Top Curr Chem (2012) 307: 1–42
DOI: 10.1007/128_2010_114

Published online: 18 January 2011

First-Principles-Based Multiscale, Multiparadigm Molecular Mechanics and Dynamics Methods for Describing Complex Chemical Processes

Andres Jaramillo-Botero, Robert Nielsen, Ravi Abrol, Julius Su, Tod Pascal, Jonathan Mueller and William A. Goddard III

Abstract We expect that systematic and seamless computational upscaling and downscaling for modeling, predicting, or optimizing material and system properties and behavior with atomistic resolution will eventually be sufficiently accurate and practical that it will transform the mode of development in the materials, chemical, catalysis, and Pharma industries. However, despite truly dramatic progress in methods, software, and hardware, this goal remains elusive, particularly for systems that exhibit inherently complex chemistry under normal or extreme conditions of temperature, pressure, radiation, and others. We describe here some of the significant progress towards solving these problems via a general multiscale, multiparadigm strategy based on first-principles quantum mechanics (QM), and the development of breakthrough methods for treating reaction processes, excited electronic states, and weak bonding effects on the conformational dynamics of large-scale molecular systems. These methods have resulted directly from filling in the physical and chemical gaps in existing theoretical and computational models, within the multiscale, multiparadigm strategy. To illustrate the procedure we demonstrate the application and transferability of such methods on an ample set of challenging problems that span multiple fields, system length- and timescales, and that lay beyond the realm of existing computational or, in some case, experimental approaches, including understanding the solvation effects on the reactivity of organic and organometallic structures, predicting transmembrane protein structures, understanding carbon nanotube nucleation and growth, understanding the effects of electronic excitations in materials subjected to extreme conditions of temperature and pressure, following the dynamics and energetics of long-term conformational evolution of DNA

A. Jaramillo-Botero (✉), R. Nielsen, R. Abrol, J. Su, T. Pascal, J. Mueller,
and W.A. Goddard III (✉)
Chemistry and Chemical Engineering, California Institute of Technology, Mail code 139-74, 1200 E California Blvd, Pasadena, CA 91125, USA
e-mail: ajaramil@wag.caltech.edu, wag@wag.caltech.edu

macromolecules, and predicting the long-term mechanisms involved in enhancing the mechanical response of polymer-based hydrogels.

Keywords Multiscale modeling, Nanotube growth, Non-adiabatic molecular dynamics, Organometallic structures, Protein structure prediction, Reactive molecular dynamics

Contents

1 First Principles-Based Multiscale, Multiparadigm Simulations

The computations required for accurate modeling and simulation of large-scale systems with atomistic resolution involve a hierarchy of levels of theory: quantum mechanics (QM) to determine the electronic states; force fields to average the electronics states and to obtain atom based forces (FF), molecular dynamics (MD) based on such an FF; mesoscale or coarse grain descriptions that average or homogenize atomic motions; and finally continuum level descriptions (see Fig. 1).

By basing computations on first principles QM it is possible to overcome the lack of experimental data to carry out accurate predictions with atomistic resolution, which would otherwise be impossible. Furthermore, QM provides the fundamental information required to describe quantum effects, electronically excited states, as well as reaction paths and barrier heights involved in chemical reactions processes. However, the practical scale for accurate QM today is $<$1,000 atoms per molecule or periodic cell (a length scale of a few nanometers) whereas the length scale for modeling supramolecular systems in biology may be in the tens of nanometers, while elucidating the interfacial effects between grains in composite materials may require hundreds of nanometers, and modeling turbulent fluid flows or shock-induced instabilities in multilayered materials may require micrometers. Thus, simulations of engineered materials and systems may require millions to billions of atoms, rendering QM methods impractical.

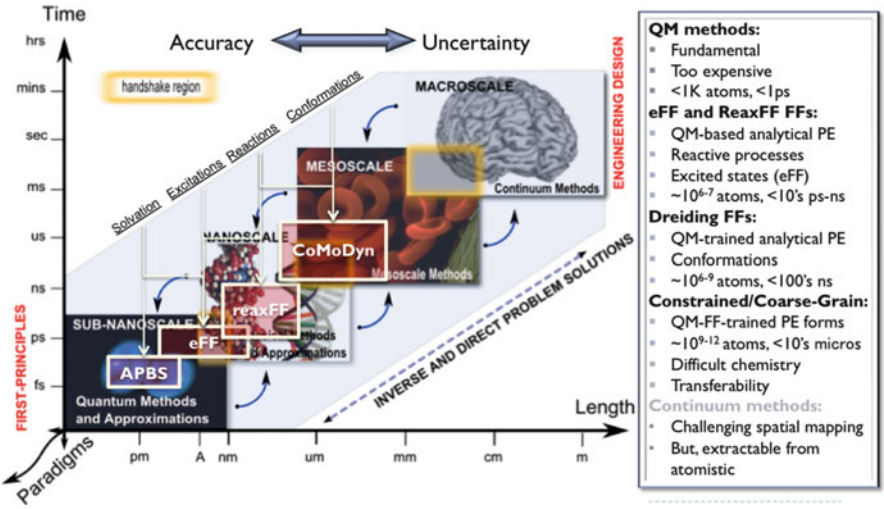

Fig. 1 Hierarchical multiscale, multiparadigm approach to materials modeling, from QM to the mesoscale, incorporating breakthrough methods to handle complex chemical processes (eFF, ReaxFF). Adapted from our multiscale group site http://www.wag.caltech.edu/multiscale

Nonetheless, QM methods are essential for accurately describing atomic-level composition, structure and energy states of materials, considering the influence of electronic degrees of freedom. By incorporating time-dependent information, the dynamics of a system under varying conditions may be explored from QM-derived forces, albeit within a limited timescale (<1 ps). The prominent challenge for theory and computation involves efficiently bridging, from QM first-principles, into larger length scales with predominantly heterogeneous spatial and density distributions, and longer timescales of simulation – enough to connect into engineering-level design variables – while retaining physicochemical accuracy and certainty. Equally challenging remains the inverse top-down engineering design problem, by which macroscopic material/process properties would be tunable from optimizing its atomic-level composition and structure. Our approach to this challenge has been to develop breakthrough methods to staple and extend hierarchically over existing ones, as well as to develop the necessary tools to enable continuous lateral (multi-paradigm) and hierarchical (multiscale) couplings, between the different theories and models as a function of their length- and timescale range – a strategy referred to here as *First-Principles-Based Multiscale-Multiparadigm Simulation*.

The ultimate goal is a reversible bottom-up, top-down approach, based on first principles QM, to characterize properties of materials and processes at a hierarchy of length and timescales. This will improve our ability to design, analyze, and interpret experimental results, perform model-based prediction of phenomena, and to control precisely the multi-scale nature of material systems for multiple applications. Such an approach is now enabling us to study problems once thought to be intractable, including reactive turbulent flows, composite material instabilities,

dynamics of warm-dense-matter and plasma formation, functional molecular biology, and protein structure prediction, among others.

In this chapter, we describe some of our progress in theory, methods, computational techniques, and tools towards first-principles-based multiscale, multiparadigm simulations, in particular, for systems that exhibit intricate chemical behavior. We map the document over the hierarchical framework depicted in Fig. 1, threading the description from QM up through mesoscale classical approximations, presenting significant and relevant example applications to different fields at each level.

2 The Role of QM in Multiscale Modeling

QM relies solely on information about the atomic structure and composition of matter to describe its behavior. Significant progress has been made in the development of QM theory and its application, since its birth in the 1920s. The following sections present an overview of some parts of this evolution, describing how it provides the foundations for our first-principles-based multiscale, multiparadigm strategy to materials modeling and simulation.

2.1 The Wave Equation for Matter

Circa 1900 Max Planck suggested that light was quantized, and soon after, in 1905, Albert Einstein interpreted Planck's quantum to be photons, particles of light, and proposed that the energy of a photon is proportional to its frequency. In 1924, Louis de Broglie argued that since light could be seen to behave under some conditions as particles [1] (e.g., Einstein's explanation of the photoelectric effect) and at other times as waves (e.g., diffraction of light), one could also consider that matter has the same ambiguity of possessing both particle and wave properties. Starting with de Broglie's idea that particles behave as waves and the fundamental (Hamilton's) equations of motion (EOM) from classical mechanics, Erwin Schrödinger [2] developed the electronic wave equation that describes the space- (and time-) dependence of quantum mechanical systems [3], for an n-particle system as

$$\left[-\hbar^2 \sum_{i=1}^{n} \frac{\nabla^2}{2m_i} + V(r_1, r_2 ... r_n, t)\right] \psi(r_1, r_2 ... r_n, t) = i\hbar \frac{\partial}{\partial t} \psi(r_1, r_2 ... r_n, t), \quad (1)$$

where the term in brackets corresponds to a linear operator that involves the kinetic (first term) and potential (second term, V) energy operators that act over the systems' wavefunction, Ψ, and the right-hand side the quantized energy operator, corresponding to the full energy of the system, acting on the same wavefunction.

The wavefunction is interpreted as the probability amplitude for different configurations, r, of the system at different times, i.e., it describes the dynamics of the n-particles as a function of space, r, and time, t. In more abstract terms, (1) may also be written as

$$\widehat{H}\psi = E\psi, \tag{2}$$

and take several different forms, depending on the physical situation.

In principle, all properties of all materials, with known atomic structure and composition, can be accurately described using (1) and one could then replace existing empirical methods used to model materials properties by a first principles or de novo computational approach design of materials and devices. Unfortunately, direct first principles applications of QM is highly impractical with current methods, mainly due to the computational complexity of solving (1) in three dimensions for a large number of particles, i.e., for systems relevant to the materials designer, with a gap of ~10^{20}!

There are numerous approaches to approximate solutions for (1), most of which involve finding the system's total ground state energy, E, including methods that treat the many-body wavefunction as an antisymmetric function of one-body orbitals (discussed in later sections), or methods that allow a direct representation of many-body effects in the wave function such as Quantum Monte Carlo (QMC), or hybrid methods such as coupled cluster (CC), which adds multi-electron wavefunction corrections to account for the many-body (electron) correlations.

QMC can, in principle, provide energies to within chemical accuracy (≈2 kcal/mol) [4] and its computational expense scales with system size as O(N^3) or better [5, 6], albeit with a large prefactor, while CC tends to scale inefficiently with the size of the system, generally O(N^6 to N!) [7].

Nevertheless, we have shown how QMC performance can be significantly improved using short equilibration schemes that effectively avoid configurations that are not representative of the desired density [8], and through efficient data parallelization schemes amenable to GPU processing [9]. Furthermore, in [10] we also showed how QMC can be used to obtain high quality energy differences, from generalized valence bond (GVB) wave functions, for an intuitive approach to capturing the important sources of static electronic correlation. Part of our current drive involves using the enhanced QMC methods to obtain improved functionals for Density Functional Theory (DFT) calculations, in order to enhance the scalability and quality of solutions to (1).

But for the sake of brevity, we will focus here on methods and applications that are unique for integrating multiple paradigms and spanning multiple length- and timescales, while retaining chemical accuracy, i.e., beyond direct use of conventional QM approaches. The following section describes the general path to classical approximations to (1), in particular to interatomic force fields and conventional MD, which sacrifice electronic contributions that drive critical chemical properties, and our departure from conventionalism to recover the missing physicochemical details.

2.2 Approximations to Schrödinger's Equation

A number of simplifications to Schrödinger's equation are commonly made to ease the computational costs; some of these are reviewed below in order to explain the nature of our methods.

2.2.1 Adiabatic Approximation (Treat Electrons Separately from the Nuclei)

An important approximation is to factor the total wavefunction in terms of an electronic wavefunction, which depends parametrically on the stationary nuclear positions, and a nuclear wavefunction, as

$$\psi_{\text{total}} = \psi^{\text{electronic}} \times \psi^{\text{nuclear}}. \tag{3}$$

This is also known as the Born–Oppenheimer [11] approximation. The underlying assumption is that since nuclei are much heavier than electrons (e.g., the proton to electron mass ratio is ~1836.153), they will also move in a much lower timescale. For a set of fixed nuclear positions, (1) is used to solve for the corresponding electronic wavefunction and electronic energies (typically in their lowest or ground-state). A sufficient set of electronic solutions, at different nuclear positions, leads to the systems' nuclei-only dependent Potential Energy Surface (PES). Modern codes can also lead directly to the inter-atomic forces, from the negative gradient of the potential energies, required for understanding the dynamics of systems.

Methods for solving the electronic equation (1) have evolved into sophisticated codes that incorporate a hierarchy of approximations that can be used as "black boxes" to achieve accurate descriptions for the PES for ground states of molecular systems. Popular codes include Gaussian [12], GAMESS [13], and Jaguar [14] for finite molecules and VASP [15], CRYSTAL [16], CASTEP [17], and Sequest [18] for periodic systems.

The simplest wavefunction involves a product of one-particle functions, or spin-orbitals, antisymmetrized to form a (Slater) determinant that satisfies the Pauli (exclusion) principle, i.e., two electrons with the same spin orbital result in no wavefunction. Optimizing these spin-dependent orbitals leads to the Hartree–Fock (HF) method, with the optimum orbitals described as molecular orbitals (MO). HF is excellent for ground state geometries and good for vibrational frequencies, but its neglect of electron correlation [19] leads to problems in describing bond breaking and chemical reactions. In addition, it cannot account for the London dispersion forces responsible for van der Waals attraction of molecular complexes. A hierarchy of methods has been developed to improve the accuracy of HF. Some of the popular methods include second-order Moller–Plesset perturbation theory (MP2) [20], CC with multiple perturbative excitations, multireference selfconsistent field (MC-SCF), and multireference configuration interaction (MR-CI) [21] methods

(see [22] for a recent review). A form of MC-SCF useful for interpreting electron correlation and bonding is the GVB method, [23–25] which leads to the best description in which every orbital is optimized for a single electron. These are referred to as ab initio methods as they are based directly on solving (1), without any empirical data. Many methods, which rely on empirical data to obtain approximate descriptions for systems too large for ab initio methods, have also been proved useful. [26]

A non-empirical alternative to ab initio methods that now provides the best compromise between accuracy and cost for solving Schrödinger's equation of large molecules is DFT. The original concept was the demonstration by Hohenberg and Kohn [27] that the ground state properties of a many-electron system are uniquely determined by the density, ρ, as a function of nuclear coordinates, r, and hence all the properties of a (molecular) system can be deduced from a functional of $\rho(r)$, i.e.,

$$E = \varepsilon[\rho(r)]. \tag{4}$$

DFT has evolved dramatically over the years, with key innovations including the formulation of the Kohn–Sham equations [28] to develop a practical one-particle approach, while imposing the Pauli principle, the Local Density Approximation (LDA) based on the exact solution of the correlation energy of the uniform electron gas, the generalized gradient approximation (GGA) to correct for the gradients in the density for real molecules, incorporating exact exchange into the DFT. This has led to methods such as B3LYP and X3LYP that provide accurate energies (~3 kcal/mol) and geometries [29] for solids, liquids, and large molecules [30, 31]. Although generally providing high accuracy, there is no prescription for improving DFT when it occasionally leads to large errors. Even so, it remains the method of choice for electronic structure calculations in chemistry and solid-state physics. We recently demonstrated improved accuracy in DFT by introducing a universal damping function to correct empirically the lack of dispersion [32].

An important area of application for QM methods has been determining and describing reaction pathways, energetics, and transition states for reaction processes between small species. QM-derived first and second derivatives of energy calculated at stable and saddle points on PES can be used under statistical mechanics formulations [33, 34] to yield enthalpies and free energies of structures in order to determine their reactivity. Transition state theory and idealized thermodynamic relationships (e.g., $\Delta G[P_0 \rightarrow P] = kT\ln[P/P_0]$) allow temperature and pressure regimes to be spanned when addressing simple gas phase and gas-surface interactions.

On the other hand, many applications involve interactions between solutes and solvent, which utterly distinguish the condensed phase from in vacuo, free energy surfaces. To tackle this challenge, we describe below a unique multiparadigm strategy to incorporate the effects of a solvent when using QM methods to determine reactivity in organic and organometallic systems.

Application Example: Solvent and pH Effects on Reactivity

Interactions critical to the rate and selectivity of reactions include the relaxation of a wavefunction or zwitter-ionic geometry in response to a polarizable solvent, hydrogen bonding, and reversible proton transfer. It is necessary in these cases to introduce solvation effects explicitly through the inclusion of solvent molecules, and/or implicitly through a continuum representation of the medium. Adding explicit solvent molecules increases the cost of already expensive QM calculations, while implicit solvation models vary in their degree of parameterization and generality.

One approach assigns an empirical surface free energy to each exposed atom or functional group in a solute. More general algorithms combine an electrostatic term based on atomic charges and solvent dielectric constant with empirical corrections specific to functional groups and solvent cavitation energies. In the Poisson–Boltzmann (PB) model [25], solvent is represented as a polarizable continuum (with dielectric ε) surrounding the solute at an interface constructed by combining atomic van der Waal radii with the effective probe radius of the solvent. Charges are allowed to develop on this interface according to the electrostatic potential of the solute and ε through the solution of the Poisson–Boltzmann equation. Charges representing the polarized solvent are then included in the QM Hamiltonian, such that the wavefunction of the complex is relaxed self-consistently with the solvent charges via iterative solution of the PB and Schrödinger equations. Implicit models offer the advantage over explicit solvation that degrees of freedom corresponding to solvent motion are thermally averaged; thus the number of particles in a QM simulation (which typically scales as N^3 or worse) is not significantly increased.

In spite of the success of implicit solvation models, it is often easier and more precise to take advantage of the tabulated free energies of solvation of small, common species such as proton, hydroxide, halide ions, and so on [35, 36]. To screen new potential homogeneous catalysts for favorable kinetics and elucidate mechanisms of existing systems, we have typically employed the following expression for free energies of species in solution:

$$G = E_{\text{elec}} + \text{ZPE} + H_{\text{vib}} - \text{TS}_{\text{vib}} + G_{\text{solv}}, \tag{5}$$

which includes an electronic energy, E_{elec}, a temperature-dependent enthalpy, TS_{vib}, entropy contributions, H_{vib}, the zero-point-energy, ZPE, and a solvation free energy, G_{solv}, provided by a PB continuum description [14].

An example of fundamental transformations that cannot be modeled without accurate accounting of changes in electronic structure (on the order of 100 kcal/mol), solvation of multiply charged species (~100 kcal/mol), and the macroscopic concentration of protons (~10 kcal/mol) is the pH-dependent oxidation of acidic metal complexes. Figure 2 compares experimentally determined pKas and oxidation potentials [33] of *trans*-$(\text{bpy})_2\text{Ru}(\text{OH}_2)_2^{2+}$ to values computed with (5). Maximum errors are 200 mV and 2 pH units, despite the large changes in the components

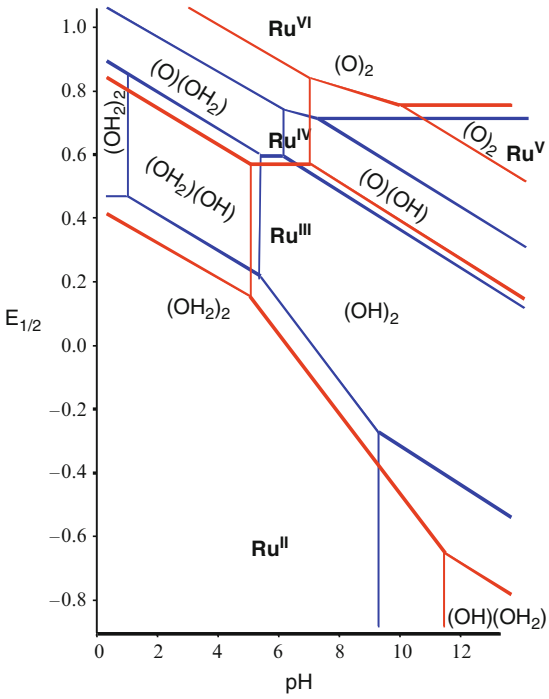

Fig. 2 Pourbaix diagram for *trans*-$(bpy)_2Ru(OH_2)_2^{2+}$ showing pKas (*vertical lines*) and oxidation potentials (*bold lines*) determined by cyclic voltammetry [33] and our calculated pKas and oxidation potentials using DFT(MO6) and Poisson–Boltzmann continuum solvation (*red*). $Ru^{III}/(OH)_2$ denotes, for example, the region of stability of *trans*-$(bpy)_2Ru^{III}(OH)_2^{1+}$

of free energy. The changes in free energy associated with redox processes determine the driving force behind many catalytic cycles. Coupled with the energies of transition states between intermediates, these tools allow *predictive* work in applications of homogeneous catalysis to problems in synthetic and energy-related reactions. Given that spin–orbit coupling corrections are important for open-shell wavefunctions of heavy elements and have been computed to useful accuracy [37], such corrections may be incorporated into (5).

Having described a hybrid approach that integrates a first-level QM-DFT approximation with a continuum-level implicit APBS solvation model, as a multiparadigm stratagem to study the effects of solvation on reactivity, we now return to describing further approximations to (1).

2.2.2 Treat the Nuclei as Classical Particles Moving on a PES

The PES found via the adiabatic approximation described in the previous section portrays the hyper landscape over which a nucleus moves, in the classical sense, while under the influence of other nuclei of a particular system. This is useful for describing vibrations or reactions. Electronic contributions have been averaged into each point on the PES, and their effect considered for that particular nuclear

conformation; therefore one might consider replacing (1) by Newton's ordinary differential equation of motion, i.e.,

$$F = -\frac{\partial V}{\partial R} = m\frac{\mathrm{d}^2 R}{\mathrm{d}t^2}, \tag{6}$$

where F represents the forces (obtained from the negative gradient of the PES with respect to nuclear positions) and m the corresponding atomic mass. Integrating (6) with respect to time leads to particle trajectories, and this is conventionally referred to as MD. Since only nuclei motions are considered, all information about the electrons is gone (e.g., quantum effects like electron tunneling, exited electronic states, and so on). Such calculations in which the forces come directly from a QM computed PES are often referred to as Car–Parrinello calculations [38]. Unfortunately, the costs of QM-MD limit such calculations to ~1,000 atoms, and at best <1 ps, so an additional simplification is to find an alternative mean to compute the PES. This is discussed next.

2.2.3 Approximate the PES with Inexpensive Analytical Forms: Force Fields

A practical solution for large systems, requiring long-term dynamics, is to describe the PES, U, in terms of a force field (FF), a superposed set of analytic functions describing the potential energy between the interacting particles (and its negative gradient, corresponding to the inter-atomic forces, F) as a function of atomic (nuclear) coordinates (x):

$$F = m_i\ddot{x}_i = -\nabla_i U(x_1, x_2, \ldots, x_n), \tag{7}$$

where U is conventionally portioned in terms of valence, or bond functions, and non-bond functions, as follows:

$$U = \left[U_\mathrm{r} + U_\theta + U_\varphi + U_\psi\right]_\mathrm{bond} + \left[U_\mathrm{vdW} + U_\mathrm{Coulomb}\right]_\mathrm{non-bond}. \tag{8}$$

Integrating (7) with respect to time, leads to a description of nuclear trajectories as a function of time.

U can take numerous forms, and since it is the key element affecting the accuracy and transferability of a force field we discuss this further below, but first a few words about the validity of the classical approximations to (1) discussed thus far.

When the thermal de Broglie wavelength is much smaller than the interparticle distance, a system can be considered to be a classical or Maxwell–Boltzmann gas (the thermal de Broglie wavelength is roughly the average de Broglie wavelength of the particles in an ideal gas at the specified temperature). On the other hand, when

the thermal de Broglie wavelength is on the order of, or larger than, the interparticle distance, quantum effects will dominate and the gas must be treated as a Fermi gas or a Bose gas, depending on the nature of the gas particles; in such a case the classical approximations discussed are unsuitable. Their use is also not recommended for very light systems such as H_2, He, Ne, or systems with vibrational frequencies $h\nu > K_BT$, systems in extreme conditions of temperature and pressure, with high energy or a large number of excited electronic states, nor for systems with two different electronic states but close nuclear energy (i.e., different ψ^n).

3 From QM to Molecular Mechanics/Dynamics: Force Fields

As mentioned previously, the definition of an empirical potential establishes its physical accuracy; those most commonly used in chemistry embody a classical treatment of pairwise particle–particle and n-body bonded interactions that can reproduce structural and conformational changes. Potentials are useful for studying the molecular mechanics (MM), e.g., structure optimization, or dynamics (MD) of systems whereby, from the ergodic hypothesis from statistical mechanics, the statistical ensemble averages (or expectation values) are taken to be equal to time averages of the system being integrated via (7).

In the following sections, we outline our first-principles-based Dreiding [39] potential, to exemplify regular force fields, which usually cannot reproduce chemical reactions, and follow up with an introduction to two of our unique force field approaches, which overcome most of the limitations in the conventional approach. In each case, we present unique applications to demonstrate their usefulness.

3.1 Conventional Force Fields

Traditionally, the bonded components are treated harmonically (see expressions in Fig. 3). There are generally two non-valence or non-bonded terms: the van der Waals term (U_{vdW}) which accounts for short-range repulsion, arising from the Pauli Principle and interacting dipoles, and for long range attractions arising from the weak London dispersions, expressed generally as

$$U_{vdW} = \sum_{\substack{R_{ij} > R\text{cut} \\ [\text{excl}(1-2,1-3)]}} \hat{U}_{vdW}(R) \cdot S\left(R_{ij}, R_{on}, R_{off}\right), \tag{9}$$

where U_{vdW} can represent different forms, and the electrostatic or Coulomb interactions, which account for the charged interactions between every *ij* pair of atoms

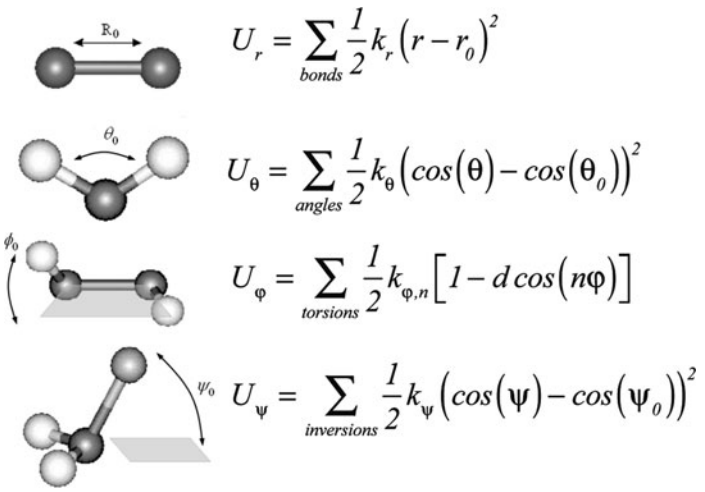

Fig. 3 Conventional (Dreiding) valence interatomic potentials. Sub-indices *0* indicate equilibrium values, *k* constants are related to force constants for vibrational frequencies, *c* constants are related to an energy barriers, and *n* refers to periodicity

flowing within a dielectric medium ($\varepsilon = 1$ in a vacuum but larger values are used for various media), expressed conventionally as

$$U_{\text{Coulomb}} = C_0 \sum_{i>j} \frac{Q_i Q_j}{\varepsilon R_{ij}} S\left(R_{ij}, R_{\text{on}}, R_{\text{off}}\right), \tag{10}$$

where C_0 corresponds to a unit conversion scalar (e.g., for energy in kcal/mol, distances in Å, and charge in electron units, $C_0 = 332.0637$), $Q_{i,j}$ to the pairwise point charges, R_{ij} to the interparticle distance, and S to a cutoff function.

One additional term included in Dreiding accounts for weak hydrogen bonded interactions, as a mixture of 3-body angles (between an H atom, and H donor and acceptor atoms) and non-bonded terms (between donor and acceptor atoms), and is given by

$$E_{\text{HB}}(R, q_{\text{AHD}}) = E_{\text{b}}(R) E_{\text{a}}(\cos(q_{\text{AHD}})). \tag{11}$$

The most time-consuming aspect of MD simulations for large systems corresponds to the calculation of long-range non-bond interactions, (7) and (8), which decrease slowly with R. This scales as O (N^2) for an N particle system (e.g., a protein with 600 residues would have ~6,000 atoms requiring ~18 million terms to be evaluated every time step). One way to reduce this cost is to allow the long-range terms to be cut off smoothly after a threshold value (S function in (9) and (10)). Alternatively, our Cell Multipole Method (CMM) [40] (and the Reduced CMM [41]) enable linear scaling, reducing the computational cost while retaining accuracy over large-scale systems.

Many useful FF have been developed over the last 30 years ([42] provides a recent review), a significant number of which are aimed at biological systems. Commonly used FF include AMBER [43], CHARMM [44], Dreiding [39], and OPLS [45]. Most of the parameters in these FF were adjusted to fit a combination of results from theory and experiments.

A key strategy in our multiscale approach has been to parameterize force fields (e.g., non-reactive Dreiding) from QM calculations on small representative systems, adjusting the FF descriptions to reproduce the structures, energetics, and dynamics from QM on nanoscale systems. This favors transferability and the predictive capability, in particular for systems with little or no existing empirical data.

With an FF it is practical to apply MD simulations to the atomic-level dynamics of large-scale systems (e.g., proteins [46]) interacting with other nanoscale components or external fields under complex conditions. Force fields allow one to carry MD simulations on systems ~10^{6-9} times larger than for QM. It is no surprise then that a particularly well-suited niche for the application of force fields is in the prediction of protein structures, in particular for membrane proteins that are otherwise impossible to crystallize in their active form using existing experimental methods. One class of membrane proteins with significant relevance is that of G-protein coupled receptors, mainly because they are involved in many diseases. This fact makes them a target of approximately 30–50% of all modern medicinal drugs. GPCRs are seven-transmembrane protein domain receptors, that sense molecules outside the cell and activate inside signal transduction pathways and, ultimately, cellular responses. GPCRs are found only in eukaryotes, choanoflagellates, and animals. The ligands that bind and activate these receptors include light-sensitive compounds, odors, pheromones, hormones, and neurotransmitters, and vary in size from small molecules to peptides to large proteins.

The following section describes our multiscale, multiparadigm modeling approach to protein structure prediction, and in particular to GPCRs.

3.1.1 Application Example: Structure Prediction of GPCRs

The activation related conformational changes in G protein-coupled receptors (GPCRs) allow cells to sense their environment and convert extracellular signals (e.g., ligand binding) into intracellular signals (through G protein and β arrestin pathways), leading to physiological responses. They are activated by a variety of molecules (including biogenic amines, peptides, lipids, nucleotides, hormones, proteins, tastants, odorants, among others) and non-molecular sensory signals (such as light, touch, and others), and thus play an important role in all major disease areas including cardiovascular, metabolic, neurodegenerative, psychiatric, cancer, and infectious diseases. There are ~370 non-sensory human GPCRs (out of ~800 human GPCRs [47, 48] but experimental crystal structures are available only for two (human β_2 and human A_{2A}), both in the inactive form. We can expect additional structures for human GPCRs to become available slowly over the next

few years, but most will also be in the inactive form. This lack of structures is contributing to the shortage of safe and efficacious drugs that target GPCRs. Availability of GPCR structures, experimental or predicted, can lay the foundation for rational structure-based drug design. So, a fast but accurate computational approach is needed that can generate structures for all important conformations of a target receptor and any other receptors implicated for off-target therapeutic side-effects and determine their ligand binding efficacies for developing highly selective drug candidates with potentially minimal side-effects. We have developed one such FF based method.

The structural topology of GPCRs consists of seven transmembrane (TM) α-helices that span the membrane and are connected by both intracellular and extracellular loops. To characterize this topology quantitatively with respect to a common reference frame, the middle of the membrane is assumed to correspond to the $z = 0$ plane or the hydrophobic plane that cuts the 7-helix bundle into two halves. Each GPCR structure can then be characterized by the six orientation parameters of the seven helices shown in Fig. 4, which shows how the helix position and tilt are defined. Helix position on the hydrophobic plane is then given by x and y. Value h corresponds to the hydrophobic center residue from the helix that will be positioned on the hydrophobic plane. Two angles, θ and ϕ, specify the tilt angles of the helix and the angle η corresponds to the helix rotation angle about its axis. The two tilt angles (θ, ϕ) and the rotation angle (η) require a definition of the helical axis which needs to account for the reality of bent helices as prolines are commonly found in the TM helices. We use a helical axis that corresponds to the least moment of inertia vector for the helix obtained by eigensolution of the moment of inertia matrix for the helix using only heavy backbone atoms.

The structural analysis of available experimental structures shows large variations in helix tilts and rotations. Considering a $\pm 10^\circ$ sampling of the θ tilt angle, and $\pm 30^\circ$ sampling of the ϕ and η angles, for each of the seven helices, leads to $(3 \times 5 \times 5)^7$, ~10 trillion possible conformations, for each of which the amino acid side chains must be optimized. To make such a huge sampling computationally feasible, we developed the SuperBiHelix sampling (BiHelix sampling only sampled the η rotation angle) procedure. As indicated by 12 double arrows in Fig. 5 (left) a typical class A GPCR template has 12 important pair-wise interactions. For each such pair of helices, we will sample all combinations of θ, ϕ, η over some grid. During this sampling, the other five helices are ignored, as indicated in Fig. 5

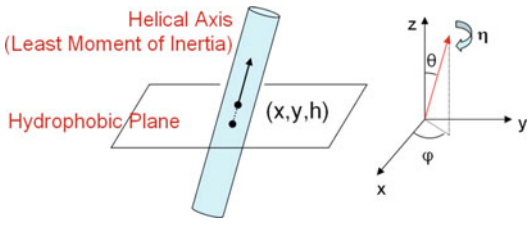

Fig. 4 Definitions of orientation parameters of a transmembrane helix

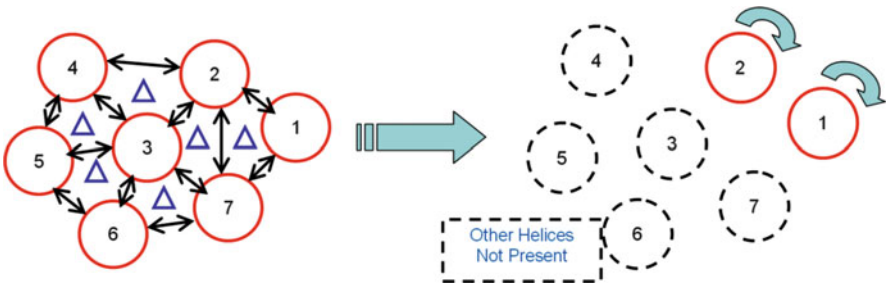

Fig. 5 *Left*: The 12 important helix pair interactions. *Right*: The BiHelix concept in which the interhelical interactions are optimized for a pair of helix rotations, while ignoring the other 5

(right) for helix 1–2 pair. For each of the $(3 \times 5 \times 5)^2 = 5{,}625$ rotational-tilt combinations of each of the 12 pairs, we optimize the side chains using SCREAM [49] with all-atom DREIDING force field [39]. SCREAM uses a library of residue conformations ranging from a CRMS diversity of 0.4–1.6 Å in conjunction with a Monte Carlo sampling using full valence, hydrogen bond, and electrostatic interactions from D3FF, but with a special flat bottom van der Waals (vdW) potential that reduces the penalty for contacts that are slightly too short while retaining the normal attractive interactions at full strength.

The total energies for each of the (12) × (5,625) helix pair combinations are used to estimate the energy for all 10 trillion 7-helix bundle conformational combinations. In a procedure called SuperComBiHelix, the top 2,000 of these helical bundles are explicitly built and the side chains reassigned, given that they will take different conformations compared to those in the bihelical model. Then the structure is minimized for ten steps. The energy ranking will be different in SuperComBiHelix than SuperBiHelix because all seven helices are present instead of just two at a time. This procedure results in an ensemble of low-lying structures. Examination of the low-lying structures shows the helix packing preferred by the receptor, which should also include conformations relevant for understanding function and activation of these proteins. This procedure was applied to A_{2A} receptor where its helices were placed in the β2 template. The starting structure was 2.0 Å from the A_{2A} crystal structure (see Fig. 6). After the SuperBiHelix/SuperCombiHelix optimization (sampling ~10 trillion conformations), the lowest energy structure was 1.3 Å from the A_{2A} crystal structure (Fig. 6), an improvement that was also critical in the identification of correct ligand binding mode in the protein. The procedure has been applied to all available crystallized receptors (humBeta2, turBeta1, bovRhod, humA2A). For each of these systems we first sampled only the helix rotation angle η over full 360° range in increments of 15°, which needs the sampling of 24^7 (~4 billion) conformations. This resulted in crystal structure ranking number 1 at the end of the procedure for all cases. The results for turBeta1 are shown in Table 1, which shows the crystal structure as ranked number 1 and other near-native structures also ranked in the top ten list.

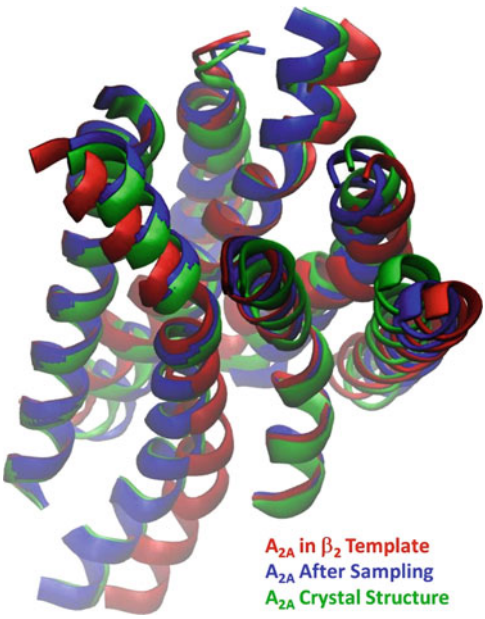

Fig. 6 Visual comparison of the starting crystal structure (*green*) and the predicted tertiary structure [before (*red*) and after optimization (*blue*)]

Table 1 Top ten predicted conformations for Turkey β1 adrenergic receptor

Eta	H1	H2	H3	H4	H5	H6	H7	TotEnergy	RMSD
Eta	0	0	0	0	0	0	0	−385	0.0
Eta	0	0	0	345	0	0	0	−379	0.4
Eta	0	0	0	0	345	0	0	−377	0.4
Eta	0	0	0	0	15	0	0	−376	0.4
Eta	0	0	0	345	15	0	0	−365	0.4
Eta	0	0	345	345	15	0	0	−364	0.5
Eta	0	0	0	15	0	0	0	−349	0.4
Eta	0	0	0	15	345	0	0	−349	0.4
Eta	0	0	0	0	0	0	345	−346	0.4
Eta	75	0	0	345	0	0	0	−341	1.1

These results show that the two-helices-at-a-time sampling method is robust, but, more important, the FF-based energies used for scoring the conformations can be trusted to resolve the near-native structures of proteins.

3.2 Simulating Complex Chemical Processes with FFs

A major drawback with MD using conventional FFs is that they are unable to describe chemical reaction processes, or other electronic structure dependent processes such as electronic excitations, and as we've already discussed, QM is not

practical for systems larger than ~1,000 atoms, and timescales longer than 1 ps. We have made significant breakthroughs in recent years towards addressing these problems, through the development of the reaxFF reactive force field [50] for describing "ground-state" reaction processes, and the electron force field (eFF) [51] for describing systems with explicit electrons in their ground or excited states.

3.2.1 The ReaxFF Force Field for Studying Reactive Processes

ReaxFF [50] provides a generally valid and accurate way to capture the barriers for various chemical reaction processes (allowed and forbidden reactions) into the force fields needed for large-scale MD simulation. ReaxFF is parameterized exclusively from QM calculations, and has been shown to reproduce the energy surfaces, structures, and reaction barriers for reactive systems at nearly the accuracy of QM but at costs nearly as low as conventional FFs.

Applications of ReaxFF have been reported for a wide range of materials, including hydrocarbons [50], nitramines [52], ceramics [53], metals and metal oxides [54, 55], metal/hydrocarbon interactions [56], and metal hydrides [57]. ReaxFF has been used to simulate chemical events in many systems, including nanotube deformation and buckyball polymerization [58, 59], thermal decomposition of polymers [60], high-energy materials initiation [61, 62], crack propagation [63], reactive oxygen and hydrogen migration in fuel cells [64], and metal/metal oxides surface catalysis [65].

Salient features of reaxFF include: (a) *Environmentally dependent charge distributions on atoms*. The charges on the atoms adjust in response to the local environment allowing them to change as bonds are broken and formed and to shield the Coulomb interaction between atoms at small distances; (b) *Bond order dependent valence terms*. A general relation is provided between bond distance and bond order and between bond order and bond energy (and hence forces). The bond orders gradually go to zero as the bond lengths increase and they gradually increase for shorter distances, finally saturating for the smallest distances (e.g., BO = 3 for CC bonds). This provides a smooth description of the valence terms during chemical reactions; (c) *Non-bond or van der Waals interactions*. ReaxFF uses a simple Morse function to account for the short-range repulsion and steric interactions arising from the Pauli principle (between *every* atom pair). The long range attraction accounts for vdW attraction; (d) *No cutoffs*. All interactions change smoothly during reactions (which are allowed to occur at any time and place) so that ReaxFF can be used with general conditions of temperature and pressure; (e) *Transferable potential*. Simple FFs provide different parameters for different atomic environments (e.g., single vs double bonds, sp3 vs sp2 geometries). ReaxFF eschews such description using only a single atom type for each element, which is necessary since bond orders and geometries change during reactions. This leads to good transferability of the FF; (f) It is *QM-based*. All parameters are optimized/derived directly from QM studies on a large number of reactions. This allows extensions to new materials where there may be no experimental data.

Recently, we have demonstrated the use of reaxFF to the challenging problem of elucidating the growth process of carbon nanotubes (CNTs). Understanding this process is critical for determining the control variables that lead to chiral-specific (with semiconducting or metallic electrical conductivity behavior) mass production of CNTs. These results are summarized in the following section.

Application Example: Dynamics of CNT Growth

Since their discovery in 1991 [66], CNTs have been widely studied. Researchers have proposed CNT applications to an ample set of technologies [67] including interconnects, transistors, and diodes for microelectronics [68], as well as electrochemical transducers [69], sensor components [70], field emission devices [71], and even gas sensors [72]. The mass production of uniform, well-characterized CNTs is crucial for realizing many of these applications. However, while CNT synthesis has been demonstrated for numerous catalysts, and a wide range of reaction conditions, complete product control has remained elusive [73]. Thus, multiple investigations aimed at elucidating the key mechanism or mechanisms of CNT growth are still being carried out, in the hope that a more fundamental understanding of the growth process will result in better synthetic control [74]. Experimental observations have shed some light on CNT growth mechanisms. Atomic force microscopy (AFM), scanning electron microscopy (SEM), and tunneling electron microscopy (TEM) have been used to support instances of tip and base growth mechanism in different synthesis procedures [75–77]. More recently time-resolved, high-resolution in situ TEM studies have highlighted the role of catalyst deformation in SWNT growth and provided direct experimental validation for a Yarmulke mechanism for nucleation [78, 79]. Nevertheless, these cutting edge techniques provide overarching, general descriptions rather than detailed, atomistic mechanisms for each stage of CNT synthesis.

To fill in these experimentally inaccessible details, mechanistic studies often appeal to atomistic simulations. DFT is now widely used to explore catalytic systems, and has been applied to simplistic models of CNT growth [74, 80, 81]. Nevertheless, the usefulness of DFT is hampered by stringent limitations on the number of atoms and especially the number of structural iterations that it is feasible to consider with current computer technology [82]. Tight binding (TB) methods, which use approximations (i.e., simplified integrals) to reduce the computational cost of handling electron–electron interactions explicitly, have been used in conjunction with MD simulations to study this problem [83]; however, the timescales necessary for observing the growth process are still beyond the reach of this approach – even though TB calculations are typically a couple orders of magnitude faster than DFT [83]. Monte Carlo methods have provided another popular means of "simulating" CNT growth [82, 84]. At best, however, Monte Carlo methods show a succession of possible snapshots from the growth process, leaving the mechanistic details hidden.

As reported previously [85, 86], we have developed a set of ReaxFF parameters describing hydrocarbon chemistry catalyzed by nickel and nickel carbide catalyst particles. This ReaxFF potential is capable of treating the adsorption and decomposition of both saturated and unsaturated hydrocarbon species on several different nickel surfaces. Of particular relevance for studying CNT growth is that a single set of ReaxFF parameters accurately describes carbon in all hybridization states and a variety of chemical environments. These states include sp, sp^2, and sp^3 hybridized carbon in various hydrocarbon molecules, carbon binding at and migrating between interstitial sites in bulk nickel, and carbon bonded to nickel surfaces strongly as an adsorbed lone adatom or in a small hydrocarbon molecule, or weakly as part of a graphene layer.

While the vast majority of theoretical studies of CNT growth starts with lone carbon atoms, assuming that decomposition has already taken place, there are conditions (e.g., low temperature growth) under which decomposition is believed to be the rate-limiting step [87]. Thus we have utilized this ReaxFF force field in a reactive dynamics (RD) study of the early stages of CNT growth. In [85] we reported on the chemisorption and decomposition of various hydrocarbon species on a nickel nanoparticle. Over the course of 100 ps of RD simulations performed, we were able map out the preferred reaction pathways for the decomposition of each hydrocarbon species studied.

The synthesis of CNTs can be broken down into three or four distinct stages. The first stage is feedstock decomposition, as discussed above. Under low temperature growth conditions, experiments suggest that feedstock decomposition is the rate-limiting step [87]. Thus our analysis of hydrocarbon decomposition pathways on nickel nanoparticles shows how the selection of different hydrocarbon species for the feedstock influences the chemisorption rate, surface coverage, and extent of carbide formation during the nanotube growth process. In particular, because we find that chemisorption is the rate limiting decomposition step for saturated hydrocarbons, the selection of unsaturated hydrocarbon species, with very small chemisorption barriers, for the feedstock, is expected to improve the growth rate where feedstock decomposition is rate limiting.

Following feedstock decomposition is the carbon transport stage, in which a hydrocarbon or carbon species is either transported along the catalyst surface or else diffuses through the catalyst bulk as carbide. Because a constant supply of carbon is needed for both nucleation and growth, carbon transport likely occurs during both the nucleation and growth stages and so is most naturally treated as a part of each of these stages taken separately. It is also possible that a partially decomposed species migrates to the nucleation or growth site where it further decomposes into the activated species. In any case, experiments indicate that there are growth conditions under which surface diffusion is the rate-limiting step [88]. ReaxFF RD simulations demonstrate the formation of nickel carbide following acetylene chemisorptions and decomposition, lending plausibility to either mechanism.

It is believed that nucleation occurs when enough carbon material accumulates on the surface for the formation of surface ring structures. The ring structures develop into a graphene island on the particle which, when it becomes large enough,

lifts its center off the particle surface in the experimentally observed Yarmulke mechanism [78, 88]. Currently, ReaxFF RD simulations beginning with several hundred gas-phase acetylene molecules surrounding a nickel nanoparticle support the early stages of this picture. Initially, as acetylene chemisorbs and decomposes on the Ni nanoparticle, the C atoms formed migrate into the bulk of the catalyst and forming carbide. After a couple nanoseconds of dynamics, the chemical potential gradient reverses and carbon begins segregating to the surface, forming carbon chains. As more carbon moves to the surface, ring structures form and clump together to form larger ring structures, resulting in multi-ring structures with tens of rings formed from a couple of hundred carbon atoms (see Fig. 7). Thus, the trajectories from these RD simulations provide an atomistically detailed picture of the early stages of CNT growth.

Following nucleation is the nanotube growth stage in which carbon is added to the end of the growing nanotube. This stage likely lasts significantly longer than the previous stages, which means that ReaxFF RD simulations of the entire growth stage are probably not computationally feasible at present. Nevertheless, a couple different strategies are available for overcoming this difficulty. The first is to use an already growing nanotube as the initial structure for ReaxFF simulations, and study just a part of the growth process. As a simple model we have used ReaxFF to consider the barriers for adding small hydrocarbon species to the edge of a graphene sheet laying on a Ni(111) surface. These simulations find the lowest carbon addition barriers for C_2 hydrocarbon species, suggesting that C_2 may be the activate form of carbon responsible for CNT growth. Unconstrained ReaxFF RD on a full-scale model of a growing CNT will provide further validation for this hypothesis.

The second option for circumventing the time limitations on ReaxFF RD is the use of a kinetic Monte Carlo procedure to bypass long periods of quasi-equilibrium dynamics between reaction events using principles from statistical mechanics and

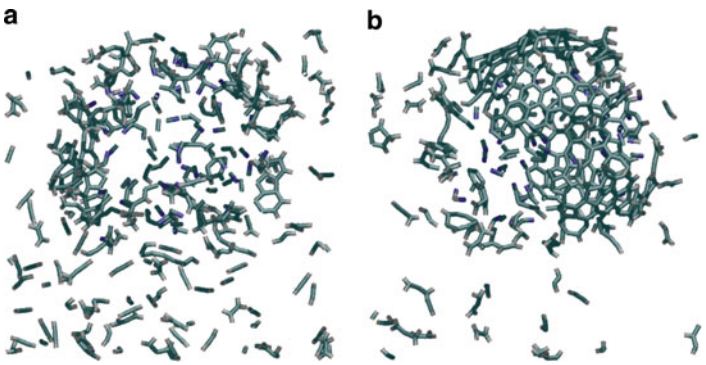

Fig. 7 ReaxFF RD simulations of acetylene adsorption and decomposition on a 468-atom nickel particle [not shown]: (**a**) after 1 ns a limited number of structured rings have formed and (**b**) after 2 ns of ReaxFF RD simulations a clear ring pattern formation appears. Simulations were performed using the parallel prototype reaxFF implementation from collaboration with H.M. Aktulga and A. Grama at Purdue, and A.C.T van Duin at Penn State

transition state theory. Because traditional kinetic Monte Carlo methods require predefined reactions and make the lattice approximation, they are not directly applicable to a complex process such as CNT growth. Nevertheless, alternative schemes have been proposed for circumventing the lattice approximation by calculating reaction barriers on the fly [89]. The bond order/bond distance relationship already present in ReaxFF would provide a natural tool for the development of an automated reaction search procedure, enabling kinetic Monte Carlo simulations within the ReaxFF framework. Such simulations would be capable of looking at CNT growth over a significantly longer timescale than ReaxFF RD.

As effective as reaxFF is for handling reactive systems and processes in their ground-state, it is unable to describe the dynamics of electrons and systems with excited electronic states. QM-MD is also limited mostly to ground-state dynamics or to a very small number of excited electronic states (see [90] for further discussion on this). The following section presents our progress in addressing this problem with a mixed quantum-classical force field method, the eFF.

3.2.2 Non-Adiabatic Excited Electronic State Dynamics with an FF

A significant number of processes involve excited electronic states, whose character – and representation in a theoretical method – depends strongly on the degree of excitation involved. Low-level electron excitations of molecules can initiate radical reactions, isomerize bonds, and induce transfers of electrons. Such processes can be studied effectively using conventional QM, using a wavefunction formed from Hartree–Fock or Kohn–Sham orbitals. At the other extreme, high-level excitations result in the formation of a weakly-coupled plasma, where bonding and chemistry vanishes, and electrons act as point particles interacting with nuclei via classical electrostatics. Such systems can be studied using classical plasma simulations techniques, i.e., particle-in-cell codes.

However, in between low and high extremes of electron excitation lies a rich variety of phenomena where the electrons are far removed from the ground state, yet remain strongly coupled to the nuclei, so that remnants of bonding and chemistry persist.

Understanding the properties of warm dense matter present in moderately excited systems is of crucial importance to developing a range of new technological enterprises. For example, in inertial confinement fusion, liquid deuterium is compressed by a shock wave, causing molecules to dissociate into atoms, atoms to ionize into plasma, and metallic conducting phases to form. Knowledge of how these phases interact could contribute to the design of improved fuel pellets.

Other examples come from the semiconductor industry, where electron beams are used to etch ultra-fine features ($<$35 nm) into silicon, the nuclear industry, where the interior of reactors must be protected from the passage of fast charged particles, and the biological community, where synchrotron radiation could enable single molecule X-ray diffraction, if the dynamics of highly excited and ionized biomolecules could be understood. In the above cases, theory could play a critical

role by elucidating the fundamental properties of chemical bonds and relating it to the performance of materials under extreme conditions.

To study such systems, our group has developed the eFF [59, 90, 91] approximation to QM, which can simulate moderate excitations (tens to hundreds of electron volts) that vary sharply over space and time, in large systems (tens of thousands of atoms) where strong couplings between nuclei and electrons exist, and chemistry occurs. In eFF, electrons are represented by wave packets, and nuclei by classical point charges moving in the time-varying field of the electrons (Ehrenfest dynamics). The overall electronic wavefunction is represented by a Hartree product of spin orbitals, where each orbital is a single Gaussian wave packet with size (s) and position (x):

$$\Psi(r) \propto \prod_i \exp\left[-\left(\frac{1}{s_i^2} - \frac{2p_{s,i}}{s_i}\right)(r - x_i)^2\right] \cdot \exp\left[ip_{x,i}r\right], \tag{12}$$

where p_x and p_s correspond to the conjugate translational and radial momenta, respectively, which represent the translational motion and radial expansion/shrinking of the wave packet over time. To account for the orbital orthogonality imposed by the Pauli principle, a spin-dependent Pauli repulsion potential is included in the Hamiltonian, which increases with the overlap between the wave packets (see Fig. 8).

The eFF method is similar to previous wave packet molecular dynamics (WPMD) approaches, but differs in the form of the Pauli potential, which we derived to account for the effect of pairwise orthogonalization on the kinetic energy of orbitals; with this change, the scope and accuracy of previous approaches is greatly extended. The only other terms present in eFF are interactions between charge distributions from classical electrostatics, and a kinetic energy term for the electron wave packets derived from QM, which provides the "kinetic energy pressure" that prevents electrons from collapsing to a point:

$$E_{\mathrm{KE}} = \frac{\hbar^2}{m_{\mathrm{e}}} \sum_i \frac{3}{2s_i^2}, \tag{13}$$

where m_{e} corresponds to the electron mass. From these simple terms, a rich array of chemical phenomena emerges – separation of core and valence electrons into

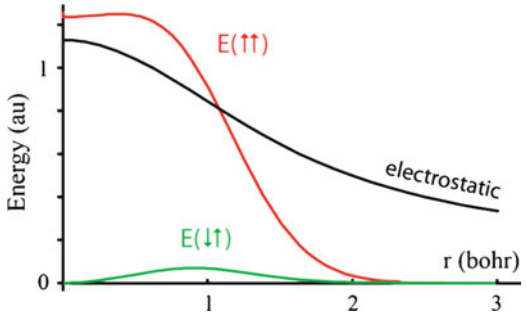

Fig. 8 Pauli repulsion between two electrons with size $s = 1$ bohr, as a function of their separation, r, and spin. These curves are described with three universal parameters adjusted to give reasonable structures and energies for CH_4, C_2H_6, LiH, and B_2H_6

separate shells, atomic hybridization, covalent, ionic, multicenter, and/or metallic bonds, and steric repulsions between bonds.

Since the interactions between particles in eFF are simply pairwise forces, the overall method is extremely fast and scales well computationally. We have simulated on a single processor tens of thousands of electrons, and on multiple processors millions of electrons [90].

Application Examples: Material Shock Hugoniots and Auger Decay

In one application of eFF we studied the thermodynamics of shock-compressed liquid hydrogen, characterizing molecular, atomic, plasma, and metallic phases at temperatures up to 200,000 K and compressions up to fivefold liquid density (see Fig. 9). We found reasonable agreement with data from both static compression (diamond anvil) and dynamic compression (shocks from explosives, magnetically pinched wires, lasers) experiments.

We have also demonstrated the capabilities of eFF for computing single-shock Hugoniots for lithium metal from dynamic shock wave experiments, via the shock wave and piston kinematics and initial and final densities of a 640,000-particle system (see Fig. 10). We also reported on the degree of ionization suffered by the material, a function of the explicit nuclear delocalization of electrons [90]. A simpler depiction of such dynamic shock experiments is shown in Fig. 11, wherein

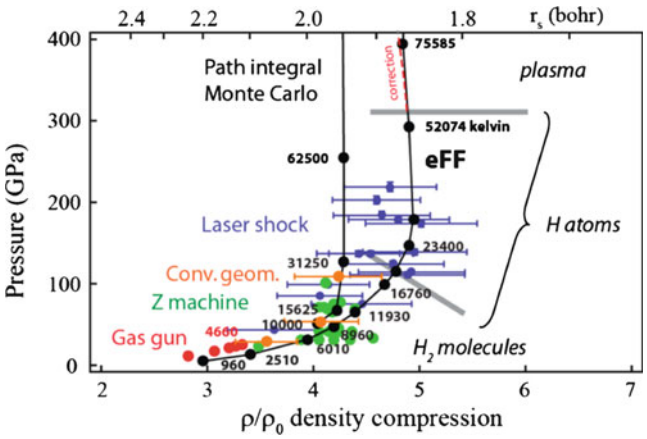

Fig. 9 Shock Hugoniot curve for liquid D2. We show here that eFF agrees well with most experiments: gas gun (*red dots*), Z machine (*green dots*), convergence geometry (*orange*), and the more recent laser data (*blue dots*) from LLNL. The PIMC results agree with eFF up to a compression of 4.2, but leads to a lower limiting compression than eFF. To compute the Hugoniot curve, we perform NVE simulations of D2, interpolating to temperatures such that the internal energy, volume, and pressure satisfy the Rankine–Hyugoniot relationship. We note that the eFF Hugoniot curve connects to an eFF low temperature starting point, while the PIMC Hugoniot curve connects to a U_0 from a separate calculation

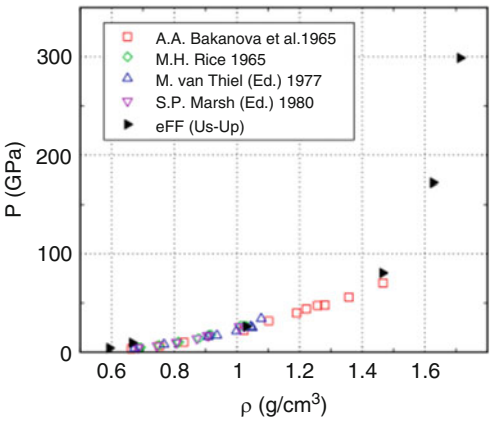

Fig. 10 Shock Hugoniot for lithium calculated directly from the planar shock velocity U_s, particle or piston velocity U_p, and initial and final densities obtained from our simulations compared to existing experiments. From [90]

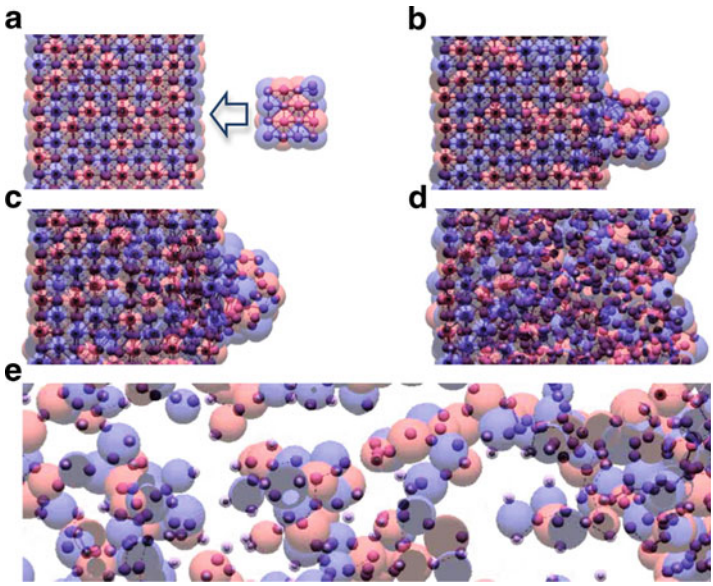

Fig. 11 Hypervelocity impact of a Li cluster on a Li metal slab. (**a**) Initial state. (**b**) Impact at $v = 2$ km/s leads to welding and no ionization. (**c**) Impact at $v = 5$ km/s leads to melting, and scattered ionized valence electrons. (**d**) Impact at $v = 10$ km/s leads to a fluid, and ~0.25 fraction of ionized valence electrons. (**e**) Impact at $v = 20$ km/s leads to a plasma with ~0.7 fraction of ionized electrons. *Dark small spheres* represent nuclei, *red/blue spheres* depict electron up/down spin, and their size represents the degree of de/localization

a small lithium cluster impacts onto a lithium metal slab at different speeds. Different material phases are observed as a function of impact velocity, as well as degrees of electron ionizations. Dynamic shock experiments enable higher compressions than static compression experiments.

In another application we have examined the Auger decay process in a diamond nanoparticle and in silicon, relevant to the etching of semiconductor substrates using low energy electron-enhanced etching processes [92]. We found, for the diamond case, that ionizing core electrons induced selective breaking of bonds via a variety of mechanisms, i.e., through direct excitation and ionization of valence electrons, or through indirect heating, or even in a small subset of cases, a billiard-ball like scattering away of valence electrons through ejection of neighboring bonding electrons (see Fig. 12). Our results were consistent with ion ejection data from photon-simulated desorption experiments performed on diamond films.

Our current development of eFF involves adding explicit electron exchange-correlation potentials, core pseudo-potentials, and extended support for systems with significant p- and d-character. Using eFF, we're now able to study the effect of highly excited electrons in the dynamics of material subjected to extreme conditions, including those described before, as well as other open problems in interfacial shock instabilities, radiation damage, to name a few.

As simulation requirements shift to larger length scales and longer times and system properties are amenable to homogenization in space and averaging in time, for example in characterizing the conformational behavior of supramolecular systems, coarse-grain methods tuned from finer scale ones (e.g., QM, MM) represent a suitable and more efficient alternative for evolving the dynamics of systems with reduced degrees of freedom. The following section discusses our progress in developing coarse-grain force fields and time-lower bound solutions to the resulting rigid multibody EOM.

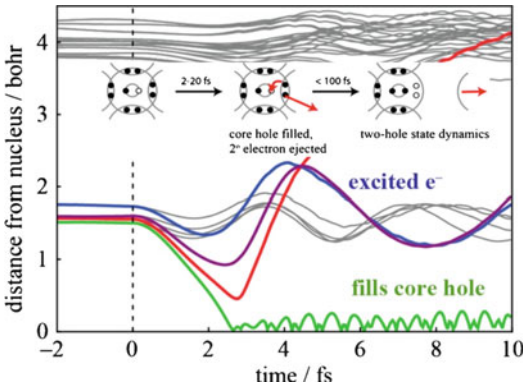

Fig. 12 Single Auger trajectory after ionization of a carbon core electron at the center of the diamond nanoparticle ($m_{elec} = m_p$). Valence electrons surrounding the core hole with the same spin as the ionized core electron are highlighted in *red*, *green*, *blue*, and *purple*. Distance of valence electrons from the core hole, showing the *green electron* filling the core hole, the *red electron* being ejected (and trapped after 20 fs, not shown), and the *blue* and *purple electrons* being excited. From [93]

4 Bridging MM/MD with the Mesoscale

Successful applications of first principles methods to supramolecular modeling requires a scale lying in between the molecular or atomistic scale (where it is convenient to describe molecules in terms of a collection of bonded atoms) and the continuum or macroscale (where it is convenient to describe systems as continuous with a finite element mesh basis) [94]. This coarse grain or meso-scale level is most important for determining the properties and performance of a wide range of different materials, including "soft condensed matter," amphiphilic fluids, colloids and polymers, gels, liquid crystals; proteins, and DNA. An important class of problems that need to be described at these scales include biological processes such as protein activation, enzymatic transformations, ribosome activity, and general diffusive motions of biomolecules on timescales of microseconds and longer [95].

Several approaches can reduce the computational costs of solving (7) for systems with a large number of atoms, including the use of explicit constraints on fast atomic motions, bead representations which join several atoms into pseudo-particles with no rotational inertia, or representations which cluster collections of atoms into rigid bodies with inertia, among others. We will refer here to those treating clusters of atoms as rigid bodies that interact with others through net forces and torques, and which use coarse-grain force fields to solve the system's dynamic behavior.

4.1 Constrained and Coarse-Grain MD

By imposing constraints on fast atomic motions, one can effectively increase the timescales of integration of the EOM, thereby enabling longer simulation times. Conventional methods for doing this on Cartesian atomistic degrees of freedom include those that compensate relative restraint forces into the particle EOM, such as SHAKE [96, 97], RATTLE [98], and the like [99–101]. Unfortunately, these methods are limited to low temperature dynamics [102] and to relatively small systems, due to the added cost of solving for the explicit constraints. To overcome this, alternative approaches have focused on simplifying the description of the system through EOM that operate only on the degrees of motion of a system, mostly using internal coordinate representations that treat clusters of atoms as rigid bodies.

Mazur et al. [103, 104] demonstrated the conformational dynamics of biomacromolecules. However, their method scaled exponentially with size and relied on an expensive expression for the inter-atomic potentials in internal coordinates. Subsequently, our group pioneered the development of internal coordinate constrained MD methods, based on ideas initially developed by the robotics community [102, 105–107], reaching O(n) serial implementations, using the Newton–Euler Inverse Mass Operator or NEIMO [108–110] and Comodyn [111] based on a variant of the Articulated Body Inertia algorithm [112], as well as a parallel implementation of O(log n) in O(n) processors using the Modified Constraint Force Algorithm

or MCFA [107, 113]. These methods can selectively handle implicit constraints through appropriate projection matrices in the EOM. Nonetheless, most have focused on the torsional degrees of freedom (DOF) which affect the conformation of a system. The general state space EOM for internal coordinate constrained MD can be written as

$$\tau = M(Q)\ddot{Q} + C\left(Q, \dot{Q}\right)\dot{Q}\,, \tag{14}$$

where τ corresponds to the vector of generalized forces (e.g., torques), M denotes the articulated body inertia matrix, C denotes the nonlinear velocity dependent terms of force (e.g., Coriolis, centrifugal and gyroscopic forces), and $Q, \dot{Q}, \ddot{Q}$ correspond to the generalized coordinates that define the state of the system. It then follows that the dynamics of motion for a microcanonical ensemble is obtained by solving for the hinge accelerations, access to increased integration time-steps, faster exchange between low- and high-frequency modes for high temperature dynamics, and faster and smoother sampling of the PES (conformational space), among others. Our rigid body MD approaches, with atomistic and coarse-grain force fields, are currently used to predict the conformational evolution of helical domains in GPCR protein bundles (see Fig. 13):

$$\ddot{Q} = M^{-1}(Q)\left[\tau - C\left(Q, \dot{Q}\right)\dot{Q}\right]. \tag{15}$$

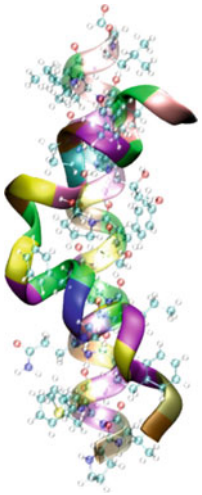

Fig. 13 Beta2 GPCR helix 7 final structure (shown in ribbons representation) after 100 ps of constrained MD-NVT, using Comodyn with a 5 fs timestep and the Dreiding force field, shows a kink about a proline amino-acid group. The kink is also observed during full-atom MD-NVT at roughly the same timescale. The original starting structure is shown in transparency, for both the full-atomistic and ribbon representations. The coarse-grain representation involved 127 clusters for a total of 133 DOF in the internal coordinate representation of the equations of motion (compared to the 1,170 DOF for the atomistic model)

These constrained MD methods rely on efficient solutions to the rigid-body EOM and on the use of atomistic, simplified atomistic (i.e., no intra-cluster valence or non-bond interactions) or coarse-grain force fields, discussed next.

Coarse grain (CG) models must carry enough information about the atomistic behavior while at the same time be efficient to scale in both time (>1 ms) and length (>100 s nm). For example, accurate models to represent solute-solvent interactions should account for solvent momentum so that its behavior can be consistent with hydrodynamics, have the correct density at the desired temperature, and be able to maintain a liquid/vapor interface over such a temperature [114].

To this end, several approaches have been used and demonstrated for phospholipids [115–123], oligosaccharides and their water mixtures [124], proteins [125], and dendrimers and polymers [126]. Coarse-grain force fields have been developed from heuristically simplified models of biomolecules (e.g., water, alkanes, lipids, etc.) and by systematic optimization procedures of a set of interaction potentials between collections of atoms treated as rigid bodies. In the former case, the fitting process relies on results from the finer atomistic MD using Monte Carlo schemes [114, 127], random search algorithms including genetic algorithms (GAs) [128, 129], and hybrid algorithms to accelerate convergence using artificial neural networks (ANNs) [128] and gradient based algorithms near local minima.

Our effort focuses on a first-principles-based strategy in order to provide not only the accuracy from finer grain calculations, but improved scalability and a seamless coupling. We exemplify this next with a coarse-grain approach for representing the dynamics of DNA.

4.1.1 Application Examples: Meso-Scale Dynamics of DNA

Our meso-scale model of DNA corresponds to a backbone-based structure, where three pseudo-atoms, with two or three pseudo-beads, represent each nucleotide. For each nucleotide, one pseudo-atom represents the phosphate group and another represents the sugar group of the sugar-phosphate backbone. There is then a pseudo-atom representing each of the different nucleosides (adenine, cytosine, guanine and thymine). Additionally, guanine and cytosine have three mass-less pseudo-beads, representing the three hydrogen bonding sites. Similarly, thymine and adenine have two pseudo-beads (see Fig. 14). These pseudo-beads move as rigid bodies with their parent pseudo-atom. This removes the need to define bond and angle terms for them.

Statistics of the potential of mean force (between sets of atoms representing each bead) is computed from various full atomistic DNA simulations. These distributions are then fitted to standard potentials in order to obtain the interaction parameters for every bead. The resulting force-field uniquely describes the interaction of each bead with every other bead in system, whose total energy is given by

$$E_{\text{Total}} = E_{\text{Bonds}} + E_{\text{Angles}} + E_{\text{Torsions}} + E_{\text{nonBond}} + E_{\text{hBond}}, \tag{16}$$

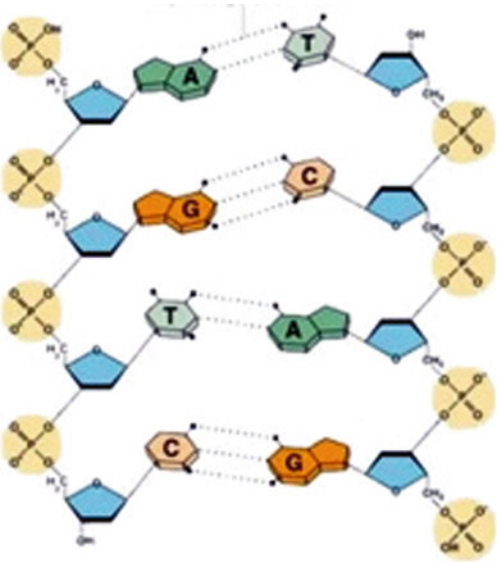

Fig. 14 Meso-scale model of DNA

where E_{Bonds}, E_{Angles}, and E_{Torsions} take the same form as those in Fig. 3, while the E_{nonBond} corresponds to a Morse type potential:

$$E_{\text{nonBond}} = D_0\left[\mathrm{e}^{-\alpha\left(-1+r_{ij}/r_0\right)} - 2\mathrm{e}^{-0.5\alpha\left(-1+r_{ij}/r_0\right)}\right], \tag{17}$$

and the E_{hBond}:

$$E_{\text{hBond}} = D_0\left[5\left(\frac{r_0}{r}\right)^{12} - 6\left(\frac{r_0}{r}\right)^{6}\right]\cos^4\theta, \tag{18}$$

where θ corresponds to the angle between a hydrogen and the corresponding donor–acceptor pair.

Coarse-Grain Parameter Optimization

The snapshots used when computing the nearest neighbor parameters are analyzed over 1,049 structures. For each snapshot, atoms comprising each pseudo-atom were determined. The center of mass of these atoms was then computed and the corresponding pseudo-atoms were placed at this center of mass.

Valence Parameters

Each snapshot was iteratively traversed and the two body bond lengths, three body angles, and four body dihedrals calculated from the positions of the pseudo-atoms.

These were tabulated and histograms of the distribution of these valence terms, for each unique combination of pseudo-atoms, were generated (see Fig. 15a). These distributions were shown to be a single-model Gaussian distribution for the bonds and angles, and multi-modal for dihedrals. The only exception occurred when considering the bond distribution between the phosphate (PHO) and sugar (SUG) pseudo-atoms, which was bi-modal. This anomaly is explained by considering that starting at the 5' end, the SUG –> PHO bond length (4.4 Å) is longer than the next PHO –> SUG bond length (4.0 Å). When constructing the meso-model from the atomistic representation, if the backbone atoms are alternatively labeled as PHO-SUG-PHP-SUS-PHO, this anomaly is effectively resolved and the PHO-SUG, PHO-SUS, PHP-SUG, and PHP-SUS distributions are uni-modal.

The parameters were obtained by performing a least squares fitting of a Gaussian curve of the desired functional form. The fitting function is an exponential with exponent equal to the relevant potential/2RT. For the bonds, the fitting function was therefore $y = a_0 e^{-0.5\kappa(r-r_0)^2/2RT}$. Table 2 gives the parameters obtained for the various bond-stretch terms.

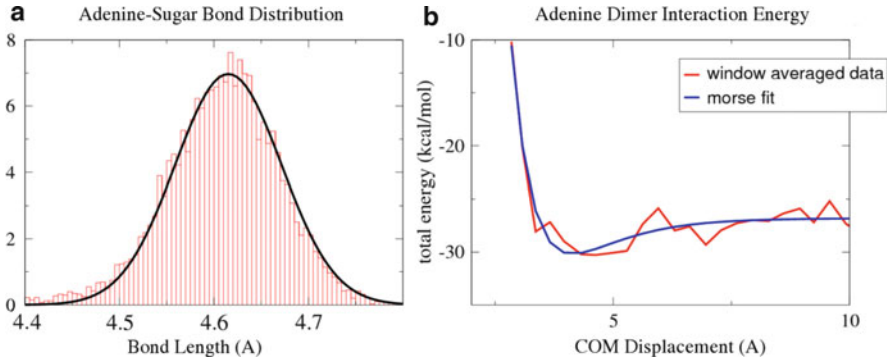

Fig. 15 Plots of (**a**) adenine–sugar bond distribution from atomistic MD and (**b**) adenine dimer interaction energy. This includes the binding energy of the dimer as well as the interaction with the water molecules

Table 2 Bond parameters for pseudo-beads

	k_0 (kcal mol^{-1} Å^{-2})	r_0 (Å)
SUG-THY	167.848	4.2136
SUG-ADE	155.974	4.6175
SUG-GUA	143.419	4.7412
SUG-CYT	193.31	4.1101
SUG-PHO	46.0059	4.0053
PHP-SUG	18.2605	4.4283

Non-Bond Parameters

Two copies of the atoms comprising each pseudo-atom were obtained, displaced by 4 Å in the *z*-direction and individually immersed in a pre-equilibrated box of TIP3 water. The charges were maintained, and hence the only set of atoms with a net charge were the five atoms comprising the phosphate (PHO/PHP) bead, with a net −1 charge per set of atoms. If necessary, the system was neutralized with Na+ ions and subjected to the same simulation protocol as before. First 100 ps of NPT dynamics was performed, with snapshots of the last 20 ps (every 2 ps) saved. The ten snapshots were then superimposed on each other and the average MD solution structure obtained as previously discussed.

A constrained MD simulation was then run, where the center of mass of each of the two sets of atoms are constrained by a harmonic potential. Every time step, the distance between the two centers of mass is checked, and if they are not equal to a specified distance, a restoring spring of magnitude $-k\Delta x$ is applied to all of the atoms in the second set. Here, Δx is the differential distance from equilibrium and k is a force constant of 500 kcal mol^{-1} Å^{-1}. The center of mass of the two sets of atoms are constrained, starting at 2.0 Å, in 0.1 Å increments, until 10 Å. After each increment, 40 ps of NPT dynamics is performed to equilibrate the structure at the restraint, followed by a further 20 ps of NPT dynamics, during which time the non-bond energy of both sets of atoms is calculated and tabulated.

The average energy per center of mass separation was then computed and plotted vs the center of mass (Fig. 15b). This function was then shifted so that its plateau was at 0, and fitted to the Morse potential in (17) using the least squares fitting procedure previously discussed. This fitting is the effective potential of mean force between both sets of atoms, in the presence of water. This potential has both electrostatic and van der Waals contributions, which is critical since none of the pseudo-atoms are charged.

Meso-Scale Simulation of B-DNA Dodecamer

Simulation Protocol

Bond stretches are the highest frequency modes in any MD simulation. The largest time step of any MD simulation is inversely proportional to the highest frequency modes, which is related to the force constant by $f = 1/2\pi\sqrt{k/\mu}$ (assuming harmonic bonds), where μ is the reduced mass of the parameter atoms being considered. A rule of thumb is that the largest time step possible should be 1/(6*f). The pseudo-atoms of this meso-scale force field are much heavier than those of regular atoms and the largest bond force constant is one-fifth that of the typical largest atomistic force constant, which allowed the use of a 10 fs timestep. Furthermore, the meso-scale system contains fewer particles than its atomistic counterpart (each nucleotide is composed of 3 pseudo-atoms, compared to about 40 for an atomistic nucleotide) and it avoids electrostatic interactions because the pseudo-atoms were

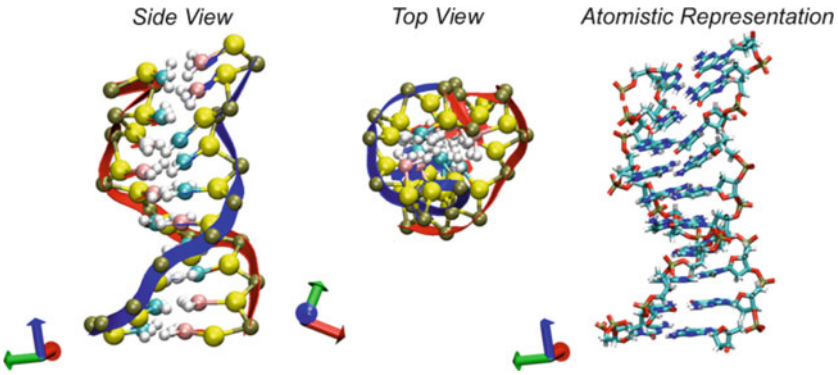

Fig. 16 Meso-dodecamer (*left* and *center*) and its atomistic reconstructured model (*right*)

chosen to be neutral, which led to an overall increase in the times spanned during simulations.

A 1.4-Å crystal structure of the B-DNA dodecamer (355D) was solvated in a TIP3 water box and neutralized. This fully atomistic structure was simulated using the usual procedure for 2 ns of NPT dynamics, with snapshots of the last 500 ps (every 5 ps) saved for thermodynamic and structural analysis. The average MD structure was calculated as before and the ions were removed. It was then converted into its mesoscale description, with the waters modeled as inflated van der Waals spheres. This water model is the same as the M3B model from our previous work in Molinero et al. [124], and was parameterized to reproduce the density, diffusivity, and cohesive energy of experimental bulk water. The mesoscale dodecamer is then simulated for 2×10^6 steps, representing 2 μs of total simulation time. Snapshots of the system during the last 50 ns of simulation were saved and used for thermodynamic and structural analysis. The average MD structure during the last 50 ns was calculated and the atomistic level description reconstructed from this average structure. This reconstructed atomistic structure was then minimized and simulated for 1,000 steps of NPT dynamics (Fig. 16).

Comparison of Meso-Scale and Atomistic Dodecamer Simulations

Timing tests indicate that the atomistic level simulation took twice as long to complete 2 ns than the meso-scale did to complete 2 μs (150 vs 69 CPU hours). This is remarkable, and represents a 2,000× speedup for the meso-scale model. This opens the door for studying DNA system in the micro-second timescale.

As a measure of the similarities between the meso, atomistic, and crystal structures, we calculate the CRMS. We find that the simulation structures are quite different from the crystal structure (5.5 Å and 3.2 Å for meso and atomistic, respectively), as well as from each other (4.7 Å difference between the two). This is further illustrated in Fig. 17, where the average MD structures of both the meso-scale and the atomistic structures are compared. A smooth backbone profile can be

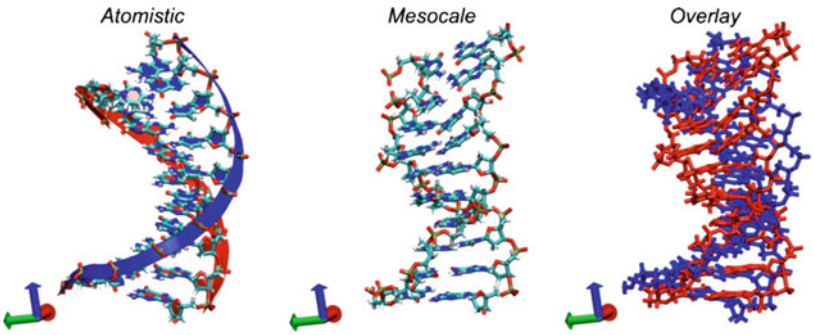

Fig. 17 Comparison of average MD structures between meso-scale and atomistic

observed in the case of the atomistic DNA, contrasted with the irregular profile for the meso-scale case. The overlay of the two structures illustrated the sharp differences in the backbone torsions of the atomistic (blue) and the meso scale (red) structures. A considerably larger distortion can be seen for the meso structure, in particular the presence of several backbone kinks. The backbone parameters for both sets of simulation are within the acceptable range of B-DNA, except for the backbone twist, where the average MD structures are under-twisted (this is a known issue with the AMBER force field). There is also significant deviation in the backbone twist for the meso dodecamer, which is a consequence of the aforementioned kinks in the structure.

From these helical analyses it is clear that further optimization of the meso-scale force field is required. In particular, the Phosphate-Sugar-Phosphate backbone twist angle and the Phosphate-Sugar-Phosphate-Sugar backbone dihedral angles need to be optimized to prevent the under-twisting of the helix seen in the dodecamer simulation. The helical rise in the meso-scale dodecamer is outside the accepted range for B-DNA (3.4 ± 0.2) which, when combined with the under-twisting of the helix, points to unwinding of the helix.

The following section describes the application of QM-parameterized atomistic and coarse-grain potentials to tissue engineering.

4.1.2 Application Example: Mechanoregulation in Polymer-Based Hydrogel Networks for Tissue Engineering

We are currently studying the nanoscale properties that lead to improved micro-scale mechanoregulatory response of polymer-based hydrogel networks for cartilage tissue scaffolding. The structural similarity of synthetic polymer-based hydrogels to the collagen ECM found in human cartilage and their ease of processability makes them ideal candidates for cartilage scaffold-supported cell therapies [131]. These hydrogels can act as a temporary artificial extra cellular matrix (ECM) to provide mechanical support, or provide the ambience for new cells to grow towards

the desired lineage, as well as enable transport or confinement of cells to/within defect sites. Yet, in spite of the significant progress in polymer-based tissue engineering during the past decade, important challenges remain to be addressed in order to restore tissues that serve a predominantly biomechanical function [132–134]. Recent experiments demonstrate the existence of high strength double network polymer-based combinations [135] that may serve as potential candidates for cartilage tissue scaffold development, yet little is known about the atomic composition and the nanometric structures that lead to the enhanced response. We have developed a multiscale approach to characterize the nature of this phenomena [130], and to provide an in silico framework for developing improved materials for cell-therapies with high mechanical loading requirements.

Our approach to characterizing the mechanisms and for tuning the mechanical response of polymer-based hydrogels involves a first principles QM-derived Dreiding force field to investigate the thermodynamic and composition conditions for the gel point near the Flory–Stockmayer transition in selected double network hydrogel combinations (poly(acrylamide)-PAAm and poly(2-acrylamido-2-methylpropanesulfonic acid)-PAMPS), with *N*,*N*′-methylene bisacrylamide (NNMBA), and a first-principles coarse-grain model for describing the long-term dynamics of percolation based on composition and density ratios between the polymeric and crosslinking agents. We found that percolation events at particular threshold ratios (between polymer components and crosslinker concentrations) are a key atomistic mechanism to promote enhanced mechanical strength in crosslinked hydrogel networks. We used atomic charges and torsional potential energy curves from QM to calculate single chain statistics (including radius of gyration as a function of degree of polymerization). Substituting the radius of gyration in an adjusted continuum model for percolation from [136] led to an estimate of the number of polymer molecules required to achieve percolation using the critical number density found from our atomistic simulations (see Fig. 18). Using our constrained MD with a coarse-grain potential based on a QM parameterized finite extensible nonlinear elastic (FENE) [137], we have been able to determine percolation thresholds as a function of composition (i.e., solvent, crosslinking agent concentration, molar ratio between polymer component, among others) on single and double solvated networks. Figure 19 shows two snapshots of our coarse-grain representation, on the left a set of linear poly(acrylamide) chains mixed at a particular proportion with solvent (water), starter, and crosslinking molecules, and on the right the appearance after running MD-NVE (from a starting temperature of 300K) of a percolated structure of poly(acrylamide). Our results are consistent with experimentally measured gel points and help explain the precipitous loss of the high fracture energy in double network hydrogels at particular crosslink densities [138]. We have also confirmed the mechanical strengthening process via calculation of elastic constants and viscoelastic response of the final networked structures, relative to the single network components. These results will be reported in a separate publication. We were able to determine the critical cross-link concentrations as a function of starting monomer concentrations and degree of polymerization required for improved

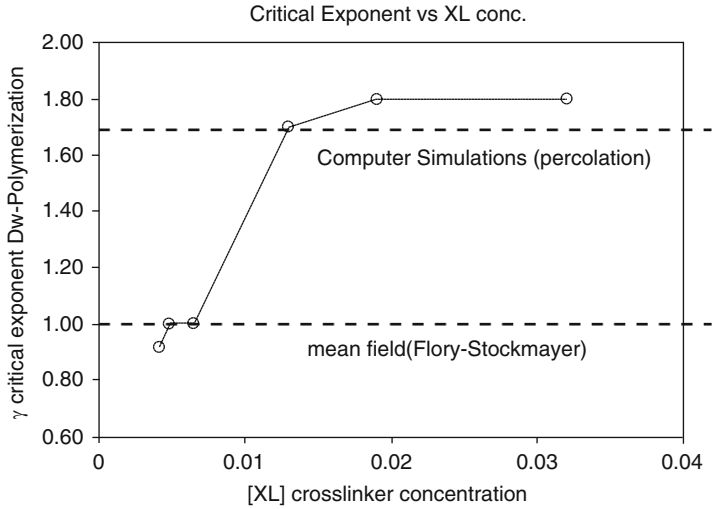

Fig. 18 Critical-exponent for the degree of polymerization in AAM gel networks. From [130]

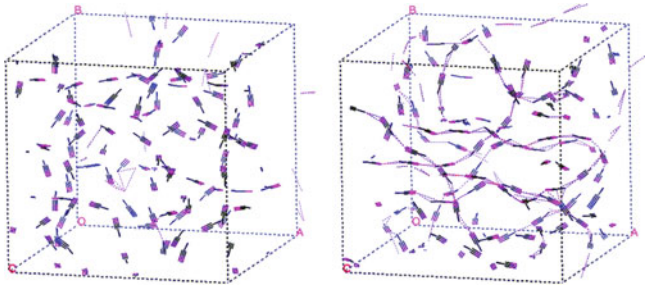

Fig. 19 Snapshots of coarse-grain unit cell model of PAAm and crosslinker mixture (*left*), and PAAm percolated structure after a few ns of MD-NVT (*right*)

mechanical strength. These findings suggest a strategy for systematic characterization and optimization of polymer-based hydrogel molecular network architectures, using design-level tuning of the structural and compositional variables that correlate to their mechanical and thermodynamic response.

5 Concluding Remarks

Seamless integration of paradigms and simulation time- and length-scales, using first-principles-based methods poses tremendous challenges that go far beyond parameter coupling between methods and scales, requiring new physical models

and novel computational schemes capable of systematic upscaling or downscaling. Nevertheless, the hierarchical strategy we have followed has spurred important breakthroughs in the development of methods that bridge paradigms and scales, without sacrificing the ability to model complex chemical processes, including the mixed QM-APBS scheme to understand the solvation effects on molecular reactivity, our reactive and non-adiabatic explicit eFF, reaxFF and eFF, and our efficient coarse-grain methods to bridge, using finer details from atomistic force fields, into the meso scale for conformational studies of systems that are defined through complex chemical interactions (including H bonding). We have demonstrated how these methods are applied to problems that have been untreatable using conventional QM or MM/MD methods.

As described for our simulations of CNT growth, ReaxFF extends our ab initio knowledge of the reactivity of small, model systems to complex, extended phenomena involving thousands of atoms throughout nanoseconds of dynamics. This sets us closer to the goal of elucidating and optimizing the fundamental mechanisms that control the chirality of CNTs during their growth. Our results using the eFF methodology, on the hypervelocity shock-induced and electron-induced chemistry of materials, demonstrate its unique ability to depict and predict accurately processes that are highly dependent on electronic contributions, enabling us to understand the real-time dynamics of systems with excited electronic states, all of which is beyond existing QM, QM-approximations (e.g., tight-binding), or MM/MD methods, and in some cases, experimental observation.

Last but not least, our parameterization of non-reactive atomistic and coarse-grain force fields has proven invaluable to understanding and predicting the structure and dynamics of conformations of large biological complexes, such as GPCRs and DNA, both of which lay beyond the capabilities of experiments, considering GPCRs have not been crystallized in their native form (embedded in the cellular membrane), and that activation and critical conformation events in these macromolecules stretch to the micro-second, and beyond, timescales.

We also described the use of our coarse-grain and rigid body MD approaches to elucidate a long-standing problem, associated with the chemical processes that lead to enhanced mechanical strength in polymer-based double network hydrogels (i.e., percolation thresholds as a function of polymer molar ratios and cross-linking concentrations near the Flory–Stockmayer transition limit in polymers gels), of significant importance in the development of scaffolds for load-bearing tissue engineering (e.g., cartilage).

In spite of our progress in methods development and the advent of more powerful computing resources, in-silico de novo design and optimization will still have to rely on the correct chemical and physical insight from modelers, especially considering the combinatorial explosion of interactions and the increased number of degrees of freedom as one approaches the continuum macro-scale. As we have shown, a first-principles-based approach does not need to propagate all degrees of freedom during upscaling, but those that are pertinent to the properties and phenomena of interest.

Acknowledgements The material on single shock Hugoniot is based upon work supported by the Department of Energy's National Nuclear Security Administration under Award Number DE-FC52-08NA28613. The material on hydrogel mechanics for tissue engineering scaffolding is based upon work supported by the National Science Foundation (CMMI 0727870). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author/s and do not necessarily reflect the views of the National Science Foundation. The early developments of these multiscale multiparadigm methods were initiated with support by DOE under the ECUT program (Prog. Mgr. Minoo Dastoor) and continued DAPRA under the PROM and ONR programs (Prog. Mgr. Carey Swartz, Judah Goldwasser, and Steve Wax). Substantial support was provided by ARO-MURI, ONR-MURI, DURIP, and ASCI projects along with Chevron, Dow-Corning, Aventis, Asahi Kasei, Intel, PharmSelex, and many other industrial labs.

References

1. de Broglie L (1924) Recherches sur la théorie des quanta
2. Schrodinger E (1926) Quantification of the eigen-value problem. Ann Phys 79(6):489–527
3. Messiah A (ed) (1999) Quantum mechanics, vol 1. Reprinted by Dover Publications, Mineola, NY
4. Anderson JB (1976) Quantum chemistry by Random-Walk – H2p, H+3d3h1a'1, H-23-Sigma+U, H-41-Sigma+G, Be1s. J Chem Phys 65(10):4121–4127
5. Williamson AJ, Hood RQ, Grossman JC (2001) Linear-scaling quantum Monte Carlo calculations. Phys Rev Lett 8724(24):246406-(4)
6. Reboredo FA, Williamson AJ (2005) Optimized nonorthogonal localized orbitals for linear scaling quantum Monte Carlo calculations. Phys Rev B 71(12):121105-(4)
7. Morokuma K et al (2001) Model studies of the structures, reactivities, and reaction mechanisms of metalloenzymes. Ibm J Res Dev 45(3–4):367–395
8. Fisher DR et al (2008) An optimized initialization algorithm to ensure accuracy in quantum Monte Carlo calculations. J Comput Chem 29(14):2335–2343
9. Anderson AG, Goddard WA, Schroder P (2007) Quantum Monte Carlo on graphical processing units. Comput Phys Commun 177(3):298–306
10. Anderson AG, Goddard WA (2010) Generalized valence bond wave functions in quantum Monte Carlo. J Chem Phys 132(16):164110-(10)
11. Born M, Oppenheimer R (1927) Quantum theory of molecules. Ann Phys 84(20):0457–0484
12. Frisch MJ et al (2004) Gaussian. Gaussian, Inc., Wallingford CT
13. Schmidt MW et al (1993) General atomic and molecular electronic-structure system. J Comput Chem 14(11):1347–1363
14. Jaguar (1991–2000) Jaguar. Schrodinger, Inc., Portland, OR
15. Kresse G, Hafner J (1993) Ab initio molecular-dynamics for liquid-metals. Phys Rev B 47(1):558–561
16. Dovesi R et al (2006) Crystal, U.o. Torino, Editor. 2006: Torino
17. Segall MD et al (2002) First-principles simulation: ideas, illustrations and the CASTEP code. J Phys Condens Matter 14(11):2717–2744
18. Schultz PA (2007) SeqQuest electronic structure code. Sandia National Laboratories, Albuquerque
19. Harvey J (2001) Molecular electronic structure. Available from: http://www.chm.bris.ac.uk/pt/harvey/elstruct/beyond_hf.html
20. Moller C, Plesset MS (1934) Note on an approximation treatment for many-electron systems. Phys Rev 46(7):0618–0622
21. Carter EA, Goddard WA (1988) Correlation-consistent configuration-interaction – accurate bond-dissociation energies from simple wave-functions. J Chem Phys 88(5):3132–3140

22. Friesner RA (2005) Ab initio quantum chemistry: methodology and applications. Proc Natl Acad Sci U S A 102(19):6648–6653
23. Goddard WA et al (1973) Generalized valence bond description of bonding in low-lying states of molecules. Acc Chem Res 6(11):368–376
24. Greeley BH et al (1994) New pseudospectral algorithms for electronic-structure calculations – length scale separation and analytical 2-electron integral corrections. J Chem Phys 101(5): 4028–4041
25. Tannor DJ et al (1994) Accurate first principles calculation of molecular charge-distributions and solvation energies from ab-initio quantum-mechanics and continuum dielectric theory. J Am Chem Soc 116(26):11875–11882
26. Bredow T, Jug K (2005) Theory and range of modern semiempirical molecular orbital methods. Theor Chem Acc 113(1):1–14
27. Hohenberg P, Kohn W (1964) Phys Rev B 136(3):864
28. Kohn W, Sham LJ (1965) Phys Rev A 140(4):1133
29. Li ZY, He W, Yang JL (2005) Recent progress in density functional theory and its numerical methods. Prog Chem 17(2):192–202
30. Foulkes WMC et al (2001) Quantum Monte Carlo simulations of solids (review article). Rev Mod Phys 73:33
31. Chen XJ, Langlois JM, Goddard WA (1995) Dual-space approach for density-functional calculations of 2-dimensional and 3-dimensional crystals using Gaussian-basis functions. Phys Rev B 52(4):2348–2361
32. Liu Y, Goddard WA (2009) A universal damping function for empirical dispersion correction on density functional theory. Mater Trans 50(7):1664–1670
33. Dobson JC, Meyer TJ (1988) Redox properties and ligand loss chemistry in aqua hydroxo complexes derived from CIS-(BPY)2RU2(OH2)2 2+ and TRANS- (BPY)2RUII(OH2) 2 2+. Inorg Chem 27(19):3283–3291
34. Hill T (1960) An introduction to statistical thermodynamics. Addison-Wesley, Reading
35. Tissandier MD et al (1998) The proton's absolute aqueous enthalpy and Gibbs free energy of solvation from cluster-ion solvation data. J Phys Chem A 102(40):7787–7794
36. Kelly CP, Cramer CJ, Truhlar DG (2006) Aqueous solvation free energies of ions and ion-water clusters based on an accurate value for the absolute aqueous solvation free energy of the proton. J Phys Chem B 110(32):16066–16081
37. Srnec M et al (2008) Effect of spin-orbit coupling on reduction potentials of octahedral ruthenium(II/III) and osmium(II/III) complexes. J Am Chem Soc 130(33):10947–10954
38. Car R, Parrinello M (1985) Unified approach for molecular-dynamics and density-functional theory. Phys Rev Lett 55(22):2471–2474
39. Mayo SL, Olafson BD, Goddard WA III (1990) DREIDING: a generic force field for molecular simulations. J Phys Chem 94:8897–8909
40. Ding HQ, Karasawa N, Goddard WA (1992) Atomic level simulations on a million particles – the cell multipole method for Coulomb and London nonbond interactions. J Chem Phys 97(6):4309–4315
41. Ding HQ, Karasawa N, Goddard WA (1992) The reduced cell multipole method for Coulomb interactions in periodic-systems with million-atom unit cells. Chem Phys Lett 196(1–2):6–10
42. Ponder JW, Case DA (2003) Force fields for protein simulations. Protein Simul 66:27–85
43. Weiner SJ et al (1986) An all atom force-field for simulations of proteins and nucleic-acids. J Comput Chem 7(2):230–252
44. MacKerell AD et al (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102(18):3586–3616
45. Jorgensen WL, Maxwell DS, TiradoRives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118(45):11225–11236
46. Karplus M, McCammon JA (2002) Molecular dynamics simulations of biomolecules. Nat Struct Biol 9(9):646–652

47. Lagerstrom MC, Schioth HB (2008) Structural diversity of G protein-coupled receptors and significance for drug discovery (vol 7, p 339, 2008). Nat Rev Drug Discov 7(6):542
48. Lagerstrom MC, Schioth HB (2008) Structural diversity of G protein-coupled receptors and significance for drug discovery. Nat Rev Drug Discov 7(4):339–357
49. Kam VWT, Goddard WA (2008) Flat-bottom strategy for improved accuracy in protein side-chain placements. J Chem Theory Comput 4(12):2160–2169
50. van Duin ACT et al (2001) ReaxFF: a reactive force field for hydrocarbons. J Phys Chem A 105(41):9396–9409
51. Su JT, Goddard WA (2007) Excited electron dynamics modeling of warm dense matter. Phys Rev Lett 99(18):185003
52. Strachan A et al (2005) Thermal decomposition of RDX from reactive molecular dynamics. J Chem Phys 122(5):054502
53. van Duin ACT et al (2003) ReaxFF(SiO) reactive force field for silicon and silicon oxide systems. J Phys Chem A 107(19):3803–3811
54. Han SS et al (2005) Optimization and application of lithium parameters for the reactive force field, ReaxFF. J Phys Chem A 109(20):4575–4582
55. Zhang Q et al (2004) Adhesion and nonwetting-wetting transition in the Al/alpha-Al2O3 interface. Phys Rev B 69(4):045423-(11)
56. Nielson KD et al (2005) Development of the ReaxFF reactive force field for describing transition metal catalyzed reactions, with application to the initial stages of the catalytic formation of carbon nanotubes. J Phys Chem A 109(3):493–499
57. Cheung S et al (2005) ReaxFF(MgH) reactive force field for magnesium hydride systems. J Phys Chem A 109(5):851–859
58. Chen N et al (2005) Mechanical properties of connected carbon nanorings via molecular dynamics simulation. Phys Rev B 72(8):085416-(9)
59. Su HB et al (2007) Simulations on the effects of confinement and Ni-catalysis on the formation of tubular fullerene structures from peapod precursors. Phys Rev B 75(13): 134107-(5)
60. Chenoweth K et al (2005) Simulations on the thermal decomposition of a poly(dimethylsiloxane) polymer using the ReaxFF reactive force field. J Am Chem Soc 127(19):7192–7202
61. Strachan A et al (2003) Shock waves in high-energy materials: the initial chemical events in nitramine RDX. Phys Rev Lett 91(9):098301-(4)
62. van Duin ACT et al (2005) Atomistic-scale simulations of the initial chemical events in the thermal initiation of triacetonetriperoxide. J Am Chem Soc 127(31):11053–11062
63. Buehler MJ, van Duin ACT, Goddard WA (2006) Multiparadigm modeling of dynamical crack propagation in silicon using a reactive force field. Phys Rev Lett 96(9):095505-(4)
64. Goddard WA et al (2006) Development of the ReaxFF reactive force field for mechanistic studies of catalytic selective oxidation processes on BiMoOx. Top Catal 38(1–3):93–103
65. Ludwig J et al (2006) Dynamics of the dissociation of hydrogen on stepped platinum surfaces using the ReaxFF reactive force field. J Phys Chem B 110(9):4274–4282
66. Iijima S (1991) Helical microtubules of graphitic carbon. Nature 354(6348):56–58
67. Endo M, Strano MS, Ajayan PM (2008) Potential applications of carbon nanotubes. Carbon Nanotubes 111:13–61
68. Kreupl F (2008) Carbon nanotubes in microelectronic applications. In: Hierold C (ed) Carbon nanotube devices, 2nd edn. Wiley VCH, Weinheim
69. Stampfer C (2008) Electromechanical carbon nanotube transducers. In: Hierold C (ed) Carbon nanotube devices, 2nd edn. Wiley VCH, Weinheim
70. Roman C (2008) Modeling the properties of carbon nanotubes for sensor-based devices. In: Hierold C (ed) Carbon nanotube devices, 2nd edn. Wiley VCH, Weinheim
71. Robertson J (2008) Carbon nanotube field emission devices. In: Hierold C (ed) Carbon nanotube devices, 2nd edn. Wiley VCH, Weinheim
72. Yeow JT (2008) Carbon nanotube gas sensors. In: Hierold C (ed) Carbon nanotube devices, 2nd edn. Wiley VCH, Weinheim

73. Joselevich E (2008) Carbon nanotube synthesis and organization. In: Jorio A, Dresselhaus G, Dresselhaus M (eds) Carbon nanotubes, 2nd edn. Springer, Berlin
74. Bolton K et al (2009) Density functional theory and tight binding-based dynamical studies of carbon-metal systems of relevance to carbon nanotube growth. Nano Res 2(10):774–782
75. Gavillet J et al (2001) Root-growth mechanism for single-wall carbon nanotubes. Phys Rev Lett 87(27):275504-(4)
76. Huang SM et al (2004) Growth mechanism of oriented long single walled carbon nanotubes using "fast-heating" chemical vapor deposition process. Nano Lett 4(6):1025–1028
77. Li YM et al (2001) Growth of single-walled carbon nanotubes from discrete catalytic nanoparticles of various sizes. J Phys Chem B 105(46):11424–11431
78. Helveg S et al (2004) Atomic-scale imaging of carbon nanofibre growth. Nature 427(6973): 426–429
79. Hofmann S et al (2007) In situ observations of catalyst dynamics during surface-bound carbon nanotube nucleation. Nano Lett 7(3):602–608
80. Raty JY, Gygi F, Galli G (2005) Growth of carbon nanotubes on metal nanoparticles: a microscopic mechanism from ab initio molecular dynamics simulations. Phys Rev Lett 95(9):096103-(4)
81. Abild-Pedersen F et al (2006) Mechanisms for catalytic carbon nanofiber growth studied by ab initio density functional theory calculations. Phys Rev B 73(11):115419-(13)
82. Amara H et al (2009) Tight-binding potential for atomistic simulations of carbon interacting with transition metals: application to the Ni-C system. Phys Rev B 79(1):014109-(17)
83. Ohta Y et al (2008) Rapid growth of a single-walled carbon nanotube on an iron cluster: density-functional tight-binding molecular dynamics simulations. Acs Nano 2(7):1437–1444
84. Moors M et al (2009) Early stages in the nucleation process of carbon nanotubes. Acs Nano 3(3):511–516
85. Mueller JE, van Duin ACT, Goddard WA (2010) Application of the ReaxFF reactive force field to reactive dynamics of hydrocarbon chemisorption and decomposition. J Phys Chem C 114(12):5675–5685
86. Mueller JE, van Duin ACT, Goddard WA (2010) Development and validation of ReaxFF reactive force field for hydrocarbon chemistry catalyzed by nickel. J Phys Chem C 114(11): 4939–4949
87. Mora E et al (2008) Low-temperature single-wall carbon nanotubes synthesis: feedstock decomposition limited growth. J Am Chem Soc 130(36):11840–11841
88. Hofmann S et al (2005) Surface diffusion: the low activation energy path for nanotube growth. Phys Rev Lett 95(3):036101-(4)
89. Henkelman G, Jonsson H (2001) Long time scale kinetic Monte Carlo simulations without lattice approximation and predefined event table. J Chem Phys 115(21):9657–9666
90. Jaramillo-Botero A et al (2010) Large-scale, long-term non-adiabatic electron molecular dynamics for describing material properties and phenomena in extreme environments. J Comput Chem. epub ahead of print.
91. Su JT, Goddard WA (2009) The dynamics of highly excited electronic systems: applications of the electron force field. J Chem Phys 131(24):244501-(20)
92. Gillis HP et al (1995) Low-energy electron-enhanced etching of Si(100) in hydrogen helium direct-current plasma. Appl Phys Lett 66(19):2475–2477
93. Su JT, Goddard WA (2009) Mechanisms of Auger-induced chemistry derived from wave packet dynamics. Proc Natl Acad Sci U S A 106(4):1001–1005
94. Goddard W III (1998) Nanoscale theory and simulation: a critical driver for and a critical challenge to commercial nanotechnology. In: WTEC workshop. World Technology Evaluation Center, Arlington, VA
95. Jaramillo-Botero A et al (2008) Multiscale-multiparadigm modeling and simulation of nanometer scale systems and processes for nanomedical applications. In: Zhang M, Xi N (eds) Nanomedicine: a systems engineering approach. Pan Stanford Publishing, Singapore

96. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical-integration of Cartesian equations of motion of a system with constraints – molecular-dynamics of n-alkanes. J Comput Phys 23(3):327–341
97. Van Gunsteren WF, Berendsen HJC (1977) Algorithms for macromolecular dynamics and constraint dynamics. Mol Phys 34(5):1311–1327
98. Andersen HC (1983) Rattle – a velocity version of the shake algorithm for molecular-dynamics calculations. J Comput Phys 52(1):24–34
99. Krautler V, Van Gunsteren WF, Hunenberger PH (2001) A fast SHAKE: algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. J Comput Chem 22(5):501–508
100. Hess B et al (1997) LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18(12):1463–1472
101. Miyamoto S, Kollman PA (1992) Settle – an analytical version of the shake and rattle algorithm for rigid water models. J Comput Chem 13(8):952–962
102. Bae DS, Haug EJ (1987) A recursive formulation for constrained mechanical system dynamics. 1. Open loop-systems. Mech Struct Mach 15(3):359–382
103. Abagyan RA, Mazur AK (1989) New methodology for computer-aided modeling of biomolecular structure and dynamics. 2. Local deformations and cycles. J Biomol Struct Dyn 6(4): 833–845
104. Mazur AK, Abagyan RA (1989) New methodology for computer-aided modeling of biomolecular structure and dynamics. 1. Non-cyclic structures. J Biomol Struct Dyn 6(4):815–832
105. Bae DS, Kuhl JG, Haug EJ (1988) A recursive formulation for constrained mechanical system dynamics. 3. Parallel processor implementation. Mech Struct Mach 16(2):249–269
106. Bae DS, Haug EJ (1988) A recursive formulation for constrained mechanical system dynamics. 2. Closed-loop systems. Mech Struct Mach 15(4):481–506
107. Jaramillo-Botero A, Lorente ACI (2002) A unified formulation for massively parallel rigid multibody dynamics of O(log(2) n) computational complexity. J Parallel Distrib Comput 62 (6):1001–1020
108. Vaidehi N, Jain A, Goddard WA (1996) Constant temperature constrained molecular dynamics: The Newton-Euler inverse mass operator method. J Phys Chem 100(25): 10508–10517
109. Jain A, Vaidehi N, Rodriguez G (1993) A fast recursive algorithm for molecular-dynamics simulation. J Comput Phys 106(2):258–268
110. Mathiowetz AM et al (1994) Protein simulations using techniques suitable for very large systems – the cell multipole method for nonbond interactions and the Newton-Euler inverse mass operator method for internal coordinate dynamics. Proteins Struct Funct Genet 20 (3):227–247
111. Jaramillo-Botero A, Liu Y, Goddard WA (2006) The computational materials design facility: CMDF. Available from: http://www.wag.caltech.edu/multiscale
112. Featherstone R (1983) The calculation of robot dynamics using articulated-body inertias. Int J Rob Res 2(1):13–30
113. Fijany A et al (1998) Novel algorithms for massively parallel, long-term, simulation of molecular dynamics systems. Adv Eng Softw 29(3–6):441–450
114. Shelley J et al (2001) A coarse grain model for phospholipid simulations. J Phys Chem B 105 (19):4464–4470
115. Groot RD (2000) Mesoscopic simulation of polymer-surfactant aggregation. Langmuir 16 (19):7493–7502
116. Groot RD, Rabone KL (2001) Mesoscopic simulation of cell membrane damage, morphology change and rupture by nonionic surfactants. Biophys J 81(2):725–736
117. Jedlovszky P (1998) Investigation of the orientational correlation of the molecules in liquid H2S with reverse Monte Carlo simulation. Mol Phys 93(6):939–946
118. Jedlovszky P et al (1996) Investigation of the uniqueness of the reverse Monte Carlo method: studies on liquid water. J Chem Phys 105(1):245–254

119. Lopez CF et al (2002) Computer simulation studies of biomembranes using a coarse grain model. Comput Phys Commun 147(1–2):1–6
120. Marrink SJ, de Vries AH, Mark AE (2004) Coarse grained model for semiquantitative lipid simulations. J Phys Chem B 108(2):750–760
121. Marrink SJ, Mark AE (2003) Molecular dynamics simulation of the formation, structure, and dynamics of small phospholipid vesicles. J Am Chem Soc 125(49):15233–15242
122. Marrink SJ, Mark AE (2004) Molecular view of hexagonal phase formation in phospholipid membranes. Biophys J 87(6):3894–3900
123. Marrink SJ et al (2007) The MARTINI force field: Coarse grained model for biomolecular simulations. J Phys Chem B 111(27):7812–7824
124. Molinero V, Goddard WA (2004) M3B: a coarse grain force field for molecular simulations of malto-oligosaccharides and their water mixtures. J Phys Chem B 108(4):1414–1427
125. Vaidehi N, Goddard WA (2000) Domain motions in phosphoglycerate kinase using hierarchical NEIMO molecular dynamics simulations. J Phys Chem A 104(11):2375–2383
126. Cagin T et al (2001) Multiscale modeling and simulation methods with applications to dendritic polymers. Comput Theor Polym Sci 11(5):345–356
127. Elezgaray J, Laguerre M (2006) A systematic method to derive force fields for coarse-grained simulations of phospholipids. Comput Phys Commun 175(4):264–268
128. Hunger J, Huttner G (1999) Optimization and analysis of force field parameters by combination of genetic algorithms and neural networks. J Comput Chem 20(4):455–471
129. Hunger J et al (1998) How to derive force field parameters by genetic algorithms: modelling tripod-Mo(CO)(3) compounds as an example. Eur J Inorg Chem 6:693–702
130. Jaramillo-Botero A et al (2010) First-principles based approaches to nano-mechanical and biomimetic characterization of polymer-based hydrogel networks for cartilage scaffold-supported therapies. J Comput Theor Nanosci 7(7):1238–1256
131. Varghese S, Elisseeff JH (2006) Hydrogels for musculoskeletal tissue engineering. In: Polymers for regenerative medicine. Springer-Verlag, Berlin, pp 95–144
132. Butler DL, Goldstein SA, Guilak F (2000) Functional tissue engineering: the role of biomechanics. J Biomech Eng Trans Asme 122(6):570–575
133. Guilak F (2000) The deformation behavior and viscoelastic properties of chondrocytes in articular cartilage. Biorheology 37(1–2):27–44
134. Guilak F, Butler DL, Goldstein SA (2001) Functional tissue engineering – the role of biomechanics in articular cartilage repair. Clin Orthop Relat Res 391:S295–S305
135. Gong JP et al (2003) Double-network hydrogels with extremely high mechanical strength. Adv Mater 15(14):1155–1158
136. Lorenz CD, Ziff RM (2001) Precise determination of the critical percolation threshold for the three-dimensional "Swiss cheese" model using a growth algorithm. J Chem Phys 114(8): 3659–3661
137. Kremer K, Grest GS (1990) Dynamics of entangled linear polymer melts – a molecular-dynamics simulation. J Chem Phys 92(8):5057–5086
138. Blanco M, Jaramillo-Botero A, Goddard W III (2010) The percolation limit near the Flory-Stockmayer transition in polymer hydrogel networks. California Institute of Technology, Pasadena