JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

Economical and Accurate Protocol for Calculating Hydrogen-Bond-Acceptor Strengths

Ahmed El Kerdawy,[†] Christofer S. Tautermann,[‡] Timothy Clark,*[,†,§,∥] and Thomas Fox[‡]

[†]Computer-Chemie-Centrum, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nägelsbachstraße 25, 91052 Erlangen, Germany

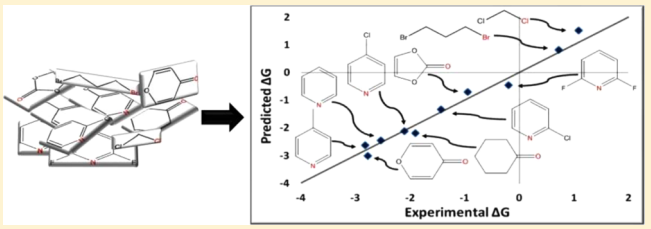
[‡]Computational Chemistry, Lead Identification and Optimization Support, Boehringer Ingelheim Pharma GmbH & Co. KG, 88397 Biberach, Germany

[§]Interdisciplinary Center for Molecular Materials, Friedrich-Alexander-Universität Erlangen-Nürnberg, Nägelsbachstraße 49, 91052 Erlangen, Germany

[∥]Centre for Molecular Design, University of Portsmouth, King Henry Building, Portsmouth PO1 2DY, United Kingdom

Supporting Information

ABSTRACT: A series of density functional/basis set combinations and second-order Møller–Plesset calculations have been used to test their ability to reproduce the trends observed experimentally for the strengths of hydrogen-bond acceptors in order to identify computationally efficient techniques for routine use in the computational drug-design process. The effects of functionals, basis sets, counterpoise corrections, and constraints on the optimized geometries were tested and analyzed, and recommendations (M06-2X/cc-pVDZ and X3LYP/cc-pVDZ with single-point counterpoise corrections or X3LYP/aug-cc-pVDZ without counterpoise) were made for suitable moderately high-throughput techniques.



INTRODUCTION

One important aspect of the drug-design process is ligand–target interaction, which is governed by noncovalent intermolecular interactions. These interactions not only dictate a specific ligand orientation during the binding scenario but also define the ligand-binding affinity. Hydrogen bonding is an important contributor to these interactions. Hence, minor changes in the ligand hydrogen-bonding motifs can have a drastic impact on the ligand-binding orientation and consequently its binding affinity.[1−3] Moreover, hydrogen bonds are not isolated interactions but are sensitive to the molecular environment; therefore, any change in the molecular constitution will affect the ligand's hydrogen-bonding capabilities.

This importance of hydrogen bonding, especially in the field of drug design, has motivated research for several decades into developing theoretical techniques for predicting hydrogen-bond strength or ranking different H-bond donors or acceptors. These techniques can be as simple as using indicator variables, such as the number of the H-bond donors and acceptors in the molecule.[4−6] The disadvantage of such methods is that they do not differentiate between the different H-bond donor and acceptor atoms and they neglect the effect of the remainder of the molecule on the hydrogen-bonding strength. Other methods are based on experimentally measured properties such as the difference between octanol–water and cyclohexane–water partition coefficients[7] or solvatochromic parameters (derived from spectroscopic data).[8−10] However, for these methods, experimental data must be available to derive the H-bond properties. On the other hand, various approaches on the basis of theoretically calculated properties such as the energies of lowest-unoccupied and highest-occupied molecular orbitals (LUMO and HOMO, respectively) and atomic charges,[11,12] self-atom polarizability and superdelocalizability,[13,14] the molecular electrostatic potential (MEP),[15,16] or combinations of these properties[17] have been proposed. However, some of these approaches treat the H-bond as an isolated interaction and do not take the influence of the molecular environment into account. They are also often restricted to a limited group of compounds, which means that they cannot be used as universal tools for predicting H-bond strength. Most of these methods also consider only the electrostatic parameters, neglecting the steric and dispersion components of the interaction.

In recent years, because of increases in computational power and the increasing confidence that computational methods are capable of accurate geometry and energy calculations for H-bonds,[18−22] attention has been directed toward using quantum mechanical (QM) calculations to predict H-bond strengths. In this context, Schwöbel et al.[23,24] used ab initio and DFT-based local molecular parameters to model and predict the H-donor and H-acceptor strengths in organic compounds. Another possibility is the supermolecular approach, which uses the

Received: October 25, 2013

calculated interaction energy of the two interacting systems as a measure for the H-bond strength[25,26] on the basis of the hypothesis that the hydrogen-bond free energy is linearly correlated to its enthalpy.[10,27,28] This approach has been shown to give a better correlation to experimental hydrogen-bonding constants than the other methods.[25] It also takes into account all of the different components of the H-bond interaction, including the steric and the dispersion components, and it considers the molecular environment, not just the isolated atoms that form the hydrogen bond.

The calculation of interaction energies requires higher levels of theory than for simple geometry optimizations because the energy is more sensitive to an accurate description than the structure itself.[29] To describe H-bonding energies accurately, high-level quantum mechanical calculations that capture a large portion of the correlation energy are required. One accurate practical method for obtaining reliable noncovalent complex geometries and energies is CCSD(T),[30,31] but its very high computational cost rules out its use as a regular procedure for calculating H-bond strengths in the drug-design process. Second-order Møller−Plesset perturbation theory (MP2)[32−37] gives interaction geometries and energy differences similar to those obtained with CCSD(T).[29,38−41] MP2 is much less computationally demanding than CCSD(T) but remains more expensive than density-functional theory (DFT) calculations; thus, DFT has become the method of choice for this purpose because of its balance of accuracy and efficiency.[38,42] It reproduces the reference experimental data and the results of the more expensive and accurate computational methods with acceptable accuracy and implicitly includes part of the dynamic correlation interaction energy contribution, which is important and not considered in the alternative Hartree−Fock (HF) method.[43] One major weakness of conventional DFT methods is that they fail to describe the dispersion component of H-bonding because nonlocal (nondynamic) correlation is not included in the local DFT energy. DFT therefore works well in cases of H-bonding interactions in which the dispersion energy component plays a minor role;[29,38,39] however, many approaches are now available that overcome this drawback. Some of these involve reparameterization of an existing DFT functional to cover the dispersion component or the addition of an empirical expression for dispersion to the DFT functional in a force-field-like manner to give the so-called DFT-D functionals.[44−48]

The most important issue of the supermolecular approach is the basis set superposition error (BSSE), which is an arithmetical error that arises when using a finite basis set to calculate the interaction energies. BSSE means that the complex system is described better than its subsystems because it can use the surplus basis functions of all of its subsystems.[29] BSSE can be decreased by extending the basis set and becomes negligible at the quadruple and quintuple zeta basis set levels if diffuse functions are used.[43] Thus, either a large basis set or a correction for BSSE should be used to overcome such problems. The most common BSSE correction is the counterpoise (CP) method proposed by Boys and Bernardi.[49] However, the use of CP is controversial. Some authors[50] have shown that the counterpoise correction systematically leads to interaction energies that differ more strongly from the real values than the uncorrected ones and that it is better to extend the basis set to minimize the BSSE rather than using CP. Others[42,51−54] have reported that a better description of H-bonding is obtained, even with moderate to low level of theory, when the geometry optimization of the complex system is carried out on a CP-corrected potential-energy surface. Yet others[55] found that there is no significant difference between the CP-corrected interaction energies and the uncorrected ones or at least that CP correction does not influence the correlation between the calculated interaction energy and the experimental hydrogen-bonding constants.[26] It has also been reported[29,38,43] that BSSE can be beneficial in some cases of medium-sized systems with double-ζ quality basis sets because it compensates the missing dispersion part in many DFT functionals, making some DFT methods capable of describing H-bonding anomalously well. Therefore, the different views on CP are the result of differences in computational methods, basis sets, and the complex systems used in the different publications, underlining the complexity of the problem.

Many studies have evaluated the ability of different DFT functional and basis set combinations to describe H-bonds.[25,26,29,38,39,42−44,51,53,54,56−60] However, many of these previous studies were carried out using specific model systems (e.g., the water dimer[42,51−54,60,61]) so that it is not obvious if the results can be generalized to sets of diverse drug-like molecules with many different functional groups. Moreover, the continuous introduction of new functionals and the influence of the basis set, CP correction, and zero-point energy (ZPE) correction complicate the appropriate selection even further.

It is important to note that all these approaches focus on a single arm of the full thermodynamic cycle for the calculation of H-bond strength necessary for the comparison with most experimental values: the one that deals with the calculation of interaction energies in vacuum so that, for instance, desolvation effects are completely ignored.

In this work, our aim is to find an appropriate level of theory that can be used as an affordable standard procedure for calculating H-bond acceptor strengths of molecules with functional groups that typically occur in drug-like molecules with acceptable accuracy. To this end, we calculate the interaction energies between a H-bond donor and various H-bond acceptors with a wide range of DFT methods and compare them to experimental values obtained from spectroscopic measurements collected in the pK_{BHX} data set.[62]

We investigate (a) what the best DFT functional/basis set combinations are for predicting H-bond strengths, (b) whether BSSE correction needs to be employed, (c) whether ZPE correction is necessary, and (d) the impact of the geometrical constraints during the calculations.

METHODS

Experimental Reference Data. As our experimental reference, we used the pK_{BHX} data set.[62] The pK_{BHX} data set reports the negative decadic logarithm of the dissociation constant (K_{BHX}) of H-bonded complexes of a series of hydrogen-bond acceptors (HBAs) with 4-fluorophenol (and/or methanol) as a reference hydrogen-bond donor (HBD) measured by Fourier transform infrared spectroscopy (FTIR). Large positive pK_{BHX} values indicate strong HBAs. This data set has several advantages over other previously published scales.[10,63−66] (a) It contains a large number of diverse compounds (~1200) and covers a large variety of HBA functional groups. (b) It gives a specific pK_{BHX} value for each HBA group in a polyfunctional HBA and not the less informative whole-molecule value.[62] This is a significant advantage for compounds with more than one HBA site, as it allows a detailed comparison with the calculated values. (c) The

dx.doi.org/10.1021/ci4006222 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

measurements were performed in CCl_4, which has no significant HBA ability and is transparent to the spectroscopic method used in the measurements. It is thus completely inert under the measurement conditions.[62] Furthermore, CCl_4 has a very low dielectric constant (ε = 2.23), making gas-phase calculations a good model for the experimental setup.

The pK_{BHX} values were converted into ΔG values in kilocalories per mole, which were then used throughout. As K (association constant) = $1/K_{BHX}$ (dissociation constant),[62] ΔG can be calculated according to the equation

$$\begin{aligned}\Delta G\ (\text{in kilocalories per mole}) &= -RT\ \ln K \\ &= -RT\ \ln 1/K_{BHX} = -2.303RT(\log 1/K_{BHX}) \\ &= -2.303RT(pK_{BHX}) = -1.363pK_{BHX} \end{aligned} \quad (1)$$

where R is the gas constant and T is the temperature, which is defined as 298 K.

Note that the stronger the HBA, the more positive the pK_{BHX} value and the more negative the ΔG value.

Selection of the DFT Functional/Basis Set Combinations. We started our investigations by assessing how well we can rank a small set of HBAs according to their H-bond strength using a common HBD.

We started with three recent publications[26,42,67] that used DFT to calculate H-bond interactions in which some DFT levels showed good agreement with either experimental data or high-level reference calculations. Nocker et al.[26] reported that the B3LYP[68−71] functional in combination with Dunning's augmented double-ζ basis set (aug-cc-pVDZ)[72−75] is able to calculate H-bond interaction energies that correlate well with experimental H-bond binding constants. They also reported that the BSSE and ZPE corrections are not necessary at this level of theory because their correlations with and without these corrections were very similar. El Kerdawy et al.[67] used the B97D[45] and ωB97xD[46] functionals, which contain empirical dispersion expressions, in combination with the aug-cc-pVDZ basis set without CP nor ZPE corrections to calculate noncovalent interaction energies. Using these energies as input for artificial neural networks (ANN), experimental interaction energies for drug-like molecules with small probe molecules such as H_2O can be reproduced. Plumley and Dannenberg[42] used the water dimer as a model system to investigate the performance of 160 different DFT functional/basis set combinations in describing H-bond energies and geometries. In addition, they employed different levels of CP correction (no CP, geometry optimization without CP followed by a single-point calculation including CP, or geometry optimization and energy calculation including CP). The calculated energies were compared to the experimental and high-level calculated water-dimer binding energy, which give a consensus value of −4.90 kcal mol^{-1}, as the reference. For this work, we selected all combinations that gave calculated interaction energies within ±0.15 kcal mol^{-1} of the reference value. However, to develop a computationally efficient protocol, we focused here only on double-ζ or very economical triple-ζ (e.g., 6-311G-based) basis sets. Likewise, we excluded Grimme's double-hybrid functionals B2PLYP[76] and B2PLYPD[77] because they incorporate MP2-like exchange-correlation contributions, resulting in significantly higher computational efforts. Furthermore, we excluded DFT functionals whose performance varied widely with the basis set used. The selected list of combinations used in our work is shown in Table 1.

Table 1. Selected DFT Functional/Basis Set Combinations from the Literature That Are Investigated in the Present Work

combination	No-CP[a]	SP-CP[b]	CP[c]
M06/cc-pVDZ			×
M06[44]/aug-cc-pVDZ	×		
M062X[44]/cc-pVDZ	×	×	
MPWB1K/6-311G(d,p)	×	×	
MPWB1K/cc-pVDZ			×
MPWB1K[78]/aug-cc-pVDZ	×		
B97D/D95++(d,p)	×	×	
B97D/6-311G(d,p)		×	×
B97D/6-311++G(d,p)	×	×	×
B97D/aug-cc-pVDZ	×	×	
ωB97xD/aug-cc-pVDZ	×	×	
B3LYP/6-311++G(d,p)	×	×	
B3LYP/aug-cc-pVDZ	×	×	
X3LYP[47]/aug-cc-pVDZ	×	×	×

[a]No-CP, calculations without CP correction. [b]SP-CP, geometrical optimization without CP correction followed by CP corrected single-point energy calculation. [c]CP, calculations with CP correction.

Computational Details. Methanol was selected as the common HBD to be used in the DFT calculations rather than 4-fluorophenol because its smaller size makes the procedures computationally cheaper. The 3D structures of methanol and the HBAs were generated using CORINA[79] followed by force-field optimization using the MMFF94x force field[80−82] implemented in MOE.[83] Initial tests showed that this procedure yields a consistent way of setting up the initial monomer geometries and that the initial geometry has no significant impact on the calculated complex interaction energies. The resulting 3D structures of the individual monomers were optimized with the chosen DFT functional/basis set combination. Finally, the H-bonded complexes of the optimized HBAs and the optimized methanol were constructed. During the DFT geometry optimization of the complexes, the methanol structure was kept frozen, and the angle between the methanol OH and the H-bond acceptor atom of the HBA was constrained to 180° (Figure 1). This constraint was used to enhance the geometry optimization convergence and to ensure consistent geometries for all complexes by avoiding additional interactions between methanol and the HBA. The restraint value was chosen to place the methanol OH hydrogen atom in the optimal position for H-bonding.[67,84]

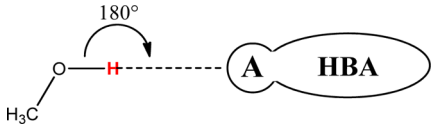


Figure 1. Schematic drawing of the geometrical constraint applied during the optimization of the methanol-HBA complex. A, H-bond acceptor atom; HBA, H-bond acceptor.

All of the calculations were performed using the Gaussian09 program[85] with default settings unless otherwise stated.

The electronic interaction energy, ΔE, was calculated as the difference between the total energy of the geometrically

dx.doi.org/10.1021/ci4006222 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

optimized complex, E_{AD}, and the energies of the individual components, E_A and E_D. If BSSE and ZPE are taken into account, then they are added to the equation.

$$\Delta E_{(\text{No-CP})} = E_{AD} - (E_A + E_D)\ [+\text{ZPE}] \quad (2)$$

$$\Delta E_{(\text{CP})} = E_{AD} - (E_A + E_D) + \text{BSSE}\ [+\text{ZPE}] \quad (3)$$

Performance Assessment. The following criteria were used as measures for the ability of the DFT functional/basis set combinations to reproduce the experimental trends in HBA strength:

(a) The correlation coefficient (R^2) between the calculated interaction energy, ΔE, and the ΔG derived from the experimental HBAs' pK_{BHX} values.

(b) The standard error of estimate (SEE) for ΔG, which measures how well the data points fit to the linear interpolation and thus indicates the reliability of the estimation of ΔG. The smaller the SEE value, the better the fit to the linear interpolation. The SEE was calculated as

$$\text{SEE} = \sqrt{\frac{\sum_{i=1}^{n} (\Delta G_i - \Delta G_i^{\text{estimated}})^2}{n-2}} \quad (4)$$

where n is the number of tested complexes and $\Delta G_i^{\text{estimated}}$ is the estimated ΔG for complex (i) based on the least-squares line equation, which describes the correlation between the calculated ΔE and the experimental ΔG.

(c) The slope (S) and intercept (I) of the least-squares (correlation) line describing the correlation between the calculated ΔE and the experimental ΔG.

RESULTS AND DISCUSSION

The goal of this work is to define a routine protocol for predicting and ranking the H-bond strengths of different drug-like HBAs. Therefore, the following four-step plan was followed:

(1) Select the most promising DFT functional/basis set combinations on the basis of the results of a small data set.

(2) Apply these promising DFT functional/basis set combinations to a larger set of 58 HBAs.

(3) Assess the effect of applying additional geometrical constraints.

(4) Validate the proposed protocol using an independent test set of HBAs.

Preliminary Evaluation of the DFT Functional/Basis Set Combinations. Using a small set of eight HBAs (Figure 2), we evaluated the selected DFT functional/basis set combinations from Table 1 to identify combinations with low performance or those that are incompatible with the calculation setup and to select those to be investigated further on a larger scale.

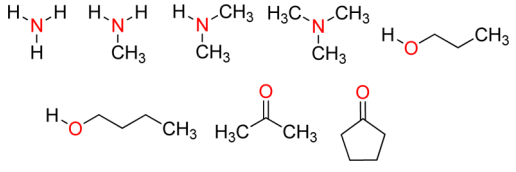


Figure 2. Selected compounds used for a preliminary evaluation of the DFT functional/basis set combinations (hydrogen-bond-acceptor atom is in red).

Table 2 shows the performance of the different DFT functional/basis set combinations. For the MPWB1K functional, we could not obtain a converged SCF for several molecules using Gaussian's default settings. Various alternatives to improve SCF convergence were considered unsuitable for a routine approach, so we excluded this functional from further consideration because accelerated convergence protocols are difficult to standardize or are computationally expensive.

Although the small data set is not very diverse and does not represent the diversity of the original data set fully and also does not allow a statistically sound selection of the best combinations, it is adequate to choose the initial set of combinations that perform well for this small data set. It was intended to select the combinations that are most compatible with the designed setup and to give consistent and stable calculation results. Therefore, for the following steps, we selected all combinations with an R^2 value of 0.85 or higher, as this criterion measures how consistently each combination can treat the set of different HBAs. The M06/aug-cc-pVDZ combination was excluded, despite its high R^2 value, because it proved to be extremely sensitive to the initial geometry of the complex.

Combinations that use CP and/or large basis sets are computationally expensive. Therefore, we sought to reduce the computational resources by reducing the CP level or using smaller basis sets. Thus, we tried to replace the combinations that employ full CP by the corresponding SP-CP or No-CP level. Moreover, for B97D, we replaced the costly aug-cc-pVDZ basis set by the less demanding Pople basis sets. We also decided to remove the B3LYP combinations, as the corresponding functional X3LYP is included; the X3LYP functional is only a slight variation of B3LYP with improved description of the van der Waals interactions. All of these selections are warranted because our initial HBA set showed that the reduction of the computational level comes at a very modest decrease in the performance, if at all. The selected eight combinations are shown in bold in Table 2.

Evaluation on a Larger Data Set. The eight DFT levels selected in the previous step were employed in a larger data set of 58 chemically diverse HBAs from the pK_{BHX} data set (see Supporting Information S1, Figure 1). They serve as prototypes for functional groups in drug-like molecules, and their pK_{BHX} values were obtained with methanol as the H-bond donor.

In addition to the selected eight combinations shown in Table 2, we decided to compare also the performance of post-Hartree−Fock techniques with similar CPU requirements. Thus, we added MP2/cc-pVDZ with No-CP and SP-CP to the list of methods. We also investigated the effect of the ZPE correction on the correlation between the calculated interaction energy ΔE and ΔG. Table 3 shows the performance of the different calculational levels.

Table 3 shows that the MP2 combinations performed less well than the DFT levels, although the CPU requirements were similar, indicating that the DFT levels can give reasonably accurate results within affordable computational times.

Three DFT levels gave R^2 values higher than 0.80, with an average SEE of 0.43 without ZPE and 0.40 with ZPE correction. These are M06-2X/cc-pVDZ(SP-CP), ωB97xD/aug-cc-pVDZ(SP-CP), and X3LYP/aug-cc-pVDZ(No-CP). The remaining DFT functional/basis set combinations yield

dx.doi.org/10.1021/ci4006222 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 2. Performance of DFT Functional/Basis Set Combinations Using Different CP Levels in the Preliminary Evaluation (Eight HBAs)[a]

	No-CP[b]				SP-CP[c]				CP[d]			
calculation level	R^2	SEE[e]	S^f	I^g	R^2	SEE[e]	S^f	I^g	R^2	SEE[e]	S^f	I^g
M06/cc-pVDZ									0.49	0.61	0.76	2.18
M06/aug-cc-pVDZ	0.87	0.29	0.91	3.73								
M06-2X/cc-pVDZ	0.76	0.39	2.87	25.14	**0.85**	**0.31**	**0.81**	**2.66**				
MPWB1K/6-311G(d,p)	0.77(1)	0.40	1.08	6.55	0.79(1)	0.38	0.84	2.41				
MPWB1K/cc-pVDZ									N/A(8)	N/A	N/A	N/A
MPWB1K/aug-cc-pVDZ	0.97(2)	0.12	1.94	9.99								
B97D/D95++(d,p)	**0.92**	**0.23**	**0.58**	**2.38**	**0.90**	**0.25**	**0.62**	**2.18**				
B97D/6-311G(d,p)					**0.95**	**0.17**	**0.71**	**2.40**	0.96	0.16	0.75	2.81
B97D/6-311++G(d,p)	**0.90**	**0.25**	**0.63**	**2.57**	**0.89**	**0.26**	**0.66**	**2.31**	0.90	0.28	0.65	2.22
B97D/aug-cc-pVDZ	0.81	0.35	0.59	2.04	0.86	0.30	0.65	2.15				
ωB97xD/aug-cc-pVDZ	0.83	0.33	0.75	3.41	**0.88**	**0.27**	**0.84**	**3.69**				
B3LYP/6-311++G(d,p)	0.88	0.28	0.93	3.63	0.84	0.32	1.01	3.49				
B3LYP/aug-cc-pVDZ	0.86	0.30	0.83	2.60	0.77	0.38	0.87	2.39				
X3LYP/aug-cc-pVDZ	**0.86**	**0.30**	**0.85**	**3.15**	0.79	0.37	0.91	3.03	0.86	0.32	0.93	3.07

[a]In parentheses are the number of compounds for which the SCF failed to converge within the given calculation setup. In bold are combinations that were selected for further investigation with the larger data set (see the text for details). [b]No-CP, calculations without CP correction. [c]SP-CP, geometry optimization without CP correction followed by CP-corrected single-point energy calculation. [d]CP, geometry optimization and energy calculation with CP correction. [e]SEE, standard error of estimation. [f]S, slope of the regression line. [g]I, intercept of the regression line with the y axis.

Table 3. Performance of the 10 Selected Calculational Methods on the Large Data Set of 58 HBAs[a]

	No-CP[b]				No-CP + ZPE[b]				SP-CP[c]				SP-CP + ZPE[c]			
combination	R^2	SEE[d]	S^e	I^f	R^2	SEE[d]	S^e	I^f	R^2	SEE[d]	S^e	I^f	R^2	SEE[d]	S^e	I^f
M06-2X/cc-pVDZ									**0.89**	**0.40**	**0.86**	**2.81**	**0.90**	**0.38**	**0.96**	**2.37**
B97D/D95++(d,p)[g]	0.73	0.59	0.66	2.97	0.71	0.61	0.73	2.65	0.78	0.53	0.72	2.83	0.77	0.55	0.80	2.52
B97D/6-311G(d,p)									0.73	0.58	0.77	2.85	0.71	0.61	0.86	2.40
B97D/6-311++G(d,p)	0.76	0.58	0.72	3.26	0.73	0.61	0.79	2.66	0.79	0.54	0.74	2.97	0.79	0.56	0.83	2.69
ωB97xD/aug-cc-pVDZ									**0.81**	**0.49**	**0.76**	**3.21**	**0.83**	**0.46**	**0.85**	**2.84**
X3LYP/aug-cc-pVDZ	**0.88**	**0.40**	**0.81**	**3.11**	**0.90**	**0.37**	**0.91**	**2.81**								
MP2/cc-pVDZ	0.74	0.60	0.53	2.80	0.75	0.60	0.59	2.63	0.70	0.65	0.84	1.76	0.61	0.74	0.86	0.87

[a]In bold, combinations selected as the best combinations (see the text for details). [b]No-CP, calculations without CP correction. [c]SP-CP, geometry optimization without CP correction followed by CP-corrected single-point energy calculation. [d]SEE, standard error of estimation. [e]S, slope of the regression line. [f]I, intercept of the regression line with the y axis. [g]Results of this combination are without the two bromine-containing compounds in the data set.

lower R^2 values (between 0.61 and 0.79) and a SEE between 0.54 and 0.74. Generally, ZPE has no significant impact on the correlation in terms of R^2 and SEE; however, it improves the slope (S) of the correlation and decreases the intercept (I) to a constant range of 2.00−3.00 kcal mol^{-1} in the different combinations. This may represent the entropic penalty for the H-bond formation.[86] The ZPE correction results in only a 10−15% increase in the computational time.

In contrast to the case of the small data set, this selection of the best combinations on the basis of the results for the large data set should be statistically meaningful. This proposal is supported by our analysis of the significance difference between the different combinations correlations using Williams' modification of the Hotelling's test[87] (Supporting Information S2).

Figure 3 shows the correlation between the ZPE-corrected interaction energy and the experimental ΔG for the M06-2X/cc-pVDZ(SP-CP) combination applied to the large HBA set. As can be seen, the combination gives interaction energies that correlate reasonably well to ΔG, with R^2 and SEE values of 0.90 and 0.38, respectively. This is true for the different types of HBAs (nitrogen, oxygen, or halogen as the HBA atom), indicating the generality of the method. For a group of compounds, (colored black in Figure 3), the calculation overestimates the interaction in all combinations investigated with the exception of the X3LYP functional (see Figure 5). These are the three nitrogen heterocycles shown in Figure 4 with alkyl substituents ortho to ring nitrogens. This suggests that with the exception of the X3LYP functional the different DFT combinations could not describe correctly the exchange repulsion energy (steric repulsion) that results from the ortho-alkyl groups.

Figure 5 shows the ability of the X3LYP/aug-cc-pVDZ (No-CP) to handle the different types of HBAs properly, including the ortho-alkyl-substituted nitrogen compounds that were not treated correctly by the other combinations.

In terms of CPU time requirements for calculating the whole data set, the best three combinations rank as M06-2X/cc-pVDZ(SP-CP) < X3LYP/aug-cc-pVDZ(No-CP) < ωB97xD/aug-cc-pVDZ(SP-CP).

To optimize the CPU requirement for the three best combinations further, we tested the effect of reducing the basis set size and/or omitting the SP-CP correction. As can be seen in Table 4, in all cases, the omission of the SP-CP correction significantly reduces the performance of the chosen DFT combination. However, reducing the basis-set size but still using

dx.doi.org/10.1021/ci4006222 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

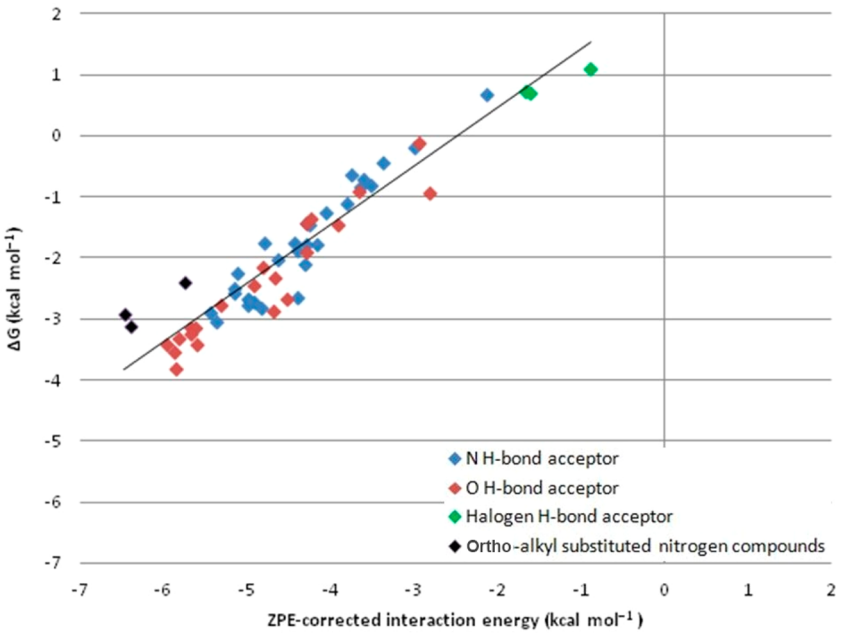


Figure 3. ZPE-corrected calculated interaction energy vs experimental ΔG for M06-2X/cc-pVDZ(SP-CP) for the large HBAs set. The color coding is according to the type of hydrogen-bond acceptor (see the legend). The black line is the regression line.

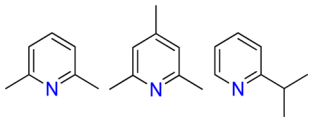


Figure 4. ortho-Alkyl nitrogen heterocycles in the large data set for which most combinations overestimate the interaction energy with methanol (black squares in Figure 3).

SP-CP brings the performance back to a level close to the original combinations at a significantly reduced computational cost, as can be seen in Table 4 from the ratio between the overall CPU times required for the energy calculation for the whole data set for each combination.

These results suggest that the calculation of the interaction energy after geometry optimization requires either a large basis set or a counterpoise correction to correct the BSSE error to achieve a good correlation with experimental data. To test this hypothesis, we carried out the geometry optimization using several functionals with the small cc-pVDZ basis set followed by a SP energy calculation with the larger aug-cc-pVTZ basis

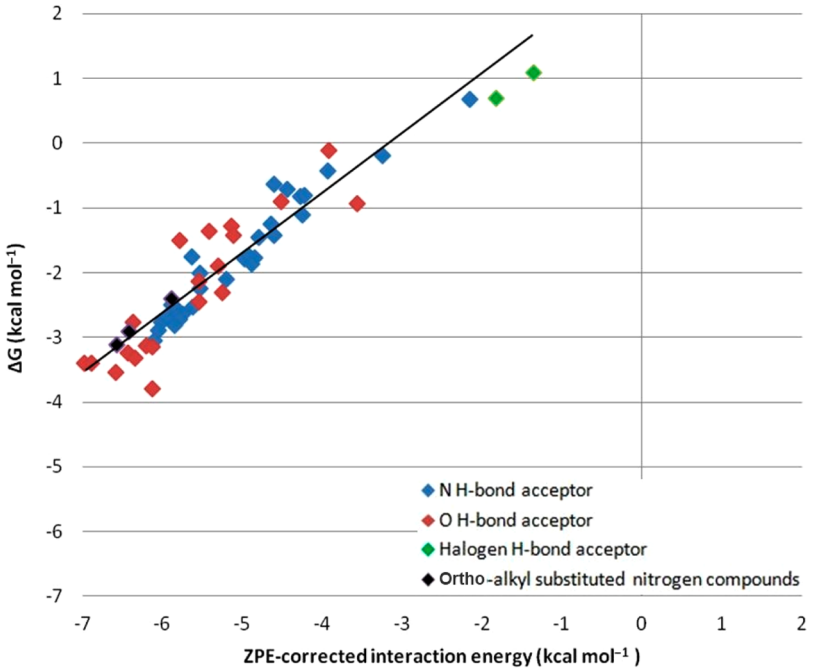


Figure 5. ZPE-corrected calculated interaction energy vs experimental ΔG for X3LYP/aug-cc-pVDZ(No-CP) for the large HBAs set. The color coding is according to the type of hydrogen-bond acceptor (see the legend). The black line is the regression line.

dx.doi.org/10.1021/ci4006222 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 4. Performance of the Best Three Combinations with Their Five Alternative Combinations on the Large Data Set with Their Relative CPU Time Requirements[a]

combination	No-ZPE R^2	No-ZPE SEE[b]	No-ZPE S[c]	No-ZPE I[d]	+ZPE R^2	+ZPE SEE[b]	+ZPE S[c]	+ZPE I[d]	CPU time ratio
M06-2X/cc-pVDZ(SP-CP)	**0.89**	**0.40**	**0.86**	**2.81**	**0.90**	**0.38**	**0.96**	**2.37**	**1.00**
M06-2X/cc-pVDZ(No-CP)	0.75	0.60	0.52	2.63	0.76	0.60	0.55	2.25	0.99
X3LYP/aug-cc-pVDZ(No-CP)	**0.88**	**0.40**	**0.81**	**3.11**	**0.90**	**0.37**	**0.91**	**2.81**	**3.70**
X3LYP/cc-pVDZ(No-CP)	0.77	0.56	0.54	2.83	0.79	0.55	0.60	2.66	0.37
X3LYP/cc-pVDZ(SP-CP)	0.87	0.42	0.91	2.60	0.85	0.46	1.01	1.96	0.38
ωB97xD/aug-cc-pVDZ(SP-CP)	**0.81**	**0.49**	**0.76**	**3.21**	**0.83**	**0.46**	**0.85**	**2.84**	**7.38**
ωB97xD/cc-pVDZ(No-CP)	0.72	0.62	0.50	2.71	0.71	0.63	0.54	2.48	1.10
ωB97xD/cc-pVDZ(SP-CP)	0.83	0.49	0.80	2.92	0.80	0.53	0.90	2.47	1.11

[a]In bold, the original DFT functional/basis set combinations. [b]SEE, standard error of estimation. [c]S, slope of the regression line. [d]I, intercept of the regression line with the y axis.

Table 5. Performance of Using Larger Basis Set for Interaction Energy Calculation

combination	No-ZPE R^2	No-ZPE SEE[a]	No-ZPE S[b]	No-ZPE I[c]	+ZPE R^2	+ZPE SEE[a]	+ZPE S[b]	+ZPE I[c]
M06-2X/aug-cc-pVTZ//M06-2X/cc-pVDZ	0.88	0.41	0.76	2.94	0.88	0.42	0.81	2.05
X3LYP/aug-cc-pVTZ//X3LYP/cc-pVDZ	0.85	0.46	0.77	2.37	0.83	0.49	0.81	1.30
ωB97xD/aug-cc-pVTZ//ωB97xD/cc-pVDZ	0.82	0.50	0.71	3.00	0.78	0.56	0.76	2.02

[a]SEE, standard error of estimation. [b]S, slope of the regression line. [c]I, intercept of the regression line with the y axis.

set. As Table 5 shows, indeed the results of these calculations are close to those obtained with the original DFT functional/basis set combinations shown in Table 4 in bold, proving our hypothesis and emphasizing that geometries obtained at a moderate level of theory (double-ζ basis set, no CP) are quite accurate.

The large-scale test thus identified five suitable (in terms of accuracy and computational requirements) DFT functional/basis set combinations that could be used regularly for ranking HBAs in drug design. These are X3LYP/cc-pVDZ(SP-CP), M06-2X/cc-pVDZ(SP-CP), ωB97xD/cc-pVDZ(SP-CP), X3LYP/aug-cc-pVDZ(No-CP), ωB97xD/aug-cc-pVDZ(SP-CP).

Effect of Additional Geometrical Constraints during Optimization. To make the process even more economical, we assessed the effect of applying additional geometrical constraints to the complex during the optimization process to test if these constraints might enhance and accelerate convergence and so decrease the computational time. Additional constraints could possibly also increase the performance by decreasing the undesirable secondary interactions because they prevent further functional groups of the HBA molecule from approaching the methanol OH functionality closely enough to form these secondary interactions. This aspect of the calculations is important because secondary interactions complicate the calculation of interaction energies because factors other than those important for hydrogen bonding may be important. In general, weak secondary interactions, which are hindered by the constraints, are a gas-phase phenomenon. In the condensed phases relevant to drug design (aqueous solution or a protein environment), secondary interactions are overridden by binding interactions to the environment.

Two additional constraints were applied to the methanol−HBA complex (colored red in Figure 6) in addition to the original constraint of the methanol OH−H-bond-acceptor atom angle to 180° (colored black in Figure 6):

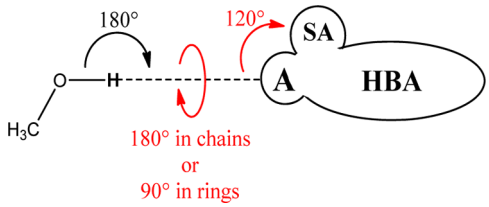


Figure 6. Original (black) and additional (red) constraints applied to the complexes. A, H-bond-acceptor atom; SA, second atom to the H-bond-acceptor atom; and HBA, H-bond acceptor.

(a) The angle between the methanol OH hydrogen atom, the H-bond-acceptor atom. and the second atom to the H-bond-acceptor atom was fixed to 120°.[84]
(b) The dihedral angle between the methanol carbon atom, the methanol OH hydrogen atom, the H-bond-acceptor atom, and the second atom to the H-bond-acceptor atom was kept at 180° for H-bond-acceptor atoms in chains and at 90° for H-bond-acceptor atoms in rings.

We used the best three levels as identified from Table 4: M06-2X/cc-pVDZ(SP-CP), X3LYP/cc-pVDZ(SP-CP), and X3LYP/aug-cc-pVDZ(No-CP). Table 6 shows the performance of the used combinations upon the additional constraints application as well as their effect on the CPU computational time.

Table 6 shows that within the protocol used additional constraints enhance and accelerate convergence of the geometry optimization. This improvement amounts to an approximately 50% decrease in the total CPU computational time required for optimizing the 58 HBAs. Despite this reduction in the computational cost, the correlation to the experimental values was almost unchanged. As was found for the calculations that used the regular constraints, ZPE corrections have no significant effect on the correlation in terms of R^2 and SEE; nevertheless, they also improve the slope and the intercept of the correlation line when the additional

dx.doi.org/10.1021/ci4006222 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 6. Effect of Multiple Constraints Application on the Combinations Performance and Their Computational Time[a]

combination	No-ZPE R^2	No-ZPE SEE[b]	No-ZPE S^c	No-ZPE I^d	+ZPE R^2	+ZPE SEE[b]	+ZPE S^c	+ZPE I^d	CPU time ratio
M06-2X/cc-pVDZ(SP-CP)	0.89	0.40	0.85	2.71	0.89	0.40	0.91	2.17	0.44
X3LYP/aug-cc-pVDZ(No-CP)	0.88	0.41	0.75	2.55	0.89	0.39	0.84	2.27	1.57
X3LYP/cc-pVDZ(SP-CP)	0.89	0.40	0.90	2.61	0.88	0.42	1.02	2.09	0.18

[a]Compare to single angle constraints results in Table 4. [b]SEE, standard error of estimation. [c]S, slope of the regression line. [d]I, intercept of the regression line with the y axis.

constraints are used. They result in an increase of 20−30% in the computational requirements.

Validation Using an Independent Test Set. Many of the experimental values in the pK_{BHX} database were generated with 4-fluorophenol as a HBD and not with methanol. Therefore, we wanted to assess if our proposed protocol, which was established with methanol as a HBD, can be transferred to the experimental data obtained with 4-fluorophenol. This was done using 17 HBAs from the pK_{BHX} data set (Supporting Information S3, Figure 2) for which the experimental pK_{BHX} values were determined using 4-fluorophenol. Most of the HBAs were chosen to differ chemically from those used in the previous steps to test the generality and transferability of the method. The least computationally demanding combination, X3LYP/cc-pVDZ(SP-CP), was used once with the single angle constraints and once with the multiple constraints. We note that methanol, not 4-fluorophenol, was used as the HBD in the calculations according to the standard protocol. Table 7 shows the performance of the tested DFT functional/basis set combination on the selected test set HBAs.

Table 7. Performance of the X3LYP/cc-pVDZ(SP-CP) Combination on the Test Set Using the Single Angle Constraint and the Multiple Constraints

Single Angle Constraint							
No-ZPE				+ZPE			
R^2	SEE[a]	S^b	I^c	R^2	SEE[a]	S^b	I^c
0.88	0.46	0.90	2.89	0.83	0.54	1.01	2.18
Multiple Constraints							
No-ZPE				+ZPE			
R^2	SEE[a]	S^b	I^c	R^2	SEE[a]	S^b	I^c
0.92	0.37	0.87	2.75	0.90	0.42	0.95	2.10

[a]SEE, standard error of estimation. [b]S, slope of the regression line. [c]I, intercept of the regression line with the y axis.

Table 7 shows a slight decrease in the performance when only the single constraint is used, whereas the R^2 and SEE are 0.88 and 0.46, respectively, compared to 0.92 and 0.37, respectively, in the case of using multiple constraints.

The difference in performance between using the single angle constraint only and using the multiple constraints is due to a single compound (benzamide). In the calculation with the single constraint, the methanol OH forms an additional interaction (Figure 7a), which lowers its interaction energy. The additional constraints prevent this geometry during the optimization as does a slightly different starting geometry of the complex (Figure 7b), giving an interaction energy much better in line with those of the other complexes.

The above results suggest that we can obtain an accurate ranking of H-bond strengths for a set of chemically diverse HBAs using the proposed protocol, which uses computationally

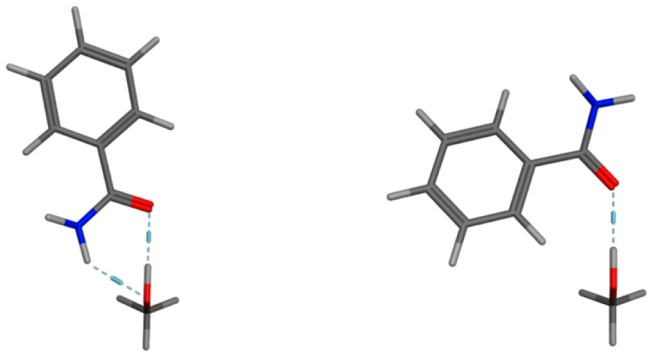


Figure 7.

economical DFT levels. It also indicates that using additional constraints not only reduces the computational time by 50% but also prevents the occurrence of additional interactions that may lower the interaction energy and thus improves the performance of the method. This emphasizes that the optimized complexes should be inspected for additional interactions, which may lead to false results.

As another means to validate our suggested protocol, we used the least-squares fit equations derived in the previous sections (with ZPE correction) for the combination used in energy calculation X3LYP/cc-pVDZ(SP-CP) in both the single angle (Table 4) and multiple constraints (Table 6) cases to predict ΔG for the interaction of the test set HBAs using their X3LYP/cc-pVDZ(SP-CP)-calculated ZPE-corrected interaction energies. Figure 8a,b shows the correlations between experimental and predicted ΔG in the single angle constraint and multiple constraints cases, respectively.

The single angle constraint and multiple constraints gave an R^2 value of 0.88 and 0.91, respectively, and root-mean-square error (rmse) of 0.48 and 0.42, respectively, indicating the ability of the protocol to predict ΔG values reliably and accurately, especially in the case where multiple constraints are used.

CONCLUSIONS

We have shown that computationally affordable levels of DFT functional/basis set can be used to assess and rank hydrogen-bond acceptors efficiently and accurately. Perhaps surprisingly, using geometrical constraints during geometry optimizations of the hydrogen-bonded complexes not only results in an approximately 50% reduction in the CPU time required but also improves the correlation with experimental data. This may be an expression of the geometrical effect of surrounding solvent in preventing the full gas-phase interaction between hydrogen-bond partners from being formed in solution. We also note that, as might be expected, increasing the size of the basis set and applying a counterpoise correction have similar effects on the results.

dx.doi.org/10.1021/ci4006222 | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

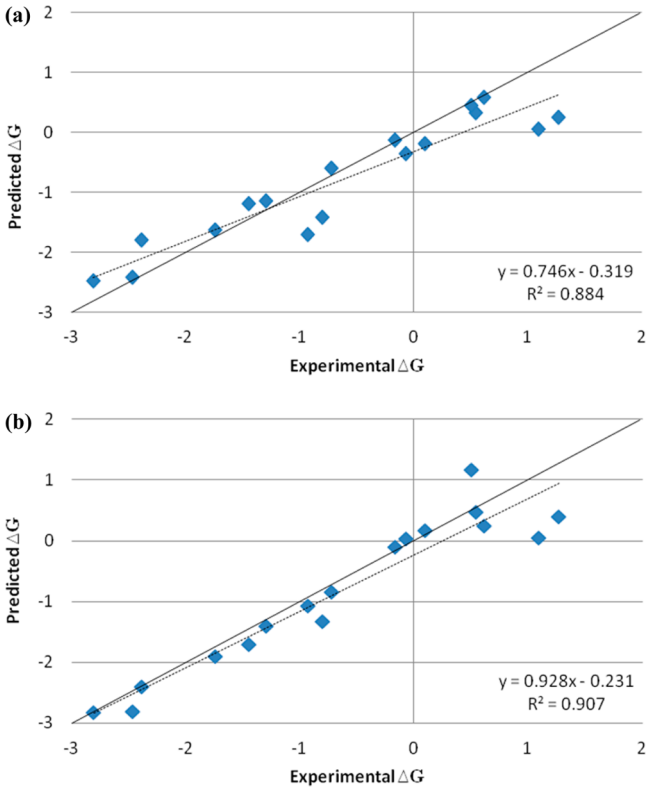


Figure 8. Experimental versus predicted ΔG using the single constraint (a) and multiple constraints (b) with the correlation line (dashed) and the 1:1 line (continuous).

The recommended DFT-levels for ranking hydrogen-bond acceptors are M06-2X/cc-pVDZ(SP-CP), X3LYP/aug-cc-pVDZ(No-CP), and X3LYP/cc-pVDZ(SP-CP), of which the third is the most economical computationally.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Set of 58 HBAs for the evaluation DFT combinations (large set), Williams−Hotelling test analysis of the correlations of the different combinations in the large data set investigation, and set of HBAs for the validation of the proposed setup. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: tim.clark@fau.de.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the Bundesministerium für Bildung und Forschung (BMBF) as part of the hpCADD project and by the grant of a fellowship to A.E.K. by the Deutscher Akademischer Austauschdienst.

■ REFERENCES

(1) Kubinyi, H. Hydrogen bonding: The last mystery in drug design? *Pharmacokinetic Optimization in Drug Research*; Verlag Helvetica Chimica Acta: Zürich, Switzerland, 2007; pp 513−524.

(2) Fang, H.; Tong, W. D.; Shi, L. M.; Blair, R.; Perkins, R.; Branham, W.; Hass, B. S.; Xie, Q.; Dial, S. L.; Moland, C. L.; Sheehan, D. M. Structure-activity relationships for a large diverse set of natural, synthetic, and environmental estrogens. *Chem. Res. Toxicol.* **2001**, *14*, 280−294.

(3) Katragadda, M.; Magotti, P.; Sfyroera, G.; Lambris, J. D. Hydrophobic effect and hydrogen bonds account for the improved activity of a complement inhibitor, compstatin. *J. Med. Chem.* **2006**, *49*, 4616−4622.

(4) Fujita, T.; Nishioka, T.; Nakajima, M. Hydrogen-bonding parameter and its significance in quantitative structure-activity studies. *J. Med. Chem.* **1977**, *20*, 1071−1081.

(5) Charton, M.; Charton, B. I. Hyper-conjugation as a parameter in correlation-analysis. *J. Org. Chem.* **1982**, *47*, 8−13.

(6) Yang, G. Z.; Lien, E. J.; Guo, Z. R. Physical factors contributing to hydrophobic constant π. *Quant. Struct.-Act. Relat.* **1986**, *5*, 12−18.

(7) Seiler, P. Interconversion of lipophilicities from hydrocarbon-water systems into octanol-water system. *Eur. J. Med. Chem.* **1974**, *9*, 473−479.

(8) Kamlet, M. J.; Doherty, R. M.; Abboud, J. L. M.; Abraham, M. H.; Taft, R. W. Linear solvation energy relationships 0.36. Molecular-properties governing solubilities of organic nonelectrolytes in water. *J. Pharm. Sci.* **1986**, *75*, 338−349.

(9) Kamlet, M. J.; Doherty, R. M.; Taft, R. W.; Abraham, M. H.; Veith, G. D.; Abraham, D. J. Solubility properties in polymers and biological media 0.8. An analysis of the factors that influence toxicities of organic nonelectrolytes to the golden Orfe fish (Leuciscus-Idus-Melanotus). *Environ. Sci. Technol.* **1987**, *21*, 149−155.

(10) Abraham, M. H.; Duce, P. P.; Prior, D. V.; Barratt, D. G.; Morris, J. J.; Taylor, P. J. Hydrogen-bonding 0.9. Solute proton donor and proton acceptor scales for use in drug design. *J. Chem. Soc., Perkin Trans. 2* **1989**, 1355−1375.

(11) Wilson, L. Y.; Famini, G. R. Using theoretical descriptors in quantitative structure-activity-relationships: Some toxicological indexes. *J. Med. Chem.* **1991**, *34*, 1668−1674.

(12) Dearden, J. C.; Ghafourian, T. Investigation of calculated hydrogen bonding parameters for QSAR. In *QSAR and Molecular Modelling: Concepts, Computational Tools and Biological Applications*; Sanz, F., Giraldo, J., Manaut, F., Eds.; Prous Science Publishers: Barcelona, Spain, 1995; pp 117−119.

(13) Dearden, J. C.; Cronin, M. T. D.; Wee, D. Prediction of hydrogen bond donor ability using new quantum chemical parameters. *J. Pharm. Pharmacol.* **1997**, *49*, 110.

(14) Gancia, E.; Montana, J. G.; Manallack, D. T. Theoretical hydrogen bonding parameters for drug design. *J. Mol. Graphics Modell.* **2001**, *19*, 349−362.

(15) Murray, J. S.; Politzer, P. Correlations between the solvent hydrogen-bond-donating parameter alpha and the calculated molecular-surface electrostatic potential. *J. Org. Chem.* **1991**, *56*, 6715−6717.

(16) Kenny, P. W. Hydrogen bonding, electrostatic potential, and molecular design. *J. Chem. Inf. Model.* **2009**, *49*, 1234−1244.

(17) Hennemann, M.; Clark, T. A QSPR-approach to the estimation of the pK_{HB} of six-membered nitrogen-heterocycles using quantum mechanically derived descriptors. *J. Mol. Model.* **2002**, *8*, 95−101.

(18) Smallwood, C. J.; McAllister, M. A. Characterization of low-barrier hydrogen bonds 0.7. Relationship between strength and geometry of short-strong hydrogen bonds. The formic acid formate anion model system. An ab initio and DFT investigation. *J. Am. Chem. Soc.* **1997**, *119*, 11277−11281.

(19) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. Hydrogen bonding in DNA base pairs: Reconciliation of theory and experiment. *J. Am. Chem. Soc.* **2000**, *122*, 4117−4128.

(20) Koch, W.; Holthausen, M. C. Hydrogen bonds and weakly bound systems. *A Chemist's Guide to Density Functional Theory*; Wiley-VCH: Weinheim, Germany, 2001; pp 213−235.

(21) Morozov, A. V.; Kortemme, T.; Tsemekhman, K.; Baker, D. Close agreement between the orientation dependence of hydrogen bonds observed in protein structures and quantum mechanical calculations. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6946−6951.

(22) Kone, M.; Illien, B.; Graton, J.; Laurence, C. B3LYP and MP2 calculations of the enthalpies of hydrogen-bonded complexes of methanol with neutral bases and anions: Comparison with experimental data. *J. Phys. Chem. A* **2005**, *109*, 11907−11913.
(23) Schwöbel, J.; Ebert, R. U.; Kuhne, R.; Schuurmann, G. Modeling the H bond donor strength of -OH, -NH, and -CH sites by local molecular parameters. *J. Comput. Chem.* **2009**, *30*, 1454−1464.
(24) Schwöbel, J.; Ebert, R. U.; Kuhne, R.; Schuurmann, G. Prediction of the intrinsic hydrogen bond acceptor strength of organic compounds by local molecular parameters. *J. Chem. Inf. Model.* **2009**, *49*, 956−962.
(25) Hao, M. H. Theoretical calculation of hydrogen-bonding strength for drug molecules. *J. Chem. Theory Comput.* **2006**, *2*, 863−872.
(26) Nocker, M.; Handschuh, S.; Tautermann, C.; Liedl, K. R. Theoretical prediction of hydrogen bond strength for use in molecular modeling. *J. Chem. Inf. Model.* **2009**, *49*, 2067−2076.
(27) Jencks, W. P. Binding energy, specificity, and enzymic catalysis: The circe effect. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1975**, *43*, 219−410.
(28) Page, M. I. Entropy, Binding-Energy, and Enzymic Catalysis. *Angew. Chem., Int. Ed. Engl.* **1977**, *16*, 449−459.
(29) Hobza, P.; Zahradnik, R.; Muller-Dethlefs, K. The world of non-covalent interactions: 2006. *Collect. Czech. Chem. Commun.* **2006**, *71*, 443−531.
(30) Scuseria, G. E.; Janssen, C. L.; Schaefer, H. F. An efficient reformulation of the closed-shell coupled cluster single and double excitation (CCSD) equations. *J. Chem. Phys.* **1988**, *89*, 7382−7387.
(31) Pople, J. A.; Head-Gordon, M.; Raghavachari, K. Quadratic configuration-interaction. A general technique for determining electron correlation energies. *J. Chem. Phys.* **1987**, *87*, 5968−5975.
(32) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. A direct MP2 gradient-method. *Chem. Phys. Lett.* **1990**, *166*, 275−280.
(33) Frisch, M. J.; Head-Gordon, M.; Pople, J. A. Semidirect algorithms for the MP2 energy and gradient. *Chem. Phys. Lett.* **1990**, *166*, 281−289.
(34) Head-Gordon, M.; Head-Gordon, T. Analytic MP2 frequencies without fifth-order storage. Theory and application to bifurcated hydrogen-bonds in the water hexamer. *Chem. Phys. Lett.* **1994**, *220*, 122−128.
(35) Head-Gordon, M.; Pople, J. A.; Frisch, M. J. MP2 energy evaluation by direct methods. *Chem. Phys. Lett.* **1988**, *153*, 503−506.
(36) Møller, C.; Plesset, M. S. Note on an approximation treatment for many-electron systems. *Phys. Rev.* **1934**, *46*, 618−622.
(37) Saebo, S.; Almlof, J. Avoiding the integral storage bottleneck in LCAO calculations of electron correlation. *Chem. Phys. Lett.* **1989**, *154*, 83−89.
(38) Cerny, J.; Hobza, P. Non-covalent interactions in biomacromolecules. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5291−5303.
(39) Daabkowska, I.; Jurecka, P.; Hobza, P. On geometries of stacked and H-bonded nucleic acid base pairs determined at various DFT, MP2, and CCSD(T) levels up to the CCSD(T)/complete basis set limit level. *J. Chem. Phys.* **2005**, *122*, 204322.
(40) Raub, S.; Marian, C. M. Quantum chemical investigation of hydrogen-bond strengths and partition into donor and acceptor contributions. *J. Comput. Chem.* **2007**, *28*, 1503−1515.
(41) Tsuzuki, S.; Uchimaru, T.; Matsumura, K.; Mikami, M.; Tanabe, K. Effects of basis set and electron correlation on the calculated interaction energies of hydrogen bonding complexes: MP2/cc-pV5Z calculations of H(2)O-MeOH, H(2)O-Me(2)O, H(2)O-H(2)CO, MeOH-MeOH, and HCOOH-HCOOH complexes. *J. Chem. Phys.* **1999**, *110*, 11906−11910.
(42) Plumley, J. A.; Dannenberg, J. J. A comparison of the behavior of functional/basis set combinations for hydrogen-bonding in the water dimer with emphasis on basis set superposition error. *J. Comput. Chem.* **2011**, *32*, 1519−1527.
(43) Muller-Dethlefs, K.; Hobza, P. Noncovalent interactions: A challenge for experiment and theory. *Chem. Rev.* **2000**, *100*, 143−167.
(44) Zhao, Y.; Truhlar, D. G. 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.
(45) Grimme, S. Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* **2006**, *27*, 1787−1799.
(46) Chai, J. D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615−6620.
(47) Xu, X.; Goddard, W. A. The X3LYP extended density functional for accurate descriptions of nonbond interactions, spin states, and thermochemical properties. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2673−2677.
(48) Grimme, S. Density functional theory with London dispersion corrections. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 211−228.
(49) Boys, S. F.; Bernardi, F. Calculation of small molecular interactions by differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553−566.
(50) Alvarez-Idaboy, J. R.; Galano, A. Counterpoise corrected interaction energies are not systematically better than uncorrected ones: comparison with CCSD(T) CBS extrapolated values. *Theor. Chem. Acc.* **2010**, *126*, 75−85.
(51) Simon, S.; Duran, M.; Dannenberg, J. J. Effect of basis set superposition error on the water dimer surface calculated at Hartree-Fock, Moller-Plesset, and density functional theory levels. *J. Phys. Chem. A* **1999**, *103*, 1640−1643.
(52) Ireta, J.; Neugebauer, J.; Scheffler, M. On the accuracy of DFT for describing hydrogen bonds: Dependence on the bond directionality. *J. Phys. Chem. A* **2004**, *108*, 5692−5698.
(53) Wieczorek, R.; Haskamp, L.; Dannenberg, J. J. Molecular orbital calculations of water clusters on counterpoise-corrected potential energy surfaces. *J. Phys. Chem. A* **2004**, *108*, 6713−6723.
(54) Hobza, P.; Havlas, Z. Counterpoise-corrected potential energy surfaces of simple H-bonded systems. *Theor. Chem. Acc.* **1998**, *99*, 372−377.
(55) Liedl, K. R. Dangers of counterpoise corrected hypersurfaces. Advantages of basis set superposition improvement. *J. Chem. Phys.* **1998**, *108*, 3199−3204.
(56) Zhao, Y.; Truhlar, D. G. Density functionals with broad applicability in chemistry. *Acc. Chem. Res.* **2008**, *41*, 157−167.
(57) Zhao, Y.; Tishchenko, O.; Truhlar, D. G. How well can density functional methods describe hydrogen bonds to π acceptors? *J. Phys. Chem. B* **2005**, *109*, 19046−19051.
(58) Zhao, Y.; Truhlar, D. G. Benchmark databases for nonbonded interactions and their use to test density functional theory. *J. Chem. Theory Comput.* **2005**, *1*, 415−432.
(59) Zhao, Y.; Truhlar, D. G. Density functionals for noncovalent interaction energies of biological importance. *J. Chem. Theory Comput.* **2007**, *3*, 289−300.
(60) Campen, R. K.; Kubicki, J. D. Interaction energy and the shift in OH stretch frequency on hydrogen bonding for the $H_2O \rightarrow H_2O$, $CH_3OH \rightarrow H_2O$, and $H_2O \rightarrow CH_3OH$ dimers. *J. Comput. Chem.* **2010**, *31*, 963−972.
(61) Xu, X.; Goddard, W. A. Bonding properties of the water dimer: A comparative study of density functional theories. *J. Phys. Chem. A* **2004**, *108*, 2305−2313.
(62) Laurence, C.; Brameld, K. A.; Graton, J.; Le Questel, J. Y.; Renault, E. The pK_{BHX} database: Toward a better understanding of hydrogen-bond basicity for medicinal chemists. *J. Med. Chem.* **2009**, *52*, 4073−4086.
(63) Taft, R. W.; Gurka, D.; Joris, L.; Schleyer, P. v. R.; Rakshys, J. W. Studies of hydrogen-bonded complex formation with p-fluorophenol. V. Linear free energy relationships with OH reference acids. *J. Am. Chem. Soc.* **1969**, *91*, 4801−4808.
(64) Abraham, M. H.; Grellier, P. L.; Prior, D. V.; Morris, J. J.; Taylor, P. J. Hydrogen-bonding Part 10. A scale of solute hydrogen-

bond basicity using log K values for complexation in tetrachloromethane. *J. Chem. Soc., Perkin Trans. 2* **1990**, 521−529.
(65) Abraham, M. H. Hydrogen-bonding Part 31. Construction of a scale of solute effective or summation hydrogen-bond basicity. *J. Phys. Org. Chem.* **1993**, *6*, 660−684.
(66) Abraham, M. H.; Platts, J. A. Hydrogen bond structural group constants. *J. Org. Chem.* **2001**, *66*, 3484−3491.
(67) El Kerdawy, A.; Wick, C. R.; Hennemann, M.; Clark, T. Predicting the sites and energies of noncovalent intermolecular interactions using local properties. *J. Chem. Inf. Model.* **2012**, *52*, 1061−1071.
(68) Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic-behavior. *Phys. Rev. A* **1988**, *38*, 3098−3100.
(69) Becke, A. D. Density-functional thermochemistry Part 3. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(70) Lee, C. T.; Yang, W. T.; Parr, R. G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron-density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.
(71) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate spin-dependent electron liquid correlation energies for local spin-density calculations: A critical analysis. *Can. J. Phys.* **1980**, *58*, 1200−1211.
(72) Dunning, T. H. Gaussian-basis sets for use in correlated molecular calculations Part 1. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007−1023.
(73) Kendall, R. A.; Dunning, T. H.; Harrison, R. J. Electron-affinities of the first row atoms revisited. Systematic basis-sets and wave-functions. *J. Chem. Phys.* **1992**, *96*, 6796−6806.
(74) Woon, D. E.; Dunning, T. H. Gaussian-basis sets for use in correlated molecular calculations Part 3. The atoms aluminum through argon. *J. Chem. Phys.* **1993**, *98*, 1358−1371.
(75) Davidson, E. R. Comment on Dunning's correlation-consistent basis sets. *Chem. Phys. Lett.* **1996**, *260*, 514−518.
(76) Grimme, S. Semiempirical hybrid density functional with perturbative second-order correlation. *J. Chem. Phys.* **2006**, *124*, 034108.
(77) Schwabe, T.; Grimme, S. Double-hybrid density functionals with long-range dispersion corrections: Higher accuracy and extended applicability. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397−3406.
(78) Zhao, Y.; Truhlar, D. G. Hybrid meta density functional theory methods for thermochemistry, thermochemical kinetics, and non-covalent interactions: The MPW1B95 and MPWB1K models and comparative assessments for hydrogen bonding and van der Waals interactions. *J. Phys. Chem. A* **2004**, *108*, 6908−6918.
(79) Sadowski, J.; Schwab, C.; Gasteiger, J. *Corina 3.4*; Molecular Networks GmbH: Erlangen, Germany.
(80) Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(81) Halgren, T. A. MMFF VI. MMFF94s option for energy minimization studies. *J. Comput. Chem.* **1999**, *20*, 720−729.
(82) Halgren, T. A. MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *J. Comput. Chem.* **1999**, *20*, 730−748.
(83) *Molecular Operating Environment (MOE)*, 2011.10; Chemical Computing Group Inc.: Montreal, QC, Canada, 2011.
(84) Ahmed, M.; Jelsch, C.; Guillot, B.; Lecomte, C.; Domagała, S. Relationship between stereochemistry and charge density in hydrogen bonds with oxygen acceptors. *Cryst. Growth Des.* **2012**, *13*, 315−325.
(85) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, revision A.02; Gaussian, Inc.: Wallingford, CT, 2009.
(86) Yu, H. T.; Rick, S. W. Free energy, entropy, and enthalpy of a water molecule in various protein environments. *J. Phys. Chem. B* **2010**, *114*, 11552−11560.
(87) Williams, E. J. *Regression Analysis*; John Wiley and Sons, Inc.: New York, 1959.