Unfolding the prospects of computational (bio)materials modeling

The Journal of Chemical Physics

Unfolding the prospects of computational (bio)materials modeling

Cite as: J. Chem. Phys. **153**, 100901 (2020); https://doi.org/10.1063/5.0019773
Submitted: 25 June 2020 . Accepted: 10 August 2020 . Published Online: 08 September 2020

G. J. Agur Sevink, Jozef Adam Liwo, Pietro Asinari, Donal MacKernan, Giuseppe Milano, and Ignacio Pagonabarraga

The Journal of Chemical Physics PERSPECTIVE scitation.org/journal/jcp

Unfolding the prospects of computational (bio)materials modeling

Cite as: J. Chem. Phys. 153, 100901 (2020); doi: 10.1063/5.0019773
Submitted: 25 June 2020 • Accepted: 10 August 2020 •
Published Online: 8 September 2020

G. J. Agur Sevink,[1,a)] Jozef Adam Liwo,[2] Pietro Asinari,[3] Donal MacKernan,[4] Giuseppe Milano,[5] and Ignacio Pagonabarraga[6,7,8]

AFFILIATIONS

[1]Leiden Institute of Chemistry, Leiden University, P.O. Box 9502, 2300 RA Leiden, The Netherlands
[2]Laboratory of Molecular Modeling, Faculty of Chemistry, University of Gdansk, Wita Stwosza 63, 80-308 Gdansk, Poland
[3]Department of Energy, Politecnico di Torino, Corso Duca degli Abruzzi, 24, 10129 Turin, Italy
[4]UCD School of Physics, University College Dublin, Dublin 4, Ireland
[5]Theoretical Physical Chemistry, Organic Materials Modeling, Department of Organic Materials Science, Yamagata University, 4-3-16 Jonan Yonezawa, Yamagata-ken 992-8510, Japan
[6]CECAM Centre Européen de Calcul Atomique et Moléculaire, École Polytechnique Fédérale de Lausanne (EPFL), Batochime, Avenue Forel 2, Lausanne CH-1015, Switzerland
[7]Departament de Física de la Matéria Condensada, Universitat de Barcelona, Martí i Franquès 1, Barcelona 08028, Spain
[8]UBICS Institute of Complex Systems, Universitat de Barcelona, Martí i Franquès 1, Barcelona 08028, Spain

Note: This paper is part of the JCP Special Topic on Classical Molecular Dynamics (MD) Simulations: Codes, Algorithms, Force Fields, and Applications.
[a)]**Author to whom correspondence should be addressed:** a.sevink@lic.leidenuniv.nl

ABSTRACT

In this perspective communication, we briefly sketch the current state of computational (bio)material research and discuss possible solutions for the four challenges that have been increasingly identified within this community: (i) the desire to develop a unified framework for testing the consistency of implementation and physical accuracy for newly developed methodologies, (ii) the selection of a standard format that can deal with the diversity of simulation data and at the same time simplifies data storage, data exchange, and data reproduction, (iii) how to deal with the generation, storage, and analysis of massive data, and (iv) the benefits of efficient "core" engines. Expressed viewpoints are the result of discussions between computational stakeholders during a Lorentz center workshop with the prosaic title *Workshop on Multi-scale Modeling* and are aimed at (i) improving validation, reporting and reproducibility of computational results, (ii) improving data migration between simulation packages and with analysis tools, (iii) popularizing the use of coarse-grained and multi-scale computational tools among non-experts and opening up these modern computational developments to an extended user community.

I. INTRODUCTION

A wealth of quantum mechanical (QM) and classical (atomistic) molecular dynamics (MD) simulation engines developed, implemented, extended, and improved during the last 50 years have enabled researchers to obtain deep insight into a plethora of intricate processes that take place at the nanoscale (<10 nm, <1 μs) associated with a relatively small ensemble of molecules. A computational study of structure formation and processes at mesoscopic (10 nm–100 nm, μs s) scales, i.e., covering detail that directly relates to emergent material properties at the macroscopic scales, has only become possible since the advent of coarse-grained (CG) computational models that are based on effective molecular descriptions (or maps) obtained from averaging over chemical detail. It should be noted that, apart from extending the time and length scales, CG models enable us to identify the origin of the formation of organized structures of biomolecules and other self-organizing systems, e.g., liquid crystals, provided that the coarse graining is carried out in

a physics-based and scale-consistent manner. Moreover, by leaving out atomistic details, it simplifies our analysis of the behavior of large systems.

In addition to methods that represent systems on a single elementary scale, coarse-grained representations are also used in combination with atomistic ones. This practice, which is appropriately gathered under the general term multi-scale (MS) modeling, extends from the QM to the continuous level and can be executed in a hierarchical or concurrent manner. One early example of the latter is the Quantum Mechanics/Molecular Mechanics (QM/MM) treatment introduced by Warshel and Levitt in 1976,[1] in which the part of the system that undergoes chemical changes is described at the QM level, and the remaining part, for instance, a surrounding protein, is treated at the classical all-atom level.[2] More recent developments such as the Adaptive Resolution Scheme (AdResS)[3–6] adopt this idea of partitioning the simulation volume into separate regions, and to dynamically treat degrees of freedom in different regions at different resolutions, in order to enable one to combine the advantages of classical atomistic simulations with those from coarse-grained simulations. Here, we focus on a discussion of various conceptual and practical aspects of computational modeling approaches that have been developed to describe physical phenomena at a single atomistic or coarse-grained scale or at multiple scales. For additional background and technical details of CG and/or MS methodology, we refer the reader to recent reviews and references therein.[7–9] Although many of the basic techniques and considerations also apply to driven systems and active processes, albeit in a somewhat adapted form, we note that non-equilibrium approaches are not a focal point of this perspective.

Outside the small community that develops, validates, and employs coarse-grained and multi-scale descriptions, the concepts of averaging as well as the consequential prospects and limitations that play an important role in the accuracy of newly developed methodology are relatively unknown. Moreover, the need for many of these CG maps to be constituted *a priori*, using intricate tools that require both the technological and conceptual skills, as well as a general lack of freely available simulation engines for many of the new developments, renders the application of most of the CG and MS technology far from straightforward for non-experts. This results in a situation where the benefits of these new computational tools and, thus, the breakthrough that they may represent in various parts of the (bio)materials research community are not exploited to the full extent.

The CG community as a whole acknowledges this situation and is in the process of discussing solutions to some of the practical issues that are encountered in computational research. Here, we present the result of a 1-week discussion meeting between experts in QM, MD, and CG modeling that took place in Leiden, The Netherlands, during the summer of 2019, on four predefined themes. The Lorentz workshop was preceded by a 1-week summer school for Ph.Ds and postdocs that covered techniques from the quantum to the macroscale, with the aim of training junior community members in a broad range of multi-scale concepts and involving them in the discussions about the future of this field. Indeed, many that attended the summer school also actively contributed to the discussions in the workshop. To complement the proposed practical solutions, we shortly comment on what we identify as future directions in CG research.

The goal of this communication is to provide practical input and action points that the materials modeling community can use to improve standards and enable both the academic and industrials users easier and more flexible access to new data and developments, including a more insightful assessment for non-experts of which the method is most suited for their personal purposes. We particularly hope that it stimulates further discussions on these topics within this community.

II. HISTORICAL PERSPECTIVE

Continuum descriptions are among the oldest and most established methods for studying physical, chemical, or biological materials, mechanisms, and phenomena by theoretical or computational means. Continuum mechanics, pioneered by Cauchy in the 19th century, is very well suited for considering material mechanics at a very coarse spatial or temporal resolution, where the material behavior is governed by effective or averaged material characteristics or correlations. With governing equations that stem from balance laws for mass, momentum, and energy, as well as kinematic relations and constitutive equations, continuum mechanics is fundamentally incapable of evaluating ensemble behavior directly from microscopic properties and capturing specific contributions that cannot be described in terms of an ensemble. During the first half of the 20th century, quantum mechanics (QM) was formulated, providing a continuum description at the very fine spatial and temporal resolution for any type of material from first principles, obviating the need for any information about effective material characteristics. In particular, QM provides a genuine reductionistic approach toward materials characterization, i.e., a description in terms of the smallest individual parts and their interactions. Whereas there is a large community of quantum physicists/chemists that applies *ab initio* and tight-binding approaches for studying a variety of intricate phenomena at the quantum level, the computational costs associated with treating electronic degrees of freedom rule them out in practice for most supra-molecular systems and/or phenomena of interest. That said, neural network based potentials using molecular dynamics—but trained on *ab initio* calculations—are becoming increasingly potent, but currently are practically limited in the number of distinct atomic species that can be treated simultaneously.[10] Should quantum computing ever become practical, everything might change.

In the 1950s, Alder and Wainright started work on self-contained descriptions of matter at an atomic level of resolution depending only on the nuclear degrees of freedom, which is now recognized as molecular dynamics. Alder had been invited by Edward Teller to Lawrence Livermore Laboratories (LLL) to work on the thermodynamic equation of state problems needed in weapons research. LLL possessed essentially the most powerful computer in the world, a machine that frequently had spare capacity, which Alder was able to use for his interesting work on the side: molecular dynamics simulations. The ground-breaking idea of MD was the recognition that the Born–Oppenheimer assumption, i.e., electronic and nuclear degrees of freedom are decoupled, is valid for many systems of practical interest. Standard or classical molecular dynamics (MD) assumes that electrons follow—adiabatically—the classical nuclear motion and can consequently be integrated out completely.

Nuclei are, thus, assumed to evolve on a single Born–Oppenheimer potential-energy surface (typically, but not necessarily, given by the electronic ground state), which is generally approximated in terms of only few-body interactions. With a global potential surface that is reconstructed from a manageable sum of analytic and additive few-body contributions, the challenge of such force field (FF) based approaches is to derive empirical parameters in the FF by fitting or mapping these contributions to effective and often non-additive many-body interactions for the same system at the QM level. This variability in the mapping procedure and the chemical identity of the molecular systems considered has resulted in the development of special FFs for diverse molecular systems that have been extensively validated and optimized. Most FFs and simulation schemes do not allow bonds to be created or broken during the course of a simulation, but this limitation is frequently acceptable. The use of neural network based potentials can avoid such limitations.[11]

Established classical FFs that have been implemented in most software packages for the molecular simulation are the Assisted Model Building and Energy Refinement (AMBER), the Chemistry at HARvard Molecular Mechanics (CHARMM), the COmputer Simulation of MOlecular Structures-Nuclear Magnetic Resonance (COSMOS-NMR), the GROningen MOlecular Simulation (GROMOS), Optimized Potential for Liquid Simulations (OPLS), the Empirical Conformational Energy Program for Peptides (ECEPP), the Quantum mechanical extension of the Consistent Force Field method to PI electron molecules (QCFF/PI), the Universal Force Field (UFF), the Consistent Force Field (CFF), the Condensed-phase Optimized Molecular Potentials for Atomistic Simulation Studies (COMPASS), the Merck Molecular Force Field (MMFF) and variants, and the Transferable Potentials for Phase Equilibria (TraPPE).[12] As these FFs mainly differ in the map that is used for determination of the empirical FF parameters, the choice of a "proper" FF is sometimes a difficult one to make. Additionally, polarizable and reactive FFs attempt to include electronic structure variations at the nuclear level.

In recent years, the success of and insight gained by classical molecular modeling, in understanding the fundamentals of complex molecular phenomena, have triggered a strong desire to go beyond the limitations of the information that can be extracted from classical MD, especially the limitations that cannot be resolved by advances in computational efficiency. One essential issue is the prohibitively long times entailed to reliably estimate time averages for large complex systems such that systems' trajectories are sufficiently ergodic. Rare-event sampling methods provide one approach to address this issue, by focusing on regions in the phase space that are difficult to sample using standard molecular dynamics simulations. A promising alternative is to exploit the methodology that is able to conceptually address the hierarchy of time and length scales that is naturally present in many phenomena. Another and somewhat complementary reason for the growing interest in such methodologies is their ability to realistically treat interactions that many localized processes experience with their extended environment. As a result, during the last two decades, there has been a continuous effort to establish and subsequently improve techniques that further limit the dimensionality of the phase space by performing additional averaging over a less significant chemical detail, inspired by the often valid assumption that well-chosen domains within a molecule, for instance, neighboring groups of heavy atoms or all atoms in a conserved secondary structural motif, express some degree of co-operativity.

Effective supra-atomic molecular representations have been developed and used for diverse molecular systems in a variety of coarse-grained (CG) and multi-scale (MS) techniques. The results of these exercises suggest that in many cases, indeed, a substantial increase in the time- and length-scales is possible without losing the essentials of the phenomena that one intends to study. To specify these developments in constituting a (physical) description for structuring that is governed by thermodynamics, CG denotes the more general situation where one disregards particular resolved degrees of freedom, which are considered less relevant at the length and time scales of interest, to develop an effective physical description in terms of the remaining (and lesser-resolved) degrees of freedom, while MS aims at treating all degrees of freedom appropriately by passing on information between effective descriptions and levels that are appropriate for a particular resolution. Moreover, in such *vertical linking* MS strategies, simulations are carried out separately for each coarse-graining level, and the connection between different levels is established by the introduction of information at the appropriate level via a mapping procedure. Historically, these strategies are also known as hierarchical multi-scale methods. Although this approach is potentially very powerful, it remains hierarchical, and different scales in the system are treated separately, i.e., they remain uncoupled, while the mapping to increasing coarser descriptions brings along several issues. Yet, owing to the hierarchical nature, the direction of the map may be reversed. In such a situation, the lesser-resolved simulations can be exploited to scan the phase space of a system under study broadly but with fewer details, followed by exploration at a higher level of resolution. An exciting new development has concentrated on extending or hybridizing popular CG approaches toward even more efficiency or options for reinserting—while maintaining the original efficiency—some of the significant factors that were initially left out—specific chemistry, compositional/structural heterogeneity, environmental factors, and realistic dynamics—in an attempt to become increasingly predictive for realistic situations. Such *horizontal linking* MS strategies (also known as concurrent or hybrid) aim at merging different levels of representation within the same physical description, either by treating different components of the systems simultaneously, but with different details, or by switching representations for the same components.

III. THE SIGNIFICANCE OF AVERAGING

To better assess the quality of new computational developments at a coarse-grained level, particularly the ones that become available through open source software packages, it is important to understand the factors that play a role in their performance and validity. The constant stream of (sometimes mutually) validated studies for most of the FF-based molecular dynamics boosts the confidence of non-expert modelers that he/she simulates something sensible as long as approved FFs are used for the system at hand. Yet, the idea that one averages over degrees of freedom even for classical MD, in order to come to an effective and more efficient physical model at the atomic scale, is not always fully appreciated. In particular, because of averaging, MD is subject to many of the subtleties that

play a role in coarsening. One of these is that the quality of the effective physical model, e.g., the FF in MD, is as good as the accuracy of the map, which can be discussed both in terms of the ability to represent or absorb the lost degrees of freedom in the resulting model and in terms of the ability of the original model to sample all relevant parts of phase space that are needed for constituting a generally valid map.

Considering atomistic FFs as the fine-grained reference description, the objective of systematic mapping routines is to map the all-atom potential-energy surfaces of the components of the systems under study to equivalent effective potential-energy terms corresponding to the coarse-grained representation. These maps, which are known under the names Inverse Monte Carlo (IMC),[13] iterative Boltzmann inversion (IBI),[14] force matching (FM),[15–17] fluctuation matching,[18] minimization of relative entropy,[19] or conditional reversible work (CRW),[20,21] estimate effective CG potentials based on an equivalence relation, either derived from particle forces or via thermodynamic and/or structural signatures. Popular approaches such as IMC and IBI, for instance, exploit the bijective relation between correlations in terms of radial distribution functions (RDFs) and pair potentials,[22] using resolved (often all-atom) RDFs as a reference (input) for a match to RDFs that are one-to-one related to effective CG pair potentials (output). In practice, however, perfect identity of two RDFs with arbitrary precision cannot be achieved, and the matching procedure suffers from being ill-posed, meaning that tiny variations of the input RDFs can lead to a dramatic change in matched pair potentials.[23] An example of two systems with significantly different potentials, a 3D Lennard-Jones (LJ) liquid and a purely repulsive reference system, and very similar RDFs was published as early as 1983.[24] In principle, one can single out realistic CG potentials by considering additional properties such as pressure, density, or other thermodynamic properties in the matching procedure. These properties can either be taken from the reference calculations or from experiments. Indeed, several CG descriptions adopt the latter option, including Martini CGMD, which in addition derives the strength of the non-bonded LJ interactions from mixing enthalpies rather than from systematic mapping. A more practical issue is the increasingly prohibitive computational effort required to extract CG potentials by such systematic mapping routines, particularly for more complex systems that entail large numbers of CG particles.

The second issue, i.e., limitations to phase space sampling, is a challenge that is shared by many computational techniques. For instance, also popular approaches in the field of machine learning are at loss when forced to interpolate in parts of the phase space that are insufficiently represented in training sets. A particular caveat in systematic coarse graining is that even long simulations at the coarser level do not always sample conformations for which the considered mapping could turn out to be inaccurate, e.g., big systems that are simulated starting from a reference configuration at a temperature that is far too low to allow the system to jump to a different energy basin or harmonic restraints are applied in simulations, meaning that the performance of a CG description is not always put to a real test. An example of such a situation is a recent protein study that employed a number of state-of-the-art force fields from the CHARMM and Amber families for massive MD simulation, illustrating that none of them can reproduce the accurate dimensions and residual secondary structure propensities for both disordered and folded proteins.[25] The possible hereditary nature of this effect should be of particular concern for systematic coarsening approaches that aim at significantly enhancing the sampling of molecular conformations by subsequent coarsening steps. It, thus, makes sense to think about enhancing the sampling capabilities, for instance, by rare-event sampling methods, on all levels of resolution.

Consequently, the quality of a computational method for a particular application should be discussed in terms of general concerns and pitfalls that are associated with averaging over more resolved degrees of freedom:

- As discussed, the validity of a set of empirical parameters for a CG description may be restricted by particular sampling limitations in the reference model. Moreover, while CG potentials are often being determined as potentials of mean force, so that they contain both the enthalpic and entropic contributions,[26] matching procedures are also restricted to a specific thermodynamic state in terms of temperature, pressure, volume, etc. Employing such a description more generally assumes transferability across different thermodynamic conditions, which is not always a valid assumption. In most cases where the CG methodology is first applied, parameter tuning is a true need.
- The approximation that is at the basis of the averaging procedure may not be valid for a system at hand, which leads to an issue of representability. For MD, this is the case for systems in electronically excited states where the energy separation between different electronic states becomes small (e.g., during photochemical events), systems in strong laser fields, in which electronic and nuclear degrees of freedom are evolving on similar characteristic time scales, or Jahn–Teller systems, in which electronic and nuclear degrees of freedom are strongly coupled. As coarsening gradually removes more and more chemical details, it eventually leads to a behavior that is determined by hydrophobicity/hydrophilicity and molecular packing. Subtle interactions such as coordination, π–π stacking interactions, and hydrogen bonding are usually even beyond resolution in mildly coarsened CGMD like Martini. However, the recently developed theory of scale-consistent effective energy terms, which include in an indirect manner the averaged out atomic details, enables these effects to be included to a large extent,[27] particularly for systems for which they are crucial for correct modeling of the structural features such as proteins.
- Removing degrees of freedom also affects the kinetics since any loss of degrees of freedom changes the distribution of thermal energy over the remaining degrees of freedom in a coarser description (following equipartition). For instance, Martini CGMD is known to accelerate kinetic processes by a factor of 4 compared to the reference classical MD, as measured from diffusion rates, despite the fact that the non-bonded interactions are of equal LJ type. Further coarsening has the effect of softening, meaning that caging may become less significant, which accelerates transport and changes the short-time kinetics in a way that cannot always be captured in a single scaling factor. At longer times, however, the majority of CG descriptions for kinetics have been shown

to be equivalent to well-known continuum descriptions. As a result, developing more accurate kinetic descriptions that are based on a separation of relevant time scales is a long-term research direction. Another challenge is to represent dynamic variations in composition. As a simple example, classical MD is unable to capture any process that is associated with a system-induced change in the electronic states.

IV. CURRENT STATUS

The variety of computational approaches that is currently available for the *in silico* study of (bio)material properties is both a luxury and a burden. On the one hand, computational and application scientists interested in using tools for quantum chemical calculations can benefit from a broad range of special purpose engines that have appeared over the last 30 years.[28] At the MD side, a small number of versatile, efficient, and well-documented open source engines such as CHARMM, NAMD, GROMACS, AMBER, GENESIS, LAMMPS, Desmond, and OpenMM are supported by considerable user communities and updated as well as extended by well-structured groups of dedicated developers. These implementations are generally easy to use for non-experts, well tested and documented, often come with online support, are flexible in terms of I/O [including graphical interfaces and built-in data visualization or external packages for visualization such as visual molecular dynamics (VMD) support their output formats], and optimized for several computer architectures including graphical processing units (GPUs). In turn, the size of these user community secures the funding (academic) or revenues (proprietary, commercial) that are required for engine maintenance and porting, as well as for proper validation, further development (e.g., basis sets or force fields), support for porting to new hardware such as GPUs, and implementation in a professional, standardized fashion.

Some of the MD engines mentioned above support more established methods for CG and MS simulations, e.g., LAMMPS and GROMACS. Yet, most scientific groups that develop, implement, validate, and apply CG and MS approaches face different conditions: they generally lack the resources for the development of easy-to-use versatile code for various platforms as well as the user base that is required for thorough validation. This situation reflects both the novelty of the field and the diversity of the approaches that are currently proposed and tested, which results in a situation where newcomers with an interest in particular techniques have to rely on the information provided in scientific publications that generally neither provide sufficient insight to judge their applicability for other systems nor contain all the information needed to get started. While some CG and/or MS simulation tools have been provided as open source engines—a successful example is Martini CGMD, which benefits from the fact that it developed a CG extension to FFs and builds on the highly efficient integrators implemented in GROMACS—their popularity is still limited compared to the classical MD. This undesired situation is due to several challenges that are discussed in more detail later, the most prominent being the required knowledge of averaging concepts and the lack of prerequisites for straightforward mapping. An illustration of this trade-off between ease-of-use and consistency is the observation that the standard CG FF in Martini frequently needs expert tuning to resolve inaccuracies even at a qualitative level.

V. FUTURE DIRECTIONS

The assessment of promising research directions is likely as diverse as the methods that have or are being developed within this community. One future challenge in modeling biological systems, and thus far only touched upon by few, is the explicit representation of electronic degrees of freedom within a computational description that is capable of capturing the length and time scales that relate to experimental observables at the emergent level, for instance, in light-activated or reactive processes in nature. On a conceptual level, five issues appear critical in the further development of current coarse-grained approaches: (1) theory, (2) parameterization, (3) validation, (4) integration of sparse experimental and bioinformatics data, and (5) use of the coarse-graining approach to understand the origin of the architecture, dynamics, and behavior of complex systems and not just as a tool for speeding up/extending the scale of simulations. Each of these issues is listed as follows:

1. **Theory development**. Coarse graining means averaging over secondary degrees of freedom, which immediately links it to the potential of mean force (PMF). However, the PMF as such is both non-transferable and often prohibitively expensive to evaluate. Therefore, splitting it into transferable force-field terms (analogous to all-atom force-field terms) is crucial. These terms are usually imported from all-atom force fields, e.g., in Martini, which results in usually poor performance because it ignores orientational dependence, which is crucial if extensive or "aggressive" coarse graining is to be performed in which multibody terms, which are crucial for the reproduction of regular structures, are ill represented. Hybrid CG approaches such as Single-Chain in Mean-Field (SCMF)[29] and MD-Self Consistent Field (MD-SCF),[30] which link discrete (particle-based) and continuum (field-based) descriptions in a single simulation volume, hold the promise to at least resolve some of these issues and are increasingly applied and validated. Two particle-based approaches have also addressed this problem: the Multi-Scale Coarse-Grained (MSCG) approach developed by the Voth group[31–34] and factorization of the PMF into Kubo's cluster cumulant functions developed by the Liwo and Scheraga groups.[27,35,36] The first approach is originally a model-free approach, but it assumes isotropic/spherical interactions. Therefore, it does not produce transferable force fields and also does not allow extensive coarse graining. In the PMF-factorization approach, non-radial symmetry is allowed, multibody terms emerge naturally, and the details of all-atom geometry, albeit not present in the coarse-grained model, are embedded in the effective potentials, resulting in appropriate dependence of these potentials on the geometry of the reduced model. It should be noted that including anisotropic/multibody terms in a force field requires additional parameterization efforts and increases the costs per evaluation of a single interaction term compared to pairwise spherical ones. Yet, owing to a large reduction of the number of interaction sites upon extensive coarse graining, a large net gain can be obtained both in terms of resources and

wall-clock time, which would not be possible when representations are restricted to pairwise radial interactions. More generally, while many assume that careful derivation of appropriate formulas does not make sense now that machine learning methods are advancing, careful linking of the effective energy surfaces to all-atom surfaces is important if extensive coarse graining is involved, especially for understanding how the simplified effective energy surfaces are linked to the parent all-atom energy surfaces, and to estimate the inaccuracy of coarse-grained force fields. In this respect, issues that remain open are (i) objective definition of coarse-grained sites, which is especially important in ultra-coarse-graining, and (ii) extension of scale-consistent derivations.

2. **Parameterization**. Arguably, the best parameterization practice in the coarse-graining community (and beyond) is a combination of bottom-up and top-down. This way of force-field parameterization combines systematic mapping of all-atom interactions onto the effective interactions between coarse-grained beads while reproducing ensemble properties. For the systematic part, PMFs of subsystems representing the effective energy terms can be calculated by numerical integration or rare-event sampling MD or can even be taken from statistical data. Since this procedure is usually rather straightforward, developments are likely to be directed toward more effective sampling methods, better error estimation, and the development of schemes as to where and how to include statistical potentials. The second, top-down part, which is also called force-field calibration, aims at reproducing thermodynamic and structural properties, as well as structures themselves. For the reasons to be given shortly, caution should be exercised when using structures or ensembles of structures solved by experimental techniques (NMR, x-ray crystallography, etc.) because they implicitly contain a bias that stems from data processing. For this reason, raw data, including low-resolution data from SAXS/SANS, XLMS, etc., should be considered instead of resolved structures. Force-field calibration, especially important while designing multi-scale simulation methodologies, can be carried out via force matching, in which coarse-grained forces are matched to all-atom forces that are mapped onto the coarse-grained representation, with the aim of achieving compatibility. While established force-matching procedures have to date been designed for CG models with spherical sites, an extension to the non-spherical case is underway.[37]
3. **Validation**. Validation should be directed to testing the reproduction of thermodynamic/ensemble-averaged properties and structural properties, as well as non-equilibrium properties such as the diffusion coefficients, rate constants, etc. At present, these two validation schemes are usually treated independently of each other. Clearly, developing an integrated measure will be worthwhile.
4. **Steered or assisted: integration with experimental data**. The increasing availability of low-resolution experimental data, which can be collected at a relatively low expense, opens up a promising direction for CG approaches as a predictive tool, especially in the area where CG descriptions are sufficiently accurate. A prominent example in proteomics is the NMR signal assignment for large proteins, where CG approaches can be employed, provided that appropriately designed target functions/selection procedures are developed. Here, the challenge is the translation of NMR-resolved separations to site–site distances. In particular, resolving the limitation of knowledge-based methods in bioinformatics, such as homology modeling, by combining them with physics-based modeling, preferably at the coarse-grained level, is a promising avenue for biomolecular simulations. Among others, such a merger will provide dynamics/ensemble averages that are not accessible by conventional bioinformatics tools.
5. **Understanding the architecture of biomolecules/self-assembling systems**. Multibody terms in the coarse-grained force field, which can be systematically derived from a factor-expansion approach, can already be used to rationalize the formation of α-helical and β-sheet domains in proteins, i.e., from an interplay between the local and backbone-electrostatic interactions,[27] and for the formation of a double-stranded DNA helix as a result of average electrostatic interaction between nucleic-acid bases.[38] A promising direction is to extend this approach to the formation of other types of structures (lipids, liquid crystals, etc.) and the design of polymers with the desired conformations, without the need of expensive simulations.

VI. DISCUSSION OUTCOMES

For the Lorentz workshop, which took place in the summer of 2019, some 35 specialists from QM, MD, CGMD, and more (general and dedicated) coarse-grained methodologies including ML were invited. It included outbreak discussion sessions in smaller groups concentrating on the following four themes:

A. Setting up the framework for a round-robin test for CG methodologies/codes.
B. How does the CG community cope with standardized tests and repositories of simulation results? Is a standard format for simulation input and output data needed and, if so, what form should it take?
C. How do we cope with massive data?
D. Standardized core engines.

In the following, we further narrow down these themes and review the outcome of the discussions.

A. Setting up the framework for a round-robin test for CG methodologies/codes

An open competition such as CASP (Critical Assessment of protein Structure Prediction)[39] could provide an opportunity to objectively test methodologies for structure prediction and, equally important, deliver an independent assessment of the state of the art in protein structure modeling to the research community and software users. Setting up a framework for testing performance of CG approaches would provide a platform with similar benefits. The desire for round-robin tests has been formulated before within this community (see, for instance, https://sites.google.com/view/emmc-uppsala-june17/program) but has yet to be implemented and would

provide multiple benefits. First, it would actively engage this community, appeals to students and young researchers, is efficient in providing a snapshot and independent assessment of state-of-the-art CG methodology, and, hence, promotes this CG field in a much broader context. Second, once operable, such a framework can be employed to attract and directly involve other stakeholders—Nvidia/Intel/Amd, compiler and simulation engine developers, and HPC centers—and offer support to computational groups with less resources.

Arguments

The idea of developing standardized tests was generally acknowledged as very useful and highly desired. Yet, it was mentioned that some sort of rewarding system must be planned in order to make this round-robin test sustainable. Such tests would serve a dual goal: (1) verification: check the implementation of new methods in simulation engines and (2) validation: clarify the performance of a method and the map used for coarse graining for its ability to reproduce known physical quantities.

Quality control of the implementation (verification) is of general concern. Whereas code development for QM/MD engines takes place in a controlled and professional setting, with a small set of dedicated scientists generating and testing core engines for a large community of users that additionally act as a testing bed for new functionalities, new CG developments are often implemented in historic development codes that are only used by developer groups for dedicated purposes. Although these groups often perform basic testing of the consistency of implementation, the community could play an active role in (i) formulating recommendations for more standardized code development, which would have the advantage of imposing a rational coding structure that would ease testing, transferability, and exchange/reuse of core parts and bring more flexibility for interfacing different codes, and (ii) developing metrics for quantifying efficiency of one type of code on different platforms.

Quantifying the performance of a method to produce realistic data (validation) is very much desired, but in practice, it is complicated both by the diversity of methods that are being developed and the various degrees of coarsening considered in these methods. This brings along the question how exact the (re)production of older data and/or data obtained by models at different resolutions should be to be acceptable and which measure should be considered to quantify agreement. In this context, it should be noted that even (exact) reproduction of simulation data by the same method/code can be problematic due to the randomness introduced by stochastic processes and computational conditions such as the setup for parallelization, which may complicate the development and implementation of automated consistency checks. For new concepts, the focus should, therefore, be on the reproduction of well-chosen thermodynamically averaged (or measurable) properties, structures, or particular mechanisms in terms of sets of consecutive states or sequences of events. One may think of partitioning coefficients, liquid state properties, and phase or conformational transitions. The initial reference values can either be obtained from first principle calculations or directly from experiments, keeping in mind that the latter can also be subject to uncertainties that come with model fits and/or indirect determination. A well-known example is the substantial variation of the area per lipids reported in the experimental literature over time.[40] Moreover, as systems are often out of equilibrium, the challenging ability to realistically account for dynamics would become a serious bottleneck. To avoid such uncertainties, which even play a role in the reference x-ray protein structures provided in the CASP competition, testing the algorithmic performance of methods that deliver the same properties should involve a comparison of directly measured structure-related properties such as distance boundaries [or, best, Nuclear Overhauser Effect (NOE) intensities] and chemical shifts from NMR measurements, distance distributions from SAXS/SANS, etc. One more advantage of using these quantities instead of solved structures is that they are ensemble averages and are, consequently, more related to the results of simulations. Free-energy landscapes (FELs), which depend on few order parameters, can also be used for the straightforward comparison of CG force fields/approaches. However, it should be noted that the FELs customarily expressed in just a few order parameters often provide only a superficial insight into the actual highly multidimensional FELs. On the other hand, the use of the principal-component analysis provides a few primary order parameters to adequately capture the FELs.

We end with a note and a warning. First, we note that there is an emerging paradigm for assessing quality of simulations and simulation results, based on verification (V), validation (V), and uncertainty quantification (UQ), namely, VVUQ; see, e.g., https://www.vecma.eu/vecmatk/. It would be worthwhile to adopt such a paradigm more systematically in the community of computational materials research. Moreover, when such a framework is installed, it should be handled with care. Sometimes, even the very modest goal of reproducing trends may be an utterly viable research objective when all existing approaches are at loss, for instance, to tackle the complicated problem of evaluating interactions at bio-interfaces, as a first and necessary step toward more quantitative descriptions. A prominent example of such a target is the evaluation of interactions between large proteins and nanoparticles in the context of nanotoxicity. In such circumstances, ranking methods purely based on quantitative testing would work counterproductive and at the end be unfair to the part of the community that develops method for new applications. Yet, in general, establishing measures for the performance of CG approaches can be very useful because they are objective, at least in theory. Given these subtleties, we advise that such an assessment procedure should be flexible and preferably not be based on a single measure of performance.

B. How does the CG community cope with standardized tests and repositories of simulation results? Can we come to a standard format for simulation input and output data and what form should it take?

Despite the fact that CG methodologies are diverse, individual groups of developers would benefit from the definition of standard tests that can be executed to judge whether a code or subroutine performs properly and efficiently. Moreover, the availability of simulation results (input and output, performance, etc.) in a standard format like the one proposed in the MODA framework of the

EMMC (https://emmc.info/moda/) and the agreement CWA 17284, see https://www.cen.eu/news/workshops/Pages/WS-2017-012.aspx, is important for reproduction, validation/comparison, and use as an input for other simulation methodologies and would add a substantial value to computational results. Except for some large computational groups, the current situation is far from optimal (little uniform I/O and data sharing via repositories) and often even does not facilitate proper comparison of performance despite the fact that many new concepts are developed toward improved computational efficiency.

Arguments

In recent years, standards for better exploiting the wealth of data and computational methods produced within the community, for both academic and commercial purposes, have been discussed at several occasions. They are usually formulated hand-in-hand with an ontology or formal description of key concepts and their relations within a particular scientific domain, which defines the organization of knowledge and is particularly useful for setting up a common vocabulary for researchers who need to share information about this domain; see https://emmc.info/emmo-info/. Examples of ontology-based standards are the MODA templates for MS simulation work flows developed by the European Molecular Modeling Council (EMMC) and the FAIR (Findable, Accessible, Interoperable, and Re-usable) principle that was developed for data-intensive science and will be adopted by the EU in the near future. Archives for computational materials science, for instance, at https://www.materialscloud.org/ by MARVEL, developed at EPFL by Nicola Marzari, and the one developed by the Novel Materials Discovery (NoMaD) Laboratory, see https://www.nomad-coe.eu/the-project/centre-of-excellence, collect and store results in a code-independent format for the purpose of mining and data-driven analysis. Currently, they use JavaScript Object Notation (JSON), a language-independent human-readable data format, and Hierarchical Data Format 5 (HDF5), a binary format for the efficient storage of large arrays and high-dimensional objects. Most, if not all, of the data stored in the latter three formats are generated by the electronic structure or MD codes. In MS and hierarchical CG treatments, information such as mapping details, trajectories (or portions thereof), and many other diverse objects are essential even for the purpose of reproduction. The community is in great need for decisions on data standards and a framework for storing, parsing, and retrieving such heterogeneous data.

It is clear that these needs can be best addressed by the CG/MS community itself to tackle the potentially unsustainable situation of a heterogeneous zoo of formats, methods, potentials/interactions, tests, and paradigms/ontologies. In particular, a solution should address this challenge on a more fundamental level than via the usual parsers and converters, for instance, as available in Babel (http://openbabel.org/wiki/Main_Page). The most efficient approach is clearly to emulate data storage concepts and/or formats proposed in related initiatives such as the NoMaD Laboratory and the Mosaic data model, which employs both a XML and a HDF5-based H5MD format. However, before such a *technical* solution can be considered, the community has to decide on a common ontology and one or a few data standards that cover the current and anticipated diversity of the CG/MS data. Another important decision to be made is a more general issue in setting up a data management plan, which is now mandatory for all EU funding. What part of the generated data can and should be stored? Should one only store data of production runs or also data generated during preparation, for instance, the result of parameter sweeps and data used to generate a mapping? And in which form? Storing raw data is a serious option, for instance, using storage solutions such as Zenodo, see https://zenodo.org/, but it makes sense to evaluate projected storage requirements for discussed formats beforehand to make sure that the massive output of many large scale CG/MS projects does not give rise to serious storage issues right from the start. Besides the community, which should propose, set up, and advocate this repository, and pursue a status similar to the Protein Data Base (PDB), journals and funding bodies can play an active role in encouraging public storage of data that is needed for the reproduction of published computational results. Unfortunately, in the current situation, all too often published secondary data (mapping details, simulation snapshots, and averaged properties) cannot be regenerated based on published information alone, which challenges the value of this work. Introducing identifiers in published work—a string that uniquely identifies repository data and additionally classifies the nature of the methodology with numerical classifiers in agreement with ontology keywords—will ease data retrieval and enable method sensitive searches. We note that such a system already exists for crystal structures (PDB) and ML, which often provide a DOI identifier to archived data. DOI identifiers are readily available for all-atom FFs via OpenKIM (https://openkim.org/), and one wonders if a similar approach could be adopted for mesoscopic (CG) FFs.

It is clear that themes A and B overlap: setting up a flexible framework for storing CG/MS data in a common format and providing testing sets for performance checks would imply options for public benchmarks. Also in the discussions during the workshop, which illustrated the diversity of objectives and current standards within the CG/MS community, the two topics were often mentioned in combination. Yet, despite this diversity, the development of a common framework was considered of utter importance for both developers and users. Being involved in generating reference datasets is another issue, which many consider as tedious work without many scientific benefits—an opinion that was particularly questioned by experts in machine learning (ML), who consider systematic data sweeps of great value. In practical terms, setting up a framework involves generating repositories for (i) storage of computational and experimental data in common data formats (theme B) and ontologies, (ii) analysis tools that enable robust and straightforward calculation of agreed measurables or observables, and (iii) reference sets for testing (theme A). It should be stressed again that particularly developing ontologies requires a tremendous effort with very little academic reward. With data already subject to data management plans, the main challenge appears to be centralization: to set up efficient and central databases that can deal with huge storage demands, are able to support several data formats, are easy to use and access, and up to date and maintained. Where most of the current method and software generation in computational soft matter research is a boundary condition or side product of a particular application study, the massive effort of building and maintaining such databases will not be easy despite their necessity. When decisions about the database layout have been

made, the first challenge will lie in securing funds for setting up the infrastructure, i.e., the design and filling of these databases, which would include collecting, archiving, and curating information that is already present in the community, particularly by groups involved in computational high-throughput screening. The second and related challenge is to find individuals that possess the expertise and show an interest in maintaining, popularizing, and updating this framework. The massive challenge when such a framework has been installed is to set up an implicit or explicit rewarding scheme for individuals to change their daily practice and adhere to such standards.

C. How do we cope with massive data?

Owing to massive parallelism, the systems that we can simulate through the electronic structure, atomistic, and CG simulation methodology have grown substantially. Most data standards developed decades ago have issues with handling such amounts of data, so how are visualization tools such as VMD going to cope? We face several additional methodological challenges, e.g., how can we efficiently treat long-ranged interactions like electrostatics in such huge systems? How to deal with such big systems/data and how one can open it up for the purpose of data mining?

Arguments

The general attitude during the workshop is that the generation of excessive data is not very useful and should be avoided. This can be promoted by (i) putting more effort in formulating a good hypothesis prior to carrying out numerical experiments, (ii) performing more analysis on the fly, and (iii) storing only raw data that are needed for post-processing. In particular, massive storage of data during simulations may even become prohibitive. When educating students and young researchers in computational research, it would be useful to put more emphasis on good research practices, as well as the basics of statistics and data science. Overall, with the continuing increase in computer power, it is questionable whether data storage is a necessity in all cases, especially when rerunning costs less time than data retrieval and interpretation. Publishing input files for versioned and benchmarked codes with back-functionality is all that is needed for reproducing such kind of data, which would seriously limit the required storage space. This is actually the approach taken by MaterialsCloud (https://www.materialscloud.org/) based on AiiDA (http://www.aiida.net/). Github (https://github.com/) offers solutions for simulation engine storage and versioning and could be more promoted as good science practice within the community. Yet, in some cases, storage of complete trajectories may be a requirement for (future) post-processing, and it should be left to the developer to make that judgment. A specification of what kind of (raw) data is required and how the ML community can better benefit from data produced by computational groups would be useful. It should be noted that since ML is not based on a physical model, big data should not replace traditional analyses.

D. Standardized core engines

Many of the current simulation engines are based on the hard-coded functionality in one of the standard computer languages (Fortran, C, C++, and Cuda). Although this is understandable from the viewpoint of computational efficiency, it requires a considerable coding effort to keep up with language evolution/versions (with the risk for engines of becoming obsolete) and meet the current shift to massive parallelism on heterogeneous platforms, e.g., mixtures of CPU/GPU. A modern way of dealing with this challenge is to separate tasks into mathematical operations and physics and reformulate the engine as a C/C++ core (covering the most compute-intensive operations, which could be open source) and physical procedures written in scripting languages such as Python or Tcl. Many of the core numerical routines are available as open source and accessible by scripts. How should one respond to these developments? One option is to refactor CG codes as scripted "plugins" on top of shared (and possibly open source) libraries, a different take on *e pluribus unum*—where diversity remains, is celebrated and is sustained through recognizing commonalities underlying them. What should these common libraries include and how can we conserve sufficient flexibility? How does one take care of ownership, dependencies, and copyright issues and which business model would guarantee CG groups and method developers continuity in such a setup?[41] The advantages of standardized software design are clear, including the benefit for smaller development groups, which is the reason why bringing up such issues cannot be avoided.

Arguments

The creation of core engines is a challenging task and not considered vital by the whole community. First of all, what functionality should such a core engine support? Common libraries already exist for General Purpose Computation on Graphics Processing Units (GPGPUs), but they do not support all functionalities, e.g., Verlet lists for inhomogeneous densities. Also for numerically intensive parts, where standardized common libraries are a true advantage, several open source solutions already exist (FFTW, BLAS/LAPACK). An immediate challenging issue is the intellectual property of such core developments. Although DOIs for modules would allow recognition of the efforts by developers and would enable keeping statistics of module/core usage, such developments would require consensus as well as a projected critical user base for developers of current engines to work on creating common libraries.

Particularly in the QM/MD domain, the current availability of highly optimized and well supported simulation engines does not immediately call for the development of new core engines, although such a core engine OpenMM (http://openmm.org/) was recently launched. Moreover, there are recent solutions for situations where an efficient coupling of existing quantum and MD codes is desired. The novel framework for Multiscale Modeling in Computational Chemistry (MiMiC[42]) enables fast data exchange between programs, through the use of Message Passing Interface (MPI) intercommunicators, based on a multiple-program multiple-data (MPMD) model with loosely coupled programs. It exploits existing parallelization strategies used by the coupled programs while maintaining a high degree of flexibility. PLUMED (https://www.plumed.org/) is another open source, community-developed library that provides tools for enhanced sampling/metadynamics simulations and can work with a wide variety of software for *ab initio*, atomistic, and coarse-grained simulations.

In the case that one chooses to support different implementations and/or functionalities by different core engines at the CG/MS level, for instance, to enable less common potentials and representations, interfacing becomes an issue. With Python being actively developed by several communities, standards and testing as well as porting between different platforms and back-functionality are becoming a serious issue. As a result, such efforts are increasingly at risk of running into conflicts. Moreover, interpreters such as Python would be much less efficient if the issue of a shared data format is not resolved. Arguably, a common interface protocol in C would be much more useful than a Python-based one and would allow a Python interface if required. The underlying data format of C++ could make it less suited for interfacing. The significant effort required for combining code interfacing and massive data treatment with educational potential calls for a broader perspective, either in the context of EU DC Connect or a Joint Research Center.

Finally, this issue can also be seen from an educational perspective. The National Science Foundation (NSF) funded http://nanoHUB.org for computational nanotechnology research, education, and collaboration. The site hosts a rapidly growing collection of simulation tools with typically an Application-Programming-Interface (API) for nanoscale phenomena that run in the cloud and are accessible through a web browser, as well as online presentations, short courses, animations, and teaching materials. This may suggest the potential benefits of including educational features such as good APIs when creating libraries of modules. A quick start in this respect might be the compilation of a (online) book of numerical modeling recipes, in analogy to numerical recipes, by the community.

VII. RECOMMENDATIONS

Coarse-grained computational modeling has reached a certain degree of maturity during the last 20 years. Yet, compared to more established QM and MM models and their community, the CG community is relatively small and unorganized and faces a number of challenges that have to be addressed in order to harvest the full potential of CG and MS methodologies in terms of breakthrough applications and funding. These challenges are both present in an academic setting, where day-to-day issues range from data exchange, conversion and storage, validation and reproducibility, and the availability of tested and efficient simulation and parameterization tools (SPTs) to a lack of key words related to CG modeling in funding schemes, but particularly also to the industrial end-users, who suffer difficulties in benefitting from these advances due to unfamiliarity with the underlying concepts, difficulties in the extraction and interpretation of available CG data, and lack of commercially available SPT. To tackle these challenges, the CG/MS communities could be well served by taking a more systematic advantage of the (lessons learned in the) MD community. One obvious link is in the choice between Python/C/C++ for interfacing, knowing that Python is becoming a preferred choice in the MD community. The lesson of PLUMED, see discussion on theme D, might be one to emulate for CG/MS, i.e., to aim at providing a large variety of CG methods in a library that runs on multiple engines. Such a strategy may at the same time facilitate improved performance on large scale machines for CG simulations and of parameterization efforts, possibly also involving ML techniques, by a better exploitation of massive parallelism.

In this perspective communication, we have made an effort to review the state of affairs and formulate a number of suggestions to address or even solve the current challenges. Ranking these suggestions based on their importance to the community and the likelihood that they can be implemented in practice would be useful, but one can argue that mutual dependencies necessitate implementing or at least considering the full set. Whatever your personal viewpoint may be, they serve a valuable goal in stimulating this timely discussion and/or defining a starting point for discussion in the broader community. In particular, we make the following recommendations (for details, see the preceding sections):

- Develop and adopt an ontology of CG models and workflows, taking the existing EMMC/EMMO framework as a starting point. Consequently, make this standard available to the community as a useful tool for documenting computational results in scientific publications.
- Set up an identification system for heterogeneous simulation data to ease data extraction.
- Select one flexible data format, for instance, the H5MD format, and define a rewarding system that stimulates the common use of this format.
- Define and adopt a framework for assessing quality, based on verification, validation, and uncertainty quantification. Validation should concern thermodynamic, dynamics, kinetics, average structure-dependent properties, or structures, with an emphasis on the kinds of properties that are intended to be reproduced. Several measures are needed to cover this heterogeneous modeling domain.
- Invest in better validation and education. The proof that the CG/MS methodology can provide (at least qualitatively) relevant results can only be readily given by experts that are actively involved in the development of such methodologies. Making this investment will also have an important educational effect. In combination with easing the access to state-of-the-art CG methodology and data, it will generate a larger user group in academics and industry and strengthen the position of the modeling community as a whole.
- Define general rules for data storage, keeping in mind that it may sometimes be more efficient to re-simulate data if input parameters files are provided, either through publication or a database, and as long as the versioned, benchmarked simulation engines with back-functionality are freely available. A system of rules can also be exploited for improving and easing data management plans.
- Set up and maintain databases for massive storage of heterogeneous simulation data. As the necessary manpower and, thus, funding will rely on proving the huge benefit of such a database, all stakeholders should be involved.
- Introduce a DOI identifier for code development.

ACKNOWLEDGMENTS

The authors are grateful to the Lorentz Center for hosting the *Summer school and Workshop on Multi-scale Modeling* and for

providing a very stimulating environment for discussion. Moreover, we are very grateful to all participants of these two events for the input that they have provided. Without it, the perspective communication in the current form would not have existed. Furthermore, the work of D.M. is supported by the European Union E-CAM center of excellence under Grant No. 676531. I.P. acknowledges support from Ministerio de Ciencia, Innovación y Universidades/FEDER UE (Grant No. PGC2018-098373-B-100), from Generalitat de Catalunya under project Grant No. 2017SGR-884, and from the Swiss National Science Foundation Project, Grant No. 200021-175719. J.A.L. acknowledges support from the National Science Center of Poland (Narodowe Centrum Nauki) via Grant Nos. UMO-2017/25/B/ST4/01026 and UMO-2017/26/M/ST4/00044, and P.A. acknowledges the support from the Italian National Project PRIN *Heat transfer and Thermal Energy Storage Enhancement by Foams and Nanoparticles* (Grant No. 2017F7KZWS).

DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

REFERENCES

[1]A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme," J. Mol. Biol. **103**, 227–249 (1976).
[2]M. H. M. Olsson, W. W. Parson, and A. Warshel, "Dynamical contributions to enzyme catalysis: Critical tests of a popular hypothesis," Chem. Rev. **106**, 1737–1756 (2006).
[3]H. Wang, C. Schütte, and L. Delle Site, "Adaptive resolution simulation (AdResS): A smooth thermodynamic and structural transition from atomistic to coarse-grained resolution and vice versa in a grand canonical fashion," J. Chem. Theory Comput. **8**, 2878–2887 (2012).
[4]H. Wang, C. Hartmann, C. Schutte, and L. Delle Site, "Grand-canonical-like molecular-dynamics simulations by using an adaptive-resolution technique," Phys. Rev. X **3**, 011018 (2013).
[5]L. Delle Site and M. Prapotnik, "Molecular systems with open boundaries: Theory and simulation," Phys. Rep. **693**, 1–56 (2017).
[6]C. Krekeler, A. Agarwal, C. Junghans, M. Praprotnik, and L. Delle Site, "Adaptive resolution molecular dynamics technique: Down to the essential," J. Chem. Phys. **149**, 024104 (2018).
[7]M. Vassaux, R. C. Sinclair, R. A. Richardson, J. L. Suter, and P. V. Coveney, "Toward high fidelity materials property prediction from multiscale modeling and simulation," Adv. Theory Simul. **3**, 1900122 (2020).
[8]K. Matou, M. G. D. Geers, V. G. Kouznetsova, and A. Gillman, "A review of predictive nonlinear theories for multiscale modeling of heterogeneous materials," J. Comput. Phys. **330**, 192–220 (2017).
[9]M. G. Guenza, M. Dinpajooh, J. McCarty, and I. Y. Lyubimov, "Accuracy, transferability, and efficiency of coarse-grained models of molecular liquids," J. Phys. Chem. B **122**, 10257–10278 (2018).
[10]B. Cheng, E. A. Engel, J. Behler, C. Dellago, and M. Ceriotti, "Ab initio thermodynamics of liquid and solid water," Proc. Natl. Acad. Sci. U. S. A. **116**(4), 1110–1115 (2019).
[11]J. Behler, "First principles neural network potentials for reactive simulations of large molecular and condensed systems," Angew. Chem., Int. Ed. **56**, 12828–12840 (2017).
[12]A. R. Leach, *Molecular Modelling: Principles and Applications* (Prentice Hall, Harlow, England, 2001).
[13]A. P. Lyubartsev and A. Laaksonen, "Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach," Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top. **52**, 3730–3737 (1995).
[14]D. Reith, M. Pütz, and F. Müller-Plathe, "Deriving effective mesoscale potentials from atomistic simulations," J. Comput. Chem. **24**, 1624–1636 (2003).
[15]S. Izvekov and G. A. Voth, "A multiscale coarse-graining method for biomolecular systems," J. Phys. Chem. B **109**, 2469–2473 (2005).
[16]W. G. Noid, J. W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen, "The multiscale coarse-graining method. I. A rigorous bridge between atomistic and coarse-grained models," J. Chem. Phys. **128**, 244114 (2008).
[17]W. G. Noid, P. Liu, Y. Wang, J. W. Chu, G. S. Ayton, S. Izvekov, H. C. Andersen, and G. A. Voth, "The multiscale coarse-graining method. II. Numerical implementation for coarse-grained molecular models," J. Chem. Phys. **128**, 244115 (2008).
[18]J.-W. Chu, S. Izveko, and G. A. Voth, "The multiscale challenge for biomolecular systems: Coarse-grained modeling," Mol. Simul. **32**, 211–218 (2006).
[19]A. Chaimovich and M. S. Shell, "Coarse-graining errors and numerical optimization using a relative entropy framework," J. Chem. Phys. **134**, 094112 (2011).
[20]E. Brini, C. R. Herbers, G. Deichmann, and N. F. A. van der Vegt, "Thermodynamic transferability of coarse-grained potentials for polymer-additive systems," Phys. Chem. Chem. Phys. **14**, 11896–11903 (2012).
[21]E. Brini and N. F. A. van der Vegt, "Chemically transferable coarse-grained potentials from conditional reversible work calculations," J. Chem. Phys. **137**, 154113 (2012).
[22]R. L. Henderson, "A uniqueness theorem for fluid pair correlation functions," Phys. Lett. A **49**, 197 (1974).
[23]R. Potestio, "Is Hendersons theorem practically useful?," JUnQ **3**, 13–15 (2013).
[24]J. D. Weeks and J. Q. Broughton, "van der Waals theory of melting in two and three dimensions," J. Chem. Phys. **78**, 4197–4205 (1983).
[25]P. Robustelli, S. Piana, and D. E. Shaw, "Developing a molecular dynamics force field for both folded and disordered protein states," Proc. Natl. Acad. Sci. U. S. A. **115**, E4758–E4766 (2018).
[26]L. Lu and G. A. Voth, "The multiscale coarse-graining method. VII. Free energy decomposition of coarse-grained effective potentials," J. Chem. Phys. **134**, 224107 (2011).
[27]A. K. Sieradzan, M. Makowski, A. Augustynowicz, and A. Liwo, "A general method for the derivation of the functional forms of the effective energy terms in coarse-grained energy functions of polymers. I. Backbone potentials of coarse-grained polypeptide chains," J. Chem. Phys. **146**, 124106 (2017).
[28]See https://en.wikipedia.org/wiki/List_of_quantum_chemistry_and_solid-state_physics_software for an extensive list of both open-source and commercial tools for quantum chemical calculations.
[29]K. C. Daoulas and M. Müller, "Single chain in mean field simulations: Quasi-instantaneous field approximation and quantitative comparison with Monte Carlo simulations," J. Chem. Phys. **125**, 184904 (2006).
[30]G. Milano and T. Kawakatsu, "Hybrid particle-field molecular dynamics simulations for dense polymer systems," J. Chem. Phys. **130**, 214106 (2009).
[31]G. S. Ayton, W. G. Noid, and G. A. Voth, "Multiscale modeling of biomolecular systems: In serial and in parallel," Curr. Opin. Struct. Biol. **17**, 192–198 (2007).
[32]J. F. Dama, A. V. Sinitskiy, M. McCullagh, J. Weare, B. Roux, A. R. Dinner, and G. A. Voth, "The theory of ultra-coarse-graining. 1. General principles," J. Chem. Theory Comput. **9**, 2466–2480 (2013).
[33]M. G. Sauders and G. A. Voth, "Coarse-graining methods for computational biology," Annu. Rev. Biophys. **42**, 73–93 (2013).
[34]G. Voth, *Coarse-Graining of Condensed Phase and Biomolecular Systems*, 1st ed. (CRC Press, Taylor & Francis Group, 2008).
[35]A. Liwo, C. Czaplewski, J. Pillardy, and H. A. Scheraga, "Cumulant-based expressions for the multibody terms for the correlation between local and electrostatic interactions in the united-residue force field," J. Chem. Phys. **115**, 2323–2347 (2001).
[36]A. Liwo, A. K. Sieradzan, A. G. Lipska, C. Czaplewski, I. Joung, W. Żmudzińska, A. Hałabis, and S. Ołdziej, "A general method for the derivation of the functional forms of the effective energy terms in coarse-grained energy functions of polymers. III. Determination of scale-consistent backbone-local and correlation potentials in the UNRES force field and force-field calibration and validation," J. Chem. Phys. **150**, 155104 (2019).

[37] A. Liwo and C. Czaplewski, "Extension of the force-matching method to coarse-grained models with axially-symmetric sites to produce transferable force fields: Application to the UNRES model of proteins," J. Chem. Phys. **152**, 054902 (2020).
[38] Y. He, M. A. Mozolewska, P. Krupa, A. K. Sieradzan, T. K. Wirecki, A. Liwo, K. Kachlishvili, S. Rackovsky, D. Jagiela, R. Slusarz, C. R. Czaplewski, S. Oldziej, and H. A. Scheraga, "Lessons from application of the UNRES force field to predictions of structures of CASP10 targets," Proc. Natl. Acad. Sci. U. S. A. **110**, 14936–14941 (2013).
[39] See http://predictioncenter.org/ for an overview of CASP activities during the last 25 years.
[40] D. Poger and A. E. Mark, "On the validation of molecular dynamics simulations of saturated and *cis*-monounsaturated phosphatidylcholine lipid bilayers: A comparison with experiment," J. Chem. Theory Comput. **6**, 325–336 (2010).
[41] See https://emmc.info/wp-content/uploads/2018/09/EMMC-CSA-D5.2_M15_WEB-PUv3.pdf for a white paper about software development standards for method description, assumptions, accuracy and limitations, testing requirements, issue resolution, version control, user documentation and continuous support, and resolution of issues.
[42] J. M. H. Olsen, V. Bolnykh, S. Meloni, E. Ippoliti, M. P. Bircher, P. Carloni, and U. Rothlisberger, "MiMiC: A novel framework for multiscale modeling in computational chemistry," J. Chem. Theory Comput. **15**(6), 3810–3823 (2019).