Noncovalent interactions in biochemistry

Kevin E. Riley[1] and Pavel Hobza[2,3]*

Noncovalent interactions are known to play a key role in biochemistry. The knowledge of stabilization (relative) energies and their components is very important for understanding the nature of these interactions. Accurate and benchmark data on interaction (relative) energies and structures can be obtained from coupled-cluster with single and double and perturbative triple excitations [CCSD(T)] calculations with extended basis of atomic orbitals or even at the complete basis set limit. These methods cannot be, however, used for systems larger than about 50 atoms. In this contribution, the applicability and performance of various recently introduced wavefunction and density functional methods are examined in detail. It is shown that a very good performance by some of these methods is obtained only by introducing empirical parameters fitted mostly to CCSD(T) benchmark data. Among the methods described, special attention is paid to two techniques. First, the symmetry-adapted perturbation technique that allows obtaining not only accurate values of total interaction energies but also their components. Results of these calculations reveal a key role of dispersion energy in stabilizing the structures of biomolecular systems. Second, the semiempirical quantum chemical parameterized model 6 method (PM6) augmented by empirical terms describing the dispersion and H-bonding energies. The method is suitable for much extended systems having several thousands of atoms and can be thus used, e.g., in the drug design. *WIREs Comput Mol Sci* 2011 1 3–17 DOI: 10.1002/wcms.8

INTRODUCTION

Historically, chemistry has been characterized, in large part, as the study of the covalent interactions occurring within molecules. Indeed, the description of the covalent bond is among the great triumphs of modern chemistry. The nature of the formation (and breaking) of covalent bonds is well understood and nowadays computational chemistry techniques can describe these processes very well, with chemical (~1 kcal/mol) or even subchemical (~0.1 kcal/mol) accuracy. Covalent bonding is thus, whether we like it or not, more or less a closed chapter (at least for main group elements). The opposite is true, however, when one considers noncovalent interactions in which we have been witnessing a dynamical growth in our knowledge and understanding in the last several years. Despite enormous progress that has been made in theory as well as experiment, we are still far from obtaining unambiguous information about noncovalent complexes, with the largest gaps in our knowledge being for extended noncovalent complexes. Why are noncovalent interactions so relevant in modern research? The answer should be sought in the role that noncovalent interactions play in biomolecular structures and nanostructures.

All life on our earth can be viewed as an application of supramolecular chemistry, with noncovalent interactions 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, and so forming a deep understanding of the nature of the stabilization of these systems is 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 three-dimensional (3D) structure of DNA (and also of other biomacromolecules such as proteins, etc.) results from a delicate balance between

*Correspondence to: pavel.hobza@uochb.cas.cz

[1]Department of Chemistry, University of Puerto Rico, Rio Piedras, Puerto Rico

[2]Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic and Center for Biomolecules and Complex Molecular Systems, Prague, Czech Republic

[3]Department of Physical Chemistry, Palacky University, Olomouc, Czech Republic

DOI: 10.1002/wcms.8

various types of noncovalent interactions acting between biomolecular building blocks and also between these blocks and the surrounding environment. Covalent bonding plays a minor role here, although it should be noted that disulfide bonds play a role in determining the structures of some proteins. Among the most important noncovalent interaction types are hydrogen bonding (H-bonding), stacking, X—H$\cdots\pi$, cation$\cdots$anion, cation$\cdots\pi$, and dispersion interactions, with other special (less common) interaction types, such as dihydrogen[1,2] and halogen[3–5] bonding, also being important. These motifs are found not only in nucleic acids and proteins but also in complexes of nucleic acids or proteins with ligands. In this context, a ligand could be a hormone, enzyme substrate, pharmaceutical compound, or environmental chemical (e.g., an industrial carcinogen). For a long time, it was believed that H-bonding is much stronger than stacking and, consequently, it was expected that H-bonding is the key player in determining biomolecular structure. Accurate quantum mechanical (QM) calculations performed recently in our laboratory have, surprisingly, shown that the stabilization of both motifs can in fact be comparable.[6,7] These calculations were difficult because of the different origins of H-bonding and stacking interactions. Although an H-bond forms due to electrostatic interactions and charge transfer, the stabilization of stacking interactions is largely governed by the London dispersion energy (although electrostatics can be responsible for the geometrical orientation of subsystems). The London dispersion energy is also, to a large extent, responsible for the stabilization in dispersion-bound and X—H$\cdots\pi$ complexes. The accurate description of H-bonding is straightforward and almost any QM level describes H-bonding acceptably well. It is an entirely different situation for the stacking interactions, in which the most accurate QM calculations are required. To properly qualitatively and quantitatively describe the stabilization of various motifs of biomolecular building blocks, it is thus necessary to obtain high-accuracy results. This may not be apparent, so let us mention just one example. The very popular density functional theory (DFT) methods (using standard functionals), which are now extensively used in the realm of chemistry and biochemistry, fail completely to describe the dispersion energy. Hence, their use cannot be recommended because they are not able to cover, e.g., stacking stabilization. Obtaining accurate absolute binding energy values would be desirable, but this is an unreachable goal for extended systems occurring in biology; reliable relative values would be sufficient for most purposes. Having the ability to obtain these values is highly important because it allows us to understand the nature of stabilization in biomolecular systems as well as the nature of key biochemical and biological processes. It must be kept in mind that this understanding is *conditio sine qua non* for the subsequent steps leading to targeted mutation of biosystems to modify these processes.

In our laboratory, and in others, it has been shown that it is possible to obtain stabilization energies for noncovalent complexes with chemical accuracy ($\sim$1 kcal/mol) 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 geometries and other properties) for all of the most important binding motifs found in biomolecules. CCSD(T)/CBS calculations thus provide benchmark data that can not only be used for investigating the nature of noncovalent interactions in various binding motifs but also in providing data for testing and/or parameterizing other, more computationally economical, (nonempirical and semiempirical) model chemistries based on wavefunction theory (WFT) and DFT as well as empirical potentials (EP). The very unique position of the CCSD(T)/CBS technique follows from the fact that it is the only theoretical *ab initio* procedure providing accurate stabilization energies for various types of noncovalent complexes. All other WFT, DFT, and EP procedures describing noncovalent interactions at least qualitatively contain one or more parameters that were most likely fit against CCSD(T)/CBS or experimental data. There is only one exception to this, the symmetry-adapted perturbation theory (SAPT), which provides accurate interaction energies determined as a sum of various energy contributions (electrostatic, repulsion, induction, dispersion, ...) having clear physical meaning. The SAPT technique and its variant, DFT–SAPT (for references see below), are genuine *ab initio* methods.

The development of faster QM procedures is highly important because 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 basically be described by EP, 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. An equally important manifestation of quantum effects is the evaluation of atomic charges. In the EP, atomic charges are assigned at the beginning of a simulation/calculation and remain constant during the entire process. QM calculations have shown, however, that these charges change dramatically depending on the neighboring groups and environment. Fast and accurate QM methods are also needed in molecular dynamics (MD) simulations in which the description of quantum effects plays a decisive role. The development of such methods is very difficult because the computational time cannot be much higher than that required by EP methods.

As mentioned above, the primary biochemical property of any system, its structure, is determined not only by various noncovalent interactions but also by interaction of the system with the environment. The role of the highly polar water environment is critical, which is not surprising, considering the fact that almost all biochemical systems are not only highly polar but also generally even charged. It is not the aim of the present review to deal with solvent effects, but it must be stressed that these effects are of key importance. The following are few examples of cases in which solvation effects play a large role: (1) The double-helical structure of DNA, which plays a key role in its function, is partially due to hydrophobic interactions of water. Gradually, diminishing the amount of water by adding (less polar) methanol causes the DNA structure to collapse. (2) The salt bridges in proteins are extremely stable (stabilization of more than 100 kcal/mol), but when the system is exposed to water, the interaction strength is strongly reduced (sometimes even eliminated). (3) Guanine and cytosine form an extremely stable H-bonded pair in the gas phase, with a stabilization energy of more than 30 kcal/mol; the less stable-stacked structure of this pair has a stabilization energy of about 18 kcal/mol. The difference in stabilization energy is so large that only the H-bonded structure can be detected (at nonzero temperatures). Passing to the water environment, only the stacked structure exists, which is due to much better solvation of this structure.

As mentioned above, today's computational chemistry techniques can determine stabilization energies, even of extended complexes, with chemical accuracy. On the contrary, the evaluation of solvation energies is connected with much higher uncertainties and it is thus a prime goal of computational chemistry to solve this problem. For further references, there are several studies investigating the effects of solvation on intermolecular interactions using WFT and DFT methods.[8–11]

In the limited space of this article, we provide review information on an extremely extended subject—noncovalent interactions in biochemistry. We consider this short review to be a supplement to two recently finished extended projects from our laboratory concerning very similar topics in which the reader can find more in-depth information. The first one is our book entitled '*Noncovalent Interactions. Theory and Experiment*'[12] in which equal importance is given to theoretical and experimental investigations of noncovalent interactions. The second one is the Chemical Review article entitled 'Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories',[13] which is focused mainly on the calculation of interaction energies. The book and review provide a large number of references, which for understandable reasons cannot be included in the present review.

The aim of the present review is to familiarize the reader with the field of calculation of biochemical noncovalent interactions and, in addition, to provide recent representative references. We also try to recommend selected techniques suitable for systems of varying size. The main focus is on biodisciplines, but all results and conclusions are, without any modification, valid in the nanoworld as well.

INTERACTION ENERGIES

Benchmark CCSD(T)/CBS Limit Calculations

The CCSD(T)/CBS level can be attained via extrapolation of self consistent field (SCF), second-order Møller–Plesset perturbation theory (MP2), and higher-order correlation energies toward the basis set limit. Each of the components mentioned is differently sensitive to the atomic orbital (AO) basis set; SCF calculations are easily carried out even for extended complexes, and for these as large a basis set as possible is recommended (e.g., aug-cc-pVTZ or cc-pVQZ). The MP2 interaction energy is the slowest converging and the larger basis set used in the extrapolation (e.g., Helgaker type) the better; the lowest acceptable level for this extrapolation is the one using the aug-cc-pVDZ and aug-cc-pVTZ (or similarly constructed) basis sets. The use of large basis sets for MP2 has become possible because one can use the density fitting/resolution-of-the-identity approximation for two-electron integrals, which lowers the computational cost of these calculations by at least an order of magnitude.[14,15] The last term, called

the ΔCCSD(T) correction term, is determined as a difference between CCSD(T) and MP2 interaction energies and converges much faster than other terms. The use of such a term is possible because, as it has been shown, MP2 and CCSD(T) binding energies converge with basis set size in a very similar way. The use of small basis sets [e.g., 6–31G*(0.25) or similar] can lead to underestimation of the ΔCCSD(T) term by not more than 10–20% and the use of the aug-cc-pVTZ basis set leads to 1–3% agreement with the CBS value. The CCSD(T)/CBS energy is constructed on the basis of Eqs (1) and (2) and is in very good agreement with CCSD(T) interaction energies extrapolated to the CBS limit directly from CCSD(T) energies.

$$\Delta E_{\text{est}}^{\text{CCSD(T)}} \approx \Delta E_{\text{CBS}}^{\text{HF}} + \Delta E_{\text{CBS}}^{\text{MP2}} + \Delta\text{CCSD(T)} \qquad (1)$$

$$\Delta\text{CCSD(T)} = (\Delta E^{\text{CCSD(T)}} - \Delta E^{\text{MP2}})|_{\text{small basis}} \qquad (2)$$

It should be noted here that one of the largest sources of errors in both WFT- and DFT-based supermolecular calculations is attributable to the incompleteness of basis sets and the resulting basis set superposition error (BSSE). The effects of BSSE can often be minimized by using the counterpoise (CP) corrections of Boys and Bernardi.[16] The other possibility for eliminating the BSSE is to work with extended AO basis sets or preferably at the CBS limit.

WFT Calculations

Despite the amount of progress in the development of hardware and software, the determination of CCSD(T)/CBS energies is, and will be in the near future, limited to systems with no more than 35 atoms. Hence, a prime goal of computational chemistry is to find accurate low-order scaling WFT methods. The price we pay for this is, however, large. It is necessary to leave the *ab initio* world (where no empirical parameters are considered) and to introduce one or more empirical parameters.

The supermolecular MP2 has long been, and still is, one of the most frequently applied WFT methods for calculation of noncovalent interactions. This method, when used along with CBS extrapolation, provides very good estimates of interaction energies for H-bonded systems but fails for stacking or, generally speaking, dispersion-bound complexes. This method generally gives stabilization energies that are strongly overestimated at potential energy minima; MP2/CBS does, however, give reasonable results for all systems at long-range distances and also describes interactions between saturated molecules (e.g., alkanes and biologically important systems such as sugars) fairly well. The use of MP2 along with small or medium basis sets can provide more reasonable energies, but this is only due to error compensation and such calculations should only be used with great care.[17] The reason for MP2/CBS overbinding is that the MP2 supermolecular interaction energies contain only the uncoupled HF dispersion energy, which is known to be overestimated by 10–20%. There are few possibilities for removing this problem. The most natural way is to remove the 'wrong' dispersion energy and to add the correct one.[18] Because of the divergence of the dispersion energy, a damping function containing empirical parameters should be introduced. Let us add here that damping functions should be used every time an empirical dispersion energy is added; the same problems appeared with the DFT-D method (see below). The method is very promising but more assessment is still necessary. The other possibility is to use the spin-component-scaled (SCS) MP2 methods pioneered by Grimme.[19] The method is based on separate scaling of the same (singlet or antiparallel) and opposite (triplet or parallel) components of the correlation energy and it is also extendible for other WFT methods (see, e.g., the very promising SCS–CCSD method by Sherrill and coworkers).[20] The scaling components for these methods are either deduced from theory or parameterized against benchmark data. Finally, the last possibility is to combine the MP2 method with a higher-order, e.g., MP3, method. It was observed that stabilizations in π–π stacking complexes are overestimated by MP2 and underestimated by MP3 by almost the same magnitude. The proposed 'MP2.5' method corrects the MP2 interaction energy using a scaled (scaling factor being 0.5) third-order MP3 correlation correction.[21] Unlike the case for other empirical approaches, such as SCS–MP2, the proper asymptotic scaling of both MP2 and MP3 is retained and MP2.5's performance is in principle never considerably worse than that of MP2 but can be significantly better.

By far, the most activity is devoted to methods of the second class and they will be mentioned briefly. The original SCS–MP2 removes the most serious problem of the MP2 method, namely the strong overestimation of stacking energies.[22] On the contrary, the technique removes the strong point of the MP2 method, namely excellent performance for H-bonding. Different modifications of the original methods have been suggested [SCS(MI)–MP2,[23] SSS(MI)–MP2,[23] SCSN–MP2,[24] SOS–MP2,[25] MOS–MP2,[26] SOS(MI)–MP2[24]] to remove this problem and they all are based on parameterization of one or both components of correlation energy against different datasets. The problem was also solved quite pragmatically by introducing the dispersion-weighted average

of the MP2 and SCS–MP2 method; for H-bonded and stacked complexes, the MP2 and SCS–MP2, respectively, techniques are adopted, whereas for 'mixed' complexes, the switching function is utilized, the S22 dataset (see later) was used for parameterization.[27]

DFT Calculations

The widely applied density functionals generally provide rather poor results for noncovalent interactions, particularly dispersion-dominated interactions. More than 20 years ago, it was shown that standard DFT calculations do not cover the dispersion energy and these methods completely failed, e.g., evaluation of stacking interactions.[28–30] It is, however, necessary to note that the DFT itself is capable of providing the exact solution to the Schrödinger equation, including the long-range correlation—the dispersion. The reason for the rather unsatisfactory description of the intermolecular interactions in DFT lies in the approximations made in the DFT functionals, not in DFT itself. The most widely used Generalized Gradient Approximation (GGA) determines exchange-correlation (XC) energy on the basis of electron density and reduced electron density gradient at a given grid point. Because the XC energy is calculated only from local properties of the density, it is necessarily local. As a consequence, the dispersion energy, which is an inherently nonlocal property, is not explicitly covered by the GGA functionals. This also holds, with only small differences, for other commonly used DFT functionals—LDA, hybrid, and meta-GGA.

DFT has several advantages over WFT (e.g., much more favorable CPU timing) and this has triggered a huge effort in the last few years to overcome the problem concerning the lack of dispersion energy. A variety of methods were suggested, ranging from the simplest but surprisingly successful addition of empirical corrections, to the nonlocal pseudopotentials in the plane wave methods, to the reparameterization of current functionals, and the construction of truly nonlocal new density functionals.

The simplest way to introduce dispersion into DFT is to add an empirical London-type dispersion correction based on the well-known asymptotic formulas. The added dispersion term should be damped by an empirical damping function that, among other things, reduces the dispersion attraction at intermediate and short distances. Let us add here that this technique is not new and was used more than 30 years ago for HF calculations[31] and about 10 years ago for approximate DFT calculations.[32] Presently, two versions of the so-called DFT-D method are broadly used, those of Grimme[33] and Jurecka et al.[34] Both approaches are similar and differ only when used for condensed phases such as in molecular crystals.[35] The damped atomic dispersion corrections provided very accurate descriptions of the noncovalent complexes as the results are statistically better than those of MP2 with medium and large basis sets.[34] The strong point of these calculations is their computational economy and the method can be recommended for general calculations on medium and large noncovalent complexes, where they provide reliable energy and geometry characteristics.

The reparameterization of current density functionals represents another simple and very effective way of improving the DFT technique. It must be kept in mind that most of the widely used density functionals, such as B3LYP, BLYP, PBE, and PW91, were designed or parameterized without consideration of noncovalent complexes and, consequently, some attraction near the van der Waals (vdW) region is missing. A suitably modified functional might, however, provide enough additional attraction in these regions, which could simulate the effect of the missing long-range correlation (with the $1/r^6$ asymptotic behavior). The problem is that this attraction will most likely be spurious and one would get 'the almost right answer for the wrong reason'. In other words, these functionals do not exhibit the expected $1/r^6$ asymptotic behavior, which makes their use at long range questionable. In the vdW region they can, however, provide quite reliable interaction energies.

This approach was pioneered by Zhao and Truhlar, who designed a series of new functionals, the most recent of which are contained in the M06[36] and M08[37] suites of functionals. For general use, the M06–2X functional exhibits good results not only for noncovalent complexes but also for other ground- and excited-state properties. It is evident that within distances of nonnegligible overlap, the empirically reparameterized functionals can provide surprisingly accurate stabilization energies for different classes on noncovalent complexes, including the most difficult class—the dispersion-bound one.

Symmetry-Adapted Perturbation Treatment

The perturbation method is naturally suited for the evaluation of the interaction energy because it is determined directly as the sum of first- (electrostatic and exchange repulsion), second- (induction and dispersion), and higher-order perturbation contributions. The main advantage of the perturbative SAPT treatment is the fact that the interaction energy is determined directly without any further problems (such as the BSSE in the supermolecular method) and it

also provides well-defined and physically meaningful energy terms. When using correlated monomer wavefunctions, the resulting energies are highly accurate but the required computation time is enormous, preventing the use of the method for extended systems. A significant improvement has been reached by utilizing a combination of the SAPT and DFT theories, resulting in DFT–SAPT,[38–43] which can even be used in combination with large basis sets (such as cc-pVTZ) for such extended complexes as DNA base pairs. The DFT–SAPT [or SAPT(DFT)] total interaction energy is determined as the sum of the polarization (electrostatic), exchange repulsion, induction, dispersion, and δ(HF) terms, the induction and dispersion components also having their exchange counterparts,

$$E_{\text{int}} = E^1_{\text{pol}} + E^1_{\text{ex}} + E^2_{\text{ind}} + E^2_{\text{ex-ind}} + E^2_{\text{disp}} + E^2_{\text{ex-disp}} + \delta\text{HF} \quad (3)$$

The δ(HF) term represents the estimated higher-order Hartree–Fock (HF) contributions and is determined as a difference between the variational HF interaction energy and the sum of electrostatic, exchange-repulsion, induction, and exchange-induction energies (up to the second perturbation order). Because of this construction, the SAPT–DFT interaction energy is not strictly BSSE free.

Semiempirical QM Calculations

Semiempirical QM methods can properly and fully describe all quantum effects, which make their use very attractive. Semiempirical methods were originally developed for theoretical studies on extended covalent systems, for which the computational cost of nonemepirical *ab initio* methods was prohibitively high. The use of these methods for noncovalent complexes (H-bonded and especially dispersion bound) was limited. Considerable improvements were made (similarly as in the case of DFT methods) by addition of empirical corrections for dispersion, and later also for H-bonding. Very promising results have been obtained with the new parameterized model 6 (PM6) semiempirical method (the method is parameterized for the entire periodic table) by Stewart[44,45] after addition of two modifications: (1) an empirical dispersion energy term, which improves the description of complexes controlled by the dispersion energy, and (2) an additional directional electrostatic term, which improves the description of H-bonded complexes. The resulting method, PM6-DH,[46,47] has an ambitious aim—to achieve standard *ab initio* chemical accuracy (~1 kcal/mol) for extended noncovalent complexes.

Empirical Force Fields (Molecular Mechanics)

Biomolecular structures are floppy and dynamic, which means that MD simulation techniques are necessary to study their behavior. Despite enormous efforts in computational sciences, the use of *ab initio* MD simulations for extended systems is impractical. The vast majority of simulations are, and will continue to be, performed with EPs. It is important to mention that even very long MD simulations, when performed with inaccurate potentials, can only yield inaccurate results. All EPs have been parameterized against experimental and theoretical values, but their validation has been, until recently, difficult, which is mainly because of the lack of suitable benchmark data.

Benchmark Database Sets

Highly accurate QM calculations provide information on the nature of processes governed by noncovalent interactions. Furthermore, and this is equally important, these results can be used for parameterization and/or validation of lower-level computational methods. We have shown that, among all computational procedures considered, only the CCSD(T)/CBS and DFT–SAPT procedures represent genuine *ab initio* techniques, and all other methods providing reliable information on noncovalent interactions should be parameterized. The benchmark database set should cover all important bonding motifs as well as all typical subsystem characteristics (such as atom hybridization) and should be easily extended in the future for new structural motifs. Among the most widely used are Zhao and Truhlar's databases[48] and the S22 and JSCH-2005 databases[49] from our laboratory. Especially, the S22 set, which contains H-bonded, dispersion-bound, and mixed complexes, has become very popular and is widely used for parameterization. It is very important to test the accuracy of the benchmark results obtained in these databases. On the basis of studies done in our laboratory and others, where the accuracy of the presently used CCSD(T)/CBS method was tested, it is possible to state that the present data are generally accurate to within about 5% of the most accurate values.[27,50–53] The S22 stabilization energies systematically represent the upper limit for stacked systems and lower limit for H-bonded pairs.

GEOMETRY

It is critical that we be able to describe the interactions within a complex not only at the potential energy

minimum but also as a function of the relative positions and orientations of interacting systems. Noncovalent interactions are extremely sensitive to geometric parameters, a characteristic that leads to fine-tuning in molecular recognition phenomena within biomolecular systems. Most studies assessing the quality of a particular method for treating noncovalent interactions focus heavily on binding energies at (or near) the potential energy minima. Here, we will describe some results obtained for varying molecular configurations of molecular complexes as well as systems containing intramolecular noncovalent interactions.

There are several reasons that it is very important to understand the ways that binding energies vary with geometric parameters. The strength of an interaction can be very sensitive to small geometric perturbations, which has large implications for the structures and stabilities of proteins, nucleic acids, and protein $\cdots$ ligand complexes. Dynamic phenomena within a biomolecule depends, of course, on the total free energy surface for that system, noncovalent interactions make large contributions to the free energy landscapes of most biomolecules. Having methods that describe geometry characteristics well at a reasonably small computational cost is very important in obtaining theoretical vibrational spectra, which can be used to interpret the results of experimental infrared spectra. It is necessary to obtain high-level computational data for noncovalent geometric dependencies to be able to assess/parameterize lower-level methods that can be used to treat large systems.

There are generally three types of studies that lend themselves to investigations of the geometric dependence of noncovalent interactions: (1) generation of potential energy curves for complexes (generally along the direct dissociation pathway), (2) geometry optimizations (generally utilizing analytical gradients), and (3) conformational studies of systems containing intramolecular interactions. Here, we will discuss two of these, namely categories (1) and (3); we will note here that there are few studies dealing with geometry optimizations at a very high level of theory.[54,55]

In the last several years, there have been a number of studies seeking to characterize potential energy curves for noncovalent interactions. Of note are studies generating potential energy curves using the (direct or estimated) CCSD(T) method, at either the CBS or using basis sets at least as large as aug-cc-pVTZ, by Hobza and coworkers,[52,56–58] Sherrill and coworkers,[59–69] Tsuzuki et al.,[70–73] Zhao and Truhlar,[74,75] Rothlisberger and coworkers,[76,77] and others.[78–80]

In a recent study, we investigated the performance of several methods [MP2, SCS(MI)–MP2, MP2.5, DFT–SAPT, DFT-D, and M06–2X] in describing potential energy curves of complexes containing the principle binding motifs in biomolecules (H-bonding, stacking, X–H$\cdots\pi$, and dispersion); benchmark data were generated using the estimated CCSD(T)/CBS technique.[57] Quite surprisingly, it was found that each of these methods is capable of describing curves for all of the interaction types (at least) at a qualitative level; this represents a great achievement for the two DFT-based methods. However, only two of the methods, DFT–SAPT and MP2.5, can be said to give consistently accurate quantitative results for all interaction types. It is also notable that the SCS(MI)–MP2 method is largely successful in its goal of improving MP2 results, giving more accurate curves for H-bonding and stacking complexes. This method, however, does very poorly for the propane dimer (the only example of a dispersion-bound complex), although it should be noted that all methods, except DFT–SAPT and MP2.5, give poor results for this complex. In the cases of SCS(MI)–MP2 and DFT-D, the poor performance for this aliphatic system may be attributable to the fact that very few aliphatic (sp^3 hybridized) carbons are contained in the S22 test set. Among the DFT-based methods, DFT-D generally produces the best performance, with M06–2X often giving potential energy curves that are of the wrong overall shape.

Intramolecular interactions are noncovalent interactions that occur between two chemical moieties located within the same chemical structure. These types of interactions are ubiquitous in biomolecular systems and are responsible for the structure, stability, and dynamics of proteins and DNA (among other biosystems). One large issue associated with the characterization of intramolecular interactions is the proper description of what has come to be known as the intramolecular BSSE. It was long believed that there could be no BSSE within an individual molecule, but it has been shown in the past years that the relative (conformational) energies of isolated systems containing intramolecular interactions can change dramatically upon extension of the basis set, indicating BSSE effects. For supermolecular complexes, BSSE can be effectively eliminated (or reduced) by using the CP corrections; this is not the case for isolated systems, for which CP corrections cannot be used. Among the most effective methods for dealing with BSSE in these types of systems is the use of local (L) methods, such as LMP2 and LCCSD(T), which converge more quickly with basis set size than conventional methods, and the use of DFT-D methods, which are often

TABLE 1 | Root Mean Square Deviations and Mean Signed Errors for Binding Energies Found Within the S22 Test Set As Computed Using Several Wavefunction and Density Functional Methods

Method	Basis Set	CP correction[1]	Ref.	S22 RMSD	S22 MSE	H-bond MSE	Disp. MSE	Mixed MSE
MP2	CBS (at-aq)[2]	Yes	22	0.94	−0.80	−0.21	−1.50	−0.58
MP2	cc-pVTZ	No	34	1.10	−1.36	−1.05	−1.92	−1.03
MP2	cc-pVTZ	Yes	34	0.85	0.30	1.20	−0.24	0.03
SCS(MI)–MP2	cc-pVTZ	Yes	23	0.30	0.03	0.05	0.11	−0.07
MP3	MP2(CBS) + Δ^3	Yes	13	0.67	0.46	0.00	1.02	0.27
MP2.5	MP2(CBS) + Δ^3	Yes	21	0.22	−0.16	−0.04	−0.25	−0.19
CCSD	MP2(CBS) + Δ^3	Yes	13	0.42	0.63	0.59	0.88	0.38
SCS(MI)–CCSD	MP2(CBS) + Δ^3	Yes	13	0.17	0.00	0.03	0.04	−0.07
DFT–SAPT	CBS(ad-at)[4]	No	13	0.49	0.37	0.78	0.22	0.14
TPSS	LP[5]	No	34	2.88	3.02	1.43	5.17	2.15
TPSS-D	LP[5]	No	34	0.38	−0.14	−0.37	−0.07	0.02
PM6		No	46	2.51	3.38	4.86	3.26	2.04
PM6-Da		No	46	2.21	1.44	3.83	0.02	0.68
PM6–DH2		No	46	0.53	0.12	−0.04	0.01	0.42

[1] Indicates use of counterpoise corrections for BSSE.
[2] Here, correction terms using the indicated method are added to the MP2/CBS result.
[3] Extrapolation to the CBS using the aug-cc-pVTZ and aug-cc-pVQZ basis sets.
[4] Extrapolation to the CBS using the aug-cc-pVDZ and aug-cc-pVTZ basis sets.
[5] LP = 6–311++G(3df,3pd).
MP2, second-order Møller–Plesset perturbation theory; CCSD, coupled-cluster with single and double and perturbative triple excitations; DFT–SAPT, density functional theory–symmetry-adapted perturbation theory ; PM6, parameterized model 6.

parameterized to account for BSSE effects without CP corrections.

Because of their relatively small size and 'floppy' nature, small peptides are among the most interesting biosystems (containing intramolecular interactions) to study. There have been many studies on peptide structure and dynamics published within the past 5 or 6 years, with contributions coming from Hobza and coworkers,[81–86] van Mourik and coworkers,[87–91] and others.[37,92–94]

In a thorough investigation, Hobza and coworkers[85] used a large number of computational methods, as well as infrared spectroscopic techniques, to characterize the conformational stabilities of the glycine–phenylalanine–alanine peptide. Here, peptide conformations were generated using MD/quenching techniques (based on both semiempirical and EP methods) and relative energies (for low-lying conformers) were evaluated using CCSD(T)/CBS, DFT-D, and M06–2X. Conformational analyses were also carried out using semiempirical and EP-based metadynamics. It was found that DFT-D gives a conformational energy ordering similar to that of CCSD(T)/CBS, whereas M06–2X does not give such good agreement. It was also found that the EP-based simulations did not generate the same near-minimum energy structures as the semiempirical MD methods.

RESULTS AND DISCUSSION

It should be clear from the above discussion that the characterization of noncovalent interactions depends heavily on the method being used. Here, we discuss the performance of several methods, some of them are popular methods that have been used for many years and some of them are newly developed methods that promise to provide improved descriptions of noncovalent interactions. Table 1 summarizes the results obtained for the S22 data set using several of these methods. Figures 1, 2, and 3 give potential energy curves for the benzene···water complex, benzene···cytosine complex, and propane dimer, respectively.

Energy and Geometry

(i) Approximations to the highly accurate CCSD(T)/CBS results can be obtained from MP2/CBS interaction energies (extrapolated from aug-cc-pVDZ and aug-cc-pVTZ, or similar, basis sets) and CCSD(T) correction terms determined with basis sets of DZ quality. These energies will generally differ by less than 5% from the accurate CCSD(T)/CBS values. More accurate values are obtained by extrapolating the MP2 energies from aug-cc-pVTZ and aug-cc-pVQZ calculations

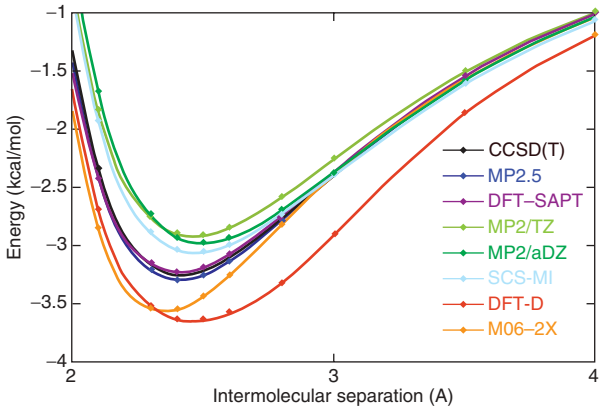

FIGURE 1 | Potential energy curves for the benzene · · · water complex using several different methods. CCSD(T) = MP2/CBS + ΔCCSD(T)/aDZ, MP2.5 = MP2/CBS + ΔMP2.5/aDZ, DFT–SAPT = DFT–SAPT/aTZ, SCS-MI = SCS(MI)–MP2/cc-pVTZ, DFT-D = DFT-D/TPSS/LP, M06–2X = DFT/M06–2X/6–311+G(2df,2pd).

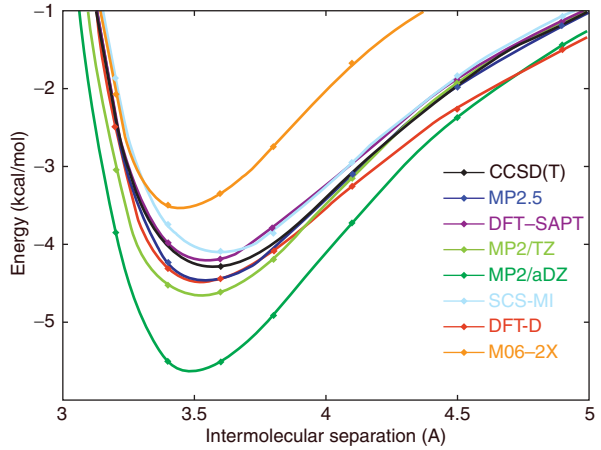

FIGURE 2 | Potential energy curves for the benzene · · · cytosine complex using several different methods.

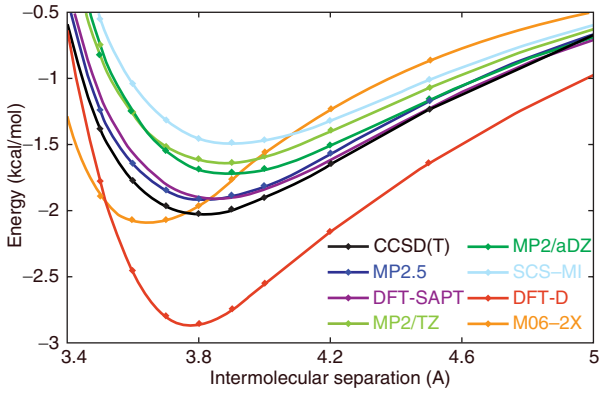

FIGURE 3 | Potential energy curves for the propane dimer using several different methods.

and/or by determining the CCSD(T) correction term at (at least) the aug-cc-pVDZ level.

(ii) DFT–SAPT interaction energies are very close to the benchmark energies, provided that an aug-cc-pVTZ (or larger) basis set is used. In the case of the smaller aug-cc-pVDZ basis set, all energy terms are close to the accurate ones with the exception of dispersion term, which is underestimated by about 10–20%.

(iii) MP2/CBS interaction energies are highly accurate for the H-bonded complexes but strongly overestimate stacked and dispersion-bound complexes. Improved MP2 results can be obtained by using small/medium basis sets such as aug-cc-pVDZ and cc-pVTZ. In the case of SCS–MP2, the situation is reversed from that of MP2/CBS—highly accurate interaction energies result for stacked and dispersion-bound complexes while H-bonding energies are underestimated. Among the other SCS methods, the SCS(MI)–MP2 technique seems to be the best balanced and provide reliable interaction energies for all noncovalent motifs at a reasonable cost. More expensive procedures, such as MP2.5 and SCS–CCSD, provide balanced and accurate descriptions of noncovalent interactions with smaller deviations from benchmark data than any method based on MP2. Generally, all WFT methods should employ the CP correction, and gradient optimizations should be carried out on the CP-corrected geometry hypersurface.

(iv) All standard (pure) DFT functionals exhibit similar deficiencies and fail to describe stacked and dispersion-bound complexes, with stabilization energies that are strongly underestimated. The description of H-bonded complexes is better, but also stabilizations are underestimated here. When an empirical dispersion correction is added to a DFT functional, the overall performance improves in all cases, this is true not only for stacked and dispersion-bound complexes but also for H-bonded ones. In many cases, such as in the cases of TPSS-D/LP and B97-D, the overall accuracy of the DFT-D method is better than the accuracy of the much more expensive MP2/(extended basis set) method. These functionals are currently probably the most recommendable DFT-based methods

for noncovalent complexes. An important point concerns the fact that most of the DFT-D schemes have been parameterized without employing the CP corrections, and should be used this way. This is particularly important in gradient geometry optimization because it makes it considerably easier (than in case when CP-corrected gradient optimization should be adopted). These methods also allow for easy treatment of systems exhibiting intramolecular interactions. Zhao and Truhlar's[36,37,48,74,75] M06–2X provides relatively accurate descriptions of different types of noncovalent interactions and is probably the best choice among the pure DFT functionals to date. A strong point of this functional is its good performance not only for noncovalent complexes but also for subsystems characteristics and, maybe surprisingly, even for electronically excited states. The local variant of the method, M06-L, provides data of lower accuracy but is much faster and can be thus used for larger systems. The advantage of truly nonlocal density functionals is that they can recover the correct physical description of the dispersion energy, but these methods are still far from being routine applications such as the DFT technique discussed above.

(v) All semiempirical QM methods provide very poor results for noncovalent complexes. This is especially true for stacked and dispersion-bound complexes, but their performance for H-bonded complexes is rather poor as well. Dramatic improvement in their overall performance results when dispersion corrections are included, which is mainly relevant for the stacked and dispersion-bound complexes. To improve the description of H-bonded complexes, a second correction, treating the electrostatic energy, should also be included. The joint dispersion plus H-bonding correction improved the performance of each of the parent semiempirical methods substantially. The best performance was found for the PM6–DH2 method, which produces a well-balanced treatment of different noncovalent interaction motifs. The performance of this method is comparable to that of considerably more expensive DFT methods. Because of its linear scaling, the method is extremely fast and can be used for extended complexes having several thousand of atoms. PM6–DH2 also yields (contrary to other semiempirical methods) reliable geometries. Consequently, it can be used not only in geometry optimization but also (and this is important) in *ab initio* on-the-fly MD simulations.

(vi) Surprisingly, empirical force fields provide better descriptions of stacked and dispersion-bound complexes than of H-bonded ones. Practically, all empirical methods underestimate H-bonding. The Cornell et al.[95] EP provides a balanced description of H-bonded, stacked, and mixed complexes not only at the equilibrium distance but also at larger distances.[96,97] Apart from reliable total interaction energies, the potential also provides reliable values of the dispersion energy well comparable those of DFT–SAPT. This is surprising in the light of the fact that dispersion and repulsion energies (vdW term) were parameterized simultaneously.

On the Role of Dispersion Energy

This section shows an example of why it is important to calculate the noncovalent interactions in biomolecules accurately and also offers a conclusion that can be obtained from these results. CCSD(T)/CBS calculations performed on DNA base pairs indicated substantial stabilization not only of H-bonded structures but also especially of stacked ones. Guanine$\cdots$cytosine and adenine$\cdots$thymine stacked pairs, not containing any H-bond, are characterized by stabilization energies of about 17 and 12 kcal/mol, respectively.[6] These stabilization energies are still smaller than the respective H-bonding (28 and 15 kcal/mol) energies, but stacking in DNA is more frequent than H-bonding and is also less screened by the aqueous environment. Several empirical models have been used to estimate the stability of DNA from its sequence and these studies, as well as a recent quantum chemical study, have surprisingly shown that stacking contributes more to the stability of DNA than H-bonding.[98] The role of stacking (or, in other words, of the dispersion energy) can also be shown using MD simulations with the energy function modified in such a way that the dispersion energy is neglected.[99] This is justified because the DFT–SAPT calculations have shown that dispersion energy is clearly the dominant energy term. Performing the MD simulations on the DNA dodecamer with such an EP leads to a rather dramatic change of the double-helical structure (Figure 4). The folded structure is almost fully transformed to the ladder-like structure characterized by considerably larger

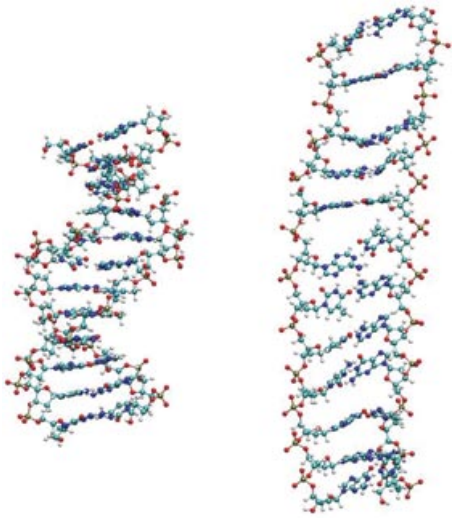

FIGURE 4 | DNA dodecamers from molecular dynamics (MD) simulations with (left) and without (right) dispersion contributions.

distances between nearly planar H-bonded pairs (from the original value of about 3.6 Å to more than 7 Å). It can be concluded that dispersion forces are essential for the structure, and full biological functionality of DNA and loss of double-helical structure upon neglecting the dispersion have fundamental biological consequences.

The situation in proteins is similar. The stabilization energies inside the hydrophobic core of rubredoxin were shown to be substantial and, similarly as in the case of the stacking of DNA bases, mostly determined by dispersion. Performing the MD simulations with a modified energy term (the dispersion energy was neglected) resulted again in a deep structural change, and the folded structure of the protein was deteriorated.[100] These findings indicate that the dispersion energy plays an important role not only in determining structure but also as a crucial driving force during the folding process.

In the previous section, we have shown the importance of dispersion energy in biomacromolecules. We considered systems with atomic structure where the dispersion energy is proportional to the sixth power of the reciprocal interatomic distance. Passing, however, to macroscopic systems, this dependence is no longer valid and the dispersion energy decays much more slowly.[101,102] For example, for the interaction of semi-infinite parallel flat slabs, the change in the free energy (which is dominated by the dispersion energy) decays as the square of the distance between these slabs.[102] The distance is measured from the edge of the sphere or cylinder. Evidently, in these cases, the dispersion energy becomes much more important. It must be mentioned, however, that for systems of this size, no reliable numerical calculations allowing for the determination of individual energy components are possible.

RECOMMENDATION AND OUTLOOK

(i) The most efficient way to evaluate noncovalent interactions (either of intra- or intermolecular origin) in biochemistry is the use of the DFT method augmented by a dispersion energy term. The method can be used both for determination of energies as well as geometries. One should never use standard DFT, which does not cover the dispersion energy; the results can be misleading.

(ii) Biochemical systems require the use of dynamics techniques (mostly via MD simulations) and up to now EPs are most widely used. Because they do not describe quantum effects, their use is limited and on-the-fly *ab initio* MD simulations will be increasingly applied. As a candidate for an energy function, the semiempirical QM methods covering the dispersion energy (tight-binding SCC-DF-TB-D or PM6–DH2) can be recommended.

(iii) If accurate energies or geometries are required, use the following techniques with the rough limitations given in parentheses: CCSD(T)/CBS (35 atoms), SCS–CCSD (50 atoms), MP2.5 (80 atoms), and SCS(MI)–MP2 (150 atoms). Generally, one should use as large a basis set as possible; the smallest reliable basis set is aug-cc-pVDZ.

(iv) Energy decomposition allows one to investigate the nature of stabilization of various biochemical systems and processes and here the DFT–SAPT/aug-cc-pVDZ technique can be recommended.

ACKNOWLEDGEMENTS

This work was a part of the 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 MSM 6198959216 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 Korea Science and Engineering Foundation (World Class University program: R32–2008-000–10180-0). K.R. gratefully acknowledges the support of the NSF EPSCORE program (EPS-0701525).

REFERENCES

1. Bakhmutov V. *Dihydrogen Bond: Principles, Experiments and Applications*. New York: Wiley-Interscience; 2008.
2. Liu Q, Hoffmann R. Theoretical aspects of a novel mode of hydrogen-hydrogen bonding. *J Am Chem Soc* 1995, 117:10108.
3. Auffinger P, Hays FA, Westhof E, Ho PS. Halogen bonds in biological molecules. *Proc Natl Acad Sci USA* 2004, 101:16789–16794.
4. Politzer P, Lane P, Concha MC, Ma YG, Murray JS. An overview of halogen bonding. *J Mol Model* 2007, 13:305–311.
5. Riley KE, Hobza P. Investigations into the nature of halogen bonding including symmetry adapted perturbation theory analyses. *J Chem Theory Comput* 2008, 4:232–242.
6. Jurecka P, Hobza P. True stabilization energies for the optimal planar hydrogen-bonded and stacked structures of guanine...cytosine, adenine...thymine, and their 9-and 1-methyl derivatives: Complete basis set calculations at the MP2 and CCSD(T) levels and comparison with experiment. *J Am Chem Soc* 2003, 125:15608.
7. Vondrasek J, Bendova L, Klusak V, Hobza P. Unexpectedly strong energy stabilization inside the hydrophobic core of small protein rubredoxin mediated by aromatic residues: Correlated ab initio quantum chemical calculations. *J Am Chem Soc* 2005, 127:2615–2619.
8. Scheiner S, Kar T. Effect of solvent upon CH...O hydrogen bonds with implications for protein folding. *J Phys Chem B* 2005, 109:3681–3689.
9. Berka K, Laskowski R, Riley KE, Hobza P, Vondrasek J. Representative Amino Acid Side Chain Interactions in Proteins. A Comparison of Highly Accurate Correlated ab Initio Quantum Chemical and Empirical Potential Procedures. *J Chem Theory Comput* 2009, 5:982–992.
10. Riley KE, Merz KM. Role of solvation in the energy stabilization inside the hydrophobic core of the protein rubredoxin. *J Phys Chem B* 2006, 110:15650–15653.
11. Riley KE, Vondrášek J, Hobza P. Performance of the DFT-D method, paired with the PCM implicit solvation model, for the computation of interaction energies of solvated complexes of biological interest. *Phys Chem Chem Phys* 2007, 9:5555–5560.
12. Hobza P, Müller-Dethlefs K. *Noncovalent Interactions. Theory and Experiment*. Cambridge: The Royal Society of Chemistry; 2009.
13. Riley KE, Jurecka P, Pitonak M, Hobza P. Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories. *Chem Rev*. 2010, 110:5023.
14. Schutz M, Manby FR. Linear scaling local coupled cluster theory with density fitting. Part I: 4-external integrals. *Phys Chem Chem Phys* 2003, 5:3349–3358.
15. Neese F, Hansen A, Wennmohs F, Grimme S. Accurate Theoretical Chemistry with Coupled Pair Models. *Acc Chem Res* 2009, 42:641–648.
16. Boys SF, Bernardi F, Calculation of Small Molecular Interactions by Differences of Separate Total Energies – Some Procedures with Reduced Errors. *Mol Phys* 1970, 19:553.
17. Riley KE, Hobza P. Assessment of the MP2 method, along with several basis sets, for the computation of interaction energies of biologically relevant hydrogen bonded and dispersion bound complexes. *J Phys Chem A* 2007, 111:8257–8263.
18. Cybulski SM, Lytle ML. The origin of deficiency of the supermolecule second-order Moller-Plesset approach for evaluating interaction energies. *J Chem Phys* 2007, 127:141102.
19. Grimme S. Improved second-order Moller-Plesset perturbation theory by separate scaling of parallel- and antiparallel-spin pair correlation energies. *J Chem Phys* 2003, 118:9095–9102.
20. Takatani T, Hohenstein EG, Sherrill CD. Improvement of the coupled-cluster singles and doubles method via scaling same- and opposite-spin components of the double excitation correlation energy. *J Chem Phys* 2008, 128:129901.
21. Pitonak M, Neogrady P, Cerny J, Grimme S, Hobza P. Scaled MP3 Non-Covalent Interaction Energies Agree Closely with Accurate CCSD(T) Benchmark Data. *Chemphyschem* 2009, 10:282–289.
22. Antony J, Grimme S. Is spin-component scaled second-order moller-plesset perturbation theory an appropriate method for the study of noncovalent

interactions in molecules? *J Phys Chem A* 2007, 111:4862–4868.

23. Distasio RA, Head-Gordon M. Optimized spin-component scaled second-order Moller-Plesset perturbation theory for intermolecular interaction energies. *Mol Phys* 2007, 105:1073–1083.
24. Hill JG, Platts JA. Spin-component scaling methods for weak and stacking interactions. *J Chem Theor Comput* 2007, 3:80–85.
25. Jung YS, Lochan RC, Dutoi AD, Head-Gordon M. Scaled opposite-spin second order Moller-Plesset correlation energy: An economical electronic structure method. *J Cheml Phys* 2004, 121:9793–9802.
26. Lochan RC, Jung Y, Head-Gordon M. Scaled opposite spin second order Moller-Plesset theory with improved physical description of long-range dispersion interactions. *J Phys Chem A* 2005, 109:7598–7605.
27. Marchetti O, Werner HJ. Accurate calculations of intermolecular interaction energies using explicitly correlated wave functions. *Phys Chem Chem Phys* 2008, 10:3400–3409.
28. Kristyan S, Pulay P. Can (semi)local density-functional theory account for the London dispersion forces. *Chem Phys Lett* 1994, 229:175.
29. Perez-Jorda JM, Becke AD. A density functional study of van-der-Waals forces – rare-gas diatomics. *Chem Phys Lett* 1995, 233:134.
30. Hobza P, Sponer J, Reschel T. Density-functional theory and molecular clusters. *J Comput Chem* 1995, 16:1315.
31. Hepburn J, Scoles G, Penco R. A simple but reliable method for the prediction of intermolecular potentials. *Chem Phys Lett* 1975, 36:451.
32. Elstner M, Hobza P, Frauenheim T, Suhai S, Kaziras E. Hydrogen bonding and stacking intercalations of nucleic acid base pairs: A density-functional-theory based treatment. *J Chem Phys* 2001, 114:5149.
33. Grimme S. Accurate description of Van Der Waals complexes by density functional theory including empirical corrections. *J Comput Chemy* 2004, 25:1463–1473.
34. Jurecka P, Cerny J, Hobza P, Salahub DR. Density functional theory augmented with an empirical dispersion term. Interaction energies and geometries of 80 noncovalent complexes compared with ab initio quantum mechanics calculations. *J Comput Chem* 2007, 28:555–569.
35. Civalleri B, Z-W CM, Valenzano L, Ugliengo P. B3LYP augmented with an empirical dispersion term (B3LYP-D*) as applied to molecular crystals. *Cryst Eng Comm* 2008, 10:405.
36. Zhao Y, Truhlar DG. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor Chem Acc* 2008, 120:215–241.
37. Zhao Y, Truhlar DG. Exploring the Limit of Accuracy of the Global Hybrid Meta Density Functional for Main-Group Thermochemistry, Kinetics, and Noncovalent Interactions. *J Chem Theor Comput* 2008, 4:1849–1868.
38. Misquitta AJ, Szalewicz K. Intermolecular forces fom asymptotically corrected density functional description of monomers. *Chem Phys Lett* 2002, 357:301.
39. Patkowski K, Szalewicz K. Frozen core and effective core potentials in symmetry-adapted perturbation theory. *J Chem Phys* 2007, 127:164103.
40. Podeszwa R, Bukowski R, Szalewicz K. Potential Energy Surface for the Benzene Dimer and Perturbational Analysis of π-π Interactions. *J Phys Chem A* 2006, 110:10345–10354.
41. Podeszwa R, Szalewicz K. Three-body symmetry-adapted perturbation theory based on Kohn-Sham description of the monomers. *J Chem Phys* 2007, 126:194101.
42. Jeziorski B, Moszynski R, Szalewicz K. Perturbation-theory approach to intermolecular potential energy surfaces of Van Der Waals complexes. *Chem Rev* 1994, 94:1887.
43. Jansen G, Hesselmann A. Comment on "Using Kohn-Sham orbitals in symmetry adapted perturbation theory to investigate intermolecular interactions". *J Phys Chem A* 2001, 105:646.
44. Stewart JJP. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements. *J Mol Model* 2007, 13:1173.
45. Stewart JJP. Application of the PM6 method to modeling the solid state. *J Mol Model* 2008, 14:499.
46. Korth M, Pitonak M, Rezac J, Hobza P. *J Chem Theory Comput* 2009, 6:344–352.
47. Rezac J, Fanfrlik J, Salahub DR, Hobza P. Semiempirical Quantum Chemical PM6 Method Augmented by Dispersion and H-Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes. *J Chem Theory Comput* 2009, 5:1749–1760.
48. Zhao Y, Truhlar DG. Benchmark databases for nonbonded interactions and their use to test density functional theory. *J Chem Theory Comput* 2005, 1:415–432.
49. Jurecka P, Sponer J, Cerny J, Hobza P. Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. *Physl Chem Chem Phys* 2006, 8:1985–1993.
50. Marchetti O, Werner H-J. Accurate calculations of intermolecular interaction energies using explicitly correlated coupled cluster wave functions and a dispersion-weighted MP2 method. *J Phys Chem A* 2009, 113:11580.

51. Pitonak M, Janowski T, Neogrady P, Pulay P, Hobza P. Convergence of the CCSD(T) Correction Term for the Stacked Complex Methyl Adenine-Methyl Thymine: Comparison with Lower-Cost Alternatives. *J Chem Theory Comput* 2009, 5:1761–1766.
52. Pitonak M, Riley KE, Neogrady P, Hobza P. Highly accurate CCSD(T) and DFT-SAPT stabilization energies of H-bonded and stacked structures of the uracil dimer. *Chemphyschem* 2008, 9:1636–1644.
53. Sherrill CD, Takatani T, Hohenstein EG. An Assessment of Theoretical Methods for Nonbonded Interactions: Comparison to Complete Basis Set Limit Coupled-Cluster Potential Energy Curves for the Benzene Dimer, the Methane Dimer, Benzene-Methane, and Benzene-H2S. *J Phys Chem A* 2009, 113:10146–10159.
54. Boese AD, Martin JML, Klopper W. Basis Set Limit Coupled Cluster Study of H-Bonded Systems and Assessment of More Approximate Methods. *J Phys Chem A* 2007, 111:11122–11133.
55. Cerny J, Pitonak M, Riley KE, Hobza P. *J Chem Theory and Comput*. Submitted for publication.
56. Morgado CA, Jurecka P, Svozil D, Hobza P, Sponer J. Balance of Attraction and Repulsion in Nucleic-Acid Base Stacking: CCSD(T)/Complete-Basis-Set-Limit Calculations on Uracil Dimer and a Comparison with the Force-Field Description. *J Chem Theory Comput* 2009, 5:1524–1544.
57. Riley KE, Pitonak M, Cerny J, Hobza P. On the Structure and Geometry of Biomolecular Binding Motifs (H-bonding, stacking, X-H...π): WFT and DFT Calculations. *J Chem Theory Comput*. Accepted for publication.
58. Ran J, Hobza P. On the Nature of Bonding in Lone Pair...π-Electron Complexes: CCSD(T)/Complete Basis Set Limit Calculations. *J Chem Theory Comput* 2009, 5:1180–1185.
59. Arnstein SA, Sherrill CD. Substituent effects in parallel-displaced pi-pi interactions. *Phys Chem Chem Phys* 2008, 10:2646–2655.
60. Hohenstein EG, Sherrill CD. Effects of Heteroatoms on Aromatic pi-pi Interactions: Benzene-Pyridine and Pyridine Dimer. *J Phys Chem A* 2009, 113:878–886.
61. Ringer AL, Figgs MS, Sinnokrot MO, Sherrill CD. Aliphatic C-H/pi interactions: Methane-benzene, methane-phenol, and methane-indole complexes. *J Phys Chem A* 2006, 110:10822–10828.
62. Ringer AL, Sinnokrot MO, Lively RP, Sherrill CD. The effect of multiple substituents on sandwich and T-Shaped pi-pi interactions. *Chem-a Eur J* 2006, 12:3821–3828.
63. Sherrill CD, Takatani T, Hohenstein EG. An Assessment of Theoretical Methods for Nonbonded Interaction: Comparison to Complete Basis Set Limit Coupled Cluster Potential Energy Curves for the Benzene Dimer, the Methane Dimer, Benzene-Methane, and benzene H2S. *J Phys Chem A* 2009, 113:10146–10159.
64. Sinnokrot MO, Sherrill CD. Substituent effects in pi-pi interactions: Sandwich and T-shaped configurations. *J Am Chem Soc* 2004, 126:7690–7697.
65. Sinnokrot MO, Sherrill CD. Highly accurate coupled cluster potential energy curves for the benzene dimer: Sandwich, T-shaped, and parallel-displaced configurations. *J Phys Chem A* 2004, 108:10200–10207.
66. Sinnokrot MO, Sherrill CD. High-accuracy quantum mechanical studies of pi-pi interactions in benzene dimers. *J Phys Chem A* 2006, 110:10656–10668.
67. Sinnokrot MO, Valeev EF, Sherrill CD. Estimates of the ab initio limit for pi-pi interactions: The benzene dimer. *J Am Chem Soc* 2002, 124:10887–10893.
68. Takatani T, Sherrill CD. Performance of spin-component-scaled Moller-Plesset theory (SCS-MP2) for potential energy curves of noncovalent interactions. *Phys Chem Chem Phys* 2007, 9:6106–6114.
69. Tauer TP, Derrick ME, Sherrill CD. Estimates of the ab initio limit for sulfur-pi interactions: The H2S-benzene dimer. *J Phys Chem A* 2005, 109:191–196.
70. Tsuzuki S, Honda K, Uchimaru T, Mikami M. Estimated MP2 and CCSD(T) interaction energies of n-alkane dimers at the basis set limit: Comparison of the methods of Helgaker et al. and Feller. *J Chem Phys* 2006, 124:7.
71. Tsuzuki S, Uchimaru T. Magnitude and physical origin of intermolecular interactions of aromatic molecules: Recent progress of computational studies. *Curr Org Chem* 2006, 10:745–762.
72. Tsuzuki S, Uchimaru T, Mikami M, Tanabe K. High level ab initio calculations of intermolecular interaction of propane dimer: Orientation dependence of interaction energy. *J Phys Chem A* 2002, 106:3867–3872.
73. Tsuzuki S, Honda K, Uchimaru T, Mikami M. Intermolecular Interactions of Nitrobenzene-Benzene Complex and Nitrobenzene Dimer: Significant Stabilization of Slipped-Parallel Orientation by Dispersion Interaction. *J Chem Phys* 2006, 125:AN 124304.
74. Zhao Y, Truhlar DG. Density functionals for noncovalent interaction energies of biological importance. *J Chem Theory Comput* 2007, 3:289–300.
75. Zhao Y, Truhlar DG. Density functionals with broad applicability in chemistry. *Acc Chem Res* 2008, 41:157–167.
76. Aeberhard PC, Arey JS, Lin IC, Rothlisberger U. Accurate DFT Descriptions for Weak Interactions of Molecules Containing Sulfur. *J Chem Theory Comput* 2009, 5:23.
77. Tapavicza E, Lin IC, von Lilienfeld OA, Tavernelli I, Coutinho-Neto MD, Rothlisberger U. Weakly bonded complexes of aliphatic and aromatic carbon compounds described with dispersion

corrected density functional theory. *J Chem Theory Comput* 2007, 3:1673–1679.

78. Janowski T, Pulay P. High accuracy benchmark calculations on the benzene dimer potential energy surface. *Chem Phys Lett* 2007, 447:27–32.
79. Tekin A, Jansen G. How Accurate is the Density Functional Theory Combined with Symmetry Adapted Perturbation Theory Approach for CH-π and π-π Interactions? A Comparison to Supermolecular Calculations for the Acetylene-Benzene Dimer. *Phys Chem Chem Phys* 2007, 9:1680–1687.
80. Fusti Molnar L, He X, Wang B, Merz KM. Further Analysis and Comparitive Study of Intermolecular Interactions using Dimers form the S22 Database. *J Chem Phys* 2009, 131:AN 065102.
81. Cerny J, Jurecka P, Hobza P, Valdes H. Resolution of identity density functional theory augmented with an empirical dispersion term (RI-DFT-D): A promising tool for studying isolated small peptides. *J Phys Chem A* 2007, 111, 1146–1154.
82. Reha D, Valdes H, Vondrasek J, Hobza P, Abu-Riziq AG, Crews B, de Vries MS, *Chem-a Eur J* 2005, 11:6803–6817.
83. Valdes H, Pluhackova K, Hobza P. Phenylalanyl-Glycyl-Phenylalanine Tripeptide: A Model System for Aromatic-Aromatic Side Chain Interactions in Proteins. *J Chem Theory Comput* 2009, 5:2248–2256.
84. Valdés H, Řeha D, Hobza P. Structure of Isolated Tryptophyl-Glycine Dipeptide and Tryptophyl-Glycyl-Glycine Tripeptide: Ab Initio SCC-DFTB-D Molecular Dynamics Simulations and High-Level Correlated Ab Initio Quantum Chemical Calculations. *J Phys Chem B* 2006, 110:6385–6396.
85. Valdes H, Spiwok C, Rezac J, Reha D, Abu-Riziq AG, de Vries MS, Hobza P. Potential-Energy and Free-Energy Surfaces of Glycyl-Phenylalanyl-Alanine (GFA) Tripeptide: Experiment and Theory. *Chem-a Eur J* 2008, 14:4886–4898.
86. Valdes H, Vojtech K, Pitonak M, Exner O, Stary I, Hobza P, Rulisek L. Evaluation of the Intramolecular Basis Set Superposition Error in the Calculations of Larger Molecules: [n]Helicenes and Phe-Gly-Phe Tripeptide. *J Comput Chem* 2008, 29:861–870.
87. Holroyd LF, van Mourik T. Insufficient Description of Dispersion in B3LYP and Large Bases Set Superposition Errors in MP2 Calculations can Hide Peptide Conformers. *Chem Phys Lett* 2007, 442:42–46.
88. Shields AE, van Mourik T. Comparison of ab Initio and DFT Electronic Structure Methods for Peptides Containing an Aromatic Ring: Effect of Dispersion and BSSE. *J Phys Chem A* 2007, 111:13272–13277.
89. Toroz D, van Mourik T. The Structure of the Gas-Phase Tyrosine-Glycine-Glycine Tripeptide. *Mol Phys* 2007, 105:209–220.
90. van Mourik T. Assessment of Density Functionals for Intrmolecular Dispersion-Rich Interactions. *J Chem Theory Comput* 2008, 4:1610–1619.
91. van Mourik T, Karamertzanis PG, Price SL. Molecular Conformations and Relative Stabilities can be as Demanding of the Electronic Structure Method as Intermolecular Calculations. *J Phys Chem A* 2006, 110:8–12.
92. Häber T, Seefeld K, Engler G, Grimme S, Kleinermanns K. IR/UV Spectra and Quantum Chemical Calculations of Trp-Ser: Stacking Interactions between Backbone and Indole Side-chains. *Phys Chem Chem Phys* 2008, 10:2844–2851.
93. Schwabe T, Grimme S. Double-Hybrid Density Functionals with Long-Range Dispersion Corrections: Higher Accuracy and Extended Applicability. *Phys Chem Chem Phys* 2007, 9:1463–9076.
94. Doslic N, Kovacevic G, Ljubic I. Signature of the Conformational Preferences of Small Peptides: a Theoretical Investigation. *J Phys Chem A* 2007, 111:8650–8658.
95. Conell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA. *J Am Chem Soc* 1995, 117:5179.
96. Kolar M, Berka K, Jurecka P, Hobza P. On the reliability of the AMBER force field and its empirical dispersion contribution for the description of noncovalent complexes. *ChemPhysChem* 2010, 11:2399.
97. Paton RS, Goodman J. Hydrogen bonding and pi-stacking: How reliable are force fields? A critical evaluation of force field descriptions of nonbonded interactions. *J Chem Inf Model* 2009, 49:944.
98. Rezac J, Hobza P. On the nature of DNA-duplex stability. *Chem-a Eur J* 2007, 13:2983.
99. Cerny J, Kabelac P, Hobza P, Double-helical -> ladder structural transition in the B-DNA is introduced by a loss of dispersion energy. *J Am Chem Soc* 2008, 130:16055.
100. Cerny J, Vondrášek J, Hobza P. Loss of dispersion energy changes the stability and folding/unfolding equilibrium of teh Trp-Cage protein. *J Phys Chem B* 2009, 113:5657.
101. Lifshitz EM. The theory of molecular attraction forces between solid bodies. *Zh Eksp Teor Fiz* 1955, 29:94.
102. Israelachvili JN. *Intermolecular and Surface Forces*. New York: Academic Press; 1985.

