ReaxFF/AMBER for Reactive Molecular Dynamics Simulations

Ali Rahnamoun[†], Mehmet Cagri Kaymak[‡], Madushanka Manathunga[†], Andreas W. Götz[ǂ],
Adri C. T. van Duin[ł], Kenneth M. Merz, Jr.[†,*] and Hasan Metin Aktulga[‡,*]

[†]Department of Chemistry and Department of Biochemistry and Molecular Biology, Michigan State University, 578 S. Shaw Lane, East Lansing, Michigan 48824-1322, United States

[‡]Department of Computer Science and Engineering, Michigan State University, 428 S. Shaw Lane, East Lansing, Michigan 48824-1322, United States

[ǂ]San Diego Supercomputer Center, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0505, United States

[ł]Department of Mechanical Engineering, Pennsylvania State University, University Park, Pennsylvania 16802, USA

KEYWORDS: AMBER, ReaxFF, PuReMD, Molecular Dynamics, Claisen Rearrangement, Umbrella Sampling

---

Abstract. Combined quantum mechanical/molecular mechanical (QM/MM) models using semiempirical and *ab initio* methods have been extensively reported on over the past few decades. These methods have been shown to be capable of providing unique insights into a range of problems, but they are still limited to relatively short time scales, especially QM/MM models using *ab initio* methods. An intermediate approach between a QM based model and classical mechanics could help fill this timescale gap and facilitate the study of a range of interesting problems. Reactive force fields represent the intermediate approach explored in this paper. A widely used reactive model is ReaxFF, which has largely been applied to materials science problems and is generally used stand-alone (*i.e.*, the full system is modeled using ReaxFF). We report a hybrid ReaxFF/AMBER molecular dynamics (MD) tool, which introduces ReaxFF capabilities to capture bond breaking and formation within the AMBER MD software package. This tool enables us to study local reactive events in large systems at a fraction of the computational costs of QM/MM models. We describe the implementation of ReaxFF/AMBER, validate this implementation using a benzene molecule solvated in water, and compare its performance against a range of similar approaches. To illustrate the predictive capabilities of ReaxFF/AMBER, we carried out a Claisen rearrangement study in aqueous solution. In a first for ReaxFF, we were able to use AMBER's potential of mean force (PMF) capabilities to perform a PMF study on this organic reaction. The ability to capture local reaction events in large systems using combined ReaxFF/AMBER opens up a range of problems that can be tackled using this model to address both chemical and biological processes.

---

1. Introduction

Modeling accurate chemistry of relatively large and complex systems for sufficiently long simulation timescales is essential for computational atomistic modeling tools. Classical molecular mechanics (MM) simulations have proven to be useful for studying complex systems in which no bond breaking or formation is involved. In such methods, empirical force fields (EFF) that can describe the relationship between the geometry and energy of a system with a set of relatively simple potential functions has been used for describing many systems around their equilibrium configuration. However, EFF based methods are unable to simulate chemical reactions due to the rigid connectivity requirement of such methods. Quantum mechanics (QM) methods can mitigate the limitations of EFF based methods such as formation or breaking of bonds and charge fluctuations due to geometry changes. However, QM models are usually applied to fragments of the regions involved in, say an enzymatic reaction, limiting the ability to explore the influence of environmental effects. Although QM based methods can be very accurate in predicting chemical reaction events, they remain limited to small systems simulated over short time scales.

Hybrid quantum mechanics/molecular mechanics (QM/MM) methods were developed to combine the best features of EFF and QM models to tackle a range of chemical problems[1, 2]. In

QM/MM methods, the total system is divided into two separate QM and MM zones. The QM zone is the chemically active region which is treated by a range of QM methods and the rest of the system is the MM zone which is treated using an EFF. Since the introduction of the QM/MM method, various approaches have been implemented and this method has found extensive application to deal with complex systems in realistic environments because of the significant reduction in the computational cost compared to pure QM methods[3-7]. Different QM/MM simulation tools have long been supported in the AMBER MD package[8, 9]. Some QM methods including semi-empirical neglect of diatomic overlap (NDDO)-type and density functional tight binding (DFTB) are built-in (*i.e.*, *internal*) and are supported natively within AMBER[10, 11]. More advanced QM methods are supported via a file based integration interface to external QM software packages[12, 13].

Despite highly innovative techniques, algorithmic improvements and fast implementations, the computational cost of the QM region still stands as the rate limiting factor in QM/MM simulations. The Reactive Force Field (ReaxFF) is a potential based on the bond length/bond order concept that bridges QM and MM methods in terms of functionality and computational costs. Importantly, ReaxFF provides a reasonable approximation of reactive phenomena at computational costs comparable to MM methods. In this paper, we report a hybrid ReaxFF/AMBER MD tool, which introduces ReaxFF capabilities to capture bond breaking and formation within the AMBER MD software package. This tool enables us to study local reactive events in large systems at a fraction of the computational costs of QM/MM models.

Another major challenge with atomistic simulations is that chemical reactions through transition states can take place on a time scale that cannot be reached by regular molecular dynamics simulations. Therefore, approaches based on enhanced sampling methods are used to locate the transition state for a chemical reaction using QM/MM methods. Umbrella sampling[14] is one of the most well-known enhanced sampling methods and can readily be used in the new ReaxFF/AMBER tool. We use it herein to map out the reaction profile of the Claisen rearrangement as a validation study.

2. Background on ReaxFF

ReaxFF is a classical MM model in spirit, which explicitly models chemical reactions based on the bond-length/bond-order concept and dynamic distribution of charges. Similar to non-reactive MM models, ReaxFF consists of two sets of terms: the bonded and nonbonded terms (van der Waals and electrostatic interactions). However, ReaxFF allows bond formation and dissociation, and, hence, has significantly different bonded terms than classical potentials.

To illustrate the philosophy of the method, we describe the determination of the bond energy using bond orders for carbon and hydrogen (other elements are similarly dependent based on their specific properties), while the structure and definition of the remaining terms can be found in the original description of ReaxFF[15]. A single atom type in ReaxFF defines each element, *e.g.*, there are no *sp*, *sp*^{2}, or *sp*^{3} hybridized carbon atoms, but only one carbon atom type. The bond energy (E_{bond}, Eq. 1) is described as a function of the sigma (BO_{ij}^{σ}), first π (BO_{ij}^{π}) and second π ($BO_{ij}^{\pi\pi}$) bond orders, as well as the corresponding D_e, $p_{be1}$and p_{bc2} parameters. The different bond orders themselves ($BO_{ij}^{\sigma}, BO_{ij}^{\pi}, BO_{ij}^{\pi\pi}$) are calculated using the pairwise distance (r_{ij}) between atoms *i-j*, the ideal bond distances ($r_o^{\sigma}, r_o^{\pi}, r_o^{\pi\pi}$) for atom types of *i* and *j*, and the force field specific parameters $p_{bo[1-6]}$ as shown in Eq. 2. All three terms in Eq. 2 are considered for a bond between two carbon atoms, while only the first term is used for the σ bond that forms between a carbon atom and a hydrogen atom. However, a pairwise distance-based representation will yield small bond orders between 1-3 atoms causing a bond order overestimation between the relevant atoms. A bond order correction (BO_{ij} in Eq. 3) is applied to minimize the long-range bond orders for such situations, where Δ_i' (Eq. 4) is the deviation of the uncorrected bond order summation from the valence state of an atom (*e.g.* carbon and hydrogen have valences of four and one, respectively).

$$E_{bond} = -D_e^{\sigma} \cdot BO_{ij}^{\sigma} \cdot exp\left[p_{be1}\left(1-\left(BO_{ij}^{\sigma}\right)^{p_{bc2}}\right)\right]$$

$$-D_e^{\pi} \cdot BO_{ij}^{\pi} - D_e^{\pi\pi} \cdot BO_{ij}^{\pi\pi} \quad (1)$$

$$BO_{ij}' = BO_{ij}^{\sigma} + BO_{ij}^{\pi} + BO_{ij}^{\pi\pi} = exp\left[p_{bo1} \cdot \left(\frac{r_{ij}}{r_o^{\sigma}}\right)^{p_{bo2}}\right] + exp\left[p_{bo3} \cdot \left(\frac{r_{ij}}{r_o^{\pi}}\right)^{p_{bo4}}\right] + exp\left[p_{bo5} \cdot \left(\frac{r_{ij}}{r_o^{\pi\pi}}\right)^{p_{bo6}}\right] \quad (2)$$

$$BO_{ij} = BO_{ij}' \cdot \mathrm{f1}\left(\Delta_i', \Delta_j'\right) * \mathrm{f4}\left(\Delta_i', BO_{ij}'\right) * \mathrm{f5}\left(\Delta_j', BO_{ij}'\right) \quad (3)$$

$$\Delta_i' = -Val_i + \sum_{j=1}^{neighbours(i)} BO_{ij}' \quad (4)$$

During an MD simulation, bond orders are evaluated at each time-step and are used to determine the atomic connectivity within a pre-defined distance cutoff (typically 5Å to Angstroms). A time step of 0.25 fs can be used for most simulations, while a smaller time-step is needed for higher temperature studies (>1500K). The energy curves are continuous throughout the simulation process, even at regions involving bond formation/breaking where favorable reactions can automatically occur without any restraints. This is ensured by the inclusion of other bond order related terms (see Eq. 5). Bond angles (E_{val}) and torsions (E_{tor}) are evaluated using similar bond order considerations. In a bond order potential, atoms often do not achieve their optimal coordination numbers. Therefore, ReaxFF requires additional abstractions such as lone pair (E_{lp}), over/under-coordination correction (E_{over} & E_{under}), 3-body penalty (E_{pen}) for systems with two double bonds sharing an atom, Three-body conjugation term (E_{coa}), Correction for C_2 (E_{C2}), Triple bond energy correction (E_{triple}) and 4-body conjugation (E_{conj}) terms. The potential is summarized below. The detailed expressions for each term can be found in the literature[15].

$$E_{system} = E_{bond} + E_{lp} + E_{over} + E_{under} + E_{val} + E_{pen} + E_{coa} + E_{C2} + E_{triple} + E_{tors} + E_{conj} + E_{H-bond} + E_{vdWaals} + E_{Coulomb} \quad (5)$$

To prevent energy jumps during bond formation/dissociation, there are nonbonded interactions between each atom pair (even for 1-2, 1-3 interactions) in ReaxFF. Electrostatic interactions are represented by a shielded Coulombic term and the van der Waals interaction uses a shielded Morse potential to prevent unrealistic values at very short distances. An important and computationally expensive pre-cursor to the electrostatic interactions is the need to dynamically determine partial charges at every MD step, which is accomplished through charge models (such as QEq[16] and EEM[17]) that require the solution of a linear system of equations.

The resulting formulation is complex, but highly flexible and transferable. This allowed ReaxFF to be broadly applicable to a wide range of challenging problems. Developed originally for hydrocarbons[15], the ReaxFF method has been extensively used to investigate complex systems in a wide range of applications including biological systems[18-22], materials[23-29], catalysts[30, 31], combustion and batteries[32].

In addition to the original Fortran based ReaxFF program[15], ReaxFF is also available in the open-source Purdue Reactive Molecular Dynamics (PuReMD) [33, 34] package and has been integrated into the open-source Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) [35]. The ReaxFF method is also available in other important simulation environments, including the Amsterdam Density Functional (ADF) [36] program and Materials Studio released under license by Biovia (https://www.3dsbiovia.com), as well as Nanohub (http://www.nanohub.org).

3. The ReaxFF/AMBER Integration

Similar to QM/MM methods, atoms are split into 3 categories in the ReaxFF/AMBER method: *i*) **ReaxFF atoms,** which include all atoms in the chemically reactive region and are handled by a ReaxFF implementation, *ii*) the **ReaxFF/MM transition atoms,** which include all atoms within a certain cutoff of the ReaxFF region and is handled by ReaxFF and AMBER collaboratively, and *iii*) **the MM atoms,** which include all remaining atoms and is handled exclusively by AMBER. These categories are illustrated in **Figure 1** with sphere-like shapes, but these regions can obviously be of any shape. In what follows, we describe the implementation of the ReaxFF/AMBER method.

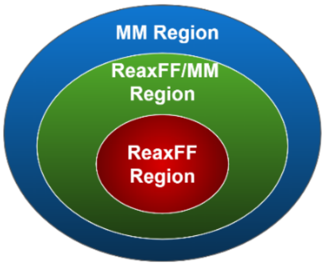

Figure 1. ReaxFF/AMBER regions in the integration implementation

3.1 Implementation

AMBER is the simulation driver in the ReaxFF/AMBER MD integration. After AMBER categorizes atoms into their respective groups, it sends all relevant information for ReaxFF and ReaxFF/MM atoms to the ReaxFF program. The ReaxFF/AMBER tool currently uses the external model interface, *i.e.*, the necessary data transfers between ReaxFF and AMBER are performed using file-based data exchange. Therefore, after AMBER completes writing the data exchange files, it launches the ReaxFF program as an external binary and waits for its output files. The ReaxFF program then runs a zero-step non-periodic simulation to calculate the dynamic charges, energies and forces on the ReaxFF and ReaxFF/MM atoms. Upon completion, it writes this information back into another file which is finally read by AMBER to complete the energy and force computations for the ReaxFF/MM and MM regions.

In implementing the interface between the ReaxFF and AMBER programs, we have adopted the following procedure for a successful hybrid model:

- Dynamic charges on ReaxFF atoms are calculated under the influence of ReaxFF/MM atoms with static charges,
- All ReaxFF interactions (as given in Eq. 5 above) between ReaxFF-ReaxFF pairs are calculated without any modifications,
- Electrostatic interactions between ReaxFF (w/dynamic charge)-ReaxFF/MM (w/static charge) atom pairs are calculated by ReaxFF,
- van der Waals interactions between ReaxFF-ReaxFF/MM atom pairs are handled by AMBER (*e.g.*, using a Lennard Jones potential),
- Interactions between MM-ReaxFF/MM and MM-MM pairs are handled by AMBER as usual.

We should note that there are some limitations of the current ReaxFF/AMBER tool. At the moment, only systems with non-covalent bonds between ReaxFF and ReaxFF/MM regions can be studied. Also, only shared memory parallelism can be leveraged for the time-being. Nevertheless, as we demonstrate below, the current implementation serves as a proof-of-concept on the feasibility and advantages of this approach.

3.2 Dynamic Charges with the Modified Electronegativity Equalization Method (mEEM)

Since we allow the statically charged MM and transition region atoms to polarize the ReaxFF atoms, the dynamic charge model used by ReaxFF needs to be modified. As mentioned previously, ReaxFF can use the charge equilibration (QEq) or electronegativity equalization method (EEM) to determine the charges. Our current implementation is based on EEM. Before describing the necessary modifications, we briefly discuss the EEM charge model that is currently used by ReaxFF/AMBER.

The EEM charge model[29] relies on the principle that charges should be distributed on atoms in order to satisfy constraints for both the net system charge and the equalized atom electronegativities. Let atomic charges be $\mathbf{q} = (q_1, q_2, \ldots\ldots q_n)$ and the positions be $\mathbf{R} = (r_1, r_2, \ldots\ldots r_n)$, where $q_i \in R$ and $r_i \in R^3$. Based on Sanderson's Electronegativity Equalization Principle[37]**,** the electronegativity of all atoms needs to be equalized

$$\epsilon_1 = \epsilon_2 = \cdots = \epsilon_i = \underline{\epsilon}$$

where ϵ_i is the electronegativity of atom *i* and $\underline{\epsilon}$ is the average molecular electronegativity. The other constraint forces the sum of the atomic charges to be equal to the given net system charge

$$\sum_{i=0}^{n} q_i = \ q_{net}.$$

The constraints and the parametrized interatomic interactions can be merged into the following linear equation where the charges **q** are the solution of:

$$\begin{bmatrix} \boldsymbol{H} & 1_n \\ 1_n^T & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{q} \\ \underline{\epsilon} \end{bmatrix} = \begin{bmatrix} -\boldsymbol{\chi} \\ q_{net} \end{bmatrix}.$$

Here, $\boldsymbol{\chi}$ is an *n* by *1* vector of atomic electronegativities and the H_{ij} values, i.e., individual elements of the *n* by *n* **H** matrix, are defined as $\delta_{ij}\eta_i + (1 - \delta_{ij}) \cdot F_{ij}$, where δ_{ij} is the Kronecker delta operator, η_i is the idempotential, and F_{ij} is defined as

$$F_{i,j} = \begin{cases} \dfrac{1}{\sqrt[3]{r_{ij}^3 + \gamma_{ij}^{-3}}}, & r_{ij} \leq r_{nonb} \\ 0, & otherwise. \end{cases}$$

In the equations above, $r_{ij} = \|r_i - r_j\|$ is the distance between atoms *i* and *j*, $\gamma_{ij} = \sqrt{\gamma_i \cdot \gamma_j}$ is a pairwise *shielding term* tuned for element types of atoms *i* and *j* to avoid unbounded electrostatic energy at short distances, and $\underline{\epsilon}$ is the dielectric constant of the medium. It has been demonstrated that the EEM model reproduces QM calculated Mulliken charges[38].

To account for the polarization effect of the transition region atoms on the core ReaxFF region, we introduce the ReaxFF/AMBER atoms as particles with fixed charges to the EEM solver. This is done by modifying the system of equations as follows. Assuming that there are *c* ReaxFF atoms and *t* transition region atoms, then $H_{core} \in R^{c \times c}$ captures the interactions within the core ReaxFF region and $H_{core-trans} \in R^{c \times t}$ captures the impact of the transition region atoms on the ReaxFF atoms.

$$\begin{bmatrix} H_{core} & H_{core-trans} & 1_c \\ H_{trans-core} & H_{trans} & 1_t \\ 1_c^T & 1_t^T & 0 \end{bmatrix} \begin{bmatrix} q_{core} \\ q_{trans} \\ \underline{\epsilon} \end{bmatrix} = \begin{bmatrix} -\chi_{core} \\ -\chi_{trans} \\ q_{net} \end{bmatrix}$$

Since a direct solution of the above linear system scales cubically with the number of total atoms, typically iterative solvers are used to obtain approximations to the optimized dynamic charges. In such an iterative scheme, we only evolve the charges on the core atoms as q_{trans} are fixed values given in the MM force field. As such, the rows corresponding to the transition region atoms can be ignored in the above linear system and we obtain:

$$\begin{bmatrix} H_{core} & H_{core-trans} & 1_c \\ 1_c^T & 1_t^T & 0 \end{bmatrix} \begin{bmatrix} q_{core} \\ \underline{\epsilon} \end{bmatrix} = \begin{bmatrix} -\chi_{core} \\ q_{net} \end{bmatrix}.$$

Since q_{trans} is fixed, it can be rearranged as

$$\begin{bmatrix} H_{core} & 1_c \\ 1_c^T & 0 \end{bmatrix} \begin{bmatrix} q_{core} \\ \underline{\epsilon} \end{bmatrix} = \begin{bmatrix} -\chi_{core} \\ q_{net} \end{bmatrix} - \begin{bmatrix} H_{core-trans} \\ 0_t^T \end{bmatrix} \begin{bmatrix} q_{trans} \end{bmatrix}$$

These modifications ensure that the existence of the fixed charges does not slow down the charge equilibration step since the fixed charges and dynamic charges are separated. See Supplementary Information (SI) for an example.

The modifications above are sufficient to run non-periodic ReaxFF/MM simulations, as well as periodic boundary simulations without long range electrostatic interactions. When running periodic boundary simulations with long-range electrostatic interactions AMBER uses the particle mesh Ewald (PME) approach which calculates interactions within a certain cutoff distance directly and approximates the rest over a mesh[39]. To be able to account for the long range interactions in EEM, we have incorporated the impact of the mesh points on ReaxFF atoms by further extending the EEM matrix, i.e., by adding a column and row containing the effects of the mesh points. This scheme does not conserve the energy in NVE simulations with periodic boundaries because ReaxFF uses tapered electrostatic interactions[40] that forces the Coulomb effects to slowly decay to 0 at the interaction cutoff radius which is typically set to 10-12Å (see SI, **Figure 1**). Tapering of the Coulomb interactions is a built in feature of the ReaxFF formulation and existing ReaxFF parameter sets have been trained with this design principle in mind. We believe that by re-training ReaxFF parameters without tapered Coulomb kernels, periodic boundary simulations with PME can successfully be enabled, but doing so goes beyond the scope of this paper, so we leave it as a topic further exploration.

3.3 Validation

For validation of the resulting ReaxFF/AMBER combination, we performed experiments using benzene-in-water systems, one with periodic boundary conditions, and one without (**Figure 2**). Benzene is not reactive in water, but the goal of these computational experiments was to validate that the ReaxFF/MM method can achieve energy conservation and produce reasonable dynamic charges.

For both simulations (non-periodic and periodic), the systems are first energy-minimized and then heated to 300K using the Berendsen thermostat in AMBER. Finally, ReaxFF/MM NVE simulations are run using a time step of 0.25 fs to check energy conservation and charges. A relatively short time step was chosen as this is the recommended setting for ReaxFF simulations, especially in the presence of H atoms. For all simulations, the SHAKE algorithm was turned off.

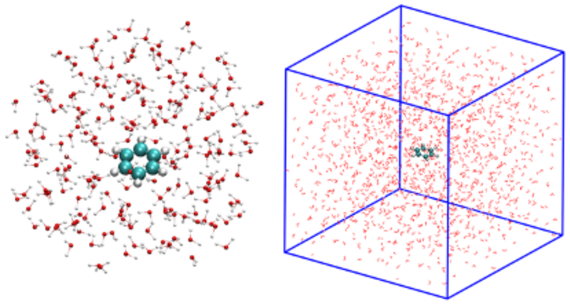

Figure 2. (Left) ReaxFF Benzene in a TIP3P water droplet with a total of 627 atoms, (Right) ReaxFF Benzene in a TIP3P water box with a total of 4398 atoms.

As shown in **Figure 3**, the energy is conserved for both simulations. **Figure 4** shows the Carbon and Hydrogen charges for Reax atoms (denoted as *Reax-C* and *Reax-H*, respectively), as well as the average across all C and H atoms in a timestep (denoted as *Reax-C-Avg* and *Reax-H-Avg*, respectively). As can be seen, dynamic charges produced by ReaxFF/MM under the influence of statically charged transition region atoms are in line with the Carbon and Hydrogen charges given in the AMBER force field for benzene (denoted as *AMBER-C* and *AMBER-H*, respectively).

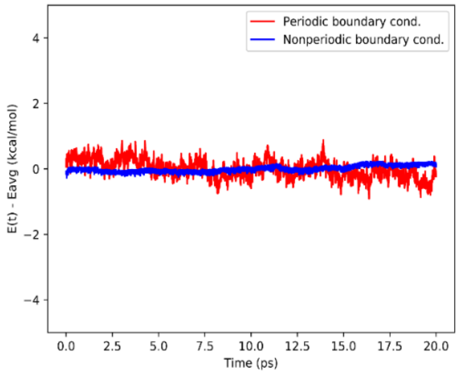

Figure 3. Simulation of a benzene molecule in water using the TIP3P water model and NVE ensemble. A time step of 0.25 fs was used. (Blue) Non-periodic boundary condition with an infinite cut-off. (Red) Periodic boundary condition with a QM cutoff of 10Å (qmcut=10Å) and ReaxFF/PME interaction turned off (qm_pme=0).

3.4 Performance Analysis

As mentioned before, the motivation for the ReaxFF/MM method is that it can significantly reduce the computational time that would be needed by a comparable QM/MM simulation. To illustrate this, we benchmarked the computational cost of separate equilibration simulations of allyl vinyl ether (AVE) which consists of 14 atoms using ReaxFF, PuReMD, SCC-DTFB, PM3, MM, the Hartree-Fock method (HF) and density functional theory (DFT) using the B3LYP functional. The HF and DFT calculations used the 6-31G* basis set. The benchmark simulations were performed using standalone ReaxFF, PuReMD, AMBER and the Gaussian software package in sequential execution mode on an Intel Xeon E5-2670 v2 CPU which runs at 2.50 GHz. The benchmarked time per simulation step are shown in Table 1.

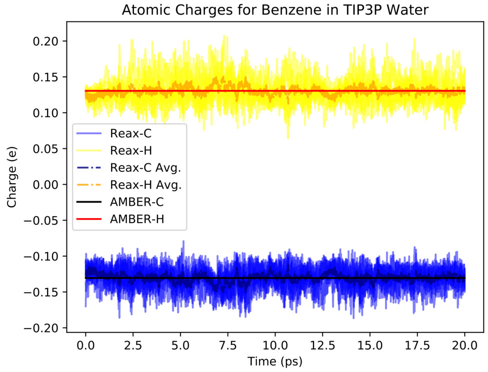

Figure 4. Dynamic ReaxFF charges and fixed AMBER charges with a ReaxFF/MM simulation of a benzene molecule in water with periodic boundary conditions using parameters qmcut=10 A and qm_pme=0.

Table 1. Sequential execution time in seconds per time step and nanoseconds per day for SCC-DFTB, PM3, ReaxFF, PuReMD, MM, HF and DFT calculations. B3LYP functional and the 6-31G* basis set were used for the HF and DFT geometry optimization calculations.

Method	**seconds/iteration (relative wrt. MM)**	**timestep(fs)**	**ns/day (relative wrt. PuReMD)**
SCC-DFTB	3.088×10^{-3} (39.8)	0.5	13.99 (0.2)
PM3	1.83×10^{-3} (23.6)	0.5	23.61 (0.4)
ReaxFF	6.25×10^{-4} (8.1)	0.25	34.56 (0.6)
PuReMD	3.63×10^{-4} (4.7)	0.25	59.50 (1.0)
MM	7.75×10^{-5} (1.0)	1	1114.84 (18.7)
HF	1,833 (23,651.6)	------	------
DFT	2.762(35,638.7)	------	------

In Table 1, PuReMD refers to a C language based efficient parallel implementation of the original ReaxFF implementation in Fortran. Time steps of 1fs for MM, 0.5fs for SCC-DFTB and PM3 and 0.25fs for ReaxFF calculations are assumed for these benchmark calculations. These benchmark data show that ReaxFF method is ~2-5 times faster than PM3 or SCC-DFTB methods, three orders of magnitude faster than HF or DFT methods and ~18 times slower than conventional molecular mechanics methods. Therefore, ReaxFF can offer an alternative for QM/MM calculations to handle larger complex systems at longer time scales.

Currently, ReaxFF/AMBER software uses the original Fortran ReaxFF code through a file-based data exchange interface. We analyzed the computational performance of the ReaxFF/AMBER software with the nonperiodic benzene-in-water simulation for 1000 steps using an infinite electrostatic interaction cut-off. The timing breakdown shown in **Figure 5** is averaged over the course of this simulation. Please see SI/Section 3 for specifications of the hardware and software used in these computational experiments.

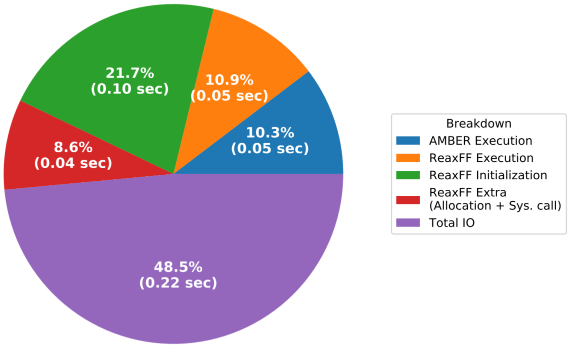

Figure 5. Breakdown of the total execution time of ReaxFF/AMBER software for the benzene-in-water simulation into its main components. The total time spent per step is 0.46 s.

Since the data exchange between ReaxFF and AMBER is file based, it requires the writing and reading of files at every simulation step. There are also the additional costs associated with launching the ReaxFF program as an external binary. Finally, the modifications to the standalone ReaxFF implementation is minimal and for that reason, ReaxFF allocates and initializes various data structures at every time step. Time spent on the IO and system calls to the external program will be removed by introducing array-based data exchanges between the relevant ReaxFF and AMBER subroutines. Additionally, by keeping the data structures in memory persistently and reusing them at each time step, we could eliminate most of the time spent on the allocation and initialization parts. When all these potential updates are considered, the ReaxFF/AMBER software would be accelerated significantly beyond what is shown here because only 10.9% of the overall time is spent on the necessary ReaxFF calculations (see **Figure 6**).

4. Claisen Rearrangement Simulations with ReaxFF/AMBER

To demonstrate the capabilities of the new ReaxFF/AMBER integration, we carried out ReaxFF/MM modeling of the classic Claisen rearrangement of allyl vinyl ether (AVE) solvated in an explicit TIP3P water model[41] as shown in **Figure 6**. Since the Claisen rearrangement is a well-studied reaction[42, 43], it was chosen to evaluate the chemical accuracy of our new method. In these simulations, the solvent is treated by AMBER using the TIP3P water model and the reactant AVE is treated with ReaxFF. No covalent bonds cross the ReaxFF/AMBER boundary.

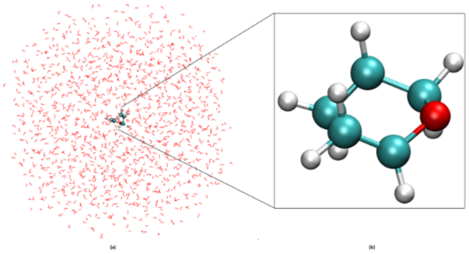

Figure 6. (a) Solvation of allyl vinyl ether (AVE) in octahedral TIP3P water with total number of 4319 atoms. (b) The molecular structure of AVE.

4.1 ReaxFF Parameter Optimization

The existing ReaxFF force field[44] was tested for simulating the Claisen rearrangement. The initial tests showed that the system got trapped in a local minimum after reaching the transition state (TS), and therefore a proper Claisen rearrangement could not be observed (**Figure 7**).

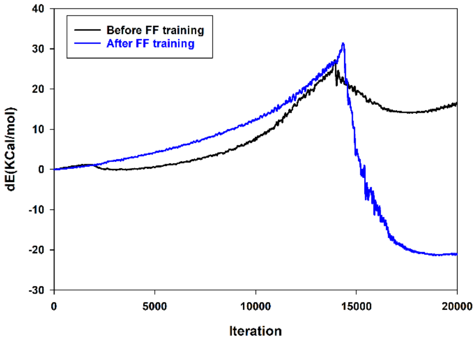

Figure 7. Claisen rearrangement of AVE in vacuum using original and trained ReaxFF force fields. The original force field leads to getting trapped in local minimum.

To resolve this, *ab initio* QM data for the Claisen rearrangement of AVE in vacuum were generated. The intrinsic reaction coordinate (IRC) method was used to obtain QM data from two different chair-like and boat-like transition states. All the optimization and IRC calculations were done with the Gaussian 16 package[45]. B3LYP functional and the 6-31G* basis set were used for the calculations. Through IRC calculations, a Claisen rearrangement transition of all chair-like and boat-like configurations and the energy changes were recorded. Comparison of these QM data against the initial ReaxFF force field results are shown in **Figure 8.a**. These initial energy differences and the failure of the initial force field in capturing the Claisen rearrangement indicated the requirement of force field training against the generated QM data. The training data used for optimizing a ReaxFF parameter set can include QM data on charges, heat of formations (kcal/mol), energy minimized geometries (in Å or °), lattice cell parameters (in Å or °) and relative energies (kcal/mol). All the training data are added to the training set and the force field is reparametrized to minimize the error function:

$$e_i = (\frac{x_{i,QM} - x_{i,ReaxFF}}{\sigma_i})^2$$

where $x_{i,QM}$ and $x_{i,ReaxFF}$ are the QM and ReaxFF values of the ith entry of the training set, respectively, and σ_i are weight parameters that determine the desired accuracy for individual training data items. Force field parameters which are designated to be tuned are defined in a separate input file. The ReaxFF training feature in the standalone Fortran code then uses a line search scheme to optimize each parameter to be tuned one at a time. The force field training input files and force field parameters resulting from this optimization procedure are supplied in the SI.

Figure 8.b shows the results of our fitting against QM data for chair-like and boat like transition states. Evaluation of the new force field after training showed satisfactory behavior in capturing the Claisen rearrangement in vacuum (**Figure 7**). Therefore, this updated ReaxFF force field was used to perform free energy calculations of the Claisen rearrangement of AVE in the presence of explicit TIP3P water using ReaxFF/AMBER.

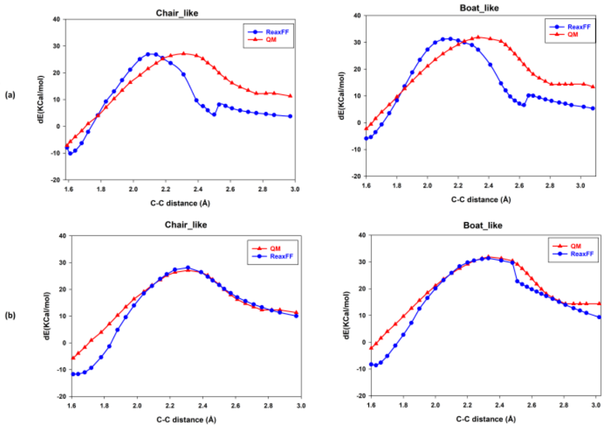

Figure 8. (a) QM (B3LYP /6-31G*) vs original ReaxFF force field results for the Claisen rearrangement of chair-like and boat-like structures. **(b)** QM (B3LYP/6-31G*) vs trained ReaxFF force field results for the Claisen rearrangement of chair-like and boat-like structures. The energies are with respect to the completely open AVE geometry optimized energy as the reference.

4.2 Free Energy Calculation with Umbrella Sampling

The MD driver in the ReaxFF/AMBER integration scheme is AMBER. Hence, we can use all of AMBER's advanced sampling techniques. One such feature is the umbrella sampling free energy calculation technique. Using the newly optimized parameter set, AMBER's umbrella sampling feature coupled with Weighted Histogram Analysis Method (WHAM) were used to generate the potential of mean force (PMF) of this reaction with ReaxFF/AMBER. The reaction coordinate chosen for the umbrella sampling simulations was the distance between two terminal AVE carbon atoms referred to as the C_1-C_5 distance (**Figure 9**).

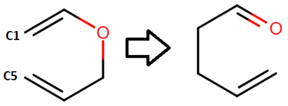

Figure 9. Claisen rearrangement of AVE while C_1-C_5 distance is changed from 3.5Å to 1.5Å.

A series of harmonic potentials were used to constrain the reaction coordinates to the defined windows. The C_1-C_5 distance was varied from 3.5Å to 1.6Å. For the regions near the transition state (1.7-2.4 Å), 1000 kcal/mol.$Å^2$ was used to obtain enough sampling for each window and for the regions far from the transition state (1.6-1.7 Å and 2.4-3.5 Å) a weaker restraint (200-700kcal/mol.$Å^2$) was used. For the 1.6-3.0 Å range, 0.02 Å window intervals were defined and for the 3.0-3.5 Å range, 0.1 Å window intervals were defined. In total, 81 windows were created for sampling, while varying the C_1-C_5 distance from 3.5Å to 1.6Å. All simulations were performed using the TIP3P rigid three site point charge water model. AVE was solvated in an octahedral box of TIP3P water molecules of 25 Å radius. All simulations were performed with periodic boundary conditions. The topology files for AVE were created using the general AMBER force field (GAFF) [46]. AMBER MM simulations using periodic boundary conditions were performed with a 10Å cutoff for the real-space nonbonded interactions and the particle mesh Ewald (PME) algorithm was incorporated to account for long-range electrostatics beyond the cutoff.

After 2000 steps minimization in each window, we ran 25 ps NPT (constant number of atoms, constant pressure and constant temperature) equilibration using constant pressure Langevin dynamics with the Berendsen barostat at 300K and 1 atm. The shake algorithm was enabled for these simulations in the MM region and the time step of 0.25 fs was chosen for both ReaxFF (no shake) and AMBER regions. Data collection for umbrella sampling was started after of the equilibration phase for 5 ps in each window. Umbrella sampling consisted of data collection from separate windows of the reaction coordinate simultaneously. By defining proper harmonic restraint constants in each window, we allowed neighboring windows to overlap and ensured there were enough windows to cover the entire reaction coordinate space. Data sampling were performed every 0.0125 ps of the production stage of umbrella sampling simulations. Finally, the PMF was calculated by combining the data from each window using WHAM.

4.3. Results and Discussion

To evaluate the potential of mean force (PMF) variations as a function of the reaction coordinate defined for the Claisen rearrangement, umbrella sampling calculations were employed. As mentioned, the simulations in different windows needed to be such that we could observe convergence on sampling to complete the umbrella sampling calculations. In order to make sure that all the windows overlap properly before performing the WHAM calculations, the histograms of all windows were generated (**Figure 10**). This plot shows that there was proper overlap and no obvious gaps were observed. Using this dataset, we used WHAM to construct the PMF of the Claisen rearrangement.

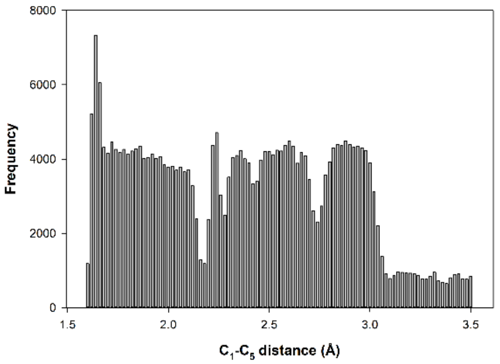

Figure 10. Histogram of C1-C5 distance samplings from the umbrella sampling windows. The C_1-C_5 distance was varied from 3.5Å to 1.6Å. For the regions far from the transition state (1.5-1.7 Å and 2.4-3.5 Å) a weaker restraint (200-700 kcal/mol.Å2) was used and for the regions near the transition state (1.7-2.4 Å) a stronger restraint (1000 kcal/mol.Å2) was used. For the 1.6-3.0 Å range, 0.02 Å interval windows were defined and for 3.0-3.5 Å windows interval of 0.1Å were defined.

The PMF obtained from umbrella sampling calculations is shown in **Figure 11**. The calculated PMF can be used to evaluate the Claisen rearrangement barrier height and the transition state configuration. A barrier height of ~32 kcal/mol was obtained from these calculations. We also utilized an implementation of the self-consistent charge density functional tight-binding (SCC-DFTB) method, which is a semiempirical method based on density functional theory (DFT) [47], and also Semi-empirical neglect of diatomic overlap (NDDO) PM3 method[48] as part of the QM/MM support in the AMBER 18 MD program to perform calculations on the Claisen rearrangement. Also, another SCC-DFTB QM/MM umbrella sampling PMF calculation using the SPC/E water model was performed to evaluate the impact of the water model on this simulation. The results of all these PMF calculations are shown in **Figure 11**. The transition configuration of these different methods are shown in supporting materials (see **Figure S3**). ReaxFF/AMBER integration successfully captured the Claisen rearrangement reaction.

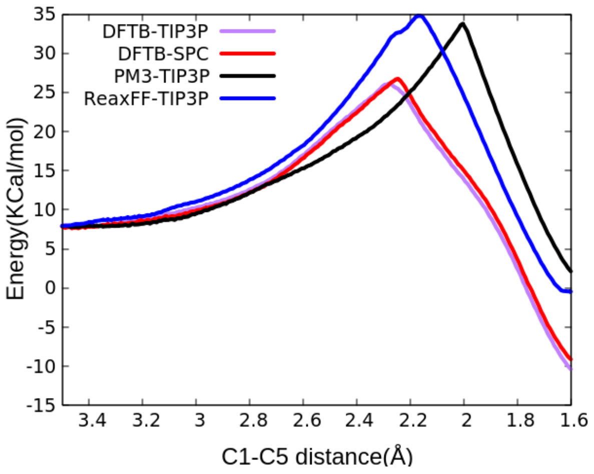

Figure 11. PMF of AVE Claisen rearrangement in TIP3P periodic water using combined ReaxFF/AMBER. The energies are with respect to the completely open AVE geometry optimized energy as the reference.

The first kinetic study of thermal rearrangement of AVE in the gas-phase by Schuler and Murphy reported an activation energy of 30.6 kcal/mol[49]. This is consistent with other experimental reports of experimental activation energies for chair transition[50-52]. This barrier height has been reported to be lower in water solvent than the value in the gas phase and nonpolar solvents[53]. The transition state bond lengths of 2.2-2.3 Å were reported using different simulation techniques[54].

5. Conclusions

We conducted umbrella sampling simulations by generating a series of configurations along the C_1-C_5 distance in AVE as the reaction coordinate, ran biasing simulations, and extracted the PMF. These results showed the capability of the ReaxFF/AMBER integration in capturing claisen rearrangement reaction despite the small inaccuracies in the reaction barrier height which can be enhanced by more thorough training of the ReaxFF force field.

This new interface will be a useful tool for modeling big biomolecular systems with local reactive regions. This combination can introduce reduced computational costs compared to available QM/MM methods. Rather than integrating ReaxFF directly into AMBER we will use the PuReMD code as the ReaxFF implementation in the next version of the ReaxFF/AMBER software suite. PuReMD contains significant performance improvements over the currently used Fortran based reference ReaxFF implementations and it has shared memory, distributed memory and GPU implementations[25-26], which will allow us to take advantage of modern massively parallel architectures.

ASSOCIATED CONTENT

Supporting text and Cartesian coordinates of the test molecules. This material is available free of charge via the Internet at http://pubs.acs.org.

ACKNOWLEDGMENT

This work was supported in part by a NIH grant (award number GM130641), a NSF CDS&E grant (award number 1807622) and a NSF OAC grant (1835144). This work used computational resources provided by the Institute for Cyber-Enabled Research at Michigan State University.

REFERENCES

1. Warshel, A.; Levitt, M., Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *Journal of Molecular Biology* **1976,** *103*, 227-249.
2. Lin, H.; Truhlar, D. G., QM/MM: what have we learned, where are we, and where do we go from here? *Theoretical Chemistry Accounts* **2006,** *117*, 185.
3. Senn, H. M.; Thiel, W., QM/MM Methods for Biomolecular Systems. *Angewandte Chemie International Edition* **2009,** *48*, 1198-1229.
4. Friesner, R. A.; Guallar, V., Ab initio quantum chemical and mixed quantum mechanics/molecular mechanics (QM/MM) methods for studying enzymatic catalysis. *Annu. Rev. Phys. Chem.* **2005,** *56*, 389-427.
5. Guallar, V.; Wallrapp, F. H., QM/MM methods: Looking inside heme proteins biochemisty. *Biophysical Chemistry* **2010,** *149*, 1-11.
6. Magalhães, R. P.; Fernandes, H. S.; Sousa, S. F., Modelling Enzymatic Mechanisms with QM/MM Approaches: Current Status and Future Challenges. *Israel Journal of Chemistry* **2020,** *n/a*.
7. Lu, X.; Fang, D.; Ito, S.; Okamoto, Y.; Ovchinnikov, V.; Cui, Q., QM/MM free energy simulations: recent progress and challenges. *Molecular Simulation* **2016,** *42*, 1056-1078.
8. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P., AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications* **1995,** *91*, 1-41.
9. Case, D. A.; Cheatham, T. E., 3rd; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J., The Amber biomolecular simulation programs. *Journal of computational chemistry* **2005,** *26*, 1668-1688.
10. Walker, R.; Crowley, M.; Case, D., The implementation of a fast and accurate QM/MM potential method in AMBER. *Journal of computational chemistry* **2008,** *29*, 1019-31.
11. Seabra, G. d. M.; Walker, R. C.; Elstner, M.; Case, D. A.; Roitberg, A. E., Implementation of the SCC-DFTB Method for Hybrid QM/MM Simulations within the Amber Molecular Dynamics Package. *The Journal of Physical Chemistry A* **2007,** *111*, 5655-5664.
12. Götz, A. W.; Clark, M. A.; Walker, R. C., An extensible interface for QM/MM molecular dynamics simulations with AMBER. *Journal of Computational Chemistry* **2014,** *35*, 95-108.
13. Isborn, C. M.; Götz, A. W.; Clark, M. A.; Walker, R. C.; Martínez, T. J., Electronic Absorption Spectra from MM and ab Initio QM/MM Molecular Dynamics: Environmental Effects on the Absorption Spectrum of Photoactive Yellow Protein. *Journal of Chemical Theory and Computation* **2012,** *8*, 5092-5106.
14. Kästner, J., Umbrella sampling. *WIREs Computational Molecular Science* **2011,** *1*, 932-942.
15. van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A., ReaxFF: A Reactive Force Field for Hydrocarbons. *The Journal of Physical Chemistry A* **2001,** *105*, 9396-9409.
16. Rappe, A. K.; Goddard, W. A., Charge equilibration for molecular dynamics simulations. *The Journal of Physical Chemistry* **1991,** *95*, 3358-3363.
17. Mortier, W. J.; Ghosh, S. K.; Shankar, S., Electronegativity-equalization method for the calculation of atomic charges in molecules. *Journal of the American Chemical Society* **1986,** *108*, 4315-4320.
18. Samieegohar, M.; Sha, F.; Clayborne, A. Z.; Wei, T., ReaxFF MD Simulations of Peptide-Grafted Gold Nanoparticles. *Langmuir* **2019,** *35*, 5029-5036.
19. Yang, S.; Zhao, T.; Zou, L.; Wang, X.; Zhang, Y., ReaxFF-based molecular dynamics simulation of DNA molecules destruction in cancer cells by plasma ROS. *Physics of Plasmas* **2019,** *26*, 083504.
20. Hubin, P. O.; Jacquemin, D.; Leherte, L.; Vercauteren, D. P., Parameterization of the ReaxFF reactive force field for a proline-catalyzed aldol reaction. *Journal of Computational Chemistry* **2016,** *37*, 2564-2572.
21. Monti, S.; Jose, J.; Sahajan, A.; Kalarikkal, N.; Thomas, S., Structure and dynamics of gold nanoparticles decorated with chitosan–gentamicin conjugates: ReaxFF molecular dynamics simulations to disclose drug delivery. *Physical Chemistry Chemical Physics* **2019,** *21*, 13099-13108.
22. Trnka, T.; Tvaroška, I.; Koča, J., Automated Training of ReaxFF Reactive Force Fields for Energetics of Enzymatic Reactions. *Journal of Chemical Theory and Computation* **2018,** *14*, 291-302.
23. Rahnamoun, A.; van Duin, A. C. T., Reactive Molecular Dynamics Simulation on the Disintegration of Kapton, POSS Polyimide, Amorphous Silica, and Teflon during Atomic Oxygen Impact Using the Reaxff Reactive Force-Field Method. *The Journal of Physical Chemistry A* **2014,** *118*, 2780-2787.
24. Raymand, D.; van Duin, A. C. T.; Baudin, M.; Hermansson, K., A reactive force field (ReaxFF) for zinc oxide. *Surface Science* **2008,** *602*, 1020-1031.
25. Chenoweth, K.; Cheung, S.; van Duin, A. C. T.; Goddard, W. A.; Kober, E. M., Simulations on the Thermal Decomposition of a Poly(dimethylsiloxane) Polymer Using the ReaxFF Reactive Force Field. *Journal of the American Chemical Society* **2005,** *127*, 7192-7202.
26. Russo, M. F.; Li, R.; Mench, M.; van Duin, A. C. T., Molecular dynamic simulation of aluminum–water reactions using the ReaxFF reactive force field. *International Journal of Hydrogen Energy* **2011,** *36*, 5828-5835.
27. Ludwig, J.; Vlachos, D. G.; van Duin, A. C. T.; Goddard, W. A., Dynamics of the Dissociation of Hydrogen on Stepped Platinum Surfaces Using the ReaxFF Reactive Force Field. *The Journal of Physical Chemistry B* **2006,** *110*, 4274-4282.
28. Yoon, K.; Rahnamoun, A.; Swett, J. L.; Iberi, V.; Cullen, D. A.; Vlassiouk, I. V.; Belianinov, A.; Jesse, S.; Sang, X.; Ovchinnikova, O. S.; Rondinone, A. J.; Unocic, R. R.; van Duin, A. C. T., Atomistic-Scale Simulations of Defect Formation in Graphene under Noble Gas Ion Irradiation. *ACS Nano* **2016,** *10*, 8376-8384.
29. Rahnamoun, A.; van Duin, A. C. T., Study of thermal conductivity of ice clusters after impact deposition on the silica surfaces using the ReaxFF reactive force field. *Physical Chemistry Chemical Physics* **2016,** *18*, 1587-1594.

30. Zou, C.; Van Duin, A., Investigation of Complex Iron Surface Catalytic Chemistry Using the ReaxFF Reactive Force Field Method. *JOM* **2012,** *64*, 1426-1437.

31. Shin, Y. K.; Kwak, H.; Vasenkov, A. V.; Sengupta, D.; van Duin, A. C. T., Development of a ReaxFF Reactive Force Field for Fe/Cr/O/S and Application to Oxidation of Butane over a Pyrite-Covered Cr2O3 Catalyst. *ACS Catalysis* **2015,** *5*, 7226-7236.

32. Senftle, T. P.; Hong, S.; Islam, M. M.; Kylasa, S. B.; Zheng, Y.; Shin, Y. K.; Junkermeier, C.; Engel-Herbert, R.; Janik, M. J.; Aktulga, H. M.; Verstraelen, T.; Grama, A.; van Duin, A. C. T., The ReaxFF reactive force-field: development, applications and future directions. *npj Computational Materials* **2016,** *2*, 15011.

33. Aktulga, H. M.; Fogarty, J. C.; Pandit, S. A.; Grama, A. Y., Parallel reactive molecular dynamics: Numerical methods and algorithmic techniques. *Parallel Computing* **2012,** *38*, 245-259.

34. Kylasa, S. B.; Aktulga, H. M.; Grama, A. Y., PuReMD-GPU: A reactive molecular dynamics simulation package for GPUs. *Journal of Computational Physics* **2014,** *272*, 343-359.

35. Plimpton, S. J.; Thompson, A. P., Computational aspects of many-body potentials. *MRS Bulletin* **2012,** *37*, 513-521.

36. Kamat, A. M.; van Duin, A. C. T.; Yakovlev, A., Molecular Dynamics Simulations of Laser-Induced Incandescence of Soot Using an Extended ReaxFF Reactive Force Field. *The Journal of Physical Chemistry A* **2010,** *114*, 12561-12572.

37. Sanderson, R., An interpretation of bond lengths and a classification of bonds. *Science* **1951,** *114*, 670-672.

38. Bultinck, P.; Langenaeker, W.; Lahorte, P.; De Proft, F.; Geerlings, P.; Waroquier, M.; Tollenaere, J., The electronegativity equalization method I: Parametrization and validation for atomic charge calculations. *The Journal of Physical Chemistry A* **2002,** *106*, 7887-7894.

39. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993,** *98*, 10089-10092.

40. Chenoweth, K.; van Duin, A. C. T.; Goddard, W. A., ReaxFF Reactive Force Field for Molecular Dynamics Simulations of Hydrocarbon Oxidation. *The Journal of Physical Chemistry A* **2008,** *112*, 1040-1053.

41. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983,** *79*, 926-935.

42. Ganem, B., The Mechanism of the Claisen Rearrangement: Déjà Vu All Over Again. *Angewandte Chemie International Edition in English* **1996,** *35*, 936-945.

43. Gao, J., Combined QM/MM Simulation Study of the Claisen Rearrangement of Allyl Vinyl Ether in Aqueous Solution. *Journal of the American Chemical Society* **1994,** *116*, 1563-1564.

44. Zhang, W.; van Duin, A., Improvement of the ReaxFF Description for Functionalized Hydrocarbon/Water Weak Interactions in the Condensed Phase. *The Journal of Physical Chemistry B* **2018,** *122*.

45. Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Petersson, G.; Nakatsuji, H., Gaussian 16. Gaussian, Inc. Wallingford, CT: 2016.

46. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and testing of a general amber force field. *Journal of Computational Chemistry* **2004,** *25*, 1157-1174.

47. Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G., Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties. *Physical Review B* **1998,** *58*, 7260-7268.

48. Stewart, J. J. P., Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *Journal of Molecular Modeling* **2007,** *13*, 1173-1213.

49. Schuler, F. W.; Murphy, G. W., The Kinetics of the Rearrangement of Vinyl Allyl Ether1. *Journal of the American Chemical Society* **1950,** *72*, 3155-3159.

50. Gajewski, J. J.; Conrad, N. D., Variable transition state structure in 3,3-sigmatropic shifts from .alpha.-secondary deuterium isotope effects. *Journal of the American Chemical Society* **1979,** *101*, 6693-6704.

51. Wiest, O.; Black, K. A.; Houk, K., Density functional theory isotope effects and activation energies for the Cope and Claisen rearrangements. *Journal of the American Chemical Society* **1994,** *116*, 10336-10337.

52. Hiersemann, M.; Nubbemeyer, U., *The Claisen rearrangement: methods and applications*. John Wiley & Sons: 2007.

53. Zhang, J.; Yang, Y. I.; Yang, L.; Gao, Y. Q., Dynamics and Kinetics Study of "In-Water" Chemical Reactions by Enhanced Sampling of Reactive Trajectories. *The Journal of Physical Chemistry B* **2015,** *119*, 14505-14514.

54. M. Guest, J.; Simon Craw, J.; A. Vincent, M.; H. Hillier, I., The effect of water on the Claisen rearrangement of allyl vinyl ether: theoretical methods including explicit solvent and electron correlation. *Journal of the Chemical Society, Perkin Transactions 2* **1997**, 71-74.

Insert Table of Contents artwork here