

Stabilization and Structure Calculations for Noncovalent Interactions in Extended Molecular Systems Based on Wave Function and Density Functional Theories

Kevin E. Riley,^{⊥,†} Michal Pitoňák,^{⊥,‡,§} Petr Jurečka,^{⊥,||} and Pavel Hobza^{*,‡,||}

Department of Chemistry, University of Puerto Rico, Rio Piedras, Puerto Rico 00931, Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems, 166 10 Prague, Czech Republic, Department of Physical and Theoretical Chemistry, Faculty of Natural Sciences, Comenius University, 842 15 Bratislava, Slovak Republic, and Department of Physical Chemistry, Palacký University, 771 46 Olomouc, Czech Republic

Received January 19, 2010

Contents

1. Introduction	5023	4.2.3. Beyond the C ₆ Term	5038
2. Scope of the Review	5025	4.2.4. Semiempirical Dispersion Correction Based on the Exchange Hole Dipole Moment	5038
3. Wave Function Theory	5026	4.3. Dispersion Correction in the Plane Wave Codes and Nonlocal Pseudopotentials	5038
3.1. CCSD(T) in the Complete Basis Set Limit: Golden Standard in Calculation of Noncovalent Interactions	5026	4.4. Reparametrization of Current Density Functionals	5039
3.1.1. Computational Demands and Applicability	5027	4.5. Double Hybrid Functionals	5039
3.1.2. Accuracy of Supermolecular CCSD(T)	5027	4.6. Truly Nonlocal Density Functionals for DFT	5040
3.1.3. Basis Set Convergence of the CCSD(T) Interaction Energy	5028	4.7. Other Approaches	5040
3.1.4. Approximate CCSD(T) Schemes	5029	5. Symmetry-Adapted Perturbation Theory	5041
3.2. MP2 in the Complete Basis Set	5029	6. Semiempirical Quantum Chemical Theories	5042
3.2.1. Computational Demands and Applicability	5029	7. JSCH-2005 and S22 Databases	5042
3.2.2. Approximate MP2 Schemes	5030	8. Results and Discussion	5043
3.2.3. Empirical Corrections to the MP2	5030	8.1. Energy Characteristics	5043
3.3. Methods Complete to the Third Order of Perturbation Theory: From MP3 to CCSD	5032	8.1.1. WFT Methods	5044
3.3.1. Computational Demands and Applicability	5032	8.1.2. DFT Methods	5046
3.3.2. Accuracy of the Supermolecular Third-Order Methods	5032	8.1.3. Symmetry-Adapted Perturbation Theory	5048
3.3.3. Empirical Third-Order Schemes	5032	8.1.4. Semiempirical QM Methods	5048
3.4. Increasing the Rate of Convergence of Interaction Energy Calculations toward the Complete Basis Set Limit	5034	8.2. Geometry Characteristics	5049
4. Density Functional Theories	5035	8.2.1. Potential Energy Curves	5050
4.1. Main Limitation of Density Functional Theory: Dispersion Interaction	5035	8.2.2. Analytical Gradients	5054
4.1.1. Locality of the Density Functionals	5036	8.2.3. Conformations of Peptides	5055
4.1.2. Terminology of Correlation Interaction in DFT	5036	9. Conclusions	5058
4.1.3. Perturbative vs Self-Consistent Treatment of the Dispersion Correction	5036	10. Outlook	5058
4.2. DFT and Empirical Dispersion Corrections	5036	11. Acknowledgments	5059
4.2.1. Role of the Damping Function	5037	12. Note Added in Proof	5059
4.2.2. Intramolecular Dispersion	5037	13. References	5060

1. Introduction

More than 20 years ago, we published in *Chemical Reviews* a paper entitled “Intermolecular Interactions between Medium-Sized Systems. Nonempirical and Empirical Calculations of Interaction Energy: Successes and Failures”.¹ The situation in calculations of noncovalent interactions at that time can be best characterized by the question we posed at the very beginning of the review: “Can quantum chemistry describe vdW (van der Waals; today we call it noncovalent) interactions as successfully as covalent interactions?” Our answer then was “unambiguously yes”. We had good reason for an optimistic “yes” since we presented the first coupled-cluster calculations including triple excitations for a (at that time) large complex—the water dimer.² We stressed the

* To whom correspondence should be addressed. E-mail: pavel.hobza@uochb.cas.cz. Phone: +420 220 410 311. Fax: +420 220 410 320.

[⊥] These authors contributed equally to this work and should be considered co-first authors.

[†] University of Puerto Rico.

[‡] Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems.

[§] Comenius University.

^{||} Palacký University.



Kevin E. Riley obtained his bachelor's degrees in chemistry and physics from the University of New Mexico in 1999 and his Ph.D. from the Pennsylvania State University in 2004 (with James B. Anderson). He then went on to do postdoctoral studies at the University of Florida (with Kenneth M. Merz, Jr.) and the Academy of Sciences of the Czech Republic (with Pavel Hobza). Dr. Riley is now an Assistant Professor in the Department of Chemistry at the University of Puerto Rico (Rio Piedras campus). His main interests are related to the computational treatment of noncovalent interactions in biomolecular systems, with special emphasis on complexes of small-molecule ligands with proteins and nucleic acids.



Petr Jurečka was born in 1976 in Valašské Meziříčí, Czech Republic, and graduated from the Charles University in 2000. In 2004 he received his Ph.D. at the Academy of Sciences of the Czech Republic (Institute of Organic Chemistry and Biochemistry with Professor Pavel Hobza). He did his postdoctoral study with Professor D. R. Salahub at the University of Calgary, Canada. He is currently a research associate at the Palacký University in Olomouc and at the Institute of Organic Chemistry and Biochemistry, Prague, Czech Republic. His research interests are in applied quantum chemistry, density functional theory, empirical potential development, weak intermolecular interactions, and nanochemistry.



Michal Pitoňák was born in 1979 in Nové Zámky, Czechoslovakia, and graduated from Comenius University in Bratislava, Slovakia, in 2003. In 2006 he received his Ph.D. (with Professor M. Urban) and from the beginning of 2007 until the present has been conducting postdoctoral studies at the Institute of Organic Chemistry and Biochemistry, Academy of Science of the Czech Republic, with Professor Pavel Hobza. Since 2009 he has also been employed as an Assistant Professor at the Department of Physical and Theoretical Chemistry at the Faculty of Natural Sciences, Comenius University in Bratislava. Dr. Pitoňák has authored about 20 scientific publications, which have dealt mainly with large-scale applications of highly accurate *ab initio* methods, mostly on molecular interactions.

importance of the triple excitations for noncovalent interactions, and in the section called Prospects, we wrote that "significant progress is highly desirable with beyond-SCF methods, where new, more accurate and efficient procedures are developed". In this respect we were right, and in the past 20 years, we have witnessed an enormous growth of interest in the fast and accurate calculation of intermolecular interactions. What is the reason for such an interest, or more generally, why are noncovalent interactions so relevant in modern research? Is it the mere existence of noncovalent complexes in the gas and liquid phases? Certainly not. The answer should be sought in the role that noncovalent interactions are playing in both bio- and nanostructures.

All life on our earth can be viewed as an application of supramolecular chemistry with noncovalent interactions



Pavel Hobza finished his Ph.D. in 1973 with Rudolf Zahradnik, and since then he has worked in the area of noncovalent interactions and their applications in chemistry and biodisciplines. Currently he is Head of the Center for Biomolecules and Complex Molecular Systems at the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic in Prague, and Professor of Physical Chemistry at Charles University in Prague, Czech Republic, and Palacký University in Olomouc, Czech Republic. Since 2009 he has also been a World Class University Professor at Postech University in Pohang, Korea. His research interests focus on theory and calculation for noncovalent interactions, development of new computational procedures, and *in silico* drug design.

playing a central role. Biomacromolecules, DNA, RNA, and proteins, all play a dominant role in our life. The function of these biomacromolecules is to a large extent determined by their structures, so forming a deep understanding of the nature of the stabilization of these systems is thus of key importance. For example, the double-helical structure of DNA is clearly linked with its function—the storage and transfer of genetic information. The 3D structure of DNA (and also of other biomacromolecules such as proteins, etc.) results from a delicate balance between various types of noncovalent interactions acting between biomolecular building blocks and also between these blocks and the surrounding environment. The balance mentioned is also responsible for what is probably the most important biological process—molecular recognition, which is the fascinating process by which one system recognizes a second one over a very long

distance, and undoubtedly directs it to form an active complex. Finally, an interaction of a ligand with a protein target, which is fully determined by various types of noncovalent interactions and solvation/desolvation processes, represents the key step in *in silico* drug design, which will undoubtedly play a more and more important role in the development of new potent drugs.

Among the most important noncovalent interaction types are hydrogen bonding (H-bonding) and stacking; these motifs are found not only in nucleic acids and proteins, but also in complexes of nucleic acids and proteins with ligands. For a long time it was believed that H-bonding was much stronger than stacking, and consequently, it was expected that H-bonding was the key player in determining biomolecular structure. At the time, the role of stacking was, in fact, not completely clear. Accurate quantum mechanical (QM) calculations performed in the past few years in our laboratory^{3,4} have, surprisingly, shown that the stabilization of both motifs can in fact be comparable. Such calculations have been, however, difficult, which is mainly due to the very different origins of each stabilization motif. While the H-bond forms due to electrostatic interactions and charge transfer, stacking is almost exclusively governed by the London dispersion energy. Accurate evaluation of the former energy contribution is straightforward, and almost any QM level describes H-bonding properly. It is an entirely different situation for the stacking interactions, where the most accurate QM calculations covering a large portion of the correlation energy combined with extended AO basis sets are required. To properly qualitatively and quantitatively describe the stabilization of biomolecular building blocks, high accuracy is required. In our laboratory, and in others, it was shown that chemical accuracy (~ 1 kcal/mol) can be obtained by performing the coupled-cluster calculation covering the single and double electron excitations iteratively and the triple electron excitations perturbatively (CCSD(T)) in combination with the complete basis set (CBS) limit extrapolations. The resulting CCSD(T)/CBS scheme provides accurate stabilization energies (as well as other properties) for various types of biomolecular motifs including H-bonding and improper blue-shifting H-bonding, stacking, electrostatic interactions, charge transfer, dihydrogen bonding, and halogen bonding.

CCSD(T)/CBS calculations thus provide benchmark data that can be used not only for investigating the nature of noncovalent interactions in various binding motifs but also for testing and/or parametrizing other, more computationally economical, *ab initio* nonempirical QM wave function and density functional theories (WFT and DFT) as well as semiempirical QM methods and empirical potentials (EPs). The accuracy of the CCSD(T) procedure for nanostructures and their models can be, however, questioned since upon delocalization of the systems (e.g., graphene) the HOMO–LUMO gap decreases and the use of perturbation methods (for evaluation of triple excitations) is no longer justified.⁵ For the same reason the use of the Møller–Plesset perturbation technique (e.g., MP2) can also be limited.

The very unique position of the CCSD(T)/CBS technique follows from the fact that it is the only theoretical *ab initio* procedure (in the sense that no empirical data are utilized) that provides accurate stabilization energies for various types of noncovalent complexes. All other WFT, DFT, and EP procedures contain one or more parameters that were most likely fitted toward the CCSD(T)/CBS or experimental data. It is true that standard WFT methods such as MP n or CCSD

do not contain any parameters, but their performance for noncovalent interactions is limited. For example, the most popular among these methods, the second-order Møller–Plesset perturbation theory, strongly overestimates the dispersion energy (see below), which causes difficulties in the description of dispersion-bound complexes. This problem can be reduced either by separate scaling of spin components or by the combination of MP2 with higher terms in the perturbation expansion (e.g., MP3). Both procedures mentioned (for references and details see below) should, however, include an empirical parameter.

The development of faster QM procedures is highly important since we need to describe larger and larger fragments of bio- and nanostructures as accurately as possible. It is especially clear that interactions within these systems are governed not only by classical short- and long-range interactions, which can be basically described by EPs, but also by effects that are clearly of quantum origin, for which EP-based methods fail. Among quantum effects that must be considered, probably the most important are charge-transfer phenomena, which are especially significant in biological processes with the participation of metals or electron donor and electron acceptor subsystems. Fast and accurate QM methods are also needed in molecular dynamic simulations where the description of quantum effects plays a decisive role. The development of such methods is difficult since here the CPU time cannot be much higher than that required by EP methods.

The literature on noncovalent interactions is broad, and many useful reviews have originated from our laboratory^{1,6–14} and in others.^{15–19}

2. Scope of the Review

In the past 10 years, an unusually large number of new QM approaches have emerged, in both the WFT and DFT families (references will be given in the subsequent paragraphs). The MP2 procedure, when combined with an extended AO basis set (or even when performed at the CBS limit), basically provides reliable data for H-bonding, but generally overestimates the stacking interactions (by as much as several kilocalories per mole). The solution to this problem came from the introduction of the spin component scaling method based on the MP2 procedure (SCS-MP2). The method solved the overestimation of MP2 for stacking but failed for the description of H-bonding, which triggered the introduction of several methods based on SCS-MP2, such as SCS(MI)-MP2, SOS-MP2, SOS(MI)-MP2, SSS(MI)-MP2, and SCSN-MP2. The above-mentioned problems of the MP2 procedure were successfully removed by passing to MP3, and the technique called scaled MP3 (or simply MP2.5) provides very promising results for various types of noncovalent complexes. It is true that MP2.5 requires about 1 order of CPU time more than the MP2 procedure, but it is still much faster than CCSD(T).

The main problem with DFT methods, which are computationally very efficient, is the fact that they do not describe the dispersion energy. It was proved rigorously that local as well as semilocal DFT techniques do not cover the dispersion.²⁰ This problem has been addressed along different lines; among the simplest of these is a very efficient procedure in which the dispersion energy is added using the traditional C_6/R^6 term. A very similar approach was used in the second half of the past century by Scoles and Ahlrichs and also by one of us (P.H.) to remove the analogous problems of the

Hartree–Fock method (for which the dispersion energy is completely missing). The approach was first applied to the tight-binding DFT procedure and later to other standard DFT procedures. During the period mentioned, the years-long pursuit for DFT methods applicable to noncovalent complexes peaked with the M06 and M08 suites of functionals and also with the first viable truly nonlocal functional. All these advancements were fueled by a growing appreciation of the importance of London dispersion for virtually all applications, mainly in biology and nanostructures.

The availability of accurate reference data contributed a great deal to this development. Substantial improvements have lately been achieved in alternative paradigms for accurate calculations not requiring the introduction of any empirical parameter, such as symmetry-adapted perturbation theory (SAPT), and especially its DFT approximation called DFT-SAPT or SAPT/DFT, and the diffusion quantum Monte Carlo (DQMC) procedure. Agreement of their results with the CCSD(T)/CBS calculations has established a consensus about the quality of the reference data (and also about the quality of both methods mentioned) and put accuracy assessment on firm ground. The SAPT procedure, when based on correlated descriptions of subsystems, provides highly accurate results, and the method is a genuine *ab initio* method. The procedure is, however, computationally expensive, which prevents its use for larger complexes. The DFT-SAPT technique allows for the treatment of much larger complexes, but only at the price of adding empirical parameters to the calculations.

In the current review we encompass the recent developments in this field, facilitate orientation among the quickly appearing treatments, and emphasize the vital importance of accurate reference calculations. In the past five years we have published about 100 papers covering the subject of the current review, and these papers, together with literature references, will make the base of the present review. We will concentrate mainly on the performance of various theoretical procedures in the computation of stabilization energies, but the evaluation of optimized geometries will also be mentioned (though to a lesser extent).

Due to the unfavorable scaling of the CCSD(T) method (roughly with the seventh power of the system size), its applicability is limited to complexes with no more than about 50 atoms. It is thus highly topical to investigate the performance of other WFT procedures that can be applied for larger complexes. As mentioned above, all of these procedures contain empirical parameters that have been fitted, for the most part, to the CCSD(T)/CBS benchmark data. The most widely used method, without doubt, is MP2, which covers the essential portion of the correlation energy and is still feasible even for complexes having more than 100 atoms. The main drawbacks of this method are the heavy dependence on the basis set as well as on the geometry of the complex. The limitations of the method will be discussed together with suggestions about how to overcome them. Here, several methods based on separate scaling of parallel and antiparallel spin contributions (SCS-MP2, SCS-CCSD) will be mentioned, and their advantages and disadvantages will be mentioned and demonstrated for selected examples. Some attention will also be paid to the scaled MP3 method, which is almost parameter-free (in its simplest version, it is an arithmetic mean of the MP2 and MP3 stabilization energies) and provides excellent stabilization energies for various types

of noncovalent complexes. Despite the fact that scaled MP3 and variants of SCS-MP2 are much faster than CCSD(T), their use for extended complexes is still limited. Furthermore, it must be mentioned that both these techniques are not based on solid theoretical grounds. The next step is thus the utilization of DFT methods.

The section on DFT should start with a brief discussion of the limitations of standard DFT functionals in the description of noncovalent interactions. Successes and failures of DFT in studies of H-bonded and stacked structures will be demonstrated on DNA base pairs and amino acid pairs. The “hereditary sin” of DFT, a failure to describe the dispersion energy, can be overcome in several different ways, which will be discussed at the beginning of the DFT paragraph. Among various techniques, two approaches, the addition of an empirical dispersion energy term combined with parametrization of the damping function and the *de novo* parametrization of the DFT functional, became by far the most successful and will be discussed in detail. Attention will be paid not only to stabilization energies but also to structures and geometries. DFT calculations are much faster than correlated WFT calculations and do not depend so heavily on the basis set size; furthermore, additional speedups can be gained by using GGA functionals instead of hybrid ones. A very promising approach entails a combination of DFT-D (dispersion energy corrected) with a GGA functional and a medium basis set. This procedure is fast and can even be used in on-the-fly *ab initio* (DFT-D) molecular dynamics simulations. This technique is a very promising tool for studies of the dynamic properties of small and medium noncovalent complexes.

Generally, semiempirical QM procedures are known not to be suitable for calculations on noncovalent interactions, with the main problem being, again, the deficiency in the description of the dispersion energy. However, if this energy term is added (in a way similar to that in DFT-D), the semiempirical QM procedures can be successfully used even in the realms of noncovalent interactions. The main attention will be paid to the PM6 method, belonging to the family of semiempirical NDDO methods; the PM6 method represents the most recent parametrization of the well-known series of PM x methods. The PM6 method, upon addition of the empirical dispersion energy and H-bonding terms (PM6-DH), provides surprisingly accurate results for various types of noncovalent complexes and, due to its favorable scaling (approaching linear, for instance, for peptides), can be used for complexes with several thousand atoms. An important advantage of the latter method is that it has been parametrized for the entire periodic table.

3. Wave Function Theory

3.1. CCSD(T) in the Complete Basis Set Limit: Golden Standard in Calculation of Noncovalent Interactions

Coupled-cluster (CC) theory was introduced into quantum chemistry more than 40 years ago.^{21–23} The concept of CC theory relies on the exponential formulation of the wave operator and its expansion into clusters of excitation operators. As a consequence, CC methods are size-extensive (the energy scales properly with the size of a system) and the convergence of the CC energy toward the full configuration interaction (FCI) value is faster compared to that of other

methods with the same asymptotic scaling with respect to system size. One of the valuable features of CC theory is that it is systematically improvable upon inclusion of a higher excitation operator,²³ providing that an adequate AO basis set for description of the system is applied, typically, but not strictly:

$$\text{CCSD} < \text{CCSD(T)} < \text{CCSDT} < \text{CCSDT}(Q_p) < \\ \text{CCSDT}(Q) < \text{CCSDTQ} < \text{FCI}$$

The CCSD(T) method, i.e., iterative inclusion of the single and double (and higher disconnected linked terms such as quadruple, hexuple, etc., resulting from the exponential ansatz) excitations with perturbative account for the effect of triple excitations in energy, is especially successful for ground-state energies and the calculation of properties for systems with single-reference character (which is generally the case in the world of noncovalent interactions). As demonstrated in several works, among all CC approximations, CCSD(T) dauntlessly stands as the “golden-standard” method due to its outstanding accuracy for the computational cost ratio (please see ref 24 and references therein).

3.1.1. Computational Demands and Applicability

Before analysis of the accuracy of the CCSD(T) method on calculations of noncovalent interactions, let us focus first on its applicability. Inclusion of single and double excitations in the CC scheme, i.e., CCSD, leads to equations with numerous terms (or “diagrams”), from which the most computationally demanding scale as $\sim N_o^2 N_v^4$ and $\sim N_o^3 N_v^3$, where N_o stands for the number of correlated occupied and N_v the number of active virtual (unoccupied or external) orbitals. The overall scaling of a CCSD iteration is usually dominated by one of these terms, depending on the N_o/N_v ratio. If no linear dependences within the underlying AO basis sets are removed, this ratio is defined by the dimension of the basis set, since the virtual orbitals are only a “byproduct” of the MO-LCAO-HF (molecular orbital linear combination of atomic orbitals Hartree–Fock) procedure, with the only condition being orthogonality to the space of occupied orbitals. The scaling of the perturbative triples calculation is by an order of magnitude steeper, i.e., $N_o^3 N_v^4$, which represents the bottleneck of CCSD(T) calculations on larger systems.

The N^7 scaling of the floating-point operations in the CCSD(T) method can be quite efficiently alleviated nowadays (but certainly not overcome) by massive parallelization, but there are other bottlenecks that have been recognized as being important only quite recently.^{25–29} The bottleneck of storage requirements resulting from storage of N^4 two-electron integrals was addressed 17 years ago by introducing the so-called integral-direct CC algorithms,³⁰ Cholesky decomposition,^{31,32} and more recently DF/RI (density fitting/resolution-of-the-identity) of two-electron integrals.^{33,34} However, the important limitations to be removed are the ones affecting the computer memory. Keeping, for instance, T_2 double excitations ($N_o^2 N_v^2$ sized array) in-core was a common praxis in “standard” CCSD(T) implementations. Storing of this “small” array can easily exceed the capacity of local memory on standard computers (typically 1–8 GB per CPU). This bottleneck has been overcome in all of the “modern” CCSD(T) implementations cited above in various ways; perhaps the ACES3²⁶ program suite serves as the best example. In this code all types of array indices, i.e., AOs,

occupied and virtual, can be, if needed, segmented to suit the memory available. Segmentation brings in certain overhead in I/O, data transfer, and arithmetic operations compared to traditional algorithms, but the scaling of this overhead with system size is typically lower than the actual work to be done in the entire algorithm by 1–2 orders of magnitude.

To clearly circumscribe the area of applicability of the CCSD(T) approach in terms of the size of the calculated system is difficult. The potential of current software implementations and the power of supercomputers have certainly not been fully exploited yet. The largest “rigorous” CCSD(T) calculations published so far have dealt with systems of about 70 atoms, such as the coronene dimer⁵ and a guanine–cytosine step from DNA.³⁵

3.1.2. Accuracy of Supermolecular CCSD(T)

In terms of perturbation theory of electron correlation, the CCSD(T) method is accurate to the fourth order in energy (with certain terms being summed up to the infinite order). Considering CCSD(T) as a “benchmark” method for molecules of about 30 (second-row and hydrogen) atoms, which nowadays can be calculated close to the CBS limit, we can still ask ourselves how good these benchmarks really are? As pointed out in several papers, interaction energies calculated at the CCSD(T)/CBS level are of, so-called, chemical accuracy (errors of less than 1 kcal/mol), but can be deficient for subchemical accuracy (errors of less than 0.1 kcal/mol) in certain cases.^{36–38}

Typically, the largest errors can be attributed to the incompleteness of the AO basis set and the resulting basis set superposition error (BSSE). This might be a problem for, say, “real-life” systems, where calculations even close to the CBS limit are not tractable due to their size. The evaluation of the BSSE error for an arbitrary AO basis set is obviously not possible; however, “bracketing” of the CBS value based on BSSE-corrected and -uncorrected interaction energies provides a reasonable estimate of the BSSE in regions reasonably close to the CBS value. The basis set convergence of CCSD(T) will be discussed in more detail in section 3.1.3.

Let us focus on the accuracy of the CCSD(T) supermolecular interaction energies in terms of the approximations made at the CC level. As stated in ref 37, certain fifth-order terms are only partially covered at the supermolecular CCSD(T) (as well as by the fully iterative CCSDT) level, and either perturbative or iterative inclusion of connected quadruple excitations is required to introduce these terms. Quantitative assessment of these effects is hindered by insufficient basis set saturation, since higher order correlation contributions tend to compensate for the one-particle basis incompleteness. Thus, any conclusion drawn from CC calculation including connected quadruple excitations (or higher) in unsaturated basis sets can be only semiquantitative. From the very few publications on this topic, the following conclusion can be drawn: The error from the perturbative approximation of the triple excitation, i.e., CCSDT vs CCSD(T), amounts to about 1% of the $\Delta\text{CCSD(T)}$ (i.e., $\text{CCSD(T)} - \text{MP2}$; see below) term for π – π stacked complexes, being repulsive. For H-bonded complexes this error can account for as much as 10–30% of the ΔCCSDT term, being repulsive as well.³⁹ Nevertheless, this alarmingly large error, on a percentile scale, affects the total interaction energy only marginally, since MP2 already describes H-bonding fairly well. The reason for such a discrepancy

between CCSD(T) and CCSDT probably originates in the coupling of T_3 with mono- and biexcitation amplitudes in the CCSDT scheme, thus affecting the dipole moment, as well as the higher electric moments of monomers. This significantly influences the intramonomer correlated electrostatic and deformation components of the interaction energy, which are dominant in H-bonding interactions. The error attributable to neglect of the connected quadruple excitations for π - π stacking complexes is about 10% of the (T) contribution,^{40,41} being repulsive, or about 5%, according to our CCSD(TQ_r)/6-31G*(0.25) calculations of several conformers of the benzene dimer,⁴² being repulsive as well.

To summarize the error evaluation, fortuitously enough, the CCSD(T) method provides the upper bound for the stabilization energies due to the opposing signs of the neglected higher order correlation terms described above.⁴⁰ The overall errors are typically expected to be less than 10% in the higher order correlation term (Δ CCSD(T)) or less than 3% in the total interaction energy (disregarding a few "pathological" complexes, such as rare-gas dimers or small diatomic dimers).

A last note should be made on the effect of core correlation, which was shown to be quite important for certain properties, such as atomization energies, excitation energies, and NMR shifts. There has not been very much attention paid to its magnitude in calculations of noncovalent interactions. Only a few in-depth works on this topic (mostly on rare-gas dimers) can be found.⁴³⁻⁴⁵ The effects of core correlation on H-bonding were studied by Boese et al.,⁴⁶ reporting that these effects can be completely neglected, being on the order of a few hundredths of a kilocalorie per mole (even for complexes containing the chlorine atom). As an illustration, frozen-core (1s orbitals of C, N, and O atoms inactive) and all-electron MP2 and CCSD(T) calculations of interaction energies in the H-bonding and stacking conformations of the uracil dimer were carried out with the aug-cc-pVDZ and aug-cc-pCVDZ⁴⁷ basis sets.⁴⁸ No additional stabilization at the MP2 level and only 0.02 kcal/mol additional stabilization at the CCSD(T) level were found for the H-bonded structure, while an additional stabilization of 0.02 kcal/mol at the MP2 level and 0.01 kcal/mol destabilization, due to higher order correlation, at the CCSD(T) level were found for the stacked structure. It can be expected that, for biologically relevant noncovalent complexes, core-correlation effects are beyond, or at, the order of magnitude of subchemical accuracy.

3.1.3. Basis Set Convergence of the CCSD(T) Interaction Energy

As already mentioned, errors resulting from the incompleteness of the AO basis set are typically the most significant ones. However, if the CCSD(T) interaction energy is decomposed into SCF, MP2, and higher order correlation corrections

$$\Delta E(\text{CCSD(T)}) = \Delta E(\text{SCF}) + \Delta E(\text{MP2}) + \Delta E(\Delta\text{CCSD(T)}) \quad (1)$$

where $\Delta E(\Delta\text{CCSD(T)})$ will be, for the sake of simplicity, further referred to as $\Delta\text{CCSD(T)}$. Different rates of convergence toward the CBS of these components can be recognized. The SCF interaction energy is the fastest converging of these components, and is usually already saturated in medium-sized, sufficiently diffuse basis sets, e.g., aug-cc-

pVTZ. If the counterpoise correction for BSSE is applied, results close to the CBS can be obtained with even smaller basis sets, such as aug-cc-pVDZ. Unlike the SCF component, the MP2 correlation contribution, clearly dominant in the CCSD(T) correlation energy expansion, is the slowest converging one. For stacked dimers of about 30–40 atoms, e.g., DNA base pairs, even basis sets as large as aug-cc-pVQZ (the use of which results in calculations with more than 2000 basis functions) still can lead to deviations of several kilocalories per mole from the CBS. A significant convergence speedup toward the CBS can be achieved by applying basis set extrapolations of energies obtained in two or more AO basis sets, by using either the optimized virtual orbital space (OVOS) or frozen natural orbital (FNO) method or by introducing explicit correlation via R12/F12 operators. All of these approaches are discussed in section 3.4.

The basis set convergence of the $\Delta\text{CCSD(T)}$ term, which is investigated in several publications,^{46,49-51} deserves special attention. Since calculating this term is the most computationally demanding part of obtaining estimated CCSD(T)/CBS results, major savings can be achieved by "optimizing" the basis set selection for this term. As shown by Jurečka and Hobza for several H-bonded and π - π stacked complexes, the total MP2 and CCSD(T) interaction energies (or their respective correlation contributions) converge roughly at the same rate, while the *difference* between the two interaction energies, i.e., $\Delta\text{CCSD(T)}$, typically converges much faster.⁴⁹ Though this assertion generally seems to be true, it is frequently used despite a lack of extensive validation. Recently cases have been reported in which $\Delta\text{CCSD(T)}$ converges slowly, such as in the case of the benzene \cdots Na⁺ complex, where the $\Delta\text{CCSD(T)}$ term is -0.13 kcal/mol in aug-cc-pVDZ (aDZ) and -1.16 kcal/mol in aug-cc-pVTZ (aTZ).⁵⁰ Sometimes the use of a small basis set leads to a qualitatively wrong description, such as in the case of the benzene \cdots water complex, 0.04 kcal/mol in aDZ and -0.31 kcal/mol in aTZ.⁵⁰ Our recent series of benchmark calculations on several stacked π - π systems revealed that the use of small basis sets (e.g., 6-31G*(0.25) or similar ones) can lead to an underestimation of the $\Delta\text{CCSD(T)}$ term by no more than 10–20% and the use of aug-cc-pVTZ basis set leads to 1–3% agreement with the CBS value.^{48,52} Further savings can be achieved by not adding the diffuse functions on hydrogen atoms, leading to the so-called "heavy augmented" correlation-consistent basis sets (haug-cc-pVxZ; x = D, T, Q), as proposed by Elsohly et al.⁵³ and Sherrill et al.,⁵¹ but used earlier in similar variants by other groups as well. The performance of these basis sets, not only in terms of $\Delta\text{CCSD(T)}$, but also at the MP2 level, is excellent, leading to errors that are typically about a few hundredths of a kilocalorie per mole, compared to those calculated using diffuse functions on hydrogen atoms. These differences with respect to the full aug-cc-pVxZ basis sets are almost completely diminished at the extrapolated level (see further text).

The fast rate of convergence of the $\Delta\text{CCSD(T)}$ term is the result of error compensation between the convergence of ΔCCSD and (T). As an example, the differences in ΔCCSD and (T) from aDZ to aTZ for the stacked uracil dimer are about 0.2 and -0.15 kcal/mol, respectively, while the difference in $\Delta\text{CCSD(T)}$ is only 0.03 kcal/mol.⁴⁸ Similarly, for the methyladenine \cdots methylthymine complex, the differences in ΔCCSD and (T) for the same basis sets are 0.45 and -0.36 kcal/mol, respectively, while the change

of $\Delta\text{CCSD(T)}$ is only 0.09 kcal/mol.⁵² Since the signs of the ΔCCSD and (T) corrections are conserved for all complexes known to us (as discussed in section 3.1.2), at least some compensation of errors between these terms for various basis sets and complexes can be expected.

It is important to realize that, at subchemical accuracy, correlation effects beyond those described by CCSD(T) and errors resulting from other approximations utilized within the CC approach might be of similar magnitude. It is thus meaningless to improve errors associated with either basis set incompleteness or higher order correlation without addressing the other problem as well. The conclusions in this section were drawn mostly from calculations on molecules with tens of atoms (benzene dimer, uracil dimer, etc.), where the basis set saturation is enhanced by overlap of basis functions of neighboring atoms. Slower rates of convergence toward the CBS can be expected in rare-gas dimers, or rather small molecules, and inclusion of basis functions with higher angular momenta might be necessary. The situation for larger (about 100 atoms or more) “compact” complexes has not, as of yet, been sufficiently described.

The importance of the $\Delta\text{CCSD(T)}$ term, as discussed in the previous paragraph, is different for H-bonded, $\pi-\pi$ stacked, and other types of noncovalent interactions. Basis set saturation in calculations of the $\Delta\text{CCSD(T)}$ term for systems dominated by dispersion interactions is important mostly for quantitative agreement, while for the H-bonded complexes or, in general, for complexes clearly dominated by electrostatic interactions (e.g., the $\text{Na}^+\cdots\text{water}$ complex), neglect of $\Delta\text{CCSD(T)}$ can sometimes lead to a more balanced treatment.⁵⁰

3.1.4. Approximate CCSD(T) Schemes

To date the majority of efforts to reduce the scaling of the correlated WFT calculations have been invested in methods with lower order scaling than CCSD(T) , mostly on HF and MP2 (see further text). Still, attempts to exploit the inherent short-range nature of electron–electron correlation in CC theory have been pursued by Werner and co-workers,^{54–57} Dolg and co-workers,^{58–60} Auer and Noojien,⁶¹ Scuseria and Ayala,⁶² Bartlett and co-workers,^{63–65} Head-Gordon and co-workers,^{66–68} and others.^{69–73}

Though some of these approaches have been quite successful, judging by their abilities to reproduce absolute correlation energies, reaction energies, or conformation energies, compared to their corresponding canonical calculations, much less is known about the accuracies of these methods for calculations on noncovalent interactions. The explicit CC treatment of only strictly spatially close localized orbitals (or domains, fragments, etc.) seems to lead to inaccuracies in descriptions of such subtle effects as the dispersion energy. Surprisingly, as pointed out by Hughes et al. and Maslen et al.,^{65,66} the more computationally demanding connected triple excitation tends to be even less local than the double excitation. Currently, two routes to linear scaling seem to be possible, one advocated by Werner and co-workers and by Dolg and co-workers (see references above), which is simply to treat only the strongly interacting region explicitly, and the other suggested by Hughes et al., which is to calculate dispersion coefficients locally and use them to construct the intermolecular interaction potential.⁶⁵ The first of these routes clearly has the capability to converge to a linear-scaling regime, but the dimension of explicitly treated regions (basically all intermolecular orbital pairs of

interacting fragments) might be computationally prohibitive.^{74,75} Additionally, in the incremental scheme of Dolg and co-workers, control over the accuracy appears to be an appealing feature. One of the positive aspects of using local correlation approaches is a decrease in the effects of basis set superposition error, both intermolecular^{76,77} and intramolecular.^{78,79} In the approach suggested by Hughes et al., a rigorous ab initio quantum chemical treatment would have to be sacrificed.⁶⁵ Rather encouraging results have come from the combination of the localized orbitals and explicit correlation (see further text) approaches by Werner,^{54,80} where the authors demonstrate an enhancement of accuracy via the symbiosis of these two approaches, rather than an accumulation of errors.

A different route to approximated coupled-cluster methods for large systems is represented by the fragment molecular orbital (FMO)⁷¹ based CCSD(T) and divide-and-conquer (DC)⁸¹ based CCSD methods. Common to these (and other related) approaches is the fact that it is unnecessary to use localized orbitals. Near linear scaling is achieved by either explicit calculation of only the (capped or isolated) monomer fragments and their respective dimers and trimers in the electrostatic field of the whole system (FMO) or by assembling the total one-electron density, which is calculated locally (DC). Judging according to a rather limited number of applications that have been published (the closest related to noncovalent interactions being calculations of the relative stability of various complexes), these methods are capable of integrating the coupled-cluster concept into calculations of hundreds of atoms with errors similar to those of local CC methods.

3.2. MP2 in the Complete Basis Set

3.2.1. Computational Demands and Applicability

Supermolecular second-order Møller–Plesset perturbation theory (MP2) has for a long time been, and still is, one of the most frequently applied WFT methods for calculations on noncovalent interactions. The reason for its relative popularity compared to other correlated methods results from its low computational cost and especially its high performance/cost ratio. Owing to the, now routinely used, density-fitting procedure^{82–87} (DF, also called the resolution-of-the-identity, RI) or the Cholesky decomposition (CD) of the two-electron integrals,^{88–93} MP2 can be applied to molecules with tens of (second-row and hydrogen) atoms in large basis sets (augmented triple- ζ or better) or up to hundreds of atoms in small basis sets (Pople’s 6-31G*, for example). In these applications MP2 is often the least expensive step, with the preceding HF (SCF) step being more costly, even though the asymptotic scaling of MP2 with the system size is higher by an order of magnitude.

Being the crudest approximation to the electron correlation problem, MP2 has its limitations in accuracy, which have now been quite well mapped after numerous “benchmark” studies, and the origins of these limitations are now, to a large extent, understood.^{94–98}

A notorious example of MP2’s “failure” is the strong overestimation of $\pi-\pi$ stacking interactions in complexes, such as the benzene dimer (overestimation by almost 80% in the parallel-displaced conformation).^{99–102} Deeper insight into the origins of this behavior can be acquired from analysis using the perturbation theory of molecular interactions (see, for instance, the review paper of Chałasiński and Szczyński¹⁰³). Among all the contributions obtained at the

supermolecular MP2 level (second-order electrostatic, dispersion, deformation, and exchange correlation), it is the dispersion contribution that is approximated with the least accuracy for these “problematic” complexes. The so-called uncoupled Hartree–Fock (UCHF) dispersion energy contained at the supermolecular MP2 level completely lacks the intramonomer correlation correction for these systems, thus overestimating binding energies by as much as 20%.^{97,98} Interaction energy components other than dispersion are satisfactorily described at the MP2 level, provided that the description of the electric properties of the monomers is also reasonable. Thus, for highly correlated systems (especially nonpolar ones with delocalized π -electrons) MP2 can be expected to perform poorly.

Another major drawback of the MP2 method is its strong basis set dependence, as discussed in section 3.1.3. For the class of strongly correlated, dispersion-dominated systems there is surprisingly close agreement of the MP2 results with benchmark data, provided that MP2 is used along with small basis sets (for instance, 6-31G*(0.25) and aug-cc-pVDZ) or with basis sets lacking diffuse functions (for instance, cc-pVTZ).¹⁰⁴ This is, however, a consequence of error compensation involving overestimation due to the inaccurate UCHF dispersion description and underestimation attributable to deficient basis sets. Unfortunately, the high accuracy of this error compensation cannot generally be relied upon. The convergence of MP2 toward the CBS limit, either by extrapolation or by introducing explicit correlation, is discussed in section 3.4.

3.2.2. Approximate MP2 Schemes

As stated above, the implementation of the DF/CD approaches represents a major breakthrough in the applicability of the MP2 method for treating medium to large systems. Unlike other approximations to be discussed, DF and CD have worked their way to becoming true “black-box” approximations, and are even accurate enough for benchmark quality applications.^{51,85,105–107}

Obviously, to really profit from the speedup of DF/CD-MP2, two-electron integral decomposition, typically leading to a 1 order of magnitude speedup in MP2, must be applied at the HF step. For noncovalent complexes with sizes similar to those found within the S22 or JSCH2005 databases (see below), errors are generally no higher than a few hundredths of a kilocalories per mole.^{51,106} In the case of DF, accuracies depend on the quality of the underlying (auxiliary) density fitting basis set. For the CD-based approaches, the quality of results is fully under control via the CD threshold, for complexes being of DF accuracy in the range of 10^{-4} to 10^{-5} au.^{42,48,52} Additional speedup of MP2 calculations can be achieved by employing the energy denominator decomposition, via either Laplace transform^{108–110} or CD,^{31,111} again with controlled accuracy.

The application of MP2 to large-scale calculation of interaction energies, with additional approximations, can be approached by several routes. Despite the fundamental limitations of spatial restrictions on excitations within the localized orbital approach, as discussed in section 3.1.4, the number of applications of LMP2 is increasing. With the combination of DF/RI/CD for two-electron integrals and energy denominator decomposition, these methods almost approach the computational efficiency of DFT methods; however, the drawbacks of the MP2 description are still retained.^{110,112} Even the partial use of electron correlation

locality in the so-called “triatomics-in-molecules” MP2 method (TRIM-MP2),^{113,114} in combination with DF, leads to a very computationally efficient, though not linear scaling, approach, with several advantages over LMP2 methods, such as continuous potential energy surfaces (PESs), etc. Approximate MP2 approaches based on molecule/complex fragmentation appear to be applicable for some of the largest complexes that can be treated using any existing computational method. Furthermore, according to Gordon et al., the combination of FMO (or another fragmentation method, such as the systematic fragmentation method (SFM),¹¹⁵ with the semiclassical effective fragment potential (EFT2; see below) retains ~ 1 kcal/mol agreement with the full MP2 calculation for just a fraction of the computational cost.^{116,117}

3.2.3. Empirical Corrections to the MP2

3.2.3.1. Spin-Component-Scaled MP2 Methods. In the pioneering work by Grimme,¹¹⁸ the idea was introduced of separately scaling the same (also referred to as the singlet or antiparallel) and opposite (or triplet or parallel) spin components of the correlation energy (also extendable to other WFT methods, if applicable) to enhance accuracy:

$$E^{\text{corr}}_{\text{SCS}} = p_{\text{S}}E^{\text{corr}}_{\text{S}} + p_{\text{T}}E^{\text{corr}}_{\text{T}} \quad (2)$$

The theoretical reasoning for the various scalings of these quantities is based on the so-called deficient short-range (dynamical) correlation, which is increased by the p_{S} parameter (p_{S} being >1), and overestimation of the long-range (nondynamical) correlation, which is damped by the p_{T} parameter (p_{T} being <1).¹¹⁹ The particular values of the scaling parameters in the original Grimme SCS-MP2 method were either deduced from theory, $p_{\text{S}} = 6/5$, or fitted to a reference QCISD(T) data set of total correlation energies, reaction and atomization energies, etc., $p_{\text{T}} = 1/3$.

SCS-MP2 has proven to be quite an improvement over MP2 in most cases. Even though the noncovalent complexes were not considered in its development, SCS-MP2 turned out to be a step in the right direction for treating the outstanding problem of the MP2 overestimation of stacked π – π complexes. However, more detailed investigations of the performance of SCS-MP2 for a wider range of noncovalent complexes indicate that clear quantitative improvement over MP2 is obtained almost exclusively for (especially nonpolar or slightly polar) stacked π – π complexes.^{74,75,120–124}

Even for stacked DNA base pairs, the overall performance of MP2 is slightly better than that of SCS-MP2.^{51,74} A fortuitous compensation of errors upon neglect of CP corrections was reported by Antony and Grimme, which makes CP-uncorrected SCS-MP2 with basis sets of triple- ξ quality more accurate than MP2, due to error compensation.¹²⁰

Hill and Platts,⁷⁴ and a year later DiStasio and Head-Gordon,¹²² came up with an idea to optimize the spin scaling coefficients against benchmark interaction energy data. Hill’s SCSN-MP2/aug-cc-pVTZ method, optimized for the best performance on DNA/RNA base pairs, utilizes only the same-spin component, p_{T} (having a value of 1.76). A similar value was also obtained when an independent training set, S22, was used by DiStasio and Head-Gordon. In their SSS(MI)-MP2 method the parameters depend on the basis set being used, with a p_{T} value of 2.54 for the (small) cc-pVDZ basis set, 1.90 for cc-pVTZ, and 1.75 for cc-pVQZ (the largest basis set considered). This last value is almost

the same as the one used by Hill and Platts, since cc-pVQZ is of similar quality for noncovalent interactions compared to aug-cc-pVTZ. DiStasio and Head-Gordon also proposed the SCS(MI)-MP2 method, which uses both scaled spin components and yields the best performance for intermolecular interactions. The increase in accuracy seen with this variant of the spin-scaled MP2 method is significant, comparing it to both MP2 and the original SCS-MP2. For the S22 benchmark data set (see further text) the rmsd (root mean square deviation) of SCS(MI)-MP2/cc-pVTZ binding energies is 0.31 kcal/mol, while for MP2 and SCS-MP2 with the same basis set these values are 0.99 and 1.45 kcal/mol, respectively. Even more pronounced is the maximum error decrease from 2.45 kcal/mol (MP2) and 2.83 kcal/mol (SCS-MP2) to 0.70 kcal/mol for SCS(MI)-MP2 with the cc-pVTZ basis set. SCS(MI)-MP2 accuracies for H-bonded complexes are still not significantly better than those of MP2, but are clearly improved over those of SCS-MP2. The following question still remains: To what extent are these optimized spin-component scaling coefficients transferable among a wider range of interaction types and, especially, for increasing system sizes and larger numbers of interacting fragments? Bachorz et al. plotted the rms deviation of SCS-MP2 with respect to the reference CCSD(T) data versus the two scaling coefficients for several stacked and H-bonded complexes of 2-pyridone and *n*-fluorobenzene ($n = 1-6$).¹²¹ They observed a very narrow range of variations of the spin-scaling coefficients (being close to those from SCSN-MP2), for which the rms values were within ~ 1 kcal/mol, having very steep valley walls. King found, while investigating the PES of the ethylene dimer, that MP2 is superior to the original (CP corrected) SCS-MP2 method.¹²³ However, refitting spin-scaling coefficients for several important ethylene dimer configurations revealed an insensitivity of the parameters to the geometry of the complex (p_S being between 1.25 and 1.30 and p_T being between 0.42 and 0.61). The SCS-MP2 method, customized for this particular complex (including CP corrections), then clearly outperformed both MP2 and Grimme's original SCS-MP2.

Attempts to reduce the computational scaling of spin-scaled MP2 methods by complete neglect of the same-spin part, leading to the so-called SOS-MP2 method^{109,125} and its variants, MOS-MP2¹²⁶ and SOS(MI)-MP2,¹²² inevitably lead to increases in errors that cannot be recovered by reparameterization. As demonstrated by Lochan et al. for rare-gas dimers, the same-spin component can sometimes be even larger than the opposite-spin one.¹²⁶ The MOS-MP2 variant proposed by Lochan et al. aims to remedy SOS-MP2's problem of wrong asymptotic scaling, where the value of the scaling parameter in the asymptotic limit should be 2 to recover the MP2 energy. This is achieved by introducing an interelectron distance-dependent function incorporating one empirical parameter, which in short-range regions (less than 2 Å) roughly recovers SOS-MP2 values, while recovering the correct asymptotics of MP2 in the long-range regions. Neglecting the same-spin component, along with the Laplace¹⁰⁹ or Cholesky¹²⁵ decomposition of the energy denominators, opens a possibility for the N^4 factorization of SOS-MP2, which, considering its scaling, might still be an interesting tool for large-scale applications, although with considerably less accurate performance.

Further increasing the number of fitting parameters to 4, as in the so-called "dispersion-weighted" MP2 (DW-MP2) proposed by Marchetti and Werner,¹²⁷ presents another route to

improving the deficiency of Grimme's original SCS-MP2 method. The authors recognized that the ratio between the HF and the MP2-F12 interaction energies for different types of noncovalent complexes (such as dispersion-bound, H-bonded, etc.) reflects the nature of this interaction (HF being repulsive for dispersion-bound complexes, while attractive for electrostatics-dominated complexes). This ratio is further correlated via a function of two parameters to a weight factor mixing the uncorrected MP2 interaction energy (considered sufficiently accurate for the H-bonded complexes) with Grimme's SCS-MP2 one (considered sufficiently accurate for the dispersion-bound complexes). The two parameters entering the weight factor were optimized against the CCSD(T*)-F12 reference data for the S22 database, presented in the same paper.¹²⁸ As demonstrated by statistical analysis based on rms, MAD, and MAX, the DW-MP2-F12 method, used with the aug-cc-pVDZ basis set, is even slightly more accurate than the two-parameter-based SCS(MI)-MP2 method extrapolated from the cc-pVTZ to the cc-pVQZ basis set.

3.2.3.2. MP2 Schemes with Improved Description of the Dispersion Interaction. Recognizing the general feature of MP2, i.e., the overestimation of stacked $\pi-\pi$ interaction and fairly good performance on H-bonded complexes, Cybulski and Lytle were led to deeper investigation of the reasons for such behavior.⁹⁷ Motivated by the excellent agreement of the DFT-SAPT results of Hesselmann et al.^{129,130} with benchmark CCSD(T) data for DNA base pairs,¹³¹ the authors focused on the possibility to improve the MP2 description of the dispersion interaction. As realized earlier,⁹⁴⁻⁹⁷ the MP2 supermolecular interaction energy contains the dispersion energy only at the uncoupled HF (UCHF) level, which differs from the coupled one by 10–20%.¹²⁹ The idea of Cybulski and Lytle was to substitute this UCHF dispersion energy in the supermolecular MP2 interaction energy by (among several options) the scaled coupled HF (from time-dependent HF (TDHF) calculation) one:

$$\Delta E = \Delta E^{\text{MP2}} - E_{\text{disp}}^{(2)}(\text{UCHF}) + E_{\text{disp}}^{(2)}(\text{M}) \quad (3)$$

The scaling factor was chosen to be the ratio between the C_6 coefficients obtained using the CCSD(T) and TDHF methods. $E_{\text{disp}}^{(2)}(\text{M})$ from eq 3 can be taken from other methods as well, as shown in an even more elaborate work by Hesselmann.⁹⁸ In this work the author tested not only the scaled TDHF dispersion energy, but also the (unscaled) coupled dispersion energy from two time-dependent DFT (TDDFT) methods, i.e., LHF/xALDA (localized HF method by Della Sala and Görling)¹³² combined with the exchange-correlation kernel from the adiabatic local density approximation (ALDA) in its exchange-only variant and PBE0AC/hybALDA (asymptotically corrected PBE0AC with the hybrid nonlocal exchange ALDA exchange-correlation kernel). As demonstrated on potential energy curves of rare-gas dimers and several H-bonded structures of the water dimer and hydrogen fluoride dimer, LHF/xALDA, though being an exchange-only approximation for calculation of the dispersion energy, outperforms all other approaches, leading to close agreement with CCSD(T) benchmark results. Further testing was performed by Pitoňák and Hesselmann (the method was, for simplicity, named MP2C, where the "C" stands for "corrected"), which confirmed the excellent accuracy also for more extended systems from the S22 data set and others.¹³³ Another important advantage of this

method, compared to the (scaled) TDHF approach, is its N^4 (N being the size of the monomer(s)) scaling, which leads to only marginal overhead to the underlying supermolecular MP2 calculation.

A different route to correcting the MP2 dispersion is followed by Tkatchenko et al.,¹³⁴ in this study the authors applied a methodology similar to that in empirical-dispersion-corrected DFT approaches (such as that of Grimme¹³⁵ or Jurečka et al.¹³⁶). An accurate interatomic potential is obtained by correcting the MP2 one via

$$V(R) = V^{\text{MP2}}(R) + F(R) \sum_{j=0} C_{2j+6} / R^{2j+6} \quad (4)$$

where ΔC_{2j+6} are the differences between C^{MP2}_{2j+6} dispersion coefficients and their accurate reference values. $F(R)$ is an exponential damping function damping the London dispersion expansion, which is equal to unity at DR_{eq} , where D is the distance damping parameter and R_{eq} is the van der Waals equilibrium distance. MP2 potentials corrected according to eq 4 represent a significant improvement with respect to the reference CCSD(T) potentials for several systems in terms of both equilibrium and stretched geometries.

3.3. Methods Complete to the Third Order of Perturbation Theory: From MP3 to CCSD

3.3.1. Computational Demands and Applicability

Computational methods correct through the third order of perturbation theory have been extensively applied to calculations of interaction energies, and their accuracies are more or less well characterized. These methods undoubtedly cover correlation effects more completely than MP2; however, their performance on certain types of noncovalent interactions (especially π - π stacking) is still unsatisfactory (see section 3.3.2). It is also important to note that, for calculation of many-body effects, at least third-order perturbation theory methods must be used to cover the nonadditive dispersion correction. Due to the N^6 scaling for iterative (from L-CCD to CCSD) and noniterative (MP3) third-order methods with the system size, the area of applicability is significantly narrowed compared to that of MP2. "Standard" approximations used in MP2, such as DF or CD of two-electron integrals, can certainly be applied, but mainly to alleviate the storage bottleneck (four virtual orbital index integrals are required). Unfortunately, due to the coupling of biexcitation amplitudes (missing in MP2), no simplified, or more efficient, factorizations, such as in MP2, are possible. Still, applications of DF/CD-based MP3 on complexes of about 100 second-row elements and hydrogen (~ 300 correlated electrons) and up to 2000 basis functions can nowadays be done routinely, mainly owing to very efficient parallelization. Similar efficiency can be obtained with iterative third-order approximate CEPA-like and CCSD methods, especially when exploiting locality and virtual-orbital space truncation schemes.³³ For complexes for which satisfactory performance of the third-order methods can be expected, more than an order of magnitude savings in computational times can be achieved compared to CCSD(T).

3.3.2. Accuracy of the Supermolecular Third-Order Methods

In terms of perturbation theory for molecular interactions, the supermolecular MP3 method provides higher order

(intramolecular) correlation corrections to electrostatics, exchange, induction, and exchange-deformation terms than MP2, and generally also gives a better description of the problematic dispersion energy. The dispersion energy is covered at both the UCHF ($\epsilon_{\text{disp}}^{(30)}$) level and the (in the case of two-body interactions, dominant) intramonomer correction to the second-order UCHF dispersion energy ($\epsilon_{\text{disp}}^{(21)}$). More sophisticated third-order methods, such as CCSD and MP4(SDQ), cover dispersion even more completely via the second-order intramonomer correlation correction, $\epsilon_{\text{disp}}^{(22)}$, but without the effect of connected triple excitations, which is crucial for certain types of interactions, as discussed further below. For complexes in which electrostatic interactions are dominant (such as in H-bonding), complexes dominated by the dispersion energy not involving delocalized π -electrons, and complexes of strongly polarizable monomer(s), the performance of the third-order methods can be expected to be satisfactory. It is important to stress that the performance of all third-order methods, from MP3 to CCSD (through all approximations in the cluster expansion), is comparable, except for complexes where the electric properties (dipole moments, polarizabilities, etc.) of the monomers exhibit strong dependence on the level of intramolecular correlation. For the most problematic cases, such as π - π stacking interactions, all the third-order methods, unlike MP2, strongly underestimate the interaction. As an example, the interaction for the benzene dimer in its parallel-displaced conformation is, according to Tsuzuki et al., repulsive by 0.02 kcal/mol for MP3 and bound only by -0.05 kcal/mol for MP4(SDQ) and -0.02 kcal/mol for CCSD, as calculated using the aug(d)-6-311G* basis set, whereas the "benchmark" CCSD(T) interaction energy is -1.02 kcal/mol (MP4(SDTQ) being -1.69 kcal/mol).^{15,102,137} Similarly, for the argon dimer using the aug-cc-pV6Z basis set, the well depth of the dissociation potential for MP3 and CCSD is 375.4 and 324.7 μE_{h} , respectively, while the MP4 (including the triple excitations) and CCSD(T) well depths are 457.3 and 441.3 μE_{h} , respectively (the MP2 well depth being overestimated, 515.4 μE_{h}).⁴⁵ Depending on the electric properties of the monomers, the MP3 third-order correction can be attractive (though usually at the same time small for polar systems), but for less polar and especially dispersion-dominated complexes it is repulsive and in good agreement with MP4(SDQ) or CCSD corrections. The values for the latter corrections are repulsive for all complexes in equilibrium geometries investigated so far.

What is missing in the third-order methods is the fourth-order effect of the connected triples, either from MP4 or, preferably, calculated from the converged excitation amplitudes from the CCSD.¹³⁸ This correction is attractive for all complexes in equilibrium geometries investigated so far, and its magnitude depends on the character of the complex. The effect of triples can range from as much as 50% of the third-order correction (several kilocalories per mole) for the π - π stacked complexes to almost zero for electrostatics-dominated complexes, such as in "typical" H-bonding.

3.3.3. Empirical Third-Order Schemes

By analyzing the behavior of the third-order correction obtained at the supermolecular MP3 level, we can recognize typical patterns for various types of interactions. According to these observations, an approximate N^6 scaling method for calculation of noncovalent interactions can be formulated. As already mentioned, electrostatics-dominated complexes

such as H-bonded ones are already well described at the MP2 level. Further increasing the quality of the description of electron correlation by, for instance, going to the CCSD(T) method yields additional stabilization typically of a few percent. Even though MP3 is not the best suited method for these types of complexes (since the intramonomer correlation corrections to the electric properties of the monomer play a crucial role), due to the relatively small value of the third-order correction term, the overall accuracy is still a slight improvement over MP2.

The situation is quite different for the π - π stacking complexes, where both MP2 (overestimated) and MP3 (underestimated) deviate from accurate CCSD(T) results by almost the same amount, but in opposite "directions" (i.e., having opposing signs).¹³⁹ According to this empirical observation, and because the fourth-order effect of triples is always attractive and its value is roughly 50% of the MP3 third-order correction, we can propose the "MP2.5" method.¹³⁹ MP2.5 corrects the MP2 interaction energy by a scaled (the scaling factor being 0.5) third-order correlation correction. Though this is purely an empirical approximation to the CCSD(T) interaction energy, its performance on various types of complexes has been recognized as being quite accurate. As for other empirical approaches, such as the SCS-MP2 method, the proper asymptotic scaling of both MP2 and MP3 is retained and the performance of MP2.5 is in principle never considerably worse than that of MP2, but can be significantly better. For complexes for which the third-order correction presents an important part of the interaction energy (in terms of percentage-wise contributions), the accuracy-determining point is the selection of the scaling factor. The optimum value of the scaling parameter has been found to be between 0.4 and 0.6, depending mostly on the polarities of the monomers and, to some extent, on the underlying AO basis set. This offers the possibility to estimate the error bars for the MP2.5 method as being roughly $\pm 10\%$ of the third-order correlation contribution.⁵² Another advantage of using the scaled third-order contribution is that its basis set dependence is very similar to that of Δ CCSD(T), which allows us to use incremental schemes such as the ones used for estimated CCSD(T)/CBS values; see eq 1.¹⁴⁰

Analogous to the SCS-MP2 idea, the scaling of different spin energy components of the CCSD correlation energy was proposed by Takatani et al.¹⁴¹ Using the same training set for optimization of the scaling parameter that was used by Grimme for SCS-MP2,¹¹⁸ significant improvements in the accuracies of SCS-CCSD values (over those of CCSD), especially for noncovalent interactions, were achieved. Further testing of SCS-CCSD on cuts through the PES of several noncovalent dispersion-bound complexes shows excellent accuracy, with errors of less than 0.2 kcal/mol (in comparison with CCSD(T)/CBS results) for several configurations of the benzene dimer, whereas errors for the SCS-MP2 variants approach almost 0.5 (SCS(MI)-MP2, SCSN-MP2) to 1.0 (SCS-MP2) kcal/mol.

Further, broader assessments of the performance of the SCS-CCSD method revealed that there is still some space for improvements in its application for noncovalent interactions. Reparametrization of SCS-CCSD toward the best performance on the S22 test set was carried out in our laboratory.¹⁴² This reparametrization is analogous to the one carried out for the SCS-MP2 method by Distasio et al., leading to the SCS(MI)-MP2 method.¹²² An appreciable improvement in the mean

unsigned (maximum) error compared to the original SCS-CCSD for the S22 set, from 0.21 (0.79) to 0.05 (0.10) kcal/mol, was obtained. Interestingly, a similar "flip" of values of the opposite- and same-spin scaling coefficients found in the case of SCS-MP2 (p_S and p_T being 1.33 and 0.33) to SCS(MI)-MP2 (p_S and p_T for cc-pV(TQ)Z being 0.40 and 1.29) was obtained. The new scaling factors for the, so-called, SCS(MI)-CCSD method (i.e., SCS-CCSD tailored toward the calculation of molecular interactions) were 1.11 (p_S) and 1.28 (p_T), i.e., in the swapped order compared to the p_S and p_T values of 1.27 and 1.13 in the original SCS-CCSD. The use of SCS- or SCS(MI)-CCSD is a very appealing option, but the iterative N^6 scaling with system size (exactly the same as for the standard CCSD), storage of the biexcitation amplitudes, and far less efficient parallelization significantly narrow the applicability of this method compared to, for instance, MP2.5. An attractive feature of the SCS- and SCS(MI)-CCSD methods is the fact that they are not based on perturbation theory.

A new approximate general-purpose family of methods formally derived from the N^6 coupled-pair approximation (CEPA/1) has recently been introduced by Neese and co-workers.^{33,143,144} Combining several approximations, such as orbital localization and truncation of the external space of modified pair-natural orbitals (i.e., LPNO-CEPA/1), the steepest scaling parts of the algorithm scale as N^5 (with a small prefactor). This very efficient algorithm (and implementation) promises applicability to systems with up to 100 atoms and about 2000 basis functions. The accuracy in the description of noncovalent interactions, as demonstrated on H-bonded (neutral and charged) complexes from the test set of Boese et al.,⁴⁶ seems to be excellent (with a mean absolute deviation (MAD) of ~ 0.2 kcal/mol).⁴⁶ The method's observed accuracy for 10 selected dispersion-bound and "mixed" complexes from the S22 database (see later) is considerably worse, MAD being ~ 0.6 for CEPA/1 (and NCPF/1) and ~ 0.3 kcal/mol for LPNO-CEPA/1, with the maximum errors being 1.40 for CEPA/1 (and NCPF/1), but only 0.58 kcal/mol for LPNO-CEPA/1. Since LPNO-CEPA/1 is an approximation to CEPA/1, some fortuitous error cancellation must be taking place, as the LPNO-CEPA/1 results are in significantly better agreement with the reference values. In a similar spirit, recently the LPNO approximations to QCISD and CCSD have been implemented in the same group.¹⁴⁵ The performance of the methods on noncovalent interactions given in the same set of 10 selected complexes from S22 revealed a slightly worse performance of the LPNO approximation than given with CEPA/1, with MAD (with respect to the reference QCISD/CCSD) being about 0.7 kcal/mol and with a maximum error of 1.2 kcal/mol (LPNO-CCSD). Considering the "orientation" of the error, for instance, for the benzene dimer in parallel-displaced conformation, errors of CCSD and LPNO approximation with respect to CCSD(T) seem to be accumulating. Furthermore, the LPNO approximation is not a "true" linear scaling local approximation (in the spirit of the work of Werner et al., Schütz et al., etc. cited above); according to results presented in ref 145, the BSSE error is not reduced compared to the full QCISD/CCSD calculation.

3.4. Increasing the Rate of Convergence of Interaction Energy Calculations toward the Complete Basis Set Limit

The systematic convergence of the (correlation) energy calculated in a series of AO basis sets with increasing size was already recognized in the 1960s (see the overview in ref 50 and references therein). The most widely used approaches for basis set extrapolation are those proposed by Helgaker et al.^{146,147} and by Truhlar.¹⁴⁸ All extrapolation techniques rely on the idea of exponential dependence of the energy on the maximum angular momentum number and its polynomial simplification:

$$E_X = E_{\text{CBS}} + B/X^\beta \quad (5)$$

where the coefficients E_{CBS} and B result from the (usually least-squares) fit of energies from a series of basis sets and parameters, X and β , which differ for each particular approximation ($X = 2, 3, 4, \dots$ for the (aug-)cc-pVxZ basis sets, β being 3 in Helgaker's scheme for the correlation energy, but 2.2 for the MP2 and 2.4 for the CCSD/CCSD(T) correlation energy in Truhlar's scheme). The Hartree–Fock energy is extrapolated separately in both schemes. While in Truhlar's scheme the same equation is used, but with the β coefficient being 3.4, in Helgaker's scheme the Hartree–Fock energy scales exponentially:

$$E_X^{\text{HF}} = E_{\text{CBS}}^{\text{HF}} + B \exp(-\beta X) \quad (6)$$

where β equals 1.54.

A rather different scheme, targeted toward the calculation of interaction energies, utilizing an almost arbitrary series of basis sets, was proposed by Kim et al.^{50,149,150} Unlike other extrapolation methods, this scheme takes advantage of the fact that the series of both BSSE-corrected and -uncorrected interaction energies converge to the same limit, i.e., the CBS. In its simplest form, for energies obtained in a sequence of basis sets with cardinal number X

$$\Delta E_{\text{CBS}} = \frac{1}{2}(\delta_X \varepsilon_{X+1} - \delta_{X+1} \varepsilon_X) / (\delta_X - \delta_{X+1}) \quad (7)$$

where δ_X and ε_X (and δ_{X+1} and ε_{X+1} analogously) are defined as

$$\delta_X = \Delta E_X^{\text{b}} - \Delta E_X^{\text{n}} \quad \varepsilon_X = \Delta E_X^{\text{b}} + \Delta E_X^{\text{n}} \quad (8)$$

where ΔE_X^{b} and ΔE_X^{n} are BSSE-corrected and -uncorrected interaction energies in a basis set of cardinality X . Contrary to the Helgaker-type extrapolation, which is based on systematically improved AO basis sets (such as, e.g., Dunning's cc-pVxZ or aug-cc-pVxZ sets), the Kim extrapolation can be based on different types of AO basis sets.

Assessment of the accuracy of all the numerous types of CBS extrapolation methods is almost impossible, even if we limit our attention to the largest systems currently treatable at the CCSD(T) level (i.e., about 30 second-row atoms and hydrogen). First, none of the schemes described or cited above are "exact" in the form that they would be derived for systems of interest here. Thus, only a statistical evaluation of these schemes would make sense; there is not, however, enough data available to carry out such an analysis. Furthermore, the "performance" of a basis set significantly depends on the method applied (i.e., HF, MP2, $\Delta\text{CCSD(T)}$, ...) and the system's size and geometry, due to the overlap of basis functions, as

discussed above. Thus, for instance, the aug-cc-pVDZ basis set, which is usually considered to be too small for quantitative predictions, gives quite accurate results when used for HF and $\Delta\text{CCSD(T)}$ calculations, but is certainly deficient for MP2 calculations.

As already mentioned, calculation of the $\Delta\text{CCSD(T)}$ term is the rate-determining step in estimated CCSD(T)/CBS calculations. On the basis of our experience, the performance of small basis sets (such as 6-31G*(0.25), etc.) on $\Delta\text{CCSD(T)}$, though sometimes surprisingly good, is in general quite unstable, thus certainly not suitable for CBS extrapolations.^{48,52} The smallest basis sets that are generally acceptable for the calculation of this term are aug-cc-pVDZ and heavy-aug-cc-pVDZ ((h)aDZ). Accurate MP2 extrapolations of small-sized to midsized systems should be based on basis sets not smaller than aug-cc-pVTZ. Use of smaller basis sets, e.g., (h)aDZ, for these systems may spoil the extrapolation and is generally not recommended. However, for larger systems, extrapolations utilizing the (h)aDZ basis set should be quite accurate due to the intramolecular overlap of the basis functions of the investigated system. Besides, using these basis sets for larger systems is often the only option due to the prohibitive computational cost or convergence problems caused by linear dependence of the AO basis. Another option, routinely available now for MP2, its SCS variant, and the double-hybrid DFT methods by Grimme,¹⁵¹ is the so-called "dual-basis" approach, which was recently shown to work well (errors on the order of hundredths of a kilocalorie per mole) for calculations of noncovalent interactions, with computational savings of almost an order of magnitude.¹⁵²

Extrapolation of HF is usually unnecessary when using basis sets of (augmented) triple- ζ quality or higher, since it is typically already converged to the CBS (to within a hundredth of a kilocalorie per mole). Generally, the performances of various X^{-3} -based procedures and Kim's type of extrapolation are quite comparable, with results usually differing by less than 0.1 kcal/mol.

A last note to be mentioned here concerns the BSSE. The most straightforward way to alleviate artificial stabilization due to the use of unequally large basis sets for the dimer and monomers (and analogously for many-body interactions) is by applying the counterpoise (CP) method as formulated by Boys and Bernardi, i.e., using the same AO basis set for the monomers and the dimer.¹⁵³ In the formalism introduced by Xantheas, the CP-corrected interaction energy is calculated as

$$\Delta E(\text{CP-corrected}) = E^{\alpha\text{U}\beta}_{\text{AB}}(\text{AB}) - E^{\alpha\text{U}\beta}_{\text{AB}}(\text{A}) - E^{\alpha\text{U}\beta}_{\text{AB}}(\text{B}) \quad (9)$$

where, for instance, $E^{\alpha\text{U}\beta}_{\text{AB}}(\text{B})$ means the energy of monomer B in the geometry of the AB complex with the unified basis set of monomers A and B.¹⁵⁴ For many-body interactions, the CP correction is slightly more complicated (see, for instance, ref 155) and will not be discussed here. Please note that in the definition of the CP-uncorrected interaction energy

$$\Delta E(\text{CP-uncorrected}) = E^{\alpha\text{U}\beta}_{\text{AB}}(\text{AB}) - E^{\alpha}_{\text{A}}(\text{A}) - E^{\beta}_{\text{B}}(\text{B}) \quad (10)$$

the energies of monomers A and B are both calculated in different basis sets, each of their "own", but also in different, so-called relaxed geometries. It is often claimed that the CP-

corrected and -uncorrected interaction energies converge to the same limit. This is, however, not always true (see refs 50 and 154 and references therein), since the CP-corrected energy lacks monomer relaxation energy. In situations where the relaxation energy is comparable to the basis set incompleteness error, using the CP correction can degenerate the results compared to uncorrected ones (for the $\text{I}^- \cdots \text{H}_2\text{O}$ complex the CP-corrected interaction energy deviates from the best reference value by 5 kcal/mol).

A remedy for this problem, which was proposed by Xantheas and also advocated elsewhere, is simply to add the missing relaxation energy of the monomers to the CP-corrected energy, $E_{\text{rel}}^\alpha(\text{A})$ and $E_{\text{rel}}^\beta(\text{B})$:^{50,154}

$$\Delta E(\text{CP-corrected}) = E_{\text{AB}}^{\alpha\text{U}\beta}(\text{AB}) - E_{\text{AB}}^{\alpha\text{U}\beta}(\text{A}) - E_{\text{AB}}^{\alpha\text{U}\beta}(\text{B}) + E_{\text{rel}}^\alpha(\text{A}) + E_{\text{rel}}^\beta(\text{B}) \quad (11)$$

The interaction energy defined according to this equation has the same CBS limit as the CP-uncorrected energy. This approach, however, has the major computation disadvantage of requiring optimization of the monomers in each of the applied basis sets if extrapolation to the CBS is desired. On the basis of the observation that the relaxation energy is the term that converges most rapidly with the basis set size, a more practical approach can be proposed. Instead of modifying the CP-corrected interaction energy, we can subtract the relaxation energy from the CP-uncorrected one, by simply calculating the modified, i.e., $\Delta E_{\text{mod}}(\text{CP-uncorrected})$:

$$\Delta E_{\text{mod}}(\text{CP-uncorrected}) = E_{\text{AB}}^{\alpha\text{U}\beta}(\text{AB}) - E_{\text{AB}}^\alpha(\text{A}) - E_{\text{AB}}^\beta(\text{B}) \quad (12)$$

This interaction energy has the same CBS limit as the CP-corrected energy in eq 9. To get the interaction energy to properly contain the relaxation of the monomers, we can a posteriori add the CBS approximation of monomer relaxation energies (calculated in a small, but reasonably selected basis set) to the CBS limit of $\Delta E(\text{CP-corrected})$ from eq 9:

$$\Delta E^{\text{CBS}} = \Delta E(\text{CP-corrected}) + E_{\text{rel}}^{\text{CBS}}(\text{A}) + E_{\text{rel}}^{\text{CBS}}(\text{B}) \quad (13)$$

A completely different route to approach the CBS interaction energies is by either utilizing the OVOS^{156,157} or FNO¹⁵⁸ technique or introducing explicit correlation into the wave function.^{159,160} The latter of these schemes is especially likely to be the future “standard” approach in high-accuracy WFT. Let us briefly discuss these approaches.

In the OVOS/FNO methods, a unitary transformation of the space of virtual orbitals (VOs) is utilized to make the expansion of correlated WFs more economical. Instead of expanding the WF into the entire space of VOs, which results as a “byproduct” of occupied orbital (OO) optimization in the Hartree–Fock SCF procedure, only a truncated subspace of VOs is used. Optimization of the VOs is carried out at the MP2 level; thus, savings can obviously be achieved for methods scaling more steeply than N^5 (e.g., MP3, CCSD, CCSD(T), etc.). The main drawback of the OVOS/FNO methods, if high accuracy is desired, is that the first sufficiently large basis set suitable for application of these techniques (for systems of about 30 atoms) is (h)TZ. On the other hand, the use of OVOS/FNO orbitals has the advantage that no CP correction of monomers is required.

The reason for this is that the contribution of VOs resulting from ghost basis functions of monomers in counterpoise-corrected calculations is almost completely absorbed into the optimized virtual space.¹⁵⁶ Several successful applications of this method have demonstrated that the use of OVOS/FNO allows for calculations using basis sets that are 1 cardinal number higher (or more, depending on the size of the basis set) compared to traditional correlated methods, for almost the same cost.

Recent major conceptual progress in the explicitly correlated theories allows for the application of R12/F12-MP2^{161–168} and R12/F12-CCSD(T)^{169–177} methods for much larger systems than were previously possible with these techniques. What is even more important is that now it is possible to use much smaller basis sets than before to yield similar results. Alleviation of the rather strong conditions imposed by the so-called “standard approximation”, by introducing the auxiliary basis along with the new correlation factors (F12), was the most important step in making the explicitly correlated method applicable for real-life examples. Pilot applications of these novel approaches in explicitly correlated theories show that MP2-F12 (in various approximations) is capable of delivering almost CBS quality results already with the aug-cc-pVTZ basis sets.¹⁷⁸ Similar quality results were demonstrated for CCSD(T)-F12 (in various approximations) calculations of parallel-displaced and “T-shaped” configurations of the benzene dimer, where the (h)TZ basis set gives results within ± 0.03 kcal/mol of the estimated CCSD(T)/CBS values by Janowski and Pulay, and were even obtained without using an incremental scheme, in the spirit of eq 1.²⁰

4. Density Functional Theories

4.1. Main Limitation of Density Functional Theory: Dispersion Interaction

A few years ago, density functional theory would hardly be regarded as a method for chemical-accuracy calculations of noncovalent interactions. The majority of density functionals, which were widely spread, provided rather poor results in this regard, in particular when speaking about the dispersion-dominated complexes.^{179–181} The weakly bonded structures are usually predicted to be less stable than in reality, and sometimes they cannot even be localized as a stationary point on the potential energy surface.¹⁸² However, although it has been clear for a long time that DFT is not suitable for high-accuracy calculations, only in the past few years have the deficiencies of the current DFT functionals become understood and appreciated.

It is necessary to note that density functional theory itself is capable of providing the exact solution to the Schrödinger equation, including long-range correlation—the dispersion. The root of the rather unsatisfactory description of the intermolecular interactions in DFT lies in the approximations made in the DFT functionals, not in density functional theory itself. The most widely used GGA approximation determines the exchange-correlation (XC) energy on the basis of the electron density and the reduced electron density gradient at a given grid point. Because the XC energy is calculated only from the local properties of the density, the total XC energy is also necessarily local. As a consequence, the dispersion energy, which requires a nonlocal description in the DFT context, is not explicitly covered by the GGA functionals. This also holds, with only minor differences,

for other currently used DFT functionals—LDA, hybrid functionals, and meta-GGA functionals.

The recognition of the importance of dispersion forces, along with the accumulated numerical evidence regarding the unsatisfactory performance of current functionals, fueled many efforts for improvement. A variety of methods were suggested to mitigate the inherent problem of locality in DFT, ranging from the simple, but surprisingly successful, empirical corrections to the nonlocal pseudopotentials in plane wave methods and reparametrization of current functionals and to the construction of truly nonlocal new density functionals. Many of these methods quickly became very successful, and nowadays DFT-based calculations are slowly gaining credit even in the area of accurate noncovalent interaction energy calculations. Thus, DFT theory, which used to have a rather bad reputation regarding noncovalent interactions, is slowly becoming a method of choice for routine calculations with good reliability. In the following paragraphs we will briefly review some of the main issues in the DFT treatment of the dispersion energy and the main features of the most successful solutions.

4.1.1. Locality of the Density Functionals

The word “local” in the DFT context is often used in different meanings, and this might cause confusion in the world of noncovalent interactions. For instance, many groups label the GGA functionals, such as BLYP or PW91, as nonlocal or semilocal functionals to emphasize the additional gradient corrections exceeding the LSDA approximation. However, the gradient corrections are in the mathematical sense local, because they describe electron density properties at only one place in space. In this review we prefer to use the term “local” for all LDA, GGA, and meta-GGA functionals. By contrast, hybrid functionals are nonlocal in their exchange part, because they contain some portion of explicit Hartree–Fock exchange. Therefore, B3LYP is called a nonlocal functional (it contains nonlocal exchange), but the correlation part of B3LYP is local. The exact density functional must be nonlocal in both its exchange and correlation parts. For the physically correct description of the dispersion interaction within the DFT framework, a nonlocal correlation functional is needed (nonlocality is introduced through its kernel). Hence, no GGA- or meta-GGA-based functional is sufficient. The clearest demonstration of the missing dispersion is the incorrect asymptotic behavior of such functionals. The nonpolar molecules exhibit r^{-6} decay of the interaction energy (London dispersion), while the local DFT interaction energy decays approximately exponentially. Upon the suggestion of a reviewer, we emphasize that the correct description of dispersion is connected with nonlocality only in the DFT context, not in general. For instance, in WFT dispersion does not require nonlocality in the above-described sense, but as a many-body effect, it requires a many-body treatment.

4.1.2. Terminology of Correlation Interaction in DFT

In this paragraph we will address the question of the long-range correlation energy, which is the energy component missing in current local or semilocal density functionals. It is not completely accurate to call this energy the dispersion energy. Strictly speaking, the term “dispersion” should only be used in the asymptote of large intermolecular separations, far from the region of electronic overlap, but we are very

often interested in the molecular systems at the van der Waals (vdW) distances, where the electronic overlap is still fairly large. Here, the long-range correlation contributes significantly; its contribution is only somewhat smaller than it would be according to the respective asymptotic dispersion formula. We will say that the dispersion interaction is damped (reduced) at short distances, and we will call it “damped dispersion”, especially in connection with the methods where a damping function is used to correct the asymptotic expansion. A very similar and basically equivalent name, “overlap-dispersive interaction”, is used by Grimme.¹⁸³ Both these terms contain the word “dispersion”, which we believe is appropriate, since it reflects the nature of this interaction resulting from the long-range correlation of instantaneous dipole moments of fluctuating electronic density. Another similar but not equivalent term, “medium-range correlation”, was introduced for the “dispersion-like” interactions in the overlap region by Grimme¹⁸⁴ and is frequently used by Zhao and Truhlar.¹⁸⁵ This term, however, does not cover the long-range nonoverlap part of the correlation energy. In the following, instead of “damped dispersion”, we will often use the term “dispersion” for short, keeping in mind that it is always damped at shorter intermolecular separations.

4.1.3. Perturbative vs Self-Consistent Treatment of the Dispersion Correction

The dispersion corrections to DFT can be applied in two different ways. In a perturbative, or postprocess, approach we first perform the DFT SCF calculation without the dispersion correction, and then, in the second step, the dispersion correction is calculated using the electron density from the first step. Methods that do not employ the density to calculate the dispersion interaction, such as the empirical dispersion correction, also belong to this category. This approach is usually fast and easy to implement when compared to the second approach, the self-consistent treatment of the density. In the self-consistent treatment, the electron density changes in the SCF procedure according to the dispersion potential, as it properly should. The dispersion potential is also needed for the correct calculation of molecular geometries. However, the self-consistent treatment of dispersion is more demanding, and the corresponding potential is sometimes not easily available. Fortunately, it has been shown that the self-consistent inclusion of the dispersion interaction changes the electron density only very little and, in most cases, the errors due to non-self-consistent treatment are negligible.¹⁸⁶ Therefore, the empirical and perturbative treatments of dispersion seem justified in most cases.

4.2. DFT and Empirical Dispersion Corrections

The simplest way to introduce dispersion into DFT is to add some sort of an empirical dispersion correction to the DFT results. The empirical correction is usually based on the asymptotic formulas valid for long-range interactions. DFT is known to provide a fairly good description of the short-range correlations. In the intermediate region, both DFT and empirical correction are valid to some extent. To avoid double counting at the intermediate distances and to adjust the empirical correction to a given functional, some sort of damping function is necessary. The damping function continuously switches off the dispersion from the intermedi-

ate to the short distances, and its form is usually chosen by trial and error. Experience in the past few years has shown that it is just the quality of the damping function that determines the accuracy of the DFT-D method.

The history of correcting a DFT or HF calculation by a simple empirical dispersion term goes back to Scoles et al., who in 1975 suggested that a damped dispersion term correction to an HF calculation can remarkably improve the potential energy curves of rare-gas atoms.¹⁸⁷ The combination of HF with a damped dispersion function was also studied by Tang and Toennies with emphasis put on the damping function.^{188,189} In the early 1980s Hobza et al. applied both damped and undamped atomic dispersion corrections to rather large systems, such as complexes of anesthetics^{190–192} and nucleic acid base pairs.¹⁹³ Over time, many researchers tailored accurate and specific (i.e., nontransferable) dispersion expressions for many small molecular complexes.^{194–196} Later some authors tested different damping functions and other variations of the dispersion treatment for several small molecules,^{197,198} with the conclusion that different combinations of dispersion and DFT functionals are suitable for different complexes.¹⁹⁸ At this point, a universal and transferable dispersion correction seemed to be difficult to reach.

One of the first more generally oriented approaches appeared in 2001, when Elstner and Hobza et al. added an atomic damped dispersion correction to the DFT-based semiempirical SCC-DFTB method by Elstner et al.,¹⁹⁹ which significantly improved performance for a wide range of molecules, from peptides to nucleic acids.²⁰⁰ (Now there are also other modifications of the SCC-DFT-D method available.²⁰¹) In 2004 Grimme presented a simple atomic dispersion scheme (termed DFT-D) that gives very good results across a wide range of weakly bonded complexes of different natures.¹³⁵ He showed that a simple scaling of the C_6 coefficients can be used to adapt the dispersion correction to various functionals, such as PBE or BLYP, which otherwise differ remarkably in their ability to deal with the weakly bonded complexes. In 2007 some of us suggested a different scaling (vdW radii are scaled instead of the C_6 coefficients) which removes the main disadvantage of Grimme's scheme.¹³⁶ Scaling of the vdW radii results in slightly smaller errors for the complexes in vacuum, but mainly it ensures that the C_6 coefficients are not scaled at the long range, where there is no electron overlap, which is more physically correct. Other researchers used this scaling before for special applications,^{202,203} and many use it now for general and transferable dispersion schemes.^{204–207} This might be important mainly in the condensed phase, such as in molecular crystals.²⁰⁴

The damped atomic dispersion corrections (we will call them DFT-D according to Grimme¹³⁵ from now on) by Grimme¹³⁵ and Jurecka et al.¹³⁶ provide a very accurate description of weakly bonded complexes, with results that are statistically better than those of MP2 calculations using medium and large basis sets.¹³⁶ Many authors followed this strategy and presented their own reparametrizations and modifications of DFT-D with similar results. Of special interest is the work of Tkatchenko and Scheffler,²⁰⁸ who introduced approximate dependence of the C_6 coefficients and damping radii on the actual electronic structure.

Although the above-mentioned DFT-D schemes were already a major improvement over pure DFT, they still suffered from certain systematic errors, one of them being average overestimation of the interaction energies of the

hydrogen-bonded complexes.¹³⁶ It was speculated that these errors come partially from the XC functionals (mainly the exchange part) and not from the dispersion correction.¹³⁶ Indeed, in 2006 Grimme reparametrized the B97 functional (a GGA functional) including the dispersion correction in the parametrization process.¹⁵¹ The resulting B97-D functional does not suffer from the above-mentioned systematic errors and is, in our opinion, the best DFT-D method with the empirical vdW radii and C_6 coefficients available; other methods will be discussed below in which C_6 and vdW radii are not empirical and are obtained from the DFT calculation, which sometimes perform even better. Other reparametrizations of the DFT functionals, with similarly good results, were recently presented by Chai and Head-Gordon (ω B97X-D)²⁰⁶ and others.²⁰⁹

4.2.1. Role of the Damping Function

Proper damping is a crucial part of a transferable and general DFT-D approach. At the equilibrium geometries, damping takes off roughly around 15% of the dispersion contribution in the case of dispersion-bonded complexes, but up to 90% in the case of the hydrogen-bonded complexes. This means that the balance between complexes with different interaction types is very sensitive to the extent of damping. The extent of damping is determined mainly by the distance at which damping takes effect, which is given by the vdW radii and their scaling factor (if they are scaled). Therefore, a good choice of the vdW radii and/or the radius scaling factors (s_R) is very important. In our experience, the overall behavior of damping and even the total dispersion contribution are more sensitive to the radius scaling than to the C_6 scaling (s_6), introduced by Grimme,¹³⁵ in the sense that a small variation in s_R brings larger changes than the same variation in s_6 . The shape of the damping function is also important. Damping functions that are too flat in connection with standard DFT functionals (i.e., functionals which are not reparametrized for use with the ED correction) lead to marked overestimation of the interaction in hydrogen-bonded complexes with respect to the dispersion-bonded complexes ($\Delta\Delta_{D-H}$ in ref 136). This is also the reason why a theoretically sound, but very flat, Tang and Toennies^{188,189} damping function cannot perform well in the DFT-D schemes. A steeper damping function gives overall better results, although it may not be efficient enough in avoiding double counting at intermediate distances. In this respect, a damping function that is very steep at short distances and rather flat at intermediate and long distances works much better.²⁰⁷ The above discussion may not apply to the DFT functionals reparametrized with dispersion included in the process,²¹⁰ where the functional form and the damping function steepness are interdependent.

4.2.2. Intramolecular Dispersion

The atomic dispersion scheme of course predicts some intramolecular dispersion contribution for larger molecules. There are, however, some disputes as to whether this is desirable. In the intermolecular perturbation theory one molecule is treated as a whole; there is just one molecular polarizability, and the intramolecular contribution to dispersion is by definition zero. Therefore, the intermolecular perturbation theory cannot give us any answer regarding the intramolecular dispersion. This, however, does not mean that there is no such effect. Grimme et al. showed that at the

LMP2 level the intramolecular and intermolecular long-range correlations have similar magnitudes in comparable cases.²¹¹ DFT-D calculations on larger molecules, such as polypeptides, show that if the intramolecular dispersion is not included, the energy ordering of conformers is wrong, but after inclusion of the dispersion correction it becomes much better.^{78,212–214} In some cases a PES without the intramolecular dispersion is so inaccurate that the global minimum cannot even be located.^{78,212–215} This all points to the conclusion that inclusion of the intramolecular dispersion correction is not only useful, but probably also theoretically sound.

4.2.3. Beyond the C_6 Term

It is well-known that the higher order terms in the perturbation expansion contribute quite a large portion of the total dispersion energy, often between 20% and 30%.^{216,217} Knowing this, it seems rather surprising that the most common empirical dispersion schemes perform so well. It is likely that this can be, to a large extent, explained by the empirical character of the C_6 coefficients and the damping function, which can “buffer” the errors due to the missing higher order contributions. Depending on the set of C_6 parameters and the scaling scheme chosen, the missing dispersion can be compensated for either by up-scaling the C_6 parameters or by down-scaling the vdW radii or both. Either option is undesirable, but the first one, up-scaling the C_6 parameters, is probably the least appropriate. This would affect the long-range dispersion, which might bring overestimation of the long-range attraction in the condensed phase, where the long-range contributions add up because of the increasing relative count of long-range contacts. Note that none of the current empirical dispersion methods considered the condensed phase in the fitting procedure. Inclusion of higher order dispersion coefficients does somewhat improve the overall DFT-D results.²⁰⁷ Another possible problem arises from the neglect of the dispersion anisotropy. Dispersion anisotropy may be very large in common molecules; for instance, in base pair dimers the C_6 coefficients can vary by a factor of 2–3 depending on the direction of the interaction.²¹⁸ However, it is still not clear what part of this anisotropy can be accounted for by the “atoms-in-molecules” approach used in DFT-D. Recently, Krishtal et al. suggested that, when anisotropy is taken into account, there is no need for the damping function,²¹⁹ but this conclusion was based on calculations on rather small complexes, and in our opinion, it needs more testing on larger molecules. In conclusion, the current simple dispersion corrections omit certain, perhaps important, contributions to the dispersion energy. Nonetheless, they provide dispersion energies that are quite accurate, which can be explained either by error cancellation or by the small importance of the missing components.

4.2.4. Semiempirical Dispersion Correction Based on the Exchange Hole Dipole Moment

The dispersion correction method by Becke and Johnson partially falls into the category of empirical dispersion corrections to DFT; only in this case the dispersion coefficients (C_6 , C_8 , C_{10}) are determined with major nonempirical input.^{220–225} The damping function remains empirical. In 2005 Becke and Johnson²²⁰ suggested a heuristic formula (revisited in a more rigorous fashion in 2007)²²³ to calculate

dispersion coefficients from the exchange hole dipole moment (the XDM model). In this way the C_6 coefficients of the atoms within the molecules can be calculated from the Kohn–Sham orbitals²²⁰ or only the density properties,²²¹ with the only empirical input being the polarizabilities of the free atoms. Later Johnson and Becke suggested a damping function for use in the overlap distances²²⁴ and extended the formalism to the higher dispersion coefficients.²²² For further discussion of the XDM model see the work by Angyan²²⁶ and Hesselmann.²²⁷ Surprisingly, the calculated dispersion coefficients are rather accurate, and tests on a set of small molecules showed fairly good results.²²⁵ In this respect the Johnson and Becke method is very promising. However, as we discussed in our paper,¹³⁶ the accuracy of the DFT-D scheme critically depends on the parametrization of the damping function. The damping used in the Johnson and Becke approach is not optimized, and in our opinion, even better results could be achieved with a more flexible damping function. This would be necessary if XDM were to be combined with different XC functionals.¹³⁶ In 2007 the authors combined the dispersion term with their XC functional in a unified DF07 functional for dynamical, nondynamical, and dispersion correlations.²²⁸ In 2009 Kong et al. derived and implemented the self-consistent XDM model and showed that the differences with respect to the non-self-consistent XDM model are rather small.¹⁸⁶ Overall it seems that the XDM method has the potential to become a very accurate DFT plus dispersion method.

4.3. Dispersion Correction in the Plane Wave Codes and Nonlocal Pseudopotentials

An interesting way of explicitly accounting for dispersion forces in DFT was suggested by Lilienfeld, Tavernelli, and Rothlisberger.²²⁹ The method is based on the pseudopotentials, and it is called DCACP, dispersion-corrected atom-centered potential. While in the DFT-D scheme the dispersion attractions are modeled as attractions between atom-centered points, in DCACP the attraction comes from an additional artificial potential that acts between electrons and nuclei. This effective atom-centered potential is optimized to reproduce the reference interaction energy, including dispersion. In this sense the DCACP method is empirical. An advantage is that the electron density reflects the “dispersion” contribution, unlike in the perturbative approaches. A disadvantage is that the method is useful only when pseudopotentials are applied, for instance, in the PW calculations. The current versions of DCACP (DCACP-MP2 and the newer DCACP-CCSD(T)) use single-channel expansion. This is enough to reproduce intermolecular potentials at the typical vdW distances, i.e., at separations smaller than ca. 5 Å, but it does not exhibit the correct asymptotic behavior ($1/R^6$).^{230,231} For correct asymptotics more channels should be used.²³² Therefore, these corrections can be very useful for correcting the major problems of DFT at the typical vdW distances, but may prove problematic when the long-range part of dispersion is non-negligible, which is often the case in solution. It is not clear whether this approach has any advantage over the conceptually simpler and, in the long range, more physical, empirical dispersion correction.

4.4. Reparametrization of Current Density Functionals

The idea of improving the performance of current density functionals for vdW interactions by adjusting their parameters, while keeping the present form, is very compelling. Most of the widely used density functionals, such as B3LYP, BLYP, PBE, PW91, etc., were designed or parametrized without regard for noncovalent interactions. It is therefore not impossible that slight parameter readjustments could lower the errors for the van der Waals complexes. Of course, we cannot forget that the LDA, GGA, meta-GGA, and hybrid functionals do not provide the desired $1/R^6$ asymptotic behavior, and any improvements can be expected only when the electron densities of the interacting molecules overlap. Still, in most situations we are interested in the interacting molecules in close contact—at the vdW distance or not very far from the vdW minimum. In such situations the electron overlap is quite appreciable. In addition, the magnitude of the correlation-related error (the missing attraction) usually reaches its maximum around the vdW distance (the contribution coming solely from the nonoverlap distances is relatively small).¹³⁶ It is easy to imagine that a suitably modified functional might provide enough additional attraction in this region, which could simulate the effect of the missing long-range correlation. However, we have to keep in mind that this attraction will most likely be spurious and we will be getting “the almost right answer for the wrong reason”. One of the other things to worry about is the quality of the resulting self-consistent electron density and especially its part originating in the long-range fluctuations. Still, if we could extend the applicability of the computationally efficient functionals of today to the overlap distances, it would be extremely helpful.

Many researchers have followed this idea with some success.^{233–238} The ad hoc BH&H functional of Becke²³⁹ has been shown to perform relatively well for the stacked complexes of DNA bases, but to overestimate the hydrogen-bonded complexes.²³³ For a long time this unbalanced description of the dispersion-bonded and hydrogen-bonded complexes seemed to be a characteristic of the functional reparametrization attempts.^{235,239} The X3LYP functional,²³⁴ which was especially designed for noncovalent interactions, failed to describe binding in these complexes qualitatively, demonstrating that parametrization and testing on rare-gas atoms do not guarantee reasonable performance for larger molecules.¹⁸² Similarly, the TPSS and TPSSh functionals, which perform well for the rare-gas dimers, fail for the larger molecules.¹³⁶

Recently, the focused efforts of Zhao and Truhlar, who have published a number of papers on empirical parametrization over the past few years, have led to functionals performing very well for a very wide variety of complexes.^{237,238,240} The authors used various types of reference data in the fitting procedure, including the interaction energies of noncovalent complexes, barrier heights, atomization energies, ionization potentials, etc. Their effort peaked in 2008 with the M06¹⁸⁵ suite of functionals (which comprises four different functionals with different areas of application) as well as the M08-HX and M08-SO²⁴¹ density functionals. For calculations on noncovalent complexes the authors recommend the M06-2X functional. It should be noted that the M06 functionals are very sensitive to the integration grid sparseness, thanks to the functional form of their kinetic energy density terms.²⁴² With the regularly used grids M06 functionals exhibit oscillations and

sometimes multiple minima on the potentials of weakly bonded complexes, so the use of ultrafine grids is recommended. This slows calculations significantly and somewhat lowers the advantage of the M06 suite with respect to the more demanding nonlocal DFT approaches (see below). Also, the long-range behavior of these functionals is necessarily unphysical and incorrect,²⁴² but as mentioned in the beginning of this section, in many applications this may not be a major problem. Unlike many other reparametrized functionals, the newer functionals of Zhao and Truhlar describe the medium-range correlation correctly using the correlation part of the XC functional.²⁴³

In conclusion, the most advanced functionals of Truhlar and co-workers show that, within the overlap distances, the empirically reparametrized functionals can provide surprisingly accurate binding energies even for noncovalent complexes.^{185,241} Although the asymptotic behavior of such reparametrized functionals is necessarily incorrect, they will likely become very useful in a number of applications.

4.5. Double Hybrid Functionals

Double hybrid functionals, like hybrid functionals, combine DFT with a certain portion of exact exchange, but in addition they also add a portion of correlation energy, usually from an MP2-type calculation. This means that the computational demands of double hybrid calculations are similar to those of MP2 calculations, i.e., significantly higher than for DFT itself. In return we can usually obtain somewhat more accurate results. Several double hybrid methods have been developed, using either Hartree–Fock unoccupied orbitals^{244,245} or DFT orbitals (B2PLYP¹⁵¹ and mPW2PLYP²⁴⁶). A computationally less demanding opposite-spin-only variant of this method was also developed (B2-OS3PLYP).²⁴⁷ For noncovalent complexes better accuracy can be achieved by adding a portion of the empirical dispersion energy on top of the double hybrid energy.²⁴⁸

4.6. Truly Nonlocal Density Functionals for DFT

DFT is an exact theory, and it can describe dispersion interactions correctly, but GGA or meta-GGA functionals are not good approximations to start with. If we want to treat the long-range correlations correctly, we need to choose from a class of truly nonlocal functionals. Several nonlocal forms have been suggested over the years, and recently, their development seems to be reaching the point of practical applicability.

The most common starting points for nonlocal and nonempirical functionals are the adiabatic connection fluctuation dissipation theorem (ACFDT) and the random phase approximation (RPA).^{249,250} Depending on the subsequent approximations, which mainly concern the frequency-dependent susceptibility, results of varying quality are obtained.

The first works on this subject were limited to the case of isolated fragments.^{251–253} They recovered the desired long-range vdW asymptotics, but the overlap effects were missing, and they diverged at short distances. Sato et al.^{254,255} suggested that when the formula of Andersson, Langreth, and Lundqvist (ALL)²⁵¹ is properly (empirically) damped at short distances, the long-range ALL correlation can be combined with a GGA-type functional. Note that it is the grid point–grid point interaction that is damped here, not the atom–atom interaction as in DFT-D. In 2009 Gräfenstein

and Cremer implemented the damped ALL formula in a quasi-self-consistent way (the density is not consistent with the dispersion interaction), allowing for faster analytical geometry optimizations and frequency calculations.²⁵⁶

In 2004 Dion et al. published a very promising functional called vdW-DF.²⁵⁷ The nonlocal correlation is expressed in terms of the density–density interaction formula, in which the interaction energy depends on densities at two different locations in space, r and r' , and the nonlocality enters through the kernel ϕ :²⁵⁷

$$E_c^{\text{nl}} = \frac{1}{2} \int d^3r d^3r' n(\vec{r}) \phi(\vec{r}, \vec{r}') n(\vec{r}') \quad (14)$$

The kernel ϕ is a formula analytically derived for a chosen approximation for a frequency-dependent response function, which depends on the density and density gradient. The entire correlation functional is nonempirical and yields the correct long-range asymptotics. This nonlocal correlation energy is then combined with an LDA correlation functional to get the total correlation, which is usually used in connection with the rev-PBE exchange functional:²⁵⁸

$$E_c[n, \nabla n] = E_c^{\text{LDA}}[n] + E_c^{\text{nl}}[n, \nabla n] \quad (15)$$

This scheme has an important advantage of being seamless, which means that the system does not have to be divided into fragments. Also the intramolecular dispersion interaction is accounted for in a natural way. The self-consistent equations and analytical dispersion gradients were derived in 2007.²⁵⁹ The available tests show that the accuracy of the vdW-DF functional predictions for intermolecular energies depends on the exchange functional with which it is combined.^{260,261}

The vdW-DF functional was also implemented self-consistently by Vydrov et al.,²⁶¹ who studied the effect of the exchange functional used on the performance of the vdW-DF scheme. In the subsequent work Vydrov et al. suggested a very interesting modification of the vdW-DF scheme, which removes the necessity to use the LDA correlation functional. Very importantly, the authors also deal with the question of adjusting the local and nonlocal correlation components (avoiding double counting) through the enhancement factor and an empirical constant.²⁶² In 2009, the same authors derived a new, simpler form of the nonlocal seamless functional based on the vdW-DF methodology, called VV09, which also performs fairly well for weakly bonded small complexes.²⁶³ A common feature of the vdW-DF-like functionals is that the total (atomic, molecular) correlation energies are worse than those predicted by current GGA functionals, and further work in this direction is necessary.^{260,263}

The vdW-DF functional evaluates grid point–grid point interactions, and therefore, the computational demands depend on the system size as n_{grid}^2 . This scaling is of lower order than the scaling of the SCF procedure ($\sim N^3$), but the prefactors are rather large, and the calculations are thus slower than a KS-DFT calculation itself for smaller molecules. Many groups are now working on faster and more efficient implementations of these methods.^{260,264,265}

The nonlocal functionals described above profit from formulas, derived on the basis of simple approximations for the response function, which are easily manageable, but there are also other ways of approaching the correlation energy, starting from the fluctuation dissipation theorem and employing the random phase approximation for the XC kernel.

Unfortunately, most of them are not as well tested, and they are often more computationally demanding. To mention at least some, we refer the reader to the works of Lein, Dobson, and Gross,²⁶⁶ Kohn et al.,²⁶⁷ Fuchs and Gonze,²⁶⁸ Furche and van Voorhis,²⁶⁹ Harl and Kresse,²⁷⁰ and Nguyen and de Gironcoli.²⁷¹

One common problem of the accurate nonlocal correlation functionals is the choice of a suitable exchange functional with which to combine it. Even if the correlation functional itself were exact, the intermolecular interaction energies would still be rather inaccurate due to the known exchange functional faults.^{260,261} Many authors argue that exact exchange is necessary for a correct description of intermolecular regions.²⁷² Range-separated approaches are often suggested as a solution.^{272–274} Nevertheless, the question of a sufficiently accurate exchange functional to be combined with the nonlocal correlation functionals still remains open.

4.7. Other Approaches

An interesting way to calculate the dispersion energy was suggested by Silvestrelli.²⁷⁵ In this method, the C_6 dispersion is calculated from the ALL formula,²⁵¹ but the electron density entering this formula is taken from the maximally localized Wannier functions (MLWFs). This allows one to perform analytic integration and to avoid the computationally intense numerical double integration over grid points, which slows most other ALL implementations. Importantly, the damping function is made dependent on the actual electronic density here, through the spread of the Wannier functions. The method is perturbative; the electron density is not consistent with the dispersion interactions. Maerzke et al., based on work by Chang et al.,²⁷⁶ suggested a method called self-consistent polarization density functional theory, SCP-DFT, in which dispersion is added using an expression consistent with second-order perturbation theory.²⁷⁷ Another simple approach by Mackie and DiLabio uses reparametrized local pseudopotentials to provide some additional attraction at vdW distances. This is not physically correct but reduces errors at short distances significantly.²⁷⁸

Wesolowski and Tran applied a bifunctional originally developed for other purposes to the weakly bonded complexes.²⁷⁹ Although his results are generally superior to those of LDA or GGA calculations, the approach is limited to the electron overlap regime (not providing the correct $1/r^6$ monomer asymptotics) and requires partitioning of the system into subsystems. Another scheme by Hesselmann calculates interaction energies only for the case of separated monomers.⁹⁸ Here, the total interaction energy is based on the supermolecular MP2 interaction energy corrected for the difference between the uncoupled and coupled second-order dispersion energy. In this case the interaction energies seem to be surprisingly accurate.

5. Symmetry-Adapted Perturbation Theory

Symmetry-adapted perturbation theory (SAPT)²⁸⁰ provides a deeper understanding of intermolecular interactions through the decomposition of the total interaction energy into physically meaningful components. The original version of SAPT was formulated as a double-perturbation series, in which the intramolecular and intermolecular perturbations were treated separately. An important advantage of the double-perturbative scheme is that it is exact, and inclusion of higher orders should bring one closer to the accurate

energy components. One of the disadvantages is its rather large computational demand.

In 2001 Williams and Chabalowski²⁸¹ proposed a new scheme in which the monomers in the SAPT calculations are represented by Kohn–Sham DFT and only the intermolecular interaction is treated perturbatively. This approach is much faster than the original SAPT, but the initial results were rather discouraging. Two groups independently followed the idea, and over the following two years developed two very similar approaches with very good results. Misquitta, Szalewicz, and Jeziorski named their method SAPT(DFT),^{282–285} and Hesselmann and Jansen named it DFT-SAPT.^{129,286–288} The key components in both approaches are the use of the asymptotically corrected DFT functionals (e.g., PBE0AC¹²⁹) and calculation of the dispersion energy from time-dependent density functional theory. The interaction energy is calculated (to the second order) as a sum of the electrostatic, $E^{(1)}_{\text{elst}}$, exchange-repulsion, $E^{(1)}_{\text{exch-rep}}$, induction, $E^{(2)}_{\text{ind}}$, exchange-induction, $E^{(2)}_{\text{exch-ind}}$, dispersion, $E^{(2)}_{\text{disp}}$, and exchange-dispersion, $E^{(2)}_{\text{exch-disp}}$, terms, and a correction for some of the higher order contributions, $\delta(\text{HF})$ (for more details see, e.g., refs 129 and 284):

$$E^{\text{SAPT}} = E^{(1)}_{\text{elst}} + E^{(1)}_{\text{exch-rep}} + E^{(2)}_{\text{ind}} + E^{(2)}_{\text{exch-ind}} + E^{(2)}_{\text{disp}} + E^{(2)}_{\text{exch-disp}} + \delta(\text{HF}) \quad (16)$$

In terms of accuracy, the original pure wave-function-based SAPT is claimed to be a reference quality method with results that are comparable to those of high-level wave-function-based calculations, such as CCSD(T). The DFT-based SAPT is naturally less reliable, partly because it depends on the quality of the density functional used. The electrostatic term and the second-order induction and dispersion terms are “potentially exact”,²⁸⁹ which means that they could be calculated exactly provided that the exact density XC functional and frequency-dependent XC kernel are known. The exchange energy is, unfortunately, not even potentially exact. Nevertheless, available numerical experience shows that DFT-based SAPT provides fairly accurate total interaction energies usually comparable with the pure SAPT results.^{129,284}

DFT-based SAPT is computationally much faster than high-level wave-function-based methods, and using currently available implementations (Molpro 2006 (also 2008 and 2009)²⁹⁰ and SAPT2008²⁹¹), complexes containing around 50 atoms can be treated using large basis sets, such as aug-cc-pVTZ, in tens of hours of CPU time.

6. Semiempirical Quantum Chemical Theories

Semiempirical QM methods can properly and fully describe all quantum effects. Because they are parametrized for covalent bonding, however, their use for noncovalent complexes is not straightforward. Let us add that, unlike the nonempirical Hartree–Fock method, which does not recover the correlation energy (and, consequently, also does not describe the London dispersion energy, which forms the dominant part of the intersystem correlation energy), all QM semiempirical methods can, in principle, recover the correlation energy.

Semiempirical methods were originally developed for theoretical studies on extended systems, for which the computational cost of nonempirical *ab initio* methods was prohibitively high.²⁹² The second-generation methods such

as MNDO,^{293,294} AM1,²⁹⁵ and PM3^{296,297} provided relatively reliable results for systems containing the main-group elements in their ground electronic states, which can be described by *s* and *p* atomic orbitals. The use of these methods for noncovalent complexes (H-bonded and dispersion-bound) was, as mentioned above, limited. The original MNDO method was not able to describe either of these interaction types. Since it was believed that H-bonding is more important than dispersion, the problem was addressed in subsequent methods (AM1, PM3, and others) by the introduction of an additional core–core term as well as by parametrization of the method toward hydrogen bonding. Another step forward was introduced in the parametrization of the PM3 method, in which not only the energies but also the geometries of H-bonded complexes were emphasized. Recently, a special parametrization of the PM3 method, PM3(BP), was introduced for applications related to nucleic acid base pairs.²⁹⁸

The first attempt to include the dispersion energy in semiempirical QM methods was made by Martin and Clark²⁹⁹ though the addition of R^{-6} , R^{-8} , and R^{-10} dispersion terms to existing semiempirical core–core repulsion appeared much earlier.³⁰⁰ The dispersion energy in ref 299 was calculated using the additive “atomic orbital” polarizability tensors, and the best values were obtained by using the Slater–Kirkwood modification of the London formula. A similar approach was used later for modification of the AM1 and PM3 methods.³⁰¹ After reparametrization of both the semiempirical method and the damping function, considerably improved characteristics of noncovalent complexes resulted. The semiempirical methods of the OM x ($x = 1, 2,$ and 3) family are known to describe noncovalent complexes better than AM1 and PM3, which is mainly due to the use of orthogonalization corrections.^{302–304} These corrections improve the description of the Pauli repulsion. Tuttle and Thiel augmented the OM2 and OM3 methods with empirical dispersion terms, but in contrast to the previous case, they used the original parametrizations (for the parent semiempirical methods).³⁰⁵ Similarly, as in the cases of AM1 and PM3, significant improvement of the original methods for noncovalent interactions was achieved. Jorgensen et al. developed the bond and group equivalents scheme, which considerably improved not only the heats of formation but also the H-bonding energies.³⁰⁶ The suggested modification of the PM3 method, called the pairwise distance-directed Gaussian (PDDG) method, provides an improvement over PM3 in the description of intermolecular interactions. Application of the method is limited to compounds containing C, H, N, and O atoms (as it is for many semiempirical QM methods). Possible future developments toward improving H-bonds within the MNDO framework are outlined in ref 307, but no further progress has been reported up to now. To summarize, we can state that all currently used semiempirical QM methods systematically underestimate the strength of H-bonds by approximately 20–30% and if the dispersion energy term is not added they fail completely for dispersion-bound complexes. However, even if the dispersion energy is added, the respective methods are still not accurate enough for most applications in noncovalent complexes. There are many reasons for this shortcoming, such as parametrization for only a limited number of atoms, strongly overestimated stabilization energies for optimized geometries of H-bonded complexes, etc.

An interesting attempt to combine the semiempirical QM and MM was made recently.³⁰⁸ The authors used the PM3 Hamiltonian in combination with the AMBER force field. Contrary to standardly used QM/MM approaches, where the reaction center is described by QM and the environment by MM, here one subsystem is described by QM and the other one by MM. The authors modified the commonly used core–core approach by Field et al.³⁰⁹ and replaced it by a new one that provides stronger H-bonding (in comparison with other QM/MM semiempirical approaches). The suggested procedure provides very promising stabilization energies for model complexes.

Recently, the new semiempirical method PM6 (parameterized model 6) was introduced and was shown to be superior to other semiempirical QM methods in various aspects.³¹⁰ It is an NDDO-based method improved by the adoption of Viotyuk's core–core diatomic interaction term and Thiel's d orbital approximation.^{302–304} These modifications allowed parametrization for 80 elements and also reduced the error for main-group elements.^{310,311} PM6 is available in the MOPAC code (versions 2007 and higher).³¹² The latest version, MOPAC 2009, introduces another interesting feature that makes the PM6 method usable for very large systems—a linear scaling variant of the SCF procedure using localized orbitals, named MOZYME.³¹² Because of favorable scaling, the code can even be used in MD simulations.

Despite all these advantages, the PM6 method (like other semiempirical QM methods) still lacks the ability to accurately describe noncovalent interactions, specifically the dispersion energy and hydrogen bonding. Even though the method yields surprisingly good geometries for all types of complexes, the interaction energies for dispersion-bound and H-bonded complexes are substantially underestimated. As mentioned above, this feature is characteristic for all semiempirical methods (see also ref 313). The approximations used simply do not allow one to simultaneously obtain good descriptions of various noncovalent interactions and covalent interactions, whose description is entirely different. Let us recall once again that this is a very demanding task, and among all widely used *ab initio* QM procedures (i.e., methods that do not use any empirical or experimental parameters), only the CCSD(T)/CBS technique satisfactorily describes all these interactions.

To improve the performance of the PM6 method, two modifications were recently introduced: (i) an empirical dispersion energy term, which improves the description of complexes controlled by the dispersion energy, and (ii) an additional electrostatic term, which improves the description of hydrogen-bonded complexes. This correction is directional and should thus provide a better description of H-bonds, including directional characteristics, than the standard semiempirical QM methods. The resulting method, PM6 with corrections for dispersion and hydrogen bonding, is named PM6-DH.³¹⁴ The aim of these modifications was ambitious—to achieve the standard *ab initio* chemical accuracy (~ 1 kcal/mol) for extended noncovalent complexes.

The first PM6-DH method provides very good estimates of interaction energies for dispersion-bound complexes, but the description of H-bonded complexes was still not completely satisfactory. A further improvement (the so-called second-generation correction) of the H-bonding corrections was suggested recently.³¹⁵ The main difference with respect to the first-generation correction³¹⁴ concerns the description

of $X-H\cdots Y-Z$ H-bonding, where six internal coordinates were considered: the $H\cdots Y$ bond length, the two bond angles $X-H\cdots Y$ and $Z-Y\cdots H$, and the corresponding three torsion angles. As a result, the new version of the H-bonding correction includes three global and five method-dependent parameters describing H-bonds involving N and O acceptor and donor atoms. These parameters were adjusted on the basis of high-level QM calculations. The improved correction was evaluated for the PM6, AM1, OM3, and SCC-DF-TB-D methods and provides a significant improvement over the original versions of the methods mentioned above.

7. JSCH-2005 and S22 Databases

There are several reasons for performing highly accurate QM calculations for model complexes; probably the most important among them is the necessity to have a suitable test set for validating lower level computational methods. We have shown that, among all computational procedures considered, only the CCSD(T)/CBS procedure represents the genuine *ab initio* technique capable of producing accurate results for medium-sized complexes. In this technique, all quantities are directly calculated and no empirical parameters are adopted. The theory behind the procedure allows for the description of all different types of noncovalent interactions occurring in molecular clusters and complex molecular systems. The computational cost of the CCSD(T)/CBS procedure for medium and extended systems is, however, very high, and despite enormous progress in computational hardware and software, use of this technique for systems with more than about 50 atoms is (and will be in the near future) still impractical. The same is true for dynamical calculations of smaller systems. Evidently, much faster computational procedures should be introduced, allowing one to perform accurate calculations for these systems.

As was shown in the previous parts of the review, the standard lower level computational procedures (both WFT and DFT) fail to describe various types of molecular complexes with quantitative accuracy, with the most frequent problem being connected with the description of the dispersion energy. The only chance for the treatment of larger systems is thus to parametrize the energy term in these (and other) methods against highly accurate computational methods. The benchmark database set should cover all important bonding motifs and should be easily extended in the future for new structural motifs. MP2/CBS and CCSD(T)/CBS interaction energies and geometries for more than 100 DNA base pairs, amino acid pairs, and other model complexes were presented in the JSCH-2005 benchmark set.¹³¹ Extrapolation to the CBS limit was done with the two-point extrapolation method using several basis set extrapolation schemes (aug-cc-pVDZ \rightarrow aug-cc-pVTZ, aug-cc-pVTZ \rightarrow aug-cc-pVQZ, cc-pVTZ \rightarrow cc-pVQZ). CCSD(T) correction terms, determined as a differences between CCSD(T) and MP2 interaction energies, were evaluated with smaller basis sets (6-31G**(0.25,0.15) and cc-pVDZ). We have seen that this procedure yields reasonable stabilization energies that are comparable to the most accurate ones obtained by direct extrapolation of the CCSD(T) energies. Two sets of complex geometries were used, optimized and experimental ones.

Apart from the large JSCH-2005 set, a smaller S22 set was also introduced.¹³¹ For the latter set, larger basis sets were used for extrapolation to the CBS limit and optimized geometries generated using the CCSD(T) and counterpoise-corrected MP2 methods were sometimes adopted. The S22

Table 1. Benchmark Stabilization Energies (ΔE) for Seven H-Bonded, Eight Dispersion-Controlled, and Seven Mixed Complexes Determined at the CCSD(T)/CBS Level^a

no.	complex (symmetry)	ΔE (MP2)	ΔE (CCSD(T)/CBS)	geometry
Hydrogen-Bonded Complexes				
1	(NH ₃) ₂ (<i>C_{2h}</i>)	-3.02 (QZ)/-3.10 (5Z)/-3.20	-3.17 (qz)	CCSD(T)/QZ
2	(H ₂ O) ₂ (<i>C_s</i>)	-4.75 (QZ)/-4.89 (5Z)/-5.03	-5.02 (qz)	CCSD(T)/QZ
3	formic acid dimer (<i>C_{2h}</i>)	-17.88 (QZ)/-18.23 (5Z)/-18.60	-18.61 (tz)	CCSD(T)/TZ
5	uracil dimer (<i>C_{2h}</i>)	-19.90 (TZ)/-0.28 (QZ)/-20.61	-20.65 (tz-fd)	MP2/TZ-CP
6	2-pyridoxine-2-aminopyridine (<i>C₁</i>)	-15.91 (TZ)/-6.77 (QZ)/-17.37	-16.71 (tz-fd)	MP2/TZ-CP
7	adenine-thymine WC (<i>C₁</i>)	-14.92 (TZ)/-5.89 (QZ)/-16.54	-16.37 (dz)	MP2/TZ-CP
Complexes with Predominant Dispersion Contribution				
8	(CH ₄) ₂ (<i>D_{3d}</i>)	-0.42 (QZ)/-0.46 (5Z)/-0.51	-0.53 (qz)	CCSD(T)/TZ
9	(C ₂ H ₄) ₂ (<i>D_{2d}</i>)	-1.43 (QZ)/-1.57 (5Z)/-1.62	-1.51 (qz)	CCSD(T)/QZ
10	benzene-CH ₄ (<i>C₃</i>)	-1.66 (QZ)/-1.75 (5Z)/-1.86	-1.50 (tz-fd)	MP2/TZ-CP
11	benzene dimer (<i>C_{2h}</i>)	-4.70 (aT)/-4.85 (aQ)/-4.95	-2.73 (adz)	MP2/TZ-CP
12	pyrazine dimer (<i>C_{2h}</i>)	-6.56 (aT)/-6.76 (aQ)/-6.90	-4.42 (tz-fd)	MP2/TZ-CP
13	uracil dimer (<i>C₂</i>)	-10.63 (TZ)/-10.99 (QZ)/-11.39	-10.12 (tz-fd)	MP2/TZ-CP
14	indole-benzene (<i>C₁</i>)	-6.44 (TZ)/-7.42 (QZ)/-8.12	-5.22 (dz)	MP2/TZ-CP
15	adenine-thymine stack (<i>C₁</i>)	-12.30 (TZ)/-13.83 (QZ)/-14.93	-12.23 (dz)	MP2/TZ-CP
Mixed Complexes				
16	ethene-ethine (<i>C_{2v}</i>)	-1.57 (QZ)/-1.62 (5Z)/-1.69	-1.53 (tz)	CCSD(T)/QZ
17	benzene-H ₂ O (<i>C_s</i>)	-3.28 (QZ)/-3.43 (5Z)/-3.61	-3.28 (tz-fd)	MP2/TZ-CP
18	benzene-NH ₃ (<i>C_s</i>)	-2.44 (QZ)/-2.57 (5Z)/-2.72	-2.35 (tz-fd)	MP2/TZ-CP
19	benzene-HCN (<i>C_s</i>)	-4.92 (aT)/-5.06 (aQ)/-5.16	-4.46 (tz-fd)	MP2/TZ-CP
20	benzene dimer (<i>C_{2v}</i>)	-3.46 (aT)/-3.55 (aQ)/-3.62	-2.74 (adz)	MP2/TZ-CP
21	indole-benzene T shape (<i>C₁</i>)	-6.16 (TZ)/-6.65 (QZ)/-7.03	-5.73 (dz)	MP2/TZ-CP
22	phenol dimer (<i>C₁</i>)	-6.71 (TZ)/-7.33 (QZ)/-7.76	-7.05 (tz-fd)	MP2/TZ-CP

^a For structures, see Figure 1. MP2 stabilization energies in smaller/larger basis sets and the CBS limit and CCSD(T)/CBS stabilization energies (kcal/mol) are given. Deformation energies of monomers are not included. The basis set abbreviations TZ, aTZ, QZ, aQZ, and 5Z (in parentheses) stand for cc-pVTZ, aug-cc-pVTZ, cc-pVQZ, aug-cc-pVQZ, and cc-pV5Z, respectively. In the modified cc-pVTZ set (tz-fd), one set of f functions and one set of d functions were removed (only the more diffuse d function was kept) and the hydrogen basis set was modified anionally. Geometries were determined with full (counterpoise-corrected) optimizations with analytical (MP2) or numerical (CCSD(T)) gradients.

set can be recommended for parametrization or for a first screening of a method, while the larger JSCH-2005 set is more appropriate for verification of a computational procedure. Table 1 contains interaction energies for the S22 set (the structures of all 22 complexes are visualized in Figure 1), which contains 7 H-bonded complexes, 8 complexes with predominant dispersion contributions, and 7 mixed complexes (containing a mixture of electrostatic and dispersion interactions). It is important that the set span a wide range of interaction strengths to represent the diversity of interactions in biomacromolecules. In each of the above-mentioned subgroups the stabilities of the complexes range between 3 and 20, 0.5 and 15, and 1.2 and 8 kcal/mol, respectively. In this set the dispersion-bound complexes are as numerous as H-bonded ones, but they contribute to the sum of the stabilization energies by less than 40%, which is in line with our previous calculations on DNA nucleic acid bases. The energies presented in Table 1 represent a very valuable set that can be used for testing different lower level methods. A strong point here is the fact that different types of noncovalent complexes covering the most important biomolecular motifs are presented. It should be mentioned that the originally published S22 set contained some errors; the present version is upgraded, and all these errors and uncertainties have been removed.

How accurate are the S22 benchmark data? The S22 set was published in 2006, and since then, higher level calculations have been performed either for selected complexes^{48,52} or for the entire set.^{127,316} On the basis of these studies, it is possible to conclude that the stabilization energies of H-bonded complexes are slightly underestimated while those of dispersion-bound complexes are slightly overestimated when compared to the most accurate CCSD(T)/CBS values. The errors are, however, in the tens of kilocalories per mole, and the largest error of about 0.5 kcal/mol was found for

the stacked benzene...indole, adenine...thymine, and methyladenine...methythymine complexes. As expected, the source of these errors is the determination of the CCSD(T) correction term with small basis sets. Evidently, in this step, a basis set of at least aug-cc-pVDZ quality should be used.

8. Results and Discussion

8.1. Energy Characteristics

To establish a consistent metric for the evaluation of the performances of various theoretical techniques for the calculation of binding energies in noncovalent complexes, we have decided to use one consistent data set of benchmark interaction energies. This data set should comprise various motifs of noncovalent complexes, with energy characteristics that are consistently evaluated at a very high theoretical level. We use our S22 data set and, to a lesser extent, the JSCH-2005 data set. The reason is evident; these benchmark sets have frequently been used in the literature in recent years and represent, without doubt, the most frequently used data set for noncovalent interactions. Table 2 comprises statistical data on the performance of various WFT, DFT, and semiempirical QM methods,^{74,120,122,133,136,139,206,210,243,255,260,315,317,318,319,320} while the preceding table gives information on the performance of various methods toward some subsets of the other data sets mentioned above.

When comparing the performances of different methods for the S22 set, we will look not only at the total rmsd's, but also at the MSEs and rmsd errors of different types of complexes separately. This gives us additional information regarding the strengths and weaknesses of each method. For instance, the average error (MSE) of the B3LYP-D/TZVP method is 0.09 kcal/mol for the entire S22 set, which is an excellent result; however, the MSEs for the hydrogen-bonded

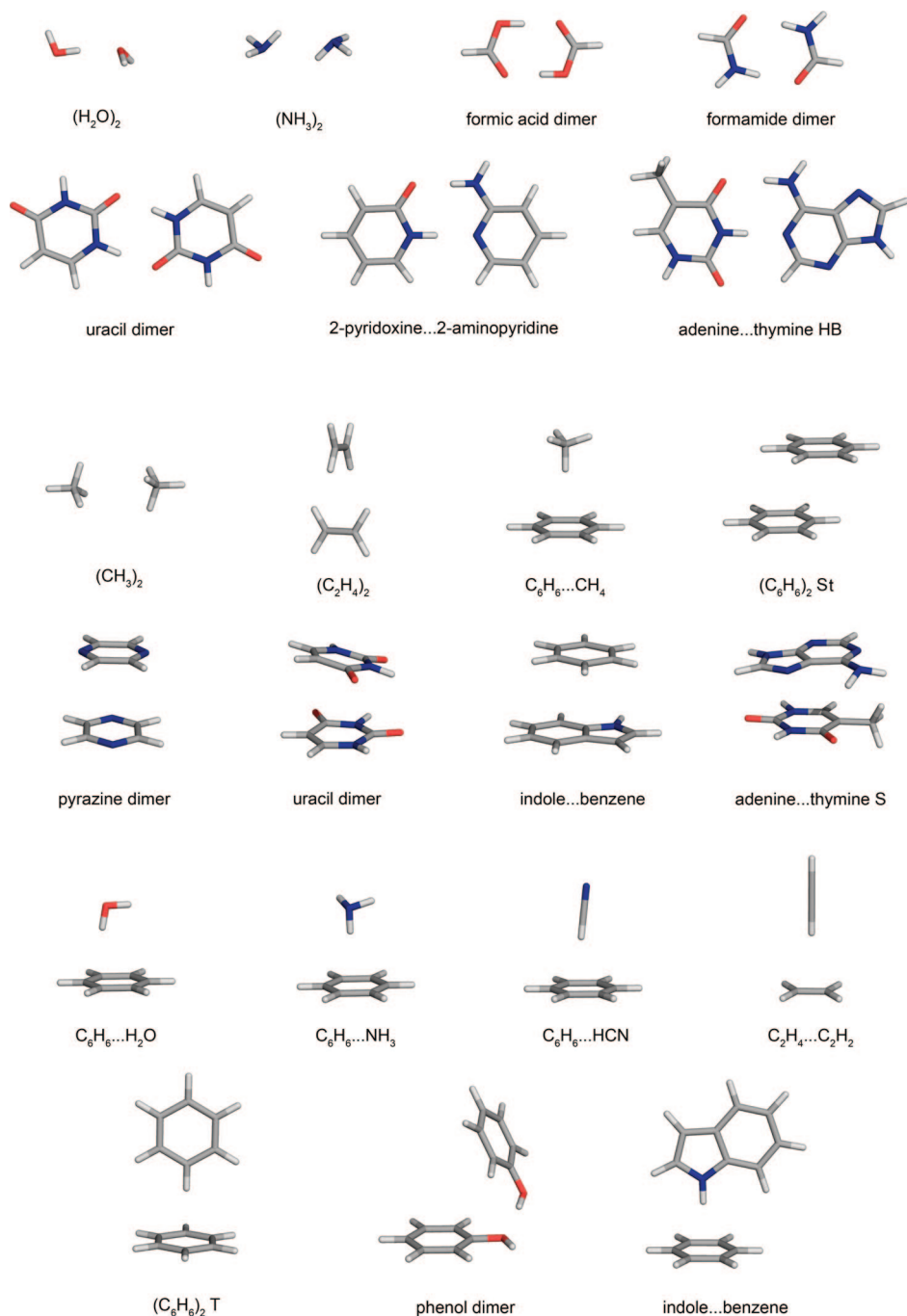


Figure 1. Structures of 22 complexes from the S22 data set.

and the dispersion-bonded complexes are -0.83 and 0.69 kcal/mol, respectively. The average difference between hydrogen-bonded and stacked complexes, $\Delta\Delta_{D-H}$, is thus $-0.83 + 0.69 = 1.52$ kcal/mol. This is a very large systematic error, which may be detrimental in a case where we want to compare, for instance, the relative stability of small peptide conformers containing both hydrogen bonds and dispersion contacts, because the typical energy difference between such conformers is rather small (usually less than 1 kcal/mol).^{212–214} Note, however, that the corresponding error in pure B3LYP is over 5 kcal/mol, which makes pure B3LYP completely useless for such applications. This example shows that more detailed analysis of the S22 errors brings valuable additional insight. Discussions will be presented separately for WFT, DFT, and semiempirical QM methods.

8.1.1. WFT Methods

8.1.1.1. MP2- and SCS-MP2-Based Methods. Comparison of MP2 results with the estimated CCSD(T)/CBS benchmark values confirms the typical error pattern discussed in section 3. MP2/CBS values (obtained by Helgaker extrapolation from aug-cc-pVTZ and aug-cc-pVQZ basis sets) for H-bonded complexes are underestimated on average by ~ 0.2 kcal/mol, while being strongly overestimated for the dispersion-dominated (primarily the $\pi-\pi$ stacked) complexes on average by as much as 1.5 kcal/mol. This unbalanced performance leads to a $\Delta\Delta_{D-H}$ value of almost -1.3 kcal/mol, thus making the MP2/CBS method unsuitable for simultaneous calculation or comparison of the stabilities of complexes and conformers with different interaction characters. The accuracy for complexes with “mixed”

Table 2. Total Root Mean Square Deviations (rmsd's), Mean Signed Errors (MSEs), and Average Differences between MSE ($\Delta\Delta_{D-H}$) (kcal/mol) for Hydrogen-Bonded, Stacked, and Mixed Complexes, Determined for 22 Complexes from the S22 Data Set Based on Various WFT, DFT, and Semiempirical QM Methods

method	basis set	ref	S22		H-bonded		dispersion-bonded		mixed		$\Delta\Delta_{D-H}$
			rmsd	MSE	rmsd	MSE	rmsd	MSE	rmsd	MSE	
MP2	cc-pVTZ	136	1.10	-1.36	0.36	-1.05	1.62	-1.92	0.58	-1.03	-0.88
MP2	cc-pVTZ/CP	136	0.85	0.30	0.59	1.20	0.75	-0.24	0.31	0.03	-1.44
MP2	CBS(at-aq)	120	0.94	-0.80	0.27	-0.21	1.24	-1.50	0.37	-0.58	-1.29
SCS-MP2	CBS(at-aq)	120	0.58	0.66	0.54	1.15	0.60	0.47	0.17	0.37	-0.68
SCS(MI)-MP2	CBS(t-q)	122	0.26	0.00	0.31	0.02	0.28	0.03	0.22	-0.05	0.00
SOS(MI)-MP2	CBS(t-q)	122	1.06	0.42	0.87	1.56	0.83	-0.29	0.21	0.07	-1.85
MP2C	MP2(CBS)+ Δ^a	133	0.18	0.01	0.19	0.05	0.17	0.10	0.13	-0.12	0.05
MP2.5	MP2(CBS)+ Δ^a	139	0.22	-0.16	0.07	-0.04	0.32	-0.25	0.14	-0.19	-0.21
MP3	MP2(CBS)+ Δ^a	this work	0.67	0.46	0.24	0.00	0.82	1.02	0.20	0.27	1.02
CCSD	MP2(CBS)+ Δ^a	this work	0.42	0.63	0.26	0.59	0.55	0.88	0.20	0.38	0.29
SCS-CCSD	MP2(CBS)+ Δ^a	this work	0.29	0.23	0.27	0.50	0.28	0.14	0.05	0.06	-0.37
SCS(MI)-CCSD	MP2(CBS)+ Δ^a	this work	0.17	0.00	0.04	0.03	0.27	0.04	0.05	-0.07	0.01
LMP2	aug-cc-pVTZ	74	1.05	-0.17	0.39	0.93	1.01	-0.99	0.43	-0.33	-1.91
SCS-LMP2	aug-cc-pVTZ	74	1.06	1.25	1.01	2.28	0.90	0.90	0.25	0.61	-1.38
SOS-LMP2	aug-cc-pVTZ	74	1.35	1.95	1.34	2.95	1.43	1.85	0.38	1.08	-1.11
SCSN-LMP2	aug-cc-pVTZ	74	0.37	-0.06	0.52	-0.18	0.23	0.10	0.29	-0.12	0.28
DFT-SAPT	aDZ	this work	0.91	1.10	0.90	1.86	0.83	0.90	0.47	0.56	-0.96
DFT-SAPT	aTZ	this work	0.56	0.52	0.53	0.95	0.56	0.42	0.34	0.21	-0.52
DFT-SAPT	CBS(ad-at)	this work	0.49	0.37	0.47	0.78	0.45	0.22	0.30	0.14	-0.56
FN-DQMC		317	0.79	-0.30	0.71	-0.74	0.52	0.27	0.81	-0.52	1.01
TPSS	LP ^c	136	2.88	3.02	1.02	1.43	3.76	5.17	1.25	2.15	3.75
TPSS-D	LP ^c	136	0.38	-0.14	0.48	-0.37	0.32	-0.07	0.23	0.02	0.30
TPSS-D	LP ^c	136	0.92	-0.59	1.22	-1.45	0.30	-0.26	0.28	-0.10	1.19
B3LYP	TZVP	136	3.28	3.35	1.16	1.06	3.86	6.09	1.40	2.49	5.03
B3LYP-D	TZVP	136	0.82	0.09	0.35	-0.83	0.48	0.69	0.67	0.32	1.52
BLYP-D	TZV(2df,2pd)	210	0.50	-0.33	0.35	-0.76	0.47	0.05	0.28	-0.35	0.82
B97-D	TZV(2df,2pd)	210	0.45	-0.01	0.57	0.17	0.22	0.10	0.40	-0.31	-0.07
B97-D	LP ^c	320	0.52	0.42	0.55	0.81	0.34	0.40	0.43	0.06	-0.42
ω B97X-D	LP ^c	320	0.25	-0.09	0.28	-0.16	0.28	-0.01	0.15	-0.12	0.15
PW86PBE+XDM(XX)	LP ^c	319	1.05	0.67	0.35	-0.05	1.36	1.36	0.54	0.59	1.41
PW86PBE+XDM(BR)	LP ^c	319	0.62	0.25	0.30	-0.25	0.75	0.65	0.31	0.29	0.90
B2-PLYP	TZV(2df,2pd)	248	1.63	1.63	0.62	0.50	1.92	2.98	0.68	1.21	2.48
B2-PLYP-D	TZV(2df,2pd)	248	0.34	0.21	0.23	0.54	0.23	-0.10	0.16	0.24	-0.64
mPW2PLYP	TZV(2df,2pd)	248	1.51	0.95	0.50	-0.28	1.65	2.28	0.58	0.67	2.56
mPW2PLYP-D	TZV(2df,2pd)	248	0.64	0.25	0.27	0.92	0.48	-0.39	0.19	0.32	-1.31
vdW-revPBE	PW, 250 Ry ^b	260	1.42	1.34	1.50	2.79	0.95	0.67	0.40	0.65	-2.12
vdW-PBE	PW, 250 Ry ^b	260	0.80	-1.17	0.57	-0.50	0.75	-1.86	0.24	-1.06	-1.37
LC-BOP+ALL	6-311+G(2df,2p)	255	0.73	0.20	0.51	-0.35	0.83	0.74	0.30	0.13	1.09
M06-2X	DIDZ ^c	318	0.64	-0.35	0.76	-0.29	0.75	-0.50	0.38	-0.24	-0.21
M06-2X	DIDZ ^c CP	318	0.53	0.23	0.84	0.39	0.32	0.11	0.35	0.20	-0.28
M06-2X	MG3S ^c	318	0.55	-0.15	0.64	0.18	0.49	-0.46	0.35	-0.16	-0.64
M06-2X	MG3S ^c CP	318	0.50	0.29	0.69	0.63	0.21	0.09	0.39	0.20	-0.55
M06-2X ^d	DIDZ ^c	318	0.78	-0.59	-0.76	-0.70	1.01	-0.71	0.49	-0.33	-0.01
M06-2X ^d	DIDZ ^c CP	318	0.55	0.02	0.77	0.01	0.51	-0.06	0.39	0.11	-0.07
M06-2X ^d	MG3S ^c	318	0.67	-0.37	0.75	-0.17	0.74	-0.66	0.43	-0.24	-0.49
M06-2X ^d	MG3S ^c CP	318	0.54	0.10	0.73	0.29	0.41	-0.09	0.43	0.12	-0.38
M05-2X	6-311+G(2df,2p)	243	0.88	0.76	0.81	0.61	1.03	1.23	0.59	0.37	0.62
ω B97X	LP ^c	206, 320	1.20	0.51	0.65	-0.29	1.46	1.40	0.55	0.30	1.69
PM6		315	2.51	3.38	3.00	4.86	2.46	3.26	1.17	2.04	-1.59
PM6-D		315	2.21	1.44	2.47	3.83	0.59	0.02	0.75	0.68	-3.81
PM6-DH2		315	0.53	0.12	0.34	-0.04	0.58	0.01	0.58	0.42	0.05
OM3		315	2.27	3.60	2.22	4.03	2.85	4.13	1.34	2.58	0.10
OM3-D		315	1.64	0.47	1.71	2.14	0.73	-0.88	0.37	0.33	-3.02
OM3-DH2		315	0.84	-0.07	0.57	0.50	0.73	-0.88	0.36	0.28	-1.38
SCC-DFTB		315	2.15	3.55	2.13	4.23	2.71	3.76	1.18	2.61	-0.47
SCC-DFTB-D		315	1.53	1.80	1.44	3.39	0.82	0.74	0.86	1.43	-2.65
SCC-DFTB-DH2		315	1.11	0.40	0.97	-0.75	0.81	0.73	0.50	1.17	1.48
AM1		315	5.50	6.81	6.48	10.56	5.11	6.63	1.55	3.27	-3.93
AM1-D		315	4.96	2.57	5.89	7.99	0.58	-0.26	0.73	0.38	-8.25
AM1-DH2		315	0.87	-0.21	1.19	-0.46	0.58	-0.27	0.81	0.12	0.19

^a These computations were made in the same way as for estimated CCSD(T)/CBS. MP2 calculations were extrapolated to the CBS, and then a correction term was added (Δ MP2C, Δ MP2.5, Δ CCSD, Δ MP3, Δ SCS-CCSD, or Δ SCS(MI)-CCSD). The basis set used for any given correction is the same as the corresponding basis used for the Δ CCSD(T) term in the original S22 paper.¹³¹ ^b Geometries fully optimized with the given method. ^c DIDZ is 6-31+G(d,p), MG3S is 6-311+G(2df,2pd), and LP is 6-311++G(3df,3pd).³²¹ ^d Structures optimized with the M06-2X/DIDZ method.

character is also quite unsatisfactory, with average errors being almost -0.6 kcal/mol. Comparing the basis set unsaturated CP-corrected MP2/cc-pVTZ with MP2/CBS, the same trends can be observed. The quality of the description for H-bonded complexes increases with the quality of the

basis set, while the overestimation of the $\pi-\pi$ stacking interaction is enhanced. Neglecting the CP correction in MP2 calculations is disadvantageous for both H-bonded and $\pi-\pi$ stacked complexes for reasons similar to those mentioned above.

The scaling of opposite- and/or same spin-components in SCS-MP2/CBS leads to appreciable improvements compared to the MP2/CBS on average; however, for the H-bonded complexes the average error increases by almost 1 kcal/mol. The use of scaling parameters optimized against molecular interactions in SCS(MI)-MP2 clearly improves the balance of the method, leading to a decrease in the SCS-MP2 $\Delta\Delta_{D-H}$ from -0.68 to almost 0 kcal/mol for SCS(MI)-MP2. Still, the deviation of the SCS(MI)-MP2 results from the CCSD(T) reference is about 0.3 kcal/mol, comparable to the accuracy of MP2/CBS for all but the dispersion-dominated complexes. Neglect of the same-spin component in SOS(MI)-MP2 leads to a less balanced treatment than that given by MP2 or SCS-MP2 with an average error comparable with that of MP2, but for a calculation speedup of about an order of magnitude.

Substitution of the UCHF dispersion energy contained in the supermolecular MP2 interaction energy by the more accurate one from TDDFT leads to major improvement in the accuracy; i.e., see the results for the MP2C method.¹³³ The rmsd of about 0.2 kcal/mol for the whole S22 is comparable with that of the SCS(MI)-MP2 method fitted for the S22 set, or that of the SCS(MI)-CCSD method, which is, however, more than an order of magnitude more expensive. The $\Delta\Delta_{D-H}$ value of 0.05 kcal/mol is one of the lowest from all the methods presented in Table 2 (except for M06-2X/DIDZ, which is fortuitously -0.01 kcal/mol, or the already mentioned SCS(MI)-MP2 method fitted for this test set, which is ~ 0 kcal/mol), making the hybrid WFT/DFT MP2C method one of the most balanced option for accurate calculation of different types of interactions.

Comparison of MP2 and LMP2 calculated in the aug-cc-pVTZ basis set indicates a certain error compensation between the overestimation of the stabilities of $\pi-\pi$ stacked complexes and the intrinsic error of the local treatment. The overall accuracy of LMP2 seems to be slightly better than that of MP2, but the balance of LMP2 is slightly worse due to larger errors in the description of H-bonded complexes.

8.1.1.2. Third-Order WFT Methods: MP2.5, MP3, CCSD, SCS-CCSD, and SCS(MI)-CCSD. Results for all methods in this paragraph were obtained analogously to the estimated CCSD(T)/CBS values used as benchmarks; i.e., higher order correlation contributions were calculated with rather small- to medium-sized basis sets in the calculation of $\Delta\text{CCSD(T)}$ (see Table 2). Let us first analyze the performance of the MP3 method. The overall accuracy and balance of MP3/CBS is better than that of MP2/CBS, comparable perhaps with that of SCS-MP2/CBS. A characteristic pattern opposing the one observed for MP2 can be recognized in the MP3 results for the dispersion-dominated complexes (again, these are mostly governed by strong, $\pi-\pi$ stacked complexes), which is a strong underestimation of binding energies by an average of about 1 kcal/mol. This is contrary to the overestimation of binding energies, by an average of about 1.5 kcal/mol, seen for MP2. The quality of the results for the H-bonded complexes is slightly improved over those of MP2, leading to an average error of about zero, probably attributable to error compensation caused by the presence of H-bonded complexes with significant dispersion contributions (such as the 2-pyridoxine \cdots 2-aminopyridine complex, for instance). The error statistics for the more sophisticated CCSD method show that inclusion of correlation effects beyond the MP3 method strongly disrupts the error cancellation for the group of H-bonded complexes, leading to systematic underestimation by an average of about

0.6 kcal/mol. The overestimation of interactions for complexes in the “dispersion-dominated” group is also reduced, thus improving the balance of CCSD over MP3 by ~ 0.7 kcal/mol, but decreasing the accuracy, as measured by the average error, by almost 0.2 kcal/mol.

The accuracy of both MP3 and CCSD is significantly increased after introduction of empirical approximations either via scaling of the third-order correlation contribution, i.e., MP2.5, or by scaling of the same- and opposite-spin components of the CCSD correlation energy, i.e., SCS-CCSD and SCS(MI)-CCSD. Scaling of the third-order correlation energy in MP2.5 to the average systematic errors of MP2 and MP3 leads to a more balanced treatment, with DDH being about -0.2 kcal/mol, and increases the accuracy to an average error of -0.16 kcal/mol. The performance of this method for the H-bonded complexes is excellent, but an average overestimation of binding energies by about 0.3 kcal/mol, attributable to the $\pi-\pi$ stacked complexes, is still not eliminated due to the inexactness of the scaling factor. The most accurate and sufficiently balanced results are obtained by the SCS(MI)-CCSD method. The overall average error is about 0.2 kcal/mol (below 0.3 kcal/mol for each group of complexes individually), and the $\Delta\Delta_{D-H}$ value is essentially zero. This indicates that the SCS(MI)-CCSD method is capable of delivering the most accurate and balanced description from all the tested methods.

8.1.2. DFT Methods

8.1.2.1. Pure DFT Functionals. Table 2 compares the performance of various DFT plus dispersion methods, and in several cases, results for the corresponding pure functional (i.e., without the dispersion correction) are shown. Without the dispersion correction, the pure DFT functionals exhibit very similar deficiencies when used along with large basis sets; these inaccuracies are well-known and have been characterized for many complexes. The interaction energies of the dispersion-bonded complexes are strongly underestimated, and the hydrogen-bonded and mixed complexes are also, on average, underestimated by the GGA, meta-GGA, and hybrid functionals. The only exception is LDA, which is well-known to overbind noncovalent complexes. The hydrogen-bonded complexes are, however, described rather well by most GGA-based functionals if a small (double- ζ quality) basis set with counterpoise corrections, or a triple- ζ basis set without counterpoise correction, is used. This fortunate error cancellation has likely contributed to DFT's popularity for intermolecular applications on systems with a prevailing polar character. It is worth noting that the differences between individual functionals are rather large. For instance, the PBE functional underestimates binding energies for the dispersion-bonded complexes by about 4.6 kcal/mol (on average), but the BLYP functional underestimates them by more than 7 kcal/mol (both with a large saturated basis). This shows that some kind of functional-dependent scaling for the dispersion correction energy is inevitable. $\Delta\Delta_{D-H}$ is usually around 4 kcal/mol, meaning that the dispersion-bonded complexes are underbound on average by about 4 kcal/mol compared to the hydrogen-bonded complexes. The hybrid functionals seem to be, in most cases, an improvement over their respective nonhybrid version. These large differences between functionals were also found in ref 322, which discusses the performance of a larger number of functionals, along with nine basis sets, for the description of hydrogen bonds.

8.1.2.2. DFT-D Methods. When an empirical dispersion correction is added to a DFT functional, the overall performance improves in all cases.³²³ Both MSE and rmsd errors tend to be smaller not only for the dispersion-dominated complexes, where they drop severalfold, but also for the hydrogen-bonded and mixed complexes. As we will see below, all types of dispersion corrections, being empirical or nonempirical, in general significantly improve the accuracy of DFT for larger molecules. In many cases, like for TPSS-D/LP, B97-D,¹⁵¹ and ω B97X-D,²⁰⁶ the overall accuracy of the DFT-D method is better than the accuracy of the, much more computationally intense, MP2/cc-pVTZ/CP and MP2/CBS methods. This is in part because the MP2 method overestimates dispersion contribution in larger bases sets, while DFT-D does not. Even the less accurate DFT-D combinations show results that are comparable to those of MP2/cc-pVTZ/CP. It is, however, necessary to use at least triple- ζ quality basis sets, because the BSSE is very large in double- ζ bases and results deteriorate progressively. Note that most DFT-D schemes are parametrized without employing the counterpoise correction and should be used this way. Although combinations of the standard DFT functionals with empirical dispersion terms represent remarkable improvements in general, these techniques exhibit systematic overbinding of the hydrogen-bonded complexes and, consequently, an imbalanced description of the relative energies of complexes of different physical natures; this imbalance is measured by $\Delta\Delta_{D-H}$ ¹³⁶ (the last column in Table 2). In terms of $\Delta\Delta_{D-H}$, the best combination with a standard DFT functional is TPSS-D/LP, with a value of about 0.3 kcal/mol; all other functionals give much worse results. This rather serious problem can be mitigated by reparametrizing the DFT functional itself in the presence of the empirical dispersion correction, as was done by Grimme (B97-D)¹⁵¹ and Chai and Head-Gordon (ω B97X-D).²⁰⁶ These functionals are currently probably the most recommendable DFT-based methods for noncovalent complexes.

8.1.2.3. Effects of the Basis Set Superposition Error. The basis set superposition error is generally much smaller in DFT than in the WFT calculations. Still, it is rather large and significantly affects the quality of results for noncovalent complexes. In the double- ζ quality bases, the BSSE amounts to about 20% of the total interaction energy.¹³⁶ When the counterpoise correction is applied, results improve, but errors remain rather large. However, using counterpoise corrections for optimizations is complicated (counterpoise-corrected gradient optimization requires considerably more CPU time), and it cannot be applied in the case of intramolecular BSSE. Therefore, most current DFT-D parametrizations are designed without counterpoise corrections and should be used this way. When the use of a double- ζ basis cannot be avoided, different parameters with down-scaled dispersion should be used.¹³⁶ With double- ζ basis sets, DFT-D accuracy is generally rather poor. This poor description of noncovalent interactions can have large consequences; for example, results for relative conformer energies in small peptides are probably useless due to small energy differences and imbalanced descriptions of different types of interactions.^{78,215} It is recommended to use at least a triple- ζ quality basis. Although BSSE is still non-negligible in triple- ζ bases, it is partially compensated by adding somewhat less dispersion (stronger damping). When almost saturated basis sets are used (of quadruple- ζ quality or plane waves in CPMD), results are usually somewhat worse than for the triple- ζ basis. This can probably

be explained by the improper behavior of current XC functionals and less efficient error cancellation.

8.1.2.4. Reparametrization of Current DFT Functionals. Table 2 also shows results for one of the most popular reparametrizations of a DFT functional from the M06 suite, M06-2X.¹⁸⁵ We will compare these results to those obtained using a very accurate DFT-D method, TPSS-D/LP,¹³⁶ both counterpoise-uncorrected in similar triple- ζ quality basis sets, MG3S and 6-311++G(3df,3pd). We have to keep in mind that the TSPP-D/LP method was parametrized using the S22 set while M06-2X was not. The overall accuracy of both methods seems to be comparable, but there are differences. While the TPSS-D method performs better overall for the dispersion-bonded complexes, M06-2X is better for the hydrogen-bonded complexes, which is also true after optimization of the geometries with the respective methods. Interestingly, while DFT-D tends to overestimate the hydrogen-bonded complexes ($\Delta\Delta_{D-H} = 0.30$ kcal/mol), M06-2X, on the contrary, underestimates hydrogen-bonded complexes with respect to the dispersion-bound ones ($\Delta\Delta_{D-H} = -0.64$ kcal/mol), as do other pure DFT functionals. It appears that M06-2X, as a pure DFT functional, retains a tendency to better deal with the hydrogen-bonded complexes than with the dispersion-bonded ones. This is probably unavoidable, because the long-range behavior of M06-2X is incorrect.²⁴² Nevertheless, M06-2X still provides a relatively accurate description of the weakly bonded complexes, and it is probably the best choice among the pure DFT functionals to date.

8.1.2.5. Double Hybrid Functionals. Table 2 shows results for two double hybrid functionals, mPW2PLYP and B2PLYP.²⁴⁶ Both functionals markedly underestimate the long-range interactions, and in terms of accuracy, they cannot compete with the DFT-D method. Note that these methods were not originally developed for the description of the intermolecular interactions. When a certain portion of the empirical dispersion is added to these double hybrids, like in mPW2PLYP-D and B2PLYP-D,²⁴⁸ the results for intermolecular interactions are much improved (B2PLYP performs significantly better for the complexes with larger dispersion contributions) and comparable to the best DFT-D results.

8.1.2.6. Truly Nonlocal Density Functionals. Truly nonlocal density functionals can recover the correct physical description of the dispersion energy. Table 2 shows results for two such functionals, vdW-DF and LC-BOP+ALL. vdW-DF was derived by Dion et al.,²⁵⁷ and it combines a standard GGA exchange energy (here revPBE²⁵⁸ and PBE³²⁴) with the LDA correlation energy and with the nonlocal correlation energy (NL). It is not necessary to use any damping function here, because damping occurs naturally. Gulans et al.²⁶⁰ implemented vdW-DF in the SIESTA³²⁵ code with self-consistent dispersion density and tested it on the S22 set. The accuracy is comparable to that of the DFT-D scheme, but without the empirical input. Apparently the vdW-DF method overestimates the stabilities of the dispersion-bound complexes relative to the hydrogen-bonded ones, $\Delta\Delta_{D-H} = -2.12$ for revPBE and -1.37 for PBE, in contrast to the opposite trend in DFT-D. It is likely that a large part of the errors originate in the nonoptimal exchange functional and much better results might be achieved if the exchange part was reparameterized using the scheme of Grimme.¹⁵¹ Unfortunately, vdW-DF is bound to use an LDA correlation functional by design, which also somewhat limits its potential

accuracy in many applications. The second example is LC-BOP+ALL,^{254,255} which is based on the ALL nonlocal functional suggested by Andersson, Langreth, and Lundqvist.²⁵¹ Because ALL is divergent at short distances (unlike vdW-DF, it is not seamless), it needs some damping, which is empirical here and quite similar to the interatomic damping in DFT-D. The presence of the adjustable empirical parameters seems to improve the accuracy, as LC-BOP+ALL results are somewhat better than the results of vdW-ALL (see Table 2). Interestingly, the damped ALL term exhibits a systematic flaw similar to that of the DFT-D method; i.e., the hydrogen-bonded complexes are overestimated relative to the dispersion-bound ones ($\Delta\Delta_{D-H} = +1.09$). This is interesting when compared with the vdW-DF functional, which exhibits the opposite trend. The main difference between those two methods is that the ALL nonlocal correlation is always negative while the vdW-DF nonlocal correlation can also be positive (repulsive) in the shorter ranges. The repulsive contribution of vdW-DF probably represents just that part of repulsion that otherwise could only be added by reparametrization of the GGA functional.^{151,206} Overall, the nonlocal DFT functionals seem to perform relatively well with errors comparable to those of the empirical dispersion methods, but at the expense of non-negligible additional time requirements. Because of their firm physical groundings, the nonlocal functionals are likely to be significantly improved and may become very important in the near future.

In conclusion, currently there are several DFT-based models able to provide fairly accurate descriptions of the noncovalent complexes (statistically speaking). However, could one of them become a method of choice for benchmarking in the future? As of now, the most accurate methods, such as DFT-D, contain empirical parameters. Although they may be extremely useful in molecular modeling, they are not good candidates for benchmark calculations because of their empirical nature. On the other hand, the nonempirical methods such as vdW-DF do not seem to be accurate enough for benchmarking purposes yet. Here, inaccuracies arise mainly in the overlap region and stem from the double counting of correlation effects and from the nonoptimal adjustment of the long-range correlation part to the short-range correlation and/or exchange. We believe that if this issue can be solved with minimal empiricism and good accuracy, the DFT-based methods can achieve the accuracy and robustness necessary for benchmarking in the near future.

8.1.3. Symmetry-Adapted Perturbation Theory

Table 2 shows DFT-SAPT interaction energies calculated using the aug-cc-pVDZ and aug-cc-pVTZ basis sets and results extrapolated to the complete basis set limit. The density functional used (PBE0AC) and other details are as in ref 129, and we used the Molpro 2006 ab initio package.²⁹⁰ The overall performance of the DFT-SAPT method with the aug-cc-pVDZ basis is relatively good, and the overall error is similar to that of the MP2/cc-pVTZ/CP calculation. In the aug-cc-pVTZ basis the errors are only slightly worse than the errors of the CCSD method in a large basis set (see seven rows above in the same table). Now we can extrapolate the dispersion component to the complete basis set limit and add it to the sum of the nondispersion components calculated in the aug-cc-pVTZ basis, similarly to that in ref 129. The results obtained in this way (the third DFT-SAPT row) are only slightly better than those in the aug-cc-pVTZ basis set, and the interaction energies remain on average underesti-

mated. This indicates that a yet larger basis set is needed for DFT-SAPT calculations. If we extrapolate also the nondispersion components and apply a somewhat more progressive extrapolation to the dispersion component (results not shown), almost all errors drop below 0.3 kcal/mol. Therefore, we believe that if a larger basis set was used, the quality of the DFT-SAPT results might be comparable to that of the CCSD(T)/CBS reference. We can conclude that the DFT-SAPT method in the aug-cc-pVTZ basis provides relatively reliable and accurate intermolecular interaction energies and it is very likely that if a better basis set was used, DFT-SAPT might compete with the quality of the wave-function-based reference for molecules within the S22 set.

8.1.4. Semiempirical QM Methods

Table 2 shows that all semimepical methods of the standard form provide poor results for noncovalent complexes. The AM1 method gives the largest average error for the entire set, while the more advanced methods give errors that are around half those of AM1. Investigating the performance of all methods for subclasses of the S22 set, we find the largest errors occur for the dispersion-bound complexes. AM1 also gives the largest average error (~ 11 kcal/mol) for these complexes, with other semiempirical methods yielding errors for this interaction type that are also very (unacceptably) large. Reviewing the performances of semiempirical methods for H-bonded complexes, we find, surprisingly, that here also these methods fail and provide very large errors.

Dramatic improvements in overall performance result when dispersion corrections are included, with the best results (by far) being obtained with the OM2-D technique, which produces an MSE that is close to zero. Performing deeper analyses, we realize, however, that the method has rather large positive errors for H-bonded complexes and rather large, but negative, errors for dispersion-bound complexes, a trend that is reflected in the large $\Delta\Delta_{D-H}$ error (-2.3 kcal/mol). Let us note here that this error is substantially smaller for the parent OM2 method. The smallest rmsd error among all dispersion-augmented methods is again for OM2-D, but for the reason discussed above the method cannot be recommended for general use. Investigating the performance of these dispersion-corrected methods for individual subclasses, we find significant improvements over the pure semiempirical methods for dispersion-bound complexes while improvement for H-bonded systems is only slight. This is the main point that led us to introduce a second correction that treats H-bonding. The joint dispersion + H-bonding correction (designated by a “-DH2” extension) improved the performance of each of the parent semiempirical methods substantially. The average error for the entire S22 set was reduced from 3.4 to 0.1 for PM6, from 3.6 to -0.1 for OM3, from 6.8 to -0.2 for AM1, and from 3.6 to 0.4 for SCC-DFT-B (all values are in kilocalories per mole). Even more important is the fact that the average errors are now small for each of the subclasses. The lowest error values were observed for PM6-DH2, whose well-balanced treatment of interaction types is reflected by the method’s very small $\Delta\Delta_{D-H}$ value, which is close to zero. Evidently the proposed scheme outperforms all other existing H-bonding- and dispersion-energy-corrected semiempirical methods by a significant margin and yields results comparable to those of the current WFT and DFT techniques for a large number of

investigated cases. This finding can be demonstrated by the values of the mean unsigned error (with respect to the benchmark CCSD(T)/CBS interaction energy, all values in kilocalories per mole) for various WFT, DFT, and semiempirical methods: MP2/CBS, 0.8; DFT-D/TPSS/TZVP, 0.6; DFT-D/TPSS/6-311+G(3df,2pd), 0.3; M06-2X/6-311+G(3df,2p), 0.4; SCC-DF-TB-DH2, 1.2; AM1-DH2, 0.7; OM3-DH2, 0.6; PM6-DH2, 0.4.

The dispersion corrections to the AM1, OMX, and PM6 methods were parametrized against the S22 set or its subsets. It is thus important to test the performance of these methods against binding energies of other noncovalent complexes not included in the parametrization set. The MSE error (kcal/mol) for the noncharged H-bonded complexes in the JSCH2005 set (37 entries) and H-bonded peptides (13 entries) from the same data set are as follows (for various methods): DFT-D/TPSS/TZVP, 0.4, -0.4; OM3-DH2, -0.8, 0.4; AM1-DH2, 0.1, 1.5; SCC-DF-TB-DH2, 2.3, -0.7; PM6-DH2, -0.7, -0.1. Evidently, all of the semiempirical methods perform well, and their performance is comparable to those of considerably more expensive DFT methods. Among all the semiempirical methods tested, the best results were obtained with PM6-DH2. When the geometries of the investigated complexes are optimized, PM6-DH2 provides comparable or even better results. This point is very important since it indicates that the method can be used in both geometry optimization and molecular dynamic simulations. Let us add that this point is also critical with other semiempirical QM methods. Geometry optimizations frequently lead to very distinct structures, considerably different from the starting ones; binding energies for these optimized structures are often unrealistically large.

8.2. Geometry Characteristics

The stabilities of most noncovalent interactions depend heavily on geometric parameters, a property that is necessary for the fine-tuning and diversity of biochemical processes. Most investigations concerned with the accurate characterization of noncovalent interactions in biomacromolecules have focused on obtaining accurate binding energies either at the potential energy minimum (as determined at some lower level of theory) or using experimentally derived complex structures (such as those obtained from X-ray crystal structures). It has long been believed that energies are more sensitive to the theoretical level used than geometries and, consequently, that lower level theoretical approaches can be applied to obtain accurate minimum energy geometric structures. Here we concentrate on studies characterizing the dependence of noncovalent interactions on geometrical parameters, and it will be shown that it is necessary to apply high-level theoretical methods to obtain accurate geometries. Continuing with one of the major themes of this review, we will largely focus on the performance of various methods compared with high-level reference data, such as those obtained using the CCSD(T)/CBS method. There are several types of studies that should be considered here: (a) studies involving potential energy curves for molecular complexes, (b) studies involving gradient optimization of molecular complexes, and (c) studies involving conformational analyses on systems whose structure is largely governed by noncovalent interactions.

There are many reasons that it is important to characterize the geometrical dependence of noncovalent interactions and not only focus on binding energies of minimum energy

structures. Despite the fact that potential energy curves for noncovalent interactions are shallower than those for covalent bonds, the noncovalent interactions are still very sensitive to geometric parameters and their strengths can often vary significantly with small geometric perturbations. This geometric sensitivity can have a tremendous influence on the structures and stabilities of proteins and nucleic acid compounds (DNA/RNA) and may be a large factor in determining whether a ligand (such as a hormone or pharmaceutical compound) successfully binds to a protein receptor. The dynamic properties of biomolecules and nanomolecular structures can be strongly affected by geometrical noncovalent properties, and formulating a deeper understanding of the behavior of these interactions as a function of spatial parameters can give us insights into the dynamics of many types of systems. One case where the accurate treatment of interactions at all points on the potential energy surface is very important is in the interpretation of vibrational (infrared) spectra of peptides, proteins, and nucleic acid compounds. Studying the geometry dependence of noncovalent interactions at a very high level for a variety of interaction types is important because it establishes reference data that can be used to assess the performance of lower level methods in terms of their ability to converge to the geometric energy minimum. This last point is very important because the structures obtained at these lower levels are often used for high-level binding energy analyses and because lower level theory is often used to obtain theoretical infrared spectra, which can potentially be useful in assigning peaks in experimentally obtained spectra. Finally, although interactions at long ranges are generally very weak, they can play a key role in complexes of extended systems where the number of contacts at these distances grows very quickly; this point is particularly significant for complex molecular systems.

There are four principal types of noncovalent interactions that play the largest roles in biomolecular and nanomolecular structures; they are hydrogen bonding, dispersion interactions, stacking interactions, and $X-H\cdots\pi$ interactions. The first type of interaction, H-bonding, is characterized by extreme sensitivity to geometric parameters because it relies not only on a proper dipole-dipole alignment between polar species but also on favorable overlap between donor and acceptor orbitals in the charge-transfer (hyperconjugation) process. In a similar fashion, $X-H\cdots\pi$ interactions, which can conceptually be considered to represent dipole-quadrupole interactions, can also be quite strongly influenced by the relative orientation of the interacting systems. Stacking interactions are defined to be stacked arrangements of aromatic moieties and, of course, must have configurations such that there is a high amount of overlap area between the interacting species. It should also be noted that the strengths of stacking interactions involving heterocyclic aromatic species are heavily influenced by the relative orientation of the subsystem's dipole moments. Dispersion, or van der Waals, interactions represent a class of noncovalent interactions that are geometrically nonspecific; that is to say, they do not depend heavily on the relative orientation of the monomers such as in the case of hydrogen bonds. Although these types of interactions are weaker than others, they are very important in biomolecular and nanomolecular structure because of their pervasiveness throughout the structures of proteins, DNA, other biostructures, and nanostructures. It should be mentioned here that very

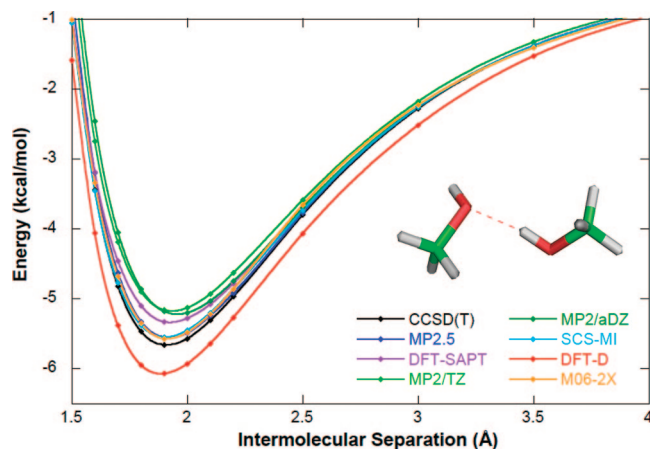


Figure 2. Potential energy curves for displacement of two H-bonded methanol molecules.

frequently stacking interactions are identified as being dispersive in nature. It is true that generally the dominant part of the stabilization in, for example, stacked nucleic acids, is due to dispersion interactions, but electrostatic (dipole-dipole) forces also play an important role in these types of interactions and are responsible for the orientation of the bases in a dimer. Dispersion interactions by themselves are important, for example, in complexes of aliphatic hydrocarbons (or in isolated hydrocarbons), where they largely determine not only the stabilization energy but also the structure of a complex.

8.2.1. Potential Energy Curves

Obtaining potential energy curves of interacting systems represents the simplest technique for analyzing the dependence of an interaction on geometric parameters. These curves are generated in one dimension along some particular geometric coordinate, usually in the direction of direct complex dissociation. It is relatively easy to compare the results of various methods for the computation of potential energy curves as the properties of these curves can generally be analyzed by simple inspection. There have been a number of studies carried out within the past several years in which high-quality potential energy curves for intermolecular interaction are produced.^{17,20,48,51,124,230,243,326–349} Of note are studies generating potential energy curves using the CCSD(T) method, either at the (estimated) CBS or using basis sets at least as large as aug-cc-pVTZ, by Hobza,^{48,326–328} Sherrill,^{17,51,124,329–336} Tsuzuki,^{337–340} Truhlar,^{243,341} Rothlisberger,^{230,348} and others.

In a recent study we compared the performance of several computational methods in generating potential energy curves for the most relevant biological interaction motifs (H-bond, stacking, dispersion, and $X-H\cdots\pi$).³²⁶ In this work reference data were obtained at the estimated CCSD(T)/CBS level and the other methods tested were estimated, MP2.5/CBS, DFT-SAPT/aug-cc-pVTZ, MP2/cc-pVTZ, MP2/aug-cc-pVDZ, MP2/6-31G*(0.25), SCS(MI)-MP2/cc-pVTZ, DFT/M06-2X/6-311+G(2df,2p), and DFT-D/TPSS/6-311++G(3df,3pd). Here several structures were studied; they are methanol dimer (H-bond; see Figure 2), methylamine dimer (H-bond), formamide dimer (H-bond), adenine \cdots benzene (stacked), cytosine \cdots benzene (stacked; see Figure 3), propane dimer (dispersion; see Figure 4), and benzene \cdots water ($X-H\cdots\pi$; see Figure 5). All of the methods, with the exceptions of DFT-D and DFT-SAPT, were used along with the counter-

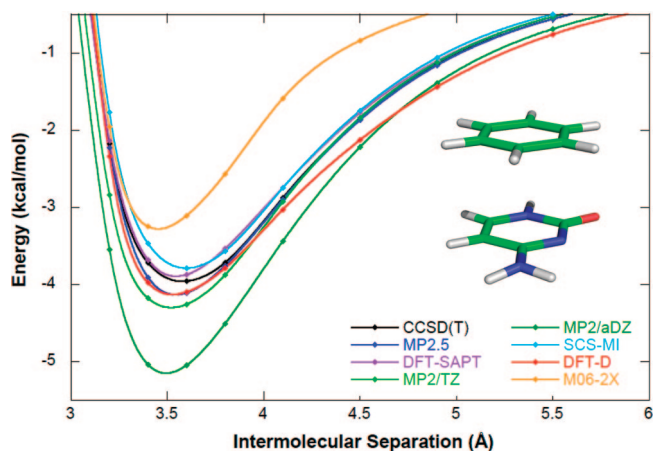


Figure 3. Potential energy curves for displacement of the stacked benzene \cdots cytosine complex.

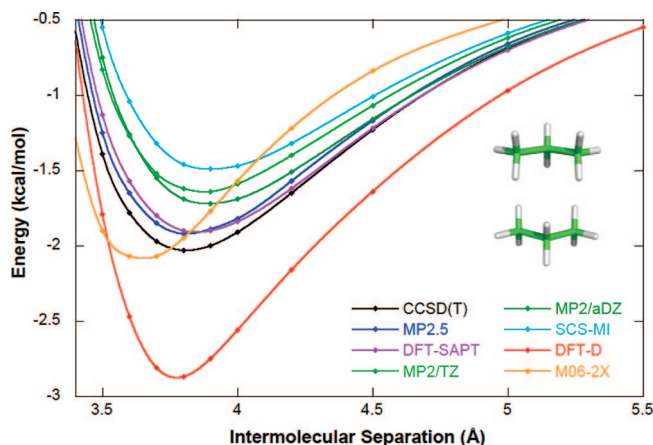


Figure 4. Potential energy curves for displacement of two propanes.

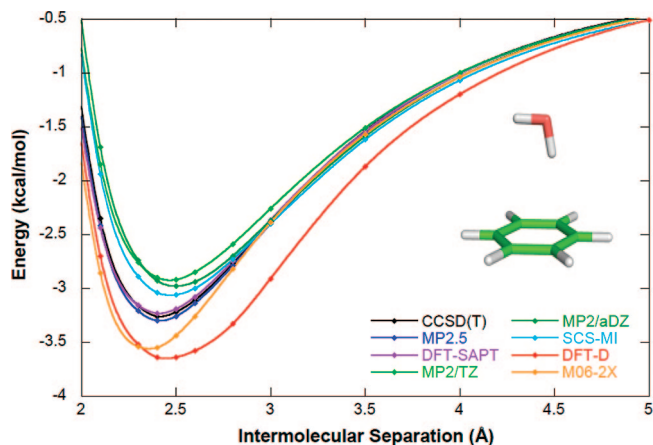


Figure 5. Potential energy curves for displacement of the $O-H\cdots\pi$ benzene \cdots water complex.

poise correction scheme to correct for the basis set superposition error.

It was found in this study that each of the computational methods is capable of describing all of the considered interaction motifs at least at a qualitative level. This is an especially impressive result for the two DFT-based methods DFT-D and M06-2X when one considers that just a few years ago it was widely acknowledged that existing DFT methods were incapable of describing noncovalent interactions whose attractive forces are largely attributable to dispersion. Only two of the tested methods, MP2.5 and DFT-SAPT, can be

said to provide accurate potential energy curves for all of the complexes considered. Unfortunately, these two methods are computationally very expensive and can only be used on complexes containing relatively few atoms (up to ~60–80). Apart from its computational expense, there are two other disadvantages of the DFT-SAPT method. First, the analytic gradients needed for optimization of geometries have not yet been implemented for this technique. Second, it is not easily possible to include the deformation energy in DFT-SAPT potential energy surface calculations; hence, only rigid monomers can be considered. Following is a brief description of the results obtained for the various types of interaction motifs.

Among all of the interaction types, H-bonds are the easiest to characterize; as such it is not surprising that they are generally well described by all of the tested computational methods (cf. Figure 2). MP2 combined with both the cc-pVTZ and aug-cc-pVDZ basis sets has a strong tendency to underbind these types of interactions but provides reliable geometries. MP2/6-31G*(0.25) has an even greater tendency to underbind the H-bonding complexes and also gives minimum energy separations that are consistently too large. The use of the spin component scaling technique (SCS(MI)-MP2) with the cc-pVTZ basis improves the performance of MP2, but produces binding energies that are still slightly too low. The estimated CCSD(T) binding energy curves for H-bonding complexes is very closely matched by those produced using the MP2.5, M06-2X, and (to a lesser extent) DFT-SAPT techniques. In general, the DFT-D method overestimates the H-bond strengths (sometimes strongly).

In terms of single-point binding energies, it has been observed that the MP2 method, when used with medium-sized basis sets, generally tends to overbind stacked structures;¹⁰⁴ this overbinding tendency is also seen for the potential energy curves of the adenine...benzene and cytosine...benzene complexes (cf. Figure 3). The trend of overbinding is especially apparent for the MP2/aug-cc-pVDZ method. When the MP2 method is used along with the 6-31G*(0.25) basis, the results for stacking interactions are greatly improved, and in the case of the cytosine...benzene complex, MP2/6-31G*(0.25) binding energies are actually slightly underbound. The SCS(MI)-MP2 method was developed to improve the performance, in terms of binding energies, of MP2, and one of the particular goals was to correct the overbinding of stacked structures. For the two stacked structures considered, this technique is largely successful, reducing the amount of overbinding significantly and, in the case of the cytosine...benzene complex, actually underbinding at the potential energy minimum. In terms of the locations of the potential energy minima, all calculations based on MP2 produce results consistent with CCSD(T) data, allowing for the structural optimization of various complexes at a rather inexpensive level, although it should be noted that the counterpoise method, which is necessary to correct for the BSSE, makes these computations more time-consuming. The cc-pVTZ basis set generally yields better geometries than aug-cc-pVDZ. This point is important since it is frequently believed that, since MP2/aug-cc-pVDZ provides larger stabilization energies, it should also provide better geometries. The MP2.5, DFT-SAPT, and DFT-D methods all yield potential energy curves that are very similar to those generated using the estimated CCSD(T) procedure. Among the tested methods, DFT/M06-2X produces the worst potential energy curves for stacked structures, being strongly

underbound for the cytosine...benzene complex, predicting incorrect minimum energy separations for both complexes, and generally producing curves with the wrong overall shape.

The propane dimer was the sole representative from the class of aliphatic dispersion-bound complexes. The MP2.5 and DFT-SAPT techniques, which are both relatively expensive, are the only methods that produce high-quality potential energy curves for this complex (cf. Figure 4). All of the MP2 techniques, including SCS(MI)-MP2, tend to strongly underestimate the binding energy of the propane dimer and produce minimum energy separations that are slightly too large. This underbinding tendency is especially pronounced for the MP2/6-31G*(0.25) method, with a minimum binding energy that is roughly half that of the reference value. Conversely, DFT-D very strongly overestimates the binding energy of this complex while giving a relatively accurate minimum energy separation. One of the likely reasons for this strong overbinding is the fact that the S22 test set used to parametrize DFT-D contains many complexes with sp^2 -hybridized carbons (as in aromatic systems) but very few sp^3 -hybridized carbons (as in aliphatic systems). Nevertheless, this problem is already fixed in some newer versions of DFT-D.^{208,350} It should be noted that the performance of the SCS(MI)-MP2 method, which was also parametrized against the S22 test set, likely suffers from the under-representation of sp^3 -hybridized carbons as well. DFT/M06-2X produces a reasonable binding energy value at the potential energy minimum, but gives a minimum energy separation that is far too small.

For the benzene...water complex, which represents the X-H... π binding motif, MP2.5 and DFT-SAPT are, once again, the only methods that produce reliably accurate results relative to those of estimated CCSD(T) (cf. Figure 5). The MP2 methods tend to underbind the benzene...water complex, as they do the hydrogen-bonding complexes and the propane dimer. All of the MP2 methods also give minimum energy intermolecular separations that are slightly too large. As for the hydrogen-bonding and dispersion-bound complexes, the MP2/6-31G*(0.25) method produces the worst MP2 results for the benzene...water complex. Conversely, both DFT-based methods tend to overbind the complex. DFT/M06-2X predicts the potential energy minimum to be at slightly too small a separation, while the minimum energy point produced by DFT-D is at slightly too large a separation, with a potential energy curve that appears to be much too broad near the bottom of the potential energy well.

The MP2 method has long served as the “workhorse” for computations on molecular complexes, traditionally being used with relatively small bases sets, such as 6-31+G* and 6-31G*(0.25), or medium-sized basis sets, such as aug-cc-pVDZ, cc-pVTZ, and aug-cc-pVTZ. In terms of potential energy curves, it was observed that the MP2 technique, when combined with the medium aug-cc-pVDZ and cc-pVTZ basis sets, produces semiquantitatively accurate results. MP2/6-31G*(0.25) results are very good for stacking complexes but are generally only qualitatively correct for all other binding motifs. The best MP2 results are given with the cc-pVTZ basis set, which is in good agreement with previous findings.³⁵¹ The SCS(MI)-MP2/cc-pVTZ method, which seeks to improve the results of MP2/cc-pVTZ, is largely successful in this task, with improved potential energy curves for all of the noncovalently bound complexes except the propane dimer. Among the DFT-based methods, which are much less

computationally expensive, DFT-D yields the best performance, giving accurate potential energy curves for hydrogen-bonding and stacking interactions, but strongly overbinding both the propane dimer and the benzene...water complexes. The M06-2X functional produces good results for hydrogen bonding and O-H... π interactions but produces curves with the wrong overall shape for both stacked and dispersion-bound complexes.

In a recent work Pitoňák et al. described both the (cyclic) hydrogen-bonding and stacking potential energy curves for the uracil dimer, the smallest nucleic acid complex, at various levels of theory including the estimated CCSD(T)/aug-cc-pVTZ level.⁴⁸ One of the main findings made in this study is that the SCS(MI)-MP2/cc-pVTZ, DFT/M06-2X/6-311+G(2df,2p), and DFT-D/TPSS/6-311++G(3df,3pd) methods produce potential energy curves for these interactions that are at least semiquantitatively accurate. The SCS(MI)-MP2 technique yielded particularly accurate results for both hydrogen-bonded and stacked systems, while the results obtained with the DFT-D and M06-2X methods were substantially better for the hydrogen-bonded complex than for the stacked one.

Sherrill and co-workers have conducted many studies characterizing potential energy curves for weakly bound complexes, with particular emphasis on aromatic...aromatic and aromatic...aliphatic interactions.^{17,51,124,329-336} Among the types of interactions that have recently been treated by these researchers are various configurations of the (substituted and unsubstituted) benzene dimer,^{17,332,334,335} the H₂S...benzene complex,³³⁶ and the methane...benzene complex.³³¹ In recent studies they have investigated the effects of aromatic substituents on the parallel-displaced conformation of the benzene dimer and the effects of aromatic nitrogen heteroatoms by studying several conformations (including stacked, T-shaped, and parallel-displaced) of the benzene dimer,^{17,51,124,329,332-335} benzene...pyridine complex,³³⁰ and pyridine dimer.³³⁰

In a recent study Takatani and Sherrill investigated the performance of the MP2, SCS-MP2, SCSN-MP2, SCS-LMP2, and SCSN-LMP2 methods for potential energy curves of the sandwich benzene dimer, T-shaped benzene dimer, methane...benzene, H₂S...benzene, and methane dimer complexes.¹²⁴ Each of the methods was used along with the aug-cc-pVxZ basis sets ($x = D, T, Q$) and complete basis set extrapolation derived from the aug-cc-pVTZ and aug-cc-pVQZ basis sets. Reference values in this study were obtained using the estimated CCSD(T)/CBS method. All nonlocal methods employed in this study, CCSD(T), SCS-MP2, SCSN-MP2, and MP2, were used along with the CP corrections to account for BSSE; the local methods, SCS-LMP2 and SCSN-LMP2, did not utilize the CP correction, as local methods exhibit very little BSSE. The fact that local methods can be used without CP corrections has implications in studies of intramolecular noncovalent interactions, for which CP corrections are not possible. It was found that each of the spin-correlated methods, when used along with the aug-cc-pVTZ or larger basis sets, produces potential energy curves that are in good agreement with those of the CCSD(T)/CBS method for all of the considered systems, with the exception of the methane dimer. All of these methods produced large improvements over MP2 for all of the complexes involving benzene. The spin-correlated methods did not yield accurate potential energy curves when used with the aug-cc-pVDZ basis. On average, the best results

were produced with the SCSN-MP2 and SCSN-LMP2 methods, with binding energy errors of no more than 6%, when used with aug-cc-pVTZ, for all systems except the methane dimer. Although the spin-correlated methods perform very well for complexes containing aromatic species, it should be noted that their poor performance for the methane dimer may be indicative of problems for other aliphatic systems. Another interesting finding in this study is that the SCSN-LMP2 result seems to converge quickly with basis set size, with aug-cc-pVTZ and aug-cc-pVQZ binding curves for the methane...benzene complex that are nearly coincident. This result indicates that the BSSE is effectively eliminated using this local method when the aug-cc-pVTZ basis set is used.

The effects of heterocyclic aromatic atoms in various configurations of the pyrimidine...benzene complex and pyrimidine dimer were thoroughly investigated in a study by Hohenstein and Sherrill.³³⁰ In this study high-level potential energy curves were generated at the estimated CCSD(T)/aug-cc-pVTZ and CCSD(T)/CBS levels. The insertion of nitrogen, which has an electronegativity higher than that of carbon, into an aromatic ring introduces a dipole moment into the aromatic system and also tends to reduce its polarizability. The addition of this electronegative heteroatom into a ring has large effects for stacked, parallel-displaced, and T-shaped dimers. For stacked systems it is found that interactions become stronger (compared to those of the benzene dimer) for the pyrimidine...benzene and antialigned pyrimidine dimer; this is due to attractive dipole-induced dipole and dipole-dipole interactions. For the aligned pyrimidine dimer (i.e., with the nitrogens on top of one another) the interaction becomes weaker because of a repulsive dipole-dipole interaction. The effects of heteroatoms on the interactions for T-shaped and parallel-displaced systems depend strongly on the relative orientations of the benzene and pyrimidine monomers; please see ref 330 for an in-depth analysis of these interactions. The performances of SCS-MP2 and SCSN-MP2, along with the aug-cc-pVTZ basis, were investigated for their abilities to reproduce accurate potential energy curves for the parallel-displaced pyrimidine...benzene and the H-bonding pyrimidine dimer. Curves for the parallel-displaced system were computed as a function of the horizontal displacement, and it was found that SCS-MP2 reproduces the reference data very well, while SCSN-MP2 is significantly overbound (by up to ~ 0.5 kcal/mol) in many places along the curve. It was also found that SCS-MP2 reliably reproduces accurate binding energies and intermolecular separations for all of the stacked systems considered in the study. The good performance of SCS-MP2 for parallel-displaced systems is consistent with the results of a study by Arnstein and Sherrill,³²⁹ where it was found that this method performs well for parallel-displaced configurations of the benzene dimer, fluorobenzene...benzene complex, and nitrobenzene...benzene complex. In contrast to the results for parallel-displaced and stacked complexes, SCSN-MP2 was found to produce accurate potential energy curves for the H-bonded configuration of the pyrimidine dimer, while SCS-MP2 significantly underbinds this interaction (by ~ 0.7 kcal/mol at the minimum). This result is consistent with the findings of other studies where it has been shown that SCS-MP2 generally underbinds H-bonding interactions.

Sherrill et al. generated very high quality potential energy curves for the three configurations of the benzene dimer

(stacked, parallel-displaced, and T-shaped), benzene...methane, benzene...H₂S, and the methane dimer using an estimated CCSD(T) procedure in which the Δ CCSD(T) term was extrapolated to the complete basis set limit (from modified aug-cc-VTZ and aug-cc-pVQZ basis sets with no diffuse functions on hydrogens).⁵¹ The performances of several methods, MP2, SCS-CCSD, SCS-MP2, SCSN-MP2, SC-S(MI)-MP2, M05-2X, M06-2X, B3LYP-D, and PBE-D, were evaluated relative to the new benchmark data. It was found that each of these methods provided a reasonable description of the interactions for each of the complexes, with the exception of the methane dimer, for which many of the methods were strongly underbound. The SCS-CCSD method provided excellent agreement with reference potential energy curves, although it should be mentioned that the computational cost for this technique is quite high (the same as that for conventional CCSD). Another focus of this study was to characterize the errors associated with the basis set used to compute the Δ CCSD(T) term for the three benzene dimer configurations. Considering results obtained with the counterpoise correction, it was found that the, relatively small, cc-pVDZ basis produces errors up to approximately 0.5 kcal/mol. The aug-cc-pVDZ basis, which is very commonly used for Δ CCSD(T) terms, was found to produce errors of up to about 0.1 kcal/mol.

Tsuzuki and co-workers recently studied the binding found within the nitrobenzene...benzene complex and the nitrobenzene dimer.³⁴⁰ Included in their study were high-level estimated CCSD(T)/CBS potential energy curves of the T-shaped nitrobenzene...benzene complex and the parallel-displaced nitrobenzene dimer, and the performances of the HF and MP2 methods for these potential energy curves are compared to the reference data. As might be expected, HF is heavily underbound throughout the curves and, in the case of the parallel-displaced complex, produces a maximum binding energy of less than 1.0 kcal/mol, which is to be compared to an estimated CCSD(T)/CBS binding energy of approximately 6.75 kcal/mol. MP2, on the other hand, strongly overbinds both of these interactions throughout the curves, which is also consistent with results described above.

Zhao and Truhlar have investigated the performance of several DFT functionals in terms of their abilities to produce accurate potential energy curves for the benzene...methane complex; included in these studies were functionals from the newly developed M05 and M06 functional suites.^{243,341} It was found that "traditional" functionals, including B3LYP, BLYP, and TPSS, failed to describe this interaction with any reliability (in some cases being completely unbound) while the M05- and M06-type functionals gave much more accurate potential energy curves relative to estimated CCSD(T)/CBS data. M06-2X, M06-HF, and (especially) M05-2X yielded particularly good results for this complex.

In a recent study Sponer and co-workers produced potential energy curves near the potential energy minima for several configurations of the uracil dimer using several electronic structure methods (including estimated CCSD(T)) and an empirical-potential-based method.³²⁷ Altogether more than 100 structures were investigated, with both long-range and close-contact DNA structures being considered. For these complexes it was observed that the DFT-D, DFT-SAPT, and SCS(MI)-MP2 methods all generated curves that were in very good agreement with reference data. A very similar study on a large number of adenine dimer structures was performed, and similar conclusions were obtained.³⁵²

Extremely high quality geometries and energies for the benzene dimer in various configurations have also been computed by Janowski and Pulay.²⁰ In this study several geometric parameters of the stacked, parallel-displaced, and T-shaped benzene dimer were manually optimized at the (explicit) QCISD(T)/aug-cc-pVTZ level (QCISD(T) being slightly more efficient but comparably accurate relative to CCSD(T)). Binding energies for the optimized systems were obtained at the QCISD(T)/CBS level, with several methods being used for CBS extrapolation. It is found that, at the highest levels of theory used here, the T-shaped conformation represents the minimum energy structure for the benzene dimer, but is only about 0.02 kcal/mol more stable than the parallel-displaced form of the dimer. In this study MP2/aug-cc-pVxZ and SCS-MP2/aug-cc-pVxZ binding energies for all benzene dimer forms are obtained at the QCISD(T)/aug-cc-pVTZ minima. It is found that both MP2 and SCS-MP2 favor the parallel-displaced configuration, with MP2 strongly favoring this structure (by about 1 kcal/mol) and SCS-MP2 favoring it less strongly (by about 0.4 kcal/mol).

Tekin and Jansen produced CCSD(T), DFT-SAPT, MP2, and SCS-MP2 (aug-cc-pVTZ for all methods) potential energy curves for several configurations of the acetylene...benzene complex, with interaction modes of both the stacked and CH... π types.³⁴⁷ It was found that DFT-SAPT reproduces the CCSD(T) data extremely well for these interactions. MP2, as might be expected with this relatively large basis set, strongly overbinds for all of these interactions. SCS-MP2 gives better potential energy curves than MP2, especially for interactions with π - π character, underbinding for all complexes.

Rothlisberger and co-workers have been heavily involved in the development of dispersion-corrected atom-centered potentials (DCACPs), which are used along with DFT methods.^{230,348} These investigators have carried out several studies comparing potential energy curves obtained using DCACPs to CCSD(T)/aug-cc-pVTZ data. In a recent study two variants of the B3LYP-based DCACPs, DCACP-MP2 and DCACP-CCSD(T) (where MP2 and CCSD(T) refer to the calibration methods), were used to produce potential energy curves for the ethane dimer, ethene dimer, ethyne dimer, and ethene...methane complex; these curves were compared to reference MP2/aug-cc-pVTZ and CCSD(T)/aug-cc-pVTZ results.²³⁰ One very interesting aspect of this study is that it includes three different hybridization states of carbon; it has been seen that the correct characterization of these hybridization states is very difficult to accomplish using methods incorporating empirical potentials. It was found in this study that both variants of the DCACPs produced potential energy curves that were in good agreement with their respective reference curves (DCACP-MP2 is compared to MP2, and DCACP-CCSD(T) is compared to CCSD(T)). For the DCACP-CCSD(T) method, the largest geometrical error occurred for the ethene dimer, whose optimum intermolecular separation was predicted to be too large by 0.19 Å. The largest energetic error for this method occurred for the ethane dimer, which was overbound by about 0.3 kcal/mol.

In an investigation by Merz and co-workers, potential energy curves for 20 of the 22 molecular complexes in the S22 data set are computed at the estimated CCSD(T)/CBS level of theory³⁵³ (see also the note added in proof). Two of the main objectives of this study are to characterize the MP2 extrapolation of binding energies from different basis sets

(with and without counterpoise corrections) and to assess the quality of several lower level methods for complex geometries and binding energies. It is found that, when counterpoise corrections are used, extrapolation of MP2 binding energies from the aug-cc-pVDZ and aug-cc-pVTZ bases results in only small errors compared to the aug-cc-pVTZ/aug-cc-pVQZ extrapolation. When the counterpoise corrections are not used, however, extrapolated binding energy values still contain some BSSE, even when the aug-cc-pVTZ/aug-cc-pVQZ basis sets are employed, as these binding energies converge much more slowly to the CBS as a function of basis set size. The use of noncounterpoise-corrected energies for extrapolation from the aug-cc-pVDZ/aug-cc-pVTZ basis sets leads to large errors compared to those of the larger basis extrapolation, and results in a greater amount of BSSE. Several types of computational techniques were tested for their abilities to describe geometric and energetic properties of molecular complexes, including DFT and semiempirical methods. Among the DFT methods tested, it was found that functionals from the M06 family give relatively accurate potential energy curves, with the M06-L functional being particularly promising because it lacks an exact exchange term and is, thus, computationally more efficient than hybrid GGA methods. As expected, semiempirical methods proved to give poor performances for all interaction types.

One of the first studies investigating the geometries of extended molecular clusters originates from our laboratory and was published in 2005. The geometries and interaction energies of stacked uracil dimers and an adenine...thymine pair were studied by means of high-level quantum chemical calculations, including estimated CCSD(T)/CBS by Dabkowska et al.³⁵¹ The results can be summarized as follows: (i) Standard MP2 geometry optimization with small basis sets (i.e., 6-31G*) provides fairly reasonable intermolecular separations (this is mainly attributable to error compensations). (ii) Geometry optimization with extended basis sets at the MP2 level underestimates the intermolecular distances compared to reference CCSD(T) results, whereas the MP2/cc-pVTZ counterpoise-corrected optimization agrees well with the reference geometries and, therefore, is recommended for improvement of geometries. (iii) The self-consistent charges density functional tight binding method, with the inclusion of the empirical dispersion energy, accurately reproduces interaction energies and geometries of dispersion-bound stacked complexes, and this method can thus be recommended for prescanning the potential energy surfaces of noncovalent complexes.

8.2.2. Analytical Gradients

Boese and co-workers performed very high level analytic gradient geometry optimizations on a large set of hydrogen-bonded complexes using the CCSD(T) method along with a basis set in which non-hydrogen atoms are described using the aug-cc-pVQZ basis and hydrogens are described using cc-pVQZ (referred to as the A'VQZ basis).⁴⁶ It should be noted that counterpoise corrections were not employed. The set of complexes used contains both neutral and charged species and reflects a large range of binding energies, spanning from 1.72 kcal/mol (CO...HF) to 33.72 kcal/mol (H₃O⁺...H₂O). Reference binding energies for these complexes were determined using the Weizmann-2 method. One of the main focuses of this study was to determine the accuracy that can be expected of several commonly used

DFT methods in computing hydrogen-bonding geometries and binding energies. Several GGA, meta-GGA, and hybrid GGA functionals, as well as MP2 (with A'VTZ), were tested in terms of their computed hydrogen bond distances, hydrogen bond shifts (affecting the Y-H bond length), and binding energies. Please see the reference paper for a description of the triple- ζ basis set used for the DFT computations.

Among the methods tested, MP2, as might be expected, proved to give the best values for all H-bond properties considered in the study. Among the DFT methods the best results for all of the H-bond properties were obtained using the hybrid functionals, B3LYP and B97-1, with B3LYP giving the best results. Interestingly, the rms error (over the entire test set) for B3LYP binding energies is very similar to that of MP2 (0.45 and 0.41 kcal/mol, respectively), but inspection of the individual binding energies shows that the range of errors for the DFT method (spanning 1.8 kcal/mol) is much greater than that of MP2 (spanning 0.9 kcal/mol). B3LYP also produces relatively accurate H-bond distances, with an rms error of 0.27 Å (the rms error for MP2 is 0.26 Å). The results obtained with most of the GGA (BP86, BLYP, HCTH/407) and meta-GGA (PBE and TPSS) methods are rather disappointing, producing binding rms errors between 0.7 and 1.5 kcal/mol and H-bond distance errors between 0.41 and 1.05 Å. It should be noted that the HCTH/407 functional yields relatively accurate binding energy results but gives the worst bond lengths of all the methods tested. Overall, the PBE functional gives the worst description of H-bond properties, with the worst binding energies (rms error of 1.5 kcal/mol) and the second worst H-bond distances (rms error of 0.95 Å).

Černý et al. carried out studies investigating the convergence of several methods to the geometric minima of several molecular complexes with various noncovalent interaction types.³⁵⁴ The systems studied here were the T-shaped and parallel-displaced acetylene dimer, the methanol dimer, the methylamine dimer, benzene...water, and the pyrrole dimer. One of the major themes of this study is to test the performance of corrected extrapolation methods, such as estimated CCSD(T), to give accurate analytic gradients and minimum energy geometries. Another goal of this work is to study the effects of the counterpoise corrections, applied to gradient optimizations, on the convergence of final optimum geometries.

There were many techniques and extrapolation methods that were tested in this study; here we will discuss results from only a few of these methods. Generally, MP2 extrapolations were done using either the cc-pVDZ \rightarrow cc-pVTZ (DT) or cc-pVTZ \rightarrow cc-pVQZ (TQ) basis set. The additional CCSD(T) and MP2.5 corrections were calculated using either the cc-pVDZ (D) or cc-pVTZ (T) basis set. Therefore, for example, a method employing the TQ extrapolation along with a CCSD(T) correction with the T basis set would be denoted as MP2/TQ + Δ CCSD(T)/T. Gradient optimizations with these extrapolated methods were carried out without counterpoise corrections. Counterpoise-corrected gradient optimizations were executed using the CCSD(T), MP2.5, and MP2 methods using the cc-pVDZ basis set, and in some cases the cc-pVTZ and cc-pVQZ basis sets were also employed. Additional gradient optimizations were conducted using the SCS(MI)-MP2/cc-pVQZ, TPSS-D/6-311++G(3df, 3pd), and M06-2X/6-311++G(3df,3pd) methods.

Here we will discuss a few key points that can be made from preliminary evaluation of the obtained data. This study represents the first attempt to conduct geometry optimizations using extrapolated methods (and extrapolated gradients) with correction terms, and one of the major conclusions to be drawn is simply that such methods do indeed work and can obtain (at least reasonably) accurate geometric configurations. Not surprisingly, it was also found that the final minimum energy geometry depends heavily on the method used, with contact distances (H-bond distances, for example) that can vary by up to 0.2 Å for most methods (results for M06-2X can be substantially worse). Some bond angles (especially dihedral angles) can vary by much larger margins (up to 30–40°), which is not surprising when one considers that the potential energy curves for these rotations are likely very shallow. Another point to be made here concerns the convergence of counterpoise-corrected and uncorrected results toward the “CBS” geometry. It has been pointed out in the past that generally counterpoise-corrected gradient optimizations in unsaturated basis sets do not converge to the same geometric minimum as uncorrected ones.³⁵⁵ Here it was found that the counterpoise-corrected minimum energy geometries are indeed very similar to noncorrected geometries when the, relatively large, cc-pVQZ basis set is used. Furthermore, in contrast to the “single-point” calculations of interaction energy, it is not so clear that the use of counterpoise correction in geometry optimization in unsaturated basis sets leads systematically to more accurate geometries.

8.2.3. Conformations of Peptides

Most commonly noncovalent interactions are considered as intermolecular interactions between two independent molecules. Indeed, intermolecular interactions are very important in many types of systems and are critical in studies of, for example, fluid properties (gas and liquid phases), protein–ligand complexes, and complex crystal structures. Intramolecular interactions are noncovalent interactions that occur between two chemical moieties located within the *same* structure. These types of interactions are extremely important in determining the overall structure of large molecules. It has been shown in many studies that intramolecular interactions play a critical role in the folding, stability, and dynamics of biomolecular systems, such as proteins and nucleic acid polymers (DNA/RNA).

A very large problem associated with the treatment of intramolecular interactions is the proper description of the BSSE.^{215,356,357} It was long believed that BSSE effects are characteristic of molecular clusters and have no effect on isolated systems. In the past decade it has been seen that extension of basis sets strongly affects the computed relative energies of isolated systems containing intramolecular interactions. This phenomenon led to the characterization of what is now termed the intramolecular basis set superposition error. As in the case of supermolecular computations, the intramolecular BSSE is more pronounced when methods including electron correlation, such as MP2 and CCSD(T), are employed.²¹⁵ In terms of molecular conformations, the effect of intramolecular BSSE is the prediction of conformations that are overstabilized by noncovalent interactions.

In the case of intermolecular complexes the BSSE can effectively be eliminated (or at least strongly reduced) for most computational methods by using the counterpoise technique. For noncovalent interactions within a single

molecule it is not possible to use the counterpoise method, which depends on a supermolecular complex description. It is also not possible to use the DFT-SAPT method, which provides BSSE-free interaction energies, as this method also relies on a supermolecular description of a complex. In the past calculations have been made where a single molecule is split into two parts to compute binding energies using the counterpoise method to account for BSSE effects. The fact that, in this approach, the covalent bond is split (and hydrogens are added to complete valence shells) brings about many problems. Unfortunately, without the ability to use the counterpoise method, there are only a few ways of overcoming the effects of intramolecular BSSE. The “brute force” approach entails accurate correlated (MP2 or CCSD(T)) computations at the CBS limit, or at least using very large basis sets. This type of methodology is successful because the BSSE disappears as one approaches the CBS limit; however, due to the large computational expense of the required calculations for all but the smallest systems, this approach is not generally feasible. Another strategy that has recently been used to deal with intramolecular BSSE is to utilize localized correlated methods, such as localized MP2 (LMP2) and localized CCSD(T) (LCCSD(T)).^{78,358,359}

Localized methods make use of a unitary transformation to the Fock matrix that yields new molecular orbitals that are spatially restricted (localized). The advantage of the LMP2 and LCCSD(T) methods is that they converge very quickly to the CBS limit, meaning that computations made with basis sets such as aug-cc-pVTZ are already very close to this limit. One possible disadvantage of localized methods is that they may not account for certain effects, such as in the cases of stacked aromatic groups and systems with large dispersive contact areas, in the proper way. The last approach for intramolecular BSSE that will be discussed here is the use of DFT-D techniques that have been parametrized to obtain accurate binding energies without counterpoise corrections. Because of the way that they are formulated, these types of methods should provide accurate results for noncovalent interactions regardless of whether those interactions are inter- or intramolecular.

There are many types of systems in which intramolecular interactions play a large role, ranging from very small organic molecules to large proteins. In this work we will focus on conformational analyses of small peptides (up to approximately three amino acid residues). There are several reasons for this constraint: (1) This is a topic in which we have been involved and that we are very well qualified to discuss. (2) Small peptides are relatively small and can be treated using both DFT-D and LMP2/LCCSD(T) techniques. (3) Conformational analyses of small peptides have lately been of particular interest in the literature. (4) Peptides are floppy systems that show an extensive conformational landscape, making this a very interesting class of molecules to study. The first step involved in protein folding is the formation of secondary structures; thus, studies on the energetics of small peptide conformations are important in elucidating the factors that lead to peptide stability.³⁶⁰ In the end the goal is to relate information based on small peptide studies to the folding and stability of proteins. Among the most interesting types of peptides to study are those containing aromatic rings (Phe, Tyr, Trp). These types of peptides are particularly fascinating because interactions involving the aromatic ring are very dispersion-rich and, thus, more challenging to characterize using computational tools. There have

been many studies on peptide structure and dynamics published within the past five or six years, with contributions coming from Hobza and co-workers,^{212,213,215,361–363} van Mourik and co-workers,^{78,356,358,359,364} and others.^{241,360,365,366}

In recent studies van Mourik and co-workers identified several stable conformations of the tyrosine–glycine dipeptide using the B3LYP/6-31+G(d) and MP2/6-31+G(d) methods.^{78,356,358} For most of these structures, the geometry optimizations were done at the DFT level and single-point energies were obtained at the MP2 level, although some MP2 optimizations were also done. Among the minimum energy structures that were found, there were several whose structures were quite different when B3LYP-optimized and MP2-optimized structures were compared. In general, the DFT method tended to favor extended structures while MP2 favored folded structures (referred to as “book” structures), containing dispersion-dominated interactions between the peptide main chain and the aromatic group of tyrosine. Two explanations are given for these structural differences. First, DFT/B3LYP is known not to describe the dispersive-type interactions and, thus, will not describe interactions occurring in the book structures properly. Second, the MP2 method is known to strongly overbind dispersion interactions when it is used with small basis sets because of the intramolecular BSSE.

Several studies were carried out by van Mourik and co-workers to characterize the errors associated with the use of traditional DFT methods, the newly developed M05- and M06-type functionals (as well as PWB6K) of Truhlar and co-workers, DFT-D (B3LYP-D), double hybrid DFT (with and without dispersion corrections), and MP2 for the computation of the structures and energetics of peptides containing dispersion-type intramolecular interactions.^{358,359} These studies are based on potential energy profiles of the book4 and book6 structures of the tyrosine–glycine peptide in which relative energies are plotted against either the glycine C_α–N angle (book4), which is a measure of the degree of “foldedness” of the system, or the distance between two carbons, one associated with the tyrosine ring and the other with the terminus of the glycine residue (book6). The reference method used in these computations was either LMP2/aug-cc-pVTZ (book4) or LCCSD(T0)/aug-cc-pVQZ (book6). Local methods converge very quickly with basis set size and, thus, should exhibit a small BSSE when reasonably large basis sets are employed.

LMP2/aug-cc-pVTZ and LCCSD(T0)/CBS computations on the book4 structure indicate that there are three local minima on the potential energy profile, located at approximately 80° (folded), 180° (extended), and 280° (intermediate), with the global minimum located at around 80° (please see Figure 6). Hybrid DFT methods, such as B3LYP and B97-1 (all with aug-cc-pVDZ as well as B3LYP/6-31+G(d)), clearly show the global minimum to be near 280° and show only very shallow minima near 80°, giving no minima around 180°. This behavior is attributable to the failure of these DFT methods to describe dispersion interactions. MP2, when used along with the (rather small) 6-31+G(d) basis, gives minima at ~80° and 280°, but fails to give any minimum near 180°. This result is likely caused by the fact that the MP2 method, when used with small basis sets, is strongly affected by the BSSE and will tend to overbind dispersion-bound structures. MP2/aug-cc-pVTZ and MP2/aug-cc-pVQZ results are in better agreement with LMP2/aug-cc-pVTZ reference values, as extension of the

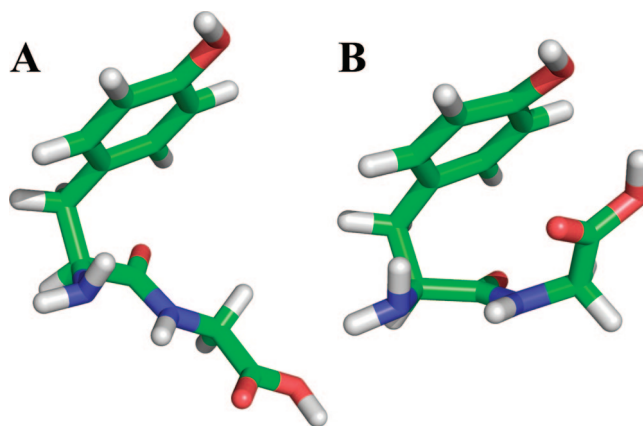


Figure 6. Energy minima of the tyrosine–glycine dipeptide as obtained with DFT (A) and MP2 (B).

basis set will tend to reduce the effects of the BSSE. The improved meta density functionals (Truhlar)^{185,241,341} give much better descriptions of the book4 potential energy profile, with M05-2X, M06-2X, and M06-L giving all of the correct minima. For all of these methods the minima at 80° and 280° are too shallow, and only M06-2X predicts the global minimum to be at ~80°. The DFT-D/B3LYP/aug-cc-pVDZ technique produces a potential energy profile with minima at the correct locations, but minima at 80° and 280° are much too shallow. The double hybrid mPW2-PLYP/aug-cc-pVDZ method yields all three local minima, but, once again, minima at 80° and 280° are too shallow. Neither mPW2-PLYP nor B3LYP-D gives the correct global minimum for this system. MPW2-PLYP results are greatly improved when this functional is augmented with a dispersion term, providing a potential energy profile that closely matches that of the reference LCCSD(T0)/CBS method, although the minimum at 280° is still slightly too shallow.

Valdes, Hobza, and co-workers have published the results of several studies investigating the nature of intramolecular interactions and intramolecular BSSE in small peptides.^{212,213,215,361–363} This research has focused on several peptides containing aromatic residues, namely, phenylalanine–glycine (FG), phenylalanine–glycine–glycine (FGG), tryptophan–glycine (WG), tryptophan–glycine–glycine (WGG), phenylalanine–glycine–phenylalanine (FGF), and glycine–phenylalanine–alanine (GFA). A number of theoretical methods have been used to characterize the properties of these systems. One of the most accurate computational techniques that can be used for molecular dynamics simulations on these peptides is the semiempirical SCC-DFTB-D method. The use of empirical force fields has been shown to be limited mainly due to assignment of atomic charges. It was shown that atomic charges differ considerably for the various structures of the peptide and even attempts to use “average” charges failed. The only reliable approach should thus be based on QM calculations where atomic charges are effectively calculated for each structure of the peptide. SCC-DFTB-D dynamics simulations have been carried out, along with quenching techniques, for several of the above listed peptides to identify the most stable conformations. The quenching technique led to a huge number of energy minima (several thousand), and only the lowest few hundred were considered for more accurate optimizations. Once minimum energy conformations were determined, subsequent geometry optimizations and single-point energy calculations were carried out using several techniques,

including DFT, DFT-D, and MP2 (with the TZVP, cc-pVDZ, and/or cc-pVTZ basis sets). In some cases estimated CCSD(T)/CBS energies are computed as well.

In a thorough investigation Hobza, de Vries, and co-workers used a large number of computational methods, as well as infrared spectroscopic techniques, to characterize the conformational stabilities of the GFA peptide.³⁶³ First, SCC-DFTB-D as well as classical (force-field based), dynamics/quenching simulations were done to identify the lowest energy conformations. It was found that the conformers obtained using SCC-DFTB-D and the force-field-based methods are quite different. This disagreement may be attributable to the improper and static assignment of charges in the classical simulation method (see above). The next step was optimization of the lowest energy structures using the MP2/cc-pVTZ, DFT-D/TPSS/6-311++G(3df,3pd), and DFT/M06-2X/6-311+G(2df,2pd) methods. The resulting structures can be categorized into four families (according to the Ramachandran terminology). The β form of the peptide is characterized by an extended backbone structure, the γ form contains a hydrogen bond between the alanine NH group and the glycine C=O group, the 3_{11} conformers have a hydrogen bond between the terminal CO₂H and NH₂ groups (HOC=O...HNH), and the γ - 3_{11} conformers exhibit properties of both 3_{11} and γ peptides. The β structures can be further subdivided into those structures in which the CO₂H group is H-bonded (CO₂H_{bonded}) and those in which this group is free (CO₂H_{free}). Interestingly, structures optimized using DFT-D and MP2 are very similar, indicating that intramolecular BSSE is not critical in these structures. It should be noted that hydrogen bonds are seemingly more important to the conformational structures than dispersion interactions involving the aromatic ring, possibly explaining the reduced amount of intramolecular BSSE exhibited by this peptide.

Estimated CCSD(T)/CBS electronic energies for the lowest energy structures were computed and compared to energies obtained using the DFT-D and M06-2X techniques. The ordering of the structures generally corresponds to the number of H-bonds present, with γ structures generally having the lowest energies, followed by β (CO₂H_{bonded}), γ - 3_{11} , 3_{11} , and β (CO₂H_{free}) structures (in that order). It was found that DFT-D conformational energies are in good agreement with CCSD(T)/CBS data, giving the same overall conformation/energy trends (including a coincident global minimum structure), although there are some minor discrepancies between the two methods. The M06-2X method does not agree with benchmark CCSD(T)/CBS data nearly as well as DFT-D, predicting the incorrect global minimum structure, and failing to favor β (CO₂H_{bonded}) structures over β (CO₂H_{free}) structures.

For peptide conformer populations to be compared with experimental (spectroscopic) results, it is necessary to compute the peptide stabilities in terms of free energies at some finite temperature. Relative enthalpies (ΔH_e) and Gibbs free energies (ΔG) for the GFA peptide were computed using both the CCSD(T)/CBS and DFT-D techniques (at 300 K). There are two main effects that are observed when passing from energies to enthalpies. The first of these is a large reduction in the energy interval by the 16 lowest energy conformers; this range is lowered from 2.14 to 0.62 kcal/mol for the CCSD(T)/CBS method. The other effect is a reordering of the minimum energy conformers, with CO₂H_{free} structures now being energetically more favorable than the CO₂H_{bonded} structures. Inclusion of entropic effects does not

have a large effect on the ordering of conformers, with CO₂H_{free} structures still being more stable than CO₂H_{bonded} ones.

Gibbs free energies are obtained using ab initio methods with the assumption that the rigid rotor harmonic oscillator ideal gas approximations (RR-HO-IG) are valid and have the potential to suffer from inaccuracies due to these approximations. Population analyses based on MD are more robust because of the fact that the entire conformational space of the peptide is sampled and because they go beyond the harmonic approximation. The GFA peptide was studied using metadynamics techniques, which are based on MD simulations, to investigate the whole conformational landscape of the peptide and to validate the minimum energy structures obtained using the technique involving MD/quenching and ab initio optimization (MD/Q + QM). The metadynamics method uses a history-based sampling potential to bias the simulation toward unsampled conformational space, allowing for shorter simulations that sample a larger portion of the free energy surface. These simulations were carried out using both SCC-DFTB-D- and force-field-based potentials; this was done to assess the quality of the force fields in obtaining low-lying conformations of the peptide. It was found that the peptide conformations obtained using SCC-DFTB-D-based metadynamics were very similar to those obtained using MD/Q + QM, indicating that anharmonic effects do not play a very large role in conformer stability. Force-field-based metadynamics gave results that were not in good agreement with those obtained using SCC-DFTB-D, and the authors recommend that classical MD methods not be used for the evaluation of thermodynamic characteristics of peptides.

Grimme, Kleinermanns, and co-workers investigated the conformational energetics and spectra of the H-Trp-Ser-OH peptide (Trp-Ser) using several computational techniques as well as IR/UV spectroscopy.³⁶⁰ The reference method used in this study is the newly developed B2PLYP-D, a double hybrid DFT method incorporating an empirical dispersion term, along with the TZVPP basis set. The use of this method for obtaining reference data is supported by the fact that dispersion-corrected double hybrid DFT methods (including B2PLYP-D) have been shown to give excellent agreement of estimated CCSD(T)/CBS data for the relative energies of various conformations of the Phe-Gly-Gly peptide. Structures were optimized at the B97-D/TZV2P level. It was found that the lowest energy structures all involved OH...O=C hydrogen bonds involving the C-terminal OH group, the serine COOH side group, and the tryptophan backbone carbonyl group. All of the lowest energy structures also exhibit some degree of stacking between the serine residue and the tryptophan aromatic ring, with the lowest energy structure showing the greatest degree of stacking. The performances of the B3LYP, B3LYP-D, B97-D, PM3-D (semiempirical), and MM3 (force field) methods were evaluated and results from these methods compared to reference B2PLYP-D conformational energies; it should be noted that separate geometry optimizations were carried out using each of these methods. It is found that the conformational assignments (energy ordering) obtained using B3LYP-D and B97-D are in very good agreement with those from the reference method (rmsd's of 0.2 and 0.5 kcal/mol, respectively), while the B3LYP method, which includes no dispersion correction, performs very poorly compared to B2PLYP-D. The PM3-D

and MM3 methods both exhibit rather large conformational energy errors (rmsd's of 2.6 and 3.0 kcal/mol, respectively), and the authors recommend that these techniques be used only for crude initial screening. It should be noted that, despite its overall inaccuracies, the PM3-D method does correctly assign the minimum energy Trp-Ser structure.

Zhao and Truhlar tested the performance of several density functional methods, including the newly developed M08-HX and M08-SO functionals, in computing accurate relative conformational energies for a test set containing five peptides (aromatic peptide conformational energies with five small peptides test set, APCE5).²⁴¹ It was found that the M05-2X, M08-HX, and M08-SO functionals yielded results that are significantly better than those of B3LYP, presumably because of their more accurate description of dispersion interactions. These functionals, however, did not produce peptide relative energies that were as accurate as those obtained using the DFT-D/TPSS/6-311++G(3df,3pd) method.

9. Conclusions

(i) In the field of WFT, the coupled-cluster method with iterative single and double excitations, and perturbative correction for the triples, CCSD(T), in the complete basis set limit still represents the benchmark method for systems with up to 30–50 second-row atoms and hydrogens. For larger noncovalent complexes, from 50 to 80 atoms, highly accurate substitutes for the CCSD(T) method are the SCS-CCSD, SCS(MI)-CCSD, and MP2.5 methods ($\sim N^6$ scaling, N being the number of AO basis functions). Outstanding accuracy for even larger complexes, i.e., of a size at the edge of applicability of the RI/DF/CD-based MP2 methods, can be obtained with the dispersion-corrected MP2 schemes of Hesselmann or Tkatchenko. From the family of the spin-component-scaled MP2 methods, especially the DW-MP2 method but also the SCS(MI)-MP2 method appears to be capable of delivering highly accurate data. Further approximations leading to $\sim N^4$ scaling methods, such as in SOS(MI)-MP2, open a possibility for calculations on systems with hundreds of atoms, but unfortunately not without a significant loss of accuracy. WFT methods exploiting the locality of the electron correlation (eventually explicitly correlated to enhance the basis set convergence rate) will certainly be developed into “black-box” implementations in the next decade and dominate the WFT methodologies used for calculations of noncovalent interaction.

(ii) Density functional theory is gradually turning into a viable tool for accurate calculations of intermolecular interactions, mainly due to advances in the description of the dispersion interaction. Currently, there are a very wide range of dispersion corrections suggested in the literature, varying in their accuracy, reliability, and time demands. Some of these approaches have already become part of the widely used computational chemistry packages, such as TurboMole (DFT-D), Gaussian 09 (DFT-D, M06 suite), Siesta (vdW-DF), NW-Chem (M06 suite), Q-Chem (DFT-D), deMon (DFT-D), and others. As of now, the DFT-D approach (a combination of current DFT functionals with an empirical dispersion correction) prevails, as it is easy to implement, is sufficiently accurate for most applications, and has negligible time demands. DFT-D will probably dominate for the next few years. However, there are many nonempirical corrections, such as nonlocal vdW-DF, whose variants may become more accurate and computationally efficient soon. These and similar methods are genuine density

functionals, and they have the potential to become benchmark quality methods.

(iii) The quality of semiempirical QM methods, augmented by second-generation corrections for dispersion and H-bonding, reaches the accuracy of DFT and WFT methods for a large number of the investigated complexes. The semiempirical methods are, however, faster by several orders. The PM6, OM3, and SCC-DF-TB methods are significantly more accurate than the AM1 method, and the best results were found for the PM6-DH2 method. The most important advantage of these techniques in comparison with empirical potentials is their ability to describe quantum effects. Additionally, these methods are fast (and simultaneously accurate) enough to be used in on-the-fly *ab initio* MD simulations.

(iv) In assessing a computational method's ability to describe noncovalent interactions, it is critical that its performance be measured not only at the potential minimum, but also at many (all the) points on the potential energy surface. A good description of the potential energy surface for all interaction types not only is important in terms of a method's ability to obtain reasonable minimum energy structures and vibrational frequencies, but also has tremendous implications in the use of QM methods for molecular dynamic simulations. Among the methods considered here, DFT-SAPT and MP2.5 can be said to produce accurate potential energy curves for a wide variety of interaction types. SCS-CCSD has also been shown to yield accurate potential energy curves for systems containing aromatic rings; however, more assessment of this method's strengths and weaknesses should still be carried out. The three techniques mentioned above are all quite computationally intensive and generally cannot be used for systems containing more than 50–80 atoms (and even fewer if geometry optimizations are to be done or if it is necessary to study large numbers of structures for the generation of potential energy curves). In terms of lower level methods, which can be used for computations on larger systems, the DFT-D method (when used with basis sets of triple- ζ quality) can be recommended for geometry optimizations and for generating potential energy curves. One important advantage of the DFT-D methods is the fact that (in their most common form) they do not require the use of counterpoise corrections. This has large implications for studies on intramolecular interactions, for which counterpoise corrections cannot be used. Finally, it should be noted that the dispersion- and hydrogen-bond-corrected semiempirical methods, such as PM6-DH2, have not been thoroughly examined in terms of their abilities to describe potential energy surfaces of noncovalent interactions. It is very important that these assessments be carried out in the near future, as these methods show such great promise for use in geometry optimizations and molecular dynamics simulations on very large systems.

10. Outlook

The recent outbreak of fast, massively parallel, and multicore computer architectures, along with new efficient algorithms that can take advantage of this hardware, has brought computational chemistry to the beginning of the “petaflop” era. It can thus be expected that the number of benchmark quality calculations on noncovalent complexes of (moderately) increasing size will grow. However, the performance of ever-improving computer technology, even in combination with the arbitrarily efficient implementations

of the present steeply scaling algorithms, such as standard CCSD(T), will always be many steps behind the desire of computational chemists. Due to this “scaling wall,” major advances in the applicability of computational methods must come from the development of new theories and techniques.

From the point of view of the “pure” WFT methods, taking advantage of the locality of the electron correlation, pure atomic-orbital-based algorithms are the best candidates for new developments. It appears to be possible to overcome problems arising from the use of incomplete or unbalanced basis sets, such as intra- and intermolecular BSSE, by introducing explicit correlation into the wave function.

Another promising route seems to be the combination of the strong points of WFT and DFT techniques into hybrid methods. For instance, the calculation of the dispersion energy from the TDDHF response functions of monomers is both cheap and accurate, and can efficiently be utilized within the WFT concept, as discussed previously. This is just one example of a successful combination of these two theories, while many others have already been published, are being tested, or are just waiting to be discovered.

As far as the approximate methods utilizing empirical parameters fitted to various databases of rather small molecules (so far) are concerned, increasing the number of benchmark calculations on extended complexes might bring a completely different point of view concerning these methods. After a certain threshold in the size of calculated complexes is passed, a new phenomenon originating from the superposition of multiple subtle long-range forces might arise. So far it is not understood how this will influence the accuracy of the approximate methods that perform well for small complexes.

DFT-based methods are usually understood to be faster alternatives to WFT calculations. When corrected for the missing dispersion interaction, the accuracy of DFT in the field of noncovalent interactions is currently becoming acceptable for most modeling applications. Over time, the fairly good basis of nonlocal approaches in DFT methodologies springing up in the literature will get implemented in commercial software packages. We can expect that DFT methods, with a relatively small number of empirical parameters, and with accuracy and reliability better than those of MP2, will become available within the next few years. However, development of benchmark quality DFT methods may take more time. In addition, it is very likely that the first benchmark quality DFT approximations will contain empirical parameters. This will make them less acceptable as reliable benchmark methods for a wider community of users. Nevertheless, judging by the growing interest in this field, and considering the rapid development it has undergone lately, it is possible that DFT methods will indeed reach benchmark quality in the next one or two decades. This certainly does not mean that DFT will replace WFT. More likely, there will be distinct sets of problems for which either DFT or WFT will provide better or more reliable results. To conclude, the prospects of DFT seem to be very bright today, but let us not forget that the era of quantum computing may be close and that some previously intractable theories may soon become comparatively faster and more accurate than current DFT approaches.

The development of WFT and DFT methods, or possibly of methods using WFT/DFT combinations, should result in the design of efficient, broadly applicable, linearly scaling QM methods suitable for calculations on extended systems

having tens of thousands of atoms, and also for performing on-the-fly *ab initio* MD simulations on systems with hundreds of atoms.

Several years ago it was believed that semiempirical QM methods were not suitable for the study of noncovalent interactions. It is now, however, clear that upon careful parametrization they can provide useful information on the structure and stability of very extended noncovalent systems. Probably their most prominent role will be in MD simulations of biomacromolecules and complex molecular systems, where they might replace empirical potentials, which are not able to describe quantum effects. In the near future we expect an advent in the evaluation of thermodynamic characteristics based on MD simulations using the QM techniques.

We refer all our readers to our next review on noncovalent interactions that we will publish in about 20 years (the previous one, ref 1, appeared in 1988), where they will find out how realistic our predictions in the present outlook are.

11. Acknowledgments

We acknowledge with gratitude the contributions of our students and co-workers, as cited in the references. Without their help we would not be able to cover such a broad area as theoretical calculations on noncovalent interactions. This work was a part of Research Project No. Z40550506 of the Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, and it was supported by Grant Nos. LC512 and MSM6198959216 from the Ministry of Education, Youth and Sports of the Czech Republic. The support of Praemium Academiae, Academy of Sciences of the Czech Republic, awarded to P.H. in 2007 is also acknowledged. This work was also supported by the Korea Science and Engineering Foundation (World Class University program, Grant R32-2008-000-10180-0) and Slovak Grant Agency VEGA under Contract Nos. 1/0428/09 and 1/0520/10. K.E.R. gratefully acknowledges the support of the NSF EPSCOR program (Grant EPS-0701525). The important part of the calculations was performed using the Chinook super-computer at EMSL, Pacific Northwest National Laboratories.

12. Note Added in Proof

In parallel to the work of Merz et al., the so called, ‘S22x5’ database (Grafova, L.; Pitonak, M.; Rezac, J.; Hobza, P., *J. Chem. Theory Comput.*, submitted for publication) was introduced in our laboratory. This database is an extension of the original S22 database containing both the equilibrium (i.e. the S22) and nonequilibrium structures of the same complexes. Complexes with frozen monomer geometries were shortened by 10%, and elongated by 20, 50, and 100% along the intermolecular interaction coordinate. CCSD(T)/CBS interaction energies were calculated to serve as the benchmarks for comparison with popular WFT (SOS- and SOS(MI)-MP2, MP2, SCS- and SCS(MI)-MP2, MP2C, MP2.5 and SCS-CCSD) and DFT (PBE, PBE-D, BLYP-D, TPSS-D, M06-2X, B2-PLYP and B2-PLYP-D) methods for noncovalent interaction calculations. Sophisticated statistical analysis, based on relative rather than absolute errors, was used to assess both the accuracy and the capability of methods to deliver balanced descriptions for both the equilibrium and stretched complex geometries. The analyses were performed for the entire S22x5 set as well as for each

interaction type and geometry distortions separately. The three different interaction types, electrostatics dominated, dispersion dominated, and mixed-character, were assigned according to DFT-SAPT interaction energy decomposition for each complex and distortion individually.

According to result obtained, among the WFT methods, the most accurate and balanced result were delivered by SCS-CCSD, MP2C, and MP2.5 and to lesser extent by the SCS(MI)-MP2 method. None of the tested DFT method was accurate and balanced enough to meet the strict statistical criteria fulfilled by the WFT methods mentioned above. For complexes dominated by electrostatic forces (mostly the H-bonded complexes) the performance of the DFT methods was satisfactory. Clearly the most problematic complexes were the dispersion-dominated ones. The conclusions in the paper indicate that the accuracy of the DFT methods in the range from equilibrium to twice-the-equilibrium monomer distance is rather unstable and the relative errors in the stretched geometries typically increased rather than converged toward zero. Despite the fact that the absolute interaction energies in mid- to long-range interaction fragment separations are decreasing, considering their abundance in large molecules, this is quite an alarming message. However, studies on a few large-scale applications discussed in the paper show that error cancellation, rather than error propagation, may take place in real-life applications.

13. References

- Hobza, P.; Zahradnik, R. *Chem. Rev.* **1988**, *88*, 871.
- Harrison, R. J.; Bartlett, R. J. *Int. J. Quantum Chem.* **1986**, *20*, 437.
- Jurecka, P.; Hobza, P. *J. Am. Chem. Soc.* **2003**, *125*, 15608.
- Vondrasek, J.; Bendova, L.; Klusak, V.; Hobza, P. *J. Am. Chem. Soc.* **2005**, *127*, 2615.
- Janowski, T.; Ford, A. R.; Pulay, P. *Mol. Phys.*, **2010**, *108*, 249.
- Hobza, P.; Sponer, J. *Chem. Rev.* **1999**, *99*, 3247.
- Hobza, P.; Havlas, Z. *Chem. Rev.* **2000**, *100*, 4253.
- Muller-Dethlefs, K.; Hobza, P. *Chem. Rev.* **2000**, *100*, 143.
- Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5291.
- Hobza, P.; Zahradnik, R.; Muller-Dethlefs, K. *Collect. Czech. Chem. Commun.* **2006**, *71*, 443.
- Chem. Rev.* **1988**, *88* (6).
- Chem. Rev.* **1994**, *94* (7).
- Chem. Rev.* **2000**, *100* (11).
- Phys. Chem. Chem. Phys.* **2008**, *10* (19).
- Rappe, A. K.; Bernstein, E. R. *J. Phys. Chem. A* **2000**, *104*, 6117.
- Sherrill, C. D. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Cundari, T. R., Eds.; Wiley: New York, 2009; Vol. 26.
- Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10656.
- Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 1009.
- Zhou, H.-X.; Gilson, M. K. *Chem. Rev.* **2009**, *109*, 4092.
- Janowski, T.; Pulay, P. *Chem. Phys. Lett.* **2007**, *447*, 27.
- Cizek, J. *J. Chem. Phys.* **1966**, *45*, 4256.
- Paldus, J.; Shavitt, I.; Cizek, J. *Phys. Rev. A* **1972**, *5*, 50.
- Bartlett, R. J.; Musial, M. *Rev. Mod. Phys.* **2007**, *79*, 291.
- Helgaker, T.; Jorgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; Wiley: New York, 2000.
- Janowski, T.; Pulay, P. *J. Chem. Theory Comput.* **2008**, *4*, 1585.
- Lotrich, V.; Flocke, N.; Ponton, M.; Yau, A. D.; Perera, A.; Deumens, E.; Bartlett, R. J. *J. Chem. Phys.* **2008**, *128*, 194104.
- Neogrady, P.; Aquilante, F.; Noga, J.; Pitoňák, M.; Urban, M. To be submitted for publication.
- Noga, J.; Valiron, P. *Mol. Phys.* **2005**, *103*, 2123.
- Olson, R. M.; Bentz, J. L.; Kendall, R. A.; Schmidt, M. W.; Gordon, M. S. *J. Chem. Theory Comput.* **2007**, *3*, 1312.
- Rendell, A. P.; Lee, T. J.; Lindh, R. *Chem. Phys. Lett.* **1992**, *194*, 84.
- Koch, H.; De Meras, A. S. *J. Chem. Phys.* **2000**, *113*, 508.
- Rendell, A. P.; Lee, T. J. *J. Chem. Phys.* **1994**, *101*, 400.
- Neese, F.; Wennmohs, F.; Hansen, A. *J. Chem. Phys.* **2009**, *130*, 114108.
- Schutz, M.; Manby, F. R. *Phys. Chem. Chem. Phys.* **2003**, *5*, 3349.
- Pitonak, M.; Neogrady, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2010**, *12*, 1369.
- Podeszwa, R.; Szalewicz, K. *J. Chem. Phys.* **2007**, *126*, 194101.
- Rode, M.; Sadlej, J.; Moszynski, R.; Wormer, P. E. S.; van der Avoird, A. *Chem. Phys. Lett.* **1999**, *314*, 326.
- Rybak, S.; Jezierski, B.; Szalewicz, K. *J. Chem. Phys.* **1991**, *95*, 6576.
- Pittner, J.; Hobza, P. *Chem. Phys. Lett.* **2004**, *390*, 496.
- Crittenden, D. L. *J. Phys. Chem. A* **2009**, *113*, 1663.
- Hopkins, B. W.; Tschumper, G. S. *J. Phys. Chem. A* **2004**, *108*, 2941.
- Pitonak, M.; Neogrady, P.; Rezac, J.; Jurecka, P.; Urban, M.; Hobza, P. *J. Chem. Theory Comput.* **2008**, *4*, 1829.
- Lee, J. S. *Theor. Chem. Acc.* **2005**, *113*, 87.
- Patkowski, K.; Szalewicz, K. *J. Chem. Phys.* **2007**, *127*, 164103.
- van Mourik, T.; Wilson, A. K.; Dunning, T. H. *Mol. Phys.* **1999**, *96*, 529.
- Boese, A. D.; Martin, J. M. L.; Klopper, W. *J. Phys. Chem. A* **2007**, *111*, 11122.
- Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.
- Pitonak, M.; Riley, K. E.; Neogrady, P.; Hobza, P. *ChemPhysChem* **2008**, *9*, 1636.
- Jurecka, P.; Hobza, P. *Chem. Phys. Lett.* **2002**, *365*, 89.
- Min, S. K.; Lee, E. C.; Lee, H. M.; Kim, D. Y.; Kim, D.; Kim, K. S. *J. Comput. Chem.* **2008**, *29*, 1208.
- Sherrill, C. D.; Takatani, T.; Hohenstein, E. G. *J. Phys. Chem. A* **2009**, *113*, 10146.
- Pitonak, M.; Janowski, T.; Neogrady, P.; Pulay, P.; Hobza, P. *J. Chem. Theory Comput.* **2009**, *5*, 1761.
- ElSohly, A. M.; Tschumper, G. S. *Int. J. Quantum Chem.* **2009**, *109*, 91.
- Adler, T. B.; Werner, H. J. *J. Chem. Phys.* **2009**, *130*, 241101.
- Hampel, C.; Werner, H. J. *J. Chem. Phys.* **1996**, *104*, 6286.
- Schutz, M.; Werner, H. J. *Chem. Phys. Lett.* **2000**, *318*, 370.
- Schutz, M.; Werner, H. J. *J. Chem. Phys.* **2001**, *114*, 661.
- Friedrich, J.; Dolg, M. *J. Chem. Theory Comput.* **2009**, *5*, 289.
- Friedrich, J.; Hanrath, M.; Dolg, M. *J. Chem. Phys.* **2007**, *126*, 154110.
- Friedrich, J.; Hanrath, M.; Dolg, M. *J. Phys. Chem. A* **2007**, *111*, 9830.
- Auer, A. A.; Nooijen, M. *J. Chem. Phys.* **2006**, *125*, 024104.
- Scuseria, G. E.; Ayala, P. Y. *J. Chem. Phys.* **1999**, *111*, 8330.
- Flocke, N.; Bartlett, R. J. *J. Chem. Phys.* **2004**, *121*, 10935.
- Hughes, T. F.; Bartlett, R. J. *J. Chem. Phys.* **2008**, *129*, 054105.
- Hughes, T. F.; Flocke, N.; Bartlett, R. J. *J. Phys. Chem. A* **2008**, *112*, 5994.
- Maslen, P. E.; Dutoi, A. D.; Lee, M. S.; Shao, Y. H.; Head-Gordon, M. *Mol. Phys.* **2005**, *103*, 425.
- Maslen, P. E.; Lee, M. S.; Head-Gordon, M. *Chem. Phys. Lett.* **2000**, *319*, 205.
- Subotnik, J. E.; Sodt, A.; Head-Gordon, M. *J. Chem. Phys.* **2006**, *125*, 074116.
- Li, S. H.; Ma, J.; Jiang, Y. *J. Comput. Chem.* **2002**, *23*, 237.
- Li, S. H.; Shen, J.; Li, W.; Jiang, Y. *J. Chem. Phys.* **2006**, *125*, 074109.
- Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2005**, *123*, 134103.
- Li, W.; Piecuch, P.; Gour, J. R.; Li, S. H. *J. Chem. Phys.* **2009**, *131*, 114109.
- Reinhardt, P.; Zhang, H.; Ma, J.; Malrieu, J. P. *J. Chem. Phys.* **2008**, *129*, 164106.
- Hill, J. G.; Platts, J. A. *J. Chem. Theory Comput.* **2007**, *3*, 80.
- Hill, J. G.; Platts, J. A. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2785.
- Runeberg, N.; Schutz, M.; Werner, H. J. *J. Chem. Phys.* **1999**, *110*, 7210.
- Schutz, M.; Rauhut, G.; Werner, H. J. *J. Phys. Chem. A* **1998**, *102*, 5997.
- Holroyd, L. F.; van Mourik, T. *Chem. Phys. Lett.* **2007**, *442*, 42.
- Valdes, H.; Klusak, V.; Pitonak, M.; Exner, O.; Stary, I.; Hobza, P.; Rulisek, L. *J. Comput. Chem.* **2008**, *29*, 861.
- Werner, H. J. *J. Chem. Phys.* **2008**, *129*, 101103.
- Kobayashi, M.; Nakai, H. *J. Chem. Phys.* **2008**, *129*, 044103.
- Dunlap, B. I.; Connolly, J. W. D.; Sabin, J. R. *J. Chem. Phys.* **1979**, *71*, 3396.
- Feyereisen, M.; Fitzgerald, G.; Komornicki, A. *Chem. Phys. Lett.* **1993**, *208*, 359.
- Vahtras, O.; Almlöf, J.; Feyereisen, M. W. *Chem. Phys. Lett.* **1993**, *213*, 514.
- Weigend, F.; Haser, M. *Theor. Chem. Acc.* **1997**, *97*, 331.
- Weigend, F.; Haser, M.; Patzelt, H.; Ahlrichs, R. *Chem. Phys. Lett.* **1998**, *294*, 143.
- Whitten, J. L. *J. Chem. Phys.* **1973**, *58*, 4496.
- Beebe, N. H. F.; Linderberg, J. *Int. J. Quantum Chem.* **1977**, *7*, 683.
- Aquilante, F.; Gagliardi, L.; Pedersen, T. B.; Lindh, R. *J. Chem. Phys.* **2009**, *130*, 154107.
- Aquilante, F.; Lindh, R.; Pedersen, T. B. *J. Chem. Phys.* **2007**, *127*, 114107.

- (91) Aquilante, F.; Pedersen, T. B.; Lindh, R. *J. Chem. Phys.* **2007**, *126*, 194106.
- (92) Koch, H.; De Meras, A. S.; Pedersen, T. B. *J. Chem. Phys.* **2003**, *118*, 9481.
- (93) Pedersen, T. B.; Aquilante, F.; Lindh, R. *Theor. Chem. Acc.* **2009**, *124*, 1.
- (94) Chalasinski, G.; Szczesniak, M. M. *Mol. Phys.* **1988**, *63*, 205.
- (95) Cybulski, S. M. *J. Chem. Phys.* **1992**, *96*, 8225.
- (96) Cybulski, S. M.; Chalasinski, G.; Moszynski, R. *J. Chem. Phys.* **1990**, *92*, 4357.
- (97) Cybulski, S. M.; Lytle, M. L. *J. Chem. Phys.* **2007**, *127*, 141102.
- (98) Hesselmann, A. *J. Chem. Phys.* **2008**, *128*, 9.
- (99) Hobza, P.; Selzle, H. L.; Schlag, E. W. *J. Phys. Chem.* **1996**, *100*, 18790.
- (100) Jaffe, R. L.; Smith, G. D. *J. Chem. Phys.* **1996**, *105*, 2780.
- (101) Tsuzuki, S.; Luthi, H. P. *J. Chem. Phys.* **2001**, *114*, 3949.
- (102) Tsuzuki, S.; Uchimarui, T.; Matsumura, K.; Mikami, M.; Tanabe, K. *Chem. Phys. Lett.* **2000**, *319*, 547.
- (103) Chalasinski, G.; Szczesniak, M. M. *Chem. Rev.* **1994**, *94*, 1723.
- (104) Riley, K. E.; Hobza, P. *J. Phys. Chem. A* **2007**, *111*, 8257.
- (105) Bostrom, J.; Aquilante, F.; Pedersen, T. B.; Lindh, R. *J. Chem. Theory Comput.* **2009**, *5*, 1545.
- (106) Jurecka, P.; Nachtigall, P.; Hobza, P. *Phys. Chem. Chem. Phys.* **2001**, *3*, 4578.
- (107) Weigend, F.; Kohn, A.; Hattig, C. *J. Chem. Phys.* **2002**, *116*, 3175.
- (108) Almlof, J. *Chem. Phys. Lett.* **1991**, *181*, 319.
- (109) Jung, Y. S.; Lochan, R. C.; Dutoi, A. D.; Head-Gordon, M. *J. Chem. Phys.* **2004**, *121*, 9793.
- (110) Kats, D.; Usvyat, D.; Schutz, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3430.
- (111) Fernandez, B.; Pedersen, T. B.; De Meras, A. S.; Koch, H. *Chem. Phys. Lett.* **2007**, *441*, 332.
- (112) Adler, T. B.; Werner, H. J.; Manby, F. R. *J. Chem. Phys.* **2009**, *130*, 054106.
- (113) Distasio, R. A.; Jung, Y. S.; Head-Gordon, M. *J. Chem. Theory Comput.* **2005**, *1*, 862.
- (114) Lee, M. S.; Maslen, P. E.; Head-Gordon, M. *J. Chem. Phys.* **2000**, *112*, 3592.
- (115) Deev, V.; Collins, M. A. *J. Chem. Phys.* **2005**, *122*, 154102.
- (116) Gordon, M. S.; Freitag, M. A.; Bandyopadhyay, P.; Jensen, J. H.; Kairys, V.; Stevens, W. J. *J. Phys. Chem. A* **2001**, *105*, 293.
- (117) Gordon, M. S.; Truhlar, D. G. *J. Am. Chem. Soc.* **1986**, *108*, 5412.
- (118) Grimme, S. *J. Chem. Phys.* **2003**, *118*, 9095.
- (119) Szabados, A. *J. Chem. Phys.* **2006**, *125*, 214105.
- (120) Antony, J.; Grimme, S. *J. Phys. Chem. A* **2007**, *111*, 4862.
- (121) Bachorz, R. A.; Bischoff, F. A.; Hofener, S.; Klopper, W.; Ottiger, P.; Leist, R.; Frey, J. A.; Leutwyler, S. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2758.
- (122) Distasio, R. A.; Head-Gordon, M. *Mol. Phys.* **2007**, *105*, 1073.
- (123) King, R. A. *Mol. Phys.* **2009**, *107*, 789.
- (124) Takatani, T.; Sherrill, C. D. *Phys. Chem. Chem. Phys.* **2007**, *9*, 6106.
- (125) Aquilante, F.; Pedersen, T. B. *Chem. Phys. Lett.* **2007**, *449*, 354.
- (126) Lochan, R. C.; Jung, Y.; Head-Gordon, M. *J. Phys. Chem. A* **2005**, *109*, 7598.
- (127) Marchetti, O.; Werner, H.-J. *J. Phys. Chem. A* **2009**, *113*, 11580.
- (128) Knizia, G.; Adler, T. B.; Werner, H.-J. *J. Chem. Phys.* **2009**, *130*, 054104.
- (129) Hesselmann, A.; Jansen, G.; Schutz, M. *J. Chem. Phys.* **2005**, *122*, 014103.
- (130) Hesselmann, A.; Jansen, G.; Schutz, M. *J. Am. Chem. Soc.* **2006**, *128*, 11730.
- (131) Jurecka, P.; Sponer, J.; Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985.
- (132) Della Sala, F.; Görling, A. *J. Chem. Phys.* **2001**, *115*, 5718.
- (133) Pitonak, M.; Hesselmann, A. *J. Chem. Theory Comput.* **2009**, *6*, 168.
- (134) Tkatchenko, A.; Distasio, R. A.; Head-Gordon, M.; Scheffler, M. *J. Chem. Phys.* **2009**, *131*, 094106.
- (135) Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463.
- (136) Jurecka, P.; Cerny, J.; Hobza, P.; Salahub, D. R. *J. Comput. Chem.* **2007**, *28*, 555.
- (137) Tsuzuki, S.; Uchimarui, T.; Tanabe, K. *Chem. Phys. Lett.* **1998**, *287*, 202.
- (138) Urban, M.; Noga, J.; Cole, S. J.; Bartlett, R. J. *J. Chem. Phys.* **1985**, *83*, 4041.
- (139) Pitonak, M.; Neogrady, P.; Cerny, J.; Grimme, S.; Hobza, P. *ChemPhysChem* **2009**, *10*, 282.
- (140) He, Y.; Cremer, D. *Theor. Chem. Acc.* **2000**, *105*, 110.
- (141) Takatani, T.; Hohenstein, E. G.; Sherrill, C. D. *J. Chem. Phys.* **2008**, *128*, 129901.
- (142) Pitonak, M.; Rezac, J.; Hobza, P. *Phys. Chem. Chem. Phys.*, submitted for publication.
- (143) Neese, F.; Hansen, A.; Wennmohs, F.; Grimme, S. *Acc. Chem. Res.* **2009**, *42*, 641.
- (144) Wennmohs, F.; Neese, F. *Chem. Phys.* **2008**, *343*, 217.
- (145) Neese, F.; Hansen, A.; Liakos, D. G. *J. Chem. Phys.* **2009**, *131*, 064103.
- (146) Halkier, A.; Helgaker, T.; Jorgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, *286*, 243.
- (147) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. *J. Chem. Phys.* **1997**, *106*, 9639.
- (148) Truhlar, D. G. *Chem. Phys. Lett.* **1998**, *294*, 45.
- (149) Lee, E. C.; Kim, D.; Jurecka, P.; Tarakeswar, P.; Hobza, P.; Kim, K. S. *J. Phys. Chem. A* **2007**, *111*, 3446.
- (150) Shin, I.; Park, M.; Min, S. K.; Lee, E. C.; Suh, S. B.; Kim, K. S. *J. Chem. Phys.* **2006**, *125*, 234305.
- (151) Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108.
- (152) Steele, R. P.; Distasio, R. A.; Head-Gordon, M. *J. Chem. Theory Comput.* **2009**, *5*, 1560.
- (153) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.
- (154) Xantheas, S. S. *J. Chem. Phys.* **1996**, *104*, 8821.
- (155) Hankins, D.; Moskowit, J. W.; Stilling, F. H. *J. Chem. Phys.* **1970**, *53*, 4544.
- (156) Dedikova, P.; Pitonak, M.; Neogrady, P.; Cernusak, I.; Urban, M. *J. Phys. Chem. A* **2008**, *112*, 7115.
- (157) Neogrady, P.; Pitonak, M.; Urban, M. *Mol. Phys.* **2005**, *103*, 2141.
- (158) Taube, A. G.; Bartlett, R. J. *Collect. Czech. Chem. Commun.* **2005**, *70*, 837.
- (159) Klopper, W.; Manby, F. R.; Ten-No, S.; Valeev, E. F. *Int. Rev. Phys. Chem.* **2006**, *25*, 427.
- (160) Kutzelnigg, W. *Theor. Chim. Acta* **1985**, *68*, 445.
- (161) Klopper, W.; Samson, C. C. M. *J. Chem. Phys.* **2002**, *116*, 6397.
- (162) Manby, F. R. *J. Chem. Phys.* **2003**, *119*, 4607.
- (163) May, A. J.; Valeev, E.; Polly, R.; Manby, F. R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2710.
- (164) Noga, J.; Kedzuch, S.; Simunek, J. *J. Chem. Phys.* **2007**, *127*, 034106.
- (165) Ten-No, S. *Chem. Phys. Lett.* **2004**, *398*, 56.
- (166) Ten-No, S.; Manby, F. R. *J. Chem. Phys.* **2003**, *119*, 5358.
- (167) Valeev, E. F. *Chem. Phys. Lett.* **2004**, *395*, 190.
- (168) Werner, H. J.; Adler, T. B.; Manby, F. R. *J. Chem. Phys.* **2007**, *126*, 164102.
- (169) Fliegl, H.; Hattig, C.; Klopper, W. *Int. J. Quantum Chem.* **2006**, *106*, 2306.
- (170) Fliegl, H.; Klopper, W.; Hattig, C. *J. Chem. Phys.* **2005**, *122*, 084107.
- (171) Noga, J.; Kedzuch, S.; Simunek, J.; Ten-No, S. *J. Chem. Phys.* **2008**, *128*, 174103.
- (172) Noga, J.; Kutzelnigg, W. *J. Chem. Phys.* **1994**, *101*, 7738.
- (173) Noga, J.; Valiron, P. *Chem. Phys. Lett.* **2000**, *324*, 166.
- (174) Shiozaki, T.; Kamiya, M.; Hirata, S.; Valeev, E. F. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3358.
- (175) Shiozaki, T.; Kamiya, M.; Hirata, S.; Valeev, E. F. *J. Chem. Phys.* **2008**, *129*, 071101.
- (176) Tew, D. P.; Klopper, W.; Hattig, C. *Chem. Phys. Lett.* **2008**, *452*, 326.
- (177) Tew, D. P.; Klopper, W.; Neiss, C.; Hattig, C. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1921.
- (178) Marchetti, O.; Werner, H. J. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3400.
- (179) Kristyan, S.; Pulay, P. *Chem. Phys. Lett.* **1994**, *229*, 175.
- (180) Perez-Jorda, J. M.; Becke, A. D. *Chem. Phys. Lett.* **1995**, *233*, 134.
- (181) Hobza, P.; Sponer, J.; Reschel, T. *J. Comput. Chem.* **1995**, *16*, 1315.
- (182) Cerny, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2005**, *8*, 1624.
- (183) Grimme, S. *Chem.—Eur. J.* **2004**, *10*, 3423.
- (184) Grimme, S. *Angew. Chem., Int. Ed.* **2005**, *45*, 4460.
- (185) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120*, 215.
- (186) Kong, J.; Gan, Z.; Proynov, E.; Freindorf, M.; Furlani, T. R. *Phys. Rev. A* **2009**, *79*, 042510.
- (187) Hepburn, J.; Scoles, G.; Penco, R. *Chem. Phys. Lett.* **1975**, *36*, 451.
- (188) Tang, K. T.; Toennies, J. P. *J. Chem. Phys.* **1977**, *66*, 1496.
- (189) Tang, K. T.; Toennies, J. P. *J. Chem. Phys.* **1978**, *68*, 5501.
- (190) Hobza, P.; Mulder, F.; Sandorfy, C. *J. Am. Chem. Soc.* **1981**, *103*, 1360.
- (191) Hobza, P.; Mulder, F.; Sandorfy, C. *J. Am. Chem. Soc.* **1982**, *104*, 925.
- (192) Hobza, P.; Sandorfy, C. *Can. J. Chem.* **1984**, *62*, 606.
- (193) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc.* **1987**, *109*, 1302.
- (194) Gianturco, F. A.; Paesani, F.; Laranjeira, M. F.; Vassilenko, V.; Cunha, M. A. *Mol. Phys.* **1998**, *94*, 605.
- (195) Gianturco, F. A.; Paesani, F.; Laranjeira, M. F.; Vassilenko, V.; Cunha, M. A. *Chem. Phys.* **1999**, *110*, 7832.
- (196) Wu, X.; Vargas, M. C.; Nayak, S.; Lotrich, V.; Scoles, G. *J. Chem. Phys.* **2001**, *115*, 8748.
- (197) Wu, Q.; Yang, W. *J. Chem. Phys.* **2002**, *116*, 515.
- (198) Zimmerli, U.; Parrinello, M.; Koumoutsakos, P. *J. Chem. Phys.* **2004**, *120*, 2693.
- (199) Elstner, M.; Porezag, D.; Jungnickel, G.; Elsner, J.; Haugk, M.; Frauenheim, T.; Suhai, S.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260.

- (200) Elstner, M.; Hobza, P.; Frauenheim, T.; Suhai, S.; Kaziras, E. *J. Chem. Phys.* **2001**, *114*, 5149.
- (201) Zhechkov, L.; Heine, T.; Patchkowskii, S.; Seifert, G.; Duarte, H. A. *J. Chem. Theory Comput.* **2005**, *1*, 841.
- (202) Hartke, B.; Flad, H.-J.; Dolg, M. *Phys. Chem. Chem. Phys.* **2001**, *3*, 5121.
- (203) Serra, S.; Iarlori, S.; Tosatti, S.; Scandolo, S.; Santoro, G. *Chem. Phys. Lett.* **2000**, *331*, 339.
- (204) Civalieri, B.; Zicovic-Wilson, C. M.; Valenzano, L.; Ugliengo, P. *Cryst. Eng. Commun.* **2008**, *10*, 405.
- (205) Goursot, A.; Mineva, T.; Kevorkyants, R.; Talbi, D. *J. Chem. Theory Comput.* **2007**, *3*, 755.
- (206) Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.
- (207) Steinmann, S. N.; Csonka, G.; Corminboeuf, C. *J. Chem. Theory Comput.* **2009**, *5*, 2950.
- (208) Tkatchenko, A.; Scheffler, M. *Phys. Rev. Lett.* **2009**, *102*, 073005.
- (209) Swart, M.; Sola, M. *J. Chem. Phys.* **2009**, *131*, 094103.
- (210) Grimme, S.; Antony, J.; Schwabe, T.; Muck-Lichtenfeld, C. *Org. Biomol. Chem.* **2007**, *5*, 741.
- (211) Grimme, S.; Muck-Lichtenfeld, C.; Antony, J. *Phys. Chem. Chem. Phys.* **2008**, *10*, 3327.
- (212) Cerny, J.; Jurecka, P.; Hobza, P.; Valdes, H. *J. Phys. Chem. A* **2007**, *111*, 1146.
- (213) Reha, D.; Valdes, H.; Vondrasek, J.; Hobza, P.; Abu-Riziq, A. G.; Crews, B.; de Vries, M. S. *Chem.—Eur. J.* **2005**, *11*, 6803.
- (214) Valdes, H.; Pluhackova, K.; Pitonak, M.; Rezac, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2747.
- (215) Valdes, H.; Vojtech, K.; Pitonak, M.; Exner, O.; Stary, I.; Hobza, P.; Rulisek, L. *J. Comput. Chem.* **2008**, *29*, 861.
- (216) Adamovic, I.; Gordon, M. S. *Mol. Phys.* **2005**, *130*, 379.
- (217) Ruzsinszky, A.; Perdew, J. P.; Csonka, G. I. *J. Phys. Chem. A* **2005**, *109*, 11015.
- (218) Haley, T. P.; Graybill, E. R.; Cybulski, S. M. *J. Chem. Phys.* **2006**, *124*, 204301.
- (219) Krishtal, A.; Vanommesleaghe, K.; Olsasz, A.; Veszpremi, T.; Van Alsenoy, C.; Geerlings, P. *J. Chem. Phys.* **2009**, *130*, 174101.
- (220) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *122*, 154104.
- (221) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2005**, *123*, 154101.
- (222) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2006**, *124*, 114104.
- (223) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2007**, *127*, 154108.
- (224) Johnson, E. R.; Becke, A. D. *J. Chem. Phys.* **2005**, *123*, 024101.
- (225) Johnson, E. R.; Becke, A. D. *Chem. Phys. Lett.* **2006**, *432*, 600.
- (226) Angyan, J. G. *J. Chem. Phys.* **2007**, *127*, 024108.
- (227) Hesselmann, A. *J. Chem. Phys.* **2009**, *130*, 084104.
- (228) Becke, A. D.; Johnson, E. R. *J. Chem. Phys.* **2007**, *127*, 124108.
- (229) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U. *Phys. Rev. Lett.* **2004**, *93*, 153004.
- (230) Tapavicza, E.; Lin, I. C.; von Lilienfeld, O. A.; Tavernelli, I.; Coutinho-Neto, M. D.; Rothlisberger, U. *J. Chem. Theory Comput.* **2007**, *3*, 1673.
- (231) von Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U. *Phys. Rev. B* **2005**, *71*, 195119.
- (232) Tavernelli, I.; Lin, I. C.; Rothlisberger, U. *Phys. Rev. B* **2009**, *79*, 045106.
- (233) Waller, M. P.; Robertazzi, A.; Platts, J. A.; Hibbs, D. E.; Williams, P. A. *J. Comput. Chem.* **2006**, *27*, 491.
- (234) Xu, X.; Goddard, W. A., III. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*.
- (235) Zhang, Y.; Vela, A.; Salahub, D. R. *Theor. Chim. Acta* **2007**, *118*, 693.
- (236) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664.
- (237) Lynch, B. J.; Fast, P. L.; Harris, M.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 4811.
- (238) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6908.
- (239) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372.
- (240) Zhao, Y.; Truhlar, D. G. *J. Chem. Phys.* **2006**, *125*, 194101.
- (241) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 1849.
- (242) Johnson, E. R.; Becke, A. D.; Sherrill, C. D.; DiLabio, G. A. *J. Chem. Phys.* **2009**, *131*, 034111.
- (243) Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2007**, *3*, 289.
- (244) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 4786.
- (245) Zhao, Y.; Lynch, B. J.; Truhlar, D. G. *Phys. Chem. Chem. Phys.* **2005**, *7*, 43.
- (246) Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2006**, *8*, 4398.
- (247) Benighaus, T.; Distasio, R. A.; Lochan, R. C.; Chai, J.-D.; Head-Gordon, M. *J. Phys. Chem. A* **2008**, *112*, 2702.
- (248) Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397.
- (249) Langreth, D. C.; Perdew, J. P. *Solid State Commun.* **1975**, *17*, 1425.
- (250) Furche, F. *J. Chem. Phys.* **2008**, *129*, 114105.
- (251) Andersson, Y.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **1996**, *76*, 102.
- (252) Dobson, J. F.; Dinte, B. P. *Chem. Phys. Lett.* **1996**, *76*, 1780.
- (253) Rapcewicz, K.; Ashcroft, N. W. *Phys. Rev. B* **1991**, *44*, 4032.
- (254) Sato, T.; Tsuneda, T.; Hirao, K. *Mol. Phys.* **2005**, *103*, 1151.
- (255) Sato, T.; Tsuneda, T.; Hirao, K. *J. Chem. Phys.* **2007**, *126*, 234114.
- (256) Grafenstein, J.; Cremer, D. *J. Chem. Phys.* **2009**, *130*, 124105.
- (257) Dion, M.; Rydberg, H.; Scroder, E.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **2004**, *92*, 246401.
- (258) Zhang, Y.; Yang, W. *Phys. Rev. Lett.* **1998**, *80*, 890.
- (259) Thonhauser, T.; Cooper, V. R.; Shen, L.; Puzder, A.; Hyldgaard, P.; Langreth, D. C. *Phys. Rev. B* **2007**, *76*, 125112.
- (260) Gulans, A.; Puska, M. J.; Nieminen, R. M. *Phys. Rev. B* **2009**, *79*, 201105.
- (261) Vydrov, O. A.; Wu, Q.; Van Voorhis, T. J. *J. Chem. Phys.* **2008**, *129*, 014106.
- (262) Vydrov, O. A.; Van Voorhis, T. J. *J. Chem. Phys.* **2009**, *130*, 104105.
- (263) Vydrov, O. A.; Van Voorhis, T. J. *Phys. Rev. Lett.* **2009**, *103*, 063004.
- (264) Sato, T.; Nakai, H. Poster presented at the ICQC Conference, Helsinki, Finland, 2009.
- (265) Sato, T.; Nakai, H. *J. Chem. Phys.* **2009**, *131*, 224104.
- (266) Lein, M.; Dobson, J. F.; Gross, E. K. *J. Comput. Chem.* **1999**, *20*, 12.
- (267) Kohn, W.; Meir, Y.; Makarov, D. E. *Phys. Rev. Lett.* **1998**, *80*, 4153.
- (268) Fuchs, M.; Gonze, X. *Phys. Rev. B* **2002**, *65*, 235109.
- (269) Furche, F.; Van Voorhis, T. J. *J. Chem. Phys.* **2005**, *122*, 164106.
- (270) Harl, J.; Kresse, G. *Phys. Rev. B* **2008**, *77*, 045146.
- (271) Nguyen, H. V.; de Gironcoli, S. *Phys. Rev. B* **2009**, *79*, 205114.
- (272) Angyan, J. G.; Gerber, I. C.; Savin, A.; Toulouse, J. *Phys. Rev. A* **2005**, *72*, 012510.
- (273) Gerber, I. C.; Angyan, J. G. *Chem. Phys. Lett.* **2005**, *416*, 370.
- (274) Goll, E.; Werner, H.-J.; Stoll, H. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3917.
- (275) Silvestrelli, P. L. *Phys. Rev. Lett.* **2008**, *100*, 053002.
- (276) Chang, D. T.; Schenter, G. K.; Garrett, B. C. *J. Chem. Phys.* **2008**, *128*, 164111.
- (277) Maerzke, K. A.; Murdachaew, G.; Mundy, C. J.; Schenter, G. K.; Siepmann, I. J. *J. Phys. Chem. A* **2009**, *113*, 2075.
- (278) Mackie, I. D.; DiLabio, G. A. *J. Phys. Chem. A* **2008**, *112*, 10968.
- (279) Wesolowski, T. A.; Tran, F. *J. Chem. Phys.* **2003**, *118*, 2072.
- (280) Jeziorski, B.; Moszynski, R.; Szalewicz, K. *Chem. Rev.* **1994**, *94*, 1887.
- (281) Williams, H. L.; Chabalowski, C. F. *J. Phys. Chem. A* **2001**, *105*, 646.
- (282) Misquitta, A. J.; Szalewicz, K. *Chem. Phys. Lett.* **2002**, *357*, 301.
- (283) Misquitta, A. J.; Jeziorski, B.; Szalewicz, K. *Phys. Rev. Lett.* **2003**, *91*, 033201.
- (284) Misquitta, A. J.; Podeszwa, R.; Jeziorski, B.; Szalewicz, K. *J. Chem. Phys.* **2005**, *123*, 214103.
- (285) Podeszwa, R.; Bukowski, R.; Szalewicz, K. *J. Chem. Theory Comput.* **2006**, *2*, 400.
- (286) Hesselmann, A.; Jansen, G. *Chem. Phys. Lett.* **2002**, *357*, 464.
- (287) Hesselmann, A.; Jansen, G. *Chem. Phys. Lett.* **2002**, *362*, 319.
- (288) Hesselmann, A.; Jansen, G. *Chem. Phys. Lett.* **2003**, *367*, 778.
- (289) Jansen, G.; Hesselmann, A. *J. Phys. Chem. A* **2001**, *105*, 11156.
- (290) MOLPRO, version 2009.1, a package of ab initio programs. Werner, H. J.; Knowles, P. J.; Lindh, R.; Manby, F. R.; Schütz, M.; Celani, P.; Korona, T.; Rauhut, G.; Amos, R. D.; Bernhardsson, A.; Berning, A.; Cooper, D. L.; Deegan, M. J. O.; Dobbyn, A. J.; Eckert, F.; Hampel, C.; Hetzer, G.; Lloyd, A. W.; McNicholas, S. J.; Meyer, W.; Mura, M. E.; Nicklass, A.; Palmieri, P.; Pitzer, P.; Schumann, U.; Stoll, H.; Stone, A. J.; Tarroni, R.; Thorsteinsson, T., 2009.
- (291) Bukowski, R.; Cencek, W.; Jankowski, P.; Jeziorska, M.; Jeziorski, B.; Kucharski, S. A.; Lotrich, V. F.; Misquitta, A. J.; Moszynski, R.; Patkowski, K.; Podeszwa, R.; Rybak, S.; Szalewicz, K.; Williams, H. L.; Wheatley, R. J.; Wormer, P. E. S.; Zuchowski, P. S. SAPT2008—An Ab Initio Program for Many-Body Symmetry-Adapted Perturbation Theory Calculations of Intermolecular Interaction Energies.
- (292) Thiel, W. *Adv. Chem. Phys.* **1996**, *93*, 703.
- (293) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4907.
- (294) Dewar, M. J. S.; Thiel, W. *J. Am. Chem. Soc.* **1977**, *99*, 4899.
- (295) Dewar, M. J. S.; Zebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
- (296) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 221.
- (297) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 209.
- (298) Giese, T. J.; Sherer, E. C.; Cramer, C. J.; York, D. M. *J. Chem. Theory Comput.* **2005**, *1*, 1275.
- (299) Martin, B.; Clark, T. *Int. J. Quantum Chem.* **2006**, *106*, 1208.
- (300) Bernal-Uruchurtu, M. I.; Martins-Costa, M. T. C.; Millot, C.; Ruiz-Lopez, M. F. *J. Comput. Chem.* **2000**, *21*, 572.
- (301) Morgado, C. A.; McNamara, J. P.; Hillier, I. H.; Burton, N. A. *J. Chem. Theory Comput.* **2007**, *3*, 1656.
- (302) Thiel, W.; Voityuk, A. *Theor. Chim. Acta* **1992**, *81*, 391.
- (303) Thiel, W.; Voityuk, A. *Theor. Chim. Acta* **1996**, *93*, 315.

- (304) Thiel, W.; Voityuk, A. *J. Phys. Chem.* **1996**, *100*, 616.
- (305) Tuttle, T.; Thiel, W. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2159.
- (306) Repasky, M. P.; Chandrasakhar, J.; Jorgensen, W. L. *J. Comput. Chem.* **2002**, *23*, 1601.
- (307) Winget, P.; Selcuki, C.; Horn, A. H. C.; Martin, B.; Clark, T. *Theor. Chim. Acta* **2003**, *110*, 254.
- (308) Wang, Q.; Bryce, R. A. *J. Chem. Theory Comput.* **2009**, *5*, 2206.
- (309) Field, H. M.; Bash, P. A.; Karplus, M. *J. Comput. Chem.* **1990**, *11*, 700.
- (310) Stewart, J. J. P. *J. Mol. Model.* **2007**, *13*, 1173.
- (311) Stewart, J. J. P. *J. Mol. Model.* **2008**, *14*, 499.
- (312) <http://OpenMOPAC.net>.
- (313) Paton, R. S.; Goodman, J. *J. Chem. Inf. Model.* **2009**, *49*, 944.
- (314) Rezac, J.; Fanfrlik, J.; Salahub, D. R.; Hobza, P. *J. Chem. Theory Comput.* **2009**, *5*, 1749.
- (315) Korth, M.; Pitonak, M.; Rezac, J.; Hobza, P. *J. Chem. Theory Comput.* **2010**, *6*, 344.
- (316) Takatani, T.; Hohenstein, E. G.; Malagoli, M.; Marshall, M. S.; Sherrill, C. D. *J. Chem. Phys.* **2010**, *132*, 144104.
- (317) Korth, M.; Luchow, A.; Grimme, S. *J. Phys. Chem. A* **2008**, *112*, 2104.
- (318) Zhao, Y.; Truhlar, D. G. *J. Phys. Chem. C* **2008**, *112*, 4061.
- (319) Kannemann, F. O.; Becke, A. D. *J. Chem. Theory Comput.* **2010**, *6*, 1081.
- (320) Chai, J.-D.; Head-Gordon, M. *J. Chem. Phys.* **2008**, *128*, 084106.
- (321) Hehre, W. J.; Radom, L.; Schleyer, P. v. R.; Pople, J. A. *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986.
- (322) Riley, K. E.; Op't Holt, B. T.; Merz, K. M. *J. Chem. Theory Comput.* **2007**, *3*, 407.
- (323) Morgado, C. A.; Vincent, M. A.; Hillier, I. H.; Shan, X. *Phys. Chem. Chem. Phys.* **2007**, *9*, 448.
- (324) Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.
- (325) Ordejon, P.; Artacho, E.; Soler, J. M. *Phys. Rev. B* **1996**, *53*, R10441.
- (326) Riley, K. E.; Pitonak, M.; Cerny, J.; Hobza, P. *J. Chem. Theory Comput.* **2010**, *6*, 66.
- (327) Morgado, C. A.; Jurecka, P.; Svozil, D.; Hobza, P.; Sponer, J. *J. Chem. Theory Comput.* **2009**, *5*, 1524.
- (328) Ran, J.; Hobza, P. *J. Chem. Theory Comput.* **2009**, *5*, 1180.
- (329) Arnstein, S. A.; Sherrill, C. D. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2646.
- (330) Hohenstein, E. G.; Sherrill, C. D. *J. Phys. Chem. A* **2009**, *113*, 878.
- (331) Ringer, A. L.; Figgis, M. S.; Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2006**, *110*, 10822.
- (332) Ringer, A. L.; Sinnokrot, M. O.; Lively, R. P.; Sherrill, C. D. *Chem.—Eur. J.* **2006**, *12*, 3821.
- (333) Sinnokrot, M. O.; Sherrill, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 7690.
- (334) Sinnokrot, M. O.; Sherrill, C. D. *J. Phys. Chem. A* **2004**, *108*, 10200.
- (335) Sinnokrot, M. O.; Valeev, E. F.; Sherrill, C. D. *J. Am. Chem. Soc.* **2002**, *124*, 10887.
- (336) Tauer, T. P.; Derrick, M. E.; Sherrill, C. D. *J. Phys. Chem. A* **2005**, *109*, 191.
- (337) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M. *J. Chem. Phys.* **2006**, *124*, 7.
- (338) Tsuzuki, S.; Uchimaru, T. *Curr. Org. Chem.* **2006**, *10*, 745.
- (339) Tsuzuki, S.; Uchimaru, T.; Mikami, M.; Tanabe, K. *J. Phys. Chem. A* **2002**, *106*, 3867.
- (340) Tsuzuki, S.; Honda, K.; Uchimaru, T.; Mikami, M. *J. Chem. Phys.* **2006**, *125*, 124304.
- (341) Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157.
- (342) Jalkanen, J. P.; Mahlanen, R.; Pakkanen, T. A.; Rowley, R. L. *J. Chem. Phys.* **2002**, *116*, 1303.
- (343) Jalkanen, J. P.; Pakkanen, T. A.; Rowley, R. L. *J. Chem. Phys.* **2004**, *120*, 1705.
- (344) Jalkanen, J. P.; Pulkkinen, S.; Pakkanen, T. A.; Rowley, R. L. *J. Phys. Chem. A* **2005**, *109*, 2866.
- (345) Rowley, R. L.; Tracy, C. M.; Pakkanen, T. A. *J. Chem. Phys.* **2006**, *125*, 154302.
- (346) Rowley, R. L.; Tracy, C. M.; Pakkanen, T. A. *J. Chem. Phys.* **2007**, *127*, 025101.
- (347) Tekin, A.; Jansen, G. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1680.
- (348) Aeberhard, P. C.; Arey, J. S.; Lin, I. C.; Rothlisberger, U. *J. Chem. Theory Comput.* **2009**, *5*, 23.
- (349) Podeszwa, R.; Bukowski, R.; Szalewicz, K. *J. Phys. Chem. A* **2006**, *110*, 10345.
- (350) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.*, in press.
- (351) Dabkowska, I.; Jurecka, P.; Hobza, P. *J. Chem. Phys.* **2005**, *122*, 204322.
- (352) Morgado, C. A.; Jurecka, P.; Svozil, D.; Hobza, P.; Sponer, J. *Phys. Chem. Chem. Phys.* **2009**, *12*, 3522.
- (353) Fusti Molnar, L.; He, X.; Wang, B.; Merz, K. M. *J. Chem. Phys.* **2009**, *131*, 065102.
- (354) Cerny, J.; Pitonak, M.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.*, to be submitted for publication.
- (355) Simon, S.; Duran, M.; Dannenberg, J. J. *J. Chem. Phys.* **1996**, *105*, 11024.
- (356) van Mourik, T.; Karamertzanis, P. G.; Price, S. L. *J. Phys. Chem. A* **2006**, *110*, 8.
- (357) Jensen, F. *Chem. Phys. Lett.* **1996**, *261*, 633.
- (358) Shields, A. E.; van Mourik, T. *J. Phys. Chem. A* **2007**, *111*, 13272.
- (359) van Mourik, T. *J. Chem. Theory Comput.* **2008**, *4*, 1610.
- (360) Häber, T.; Seefeld, K.; Engler, G.; Grimme, S.; Kleinermanns, K. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2844.
- (361) Valdes, H.; Pluhackova, K.; Hobza, P. *J. Chem. Theory Comput.* **2009**, *5*, 2248.
- (362) Valdés, H.; Řeha, D.; Hobza, P. *J. Phys. Chem. B* **2006**, *110*, 6385.
- (363) Valdes, H.; Spiwok, C.; Rezac, J.; Řeha, D.; Abu-Riziq, A. G.; de Vries, M. S.; Hobza, P. *Chem.—Eur. J.* **2008**, *14*, 4886.
- (364) Toroz, D.; van Mourik, T. *Mol. Phys.* **2007**, *105*, 209.
- (365) Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2007**, *9*, 1463.
- (366) Doslic, N.; Kovacevic, G.; Ljubic, I. *J. Phys. Chem. A* **2007**, *111*, 8650.