

Feature article

Towards a “next generation” neglect of diatomic differential overlap based semiempirical molecular orbital technique

Paul Winget, Cenk Selçuki, Anselm H.C. Horn, Bodo Martin, Timothy Clark

Computer-Chemie-Centrum, Friedrich-Alexander-Universität Erlangen–Nürnberg, Nögelsbachstrasse 25, 91052, Erlangen, Germany

Received: 30 January 2003 / Accepted: 29 April 2003 / Published online: 20 October 2003
© Springer-Verlag 2003

Abstract. We discuss problems and features of current semiempirical molecular orbital techniques and test some of the approximations and assumptions used. Prerequisites for a “next generation” technique include orthogonalization corrections, effective core potentials and an implicit dispersion term. However, validation of experimental parameterization data using density functional theory or the Gaussian 2 approach reveals significant errors in some cases. Developers of future methods will need to validate all their parameterization data and may no longer be able to parameterize for heats of formation at 298 K, but may need to use Born–Oppenheimer binding energies. We also suggest that there is no inherent reason that semiempirical techniques should not reproduce hydrogen bonding and show that the Gaussian potentials added to the core–core terms in AM1 and the PM n methods actually weaken hydrogen bonding, rather than strengthening it.

Keywords: Molecular orbital theory – Modified neglect of differential overlap – MNDO/d – AM1 – PM3

Introduction

Semiempirical molecular orbital (MO) theory in its current incarnation is generally based on the neglect of diatomic differential overlap (NDDO) formalism [1]. In essence, current techniques such as AM1 [2, 3] and PM3 [4, 5] are minor modifications of Dewar and Thiel’s modified neglect of differential overlap (MNDO) [6, 7]. Modifications to the original MNDO formalism have tended to be in the form of a change in the parameterization strategy [4, 5] and the addition of “correcting” terms to two-center energies, usually the core–core term

[2, 3, 4, 5, 8], but also as an additional classical torsional term to increase amide C–N bond rotation barriers [4, 5]. Significant advances in NDDO theory include the addition of d orbitals to the minimal s,p Slater basis [9, 10] and the formulation of an orthogonalization correction [11], both of which were introduced by Thiel and his group. An interesting intermediate neglect of diatomic overlap (INDO) [12] based technique, SINDO/1 [13], has been described in detail and appears to perform well, but has not found the wide acceptance of the NDDO-based techniques.

Recently, extensions of established techniques to new elements (AM1, PM3 [14]), a new parameterization (PM5 [14]) and an extension of PM3 with d orbitals (PM3-tm [15]) have appeared, but details have not yet been published. A further essentially unpublished technique, SAM1 [16], apparently differs significantly from the original MNDO ansatz. In laudable contrast to these undocumented methods, Voityuk and Rösch [17] have published an AM1(d) parameter set for molybdenum, Hutter et al. [18] a new magnesium parameterization for AM1, Brothers and Merz [19] sodium parameters for AM1 and PM3 and Jorgensen et al. [20] a new additivity scheme for heats of formation. We are now in the interesting situation that the probably most widely distributed (at least in its earlier versions) semiempirical program, MOPAC, offers unpublished AM1 and PM3 parameters for sodium but not the published parameter sets. The methods and elements available in our semiempirical program VAMP are listed with their literature references in Table 1.

We have previously [21] speculated on the future of semiempirical MO theory in the light of the consistent reports of its death. It suffices here to point out that there is really no competition between semiempirical theory and more “respectable” methods such as ab initio or density functional theory (DFT) because semiempirical methods are generally used for applications that would not be accessible to the more expensive methods. In good research, semiempirical theory is either used as a preliminary

Correspondence to: T. Clark
e-mail: clark@chemie.uni-erlangen.de

Table 1. Parameters implemented inVAMP

Hamiltonian	Elements
MNDO	H [6], He [67], Li [68], Be [69], B [70], C [6], N [6], O [6], F [71], Mg [72], Al [73], Si [74], P [75], S [76], Cl [77], K [78], Ca [79], Zn [80], Ge [81], Br [82], Sn [83], I [84], Hg [85], Pb [86]
MNDO/c [87]	H, C, N, O
AM1	H [2], B [88], C [2], N [2], O [2], F [89], Na [19], Mg [18], Al [90], Si [91], P [92], S [93], Cl [89], Zn [94] ^a , Ge [95], Br [89], Sn [96], I [89], Hg [97]
PM3	H [4], Li [98], Be [99], B [100], C [4], N [4], O [4], F [4], Na [19], Mg [99], Al [4], Si [4], P [4], S [4], Cl [4], Ca [100], Zn [99], Ga [99], Ge [99], As [99], Se [99], Br [4], Cd [99], In[99], Sn [99], Sb [99], Te [99], I [4], Hg[99], Tl [99], Pb [99], Bi [99]
MNDO/d	H [6], He [67], Li [68], Be [69], B [70], C [6], N [6], O [6], F [71], Na [10], Mg [10], Al [10] ^b , Si [101], P [10], S [10], Cl [10], Zn [10], Br [10], Cd [10], I [10], Hg [10]
AM1*	H [2], C [2], N [2], O [2], F [89], P [37], S [37], Cl [37]

^aAlternative parameters are also available for Zn [15]

^bThe reported parameters are in error. We thank Alexander Voityuk for the correct parameters

investigative tool or on systems for which ab initio or DFT would be far too slow, although this is not always true. There are some alternatives to semiempirical theory for large systems. Self-consistent-charge density functional tight binding theory [22] looks promising for proteins, but is not yet available for many elements. Empirical valence bond theory [23] has also recently enjoyed a wave of popularity as a means of treating reactions within an essentially classical mechanical framework. However, semiempirical techniques have played an important role in the development of continuum solvation models [24, 25] and are particularly promising for very large applications using either hybrid quantum mechanics/molecular mechanics [26, 27, 28] or linear scaling divide-and-conquer [29, 30] or localized orbital [31] techniques.

Because of the unsatisfactory situation that new semiempirical parameterizations are not being published for commercial reasons, that very few groups are actually working on improvements to the basic NDDO theory and that, on the contrary, modifications to AM1 and PM3 often involve a classical two-center potential added to the quantum mechanical technique, we decided to investigate the behavior and characteristics of NDDO-based theory and of possible improvements that are theoretically justifiable. MNDO forms the basis of these investigations because it is largely free of “correcting” influences, such as the Gaussian modification to the core–core energy, and thus offers the best chance to observe the true characteristics of the underlying theoretical framework. In particular, our work is directed towards establishing a “next generation” NDDO-based technique, which will not be presented here. Rather, the aim of the current work is to provide the foundation for the development of such a method, in order to be able to plan the theoretical extensions to MNDO. First, however, we review the problems of current methods and their possible solutions.

Systematic errors of current methods

Rotation barriers and conformations

Rotation barriers for both normal single bonds and formally single bonds in conjugated π systems are con-

sistently too low in MNDO, AM1, PM3, etc. AM1 performs slightly better for π -rotation barriers than the other two methods, but still underestimates them significantly. The problem is particularly relevant for amide bonds, for which PM3 only gives a rotation barrier of about 3 kcal mol⁻¹. This problem was “fixed” by adding an optional classical torsional potential for these bonds in MOPAC.

A related problem is that of the relative stabilities of alternative conformations at single bonds, either for problem systems such as peroxides and hydrazines or for more basic problems like *gauche*-butane and *trans*-butane. The latter problem is very fundamental because the *gauche*–*trans* energy difference is the basis of much of the conformational behavior of cycloalkanes and their derivatives. The *gauche*–*trans* energy difference (experimentally 0.9 kcal mol⁻¹) is calculated to be 0.6, 0.7 and 0.5 kcal mol⁻¹ for MNDO, AM1 and PM3, respectively. Although the three methods only underestimate the energy difference slightly, they make the carbon skeleton of cyclobutane planar and cyclopentane and cyclohexane flatter than their experimental structures. MNDO also calculates a planar ring skeleton for cyclopentane.

The orthogonalization correction introduced by Kolb, Thiel and Weber [11] is derived in terms of an atom-based pseudopotential that is added to the one-electron matrix, H^{core} . The final semiempirical method contains a two-center and a three-center energy correction; the latter is given here for review purposes:

$$\begin{aligned}
 H'_{\mu\lambda} &= \sum_C V_{\mu\lambda,C}(\text{ORT}), \\
 V_{\mu\lambda,C}(\text{ORT}) &= -\frac{1}{2} G_1^{\text{AB}} \sum_{\rho}^C (S_{\mu\rho} \beta_{\rho\lambda} + \beta_{\mu\rho} S_{\rho\lambda}) + \frac{1}{8} G_2^{\text{AB}} \\
 &\quad \times \sum_{\rho}^C S_{\mu\rho} S_{\rho\lambda} (H_{\mu\mu}^{\text{loc}} + H_{\lambda\lambda}^{\text{loc}} - 2H_{\rho\rho}^{\text{loc}}), \\
 \beta_{\mu\lambda} &= \pm \frac{1}{2} (\beta_{\mu}^{\text{A}} + \beta_{\lambda}^{\text{B}}) \sqrt{R_{\text{AB}}} \exp \left[-(\alpha_{\mu}^{\text{A}} + \alpha_{\lambda}^{\text{B}}) R_{\text{AB}}^2 \right].
 \end{aligned}
 \tag{1}$$

μ , λ and ρ refer to atomic orbitals at centers A, B and C. $S_{\mu\lambda}$ is the overlap between the two atomic orbitals, $\beta_{\mu\lambda}$ is

the resonance integral, an empirical function with the four parameters β_μ^A , β_λ^B , α_μ^A and α_λ^B . G_1^{AB} and G_2^{AB} are the arithmetic mean values of atomic parameters. Finally, H_{xx}^{loc} is the matrix element of \mathbf{H}^{core} in a local coordinate system. The two-center correction has the same functional form with different atomic parameters.

This correction significantly improves the performance of MNDO-type methods for the conformations of alkanes and cycloalkanes, but it is not clear whether such extensions also cure the problem for π systems. It is also possible that paying careful attention to rotation barriers during the parameterization may be an adequate solution to this problem. As far as we know, rotation barriers have never been addressed specifically during parameterization.

Nitrogen pyramidal

The pyramidal and inversion barriers at nitrogen generally present problems for the current methods. AM1, for instance, finds amide nitrogens to be planar, but also makes amines less pyramidal than they should be and underestimates their barriers to inversion. This problem is particularly severe for tertiary amines. PM3, on the other hand, often makes amide nitrogens pyramidal. We discuss this problem later.

Hydrogen bonds

MNDO does not reproduce hydrogen bonding. This was remedied in MNDO/H [8] by additional Gaussian terms in the core–core repulsion, a practice that was also adopted for AM1 and PM3. The expression for the core–core repulsion, $E_{\text{core-core}}(i-j)$, between atoms i and j then takes the form

$$E_{\text{core-core}}(i-j) = Z_i Z_j \gamma_{\text{ss}} [1 + \exp(-\alpha_i R_{ij}) + \exp(-\alpha_j R_{ij}) + G_i + G_j], \quad (2)$$

where Z_i and Z_j are the core charges of i and j , respectively, γ_{ss} is the monopole–monopole integral $\langle ii | jj \rangle$ as defined for MNDO [6, 7], R_{ij} is the distance between the two atoms, α_i and α_j are element-specific parameters and G_i and G_j are the element-specific Gaussian terms, which are defined as

$$G_i = \sum_n K_{i_n} \exp [L_{i_n} (R_{ij} - M_{i_n})^2] Q, \quad (3)$$

where n is the number of Gaussian functions defined for element i (usually 2–4), K_{i_n} , L_{i_n} and M_{i_n} are element-specific parameters that define the magnitude, position and width of the n th Gaussian function for element i .

However, AM1 does not reproduce the geometries of hydrogen-bonded systems correctly, tending to give bifurcated structures. PM3 reproduces these geometries

better, but underestimates the strengths of hydrogen bonds. Jug and Geudtner [32] have found that additional p orbitals for hydrogen fix this problem in SIN-DO/1, but it is not yet clear that this is a universal solution. Extending the basis set for hydrogen would be a computationally expensive solution, but would be affordable in view of the progress that can be expected in hardware performance. Other approaches to the problem of hydrogen bonding include a specific parameterization of PM3, known as PM3_{BP} [33] for nucleic acid base pairs and by a different type of modification of the core–core repulsion term [34, 35] that avoids the problems inherent in the use of Gaussian-modified terms.

Weak interactions

A weakness that semiempirical methods share with DFT is their inability to reproduce weak interactions (i.e. dispersion). This may be part of the problem with hydrogen bonding (see later), but is unlikely to be the whole cause. Apart from the obvious effect that van der Waals complexes are not bound, the lack of dispersion may also lead to an inherent size-consistency problem (see later). Dispersion energy can become a significant contribution to the stability of large molecules. A classical van der Waals potential is apparently included in PM5 as a simple two-center classical term [14].

Phosphorus

Even the extensive reparameterization apparently used to develop PM5 [14] was not able to improve the performance for phosphorus compounds. The solution to this problem is, however, known. MNDO/d performs significantly better than the s,p techniques for phosphorus, so that in this case the inclusion of d orbitals in the phosphorus basis is necessary. Phosphorus, however, is certainly not the only element for which this is true (see later). It is also now clear that some of the experimental data on which the PM3 parameterization (and possibly also that for PM5) was based are in error by roughly 140 kcal mol⁻¹ in the heat of formation [36]. Bogus experimental data of this type can skew parameterizations significantly, although this situation should be detectable by inspection of the results of the parameterization.

Transition metals

Transition metals remain the holy grail of NDDO-based semiempirical MO theory. An extension of MNDO/d to transition metals has not yet appeared. PM3-tm [15] is apparently a simple extension of PM3 with d orbitals to transition metals. However, it was only parameterized to reproduce X-ray structures without the usual energies, ionization potentials and dipole moments. Energies usually dominate the error function in semiempirical

parameterizations, so the success of this parameterization remains uncertain. Voityuk and Rösch [17] report good results for AM1(d) for molybdenum, but molybdenum remains the only metal for which this technique has been reported. However, our recent work on AM1* [37] suggests that the Voityuk–Rösch parameter set for molybdenum is of very high quality.

Perhaps surprisingly, the INDO-based methods SINDO/1 [13] and ZINDO [38] have been the most successful for transition metals. However, the latter is unsuitable for geometry optimizations, which seriously limits its usefulness. The success of these two methods, however, suggests strongly that good performance for transition metals can also be achieved with NDDO-based techniques.

Factors in the development of new methods

Experimental data

Tabulated values of experimental data are not always correct. Spurious experimental data points, especially for heats of formation, can dramatically affect the parameterization of a semiempirical method. A selection of 21 compounds for which we have performed either large basis set DFT [B3LYP/6-31+G(3df,2dp)//B3LYP/6-31+G(d) in this paper] or G2 [39] calculations in order to check the reliability of experimental data is given in Table 2.

Table 2 demonstrates several potential problems with experimental data, most of which can, however, be

solved by validation with higher levels of theory. The first example shown in Table 2 is cubane, which has been a consistent problem in semiempirical parameterizations, the calculated heat of formation being typically in error by about 40 kcal mol⁻¹, leading to some speculation about the quantity's accuracy. Our G2 result supports the experimental result and points to a systematic weakness of current MNDO-type methods. In cases such as this, it is important to confirm the reliability of the experimental datum before concluding that there is a fundamental flaw in the calculational method. In contrast, there are also plenty of examples of compounds for which the experimental data are very likely to be in error, sometimes by relatively large amounts. Examples of these shown in Table 2 are Cl₂O₂, HSO₃, SO₂F, SOF₃ and SOF₄, SCl₆, SO₂Cl, P₄O₁₀, (OCN)₃PO, PCl₄, F₂PO and Cl₂PO. The frequent occurrence of sulfur and phosphorus compounds in this exemplary list perhaps explains the poor performance of semiempirical methods for these compounds.

The second class of compounds is that for which there is more than one experimental measurement, and for which sometimes different values have been used for different parameterizations. Perhaps the most extreme example is P₄O₆, for which experimental heats of formation of -529.2 [40], -375.5 [41] and -378.01 kcal mol⁻¹ [42] have been reported. Thus, PM3, for which $\Delta H_f^0 = -529.2$ kcal mol⁻¹ was used in the parameterization, gives a heat of formation of -511 kcal mol⁻¹, whereas AM1 and MNDO/d give -321.5 and -375.9 kcal mol⁻¹, respectively. Engels et al. [36] re-

Table 2. Comparison of experimental and calculated heats of formation (kcal mol⁻¹) for selected compounds. *M*=MNDO, *Md*=MNDO/d, *A*=AM1, *P*=PM3

Formula	Compound	Used for Parameterizing	Exp. 1	Exp. 2	Exp. 3	G2	DFT	Rec.	Change	Comments
C ₈ H ₈	Cubane	M:A:P	148.7			148.1		148.7		G2, see also Refs. [58, 59, 60]
HCl ₂ ⁻	Hydrogen dichloride anion	P	-142.0			-101.4	-103.2	-100.2	-41.8	NIST
ClO ₃	O ₃ Cl	Md	37.0			50.5	53.1	43.1	-6.1	Ref. [61]
Cl ₂ O ₂	ClOOCl	M	12.7			29.7	32.8	29.7	-17.0	G2, this work
Cl ₂ O ₇	O ₃ ClOClO ₃	Md	65.0			82.1	102.9	61.5	3.5	Ref. [61]
CF ₂ Cl ₂	Difluorodichloromethane	M:A:P:Md	-114.1	-117.5	-117.9	-122.9	-111.8	?		
CFCl ₃	Fluorotrichloromethane	M:A:P:Md	-64.0	-64.1	-68.1	-73.0	-59.8	?		
HSO ₃	HOSO ₂	Md	-98.0			-59.3	-60.5	-59.3	-38.7	G2, this work
CH ₄ N ₂ S	Thiourea	A	-6.0			4.8	0.9	5.5	-11.5	NIST
F ₃ S	Sulfur trifluoride	P	-130.0			-103.6	-107.4	-116.7	-9.8	Ref. [62]
F ₂ S ₂	FSSF	M:P:Md	-54.5	-68.4	-80.4	-72.6	-72.2	-68.4	13.9	As used for MNDO/d
SO ₂ F	SO ₂ F	P:Md	-102.3	-113.2		-87.2	-86.0	-87.2	-15.1	G2, this work
SOF ₃	SOF ₃	P	-185.1			-137.2	-136.7	-137.2	-47.9	G2, this work
SOF ₄	SOF ₄	P:Md	-228.0	-235.5		-207.0	-195.5	-207.0	-21.0	G2, this work
SCl ₆	Sulfur hexachloride	P	-19.8			21.9	45.5	21.9	-41.7	G2, this work
SO ₂ Cl	SO ₂ Cl	P	-66.4			-45.9	-45.1	-45.9	-20.5	G2, this work
P ₄ O ₆	Phosphorus trioxide	A:P:Md	-529.2	-375.9	-378.01		-367.44	-375.9	-153.3	See Ref. [32]
P ₄ O ₁₀	Phosphorus pentoxide dimer	A:P:Md	-672.0	-694.1			-632.3	-632.3	-39.7	DFT, this work
O ₃ C ₃ N ₃ PO	(OCN) ₃ PO	A:Md	-158.6				-64.3	-64.3	-94.3	DFT, this work
OF ₂ P	OPF ₂	P	-213.6			-171.1	-171.5	-171.1	-42.5	G2, this work
PCl ₄	Phosphorous tetrachloride	P	-80.5			-55.9	-53.8	-55.9	-24.7	G2, this work
OPCl ₂	OPCl ₂	P	-109.9			-70.9	-71.3	-70.9	-39.0	G2, this work

^aAs discussed in the text, given the uncertainty of the calculations and experimental data, it is unclear what the recommended value should be

cently reported ab initio calculations on P_4O_6 from which a heat of formation around $-350 \text{ kcal mol}^{-1}$ can be deduced. Our DFT calculations give a value of $-367.4 \text{ kcal mol}^{-1}$. It is therefore extremely likely that PM3 was parameterized using an experimental datum that is in error by about $140 \text{ kcal mol}^{-1}$. Another such example, but less extreme, is FSSF. However, when there is relatively little spread in the experimental values, high-level calculations can only be valuable if the high-level methods are known to be reliable for the given class of compounds. The currently available techniques have, for instance, been parameterized with target values ranging from -114.1 to $-117.9 \text{ kcal mol}^{-1}$ for CF_2Cl_2 and from -64 to $-69 \text{ kcal mol}^{-1}$ for CFCl_3 . Such cases could lead to an inability to reproduce valid trends in the data, even though the magnitude of the error is small. As has been noted previously, there are unfortunately limitations to G2 and G3 methods as well, specifically in the case of halogenated compounds. Even the inclusion of a spin-orbit coupling term in the G2 and G3 theories does not completely correct this problem, making predictions for these compounds unreliable. There are thus still limits to the theoretical validation of experimental data.

The final class of compounds is that for which new experimental values are available (Cl-H-Cl^- , Cl_2O_7 , thiourea and SF_3 in Table 2). In most cases, but not all, the new experimental value agrees well with the G2 calculations.

The examples shown in Table 2 are admittedly among the most extreme that we have found. Nevertheless, it is clear that almost all experimental data must be validated if we are to improve the accuracy of current semiempirical techniques. There are some cases, such as the chlorofluorocarbons, for which we cannot assign reliable values, but theoretical validation can at least correct gross errors. Currently, however, we cannot create a dataset of heats of formation with a reliability better than about $\pm 3 \text{ kcal mol}^{-1}$. However, if asked, almost all computational chemists would answer that they would be very happy with a semiempirical method that performs as well as a moderately sized ab initio or DFT calculation, for example, B3LYP/6-31 + G(3df,2dp)//B3LYP/6-31 + G(d). However, the mean unsigned error in the calculated heats of formation at this level compared with experiment for our dataset of 382 silicon, phosphorus, sulfur and chlorine compounds is $11.1 \text{ kcal mol}^{-1}$, compared with $11.8 \text{ kcal mol}^{-1}$ for PM3. That much of this error can be attributed to the experimental data is demonstrated by the fact that the mean unsigned error between G2 and DFT for the 245 compounds of our dataset for which we currently have both data available is only $5.3 \text{ kcal mol}^{-1}$, compared with $12.2 \text{ kcal mol}^{-1}$ for PM3.

Parameterization for heats of formation

An important approximation lies at the heart of most parameterizations of semiempirical techniques. What, of

course, is calculated is the Born–Oppenheimer total energy, E_{TOT} . This is converted to a Born–Oppenheimer total binding energy, E_{B} , by subtracting the Born–Oppenheimer energies of the constituent atoms:

$$E_{\text{B}} = E_{\text{TOT}} - \sum_n E_n, \quad (4)$$

where n is the number of atoms in the molecule and E_n is the Born–Oppenheimer energy of atom n .

In a breathtaking approximation, E_{B} is then equated to the enthalpy of atomization at 298 K, $\Delta H_{\text{ATOM}}^{298}$, which is given by

$$\Delta H_{\text{ATOM}}^{298} = \Delta H_{\text{f}}^0 - \sum_n \Delta H_{\text{f}}^{298}(n), \quad (5)$$

where ΔH_{f}^0 is the standard heat of formation of the molecule and $\Delta H_{\text{f}}^{298}(n)$ is the heat of formation of atom n . The calculated heat of formation is then given as

$$\Delta H_{\text{f}}^0 = E_{\text{TOT}} - \sum_n [E_n - \Delta H_{\text{f}}^{298}(n)]. \quad (6)$$

Implicit in this equation is the assumption that the zero-point vibrational energy and the thermal energy of a molecule at 298 K can be treated as the sum of a series of increments assigned to each atom. Even if this assumption is justified, the parameterization procedure must be able to compensate for the implicit vibrational contributions present in ΔH_{f}^0 but not in E_{B} . Hicks and Thiel [43] tested this approximation in 1986 for a limited (36 compounds) set of small hydrocarbons for which experimental frequencies were available and found that it did not introduce additional errors. However, advances in hardware and software and the advent of hybrid DFT techniques that allow the accurate calculation of vibrational frequencies for a wide set of functional groups now allow us to investigate this question more closely.

We calculated the B3LYP/6-31 + G^(d) [44, 45] vibrational frequencies for a dataset of 273 compounds containing H, C, N and O and tested the validity of the atom-additivity assumption by multiple linear regression. The resulting regression equation

$$\text{ZPE} = 6.801N_{\text{H}} + 3.672N_{\text{C}} + 2.781N_{\text{N}} + 2.459N_{\text{O}}, \quad (7)$$

where N_{H} , N_{C} , N_{N} and N_{O} are the numbers of hydrogen, carbon, nitrogen and oxygen atoms in the molecule, respectively, reproduces the DFT zero-point energies with a standard deviation of $2.3 \text{ kcal mol}^{-1}$, superficially confirming that ΔH_{f}^0 can be used for parameterization without much loss of accuracy.

The correlation between B3LYP/6-31 + G^(d) calculated zero-point energies and those given by Eq. (7) is shown in Fig. 1. The figure clearly shows that there is a problem with the approximation embodied by Eq. (7). The slope of the line is 5% too low and the intercept is around $3.6 \text{ kcal mol}^{-1}$. In fact, if we allow a constant term in the regression equation, the standard deviation

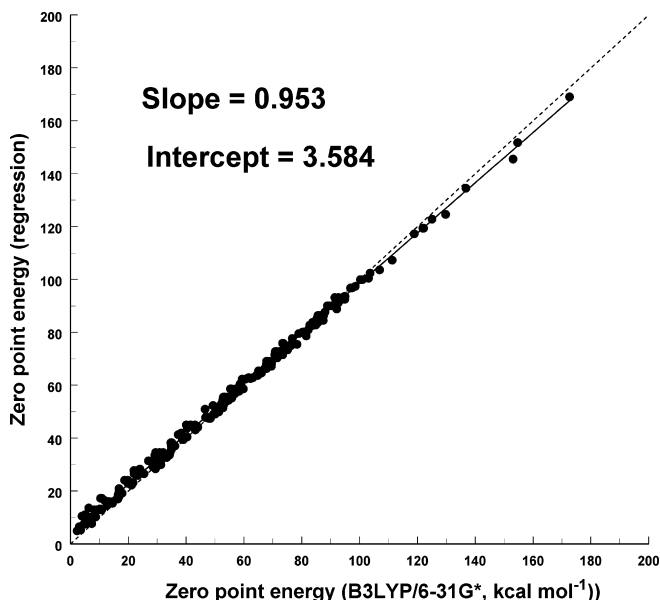


Fig. 1. The correlation between B3LYP/6-31 + G^(d) calculated zero-point energies and those given by Eq. (7) for the 273 compounds of the C, H, N, O dataset. The *solid line* shows the regression between the two and the *dashed line* perfect agreement

between the B3LYP and the additive zero-point energies decreases by about 1 kcal mol⁻¹:

$$\text{ZPE} = -4.588 + 7.169N_{\text{H}} + 3.900N_{\text{C}} + 3.712N_{\text{N}} + 2.856N_{\text{O}}. \quad (8)$$

The standard deviation given by Eq. (8) is 1.27 kcal mol⁻¹, the slope 1.000 and the intercept 0.000. Fig. 1 therefore suggests that the heat of formation approximation should lead to the highest accuracy around the center of the size (around the zero-point energy) range of the parameterization dataset, but that there should be systematic deviations for very large or very small compounds (i.e. those with zero-point energies close to zero or near or above the upper limit of the parameterization dataset).

A further feature of the data shown by Fig. 2 is that certain classes of compounds, in this case all those with a nitro group, deviate systematically from the regression equation. This must lead to systematic deviations in any parameterization based on the heat of formation or, even worse, to a skewing of the optimized parameters to accommodate the outlying class of compounds. As shown later, this is particularly relevant for nitro compounds.

Moreover, such a parameterization is applicable only for cases where the molecular structure is a minimum and thus is not applicable for other points on the potential-energy surface, such as transition states.

Parameterization for E_B

The dataset of 273 C, H, N and O compounds was used for two identical parameterizations, except that in one

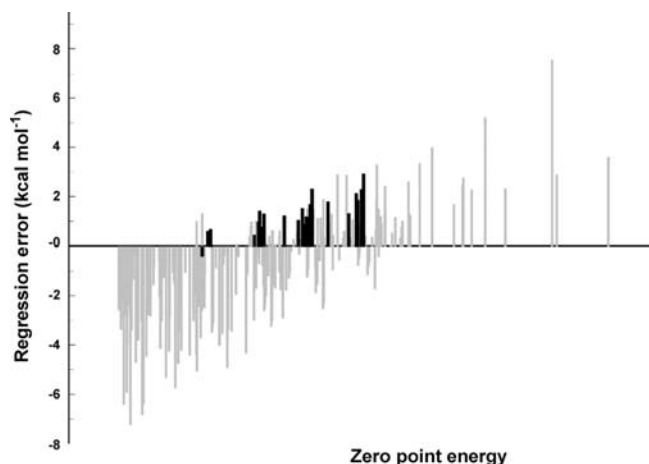


Fig. 2. The residuals between the B3LYP zero-point energies and those calculated by Eq. (7) against the value of the zero-point energy (B3LYP). Nitro compounds are shown as *black bars*, all others as *gray bars*

case the target energies were heats of formation and in the other E_B values. The results of the two parameterizations are shown in Table 3 and in Fig. 3.

Table 3 suggests that there is no advantage in using E_B , rather than ΔH_f^0 . This result agrees with those of Hicks and Thiel [43]. The mean unsigned error and the standard deviation between experimental and calculated energies are actually larger for the E_B dataset than for ΔH_f^0 . However, the most negative deviation and the mean signed error are all smaller for the E_B dataset. It is tempting to interpret these numbers as meaning that the E_B parameterization is “physically more reasonable” (as it should be purely on the basis of the physics behind it), but this is not really justified. However, a careful analysis of the data reveals patterns that suggest that using E_B is after all the better strategy.

Fig. 3 allows a comparison of the errors for individual compound in the two datasets. The points have been color coded from light gray (low) to black (high) according to the zero-point energies calculated for the corresponding compounds. In this case the zero-point energies are essentially a measure of the size of the compound. The vertical and horizontal reference lines give roughly the limits of the errors attributable to the zero-point-energy error for this dataset. Figure 3 allows us to draw several important conclusions.

Firstly, the errors given by the heat of formation model correlate fairly well ($r^2 = 0.75$) with those given by the E_B model. The range of errors is roughly a factor of 5 larger than that expected to arise from the heat of formation approximation. The compounds that give the largest improvement between the ΔH_f^0 and E_B models are given in Table 4.

The largest outliers of the ΔH_f^0 model are improved significantly in the E_B model. The error for cubane, for instance, is reduced by 10 kcal mol⁻¹, but still remains the largest outlier. Both states of O₂, peroxides and

Table 3. Statistics of the deviations between the parameterizations for ΔH_f^0 and E_B . All energies are given in kcal mol⁻¹

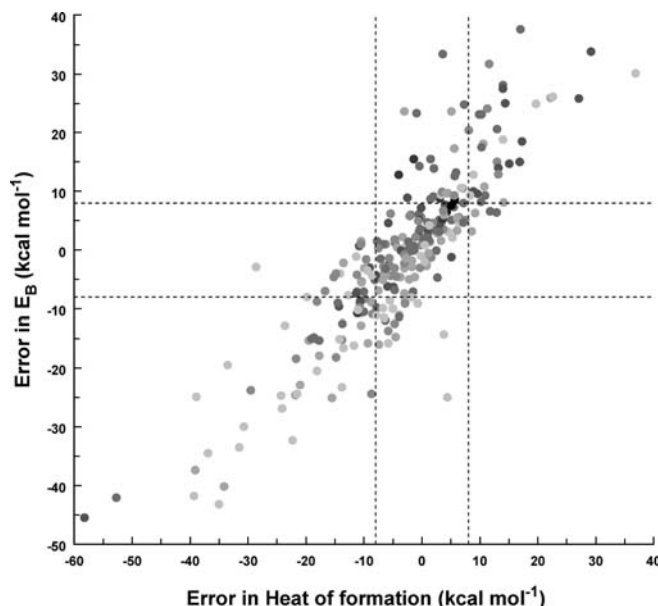
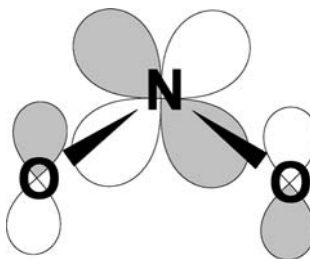
	ΔH_f^0	E_B
Most positive deviation	36.9	37.6
Most negative deviation	-58.2	-45.5
Mean unsigned error	9.0	10.3
Mean signed error	-3.3	-0.9
Standard deviation	12.4	14.0

compounds in which nitrogen is bound to three oxygens all improve significantly, as do six-membered nitrogen aromatics, anhydrides, lactones and cyanides. Many of these compounds are well-known problem cases for semiempirical theory. More interesting, however, are the compounds that become significantly worse in the E_B model than in the ΔH_f^0 model; these are shown in Table 5.

Aromatic compounds, water and the hydronium ion, hydrazine and two rather puzzling examples, ethylene glycol and glycerine trinitrate, all give larger errors in the ΔH_f^0 model, but all other compounds in this class contain at least one nitro group. This is clearly because the nitro compounds are outliers in the zero-point-energy correlation, as shown in Fig. 2. One of the great ironies of semiempirical theory is that a large number of nitro compounds were placed in the PM3 parameterization set [4, 5] in order to improve performance for these compounds; however, these compounds introduce significant errors if the ΔH_f^0 approximation is used. In the case when the ΔH_f^0 approximation is used, it compensates the inherent tendency of *s,p*-basis NDDO theory to underestimate the stability of these compounds by a systematic deviation from the additive zero-point-energy scale.

What is the reason for the inability of *s,p*-basis NDDO theory to describe nitro groups? A clue is given by ab initio calculations with and without *d* orbitals in the basis set. Although such comparisons have in the past caused long specious discussions as to “whether *d* orbitals are involved in bonding”, for instance in sulfur compounds [46], they do provide strong evidence as to whether polarization functions are necessary to describe particular bonding patterns correctly. The N–O bond lengths in nitromethane, for instance, are calculated to be 0.03 Å longer at RHF/6-31G^(d) than at RHF/6-31G. The total *d*-orbital population at nitrogen is 0.3e⁻ using a Mulliken population analysis [47] and 0.1e⁻ using a natural population analysis [48], far higher than for the carbon and the oxygens. Qualitatively, it is easy to see that a strong π contribution of a nitrogen *d* orbital to the highest π MO of the nitro group can provide significant stabilization, as shown in Scheme 1.

If the previous comments are true, the results for carboxylate anions should also become worse when E_B is used rather than the ΔH_f^0 approximation. This is found to be the case: the error for the formate and acetate anions increases from 3.3 and 10.6 kcal mol⁻¹, respectively for the ΔH_f^0 approximation to 9.4 and 18.1 using E_B .

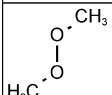
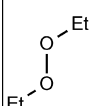
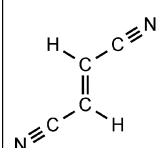


**Fig. 3.** Scatter plot of the errors between calculated and experimental energies for ΔH_f^0 and E_B . The target energies were used for two different parameterization runs in which all other parameterization data and optimization variables were the same. The vertical and horizontal reference lines give roughly the limits of the errors attributable to the zero-point-energy error for this dataset**Scheme 1.** Possible involvement of nitrogen polarization functions in the bonding in nitro compounds


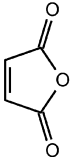
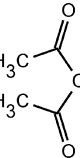

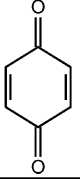

The idea that *d* orbitals might be necessary for nitrogen (and by analogy carbon and oxygen) is attractive from the known weaknesses of current methods, although it would mean considerable extra computational expense. The problems outlined earlier concerning the pyramidity of nitrogen centers, for instance, are known from the early days of ab initio theory and were cured by adding polarization functions to the basis set [49].

Treatment of the nucleus and nonvalence electrons

The treatment of the nucleus and its associated nonvalence electron shells has been the subject of many modifications over the years. The traditional approximation as used in most NDDO-based techniques simply treats the frozen core as a nuclear charge reduced in magnitude by the number of nonvalence electrons. It has

Table 4. Compounds for which the E_B parameterization performs significantly better than that for ΔH_f^0 . ΔE_B and Δ_f^0 are the errors (kcal mol⁻¹) for the E_B and ΔH_f^0 parameterizations, respectively

Compound	ΔE_B	$\Delta \Delta H_f^0$
³ O ₂	-25	-39
¹ O ₂	-20	-34
N ₂ O ₅	-8	-20
NO ₃ ⁻	-1	-11
CH ₃ ONO ₂	-3	-10
C ₂ H ₅ ONO ₂	0	-8
	-7	-17
	1	-11
N≡C—C≡C—C≡N	-13	-24
	-2	-14
N≡C—C≡N	-4	-14
	-2	-10
	-5	-15

	-9	-18
	2	-11
	1	-11
	-4	-11
	-4	-15
	-42	-53

been clear since the development of MNDO [6, 7] that this approximation leads to poor results and so the core-core term is traditionally modified. More specific modifications were introduced for MNDO/H [8], AM1 [2, 3] and PM3 [4, 5] in the form of varying numbers of Gaussian functions and for MNDO/d [9, 10] modifications of the core-core term for specific combinations of elements are used. The latter technique appears to give better results, but in some ways represents a step back towards the element-pair specific parameterization of MINDO/3 [50].

Probably the most promising improvement to current theories in this respect is to use effective core potentials [51] as, for instance, in SINDO/1 [13] or in the OM n techniques introduced by Thiel, Kolb and Weber [11].

Dispersion




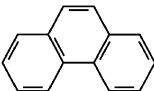
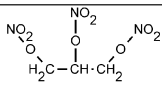
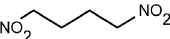
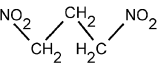
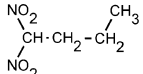



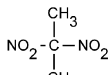





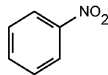

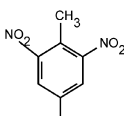
Intermolecular van der Waals interactions are obviously important for weak complexes. However, intramolecular dispersion energies may also become significant for large molecules. This is demonstrated by the data shown in Fig. 3. There is a clear trend for compounds with low zero-point energies (i.e. small molecules, those color coded in light gray) to be cal-

culated to be too stable, whereas larger molecules (dark gray to black points) cluster in the region where both the ΔH_f^0 and the E_B models calculate the compounds to be too unstable.

There are several potential reasons for this, but the most likely is the effect that we know to be missing from the theory, dispersion. However, rather than use a purely classical two-center potential, as has apparently been added to PM5 [14], we prefer to add a London-type term to the Hamiltonian. This can either be added post-self-consistent field (SCF), as in our current implementation, or included in the SCF iterations. Our dispersion correction uses our parameterized version [52] of the variational polarizability treatment suggested by Rivail and coworkers [53]. We have introduced a partitioning scheme that allows us to calculate additive atomic polarizabilities [54] and can easily be extended to "atomic orbital polarizabilities" [55]. While this additive polarizability model is arbitrary in the same way that population analyses are, it does provide useful access to a dispersion correction via the London formula for the interaction energy between two atoms A and B [56]:

$$U_{\text{disp}}^{(6)} \approx -\frac{U_A U_B}{4(U_A + U_B)} T_{\alpha\beta} T_{\alpha\beta} \overline{\alpha^A \alpha^B}, \quad (9)$$

Table 5. Compounds for which the E_B parameterization performs significantly worse than that for ΔH_f^0 . ΔE_B and Δ_f^0 are the errors (kcal mol⁻¹) for the E_B and ΔH_f^0 parameterizations, respectively

Compound	ΔE_B	$\Delta \Delta H_f^0$
H ₂ O	-14	4
H ₃ O ⁺	-25	4
H ₂ N-NH ₂	-15	-5
	24	-9
	6	0
	9	-3
	16	-1
	23	-1
	25	14
	23	10
	28	14
	17	6
	20	8
	23	10
	28	14
	24	5
	31	12
	38	17
	24	-3
	14	0
	16	2
	25	7
	33	4

where $\bar{\alpha}$ represents the mean polarizability of atom A or B, $T_{\alpha\beta}$ is an interaction tensor and the factors U can be estimated using the Slater–Kirkwood approach

$$U_A \approx \sqrt{\frac{N_A}{\bar{\alpha}_A}}, \quad (10)$$

where N_A is the effective number of valence electrons in atom A and α_A is its mean polarizability¹. We [55]

¹For exact definitions of T and a detailed description of the London formula and alternatives, see Ref. [57]

have implemented such a scheme as a post-SCF addition to the electronic energy within the usual MNDO-like framework and are investigating a self-consistent scheme in which the dispersion interaction is built into the SCF iterations. An important advantage of this ansatz compared with simple isotropic two-body classical potentials is that it takes the effects of polarizability anisotropy into account. The dispersion interactions must be scaled at close range, where the London formula breaks down, using a scaling function, $f(r)$, suggested by Elstner et al. [58] and given by



Fig. 4. Calculated structures and energies of the methane:ethane dimer

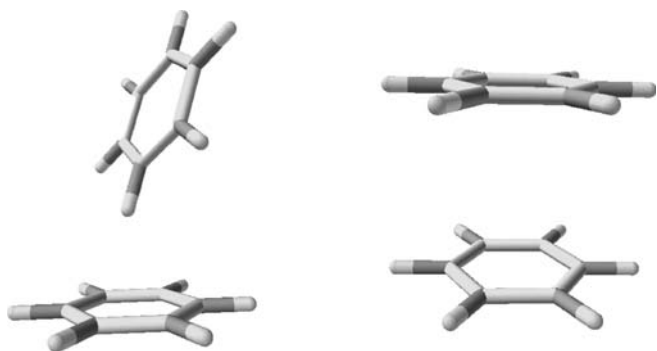


Fig. 5. Calculated structures and energies of the benzene dimer

$$f(r) = \left\{ 1 - \exp \left[-3.0 \left(\frac{r}{r_0} \right)^7 \right] \right\}^4, \quad (11)$$

where r is the distance and r_0 is defined by a cubic mean rule based on standard radii for the elements.

Thus, after a simple parameterization, the energies and geometries of hydrocarbon van der Waals complexes, for which high level ab initio data are available, can be reproduced well. The minimum-energy geometries for two different orientations of the methane:ethane dimer compared to published ab initio data [59] are shown in Fig. 4 and the geometries and energies of two alternative conformations of the benzene dimer, again compared to ab initio data [60], are shown in Fig. 5.

The influence of d orbitals

Thiel and Voityuk [9, 10] have demonstrated the need for d orbitals in the basis in order to describe heavier elements correctly. This need is confirmed impressively by the poor performance of PM5, which presumably does not have d orbitals, for phosphorus compounds [14].

One-electron energies

In principle, the one-electron energies for s , p , and d orbitals for minimal basis set NDDO calculations need not be parameterized, as they are known spectroscopically. However, these parameters have traditionally been

treated as variables in the modified NDDO methods, so they may be considerably different from their experimental values. An important example is hydrogen, for which the experimental value for U_{ss} is -13.59844 eV, compared with parameterized values shown in Table 6.

Thus, in all cases U_{ss} for hydrogen is significantly (from 2.2 to 0.5 eV) less negative than the experimental value. This makes hydrogen a worse acceptor than it should be. Hydrogen that is a bad acceptor cannot be expected to make proper hydrogen bonds, a systematic weakness of current NDDO-based semiempirical techniques. However, this does not seem to be the reason for the lack of hydrogen bonds in MNDO (see later).

It is useful to consider the history of the hydrogen-bond problem in semiempirical MO-theory. MNDO ($U_{ss} = -11.9$ eV) was unable to reproduce hydrogen bonds at all. In order to correct this problem without altering the rest of the parameterization, Burschkin and Isaev [8] introduced additional Gaussian terms into the core-core energy to give MNDO/H. This modification was adopted by Dewar for AM1. Although AM1 ($U_{ss} = -11.4$ eV) yielded hydrogen bonds, it did not reproduce the linear hydrogen-bond linkage that results from the lone pair to $\sigma^*_{(H-X)}$ antibonding donor-acceptor interaction. We propose that the hydrogen-bonding energy in AM1 results largely from the (anisotropic) modification in the core-core energy. When the geometry of hydrogen bonds was emphasized in the PM3 parameterization, U_{ss} decreased to -13.1 eV, the lowest value found for all common NDDO-based methods.

However, U_{ss} proves not to be the sole culprit of this failure when we investigate hydrogen-bonding behavior more closely. We removed the Gaussian core-core functions for oxygen and hydrogen from the standard AM1 parameterization for this review. The structure obtained for the water dimer is shown in Fig. 6.

The quasi-linear geometry of the $O \cdots H-O$ linkage is reproduced (the $O \cdots H-O$ angle is 180.0°) and the $O \cdots H$ distance is short (1.349 Å; the best current estimate is about 1.95 Å [61]). The dimerization energy is 10 kcal mol^{-1} , considerably larger than the most recent experimental value [61] of 4.85 kcal mol^{-1} . Thus, the Gaussian terms in AM1 weaken hydrogen bonds, and do not strengthen them, and make AM1 give the wrong geometry. The situation is similar for PM3: removing the Gaussian terms from the standard PM3 parameter set shortens the $O \cdots H$ distance to 1.558 Å and increases

Table 6. $U_{ss}(\text{H})$ and $\zeta_s(\text{H})$ values for different parameterizations and their performance for hydrogen bonds

Parameterization	$U_{ss}(\text{H})$ (eV)	$\zeta_s(\text{H})$	Hydrogen bonds
MNDO	-11.906	1.332	Absent
MNDO/c	-12.114	1.360	Absent
AM1	-11.396	1.188	Bifurcated
PM3	-13.073	0.968	Geometry correct, too weak

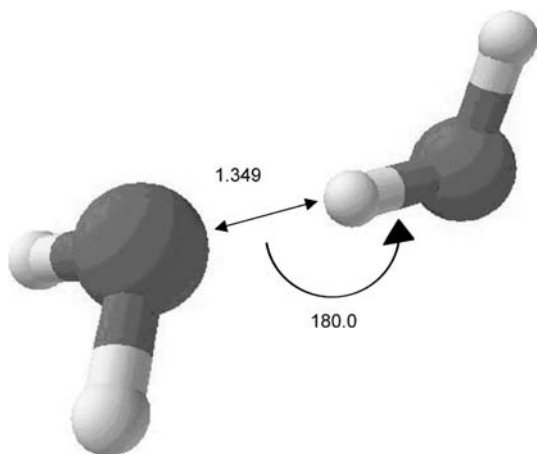


Fig. 6. The structure obtained for the water dimer using the standard AM1 parameters but with the experimental value for U_{ss} for hydrogen and without any Gaussian terms to modify the core-core potential

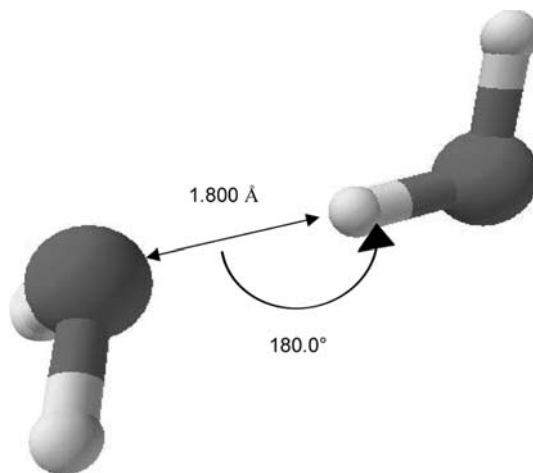


Fig. 7. The structure obtained for the water dimer using the standard MNDO parameters for oxygen with the modified hydrogen parameters specified in the text

the dimerization energy to $-6.5 \text{ kcal mol}^{-1}$. Thus, paradoxically, the Gaussian terms introduced to be able to reproduce hydrogen bonds in MNDO/H actually weaken the hydrogen bonding in PM3 and AM1.

What then is the reason why MNDO cannot reproduce hydrogen bonds (even if we modify the core-core terms in the same way as in MNDO/d) but why AM1 and PM3 actually overestimate hydrogen-bond strengths in their “virgin” incarnations without Gaussian functions? As already mentioned, our first candidate was U_{ss} for hydrogen; however, this proved not to be the only important parameter. Rather, the exponent of the hydrogen $1s$ orbital follows (or causes) the trends in hydrogen-bonding behavior for the different methods. Table 6 also shows the values used for ζ_s for hydrogen in different parameterizations and relates them to how the method reproduces hydrogen bonds.

The data appear to reveal a trend. In order to test this, we used the spectroscopic value for $U_{ss}(\text{H})$, reduced the value of ζ_s for hydrogen in the standard MNDO-parameterization to 1.0 and compensated by adding an MNDO/d-like term to the core-core repulsion, which is shown in Eq. (12):

$$E^{\text{core}}(i-j) = Z_i Z_j \rho_{ss}^0 \exp(-\alpha_{ij} r_{ij}). \quad (12)$$

We set the value of α for O–H interactions to 3.0. The resulting structure for the water dimer is shown in Fig. 7. The O–H–O angle (180°) and the H...O distance (1.80 Å) are now quite reasonable. The dimerization energy is $-4.7 \text{ kcal mol}^{-1}$.

Furthermore, we can tune the structure of the water dimer simply by changing the value of ζ_s for hydrogen. Very compact hydrogen $1s$ orbitals prevent the formation of hydrogen bonds, which are reproduced correctly with a more diffuse hydrogen basis set. This observation is consistent with Jug’s finding that p orbitals on hydrogen (which are more diffuse than s) reproduce

hydrogen bonding. Our results suggest, however, that we may not need polarization functions on hydrogen. Our conclusion is that there is no fundamental reason why semiempirical methods should not perform well for hydrogen bonds.

Conclusions and outlook

In this article, we have tried to examine the current situation and the potential of semiempirical methods objectively. Our discussion suggests that there are no fundamental reasons why semiempirical MO theory is inherently unable to treat hydrogen bonding or transition metals, the two main problems faced by current techniques. On the contrary, attributing the poor performance of MNDO for hydrogen bonds to a deficit in the core-core term seems to have created the “problem” hydrogen bonding. The Gaussian terms in the AM1 and PM3 core-core terms do not help make hydrogen bonds, they actually weaken them. Similarly, the success of ZINDO [38], SINDO/1 [13] and more recently Voityuk and Rösch’s molybdenum parameters for AM1(d) [17] suggest that transition metals may also not be inherently difficult.

Next generation methods, however, will need some improvement over the current MNDO-generation. We suggest that the following factors may play a significant role in the next generation of NDDO-based methods:

1. The orthogonalization corrections introduced by Thiel and his groups [11] for the OM n methods are an important factor in improving performance for rotation barriers, conformational equilibria, structural details and many other aspects.
2. Extending the basis set with d orbitals will certainly be necessary for elements heavier than silicon, and may considerably increase the accuracy when added to some first-row elements.

3. Ideally, next-generation methods will pay more attention to the magnitudes of the one-electron energies of the atomic orbitals, and perhaps even fix them at the spectroscopic values. There is no hard evidence that this is essential, but current methods make, for instance, hydrogen up to 2 eV too electropositive, which must affect performance for all other elements. The success of ZINDO [38] and the AM1(d) molybdenum parameterization [17], both of which use spectroscopic values, supports this view.
4. Effective core potentials, already in use for SINDO/1 [13] and OM*n* [11], should improve performance for heavier elements.
5. Some form of extra treatment of dispersion will be necessary, not only to improve performance for weak interactions, but also to remove the inherent size inconsistency outlined earlier.
6. It may be of advantage for accurate methods to parameterize for E_B , rather than for ΔH_f^0 .
7. Experimental data may already not be accurate enough for highly parameterized methods, especially for the elements of the second row. Careful validation, at least for geometries and energies, using either DFT or an extrapolation scheme like G2 or G3, is absolutely essential for future accurate methods.

We feel strongly that a next-generation NDDO technique can best be developed by a concerted effort of the relatively few groups working in this area. To this end, we have undertaken to publish all of our methods completely and will also make our parameterization data freely available.

Acknowledgements This work was supported by the Fonds der Chemischen Industrie. C.S. thanks the Alexander-von-Humboldt-Stiftung for the award of a Fellowship. We thank Jan Andzelm, George Fitzgerald, Kenny Merz, Walter Thiel and Darrin York for fruitful discussions at various stages of the conception of this manuscript and the referees for helpful comments.

References

1. Pople JA, Santry DP, Segal GA (1965) *J Chem Phys* 43:S129–S135
2. Dewar MJS, Zoebisch EG, Healy EF, Stewart JJP (1985) *J Am Chem Soc* 107:3902–3909
3. Holder AJ (1998) In: Schleyer PvR, Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer HF III, Schreiner PR (eds) *Encyclopedia of computational chemistry*, vol 1. Wiley, Chichester, p 8
4. Stewart JJP (1989) *J Comput Chem* 10:209–221
5. Stewart JJP (1998) In: Schleyer PvR, Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer HF III, Schreiner PR (eds) *Encyclopedia of computational chemistry*, vol 3. Wiley, Chichester, p 2080
6. Dewar MJS, Thiel W (1977) *J Am Chem Soc* 99:4899–4907
7. Thiel W (1998) In: Schleyer PvR, Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer HF III, Schreiner PR (eds) *Encyclopedia of computational chemistry*, vol 3. Wiley, Chichester, p 1599
8. Burstein KY, Isaev AN (1984) *Theor Chim Acta* 64:397–401
9. Thiel W, Voityuk AA (1992) *Theor Chim Acta* 81:391–404; (b) Thiel W, Voityuk AA (1996) *Theor Chim Acta* 93:315–315; (c) Thiel W, Voityuk AA (1994) *Int J Quantum Chem* 44:807; (d) Thiel W, Voityuk AA (1994) *J Mol Struct* 313:141–154; (e) Thiel W (1998) In: Schleyer PvR, Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer HF III, Schreiner PR (eds) *Encyclopedia of computational chemistry*, vol 3. Wiley, Chichester, p 1604
10. Thiel W, Voityuk AA (1996) *J Phys Chem* 100:616–626
11. (a) Kolb M, Thiel W (1993) *J Comput Chem* 14:775–789; (b) Weber W, Thiel W (2000) *Theor Chem Acc* 103:495–506
12. Pople JA, Beveridge DL, Dobosh PA (1967) *J Chem Phys* 47:2026–2033
13. (a) Ahlswede B, Jug K (1999) *J Comput Chem* 20:563–571; (b) Ahlswede B, Jug K (1999) *J Comput Chem* 20:572–578, and references therein
14. <http://www.cachesoftware.com/mopac/index.shtml>
15. (a) <http://www.wavefun.com/>; (b) <http://kripto.qui.ub.es/~rbosque/PM3tm.html>
16. (a) Dewar MJS, Jie C, Yu J (1993) *Tetrahedron* 49:5003–50038; (b) Dewar MJS, Jie C, Yu J (1993) *Tetrahedron* 49:7393; (c) Holder AJ (1998) In: Schleyer PvR, Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer HF III, Schreiner PR (eds) *Encyclopedia of computational chemistry*, vol 4. Wiley, Chichester, p 2542
17. Voityuk AA, Roesch N (2000) *J Phys Chem A* 104:4089–4094
18. Hutter MC, Reimers JR, Hush NS (1998) *J Phys Chem B* 102:8080–8090
19. Brothers EN, Merz KM Jr (2002) *J Phys Chem B* 106:2779–2785
20. Repasky MP, Chandrasekhar J, Jorgensen WL (2002) *J Comput Chem* 23:1601–1622
21. Clark T (2000) *J Mol Struct (THEOCHEM)* 530:1–10
22. Elstner M, Frauenheim T, Kaxiras E, Seifert G, Suhai S (2000) *Phys Status Solidi* 217:357–376
23. Warshel A, Weiss RM (1980) *J Am Chem Soc* 102:6218–6226
24. Tomasi J, Persico M (1994) *Chem Rev* 94:2027–2094
25. Cramer CJ, Truhlar DJ (1999) *Chem Rev* 99:2161–2200
26. Bakowies D, Thiel W (1996) *J Phys Chem* 100:10580–10594
27. Aqvist J, Warshel A (1993) *Chem Rev* 93:2523–2544
28. Field MJ, Bash P, Karplus M (1990) *J Comput Chem* 11:700–733
29. Pan W, Lee T-S, Yang W (1998) *J Comput Chem* 19:1101–1109
30. Van Der Vaart A, Gogonea A, Dixon SL, Merz KM Jr (2000) *J Comput Chem* 21:1494–1504
31. Stewart JJP (1996) *Int J Quantum Chem* 58:133–146
32. Jug K, Geudtner G (1993) *J Comput Chem* 14:639–646
33. Sherer EC, York DM, Cramer CJ (2003) *J Comput Chem* 24:57–67
34. Bernal-Uruchurtu MI, Martins-Costa MTC, Millot C, Ruiz-López MF (2000) *J Comput Chem* 21:572–581
35. Bernal-Uruchurtu MI, Ruiz-López MF (2000) *Chem Phys Lett* 330:118–124
36. Engels B, Valentim AR, Peyerimhoff SD (2001) *Angew Chem Int Ed Engl* 40:378–381
37. Winget P, Selçuki C, Horn AHC, Martin B, Clark T (2003) *J Mol Model*: available online.
38. Zerner MC (1991) In: Lipkowitz KB, Boyd DB (eds) *Reviews in computational chemistry*, vol 2. VCH, New York, p 45, and references therein
39. Curtiss LA, Raghavachari K, Pople JA (1995) *J Chem Phys* 103:4192–4200
40. Chase MW Jr (1998) *J Phys Chem Ref Data*, monograph 9, pp 1–1951
41. Pilcher G (1990) In: Hartly FR (ed) *The chemistry of organophosphorus compounds*, vol 1. Wiley, New York, pp 127–136
42. Muenow DW, Uy OM, Margrave JL (1970) *J Inorg Nucl Chem* 32:3459–3467
43. Hicks MG, Thiel W (1986) *J Comput Chem* 7:213–218
44. Stephens PJ, Devlin FJ, Chabalowski CF, Frisch MJ (1994) *J Phys Chem* 98:11623–11627
45. Hariharan PC, Pople JA (1972) *Chem Phys Lett* 66:217–219

46. Kwart H, King KG (1977) In: *Reactivity and structure. Concepts in organic chemistry, vol 3. d-orbitals in the chemistry of silicon, phosphorus and sulfur.* Springer, Berlin Heidelberg, New York
47. Mulliken RS (1955) *J Chem Phys* 23:1833–1840
48. Reed AE, Weinstock RB, Weinhold F (1985) *J Chem Phys* 83:735–746
49. Alcamí M, De Paz JLG, Yanez M (1989) *J Comput Chem* 10:468–478
50. Bingham RC, Dewar MJS, Lo DH (1975) *J Am Chem Soc* 97:1285–1293
51. Kahn LR, Baybutt P, Truhlar DG (1976) *J Chem Phys* 65:3826–3853
52. Schürer G, Gedeck P, Gottschalk M, Clark T (1999) *Int J Quantum Chem* 75:17–31
53. (a) Rivail JL, Carter A (1978) *Mol Phys* 36:1085–1097; (b) Rinaldi D, Rivail JL (1974) *Theor Chim Acta* 32:243–251
54. Martin B, Gedeck P, Clark T (2000) *Int J Quantum Chem* 77:473–497
55. Martin B (2003) PhD thesis. Universität Erlangen–Nürnberg
56. London F (1937) *Trans Faraday Soc* 33:8–26
57. Stone AJ (1996) *The theory of intermolecular forces.* Clarendon, Oxford
58. Elstner M, Hobza P, Frauenheim T, Suhai S, Kaxiras E (2001) *J Chem Phys* 114:5149–5155
59. Tsuzuki S, Honda K, Uchimura T, Tanabe K (1999) *J Phys Chem A* 103:8265–8271
60. Tsuzuki S, Lüthi HP (2001) *J Chem Phys* 114:3949–3957
61. Goldman N, Fellers RS, Brown MG, Braly LB, Keoshian CJ, Leforestier C, Saykally RJ (2002) *J Chem Phys* 116:10148–10163, and references therein
62. Jursic BS (2000) *THEOCHEM* 499:137–140
63. Castano O, Notario R, Abboud JLM, Gomperts R, Palmeiro R, Frutos LM (1999) *J Org Chem* 64:9015–9018
64. Ball DW (1996) *THEOCHEM* 364:183–188
65. Sicre JE, Cobos CJ (2003) *J Mol Struct (THEOCHEM)* 620:215–226
66. Herron JT (1987) *J Phys Chem Ref Data* 16:1–6
67. Kolb M, Thiel W (1993) *J Comput Chem* 14:37–44
68. Thiel W (1982) *QCPE Bull* 2:63
69. Dewar MJS, Rzepa HS (1978) *J Am Chem Soc* 100:777–784
70. Dewar MJS, McKee ML (1977) *J Am Chem Soc* 99:5231–5241
71. Dewar MJS, Rzepa HS (1978) *J Am Chem Soc* 100:58–67
72. Voityuk AA (1987) *Zh Strukt Khim* 28:128–131
73. Davis LP, Guidry RM, Williams JR, Dewar MJS, Rzepa HS (1981) *J Comput Chem* 2:433–445
74. Dewar MJS, Healy EF, Stewart JJP, Friedheim J, Grady GL (1986) *Organometallics* 5:375–379
75. Dewar MJS, McKee ML, Rzepa HS (1978) *J Am Chem Soc* 100:3607
76. Dewar MJS, Reynolds CH (1986) *J Comput Chem* 7:140–143
77. Dewar MJS, Rzepa HS (1983) *J Comput Chem* 4:158–169
78. Havlas Z, Nick S, Bock H (1992) *Int J Quantum Chem* 44:449–467
79. Bliznyuk AA, Voityuk AA (1988) *Zh Strukt Khim* 29:156–157
80. Dewar MJS, Merz Jr KM (1986) *Organometallics* 5:1494–1496
81. Dewar MJS, Grady GL, Healy EF (1987) *Organometallics* 6:186–189
82. Dewar MJS, Healy EF (1983) *J Comput Chem* 4:542–551
83. Dewar MJS, Grady GL, Stewart JJP (1984) *J Am Chem Soc* 106:6771–6773
84. Dewar MJS, Healy EF, Stewart JJP (1984) *J Comput Chem* 5:358–362
85. Dewar MJS, Grady GL, Merz KM Jr, Stewart JJP (1985) *Organometallics* 4:1964–1966
86. Dewar MJS, Holloway MK, Grady G, Stewart JJP (1985) *Organometallics* 4:1973–1980
87. Thiel W (1981) *J Am Chem Soc* 103:1413–1420
88. Dewar MJS, Jie C, Ziebis EG (1988) *Organometallics* 7:513–521
89. Dewar MJS, Ziebis EG (1988) *THEOCHEM* 180:1–21
90. Dewar MJS, Holder AJ (1990) *Organometallics* 9:508–511
91. Dewar MJS, Jie C (1987) *Organometallics* 6:1486–1490
92. Dewar MJS, Jie C (1989) *THEOCHEM* 187:1–13
93. Dewar MJS, Yuan YC (1990) *Inorg Chem* 29:3881–3890
94. Dewar MJS, Merz KM Jr (1988) *Organometallics* 7:522–524
95. Dewar MJS, Jie C (1989) *Organometallics* 8:1544–1547
96. Dewar MJS, Healy EF, Kuhn DR, Holder AJ (1991) *Organometallics* 10:431–435
97. Dewar MJS, Jie C (1989) *Organometallics* 8:1547–1549
98. Anders E, Koch R, Freunsh P (1993) *J Comput Chem* 14:1301–1312
99. Stewart JJP (1991) *J Comput Chem* 12:320–341
100. Yu J, Hehre WJ (1995) *Wavefunction*, Irvine, CA
101. Thiel W, Voityuk AA (1994) *J Mol Struct* 313:141–154