

## NO-MNDO: Reintroduction of the Overlap Matrix into MNDO

Kurt W. Sattelmeyer, Ivan Tubert-Brohman, and William L. Jorgensen\*

*Department of Chemistry, Yale University, 225 Prospect Street,  
New Haven, Connecticut 06520*

Received July 18, 2005

**Abstract:** The effect of reintroducing the overlap matrix into the secular equations for an NDDO (neglect of diatomic differential overlap)-based semiempirical molecular orbital method has been investigated. The modification is expected to improve the description of interactions between electron pairs. The idea has been tested by implementation and evaluation of a nonorthogonal version of the MNDO method (NO-MNDO) with parametrization for hydrogen, carbon, nitrogen, and oxygen. Overall, the accuracy of NO-MNDO for heats of formation is nearly identical to that for the more highly parametrized AM1 method. The mean absolute error (MAE) for heats of formation of a comprehensive set of 622 neutral, closed-shell molecules is reduced from 8.4 kcal/mol with MNDO to 6.8 kcal/mol with NO-MNDO. In addition, the performance for conformational equilibria and torsional barriers is significantly improved with NO-MNDO, presumably owing to the improved description of closed-shell interactions. For molecular geometries, the usual training and test sets have been expanded through use of MP2/6-31G(d) results for consistent comparisons. The performance of NO-MNDO for bond lengths, bond angles, and dihedral angles remains good with MAEs of 0.017 Å, 2.5°, and 4.5°. Additionally, NO-MNDO corrects severe errors by MNDO for  $R^\bullet + H-R'$  hydrogen-atom transfers, while testing for activation barriers for nine pericyclic reactions reveals only modest improvement.

### 1. Introduction

The speedup afforded by the Neglect of Diatomic Differential Overlap (NDDO)<sup>1–3</sup> approximation has made the semiempirical molecular orbital (SMO) methods based on it, including MNDO,<sup>4,5</sup> AM1,<sup>6</sup> PM3,<sup>7,8</sup> and MNDO/d,<sup>9</sup> valuable tools when a more rigorous approach is precluded by either the size of the system or the number of computations required. While mean absolute errors (MAEs) using these schemes do not reach chemical accuracy (ca. 1.0 kcal/mol, as can be approached using the best available  $N^7$  ab initio methods, such as CCSD(T)<sup>10,11</sup>), the structures and energetics from SMO methods are often acceptable for many applications. For example, MNDO, AM1, and PM3 give MAEs for heats of formation of 6.8, 5.1, and 4.1 kcal/mol for the 56 molecules in the combined G2-1 and G2-2 sets,<sup>12,13</sup> which contain only H, C, N, and O atoms. Nevertheless, SMO

methods suffer from a number of problems. Common errors include prediction of straight chain hydrocarbons to be more stable than branched isomers, underestimation of rotational barriers, overestimation of activation energies for pericyclic reactions, and significant energetic errors for molecules containing adjacent heteroatoms or small rings.<sup>14,15</sup>

Recent attempts to improve upon these semiempirical methods have largely centered upon corrections to the core repulsion formula (CRF). Specifically, the Pairwise Distance Directed Gaussian (PDDG)<sup>16</sup> extension in eq 1 to the MNDO and PM3 CRFs and subsequent reparametrizations

$$\text{PDDG}(A,B) = \text{CRF}_{\text{MNDO}} + \frac{1}{n_A + n_B} \left[ \sum_{i=1}^2 \sum_{j=1}^2 (n_A P_{Ai} + n_B P_{Bj}) e^{-10\text{\AA}^{-2}(R_{AB}-D_{Ai}-D_{Bj})^2} \right] \quad (1)$$

yield large improvements for heats of formation and isomerization energies without significantly degrading other proper-

\* Corresponding author e-mail: william.jorgensen@yale.edu.

ties. The PDDG correction to MNDO lowers the MAEs for a comprehensive set of 622 neutral, closed-shell molecules containing only H, C, N, and O from 8.4 to 5.2 kcal/mol. It is not surprising that this result is somewhat worse than the 4.4 kcal/mol MAE of PM3 (the CRF for PM3 is represented in eq 2), as the number of parameters per atom for the added Gaussians has been reduced

$$\text{PM3}(A,B) = \text{CRF}_{\text{MNDO}} + \frac{Z_A Z_B}{R_{AB}} \left( \sum_{i=1}^2 a_{Ai} e^{-b_{Ai}(R_{AB}-c_{Ai})^2} + \sum_{i=1}^2 a_{Bi} e^{-b_{Bi}(R_{AB}-c_{Bi})^2} \right) \quad (2)$$

from six to four, with the discrepancy arising from the  $b$  parameters in eq 2 being taken as constants in PDDG/MNDO. It is also not surprising that the combined PDDG/PM3 method gives the lowest MAE, 3.2 kcal/mol. Additionally, Thiel et al. have pursued a different approach in their OM1 and OM2 methods<sup>17</sup> through modifications to the NDDO version of the Roothaan–Hall equations.<sup>18,19</sup> In their work, the lack of Pauli repulsion has been addressed by the addition of three-center terms into the two-center core Hamiltonian, yielding a better description of barriers to internal rotation.

While these methods have been largely successful in addressing certain problems with specific implementations of the NDDO formalism, ad hoc modifications have thus far not led to a general purpose method capable of correcting all of the problems simultaneously. With this in mind, we have been exploring simple modifications to yield a general, improved SMO method that can be easily implemented in existing SMO codes and that is particularly appropriate for QM/MM calculations, while still retaining the favorable  $N^3$  scaling. In this paper, one such modification, namely, an alternative treatment of the Pauli repulsion issue, is considered by reintroduction of the overlap matrix into the secular equations.

## 2. Nonorthogonalized MNDO (NO-MNDO) Formalism

Much of the motivation behind Thiel's OM1 and OM2 is the improper, even splitting of mixing orbitals that arises from neglecting the overlap integrals in the secular equations for ZDO (zero differential overlap) theories and by extension for the NDDO-based methods. Instead, the lower-lying orbital should fall by less than the higher-lying orbital increases in energy. Thus, when both orbitals are doubly occupied, there is an intrinsic repulsion, the "Pauli repulsion," which increases as the overlap of the orbitals increases. This has nothing to do with two-electron integrals and simply results from deriving secular equations for one- or many-electron systems. Neglect of Pauli repulsion can be expected to contribute to the underestimation of rotational barriers and problems with the treatment of adjacent heteroatoms with lone pairs and electronic excitation energies, which are common with SMO methods. Although Thiel's OM methods represent a reasonable palliation, the orthogonalization issue has been addressed here more simply by reintroducing the overlap matrix ( $S$ ) into the secular equations. Therefore, we have returned to solving  $FC=SCE$ . This modification does

**Table 1.** Molecular Properties Used in Construction of the Error Function for Parameter Optimization

| data type            | $N$ | weighting factor      |
|----------------------|-----|-----------------------|
| heat of formation    | 355 | 1 mol/kcal            |
| bond length          | 153 | 100 Å <sup>-1</sup>   |
| bond angle           | 93  | 2/3 deg <sup>-1</sup> |
| dihedral angle       | 15  | 1/3 deg <sup>-1</sup> |
| ionization potential | 66  | 10 eV <sup>-1</sup>   |
| dipole moment        | 42  | 20 D <sup>-1</sup>    |

not increase the scaling of the method or the number of parameters versus MNDO. The  $S$  matrix is always available in SMO calculations as it is used in the computation of the one-electron, two-center resonance integrals,  $\beta_{\mu\nu}$ . In calculating heats of formation, the PDDG approach has also been used here, i.e., the electronic energy of an atom (*isol*) is treated as an optimizable parameter and not as one derived from calculations on the atom. This parameter is simply set to minimize the MAEs.

It can be speculated that the motivation to remove the overlap terms from the secular equations for the seminal ZDO-based SMO method, CNDO,<sup>20</sup> by Pople et al. in 1965 was somewhat influenced by practicality since the requisite second matrix diagonalization in solving the nonorthogonal eigenvalue problem essentially doubles the required computer time. Though overlap distributions  $\varphi_\mu\varphi_\nu$  were neglected in two-electron integrals, they were, of course, never neglected for one-electron, two-center resonance integrals, and their neglect in the secular equations was arbitrary from a theoretical standpoint and physically incorrect. Furthermore, it should be noted that simple Hückel theories also neglect the overlap integrals in the secular equations, while they are included in Hoffmann's extended Hückel method (EHT) from 1963.<sup>21</sup> The practical difference here is that EHT calculations are noniterative, so the two matrix diagonalizations are only performed once, while SMO calculations require a normal SCF cycle. The inclusion of Pauli repulsion in EHT has been known for many years to be essential to its qualitative success in describing orbital interactions, aromaticity, and rotational barriers.<sup>22</sup>

## 3. Optimization of the Parameters

The impact of the reintroduction of the overlap integrals into the secular equations has been tested by (1) making the modification to the MNDO method to yield nonorthogonal MNDO (NO-MNDO), (2) parametrizing the method for molecules containing C, H, N, and O atoms, and (3) comparing the results with those from MNDO and other SMO methods. The required computer time for a NO-MNDO calculation is ca. twice that for an MNDO calculation owing to the second matrix diagonalization, as expected. It should be noted that NO-MNDO is not a method that we intend to utilize further; it was simply pursued to gauge the importance of the orthogonality issue for future SMO development.

All original MNDO parameters ( $U_{ss}$ ,  $\beta_s$ ,  $\zeta_s$ , and  $\alpha$  for hydrogen and  $U_{ss}$ ,  $U_{pp}$ ,  $\beta_s$ ,  $\beta_p$ ,  $\zeta_s$ ,  $\zeta_p$ , and  $\alpha$  for carbon, nitrogen, and oxygen) were reoptimized. As previously explained<sup>16,23</sup> the optimization process consists of three

**Table 2.** MNDO and NO-MNDO Parameters<sup>a</sup>

|              | MNDO <sup>b</sup> |             |             |             | NO-MNDO    |             |             |             |
|--------------|-------------------|-------------|-------------|-------------|------------|-------------|-------------|-------------|
|              | H                 | C           | N           | O           | H          | C           | N           | O           |
| $U_{ss}$     | -11.906276        | -52.279745  | -71.932122  | -99.64309   | -10.880363 | -50.189763  | -69.782951  | -96.705658  |
| $U_{pp}$     |                   | -39.205558  | -57.172319  | -77.797472  |            | -39.547267  | -56.981889  | -76.391762  |
| $\beta_s$    | -6.989064         | -18.985044  | -20.495758  | -32.688082  | -9.364858  | -16.208034  | -24.905520  | -35.477596  |
| $\beta_p$    |                   | -7.934122   | -20.495758  | -32.688082  |            | -10.637421  | -21.291958  | -28.881783  |
| $\zeta_s$    | 1.331967          | 1.787537    | 2.255614    | 2.699905    | 1.061597   | 1.925428    | 2.351138    | 2.455548    |
| $\zeta_p$    |                   | 1.787537    | 2.255614    | 2.699905    |            | 1.727933    | 1.951819    | 2.537964    |
| $\alpha$     | 2.544134          | 2.546380    | 2.861342    | 3.160604    | 2.687705   | 2.484460    | 2.658599    | 2.946645    |
| <i>eisol</i> | -11.906276        | -120.500606 | -202.566201 | -317.868506 | -13.160122 | -119.594403 | -202.243601 | -304.341294 |
| <i>DD</i>    |                   | 0.807466    | 0.639904    | 0.534602    |            | 0.784403    | 0.656525    | 0.577707    |
| <i>QQ</i>    |                   | 0.685158    | 0.542976    | 0.453625    |            | 0.708792    | 0.627489    | 0.482570    |
| $\rho_0^c$   | 1.058920          | 1.112429    | 1.001103    | 0.882296    | 1.058920   | 1.112428    | 1.001103    | 0.882296    |
| $\rho_1^c$   |                   | 0.813078    | 0.637459    | 0.521237    |            | 0.800239    | 0.646434    | 0.543492    |
| $\rho_2^c$   |                   | 0.747842    | 0.615275    | 0.526541    |            | 0.765271    | 0.679156    | 0.549476    |

<sup>a</sup> Units are as follows: (eV)  $U_{ss}$ ,  $U_{pp}$ ,  $\beta_s$ ,  $\beta_p$ , *eisol*; (au)  $\zeta_s$ ,  $\zeta_p$ ; (Bohr) *DD*, *QQ*,  $\rho_0$ ,  $\rho_1$ ,  $\rho_2$ ; (Å)  $\alpha$ . <sup>b</sup> References 4 and 5. <sup>c</sup> For use in MOPAC 6,  $\rho_0 = 0.5/AM$ ,  $\rho_1 = 0.5/AD$ ,  $\rho_2 = 0.5/AQ$ .

stages. First, random displacements of the parameters are generated, and simulated annealing is used to minimize an error function constructed from the properties listed in Table 1 and from the gradients for bond lengths, bond angles, and dihedral angles on the nonoptimized structures of a training set of 126 molecules. All reference geometry parameters have now been taken from MP2/6-31G\* calculations, which allowed a large expansion in the size of the training set. Additionally, heats of formation for a total of 355 molecules are present in this training set, including heats of formation of several transition state structures. Subsequently, promising parameter sets (as determined by their small error functions) are optimized with full geometry optimization for all molecules, neglecting the (zero) contribution of gradients to the error function. Final values of  $E_{el}^A$  or “*eisol*”, the electronic energy for each element A, are determined by minimizing the MAE in heats of formation for 473 molecules (464 neutral, closed-shell molecules and nine transition states of pericyclic reactions). Equation 3 provides the relationship between a molecule’s  $\Delta H_f$ , electronic energy,  $E_{mol}$ , and  $E_{el}^A$ .

$$\Delta H_f = E_{mol} + \sum_A (\Delta H_f^A - E_{el}^A) \quad (3)$$

The final testing for heats of formation was carried out on the full set of 622 molecules that was used previously.<sup>16</sup> All SMO calculations have been executed with a local version of MOPAC 6.<sup>24</sup> Detailed results for all molecules and transition structures are presented in the Supporting Information.

## 4. Results and Discussion

**Energetics.** Table 2 lists the newly optimized parameters for NO-MNDO as well as those for MNDO, and Table 3 shows the performance of NO-MNDO and the other semi-empirical methods for heats of formation. As expected, due to the global search method employed and the change in methodology, a significantly different parameter emerged. The most striking change is in the  $\beta_s$  value for hydrogen, which is more than 30% lower in NO-MNDO. Overall, the 6.8 kcal/mol MAE for NO-MNDO represents a 1.6 kcal/

**Table 3.** Mean Absolute Errors for Heats of Formation of Neutral, Closed-Shell Molecules (kcal/mol)

| molecules | N   | standard NDDO |     |     | PDDG |     | nonorthog<br>NO-MNDO |
|-----------|-----|---------------|-----|-----|------|-----|----------------------|
|           |     | MNDO          | AM1 | PM3 | MNDO | PM3 |                      |
| all       | 622 | 8.4           | 6.7 | 4.4 | 5.2  | 3.2 | 6.8                  |
| HC        | 254 | 8.0           | 5.6 | 3.6 | 5.1  | 2.6 | 5.8                  |
| HCN       | 89  | 6.3           | 7.3 | 4.7 | 5.7  | 4.2 | 10.5                 |
| HCO       | 238 | 8.7           | 7.2 | 4.6 | 5.0  | 3.2 | 6.2                  |
| HCNO      | 41  | 13.4          | 9.5 | 7.0 | 4.9  | 4.5 | 9.3                  |

mol improvement over MNDO. NO-MNDO shows greater than 2 kcal/mol improvements for CH and CHO containing compounds, but it currently does less well than MNDO with CHN containing compounds. It is possible that further parameter search would correct this anomaly. However, in comparison to MNDO, NO-MNDO benefits from the optimization of the *eisol* values as well as the inclusion of the overlap matrix. Specifically, we previously found that optimization of the *eisol* values in conjunction with a complete reoptimization of the other MNDO parameters yields an MNDO version with an MAE of 7.3 kcal/mol for the 622 molecules.<sup>16</sup> Thus, 0.5 kcal/mol of the remaining error is removed in proceeding to the current NO-MNDO. This is encouraging, particularly if a similar gain could be made starting from PM3, as the new method retains the same scaling properties without introducing any new optimizable parameters. It is also notable that nearly identical MAEs are obtained with NO-MNDO and AM1; however, two to four Gaussian functions per element are added to the core repulsion formula for AM1 along with 36 additional parameters for coverage of C, H, N, and O. It is apparent that much of the success of AM1 over MNDO comes from the reparametrization including independent optimization of the orbital exponents  $\zeta_s$  and  $\zeta_p$  rather than from the addition of the Gaussians to the CRF. The optimization of MNDO was clearly constrained by the modest computer resources available in the mid-1970s.<sup>4,5</sup>

A related gauge of success is the relative abundance of outliers, i.e., molecules for which the computed heat of

**Table 4.** Number of Molecules (out of 622) Computed to Have Heats of Formation Differing from Experimental Values by More than 15 and 30 Kcal/Mol

|               | standard NDDO |     |     | PDDG |     | nonorthog<br>NO-MNDO |
|---------------|---------------|-----|-----|------|-----|----------------------|
|               | MNDO          | AM1 | PM3 | MNDO | PM3 |                      |
| > 15 kcal/mol | 97            | 49  | 16  | 20   | 7   | 55                   |
| > 30 kcal/mol | 18            | 7   | 2   | 4    | 1   | 4                    |

formation differs from the experimental reference value by more than some large value. The numbers of these are listed in Table 4 using the arbitrary cutoffs of 15 and 30 kcal/mol. Again, NO-MNDO demonstrates a significant improvement over MNDO and is similar in performance to AM1. As detailed in Table 5, out of the 622 minimum-energy structures considered in this work, NO-MNDO is not able to reproduce the experimental heat of formation to within 30 kcal/mol for diazirine, carbon suboxide, isophthalamide, and N<sub>2</sub>. Stewart has recently suggested that there is an error in the experimentally reported heat of formation of isophthalamide,<sup>25</sup> which is also greatly overpredicted by AM1 and PM3. Using his value of −70.3 kcal/mol reduces the errors for isophthalamide by 21.1 kcal/mol. The other poorest performing cases for NO-MNDO are all small and have unique bonding characteristics. Cubane, which suffers from the additive errors of multiple four-membered rings, populates Table 5 for all SMO methods except AM1 and NO-MNDO. In general, the greatest problems for NO-MNDO occur with acetylenes, nitrogen-containing aromatic heterocycles, and compounds containing nitrogen–nitrogen multiple bonds.

A well-known problem for MNDO and AM1 is that they erroneously find branched isomers to be less stable than

straight-chain ones. MNDO's most severe branching problems (tri-*tert*-butyl methane, 2,2,3,3-tetramethylpentane, 2,3,3,4-tetramethylpentane, etc.) are largely corrected in NO-MNDO by the use of a larger  $\alpha$  in the CRF for hydrogen. However, there is a fine balance here with the limited number of parameters available, as choosing too large an exponent results in a marked contraction of H–C bond lengths. Less extreme cases, such as the pentane/neopentane enthalpy difference of −5.0 kcal/mol, also show some improvement with NO-MNDO. MNDO and AM1 predict +9.8 and +5.2 kcal/mol, while NO-MNDO yields +3.5 kcal/mol. PM3 does better at −1.3 kcal/mol, while the expanded core repulsion formula with PDDG/PM3 adequately solves the problem (−7.2). For butane vs isobutane, NO-MNDO gives +0.8 kcal/mol, while the experimental, MNDO, AM1, PM3, and PDDG/PM3, numbers are −2.0, +2.9, +1.7, −0.4, and −2.5 kcal/mol, respectively.

Another area of potential improvement for SMO methods is in the description of the transition-state energetics of prototypical pericyclic reactions, especially in comparison to their overall treatment of hydrocarbons. The consistent trend is an overestimation of the reaction barriers. As shown in Table 6, AM1 is best able to reproduce the experimental barriers for the nine representative reactions, which have been thoroughly studied by Guner et al.,<sup>26,27</sup> giving an average error of 6.8 kcal/mol. In comparison, HF/6-31G\*, MP2/6-31G\*, KMLYP/6-31G\*, and B3LYP/6-31G\* have MAEs of 18.7, 4.6, 3.2, and 1.7 kcal/mol, respectively. Due to the current deficiencies, the activation barriers for these nine prototypical pericyclic reactions were explicitly included in the parametrization of NO-MNDO. As demonstrated in Table 6, this led to only modest improvement over MNDO, even though the contribution of these errors to the overall error

**Table 5.** Problematic Heats of Formation and Their Differences from Experimental Values

| standard NDDO                           |   |                        | PDDG                   |                | nonorthog<br>NO-MNDO    |
|---|---|------------------------|------------------------|----------------|-------------------------|
| MNDO                                    | AM1                                     | PM3                    | MNDO                   | PM3            |                         |
| ozone (+94.6)                           | isophthalamide (+38.2)                  | isophthalamide (+35.8) | ozone (−51.4)          | cubane (−39.1) | diazirine (+60.4)       |
| tri- <i>tert</i> -butylmethane (+88.8)  | di- <i>tert</i> -butyl peroxide (+34.3) | cubane (−34.9)         | isobutylamine (−45.4)  |                | carbon suboxide (−50.6) |
| cubane (−49.6)                          | bicyclo[1.1.1] pentane (+33.4)          |                        | diadamantanone (+41.6) |                | isophthalamide (+41.8)  |
| 3,3,4,4-tetramethyl-2-pentanone (+45.7) | 5-methylisoxazole (+31.8)               |                        | cubane (−39.2)         |                | nitrogen (+39.2)        |
| di- <i>tert</i> -butyl peroxide (+45.1) | 3,5-dimethylisoxazole (+31.7)           |                        |                        |                |                         |

**Table 6.** Activation Enthalpies for Selected Pericyclic Reactions (kcal/mol)

| reaction                      | expt <sup>a</sup> | standard NDDO |      |      | PDDG |      | nonorthog<br>NO-MNDO |
|-------------------------------|-------------------|---------------|------|------|------|------|----------------------|
|                               |                   | MNDO          | AM1  | PM3  | MNDO | PM3  |                      |
| cyclobutene opening           | 31.9              | 49.8          | 35.3 | 40.6 | 44.8 | 41.3 | 51.3                 |
| 1,3,5-hexatriene closure      | 30.2              | 40.0          | 31.0 | 31.2 | 38.7 | 36.6 | 46.3                 |
| <i>o</i> -xylylene closure    | 28.1              | 40.3          | 38.8 | 38.9 | 43.6 | 41.2 | 46.4                 |
| 1,3-pentadiene [1,5]-H shift  | 36.8              | 57.4          | 39.8 | 36.4 | 46.1 | 32.4 | 39.1                 |
| cyclopentadiene [1,5]-H shift | 23.7              | 48.9          | 39.5 | 37.7 | 47.5 | 30.7 | 42.1                 |
| 1,5-hexadiene Cope            | 34.5              | 40.9          | 37.6 | 41.8 | 42.8 | 45.9 | 53.8                 |
| ethylene + 1,3-butadiene DA   | 25.0              | 45.3          | 23.8 | 27.0 | 41.1 | 30.0 | 44.0                 |
| ethylene + cyclopentadiene DA | 23.7              | 50.5          | 28.5 | 32.1 | 42.1 | 33.0 | 51.0                 |
| cyclopentadiene dimerization  | 15.9              | 50.0          | 34.2 | 37.4 | 44.3 | 38.3 | 26.6                 |
| MAE                           |                   | 19.3          | 6.8  | 8.2  | 15.7 | 9.8  | 16.7                 |

<sup>a</sup> References 26 and 27.



**Table 7.** Barrier Heights for Hydrogen Transfer Reactions (kcal/mol)

|  | consensus <sup>a</sup> | standard NDDO <sup>a</sup> |       |       | PDDG <sup>a</sup> |       | nonorthog<br>NO-MNDO |
|--|------------------------|----------------------------|-------|-------|-------------------|-------|----------------------|
|  |                        | MNDO                       | AM1   | PM3   | MNDO              | PM3   |                      |
| CH <sub>3</sub> <sup>•</sup> + CH <sub>4</sub>                             | 17.53                  | 28.59                      | 13.49 | 10.14 | 20.36             | 6.51  | 9.55                 |
| CH <sub>3</sub> <sup>•</sup> + C <sub>2</sub> H <sub>6</sub>               | 15.36                  | 79.59                      | 12.01 | 7.29  | 18.71             | 3.45  | 6.92                 |
| C <sub>2</sub> H <sub>5</sub> <sup>•</sup> + CH <sub>4</sub>               | 18.99                  | 85.89                      | 17.81 | 15.60 | 25.17             | 12.07 | 16.45                |
| C <sub>2</sub> H <sub>5</sub> <sup>•</sup> + C <sub>2</sub> H <sub>6</sub> | 16.69                  | 32.15                      | 16.02 | 11.99 | 23.41             | 8.29  | 13.08                |
| C <sub>3</sub> H <sub>7</sub> <sup>•</sup> + C <sub>3</sub> H <sub>8</sub> | 16.04                  | 32.45                      | 15.56 | 12.47 | 22.80             | 7.33  | 16.89                |
| MAE  |                        | 34.8                       | 1.9   | 5.4   | 5.2               | 9.4   | 4.7                  |

<sup>a</sup> Reference 28.**Table 8.** Conformational and Isomerization Energies (kcal/mol) for Prototypical Organic Molecules

| molecule                       |          | ref ΔE            | standard NDDO |      |      | PDDG |      | nonorthog<br>NO-MNDO |
|--------------------------------|----------|-------------------|---------------|------|------|------|------|----------------------|
|                                |          |                   | MNDO          | AM1  | PM3  | MNDO | PM3  |                      |
| butane (trans)                 | skew     | 3.6 <sup>a</sup>  | 1.4           | 1.5  | 1.6  | 2.2  | 1.6  | 1.0                  |
|                                | gauche   | 0.7               | 0.6           | 0.7  | 0.6  | 1.3  | 0.3  | −0.4                 |
|                                | cis      | 5.7               | 3.2           | 3.3  | 4.0  | 5.5  | 3.9  | 4.4                  |
| ethane (staggered)             | eclipsed | 2.8 <sup>b</sup>  | 1.0           | 1.2  | 1.4  | 2.2  | 1.1  | 1.4                  |
| methylcyclohexane (equatorial) | axial    | 1.8 <sup>c</sup>  | 6.6           | 1.4  | 1.1  | 4.4  | 0.9  | 0.3                  |
| propene (eclipsed)             | bisected | 2.0 <sup>d</sup>  | 0.2           | 0.6  | 0.7  | 0.1  | 0.7  | 1.0                  |
| 2-butene (trans)               | cis      | 1.0 <sup>e</sup>  | 0.8           | 1.1  | 0.2  | 0.8  | 1.5  | 0.9                  |
| 1,3-butadiene (trans)          | skew     | 2.49 <sup>e</sup> | 0.3           | 0.8  | 0.7  | 1.4  | 0.7  | 1.7                  |
| 1-butene (skew)                | cis      | 0.53 <sup>e</sup> | 1.3           | 0.7  | 0.9  | 2.3  | 0.7  | 0.8                  |
| propanal (cis)                 | skew     | 0.95 <sup>e</sup> | −0.3          | −0.6 | −0.7 | −1.5 | −1.1 | 0.6                  |
| N-methylacetamide (Z)          | E        | 2.3 <sup>e</sup>  | 1.0           | 1.6  | 0.4  | 1.3  | 1.9  | 2.2                  |
| acrolein (trans)               | cis      | 2.0 <sup>e</sup>  | −0.4          | 0.2  | 0.4  | 0.7  | 0.8  | 0.3                  |
| methyl formate (Z)             | E        | 4.75 <sup>e</sup> | 2.9           | 5.6  | 1.9  | 0.2  | 1.8  | 6.7                  |
| MAE                            |          | 0                 | 1.8           | 1.1  | 1.4  | 1.5  | 1.3  | 1.1                  |

<sup>a</sup> References 29 and 30. <sup>b</sup> References 31–34. <sup>c</sup> References 35 and 36. <sup>d</sup> Reference 37. <sup>e</sup> Reference 38.

function is not small. This suggests that the standard MNDO formalism does not allow enough flexibility to describe these activation barriers accurately. Indeed, inspection of the core repulsion formula of AM1 for carbon reveals two attractive Gaussians centered at 2.05 and 2.65 Å, thereby allowing AM1 to perform better due to the more favorable C–C interactions in this range. This result is somewhat spurious, though, as AM1 benefits from erroneous overestimation of the heats of formation of the reactants in these reactions; it is not noticeably better for the overall heats of reaction, giving an MAE of 7.1 kcal/mol for the six reactions with nonzero enthalpy changes compared to 8.6, 4.8, 3.9, 3.3, and 8.0 kcal/mol for MNDO, PM3, PDDG/MNDO, PDDG/PM3, and NO-MNDO, respectively.

We also examined the barrier heights of several hydrogen transfer reactions. The results are listed in Table 7 and accompany the recent, best estimates from Dybala-Defratyka et al.<sup>28</sup> In these cases, NO-MNDO does very respectably with an MAE of only 4.7 kcal/mol. The largest error is for the CH<sub>3</sub><sup>•</sup> + C<sub>2</sub>H<sub>6</sub> reaction, where the barrier is underestimated by 8.4 kcal/mol. Furthermore, NO-MNDO is seen to correct the serious problems in the MNDO results, indicating that the standard MNDO formalism augmented with the Pauli repulsions is adequate here.

As a final energetic issue, Table 8 compares conformational and isomerization energies for prototypical molecules from the SMO methods with experimental and high-level, computed values from the literature.<sup>29–38</sup> NO-MNDO and AM1 perform the best among the SMO methods. Although the gauche structure of butane is predicted to be 0.4 kcal/

**Table 9.** Mean Absolute Errors in Bond Lengths (Å)

|              | N   | standard NDDO |       |       | PDDG  |       | nonorthog<br>NO-MNDO |
|--------------|-----|---------------|-------|-------|-------|-------|----------------------|
|              |     | MNDO          | AM1   | PM3   | MNDO  | PM3   |                      |
| training set | 153 | 0.013         | 0.016 | 0.012 | 0.014 | 0.011 | 0.016                |
| test set     | 65  | 0.018         | 0.019 | 0.013 | 0.020 | 0.017 | 0.020                |
| all          | 218 | 0.015         | 0.017 | 0.012 | 0.016 | 0.013 | 0.017                |

mol more stable than anti using NO-MNDO, the overall improvement versus MNDO is apparent. While the anti to cis energy difference for butane is underestimated using MNDO by 2.5 kcal/mol and the energy difference between equatorial and axial methylcyclohexane is overestimated by 4.8 kcal/mol, the errors with NO-MNDO are just 1.3 and 1.5 kcal/mol, respectively. NO-MNDO is also the only method to give the cis structure of propanal as the minimum, as the other SMO methods have the skew structure lower by 0.3–1.5 kcal/mol. Thus, it seems likely that the addition of the Pauli repulsions in NO-MNDO has helped in this area.

**Structure.** Despite the improvements of NO-MNDO over MNDO with respect to energetics, it is important that the results for molecular geometries remain reasonable. In fact, as summarized in Tables 9–11, the overall quality of geometrical results is similar for all of the SMO methods. As previously mentioned, this work uses results of geometry optimizations at the MP2/6-31G\* level for the reference values. This allowed expansion of the training set for bond lengths, bond angles, and dihedral angles, and comparisons can now be made in a more consistent manner, e.g., using

**Table 10.** Mean Absolute Errors in Bond Angles (deg)

|              | N   | standard NDDO |     |     | PDDG |     | nonorthog<br>NO-MNDO |
|--------------|-----|---------------|-----|-----|------|-----|----------------------|
|              |     | MNDO          | AM1 | PM3 | MNDO | PM3 |                      |
| training set | 93  | 1.8           | 1.4 | 1.7 | 2.5  | 1.8 | 2.5                  |
| test set     | 33  | 2.0           | 1.7 | 1.6 | 2.3  | 1.9 | 2.4                  |
| all          | 126 | 1.9           | 1.5 | 1.7 | 2.4  | 1.9 | 2.5                  |

**Table 11.** Mean Absolute Errors in Dihedral Angles (deg)

|              | N  | standard NDDO |     |     | PDDG |     | nonorthog<br>NO-MNDO |
|--------------|----|---------------|-----|-----|------|-----|----------------------|
|              |    | MNDO          | AM1 | PM3 | MNDO | PM3 |                      |
| training set | 15 | 4.6           | 2.3 | 2.8 | 2.8  | 3.5 | 4.5                  |
| test set     | 19 | 3.1           | 3.2 | 3.5 | 4.9  | 3.9 | 4.4                  |
| all          | 34 | 3.8           | 2.8 | 3.2 | 4.0  | 3.7 | 4.5                  |

$r_e$  values for bond lengths and not a collection of  $r_0$ ,  $r_s$ , etc., depending on availability.

The greatest bond-length errors for NO-MNDO are for triple bonds. Those present in acetylene, propyne, isocyanomethane, 2-butyne, 1,3-butadiyne, vinylacetylene, and cyanogen are each underestimated by between 0.026 and 0.048 Å. In contrast, with an average error of 0.014 Å, the lengths of carbon–carbon single bonds are generally accurate and do not show systematic discrepancies. The largest errors here include an overestimate of the carbon–carbon single bonds of azirane by 0.046 Å and underestimates of those for propyne and 2-butyne by 0.041 and 0.040 Å. Other substantial deviations include underestimations of the bond length in molecular hydrogen by 0.099 Å and the nitrogen–nitrogen bond in hydrazine by 0.067 Å.

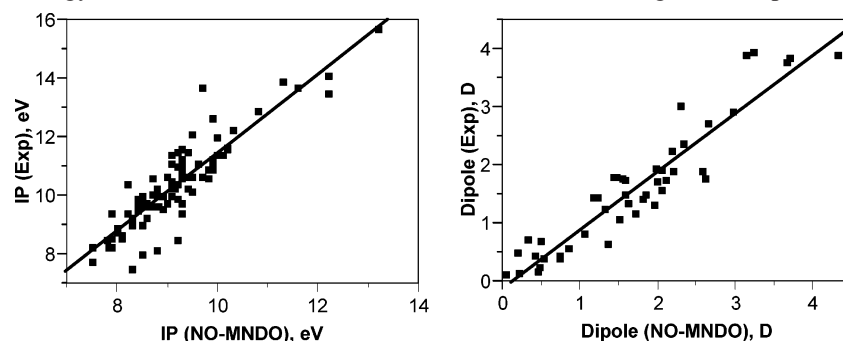
While NO-MNDO yields an MAE for bond angles slightly higher than the other SMO methods, it is seen to have its significant errors in many of the same situations as MNDO and PDDG/MNDO, specifically, the overestimation of angles where the central atom is oxygen. For example, the MNDO, PDDG/MNDO, NO-MNDO, and MP2/6-31G\* COH bond angles for formic and acetic acid are 116.2°, 123.5°, 118.5°, and 106.1° and 115.6°, 122.9°, 117.8°, and 105.4°, respectively, while those for the COC angle of methyl formate are 125.7°, 127.9°, 121.2°, and 113.9°. Table 11 shows that the dihedral angle results from NO-MNDO are comparable to those of the other methods. AM1 appears to be the best performer in this area, though neither the data set nor the margin is large. The greatest sources of errors are also consistent across the methods and correspond to cases with relatively flat torsional energy surfaces.

**Ionization Potentials and Dipole Moments.** Ionization potentials from Koopman's theorem and dipole moments are also traditionally examined in papers reporting SMO methods. Though these properties were not emphasized in this study, the results with NO-MNDO compare reasonably well with those from the alternative SMO methods. The increase of the one-electron energy,  $U_{ss}$ , for hydrogen in Table 2 causes ionization potentials to generally be underestimated by ca. 1–2 eV for hydrocarbons with NO-MNDO; the largest error occurs for methane. For the 96 compounds that were studied, the average errors in ionization potentials are 0.72, 0.53, 0.59, 0.65, 0.56, and 1.20 eV for MNDO, AM1, PM3, PDDG/MNDO, PDDG/PM3, and NO-MNDO. However, in plots of the experimental and SMO results, as in Figure 1, the correlation coefficients ( $r^2$ ) for the ionization potentials are 0.75, 0.86, 0.81, 0.82, 0.82, and 0.80 from MNDO, AM1, PM3, PDDG/MNDO, PDDG/PM3, and NO-MNDO, and rms errors from the linear fits are 0.74, 0.55, 0.64, 0.62, 0.62, and 0.66 eV, respectively.

For gas-phase dipole moments, 47 molecules were considered. The average errors are 0.29, 0.22, 0.25, 0.20, 0.23, and 0.31 D from MNDO, AM1, PM3, PDDG/MNDO, PDDG/PM3, and NO-MNDO. For the correlations of the experimental and computed values, the corresponding  $r^2$  values are 0.88, 0.90, 0.91, 0.93, 0.92, and 0.88, and the rms errors are 0.38, 0.33, 0.33, 0.30, 0.31, and 0.37 D, respectively.

## 5. Conclusions

The effect of introduction of the overlap matrix in the secular determinant has been evaluated starting from the MNDO method. The implementation featured parametrization for molecules containing C, H, N, and O atoms, and the resultant nonorthogonal method was designated NO-MNDO. Testing included computation of a large number and variety of energetic quantities, ionization potentials, and dipole moments. Any study of this type may be incomplete since additional testing, e.g., for ion energetics or hydrogen bonding, could be performed and because the optimal parameter sets may not have been found, as suspected here for nitrogen with NO-MNDO. The present results indicate that the NO-MNDO modification coupled with optimization of the atomic energies, *isol*, provides significantly improved



**Figure 1.** Correlation of experimental ionization potentials (left) and gas-phase dipole moments (right) with results from NO-MNDO calculations.

energetic results over those from MNDO. Overall, the accuracy of NO-MNDO is very similar to that of the AM1 method, which utilizes more than 30 additional, optimized parameters. Therefore, it is apparent that one can devise an MNDO variant that has similar quality as AM1 but does not require the addition of the AM1 Gaussians to the core repulsion formula. Notable improvements for NO-MNDO over MNDO are obtained for rotational barriers about single bonds and for the barriers for hydrogen-atom transfer reactions. The characteristic branching errors for isomers from MNDO and AM1 were also relieved. It is reiterated that the present work was not carried out to introduce a new SMO method; its sole purpose was to test the impact of including the overlap matrix in the secular equations for an MNDO-based method. The associated improvements are significant enough to warrant consideration of the methodological change in the development of future semiempirical MO methods.

**Acknowledgment.** The authors wish to thank Dr. Cristiano R. W. Guimarães for helpful discussion and the National Science Foundation (CHE-0446920) for financial support.

**Supporting Information Available:** Complete listing of all computed and experimental data. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## References

- (1) Zerner, M. C. In *Semiempirical Molecular Orbital Methods*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1991; Vol. 2, p 313.
- (2) Stewart, J. J. P. In *Semiempirical Molecular Orbital Methods*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, 1991; Vol. 1, p 70.
- (3) Pople, J. A.; Beveridge, D. L. In *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*, 209.
- (8) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 221.
- (9) Thiel, W.; Voityuk, A. A. *J. Phys. Chem.* **1996**, *100*, 616.
- (10) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.
- (11) Bartlett, R. J.; Watts, J. D.; Kucharski, S. A.; Noga, J. *Chem. Phys. Lett.* **1990**, *165*, 513.
- (12) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1991**, *94*, 7221.
- (13) Curtiss, L. A.; Raghavachari, K.; Trucks, G. W.; Pople, J. A. *J. Chem. Phys.* **1997**, *106*, 1063.
- (14) Dewar, M. J. S.; Dieter, K. M. *J. Am. Chem. Soc.* **1986**, *108*, 8.
- (15) Burk, P.; Herodes, K.; Koppel, I.; Koppel, I. *Int. J. Quantum Chem. Symp.* **1993**, *27*, 633.
- (16) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601.
- (17) (a) Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495. (b) Mohle, K.; Hofmann, H.-J.; Thiel, W. *J. Comput. Chem.* **2001**, *22*, 509. (c) Kolb, M.; Thiel, W. *J. Comput. Chem.* **1993**, *14*, 775.
- (18) Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69.
- (19) Hall, G. G. *Proc. R. Soc. London, Ser. A* **1951**, *205*, 541.
- (20) Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S129.
- (21) Hoffmann, R. *J. Chem. Phys.* **1963**, *39*, 1397.
- (22) Jorgensen, W. L.; Borden, W. T. *J. Am. Chem. Soc.* **1973**, *95*, 6649.
- (23) (a) Tubert-Brohman, I.; Guimarães, C. R. W.; Repasky, M. P.; Jorgensen, W. L. *J. Comput. Chem.* **2004**, *25*, 138. (b) Tubert-Brohman, I.; Guimarães, C. R. W.; Jorgensen, W. L. *J. Chem. Theory Comput.* In press.
- (24) Stewart, J. J. P. *J. Comput.-Aided Mol. Des.* **1990**, *4*, 1.
- (25) Stewart, J. J. P. *J. Mol. Model.* **2004**, *10*, 6.
- (26) Guner, V.; Khuong, K. S.; Leach, A. G.; Lee, P. S.; Bartberger, M. D.; Houk, K. N. *J. Phys. Chem.* **2003**, *107*, 11445.
- (27) Ess, H. D.; Houk, K. N., personal communication.
- (28) Dybala-Defratyka, A.; Paneth, P.; Pu, J.; Truhlar, D. G. *J. Phys. Chem.* **2004**, *108*, 2475–2486.
- (29) Herrebout, W. A.; van der Veken, B. J.; Wang, A.; Durig, J. R. *J. Phys. Chem.* **1995**, *99*, 578.
- (30) (a) Murcko, M. A.; Castejon, H.; Wiberg, K. B. *J. Phys. Chem.* **1996**, *100*, 16162. (b) Allinger, N. L.; Fermann, J. T.; Allen, W. D.; Schaefer, H. F., III *J. Chem. Phys.* **1997**, *106*, 5143.
- (31) Weiss, S.; Leroi, G. *J. Chem. Phys.* **1968**, *48*, 962.
- (32) Csaszár, A. G.; Allen, W. D.; Schaefer, H. F., III *J. Chem. Phys.* **1998**, *108*, 9751.
- (33) Moazzen-Ahmadi, N.; Gush, H. P.; Halpren, M.; Jagannath, H.; Leung, A.; Ozier, I. *J. Chem. Phys.* **1988**, *88*, 563.
- (34) Fantoni, R.; van Hellvoort, K.; Knippers, W.; Reuss, J. *Chem. Phys.* **1986**, *110*, 1.
- (35) Booth, H.; Everett, J. R. *J. Chem. Soc., Perkin Trans.* **1980**, *2*, 255.
- (36) Abraham, R. J.; Ribeiro, D. S. *J. Chem. Soc., Perkin Trans.* **2001**, *2*, 302.
- (37) Wiberg, K. B.; Martin, E. *J. Am. Chem. Soc.* **1985**, *107*, 5035.
- (38) Murphy, R. B.; Beachy, M. D.; Friesner, R. A.; Ringnalda, M. N. *J. Chem. Phys.* **1995**, *103*, 1481.

CT050174C