PCCP

Testing the quality of some recent water–water potentials†

J. G. C. M. van Duijneveldt-van de Rijdt,[a] W. T. M. Mooij‡[b] and F. B. van Duijneveldt[a]

[a] *Debye Institute, Theoretical Chemistry Group, Padualaan 14, Utrecht University, 3584 CH, Utrecht, The Netherlands*
[b] *Bijvoet Institute, Structural and Crystal Chemistry Group, Padualaan 8, Utrecht University, 3584 CH, Utrecht, The Netherlands*

Received 31st May 2002, Accepted 20th January 2003
First published as an Advance Article on the web 6th February 2003

Accurate counterpoise-corrected SCF + MP2 total and partitioned interaction energies are reported at ten stationary points on the water-dimer potential energy surface. For further analysis, separate energy components were obtained using symmetry-adapted perturbation theory (SAPT). An interaction optimized basis set (IOM) of 136 functions was employed. At the global equilibrium geometry the binding energy for this basis is 1 kJ mol^{-1} less than the value at the basis set limit, but the relative energies at the remaining stationary geometries are much more precise. The new IOM data have been used to test the quality of several water–water potentials including SIBFA, ASP-W4, TAP, SAPT-5s and SAPT-5st. Focussing on the relative energies with respect to that of the global minimum, the r.m.s. deviations from the IOM data are found to vary from 0.4 to 3.5 kJ mol^{-1}, those of the SAPT-5s(t) potentials being much the smallest. Inspection of the separate components of the SAPT-5s(t) potentials shows, however, that their components do not properly reflect the physics of the interaction, due to shortcomings in the chosen functional forms. It is suggested that a combination of successful ingredients of the various potentials may yield a water potential that is fairly simple, transferable to larger water clusters and yet more accurate than those currently available.

1. Introduction

More analytical potentials have probably been developed for water than for any other system and, as we will show, the accuracy of some recent potentials is remarkable. Nevertheless, there is still room for improvement. Let us first analyze the present situation.

The long range part of the potentials is usually modelled using monomer properties. These are now known very precisely, and the absolute and relative errors in this part of the potential can be kept very small. The short range part is much less well known. This is where the wavefunctions of different molecules overlap significantly, giving rise to penetration terms in the long-range components of the interaction energy, as well as to exchange energies arising from the Pauli principle. To guide model development in this region, some authors use water dimer *ab initio* interaction energies or components thereof. Other authors resort to the use of experimental data such as spectral information or virial coefficients.

Clearly there is a large variety of design strategies, and the need arises to compare the quality of potentials in some objective manner. It was recognized by Millot and Stone[1] that the set of stationary points on the water dimer potential energy surface constitutes a convenient benchmark for such comparisons, and they employed the ten stationary points identified in SCF + MPn calculations by Smith *et al.*[2] for this purpose. These show a variety of O–H···O contacts: open, cyclic, bifurcated, and multiply hydrogen bonded. All structures represent minima along the R coordinate which separates the molecules, and so provide information on the short-range overlap effects. Whether we are dealing with the global minimum (Hessian index $N = 0$), a saddle point ($N = 1$) or a higher order stationary point is determined entirely by the mutual orientations of the molecules and so the anisotropy of the short-range effects is sampled as well.

The structures 2, 4 and 9, which are the saddle points on the surface, govern the dynamical processes in a dimer as well as in larger water clusters. They characterize the lowest-energy paths for interconversions among the eight equivalent $N = 0$ structures and their barrier heights determine the relative ease of these interconversions.[2] Tunneling processes occur along these paths, giving rise to the tunneling splittings observed in recent years in the rotational–vibrational-tunneling (VRT) spectra of the dimer[3] and this information can be used to tune a potential if one desires to describe such spectra accurately.

Since the value of kT at room temperature corresponds to 2.5 kJ mol^{-1} it seems reasonable to require that the errors in relative energies produced by analytical model potentials do not exceed 1 kJ mol^{-1}. In extensive comparisons of a large number of analytical potentials with Smith's data, Millot and coworkers[4] found errors in most of the then available potentials much larger than this. Indeed, the number of stationary points and their Hessian indices differed for some model potentials and even the potentials derived on the basis of *ab initio* calculations were not consistent with each other. It was not clear whether the potentials or the *ab initio* calculations were to blame for the discrepancies. Three different basis sets were used in Smith's *ab initio* work, but only the smaller two (*viz.* 6-31+G(d,p) and 6-31+G(2d,2p)) were applied at all ten geometries. In the latter calculations the basis set superposition error (BSSE) was not corrected for, even though the (presumably smaller) BSSE in the third and largest basis set was found to affect the energy differences between different stationary points. Thus this study was not systematic enough to provide a reliable set of relative energies for the stationary points.[4,5]

† Electronic supplementary information (ESI) available: Cartesian coordinates for the ten stationary point geometries of the water dimer. See http://www.rsc.org/suppdata/cp/b2/b205307a/

‡ Present address: Astex Technology, 250 Cambridge Science Park, Milton Road, Cambridge CB4 0WE, UK.

This journal is © The Owner Societies 2003

Our aims in the present study are as follows. The first is to present more accurate SCF + MP2 *ab initio* interaction energies at Smith's geometries. In addition, to analyze the weak and strong points of the analytical potentials we present accurate *ab initio* results for separate components of the interaction energy. Unlike the total energies, these separate energy components may depend strongly upon the precise geometries used for the various stationary points, and this was the reason for using a fixed set of geometries (*viz.* those of Smith *et al.*) rather than the optimum stationary geometries corresponding to a given potential model.

Next, on the basis of these data we assess the accuracy of some recent water–water potentials, most of which were not covered in the study by Millot *et al.*, *viz.* SIBFA,[6–8] ASP-W4,[4] SW,[5] TAP,[9] VRT-ASPW,[10] SAPT-5s[11] and SAPT-5st.[12] The *ab initio* results that we present employed Mooij's interaction optimized basis set labelled IOM. Its composition is O: 5s3p3d2f1g; H: 2s2p. This basis was chosen for a variety of reasons. Firstly, this basis was optimized by Mooij *et al.*[9] for the methanol dimer and then used to derive the TAP potential. Hence, by comparing the TAP data and the IOM *ab initio* data for the water dimer we have a direct check of the accuracy and the transferability of the TAP model. Secondly, a comparison of water dimer binding energies at the global minimum for a series of basis sets of increasing accuracy showed[9] that the IOM results are of the same accuracy as those obtained using the cc-pVQZ and the aug-cc-pVTZ basis sets, which are significantly larger than IOM. In particular, all three basis sets yield SCF interaction energies within 0.2 kJ mol^{-1} from the Hartree–Fock limit, and all three yield MP2 interaction energies that are about 1.0 kJ mol^{-1} above the MP2-limit, the most important shortcoming of these sets being that they underestimate the dispersion part of the MP2 energy by about 10%. Since this underestimation will occur for all orientations the relative energies of the stationary points will be little affected by this error. We return to this point in section 3 where we will show that the average error in MP2/IOM relative energies compared to the best *ab initio* data currently available[13,14] is only 0.2 kJ mol^{-1}. We conclude that the IOM results are adequate as a reference to test the quality of analytical model potentials. Finally, the small size of the IOM basis was an advantage in subsequent work[14] in which we performed counterpoise-corrected SCF + MP2 geometry optimizations for the water dimer stationary points.

After completion of the present study, Burnham and Xantheas[15] published a paper which complements our work. These authors compared some analytical water–water potentials, including the ASP potentials, to MP2 *ab initio* energies in the aug-cc-pVTZ basis. This was done by mapping the optimal energies and structures for dimers restricted to C_s, C_2, C_i or C_{2v} symmetry as a function of the intermolecular O–O separation. The ASP family of potentials was found to fit the *ab initio* data best. Their *ab initio* results at the stationary points within each symmetry will be compared to the present data in section 3.

2. *Ab initio* methods employed in this work

Geometries for the stationary points 1–10 (*cf.* Fig. 1) were taken from Smith *et al.*[2] However, instead of using their optimized monomer geometries, which differ for each stationary point, fixed experimental monomer geomtries were used throughout (OH 0.9572 Å and HOH 104.52°). The same choice is usually made in defining the analytical potentials. Cartesian coordinates defining the present geometries are available as ESI.†

In discussing our results we employ the following notations. Starting the dimer SCF calculation from monomer SCF wavefunctions one obtains the zeroth iteration Heitler–London total dimer energy. Subtracting the monomer energies from this one obtains the Heitler–London interaction energy ΔE^{HL}. This may be partitioned into a coulomb and an exchange repulsion part by separately calculating the electrostatic energy E_{coul} between the monomer SCF charge distributions:

$$\Delta E^{HL} = E_{coul} + E_{exch}$$

Next, iterating the dimer to self-consistency one obtains the delocalization energy ΔE^{deloc} which may be partitioned in induction and charge transfer parts according to

$$\Delta E^{deloc} = E_{ind} + E_{ct}$$

by separately calculating an induction energy E_{ind}.

The total SCF interaction energy is then

$$\Delta E^{SCF} = \Delta E^{HL} + \Delta E^{deloc}$$

Treating electron correlation at the MP2 level gives an extra interaction energy, ΔE^{MP2}, which includes a dispersion energy, E_{disp}, as well as correlation corrections to all previous terms (E_{coul}, E_{exch}, E_{ind} and E_{ct}). The final total interaction energy becomes $\Delta E^{SCF+MP2}$ or ΔE for short.

The SAPT calculations on the separate components of the energy yield results such as $E^{12}_{pol,resp}$, where the superscripts refer to the order of this term in the intermolecular potential and in the intramolecular fluctuation potential, respectively. The subscript 'pol' refers to the coulomb term and 'resp' indicates that monomer orbital relaxation is taken into account.[16,17]

Various additional choices have to be made in actual *ab initio* calculations of the above quantities. In the present work the monomer wavefunctions needed to partition ΔE^{SCF} to an Heitler–London part and a delocalization part were obtained using the full basis set of the dimer ('DCBS'). This has the advantage that ΔE^{HL} is closer to the basis set limit than

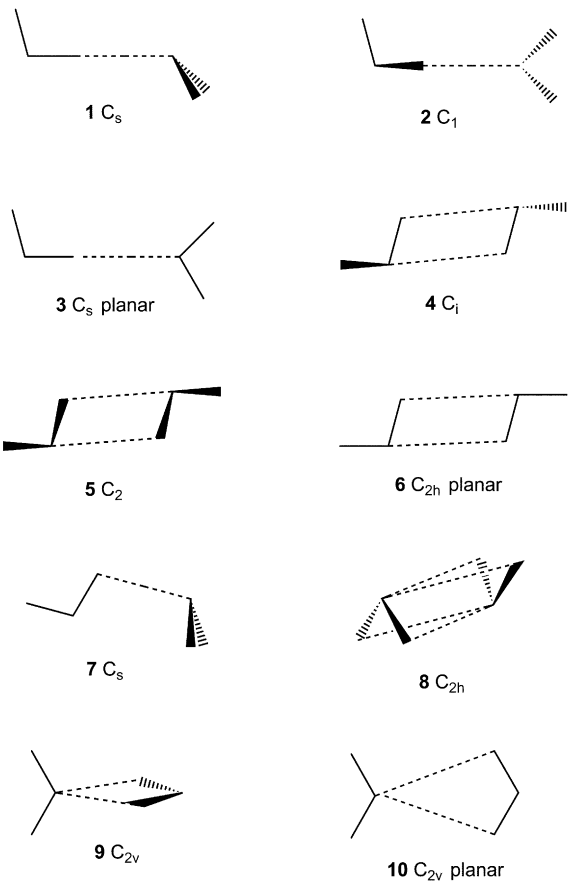

Fig. 1 Sketch of the ten water dimer stationary geometries as determined by Smith *et al.*[2]

when the monomers employ only their own basis set.[18] Moreover, by starting from DCBS descriptions, no BSSE enters ΔE^{HL} or ΔE^{deloc}, and no BSSE correction is necessary. The calculation of ΔE^{MP2} likewise employs monomer MP2 energies in DCBS, and again no BSSE correction is necessary. That is, we employ the counterpoise method which we denote by the label CP. In a few cases we have, in addition, generated interaction energies uncorrected for BSSE ('noCP'). Here the monomer energies were calculated using their own basis set only.

The present calculations were carried out using the ATMOL system of programs,[19] including its local extensions SERVEC for vector manipulations[20] and INTACAT for MP2 and E^{20}_{disp} dispersion energy calculations[21] (*cf.* section 6). The TURTLE program[22] was used to calculate the E^{SMMO}_{ind} induction energies of section 7. The SAPT96 program[23] was used to evaluate the SCF level and correlated components of the interaction energy. Except where noted otherwise, basis sets with five d, seven f, nine g sets of polarization functions were employed and no frozen-core options were invoked.

3. *Ab initio* results at the stationary geometries

The *ab initio* results computed in this work are summarized in Table 1. We started by recalculating Smith's 6-31+G(d,p) data using the present geometrical choices. The total ΔEs are given in column 2. As in Smith *et al.*, these data are not corrected for BSSE (noCP). To remedy this, we provide the corresponding counterpoise corrected data in column 3 (CP). To show the role of using a slightly larger basis set we quote in column 4 the results obtained by Smith *et al.* at the MP4/6-31 + G(2d,2p) level (no counterpoise). (No MP2 results were reported for this basis set, but the difference between ΔE^{MP4} and ΔE^{MP2} is small for the water dimer). Finally, in columns 5–8 we give the total ΔE for the IOM basis set, as well as some partitioned energies that we will need in later sections. For completeness, column 9 lists the correlation contribution E^{corr}_{SAPT} to the SAPT interaction energy $E_{int} = \Delta E^{SCF} + E^{corr}_{SAPT}$ as it is provided by the SAPT96 program.

The energy at the global minimum is significantly different in columns 2–5. However, for most applications in which one employs analytical model potentials only relative energies are important. We therefore focus on the relative energies with respect to that of the global minimum, *i.e.* in each column we take the energy of the global minimum as reference energy. Root mean square deviations (rmsd) between the relative energies in a given column and those in the IOM column are given in the lower lines of Table 1.

The results confirm that Smith's data are not precise enough. Although the rmsd is reduced by applying the counterpoise method (column 3) or by using a larger set of polarization functions (column 4) the resulting rmsd of 1.1 kJ mol^{-1} are still not good enough to use these energies as a reference for testing model potentials, because, as we will show below, some of the newer model potentials are at least as accurate.

The high accuracy of some recent water–water models raises the question whether the present MP2/IOM data are accurate enough to assess the quality of these models. In Table 2 we therefore compare the MP2/IOM data to the best *ab initio* data that were available to us. The data are partly taken from a forthcoming paper on the geometries of the water dimer stationary points,[14] since this paper reports IOM results and 'best' results at the same geometries. In addition to total ΔE at geometry 1, calculated by a variety of methods, Table 2 lists relative energies at geometries 2–10, taking the ΔE at geometry 1 as reference. Columns 2 and 3 show CP and noCP MP2 energies in the IOM basis set, using geometries optimized at the CP MP2/IOM level of theory.[14] The latter geometries were also used in column 4, which shows the results of MP2-R12 calculations, and in column 5 which lists the 'best *ab initio*' data from ref. 14, based on a combination of the MP2-R12 data and CCSD(T) results in the IOM basis set. Similar 'best *ab initio*' data (but at differently optimized geometries) have recently been reported by Tschumper *et al.*[13] and these are shown in column 6. The table concludes with the noCP MP2/aug-cc-pVTZ results recently obtained by Burnham and Xantheas.[15]

Let us first consider our choice to use the counterpoise procedure for calculating MP2 interaction energies when using the IOM basis set. There is evidence that CP and noCP MP2 interaction energies calculated for hydrogen-bonded systems in the aug-cc-pVXZ (X = TZ through 6Z) series of basis sets lie about equally far from the MP2 complete basis set (CBS) limit.[24] Taking the agreement with CBS as the criterion of accuracy it was concluded that removal of BSSE by applying the counterpoise principle does not bring about an increase of accuracy and so may as well be omitted in this class of applications. This conclusion was reinforced by the observation that for the aug-cc-pVDZ basis, the smallest member of the series, the noCP result is often markedly closer to CBS than the CP result.

Table 1 *Ab initio* SCF + MP*n* interaction energies (kJ mol^{-1}) in several basis sets at the ten stationary points of the water dimer;[a] the rmsd and maximum deviations given in the lower lines refer to the relative energies for geometries 2–10, using the entry for geometry 1 as reference

	MP2 6-31 + G(d,p) noCP	MP2 6-31 + G(d,p) CP	MP4 6-31 + G(2d,2p) noCP	MP2 IOM CP	Partitioning of MP2(IOM)			
					ΔE^{HL}	ΔE^{deloc}	ΔE^{MP2}	E^{corr}_{SAPT} [b]
1	−25.97	−19.53	−22.42	−19.51	−5.40	−9.56	−4.55	−4.96
2	−23.26	−17.53	−19.62	−17.53	−5.30	−8.11	−4.12	−4.49
3	−23.19	−17.71	−19.32	−17.42	−5.63	−7.85	−3.94	−4.28
4	−21.00	−17.07	−18.24	−16.50	−6.36	−4.69	−5.45	−5.87
5	−19.71	−16.22	−16.90	−15.66	−6.22	−4.04	−5.40	−5.62
6	−19.58	−16.27	−16.69	−15.53	−6.41	−3.81	−5.31	−5.49
7	−17.16	−12.80	−14.72	−12.29	−6.24	−2.31	−3.75	−4.75
8	− 7.45	− 4.40	− 6.73	−4.61	−2.46	−0.53	−1.63	−2.45
9	−19.06	−13.49	−14.93	−11.40	−5.58	−2.82	−2.99	−4.07
10	−13.83	−10.20	−10.54	−8.10	−4.76	−1.63	−1.71	−2.56
Rmsd (rel)	2.0	1.1	1.1	[0.0]				
Max. dev.	3.6	2.1	1.8	[0.0]				

[a] The MP4 results are frozen-core data taken from ref. 2. They refer to the geometries as determined by Smith *et al.* All other entries were calculated in the present work, correlating all electrons and employing fixed monomer geometries as described in the text. [b] $E^{corr}_{SAPT} = \varepsilon^{1}_{pol,CCSD} + \varepsilon^{1}_{exch}(CCSD) + {}^{t}E^{22}_{ind} + {}^{t}E^{22}_{exch-ind} + E^{20}_{disp} + E^{20}_{exch,disp} + \varepsilon^{2}_{disp}(2)$.

Table 2 Comparison of the *ab initio* MP2 interaction energies (kJ mol^{-1}) in the IOM basis set at the ten stationary points of the water dimer to the best *ab initio* results available. For geometries 2–10 the energies relative to those at geometry 1 are shown, and the rmsd refer to these energies only; high-level optimized geometries were used, as discussed in the papers quoted

	MP2 IOM noCP Ref. 14	MP2 IOM CP Ref. 14	MP2-R12 Ref. 14	Best *ab initio* CP Ref. 14	Best *ab initio* noCP Ref. 13	MP2 aug-cc-pVTZ noCP Ref. 15
1	−22.93	−19.66	−20.52	−20.65	−20.88	−21.67[a]
2	2.28	1.98	2.09	2.18	2.17	–
3	2.50	2.15	2.26	2.37	2.37	–
4	3.72	2.96	2.95	3.10	2.93	3.47
5	4.60	3.84	3.86	3.95	3.98	3.69
6	4.69	3.99	3.99	4.14	4.16	–
7	9.23	7.82	8.12	7.78	7.59	–
8	16.55	14.67	15.26	14.87	14.94	–
9	9.42	7.72	8.05	7.72	7.48	8.37
10	13.32	11.27	11.75	11.47	11.34	–
Rmsd (rel)	1.0[b]	0.3[b], 0.2[c], 0.2[d]				
Max.dev.	1.6[b]	0.6[b], 0.2[c], 0.3[d]				

[a] The corresponding CP corrected result is −19.71 kJ mol^{-1}.[24] [b] Deviations from the MP2-R12 data in column 4. [c] Deviations from the best *ab initio* data in column 5. [d] Deviations from the best *ab initio* data in column 6.

Turning now to the present calculations, and taking the MP2-R12 results in column 4 to represent the MP2 CBS limit one notes that the CP MP2/IOM result at the global minimum (column 3, line 1) lies 0.9 kJ mol^{-1} above the CBS limit while the noCP result (column 2, line 1) lies as much as 2.4 kJ mol^{-1} below it. Thus for the IOM basis set the interaction energy is distinctly improved by applying the counterpoise procedure.

A second (and more fundamental) reason for using CP-corrected interaction energies in the present paper is that, according to the counterpoise theorem,[25] they precisely represent the physics of the interaction, as rendered by the method and basis set that one employs and so they offer a more rigorous basis for judging the physics as described by an analytical model potential than interaction energies contaminated by BSSE.

The advantages of using counterpoise-corrected data become even more apparent when one looks at the relative interaction energies at geometries 2–10 which, after all, are the key data in this paper. Comparing the CP-corrected MP2 relative energies for the IOM basis set (column 3) to the MP2-R12 results in column 4 one notes a striking agreement. The rmsd between these energies is only 0.3 kJ mol^{-1}, and the maximum deviation is only 0.6 kJ mol^{-1}. These errors are smaller, as we shall see, than the errors in the analytical potentials that we want to examine. On the other hand, the errors between the noCP IOM results (column 2) and the MP2-R12 data are about three times larger (rmsd = 1.0 kJ mol^{-1} and max. deviation 1.6 kJ mol^{-1}) so these noCP data are about as poor as Smith's noCP MP4 data quoted in Table 1.

Comparing the CP IOM MP2 data (column 3) to the 'best *ab initio*' results from ref. 14 (column 5) one might have expected some deterioration of the agreement, since there is no reason why MP2 results should be uniformly close to data based on CCSD(T) results. However, the agreement actually becomes slightly better (rmsd = 0.2 kJ mol^{-1}, max. deviation 0.3 kJ mol^{-1}). Very similar errors are noted when comparing with the 'best *ab initio*' data reported by Tschumper *et al.*[13] (column 6). We conclude that our use of CP MP2/IOM data as benchmark data at Smith's geometries is justified.

An alternative set of benchmark data could have been the SAPT total interaction energies $E_{\rm int}$ which became available in the later stages of this work. These differ from the MP2/IOM ΔE by the differences between columns 8 and 9 of Table 1. The differences resemble the change that occurs when going from MP2 to CCSD(T) (*cf.* Table 2 columns 4 and 5), confirming the view[11] that the high-level SAPT results in Table 1, column 9 are roughly of MP4 quality. However, such an improvement in the correlation level without simultaneously increasing the quality of the basis actually deteriorates the agreement with 'best' *ab initio* such as those in columns 5 and 6 of Table 2 and so this alternative was not considered further.

The final column of Table 2 shows the relative energies at geometries 4, 5 and 9 which were recently obtained in noCP MP2/aug-cc-pVTZ calculations by Burnham and Xantheas.[15] Keeping in mind that geometrical differences prevent us from making a rigorous comparison one may tentatively conclude that these results are somewhat closer to the MP2-R12 data than the noCP data for the IOM basis set, but not as close as the CP IOM data. It may well be that for the aug-cc-pVTZ basis too, even though its BSSE is less than for the IOM basis, the accuracy of relative energies is markedly improved when the data are counterpoise corrected. Evidence for this comes from the study of Xantheas *et al.*[24] on the structures and energies of water clusters. They reported MP2 CBS relative energies for the cage, book and ring structures of the water hexamer, relative to that of the prism structure, of 0.03, 0.25 and 1.0 kcal mol^{-1}, respectively. The corresponding results for the aug-cc-pVTZ basis were mediocre (0.11, 0.47 and 1.57, respectively) when noCP data were used, but the CP data were accurate (0.10, 0.28 and 1.03). The CP aug-cc pVDZ relative energies were also accurate (0.13, 0.26 and 0.95 kcal mol^{-1}, respectively) while the noCP relative energies were poor. The basic reason for these CP data performing so well is probably that in comparing relative energies of different equilibrium structures or higher-order stationary points of a given system one is comparing the energies of structures which have similar dispersion energies. While underestimation of the dispersion energy is often one of the main errors in the MP2 interaction energy, this error will tend to cancel in calculating relative energies and so these may be quite accurate provided that BSSE (which apparently varies more randomly) is eliminated. Similar conclusions have been reached in a recent study of the barrier to rotation in the $C_2H_4\cdots SO_2$ complex.[26]

4. Model potentials considered

Of the many available water–water model potentials, several were already compared in Millot *et al.*[4] Of these, we here included only the ASP-W4 potential, which was among the best in their comparison. In addition we tested the SIBFA and SW potentials as well as some more recent ones. A brief description of the characteristics of these potentials is as follows.

SIBFA, 'Sum of interactions between fragments *ab initio* computed'. This is an all-atom transferable potential that has been parametrized for several types of atoms by fitting on *ab initio* (mostly SCF) energies. It was first presented in 1986[6] and has been repeatedly refined in recent years.[7,8] The SIBFA model has been applied to numerous H-bonded complexes and generally provides equilibrium geometries and equilibrium binding energies in good agreement with the corresponding *ab initio* data.

ASP, 'Anisotropic site potential'. This is a family of water–water potentials fitted on *ab initio* dimer energies. It was first presented in 1992[1] and refined more recently. We considered the most accurate version, ASP-W4.[4] The ASP-W4 model has been shown to yield equilibrium geometries and binding energies for water clusters in good agreement with *ab initio* data.[27] We also tested the VRT-ASPW potential. This is a variant of ASP-W4 in which the repulsion term was tuned to fit the water dimer VRT spectrum.[28]

SW, 'Systematic potential for water'. This is one of the 'systematic' potentials proposed by Wheatley.[5] Their aim is to achieve good accuracy while relying as little as possible on dimer *ab initio* energies.

TAP, 'Transferable *ab initio* intermolecular potential'. This all-atom potential was derived by Mooij *et al.*[9] by fitting to methanol dimer and trimer *ab initio* data. The TAP model has successfully been applied in the '*ab initio*' prediction of the crystal structures of alcohols and saccharides.[29] We here employ only the O and H parameters of this potential, in combination with MP2 atomic multipole moments specifically derived for water, as described in ref. 9.

SAPT-5s(t), 'Symmetry adapted perturbation theory potential with five distinct sites per molecule'. The SAPT-5s potential[11] was fitted on over 2000 water dimer energies calculated by SAPT theory. Vibrationally averaged monomer geometries were employed. This leads to a slight increase in the interaction energies, but we expect the relative energies at the stationary points to be insensitive to this. In the tuned version SAPT-5st the repulsion term was empirically adapted to fit the water dimer far-IR VRT spectrum.[12]

In each of these models, the interaction energy is a sum of several components

$$\Delta E = E_{\mathrm{coul}} + E_{\mathrm{rep}} + E_{\mathrm{ind}} + E_{\mathrm{ct}} + E_{\mathrm{disp}}$$

In the ASP-W4 and the SIBFA potentials each term is fitted separately, but in the TAP, SW, and the SAPT-5s(t) potentials the charge transfer effects are absorbed in the repulsion term. The size of the repulsion term is fixed by fitting to the total *ab initio* ΔE for TAP and SAPT5s(t) and to second virial coefficients for SW. In SAPT-5s(t) the induction and dispersion terms are fitted together as 'long-range' $= E_{\mathrm{ind}} + E_{\mathrm{disp}}$. There are numerous further differences between the models, for instance in the number of sites employed and in the quality of the *ab initio* data used in the fitting procedure. Some of these differences will be considered in more detail in the following sections, where we compare the separate energy components of the various models to suitable *ab initio* counterparts. Hartree–Fock level *ab initio* descriptions and correlated *ab initio* descriptions (calculated using SAPT) were generated for this purpose.

5. Coulomb energy

The first four columns of Table 3 list some *ab initio* coulomb energy results that may serve to judge the quality of the coulomb models in columns 5–8. The first column shows the E_{coul} results obtained by Gresh from partitioned SCF calculations using the CEP 4–31G(2d) basis set.[30] In the terminology of SAPT, these are E^{10}_{pol} values. The monomers were described using their own basis set ('MCBS'). The next column has E^{10}_{pol} values as well, but here the IOM basis was employed, and the monomer SCF wavefunctions were obtained using the full dimer basis ('DCBS'). (The use of DCBS orbitals is the preferred choice in SAPT *ab initio* calculations and we adopted it for all other SAPT calculations as well.) The third column includes a correction for intramolecular correlation (*viz.* $E^{12}_{\mathrm{pol,resp}} + E^{13}_{\mathrm{pol,resp}}$) which leads to a description of the coulomb energy compatible with supermolecular calculations at the MP3 level of theory.[31] Similarly, column 4 lists higher-level correlated results compatible with CCSD supermolecular calculations. In basis IOM, going from the SCF level to the correlated MP2 level the monomer dipole moment is reduced from 0.780 to 0.735 a.u. This reduces the dipole–dipole component of the MP3 coulomb energy by 12%, which is about +2 kJ mol^{-1} at the global minimum (geometry 1). The *ab initio* results in columns 2 and 3 show a difference of only about 1 kJ mol^{-1} for this geometry. This may indicate that in addition to the dipole change there is an increase of the intermolecular overlap at the correlated level causing an increase in the attractive penetration component of E_{coul} (*cf.* ref. 16). This is plausible since the first-order exchange energy, which also depends on intermolecular overlap, increases significantly when intramolecular correlation is accounted for (*cf.* Table 7, columns 8 and 9).

In Fig. 2 the various model coulomb energies are plotted *versus* the *ab initio* results of column 4. Since the ASP-W4, TAP and SIBFA models do not allow for penetration effects, no rigorous agreement with the *ab initio* energies is to be expected. Nevertheless, the correlation coefficients between these model data and all four *ab initio* columns are high (linear correlation coefficients $r = 0.997$ or higher. We employ trend lines that pass through the origin). If the multipole component

Table 3 Coulomb energies E_{coul} (kJ mol^{-1}) for the model potentials at the ten stationary points of the water dimer compared to SCF and correlated *ab initio* data calculated using SAPT

Ab initio[a]	*Ab initio*[b]	*Ab initio*[c]	*Ab initio*[d]	SIBFA	TAP	ASP-W4	SAPT-5s(t)
−32.87	−34.71	−33.81	−33.57	−28.07	−26.26	−27.19	−14.42
−27.28	−29.63	−28.51	−28.14	−23.01	−21.75	−22.42	−12.95
−26.40	−28.90	−27.62	−27.17	−22.13	−21.02	−21.61	−13.12
−27.49	−28.66	−28.17	−28.03	−23.01	−21.19	−21.85	−22.06
−24.10	−24.99	−24.17	−23.80	−20.04	−18.44	−19.37	−21.32
−23.05	−23.79	−22.77	−22.29	−19.04	−17.52	−18.52	−21.53
−19.92	−20.65	−20.60	−20.77	−16.44	−15.20	−15.57	−19.24
−5.77	−6.25	−6.33	−6.47	−5.06	−4.45	− 4.70	−10.06
−19.00	−20.82	−20.22	−20.12	−15.69	−14.31	−15.78	−13.91
−11.13	−12.32	−11.56	−11.28	−9.46	− 8.34	−9.56	−10.16

[a] E^{10}_{pol} calculated using the CEP4-31G(2d) basis set (MCBS).[30] [b] E^{10}_{pol} calculated using the IOM basis set (DCBS). [c] $E^{1}_{\mathrm{pol}} = E^{10}_{\mathrm{pol}} + E^{12}_{\mathrm{pol,resp}} + E^{13}_{\mathrm{pol,resp}}$ calculated using the IOM basis set (DCBS). [d] $E^{10}_{\mathrm{pol}} = E^{10}_{\mathrm{pol}} + \varepsilon^{1}_{\mathrm{pol,CCSD}}$ calculated using the IOM basis set (DCBS).

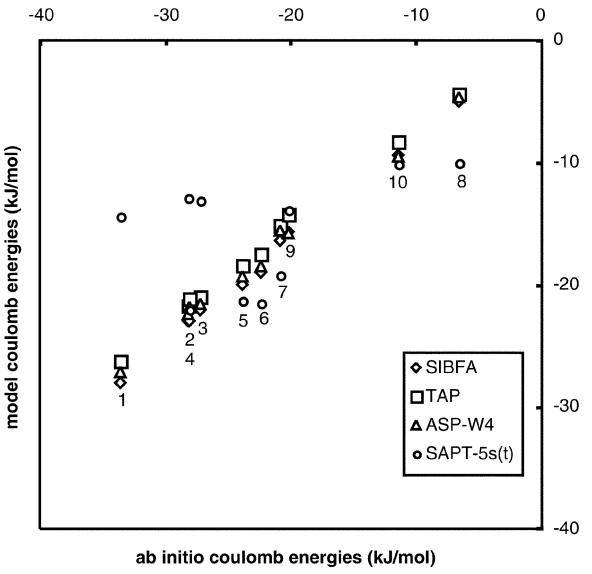

Fig. 2 Model *versus ab initio* coulomb energies of the water dimer at the stationary points determined by Smith *et al.* The *ab initio* energies are the SAPT results from column 4 of Table 3, which include the effects of intramolecular correlation at the CCSD level. Of the analytical models only the SAPT-5s(t) model allows for charge penetration effects.

of the models carries very small errors the high r-values would imply that the penetration component in the *ab initio* results is roughly proportional to the multipole component for these ten geometries, but in the absence of rigorous ways to assess the multipole results we cannot exclude other explanations, such as fortuitous cancellations of errors. The differences between the three model columns are in keeping with the differences between the dipole moments of the various models. The SIBFA model is based on SCF moments, whereas correlated moments are used in TAP and ASP-W4. Thus, as expected, at geometry 1 the TAP result lies about 2 kJ mol^{-1} above SIBFA. Likewise, the ASP-W4 dipole value (0.765 a.u.) is higher than the TAP moment which rationalizes that the ASP-W4 model coulomb energies lie generally in between the TAP and SIBFA values. Thus the three models behave as expected, suggesting that the coulomb part of these models (using sets of multipoles at a number of sites) was adequate. The TAP model, using multipoles up to the quadrupole at atomic sites only, is the simplest to use in practice.

The penetration contribution to the coulomb energy is not accounted for in these three models. Comparing the SIBFA model result to the *ab initio* E_{coul} value of column 1, or the TAP model energies to the MP3-level *ab initio* energies in column 3, one notes that the penetration energy is always attractive (by up to 5 or 6 kJ mol^{-1}) for the present geometries (*cf.* Wheatley[5]). Thus, the penetration contribution to the coulomb interaction is too large to ignore. The present models take it into account by absorbing it in the repulsion term of the model.

The SAPT-5s coulomb results show a large spread. The modeling of the coulomb term in this potential differs from that of the other potentials by the presence of damping factors that were included to account for penetration effects. Since the penetration energy is attractive for the present geometries one would expect the damping to make the SAPT-5s coulomb energies more negative than those of the other models, which are not damped. The results show, however, that for some geometries (and especially those with short O···H contacts, such as 1, 2 and 3) the SAPT-5s coulomb energy is actually much *less* attractive. The anomaly may partly be due to the fact that the damping parameters were optimized in a global fit of all adjustable parameters of the SAPT-5s model on the total *ab initio* SAPT interaction energy, rather than in a fit using only the *ab initio* coulomb energies. However, the anomaly seems primarily due to the fact that in the SAPT-5s(t) model for E_{coul} all atom–atom contributions are reduced by the damping algorithm. Consequently, the H···H and O···O atom–atom repulsions become less repulsive, which is qualitatively correct. On the other hand, the H···O atom–atom attractions are reduced as well. This is qualitatively wrong since charge-overlap effects will actually increase such attractions at the distances under consideration.[32] Thus whereas the somewhat similar damping recipe that was used in developing the SAPT-pp potential[33] led to a good fit to the SAPT-theory coulomb energy, which includes penetration effects, it seems that to account for penetration effects in the SAPT-5s(t) site–site coulomb model a more refined damping scheme than the one used here is called for. We should emphasize that the erroneous E_{coul} behaviour does not affect the quality of the SAPT-5s(t) model total energy since, as we will show, the flexible form assumed for the SAPT-5s(t) repulsion term is capable of compensating errors in the other terms of the potential.

6. Dispersion energy

Using SAPT theory together with the IOM basis set we have calculated several *ab initio* dispersion energy results that may serve to judge the quality of the dispersion models. A selection of these data is shown in the first four columns of Table 4. The first column shows E^{20}_{disp} results. The molecular orbitals employed were the monomer MCBS occupied and virtual SCF MOs. The next column also shows E^{20}_{disp} results, but here (and in the next two columns) DCBS monomer orbitals were employed. The change from MCBS to DCBS introduces charge-transfer contributions that are seen to increase the dispersion energy by some 20%. Most of this change is quenched when one adds the $E^{20}_{exch-disp}$ term (column 3). The first three columns give progressively more complete descriptions of the dispersion energy that is present in a supermolecular MP2 calculation. Finally, the fourth column includes the $\varepsilon^2_{disp}(2)$ term that allows for the effect of intramolecular correlation on the dispersion energy. The $\varepsilon^2_{disp}(2)$ term is seen to increase the dispersion attraction by 15–20%, depending on the stationary point considered. Exchange effects may reduce this, but the corresponding exchange correction has not yet been implemented in the SAPT program. Column 4 provides a description of dispersion energy roughly equivalent with that of the MP4 supermolecular approach.

Table 4 Dispersion energies E_{disp}(kJ mol^{-1}) of the SIBFA, TAP and ASP-W4 models at the ten stationary points of the water dimer; the first four columns list selected *ab initio* dispersion energy results calculated using SAPT theory and the IOM basis set

Ab initio[a]	*Ab initio*[b]	*Ab initio*[c]	*Ab initio*[d]	SIBFA	TAP	ASP-W4
−8.68	−10.35	−8.49	−10.20	−7.66	−7.93	−9.39
−7.98	−9.37	−7.82	−9.26	−6.99	−7.41	−8.35
−7.86	−9.18	−7.70	−9.07	−6.90	−7.35	−8.10
−8.42	−9.96	−8.37	−9.82	−7.11	−8.08	−9.66
−8.13	−9.41	−8.09	−9.26	−6.90	−8.15	−9.27
−8.06	−9.25	−8.03	−9.09	−6.90	−8.21	−9.08
−6.50	−7.78	−6.57	−8.02	−5.44	−5.94	−8.40
−2.73	−3.24	−2.87	−3.73	−2.76	−2.36	−3.93
−6.24	−7.55	−6.32	−7.74	−5.15	−5.47	−7.81
−4.23	−4.98	−4.32	−5.29	−3.77	−3.73	−5.25

[a] E^{20}_{disp}(MCBS). [b] E^{20}_{disp}(DCBS). [c] $E^{20}_{disp} + E^{20}_{exch-disp}$ (DCBS). [d] $E^{20}_{disp} + E^{20}_{exch-disp} + \varepsilon^2_{disp}$ (2) (DCBS).

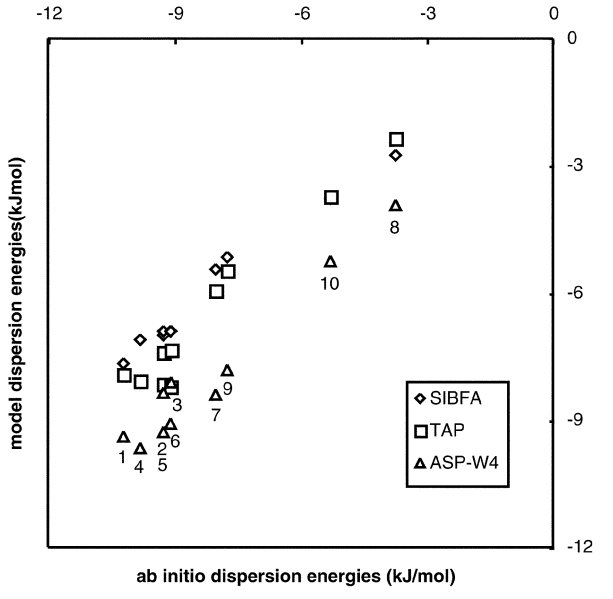

Fig. 3 Model *versus ab initio* dispersion energies for water dimer. The *ab initio* energies are the SAPT results from column 4 of Table 4, which include exchange effects as well as the effect of intramolecular correlation.

To visualize the trends in the data, in Fig. 3 the model dispersion terms are compared to the *ab initio* results of column 4 of Table 4. Since the TAP model was originally fitted to E^{20}_{disp} (MCBS) energies for methanol dimer in the IOM basis set, it is satisfying that of the present *ab initio* options the TAP model agrees best with the E^{20}_{disp} (MCBS) values (Table 4, column 1). The correlation coefficient is 0.986, but the TAP values are about 6% too small. This may be due to a lack of transferability from methanol to water or it may point to a systematic underestimation by the TAP model of the dispersion energy at the rather short separations we consider here. Another shortcoming of the TAP model is that it fails to reproduce the trend in the dispersion energies for geometries 4, 5 and 6. Whereas TAP has only one C_6 term for each atom–atom pair, the SIBFA model has C_6, C_8 and C_{10} dispersion terms. This should increase its reliability when comparing structures with different intermolecular distances. Like the TAP model, the SIBFA model was fitted on E^{20}_{disp} (MCBS) dispersion energies and SIBFA too agrees best with the present E^{20}_{disp} (MCBS) *ab initio* data (column 1, Table 4). The SIBFA model underestimates these energies by more than 10%, probably because it was derived for data that used a doubly polarized basis less complete than IOM, but it performs better for geometries 4, 5 and 6, and its correlation coefficient r is as high as 0.992.

Unlike the TAP and SIBFA models, which are all-atom models, the ASP-W4 dispersion model employs a large set of C_n/R^n terms referring to a single site in each molecule. The highly-correlated large-basis C_n coefficients of Wormer and Hettema[34] were used with some minor modifications. As a result, the ASP-W4 model gives considerably more dispersion energy than the TAP and SIBFA models. Moreover, the ASP-W4 show a better correlation with the correlated *ab initio* data in column 4 of Table 4 than with the lower level data in columns 1–3. However, the ASP-W4 data do not follow the column 4 *ab initio* data very closely ($r = 0.975$). In particular, the ASP-W4 energies for geometries 1, 2 and 3 are too small, while those for 7 and 8 are rather large. These discrepancies were already noted by Millot and Stone.[1] They thought the damping model to be responsible for this, but the discrepancies may also be due to the use of a single-site model.

The SAPT-5s(t) potentials do not have a separate dispersion term, and so could not be included in this comparison.

7. Induction energy

The last three columns of Table 5 show selected results for the induction energy taken from our SAPT calculations using the IOM basis set. As before, DCBS molecular orbitals were used, which implies that charge transfer contributions are included. Thus, these data are unsuitable as a reference for the model potentials since the latter are designed to treat local polarization effects only. We nevertheless included the SAPT data since they were available anyway and since they allow some relevant observations to be made. Firstly, the uncorrelated results including exchange effects (column 6) should correspond to the $\Delta E^{\mathrm{deloc}}$ values of Table 1, but one notes that SAPT seriously underestimates these energies. This behaviour has been documented previously.[17] It arises from the way that exchange effects are treated in SAPT. Now, by using MCBS instead of DCBS molecular orbitals it is possible to generate SAPT induction energies that treat local effects only, and this might have been a way to obtain reference data suitable for our present purpose. However, a comparison of previously reported[35,36] MCBS water dimer induction energies with the SMMO results presented below suggests that at the MCBS level, too, the induction energies would be underestimated. We have therefore not explored this option further.

The second observation is that the intramolecular correlation correction (included in column 7) is not small. It tends to increase the induction energy at all ten geometries. This is contrary to what one expects for the long range (see below), so it points to some (as yet unidentified) short-range attraction mechanism at the correlated level. Possible candidates might be increases of penetration effects and of charge-transfer effects at the correlated level.

In view of the above considerations we looked for an alternative *ab initio* description of the induction energy in which molecules polarize each other, but charge transfer is prevented. Moreover, we need a method that applies antisymmetrization from the outset. We adopted Cullen's Strictly Monomer Molecular Orbital (SMMO) method[37] for this purpose. In this SCF method, charge transfer is prevented since the molecules are allowed to employ only their own basis set when being polarized by the other molecules. This requires an approach in which the MOs of the molecules are not orthogonal to each other. In principle the prevention of charge transfer is not rigorous, since the use of truly complete basis sets on the individual molecules would allow SMMO to recover the full SCF interaction energy of the complex. However, in practice (for

Table 5 Induction energies (kJ mol^{-1}) for the SIBFA, TAP and ASP-W4 models at the ten stationary points compared to various *ab initio* results (using the IOM basis set)

Ab initio[a]	SIBFA	TAP	ASP-W4	*Ab initio*[b]	*Ab initio*[c]	*Ab initio*[d]
−5.31	−4.85	−4.81	−4.04	−12.26	−5.56	−6.42
−4.76	−3.89	−3.96	−3.53	−10.03	−4.85	−5.52
−4.68	−3.31	−3.76	−3.40	−9.56	−4.73	−5.35
−2.70	−2.22	−2.28	−2.28	−8.39	−2.92	−3.47
−2.43	−2.26	−2.05	−2.19	−6.55	−2.64	−3.04
−2.33	−1.88	−1.94	−2.16	−5.85	−2.52	−2.86
−1.37	−0.92	−1.21	−1.27	−5.46	−1.42	−1.82
−0.35	−0.25	−0.41	−0.37	−1.41	−0.40	−0.55
−1.89	−1.30	−2.04	−1.92	−5.45	−1.83	−2.23
−1.19	−0.84	−1.37	−1.26	−2.83	−1.19	−1.44

[a] SMMO theory (see text). [b] $E^{20}_{\mathrm{ind,resp}}$ (DCBS). [c] $E^{20}_{\mathrm{ind,resp}} + E^{20}_{\mathrm{exch-ind,resp}}$ (DCBS). [d] $E^{20}_{\mathrm{ind,resp}} + E^{20}_{\mathrm{exch-ind,resp}} + {}^{\mathrm{t}}E^{22}_{\mathrm{ind}} + {}^{\mathrm{t}}E^{22}_{\mathrm{exch-ind}}$ (DCBS).

conventional, atom-centred basis sets) the SMMO energy readily converges to a numerically well-defined limit less negative than the full SCF dimer energy when enlarging the basis set,[38] because atom-centered basis functions are unsuitable for describing wavefunction details at a neighbouring molecule. We therefore identify the difference between the SMMO and Heitler–London total energy with the induction energy $E_{\rm ind}^{\rm SMMO}$. The SMMO description of the polarization process is more complete than that of related methods such as the restricted variational space (RVS) method[39] because it includes the higher-order interactions between the induced moments. Although SMMO is an SCF method, its induction energies for water dimer should (in the long range) be numerically similar to those in correlated descriptions because here, as it happens, correlation reduces the dipole field F by 7% and enlarges the average polarizability α by 14% so the leading $\frac{1}{2}(\alpha F^2)$ long-range induction terms are little affected. Our SMMO calculations employed the IOM basis set, which should give results close to the basis set limit for this component.

The ASP-W4 and SIBFA induction models were designed using polarizabilities obtained from monomer calculations and the interaction is damped by a factor based on *ab initio* (antisymmetrized) perturbation theory or SCF induction energy data. The SIBFA model has anisotropic dipole polarizabilities on the charge centres of the localized MOs, while in ASP-W4 there are several anisotropic polarizabilities, located on a single site in the molecule. The SIBFA model is defined at the SCF level, whereas ASP-W4 uses experimental as well as correlated quantities. To allow for higher-order induction effects, the SIBFA and ASP-W4 induction energies are obtained in an iterative fashion.

In the TAP model there is one isotropic dipole-polarizability at each atom. The size of these polarizabilities was found by fitting the model (which also included a damping factor) to the non-additivities of the SCF delocalization energies in a set of methanol trimers. The fit was surprisingly accurate (rmsd 0.2 kJ mol^{-1}), confirming[40] that a simple induction model can give an accurate description of the energy non-additivity of larger clusters of polar molecules. Conversely, it has not yet been verified whether this model gives a reasonable description of dimer induction energies as well. By construction, the TAP atomic dipole polarizabilities are effective polarizabilities, which allow for some higher-order induction effects. For this reason it would be inappropriate to apply the TAP induction model in an iterative fashion.

The results of the model potentials are shown in Table 5 and they are plotted against the $E_{\rm ind}^{\rm SMMO}$ data in Fig. 4. The model potentials reproduce the *ab initio* results reasonably well, the *r*-factors being 0.981, 0.986 and 0.970 for SIBFA, TAP and ASP-W4, respectively. The largest discrepancies with the *ab initio* data occur for the first three geometries. These have near-linear hydrogen bonds and short H-bond distances, and the results suggest that all models (and especially the single-site ASP-W4 model) have difficulty in accounting for the resulting field inhomogeneity. The errors in the models are not alarmingly large in the present system where induction gives only a modest contribution to the total interaction energy. However in clusters the relative importance of the induction energy may be much larger as a result of non-additivity. There remains a need for induction models that perform better in the short-range region.

8. Charge transfer energy

The $E_{\rm ind}^{\rm SMMO}$ energy accounts only for the local polarization part of the full SCF delocalization energy. The remainder is due to polarization involving orbitals on the other molecule, so it represents the charge-transfer component of $\Delta E^{\rm deloc}$

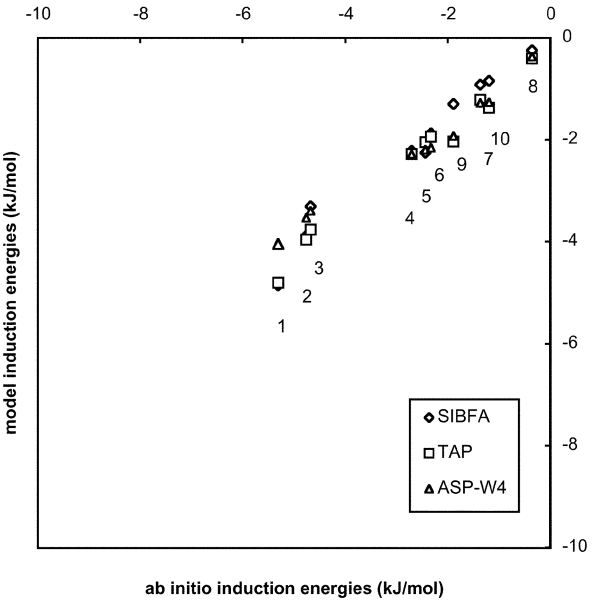

Fig. 4 Model *versus ab initio* induction energies for water dimer. The *ab initio* energies were calculated using the SMMO method in the IOM basis set.

$$E_{\rm ct} = \Delta E^{\rm deloc} - E_{\rm ind}^{\rm SMMO}$$

Our *ab initio* $E_{\rm ct}$ values in the IOM basis set are shown in the first column of Table 6. In agreement with earlier observations,[39] $E_{\rm ct}$ is about as large as $E_{\rm ind}^{\rm SMMO}$ for all ten geometries studied here. It is too large to ignore. Since $E_{\rm ct}$ is known to decrease exponentially with distance, the TAP and SAPT-5s(t) authors chose to account for $E_{\rm ct}$ by including it in the repulsion term of the model. On the other hand, SIBFA and ASP-W4 have an explicit $E_{\rm ct}$ term. The SIBFA model is very detailed and its results are in good agreement with the *ab initio* results, with differences smaller than 0.4 kJ mol^{-1}. The ASP-W4 results are systematically too small. This probably indicates that the perturbation estimates of $E_{\rm ct}$ that were used to fit this model were too small compared to the present *ab initio* data.

9. Long-range term

Instead of separate induction and dispersion terms, the SAPT-5s(t) models have a combined term labelled 'long range'. These long-range energies are compared to the corresponding

Table 6 Charge transfer energies $E_{\rm ct}$ (kJ mol^{-1}) for the ten stationary points obtained in the SIBFA and ASP-W4 models compared to the charge-transfer component of the SCF delocalization energy (using the IOM basis set)

Ab initio[a]	SIBFA	ASP-W4
−4.25	−3.85	−2.70
−3.35	−3.35	−2.21
−3.17	−3.35	−2.11
−1.99	−1.76	−1.40
−1.61	−1.51	−1.32
−1.48	−1.46	−1.28
−0.94	−0.84	−0.81
−0.18	−0.17	−0.16
−0.93	−1.30	−0.84
−0.44	−0.71	−0.43

[a] $E_{\rm ct} = \Delta E^{\rm deloc} - E_{\rm ind}^{\rm SMMO}$.

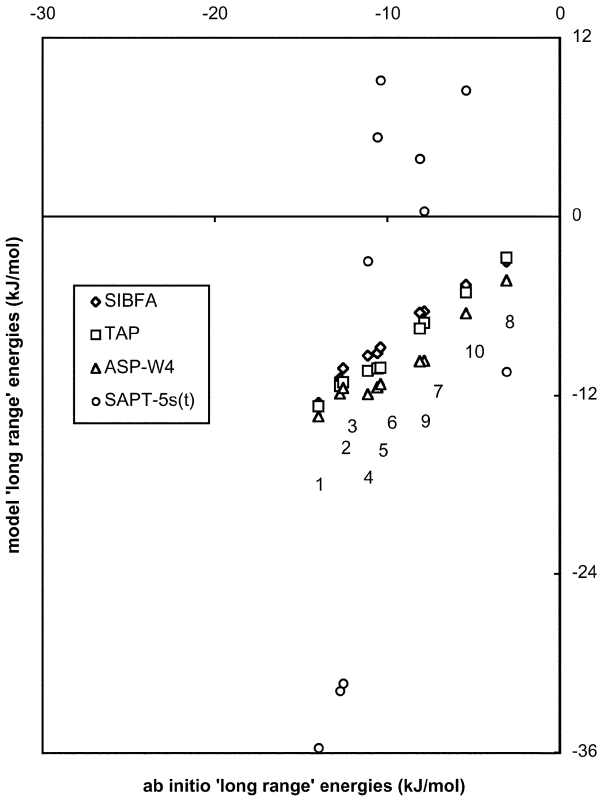

Fig. 5 Model *versus ab initio* 'long range' energies for water dimer. The SIBFA, TAP and ASP-W4 data represent the sum of the corresponding induction and dispersion data shown in Tables 4 and 5. The *ab initio* energies were calculated as the sum of the induction and dispersion *ab initio* data in the first columns of these Tables.

ab initio and model results in Fig. 5. As expected from the previous discussions, the SIBFA, ASP-W4 and TAP models closely follow the *ab initio* data. By contrast, the SAPT results are completely different, with deviations of up to 20 kJ mol^{-1}. . Some of the long-range 'attractions' are even repulsive.

The SAPT-5s(t) long-range model was fitted on data at relatively long R. This does not guarantee that the model is also valid at short R. The main problem is that the simple (isotropic) form of the atom–atom contributions in the model cannot cope with the different anisotropies of the induction and dispersion terms. This leads to fitting coefficients for the O···H atom pairs which are unphysical, such as a positive C_6 value. This explains the anomalous results for the long-range term such as the too small attractions obtained for geometries 1–3, since these are dominated by their short O···H contacts. As in the case of the coulomb term, the anomaly in the SAPT-5s(t) long-range term does not affect the quality of the overall SAPT-5s(t) potential because the errors were compensated by adjusting the repulsion term in the final fitting step.

10. Repulsion energy

Two different approaches have been used to model the terms that cause the short-range repulsion. In SIBFA and ASP-W4 one models the overlap-dependent part E_{rep} of the Heitler–London first-order interaction energy:

$$E_{rep} = \Delta E^{HL} - E_{coul} \quad (1)$$

here ΔE^{HL} is calculated from Hartree–Fock monomer wavefunctions, while E_{coul} is the (multipolar) coulomb term of the SIBFA or ASP-W4 model. Thus, in addition to the first-order exchange repulsion the quantity E_{rep} implicitly allows for the penetration part of the coulomb interaction energy.

The TAP and SAPT-5s(t) models are based on the view that there are several additional overlap-dependent terms that ought to be taken into account (unless the model already considers them explicitly) such as the exchange-dispersion term, the correlation correction to the first-order exchange energy and the charge-transfer energy. It is difficult to develop convincing models for each of these separately. Moreover, the various long-range terms that are explicitly included in a model are usually somewhat uncertain at short range. Hence in these models one defines the 'repulsion' energy as the difference that remains between the total *ab initio* interaction energy ΔE and the coulomb, induction and dispersion terms already present in the model:

$$E_{rep} = \Delta E - E_{coul} - E_{ind} - E_{disp} \quad (2)$$

and the E_{rep} term in the model is fitted to a set of these *ab initio* E_{rep} values.

A selection of E_{rep} results is shown in Table 7. Column 1 shows the *ab initio*/SIBFA repulsion data calculated according to eqn. (1). Column 2 is the SIBFA model fitted to these data. Column 3 gives the E_{rep} term of the ASP-W4 model. The SIBFA model was developed for the data shown in column 1. A good fit might have been expected, but especially for geometries 4, 5 and 6 the model fails to reproduce the *ab initio* data. The ASP-W4 repulsion energies correlate much better ($r = 0.993$) with the *ab initio* data, but the slope is larger. This may partly be due to the better basis set used in the ASP-W4 modelling, for this is known to increase the exchange repulsion.

Column 4 of Table 7 shows the *ab initio*/TAP repulsion data according to eqn. (2). The TAP model (which was fitted on

Table 7 Repulsion terms (kJ mol^{-1}) of various model potentials at the ten stationary points compared to the corresponding *ab initio* repulsion energies as defined in the text. The last two columns list first-order exchange repulsion terms calculated using SAPT (DCBS)

Ab initio[a]	SIBFA	ASP-W4	*Ab initio*[b]	TAP	SAPT-5s	SAPT-5st	E^{10}_{exch}	E^{1}_{exch} [c]
21.50	21.38	23.89	19.49	19.04	30.01	29.30	29.03	34.00
17.57	17.99	20.70	15.59	15.59	26.46	26.39	24.13	28.08
16.69	17.70	20.23	14.71	14.93	26.13	26.12	23.09	26.78
15.14	18.45	18.10	15.05	15.22	7.17	7.31	22.12	25.99
12.97	16.57	15.95	12.98	14.31	−0.87	−0.69	18.64	21.48
12.09	16.74	15.66	12.14	14.04	−4.44	−4.32	17.24	19.68
8.62	9.67	10.91	10.06	8.13	6.06	6.43	14.33	18.11
1.75	1.88	2.46	2.61	1.46	15.49	15.91	3.78	5.43
9.41	9.29	11.29	10.42	7.09	5.73	5.82	15.14	18.51
4.36	4.56	5.58	5.34	3.49	9.23	8.94	7.53	9.48

[a] $E_{rep} = \Delta E^{HL} - E_{coul}$ (SIBFA). ΔE^{HL} was obtained in the CEP4-31G(2d) basis.[30] [b] $E_{rep} = \Delta E^{SCF+MP2}$(IOM) $-(E_{coul} + E_{ind} + E_{disp}$, taken from the TAP potential). [c] $E^{1}_{exch} = E^{10}_{exch} + \varepsilon^{1}_{exch}$(CCSD).

such data for methanol) reproduces these data fairly well (column 5) but since the repulsion energies are very large, the absolute errors are not small. The largest discrepancies occur for geometries that were not present in the methanol data set, such as 9 and 10. It is worth noting that the ASP-W4 model actually correlates better with the *ab initio*/TAP data ($r = 0.984$) than the TAP potential itself ($r = 0.959$), even though the ASP model was not optimized for this kind of data. The SAPT-5s(t) repulsion results in columns 6 and 7 are seen to be completely different. This is because these were fitted to eqn. (2) data based on SAPT-5s(t) components. By construction, the success of the E_{rep} fit determines the final quality of the model, and it is in this way, as we shall see in the next section, that the SAPT-5s(t) models, by using a very flexible fit of eqn. (2), succeed in getting a good representation of the total *ab initio* interaction energy, in spite of their very poor representation of the coulomb, induction and dispersion terms. The resulting SAPT-5s(t) repulsion energies have no physical meaning, because they are heavily contaminated by error corrections. The repulsions even become attractive for geometries 5 and 6.

Conversely, for SIBFA and ASP-W4 no such error correction takes place and, moreover, the terms which are not explicitly modelled enter the final fitting error in their entirity. Both models have an explicit charge-transfer term, and SIBFA moreover has an exchange-dispersion term, but other overlap dependent terms, which may be large as well, are still missing. Notably, according to the SAPT exchange-repulsion energies in columns 8 and 9 of Table 7 the correlation correction to the first-order exchange repulsion is as large as +5.0 kJ mol^{-1} for geometry 1. We have quoted the latter SAPT results in order to complete the documentation of the present SAPT results and because they show that the correlation correction to the exchange energy is as large as 15–40% of the SCF-level exchange term E^{10}_{exch}. This suggests, as mentioned before, that intermolecular overlap effects are enhanced at the correlated level.

11. Quality of the total interaction energies

The results for the total interaction energies are shown in Table 8 (Fig. 6 and 7). Rmsds between the model potentials and the MP2/IOM data are shown in Table 9. These were calculated in two different ways. Those in column 1 are based on the deviations in the relative total energies using the energy of the global minimum as reference. The rmsds in column 2 are calculated from deviations in the relative total energies taking the mean total energy as reference. In the latter rmsds the errors at the ten geometries are treated on an equal footing. Conversely, the rmsds of column 1 are a better measure of the error in the barrier heights for isomerizations on the water dimer potential energy surface. The rmsds in column 1 are larger than those in column 2 for most potentials and it depends on the application which of the two is the more relevant error measure. The rmsds vary from 0.37 to 3.47 kJ mol^{-1}. The best results are obtained for the SAPT-5s(t) potentials. These reproduce the present *ab initio* (relative) energies very well, even though they were fitted to SAPT energies on water dimer in another basis set. Among the next best results are those of the TAP potential, even though it was developed for methanol rather than for water. This ranking reflects the fact that the SAPT-5s(t) and TAP potentials fit their E_{rep} in such a way that shortcomings in the model are compensated to some extent, depending on the flexibility of the E_{rep} fit function. The very small rmsd value obtained in the SAPT 5s potential shows that their E_{rep} is well suited for this purpose. In the SAPT-5st potential the E_{rep} term was tuned to the water dimer far-IR VRT spectrum to reproduce the tunnel splittings. These are governed by particular relative energies, *viz.* those of 2, 4 and 9, but the tuning somewhat spoils the overall rmsd, as reported in Table 9. The much simpler E_{rep} model of the TAP potential is clearly less successful and the aim of keeping the errors smaller than 1 kJ mol^{-1} is not met by this potential.

Table 8 Total interaction energies (kJ mol^{-1}) for several model potentials at the ten stationary points compared to the *ab initio* $\Delta E^{SCF+MP2}$ in the IOM basis set

Ab initio	TAP	ASP-W4	VRT (ASP-W)	SW	SIBFA	SAPT-5s	SAPT-5st
−19.51	−19.96	−19.43	−20.30	−19.8	−23.05	−20.08	−20.79
−17.53	−17.52	−15.82	−17.88	−17.5	−19.29	−18.32	−18.39
−17.42	−17.21	−14.99	−17.39	−17.2	−17.99	−18.34	−18.35
−16.50	−16.33	−17.10	−17.42	−18.5	−15.69	−17.87	−17.73
−15.66	−14.32	−16.20	−16.27	−17.7	−14.14	−16.85	−16.67
−15.53	−13.64	−15.38	−15.35	−17.5	−12.55	−16.82	−16.70
−12.29	−14.23	−15.15	−14.96	−15.6	−13.98	−12.79	−12.42
−4.61	−5.75	−6.72	−6.50	−6.8	−6.40	−4.97	−4.55
−11.40	−14.73	−15.05	−15.71	−15.7	−14.10	−12.08	−12.00
−8.10	−9.94	−10.92	−10.90	−11.2	−10.21	−9.42	−9.71

Table 9 Rmsds in relative total energies (kJ mol^{-1}) for a number of water–water potentials, taking the $\Delta E^{SCF+MP2}$ results in the IOM basis set as reference; maximum deviations are given in parentheses

Ab initio or model potential	Rmsd[a]	Rmsd(mean)[b]
MP2/6-31 + G(d,p)	1.95 (3.61)	1.39 (2.50)
Idem, but CP corrected	1.06 (2.08)	0.81 (1.44)
SAPT-5s	0.51 (0.75)	0.37 (0.54)
SAPT-5st	0.67 (1.34)	0.52 (0.94)
TAP	1.59 (2.88)	1.58 (2.82)
VRT-ASPW	1.60 (3.52)	1.46 (2.90)
ASP-W4	2.23 (3.73)	2.02 (3.25)
SW (data taken from ref. 5)	2.26 (4.01)	1.50 (2.41)
SIBFA (data from ref. 30)	3.47 (6.52)	2.05 (3.86)

[a] Using the total interaction energy for geometry 1 as reference. [b] Using the mean interaction energy for geometries 1–10 as reference.

The SIBFA and ASP-W4 models lack a final compensating fit, so the rmsds reflect the shortcomings of the fits of the individual terms as well as the terms that are missing altogether. In ASP-W4 the final error is mainly due to the errors in the induction and dispersion energies, which were described using single-site models. In SIBFA most of the separate terms were modelled very well, and the poor overall rmsd is caused by the large errors in the E_{rep} fit.

The ASP-W4 potential gives a poor representation of the VRT water dimer spectrum, because the crucial relative energies are poorly represented. This was rectified to some extent in the VRT-ASPW version, leading to better overall rmsd values in Table 9.

Judging by its rmsd values the SW potential is of better quality than ASP-W4. Individual components were not available for this potential, but it is possible that the single-site models for induction and dispersion, as in the case of ASP-W4, affect the quality adversely.

12. Conclusions

1. Counterpoise corrected SCF + MP2 interaction energies have been computed in the IOM basis set at ten stationary geometries of the water dimer. Compared to the most accurate data currently available the r.m.s. error of the IOM data was found to be only about 0.3 kJ mol^{-1} (that is, for ΔEs relative to ΔE at the global minimum). Thus the IOM data are of sufficient accuracy to use them as reference in a test of the quality

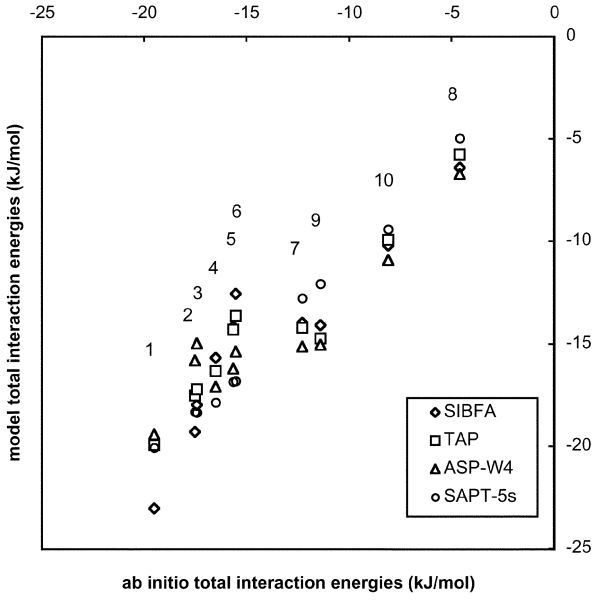

Fig. 6 Selected model *versus ab initio* total interaction energies at the water dimer stationary points. The *ab initio* energies are MP2 CP-corrected supermolecule interaction energies in the IOM basis set.

of analytical water–water potentials. Note that for most applications where such potentials are used it is relative energies that are important.

2. The quality of some recent analytical water–water potential functions has been tested by comparing their individual components as well as their relative total ΔEs to the IOM data. In some cases the *ab initio* data were reproduced surprisingly well. The SAPT-5s(t) potentials gave the best fit to the relative total ΔEs (*e.g.* SAPT-5s, rmsd 0.5 kJ mol^{-1}). Some defects were identified in the individual coulomb and long-range terms of these potentials, but clearly the very flexible form of the repulsion term used in the final fitting was able to compensate for these problems. Other potentials (*viz.* SIBFA, TAP and ASP-W4) were successful in fitting one or more of the separate components, *i.e.* in fitting (part of) the physics of the interaction, but the error in the total relative ΔEs was larger (1.6–3.5 kJ mol^{-1}). The remaining potentials (*i.e.* SW, VRT-ASPW) had similar errors in their total ΔEs.

There are several possible causes for the larger error in these cases. The *ab initio* data used in deriving the SIBFA and ASP-W4 potentials were most probably less accurate than the IOM data used in the present test. Moreover, neither SIBFA nor ASP-W4 has a final fitting step to the total ΔE which could have compensated for lingering errors in these potential models. The TAP model has such a final step, and was developed using IOM *ab initio* energies, but it was developed for methanol dimer and trimer data, and except for the multipole description of the water monomer the methanol TAP model was tranferred to water without any adjustment. In short, all of these models can probably be somewhat improved by performing a reoptimization using the IOM water data. However, it seems likely that some error would remain even then, due to intrinsic limitations of these models.

3. An important aspect of a potential model is its transferability, *i.e.* the degree to which it can be used on related systems without changes in the model. Water dimer models that reproduce the values of dimer binding energies but not the underlying physical mechanisms can be used to describe the pair interactions in a larger water cluster, but they will not recover the pronounced many-body non-additivity energies of such systems, unless explicit non-additivity terms are added to the model. Such models are therefore not transferable to larger water clusters. As illustrated in the present work, the SAPT-5s(t) models are in this class. Now, water cluster energy non-additivities have been shown to arise primarily from the non-pairwise additivity of the induction energy. The physics involved is that one should add the polarizing fields due to all other molecules before the induction energy is evaluated. The remaining water models do have such an induction model and will therefore recover some of the non-additivity effects in larger water clusters. Since in the present work we studied dimer interactions only, we cannot judge the quality of the various models on this aspect. The ASP-W4 and SIBFA models have been explicitly tested on larger water clusters, and were found to reproduce the non-additivity effects on energies and geometries adequately.[27,41] Conversely, the TAP model was designed to reproduce trimer non-additivities and here a remaining question was how well this model describes the two-body induction term. The present calculations have shown TAP to be adequate in this respect.

4. Should errors of a few kJ mol^{-1} in total ΔEs be considered small or large? This clearly depends on the application envisaged for a particular potential. In most biological applications one would like the error in the potential to be small compared to kT at room temperature, *i.e.* smaller than 1 kJ mol^{-1}. None of the transferable potentials actually achieves such accuracy. However, the discussion in previous sections suggests that an improved water–water potential could be assembled by combining the most successful elements of the SIBFA, ASP-W4 and TAP potentials. The coulomb term could be taken from TAP, since it is the simpler of the three versions, and yet very accurate. None of the three induction components is entirely satisfactory. It would be nice if the simplicity of the TAP model (atom–atom terms only) could be retained, but it is necessary to account for inhomogeneity of the inducing field and, moreover, an iterative process is needed to allow for higher-order induction effects. For the dispersion term the form of the SIBFA model seems adequate, but it will have to be refitted on more accurate *ab initio* data. For the repulsion term we advocate the anisotropic ASP-W4 model. One would hope that this model (or slight extensions of it) is flexible enough to provide an accurate final fit of the kind applied in the TAP and SAPT-5s(t) fitting procedures. One could also include explicit terms for some of the larger missing

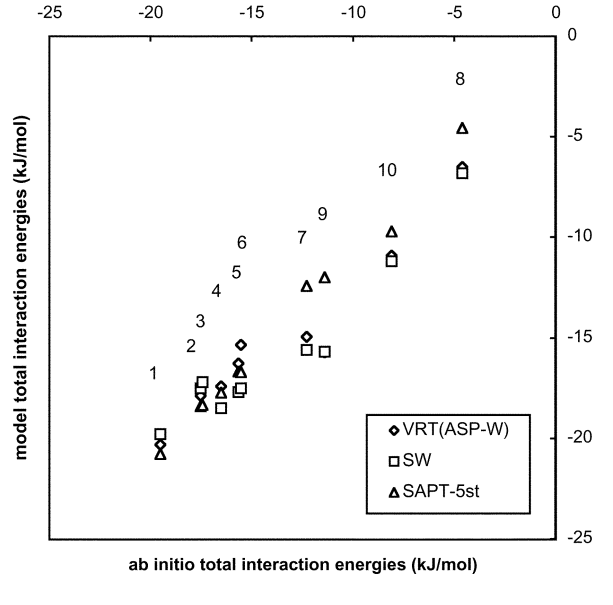

Fig. 7 Further model *versus ab initio* total interaction energies at the water dimer stationary points. The *ab initio* energies are MP2 CP-corrected supermolecule interaction energies in the IOM basis set.

terms, such as the charge-transfer term, for which the SIBFA model is adequate.

Acknowledgements

We thank Dr P. E. S. Wormer for providing Euler angles for Smith's geometries and for assistance in running SAPT96. We thank Dr W. M. Klopper for generating the SAPT-5s(t) potential data at these geometries. We thank Dr N. Gresh for providing the SIBFA data, as well as the corresponding SCF *ab initio* results. We also thank our first-year students Alex de Beer, Erik Hesselink, Arie van Houselt and Tim van Wijngaarden for their contribution in the early stages of this work.

References

1 C. Millot and A. J. Stone, *Mol. Phys.*, 1992, **77**, 439.
2 B. J. Smith, D. J. Swanton, J. A. Pople, H. F. Schaefer III and L. Radom, *J. Chem. Phys.*, 1990, **92**, 1240.
3 L. B. Braly, J. D. Cruzan, K. Liu, R. S. Fellers and R. J. Saykally, *J. Chem. Phys.*, 2000, **112**, 10 293.
4 C. Millot, J.-C. Soetens, M. T. C. Martins Costa, M. P. Hodges and A. J. Stone, *J. Phys. Chem. A*, 1998, **102**, 754.
5 R. J. Wheatley, *Mol. Phys.*, 1996, **87**, 1083.
6 N. Gresh, P. Claverie and A. Pullman, *Int. J. Quantum Chem.*, 1986, 101.
7 N. Gresh, M. Leboeuf and D. Salahub, *ACS Symp. Ser.*, 1994, **569**, 82.
8 N. Gresh, *J. Comput. Chem.*, 1995, **16**, 856.
9 W. T. M. Mooij, F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt and B. P. van Eijck, *J. Phys. Chem. A*, 1999, **103**, 9872.
10 R. S. Fellers, C. Leforestier, L. B. Braly, M. G. Brown and R. J. Saykally, *Science*, 1999, **284**, 945.
11 E. M. Mas, R. Bukowski, K. Szalewicz, G. C. Groenenboom, P. E. S. Wormer and A. van der Avoird, *J. Chem. Phys.*, 2000, **113**, 6687.
12 G. C. Groenenboom, P. E. S. Wormer, A. van der Avoird, E. M. Mas, R. Bukowski and K. Szalewicz, *J. Chem. Phys.*, 2000, **113**, 6702.
13 G. S. Tschumper, M. L. Leininger, B. C. Hoffman, E. F. Valeev, H. F. Schaefer III and M. Quack, *J. Chem. Phys.*, 2002, **116**, 690.
14 J. G. C. M. van Duijneveldt-van de Rijdt, P. J. A. Ruttink, W. M. Klopper, W. T. M. Mooij and F. B. van Duijneveldt, in preparation.
15 C. J. Burnham and S. S. Xantheas, *J. Chem. Phys.*, 2002, **116**, 1479.
16 B. Jeziorski, R. Moszynski and K. Szalewicz, *Chem. Rev.*, 1994, **94**, 1887.
17 K. Szalewicz, B. Jeziorski, in *Molecular Interactions*, ed. S. Scheiner Wiley, New York, 1997.
18 J. G. C. M. van Duijneveldt-van de Rijdt and F. B. van Duijneveldt, *J. Chem. Phys.*, 1992, **97**, 5019.
19 V. Saunders and M. F. Guest, *ATMOL Program Package*, SERC Daresbury Laboratory, UK, 1986.
20 J. H. van Lenthe, *SERVEC, ATMOL Vector Service Program*, Utrecht University, 1988.
21 J. H. van Lenthe, *INTACAT program package*, Utrecht University, 1988.
22 J. Verbeek, J. H. Langenberg, C. P. Byrman, F. Dijkstra and J. H. van Lenthe, *TURTLE, a VB/VBSCF ab initio program*, Utrecht University, 1998.
23 R. Bukowski, P. Jankowski, B. Jeziorski, M. Jeziorska, S. A. Kucharski, R. Moszynski, S. Rybak, K. Szalewicz, H. L. Williams and P. E. S. Wormer, *SAPT96*, University of Warsaw and University of Delaware, 1996.
24 S. S. Xantheas, C. J. Burnham and R. J. Harrison, *J. Chem. Phys.*, 2002, **116**, 1493.
25 F. B. van Duijneveldt, J. G. C. M. van Duijneveldt-van de Rijdt and J. H. van Lenthe, *Chem. Rev.*, 1994, **94**, 1873.
26 S. M. Resende, W. B. De Almeida, J. G. C. M. van Duijneveldt-van de Rijdt and F. B. van Duijneveldt, *J. Chem. Phys.*, 2001, **115**, 2476.
27 M. P. Hodges, A. J. Stone and S. S. Xantheas, *J. Phys. Chem. A*, 1997, **101**, 9163.
28 R. S. Fellers, L. B. Braly and R. J. Saykally, *J. Chem. Phys.*, 1999, **110**, 6306.
29 B. P. van Eijck, W. T. M. Mooij and J. Kroon, *J. Phys. Chem. B*, 2001, **105**, 10 573.
30 N. Gresh, personal communication 2001.
31 G. Chalasinski and M. M. Szczesniak, *Mol. Phys.*, 1988, **63**, 205.
32 A. J. Stone, *The Theory of Intermolecular Forces*, Clarendon Press, Oxford, 1996, p. 95.
33 E. M. Mas, K. Szalewicz, R. Bukowski and B. Jeziorski, *J. Chem. Phys.*, 1997, **107**, 4207.
34 P. E. S. Wormer and H. J. Hettema, *J. Chem. Phys.*, 1992, **97**, 5592.
35 H. Williams, E. M. Mas, K. Szalewicz and B. Jeziorski, *J. Chem. Phys.*, 1995, **103**, 7374.
36 M. Torheyden and G. Jansen, *Theor. Chim. Acta*, 2000, **104**, 370.
37 J. M. Cullen, *Int. J. Quantum Chem. Symp.*, 1991, **25**, 193.
38 A. Famulari, M. Raimondi, M. Sironi and E. Gianinetti, *Chem. Phys.*, 1998, **232**, 275.
39 W. J. Stevens and W. Fink, *Chem. Phys. Lett.*, 1987, **139**, 15.
40 E. Clementi, W. Kolos, G. C. Lie and G. Ranghino, *Int. J. Quantum Chem.*, 1980, **17**, 377.
41 M. Masella, N. Gresh and J.-P. Flament, *J. Chem. Soc., Faraday Trans.*, 1998, **94**, 2745.