Semiempirical quantum-chemical methods

Walter Thiel
Max-Planck-Institut für Kohlenforschung
Kaiser-Wilhelm-Platz 1, 45470 Mülheim, Germany
thiel@kofo.mpg.de

Abstract

The semiempirical methods of quantum chemistry are reviewed, with emphasis on established NDDO-based methods (MNDO, AM1, PM3) and on the more recent orthogonalization-corrected methods (OM1, OM2, OM3). After a brief historical overview, the methodology is presented in non-technical terms, covering the underlying concepts, parameterization strategies, and computational aspects, as well as linear scaling and hybrid approaches. The application section addresses selected recent benchmarks and surveys ground-state and excited-state studies, including recent OM2-based excited-state dynamics investigations.

Introduction

Quantum mechanics provides the conceptual framework for understanding chemistry and the theoretical foundation for computational methods that model the electronic structure of chemical compounds. There are three types of such approaches: Quantum-chemical *ab initio* methods provide a convergent path to the exact solution of the Schrödinger equation and can therefore give "the right answer for the right reason", but they are costly and thus restricted to relatively small molecules (at least in the case of the highly accurate correlated approaches). Density functional theory (DFT) has become the workhorse of computational chemistry because of its favourable price/performance ratio, allowing for fairly accurate calculations on medium-size molecules, but there is no systematic path of improvement in spite of the first-principles character of DFT. Quantum-chemical semiempirical methods are the simplest variant of electronic structure theory, involving integral approximations and parameterizations that limit their accuracy but make them very efficient, so that large molecules can be modelled in a realistic manner.

These semiempirical methods adhere to a simple strategy. They start out from an *ab initio* or first-principles formalism and then introduce rather drastic assumptions to speed up the calculations, typically by neglecting many of the less important terms in the underlying equations. In order to compensate for the errors caused by these approximations, empirical parameters are incorporated into the formalism and calibrated against reliable experimental or theoretical reference data. If the chosen semiempirical model retains the essential physics to describe the properties of interest, the parameterization may account for all other effects in an average sense, and it is then a matter of validation and benchmarking to establish the numerical accuracy of such methods.

In this review, we consider only semiempirical methods that are based on molecular orbital (MO) theory and make use of integral approximations and parameters already at the MO level. In the following sections, we give a brief historical overview, survey the methodology, summarize selected applications, and offer an outlook. The presentation is deliberately non-technical, and the readers are referred to the original publications for more detailed information, especially with regard to the formalism, implementation, and validation of these methods.

Notation and Scope

Standard dictionaries define the term "*semiempirical"* as "*involving assumptions, approximations, or generalizations designed to simplify calculation or to yield a result in accord with observation*". Semiempirical approaches to quantum chemistry are thus characterized by the use of empirical parameters in a quantum mechanical framework. Strictly speaking, some contemporary *ab initio* and many advanced DFT methods could thus be labelled as "*semiempirical*", because they include a – sometimes substantial – number of empirical parameters. We do not cover such approaches here, but follow the conventional classification of electronic structure methods.

HISTORICAL OVERVIEW

One of the earliest semiempirical approaches in quantum chemistry was the π-electron method of Hückel (1931),[1] which generates MOs essentially from the connectivity matrix of a molecule and provides valuable qualitative insights into the structure, stability, and spectroscopy of unsaturated molecules. The extended Hückel theory by Hoffmann (1963)[2] includes all valence electrons and has been applied in many qualitative studies of inorganic and organometallic compounds. These early semiempirical methods have had a lasting impact on chemical thinking because they guided the development of qualitative MO theory which is commonly employed for rationalizing chemical phenomena in terms of orbital interactions.

Hückel-type methods include only one-electron integrals and are therefore non-iterative. Two-electron interactions are taken into account explicitly in semiempirical self-consistent-field (SCF) methods. Again, the first such approaches were restricted to π electrons, most notably the Pariser-Parr-Pople (PPP) method (1953),[3,4] which reliably describes the electronic spectra of unsaturated molecules. The generalization to valence electrons was proposed by Pople (1965),[5] who introduced a hierarchy of integral approximations that satisfy rotational invariance and other consistency criteria (CNDO complete neglect of differential overlap, INDO intermediate neglect of differential overlap, NDDO neglect of diatomic differential overlap).

The original parameterization of these all-valence-electron MO methods was designed to reproduce *ab initio* Hartree-Fock (HF) results obtained with a minimal basis set. This gives rise to approximate MO treatments that can at best reach the accuracy of the target *ab initio* HF methods (meanwhile known to be rather poor). Prominent examples are the original CNDO/2 and INDO methods.[6,7]

A different parameterization strategy was pursued by Dewar (1967-1990). Aiming at a realistic description of ground-state potential surfaces, particularly for organic molecules, he advocated calibration against experimental reference data. This work culminated in an INDO-based method named MINDO/3 (1975)[8] and two NDDO-based methods labelled MNDO (1977)[9,10] and AM1 (1985).[11] A later parameterization of the MNDO model gave rise to PM3 (1989).[12,13] Formally, AM1 and PM3 differ from MNDO only in the choice of the empirical core repulsion function, and they can therefore be viewed as attempts to explore the limits of the MNDO electronic structure model through extensive parameterization.

In the time before 1990, there are two other noteworthy INDO-based developments. In the work by Jug (1973-1990), orthogonalization corrections to the one-electron integrals were included in an INDO scheme, and parameterization against ground-state properties then led to the SINDO1 method,[14,15] which was later modified and upgraded to MSINDO.[16,17] The INDO/S method[18,19] developed by Zerner (1973-1990) targets the calculation of electronic spectra, in particular vertical excitation energies, using configuration interaction with single excitations (CIS). INDO/S was parameterized at the CIS level and turned out to be remarkably successful in spectroscopic and related applications.[20]

In the time since 1990, the MNDO model has been generalized from an *sp* basis[9,21] to an *spd* basis (1992-1996),[22,23] which enabled the treatment of heavier elements and led to improved MNDO/d results, especially in the case of hypervalent main-group elements. This extension to an *spd* basis has been widely adopted, in particular also in the latest general-purpose parameterizations giving rise to PM6 (2007) and PM7 (2013).[24,25] These two methods cover essentially the whole periodic table and can be used to compute both molecular and solid-state properties.[24,25] Another general-purpose parameterization of the MNDO model employed a modified empirical core repulsion function using Pairwise Distance Directed Gaussians (PDDG) and yielded the rather accurate PDDG/MNDO and PDDG/PM3 variants (2002-2004).[26,27] The strategy of changing or enhancing the core repulsion function in the MNDO model was also adopted in a number of special-purpose parameterizations (see next section), for example in recent work addressing dispersion and hydrogen bonding interactions (since 2008).[28,29] Finally, we note that a straight re-parameterization of the AM1 method with a much larger set of reference data afforded the general-purpose RM1 variant with improved results (2006).[30]

Going beyond the MNDO model, a series of orthogonalization models (OM1, OM2, and OM3) have been proposed that include orthogonalization corrections in the one-electron terms of the NDDO Fock matrix to account for the effects of Pauli exchange repulsion.[31-34] OM1 (1993) incorporates such corrections only in the one-center one-electron part, while OM2 (1996) and OM3 (2003) include them also in the two-center one-electron terms. OM3 disregards some of the smaller corrections, without significant loss of accuracy. The explicit representation of Pauli exchange repulsion in the OMx methods was shown to improve the description of conformational properties, non-covalent interactions, and electronically excited states.

Conceptually, the semiempirical methods outlined above are simplified *ab initio* MO treatments. Semiempirical tight-binding (TB) versions of DFT methods are also available, most notably the original DFTB approach (1986)[35,36] and the self-consistent-charge (SCC) DFTB method (1998).[37,38] Even though the conceptual origin and the derivation of tight-binding DFTB approaches seem quite different from those of conventional semiempirical methods, the implementations and the actual computational procedures share many similarities.[39] Since the DFTB methods employ severe integral approximations and extensive parameterization, it is appropriate to consider them as semiempirical methods on par with the traditional ones.[39]

In the present overview, we focus on the methods that are nowadays of practical relevance. Historically, mainly during the 1980s and 1990s, the MNDO-type methods served as workhorse for quantum-chemical computations, especially on medium-size and large molecules, as can be seen from the current combined citation count of more than 32,000 for the basic MNDO, AM1, and PM3 papers.[9-13] At present, these methods are still widely used, along with more recent versions like PM6,[24] even though DFT calculations have become dominant overall. INDO/S continues to be useful in many applications.[20] The OMx methods appear to be the most accurate among the available semiempirical approaches, with particular merits for electronically excited states.[39-41] DFTB methods are popular in studies on biochemistry and materials science.[36,38] Recent reviews in this journal have covered INDO/S,[20] DFTB,[36] and SCC-DFTB,[38] and therefore we concentrate on the MNDO-type and OMx treatments in the following.

METHODOLOGY

In this section, we provide a broad and non-mathematical review on the MNDO-type and OMx methods. The formalism is described in more detail in the original publications and several comprehensive review articles.[42-49]

Basic concepts

A semiempirical *model* is defined by the underlying theory and the integral approximations that determine the types of interactions included. The MNDO and OMx models employ a Hartree-Fock SCF-MO treatment for the valence electrons with a minimal basis set. The core electrons are taken into account through a reduced nuclear charge (assuming complete shielding) and, in addition, through an effective core potential at the OMx level. Electron correlation is treated explicitly only if this is necessary for an appropriate zero-order description. Dynamic correlation effects are included in an average sense by a suitable representation of the two-electron integrals and the overall parameterization.

The standard Hartree-Fock SCF-MO equations are simplified by *integral approximations* that are designed to neglect all three-center and four-center two-electron integrals. The CNDO, INDO, and NDDO schemes have been introduced for this purpose.[5] The MNDO-type and OMx methods make use of NDDO, which is the most refined of these schemes and does not require any spherical averaging for maintaining rotational invariance. NDDO retains the higher multipoles of charge distributions in the two-center interactions (unlike CNDO and INDO which truncate after the monopole), and therefore accounts for anisotropies in these interactions. The NDDO approximation is applied to all integrals that involve Coulomb interactions, and to the overlap integrals that appear in the Hartree-Fock secular equations. The eigenvector matrix **C** and the diagonal eigenvalue matrix **E** are thus obtained from simplified secular equations (**FC**=**CE**) through diagonalization of the Fock matrix **F**.

The *implementation* of a semiempirical model specifies the evaluation of all non-vanishing integrals and introduces the associate parameters. The integrals are either determined directly from experimental data or calculated exactly from the corresponding analytic formulas or represented by suitable parametric expressions. The first option is generally only feasible for the one-center integrals which may be derived from atomic spectroscopic data. The selection of appropriate parametric expressions is guided by an analysis of the corresponding analytic integrals or by intuition.

The *parameterization* of a given implementation serves to determine optimum parameter values by calibrating against suitable reference data. The MNDO-type and OMx methods generally adhere to the semiempirical philosophy and attempt to reproduce experiment. In the absence of reliable experimental reference data, accurate theoretical data (e.g., from high-level *ab initio* calculations) are accepted as substitutes. The quality of the results is strongly influenced by the effort put into the parameterization.

MNDO-type methods

The MNDO-type methods employ Slater-type atomic orbitals (AOs) as basis functions. In the Fock matrix, the one-center integrals are derived from atomic spectroscopic data, with the refinement that slight adjustments may be allowed in the parameterization (to a different extent in different implementations) . The one-center two-electron integrals provide the one-center limit of the two-center two-electron integrals, while the asymptotic limit at large distances is determined by classical electrostatics. These limits are satisfied by evaluating the two-center two-electron integrals from semiempirical multipole-multipole interactions damped according to the Klopman-Ohno formula.[21] At small and intermediate distances, these integrals are smaller than their analytic counterparts, which reflects some average inclusion of electron correlation effects. Aiming for a reasonable balance between electrostatic attractions and repulsions within a molecule, the two-center core-electron attractions and core-core repulsions are expressed in terms of appropriate two-electron integrals. An effective atom-pair potential (with an essentially exponential repulsion) is added to the core-core term in an attempt to account for Pauli exchange repulsions and also to compensate for errors introduced by other assumptions. Finally, the two-center one-electron resonance integrals are taken to be proportional to the corresponding overlap integrals.

The parameterization of the original MNDO method[9] focused on ground-state properties, mainly heats of formation and geometries, using ionization potentials and dipole moments as additional reference data. The choice of heats of formation as reference data implies that the parameterization must account for zero-point vibrational energies and for thermal corrections between 0 K and 298 K in an average sense. This is not satisfactory theoretically, but it has been shown empirically that the overall performance of MNDO is not affected much by this choice.

AM1[11] and PM3[12,13] are based on exactly the same model as MNDO, and they differ from MNDO only in one aspect of the implementation: the effective atom-pair potential in the core-core repulsion function is represented by a more flexible function with several additional adjustable parameters. The corresponding additional Gaussian terms are not derived theoretically, but justified empirically as providing more opportunities for fine tuning. The parameterization in AM1 and PM3 followed the same philosophy as in MNDO, but was more extensive: additional terms were treated as adjustable parameters so that the number of optimized parameters per element was typically increased from 5-7 in MNDO to 18 in PM3.

MNDO, AM1, and PM3 employ an *sp* basis without *d* orbitals in their original implementation. Therefore, they cannot be applied to most transition metal compounds, and there may be inadequacies for hypervalent compounds of main-group elements where the importance of *d* orbitals for quantitative accuracy is well documented at the *ab initio* level. In the MNDO/d extension, the established MNDO formalism and parameters remain unchanged for hydrogen, helium, and the first-row elements. The inclusion of *d* orbitals for the heavier elements requires a generalized semiempirical treatment of the two-electron interactions. The two-center two-electron integrals for an *spd* basis are calculated by a extension[22] of the original point-charge model for an *sp* basis,[21] with a truncation of the semiempirical multipole-multipole interactions at the quadrupole level. All nonzero one-center two-electron integrals are retained to ensure rotational invariance. The implementation and parameterization of MNDO/d are analogous to MNDO, with only very minor variations.[22,23] The two-electron integral scheme for an *spd* basis[22] from MNDO/d can be implemented in combination with any MNDO-type method. It is used, for example, in the latest general-purpose methods, PM6[24] and PM7.[25]

Most of the MNDO-type methods have been parameterized for many elements and are thus broadly applicable. However, MNDO/d parameters are available for the second-row elements, the halogens, and the zinc group elements. On the other extreme, PM6 parameters have been determined for 70 elements, thus covering most of the Periodic Table.

OMx methods

The *ab initio* SCF-MO secular equations include overlap and require a transformation from the chosen non-orthogonal to an orthogonal basis to arrive at the standard eigenvalue problem in the orthogonal basis, $\mathbf{F}\,\mathbf{C} = \mathbf{C}\,\mathbf{E}$. The semiempirical integral approximations yield such secular equations without overlap directly (see above). This suggests that the semiempirical Fock matrix implicitly refers to an orthogonalized basis and that the semiempirical integrals should thus be associated with theoretical integrals in an orthogonalized basis.

In the case of the two-electron integrals, this provides the traditional justification for the NDDO approximation.[45] The one-electron integrals are usually affected strongly by the orthogonalization. In an *ab initio* framework, the Pauli exchange repulsions and the asymmetric splitting of bonding and antibonding orbitals arise from these corrections. When using the standard semiempirical integral approximations, these and related effects are formally neglected, and many deficiencies of semiempirical methods may be attributed to this cause.[45]

As mentioned above, the MNDO-type methods attempt to incorporate the effects of Pauli exchange repulsion in an empirical manner, through an effective atom-pair potential that is added to the core-core repulsion. The OMx methods include the underlying orthogonalization corrections explicitly in the electronic calculation and thus do not have an effective atom-pair potential in the core-core repulsion. In a semiempirical context, the dominant one-electron corrections can be represented by parametric functions that reflect the second-order expansions of the Löwdin orthogonalization transformation in terms of overlap and that may be adjusted during the parameterization process.

Following previous INDO-based work,[14,15] these basic ideas were implemented at the NDDO level in three steps.[31-34] First, the valence-shell orthogonalization corrections were introduced only in the one-center part of the core Hamiltonian (OM1).[31] In the second step, they were also included in the two-center part of the core Hamiltonian, i.e., in the resonance integrals (OM2).[32,33] In the third step, less important correction terms were omitted (OM3).[34] OM1 contains only one-center and two-center terms, whereas OM2 and OM3 include three-center contributions in the corrections to the resonance integrals, which reflect the stereochemical environment of each electron pair bond and should thus be important for modeling conformational properties.

Concerning the implementation, the MNDO and OMx models share many features, but there are also some distinctions. For example, the OMx methods use Gaussian basis orbitals and a flexible empirical representation of the resonance integrals (no longer proportional to the overlaps). The

OMx parameterization followed the same strategy as in the MNDO case, using experimental ground-state reference data (see above). Up to now, OM1, OM2 and OM3 have been parameterized for the elements H, C, N, O, and F.

The OMx methods perform well for ground-state properties and offer consistent improvements over the established MNDO-type methods in statistical validations.[31-34,39,41] More significant qualitative advances are found in several other areas where the explicit inclusion of Pauli exchange repulsions is expected to be important, in particular for excited states, conformational properties, and hydrogen bonds.[31-34,39-41] In view of this good overall performance, it is clearly desirable to extend the OMx parameterization to other elements.

Special-purpose parameterization of MNDO-type and OMx methods

General-purpose semiempirical methods attempt to describe all classes of compounds and many properties simultaneously and equally well. It is obvious that compromises cannot be avoided in such an ambitious endeavour. One way forward is to develop specialized semiempirical methods for certain classes of compounds or specific properties, anticipating that such methods should be more accurate in their area of applicability than the general-purpose methods.

Specialized semiempirical treatments of this kind exist (see Ref 45 for a review on early work). Many of them are based on the MNDO model in one of its standard implementations. For example, there are several early MNDO, AM1, and PM3 variants with a special treatment of hydrogen bonds,[45,50] which have recently been refined with elaborate PM6-based hydrogen-bond potential functions.[28,29] These special approaches exploit the flexibility offered in MNDO-type methods by the presence of the effective core-core repulsion term, which can be modified for fine tuning.

In a similar vein, dispersion corrections have recently been included both in MNDO-type and OMx methods.[28,29,51] Following the pioneering work of Grimme at the DFT level,[52,53] damped attractive dispersion terms with asymptotic R^{-6} behavior were added to the core-core term, with very limited parameterization (typically only 2 global parameters, without the need to change the standard semiempirical parameters). Adding such physically sound dispersion corrections to the energy expression gave significant improvements in the OMx results for systems with strong non-covalent interactions.[51] We note in this context that PM7, being the most recent general-purpose method, includes both dispersion and hydrogen bond corrections.[25]

Other special-purpose approaches focus on chemical reactions and employ NDDO-based methods with specific reaction parameters (NDDO-SRP).[54,55] This concept has been adopted by a number of groups for direct dynamics calculations. Typically, the parameters in a standard method such as AM1 are adjusted to optimize the potential surface for an individual reaction or a set of related reactions (allowing moderate parameter variations up to 10%). This is done by fitting against experimental data (reaction energies, barriers) and/or high-level *ab initio* data (relevant points on the potential surface of a suitable model system). The NDDO-SRP scheme serves as a robust and economic protocol for generating realistic potential surfaces in a cost-effective manner. A recent example[56] is the re-parameterization of AM1 for modelling the reduction of 7,8-dihydrofolate by nicotinamide adenine dinucleotide phosphate hydride (NADPH) in the enzyme dihydrofolate reductase (DHFR):

the corresponding SRP-AM1 Hamiltonian was used in QM/MM simulations of the DHFR-catalyzed reaction to compute kinetic isotope effects (using a path-integral approach) in excellent agreement with experiment. Another example[57] is provided by re-parameterized semiempirical models for proton transfer in water, which are particularly accurate at the OM3 level and may be used in QM/MM simulations of proton transfer in solvated biological systems.

Special parameterizations are also available for a number of properties including electrostatic potentials and effective atomic charges for use in biomolecular modeling.[45] The semiempirical calculation of NMR chemical shifts at the GIAO-MNDO level also requires special parameterization. In this approach,[58] the NMR shielding tensor is evaluated in MNDO approximation using gauge-including atomic orbitals (GIAO) and analytic derivative theory. GIAO-MNDO calculations with standard MNDO parameters overestimate the variation of the paramagnetic contribution to the NMR chemical shifts because of the systematic underestimation of excitation energies in MNDO. This can be rectified by a re-parameterization[58] where some of the usual MNDO parameters are adjusted to increase the gap between occupied and unoccupied MOs. The resulting decrease of the paramagnetic contributions leads to significant improvements: the mean absolute errors of the computed NMR shifts drop to less than 5% of the total chemical shift range of a given element (e.g., to 8 ppm for C). The GIAO-MNDO approach also provides realistic nucleus-independent chemical shifts that are often used as a magnetic criterion for aromaticity: the aromatic or antiaromatic character can normally be assigned correctly on the basis of the GIAO-MNDO results.[59]

It is obvious from these examples that special-purpose parameterizations of semiempirical models can be used in a pragmatic matter to enhance their accuracy for specific applications.

Computational aspects

Semiempirical SCF-MO methods are designed to be efficient. The commonly applied integral approximations make integral evaluation scale as $O(N^2)$ for N basis functions. In OM2 and OM3, the two-center one-electron orthogonalization corrections formally scale as $O(N^3)$, but they contain products of terms that decrease exponentially with distance, and hence the computational effort can effectively be reduced to $O(N^2)$ by pre-screening. For large molecules, the time for integral evaluation becomes negligible compared with the $O(N^3)$ steps in the solution of the secular equations and the formation of the density matrix. Full diagonalizations of the Fock matrix are avoided – whenever possible – by adopting a pseudo-diagonalization procedure[60] that involves transformation of the Fock matrix from the AO basis to the MO basis and subsequent non-iterative annihilation of matrix elements in the occupied-virtual block through 2x2 Jacobi rotations. Therefore, most of the work in the $O(N^3)$ steps occurs in matrix multiplications. The small number of nonzero integrals implies that there are usually no input/output bottlenecks. The calculations can normally be performed completely in memory, with memory requirements that scale as $O(N^2)$.

Given these characteristics, it is evident that large-scale semiempirical SCF-MO calculations are ideally suited for vectorization and shared-memory parallelization. The dominant matrix multiplications can be performed very efficiently by BLAS library routines, and the remaining minor tasks of integral evaluation and Fock matrix construction can also be handled well on parallel vector processors with shared memory.[47] MNDO calculations on large fullerenes run with a speed close to the hardware limit on shared-memory machines.[61] The situation is less advantageous for massively parallel (MP) systems with distributed memory. Several fine-grained parallel implementations of semiempirical SCF-MO codes on MP hardware are available, but satisfactory overall speedups are normally obtained only for relatively small numbers of nodes (see Ref 47 for details).

In recent years, graphics processing units (GPUs) are increasingly used as coprocessors on hybrid multicore CPU-GPU platforms, to accelerate numerically intensive computations. The most time-consuming parts of semiempirical SCF-MO calculations (see above) run exceedingly fast on GPUs. The port of the MNDO code to a CPU-GPU platform made use of vendor-provided library functions for the $O(N^3)$ matrix operations, supplemented with a special GPU kernel for the non-iterative Jacobi rotations during pseudo-diagonalization.[62] The overall computation times for single-point energy evaluations and geometry optimizations of large molecules were reduced by an order of magnitude, both for MNDO-type and OMx methods.[62] An independent port of the MOPAC code also utilized library functions for matrix operations to speed up MNDO-type calculations.[63]

Efficient explorations of potential surfaces require the derivatives of the energy with respect to the nuclear coordinates. The analytic derivatives generally contain contributions from integral derivatives and from density matrix derivatives (CPHF terms from a coupled-perturbed Hartree-Fock treatment). At the semiempirical level, the evaluation of the former normally involves little computational effort since there are only $O(N^2)$ integral derivatives due to the neglect of most integrals. Efficient analytic derivative codes are available for computing the gradient[64,65] and the harmonic force constants[66] in MNDO-type methods. In the case of the semiempirical CI gradient, the original $O(N^4)$ CPHF implementation[67] was replaced by a procedure that makes use of the **Z**-vector method[68] and scales as $O(N^3)$, with speedups by a factor of about N. The analytic CI gradient code[65] is general and can be used both with minimal CI treatments (traditionally applied in semiempirical work) and with elaborate semiempirical multi-reference MR-CI approaches.[69,70] In both cases, the evaluation of the analytic CI gradient is significantly faster than the underlying SCF and CI calculations.[65,70] These computational advances greatly facilitate semiempirical CI studies of electronically excited states of large molecules, for MNDO-type as well as OMx methods.[70,71]

Linear scaling and hybrid methods

In practice, conventional semiempirical SCF-MO calculations are easily done on current hardware for molecules containing up to about 1000 non-hydrogen atoms (and even more with the use of GPUs). For much larger molecules, it is advisable to employ alternative approaches, most notably linear scaling methods or hybrid quantum mechanics/molecular mechanics (QM/MM) methods.

In the former, one attempts to achieve a linear scaling of the computational effort with system size, by exploiting the local character of most relevant interactions and the sparsity of the associate matrices. In semiempirical quantum chemistry, the primary objective of linear scaling methods is to avoid the bottlenecks related to diagonalization, i.e., to avoid the $O(N^3)$ steps. Three different approaches are available for this purpose. In the localized MO (LMO) approach,[72] Jacobi rotations are applied to annihilate the interactions between pairs of occupied and virtual LMOs that are located within a certain cutoff radius, whereas all other interactions are considered to be negligible. The divide-and-conquer methods[73-75] are based on a partitioning of the density matrix: the overall

SCF-MO calculation is decomposed into a series of relatively inexpensive, standard calculations for a set of smaller, overlapping subsystems, and a global description of the full system is then obtained by combining the information from all subsystem density matrices. The conjugate gradient density matrix search (CG-DMS)[76-79] avoids diagonalizations by direct optimization of the density matrix, under the constraints that it must be normalized, idempotent, and commuting with the Fock matrix after SCF convergence.

All these linear scaling methods introduce some numerical approximations so that the results will show some deviations from the conventional results, which can be controlled by the choice of suitable cutoffs. Semiempirical applications have mainly targeted large biomolecules, with system sizes up to about 20,000 atoms.[78] The merits of such calculations are most obvious for large systems with long-range charge transfer or long-range charge fluctuations since such effects are best captured by quantum-chemical approaches that cover the complete system.

In many large systems, the electronic processes of interest are localized in a small region, for example chemical reactions in the active site of an enzyme or electronic excitation of a chromophore in a large biomolecule. In such cases, QM/MM methods[80-82] are attractive, because they describe the electronically active region at the QM level (as accurately as needed – using semiempirical, DFT, or ab initio methods), and the environment at the MM level (with a classical force field). The QM/MM methods thus offer a versatile approach that can be tailored to the specific systems studied, by a suitable choice of the QM/MM partitioning and of the applied QM and MM methods. QM/MM calculations are generally significantly faster than corresponding linear scaling QM calculations, in the case of semiempirical QM components often by several orders of magnitude.[79,83] In practice, semiempirical QM/MM methods can be useful for extensive potential energy explorations and molecular dynamics (MD) simulations in biomolecular systems, while linear scaling semiempirical QM single-point calculations may serve to check the validity of the QM/MM results (e.g., with regard to charge transfer and charge fluctuations). While both approaches are thus complementary, QM/MM studies have dominated the field over the past decade, by virtue of being more versatile and efficient as well as potentially more accurate (through the use of high-level QM components).[84]

APPLICATIONS

Over the past decades, there have been a myriad of computational studies using semiempirical methods. The readers are referred to previous reviews for surveys of such work.[42-49,81,82] In this section, we address recent comparative benchmarks, give a brief summary of ground-state studies, and review recent semiempirical work on electronically excited states.

Validation and benchmarks

When a novel semiempirical method is introduced, the corresponding publication normally contains a validation section that reports numerical results for a large set of test molecules and properties, together with a statistical evaluation of the errors. The performance of a given method can already be assessed on this basis. For a more detailed assessment, comprehensive comparisons between different methods are desirable.

One such study compared the performance of seven semiempirical methods (MNDO, AM1, PM3, OM1, OM2, OM3, and SCC-DFTB) for standard test sets that are in common use.[39] The overall accuracy of these methods was found to be in the same range, with rankings depending on the properties or compound classes considered and with an overall tendency AM1 < SCC-DFTB < OM2. For example, considering the standard *ab initio* G2 (G3) sets, the following mean absolute deviations were reported[39] for heats of formation (in kcal/mol): AM1 7.37 (6.27); SCC-DFTB 9.19 (4.50); OM2 3.36 (3.15); DFT-B3LYP 2.35 (7.12). We note in this context that the performance of DFT-B3LYP for a large test set of 622 neutral closed-shell organic molecules turned out to be similar to PDDG/PM3, with no significant difference in quality between the B3LYP-based results and those from PDDG/PM3 for heats of formation and isomerization energies.[85]

A more recent benchmark[41] addressed the performance of six semiempirical methods (AM1, PM6, OM1, OM2, OM3, SCC-DFTB) and two DFT methods (B3LYP, PBE) using the CHNO-subset of the comprehensive GMTKN24 database[86] for general main-group thermochemistry, kinetics, and non-covalent interactions. The OMx methods were found to outperform AM1, PM6, and SCC-DFTB by a significant margin, with a substantial gain in accuracy especially for OM2 and OM3. These latter two methods were even quite accurate in comparison with DFT, and remarkably robust with regard to the unusual bonding situations encountered in one batch of test molecules. The overall mean absolute deviations for the energies in the whole data data set were as follows (in kcal/mol): PBE 6.60; B3LYP 4.82; OM3 7.86; OM2 8.33; OM1 10.93; PM6 18.19; AM1 14.52; SCC-DFTB 13.87 (SCC-DFTB without triplets and quartets). Inclusion of dispersion corrections changed these values only slightly.[41]

Concerning electronically excited states, a recent benchmark covered seven standard semiempirical methods (MNDO, AM1, PM3, OM1, OM2, OM3, INDO/S).[40] Semiempirical CI calculations were carried out for a standard set of 28 medium-sized organic molecules (104 singlet and 63 triplet excited states), with reference data being taken from accurate high-level *ab initio* calculations.[87] All applied semiempirical methods tended to underestimate the vertical excitation energies, but the errors were much larger in the case of the MNDO-type methods. Overall, the mean absolute deviations relative to the theoretical best estimates were lowest for OM3, and only slightly higher for OM1 and OM2. INDO/S performed similar to OM2 for the singlet excited states, but deteriorated for triplet states. The following mean absolute deviations in the vertical excitation energies for singlets (triplets) were obtained (in eV): MNDO 1.35 (1.55); AM1 1.19 (1.30); PM3 1.41 (1.42); OM1 0.45 (0.49); OM2 0.50 (0.47); OM3 0.45 (0.45); INDO/S 0.51 (0.65). SCC-DFTB was not tested since it is known to yield unsatisfactory results for electronically excited states. For comparison, time-dependent (TD) DFT calculations with the standard BP86 and B3LYP functionals yielded the following mean absolute deviations: TD-BP86 0.52 (0.53); B3LYP 0.27 (0.45).[88]

Taken together, these recent benchmark studies indicate that OM2 and OM3 generally show the best overall performance among the currently popular semiempirical methods. They often even approach the accuracy of standard DFT methods, despite the fact that the computational effort is lower by typically three orders of magnitude.

Ground-state studies

For studies on small and medium-size molecules, first-principles calculations (mostly DFT) are generally the favoured choice. Semiempirical methods are considered when a project involves very large molecules or very many calculations (for example, in MD simulations) or when it is desirable to get an initial overview followed by higher-level computations (for example, in the exploration of complicated potential energy surfaces). In a previous review,[49] typical target areas for semiempirical studies were identified that include large biomolecules (e.g., reaction mechanisms), medicinal chemistry and drug design (quantitative structure-property and structure-activity relationships), nanoparticles (e.g., fullerenes and nanotubes, molecular electronics), solid-state and surface chemistry, as well as direct reaction dynamics (possibly using special-purpose parameterizations).

Over the past decade, there have been many semiempirical QM/MM studies on biomolecules (see Refs 81 and 82 for detailed overviews and references). Again, there is a trend towards first-principles QM components in such investigations, but semiempirical QM methods are often the only practical option, for example in QM/MM MD simulations and free energy calculations.[38,82,89,90]

Finally, we mention the recent development of a theoretical procedure for computing electron impact mass spectra (EI-MS) of medium-size organic molecules, based on a combination of fast quantum chemical methods, molecular dynamics, and stochastic preparation of hot primary ions.[91] In this procedure, QM MD simulations in the ps range (up to 1 ns) of the initially generated ions are carried out to identify the fragmentation mechanisms and products. Many (often thousands) of MD runs are necessary for proper statistics (with millions of energy and gradient evaluations). DFT methods were tested in proof-of-principle calculations, but were too costly in practice. Among the semiempirical methods, OM2 was the preferred choice. It was concluded that routine calculations of EI-MS are possible with OM2 for many organic molecules, with an accuracy that should be sufficient for practical purposes.[91]

Excited-state studies

For decades, INDO/S has been the favoured semiempirical method for excited-state calculations.[18-20] It has been applied to compute the spectra of many large systems including the bacteriochlorophyll *b* dimer of *Rhodopseudomonas viridis* (QM/MM study with 325 QM atoms)[92] and aggregates of bacteriochlorophylls that occur in light-harvesting complexes of photosynthetic bacteria (QM studies with up to 704 atoms for a model of the hexadecamer).[93] By its design, INDO/S describes vertical excitation processes reliably, but it is not recommended for exploring excited-state potential energy surfaces (PES).[20]

MNDO-type methods can be used for PES exploration, but they underestimate excited-state energies severely because of the inherently symmetric splitting of bonding and antibonding orbitals (see above), and they are thus not useful for excited-state calculations. The correlated MNDOC variant[94] gave some improvement of the computed excited-state energies,[95] but was still too inaccurate to be recommended.

In the OMx methods, the explicit inclusion of Pauli exchange repulsion in the Fock matrix leads to an asymmetric splitting of the orbitals, with the antibonding combination being raised more than the bonding one is lowered, which overcomes the inherent deficiency of MNDO-type methods to underestimate excitation energies. As confirmed by extensive benchmarks (see above), the OMx methods describe both ground and excited states well, and it is possible to employ them for PES explorations.

Given these encouraging features and the availability of an efficient MRCI code with analytic derivatives,[65,70,71] several OM2/MRCI studies have recently been performed on the excited-state dynamics of medium-size organic molecules using trajectory surface hopping (TSH).[96,97] These TSH simulations have provided insight into the photostability of DNA bases in different environments (gas phase, aqueous solution, single-stranded and double-stranded DNA oligomers),[98-102,107,108] the mechanism of photoswitches and photoinduced molecular rotors,[104,110] the complete photochemical cycle of a GFP chromophore with ultrafast excited-state proton transfer,[105] the chiral pathways and mode-specific tuning of photoisomerization in azobenzenes,[103,109] and the competition between concerted and stepwise mechanisms in the ultrafast photoinduced Wolff rearrangement of 2-diaza-1-naphthoquinone.[112] The results of these OM2/MRCI studies are generally consistent with the available experimental data and high-level static calculations, but the dynamics often detects pathways and preferences between pathways that are not obvious from static calculations. In view of the cost of accurate *ab initio* treatments and the limitations of TD-DFT,[113,114] the OM2/MRCI approach appears to be a promising tool for investigating the excited states and the photochemistry of large molecules.

Conclusion

Among the available semiempirical approaches, the MNDO-type methods are well established and continue to be improved by general-purpose and special-purpose parameterizations. The OMx methods are based on a better theoretical model that includes explicit orthogonalization corrections to account for Pauli exchange repulsion. They show superior performance in comprehensive benchmarks, with significant advances in the treatment of conformational properties, non-covalent interactions, and electronically excited states.

In the hierarchy of computational chemistry, semiempirical methods are the simplest electronic structure theory. They are less robust and generally less accurate than DFT treatments, but typically also about 1,000 times faster than DFT. On the other hand, they are several orders of magnitude slower than MM treatments, but unlike classical force fields, they are capable of treating electronic events such as chemical reactions and electronic excitations that cannot be described at the MM level. Filling the gap between DFT and classical force fields, semiempirical methods enable realistic electronic structure calculations, with useful accuracy, on large complex systems in all branches of chemistry. Particularly when applied pragmatically in a multi-method strategy, they are useful tools in theoretical studies of large molecules.

References

1. Hückel E. Quantum contributions to the benzene problem. *Z Phys* 1931, 70:204–286.
2. Hoffmann R. An Extended Hückel Theory. I. Hydrocarbons. *J Chem Phys* 1963, 39:1397–1412.
3. Pariser R, Parr RG. A Semi-Empirical Theory of the Electronic Spectra and Electronic Structure of Complex Unsaturated Molecules. *J Chem Phys* 1953, 21:466–471.
4. Pople JA. Electron interaction in unsaturated hydrocarbons. *Trans Faraday* Soc 1953, 49:1375–1385.
5. Pople JA, Santry DP, Segal GA. Approximate Self-Consistent Molecular Orbital Theory. I. Invariant Procedures. *J Chem Phys* 1965, 43:S129–S135.
6. Pople JA, Segal GA. Approximate Self-Consistent Molecular Orbital Theory. III. CNDO Results for AB_2 and AB_3 Systems. *J Chem Phys* 1966, 44:3289–3296.
7. Pople JA, Beveridge DL, PA Dobosh PA. Approximate Self-Consistent Molecular-Orbital Theory. V. Intermediate Neglect of Differential Overlap. *J Chem Phys* 1967, 47:2026–2033.
8. Bingham RC, Dewar MJS, Lo DH. Ground States of Molecules. XXV. MINDO/3. Improved Version of the MINDO Semiempirical SCF-MO Method. *J Am Chem* Soc 97 1975, 97:1285–1293.
9. Dewar MJS, Thiel W. Ground States of Molecules. 38. The MNDO Method. Approximations and Parameters. *J Am Chem Soc* 1977, 99:4899–4907.
10. Dewar MJS, Thiel W. Ground States of Molecules. 39. MNDO Results for Molecules Containing Hydrogen, Carbon, Nitrogen and Oxygen. *J Am Chem Soc* 1977, 99:4907–4917.
11. Dewar MJS, E Zoebisch, EF Healy, Stewart JJP. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: a New General Purpose Quantum Mechanical Molecular Model. *J Am Chem Soc* 1985, 107:3902– 3909.
12. Stewart JJP. Optimization of Parameters for Semiempirical Methods I. Method. *J Comput Chem* 1989, 10:209–220.
13. Stewart JJP. Optimization of Parameters for Semiempirical Methods II. Applications. *J Comput Chem* 1989, 10:221–264.
14. Nanda DN, Jug K. SINDO1. A Semiempirical SCF MO Method for Molecular Binding Energy and Geometry. I. Approximations and Parametrization . *Theor Chim Acta* 1980, 57:95–106.
15. Jug K, Iffert R, Schulz J. Development and Parametrization of SINDO1 for Second-row Elements . *Int J Quantum Chem* 1987, 32:265–277.
16. Ahlswede B, Jug K. Consistent Modifications of SINDO1: I. Approximations and Parameters. *J Comput Chem* 1999, 20:563–571.
17. Ahlswede B, Jug K. Consistent Modifications of SINDO1: II. Applications to First- and Second-Row Elements. *J Comput Chem* 1999, 20:572–578.
18. Ridley J, Zerner MC. An Intermediate Neglect of Differential Overlap Technique for Spectroscopy: Pyrrole and the Azines. *Theor Chim Acta* 1973, 32:111–134.
19. Bacon AD, Zerner MC. An Intermediate Neglect of Differential Overlap Theory for Transition Metal Complexes: Fe, Co and Cu Chlorides. *Theor Chim Acta* 1979, 53:21–54.
20. Voityuk AA, Intermediate neglect of differential overlap for spectroscopy. *WIREs Comp Mol Sci* 2013, doi:10.1002/wcms.1141.
21. Dewar MJS, Thiel W. A Semiempirical Model for the Two-Center Repulsion Integrals in the NDDO Approximation. *Theor Chim Acta* 1977, 46:89–104.
22. Thiel W. Voityuk AA. Extension of the MNDO formalism to *d* orbitals: Integral approximations and preliminary numerical results. *Theor Chim Acta* 1992, 81:391–404; 1996, 93:315.

23. Thiel W, Voityuk AA. Extension of MNDO to d Orbitals: Parameters and Results for the Second-Row Elements and for the Zinc Group. *J Phys Chem* 1996, 100:616–626.
24. Stewart JJP. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J Mol Model* 2007, 13:1173–1213.
25. Stewart JJP. Optimization of parameters for semiempirical methods VI: More modifications to the NDDO approximations and re-optimization of parameters. *J Mol Model* 2013, 19:1–32.
26. Repasky MP, Chandrasekhar J, Jorgensen WL. PDDG/PM3 and PDDG/MNDO: Improved Semiempirical Methods. *J Comput Chem* 2002, 23:1601–1622.
27. Tubert-Brohman I, Guimaraes CRW, Repasky MP, Jorgensen WL. Extension of the PDDG/PM3 and PDDG/MNDO Semiempirical Molecular Orbital Methods to the Halogens. *J Comput Chem* 2004, 25:138–150.
28. Korth M, Pitonak M, Rezac J, Hobza P. A Transferable H-Bonding Correction for Semiempirical Quantum-Chemical Methods. *J Chem Theory Comput* 2010, 6:344-352.
29. Korth M. Empirical hydrogen-bond potential functions – an old hat reconditioned. *ChemPhysChem* 2011, 12:3131-3142, and references cited therein.
30. Rocha GB, Freire RO, Simas AM, Stewart JJP. RM1: A reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J Comput Chem* 2006, 27:1101–1111.
31. Kolb M, Thiel W. Beyond the MNDO Model: Methodical Considerations and Numerical Results. *J Comput Chem* 1993, 14:775–789.
32. Weber W. Ein neues semiempirisches NDDO-Verfahren mit Orthogonalisierungskorrekturen: Entwicklung des Modells, Parametrisierung und Anwendungen. PhD Thesis, Universität Zürich, 1996.
33. Weber W, Thiel W. Orthogonalization Corrections for Semiempirical Methods. *Theor Chem Acc* 2000, 103:495–506.
34. Scholten M. Semiempirische Verfahren mit Orthogonalisierungskorrekturen: Die OM3 Methode. PhD Thesis, Universität Düsseldorf, 2003.
35. Seifert G, Eschrig H, Bieger W. An approximation variant of LCAO-Xα methods. *Z Phys Chem* 1986, 267:529–539.
36. Seifert G, Joswig J-O. Density functional tight-binding – an approximate DFT method. *WIREs Comp Mol Sci* 2012, 2:456–465.
37. Elstner M, Porezag D, Jungnickel G, Elsner J, Haugk M, Frauenheim T, Suhai S, Seifert G. Self-consistent-charge density functional method for simulations of complex materials properties. *Phys Rev B* 1998, 58:7260–7268.
38. Elstner M, Gaus M, Cui Q. Density functional tight binding: Application to organic and biological molecules. *WIREs Comp Mol Sci* 2013, submitted.
39. Otte N, Scholten M, Thiel W. Looking at Self-Consistent-Charge Density Functional Tight Binding from a Semiempirical Perspective. *J Phys Chem A* 2007, 111:5751–5755.
40. Silva-Junior MR, Thiel W. Benchmark of Electronically Excited States for Semiempirical Methods: MNDO, AM1, PM3, OM1, OM2, OM3, INDO/S, and INDO/S2. *J Chem Theory Comput* 2010, 6:1546–1564.
41. Korth M, Thiel W. Benchmarking Semiempirical Methods for Thermochemistry, Kinetics, and Noncovalent Interactions: OMx Methods are almost as Accurate and Robust as DFT-GGA Methods for Organic Molecules. *J Chem Theory Comput* 2011, 7:2929–2936.
42. Thiel W. Semiempirical Methods: Current Status and Perspectives. *Tetrahedron* 1988, 44:7393–7408.

43. Stewart JJP. Semiempirical Molecular Orbital Methods. In Lipkowitz KB, Boyd DB, eds. Reviews in Computational Chemistry, Vol 1. VCH Publishers, New York, 1990, pp 45–81.
44. Stewart JJP. MOPAC: A semiempirical molecular orbital program. *J Comp-Aided Mol Design* 1990, 4:1–103.
45. Thiel W. Perspectives on Semiempirical Molecular Orbital Theory. *Adv Chem Phys* 1996, 93:703–757.
46. Clark T. Quo vadis semiempirical MO theory. *J. Mol. Struct. (THEOCHEM)* 2000, 530:1-10.
47. Thiel W. Semiempirical methods. In Grotendorst J, ed. Modern methods and algorithms of quantum chemistry, NIC Series, Vol 3. John von Neumann Institute for Computing, Jülich, Germany, 2000, pp 261–283.
48. Bredow T, Jug K. Theory and range of modern semiempirical molecular orbital methods. *Theor Chem Acc* 2005, 113:1–14.
49. Thiel, W. Semiempirical quantum-chemical methods in computational chemistry. In Dykstra CE, Kim KS, Frenking G, Scuseria, GE, eds. Theory and Applications of Computational Chemistry: The First 40 Years. Elsevier, Amsterdam, 2005, pp 559-580.
50. Bernal-Uruchurtu MI, Martins-Costa MTC, Millot C, Ruiz-Lopez MF. Improving the description of hydrogen bonds at the semiempirical level: Water-water interactions as test case. *J Comput Chem* 2000, 21:572–581.
51. Tuttle T, Thiel W. OMx-D: Semiempirical methods with orthogonalization and dispersion corrections. Implementation and biochemical application. *Phys Chem Chem Phys* 2008, 10:2159–2166.
52. Grimme S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J Comput Chem* 2006, 27:1787-1799.
53. Grimme S, Antony J, Ehrlich S, Krieg H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J Chem Phys* 2010, 132:154104/1-19.
54. Gonzalez-Lafont A, Truong TN, Truhlar DG. Direct Dynamics Calculations with Neglect of Diatomic Differential Overlap Molecular Orbital Theory with Specific Reaction Parameters. *J Phys Chem* 1991, 95:4618–4627.
55. Rossi I, Truhlar DG. Parameterization of NDDO wave functions using genetic algorithms. An evolutionary approach to parameterizing potential energy surfaces and direct dynamics calculations for organic reactions. *Chem Phys Lett* 1995, 233:231–236.
56. Doron D, Major DT, Kohen A, Thiel W, Wu X. Hybrid Quantum and Classical Simulations of the Dihydrofolate Reductase Catalyzed Hydride Transfer Reaction on an Accurate Semi-Empirical Potential Energy Surface*. J. Chem. Theory Comput.* 2011, 7: 3420-3437.
57. Wu X, Thiel W, Pezeshki S, Lin H. Specific Reaction Path Hamiltonian for Proton Transfer in Water: Re-parameterized Semi-Empirical Models. *J. Chem. Theory Comput.* 2013, submitted.
58. Patchkovskii S, Thiel W. NMR Chemical Shifts in MNDO Approximation: Parameters and Results for H, C, N, and O. *J Comput Chem* 1999, 20:1220–1245.
59. Patchkovskii S, Thiel W. Nucleus-Independent Chemical Shifts from Semiempirical Calculations. *J Mol Model* 2000, 6:67–75.
60. Stewart JJP, Csaszar P, Pulay P. Fast Semiempirical Calculations. *J Comput Chem* 1982, 3:227–228.
61. Bakowies D, Thiel W. MNDO Study of Large Carbon Clusters. *J Am Chem Soc* 1991, 113:3704–3714.

62. Wu X, Koslowski A, Thiel W. Semiempirical Quantum Chemical Calculations Accelerated on a Hybrid Multicore CPU-GPU Computing Platform. *J Chem Theory Comput* 2012, 8:2272–2281.
63. Maia JDC, Carvalho GAU, Mangueira CP, Santana SR, Cabral LAF, Rocha GB. GPU Linear Algebra Libraries and GPGPU Programming for Accelerating MOPAC Semiempirical Quantum Chemistry Calculations. *J Chem Theory Comput* 2012, 8:3072–3081.
64. Patchkovskii S, Thiel W. Analytical first derivatives of the energy in the MNDO half-electron open-shell treatment. *Theor Chim Acta* 1996, 93:87–99.
65. Patchkovskii S, Thiel W. Analytical first derivatives of the energy for small CI expansions. *Theor Chem Acc* 1997, 98:1–4.
66. Patchkovskii S, Thiel W. Analytical Second Derivatives of the Energy in MNDO Methods. *J Comput Chem* 1996, 17:1318–1327.
67. Dewar MJS, Liotard DA. An efficient procedure for calculating the molecular gradient using SCF-CI semiempirical wave functions with a limited number of configurations. *J Mol Struct (THEOCHEM)* 1990, 206:123–133.
68. Handy NC, Schaefer HF. On the evaluation of analytic energy derivatives for correlated wave functions. *J Chem Phys* 1984, 81:5031–5033.
69. Strodel P, Tavan P. A revised MRCI-algorithm. I. Efficient combination of spin adaptation with individual configuration selection coupled to an effective valence-shell Hamiltonian. *J Chem Phys* 2002, 117:4667–4676.
70. Koslowski A, Beck ME, Thiel W. Implementation of a General Multireference Configuration Interaction Procedure with Analytic Gradients in a Semiempirical Context Using the Graphical Unitary Group Approach. *J Comput Chem* 2003, 24:714–726.
71. Patchkovskii S, Koslowski A, Thiel W. Generic implementation of semi-analytical CI gradients for NDDO-type methods. *Theor Chem Acc* 2005, 114:84–89.
72. Stewart JJP. Application of Localized Molecular Orbitals to the Solution of Semiempirical Self-Consistent Field Equations. *Int J Quantum Chem* 1996, 58:133–146.
73. Yang W, Lee T-S. A density-matrix divide-and-conquer approach for electronic structure calculations of large molecules. *J Chem Phys* 1995, 103:5674–5678.
74. Lee TS, York DM, Yang W. Linear-scaling semiempirical quantum calculations for macromolecules. *J Chem Phys* 1996, 105:2744–2750.
75. Dixon SL, Merz KM. Semiempirical molecular orbital calculations with linear system size scaling. *J Chem Phys* 1996, 104:6643–6649.
76. Li X-P, Nunes RW, Vanderbilt D. Density-matrix electronic-structure method with linear system-size scaling. *Phys Rev B* 1993, 47:10891–10894.
77. Millam JM, Scuseria GE. Linear scaling conjugate gradient density matrix search as an alternative to diagonalization for first principles electronic structure calculations. *J Chem Phys* 1997, 106:5569–5577.
78. Daniels AD, Millam JM, Scuseria GE. Semiempirical methods with conjugate gradient density matrix search to replace diagonalization for molecular systems containing thousands of atoms. *J Chem Phys* 1997, 107:425–431.
79. Kevorkiants R. Linear Scaling Conjugate Gradient Density Matrix Search: Implementation, Validation and Application with Semiempirical MO Methods. PhD thesis, Universität Düsseldorf, 2003.
80. Warshel A, Levitt M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J Mol Biol* 1976, 103:227–249.

81. Senn HM, Thiel W. QM/MM Methods for Biological Systems. *Top Curr Chem* 2007, 268:173–290.
82. Senn HM, Thiel W. QM/MM Methods for Biomolecular Systems. *Angew. Chem. Int. Ed.* 2009, 48:1198–1229.
83. Titmuss SJ, Cummins PL, Bliznyuk AA, Rendell AP, Gready JE. Comparison of linear scaling semiempirical methods and combined quantum mechanical/molecular mechanical methods applied to enzyme reactions. *Chem. Phys. Lett.* 2000, 320:169–176.
84. Claeyssens F, Harvey JN, Manby FR, Mata RA, Mulholland AJ, Ranaghan KE, Schütz M, Thiel S, Thiel W, Werner H-J. High-Accuracy Computation of Reaction Barriers in Enzymes. *Angew Chem Int Ed* 2006, 45:6856–6859.
85. Tirado-Rives J, Jorgensen, WL. Performance of B3LYP Density Functional Methods for a Large Set of Organic Molecules. *J Chem Theory Comput* 2008, 4:297–306.
86. Goerigk L, Grimme S. A General Database for Main Group Thermochemistry, Kinetics, and Noncovalent Interactions - Assessment of Common and Reparameterized (meta-)GGA Density Functionals. *J. Chem. Theory Comput*. 2010, 6:107.
87. Schreiber M, Silva-Junior MR, Sauer SPA, Thiel W. Benchmarks for electronically excited states: CASPT2, CC2, CCSD, and CC3. *J Chem Phys* 2008, 128:131410/1–25.
88. Silva-Junior MR, Schreiber M, Sauer SPA, Thiel W. Benchmarks for electronically excited states: Time-dependent density functional theory and density functional theory based multireference configuration interaction. *J Chem Phys* 2008, 129:104103/1–14.
89. Senn HM, Kästner J, Breidung J, Thiel W. Finite-temperature effects in enzymatic reactions – Insights from QM/MM free-energy eimulations. *Can J Chem* 2009, 87:1322–1337.
90. Polyak I, Reetz MT, Thiel W. QM/MM Study on the Mechanism of the Enzymatic Baeyer-Villiger Reaction. *J Am Chem Soc* 2012, 134:2732–2741.
91. Grimme S. Towards First Principles Calculation of Electron Impact Mass Spectrometry of Molecules. *Angew Chem Int Ed* 2013, accepted, doi:10.1002/anie.201300158.
92. Thompson MA, Schenter GK. Excited States of the Bacteriochlorophyll *b* Dimer of *Rhodopseudomonas viridis*: A QM/MM Study of the Photosynthetic Reaction Center That Includes MM Polarization. *J Phys Chem* 1995, 99:6374–6386.
93. Cory MG, Zerner MC, Hu X, Schulten K. Electronic Excitations in Aggregates of Bacteriochlorophylls. *J Phys Chem B* 1998, 102:7640–7650.
94. Thiel W. The MNDOC Method, a Correlated Version of the MNDO Model. *J Am Chem Soc* 1981, 103:1413–1420.
95. Schweig A, Thiel W. MNDOC Study of Excited States. *J Am Chem Soc* 1981, 103:1425–1431.
96. Tully JC. Molecular dynamics with electronic transitions. *J. Chem. Phys.* 1990, 93:1061–1071.
97. Fabiano E, Keal TW, Thiel W. Implementation of surface hopping molecular dynamics using semiempirical methods. *Chem Phys* 2008, 349:334–347.
98. Fabiano E, Thiel W. Nonradiative Deexcitation Dynamics of 9H-Adenine: An OM2 Surface Hopping Study. *J Phys Chem A* 2008, 112:6859–6863.
99. Lan Z, Fabiano E, Thiel W. Photoinduced Nonadiabatic Dynamics of Pyrimidine Nucleobases: On-the-fly Surface-Hopping Study with Semiempirical Methods. *J Phys Chem B* 2009, 113:3548–3555.
100. Lan Z, Fabiano E, Thiel W. Photoinduced Nonadiabatic Dynamics of 9*H*-Guanine. *ChemPhysChem* 2009, 10:1225–1229.
101. Lan Z, Lu Y, Fabiano E, Thiel W. QM/MM Nonadiabatic Decay Dynamics of 9*H*-Adenine in Aqueous Solution. *ChemPhysChem* 2011, 12:1989–1998.

102. Lu Y, Lan Z, Thiel W. Hydrogen Bonding Regulates the Monomeric Nonradiative Decay of Adenine in DNA Strands. *Angew Chem Int Ed* 2011, 50:6864–6867.
103. Weingart O, Lan Z, Koslowski A, Thiel W. Chiral Pathways and Periodic Decay in *cis*-Azobenzene Photodynamics. *J Phys Chem Lett* 2011, 2:1506–1509.
104. Kazaryan A, Lan Z, Schäfer LV, Thiel W, Filatov M. Surface Hopping Excited-State Dynamics Study of the Photoisomerization of a Light-Driven Fluorene Molecular Rotary Motor. *J Chem Theory Comput* 2011, 7:2189–2199.
105. Cui G, Lan Z, Thiel W. Intramolecular Hydrogen Bonding Plays a Crucial Role in the Photophysics and Photochemistry of the GFP Chromophore. *J Am Chem Soc* 2012, 134:1662–1672.
106. Lan Z, Lu Y, Weingart O, Thiel W. Nonadiabatic Decay Dynamics of a Benzylidene Malononitrile. *J Phys Chem A* 2012, 116:1510–1518.
107. Lu Y, Lan Z, Thiel W. Monomeric Adenine Decay Dynamics Influenced by the DNA Environment. *J Comput Chem* 2012, 33:1225–1235.
108. Heggen B, Lan Z, Thiel W. Nonadiabatic decay dynamics of 9*H*-guanine in aqueous solution. *Phys Chem Chem Phys* 2012, 14: 8137–8146.
109. Gamez JA, Weingart O, Koslowski A, Thiel W. Cooperating Dinitrogen and Phenyl Rotations in *trans*-Azobenzene Photoisomerization. *J Chem Theory Comput* 2012, 8:2352–2358.
110. Schönborn JB, Koslowski A, Thiel W, Hartke B. Photochemical dynamics of *E*-iPr-furylfulgide. *Phys Chem Chem Phys* 2012, 14:12193–12201.
111. Cui G, Thiel W. Nonadiabatic dynamics of a truncated indigo model. *Phys Chem Chem Phys* 2012, 14:12378–12384.
112. Cui G, Thiel W. Photoinduced Ultrafast Wolff Rearrangement: A Non-Adiabatic Dynamics Perspective. *Angew Chem Int Ed* 2013, 52 :433–436.
113. Dreuw A, Head-Gordon M. Failure of Time-Dependent Density Functional Theory for Long-Range Charge-Transfer Excited States: The Zincbacteriochlorin-Bacterlochlorin and Bacteriochlorophyll-Spheroidene Complexes. *J Am Chem Soc* 2004, 126:4007–4016.
114. Dreuw A, Head-Gordon M. Single-Reference ab Initio Methods for the Calculation of Excited States of Large Molecules. *Chem Rev* 2005, 105:4009–4037.