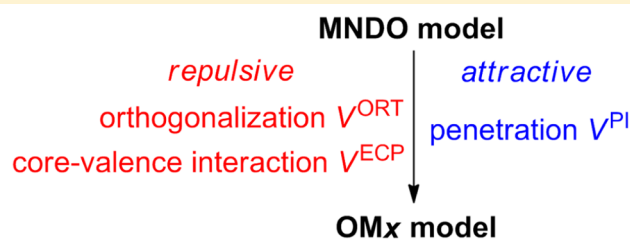Article

# Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Theory, Implementation, and Parameters

Pavlo O. Dral, Xin Wu, Lasse Spörkel, Axel Koslowski, Wolfgang Weber,[†] Rainer Steiger,[‡] Mirjam Scholten,[§] and Walter Thiel*

Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany

**S** Supporting Information

**ABSTRACT:** Semiempirical orthogonalization-corrected methods (OM1, OM2, and OM3) go beyond the standard MNDO model by explicitly including additional interactions into the Fock matrix in an approximate manner (Pauli repulsion, penetration effects, and core−valence interactions), which yields systematic improvements both for ground-state and excited-state properties. In this Article, we describe the underlying theoretical formalism of the OM$x$ methods and their implementation in full detail, and we report all relevant OM$x$ parameters for hydrogen, carbon, nitrogen, oxygen, and fluorine. For a standard set of mostly organic molecules commonly used in semiempirical method development, the OM$x$ results are found to be superior to those from standard MNDO-type methods. Parametrized Grimme-type dispersion corrections can be added to OM2 and OM3 energies to provide a realistic treatment of noncovalent interaction energies, as demonstrated for the complexes in the S22 and S66×8 test sets.

## 1. INTRODUCTION

Over the past decades, semiempirical quantum-chemical (SQC) methods have been widely used in computational chemistry to study molecular structure, stability, spectroscopy, and reactivity.[1] They are computationally much less demanding than *ab initio* or density functional theory (DFT) methods and can therefore be applied to investigate huge molecules or large sets of molecules and to perform large-scale molecular dynamics simulations, also in the context of quantum mechanics/molecular mechanics (QM/MM) approaches.[1,2]

Formally, SQC methods are simplified molecular orbital (MO) treatments. Compared with *ab initio* MO methods, their computational speed comes from the neglect of many small (mostly two-electron) integrals. To compensate for the associated errors, the remaining integrals are usually represented by functions containing parameters, which are then fitted against experimental or high-level theoretical data to achieve a reasonable compromise between computational speed and accuracy.[1,3,4] Many recent benchmark studies actually show that the accuracy of modern SQC methods may approach or sometimes even exceed that of standard DFT methods, especially for ground-state properties of organic molecules,[3−10] despite the fact that for many years method development has been much less intense in the SQC field than in the *ab initio* and DFT realm.[11]

Most established modern SQC methods make use of the NDDO (neglect of diatomic differential overlap) integral approximation and are based on the general-purpose MNDO (modified neglect of diatomic overlap) model introduced in 1977.[12,13] Since then, there have been many refinements aimed at enhancing the accuracy and the application range of the original

MNDO method,[1,3] for example by increasing the size and quality of the training data during parametrization, by modifying the empirical core repulsion function, and by extending the basis set from sp to spd. The resulting MNDO-type methods include the following SQC treatments (standard acronyms, in parentheses year of publication): AM1[14] (1985), PM3[15,16] (1989), MNDO/d[17−19] (1992), PDDG/MNDO and PDDG/PM3[20,21] (2002), AM1*[22] (2003), RM1[23] (2006), PM6[5] (2007), and PM7[4] (2013). To enable the treatment of solid-state systems, PM7 uses modified parametric expressions for the electrostatic interactions.[4] For the latest PM$x$ methods, parameters are available for most elements in the Periodic Table.[4,5]

Another line of SQC development at the NDDO level goes beyond the MNDO model by including explicit orthogonalization corrections into the one-electron terms of the Fock matrix.[1] This gives rise to a series of orthogonalization-corrected methods (OM$x$): OM1[24,25] (1991−1993), OM2[26,27] (1996−2000), and OM3[28] (2003). They differ in the extent to which these corrections are applied to the core Hamiltonian part of the Fock matrix (see section 2.2). The orthogonalization terms represent Pauli repulsions, which are formally missing in the Fock matrix of MNDO-type methods where they are mimicked by the empirical repulsive terms in the core repulsion function.[1,29] Their explicit inclusion in the OM$x$ methods overcomes some characteristic shortcomings of MNDO-type methods. For example, it offers significant improvements in the description of conformational properties, hydrogen bonds, and transition states, and it leads to a

qualitatively correct asymmetric splitting of bonding and antibonding orbitals (as opposed to the symmetric splitting in MNDO-type methods).[1] Statistical evaluations in benchmark studies show that OM$x$ methods are overall more accurate than MNDO-type methods both for ground-state and excited-state properties.[6,30] However, OM$x$ parameters are currently only available for the elements H, C, N, O, and F, which limits their applicability. The benefits of an explicit SQC treatment of orthogonalization have also been demonstrated in an independent proof-of-principle study at the MNDO level (NO-MNDO method, 2006).[31]

Many recent applications of OM$x$ methods focus on electronically excited states of organic molecules and their dynamics. For the exploration of excited-state potential energy surfaces the OM$x$ methods are clearly superior to standard MNDO-type approaches;[30] the semiempirical INDO/S method is adequate only for vertical transitions;[32] time-dependent density functional theory (TD-DFT) has well-known limitations for charge-transfer states and for states of multireference character;[33,34] and accurate correlated *ab initio* approaches are often too costly for larger systems. To account for static correlation effects, OM$x$ studies of excited states employ a multireference configuration interaction (MRCI) treatment,[35] with a properly chosen, relatively small active space and with an efficient implementation of (semi)-analytic derivatives.[36,37] To quote a few examples, OM$x$/MRCI excited-state dynamics simulations using surface hopping[38] have been reported for azobenzenes[39,40] and other photoswitches,[41−43] for a molecular rotary motor,[44] for a variant of the GFP chromophore,[45] for DNA bases in different environments,[46−52] and for the mechanism of the photoinduced ultrafast Wolff rearrangement.[53] Regardless of these successful excited-state studies, we will focus in the remainder of this Article on the OM$x$ performance for ground-state properties, since comprehensive excited-state validations have been presented elsewhere.[30]

Dispersion interactions play an important role in many parts of chemistry and biochemistry, for example in molecular recognition and noncovalent complex formation.[54] Dispersion can be treated adequately by correlated *ab initio* methods but is formally not included in single-determinant MO approaches (at the *ab initio*, standard DFT, and SQC levels). In recent years, there have been many efforts to remedy this deficiency, especially by the development of simple empirical dispersion corrections, as exemplified by the work of Grimme on DFT methods.[55−58]

Such corrections have also been applied to various SQC approaches, sometimes combined with further empirical corrections for other types of noncovalent interactions such as hydrogen bonding, which are often described not too well by standard SQC methods.[3] General-purpose dispersion-corrected SQC methods include treatments labeled AM1-D and PM3-D;[59] PM6-DH;[60] AM1, PM6, and OM3 augmented with DH2 corrections;[61] AM1, PM6, and OM3 augmented with DH+ corrections;[62] AM1, RM1, PM6, and OM3 with D3H4 corrections;[63] PM6-D3H;[64] PM7;[4] and PM6-D3H+;[65] as well as OM2-D3 and OM3-D3[10] (for a recent review see ref 66). In most of these approaches,[66] the parameters of the core SQC methods were kept fixed, and only the correction terms were parametrized. A notable exception is the recently proposed PM7 method, for which a full parametrization was performed (with dispersion and all other corrections included simultaneously).[4]

In our own initial work in this area, we augmented the OM$x$ methods with partially reoptimized Grimme dispersion corrections (without changing any of the basic OM$x$ parameters)

and showed that this gave much improved results for noncovalent complexes and allowed for realistic QM and QM/MM calculations on the binding between antibody 34E4 and a hapten.[67] At a later stage, we reoptimized the Grimme DFT-D2 corrections more carefully for OM2 and OM3 using larger training sets;[68] in the following, the resulting treatments are denoted as OM2-D2 and OM3-D2, respectively. The D3(BJ) dispersion corrections were recently parametrized by the Grimme group for OM2 and OM3.[10] These OM2-D3 and OM3-D3 treatments appear to perform similarly well as many of the computationally more expensive dispersion-corrected DFT approaches.[10]

It is obvious from the preceding literature survey that the OM$x$ methods have been available and in use for some time. However, in the case of the more advanced OM2 and OM3 methods, the formalism and the parameters have been fully described only in two Ph.D. theses.[26,28] The main purpose of this Article is to document the underlying methodology, the implementation, and the optimized parameters of the OM$x$ methods in full detail to make them more easily accessible. In addition, we present and discuss the dispersion corrections that can be used in combination with the OM$x$ methods. In terms of validation, we provide numerical results and statistical evaluations for the test sets that have been used during the development of the OM$x$ methods and the associated dispersion corrections. We have performed further extensive benchmarks of the OM$x$ and OM$x$-D$n$ methods using many of the recently published validation and test sets. The corresponding results are reported in the following Article on ground-state properties including noncovalent interactions.[69]

## 2. METHODOLOGY

**2.1. Overview.** The most popular semiempirical methods for studying ground-state molecular properties, e.g., AM1,[14] PM3,[15] PM6,[5] PM7,[4] and the PDDG variants,[20,21] are based on the MNDO model.[12] In matrix notation, the basic equations of this model are as follows:

$$\psi = \phi \mathbf{C} \qquad (1)$$

$$\mathbf{FC} = \mathbf{CE} \qquad (2)$$

$$\mathbf{F} = \mathbf{H} + \mathbf{G}(\mathbf{P}) \qquad (3)$$

$$\mathbf{P} = 2\mathbf{C}_o \mathbf{C}_o^T \qquad (4)$$

$$E_{tot} = E_{el} + \sum_{A<B} E_{AB}^{core} \qquad (5)$$

The MOs $\psi$ are expressed in a minimal valence basis as linear combinations of atomic orbitals $\phi$ (eq 1). The MO coefficients $\mathbf{C}$ are obtained by solving the secular equation in zero-differential-overlap (ZDO) approximation;[70] $\mathbf{F}$ is the Fock matrix and $\mathbf{E}$ denotes the diagonal matrix of MO energies (eq 2). The Fock matrix is the sum of the one-electron core Hamiltonian matrix $\mathbf{H}$ and a two-electron part $\mathbf{G}$ that depends on the density matrix $\mathbf{P}$ (eq 3). The NDDO integral approximation is invoked in the computation of the Fock matrix, which leads to the neglect of many small integrals (e.g., all three-center and four-center two-electron integrals). The dependence of $\mathbf{G}$ on $\mathbf{P}$ necessitates an iterative self-consistent-field (SCF) solution of the secular equation, since $\mathbf{P}$ is determined by the coefficients $\mathbf{C}_o$ of the occupied MOs (eq 4, closed-shell case; superscript "T" denotes the transpose). The total energy $E_{tot}$ is obtained as the sum of the converged electronic energy $E_{el}$ and the core−core repulsions

$E_{AB}^{core}$ between atoms $A$ and $B$ (eq 5). Following semiempirical convention, the total energy of a molecule can be converted into its heat of formation by subtracting the total energies of the constituent atoms (computed at the same level) and adding the experimental heats of formation of the atoms.[12]

All of these basic features of the MNDO model are retained in the OM$x$ methods. In addition, all one-center integrals are treated in exactly the same manner. The one-center one-electron energies are considered to be adjustable parameters that are optimized in the parametrization procedure, while the one-center two-electron integrals are derived from experimental data (atomic spectra) and kept fixed at the values used in MNDO.[12] The OM$x$ methods differ from the MNDO model in the treatment of the two-center terms entering the core Hamiltonian and the Fock matrix.

Conceptually, the MNDO electronic structure model formally neglects the Pauli exchange repulsions, although their effects on energies may be partly retrieved in MNDO-type methods through additional core repulsion potentials and versatile parametrizations.[4,5,14,15,20,21,71] However, intrinsic deficiencies of the MNDO-type methods are evident in many studies;[1,24−29,72−75] e.g., the gaps between bonding and antibonding MOs and the corresponding excitation energies are significantly underestimated. The OM$x$ methods include explicit orthogonalization corrections in the core Hamiltonian to model Pauli exchange repulsions in the electronic structure calculations.[24−28] For the sake of consistency, penetration effects and core−valence interactions are also incorporated into the core Hamiltonian, and the resonance integrals are represented by a nontraditional empirical expression.[25]

At the implementation stage, there is another major distinction. The two-center Coulomb interactions are represented as parametric functions in MNDO-type methods, while they are computed analytically at the OM$x$ level followed by a uniform Klopman−Ohno scaling.[76,77] The Klopman−Ohno scale factor, $f_{KO}$, is chosen such that the resulting two-center two-electron integrals $(s^A s^A, s^B s^B)$ are the same as in the MNDO model[12,78] at all distances $R_{AB}$. This ensures that the two-center two-electron integrals converge to the correct one-center limit (i.e., to the MNDO values of the one-center two-electron integrals; see earlier discussion) and that they show the correct asymptotic behavior at large distances; at intermediate distances, the Klopman−Ohno scaling will account for dynamic correlation effects in an average manner. To achieve a balanced treatment of all two-center Coulomb interactions, the Klopman−Ohno scaling is applied not only to the electron−electron repulsions but also to the core−electron attractions and core−core repulsions.

In the next two sections, we present the distinguishing features of the OM$x$ methods and their implementation in more detail.

**2.2. OM$x$ Formalism: Core Hamiltonian.** The secular equation without overlap (eq 2) is valid in the symmetrically orthogonalized Löwdin basis,[79,80] and hence the Fock matrix should correspond to such an orthogonalized basis in SQC methods. The Löwdin orthogonalization is known to affect the one-electron integrals in the Fock matrix more strongly than the two-electron integrals, and there is numerical evidence that the two-electron integrals surviving under the NDDO approximation are not too different when evaluated in the original nonorthogonal and the orthogonalized basis. In the OM$x$ methods, we therefore decided not to modify the two-electron integrals by orthogonalization corrections.

The one-electron core Hamiltonian of the OM$x$ methods is given by

$$H_{\mu\nu}^{core} = U_{\mu\mu}\delta_{\mu\nu} + \sum_B [V_{\mu\nu,B}^{s} + V_{\mu\nu,B}^{ORT} + V_{\mu\nu,B}^{PI} + V_{\mu\nu,B}^{ECP}] \quad (6)$$

$$H_{\mu\lambda}^{core} = \beta_{\mu\lambda} + \sum_C V_{\mu\lambda,C}^{ORT} \quad (7)$$

where the subscripts $\mu$ and $\nu$ refer to orbitals centered at the same atom $A$, $\lambda$ designates an orbital at a different atom $B$ $(B \neq A)$, and $C$ is an atom that is neither $A$ nor $B$; $\delta_{\mu\nu}$ denotes the Kronecker delta. The one-electron energies $U_{\mu\mu}$ are adjustable parameters. The integrals $V_{\mu\nu,B}^{s}$ represent semiempirical core−electron attractions (superscript "s") expressed in terms of related two-center two-electron integrals (see later text).[12]

The most important enhancement of the OM$x$ methods over the MNDO model is the inclusion of the valence orthogonalization corrections $(V^{ORT})$. Their introduction necessitates the incorporation of two other terms that can be of similar magnitude and exhibit a similar distance dependence: the penetration integrals $(V^{PI})$ account for differences between corresponding core−electron attractions and electron−electron repulsions, while the effective core potentials (ECPs) capture the influence of the inner core electrons $(V^{ECP})$. The sum of the repulsive orthogonalization corrections $(V^{ORT})$, the attractive penetration integrals $(V^{PI})$, and the repulsive effective core potentials $(V^{ECP})$ normally gives rise to an overall repulsive contribution to the corresponding element of the core Hamiltonian. All these terms are formally neglected in MNDO-type methods, where they are modeled by an empirical effective pair potential in the core repulsion function.

We now present explicit expressions for these terms in the OM$x$ formalism which have been derived previously.[24−27] $V_{\mu\nu,B}^{ORT}$ and $V_{\mu\lambda,C}^{ORT}$ denote the two-center (atoms $A$ and $B$) and three-center (atoms $A$, $B$, and $C$) orthogonalization corrections to the core Hamiltonian, respectively. They are given by

$$V_{\mu\nu,B}^{ORT} = -\frac{1}{2}F_1^A \sum_{\rho \in B}(S_{\mu\rho}\beta_{\rho\nu} + \beta_{\mu\rho}S_{\rho\nu})$$
$$+ \frac{1}{8}F_2^A \sum_{\rho \in B}S_{\mu\rho}S_{\rho\nu}(H_{\mu\mu,B}^{loc} + H_{\nu\nu,B}^{loc} - 2H_{\rho\rho,A}^{loc}) \quad (8)$$

$$V_{\mu\lambda,C}^{ORT} = -\frac{1}{2}G_1^{AB} \sum_{\rho \in C}(S_{\mu\rho}\beta_{\rho\lambda} + \beta_{\mu\rho}S_{\rho\lambda})$$
$$+ \frac{1}{8}G_2^{AB} \sum_{\rho \in C}S_{\mu\rho}S_{\rho\lambda}(H_{\mu\mu,C}^{loc} + H_{\lambda\lambda,C}^{loc} - H_{\rho\rho,A}^{loc} - H_{\rho\rho,B}^{loc}) \quad (9)$$

$$H_{\mu\mu,X}^{loc} = U_{\mu\mu} + V_{\mu\mu,X}^{s} \quad (10)$$

$$G_1^{AB} = \frac{1}{2}(G_1^A + G_1^B) \quad (11)$$

$$G_2^{AB} = \frac{1}{2}(G_2^A + G_2^B) \quad (12)$$

where $S_{\mu\rho}$ and $\beta_{\rho\nu}$ are elements of the overlap matrix and of the matrix of resonance integrals (defined later), respectively. $H_{\mu\mu,X}^{loc}$ is an element of the local core Hamiltonian restricted to atom pair $A−X$;[24−26] it is defined as the sum of the one-electron energy $U_{\mu\mu}$ (atom $A$) and the semiempirical core−electron attraction $V_{\mu\mu,X}^{s}$ to the second atom $X$ in the pair considered. To avoid rotational invariance issues, average local matrix elements are employed for

the three $p$ functions at any given atom.[26] $F$ and $G$ are parameters that are used to adjust the magnitudes of the orthogonalization corrections to the one- and two-center core Hamiltonian, respectively; the subscripts 1 and 2 specify the type of these corrections, while the superscripts A and B denote the associated atoms.

The orthogonalization corrections are incorporated into the different OM$x$ methods to a different extent. The earliest variant (OM1) only contains the corrections to the one-center terms, *i.e.*, $V_{\mu\nu,B}^{ORT}$ (eq 8). The OM2 method fully includes all correction terms listed previously. However, further analysis showed that the contributions from the second sum in eqs 8 and 9 are often small and of varying sign, and they normally have only a minor influence on the results, while notably increasing the computational effort for large molecules. Therefore, they are neglected in the OM3 method, *i.e.*, $F_2 \equiv 0$ and $G_2 \equiv 0$; this is the only formal distinction between OM2 and OM3.

The penetration integrals[81] are treated in the same manner in all OM$x$ methods, as specified in the original OM1 work.[25] Briefly, following standard conventions, the semiempirical core−electron attractions $V_{\mu\nu,B}^{s}$ (superscript "s") are expressed in terms of related two-center two-electron integrals $(\mu^A\nu^A,s^B s^B)^s$ involving the ss charge distribution at atom B. The penetration integrals $V_{\mu\nu,B}^{PI}$ are the difference between the actual core−electron attractions $(V_{\mu\nu,B}^{s} + V_{\mu\nu,B}^{PI})$ and their semiempirical values $V_{\mu\nu,B}^{s}$. In the OM$x$ methods the former are obtained from the analytic core−electron attraction integrals $V_{\mu\nu,B}^{a}$ (superscript "a") by uniform scaling[25] with $f_{KO}$ (see as follows).

$$V_{\mu\nu,B}^{s} = -Z_B(\mu^A\nu^A, s^B s^B)^s \tag{13}$$

$$V_{\mu\nu,B}^{s} + V_{\mu\nu,B}^{PI} = f_{KO} V_{\mu\nu,B}^{a} \tag{14}$$

where $Z_B$ is the charge of the core of atom B.

In early semiempirical work, the attractive penetration integrals had been included while neglecting the repulsive orthogonalization corrections, which gave much too short bond lengths and much too large binding energies.[70,82] In most of the later semiempirical methods, both types of interactions were neglected on heuristic grounds,[12,82] assuming that these two terms will cancel and that any remaining effects can be taken into account through small empirical corrections and careful parametrization. The OM$x$ methods treat these terms explicitly, with the motivation to recover electronic effects resulting from an incomplete cancelation of these terms.

In the MNDO model, the core electrons are treated as if they were incorporated into the corresponding nucleus, with a corresponding reduction of the effective nuclear charge experienced by the valence electrons (assuming complete screening); the repulsions between the inner and valence shells arising from orthogonality and exchange effects are also disregarded. By contrast, the OM$x$ methods attempt to incorporate core−valence interactions by means of ECPs for the core electrons, in analogy to *ab initio* and DFT approaches.[83−86] The ECP contributions are treated differently in the OM$x$ variants. In OM1 the integrals $V_{\mu\nu,B}^{ECP}$ are evaluated analytically (*ab initio*) and then subjected to Klopman−Ohno scaling (like all Coulomb interactions).[25] The OM2 and OM3 methods employ the following semiempirical ECP formulas derived previously:[26,27]

$$V_{\mu\nu,B}^{ECP} = -\sum_{\alpha\in B} (S_{\mu\alpha}G_{\alpha\nu} + G_{\mu\alpha}S_{\alpha\nu} + S_{\mu\alpha}S_{\alpha\nu}F_{\alpha\alpha}) \tag{15}$$

where $G_{\mu\alpha}$ is an empirical function that is chosen to be of the same form[26,27] as the empirical resonance integral (see eq 17):

$$G_{\mu\alpha} = \frac{1}{2}(\beta_\mu^A + \beta_\alpha^B)\sqrt{R_{AB}} \exp[-(\alpha_\mu^A + \alpha_\alpha^B)R_{AB}^2] \tag{16}$$

These expressions contain several adjustable parameters related to the core electrons (index $\alpha$), namely, $\alpha_\alpha$, $\beta_\alpha$, $F_{\alpha\alpha}$, and the core orbital exponent $\zeta_\alpha$ needed to evaluate the core−valence overlap integrals $S_{\mu\alpha}$. Extensive numerical tests indicate reasonable qualitative agreement between the *ab initio* and semiempirical ECPs when using properly adjusted parameters.[26,27]

The explicit inclusion of orthogonalization corrections in the OM$x$ methods implies the need for a different treatment of the resonance integrals, which has been introduced and discussed in the original OM1 work.[25] In MNDO-type methods, the two-center resonance integrals are taken to be proportional to the overlap integrals.[4,5,12,14,15,71] This assumption cannot be well justified for an orthogonalized atomic basis. Therefore, on the basis of comparisons with corresponding *ab initio* matrix elements,[24,25] the following empirical formula is adopted for the resonance integrals in the OM$x$ methods:[24−28]

$$\beta_{\mu\lambda}^{loc} = \frac{1}{2}(\beta_\mu^A + \beta_\lambda^B)\sqrt{R_{AB}} \exp[-(\alpha_\mu^A + \alpha_\lambda^B)R_{AB}^2] \tag{17}$$

where $\alpha^A$ and $\beta^A$ are adjustable parameters for the various orbital types of atoms $A$ and $B$. The preceding expression refers to a local diatomic coordinate system and contains an implicit phase factor.[25] The resonance integrals in the molecular coordinate system are obtained by standard rotational transformations.

To summarize, the most important improvement in the OM$x$ methods is the inclusion of valence orthogonalization corrections so that the repulsive Pauli exchange effects can be taken into account in an explicit manner. Penetration integrals for the valence shell and ECPs for the core−valence interactions are also included in the OM$x$ methods for the sake of consistency, in order to ensure a balanced overall treatment. As a further consequence, the two-center resonance integrals are represented by a different kind of empirical formula to better mimic the behavior of the corresponding *ab initio* matrix elements. Accordingly, the OM$x$ methods go significantly beyond the established MNDO model by incorporating additional interactions in the electronic structure calculation.

**2.3. Implementation.** The OM$x$ methods employ a minimal valence basis of contracted Gaussian functions (STO-3G for hydrogen and ECP-3G for carbon, nitrogen, oxygen, and fluorine), which has been fully documented previously.[24,25] To allow for fine-tuning, there is one adjustable parameter $\zeta$ per element that scales the exponents of all Gaussian primitives in the usual manner (multiplication by $\zeta^2$). These scale factors $\zeta$ are optimized in the parametrization of the OM$x$ methods but are expected to remain close to 1. All analytic integral evaluations are performed with these scaled basis functions.

The elements of the one-electron core Hamiltonian matrix **H** are fully specified in the preceding section (see eqs 6−17). Analytic integral evaluations are performed for the valence-shell overlap integrals $S_{\mu\nu}$ (see eqs 8 and 9) and the core−electron attractions $V_{\mu\nu,B}^{a}$ (see eq 14). In OM1 the core−valence interactions $V_{\mu\nu,B}^{ECP}$ are also evaluated analytically,[25] while they are represented in OM2 and OM3 by an empirical expression containing core−valence overlap integrals $S_{\mu\alpha}$ (see eq 15). The latter are computed analytically using standard STO-3G basis functions for the core orbitals, with adjustable scale factors $\zeta_\alpha$ that

are optimized in the parametrization of OM2 and OM3 for each element; their values are expected to be close to the optimum core orbital exponents for Slater-type orbitals.

The two-electron integrals (entering the Fock matrix via **G**) are treated in the same manner in all OM$x$ methods.[25] The semiempirical scheme used in MNDO-type methods for evaluating the two-center two-electron integrals, in terms of a point-charge representation of multipole—multipole interactions,[78] is replaced by the analytic computation of these integrals over the contracted Gaussian valence orbitals,[24,25] followed by a uniform Klopman—Ohno scaling.[76,77]

$$(\mu^A \nu^A, \lambda^B \sigma^B)^s = f_{KO} (\mu^A \nu^A, \lambda^B \sigma^B)^a \tag{18}$$

$$f_{KO} = (s^A s^A, s^B s^B)^{MNDO} / (s^A s^A, s^B s^B)^a \tag{19}$$

$f_{KO}$ is defined such that the semiempirical two-center two-electron integrals $(s^A s^A, s^B s^B)^s$ are the same as in the MNDO model[12,78] (as required in the one-center limit). The other two-center two-electron integrals $(\mu^A \nu^A, \lambda^B \sigma^B)^s$ will be of similar magnitude as in MNDO but will generally differ slightly. To achieve a balanced treatment of all two-center Coulomb interactions, the Klopman—Ohno scaling is applied not only to the core—electron attractions (eq 14) and electron—electron repulsions (eq 18) but also to the core—core repulsions $E_{AB}^{core}$ between atoms $A$ and $B$:

$$E_{AB}^{core} = f_{KO} Z_A Z_B / R_{AB} \tag{20}$$

This completes the description of integral evaluation in the OM$x$ methods. Some further comments seem appropriate concerning the practical implementation of the three-center orthogonalization corrections in OM2 and OM3 (eq 9). Generally speaking, all two-center integrals in the OM$x$ methods are evaluated in a local coordinate system and are then converted to the molecular coordinate system using standard rotational transformations. This is no longer feasible for the three-center orthogonalization corrections which need to be assembled directly in the molecular coordinate system. For this purpose, the overlap matrix **S** and the matrix of resonance integrals $\boldsymbol{\beta}$ are precomputed in the molecular coordinate system, as well as the required local core Hamiltonian matrix elements of the type $H_{\mu\mu,X}^{loc}$ (eqs 9 and 10); the one-center matrix elements of $\boldsymbol{\beta}$ are defined to be zero.

For $N$ basis functions, the computation of the three-center orthogonalization corrections (eq 9) formally scales as $O(N^3)$, which may raise some concern since all other integral evaluations in SQC methods generally scale as $O(N^2)$. However, a careful implementation that exclusively targets the nonzero contributions already allows an efficient evaluation of the three-center orthogonalization corrections (consuming only a fraction of the overall computation time).

Moreover, the contributions for a given combination of atoms ($A$, $B$, $C$) fall off quickly with the interatomic distances $R_{AC}$ and $R_{BC}$ because of the steep decline of the corresponding elements of **S** and $\boldsymbol{\beta}$ that appear in eq 9. It is thus possible to use prescreening techniques to minimize the computational effort, by utilizing the precalculated matrices **S** and $\boldsymbol{\beta}$. Since the resonance integrals (eq 17) normally fall off faster than the corresponding overlap integrals, a conservative prescreening criterion may be based on the product of the ss overlap integrals for atom pairs $A-C$ and $B-C$: for example, if this product is smaller than a user-defined cutoff $\tau_1$ the three-center contributions for the combination ($A$, $B$, $C$) can be neglected entirely. This cutoff is available in our code (others can easily be

envisioned) but is used only rarely for energy evaluation, in order to avoid sacrificing numerical precision for a rather small computational gain overall.

The three-center orthogonalization corrections for OM3 are formally of first order in overlap. They can be written in matrix form by expressing the first sum in eq 9 as the sum of the matrix product $\tilde{\boldsymbol{S}}\boldsymbol{\beta}$ and its transpose;[68] the matrix $\tilde{\boldsymbol{S}}$ is the overlap matrix with the diagonal elements set to zero.[25]

$$X_{\mu\lambda} = \sum_{\rho \in C} (S_{\mu\rho} \beta_{\rho\lambda} + \beta_{\mu\rho} S_{\rho\lambda}) \tag{21}$$

$$\mathbf{X} = (\tilde{\boldsymbol{S}}\boldsymbol{\beta}) + (\tilde{\boldsymbol{S}}\boldsymbol{\beta})^T \tag{22}$$

Multiplication of $X_{\mu\lambda}$ with the appropriate prefactor yields the three-center correction $V_{\mu\lambda,C}^{ORT}$ for OM3 (eq 9). The use of optimized library routines for matrix multiplication makes this approach an extremely efficient alternative to applying cutoffs in OM3, especially when using graphics processing units (GPUs).[68]

For a direct comparison of the different implementations, wall clock computing times were measured for the evaluation of the OM3 three-center corrections for seven proteins ranging in size from 1097 to 3558 atoms, using a platform with two Intel Xeon X5690 CPUs and two NVIDIA Tesla M2090 GPUs.[68] The time for the full explicit calculation on a single CPU core (no GPU) served as reference. The measured speedups ranged from 14.0 to 18.0 for the explicit calculation with the cutoff $\tau_1 = 10^{-6}$, from 8.9 to 13.1 for the evaluation via matrix multiplication (eq 22) on a single CPU core (no GPU), and from 241 to 351 for the same evaluation using one CPU core and one GPU.[68]

In standard MNDO-type methods, the computational effort for a full SCF energy evaluation is dominated by matrix operations that scale as $O(N^3)$ and need to be performed in each SCF iteration (diagonalization or pseudodiagonalization of the Fock matrix to solve eq 2, density matrix computation via eq 4, and possibly also DIIS extrapolation). In SCF calculations at the OM$x$ level, the same matrix operations dominate the computational effort, even though the evaluation of the three-center orthogonalization corrections in OM2 and OM3 also scales formally as $O(N^3)$; however, this evaluation is required only once at the beginning to construct the core Hamiltonian (and not again in each SCF iteration), and it can be made very efficient as outlined previously. Actual benchmarks on fullerenes up to $C_{960}$ and on clusters up to 1200 water molecules confirm that the evaluation of the core Hamiltonian matrix takes only a minor fraction of the time for a full SCF calculation at the OM$x$ level— less than 10% even in the worst case (largest water cluster, OM2 with full explicit calculation of the three-center orthogonalization corrections).[68,87] In these and other benchmarks, standard MNDO-type and OM$x$ SCF calculations generally take about the same time for all systems considered,[68] simply because the dominant matrix operations are the same in both cases. Previous work achieved substantial speedups of MNDO-type calculations through shared-memory and distributed-memory message-passing parallelization of these matrix operations (and of integral evaluation and Fock matrix construction);[73−75,88] these implementations can be applied likewise at the OM$x$ level. Finally, MNDO-type and OM$x$ computations benefit from hybrid CPU-GPU computing in the same manner, with large speedups of comparable size.[68,87] In conclusion, MNDO-type and OM$x$ SCF calculations on large molecules are essentially equally fast and equally facile with our current code (for further detailed timing data on large molecules, see ref 87).

In semiempirical SCF methods, the energy expression can be differentiated analytically with respect to the nuclear coordinates.[89,90] The resulting analytic gradient expressions contain integral derivatives that can be evaluated analytically or by finite difference. In our implementation, the fully analytic approach is available for MNDO-type methods,[89,90] while the integral derivatives are computed by finite difference for the OM$x$ methods.[37] In closed-shell SCF calculations with MNDO-type and OM$x$ methods, one may also use a simple finite-difference procedure that involves computing the change in the energy at displaced geometries (with constant density matrix and recalculated values of the integrals that change upon displacement). All these finite-difference techniques are efficient at the SQC level as long as only two-center integrals need to be treated (MNDO-type methods and OM1): the computational effort then scales as $O(N^2)$, and gradient evaluation takes only a minor fraction of the time needed for calculating the energy (typically much less than 10%).

The situation becomes less favorable in the case of OM2 and OM3 because of the presence of the three-center orthogonalization corrections, which cause an $O(N^3)$ scaling of the finite-difference computation of the corresponding gradient contributions. Even when care is taken to target only the nonvanishing contributions, the computational effort increases rapidly with system size so that it becomes more costly for large systems to evaluate the gradient than the energy. In this situation, prescreening becomes essential for efficiency. In addition to the test on the ss overlap integrals described previously (cutoff $\tau_1$), our current implementation offers a second test: the contribution of the atom triple ($A$, $B$, $C$) to the overall three-center orthogonalization correction is known at the reference geometry (in terms of energy), and if this is only a small fraction (less than a cutoff $\tau_2$) the corresponding contributions to the gradient can be neglected safely.

To illustrate the effects of prescreening, we measured wall clock computing times for OM2 and OM3 calculations on our seven test proteins (1097−3558 atoms, 2699−8727 basis orbitals) on the same platform as before. Without any cutoffs, gradient evaluation was slower than energy evaluation on a single CPU typically by factors of 3−4 (2−3) for OM2 (OM3). Applying the cutoffs $\tau_1 = \tau_2 = 10^{-8}$ led to negligible changes in the Cartesian gradient norm (typically less than 0.02 kcal/(mol·Å)) while accelerating gradient evaluation on a single CPU by factors of 17−24 (11−16) for OM2 (OM3) so that it always took only a fraction of the time needed for energy evaluation with the same cutoffs (15−27%). On the hybrid platform with six CPU threads and two GPUs, combined energy and gradient calculations on the test proteins with the same cutoffs were faster by factors of 7−10 compared with runs on a single CPU, both for OM2 and OM3; wall clock times for the largest protein (3558 atoms) were 1530 (1280) s for OM2 (OM3).

To summarize, in our current implementation of the three-center orthogonalization corrections in OM2 and OM3, prescreening is not required for computing the energy (particularly in OM3) but strongly recommended for gradient evaluation in large systems (default cutoffs, $10^{-8}$).

**2.4. Parameters.** The parameters of the OM$x$ methods were optimized by an iterative nonlinear least-squares fitting procedure, which minimizes the sum-of-squares (SSQ) of the weighted errors for the chosen properties of reference molecules. For a given set of trial parameters, the Levenberg−Marquardt algorithm[91,92] was used to define a search vector, and a line search along this direction was performed to minimize the SSQ value and thus to determine an improved set of parameters; these steps were iterated until the SSQ value could not be lowered further. The Levenberg−Marquardt algorithm was chosen because it allows for a gradual transition from a steepest descent procedure at the beginning to a Gauss−Newton method when approaching the minimum. It requires knowledge of the Jacobi matrix containing the partial derivatives of the individual errors of the reference properties with respect to the parameters being optimized. The Jacobi matrix was computed by finite difference in the first few iterations and thereafter in regular intervals to limit the computational effort, with approximate updates in between. Further technical details on the parametrization procedure are available in the original reports.[24,26,28]

The chosen parametrization algorithm provides a derivative-based method for local optimization of the parameters with respect to the SSQ value. The parameter surface of SQC methods is known to be quite rugged, and the optimization can thus converge to different local minima when starting from different initial sets of parameters. Therefore, many separate parametrization runs with different initial conditions were carried out for each of the OM$x$ methods.[24,26,28]

Following semiempirical tradition,[12] the reference properties mainly included heats of formation, geometries, ionization potentials (computed via Koopmans' theorem), and dipole moments (computed in ZDO approximation). The weights were chosen such that the heats of formation contributed most to the overall SSQ value (followed by the geometries). For each of the OM$x$ methods, parametrizations were first run for hydrocarbons to obtain reasonable CH parameters, which were then kept fixed in subsequent optimizations of the N and O parameters. This provided good starting sets for final simultaneous refinements of all CHNO parameters. The parameters for F were determined separately (with fixed CHNO parameters).

The scope of the CHNO parametrization of OM1 (early 1990s), OM2 (mid 1990s), and OM3 (early 2000s) can be judged from the number of reference molecules employed: OM1, 10−30 initially, 102 for the last refinement;[24] OM2, 35−40 typically;[26] and OM3, 227 at the final stage.[28] Most of the reference molecules were taken from the CHNO and FLUOR training sets of neutral molecules described in the next section.

For the assessment of the resulting parameter sets, test calculations were done for other types of molecules (e.g., in the case of OM3 for radicals, cations, anions, hydrogen-bonded complexes, and peptides) as well as other properties (e.g., vibrational wavenumbers, higher ionization potentials, vertical excitation energies, transition-state energies, and geometries) to compare with corresponding reference data from experiment or from accurate *ab initio* calculations. In some cases, this motivated the inclusion of other reference molecules into the training set.[24,26,28]

To summarize, the parametrizations of the OM$x$ methods did not follow a strictly defined protocol, but there were some common characteristic features. They employed local derivative-based optimization methods, a rather limited number (by current standards) of reference molecules, and ground-state reference properties with focus on energies (heats of formation). The parametrization runs normally yielded several parameter sets with comparable SSQ values, and hence the selection of the final parameters involved some subjective judgment taking into account also the overall performance for other molecules and

## Table 1. OM1 Parameters

|  | H | C | N | O | F |
|---|---|---|---|---|---|
| | | | Orbital Exponent: Scale Factor | | |
| $\zeta$ (au) | 1.20948930 | 1.13551142 | 1.16081665 | 1.10190209 | 1.16498140 |
| | | | One-Center One-Electron Energies | | |
| $U_{ss}$ (eV) | −12.83851861 | −50.15945000 | −71.33505463 | −93.04158571 | −121.69518463 |
| $U_{pp}$ (eV) | | −38.76257345 | −56.58315267 | −77.59792792 | −106.37310634 |
| | | | Resonance Integrals | | |
| $\beta_s$ (eV bohr$^{-1/2}$) | −4.89312435 | −7.58632270 | −12.00586167 | −6.22223757 | −5.73558167 |
| $\beta_p$ (eV bohr$^{-1/2}$) | | −4.49894163 | −9.64950408 | −9.94028730 | −16.36168108 |
| $\beta_\pi$ (eV bohr$^{-1/2}$) | | −5.91210138 | −10.16405908 | −11.29342651 | −17.22481680 |
| $\alpha_s$ (au) | 0.09653898 | 0.09325105 | 0.10185884 | 0.10891616 | 0.19374572 |
| $\alpha_p$ (au) | | 0.05398748 | 0.08540515 | 0.09666556 | 0.13034754 |
| $\alpha_\pi$ (au) | | 0.10477244 | 0.14350678 | 0.15255321 | 0.22033729 |
| $\beta_s$(X−H) (eV bohr$^{-1/2}$) | | | −8.08332477 | −6.45960578 | |
| $\beta_p$(X−H) (eV bohr$^{-1/2}$) | | | −11.48923075 | −12.47451386 | |
| $\alpha_s$(X−H) (au) | | | 0.07514671 | 0.07953217 | |
| $\alpha_p$(X−H) (au) | | | 0.11556228 | 0.12908958 | |
| | | | Orthogonalization Factors | | |
| $F_1$ | 0.54128873 | 0.50383851 | 0.63476395 | 0.68193417 | 1.19938976 |
| $F_2$ | 0.84668969 | 0.66944409 | 0.31135759 | 0.47652748 | 0.49484369 |

## Table 2. OM2 Parameters

|  | H | C | N | O | F |
|---|---|---|---|---|---|
| | | | Orbital Exponent: Scale Factor | | |
| $\zeta$ (au) | 1.47386481 | 1.42036892 | 1.33175233 | 1.55214516 | 1.45216726 |
| | | | One-Center One-Electron Energies | | |
| $U_{ss}$ (eV) | −12.64890000 | −51.65550844 | −74.37638240 | −101.82723464 | −120.62785370 |
| $U_{pp}$ (eV) | | −39.74369825 | −57.60067613 | −78.92823923 | −107.27105397 |
| | | | Resonance Integrals | | |
| $\beta_s$ (eV bohr$^{-1/2}$) | −3.41998220 | −7.21406021 | −10.84303446 | −10.64436974 | −6.25438426 |
| $\beta_p$ (eV bohr$^{-1/2}$) | | −4.14394503 | −7.62373736 | −8.63610952 | −13.93492471 |
| $\beta_\pi$ (eV bohr$^{-1/2}$) | | −5.97107657 | −9.27936312 | −9.21201190 | −18.73205761 |
| $\alpha_s$ (au) | 0.06607903 | 0.09045297 | 0.08974553 | 0.13062089 | 0.26624434 |
| $\alpha_p$ (au) | | 0.05452192 | 0.08759680 | 0.09626876 | 0.12261412 |
| $\alpha_\pi$ (au) | | 0.10204903 | 0.13172314 | 0.13071747 | 0.21684388 |
| $\beta_s$(X−H) (eV bohr$^{-1/2}$) | | −6.30164062 | −9.49567107 | −6.54238767 | −6.25104378 |
| $\beta_p$(X−H) (eV bohr$^{-1/2}$) | | −4.04444703 | −8.51180846 | −10.11307271 | −13.94492971 |
| $\alpha_s$(X−H) (au) | | 0.09668329 | 0.11429048 | 0.11112738 | 0.44713918 |
| $\alpha_p$(X−H) (au) | | 0.05283694 | 0.10673732 | 0.11891861 | 0.15648906 |
| | | | Orthogonalization Factors | | |
| $F_1$ | 0.29566861 | 0.49949211 | 0.64073384 | 1.26450169 | 2.11499396 |
| $F_2$ | 1.40190659 | 0.72261226 | 0.19580808 | 1.14847352 | 1.09156321 |
| $G_1$ | 0.65271563 | 0.21284361 | 0.13946233 | 0.28309603 | 0.31704089 |
| $G_2$ | 0.90843670 | 0.99250289 | 0.84373060 | 0.78414131 | 0.02140504 |
| | | | Effective Core Potentials | | |
| $\zeta_\alpha$ (au) | | 5.16802668 | 6.93980600 | 7.58579774 | 8.71226515 |
| $F_{\alpha\alpha}$ (eV) | | −305.68646337 | −407.39202305 | −514.45812327 | −685.41988599 |
| $\beta_\alpha$ (eV bohr$^{-1/2}$) | | −9.07185084 | −9.97910210 | −14.16551053 | −9.17960365 |
| $\alpha_\alpha$ (au) | | 0.16985745 | 0.16173024 | 0.34390559 | 0.99971548 |

properties not included in the training set. The final parameters for the OM1, OM2, and OM3 methods are listed in Tables 1−3, respectively.

Generally speaking, the parameters for a given element can be classified into five groups: the scale factor for the Gaussian-type basis functions, the one-center one-electron energies, the parameters for the resonance integrals, the coefficients for the orthogonalization corrections, and the parameters for the semiempirical ECPs in the case of OM2 and OM3.

The exponents of all Gaussian primitives in a given contracted basis were allowed to be scaled by a common factor $\zeta^2$. The optimized scale factors are generally somewhat larger than 1 indicating some contraction (as also encountered in the *ab initio* case for the hydrogen STO-3G basis).

The one-center one-electron energies $U_{ss}$ and $U_{pp}$ are treated as adjustable parameters in the OM*x* and MNDO-type methods. Their optimized values for a given element are generally of similar magnitude in all methods.

**Table 3. OM3 Parameters**

| | H | C | N | O | F |
|---|---|---|---|---|---|
| | | Orbital Exponent: Scale Factor | | | |
| $\zeta$ (au) | 1.25906452 | 1.27811536 | 1.30965521 | 1.20838191 | 1.20564838 |
| | | One-Center One-Electron Energies | | | |
| $U_{ss}$ (eV) | −12.45828647 | −50.55997310 | −75.98413465 | −105.79319826 | −120.65477058 |
| $U_{pp}$ (eV) | | −39.60463506 | −57.38630489 | −78.90502490 | −107.50304443 |
| | | Resonance Integrals | | | |
| $\beta_s$ (eV bohr$^{-1/2}$) | −3.40064659 | −7.15007507 | −13.42485887 | −14.42839639 | −6.19918959 |
| $\beta_p$ (eV bohr$^{-1/2}$) | | −4.00965991 | −5.69143961 | −8.77114206 | −13.82075048 |
| $\beta_\pi$ (eV bohr$^{-1/2}$) | | −5.63958651 | −8.25767437 | −12.94995697 | −18.96710976 |
| $\alpha_s$ (au) | 0.06931667 | 0.09197146 | 0.09461210 | 0.12962541 | 0.31128286 |
| $\alpha_p$ (au) | | 0.05274021 | 0.06941595 | 0.09275135 | 0.12448702 |
| $\alpha_\pi$ (au) | | 0.09864674 | 0.10511596 | 0.16086067 | 0.21596283 |
| $\beta_s$(X−H) (eV bohr$^{-1/2}$) | | −6.19914817 | −11.40440049 | −13.56403003 | −8.06286638 |
| $\beta_p$(X−H) (eV bohr$^{-1/2}$) | | −4.23218526 | −7.87479008 | −9.42200507 | −13.92380910 |
| $\alpha_s$(X−H) (au) | | 0.10023679 | 0.11356707 | 0.14516313 | 0.32623663 |
| $\alpha_p$(X−H) (au) | | 0.05492720 | 0.09244486 | 0.10981869 | 0.15497349 |
| | | Orthogonalization Factors | | | |
| $F_1$ | 0.25393975 | 0.41151269 | 0.58223498 | 0.55266327 | 1.03522357 |
| $G_1$ | 0.35600772 | 0.10398816 | 0.05928617 | 0.06226814 | 0.14035077 |
| | | Effective Core Potentials | | | |
| $\zeta_\alpha$ (au) | | 5.70000000 | 6.73673665 | 7.74239776 | 8.70367110 |
| $F_{\alpha\alpha}$(eV) | | −283.81699000 | −380.94261410 | −512.51900000 | −685.41999336 |
| $\beta_\alpha$ (eV bohr$^{-1/2}$) | | −22.48815939 | −22.78185605 | −26.13251784 | −10.62649349 |
| $\alpha_\alpha$ (au) | | 0.15323932 | 0.15915338 | 0.18281098 | 0.00010713 |

The two-center resonance integrals are highly parametrized in the OM$x$ methods, generally more so than in standard MNDO-type methods. The relevant parameters are distinguished by the subscripts s, p, and $\pi$ assigning them to the s, p$_\sigma$, and p$_\pi$ orbitals, respectively, in the local coordinate system used for evaluating the resonance integrals. For X−H bonds between a heavy atom and a hydrogen atom, the OM$x$ formalism allows the use of separate heavy-atom parameters, i.e., $\beta_s$(X−H), $\beta_p$(X−H), $\alpha_s$(X−H), and $\alpha_p$(X−H), to enable some fine-tuning if this is deemed to be desirable for a given element. The various resonance parameters are interdependent to some extent, and it is therefore difficult to interpret their optimized values. Meaningful comparisons are best done by plotting the relevant resonance integrals in the local diatomic coordinate system as functions of the interatomic distance. Such plots are available and show qualitatively reasonable behavior both in the bonding region and asymptotically.[24,25,28]

The functional form of the orthogonalization corrections in the OM$x$ methods originates from formal expansions of the *ab initio* corrections (arising from symmetric Löwdin orthogonalization) in terms of overlap.[24−28] The OM$x$ corrections are scaled by prefactors (e.g., $F_1$, $F_2$, $G_1$, and $G_2$ in OM2) that are adjusted in the parametrization. Their optimum values are thus expected to be of the order of 1. While this holds roughly, most of these values tend to be somewhat smaller.

The parameters for the semiempirical ECPs used in OM2 and OM3 are listed in Tables 2 and 3, respectively. As expected, the optimized core orbital exponents $\zeta_\alpha$ are in most cases reasonably close to the optimum exponent for the corresponding Slater-type core orbital (i.e., atomic number minus 0.3 to account for partial screening), and the atomic Fock matrix elements $F_{\alpha\alpha}$ for the core orbitals represented by ECPs[26,27] adopt optimized values that are of similar magnitude as the corresponding experimental core ionization potentials. The core orbital parameters $\alpha_\alpha$ and $\beta_\alpha$ entering the empirical ECP expressions are more difficult to

assess; it is again best to inspect plots of all relevant ECP integrals to confirm that semiempirical and corresponding *ab initio* ECP integrals match reasonably well.[26,28]

Further details about the parametrizations for the OM$x$ methods can be found in the original reports.[24,26,28]

**2.5. Validation.** Here we provide a brief assessment of the accuracy of the OM$x$ methods for the standard sets of reference molecules and reference properties used most regularly for parametrization (see preceding discussion). These sets are fully specified in the Supporting Information (Tables S1−S6). The standard CHNO set consists of 140 molecules containing only hydrogen, carbon, nitrogen, and/or oxygen; it is similar to the set used in the original MNDO development in the 1970s.[12,13] The standard FLUOR set comprises 48 fluorine-containing molecules. Both standard sets include only ground-state reference properties, mainly heats of formation, geometry data (bond lengths, bond angles, and dihedral angles), first ionization potentials, and dipole moments.[7,24−28,93]

Detailed numerical results for all reference molecules and reference properties are documented in the Supporting Information (Tables S1−S6) for some of the established MNDO-type methods (MNDO, AM1, PM3, PM6, and PM7) and for the OM$x$ methods (OM1, OM2, and OM3). A statistical evaluation of these results is given in Tables 4 and 5 for the CHNO and FLUOR sets, respectively. These tables provide mean absolute errors for the entire set and for chemically defined element-specific subsets.

The statistical evaluations suggest the following overall assessment. For the calculated heats of formation, the mean absolute errors (MAEs) are lowest overall for the OM$x$ methods; the most recent MNDO-type method PM7 that has been fitted on a much larger data set is closest in accuracy to the OM$x$ methods, but it has larger outliers (see Figure 1, Supporting Information Tables S1 and S5). Bond lengths and bond angles are reproduced similarly well by most SQC methods investigated

**Table 4. Mean Absolute Errors in Calculated Heats of Formation, Relative Energies and Barriers (kcal/mol), Bond Lengths (Å), Bond Angles (deg), Ionization Potentials (eV), and Dipole Moments (D) for the CHNO Set and Its Subsets: MNDO, AM1, PM*x*, and OM*x***

| subset | N | MNDO | AM1 | PM3 | PM6 | PM7 | OM1 | OM2 | OM3 |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | method | | | | |
| | | | | | Heats of Formation | | | | |
| overall | 140 | 6.34 | 5.51 | 4.17 | 4.80 | 3.80 | 3.46 | 3.10 | 3.05 |
| CH | 57 | 5.90 | 4.89 | 3.63 | 4.74 | 3.50 | 2.49 | 1.72 | 1.63 |
| CHN | 32 | 6.24 | 4.65 | 5.02 | 4.30 | 4.11 | 4.27 | 3.92 | 3.80 |
| CHO | 39 | 4.80 | 5.54 | 3.69 | 4.53 | 3.42 | 3.41 | 4.53 | 4.20 |
| CHNO | 4 | 13.52 | 8.36 | 3.00 | 2.59 | 1.48 | 6.18 | 1.96 | 3.24 |
| HNO | 8 | 13.73 | 11.93 | 7.54 | 9.63 | 7.71 | 5.94 | 3.28 | 4.53 |
| | | | | | Bond Lengths | | | | |
| overall | 242 | 0.014 | 0.017 | 0.011 | 0.012 | 0.011 | 0.012 | 0.016 | 0.019 |
| CH | 113 | 0.011 | 0.013 | 0.011 | 0.009 | 0.009 | 0.012 | 0.010 | 0.009 |
| CHN | 49 | 0.011 | 0.015 | 0.010 | 0.010 | 0.009 | 0.009 | 0.015 | 0.027 |
| CHO | 57 | 0.017 | 0.019 | 0.011 | 0.013 | 0.011 | 0.013 | 0.018 | 0.022 |
| CHNO | 5 | 0.015 | 0.017 | 0.016 | 0.011 | 0.010 | 0.013 | 0.018 | 0.033 |
| HNO | 18 | 0.036 | 0.038 | 0.016 | 0.036 | 0.036 | 0.019 | 0.049 | 0.043 |
| | | | | | Bond Angles | | | | |
| overall | 101 | 2.56 | 1.90 | 2.08 | 2.14 | 2.31 | 1.82 | 2.24 | 1.85 |
| CH | 38 | 1.81 | 1.36 | 1.57 | 1.64 | 1.88 | 1.45 | 1.46 | 1.23 |
| CHN | 22 | 2.03 | 1.86 | 2.04 | 1.86 | 2.04 | 1.85 | 2.30 | 1.82 |
| CHO | 31 | 3.11 | 2.17 | 2.43 | 2.38 | 2.69 | 1.95 | 2.45 | 2.03 |
| HNO | 10 | 4.83 | 3.17 | 3.04 | 3.91 | 3.30 | 2.76 | 4.42 | 3.76 |
| | | | | | Ionization Potentials | | | | |
| overall | 52 | 0.46 | 0.36 | 0.42 | 0.45 | 0.42 | 0.32 | 0.26 | 0.44 |
| CH | 22 | 0.38 | 0.27 | 0.34 | 0.31 | 0.37 | 0.24 | 0.24 | 0.37 |
| CHN | 13 | 0.54 | 0.37 | 0.52 | 0.53 | 0.52 | 0.33 | 0.22 | 0.39 |
| CHO | 14 | 0.57 | 0.45 | 0.51 | 0.48 | 0.41 | 0.41 | 0.34 | 0.61 |
| HNO | 3 | 0.24 | 0.52 | 0.16 | 0.96 | 0.52 | 0.33 | 0.23 | 0.45 |
| | | | | | Dipole Moments | | | | |
| overall | 63 | 0.36 | 0.24 | 0.28 | 0.38 | 0.38 | 0.23 | 0.25 | 0.26 |
| CH | 20 | 0.30 | 0.16 | 0.16 | 0.24 | 0.24 | 0.08 | 0.11 | 0.10 |
| CHN | 16 | 0.54 | 0.48 | 0.39 | 0.54 | 0.54 | 0.40 | 0.27 | 0.33 |
| CHO | 19 | 0.22 | 0.10 | 0.25 | 0.39 | 0.33 | 0.23 | 0.31 | 0.26 |
| HNO | 6 | 0.40 | 0.32 | 0.31 | 0.38 | 0.56 | 0.31 | 0.49 | 0.58 |
| | | | | | Relative Energies | | | | |
| overall | 17 | 9.01 | 5.70 | 4.17 | 2.97 | 1.97 | 5.25 | 1.96 | 2.83 |
| CH | 9 | 4.34 | 2.62 | 1.83 | 1.19 | 1.39 | 3.20 | 0.52 | 1.08 |
| CHN | 3 | 9.48 | 7.37 | 4.71 | 2.91 | 2.17 | 6.04 | 4.09 | 5.65 |
| CHO | 3 | 13.22 | 7.78 | 9.30 | 4.17 | 3.41 | 6.63 | 3.63 | 3.91 |
| | | | | | Barriers | | | | |
| overall | 60 | 3.48 | 2.12 | 3.01 | 2.55 | 2.46 | 2.31 | 1.63 | 1.61 |
| CH | 20 | 4.27 | 3.31 | 3.25 | 3.65 | 3.32 | 2.37 | 1.69 | 2.01 |
| CHN | 10 | 3.20 | 1.32 | 3.17 | 3.39 | 3.17 | 3.03 | 1.91 | 1.51 |
| CHO | 25 | 2.42 | 1.42 | 2.07 | 1.46 | 1.56 | 1.61 | 1.50 | 1.45 |
| CHNO | 3 | 7.98 | 3.16 | 7.70 | 2.35 | 2.46 | 5.42 | 1.43 | 0.68 |

presently. Ionization potentials are predicted best by OM2 for the CHNO set and by OM1 and OM2 for the FLUOR set. Dipole moments are again computed similarly well in general (please note that the OM*x* dipole moments are calculated in ZDO approximation without orthogonalization corrections, the effect of which should be studied in future work). The most accurate results for relative energies and barriers are obtained from OM2 and PM7 and from OM2 and OM3, respectively.

Upon the request of a reviewer, we have carried out analogous RM1 calculations. The statistical evaluations of the RM1 results are given in the Supporting Information (Tables S11 and S12) along with a brief assessment. RM1 tends to be generally more accurate than AM1, about as accurate as the PM*x* methods, and somewhat less accurate than the OM*x* methods.

We have performed extensive benchmarks on many other validation sets from the literature, to compare the performance of OM*x* and MNDO-type methods more systematically. These benchmarks are reported in a separate Article.[69]

## 3. DISPERSION CORRECTIONS

In this section, we discuss the effects of including dispersion corrections in the OM*x* methods, without changing any of the basic OM*x* parameters presented in Tables 1−3. Contrary to other recent semiempirical studies (see Introduction), we do not introduce any hydrogen-bond corrections because the OM*x*

**Table 5. Mean Absolute Errors in Calculated Heats of Formation (kcal/mol), Bond Lengths (Å), Bond Angles (deg), Ionization Potentials (eV), and Dipole Moments (D) for the FLUOR Set and Its Subsets: MNDO, AM1, PM*x*, and OM*x***

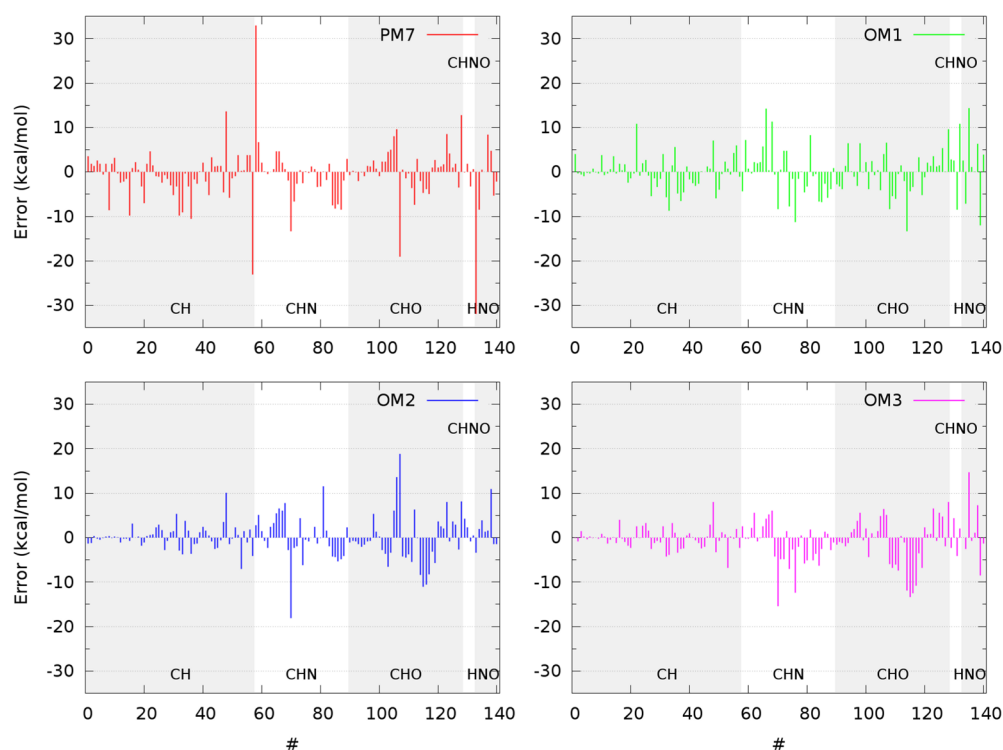| subset | N | method | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | MNDO | AM1 | PM3 | PM6 | PM7 | OM1 | OM2 | OM3 |
| | | | | Heats of Formation | | | | | |
| overall | 48 | 10.81 | 7.75 | 5.87 | 5.21 | 5.82 | 4.92 | 3.41 | 3.70 |
| CHF | 39 | 9.89 | 7.37 | 4.96 | 4.92 | 5.57 | 3.67 | 3.72 | 3.88 |
| HNOF | 9 | 14.79 | 9.42 | 9.81 | 6.48 | 6.90 | 10.32 | 2.08 | 2.93 |
| | | | | Bond Lengths | | | | | |
| overall | 125 | 0.037 | 0.028 | 0.022 | 0.016 | 0.017 | 0.020 | 0.023 | 0.024 |
| CHF | 104 | 0.030 | 0.026 | 0.020 | 0.014 | 0.015 | 0.018 | 0.019 | 0.021 |
| CHNOF | 3 | 0.029 | 0.052 | 0.021 | 0.013 | 0.017 | 0.024 | 0.020 | 0.014 |
| HNOF | 17 | 0.080 | 0.037 | 0.037 | 0.026 | 0.026 | 0.032 | 0.043 | 0.044 |
| | | | | Bond Angles | | | | | |
| overall | 69 | 3.00 | 3.16 | 2.75 | 2.76 | 2.55 | 1.97 | 2.23 | 1.78 |
| CHF | 56 | 2.59 | 2.93 | 2.69 | 2.64 | 2.45 | 1.88 | 2.06 | 1.61 |
| CHNOF | 3 | 2.89 | 3.76 | 3.41 | 3.76 | 2.61 | 1.30 | 2.44 | 1.75 |
| HNOF | 9 | 5.56 | 4.59 | 3.05 | 3.50 | 3.37 | 2.43 | 2.91 | 2.68 |
| | | | | Ionization Potentials | | | | | |
| overall | 39 | 0.34 | 0.53 | 0.39 | 0.50 | 0.46 | 0.23 | 0.26 | 0.32 |
| CHF | 29 | 0.33 | 0.47 | 0.32 | 0.36 | 0.40 | 0.20 | 0.25 | 0.29 |
| HNOF | 9 | 0.38 | 0.70 | 0.57 | 1.00 | 0.68 | 0.30 | 0.29 | 0.42 |
| | | | | Dipole Moments | | | | | |
| overall | 39 | 0.38 | 0.31 | 0.29 | 0.33 | 0.29 | 0.27 | 0.31 | 0.25 |
| CHF | 30 | 0.41 | 0.29 | 0.25 | 0.31 | 0.26 | 0.31 | 0.33 | 0.24 |
| HNOF | 8 | 0.29 | 0.44 | 0.47 | 0.42 | 0.41 | 0.15 | 0.26 | 0.29 |



**Figure 1.** Error distribution for the heats of formation from OM*x* and PM7 calculations on the CHNO set. The CH, CHN, CHO, CHNO, and HNO subsets are shown in this order from left to right and are marked with alternating gray and white backgrounds.

methods normally describe hydrogen bonds already quite well by themselves, even though further improvements can be achieved by specific re-parameterization.[94] We have previously shown that simple dispersion-corrected OM2 and OM3 approaches are capable of yielding acceptable results for hydrogen-bonded complexes.[67]

The functional form of the dispersion corrections is taken from the work by Grimme, specifically from DFT-D2[56] and DFT-D3(BJ).[57,58] Our notation is as follows: D2 is used for re-parametrized versions of the DFT-D2 corrections,[56] while D3 denotes the parametrization of the DFT-D3(BJ) correction[57,58] that has recently been performed by Grimme for

OM2 and OM3.[10] The dispersion corrections are added to the SCF energy. For ease of reference, we first outline the relevant formalism before presenting the results.

**3.1. D2 Corrections.** The D2 dispersion correction $E_{disp}^{D2}$ is given by[56,68]

$$E_{disp}^{D2} = -s_6 \sum_{A<B} \frac{C_6^{AB}}{R_{AB}^6} f_{damp}(R_{AB}) \tag{23}$$

Here $R_{AB}$ is the distance between atoms $A$ and $B$, $C_6^{AB}$ is the corresponding dispersion coefficient, $s_6$ is an element-independent (global) scale factor, and $f_{damp}(R_{AB})$ is a damping function that avoids an unphysical overbinding between atoms $A$ and $B$ that are very close to each other. The cross-terms $C_6^{AB}$ are obtained from

$$C_6^{AB} = \sqrt{C_6^A C_6^B} \tag{24}$$

where $C_6^A$ and $C_6^B$ are atomic coefficients derived from the London dispersion formula. The damping function $f_{damp}(R_{AB})$ is defined by[56,68]

$$f_{damp}(R_{AB}) = \frac{1}{1 + e^{-d[R_{AB}/(R_0^A + R_0^B) - 1]}} \tag{25}$$

where $R_0$ is an element-specific parameter that represents the van der Waals radius,[56] and $d$ is an element-independent (global) parameter that determines the steepness of the damping function.[55] Tests of an alternative damping function[95] in the OMx-D2 framework yield results close to those obtained with the aforementioned standard DFT-D2 damping function;[68] for the sake of brevity, we only present and discuss the latter in the following sections.

**3.2. D3 Corrections.** Recently, Risthaus and Grimme derived and successfully applied D3(BJ) corrections to OM2 and OM3.[10] The resulting dispersion-corrected methods are labeled as OM2-D3 and OM3-D3 both in the original work[10] and here. The D3(BJ) corrections are defined by[58]

$$E_{disp}^{D3(BJ)} = -\sum_{A<B} \left[ s_6 \frac{C_6^{AB}}{R_{AB}^6 + f(R_{AB}^0)^6} + s_8 \frac{C_8^{AB}}{R_{AB}^8 + f(R_{AB}^0)^8} \right] \tag{26}$$

where $s_6$ and $s_8$ are element-independent (global) parameters, $C_6^{AB}$ and $C_8^{AB}$ are dispersion parameters, and $f(R_{AB})$ denotes the Becke–Johnson-type (BJ) damping function[96–98] that contains adjustable parameters $a_1$ and $a_2$. The dispersion coefficients $C_6^{AB}$ are derived from TD-DFT calculations of average dipole polarizabilities.[57] The dispersion coefficients $C_8^{AB}$ are obtained recursively from $C_6^{AB}$ and DFT-derived quantities.[57]

Risthaus and Grimme[10] also applied three-body dispersion corrections ($E_{disp}^{ABC}$)[57] on top of the preceding two-body corrections. These three-body corrections are important for large dense systems.[99,100] They are defined explicitly in the literature.[10,57] Inclusion of the $E_{disp}^{ABC}$ corrections has been shown to improve the accuracy of OM2-D3 and OM3-D3 noncovalent interaction energies of large complexes on average by 1.4 kcal/mol[10] but is expected to have little effect on smaller systems.[10,100] We denote OM2-D3 and OM3-D3 calculations that include three-body corrections by OM2-D3T and OM3-D3T, respectively.

**3.3. Parameters.** The global parameters $s_6$ and $d$ as well as the element-specific parameters $C_6$ and $R_0$ that appear in the D2 dispersion corrections were fitted by optimizing the complexes of the S22 benchmark set[101,102] and using the respective reference interaction energies.[68] During this fitting, $C_6$ and $R_0$ were not allowed to change by more than 10% to ensure that they retain

their physical meaning.[68] The resulting element-specific parameters for H, C, N, and O do not differ significantly from the respective DFT-D2 parameters,[56,68] and therefore we decided to adopt for F the published DFT-D2 parameters (note that there are no fluorine-containing complexes in the S22 set). The final D2 parameters are listed in Table 6.

**Table 6. Parameters for the D2 and D3 Dispersion Corrections for OM2 and OM3: $s_6$, $s_8$, $a_1$, $a_2$, $d$, $C_6$ (J·nm$^6$·mol$^{-1}$), and $R_0$ (nm) (See Text)**

| Parameter | Element | | | | |
|---|---|---|---|---|---|
| | H | C | N | O | F |
| | | | OM2-D2$^a$ | | |
| $s_6$ | | | 0.5 | | |
| $d$ | | | 12.0 | | |
| $C_6$ | 0.13349720 | 1.74649564 | 1.22649173 | 0.68974720 | 0.75000000$^b$ |
| $R_0$ | 0.10685589 | 0.14169369 | 0.14069819 | 0.13069564 | 0.12870000$^b$ |
| | | | OM3-D2$^a$ | | |
| $s_6$ | | | 0.6 | | |
| $d$ | | | 18.6 | | |
| $C_6$ | 0.14146973 | 1.75148586 | 1.23148586 | 0.70136073 | 0.75000000$^b$ |
| $R_0$ | 0.10158635 | 0.15275442 | 0.14115458 | 0.13573278 | 0.12870000$^b$ |
| | | | OM2-D3$^c$ | | |
| $s_6$ | | | 1.000 | | |
| $s_8$ | | | 0.531 | | |
| $a_1$ | | | 0.690 | | |
| $a_2$ | | | 3.446 | | |
| | | | OM3-D3$^c$ | | |
| $s_6$ | | | 1.000 | | |
| $s_8$ | | | 0.501 | | |
| $a_1$ | | | 0.613 | | |
| $a_2$ | | | 3.258 | | |

$^a$Taken from ref 68. $^b$Taken from ref 56., see text. $^c$Taken from ref 10.

The OM2-D3 and OM3-D3 parameters are taken from the literature.[10] As commonly done, $s_6$ was fixed to unity,[10,58] while the short-range D3(BJ) parameters ($a_1$, $a_2$, and $s_8$) were fitted against reference data from the S66×8[103] set.[10] The optimized D3 parameters are included in Table 6.

**3.4. Validation.** Here we provide an initial assessment of the dispersion-corrected OMx-Dn results for the S22 and S66×8 training sets used during parametrization. Extensive further benchmarks have been performed for several other test sets, which will be reported in a separate Article.[69]

The S22 set is comprised of 22 small noncovalently bound complexes. Most of the reference geometries were obtained at the MP2/cc-pVTZ level with counterpoise corrections, and some also at the CCSD(T) level without counterpoise corrections.[101,102] The standard S22 reference energies have been updated in this study by including more accurate CCSD(T)/CBS values.[104] The S66×8 set samples intermolecular distances in 66 small noncovalently bound complexes, which were optimized at the MP2/cc-pVTZ level with counterpoise corrections;[103] thereafter the equilibrium distance between the monomers was scaled by factors of 0.9, 0.95, 1.05, 1.1, 1.25, 1.5, and 2.0, and reference energies were calculated at the CCSD(T)/CBS level for the resulting eight geometries.[103] In both sets, the complexes are divided into three types: electrostatics-dominated (hydrogen bonded), dispersion-dominated, and mixed electrostatics/dispersion.[101,103]

We performed single-point calculations at the reference geometries in order to be able to compare the results from different methods. Detailed numerical results are documented in the Supporting Information (Tables S9 and S10) for MNDO-type methods (MNDO, AM1, PM3, PM6, and PM7), for the OMx methods (OM1, OM2, and OM3), and in the case of OM2

**Table 7. Mean Absolute Errors of the Interaction Energies (kcal/mol) for the S22 and S66×8 Benchmark Sets from Single-Point PM7 and OMx-Dn Calculations at the Reference Geometries**

| | | | method | | | | | |
| | | | OM2 | | | OM3 | | |
| subset | N | PM7 | D2 | D3 | D3T | D2 | D3 | D3T |
|---|---|---|---|---|---|---|---|---|
| | | | | S22 | | | | |
| overall | 22 | 0.76 | 1.16 | 0.91 | 0.94 | 1.25 | 0.99 | 0.97 |
| hydrogen bonded | 7 | 0.73 | 2.32 | 2.17 | 2.15 | 2.65 | 2.30 | 2.28 |
| mixed | 7 | 0.70 | 0.40 | 0.27 | 0.28 | 0.79 | 0.47 | 0.50 |
| dispersion | 8 | 0.85 | 0.81 | 0.36 | 0.47 | 0.44 | 0.29 | 0.24 |
| | | | | S66×8 | | | | |
| overall | 528 | 0.73 | 0.96 | 0.78 | 0.79 | 0.88 | 0.69 | 0.71 |
| electrostatic | 184 | 0.91 | 1.54 | 1.43 | 1.43 | 1.48 | 1.30 | 1.30 |
| mixed | 160 | 0.55 | 0.52 | 0.37 | 0.37 | 0.53 | 0.33 | 0.34 |
| dispersion | 184 | 0.69 | 0.77 | 0.49 | 0.52 | 0.59 | 0.41 | 0.43 |

**Table 8. Mean Absolute Errors in Selected Interatomic Distances (Å) and Angles (deg) from PM7 and OMx-Dn Optimizations for the S22 Set**

| | | | method | | | | | |
| | | | OM2 | | | OM3 | | |
| subset | N | PM7 | D2 | D3 | D3T | D2 | D3 | D3T |
|---|---|---|---|---|---|---|---|---|
| | | | Selected Interatomic Distances | | | | | |
| overall | 105 | 0.565 | 0.287 | 0.281 | 0.424 | 0.265 | 0.281 | 0.285 |
| | | | Selected Angles | | | | | |
| overall | 14 | 10.76 | 2.47 | 2.63 | 2.60 | 1.08 | 0.86 | 0.85 |

and OM3 also for the dispersion-corrected variants (D2, D3, and D3T). Statistical evaluations are given in Table 7 for the PM7 and OMx-Dn methods and in Supporting Information Table S7 for the other SQC methods. These tables provide mean absolute errors of the interaction energies for the S22 and S66×8 sets and for the corresponding subsets.

In the case of the S22 set, the overall mean absolute error (MAE) is lowest for PM7 (0.76 kcal/mol); the corresponding MAEs for the OMx-Dn methods are slightly higher by 0.15−0.49 kcal/mol. However, the latter are more accurate for the dispersion-dominated and mixed-type complexes, with MAEs of 0.27−0.47 kcal/mol for OM2-D3 and OM3-D3. The largest outlier in the OMx-Dn results is the interaction energy of formic acid dimer which is underestimated by 3.96−5.87 kcal/mol.

Similar trends are found for the interaction energies in the S66×8 set. The overall MAEs for PM7, OM2-D3, and OM3-D3 are close to each other (0.73, 0.78, and 0.69 kcal/mol, respectively), and the OMx-Dn methods again outperform PM7 for dispersion-dominated and mixed-type complexes. The largest outlier in the OMx-Dn results is the interaction energy of acetic acid dimer which is underestimated by up to 7.0 kcal/mol.

Since all complexes in the S22 and S66×8 sets are quite small, it is not surprising that the role of the three-body corrections is negligible (see the D3 and D3T results).

To check the quality of geometry predictions for these noncovalent complexes, we performed full geometry optimizations for the S22 set at the PM7 and OMx-Dn levels and compared the results for 105 selected interatomic distances and for 14 selected bond angles with the *ab initio* reference values.[68] Statistical evaluation of the results are given in Table 8 for the PM7 and OMx-Dn methods and in Supporting Information Table S8 for other SQC methods. It is obvious that the MAEs for the selected distances and angles are quite large for all SQC

methods investigated presently. Still, the MAEs are significantly lower for the OMx-Dn methods than for PM7 (see Table 8).

For most of the complexes of the S22 set, the PM7 and OMx-Dn methods give qualitatively correct geometry predictions, but there are also cases of failure in which some of the investigated SQC methods yield a qualitatively wrong structure. For example, the reference conformation of the methane dimer is well-reproduced by the OM3-Dn methods but not by OM2-D2 and PM7 (Supporting Information Figure S1). The reference geometry of the T-shaped benzene dimer is predicted correctly by OM2-D2 and OM2-D3, while optimizations with OM3-Dn yield other types of conformers including some that resemble the parallel displaced structure (Supporting Information Figure S2). The optimization of the stacked indole···benzene complex leads to a reasonable structure with PM7, whereas OMx-Dn give a conformer in which benzene is too strongly shifted compared with the reference geometry (Supporting Information Figure S3) For the T-shaped indole···benzene complex, similar problems are encountered as in the case of the T-shaped benzene dimer: for example, optimizations at the OM3-Dn levels lead to a stacked geometry (Supporting Information Figure S4). Finally, in the case of the hydrogen-bonded formic acid dimer, the reference, PM7, and OM2-Dn geometries are similar (with each of the monomers containing an O−H bond), whereas geometry optimizations with OM3 and OM3-Dn converge to a cyclic $D_{2h}$ structure with a symmetric hydrogen bond characterized by equal O−H distances toward both monomers (Supporting Information Figure S5).

Although the three-body corrections are generally small for the S22 set, their inclusion in the OM2-D3T optimization of the T-shaped benzene dimer produces another conformer of the dimer (Supporting Information Figure S2). This causes the outlier in the MAE value for the OM2-D3T distances, which is much higher than that for OM2-D3 (Table 8).

## 4. COMPUTATIONAL DETAILS

All MNDO, AM1, PM3, and OM$x$ calculations were carried out with our locally modified `MNDO2005` program,[105] while the PM6, PM7, and RM1 calculations were done with the `MOPAC2012` program.[106,107] Molecules were visualized with Chemcraft 1.7.[108]

The SCF energy convergence criterion was set to $10^{-9}$ eV. In addition, the diagonal elements of the density matrix were required to change by less than $10^{-9}$ to achieve SCF convergence in `MNDO2005`.

By default, geometry optimizations were performed until the gradient norm was smaller than 0.01 kcal/(mol·Å). The optimizations employed the BFGS algorithm in `MNDO2005` and the eigenvector following method in `MOPAC2012`. In the latter case, the full Hessian matrix was constructed using single-sided derivatives and recalculated every 10 steps. In many cases, `MOPAC2012` stopped when the heat of formation remained essentially constant. In these cases the gradient norm was usually still smaller than 0.1 kcal/(mol·Å).

Open-shell systems up to triplet spin state were treated with the half-electron approach.[109] Molecular mechanics corrections for PM6 and PM7 were not used. No cutoffs were applied for the three-center orthogonalization corrections in OM$x$ and OM$x$-D$n$ calculations.

D3 dispersion corrections were implemented into the `MNDO2005` program by adapting code from the `DFT-D3` stand-alone program by Grimme (version 3.0 Rev 2).[110] The D3 corrections were computed using Becke−Johnson damping and default cutoffs (95 au for two-body terms and 40 au for coordination numbers). The D3T three-body terms were implemented starting from version 3.1 Rev 0 of the same program;[110] in this case, we increased the default cutoffs (95 au for two-body terms, coordination numbers, and three-body terms) which had almost no impact on computation times.

## 5. CONCLUSIONS

In previous work we have introduced orthogonalization-corrected SQC methods (OM1, OM2, and OM3) and applied them in various studies, especially in the area of excited-state dynamics. The OM$x$ methods go beyond the MNDO model by including orthogonalization terms, penetration integrals, and ECP integrals in the Fock matrix. In this Article, we describe the underlying theoretical formalism and its implementation in full detail and report all relevant parameters.

The performance of the OM$x$ methods for ground-state properties is demonstrated for a standard set of molecules that has commonly been used in semiempirical method development. In the majority of cases, the OM$x$ results are found to be superior to those from established MNDO-type methods. Further extensive benchmarks have been carried out and are reported separately.[69]

Dispersion interactions are not covered in the OM$x$ formalism but need to be included for a realistic description of noncovalent interactions. This has been done by adding parametrized Grimme-type dispersion corrections to OM2 and OM3 energies (computed without changing the standard OM2 and OM3 parameters). This treatment provides a much improved description of noncovalent interactions in the established S22 and S66×8 test sets. Further advances may be possible by a full reparametrization of the OM2-D3 and OM3-D3 approaches (work in progress).

## ■ ASSOCIATED CONTENT

### ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b01046.

> Detailed numerical results for the ground-state properties of all molecules in the CHNO and FLUOR sets for all SQC methods considered, individual interaction energies in the noncovalent complexes of the S22 and S66×8 set for all SQC methods considered, and plots of the optimized geometries of selected noncovalent complexes from the S22 set (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Author
*E-mail: thiel@kofo.mpg.de.

### Present Addresses
†BASF, Rue Marie de Bourgogne 58, 1000 Brussels, Belgium.
‡Arbentalstrasse 204, 8045 Zürich, Switzerland.
§Leberberg 11, 65193 Wiesbaden, Germany

### Notes
The authors declare no competing financial interest.

## ■ REFERENCES

(1) Thiel, W. *WIREs: Comput. Mol. Sci.* **2014**, *4*, 145−157 (see also references therein).

(2) Senn, H. M.; Thiel, W. *Angew. Chem., Int. Ed.* **2009**, *48*, 1198−1229.

(3) Clark, T.; Stewart, J. J. P. MNDO-Like Semiempirical Molecular Orbital Theory and Its Application to Large Systems. In *Computational Methods for Large Systems: Electronic Structure Approaches for Biotechnology and Nanotechnology*; Reimers, J. R., Ed.; John Wiley & Sons: Hoboken, NJ, USA, 2011; pp 259−286.

(4) Stewart, J. J. P. *J. Mol. Model.* **2013**, *19*, 1−32.

(5) Stewart, J. J. P. *J. Mol. Model.* **2007**, *13*, 1173−1213.

(6) Korth, M.; Thiel, W. *J. Chem. Theory Comput.* **2011**, *7*, 2929−2936.

(7) Otte, N.; Scholten, M.; Thiel, W. *J. Phys. Chem. A* **2007**, *111*, 5751−5755.

(8) Tirado-Rives, J.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2008**, *4*, 297−306.

(9) Dral, P. O.; Clark, T. *J. Phys. Chem. A* **2011**, *115*, 11303−11312.

(10) Risthaus, T.; Grimme, S. *J. Chem. Theory Comput.* **2013**, *9*, 1580−1591.

(11) Clark, T. *J. Mol. Struct.: THEOCHEM* **2000**, *530*, 1−10.

(12) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899−4907.

(13) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4907−4917.

(14) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(15) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209−220.

(16) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 221−264.

(17) Thiel, W.; Voityuk, A. A. *Theor. Chim. Acta* **1992**, *81*, 391−404.

(18) Thiel, W.; Voityuk, A. A. *Theor. Chim. Acta* **1996**, *93*, 315.

(19) Thiel, W.; Voityuk, A. A. *J. Phys. Chem.* **1996**, *100*, 616−626.

(20) Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601−1622 (see also references therein and in the Supporting Information).

(21) Tubert-Brohman, I.; Guimarães, C. R. W.; Repasky, M. P.; Jorgensen, W. L. *J. Comput. Chem.* **2004**, *25*, 138−150.

(22) Winget, P.; Horn, A. H. C.; Selçuki, C.; Martin, B.; Clark, T. *J. Mol. Model.* **2003**, *9*, 408−414.

(23) Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. *J. Comput. Chem.* **2006**, *27*, 1101−1111.

(24) Kolb, M. Ein neues semiempirisches Verfahren auf Grundlage der NDDO-Näherung: Entwicklung der Methode, Parametrisierung und

Anwendungen. Ph.D. thesis, Bergische Universität-Gesamthochschule Wuppertal, Wuppertal, Germany, 1991.

(25) Kolb, M.; Thiel, W. *J. Comput. Chem.* **1993**, *14*, 775−789.

(26) Weber, W. Ein neues semiempirisches NDDO-Verfahren mit Orthogonalisierungskorrekturen: Entwicklung des Modells, Parametrisierung und Anwendungen. Ph.D. thesis, Universität Zürich, Zürich, Switzerland, 1996.

(27) Weber, W.; Thiel, W. *Theor. Chem. Acc.* **2000**, *103*, 495−506.

(28) Scholten, M. Semiempirische Verfahren mit Orthogonalisierungskorrekturen: Die OM3 Methode. Ph.D. thesis, Universität Düsseldorf, Düsseldorf, Germany, 2003.

(29) Thiel, W. *Adv. Chem. Phys.* **1996**, *93*, 703−757.

(30) Silva-Junior, M. R.; Thiel, W. *J. Chem. Theory Comput.* **2010**, *6*, 1546−1564.

(31) Sattelmeyer, K. W.; Tubert-Brohman, I.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2006**, *2*, 413−419.

(32) Voityuk, A. A. *WIREs Comput. Mol. Sci.* **2013**, *3*, 515−527.

(33) Dreuw, A.; Head-Gordon, M. *J. Am. Chem. Soc.* **2004**, *126*, 4007−4016.

(34) Dreuw, A.; Head-Gordon, M. *Chem. Rev.* **2005**, *105*, 4009−4037.

(35) Koslowski, A.; Beck, M. E.; Thiel, W. *J. Comput. Chem.* **2003**, *24*, 714−726.

(36) Patchkovskii, S.; Thiel, W. *Theor. Chem. Acc.* **1997**, *98*, 1−4.

(37) Patchkovskii, S.; Koslowski, A.; Thiel, W. *Theor. Chem. Acc.* **2005**, *114*, 84−89.

(38) Fabiano, E.; Keal, T.; Thiel, W. *Chem. Phys.* **2008**, *349*, 334−347.

(39) Weingart, O.; Lan, Z.; Koslowski, A.; Thiel, W. *J. Phys. Chem. Lett.* **2011**, *2*, 1506−1509.

(40) Gámez, J. A.; Weingart, O.; Koslowski, A.; Thiel, W. *J. Chem. Theory Comput.* **2012**, *8*, 2352−2358.

(41) Schönborn, J. B.; Koslowski, A.; Thiel, W.; Hartke, B. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12193.

(42) Spörkel, L.; Cui, G.; Thiel, W. *J. Phys. Chem. A* **2013**, *117*, 4574−4583.

(43) Spörkel, L.; Jankowska, J.; Thiel, W. *J. Phys. Chem. B* **2015**, *119*, 2702−2710.

(44) Kazaryan, A.; Lan, Z.; Schäfer, L. V.; Thiel, W.; Filatov, M. *J. Chem. Theory Comput.* **2011**, *7*, 2189−2199.

(45) Cui, G.; Lan, Z.; Thiel, W. *J. Am. Chem. Soc.* **2012**, *134*, 1662−1672.

(46) Fabiano, E.; Thiel, W. *J. Phys. Chem. A* **2008**, *112*, 6859−6863.

(47) Lan, Z.; Fabiano, E.; Thiel, W. *J. Phys. Chem. B* **2009**, *113*, 3548−3555.

(48) Lan, Z.; Fabiano, E.; Thiel, W. *ChemPhysChem* **2009**, *10*, 1225−1229.

(49) Lan, Z.; Lu, Y.; Fabiano, E.; Thiel, W. *ChemPhysChem* **2011**, *12*, 1989−1998.

(50) Lu, Y.; Lan, Z.; Thiel, W. *Angew. Chem., Int. Ed.* **2011**, *50*, 6864−6867.

(51) Lu, Y.; Lan, Z.; Thiel, W. *J. Comput. Chem.* **2012**, *33*, 1225−1235.

(52) Heggen, B.; Lan, Z.; Thiel, W. *Phys. Chem. Chem. Phys.* **2012**, *14*, 8137.

(53) Cui, G.; Thiel, W. *Angew. Chem., Int. Ed.* **2013**, *52*, 433−436.

(54) Grimme, S. *WIREs Comput. Mol. Sci.* **2011**, *1*, 211−228.

(55) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463−1473.

(56) Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787−1799.

(57) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132*, 154104.

(58) Grimme, S.; Ehrlich, S.; Goerigk, L. *J. Comput. Chem.* **2011**, *32*, 1456−1465.

(59) McNamara, J. P.; Hillier, I. H. *Phys. Chem. Chem. Phys.* **2007**, *9*, 2362−2370.

(60) Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P. *J. Chem. Theory Comput.* **2009**, *5*, 1749−1760.

(61) Korth, M.; Pitoňák, M.; Řezáč, J.; Hobza, P. *J. Chem. Theory Comput.* **2010**, *6*, 344−352.

(62) Korth, M. *J. Chem. Theory Comput.* **2010**, *6*, 3808−3816.

(63) Řezáč, J.; Hobza, P. *J. Chem. Theory Comput.* **2012**, *8*, 141−151.

(64) Grimme, S. *Chem. - Eur. J.* **2012**, *18*, 9955−9964.

(65) Kromann, J. C.; Christensen, A.; Steinmann, C.; Korth, M.; Jensen, J. H. *PeerJ* **2014**, *2*, e449.

(66) Yilmazer, N. D.; Korth, M. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 169−175.

(67) Tuttle, T.; Thiel, W. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2159−2166.

(68) Wu, X. Semiempirical Quantum Chemistry on High-Performance Heterogeneous Computers. Ph.D. thesis, Universität Düsseldorf, Düsseldorf, Germany, 2013.

(69) Dral, P. O.; Wu, X.; Spörkel, L.; Koslowski, A.; Thiel, W. *J. Chem. Theory Comput.*, following paper in this issue (DOI: 10.1021/acs.jctc.5b01047).

(70) Pople, J. A.; Santry, D. P.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S129−S135.

(71) Stewart, J. J. P. *J. Mol. Model.* **2004**, *10*, 6−12.

(72) Thiel, W. *Tetrahedron* **1988**, *44*, 7393−7408.

(73) Thiel, W. *J. Mol. Struct.: THEOCHEM* **1997**, *398−399*, 1−6.

(74) Thiel, W. Semiempirical methods. In *Modern methods and algorithms of quantum chemistry*; Grotendorst, J., Ed.; NIC Series; John von Neumann Institute for Computing: Jülich, Germany, 2000; Vol. 3; pp 261−283.

(75) Thiel, W. Semiempirical Quantum-Chemical Methods in Computational Chemistry. In *Theory and Applications of Computational Chemistry: The First Forty Years*; Dykstra, C. E., Frenking, G., Kim, K. S., Scuseria, G. E., Eds.; Elsevier: Amsterdam, 2005; pp 559−580.

(76) Klopman, G. *J. Am. Chem. Soc.* **1964**, *86*, 4550−4557.

(77) Ohno, K. *Theor. Chim. Acta* **1964**, *2*, 219−227.

(78) Dewar, M. J. S.; Thiel, W. *Theor. Chim. Acta* **1977**, *46*, 89−104.

(79) Löwdin, P. *J. Chem. Phys.* **1950**, *18*, 365−375.

(80) Löwdin, P.-O. *Adv. Quantum Chem.* **1970**, *5*, 185−199.

(81) Goeppert-Mayer, M.; Sklar, A. L. *J. Chem. Phys.* **1938**, *6*, 645−652.

(82) Pople, J. A.; Segal, G. A. *J. Chem. Phys.* **1965**, *43*, S136−S151.

(83) Krauss, M.; Stevens, W. J. *Annu. Rev. Phys. Chem.* **1984**, *35*, 357−385.

(84) Stevens, W. J.; Basch, H.; Krauss, M. *J. Chem. Phys.* **1984**, *81*, 6026−6033.

(85) Wadt, W. R.; Hay, P. J. *J. Chem. Phys.* **1985**, *82*, 284−298.

(86) Cao, X.; Dolg, M. *WIREs Comput. Mol. Sci.* **2011**, *1*, 200−210.

(87) Wu, X.; Koslowski, A.; Thiel, W. *J. Chem. Theory Comput.* **2012**, *8*, 2272−2281.

(88) Thiel, W.; Green, D. G. The MNDO94 Code: Parallelization of a Semiempirical Quantum-chemical Program. In *Methods and Techniques in Computational Chemistry: METECC-95*; Clementi, E., Corongiu, G., Eds.; NIC Series; STEF: Cagliari, Italy, 1995; Vol. 3; pp 141−168.

(89) Patchkovskii, S.; Thiel, W. *Theor. Chim. Acta* **1996**, *93*, 87−99.

(90) Patchkovskii, S.; Thiel, W. *J. Comput. Chem.* **1996**, *17*, 1318−1327.

(91) Levenberg, K. *Quart. Appl. Math.* **1944**, *2*, 164−168.

(92) Marquardt, D. W. *J. Soc. Ind. Appl. Math.* **1963**, *11*, 431−441.

(93) Steiger, R. Quantenchemische Programmenentwicklung: Automatische Erzeugung von Ableitungen und Parametrisierung semiempirischer Methoden. Ph.D. thesis, Universität Zürich, Zürich, Switzerland, 2003.

(94) Wu, X.; Thiel, W.; Pezeshki, S.; Lin, H. *J. Chem. Theory Comput.* **2013**, *9*, 2672−2686.

(95) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaxiras, E. *J. Chem. Phys.* **2001**, *114*, 5149−5155.

(96) Johnson, E. R.; Becke, A. D. *J. Chem. Phys.* **2005**, *123*, 024101.

(97) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *123*, 154101.

(98) Johnson, E. R.; Becke, A. D. *J. Chem. Phys.* **2006**, *124*, 174104.

(99) Tkatchenko, A.; von Lilienfeld, O. A. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2008**, *78*, 045116.

(100) von Lilienfeld, O. A.; Tkatchenko, A. *J. Chem. Phys.* **2010**, *132*, 234109.

(101) Jurečka, P.; Šponer, J.; Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985−1993.

(102) *BEGDB: Benchmark Energy and Geometry DataBase*. Accessed on Aug. 4, 2014. http://www.begdb.com/.

(103) Řezáč, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2011**, 7, 2427−2438.

(104) Takatani, T.; Hohenstein, E. G.; Malagoli, M.; Marshall, M. S.; Sherrill, C. D. *J. Chem. Phys.* **2010**, 132, 144104.

(105) Thiel, W. *MNDO2005*, version 7.0. Max-Planck-Institut für Kohlenforschung: Mülheim an der Ruhr, Germany, 2005.

(106) Stewart, J. J. P. *MOPAC2012*, versions 13.326L and 15.027L; Stewart Computational Chemistry, Colorado Springs, CO, USA; http://OpenMOPAC.net.

(107) Maia, J. D. C.; Carvalho, G. A. U.; Mangueira, C. P., Jr.; Santana, S. R.; Cabral, L. A. F.; Rocha, G. B. *J. Chem. Theory Comput.* **2012**, 8, 3072−3081.

(108) Andrienko, G. A. *ChemCraft*; http://www.chemcraftprog.com, 2014.

(109) Dewar, M. J. S.; Hashmall, J. A.; Venier, C. G. *J. Am. Chem. Soc.* **1968**, 90, 1953−1957.

(110) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *DFT-D3*, versions 3.0 and 3.1; Universität Bonn, Bonn, Germany, 2014; http://www.thch.uni-bonn.de/tc/index.php?section=downloads&subsection=getd3.