Advanced Review

Challenges in large scale quantum mechanical calculations

Laura E. Ratcliff,[1] Stephan Mohr,[2] Georg Huhs,[2] Thierry Deutsch,[3,4] Michel Masella[5] and Luigi Genovese[3,4]*

During the past decades, quantum mechanical methods have undergone an amazing transition from pioneering investigations of experts into a wide range of practical applications, made by a vast community of researchers. First principles calculations of systems containing up to a few hundred atoms have become a standard in many branches of science. The sizes of the systems which can be simulated have increased even further during recent years, and quantum-mechanical calculations of systems up to many thousands of atoms are nowadays possible. This opens up new appealing possibilities, in particular for interdisciplinary work, bridging together communities of different needs and sensibilities. In this review we will present the current status of this topic, and will also give an outlook on the vast multitude of applications, challenges, and opportunities stimulated by electronic structure calculations, making this field an important working tool and bringing together researchers of many different domains.

How to cite this article:
WIREs Comput Mol Sci 2017, 7:e1290. doi: 10.1002/wcms.1290

INTRODUCTION

The fundamental laws for a quantum mechanical (QM) description of atomistic systems up to the nanoscale are known and have been well established for a little less than a century. Yet, there are many challenges related to the QM treatment of large systems. In the vast majority of cases, we are still unable to solve the fundamental Schrödinger equation for systems of realistic sizes in such a way that the results satisfy 'universal' requirements of accuracy, precision, and especially predictability. Unfortunately, this also implies that we are still far from being able to quantitatively predict experimental results at the nanoscale.

The problems are not only related to the computational complexity needed to solve the equations of QM, there are also intrinsic obstacles. To give an example, let us remind the so-called 'Coulson's challenge.' In 1960, Coulson[1] noticed that the most compact object needed to characterize quantum mechanically an N-electron system (at least in its ground state) is the two-body reduced density matrix (2RDM). However, it turns out that *we do not know* all the necessary conditions for the 2RDM to be *N-representable*, i.e., coming from an anti-symmetric wavefunction of an N-electron system. Thus, even if a compact (and, in principle, computationally accessible) object exists, theoretical and algorithmic bottlenecks hinder its practical usage. In 1964[2] and 1965,[3] Kohn, Hohenberg, and Sham further reduced the complexity by showing that the electronic density is in a one-to-one correspondence with the ground state energy of a system of interacting electrons, and that such an interacting system can be replaced by a mean-field problem of N noninteracting fermions that provide the same distribution of the density. These are the fundamental ideas of Density Functional Theory (DFT).

*Correspondence to: luigi.genovese@cea.fr

[1]Argonne Leadership Computing Facility, Argonne National Laboratory, Lemon, IL, USA

[2]Department of Computer Applications in Science and Engineering, Barcelona Supercomputing Center (BSC-CNS), Barcelona, Spain

[3]University Grenoble Alpes, INAC-MEM, Grenoble, France

[4]CEA, INAC-MEM, Grenoble, France

[5]Laboratoire de Biologie Structurale et Radiologie, Service de Bioénergétique, Biologie Structurale et Mécanisme, Institut de Biologie et de Technologie de Saclay, CEA Saclay, Gif-sur-Yvette Cedex, France

Conflict of interest: The authors have declared no conflicts of interest for this article.

DFT has been, for more than 20 years, the workhorse method for simulations within the solid state community. Moreover, in spite of the fact that DFT drastically reduces the complexity with respect to the *ab initio* methods of Quantum Chemistry, the success of such a treatment in the latter community is undeniable. This is mainly due to the fact that, on one hand, the quality of the exchange and correlation functionals available permits the calculation of certain properties with almost chemical accuracy, and, on the other hand, there are numerous software packages, which are relatively easy to use, that have contributed to the diffusion of the computational approach.

During the past years, there has been a multiplication of DFT software packages that are able to treat systems of increasingly large size. This has, on one hand, been enabled by both the advances in supercomputing architectures and the code developers continuously improving their codes to exploit the steadily increasing performance provided, but it is at the same time also motivated by various scientific needs. This fact clearly extends the range of possible applications to new fields, and to communities traditionally focused on larger systems. In a similar manner to the uptake of DFT in the Quantum Chemistry community, things are progressing as if we are entering a 'second era' of DFT calculations, where DFT and, more generally, large-scale QM treatments, are susceptible to wide diffusion in other communities.

In this review paper, we will present some of the motivations that led computational physicists and quantum chemists into this second era. The different aspects will be separated into various subcategories, while trying to give a general overview, and will be completed by notable examples in the literature. This inspection of the state-of-the-art will provide the reader with an outlook on the present capabilities of QM approaches. We will then continue our discussion by presenting the key concepts that have emerged in the last decade. These concepts are often specific to QM calculations at large scale and are rather different from those which are typical of traditional calculations, where the systems' sizes are limited to a few hundreds of orbitals. These concepts are therefore of high importance for potential users of such advanced DFT methods.

The Need for Large-scale QM

Given the unbiased predictive power of Quantum Mechanics, there is obviously no need to explain why systems containing only a few atoms should be modeled using this approach. However, for simulations of systems at the nanoscale, composed of many thousands of atoms, the question of the need for a QM treatment might appear legitimate. For systems of these sizes, the electronic degrees of freedom are seldom of interest, and the interatomic potential might be described by more compact approaches like Force Fields, possibly fine-tuned for describing experimentally known structural and dynamical or polarizability properties. In other terms, the intimate nature of the problem changes: instead of focusing solely on the correct estimation of interactions and correlation between electrons, one rather has to concentrate on the exhaustive *sampling* of the configuration space, thereby losing the need for an intrinsic QM description.

In addition we know that, even if a QM approach were feasible, this would not necessarily lead to a better description. Although the complexity of the model is certainly higher and the description is less biased, there are still many approximations which are hidden in a QM calculation; therefore we are—even for a system containing only a handful of atoms—in general still far from chemical accuracy. The situation for large systems will be the same or even worse, since more severe approximations have to be adopted. However, the need for QM calculations of large systems does not solely come from a quest for accuracy. Indeed there are other reasons why an *ab initio* description for large systems is desirable or even crucial, and one of the purposes of this review paper is to identify and discuss some of these aspects.

The present-day scenario of the available methodological techniques to study systems at the atomistic level can be sketched in Figure 1: here various methods are illustrated within the typical scales where they have been usually applied. It is interesting to notice that the size where typical QM approaches are developed and improved is of the order of few atoms, even though some of these concepts are then also applied to larger systems. John Perdew introduced the renowned metaphor of *Jacob's Ladder*,[5] where the computational complexity of the implementation of the DFT exchange and correlation functional is (in principle) directly related to the accuracy of the description, aiming at the 'heaven' of chemical accuracy.

Likewise, for more than 10 years, a lot of work has been done to extend the range of applicability of QM methods to larger systems. The so-called *nearsightedness property*[6] suggested that, at least in principle, one could exploit locality to build linear-scaling methods that are able to reach larger scales. Initially, the development of such computational methods was

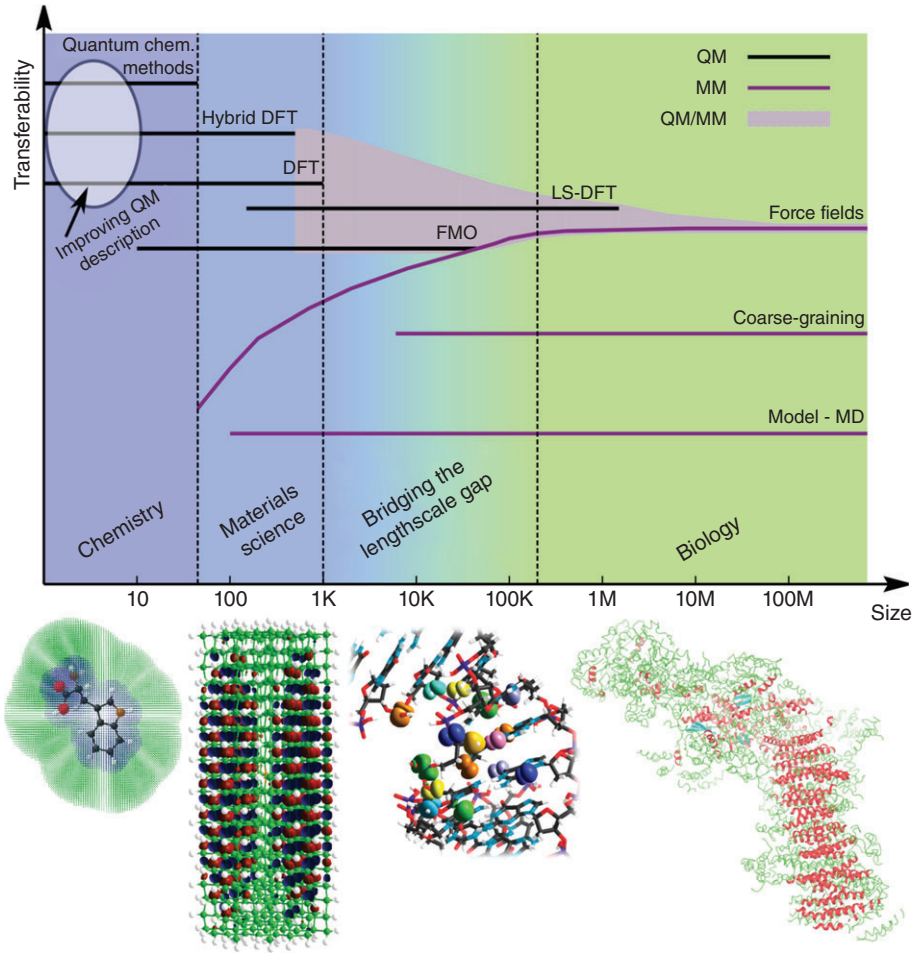

FIGURE 1 | Overview of the popular methods used in simulations of systems with atomistic resolution, showing the typical length scales over which they are applied as well as the degree of transferability of each method, i.e., the extent to which they give accurate results across different systems without re-tuning. On the left hand side we have the Quantum Chemistry methods which are highly transferable but only applicable to a few tens of atoms; on the right hand side we see the less transferable (semi-)empirical methods, which can however express reliable results (as they are parameterized for) for systems containing millions of atoms; and in the middle we see the methods—in particular linear-scaling DFT—which can bridge the gap between the two regimes. The vertical divisions and corresponding background colors give an indication of the fields in which the methods are typically applied, namely chemistry, materials science, biology, and an intermediary regime ('bridging the length scale gap') between materials science and biology. The line colors indicate whether a method is QM or MM, while the typical regime for QM/MM methods is indicated by the shaded region. In the top left the region wherein efforts to improve the quantum mechanical treatment are focussed, that is the quest to climb 'Jacob's ladder' by developing new and improved exchange-correlation functionals, is also highlighted. Some representative systems for the different regimes are depicted along the bottom: the amino acid tryptophan with a multi-resolution grid, a defective Si nanotube with an extended KS wavefunction, DNA with localized orbitals, and the protein mitochondrial NADH:ubiquinone oxidoreductase.[4]

driven by the 'academic' purpose of verifying the computational consequences of nearsightedness. In *Large Scale QM: Methodological and Computational Approaches* section we will overview the most important advancements in this topic and some of the established computational approaches in large-scale QM. This is by no means new and there are a number of valid review papers on the topic, to which we will also refer. Our aim is not to be fully exhaustive on this topic as there has been many research studies in this direction. However we would like to put the emphasis on the fact that nowadays the panorama is so rich and there is enough diversity in the computational approaches to claim that such a discipline is now mature enough to be largely diffused also among nonspecialists.

The reason for this diffusion is related to the *opportunities* that a QM approach opens for systems composed of thousands, if not hundreds of thousands, atoms. On the one hand, there are quantities which are intrinsically only accessible using QM, e.g., all investigations dealing with electronic

excitations[7]; we will present some more examples in *Large-Scale QM Applications* and *Multiscale Linked Together: An Example* sections. On the other hand, large QM calculations are also needed to access *error bars and statistics* of the results. An example is the need to get good statistics among different constituents in a morphology—a task which is not possible by implicit, classical modeling of the environment. In addition, another aspect where first-principles QM approaches are important is the need for *validation* of non-QM approaches.

For all of these tasks, there is a typical length scale, ranging from a few hundred to many thousands of atoms, where it is important to master *both* QM and classical approaches. As discussed in the *Introduction* section, it was not possible in the initial implementations of DFT software packages—for various reasons, including the available computational resources—to reach such large length scales, i.e., there was a 'length scale gap' between the maximum scale which was accessible to QM and the typical scale at which classical approaches are applied. QM computational paradigms had to bridge this gap in order to be used as investigation tools for systems of many thousand atoms. However, since a few years ago and mostly driven by the development of linear-scaling QM methods, this gap has vanished. Thus, intensive research and investigation in this range will allow the set up of new, powerful computational approaches in various disciplines such as soft matter, biology, and life sciences.

LARGE SCALE QM: METHODOLOGICAL AND COMPUTATIONAL APPROACHES

The problem of treating large systems with DFT is not a new one; indeed research into this area goes back more than two decades.[8,9] This work focused on developing new methods with reduced scaling, leading to the different linear-scaling DFT (LS-DFT) codes which exist today. The emphasis was initially on academic interest, that is to say the focus was on the methods themselves and finding new and better ways to accelerate calculations of ever larger systems, rather than on the application to major scientific problems. Indeed, until more recently, the vast majority of applications were limited to proof-of-concept calculations, which served to demonstrate the capability of these algorithms to treat ever larger systems, while hinting at future possibilities for production calculations. Nonetheless, without this pioneering work, we would not be in the position today to tackle large and challenging systems such as those discussed in more detail below.

The development of reduced scaling methods was also naturally coupled with the availability of high performance computing (HPC) resources; thanks to both the increase in computing power of the fastest supercomputers and the widespread availability of commodity clusters, LS-DFT can now not only be applied to very large systems indeed, but it can also do so while maintaining the same accuracy as more traditional cubic-scaling approaches.

As a result of the complexities involved in such methods, their usage was initially mostly limited to experts within the community. This is no longer entirely the case, however, there remain a number of additional concepts with which interested users must familiarize themselves before attempting practical calculations. In this section, we give an overview of some of these important concepts, notably the quantum-mechanical principle of nearsightedness, which provides the justification for linear-scaling methods, and the codes within which they are implemented. This is not intended to be a fully exhaustive list, rather the aim is to highlight the most popular approaches and some of the key achievements within the field. For a more thorough discussion, the reader is encouraged to refer to other, more extensive reviews of the subject.[8–11]

Nearsightedness and Linear Scaling

In the context of DFT, the tendency of the Kinetic Energy (KE) operator to favor the delocalisation of the Kohn-Sham orbitals means that they are in general extended over the entire system. This nonlocality leads to an unfavorable cubic scaling, meaning that an increase of the system size by a factor often leads to a computational effort which is 1000 times greater. Even though this is considerably better than the scaling of other popular Quantum Chemistry methods, which ranges from $\mathcal{O}(N^4)$ for Hartree Fock (HF) to $\mathcal{O}(N^5)$ for MP2, $\mathcal{O}(N^6)$ for MP3, and $\mathcal{O}(N^7)$ for MP4, CISD(T) and CCSD(T), it still makes large scale simulations prohibitive.

On the other hand, the density matrix $F(\mathbf{r}; \mathbf{r}')$, which is an integrated quantity that is invariant under unitary transformations of the Kohn-Sham orbitals, does not reflect this nonlocality. Indeed it can be shown that the elements of the density matrix decay rapidly with respect to the distance between $\mathbf{r}$ and $\mathbf{r}'$: for insulators and metals at finite temperature exponentially,[12–18] and for metals at zero temperature algebraically.[19] Kohn has coined the term 'nearsightedness' for this effect,[20] and this concept is

the key towards calculations of very large systems: by truncating elements beyond a given cutoff radius it is possible to reach an algorithm which scales only linearly with respect to the size of the system.

An illustration of this effect is shown in Figure 2 for the case of a water droplet containing 1500 atoms. Here we plot on the left side the isosurface of an extended Kohn-Sham orbital, and on the right side the density matrix of the system in the x dimension, i.e., $F(x, y_0, z_0; x', y_0, z_0) = \sum_i f(\varepsilon_i)\psi_i(x, y_0, z_0)\psi_i(x', y_0, z_0)$, where ψ_i are the Kohn-Sham orbitals and $f(\varepsilon_i)$ their occupation numbers. As can be seen, the summation of the extended orbitals nevertheless leads to a localized quantity, meaning that the nonlocal contributions are canceled due to interference effects.

In a linear-scaling DFT approach, this locality must be taken advantage of, which can be achieved by building the algorithm directly on the density matrix, rather than the Kohn-Sham orbitals. Since this may introduce an additional computational overhead, these $\mathcal{O}(N)$ algorithms are usually slower for small systems than traditional approaches and only outperform the latter ones beyond a critical system size, the so-called crossover point. This crossover point is dependent not only on the details of the method used, but also the properties, in particular the dimensionality of the system being studied. In many linear-scaling DFT approaches, such as ONETEP,[21] CONQUEST,[22] QUICKSTEP,[23] and BIGDFT,[24,25] the density matrix is written in separable form as

$$F(\mathbf{r},\mathbf{r}') = \sum_{\alpha,\beta} \Phi_\alpha(\mathbf{r}) K^{\alpha\beta} \Phi_\beta(\mathbf{r}') \qquad (1)$$

with a set of so-called support functions $\varphi_\alpha(\mathbf{r})$ and the density kernel $\mathbf{K}$. In order to reach a linear complexity, the support functions are strictly localized and the density kernel is enforced to be sparse, meaning that elements are set to zero beyond a given cutoff radius. Different approaches can be used to find the ground state density matrix, which are discussed below.

Reduced-scaling Approaches and Established Codes

In the following we describe both pioneering early approaches to LS-DFT and modern, state of the art methods currently being used for applications. Since this review is intended to be of practical use rather than purely theoretical, where appropriate, we categorize the various approaches by the code in which they are implemented. It should be noted that many, though not all, of the approaches to LS-DFT described below are valid only for systems with a band gap, since, as mentioned above, the density matrix decays only algebraically at zero temperature for metals, rather than exponentially. Exponential decay, is, however, recovered for metals at finite temperature, thereby providing one avenue for LS-DFT with metals. Where relevant, we mention if the codes are capable of treating metallic systems.

Pioneering Order N Methods

The earliest LS-DFT method was the divide and conquer approach of Yang.[26] As the name implies, in this approach the system of interest is divided into a number of smaller subsystems which can be treated independently using a local approximation to the Hamiltonian. The KS energy for the full system is extracted from the subsystems, which are coupled via the local potential and Fermi energy. Since the size of the subsystems is independent of the total system

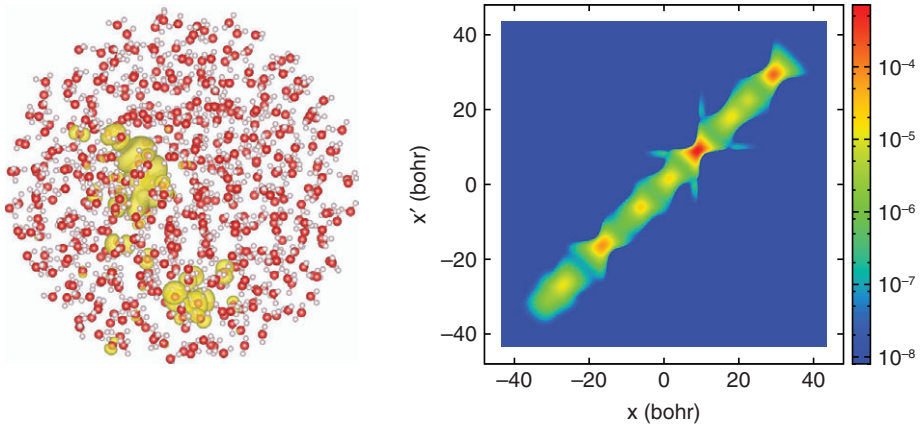

FIGURE 2 | Left: Isosurface of one Kohn-Sham orbital for a water droplet consisting of 1500 atoms. Right: density matrix in the x dimension, i.e., $F(x, y_0, z_0; x', y_0, z_0)$, for the same system.

size, the method scales linearly, and is also straightforward to parallelize, however, the crossover point can be rather high.

Another pioneering work proposed an order N method to calculate the density of states and the band structure by means of the Green function and a recursion method to calculate the moments of the electronic density[27] in real space using a finite difference scheme. The evaluation of the moments of the electronic density by means of random vectors was also used to treat systems up to 2160 atoms.[28] Many density matrix minimization methods were also developed; the Fermi Operator Expansion (FOE),[29,30] LNV (Li-Nunes Vanderbilt)[31,32] and other approaches related to the purification transformation[33] are currently used in various codes today. The authors refer to the comprehensive review of S. Goedecker[9] for a full description of these different methods.

SIESTA

The first widely used code with a linear-scaling method was SIESTA,[34–36] which is based on numerical atomic orbitals. The original linear-scaling approach is based on the minimization of the functional of Kim et al.,[37] which avoids explicit orthogonalization. More recently a divide-and-conquer algorithm has also been implemented.[38] There are many real applications using SIESTA for large systems, but these calculations use a traditional cubic-scaling scheme based on the diagonalization of the Hamiltonian. In 2000, λ-DNA of 715 atoms was calculated using the linear-scaling method to show the absence of DC-conduction.[39] In 2006, the calculation of some CDK2 inhibitors was done using SIESTA, with also a comparison to ONETEP.[40] Recently SIESTA has been coupled[41] with the PEXSI library,[42] which avoids the cubic-scaling diagonalization of the Hamiltonian by taking advantage of its sparsity in the localized basis. This reduces the computational complexity without the need for nearsightedness or other simplifications, thus allowing considerably larger systems to be tackled without requiring any explicit truncation of the density matrix. The first published scientific application of SIESTA-PEXSI examines carbon nanoflakes up to a size of 11,700 atoms.[43] In addition SIESTA allows one to perform electron transport calculations using the TranSIESTA tool[44] providing a tight-binding Hamiltonian and can also be used for QM/MM simulations[45]

ONETEP

The Order-N Electronic Total Energy Package ONETEP[21,46–48] is a LS-DFT code which employs a density matrix approach, wherein the strictly localized support functions, termed Nonorthogonal Generalized Wannier Functions (NGWFs), are represented in a basis of periodic sinc (psinc) functions and optimized *in situ*, adapting themselves to the chemical environment. Since the psinc basis can be directly related to plane-waves, the NGWFs form a localized minimal basis with the same accuracy as a plane-wave calculation. The density kernel is calculated primarily using the LNV approach in combination with other methods.[49] A number of functionalities have been implemented in ONETEP such as DFT+U,[50] the calculation of optical spectra,[51] including via time-dependent (TD) DFT,[52,53] constrained DFT,[54] electronic transport,[55] natural bond orbital analysis,[56] and implicit solvents.[57] A method to treat metallic systems at finite temperature has also been implemented.[58]

Many large calculations have been performed with ONETEP, such as on DNA (2606 atoms),[21] carbon nanotubes (4000 atoms),[59] a silicon crystal (4096 atoms),[60] and point defects in Al_2O_3.[59] ONETEP was also used in biology to study the binding process within a 1000-atom QM model of the myoglobin metalloprotein[61] and also, in a QM/MM approach, the transition state (TS) optimization of some enzyme-catalyzed reactions.[62] Some of the applications with ONETEP have clearly highlighted the need for and challenges associated with large scale QM calculations. For example there is a clear need for methods capable not only of incorporating and analyzing electronic effects on a large scale in proteins including the solvent effect, at least implicitly, as demonstrated by the study of a 2615-atom protein–ligand complex,[57] but also of optimizing TS structures in this context. In another study, ONETEP was used to put in evidence the importance of preparing systems correctly to avoid the problem of the vanishing gap for large systems (proteins and water clusters).[63]

OpenMX

The OPENMX code[64,65] has both a linear-scaling version, based on the divide and conquer approach defined in a Krylov subspace,[64] and a cubic-scaling version which uses diagonalization. It uses a basis set of pseudo-atomic orbitals (PAOs) and a number of functionalities have been implemented, such as DFT+U,[66] electronic transport[67] and the calculation of natural bond orbitals.[68] This latter capability was used to analyze a molecular dynamics (MD) simulation on a liquid electrolyte bulk model, namely propylene carbonate + $LiBF_4$ in a model containing 2176 atoms.[68]

FHI-aims

The Fritz-Haber-Institute *ab initio* molecular simulations package[69,70] uses explicit confining potentials to construct numerical atom-centered orbital basis functions; around 50 basis functions per atom are needed to have an accurate solution of less than one meV per atom. This scheme can be used naturally to achieve quasi-linear scaling for the grid based operations[71] with a demonstrated $O(N^{1.5})$ overall scaling for a linear system of polyalanine up to 603 atoms. The authors of FHI-aims have also developed a massively parallel eigensolver, ELPA,[72] for large dense matrices based on a two-step procedure (full matrix to a banded one, and banded matrix to a tridagonal one). Traditional DFT and embedded-cluster DFT[73] calculations can be done not only on molecules,[74] but also on periodic systems. Hybrid functionals,[74] RPA, MP2, and GW methods are also implemented using a resolution of identity[75] based on auxiliary basis functions.

CONQUEST

CONQUEST[10,22,76,77] uses an approach based on support functions and density matrix minimization. The support functions can be represented either in a systematic B-spline basis, or in a basis of PAOs, according to the user's preference. There is also a choice of using a linear-scaling approach wherein the density matrix is optimized using LNV or a cubic-scaling approach using diagonalization. Constrained DFT[78] and multisite support functions are implemented, wherein the support functions are associated with more than one atom.[79] Scaling tests have been performed on up to 2 million atoms of bulk Si[22] and the approach has also been applied to Ge hut clusters on Si, for systems of up to 23,000 atoms.[80] Other examples of calculations with CONQUEST include 3400 atom simulations of hydrated DNA[81] and MD simulations of over 30,000 atoms of crystalline Si[82] using the extended Lagrangian Born-Oppenheimer method.[83]

BIGDFT

The BIGDFT code[84] emerged as an outcome of an EU project in 2008. One of the most particular features of this code is the basis set it uses, Daubechies wavelets.[85] These functions have the remarkable property of—at the same time—being orthonormal, having compact support in both real and reciprocal space and forming a complete basis set. Such a basis set offers optimal properties for DFT at large scale. The code was first designed following a traditional cubic-scaling approach,[86] and later complemented with a linear-scaling algorithm.[24,25] Since wavelets form a very accurate basis set, BIGDFT is—in conjunction with elaborate pseudopotentials—capable of yielding a very high precision[87] at maintainable computational costs. This is also true for the linear-scaling version, where the support functions are expanded in the wavelet basis and can thus be adapted *in situ*. The main approach used to optimize the density matrix and thereby achieve linear scaling is FOE.[9]

Some features implemented in BIGDFT are, among others, time-dependent DFT[88] and constrained DFT, which has been implemented based on a fragment approach,[89] along a similar spirit to the fragment molecular orbital (FMO) approach described in more detail below. In addition BIGDFT incorporates a very efficient Poisson Solver based on interpolating scaling functions,[90–93] which solves the electrostatic problem with a low $\mathcal{O}(N\log N)$ complexity and a small prefactor and can thus also be used for large scale applications. BIGDFT was also one of the first DFT codes taking benefit of accelerators used in HPC systems, such as Graphic Processing Units.[94] Some of the code developers are among the authors of the present review, therefore some illustrative examples that will be given in the following sections originate from runs with BIGDFT.

ERGOSCF

ERGOSCF[95,96] is a quantum chemistry code for large scale HF and DFT calculations, which has a variety of pure hybrid functionals available and is an all-electron approach based on Gaussian basis sets. It uses a trace-correcting purification method in conjunction with fast multipole methods, hierarchic sparse matrix algebra, and efficient integral screening to achieve linear scaling. It has been applied to protein calculations, using both explicit and implicit solvents.[97]

FREEON

Formerly mondoscf, FREEON is a suite of linear-scaling experimental chemistry programs[98] which performs HF, pure DFT, and hybrid HF/DFT calculations in a Cartesian-Gaussian LCAO basis. All algorithms are $O(N)$ or $O(NlogN)$ for nonmetallic systems. Different purification and density matrix minimization approaches have been implemented and compared in the code.[99]

QUICKSTEP

The QM part of the CP2K[100] package, QUICKSTEP,[23] uses traditional Gaussian basis sets to expand the orbitals, whereas the electronic density is expressed in plane waves to perform HF, DFT, hybrid HF/DFT

and MP2[101] calculations. The Kohn-Sham energy and Hamiltonian matrix is calculated using a linear-scaling approach with screening techniques.[23]

FEMTECK

The Finite Element Method based Total Energy Calculation Kit FEMTECK code[102,103] uses an adaptive finite element basis to represent Wannier functions, in conjunction with the augmented orbital minimization method (OMM), which imposes additional constraints on the orbitals to guarantee linear independence in order to overcome the slow convergence and local minima usually associated with the standard OMM method. FEMTECK has been used for MD simulations of liquid ethanol with 1125 atoms,[104] as well as for the study of fast-ionic conductivity of Li ions in the high-temperature hexagonal phase of $LiBH_4$, in MD simulations of 1200 atoms.[105]

RMGDFT

The real space multigrid based DFT electronic structure code[106] (RMGDFT) uses a multigrid[107] or a structured nonuniform mesh.[108] A linear-scaling method[109] was also developed. Using maximally localized Wannier functions expressed on a uniform finite difference mesh, Osei-Kuffuor and Fattebert[110] performed a MD simulation up to 101,952 atoms of polymers to demonstrate the scalability of their algorithms.

PROFESS

As the imposition of orthonormality constraints on the KS orbitals is one of the factors which dominates the cubic-scaling of standard DFT, one strategy to achieve linear-scaling is to eliminate the need for the orbitals. This so-called orbital-free (OF) DFT approach does so by defining a KE functional, for which several forms have been proposed, see Refs 111–113. Such an approach has been implemented in PROFESS (Princeton Orbital-Free Electronic Structure Software),[114–117] which offers a choice between several implemented KE functionals using a grid-based approach to represent the density. The code requires the use of local pseudopotentials, which are provided for certain elements only, including Mg, Si, and Al. The approach only achieves the same accuracy as KS-DFT for main group elements in metallic states, but recent work developing new KE functionals for semiconductors and transition metals[113,118,119] allow some properties of semiconductors to also be reproduced well. Despite this limitation, large defects in crystals (e.g., dislocations, grain boundaries) and large nanostructures (e.g., nanowires, quantum dots) are too computationally costly to treat with most first principles approaches, and so OF-DFT offers an appealing alternative for such systems. The code has been used to simulate more than 1 million atoms of bulk Al[120] and to study melting of Li using MD.[121]

Quantum Chemistry

Reduced-scaling electronic structure methods are a domain where the ultimate goal is to have a linear-scaling approach with chemical accuracy. Based on pair natural orbitals, a coupled cluster theory method[122] has been developed which scales up to 1000 atoms claiming that chemical accuracy was achieved. A Quantum Monte Carlo method is also being developed for large chemical systems with some calculations on peptides[123,124] up to 1731 electrons.

Machine Learning

Finally we wish to mention the various works on neural networks and other machine learning techniques where the goal is to obtain interatomic potentials with the same accuracy as DFT or even quantum chemistry. The first works[125–127] of J. Behler using a high-dimensional neural network give a way of calculating potential-energy surfaces in order to perform metadynamics. Another approach is to use the electronic charge density coming from DFT to build interatomic potentials for ionic systems.[128] By using GPUs it is possible to speed up the neural network performance by two orders of magnitude, which permits a large computing capacity within a single workstation. Rupp et al. considerably improved the predictive precision and transferability of spectroscopically relevant observables and atomic forces for molecules using kernel ridge regression.[129] Once such a system is trained, the cost for new calculations is orders of magnitude smaller than for corresponding DFT calculations.

Towards Coarse-graining Modeling of Large Systems: FMO and DFTB Approaches

Fragment Molecular Orbitals

One important method for proteins and other biological molecules, which could be considered an extension of the divide-and-conquer approach, is the FMO approach.[130,131] In this approach, the molecule is divided into fragments—whose definition is based on chemical intuition—which are each assigned a number of electrons. The size of each fragment might

therefore vary depending on the system in question, e.g., 2D pi-conjugated systems would need large fragments for accuracy. The molecular orbitals (MOs) are then calculated for each fragment, under the constraint that they remain localized within the fragment. The distinguishing feature of FMO compared with divide and conquer is that the MOs for the fragments are calculated in the Coulomb field coming from the rest of the system (i.e., the environment), so that long range electrostatics are included. The fragment MOs must be updated iteratively to ensure self-consistency of this environment electrostatic potential. Different levels of approximation can be used: the most basic is FMO1, which only explicitly calculates MOs of single fragments (referred to as monomers) and constructs the total energy from these results. The next, and most common level, is FMO2, which also incorporates explicit dimer calculations (i.e., between pairs of fragments) into the total energy. There is also FMO3,[132] which also adds trimers, and even FMO4, which incorporates also 4-body terms.[133] The accuracy of the approximation, but also the cost increases with the addition of higher order terms. FMO2 is often sufficiently accurate for many applications, but there are some cases where higher order interactions are required, e.g., at least three-body terms were found to be necessary for MD of water;[134] geometry optimizations of open shell systems may also require higher order terms, or, where possible, larger fragments in order to ensure good convergence.[135] FMO has been implemented in GAMESS (General Atomic and Molecular Electronic Structure System),[136–138] with an implementation which is designed to exploit massively parallel machines; ABINIT-MP,[139,140] which has also been designed for massively parallel calculations;[141] and a version of NWChem.[142]

FMO belongs to a wider class of fragment-based methods, such as the molecular tailoring approach.[143] FMO, molecular tailoring and other related approaches have been reviewed in detail elsewhere, along with a number of example applications.[131,144–146] Here we highlight a few examples for large systems. Aside from DFT, FMO may also be used for HF and MP2 calculations; e.g., the GAMESS implementation has been benchmarked for water clusters containing around 12,000 atoms at the MP2 level of theory.[147] FMO has also been used for geometry optimizations of large systems, e.g., the prostanglandin synthase in complex with ibuprofen, containing around 20,000 atoms, was optimized using B3LYP and restricted HF (RHF) for different domains of the system.[148] Another example application is the study of the influenza virus hemagglutinin, where QM calculations of up to 24,000 atoms[149,150] have been performed using FMO-MP2 in combination with the polarizable continuum model (PCM).[151,152] More than 20,000 atoms were also included in a RHF simulation of the photosynthetic reaction center of rhodopseudomonas viridis, which required around 1400 fragments.[153] While less common, FMO can also be applied to solids, surfaces and nanomaterials. For example a new fragmentation scheme for fractioned bonds was developed and applied to the adsorption of toluene and phenol on zeolite;[154] Si nanowires have also been studied using FMO.[155] FMO may also be used for excited calculations, e.g., in combination with TDDFT, which has been tested for solid state quinacridone.[156]

DFTB

The Density Functional Tight-Binding approach was first notably applied in carbon-based systems. The idea was, at the first-order, to use a frozen density from atoms. DFTB, at the second order, has been extended in order to include a Self-Consistent Charge (SCC) correction,[157] accounting for valence electron density redistribution due to the interatomic interactions. A third-order DFTB3[158] was also developed which has introduced an additional term with coupling between charges. Parameters for the whole periodic table are available.[159] A confinement potential was used to tighten the Kohn-Sham orbitals. The solution conformations of biologically mono- and di-α-D-arabinofuranosides were investigated[160] by means of MD using dispersion-corrected self-consistent DFTB and compared to the results from the AMBER ff99SB force field[161] with the GLYCAM (version 04f) parameter set for carbohydrates[162,163] as well as to NMR experiments. There are also some extensive tests on hydroxide water clusters and aqueous hydroxide solutions.[164]

FMO-DFTB

Recently, the FMO approach has been combined with DFTB,[165] with the aim being to reduce the cost of DFTB to a few seconds, in order to perform MD simulations. The accuracy of FMO-DFTB is very close to that of DFTB, while excellent speedups have also been achieved: for an MD simulation of 768 atoms of water, the speedup compared to DFTB was shown to be more than 100.[166] It has been used to optimize an 11,000 atom nanoflake of cellulose $I\beta$,[158] as well as for MD simulations of liquid hydrogen halides containing 2000 atoms,[166] for which the speedup was an order of magnitude greater than the above example. FMO-DFTB could therefore be a

very promising approach of QM-MD of large systems.

HPC Concepts/Performance

The effort for the development of the above mentioned computer codes has also contributed to another improvement in the community: the ability to exploit HPC resources. This has become a very important aspect with the advent of petaflop supercomputers. New science can be done on these machines only if the code developers are able to profit from such large scale supercomputers. However, the development of accelerated methods for large systems is not meant to replace the exploitation of powerful HPC platforms, rather the two go hand in hand: in order to execute calculations for systems as large as the range for which they are designed, LS-DFT codes require large computing resources; and parallel compute clusters are most efficiently exploited if they are used to treat large problem sizes, rather than to compensate for the cubic scaling of standard DFT codes.

In the ideal case, for a method which exhibits both perfect linear scaling with respect to the number of atoms and ideal parallel scaling with respect to the number of computing cores, the time taken for a single point calculation should remain constant if the ratio of atoms to cores remains constant, that is, a so-called weak scaling curve would be flat. This allows for the definition of the concept of *CPU minutes per atom* and, correspondingly, *memory per atom*. Since both the time and memory requirements for a small system running on a few cores would be approximately the same as a large system running on many cores, this value for a particular code is a function of the system, which depends on the dimensionality of the material, the atomic species and various user-defined quantities, such as the grid spacing and the localization radii beyond which localized basis functions are truncated.

In other terms, we might say that the values of per-atom computing resources for a given code are functionals of the input parameters of the code and of the computing architecture employed. However, even though they cannot be predicted beforehand and have to be evaluated, it is interesting to compare values of CPU minutes and memory per atom between different systems. This is especially useful because such a viewpoint provides a quantitative method for estimating the cost of a large simulation based on a small representative calculation, i.e., a smaller but equivalent simulation domain running on the same computing architecture. For example, when one has a fixed number of cores available, one could estimate the total run time for a large calculation using the CPU minutes per atom value obtained from the small calculation. Alternatively, in the case where one has many cores but limited memory availability, one could use the memory per atom value from the small calculation to determine the minimum number of cores needed to fit the large problem in memory. Although in practice one might not achieve perfect parallel scaling, the validity of these quantities has been demonstrated in the context of the BigDFT code, where it has been tested for DNA fragments and water droplets.[25] A similar concept was also demonstrated in the context of coupled cluster theory.[122]

It is however fundamental that all these performance achievements come together with the *robustness* of the approach, or more precisely its implementation. It is easy to imagine that QM methods become technically very complicated at large scale, with a multiplication of the input variables and troubleshooting techniques which are typical of the algorithm employed. Code developers have to provide robust and reliable algorithms for nonspecialists (failsafe mode), even at the cost of lower performance.

LARGE-SCALE QM APPLICATIONS

As already mentioned, first principle calculations are *a priori* the most accurate approach to any atomistic simulation. Unfortunately an exact analytical or computational solution to the fundamental QM equations is only possible for a few, rare cases at present; for all other systems one either has to introduce approximations, solve the equations numerically, or both. Owing to these approximations there might thus even be situations where an empirical approach, which is tuned for one particular property, might yield more precise results than a first principles calculation. One particular, but important example is water, where traditional *ab initio* approaches like DFT do not come as close to the experimental values (see, for instance, Ref 167 and references therein) as empirical force fields.[168,169] On the other hand such empirical approaches will only work for systems which are very similar to the ones which were used for the tuning and will in general fail for systems which are distinct, making the simulation of unknown materials tricky. In addition, traditional force fields do not in general allow for bond breaking and forming, which is however abundant in chemical reactions. This is in strong contrast to *ab initio*

approaches, which are less biased and are thus expected to yield the correct tendencies over a much larger range of systems.

Moreover there are situations where a first principles description is not only desirable, but also essential. Obviously this is the case when quantities are needed which are not accessible with classical force field approaches. For instance, a major shortcoming of classical approaches is that they cannot provide direct insights into electronic charge rearrangements. This is however necessary if one wants to analyze charge transfers, which play an important role for instance in biology. Since such electronic charge transfers can occur over a large distance and over a long time frame, such simulations would quickly go beyond the scope of pure *ab initio* calculations. A possible solution is to let the system evolve according to a classical approach, which is orders of magnitudes faster, and only analyze certain snapshots or averages on a QM level. An example is the work of Livshitz et al.,[170] which investigates the charge transfer properties of DNA molecules adsorbed onto a mica surface, or the work of Lech et al.[171] which investigates the electron–hole transfer in various stacking geometries of nuclear acids. With respect to DNA, *ab initio* calculations have also been used to investigate the molecular interactions of nucleic acid bases[172] and to study the impact of ion polarization.[173] A nice overview over the various approaches for DNA can be found in[174]

The volume of an atom is also an example of such an intrinsic QM quantity, which is most straightforwardly defined using its charge distribution, but cannot be accessed directly using force fields; consequently other models must be adopted.[175] Another case where first principles calculations are needed is the determination of photophysical and spectroscopic properties. These calculations do not only require the determination of the ground state, but also of excited states, which is not possible with classical approaches based on empirical force fields. This application is also very demanding from the point of view of the QM method, since popular approaches like HF or DFT are ground state theories and therefore usually give rather poor results for excited states. A popular solution to this problem is to use TDDFT for the calculation of the excitations.[88]

Highly accurate QM methods are also required in the determination of small energy differences, for instance, the calculation of activation energy barriers in chemical reactions. The problem is that, in particular for biological systems, the reaction is often catalyzed by an environment which can be much larger than the actual active site, making a fully *ab initio* treatment impossible. However, if there is no charge transfer between the active site and its environment, it is possible to use so-called QM/MM schemes, where only the active site is treated on a highly accurate *ab initio* level and the environment is handled using force fields. As an example, such an approach was used, among others, to investigate catalysts for the Kemp eliminase,[176] and QM calculations in general are an important ingredient for the computational design of enzymes.[177]

Obviously such a QM/MM strategy raises the question of how the coupling between QM and MM regions can be done;[178] for an interesting review on the subject. Typically one distinguishes between three different setups, namely mechanical embedding, electronic embedding and polarized embedding.[179–181] In the first case, the interaction between the QM and MM region is treated in the same way as the interaction within the MM region itself; in the second case, the MM environment is incorporated into the QM Hamiltonian, thus leading to a polarization of the electronic charge density; and in the third case, the MM environment is also polarized by the QM charge distribution. The second method is the most popular one, and the coupling between QM and MM region can for instance be done using a multipole representation which is fitted to the exact electrostatic potential.[182]

Unfortunately electronic embedding is known to exhibit the shortcoming of 'overpolarization' at the boundary between the QM and MM region,[181,183] in particular if covalent bonds are cut. This overpolarization problem is due to the fact that for the electrons of the MM atoms the Pauli repulsion is not accounted for, resulting in an incorrect description of the short range interaction at the QM/MM interface. In particular, positive atoms on the MM side might act as traps for QM electrons, leading to an excessive polarization. However, there exist several approaches to address this issue, for instance the use of a delocalized charge distribution for the MM atoms,[184–186] and indeed it can be shown that a careful implementation allows the correct description of polarization effects within a QM/MM approach.[187] Another, more straightforward solution to this problem is to increase the size of the QM region, keeping the problematic boundary farther away from the active site. Since in this way the QM region easily contains hundreds or thousands of atoms, the use of a linear-scaling algorithm for the QM part is indispensable. In a ONETEP application by Zuehlsdorff et al.[188] they even found that an explicit inclusion of the solvent into the QM region was required to get a reliable description.

In another recent work employing ONETEP, Lever et al.[62] used this LS-DFT code to investigate the transition states (TS) in enzyme-catalyzed reactions. The same reaction has already been investigated earlier with a QM/MM approach employing for the QM part highly accurate quantum chemistry methods such as MP2, LMP2, and LCCSD(T0).[189] Even though the development of reduced scaling algorithms[190–196] made their use somewhat affordable also for larger systems, the cost of these methods is still very high. On the other hand they have the advantage that they yield very accurate estimates for activation barriers, in contrast to DFT, which in general tends to underestimate these values. Indeed another study by Mlýnský et al. demonstrated the need for using appropriate methods for the QM treatment within a QM/MM approach. Whereas all methods (quantum chemistry, DFT and semi-empirical) gave similar reaction barriers, the reaction pathways were considerably different for the semi-empirical calculations.[197] Recent developments also allow the embedding of a small region which is treated by quantum chemistry methods within a larger region which is treated by DFT, and both regions can then also be used within a QM/MM approach.[198,199]

Instead of falling back to expensive quantum chemistry methods it is also possible to improve the accuracy of DFT calculations in a cheap way by including dispersion corrections. A study by Lonsdale et al. found that this considerably improved the values of the calculated energy barriers.[200] Generally speaking, it is recommendable to include such dispersion corrections in any large scale QM calculation, as they add only a small overhead, but may improve the physical description considerably.

The QM/MM philosophy is also useful to calculate, for a given subsystem, quantities which are intrinsically only possible with an *ab initio* approach, but influenced by a surrounding which does not require a strict first principles treatment. For instance, excitation energies and the absorption spectrum of DNA were calculated by Spata et al.[201] using an electrostatic embedding QM/MM approach and by Gattuso et al.[202] using a QM/MM approach based on the Local Self Consistent Field (LSCF) method.[203] Also heavy atoms like actinides can be considered in this scheme.[204]

Additionally, as QM/MM calculations try to couple various levels of theory, see e.g., Ref 205. It might be necessary in some cases to manually adjust some force field parameters, as for instance done by Pentikäinen et al. for the QM/MM simulation nucleic acid bases,[206] and to carefully check the compatibility of the chosen methods.[207] For some applications it might even be the case that a 'traditional' QM/MM approach (i.e., involving two levels of description) is not sufficient in order to cover the entire length scale. Thus one might have to use additional levels of coarse graining and abstraction, together with a coupling between them, as has for instance been done by Lonsdale et al.[208]

To summarize, QM/MM approaches seem to give, at least qualitatively, very useful results, and the main source of error is rather due to a lack of physical correctness in the QM model than in the QM/MM partitioning.

The calculation of the partial density of states is another example intrinsically requiring a QM treatment, which we will demonstrate for the system depicted in Figure 3, showing a small fragment of DNA in a water-Na solution consisting in total of 15,613 atoms. The determination of the electronic structure is only possible using a QM method, but the influence of the environment on the DNA can also be modeled with a less expensive classical approach. In Figure 4, we compare the outcome of a full QM calculation with a static QM/MM approach, where all the solvent except for a small shell around the DNA has been replaced by a multipole expansion up to quadrupoles, leaving in total only 1877 atoms in the QM region. Both calculations were done with BIGDFT, and the multipoles were calculated as a post-processing of the full QM calculation. As can be seen from the plot, the two curves are virtually identical, but the QM/MM approach had to treat about 8 times fewer atoms on a QM level and was thus computationally considerably cheaper.

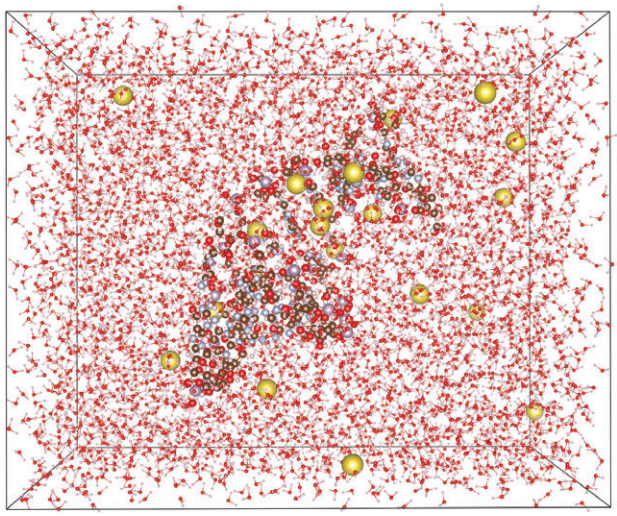

FIGURE 3 | Visualization[209] of a DNA fragment containing 11 base pairs, surrounded by a solvent of water and Na ions (giving in total 15,613 atoms), with periodic boundary conditions.

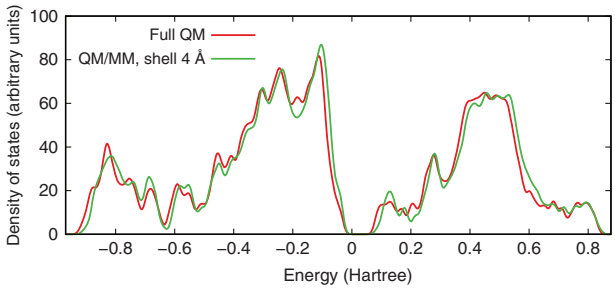

FIGURE 4 | Partial density of states for the DNA within the system depicted in Figure 3. The red curve was generated treating the entire system on a QM level, whereas the green curve only treated the DNA plus a shell of 4 Å on a QM level, with the remaining solvent atoms replaced by a multipole expansion. In order to allow for a better comparison, the QM/MM curve was shifted such that its HOMO energy coincides with the one of the full QM approach.

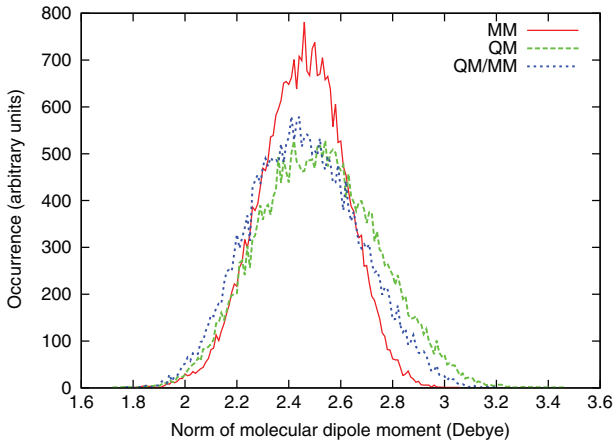

FIGURE 5 | Dispersion of the molecular dipole moment of water molecules within a droplet of 1800 atoms, with statistics taken over 50 snapshots of an MD simulation. The dipole is calculated based on the atomic monopoles and dipoles, and these were obtained from (a) a classical simulation using POLARIS(MD), (b) a DFT simulation using BIGDFT, and (c) a combined QM/MM approach.

Another field of application for QM methods is the parameterization of force fields. Many of the widely used force fields are fitted to reproduce experimental data of a certain test set of structures. However the outcome of this fitting procedure is not necessarily transferable to other compounds.[210] A more severe problem is the lack of applicability to different physicochemical conditions, such as pressure or temperature. This approach might thus lead to bad results when these force fields are applied to systems or conditions which are considerably different than the ones used for the parameterization. A possible solution is to parameterize a force field using results from *ab initio* calculations, which widens the range of possible applications. For instance, certain versions of the AMBER force field have been parameterized using atomic charges derived from *ab-initio* calculations, as for instance those described by Weiner et al.,[211] Cornell et al.[212] or Wang et al.;[213] for the last two, charges derived from the restrained electrostatic potential (RESP) approach[214] have been used. *Ab initio* results were also included—among experimental results—into the parameterization of the CHARMM22 force field.[215] There have also been attempts to develop force fields which determine the optimal set of parameters in an automatic way, using *ab initio* results as target data.[216] A logical continuation of this line uses statistical learning and big data analytics as envisioned in the European project NOMAD[217] which has the goal of using these techniques on top of a large computational material database.

Finally we also highlight the advantages of *large scale* QM simulations. Sometimes one is interested in atomistic characteristics averaged over a large number of samples, in this way generating the macroscopic behavior. An example is the dipole moment of liquid water, which is a macroscopic observable with a microscopic origin. In order to calculate it accurately it is not sufficient to simply compute the dipole moment of one water molecule in vacuum. Instead one has to take into account the polarization effects generated by the other surrounding water molecules. Owing to thermal fluctuations each molecule will however yield a different value, and the macroscopic observable result (keeping in mind that this can only be determined indirectly and is thus itself subject to fluctuations) can therefore only be obtained by averaging over all molecules, thereby requiring a truly large scale first principles simulation. The outcome of such a simulation, carried out using the MM code POLARIS(MD)[218] and the QM code BIGDFT is shown in Figure 5. Here we plot the dispersion of the molecular dipole moments, calculated based on atomic monopoles (i.e., atomic charges and dipoles) of a water droplet consisting of 600 molecules at ambient conditions and taking 50 snapshots of an MD simulation. As can be seen, there is a wide dispersion of the molecular dipole moments, which however yield a mean value in line with other theoretical and experimental studies.[219]

MULTISCALE LINKED TOGETHER: AN EXAMPLE

The above presented studies, linking together different models and length scales, are of course only a small set of representatives of the ongoing works in

the literature. The need for a connection of models in the common scale regimes is not only related to the QM/semiclassical regime. Such multi-method schemes can be applied also to larger length scales, up to sizes of interest for actual industrial applications. There is therefore a direct implication of large-scale QM methods on present-day *technological* challenges.

As an illustrative example we present the European project H2020 EXTMOS. The objective of this project is to build a model simulating organic light emitting diodes (OLEDs) in order to calculate their efficiency, where the only inputs are the organic molecule components. In the OLED realization process, acceptor and donor molecules, together with some dopant molecules, are mixed in a thin film; see Figure 6. It can be easily imagined that the investigation of such a process requires a multiscale approach with a coupling between different description levels. We will briefly describe them here.

Phase Organizations

Calculating the morphology of the organic film is the first step, which is done by means of molecular mechanics or MD at room temperature using an appropriate polarizable force field (PFF).[220] These PFFs are fitted with a charge analysis coming from DFT in order to reproduce the electrostatic potential. This step is crucial especially when considering different dopant molecules in the process. Systems of a few hundred atoms are simulated in QM, calculating the atomic forces and electrostatic potential[214] which are compared with those coming from PFFs. As soon as the PFF is fitted, morphologies of the organic film can be calculated and correct statistics of many thousands of atoms with their atomic positions can be easily generated even at different temperatures. Since the device will only operate within a limited temperature range—in particular only within one phase—the PFF parameters should be transferable without notable loss of accuracy.

Determination of the Electronic Properties

As soon as atomic configurations are determined, quantities of interest for the electronic properties of the molecules need to be extracted. Other QM methods coming from many-body perturbation theory (MBPT) such as GW[221] and Bethe-Salpeter methods[222] can be used to calculate the intrinsic properties of the organic molecules.[223,224] The challenge is to use such methods within an environment, modeled by adequate electrostatic degrees of freedom to describe the morphology of the organic film.[225]

Hopping Integrals

The previous step permits the calculation of the charge transfer of a few organic molecules in a given embedding environment. Since configurational statistics are important to represent correctly an organic film, constrained DFT[226] is well suited to understanding the influence of the environmental degrees of freedom[227] as well as to impose the correct charge

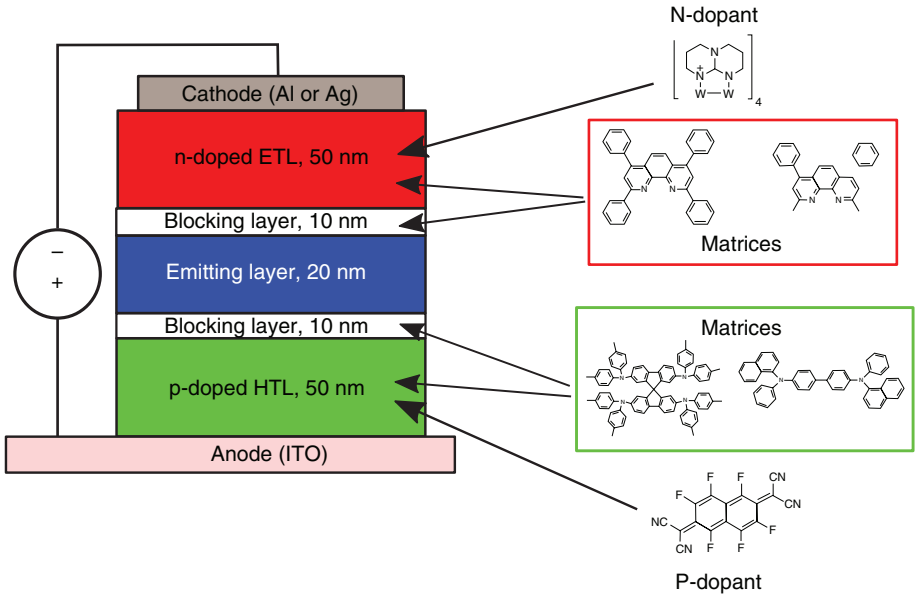

FIGURE 6 | Organic light-emitting diodes (OLED) device configuration illustrating the target goal of the EXTMOS project: simulating the full device from the molecular composition of the different layers.

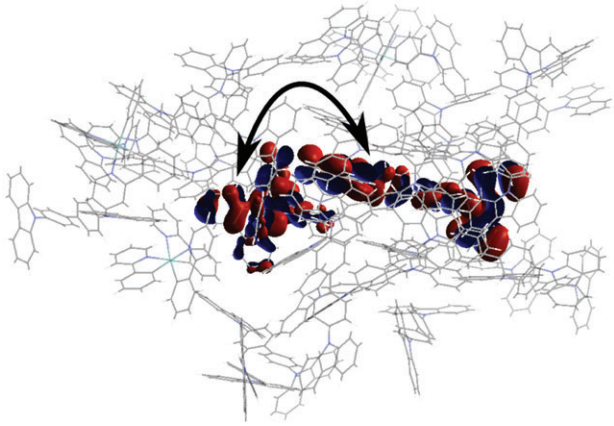

FIGURE 7 | Plot showing the HOMOs of two neighboring molecules calculated using a fragment approach. Their nearest neighbors extracted from a large disordered host-guest morphology are also depicted. Using this setup, one can calculate transfer integrals which take into account the environment.[228]

transfer and to calculate the statistics of hopping and site integrals[228] over an ensemble of molecules from the morphology (see Figure 7 for an illustration). Here again, another important quantity is the *dispersion* of the results provided by the morphologies. The QM fragment approach is well suited to calculate a set of hundreds of molecules in different orientations and environments.

Towards Device Simulation

The hopping and site integral parameters are finally used to calculate the efficiency[229] of the organic film using a Kinetic Monte Carlo method to predict charge and exciton transport processes through a random walk simulation. Transport parameters and device characteristics are deduced from the trajectories. Finally these parameters are included in a drift diffusion simulation in order to simulate larger region sizes and determine a circuit model.

In this example, the role of QM is important to determine correctly the film morphology and also the electronic properties. Nevertheless, QM needs to be used in collaboration with complementary methods: force fields for tractable MD and kinetic Monte Carlo methods to deal with larger systems.

CONCLUSION AND OUTLOOK

Advanced atomistic simulation techniques of many different flavors have found widespread applicability during the past years. Out of this plethora, we have seen the features of some QM codes that are now able to deal with systems with many thousands of atoms. Most of these techniques were invented more than a decade ago, however the approach to large-scale QM calculations is changing in the present day. We might even say that we are entering a 'second era' of DFT and, more generally, of QM methods in computational science.

On the one hand, the large research effort within the Quantum Chemistry and materials science communities is still ongoing with a focus on small scale systems, trying to achieve very high accuracy (e.g., novel exchange-correlation functionals, MBPT methods) and to improve the precision and reliability of the various codes and approaches.[230] However, as this ongoing work concentrates on small scale systems, the QM methods which are nowadays able to arrive at large scales rely on slightly more mature approaches and are thus forcibly less accurate than state-of-the-art QM methods. In other words, the fact that we have nowadays the ability to efficiently treat big systems does not mean that all problems at lower scale are solved.

On the other hand, we have seen a considerable effort of the community to enlarge the accessible length scales of QM simulations. These developments did not aim at developing new approaches to solve the fundamental QM equations, but rather tried to translate existing concepts into new domains. We have seen that this transition was driven by various aspects.

The first important point is related to the reliability of a calculation. One might raise the question whether a QM treatment is still appropriate above 'traditional' length scales: as already stated, a calculation which is more complex is not necessarily more accurate. But a QM approach is definitely less biased, leading therefore to considerably less arbitrariness. This is in strong contrast to established approaches such as force fields, where the output of a calculation depends strongly on the input of the calculation, for instance the chosen parameterization. When possible, it is helpful and important to use QM approaches also for large systems, in order to get unbiased insights into the effects of *realistic* experimental conditions on the values of interest, thereby yielding a deeper understanding of fundamental descriptions and trends. It is therefore important to have the possibility to extend already established QM models to large sizes, in order to have an idea of the effects of such realistic conditions. This leads to a *statistical* approach to large-scale calculations.

These considerations come at hand with the obvious observation that we *have* to abandon the QM treatment above a given length scale where a quantum description will be unnecessary. We used

on purpose 'unnecessary' instead of 'impossible' or 'not affordable'; at large scale, a QM calculation is justified *only* if there is the *need* to perform it. There will be no point in obtaining, with a QM treatment, results that could have been obtained with a more compact description like Force Fields or Coarse-Grained Models, unless these need to be validated first. This means that we have to provide strategies to couple the QM description with the modeling methods above this maximum length scale. In other terms, we must be able to provide, eventually, a *reduction* of the complexity of the description, implying that a good QM method at large scale has to provide different levels of theory and precision that can be linked to mesoscopic scales (e.g., atomic charges, Hamiltonian matrix elements, basis set multipoles, second principles).[231] We have presented in *Multiscale Linked Together: An Example* section one example where such a multi-method approach, completed with a modern QM treatment for the electronic excitations, can lead to results with potential technological implications.

This is also important in those cases where the QM level of theory alone is not able to correctly describe the properties of the system and must be complemented with other approaches. Thus, the large-scale QM methods described above are important to 'bridge' the length scale gap with non-QM methods; only if we can perform QM and post-QM approaches for systems with the same size, are we able to see if the trends—if not the actual quantities—are similar, in this way validating the respective levels of theory. A fundamental aspect for this task is the *systematicity* of the investigation. The ability to refine coarse-grained results at a QM level would help at least to *identify* if a refinement of the description might provide different trends. With respect to this task, the diversity of available QM approaches is thus essential.

The approach to large-scale QM calculations is not a mere question of a 'good software'; rather, it represents an opportunity to work in connection with different sensibilities. This will help in establishing a cross-disciplinary community, working at large scales and connecting together researchers with different sensibilities working with different computational methods and know-how. This point will be beneficial in both directions. Specialists of QM methods will learn to deal with the typical problems related to simulations at the million atom scale, taking advantage of the large experience acquired through the well established classical approaches over the past decades. For people with a background in classical approaches, tight collaborations with the electronic structure community will offer access to quantities and descriptions that are out of reach without the sensibility and experience of researchers working in QM methods.

Owing to all these reasons the field of large scale QM calculations might attract much attention during the forthcoming years. The topic presents big challenges, but offers even greater opportunities.

ACKNOWLEDGMENTS

We would like to thank Modesto Orozco and Hansel Gómez for fruitful discussions and Fátima Lucas for providing various test systems and helping with some visualizations. This work was supported by the EXTMOS project, grant agreement number 646176, and the Energy oriented Centre of Excellence (EoCoE), grant agreement number 676629, funded both within the Horizon2020 framework of the European Union. This research used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-06CH11357.

REFERENCES

1. Coulson CA. Present state of molecular structure calculations. *Rev Mod Phys* 1960, 32:170.
2. Hohenberg P, Kohn W. Inhomogeneous electron gas. *Phys Rev* 1964, 136:B864.
3. Kohn W, Sham LJ. Self-consistent equations including exchange and correlation effects. *Phys Rev A* 1965, 140:1133.
4. Zickermann V, Wirth C, Nasiri H, Siegmund K, Schwalbe H, Hunte C, Brandt U. Mechanistic insight from the crystal structure of mitochondrial complex I. *Science* 2015, 5:4.
5. Perdew JP, Schmidt K. Jacob's ladder of density functional approximations for the exchange-correlation energy. In: *AIP Conference Proceedings*, 577, 1, 2001.
6. Kohn W. Density-functional theory for systems of very many atoms. *Int J Quantum Chem* 1995, 56:229.
7. Severo Pereira Gomes A, Jacob CR. Quantum-chemical embedding methods for treating local electronic

excitations in complex chemical systems. *Annu Rep Prog Chem C Phys Chem* 2012, 108:222.

8. Galli G. Linear scaling methods for electronic structure calculations and quantum molecular dynamics simulations. *Curr Opin Solid State Mater Sci* 1996, 1:864.
9. Goedecker S. Linear scaling electronic structure methods. *Rev Mod Phys* 1999, 71:1085.
10. Bowler DR, Miyazaki T, Gillan MJ. Recent progress in linear scaling ab initio electronic structure techniques. *J Phys Condens Matter* 2002, 14:2781.
11. Bowler DR, Miyazaki T. O(N) methods in electronic structure calculations. Reports on progress in physics. *Phys Soc (Great Britain)* 2012, 75:036305.
12. Cloizeaux JD. Energy bands and projection operators in a crystal: analytic and asymptotic properties. *Phys Rev* 1964, 135:A685.
13. Cloizeaux JD. Analytical properties of n-dimensional energy bands and Wannier functions. *Phys Rev* 1964, 135:A698.
14. Kohn W. Analytic properties of bloch waves and Wannier functions. *Phys Rev* 1959, 115:809.
15. Baer R, Head-Gordon M. Sparsity of the density matrix in Kohn-sham density functional theory and an assessment of linear system-size scaling methods. *Phys Rev Lett* 1997, 79:3962.
16. Ismail-Beigi S, Arias TA. Locality of the density matrix in metals, semiconductors, and insulators. *Phys Rev Lett* 1999, 82:2127.
17. Goedecker S. Decay properties of the finite-temperature density matrix in metals. *Phys Rev B* 1998, 58:3501.
18. He L, Vanderbilt D. Exponential decay properties of Wannier functions and related quantities. *Phys Rev Lett* 2001, 86:5341.
19. March N, Young W, Sampanthar S. *The Many-Body Problem in Quantum Mechanics*. New York: Dover Publications, Incorporated; 1967.
20. Kohn W. Density functional and density matrix method scaling linearly with the number of atoms. *Phys Rev Lett* 1996, 76:3168.
21. Skylaris C-K, Haynes PD, Mostofi AA, Payne MC. Introducing ONETEP: linear-scaling density functional simulations on parallel computers. *J Chem Phys* 2005, 122:84119.
22. Bowler DR, Miyazaki T. Calculations for millions of atoms with density functional theory: linear scaling shows its potential. *J Phys Condens Matter* 2010, 22:074207.
23. VandeVondele J, Krack M, Mohamed F, Parrinello M, Chassaing T, Hutter J. QUICKSTEP: fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput Phys Commun* 2005, 167:103.
24. Mohr S, Ratcliff LE, Boulanger P, Genovese L, Caliste D, Deutsch T, Goedecker S. Daubechies wavelets for linear scaling density functional theory. *J Chem Phys* 2014, 140:204110.
25. Mohr S, Ratcliff LE, Genovese L, Caliste D, Boulanger P, Goedecker S, Deutsch T. Accurate and efficient linear scaling DFT calculations with universal applicability. *Phys Chem Chem Phys* 2015, 17:31360.
26. Yang W. Direct calculation of electron density in density-functional theory. *Phys Rev Lett* 1991, 66:1438.
27. Baroni S, Giannozzi P. Towards very large-scale electronic-structure calculations. *Europhys Lett* 1992, 17:547.
28. Drabold D, Sankey O. Maximum entropy approach for linear scaling in the electronic structure problem. *Phys Rev Lett* 1993, 70:3631.
29. Goedecker S, Colombo L. Efficient linear scaling algorithm for tight-binding molecular dynamics. *Phys Rev Lett* 1994, 73:122.
30. Goedecker S, Teter M. Tight-binding electronic structure calculations and tight-binding molecular dynamics with localized orbitals. *Phys Rev B* 1995, 51:9455.
31. Li X-P, Nunes RW, Vanderbilt D. Density-matrix electronic-structure method with linear system-size scaling. *Phys Rev B* 1993, 47:10891.
32. Nunes RW, Vanderbilt D. Generalization of the density-matrix method to a nonorthogonal basis. *Phys Rev B* 1994, 50:17611.
33. McWeeny R. Some recent advances in density matrix theory. *Rev Mod Phys* 1960, 32:335.
34. Artacho E, Sánchez-Portal D, Ordejón P, García A, Soler JM. Linear-scaling ab-initio calculations for large and complex systems. *Phys Status Solidi B* 1999, 215:809.
35. Soler JM, Artacho E, Gale JD, García A, Junquera J, Ordejón P, Sánchez-Portal D. The SIESTA method for ab initio order-N materials simulation. *J Phys Condens Matter* 2002, 14:2745.
36. ICMAB. Available at: http://departments.icmab.es/leem/siesta/. (Accessed October 17, 2016).
37. Kim J, Mauri F, Galli G. Total-energy global optimizations using nonorthogonal localized orbitals. *Phys Rev B* 1995, 52:1640.
38. Cankurtaran BO, Gale JD, Ford MJ. First principles calculations using density matrix divide and- conquer within the SIESTA methodology. *J Phys Condens Matter* 2008, 20:294208.
39. De Pablo PJ, Moreno-Herrero F, Colchero J, Gómez Herrero J, Herrero P, Baró AM, Ordejón P, Soler JM, Artacho E. Absence of dc-conductivity in λ-DNA. *Phys Rev Lett* 2000, 85:4992.

40. Heady L, Fernandez-Serra M, Mancera RL, Joyce S, Venkitaraman AR, Artacho E, Skylaris C-K, Ciacchi LC, Payne MC. Novel structural features of CDK inhibition revealed by an ab initio computational method combined with dynamic simulations. *J Med Chem* 2006, 49:5141.
41. Lin L, García A, Huhs G, Yang C. SIESTA-PEXSI: massively parallel method for efficient and accurate ab initio materials simulation without matrix diagonalization. *J Phys Condens Matter* 2014, 26:305503.
42. Lin L, Chen M, Yang C, He L. Accelerating atomic orbital-based electronic structure calculation via pole expansion and selected inversion. *J Phys Condens Matter* 2013, 25:295501.
43. Hu W, Lin L, Yang C, Yang J. Electronic structure and aromaticity of large-scale hexagonal graphene nanoflakes. *J Chem Phys* 2014, 141:214704.
44. Stokbro K, Taylor J, Brandbyge M, Ordejón P. TranSIESTA: a spice for molecular electronics. *Ann N Y Acad Sci* 2003, 1006:212.
45. Sanz-Navarro CF, Grima R, García A, Bea EA, Soba A, JM, Ordejón P. An efficient implementation of a QM-MM method in SIESTA. *Theor Chem Acc* 2011, 128:825.
46. Skylaris C-K, Haynes PD, Mostofi AA, Payne MC. Recent progress in linear-scaling density functional calculations with plane waves and pseudopotentials: the ONETEP code. *J Phys Condes Matter* 2008, 20:64209.
47. Hine N, Haynes P, Mostofi A, Skylaris C-K, Payne M. Linear-scaling density-functional theory with tens of thousands of atoms: expanding the scope and scale of calculations with ONETEP. *Comput Phys Commun* 2009, 180:1041.
48. http://www.onetep.org. (Accessed October 17, 2016).
49. Haynes PD, Skylaris C-K, Mostofi AA, Payne MC. Density kernel optimization in the ONETEP code. *J Phys Condens Matter* 2008, 20:294207.
50. O'Regan DD, Hine NDM, Payne MC, Mostofi AA. Linear-scaling DFT+U with full local orbital optimization. *Phys Rev B* 2012, 85:085107.
51. Ratcliff LE, Hine NDM, Haynes PD. Calculating optical absorption spectra for large systems using linear-scaling density functional theory. *Phys Rev B* 2011, 84:165131.
52. Zuehlsdorff TJ, Hine NDM, Spencer JS, Harrison NM, Riley DJ, Haynes PD. Linear-scaling time-dependent density-functional theory in the linear response formalism. *J Chem Phys* 2013, 139:064104.
53. Zuehlsdorff TJ, Hine NDM, Payne MC, Haynes PD. Linear-scaling time-dependent density-functional theory beyond the Tamm-Dancoff approximation: obtaining efficiency and accuracy with in situ optimised local orbitals. *J Chem Phys* 2015, 143:204107.
54. Turban DHP, Teobaldi G, O'Regan DD, Hine NDM. Supercell convergence of charge-transfer energies in pentacene molecular crystals from constrained DFT, ArXiv e-prints, 2016, arXiv:1603.02174 [physics.chem-ph].
55. Bell RA, Dubois SMM, Payne MC, Mostofi AA. Electronic transport calculations in the ONETEP code: implementation and applications. *Comput Phys Commun* 2015, 193:78.
56. Lee LP, Cole DJ, Payne MC, Skylaris C-K. Natural bond orbital analysis in the ONETEP code: applications to large protein systems. *J Comput Chem* 2013, 34:429.
57. Dziedzic J, Helal HH, Skylaris C-K, Mostofi AA, Payne MC. Minimal parameter implicit solvent model for ab initio electronic-structure calculations. *Europhys Lett* 2011, 95:43001.
58. Ruiz-Serrano A, Skylaris C-K. A variational method for density functional theory calculations on metallic systems with thousands of atoms. *J Chem Phys* 2013, 139:054107.
59. Hine NDM, Haynes PD, Mostofi AA, Payne MC. Linear-scaling density-functional simulations of charged point defects in Al_2O_3 using hierarchical sparse matrix algebra. *J Chem Phys* 2010, 133:1.
60. Hine N, Robinson M, Haynes P, Skylaris C-K, Payne M, Mostofi A. Accurate ionic forces and geometry optimization in linear-scaling density-functional theory with local orbitals. *Phys Rev B* 2011, 83:195102.
61. Weber C, Cole DJ, O'Regan DD, Payne MC. Renormalization of myoglobin-ligand binding energetics by quantum many-body effects. *Proc Natl Acad Sci U S A* 2014, 111:5790.
62. Lever G, Cole DJ, Lonsdale R, Ranaghan KE, Wales DJ, Mulholland AJ, Skylaris C-K, Payne MC. Large-scale density functional theory transition state searching in enzymes. *J Phys Chem Lett* 2014, 5:3614.
63. Lever G, Cole DJ, Hine NDM, Haynes PD, Payne MC. Electrostatic considerations affecting the calculated HOMO–LUMO gap in protein molecules. *J Phys Condens Matter* 2013, 25:152101.
64. Ozaki T. O(N) Krylov-subspace method for large-scale ab initio electronic structure calculations. *Phys Rev B* 2006, 74:245101.
65. http://www.openmx-square.org. (Accessed October 17, 2016).
66. Han MJ, Ozaki T, Yu J. O(N) LDA+U electronic structure calculation method based on the nonorthogonal pseudoatomic orbital basis. *Phys Rev B* 2006, 73:045110.
67. Ozaki T, Nishio K, Kino H. Efficient implementation of the nonequilibrium Green function method for

electronic transport calculations. *Phys Rev B* 2010, 81:035116.

68. Ohwaki T, Otani M, Ozaki T. A method of orbital analysis for large-scale first-principles simulations. *J Chem Phys* 2014, 140:244105.
69. Blum V, Gehrke R, Hanke F, Havu P, Havu V, Ren X, Reuter K, Scheffler M. Ab initio molecular simulations with numeric atom-centered orbitals. *Comput Phys Commun* 2009, 180:2175.
70. https://aimsclub.fhi-berlin.mpg.de. (Accessed October 17, 2016).
71. Havu V, Blum V, Havu P, Scheffler M. Efficient O(N) integration for all-electron electronic structure calculation using numeric basis functions. *J Comput Phys* 2009, 228:8367.
72. Marek A, Blum V, Johanni R, Havu V, Lang B, Auckenthaler T, Heinecke A, Bungartz H-J, Lederer H. The ELPA library: scalable parallel eigenvalue solutions for electronic structure theory and computational science. *J Phys Condens Matter* 2014, 26:213201.
73. Berger D, Logsdail AJ, Oberhofer H, Farrow MR, Catlow CRA, Sherwood P, Sokol AA, Blum V, Reuter K. Embedded-cluster calculations in a numeric atomic orbital density-functional theory framework. *J Chem Phys* 2014, 141:024105. doi:10.1063/1.4885816, arXiv:arXiv:1404.2130v1.
74. Schubert F, Rossi M, Baldauf C, Pagel K, Warnke S, von Helden G, Filsinger F, Kupser P, Meijer G, Salwiczek M, et al. Exploring the conformational preferences of 20-residue peptides in isolation: Ac-Ala_{19}-Lys + H^+ vs. Ac-Lys-Ala_{19} + H^+ and the current reach of DFT. *Phys Chem Chem Phys* 2015, 17:7373.
75. Ren X, Rinke P, Blum V, Wieferink J, Tkatchenko A, Sanfilippo A, Reuter K, Scheffler M. Resolution-of-identity approach to Hartree-Fock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions, New Journal of Physics 14 (2012), 10.1088/1367-2630/14/5/053020, arXiv:1201.0655.
76. Bowler DR, Bush IJ, Gillan MJ. Practical methods for ab initio calculations on thousands of atoms. *Int J Quantum Chem* 2000, 77:831.
77. http://www.order-n.org (Accessed October 17, 2016).
78. Sena AMP, Miyazaki T, Bowler DR. Linear scaling constrained density functional theory in CONQUEST. *J Chem Theory Comput* 2011, 7:884.
79. Nakata A, Bowler DR, Miyazaki T. Efficient Calculations with Multisite Local Orbitals in a Large-Scale DFT Code CONQUEST. *J Chem Theory Comput* 2014, 10:4813.
80. Miyazaki T, Bowler DR, Gillan MJ, Ohno T. The energetics of hut-cluster self-assembly in Ge/Si(001) from linear-scaling DFT calculations. *J Physical Soc Japan* 2008, 77:123706.
81. Otsuka T, Miyazaki T, Ohno T, Bowler DR, Gillan M. Accuracy of order-N density-functional theory calculations on DNA systems using CONQUEST. *J Phys Condens Matter* 2008, 20:294201.
82. Arita M, Bowler DR, Miyazaki T. Stable and efficient linear scaling first-principles molecular dynamics for 1000+ atoms. *J Chem Theory Comput* 2014, 10:5419.
83. Niklasson AMN. Extended Born-Oppenheimer molecular dynamics. *Phys Rev Lett* 2008, 100:123004.
84. http://www.bigdft.org. (Accessed October 17, 2016).
85. Daubechies I. Ten lectures on wavelets. In: *CBMS-NSF Regional Conference Series in Applied Mathematics*, 61, SIAM, 1992.
86. Genovese L, Neelov A, Goedecker S, Deutsch T, Ghasemi SA, Willand A, Caliste D, Zilberberg O, Rayson M, Bergman A, et al. Daubechies wavelets as a basis set for density functional pseudopotential calculations. *J Chem Phys* 2008, 129:014109.
87. Willand A, Kvashnin YO, Genovese L, Vázquez-Mayagoitia Á, Deb AK, Sadeghi A, Deutsch T, S. Norm-conserving pseudopotentials with chemical accuracy compared to all-electron calculations. *J Chem Phys* 2013, 138:104109.
88. Natarajan B, Genovese L, Casida ME, Deutsch T, Burchak ON, Philouze C, Balakirev MY. Wavelet-based linear-response time-dependent densityfunctional theory. *Chem Phys* 2012, 402:29.
89. Ratcliff LE, Genovese L, Mohr S, Deutsch T. Fragment approach to constrained density functional theory calculations using Daubechies wavelets. *J Chem Phys* 2015, 142:234105.
90. Genovese L, Deutsch T, Neelov A, Goedecker S, Beylkin G. Efficient solution of Poisson's equation with free boundary conditions. *J Chem Phys* 2006, 125:074105.
91. Genovese L, Deutsch T, Goedecker S. Efficient and accurate three-dimensional Poisson solver for surface problems. *J Chem Phys* 2007, 127:054704.
92. Cerioni A, Genovese L, Mirone A, Sole VA. Efficient and accurate solver of the three-dimensional screened and unscreened Poissons equation with generic boundary conditions. *J Chem Phys* 2012, 137:134108.
93. Fisicaro G, Genovese L, Andreussi O, Marzari N, Goedecker S. A generalized Poisson and Poisson-Boltzmann solver for electrostatic environments. *J Chem Phys* 2016, 143:014103.
94. Genovese L, Ospici M, Deutsch T, Méhaut J-F, Neelov A, Goedecker S. Density functional theory calculation on many-cores hybrid central processing unit-graphic processing unit architectures. *J Chem Phys* 2009, 131:034103.

95. Rudberg E, Rubensson EH, Sałek P. Kohn–Sham density functional theory electronic structure calculations with linearly scaling computational time and memory usage. *J Chem Theory Comput* 2011, 7:340.
96. http://ergoscf.org/. (Accessed October 17, 2016).
97. Rudberg E. Difficulties in applying pure Kohn-Sham density functional theory electronic structure methods to protein molecules. *J Phys Condens Matter* 2012, 24:072202.
98. Bock N, Challacombe M, Gan CK, Henkelman G, Nemeth K, Niklasson AMN, Odell A, Schwegler E, Tymczak CJ, Weber V. FreeON, Los Alamos National Laboratory, 2012, http://www.freeon.org.
99. Jordan DK, Mazziotti DA. Comparison of two genres for linear scaling in density functional theory: purification and density matrix minimization methods. *J Chem Phys* 2005, 122:084114.
100. https://www.cp2k.org/quickstep. (Accessed October 17, 2016).
101. Del Ben M, Hutter J, Vandevondele J. Forces and stress in second order Møller-Plesset perturbation theory for condensed phase systems within the resolution of identity Gaussian and plane waves approach. *J Chem Phys* 2015, 143:102803.
102. Tsuchida E, Tsukada M. Large-scale electronic structure calculations based on the adaptive finite element method. *J Physical Soc Japan* 1998, 67:3844.
103. Tsuchida E. Augmented orbital minimization method for linear scaling electronic structure calculations. *J Physical Soc Japan* 2007, 76:034708.
104. Tsuchida E. Ab initio molecular dynamics simulations with linear scaling: application to liquid ethanol. *J Phys Condens Matter* 2008, 20:294212.
105. Ikeshoji T, Tsuchida E, Morishita T, Ikeda K, Matsuo M, Kawazoe Y, Orimo S-i. Fast-ionic conductivity of Li+ in LiBH4. *Phys Rev B* 2011, 83:144301.
106. http://rmgdft.sourceforge.net/. (Accessed October 17, 2016).
107. Fattebert J-L, Bernholc J. Towards grid-based O(N) density-functional theory methods: optimized nonorthogonal orbitals and multigrid acceleration. *Phys Rev B* 2000, 62:1713.
108. Fattebert J, Hornung R, Wissink A. Finite element approach for density functional theory calculations on locally-refined meshes. *J Comput Phys* 2007, 223:759.
109. Fattebert J-L, Gygi F. Linear scaling first-principles molecular dynamics with controlled accuracy. *Comput Phys Commun* 2004, 162:24.
110. Osei-Kuffuor D, Fattebert J-L. Accurate and scalable N algorithm for first-principles molecular-dynamics computations on large parallel compute. *Phys Rev Lett* 2014, 112:046401.
111. Wang L-W, Teter MP. Kinetic-energy functional of the electron density. *Phys Rev B* 1992, 45:13196.
112. García-Aldea D, Alvarellos JE. Kinetic energy density study of some representative semilocal kinetic energy functionals. *J Chem Phys* 2007, 127:144109.
113. Huang C, Carter EA. Nonlocal orbital-free kinetic energy density functional for semiconductors. *Phys Rev B* 2010, 81:45206.
114. Ho GS, Lignères VL, Carter EA. Introducing PROFESS: a new program for orbital-free density functional theory calculations. *Comput Phys Commun* 2008, 179:839.
115. Hung L, Huang C, Shin I, Ho GS, Lignères VL, Carter EA. Introducing PROFESS 2.0: a parallelized, fully linear scaling program for orbital-free density functional theory calculations. *Comput Phys Commun* 2010, 181:2208.
116. Chen M, Xia J, Huang C, Dieterich JM, Hung L, Shin I, Carter EA. Introducing PROFESS 3.0: an advanced program for orbital-free density functional theory molecular dynamics simulations. *Comput Phys Commun* 2015, 190:228.
117. https://carter.princeton.edu/research/software/. (Accessed October 17, 2016).
118. Shin I, Carter EA. Enhanced von Weizsäcker wang-govind-carter kinetic energy density functional for semiconductors. *J Chem Phys* 2014, 140:18A531.
119. Huang C, Carter EA. Toward an orbital-free density functional theory of transition metals based on an electron density decomposition. *Phys Rev B* 2012, 85:045126.
120. Hung L, Carter EA. Accurate simulations of metals at the mesoscale: explicit treatment of 1 million atoms with quantum mechanics. *Chem Phys Lett* 2009, 475:163.
121. Chen M, Hung L, Huang C, Xia J, Carter EA. The melting point of lithium: an orbital-free firstprinciples molecular dynamics study. *Mol Phys* 2013, 111:3448.
122. Riplinger C, Pinski P, Becker U, Valeev EF, Neese F. Sparse maps—a systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory. *J Chem Phys* 2016, 144:024109.
123. Scemama A, Caffarel M, Oseret E, Jalby W. Quantum Monte Carlo for large chemical systems: implementing efficient strategies for petascale platforms and beyond. *J Comput Chem* 2013, 34:938.
124. Scemama A, Caffarel M, Oseret E, Jalby A. QMC=Chem: a quantum Monte Carlo Program for large-scale simulations in chemistry at the petascale level and beyond, high performance computing for computational science. *Vecpar* 2013, 2012:118.

125. Behler J, Parrinello M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys Rev Lett* 2007, 98:146401.

126. Behler J, Martoňák R, Donadio D, Parrinello M. Pressure-induced phase transitions in silicon studied by neural network-based metadynamics simulations. *Phys Status Solidi B* 2008, 245:2618.

127. Behler J, Martoňák R, Donadio D, Parrinello M. Metadynamics simulations of the high-pressure phases of silicon employing a high-dimensional neural network potential. *Phys Rev Lett* 2008, 100:185501.

128. Ghasemi SA, Hofstetter A, Saha S, Goedecker S. Interatomic potentials for ionic systems with density functional accuracy based on charge densities obtained by a neural network. *Phys Rev B* 2015, 92:045131.

129. Rupp M, Ramakrishnan R, von Lilienfeld OA. Machine learning for quantum mechanical properties of atoms in molecules. *J Phys Chem Lett* 2015, 6:3309.

130. Kitaura K, Ikeo E, Asada T, Nakano T, Uebayasi M. Fragment molecular orbital method: an approximate computational method for large molecules. *Chem Phys Lett* 1999, 313:701.

131. Fedorov DG, Kitaura K. Extending the power of quantum chemistry to large systems with the fragment molecular orbital method. *J Phys Chem A* 2007, 111:6904.

132. Fedorov DG, Kitaura K. The importance of three-body terms in the fragment molecular orbital method. *J Chem Phys* 2004, 120:6832.

133. Nakano T, Mochizuki Y, Yamashita K, Watanabe C, Fukuzawa K, Segawa K, Okiyama Y, Tsukamoto T, Tanaka S. Development of the four-body corrected fragment molecular orbital (fmo4) method. *Chem Phys Lett* 2012, 523:128.

134. Pruitt SR, Nakata H, Nagata T, Mayes M, Alexeev Y, Fletcher G, Fedorov DG, Kitaura K, Gordon MS. Importance of three-body interactions in molecular dynamics simulations of water demonstrated with the fragment molecular orbital method. *J Chem Theory Comput* 2016, 12:1423–1435.

135. Pruitt SR, Fedorov DG, Gordon MS. Geometry optimizations of open-shell systems with the fragment molecular orbital method. *J Phys Chem A* 2012, 116:4965.

136. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su S, et al. General atomic and molecular electronic structure system. *J Comput Chem* 1993, 14:1347.

137. Gordon MS, Schmidt MW. Chapter 41 – Advances in electronic structure theory: GAMESS a decade later. In: Scuseria CEDFSKE, ed. *Theory and Applications of Computational Chemistry*. Amsterdam: Elsevier; 2005, 1167–1189.

138. http://www.msg.ameslab.gov/gamess/. (Accessed October 17, 2016).

139. Nakano T, Mochizuki Y, Fukuzawa K, Amari S, Tanaka S. Chapter 2—developments and applications of abinit-mp software based on the fragment molecularorbital method. In: Starikov E, Lewis J, Tanaka S, eds. *Modern Methods for Theoretical Physical Chemistry of Biopolymers*. Amsterdam: Elsevier Science; 2006, 39–52.

140. http://moldb.nihs.go.jp/abinitmp/. (Accessed October 17, 2016).

141. Mochizuki Y, Yamashita K, Murase T, Nakano T, Fukuzawa K, Takematsu K, Watanabe H, Tanaka S. Large scale FMO-MP2 calculations on a massively parallel-vector computer. *Chem Phys Lett* 2008, 457:396.

142. Sekino H, Sengoku Y, Sugiki S, Kurita N. Molecular orbital analysis based on fragment molecular orbital scheme. *Chem Phys Lett* 2003, 378:589.

143. Ganesh V, Dongare RK, Balanarayan P, Gadre SR. Molecular tailoring approach for geometry optimization of large molecules: energy evaluation and parallelization strategies. *J Chem Phys* 2006, 125:104109.

144. Gordon MS, Mullin JM, Pruitt SR, Roskop LB, Slipchenko LV, Boatz JA. Accurate methods for large molecular systems. *J Phys Chem B* 2009, 113:9646.

145. Gordon MS, Fedorov DG, Pruitt SR, Slipchenko LV. Fragmentation methods: a route to accurate calculations on large systems. *Chem Rev* 2012, 112:632.

146. Pruitt SR, Bertoni C, Brorsen KR, Gordon MS. Efficient and accurate fragmentation methods. *Acc Chem Res* 2014, 47:2786.

147. Fletcher GD, Fedorov DG, Pruitt SR, Windus TL, Gordon MS. Large-scale MP2 calculations on the Blue Gene architecture using the fragment molecular orbital method. *J Chem Theory Comput* 2012, 8:75.

148. Fedorov DG, Alexeev Y, Kitaura K. Geometry optimization of the active site of a large system with the fragment molecular orbital method. *J Phys Chem Lett* 2011, 2:282.

149. Sawada T, Fedorov DG, Kitaura K. Role of the key mutation in the selective binding of avian and human influenza hemagglutinin to sialosides revealed by quantum-mechanical calculations. *J Am Chem Soc* 2010, 132:16862.

150. Sawada T, Fedorov DG, Kitaura K. Binding of influenza a virus hemagglutinin to the sialoside receptor is not controlled by the homotropic allosteric effect. *J Phys Chem B* 2010, 114:15700.

151. Fedorov DG, Kitaura K, Li H, Jensen JH, Gordon MS. The polarizable continuum model (PCM) interfaced with the fragment molecular orbital method (fmo). *J Comput Chem* 2006, 27:976.

152. Barone V, Cossi M, Tomasi J. A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. *J Chem Phys* 1997, 107:3210.

153. Ikegami T, Ishida T, Fedorov DG, Kitaura K, Inadomi Y, Umeda H, Yokokawa M, Sekiguchi S. Full electron calculation beyond 20,000 atoms: ground electronic state of photosynthetic proteins, supercomputing, 2005. In: *Proceedings of the ACM/IEEE SC 2005 Conference*, 10, 2005.

154. Fedorov DG, Jensen JH, Deka RC, Kitaura K. Covalent bond fragmentation suitable to describe solids in the fragment molecular orbital method. *J Phys Chem A* 2008, 112:11808.

155. Fedorov DG, Avramov PV, Jensen JH, Kitaura K. Analytic gradient for the adaptive frozen orbital bond detachment in the fragment molecular orbital method. *Chem Phys Lett* 2009, 477:169.

156. Fukunaga H, Fedorov DG, Chiba M, Nii K, Kitaura K. Theoretical analysis of the intermolecular interaction effects on the excitation energy of organic pigments: solid state quinacridone. *J Phys Chem A* 2008, 112:10887.

157. Elstner M, Porezag D, Jungnickel G, Elsner J, Haugk M, Frauenheim T, Suhai S, Seifert G. Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Phys Rev B* 1998, 58:7260.

158. Nishimoto Y, Fedorov DG, Irle S. Third-order density-functional tight-binding combined with the fragment molecular orbital method. *Chem Phys Lett* 2015, 636:90.

159. Wahiduzzaman M, Oliveira AF, Philipsen P, Zhechkov L, Van Lenthe E, Witek HA, Heine T. DFTB parameters for the periodic table: part 1, electronic structure. *J Chem Theory Comput* 2013, 9:4006.

160. Islam SM, Roy P-N. Performance of the SCCDFTB model for description of five-membered ring carbohydrate conformations: comparison to force fields, high-level electronic structure methods, and experiment. *J Chem Theory Comput* 2012, 8:2412.

161. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Struct Funct Bioinf* 2006, 65:712, arXiv:0605018 [q-bio].

162. Woods RJ, Dwek RA, Edge CJ, Fraser-Reid B. Molecular mechanical and molecular dynamic simulations of glycoproteins and oligosaccharides. 1. GLYCAM 93 parameter development. *J Phys Chem* 1995, 99:3832.

163. Case DA, Cheatham TE, Darden T, Gohlke H, Luo R, Merz KM, Onufriev A, Simmerling C, Wang B, Woods RJ. The Amber biomolecular simulation programs. *J Comput Chem* 2005, 26:1668, arXiv:NIHMS150003.

164. Choi TH, Liang R, Maupin CM, Voth GA. Application of the SCC-DFTB method to hydroxide water clusters and aqueous hydroxide solutions. *J Phys Chem B* 2013, 117:5165.

165. Nishimoto Y, Fedorov DG, Irle S. Density- functional tight-binding combined with the fragment molecular orbital method. *J Chem Theory Comput* 2014, 10:4801.

166. Nishimoto Y, Nakata H, Fedorov DG, Irle S. Large-scale quantum-mechanical molecular dynamics simulations using density-functional tight-binding combined with the fragment molecular orbital method. *J Phys Chem Lett* 2015, 6:5034.

167. Zen A, Luo Y, Mazzola G, Guidoni L, Sorella S. Ab initio molecular dynamics simulation of liquid water by quantum Monte Carlo. *J Chem Phys* 2015, 142:144111.

168. Vega C, Abascal JLF. Simulating water with rigid non-polarizable models: a general perspective. *Phys Chem Chem Phys* 2011, 13:19663.

169. Kiss PT, Baranyai A. Density maximum and polarizable models of water. *J Chem Phys* 2012, 137:084506.

170. Livshits GI, Stern A, Rotem D, Borovok N, Eidelshtein G, Migliore A, Penzo E, Wind SJ, Di Felice R, Skourtis SS, et al. Long-range charge transport in single G-quadruplex DNA molecules. *Nat Nanotechnol* 2014, 9:1040.

171. Lech CJ, Phan AT, Michel-Beyerle ME, Voityuk AA. Electron–hole transfer in G-quadruplexes with different tetrad stacking geometries: a combined QM and MD study. *J Phys Chem B* 2013, 117:9851.

172. Sponer J, Leszczynski J, Hobza P. Electronic properties, hydrogen bonding, stacking, and cation binding of DNA and RNA bases. *Biopolym (Nucl Acid Sci)* 2002, 61:3.

173. Gkionis K, Kruse H, Platts JA, Mládek A, Koča J, Šponer J. Ion binding to quadruplex DNA stems. Comparison of MM and QM descriptions reveals sizable polarization effects not included in contemporary simulations. *J Chem Theory Comput* 2014, 10:1326.

174. Dans PD, Walther J, Gómez H, Orozco M. Multiscale simulation of DNA. *Curr Opin Struct Biol* 2016, 37:29.

175. Gaines JC, Smith WW, Regan L, O'Hern CS. Random close packing in protein cores. *Phys Rev E* 2016, 93:032415.

176. Kiss G, Röthlisberger D, Baker D, Houk KN. Evaluation and ranking of enzyme designs. *Protein Sci* 2010, 19:1760.

177. Kiss G, Çelebi-Ölçüm N, Moretti R, Baker D, Houk KN. Computational enzyme design. *Angew Chem Int Ed* 2013, 52:5700.

178. Jacob CR, Neugebauer J. Subsystem density functional theory. *WIREs Comput Mol Sci* 2014, 4:325.

179. Bakowies D, Thiel W. Hybrid models for combined quantum mechanical and molecular mechanical approaches. *J Phys Chem* 1996, 100:10580.

180. Lin H, Truhlar DG. QM/MM: what have we learned, where are we, and where do we go from here? *Theor Chem Acc* 2007, 117:185.

181. Senn HM, Thiel W. QM/MM methods for biomolecular systems. *Angew Chem Int Ed* 2009, 48:1198.

182. Ferré N, Ángyán JG. Approximate electrostatic interaction operator for QM/MM calculations. *Chem Phys Lett* 2002, 356:331.

183. Neugebauer J. Subsystem-based theoretical spectroscopy of biomolecules and biomolecular assemblies. *ChemPhysChem* 2009, 10:3148.

184. Eichinger M, Tavan P, Hutter J, Parrinello M. A hybrid method for solutes in complex solvents: density functional theory combined with empirical force fields. *J Chem Phys* 1999, 110:10452.

185. Das D, Eurenius KP, Billings EM, Sherwood P, Chatfield DC, Hodošček M, Brooks BR. Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method. *J Chem Phys* 2002, 117:10534.

186. Biswas PK, Gogonea V. A regularized and renormalized electrostatic coupling Hamiltonian for hybrid quantum-mechanical-molecular-mechanical calculations. *J Chem Phys* 2005, 123:1.

187. Senthilkumar K, Mujika JI, Ranaghan KE, Manby FR, Mulholland AJ, Harvey JN. Analysis of polarization in QM/MM modelling of biologically relevant hydrogen bonds. *J R Soc Interface* 2008, 5-(Suppl 3):S207.

188. Zuehlsdorff TJ, Haynes PD, Hanke F, Payne MC, Hine NDM. Solvent effects on electronic excitations of an organic chromophore. *J Chem Theory Comput* 2016, 12:1853.

189. Claeyssens F, Harvey JN, Manby FR, Mata RA, Mulholland AJ, Ranaghan KE, Schütz M, Thiel S, Thiel W, Werner H-J. High-accuracy computation of reaction barriers in enzymes. *Angew Chem Int Ed* 2006, 45:6856.

190. Schütz M, Hetzer G, Werner H-J. Low-order scaling local electron correlation methods. I. Linear scaling local MP2. *J. Chem. Phys.* 1999, 111:5691.

191. Hetzer G, Schütz M, Stoll H, Werner H-J. Loworder scaling local correlation methods II: splitting the Coulomb operator in linear scaling local second-order Møller-Plesset perturbation theory. *J Chem Phys* 2000, 113:9443.

192. Schütz M. Low-order scaling local electron correlation methods. III. Linear scaling local perturbative triples correction (T). *J Chem Phys* 2000, 113:9986.

193. Schütz M, Werner H-J. Low-order scaling local electron correlation methods. IV. Linear scaling local coupled-cluster (LCCSD). *J Chem Phys* 2001, 114:661.

194. Schütz M. Low-order scaling local electron correlation methods. V. Connected triples beyond (T): linear scaling local CCSDT-1b. *J Chem Phys* 2002, 116:8772.

195. Schütz M, Werner H-J. Local perturbative triples correction (T) with linear cost scaling. *Chem Phys Lett* 2000, 318:370.

196. Werner HJ, Manby FR, Knowles PJ. Fast linear scaling second-order Moller-Plesset perturbation theory (MP2) using local and density fitting approximations. *J Chem Phys* 2003, 118:8149.

197. Mlýnský V, Banáš P, Šponer J, van der Kamp MW, Mulholland AJ, Otyepka M. Comparison of ab initio, DFT, and semiempirical QM/MM approaches for description of catalytic mechanism of hairpin ribozyme. *J Chem Theory Comput* 2014, 10:1608.

198. Manby FR, Stella M, Goodpaster JD, Miller TF. A simple, exact density-functional-theory embedding scheme. *J Chem Theory Comput* 2012, 8:2564.

199. Bennie SJ, van der Kamp MW, Pennifold RCR, Stella M, Manby FR, Mulholland AJ. A projector-embedding approach for multiscale coupled-cluster calculations applied to citrate synthase. *J Chem Theory Comput* 2016, 12:2689.

200. Lonsdale R, Harvey JN, Mulholland AJ. Effects of dispersion in density functional based quantum mechanical/molecular mechanical calculations on cytochrome P450 catalyzed reactions. *J Chem Theory Comput* 2012, 8:4637.

201. Spata VA, Matsika S. Role of excitonic coupling and charge-transfer States in the absorption and CD spectra of adenine-based oligonucleotides investigated through QM/MM simulations. *J Phys Chem A* 2014, 118:12021.

202. Gattuso H, Assfeld X, Monari A. Modeling DNA electronic circular dichroism by QM/MM methods and Frenkel Hamiltonian. *Theor Chem Acc* 2015, 134:1.

203. Monari A, Rivail J-L, Assfeld X. Theoretical modeling of large molecular systems. Advances in the local self consistent field method for mixed quantum mechanics/molecular mechanics calculations. *Acc Chem Res* 2013, 46:596.

204. Gomes ASP, Jacob CR, Real F, Visscher L, Vallet V. Towards systematically improvable models for actinides in condensed phase: the electronic spectrum of uranyl in Cs2UO2Cl4 as a test case. *Phys Chem Chem Phys* 2013, 15:15153.

205. Houriez C, Ferré N, Masella M, Siri D. Prediction of nitroxide hyperfine coupling constants in solution

from combined nanosecond scale simulations and quantum computations. *J Chem Phys* 2008, 128:244504. doi:10.1063/1.2939121.

206. Pentikäinen U, Shaw KE, Senthilkumar K, Woods CJ, Mulholland AJ. Lennard–Jones parameters for B3LYP/CHARMM27 QM/MM modeling of nucleic acid bases. *J Chem Theory Comput* 2009, 5:396.
207. Shaw KE, Woods CJ, Mulholland AJ. Compatibility of quantum chemical methods and empirical (MM) water models in quantum mechanics/molecular mechanics liquid water simulations. *J Phys Chem Lett* 2010, 1:219.
208. Lonsdale R, Rouse SL, Sansom MSP, Mulholland AJ. A multiscale approach to modelling drug metabolism by membrane-bound cytochrome P450 enzymes. *PLoS Comput Biol* 2014, 10:1.
209. Momma K, Izumi F. VESTA 3 for three-dimensional visualization of crystal, volumetric and morphology data. *J Appl Crystallogr* 2011, 44:1272.
210. Mackerell AD. Empirical force fields for biological macromolecules: overview and issues. *J Comput Chem* 2004, 25:1584.
211. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weinerl P. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J Am Chem Soc* 1984, 106:765.
212. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* 1995, 117:5179, arXiv:z0024.
213. Wang J, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules. *J Comput Chem* 2000, 21:1049.
214. Bayly CI, Cieplak P, Cornell W, Kollman PA. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem* 1993, 97:10269.
215. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 1998, 102:3586.
216. Huang L, Roux B. Automated force field parameterization for nonpolarizable and polarizable atomic models based on ab initio target data. *J Chem Theory Comput* 2013, 9:3543.
217. The H2020 Center of Excellence project NOMAD (http://www.nomad-coe.eu) takes its data from the online database (http://www.nomad-repository.eu).
218. Réal F, Vallet V, Flament JP, Masella M. Revisiting a many-body model for water based on a single polarizable site: from gas phase clusters to liquid and air/liquid water systems. *J Chem Phys* 2013, 139:114502.
219. Gubskaya AV, Kusalik PG. The total molecular dipole moment for liquid water. *J Chem Phys* 2002, 117:5290.
220. Baker CM. Polarizable force fields for molecular dynamics simulations of biomolecules. *WIREs Comput Mol Sci* 2015, 5:241.
221. Hedin L. New method for calculating the one-particle Green's function with application to the electron-gas problem. *Phys Rev* 1965, 139:A796.
222. Onida G, Reining L, Rubio A. Electronic excitations: density-functional versus many-body Green's function approaches. *Rev Mod Phys* 2002, 74:601.
223. Blase X, Attaccalite C, Olevano V. First-principles GW calculations for fullerenes, porphyrins, phthalocyanine, and other molecules of interest for organic photovoltaic applications. *Phys Rev B* 2011, 83:1.
224. Duchemin I, Deutsch T, Blase X. Short-range to long-range charge-transfer excitations in the zincbacteriochlorin-bacteriochlorin complex: a Bethe-Salpeter study. *Phys Rev Lett* 2012, 109:167801.
225. D'Avino G, Muccioli L, Zannoni C, Beljonne D, Soos ZGZ, D'Avino G, Muccioli L, Zannoni C, Beljonne D, Soos ZGZ. Electronic polarization in organic crystals: a comparative study of induced dipoles and intramolecular charge redistribution schemes. *J Chem Theory Comput* 2014, 10:4959.
226. Wu Q, Van Voorhis T. Extracting electron transfer coupling elements from constrained density functional theory. *J Chem Phys* 2006, 125:164105.
227. Schober C, Reuter K, Oberhofer H. Critical analysis of fragment-orbital DFT schemes for the calculation of electronic coupling values. *J Chem Phys* 2016, 144:054103.
228. Ratcliff LE, Grisanti L, Genovese L, Deutsch T, Neumann T, Danilov D, Wenzel W, Beljonne D, Cornil J. Toward fast and accurate evaluation of charge on-site energies and transfer integrals in supramolecular architectures using linear constrained density functional theory (CDFT)-based methods. *J Chem Theory Comput* 2015, 11:2077.
229. Cornil J, Verlaak S, Martinelli N, Mityashin A, Olivier Y, Van Regemorter T, D'Avino G, Muccioli L, Zannoni C, Castet F, et al. Exploring the energy landscape of the charge transport levels in organic semiconductors at the molecular scale. *Acc Chem Res* 2013, 46:434.
230. Lejaeghere K, Bihlmayer G, Björkman T, Blaha P, Blügel S, Blum V, Caliste D, Castelli IE, Clark SJ, Dal Corso A, et al. Reproducibility in density functional theory calculations of solids. *Science* 2016, 351:1415.
231. Wojdel JC, Junquera J. Second-principles method including electron and lattice degrees of freedom, Arxiv preprint, 1, 2015, arXiv:arXiv:1511.07675v1.

