

Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment

Marcus Elstner^{a)}

Physics Department, Harvard University, Cambridge, Massachusetts 02138
Department of Theoretical Physics, University of Paderborn, D-33098 Paderborn, Germany, and
Department Molecular Biophysics, German Cancer Research Center, D-69120 Heidelberg, Germany

Pavel Hobza

J. Heyrovsky Institute of Physical Chemistry, Academy of Sciences of the Czech Republic and Center for Complex Molecular Systems and Biomolecules, Doleskova 3, 18223 Prague, Czech Republic

Thomas Frauenheim

Department of Theoretical Physics, University of Paderborn, D-33098 Paderborn, Germany

Sándor Suhai

Department of Molecular Biophysics, German Cancer Research Center, D-69120 Heidelberg, Germany

Efthimios Kaxiras

Physics Department, Harvard University, Cambridge, Massachusetts 02138

(Received 4 August 2000; accepted 11 October 2000)

We extend an approximate density functional theory (DFT) method for the description of long-range dispersive interactions which are normally neglected by construction, irrespective of the correlation function applied. An empirical formula, consisting of an R^{-6} term is introduced, which is appropriately damped for short distances; the corresponding C_6 coefficient, which is calculated from experimental atomic polarizabilities, can be consistently added to the total energy expression of the method. We apply this approximate DFT plus dispersion energy method to describe the hydrogen bonding and stacking interactions of nucleic acid base pairs. Comparison to MP2/6-31G*(0.25) results shows that the method is capable of reproducing hydrogen bonding as well as the vertical and twist dependence of the interaction energy very accurately. © 2001 American Institute of Physics. [DOI: 10.1063/1.1329889]

I. INTRODUCTION

The interaction between closed shell molecules plays an important role in physics, chemistry, and especially in biology. Molecular interactions affect many biophysical and biochemical processes such as, for example, molecular recognition. Interaction between nucleic acid bases, responsible for the structure of DNA, which is of key importance for its function, is another typical and important example.

Much of our knowledge of the DNA structure has been obtained from oligonucleotide crystal structure studies.^{1,2} Other sources of information are nuclear magnetic resonance studies of DNA in solution.^{3,4} In recent years experimental techniques have been increasingly supplemented by computational and theoretical studies. This concerns application of quantum chemical as well as statistical methods. The main goal of quantum chemical calculations is to complement experiments and provide information and predictions which are not easily accessible by experimental techniques, in order to elucidate the nature of the processes studied.

DNA is a challenging system for quantum chemical studies because an accurate description of weak intermolecular interactions, manifested in hydrogen bonding (H bonding) and stacking interactions of DNA base pairs, is neces-

sary. Recently, *ab initio* methods were intensely applied to study hydrogen bonded and stacked nucleic acid dimers^{5,6} (for a review, see 7 and 8). H bonding can already be described to a reasonable accuracy at the Hartree-Fock (HF) level of theory using medium-sized basis sets. This is due to cancellation of errors concerning the basis set incompleteness error, the basis set superposition error, and the neglect of correlation (this has been described in great detail, e.g., for the water dimer⁹). Description of dispersion interactions responsible for the stabilization of the stacked pairs requires inclusion of the effects of correlation. Unfortunately, the convergence of the interaction energies with the basis set size and the order in perturbation theory is rather slow.¹⁰ Very large basis sets and treatment of correlation at least at the fourth-order Møller-Plesset or even coupled-cluster single double triple level of theory restricts the systems to be considered to the size of a few atoms. However, a careful analysis of the concurring effects allows the choice of a cheaper computational method, since the influence of basis set effects and higher orders of correlation on the stabilization energy have opposite sign. Therefore, the MP2/6-31G*(0.25) level has been proposed^{5,6} to describe the stacking interactions of DNA bases with good accuracy. Still, this method is limited to systems not larger than base trimers or tetramers due to its computational cost.

Semiempirical methods like the AM1¹¹ and PM3¹² mod-

^{a)}Electronic mail: m.elstner@dkfz.de

els are widely used for electronic structure calculations, since they are several orders of magnitude faster than *ab initio* methods. They are suitable methods for systems containing up to 1000 atoms which would allow treatment of larger fractions of DNA. Formally, these methods are an approximation to HF theory, but correlation effects are incorporated into the methods by the fitting procedure applied to determine the values of the parameters, which in principle would allow an application to DNA oligomers. However, while H bonding of base pairs is reasonably described using a special parametrization for H-bonded systems [modified neglect of differential overlap (MNDO)/M¹³], these methods fail in the stacking problem, predicting repulsive instead of attractive interactions between the base pairs.⁶

In contrast, several empirical force fields have been shown to describe H bonding and stacking very accurately.⁶ Therefore, recent force fields represent a considerable step forward toward an accurate description of nonbonding interactions and allow realistic simulations of larger DNA fragments. There has been enormous improvement of the quality of empirical force fields which formerly yielded unrealistic results for hydrated oligonucleotides in terms of structure and energy.¹⁴ Additional improvements seem to be necessary in order to describe, e.g., the pyramidalization of amino groups, which allows for an additional flexibility in H bonding. Moreover, polarization is not covered in recent force fields, which is important to describe, e.g., protonated base pairs, polarization of the bases due to interaction with metal cations, or the nonadditivity of base pair stacking along the double helix (for a review, see Ref. 8).

All the aforementioned effects can be captured by quantum mechanical methods and, although there are attempts to include polarization in empirical force fields, it would be desirable to have a quantum mechanical method suitable for applications to larger systems. Density functional theory (DFT) has become popular in the last decade due to its lower computational cost compared to *ab initio* methods. Since correlation is included from the very beginning, it provides highly accurate results for many molecular properties. While the inclusion of gradient corrected approximation (GGA) terms into the exchange–correlation functional E_{xc} significantly improves the description of H bonding over the local density approximation (LDA) formalism, the situation is very different for application to complexes stabilized by dispersion interactions. Although the exact E_{xc} must also include dispersion, currently available LDA and GGA functionals do not cover the correlation between distant dipolar density fluctuations, i.e., they do not reproduce the well-known R^{-6} behavior of two separate neutral (nonoverlapping) fragments.¹⁵ Therefore, novel approaches try to incorporate dispersion into DFT by developing new functionals.^{16–19} An alternative approach for the calculation of the corresponding dispersion coefficient C_6 is provided by determining the frequency dependent polarizabilities with time-dependent DFT.²⁰

Although the dispersion interaction arises from the correlation energy, Zhang and Yang²¹ showed that the behavior of the exchange functional for small densities and large density gradients is crucial for binding in van der Waals (vdW)

systems. While LDA overbinds rare gas diatomic molecules, some GGA functionals result in repulsive interactions.²¹ Similar results have been described for the interaction of benzene with organic dimers.²² For the systems considered in these studies, LDA seems to overbind, the BECKE88²³ exchange leads to repulsive interactions,^{24,25,21,22,26–28} while other exchange correlation functionals like the PW91²⁹ and PBE³⁰ lead in some cases to good agreement with post-HF results. Using the HF exchange functional, Perez-Jorda *et al.* tested the performance of several correlation functionals for the binding of rare gas dimers.³¹ Except for one case, all other tested functionals lead to significant overbinding or underbinding. Attempts to improve the long-range behavior of the exchange functionals by modifying existing functionals could improve the performance significantly.^{27,32} Kafafi³² suggested a modification and combination of existing functionals, the K2-BVWN methodology, where the gradient corrections to LDA are eliminated and an approximate exchange term is combined with exact exchange. The important point is that the resulting correlation functional has the right distance dependence of R^{-n} ($n=4,6,8$), that is, the London dispersion interaction is recovered. The new functional is able to describe qualitative trends in binding energies of rare gas pairs, however, quantitatively the functional still underestimates or overestimates the energies depending on the basis set used. For the application of a DFT approach to larger scale calculations on DNA oligomers, several obstacles seem to be relevant. First, despite some success in several cases, the description of vdW complexes is still qualitative, and there is no clear picture to which extent the dispersion is covered by the particular exchange–correlation functional employed. Further, the convergence of the interaction energies with basis set size is rather slow. Fairly large basis sets are needed which in turn limit the system size significantly.

In order to apply density functional theory to study DNA oligomers, we decided to follow an alternative approach. We augment an approximate DFT method with an empirical dispersion term for the description of vdW complexes. A similar approach has been followed by Lewis and Sankey³³ in the framework of an approximate DFT-LDA method, and Meijer and Sprik tested the performance of an additional empirical dispersion contribution for the interaction of two benzene molecules using both LDA and GGA functionals.³⁴ Gianturco *et al.* have added dispersion forces from results of perturbation theory to DFT and suggested a procedure to scale these contributions for small distances.³⁵

The approximate DFT method has been described in detail elsewhere.^{36–38} It is comparable in computational speed to the AM1 and PM3 methods and is derived from DFT by an expansion of the DFT total energy up to second order in the charge density fluctuations around a reference density. The subsequent approximations lead to a generalized eigenvalue problem, which has to be solved self-consistently for atomic charges. The method can be seen as a general extension of so-called tight binding (TB) methods to charge self-consistency. All parameters of this model are calculated from DFT, and the method is therefore called a self-consistent-charge, density-functional tight-binding method (SCC-

DFTB). The results for reaction energies, geometries and frequencies for small organic molecules have been presented elsewhere.^{36,37} The mean average deviations from experimental values are comparable to DFT calculations. The method has also been benchmarked for biologically relevant molecules, H-bonded complexes, small peptides, and DNA H-bonding interactions.^{38,39} The benchmarks performed so far have been quite satisfactory, showing that the SCC-DFTB method is able to give a reliable description of several biological molecules. In this work, we describe an extension of the SCC-DFTB method by augmenting the model with a London dispersion term and show the results for H-bonded and stacked DNA base pairs.

The paper is organized as follows: In Sec. II we describe briefly the SCC-DFTB model only to the extent necessary to derive the extension for the dispersion term. In Sec. III we describe the choice of the parameters entering the dispersion energy and the damping of the R^{-6} term for small interatomic distances, while in Sec. IV we present results for H bonded and stacked base pairs. We conclude in Sec. V.

II. METHOD

The SCC-DFTB model is derived from density functional theory (DFT) by a second-order expansion of the DFT total energy functional with respect to the charge density fluctuations δn around a given reference density n_0 :

$$E = \sum_i^{\text{occ}} \langle \psi_i | \hat{H}^0 | \psi_i \rangle + \frac{1}{2} \int \int \left(\frac{1}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta^2 E_{\text{xc}}}{\delta n(\mathbf{r}) \delta n(\mathbf{r}')} \Big|_{n_0} \right) \times \delta n(\mathbf{r}) \delta n(\mathbf{r}') d\mathbf{r} d\mathbf{r}' - \frac{1}{2} \int \int \frac{n_0(\mathbf{r}) n_0(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + \int [E_{\text{xc}}[n_0] - V_{\text{xc}}[n_0]] n_0(\mathbf{r}) d\mathbf{r} + E_{ii}. \quad (1)$$

$\hat{H}^0 = \hat{H}[n_0]$ is the effective Kohn–Sham Hamiltonian evaluated at the reference density and the ψ_i are Kohn–Sham orbitals. E_{xc} and V_{xc} are the exchange–correlation energy and potential, respectively, and E_{ii} is the core–core repulsion energy.

To derive the total energy of the SCC-DFTB method, the energy contributions in Eq. (1) are further subjected to the following approximations.

(a) The Hamiltonian matrix elements $\langle \psi_i | \hat{H}^0 | \psi_i \rangle$ are represented in a minimal basis of confined, pseudoatomic orbitals ϕ_μ ,

$$\psi_i = \sum_\mu c_\mu^i \phi_\mu.$$

To determine the basis functions ϕ_μ , we solve the DFT atomic problem by adding an additional harmonic potential⁴⁰ to confine the basis functions. The Hamiltonian matrix elements in this basis, $H_{\mu\nu}^0$, are then calculated as follows: The diagonal elements $H_{\mu\mu}^0$ are taken to be the atomic eigenvalues and the nondiagonal elements $H_{\mu\nu}^0$ are calculated in a two-center approximation,

$$H_{\mu\nu}^0 = \langle \phi_\mu | \hat{T} + V_{\text{eff}}[n_\alpha^0 + n_\beta^0] | \phi_\nu \rangle, \quad \mu \in \alpha, \nu \in \beta,$$

which are tabulated together with the overlap matrix elements $S_{\mu\nu}$ with respect to the interatomic distance $R_{\alpha\beta}$. V_{eff} is the effective Kohn–Sham potential and n_α^0 are the densities of the neutral atoms α ; \hat{T} is the kinetic energy operator.

(b) The charge density fluctuations δn are written as a superposition of atomic contributions δn_α ,

$$\delta n = \sum_\alpha \delta n_\alpha,$$

which are approximated by the charge fluctuations at the atoms α , $\Delta q_\alpha = q_\alpha - q_\alpha^0$. q_α^0 is the number of electrons of the neutral atom α and the q_α are determined from a Mulliken population analysis. The second order term in δn of the energy in Eq. (1) is then approximated by $\sum_{\alpha\beta} \gamma_{\alpha\beta} \Delta q_\alpha \Delta q_\beta$. For $\alpha \neq \beta$, $\gamma_{\alpha\beta}$ is determined analytically from the Coulomb interaction of two spherical charge distributions, located at R_α and R_β . The onsite contributions $\gamma_{\alpha\alpha}$ are calculated as $\partial E_\alpha^2 / \partial^2 q_\alpha$.

(c) The double counting terms and the core–core repulsion energy are represented in the term E_{rep} , which is approximated as a sum over all pairs of a two-body interaction potential $U(r)$:

$$E_{\text{rep}} = \sum_{\alpha \neq \beta} U[R_{\alpha\beta}].$$

With these definitions and approximations, the SCC-DFTB total energy finally reads:

$$E_{\text{tot}} = \sum_{i\mu\nu} c_\mu^i c_\nu^i H_{\mu\nu}^0 + \frac{1}{2} \sum_{\alpha\beta} \gamma_{\alpha\beta} \Delta q_\alpha \Delta q_\beta + E_{\text{rep}}. \quad (2)$$

In the determination of the Hamiltonian matrix elements we use the Perdew–Burke–Ernzerhof²⁹ exchange–correlation functional, which does not show excessive repulsion for vdW complexes.²¹

Applying the variational principle, one obtains a set of approximate Kohn–Sham equations, which have to be solved iteratively for the wave function expansion coefficients c_μ^i , since the Hamiltonian matrix elements depend on the c_μ^i due to the Mulliken charges. Analytic first derivatives are readily obtained, and second derivatives of the energy with respect to atomic positions are calculated numerically. The two-body interaction potential $U(r)$ is calculated by subtracting the DFT total energy from the SCC-DFTB electronic energy as a function of the bond distance $r = R_{\alpha\beta}$ for a suitable reference system.

III. DISPERSION ENERGY

To discuss the extension of the method, we consider for simplicity the case of vanishing charge fluctuations $\Delta q_\alpha = 0$ and assume for the moment E_{xc} to be the true exchange–correlation functional, where the dispersion interaction would be included. In this case the total energy is given by

$$E_{\text{tot}} = \sum_{i\mu\nu} c_\mu^i c_\nu^i H_{\mu\nu}^0 + E_{\text{rep}} = \sum_i^{\text{occ}} \epsilon_i + E_{\text{rep}} \quad (3)$$

with

$$E_{\text{rep}} = -\frac{1}{2} \int \int \frac{n_0(\mathbf{r}n_0(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r} d\mathbf{r}' + \int [E_{\text{xc}}[n_0] - V_{\text{xc}}[n_0]]n_0(\mathbf{r})d\mathbf{r} + E_{ii},$$

the ϵ_i are the eigenvalues of the approximate Kohn–Sham equations. Inspection of Eq. (3) shows that the asymptotic R^{-6} behavior in the case of nonoverlapping densities of the atomic fragments for the true exchange–correlation functional comes from $E_{\text{xc}}[n_0]$. In the SCC-DFTB model, this is contained in E_{rep} , since $\sum_i \epsilon_i$ decays exponentially with increasing interatomic distance and is zero in the case of vanishing overlap. As described previously, E_{rep} is determined from DFT calculations by taking the difference of the SCC-DFTB and DFT binding curves. In practice, E_{rep} is assumed to be short ranged and purely repulsive, it approaches zero for distances between first and second nearest neighbor distances of the considered atom pairs, which is about 2.5–3.0 Å for atoms of the first row of the Periodic Table. On the other hand, E_{rep} containing the true E_{xc} functional should have an R^{-6} behavior for nonoverlapping atomic densities, that is, the case of atoms of the first row for $R \geq 4-5$ Å.

Practically, it is not possible to obtain an E_{rep} containing the R^{-6} interaction with the procedure described previously, since DFT does not cover dispersion properly. Therefore, we decided to add a London-type dispersion energy for large distances, and include a damping of this term with the onset of overlap of the charge density. Similar arguments can be applied for an extension of DFT methods using current LDA or GGA functionals where E_{xc} vanishes for zero overlap of the two atomic or molecular fragments.¹⁵

The London⁴¹ dispersion energy is given by

$$E_{\text{dis}} = \frac{3}{2} \frac{I_\alpha I_\beta p_\alpha p_\beta}{(I_\alpha + I_\beta) R_{\alpha\beta}^6}, \quad (4)$$

where I_α is the ionization potential and p_α is the polarizability of atom α . Kang and Jhon⁴⁴ determined atomic polarizabilities for several elements and showed that they can describe molecular polarizabilities very accurately assuming additivity. However, vdW coefficients are slightly underestimated when calculated with Eq. (4). Recently Halgren⁴² has reviewed the treatment of vdW interactions in the framework of empirical potentials and has suggested a scheme to calculate vdW coefficients (C_6) from atomic polarizabilities. The advantage of this scheme is that only the atomic polarizabilities are needed, which makes it easily applicable to systems containing any type of atoms.

A. Calculation of C_6 coefficients

Given the atomic polarizabilities, the C_6 coefficients are calculated by

$$C_6^\alpha = 0.75 \sqrt{N_\alpha p_\alpha^3} \quad (5)$$

with N_α the Slater–Kirkwood effective number of electrons. Halgren⁴² proposes an algorithm for calculating N_α with $N_\alpha = 0.8$ for hydrogen and

$$N_\alpha = 1.17 + 0.33n_v^\alpha \quad (6)$$

TABLE I. Static atomic polarizabilities (in Å³) from Miller (Ref. 43) and Kang and Jhon (Ref. 44), and ionization potentials I_p (eV). The types of atoms follow the definitions of Ref. 43.

	H	CTR	CTE	NTR2	NPI2	O=C
Reference 43	0.387	1.352	1.061	1.030	1.090	0.569
Reference 44	0.386	1.382	1.064	1.030	1.090	0.46
I_p	13.61	11.22	14.57	14.51	12.25	17.25

for C to Ne atoms, where n_v^α is the number of valence electrons for atom α . For diatomic coefficients $C_6^{\alpha\beta}$ Halgren obtained the best results using the Slater–Kirkwood combination rule:

$$C_6^{\alpha\beta} = \frac{2C_6^\alpha C_6^\beta p_\alpha p_\beta}{p_\alpha^2 C_6^\alpha + p_\beta^2 C_6^\beta}. \quad (7)$$

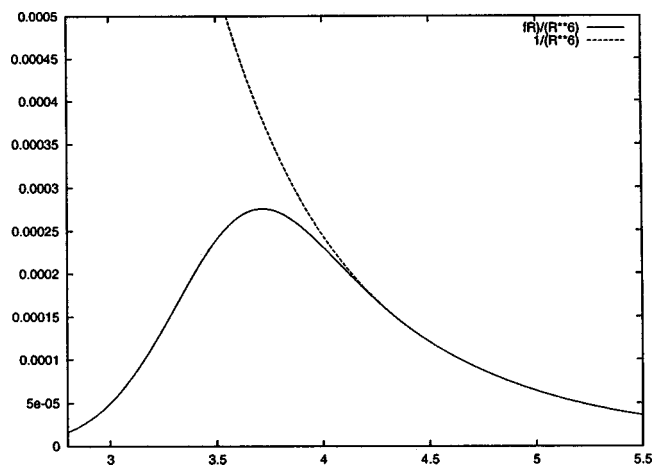
Molecular polarizabilities can be represented with good accuracy as a sum of atomic contributions, when the hybrid states of atoms are taken into account (see, e.g., Refs. 43 and 44). For the atom types of interest here, the atomic polarizabilities from Miller⁴³ and Kang and Jhon⁴⁴ are comparable, as shown in Table I. Table II shows the C_6^α values calculated from Eq. (4) using the data from Kang and Jhon⁴⁴ and from Eqs. (5) and (6) using the polarizabilities given by Miller.⁴³ The latter are slightly larger, which should be in better accord with experimental values as mentioned before.⁴⁴ Further, the latter scheme has the advantage of being easily extended to other types of atoms, since only the atomic polarizabilities are needed and similar prescriptions as Eq. (6) are given for elements in the second and third rows of the Periodic Table.

B. Damping of the R^{-6} function

The London dispersion formula, i.e., the R^{-6} dependence of the interaction energy, has been derived for the case of the interaction of two systems with nonoverlapping densities and is, therefore, strictly valid only in this regime. Further, the current LDA and GGA correlation functionals give zero contribution to the binding energy in this case of vanishing electron density overlap, while the effects of correlation are covered for smaller interatomic distances, i.e., in the binding region. Concerning the SCC-DFTB method, the latter effects are included in E_{rep} [Eq. (3)], while the long-range R^{-6} dependence is clearly missing. Therefore, the R^{-6} term should be damped for small distances, where the electronic densities start to overlap. An appropriate damping function is given by

TABLE II. C_6^α coefficients (eVÅ⁶): SLKI: C_6 calculated from atomic polarizabilities (Ref. 43) using the Slater–Kirkwood formula. LOND: C_6 calculated from London dispersion formula using values from Kang and Jhon (Ref. 44) for the ionization potentials and polarizabilities. The types of atoms follow the definitions of Ref. 43.

Conf.	H	CTR	CTE	NTR2	NPI2	O=C
SLKI	1.61	18.56	12.93	13.16	14.33	5.71
LOND	1.53	16.07	12.37	11.55	10.92	4.15

FIG. 1. Damping function $f(r)$ as described in the text.

$$f(R) = [1 - \exp(-d^*(R/R_0)^N)]^M. \quad (8)$$

A similar damping function has been used in earlier HF plus dispersion calculations (for a review see Ref. 42, and references therein) and recently for the development of intermolecular potentials.⁴⁵ Figure 1 shows the function $f(R)R^{-6}$ and R^{-6} with the values $d=3.0$, $N=7$, $M=4$, and $R_0=3.8$ Å. It can be seen, that $f(R)R^{-6}$ starts deviating from R^{-6} at R_0 and approaches zero for $R \approx 3$ Å, the desired behavior as discussed previously. We use Eq. (8) with the same values d , N , and M for all types of atoms. R_0 is defined by the range of the overlap of two atoms. We use $R_0=3.8$ Å for first row elements and $R_0=4.8$ Å for second row elements. Combinations of R_0 are given by the cubic mean rule,⁴²

$$R_0^{\alpha\beta} = \frac{(R_0^\alpha)^3 + (R_0^\beta)^3}{(R_0^\alpha)^2 + (R_0^\beta)^2}.$$

With Eqs. (7) and (8), the total energy of the SCC-DFTB model augmented with a damped R^{-6} finally becomes

$$E_{\text{tot}} = \sum_{i\mu\nu} c_{\mu}^i c_{\nu}^i H_{\mu\nu}^0 + \frac{1}{2} \sum_{\alpha\beta} \gamma_{\alpha\beta} \Delta q_{\alpha} \Delta q_{\beta} + E_{\text{rep}} + E_{\text{dis}}, \quad (9)$$

$$E_{\text{dis}} = - \sum_{\alpha\beta} f(R_{\alpha\beta}) C_{\alpha\beta}^{\alpha\beta} (R_{\alpha\beta})^{-6}.$$

TABLE III. Interaction energies (kcal/mol) of H-bonding base pairs at the SCC-DFTB, MP2, and SCC-DFTB + E_{dis} levels of theory, as described in the text.

	SCC-DFTB	MP2	SCC-DFTB + E_{dis}		SCC-DFTB	MP2	SCC-DFTB + E_{dis}
gcwc	21.0	25.4	23.9	atwc	9.7	12.4	11.8
gg1	22.5	24.0	24.7	atrwc	9.5	12.4	11.7
cc	13.6	18.8	15.8	aa1	8.7	11.5	10.9
gg3	12.5	17.1	15.2	ga4	8.4	11.1	10.8
ga1	12.3	15.7	14.8	tc2	9.1	11.8	11.3
gt1	14.4	14.7	16.3	tc1	8.8	11.6	10.9
gt2	13.9	14.3	15.7	aa2	7.7	11.0	10.3
ac1	10.5	14.3	12.7	tt2	10.4	10.6	11.5
gc1	10.3	13.9	12.9	tt1	10.3	10.6	12.0
ac2	9.5	14.1	12.4	tt3	10.2	10.5	11.8
ath	9.5	13.3	11.9	ga2	7.5	10.4	10.3
ga3	10.4	15.2	13.2	gg4	6.9	10.3	9.7
atrh	9.5	13.2	11.8	aa3	6.8	10.0	9.7

With this formulation, the calculation of energy derivatives with respect to atomic positions, which give the forces acting on the atoms, is straightforward.

IV. RESULTS

A. H-bonded DNA base pairs

In a previous publication we presented the interaction energies for 26 H-bonded base pairs³⁸ and compared them with the results from Hobza *et al.*,⁶ who optimized the geometries of all complexes at the HF level of theory, while the interaction energies were calculated at the MP2/6-31G*(0.25) level of theory with subsequent correction for the basis set superposition error (BSSE).⁵ In Ref. 6, a set of empirical force fields have been tested as well as the semi-empirical AM1 and PM3 methods. While most of the force field methods show a very small average error compared to the second-order Møller–Plesset (MP2) reference interaction energies of the 26 base pairs, leading to mean average errors of 0.9–2.4 kcal/mol, the semiempirical methods AM1 and PM3 show much larger deviations. The mean average errors are 7.3 and 6.3 kcal/mol for the AM1 and PM3 methods, respectively. However, the MNDO/M is in good agreement with the MP2 results with a mean average error of only 2.5 kcal/mol. The SCC-DFTB interaction energies evaluated with Eq. (2) are slightly underestimated, but the relative stabilities are reproduced well, leading to a mean average error of 2.8 kcal/mol. Introduction of the additional dispersion energy according to Eq. (9) increases the interaction energies. The SCC-DFTB, MP2, and SCC-DFTB + E_{dis} results are presented in Table III. The mean average error with respect to the MP2 values decreases to 1.1 kcal/mol. The geometries of the H-bonded pairs were determined by means of unconstrained geometry optimizations. All intra- as well as intermolecular degrees of freedom have been optimized.

B. Stacked DNA base pairs

To calculate interaction energies for the stacked pairs, we followed the procedure of Ref. 6. We consider undisplaced dimers, i.e., the centers of both monomers are stacked one above the other with coplanar bases. The subsystem ge-

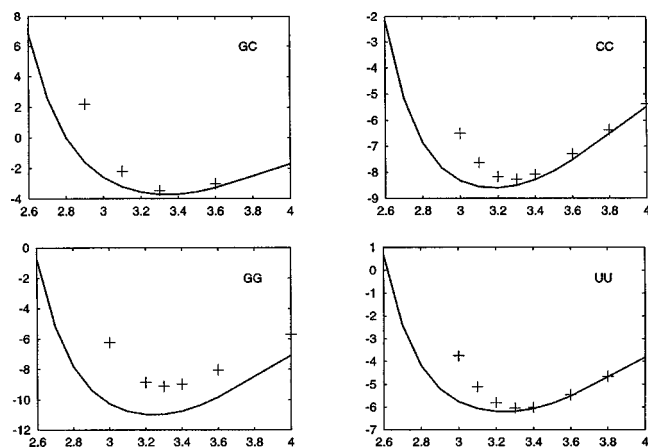


FIG. 2. Dependence of the interaction energy (in kcal/mol) on the vertical separation (in Å) for the different base pairs. The solid line corresponds to the SCC-DFTB+ E_{dis} model; the points relate to the MP2 energies as described in the text.

ometries were frozen from MP2/6-31G* optimization. We investigated the dependence of stacking energy on the vertical distance and on the twist angle. The twist between two bases is defined in the following way: The two bases are initialized in such a way that their centers of mass coincide and the glycosidic bond vectors are parallel. Then a twist is a counter-rotation of the two bases around the axis passing through the centers of mass of the bases and perpendicular to their plane. A twist of 0° corresponds to a parallel, a twist of 180° corresponds to an antiparallel arrangement of the bases. The twist dependence was investigated for a vertical distance of 3.4 Å. Stacking energies have been evaluated at the MP2/6-31G*(0.25) level of theory with a *posteriori* correction for BSSE at geometries resulting from MP2/6-31G* optimization of the respective monomers.⁶

The dependence of the stacking energies on the vertical distance for a fixed twist angle of 180° (antiparallel arrangement of the bases) is given in Fig. 2. Without the dispersion correction, SCC-DFTB interactions are sometimes completely repulsive, while the E_{dis} correction lowers the energies leading to very good agreement with the MP2 results (Fig. 3). The main difference appears in the curvature of the total energy: the SCC-DFTB+ E_{dis} total-energy curves are slightly flatter than the MP2 ones and the equilibrium distances are about 0.1–0.15 Å shorter. The overall agreement with the MP2 data, however, is surprisingly good.

Figures 3 and 4 show the dependence of the base stacking energies of all ten stacked dimers with respect to the twist angle. All three curves (SCC-DFTB, SCC-DFTB+ E_{dis} , MP2) exhibit a similar twist dependence. This is not surprising since the twist dependence is mainly governed by the electrostatic energy, which qualitatively can be described by a simple point charge model.⁴⁶ The SCC-DFTB energy curve is too repulsive since the dispersion energy is not covered at this level of theory. SCC-DFTB+ E_{dis} overestimates the stabilization energy with respect to the MP2 values (by about 1.5 kcal/mol), but the overall agreement with the MP2 data is again very good.

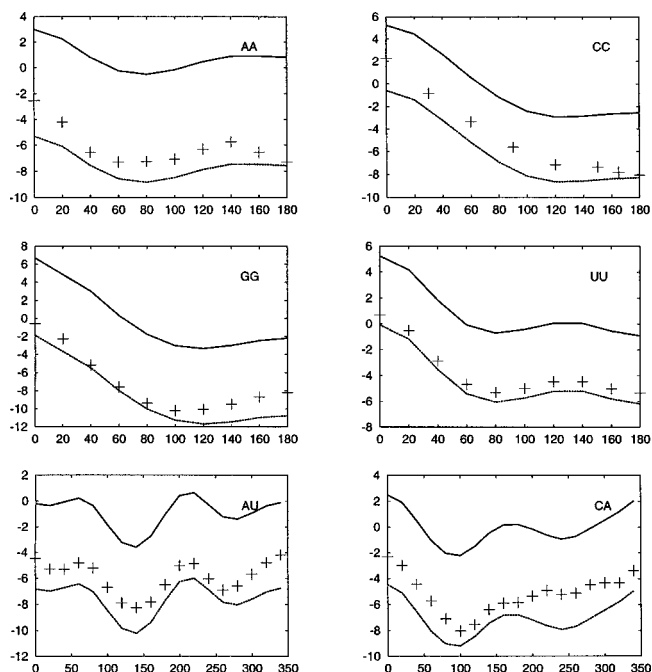


FIG. 3. Dependence of the interaction energy (in kcal/mol) on the twist angle (in deg) for the different base pairs. The upper solid line (higher in energy) corresponds to the SCC-DFTB, the lower one to the SCC-DFTB+ E_{dis} model, and the crosses relate to the MP2 energies as described in the text.

V. CONCLUSION

We have augmented the approximate DFT method (SCC-DFTB) with a pairwise dispersion term to account for van der Waals interactions. The R^{-6} function is truncated for small distances with an appropriate damping function which is similar to the HF+ E_{dis} approach. For the calculation of the corresponding C_6 coefficients we utilize a simple prescription using static atomic polarizabilities. In principle, a similar procedure can be applied to full DFT calculations. We applied this extension to H bonding and stacking of DNA base pairs. In the case of H-bonding energies, we found a

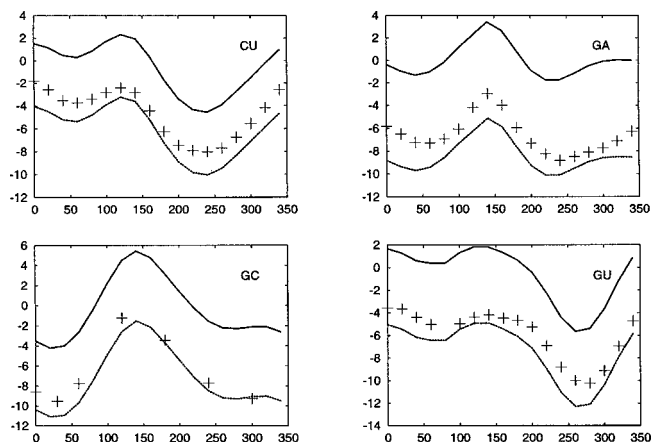


FIG. 4. Dependence of the interaction energy (in kcal/mol) on the twist angle (in deg) for the different base pairs. The upper solid line (higher in energy) corresponds to the SCC-DFTB, the lower one to the SCC-DFTB+ E_{dis} model, and the points relate to the MP2 energies as described in the text.

slight improvement compared to the values evaluated with the SCC-DFTB model. For the stacking, however, the correction changes the performance of the method qualitatively, since the SCC-DFTB gives unphysical results for the base stacking, which is unstable (repulsive interactions). After inclusion of the approximate dispersion energy we obtain very good agreement with the MP2 reference calculations for the dependence of the energy on the vertical distance as well as the twist angle. This approximate DFT + E_{dis} method thus represents a DFT-based method applicable to DNA stacking. It is expected that the computational efficiency of this method in combination with QM/MM⁴⁷ and $O(N)$ ⁴⁸ algorithms will allow its application to studies of the structure and dynamics of DNA and the interaction of DNA with other molecules.

ACKNOWLEDGMENTS

We would like to thank G. Seifert and H. Liu for helpful discussions. P.H. thanks the Ministry of Education of the Czech Republic for support through Project No. LN00A032 (Center for Complex Molecular Systems and Biomolecules).

- ¹O. Kennard and W. N. Hunter, *Angew. Chem. Int. Ed. Engl.* **30**, 1254 (1991).
- ²C. Wahl and M. Sundaralingham, *Biopolymers* **44**, 45 (1997).
- ³F. W. Smith and J. Feigan, *Biochemistry* **32**, 8682 (1993).
- ⁴G. D. Strahan, M. A. Keniry, and R. H. Shafer, *Biophys. J.* **75**, 968 (1998).
- ⁵J. Spomer, J. Leszczynski, and P. Hobza, *J. Phys. Chem.* **100**, 5590 (1996).
- ⁶P. Hobza *et al.*, *J. Comput. Chem.* **18**, 1136 (1997).
- ⁷J. Spomer, J. Leszczynski, and P. Hobza, *J. Biomol. Struct. Dyn.* **14**, 117 (1996).
- ⁸P. Hobza and J. Spomer, *Chem. Rev.* **99**, 3247 (1999).
- ⁹J. G. C. M. van Duijneveldt-van de Rijdt, and F. B. van Duijneveldt, in *Theoretical Treatment of Hydrogen Bonding*, edited by D. Hadzi (Wiley, New York, 1997).
- ¹⁰J. Spomer and P. Hobza, *Chem. Phys. Lett.* **267**, 263 (1997).
- ¹¹J. S. Dewar, E. Zebisch, E. F. Healy, and J. J. P. Stewart, *J. Am. Chem. Soc.* **107**, 3902 (1985).
- ¹²J. J. P. Stewart, *J. Comput. Chem.* **10**, 209 (1989); **10**, 221 (1989).
- ¹³A. A. Voityuk and A. A. Blizniuk, *Theor. Chim. Acta* **72**, 223 (1987).
- ¹⁴P. Cieplak, in *The Encyclopedia of Computational Chemistry*, edited by P. R. Schleyer, N. L. Allinger, T. Clark, J. Gasteiger, P. A. Kollman, H. F. Schäfer III, and P. R. Schreiner (Wiley, Chichester, 1998), p. 1922.
- ¹⁵J. F. Dobson, B. P. Dinte, and J. Wang, in *Electronic Density Functional Theory: Recent Progress and New Directions*, edited by J. F. Dobson, G. Vignale, and M. P. Das (Plenum, New York, 1997).
- ¹⁶B. I. Lundquist, Y. Andersson, S. Shao, S. Chan, and D. C. Langreth, *Int. J. Quantum Chem.* **56**, 247 (1995); T. Andersson, D. C. Langreth, and B. I. Lundquist, *Phys. Rev. Lett.* **76**, 102 (1996); E. Hult, H. Rydberg, B. I. Lundquist, and D. C. Langreth, *Phys. Rev. B* **59**, 4708 (1999).
- ¹⁷J. F. Dobson and J. Wang, *Phys. Rev. Lett.* **82**, 2123 (1999).
- ¹⁸M. Lein, J. F. Dobson, and E. K. U. Gross, *J. Comput. Chem.* **20**, 12 (1999).
- ¹⁹W. Kohn, Y. Meir, and D. E. Markov, *Phys. Rev. Lett.* **80**, 4153 (1998).
- ²⁰V. P. Osinga, S. J. A. van Ginsbergen, J. G. Snijders, and E. J. Baerends, *J. Chem. Phys.* **106**, 5091 (1997).
- ²¹Y. Zhang, W. Pan, and W. Yang, *J. Chem. Phys.* **107**, 7921 (1997).
- ²²T. A. Weslowski, O. Parisel, Y. Ellinger, and J. Weber, *J. Phys. Chem. A* **101**, 7818 (1997).
- ²³A. D. Becke, *Phys. Rev. A* **38**, 3098 (1988).
- ²⁴S. Kristyan and P. Pulay, *Chem. Phys. Lett.* **229**, 175 (1994).
- ²⁵J. Perez-Jorda and A. D. Becke, *Chem. Phys. Lett.* **233**, 134 (1995).
- ²⁶S. Tsuzuki, T. Uchimaru, and K. Tanabe, *Chem. Phys. Lett.* **287**, 202 (1998).
- ²⁷C. Adamo and V. Barone, *J. Chem. Phys.* **108**, 664 (1998).
- ²⁸P. Hobza, J. Spomer, and T. Reschel, *J. Comput. Chem.* **16**, 1315 (1995).
- ²⁹J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, *Phys. Rev. B* **46**, 6671 (1992).
- ³⁰J. P. Perdew, K. Burke, and M. Ernzerhof, *Phys. Rev. Lett.* **77**, 3865 (1996).
- ³¹J. M. Perez-Jorda, E. San-Fabian, and A. J. Perez-Jimenez, *J. Chem. Phys.* **110**, 1916 (1999).
- ³²S. A. Kafafi, *J. Phys. Chem. A* **102**, 10404 (1998).
- ³³J. P. Lewis and O. F. Sankey, *Biophys. J.* **69**, 1068 (1995).
- ³⁴E. J. Meijer and M. Sprik, *J. Chem. Phys.* **105**, 8684 (1996).
- ³⁵F. A. Gianturco, F. Paesani, M. F. Laranjeira, V. Vassilenko, and M. A. Cunha, *J. Chem. Phys.* **110**, 7832 (1999).
- ³⁶M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, *Phys. Rev. B* **58**, 7260 (1998).
- ³⁷M. Elstner, D. Porezag, G. Jungnickel, T. Frauenheim, S. Suhai, and G. Seifert, in *Tight-binding Approach to Computational Materials Science*, edited by P. Turchi, A. Gonis, and L. Colombo, [Mater. Res. Soc. Symp. Proc. **491**, 131 (1998)].
- ³⁸M. Elstner, D. Porezag, T. Frauenheim, S. Suhai, and G. Seifert, in *Multiscale Modelling of Materials*, edited by T. Diaz de la Rubia, T. Kaxiras, V. Bulatov, N. M. Ghoniem, and R. Phillips, [Mater. Res. Soc. Symp. Proc. **538**, 243 (1999)].
- ³⁹M. Elstner, T. Frauenheim, E. Kaxiras, G. Seifert, and S. Suhai, *Phys. Status Solidi B* **217**, 357 (2000).
- ⁴⁰D. Porezag, T. Frauenheim, T. Köhler, G. Seifert, and R. Kaschner, *Phys. Rev. B* **51**, 12947 (1995).
- ⁴¹F. London, *Z. Phys. Chem. Abt. B* **11**, 22 (1930).
- ⁴²T. A. Halgren, *J. Am. Chem. Soc.* **114**, 7827 (1992).
- ⁴³K. J. Miller, *J. Am. Chem. Soc.* **112**, 8533 (1990).
- ⁴⁴Y. K. Kang and M. S. Jhon, *Theor. Chim. Acta* **61**, 41 (1982).
- ⁴⁵W. T. M. Mooij, F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt, and B. P. van Eijck, *J. Phys. Chem.* **103**, 9872 (1999).
- ⁴⁶J. Spomer, J. Leszczynski, and P. Hobza, *J. Comput. Chem.* **17**, 841 (1996).
- ⁴⁷W. Han, M. Elstner, K. J. Jalkanen, T. Frauenheim, and S. Suhai, *Int. J. Quantum Chem.* **78**, 459 (2000); Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim, and M. Karplus *J. Phys. Chem. B* **105**, 569 (2001).
- ⁴⁸H. Liu, M. Elstner, E. Kaxiras, T. Frauenheim, J. Hermans, and W. Yang (unpublished).