A QM/MM Implementation of the Self-Consistent Charge Density Functional Tight Binding (SCC-DFTB) Method

Qiang Cui,[†] Marcus Elstner,[‡,§] Efthimios Kaxiras,[‡] Thomas Frauenheim,[§] and Martin Karplus*[,†,⊥,||]

Department of Chemistry and Chemical Biology and Department of Physics, Harvard University, Cambridge, Massachusetts 02138; Department of Theoretical Physics, University of Paderborn, D-33098 Paderborn, Germany; Laboratoire de Chimie Biophysique, ISIS Université Loius Pasteur, 67000 Strasbourg, France; and New Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QT U.K.

Received: August 10, 2000; In Final Form: October 27, 2000

A quantum mechanical/molecular mechanical (QM/MM) approach based on an approximate density functional theory, the so-called self-consistent charge density functional tight binding (SCC-DFTB) method, has been implemented in the CHARMM program and tested on a number of systems of biological interest. In the gas phase, SCC-DFTB gives reliable energetics for models of the triosephosphate isomerase (TIM) catalyzed reactions. The rms errors in the energetics compared to B3LYP/6-31+G(d,p) are about 2−4 kcal/mol; this is to be contrasted with AM1, where the corresponding errors are 9−11 kcal/mol. The method also gives accurate vibrational frequencies. For the TIM reactions in the presence of the enzyme, the overall SCC-DFTB/CHARMM results are in somewhat worse agreement with the B3LYP/6-31+G(d,p)/CHARMM values; the rms error in the energies is 5.4 kcal/mol. Single-point B3LYP/CHARMM energies at the SCC-DFTB/CHARMM optimized structures were found to be very similar to the full B3LYP/CHARMM values. The relative stabilities of the α_R and 3_{10} conformations of penta- and octaalanine peptides were studied with minimization and molecular dynamics simulations in vacuum and in solution. Although CHARMM and SCC-DFTB give qualitative different results in the gas phase (the latter is in approximate agreement with previous B3LYP calculations), similar behavior was found in aqueous solution simulations with CHARMM and SCC-DFTB/CHARMM. The 3_{10} conformation was not found to be stable, and converted to the α_R form in about 15 ps. The α_R conformation was stable in the simulation with both SCC-DFTB/CHARMM and CHARMM. The $i,i+3$ CO···HN distances in the α_R conformation were shorter with the SCC-DFTB method (2.58 Å) than with CHARMM (3.13 Å). With SCC-DFTB/CHARMM, significant populations with $i,i+3$ CO···HN distances near 2.25 Å, particularly for the residues at the termini, were found. This can be related to the conclusion from NMR spectroscopy that the 3_{10} configuration contributes for alanine-rich peptides, especially at the termini.

I. Introduction

Although dramatic progress has been made in the field of quantum chemistry in recent years,[1] highly accurate ("chemical accuracy") calculations are still limited to systems of less than 10 heavy atoms.[2] Nevertheless, meaningful calculations have been made for significantly larger systems. One example is the application of linear scaling methods[3] in a semiempirical framework to proteins and nucleic acids.[4] A problem where such methods can be useful is that of long-range electron transfer.[5] In most other cases of interest, treating the entire protein or nucleic acid by a quantum mechanical approach is not necessary; i.e., only part of the system needs to be treated quantum mechanically and that part may require high-level density functional or ab initio methods. Combined quantum mechanical/molecular mechanical (QM/MM) methods[6] are well suited for such problems; examples are reactions in solution or in enzyme.[7] Both with semiempirical and ab initio quantum mechanics, QM/MM methods have been shown to provide fundamental insights into the mechanism of chemical and enzymatic reactions.[6,7] For a QM region of limited size (<50 atoms), the QM/MM method can be applied to systems containing thousands of atoms. Most studies have concentrated on the potential energy surface, though solvation effects on pK_a values,[8] solvent-induced spectral shifts in UV absorption of organic molecules,[9] and charge polarization have also been studied.[10] Recently, we have developed and applied QM/MM methodology to the calculation of infrared spectra[11] and NMR chemical shifts.[12]

Since high-level quantum mechanical calculations in the QM/MM approach are computationally intensive, it is very time consuming to do the full dynamical simulations required to obtain a potential of mean force for an enzymatic reaction, for example. This means that semiempirical QM methods have to be employed. To achieve the required accuracy, it is possible to use an EVB[13] or MNDO/AM1[14,20] framework with parameters optimized for the system of interest. An attractive alternative would be a general semiempirical approach that gives higher accuracy than the standard methods. A promising new approach, called the self-consistent charge density functional tight binding (SCC-DFTB) method, has been developed recently.[15] It is based on a second-order expansion of the Kohn−Sham total energy with respect to charge density variations in the LCAO (tight-binding) framework. In contrast to other non-

* To whom correspondence should be addressed.
[†] Department of Chemistry and Chemical Biology, Harvard University.
[‡] Department of Physics, Harvard University.
[§] University of Paderborn.
[⊥] ISIS Université Loius Pasteur.
[||] University of Oxford.

10.1021/jp0029109 CCC: $20.00
Published on Web 12/16/2000

self-consistent DFTB approaches,[16] this method introduces relaxation of the charge density at the level of Mulliken population, which has greatly decreased the dependence on the initial density and increased the transferability of parameters. The method has been applied to calculate the energies, geometries, and vibrational frequencies of small organic molecules, and the resulting mean average deviations from the experimental values were comparable to those of full density functional theory (DFT) calculations with a double-ζ plus polarization basis set. The method has been tested for biologically relevant molecules, including structures of hydrogen-bonded complexes (e.g., HCOOH dimer, $NH_4^+-NH_3$), small peptides, and base pairing and stacking interactions.[17] The geometries of the hydrogen-bonded complexes as well as the geometries and relative energies for various conformers of the capped peptide *N*-acetyl-(L-Ala)$_n$-*N*′-methylamide compared well with B3LYP/6-31G(d) and MP2/6-31G(d) results. The binding energies of hydrogen-bonded complexes are underestimated with the SCC-DFTB approach by about 1−2 kcal/mol, compared to ab initio calculations at the MP2 and MP4 levels with double-ζ plus polarization quality basis sets. Recently, Bohr et al.[18] have calculated the vibrational absorption and vibrational circular dichroism spectra of the two lowest energy conformers of *N*-acetyl-(L-Ala)$_n$-*N*′-methylamide with a hybrid DFT/SCC-DFTB approach. The Hessian matrixes at the SCC-DFTB level were used to obtain the frequencies and normal modes, and the dipole derivatives and VCD tensors at the B3LYP/6-31G(d) level were used to obtain the IR and VCD intensities. The results were found to be in good overall agreement with those from MP2/6-31G(d) and B3LYP/6-31G(d) calculations.

Given these promising results, we have implemented the QM/MM approach in the CHARMM program[19] with SCC-DFTB as the QM method and with the MM potential functions from CHARMM. In section II, we briefly review the SCC-DFTB approach and describe its implementation in the CHARMM program. In section III, we illustrate the SCC-DFTB/CHARMM approach for a number of systems, including the reactions catalyzed by triosephosphate isomerase (TIM) and the dynamics of small peptide helices in water. For TIM, comparisons are made with our previous DFT and DFT/CHARMM calculations.[20] For the helical systems, the results are compared with MM based CHARMM calculations. The conclusions are presented in section IV.

II. Theory and CHARMM Implementation

a. The SCC-DFTB Formulation. The SCC-DFTB approach has been described in some detail,[15] so it is only reviewed briefly here. The method is derived from density functional theory (DFT) by a second-order expansion of the DFT total energy with respect to charge density variation, $\delta\rho$, relative to a chosen reference density, ρ_0. The resulting equation for the total energy is

$$E = \sum_i^{\mathrm{occ}} \langle \phi_i | \hat{H}^0 | \phi_i \rangle + \frac{1}{2} \int\int \left(\frac{1}{|\vec{r} - \vec{r}'|} + \left. \frac{\delta^2 E_{\mathrm{ex}}}{\delta\rho\delta\rho'} \right|_{\rho_0} \right) \times \delta\rho(\vec{r})\delta\rho'(\vec{r}')\, \mathrm{d}\vec{r}\, \mathrm{d}\vec{r}' + \left\{ -\frac{1}{2} \int\int \frac{\rho_0'(\vec{r}')\rho_0(\vec{r})}{|\vec{r} - \vec{r}'|} \mathrm{d}\vec{r}\, \mathrm{d}\vec{r}' + E_{\mathrm{xc}}[\rho_0(\vec{r})] - \int V_{xc}[\rho_0]\rho_0(\vec{r})\, \mathrm{d}\vec{r} \right\} + E_{\mathrm{core}} \quad (1)$$

The ϕ_i are the Kohn−Sham orbitals, E_{xc} and V_{xc} represents the exchange-correlation energy functional and potential, respectively. The symbol $\hat{H}^0$ stands for the effective Kohn−Sham Hamiltonian, which depends only on the reference density, ρ_0

$$\hat{H}_0 = -\frac{1}{2}\nabla^2 - \sum_k \frac{Z_k}{|\vec{R}_k - \vec{r}|} + \int \frac{\rho_0(r')}{|\vec{r} - \vec{r}'|} \mathrm{d}\vec{r}' + V_{\mathrm{xc}}[\rho_0(\vec{r})] \quad (2)$$

where Z_k is the charge of the kth nucleus.

The second term in eq 1 represents the contribution of the density variations to the total energy; the third term with the curly brackets corresponds to the correction for the double counting terms from the Coulombic and exchange-correlation contributions included in the $\hat{H}_0$ matrix elements; and the last term, E_{core}, is the core−core repulsion energy, which plays an important role in the parametrization (see below).

To obtain the SCC-DFTB energy expression, eq 1 is subjected to a number of approximations. Like most other semiempirical methods, only the valence electrons are treated explicitly and the nuclear charges, Z_k in eq 2, are adjusted accordingly. The Hamiltonian matrix elements $\langle \phi_i | \hat{H}^0 | \phi_i \rangle$ are represented in a *minimal basis* of localized pseudo-atomic Slater orbitals, χ_μ

$$\phi_i = \sum_\mu c_\mu^i \chi_\mu \quad (3)$$

To determine χ_μ, the neutral atom valence electron DFT problem is solved with an additional harmonic potential, in the form of $(r/r_0)^2$, to localize the orbitals.[21] For the determination of the confinement radius, r_0, a variational principle can be applied;[22] trial calculations have shown that the choice $r_0 = (1.8-2.0)r_{\mathrm{cov}}$ (where r_{cov} is the covalent radius of the element, as obtained from ref 23) is satisfactory. The Hamiltonian matrix elements in this LCAO basis, $H^0_{\mu\nu} \equiv \langle \chi_\mu | \hat{H}^0 | \chi_\nu \rangle$, are then approximated as follows. The diagonal elements, $H^0_{\mu\mu}$, are taken to be the calculated atomic eigenvalues, and the nondiagonal elements are calculated with a two-center approximation

$$H^0_{\mu\nu} = \langle \chi_\mu | \hat{T} + V_{\mathrm{eff}}[\rho_0^{\mathrm{A}} + \rho_0^{\mathrm{B}}] | \chi_\nu \rangle, \quad \mu \in \mathrm{A}, \nu \in \mathrm{B} \quad (4)$$

where V_{eff} is the effective Kohn−Sham potential (Coulombic plus exchange-correlation) and ρ_0^{A} is the electron density of the neutral atom A; A and B define the two atoms. The Hamiltonian matrix elements, $H^0_{\mu\nu}$, and the overlap matrix elements, $S_{\mu\nu}$, are computed and tabulated with respect to the interatomic distances between A and B. The values of these matrix elements are obtained by interpolation of the tabulated data. This is one of the factors that make the SCC-DFTB calculation fast.

The last four terms on the second line in eq 1 involve the input neutral electronic density ρ_0 and E_{core}. They are combined in a form E_{rep}, which is approximated as the sum of a set of pairwise atom−atom potentials, which are then fitted for selected reference systems and assumed to be transferable in calculations for more complicated molecules. Typical reference systems include H_3C-CH_3, $H_2C=CH_2$, and $HC\equiv CH$ for the determination of the C−C interaction and $HO-CH_3$ and $O=CH_2$ for C−O. In the calculations, low spin states are always assumed and described with the spin-restricted approach.

$$E_{\mathrm{rep}} = \sum_{\mathrm{A,B}} U^{\mathrm{A-B}}(R_{\mathrm{AB}}) \quad (5)$$

The pairwise potential in eq 5 is fitted with polynomials or spline functions for the chosen reference systems at a series of interatomic distances; it is taken as the difference between the DFT total energy and the SCC-DFTB electronic energy (the

first two terms on the rhs of eq 7 below). In the current implementation, the exchange-correlation functional used in SCC-DFTB is the one derived by Perdew, Burke, and Ernzerhof;[24] while the DFT method used as the standard for the fitting was the B3LYP method associated with the double-ζ plus polarization quality basis sets. An energy shift was applied to account for the difference in the number of electrons included in the DFT and SCC-DFTB calculations, based on the dissociation limit of the reference system.

The charge density variation, $\delta\rho$, relative to the reference density, ρ_0, is written as a superposition of atomic contributions, $\delta\rho^A$

$$\delta\rho = \sum_A \delta\rho^A \tag{6}$$

which are approximated by the charge fluctuations at the atoms A, $\Delta q^A = q^A - q_0^A$. The quantity q_0^A is equal to the number of valence electrons of the neutral atom A, and q^A is the Mulliken population resulting from the sum over the orbitals. The second derivative of the total energy with respect to the charge density fluctuations, $(\partial^2 E/\delta\rho\delta\rho')|_{\rho_0}$, is approximated by a function γ_{AB}. In the limit of A = B, the quantity $\gamma_{AA}/2$ is approximated[25] by the chemical hardness[26] of the atomic species, and is calculated with PBE.[15] In the case of A $\neq$ B, commonly used functional forms for γ_{AB} have been presented by Ohno, Klopman, and Mataga-Nishimoto, for example.[27] However, these functionals were not adopted in the current implementation because the SCC-DFTB method was originally designed for periodic systems for which these functional can cause severe numerical problems. A new functional form for γ_{AB} was proposed, based on the analysis of Coulombic interaction between two exponentially decaying spherical charge distributions centered on the two nuclei;[15] the form is

$$\gamma_{AB} = \int\int \frac{s(\tau_A,\vec{R}_A)s(\tau_B,\vec{R}_B)}{|\vec{r} - \vec{r}'|}\,dr\,dr' = \frac{1}{R_{AB}} - [e^{-\tau_A R_{AB}}K(\tau_A,\tau_B,R_{AB}) + e^{-\tau_B R_{AB}}K(\tau_B,\tau_A,R_{AB})] \tag{7}$$

where $s(\tau_A, \vec{R}_A)$ is a 1s Slater orbital centered on $\vec{R}_A$ with exponent τ_A, and $K(\tau_A,\tau_B,R_{AB})$ is given by[15]

$$K(\tau_A,\tau_B,R_{AB}) = \frac{\tau_B^4\tau_A}{2(\tau_A^2 - \tau_B^2)^2} - \frac{\tau_B^6 - 3\tau_B^4\,\tau_A^2}{(\tau_A^2 - \tau_B^2)^3 R_{AB}} \tag{8}$$

The function γ_{AB} has the asymptotic form $(1/R_{AB})$ at long distance, and approaches γ_{AA} at the limit of $R_{AB} \rightarrow 0$.[15] The latter condition relates the decaying exponent of the spherical charge distribution to the chemical hardness (U_A);[26] that is

$$\tau_A = \frac{16}{5}U_A \tag{9}$$

With the above approximations, the total SCC-DFTB energy is written in the following form

$$E^{SCC-DFTB} = \sum_i^{occ}\langle\phi_i|\hat{H}^0|\phi_i\rangle + \frac{1}{2}\sum_{A,B}\gamma_{AB}\Delta q^A\Delta q^B + E_{rep} \tag{10}$$

By applying the variational principle, one obtains Kohn−Sham type equations, which have to be solved iteratively since the energy $E^{SCC-DFTB}$ depends on the orbital coefficients, c_μ^i, due to the Mullikan charges in eq 10. Analytic first derivatives of the total energy, which are essential in geometry optimization and molecular dynamics, are readily obtained at the SCC-DFTB level[15]

$$\frac{\partial E^{SCC-DFTB}}{\partial \vec{R}_A} = \sum_i n_i \sum_{\mu\nu} c_\mu^i\, c_\nu^i \times \left[\frac{\partial H_{\mu\nu}^0}{\partial \vec{R}_A} - \left(\epsilon_i - \frac{1}{2}\sum_c(\gamma_{AC} + \gamma_{BC})\Delta q^c\right)\frac{\partial S_{\mu\nu}}{\partial \vec{R}_A}\right] + \Delta q^A \sum_c \frac{\partial \gamma_{AC}}{\partial \vec{R}_A}\Delta q^c + \frac{\partial E_{rep}}{\partial \vec{R}_A} \quad \mu \in A;\ \nu \in B \tag{11}$$

The n_i and ϵ_i are the occupation number and eigenvalue for the ith molecular orbital, respectively, and $S_{\mu\nu}$ is the tabulated overlap matrix element. The second derivatives of the energy with respect to nuclear positions, which are used in vibrational frequency calculations, can in principle be formulated analytically. In the current implementation, however, they are calculated by numerical differentiation of the first derivatives. The two-center approximation (eq 4) and tabulated matrix elements make the calculation very fast, similar in speed to other semiempirical methods, such as AM1 and PM3. The remaining time-consuming step in SCC-DFTB, AM1 and PM3, which is the diagonalization of the Fock matrix and can be speeded up using methods such as the divide-and-conquer approach,[3,65] is the same.

b. QM/MM Implementation. In this section, we describe the implementation of SCC-DFTB in the CHARMM program.[19] The method follows that used previously for QM/MM implementations with semiempirical,[28] ab initio Hartree−Fock,[29] and DFT[30] as the QM method. The QM atoms interact with the MM atoms through electrostatic and van der Waals terms. The total energy of the system is written in the form

$$E^{tot} = \langle\Psi|\hat{H}^{QM} + \hat{H}_{el}^{QM/MM}|\Psi\rangle + E_{van}^{QM/MM} + E^{MM} \tag{12}$$

where the first term gives the electronic energy of the QM system including the electrostatic interaction from the MM part with $\langle\Psi|\hat{H}^{QM}|\Psi\rangle$ given by eq 10; the second term is the van der Waals interaction between the QM and MM atoms, which corresponds to the Pauli repulsion at short distances and dispersion at long distances; the last term is the empirical energy of the MM atoms. Formally, the Coulombic interaction between the MM partial charges and the QM electrons and nuclei is described by the one-electron operator

$$\hat{H}_{el}^{QM/MM} = \sum_{A\in MM}\sum_{B\in QM}\frac{Q_A Z_B}{|\vec{R}_A - \vec{R}_B|} - \sum_{A\in MM}\sum_{i=1}^{N_{el}}\frac{Q_A}{|\vec{R}_A - \vec{r}_i|} \tag{13}$$

where Q_A is the partial charge on MM atom A, Z_B is the nuclear charge of QM atom B, and N_{el} is the number of QM electrons. If this operator were treated analytically, which was done in previous QM/MM implementations for ab initio[29] and DFT[30] QM, two- and three-center integrals of the form

$$\left\langle \chi_\mu \left| \frac{Q_A}{|\vec{R}_A - \vec{r}_i|} \right| \chi_\nu \right\rangle$$

would have to be computed. This would be computationally intensive. To avoid this problem, the QM/MM electrostatic interaction is approximated in the spirit of the SCC-DFTB method by the Coulombic interaction between the Mulliken

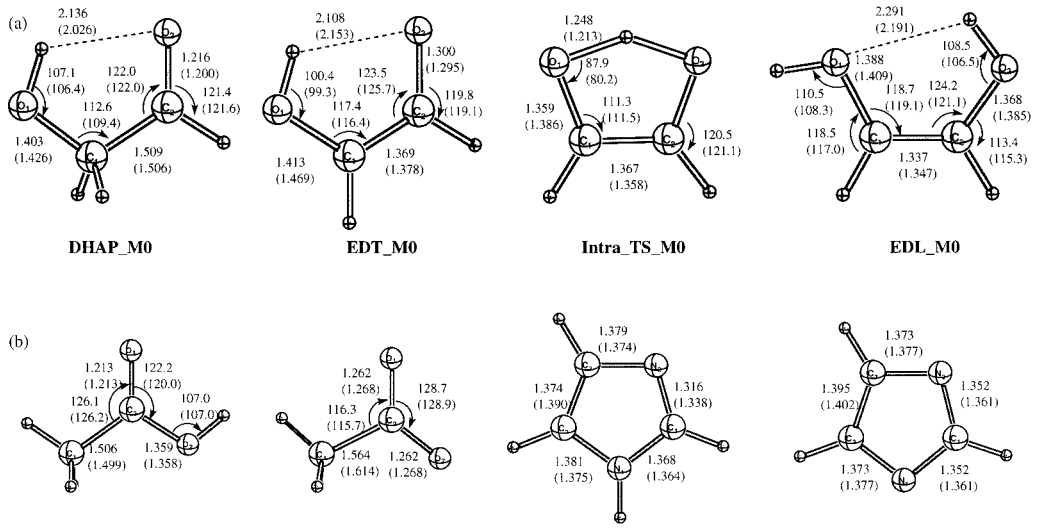


Figure 1. Gas-phase QM calculations for simplified models of substrate and catalytic residues in TIM-catalyzed reactions (M_0). (a) Optimized structures of substrates in different forms, plus the transition state for the intramolecular proton transfer in the enediolate. (b) Optimized structures for the models of the side chains of Glu165 and His95 in different protonation states. The numbers without parentheses were obtained with B3LYP/6-31+G(d,p), and those with parentheses were optimized with SCC-DFTB. The distances are in Å.

charge of the QM atoms and the MM partial charges. Higher order multipolar interactions between the QM electrons and MM partial charges are neglected; that is

$$\hat{H}^{\mathrm{el}}_{\mathrm{QM/MM}} \approx \sum_{\mathrm{A}\in\mathrm{MM}} \sum_{\mathrm{B}\in\mathrm{QM}} \frac{Q_{\mathrm{A}}\Delta q^{\mathrm{B}}}{|\vec{R}_{\mathrm{A}} - \vec{R}_{\mathrm{B}}|} \tag{14}$$

This approximation is different from that used in previous QM/MM implementation with a semiempirical QM.[6b] There the exact integral for MM charge−electron interaction in eq 13 was found to be proportionately too large and therefore was approximated by a term of the form $-Q_{\mathrm{A}}\langle\mu_{\mathrm{B}}\nu_{\mathrm{B}}|s_{\mathrm{A}}s_{\mathrm{A}}\rangle$, where a 1s-type Slater orbital was placed on the MM atom A.

Substituting eq 14 into eq 10, and applying the variational principle, one obtains the eigenvalue equations

$$\sum_{\nu}(H_{\mu\nu} - \epsilon_i S_{\mu\nu})c^i_{\nu} = 0 \tag{15}$$

where the matrix elements including the MM partial charges have the form

$$H_{\mu\nu} = H^0_{\mu\nu} + \frac{1}{2}S_{\mu\nu}\sum_{\mathrm{B}\in\mathrm{QM}}(\gamma_{\mathrm{CB}} + \gamma_{\mathrm{DB}})\Delta q^{\mathrm{B}} + \frac{1}{2}S_{\mu\nu}\sum_{\mathrm{A}\in\mathrm{MM}}\left(\frac{Q_{\mathrm{A}}\Delta q^{\mathrm{C}}}{|\vec{R}_{\mathrm{C}} - \vec{R}_{\mathrm{A}}|} + \frac{Q_{\mathrm{A}}\Delta q^{\mathrm{D}}}{|\vec{R}_{\mathrm{D}} - \vec{R}_{\mathrm{A}}|}\right) \quad \mu \in \mathrm{C};\ \nu \in \mathrm{D} \tag{16}$$

The C and D denote the QM atoms on which the atomic basis functions χ_μ and χ_ν, respectively, are located.

Preliminary test calculations with SCC-DFTB/AMBER[31] indicated that the approximation by eq 14 works rather well for a number of hydrogen-bonded complexes such as alanine peptides and water clusters.

III. Test Calculations

Several different types of calculations were performed to test the SCC-DFTB and SCC-DFTB/CHARMM methods in the context of biologically interesting systems. The first concerns the use of the SCC-DFTB method *per se*, for treating transition states as well as stable structures. For this purpose, model systems related to the triosephosphate isomerase (TIM) reactions were examined because values for the structures and energetics had been determined previously at various levels of the QM method.[20] The SCC-DFTB/MM formulation, eq 14, was then tested using the TIM related model systems and the TIM-catalyzed reactions in an enzymatic environment. The results were compared to the available values at different QM levels, including AM1-SRP/CHARMM and B3LYP/6-31+G(d,p)/CHARMM.[20] The final application examines the relative stability of 3_{10} and α_{R} helices in solution. A QM/MM molecular dynamics simulation was performed, in which the helices were treated with the SCC-DFTB method and the water molecules were treated with the standard TIP3P MM potential.[32]

III.1. Gas-Phase Models for Species Involved in the TIM Catalysis: Test of the SCC-DFTB Method for Stable and Transition States. The reaction catalyzed by TIM is the interconversion of dihydroxyacetone phosphate(DHAP) and D-glyceraldehyde 3-phosphate (GAP).[33] In earlier work, several model systems of interest for the TIM-catalyzed reactions were studied (see ref 20 for details). In the present paper, we use the earlier calculations on some of these as tests of the SCC-DFTB and the SCC-DFTB/CHARMM methods. Three sets of models are considered in this section. The first set is the minimal model (M_0 set), which includes different states of the substrate without the phosphate group plus models for the catalytic residues. The second set (M_{2NP}) includes different states of the substrates with a neutralized phosphate group; these permit a test of the SCC-DFTB parameters for phosphate. As explained previously,[20] we use a neutralized phosphate group despite the fact that the realistic substrate is doubly anionic because the negative charge is expected to be stabilized in the enzyme by (charged) residues, such as Lys12 in the active site. The third set is a "basic" model (M_{3b}), which includes the substrate and at most one catalytic residue, either Glu165 or His95. The choice of residue is based on the one involved explicitly in the proton-transfer step. We compare results obtained with structures optimized at the AM1, PM3, SCC-DFTB, and B3LYP/6-31+G(d,p) levels. The structures are shown in Figures 1 and 2 for the M_0 and M_{2NP} systems, respectively. For the M_0 set, single-point MP2 and CCSD energies have also been calculated with the 6-311+G(d,p) basis set at the B3LYP/6-31+G(d,p) geometries. The energies of

TABLE 1: Comparison of Quantum Mechanical Results for Species of Interest Involved in TIM[a]

species	AM1	PM3	SCC-DFTB	B3LYP/6-31+G(d,p)	MP2/6-311+G(d,p)[b]	CCSD/6-311+G(d,p)[b]
DHAP_M0	*−0.13877*	*−0.13222*	*−11.5642*	*−229.06152*	*−228.52206*	*−228.54424*
EDL_M0[b]	7.3	8.1	8.7	8.3	9.4	9.4
EDT_M0[b]	365.2	362.0	373.7	366.3	368.9	374.8
Intra_TS_M0[c]	398.9(33.8)	393.2(31.2)	386.2(12.5)	374.8 (8.5)	376.1 (7.2)	384.9 (10.1)

[a] For identification of the species, see Figure 1. Their roles in the TIM-catalyzed reactions are described in ref 20. The total energies for DHAP_M0 at different levels are given in hartrees in italics; other energies are in kcal/mol. [b] Calculated at the B3LYP/6-31+G(d,p) optimized structures. Zero point corrections (ZPC) are not included; this is true for all the tables. [c] The numbers without parentheses are calculated relative to DHAP_M0; the numbers in parentheses are the barrier height for intramolecular proton transfer measured from EDT_M0.

TABLE 2: Comparison of Quantum Mechanical Results for Simple Models of the Catalytic Residues in TIM[a]

species	AM1	PM3	SCC-DFTB	B3LYP/6-31+G(d,p)	MP2/6-311+G(d,p)[b]	CCSD/6-311+G(d,p)[b]
CH_3COOH	*−0.16422*	*−0.16264*	*−11.60418*	*−229.10583*	*−228.56784*	*−228.58734*
imidazole	*0.08088*	*0.04977*	*−11.383622*	*−226.23539*	*−225.63919*	*−225.65714*
CH_3COO^- [b]	354.8(10.4)	349.6(12.4)	356.2(17.5)	352.9 (13.4)	355.1 (13.8)	359.4 (15.4)
imidazolate[b]	347.7(10.2)	341.9(12.0)	361.2(3.7)	356.7 (1.4)	353.4 (6.2)	359.2 (6.3)

[a] The total energies for CH_3COOH and imidazole at different levels are given in hartrees in italics; other energies are in kcal/mol. [b] The numbers without parentheses are the deprotonation energies, and those in parentheses are exothermicities for the proton exchange reaction: DHAP_M0/EDL_M0 + CH_3COO^-/imidazolate → EDT_M0 + CH_3COOH/imidazole.

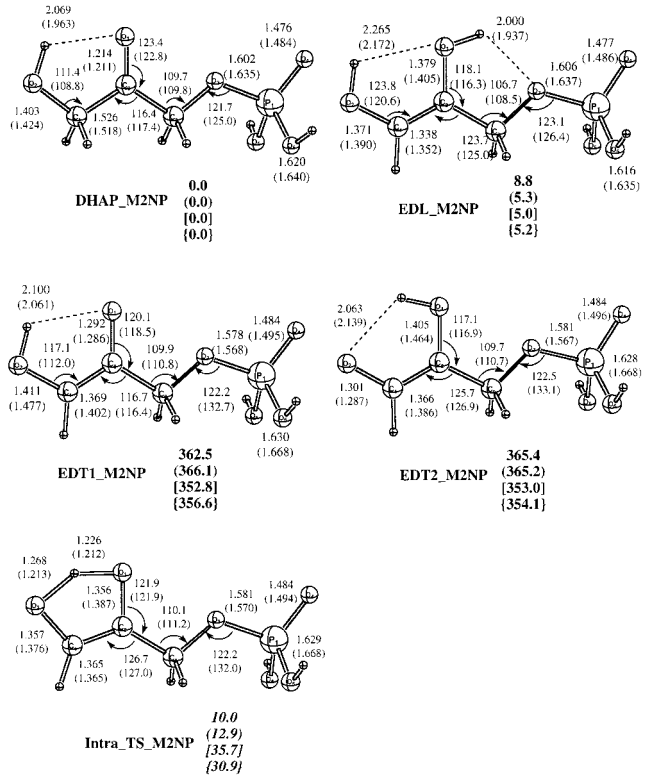


Figure 2. Gas-phase QM calculations of structures for the substrate and reaction intermediates with neutralized phosphate group (M_{2NP}). The numbers without parentheses were obtained with B3LYP/6-31+G-(d,p), and those with parentheses were optimized with SCC-DFTB. The bold numbers are energies measured relative to DHAP_M2NP, the exception is Intra_TS_M2NP for which the barrier height measured relative to EDT1_M2NP is shown in italics. The values with square and curled brackets were obtained at the AM1 and PM3 level, respectively. The distances are in Å.

various structures for the M_0 and M_{2NP} sets are summarized in Tables 1 and 2, Figure 2, respectively. For the M_{3b} set, only rms errors in the energetics are included in Table 3; selected structures are shown in Figure 3; the full list of structures and the energetics are given in Supporting Information.

Figure 1 shows that the geometries obtained with the B3LYP/6-31+G(d,p) and SCC-DFTB QM methods generally agree within 0.05 Å for bond lengths (mostly less than 0.02 Å) and 1° for bond angles; exceptions are the C−O distance in

TABLE 3: Rms Error in Relative Energetics (in kcal/mol) for Different Calculations on TIM-Related Models Compared to Full B3LYP or B3LYP/CHARMM Results[a]

model	AM1	AM1-SRP	PM3	SCC-DFTB
M_{3b}	11.1	3.9	11.7	4.0
M_{3a}[b]	9.7	1.3		2.4(2.2)
enzyme	8.4	4.5		5.3(5.5)[2.5][c]

[a] M_{3b} and M_{3a} were calculated in the gas phase, and the rms errors are defined relative to full B3LYP/6-31+G(d,p) results; the enzyme model was treated with the QM/MM approach as described in the text, for which the rms error was calculated relative to B3LYP/6-31+G(d,p)/CHARMM values. [b] The numbers without parentheses are rms errors for the QM/MM calculations in which model His95 was treated with the CHARMM22 force field; the value in parentheses is the rms error for pure QM results. [c] The number with and without the parentheses is obtained with the large and small QM partition, respectively; the value in the brackets is obtained with B3LYP/6-31+G(d,p)/CHARMM single-point energetic calculations at SCC-DFTB/CHARMM geometries with the small QM partition.

EDT_M0 and the H−O−C angle in Intra_TS_M0. The energies of various stable structures are shown in Tables 1 and 2 at the AM1, PM3, and SCC-DFTB levels, and they are all rather close to the higher level B3LYP, MP2, and CCSD calculations. The exception is that AM1 and PM3 give too small proton affinities for imidazole. As a result, the exothermicity for the reaction EDL_M0 + imidazolate → EDT_M0 + imidazole at the AM1 and PM3 levels, 10.2 and 12.0 kcal/mol, respectively, deviates substantially from the results obtained from other calculations. For the intramolecular proton transfer barrier height, the SCC-DFTB result of 12.5 kcal/mol is larger than the value of 8.5 kcal/mol at the B3LYP/6-31+G(d,p) level. The AM1 and PM3 calculations give barrier heights that are too high by more than 20 kcal/mol. Overall, the rms error of AM1, PM3, and SCC-DFTB compared to B3LYP/6-31+G(d,p) is 12.1, 11.4, and 3.9 kcal/mol, respectively. Compared to the CCSD/6-311+G(d,p) results, the rms error at the AM1, PM3, and SCC-DFTB level is 11.8, 11.4, and 2.0 kcal/mol, respectively; the rms error of B3LYP/6-31+G(d,p) is 5.6 kcal/mol. The SCC-DFTB method not only gives reliable structures and energetics, it also give vibrational frequencies of reasonable accuracy. For instance, the imaginary frequency for Intra_TS_M0 is 1673.8*i* cm^{-1} at the SCC-DFTB level, about 300 cm^{-1} larger than the value of 1383.8*i* cm^{-1} at the B3LYP/6-31+G(d,p) level; the AM1 and PM3 calculations, in contrast, give significantly higher values,

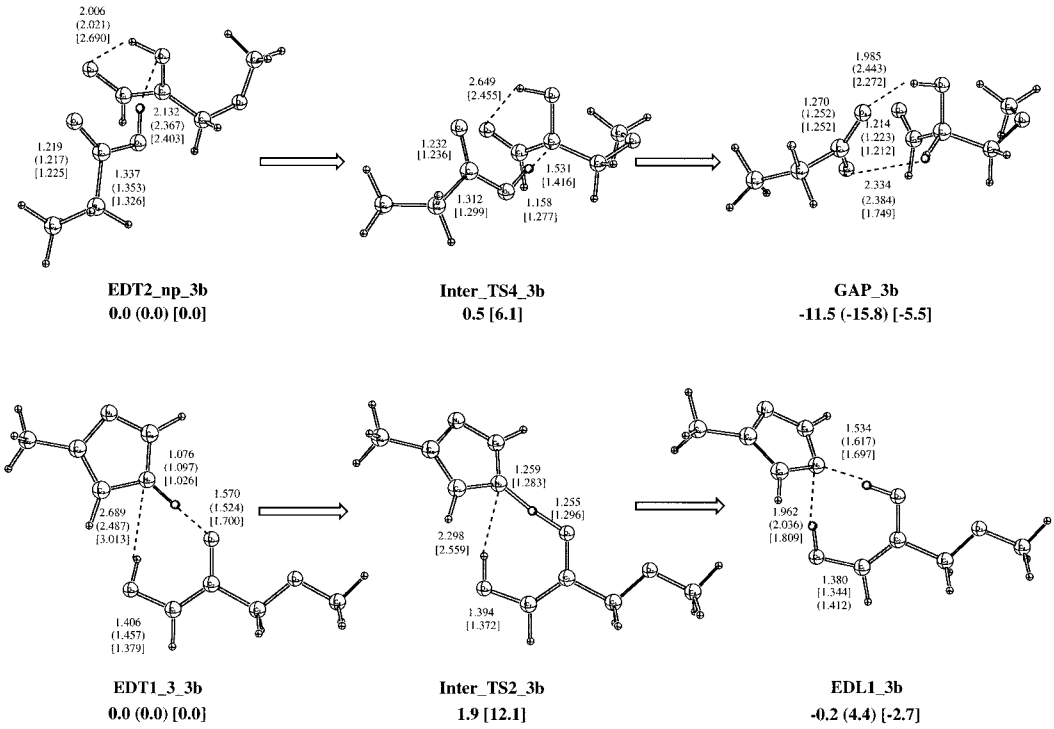


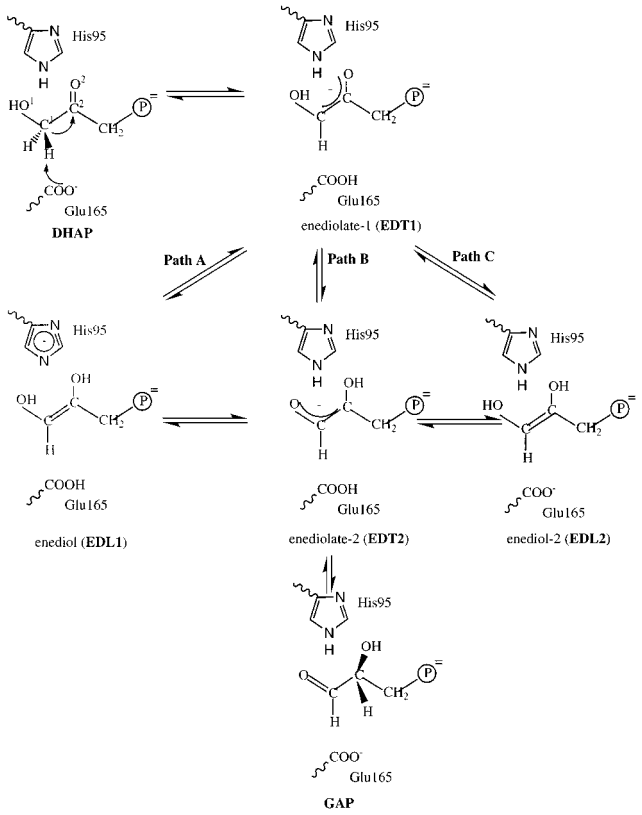
Figure 3. Selected structures from the M3b model for the TIM-catalyzed reactions. The numbers without parentheses were obtained with B3LYP/6-31+G(d,p), and these with parentheses and brackets were optimized with SCC-DFTB and PM3, respectively. The distances are in Å. The numbers in bold are energetics in kcal/mol, without zero-point corrections. The transition states were not located at the SCC-DFTB level, presumably because the barriers would be even lower than those at the B3LYP/6-31+G(d,p) level, which are already very low.

2175.0*i* and 2698*i* cm^{-1}, respectively, which are consistent with the overestimated barrier heights.

Figure 2 shows gas phase structures for the substrate and intermediate with neutralized phosphate group (NP). These permit a test of the SCC-DFTB parameters for phosphate. Corresponding trends are observed for the various calculations. The B3LYP and SCC-DFTB values generally agree within 0.05 Å for bond lengths (mostly less than 0.02 Å) and 1° for bond angles. The exception is that the C−O distances are overestimated by SCC-DFTB by about 0.06 Å in the two enediolate structures. We note that the phosphate SCC-DFTB parameters give satisfactory P−O distances compared to the B3LYP/6-31+G(d,p) values, with the largest error equal to 0.04 Å. The AM1, PM3, and SCC-DFTB methods give satisfactory energetics for the minimum structures, while only SCC-DFTB gives a reliable intramolecular barrier height; it is 12.9 kcal/mol compared to the value of 10.0 kcal/mol at the B3LYP/6-31+G(d,p) level.

The third set includes the substrate and one catalytic residue, either Glu165 or His95; see Scheme 1 for reactions involved. The residue included is the one involved explicitly in the proton transfer step. The covalent bond distances and angles at the SCC-DFTB level are in good agreement with the B3LYP/6-31+G(d,p) results, similar to the two smaller TIM models described above. For the hydrogen bond distances, the SCC-DFTB method gives shorter values for cases involving negatively charged atoms; e.g., the hydrogen bond between EDL2 and Glu165 is 1.914 Å at B3LYP/6-31+G(d,p) level and 1.830 Å at the SCC-DFTB level. Overall, the rms difference in hydrogen bond distances is 0.16 and 0.35 Å at the SCC-DFTB and PM3 levels, respectively, as compared to B3LYP/6-31+G(d,p) values. For all the critical geometrical parameters, the rms difference is 0.10 and 0.22 Å for SCC-DFTB and PM3, respectively, compared to B3LYP/6-31+G(d,p). The SCC-DFTB approach gives structures similar to the B3LYP results for most transition states. The few exceptions are shown in Figure 3. These include Inter_TS4_3b (the TS between EDT2_np_3b and GAP_3b) and Inter_TS2_3b (the TS between EDT1_3b and EDL1_3b), for which the SCC-DFTB calculations

SCHEME 1: Proposed Catalytic Mechanisms for TIM-Catalyzed Reactions



collapsed to EDT2_np_3b and EDL1_3b, respectively. This appears to be a consequence of the fact that barriers for these two steps are very small; they equal 1.9 and 0.5 kcal/mol, respectively, at the B3LYP/6-31+G(d,p) level (see Scheme 1 for the corresponding processes). The energetics at the SCC-DFTB level are in much better agreement with the B3LYP/6-31+G(d,p) results as compared with AM1 and PM3, especially for the barrier heights. Similar to the observation in the M_0 and M_{2NP} sets, the intramolecular barrier heights, measured at Intra_PT1_3b (the TS between EDT1_3b and EDT2_3b) and Intra_PT2_3b (the TS between EDT1_3_3b and EDT2_2_3b) are overestimated by more than 20 kcal/mol at both the AM1 and PM3 level; SCC-DFTB gives the value of 9.1 kcal/mol, which is rather close to the value of 4.8 kcal/mol at the B3LYP/6-31+G(d,p) level. Another example is the barrier height measured at Inter_TS5_3b (the TS between EDT1_2_3b and EDL2_3b) relative to EDT1_2_3b, for which the values are 19.1, 17.3, 6.7 and 4.4 kcal/mol, at the AM1, PM3, SCC-DFTB and B3LYP/6-31+G(d,p) levels, respectively. Overall, the rms error of AM1, PM3 and SCC-DFTB relative to B3LYP/6-31+G(d,p) is 11.1, 11.7 and 4.0 kcal/mol, indicating that SCC-DFTB is in much better agreement with B3LYP/6-31+G(d,p).

III.2. Tests of QM/MM Formulation. One would expect that the major difference between the different QM/MM results, where only the QM method is varied and the MM part of the calculations are done with the CHARMM22 force field, would arise from the accuracy of the former. In section III.1, we showed that SCC-DFTB gives results similar to B3LYP for reactions involving small molecules and transition states in the gas phase. Here we carry out SCC-DFTB/MM calculations to demonstrate that the environmental effect can be well captured with the approximate QM/MM interaction, eq 14. Calculations were carried out for a number of species in a set of model structures for TIM, which were studied previously by QM and QM/MM approaches with QM corresponding to AM1, AM1-SRP and B3LYP/6-31+G(d,p).[20] These structures include the model substrate, Glu165 and His95. As explained in ref 20, the structures were optimized with HF/3-21+G. Single point energies were calculated at a number of levels, including B3LYP/6-31+G(d,p), SCC-DFTB, B3LYP/6-31G(d,p)/CHARMM and SCC-DFTB/CHARMM. In the latter two QM/MM type of calculations, the QM part includes the substrate and Glu165. The His95 was treated with the CHARMM22 force field.[34] The rms errors are given in Table 3, the detailed values are given in the Supporting Information.

With full QM calculations, the SCC-DFTB results are in very good agreement with the B3LYP/6-31+G(d,p) results, as expected. This applies not only to the relative energetics of different minima, but also to the barrier heights. The rms difference between the two methods is 2.2 kcal/mol. The QM/MM calculations at both QM levels agree very well with the corresponding full QM results. The rms difference is 1.5 kcal/mol for B3LYP and 1.2 kcal/mol for SCC-DFTB. Compared to B3LYP/6-31+G(d,p), the SCC-DFTB/CHARMM calculations have a rms difference of 2.4 kcal/mol, similar to that of full SCC-DFTB; the corresponding error is 9.7 and 1.3 kcal/mol for AM1/CHARMM and AM1-SRP/CHARMM, respectively. Test calculations (results not shown) with the partial charges on the MM atoms zeroed out suggested that the effect of these MM atoms on the relative energetics range from 2 to nearly 10 kcal/mol.

Evidently, the present combined SCC-DFTB/CHARMM method works well with the two-center QM/MM interaction approximation in eq 14. The obtained results are not only close to full SCC-DFTB values, but also to those at the B3LYP level.

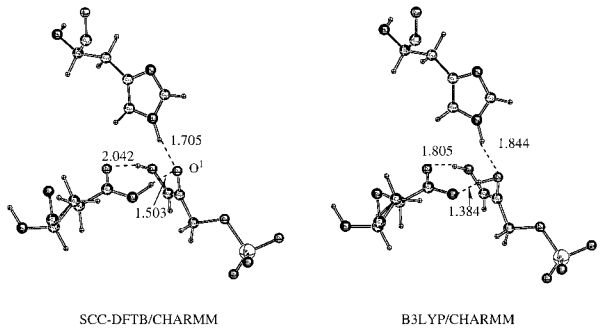


Figure 4. Active site structure of EDL2 involved in the TIM-catalyzed reactions obtained at the SCC-DFTB/CHARMM and B3LYP/6-31+G(d,p)/CHARMM level. The distances are in Å.

III.3. Reactions in Enzyme: The TIM-Catalyzed Reactions. In this subsection, we use the SCC-DFTB/CHARMM method to study the TIM-catalyzed reactions, and compare the results to these from previous studies[20] at the AM1/CHARMM, AM1-SRP/CHARMM, and B3LYP/6-31+G(d,p)/CHARMM levels. A number of different sets of calculations were carried out. In the first set, the same QM/MM partition (QM partition-I) as the previous work has been used. The QM region includes the substrate and the side chain of Glu165 in the reaction path scans for path B and C. In the calculations for path A, the QM region includes the substrate and the side chain of His95. In the second set of calculations, the substrate and the side chains of both Glu165 and His95 (QM partition-II) are treated with SCC-DFTB. Finally, B3LYP/6-31+G(d,p)/CHARMM single-point energies were carried out with the QM partition-I at the SCC-DFTB/CHARMM optimized structures using the QM partition-II. In the SCC-DFT/CHARMM calculations, the van der Waals parameters for the substrate atoms were taken from previous studies optimized for B3LYP/6-31+G(d,p). The choice is made because SCC-DFTB gave results rather similar to those at the B3LYP/6-31+G(d,p) level in the gas phase (section III.1), and that this set of van der Waals parameters yield good SCC-DFTB/CHARMM results for models in section III.2. To simplify the description of the results, only the rms errors in the energetics from the different calculations relative to the B3LYP/6-31+G(d,p)/CHARMM values are shown in Table 3; the actual values are tabulated in the Supporting Information.

The structural parameters are in good agreement with both AM1-SRP/CHARMM and B3LYP/6-31+G(d,p)/CHARMM results. The rms deviation in the geometries of active site atoms (including the substrate, Asn10, Lys12, His95, and Glu165) is 0.18 and 0.21 Å for SCC-DFTB/CHARMM and AM1-SRP/CHARMM, respectively, compared to B3LYP/6-31+G(d,p)/CHARMM values. As to the energetics, for most structures, the SCC-DFTB/CHARMM results depend little on the size of the QM partition. The rms error of SCC-DFTB/CHARMM results compared to B3LYP/6-31+G(d,p)/CHARMM values is 5.3 and 5.5 kcal/mol for QM partition I and II, respectively. These values are close to the rms error for AM1-SRP/CHARMM, 4.5 kcal/mol, and are smaller than the corresponding value for AM1/CHARMM, which is 8.4 kcal/mol. The single-point B3LYP/6-31+G(d,p)/CHARMM energies at the SCC-DFTB/CHARMM optimized structures using the QM partition-II are even closer to the full B3LYP/CHARMM results with a rms error of 2.5 kcal/mol. For EDL2, however, a significant difference between SCC-DFTB/CHARMM and B3LYP/CHARMM is observed, as illustrated in Figure 4. At the SCC-

DFTB/CHARMM level, the substrate in the optimized structure is actually in the enediolate form rather than the enediol form in the B3LYP/CHARMM structure. This difference is consistent with the observation that the stability of EDL2_3b in the gas phase M_{3b} is underestimated by about 6 kcal/mol at the SCC-DFTB level. In both the SCC-DFTB/CHARMM and B3LYP/CHARMM structures, however, the hydrogen bond distance between the side chain of Glu165 and the substrate is very short, 1.503 and 1.384 Å, respectively. This suggests that the proton transfer between the carboxylic oxygen in Glu165 and O^1 in the substrate is facile. Indeed, the structure optimized at the SCC-DFTB/CHARMM level with the substrate constrained in the enediol form was found to be only 2 kcal/mol higher than the fully optimized EDL2. Another difference concerns with the other enediol species, EDL1 in path A. As discussed in previous work, EDL1_A is not a stationary point at the B3LYP/6-31+G(d,p)/CHARMM level, and the proton transfer from the hydroxyl group back to the deprotonated His95 is barrierless. At the SCC-DFTB/CHARMM level, EDL1_A does exist as a stationary point. The barrier height for the subsequent proton transfer, however, is very low, 1.6 kcal/mol with the QM partition-II without zero-point corrections.

We briefly compare the speed of a number of QM/MM methods. For a single-point energy evaluation for TIM with the small QM partition, the B3LYP/6-31+G(d,p)/CHARMM calculation, which includes 339 basis functions, took about 4 CPU hours on a single node of an Origin 2000; a HF/CHARMM calculation with the same basis set was slightly faster and took approximately 3 CPU hours. By contrast, the SCC-DFTB/CHARMM calculation took about 4 s, and the AM1-SRP/CHARMM method took 10 s. As mentioned in section II, the SCC-DFTB is slightly faster than AM1 (or PM3) mainly because the matrix elements in the AO basis (e.g., eq 4) are tabulated in SCC-DFTB rather than computed, as in AM1 (or PM3). The remaining time-consuming step, which is the diagonalization of the Fock matrix and can be speeded up using methods such as the divide-and-conquer approach,[3,65] takes the same time.

We have shown that SCC-DFTB/CHARMM is a reasonably accurate method that can be applied to study the mechanism of enzymatic reactions. The results are in approximate agreement with the far more expensive DFT/CHARMM method, and significantly more accurate than the AM1/CHARMM approach, at least for the systems studied. To obtain more quantitative energetics, single-point energy DFT/CHARMM calculation at the SCC-DFTB/CHARMM optimized structure should be satisfactory.

III.4. Molecular Dynamics with QM/MM: Stability of Helices in Solution. In this subsection, we illustrate the usefulness of the SCC-DFTB/CHARMM approach by studying the stability of different helical conformations of small peptides in solution.[35] We consider two common types of helices in polypeptides: α_R-helix, characterized by $i,i+4$ C**O**$\cdots$**H**N hydrogen bonds and 3_{10}-helix, characterized by $i,i+4$ C**O**$\cdots$**H**N hydrogen bonds.[36] Although most helices in proteins adopt the α_R conformation, approximately 20% adopt the 3_{10} conformation;[37] among the 362 helices examined by Barlow and Thornton,[37] 71 were 3_{10}. The tightly pitched 3_{10} helices have a mean length of 3.3 residues, while the α_R helices are significantly longer with an average length of 12.2 residues. Some experimental studies have been interpreted as suggesting that 3_{10} helices are formed in early stages of protein and peptide unfolding.[38] This has also been observed in several theoretical studies.[39,40] Transitions between 3_{10} and α_R were found to be involved in domain motions of a number of proteins including aspartate transferase,[41] lactate dehydrogenase,[42] for which structures are available. For a number of designed alanine-rich polypeptides, which are good helix formers,[43] both double label electron spin resonance (ESR),[44−46] and solution[47]/solid-state[48] nuclear magnetic resonance (NMR) experiments have been used to probe the dominant helical conformation. In the double-label ESP experiment, two labeled molecules with unpaired electrons were attached at selected sites in the peptide, separated by one, two, or three residues. The observed ESR signals were related to the relative distances between the labels, which in turn, was used to characterize the backbone helical conformation.[44] On the basis of such a technique, Millhauser et al. concluded that short alanine-based peptide (capped hexa-alanine) favor the 3_{10} population,[46] while longer (>20) helices favor the α_R conformation.[49] For helices of intermediate length (16−17 residues), the result from ESR experiments was found to depend sensitively on the flexibility of the spin labels. Using pairs of nitroxide spin-labeled cysteins, which are rather flexible, a mix of α_R and 3_{10} conformations were detected in a series of 16-residue alanine-rich peptide.[44] However, with a more rigid label, Smythe et al.[45] found α_R to be the dominant conformation for a similar alanine-rich peptide in solution. Millhauser's group came to the similar conclusion with the same spin label.[46] Solution NMR and hydrogen exchange measurements were also performed by Millhauser et al. to probe the relative population of 3_{10} and α_R conformations of a 16-residue alanine based peptide.[47] A significant population of 3_{10} helices ($\sim$50%) was estimated near the termini based on NOE data (see discussion below). In the solid-state NMR study of Tycko et al.,[48] the (ϕ,ψ) angles of a similar 17-residue alanine-rich peptide were probed with two-dimensional NMR. The fractional population of α_R, 3_{10}, and random coils were estimated by analyzing the NMR data based on constrained molecular dynamics ($i,i+4$ and $i,i+3$ hydrogen bonds were constrained for the α_R and 3_{10}, respectively), which generated model (ϕ,ψ) angle distributions. The α_R helix was found to be the dominant conformation in glycerol/water at low temperature (−140 °C), and it converted to a 3_{10} helix upon the addition of urea.

From the above summary, the factors affecting the relative stabilities of α_R and 3_{10} helices appear to be complex, suggesting that simulations would be of interest. Molecular dynamics simulation studies using empirical force fields and explicit water solvent found that the α_R conformation was strongly favored for short peptides in solution. The α_R conformations of deca-alanine[50] and undecaalanine[51] peptides were found to be stable in the simulations, whereas 3_{10} conformation of the undecaalaine was found to convert spontaneously to the α_R form in less than 15 ps. Free energy simulations suggest that α_R is more stable than the 3_{10} conformation in water by about 16 kcal/mol for decaalanine[50] with the Ceder all-atom force field, and 10.6 kcal/mol for undecaalanine with the OPLS[52] force field. A similar value of 11.3 kcal/mol was found for a methyl-capped alanine decamer using the AMBER force field.[53] In vacuum, the α_R form was also found to be more stable than the 3_{10} form, although to a lesser extent, i.e., the AMBER estimates are 3.3 kcal/mol for the MeA decamer and 8.0 kcal/mol for the deca-alanine.[53] This means that the aqueous solvent stabilizes the α_R form more than the 3_{10} conformation. A strong dependence of the relative stability of the α_R and the 3_{10} form on the peptide size has been reported for MeA peptides in vacuum.[54] For seven residues, the α_R helix is 0.55 kcal/mol more stable than the 3_{10} helix; for a nonamer, the free energy difference is 1.1 kcal/mol and the potential energy difference is 5.3 kcal/mol favoring the

α_R conformation. The estimated entropic effect is comparable to the results of Zhang et al.[50] for decaalanine.

Interestingly, QM calculations seem to give the opposite trend for the relative stability of the two conformations in the gas phase. At the AM1 level,[55] the 3_{10} form was found to be more stable by about 1 kcal/mol for the alanine heptamer, and both conformations were found to be equally stable for the nonaalanine. With a dielectric continuum model, it was found that solvation favors the α_R over the 3_{10} form by 8.5 kcal/mol for heptaalanine. Recently, Elstner et al.[56] studied *N*-acetyl-(L-Alanyl)$_n$ *N*′-methylamide at the B3LYP/6-31G* level for n = 3, 5, 8, and 11, and at the SCC-DFTB level for n = 3, 5, 8, 11, 14, 17, and 20 in the absence of solvent. The intrinsic relative stabilities of the α_R and the 3_{10} conformation were estimated by fixing the (ϕ,ψ) angles at the corresponding ideal values, $(-60°,-30°)$ and $(-57°,-47°)$ for the 3_{10} and the α_R conformation, respectively. It was found that the 3_{10} form is more stable than the α_R conformation for all the helices studied with length up to n = 20. The relative energetics of α_R and the 3_{10} conformations at the SCC-DFTB and the B3LYP/6-31G* levels are rather similar, while the differences are smaller in magnitude at the AM1 level. For example, the 3_{10} conformation is more stable than the α_R form by 6.1 kcal/mol for n = 8 and 4.1 kcal/mol for n = 11 at the B3LYP/6-31G(d) level; the corresponding values are 6.1 and 5.3 kcal/mol for SCC-DFTB and 2.3 and 1.3 kcal/mol for AM1, respectively. Upon unconstrained minimization, it was found that the α_R helix was not stable until a peptide of 8 residues was reached, and that all the shorter helices adopted the 3_{10} conformation. For longer helices, a significant amount of 3_{10} conformation was found near the termini of the peptide.

The above results suggest that the empirical force fields overestimate the stability of the α_R conformation, relative to the 3_{10} conformation. This might explain the discrepancy in terms of the relative stability of the α_R and the 3_{10} conformations of short peptides between the MD studies employing empirical force fields[50,51,53] and experimental observations.[47] To see if this is the case, we chose to study the capped penta- and octaalanine peptide, *N*-acetyl-(L-alanyl)$_n$ *N*′-methylamide (n = 5, 8), with both CHARMM and SCC-DFTB/CHARMM methods. Minimizations were carried out in the gas phase and solution for both the α_R and 3_{10} conformations. Molecular dynamics were performed in water for the octaalanine peptide starting from the two helical conformations. The entire peptide was treated with SCC-DFTB in the SCC-DFTB/CHARMM simulations, and the water molecules are described with the standard TIP3P model,[32] which is used within the CHARMM force field.

To test the performance of the SCC-DFTB/CHARMM approach for the peptide−water system, a number of small systems were studied which include NMA (*N*-methylacetamide)−H_2O, NMA−NMA, and a dialanine peptide with four hydrogen-bonded water molecules (Figure 5). At the ab initio level, the structures were minimized with B3LYP/6-31G(d,p), and the interaction energies were calculated with single point B3LYP/6-311++G(d,p). Calculations were also carried out with SCC-DFTB, CHARMM, and combined SCC-DFTB/CHARMM; water molecules were described with the TIP3P model in the SCC-DFTB/CHARMM calculations. For the NMA complexes, the SCC-DFTB approach on average gave shorter (by 0.05−0.1 Å) hydrogen bond distances compared to the B3LYP/6-31G(d,p) values, and both SCC-DFTB/CHARMM and CHARMM calculations gave even shorter values (by 0.1−0.2 Å). The latter is in accord with the CHARMM parametrization[34] and appears appropriate for obtaining the correct solution properties. The

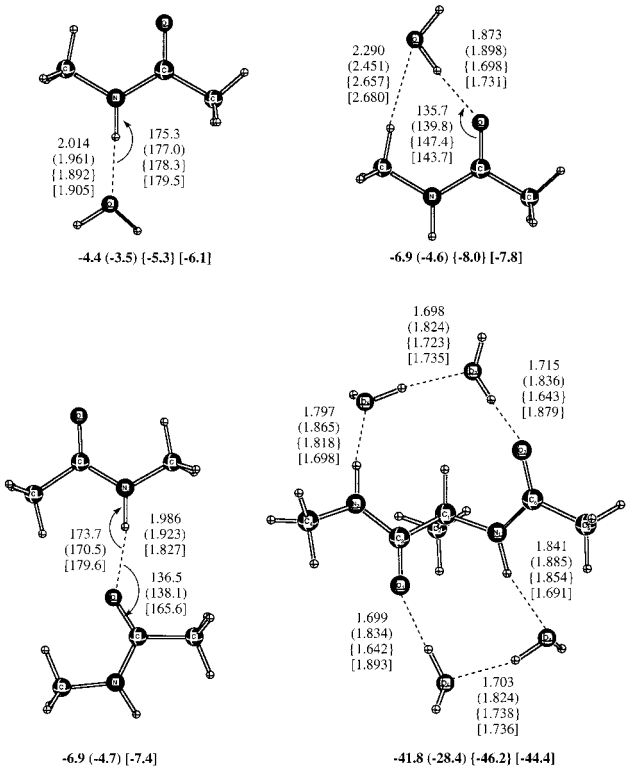


Figure 5. Optimized structure for NMA-H_2O, NMA-NMA, and a dialanine peptide interacting with four water molecules. The geometrical parameters (distances in Å, angles in degrees) without parentheses were obtained at the B3LYP/6-31G(d,p) level; the binding energies (in kcal/mol, in bold) were obtained with B3LYP/6-311++G(d,p) single-point calculations. The values with parentheses were obtained from SCC-DFTB calculations; those with curled parentheses were obtained with the SCC-DFTB/CHARMM calculations in which the water molecules were treated withTIP3P calculations. The numbers in brackets were calculated with the CHARMM22 force field.

hydrogen-bonding energies were underestimated by 1−2 kcal/mol at the SCC-DFTB level, a trend that was also observed in a more systematic study on the performance of SCC-DFTB for hydrogen bonding interactions.[17] Both SCC-DFTB/CHARMM and CHARMM overestimated hydrogen-bonding energies compared to the B3LYP/6-311++G(d,p) values, and the differences are about 1 kcal/mol; again this is consistent with the CHARMM parametrization, which assumes that in solution the mean dipole moment of the water molecules is increased relative to the gas phase.[34] For the dipeptide−water system, it is interesting to observe that the hydrogen bond distances at the B3LYP level are much shorter than those in the NMA−water complexes. Such a cooperative effect due to hydrogen-bonded networks has been observed previously.[57] The effect was qualitatively described in both the SCC-DFTB and SCC-DFTB/CHARMM levels, as reflected by the shortening of the hydrogen bond distances compared to the NMA−water complexes. The binding energy of the four water molecules to the dipeptide is 28.4 kcal/mol at the SCC-DFTB level, this is a significant underestimate by about 2 kcal/mol per hydrogen bond, including two water−water hydrogen bonds, compared to the binding energy of 41.8 kcal/mol at the B3LYP/6-311++G(d,p) level. Encouragingly, both SCC-DFTB/CHARMM (46.2 kcal/mol) and CHARMM (44.4 kcal/mol) gave binding energies that are very close to the B3LYP/6-311++G(d,p) value.

By optimizations in the gas phase, an initial geometry was obtained, starting with constrained ideal dihedral angles of $(-60°,-30°)$ and $(-57°,-47°)$ for (ϕ,ψ) angles in the 3_{10} and the α_R conformations, respectively. These dihedral angles were

TABLE 4: Dihedral Angles (ϕ,ψ) in Different Helices Obtained with SCC-DFTB/CHARMM and CHARMM[a]

	5−3_{10}	5-α_R	5-α_{R_MM}	8−3_{10}	8-α_R	8-α_{R_MM}
ϕ1	−48.2 (−39.5)	(−54.4)	−55.1 (−61.1)	−46.5 (−52.1) [−55.3/19.4]	−43.4 (−62.0) [−59.5/20.9]	−55.9 (−72.7) [−73.1/18.9]
ψ1	−40.0 (−59.8)	(−51.2)	−43.7 (−60.9)	−42.2 (−34.4) [−49.6/11.0]	−52.2 (−42.6) [−52.0/9.9]	−45.4 (−39.8) [−48.1/11.4]
ϕ2	−53.0 (−71.4)	(−73.7)	−64.4 (−65.9)	−49.3 (−52.9) [−58.3/10.8]	−57.1 (−61.3) [−57.1/10.6]	−64.1 (−65.5) [−68.8/10.7]
ψ2	−30.9 (−8.9)	(−20.0)	−36.4 (−29.0)	−34.9 (−32.0) [−41.7/10.5]	−40.8 (−41.7) [−45.6/9.9]	−38.8 (−38.5) [−36.1/11.1]
ϕ3	−56.8 (−63.6)	(−75.0)	−74.9 (−65.7)	−51.8 (−59.8) [−54.7/12.1]	−64.6 (−58.4) [−58.6/11.3]	−69.3 (−65.3) [−63.0/11.0]
ψ3	−27.2 (−26.1)	(−47.6)	−36.7 (−46.3)	−32.9 (−23.5) [−41.2/9.9]	−41.3 (−46.7) [−40.5/13.1]	−43.1 (−44.8) [−45.0/9.7]
ϕ4	−66.5 (−70.9)	(−61.1)	−68.5 (−69.1)	−54.4 (−63.5) [−60.8/10.8]	−52.8 (−60.6) [−59.7/13.2]	−60.5 (−62.8) [−66.1/9.4]
ψ4	−5.0 (−5.8)	(−27.1)	−37.6 (−43.6)	−28.0 (−25.8) [−43.0/9.9]	−47.5 (−25.2) [−44.6/8.4]	−45.2 (−33.9) [−39.5/9.9]
ϕ5	−89.4 (−96.0)	(−106.5)	−74.6 (−73.0)	−58.2 (−62.1) [−57.1/11.3]	−54.5 (−83.6) [−59.6/10.7]	−61.3 (−71.4) [−65.2/9.7]
ψ5	6.0 (7.3)	(−33.8)	−44.4 (−46.0)	−26.4 (−16.9) [−44.3/8.6]	−46.1 (−23.2) [−42.1/9.6]	−44.8 (−37.3) [−40.9/9.8]
ϕ6				−61.4 (−84.2) [−58.8/10.3]	−55.3 (−78.3) [−59.8/10.8]	−64.1 (−68.8) [−67.6/9.4]
ψ6				−23.0 (−2.8) [−36.9/10.8]	−35.0 (−42.1) [−44.7/10.5]	−39.3 (−42.8) [−37.5/11.1]
ϕ7				−72.0 (−84.9) [−57.5/12.7]	−90.6 (−72.8) [−59.2/13.1]	−69.8 (−69.0) [−70.0/10.6]
ψ7				1.0 (7.9) [−36.5/12.2]	0.5 (−12.4) [−34.6/11.7]	−36.9 (−24.7) [−46.0/15.6]
ϕ8				−97.6 (−78.2) [−69.4/19.7]	−118.7 (−86.3) [−68.0/14.7]	−73.9 (−74.1) [−75.6/13.8]
ψ8				7.3 (−7.1) [−37.8/13.3]	7.4 (−52.0) [−38.9/13.6]	−43.0 (−53.5) [−51.6/16.2]

[a] The ideal dihedral angles are (−60.0°,−30.0°) and (−57.0°,−47.0°) for the 3_{10} and α_R conformation, respectively. The numbers without parentheses are gas-phase values, and those with parentheses are the values optimized in water. The numbers in brackets are the averaged values and fluctuations along the trajectories in water. The "5" and "8" in the labels indicates the length of the peptide. The numbers with "MM" are from the CHARMM calculations, and the others are from the SCC-DFTB/CHARMM simulations.

not constrained in the subsequent 200 steps of minimization with the adapted basis Newton−Rhapson (ABNR) algorithm; the resulting values are summarized in Table 4. For the optimization in water, the peptide with the gas-phase geometry was solvated by a 16 Å water sphere with a deformable stochastic boundary potential.[58] In cases where the gas-phase minimized structure went to the other conformation (e.g., α_R penta-ala, which went to 3_{10} at the SCC-DFTB level), the ideal (ϕ,ψ) angles were used. The water molecules were first relaxed using 200 steps of steepest descent minimization with the peptide held fixed, and then the entire system was minimized with 400 steps of ABNR. In the molecular dynamics simulations, a time step of 2 fs was used with all the bonds involving hydrogen constrained with the SHAKE algorithm.[59] Water molecules beyond 12 Å from the center of the sphere were treated with the Langevian dynamics and Newtonian dynamics was used for all other atoms; the LD/MD boundary was updated every 25 integration steps. A nonbonded cutoff of 13.0 Å was used, and the nonbonded list is updated with a heuristic algorithm in CHARMM.[60] The initial temperature of the system was set to 100 K, which was raised to 300 K by increasing temperature of the Langevian bath during the 15 ps of equilibration. A production run was then carried out for another 15 ps at the SCC-DFTB/CHARMM level and 40 ps at the full CHARMM level. Even at the SCC-DFTB/CHARMM level, the present calculations are rather expensive since 92 atoms are treated quantum mechanically; i.e., 1 ps of molecular dynamics takes about 2 h of CPU time on a single node of a SGI Origin2000 machine. In the calculations, converged orbital coefficients were saved after each iteration and used as starting values for the subsequent calculations. The convergence criterion for SCF is 10^{-8} hartree in energy. To compare with the NMR experiments, NOE intensity ratios were computed for different helical structures. As pointed out in the literature (see also discussion below),[47,61] the α_R and 3_{10} helices can be characterized by the ratio of two long-range NOEs, $I_{\alpha\beta(i,i+3)}/I_{\alpha N(i,i+3)}$, which is given by

$$\frac{I_{\alpha\beta(i,i+3)}}{I_{\alpha N(i,i+3)}} = \frac{R^6_{\alpha N(i,i+3)}}{\min\{R^6_{\alpha\beta(i,i+3)}\}} \qquad (17)$$

where $R_{\alpha N(i,i+3)}$ ($R_{\alpha\beta(i,i+3)}$) is the distance between the Cα proton in the ith residue and the amide (Cβ) proton in the (i+3)th residue.

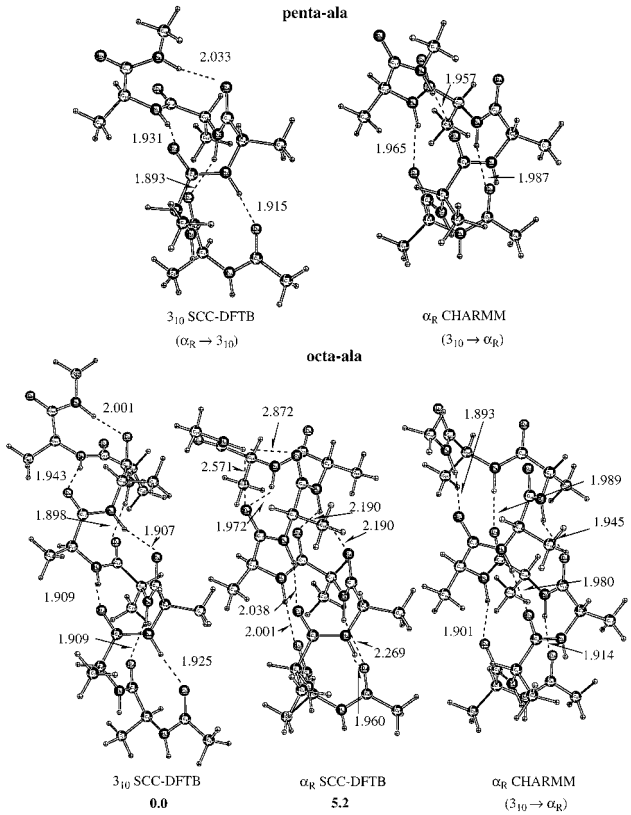


Figure 6. Optimized structures for capped pentaalanine and octaalanine peptides in the gas phase with SCC-DFTB and CHARMM calculations. The distances are in Å.

Optimized structures in the gas phase and water are shown in Figures 6 and 7, respectively, and the (ϕ,ψ) angles are listed in Table 4. Calculated NOE ratios are listed in Table 5. In Figure 8, the end-to-end distance of the helix (defined as the distance between the methyl carbon atom in the capped N- and C-termini), the i,i+3 (C**O**···**H**N) and i,i+4 (C**O**···**H**N) distances are shown. Distributions of i,i+3 and i,i+4 C**O**···**H**N distances along the trajectory for the α_R conformation are shown in Figures 9 and 10, respectively. The distributions of selected i,i+3 and i,i+4 C**O**···**H**N angles are shown in Figure 11.

In the gas phase, there is a significant difference in the stability of the α_R and 3_{10} conformation between SCC-DFTB and CHARMM (Figure 6). The α_R conformation for the capped

TABLE 5: Computed Long-Range (*i*,*i*+3) NOE Intensity for Residues in Different Helices Obtained with SCC-DFTB/CHARMM and CHARMM[a]

resi	$5-3_{10}$	$5-\alpha_R$	$5-\alpha_{R\_MM}$	$8-3_{10}$	$8-\alpha_R$	$8-\alpha_{R\_MM}$
1	0.7 (0.8)	(3.1)	5.9 (3.0)	0.6 (0.5) [4.4/3.4]	3.5 (3.2) [5.5/3.4]	6.6 (4.3) [3.5/2.6]
2	0.7 (1.6)	(7.8)	4.4 (9.9)	0.6 (0.9) [2.9/2.1]	2.9 (2.4) [3.8/2.9]	6.1 (4.3) [6.1/3.9]
3				0.5 (0.8) [4.7/3.3]	2.8 (2.5) [4.3/3.2]	5.2 (5.0) [4.3/2.7]
4				0.5 (0.6) [3.9/3.2]	2.1 (3.7) [4.9/3.1]	5.0 (5.7) [5.1/3.1]
5				0.7 (0.6) [2.4/2.2]	3.3 (4.9) [4.0/3.0]	4.5 (5.1) [3.2/2.9]

[a] The ratios were calculated using eq 17 in the text. The notations are the same as those in Table 4.

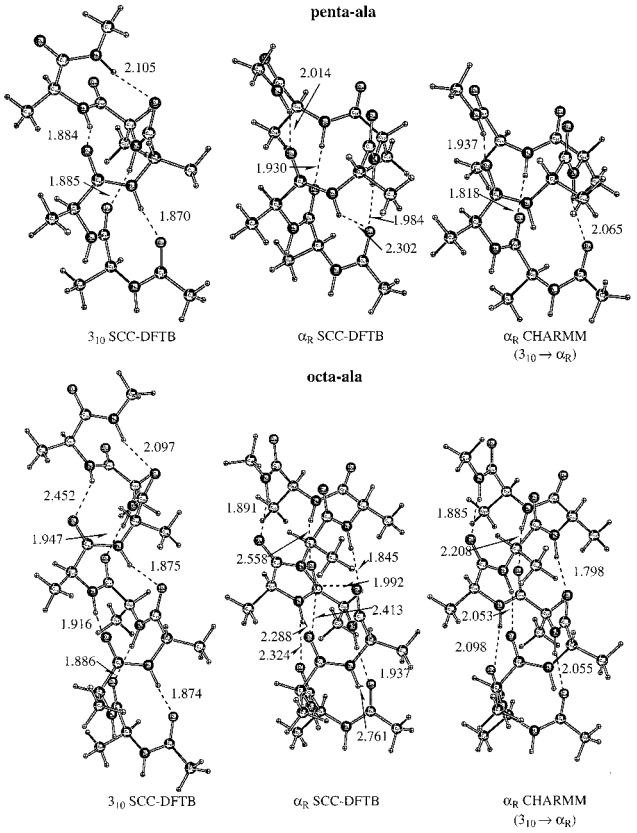


Figure 7. One set of optimized structures for capped pentaalanine and octaalanine peptides in a 16 Å water sphere with SCC-DFTB and CHARMM calculations starting from the ideal (ϕ,ψ) angles. The distances are in Å.

pentaalanine is not stable at the SCC-DFTB level, and optimization yields a 3_{10} conformation, which is shown in Figure 6. With the CHARMM force field, by contrast, the 3_{10} conformation is not stable and is converted to the α_R conformation during geometry optimization (Figure 6). For the octaalanine, both the α_R and 3_{10} conformations are stable at the SCC-DFTB level, with the latter more stable by 5.2 kcal/mol. At the full CHARMM level, however, the 3_{10} conformation is not stable and is converted to the α_R conformation upon geometry optimization. In the α_R structure at the SCC-DFTB level, there are *i*,*i*+3 hydrogen bonds at both the N- and C-termini, which are not present in the CHARMM optimized structure. Correspondingly, the (ϕ,ψ) angles (see Table 4) for the α_R conformation octa-alanine optimized at the SCC-DFTB and CHARMM levels also differ more at the N,C-termini than in the middle. This suggests the CHARMM force field overestimates the stability of the α_R over the 3_{10} conformation in the gas phase.

The computer NOE ratios (see Table 5) for the 3_{10} conformers range from 0.5 to 0.7, which are close to the values reported in the literature.[47,62] The NOE ratios for the α_R helices show larger variations, ranging from 2.1 to 6.6; previous molecular dynamics simulations gave values ranging from 2.7 to 4.4.[63] The estimated experimental values ranges from 4.0 to 5.6.[47,64] It should be noted that due to the weak intensities of those long-range NOEs, it is very difficult to determine the experimental NOE ratio up to high accuracy. The NOE ratios from CHARMM minimized structure are somewhat larger than the values obtained using the SCC-DFTB structure. This is due to the fact that $R_{\alpha N(i,i+3)}$ is longer and $R_{\alpha\beta(i,i+3)}$ is shorter in the CHARMM minimized structure.

In water, the α_R conformation is stabilized over the 3_{10} conformation according to previous theoretical analyses.[53] This is due partly to the fact that the α_R conformation has on average one more free C═O and N−H to form hydrogen bonds with the water environment. Thus, only the α_R conformations are found to be stable with CHARMM for both capped penta- and octaalanines, while both α_R and 3_{10} are found to be stable with SCC-DFTB/CHARMM. In the SCC-DFTB/CHARMM optimized α_R structures, some bifurcated *i*,*i*+3/*i*,*i*+4 hydrogen bonds occur for both the penta- and octaalanine; neither is present in the CHARMM structures (Figure 7). Compared to the gas-phase structures, larger variations in the (ϕ,ψ) angles occur in the α_R conformation, especially at the N,C-termini (see Table 4). Correspondingly, the NOE ratios also show larger variations in the α_R conformations upon solvation. The trend is that the NOE ratio decreases near the N terminus but increases near the C terminus. This is due to the fact that upon solvation, $R_{\alpha N(i,i+3)}$ decreases more at the N terminus while $R_{\alpha\beta(i,i+3)}$ decreases more at the C terminus.

In the SCC-DFTB/CHARMM dynamics starting with the 3_{10} conformation of the capped octaalanine, the end to end distance is stable around 15.5 Å for about a few picoseconds (not including equilibration), and then decreases quickly to about 14 Å. With the ideal dihedral angles, the end-to-end distances in the α_R and 3_{10} conformation is 17.2 and 14.3 Å, respectively. At the same time, two *i*,*i*+4 C**O**···**H**N distances decreased from about 4 Å to about 2 Å. The rest *i*,*i*+4 C**O**···**H**N distances were already around 2.5 Å in the beginning of the production run. The averaged (ϕ,ψ) angles also approached the values in the trajectory for the α_R conformation. These suggest a transition from the 3_{10} to α_R conformation, which indicates that 3_{10} conformation is less stable than α_R for octaalanine in water, and that there is no significant barrier for the transition from 3_{10} to α_R.

During molecular dynamics, the α_R conformation is stable at the CHARMM level in agreement with previous simulations.[51,53] The *i*,*i*+4 C**O**···**H**N distances fluctuate around 2.1 Å, while the *i*,*i*+3 C**O**···**H**N distances fluctuate about 3.1 Å, and the end-to-end distance fluctuates about 14 Å. The averaged (ϕ,ψ) angles are similar to the optimized values, and the fluctuations are about 10°, except for the termini which have larger fluctuations. The *i*,*i*+4 C**O**···**H**N angles have similar averaged values around 140°, while the *i*,*i*+3 C**O**···**H**N angles average about 80°−90°. These results suggest that the peptide conformation is far from 3_{10} throughout the dynamics. Similarly, the α_R conformation at the SCC-DFTB/CHARMM level is

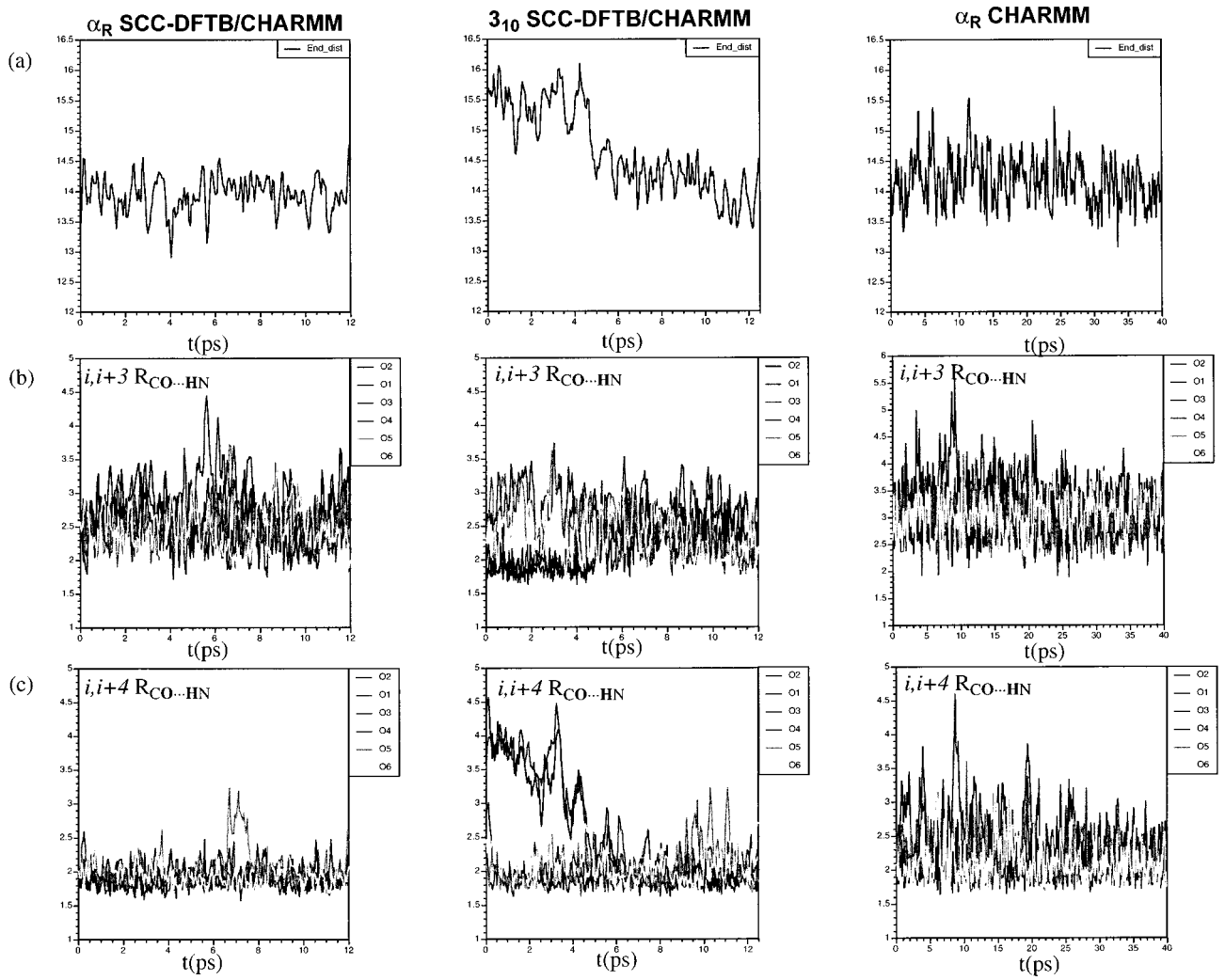


Figure 8. Properties of a capped octaalanine peptide along molecular dynamics simulations at SCC-DFTB/CHARMM and pure CHARMM levels. The equilibration stage is not included, which is 10 ps for SCC-DFTB/CHARMM and 20 ps for pure CHARMM. In the SCC-DFTB/CHARMM simulations, the peptide is treated with SCC-DFTB and the water molecules are described with the TIP3P model. Simulations starting from optimized α_R and 3_{10} conformations (Figure 7) were carried out at the SCC-DFTB level, while only one simulation starting from the α_R conformation was done with the pure CHARMM force field. (a) End to end distance of the peptide measured from the methyl carbon atoms in the N and C termini. (b) The distance between the carbonyl oxygen and the amide proton two residues away (*i*,*i*+3 C**O···H**N). (c) The distance between the carbonyl oxygen and the amide proton three residues away (*i*,*i*+4 C**O···H**N).

stable during the MD simulation, although with somewhat smaller magnitude of fluctuations for the *i*,*i*+4 C**O···H**N distances. For the α_R conformation, the SCC-DFTB/CHARMM simulation generally gave shorter *i*,*i*+4 and *i*,*i*+3 C**O···H**N distances in comparison with the CHARMM calculations, as shown in the distribution plots in Figures 9 and 10. The difference is more dramatic for the *i*,*i*+3 C**O···H**N distances, which have an average value of 2.58 Å at the SCC-DFTB/CHARMM level, in contrast to the value of 3.13 Å with the CHARMM force field. All the oxygen atoms, particularly these at the end of peptide, spend considerable time within 2.25 Å of the amide proton two residues away (*i*,*i*+3 interaction) in the SCC-DFTB/CHARMM simulations; by contrast, they are rarely within 2.5 Å of the $i + 3$ amide protons in the CHARMM calculations. In addition, the SCC-DFTB/CHARMM and CHARMM simulations gave significantly different distribution for both the *i*,*i*+3 C**O···H**N and the *i*,*i*+3 **CO···H**N angles at the termini, but yielded similar results for the middle of the peptide (Figure 11). The *i*,*i*+3 C**O···H**N angles at the termini in the SCC-DFTB/CHARMM calculation were often found in the range over 100°, while this rarely occurred in the CHARMM simulation. At the same time, the average (ϕ,ψ) angles at the SCC-DFTB/CHARMM level differ somewhat from those in the CHARMM calculations, especially around the termini. For instance, ψ7 and ψ8 at the SCC-DFTB/CHARMM level is −34.6° and −38.9°, respectively; the corresponding values from the CHARMM calculation is −46.0° and −51.6°, respectively. The SCC-DFTB/CHARMM results are closer to the ideal values of −30.0° in the 3_{10} conformation, while the CHARMM values are closer to the ideal value of −47.0° in the α_R conformation. In Figure 12, the averaged structures from SCC-DFTB/CHARMM and CHARMM MD simulations are shown. It is seen that the difference mainly occurs at the C-termini, for which SCC-DFTB/CHARMM gives an *i*,*i*+3 hydrogen bond and CHARMM gives the *i*,*i*+4 hydrogen bond expected for α_R conformation.

The calculation results can be compared with the NMR study of Milhauser et al.[47] They studied the NOE pattern of two 16-residue alanine-rich peptides, Ac-AAAAKAAAAKAAAAKA-NH_2 (3K) and Ac-AMAAKAWAAKAAAARA-NH_2 (MW). Although the $C_\alpha H(i)$ and NH(*i*+3) proton distances are very similar (3.3−3.4 Å) in the 3_{10} and α_R helices, the distances between $C_\alpha H(i)$ and $C_\beta H(i+3)$ protons are different for the two conformers; they range from 3.1 to 5.1 Å and 2.4 to 4.4 Å for

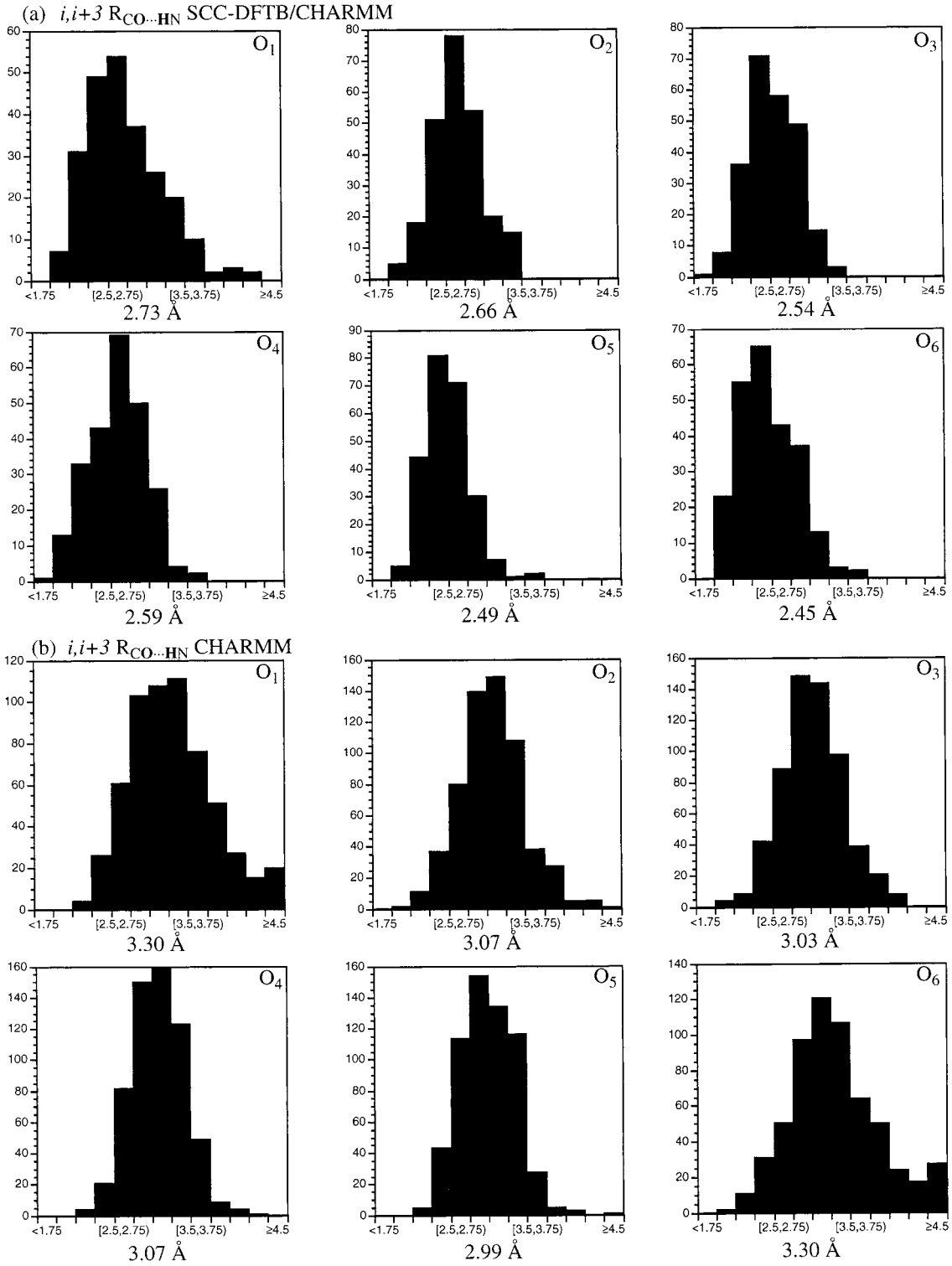


Figure 9. Distribution of *i*,*i*+3 C**O**···**H**N distances (in Å) along trajectories starting from the α_R conformation at he SCC-DFTB/CHARMM and pure CHARMM levels (see Figure 8). The values under the plots are the average distances.

the 3_{10} and α_R structures, respectively.[61] Therefore, the relative intensities, *I*, of the $\alpha N(i,i+3)$ and $\alpha\beta(i,i+3)$ NOEs can be related to the relative population of the 3_{10} and α_R helices. Using the $I_{\alpha\beta(i,i+3)}/I_{\alpha N(i,i+3)}$ ratios of 4.0 and 0.67 for α_R and 3_{10} helix, respectively, calculated for the ideal structures, it was estimated that the lower bound of the 3_{10} population is approximately 50% at the termini and 25% in the middle of the peptide. It was observed through hydrogen−deuterium exchange kinetics measurement that the third amide protons of the 3K peptide is partially protected from exchange, which also suggests that this proton is involved in the *i*,*i*+3 hydrogen bonding with the carbonyl of the acetyl capping group; it would not be protected if this region were α helical. These observations seem to be more consistent with the SCC-DFTB/CHARMM calculations, which found 3_{10} characters (short *i*,*i*+3 C**O**···**H**N distances and smaller ψ values) around the helical termini.

The computed NOE ratios (see Table 5) along the trajectories show considerable fluctuations (on the order of 3) in both SCC-DFTB/CHARMM and CHARMM simulations. The values computed from the trajectory starting from the 3_{10} conformer

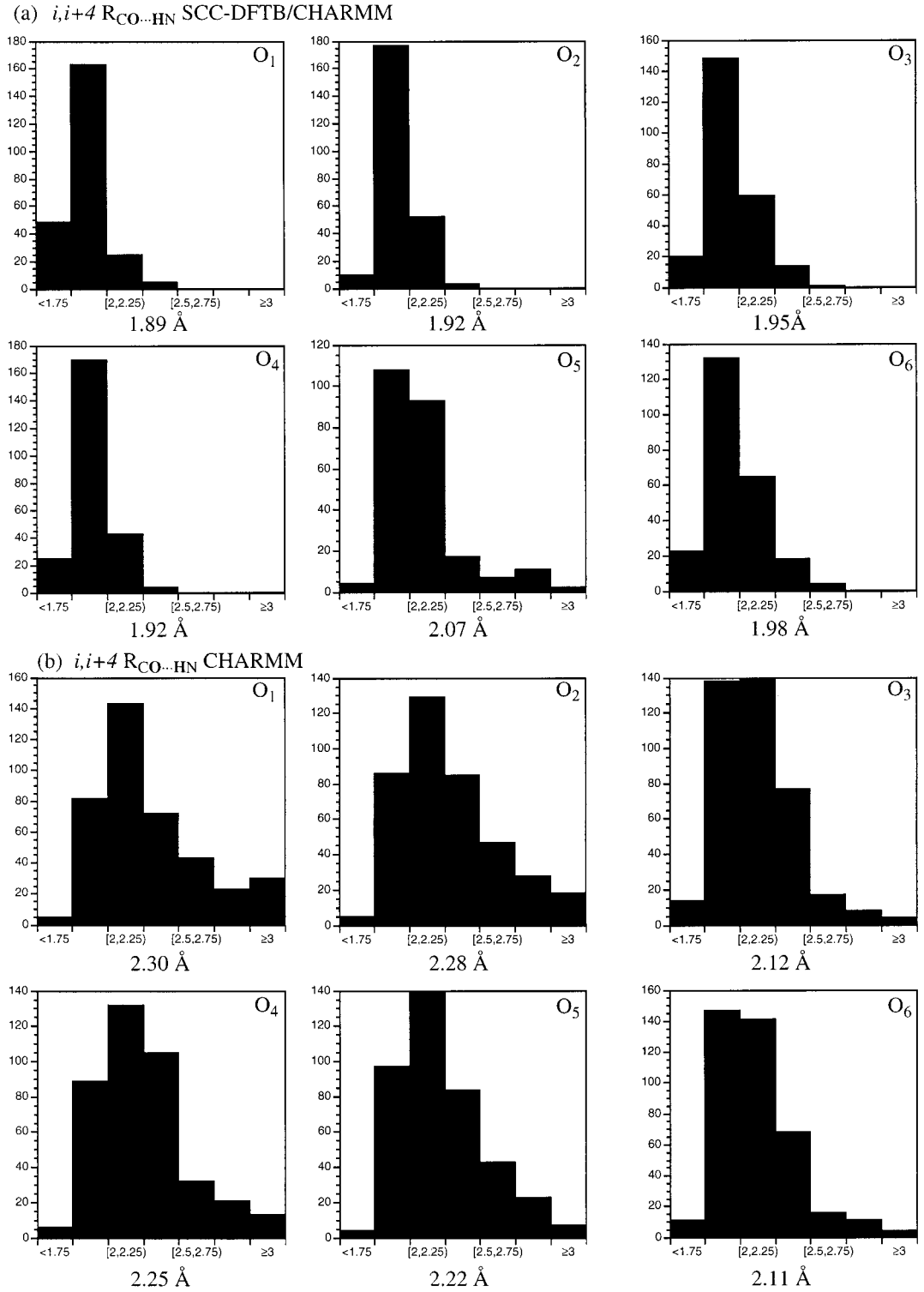


Figure 10. Distribution of *i,i*+4 C**O**···**H**N distances (in Å) along trajectories starting from the α_R conformation at the SCC-DFTB/CHARMM and pure CHARMM levels (see Figure 8). The values under the plots are the average distances.

are generally smaller than those from the α_R form, because the helix was in the 3_{10} conformation in the early stage of the simulation. For the α_R octaalanine, the average values are similar for the two simulation methods, especially for $i = 3$, 4, and 5. For the two residues close to the N-terminus, there are somewhat larger differences between the SCC-DFTB/CHARMM and CHARMM results. For $i = 1$, the SCC-DFTB/CHARMM value is larger, while for $i = 2$, the CHARMM value is considerably larger. Nevertheless, all the average values are much larger than that for a typical 3_{10} conformation observed above (~0.7). This also applies to most instantaneous values along the trajectory. Therefore, the 3_{10} characters (short *i,i*+3 C**O**···**H**N distances and smaller ψ values) observed in the SCC-DFTB/CHARMM simulations could not be seen clearly from the NOE data. It appears that molecular motions have considerable influence on the resolution of the long-range NOE ratio for distinguishing the two helical forms. A similar conclusion concerning the effect of dynamics on the quantitative interpretation of the $\alpha\beta(i,i+3)$ NOEs' was obtained in a previous molecular dynamics simulation on a ribonuclease C-peptide analogue.[63]

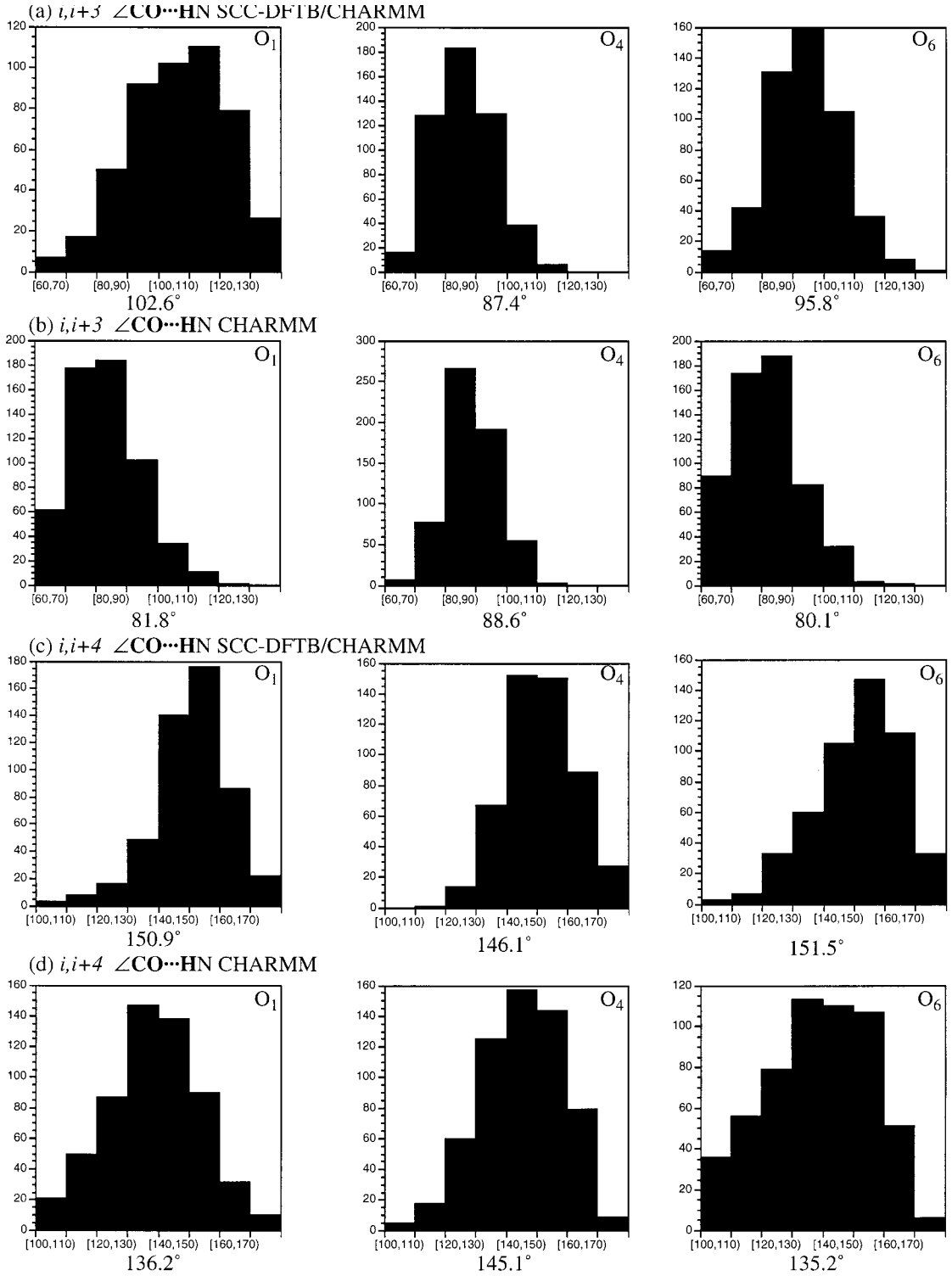


Figure 11. Distributions of selected *i,i*+3 C**O···H**N and *i,i*+4 C**O···H**N angles (in degrees) at the termini and in the middle of the peptide along trajectories starting from the α_R conformation at the SCC-DFTB/CHARMM and pure CHARMM levels (see Figure 8). The values under the plots are the average angles.

To further test the SCC-DFTB/CHARMM model, we carried out single-point calculations with a larger SCC-DFTB partition for 100 structures taken from the trajectories. The peptide and water molecules within 2.4 Å of the peptide were treated with SCC-DFTB, and the remainders of the water molecules were described with the TIP3P model. The energies relative to the starting geometry are plotted in Figure 13, against the corresponding values from the calculations with only the peptide treated with SCC-DFTB. Good correlation was found between the two sets of results. Therefore, we believe that the behavior found in the current SCC-DFTB/CHARMM-MD calculations is at least qualitatively correct.

Although CHARMM and SCC-DFTB behave differently in the gas phase (the latter gave results in approximate agreement with B3LYP calculations[56]), the qualitative behavior of polyalanine peptide in water is the similar with the two methods. Since the CHARMM22 force field was developed to reproduce properties of molecules in the solution rather than gas phase,[34] the result supports the reliability of the SCC-DFTB/CHARMM model. The current calculation is in accord with the earlier

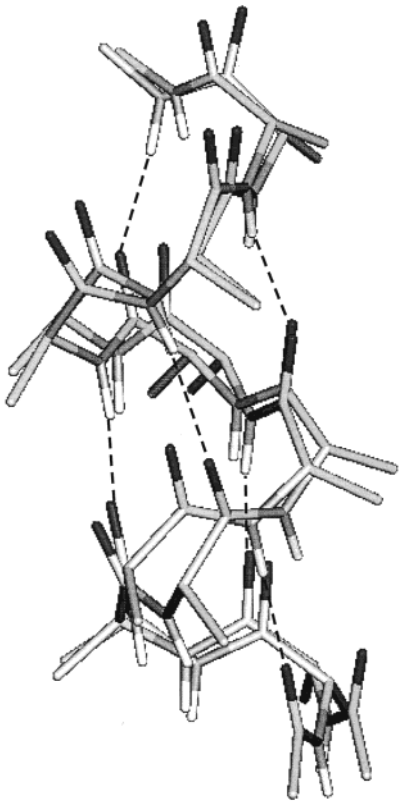


Figure 12. Comparison of averaged α_R octaalanine from SCC-DFTB/CHARMM and CHARMM molecular dynamics calculations (see Figure 8).

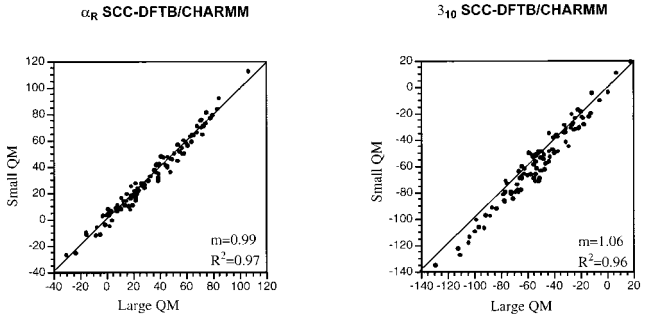


Figure 13. Comparison of single-point energetics for configurations along the trajectories with different SCC-DFTB partition. The small partition is the one used to propagate the trajectory and treats only the peptide with SCC-DFTB. The large partition treats also the solvent molecules with 2.4 Å from the peptide with SCC-DFTB.

findings that 3_{10} is not a stable conformation for short alanine peptide in solution.[51,53] However, the *i*,*i*+3 C**O**···**H**N distances, particularly these involving the termini groups, were found to be less than 2.25 Å for a significant fraction of the time along the trajectory at the SCC-DFTB/CHARMM level. At the same time, the average $\psi 7$ and $\psi 8$ values are close to the ideal values for the 3_{10} conformation. At the CHARMM level, however, the *i*,*i*+3 C**O**···**H**N distances are always far beyond typical hydrogen bond distances and the average $\psi 7$ and $\psi 8$ values are close to the ideal values for the α_R conformation. The SCC-DFTB/CHARMM results apparently are consistent with the NMR study of Milhauser et al.,[47] who proposed significant 3_{10} populations in alanine-rich peptides, particularly at the termini, based on NOE data. However, we also found that the molecular motions have considerable influence on the long-range NOE ratio and its significance for distinguishing the two helical forms. This confirms that is necessary to be very cautious in applying the approach to estimate the populations of helices.[63]

IV. Conclusions

To extend the utility of QM/MM methods, it is necessary to improve their accuracy without significantly increasing the CPU cost. In the present work, we describe the implementation of a QM/CHARMM approach based on the self-consistent charge density functional tight binding (SCC-DFTB) as the QM method in the CHARMM program and show that it satisfies this condition. The SCC-DFTB approach is a DFT based tight-binding model, which differs from standard tight-binding methods in that charge variations are introduced through a self-consistent treatment at the monopole level. The method is similar to AM1 and PM3 in speed and generally gives more accurate results, as compared with high-level DFT calculations, in the applications that have been presented here; i.e., the current work indicates that SCC-DFTB works better than AM1 and PM3 in most cases for the TIM models, relative to B3LYP with double-ζ plus polarization functions and is comparable with a version of AM1 optimized for the system.[20] However, there are exceptions; e.g., one of the enediol species (EDL2) was found to be too high in energy. This means that the user should test the SCC-DFTB method by comparing the results on model systems with high-level ab initio or DFT values for each type of application.

The SCC-DFTB and SCC-DFTB/CHARMM methods were tested on several systems of biological interest. In the gas phase, SCC-DFTB gives reliable energetics for models of TIM-catalyzed reactions. The rms errors in the energetics compared to B3LYP/6-31+G(d,p) are about 2−4 kcal/mol; the corresponding values for AM1 and PM3 are 9−11 kcal/mol. The method also gives accurate vibrational frequencies. For the TIM reactions in the presence of the enzyme, the rms error of the SCC-DFTB/CHARMM energies compared to B3LYP/6-31+G(d,p)/CHARMM values is 5.4 kcal/mol, comparative with a version of AM1 optimized for TIM.[20] Single-point B3LYP/CHARMM calculations at the SCC-DFTB/CHARMM optimized structures were found to be very similar to the full B3LYP/CHARMM values. The relative stabilities of the α_R and 3_{10} conformations of penta- and octaalanine peptides were studied with minimizations and molecular dynamics simulations in vacuum and in solution. Although CHARMM and SCC-DFTB gave qualitatively different results in the gas phase (the latter were in approximate agreement with B3LYP calculations[56]), similar behavior in aqueous solution was found for the CHARMM and SCC-DFTB/CHARMM simulations. This is in accord with the CHARMM parametrization philosophy, which is based on values that reproduce condensed phase results.[34] The 3_{10} conformations was not found to be stable in solution and converted to the α_R form in about 15 ps. The α_R conformation was stable in MD simulations with both SCC-DFTB/CHARMM and the CHARMM force field. The *i*,*i*+3 and *i*,*i*+4 C**O**···**H**N distances were significantly shorter with the SCC-DFTB/CHARMM approach. The effect is more dramatic for the *i*,*i*+3 C**O**···**H**N distances, which have an average value of 2.58 Å at the SCC-DFTB/CHARMM level in contrast to the value of 3.13 Å with the CHARMM force field. In the SCC-DFTB/CHARMM simulations, the populations obtained along the trajectory can have *i*,*i*+3 C**O**···**H**N distances around 2.25 Å, particularly for the residues at the termini. This can be related to the NMR observation of significant contribution of the 3_{10} configuration for alanine-rich peptides, especially at the termini.[47]

With the development of linear-scaling SCC-DFTB,[65] time-dependent treatment of electronically excited states,[66] implicit

solvation models and more parameters for metal ions for which DFT methods are well suited,[67] the applicability of the present approach for studies of biologically interesting systems will be extended.

Acknowledgment. Part of the computations were done on the J90 machines at NERSC, and the IBM SP2 machines at Argonne National Lab. The research was supported in part by the Department of Energy and the National Science Foundation.

Supporting Information Available: Energetics for the model structures in TIM catalyzed reactions, QM/MM and QM energetics for syn-model-3a, QM/MM energetics for TIM catalyzed reactions in enzyme, and gas phase QM structures. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) See, for example: Head-Gordon, M. *J. Phys. Chem*. **1996**, *100*, 13213.

(2) (a) Curtiss, L. A.; Raghavachari, K.; Pople, J. A. *J. Chem. Phys*. **1995**, *103*, 4192. (b) Petersson, G. A.; Malick. D. K.; Wilson, W. G.; Ochterski, J. W.; Montgomery, J. A.; Frisch, M. J. *J. Chem. Phys*. **1998**, *109*, 10570.

(3) See, for example: (a) White, C. A.; Johnson, B. G.; Gill, P. M. W.; Head-Gordon, M. *Chem. Phys. Lett*. **1994**, *230*, 8. (b) Strain, M. C.; Scuseria, G. E.; Frisch, M. J. *Science* **1996**, *271*, 51. (c) Yang, W. *Phys. Rev. Lett*. **1991**, *66*, 1438.

(4) York, D. M.; Lee, T. S.; Yang, W. *Phys. Rev. Lett.* **1998**, *80*, 5011. (b) Lee, T. S.; York, D. M.; Yang, W. *J. Chem. Phys*. **1996**, *105*, 2744. (c) Daniels, A. D.; Millam, J. M; Scuseria, G. E. *J. Chem. Phys*. **1997**, *107*, 425.

(5) See, for example: (a) Beratan, D. N.; Onuchic, J. N.; Winkler, J. R.; Gray, H. B. *Science*, **1992**, *258*, 1740. (b) Kelley, S. O.; Barton, J. K. *Science* **1999**, *283*, 375.

(6) (a) Warshel, A.; Karplus, M. *J. Am. Chem. Soc*. **1972**, *94*, 5612. (b) Field, M. J.; Bash, P. A.; Karplus, M. *J. Comput. Chem*. **1990**, *11*, 700. (c) Maseras, F.; Morokuma, K. *J. Comp. Chem*. **1995**, *16*, 1170. (d) Matsubara, T.; Maseras, F.; Koga, N.; Morokuma, K. *J. Phys. Chem*. **1996**, *100*, 2573. (e) Gao, J. In *Reviews in Computational Chemistry*; Lipkowitza, K. B., Boyd, D. B., Eds.; VCH: New York, 1996; Vol. 7, p 119. (f) Sighn, U. C.; Kollman, P. A. *J. Comput. Chem.* **1996**, *7*, 718. (g) Bakowies, D.; Thiel, W. *J. Phys. Chem*. **1996**, *100*, 10580.

(7) See for example: (a) Bash, P. A.; Field, M. J.; Davenport, R. C.; Petsko, G. A.; Ringe, D.; Karplus, M. *Biochemistry* **1991**, *30*, 5826. (b) Bash, P. A.; Field, M. J.; Karplus, M. *J. Am. Chem. Soc.* **1987**, *109*, 8092. (c) Gao, J. *Acc. Chem. Res*. **1996**, *29*, 298. (d) Thompson, M. A.; Schenter, G. K. *J. Phys. Chem*. **1995**, *99*, 6374. (e) Froese, R. D.; Humbel, S.; Svensson, M.; Morokuma, K. *J. Phys. Chem*. *A* **1997**, *101*, 227. (f) Humbel, S.; Sieber, S.; Morokuma, K. *J. Chem. Phys.* **1996**, *105*, 1959. (g) Mulholland, A. J. Lyne, P. D.; Karplus, M. *J. Am. Chem. Soc.* **2000**, *122*, 534.

(8) (a) Bash, P. A.; Ho, L. L.; Mackerell, A. D.; Levine, D.; Hallstrom, P. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 3698. (b) Li, J.; Fisher, C. L.; Chen, J. L.; Bashford, D.; Noodleman, L. *Inorg. Chem*. **1996**, *35*, 4694.

(9) Gao, J. *J. Am. Chem. Soc*. **1994**, *116*, 9324.

(10) York, D. M.; Lee, T. S.; Yang, W. *J. Am. Chem. Soc*. **1996**, *118*, 10940.

(11) Cui, Q.; Karplus, M. *J. Chem. Phys*. **2000**, *112*, 1133.

(12) Cui, Q.; Karplus, M. *J. Phys. Chem. B* **2000**, *104*, 3721.

(13) (a) Åqvist, J.; Warshel, A. *Chem. Rev*. **1993**, *93*, 2523. (b) Neria, E.; Karplus, M. *Chem. Phys. Lett*. **1997**, *267*, 23.

(14) (a) Rossi, I.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *233*, 231. (b) Lopez, X.; York, D. M.; Karplus, M. Unpublished work.

(15) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260.

(16) Goringe, C. M.; Bowler, D. R.; Hernandez, E. *Rep. Prog. Phys*. **1997**, *60*, 1447.

(17) (a) Elstner, M.; Porezag, D.; Frauenheim, T.; Suhai, S.; Seifert, G. Materials Research Society, Pittsburgh, PA, 1999. (b) Elstner, M.; Frauenheim, T.; Kaxiras, E.; Seifert, G.; Suhai, S. *Phys. Status Solidi B* **2000**, *217*, 357.

(18) Bohr, H. G.; Frimand, K.; Jalkaanen, K. J.; Elstner, M.; Suhai, S. *Chem. Phys*. **1999**, *246*, 13.

(19) Brooks, B. R.; Burccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comp. Chem*. **1983**, *4*, 187.

(20) Cui, Q.; Karplus, M. *J. Am. Chem. Soc*. Submitted for publication.

(21) Porezag, D.; Frauenheim, T.; Köhler, T.; Seifert, G.; Kaschner, R. *Phys. Rev. B* **1995**, *51*, 12947.

(22) Koepernik, K.; Eschrig, H. *Phys. Rev. B* **1999**, *59*, 1743.

(23) *Handbook of Chemistry and Physics*, 78th ed.; CRC Press: New York, 1997.

(24) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.

(25) Pariser, R. *J. Chem. Phys*. **1956**, *24*, 250.

(26) Parr, R. G.; Pearson, R. G. *J. Am. Chem. Soc*. **1983**, *105*, 7512.

(27) (a) Ohno, K. *Theor. Chim. Acta*. **1964**, *2*, 219. (b) Klopman, G. *J. Am. Chem. Soc*. **1964**, *86*, 4550. (c) Mataga, N.; Nishimoto, K. *Z. Phys. Chem. (Frankfurt)* **1957**, *13*, 140.

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

(29) Eurenius, K. P.; Chatfield, D. C.; Brooks, B. R.; Hodoscek, M. *Int. J. Quantum Chem*. **1996**, *60*, 1189.

(30) Lyne, P.; Hodoscek, M.; Karplus, M. *J. Phys. Chem*. **1999**, *103*, 3462.

(31) Han, W.; Elstner, M.; Jalkanen, K. J.; Frauenheim, T.; Suhai, S. *Int. J. Quantum Chem.* **2000**, *78*, 459.

(32) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. P. *J. Chem. Phys.* **1983**, *79*, 926.

(33) For a recent review, see: Knowles, J. R. *Nature* **1991**, *350*, 121.

(34) MacKerell, A., et al. *J. Phys. Chem. B* **1998**, *102*, 3586.

(35) For a recent review, see: Bolin, K. A.; Millhauser, G. L. *Acc. Chem. Res.* **1999**, *32*, 1027.

(36) Voet, D., Voet, J. G., Eds. *Biochemistry*, 2nd ed.; J. Wiley & Sons, Inc.: New York, 1995.

(37) Barlow, D. J.; Thornton, J. M. *J. Mol. Biol.* **1988**, *201*, 601.

(38) (a) Sundaralingam, Sekharudu, Y. C. *Science* **1989**, *244*, 1333. (b) Millhauser, G. L. *Biochemistry* **1995**, *34*, 3873.

(39) Sheinerman, F. B.; Brooks, C. L. *J. Am. Chem. Soc.* **1995**, *117*, 10098.

(40) (a) Tirado-Rives, J.; Jorgensen, W. L. *Biochemistry* **1991**, *30*, 3864. (b) Sung, S. S. *Biophysics* **1995**, *68*, 826.

(41) McPhalen, C. A.; Vincent, M. G.; Picot, D.; Jansonius, J. N.; Lesk, A. M.; Chothia, C. *J. Mol. Biol.* **1992**, *227*, 197.

(42) Gerstein, M.; Chothia, C. *J. Mol. Biol*. **1991**, *220*, 133.

(43) (a) Marqusee, S.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 8898. (b) Marqusee, S.; Robbins, V. H.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5286.

(44) (a) Miick, S. M.; Martinez, G. V.; Fiori, W. R.; Todd, A. P.; Millhauser, G. L. *Nature*, **1992**, *359*, 653. (b) Fiori, W. R.; Lundberg, K. M.; Millhauser, G. L. *Nat. Struct. Biol*. **1994**, *1*, 374.

(45) Smythe, M. L.; Nakaie, C. R.; Marshall, G. R. *J. Am. Chem. Soc.* **1995**, *117*, 10555.

(46) (a) Hanson, P.; Martinez, G.; Millhauser, G. L.; Formaggio, F.; Crisma, M.; Toniolo, C.; Vita, C. *J. Am. Chem. Soc.* **1996**, *118*, 271. (b) Hanson, P.; Martinez, G.; Millhauser, G. L.; Formaggio, F.; Crisma, M.; Toniolo, C. *J. Am. Chem. Soc.* **1996**, *118*, 7618.

(47) Millhauser, G. L.; Stenland, C. J.; Hanson, P.; Bolin, K. A.; van de Ven, F. J. M. *J. Mol. Biol.* **1997**, *267*, 963.

(48) Long, H. W.; Tycko, R. *J. Am. Chem. Soc.* **1998**, *120*, 7039.

(49) Fiori, W. R.; Miick, S. M.; Millhauser, G. L. *Biochemistry* **1993**, *32*, 11957.

(50) Zhang, L.; Hermans, J. *J. Am. Chem. Soc*. **1994**, *116*, 11915.

(51) Tirado-Rives, J.; Maxwell, D. S.; Jorgensen, W. L. *J. Am. Chem. Soc*. **1993**, *115*, 11590.

(52) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc*. **1988**, *110*, 1657.

(53) Smythe, M. L.; Huston, S. E.; Marshall, G. R. *J. Am. Chem. Soc*. **1993**, *115*, 11594.

(54) Huston, S. E.; Marshall, G. R. *Biopolymers* **1994**, *34*, 75.

(55) Alemän, C.; Roca, R.; Luque, F. J.; Orozco, M. *Proteins* **1997**, *28*, 83.

(56) Elstner, M.; Jalkanen, K.; Knapp-Mohammady, M.; Frauenheim, T.; Suhai, S. *Chem. Phys*. **2000**, *256*, 15.

(57) Guo, H.; Karplus, M. *J. Phys. Chem.* **1994**, *98*, 7104.

(58) Brooks, C. L., III; Karplus, M. *J. Chem. Phys.* **1983**, *79*, 6312.

(59) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys*. **1977**, *23*, 327.

(60) Fincham, D.; Ralston, B. J. *Comput. Phys. Commun*. **1981**, *23*, 127.

(61) *NMR of Proteins and Nucleic Acids*; Wüthrich, K. J., Ed.; Wiley & Sons, Inc.: New York, 1986.

(62) Liu, H.; Thomas, P. D.; James, T. L. *J. Magn. Reson*. **1992**, *98*, 163.

(63) Soman, K. V.; Karimi, A.; Case, D. A. *Biopolymers* **1993**, *33*, 1567.

(64) Lockhart, D. J.; Kim, P. S. *Science* **1993**, *260*, 198.

(65) Liu, H.; Elstner, M.; Yang, W.; Kaxiras, E.; Frauenheim, T. To be submitted for publication.

(66) Elstner, M.; Neihaus, T.; Frauenheim, T.; Suhai, S. Manuscript in preparation.

(67) Cui, Q.; Elstner, M.; Kaxiras, E.; Frauenheim, T.; Karplus, M. Manuscript in preparation.