Article

New Approach for Correcting Noncovalent Interactions in Semiempirical Quantum Mechanical Methods: The Importance of Multiple-Orientation Sampling

Sergio Pérez-Tabero, Berta Fernández, Enrique M. Cabaleiro-Lago, Emilio Martínez-Núñez, and Saulo A. Vázquez*

Read Online

ACCESS | Metrics & More | Article Recommendations | SI Supporting Information

ABSTRACT: A new approach is presented to improve the performance of semiempirical quantum mechanical (SQM) methods in the description of noncovalent interactions. To show the strategy, the PM6 Hamiltonian was selected, although, in general, the procedure can be applied to other semiempirical Hamiltonians and to different methodologies. A set of small molecules were selected as representative of various functional groups, and intermolecular potential energy curves (IPECs) were evaluated for the most relevant orientations of interacting molecular pairs. Then, analytical corrections to PM6 were derived from fits to B3LYP-D3/def2-TZVP reference−PM6 interaction energy differences. IPECs provided by the B3LYP-D3/def2-TZVP combination of the electronic structure method and basis set were chosen as the reference because they are in excellent agreement with CCSD(T)/aug-cc-pVTZ curves for the studied systems. The resulting method, called PM6-FGC (from functional group corrections), significantly improves the performance of PM6 and shows the importance of including a sufficient number of orientations of the interacting molecules in the reference data set in order to obtain well-balanced descriptions.
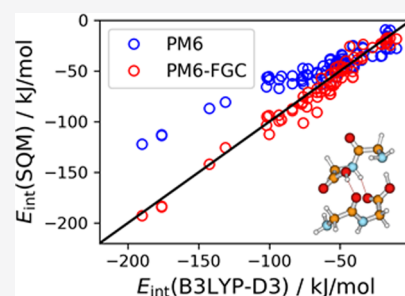
1. INTRODUCTION

One of the well-known problems inherent to semiempirical quantum mechanical (SQM) methods is the poor performance in describing noncovalent interactions.[1,2] Over the last years, much effort has been devoted to improve the accuracy of SQM methods for noncovalent interactions, particularly those based on the neglect of diatomic differential overlap (NDDO) approximation.[3,4] The most common strategy used to ameliorate the performance of SQM methods in calculations of intermolecular interactions has been the inclusion of empirical corrections.[5−21]

Řezáč, Hobza, and their co-workers developed several generations of corrections for dispersion,[6,7,9] hydrogen bond,[7,9] and halogen bond[8] interactions and parameterized them within the PM6 method[22] as well as for other SQM methods. Contributions to this series of generations were also made by Korth[10] and Jensen and co-workers.[11] The final version of this series of corrections is called D3H4X, in reference to the third-generation dispersion correction, fourth-generation hydrogen-bonding correction, and halogen-bonding correction. In this version, the dispersion correction is the D3 proposed by Grimme et al. for density functional theory (DFT),[23] but without including the $1/r^8$ term, which was considered to yield no significant improvement in the case of SQM methods.[9] For these methods, Řezáč and Hobza found a specific error in the description of interactions between hydrocarbons, namely, the overestimation of interaction

energies and the underestimation of equilibrium distances.[9] To solve this problem, they included a repulsive term for all pairs of hydrogen atoms. The function used to improve the description of hydrogen bonding includes a polynomial function of degree 7 in the donor−acceptor distance, which is scaled by an angular term (dependent on the acceptor-hydrogen-donor angle) and a proton transfer term that varies with the hydrogen position. If the system contains charged groups, an additional factor is included to increase the strength of the correction. Finally, the correction used for halogen bonding consists of an exponential term.[8] The D3H4X correction and other generations of corrections have been implemented in the MOPAC2016 program.[24]

The procedure adopted by Řezáč and Hobza to parameterize the D3H4 corrections was as follows:[9] First, they fitted the hydrogen-bonding correction, including the contribution from dispersion in the calculated energies of the considered hydrogen-bonded complexes. For the fittings, they performed least-squares optimizations, minimizing the root-mean-square

error of the interaction energy when compared to reference data obtained at the coupled cluster singles and doubles with perturbative triples correction/complete basis set (CCSD(T)/ CBS) level of calculation. Specifically, as a benchmark set, they used the S66 database,[25−27] which includes dissociation curves for 66 noncovalent complexes that exhibit dispersion, hydrogen bonds, and mixed dispersion/electrostatic interactions.

The work of Thiel and co-workers directed to improve the reliability of SQM methods also deserves some attention. They developed the orthogonalization-corrected methods OMx[33−35] and ODMx,[17] which include significant improvements in the semiempirical Hamiltonian, thus leading, in general, to better results in comparison with NDDO-based methods that make use of the modified neglect of diatomic overlap (e.g., AM1 or PM6). These semiempirical Hamiltonians needed to incorporate dispersion corrections to improve the description of noncovalent interactions. In particular, they include Grimme's D3 dispersion correction[23,36] with the Becke−Johnson damping function,[37−39] as well as Axilrod−Teller−Muto three-body terms,[23,40] which ameliorate the description of large dense systems.[41,42] For the recent ODMx methods, several training sets were used, including the abovementioned S66 data set.[25−27] Parameterization of semiempirical Hamiltonians and correction potentials for noncovalent interactions is a key issue, and the procedure followed within the ODMx methods is extensively discussed in the recent article by Dral et al.[17]

The abovementioned studies led to remarkable improvements in SQM methods for the evaluation of noncovalent interactions. In general, the corrections for dispersion and hydrogen bonding interactions are modeled by potential functions based on physically sound formulas. In addition, the training sets used for parameterizations are quite large, which may ensure a wide range of applicability. However, and as it will be shown later in the present work, the errors in the description of noncovalent interactions may be significant, depending on the relative orientation of the interacting molecules. This can be a consequence of possible shortcomings in popular data sets, which in general only include the most relevant configurations of interacting molecules.

In this paper, we present, as a proof-of-concept study, an alternative way to develop analytical corrections for SQM methods to improve the description of noncovalent

interactions. The idea is based on previous chemical dynamics studies in which pairwise intermolecular potentials were parameterized through fittings to a series of intermolecular potential energy curves (IPECs) that emphasize the different atom-pair potentials exhibited by the interacting molecules.[43−47] Following the strategy used to develop potentials for interactions of peptides with self-assembled monolayers of perfluorinated alkanes,[47] we selected small molecules as representatives for typical functional groups and evaluated IPECs for all possible molecular pairings. Specifically, for this proof-of-concept study, we have chosen methane, formic acid, and ammonia, which give six different pair combinations: the three dimers and the CH_4/HCO_2H, CH_4/NH_3, and NH_3/HCO_2H pairs of molecules. We developed empirical corrections for the PM6 method, which in principle can be used for interactions between hydrocarbons, carboxylic acids, and amines. To assess the performance of the method as well as the transferability of the corrections to other systems, we applied them to evaluate interaction energies for several complexes of the S66,[25] A24,[48] and ADIM6[49,50] data sets, as well as for a collection of different conformers of the diglycine dimer and trimer, and the dialanine dimer, obtained through automated exploration of the corresponding potential energy surfaces (PESs). The novelty of our approach is the introduction of two important features. First, and most important, the inclusion of several orientations of the interacting molecules in the database, which is crucial to obtain well-balanced corrections. Second, the use of general corrections to take into account that SQM methods have significant limitations to accurately describe not only dispersion interactions but also electrostatics, induction, and exchange repulsion. We notice that the purpose of this paper is not to present corrections with final parameters for universal applicability, but to show a strategy to develop corrections that can satisfactorily model noncovalent interactions for all orientations of interacting molecules.

2. METHODS

Intermolecular potential energy curves for the six pairs of molecules indicated above were calculated using CCSD(T)[51] and the augmented correlation consistent polarized valence triple-zeta basis set aug-cc-pVTZ.[52] The IPECs were also evaluated employing DFT with the B3LYP functional,[53−55] including the D3 dispersion correction with the Becke−Johnson damping scheme[37−39] and with the valence triple-zeta polarization def2-TZVP basis set.[56] The IPECs were computed using the supermolecular approach with frozen intramolecular geometries and correcting the interaction energy for basis set superposition error (BSSE) through the counterpoise method.[57,58] The intramolecular geometries were obtained by B3LYP-D3/def2-TZVP optimizations. Several orientations of the interacting molecules were selected to stress the different pair-type interactions. Specifically, for each pair of molecules, the number of orientations was at least equal to the number of the different pair-type interactions. A proper selection of orientations is crucial to obtain well-balanced corrections. These electronic structure calculations were performed with the ORCA 4.0 program and the default frozen core approximation.[59,60]

The general expression of the noncovalent potential−energy correction developed in this work for the PM6 method is written as a pairwise sum of the form

$$E_{corr} = \sum_i \sum_j f_{cut}(r_{ij}) \times \left\{ A_{ij} e^{-B_{ij} r_{ij}} + \frac{C_{ij}}{r_{ij}^{n_{ij}}} \right\} \tag{1}$$

where indexes i and j refer to atoms belonging to different interacting molecules, and r_{ij} is the interatomic distance between atoms i and j. Parameters A_{ij}, B_{ij}, and C_{ij} (real numbers) as well as n_{ij} (integers) depend on the nature of the considered pair of atoms. $f_{cut}(r_{ij})$ is a cutoff function introduced to remove the correction at very short r_{ij} distances

$$f_{cut}(r_{ij}) = (1 + \tanh(s_{ij}(r_{ij} - d_{ij})))/2 \tag{2}$$

where s_{ij} is a parameter that controls the strength of the damping for the interaction between atoms i and j, and d_{ij} is the distance at which the cutoff function takes the value 1/2. The n_{ij} parameters were not fixed to 6; rather, they were allowed to vary around this value. Also, the A_{ij} and C_{ij} parameters may be either positive or negative. We notice that eq 1 should be regarded as a practical correction, without any physical interpretation. However, one may expect the functional form given by eq 1, based on Buckingham's potential,[61] to work reasonably well because this potential can model intermolecular interactions with pretty good accuracy.

The abovementioned parameters were obtained through fittings to differences between the interaction energies calculated at the reference level and those computed with the PM6 method. As described in the next section, the IPECs obtained by B3LYP-D3/def2-TZVP calculations are in very good agreement with those determined at the CCSD(T)/aug-cc-pVTZ level of theory, showing the efficiency of the combination of the B3LYP-D3 density functional and the def2-TZVP basis set in providing acceptable and inexpensive IPECs. For our approach, the use of an accurate and inexpensive reference methodology is important because, in general, we need a thorough exploration of potential energy surfaces for both parameterization and validation processes. For this reason and considering that future work, for extending the method to other types of functional groups, will involve a large amount of calculations, we have selected the B3LYP-D3/def2-TZVP level as the reference for the fittings. Furthermore, for the molecular systems considered in this study, the errors of the fits are larger than the B3LYP-D3−CCSD(T) potential energy differences. The SQM calculations were carried out with the MOPAC2016 program.[24] We used a least-squares nonlinear fitting procedure based on a genetic algorithm, as implemented in our GAFit code,[62] and with the following objective function, χ^2

$$\chi^2(a) = \sum_{i=1}^{N} [y_i - f(x_i; a)]^2 \times w_i \tag{3}$$

where (x_i, y_i) represents one of the N data points, a is the collective variable formed by the total number of fitting parameters, and $f(x_i; a)$ is the value of the model function at x_i (i.e., a particular geometry of the interacting molecules). The square of the difference between y_i (i.e., a B3LYP-D3−PM6 energy difference) and the corresponding model value, calculated with eqs 1 and 2, may be multiplied by a weighting factor (w_i) assigned to each data point.

To validate our model function and parameterization strategy, as well as to explore the transferability of the corrections, we applied them to several complexes of the S66,[25]

A24,[48] and ADIM6[49,50] databases. Moreover, we checked the performance of the corrections on a data set formed by a collection of different conformations of diglycine and dialanine complexes, obtained by an automated exploration of the PESs at the PM6-D3H4 level, using the AutoMeKin package,[63−65] which has an interface with the MOPAC2016 program.[24] Specifically, we considered the diglycine dimer and trimer, as well as the dialanine dimer. Although AutoMeKin has mainly been designed to discover and simulate chemical reaction mechanisms, it includes an option for obtaining stationary points for intermolecular complexes.[66] As for the fittings, the benchmark level for this validation was B3LYP-D3/def2-TZVP, correcting the interaction energies for BSSE.

3. RESULTS AND DISCUSSION

3.1. Formic Acid Dimer. The formic acid molecule has five chemically nonequivalent atoms, that is, all the atoms are nonequivalent. Therefore, for this system, there are 15 different types of pairwise interactions. Consequently, for the fittings, we included 15 orientations that emphasize the distinct pairwise interactions, as well as an additional orientation corresponding to the global minimum of the formic acid dimer, which exhibits double hydrogen bonding. These 16 orientations are depicted in Figure 1.
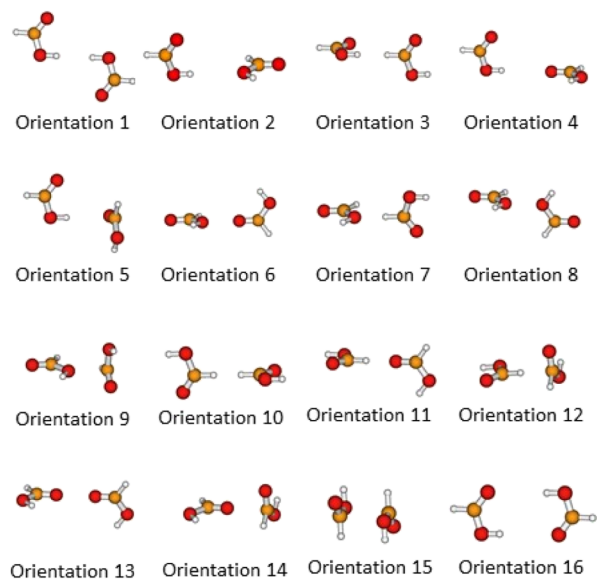
Figure 1. Orientations of the formic acid dimer considered for the fitting.

As pointed out in the previous section, for the selected orientations, we calculated intermolecular potential energy curves at the CCSD(T)/aug-cc-pVTZ and B3LYP-D3/def2-TZVP levels. As shown in the Supporting Information, Figures S1−S7, the agreement between the CCSD(T) and DFT curves is excellent. To develop corrections for the PM6 method, using the functional form specified in eqs 1 and 2, we used the differences between the B3LYP-D3/def2-TZVP interaction energies and the corresponding PM6 values as the data for the fittings. These energy differences have the typical forms displayed in Figure 2. In these four plots, r corresponds to the distance between attacking atoms, that is, the two carboxylic hydrogens in orientation 1, for example. The form of the energy difference as a function of r for this orientation
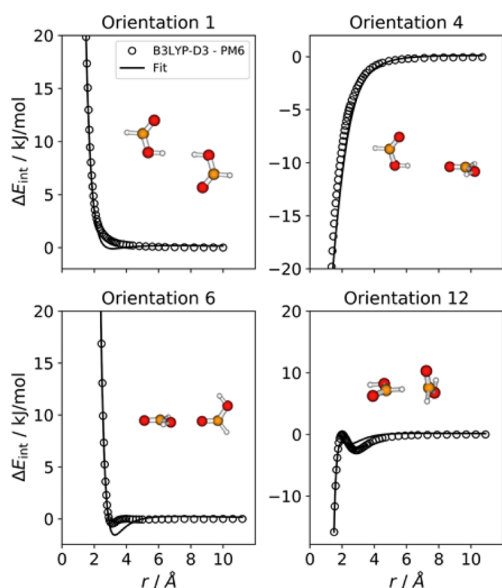
Figure 2. B3LYP-D3−PM6 interaction energy differences (open circles) for selected orientations of the formic acid dimer. The black lines correspond to the fit (see text).

resembles a typical repulsive potential, indicating that PM6 has a less repulsive IPEC than that of the reference method. By contrast, for orientation 4, the form of the DFT−PM6 plot behaves as a decaying exponential with negative amplitude (A_{ij} in eq 1), thus pointing out a stronger repulsion character of the PM6 interaction potential. Most of the orientations show a well followed by a pronounced increase of the energy difference as the distance between the attacking atoms becomes shorter, as can be seen for orientation 6 (carbonyl oxygen−hydroxyl oxygen attack). The presence of a well in these plots does not mean that the DFT and PM6 IPECs exhibit potential minima (although in most cases they do). Actually, for orientation 6, the DFT and PM6 IPECs are repulsive in nature, as can be seen in Figure 3. In very few cases (only one for the formic acid dimer), the B3LYP-D3−PM6 energy differences display a more complex form, showing both a minimum and a maximum, as for orientation 12, where the hydrogen attached to the carbonyl carbon attacks the carbon atom of the other molecule. Although it could be expected that eq 1 would be valid as a practical and simple correction for SQM methods, these plots further justify its use.

Using GAFit[62] and the geometries and energy differences corresponding to the 16 orientations of the formic acid dimer, we simultaneously fitted the parameters involved in eqs 1 and 2. There is not a universal, objective way to conduct a parametrization and, furthermore, with the use of genetic algorithms, one may obtain many solutions than can be equally valid. For general discussions on parameterizations for SQM methods, the reader may consult the studies on the development of the PMO2a[16] and ODMx[17] methods. For the formic acid dimer, because there are 15 different types of pairwise interactions, the total number of parameters is 60, without including those associated with the cutoff function given by eq 2. All the 60 parameters were fitted simultaneously. We found that including the cutoff parameters into the parameter spectrum explored by the genetic algorithm did not improve the fittings significantly. For this reason, after some analyses and to avoid overparameterization, we have chosen a value of 10 for all the s_{ij} parameters and different values for the
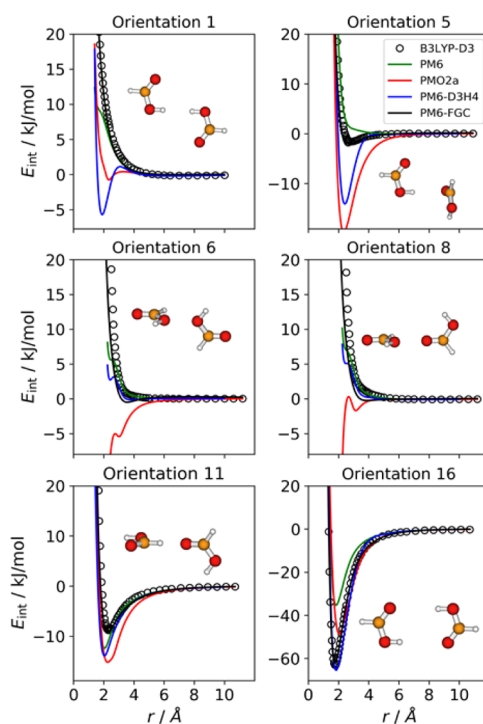
Figure 3. Comparison of IPECs for six selected orientations of the formic acid dimer.

d_{ij} parameters, depending on the nature of atoms i and j. The parameters obtained from our best fit are collected in Table 1,

Table 1. Parameters Obtained for the Formic Acid Dimer[a]

atom pair	A	B	C	n	d
C−C	87222.68	4.52	−2614.24	6	1.8
C−O	18423.16	3.22	−4.76	10	1.7
C−OH	101988.82	3.76	−801.90	6	1.7
C−HO	−118682.03	5.31	49.92	4	1.2
C−HCO	−91515.91	5.31	5.68	2	1.2
O−O	157533.31	3.84	−249.70	9	1.7
O−OH	229576.11	3.74	−1494.51	6	1.7
O−HO	−12321.28	5.15	−106.36	3	1.0
O−HCO	7911.47	3.45	−735.00	9	1.2
OH−OH	282168.52	3.74	−1073.49	5	1.7
OH−HO	−3746.89	3.44	−49.80	4	1.0
OH−HCO	3900.44	2.91	−380.87	8	1.2
HO−HO	2029.79	3.07	27.04	2	1.0
HO−HCO	10653.65	4.26	−396.77	11	1.0
HCO−HCO	9870.51	3.49	−289.72	5	1.0

[a]The units are such that the potential energy is in kJ/mol and distances in Å.

and the fit results are depicted in Figure 2 for some selected orientations. For simplicity, for parameters A, B and C, we only show two decimals; the high precision parameters are included in Table S1 in the Supporting Information. Notice that we have defined atom types in much the same way as in molecular mechanics force fields. The symbols chosen in this work are shown in Figure 4. Adding to the PM6 interaction potential the corrections calculated with eqs 1 and 2, and the parameters fitted in this work, results in our PM6-FGC method. We have chosen the FGC acronym, from functional group corrections,
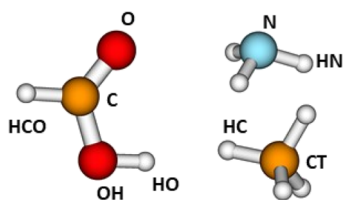
Figure 4. Atom types defined for ammonia, formic acid, and methane.

to emphasize the idea of specific parameters being obtained for different functional groups.

Figure 3 compares the reference IPECs with the PM6-FGC interaction curves for six selected orientations of the formic acid dimer. The global minimum corresponds to orientation 16. The IPECs for the remaining orientations are displayed in Figure S8 in the Supporting Information. For comparison, we include the PM6 curves as well as those obtained with the PM6-D3H4[9] and PMO2a[16] methods, which are implemented in the freely distributed MOPAC2016[24] and MOPAC 5.022mn[32] programs, respectively. It would be interesting to include results of calculations using the ODMx method;[17] however, to our knowledge, the code in which this method is implemented is not freely available.[67] As can be seen from Figures 3 and S8, the PM6-FGC curves (black lines) agree well with the B3LYP-D3/def2-TZVP data (black open circles).

The IPECs calculated with the PMO2a method (red lines) display remarkable discrepancies with the reference curves. Strikingly, for orientations 6 and 8, as well as for orientation 13 (see Supporting Information), this method shows an unphysical behavior, because the interaction energy in the repulsive region decreases as the distance between the attacking atoms decreases. These orientations correspond to configurations that emphasize the interaction between oxygen atoms. Clearly, a revision of this method is required to improve its performance. For this reason, for the remaining systems under investigation here, we have not considered the PMO2a method any further.

The PM6-D3H4 potential energy curves are displayed as blue lines in the figures. The D3H4 corrections were developed using a training set based on CCSD(T)/CBS data, so that slight deviations may be expected because our IPECs were obtained with B3LYP-D3/def2-TZVP calculations, although, as already mentioned, they are in excellent agreement with the corresponding CCSD(T)/aug-cc-pVTZ curves. As can be seen, for orientation 1, which corresponds to the attack between carboxyl hydrogens, the PM6-D3H4 method exhibits a clear minimum. This orientation is predicted to be repulsive with both B3LYP-D3 and CCSD(T) methods. Also, for orientation 5, the PM6-D3H4 method gives a significant minimum, which contrasts with the small well depth predicted by the reference calculations. It is also worth mentioning that, for orientations 6 and 8, PM6-D3H4 and PM6 exhibit an unphysical behavior in the repulsive region (around 5 kJ/mol). Although the S66x8 data set[25] employed by Hobza and co-workers comprises a wide range of complexes, and the D3H4 corrections are able to describe noncovalent interactions for the most relevant orientations of interacting molecules, our results point out some deficiencies in these corrections, which may be especially problematic for dynamics studies, where all orientations may be sampled. The source of these deficiencies comes from an important drawback of the S66x8 database, namely, the fact that, in general, it only includes the most relevant orientation for each selected

complex. For carboxylic acids, this database includes the acetic acid dimer in its most attractive orientation, that is, that exhibiting a double hydrogen bond (the equivalent of orientation 16 for the formic acid dimer, Figure 1).

3.2. Ammonia Dimer. This dimer shows three different types of pairwise interactions, and therefore, we need at least three different orientations. In this work, we considered the four orientations depicted in Figure 5, which compare the
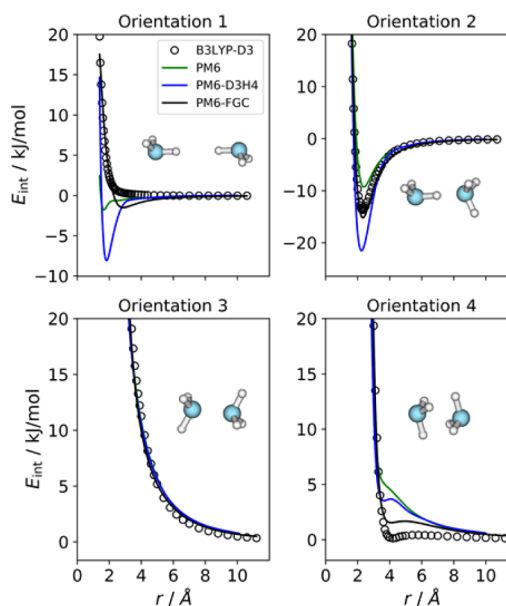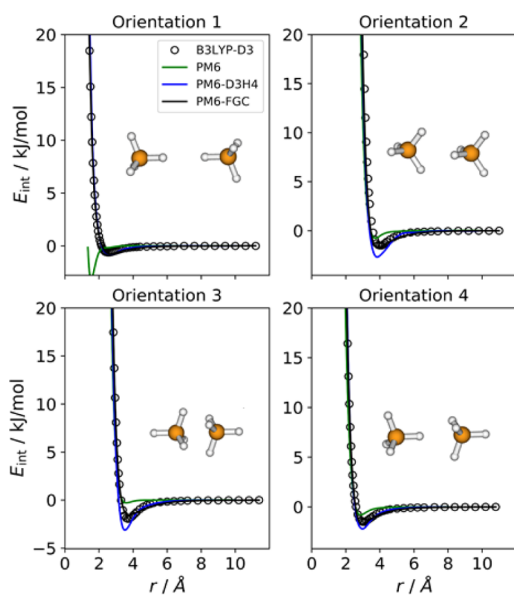

Figure 5. Comparison of IPECs for the considered orientations of the ammonia dimer.

IPECs obtained with PM6 (green lines), PM6-D3H4 (blue lines), and PM6-FGC (black lines), with those determined with the reference method (open circles). As can be seen from Figure 5, the PM6-D3H4 method exhibits substantial deficiencies, similar to those encountered in the formic acid dimer. Specifically, it predicts a remarkable minimum for orientation 1 (H···H attack), which is clearly repulsive at the reference level. In addition, for orientation 2, which exhibits hydrogen bonding, the PM6-D3H4 method clearly overestimates the strength of the interaction.

Among the systems considered in this study, the ammonia dimer was the most challenging. Actually, using our simple expression for the analytical correction (eqs 1 and 2), we were not successful at obtaining a good fit, as reflected in Figure 5. Particularly, for orientation 4, PM6-FGC as well as PM6 and PM6-D3H4 show curves more repulsive than that predicted with the benchmark method. One way to improve the fit is to add a pseudoatom to model the effect of the nitrogen lone pair, as it is done in several force fields, but for this proof-of-concept presentation, we wanted to keep the correction scheme as simple as possible. The parameters obtained from the ammonia dimer fit are collected in Table S1.

3.3. Methane Dimer. The IPECs evaluated for this dimer are displayed in Figure 6, which also describes the orientations selected in this work. In principle, this system appears to be the simplest one among those studied here. As can be seen, both PM6-D3H4 and PM6-FGC, using the parameters shown in Table S1, exhibit IPECs in satisfactory agreement with the reference curves. The underestimation of the dispersion interaction in the PM6 method is clear, but the worse

Figure 6. Comparison of IPECs for the considered orientations of the methane dimer.

performance predicted with this method appears for orientation 1 (i.e., H⋯H attack), which shows a significant minimum at a quite short H⋯H distance.

3.4. Ammonia−Formic Acid Complex. For this complex, we considered 10 orientations, that is, the same number as that of the different types of pairwise interactions. The best fit for this system led to the parameters reported in Table S1. Figure 7 depicts four selected orientations, together with their IPECs.
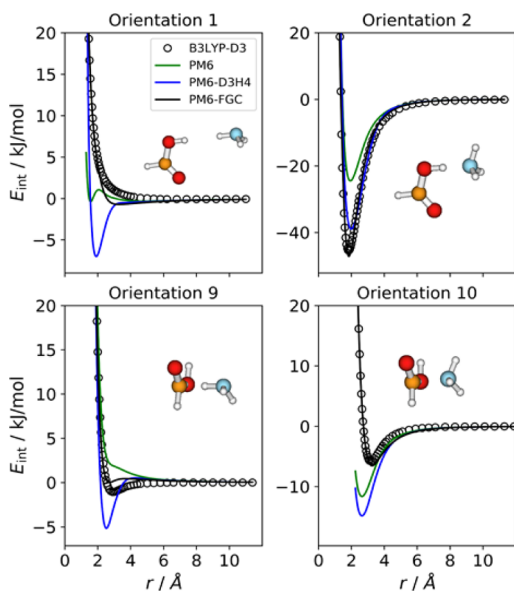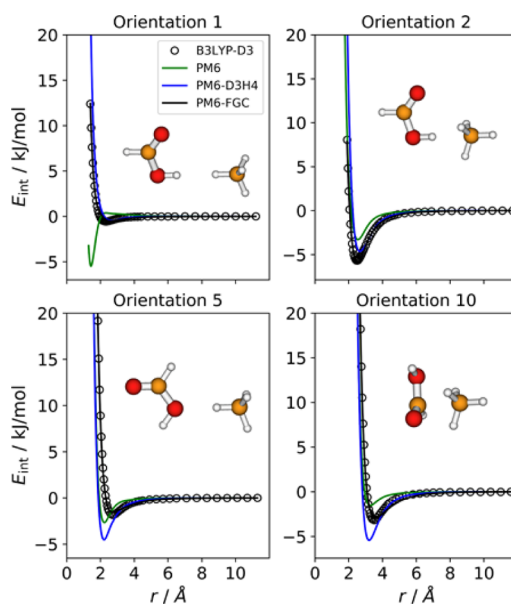

Figure 7. Comparison of IPECs for four selected orientations of the HCOOH/NH$_3$ complex.

The plots for the remaining orientations are shown in Figure S9. The most attractive orientation exhibits a hydrogen bond between the carboxylic hydrogen and the ammonia nitrogen (orientation 2). Both the PM6-D3H4 (blue line) and the PM6-FGC (black line) methods satisfactorily describe the interaction for this orientation. However, for several other orientations (1, 5, 7, 9, and 10), the PM6-D3H4 method

predicts minima with potential well depths larger than those obtained through CCSD(T) and B3LYP-D3 calculations.

For orientation 1, that is, the attack between carboxylic and ammonia hydrogens, the behavior of the PM6-D3H4 curve resembles that found for orientation 1 in the formic acid and ammonia dimers. In these three cases, the reference IPECs clearly exhibit a repulsive character. The PM6-FGC curve shows a small deviation from the reference IPEC, similar to that exhibited in the ammonia dimer. For orientations 9 and 10, which correspond to the attack of the carbonyl carbon to ammonia hydrogen and nitrogen, respectively, the PM6-D3H4 (and PM6) curves also show remarkable discrepancies with respect to the reference curves. For orientation 9, the PM6-FGC curve exhibits a small deviation from the benchmark IPEC. Overall (see also Figure S9), the PM6-FGC method gives a satisfactory description of the noncovalent interaction in the ammonia-formic acid complex.

3.5. Methane−Formic Acid Complex. For this complex, we considered 10 orientations, and four of them are depicted in Figure 8, together with their corresponding IPECs. The

Figure 8. Comparison of IPECs for four selected orientations of the HCOOH/CH$_4$ complex.

IPECs of the remaining orientations are shown in Figure S3, and the parameters obtained in the fit are displayed in Table S1. As can be seen, the PM6-FGC curves agree well with the corresponding benchmark IPECs. The PM6-D3H4 method shows, in general, satisfactory performance, although for several orientations (e.g., 5 and 10), it predicts more attractive interactions. As expected, the PM6 interaction energies are very inaccurate, and for orientations 1 and 3 (attacks of hydrogen atoms), the corresponding IPECs display remarkable minima.

3.6. Ammonia−Methane Complex. Four different orientations were considered for this complex, and they are displayed in Figure 9, together with their IPECs. The agreement between the IPECs obtained with the PM6-FGC method and those evaluated at the reference level reflects the good quality of the fit. The PM6-D3H4 method also predicts satisfactory interaction energies, except for orientation 4, for which it provides a significant potential well depth.
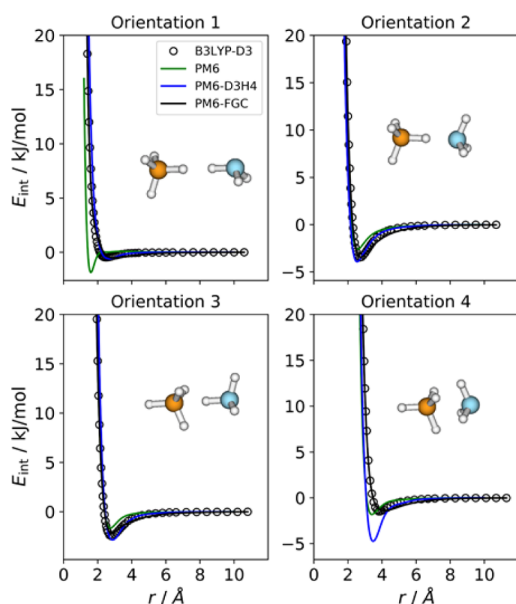
Figure 9. Comparison of IPECs for the considered orientations of the CH_4/NH_3 complex.

3.7. Validation and Critical Assessment of the FGC Approach. There are two critical issues that we need to consider in order to validate our method: (1) the possibility of overfitting and (2) parameter transferability. Overfitting occurs when the model function uses more parameters than are necessary or exhibits a more complex form than is needed.[68] Considering the drastic approximations inherent to SQM methods, it is clear that these approaches fail in describing not only dispersion interactions but also electrostatic, induction, and exchange-repulsion. Our analytical correction consists of a pairwise sum over interacting atoms, with four parameters per interaction type. The form of the energy differences between the interaction energies calculated at the reference level and those obtained with the PM6 Hamiltonian, as exemplified in Figure 2 for selected orientations of the formic acid dimer, suggests that our approach is reasonable. Therefore, the remaining question is whether our model function and parameter dimensionality are appropriate to correct the PM6 deficiencies, without leading to overfitting. At this point, we want to remark the importance of including many orientations in the training set. Using a single orientation seems to be insufficient to parameterize well-balanced corrections.

To analyze the possibility of overfitting in our approach, let us focus on the formic acid dimer. For the fittings, we have considered that all the atoms are different, which led to 60 parameters in all, and they were simultaneously fitted using IPECs for 16 different orientations. The additional 15 parameters of the cutoff function were set, rather than fitted, after some exploratory fittings. To investigate overfitting, we have performed two sets of additional fittings: (i) considering that all oxygen atoms share the same parameters (4 atom types) and (ii) considering that, in addition, the hydrogen atoms are equivalent to each other (3 atom types). These simplifications lead to 40 and 24 parameters, respectively. For the fittings, we used the same number of points and weights as for the fittings described previously for this system (in particular, each data point had a weight of unity). The data points extended up to repulsive energies ranging from 23 to 141 kJ/mol, depending on the orientation. With the corrections obtained from the best fits (parameters shown in Tables S2 and S3), we evaluated the IPECs for the 16 orientations of the formic acid dimer, and they are plotted in Figures S11 and S12, together with the B3LYP-D3 curves and those determined with the parameters of Table 1 (5 atom types). As can be seen, when we use 4 atom types, the calculated IPECs are very similar to those evaluated with the parameters of Table 1 (i.e., five atom types). This indicates that the two oxygen atoms are quite similar two each other, as far as interaction potentials are concerned. However, when we use only 3 atom types, the corrections added to the PM6 Hamiltonian cannot accurately describe the noncovalent interactions in this dimer because significant deviations from the reference curves are found for several orientations. Therefore, the HO and HCO atom types differ significantly from each other. To quantify the deviations from the B3LYP-D3 curves, we calculated the mean absolute errors (MAEs), for the three PM6-corrected IPECs, obtaining 0.7, 0.9, and 2.8 kJ/mol for the curves determined using 5, 4, and 3 atom types, respectively.

In addition to the previous analysis, we evaluated IPECs for 10 random orientations of the formic acid dimer, which were not included in the fittings. The comparison of these IPECs is shown graphically in Figure S13. As can be seen, the PM6-FGC curves (using the parameters given in Table S1) are, in general, in good agreement with the DFT curves. For orientation 4, the PM6-FGC curve displays a potential well less pronounced than that of the reference level. The most important deviations are exhibited by the PM6-D3H4 curves
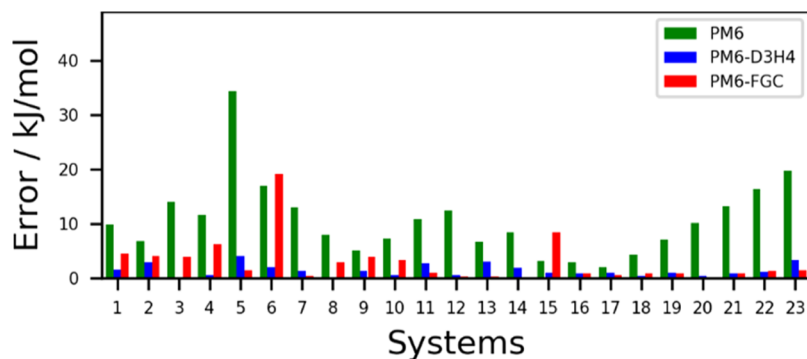
Figure 10. Absolute errors calculated for 23 complexes of well-known data sets. The complex associated with each system number is specified in Table S4. The benchmark interaction energies were taken from the BEGDB web page.[69]

for orientations 3 and 7. Notice that, for several orientations, both PM6 and PM6-D3H4 curves deviate significantly from the reference curves in the repulsive region. Finally, we notice that the MAE calculated for the PM6-FGC curves, including all the points with interaction energies up to 100 kJ/mol, is 1.1 kJ/mol, slightly smaller than that calculated employing the FGC parameters obtained using 4 atom types (1.3 kJ/mol), thus resembling the results reached for the training set. From the discussion in this and the previous paragraph, we conclude that our parameterization model does not exhibit overfitting, at least at a significant level.

As a further validation of our method, and particularly to analyze the transferability of the corrections to related systems, we calculated the interaction energies for 23 complexes defined in well-known databases[25] and compared the results with benchmark data taken from the GMTKN55[50] and BEGDB[69] web pages. All the benchmark interaction energies correspond to CCSD(T)/CBS calculations (with BSSE corrections). Table S4 specifies the complexes considered for this comparison and gives the values of the benchmark interaction energies as well as those calculated with the PM6, PM6-D3H4 and PM6-FGC approaches. The corresponding absolute errors are displayed in Figure 10, and the calculated MAEs are 10.6, 1.4, and 2.9 kJ/mol for PM6, PM6-D3H4, and PM6-FGC, respectively. Our method significantly improves the results of the PM6 Hamiltomian; but in general, it shows worse performance than the PM6-D3H4 method, although we notice that most of the complexes included in this comparison were used in the parameterization of the PM6-D3H4 method. Anyway, this comparison shows important information to assess the transferability of our corrections and provides guidelines for future improvements. We analyze these issues in the next paragraphs.

As shown in Table S4, systems 1−4 correspond to complexes of methylamine and *N*-methylacetamide (denoted in BEGDB as peptide). For these systems, our results are worse than the PM6-D3H4 results. Notice that the training set used for the parameterization of the PM6-D3H4 method was the S66x8 database,[25] which includes systems 1−14. For system 6, the acetamide dimer, the performance of our method is clearly unsatisfactory. Interestingly, the PM6-FGC interaction energy calculated for the complex of pentane/acetamide (−14.8 kJ/mol), system 14 in Figure 10, is in excellent agreement with the S66 benchmark data (−14.7 kJ/mol). The interactions of pentane with *N*-methylacetamide and with acetic acid, systems 12 and 13, respectively, are also very well described by our method. These results are in line with previous work[47] wherein pairwise analytical potentials obtained from fittings involving the CF_4/NH_3 and the $CF_4/HCOOH$ systems reproduced ab initio IPECs calculated for the $CF_4/HCONH_2$ complex.

The results discussed in the previous paragraph clarify an interesting feature regarding transferability in the studied systems. Although parameters obtained from NH_3/CH_4 fittings may be transferable to model $-NH_2$/alkane interactions, other parameters obtained with NH_3 may not be transferable to amines and amides. The description of noncovalent interactions involving the $-NH_2$ group in amines and amides could be improved by including new representative compounds, namely, methylamine and acetamide. The water dimer is another challenging case; we did not succeed in developing corrections as defined in eqs 1 and 2. One way to overcome this problem involves the inclusion of pseudoatoms. These issues will be considered in future work.

It is important to notice that the interaction energies calculated at the B3LYP-D3/def2-TZVP level for methylamine and acetamide dimers (−18.0 and −70.1 kJ/mol, respectively) are in very good agreement with the reference CCSD(T)/CBS(haTZ) values, (−17.5 and −68.8 kJ/mol, respectively).[69] For the sake of comparison, the corresponding CCSD(T)/aug-cc-pVTZ interaction energies are −16.7 and −65.9 kJ/mol, respectively. These results reinforce the use of B3LYP-D3/def2-TZVP data as the benchmark for our parameterizations.

For the hydrocarbon systems considered in Figure 10 (systems 7−11 and 16−23; see also Table S4), our method predicts reasonably good interaction energies. There are, however, some deviations for complexes involving neopentane (8 and 9). Finally, for system 15, which corresponds to the formaldehyde dimer, the interaction energy calculated with our method has an absolute error larger than that obtained with PM6. This indicates that the parameters of the carbonyl group developed from fittings on the formic acid dimer are not transferable to aldehydes and ketones, which supports the idea of using corrections for specific functional groups.
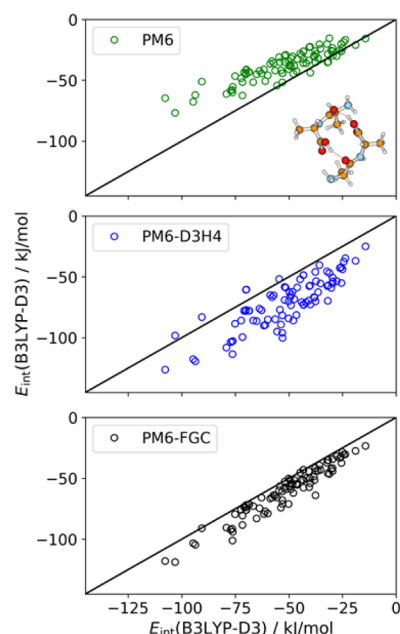
Even though the present corrections cannot accurately describe some interactions involving the $-NH_2$ group, it is important to assess the performance of the method on larger systems. Because our purpose is to derive corrections for biological systems, here, we have studied complexes of diglycine and dialanine. Specifically, we have considered the dimers of these dipeptides as well as the trimer of diglycine. In this work, conformers of these complexes were found by automated exploration of their potential energy surfaces at the PM6-D3H4 level, using AutoMeKin,[63−66] which has an interface with the MOPAC2016 program.[24] These searches involved changes in both intramolecular and intermolecular conformations and led to 77 and 90 different conformers for diglycine and dialanine dimers, respectively, and 146 conformers for the diglycine trimer. Figure 11 shows linear correlations between the interaction energies calculated at the B3LYP-D3/def2-TZVP level and those evaluated with the SQM methods considered here, for the complexes of the diglycine dimer. The diagonal straight lines represent the case of perfect correlation. As can be seen, our corrections significantly improve the performance of the PM6 Hamiltonian (shown as green open circles in Figure 11). As shown in Table 2, the mean absolute errors (MAE) and mean bias errors (MBE) calculated for both PM6-FGC parameter sets are substantially smaller than those computed for PM6. The bias error is calculated as the mean of the differences between the reference values and the SQM values. In the case of the diglycine dimer, the MAE calculated for PM6-D3H4 is very similar to that of PM6 (∼18 kJ/mol). Likewise, the bias values for these two methods are very similar to each other in absolute value. However, the PM6 method underestimates the strength of the interactions (negative MBE), especially for the most attractive conformers of the complexes, whereas the PM6-D3H4 method exhibits overestimation (positive MBE), especially for conformers showing from low to moderate interaction strengths.

The linear correlations evaluated for the conformers of the dialanine dimer are depicted in Figure 12. The MAE and MBE values (Table 2) calculated for PM6-D3H4 are somewhat larger than those computed for PM6. Again, the PM6 and the PM6-D3H4 methods lead to underestimation and overestimation, respectively, of the interaction energies. Finally, Figure 13 depicts the linear correlations for the diglycine
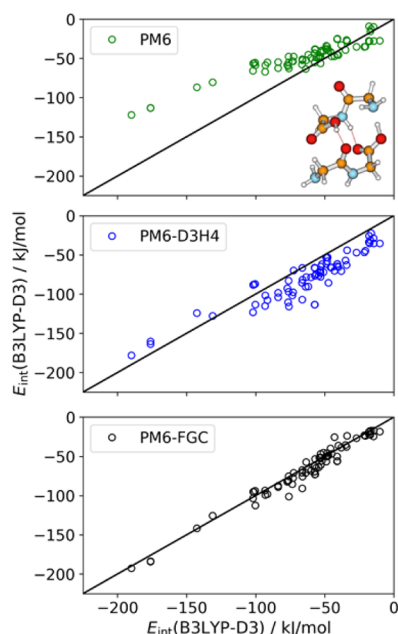
Figure 11. Linear correlations between the B3LYP-D3/def2-TZVP interaction energies and those calculated with PM6, PM6-D3H4, and PM6-FCG, for conformers of the diglycine dimer. The molecular drawing corresponds to the conformer with the largest interaction strength as calculated at the DFT level.

Table 2. Statistical Parameters[a] of the Linear Correlations between B3LYP-D3/def2-TZVP Interaction Energies and the Different PM6 SQM Interaction Energies Calculated for the Conformers of the Diglycine and Dialanine Dimers, as Well as the Diglycine Trimer

	diglycine dimer		dialanine dimer		diglycine trimer	
	MAE	MBE	MAE	MBE	MAE	MBE
PM6	17.9	−15.0	14.3	−14.0	10.8	−5.1
PM6-D3H4	18.6	15.5	20.3	19.5	25.3	25.3
PM6-FGC	6.0	3.2	8.7	8.7	9.1	8.7

[a]MAE and MBE values are given in kJ/mol.

trimer. For this system, the calculated interaction energies correspond to the difference between the energy of the trimer and that of separated dimer and monomer. As can be seen, the figure shows the same trend observed in Figures 11 and 12. The PM6 method underestimates the strength of the interaction energies whereas PM6-FGC and, especially, PM6-D3H4 overestimate it. The MAE calculated for PM6-D3H4 is more than twice the PM6 or PM6-FGC MAE. However, the errors in the PM6 interaction energies for the most attractive complexes are remarkable. Although a direct comparison with PM6-D3H4 results should be taken with caution because the latter corrections were determined from a CCSD(T)/CBS reference; the results of the present work provide evidence of deficiencies in the PM6-D3H4 method (and other SQM methods). These deficiencies are mainly the result of using, for the parameterization scheme, a data set (i.e., the S66 database[25,27]) that does not include sufficient orientations of the interacting molecules.

Finally, we would like to point out that our corrections may be problematic in the study of chemical reactions, since, in general, SQM methods do not predict accurate reaction barriers. To overcome this problem, we recommend the use of

Figure 12. Linear correlations between the B3LYP-D3/def2-TZVP interaction energies and those calculated with PM6, PM6-D3H4, and PM6-FCG, for conformers of the dialanine dimer. The molecular drawing corresponds to the conformer having the largest interaction strength as calculated at the DFT level.
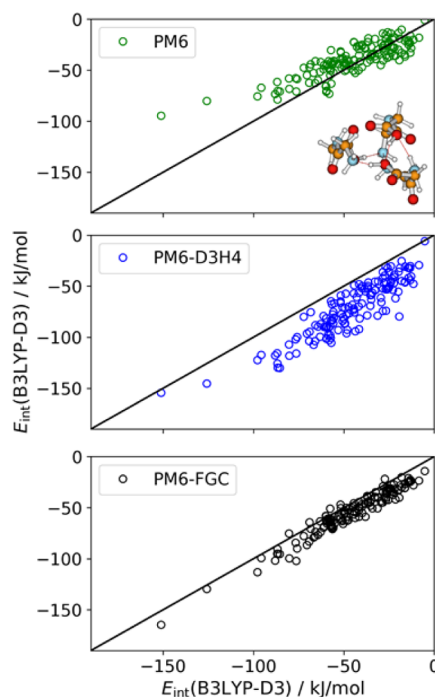

Figure 13. Linear correlations between the B3LYP-D3/def2-TZVP interaction energies and those calculated with PM6, PM6-D3H4, and PM6-FCG, for conformers of the diglycine trimer. The molecular drawing corresponds to the conformer having the largest interaction strength as calculated at the DFT level.

specific reaction parameters (SRPs), following the strategy pioneered by Truhlar and co-workers.[70] SRPs should be developed and applied to the atoms involved in the reaction, and our corrections should not be used for these atoms.

4. CONCLUSIONS

We have presented a new strategy, that is, the PM6-Functional Group Corrections (PM6-FGC) method, to develop analytical corrections for semiempirical quantum mechanical methods, aimed at improving the description of noncovalent interactions. For this proof-of-concept presentation, we have selected the PM6 SQM method. Employing this Hamiltonian and the B3LYP-D3/def2-TZVP level for benchmarking, we calculated intermolecular potential energy curves for several orientations of pairs of small molecules, which are selected as representatives of different functional groups. Specifically, we have considered ammonia, formic acid, and methane, which result in a total of six molecular pairings. A simple mathematical expression, which should be considered as a practical way, without any physical meaning, is used to represent the corrections to improve the performance of SQM methods. The parameters of the analytical corrections were evaluated from fits to differences between interaction energies calculated at the reference level and those evaluated at the SQM level.

Intermolecular potential energy curves (IPECs) were also evaluated at the CCSD(T)/aug-cc-pVTZ, including counterpoise correction for BSSE. The agreement between the B3LYP-D3 and the corresponding CCSD(T) curves was excellent. This result together with the fact that we will need to perform extensive benchmark calculations to extend our method to other functional groups, in order to develop a universally applicable method for peptides and other biological systems, led us to select the DFT method as the benchmark level for the parameterizations.

In general, the IPECs obtained with the method parameterized in this work, namely, PM6-FGC, are in good agreement with those determined at the DFT level, and significantly improve those provided by successfully corrected SQM methods. In this way, this work emphasizes the importance of including, in the databases, sufficient orientations of the interacting molecules; this fact is crucial to develop well-balanced corrections.

Even though the corrections proposed in this work are based on general pairwise functions that include a significant number of parameters, several validation tests suggest that our parameterization strategy is not affected by overfitting, at least at a significant level. A comparison of interaction energies calculated for a series of complexes of well-known data sets has revealed important warnings concerning parameter transferability. In particular, we found that the ammonia dimer is not a good representative for the parameterization of corrections for some interactions involving the $-NH_2$ group in amines and amides. This issue will be faced in future work. However, using the present set of parameters, we evaluated interaction energies for a large number of conformers of diglycine and dialanine dimers, and diglycine trimer, obtaining very good results in comparison with those predicted by PM6 and PM6-D3H4.

Although the method described here involves simple pairwise functions, more accurate corrections could be developed by using or adding alternative functions, as well as by introducing three-body terms of the type of the Axilrod−Teller−Muto potential.[23,40] We are continuing with our efforts to improve and extend our corrections to other functional groups, relevant to biological compounds, and we intend to implement them in the MOPAC program.[24] Meanwhile, a Python script to compute PM6-FGC corrections is available upon request.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.1c00365.

> Parameters obtained from fittings to B3LYP-D3/def2-TZVP−PM6 interaction energy differences; parameters obtained for the formic acid dimer using 4 atom types and 3 atom types; interaction energies for several complexes of well-known data sets; comparisons of IPECs calculated at the CCSD(T)/aug-cc-pVTZ and B3LYP-D3/def2-TZVP; comparison of IPECs for orientations not shown in the article ; comparison of IPECs for the formic acid dimer obtained with corrections determined using 5, 4, and 3 atom types; and comparison of IPECs for 10 random orientations of the formic acid dimer (PDF)

AUTHOR INFORMATION

Corresponding Author

Saulo A. Vázquez − *Departamento de Química Física, Facultade de Química, Universidade de Santiago de Compostela, Santiago de Compostela 15782, Spain;* ⓘ orcid.org/0000-0002-2473-4557; Email: saulo.vazquez@usc.es

Authors

Sergio Pérez-Tabero − *Departamento de Química Física, Facultade de Química, Universidade de Santiago de Compostela, Santiago de Compostela 15782, Spain*

Berta Fernández − *Departamento de Química Física, Facultade de Química, Universidade de Santiago de Compostela, Santiago de Compostela 15782, Spain;* ⓘ orcid.org/0000-0001-6686-6534

Enrique M. Cabaleiro-Lago − *Departamento de Química Física, Facultade de Química, Universidade de Santiago de Compostela, Santiago de Compostela 15782, Spain;* ⓘ orcid.org/0000-0001-5848-6523

Emilio Martínez-Núñez − *Departamento de Química Física, Facultade de Química, Universidade de Santiago de Compostela, Santiago de Compostela 15782, Spain;* ⓘ orcid.org/0000-0001-6221-4977

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.1c00365

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Christensen, A. S.; Kubař, T.; Cui, Q.; Elstner, M. Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications. *Chem. Rev.* **2016**, *116*, 5301−5337.

(2) Husch, T.; Vaucher, A. C.; Reiher, M. Semiempirical molecular orbital models based on the neglect of diatomic differential overlap approximation. *Int. J. Quantum Chem.* **2018**, *118*, No. e25799.

(3) Pople, J. A.; Santry, D. P.; Segal, G. A. Approximate Self-Consistent Molecular Orbital Theory. I. Invariant Procedures. *J. Chem. Phys.* **1965**, *43*, S129−S135.

(4) Husch, T.; Reiher, M. Comprehensive Analysis of the Neglect of Diatomic Differential Overlap Approximation. *J. Chem. Theory Comput.* **2018**, *14*, 5169−5179.

(5) McNamara, J. P.; Hillier, I. H. Semi-empirical molecular orbital methods including dispersion corrections for the accurate prediction of the full range of intermolecular interactions in biomolecules. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2362−2370.

(6) Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P. Semiempirical Quantum Chemical PM6 Method Augmented by Dispersion and H-Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes. *J. Chem. Theory Comput.* **2009**, *5*, 1749−1760.

(7) Korth, M.; Pitoňák, M.; Řezáč, J.; Hobza, P. A Transferable H-Bonding Correction for Semiempirical Quantum-Chemical Methods. *J. Chem. Theory Comput.* **2010**, *6*, 344−352.

(8) Řezáč, J.; Hobza, P. A halogen-bonding correction for the semiempirical PM6 method. *Chem. Phys. Lett.* **2011**, *506*, 286−289.

(9) Řezáč, J.; Hobza, P. Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *J. Chem. Theory Comput.* **2012**, *8*, 141−151.

(10) Korth, M. Third-Generation Hydrogen-Bonding Corrections for Semiempirical QM Methods and Force Fields. *J. Chem. Theory Comput.* **2010**, *6*, 3808−3816.

(11) Kromann, J. C.; Christensen, A. S.; Steinmann, C.; Korth, M.; Jensen, J. H. A third-generation dispersion and third-generation hydrogen bonding corrected PM6 method: PM6-D3H+. *PeerJ* **2014**, *2*, No. e449.

(12) Fiedler, L.; Gao, J.; Truhlar, D. G. Polarized Molecular Orbital Model Chemistry. 1. Ab Initio Foundations. *J. Chem. Theory Comput.* **2011**, *7*, 852−856.

(13) Zhang, P.; Fiedler, L.; Leverentz, H. R.; Truhlar, D. G.; Gao, J. Polarized Molecular Orbital Model Chemistry. 2. The PMO Method. *J. Chem. Theory Comput.* **2011**, *7*, 857−867.

(14) Zhang, P.; Fiedler, L.; Leverentz, H. R.; Truhlar, D. G.; Gao, J. Erratum: Polarized Molecular Orbital Chemistry. 2. The PMO Method. *J. Chem. Theory Comput.* **2012**, *8*, 2983.

(15) Isegawa, M.; Fiedler, L.; Leverentz, H. R.; Wang, Y.; Nachimuthu, S.; Gao, J.; Truhlar, D. G. Polarized Molecular Orbital Model Chemistry 3. The PMO Method Extended to Organic Chemistry. *J. Chem. Theory Comput.* **2013**, *9*, 33−45.

(16) Fiedler, L.; Leverentz, H. R.; Nachimuthu, S.; Friedrich, J.; Truhlar, D. G. Nitrogen and Sulfur Compounds in Atmospheric Aerosols: A New Parametrization of Polarized Molecular Orbital Model Chemistry and Its Validation against Converged CCSD(T) Calculations for Large Clusters. *J. Chem. Theory Comput.* **2014**, *10*, 3129−3139.

(17) Dral, P. O.; Wu, X.; Thiel, W. Semiempirical Quantum-Chemical Methods with Orthogonalization and Dispersion Corrections. *J. Chem. Theory Comput.* **2019**, *15*, 1743−1760.

(18) Tuttle, T.; Thiel, W. OMx-D: semiempirical methods with orthogonalization and dispersion corrections. Implementation and biochemical application. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2159−2166.

(19) Wang, Q.; Bryce, R. A. Improved Hydrogen Bonding at the NDDO-Type Semiempirical Quantum Mechanical/Molecular Mechanical Interface. *J. Chem. Theory Comput.* **2009**, *5*, 2206−2211.

(20) Morgado, C. A.; McNamara, J. P.; Hillier, I. H.; Burton, N. A.; Vincent, M. A. Density Functional and Semiempirical Molecular Orbital Methods Including Dispersion Corrections for the Accurate Description of Noncovalent Interactions Involving Sulfur-Containing Molecules. *J. Chem. Theory Comput.* **2007**, *3*, 1656−1664.

(21) Foster, M. E.; Sohlberg, K. Empirically corrected DFT and semi-empirical methods for non-bonding interactions. *Phys. Chem. Chem. Phys.* **2010**, *12*, 307−322.

(22) Stewart, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J. Mol. Model.* **2007**, *13*, 1173−1213.

(23) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.

(24) Stewart, J. J. P. MOPAC2016, 16.307; Steward Computational Chemistry. 2016, web-site, http://OpenMOPAC.net (accessed April 1, 2021).

(25) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427−2438.

(26) Řezáč, J.; Riley, K. E.; Hobza, P. Erratum to "S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures". *J. Chem. Theory Comput.* **2014**, *10*, 1359−1360.

(27) Řezáč, J.; Riley, K. E.; Hobza, P. Extensions of the S66 Data Set: More Accurate Interaction Energies and Angular-Displaced Nonequilibrium Geometries. *J. Chem. Theory Comput.* **2011**, *7*, 3466−3470.

(28) Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comp. Chem.* **2004**, *25*, 1463−1473.

(29) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comp. Chem.* **2006**, *27*, 1787−1799.

(30) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. [Erratum to document cited in CA103(2):11627f]. *J. Am. Chem. Soc.* **1993**, *115*, 5348.

(31) Stewart, J. J. P. Optimization of parameters for semiempirical methods I. Method. *J. Comp. Chem.* **1989**, *10*, 209−220.

(32) Stewart, J. J. P.; Fiedler, L. J.; Zhang, P.; Zheng, J.; Rossi, I.; Hu, W.-P.; Lynch, G. C.; Liu, Y.-P.; Chuang, Y.-Y.; Pu, J.; Li, J.; Cramer, C. J.; Fast, P. L.; Truhlar, D. G. *MOPAC 5.022mn*; Department of Chemistry and Supercomputing Institute, University of Minnesota: Minneapolis, 2015.

(33) Kolb, M.; Thiel, W. Beyond the MNDO model: Methodical considerations and numerical results. *J. Comp. Chem.* **1993**, *14*, 775−789.

(34) Weber, W.; Thiel, W. Orthogonalization corrections for semiempirical methods. *Theor. Chem. Acc.* **2000**, *103*, 495−506.

(35) Dral, P. O.; Wu, X.; Spörkel, L.; Koslowski, A.; Weber, W.; Steiger, R.; Scholten, M.; Thiel, W. Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Theory, Implementation, and Parameters. *J. Chem. Theory Comput.* **2016**, *12*, 1082−1096.

(36) Grimme, S.; Ehrlich, S.; Goerigk, L. Effect of the damping function in dispersion corrected density functional theory. *J. Comp. Chem.* **2011**, *32*, 1456−1465.

(37) Johnson, E. R.; Becke, A. D. A post-Hartree-Fock model of intermolecular interactions. *J. Chem. Phys.* **2005**, *123*, 024101.

(38) Becke, A. D.; Johnson, E. R. A density-functional model of the dispersion interaction. *J. Chem. Phys.* **2005**, *123*, 154101.

(39) Johnson, E. R.; Becke, A. D. A post-Hartree-Fock model of intermolecular interactions: Inclusion of higher-order corrections. *J. Chem. Phys.* **2006**, *124*, 174104.

(40) Axilrod, B. M.; Teller, E. Interaction of the van der Waals Type Between Three Atoms. *J. Chem. Phys.* **1943**, *11*, 299−300.

(41) Tkatchenko, A.; von Lilienfeld, O. A. Popular Kohn-Sham density functionals strongly overestimate many-body interactions in van der Waals systems. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *78*, 045116.

(42) Anatole von Lilienfeld, O.; Tkatchenko, A. Two- and three-body interatomic dispersion energy contributions to binding in molecules and solids. *J. Chem. Phys.* **2010**, *132*, 234109.

(43) Meroueh, O.; Hase, W. L. Dynamics of Energy Transfer in Peptide−Surface Collisions. *J. Am. Chem. Soc.* **2002**, *124*, 1524−1531.

(44) Wang, J.; Hase, W. L. Intermolecular Potential To Represent Collisions of Protonated Peptide Ions with Fluorinated Alkane Surfaces†. *J. Chem. Phys. B* **2005**, *109*, 8320−8324.

(45) Deb, B.; Hu, W.; Song, K.; Hase, W. L. An analytical potential energy function to model protonated peptide soft-landing experiments. The CH3NH3+/CH4 interactions. *Phys. Chem. Chem. Phys.* **2008**, *10*, 4565−4572.

(46) Nogueira, J. J.; Sánchez-Coronilla, A.; Marques, J. M. C.; Hase, W. L.; Martínez-Núñez, E.; Vázquez, S. A. Intermolecular potentials for simulations of collisions of SiNCS+ and (CH3)2SiNCS+ ions with fluorinated self-assembled monolayers. *Chem. Phys.* **2012**, *399*, 193−204.

(47) Pratihar, S.; Kohale, S. C.; Vázquez, S. A.; Hase, W. L. Intermolecular Potential for Binding of Protonated Peptide Ions with Perfluorinated Hydrocarbon Surfaces. *J. Chem. Phys. B* **2014**, *118*, 5577−5588.

(48) Řezáč, J.; Hobza, P. Describing Noncovalent Interactions beyond the Common Approximations: How Accurate Is the "Gold Standard," CCSD(T) at the Complete Basis Set Limit? *J. Chem. Theory Comput.* **2013**, *9*, 2151−2155.

(49) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M. Estimated MP2 and CCSD(T) interaction energies of n-alkane dimers at the basis set limit: Comparison of the methods of Helgaker et al. and Feller. *J. Chem. Phys.* **2006**, *124*, 114304.

(50) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A look at the density functional theory zoo with the advanced GMTKN55 database for general main group thermochemistry, kinetics and noncovalent interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184−32215.

(51) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. Quadratic configuration interaction. A general technique for determining electron correlation energies. *J. Chem. Phys.* **1987**, *87*, 5968−5975.

(52) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796−6806.

(53) Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.

(54) Becke, A. D. Density-functional thermochemistry. I. The effect of the exchange-only gradient correction. *J. Chem. Phys.* **1992**, *96*, 2155−2160.

(55) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785.

(56) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.

(57) Boys, S. F.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553−566.

(58) Simon, S.; Duran, M.; Dannenberg, J. J. How does basis set superposition error change the potential surfaces for hydrogen-bonded dimers? *J. Chem. Phys.* **1996**, *105*, 11024−11031.

(59) Neese, F. The ORCA program system. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 73−78.

(60) Neese, F. Software update: the ORCA program system, version 4.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, No. e1327.

(61) Buckingham, R. A. The classical equation of state of gaseous helium, neon and argon. *Proc. R. Soc. London, Ser. A* **1938**, *168*, 264−283.

(62) Rodríguez-Fernández, R.; Pereira, F. B.; Marques, J. M. C.; Martínez-Núñez, E.; Vázquez, S. A. GAFit: A general-purpose, user-friendly program for fitting potential energy surfaces. *Comput. Phys. Commun.* **2017**, *217*, 89−98.

(63) Martínez-Núñez, E. An automated method to find transition states using chemical dynamics simulations. *J. Comp. Chem.* **2015**, *36*, 222−234.

(64) Martínez-Núñez, E. An automated transition state search using classical trajectories initialized at multiple minima. *Phys. Chem. Chem. Phys.* **2015**, *17*, 14912−14921.

(65) Rodríguez, A.; Rodríguez-Fernández, R.; Vázquez, S. A.; Barnes, G. L.; Stewart, J. J. P.; Martínez-Núñez, E. tsscds2018: A code for automated discovery of chemical reaction mechanisms and solving the kinetics. *J. Comp. Chem.* **2018**, *39*, 1922−1930.

(66) Kopec, S.; Martínez-Núñez, E.; Soto, J.; Peláez, D. vdW-TSSCDS—An automated and global procedure for the computation of stationary points on intermolecular potential energy surfaces. *Int. J. Quantum Chem.* **2019**, *119*, No. e26008.

(67) Dral, P. O. Personal Communication.

(68) Hawkins, D. M. The Problem of Overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1−12.

(69) Řezáč, J.; Jurečkab, P.; Riley, K. E.; Černý, J.; Valdes, H.; Pluháčková, K.; Berka, K.; Řezáč, T.; Pitoňák, M.; Vondrášek, J.; Hobza, P. Quantum Chemical Benchmark Energy and Geometry Database for Molecular Clusters and Complex Molecular Systems (www.begdb.com): A Users Manual and Examples. *Collect. Czech. Chem. Commun.* **2008**, *73*, 1261−1270.

(70) Gonzalez-Lafont, A.; Truong, T. N.; Truhlar, D. G. Direct dynamics calculations with NDDO (neglect of diatomic differential overlap) molecular orbital theory with specific reaction parameters. *J. Chem. Phys.* **1991**, *95*, 4618−4627.