

Article

The PM6-FGC Method: Improved Corrections for Amines and Amides

Martiño Ríos-García , Berta Fernández , Jesús Rodríguez-Otero , Enrique M. Cabaleiro-Lago and Saulo A. Vázquez * 

Departamento de Química Física, Universidade de Santiago de Compostela, 15782 Santiago de Compostela, Spain; martino.rios@rai.usc.es (M.R.-G.); berta.fernandez@usc.es (B.F.); r.otero@usc.es (J.R.-O.); caba.lago@usc.es (E.M.C.-L.)

* Correspondence: saulo.vazquez@usc.es

Abstract: Recently, we reported a new approach to develop pairwise analytical corrections to improve the description of noncovalent interactions, by approximate methods of electronic structures, such as semiempirical quantum mechanical (SQM) methods. In particular, and as a proof of concept, we used the PM6 Hamiltonian and we named the method PM6-FGC, where the FGC acronym, corresponding to Functional Group Corrections, emphasizes the idea that the corrections work for specific functional groups rather than for individual atom pairs. The analytical corrections were derived from fits to B3LYP-D3/def2-TZVP (reference). PM6 interaction energy differences, evaluated for a reduced set of small bimolecular complexes, were chosen as representatives of saturated hydrocarbons, carboxylic, amine and, tentatively, amide functional groups. For the validation, the method was applied to several complexes of well-known databases, as well as to complexes of diglycine and dialanine, assuming the transferability of amine group corrections to amide groups. The PM6-FGC method showed great potential but revealed significant inaccuracies for the description of some interactions involving the $-NH_2$ group in amines and amides, caused by the inadequate selection of the model compound used to represent these functional groups (an NH_3 molecule). In this work, methylamine and acetamide are used as representatives of amine and amide groups, respectively. This new selection leads to significant improvements in the calculation of noncovalent interactions in the validation set.

Keywords: semiempirical methods; PM6 Hamiltonian; PM6-FGC approach; noncovalent interactions; potential energy curves



Citation: Ríos-García, M.; Fernández, B.; Rodríguez-Otero, J.; Cabaleiro-Lago, E.M.; Vázquez, S.A. The PM6-FGC Method: Improved Corrections for Amines and Amides. *Molecules* **2022**, *27*, 1678. <https://doi.org/10.3390/molecules27051678>

Academic Editor: Leonardo Belpassi

Received: 28 January 2022

Accepted: 1 March 2022

Published: 3 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The development of corrections for the description of noncovalent interactions by semiempirical quantum mechanical (SQM) methods has attracted much attention in the last two decades. It is worth mentioning the work carried out by Hobza et al. [1–4], through the development of several generations of corrections for dispersion [1,2,4], hydrogen bond [2,4] and halogen bond [3] interactions, and their parameterization within the PM6 method [5]. Korth [6] and Jensen and co-workers [7] contributed to the third generation of this series.

Truhlar, Gao, and their co-workers, implemented the polarized molecular orbital (PMO) method [8–11], based on a Neglect of Diatomic Differential Overlap (NDDO) Hamiltonian, which includes polarization functions on hydrogen atoms. Additionally, they improved the description of dispersion interactions by adding the first damped dispersion term developed by Grimme [12,13]. The final versions of the PMO method, namely, PMO2 [10] and PMO2a [11], may accurately describe polarization effects, as well as noncovalent complexation energies for some specific systems.

Significant contributions were also made by Thiel et al., who improved the semiempirical Hamiltonian through the orthogonalization-corrected methods OMx [14–16] and

ODMx [17], which led to better results in comparison to NDDO-based methods (e.g., AM1 [18] or PM6 [5]). These semiempirical Hamiltonians needed to incorporate dispersion corrections to improve the description of noncovalent interactions. For the recent ODMx methods, several training sets were considered, including S66 dataset [19,20], and the followed procedure is extensively discussed in the work by Dral et al. [17].

In general, the corrections included in the above studies for dispersion and hydrogen bonding interactions are modelled on potential functions, based on physically sound formulas and parameterized using quite large training sets, which when properly selected, can ensure wider applicability range. The above studies considerably improved the evaluation of noncovalent interactions using SQM methodology, but they still show significant deficiencies in some cases [21]. Often these errors depend on the relative orientation of the interacting molecules and can be attributed to possible shortcomings in popular data sets, which in general only include the most relevant interacting configurations of the molecules.

Very recently, we developed the PM6 Functional Group Corrections (PM6-FGC) approach [21] to correct the semiempirical PM6 Hamiltonian and improve the description of noncovalent interactions. Although in general, the subsequent PM7 method [22] provides higher accuracy than its predecessor, we used the PM6 Hamiltonian, following Grimme and co-workers' recommendation [23]. Specifically, they reported significant errors in interaction energies calculated with PM7 for large systems, due to the damping and truncation of the dispersion energy at long range in this Hamiltonian [24,25]. Our previous work [21], presented as a proof-of-concept study, provides an alternative way to obtain analytical corrections for SQM methods. The improvement of the description of noncovalent interactions in SQM methods is particularly important in quantum mechanics/molecular mechanics (QM/MM) simulations, since the efficiency of the quantum mechanical method is crucial for practical applications. Furthermore, the use of a semiempirical Hamiltonian may be convenient for investigating chemical reactions because the semiempirical parameters may be specifically adjusted for individual reactions [26].

In our previous work [21], the corrections are obtained by fits involving intermolecular potential energy curves (IPECs) of bimolecular complexes of small molecules that contain functional groups of relevance for biological chemistry. Specifically, we considered methane, formic acid, and ammonia as representatives of saturated hydrocarbons, carboxylic acids, and $-NH_2$ groups in amines and, tentatively, in amides. We evaluated IPECs for the six different pair combinations of these molecules, considering orientations of the interacting molecules that emphasize the atom-pair interactions in the complexes.

The reference data for the fittings were obtained by calculations using Density Functional Theory (DFT), with the B3LYP functional [27–29], including the D3 dispersion correction with the Becke–Johnson damping scheme [30–32], and the valence triple-zeta polarization def2-TZVP basis set [33]. This combination of density functional and basis set was selected after benchmarking the results with those obtained at the CCSD(T) level [34], together with the augmented correlation consistent polarized valence triple-zeta basis set aug-cc-pVTZ [35]. The differences between the DFT and the coupled cluster IPECs were negligible for our purposes and showed the efficiency of the combination of the B3LYP-D3 density functional and the def2-TZVP basis set, in providing acceptable and inexpensive interaction energies. Since we plan to extend the study to larger molecules and other types of functional groups, the use of an efficient and inexpensive reference methodology is important and, therefore, we have selected the B3LYP-D3/def2-TZVP level as the reference.

Our analytical pairwise corrections are parameterized by fits to a series of curves obtained by the difference between the IPECs calculated at the B3LYP-D3/def2-TZVP level and those computed with the PM6 method, following a strategy similar to that employed in the derivation of intermolecular potentials for dynamics studies [36–40]. Our previous work [21] emphasized the importance of including all the significant orientations of the interacting molecules in the databases, a fact that is crucial in the development of well-balanced corrections. Furthermore, the use of general corrections is important to take into account that SQM methods have significant limitations, not only to accurately describe

dispersion interactions, but also other contributions, such as electrostatics, induction, and exchange repulsion.

The IPECs obtained with the PM6-FGC method for the bimolecular complexes of methane, formic acid, and ammonia acceptably reproduced those determined at the DFT level [21], and considerably improved the results provided by previously corrected SQM methods (particularly, PMO2 and PM6-D3H4). To assess the performance of the method, as well as the transferability of the corrections to other systems, we carried out the PM6-FGC evaluation of interaction energies for several complexes of the S66 [19], A24 [41], and ADIM6 [42,43] 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). Although the performance of the PM6-FGC approach was acceptable in general, we found significant inaccuracies in the description of some interactions involving the $-NH_2$ group in amines and amides (see [21] for extra details). The aim of the present study is to improve the method by handling and solving these shortcomings.

The manuscript is organized as follows: in Section 2, the results are presented and discussed, in Section 3, we describe the method and give the computational details, and in the last Section, we summarize and conclude.

2. Results and Discussion

In this work, for the parameterization of our corrections, we have selected the following model compounds: acetamide, methylamine, methane, formic acid and acetic acid. We considered the different atom types defined in Figure 1a. For parameterizations involving the carboxylic functional group, we used either formic acid, as in our previous work [21], or acetic acid. The latter compound was selected to explore different alternatives in our parameterization strategy, as explained later. As discussed in ref. [21], our main purpose is to develop corrections to improve the description of noncovalent interactions in peptides and other biological systems. As for the development of molecular mechanics force fields, we assume that the corrections developed from fits involving small model compounds may be transferable to larger systems containing the corresponding functional groups. Figure 1b depicts a conformation of dialanine showing the atom types defined in this study.

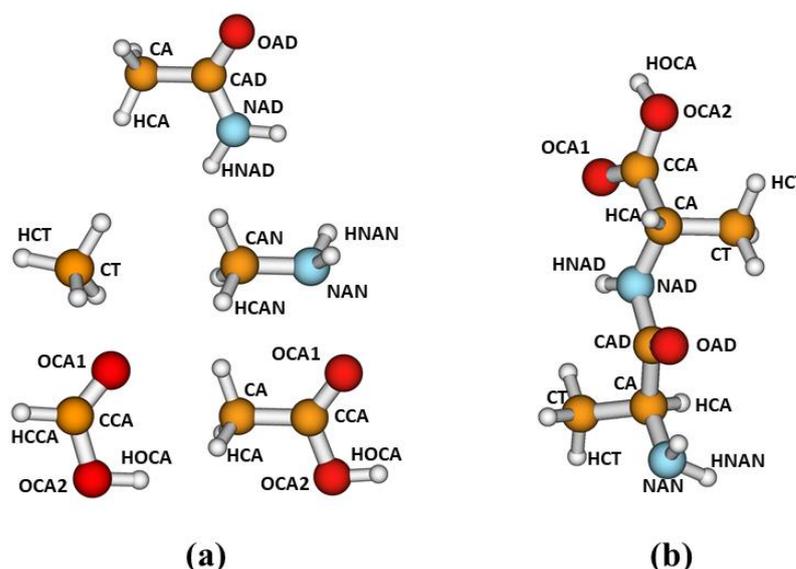


Figure 1. (a) Model compounds and atom types considered for the parameterizations, and (b) atom-type transferability to dialanine.

We have performed fits for the following complexes: $CH_3NH_2-CH_3NH_2$, $CH_3CONH_2-CH_3CONH_2$, $CH_3NH_2-CH_3CONH_2$, $CH_3NH_2-CH_4$, $CH_3NH_2-HCOOH$, CH_3CONH_2-

CH_4 , and $\text{CH}_3\text{CONH}_2\text{--CH}_3\text{COOH}$. For each complex, we carried out a series of independent fits, adjusting all the parameters simultaneously. The final parameters, resulting from our best fits, are given in the Supplementary Materials, in the form of a Python dictionary. Notice that, for each complex, the total number of parameters involved in the fits is five times the number of different types of pairwise interactions.

2.1. Methylamine Dimer

The methylamine molecule has four different types of atoms; therefore, the dimer exhibits ten different types of pairwise interactions. For the fits, we considered ten orientations that emphasize the different types of two-body interactions, as well as a cyclic configuration that facilitates the simultaneous interaction between most of the atoms (orientation 11). Four of these orientations are depicted in Figure 2, together with the corresponding intermolecular potential energy curves. The remaining five orientations and the associated IPECs are shown in Figure S1, in the Supplementary Materials. In these and the following figures, the B3LYP-D3/def2-TZVP data are displayed as open, black circles, and r refers to the distance between the selected pair of attacking atoms. The IPECs calculated by the PM6 method are shown as red solid lines. For comparison, we include the results obtained with the PM6-D3H4 method (blue lines), which is one of the most successfully applied SQM methods. For this system, 50 independent parameters were fitted simultaneously. The parameters of the best fit are collected in the Supplementary Materials, in the form of a Python dictionary. Adding the corrections to the PM6 Hamiltonian results in the PM6-FGC method, the corresponding IPECs are shown as green lines in the figures.

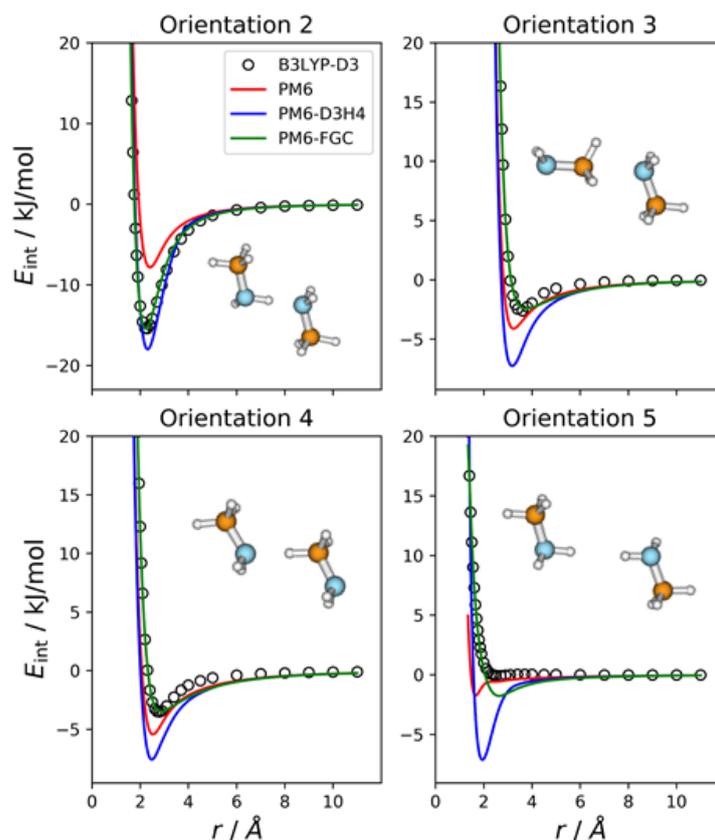


Figure 2. Comparison of IPECs for four selected orientations of the CH_3NH_2 dimer.

In orientation 2, an amine hydrogen of one of the molecules interacts with the nitrogen of the other molecule in a way that favors the formation of a hydrogen bond. As expected, this orientation leads to the deepest well among those considered in the present study. The B3LYP-D3 interaction energy is -15.4 kJ/mol at 2.3 Å (distance between the attacking

atoms, i.e., H and N). Both corrected methods, that is, PM6-D3H4 and PM6-FGC, reproduce the DFT IPEC quite well. The PM6 method underestimates the strength of the hydrogen bond, which is one of the well-known deficiencies of this method. The PM6-D3H4 method clearly overestimates the strength of the interaction in the well region for orientations 3, 4, and 5. Hobza and co-workers used the S66 database [19,20] for the parameterization of the D3H4 corrections. In general, the interaction energies included in this database were obtained by CCSD(T) calculations, with complete basis set extrapolation and, therefore, this should be borne in mind when the PM6-D3H4 IPECs are compared with our DFT data. However, as previously shown [21], for the type of model compounds considered in our work, the results of B3LYP-D3/def2-TZVP calculations are, in general, in very good agreement with those determined at the CCSD(T)/CBS level (see also Section 2.8).

The case of orientation 5 is particularly interesting. This orientation stresses the interaction between amine hydrogens, which is repulsive at the B3LYP-D3 level. The PM6-D3H4 method predicts a significant well depth, similar to that found for the ammonia dimer [21]. Our method also gives a minimum, but the well depth is rather small (1.8 kJ/mol). Regarding the methylamine dimer, this is perhaps the orientation for which our method shows the most significant deviation from the reference data.

In general, for the 11 orientations considered here, the performance of our method is quite good and the improvement over the PM6 and PM6-D3H4 is clear. This may be quantified by the values of the mean absolute errors (MAEs), which are collected in Table S1 in the Supplementary Materials. For interaction energies below 5 kJ/mol, the calculated MAEs are 1.65 kJ/mol (PM6), 1.32 kJ/mol (PM6-D3H4), and 0.34 kJ/mol (PM6-FGC). Increasing the maximum interaction energy up to 20 kJ/mol, we obtained 2.69 kJ/mol (PM6), 2.27 kJ/mol (PM6-D3H4), and 0.38 kJ/mol (PM6-FGC). The significant increase in the PM6 and PM6-D3H4 MAEs, obtained by increasing the range of the interaction energies, indicates that the accuracy of these methods for the repulsive region of the intermolecular potential is very low, as could be anticipated by close inspection of Figure 2 and Figure S1. The performance of the PM6-FGC is quite good, even up to interaction energies around 100 kJ/mol; at this energy, the calculated MAEs are 4.79 kJ/mol (PM6), 5.15 kJ/mol (PM6-D3H4), and 0.41 kJ/mol (PM6-FGC). For higher repulsive energies, the PM6-FGC may fail dramatically with the present corrections, although the development of specific corrections for high interaction energies is straightforward. For most practical applications, however, the attractive region of the potential energy is the important one.

2.2. Acetamide Dimer

For acetamide, we defined five different atom types, as shown in Figure 1. Consequently, the dimer has 21 different types of pairwise interactions. For the fits, we considered 22 orientations. The additional one (i.e., orientation 22) corresponds to the most stable conformation of the dimer. For this orientation, we performed a full optimization at the B3LYP-D3/def2-TZVP level. Then, we calculated the associated IPEC by varying one of the O...H distances along the direction defined by these two atoms. In Figure 3, we display eight selected orientations with the corresponding IPECs. The IPECs of the remaining orientations are shown in Figure S2 of the Supplementary Materials. For the most stable configuration of the dimer, both corrected methods reproduce the reference IPEC very well. The interaction energy at the minimum is about -70 kJ/mol. The underestimation of the hydrogen bonding interaction by the PM6 method is apparent in the figure.

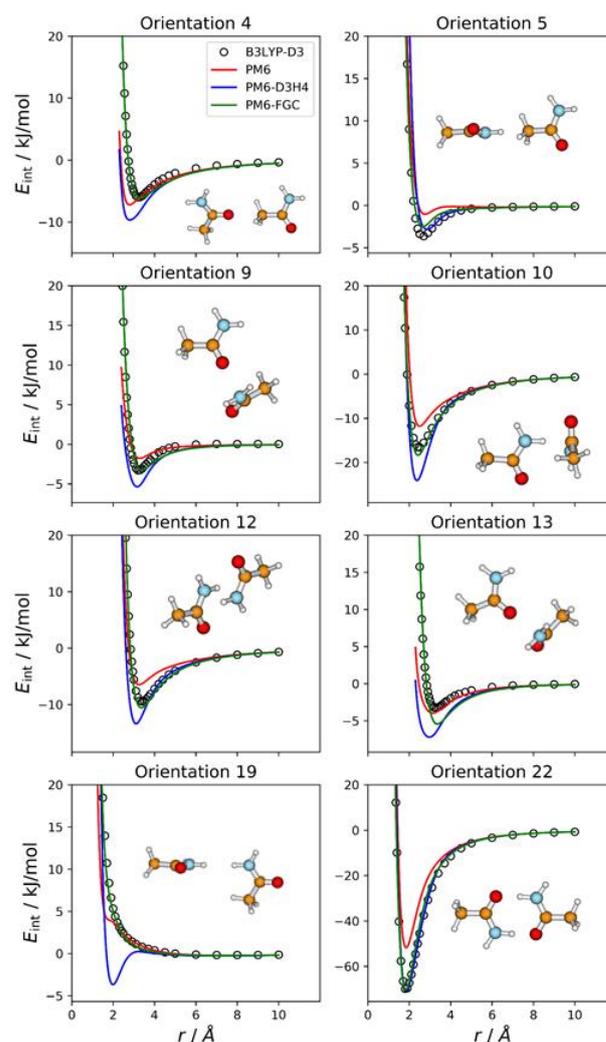


Figure 3. Comparison of IPECs for eight selected orientations of the CH_3CONH_2 dimer.

For some orientations (e.g., 4, 10, or 13), the PM6-D3H4 method overestimates the well depth of the minima. In addition, and resembling the results obtained for the methylamine dimer, the PM6-D3H4 method predicts a significant minimum for the orientation in which amide hydrogens face each other (orientation 19; cf. orientation 5 in Figure 2). The MAEs calculated for this complex for interaction energies below 5 kJ/mol are 2.51 kJ/mol (PM6), 1.55 kJ/mol (PM6-D3H4), and 0.54 kJ/mol (PM6-FGC). Again, the MAEs increase as we increase the upper limit of the interaction energy (see Table S1). For the PM6-FGC method, the increase in the MAE is small; specifically, for interaction energies below 40 kJ/mol, the calculated MAE is 0.76 kJ/mol. In general, the performance of the PM6-FGC is quite good. However, for orientation 13, in which the oxygen of one of the molecules faces the nitrogen atom of the other molecule, our method significantly overestimates the interaction in the attractive region.

2.3. Methylamine—Acetamide Complex

The number of orientations considered for the $\text{CH}_3\text{NH}_2\text{—CH}_3\text{CONH}_2$ complex is 24, which is equal to the number of different types of pairwise interactions in this system. We show in Figure 4 the IPECs of eight orientations; the remaining IPECs are depicted in Figures S3 and S4 of the Supplementary Materials. Among the orientations considered for this complex, orientation 13 exhibits the strongest attractive interaction, with a well depth of 28.9 kJ/mol at the DFT level, which is consistent with formation of a hydrogen bond between an amide hydrogen and the amine nitrogen of the partner molecule. For this

orientation, both corrected methods reproduce the reference IPEC quite well. By contrast, and as expected, the PM6 method significantly underestimates the strength of the hydrogen bonding interaction.

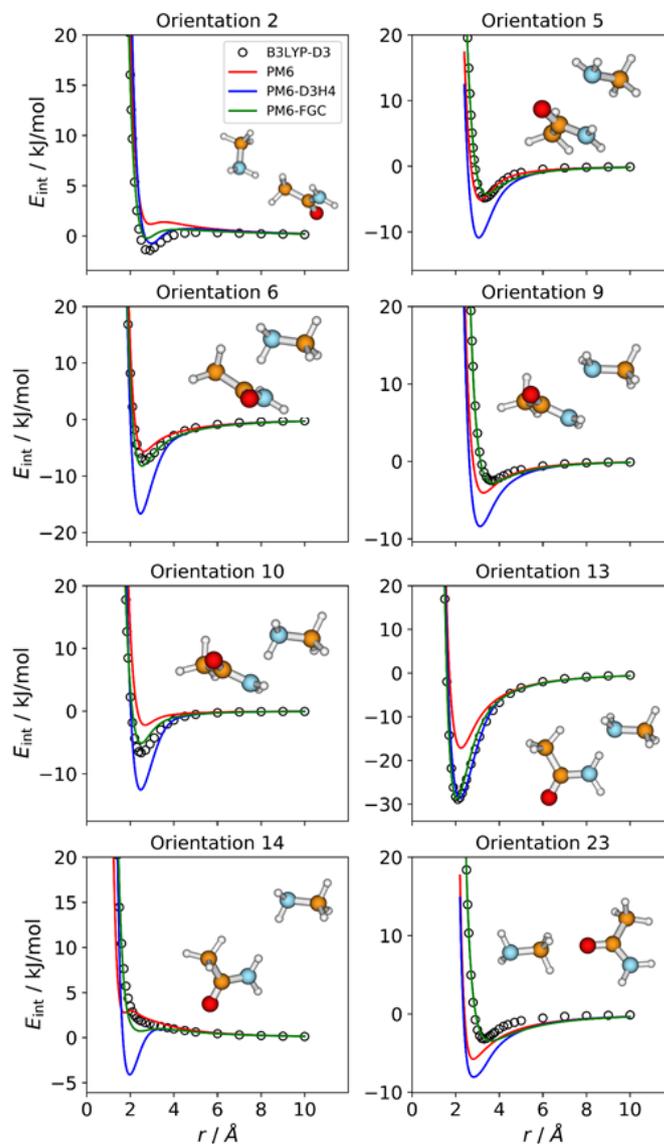


Figure 4. Comparison of IPECs for eight selected orientations of the $\text{CH}_3\text{NH}_2\text{-CH}_3\text{CONH}_2$ complex.

For several orientations, the PM6 and PM6-D3H4 IPECs deviate markedly from the reference curves. Regarding the PM6-D3H4 method, most of the deviations involve overestimation of minimum well depths (see Figure 4), which is the opposite of the trend shown by the PM6 method (see Figures S3 and S4). For orientation 14, in which an amine hydrogen faces an amide hydrogen, the PM6-D3H4 method predicts a significant minimum, which contrasts with the repulsive nature of this orientation, as predicted by the B3LYP-D3/def2-TZVP calculations. This result resembles those found for orientations 5 and 19 in the methylamine and acetamide dimers, respectively. The most significant deviations observed for our method correspond to orientations 2, 23 (Figure 4), 8 (Figure S3), and 19 (Figure S4), but they are rather small. These results can be quantified by the MAEs (Table S1). For interaction energies below 5 kJ/mol, the calculated MAEs are 1.80 kJ/mol (PM6), 1.71 kJ/mol (PM6-D3H4), and 0.52 kJ/mol (PM6-FGC).

2.4. Methylamine—Methane Complex

Since methylamine and methane have four and two different atom types, respectively, the number of different types of pairwise interactions in the $\text{CH}_3\text{NH}_2\text{--CH}_4$ complex is eight. This is the number of orientations we considered for this complex. Figure 5 shows the IPECs for four selected orientations. The IPECs for the remaining orientations are depicted in Figure S5 of the Supplementary Materials. As can be seen, although the PM6-D3H4 curves for orientations 1 and 5 show small disagreement with the reference IPECs, both corrected methods describe the intermolecular interaction satisfactorily. The PM6 curves deviate significantly from the reference DFT data, especially in orientations 4 and 8, in which hydrogen atoms face each other.

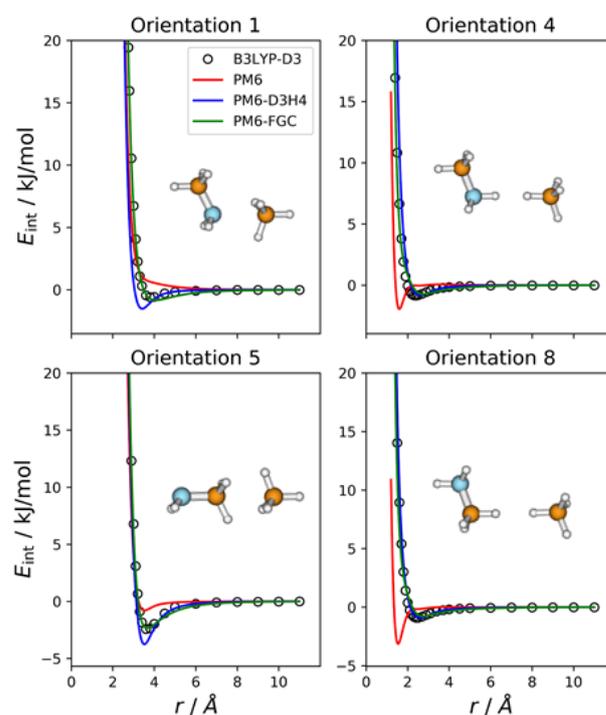


Figure 5. Comparison of IPECs for four selected orientations of the $\text{CH}_3\text{NH}_2\text{--CH}_4$ complex.

The MAEs calculated for interaction energies below 5 kJ/mol are 0.63 kJ/mol (PM6), 0.47 kJ/mol (PM6-D3H4), and 0.18 kJ/mol (PM6-FGC). If we increase the energy range for the calculation of MAEs, the values for the PM6 and PM6-D3H4 methods increase significantly (see Table S1), thus, following the same trend as that found for the previous complexes.

2.5. Methylamine—Formic Acid Complex

For this complex, we considered 20 different orientations, which is the number of different types of pairwise interactions exhibited in this system. In Figure 6, we display the IPECs for six orientations of this complex. The remaining 14 orientations are depicted in Figure S6 of the Supplementary Materials. Orientation 4 shows the strongest attractive interaction, as expected, since it leads to hydrogen bonding between the carboxylic hydrogen and the amine nitrogen. The well depth calculated at the B3LYP-D3/def2-TZVP level is 43.5 kJ/mol, with the minimum being at an $\text{H}\cdots\text{N}$ distance of 1.8 Å. In general, the performance of the PM6-FGC is quite good; the most significant deviations from the reference curves, for example, those occurring for orientations 6 and 19, are rather small.

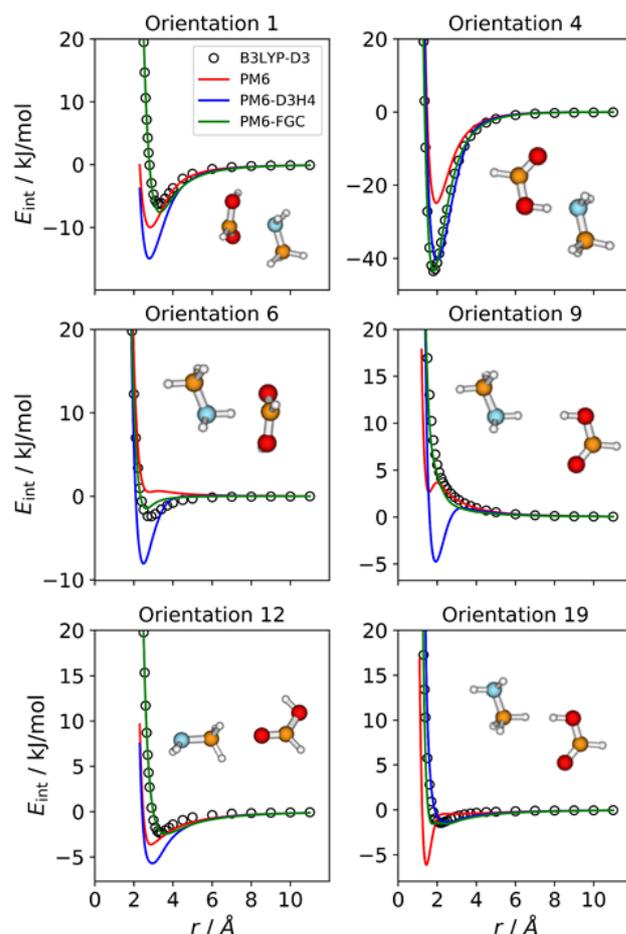


Figure 6. Comparison of IPECs for six selected orientations of the $\text{CH}_3\text{NH}_2\text{-HCOOH}$ complex.

For several orientations, the PM6-D3H4 curves show remarkable discrepancies with the reference IPECs. This is the case for orientations 1 and 6, in which the carboxylic carbon faces the nitrogen atom and the amine hydrogen, respectively. Resembling the results found in previous complexes, wherein hydrogen bonding can take place, the most striking result appears for orientation 9, in which the carboxylic hydrogen faces one of the amine hydrogens. This orientation is repulsive at the B3LYP-D3/def2-TZVP level, but the PM6-D3H4 method predicts a clear minimum (4.8 kJ/mol at an $\text{H}\cdots\text{H}$ distance of 1.9 Å). We found the same behaviour for the $\text{NH}_3\text{-HCOOH}$ complex in our previous work [21]. As for the complexes discussed earlier in this paper, the improvement of our approach over the PM6 and PM6-D3H4 methods is clear. Considering all the orientations of this complex, the MAEs calculated for interaction energies below 5 kJ/mol are 1.76 kJ/mol (PM6), 1.88 kJ/mol (PM6-D3H4), and 0.38 kJ/mol (PM6-FGC). For this system, the PM6 value is slightly lower than that calculated for the PM6-D3H4 method, and the same occurs for the other upper limits of the interaction energies (Table S1).

2.6. Acetamide—Methane Complex

This complex shows 12 different types of pairwise interactions. Accordingly, we considered 12 orientations, selected in such a way that they emphasize the different pair interactions. The IPECs for six of these orientations are exhibited in Figure 7; the IPECs of the remaining orientations are shown in Figure S7 of the Supplementary Materials. As for the $\text{CH}_3\text{NH}_2\text{-CH}_4$ complex, both corrected methods afford results in good or reasonably good agreement with the reference IPECs. The improvement of PM6-FGC over the PM6-D3H4 method is reflected in the values of the calculated MAEs (Table S1). Considering all the

orientations and interaction energies below 5 kJ/mol, the MAE of our method (0.15 kJ/mol) is four times smaller than that of the PM6-D3H4 method (0.61 kJ/mol).

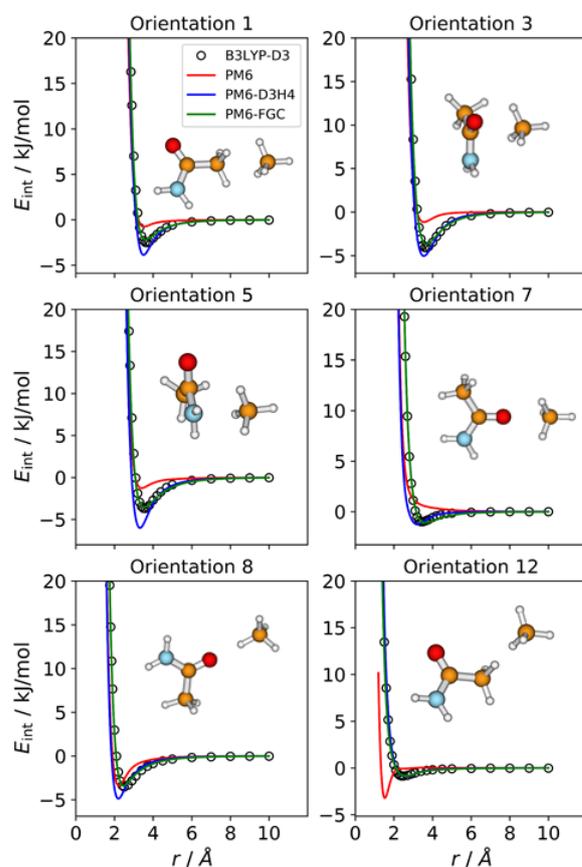


Figure 7. Comparison of IPECs for six selected orientations of the $\text{CH}_3\text{CONH}_2/\text{CH}_4$ complex.

2.7. Acetamide—Acetic Acid Complex

For this complex, we explored three different parameterization strategies. In one of them, the methyl group of the acetic acid molecule was considered to be equivalent to that of acetamide (see Figure 1a), and the parameters for the CA–CA, CA–HCA, and HCA–HCA interactions (15 in all) were taken from the parameterization of the acetamide dimer. With this definition of atom types, there are 35 different types of pairwise interactions and a total of 175 parameters, so that the fittings involved the simultaneous adjustment of 160 parameters. In the second one, in addition to the methyl group, the carbon and oxygen atoms of the carbonyl group of the acetic acid molecule were considered to be equivalent to the corresponding atoms in acetamide. This leads to 30 different types of pairwise interactions and a total of 150 parameters. Half of the distinct types of pairwise interactions are exhibited in the acetamide dimer, and the corresponding 75 parameters were taken from that system. Consequently, only 75 parameters were fitted in this case. Finally, in the third strategy, all the atom types in acetic acid were different from those of acetamide, which resulted in 36 distinct types of pairwise interactions and 180 parameters in all. Preliminary fits showed that the first and the third strategies led to similar values of the objective function (see the section on Methods and Computational Details), whereas the second strategy resulted in larger values. This may suggest that the carbonyl group in the acetic acid molecule behaves somewhat different from that in acetamide. For this reason, we chose the first strategy.

Even though for this system we followed the first parameterization strategy described in the previous paragraph, we considered 37 different orientations, 36 of which emphasize the distinct types of pairwise interactions in the complex, assuming no transferability of atom types between the two molecules (i.e., the assumption followed in the third

parameterization strategy). The remaining orientation corresponds to the most stable configuration of the complex, which facilitates the formation of a double hydrogen bond (orientation 37). Figure 8 displays the IPECs of eight selected orientations; the results for the remaining orientations are shown in Figures S8 and S9 in the Supplementary Materials. For many orientations, the failure of the PM6 and PM6-D3H4 methods is apparent. As for the previous complexes in which electrostatic interactions are important, for many orientations, the PM6 method underestimates the strength of attractive interactions in the well region (see Figures 8, S8 and S9).

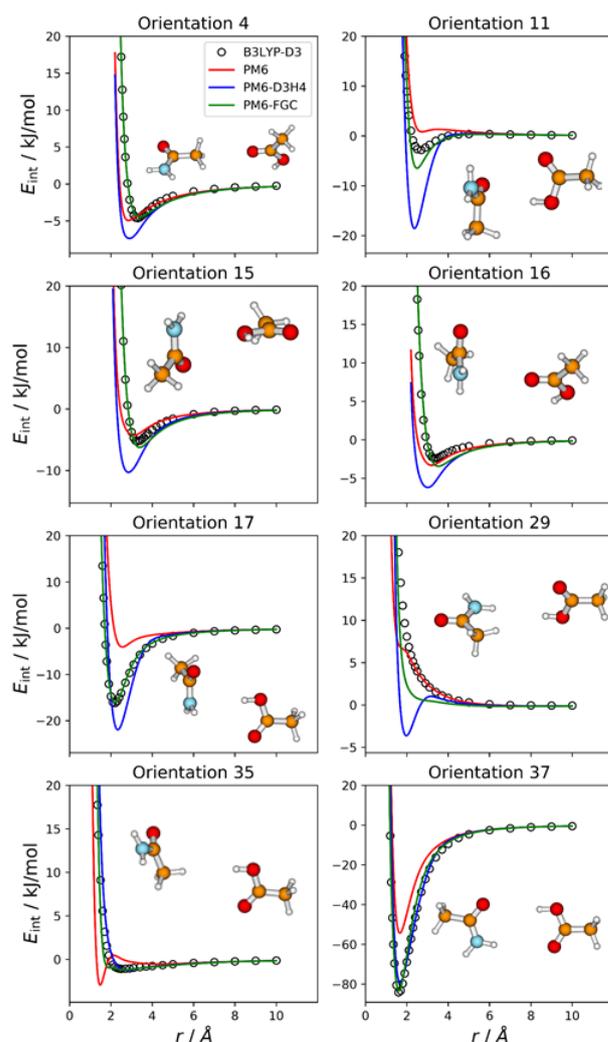


Figure 8. Comparison of IPECs for eight selected orientations of the $\text{CH}_3\text{CONH}_2/\text{CH}_3\text{COOH}$ complex.

The PM6-D3H4 method significantly overestimates the well depth for several orientations. This is the case for orientation 11, in which the carboxyl hydrogen attacks the carbon atom of the amide group. For this orientation, the PM6-D3H4 predicts a well depth 18.5 kJ/mol, whereas the DFT value is 2.8 kJ/mol. The PM6-FGC curve, with a well depth of 6.5 kJ/mol, also deviates significantly from the reference IPEC, although not as much as the PM6-D3H4 curve. Another significant failure of our method occurs for orientation 29, in which the carboxyl hydrogen faces one of the amide hydrogens. Although the PM6-FGC curve is repulsive, it is less repulsive than that predicted by the B3LYP-D3/def2-TZVP calculations. For this orientation, the PM6-D3H4 method shows a clear minimum, thus, resembling the behaviour found for other complexes (e.g., orientation 9 in the $\text{CH}_3\text{NH}_2\text{-HCOOH}$ complex, Figure 6). For orientation 35, wherein the carboxyl hydrogen faces a methyl hydrogen of the partner molecule, the PM6-FGC curve around an $\text{H}\cdots\text{H}$ distance of 2 Å deviates somewhat from the reference curve. This deviation results from the strange

behaviour of the PM6 Hamiltonian, whose curve shows a maximum at an H...H distance of 2.1 Å and a minimum at 1.5 Å. Nevertheless, in general, the performance of our method is reasonably good, and the improvement over the other SQM approaches is clear. The calculated MAEs, for energies below 5 kJ/mol, are 2.71, 1.72, and 0.51 kJ/mol for PM6, PM6-D3H4, and PM6-FGC, respectively (see Table S1 of the Supplementary Materials).

2.8. Validation of the PM6-FGC Method

One of the issues that deserves special attention in model fitting is the possibility of overfitting [44]. As discussed in our previous work [21], our tests suggested that our parameterization model does not show overfitting, at least at a significant level. Since this issue was analysed in detail in our previous work, here, we will only focus on the validation of the new corrections, by applying them to several complexes of the S66 database [19] that involve amine and amide groups, for which our previous corrections were found to be inaccurate, as well as to complexes of diglycine and dialanine, for which our method showed substantial improvement over the PM6 and the PM6-D3H4 methods.

Table 1 displays the interaction energies calculated for several complexes included in the S66 database, which contain amine and amide groups. The structures of these complexes are depicted in Figure S10 in the Supplementary Materials. Notice that, in the S66 database, “peptide” refers to *N*-methylacetamide. In addition to the CCSD(T)/CBS benchmark values reported in ref. [19], we include the interaction energies calculated at the B3LYP-D3/def2-TZVP level, together with the values obtained with the PM6, PM6-D3H4, and PM6-FGC methods. For the PM6-FGC calculations, the atom types considered for the carbon and hydrogen atoms of the methyl group, attached to the nitrogen atom of *qu*-methylacetamide, were “CAN” and “HCAN”, respectively (see Figure 1). For our method, we also show the values obtained in our previous work. The first important observation is the very good agreement between the CCSD(T)/CBS data and the B3LYP-D3/def2-TZVP values, as previously mentioned. The corresponding MAE is below 1 kJ/mol. A further remarkable finding is the small MAE calculated for the PM6-D3H4 data. This is somewhat expected, since these structures were included in the training set used to parameterize this method. An analysis of the PM6-FGC results reveals a clear, general improvement of the corrections. The MAE calculated with the old corrections is more than twice the value obtained with the present corrections. Notice that in our previous work, we used NH₃ as the representative for both amine and amide groups, which resulted in a drastic approximation for amides. Particularly, this led to a remarkable error for the global minimum of the acetamide dimer. It is important to mention that, in this work, we have included the most stable configuration of the acetamide dimer in the training set, but not the orientation of the methylamine dimer considered in the S66 database (complex 10, see Figure S10). Finally, the interaction energies calculated for complexes 46 and 62 with our previous corrections were found to be in better agreement with the benchmark data than those evaluated with the present corrections. This is probably a result of error compensation.

Table 1. Interaction energies (in kJ/mol) for selected complexes of the S66 database.

Complex ^a	CCSD(T)/CBS ^b	B3LYP-D3	PM6	PM6-D3H4	PM6-FGC ^c	PM6-FGC ^d
(10) CH ₃ NH ₂ –CH ₃ NH ₂	−17.41	−18.04	−7.70	−19.00	−13.10	−14.10
(11) CH ₃ NH ₂ –peptide	−22.68	−22.71	−16.11	−25.69	−18.79	−19.02
(14) peptide–CH ₃ NH ₂	−31.17	−32.98	−17.49	−31.42	−27.61	−32.32
(15) peptide–peptide	−36.11	−36.74	−24.73	−36.82	−30.12	−37.92
(21) CH ₃ CONH ₂ –CH ₃ CONH ₂	−68.03	−70.09	−51.80	−70.71	−49.58	−70.95
(46) peptide–pentane	−17.82	−17.21	−5.27	−17.20	−17.57	−15.06
(62) pentane–CH ₃ CONH ₂	−14.77	−14.46	−6.44	−16.48	−14.81	−13.47
MAE		0.87	11.21	1.51	5.21	2.42

^a Identification numbers used in the S66 database are given in parentheses. ^b Ref. [19]. ^c Previous work [21].

^d This work.

In the development of analytical corrections for biological compounds, we first focus on peptides. For this reason, for the validation of the PM6-FGC corrections, in this and the previous study, we have selected complexes of small peptides, specifically, the dimers of diglycine and dialanine, as well as the trimer of diglycine. The conformational space of these complexes was explored in our previous work, using AutoMeKin [45–49], software designed to predict and simulate the kinetics of complex reaction mechanisms. This program contains a module that allows one to locate minima for intermolecular complexes in an automated fashion. The search involves both intramolecular and intermolecular degrees of freedom, and for our study, we used the PM6-D3H4 method, available through the interface between AutoMeKin and the MOPAC2016 program [50].

As described in our previous study [21], for the dialanine dimer we obtained 90 different conformers. For these conformers, the interaction energies were evaluated by single-point B3LYP-D3/def2-TZVP calculations, using counterpoise corrections, and the data were used to compute linear correlations with the corresponding SQM values. The results are shown graphically in Figure 9. The black straight lines correspond to the case of perfect correlation. In addition, in seen in Table 2, we collected the calculated MAEs, as well as the mean bias errors (MBEs). The MBE is just the mean value of the differences between the reference and the model prediction (i.e., the SQM results). A positive MBE value indicates that the SQM method overestimates the strength of the interaction whereas a negative value indicates underestimation. As can be seen from the figure and the statistical values given in Table 2, our approach provides an improvement in the accuracy of SQM methods for the calculation of noncovalent interactions in this system. The MAE and MBE values calculated for PM6 and PM6-D3H4 are much larger than those computed for our method, using either the present or the previous corrections. The MAE calculated for the data obtained with the present corrections (4.9 kJ/mol) is significantly smaller than that determined in our previous study (8.7 kJ/mol). To a large extent, this is a consequence of the very small bias displayed by the interaction energies calculated with the present corrections.

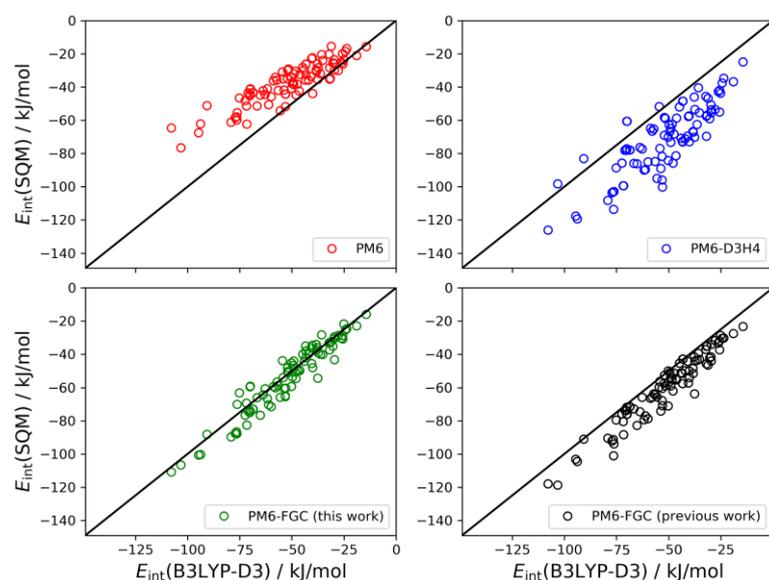


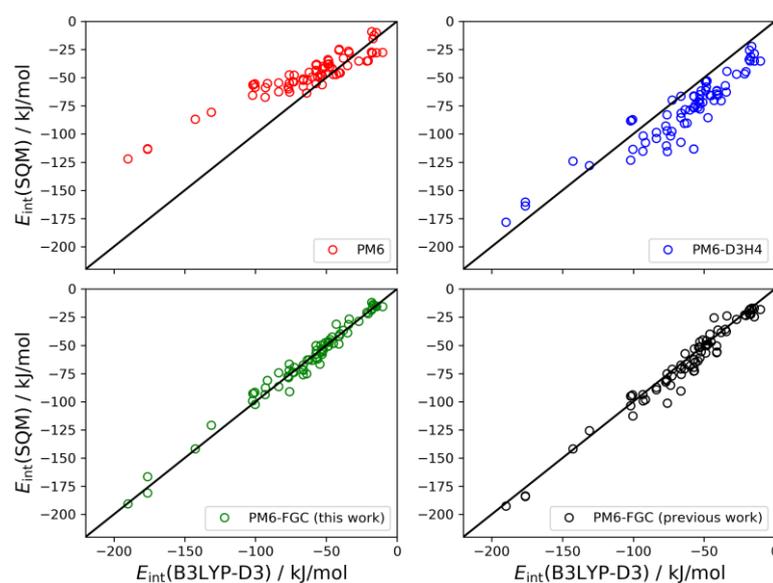
Figure 9. Linear correlations obtained for the dialanine dimer. For comparison, we include the PM6-FGC results evaluated with our previous corrections [21].

Table 2. Statistical parameters ^a of the linear correlations using B3LYP-D3/def2-TZVP interaction energies as the reference.

	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	4.0 (6.0)	−0.6 (3.2)	4.9 (8.7)	1.9 (8.7)	5.9 (9.1)	4.6 (8.7)

^a MAE and MBE values are given in kJ/mol. Values in parentheses correspond to the parameterization performed in our previous work [21].

The set of conformers of the diglycine dimer contain 77 different structures [21]. The linear correlations are depicted in Figure 10. The results for this system follow the same trends observed for the dialanine dimer. The PM6 method underestimates the strength of the noncovalent interactions (MBE = −15.0 kJ/mol) and, on the contrary, the PM6-D3H4 method substantially overestimates the interaction strength (MBE = 15.5 kJ/mol). Both methods give MAEs around 18 kJ/mol. Our method improves the statistics remarkably. With the present corrections, the calculated MAE is 4.0 kJ/mol, which is 33% smaller than our previous value. Furthermore, the MBE value is almost negligible (−0.6 kJ/mol), thus, pointing out a significant improvement over the previous results.

**Figure 10.** Linear correlations obtained for the diglycine dimer. For comparison, we include the PM6-FGC results evaluated with our previous corrections [21].

The last system considered here for the validation of our method is the trimer of diglycine. In this case, the interaction energy is calculated as the difference between the energy of the non-interacting monomer and dimer and the energy of the trimer. In our previous study, we found 146 different conformers for the trimer [21]. The linear correlations are depicted in Figure 11. For this system, the results obtained with the present corrections show a small bias (MBE = 4.6 kJ/mol) towards higher strength of the noncovalent interaction. However, the improvement over the previous parameterization is clear; the MAE and MBE values are almost half those determined in our previous work. In the analysis of these correlations, it should be kept in mind that the PM6-D3H4 method was parameterized using CCSD(T)/CBS data in general. Nonetheless, for these types of peptides, the higher accuracy of the PM6-FGC method is apparent.

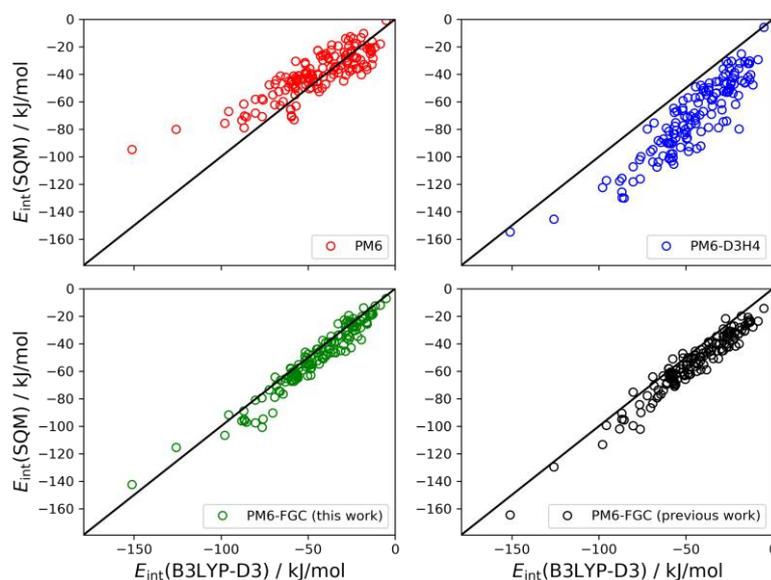


Figure 11. Linear correlations obtained for the diglycine trimer. For comparison, we include the PM6-FGC results evaluated with our previous corrections [21].

3. Methods and Computational Details

We followed the procedure described in our previous work [21], where we presented the PM6-FGC method, whose aim was to show a strategy to develop corrections able to satisfactorily model noncovalent interactions for all orientations of interacting molecules. In that work [21], to parameterize corrections involving N atoms, we selected the NH_3 molecule as representative. We chose three molecules in all, specifically, methane, formic acid, and ammonia, which resulted in six bimolecular complexes needed to get the parameters. The corrections obtained from NH_3/CH_4 fittings could model $-\text{NH}_2$ /alkane interactions, but the use of ammonia as representative proved not to be adequate enough to reproduce other interactions of the $-\text{NH}_2$ group in amines and amides (see for instance the results shown in Table 1 for some complexes of methylamine, *N*-methylacetamide and acetamide). Therefore, to try to solve these limitations, in the present investigation, we decided to use methylamine and acetamide as representatives, instead of ammonia. In addition, and with the purpose of exploring parameter transferability, we also considered the acetic acid molecule.

Since in ref. [21], we selected B3LYP-D3/def2-TZVP intermolecular potential energy curves for the training set, due to their good agreement with the corresponding CCSD(T)/aug-cc-pVTZ results, here, we used the same level of electronic structure for the evaluation of interaction energies in the new bimolecular complexes. As before, we employed the supermolecular approach and corrected for basis set superposition error (BSSE) with the counterpoise method [51,52]. The intramolecular geometries were obtained from B3LYP-D3/def2-TZVP optimizations and kept frozen in the interaction energy calculations. As shown in our previous work [21], the number and types of orientations play a crucial role in developing adequate well-balanced corrections. In this way, for each pair of molecules, the number of orientations should be at least equal to the number of the different pair-type interactions. The electronic structure calculations were performed with the ORCA 4.0 and 5.0 programs using the frozen core approximation [53,54].

The general expression for the noncovalent potential energy correction to the PM6 Hamiltonian is given as a pairwise sum of the form:

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

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 . The parameters A_{ij} , B_{ij} , C_{ij} and D_{ij} depend on the nature of the considered pair of atoms. The parameters A_{ij} and C_{ij} may be either positive or negative. In our previous work [21], we constrained the D_{ij} parameters to be integers. In this study, to get additional, although small, flexibility, these parameters were considered real (positive) numbers instead. $f_{\text{cut}}(r_{ij})$ is a cutoff function introduced to remove the correction at very short r_{ij} distances:

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

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. As in our previous study, we set $s_{ij} = 10$. The functional form of our corrections, i.e., Equation (1), was justified in our proof-of-concept work [21]. As shown in that paper, the differences between the reference and the PM6 IPECs have, in general, the form of typical intermolecular potential energy curves or the form of decaying exponentials with negative amplitudes, in agreement with Equation (1).

As already mentioned, the parameters were evaluated by fits to differences between the interaction energies calculated at the B3LYP-D3/def2-TZVP level and those computed with the PM6 method. The SQM calculations were carried out with the MOPAC2016 program [50]. We used a least-squares nonlinear fitting procedure based on a genetic algorithm [55,56], as implemented in our GAFit code [57], with the following objective function, χ^2 :

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

where (x_i, y_i) represents one of the N data points, \mathbf{a} is the collective variable formed by the total number of fitting parameters and $f(x_i; \mathbf{a})$ 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 Equations (1) and (2), may be multiplied by a weighting factor (w_i), assigned to each data point. The use of a genetic algorithm has the advantage of efficiently exploring the search space and provides near optimal solutions when the number of fitting parameters is large. Since genetic algorithms may lead to many solutions that can be equally valid, our corrections should be viewed more as whole functional group corrections, rather than as individual pairwise corrections.

4. Summary and Conclusions

Bearing in mind the key role noncovalent interactions play in chemistry and biology, we recently presented a new approach, i.e., the PM6-FGC method, to improve the performance of SQM methods in the description of noncovalent interactions [21]. The central idea is to develop pairwise analytical corrections, using a simple functional form that resembles Buckingham's potential [58], but enhancing the parameter spectrum substantially. The parameters of the model function are adjusted by fits to differences between interaction energies calculated at the reference level and those evaluated with the SQM method (PM6 in our case) for complexes of small molecules, selected as representatives of functional groups. For each complex, we consider different orientations of the molecules, and include a series of configurations at several intermolecular distances. The inclusion of different orientations in the database is important to obtain well-balanced corrections. The corrections obtained should be considered as group corrections rather than as individual atom–pair corrections. To emphasize this, we chose the acronym FGC (Functional Group Corrections).

In our previous study [21], we used methane, formic acid, and ammonia as representatives to develop corrections for saturated hydrocarbons, carboxylic acids, amines, and, tentatively, amides. Although, in general, the PM6-FGC method provided quite good results, we found significant inaccuracies for some interactions involving $-\text{NH}_2$ groups. For this reason, in the present study, we performed new parameterizations, using methylamine

and acetamide as representatives of the amine and amide functional groups. The results of this work show a clear improvement over our previous parameterization and reinforce the importance of considering sufficient orientations of the interacting molecules in the reference database. We plan to extend the method to other functional groups relevant to biological compounds and implement the corrections in the MOPAC program. A Python script to calculate PM6-FGC corrections is available upon request.

Supplementary Materials: The following supplementary data are available online. Table S1: Mean absolute errors (MAEs) calculated for interaction energies below 5, 20 and 40 kJ/mol; Figures S1–S9: orientations of the bimolecular complexes considered in the fits and not shown in the manuscript; Figure S10: Complexes of the S66 database considered for validation.

Author Contributions: Conceptualization, methodology and writing—original draft preparation, S.A.V.; writing—review and editing, B.F., J.R.-O., E.M.C.-L. and S.A.V.; investigation, M.R.-G., B.F., J.R.-O., E.M.C.-L. and S.A.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ministerio de Ciencia e Innovación, grant numbers PID2019-107307RB-100 and PID2020-117605GB-100, and Xunta de Galicia, grant number ED431C 2021/40.

Data Availability Statement: Data are contained within the article and the Supplementary Materials.

Acknowledgments: We thank CESGA for computational time.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Sample Availability: Samples of the compounds are not available from the authors.

References

1. 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. [[CrossRef](#)]
2. 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. [[CrossRef](#)] [[PubMed](#)]
3. Řezáč, J.; Hobza, P. A halogen-bonding correction for the semiempirical PM6 method. *Chem. Phys. Lett.* **2011**, *506*, 286–289. [[CrossRef](#)]
4. Řezáč, J.; Hobza, P. Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods. *J. Chem. Theory Comput.* **2012**, *8*, 141–151. [[CrossRef](#)]
5. 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. [[CrossRef](#)]
6. Korth, M. Third-Generation Hydrogen-Bonding Corrections for Semiempirical QM Methods and Force Fields. *J. Chem. Theory Comput.* **2010**, *6*, 3808–3816. [[CrossRef](#)]
7. 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*, e449. [[CrossRef](#)] [[PubMed](#)]
8. Fiedler, L.; Gao, J.; Truhlar, D.G. Polarized Molecular Orbital Model Chemistry. 1. Ab Initio Foundations. *J. Chem. Theory Comput.* **2011**, *7*, 852–856. [[CrossRef](#)] [[PubMed](#)]
9. 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, Erratum in *J. Chem. Theory Comput.* **2012**, *8*, 2983–2983. [[CrossRef](#)]
10. 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. [[CrossRef](#)]
11. 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. [[CrossRef](#)] [[PubMed](#)]
12. Grimme, S. Accurate description of van der Waals complexes by density functional theory including empirical corrections. *J. Comp. Chem.* **2004**, *25*, 1463–1473. [[CrossRef](#)] [[PubMed](#)]
13. Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comp. Chem.* **2006**, *27*, 1787–1799. [[CrossRef](#)] [[PubMed](#)]
14. Kolb, M.; Thiel, W. Beyond the MNDO model: Methodical considerations and numerical results. *J. Comp. Chem.* **1993**, *14*, 775–789. [[CrossRef](#)]

15. Weber, W.; Thiel, W. Orthogonalization corrections for semiempirical methods. *Theor. Chem. Acc.* **2000**, *103*, 495–506. [[CrossRef](#)]
16. 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. [[CrossRef](#)]
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. [[CrossRef](#)]
18. Dewar, M.J.S.; Zebisch, 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. [[CrossRef](#)]
19. 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, Erratum in *J. Chem. Theory Comput.* **2014**, *10*, 1359–1360. [[CrossRef](#)]
20. 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. [[CrossRef](#)]
21. Pérez-Tabero, S.; Fernández, B. Cabaleiro-Lago, E.M.; Martínez-Núñez, E.; Vázquez, S.A., New Approach for Correcting Noncovalent Interactions in Semiempirical Quantum Mechanical Methods: The Importance of Multiple-Orientiation Sampling. *J. Chem. Theory Comput.* **2021**, *17*, 5556–5567. [[CrossRef](#)] [[PubMed](#)]
22. Hostaš, J.; Řezáč, J.; Hobza, P. On the performance of the semiempirical quantum mechanical PM6 and PM7 methods for noncovalent interactions. *Chem. Phys. Lett.* **2013**, *568*, 161–166. [[CrossRef](#)]
23. Grimme, S.; Hansen, A.; Brandenburg, J.G.; Bannwarth, C. Dispersion-Corrected Mean-Field Electronic Structure Methods. *Chem. Rev.* **2016**, *116*, 5105–5154. [[CrossRef](#)] [[PubMed](#)]
24. Sure, R.; Grimme, S. Comprehensive Benchmark of Association (Free) Energies of Realistic Host–Guest Complexes. *J. Chem. Theory Comput.* **2015**, *11*, 3785–3801. [[CrossRef](#)]
25. Brandenburg, J.G.; Hochheim, M.; Bredow, T.; Grimme, S. Low-Cost Quantum Chemical Methods for Noncovalent Interactions. *J. Phys. Chem. Lett.* **2014**, *5*, 4275–4284. [[CrossRef](#)]
26. 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. [[CrossRef](#)]
27. Becke, A.D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652. [[CrossRef](#)]
28. Becke, A.D. Density-functional thermochemistry. I. The effect of the exchange-only gradient correction. *J. Chem. Phys.* **1992**, *96*, 2155–2160. [[CrossRef](#)]
29. Becke, A.D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **1988**, *38*, 3098–3100. [[CrossRef](#)]
30. Johnson, E.R.; Becke, A.D. A post-Hartree–Fock model of intermolecular interactions. *J. Chem. Phys.* **2005**, *123*, 024101. [[CrossRef](#)]
31. Becke, A.D.; Johnson, E.R. A density-functional model of the dispersion interaction. *J. Chem. Phys.* **2005**, *123*, 154101. [[CrossRef](#)] [[PubMed](#)]
32. 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. [[CrossRef](#)] [[PubMed](#)]
33. 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. [[CrossRef](#)]
34. 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. [[CrossRef](#)]
35. Kendall, R.A.; Dunning, J.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. [[CrossRef](#)]
36. Meroueh, O.; Hase, W.L. Dynamics of energy transfer in peptide–Surface collisions. *J. Am. Chem. Soc.* **2002**, *124*, 1524–1531. [[CrossRef](#)]
37. 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. [[CrossRef](#)]
38. 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. [[CrossRef](#)]
39. 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. [[CrossRef](#)]
40. Pratihari, 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. [[CrossRef](#)]
41. 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. [[CrossRef](#)] [[PubMed](#)]
42. 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. [[CrossRef](#)] [[PubMed](#)]

43. 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. [CrossRef]
44. Hawkins, D.M. The Problem of Overfitting. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1–12. [CrossRef]
45. Martínez-Núñez, E. An automated method to find transition states using chemical dynamics simulations. *J. Comp. Chem.* **2015**, *36*, 222–234. [CrossRef] [PubMed]
46. Martínez-Núñez, E. An automated transition state search using classical trajectories initialized at multiple minima. *Phys. Chem. Chem. Phys.* **2015**, *17*, 14912–14921. [CrossRef]
47. Rodríguez, A.; Rodríguez-Fernández, R.; Vázquez, S.A.; Barnes, G.L.; Stewart, J.J.P.; Martínez-Núñez, E. tssc2018: A code for automated discovery of chemical reaction mechanisms and solving the kinetics. *J. Comp. Chem.* **2018**, *39*, 1922–1930. [CrossRef] [PubMed]
48. 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*, e26008. [CrossRef]
49. Martínez-Núñez, E.; Barnes, G.L.; Glowacki, D.R.; Kopec, S.; Peláez, D.; Rodríguez, A.; Rodríguez-Fernández, R.; Shannon, R.J.; Stewart, J.J.P.; Tahoces, P.G.; et al. AutoMeKin2021: An open-source program for automated reaction discovery. *J. Comp. Chem.* **2021**, *42*, 2036–2048. [CrossRef]
50. Stewart, J.J.P. MOPAC2016, 16.307; Stewart Computational Chemistry. Available online: <http://openmopac.net/> (accessed on 1 April 2021).
51. 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. [CrossRef]
52. 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. [CrossRef]
53. Neese, F. Software update: The ORCA program system, version 4.0. *WIREs Comput. Mol. Sci.* **2018**, *8*, e1327.
54. Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA quantum chemistry program package. *J. Chem. Phys.* **2020**, *152*, 224108. [CrossRef] [PubMed]
55. Marques, J.M.C.; Prudente, F.V.; Pereira, F.B.; Almeida, M.M.; Maniero, A.M.; Fellows, C.E. A new genetic algorithm to be used in the direct fit of potential energy curves to ab initio and spectroscopic data. *J. Phys. B At. Mol. Opt. Phys.* **2008**, *41*, 085103. [CrossRef]
56. Almeida, M.M.; Prudente, F.V.; Fellows, C.E.; Marques, J.M.C.; Pereira, F.B. Direct fit of spectroscopic data of diatomic molecules by using genetic algorithms: II. The ground state of RbCs. *J. Phys. B At. Mol. Opt. Phys.* **2011**, *44*, 225102. [CrossRef]
57. 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. [CrossRef]
58. Buckingham, R.A. The classical equation of state of gaseous helium, neon and argon. *Proc. R. Soc. Lond. A* **1938**, *168*, 264–283.