

# MoD-QM/MM Structural Refinement Method: Characterization of Hydrogen Bonding in the *Oxytricha nova* G-Quadruplex

Junming Ho,<sup>†</sup> Michael B. Newcomer,<sup>†</sup> Christina M. Ragain,<sup>†</sup> Jose A. Gascon,<sup>†,§</sup> Enrique R. Batista,<sup>‡</sup> J. Patrick Loria,<sup>†,#</sup> and Victor S. Batista\*,<sup>†</sup>

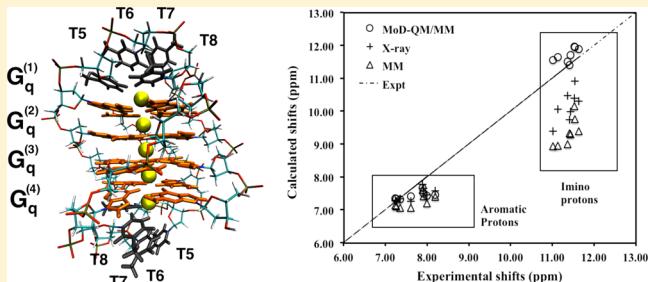
<sup>†</sup>Department of Chemistry, Yale University, P.O. Box 208107, New Haven, Connecticut 06520-8107, United States

<sup>‡</sup>Theoretical Division, Los Alamos National Laboratory, MS-B214, Los Alamos, New Mexico 87545, United States

<sup>#</sup>Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, United States

## S Supporting Information

**ABSTRACT:** A generalization of the Moving-Domain Quantum Mechanics/Molecular Mechanics (MoD-QM/MM) hybrid method [Gascon, J. A.; Leung, S. S. F.; Batista, E. R.; Batista, V. S. *J. Chem. Theory Comput.* **2006**, *2*, 175–186] is introduced to provide a self-consistent computational protocol for structural refinement of extended systems. The method partitions the system into molecular domains that are iteratively optimized as quantum mechanical (QM) layers embedded in their surrounding molecular environment to obtain an ab initio quality description of the geometry and the molecular electrostatic potential of the extended system composed of those constituent fragments. The resulting methodology is benchmarked as applied to model systems that allow for full QM optimization as well as through refinement of the hydrogen bonding geometry in *Oxytricha nova* guanine quadruplex for which several studies have been reported, including the X-ray structure and NMR data. Calculations of <sup>1</sup>H NMR chemical shifts based on the gauge independent atomic orbital (GIAO) method and direct comparisons with experiments show that solvated MoD-QM/MM structures, sampled from explicit solvent molecular dynamics simulations, allow for NMR simulations in much improved agreement with experimental data than models based on the X-ray structure or those optimized using classical molecular mechanics force fields.



## INTRODUCTION

Many biological processes, including proton transport,<sup>1–5</sup> electron transfer,<sup>6</sup> enzyme catalysis,<sup>7</sup> and ligand binding<sup>8</sup> are strongly affected by electrostatic interactions. These interactions also influence the molecular geometry, protonation states of proteins, and contribute significantly to the overall energetics. Therefore, an accurate description of the molecular electrostatic potential is essential for realistic simulations of a wide range of systems. This paper introduces a self-consistent computational protocol for structural refinement of extended systems, as a generalized version of the linear scaling Moving-Domain Quantum-Mechanics/Molecular-Mechanics (MoD-QM/MM) hybrid method.<sup>9–11</sup>

The computational effort associated with conventional electronic structure methods scales steeply with the size of the molecular system, typically hindering direct ab initio calculations of systems with hundreds or thousands of atoms. Therefore, approximate approaches are usually implemented for calculations of molecular electrostatic potentials (MEP) of large biological molecules such as semiempirical methods,<sup>12–14</sup> or in a mean field manner via static point charge distributions (e.g., AMBER,<sup>15</sup> CHARMM,<sup>16</sup> and OPLS<sup>17,18</sup> force fields). However, standard molecular mechanics (MM) force fields do

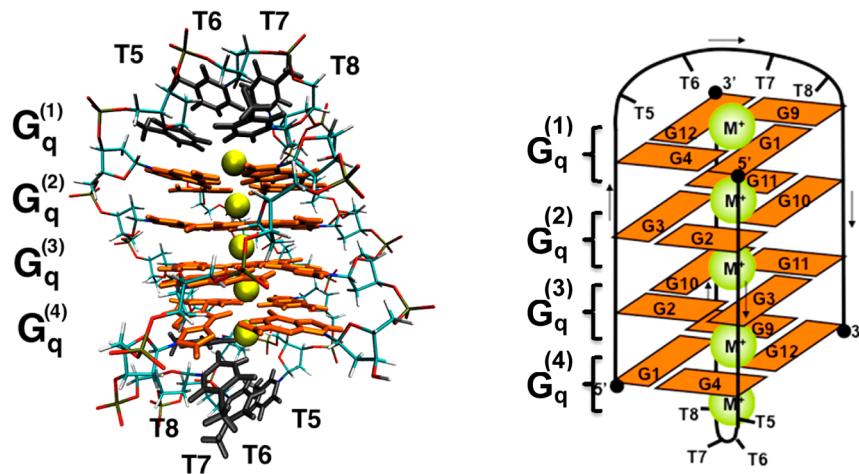
not provide ab initio quality MEPs and do not account for effects of induced electronic polarization between atoms. Significant research efforts have been focused on resolving this problem by developing polarizable MM force fields,<sup>19–26</sup> which are still computationally demanding and not as extensively implemented as classical force fields based on static point charge distributions. This is partially due to the computational cost of polarizable force fields and the intrinsic difficulty of polarization effects, which are system dependent.

The development of linear scaling quantum mechanical methods offers a promising strategy for describing the MEP and other properties (e.g., energies) of large systems at a significantly reduced computational cost. Fragment-based approaches partition the molecule into subsystems and subsequently combine the calculations of the fragments to predict the corresponding properties for the whole system. Both density- and energy-based fragmentation methods have been proposed. Examples of the former include the “divide-and-conquer” method developed by Yang,<sup>27</sup> the adjustable density matrix assembler approach of Exner and Mezey,<sup>28–32</sup>

Received: July 3, 2014

Published: September 24, 2014





**Figure 1.** (Left) Representation of the DNA quadruplex of *Oxytricha nova* d(GGGGTTTTGGGG) (PDB code: 1JPQ<sup>76</sup>). The guanine quartets ( $G_q^{(1)}$ ,  $G_q^{(2)}$ ,  $G_q^{(3)}$ , and  $G_q^{(4)}$ ) are shown in orange. Yellow spheres represent the potassium ions in the central channel. Thymine ( $T_4$ ) loops are shown in gray. (Right) Schematic representation with labels for individual guanines (G1–G4, G9–G12) arranged in quartets ( $G_q^{(1)}\text{--}G_q^{(4)}$ ) and thymines (T5–T8) in the loops. Arrows indicate the  $5' \rightarrow 3'$  phosphodiester backbone direction.

and the fragment molecular orbital approach of Kitaura and co-workers.<sup>33–35</sup> Energy-based fragmentation methods include the molecular tailoring approach (MTA),<sup>36,37</sup> the kernel energy method,<sup>38,39</sup> the molecular fractionation with conjugate caps (MFCC) method,<sup>40,41</sup> the generalized energy-based fragmentation method,<sup>42</sup> and systematic fragmentation methods.<sup>43–48</sup> Other related approaches include the effective fragment potential method,<sup>49–51</sup> as well as the explicit polarization (X-Pol) potential,<sup>52–56</sup> which are essentially electronic structure based force fields. Some of these methods have already been successfully applied to evaluate the MEP of large molecular systems.<sup>37,46,57</sup> Also, a comprehensive discussion of fragmentation methods has recently been reviewed.<sup>58</sup>

An alternative strategy to model large molecular systems is to use hybrid Quantum Mechanics/Molecular Mechanics (QM/MM) methods,<sup>6,59–63</sup> such as the ONIOM method.<sup>59–61</sup> QM/MM methods model a fragment of the system at the QM level, and the rest is treated more approximately, either through MM force fields, semiempirical methods, or an inexpensive ab initio method (e.g., Hartree–Fock (HF) or density functional theory (DFT) with a modest basis set). In these methods, one can take advantage of electronic embedding whereby atomic charges of the MM region are incorporated into the quantum mechanical Hamiltonian (i.e., the QM subsystem is polarized by the surrounding MM point charges) so as to provide an improved description of the electrostatic interaction between the QM and MM regions. Variants of this method such as QM:QM embedding using electron densities,<sup>64</sup> Mulliken atomic charges,<sup>65</sup> many-body QM:QM methods,<sup>66–69</sup> as well as the molecules-in-molecules (MIM)<sup>70</sup> and extended ONIOM (XO)<sup>71</sup> methods have also been developed.

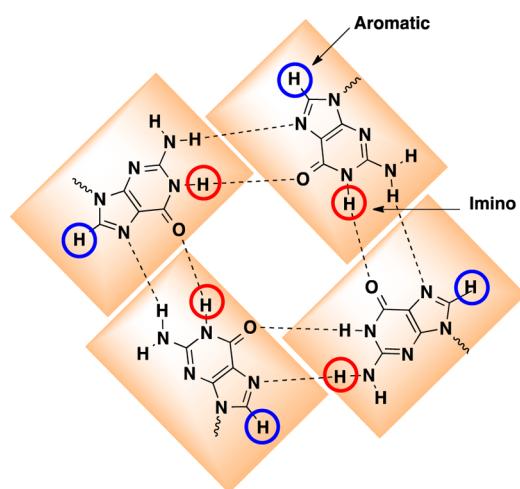
The Moving-Domain QM/MM (MoD-QM/MM) method<sup>9,10</sup> combines the fragmentation approach of linear scaling methods and the QM/MM strategy to obtain ab initio MEP of large biological macromolecules (e.g., proteins, DNA, etc.) in a given configuration. Under the MoD-QM/MM protocol, the system is partitioned into molecular domains according to a space-domain decomposition scheme. Atomic charges of the constituent QM domains are computed sequentially by fitting them to reproduce the electrostatic potential produced by the QM charge density in the surrounding MM environment. After

updating the distribution of atomic charges in each domain the whole cycle is iterated until obtaining a self-consistent point-charge model of the MEP. Such an iterative scheme accounts for mutual polarization effects and usually converges within a few iteration-cycles, (i.e., 4 or 5 cycles) scaling linearly with the size of the system. Therefore, the method bypasses the enormous demands of memory and computational resources that would be required by a “brute-force” quantum chemical calculation of the complete system. The MoD-QM/MM method has been successfully implemented for the description of several systems, including photosystem II,<sup>72</sup> the KscA potassium ion channel and green fluorescent protein<sup>9</sup> as well as in calculations of Stark shifts.<sup>73</sup>

This paper generalizes the MoD-QM/MM methodology to obtain not only self-consistent charges but also self-consistent geometries of each domain upon convergence of the iterative cycle. The resulting approach exploits the computational efficiency of QM/MM, as applied to each constituent fragment, allowing for structural refinement of extended systems at the ab initio level. We illustrate the generalized MoD-QM/MM method as applied to structural refinement of a guanine quadruplex from the telomeric sequence of the ciliate *Oxytricha nova*. The quadruplex involves a dimer of short guanine-rich DNA sequences d(GGGGTTTTGGGG) folded into a stack of four quartets ( $G_q^{(1)}$  to  $G_q^{(4)}$ ) with hydrogen-bonded guanine nucleotides (Figure 1). The quadruplex requires monovalent cations (e.g.,  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{NH}_4^+$ ) for maintenance of structural integrity.<sup>74</sup> When formed in a telomere at the end of a chromosome, the folded quadruplex prevents the action of the telomerase enzyme that enables cancerous growth by adding TTAGGG repeats to the 3' end of DNA. Therefore, there is significant interest in understanding the structural stability of quadruplexes in telomeres and how to stabilize the folded conformation by using small molecules as specific chemotherapeutic agents.<sup>75</sup>

Several theoretical studies have focused on the nature of hydrogen bonds that stabilize the guanine quartets in DNA quadruplexes (Scheme 1).<sup>77–85</sup> For example, Bickelhaupt and co-workers have carried out energy decomposition analyses to investigate whether cooperativity originates from charge separation due to donor–acceptor orbital interactions of  $\sigma$ -

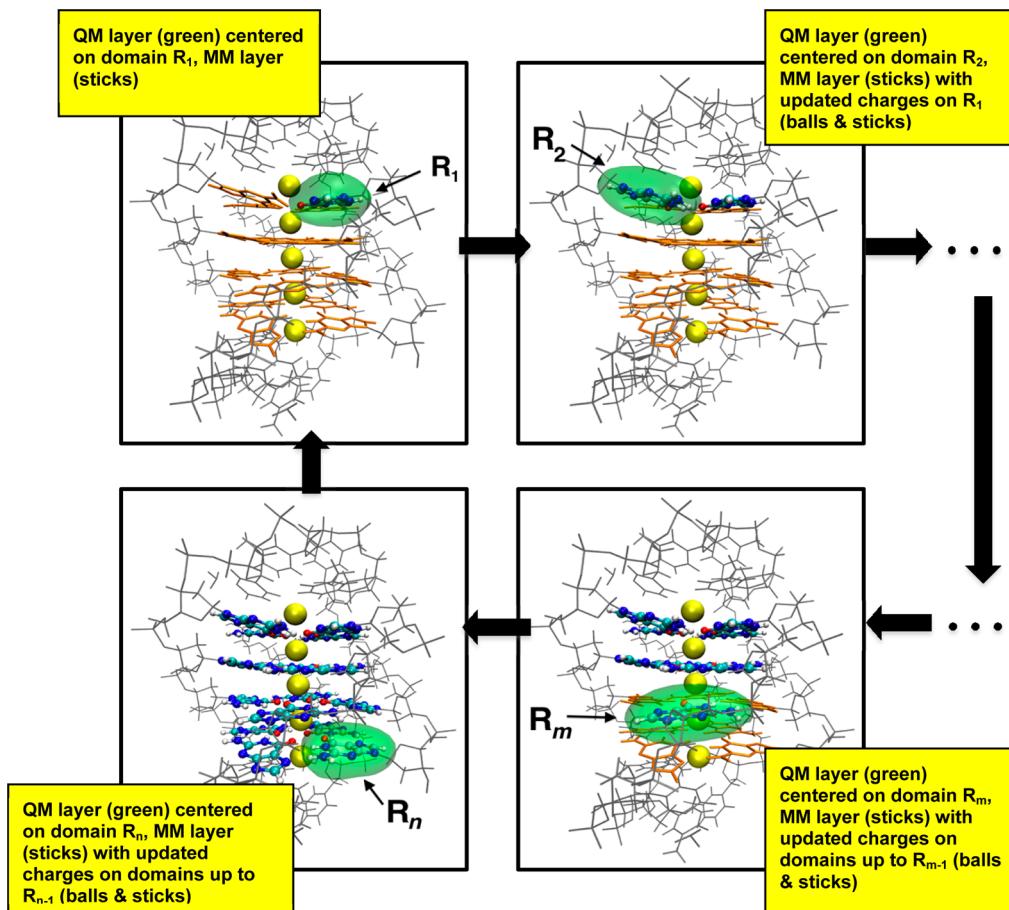
**Scheme 1. Aromatic and Imino Protons in a Guanine Quartet**



electrons, rather than resonance assistance of the  $\pi$ -electron system.<sup>83</sup> Marek and co-workers employed various techniques including the Quantum Theory of Atoms in Molecules (QTAIM), natural bond orbital (NBO) analysis, energy decomposition analysis (EDA), and noncovalent interaction analysis (NCA) to examine the contributions from hydrogen bonding,  $\pi-\pi$  stacking and ion coordination to the stability of a

model system of guanine quadruplex.<sup>85</sup> In a recent study, Grimme and co-workers carried out large scale DFT-D3 computations to predict energetic differences between different arrangements of guanine quadruplex stems.<sup>86</sup> Based on a model system, the authors found large systematic differences between the MM (AMBER) and QM predictions of relative stabilities of different guanine quadruplex stem topologies, attributed to the neglect of polarization effects in the MM calculations.

Many ab initio studies of DNA quadruplexes are based on model systems where the backbone groups (i.e., thymine, anionic phosphate, and sugar moieties), solvent, and counterions are omitted. On the other hand, full structural models of DNA quadruplexes have been studied using classical molecular dynamics simulations.<sup>87–94</sup> Several recent studies have found that polarization effects are critical for describing the dynamical structure of the quadruplex.<sup>94–96</sup> Šponer and co-workers have reported explicit solvent molecular dynamics simulations of DNA quadruplexes using various classical force fields and observed that none of the presently available force fields accurately describe the quadruplex loops. Notably, the cations in the loop region rapidly escaped into the bulk solution at early stages of the simulation as a result of missing polarization effects in classical force fields.<sup>93</sup> In a recent MD simulation study, Zhang and co-workers have employed a new charge model that has been “pre-polarized” based on a reference structure of a DNA quadruplex.<sup>94</sup> The incorporation of polarized MM charges was critical for retaining the cations in



**Figure 2.** Representation of the MoD-QM/MM method for structural refinement. Green surfaces represent the QM region for each step of the cycle. Colored balls and sticks represent domains with geometries and charges updated in previous steps of the cycle.

the thymine loops and properly modeling the G-quadruplex structure.

In addition to previous theoretical studies, experimental structural information has been reported including the X-ray structure at 1.6 Å resolution<sup>97</sup> and NMR data.<sup>74</sup> Therefore, the guanine quadruplex offers an ideal benchmark system for studies of electrostatic and structural refinement based on the MoD-QM/MM method. This paper is organized as follows: We first describe the generalized MoD-QM/MM methodology and benchmark calculations where we apply MoD-QM/MM to the description of individual guanine quartets embedded in the *Oxytricha nova* guanine quadruplex. Having validated the method as directly compared to full QM calculations of quartets, we apply explicit solvent molecular dynamics simulations on the full structural model of guanine quadruplex<sup>97</sup> for obtaining a representative ensemble of solvated configurations, which were then refined by the MoD-QM/MM protocol. Finally, we employ the gauge independent atomic orbital (GIAO) method<sup>88,89</sup> to compare the calculated <sup>1</sup>H NMR chemical shifts of aromatic and imino protons (Scheme 1), which are sensitive probes of both hydrogen bonds and stacking in the quadruplex, to readily available experimental data.<sup>74</sup> We show that the ensemble of configurations refined by the MoD-QM/MM methodology leads to much better agreement with experiment than configurations generated with popular molecular mechanics force fields such as AMBER or the X-ray structure.

## ■ COMPUTATIONAL DETAILS

**Generalized MoD-QM/MM Method.** A detailed description of the MoD-QM/MM methodology for ab initio computations of molecular electrostatic potentials of extended systems in predefined configurations has already been reported.<sup>10,11</sup> Here, we describe a straightforward generalization of the MoD-QM/MM method to obtain self-consistent geometries as well as molecular electrostatic potentials of extended systems in relaxed configurations (Figure 2). As in the original formulation, the method requires fragmentation of the system into molecular domains and implements an iterative cycle of QM/MM calculations.

The generalized MoD-QM/MM methodology can be implemented with any QM/MM method that explicitly considers polarization of the QM layer as influenced by the surrounding MM environment. In this study, we have used the ONIOM-electronic embedding (EE) approach introduced by Morokuma and co-workers,<sup>59–61</sup> as implemented in Gaussian09<sup>98</sup> with the standard hydrogen link atom scheme<sup>99</sup> for the boundary between QM and MM regions. In the two-layer ONIOM(QM:MM) method, the region of interest is treated with rigorous ab initio methods (denoted *X*), while the rest of the molecule (denoted region *Y*) is treated with a MM force field. The ONIOM energy expression is

$$E^{\text{ONIOM}(X:Y)} = E_X^{\text{QM}} + E_{X+Y}^{\text{MM}} - E_X^{\text{MM}} \quad (1)$$

where  $E_{X+Y}^{\text{MM}}$  is the energy calculated for the whole system using a MM force field.  $E_X^{\text{QM}}$  and  $E_X^{\text{MM}}$  are calculated for the region *X* at the QM and MM levels of theory, respectively. With electronic embedding, the electrostatic interactions between the layers are included in the QM calculations by adding the Coulombic interactions between the QM electrons and nuclei with partial charges in the MM region in the Hamiltonian.<sup>60</sup> This allows the QM charge distribution to be polarized,

providing an improved description of electrostatic interactions between the two layers. The interactions between the two layers are treated at the MM level, modeled according to a classical force field such as AMBER:<sup>100</sup>

$$E_{\text{int}}^{\text{MM}} = \sum_{i < j} \left[ \alpha_{ij}^{\text{vdW}} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right) + \alpha_{ij}^q \frac{q_i q_j}{\epsilon r_{ij}} \right] \quad (2)$$

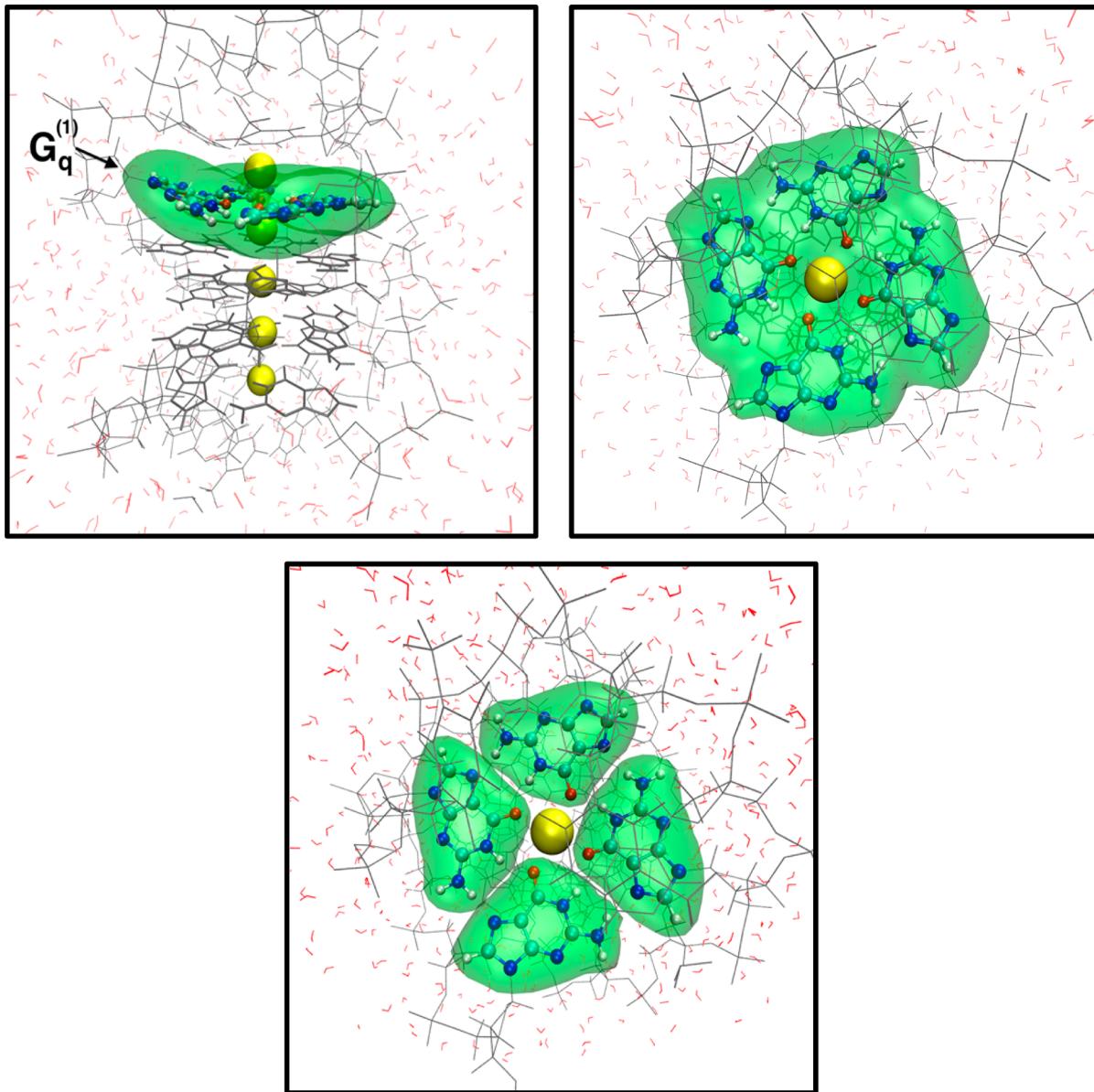
As described above, the MoD-QM/MM procedure fragments the QM layer (*X*) into smaller domains ( $X_1, X_2, \dots, X_N$ ) and the energy of the system is obtained from the resulting multilayer  $\text{QM}_1:\text{QM}_2:\dots:\text{QM}_N/\text{MM}$  expression (eq 3). Polarization of the fragments is thus described at the QM level like in other QM:QM methods,<sup>66–69</sup> molecules-in-molecules (MIM),<sup>70</sup> the extended ONIOM (XO),<sup>71</sup> and the hybrid many-body interaction (HMBI).<sup>101</sup>

$$\begin{aligned} E^{\text{ONIOM}(X_1, X_2, \dots, X_N:Y)} &= E_{X_1, X_2, \dots, X_N}^{\text{QM}} + E_{X_1, X_2, \dots, X_N+Y}^{\text{MM}} - E_{X_1, X_2, \dots, X_N}^{\text{MM}} \\ &\sim E_{X_1}^{\text{QM}} + E_{X_2}^{\text{QM}} + \dots + E_{X_N}^{\text{QM}} + E_{X_1, X_2, \dots, X_N+Y}^{\text{MM}} - E_{X_1}^{\text{MM}} \\ &\quad - E_{X_2}^{\text{MM}} - \dots - E_{X_N}^{\text{MM}} = \sum_{i=1}^N E^{\text{ONIOM}(X_i: \bar{X}_i Y)} \\ &\quad - (N-1) E_{X_1, X_2, \dots, X_N+Y}^{\text{MM}} \end{aligned} \quad (3)$$

$$\begin{aligned} \frac{\partial E^{\text{ONIOM}(X_1, X_2, \dots, X_N:Y)}}{\partial q} &\sim \sum_{i=1}^N \frac{\partial E^{\text{ONIOM}(X_i: \bar{X}_i Y)}}{\partial q} \\ &\quad - (N-1) \frac{\partial E_{X_1, X_2, \dots, X_N+Y}^{\text{MM}}}{\partial q} \end{aligned} \quad (4)$$

The simplest implementation of MoD-QM/MM geometry optimization requires a calculation of the QM gradients for all domains,  $X_1-X_N$  (eq 4), and then update of the configuration of the domains by steepest descent based on the calculated gradients. With the updated configuration, a second round of QM/MM calculations is required to update the ESP atomic charges of the domains before computing the gradients for the next optimization step. The process is then repeated until the gradients are sufficiently small and the configuration is converged. Alternatively, one can compute the gradients of each fragment and update their coordinates and charges sequentially, using the most updated description of the fragments previously considered in the sequence. We implement this sequential approach, known as the (block) coordinate descent algorithm,<sup>102,103</sup> which has been successfully applied in robotics and in protein loop prediction.<sup>104</sup> It is also similar in spirit to the ONIOM geometry optimization scheme implemented in Gaussian 09, where a series of microiterations is employed to fully optimize the MM region before each optimization step in the QM region.<sup>105</sup> While the coordinate descent algorithm is usually an effective optimization method, it does have limitations since it achieves only local optimization relative to a reference configuration. Therefore, it should work best only for small to moderate rearrangements.

Figure 2 illustrates the generalized MoD-QM/MM procedure as implemented for the *Oxytricha nova* guanine quadruplex. For simplicity, each guanine defines an individual QM domain although larger domains (e.g., pairs of guanines) could also be defined. Starting with a QM layer centered on the molecular domain  $R_1$ , a QM/MM optimization is performed subject to the constraint of fixed nuclear configuration for the



**Figure 3.** Illustration of a guanine quartet model system used in benchmark calculations of MoD-QM/MM optimization as compared to reference QM/MM optimization where all four guanines in the quartet are included in the QM layer (green) while the rest of the system is modeled at the MM (AMBER) level. Top left: Side view. Top right: Top view. Yellow spheres represent the potassium ions. Bottom: The four QM domains defined in the corresponding MoD-QM/MM calculation.

other fragments in the surrounding environment. Upon convergence, the ESP charges of  $R_1$  are updated. In the next step, the QM domain is moved to  $R_2$  and its geometry and ESP charges are updated by QM/MM analogously. The rest of the system defines the MM layer, including  $R_1$  with its nuclear configuration and distribution of ESP charges as computed in the previous step. The procedure is analogously applied to all fragments and the entire computational cycle is iterated several times until reaching self-consistency in the description of the geometry and distribution of atomic charges for all domains. We have employed atomic charges that have been fitted to reproduce the electrostatic potential at points selected according to the Merz–Singh–Kollman scheme<sup>106,107</sup> subject to the constraint of zero charge for the link hydrogen atom.<sup>10</sup> For comparison, we have also examined the use of restrained ESP charges (RESP),<sup>108,109</sup> and its impact on the resulting MoD-QM/MM optimized geometry. The  $\omega$ B97XD density

functional theory method<sup>110</sup> and the AMBER-99 force field<sup>111</sup> were applied for the description of the QM and MM layers, respectively. NMR calculations were carried out by using the gauge independent atomic orbital (GIAO) method<sup>112,113</sup> in conjunction with DFT methods ( $\omega$ B97XD, B3LYP, and mPW1PW91). The MoD-QM/MM procedure was carried out using our in-house program MODQ3M.<sup>114</sup>

**System Preparation.** The crystal structure of the *Oxytricha nova* guanine quadruplex (PDB code: 1JPQ at 1.6 Å resolution)<sup>76</sup> was used as our starting model structure. The bare G-quadruplex carries a net charge of  $-17\text{ e}$ ; therefore, solvation and inclusion of counterions are essential for maintaining structural stability. The AmberTools12 package<sup>115</sup> was used to add hydrogen atoms and 17 potassium counterions near the phosphate groups to prepare an overall electrically neutral model. The potassium Lennard-Jones parameters (radius 1.705 Å and well depth 0.194 kcal mol<sup>-1</sup>), optimized

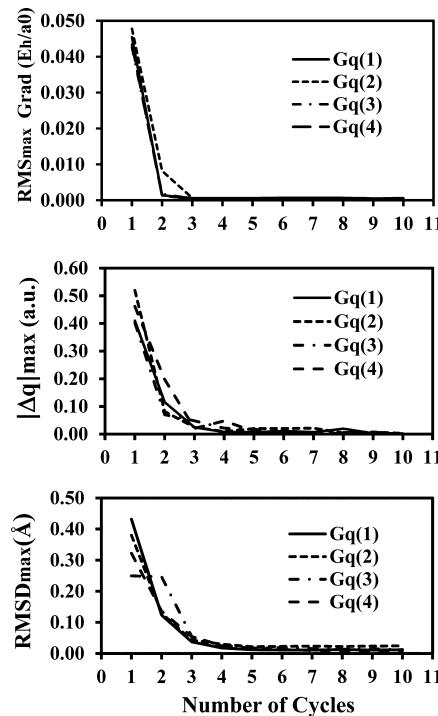
by Cheatham and co-workers, were used in our work.<sup>116</sup> Classical molecular dynamics simulations in a 5.5 nm cubic box of TIP3P water with periodic boundary conditions were carried out for equilibration of the guanine quadruplex structure. To maintain the structural integrity of the complex, the DNA atoms were restrained at their crystallographic positions during the simulation. After equilibration for 2 ns at 300 K, configurations that include the G-quadruplex and all water molecules and counterions within 20 Å of the guanine center were subsequently relaxed at the DFT level using the MoD-QM/MM method, and also at the MM level (vide infra). The <sup>1</sup>H NMR chemical shifts of aromatic and imino protons were computed for each relaxed configuration in order to estimate the ensemble-averaged chemical shifts that could be directly compared to experiment. All MD simulations were carried out by using the NAMD package.<sup>117</sup> Visualization and analysis of the MD simulations and MoD-QM/MM optimizations were performed using the VMD program.<sup>118</sup>

## RESULTS AND DISCUSSION

**Benchmark Calculations on Model Systems.** Our benchmark calculations aimed to address several questions regarding performance of the generalized MoD-QM/MM procedure, including accuracy of the iterative scheme based on individual guanines as compared to calculations where the QM layers include entire quartets, rate of convergence, and performance based on ESP versus restrained ESP (RESP) atomic charges.

The individual guanine quartets,  $G_q^{(1)}$ ,  $G_q^{(2)}$ ,  $G_q^{(3)}$ , and  $G_q^{(4)}$ , depicted in Figure 1, were employed as model systems in our benchmarking study. We compared the relaxed structures of each quartet obtained by implementing the MoD-QM/MM method with QM domains defined by individual guanines to the corresponding relaxed geometry of the quartet obtained at the QM/MM level using a QM layer that spans the entire quartet (Figure 3). The MoD-QM/MM method treats the interactions between individual guanines at the MM level, albeit using geometries and atomic charges obtained via the MoD-QM/MM methodology. Computations were carried out at the ONIOM( $\omega$ B97XD/6-31G(d):AMBER)-EE level of theory, where the long-range corrected hybrid functional  $\omega$ B97XD includes empirical atom–atom dispersion corrections (Grimme's D2 dispersion model),<sup>110</sup> shown to accurately describe noncovalent interactions.<sup>119,120</sup> The specific solvent environment surrounding the guanine quadruplex was sampled from molecular dynamics simulations, after equilibration of the system to include all water molecules and counterions that are within 20 Å of the guanine quadruplex center. All solvent molecules, counterions, and DNA atoms in the MM layer (except QM domains) were allowed to relax during the QM/MM optimization.

Figure 4 shows typical convergence rates for the computed energy gradients, atomic charges, and geometries of the guanine quartets, as quantified by the maximum RMS gradient, maximum change in the atomic charges ( $|\Delta q|_{\max}$ ), and maximum root-mean-square deviations (RMSD<sub>max</sub>) of nuclear configurations for the QM domains along the iterative cycles of MoD-QM/MM optimization. Figure 4 shows that all the RMS gradients, atomic charges, and RMSDs converge rapidly to very small values (below 0.0005 E<sub>h</sub>/a<sub>0</sub>, 0.005 e, and 0.01 Å respectively) in about 5 cycles. This rate of convergence is comparable to the performance of MoD-QM/MM calculations of ESP charges for systems with fixed nuclear configurations,<sup>9</sup>



**Figure 4.** Convergence of benchmark parameters. Top: Maximum RMS gradient (in au). Middle: Maximum change in atomic charges (in au). Bottom: Maximum RMSD of the nuclear configurations (in Å) of quartets  $G_q^{(1)} - G_q^{(4)}$ , along successive iterations of MoD-QM/MM optimization cycle.

showing that including geometry optimization does not significantly affect the number of cycles necessary for convergence. As described in the next section, a very similar rate of convergence is observed when the MoD-QM/MM is applied to the full model of *Oxytricha nova* guanine quadruplex composed of 16 domains. Accordingly, the computational time associated with this procedure is  $\tau = N_c \tau_0 n$ , where  $N_c$  is the number of MoD-QM/MM cycles (ca. 5) needed for convergence,  $\tau_0$  is the average computational time required for the ONIOM geometry optimization of an individual QM domain, and  $n$  is the number of molecular domains in the system. In the present case,  $\tau_0$  is approximately 7 min (based on a single CPU), and the computational time associated with the MoD-QM/MM procedure is approximately 2.5 h. By comparison, the benchmark ONIOM( $\omega$ B97XD/6-31G(d):AMBER)-EE calculation on the entire quartet takes about 12 h, indicating a 5-fold savings in computational time. The savings in computational time and memory requirements are even more dramatic when comparing to calculations with larger domains (e.g., a domain including all four quartets).

Table 1 reports RMSDs between the geometry of each quartet relative to the reference minimum energy configuration optimized by including complete quartets in the QM layer. For completeness, Table 1 compares results for the MoD-QM/MM structure obtained with electronic embedding (EE) and individual guanines in the QM domains, to the analogous results for the X-ray crystal structure and the minimum energy structure obtained by AMBER as well as mechanical embedding (ME), rather than electronic embedding (EE) (i.e., neglecting polarization effects). The results show that MoD-QM/MM optimization with EE displays the smallest averaged RMSD, whereas all other structures display values that are significantly

**Table 1.** RMSDs ( $\text{\AA}$ ) for Each Quartet, Relative to the ONIOM( $\omega$ B97XD/6-31G(d):AMBER)-EE Reference Structure

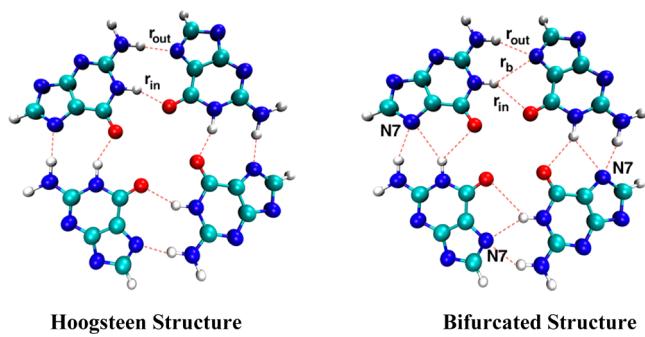
	ONIOM-ME <sup>a</sup>	MoD-QM/MM	AMBER	X-ray <sup>c</sup>
$G_q^{(1)}$	0.07	0.05	0.11	0.16
$G_q^{(2)}$	0.09	0.06	0.38	0.22
$G_q^{(3)}$	0.53	0.04	0.26	0.24
$G_q^{(4)}$	0.11	0.08	0.17	0.18
$\langle \text{RMSD} \rangle^b$	0.20	0.06	0.23	0.20

<sup>a</sup>The QM layer includes all four guanines in the quartet. <sup>b</sup>RMSD averaged over four quartets. <sup>c</sup>Only hydrogen atoms are relaxed using the AMBER force field.

higher since they were obtained by neglecting polarization effects or the structural perturbation due to the surrounding solvent.

Structures optimized at the ONIOM-ME level of theory, or with a classical force field (e.g., AMBER) were observed to display RMSDs that are higher than the X-ray model relative to the reference structure. This is due to the structure of hydrogen bonding in those models where quartets  $G_q^{(2)}$  and  $G_q^{(3)}$  are stabilized by a bifurcated hydrogen bond configuration (Scheme 2 right) with an additional hydrogen bond ( $r_b$ )

**Scheme 2.** The Bifurcated Geometry (Right) Has an Additional Hydrogen Bond ( $r_b$ ) between the N1 and N7 Atoms



between N1 and N7. Such a hydrogen bond motif is not seen in the X-ray, MoD-QM/MM, or reference structures (Scheme 2 left), and leads to deviations in hydrogen bond distances and angles by as much as 1.2  $\text{\AA}$  and 27°, relative to the reference structure. Specifically, the Hoogsteen geometry in the reference structure has both inner ( $r_{in}$ ) and outer ( $r_{out}$ ) hydrogen bonds of about 1.9  $\text{\AA}$  while in the bifurcated configuration the corresponding hydrogen bonds are of about 2.8 and 2.0  $\text{\AA}$ , respectively. In contrast, the hydrogen bond distances and angles in the MoD-QM/MM structures are generally within 0.1  $\text{\AA}$  and 3° relative to the reference configuration (Full structural data provided in Tables S1–S3 in Supporting Information). These are important results, considering the efficiency of the MoD-QM/MM methodology when implemented with one guanine per domain and its remarkable ability to provide a proper description of hydrogen bonding.

To verify that the bifurcated configuration is not in fact a local minimum, we have also performed ONIOM( $\omega$ B97XD/6-31G(d):AMBER)-EE and MoD-QM/MM calculations starting from the bifurcated geometry and we found that the geometry converges to the Hoogsteen configuration, while optimizations with the MM force fields or ONIOM-ME converge to the

bifurcated configuration, even when starting with the Hoogsteen hydrogen-bonding configuration. This is likely due to the lack of polarization effects in standard MM force fields and mechanical embedding calculations. Similar observations were also observed in several earlier studies employing classical force field calculations<sup>89,94</sup> and ascribed as an artifact of the nonpolarizable force-field which underestimates the cation-base interaction.

An important technical aspect that should be noted is that unrestrained ESP charges typically overestimate bond polarities.<sup>108</sup> Therefore, we have also examined the effect of using restrained atomic charges in our MoD-QM/MM calculations through the inclusion of a hyperbolic restraint function during the fitting of the classical electrostatic potential. Restraint weights ( $k_{\text{str}}$ ) with values ranging from 0.001 to 0.01, as recommended for calculations of atomic charges for polar molecules,<sup>108,109</sup> have been tested (ESP and RESP atomic charges obtained for various values of  $k_{\text{str}}$  are reported in Table S5 of the Supporting Information). Increasing  $k_{\text{str}}$  leads to charges that are smaller in amplitude but increases the RMSD between the QM and fitted electrostatic potentials. Table 2

**Table 2.** Deviations in Hydrogen Bond Distances ( $\text{\AA}$ ) and Angles (deg), Relative to the Reference Structure, for Various Values of the Restraint Weight ( $k_{\text{str}}$ )

$k_{\text{str}}^a$	0.00 (ESP)	0.0005	0.001	0.005	0.01
$\langle \Delta r \rangle^b$	0.08	0.07	0.06	0.04	0.04
$\langle \Delta \theta \rangle^b$	2.21	1.24	0.82	0.62	1.03
$\Delta r_{\text{max}}$	0.15	0.13	0.12	0.10	0.08
$\Delta \theta_{\text{max}}$	4.01	3.02	1.53	2.35	2.63

<sup>a</sup>Charges on non-hydrogen atoms have been restrained to a value of zero. <sup>b</sup>Values obtained from averaging over four quartets.

shows that  $k_{\text{str}} = 0.005$  provides the best description of hydrogen bonding as compared to the reference structure, as described by the average absolute deviation in hydrogen bond distances ( $\Delta r$ ) and angles ( $\Delta \theta$ ) for MoD-QM/MM structures obtained with various different values of  $k_{\text{str}}$ . The definition of the geometrical parameters and full data are provided in Table S1 of the Supporting Information. Notably, we observed that MoD-QM/MM optimized geometries employing ESP charges led to hydrogen bonds that are typically 0.1  $\text{\AA}$  shorter than the benchmark structure presumably due to overpolarization. As such, the value of  $k_{\text{str}} = 0.005$  has been adopted for all MoD-QM/MM calculations presented in this study.

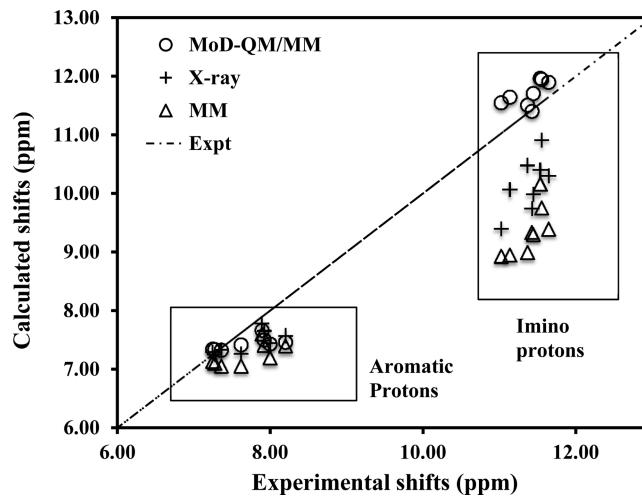
**Structural Refinement and NMR Assessment.** The structural refinement of the X-ray structure of *Oxytricha nova* G-quadruplex (PDB code: 1JPQ)<sup>97</sup> has been performed by using the validated MoD-QM/MM methodology, as implemented with the efficient partitioning scheme including one guanine per QM domain and the RESP charge fitting procedure described in the previous section. With 4 quartets in the quadruplex, each MoD-QM/MM cycle involves 16 QM/MM optimizations. Nevertheless, the convergence rate of the iterative procedure is comparable to that observed in model systems (Figure 4), suggesting that the overall convergence is relatively insensitive to the size of the system.

The resulting MoD-QM/MM refined structure was validated through calculations of  $^1\text{H}$  NMR chemical shifts of imino and aromatic protons (Scheme 1) and direct comparisons with experimental data. Thymine (T4) loops were not compared to

experimental data since previous NMR and crystallographic studies<sup>121,122</sup> have indicated the presence of conformational exchange suggesting that the loop conformation observed in the crystal structure is one of several that exist in solution. For reliable comparisons of NMR chemical shifts, it is important to consider the influence of explicit solvent molecules and counterions.<sup>123–125</sup> In this work, solvated configurations were sampled by using *NPT* explicit solvent molecular dynamics (see Computational Details). After equilibration, snapshots were randomly sampled from a 2 ns trajectory, including all water molecules and counterions within 20 Å of the center of the guanine quadruplex. The sampled configurations typically included 900–1000 waters in addition to the DNA quadruplex and the counterions.

Each configuration was subsequently optimized by using the MoD-QM/MM algorithm, where all the DNA backbone atoms, solvent molecules, and counterions in the MM layer (with the exception of the QM domains) were allowed to relax during the minimization. The isotropic shielding constants for aromatic and imino protons were computed for all relaxed configurations, using the GIAO method<sup>112,113</sup> at the ONIOM-( $\omega$ B97XD/6-31G(d,p):AMBER)-EE level. Comparisons of chemical shifts obtained by using the B3LYP and mPW1PW91 methods showed that both functionals give very similar results (see Tables S5–S7 in Supporting Information) when using an extended QM layer including all 16 guanines (Figure 1; QM layer shown in orange) to include shielding effects due to  $\pi$ -electrons in all quartets. The calculated isotropic shielding constants were averaged over all sampled configurations and subtracted from the corresponding shielding constant of tetramethylsilane (TMS) computed at the same QM level of theory.<sup>125,126</sup> In this way, the combined effects of solvent and its influence on the local geometry of the G-quadruplex, on calculated chemical shifts, are taken into account. For comparison, we have also carried out analogous calculations based on configurations obtained by classical optimization at the AMBER MM level and for the X-ray structure. For the X-ray model, only the positions of added hydrogen atoms and explicit water molecules from the MD snapshots were optimized by using the AMBER MM force field, keeping fixed all backbone atoms at their crystallographic positions. For the MM model, all the atoms are relaxed using the AMBER force field. The AMBER library charges were employed for the ONIOM NMR calculations in both models.

Figure 5 compares the experimental values<sup>74,122</sup> of  $^1\text{H}$  NMR chemical shifts to the ensemble averages computed at the ONIOM( $\omega$ B97XD/6-31G(d,p):AMBER)-EE level of theory for structures relaxed according to the MoD-QM/MM, X-ray and AMBER optimization. Standard deviations of the averages are 0.05–0.07 ppm for structures relaxed at the MoD-QM/MM level and AMBER MM force field and 0.03 ppm for the X-ray structure. As shown in Figure 5, the calculated chemical shifts for aromatic protons in all models are generally in very good agreement with experimental data, with mean absolute deviations of 0.29, 0.25, and 0.44 ppm for the MoD-QM/MM, X-ray, and AMBER minimized structures, respectively. However, the calculated chemical shifts for imino protons are in agreement with experimental data (MAD 0.32 ppm) only when computed with structures relaxed at the MoD-QM/MM level, within the errors of typical ab initio calculations of NMR chemical shifts of proteins and related macromolecular systems at similar levels of theory.<sup>124–127</sup> We have analyzed the effect of updating the charges, keeping fixed the X-ray structure to show



**Figure 5.** Comparison of experimental  $^1\text{H}$  NMR chemical shifts to ensemble averages computed at the ONIOM( $\omega$ B97XD/6-31G(d,p):AMBER)-EE level of theory for structures relaxed according to the MoD-QM/MM method (circles), classical AMBER force field (triangles), and the X-ray structure (crosses). A perfect correlation is given by the dashed line.

how much of the correction of NMR chemical shifts comes from updating charges and how much from geometrical optimization. It turns out that the correction of charges has a very small effect on the computed NMR shifts compared to using library AMBER charges (see Table S8 in the Supporting Information) presumably because the predominant electrostatic effects are already included in the QM layer of our ONIOM NMR calculations. Therefore, the corrections shown in Figure 5 are mostly due to the improvement in the geometry provided by the MoD-QM/MM method.

The comparative analysis of stacking and hydrogen bonds in the three sets of structures provides insights on the origin of deviations in the calculated imino  $^1\text{H}$  NMR, relative to experimental data. To estimate the stacking distances between the guanine quartets, the channel of potassium ions (see Figure 1) was first aligned with the z-axis, and the z-coordinate associated with each quartet is estimated from the average of z-coordinates of all atoms in that plane. We found that the ensemble-averaged stacking distances are very similar for the three sets of structures (3.42, 3.41, and 3.47 Å for the MoD-QM/MM, X-ray, and AMBER structures, respectively) while the hydrogen bonds are significantly different for the X-ray and AMBER models. Specifically, formation of the bifurcated hydrogen bonding geometry (Scheme 2) in the AMBER models stretches the hydrogen bonds of imino hydrogens, leading to significant changes in their magnetic environment. In contrast, MoD-QM/MM structures display Hoogsteen hydrogen bonding with shorter imino hydrogen bonds, giving much better agreement with the experimental NMR data. For comparison, the imino hydrogen bonds ( $r_{\text{in}}$  in Scheme 2) averaged over all guanine quartets are 1.83, 1.93, and 2.18 Å for the MoD-QM/MM, X-ray, and AMBER optimized structures, respectively. On the other hand, the aromatic protons are not involved in hydrogen bonding and their chemical environment are rather similar for both bifurcated and Hoogsteen hydrogen bonding configurations, giving consistent agreement with the experimental  $^1\text{H}$  NMR chemical shifts.

Collectively, the reported results demonstrate the capabilities of the MoD-QM/MM method for efficient calculations of ab

initio quality geometries of extended systems and DFT electrostatic potentials through calculations of RESP atomic charges that account for mutual polarization effects. The MoD-QM/MM methodology thus provides a valuable approach for the structural refinement of large biomolecular systems.

## CONCLUSIONS

We have generalized the MoD-QM/MM methodology to provide an ab initio quality computational protocol for structural refinement of extended systems. The method partitions the system into molecular domains that are iteratively optimized as QM layers embedded in their surrounding molecular environment and their RESP atomic charges are iteratively computed until reaching self-consistency in the description of the overall geometry and molecular electrostatic potential. The resulting methodology is benchmarked with model systems that allow for full QM optimization and applied to the structural refinement of the *Oxytricha nova* guanine quadruplex. Calculations of  $^1\text{H}$  NMR chemical shifts based on the gauge independent atomic orbital (GIAO) method and direct comparisons with experiments show that solvated MoD-QM/MM structures, sampled from explicit solvent molecular dynamics simulations, allow for a proper description of stacking and hydrogen bonding giving good agreement of calculated  $^1\text{H}$  NMR chemical shifts and experimental data.

The MoD-QM/MM methodology presents a simple approach for incorporating polarization effects into the calculations based on a distribution of MM atomic charges. The approach is an effective means of obtaining relaxed geometries that are consistent with atomic charges providing ab initio quality electrostatic potentials. The method could thus serve to analyze polarization effects along conformational changes. For example, it would be interesting to examine the effect of MoD-QM/MM charge reparameterization in proteins, DNA, and related biomolecules during classical molecular dynamics simulations. The reparameterized charges may also be used to improve free energy calculations of  $\text{p}K_a$  values and protein–ligand binding energies, as we are currently investigating.

## ASSOCIATED CONTENT

### Supporting Information

Geometrical data for the MoD-QM/MM optimized G-quartet model systems, ESP and RESP atomic charges for a guanine base, as well as the NMR chemical shifts for *Oxytricha nova* G-quadruplex calculated at various levels of theory and corresponding experimental values. This material is available free of charge via the Internet at <http://pubs.acs.org/>.

## AUTHOR INFORMATION

### Corresponding Author

\*Email: victor.batista@yale.edu.

### Present Address

<sup>\$</sup>Department of Chemistry, University of Connecticut, 55 North Eagleville Road, University of Connecticut, Storrs, CT 06269, United States.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

The authors thank the two reviewers of the manuscript for their valuable comments. V.S.B. acknowledges supercomputer time

from the National Energy Research Scientific Computing Center (NERSC) and support from the National Science Foundation (NSF CHE-0911520). Development of the MoD-QM/MM approach was partially funded by the National Institutes of Health (NIH) grant 1R01GM10621-01A1. J.H. acknowledges the Agency for Science, Technology, and Research for support. C.M.R. acknowledges support from the NIH Biophysical Training Grant. J.A.G. acknowledges support from NSF (CHE-0847340). J.P.L. acknowledges support from NSF MCB 1121372. M.B.N. acknowledges support from the LANL CNLS Summer Program and E.R.B. from the LDRD program at Los Alamos National Laboratory (LANL). LANL is operated by Los Alamos National Security, LLC, for the National Nuclear Security Administration of the U.S. Department of Energy under Contract DE-AC5206NA25396

## REFERENCES

- (1) Okamura, M. Y.; Feher, G. *Annu. Rev. Biochem.* **1992**, *61*, 861.
- (2) Burykin, A.; Warshel, A. *FEBS Lett.* **2004**, *570*, 41.
- (3) Pisliakov, A. V.; Sharma, P. K.; Chu, Z. T.; Haranczyk, M.; Warshel, A. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 7726.
- (4) Warshel, A.; Dryga, A. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 3469.
- (5) Bondar, A. N.; Fischer, S.; Smith, J. C.; Elstner, M.; Suhai, S. *J. Am. Chem. Soc.* **2004**, *126*, 14668.
- (6) Kamerlin, S. C. L.; Haranczyk, M.; Warshel, A. *J. Phys. Chem. B* **2009**, *113*, 1253.
- (7) Warshel, A.; Sharma, P. K.; Kato, M.; Xiang, Y.; Liu, H.; Olsson, M. H. M. *Chem. Rev.* **2006**, *106*, 3210.
- (8) Simonson, T.; Archontis, G.; Karplus, M. *Acc. Chem. Res.* **2002**, *35*, 430.
- (9) Gascon, J. A.; Leung, S.; Batista, E. R.; Batista, V. *Abstr. Pap. Am. Chem. Soc.* **2005**, *230*, U1403.
- (10) Gascon, J. A.; Leung, S. F. S.; Batista, E. R.; Batista, V. S. *J. Chem. Theory Comput.* **2006**, *2*, 175.
- (11) Menikarachchi, L. C.; Gason, J. A. *J. Mol. Mod.* **2008**, *14*, 1.
- (12) Horn, A. H. C.; Lin, J.-H.; Clark, T. *Theor. Chem. Acc.* **2005**, *114*, 159.
- (13) Bikadi, Z.; Hazai, E. *J. Cheminf.* **2009**, *1*, 1.
- (14) Tsai, K.-C.; Chen, Y.-C.; Hsiao, N.-W.; Wang, C.-L.; Lin, C.-L.; Lee, Y.-C.; Li, M.; Wang, B. *Eur. J. Med. Chem.* **2010**, *45*, 1544.
- (15) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. *Comput. Mol. Sci.* **2013**, *3*, 198.
- (16) MacKerell, A. D., Jr. In *Computational Biochemistry and Biophysics*; Becker, O. M., MacKerell, A. D., Jr., Roux, B., Watanabe, M., Eds.; Marcel Dekker, Inc.: New York, 2001; p 7 to 38.
- (17) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
- (18) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474.
- (19) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; Di, S.; Robert, A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. *J. Phys. Chem. B* **2010**, *114*, 2549.
- (20) Patel, S.; MacKerell, A. D. J.; Brooks, C. L. I. *J. Comput. Chem.* **2004**, *25*, 1504.
- (21) Patel, S.; Brooks, C. L. I. *J. Comput. Chem.* **2003**, *25*, 1.
- (22) Lamoureux, G.; MacKerell, A. D. J.; Roux, B. *J. Chem. Phys.* **2003**, *119*, 5185.
- (23) Lopes, P. E. M.; Huang, J.; J, S.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D. *J. J. Chem. Theory Comput.* **2013**, *9*, 5430.
- (24) Banks, J. L.; Kaminski, G. A.; Zhou, R.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. *J. Chem. Phys.* **1999**, *110*, 741.
- (25) Harder, E.; Kim, B.; Friesner, R. A.; Berne, B. J. *J. Chem. Theory Comput.* **2005**, *1*, 169.
- (26) Warshel, A.; Kato, S.; Pisliakov, A. V. *J. Chem. Theory Comput.* **2007**, *3*, 2034.

- (27) Yang, W. T.; Lee, T. S. *J. Chem. Phys.* **1995**, *103*, 5674.  
 (28) Eckard, S.; Exner, T. E. *Z. Phys. Chem.* **2006**, *220*, 927.  
 (29) Exner, T. E.; Mezey, P. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 4061.  
 (30) Exner, T. E.; Mezey, P. G. *J. Phys. Chem. A* **2004**, *108*, 4301.  
 (31) Exner, T. E.; Mezey, P. G. *J. Comput. Chem.* **2003**, *24*, 1980.  
 (32) Exner, T. E.; Mezey, P. G. *J. Phys. Chem. A* **2002**, *106*, 11791.  
 (33) Fedorov, D. G.; Kitaura, K. *The Fragment Molecular Orbital Method: Practical Applications to Large Molecular Systems*; Taylor & Francis: Boca Raton, FL, 2009.  
 (34) Fedorov, D. G.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 6904.  
 (35) Fedorov, D. G.; Ishida, T.; Uebayasi, M.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 2722.  
 (36) Babu, K.; Ganesh, V.; Gadre, S. R. *Theor. Chem. Acc.* **2004**, *111*, 255.  
 (37) Babu, K.; Gadre, S. R. *J. Comput. Chem.* **2003**, *24*, 484.  
 (38) Huang, L.; Massa, L.; Karle, J. *Int. J. Quantum Chem.* **2006**, *106*, 447.  
 (39) Huang, L.; Massa, L.; Karle, J. *Int. J. Quantum Chem.* **2005**, *103*, 808.  
 (40) Zhang, D. W.; Chen, X. H.; Zhang, J. Z. H. *J. Comput. Chem.* **2003**, *24*, 1846.  
 (41) Zhang, D. W.; Zhang, J. Z. H. *J. Chem. Phys.* **2003**, *119*, 3599.  
 (42) Li, W.; Li, S.; Jiang, Y. *J. Phys. Chem. A* **2007**, *111*, 2193.  
 (43) Addicoat, M. A.; Collins, M. A. *J. Chem. Phys.* **2009**, *131*, 104103/1.  
 (44) Collins, M. A.; Deev, V. A. *J. Chem. Phys.* **2006**, *125*, 104104/1.  
 (45) Le, H.-A.; Tan, H.-J.; Ouyang, J. F.; Bettens, R. P. A. *J. Chem. Theory Comput.* **2012**, *8*, 469.  
 (46) Le, H.-A.; Lee, A. M.; Bettens, R. P. A. *J. Phys. Chem. A* **2009**, *113*, 10527.  
 (47) Lee, A. M.; Bettens, R. P. A. *J. Phys. Chem. A* **2007**, *111*, 5111.  
 (48) Bettens, R. P. A.; Lee, A. M. *J. Phys. Chem. A* **2006**, *110*, 8777.  
 (49) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293.  
 (50) Gordon, M. S.; Slipchenko, L.; Li, H.; Jensen, J. H. *Annu. Reports in Comput. Chem.* **2007**, *3*, 177.  
 (51) Gordon, M. S.; Smith, Q. A.; Xu, P.; Slipchenko, L. *Annu. Rev. Phys. Chem.* **2013**, *64*, 553.  
 (52) Wang, Y.; Sosa, C.; Cembran, A.; Truhlar, D. G.; Gao, J. *J. Phys. Chem. B* **2012**, *116*, 6781.  
 (53) Xie, W.; Gao, J. *J. Chem. Theory Comput.* **2007**, *3*, 1890.  
 (54) Xie, W.; Orozco, M.; Truhlar, D. G.; Gao, J. *J. Chem. Theory Comput.* **2009**, *5*, 459.  
 (55) Field, M. J. *Mol. Phys.* **1997**, *91*, 835.  
 (56) Gao, J. *J. Chem. Phys.* **1998**, *109*, 2346.  
 (57) Reid, D. M.; Collins, M. A. *J. Chem. Phys.* **2013**, *139*, 184117.  
 (58) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. *Chem. Rev.* **2012**, *112*, 632.  
 (59) Chung, L. W.; Hirao, H.; Li, X.; Morokuma, K. *Comput. Mol. Sci.* **2012**, *2*, 327.  
 (60) Vreven, T.; Byun, K. S.; Komáromi, I.; Dapprich, S.; Montgomery, J. A.; Morokuma, K.; Frisch, M. J. *J. Chem. Theory Comput.* **2006**, *2*, 815.  
 (61) Vreven, T.; Morokuma, K. *J. Comput. Chem.* **2000**, *21*, 1419.  
 (62) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198.  
 (63) van der Kamp, M. W.; Mulholland, A. J. *Biochemistry* **2013**, *52*, 2708.  
 (64) Hratchian, H. P.; Kruckau, A. V.; Parandekar, P. V.; Frisch, M. J.; Raghavachari, K. *J. Chem. Phys.* **2011**, *135*, 014105.  
 (65) Hratchian, H. P.; Parandekar, P. V.; Raghavachari, K.; Frisch, M. J.; Vreven, T. *J. Chem. Phys.* **2008**, *128*, 034107.  
 (66) Howard, J. C.; Tschumper, G. S. *J. Chem. Phys.* **2013**, *139*, 184113.  
 (67) Bates, D. M.; Smith, J. R.; Tschumper, G. S. *J. Chem. Theory Comput.* **2011**, *7*, 2753.  
 (68) Elsohly, A. M.; Shaw, C. L.; Guice, M. E.; Smith, B. D.; Tschumper, G. S. *Mol. Phys.* **2007**, *105*, 2777.  
 (69) Tschumper, G. S. *Chem. Phys. Lett.* **2006**, *427*, 362.  
 (70) Mayhall, N.; Raghavachari, K. *J. Chem. Theory Comput.* **2011**, *7*, 1336.  
 (71) Guo, W.; Wu, A.; Xu, X. *Chem. Phys. Lett.* **2010**, *498*, 203.  
 (72) Sproviero, E. M.; Gascon, J. A.; McEvoy, J. P.; Brudvig, G. W.; Batista, V. S. *J. Chem. Theory Comput.* **2006**, *2*, 1119.  
 (73) Sandberg, D. J.; Rudnitskaya, A. N.; Gascon, J. A. *J. Chem. Theory Comput.* **2012**, *8*, 2817.  
 (74) Schultze, P.; Hud, N. V.; Smith, F. W.; Feigon, J. *Nucleic Acids Res.* **1999**, *27*, 3018.  
 (75) Patel, D. J.; Phan, A. T.; Kuryavyi, V. *Nucleic Acids Res.* **2007**, *35*, 7429.  
 (76) Haider, S.; Parkinson, G. N.; Neidle, S. *J. Mol. Biol.* **2002**, *320*, 189.  
 (77) Gu, J.; Leszczynski, J. *J. Phys. Chem. A* **2000**, *104*, 6308.  
 (78) Gu, J.; Leszczynski, J. *J. Phys. Chem. A* **2002**, *106*, 529.  
 (79) Meyer, M.; Brandl, M.; Suehnel, J. *J. Phys. Chem. A* **2001**, *105*, 8223.  
 (80) Meyer, M.; Steinke, T.; Brandl, M.; Suhnel, J. *J. Comput. Chem.* **2000**, *22*, 109.  
 (81) van Mourik, T.; Dingley, A. *J. Chem.—Eur. J.* **2005**, *11*, 6064.  
 (82) Fonseca Guerra, C.; van der Wijst, T.; Poater, J.; Swart, M.; Bickelhaupt, F. M. *Theor. Chem. Acc.* **2010**, *125*, 245.  
 (83) Fonseca Guerra, C.; Zijlstra, H.; Paragi, G.; Bickelhaupt, F. M. *Chem.—Eur. J.* **2011**, *17*, 12612.  
 (84) Liu, H.; Gauld, J. W. *Phys. Chem. Chem. Phys.* **2009**, *11*, 278.  
 (85) Yurenko, Y. P.; Novotny, J.; Sklenar, V.; Marek, R. *Phys. Chem. Chem. Phys.* **2014**, *16*, 2072.  
 (86) Šponer, J.; Mládek, A.; Špačková, N.; Cang, X.; Cheatham, T. E., III; Grimme, S. *J. Am. Chem. Soc.* **2013**, *135*, 9785.  
 (87) Ross, W. S.; Hardin, C. C. *J. Am. Chem. Soc.* **1994**, *116*, 6070.  
 (88) Strahan, G. D.; Keniry, M. A.; Shafer, R. H. *Biophys. J.* **1998**, *75*, 968.  
 (89) Špačková, N.; Berger, I.; Šponer, J. *J. Am. Chem. Soc.* **1999**, *121*, 5519.  
 (90) Read, M. A.; Neidle, S. *Biochemistry* **2000**, *39*, 13422.  
 (91) Haider, S.; Parkinson, G. N.; Neidle, S. *Biophys. J.* **2008**, *95*, 296.  
 (92) Akhshi, P.; Acton, G.; Wu, G. *J. Phys. Chem. B* **2012**, *116*, 9363.  
 (93) Fadrna', E.; Špačková, N.; Sarzyn'ska, J.; Koča, J.; Orozco, M.; Cheatham, T. E., III; Kulinski, T.; Šponer, J. *J. Chem. Theory Comput.* **2009**, *5*, 2514.  
 (94) Song, J.; Ji, C.; Zhang, J. Z. H. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3846.  
 (95) Gkionis, K.; Kruse, H.; Platts, J. A.; Mládek, A.; Koča, J.; Šponer, J. *J. Chem. Theory Comput.* **2014**, *10*, 1326.  
 (96) Šponer, J.; Mláek, A.; Šponer, J. E.; Svozil, D.; Zgarbova, M.; Banas, P.; Jurecka, P.; Otyepka, M. *Phys. Chem. Chem. Phys.* **2012**, *14*, 15257.  
 (97) Haider, S.; Parkinson, G. N.; Neidle, S. *J. Mol. Biol.* **2002**, *320*, 189.  
 (98) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian09*, Revision D.01; Gaussian, Inc.: Pittsburgh, PA, 2009.  
 (99) Maseras, M.; Morokuma, K. *J. Comput. Chem.* **1995**, *16*, 1170.

- (100) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Guld, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
- (101) Beran, G. J.; Nanda, K. *J. Phys. Chem. Lett.* **2010**, *1*, 3480.
- (102) Wang, L. T.; Chen, C. C. *IEEE Trans. Robotics Automation* **1991**, *7*, 489.
- (103) Tseng, P. *J. Opt. Theory Appl.* **2001**, *109*, 475.
- (104) Canutescu, A. A.; Dunbrack, R. L. *J. Protein Sci.* **2003**, *12*, 963.
- (105) Vreven, T.; Morokuma, K.; Farkas, O.; Schlegel, H. B.; Frisch, M. J. *J. Comput. Chem.* **2003**, *24*, 760.
- (106) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1984**, *5*, 129.
- (107) Besler, B. H.; Merz, J. K. M. *J. Comput. Chem.* **1990**, *11*, 431.
- (108) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.
- (109) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620.
- (110) Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.
- (111) Ponder, J. W.; Case, D. A. *Adv. Prot. Chem.* **2003**, *66*, 27.
- (112) Ditchfield, R. *Mol. Phys.* **1974**, *27*, 789.
- (113) Wolinski, K.; Hilton, J. F.; Pulay, P. *J. Am. Chem. Soc.* **1990**, *112*, 8251.
- (114) Available for download from <http://gascon.chem.uconn.edu/software>.
- (115) Case, D. A.; Darden, T. A.; Cheatham, T.E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER12*; University of California: San Francisco, 2012.
- (116) Joung, I. S.; Cheatham, T. E., III *J. Phys. Chem. B* **2008**, *112*, 9020.
- (117) Philips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781.
- (118) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33.
- (119) Thanthiriwatte, K. S.; Hohenstein, E. G.; Burns, L. A.; Sherrill, C. D. *J. Chem. Theory Comput.* **2011**, *7*, 88.
- (120) Minenkov, Y.; Singstad, A.; Occhipinti, G.; Jensen, V. R. *Dalton Trans.* **2012**, *14*, 5526.
- (121) Gill, M. L.; Strobel, S. A.; Loria, J. P. *Nucleic Acids Res.* **2006**, *34*, 4506.
- (122) Gill, M. L.; Strobel, S. A.; Loria, J. P. *J. Am. Chem. Soc.* **2005**, *127*, 16723.
- (123) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. *Chem. Rev.* **2012**, *112*, 1839.
- (124) Zhu, T.; Zhang, J. Z. H.; He, X. *J. Chem. Theory Comput.* **2013**, *9*, 2104.
- (125) Exner, T. E.; Frank, A.; Onila, I.; Möller, H. M. *J. Chem. Theory Comput.* **2012**, *8*, 4818.
- (126) Dracinsky, M.; Möller, H. M.; Exner, T. E. *J. Chem. Theory Comput.* **2013**, *9*, 3806.
- (127) Frank, A.; Möller, H. M.; Exner, T. E. *J. Chem. Theory Comput.* **2012**, *8*, 1480.