PDDG/PM3 and PDDG/MNDO:
Improved Semiempirical Methods

MATTHEW P. REPASKY, JAYARAMAN CHANDRASEKHAR,* WILLIAM L. JORGENSEN

Department of Chemistry, Yale University, 225 Prospect St.,
New Haven, Connecticut 06520-8107

Abstract: Two new semiempirical methods employing a Pairwise Distance Directed Gaussian modification have been developed: PDDG/PM3 and PDDG/MNDO; they are easily implemented in existing software, and yield heats of formation for compounds containing C, H, N, and O atoms with significantly improved accuracy over the standard NDDO schemes, PM5, PM3, AM1, and MNDO. The PDDG/PM3 results for heats of formation also show substantial improvement over density functional theory with large basis sets. The PDDG modifications consist of a single function, which is added to the existing pairwise core repulsion functions within PM3 and MNDO, a reparameterized semiempirical parameter set, and modified computation of the energy of formation of a gaseous atom. The PDDG addition introduces functional group information via pairwise atomic interactions using only atom-based parameters. For 622 diverse molecules containing C, H, N, and O atoms, mean absolute errors in calculated heats of formation are reduced from 4.4 to 3.2 kcal/mol and from 8.4 to 5.2 kcal/mol using the PDDG modified versions of PM3 and MNDO over the standard versions, respectively. Several specific problems are overcome, including the relative stability of hydrocarbon isomers, and energetics of small rings and molecules containing multiple heteroatoms. The internal consistency of PDDG energies is also significantly improved, enabling more reliable analysis of isomerization energies and trends across series of molecules; PDDG isomerization energies show significant improvement over B3LYP/6-31G* results. Comparison of heats of formation, ionization potentials, dipole moments, isomer, and conformer energetics, intermolecular interaction energies, activation energies, and molecular geometries from the PDDG techniques is made to experimental data and values from other semiempirical and *ab initio* methods.

© 2002 Wiley Periodicals, Inc.   J Comput Chem 23: 1601–1622, 2002

Key words: semiempirical MO; heats of formation; composite *ab initio;* PM3; NDDO

Introduction

Modern semiempirical methods based on the Neglect of Diatomic Differential Overlap (NDDO) formalism[1–3] were introduced in 1977 with the development of MNDO.[4,5] Continuing effort in the field led to the advanced methods, AM1,[6] PM3,[7,8] SAMI,[9] and PM5,[10] which continue to find widespread usage in a variety of applications. However, due to their limited accuracy in reproducing experimental data for a diverse collection of quantum mechanical observables, they are in danger of being displaced by *ab initio* and density functional theory (DFT) techniques as greater computational power becomes readily available. Of particular concern to chemists are the substantial errors in semiempirically calculated heats of formation for "organic" molecules consisting of hydrogen, carbon, nitrogen, and/or oxygen atoms. For instance, over 56 neutral, closed-shell ground state molecules in the combined G2-1[11] and G2-2[12] sets containing only H, C, N, and O atoms (not including diatomics or CO_2), mean absolute errors (MAEs) in

heats of formation are found to be 0.6, 1.3, 1.8, 2.3, and 3.3 kcal/mol by the *ab initio* G3,[13] G2,[11] G2(MP2),[14] B3LYP,[12] and CBS-4,[15] composite methods, respectively. With commonly used semiempirical methods, MAEs for the same set of molecules are found to be 4.1, 5.1, and 6.8 kcal/mol with PM3, AM1, and MNDO. In addition to larger MAEs, those in semiempirically calculated heats of formation are often unsystematic, making the comparison of properties for groups of related molecules unreli-

able. Furthermore, a large number of common errors that limit the applicability of these methods to chemically interesting systems have been found.[16] Examples of deficiencies in NDDO calculated results include, (1) erroneous prediction of branched chain hydrocarbons being less stable than their straight chain analogs, (2) poor treatment of the energetics of bonds in which one or more atomic component possess lone-pair electrons, (3) flawed energetics for molecules with multiple rings, (4) underestimation of the stabilization energy of delocalized systems, (5) general overestimation of the energy associated with the methylene fragment, which rapidly skews heats of formation of molecules containing long alkyl chains, (6) inaccurate description of intermolecular interactions, especially hydrogen bonding, (7) poor treatment of conformational energetics including *cis-trans* isomers and rotomeric energies, (8) incorrectly predicted molecular geometries for a number of molecules including hydrogen peroxide and cyclobutane and, (9) significant errors in proton affinities.[17,18]

A number of attempts have been made in recent years to overcome the major shortcomings of the commonly used NDDO based methods without diminishing their strengths, namely their speed and applicability to a wide range of chemically interesting problems. The lack of orbital orthogonality and electron correlation, and parameterization issues have been suggested as the primary culprits in generating errors in NDDO methods, and several formalism improvements have been recommended.[19,20] Many of these formalism modifications have focused on improving molecular energies by reducing the number of approximations in the NDDO formalism. Orbital orthogonality has been restored in an approximate manner in Thiel's OM1[21] and OM2[22] methods, which have been suggested to be the most accurate NDDO-based methods for computing heats of formation to date.[23] They have also been shown to be the semiempirical method of choice for treating amino acids.[24] Potential pitfalls in parameterization have been bypassed with the specific reaction parameterization concept (SRP)[25] of Truhlar, which has enjoyed considerable success in generating dependable results for describing the potential energy surfaces of specific reactions[18,26–30] and intermolecular complexes.[31,32] These techniques, while more accurate than existing NDDO methods, require some sacrifice of speed or wide applicability. OM1 and OM2 require significant modification to the semiempirical formalism, and thus are not yet widely available, while the SRP treatment requires a separate parameterization for each system of interest, making this scheme only applicable to a specific problem.

As a further approach, in our lab we have recently developed the Bond and Group Equivalents (BGE) scheme,[33] which provides vastly improved heats of formation for semiempirical methods. The BGE method uses a number of bond and group equivalents to convert semiempirical molecular energies to heats of formation through the equation:

$$\Delta H_f = E_{\text{mol}} + T/R + \sum_j a_j n_j \qquad (1)$$

Here, the total molecular energy, E_{mol}, is converted to a heat of formation using translational and rotational corrections (T/R), a value of 4RT for nonlinear molecules at 298 K, in addition to a sum of the products of equivalents, a_j, and the number of times they occur in a given molecule, n_j. Two types of equivalents are defined: bond

equivalents, which account for the energy of atoms bound to each other, and group equivalents, which are used to account for variation in the chemical environment surrounding an atom. Using the BGE scheme with standard semiempirical molecular energies, mean absolute errors are reduced for AM1, PM3, and MNDO from 6.6, 4.2, and 8.2 kcal/mol to 2.3, 2.2, and 3.0 kcal/mol, respectively, for a set of 583 neutral, closed-shell molecules. Furthermore, a number of the systematic NDDO problems previously introduced are also overcome by this scheme, including the relative stability of branched hydrocarbons, energetics of conjugated systems, heats of formation for long chain hydrocarbons, and enthalpies of molecules containing multiple heteroatoms. Because of its reliance on user or algorithm defined fragment or functional group information, the BGE scheme may not be applied generally, rather only for molecules at stationary points on the molecular surface with conventional bonding. Thus, we have sought improvements in calculated heats of formation of a similar magnitude through development of general semiempirical methods by introducing fragment or functional group information in an atom-centered approach.

The most suitable tactic for introducing such information into semiempirical formalisms without adversely affecting molecular electronic energies is through modification of the core repulsion function (CRF). There is precedence in the modification of semiempirical CRFs to generate improved techniques, although efforts in this arena should be undertaken with care as this type of adjustment has been found to be problematic.[34,35] The only difference in formalism between the advanced PM3 and AM1 methods, and the older MNDO scheme is the addition of a number of Gaussian functions, which were added to the CRF to reduce excessive core–core repulsions just outside bonding distances.

$$\text{CRF}_{\text{AM1,PM3}}(A, B) = \text{CRF}_{\text{MNDO}}(A, B) + \left(\frac{Z_A Z_B}{R_{AB}} \right)$$

$$\times \left[\sum_k a_{kA} e^{-b_{kA}(R_{AB} - c_{kA})^2} + \sum_k a_{kB} e^{-b_{kB}(R_{AB} - c_{kB})^2} \right] \qquad (2)$$

Here, the core repulsion function for the atom pair consisting of atoms A and B is calculated by adding the MNDO CRF to a number of gaussians with three parameters, a_k, b_k, and c_k. PM3 has exactly two Gaussians per element ($k = 2$), while AM1 has between two and four, depending on the element. These Gaussians may be thought of as a first attempt at introducing atom-based functional group information in a general manner into the CRF. However, the AM1 and PM3 Gaussians are held too closely to the atoms upon which they are based to be used effectively to differentiate between functional groups. Furthermore, as the location of Gaussians is only directed by the single distance parameter, c_k, this functional form is again inappropriate for trying to distinguish between various functional groups.

Another addition to the CRF was proposed by Bernal–Uruchurtu et al. in their method for improving the energetics and structures of hydrogen bonding.[36,37] For intermolecular atomic pairs, they replaced the Gaussians of PM3 with the pairwise parameterizable interaction function:

$$\text{CRF} = \text{CRF}_{\text{MNDO}} + \sum_{A,B}^{\text{inter}} \alpha_{AB} e^{-\beta_{AB} R_{AB}} + \frac{\chi_{AB}}{R_{AB}^6} + \frac{\delta_{AB}}{R_{AB}^8} + \frac{\varepsilon_{AB}}{R_{AB}^{10}} \quad (3)$$

Using this function, which employs bond-based empirical parameters (α_{AB}, β_{AB}, χ_{AB}, δ_{AB}, and ε_{AB}), they vastly improved the energetics and structures of water dimers and clusters. The significant advantages of AM1 and PM3 over MNDO, and the greatly improved treatment of hydrogen bonds suggests that by the use of an appropriate function added to existing semiempirical CRFs, an improved approach that generates accurate heats of formation by following the BGE example may be possible.

The BGE scheme additionally illustrated the importance of the electronic energy of formation of a gaseous atom, the atomic eisol parameter, in obtaining accurate heats of formation. In all NDDO schemes, eisol is the electronic energy of an element calculated with a restricted single-determinant wave function using the desired semiempirical formalism and parameter set.[38] Eisol is a derived parameter in that its value is dependent in a nonsystematic way on the values of all one-center parameters within the semiempirical formalism. As semiempirical methods are parameterized to reproduce properties for molecules that generally consist of several heavy atoms, the values of eisol which arise are most likely erroneous. This could help explain the generally poor energetic behavior of semiempirical methods in treating diatomics and other small molecules. In the BGE study, values of the eisol terms were reevaluated for the common organic elements using through-origin multiple regression over a large number of experimental heats of formation. Fairly significant differences were found between the original-derived eisol values and those that arose from the regression. Thus, following the example set forth by the BGE scheme, reevaluation of eisol terms in addition to adding functional group information into the CRF should produce a general method capable of generating significantly improved heats of formation.

By adding functional group information into the CRF and reevaluating the procedure for deriving eisol parameters, we have developed a novel semiempirical formalism that simultaneously reduces overall errors in heats of formation and limits systematic errors in semiempirically calculated energies. Errors in other observables such as geometries, ionization potentials, dipole moments, etc., in the advanced PM3 and AM1 schemes have been deemed acceptable, and effort was made only to avoid significant degradation in their results. Following application of the formalism and reparameterization of the semiempirical parameter sets, the goal of developing improved semiempirical methods with only a negligible increase in computation time has been realized and embodied in the newly developed Pairwise Distance Directed Gaussian methods, PDDG/PM3 and PDDG/MNDO. Results for a number of important observables are presented in this article.

Development of the PDDG Formalism

The most effective form of the function used in conjunction with semiempirical methods to distinguish between various functional groups is unknown. There are a number of considerations of importance in designing this term.

1. The interactions introduced by this function must make small contributions to the overall molecular energy, or they may overwhelm the electronic portion and adversely alter optimized molecular geometries.
2. The individual terms must be able to differentiate between a wide range of functional groups based on molecular geometries with a limited number of parameters.
3. Bond specific parameters must not be used to avoid the trap of an exponentially expanding parameter set.
4. No parameters must be introduced for specific functional groups or interactions.
5. The corrections must not introduce significant errors in molecules with nonstandard bonding, such as charged species and transition states.

Testing revealed the most successful function, which fulfilled all the previously mentioned criteria to be composed of four weighted Gaussians for heterodimer atom pairs ($A \neq B$) and three weighted Gaussians for homodimers ($A = B$): the Pairwise Distance Directed Gaussian function (PDDG).

$$\text{PDDG}(A, B) = \sum_A \sum_{B>A} \frac{1}{n_A + n_B}$$

$$\times \left[\sum_{i=1}^{2} \sum_{j=1}^{2} (n_A P_{A_i} + n_B P_{B_j}) \exp(-10\text{Å}^{-2}(R_{AB} - D_{A_i} - D_{B_j})^2) \right]$$

(4)

Here, the sum is over all distinct atom pairs formed by looping over atoms *A* and *B*. Each element requires a total of four optimizable parameters for the PDDG function, two preexponential factors, P_{A_1} and P_{A_2}, and two distance terms, D_{A_1} and D_{A_2}. The function is weighted using the number of valence electrons on atoms A and B, n_A and n_B, in an effort to avoid problems between atom pairs of large and small atomic numbers. This functional form exhibits the necessary flexibility and shape to differentiate between various functional groups using only a small number of atom-based parameters. By using the difference in atomic distance terms to locate Gaussians and the sum of atomic preexponential terms to give their amplitude, the PDDG function utilizes only atom-based terms to generate bond-specific Gaussians. Furthermore, because of its generality this function may be used to enhance any NDDO-based semiempirical methods without significant effort or increase in computation time. The PDDG function has been added to CRFs arising from PM3 and MNDO within the MOPAC 6.0 program[38] generating, after reparameterization, the new PDDG/PM3 and PDDG/MNDO methods. For comparison purposes, the AM1, PM3, and MNDO Hamiltonians within MOPAC 6.0 were used while PM5 results were generated via the CAChe 5.0 program from Fujitsu.

Parameterization of PDDG Methods

A semiempirical method is comprised of both a mathematical formalism and an optimized set of parameters that best reproduce experimental data within that formalism. Having completely de-

Table 1. Weighting Factors Used in Parameterization of the PDDG Methods.

Reference function	Weighting factor
Heats of formation	1 mol/kcal
Ionization potentials	10 eV^{-1}
Dipole moments	20 debye^{-1}
Bond lengths[a]	100 Å^{-1}
Bond angles[a]	2/3 deg.$^{-1}$
Dihedral angles[a]	1/3 deg.$^{-1}$
Gradients for bond lengths[b]	0.1 mol/kcal Å
Gradients for bond angles[b]	0.1 mol/kcal deg.
Gradients for dihedral angles[b]	0.1 mol/kcal deg.

[a]For parameterization using flexible molecular geometries.
[b]For parameterization using fixed molecular geometries.

fined the PDDG modification to standard NDDO methods above, a set of optimized parameters that take advantage of the PDDG function is needed. There are a number of different approaches to follow in developing optimized semiempirical parameter sets, of which the two primary schemes are: (1) follow the PM3 ideal and reoptimize all possible parameters (including one-center two-electron repulsion integrals) with an emphasis on reducing overall

errors,[8] and (2) follow the AM1/MNDO ideal and optimize a smaller set of parameters, keeping the remaining terms at values suggested by experimental data, with an emphasis not solely on limiting overall errors, but on utilizing chemical intuition to generate a reliable method.[4] We believe a combination of the two approaches is most appropriate. The set of parameters for which optimal values were determined in Dewar's methods (U_{ss}, U_{pp}, β_s, β_p, ζ_s, ζ_p, α, and all Gaussian preexponential and distance terms) in addition to parameters introduced by the PDDG function (P_{A_1}, P_{A_2}, D_{A_1}, D_{A_2}) are optimized, retaining the values of all one-center two-electron integrals (G_{ss}, G_{pp}, G_{sp}, G_{p2}, H_{sp}) at their standard AM1/MNDO or PM3 values. The optimization procedure was designed to reproduce a wide range of experimental data, while simultaneously aiming to eliminate systematic problems and minimize overall errors in heats of formation.

An optimized parameter set is determined by optimizing the value of an error function, S, defined as,

$$S = \sum_i w_i^2 (q_i^{\text{calc}} - q_i^{\text{exp}})^2 + \sum_j w_j \frac{\delta E}{\delta x_j} + (\chi^2)^4 \tag{5}$$

Here, the reference functions, q_i, are experimental and calculated heats of formation, ionization potentials, dipole moments, and geometrical variables such as bond lengths, bond angles, or dihe-

Table 2. Comparison of the PDDG/PM3 and PM3 Parameter Sets.[a]

	PM3 parameters[b]				PDDG/PM3 parameters			
	H	C	N	O	H	C	N	O
U_{ss}	−13.073321	−47.270320	−49.335672	−86.993002	−12.893272	−48.241241	−49.454546	−87.412505
U_{pp}		−36.266918	−47.509736	−71.879580		−36.461256	−47.757406	−72.183070
β_s	−5.626512	−11.910015	−14.062521	−45.202651	−6.152654	−11.952818	−14.117230	−44.874553
β_p		−9.802755	−20.043848	−24.752515		−9.922411	−19.938509	−24.601939
ζ_s	0.967807	1.565085	2.028094	3.796544	0.972786	1.567864	2.035807	3.814565
ζ_p		1.842345	2.313728	2.389402		1.846659	2.324327	2.318011
α	3.356386	2.707807	2.830545	3.217102	3.381686	2.725772	2.849124	3.225309
eisol	−13.073321	−111.229917	−157.613776	−289.342207	−13.120566	−113.428242	−158.416205	−292.188766
DD		0.833240	0.657701	0.408617		0.831413	0.654855	0.403741
QQ		0.664775	0.529338	0.512574		0.663222	0.526924	0.528360
ρ_0^c	0.919616	1.214657	1.142818	0.863494	0.919616	1.214657	1.142818	0.863494
ρ_1^c		0.849494	0.993838	0.943462		0.848467	0.991235	0.936266
ρ_2^c		0.653794	0.678893	0.611275		0.652785	0.676704	0.624291
a_1	1.128750	0.050107	1.501674	−1.131128	1.122244	0.048906	1.513320	−1.138455
b_1	5.096282	6.003165	5.901148	6.002477	4.707790	5.765340	5.904394	6.000043
c_1	1.537465	1.642214	1.710740	1.607311	1.547099	1.682232	1.728376	1.622362
a_2	−1.060329	0.050733	−1.505772	1.137891	−1.069737	0.047697	−1.511892	1.146007
b_2	6.003788	6.002979	6.004658	5.950512	5.857995	5.973721	6.030014	5.963494
c_2	1.570189	0.892488	1.716149	1.598395	1.567893	0.894406	1.734108	1.614788
P_{A_1}					0.057193	−0.000743	−0.003160	−0.001000
P_{A_2}					−0.034823	0.000985	0.012501	−0.001522
D_{A_1}					0.663395	0.836915	1.004172	1.360685
D_{A_2}					1.081901	1.585236	1.516336	1.366407

[a]Parameter units are as follows: (eV) U_{ss}, U_{pp}, β_s, β_p, eisol, a_1, a_2, P_{A_1}, P_{A_2}; (au) ζ_s, ζ_p; (Bohr) DD, QQ, ρ_0, ρ_1, ρ_2; (Å) c_1, c_2, D_{A_1}, D_{A_2}; (Å^{-1}) α, b_1, b_2.
[b]Ref. 8.
[c]For use in MOPAC 6, $\rho_0 = 0.5/AM$, $\rho_1 = 0.5/AD$, $\rho_2 = 0.5/AQ$.

Table 3. Comparison of the PDDG/MNDO and MNDO Parameter Sets.[a]

	MNDO parameters[b]				PDDG/MNDO parameters			
	H	C	N	O	H	C	N	O
U_{ss}	−11.906276	−52.279745	−71.932122	−99.644309	−11.724114	−53.837582	−71.871894	−97.884970
U_{pp}		−39.205558	−57.172319	−77.797472		−39.936409	−58.216617	−77.342674
β_s	−6.989064	−18.985044	−20.495758	−32.688082	−7.493504	−18.841334	−20.375774	−33.606336
β_p		−7.934122	−20.495758	−32.688082		−7.922234	−21.085373	−27.984442
ζ_s	1.331967	1.787537	2.255614	2.699905	1.322431	1.809817	2.231424	2.569172
ζ_p		1.787537	2.255614	2.699905		1.825008	2.253460	2.697152
α	2.544134	2.546380	2.861342	3.160604	2.491813	2.555522	2.843678	3.238842
eisol	−11.906276	−120.500606	−202.566201	−317.868506	−12.015956	−123.864412	−206.466626	−310.879745
DD		0.807466	0.639904	0.534602		0.794158	0.643624	0.547344
QQ		0.685158	0.542976	0.453625		0.671090	0.543495	0.454088
ρ_0^c	1.058920	1.112429	1.001103	0.882296	1.058920	1.112429	1.001103	0.882296
ρ_1^c		0.813078	0.637459	0.521237		0.805697	0.639479	0.527927
ρ_2^c		0.747842	0.615275	0.526541		0.737353	0.615679	0.526913
P_{A_1}					−0.108861	−0.006889	0.035027	0.086344
P_{A_2}					−0.024706	−0.027751	−0.001721	0.030403
D_{A_1}					0.460721	1.192456	1.011630	0.725408
D_{A_2}					1.298731	1.329522	2.278423	0.709728

[a]Parameter units are as follows: (eV) U_{ss}, U_{pp}, β_s, β_p, eisol, P_{A_1}, P_{A_2}; (au) ζ_s, ζ_p; (Bohr) DD, QQ, ρ_0, ρ_1, ρ_2; (Å) D_{A_1}, D_{A_2}; (Å^{-1}) α.
[b]Ref. 4.
[c]For use in MOPAC 6, $\rho_0 = 0.5/AM$, $\rho_1 = 0.5/AD$, $\rho_2 = 0.5/AQ$.

dral angles, as well as more complicated functions giving isomerization or intermolecular interaction energies. A total of 384 reference functions were used encompassing 134 neutral, closed-shell molecules, all of which were employed in the regression to determine values for eisol parameters. Additional energetic reference functions included six charged or open-shell molecules (CHO cation, CHO radical, hydroxide anion, benzyl cation, azide radical, and 2-propyl radical) and six isomer energies including axial/equitorial energy differences for cyclohexanol and 2-methoxytetrahydropyran, E/Z isomer energies for 1,3-butadiene, methyl acetate, and *N*-methylacetamide, and the staggered/eclipsed isomer energy for *n*-butane. All neutral, closed-shell molecules whose heats of formation were included as reference functions are listed in Table S1 of the Supplementary Material as being included in the fit set. Reference functions arising from dipole moments, ionization potentials, and molecular geometries were limited to arising from molecules in the fit set. Default weighting factors (w_i, w_j) are listed in Table 1, although these were adjusted for a small number of individual reference functions to emphasize or reduce the role they played in the error function. Chi squared (χ^2) provides an estimate of the quality of the through-origin multiple regression fit that generates eisol parameters,[39] and is taken to the fourth power to harshly penalize parameter sets that yield poor energetic results over the entire fit set of molecules. Close correlation between the quality of this fit and the ability of the new semiempirical methods to reproduce experimental data makes its inclusion in the error function useful. Due to the large number of molecules for which calculations had to be performed, initial optimization of parameters was accomplished using single-point geometries with molecules at their original semiempirical geometric minima. Exclu-

sively in these cases, gradients of the energy with respect to molecular geometry were multiplied by a weighting factor and added to the error function.

Only a single formalism modification to the original semiempirical methods was made. Therefore, the differences between current semiempirical parameter sets and those optimized using the PDDG function should be relatively small. Thus, local minima of *S* relative to the original parameter sets for PM3 and MNDO were sought, ultimately leading to the PDDG/PM3 and PDDG/MNDO methods. This was accomplished by a multistep parameterization procedure beginning with the two initial NDDO methods utilizing a combination of gradient based and simulated annealing optimization algorithms. First, 648 different combinations[40] for the Gaussian distance terms (D_{A_1}, D_{A_2}) were initially scanned using semiempirically determined single-point molecular geometries. Previous testing had shown that these distances were most likely to produce effective functional group differentiation, although other distances are certainly capable of yielding similar effects. During this procedure the distance terms and semiempirical parameters were fixed, and only values for the preexponential terms (P_{A_1}, P_{A_2}) were optimized. Taking the optimized P_{A_1} and P_{A_2} parameters from this initial scan, they were then reoptimized over the 648 combinations of the distance terms to minimize the error function using fully flexible molecular geometries. All semiempirical parameters were then optimized starting with the 648 sets of optimized P_{A_1} and P_{A_2} parameters in addition to the two distance parameters per element using single-point molecular geometries. Finally, beginning with the optimized parameter set found using single-point geometries, the semiempirical parameters and four PDDG parameters per element were reoptimized a final
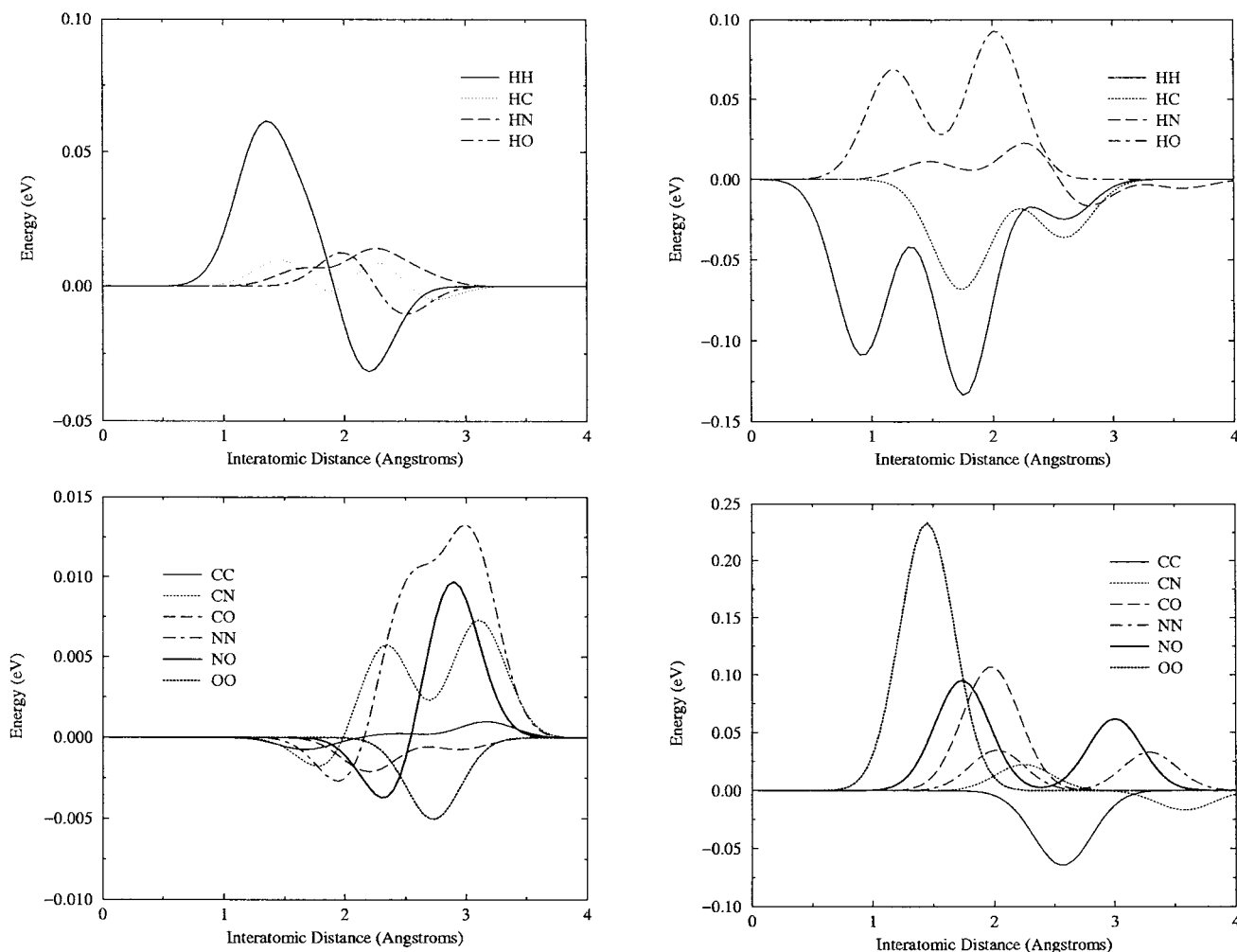
**PDDG/PM3 Gaussians   PDDG/MNDO Gaussians**

Figure 1. Plots of PDDG/PM3 and PDDG/MNDO Gaussians [eq. (4)] as functions of the interatomic distance.

time using flexible molecular geometries. Because of the large number of parameters being simultaneously optimized this was a very difficult undertaking, and certainly illustrated what other groups have previously discovered; parameter development for semiempirical methods is a taxing "art form." [27]

Results and Discussion

Optimized values for the PDDG/PM3 parameter set are shown in Table 2, along with standard PM3 terms, while PDDG/MNDO and MNDO parameters are compared in Table 3. The similarity of the PDDG and standard semiempirical parameter sets suggests our attempts to locate local minima near the original parameter sets were successful. Only three PDDG/PM3 parameters deviated by

more than 5% from standard PM3 terms, hydrogen β_s, hydrogen b_1, and carbon a_2. For PDDG/MNDO, three parameters were found to differ by >5% from standard MNDO values, hydrogen β_s, oxygen β_p, and oxygen ζ_s. The largest deviations from MNDO and PM3 were in the two-center one-electron core resonance integrals (β), with differences of 14.4% (O β_p) and 8.6% (H β_s), respectively. This is reasonable as the two-center one-electron integrals largely account for chemical bonding energies; thus, they are expected to be affected by the PDDG Gaussians, which should exhibit some 1,2-interaction character. The eisol parameters in Tables 2 and 3 are those derived via through-origin multiple regression and differ somewhat from eisol values determined using the standard procedure with the PDDG optimized parameter sets (PDDG/PM3 derived eisol terms (eV): H −12.893272, C −113.560435, N −158.594535, and O −291.395172; PDDG/
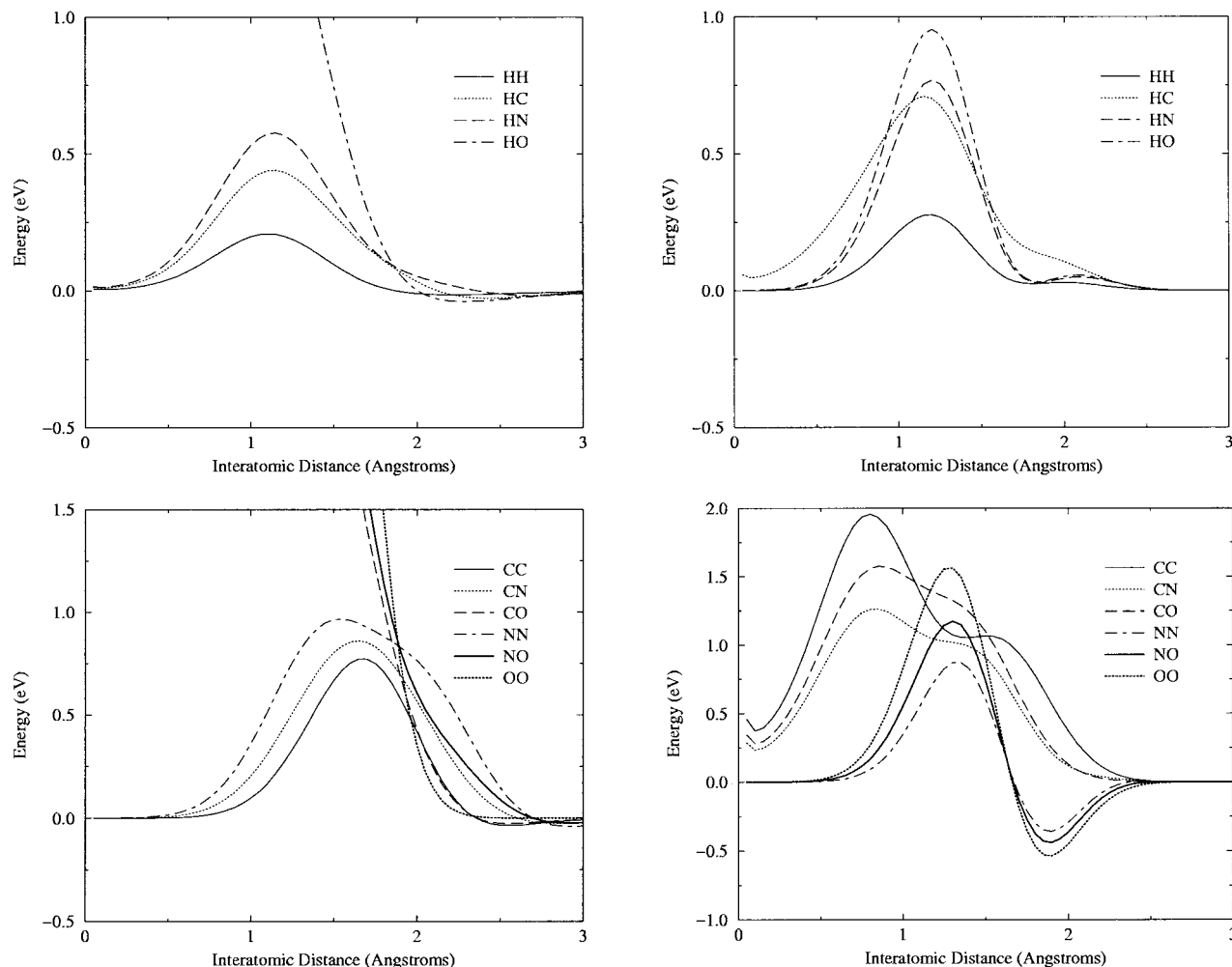
AM1 Gaussians

PM3 Gaussians

Figure 2. Plots of AM1 and PM3 Gaussians [eq. (2) neglecting $CRF_{MNDO}(A, B)$] as functions of the interatomic distance.

MNDO derived eisol terms (eV): H -11.724114, C -125.077983, N -205.578639, and O -312.530636).

Optimized values of the PDDG distance parameters, D_{A_1} and D_{A_2}, are well positioned to differentiate between various functional groups. The distance parameters for H, C, N, and O range from 0.66 to 1.59 Å in PDDG/PM3 and from 0.46 to 2.28 Å in PDDG/MNDO. Given that the PDDG Gaussian distances are a combination of two atomic D_{A_1} and/or D_{A_2} values, the range of Gaussian locations spans from 1.32 to 3.18 Å and from 0.92 to 4.56 Å, respectively. This is sufficient to account for 1,2-, 1,3-, and 1,4-interactions. Functional forms of the pairwise PDDG corrections [eq. (4)] are illustrated in Figure 1 for all possible pairs of H, C, N, and O atoms, while AM1 and PM3 Gaussians [eq. (2) neglecting $CRF_{MNDO}(A, B)$] are presented in Figure 2. Off-scale AM1 maxima for HO, CN, NO, and OO atomic pairs are located at interatomic separations of 0.8, 0.7, 0.7, and 0.7 Å with magnitudes of 2.53, 8.70, 10.91, and 26.08 eV, respectively. Comparison

of PDDG/MNDO and AM1 Gaussians shows the stark contrast between our approach of adding functional group information into the CRF and Dewar's "brute force" correction of some glaring deficiencies of MNDO.[18] The amplitude of the PDDG Gaussians, which is directly controlled by the preexponential P_{A_1} and P_{A_2} terms, is at least an order of magnitude smaller than AM1 and PM3 Gaussians. The small magnitude of PDDG Gaussians is very favorable, as previous attempts at including additional functions into the CRF, other than the standard adjustment for the net electrostatic repulsion made in MNDO, have lead to unphysical behaviors.[34,35] For PM3, the addition of Gaussians with oscillations introduced spurious minima into the potential energy surface. Although PDDG Gaussians show oscillatory behavior, it is doubtful that the contribution from any given atom pair would be able to overwhelm the electronic energy and create a false minimum thanks to the very small magnitude of our Gaussians. Note that as the PDDG/PM3 method is built upon PM3 with only a local

Table 4. Mean Absolute Errors in Calculated Heats of Formation (kcal/mol).

Molecules	PDDG		Standard NDDO methods				Num. mol.
	PM3	MNDO	PM5	PM3	AM1	MNDO	
Overall	3.2	5.2	4.8	4.4	6.7	8.4	622
Training set[a]	2.7	3.6	4.5	4.3	6.1	7.5	134
Test set	3.3	5.6	4.9	4.4	6.8	8.6	488
Acyclic alkanes	1.2	3.9	5.2	3.1	5.3	17.2	48
Acyclic alkenes	1.7	4.7	2.0	2.1	3.1	5.0	63
Acyclic alkynes	3.0	9.2	8.2	5.0	5.6	6.4	25
Aromatics	2.7	3.9	2.1	2.7	3.5	3.1	26
Cyclic alkanes	3.7	5.6	7.1	5.0	8.7	7.2	62
Cyclic alkenes	3.2	4.9	5.7	3.9	7.0	6.7	28
All cyclic hydrocarbons	3.5	5.4	6.7	4.7	8.1	7.0	90
All acyclic hydrocarbons	1.8	5.2	4.2	3.0	4.3	9.5	137
All HC mol.	2.6	5.1	4.9	3.6	5.6	8.0	254
Alcohols	2.6	4.1	2.3	2.7	7.4	6.0	52
Ethers	3.2	3.9	5.4	4.8	10.6	8.5	50
Ketones	3.6	6.5	5.8	4.0	7.1	11.6	47
Aldehydes	3.3	5.3	2.1	4.5	5.2	3.3	14
Acids	2.7	4.1	3.4	3.1	5.4	6.0	18
Esters	3.9	3.4	4.4	7.2	5.0	8.8	44
Peroxides	2.3	10.6	5.2	5.2	7.9	13.0	10
All HCO mol.	3.2	5.0	4.6	4.6	7.2	8.7	238
Aliphatic amines	2.3	5.6	3.9	3.0	4.2	5.7	28
Nitriles	3.0	2.4	4.1	4.3	4.2	5.4	12
Aromatic N mol.	5.4	4.6	4.7	4.2	10.3	6.7	31
All HCN mol.	4.2	5.7	5.2	4.6	7.3	6.3	89
Amides	4.9	3.5	4.1	8.0	8.5	12.0	18
Nitro and nitrates	3.1	7.8	4.3	3.6	7.7	21.5	13
All HCNO mol.	4.5	4.9	4.9	7.0	9.5	13.4	41

[a]Neutral, closed-shell molecules only.

optimization of the parameter set, spurious minima introduced by the PM3 CRF Gaussians are retained.

Improved Accuracy in Predicted Heats of Formation

Using the PDDG/PM3 and PDDG/MNDO methods, computed heats of formation show mean absolute errors of 3.2 and 5.2 kcal/mol, respectively, for a set of 622 neutral, closed-shell molecules (Table 4). For comparison of results, PDDG/PM3 is most similar to PM3 in terms of values of the semiempirical parameters. The number of Gaussians in the CRF expands from four per atom pair in PM3 to seven or eight in PDDG/PM3. PDDG/MNDO is similar in terms of semiempirical parameters to both AM1 and MNDO and utilizes three or four Gaussians per atom pair. AM1 has between four and eight Gaussians per atom pair for compounds containing H, C, N, and O elements, while MNDO has none. Thus, PDDG/MNDO and AM1 are directly comparable due to similar semiempirical parameter sets and a comparable number of Gaussians per atom pair. Variation in the results of AM1 and PDDG/MNDO will reflect the underlying differences arising from the use of atom-centered Gaussians vs. pairwise PDDG Gaussians.

The PDDG/PM3 method provides improvements of 27 and 33% over PM3 and PM5 heats of formation, respectively, for the entire set of molecules containing H, C, N, or O atoms. The PDDG/MNDO method is a 22% improvement over AM1 and an impressive 38% better than MNDO. As expected, mean absolute errors from training set molecules were smaller than errors arising from the test set. Thus, the error function as implemented is not directly proportional to the quality of results generated by the optimized parameter set over a collection of molecules that differ from the fit set. Further improvements in the error function are possible to reduce this dependence on the molecules selected to comprise the fit set. To evaluate the importance of contributions from the PDDG Gaussians and from modification to the treatment of the eisol parameter, a full parameter optimization was performed in which no PDDG functions were employed. An overall MAE for the set of 622 molecules was then found to be 7.3 kcal/mol, only a modest 13% improvement over standard MNDO. Thus, the large improvements observed from the PDDG methods are not based solely on the eisol modification. Nor do improvements in heats of formation arise simply from increasing the number of Gaussians in the CRF, indicated by the 22% improvement of PDDG/MNDO over AM1, because PDDG/MNDO uses fewer Gaussians than AM1. Thus, the large improvements in heats

Table 5. Comparison of Mean Absolute Errors in Heats of Formation from *Ab Initio* and Semiempirical Methods (kcal/mol).

Molecule sets	Composite *ab initio* methods						PDDG		Standard NDDO			
	G2[a]	G3[b,c]	G2(MP2)[a]	BLYP[a]	B3LYP[a,d]	CBS-4[e]	PM3	MNDO	PM5	PM3	AM1	MNDO
G2-1	1.0	0.7	1.3	5.2	1.7	1.7	4.0	6.5	8.1	6.0	6.8	7.3
G2-2	1.4	0.7	1.9	8.1	2.4	3.6	3.0	5.1	4.6	4.2	5.2	7.1
G2-1 and G2-2	1.3	0.7	1.8	7.4	2.2	3.2	3.2	5.4	5.4	4.6	5.6	7.2
G3		0.9			7.2		3.2	4.6	5.1	4.4	6.2	7.3
G2-1, G2-2, and G3		0.8			4.4		3.2	5.1	5.3	4.5	5.9	7.2

[a] Ref. 12.
[b] Ref. 49.
[c] Ref. 13.
[d] Basis set is 6-311+G(3df,2p).
[e] Ref. 15.

of formation observed are a result of the PDDG function acting in concert with the eisol modification.

For hydrocarbons, PDDG/PM3 mean absolute errors are reduced by 28 and 47% relative to standard PM3 and PM5 methods, while PDDG/MNDO errors show a reduction of 36 and 9% over MNDO and AM1. Molecules consisting of carbon, hydrogen, and oxygen atoms show significant decreases in MAEs of 30% for PDDG/PM3 over both PM3 and PM5. PDDG/MNDO shows even larger improvements of 43 and 31% over MNDO and AM1, respectively. For molecules consisting of carbon, hydrogen, and nitrogen atoms, PDDG/PM3 reduces MAEs by 9 and 19% relative to PM3 and PM5, while PDDG/MNDO generates improvements of 10 and 22% over MNDO and AM1. Finally, for molecules consisting of carbon, hydrogen, nitrogen, and oxygen atoms, PDDG/PM3 MAEs are reduced by 36 and 8% over PM3 and PM5, while PDDG/MNDO shows a reduction of 63 and 48% over MNDO and AM1, respectively.

Comparison with the OM2 method was made using molecules common to both our and Thiel's evaluation sets.[41] For this limited set of 129 H, C, N, and O containing molecules, MAEs for the OM2 and PDDG/PM3 methods were found to be 3.09 and 3.31 kcal/mol. If only two molecules are removed from the set, H_2 and cubane, the MAEs are reduced to 3.08 and 2.88 kcal/mol, respectively. Overall, energetic results obtained using the PDDG methods show significant improvements over those calculated by standard semiempirical methods for a wide range of functional groups. Furthermore, no standard semiempirical method shows a similar consistency in maintaining low mean absolute errors across a number of functional groups as exhibited by the PDDG/PM3 scheme (Table 4). All PDDG/PM3 functional group errors are less than 5 kcal/mol with the exception of aromatic compounds containing nitrogen atoms, while PM3 has five functional groups with errors >5 kcal/mol (acyclic alkynes, cyclic alkanes, esters, peroxides, and amides) and PM5 shows seven groups with errors >5 kcal/mol (acyclic alkanes, acyclic alkynes, cyclic alkanes, cyclic alkenes, ethers, ketones, and peroxides).

A comparison of the PDDG methods with several composite *ab initio* schemes for computing heats of formation is shown in Table 5. Comparison was made for the G2-1, G2-2, and G3 sets of molecules. All ground-state, closed-shell, neutral molecules containing C, H, N, and O atoms were included, specifically 14, 46, and 45 molecules in the G2-1, G2-2, and G3 sets. The PDDG methods are outperformed by the G2, G3, and G2(MP2) methods, while the PDDG MAEs in heats of formation are more than competitive with those from DFT methods with a very large basis set and from CBS-4. Of course, the semiempirical calculations enjoy a speed advantage of several orders of magnitude over these alternatives. Furthermore, consideration of the size consistency of results should be made. In going from the G2 to the G3 molecule set, where the average number of heavy atoms per molecule changes from four to six, errors for B3LYP significantly expand, reflecting size consistency problems with the atom equivalents used to convert the molecular energies into heats of formation. This is a problem that appears to have been avoided in the PDDG methods by including molecules of various sizes in the training set.

Energetic results for ions and radicals are presented in Table 6. For these species, large uncertainties in the experimental data undoubtedly contribute to the significant overall errors found for all methods. For the set of 71 species, MAEs of 11.3 and 13.3 kcal/mol are found for PDDG/PM3 and PDDG/MNDO. MNDO, AM1, and PM3 give mean absolute errors of 12.5, 10.0, and 10.8 kcal/mol for the same set. In comparing the PDDG and standard NDDO techniques, the PDDG/PM3 method generates MAEs that are 4% larger than PM3, while PDDG/MNDO generates errors that are 25 and 6% larger than AM1 and MNDO. PDDG trends in the magnitude of MAEs for cations, anions, and radicals follow the trends of their NDDO precursors. For all methods, largest errors are found for radical species with MAEs of 13.0, 19.2, 15.4, 12.9, and 12.5 kcal/mol for PDDG/PM3, PDDG/MNDO, MNDO, AM1, and PM3, respectively. Although errors remain large for this class of molecules, those arising from the PDDG methods are similar to errors in the standard NDDO schemes.

Improved Self-Consistency of Energetics

The lack of self-consistency in heats of formation across different functional groups and related molecules has proven to be a prob-

Table 6. Heats of Formation of Ions and Radicals (kcal/mol).

	PDDG		Standard NDDO			
	PM3	MNDO	PM3	AM1	MNDO	Exp.
Methyl cation	256.7	247.8	256.6	252.4	243.9	262.9[b]
Ethyl cation	220.3	220.4	222.5	216.8	219.7	215.6[c]
Isopropyl cation	192.9	197.4	197.3	191.9	200.7	190.9[b]
tert-butyl cation	171.5	175.2	177.8	173.9	187.3	176.0[d]
Allyl cation	231.9	227.4	232.7	226.2	221.4	226.0[b]
Tropylium cation	222.4	222.4	221.0	210.5	207.7	203.0[b]
Benzyl cation	226.7	224.1	227.4	222.1	218.0	215.0[b]
OH_3^+	157.1	126.4	159.1	143.5	134.2	139.0[d]
CHO^+	175.4	188.9	176.9	187.5	184.9	199.0[d]
$CH_2{=}OH^+$	166.0	160.6	166.3	161.3	155.6	167.0[d]
H_3^+	192.7	247.6	215.2	223.9	259.6	264.5[b]
1-methylcyclopentyl cation	171.3	176.6	174.5	167.5	178.7	167.0[e]
Cyclohexyl cation (chair)	185.6	187.5	186.1	174.2	187.0	175.0[b]
Vinyl cation	261.0	266.7	263.9	261.5	265.7	268.9[b]
Methyleneoxonium ion	162.1	159.7	165.1	158.1	156.0	168.0[b]
NH_4^+	149.6	154.9	153.4	150.6	164.6	155.0[d]
$CH_2NH_2^+$	182.4	185.8	185.3	176.3	186.8	178.0[d]
NO_2^+	201.2	236.3	208.4	221.1	240.6	233.0[d]
NO^+	235.7	253.2	238.2	228.2	230.6	237.0[d]
Methyl radical	24.1	17.9	28.0	30.0	24.6	34.8[a]
Ethyl radical	13.0	7.0	14.5	15.5	10.5	25.0[d]
Isopropyl radical	1.6	−4.8	2.3	3.6	−1.5	22.3[b]
tert-butyl radical	−9.9	−16.3	−9.3	−6.1	−10.1	11.0[b]
OH radical	12.1	−12.8	2.8	0.6	0.2	9.3[b]
Methoxy radical	−1.9	−16.8	−10.5	−9.2	−5.8	−0.5[d]
CHO radical	−12.5	−11.5	−10.2	−2.0	−1.4	10.4[d]
NO radical	9.0	3.0	14.7	0.9	−0.5	21.6[d]
NO_2 radical	−8.3	−21.4	−1.9	−16.6	−6.1	7.9[d]
Azide radical	90.1	119.1	101.7	102.1	97.1	99.0[d]
C_2 (triplet)	222.3	202.2	231.2	221.9	231.9	200.2[a]
C_3 (carbon trimer)	196.9	186.1	206.7	212.4	220.3	196.0[a]
CH	141.2	126.0	144.8	144.6	143.3	142.4[b]
Carbene (singlet)	91.1	80.4	95.1	95.3	91.9	99.8[d]
Carbene (triplet)	67.2	62.5	71.6	77.2	73.9	92.3[a]
n-propyl radical	9.3	2.2	10.3	9.6	5.1	16.8[b]
sec-butyl radical	−3.4	−11.3	−3.2	−3.3	−6.4	17.0[b]
Vinyl radical	55.2	52.5	57.8	60.5	59.0	63.4[b]
Allyl radical	28.5	22.3	31.2	30.2	25.4	40.0[b]
Amidyl radical (NH_2)	26.6	45.4	35.2	37.7	36.4	45.1[b]
Cyanide radical	127.6	121.7	136.6	112.3	143.6	104.0[b]
Cyanomethyl radical	53.0	49.8	56.8	52.3	48.7	59.0[b]
Aminomethyl radical	22.5	15.1	26.5	23.2	22.0	38.0[b]
Methylamidyl radical	23.3	37.9	25.7	30.6	29.4	43.6[b]
Oxygen (triplet O_2)	−6.5	−59.7	−4.2	−27.7	−16.0	0.0
Cyanate radical	51.8	56.4	60.7	67.4	63.0	38.1[a]
Methoxy anion	−28.3	−29.7	−37.9	−38.5	−39.7	−33.2[b]
Ethoxy anion	−38.4	−38.7	−44.8	−45.5	−45.3	−47.5[f]
Phenol anion	−45.3	−47.6	−44.1	−41.0	−42.2	−40.5[f]
Formic acid anion	−108.1	−106.7	−110.9	−109.4	−101.6	−110.9[b]
Acetic acid anion	−121.6	−120.8	−119.6	−115.4	−110.0	−122.5[f]
Hydroxide ion	−14.6	−7.2	−17.5	−14.1	−5.8	−32.7[b]
Methyl amine anion	19.8	25.3	21.7	33.1	23.5	30.5[f]
Dimethyl amine anion	12.1	13.8	7.8	22.4	8.5	24.7[f]
Pyrrole anion	8.0	9.8	11.5	28.1	14.3	18.9[b]
Acetonitrile anion	20.6	19.2	28.5	30.8	32.0	24.1[f]
Nitromethane anion	−48.7	−41.6	−43.3	−29.2	−14.9	−26.4[f]

(continued)

Table 6. *(Continued)*

	PDDG		Standard NDDO			
	PM3	MNDO	PM3	AM1	MNDO	Exp.
Methyl anion	43.3	39.4	51.5	57.7	56.8	33.2[b]
Ethyl anion	31.3	22.8	31.7	34.5	27.4	35.1[b]
Isopropyl anion	18.6	7.1	15.1	16.9	5.0	28.2[b]
tert-butyl anion	4.9	−8.6	−0.4	2.7	−11.9	16.0[b]
Cyclopropyl anion	54.9	54.5	60.2	67.6	57.0	57.5[b]
Vinyl anion	55.1	55.7	61.7	67.8	71.1	52.9[b]
Allyl anion	20.6	16.2	25.2	27.6	26.0	29.9[b]
Diphenylmethyl anion	27.2	24.3	26.8	29.2	31.0	31.3[b]
Cyclopentadienate anion	13.6	16.2	15.9	25.2	18.9	19.6[b]
Cycloheptatrienate ion	45.9	33.7	43.1	40.3	30.3	53.3[b]
Monoacetylide ion (C_2H^-)	81.3	82.6	90.0	101.5	104.7	65.5[b]
Methyoxymethyl anion	−8.2	−16.1	−5.0	−4.7	−6.7	−2.6[b]
Phenate ion	−45.3	−47.6	−44.1	−41.0	−42.2	−39.4[c]
Cyanide ion	13.8	24.2	27.7	44.0	55.3	17.7[b]
Nitrate ion	−96.7	−98.0	−93.3	−88.8	−67.0	−73.4[b]

[a]Ref. 50.
[b]Ref. 51.
[c]Ref. 52.
[d]Ref. 4.
[e]Ref. 9.
[f]Ref. 53.

Table 7. Isomerization Enthalpies (kcal/mol).

	PDDG		Standard NDDO				
	PM3	MNDO	PM5	PM3	AM1	MNDO	Exp.[a]
(alkyne →)	5.9	4.7	4.8	6.8	2.7	2.5	1.2
(alkyne → cyclopropane)	21.6	29.5	28.9	28.0	31.4	26.9	21.8
(→ cyclopropane)	8.1	9.3	9.9	9.2	10.7	6.1	7.9
(→)	1.7	0.8	0.9	0.2	1.1	1.2	1.1
(→)	0.6	−0.8	1.0	1.7	3.3	3.7	−1.3
(→)	5.4	6.3	3.6	5.2	4.4	5.7	7.1
(→)	−7.2	−4.8	−1.6	−1.4	5.2	9.8	−5.1
(→)	5.0	4.2	3.1	5.2	3.6	4.9	2.8
$C_2H_5OH \rightarrow (CH_3)_2O$	9.1	10.2	11.1	8.5	9.5	11.8	12.3
(→)	33.1	34.7	31.7	36.1	32.6	26.7	27.2
(→)	16.2	19.5	15.5	15.0	11.9	15.6	18.3
$C_2H_5NH_2 \rightarrow (CH_3)_2NH$	6.6	6.4	9.1	3.2	7.9	7.0	6.9
Overall MAEs	2.1	2.3	2.8	3.7	4.0	2.9	

[a]References for individual molecules may be found in the supplemental data tables.
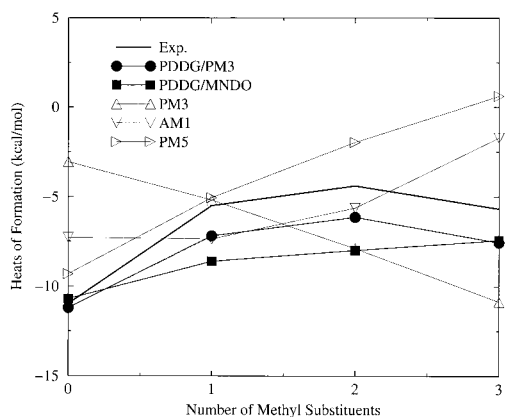
Figure 3. Amine energetic trends with increasing methyl substitution.

lem with standard NDDO methods. The seeming randomness of errors in semiempirical methods makes their untested application unreliable; thus, it is reassuring to see that the PDDG methods exhibit greater self-consistency across both different functional groups and related molecules than observed in the standard semiempirical schemes. In Table 7, a number of isomerization energies are compared, enabling evaluation of the self-consistency of energies across different functional groups. Overall, PDDG/PM3 reflects a 43% improvement over PM3 and a 25% improvement over PM5 in the calculation of isomerization energies. PDDG/MNDO generates isomerization energies that are improved by 21 and 43% over MNDO and AM1, respectively.

The chemistry of amines is of importance in the study of a large number of biochemical processes. Thus, the inability of standard NDDO methods to reproduce energetic trends with increasing methyl substitution at nitrogen centers limits their

Table 8. Heats of Formation of Molecules Containing Three-, Four-, and Five-Membered Rings (kcal/mol).
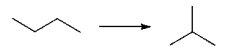
Molecules	PDDG		Standard NDDO				Exp.[a]
	PM3	MNDO	PM5	PM3	AM1	MNDO	
▷	13.50	13.17	16.85	16.27	17.78	11.22	12.73
☐	−3.23	−5.40	−3.28	−3.79	−0.99	−11.89	6.78
⬠	−20.83	−21.55	−24.86	−23.78	−28.77	−30.37	−18.44
▷⟋	2.63	2.15	3.63	5.73	6.47	1.97	1.10
☐⟋	−16.00	−19.48	−15.90	−15.79	−13.44	−21.05	−6.29
⬠⟋	−33.23	−35.27	−36.45	−34.89	−40.31	−36.59	−30.34
▷|	62.69	68.25	66.53	68.17	74.81	68.30	66.20
☐|	35.95	33.93	35.89	37.67	45.76	31.03	37.45
⬠	4.41	4.51	2.66	3.02	2.98	−0.26	8.23
▷N	30.72	30.13	31.59	31.61	33.13	25.09	30.24
⬠N	−6.46	−6.34	−10.62	−11.38	−10.43	−15.84	−0.81
▷O	−10.47	−8.99	−8.32	−8.13	−8.96	−15.53	−12.58
☐O	−25.46	−27.38	−23.87	−26.71	−25.55	−37.16	−19.25
⬠O	−47.72	−47.28	−51.66	−51.29	−58.40	−59.28	−44.02
⬠	30.64	33.82	33.00	31.76	37.06	32.09	31.94
⬠N	24.66	29.36	32.68	27.11	39.88	32.45	25.88
⬠O	−12.03	−8.99	−0.98	−4.03	2.96	−8.61	−8.30
Overall MAEs	3.56	4.17	5.24	4.84	7.94	7.90	

[a]References for individual molecules may be found in the supplemental data tables.

Table 9. Heats of Isomerization for Alkanes (kcal/mol).

Reaction	PDDG		Standard NDDO				Exp.[a]
	PM3	MNDO	PM5	PM3	AM1	MNDO	
(structure)	−2.57	−1.57	−0.56	−0.47	1.76	2.93	−2.04
(structure)	−7.15	−4.82	−1.58	−1.35	5.20	9.81	−5.06
(structure)	−2.14	−0.73	0.12	0.26	2.68	4.46	−1.84
(structure)	−2.07	−0.85	0.19	0.23	2.63	4.57	−1.82
(structure)	−1.42	1.21	0.97	1.15	3.58	6.15	−1.18
(structure)	−3.08	−0.11	0.95	0.68	5.99	11.41	−2.54
(structure)	−5.69	−1.85	−0.03	0.21	7.22	13.10	−4.53
Overall MAEs	0.73	1.47	2.72	2.82	6.87	10.21	

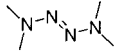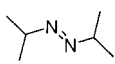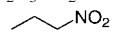[a]References for individual molecules may be found in the supplemental data tables.
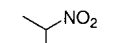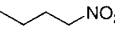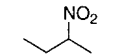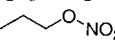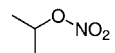
applicability to this broad field. Energetic trends across a group of methyl substituted amines were examined with results presented in Figure 3. Only with the PDDG methods are the experimental trends qualitatively reproduced. For the PDDG/ PM3 method, the ordering of experimental energies is exactly reproduced, while PDDG/MNDO only errs in predicting too little stabilization in going from dimethylamine to trimethylamine. The standard NDDO methods all predict incorrect qualitative trends. PM3 predicts linear stabilization with increasing methyl group substitution, while PM5 shows a linear destabilization. AM1 predicts energy orderings almost exactly opposite to the experimental results. Additional evidence of greater internal self-consistency is obtained by investigating the problematic energetics of hydrocarbon and heterocyclic small rings (Table 8). Overall, PDDG/PM3 represents a 26 and 32% improvement over PM3 and PM5, respectively, for a set of 17 molecules containing three-, four-, and five-membered rings. PDDG/MNDO realizes slightly larger improvements, 47% better than both AM1 and MNDO. Three-membered rings see significant improvement with mean absolute errors of only 1.68 and 1.45 kcal/mol for PDDG/PM3 and PDDG/MNDO compared to 3.19, 2.52, 5.11, and 2.52 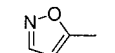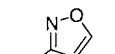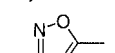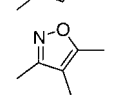kcal/mol for PM3, PM5, AM1, and MNDO, respectively. Five-membered rings benefit most from the PDDG schemes with mean absolute errors of only 2.98 and 3.27 kcal/mol for PDDG/PM3 and PDDG/ MNDO, significant improvements over PM3, PM5, AM1, and MNDO errors of 4.78, 6.45, 10.67, and 7.93 kcal/mol, respectively. Little improvement over the standard schemes is experienced for four-membered rings.

An additional systematic problem with standard semiempirical methods that is effectively corrected by use of the PDDG methodology involves the stability of branched molecules relative to their straight chain analogs (Table 9). AM1 and MNDO incorrectly predict all branched compounds to be less stable than their straight chain analogs, while PM3 and PM5 drastically reduce the energy difference between branched and normal molecules making their average errors appear smaller. The use of PDDG methods leads to large improvements in all cases with branched compounds being more stable than their normal analogs with only a single exception for PDDG/MNDO. For example, absolute errors in the isomerization energy for the conversion of *n*-pentane to neopentane are only 2.09 and 0.24 kcal/mol from PDDG/PM3 and PDDG/MNDO, while for the standard NDDO methods, errors of 3.48, 3.71, 10.26, and 14.87 kcal/mol are found for PM5, PM3, AM1, and MNDO, respectively. MAEs for PDDG/PM3 over the set of seven compounds are reduced by 74 and 73% relative to PM3 and PM5, respectively. Additionally, the experimental energetic ordering is very well reproduced by the PDDG/PM3 method. Even more impressive improvements are introduced by the PDDG/MNDO method relative to AM1 and MNDO where mean absolute errors were reduced by 79 and 86%.

The zero-differential overlap approximation as implemented in NDDO methods leads to underestimation of repulsion in the treatment of molecules with adjacent lone-pair electrons. As previously noted, this issue is especially evident in the case of molecules containing NN and NO bonds. The PDDG/PM3 method provides a 9% increase in MAEs for molecules with NN bonds relative to PM3, yet a 32% improvement over the more recent PM5 method (Table 10). The PDDG/MNDO method enjoys overall decreases in MAEs of 20 and 18% over AM1 and MNDO, respectively. For NO bonds, greater consistency in the improvement introduced by the PDDG methods is observed with PDDG/PM3 realizing 31 and 20% improvements over PM3 and PM5, while PDDG/MNDO is a surprising 62 and 73% better than AM1 and MNDO, respectively.

Table 10. Heats of Formation of Molecules Containing NN and NO Bonds (kcal/mol).

Molecule	PDDG		Standard NDDO				Exp.[a]
	PM3	MNDO	PM5	PM3	AM1	MNDO	
$H_2N—NH_2$	18.17	16.93	16.57	20.67	13.68	14.21	22.79
$CH_3NH—NH_2$	19.41	17.85	19.12	17.90	17.33	14.47	22.60
$CH_3NH—NHCH_3$	21.57	21.06	22.45	16.46	24.02	15.10	21.99
$(CH_3)_2N—NH_2$	18.62	19.31	21.49	15.11	23.97	18.14	20.06
$CH_2{=}N—N{=}CH_2$	54.98	51.65	56.47	57.00	58.16	46.19	58.00
(tetrazene structure)	68.96	57.56	76.84	62.28	88.59	47.84	64.70
CH_3N_3	63.46	83.57	72.95	70.56	77.64	67.27	67.00
trans-HN=NH	29.67	45.29	42.81	37.75	31.55	31.85	36.00
(azoisopropane structure)	4.13	9.77	16.74	6.06	14.76	2.64	8.55
CH_3NO_2	−18.83	−13.54	−11.57	−15.94	−9.93	3.32	−17.86
$C_2H_5NO_2$	−23.23	−19.75	−18.50	−21.40	−16.77	−3.31	−24.38
(propyl nitro structure)	−28.35	−25.78	−23.98	−26.77	−23.60	−8.12	−29.98
(isopropyl nitro structure)	−30.36	−26.58	−25.73	−27.10	−21.63	−6.34	−33.21
(butyl nitro structure)	−33.51	−31.79	−29.47	−32.14	−30.41	−12.83	−34.39
(sec-butyl nitro structure)	−34.79	−31.27	−31.00	−31.82	−28.16	−10.11	−39.11
$HONO_2$	−39.39	−38.34	−30.26	−38.01	−37.46	−17.54	−32.10
CH_3ONO_2	−34.97	−33.91	−28.74	−32.42	−31.32	−12.43	−29.11
$C_2H_5ONO_2$	−40.44	−40.05	−34.44	−36.38	−37.27	−17.90	−36.83
(propyl nitrate structure)	−45.45	−46.17	−39.87	−41.62	−44.02	−22.61	−41.56
(isopropyl nitrate structure)	−49.47	−47.54	−42.22	−42.10	−42.13	−19.71	−45.64
(isoxazole structure 1)	16.82	6.74	18.54	26.13	35.35	8.18	3.60
(isoxazole structure 2)	17.28	8.21	18.96	25.69	34.99	7.50	4.80
(isoxazole structure 3)	8.09	−4.85	9.93	16.74	27.44	−3.54	−4.30
(isoxazole structure 4)	−0.32	−15.51	1.72	7.42	19.53	−13.18	−4.80
NN MAEs	3.47	5.87	5.13	3.17	7.30	7.17	
NO MAEs	5.26	4.59	6.55	7.58	12.03	16.87	

[a]References for individual molecules may be found in the supplemental data tables.
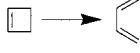
Activation, Conformer, and Intermolecular Interaction Energies

The PDDG methods share an identical formalism for calculating electronic energies with AM1, PM3, and MNDO. As such, they are also subject to issues that arise from deficiencies in the formalism. Two of the most significant omissions in the NDDO formalism are the lack of electron correlation and orbital orthogonality (no off-diagonal overlap terms in the secular determinant). Electron correlation, due to its bond-based nature, has most likely been at least partially accounted for via the PDDG scheme, as evidenced by examining the energies of branched compounds relative to their straight chain analogs. Orbital orthogonality, however, is not accounted for within the PDDG formalism, and is a likely culprit for difficulties in treating activation energies, conformer energetics, and intermolecular interaction energies.

Activation energies for five pericyclic reactions are presented in Table 11. The experimental results are almost uniformly overestimated, and the AM1, PM3, and PDDG/PM3 predictions have similar overall quality for this small set of reactions. Further study

Table 11. Activation Barriers (kcal/mol).

	PDDG		Standard NDDO			
	PM3	MNDO	PM3	AM1	MNDO	Exp.
	29.97	41.13	27.05	23.77		25.1[a]
	45.59	42.83	41.81	37.85	41.43	33.3[b]
	39.74	36.33	35.06	32.02	39.81	25.4[c]/30.6[d]/34.1[e]
	41.33	44.75	40.62	35.33	49.78	32.9[f]
	18.47	40.21	25.70	22.06	38.96	4–5[g]

[a]Ref. 54.
[b]Ref. 9.
[c]Ref. 55.
[d]Ref. 56.
[e]Ref. 57.
[f]Ref. 58.
[g]Ref. 59. *Ab initio* estimates range from 3–6 kcal/mol.

is needed, particularly to assess the alternative methods' abilities to reproduce substituent effects on the activation energies and to provide detailed analysis of transition structures.

Conformational energetics for a wide range of isomers were evaluated and results are shown in Table 12. Qualitatively, all the semiempirical methods show significant deficiencies. In about 30% of the conformer comparisons investigated using PDDG/PM3, PM3, and AM1, the lower energy conformer is incorrectly predicted. MNDO and PDDG/MNDO perform better with only five and six incorrect energy orderings, respectively. Often the target energy differences are small, but the neglect of the off-diagonal overlaps in NDDO methods is problematic for proper evaluation of rotational barriers and for the related conformational energetics.[41] The marginal treatment of conformer energetics by NDDO theories finds company with many molecular mechanics methods, as evidenced by a recent comparison of various force field, experimental, and *ab initio* data.[42] For improved conformer energetics, merger of the PDDG approach with the OM2 method[22] might be fruitful.

The treatment of intermolecular interactions has proven to be a difficult task for semiempirical methods, as illustrated by a comparison of calculated interaction energies to those from high-level *ab initio* data (Table 13). Semiempirical hydrogen bond energies are typically half the strength of the corresponding *ab initio* values. MNDO predicts virtually no stabilization from hydrogen bonding, and its only significant interaction energies arise from hydrogen bonds to charged species ($H_3(CH_3)N^+$ and $CH_3CO_2^-$). AM1 and PM3 show similar levels of stabilization with two exceptions: hydrogen bonds to $HCONH_2$, and interactions between water and methanol. In both cases PM3 predicts a stabilization energy less than half the AM1 value. The PDDG/MNDO method shows similar behavior

to MNDO with slightly larger interaction energies, while PDDG/PM3 is an improvement over PM3 in the treatment of intermolecular interactions. PDDG/PM3 provides interaction energies on par with AM1 values, and is the only semiempirical method studied to correctly predict stabilization in all three geometries of the benzene dimer. For the three benzene dimers (parallel, slipped-parallel, and T-shaped) it has been noted that the addition of electron correlation into high level *ab initio* calculations leads to significant increases in attraction, providing further evidence that the PDDG/PM3 method includes electron correlation in a more reliable fashion than other semiempirical methods.[43,44] As π-stacking interactions are increasingly found to be of importance in fields ranging from biochemistry to material science, the qualitatively correct treatment of the benzene dimer by a semiempirical method is a step forward.

Interaction energies for a series of water dimers are illustrated in Figure 4, and reveal the difficulties present in trying to reproduce weak intermolecular interactions. Structures for the nine dimers were fully optimized starting from MP2/6-31G** optimized geometries. MNDO finds little stabilization to arise from any of the dimer geometries. AM1, which was parameterized to reproduce the experimental energy of the water dimer does well in getting a most favorable interaction energy of −5.46 kcal/mol, very close to the *ab initio* value of −5.36 kcal/mol; however, the geometry of interaction is incorrect. In the parameterization of PM3, effort appears to have been made to simultaneously reproduce the geometry of the water dimer and reproduce energetic data. This is seen as PM3 finds nearly isoenergetically weak minima for all nine structures, with the exception of **1** and **2**. Furthermore, it has been demonstrated that the geometry of water dimer **1** is generated due to an artifact in the PM3 CRF.[35] PDDG/

Table 12. Conformer Energetics (kcal/mol).

	PDDG		Standard NDDO			
	PM3[c]	MNDO	PM3[c]	AM1	MNDO	Exp.
Ethane *anti* vs. *eclip.*	1.14	2.16	1.43	1.25	1.01	2.9[a]
Propene *anti* vs. *eclip.*	0.69	0.13	0.67	0.55	0.18	2[b]
Methylcyclohexane *ax.* vs. *eq.*	0.91	4.14	1.14	1.44	1.04	1.75[b]
Ethylene *orthog.* vs. *planar*	65.31	59.14	65.32	65.93	62.47	65[a]
Methylamine *anti* vs. *eclip.*	1.30	0.81	1.18	1.26	1.08	2.0[a]
Formamide *anti* vs. *eclip.*	9.88	8.90	9.46	15.09	7.05	~20[a]
Ammonia *planar* vs. *pyr.*	8.79	1.61	9.98	4.24	11.58	~6[a]
Methanol *anti* vs. *eclip.*	1.48	0.37	0.90	1.04	0.74	1.1[a]
Acrolein *cis* vs. *trans*	0.83	0.66	0.40	0.22	−0.44	1.70[b]
1,3-butadiene *gauche* vs. s-*trans*	0.73	1.31	0.73	0.77	0.30	2.89[b]
n-butane to *gauche*	0.23	1.24	0.55	0.71	0.66	0.67[b]
n-butane to *gauche eclip.*	2.10	2.80	1.68	1.53	1.49	3.4
n-butane to *eclip.*	3.85	5.51	4.00	3.28	3.23	4–6[a]
2,3-dimethylbutane vs. C_2	−0.26	0.58	0.83	0.05	−1.28	0.05[b]
Cyclohexanol *ax.* vs. *eq.*	0.79	0.18	1.05	1.27	0.21	0.58[b]
Cyclohexylamine *ax.* vs. *eq.*	2.42	2.79	1.37	0.82	0.28	1.15[b]
Ether *tt* vs. *tg*	−1.86	0.49	−1.08	−0.35	1.00	1.14[b]
Ethyl methyl ether *anti* vs. *gauche*	−1.80	0.43	−1.04	−0.37	0.96	1.5[b]
Methyl vinyl ether *cis* vs. *skew*	2.45	0.50	1.66	2.06	−0.61	1.7[b]
N-methylacetamide *Z* vs. *E*	−0.87	0.61	−0.32	−0.29	0.91	2.3[b]
N-methylformamide *Z* vs. *E*	−1.42	−1.76	−1.15	−0.80	−0.47	1.4[b]
Glyoxylic acid *tE* vs. *tZ*	0.01	1.84	0.00	4.30	−0.27	1.2[b]
Methylacetate *cis* vs. *trans*	0.43	1.09	1.16	5.32	5.22	8.5[b]
Methylformate *cis* vs. *trans*	1.83	0.17	1.86	5.59	2.89	4.75[b]
Ethanol *anti* vs. *gauche*	−2.13	−2.20	−1.87	−1.55	0.37	0.12[b]
Formic acid *Z* vs. *E*	3.30	2.50	4.29	7.42	3.74	3.90[b]
2-methyl-1,3-butadiene *gauche* vs. s-*trans*	−0.07	−1.34	−0.39	−0.12	−1.74	2.65[b]
2-propanol HCOH *anti* vs. *gauche*	−2.04	−1.77	−1.81	−1.50	0.33	0.28[b]
N-methyl piperidine *ax.* vs. *eq.*	−2.61	0.14	−1.34	−1.41	1.30	3.15[b]
2-methoxytetrahydropyran *ax.* vs. *eq.*	−0.07	0.07	0.55	2.53	1.96	1.05[b]
Piperidine *ax.* vs. *eq.*	−2.99	−3.10	−2.25	−2.73	0.49	0.53[b]

[a]Ref. 6.
[b]Ref. 42.
[c]The MOPAC MMOK correction[38] was not applied.

MNDO generates weak interaction energies, although two minima were found in addition to those identified by MNDO. PDDG/PM3 generates the most accurate interaction energies for the water dimers, although no minima were identified corresponding to structures **2** and **4.** The ability to reproduce *ab initio* energies over a number of water dimer geometries should make PDDG/PM3 of interest in investigating the properties of water clusters and in hybrid QM/MM techniques.

Recently, questions have been raised about the nature of semiempirical interaction energies.[45] Energy decomposition results find that coulombic energies of interaction are generally repulsive, and that stabilization is primarily a result of the deformation terms as defined by van der Vaart and Merz.[46,47] This stands in marked contrast to decomposition results of *ab initio* calculations, which reveal significant electrostatic stabilization. Thus, there is some question whether interaction energies are capable of being accurately reproduced by the formalism underlying semiempirical methods.

Treatment of Other QM Observables

Ionization potentials, dipole moments, and molecular geometries arising from the standard NDDO methods are generally acceptable for most problems. Thus, during parameterization of the PDDG methods, effort was made only to see no significant degradation in results for these observables. The design of the PDDG function and desire to examine only local minima to the standard semiempirical parameter sets suggests that these terms should not be significantly altered from their standard values. This was observed in examining ionization potentials (Table 14), dipole moments (Table 15), and optimized molecular geometries (Table 16). Mean absolute errors over a set of 97 ionization potentials reveal PDDG/PM3 is 3% more accurate than PM3 and 12% less accurate than PM5. PDDG/MNDO is 19% less accurate than AM1 but 10% more accurate than MNDO. All the semiempirical methods presented do a satisfactory job of reproducing experimental ionization potential data. Over the set of 47 dipole moments, MAEs of 0.24
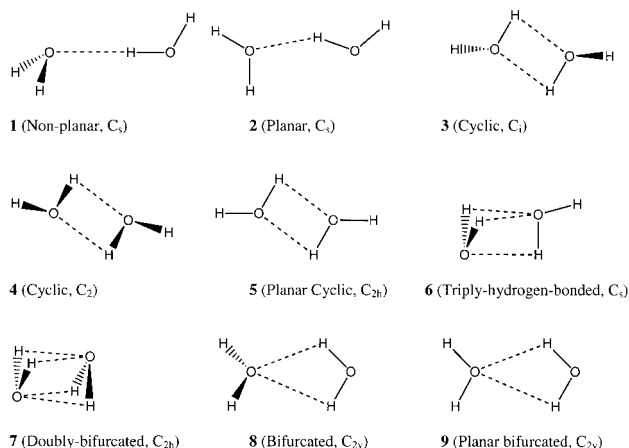
Table 13. Intermolecular Interaction Energies for Hydrogen Bond and van der Waals Complexes (kcal/mol).

	PDDG		Standard NDDO			
	PM3	MNDO	PM3	AM1	MNDO	*Ab initio*
$H_2O \cdots H_2O$	−3.72	−1.04	−3.50	−5.41	−0.96	−4.80[a]
$H_2O \cdots MeOH$	−2.63	−0.74	−1.65	−5.09	−0.09	−4.90[a]
$H_2O \cdots Me_2O$	−3.37	−0.39	−1.72	−2.24	−0.77	−5.51[a]
$MeOH \cdots MeOH$	−3.15	−0.72	−1.69	−2.73	−0.62	−5.45[a]
$H_2O \cdots H_2CO$	−1.75	−0.74	−1.51	−2.74	−0.78	−5.17[a]
$HCOOH \cdots HCOOH$	−3.95	−1.87	−8.58	−6.43	−1.69	−13.93[a]
$NH_3 \cdots NH_3$	−1.33	−0.76	−1.37	−1.39	−0.64	−2.94[a]
$NH_3 \cdots H_2O$	−1.76	−1.12	−1.51	−1.99	−0.57	−6.36[a]
$HCONH_2 \cdots H_2O$	−3.64	−1.07	−2.20	−7.54	−0.98	−8.88[a]
$HCONH_2 \cdots HCONH_2$	−5.44	−1.43	−3.25	−8.13	−1.73	−13.55[a]
$H_2O \cdots H_3(CH_3)N^+$	−13.23	−7.53	−12.68	−13.06	−6.93	−18.76[b]
$H_2O \cdots CH_3CO_2^-$	−16.15	−7.10	−15.78	−15.90	−6.80	−19.22[b]
C_6H_6 parallel	−3.29	0.08	0.16	0.23	0.05	−1.21[a]
C_6H_6 slipped-parallel	−0.96	0.06	0.07	0.05	0.05	−2.01[a]
C_6H_6 T-shaped	−1.74	1.90	−0.54	0.14	0.03	−2.34[a]

[a]Estimated CCSD(T) interaction energies at the basis set limit, from ref. 43.
[b]RHF/6-31G* basis set superposition error corrected values, from ref. 60.

and 0.22 Debye are found for PDDG/PM3 and PDDG/MNDO, while errors of 0.36, 0.31, 0.24, and 0.27 kcal/mol are found for PM5, MNDO, AM1, and PM3. Thus, the PDDG methods are a

1 (Non-planar, C_s)   **2** (Planar, C_s)   **3** (Cyclic, C_i)

4 (Cyclic, C_2)   **5** (Planar Cyclic, C_{2h})   **6** (Triply-hydrogen-bonded, C_s)

7 (Doubly-bifurcated, C_{2h})   **8** (Bifurcated, C_{2v})   **9** (Planar bifurcated, C_{2v})

	PDDG[a]		Standard NDDO[a]			
Structure	PM3	MNDO	MNDO	AM1	PM3	MP4/6-31+G**[b]
1	-3.72	*9*	*9*	*7*	-3.50	-5.36
2	*1*	*9*	*8/9*	*8/9*	-2.81	-4.62
3	-2.90	-0.78	*8/9*	*8*	-1.85	-4.36
4	*7*	*6*	*8/9*	*7*	-1.93	-4.04
5	-2.46	-0.77	*8/9*	-3.6	-1.52	-3.99
6	-3.31	-0.50	-0.50	-5.46	-1.93	-3.52
7	-3.37	-0.57	-0.52	-5.36	-1.71	-1.61
8	-2.64	-1.08	-0.98	-5.00	-1.99	-3.57
9	-2.45	-1.06	-0.91	-4.61	-1.81	-2.52

[a]Fully optimized semiempirical geometries where structure numbers for molecules that collapse into an alternative minima are given in italics. Optimized structures with an identical interaction mode as structures 8 and 9, but HO-OH dihedral angles between 0° and 90°, are indicated as *8/9*.
[b]Ref. 61.

Figure 4. Water dimer geometries and interaction energies.

slight improvement over the standard schemes in the calculation of dipole moments. Finally, for molecular geometries, MAEs for the PDDG/PM3 method lie within 6, 9, and 3% of those from PM3 for bond lengths, bond angles, and dihedral angles. PDDG/MNDO exhibits mean absolute errors for bond lengths and bond angles that are within 15 and 14% of MNDO calculated geometries, while nearly identical dihedral angle MAEs were found. Mean absolute errors in bond length per atom pair are shown in Table 17, and reveal that use of the PDDG methods does not introduce significant changes compared to the standard NDDO schemes. In atom pairs for which a reasonable amount of experimental data was found (i.e., six molecules or more), largest errors of 0.018 and 0.019 Å were found for PDDG/PM3 and PDDG/MNDO. This compares favorably with the largest errors from MNDO, AM1, and PM3, 0.016, 0.020, and 0.016 Å, respectively. Thus, molecular geometries from the PDDG methods are very similar to their formalism precursors, and do not appear to exhibit any adverse effects from having CRF Gaussians that may be located near bonding distances.

Scope and Limitations of the PDDG Methods

It is useful to highlight the advantages and limitations of the proposed PDDG methods in addition to presenting information regarding their implementation into standard semiempirical software packages. Only two changes must be made to existing semiempirical codes to facilitate use of PDDG/PM3 and PDDG/MNDO. First, the PDDG function must be added to the CRF. As both the PDDG function and existing semiempirical CRFs loop over all distinct atom pairs this is trivially accomplished. Second, the PDDG/PM3 and PDDG/MNDO parameters must be added. A detailed step-by-step example of the addition of the

Table 14. Selected Ionization Potentials from PDDG and Standard NDDO Methods (eV).

	PDDG		Standard NDDO				
	PM3	MNDO	PM5	PM3	AM1	MNDO	Exp.[a]
Hydrogen	16.4	16.0	13.8	16.1	14.9	15.7	15.4
Methane	13.7	13.8	12.7	13.6	13.3	13.9	13.6
Acetylene	11.7	11.4	11.5	11.6	11.5	11.0	11.4
Ethylene	10.7	10.4	10.5	10.6	10.6	10.2	10.5
Ethane	11.9	12.5	11.2	12.0	11.8	12.7	12.0
Allene	10.2	10.3	10.0	10.2	10.1	10.0	10.1
Cyclopropene	9.8	10.0	9.6	9.9	9.8	9.9	9.9
Propyne	10.9	10.9	10.6	10.9	10.7	10.7	10.4
Cyclopropane	11.8	11.4	11.2	11.8	11.5	11.4	11.0
Propene	10.1	10.0	9.8	10.1	10.0	10.0	9.9
n-propane	11.3	12.1	10.8	11.5	11.3	12.3	11.5
1,3-butadiyne	10.6	10.5	10.4	10.5	10.4	10.0	10.2
1,2-butadiene	9.7	9.9	9.4	9.7	9.7	9.8	9.2
1-butyne	10.8	10.8	10.5	10.8	10.6	10.7	10.2
2-butyne	10.3	10.5	10.0	10.3	10.2	10.5	9.6
Cyclobutene	9.7	9.8	9.6	9.8	9.7	9.8	9.4
1,4-pentadiene	9.5	9.4	9.3	9.5	9.3	9.1	9.1
1-butene	10.1	10.0	9.9	10.2	10.0	10.0	9.7
Cyclobutane	10.8	11.6	10.4	11.0	11.0	11.8	10.7
n-butane	11.2	12.0	10.6	11.4	11.2	12.2	11.2
Isobutane	11.5	11.9	10.8	11.6	11.3	12.1	11.4
Cyclopentadiene	9.2	9.2	9.0	9.2	9.1	9.0	8.6
Cyclopentene	9.4	9.7	9.2	9.5	9.4	9.7	9.2
1-pentene	10.1	10.0	9.9	10.1	10.0	10.0	7.9
2-methyl-1-butene	9.8	9.8	9.5	9.9	9.7	9.8	7.4
2-methyl-2-butene	9.2	9.5	9.0	9.3	9.2	9.6	8.7
3-methyl-1-butene	10.2	10.0	10.0	10.2	10.0	10.0	9.6
Cyclopentane	10.8	11.8	10.4	11.1	11.0	12.1	10.5
Isopentane	11.3	11.7	10.7	11.5	11.2	12.0	10.3
Neopentane	11.9	11.8	11.2	12.1	11.5	12.1	11.3
n-pentane	11.1	11.8	10.5	11.3	11.1	12.2	10.3
Benzene	9.7	9.5	9.5	9.8	9.7	9.4	9.3
Cyclohexene	9.5	9.7	9.3	9.6	9.5	9.8	10.3
Cyclohexane	11.1	11.4	10.5	11.3	10.9	11.7	10.3
Cycloheptatriene	9.0	8.9	8.7	8.9	8.7	8.7	8.5
Toluene	9.4	9.3	9.1	9.4	9.3	9.3	8.8
Ethylbenzene	9.3	9.3	9.2	9.4	9.3	9.3	8.8
Cyclo[2.2.2] octane	10.7	11.1	10.1	11.0	10.6	11.4	9.5
Napthalene	8.8	8.8	8.6	8.8	8.7	8.6	8.2
Adamantane	10.6	10.9	10.0	10.8	10.4	11.3	9.6
Anthracene	8.3	8.3	8.1	8.2	8.1	8.0	8.2
Water	12.6	11.5	11.7	12.3	12.5	12.2	12.6
Carbon monoxide	13.4	13.9	12.2	13.0	13.3	13.4	14.0
Formaldehyde	10.6	10.9	10.4	10.6	10.8	11.0	10.1
Methanol	11.1	10.9	10.6	11.1	11.1	11.4	11.0
Ketene	9.6	9.6	9.7	9.5	9.6	9.3	9.6
Acetaldehyde	10.7	10.7	10.4	10.7	10.7	10.9	10.2
Ethylene oxide	11.3	11.1	10.7	11.3	11.3	11.5	10.6
Ethanol	10.8	10.8	10.3	10.9	10.9	11.3	10.6
Methyl ether	10.5	10.5	10.1	10.7	10.6	11.0	10.0
Acetone	10.8	10.6	10.5	10.8	10.7	10.8	9.7
Furan	9.4	9.4	9.4	9.4	9.3	9.1	8.9
2-butenal	10.4	10.3	10.3	10.4	10.4	10.4	9.9
Butanal	10.6	10.6	10.3	10.6	10.6	10.8	9.8
Ether	10.3	10.3	9.9	10.5	10.4	10.9	9.6
3-pentanol	10.9	10.6	10.3	11.0	10.8	11.1	10.2
Benzaldehyde	10.0	9.8	9.9	10.1	10.0	9.7	9.7

(continued)

Table 14. *(Continued)*

	PDDG		Standard NDDO				
	PM3	MNDO	PM5	PM3	AM1	MNDO	Exp.[a]
Anisole	9.0	8.9	8.8	9.1	9.0	8.8	8.4
Carbon dioxide	13.0	12.4	12.9	12.7	13.2	12.8	13.8
Formic acid	11.7	11.5	11.5	11.6	11.8	11.7	11.5
Glyoxal	10.6	10.7	10.5	10.6	10.7	10.8	10.6
Acetic acid	11.5	11.3	11.3	11.4	11.6	11.6	10.8
Methyl formate	11.4	11.2	11.2	11.3	11.6	11.6	11.0
Dimethyl peroxide	10.4	10.1	9.9	10.8	10.9	10.5	10.6
β-propiolactone	11.3	11.1	11.2	11.3	11.4	11.4	10.6
Propionic acid	11.4	11.2	11.2	11.4	11.5	11.5	10.5
Methyl acetate	11.3	11.1	11.1	11.3	11.4	11.5	10.6
Acetylacetone	10.8	10.5	10.5	10.5	10.4	10.6	8.4
Benzoic acid	10.1	9.9	10.0	10.1	10.1	9.8	9.8
Ozone	13.0	12.3	10.0	12.7	13.1	10.8	12.8
Maleic anhydride	11.8	11.6	11.8	11.7	12.0	11.7	10.8
Oxalic acid	11.7	11.4	11.4	11.6	11.8	11.7	11.2
Ammonia	9.8	10.9	10.2	9.7	10.4	11.2	10.9
Hydrogen cyanide	12.7	14.2	13.0	12.6	13.7	13.4	13.6
Methylamine	9.4	10.4	9.5	9.4	9.8	10.6	9.6
Acetonitrile	12.4	13.3	12.3	12.3	12.5	12.8	12.2
Isocyanomethane	12.0	12.6	11.6	11.7	12.0	12.2	11.3
Azirane	9.8	10.7	9.9	9.9	10.3	10.7	9.9
Ethylamine	9.3	10.3	9.4	9.4	9.7	10.5	9.5
Dimethylamine	9.1	10.0	9.1	9.2	9.4	10.0	8.9
Acrylonitrile	11.0	11.1	10.9	10.9	10.9	10.6	10.9
Propionitrile	12.0	12.9	11.7	12.0	12.0	12.6	11.9
Trimethylamine	8.9	9.7	8.9	9.1	9.1	9.6	8.5
Pyrrole	8.9	8.8	8.6	8.9	8.7	8.6	8.2
Pyridine	10.1	9.9	9.9	10.1	9.9	9.7	9.7
Aniline	8.6	8.6	8.5	8.6	8.5	8.8	7.7
Phenyl cyanide	10.1	10.1	9.9	10.1	10.0	9.8	9.7
Isocyanic acid	10.8	11.6	11.2	10.6	11.2	11.1	11.6
Nitromethane	12.5	10.9	11.6	12.2	12.0	11.5	11.3
Methyl nitrite	10.5	11.3	10.4	10.6	11.2	11.4	11.0
L-alanine	9.8	10.5	10.1	9.9	10.2	10.9	8.1
Nitrogen	14.0	15.7	12.9	13.8	14.3	14.9	15.6
Methylhydrazine	8.8	9.6	8.8	8.9	9.0	9.7	9.3
Cyanogen	13.1	14.3	13.4	12.9	13.3	13.2	13.4
Pyridazine	9.9	10.9	10.2	9.9	10.7	10.5	9.3
Pyrimidine	10.3	10.6	10.1	10.3	10.6	10.4	9.7
Pyrazine	10.1	10.3	9.9	10.2	10.3	10.0	9.9

[a]Ref. 7.

PDDG methods to MOPAC 6.0 has been made, which should be relatively transferable for any semiempirical algorithm library.[48] Note that no modification of the procedure for calculating electronic energies has been introduced; thus, the calculation of gradients of the energy with respect to molecular geometries remains unchanged.

The PDDG methodology is currently limited to treating molecules containing the elements H, C, N, and O. This will be addressed with continuing parameterization for additional elements including the halogen series and silicon, sulfur, and phosphorous. Because the PDDG and standard semiempirical parameter sets are similar (only local minimum of the error function were sought) existing issues with a basis in NDDO approximations that are not bond centered are most likely maintained in PDDG/PM3 and PDDG/MNDO, including activation energies and conformer energetics. Furthermore, problems in the treatment of molecular geometries such as the incorrect dihedral angles predicted for hydrogen peroxide and cyclobutane remain. Finally, both the PDDG and NDDO methods generate large MAEs for heats of formation of ions and radicals.

The most significant advantage of the PDDG schemes is that the quality and self-consistency of energetic results is vastly en-

Table 15. Selected Dipole Moments from PDDG and Standard NDDO Methods (Debye).

	PDDG		Standard NDDO				
	PM3	MNDO	PM5	PM3	AM1	MNDO	Exp.[a]
Cyclopropene	0.31	0.31	0.32	0.39	0.36	0.48	0.45
Propyne	0.63	0.38	0.49	0.36	0.40	0.12	0.78
Propene	0.35	0.19	0.26	0.21	0.21	0.06	0.37
n-propane	0.03	0.02	0.01	0.01	0.00	0.00	0.08
Bicyclobutane	0.41	0.40	0.42	0.43	0.42	0.41	0.68
Cyclobutene	0.26	0.11	0.18	0.15	0.17	0.08	0.13
Cyclopentadiene	0.64	0.35	0.49	0.53	0.53	0.18	0.42
Cyclopentene	0.24	0.14	0.15	0.15	0.17	0.06	0.20
Fulvene	0.55	0.44	0.43	0.66	0.69	0.70	0.42
Toluene	0.43	0.27	0.32	0.26	0.27	0.05	0.36
Water	1.84	1.88	1.93	1.74	1.86	1.78	1.85
Carbon monoxide	0.23	0.13	0.30	0.19	0.06	0.20	0.11
Formaldehyde	2.33	2.21	2.63	2.16	2.32	2.17	2.33
Methanol	1.62	1.71	1.73	1.49	1.62	1.48	1.70
Ketene	1.11	1.22	2.06	1.06	1.35	1.04	1.42
Acetaldehyde	2.71	2.46	3.22	2.54	2.69	2.38	2.69
Ethylene oxide	1.84	2.00	2.44	1.77	1.91	1.92	1.89
Ethanol	1.57	1.62	1.70	1.45	1.55	1.40	1.69
Methyl ether	1.35	1.59	1.57	1.25	1.43	1.27	1.30
Acetone	3.03	2.67	3.57	2.78	2.92	2.52	2.88
Furan	0.20	0.55	1.00	0.22	0.49	0.42	0.66
Ether	1.24	1.37	1.41	1.15	1.25	1.10	1.15
Phenol	1.18	1.25	1.40	1.15	1.23	1.17	1.45
Anisole	1.22	1.35	1.39	1.08	1.25	1.07	1.38
Formic acid	1.63	1.46	1.86	1.51	1.48	1.49	1.41
Acetic acid	1.97	1.71	2.26	1.84	1.89	1.68	1.74
Methyl formate	1.71	1.60	1.98	1.59	1.51	1.63	1.77
Propionic acid	1.88	1.65	2.18	1.80	1.84	1.64	1.75
Methyl acetate	1.97	1.73	2.19	1.83	1.78	1.75	1.72
Ozone	1.70	1.01	0.00	1.71	1.20	0.00	0.53
Ammonia	1.70	1.66	2.09	1.55	1.85	1.75	1.47
Hydrogen cyanide	2.85	3.00	2.67	2.70	2.36	2.50	2.98
Methylamine	1.55	1.53	1.74	1.40	1.50	1.48	1.31
Acetonitrile	3.60	3.57	3.42	3.21	2.89	2.63	3.92
Isocyanomethane	3.97	2.70	3.62	3.69	2.83	2.17	3.85
Azirane	1.80	2.07	2.03	1.70	1.75	1.75	1.90
Ethylamine	1.50	1.46	1.68	1.36	1.47	1.42	1.22
Dimethylamine	1.42	1.35	1.47	1.27	1.23	1.17	1.03
Acrylonitrile	3.55	3.77	3.53	3.25	3.01	2.97	3.87
Trimethylamine	1.32	1.26	1.29	1.15	1.02	0.75	0.61
Pyrrole	2.09	1.71	1.85	2.18	1.95	1.81	1.74
Pyridine	2.05	2.34	2.28	1.93	1.98	1.97	2.22
Aniline	1.35	1.37	1.61	1.30	1.54	1.46	1.53
Formamide	3.27	3.27	4.25	3.11	3.70	3.36	3.73
Dimethyl-formamide	3.31	3.45	4.39	3.07	3.69	3.19	3.82
Nitrous acid	2.31	2.23	4.39	2.09	2.31	2.28	1.86
Methylhydrazine	0.57	0.34	0.54	0.27	0.72	0.45	1.66

[a]Ref. 7.

hanced over the set of molecules that can be studied. This improvement in energetics is obtained without an increase in the required computation time and as part of a fully general semiempirical method. The energetics of an overwhelming majority of functional groups in addition to problematic molecules with rings and multiple heteroatoms are obtained consistently better with the PDDG methods as are isomerization energies and trends across groups of related molecules. The PDDG methodology provides a template, embodied in the PDDG/PM3 and PDDG/MNDO methods, for introducing significant improvement in any NDDO-based

Table 16. Mean Absolute Errors in Molecular Geometries.

Geometric parameters	PDDG		Standard NDDO			No. mol.
	PM3	MNDO	PM3	AM1	MNDO	
Bond lengths (Å)	0.016	0.020	0.015	0.020	0.017	153
Bond angles (deg.)	2.3	3.5	2.1	2.0	3.0	54
Dihedral angles (deg.)	29.0	26.7	28.0	28.7	26.6	6

semiempirical treatment with the addition of only a small number of parameters per element.

Conclusions

Addition of the PDDG function and reparameterization of the standard semiempirical parameter set using an alternative approach for determining the electronic energy of formation of a gaseous atom has lead to the development of the PDDG/PM3 and PDDG/MNDO methods. The PDDG methods are significant improvements over the standard semiempirical methods (AM1, PM3, PM5, and MNDO) with no increase in required computation time. For a set of 622 molecules containing H, C, N, and O atoms, MAEs in heats of formation of only 3.2 and 5.2 kcal/mol are obtained with PDDG/PM3 and PDDG/MNDO. The errors with PDDG/PM3 are not only smaller than those of standard semiempirical methods, PM3 (4.4), PM5 (4.8), AM1 (6.7), and MNDO (8.4), but of density functional theories as well (Table 5). The PDDG/MNDO improvement over AM1 is also notable because PDDG/MNDO uses a simpler core repulsion formula. In addition, greater internal consistency in the semiempirical results has been introduced, as evidenced by the improved treatment of the stability of branched hydrocarbons, the energetics of rings and systems with multiple heteroatoms, and the energetic trends for substituted amines. However, little or no improvement was observed in situations where the underlying electronic semiempirical formalism is known to be deficient including for conformational energetics. Improvements are also desirable for interaction energies, although

PDDG/PM3 produces the most reliable hydrogen-bond energies for a set of nine water dimer geometries and far better energetics for the benzene dimer (Table 13). The inclusion of atom-pair information in the CRF via the PDDG function represents a valuable addition to the family of semiempirical methods. It is easy to implement, and it can be considered for future semiempirical MO methods. Work is underway to extend the PDDG methods to treat compounds that contain halogens, silicon, sulfur, and phosphorous.

Supplementary Material

A table is provided listing heats of formation for the 622 molecules from experiment and the AM1, PM3, MNDO, PM5, PDDG/MNDO, and PDDG/PM3 calculations. Corresponding results for bond lengths, bond angles, and dihedral angles of selected molecules are provided in a second table.

Acknowledgments

Table 17. Mean Absolute Errors in Bond Lengths (Å).

Bonding pairs	PDDG		Standard NDDO			No. mol.
	PM3	MNDO	PM3	AM1	MNDO	
H and H	0.060	0.082	0.042	0.064	0.078	1
H and C	0.018	0.016	0.016	0.020	0.016	38
H and N	0.027	0.014	0.019	0.012	0.005	4
H and O	0.006	0.014	0.010	0.013	0.012	6
C and C	0.016	0.019	0.016	0.018	0.016	72
C and N	0.011	0.011	0.013	0.013	0.012	13
C and O	0.010	0.019	0.010	0.019	0.016	16
N and N	0.028	0.070	0.009	0.070	0.053	1
N and O	0.018	0.044	0.009	0.046	0.035	2
O and O	0.070	0.199	0.007	0.165	0.179	1

References

1. Zerner, M. C. In Semiempirical Molecular Orbital Methods; Lipkowitz, K. B.; Boyd, D. B., Eds.; VCH Publishers: New York, 1991; vol. 2.

2. Stewart, J. J. P. In Semiempirical Molecular Orbital Methods; Lipkowitz, K. B.; Boyd, D. B., Eds.; VCH Publishers: New York, 1990; vol. 1.

3. Pople, J. A.; Beveridge, D. L. Approximate Molecular Orbital Theory; McGraw-Hill: New York, 1970.

4. Dewar, M. J. S.; Thiel, W. J Am Chem Soc 1977, 99, 4899.

5. Dewar, M. J. S.; Thiel, W. J Am Chem Soc 1977, 99, 4907.

6. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. J Am Chem Soc 1985, 107, 3902.

7. Stewart, J. J. P. J Comput Chem 1989, 10, 221.

8. Stewart, J. J. P. J Comput Chem 1989, 10, 209.

9. Dewar, M. J. S.; Jie, C.; Yu, J. Tetrahedron 1993, 49, 5003.

10. Stewart, J. P. MOPAC; Fujitsu, 2002.

11. Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. J Chem Phys 1991, 94, 7221.

12. Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. J Chem Phys 1997, 106, 1063.

13. Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. J Chem Phys 1998, 109, 7764.

14. Curtiss, L. A.; Raghavachari, K.; Pople, J. J Chem Phys 1993, 98, 1293.

15. Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Stefanov, B. B. J Chem Phys 1998, 108, 692.

16. Anh, N. T.; Frison, G.; Solladie–Cavallo, A.; Patrick, M. Tetrahedron 1998, 54, 12841.

17. Burk, P.; Herodes, K.; Koppel, I.; Koppel, I. Int J Quantum Chem Symp 1993, 27, 633.

18. Dewar, M. J. S.; Dieter, K. M. J Am Chem Soc 1986, 108, 8075.

19. Thiel, W. Tetrahedron 1988, 44, 7393.

20. Clark, T. J Mol Str (Theochem) 2000, 530, 1.

21. Kolb, M.; Thiel, W. J Comput Chem 1993, 14, 775.

22. Weber, W.; Thiel, W. Theor Chem Acc 2000, 103, 495.

23. Thiel, W. In Computational Thermochemistry. Prediction and Estimation of Molecular Thermodynamics; Irikura, K. K. and Frurip, D. J., Eds.; American Chemical Society: Washington, DC, 1998.

24. Mohle, K.; Hofmann, H.; Thiel, W. J Comput Chem 2001, 22, 509.

25. Gonzalez–Lafont, A.; Truong, T. N.; Truhlar, D. G. J Phys Chem 1991, 95, 4618.

26. Rossi, I.; Truhlar, D. G. Chem Phys Lett 1995, 233, 231.

27. Sekusak, S.; Corey, M. G.; Bartlett, R. J.; Sabljic, A. J Phys Chem A 1999, 103, 11394.

28. Liu, Y.; Lu, D.; Gonzalez–Lafont, A.; Truhlar, D. G.; Garrett, B. C. J Am Chem Soc 1993, 115, 7806.

29. Chuang, Y.; Radhakrishnan, M. L.; Fast, P. L.; Cramer, C. J.; Truhlar, D. G. J Phys Chem A 1999, 103, 4893.

30. Viggiano, A. A.; Paschkewitz, J. S.; Morris, R. A.; Paulson, J. F.; Gonzalez–Lafont, A.; Truhlar, D. G. J Am Chem Soc 1991, 11, 9404.

31. Khan, A. Chem Phys Lett 1996, 253, 299.

32. Khan, A. J Phys Chem A 1999, 103, 1260.

33. Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. J Comput Chem 2002, 23, 498.

34. Csonka, G. I. J Comput Chem 1993, 14, 895.

35. Csonka, G. I.; Angyan, J. G. J Mol Str (Theochem) 1997, 393, 31.

36. Bernal–Uruchurtu, M. I.; Ruiz–Lopez, M. F. Chem Phys Lett 2000, 330, 118.

37. Bernal–Uruchurtu, M. I.; Martins–Costa, M. T. C.; Millot, C.; Ruiz–Lopez, M. F. J Comput Chem 2000, 21, 572.

38. Stewart, J. J. P. J Comput Aided Mol Des 1990, 4, 1.

39. Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. General Linear Least Squares; Press Syndicate of the University of Cambridge: New York, 1996; vol. 1, p 665.

40. Initial distances are: $D_{H_1} = 0.7$; $D_{H_2} = 1.1, 1.2, 1.3$; $D_{C_1} = 1.1, 1.2$; $D_{C_2} = 1.3, 1.4, 1.5$; $D_{N_1} = 1.0, 1.1$; $D_{N_2} = 1.3, 1.4, 1.5$; $D_{O_1} = 1.1, 1.2$; $D_{O_2} = 1.3, 1.4, 1.5$, all distances in Angstroms.

41. Weber, W. Ein Neues Semiempirisches NDDO-Verfahren mit Orthogonalisierungskorrekturen: Entwicklung des Modells, Implementierung, Parametrisierung and Anwendungen; PhD thesis; Universitat Zurich, 1996.

42. Halgren, T. A. J Comput Chem 1999, 20, 730.

43. Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M.; Tanabe, K. J Am Chem Soc 2002, 124, 104.

44. Tsuzuki, S.; Luthi, H. P. J Chem Phys 2001, 114, 3949.

45. Cummins, P. L.; Titmuss, S. J.; Jayatilaka, D.; Bliznyuk, A. A.; Rendell, A. P.; Gready, J. E. Chem Phys Lett 2002, 352, 245.

46. van der Vaart, A.; Mertz, K. M. J. J Phys Chem A 1999, 103, 3321.

47. van der Vaart, A.; Mertz, K. M. J. J Am Chem Soc 1999, 121, 9182.

48. See http://zarbi.chem.yale.edu.

49. Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Pople, J. A. J Chem Phys 2000, 112, 7374.

50. Lias, S. G.; Bartmess, J. E.; Liebman, J. F.; Holmes, J. L.; Levin, R. D.; Mallard, W. G. J Phys Chem Ref Data 1988, 17.

51. Wagman, D. D.; Evans, W. H.; Parker, V. B.; Schum, R. H.; Hallow, I.; Bailey, S. M.; Churney, K. L.; Nuttall, R. L. J Phys Chem Ref Data 1982, 11, Supplement 2.

52. Thiel, W.; Voityuk, A. A. Int J Quantum Chem 1992, 44, 807.

53. Bartmess, J. E.; McIver, R. T., Jr. Gas Phase Ion Chemistry; Academic Press: New York, 1979; vol. II.

54. Li, Y.; Houk, K. N. J Am Chem Soc 1993, 115, 7478.

55. Burrows, C. J.; Carpenter, B. K. J Am Chem Soc 1981, 103, 6983.

56. Coates, R. M.; Rogers, B. D.; Hobbs, S. J.; Peck, D. R.; Curran, D. J Am Chem Soc 1987, 109, 1160.

57. Yoo, H. Y.; Houk, K. N. J Am Chem Soc 1997, 119, 2877.

58. Carr, R. W.; Walters, W. D. J Phys Chem 1965, 69, 1073.

59. Baughcum, S. L.; Smith, Z.; Wilson, E. B.; Duerst, R. W. J Am Chem Soc 1984, 106, 2260.

60. Halgren, T. A. J Comput Chem 1996, 17, 520.

61. Smith, B. J.; Swanton, D. J.; Pople, J. A.; Schaefer, H. F., III; Radom, L. J Chem Phys 1990, 92, 1240.