JCTC Journal of Chemical Theory and Computation

Semiempirical Quantum Chemical PM6 Method Augmented by Dispersion and H-Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes

Jan Řezáč,[†,‡] Jindřich Fanfrlík,[†] Dennis Salahub,*[,‡] and Pavel Hobza*[,†,§]

Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Systems, 166 10 Prague 6, Czech Republic, Department of Chemistry, Institute for Biocomplexity and Informatics (IBI) and Institute for Sustainable Energy, Environment and Economy (ISEEE), University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4, and Department of Physical Chemistry, Palacky University, Olomouc, 771 46 Olomouc, Czech Republic

Received February 23, 2009

Abstract: Because of its construction and parametrization for more than 80 elements, the semiempirical quantum chemical PM6 method is superior to other similar methods. Despite its advantages, however, the PM6 method fails for the description of noncovalent interactions, specifically the dispersion energy and H-bonding. Upon inclusion of correction terms for dispersion and H-bonding, the performance of the method was found to be dramatically improved. The former correction included two parameters in the damping function that were parametrized to reproduce the benchmark interaction energies [CCSD(T)/complete basis set (CBS) limit] of the dispersion-bonded complexes from the S22 data set. The latter correction was parametrized on an extended set of H-bonded stabilization energies determined at the MP2/cc-pVTZ level. The resulting PM6-DH method was tested on the S22 data set, for which chemical accuracy (error < 1 kcal/mol) was achieved, and also on the JSCH2005 set, for which significant improvement over the original PM6 method was also obtained. Implementation of analytical gradients allows very efficient geometry optimization, which, for all complexes, provides better agreement with the benchmark data. Excellent results were also achieved for small peptides, and here again, chemical accuracy was obtained (i.e., the error with respect to CCSD(T)/CBS results was smaller than 1 kcal/mol). The performance of the technique was finally demonstrated on extended complexes, namely, the porphine dimer and various graphene models with DNA bases and base pairs, where the PM6-DH stabilization energies agree very well with available benchmark data obtained with DFT-D, SCS-MP2, and MP2.5 methods. The PM6-DH calculations are very efficient and can be routinely applied for systems of up to 1000 atoms. For nonaromatic systems, the use of a linear scaling version of the SCF procedure based on localized orbitals speeds up the method significantly and allows one to investigate systems with several thousand atoms. The method can thus replace force fields, which face basic problems for the description of quantum effects, in many applications.

Introduction

Noncovalent interactions are of fundamental importance for chemistry and molecular biology disciplines. Theoretical description of these interactions is difficult, mainly because they are much weaker than covalent interactions and also because of the key role played by the London dispersion energy. The proper description of these interactions thus requires recovering a large portion of the correlation energy and the use of an extended atomic orbital (AO) basis set. It is now evident that the coupled-cluster method considering the single and double electron excitations iteratively and the triple excitations perturbatively [CCSD(T)] using an extended

* Corresponding authors e-mail: pavel.hobza@uochb.cas.cz (P.H.), dennis.salahub@ucalgary.ca (D.S.).

† Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Systems.

‡ University of Calgary.

§ Palacky University.

10.1021/ct9000922 CCC: $40.75
Published on Web 05/26/2009

basis set, or even performed at the complete basis set (CBS) limit,[1,2] provides accurate energies and other characteristics for different types of noncovalent complexes.[3] The CCSD(T)/CBS method is a genuine ab initio method (i.e., no empirical or experimental characteristics are utilized), but its use for larger systems is limited because of its N^7 scaling with the size of system (N is the number of AOs). All other nonempirical wave function (WF) and density functional theory (DF) quantum mechanical (QM) theories applicable in this field use one or more empirical characteristics mostly parametrized to the benchmark CCSD(T) data. These theories provide static characteristics of noncovalent interactions, and despite better scaling with the size of system, their use is limited to complexes having a maximum of several dozens to several hundreds of atoms. For the description of systems with several thousands of atoms as well as for the understanding of their dynamics, much faster computational procedures should be applied, and molecular mechanics (MM) methods (also called empirical potentials) play an indispensable role here. These methods are efficient enough and provide surprisingly reliable characteristics for various types of noncovalent complexes. The serious drawback of these methods is the fact that they cannot describe quantum effects. The most important among these effects are breaking and formation of a covalent bond, changes in electronic structure of various conformers of complex molecular systems, cooperativity effects, charge transfer, and chemical reactions.

Two different approaches can be utilized to solve these problems. The first represents the use of quantum mechanics coupled with molecular mechanics (QM/MM). This approach, however, is still not efficient enough for extensive sampling of the configuration space of complex molecular systems, and further, it is not free of problems related to the cutting of the chemical bond at the QM/MM boundary. The application of semiempirical QM methods represents another approach. Semiempirical QM methods properly and fully describe all quantum effects mentioned above. Because they were parametrized for covalent bonding, however, their use for noncovalent complexes is not straightforward. Also, in contrast to the nonempirical Hartree−Fock method, which does not recover the correlation energy (and, consequently, also does not recover the London dispersion energy, which forms the dominant part of the intersystem correlation energy), all QM semiempirical methods do account for the dynamic correlation via the scaling of the two-electron integrals, but they do not [without configuration interaction (CI)] include a nondynamic correlation. This situation is identical to the DFT method, where the empirical dispersion correction is applied with great success.

Hydrogen bonding is another important issue in semiempirical QM methods. The original MNDO[4,5] (modified neglect of differential overlap) method was not able to describe H-bonding at all. This serious problem was addressed in later MNDO-based methods (AM1,[6] PM3,[7,8] and others) through the introduction of additional core−core interaction terms and parametrization of the method to reproduce hydrogen bonding. Such a treatment is only empirical, however, and does not solve the problem completely. AM1 was able to describe the interaction, but it yielded incorrect geometries, often featuring bifurcated hydrogen bonds. Another step forward was introduced in the parametrization of the PM3 method, where the geometry of H-bonded complexes was emphasized. Recently, a special parametrization of the PM3 method, PM3(BP), was introduced[9] for application of the method to nucleic acid base pairs. Density functional tight-binding,[10] a semiempirical QM method with parameters adjusted according to density functional considerations, also has problems with hydrogen bonding. These are discussed and addressed in ref 11. An overview of the problem is provided in ref 12. A possible future development toward improving H-bonds within the MNDO framework is outlined in this work, but no further progress has been reported up to now. In general, currently used semiempirical QM methods systematically underestimate the strength of H-bonds by approximately 20−30%. Surprisingly, the problems of semiemprical QM methods with the lack of dispersion energy, which were believed to be more serious, were removed rather easily. Already in 2001,[13] Elstner and one of us (P.H.) modified the semiempirical tight-binding DFT technique by the simple addition of an empirical London dispersion energy, which dramatically improved the performance of the method toward the dispersion-bound noncovalent complexes (e.g., stacked DNA base pairs). Martin and Clark[14] introduced an additional term to treat the dispersion in NDDO-based semiempirical quantum chemical techniques (where NDDO is neglect of diatomic differential overlap). The dispersion energy was calculated using additive "atomic orbital" polarizability tensors. A similar procedure was used later for the modification of AM1, PM3, and OM-x semiempirical methods,[15,16] and in all of these cases as well, much better performance toward dispersion-bound complexes resulted. Unfortunately, for various reasons (parametrization for only a limited number of atoms, strongly overestimated stabilization energies for optimized geometries of H-bonded complexes), these methods are still not accurate enough for most applications in complex molecular systems.

Recently, the new semiempirical method PM6 (parameterized model 6) was introduced,[17] which is superior to other semiempirical QM methods in various aspects. It is an NDDO-based method improved by the adoption of Viotyuk's core−core diatomic interaction term[18] and Thiel's d-orbital approximation.[19−21] These modifications allowed parametrization of 80 elements and also reduced the error for main-group elements.[17,22] The good performance of this method and its applicability to a wide range of problems are the reasons why we selected the PM6 method for further improvement in the direction of noncovalent interactions.

PM6 is available in the MOPAC code[23] from version 2007 and also in the VAMP 10.0 program.[24] The latest version of MOPAC, MOPAC 2009, introduces another interesting feature that makes the PM6 method usable for very large systems: a linear scaling version of the self-consistent field (SCF) procedure using localized orbitals, named MOZYME.[23]

Despite all these advantages, the PM6 method still lacks the ability to accurately describe noncovalent interactions,

specifically the dispersion energy and hydrogen bonding. Even though the method yields surprisingly good geometries of all types of complexes, the interaction energy for dispersion-bound and H-bonded complexes is substantially underestimated. We believe that common NDDO parametrizations do not reproduce hydrogen bonds well. General suggestions as to how to improve semiempirical methods toward reliable description of H-bonds are provided in ref 32. Jug and Geudtner[25] also reported a variant of the SINDO (symmetrically orthogonalized intermediate neglect of differential overlap) method using a p-polarization function on hydrogen to improve the description of hydrogen bonds. However, there is no readily available and applicable method that would give acceptable results for noncovalent interactions. Let us recall once again that, among all widely used ab initio QM procedures (i.e., methods that do not use any empirical or experimental parameters), it is only the CCSD(T)/CBS technique that satisfactorily describes both of these interactions.

In the present study, we introduce an extension of the PM6 method in two directions: (i) including an empirical dispersion energy term that improves the description of complexes controlled by the dispersion energy and (ii) introducing an additional electrostatic term that improves the description of hydrogen-bonded complexes. The resulting method, PM6 with corrections for dispersion and hydrogen bonding, is named PM6-DH. The aim is ambitious: for extended noncovalent complexes to achieve standard ab initio chemical accuracy ($\sim$ 1 kcal/mol). Because of favorable scaling of the code, we would also like to use it in MD simulations. Therefore, in addition to single-point calculations, we also carefully tested the performance of the method when the full gradient optimization was adopted. It should be remembered that just this point was critical for the application of other semiempirical QM methods.

This article also presents benchmarks of the method and a comparison to other methods with a comparable range of applications. The first tests concerned the stacking interactions. Specifically, we investigated the interaction of two porphine molecules, as well as the interaction of various graphene models with nucleic acid bases and base pairs. In the next step, we compared a more complex system that contained both characteristic interaction types, stacking as well as H-bonding. We studied the interaction of DNA with 4′,6-diamidino-2-phenylindole (DAPI), a fluorescent dye that can bind both in the minor groove of the double helix and as an intercalator. Interaction energies obtained with PM6-DH were compared with benchmark calculations in these examples. Finally, as a last example, a DNA tetramer was optimized using the PM6 and PM6-DH methods. The structure of the DNA fragment was determined by both stacking and H-bonding, and this test example was selected to demonstrate the performance of PM6-DH toward the important world of nucleic acids. Another reason supporting this point is the recent finding that the description of the DNA double-helical structure requires inclusion of the dispersion energy.[26]

Methods

Dispersion Correction. The first step in improving the PM6 method was the addition of an empirical dispersion term. This was not a difficult task, because the London dispersion term is well separated from the QM calculation and thus transferable between various methods. Using our experience with empirical dispersion in the Hartree−Fock and density functional theory (DFT) methods,[27−30] including semiempirical tight-binding DFT,[13] we have adopted the formalism described in the work of Jurecka et al.[30]

The correction has the form of a pairwise interatomic force field

$$\sum E_{\mathrm{dis}} = -\sum f_{\mathrm{damp}}(r_{ij}, R_{ij}^{0}) C_{6ij} r_{ij}^{-6} \qquad (1)$$

It consists of the physically sound r^{-6} term, damped at short distances to avoid interfering with the underlying QM potential, that describes the short-range repulsion correctly. Atomic parameters, namely, atomic van der Waals radii (R_{ij}^{0}) and C_6 coefficients, are independent of the QM method and were adopted from the original work. Two parameters in the damping function, the scaling factor for the radii s_r and the exponent α that affects the slope of the damping, were optimized to reproduce interaction energies of dispersion-bonded complexes. A subset of the S22 benchmark data set[31] with hydrogen-bonded complexes removed was used in the fit. Omitting hydrogen-bonded complexes from the training set should yield a better description of the dispersion-bonded complexes; the errors in hydrogen bonds are corrected separately.

The PM6 method with the dispersion correction only is abbreviated as PM6-D in this article.

Hydrogen-Bond Correction. Improving hydrogen-bonding interactions is not as straightforward as correcting dispersion interactions because the electrostatic term, mainly responsible for the description of H-bonding, was included in the original parametrization. Previous attempts to improve the hydrogen-bond description were done at the level of modifying the semiempirical method itself or reparameterizing it.[11,12,32] Because these attempts have not solved the problem, we decided to take another approach. Inspired by the success of the dispersion correction, we aimed to introduce a specific correction that would affect only hydrogen bonds, added on top of an unmodified semiempirical calculation.

Training Set of 104 Hydrogen-Bonded Complexes. The first step in this work was the preparation of an extensive training set of model hydrogen-bonded complexes. We designed the set to cover different types of hydrogen bonds present in biomolecules and organic compounds. Therefore, the training set was composed of 104 hydrogen-bonded complexes formed from the proton donors acetylamine, acetic acid, dimethylamine, phenol, methanol, methylamine, phenylamine, peptide bond, pyrrol, uracil, water molecule, and 1-imino-3-aminocyclohexane, along with the proton acceptors acetic acid, dimethyl amine, phenylamine, phenylm-

ethylamine, furan, methanol, methylamine, propanol, peptide bond, pyrazine, uracil, water molecule, and 1-imino-3-aminocyclohexane. The reference geometries were obtained by RI-MP2/cc-pVTZ gradient optimization, which is known[33] to provide reliable results for various types of noncovalent complexes, including the presently investigated H-bonded ones.

The dissociation curve (scan of the interaction energy as a function of the hydrogen-bond distance) of each of these complexes was calculated using the accurate SCS(MI)-MP2/cc-pVTZ method,[34,35] with the counterpoise correction to eliminate basis set superposition error. This method is parametrized to yield highly accurate interaction energies, while being efficient enough for this task. It should be recalled that the SCS(MI)-MP2 method provides reliable interaction energies not only for stacked complexes (similarly to the original SCS-MP2 technique) but also for H-bonded ones where the SCS-MP2 technique failed. With 14 points per dissociation curve, about 1500 interaction energies had to be calculated to prepare the training set. The calculations were performed using Turbomole 5.9.[36]

The same curves were then calculated using PM6 with the dispersion correction. With this information, we then analyzed the distance-dependent behavior of the error to design the form of the correction.

Selection of Possible H-Bonds. Prior to the calculations, all possible hydrogen bonds were determined from the topology. All possible combinations of hydrogens bonded to an electronegative atom (elements N and O) and electronegative hydrogen acceptors (N, O) were listed. Then, all pairs in the 1−4 configuration (i.e., in the H−X−Y−acceptor pattern) were removed from the list, because they cannot form hydrogen bonds.

The correction was calculated for all of these possible pairs, with the form of the correction ensuring that only the actual hydrogen bonds contributed significantly to the total energy.

Atom Types. An important finding was that there is a significant difference between hydrogen groups with nitrogen as the acceptor and those with oxygen as the acceptor. Further sorting of the error curves showed that there are differences in hydrogen bonds between the elements and that these differences can be correlated with the valence state and environment of the atoms. Later, optimization of different models confirmed that the introduction of atom types is necessary to describe all types of hydrogen bonds accurately.

Rather than using atom types directly, we selected only their combinations that show different behavior and introduced "hydrogen-bond types" to reduce the number of parameters in the model. This step, of course, makes the correction rather empirical and is a source of several limitations, but it is, in our opinion, worth the improved accuracy.

The hydrogen-bond types are as follows:

(1) nitrogen with no hydrogens bonded to it (mostly in aromatic rings) interacting with any hydrogen,

(2) nitrogen with one hydrogen (secondary amines) interacting with any hydrogen,

(3) nitrogen with two or more hydrogens (primary amines, ammonia) interacting with any hydrogen,

(4) oxygen except carbonyl interacting with HN,

(5) carbonyl oxygen interacting with HN,

(6) oxygen interacting with HO hydrogen different from 7 and 8,

(7) oxygen interacting with H in a water molecule, and

(8) oxygen interacting with H in a carboxyl group.

The first limitation that arises from the introduction of hydrogen-bond types is that the method can be applied only to hydrogen bonds for which it was explicitly parametrized. In this work, we attempted to cover all hydrogen bonds in organic compounds involving nitrogen and oxygen. The extension of this set would be straightforward, as there is no reason to believe that the same correction cannot be applied to other elements. We plan to further improve the method, and extending the set of parameters is one of the goals toward this end. Specifically, the addition of parameters for sulfur is necessary for full coverage of the interactions in proteins.

The second limitation directly connected to the hydrogen-bond types in the current implementation is that the bond types are determined only once (at the beginning of the calculation), even for calculations where the geometry can change (i.e., in optimization). This prevents the method from being used to study processes where the valence states of the atoms involved in hydrogen bonds change, because the valence state is what determines the bond type used. The most obvious example of such a reaction is a proton transfer. A simple re-evaluation of the bond types in each step of the calculation would not correct this problem, because the potential energy surface would become discontinuous. On the other hand, this limitation merely prevents the method from being used to study the intermediates of such a reaction. When the reactants and products are studied separately, the different bond types can be readily assigned.

Form of the Correction. Most of the error curves (constructed as energy differences between the benchmark and PM6-D) are similar to the function $-1/r$, which is in agreement with the Coulomb nature of hydrogen bonds. It is obvious that the form of the correction should be similar to the formula for an electrostatic interaction. In some cases, especially in hydrogen bonds of the XH···O type, there is a repulsion at shorter distances, and the correction sometimes becomes positive. An exponential term was added to describe this effect.

Hydrogen bonds differ in their strengths, and so does the correction. Models that do not take this fact into account are not very accurate. To cover these differences, we used partial atomic charges from the PM6 calculations (obtained using Mulliken population analysis) on the hydrogen and on the acceptor. The product of these two charges correlates reasonably well. We tested many different formulas, but the one based on the interaction energy of a point charge and a dipole worked best. Unlike charge−charge interactions, which have an r^{-1} dependence, the distance dependence in the present formula is r^{-2}, and there is an angular part calculated from the angle of the three atoms involved in the hydrogen bond. This angular part is crucial, because the

core−core interaction introduced in semiempirical methods (including PM6) to improve hydrogen bonding is not directional, whereas in reality, H-bonds are very sensitive to the spatial arrangement. Combined together, the final formula for the correction energy (E_{HB}) term in hydrogen bond XH···Y is

$$E_{HB} = c[q(H)\, q(Y)/r^2 \cos(\theta) + c_{rep}A^{-r}] \qquad (2)$$

where r is the H···Y distance and θ denotes the angle XHY. c, c_{rep}, and A are parameters fitted to obtain the best results over the training set.

It must be emphasized that many different models of the correction term were tested, but the one presented here was found to yield the best results.

To prevent problems arising from calculation of pairs in the Y···XH geometry (for example, within a single molecule with more electronegative atoms), an additional rule was added, and pairs with $\theta < 90°$ were not calculated, because they certainly are not hydrogen-bonded.

Optimization Procedure. Different types of hydrogen bonds cannot be parametrized separately, because the contribution of each possible hydrogen bond is calculated in the complex. To gain more control over the fitting procedure, we used gradient optimization of an arbitrary error function, based on numerical gradients of the fitted coefficients. The Broyden−Fletcher−Goldfarb−Shanno (BFGS) algorithm was used to make the optimization efficient. This setup, in contrast to the least-squares method normally used in similar applications, allowed us to experiment with different goals of the optimization. We used two measures of the quality of the fit: the mean absolute error, e_{avg}, and the maximum absolute error, e_{max}. The error function that was minimized by the optimization procedure was a combination of these error measures in variable ratio. We did multiple optimizations with different ratios of e_{avg} and e_{max} to find the best compromise in usability of the fitted method.

Implementation. The basis for our work was the PM6 method implemented in the MOPAC2007 package.[23] The corrections were calculated on top of the finished PM6 calculation, which makes them independent of the PM6 code itself. For the purpose of development of the method, we developed our own code that calls MOPAC for the PM6 calculation and adds the corrections to the result. The latest version of this software has been made available to the public and can be obtained from the authors upon request (http://www.molecular.cz/~rezac/pm6dh.html).

To make the method more useful, we calculated the gradients of both corrections. For the dispersion correction, this was a straightforward task, because the correction was completely separated from the underlying PM6 calculation. The situation was more complicated in the case of the H-bonding correction, which used atomic charges from the PM6 calculation. Determination of the true analytical gradient would require calculation of the derivatives of these charges with respect to nuclear coordinates, which would make the calculation more expensive. In addition, such a calculation is not possible in the current implementation. We made use of the fact that the charges varied only slightly and considered them to be constant to calculate an estimate of the gradient. To justify this approach, we compared this estimate to a full numerical evaluation of the gradients. We found that the differences in the gradient itself and, more importantly, in the optimized energy and geometry were very small and below the convergence limits used in the optimization.

The BFGS optimizer working in Cartesian coordinates is part of the final code.

Validation Sets. The PM6-DH method was tested on multiple sets of benchmark data available from the BEGDB database.[37] All of these data sets feature high-quality geometries and energies extrapolated to the CCSD(T)/CBS level.

First, the results were compared to the S22 data set,[31] which includes model complexes covering hydrogen-bonding and dispersion interactions. In this case, both energies and optimized geometries were utilized. The fact that geometries were tested along with the energies is important because all previous semiempirical QM procedures were tested only for stabilization energies and, on the basis of our experience, their application to geometry optimization is not straightforward. Second, the JSCH2005 database[31] of DNA base pairs in various geometries and some amino acid pairs was used as an example of biologically relevant complexes. Finally, we tested the PM6-DH method in an application different from the calculation of interaction energies (i.e., molecular clusters). The last test was therefore the calculation of the relative stability of conformers of small peptides.[38] In the original work featuring this data set, it was shown that this is a very sensitive measure of the quality of computational methods, and a comparison with benchmark energies is available.

Results

Parameterization of Dispersion Corrections. There were only two parameters to be adjusted in the damping function in the empirical dispersion term. Calculation of the correction was very fast, so we performed a complete 2D scan instead of optimization to obtain the values corresponding to minimal error, expressed as the average absolute value of the difference between PM6-D and the benchmark energy over the training set. This procedure led to $s_r = 1.07$ and $\alpha = 11$ with $e_{avg} = 0.4$ kcal/mol for stacked structures within the S22 benchmark data set.[31]

Parameterization of H-Bond Correction. Optimization of the parameters in the H-bond correction was done in two steps to increase the efficiency of the process. In the first round, only optimal geometries of the complexes in the training set were used in the optimization. The resulting set of coefficients was then used as a starting point for a second round, in which all geometries along the dissociation curve were taken into account.

We tested error functions (the optimized value) with various ratios of the mean and maximum errors, observing the impact on the results. The best error function, which was used to derive the presented parameters, consisted of 90% of the mean error and 10% of the maximum error. Increasing the fraction of the mean error brought no significant

Table 1. Fitted Coefficients in the H-Bond Correction for Hydrogen-Bond Types Discussed in the Text[a]

bond type no.	c	c_{rep}	A
1	14.4209	-1.3273×10^{-2}	7.2847
2	73.3566	-5.3979×10^{-4}	7.0920
3	48.7161	2.9844×10^{-4}	6.4259
4	29.8036	2.1262×10^{-3}	6.9768
5	−6.4578	7.3142×10^{-3}	7.8379
6	23.1582	-4.8015×10^{-5}	6.9382
7	15.3029	2.0789×10^{-3}	7.0365
8	14.8668	-4.6652×10^{-3}	6.9111

[a] Coefficients derived using the following units in eq 2: elementary charge, Å, and kcal/mol.

improvement in e_{avg}, but resulted in an increase in e_{max}. For the opposite change, e_{max} was improved only marginally, whereas e_{avg} became too large. A mean error of 0.8 kcal/mol was achieved for the training set.

Parameters obtained from the optimization are listed in Table 1. Note that the coefficient scaling the repulsion term is negative in some cases. The exponential term is thus no longer repulsive, but it further enhances the attraction, improving the potential of the electrostatic correction where needed. We tried to constrain these coefficients to be nonnegative, but the final results were worse. This issue reflects the major difference between oxygen and nitrogen hydrogen bonds in the PM6 method. (Data demonstrating their different natures are provided in the Supporting Information.) We understand that changing the sign of a term that was originally designed to be repulsive makes our correction further decline from a physically sound formulation, but our goal was to develop an empirical correction, focusing mainly on the final accuracy.

The validation sets yielded a slight increase in the error when the H-bond correction was added to dispersion-bonded complexes that contained heteroatoms that could be involved in hydrogen bonds. We tried to address this problem by a more localized variant of the correction, with a cutoff that would eliminate all contributions that were not real hydrogen bonds. However, the overall results were worse.

Geometry Optimization. After the correction energy term was established, it was tested in geometry optimization to verify its performance. The method was found to yield good geometries, as discussed below, with one exception. In the H-bonded adenine−thymine base pair, we observed a proton transfer along one of the hydrogen bonds, as the gradient of the H-bond correction overpowered the reaction barrier. This happened because the possible hydrogen bonds were determined from the geometry of the initial state, so there was a force driving the hydrogen in one direction but not in the other. The gradient was very steep in this region; here, it became an incorrect extrapolation from the potential at larger distances.

To solve this problem, we had to modify our potential. It had the original form as long as the YH distance (in the Y···HX arrangement) was above a certain limit and remained constant at shorter distances. We set the limit at 1.8 Å, the YH distance in the shortest H-bonds in our training set. This can be easily achieved by using the following rule to determine the value of r in eq 2

$$r = \begin{cases} r & \text{for } r > 1.8 \\ 1.8 & \text{otherwise} \end{cases} \quad (3)$$

The resulting potential has a constant energy and a zero gradient at very short distances, which prevents the anomalous behavior observed before. We tested this modification extensively and found it to be safe. It does not affect calculations in a fixed geometry; it only eliminates the rare problem of geometry optimization described above. This modification was used for all calculations presented here and was implemented in the PM6-DH code.

Tests on Benchmark Data. *S22 Data Set.* The S22 data set is our first choice for evaluating the accuracy of a method in describing noncovalent interactions. It contains complexes featuring hydrogen bonds, dispersion interactions, and their combinations. Biomolecules are represented here by base pairs in both stacked and Watson−Crick (WC) arrangements.

Our first test was an analysis of interaction energies calculated with the PM6, PM6-D, and PM6-DH techniques on the S22 data set. The results are summarized in Table 2, and their differences from the benchmark CCSD(T)/CBS data are plotted in Figure 1. For comparison, we include interaction energies calculated at the much more expensive MP2/cc-pVTZ level (corrected for basis set superposition error).

The performance of the unmodified PM6 is rather poor. The mean absolute error (e_{avg}) is 3.4 kcal/mol, and the maximum absolute error (e_{max}) is 7.5 kcal/mol, occurring in the formic acid dimer. PM6-D brings a great improvement in complexes dominated by dispersion, which reduces e_{avg} to 1.4 kcal/mol, but the maximum error remains high, 6.5 kcal/mol. Finally, by combining the dispersion correction with the correction for hydrogen bonds in PM6-DH, both sources of error are addressed, and the results improve significantly. The mean absolute error is 0.6 kcal/mol, and e_{max} is reduced to 1.8 kcal/mol. These results are even better than in the MP2/cc-pVTZ calculation, for which the mean error is 0.7 kcal/mol and the maximum error is 1.9 kcal/mol.

The errors worsen slightly for the calculation of interaction energies on structures optimized with PM6-DH itself, with $e_{avg} = 0.8$ kcal/mol and $e_{max} = 2.8$ kcal/mol. A possible source of this error is discussed later in this article. However, this behavior is specific only to the strongest hydrogen bonds, and the optimization generally improves the results. Importantly, geometry optimizations performed with other semiempirical QM techniques led to considerably worse results.

It is clear from the results presented that the correction for hydrogen bonding is less perfect than that for dispersion. The complexes with the largest errors are hydrogen-bonded, and the errors can be both positive and negative. The hydrogen-bond correction also introduces some small error into some dispersion-bonded complexes. These are the limitations of the simple form of our correction and its empirical nature. Nevertheless, these errors are very small compared to those obtained with the uncorrected PM6 method, and the overall accuracy is on a par with that of expensive correlated ab initio calculations.

Finally, we tested PM6 and its modifications for the optimization of the S22 complexes. In general, the results

Table 2. Interaction Energy Errors (in kcal/mol) for the S22 Set of Complexes, Calculated as Differences between the Studied Method and the Benchmark CCSD(T)/CBS Results, in kcal/mol[a]

	MP2/cc-pVTZ	PM6	PM6-D	PM6-DH	PM6-DH optimized[b]
ammonia dimer	0.40	0.86	0.33	−0.57	−0.75
water dimer	0.59	1.08	0.70	0.35	0.29
formic acid dimer	1.72	7.47	6.47	1.22	−0.50
formamide dimer	1.71	3.41	2.19	0.57	0.95
uracil dimer (H-bonded)	1.91	7.33	5.55	1.81	1.10
2-pyridoxine···2-aminopyridine	0.80	6.73	4.51	−0.64	−1.79
adenine···thymine (WC)	1.45	7.31	4.90	−1.46	−2.75
methane dimer	0.21	0.47	−0.20	−0.20	−0.20
ethene dimer	0.36	1.11	−0.01	−0.01	−0.02
benzene···methane	0.09	1.03	−0.25	−0.25	−0.38
benzene dimer (stacked)	−1.03	2.86	−0.89	−0.89	−0.86
pyrazine dimer	−1.02	2.61	−0.99	−0.99	−1.32
uracil dimer (stacked)	1.03	5.66	0.53	0.42	0.09
indole···benzene (stacked)	−1.22	5.29	0.02	0.02	−0.77
adenine···thymine (stacked)	−0.07	7.29	−0.04	−0.55	−1.38
ethene···ethine	0.10	0.98	0.42	0.42	0.36
benzene···water	0.35	1.00	−0.13	−0.13	−0.67
benzene···ammonia	0.22	0.82	−0.42	−0.42	−1.47
benzene···HCN	−0.14	2.48	1.26	1.26	1.25
benzene dimer (T-shaped)	−0.27	1.99	−0.10	−0.10	−0.11
indole···benzene (T-shaped)	−0.43	3.33	0.43	0.43	0.51
phenol dimer	0.34	3.67	1.33	0.32	−0.41
e_{avg}[c]	0.7	3.4	1.4	0.6	0.8
e_{max}[c]	1.9	7.5	6.5	1.8	2.8

[a] High-quality geometries from the S22 database used unless otherwise noted. [b] Geometries optimized using the PM6-DH method. [c] Mean and maximum absolute errors used as a measure of the quality of the method.

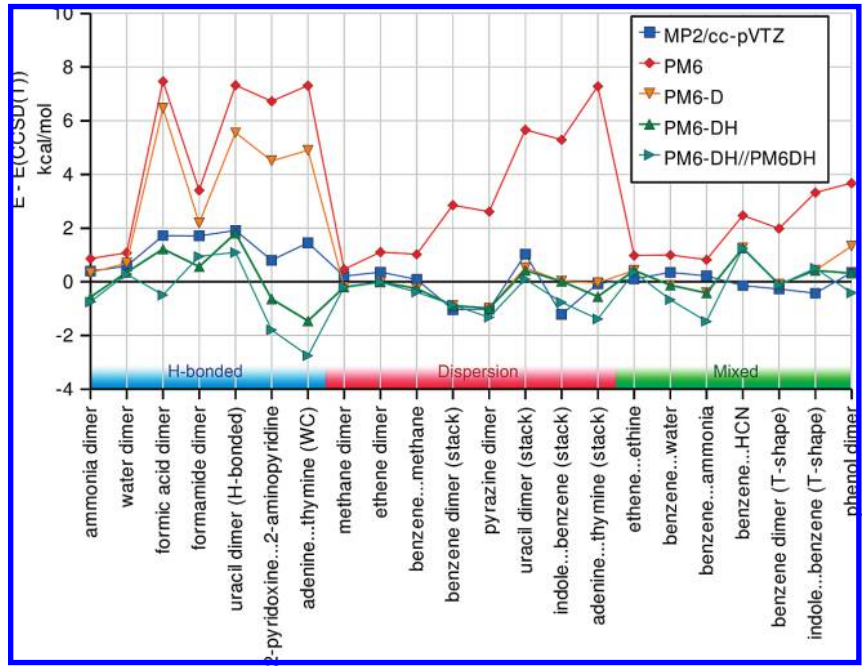

Figure 1. Interaction energy errors in the S22 set of complexes, plotted as the difference between the studied method and benchmark CCSD(T)/CBS results.

were very good even for PM6, and they did not improve when the corrections for dispersion and H-bonding were introduced. The root-mean-square deviation (rmsd) upon optimization, starting from the benchmark geometry, was 0.14 Å for PM6, 0.13 Å for PM6-D, and 0.15 Å for PM6-DH. The hydrogen-bond correction actually made the geometries slightly worse. This effect was more pronounced in complexes featuring very strong H-bonds, such as the formic acid dimer. The problem can be attributed to the form of the correction and the description of hydrogen bonds in the semiempirical method itself. In both cases, the potential responsible for hydrogen bonding acts between the nuclei, but in reality, the hydrogen bond forms between the lone pair on the proton acceptor and an antibonding orbital of the X—H bond of the proton donor. An important feature of H-bonding, namely, its directionality, is due to this point.

Table 3. Analysis of Uncorrected and Corrected PM6 Results for the JSCH2005 Database Featuring Biomolecular Complexes[a]

complex	PM6	PM6-D	PM6-DH
	e_{avg} (kcal/mol)		
all neutral	4.5	2.2	1.6
DNA bases	4.7	2.3	1.6
DNA bases, H-bonds	7.7	5.3	1.9
DNA bases, stacked	3.6	1.2	1.5
amino acids, neutral	3.1	0.9	1.0
amino acids, charged	20.5	17.7	17.5
	e_{avg} (%)		
all neutral	52	25	18
DNA bases	51	25	17
DNA bases, H-bonds	42	29	10
DNA bases, stacked	61	20	26
amino acids, neutral	67	20	22
amino acids, charged	24	21	20
	e_{max} (kcal/mol)		
all neutral	12.1	7.9	6.1
DNA bases	12.1	7.9	6.1
DNA bases, H-bonds	10.7	7.9	4.3
DNA bases, stacked	12.1	6.5	6.1
amino acids, neutral	6.4	3.2	3.3
amino acids, charged	30.1	26.0	25.7

[a] Errors compare PM6-based calculations to benchmark CCSD(T)/CBS results.

Our simple model does not cover this difference while making the interaction stronger, which leads to slightly distorted geometries. This fact should be considered when PM6-DH is used, but the benefits of more accurate interaction energies are much more important. Proper energetics for noncovalent interactions would also improve geometries in more complex structures where intramolecular and/or intermolecular interactions and their balance with other forces are crucial.

JSCH2005. A further step in the evaluation of PM6-DH was to extend the validation set. We used the JSCH2005 data set[31] for this purpose. It consists of DNA base pairs in various geometries and some complexes of amino acids in arrangements found in proteins. In this step, only the interaction and/or relative energies were considered (i.e., no geometry optimization was performed). These complexes were divided into several groups for the analysis. First, we examined separately all complexes of neutral molecules and all charged complexes, because the magnitudes of the interactions are different. In addition, we evaluated the error measures in the groups of hydrogen-bonded base pairs, stacked base pairs, and amino acid complexes. The results, again in terms of average and maximum absolute difference from CCSD(T)/CBS value, are presented in Table 3.

Several important conclusions can be drawn from these results. The first is the progressive improvement of the results when the corrections are added to PM6 calculations, with the PM6-DH method reaching an average error of only 1.6 kcal/mol for the neutral complexes. The error for the charged complexes, the amino acid ion pairs, is much larger, but this is because the interaction energy itself is very large. When the errors are made relative to the magnitude of the interaction, the percentage errors become very similar.

Table 4. Mean and Maximum Absolute Errors (in kcal/mol) of PM6-Based Methods Compared to CCSD(T)/CBS Conformer Energies in a Set of 76 Peptides

	method				
	MP6	PM6-D	PM6-DH	PM6-DH optimized	MP2/cc-pVTZ
e_{avg} (kcal/mol)	1.66	1.76	1.04	0.89	0.92
e_{max} (kcal/mol)	4.69	4.36	5.46	3.31	2.64

Second, these data show that the corrections improve interaction energies in complexes of charged molecules, even though they were parametrized only for neutral ones.

Finally, this large validation set shows that the descriptions of different types of interactions are similar; the method can thus handle both hydrogen bonds and dispersion in a balanced way.

It should be noted that these complexes are the most difficult ones to describe with our method. The hydrogen-bonded base pairs feature strong, cyclic hydrogen bonds for which additional cooperativity of the bonds is important, whereas the correction was parametrized on model complexes featuring single hydrogen bonds. The same applies for stacked bases here, because the molecules feature multiple sites that could be involved in hydrogen bonds. The dispersion energy in nonpolar molecules is easier to describe, and the error can be much smaller, as demonstrated for the S22 set.

Peptides. In our previous work,[38] we thoroughly investigated the performance of a wide range of computational methods for the calculation of conformational energies of small peptides. This turned out to be a very sensitive test case, and not many methods were found to yield satisfactory accuracy. Standard force field methods had problems[38] with atomic charges because these charges depend on the peptide conformation. The stability of the conformers is often determined by intramolecular noncovalent interactions, mainly hydrogen bonds, when present, and the dispersion energy, in peptides containing aromatic amino acid residues. The high sensitivity of the description of noncovalent interactions and the availability of benchmark energies and geometries make this data set very important for the evaluation of the PM6-DH method. The application of semiempirical QM methods

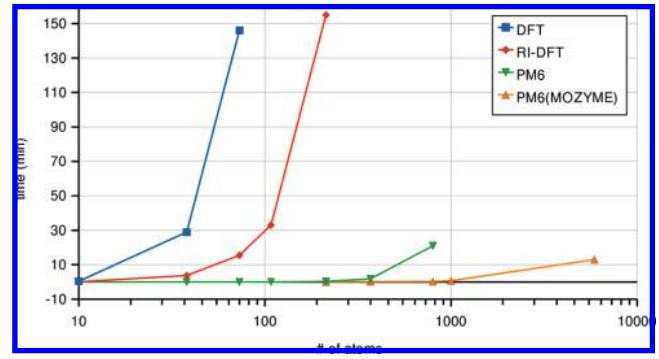

Figure 2. Comparison of DFT, RI-DFT, PM6, and PM6-MOZYME computational resources required for one SCF cycle on a single protein molecule. Note that the horizontal axis is logarithmic to accommodate the studied range of molecule sizes.

Table 5. Stabilization Energies (in kcal/mol) of Stacked (S), Parallel-Stacked (PS), and T-Shaped (T) Structures of the Porphine Dimer Determined by the PM6-DH, DFT-D/TPSS/TZVP, DFT-D/B97/TZV(2df,2pd), SCS-MP2/ TZV(2df,2pd), and MP2.5/CBS[a] Techniques

structure	DFT-D/TPSS/TZVP	DFT-D/B97/TZV	SCS-MP2	PM6	PM6-DH	MP2.5
A(S)	10.6	13.5	17.7	−0.8	16.9	
B(S)	11.2	14.5	19.0	0.4	18.0	17.6
C(S)	16.0	18.3	24.6	0.2	21.1	
D(PS)	16.6	20.0	25.6	0.2	22.0	
E(PS)	16.9	20.5	25.8	1.5	23.3	23.0
F(PS)	17.1	20.6	26.0	0.9	23.2	
G(T)	7.2	8.4	7.3	3.2	8.9	
H(T)	7.3	9.7	9.1	2.3	9.4	

[a] $E^{(2)}$ correction from CBS and scaled $E^{(3)}$ correction from TZV.

in this field is important because empirical potentials, which are usually applied, have failed to describe the variability of atomic charges for different conformers.

On the set of 76 structures of FGG, GFA, GGF, WG and WGG peptides, the relative energies of the conformers (taking the average energy for the same peptide as the zero level) were calculated using PM6, PM6-D, and PM6-DH with the original geometries and also using geometries optimized with PM6-DH. Mean and maximum absolute errors [compared to CCSD(T) relative energies] are listed in Table 4. MP2/cc-pVTZ results from the original work are included for reference.

Achieving an average error of 1.04 kcal/mol with PM6-DH is a very encouraging result. Even more importantly, the errors were reduced by geometry optimization. In the optimized structures, we achieved the limit of chemical accuracy, 1 kcal/mol. Note that these results are comparable to those of the considerably more expensive MP2/cc-pVTZ calculations.

Timing. A comparison of the timing of DFT, the resolution of the identity (RI) approximation to DFT, PM6, and PM6-MOZYME was done on model peptides and proteins. In-silico-built poly(glycine) helices covered the smaller testing systems; experimental protein geometries (PDB numbers 2BEG and 1AFW) served as the largest testing systems. All calculations were run on the same computer featuring a single 2.4 GHz Intel processor. The results are shown in Figure 2. The Meta-GGA functional TPSS and the TZVP basis set were used for both DFT and RI-DFT calculations. Obviously, the current limit of the DFT method is placed at around 100 atoms. The RI approximation[39] to the DFT method allows for the treatment of several hundreds atoms. The NDDO approximation within the PM6 method reduces the computational cost and places the limit of semiempirical methods at around 1000 atoms. The use of the localized molecular orbital method (MOZYME) speeds up the PM6 method significantly and allows for the calculation of systems having several thousand atoms. Following expectations, the localized molecular orbital method was found to be effective only in the case of systems with localized electrons (e.g., amino acids, peptides, proteins). For aromatic systems (such as DNA bases), the acceleration of the calculation resulting from the use of MOZYME option was less dramatic. Notice that the largest models (real proteins) contained aromatic groups, such as phenylalanine. The number of such residues, however, is considerably smaller than the number of nonaromatic ones. In the case of DNA, the ratio of aromatic residues is much higher.

Table 6. Stabilization Energies (in kcal/mol) of Various Graphene Models with Nucleic Acid Bases and Base Pairs Determined by the PM6-DH, DFT-D/B97/TZV(2d, 2p), DFT-D/TPSS/TZVP, and SCS-MP2/ TZV(2df,2pd) Techniques

C		A	T	U	C	G	A···T	A···U	G···C
24	PM6-DH	14.8	15.4	13.6	13.7	18.2	19.7	18.9	20.4
	PM6	2.3	3.9	3.9	3.0	5.3	1.9	1.9	2.4
	DFT-D/TPSS/TZVP	12.1	12.7	11.4	11.6	16.2	15.3	14.7	15.8
	DFT-D/B97/TZV	14.2	15.4	13.2	13.7	18.3	17.7	17.0	18.3
	SCS-MP2	15.1	14.9	13.0	13.4	18.0			
54	PM6-DH	19.0	17.8	15.5	16.9	21.5	29.3	28.2	31.2
	PM6	2.9	3.2	3.2	3.4	4.2	2.9	3.0	3.9
	DFT-D/TPSS/TZVP	15.9	14.8	13.4	15.4	19.8	24.4	23.6	27.4
	DFT-D/B97/TZV	19.3	18.0	15.4	17.8	23.3	29.9	28.5	32.0
	SCS-MP2	20.8	18.4	16.1	18.2	24.1			
96	PM6-DH	19.7	18.0	15.6	17.5	22.4	35.2	33.1	36.5
	PM6	2.9	2.6	2.5	3.4	4.4	4.4	4.5	5.5
	DFT-D/TPSS/TZVP	16.6	15.5	14.0	16.1	20.7	29.8	28.1	32.1
	DFT-D/B97/TZV	20.2	19.0	16.3	18.5	24.2	36.5	34.0	37.9
150	PM6-DH	19.8	18.1	15.8	17.7	22.9	36.7	34.4	37.7
	PM6	2.8	2.4	2.4	3.3	4.2	4.9	4.9	6.0
	DFT-D/TPSS/TZVP	16.7	15.8	14.3	16.4	21.1	30.5	28.8	33.0
	DFT-D/B97/TZV	20.3	19.3	16.5	18.8	24.9	–	34.9	38.7

Application Examples. *Porphine Dimers.* Stabilization energies of various structures of the porphine (the simpliest porphyrine macrocycle) dimer were determined using the PM6-DH technique and these energies were compared to reference data from the work of Muck-Lichtenfeld and Grimme,[40] calculated using DFT-D/B97 and SCS-MP2 [with application of the TZV(2df,2pd) basis set in both cases], as well as our DFT-D TPSS/TZVP and MP2.5/CBS [$E^{(2)}$ correction from CBS and scaled $E^{(3)}$ correction from TZV] calculations.[41] Table 5 shows that the PM6-DH stabilization energies agree very well with the most reliable MP2.5 results, with an average error for two typical structural types of less than 3%. The DFT-D values are systematically (with the exception of structure H) underestimated (by about 13% for the B97 functional, more with TPSS). The SCS-MP2 values are overestimated (by about 11%) for all stacked structures whereas they are underestimated for both T-shaped structures.

Graphene···Nucleic Acid Bases and Base Pairs. PM6-DH stabilization energies of various sheet models of

graphene (having 24, 54, 96, and 150 carbon atoms) with nucleic acid bases as well with H-bonded base pairs were compared with the reference data obtained by the DFT-D/B97/TZV(2d,2p) and SCS-MP2/TZV(2df,2pd) techniques.[42] Our own calculation at the DFT-D/TPSS-TZVP level is included for consistency with the other results presented here. Table 6 shows that PM6-DH results for complexes of graphene with isolated bases agree very well with reference energies for both relative and absolute stabilization energies. A similar conclusion can be made about complexes of graphene with DNA base pairs.

DNA···DAPI. Stabilization energies for the DNA dimer with 4′,6-diamidino-2-phenylindole (DAPI) bound as an intercalator and for the DNA trimer with DAPI bound to the minor groove determined with the PM6-DH method are 58.6 and 88.5 kcal/mol, respectively. Reference stabilization energies obtained using DFT-D/TPSS/TZVP are smaller than the PM6-DH results (40.4 and 67.8 kcal/mol, respectively). Considering, however, the results of the previous applications (i.e., porphine dimers and complexes of graphene with nucleic acid bases), the DFT-D results represent a lower limit of the real (unknown) stabilization energy. The present PM6-DH results can be thus considered as a satisfactory estimate of the respective stabilization energies.

Optimization of the DNA Tetramer. The DNA tetramer was optimized with the PM6 and PM6-DH methods, and the resulting geometries were compared to the DFT-D/TPSS/SVP ones. To demonstrate the flexibility of the method, optimizations were performed in the gas phase as well as in an environment represented by the COSMO implicit solvent model. The structure of the DNA tetramer collapsed during optimization using the PM6 method, and the rmsd for the resulting structure was 2.4 Å (compared to the DFT-D/TPSS/SVP geometry). The D and H corrections to the PM6 method improved its performance. The DNA tetramer retained its characteristic features during optimization with the PM6-DH method, and the rmsd decreased to 1.6 Å. Consideration of an implicit water model in PM6-DH (and also in the reference DFT-D calculations) further decreased the rmsd to an acceptable value of 1.3 Å.

Conclusions

Herein, we present a novel approach to improve semiempirical methods toward a more accurate description of noncovalent interactions acting in molecular clusters as well as in complex molecular systems. London dispersion and hydrogen bonds are treated separately by empirical corrections added to the QM calculation.

(1) Our method (PM6-DH) is based on the recent semiempirical method PM6, which currently represents the state of the art in the development of semiempirical QM methods. The same approach can be used with other QM and semiempirical methods.

(2) We adopted the formalism of a dispersion correction used previously in DFT and reparameterized it for PM6. It systematically improves all dispersion-bonded complexes, yielding an accuracy close to that of high-level correlated QM methods.

(3) The correction of hydrogen bonds has the form of an additional electrostatic term applied to all possible hydrogen bonds in the system. Differences in the nature and magnitude of the error found in PM6 calculations required the introduction of atom types to differentiate several types of hydrogen bonds. The correction is directional and should thus provide a better description of the hydrogen bonds than the standard core−core interaction used in semiempirical methods.

(4) The resulting PM6-DH method was tested on multiple sets of high-quality benchmark data. The results are superior to those obtained with the PM6 method alone. It has to be stressed that the accuracy of the method is close to that of correlated ab initio methods. The long-sought target accuracy for semiempirical QM methods, so-called chemical accuracy (error < 1 kcal/mol), was achieved for the S22 set. On the larger set of biomolecular complexes, the JSCH2005 database, it was shown that the description of different types of interactions is consistent and brings a significant improvement when compared to the uncorrected PM6 method.

(5) Although the method was derived for interaction energies in molecular complexes, the corrections are important also in isolated molecules featuring intramolecular noncovalent interactions. Excellent results were achieved in the description of conformational energies of small peptides, with a mean absolute error [compared to CCSD(T)/CBS results] amounting to 0.9 kcal/mol for a set of 76 structures. This accuracy is comparable to that of MP2/cc-pVTZ calculations.

(6) Implementation of a gradient makes this method useful for the optimization of systems with geometries determined by noncovalent interactions. The analytical calculation of gradients of the H-bond correction is not exact but uses an approximation. Nevertheless, the method performed well in geometry optimization tests, yielding geometries close to those obtained using the best QM methods available. There is a minor problem in the geometries of the strongest H-bonds associated with the form of the correction term, but it is outweighed by the other benefits. Optimization of the structure by PM6-DH leads to improvements in the relative energies of peptide conformers. This is, to the best of our knowledge, a unique feature because geometry optimization with other semiempirical QM methods usually strongly deteriorates the quality of the geometries obtained. Because hydrogen-bond types are determined only once (at the beginning of the calculation), the H-bond correction cannot be used for a continuous description of a reaction that changes valence states of the atoms involved in hydrogen bonds.

(7) The PM6-DH technique was further tested for various extended stacked complexes (porphine dimer, graphene models with DNA bases, and base pairs). It was shown that the method provides excellent stabilization energies that agree very closely with the benchmark values obtained by much more expensive DFT-D, SCS-MP2, or even MP2.5 methods. Finally, for the example of a DNA tetramer, we showed that PM6-DH can be used for geometry optimizations of rather large biomolecules.

(8) The PM6-DH calculations are very efficient and can be routinely applied to systems of up to 1000 atoms. Using the linear scaling MOZYME algorithm available in the PM6 implementation in MOPAC2009, nonaromatic systems with several thousand atoms can be calculated. With this performance, the method alone can replace molecular mechanics in calculations of smaller systems. In contrast to molecular mechanics, the present PM6-DH method fully and properly includes quantum effects. This fact and the efficiency of the method make its application in biological disciplines (concerning static or dynamic descriptions) extremely attractive for use in MD simulations.

Acknowledgment. This work was a part of Research Project Z40550506 of the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, and it was supported by Grants LC512 and MSM6198959216 from the Ministry of Education, Youth and Sports of the Czech Republic. The support of Praemium Academiae, Academy of Sciences of the Czech Republic, awarded to P.H. in 2007 is also acknowledged. D.R.S. is grateful to NSERC-Canada for ongoing Discovery Grant support. We also thank Martin Lepšík and Michal Pitoňák for inspiring discussions. We appreciate the discussion with J. J. P. Stewart, and we are grateful for his assistance and help with the original code. We are also grateful to the reviewers of our article for their insightful comments.

Supporting Information Available: Studying the error in PM6 results for hydrogen bonds and its dependence on the H-bond distance, we found a significant difference between hydrogen bonds with oxygen and nitrogen as the acceptor. This, of course, requires a different form of the correction that, after optimization, differs in the sign of the exponential term. It works as a repulsive term that lowers the magnitude of the correction for hydrogen bonds of oxygen, but it becomes a dominant attractive term for nitrogen. The demonstration of this issue on two typical systems from our training set (methanol···water and pyrazine···water complexes) is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286* (3−4), 243–252.

(2) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Olsen, J. *Chem. Phys. Lett.* **1999**, *302* (5−6), 437–446.

(3) Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2007**, *9* (39), 5291–5303.

(4) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, 4899–4907.

(5) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, 4907–4917.

(6) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107* (13), 3902–3909.

(7) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10* (2), 221–264.

(8) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10* (2), 209–220.

(9) Giese, T. J.; Sherer, E. C.; Cramer, C. J.; York, D. M. J. *Chem. Theory Comput.* **2005**, *1* (6), 1275–1285.

(10) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58* (11), 7260–7268.

(11) Yang, Y.; Yu, H. B.; York, D.; Cui, Q.; Elstner, M. *J. Phys. Chem. A* **2007**, *111* (42), 10861–10873.

(12) Winget, P.; Selcuki, C.; Horn, A. H. C.; Martin, B.; Clark, T. *Theor. Chem. Acc.* **2003**, *110* (4), 254–266.

(13) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114* (12), 5149–5155.

(14) Martin, B.; Clark, T. *Int. J. Quantum Chem.* **2006**, *106* (5), 1208–1216.

(15) Morgado, C. A.; McNamara, J. P.; Hillier, I. H.; Burton, N. A. *J. Chem. Theory Comput.* **2007**, *3* (5), 1656–1664.

(16) Tuttle, T.; Thiel, W. *Phys. Chem. Chem. Phys.* **2008**, *10* (16), 2159–2166.

(17) Stewart, J. J. P. *J. Mol. Model.* **2007**, *13* (12), 1173–1213.

(18) Voityuk, A. A.; Rosch, N. *J. Phys. Chem. A* **2000**, *104*, 4089–4094.

(19) Thiel, W.; Voityuk, A. A. *Theor. Chim. Acta* **1992**, *81* (6), 391–404.

(20) Thiel, W.; Voityuk, A. A. *Theor. Chim. Acta* **1996**, *93* (5), 315.

(21) Thiel, W.; Voityuk, A. A. *J. Phys. Chem.* **1996**, *100* (2), 616–626.

(22) Stewart, J. J. P. *J. Mol. Model.* **2008**, *14* (6), 499–535.

(23) Stewart, J. J. P. *MOPAC2007*; Stewart Computational Chemistry: Colorado Springs, CO, 2007; available at http://OpenMOPAC.net(accessed Nov 10, 2008).

(24) VAMP. In *Materials Studio 4.4*; Accelrys: San Diego, CA, 2008; available at http://accelrys.com/products/materials-studio/modules/VAMP.html(accessed Apr 22, 2009).

(25) Jug, K.; Geudtner, G. *J. Comput. Chem.* **1993**, *14* (6), 639–646.

(26) Černý, J.; Kabeláč, M.; Hobza, P. *J. Am. Chem. Soc.* **2008**, *130* (47), 16055–16059.

(27) Hobza, P.; Mulder, F.; Sandorfy, C. *J. Am. Chem. Soc.* **1981**, *103* (6), 1360–1366.

(28) Hobza, P.; Mulder, F.; Sandorfy, C. *J. Am. Chem. Soc.* **1982**, *104* (4), 925–928.

(29) Hobza, P.; Sandorfy, C. *Can. J. Chem.* **1984**, *62* (3), 606–609.

(30) Jurečka, P.; Černý, J.; Hobza, P.; Salahub, D. R. *J. Comput. Chem.* **2007**, *28* (2), 555–569.

(31) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8* (17), 1985–1993.

(32) Clark, T. *J. Mol. Struct. (THEOCHEM)* **2000**, *530* (1−2), 1–10.

(33) Dąbkowska, I.; Jurečka, P.; Hobza, P. *J. Chem. Phys.* **2005**, *122* (20), 204322.

(34) Distasio, R. A.; Head-Gordon, M. *Mol. Phys.* **2007**, *105* (8), 1073–1083.

(35) Grimme, S. *J. Chem. Phys.* **2003**, *118* (20), 9095–9102.

(36) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. *Chem. Phys. Lett.* **1989**, *162* (3), 165–169.

(37) Řezáč, J.; Jurečka, P.; Riley, K. E.; Černý, J.; Valdes, H.; Pluháčková, K.; Berka, K.; Řezáč, T.; Pitoňák, M.; Vondrášek, J.; Hobza, P. *Collect. Czech. Chem. Commun.* **2008**, *73* (10), 1261−1270. See also www.BEGDB.com.

(38) Valdes, H.; Pluháčková, K.; Pitoňák, M.; Řezáč, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2008**, *10* (19), 2747–2757.

(39) Feyereisen, M.; Fitzgerald, G.; Komornicki, A. 1. *Chem. Phy. Lett.* **1993**, *208* (5−6), 359–363.

(40) Muck-Lichtenfeld, C.; Grimme, S. *Mol. Phys.* **2007**, *105* (19−22), 2793–2798.

(41) Pitoňák, M.; Neogrdy, P.; Černý, J.; Grimme, S.; Hobza, P. *ChemPhysChem* **2009**, *10* (1), 282–289.

(42) Antony, J.; Grimme, S. *Phys. Chem. Chem. Phys.* **2008**, *10* (19), 2722–2729.

CT9000922