

Accurate Theoretical Chemistry with Coupled Pair Models

FRANK NEESE,^{*,†,‡} ANDREAS HANSEN,[†] FRANK WENNMOHS,[†]
AND STEFAN GRIMME[§]

[†]Lehrstuhl für Theoretische Chemie, Institut für Physikalische and Theoretische Chemie, Universität Bonn, Wegelerstr. 12, D-53115 Bonn, Germany, [‡]Max-Planck Institut für Bioanorganische Chemie, Stiftstr. 34-36, D-45470 Mülheim an der Ruhr, Germany, [§]Theoretische Organische Chemie, Organisch-Chemisches Institut der Universität Münster, Corrensstrasse 40, D-48149 Münster, Germany

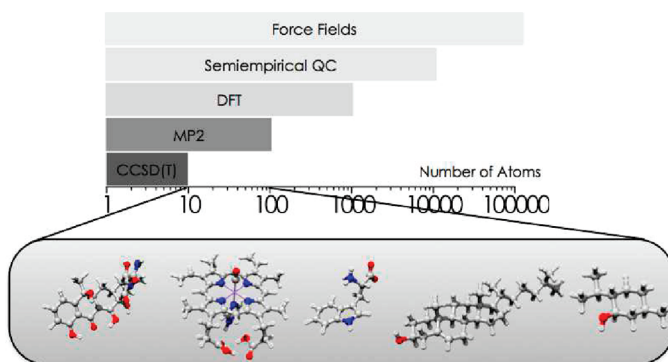
RECEIVED ON OCTOBER 20, 2008

CON SPECTUS

Quantum chemistry has found its way into the everyday work of many experimental chemists. Calculations can predict the outcome of chemical reactions, afford insight into reaction mechanisms, and be used to interpret structure and bonding in molecules. Thus, contemporary theory offers tremendous opportunities in experimental chemical research. However, even with present-day computers and algorithms, we cannot solve the many particle Schrödinger equation exactly; inevitably some error is introduced in approximating the solutions of this equation. Thus, the accuracy of quantum chemical calculations is of critical importance.

The affordable accuracy depends on molecular size and particularly on the total number of atoms: for orientation, ethanol has 9 atoms, aspirin 21 atoms, morphine 40 atoms, sildenafil 63 atoms, paclitaxel 113 atoms, insulin nearly 800 atoms, and quaternary hemoglobin almost 12,000 atoms. Currently, molecules with up to ~10 atoms can be very accurately studied by coupled cluster (CC) theory, ~100 atoms with second-order Møller–Plesset perturbation theory (MP2), ~1000 atoms with density functional theory (DFT), and beyond that number with semiempirical quantum chemistry and force-field methods. The overwhelming majority of present-day calculations in the 100-atom range use DFT. Although these methods have been very successful in quantum chemistry, they do not offer a well-defined hierarchy of calculations that allows one to systematically converge to the correct answer. Recently a number of rather spectacular failures of DFT methods have been found—even for seemingly simple systems such as hydrocarbons, fueling renewed interest in wave function-based methods that incorporate the relevant physics of electron correlation in a more systematic way. Thus, it would be highly desirable to fill the gap between 10 and 100 atoms with highly correlated ab initio methods. We have found that one of the earliest (and now almost forgotten) of this class of methods, the coupled-electron pair approximation (CEPA), performs exceedingly well in chemical applications.

In this Account, we examine the performance of CEPA in chemical applications. One attractive feature of CEPA, in addition to its surprising accuracy that surpasses that of DFT and MP2 theory, is a simplicity that allows for straightforward and very efficient approximations and extensions to be developed; these are much more difficult or even impossible with the more rigorous CC theory. Thus, approximate CEPA methods can be implemented efficiently enough to allow for calculations on molecules of 50–100 atoms, perhaps the most common range in contemporary chemical research.



1. Introduction

The steep rise of the impact of quantum chemistry has certainly been one of the most important developments in contemporary chemical research. Nowadays, it is commonplace to publish experimental data alongside theoretical calculations, and in many cases, referees even demand the back up of conclusions from experimental investigations with quantum chemical calculations. One central issue that arises in this context is the question: How accurate and reliable are the theoretical results? In order to answer this question, it is obviously necessary to first define what we mean by the term theoretical calculation. In principle, the *ab initio* methods of quantum chemistry are able to solve the time-independent *N*-particle Schrödinger equation to almost arbitrary precision. To this end, one usually divides the problem into the solution of the mean-field equation (the Hartree–Fock model) and the remainder (the correlation energy). Second, one needs to introduce a finite one-particle basis set in order to solve the quantum chemical equations. If both, the one-particle basis set and the *N*-particle correlation space, approach mathematical completeness, then exact solutions to the *N*-particle Born–Oppenheimer (BO) Hamiltonian are approached as well. At this point, one would obtain a model chemistry that would describe the majority of chemical problems with high accuracy. Spectacular progress toward nearly exact BO eigenenergies have been made on the basis of the powerful hierarchy of coupled-cluster (CC) theory, which is able to deliver sub kJ/mol accuracy.^{1,2} However, the associated theoretical and computational apparatus is very heavy and so far remains limited to very small molecular systems.

However, the popularity of theory in chemical investigations on real-life molecules (typically in the range 10–200 atoms) is due to density functional theory (DFT) that strongly dominates applied theoretical chemistry.^{3,4} Unlike the case for CC theory, the realizations of DFT are not meant to approach exact *N*-particle solutions but rather represent pragmatic models of molecular electronic structure. The price–performance ratio of DFT is certainly very good such that many questions of chemical interest can be successfully answered on the basis of such calculations. However, uncritical trust in the results of DFT investigations is as unjustified as the previously dominant reluctance of the experimental community to accept theoretical results at all.

Despite the many successful DFT studies, in recent years some unexpected failures have become apparent. For example, DFT often shows large errors (exceeding by far the accepted error limits of 2–3 kcal/mol) in application to the

seemingly most straightforward systems: closed shell hydrocarbons.^{3,4} Rather than trying to repair these errors through more or less physically motivated corrections and additions to the commonly used functionals, there is a renewed interest in wave function theory. While already the simplest post-HF approach (second order Møller–Plesset perturbation theory, MP2), has its merits for certain classes of systems, it is neither accurate nor robust enough for more challenging applications such as those met in open-shell transition metal chemistry. Considerable improvements based on spin-component scaling (SCS-MP2) and higher order perturbation theory (SCS-MP3) have been discussed previously by one of the authors.^{16,33} Because of extensive computational developments such as, e.g., the resolution-of-the-identity (RI),⁶ MP2 energies can nowadays be obtained at a computational cost that is not much larger than that of (hybrid) DFT.

In order to significantly proceed beyond the accuracy obtainable with DFT or MP2 methods, it is necessary to introduce higher-order correlation effects. Unfortunately, all of the known methods that achieve this goal are very much more expensive than DFT or MP2 itself. While MP2 shows a noniterative $O(N^5)$ scaling, higher order correlation methods of the CC type are at least of iterative $O(N^6)$ scaling and hence typically a factor of 100 more expensive. Linear scaling variants of CC methods have been extensively developed by Werner, Schütz, and co-workers⁷ based on concepts of Pulay⁸ and may change this situation if they find their way into the standard arsenal of computational chemists.

It is our hypothesis that methods that are designed to compete with DFT must fulfill the following criteria: (a) they need to be systematically more accurate than DFT and MP2, (b) they must be of a black-box nature, (c) they must provide an efficiency comparable to that of DFT or MP2 (i.e., the increase in computation time should not be more than a factor of 2–5; ideally, should be even less), and (d) they must be robust (i.e., applicable to a wide range of systems and difficult bonding situations).

In our recent work, we have asked the question whether it is really necessary to take on the considerable challenge to approximate the very intricate CC equations or whether it would not be possible to approximate somewhat simpler equations that can be implemented with outstanding efficiency. We have identified that the coupled electron pair approaches (CEPA), the earliest (now almost forgotten) methods of quantum chemistry that have delivered accurate results, do fulfill the above-mentioned criteria. They have been developed in the 1970s and 1980s by Meyer, Kutzelnigg, Ahlrichs, Hurley, and Taylor among others.^{9,10,35} The CEPA methods

are based on straightforward size-consistent modifications of the basic configuration interaction method with single- and double-excitations (CISD). We will show elsewhere in detail that the HF and CEPA equations can be solved with a computational effort that is not much higher than that of current hybrid DFT programs and can be applied to fairly large systems (100 atoms, 2000 basis functions). Following up on previous work on geometries and harmonic frequencies,¹¹ we show here for the first time that CEPA-type methods in conjunction with large and flexible basis sets deliver competitive accuracy in thermochemical applications. Hence, we believe that a multitude of successful chemical applications can be based on the CEPA approach.

2. Theory of the Coupled Pair Approach

The details of the various flavors of CEPA style methods in the generator state formalism have been discussed in detail in a recent paper that has also provided a detailed review of the historical context and the physical motivation for these methods.¹¹ Briefly, we start from the variational principle $E \leq \langle \Psi | H | \Psi \rangle / \langle \Psi | \Psi \rangle$ with a wave function of the CISD type (the C 's are the CI coefficients; Ψ_i^a and Ψ_{ij}^{ab} are singly- and double-excited configuration states functions; the labels i, j refer to occupied orbitals in the HF determinant and a, b to unoccupied ones; we will only treat closed shell systems here):

$$\Psi = \Psi_{HF} + \sum_{ia} C_a^i \Psi_i^a + \sum_{i \leq j} \sum_{ab} C_{ab}^{ij} \Psi_{ij}^{ab} \quad (1)$$

with the $2n$ -electron HF determinant $\Psi_{HF} = |\psi_1 \bar{\psi}_1 \dots \psi_n \bar{\psi}_n|$, where $\{\psi\}$ is the set of (possibly localized) HF orbitals. Variation with respect to the CI coefficients yields the CISD equations:

$$E = \langle \Psi_{HF} | H | \Psi \rangle = E_{HF} + E_C \quad (2)$$

$$\sigma_{ab}^{ij} = \langle \Psi_{ij}^{ab} | H - E_{HF} - E_C | \Psi \rangle = 0 \quad (3)$$

The CEPA method simulates the size-consistency restoring effects of disconnected quadruple excitations by replacing the correlation energy (E_C) by an excitation dependent shift Δ_{ab}^{ij} that differs for different flavors of CEPA. The individual definitions of the shift for closed shell systems have been discussed in detail previously.¹¹ We prefer the CEPA/1 variant in our work. Through this replacement, one restores size consistency but loses the variational property of the equations as well as the unitary invariance with respect to rotations of the occupied orbitals. We fix the latter ambiguity by always referring to localized occupied MOs. Ahlrichs has shown how to restore the stationary character of the wave function ampli-

tudes \mathbf{C} in the development of the coupled-pair functional (CPF) that provides results very similar to those of CEPA/1.¹² Our slight modification of it (NCPF/1) has been discussed previously¹¹ and will be used in this work as well.

3. Thermochemistry with CEPA Methods

In order to assess the potential of CEPA style methods in computational chemistry, we have applied it to several sets of benchmark molecules that were chosen to represent: (a) typical reactions, (b) difficult systems, (c) transition states, and (d) noncovalent interactions. We have been less concerned with atomization energies and follow the arguments brought forward in ref 13 that they are of limited relevance for the evaluation of quantum chemical methods. This is not to say that CEPA methods are particularly poor in this respect. Indeed, all CEPA methods that we have tested are superior to QCISD or CCSD. CEPA/2 is even approaching the same type of mean-average deviation as QCISD(T) (3.8 kcal/mol). All results can be found in Supporting Information.

3.1. Reaction Energies. The investigated test set (for details, see Supporting Information) includes 21 closed shell reactions built from the G2/97 test set of Curtiss et al.¹⁴ We have provided two sets of data that refer to the deviations from experiment and the deviations from CCSD(T) in an extended Leerzeichen quadruple- ζ basis set. It is known that once all basis set, relativistic, and vibrational effects are properly taken care of, CCSD(T) yields chemical accuracy for this test set.⁵ Hence, we argue that the CCSD(T) electronic energies provides the most convincing reference for evaluating the CEPA methods. Deviations from the CCSD(T) energetics then directly measure the quality of the electronic energies and leave less room for error cancellation than comparison of a (incomplete) calculation with the experimental data that also have limited accuracy. If one aims at producing the right answer for the right reason, all of the above-mentioned effects should be included if comparison with experiment is the ultimate goal.¹⁵

With this in mind, the performance of the CEPA/1 and NCPF/1 methods is outstanding (mean average errors of only ~ 1 kcal/mol and maximum errors of 2.5–3 kcal/mol). This is obviously much better than the more expensive QCISD and CCSD methods that show average absolute and maximum errors of ~ 3 and ~ 13 kcal/mol, respectively. The CEPA/1 and NCPF/1 methods also behave better than the popular B3LYP method, which in turn is inferior to the more recently developed SCS-MP2 wave function method. Compared with conventional MP2 calculations, the results of the SCS-MP2 method

TABLE 1. Statistical Data (in kcal mol⁻¹) for 21 Reaction Energies (Small Molecules) Calculated with Different Quantum Chemical Methods with the def2-QZVPP Basis Set^a

	deviation from CCSD(T) (exp.) reference		
	MAD	MD	MAX
CCSD(T)	(1.76)	(-0.82)	(6.73)
CCSD	3.41 (4.78)	-3.02 (-3.84)	14.13 (17.23)
QCISD	3.07 (4.44)	-2.71 (-3.53)	12.11 (16.27)
CEPA/1	1.08 (2.53)	-0.35 (-1.17)	2.99 (7.14)
NCPF/1	1.03 (2.42)	-0.26 (-1.08)	2.44 (6.84)
SCS-MP3	2.22 (3.01)	-1.21 (-2.03)	7.06 (11.91)
SCS-MP2	3.41 (2.78)	-0.34 (-1.17)	9.58 (15.36)
MP2	5.54 (4.99)	-2.76 (-3.58)	26.39 (33.12)
B2PLYP	2.68 (2.45)	0.55 (-0.28)	6.33 (7.93)
B3LYP	3.18 (2.91)	1.72 (0.90)	8.72 (7.26)

^a Individual results are collected in Supporting Information. Numbers in parenthesis refer to deviations from the experimental data and otherwise to deviations from CCSD(T). Shown are the mean deviation (MD), mean absolute deviation (MAD), and the maximum absolute deviation (MAX) from the experimental reference data, which have an accuracy of ± 1 kcal/mol (nonrelativistic atomization energies back-corrected with unscaled PBE/def2-TZVP vibrations and atomic heats of formation) that have been provided by one of the authors (S.G.). The test set includes 21 reactions built up from 39 closed shell molecules (see Supporting Information) out of the G2/97 test set of Curtiss et al.¹⁴ All calculations have been performed with a development version of the ORCA program¹⁷ using restricted orbitals; the SCF iterations were converged to 10^{-8} Eh in the total energy (ORCA keyword VeryTightSCF). The geometries were optimized at the B3LYP/TZVP level. For the single point energy calculations, we used the def2 version of the split-valence quadruple- ζ (QZV¹⁸) basis set supplemented with the appropriate polarization functions from the TurboMole library.¹⁹ For the RI-SCS-MP2 calculation, the corresponding def2 fit basis was used. SCS-MP3 calculations were performed without any integral approximation. The basis sets were truncated after $L = 4$ due to technical constraints. In all correlation calculations, only the valence electrons have been correlated. The resulting error is usually not larger than 0.5–1 kcal mol⁻¹.¹⁶ The mean value of the absolute CCSD(T) reference values is 99.32 kcal/mol, and that of the experimental reference values is 98.50 kcal/mol.

are much more accurate, which was also reported for other test sets.¹⁶ The accuracy of the SCS-MP3 method is seen to be intermediate between SCS-MP2 and the CEPA approaches, which is also true for its computational cost. Except for CEPA/1 and NCPF/1, the B2PLYP double hybrid functional performs better than the other methods under investigation. The B3LYP, B2PLYP, SCS-MP2, and SCS-MP3 methods have been parametrized using closely related molecules in the training set. The CEPA/1 and NCPF/1 methods have no adjustable parameters at all and obviously achieve high and systematic accuracy intrinsically.

3.2. Complicated and Challenging Reactions. Obviously, the molecules of the G2/97 test set represent generally well-behaved, small systems that are still not of outstandingly high relevance for actual chemical applications. In particular, they cover up some of the problems that especially the B3LYP method has for larger molecules.⁵ We have therefore applied the same methods to a more challenging set of larger molecules. Since accurate experimental data are scarce and CCSD(T) calculations with sufficiently saturated basis sets are impossible for this test set, we have decided to

use CCSD(T)/TZVP results as our reference data. The quality of this basis set was investigated previously and found to be adequate for this kind of molecules.⁵ The investigated test set (23 closed shell reactions) consists of 16 molecules from the isomerization reaction test set of Jørgensen,²¹ which was extensively discussed in a previous study⁵ and supplemented with 7 more complicated reactions. All reactions as well as the optimized geometries of the molecules involved in the more complicated reactions can be found in the Supporting Information.

It is pleasing to observe that the CEPA/1 and NCPF/1 methods remain beautifully accurate and represent mean-average errors of below 1 kcal/mol, again superior to QCISD and CCSD. For these more difficult systems, CEPA/1 and NCPF/1 also clearly outperform B2PLYP and SCS-MP2. The latter one yields more accurate results than CCSD. In coincidence with previous studies,^{5,13} SCS-MP2 and also B2PLYP are still clear improvements over B3LYP that shows a rather large mean absolute error of more than 7 kcal/mol. We believe that the latter value comes closer to the intrinsic accuracy of B3LYP than the frequently quoted 2–3 kcal/mol that was derived from the G2 test set. Once more, this demonstrates that the G2 test set based results are of limited relevance if the goal is to judge the quality of a given theoretical method for actual chemical applications.

We note that CEPA/1 and NCPF/1 also outperform QCISD and CCSD with respect to the number of outliers in the data set: both methods only show two outliers (for the reactions 2,2,3-dimethylbut-2-ene \rightarrow octamethylcyclobutane and 2 paraxylene \rightarrow [2,2]paracyclophane + 2 H₂), while QCISD and CCSD fail in 5 and 9 cases, respectively to reproduce the CCSD(T) reference values to better than 2 kcal/mol accuracy. The SCS-MP2 method also works rather well (5 outliers); only the reaction heptahexaene \rightarrow heptatriyne is not described properly (-9.4 kcal/mol deviation from the CCSD(T) value) and again shows clear improvements over conventional MP2 (13 outliers). Here, SCS-MP3 only offers at best modest improvements over SCS-MP2. The DFT based methods have problems for larger molecules, in particular if large geometry changes occur in the reaction. This is apparent for B3LYP (17 outliers) and to a lesser extent also for B2PLYP (11 outliers). In four and two cases, respectively, B3LYP and B2PLYP even predicted the wrong sign of the reaction energy, while this was not observed in any of the CEPA/1 or NCPF/1 calculations.

3.3. H-Bonded Dimers and van der Waals Interactions. In order to assess the quality of the CEPA-style methods for noncovalent interactions, we have studied the interaction energies of 7 neutral and 4 charged H-bonded

TABLE 2. Statistical Data (in kcal mol⁻¹) for the Reaction Energies of 23 Reactions Containing Larger Molecules Calculated with Different Quantum Chemical Methods and the def2-TZVP Basis Set^a

	deviation from the CCSD(T) value		
	MAD	MD	MAX
CCSD	1.81	0.33	6.60
QCISD	1.06	0.53	6.60
CEPA/1	0.70	0.47	4.93
NCPF/1	0.73	0.46	4.89
SCS-MP3	1.43	0.41	7.84
SCS-MP2	1.52	0.31	9.35
MP2	2.85	-0.36	9.61
B2PLYP	3.94	1.95	16.17
B3LYP	7.13	2.96	29.49

^a Shown are the mean deviation (MD), mean absolute deviation (MAD), and the maximum absolute deviation (MAX) from the corresponding CCSD(T) values for a test set containing 23 closed shell reactions, mostly isomerization reactions of organic molecules from a test set introduced by Jørgensen²¹ supplemented with 7 more difficult cases also including larger molecules provided by one of us (S.G.). All calculations have been performed with restricted orbitals, and the SCF iterations were converged to 10⁻⁸ Eh in the total energy (ORCA keyword VeryTightSCF). The geometries were optimized at the B3LYP/TZVP level (see Supporting Information). For the single point energy calculations, we used the def2 version of the split-valence triple- ζ (TZV²⁰) basis set supplemented with the appropriate polarization functions (TZVP) from the TurboMole library.¹⁹ For the RI-SCS-MP2 calculation, the corresponding def2 fit basis was used. In all correlation calculations, only the valence electrons have been correlated. The mean value of the absolute CCSD(T) reference values is 21.03 kcal/mol.

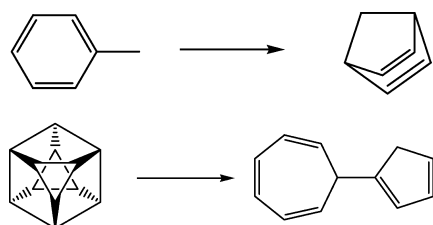


FIGURE 1. Two examples for isomerization reactions (toluene \rightarrow norbornadiene and octahedrane \rightarrow tropenylcyclopentadienyl). For both reactions, B3LYP yields large deviations (13.97 and -18.21 kcal/mol) from the CCSD(T) reference values (42.79 and 18.22 kcal/mol), while CEPA/1 (0.04 and -0.24 kcal/mol) and NCPF/1 (0.01 and -0.35 kcal/mol) are almost on top of the reference values.

dimers from a test set used by Boese, Martin, and Klopper.²² Once more, CEPA/1 and NCPF/1 almost quantitatively reproduce the computationally much more demanding W2 reference values. Only the CN⁻...H₂O dimer shows a slightly larger error (0.61 and 0.64 kcal/mol for CEPA/1 and NCPF/1, respectively). The average errors of the two coupled pair approaches are again superior to QCISD, CCSD, SCS-MP2, SCS-MP3, MP2, B2PLYP, and B3LYP as well as, and probably to some extent fortuitously so, CCSD(T). Hence, H-bonds are also described well by the coupled pair methods.

Furthermore, we have tested the performance of the coupled pair methods for noncovalent bonded molecules with dominant dispersion interaction (van der Waals complexes). Therefore, we used a subset containing 10 molecules out of

TABLE 3. Statistical data (in kcal mol⁻¹) for 7 Neutral and 4 Charged (Values in Parentheses) H-Bond Dimer Interaction Energies Calculated with Different Quantum Chemical Methods and the def2-QZVPP Basis Set^a

	deviation from the W2 reference value		
	MAD	MD	MAX
CCSD(T)	0.11 (0.50)	0.11 (0.50)	0.17 (1.07)
CCSD	0.17 (0.32)	-0.17 (-0.16)	0.29 (0.50)
QCISD	0.14 (0.33)	-0.14 (-0.15)	0.21 (0.48)
CEPA/1	0.06 (0.21)	-0.05 (0.13)	0.14 (0.61)
NCPF/1	0.06 (0.21)	-0.05 (0.14)	0.14 (0.64)
SCS-MP3	0.29 (0.45)	-0.29 (-0.17)	0.41 (0.54)
SCS-MP2	0.24 (0.39)	-0.22 (-0.09)	0.36 (0.60)
MP2	0.26 (0.94)	0.20 (0.74)	0.51 (1.52)
B2PLYP	0.28 (0.88)	0.14 (0.68)	0.39 (1.34)
B3LYP	0.22 (0.85)	0.03 (0.85)	0.47 (1.68)

^a Shown are the mean deviation (MD), mean absolute deviation (MAD), and the maximum absolute deviation (MAX) from the reference values given by Boese, Martin, and Klopper (W2 calculations, values back-corrected from relativistic corrections, core correlation, complete basis set extrapolation, and BSSE corrections).²² The investigated test set includes interaction energies of 7 neutral and 4 charged H-bonded dimers (see Supporting Information) from the same reference. The optimized CCSD(T)/aug-cc-pVQZ geometries have been provided by Martin.²³ All calculations have been performed with restricted orbitals and the same computational setup as that for the reaction energies of the small molecules (def2-QZVPP basis set, see section 3.1.). Since the authors²² pointed out that for these systems the BSSE correction is almost canceled by the complete basis set correction we have not applied any BSSE procedure. The mean value of the absolute reference values for the neutral systems is 4.65 kcal/mol, and that of the charged systems is 21.94 kcal/mol.

the S22 training set of noncovalent interactions proposed by Hobza et al.²⁴ Since some larger systems are included in this test set and the use of diffuse basis functions is mandatory in order to describe dispersion interactions, we have used a smaller basis set (aug-cc-pVDZ²⁵). The CCSD(T) values calculated with the same basis set serve as reference values. Again, CEPA/1 and NCPF/1 outperform CCSD, QCISD and MP2; only for the benzene dimer (C_{2h} symmetry) and the pyrazine dimer (C_s symmetry), errors >1 kcal/mol have been observed. These two π -stacked complexes are particularly well described by the SCS-MP2 method that performs best among the tested methods for this set. Once more, SCS-MP3 does not improve further upon the SCS-MP2 values. As expected and in agreement with previous studies,²⁶ the B3LYP method yields large errors for almost all tested systems and should not be applied for calculations of van der Waals-bonded molecular complexes. Although the B2PLYP method performs much better than B3LYP, the results are not accurate enough without applying further corrections such as B2PLYP-D.¹³ By contrast, our results show that noncovalent interactions are reasonably well described by the coupled pair methods.

3.4. Barrier Heights. An important application for chemistry is the calculation of reactions barriers. In this area, DFT appears to have significant problems, and many (mostly ad hoc) modified procedures have appeared in the literature.^{27,28} The DFT problems are acknowledged to be dominated by the

TABLE 4. Statistical Data (in kcal mol⁻¹) for Interaction Energies of 10 Van der Waals-Bonded Molecules Calculated with Different Quantum Chemical Methods and the aug-cc-pVDZ Basis Set^a

	deviation from the CCSD(T) value		
	MAD	MD	MAX
CCSD	0.68	0.68	1.72
QCISD	0.71	0.71	1.69
CEPA/1	0.59	0.59	1.40
NCPF/1	0.59	0.59	1.40
SCS-MP3	0.52	0.52	0.83
SCS-MP2	0.29	0.21	0.40
MP2	0.67	0.66	2.44
B2PLYP	2.48	-2.48	5.11
B3LYP	4.25	4.25	9.15

^a Shown are the mean deviation (MD), mean absolute deviation (MAD), and the maximum absolute deviation (MAX) from the CCSD(T) reference values. The investigated test set includes interaction energies of 10 Van der Waals-bonded complexes (see Supporting Information) out of the S22 test set of Hobza et al.²⁴ The optimized geometries can be found in the Supporting Information of the same reference. All calculations have been performed with restricted orbitals and the aug-cc-pVDZ basis set.²⁵ The SCF iterations were converged to 10⁻⁷ Eh in the total energy (ORCA keyword TightSCF). Only the valence electrons have been correlated. We did not apply any CBS or BSSE procedure. The mean value of the absolute reference values is 4.23 kcal/mol.

self-interaction error, which is absent in Hartree–Fock based correlation methods. In order to assess the performance of the coupled pair type approaches, we have calculated barrier heights of the 12 reactions contained in Truhlar’s DBH24 test set.²⁹ Since most of the transition states are open shell molecules, we used the unrestricted versions of the CEPA methods. The most accurate of the tested methods, as expected, is QCISD(T) with an error of below 1 kcal/mol. Since there is no unrestricted NCPF/1 version yet, we have replaced it by the nearly identically performing NCEPA/1 variant.¹¹ NCEPA/1 also turns out to be the most accurate of the currently implemented coupled pair methods for an unrestricted reference determinant. It also profits from the enhanced stability of the ‘N’ variants proposed in the same paper. Both, the CEPA/1 and QCISD methods also give acceptable small errors (<2 kcal/mol). SCS-MP2 and B3LYP yield much larger errors, and the MP2 method completely fails for this test set. The relatively large error of the SCS-MP2 and MP2 method is mainly due to significant spin contamination of the investigated open shell transition states. For closed shell transition states, both MP2 variants yield more reliable results. BP86 is even worse than B3LYP since no HF exchange is included in the functional. Given that for each 1.4 kcal/mol error in the calculated barrier one predicts a reaction rate that is in error by another factor of 10, it is obvious that DFT predicts reaction rates that are typically off by several if not many orders of magnitude. A clear improvement is the double hybrid functional B2PLYP, which is fairly stable to more complicated situations such as spin contaminated open shell molecules. A similar enhanced stability was also reported for the calculation of atomization

TABLE 5. Statistical Data (in kcal mol⁻¹) for 12 Forward and 12 Backward Barrier Heights Calculated with Different Quantum Chemical Methods and the def2-QZVPP Basis Set^a

	deviation from DBH24 reference value		
	MAD	MD	MAX
QCISD(T)	0.68	-0.32	4.07
QCISD	1.64	1.51	7.52
CEPA/1	1.77	-1.00	5.12
NCEPA/1	1.32	-0.36	6.66
SCS-MP2	6.28	6.02	18.94
MP2	9.35	8.70	32.60
B2PLYP	2.82	-0.97	6.27
B3LYP	4.63	-4.60	10.18
BP86	9.00	-9.00	27.18

^a Shown are the mean deviation (MD), mean absolute deviation (MAD), and the maximum absolute deviation (MAX) from reference values for the DBH24 test set (see Supporting Information) of Truhlar et al.²⁹ The optimized geometries (QCISD/MG3 level) and the reference values (semiexperimental and calculated W1,W2 values) were taken from the same reference. All calculations have been performed with unrestricted orbitals and the same computational setup as that for the reaction energies of the small molecules (def2-QZVPP basis set; see section 3.1.). The mean value of the absolute reference values is 21.12 kcal/mol.

energies.¹³ However, the coupled pair methods again clearly outperform the density functional methods.

4. Conclusions

In this Account, we have evaluated the thermochemical and kinetic performance of the coupled pair approaches in an attempt to revive interest in them for large-scale chemical use. Indeed, our results conclusively show that these methods are highly suitable for chemical applications. They, of course, do not reach the accuracy of QCISD(T) and CCSD(T), but have consistently outperformed B3LYP, MP2, QCISD, CCSD, and to a large extent also SCS-MP2, in agreement with the results of ref 11. The attractive feature of the CEPA methods is their appealing simplicity. The superiority of CEPA and CPF relative to CCSD is not easily understood since the latter is based on a more rigorous derivation of the many body effects. It has frequently been argued that the slight overshooting of the many body effects in CEPA (and CPF) simulates the presence of connected triples.^{9,10} Unlike genuine coupled-cluster methods, they can easily be made stationary with respect to variations in the CI coefficients¹² and straightforwardly extended to the multireference and restricted open-shell cases. Moreover, we found it easier to derive approximate CEPA schemes that are suitable for efficient parallel implementation than to approximate coupled-cluster schemes. Our implementation (into the ORCA program¹⁷) has reached the stage where we can efficiently perform calculations with up to about 2200 basis functions.³² For normal sized molecules, the approximate CEPA code (in conjunction with a recently devised efficient approximate Hartree–Fock exchange³⁰ and the Split-RI-J algorithm

for the Coulomb matrix³¹) allows for calculations with extended basis sets in computation times that are frequently not much longer than those required for hybrid DFT calculations with popular program packages.³² The method and implementation are completely of black box character. These developments will be described in full detail elsewhere,³² and the program will be made accessible to the chemical community in due course. Now that extended accuracy tests have been successfully performed, we see no reason that would prevent the use of CEPA style methods for large-scale chemical applications. Work in this direction is underway in our laboratories.

We gratefully acknowledge the financial support of our research by the German Science Foundation, the collaborative research centers SFB 624, SFB 663, SFB 813, the priority program SPP 1137, The university of Bonn, the Max-Planck Society, and the German-Israeli Foundation.

Supporting Information Available. Individual results of all calculated atomization energies, reaction energies, dimer interaction energies and barrier heights as well as the cartesian coordinates of 16 molecules involved in the larger molecule reaction energy test set can be found in the Supporting Information. This material is available free of charge via the Internet at <http://pubs.acs.org>.

BIOGRAPHICAL INFORMATION

Frank Neese studied biology at the University of Konstanz and did his Ph.D. thesis in biochemistry (with Peter Kroneck, 1997). He then moved to Stanford University as a postdoctoral fellow with Edward Solomon where he focused on the spectroscopy and reactivity of nonheme iron centers. After returning to Germany, he did his habilitation at the University of Konstanz (2001) where he started the development of the ORCA quantum chemistry software. After five years as a group leader at the Max-Planck Institute for Bioinorganic Chemistry (with Karl Wieghardt), he accepted the offer for the chair of theoretical chemistry in Bonn (2006) where he now focuses on quantum chemical method development, spectroscopy, magnetism, and applications in transition metal and bioinorganic chemistry. He also holds a part-time appointment as Max-Planck fellow at the MPI for bioinorganic chemistry.

Andreas Hansen studied chemistry at the TU Chemnitz and the University of Bonn with special focus on theoretical chemistry. He received his diploma degree in 2007 for a thesis about open shell ab initio single reference methods, after which he applied for a Ph.D. under the supervision of Frank Neese and continues his research on efficient single reference wavefunction methods.

Frank Wennmohs studied biochemistry at the University of Bochum, where he did his diploma thesis on molecular dynamics simulations. He then moved to BayerCropScience for his Ph.D.

work in theoretical chemistry on weak hydrogen bonds. Afterwards, he joined the MPI for bioinorganic chemistry at Mühlheim a. d. Ruhr for the development of high level ab initio methods in ORCA. Since 2006, he has been working on his habilitation at the University of Bonn with the focus on noncovalent interactions in biological systems.

Stefan Grimme studied chemistry at TU Braunschweig and finished his Ph.D. in 1991 in physical chemistry on a topic in laser spectroscopy. He then moved to Bonn where he did his habilitation in theoretical chemistry in the group of Sigrid Peyerimoff. In 2000, he got the C4 chair for theoretical organic chemistry at the University of Münster. His research interests are the development of quantum chemical methods for large systems, density functional theory, electronic spectroscopy/excited states, and the properties of chiral systems.

FOOTNOTES

* To whom correspondence should be addressed. neese@thch.uni-bonn.de.

REFERENCES

- Karton, A.; Rabinovich, E.; Ruscic, B.; Martin, J. M. L. W4 theory for computational thermochemistry: In pursuit of confident sub-kJ/mol predictions. *J. Chem. Phys.* **2006**, *125*, 144108.
- Wood, G. P. F.; Radom, L.; Petersson, G. A.; Barnes, E. C.; Frisch, M. J.; Montgomery, J. A. A. restricted-open shell complete-basis-set model chemistry. *J. Chem. Phys.* **2006**, *125*, 094106.
- Kohn, W.; Becke, A. D.; Parr, R. G. Density functional theory of electronic structure. *J. Phys. Chem.* **1996**, *100*, 12974–12980.
- Ziegler, T. Density functional theory as a practical tool in studies of organometallic energetics and kinetics. Beating the heavy metal blues with DFT. *Can. J. Chem.* **1995**, *73*, 743–761.
- Grimme, S.; Steinmetz, M.; Korth, M. How to compute isomerization energies of organic molecules with quantum chemical methods. *J. Org. Chem.* **2007**, *72*, 2118–2126.
- Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002**, *116*, 3175–3183.
- Schütz, M.; Werner, H.-J. Local perturbative triples correction (T) with linear cost scaling. *Chem. Phys. Lett.* **2000**, *318*, 370–378.
- Saebo, S.; Pulay, P. Local treatment of electron correlation. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213–236.
- (a) Meyer, W. *Configuration Expansion by Means of Pseudonatural Orbitals in Methods of Electronic Structure Theory*; Schaefer, H. F., III, Ed.; Plenum Press: New York, 1977, pp 413–446. (b) Meyer, W. PNO-CI studies of electron correlation effects. 1. Configuration expansion by means of nonorthogonal orbitals, and application to ground-state and ionized states of methane. *J. Chem. Phys.* **1973**, *58*, 1017–1035.
- (a) Ahlrichs, R. Many-body perturbation calculations and coupled electron pair models. *Comp. Phys. Comm.* **1979**, *17*, 31–45. (b) Ahlrichs, R.; Driessler, F.; Lischka, H.; Staemmler, V.; Kutzelnigg, W. PNO-CI (pair natural orbital configuration interaction) and CEPA-PNO (coupled electron pair approximation with pair natural orbitals) calculations of molecular systems. 2. Molecules BeH₂, BH, BH₃, CH₄, CH₃⁻, NH₃ (planar and pyramidal), H₂O, OH⁺³, HF and Ne atom. *J. Chem. Phys.* **1975**, *62*, 1235–1247.
- Wennmohs, F.; Neese, F. A comparative of single reference correlation methods of the coupled-pair type. *Chem. Phys.* **2008**, *343*, 217–230.
- Ahlrichs, R.; Scharf, P.; Ehrhardt, C. The coupled pair functional (CPF). A size consistent modification of the CI(SD) based on an energy functional. *J. Chem. Phys.* **1985**, *82*, 890–898.
- Schwabe, T.; Grimme, S. Theoretical Thermodynamics for Large Molecules: Walking the thin line between accuracy and computational cost. *Acc. Chem. Res.* **2008**, *41*, 569–578.
- Curtiss, L. A.; Raghavachari, K.; Redfern, P. C. Assessment of Gaussian-2 and density functional theories for the computation of enthalpies of formation. *J. Chem. Phys.* **1997**, *106*, 1063–1079.

- 15 Parthiban, S.; Martin, J. M. L. Assessment of W1 and W2 theories for the computation of electron affinities, ionization potentials, heats of formation, and proton affinities. *J. Chem. Phys.* **2001**, *114*, 6014–6029.
- 16 Grimme, S. Improved second-order Møller-Plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies. *J. Chem. Phys.* **2003**, *118*, 9095–9102.
- 17 Neese, F. *ORCA, an Ab Initio, Density Functional and Semi-Empirical Program Package*; University of Bonn: Germany, 2008. <http://www.thch.uni-bonn.de/tc/orca>.
- 18 Weigend, F.; Furche, F.; Ahlrichs, R. Gaussian basis sets of quadruple- ζ valence quality for atoms H-Kr. *J. Chem. Phys.* **2003**, *119*, 12753–12762.
- 19 Ahlrichs, R., Basis sets were obtained from the ftp server of the Karlsruhe quantum chemistry group: <ftp://chemie.uni-karlsruhe.de/pub/bases>.
- 20 Schäfer, A.; Huber, C.; Ahlrichs, R. Fully optimized contracted Gaussian basis sets of triple- ζ valence quality for atoms Li to Kr. *J. Chem. Phys.* **1994**, *100*, 5829–5835.
- 21 Repasky, M. P.; Chandrasekhar, J.; Jørgensen, W. L. PDDG/PM3 and PDDG/MNDO: Improved semiempirical methods. *J. Comput. Chem.* **2002**, *23*, 1601–1622.
- 22 Boese, A.; Daniel, M.; Jan, M. L.; Klopper, W. Basis set limit coupled cluster study of H-bonded systems and assessment of more approximative methods. *J. Phys. Chem. A* **2007**, *111*, 11122–11133.
- 23 Martin, J. M. L., personal communication, 2008.
- 24 Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.
- 25 Dunning, T. H., Jr. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
- 26 Zhao, Y.; Truhlar, D. G. A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions. *J. Chem. Phys.* **2006**, *125*, 194101.
- 27 Johnson, E. R.; Becke, A. D. A. unified density-functional treatment of dynamical, nondynamical, and dispersion correlations. II. Thermochemical and kinetic benchmarks. *J. Chem. Phys.* **2008**, *128*, 124105.
- 28 Tomopolsky, A.; Karton, A.; Sertchook, R.; Vuzman, D.; Martin, J. M. L. Double-hybrid functionals for thermochemical kinetics. *J. Phys. Chem. A* **2008**, *112*, 3–8.
- 29 Zheng, J.; Zhao, Y.; Truhlar, D. G. Representative benchmark suites for barrier heights of diverse reaction types and assessment of electronic structure methods for thermochemical kinetics. *J. Chem. Theory Comput.* **2007**, *3*, 569–582.
- 30 Neese, F.; Hansen, A.; Wennmohs, F.; Becker, U. Efficient, approximate, parallel Hartree-Fock and hybrid DFT calculations. A linear scaling 'Chain-of-Spheres' algorithm for the Hartree-Fock exchange. *Chem. Phys.* **2009**, *356*, 98–109.
- 31 Neese, F. An improvement of the resolution of the identity approximation for the calculation of the coulomb matrix. *J. Comput. Chem.* **2003**, *24*, 1740–1747.
- 32 Neese, F.; Wennmohs, F.; Hansen, A. Efficient and accurate local approximations to coupled electron pair approaches. An attempt to revive the pair-natural orbital method. *J. Chem. Phys.*, in press.
- 33 Grimme, S. Improved third-order Møller-Plesset perturbation theory. *J. Comput. Chem.* **2003**, *24*, 1529–1537.
- 34 Grimme, S. Seemingly simple stereoelectronic effects in alkane isomers and the implications for Kohn-Sham density functional theory. *Angew. Chem., Int. Ed.* **2006**, *118*, 4571–4575; **2006**, *45*, 4460–4464.
- 35 Taylor, P.; Bacskay, G. B.; Hush, N. S.; Hurley, A. C. Coupled-pair approximation in a basis of independent-pair natural orbitals. *Chem. Phys. Lett.* **1976**, *41*, 444–449.