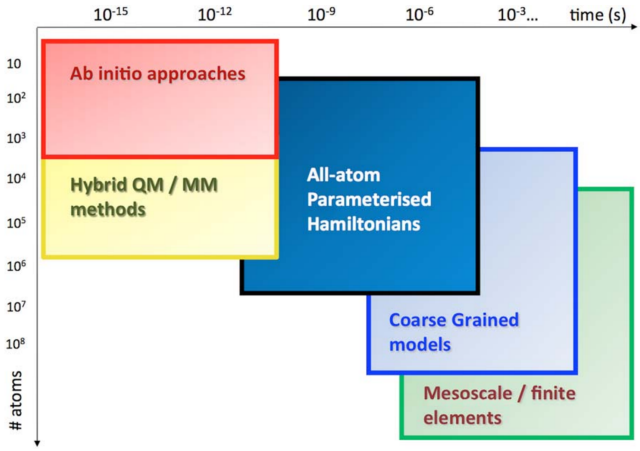

Figure 1.1: Accessible time- and size-scales for different computational methods in routine studies using state of the art implementations and architectures. Reprinted from [2].

The bottom-up approaches of computational chemistry are constituted by a set of methods for modeling of biomolecular phenomena at different resolutions (Figure 1.1). At the most fundamental level, *ab initio* approaches aim to numerically solve the quantum mechanical problem of the *Schrödinger equation* for the electrons of atoms composing the molecular system. Within the limitations of the validity of the Schrödinger equation (nonrelativistic approximation), ab initio approaches offer a route with very few assumptions. The major advances in both methodology (primarily computational cheap *Density functional theory* (DFT) and highly reliable *Coupled-cluster methods*) and software (such as Gaussian [3]), have resulted in an extensive adoption of ab initio modelling. Nevertheless, solving the electronic structure problem for systems beyond thousands of atoms acting over nanoseconds is still computationally expensive. Consequently, even a single biological molecule interacting with solvent quickly becomes intractable[1].

To treat larger molecules interacting with solvent, *all atom parameterized models* (AA) are more appropriate [5]. In AA, the *Born–Oppenheimer approximation* is used to consider the nuclei as point particles, and the interactions between atoms are approximated by simple interaction potentials (force-field). Even though AA reduces the number of degrees of freedom by about one order of magnitude, well parameterized potentials, such as the CHARMM force-field for proteins [6], are sufficiently reliable to predict the folding of proteins, and are used by pharmaceutical companies in drug discovery. Sometimes even, as is the case for liquid water models at room temperature, AA models achieve better modeling water-structure than the ab initio DFT [5].

To date, world-leading groups in molecular simulations have pushed molecular dynamics (MD) studies of proteins to reach time scales in the millisecond range and sizes as big as 10^7 atoms. These technical progresses have allowed researchers to investigate, for example, the folding of several globular proteins [7, 8], to elucidate signal transduction in G-protein coupled receptors [9], and to achieve structural refinement of low-resolution Cryogenic electron microscopy images of the HIV-1 capsid [10]. Nonetheless, these dimensionalities are on one hand not accessible to the broad computational scientific community, and on the other hand not sufficient to cover the scales pertinent to large in vivo macromolecular complexes.

Similarly for lipid bilayers, the progress in both software and hardware has allowed for the expansion of time scales reached by all-atom MD simulations of membranes from the hundreds of picoseconds, reported in the first studies of fully solvated phospholipid bilayers [11], up to microseconds in current simulations [12]. However, the lateral dimensions of MD simulations of lipid bilayers have only marginally increased, remaining confined to box sizes of approximately $10\,\mathrm{nm} \times 10\,\mathrm{nm}$. This limitation not only prevents investigation of large-scale membrane remodeling phenomena that are crucial in cellular pro-

[1] Such methods play however an important role in multiresolution methods, including the QM/MM method [4], where a part of the molecular system is described by quantum mechanics and the rest by classical physics. Multiresolution approaches allow us to simulate much larger system sizes (see Figure 1.1).

cesses, but also does not permit a direct comparison between all-atom bottom-up numerical simulations and continuum theories that historically have been successful in investigating membrane properties at larger scales. Consequently, this thesis considers coarse-grained (CG) modeling [13–15].

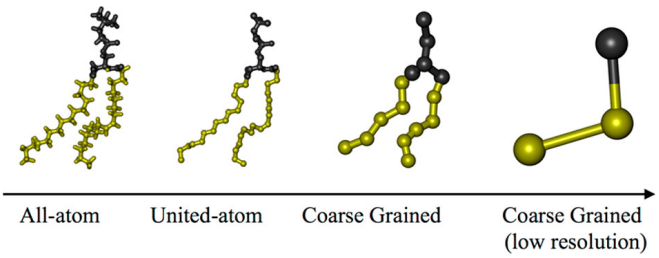

Figure 1.2: Resolutions of description of a Lipid, from AA to CG. Reprinted from [16]

In CG models, a lower resolution representation of the molecular system of interest is obtained by clustering atoms into beads (Figure 1.2), which interact through effective model potentials. While the detailed atomic resolution is lost, some information on the topological structure of the molecular assembly is retained. Computational efficiency is gained by lowering the number of degrees of freedom, which reduces the amount of interactions needed to be computed. Furthermore, the CG-procedure generally leads to a filtering out of high frequency modes present in the AA, thereby speeding up the dynamics. These models can efficiently represent molecular systems composed of several millions of atoms, for effective times that can reach the second scale; they are therefore well-adapted to investigate the structure and dynamics of large macromolecular assemblies and multi-phase systems. Consequently, CG modeling opens up the possibility of bridging the all-atom and mesoscopic scales. However, for the problems of computational efficiency of AA that CG solves, CG introduces new problems to the theoretical modeling: how should the the CG-beads be constructed, and how do we model interactions among the CG-beads?

These questions are partly answered by *statistical physics*. Generally, the statistical properties of a molecular system is determined by the partition function:

$$Z = \int \mathrm{d}\mathbf{\Gamma} \;\; \exp\left[-\beta H\left(\mathbf{\Gamma}\right)\right], \tag{1.1}$$

where $\mathbf{\Gamma}$ is the full set degrees of freedom describing a system subject to Hamiltonian H and $\beta \equiv 1/k_{\mathrm{b}}T$. A rigorous coarse-grained simulation aims at approximating the partition function:

$$Z \simeq \int \mathrm{d}\mathbf{\Gamma}_{\mathrm{CG}} \;\; \exp\left[-\beta H_{\mathrm{CG}}\left(\mathbf{\Gamma}_{\mathrm{CG}}\right)\right], \tag{1.2}$$

where $\mathbf{\Gamma}_{\mathrm{CG}}$ are now the degrees of freedom in the coarse grained representation. Theoretically, an exact relationship exists. Let $\mathbf{M}(\mathbf{\Gamma})$ be the mapping of the

fine grained positions to a CG-site. Using the δ function, we have

$$\begin{aligned} Z &= \int \mathrm{d}\mathbf{\Gamma}\mathrm{d}\mathbf{\Gamma}_{\mathrm{CG}}\ \delta(\mathbf{M}(\mathbf{\Gamma}) - \mathbf{\Gamma}_{\mathrm{CG}}) \exp\left[-\beta H\left(\mathbf{\Gamma}\right)\right] \\ &= \int \mathrm{d}\mathbf{\Gamma}_{\mathrm{CG}}\ \exp\left[-\beta H_{\mathrm{CG}}\left(\mathbf{\Gamma}\right)\right], \end{aligned} \tag{1.3}$$

where we define H_{CG} by:

$$\exp\left[-\beta H_{\mathrm{CG}}\left(\mathbf{\Gamma}_{\mathrm{CG}}\right)\right] \equiv \int \mathrm{d}\mathbf{\Gamma}\ \delta(\mathbf{M}(\mathbf{\Gamma}) - \mathbf{\Gamma}_{\mathrm{CG}}) \exp\left[-\beta H\left(\mathbf{\Gamma}\right)\right], \tag{1.4}$$

or

$$H_{\mathrm{CG}}(\mathbf{\Gamma}_{\mathrm{CG}}) = -\frac{1}{\beta} \ln\left[\int \mathrm{d}\mathbf{\Gamma}\ \delta(\mathbf{M}(\mathbf{\Gamma}) - \mathbf{\Gamma}_{\mathrm{CG}}) \exp\left[-\beta H\left(\mathbf{\Gamma}\right)\right]\right]. \tag{1.5}$$

Although (1.5) is a simple formula from a simple derivation, this formula illustrates many of the complications faced when coarse-graining. Firstly, obtaining H_{CG} requires integrating out microscopic degrees of freedom. This means that entropic contributions to the free energy are absorbed into H_{CG}. Secondly, H_{CG} is a state-dependent function as it depends on temperature, pressure and other system properties, such as concentration of salt and the pH. Finally, for CG to be useful, H_{CG} needs to be simpler and more efficient to compute than H. All of these factors force us to apply approximations and modeling in the search for H_{CG}.

Approaches to CG can be divided into two classes: top-bottom and bottom-up. In top-bottom approaches, H_{CG} is parameterized to reproduce properties of a higher level, such as experimental, thermodynamic and/or structural properties. For example, in the *MARTINI* [17], which is the most widely applied CG-force-field, the *Lennard-Jones potentials* used to model nonbonded interactions, are parameterized to reproduce water/oil-partitioning coefficients. Bottom-up approaches use information from a lower molecular scale to construct H_{CG}. In implicit solvent models (no beads for solvent), salt-salt interactions can be derived to reproduce the radial distribution function by *iterative Boltzmann Inversion* [18] or by *Inverse Monte Carlo* [19]. Also, widely used is the *force-matching* technique developed by, among others, the Voth group [14], where H_{CG} is parameterized to reproduce forces of the lower molecular scale, either from electronic structure calculations or all-atom force fields.

While the usefulness of CG is related to how well H_{CG} is able to accurately approximate (1.2), it is also important to consider its computational efficiency. An introduction of many-body interactions to model H_{CG} can quickly become less efficient than AA. For this reason, H_{CG} is often limited to two-body pair interactions, ignoring many-body terms. For example, the MARTINI force-field models nonbonded interactions with all-atom Lennard-Jones potentials, despite there being little physical justifications for it. Here, the main advantage lies in the computational efficiency which the all-atom MD softwares, such as GROMACS [20], can compute such interactions.

Similar in spirit, and going even one step further in using computational cheap potentials, is the *single chain in mean field method* (SCMF), which is actively being developed in the group of Marcus Müller [21, 22]. The SCMF method models polymers by particle-based coarse-grained polymer chains, subject to intramolecular interactions. Inspired by *self-consistent field theory* (SCFT) for polymers [23–25], the intermolecular interactions between chains are modeled by a purely density-dependent interaction energy. This gives rise to a species-dependent instantaneous inhomogenous external potential, which acts on the chains. Statistical sampling is achieved by Monte Carlo (MC) moves. Since there are no pair interactions, the polymer chains can be efficiently divided among processors, needing only to communicate for the update of the external potential. Because the external potential is a slow changing variable, a multiple time-step approach can be used, where the external potential is kept constant across many MC moves. The achieved net effect is excellent scaling with processors for small and large systems. Recently, a GPU-based implementation reached the milestone of 10 billion particles, outperforming in relaxation of copolymer melts, the state of the art Lennard-Jones based HOOMD-blue software [26, 27].

Inspired by the SCMF method, Giuseppe Milano proposed the hybrid Particle-Field molecular dynamics method (hPF-MD) [28]. Instead of using MC moves, molecular dynamics is performed by integrating equations of motion, with forces computed from spatial derivatives of the external potential (particle-field forces). This procedure retains many of the computational advantages present in the SCMF method. From the dense polymer melts [29], that are prototypical applications of SCFT and SCMF, hPF-MD has found a wide range of applications, including non-lamellar and lamellar phases of lipids [30–32], vesicles [33] and percolation phenomena for carbon nanotubes in polymer melts [34], all paving a way towards closing the gap between all-atom and mesoscopic dimensionalities for biomolecular systems [2, 16, 35].

1.2 Scope, challenges and objectives

The scope of this thesis has been to develop new hPF-MD methods and models that can be used in the study of macromolecular biological systems. In particular, the thesis aims to extend the capability of hPF-MD to represent:

- Polypeptides
- Electrostatics
- Multiphase electrolytic systems
- Constant-pressure simulations

CG modelling of polypeptides is generally considered a challenging task. Contrary to simpler organic polymers, proteins exhibit complex local *secondary* structures, for example α-helices and β-sheets, that assemble in more complex

tertiary folded structures. These motifs, that can involve only a few amino acids, as well as tens to hundreds, are crucial not only as structural scaffold, but can be directly involved in the protein function. As the folding of secondary structure elements depends on the balance of both local and nonlocal interactions, their stability is very difficult to capture when coarse-graining, as important information can easily be lost. Given these challenges, the main research objective for the work on polypeptides is to provide:

- A proof of principle hPF-MD model that can represent conformational dynamics of polypeptides.

The polyelectrolyte nature of many biological molecules, such as proteins and DNA, plays a crucial role in determining their function. Therefore, the modeling of a large number of biological molecules requires an adequate description of electrostatics. The efficiency of the hPF-MD approach requires not using pair interactions, hence specialized methodology ans software is required. With this in mind, one of the main research objectives is to develop:

- Efficient hPF-MD software for computing electrostatic interactions.

Achieving accurate modeling of electrostatics in CG models is particularly challenging, as the molecular organization on a detailed level, which gives rise to dielectrics and screening of electrostatic interactions, is lost in the coarse-grained representation. This is especially detrimental for multiphase systems, where the value of relative dielectric, can vary by almost two orders of magnitude. Therefore a particular focus of this thesis is to develop:

- Accurate modeling of electrostatics in multiphase systems.

The utility of methodology and software development depend on applications. The hPF-MD method is particularly well suited to study multiphase phenomena, such as aggregation and phase-separation. With electrostatics in hPF-MD, we can model multiphase systems involving polyelectrolytes. Therefore, a key research objective is to develop:

- New models for polyelectrolytes that can assist in understanding multiphase-phenomena.

hPF-MD has yet to incorporate constant-pressure simulations. This is problematic for many biological systems, and in particular lipid membranes. The properties of lipid membranes are typically defined at a given surface tension, and this can only be obtained with constant-pressure simulations. With these considerations in mind, it is crucial for the applicability of hPF-MD to develop:

- Methodology for constant-pressure simulations with hPF-MD.

1.3 Structure of the thesis

The thesis is composed of four chapters. Chapter 1 aims at providing a context, as well as a motivation, for the work presented in this thesis. Chapter 2 introduces hPF-MD methodology in detail. Chapter 3 presents the papers which have resulted from the work of this thesis. Finally, chapter 4 sets out the main conclusion and outlook of this thesis.

Chapter 2

Methods

In this chapter we introduce the methods used throughout the thesis. While *molecular dynamics* is a central part of this thesis, the novelty of the work lies mostly in the use of density-field formalism to model intermolecular interactions among molecules. Hence the focus is put on the hybrid particle-field (hPF) part. For a thorough introduction into the foundations of molecular dynamics, the interested reader is referred to the excellent books by Dennis C. Rapaport [36] and by Daan Frenkel and Berend Smit [37].

2.1 Hybrid particle-field molecular dynamics

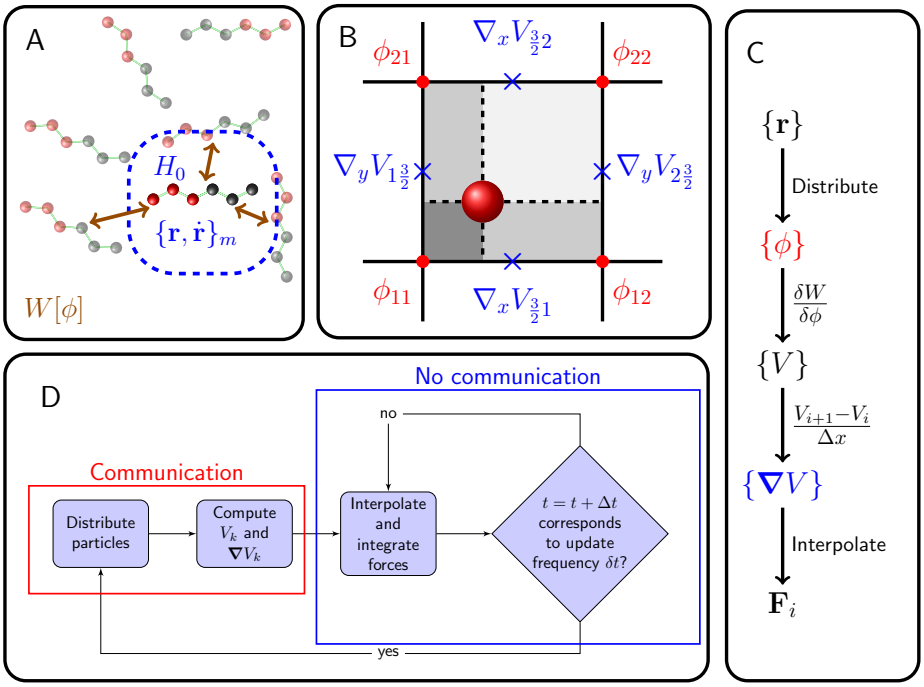

Figure 2.1: Illustration of the hybrid particle-field molecular dynamics. (A) A Polymer melt in which the polymer is subject to single-molecule Hamiltonian H_0 and is coupled to its environment through $W[\phi]$. (B) Particle-mesh routine for obtaining forces. (C) Relationship between the different variables needed for computing forces on particles. (D) Flow-chart for the molecular dynamics.

The essence of the hPF-method (see Figure 2.1) is contained in the separation

of the total energy of a molecular system into two terms:

$$H = \sum_{m=1}^{N_m} H_0(\{\mathbf{r}, \dot{\mathbf{r}}\}_m) + W([\{\phi\}]). \tag{2.1}$$

The first term is the sum over all N_m molecules of the *single-molecule Hamiltonian* $H_0(\{\mathbf{r}, \dot{\mathbf{r}}\}_m)$, which has an explicit dependence on particle positions $\{\mathbf{r}\}_m$ and velocities $\{\dot{\mathbf{r}}\}_m$. This includes the kinetic energy and the intramolecular energy. Intermolecular interactions are modeled through the second term, an *interaction energy functional* $W([\{\phi\}])$. The interaction energy functional is dependent on particle positions *only* through their set of local number densities $\{\phi\} = \{\phi_1 \dots \phi_k \dots \phi_M\}$ of the M particle species. The net effect of the interaction energy is a particle-specific external potential, which is given by (to be explained in 2.1.2):

$$V_k(\mathbf{r}) = \frac{\delta W[\{\phi\}]}{\delta \phi_k(\mathbf{r})}, \tag{2.2}$$

and a corresponding force-contribution on particle i of type k:

$$\mathbf{F}_i = -\boldsymbol{\nabla} V_k(\mathbf{r}_i). \tag{2.3}$$

Statistical properties for molecular systems with total energy expressed as (2.1), can be estimated by applying sampling methods. The single chain in mean field method (SCMF) [21] samples with *Monte Carlo* (MC), while hPF-MD method samples with *molecular dynamics* (MD).

Remarks on MD, MC and dynamics In general, MD has the advantage over MC, that it in addition to providing sampling, also can represent dynamics. However, with hPF-MD care should be taken before concluding on dynamics. One particular limitation lies within its limited ability to model sterics. Contrary to ordinary MD with pair interactions (particle-particle), where the intermolecular interactions keep particles from overlapping, sterics can only be modeled in a limited sense[1], and particles in hPF-MD can overlap. This implies that effects on dynamics, by for instance entanglement[2] or crowding, are not captured. In some sense this is problematic, because such effects are highly important for functionality. However, given the already ambitious goal of hPF-methods achieving good statistics on large macromolecular systems, dynamics should be addressed in future work. In some respects, the inaccurate modeling of steric effects is useful, as sterics can slow down the dynamics and thereby sampling. In fact, one of the keys to the success of SCMF is that molecules are more easily moved by MC (less rejection) than in particle-particle based methods [39].

[1] Due to the compressibility term, radial distribution functions obtained with hPF-MD simulations generally show a low probability at overlap-distances [31].

[2] We note that one can add on top of hPF specific models for entanglement, such as in ref. [38].

2.1.1 Molecular dynamics

In MD, sampling is done by integration of Newton's equations for the particles:

$$m_i \frac{\mathrm{d}^2 \mathbf{r}_i}{\mathrm{d}t^2} = \mathbf{F}_i, \tag{2.4}$$

where $\mathbf{F}_i$ is the total force acting on the particle i, and m_i is its mass. The integration of (2.4) is done by discretization of time into time steps. Using specialized integrators[3], such as *velocity Verlet*:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \dot{\mathbf{r}}_i(t)\Delta t + \frac{1}{2m_i}\mathbf{F}_i(t)\Delta t^2, \tag{2.5a}$$

$$\dot{\mathbf{r}}_i(t + \Delta t) = \dot{\mathbf{r}}_i(t) + \frac{1}{2m_i}\left(\mathbf{F}_i(t) + \mathbf{F}_i(t + \Delta t)\right)\Delta t, \tag{2.5b}$$

the dynamics of the particles is propagated. Assuming the *ergodic theorem*, estimates of averages of observables $\mathcal{O}$ are computed by time averaging:

$$\langle \mathcal{O} \rangle = \frac{1}{t}\int_0^t \mathrm{d}t' \ \mathcal{O}(t'), \tag{2.6}$$

when the simulation time $t \to \infty$.

2.1.2 Derivation of hPF-MD and relationship with self-consistent field theory

Equation (2.2) shows the relationship between the interaction energy and the external potential which acts on the particles. This formula, which is at the basis of hPF-MD, was first derived by following the same procedure as in self-consistent field theory (SCFT) [24, 40]. Starting from the Hamiltonian [40, p. 127]:

$$\hat{H}(\mathbf{\Gamma}) = \hat{H}_0(\mathbf{\Gamma}) + \hat{W}(\mathbf{\Gamma}), \tag{2.7}$$

where $\mathbf{\Gamma}$ specifies the microstate of the system, $\hat{H}_0$ is the energy of a noninteracting molecule, and $\hat{W}$ is the interaction energy. In the canonical ensemble, we have:

$$Z = \int \mathrm{d}\mathbf{\Gamma} \ \exp\left[-\beta\left(\hat{H}_0(\mathbf{\Gamma}) + \hat{W}(\mathbf{\Gamma})\right)\right], \tag{2.8}$$

where the integral over $\mathbf{\Gamma}$ is the the integral over the whole phase space:

$$\mathrm{d}\mathbf{\Gamma} = \prod_i^N \mathrm{d}\mathbf{r}_i \mathrm{d}\mathbf{p}_i. \tag{2.9}$$

We then assume

$$\hat{W}(\mathbf{\Gamma}) = \hat{W}(\hat{\phi}), \tag{2.10}$$

[3]Integrators for MD are designed to conserve the symplectic structure of the Hamiltonian dynamics. Most importantly, this entails conservation of the total energy.

where $\hat{\phi}$ are particle densities:

$$\hat{\phi}(\mathbf{r}, \mathbf{\Gamma}) = \sum_{i}^{N} \delta(\mathbf{r} - \mathbf{r}_i). \tag{2.11}$$

Using the δ function property:

$$\int [\mathrm{D}g(\mathbf{r})]\ \delta(f(\mathbf{r}) - g(\mathbf{r}))F[g(\mathbf{r})] = F[f(\mathbf{r})], \tag{2.12}$$

we can write (2.8) as:

$$Z = \int [\mathrm{D}\varphi(\mathbf{r})]\ \int \mathrm{d}\mathbf{\Gamma}\ \delta[\varphi(\mathbf{r}) - \hat{\phi}(\mathbf{r}, \mathbf{\Gamma})] \exp\left[-\beta\left(\hat{H}_0(\mathbf{\Gamma}) + \hat{W}(\varphi(\mathbf{r}))\right)\right]. \tag{2.13}$$

The δ function can be rewritten by Fourier transformation as follows:

$$\delta[\varphi(\mathbf{r}) - \hat{\phi}(\mathbf{r}, \mathbf{\Gamma})] = \int [\mathrm{D}w(\mathbf{r})]\ \exp\left[i \int \mathrm{d}\mathbf{r}\ w(\mathbf{r})\left(\varphi(\mathbf{r}) - \hat{\phi}(\mathbf{r}, \mathbf{\Gamma})\right)\right]. \tag{2.14}$$

Inserting (2.14) into (2.13), we get:

$$Z = \int \mathrm{d}\mathbf{\Gamma}\ \int [\mathrm{D}\varphi(\mathbf{r})]\ \int [\mathrm{D}w(\mathbf{r})]\ \exp\left[i \int \mathrm{d}\mathbf{r}\ w(\mathbf{r})\left(\varphi(\mathbf{r}) - \hat{\phi}(\mathbf{r}, \mathbf{\Gamma})\right)\right] \exp\left[-\beta\left(\hat{H}_0(\mathbf{\Gamma}) + \hat{W}(\varphi(\mathbf{r}))\right)\right] \tag{2.15}$$

We define

$$V(\mathbf{r}) \equiv i/\beta w(\mathbf{r}), \tag{2.16}$$

and the partition function of the molecules subject to $V(\mathbf{r})$ as:

$$z(V(\mathbf{r})) \equiv \int \mathrm{d}\mathbf{\Gamma}\ \exp\left[-\beta\left(\hat{H}_0(\mathbf{\Gamma}) + \int \mathrm{d}\mathbf{r}\ \hat{\phi}(\mathbf{r}, \mathbf{\Gamma})V(\mathbf{r})\right)\right]. \tag{2.17}$$

With definition (2.17) we get:

$$Z = \int [\mathrm{D}\varphi(\mathbf{r})]\ \int [\mathrm{D}w(\mathbf{r})] \exp\left[-\beta\left(-\frac{1}{\beta}\ln z + W(\varphi(\mathbf{r})) - \int \mathrm{d}\mathbf{r}\ V(\mathbf{r})\varphi(\mathbf{r})\right)\right] \tag{2.18}$$

or

$$Z = \int [\mathrm{D}\varphi(\mathbf{r})]\ \int [\mathrm{D}V(\mathbf{r})] \exp\left[-\beta\mathcal{F}([\varphi(\mathbf{r}), V(\mathbf{r})])\right], \tag{2.19a}$$

$$\mathcal{F}([\varphi(\mathbf{r}), V(\mathbf{r})]) \equiv -\frac{1}{\beta}\ln z + W(\varphi(\mathbf{r})) - \int \mathrm{d}\mathbf{r}\ V(\mathbf{r})\varphi(\mathbf{r}). \tag{2.19b}$$

The partition function can be sampled by *Field theoretic methods*, such as the ones being developed in the Fredrickson group [41]. In the self-consistent field theory, the sum over the canonical ensemble is approximated by a Gaussian integral around the most probable state that minimizes the argument of the exponential function, also referred to as the method of steepest descent. The condition determining the most probable state is given by:

$$\frac{\delta \mathcal{F}}{\delta \varphi(\mathbf{r})} = 0, \quad \frac{\delta \mathcal{F}}{\delta V(\mathbf{r})} = 0, \tag{2.20a}$$

which gives

$$V(\mathbf{r}) = \frac{\delta W[\phi]}{\delta \phi(\mathbf{r})}, \text{ and } \quad \varphi(\mathbf{r}) = -\frac{1}{\beta}\frac{\delta z}{\delta V(\mathbf{r})} = \left\langle \hat{\phi}(\mathbf{r}, \mathbf{\Gamma}) \right\rangle = \phi(\mathbf{r}). \tag{2.21}$$

Although the derivation above was done with a single component, it is easily generalized to multicomponent systems:

$$V_k(\mathbf{r}) = \frac{\delta W[\phi]}{\delta \phi_k(\mathbf{r})}, \text{ and } \quad \phi_k(\mathbf{r}) = -\frac{1}{\beta}\frac{\delta z}{\delta V_k(\mathbf{r})}, \tag{2.22}$$

where k denotes the species.

The procedure for solving SCFT is illustrated in Figure 2.2. By using densities $\phi_k(\mathbf{r})$, the external potential $V_k(\mathbf{r})$ is computed. From the external potential, the partition function for independent chains z is computed. From z, a new set of averaged out densities ϕ_k is computed, and the procedure is repeated until self-consistency is reached, corresponding to the minimal free energy of the system at field configuration $\{\phi_k^*, V_k^*\}$ [23, p. 204].

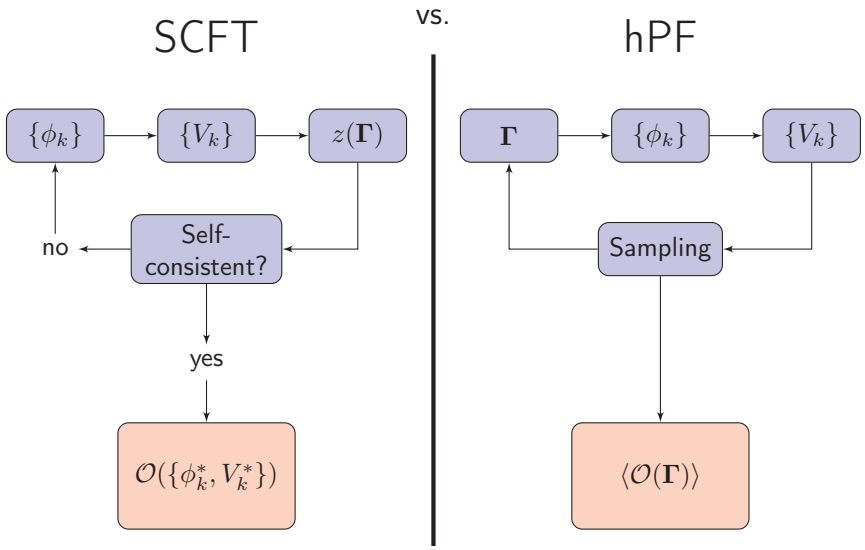

Figure 2.2: SCFT-procedure compared against hPF.

While hPF and SCMF are similar to SCFT, they are not equivalent. The use of the saddle-point approximation in SCFT amounts to considering *mean-field interactions*, because the density-fields are averaged over the single molecule partition function z. Therefore, one obtains only contributions from a single field-configuration $\{\phi_k^*, V_k^*\}$ when computing observables $\mathcal{O}(\{\phi_k^*, V_k^*\})$ [23, pp 204]. Instead, hPF-MD and SCMF consider the *instantaneous external potential* from a single molecular configuration $\mathbf{\Gamma}$, and by sampling different molecular configurations, we obtain averages of $\mathcal{O}(\mathbf{\Gamma})$. Consequently, fluctuations that are not present in SCFT, can be described by SCMF and hPF-MD [21].

An alternative manner of viewing hPF-MD is by considering the force directly from [38]:

$$\mathbf{F}_i = -\frac{\partial W}{\partial \mathbf{r}_i}, \tag{2.23}$$

which gives:

$$\mathbf{F}_i = -\int \mathrm{d}\mathbf{r}\ \frac{\delta W}{\delta \phi_k(\mathbf{r})} \frac{\partial \phi_k(\mathbf{r})}{\partial \mathbf{r}_i}, \tag{2.24}$$

or

$$\mathbf{F}_i = -\int \mathrm{d}\mathbf{r}\ V_k(\mathbf{r}) \frac{\partial \phi_k(\mathbf{r})}{\partial \mathbf{r}_i}, \tag{2.25}$$

Here the force is related to an integral over the same external potential found in our derivation by SCFT. There are multiple ways of computing this integral (such as in [38]); in the hPF-MD procedure proposed in [28], it is estimated by:

$$\mathbf{F}_i \simeq -\boldsymbol{\nabla} V_k(\mathbf{r}_i), \tag{2.26}$$

i.e. the derivative of the external potential at the position of the particle. In practice this is done by interpolation of derivatives of the external potential onto the particles (see 2.3.1).

2.2 Interaction energies

The most important interaction-energies employed in hPF-MD are in the form of:

$$W[\{\phi(\mathbf{r})\}] = \int \mathrm{d}\mathbf{r}\ w(\{\phi(\mathbf{r})\}), \tag{2.27}$$

where $w(\{\phi(\mathbf{r})\})$ is the *local interaction energy density*. Interaction energies of this particular form give rise to external potentials:

$$V_k(\mathbf{r}) = \frac{\partial w(\{\phi(\mathbf{r})\})}{\partial \phi_k(\mathbf{r})}. \tag{2.28}$$

2.2.1 Partitioning: The Flory-Huggins term

In analogy with Flory-Huggins lattice theory, and assuming each particle occupies a space of $v_0 = 1/\rho_0$, we describe interaction between densities with the

following interaction energy density [23, p. 151]:

$$w_{\tilde{\chi}}(\{\phi\}) = \frac{1}{2\rho_0} \sum_{k\ell} \tilde{\chi}_{k\ell} \phi_k(\mathbf{r}) \phi_l(\mathbf{r}), \tag{2.29}$$

where $\tilde{\chi}_{k\ell}$ is analogous to the Flory χ parameter[4]. However, unlike in ideal theory, where $\tilde{\chi}_{kl}$ is related to the potential of mean force or the partition coefficient between two species, here it is purely energetic and a phenomenological constant used to model contacts between particles. The corresponding external potential of $w_{\tilde{\chi}}$ is given by:

$$V_{\tilde{\chi},k}(\mathbf{r}) = \frac{1}{\rho_0} \sum_{\ell} \tilde{\chi}_{k\ell} \phi_l(\mathbf{r}). \tag{2.30}$$

2.2.2 Homogeneity: The excluded volume term

To control local fluctuations and avoid nonphysical accumulation of particles, a local energy density which is dependent on the sum of particle densities is often used [40, p. 164]:

$$w_{\kappa}(\phi) = \frac{1}{2\rho_0 \kappa} \left(\sum_{\ell} \phi_\ell - \rho_0 \right)^2, \tag{2.31}$$

where κ is referred to as a compressibility parameter. The external potential of (2.31) felt by a particle of species k, is given by:

$$V_{\kappa,k} = \frac{1}{\rho_0 \kappa} \left(\sum_{\ell} \phi_\ell - \rho_0 \right), \tag{2.32}$$

Note that the second term is only a constant, and therefore it does not contribute to forces. Consequently, we can relate the two terms to the $V_{\tilde{\chi}}(\mathbf{r})$ through the following relation:

$$\tilde{\chi}_{k\ell} = \frac{1}{\kappa}, \tag{2.33}$$

where the κ term is effectively an added constant to the whole $\tilde{\chi}_{kl}$ matrix.

2.3 Computational procedures for hPF-MD

Having introduced the hPF-MD formalism, we now consider the computational procedures needed to perform hPF-MD. All the procedures that are presented here have been implemented in the hPF-MD code OCCAM, which was used for obtaining the results of thesis. As is emphasized later in 2.3.4, these procedures are not unique, but they are designed to exploit the efficiencies which the hPF-MD formalism allow for.

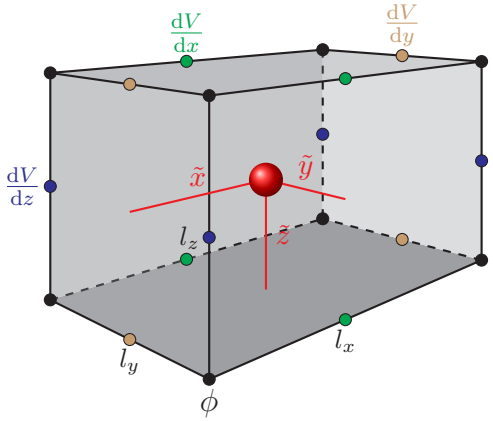

Figure 2.3: Distribution of a particle with CIC-procedure. Grid of density is indicated by black balls, while the staggered grid of the derivatives of the external potential is indicated by coloured balls.

2.3.1 Computation of densities and forces

In hPF-MD the forces due to the density-dependent interaction potential are computed through a particle-mesh approach [42]. First, the simulation box of $L_x \times L_y \times L_z$ is divided into $m_x \times m_y \times m_z$ cells (regular grid) of size $l_x = L_x/m_x, l_x = L_x/m_y, l_y = L_z/m_z$. The N particles are then distributed onto M grids for each of the M types of particles. In particle-mesh methods, the most commonly used ways of distributing particles are the nearest-grid-point (NGP) and cloud-in-cell (CIC), which differ by considering nearest grid point and grid points of the cell, respectively. In the current version of OCCAM, the CIC method, as illustrated in Figure 2.3, is used:

$$W(\tilde{x},\tilde{y},\tilde{z}) = w(\tilde{x})w(\tilde{y})w(\tilde{z}), \quad w(\tilde{x}) = 1-\tilde{x}, \quad \tilde{x} \equiv x/l_x - \text{floor}\ (x/l_x), \quad (2.34)$$

where $W(\tilde{x},\tilde{y},\tilde{z})$ is the weight prescribed to neighboring vertices. The discretized densities are then obtained by summing the contributions of all the particles onto discretized densities $\phi_k^{i_x i_y i_z}$. Using $\phi_k^{i_x i_y i_z}$, the corresponding external potential $V_k^{i_x i_y i_z}$ is computed. The force on particle i requires the gradient of the external potential at position $\mathbf{r}_i$. These gradients are computed on a staggered grid with central finite-difference approximation at the mid-points of the edges (see Figure 2.3):

$$\left.\frac{dV_k}{dx}\right|_{i_x+1/2,i_y,i_z} = \frac{1}{l_x}\left(V_k^{i_x+1,i_y,i_z} - V_k^{i_x,i_y,i_z}\right) \quad (2.35)$$

[4]Throughout this thesis we use $\tilde{\chi}$ with unit energy, which in Flory-Huggins-theory would correspond to $\tilde{\chi} = k_b T \cdot \chi$.

Finally, forces can be computed directly from the staggered grid as follows [28, 43]:

$$F_{i,x} = -w(\tilde{x}_i + 1/2)\frac{\mathrm{d}V_k}{\mathrm{d}x}_{i_x+1/2,i_y,i_z} - w(1/2 - \tilde{x}_i))\frac{\mathrm{d}V_k}{\mathrm{d}x}_{i_x-1/2,i_y,i_z}, \quad (2.36a)$$

$$F_{i,y} = -w(\tilde{y}_i + 1/2)\frac{\mathrm{d}V_k}{\mathrm{d}y}_{i_x,i_y+1/2,i_z} - w(1/2 - \tilde{y}_i)\frac{\mathrm{d}V_k}{\mathrm{d}y}_{i_x,i_y-1/2,i_z}, \quad (2.36b)$$

$$F_{i,z} = -w(\tilde{z}_i + 1/2)\frac{\mathrm{d}V_k}{\mathrm{d}z}_{i_x,i_y,i_z+1/2} - w(1/2 - \tilde{z}_i)\frac{\mathrm{d}V_k}{\mathrm{d}z}_{i_x,i_y,i_z-1/2}. \quad (2.36c)$$

2.3.2 Quasi-instantaneous external potential

One of the most important rules for numerical solution of partial differential equations is the *Courant–Friedrichs–Lewy condition* (CFL) [44]. It states that the time step used for propagating explicit time-integration schemes is limited by

$$\Delta t < C\Delta x, \quad (2.37)$$

where C is a number describing how fast the solution travels in space. Analogously for hPF-MD and SCMF, the densities are computed on a coarse grid, and their speed "C" is low. The *quasi-instantaneous* approximation exploits this fact by keeping the external potential constant over multiple time steps. Systematic benchmarks, and in particular the extensive phospholipid simulations in [29], show that the computed properties remain largely unaltered for hundreds of time steps, depending on the size of the grid. The quasi-instantaneous approximation significantly boots the efficiency of hPF-MD by reducing the amount of computation per time step. Moreover, it allows for a more efficient parallelization.

2.3.3 Parallelization

The main advantage of the hPF-MD compared to ordinary MD methods lies within the computational efficiencies that the particle-field formalism allows for. This is best understood when contrasted by MD. In ordinary MD, the most expensive routines are generally the computation of intermolecular interactions. The main computational cost lies in the computation pair interactions which are of order N^2. Years of development has resulted in methods that overcome this scaling, such as Verlet lists [37, p. 545], truncation schemes [37, p. 98] and Ewald summation-schemes [45], reducing it to $\mathcal{O}(N\log(N))$. State of the art MD-Packages, such as NAMD [46] and GROMACS [20], use *domain decomposition* as a parallelization strategy. In domain decomposition, the simulation box is divided into spatial domains, and molecules are assigned to processors according to which domain they reside. This not only reduces the memory usage but also communication among processors, as long-range interactions are computed by only considering the neighboring domains. Nonetheless, computing pair-interactions still remains the most expensive part of the simulation.

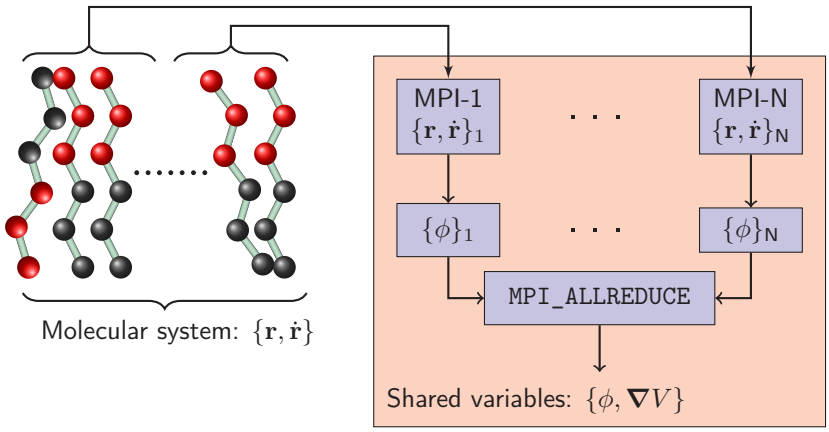

Figure 2.4: Parallelization strategy for hPF-MD. Molecules are assigned to MPI-tasks at the start of the simulation. Each MPI-task n contains only information $\{\mathbf{r}, \dot{\mathbf{r}}\}_n$ on its molecules throughout the whole simulation. The global density-field is obtained by computing density-contributions from each MPI-Task $\{\phi\}_n$, which are then combined by a single `MPI_ALLREDUCE` command. From the shared densities, the external potential and its derivative are computed on each MPI-task and are identical for all MPI-tasks.

In hPF-MD, since H_0 only involves intramolecular interactions and intermolecular interactions are computed by a particle-mesh approach, molecules naturally decouples, significantly reducing the amount of needed communication. In fact, domain decomposition is not necessary, and a simpler strategy where each MPI-task has a copy of global variables, such as the density grid, can be adopted. With this strategy, molecules are divided and assigned to MPI-tasks at the start of the simulation (Figure 2.4). Each MPI-task only contains information on its molecules (the positions and velocities of the particles). On each MPI-task, densities are computed, and by a single `MPI_allreduce` command, the densities from all the MPI-tasks are combined. This is the only major event requiring communication among all of the processors in use.

The algorithm for propagating hPF-MD and the effect of the parallelization strategy are illustrated in Figure 2.5. The algorithm is separated into two parts: one requiring communication and the second not requiring communication. This separation captures the essential features determining the overall computational efficiency. The cost of the communication part is controlled by the quasi-instantaneous update frequency $\delta t \equiv m\Delta t$ and is reduced as $1/m$, where commonly $m \sim 100$. The second part, by requiring no communication, is *trivially parallelizable*, and its contribution to computation-time is reduced as $1/N$, where N is number of MPI-tasks. This algorithm formally[5] exhibits

[5]One of the limitations of domain decomposition based codes is that the smaller the system, the fewer CPUs can be used before performance levels off. This is caused by the domains becoming too small, and resulting in increased communication.

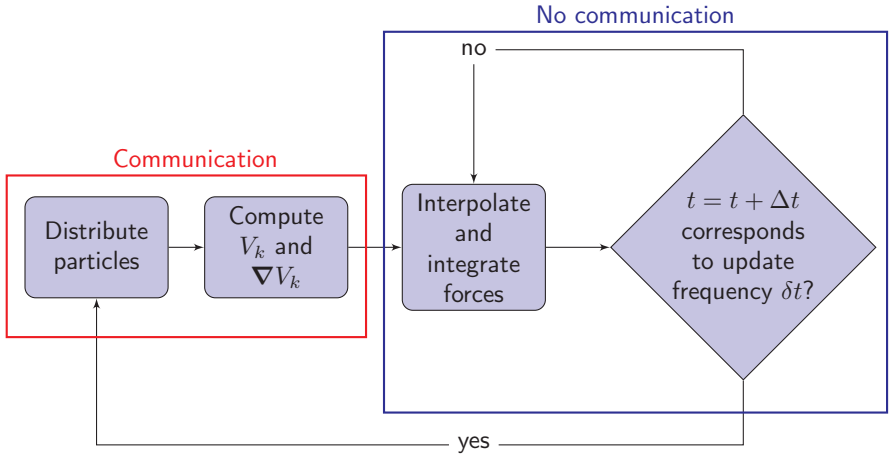

Figure 2.5: Algorithm for time propagation of hPF-MD and which parts that require communication.

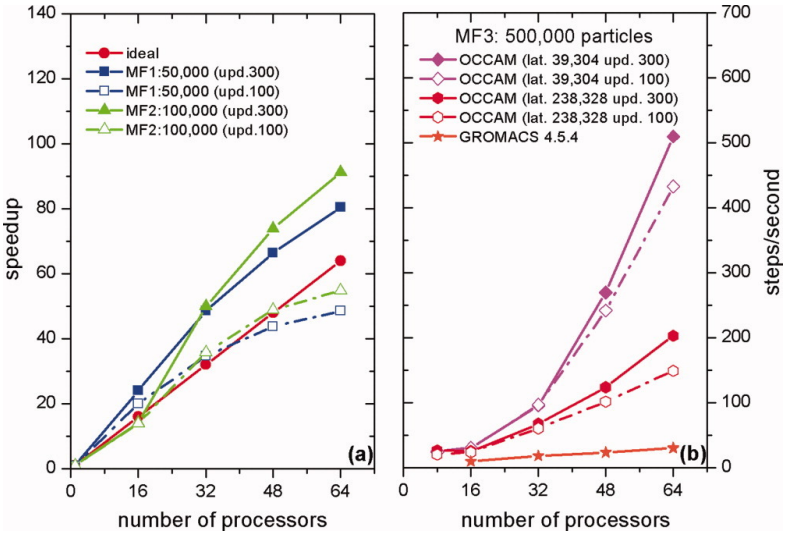

Figure 2.6: Benchmark of hPF-MD (particle-field) against GROMACS (Lennard-Jones) for monoatomic fluid systems. Left) Speedup of OCCAM with the number of processors for 50 000 particles (MF1) and 100 000 particles (MF2) and update frequency 100 and 300. Right) Performances of OCCAM program as steps/s for a system of 500 000 particles (MF3) in comparison to a Lennard-Jones fluid simulation with GROMACS 4.5.4. Results of OCCAM were obtained by using 39 304 and 238 328 lattice points with update frequency of 100 and 300. Figure is reprinted from ref. [47].

strong scaling: linear scaling for large and small systems [27]. However, for many MPI-tasks and large systems involving large grids, combining densities can become computationally expensive. Secondly, operations on the grid, such as finding the derivative of the external potential, are also computationally costly as these operations are only performed by a single MPI-task. For small systems, such operations are cheap, while for larger systems such computation can lead to a flatting out of performance. A benchmark plot of hPF-MD implemented into OCCAM [47] is presented in Figure 2.6. We recognize several of the main features which is formally exhibited by hPF-MD: close to linear scaling, and increased efficiency with lower update frequency and fewer grid points.

2.3.4 The OCCAM code

All the methods presented in this thesis were implemented into the OCCAM code[6]. The OCCAM code is a specialized molecular dynamics software for running hPF-MD. The first serial version was developed by Milano and Kawakatsu in 2009 [28] and included all the basic ingredients needed for MD and force calculation. The code was eventually extended to MPI parallelization, where molecules are divided among processors [47]. As a result of the work in this thesis, the code has been extended to include electrostatics and constant-pressure simulations. Additionally, there is an unofficial version of OCCAM including multiple-particle collisions [43].

Similar approaches and codes While we have stressed the connection between hPF-MD with SCMF, there are additional related methods and codes. GALAMOST [48], a GPU-based MD package aimed at coarse-grained simulations, has an implementation of hPF-MD with the interaction energies described in 2.2 and electrostatics [49]. Recently, Doros Theodorou [38] has developed a hPF-BD/kMC method where particles are propagated by Brownian dynamics or kinetic Monte Carlo.

[6] `http://www.occammd.org/`

Chapter 3

Introduction to the papers

This thesis contains six papers, three of which have been published in peer-reviewed journals, one has been accepted after peer-review and two are manuscripts in preparation. The papers are introduced in thematic order.

3.1 Polypeptides

Paper I: *Hybrid Particle-Field Model for Conformational Dynamics of Peptide Chains*
Sigbjørn Løland Bore, Giuseppe Milano and Michele Cascella
Journal of Chemical Theory and Computation **14**, 1120–1130 (2018)

Summary

Conformational dynamics of proteins is very important in biology and medicine (protein folding, protein regulation, peptide-mediated signaling, antimicrobial peptide action, etc.), but it is also very difficult to represent by all-atom simulations. Speeding up peptide studies by reliable CG models is crucial. Paper I proposes the first ever hPF model for polypeptides. In accordance with the hPF formalism, an underlying model for the single-molecule Hamiltonian H_0 is required. The model for H_0 is based on a two-bead representation of each amino acid, with one bead being placed at the Cα-position and another placed at the center of mass of the sidechain. It is known that directional interactions are necessary to stabilize protein secondary structures [50]. To introduce this into our model we employ the dipole reconstruction method, developed by Cascella and coworkers [51], for the dipole moment of the peptide bond. Due to the rigidity of the peptide bonds, the fundamental degrees of freedom determining the conformation of the polypeptide, are the bending and torsional angles of the Cαs. It has been shown by Tozzini et al. [50] that these angles exhibit strong correlations. To take this into account, we employ a combined torsional-bending potential. On top of this, we introduce a propensity potential which models the specific amino-acid propensity towards certain conformations. The intermolecular interactions are modeled by the interaction energy with mixing and compressibility terms as described in Section 2.2. In particular, we adopt a hydrophobic-polar (HP) model, in which side chains are categorized as either hydrophobic or polar by their $\tilde{\chi}$ parameter with water. The model was implemented into OCCAM.

Through a series of test cases, we demonstrated that the model is able to reproduce key elementary structural elements, such as α-helices and β-structures. Furthermore, we explored the phase diagram of homo sequence polypeptide in

terms of the propensity and $\tilde{\chi}$ parameters, and showed that the $\tilde{\chi}$ parameter is able to signal environmental effects on conformations. Using amphiphilic sequences, we showed that the $\tilde{\chi}$ parameter can facilitate super secondary structures, with tertiary and quaternary structures requiring the $\tilde{\chi}$ interactions. Finally, when combining the polypeptide model with the model for dioleoylphosphatidylcholine (DOPC) [31], we found that by only having $\tilde{\chi}$ interactions between the polypeptide and the membrane, the multiphase environment stabilizes secondary structures that are unstable in a homogeneous environment.

Statement of significance

In the past, the hPF formalism has predominantly been applied on simpler molecules. Therefore, its application onto polypeptides constitute a major widening of the scope of hPF modeling towards complex biological molecules. At the present stage, the model is a toy model and should not be used to model specific sequences. However, with careful parameterization of propensity and the $\tilde{\chi}$ matrix, chemically specific models can be made. Lastly, by having this model within the hPF approach opens up for studying large assemblies of polypeptides and combining them with other hPF models.

Contributions

Paper I The model was primarily developed by Michele Cascella and me, and I implemented it into OCCAM. The paper was written by Michele Cascella and me with corrections and suggestions from Giuseppe Milano.

3.2 Electrostatics

Paper II: *Hybrid Particle-Field Molecular Dynamics Simulations of Charged Amphiphiles in Aqueous Environment*
Hima Bindu Kolli, Antonio De Nicola, **Sigbjørn Løland Bore**, Ken Schäfer, Gregor Diezemann, Jürgen Gauss, Toshihiro Kawakatsu, Zhongyuan Lu, You-Liang Zhu, Giuseppe Milano and Michele Cascella
Journal of Chemical Theory and Computation **14**, 4928–4937 (2018)

Paper III: *Mesoscale Electrostatics Driving Particle Dynamics in Nonhomogeneous Dielectrics*
Sigbjørn Løland Bore, Hima Bindu Kolli, Toshihiro Kawakatsu, Giuseppe Milano and Michele Cascella
Journal of Chemical Theory and Computation **15**, 2033 (2019)

Summary

The polyelectrolytic nature of biological molecules, such as polypeptides, DNA, charged surfactants, and lipids, is an important factor determining key prop-

erties, including phase behavior, structure and function. Since the use of pair interactions slow down the hPF approach, specialized methods which model electrostatics with particle-field interactions are needed. Paper II and III concern two complementary models for electrostatics within the hPF-approach.

Paper II is on an implementation and application of the hPF method for computing electrostatic interactions. The method was first developed by Zhu et al. [49] and is an adaptation of the Particle-Mesh-Ewald approach [52, 53], which replaces short-range pair interactions with short-range particle-density interactions. The combination of long and short-range interactions results in an external potential which acts on charged particles. A benchmark of the method and the implementation was carried out through applications on palmitoyloleoylphosphatidylglycerol (POPG) lipid membrane and sodium dodecyl sulfate (SDS) surfactants. We demonstrated that upon proper calibration of simulation parameters, in particular the relative dielectric constant, we achieve an excellent POPG lipid bilayer structure and concentration dependence of the SDS assembly into microtubular aggregates.

For multiphase systems, and in particular lipid bilayers, the effective relative dielectrics can change by almost two orders of magnitude. This has a profound effect on the screening of electrostatic interactions. Therefore, the use of a constant dielectrics constitutes a major approximation. Paper III expands on Paper II by considering a density dependent dielectrics and a electrostatic potential governed by the *generalized Poisson equation*. From the total electrostatic energy of the system and following the procedure as described in Section 2.1, the external potential is obtained. The external potential acting on particles contains, not only the Coulomb term, but also a polarization term acting on all particles. Through a series of applications, we verified the model's ability to correctly reproduce partition phenomena of ions in multiphase systems, benchmarking against the Born model for solvation of ideal ions. Lastly, applications on ion distribution around lipid bilayers show that the use of density-dependent dielectrics correctly predicts low penetrability of ions into the POPG lipid bilayer.

Statement of significance

The extension of software and methodology to polyelectrolytes significantly expands the applicability of the hPF approach. This is evidenced by the applications presented in Paper V on lipid A and in study in Paper VI on SDS, which before this work was beyond the scope for the OCCAM code. While the density-dependent dielectric method of Paper III was derived within the hPF formalism, it is applicable to all coarse-grained models employing explicit solvation. In particular, it could be used with MARTINI [17] on the standard nonpolarizable water model.

Contributions

Paper II My main contribution to the work presented in Paper II was to test the code by computing the forces between two charged point particles against analytic forces. I discovered a major bug in the original implementation which was related to the ordering of wavenumbers in the FFTW library [54] (fast Fourier transform). Finally, I produced the data for the benchmark of forces and assisted in writing the manuscript.

Paper III The method was derived by me and Michele Cascella. I implemented the method, on top of code developed by Hima Bindu Kolli, in addition to conducting the simulations. The manuscript was written by me and Michele Cascella with corrections from the co-authors.

3.3 Multiphase electrolytic systems

Paper IV: *Aggregation of Lipid A Variants: a Hybrid Particle-Field Model*
Antonio De Nicola, Thereza A. Soares, Denys E. S. Santos, **Sigbjørn Løland Bore**, G. J. Agur Sevink, Michele Cascella and Giuseppe Milano
Accepted for publication in BBA – General Subjects (2020)

Paper V: *Beyond the Molecular Packing Model: Understanding Morphological Transitions of Charged Surfactant Micelles*
Ken Schäfer, Hima Bindu Kolli, Mikkel Killingmoe Christensen, **Sigbjørn Løland Bore**, Gregor Diezemann, Jürgen Gauss, Giuseppe Milano, Reidar Lund and Michele Cascella
(In preparation)

Summary

Despite the growing sophistication and availability of experimental methods for probing multiphase soft matter and biological systems, primarily scattering experiments, it is at present, beyond regular or simple shapes, difficult to interpret what the assembled structure is like on a molecular level. From the standpoint of a theoretical chemist, multiphase electrolytic systems are particularly challenging. First, their multiphase nature is caused by *collective interactions*, and therefore an understanding cannot be ascertained by only considering the constituents in isolation. Second, the phase behavior is dictated by the long-range electrostatics, and therefore nonlocal effects and salt dependency must be accounted for. The mesoscale length and time scales prohibit the study by high resolution computational methods, such as quantum mechanical and even all-atom approaches, but are however perfectly suited for CG modeling.

Paper IV develops a hPF model for Lipid A. This lipid is particularly relevant in biology, as it is one of three components of bacterial lipopolysaccharides which comprise the outer membrane of Gram-negative bacteria. It is a complex lipid

with four tails and two negatively charged heads. We use the electrostatics code developed in Paper II and parameterize the model to be consistent with the all-atom lipid A model developed by Thereza A. Soares [55]. As parameterization procedure, we compared density profiles of all-atom simulations of solvated lipid A bilayer and tuned model parameters, mainly the relative dielectric and the $\tilde{\chi}$ matrix, to reproduce the density profiles. Upon calibration, we found that our model reproduces the structural properties of all-atom simulations and that it gives a description of phase behavior which is consistent with experiments.

Paper V continues the work of Paper II by considering in detail the phase diagram of SDS in terms of concentration of SDS and salt. In particular, we use SDS as a model system to study the validity of the packing model in systems with long range electrostatic interactions. The phase diagram is explored by scattering experiments, SAXS and SANS, and simulations with hPF-MD using the code and model developed in Paper II. The simulations and experiments are in qualitative agreement, exhibiting the same phase diagram, but with transitions from spherical to tubular micelles at quantitatively different concentrations. We find by examining the molecular structures in the simulations that the transition is not signaled by a change in packing parameter, but rather a change in the distribution of counterions near the charged heads of the SDS. On this basis, we developed a simple electrostatics model for the coordination of ions close to the SDS heads. The model suggests that at high concentration of salt, counterions transition from localized to unlocalized binding between the heads of SDS. From this insight, we propose a mechanism for the transition where the fundamental packing piece changes from a single SDS (packing parameter of a cone) to two SDS (packing parameter of cake piece). With such a transition, the packing model adequately explains the transition.

Statement of significance

Lipid A is the prototypic example of a complex, slow-diffusing lipid, for which an accelerated method like hPF is particularly suited. Having developed and benchmarked our model, it is now ready to be further applied on larger systems and to explore the phase diagram in detail. One possibility is to couple this new model with a density-dependent dielectric. This will likely require a reparameterization of the $\tilde{\chi}$ matrix, but opens up for treating electrostatics more accurately.

Paper V further strengthens the validity of the SDS model, proposed in Paper II, in accurately representing the phase diagram of SDS. While this molecule has been studied for more than a century, the molecular mechanism dictating their morphology is unknown. This paper proposes a simple model for the transition for SDS and gives a general insight to how electrostatics can affect the morphological transitions of small charged copolymers.

Contributions

Paper IV My main contribution to Paper IV was in supervising the use of the electrostatics code, and incorporating and testing of the rotational invariant gradient used for the vesicle simulations. I also contributed to the writing of the paper, with particular focus on the methods part.

Paper V I supervised the use of electrostatics code in the simulations by Ken Schäfer, and contributed with corrections and suggestions to the manuscript.

3.4 Constant-pressure simulations

Paper VI: *Hybrid Particle-Field Molecular Dynamics Under Constant Pressure*
Sigbjørn Løland Bore, Hima Bindu Kolli, Antonio De Nicola, Maksym Byshkin, Toshihiro Kawakatsu, Giuseppe Milano and Michele Cascella
(In preparation)

Summary

The investigation of multiphase systems often requires probing the system at constant pressure. For example, lipid membranes are typically studied under constant tension conditions and is only obtained with a barostat. So far, hPF simulations have been limited to constant-volume conditions. Paper VI presents the first ever constant-pressure simulations with hPF. This development was enabled by reformulating the interaction energy, by introducing an equation of state parameter and a square-gradient density term which models interfacial energy. We first considered the model for water (four water molecules per CG-bead) and parameterized the equation-of-state parameter, such that it reproduced the correct density of water at ambient conditions. Using this parameterization, we explored the behavior of binary fluids, demonstrating that $\tilde{\chi}$ interaction results in strong excess volume effects and that the square-gradient term controls the shape of a phase separated droplet. Moreover, we considered the already developed model for DPPC by de Nicola [31], and added a square-gradient interaction between carbon and water. We found that the square-gradient term is necessary to obtain the correct area per lipid for a DPPC lipid bilayer. Moreover, initial simulations of a small lipid vesicle indicate that it improves the agreement of all properties with what has been reported for the MARTINI model [56].

Statement of significance

Paper VI expands the applicability of hPF simulations. With this new formalism, one can equilibrate simulation boxes and study a range of phenomena requiring constant-pressure simulations. The incorporation of square-gradient

has a large potential for extending the capability of hPF-models to more accurately represent surface phenomena. Despite the very simplified parameterization with square-gradient interactions between only carbon and water, we obtained remarkable improvement on many properties of the lipid model. A full reparameterization is a promising route towards hPF models with chemical specificity and accurate modeling of surface phenomena. Finally, since the method was derived for hPF-methods in general, it is also applicable for the *single chain mean field* method.

Contributions

Paper VI The new formalism, derivations and its implementation, that allowed for constant-pressure simulations with hPF, were developed by me. The applications were conducted by me. The paper was written by Michele Cascella and myself, with corrections from the other coauthors.

Chapter 4

Conclusion and outlook

This thesis has considered hPF methods for biological systems. The primary research output of this work lies in its advancement of *computational methodology*, *modeling* and *software*. These advancements allow us to use the hPF method to study new biological systems with complex electrostatics and multiphase behavior, which prior this work were beyond the scope of existing methods and implementations. This is demonstrated by the new applications of hPF-MD on a wide range of new systems, including polypeptides, charged membranes, SDS and Lipid A.

However, since the hPF-MD method is quite new, there are still many avenues to explore. In addition to the new developments, much experience has been gained on the limitations of hPF-MD. From this I strongly believe that the following lines of research can extend the capabilities of the hPF-MD approach.

Systematic parameterization procedures One of the major challenges when developing new models within the hPF-MD approach, lies in choosing the $\tilde{\chi}$ matrix for the coarse-grained beads. In the models that have been developed thus far, two main approaches have been applied for obtaining $\tilde{\chi}$ matrix: from Flory-Huggins lattice theory using data (either experimental or simulation) or from optimization by hand to reproduce some statistics. It is difficult for a specific CG bead, such as a part of the protein, to estimate the value of $\tilde{\chi}$. Moreover, the value of $\tilde{\chi}$ depends on the modeling, for example whether there is electrostatic interactions or not. I have therefore started to explore systematic procedures for choosing $\tilde{\chi}$ in collaboration with PhD student Morten Ledum, who is continuing this work. The idea is to focus less on the theory and interpretation behind the simulation parameters and more on the output of these parameters in terms of statistical properties which determine the fitness of the CG model. In particular, we have started a project on developing global optimization routines, based on machine learning of the simulation parameters, focusing on $\tilde{\chi}$ with the test case of phospholipids. This is also a promising route for obtaining chemically specific parameterization of polypeptides, lipids and ions. Moreover, it will make it less difficult to incorporate new methods and develop new models in the future when using this approach.

New interaction energies In the hPF models discussed in this thesis, we have employed mixing and compressibility terms. These terms are on one hand very simple, but on the other hand limited in their capability to represent physical phenomena. In particular, the compressibility term promotes all particles to occupy the same volume, which can be detrimental for multiphase systems where particles do not necessarily occupy the same volume. One worthwhile route (for

NVT-simulations) is to use the equivalence between the κ and the $\tilde{\chi}$ in (2.33) to add diagonal terms to the $\tilde{\chi}$ matrix. Such an approach should be able to capture particles occupying different molar volumes. Another approach is to leave behind this simplified compressibility model and move towards a more complex functional dependency on densities for the interaction energies, such as in the hPF method of Theodorou [38]. These models are also likely to be more suitable to studying the response of a system under changing thermodynamic conditions.

Rigorous force computation The computation of the forces in hPF-MD is in my opinion a topic which should be examined in detail. It is known that the shape of large vesicles are affected by the grid. For example using the standard central finite differences produces for a large vesicle with a cubic shape oriented according to the grid [43]. This can be remedied by using finite-differences stencils with rotational invariance. Derivatives computed with the fast Fourier transform might produce even better results, but it is dependent on a smooth density to avoid producing spurious oscillations. Also important is the interpolation of densities and forces. There is an extensive literature on particle-mesh methods [42], which can serve as starting point for this. A benchmark of the force computation should determine what is needed to obtain forces on particles (due to W) which are translation invariant, rotational invariant and conserve the energy. Improving these aspects will likely result in an improved method with less numerical artifacts. Also, the quasi-instantaneous approximation should be revisited more in detail. Which artifacts does it produce? Can we make it more efficient by not interpolating forces onto particles every time step? A multiple time step algorithm separating bonded forces and particle-field forces is a promising route for increasing the computational efficiency of the hPF-MD.

Coupling differential equations to hPF The procedure that was used to derive the forces on the particles in density-dependent dielectric method of Paper III, is generalizable. It tells us how we, given an interaction energy and a corresponding partial differential equation, compute the forces on the particles composing the density. While solving partial differential equations with high frequency appears computationally expensive on the surface, there are several factors which reduces the computational cost. First, the density of the particles is not an unknown, thus the complexity of many partial differential equations is reduced. For example, we only had to consider the generalized Poisson equation and not the Poisson-Boltzmann equation in Paper III. Lastly, since most methods for solving partial differential equations are based on iterations on an initial trial solution, one can reuse the previous solution to reduce the amount of needed iterations. In particular, it would be very interesting to obtain correlation between the motion of particles through hydrodynamic interactions.

Improved software The parallelization in Zhao et al. [47] drastically increased the capability of OCCAM to simulate large systems. However, given the huge

increase of computational resources, it is now in need of modernization. The newly developed SCMF code by Schneider and Müller [27] tells us how this can be carried out. First and foremost, the current version of OCCAM uses an MPI-parallelization with one CPU per MPI-task. This is problematic for large systems, as some large operations are conducted on a single CPU. This can to a large degree be fixed by an hybrid MPI and openMP parallelization. In this strategy, each MPI-task uses many CPUs (typically a whole node) and are parallelized by openMP. This alone will improve the capability of OCCAM to simulate systems that are about 40 times larger (depending on the amount of CPUs per node). Another promising route is to develop a GPU-version. The low level communication in the hPF-MD approach will make such an implementation easier than for particle-particle approaches and indeed more efficient. Such a code will allow for very cheap investigation of large systems.

Closing remarks

Limiting ourselves to the hPF-MD, we now revisit the question posed at the beginning of this thesis: "Can we understand complex biological systems, such as bacterias, organelles and cells, by considering their constituents, the atoms?" First, whether hPF-MD simulations can provide an understanding on these systems is dependent on its capability of describing the chemistry dictating the behavior of these systems. The developments presented in this thesis, in particular on electrostatics, surface tension, and in models for polypeptides and Lipid A, provide us with many of the necessary ingredients needed to model, though simplified, such systems. However, whether these models provide an understanding is dependent on their quality, and, in particular, the polypeptide model will require further parameterization. Second, in terms of size these systems are on the micrometer scale, and involve of the order billions to trillions of atoms. By the level of coarse-graining that has been adopted in this thesis, the number of degrees of freedom is reduced by a factor of about 12. With the current version of OCCAM, we have simulated systems up 20 million beads or 200 million atoms. Therefore, these systems are currently too large. However, with the development of the software as outlined in the previous paragraph, we should be able to approach dimensions of relevance for organelles, viruses, and even parts of a cell. To conclude, we are not there yet, but the work of this thesis has led us closer, and many extremely interesting systems are within the horizon of hPF-MD.

Bibliography

[1] P. A. M. Dirac, *Quantum mechanics of many-electron systems*, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **123**, 714–733 (1929).

[2] M. Cascella and S. Vanni, *Toward accurate coarse-graining approaches for protein and membrane simulations*, Chemical Modelling **12**, 9781782622703–00001 (2015).

[3] M. J. F. et al., *Gaussian 16 Revision C.01*, Gaussian Inc. Wallingford CT, 2016.

[4] P. Carloni, U. Rothlisberger, and M. Parrinello, *The role and perspective of ab initio molecular dynamics in the study of biological systems*, Accounts of Chemical Research **35**, 455–464 (2002).

[5] K. Meier, A. Choutko, J. Dolenc, A. P. Eichenberger, S. Riniker, and W. F. Van Gunsteren, *Multi-resolution simulation of biomolecular systems: a review of methodological issues*, Angewandte Chemie International Edition **52**, 2820–2834 (2013).

[6] A. D. MacKerell Jr, D. Bashford, M. Bellott, R. L. Dunbrack Jr, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H Guo, S. Ha, et al., *All-atom empirical potential for molecular modeling and dynamics studies of proteins*, The Journal of Physical Chemistry B **102**, 3586–3616 (1998).

[7] V. A. Voelz, G. R. Bowman, K. Beauchamp, and V. S. Pande, *Molecular simulation of ab initio protein folding for a millisecond folder ntl9 (1- 39)*, Journal of the American Chemical Society **132**, 1526–1528 (2010).

[8] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, *How fast-folding proteins fold*, Science **334**, 517–520 (2011).

[9] R. O. Dror, H. F. Green, C. Valant, D. W. Borhani, J. R. Valcourt, A. C. Pan, D. H. Arlow, M. Canals, J. R. Lane, R. Rahmani, et al., *Structural basis for modulation of a g-protein-coupled receptor by allosteric drugs*, Nature **503**, 295 (2013).

[10] G. Zhao, J. R. Perilla, E. L. Yufenyuy, X. Meng, B. Chen, J. Ning, J. Ahn, A. M. Gronenborn, K. Schulten, C. Aiken, et al., *Mature hiv-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics*, Nature **497**, 643 (2013).

[11] D. Bassolino-Klimas, H. E. Alper, and T. R. Stouch, *Solute diffusion in lipid bilayer membranes: an atomic level study by molecular dynamics simulation*, Biochemistry **32**, 12624–12637 (1993).

[12] C. Hong, D. P. Tieleman, and Y. Wang, *Microsecond molecular dynamics simulations of lipid mixing*, Langmuir **30**, 11993–12001 (2014).

[13] G. A. Voth, *Coarse-graining of condensed phase and biomolecular systems* (CRC press, 2008).

[14] S. Izvekov and G. A. Voth, *A multiscale coarse-graining method for biomolecular systems*, The Journal of Physical Chemistry B **109**, 2469–2473 (2005).

[15] H. I. Ingólfsson, C. A. Lopez, J. J. Uusitalo, D. H. de Jong, S. M. Gopal, X. Periole, and S. J. Marrink, *The power of coarse graining in biomolecular simulations*, Wiley Interdisciplinary Reviews: Computational Molecular Science **4**, 225–248 (2014).

[16] T. A. Soares, S. Vanni, G. Milano, and M. Cascella, *Toward chemically resolved computer simulations of dynamics and remodeling of biological membranes*, The Journal of Physical Chemistry Letters **8**, 3586–3594 (2017).

[17] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. De Vries, *The martini force field: coarse grained model for biomolecular simulations*, The Journal of Physical Chemistry B **111**, 7812–7824 (2007).

[18] D. Reith, M. Pütz, and F. Müller-Plathe, *Deriving effective mesoscale potentials from atomistic simulations*, Journal of Computational Chemistry **24**, 1624–1636 (2003).

[19] A. P. Lyubartsev, in *Coarse-grained modeling of biomolecules* (CRC Press, 2017), pp. 29–54.

[20] B. Hess, C. Kutzner, D. Van Der Spoel, and E. Lindahl, *Gromacs 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation*, Journal of Chemical Theory and Computation **4**, 435–447 (2008).

[21] K. C. Daoulas and M. Müller, *Single chain in mean field simulations: quasi-instantaneous field approximation and quantitative comparison with monte carlo simulations*, The Journal of Chemical Physics **125**, 184904 (2006).

[22] K. C. Daoulas, M. Müller, J. J. De Pablo, P. F. Nealey, and G. D. Smith, *Morphology of multi-component polymer systems: single chain in mean field simulation studies*, Soft Matter **2**, 573–583 (2006).

[23] G. Fredrickson et al., *The equilibrium theory of inhomogeneous polymers*, Vol. 134 (Oxford University Press on Demand, 2006).

[24] F. Schmid, *Self-consistent-field theories for complex fluids*, Journal of Physics: Condensed Matter **10**, 8105 (1998).

[25] F. Schmid, *Theory and simulation of multiphase polymer systems*, Handbook of multiphase polymer systems, 31–80 (2011).

[26] J. A. Anderson, C. D. Lorenz, and A. Travesset, *General purpose molecular dynamics simulations fully implemented on graphics processing units*, Journal of Computational Physics **227**, 5342–5359 (2008).

[27] L. Schneider and M. Müller, *Multi-architecture monte-carlo (mc) simulation of soft coarse-grained polymeric materials: soft coarse grained monte-carlo acceleration (soma)*, Computer Physics Communications **235**, 463–476 (2019).

[28] G. Milano and T. Kawakatsu, *Hybrid particle-field molecular dynamics simulations for dense polymer systems*, The Journal of Chemical Physics **130**, 214106 (2009).

[29] A. De Nicola, T. Kawakatsu, and G. Milano, *Generation of well-relaxed all-atom models of large molecular weight polymer melts: a hybrid particle-continuum approach based on particle-field molecular dynamics simulations*, Journal of Chemical Theory and Computation **10**, 5651–5667 (2014).

[30] A. De Nicola, Y. Zhao, T. Kawakatsu, D. Roccatano, and G. Milano, *Validation of a hybrid md-scf coarse-grained model for dppc in non-lamellar phases*, Theoretical Chemistry Accounts **131**, 1167 (2012).

[31] A. De Nicola, Y. Zhao, T. Kawakatsu, D. Roccatano, and G. Milano, *Hybrid particle-field coarse-grained models for biological phospholipids*, Journal of Chemical Theory and Computation **7**, 2947–2962 (2011).

[32] G. Milano, T. Kawakatsu, and A. De Nicola, *A hybrid particle–field molecular dynamics approach: a route toward efficient coarse-grained models for biomembranes*, Physical Biology **10**, 045007 (2013).

[33] A. De Nicola, T. Kawakatsu, C. Rosano, M. Celino, M. Rocco, and G. Milano, *Self-assembly of triton x-100 in water solutions: a multiscale simulation study linking mesoscale to atomistic models*, Journal of Chemical Theory and Computation **11**, 4959–4971 (2015).

[34] Y. Zhao, M. Byshkin, Y. Cong, T. Kawakatsu, L. Guadagno, A. De Nicola, N. Yu, G. Milano, and B. Dong, *Self-assembly of carbon nanotubes in polymer melts: simulation of structural and electrical behaviour by hybrid particle-field molecular dynamics*, Nanoscale **8**, 15538–15552 (2016).

[35] S. J. Marrink, V. Corradi, P. C. Souza, H. I. Ingólfsson, D. P. Tieleman, and M. S. Sansom, *Computational modeling of realistic cell membranes*, Chemical Reviews **119**, 6184–6226 (2019).

[36] D. C. Rapaport, *The art of molecular dynamics simulation* (Cambridge university press, 2004).

[37] D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, Vol. 1 (Elsevier, 2001).

[38] G. G. Vogiatzis, G. Megariotis, and D. N. Theodorou, *Equation of state based slip spring model for entangled polymer dynamics*, Macromolecules **50**, 3004–3029 (2017).

[39] M. Müller, *Studying amphiphilic self-assembly with soft coarse-grained models*, Journal of Statistical Physics **145**, 967–1016 (2011).

[40] T. Kawakatsu, *Statistical physics of polymers: an introduction* (Springer Science & Business Media, 2013).

[41] G. H. Fredrickson, V. Ganesan, and F. Drolet, *Field-theoretic computer simulation methods for polymers and complex fluids*, Macromolecules **35**, 16–39 (2002).

[42] R. W. Hockney and J. W. Eastwood, *Computer simulation using particles* (crc Press, 1988).

[43] G. Sevink, F Schmid, T Kawakatsu, and G Milano, *Combining cell-based hydrodynamics with hybrid particle-field simulations: efficient and realistic simulation of structuring dynamics*, Soft Matter **13**, 1594–1623 (2017).

[44] R. Courant, K. Friedrichs, and H. Lewy, *On the partial difference equations of mathematical physics*, IBM journal of Research and Development **11**, 215–234 (1967).

[45] M. Deserno and C. Holm, *How to mesh up ewald sums. i. a theoretical and numerical comparison of various particle mesh routines*, The Journal of Chemical Physics **109**, 7678–7693 (1998).

[46] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, *Scalable molecular dynamics with namd*, Journal of Computational Chemistry **26**, 1781–1802 (2005).

[47] Y. Zhao, A. De Nicola, T. Kawakatsu, and G. Milano, *Hybrid particle-field molecular dynamics simulations: parallelization and benchmarks*, Journal of Computational Chemistry **33**, 868–880 (2012).

[48] Y.-L. Zhu, H. Liu, Z.-W. Li, H.-J. Qian, G. Milano, and Z.-Y. Lu, *Galamost: gpu-accelerated large-scale molecular simulation toolkit*, Journal of Computational Chemistry **34**, 2197–2211 (2013).

[49] Y.-L. Zhu, Z.-Y. Lu, G. Milano, A.-C. Shi, and Z.-Y. Sun, *Hybrid particle–field molecular dynamics simulation for polyelectrolyte systems*, Physical Chemistry Chemical Physics **18**, 9799–9808 (2016).

[50] V. Tozzini, *Minimalist models for proteins: a comparative analysis*, Quarterly reviews of biophysics **43**, 333–371 (2010).

[51] D. Alemani, F. Collu, M. Cascella, and M. Dal Peraro, *A nonradial coarse-grained potential for proteins produces naturally stable secondary structure elements*, Journal of Chemical Theory and Computation **6**, 315–324 (2009).

[52] T. Darden, D. York, and L. Pedersen, *Particle mesh ewald: an $N \cdot \log(N)$ method for ewald sums in large systems*, The Journal of chemical physics **98**, 10089–10092 (1993).

[53] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, *A smooth particle mesh ewald method*, The Journal of chemical physics **103**, 8577–8593 (1995).

[54] M. Frigo and S. G. Johnson, "Fftw: an adaptive software architecture for the fft", in Proceedings of the 1998 ieee international conference on acoustics, speech and signal processing, icassp'98 (cat. no. 98ch36181), Vol. 3 (IEEE, 1998), pp. 1381–1384.

[55] K. N. Kirschner, R. D. Lins, A. Maass, and T. A. Soares, *A glycam-based force field for simulations of lipopolysaccharide membranes: parametrization and validation*, Journal of chemical theory and computation **8**, 4719–4731 (2012).

[56] H. J. Risselada and S. J. Marrink, *Curvature effects on lipid packing and dynamics in liposomes revealed by coarse grained molecular dynamics simulations*, Physical Chemistry Chemical Physics **11**, 2056–2067 (2009).