phys. stat. sol. (b) **217**, 357 (2000)

Subject classification: 71.15.Fv; 71.15.Mb; 78.30.Jw; S12

A Self-Consistent Charge Density-Functional Based Tight-Binding Scheme for Large Biomolecules

M. ELSTNER (a, b), TH. FRAUENHEIM (b), E. KAXIRAS (a), G. SEIFERT (b), and S. SUHAI (c)

(a) Department of Physics, Harvard University, Cambridge MA 02138, USA

(b) Theoretische Physik, Universität Paderborn, D-33098 Paderborn, Germany

(c) Molekulare Biophysik, Deutsches Krebsforschungszentrum, D-69120 Heidelberg, Germany

(Received August 10, 1999)

A common feature of traditional tight-binding (TB) methods is the non-self-consistent solution of the eigenvalue problem of a Hamiltonian operator, represented in a minimal basis set. These TB schemes have been applied mostly to solid state systems, containing atoms with similar electronegativities. Recently self-consistent TB schemes have been developed which now allow the treatment of systems where a redistribution of charges, and the related detailed charge balance between the atoms, become important as e.g. in biological systems. We discuss the application of such a method, a self-consistent charge density-functional based TB scheme (SCC-DFTB), to biological model compounds. We present recent extensions of the method: (i) The combination of the tight binding scheme with an empirical force field, that makes large scale simulations with several thousand atoms possible. (ii) An extension which allows a quantitative description of weak-bonding interactions in biological systems. The latter include an improved description of hydrogen bonding achieved by extending the basis set and improved molecular stacking interactions achieved by incorporating the dispersion contributions empirically. In applying the method, we present benchmarks for conformational energies, geometries and frequencies of small peptides and compare with ab initio and semiempirical quantum chemistry data. These developments provide a fast and reliable method, which can handle large scale quantum molecular dynamic simulations in biological systems.

1. Introduction

Biomolecules are challenging systems for computational methods for several reasons:

(i) they usually contain many different types of atoms with very different electronegativities and properties, such as H, C, N, O, P, S as well as several types of metal atoms. The molecules can occur in multiply charged states and the different electronegativities lead to large inter- and intra-molecular charge transfer.

(ii) Biomolecules exhibit a large variety of different bonding types, ranging from covalent and ionic bonding to hydrogen bonding and van der Waals-type interactions. Compared to covalent bond energies, the hydrogen bonds are much weaker, ranging from 2 to 20 kcal/mol, which puts considerable demands on the accuracy of computational methods.

(iii) The potential energy surfaces (PES) of biomolecules covering e.g. H bonded complexes, polypeptides, etc. are highly complex, exhibiting many local minima with small energy differences and often separated by small energy barriers. These minima

can be very shallow, where large geometrical changes can occur with only a small change in energy.

(iv) The description of chemical reactions and transfer processes of protons and electrons is of special interest in the theoretical modelling. Therefore, the methods should address issues of bond breaking and formation properly. Further, it is desirable that reaction energies, transition states and reaction pathways are determined with high accuracy, i.e. within a few kcal/mol.

(v) Many biomolecules of interest are very large, while chemical reactions depend sensitively on the protein enviroment and on solvent effects. Therefore, simple model calculations of only one part of the system are often of limitied value. For predictive calculations, realistically large systems containing up to several thousand atoms must be considered. Furthermore, in order to compare to experimental situations, free energies rather than potential energies have to be calculated. This would involve molecular dynamic (MD) simulations over rather long time scales, ranging form several hundreds of picoseconds up to milliseconds. In the latter cases, special sampling methods may allow a considerable reduction in the simulation time. Note, however, that many structure forming processes appear only for long time scales, like the folding of a polypeptide into its three dimensional proteine structure.

Several of these features do not only occur in biomolecules, but are common to many organic and inorganic systems (molecules, clusters, solids, surfaces, adsorbates, ...) addressed by computational methods. Presently, there is no single method available which covers all different chemical bonding types at the desired level of accuracy to satisfy all the demands. For this reason, it is common practice that different methods are used for different purposes.

Methods using classical empirical potential energy functions can perform MD simulations for several thousands of atoms and can reach time-scales in the nano- and microsecond region. However, they usually do not allow for bond breaking and reformation. Further, they are generally parametrized to experimental data for equlibrium conformations of molecules and solids. Little information is included for the regions of the PES far from those equilibrium structures or for that matter for variable chemical environments.

The so called ab initio methods, on the other hand, have been shown to be highly accurate and predictive, but at a very high computational cost. This limits their applications to relatively small systems containing up to 100 atoms and MD time scales of about 1 ps. Even within the ab initio schemes, the accuracy varies with the special method used and the basis set applied. Hartree-Fock (HF) calculations can be applied for certain classes of systems; of more general applicability are post-Hatree-Fock methods like MP2, but they are limited to even smaller systems sizes of about 20 to 40 atoms. Density functional theory (DFT) methods have been increasingly applied to biomolecules yielding an accuracy comparable to MP2 at much lower cost [1], but those are still too computationally demanding for many interesting systems.

Semi-empirical methods provide a compromise: They are less accurate than the ab initio methods, but about two to three orders of magnitude faster, and by the same factor slower than empirical potential methods. Quantum chemical semi-empirical methods have been developed for several decades. The most widely used methods today, the AM1 [2] and PM3 [3] methods, are an approximation to the Hartree-Fock theory: they neglect certain types of integrals and determine the remaining integrals from experi-

mental data and/or by fitting them to reproduce properties of organic molecules, like heats of formation, geometries, dipole moments etc. [4]. These methods have been sucessfully applied to biomolecular systems, and a variety of attempts have been made to further improve their accuracy.

Another type of semi-empirical method, the so-called tight binding (TB) approach, has been developed mostly in the context of solid state theory. The standard TB method works by expanding eigenstates of a Hamiltonian in an orthogonalized basis set of atomic-like wavefunctions, and by representing the exact Hamiltonian operator with a parametrized matrix. A common feature of these methods is the non-self-consistent solution of the correponding eigenvalue problem, which does not take into account the charge redistribution in a molecule. This is of minor importance for systems which contain atoms with similar electronegativies. TB methods are therefore mainly applied to such systems which contain only one or two types of atoms. This approach fails when applied to organic or biological molecules consisting of carbon, hydrogen, oxygen and nitrogen. For a more detailed discussion of these methods, see the article of Frauenheim et al. [5], this volume.

Recently, we have presented the development of a self-consistent-charge tight binding scheme (SCC-DFTB), based on the density functional theory [6, 7]. This method is derived by a second order expansion of the DFT total energy functional with respect to the charge density fluctuations $\Delta\varrho$ around a given reference density ϱ_0. The second order terms in the density fluctuations are approximated by a simple distribution of atom-centered point charges $\Delta q_\alpha = q_\alpha - q_\alpha^0$, estimated by a Mulliken charge analysis, while all other terms maintain the standard functional form of the tight-binding approach. Hence, this method can be applied in a straight-forward manner to any tight binding scheme in which the total energy is written in terms of the band structure energy $E_{\mathrm{bs}} = \sum_i^{\mathrm{occ}} \sum_{\mu\nu} c_\mu^i c_\nu^i H_{\mu\nu}[\varrho_0]$ and a short-range repulsive pair potential, E_{rep}, both determined at the reference density ϱ_0. The approximate DFT energy functional, explicitly containing the second order terms becomes

$$E_{\mathrm{tot}} = \sum_i^{\mathrm{occ}} \sum_{\mu\nu} c_\mu^i c_\nu^i H_{\mu\nu}[\varrho_0] + E_{\mathrm{rep}}[\varrho_0] + \tfrac{1}{2} \sum_{\alpha\beta} \Delta q_\alpha \, \Delta q_\beta \gamma_{\alpha\beta} \,. \tag{1}$$

The third term on the right hand side represents the longe-range Coulomb interactions between point charges at different sites and includes the self-interaction contributions of the single atoms [6]. Making this approximate Kohn-Sham functional subject to a variational principle within a LCAO representation for the single-particle electronic states, we derive a Kohn-Sham equation by minimizing the second order energy functional which is self-consistent in the Mulliken charge distribution through the modification of the Hamitonian matrix [6, 5]. We implemented this self-consistent charge extension into our density-functional based tight-binding method [8], where the $H_{\mu\nu}[\varrho_0]$ are calculated within DFT-GGA in a two-center approximation using a minimal basis of atomic-like wavefunctions ϕ_μ. In order to apply the method to biological molecules, we have derived tight-binding integrals as a function of distance between atom types S, O, N, C, and H. The parametrization for P, Zn and other metals like Mg or Na, relevant to biological processes, is in progress.

Before an approximate method, like AM1, PM3 or SCC-DFTB, can be applied to biological systems, its reliability has to be tested. A first test is to benchmark the meth-

od for properties of small organic molecules, to examine reaction energies, structures, vibrational frequencies etc. Such tests for the SCC-DFTB method have been presented elsewhere [6, 7]. They show that this method has an accuracy which is comparable to that of fully self-consistent DFT methods. However, the performance for the properties of small organic molecules is not a sufficient test of the method for larger, biologically relevant structures. Important interactions and chemical situations, present in larger molecules, as e.g. H bonding and molecular stacking interactions, are not covered by the tests on small molecules. Further, the structure and energetics of larger molecules can be determined by the energy landscape given, e.g., by rotations around covalent bonds. These can be very different from the ground state of small organic molecules. Additionally, a different chemical environment can alter the energy landscape. This might be the environment in a protein, or the effect of aqueus solution. Typical examples are polypeptides, where different conformations like α- and 3_{10}-helical structures or β-sheets, result from rotations around covalent bonds with low energy barriers. Therefore, to examine the reliability of an approximate method, one must test it for molecules and molecular models that represent typical situations in biological systems. If a method performs well for representatives of a certain class of molecules, it can be used for a prediction of yet unknown properties. The approximate methods should also be examined for typical reactions, transition states, etc.

Another issue is the computational cost. Although the approximate methods mentioned above are two to three orders of magnitude faster than ab initio methods, they are still limited to system sizes of several hundred atoms. There are basically two approaches suitable for reaching larger system sizes in the framework of quantum mechanical methods: The first is to circumvent the computationally costly solution of the generalized eigenvalue problem, which exhibits a cubic scaling with increasing system size; the second consists of combining quantum mechanical (QM) methods with empirical molecular mechanical (MM) force fields (QM/MM).

The outline of the paper is as follows: In Section 2, we describe the combination of the SCC-DFTB with a MM method. In Section 3, we discuss the challenges of describing H bonding within an approximate method and summarize the results of the SCC-DFTB method. In Section 4, we focus on the investigation of typical peptide structures and examine the performance of the SCC-DFTB in comparison with ab initio methods for energetic, structural and vibrational properties. Molecular stacking interactions, as they appear for example in the DNA double helix, seem to be even more challenging to theoretical modelling. They are poorly described even within more elaborate methods, such as DFT. Therfore, since the SCC-DFTB method is an approximation to DFT, we included these interactions empirically into the SCC-DFTB scheme; this is described in Section 5. Finally, Section 6 contains conclusions and outlook.

2. Hybrid SCC-DFTB/Molecular-Mechanical Coupling

The computational cost for large molecules (containing more than 100 atoms) within the SCC-TB method is determined by the diagonalization of the Hamitonian matrix, which exhibits N^3 scaling with increasing system size N. The method allows routine application to systems containing several hundred atoms on a workstation. However, for extended molecular dynamics runs or for the study of very large systems (containing several 1000 atoms), the SCC-DFTB method is computationally too costly. One way

to deal with this is through the so called O(N) methods, which circumvent the matrix diagonalization for the solution of the generalized eigenvalue problem. Such methods are described and applied to large scale simulations in the articles by Galli and Ordejon, in Refs. [9, 10], this Special Issue. Here we focus on an alternative approach, which has become popular in quantum chemistry in the last few years: the combination of a quantum mechanical method with an empirical force field. The idea behind the combined quantum mechanical/empircal force field methods (QM/MM) is to describe a part of the molecule quantum mechanically and the rest of the system within the computationally much faster empirical force field approach. In this approach, the total energy is usually written as

$$E_{\mathrm{tot}} = E_{\mathrm{QM}} + E_{\mathrm{MM}} + E_{\mathrm{QM-MM}}\,, \tag{2}$$

where E_{QM} is the energy of the QM part of the subsystem represented by the SCC-DFB energy Eq. (1), E_{MM} is the energy of the MM subsystem given by the energy function of the empirical force field, and $E_{\mathrm{QM-MM}}$ describes the coupling of the two subsystems.

If the boundary of the QM and MM regions intersects a covalent bond, the combination of those methods is not straight-forward. Several suggestions have been made to tackle this problem. A popular approach is the so called link-atom approach, where the quantum system is saturated with a fictitious atom for the QM calculation only, while the bond across the QM/MM boundary is modeled by the bonding interaction of the empirical force field. QM/MM approaches have been reviewed recently [11], and we will not discuss details of their implementation. We will only present the main ideas for the SCC-DFTB/MM-coupling [12, 13].

$E_{\mathrm{QM-MM}}$ consists of Coulomb and van der Waals (vdW) interactions between the two subsystems. The vdW interaction is modeled by the interaction terms present in the empirical force field method, while the Coulomb term is approximated by the interactions of the point charges between the subsystems, where the QM charges are given by the Mulliken charges Δq_α of the SCC-DFTB method and the MM charges Q_β are given by the force field parameters:

$$E_{\mathrm{QM-MM}} = -\sum_{\alpha\in\mathrm{QM},\,\beta\in\mathrm{MM}} \frac{\Delta q_\alpha\, Q_\beta}{R_{\alpha\beta}} + E_{\mathrm{vdW}}\,. \tag{3}$$

The total energy is

$$E_{\mathrm{tot}} = E_{\mathrm{QM}} + E_{\mathrm{MM}} - \sum_{\alpha\in\mathrm{QM},\,\beta\in\mathrm{MM}} \frac{\Delta q_\alpha\, Q_\beta}{R_{\alpha\beta}} + E_{\mathrm{vdW}}\,. \tag{4}$$

Applying the variational principle to this energy expression, we arrive at the generalized eigenvalue problem

$$\sum_\nu c^i_{n\nu}(H_{\nu\mu} - \epsilon_i S_{\nu\mu}) = 0 \tag{5}$$

with the matrix elements

$$H_{\nu\mu} = H^0_{\nu\mu} + \tfrac{1}{2} S_{\nu\mu} \sum_\delta (\gamma_{\nu\delta} + \gamma_{\mu\delta})\,\Delta q_\delta - \tfrac{1}{2} S_{\nu\mu} \sum_{\beta\in\mathrm{MM}} \left(\frac{1}{R_{\nu\beta}} + \frac{1}{R_{\mu\beta}}\right) Q_\beta\,. \tag{6}$$

$S_{\nu\mu}$ is the overlap matrix, $R_{\mu\beta}$ the distance between the corresponding QM and MM atoms. The generalized eigenvalue problem therefore has to be solved in the presence

of the "external" charges Q_β. We have tested this method extensively for H bonded compounds, where one molecule is treated quantum mechanically and the other is treated with the force field method. The results are very promising: geometries and energies compare well with higher level calculations and the relative ordering of the energies of several conformers is well reproduced [13].

The effects of external electrical fields can be included similarily in an approximate way, i.e. making use of the monopole approximation for the SCC-DFTB charges and coupling the electric field to the SCC-DFTB charges [12]. We have used this extension to study proton transport in linear water filaments, driven by the external electric field [14].

3. The Description of H Bonding

Hydrogen bonds are a common bonding pattern in biolocial structures which play a crucial role for determining the geometries, the energetics and other properties of biomolecules. Two or more molecules can be bound together by an inter-molecular H bond, or the conformation of one molecule can be stabilized by intra-molecular H bonds. Intra-molecular H bonds e.g. stabilize the three-dimensional structure of polypeptides and proteins. In these systems, different conformations may exhibit different H bonding patterns, as will be discussd in the next section.

H bonds can be described as van der Waals (vdW)-type interactions, as a combination of electrostatic attraction, charge transfer effects and repulsion and dispersion interactions. The dispersion interaction contributes significantly to the binding energy and strongly reduces the H bond lengths.

Although ab initio methods are computationally too demanding for many biomolecules of interest, they can prove helpful in gaining insight into the physics of the H bond by studying small model systems. They can also be used to build up a data set for either parametrizing the approximate methods or for testing their performance.

Most ab initio studies of H bonded complexes have been performed at the HF and post-HF (MP2 and higher) level of theory using localized basis functions. In calculating H bond energies, three major sources of error have to be considered [15]. The first is called the basis set superposition error (BSSE), while the second is referred to as the basis set incompleteness error (BSIE) and the third is related to an insufficient treatment or even neglect (HF) of the dispersion interaction. The first error stems from the manner in which binding energies are evaluated by subtracting the energies of the monomers from the energy of the compound. By doing this, the energy of the monomers is calculated in the monomer basis sets, while in the calculation of the compound energy, each monomer of the compound is described in the basis set of the whole compound. This results in lowering the monomer enery in the compound relative to the energy of the isolated one due to the variational principle. Therefore, BSSE errors lead to an overestimation of the binding energy; typical errors are in the range of 0.5 to 3 kcal/mol, depending on the system and the basis set. The BSSE error decreases with increasing size of the basis set, because when a monomer is described in a satisfactory basis the additional basis functions from neighboring molecules have neglegible effect on its total energy. There are several ways to account for the BSSE error; the most widley used method is given by the counterpoise procedure (CP) of Boys and Bernardi [16] where the monomer energies are calculated in the basis set of the compound, that is the bonded and isolated monomers are treated within the same basis set. The BSIE

also leads to an overestimation of the binding energy; the error again decreases for larger basis sets. The effect of correlation is also to increase the binding energy as a function of the size of the basis set.

These three factors seem to show typical trends for a wide class of H bonded systems and have been examined in detail for the water dimer [15] (see Fig. 1). As a result of these three factors, calculations at lower theoretical levels can lead accidentally to good binding energies. For example, the binding energy of the water dimer is 3.6 kcal/mol and 4.9 kcal/mol at the HF and MP2 level within the basis set limit. The experimental value is 5.4 kcal/mol [17]. Because the neglect of correlation lowers the binding energy and the BSIE overestimates it, a HF calculation using singly polarized basis sets like 6-31G(d,p) finds a surprising accurate value of 5.5 kcal/mol, while it would be overestimated by 30% at the MP2 level using the same basis set [15].

While the H bonding energies are clearly overestimated within the the DFT-LDA approach, gradient corrected DFT methods seem to perform very well for H bonded systems too: the binding energy of the water dimer is slightly understimated by about 0.5 to 1 kcal/mol, depending on the basis set and the exchange–correlation functional used [18].

Concerning the approximate (semi-empirical) methods it is difficult to account properly for the sources of errors as discussed above for the ab initio methods. The lack of extended basis sets or appropriate consideration of correlation effects (dispersion interaction) is partially compensated by a proper choice of empirical parameters. Semi-empirical methods show the overall tendency to understimate H bond strenghs (for a review, see [19]).

The same tendency is found for the SCC-DFTB method [20, 13]. The geometries, especially the H bond distances, are in very good agreement with MP2 and DFT calculations, with the ordering in energy of different conformations well reproduced. But the SCC-DFTB binding energies of weakly bonded complexes are consistently underestimated by 1 to 2 kcal/mol when compared to ab initio methods.

Several attempts to refine the semi-empirical methods for H bonded systems have been made. One strategy is to modify the core–core repulsion term. This term not only contains the ion–ion repulsion, but also an additional effective energy contribution. In the AM1 method, this part has been modified compared to its predecessor MNDO [21], which improves H bonding significantly. However, an accurate description of H bonds seems to require a specific parametrization for these systems, as has been done in the MNDO/M [22] approach. In that approach, new functions for the core–core repulsion energy have been introduced and explicitly parametrized to reproduce binding energies and geometries of simple H bonded complexes. A different strategy has been followed in the framework of the SINDO1 model [23], where polarization func-

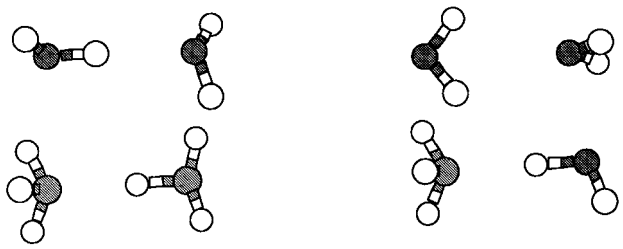

Fig. 1. Linear and bifurcated water dimer configurations (upper left and upper right), linear ammonia dimer (lower left) and ammonia–water complex (lower right)

tions have been introduced for the hydrogen atoms to increase the inter-molecular bonding. But this approach did not succeed without introducing an additional empirical function, which damps the effect of the p-functions for distances smaller and larger than the typical hydrogen bond distance.

To improve the description of H bonding in the SCC-DFTB model, we extended the clearly very limited basis for hydrogen (only 1s) by including also 2p basis functions. This should improve the description of the charge transfer effects in the H bonding region ($\approx$2 ÅX–H distance) and increase the H bonding energy. However, since the 2p atomic wavefunction is quite diffuse, the long range decay of the corresponding matrix elements had to be modified that they approach smoothly to zero at about 5 to 6 Å.

A large variety of hydrogen bonding complexes have been studied with this extension of the method. We compared the results of different levels of theory, including empirical force-field and semiempirical methods, DFT, HF and post-HF methods and the minimal basis DFTB approach. Here we focus on discussing only a few weakly bonded complexes with interaction energies ranging from 2 to 6 kcal/mol, which are the least satisfactorily described systems within SCC-DFTB (see Fig. 1).

The global minimum of the water dimer is a linear structure, where one H bond is formed between the water monomers (structure 1 in Table 1). The bifurcated structure, where both hydrogens of one water molecule form H bonds with the oxygen of the other water molecule (structure 2), is about 2 kcal/mol lower in energy at the MP2/6-311+G** level of theory [24]. The NH_3 dimer considered here forms one N–H H bond, and in the conformer NH_3–H_2O-1, the water molecule acts as a donor forming one H bond, while in conformer NH_3–H_2O-2 both hydrogens form H bonds with the nitrogen, similar to the case in the conformer H_2O–H_2O-2. In the H_3COH–H_2O-1 complex, the H_3COH acts as a donor, while in the H_3COH–H_2O-2 the reverse is true. With the exception of the NH_3–H_2O-1 conformer, where the binding energy is underestimated by more than 1 kcal/mol, the results in Table 1 show that the SCC-DFTB method agrees quite well with the higher level calculations.

SCC-DFTB geometries compare very well with higher level results, and the energetic ordering of different conformers of the complexes (not discussed here in detail) is also reproduced (which is also true for the SCC-DFTB method applying the minimal basis set only [13]). H bond lengths are not given here in detail, but they deviate with respect to the DFT and MP2 results by about 0.05 to 0.1 Å. For example, the O–O

Table 1

H bonding energies (kcal/mol) of H bond complexes; min denotes SCC-DFTB calculations with the minimal LCAO basis set. AI denotes ab initio calculations as dicussed in the text

complex	SCC-DFTB	min	AI	complex	SCC-DFTB	min	AI
H_2O–H_2O-1	5.0	3.3	5.4[1])	H_3CNH_2–H_2O	4.9	3.3	6.5[4])
H_2O–H_2O-2	3.6	2.1	3.5[1])	NH_3–H_2O-1	4.9	3.4	5.8[3])
NH_3–NH_3	3.4	1.8	3.1[2])	NH_3–H_2O-2	3.3	1.4	3.2[3])
H_3COH–H_2O-1	5.3	3.5	5.6[4])	HCOOH–H_2O	9.0	7.1	10.8[4])
H_3COH–H_2O-2	5.1	3.1	5.6[4])				

[1]) MP2/6-311+G** [24], [2]) HF/6-31+G(2d,2p) [29], [3]) MP2/6-311+G** corrected for BSSE [13], [4]) HF/6-31G*[28]

distance in H_2O–H_2O-1 is 2.85 Å, whereas it is 2.9 Å at the MP2 and MP4 level of theory [25].

The performance of semi-empirical methods like AM1, PM3 and MNDO/M for H-bonded complexes has been evaluated recently [19]. Most semi-empirical methods underestimate H-bond strengths when there is no special emphasis on H bonds in the parameter determination procedure, as e.g. in the MNDO/M method. The performance of the MNDO/M method seems to be very good for H-bonding interactions: it has been shown to reproduce the stabilization energies of DNA H-bonded base pairs very well [27], while for other H-bonding complexes it shows a tendency to overestimate interaction enthalpies [19]. AM1 often does not predict the right ground state structure, e.g. bifurcated H bonds are favored against linear ones [19]. Another example is given by the H_3O^+–H_2O complex, which is symmetric, with the proton centered between the oxygens. Both, AM1 and PM3 fail to reproduce this global minimum: PM3 yields an asymmetric structure and AM1 predicts bifurcated binding while SCC-DFTB reproduces the correct symmetric structure.

The H_2O–OH^- compound has an asymmetric structure with two different O–H bond lengths at the HF, MP2 and MP4 levels of theory, while DFT predicts this to be a symmetric structure [26]. The SCC-DFTB follows here the DFT results, also predicting a symmetric conformation. The energy difference between the symmetric and asymmetric conformation is only a few tenths of kcal/mol at the post-HF level of theory.

Finally, we compare the interaction energies for H bonded DNA base pairs with MP2 results [27], as we have already done for the SCC-DFTB method without H-p type orbitals [20]. Hobza et al. [27] investigated the H-bonding energies of 26 base pair conformations of the bases adenine (a), cytosine (c), guanine (g) and thymine (t). In those compounds the interaction energies are underestimated in the SCC-DFTB (with the H-s basis only) compared to the MP2 values, yielding a mean average error of 2.8 kcal/mol compared to MP2. Empirical force fields and the semi-empirical methods like AM1 [2], PM3 [3] and MNDO/M [22] have been tested for this molecule set as well [27]. The empirical force fields show mean average errors of 0.9 to 2.4 kcal/mol, whereas the semi-emprical methods have mean average errors of 7.3 kcal/mol (AM1), 6.3 kcal/mol (PM3) and 2.5 kcal/mol (MNDO/M). The results for the SCC-DFTB including H-p type orbitals as given in Table 2 (for the abreviations see [27]) show similar trends as with s-basis only [20]: For instance, the interaction energy of the gg1 base pair is higher than that of the gcwc base pair. The tt and gt base pair interaction energies were nearly equal to the MP2 values in the H-s only basis, whereas other base pair interactions were underestimated on average by 3 kcal/mol. These base pair interaction energies are higher than the corresponding MP2 values. With the new H-p basis, we find a mean average error of 1.5 kcal/mol with respect to the MP2 values.

Clearly, none of the approximate methods, including SCC-DFTB, are able to produce highly accurate results for all the H bonded systems considered. But many systems also put high computational demands on the ab initio methods, where correlation, large basis sets and correction for BSSE have to be taken into account to provide the needed high accuracy. This makes the computations extremely demanding, so that any application to larger molecular complexes will be prohibitive. The SCC-DFTB scheme is able to reproduce geometries and energies very well, showing deviations from higher level calculations which are of the same order as those at the ab initio level, when different methods and basis sets are compared. The compounds tested so far (not all of which

Table 2
Interaction energies (kcal/mol) of H bonded base pairs, as described in the text. For the notation identifying the base pairs see Ref. [27]

base pair	SCC-TB	MP2	base pair	SCC-TB	MP2
gcwc	24.0	25.4	atwc	11.6	12.4
gg1	24.4	24.0	atrwc	11.5	12.4
cc	16.2	18.8	aa1	9.7	11.5
gg3	15.0	17.1	ga4	9.7	11.1
ga1	14.7	15.7	tc2	11.2	11.8
gt1	16.2	14.7	tc1	11.0	11.6
gt2	15.8	14.3	aa2	9.0	11.0
ac1	12.2	14.3	tt2	11.9	10.6
gc1	12.2	13.9	tt1	11.9	10.6
ac2	11.8	14.1	tt3	11.9	10.5
ath	11.3	13.3	ga2	9.0	10.4
ga3	12.8	15.2	gg4	8.6	10.3
atrh	11.3	13.2	aa3	8.3	10.0

are discussed here), were chosen to be representative of patterns that occur in biomolecules. Since these compounds are described satisfactorily, the SCC-DFTB method is expected to be successful in describing H bonding in biological systems.

4. Polypeptides: Conformational Energies and Geometries

The three-dimensional structure of peptides and proteins is given by the spacial arrangement of simple, so-called secondary structural elements, like α-helices, β-sheets (shown in Fig. 2) or turn structures. The pure sequence of amino acids, like glycine,

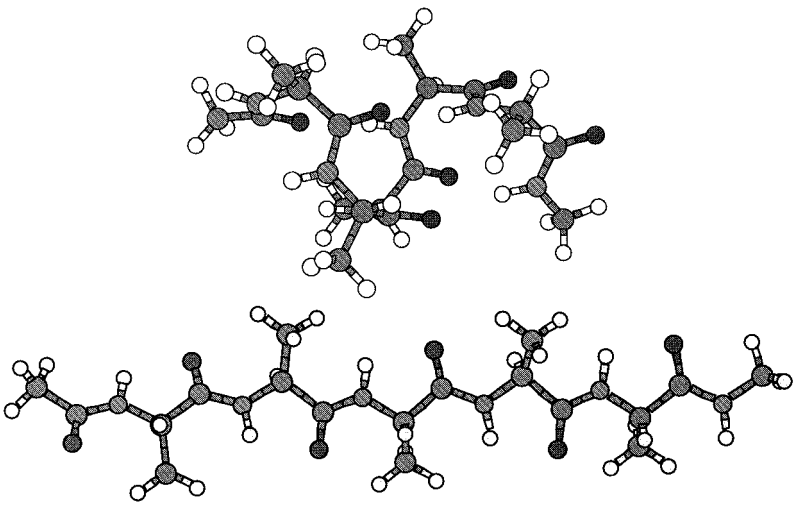

Fig. 2. α-helical and β-sheet (extended) conformations of a polypeptide model containing five alanine residues

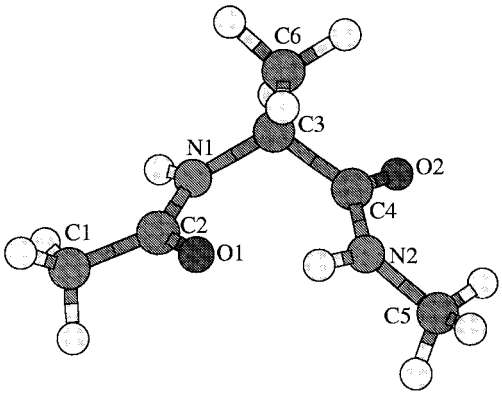

Fig. 3. The C_7^{ax} conformer of NA-LA-NMA (see text)

alanine etc., which build up secondary and tertiary structure, is called the primary structure. Since protein folding from a theoretical viewpoint can be understood as a hierarchical process, where secondary structure elements are formed first and then assembled to build up the three dimensional system [30], it is of great interest to study the fundamental structure and energetics of such elements with theoretical methods.

Molecules like those shown in Fig. 2, which contains five alanine residues, truncated by the "capping" groups $COCH_3$ at the left end and $NHCH_3$ (at the right end), are still very large for extensive ab initio studies. This is why smaller model peptides, like the N-acetyl-L-alanine-N′-methylamide (NA-LA-NMA) (Fig. 3) molecule have been studied at the HF, MP2 and DFT levels of theory in order to understand the geometric structure and the detailed energetics of polypeptides.

The NA-LA-NMA peptide model is derived from the alanine molecule by subsituting the H at N1 in the alanine molecule by the $C2O1C1H_3$ (acetyl) group and the OH group at C4 in the alaline molecule by the $N2HC5H_3$ (methylamide) group. These substitutions establish the Ψ and Φ dihedral angles, which characterize the peptide backbone conformations in polypeptides and proteins. The Φ angle is defined by rotations about the N1–C3 bond, Φ(C2–N1–C3–C4), whereas the Ψ angle is defined by rotations of the C3–C4 bond, Ψ(N1–C3–C4–N2), see Fig. 3.

In glycine based peptides the $C6H_3$ group is substiuted by a hydrogen atom. Other types of aminoacid residues result from substitution of this group by other organic rest-groups. These side-chains can themselves have rotational degrees of freedom, which are important for the tertiary structure, since they may allow for additional binding between amino-acid residues, as for example in sulfid bridges. Here, we will concentrate only on alanine based polypeptides.

NA-LA-NMA has six stable conformers on the DFT-B3LYP/6-31G* and MP2/6-31G* potential energy surfaces (PES) [31], which have been taken as starting structures for further geometry optimizations with the SCC-DFTB, AM1 and PM3 methods. The relative energies of these conformers for different methods are shown in Table 3. The three lowest energy conformers form internal H bonds, whereas the higher energy conformers do not. The C_7^{ax} structure is shown in Fig. 3. The C_7^{ax} and C_7^{eq} conformations differ in the orientation of the methyl group attached to the central C3 atom. In the C_7^{ax} conformer, this methyl group is perpendicular to the seven membered ring which forms the H bond, whereas in the C_7^{eq} this methyl group is in the plane of the seven mem-

Table 3

Relative energies (kcal/mol) of the different conformers of NA-LA-NMA for different methods as described in the text. DFT-GGA refers to the method B3LYP. The geometries at the DFT, MP2 and HF level have been determined with the 6-31G* basis set

conf.	DFT-GGA	MP2	HF	SCC-DFTB	PM3	AM1
C_7^{eq}	0.00	0.00	0.00	0.00	0.00	0.00
C_5^{ext}	1.43	1.76	0.41	0.99	−1.55*)	1.72*)
C_7^{ax}	2.58	2.61	2.82	1.03	0.87	0.72
β_2	3.18	3.37	2.58	2.20+)	–	–
α_L	5.82	4.60	4.72	3.70	3.53	–
α_P	6.85	6.34	5.74	4.78	1.10	3.29

+) The β_2 conformer is not stable within the SCC-DFTB, but the maximum force at the B3LYP geometry is very small. The energy is given for a geometry, in which the forces are smaller than 0.00065 a.u.

*) Distorted structure.

bered ring. In the C_5^{ext} conformer, the hydrogen attached to N1 forms a H bond with O2, leading to a five membered ring.

The α-helical structures are predicted to be high energy conformers on the PES. Since these two structures form no internal hydrogen bonds, the smaller energy difference with respect to the ground state could be due to the fact that the energy of the H bonds may be considered to be underestimated in the SCC-DFTB method. Therefore, the stabilization of the C_7 and C_5 structures is underestimated compared to the α-helical structures. The β_2 conformer is found to be unstable in the SCC-DFTB model, but the maximum force at the dihedral angles (as shown in Table 4) is very small. This conformer may be stabilized due to internal H bonds in larger polypeptides.

Both, the AM1 and PM3 methods find a very distorted C_5^{ext} conformation, where the internal H bond is broken, as can be seen from the dihedral angles in Table 4. The β_2 conformer is also not stable in both methods and the α_L is unstable in the AM1 model.

The effects of solvation and peptide length are expected to play a crucial role for stabilizing the different conformers. Therefore, relative stabilities of the various conformers estimated within quantum mechanical calculations may change when the effects of solvents are included and longer peptides are considered. By applying a quantum che-

Table 4

Dihedral angles (in degrees) of the NA-LA-NMA conformers for different methods as described in the text. DFT-66A refers to the B3LYP method

conformer angles	C_7^{eq}		C_7^{ax}		C_5^{ext}		β_2		α_L		α_P	
	Φ	Ψ	Φ	Ψ	Φ	Ψ	Φ	Ψ	Φ	Ψ	Φ	Ψ
DFT-66A	−81.9	72.3	73.8	−60.0	−157.3	165.3	−135.9	23.4	68.5	24.5	−169.4	−37.8
SCC	−81.3	72.0	74.6	−66.1	−153.2	176.6	−136.7	24.9+)	65.6	13.0	−172.5	−51.1
AM1	−84.4	68.5	76.6	−64.0	−117.7	141.5	–	–	–	–	−115.5	−55.2
PM3	−71.4	77.7	68.8	−67.9	− 93.9	147.9	–	–	62.3	39.6	−137.6	−60.5

+) The β_2 conformer is not stable within the SCC-DFTB method. The Φ, Ψ values refer to a conformation, where the maximum force is lower than 0.00065 a.u.

mical reaction field model at the RHF/6-31G* level of theory [32], the α_R conformation is shown to be considerably stabilized relative to the C_7^{eq} conformer. However, the conformation is still not a local minimum on the PES. Recently, it was shown that the α_R conformer will become stabilized only by explicitly including water molecules on a quantum theory level [33], supporting free energy calculations with emprirical force fields which also stabilize this conformer in solution [34]. We have shown that the SCC-DFTB model is also able to describe the changes of the PES due to the solvent effects [13].

While NALANMA does not show secondary structure motifs, like α-helical or turn-like conformers, such structural elements may be stabilized for larger polypeptides due to internal H bond formation. In addition, the relative stability of extended versus helical structures is expected to change due to cooperative effects which are much more pronounced in helical conformations. This has been examined with ab initio calculations, confirming large cooperative effects for the α_R helix compared to corresponding linear structures [35]. This in turn may result in an increased stability of helical conformers compared to extended ones with increasing peptide size.

We discuss next the tripeptide model NA-LA_2-NMA, containing two amino acids "capped" with the acetyle and methylamide groups, where the formation of turn structures is possible. The so-called β turn structures reverse the direction of a polypeptide via four amino acids and are therefore a common structural motif in proteins. In the tripeptide model (see Fig. 4), the oxygen atom of the acetyle goup can form a H bond with the nitrogen atom of the methylamide group, a so called $i \rightarrow i+3$ H bond, since this H bond connects the *i*-th residue (represented by the acetyle group) along a polypeptide chain with the residue $i+3$ (represented by the methylamide group). Protein α-helices form $i \rightarrow i+4$ H-bond patterns, therefore they cannot appear in this tripeptide model. Longer polypeptides, starting from the NA-LA_3-NMA model, would allow the formation of such H bonds.

Turn structures have been studied at the HF/3-21G [36], HF and MP2 (6-31G*) level of theory [37, 38]. Here we focus on the C_7^{eq}, C_5^{ext} linear repeat conformers and two β turn structures, the type I (βI) and type II (βII). These are classified by idealized backbone dihedral angles [39]. The dihedral angles are defined, as above in the NA-LA-NMA molecule, as rotations around the C_α–N and C_α–C bonds, where C_α is defined as the C atom to which the side chains (in the case of alanine the CH_3 group) are attached. A βI type turn structure is characterized by the ideal values $\Phi_1 = -60^\circ$ and $\Psi_1 = -30^\circ$, $\Phi_2 = -90^\circ$ and $\Psi_2 = 0^\circ$ respectively. A βII type turn is classified by the ideal dihedral

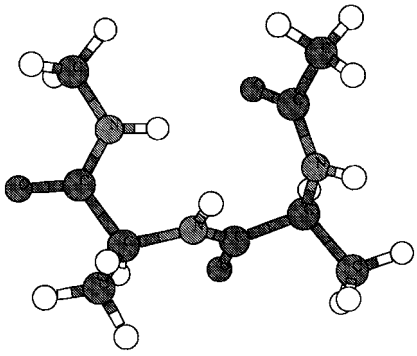

Fig. 4. The βI conformer of NA-LA_2-NMA (see text)

angles $\Phi_1 = -60°$, $\Psi_1 = 120°$, $\Phi_2 = 80°$ and $\Psi_2 = 0°$. A βIII type turn structure is characterized by the ideal values $\Phi_1 = -60°$ $\Psi_1 = -30°$ $\Phi_2 = -60°$ and $\Psi_2 = -30°$, but this turn is not a stable conformer in this tripeptide model. All these turn structures form $i \to i+3$ H bonds, but only the type βIII turns are usually referred to as 3_{10} helices. Table 5 shows the dihedral angles evaluated at the SCC-DFTB and B3LYP/6-31G* (using the GAUSSIAN98 program package [43]) levels of theory for the four conformations. The dihedral angles of SCC-DFTB compared to B3LYP show deviations of up to 20° as in the case of the βII conformer. However, the PES for these molecules are very shallow and, consequently, deviations in this range can be expected even when comparing ab initio calculations at various levels of theory. The relative energy differences are consistently underestimated in the SCC-DFTB model, but the relative stabilities are reproduced well.

In the SCC-DFTB method, dipole moments are calculated by using the SCC-DFTB Mulliken charges. Despite this approximation, the SCC-DFTB dipole moments compare quite well with those at the B3LYP level of theory, as shown in Table 5.

We also have investigated larger polyalanine molecules, containing up to 11 alanine residues at the B3LYP/6-31G*, SCC-DFTB, AM1 and PM3 levels of theory [41, 44]. For the SCC-DFTB model we found similar trends as described above: the relative energetic ordering of different conformers is reproduced well, although the relative energy differences are underestimated. Geometries are in good agreement with B3LYP results, showing similar deviations as dicussed above. The SCC-DFTB model, therefore, seems to give a reliable description of structures and a semi-quantitative estimate of the energetics. Secondary structural motives, like β-sheets, helices and turn structures, are predicted to be stable conformers in agreement with B3LYP results. For approximate methods, this good performance cannot be expected a priori. It has been shown that some empirical force field methods, AM1 and even HF/3-21G calculations, do not predict correctly turn structures to be stable conformers [37]. Further, semi-empirical methods like AM1 and PM3 do not reproduce satisfactorily the relative ordering in energy and structural properties of these secondary structural motifs [41, 44]. β-sheet structures show large distortions in both models. While PM3 is not able to describe helical structures, it unwinds helices and breaks the internal H bonds, AM1 seems to favor helices which are in between the 3_{10} and α_R conformations. In the latter case bifurcated H bonds (of both $i \to i+3$ and $i \to i+4$ type) are formed.

Table 5

Dihedral angles (in degrees), relative energies ΔE (in kcal/mol) and dipole moments (in Debeye) of NA-LA$_2$NMA at different levels of theory. DFT denotes B3LYP/6-31G* calculations

conformer	C_7^{eq}		C_5		βI		βII	
method	DFT	SCC-DFTB	DFT	SCC-DFTB	DFT	SCC-DFTB	DFT	SCC-DFTB
Φ_1	−82.7	−81.3	−158.7	−157.4	− 74.7	− 70.6	−60.8	−59.2
Ψ_1	69.5	68.8	165.5	175.7	− 12.3	− 5.4	128.8	110.2
Φ_2	−84.6	−83.3	−159.4	−160.5	−105.5	−111.0	69.8	64.1
Ψ_2	70.0	68.7	165.0	178.5	13.1	18.1	15.1	20.9
dipole	5.8	5.3	6.3	5.8	8.1	7.5	6.6	6.3
ΔE	0.0	0.0	2.13	1.51	2.59	1.76	4.32	2.71

Recently, we also studied the vibrational frequencies, infrared absorption (IR) and vibrational circular dichroism (VCD) intensities of NA-LA-NMA with the SCC-DFTB method in comparision to HF, B3LYP and MP2 calculations [45]. In this work, in order to estimate the intensities at the SCC-DFTB level we used an SCC-DFTB – DFT hybrid approach. Ground state geometries and second energy derivatives were calculated with the SCC-DFTB method while the dipole derivatives and VCD tensors were calculated with the B3LYP method by using the SCC-DFTB geometries. Frequencies for the C_7^{eq} and C_5 conformers at the B3LYP, MP2, HF (6-31G*) and SCC-DFTB level of theory are given in Ref. [45] in detail. The SCC-DFTB values compare very well with the higher level calculations; this also holds for the C_7^{ax}, α_L and α_P structures, but these results will not be given here explicitly. As one example, for the C_7^{eq} and C_5 conformers the standard deviation from experimental frequencies is 3.0%, 4.4% and 6.7% for the MP2, B3LYP and SCC-DFTB methods respectively. The IR and VCD intensities estimated with the hybrid approach compare satisfactorily with those of the higher level calculations [45] for the two conformers investigated, the C_7^{eq} and C_5^{ext}.

This good agreement encouraged us to calculate the IR intensities fully at the SCC-DFTB level of theory. As exploited above, the dipole moments from the SCC-DFTB model compare reasonably well with the full DFT results. We used the derivatives of the SCC-DFTB dipolmoments (calculated from Mullikan charges) with respect to the atomic coordinates as an approximate way to calculate the IR intensities.

Important modes for the characterization of peptide conformations are the N–H stretch (amide A) mode, the C=O stretch (amide I) and the N–H bend (in combination with C–N stretch) (amide II) modes, located around 3400 to 3500 cm^{-1}, 1700 cm^{-1} and 1500 to 1550 cm^{-1}, respectively (for the assignment of experimental frequencies see Ref. [31] and references therein).

The N–H and C=O bonds occur twice in NA-LA-NMA (Fig. 3), but the corresponding vibrational modes are not degenerate, since only one of the two N–H and C=O bonds is involved in a H bond. This alters the vibrational frequency and leads to a splitting of the amide A, I and II modes. The two modes corresponding to each bond type will be labeled by A_a and A_b, I_a and I_b etc. In Fig. 5 we show the IR spectra of the C_7^{eq} conformer of NA-LA-NMA estimated on the SCC-DFTB and the B3LYP/6-31G* level of theory.

The SCC-DFTB mainly overestimates the intensities of the more intense modes, i.e. the C=O stretch, relative to the less intense modes. This is mainly due to the approximation of the dipole derivatives, since in the hybrid SCC-DFTB/B3LYP approach the relative intensities compare much better with the full ab-initio data [45]. As can be seen from Fig. 5, the SCC-DFTB reproduces the frequency splitting of the N–H stretch mode at 3500 cm^{-1} and of the C=O stretch mode at 1750 cm^{-1} reasonably well. Only for the N–H bend mode the splitting is smaller than at the B3LYP level of theory, so that only one line occurs at 1600 cm^{-1} in the SCC-DFTB spectrum.

Next, we analyze the frequency splitting for the five stable conformers at the SCC-DFTB potential energy surface in more detail. Table 6 shows the values for the splitting of the amide A, I and II modes (experimental values and assingnement and B3LYP/6-31G* data are from Ref. [31]), which are calculated as the difference of the two frequencies of the modes A, I and II, e.g. $\Delta\nu_A = \nu_{A_a} - \nu_{A_b}$ (ν_{A_a} is larger than ν_{A_b}). As can be seen from the values for the C_7^{eq} and C_5 conformers, both theoretical methods show deviations from the experimental values of up to 30 cm^{-1}. Both methods

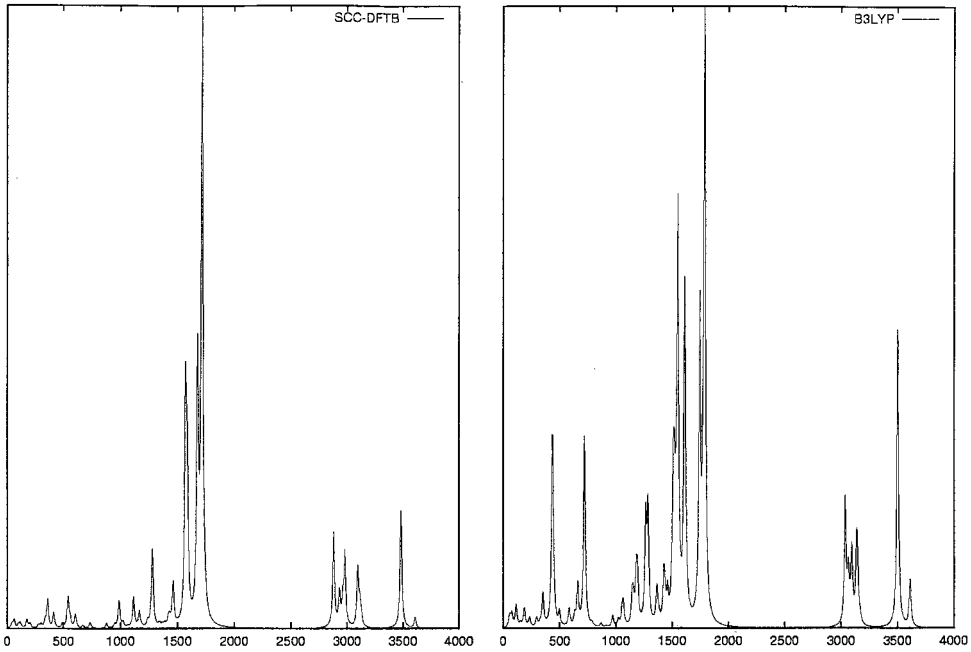

Fig. 5. IR intesities for the C_7^{eq} conformer of NA-LA-NMA at the SCC-DFTB (left) and B3LYP level of theory, see text. Intensities are in arbitray units, frequencies in cm^{-1}

reproduce the trends in the splitting for the amide A and I modes when going from the C_7^{eq} to the C_5 conformer. For the amide II mode, the SCC-DFTB does not reproduce the experimental trend, wheras B3LYP does. To discuss relative intensities of the a and b modes in more detail, we take the ratio of the intensities I_a^X and I_b^X:

$$Q = \frac{I_a^X}{I_b^X},$$

Table 6

Splitting of the amide A, I and II modes (in cm^{-1}) for the NA-LA_1NMA conformers at the B3LYP and SCC-DFTB level of theory respectively in comparison to experiment (see text)

	B3LYP	SCC-DFTB	EXP	B3LYP	SCC-DFTB	EXP	B3LYP	SCC-DFTB
	C_7^{eq}			C_5			α_P	
A	105	141	115	43	89	71	12	23
I	40	50	25	16	11	17	3	5
II	60	12	42	41	40	34	34	62
	C_7^{ax}			α_L				
A	45	33		34	20			
I	37	46		7	12			
II	59	13		27	46			

Table 7
Relative intensities of amide A (Q_A), I (Q_I) and II (Q_{II}) modes for the NA-LA$_1$NMA conformers at the B3LYP, SCC-DFTB and SCC-B3LYPB/B3LYP hybride (HYB) levels of theory, respectively (see text)

	B3LYP	SCC-DFTB	HYB	B3LYP	SCC-DFTB	HYB	B3LYP	SCC-DFTB
	C_7^{eq}			C_5			α_P	
A	1:6.4	1:9.7	1:6.5	1:4.0	1:6.3	1:3.6	1:1.2	1:1.1
I	2.1:1	2.1:1	2.2:1	1:4.7	1:1.3	1:1.1	1:1.1	1:1.0
II	1:1.3	1:1.8	1:1.6	1:2.7	1:3.0	1:2.9	1.2:1	1:1.7
	C_7^{ax}			α_L				
A	1:9.8	1:17.1		1.7:1	1.7:1			
I	2.7:1	3.1:1		1.1:1	1:1.2			
II	1.2:1	1:1.2		1.3:1	1:1.6			

where X labels the modes (X = A, I and II). The Q values are given in Table 7 (values for B3LYP and the hybrid SCC-DFTB/B3LYP are taken from Ref. [31] and Ref. [45], respectively). Compared to B3LYP the relation of the intensities is reversed for some modes. Further, the amide I mode of the C_5 conformer shows a much smaller ratio at the SCC-DFTB level than at the B3LYP level of theory, which is in agreement with MP2 results [45]. However, at the HF/6-31G* level of theory deviations in these ratios from the B3LYP or MP2 results are similar, whereas, when smaller basis sets are used in HF, like 4-31G, even larger deviations are obtained [45, 31]. The higher ratios of the intense modes at the SCC-DFTB level indicate that more intense modes are indeed overestimated compared to the modes with lower intensity. This shows that there is a systematic overestimation of the intense modes (or underestimation of the modes with low intensity), which possibly might be corrected by scaling the intensities.

5. DNA: H Bonding and Stacking

Base pair stacking is an even more challenging test than H bonding (see Fig. 6), since correlation is responsible for the stacking stabilization energies to a large extent and very diffuse functions are of primary importance for a proper description. Further, HF has been shown to reproduce stacking energies poorly and MP2 seems to overestimate correlation contributions by 15 to 30% [46]. Recently, there have been suggestions for constructions of density functionals to include the van der Waals interaction (see Ref. [47] and references therein).

Since within the SCC-DFTB model the dispersion interaction is clearly left out, we have chosen to include it empirically. To our knowledge, a first attempt in this dircetion was based on the Slater-Kirkwood approximation [48] as implemented by Lewis and Sankey [49]. We choose to add the London dispersion formula

$$E_{\text{dis}} = \sum_{i,j} \frac{(I_i \alpha_i)(I_j \alpha_j)}{(I_i + I_j) R_{ij}^6} \tag{7}$$

to the total energy Eq. (1). I_i and α_i are experimental values for the ionization potential and the polarizability of atom i, R_{ij} is the distance between atoms i and j. However,

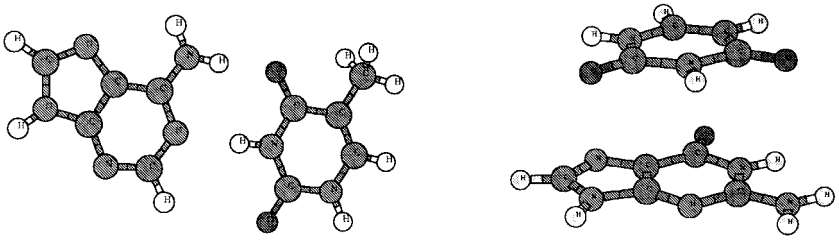

Fig. 6. The H-bonded ac base pair and the gu stacked base pair

in the intermediate distance region in the vicinity of the potential minimum (of the base–base interaction) the London formula is no longer valid due to the overlap of the charge densities, i.e. the interaction becomes too attractive. Therefore, we have chosen a scaling function in order to make E_{dis} vanish for small distances compared to the base pair equilibrium distance. Details will be given in a future publication [51]. The expression of E_{dis} has the advantage that it can be applied to all atoms in the system, and does not need to be restricted to inter-base pair interactions. Since analytic expressions for atomic forces are easily evaluated, geometry optimizations and molecular dynamic simulations can be performed in the standard way.

Stacking energies for DNA-base pairs have been evaluated at the MP2 level of theory with an a posteriori correction for BSSE using geometries resulting from empirical force field optimizations [40]. Empirical force fields are capable of satisfactorily describing the stacking interactions, although there are discrepancies between the force fields themselves and with respect to the MP2 results of up to 100% in the interaction energies [27]. The AM1 and PM3, as well as the MNDO/M method, have been shown to be unable to reproduce the attractive stacking interactions; for these methods, the interactions are erroneously found to be repulsive (2 to 10 kcal/mol), leading to a destabilization of the stacked base pairs [27].

At the SCC-DFTB level of theory, the interaction energies are attractive, although significantly underestimated [20]. The empirical correction leads to interaction energies which compare well with the MP2 results, see Table 8. We also investigated the radial and torsional dependence of the stacking interactions. The empirically extended SCC-DFTB model is able to reproduce these dependencies very accurately compared to the MP2 values [42].

Table 8

Stacking energies (kcal/mol) of base pairs for the SCC-DFTB, MP2 and SCC-DFTB with empirical inclusion of dispersion (see text)

	SCC-DFTB	MP2	SCC-DFTB + dispersion		SCC-DFTB	MP2	SCC-DFTB + dispersion
ga	2.6	11.2	9.0	gg	5.6	11.3	11.1
gu	6.5	10.6	11.4	aa	1.5	8.8	6.9
ac	3.0	9.5	8.1	cc	2.6	8.3	8.9
gc	6.0	9.3	10.7	uu	3.3	6.5	5.1
au	3.8	9.1	8.9	cu	5.3	8.5	9.3

6. Conclusions

We have presented several extensions of an approximate DFT scheme (SCC-DFTB) in order to improve the accuracy for biologically relevant molecular interactions with the aim to develop a highly efficient method for accurate large-scale simulations of biomolecules. The SCC-DFTB has been implemented into a QM/MM scheme, which has been shown to yield satifactory results for intermolecular H-bonded compounds. We have also extended the SCC-DFTB minimal basis set to include p-type basis functions on the hydrogen atoms. Tests for a large set of compounds relevant to biological molecules show a clear improvement in the description of H bonding energies, which now are in quantitative agreement with higher level ab initio results. The SCC-DFTB method has been benchmarked for different conformers of small model peptides. The geometries are shown to be reliably determined and the relative energies show the right ordering compared to B3LYP/6-31G* reference calculations, although the energy differences are slightly underestimated. We further discussed the performance of the SCC-DFTB method for vibrational frequencies and IR absorption for the conformers of the model peptide NA-LA-NMA. The determination of IR intensities is based on the Mulliken approximation for atomic point charges, which shows an overestimation of the more intense peaks. Trends in the absorption spectra are reproducible to some extent and a correction of the intensities via a scaling might be possible. Finally, we described briefly the inclusion of the dispersion interaction in an empirical manner. This extension leads to interaction energies comparable to those obtained at the MP2 level of theory. In summary, we have shown that the SCC-DFTB method is able to describe biologically relevant structures with high accuracy, while it is several orders of magnitude faster than ab initio methods.

References

[1] A. St-Amant, Density Functional Methods in Biomolecular Modeling, in: Reviews in Computational Chemistry, Vol. 7, Eds. K. B. Lipkowitz and D. B. Boyd, New York 1996 (p. 217).
[2] J. S. Dewar, E. Zoebisch, E. F. Healy, and J. J. P. Stewart, J. Amer. Chem. Soc. **107**, 3902 (1985).
[3] J. J. P. Stewart, J. Comput. Chem. **10**, 209, 221 (1989).
[4] M. C. Zerner, Semiempirical Molecular Orbital Methods, in: Reviews in Computational Chemistry, Eds. K. B. Lipkowitz and D. B. Boyd, New York 1990 (p. 45).
[5] Th. Frauenheim et al., phys. stat. sol. (b) **217**, 41 (2000).
[6] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, Phys. Rev. B **58**, 7260 (1998).
[7] M. Elstner, D. Porezag, G. Jungnickel, T. Frauenheim, S. Suhai, and G. Seifert, in: Tight-Binding Approach to Computational Materials Science, Eds. P. Turchi, A.Gonis, and L. Colombo, Mater. Res. Soc. Symp. Proc. **491**, 131 (1998).
[8] D. Porezag, T. Frauenheim, T. Köhler, G. Seifert, and R. Kaschner, Phys. Rev. B **51**, 12947 (1995).
[9] G. Galli, phys. stat. sol. (b) **217**, 231 (2000).
[10] P. Ordejón, phys. stat. sol. (b) **217**, 335 (2000).
[11] J. Gao, Methods and Applications of Combined Quantum Mechanical and Molecular Mechanical Potentials, in: Reviews in Computational Chemistry, Vol. 7, Eds. K. B. Lipkowitz and D. B. Boyd, New York 1996 (p. 119).
[12] M. Elstner, Ph.D. Thesis, University Paderborn, Paderborn (Germany) 1998.
[13] W. Han, M. Elstner, K. J. Jalkanen, T. Frauenheim, and S. Suhai, submitted to Internat. J. Quant. Chem.
[14] M. Elstner, S. M. Lee, Y. H. Lee, E. Kaxiras, and T. Frauenheim, in preparation.

[15] J. G. C. M. van Duijneveldt-van de Rijdt and F. B. van Duijneveldt, Ab initio Methods Applied to Hydrogen-Bonded Systems, in: Theoretical Treatment of Hydrogen Bonding, Ed. D. Hadzi, Wiley, New York 1997.
[16] S. F. Boys and F. Bernardi, Mol. Phys. **19**, 553 (1970).
[17] L. A. Curtiss, D. J. Frurip, and M. J. Lander, Chem. Phys. **71**, 2703 (1979).
[18] H. Guo, S. Sirois, E. I. Proynov, and D. R. Salahub, Density Functional Theory and Its Application to Hydrogen-Bonded Systems, in: Theoretical Treatment of Hydrogen Bonding, Ed. D. Hadzi, Wiley, New York 1997.
[19] D. Hadzi and J. Koller, Hydrogen Bonding by Semi-Empirical Molecular Orbital Methods, in: Theoretical Treatment of Hydrogen Bonding, Ed. D. Hadzi, Wiley, New York 1997.
[20] M. Elstner, D. Porezag, T. Frauenheim, S. Suhai, and G. Seifert, in: Multiscale Modelling of Materials, Eds. T. Diaz de la Rubia, T. Kaxiras, V. Bulatov, N. M. Ghoniem, and R. Phillips, Mater. Res. Soc. Symp. Proc. **538**, 243 (1999).
[21] M. J. S. Dewar and W. J. Thiel, J. Amer. Chem. Soc. **99**, 4899, 4907 (1977).
[22] A. A. Voityuk and A. A. Blizniuk, Theor. Chim. Acta **72**, 223 (1987).
[23] K. Jug and G. Geudtner, J. Comput. Chem. **14**, 639 (1993).
[24] B. K. Smith, D. J. Swanton, J. A. Pople, H. F. Schaefer, and L. Randon, J. Chem. Phys. **92**, 1240 (1990).
[25] S. Suhai, J. Phys. Chem. **99**, 1172 (1995).
[26] R. V. Stanton and K. M. Merz, J. Chem. Phys. **101**, 6658 (1994).
[27] P. Hobza et al., J. Comput. Chem. **18**, 1136 (1997).
[28] Y.-J. Zheng and K. M. Merz, J. Comput. Chem. **13**, 1151 (1992).
[29] J. E. Del Bene, J. Comput. Chem. **10**, 603 (1998).
[30] M. Karplus and D. L. Weaver, Protein Sci. **3**, 650 (1994).
[31] K. J. Jalkanen and S. Suhai, Chem. Phys. **208**, 81 (1996).
[32] K. Rommel-Möhle and H.-J. Hofmann, J. Mol. Struc. (Theochem) **285**, 211 (1993).
[33] W. Han, K. J. Jalkanen, M. Elstner, and S. Suhai, J. Phys. Chem. B **102**, 2587 (1998).
[34] C. L. Brooks III and D. A. Case, Chem. Rev. **93**, 2487 (1993).
[35] P. T. van Duijnen and B. T. Thole, Biopolymers **21**, 1749 (1982).
[36] A. Perczel, M. A. McAllister, P. Csaszar, and I. C. Csizmadia, J. Amer. Chem. Soc. **115**, 4849 (1993).
[37] H.-J. Böhm, J. Amer. Chem. Soc. **115**, 6152 (1993).
[38] K. Möhle, M. Gussmann, A. Rost, R. Cimiraglia, and H.-J. Hofmann, J. Phys. Chem. **101**, 8571 (1997).
[39] J. Richardson, Adv. Protein Chem. **43**, 167 (1991).
[40] J. Sponer, J. Leszczynski, and P. Hobza, J. Phys. Chem. **100**, 5590 (1996).
[41] K. Jalkanen, M. Elstner, S. Suhai, and T. Frauenheim, to be published.
[42] P. Hobza, M. Elstner, T. Frauenheim, E. Kaxiras, and S. Suhai, to be published.
[43] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, Gaussian 98, Revision A.5, Gaussian, Inc., Pittsburgh (PA) 1998.
[44] K. Jalkanen, M. Elstner, T. Frauenheim, and S. Suhai, submitted to Chem. Phys.
[45] H. G. Bohr, K. Frimand, K. J. Jalkanen, M. Elstner, and S. Suhai, Chem. Phys. **246**, 13 (1999).
[46] J. Sponer and P. Hobza, Chem. Phys. Lett. **267**, 263 (1997).
[47] J. F. Dobson and J. Wang, Phys. Rev. Lett. **82**, 2123 (1999).
M. Lein, J. F. Dobson, and E. K. U. Gross, J. Comput. Chem. **20**, 12 (1999).
[48] T. A. Halgren, J. Amer. Chem. Soc. **114**, 7827 (1992).
[49] J. P. Lewis and O. F. Sankey, Biophys. J. **69**, 1068 (1995).