J Mol Model (2009) 15:1051–1060
DOI 10.1007/s00894-009-0459-4

ORIGINAL PAPER

Performance of Becke's half-and-half functional for non-covalent interactions: energetics, geometries and electron densities

Konstantinos Gkionis · J. Grant Hill · Steven P. Oldfield · James A. Platts

Received: 29 October 2008 / Accepted: 11 December 2008 / Published online: 11 February 2009

Abstract The performance of Becke's half-and-half functional, BHandH, for description of non-covalent interactions is reported, using high-level *ab initio* results as benchmarks. Binding energies are found to be well reproduced for complexes that are bound predominantly by dispersion, whereas significant and consistent overestimation is observed for hydrogen bonded complexes. Overall, the mean average error is around 2 kcal mol^{-1}, for all basis sets considered. The effect of changing the proportion of exact and Slater exchange in the functional is shown to alter the balance of description of hydrogen bonded and dispersion bound complexes, but does not improve the overall performance. However, a simple multiplicative scaling of binding energies is possible, and reduces the mean average error to less than 1 kcal mol^{-1}. The performance of the BHandH functional for geometry optimization was also studied, and in almost all cases the difference from *ab initio* geometries is small, with root mean square deviations of between 0.05 and 0.20 Å. Harmonic frequency calculation allow us to check whether optimized geometries are true minima at this level, and to estimate the zero point vibrational energy change on binding.

Keywords Density functional theory · Electron density analysis · Hydrogen bonding · π-stacking

K. Gkionis · J. G. Hill · S. P. Oldfield · J. A. Platts (✉)
School of Chemistry,
Cardiff University, Park Place,
Cardiff CF10 3AT, UK
e-mail: platts@cf.ac.uk

Introduction

Non-covalent interactions such as hydrogen bonding and π-stacking play a crucial role in a wide range of phenomena, not least in the structure and function of biomolecules, including peptides, proteins and nucleic acids [1]. Hydrogen bonding controls the backbone structure of protein chains, and is also responsible for correct recognition of bases within DNA, while π-stacking interactions stabilize the double-helix structure of DNA. Theoretical interest in characterizing such interactions is therefore longstanding and intense. The energetic stabilization due to non-covalent interactions is clearly of prime importance, and has probably received most attention. Other aspects, including geometrical preference, origins of stabilization, and characterization within the context of multiple interactions, are also important and have been studied in depth.

The current state of the art in theoretical study of non-covalent interactions is summarized well by the recent work of Hobza and co-workers [2]. Correct *ab initio* description of both hydrogen bonding and π-stacking requires large basis sets and at least some degree of electron correlation, as well as some account of basis set superposition error (BSSE). The energy and geometry of hydrogen bonds is well described using second order Møller-Plesset perturbation theory (MP2) and moderately large basis sets (Dunning's augmented triple-ζ or similar) [3]. In contrast, MP2 fails to give a proper quantitative description of π-stacking, even after extrapolation to the basis set limit. Instead, coupled cluster methods such as CCSD(T) are required for accuracy, which is typically achieved through use of corrections with smaller basis sets. In establishing these requirements, Hobza et al. have published databases of interaction energies for complexes incorporating all the important classes of interactions, which form the basis of the tests reported in this work [4].

In recent years, our group (amongst many others) has reported methods that attempt to allow study of non-covalent interactions without the need for very large basis sets or computationally expensive coupled-cluster methods. Density fitting (alternatively termed resolution of the identity by some authors) replaces expensive 4-index 2-electron integrals with a combination of 2- and 3-index ones, thus reducing the time required for the MP2 calculation by around an order of magnitude [5–7]. Local correlation methods use localized orbitals to restrict the excitations to sets of virtual orbitals (domains) that are spatially close to the occupied orbitals, and in doing so further reduce computational cost [8–12]. Additionally, the local construction also effectively eliminates BSSE when used with moderately large basis sets. The combination of density fitting (DF) and local electron correlation (L) MP2 methods, known as DF-LMP2, reproduces counterpoise corrected canonical MP2 potential energy surfaces for the benzene dimer [13]. Spin-component-scaling MP2 (SCS-MP2) methods scale the same-spin and opposite-spin electron pair contributions to the correlation energy by different factors [14], with a set of scaling parameters optimized for stacking interactions in nucleic acids known as SCSN [15]. This DF-SCSN-LMP2 method agrees with estimated CCSD(T) data for Hobza's S22 set of non-covalent interactions with a mean error of 0.27 kcal mol^{-1}.

A second strand of research that shows promise is the use of density functional theory (DFT). Standard methods of DFT perform well for hydrogen bonding, but fail completely in their description of π-stacking [16]. This failure is generally ascribed to the lack of dispersion in Kohn-Sham DFT. Several authors have therefore constructed methods that include dispersion-like terms in DFT, *via* Lennard-Jones or similar parametric forms [17–19]. Another approach varies the description of exchange within the exchange-correlation functional to mimic the effects of dispersion [20–22]. Within the various methods to take this approach, we showed that Becke's half-and-half functional, BHandH [23], performs surprisingly well in describing π-stacking interactions, performing better than MP2 on average across a wide range of stacked complexes [24]. However, BHandH consistently overestimates the strength of hydrogen bonds, leading to an imbalance in description of structures where both stacking and H-bonding are present. Previous work has shown the importance of environment (water, counterions *etc.*) in reproducing experimental data for the interactions within and between DNA strands [25, 26].

Finally, we have also employed electron densities within the framework of Bader's quantum theory of atoms in molecules (QTAIM) [27, 28], to further characterize and predict the strengths of non-covalent interactions. This is based on the concept of critical points in the electron density, *i.e.* points where $\nabla \rho = 0$, which can be used to unambiguously identify bonding interactions, including those in non-covalent interactions. As shown in many previous studies [29, 30], hydrogen bond strengths are closely related to properties at critical points of bonds involved in the H-bonds. In testing the BHandH method, we observed similar relations for π-stacked complexes [24]. For instance, the number of bond critical points observed increases with interaction energy, as does the total electron density summed at these points. A potential advantage of this approach is that the shortcomings of BHandH for H-bonding could be avoided by fitting DFT density properties to MP2 or CCSD(T) quality energies.

In this work, we test the performance of BHandH in calculating the energy, geometry and electron density of complexes bound by non-covalent interactions.

Methods

All DFT calculations were performed using the Gaussian03 suite of programs [31], while DF-LMP2 (and SCS variants thereof) benchmark data were produced with the Molpro package of *ab initio* programs [32]. Initially, energies and electron densities for the "S22" benchmark data set were calculated at the geometries reported in ref. [4] without modification, incorporating the point group symmetry indicated in that work. BSSE was accounted for using the counterpoise method [33]. Basis sets employed were 6-31G(d), 6-31+G(d,p), 6-311++G(d,p) [34, 35], and aug-cc-pV*n*Z (*n*=D, T) [3]. Subsequently, the geometry of each complex was optimized using BHandH/6-311++G(d,p), starting from the literature structure, and the resulting structure checked *via* harmonic frequency calculation using ultra fine grid for integration. Root mean square deviation (RMSD) between these structures was calculated using the Chemcraft package [36].

All DF-LMP2 calculations were carried out following the same procedure as that outlined in ref. [7] and employed the aug-cc-pVTZ orbital basis sets [3]. Density fitting approximations were enabled in the MP2 and HF steps of the calculations with the use of the aug-cc-pVTZ MP2-fitting auxiliary basis sets of Weigend *et al.* [37] and the cc-pVTZ JK-fitting auxiliary basis sets of Weigend, respectively [38]. Orbital localization was performed with the method of Pipek and Mezey [39] implemented with a Newton-Raphson algorithm to ensure convergence. Further problems in the generation of localized orbitals were avoided by eliminating the contribution of the diffuse basis functions to the convergence criterion. The local correlation domains were selected following the procedure of Boughton and Pulay with a completeness criterion of 0.985 [40] and the merging of rotationally invariant

domains in delocalized π-domains. DF-LMP2 binding energies were calculated at both literature and optimized geometries.

Atoms in molecules analysis on BHandH calculated electron densities was carried out using the AIM2000 package [41], allowing visual identification of all bond critical points (bcp's) involved in intermolecular interactions. Properties at these bcp's, most notably the electron density, were calculated and used as descriptors of binding energy, as previously demonstrated [24].

Results and discussion

Table 1 reports the mean and mean absolute errors (ME and MAE) of binding energies across the whole S22 set of model complexes, using BHandH with various basis sets at the literature geometries. In addition, this set can be broken down into seven complexes dominated by hydrogen bonding (HB), eight bound largely by dispersion forces (Disp), and seven where both types of interaction are important (Mix). This data shows that the average absolute error across this varied set of complexes varies only slightly with basis set, in a range of 2 to 2.5 kcal mol^{-1}. Inclusion of diffuse functions improves predictions for HB and Disp complexes, but not for Mix, while *f*-type basis functions give slightly improved predictions for all three classes. The average error is uniformly positive, indicating that BHandH tends to over-bind the complexes.

Errors in the binding energies of hydrogen-bonded complexes are large for BHandH, on average around 5 kcal mol^{-1}, but are much smaller for dispersion bound and mixed complexes. To put these values in context, MP2 binding energies extrapolated to the basis set limit (taken from ref. 4) gives MAE=0.80 (all), 0.15 (HB), 1.51 (Disp), and 0.64 (Mix). It should be noted that the MP2/CBS estimates were recently updated by Marchetti and Werner, using explicitly correlated methods and accounting for errors related to the uracil dimer that were present in the original paper (see ref. 42 for further discussion). Thus, BHandH provides binding energies of much poorer quality than MP2 for H-bonded complexes, but performs rather better than MP2 for dispersion bound complexes, where MP2 is known to over-bind [2, 23], while the two methods give comparable performance for mixed complexes. In general, the smallest basis set 6-31G(d) gives significantly larger errors than all others, in accord with our first paper on BHandH, in which at least one diffuse function on heavy atoms was required for reasonable accuracy. Mean counterpoise corrections are small, varying from 1.94 kcal mol^{-1} for 6-31G(d) to 0.49 kcal mol^{-1} for 6-311++G(2df,2p).

In our initial study of BHandH [24] it was concluded that the good performance for π-stacking stems from the use of local exchange, since the purely local density approximation (LDA) also gave reasonable results for the benzene dimer. This suggests that the performance of BHandH-like hybrid functionals may be tuned by altering the amounts of exact and Slater exchange employed. Table 2 reports errors for various combinations of exact and Slater exchange in the functional, keeping the LYP correlation functional throughout. This analysis reveals that increasing the amount of LDA exchange leads to worse predictions for all three classes of interaction. Increasing the amount of HF exchange improves performance for H-bonded and mixed complexes, but degrades that for dispersion bound ones. Indeed, Table 2 makes it clear that BHandH as originally defined includes close to the optimal amount of HF and LDA exchange for the S22 set. The overall error is reduced slightly with increased fractions of HF exchange, but the gains are small.

The fact that BHandH consistently over-binds non-covalently bound complexes in general (just 2 out of 22 complexes are predicted to be under-bound) suggests that prediction of binding energy made using this functional can be simply scaled by a multiplicative factor to improve performance. As shown in Fig. 1, this is indeed the case: the line of best fit is some way from the ideal $y=x$ line, but fits the data well, with $R^2=0.971$ and MAE=0.83 kcal mol^{-1}. Thus, this scaling process reduces the overall error by more than half. It is also possible to individually scale the

Table 1 Mean and mean absolute errors for S22 binding energies at literature geometry, using BHandH with various basis sets (kcal mol^{-1})

	ME	MAE	MAE HB	MAE Disp	MAE Mix	Mean BSSE
6-31G(d)	+1.73	2.40	5.22	1.08	1.10	1.94
6-31+G(d,p)	+1.90	2.12	4.89	0.60	1.10	0.59
6-311++G(d,p)	+1.93	2.10	4.81	0.54	1.16	0.70
6-311++G(2df,2p)	+1.81	1.99	4.58	0.55	1.05	0.49
aug-cc-pVDZ	+1.79	2.00	4.65	0.56	1.00	0.73

Table 2 Effect of variation of the amount of HF exchange contained in the BHandH functional for S22 binding energies, with 6-31+G(d,p) (kcal mol^{-1})

%HF, LDA	ME	MAE	MAE HB	MAE Disp	MAE Mix
30, 70	+3.06	3.06	6.30	1.31	1.83
40, 60	+2.47	2.52	5.57	0.79	1.46
50, 50	+1.90	2.12	4.89	0.60	1.10
60, 40	+1.36	1.93	4.25	0.89	0.80
70, 30	+0.84	1.88	3.65	1.36	0.72
80, 20	+0.35	1.86	3.09	1.86	0.64
90, 10	-0.11	1.89	2.56	2.40	0.63

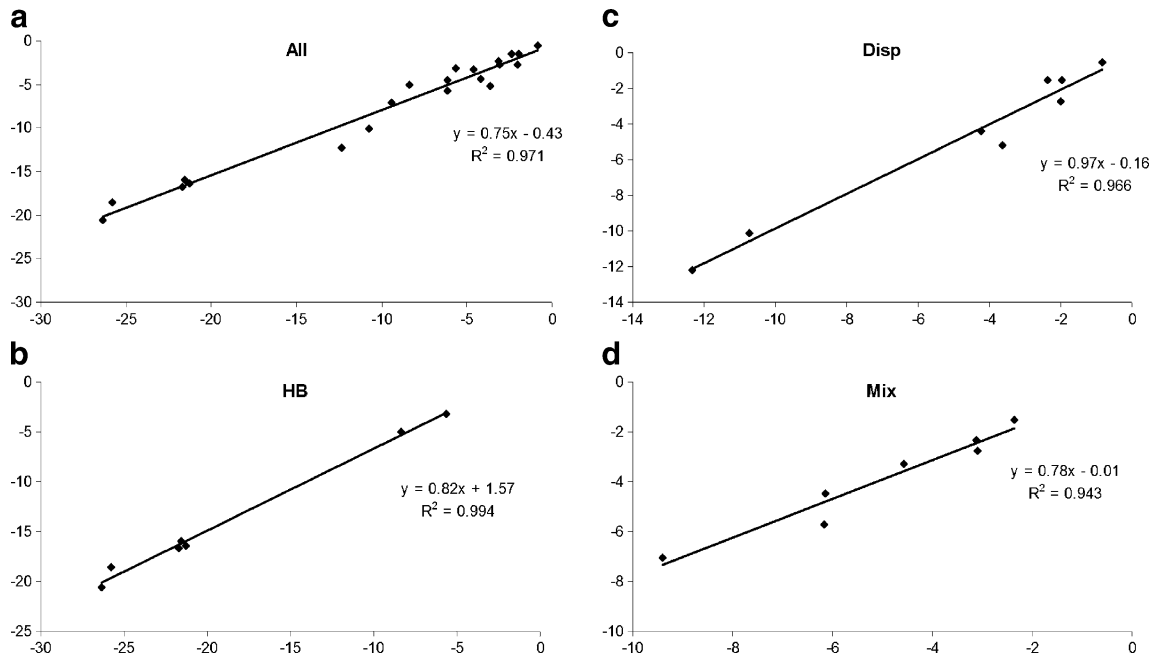

Fig. 1 Plot of BHandH/6-31+G(d,p) (abscissa) vs. literature CBS(T) (ordinate) binding energies (kcal mol^{-1}) for the S22 set of model complexes

different classes of complex. The best fit for dispersion bound complexes is very close to $y=x$, while for H-bonded and mixed complexes this is closer to that shown in Fig. 1. These fits give mean absolute errors of 0.60, 0.43, and 0.36 kcal mol^{-1} for dispersion, H-bonded, and mixed complexes, respectively.

In addition to direct calculation of interaction *via* the supermolecular approach, our first paper on BHandH [24] set out a surprisingly accurate correlation between interaction energy and the sum of the electron density at all bcp's found in the electron density of stacked complexes. This relation is summarized in Eq. (1)

$$\Delta E_{\pi} = -173.18 * \sum \rho_{\pi} - 0.02 \quad (1)$$

where ΔE_{π} is the interaction energy, in kcal mol^{-1}, and $\Sigma\rho_{\pi}$ is the sum of electron density at all bond critical points located between stacked molecules, in atomic units. This relation gave an r^2 value of 0.950 and a root mean square error of just 0.48 kcal mol^{-1}. Similar relations have long been established for hydrogen bonding, with the electron density at the H-bond bcp and/or the change in density at the donor bcp particularly widely used [29, 30].

However, the direct analogue of Eq. (1) would not be particularly useful, given the errors noted for H-bonds in Table 1. Instead, we have taken an alternative approach, whereby BHandH electron density data is plotted against MP2 interaction energy. This interaction energy was evaluated for 26 small and medium sized H-bonded complexes, using counterpoise corrected MP2/aug-cc-pVTZ methods. This gives rise to the equation shown as Eq. (2)

$$\Delta E_{HB} = -162.55 * \rho_{HB} + 0.44 \quad (2)$$

where units are the same as for Eq. (1). This correlation has r^2=0.98 and RMSE=1.07 kcal mol^{-1}. In this way, we aim to reproduce MP2 quality hydrogen bond energies from BHandH calculated electron densities, by identifying all intermolecular bcp's and applying the relevant linear correlation.

Data for application of these relations is reported in Table 3: at the literature geometry, agreement for dispersion bound complexes is reasonable, but less good for H-bonded and mixed complexes. Performance seems particularly poor for H-bonded complexes containing multiple H-bonds (*i.e.* all except ammonia and water dimers), which can be in error by almost 10 kcal mol^{-1}. Equation (2) was trained on simple complexes containing just one H-bond, and clearly struggles to describe these multiple H-bonded complexes. Table 3 also contains AIM analysis at the BHandH fully optimized geometry (see below for details), for which the performance is rather better, with MAE=1.80 kcal mol^{-1}, comparable to direct evaluation of interaction energy. This error is rather constant across the different types of complexes (1.63, 2.07 and 1.74 kcal mol^{-1}).

The similarity between the slopes of Eq. (1) and (2) are striking: indeed, within statistical error they are effectively identical. This suggests that the electron density properties of non-covalent interactions are similar, no matter the type of interaction involved. To the best of our knowledge, this

Table 3 Estimated ΔE using Eq. (1) and (2) (kcal mol^{-1}) for the S22 set of complexes

	CBS(T)	BHandH Geometry	Hobza Geometry
Ammonia dimer	-3.17	-2.32	-1.38
Water dimer	-5.02	-5.14	-3.42
Formic acid dimer	-18.61	-19.37	-14.26
Formamide dimer	-15.96	-12.46	-9.50
Uracil dimer	-20.65	-15.05	-11.14
Pyridoxine-aminopyridine	-16.71	-14.13	-10.14
Adenine-thymine WC	-16.37	-15.29	-10.28
Methane dimer	-0.53	-0.95	-0.48
Ethene dimer	-1.51	-2.60	-1.06
Benzene-methane	-1.50	-3.86	-2.42
Benzene dimer	-2.73	-2.44	-2.30
Pyrazine dimer	-4.42	-3.87	-3.40
Uracil dimer	-10.12	-7.13	-7.55
Indole-benzene	-5.22	-2.80	-4.58
Adenine-thymine	-12.23	-9.32	-10.66
Ethene-ethyne	-1.53	-1.11	-0.65
Benzene-water	-3.28	-1.91	-0.85
Benzene-ammonia	-2.35	-1.07	-0.61
Benzene-HCN	-4.46	-1.19	-0.83
Benzene dimer	-2.74	-1.75	-1.38
Indole-benzene T	-5.73	-1.79	-1.25
Phenol dimer	-7.05	-7.93	-4.86

possibility has not been noted before. Such a property would be highly desirable, since separation of critical points into H-bond and stacking classes is time consuming and can be ambiguous. We therefore attempted to combine models of this form, taking the S22 literature binding energy as a reference, and testing correlations similar to Eq. (1), (2) for these. Taking BHandH/6-31+G(d,p) calculated electron densities, a fit with R^2=0.926 and RMSE=1.78 kcal mol^{-1} was obtained, while the larger 6-311++G(d,p) basis gave marginal improvement to R^2=0.931 and RMSE=1.72 kcal mol^{-1}. The errors involved in fitting to the entire S22 dataset are therefore rather worse than for the individual sets used to train Eq. (1) and (2), but are comparable to those found for direct evaluation of interaction energy. As a further test, MP2/6-311++G(d,p) bcp properties were calculated, giving R^2=0.925 and RMSE=1.78 kcal mol^{-1}. Thus, it seems clear that neither method nor basis is crucial to calculated electron density properties, and that BHandH/6-31+G(d,p) is sufficient for such analysis.

To further investigate these correlations, we employed the H-bonded nucleic acid pairs of the rest of the JSCH-2005 database, which includes both experimental and optimized geometries. A trend was evident for this data, although the fit quality is not as good as for the S22 data (R^2=0.806, MAE=2.84 kcal mol^{-1}). Including only optimized geometries improved results, lowering the mean unsigned error by 0.4 kcal mol^{-1}. Closer examination revealed two distinct trends: one for nucleic acid bases paired by two H-bonds, and another for base pairs with three H-bonds. Separating these classes resulted in better correlation for the doubly H-bonded pairs (R^2 = 0.901, MAE=0.86 kcal mol^{-1}), although this was not the case for the base pairs with three H-bonds for which a rather poor fit was observed. However, there is indication that donor (D) – acceptor (A) patterns within the complexes with three H-bonds play a significant role in the resulting trends. Distinguishing the complexes in those following a DDA pattern and those having an ADA pattern resulted in separate fits of similar quality, but with significantly different slopes. From these results it seems that synergy in H-bond effects might play a role in the resulting binding energies, and hence that correlations between bcp properties and binding energies are family dependent.

As well as binding energies, the location of optimal geometries for complexes involved in non-covalent interactions is an important area, and one for which DFT offers several advantages. The speed of DFT over MP2+ΔCCSD (T) will rapidly come to the fore when many optimization steps are required. First and second derivatives of energy with respect to nuclear coordinates for DFT methods are

Table 4 Root mean square deviation (RMSD) from literature geometry and energy difference (|ΔE|) between literature and BHandH optimized geometries for the S22 set of model complexes

	Rigid monomer RMSD (Å)	Full opt RMSD (Å)	|ΔE| (kcal mol^{-1})
Ammonia dimer	0.11	0.11	0.58
Water dimer	0.08	0.07	0.46
Formic acid dimer	0.05	0.06	0.97
Formamide dimer	0.05	0.06	0.36
Uracil dimer	0.05	0.07	0.07
Pyridoxine-aminopyridine	0.05	0.20	0.76
Adenine-thymine WC	0.07	0.11	0.15
Methane dimer	0.20	0.20	0.35
Ethene dimer	0.13	0.13	0.43
Benzene-methane	0.07	0.07	0.10
Benzene dimer	0.08	0.04	0.07
Pyrazine dimer	0.07	0.05	0.07
Uracil dimer	0.11	0.13	0.03
Indole-benzene	0.14	0.23	0.26
Adenine-thymine stack	0.09	0.09	0.31
Ethene-ethyne	0.12	0.10	0.26
Benzene-water	0.09	0.26	0.54
Benzene-ammonia	0.08	0.10	0.24
Benzene-HCN	0.06	0.06	0.22
Benzene dimer	0.04	0.06	0.09
Indole-benzene T	0.05	0.09	0.23
Phenol dimer	0.67	0.59	0.84
Average	0.11	0.13	0.34

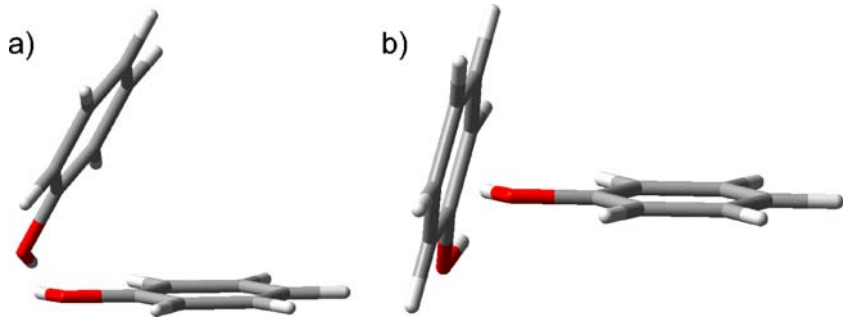

Fig. 2 Phenol dimer geometry from a) literature, and b) BHandH/6-311++G(d,p) optimization

available in most computational packages, so location and characterization of minima should be straightforward. Also, counterpoise corrections are generally much smaller in single-determinant Kohn-Sham DFT than in *ab initio* wavefunction theory: omission of counterpoise corrections from geometry optimisation would further simplify this procedure.

Two optimization strategies have been employed: firstly, a rigid monomer approach, in which both monomers are fixed at the geometry given by Hobza et al., and only the six intermolecular degrees of freedom allowed to vary. Subsequently, a fully flexible optimization including all intra- and intermolecular parameters was performed. The first set of data, reported in Table 4, indicates that in 14 of 22 cases, the optimized structure has an RMSD within 0.1 Å of the literature geometry, and in a further seven within 0.2 Å. Despite the clear overestimation of the strength of hydrogen bonding, it appears that BHandH gives a reasonable estimation of the geometry of H-bonded complexes, with an average RMSD of 0.06 Å for these six complexes. Only in one case, the phenol dimer, is a rather large change from literature data observed. This stems from a change in relative orientation of the rings, as shown in Fig. 2, in which the interplanar angle changes from 60.5° to 78.2°.

RMSD data for the full optimization of all geometrical parameters are also included in Table 4. In most cases, these are identical to or slightly larger than the rigid monomer values. One case stands out from this trend, namely benzene-water, in which the orientation of the water molecule alters to give two O—H...π H-bonds rather than one, as shown in Fig. 3. As in the frozen monomer case, the phenol dimer, shows a rather larger change of almost 0.6 Å, which represents a change in the relative orientation of the aromatic rings and formation of a C—H...π interaction that is not present in the literature geometry.

To check the energetic consequences of these geometry changes, DF-LMP2 calculations of binding energy were carried out at the literature and fully optimized geometries. Non-covalent interactions are generally associated with rather flat potential energy surfaces, such that relatively large changes in geometry can result in very small energy changes. Table 4 confirms that this is indeed the case, with an average difference in binding energy of 0.34 kcal mol^{-1} between literature and optimized geometries. In general, energy changes are larger for H-bonded complexes than for dispersion bound or mixed complexes, despite the smaller changes in geometry observed. The largest changes are seen for the dimers of formic acid, which is strongly H-bonded, and phenol, which undergoes the largest change in geometry during full optimization. Even these changes are less than 1 kcal mol^{-1}, and on average the change in energy is 0.48 kcal mol^{-1} for all six H-bonded complexes, compared with 0.20 and 0.35 kcal mol^{-1} for the dispersion bound and mixed complexes, respectively.

Initial frequency calculations were performed on BHandH/6-311++G(d,p) optimized geometries using the same method and basis set as for the optimization, with the "ultrafine" grid option for numerical integration selected. These showed that 14 complexes are true minima at this level, the remaining eight complexes having one or two imaginary frequencies. Re-optimization using tight SCF and force/displacement criteria removed imaginary frequencies for a further two complexes, namely the stacked indole-benzene and the benzene-HCN complexes. The number of imaginary frequencies after tight optimization

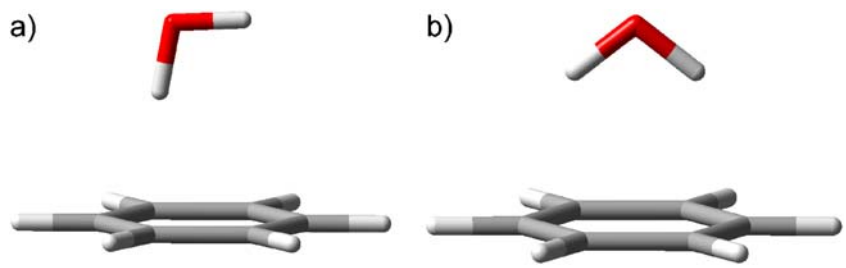

Fig. 3 Benzene-water geometry from a) literature, and b) BHandH/6-311++G(d,p) optimization

for each complex is shown in Table 5. The values of the imaginary frequencies present are generally small, ranging from 58*i* to 5*i* cm^{-1}, and are illustrated in Fig. 4. The H-bonded ammonia dimer has a single imaginary frequency (20.5*i* cm^{-1}), which corresponds to a "rocking" motion of both ammonia molecules that breaks C_{2h} symmetry and increases the linearity of the N—H...N hydrogen bond. Benzene-methane displays two degenerate imaginary frequencies at 26.4*i* cm^{-1}, both of which correspond to rotation of methane relative to benzene, moving C—H off the C_6 axis of benzene to remove the C_3 symmetry of the complex. The stacked uracil dimer has a single imaginary frequency (15.9*i* cm^{-1}), consisting of a "rocking" motion of each uracil to break the C_2 symmetry. Benzene-water and benzene-ammonia both have one imaginary frequency (11.3*i* cm^{-1} and 57.8*i* cm^{-1}, respectively). In the former, this is a rotation of H_2O away from C_s symmetry, while in the latter is similar to benzene-methane, moving N–H off the C_6 axis of benzene. The T-shaped benzene dimer displays two imaginary frequencies (28.6*i* and 5.1*i* cm^{-1}), both of which are rotations of one benzene relative to the other, moving the donor C–H off the C_2 axis of the C_{2v} complex.

This analysis cannot reveal whether the fact that six complexes are saddle points rather than true minima at this level is due to the inherent shortcomings of the BHandH functional, or to the symmetry of the literature geometry. It is notable that all complexes with imaginary frequencies do have some symmetry, and that in ref. [4] this is assumed with no further comment or test. Harmonic frequency calculation with suitably high-level *ab initio* methods is a difficult task, but the speed of DF-LMP2 makes this feasible, although the lack of analytical second derivatives means that numerical differentiation is required, limiting this to smaller complexes at this stage [43]. In the case of the ammonia dimer, DF-LMP2/aug-cc-pVTZ optimization and frequency confirms that the literature C_{2h} geometry is appropriate and hence that BHandH is in error, presumably due to its tendency to overestimate the strength of H-bonding. However, benzene-water is a first order saddle point with DF-LMP2 as well as with BHandH in the literature C_s geometry, with a similar normal mode of imaginary frequency corresponding to rotation of water off the symmetry plane of benzene. A more complete analysis is underway, and will be reported in a subsequent publication.

Table 5 Harmonic frequency analysis for the S22 set of model complexes

Ammonia dimer	#Imag Freq	ΔZPVE * (kcal mol^{-1})
	1	1.77 (1.70)
Water dimer	0	2.51 (2.37)
Formic acid dimer	0	1.67
Formamide dimer	0	2.27
Uracil dimer	0	1.06
2-Pyridoxine…2-aminopyridine	0	0.89
Adenine-thymine WC	0	0.92
Methane dimer	0	0.61 (1.16)
Ethene dimer	0	0.98
Benzene-methane	2	0.54
Benzene dimer stack	0	0.18
Pyrazine dimer	0	0.52
Uracil dimer stack	1	0.97
Indole-benzene stack	0	0.33
Adenine-thymine stack	0	1.12
Ethene-ethyne	0	0.60 (0.71)
Benzene-water	1	0.74 (1.06)
Benzene-ammonia	1	0.70
Benzene-HCN	0	0.60
Benzene dimer T	2	0.34
Indole-benzene T	0	0.43
Phenol dimer	0	1.54

*BHandH/6-311++G(d,p) values at minima following re-optimization, DF-LMP2/aug-cc-pVTZ values in parenthesis where available

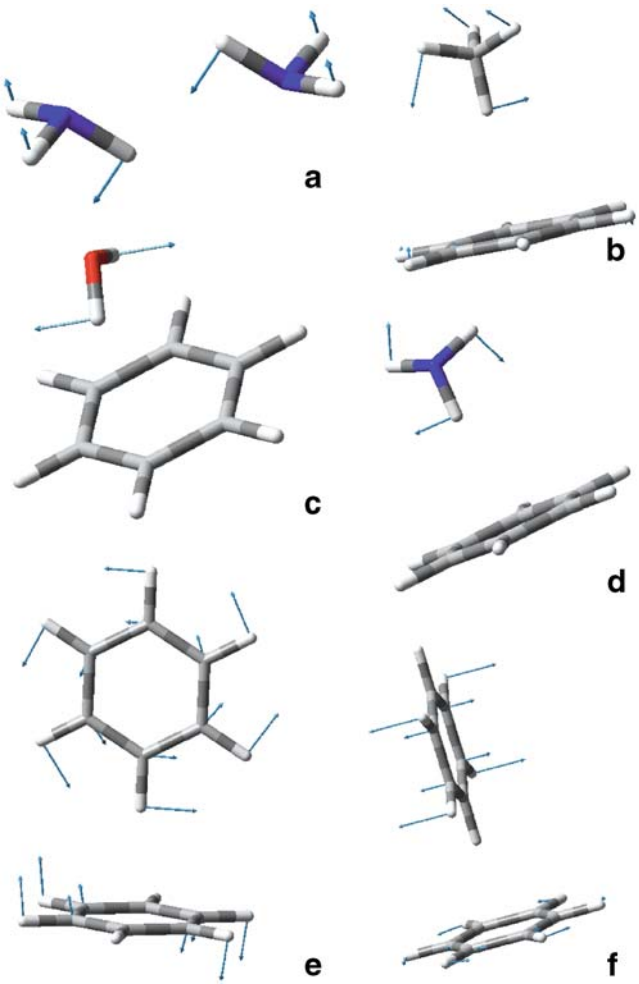

Fig. 4 Displacement vectors for normal modes corresponding to imaginary frequencies after tight geometry optimization

Following this analysis, each complex with imaginary frequencies at the BHandH level was perturbed along one of the associated normal modes, then re-optimised and harmonic frequencies checked again to ensure a true minimum (at this level) was located. Changes are generally very small, reflecting the very flat nature of the potential energy surfaces associated with non-covalent interactions. In most cases, the difference in energy between the high symmetry saddle points and lower symmetry minima is less than 0.1 kcal mol^{-1}. However, in the T-shaped benzene dimer a slightly larger change is evident, with a reduction of 0.23 kcal mol^{-1} on moving to the lower symmetry case. This is accompanied by a significant re-orientation of the complex from C_{2v} to C_s symmetry, as shown in Fig. 5. This aspect of the benzene dimer's structure has been noted before using *ab initio* and DFT-SAPT (symmetry adapted perturbation theory) methods, and indeed Wang et al. [44, 45] employed BHandH to check for blue-shifts in C–H stretches on dimer formation.

Table 5 also reports the calculated change in zero-point vibrational energy accompanied by complex to the binding energies (ΔZPVE). For those compounds initially with imaginary frequencies, this data was calculated at the re-optimized geometry. This data shows that ΔZPVE is small for the dispersion bound complexes, values rising above 1 kcal mol^{-1} only for stacked adenine-thymine, whereas ΔZPVE values are rather larger for the hydrogen bonded complexes. This is to be expected, since the force constants for displacement of hydrogen bonds are generally larger than for displacement of stacking interactions. Within the mixed complexes values are small except for the phenol dimer, which contains a relatively strong O–H...O H-bond. For selected complexes, the zero-point vibrational energy corrections were also calculated using the DF-LMP2 method. These are reported in parenthesis in Table 5, and where comparison is possible the agreement is generally excellent.

As an independent test of the conclusions drawn on the basis of Hobza's S22 dataset, we checked the performance of BHandH and related methods for the toluene dimer

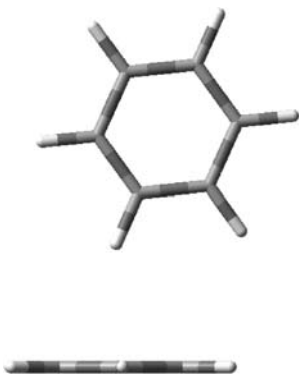

Fig. 5 Re-optimized geometry of "T-shaped" benzene dimer

Table 6 Binding energy for three orientations of toluene dimer (kcal mol^{-1})

	CCSD(T)	BHandH unscaled	BHandH scaled	AIM[a]	AIM[b]
Parallel	-2.71	-3.79	-3.27	-2.18	-1.27
Antiparallel	-3.47	-5.40	-4.48	-2.69	-1.68
Cross	-3.24	-4.58	-3.87	-2.91	-1.25

[a] At literature CCSD(T) geometry;
[b] At BHandH optimized geometry

(Table 6). Structures and binding energies for this complex in three distinct orientations, termed parallel, antiparallel, and cross, were reported by Wright [46], using CCSD(T) methods. At the geometries reported in ref. [31], BHandH overestimates binding, with an average absolute error of 1.45 kcal mol^{-1}. Scaling these binding energies following Fig. 1 leads to rather better predictions (MAE=0.73 kcal mol^{-1}). Importantly, BHandH reproduces the relative energy of the three forms, indicating that "antiparallel" is the most stable form, following by the "cross" form, with the "parallel" orientation the least stable. Optimization using BHandH again leads to small changes in structure (RMSD=0.13, 0.21 and 0.14, respectively) and in energy (0.21, 0.12, and 0.19 kcal mol^{-1}, respectively using DF-LMP2). AIM results are less accurate, underestimating binding throughout and, at the literature geometry, ordering the stability of complexes incorrectly.

Finally, recent work has shown that most common DFT functionals fail to recover the relative stability of linear and branched alkanes, a shortcoming that was ascribed to the poor description of weak interactions, most notably dispersion, in such functionals [47]. As a simple test of the performance of BHandH in this regard, geometries of *n*-octane and 2,2,3,3-tetramethylbutane were optimized using BHandH with both 6-31+G(d,p) and 6-311++G(d,p) basis sets. At these minima, possessing C_{2h} and D_{3d} symmetry, respectively, the branched isomer is 2.06 (with the smaller basis set) and 2.93 (larger basis) kcal mol^{-1} more stable than the linear one, compared with an experimental value of 1.9±0.5 kcal mol^{-1}. We note that this is in stark contrast to all other functionals considered in ref. [32], which predict the linear form to be the more stable one, with errors compared to experiment and/or *ab initio* data of between 5 and 12 kcal mol^{-1}.

Conclusions

We have tested the performance of a simple hybrid density functional for description of non-covalent interactions, using a database of high-level *ab initio* results as the principal benchmark. As has been noted before,

BHandH consistently overestimates the strengths of hydrogen bonds, but performs rather better for dispersion bound and "mixed" complexes. The errors associated with this method can be ameliorated somewhat by scaling of calculated interaction energies, but cannot be improved overall by altering the make-up of the exchange-correlation functional. We note that all tests presented here are for dimers; it is possible, likely even, that the errors associated with approximate DFT methods such as BHandH may accumulate for larger systems, such as trimers or tetramers. In the absence of reliable benchmark data for these, this cannot be tested, but should be borne in mind for applications of such methods. Empirical relationships between electron density properties and interaction energies were tested using the same database, and resulted in similar overall accuracy. Patterns for complexes containing multiple hydrogen bonds were also identified, with markedly different relations depending on whether donor and acceptor atoms are located in the same or different molecules.

The utility of the BHandH functional as a practical method for geometry optimization and harmonic frequency calculation was also tested, and shown to give reasonable results in almost all cases, with an average root mean square deviation from literature geometries of 0.11 Å. The difference in energy between literature and BHandH optimized geometries was tested using *ab initio* DF-LMP2 methods, and found to be less than 1 kcal mol^{-1} in all cases, and 0.34 kcal mol^{-1} on average. In six out of 22 cases, the optimized geometries that result were found not to be true minima if the point group symmetry of literature structures is conserved. For one complex, this seems to be an artifact of the shortcomings of BHandH, but for another two this is supported either by literature studies or by *ab initio* data: tests for the remaining complexes are ongoing. Such analysis also allows us to simply calculate the zero-point correction to interaction energies, an important point for comparison with experiment. Finally, the same methods were tested against high-level data for three isomers of the toluene dimer, and for two isomers of octane, confirming the overall conclusions gleaned from the larger database.

References

1. See for example Meyer EA, Castellano RK, Diedrich F (2003) Angew Chem Int Ed42:1210
2. Šponer J, Riley KE, Hobza P (2008) Phys Chem Chem Phys 10:2595
3. Dunning TH (1989) J Chem Phys 90:1007
4. Jurečka P, Šponer J, Černy J, Hobza P (2006) Phys Chem Chem Phys 8:1985
5. Vahtras O, Almlöf J, Feyereisen MW (1993) Chem Phys Lett 213:514
6. Hättig C, Weigend F (2000) J Chem Phys 113:5154
7. Schütz M, Manby FR (2003) Phys Chem Chem Phys 5:3349
8. Pulay P (1983) Chem Phys Lett 100:151
9. Saebø S, Pulay P (1993) Annu Rev Phys Chem 44:213
10. Hampel C, Werner HJ (1996) J Chem Phys 104:6286
11. Schütz M, Hetzer G, Werner HJ (1999) J Chem Phys 111:5691
12. Hetzer G, Schütz M, Stoll H, Werner HJ (2000) J Chem Phys 113:9443
13. Hill JG, Platts JA, Werner HJ (2006) Phys Chem Chem Phys 8:4072
14. Grimme S (2003) J Chem Phys 118:9095
15. Hill JG, Platts JA (2007) J Chem Theor Comput 3:80
16. Johnson ER, Wolkow RA, DiLabio GA (2004) Chem Phys Lett 394:334
17. Grimme S (2004) J Comput Chem 25:1463
18. Grimme S (2006) J Comput Chem 27:1787
19. Jurečka P, Černy J, Hobza P, Salahub DR (2007) J Comput Chem 28:555
20. Zhao Y, Truhlar DG (2008) Theor Chem Acc 120:215
21. Zheng JJ, Zhao Y, Truhlar DG (2007) J Chem Theory Comput 3:569
22. Swart M, van der Wijst T, Fonseca Guerra C, Bickelhaupt FM (2007) J Mol Model 13:1245
23. Becke AD (1993) J Chem Phys 98:1372
24. Waller MP, Robertazzi A, Platts JA, Hibbs DE, Williams PA (2006) J Comput Chem 27:491
25. Fonseca Guerra C, Bickelhaupt FM, Snijders JG, Baerends EJ (2000) J Am Chem Soc 122:4117
26. van der Wijst T, Fonseca Guerra C, Swart M, Bickelhaupt FM (2006) Chem Phys Lett 426:415
27. Bader RFW (1990) Atoms in Molecules, A Quantum Theory. Clarendon Press, Oxford
28. Popelier PL (2000) Atoms in Molecules. An Introduction. Prentice Hall
29. Boyd RJ, Choi SJ (1985) Chem Phys Lett 120:80
30. Carroll MT, Bader RFW (1988) Mol Phys 65:695
31. Gaussian 03, Revision C.02, Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C and Pople JA, Gaussian, Inc., Wallingford CT, 2004
32. MOLPRO, version 2006.4, a package of ab initio programs designed by Werner HJ, Knowles PJ, Lindh R, Manby FR, Schütz M, Celani P, Korana T, Mitrushenkov A, Rauhut G, Adler TB, Amos RD, Bernhardsson A, Berning A et al.
33. Boys SF, Bernardi F (1970) Mol Phys 19:553
34. Hehre WJ, Ditchfie R, Pople JA (1972) J Chem Phys 56:2257
35. Frisch MJ, Pople JA, Binkley JS (1984) J Chem Phys 80:3265
36. http://www.chemcraftprog.com
37. Weigend F, Köhn A, Hättig C (2002) J Chem Phys 116:3175
38. Weigend F (2002) Phys Chem Chem Phys 4:4285
39. Pipek J, Mezey PG (1989) J Chem Phys 90:4916
40. Boughton JW, Pulay PJ (1993) J Comput Chem 14:736
41. Biegler-König F, Schönbohm J (2002) J Comp Chem 23:1489

42. Marchetti O, Werner HJ (2008) Phys Chem Chem Phys 10:3400–3409
43. DF-LMP2/aug-cc-pVTZ optimisation and frequency calculations, commencing from the literature geometry, were performed for smaller complexes. Analogous calculations for the remaining complexes are in progress, and will reported: Hill JG, Platts JA, *in preparation*
44. Podeszwa R, Bukowski R, Szalewicz K (2006) J Phys Chem A 110:10345
45. Wang W, Pitoňák M, Hobza P (2007) Chem Phys Chem 8:2107
46. Rogers DM, Hirst JD, Lee EPF, Wright TG (2006) Chem Phys Lett 427:410
47. Grimme S (2006) Angew Chem Int Ed Engl 45:4460