JCTC Journal of Chemical Theory and Computation

Optimization of Capping Potentials for Spectroscopic Parameters in Hybrid Quantum Mechanical/Mechanical Modeling Calculations

Sittipong Komin† and Daniel Sebastiani*

Max-Planck-Institute for Polymer Research, Ackermannweg 10, 55128 Mainz, Germany

Received December 2, 2008

Abstract: We present a capping scheme for hybrid calculations which is designed for a systematic optimization to reproduce the molecular structure, frontier bond potential, and spectroscopic properties for the quantum subsystem. Our technique is capable of reducing the perturbations of the electronic structure which are normally caused by conventional link atoms between quantum and classical regions. Specifically, we propose analytic effective core potentials with a small set of adjustable parameters, which are optimized to reproduce the full-quantum-mechanical (full-QM) properties in the direct environment of the bond cleavage. The capping potentials are conceptually simple and easy to employ in most instances without significant code modifications. They do not require any further external geometry constraints and yield also reasonable results for the potential energy surface. We benchmark these potentials for a series of chemically and biologically relevant molecules calculating NMR chemical shifts, protonation energies, and optimized geometries. Our optimized QM/mechanical modeling (MM) potentials are another step toward a realistic first-principles prediction of spectroscopic parameters in complex chemical environments using hybrid QM/MM calculations.

1. Introduction

The determination of the detailed microscopic structure and dynamics of complex supramolecular systems is still a challenge for modern physics and chemistry. The interplay of intramolecular and intermolecular interactions is crucial for a broad range of chemical, biological, and physical processes that occur in nature.[1–4] To obtain structural data of supramolecular systems, the combination of spectroscopic experiments with advanced theoretical predictions and computer simulations is becoming increasingly popular, because this combination often yields a predictive power above the sum of the individual techniques.[5–7]

With the recent advances in computational methodology as well as computer hardware, the first-principles prediction of such noncovalent effects on the structure and experimentally observable spectra has come into reach for many systems of technological and fundamental scientific interest.[8–10] Several methods exist to incorporate the influence of the chemical environment into such electronic structure calculations. The explicit consideration of a large number of neighboring molecules is in principle most accurate, but computationally very demanding and thus only applicable in simple cases.[11–14]

Alternatively, one can resort to embedding schemes, which can treat the environment at various levels of approximation. In this context, a hybrid method is often adopted which splits the total system into a smaller part, which is treated quantum-mechanically (QM) using electronic structure methods, and the remaining part, which is described via parametrized potentials (MM).[15–17] One of the difficulties of such a hybrid quantum mechanical/mechanical modeling (QM/MM) approach is the transition region between the two different parts. Often, chemical bonds are "broken"; i.e., one of the atoms involved in the covalent bond is in the quantum (QM) part, the other in the classical (MM) one. This situation is sketched

* Corresponding author phone: 49-6131-379-260; fax: 49-6131-379-100; e-mail: sebastia@mpip-mainz.mpg.de.

† Department of Physics, Faculty of Science, Ubonratchathanee University, Ubonratchathanee, Thailand.

10.1021/ct800525u CCC: $40.75
Published on Web 05/04/2009

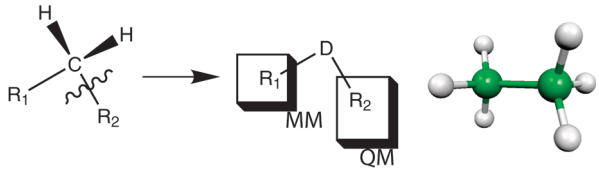

Figure 1. General principle of the repartitioning scheme for a QM/MM calculation in which a chemical bond (here C−R_2) crosses the QM/MM border and is hence cleaved. The link atom that saturates the resulting dangling bond (•R_2) is denoted D. Also shown is the ethane molecule, which serves as the reference molecule for optimizing the parameters of the pseudopotential by which the dummy atom is implemented.

in Figure 1. Similar problems arise when MM atoms are located near a QM region, because the QM and MM descriptions are not genuinely compatible. Thus, a suitable interface has to be used, which can mutually couple the two schemes in a realistic way.

In this work, we address the perturbing effect of a bond cleavage that occurs if a part of a given molecule is treated using quantum mechanics and another part is modeled classically (Figure 1). There are several approaches for tackling such a situation, where most commonly a carbon−carbon bond is cut. In the following, we will refer to the resulting pseudoatom synonymously as the "dummy atom" or capping potential (the "D" atom in Figure 1). There are many QM/MM implementations already available in quantum chemistry software packages; many groups have further developed specific improvements to the QM/MM idea.[15,16,18–25] In particular, there are several approaches to tackle the bond saturation problem arising from a bond cleavage by the QM/MM repartitioning as mentioned above. Among them are the following.

(a) Hydrogen capping: The dummy atom in Figure 1 is represented by a regular hydrogen atom.[26] This relatively straightforward solution has known disadvantages, but it is nevertheless used very often. Obviously, the C−H bond length is shorter than the original C−C bond, and the vibrational frequencies are different. The smaller electronegativity of the hydrogen furthermore changes the electronic structure of the quantum subsystem in the vicinity of the border region considerably. This perturbation can reach over several C−C bonds in the QM subsystem.

(b) Fluorine capping: The saturation of the dangling bond is done via a seven-valent termination atom, for instance, a fluorine, instead of a hydrogen. While this solution, which was originally developed as a pseudobond approach,[23] provides a better bond distance agreement ($d_{C-F} \approx d_{C-C}$), the electronegativity of fluorine is significantly higher than that of carbon. Thus, the electronic subsystem can be perturbed somewhat stronger compared to that with hydrogen capping.

(c) Frozen orbitals: An alternative method relies on precomputed atomic orbitals that are placed at the link atom to ensure an adequate electrostatic interaction and an accurate orthogonality of the terminal chemical bond of the QM subsystem.[27] This frozen-orbital scheme has also been employed for the calculation of NMR shielding constants.[28] A related approach has been developed by the Truhlar group, where auxiliary hybrid orbitals are used to provide an optimal directionality of the termination of the last QM bond.[29,30] While this class of approaches is one of the more accurate ones, it involves a higher coding effort for the incorporation of the frozen orbitals, even though they are excluded from the actual SCF optimization.

(d) Effective fragment potential: Originally designed as a discrete solvation approach to treat chemical reactions in solution,[31] it has been extended to study covalently bound clusters and bulk properties.[32–35] In this method, the total system is divided into a QM region and the environment (the fragment) which interacts with the QM region via a set of one-electron potentials. All important physical interactions between the two fragments (which can be either covalently or noncovalently bonded) are considered explicitly, in particular electrostatic interactions, charge penetration, and polarization effects. Also the effect of exchange repulsion can be incorporated into the scheme. While this effective fragment potential provides a highly accurate description of the original quantum-mechanical interactions, it is not designed to be transferable between different types of fragments. Furthermore, it requires a considerable additional effort for both the design and implementation of the fragment potentials, and it increases the computational effort at runtime compared to conventional QM/MM approaches based on empirical force fields for the MM part. A method that has similar characteristics is known as the effective group potential method,[36,37] but has been used less frequently than the original effective fragment potential approach.

(e) Field-adapted adjustable density matrix assembler (FA-ADMA): A related technique exists in which the target macromolecule is divided into fragments for which conventional quantum chemical calculations are performed.[38–40] Both the fragment and its local environment up to a certain distance are included in these calculations, and the rest of the macromolecule is incorporated via point charges. This approach is hence a regular QM/MM method, with the difference that the QM region is made somewhat larger than really necessary to remove the problems related to the QM/MM boundary region.

(f) Quantum capping potentials: The saturation of dangling bonds with effective potentials has already been attempted by DiLabio et al.[41–45] in an approach that is similar to the one proposed in this work. A conventional pseudopotential is used to truncate the quantum region, using a local part and nonlocal angular-momentum-dependent projectors. These effective capping potentials, however, are not specifically tuned to reproduce the full-QM spectroscopic properties in the QM/MM calculations. Instead, they are built in analogy to the generation of regular atomic pseudopotentials, focusing on the capping atom's orbitals and their energy levels.

In this work, we go one step beyond the QM/MM capping approaches presented above, by using specially designed capping potentials. We present the results of an optimization scheme designed to improve such special potentials within a density functional theory based approach. Specifically, our work is based on analytical effective core potentials (pseudopotentials) of the Goedecker type,[46,47] in line with previous

QM/MM studies.[10,16,48] Our goal is to optimize the pseudopotential parameters in such a way that the change of the electronic density in the quantum part of a QM/MM calculation is minimal with respect to a "full-QM" calculation. In this way, we also ensure that structural parameters and spectroscopic properties in the direct neighborhood of a QM/MM bond cleavage are modeled with a high degree of reliability.

To achieve this aim, we define a penalty functional that quantifies the deviation of the electronic density in a molecular fragment from the corresponding density in the complete molecule, while simultaneously penalizing changes in the equilibrium bond distance and frequency. The penalty functional is minimized iteratively by varying the coefficients of the capping potential placed at the bond cleavage site. This approach is similar to the recently developed heptavalent potential,[49] where we variationally optimized effective atom-centered potentials to describe the methyl group in acetic acid. However, we found that this potential does not always optimally reproduce the spectroscopic parameters of the full molecule. Our capping potentials can be used as link atoms replacing a carbon and involve no further external geometry constraints. They also give reasonable results for potential energy surfaces of the C−C bond. We characterize the perturbative effect of the bond cleavage by means of NMR chemical shifts, which are known to be particularly sensitive to both the intramolecular electronic structure and intermolecular effects such as hydrogen bonding.[50–54] Hence, we can not only gauge the direct perturbing effect of the cleaved bond on the electronic structure of the remaining part of a molecule, but also quantitatively describe how strongly its response properties are tainted by the QM/MM bond cleavage.

2. Methods and Computational Details

2.1. Goal of the Optimization. The purpose of dummy atoms in QM/MM calculations is to enable a saturation of the last covalent bond of the quantum region, i.e., the bond which is cleaved by the QM/MM repartitioning. The central difficulty regarding the quantum region is that the true character of the bond cannot be reproduced easily by a simple terminal atom. Especially spectroscopic parameters react very sensitively to small deviations in the electronic structure around the cleaved bond.

The aim of our optimization scheme is to provide a tool which allows tuning of the properties of the terminal dummy atom in such a way as to make the electronic density in the QM part of the molecule (ρ_D) as similar as possible to the reference electron density (ρ_{QM}), i.e., the density when the entire molecule is treated quantum mechanically. This will eventually lead to an improvement in the spectroscopic properties of the system in the QM/MM description. We further aim at preserving the C−C equilibrium bond length in the dummy calculation to allow an easy coupling of the "first" classical MM atom and to avoid the need for additional geometric constraints.

To this aim, we define a penalty functional which expresses the deviation of these properties from their target values obtained in a full-QM calculation via

$$\mathscr{P} = \int_{\Omega} \mathrm{d}^3 r\, [\rho_{\mathrm{QM}}(\mathbf{r}) - \rho_{\mathscr{D}}(\mathbf{r})]^2 + \sum_{J}^{N_{\mathrm{geom}}} \left\{ w_F \sum_{I}^{N_{\mathrm{ions}}} [F_I^{\mathrm{QM}}(R_J) - F_I^{\mathscr{D}}(R_J)]^2 + w_E[\Delta E^{\mathrm{QM}}(R_J) - \Delta E^{\mathscr{D}}(R_J)]^2 \right\} \quad (1)$$

The integration volume Ω is used to restrict the penalization region to areas in which an improvement is physically meaningful. In our case, this volume corresponds to the union of spheres of 1 Å radius around all QM atoms except the carbon which immediately follows the dummy atom. This definition ensures that the covalent dummy−carbon bond is *not* included in the penalty integration volume, while all other bonds of the first carbon are fully incorporated. w_F and w_E are weighting factors to ensure that an adequate relative importance is given to deviations of the electronic density and the forces and total energy, respectively. Several different molecular geometries (here $N_{\mathrm{geom}} = 3$) are incorporated into the force and energy terms of eq 1 to ensure that not only the equilibrium configuration of the molecule is taken into account. Typically, these conformations will correspond to variations of the bond length of the capping potential.

2.2. Functional Form of the Capping Potential. The capping potentials are represented in the form of analytical effective core potentials of the Goedecker type,[47,49] consisting of a local and a nonlocal part. For a carbon atom, the local potential reads

$$V_{\mathrm{loc}}(\mathbf{r}) = \frac{-Z_{\mathrm{ion}}}{|\mathbf{r}|} \mathrm{erf}[\rho] + \mathrm{e}^{-\rho^2}(C_1 + C_2\rho^2) \quad (2)$$

with the reduced radius $\rho = |\mathbf{r}|/2^{1/2}r_{\mathrm{loc}}$ and the valence charge Z_{ion}, which would be $Z_{\mathrm{ion}} = 4e$ for a regular carbon pseudopotential. The local radius r_{loc} characterizes both the Gaussian smearing of the nuclear charge density resulting in the error function and the decay of the local potential in eq 2. The nonlocal part of the carbon capping potential consists of one s-type and one p-type projector:

$$V_{\mathrm{nl}}(\mathbf{r},\mathbf{r}') = h_{\mathrm{s}} \frac{1}{2\pi^2 r_{\mathrm{s}}^3} \exp\left(-\frac{\mathbf{r}^2 + \mathbf{r}'^2}{2r_{\mathrm{s}}^2}\right) + h_{\mathrm{p}} \frac{32}{225\pi} \frac{\mathbf{r}\mathbf{r}'}{r_{\mathrm{p}}^5} \exp\left(-\frac{\mathbf{r}^2 + \mathbf{r}'^2}{2r_{\mathrm{p}}^2}\right) \sum_{m=0,\pm1} \bar{Y}_1^m(\hat{\mathbf{r}}) Y_1^m(\hat{\mathbf{r}}') \quad (3)$$

with additional characteristic radii r_{s} and r_{p} and the amplitudes h_{s} and h_{p} of one s-type and one p-type projector, respectively. The starting point for the optimization of the capping potential parameters (C_1, C_2, r_{loc}, r_{s}, h_{s}) was the regular carbon pseudopotential with an adjusted valence charge ($Z_{\mathrm{ion}} = 1$).

2.3. Optimization Scheme. Common effective core potentials are often generated by means of a direct inversion of the electronic Schrödinger equation for an isolated atom, with the help of its all-electron orbitals.[55] An alternative approach

consists in iteratively minimizing a penalty functional that expresses the deviations of the pseudo wave function from its all-electron counterpart; this method is commonly used for analytic potentials of the Goedecker type.[46,47]

In analogy to this concept, we optimize our potentials by an iterative Nelder−Mead downhill simplex minimization[56] of the penalty function in eq 1. All seven parameters of the analytic expression in eqs 2 and 3 are varied until the penalty functional becomes stationary. While the derivative of the force and energy terms of the penalty functional with respect to the capping parameters is done via a three-point finite difference, the derivative of the density deviation is done analytically via perturbation theory. On the example of the radius of the s-channel of the potential, this can be achieved according to

$$\frac{\mathrm{d}\mathscr{P}}{\mathrm{d}r_s} = 2\int_\Omega \mathrm{d}^3r\,[\rho_{QM}(\mathbf{r}) - \rho_{\mathscr{D}}(\mathbf{r})]\,\frac{\mathrm{d}\rho_{\mathscr{D}}(\mathbf{r})}{\mathrm{d}r_s} + \ldots \quad (4)$$

in which the term $\mathrm{d}\rho_{\mathscr{D}}(\mathbf{r})/\mathrm{d}r_s$ is computed as the first-order density response of the system with respect to the "perturbation" that is induced by varying the s-channel radius r_s in the capping potential. In this context, $\mathscr{H}^{(1)} = \mathrm{d}V_{nl}/\mathrm{d}r_s$ represents a perturbation Hamiltonian, as would be an external electric or magnetic field in the case of an external perturbation.[57–59]

2.4. Computational Details. Our calculations are done within density functional theory[60–62] using the BLYP[63,64] exchange-correlation functional, as implemented in the CPMD package.[65,66] We use standard norm-conserving pseudopotentials[46,47] and a 70 Ry energy cutoff for the plane-wave expansion of the Kohn−Sham orbitals. To simplify the problem of the bond cleavage and to eliminate the corresponding degrees of freedom, we have not assigned any point charges to the atoms in the classical fragments.

The calculation of magnetic resonance properties (NMR chemical shifts) are done within density functional perturbation theory as implemented in the CPMD package.[57,59,67] Following the experimental convention, we quote chemical shifts relative to computed nuclear shieldings of standard reference systems tetramethylsilane and nitromethane for ^{13}C, ^{1}H, and ^{15}N according to eq 5; all sp^2-hybridized carbons are actually referenced indirectly to TMS via the experimental shift and the computed shieldings of benzene ($\delta^{exptl}_{(C_6H_6)} = 128.4$ ppm[68]) according to eq 6. Chemical shift anisotropies were computed as $\Delta\sigma = \sigma_{33} - {}^1/_2(\sigma_{11} + \sigma_{22})$ using the convention $\sigma_{11} < \sigma_{22} < \sigma_{33}$ for the principal values σ_{ii} of the nuclear shielding tensor.

$$\delta^{calcd}_{(X)} = \frac{1}{3}\,\mathrm{Tr}[\sigma^{calcd}_{(TMS/NMe)} - \sigma^{calcd}_{(X)}] \quad (5)$$

$$\delta^{calcd}_{(X)} = \delta^{exptl}_{(C_6H_6)} + \frac{1}{3}\,\mathrm{Tr}[\sigma^{calcd}_{(C_6H_6)} - \sigma^{calcd}_{(X)}] \quad (6)$$

3. Results and Discussion

3.1. Capping Potentials in the Reference Molecule. The ethane molecule serves as our reference molecule for the optimization of the pseudopotential parameters of the carbon dummy atom. As C−C bonds are the most common bond type within biomolecules, a controllable way of cutting is highly desirable. For the optimization process, one of the methyl groups is replaced by a capping pseudopotential, whose parameters are varied until the penalty functional in eq 1 becomes stationary. We used three different geometries corresponding to a stretching and shortening of the C−$\mathscr{D}$ bond by ±0.24 Å in the penalty functional. Together with relative weights of $w_F = w_E = 1$, this setup turned out to yield a good compromise between geometric and electronic properties. Figure 2 shows the evolution of the parameters r_{loc}, r_s, and r_p during the progress of the optimization.

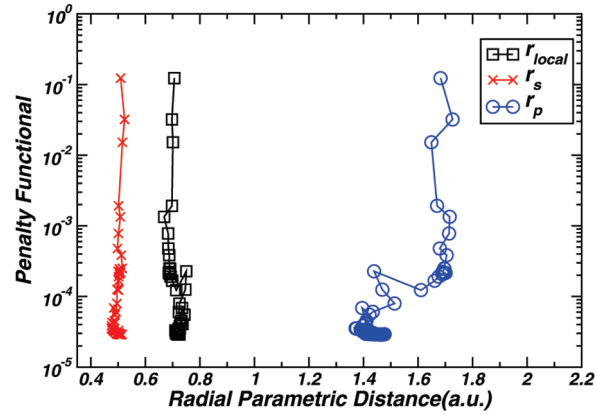

Figure 2. History of the penalty function depending on r_{loc}, r_s, and r_p.

Table 1. Pseudopotenial Parameters of the Regular Carbon Atom and the Optimized Dummy Capping Potential ($\mathscr{D}_C$)[a]

	r_{loc}	C_1	C_2	r_s	h_s	r_p	h_p
regular C	0.3376	–9.1285	1.4251	0.3025	9.6507		
$\mathscr{D}_C$	0.7221	9.9068	–2.5466	0.5120	–3.5081	1.4664	0.2316

[a] Length parameters (r_{loc}, r_s, r_p) are given in units of bohr and energy coefficients (C_1, C_2, h_s, h_p) in hartrees.

The final results for the optimized parameters of the capping potential are shown in Table 1. Note that also it turns out that these values have strongly changed compared with the original carbon pseudopotential from which the optimization was started. Note that the valence charge has been switched from $Z_v = 4$ to $Z_v = 1$. The increase of r_{loc} corresponds to a considerably broader Gaussian smearing of the nuclear charge density, reaching far into the covalent bonding region. The positive C_1 and the negative C_2 coefficients have the effect of further pushing the bonding electron away from the dummy position and attracting it to the middle of the C−$\mathscr{D}$ bond; together, these changes can be seen as a considerably reduced electronegativity of the dummy. The s-channel projector in turn has become attractive, which somewhat compensates the repulsive effect of the local potential (C_1 and C_2). Note that the original carbon pseudopotential had no projector in the p-channel.

3.2. Improvement of Electronic Densities with $\mathscr{D}_{opti}$. In Figure 3, the improvements obtained due to the optimization process for ethane are illustrated in terms of electron density differences. We have compared the density in the full molecule to the density of the dummy-substituted one, using the initial values for the pseudopotential ($\mathscr{D}_{ini}$) and the optimized capping parameters ($\mathscr{D}_{opti}$). We recall that our initial values are the pseudopotential parameters for a regular

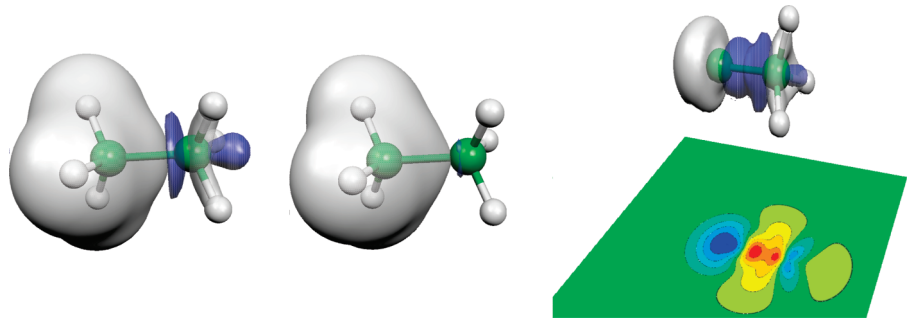

Figure 3. Electron density difference maps for two dummy-substituted ethane molecules. Left: $\rho^{\mathscr{D}_{ini}-CH_3} - \rho^{H_3C-CH_3}$; center: $\rho^{\mathscr{D}_{opti}-CH_3} - \rho^{H_3C-CH_3}$; right: $\rho^{\mathscr{D}_{opti}-CH_3} - \rho^{\mathscr{D}_{ini}-CH_3}$.

Table 2. Calculated ^{1}H and ^{13}C NMR Chemical Shifts δ and Anisotropies $\Delta\sigma$ (ppm) of Ethane before and after the Substitution of the Methyl Group by Dummy Atoms[a]

	(full-QM) R → CH_3	R → $\mathscr{D}_{ini}$	R → $\mathscr{D}_{7v}$	R → $\mathscr{D}_{opti}$	R → H	exptl[69]
δ^H	1.13	–0.72	3.34	0.94	2.21	0.86
$\Delta\sigma^H$	8.6	11.4	13.1	7.9	7.1	
δ^C	10.97	–22.83	–0.31	11.68	28.88	7.00
$\Delta\sigma^C$	17.7	9.4	10.6	23.61	49.0	

[a] In addition to our initial and optimized capping potentials ($\mathscr{D}_{ini}$ and $\mathscr{D}_{opti}$), the heptavalent potential developed by Lilienfeld et al.[49] ($\mathscr{D}_{7v}$) and a simple hydrogen atom were used.

carbon atom (except for the valence charge, which is reduced to 1). Finally, a direct comparison of $\mathscr{D}-CH_3$ between the initial and optimized dummy link atom $\mathscr{D}$ is shown in Figure 3, along with its projection in two dimensions.

When the unoptimized dummy $\mathscr{D}_i$ is used, the deviations of the electronic density with respect to the corresponding full-QM calculation reach somewhat beyond the cleavage bond ($\mathscr{D}-C$), with regions of both increased and decreased electron density (blue and white clouds in Figure 3). The optimized dummy yields somewhat lower density differences with respect to the unperturbed molecule beyond the first regular carbon atom. When comparing the density differences between the initial and optimized dummy atoms directly (rightmost plot in Figure 3), the strongest effect is located at the dummy itself. Nevertheless, also at the methyl protons, the density redistribution is still considerable.

3.3. NMR Chemical Shifts with the Optimized Capping Potential. We have benchmarked the accuracy of the optimized dummy atoms by calculating NMR chemical shifts, which represent the electronic response to an external magnetic field. These NMR parameters offer a unique reduction of the complex electronic structure in the vicinity of a nucleus into a single number, and they are highly sensitive to small changes in the electronic orbitals. In this way, they offer a local orbital-based probe, complementary to the penalty functional itself that is based only on the total density and geometric quantities. The isotropic NMR chemical shifts of the dummy-substituted molecule are compared to those of the reference molecule in Table 2. Both the optimized ($\mathscr{D}_{opti}$) and the unoptimized ($\mathscr{D}_{ini}$) monovalent dummy potentials are used, as well as the seven-valent one ($\mathscr{D}_{7v}$) by Lilienfeld et al.,[49] which was developed to reproduce the electronic density in acetic acid. For the sake of completeness, we have also added the results for a simple hydrogen capping. To exclude the effects of conformational changes on the NMR chemical shifts, we have always used the optimized geometries of the full molecule.

Table 3. Optimized Bond Lengths (Å) and Vibrational Frequencies of the C−$\mathscr{D}$ Bond for the Ethane Reference Molecule before and after Substitution of the Methyl Group

	(full-QM) R → CH_3	R → $\mathscr{D}_{ini}$	R → $\mathscr{D}_{opti}$	R → $\mathscr{D}_{7v}$	R → H
d(R−C) (Å)	1.54	1.68	1.54	2.04	1.10
d(C−H) (Å)	1.10	1.10	1.10	1.09	1.10
θ_{H-C-R} (deg)	111.3	110.3	112.8	106.8	109.4
θ_{H-C-H} (deg)	107.6	108.6	105.9	112.0	109.4
$\tilde{\nu}_{C-\mathscr{D}}$ (cm^{-1})	909	483	809	682	1103

It turns out (see Table 2) that the chemical shifts from our optimized capping atom are generally in better agreement with the all-QM calculation than for the initial ($\mathscr{D}_{ini}$) and seven-valent ($\mathscr{D}_{7v}$) capping potentials. The initial potential results in significantly lower chemical shifts, while the heptavalent dummy overestimates the ^{1}H shifts and underestimates the ^{13}C shifts. Hydrogen capping in turn always results in too positive chemical shifts. Only the optimized monovalent substitution yields values close to the all-quantum calculation for both nuclei. The deviations between the reference and $\mathscr{D}_{opti}$ are below 0.3 ppm for protons and 1 ppm for carbon, which are errors that can certainly be tolerated for nuclei that are only one or two bonds away from the capping atom. This is not the case for all other capping variants, which exhibit deviations in the NMR chemical shifts of more than 1 ppm (^{1}H) and 20−30 ppm (^{13}C). In conclusion, the NMR resonances of our dummy-substituted ethane show indeed a very good agreement with the corresponding full-quantum calculations, even for the atom directly connected to the bond cleavage.

3.4. Energetic and Geometric Properties of the $\mathscr{D}$−C Bonds. We have optimized the geometry of the reference ethane with the methyl group substituted by the optimized dummy atom. The results for selected distances and angles are shown in Table 3, and Figure 4 shows the potential energy profile of ethane as a function of the $\mathscr{D}$−C

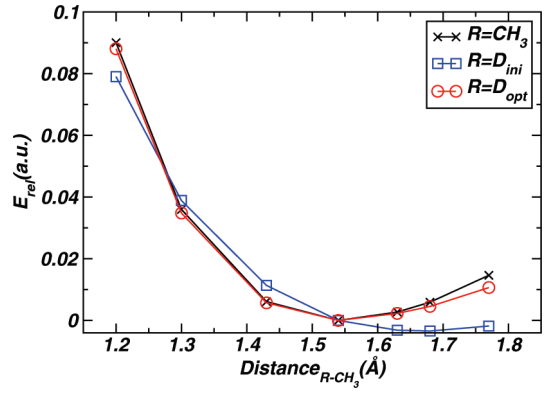

Figure 4. Potential energy curve for rigid stretching/compression of the C−C bond in ethane and $\mathscr{D}$−CH_3 with the original and the optimized pseudopotentials.

bond length. The equilibrium bond length of $\mathscr{D}_{ini}$−CH_3 is somewhat longer than the full QM value. The optimized capping potential $\mathscr{D}_{opti}$, in turn, improved this distance considerably. Also compared to the heptavalent dummy ($\mathscr{D}_{7v}$), we find a clear improvement of the equilibrium bond distances, which is most likely due to the incorporation of the atomic forces into our new penalty functional, eq 1.

Last but now least, the harmonic stretch frequency of the $\mathscr{D}_{opti}$−C bond is almost perfectly reproduced, while the $\mathscr{D}_{ini}$−C and $\mathscr{D}_{7v}$−C variants show deviations. Not surprisingly, the harmonic frequency of the hydrogen capping cannot accurately reproduce the C−C frequency either (although the carbon mass was used instead of the hydrogen one in the calculation of the dynamical matrix).

The full potential energy curve for the $\mathscr{D}$−C bond is shown in Figure 4. The agreement of the optimized dummy potential (red) with the full-QM dissociation curve (black) is remarkable, while the initial capping potential shows an underestimated dissociation energy and a somewhat extended equilibrium bond length. For comparison, we also optimized the capping potential using somewhat lower weights for the geometric penalty contributions (w_F and w_E, data not shown). This calculation resulted in a capping potential that was numerically more similar to the original one (the regular C potential; see Table 1), at the expense of a considerably worse C−$\mathscr{D}$ bond distance and vibrational frequency.

3.5. Histidine. The good agreement obtained in the previous section might have been fortuitous, as the dummy potentials were optimized for the very specific molecule that was subsequently benchmarked there. Thus, we have checked the transferability of our dummy potentials by applying them to a different molecule, namely, histidine. Two selected bond lengths are listed in Table 4, comparing the initial and final capping potentials to the full-QM results. In analogy to the situation encountered for the ethane molecule, the initial $\mathscr{D}_{ini}$ capping yields a stretched $\mathscr{D}$−C_2 bond length, while C_2−C_3 is slightly shortened. Both deficiencies are considerably improved upon by $\mathscr{D}_{opti}$.

Table 5 shows the NMR chemical shifts of the full histidine molecule and its imidazole fragment within a QM/MM description, always using the optimized geometry of the full histidine molecule. As expected, the strongest deviations are observed for carbon C_2 directly involved in the bond cleavage. Here, the capping optimization scheme results in the reduction of the error by almost an order of magnitude compared to the unoptimized capping atom. Similarly, a considerable improvement is obtained for the next carbons C_3 and C_6, as well as nitrogen N_7. In all these cases, the optimized monovalent potential also performs better than the heptavalent dummy atom. For C_9 and N_8, which are further away from the bond cleavage, the situation

Table 4. Geometric Data as Well as Computed and Experimental Proton Affinities $\Delta E = E_{DFT}(X) - E_{DFT}(X{-}H^+)$ for Histidine and Lysine as Well as their Dummy-Substituted Fragments[a]

	full-QM	R → $\mathscr{D}_{ini}$	R → $\mathscr{D}_{opti}$	R → H
		Histidine		
$d_{\mathscr{D}-C_2}$ (Å)	1.57	1.69	1.55	1.1
$d_{C_2-C_3}$ (Å)	1.51	1.49	1.51	1.50
ΔE (kcal/mol)	238.1	245.8	239.4	235.8
		Lysine		
$d_{\mathscr{D}-C_1}$ (Å)	1.56	1.69	1.55	1.1
$d_{C_1-C_2}$ (Å)	1.55	1.55	1.55	1.55
ΔE (kcal/mol)	225.6	232.8	228.3	76.96

[a] For the atom numbering, see Figure 5. R represents the classical part of the molecule, i.e., the amino and carboxylic acid groups.

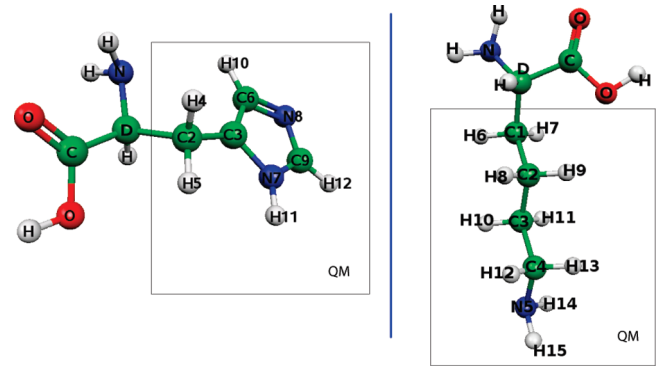

Figure 5. Atom numbering and bond cutting scheme for the histidine (left) and lysine (right) molecules. The molecular fragments outside the "QM" boxes are substituted by means of the capping potential (the dummy atom).

Table 5. Calculated ^1H, ^{13}C, and ^{15}N NMR Chemical Shifts (ppm) of Histidine and Its Dummy-Substituted Fragment[a]

	full-QM	R → $\mathscr{D}_{opti}$	R → $\mathscr{D}_{ini}$	R → $\mathscr{D}_{7v}$	R → H	exptl[69]
C_2	32.46	27.03	−2.72	11.11	43.03	30.78
$\Delta\sigma(C_2)$	25.72	27.05	41.46			
C_3	136.90	140.90	142.05	141.54	137.35	134.67
$\Delta\sigma(C_3)$	80.40	84.40	84.04			
C_6	137.37	133.65	132.52	135.03	138.45	119.55
C_9	138.05	136.37	135.96	137.37	137.85	138.97
N_7	−202.97	−202.74	−200.70	−204.79	−202.73	-
N_8	−138.91	−139.26	−139.60	−139.60	−137.12	-
H_4	3.40	3.66	2.62	5.99	5.01	3.12
$\Delta\sigma(H_4)$	6.39	7.84	10.09			
H_5	3.56	3.14	2.03	5.26	4.46	3.23
$\Delta\sigma(H_5)$	6.53	6.73	8.63			
H_{10}	7.60	6.86	6.56	7.28	7.06	7.06
H_{11}	9.93	9.82	9.66	9.93	9.84	
H_{12}	7.33	7.17	7.09	7.20	7.25	7.80

[a] For the atom numbering, see Figure 5. Data for both the initial and optimized dummy potentials ($\mathscr{D}_{ini}$ and $\mathscr{D}_{opti}$) as well as the heptavalent one ($\mathscr{D}_{7v}$)[49] are shown.

Table 6. Calculated ^{1}H, ^{13}C, and ^{15}N NMR Chemical Shifts (ppm) of Lysine and Its Dummy-Substituted Fragment[a]

	full-QM	R → $\mathscr{D}_{ini}$	R → $\mathscr{D}_{opti}$	R → H	exptl[69]
C_1	37.54	7.33	37.27	53.11	32.60
$\Delta\sigma(C_1)$	35.69	40.08	38.11		
C_2	28.81	30.88	30.79	28.66	24.13
C_3	35.27	43.81	36.98	35.15	29.11
C_4	48.10	48.27	47.81	47.58	41.75
N_5	–293.21	–292.82	–292.95	–293.25	
$H_{6,7}$	2.25	1.13	2.24	3.56	1.9
$\Delta\sigma(H_{6,7})$	4.17	7.88	6.04		
$H_{8,9}$	2.25	2.13	2.24	2.36	1.5
$H_{10,11}$	2.10	1.99	1.91	1.93	1.7
$H_{12,13}$	3.72	3.59	3.65	3.69	3.0
H_{14}	1.95	1.88	1.90	1.93	
H_{15}	2.78	2.70	2.74	2.77	

[a] For the atom numbering, see Figure 5. R represents the remaining fragments in the classical regions.

is less drastic, and all three choices yield similar—yet small—discrepancies with respect to the full-QM calculation.

The hydrogens exhibit smaller absolute deviations, which is because their NMR chemical shift spectrum spans a range that is about an order of magnitude smaller than that of C and N nuclei. Nevertheless, the shifts of the hydrogens adjacent to the bond cut (H_4 and H_5) are considerably better. The problem for H_4 might be due to the effect of the lone electron pairs of the nearby nitrogen, which are missing entirely when the left fragment is replaced by the capping potential. We think that the same phenomenon applies to H_{10}, which is in somewhat better agreement after the optimization. The deviations of H_{11} and H_{12} are quite small and do not obviously correlate to the choice of the capping potential. In most cases, the monovalent capping potential outperforms the heptavalent one.

In addition to the NMR chemical shifts, we have used the optimized capping potential to compute the proton affinities, with N_8 as the protonation site. The energy differences between the neutral and charged histidine molecules (also given in Table 4) from the $\mathscr{D}_{opti}$ capping model are quite accurate when compared to the full-QM results. The error from the full-QM proton affinity is reduced from $\Delta E = 7.7$ kcal/mol to $\Delta E = 1.3$ kcal/mol, which corresponds to less than 1% of the total affinity. In conclusion, the capping potential which was optimized for the ethane molecules yields a very good overall accuracy when transferred to histidine.

3.6. Lysine. Our second QM/MM application of the capping potential is the amino acid lysine, which was split into QM/MM fragments as illustrated in Figure 5. In our setup, the amino group was replaced by our initial and optimized dummy potentials; note that also here the $\mathscr{D}_{opti}$ parameters were taken from the ethane-based optimization without any further change. The results of the geometry optimization using our initial and optimized capping potentials $\mathscr{D}_{ini}$ and $\mathscr{D}_{opti}$ are shown in Table 4. While the C_1-C_2 bond is not affected by the bond cleavage, the $\mathscr{D}-C_1$ bond length is notably different for the unoptimized capping potential. A similar improvement is observed for the proton affinity of lysine.

The NMR chemical shifts obtained via our full-QM and QM/MM calculations are summarized in Table 6, again using the optimized geometry of the full lysine. It turns out that the results for the $\mathscr{D}_{opti}$ capping potential are in better agreement with the full-QM values than that those from the hydrogen capping and the unoptimized dummy $\mathscr{D}_{ini}$. The deviation of the C_1 chemical shift is decreased from 30 to <1 ppm, and for H_6 and H_7, we find improvements from $\Delta\delta = 1.1$ ppm to $\Delta\delta < 0.1$ ppm. This suggests that the nonoptimized dummy potential leads to a distortion of the electronic density that significantly affects the atoms near the "broken" QM/MM bond. Inspecting the shifts of C_2, C_3, and C_4 under the $\mathscr{D}_{ini}$ capping, we find that the influence of the bond cleavage is not negligible even for the atoms which are located several bonds away from the QM/MM border. Nevertheless, the new $\mathscr{D}_{opti}$ agrees quite well with the full-QM calculations for all atoms.

4. Conclusion

We have presented improved monovalent capping potentials for hybrid QM/MM calculations within density functional theory. The parameters of analytic effective pseudopotentials are optimized such as to reproduce the electronic density, proton affinities, atomic forces, and geometries as closely as possible with respect to the corresponding full-QM quantities. Particular focus is put on the reliability of NMR chemical shifts as highly sensitive probes of the ground-state and response properties of the electronic orbitals. The resulting analytic capping potentials are shown to have a high transferability for different molecules. An important advantage resulting from the improved electronic structure of the optimized capping potentials is that our $\mathscr{D}_{opti}$ can help to reduce the QM box size significantly, since the perturbation of the QM/MM bond cleavage is essentially undetectable in the QM region beyond one single chemical bond.

Acknowledgment. S.K. thanks the Ubon Ratjatanee University. D.S. acknowledges support from the Deutsche Forschungsgemeinschaft (DFG) under Grants SE-1008/5 and SE-1008/6. We are grateful to Jochen Schmidt for useful discussions and for his precious help.

References

(1) Lehn, J.-M. *Rep. Prog. Phys.* **2004**, *67*, 249–265.

(2) Tolstoy, P. M.; Schah-Mohammedi, P.; Smirnov, S. N.; Golubev, N. S.; Denisov, G. S.; Limbach, H. H. *J. Am. Chem. Soc.* **2004**, *126*, 5621–5634.

(3) Meng, S.; Xu, L. F.; Wang, E. G.; Gao, S. *Phys. Rev. Lett.* **2002**, *89*, 176104.

(4) Chen, B.; Ivanov, I.; Klein, M. L.; Parrinello, M. *Phys. Rev. Lett.* **2003**, *91*, 215503.

(5) Rapp, A.; Schnell, I.; Sebastiani, D.; Brown, S. P.; Percec, V.; Spiess, H. W. *J. Am. Chem. Soc.* **2003**, *125*, 13284–13297.

(6) Goward, G.; Sebastiani, D.; Schnell, I.; Spiess, H. W. *J. Am. Chem. Soc.* **2003**, *125*, 5792–5800.

(7) Lee, Y.; Murakhtina, T.; Sebastiani, D.; Spiess, H. *J. Am. Chem. Soc.* **2007**, *129*, 12406–12407.

(8) Gervais, C.; Dupree, R.; Pike, K. J.; Bonhomme, C.; Profeta, M.; Pickard, C. J.; Mauri, F. *J. Phys. Chem. A* **2005**, *109*, 6960–6969.

(9) Murakhtina, T.; Delle Site, L.; Sebastiani, D. *ChemPhysChem* **2006**, *7*, 1215–1219.

(10) Rohrig, U.; Guidoni, L.; Laio, A.; Frank, I.; Rothlisberger, U. *J. Am. Chem. Soc.* **2004**, *126*, 15328–15329.

(11) Sebastiani, D.; Parrinello, M. *ChemPhysChem* **2002**, *3*, 675.

(12) Murakhtina, T.; Heuft, J.; Meijer, J.-E.; Sebastiani, D. *ChemPhysChem* **2006**, *7*, 2578–2584.

(13) Bühl, M.; Grigoleit, S.; Kabrede, H.; Mauschick, F. T. *Chem.—Eur. J.* **2006**, *12*, 477–488.

(14) Hansen, M. R.; Sekharan, S.; Graf, R.; Sebastiani, D. *J. Am. Chem. Soc.* **2009**, *131*, 5251–5256.

(15) Deng, R. Z.; Martyna, G. J.; Klein, M. L. *Phys. Rev. Lett.* **1993**, *71*, 267.

(16) Laio, A.; VandeVondele, J.; Roethlisberger, U. *J. Chem. Phys.* **2002**, *116*, 6941.

(17) Komin, S.; Gossens, C.; Tavernelli, I.; Röthlisberger, U.; Sebastiani, D. *J. Phys. Chem. B* **2007**, *111*, 5225–5232.

(18) Wei, D.; Salahub, D. *Chem. Phys. Lett.* **1994**, *224*, 291.

(19) Stanton, R. V.; Little, L. R.; Merz, K. M. *J. Phys. Chem.* **1996**, *99*, 11266.

(20) Eichinger, M.; Tavan, P.; Hutter, J.; Parrinello, M. *J. Chem. Phys.* **1999**, *21*, 10452.

(21) Lyne, P.; Hodosceck, M.; Karplus, M. *J. Phys. Chem. A* **1999**, *103*, 3462.

(22) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.

(23) Zhang, Y.; Lee, T.-S.; Yang, W. *J. Phys. Chem.* **1999**, *110*, 46–54.

(24) Brancato, G.; Rega, N.; Barone, V. *J. Chem. Phys.* **2008**, *128*, 144501.

(25) Cui, Q. *J. Chem. Phys.* **2002**, *117*, 4720–4728.

(26) Singh, U. C.; Kollman, P. A. *J. Comput. Chem.* **1986**, *7*, 718.

(27) Assfeld, X.; Rivail, J.-L. *Chem. Phys. Lett.* **1996**, *263*, 100–106.

(28) Jacob, C. R.; Visscher, L. *J. Chem. Phys.* **2006**, *125*, 194104.

(29) Pu, J.; Gao, J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 632–650.

(30) Jung, J.; Choi, C. H.; Sugita, Y.; Ten-no, S. *J. Chem. Phys.* **2007**, *127*, 204102.

(31) Ohta, K.; Yoshioka, Y.; Morokuma, K.; Kitaura, K. *Chem. Phys. Lett.* **1983**, *101*, 12–17.

(32) Day, P. N.; Jensen, J. H.; Gordon, M. S.; Webb, S. P.; Stevens, W. J.; Krauss, M.; Garmer, D. *J. Chem. Phys.* **1996**, *105*, 1968.

(33) Adamovic, I.; Freitag, M. A.; Gordon, M. S. *J. Chem. Phys.* **2003**, *118*, 6725–6732.

(34) Netzloff, H. M.; Gordon, M. S. *J. Chem. Phys.* **2004**, *121*, 2711–2714.

(35) Adamovic, I.; Gordon, M. S. *J. Phys. Chem. A* **2006**, *110*, 10267–10273.

(36) Poteau, R.; Ortega, I.; Alary, F.; Solis, A. R.; Barthelat, J.-C.; Daudey, J.-P. *J. Phys. Chem. A* **2001**, *105*, 198–205.

(37) Poteau, R.; Alary, F.; Makarim, H. A. E.; Heully, J.-L.; Barthelat, J.-C.; Daudey, J.-P. *J. Phys. Chem. A* **2001**, *105*, 206–214.

(38) Exner, T. E.; Mezey, P. G. *J. Comput. Chem.* **2003**, *24*, 1980–1986.

(39) Exner, T. E.; Mezey, P. G. *Phys. Chem. Chem. Phys.* **2005**, *24*, 4061–4069.

(40) Eckard, S.; Exner, T. E. *Z. Phys. Chem.* **2006**, *220*, 927–944.

(41) Jardilliera, N.; Goursot, A. *Chem. Phys. Lett.* **2008**, *454*, 65–69.

(42) Mallik, A.; Taylor, D. E.; Runge, K.; Dufty, J. W. *Int. J. Quantum Chem.* **2004**, *100*, 1019–1025.

(43) DiLabio, G. A.; Wolkow, R. A.; Johnson, E. R. *J. Chem. Phys.* **2005**, *122*, 044708.

(44) Slavicek, P.; Martinez, T. J. *J. Chem. Phys.* **2006**, *124*, 084107.

(45) DiLabio, G. A.; Hurley, M. M.; Christiansen, P. A. *J. Chem. Phys.* **2002**, *116*, 9578–9584.

(46) Goedecker, S.; Teter, M.; Hutter, J. *Phys. Rev. B* **1996**, *54*, 1703.

(47) Hartwigsen, C.; Goedecker, S.; Hutter, J. *Phys. Rev. B* **1998**, *58*, 3641.

(48) Rohrig, U. F.; Sebastiani, D. *J. Phys. Chem. B* **2008**, *112*, 1267–1274.

(49) von Lilienfeld-Toal, A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *J. Chem. Phys.* **2005**, *122*, 014113.

(50) Brown, S. P.; Spiess, H. W. *Chem. Rev.* **2001**, *101*–4125.

(51) Spiess, H. W. *Macromol. Chem. Phys.* **2003**, *204*, 340–346.

(52) Schulz-Dobrick, M.; Metzroth, T.; Spiess, H. W.; Gauss, J.; Schnell, I. *ChemPhysChem* **2005**, *6*, 315–327.

(53) Ochsenfeld, C.; Brown, S. P.; Schnell, I.; Gauss, J.; Spiess, H. W. *J. Am. Chem. Soc.* **2001**, *123*, 2597–2606.

(54) Bühl, M.; Kabrede, H.; Diss, R.; Wipff, G. *J. Am. Chem. Soc.* **2006**, *128*, 6357–6368.

(55) Pickett, W. E. *Comput. Phys. Rep.* **1989**, *9*, 115–198.

(56) Press, W. H.; Teukoldky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes*, 2nd ed.; Cambridge University Press: Cambridge, U.K., 1992.

(57) Putrino, A.; Sebastiani, D.; Parrinello, M. *J. Chem. Phys.* **2000**, *113*, 7102–7109.

(58) Putrino, A.; Parrinello, M. *Phys. Rev. Lett.* **2002**, *88*, 176401.

(59) Sebastiani, D.; Parrinello, M. *J. Phys. Chem. A* **2001**, *105*, 1951.

(60) Hohenberg, P.; Kohn, W. *Phys. Rev.* **1964**, *136*, B864.

(61) Kohn, W.; Sham, L. J. *Phys. Rev.* **1965**, *140*, A1133.

(62) Jones, R. O.; Gunnarsson, O. *Rev. Mod. Phys.* **1989**, *61*, 689–746.

(63) Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098.

(64) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

(65) Hutter, J.; Marx, D.; Focher, P.; Tuckerman, M.; Andreoni, W.; Curioni, A.; Fois, E.; Rothlisberger, U.; Giannozzi, P.; Deutsch, T.; Alavi, A.; Sebastiani, D.; Laio, A.; VandeVondele, J.; Seitsonen, A.; Billeter, S.; Parrinello, M. *Computer Code CPMD*, version 3.12; copyright IBM Corp. and MPI-FKF Stuttgart 1990−2007; http://www.cpmd.org.

(66) Hutter, J.; Curioni, A. *ChemPhysChem* **2005**, *6*, 1788–1793.

(67) Sebastiani, D.; Goward, G.; Schnell, I.; Spiess, H. W. *J. Mol. Struct.: THEOCHEM* **2003**, *625*, 283–288.

(68) Gottlieb, H. E.; Kotlyar, V.; Nudelman, A. *J. Org. Chem.* **1997**, *62*, 7512–7515.

(69) Ulrich, E. L.; Akutsu, H.; Doreleijers, J. F.; Harano, Y.; Ioannidis, Y. E.; Lin, J.; Livny, M.; Mading, S.; Maziuk, D.; Miller, Z.; Nakatani, E.; Schulte, C. F.; Tolmie, D. E.; Wenger, R. K.; Yao, H.; Markley, J. L. *Nucleic Acids Res.* **2007**, *36*, D402−D408.

CT800525U