

Cite this: *Phys. Chem. Chem. Phys.*, 2012, **14**, 955–963

www.rsc.org/pccp

PAPER

Polarization charge densities provide a predictive quantification of hydrogen bond energies†

Andreas Klamt,^{*ab} Jens Reinisch,^a Frank Eckert,^a Arnim Hellweg^a and Michael Diederhosen^a

Received 17th August 2011, Accepted 8th November 2011

DOI: 10.1039/c1cp22640a

A systematic density functional theory based study of hydrogen bond energies of 2465 single hydrogen bonds has been performed. In order to be closer to liquid phase conditions, different from the usual reference state of individual donor and acceptor molecules in vacuum, the reference state of donors and acceptors embedded in a perfect conductor as simulated by the COSMO solvation model has been used for the calculation of the hydrogen bond energies. The relationship between vacuum and conductor reference hydrogen bond energies is investigated and interpreted in the light of different physical contributions, such as electrostatic energy and dispersion. A very good correlation of the DFT/COSMO hydrogen bond energies with conductor polarization charge densities of separated donor and acceptor atoms was found. This provides a method to predict hydrogen bond strength in solution with a root mean square error of 0.36 kcal mol⁻¹ relative to the quantum chemical dimer calculations. The observed correlation is broadly applicable and allows for a predictive quantification of hydrogen bonding, which can be of great value in many areas of computational, medicinal and physical chemistry.

Introduction

Hydrogen bonding is the strongest intermolecular interaction mode for many molecules especially in organic and biochemistry.¹ Moreover, it is responsible for most recognition and structure formation phenomena in biological systems.² Therefore the proper quantification of hydrogen bond energies is of crucial importance for the modeling and understanding of many important systems in physical chemistry, life-science, formulation science, and chemical engineering. Any predictive assessment of physicochemical properties such as solubility, partitioning and phase separation, drug–receptor binding, crystal structure, *etc.* would fail, if hydrogen bond (HB) energies are not known with an adequate accuracy. However, due to the complex nature of hydrogen bonding, even the definition of hydrogen bonding is still an issue of discussion,³ and reasonably simple and accurate methods for the predictive quantification of hydrogen bond energies are still unavailable. This was recently demonstrated by Wendler *et al.*,⁴ testing several predictive and retrospective

quantification schemes for hydrogen bond energies, and ending up with root mean square deviations (RMSD) in the order of 1.5 kcal mol⁻¹. Without any doubt, nowadays high-level quantum chemical (QC) calculations can be useful to quantify and analyze hydrogen bond energies for small reference systems of isolated donor–acceptor pairs,⁵ but they are far too expensive to serve as standard tools for practical hydrogen bond energy quantification in most life science and fluid phase thermodynamics applications.

While almost all of the large number of previous quantum chemical studies of hydrogen bonding are performed for molecules and hydrogen bonded clusters in vacuum, it is the purpose of this paper to introduce the state of molecules and hydrogen bonded dimers in a virtual conductor environment, as emulated by the COSMO continuum solvation model,⁶ as a proper starting point and reference system for the quantification and analysis of hydrogen bonding. Apart from being closer to the real environment of molecules in polar solvents and crystals, *i.e.* being closer to the realistic conditions under which most of the important hydrogen bond phenomena take place, the conductor reference state suppresses the long-range electrostatics and thus reduces the hydrogen bond interactions to short range quantum chemical energy contributions. The suppression of the long-range electrostatics results from the polarization charges, which are generated on the conductor–solute interface in order to avoid the penetration of the solute electric field into the bulk conductor. These polarization charges are self-consistently calculated in the COSMO solvation

^a COSMOlogic GmbH&CoKG, Burscheider Str. 515, 51381 Leverkusen, Germany. E-mail: klamt@cosmologic.de; Tel: +49 2171 731681

^b Institute of Physical and Theoretical Chemistry, University of Regensburg, Germany

† Electronic supporting information available: Table of the COSMO radii (SII) and the complete table of the data for the 2465 hydrogen bond complexes, and an algorithm to generate the vdW-distance HB clusters. All structures used in the study are supplied in xyz format in a separate zip file. See DOI: 10.1039/c1cp22640a

model along with the quantum chemical equations and hence they are readily available as a result of a COSMO calculation. The local conductor surface polarization charge densities σ , *i.e.* the polarization charge per unit area of solute–conductor interface, has been shown to be of great value for the quantification of molecular interactions.^{7,8} In this paper we demonstrate that the conductor polarization charge densities of HB-donors and HB-acceptors are excellent descriptors for a combinatorial, predictive quantification of the HB energy. A very similar, but purely empirical quantification of hydrogen bond free energies in solution based on conductor polarization charge densities has already been successfully established since long as part of the surface interaction energy functional of the COSMO-RS fluid phase thermodynamics model,^{7,8} but in this paper for the first time that approach is justified based on quantum chemical HB cluster calculations.

It should be noted that it is not the purpose of the present paper to provide a calculation scheme for absolute energies or enthalpies of hydrogen bonding, but to develop a systematic method for the relative quantification of hydrogen bonding. We are aware of the fact that even this goal is only partly achieved here, because the present study is restricted to quantum chemical HB energies in a virtual conductor environment and thus treating cooperative effects only on a continuum level, and since it is neglecting any entropic contributions, while the latter are known to be of crucial importance for HB thermodynamics due to the large loss of entropy going along with HB formation. We focus on single hydrogen bonded complexes of neutral species with minimum steric hindrance and minimum conformational changes, but covering a broad range of hydrogen bond strength and chemical diversity. By this focus and those limitations we try to work out some small but systematic piece of insight which may help to solve the complicated puzzle of hydrogen bonding.

The rest of the paper will be organized as follows: we start with a discussion of the accuracy of different QC calculation levels for gas phase HB Energies. From this we derive a justification of a suitable QC calculation scheme for a big dataset of HB complexes considered later. The next section gives a comparison of gas phase and COSMO reference state HB energies and some conclusions which will be drawn based on those regarding the contributions of different interaction modes. Then we will present and discuss the results of a large scale study on HB energies in the conductor reference state and the correlation of these energies with the conductor surface polarization charge densities σ of the donors and acceptors. This is followed by a final discussion.

Quantum chemical accuracy for gas phase hydrogen bond energies

Jurečka *et al.*⁵ have published QM reference HB energies for a small set of seven biological relevant predominantly hydrogen bond complexes on a CCSD(T)/CBS level, which meanwhile are widely accepted as the most accurate estimates of hydrogen bond interaction energies of small and medium hydrogen bonded complexes. Their calculations are performed *in vacuo* and give the total interaction energy, *i.e.* the difference of the sum of the individual QC energies of the two reactant molecules and the final QC energy of the hydrogen bonded

complex. Due to the CCSD(T) and CBS corrections their interaction energies do include at one of the best currently practical levels all interaction energy contributions as long-range electrostatics, dispersion, geometric relaxation, and the “covalent hydrogen bond energy”, *i.e.* the essentially quantum chemical contributions resulting from the reorganization of the electronic wave function in the complex compared to the educts. They do not include any zero-point or finite temperature vibrational, rotational, or translational contributions, and thus in a thermodynamic sense they are neither complete gas phase interaction enthalpies nor free energies of hydrogen bonding. We will refer to their data as JSCH7 data further on. In order to find an adequate, computationally affordable QC level for our large scale study on hydrogen bonded complexes, we tested a series of density functional theory (DFT) methods *versus* the JSCH7 data. Level 1 (L1) consists in BP^{9,10} DFT calculations with a def-TZVP^{11,12} basis set. This level was our default level for our liquid phase thermodynamics calculations throughout the past 12 years and proved to yield robust results. Level L2, BP/def-TZVP//BP/def2-TZVPD, improves L1 by final single-point energy corrections with a larger basis set def2-TZVPD,¹³ which supplements TZVP by diffuse functions. In order to improve DFT with respect to dispersion, in level L3, BP/def-TZVP//BP-D3/def-TZVP, we add Grimme’s D3-dispersion correction¹⁴ to the L1 level. Level L4, BP/def-TZVP//BP-D3/def-TZVP, is the same as L2 with added D3 dispersion, and Level L5, BP-D3/def2-TZVPD, is the same as L4, but with complete geometry optimization on the larger basis set and D3-dispersion. Level L6 is identical with L3, but with counter Poise (CP) corrections, and L7 is the CP corrected L4 level. All calculations have been performed with the TURBOMOLE6.3 program package.¹⁵ All results are given in Table 1 and graphically displayed in Fig. S11 (ESI†).

Table 1 clearly demonstrates that all DFT based results strongly improve by addition of dispersive contributions *via* the D3 correction. While the larger basis set seems to increase the deviations to the reference in the step from L1 to L2, it causes a large reduction of the RMSD if dispersion is included (see L4 *vs.* L2). Level L5, *i.e.* complete geometry optimization including D3-dispersion and the larger basis set, does not appear to yield any improvement over L4, *i.e.* over single-point corrections based on BP-TZVP geometries. CP corrections yield a considerable reduction (0.45 kcal mol⁻¹) of the RMSD for the def-TZVP basis set (see L6 *vs.* L3), but the improvement is only 0.06 kcal mol⁻¹ on the larger def2-TZVPD basis set (L7 *vs.* L4). Based on these data we consider L4, *i.e.* BP/def-TZVP//BP-D3/def2-TZVPD with a RMSD of 0.75 kcal mol⁻¹ as a good level for reliable estimates of HB-interaction energies at affordable computational costs. It is interesting to note, that all of the tested levels yield very good correlative results. After linear regression, the *r*²-coefficient for all levels is close to 0.99, and the RMSD varies only between 0.61 and 0.78 kcal mol⁻¹. This observation agrees well with the data reported by Paton and Goodman,¹⁶ who report BH&HLYP/aug-ccpVTZ DFT results for the same data set, which have a large RMSD of 5.2 kcal mol⁻¹, but which yield the same *r*² of 0.99 by linear regression. A correlation with *r*² of 0.99 of DFT HB energy on BLYP-D level *vs.* CCSD(T) results was also reported by Wendler *et al.*⁴ on a very different set of

Table 1 Hydrogen bond interactions energies in vacuum (in kcal mol⁻¹) for 7 hydrogen bonded clusters calculated on different QC levels: L1 = BP/def-TZVP, L2 = BP/def-TZVP//BP/def2-TZVPD, L3 = BP/def-TZVP//BP-D3/def-TZVP, L4 = BP/def-TZVP//BP-D3/def2-TZVPD, L5 = BP-D3/def2-TZVPD, L6 = L3 (counterpoise corrected), L7 = L4 (counterpoise corrected)

HB-complex	JSCH7	L1	L2	L3	L4	L5	L6	L7
(NH ₃) ₂	3.17	3.05	2.03	4.1	3.07	3.11	3.72	2.92
(H ₂ O) ₂	5.02	5.69	4.31	6.5	5.14	5.20	5.83	5.06
Formic acid dimer	18.61	16.30	15.79	19.1	18.56	18.63	18.22	18.42
Formamide dimer	15.96	13.80	13.13	16.8	16.09	16.14	16.09	15.97
Uracil dimer	20.65	16.88	16.24	20.6	20.01	20.10	19.96	19.81
2-Pyridoxine...2-aminopyridine	16.71	14.53	13.62	19.0	18.13	18.23	18.38	17.92
A...T Watson-Crick	16.37	13.37	12.43	18.4	17.42	17.55	17.66	17.18
Statistics								
RMSD		2.34	2.98	1.38	0.71	0.76	0.93	0.65
Mean		-1.84	-2.71	1.14	0.28	0.35	0.48	0.11
Correlation coefficient <i>r</i> ²		0.99	0.99	0.99	0.99	0.99	0.98	0.99
Correlation RMSD		0.63	0.61	0.77	0.65	0.66	0.78	0.63
Correlation slope		1.26	1.21	1.01	0.98	0.98	1.03	0.99
Correlation constant		-1.31	0.37	-1.30	0.00	-0.04	-0.84	0.09

5 small-compound hydrogen bond clusters, which included a broader range of elements, *i.e.* C, H, O, F, Cl and S, while the JSCH7 set does only include C, H, O, and N. Summarizing we may conclude that apparently the trends of hydrogen bond energies are very well described by standard DFT calculations, and good quantitative agreement with current QC accuracy limits can be achieved with sufficiently large basis sets and addition of empirical dispersion corrections.

***In vacuo* vs. conductor reference state for hydrogen bond energies**

All of the above mentioned studies on HB energies employed the standard reference state of quantum chemistry, *i.e.* molecules *in vacuo*. While from a theoretical point of view this is the simplest and best defined reference, it nevertheless causes a number of problems. One problem is that such definition of HB energy leads to a mixing of contributions arising from long-range-electrostatics, dispersion, and covalent HB contributions, *i.e.* that part of the hydrogen bond energy which results from the electronic reorganization which goes along with the penetration of the polar hydrogen into the electron density of the HB acceptor atom. Obviously the latter part, which we may abbreviate as covalent-HB energy, is hard to define exactly, but we can try to approximately define it by subtracting those energy contributions from the total HB energy which clearly have nothing to do with donor-acceptor interpenetration. Since HB donors and HB-acceptors always have considerable local dipole moments, a considerable Coulomb energy is already gained during the approach of the initially infinitely distant donor and acceptor down to a typical van der Waals (vdW) contact distance, *i.e.* the typical distance of non-hydrogen bonded molecules. This distance approximately can be described by the sum of the element specific van der Waals radii as collected by Bondi.¹⁸ In this paper we will employ a very similar definition of such contact distances based on radii resulting from the COSMO solvation model. Obviously, the net gain at vdW-distance would also include some contribution arising from the mutual polarization of donor and acceptor occurring in the last stage of the approach, but—although going along with small electron reorganization—this effect arising from polarity and polarizability

typically would not be counted as a true hydrogen bond contribution, because it also occurs for polar molecules which do not form hydrogen bonds. The same is true for dispersion energy. Most of the dispersive energy is gained along the approach from infinite to vdW-contact distance and thus has little to do with hydrogen bonding itself. From these considerations we hereby suggest to split the total HB energies into a non-covalent HB energy part and a covalent HB energy by introducing an intermediate vdW-HB complex, which can be constructed by moving the HB educts towards each other until the HB-donor and HB-acceptor atoms are at vdW-contact distance and have the same orientation to each other as in the final HB complex. A detailed algorithm to generate such a complex is described in the ESI.†

With this definition of the vdW-HB-complex and employing the QC level L4 introduced above we have analyzed the non-covalent HB energy and covalent HB energy contributions for the JSCH7 dataset and for 7 additional clusters of small molecules in order to cover a broader range of donors and acceptors. The results are shown in Table 2 and Fig. 1. The two components of the total HB energy *in vacuo* appear to be correlated, which is not completely surprising, since both parts are in some way proportional to the polarities of the donors and acceptors. Nevertheless, for weak hydrogen bond complexes we find that essentially the complete interaction energy is already included in the non-covalent HB energy, *i.e.* that there is almost no covalent contribution. Remarkably, for the water dimer, the covalent part of the HB energy is only ~1/3 of the total HB energy, and even for very strong HBs as the HF-NH₃ bond almost 40% of the interaction energy seems to arise from non-covalent contributions.

Another problem of the usage of the vacuum reference state arises from the fact that most of the important hydrogen bonding phenomena occur in polar condensed phases, most often in polar solvents or crystals. In such systems the hydrogen bond reactant molecules are surrounded by polar and maybe already hydrogen bonding neighbors. As a result they are usually already much more polarized and the energy gain resulting from a certain HB under consideration merely is the net energy gain of a reaction in which the pre-polarized educts form a hydrogen bond while removing their original interaction partners on the hydrogen bond contact surface area. Obviously it is impossible to take such complicated situation into account

Table 2 Hydrogen bond interactions energies (in kcal mol⁻¹) of 14 HB dimers *in vacuo*, and in the conductor reference state, analyzed with respect to covalent and non-covalent contributions

HB-dimer	<i>In vacuo</i>			In conductor			In conductor without DFT-D3		
	Total HB energy	Non-covalent HB energy	Covalent HB energy	Total HB energy	Non-covalent HB energy	Covalent HB energy	Total HB energy	Non-covalent HB energy	Covalent HB energy
Formic acid	-18.56	-6.24	-12.32	-9.71	-0.27	-9.44	-6.96	1.12	-8.08
Formamide	-16.09	-8.43	-7.66	-5.61	-0.96	-4.65	-2.70	0.78	-3.48
Uracil	-20.01	-9.19	-10.82	-7.38	0.01	-7.38	-3.64	1.92	-5.56
2-Pyridoxine...2-aminopyridine	-18.13	-8.22	-9.91	-8.59	-0.87	-7.72	-4.08	1.27	-5.35
Adenine...thymine	-17.42	-7.32	-10.11	-8.55	-0.83	-7.72	-3.71	1.35	-5.06
nh3...nh3	-3.07	-2.98	-0.09	0.19	0.37	-0.18	1.07	0.92	0.15
h2o...h2o	-5.14	-3.00	-2.14	-3.24	-0.26	-2.98	-2.40	0.28	-2.68
Formic acid monohb	-6.26	-3.16	-3.09	-4.16	0.00	-4.15	-3.09	0.60	-3.69
h2s...h2s	-2.31	-2.02	-0.29	-1.22	-0.81	-0.42	0.03	0.23	-0.20
hcl...h2o	-6.48	-2.58	-3.90	-7.05	0.36	-7.41	-6.09	0.43	-6.53
hf...h2o	-9.39	-3.32	-6.08	-9.04	0.26	-9.30	-8.42	0.67	-9.10
hf...nh3	-14.57	-3.89	-10.68	-18.10	0.69	-18.79	-17.37	1.00	-18.37
hcn...hcn	-4.50	-3.50	-1.00	-0.90	0.40	-1.30	-0.13	0.95	-1.08
h2o...nh3	-7.41	-3.60	-3.81	-6.27	-0.36	-5.91	-5.17	0.14	-5.31

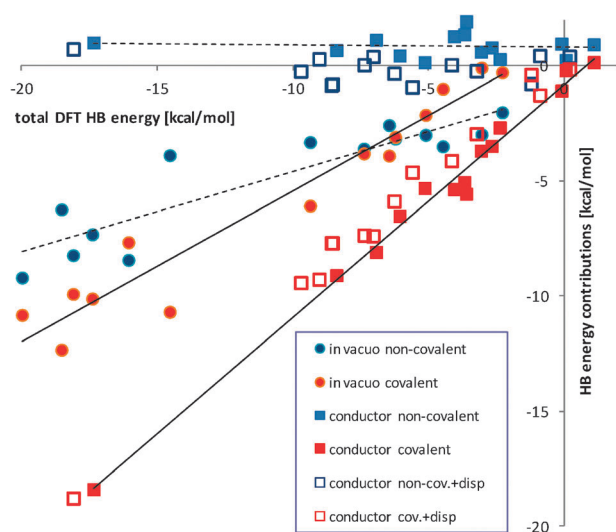


Fig. 1 HB energy components total HB energies for 14 compounds *in vacuo* and in the conductor reference state.

in detail in a systematic study. Nevertheless, we may introduce the conductor reference state as an alternative starting point for HB energy considerations. The state of a molecule embedded in a virtual conductor was originally introduced in the context of the conductor-like screening model COSMO,⁶ which is a modification of dielectric continuum models. Later this state has been declared as energetic reference state for the COSMO-RS method,^{7,8} which is a statistical thermodynamics treatment of solvation energies based on the surface polarization charge densities σ arising on the molecular surfaces when embedded in a virtual conductor. It should be noted that the conductor reference is identical with the infinite dielectric reference state and hence can be almost identically achieved within other dielectric continuum solvation models.¹⁷ We will denote it as conductor state further on. Basically the conductor reference state is a clean reference because the electric field of each molecule is completely compensated by the surface polarization charge density σ , which is self-consistently calculated throughout the QM/COSMO procedure. The only problem is

the proper definition of the molecular surface. During the past 15 years some consensus had been achieved within the implicit solvation modeling community that at least for neutral compounds reasonable molecular surfaces can be constructed based on atom centered spheres using solvation radii, which are roughly 15%–20% larger than the radii introduced by Bondi.¹⁸ In this paper we will use the element specific set of COSMO radii optimized within the COSMO-RS method, which is sometimes referenced as “Klamt radii” in the literature. These radii are about 17% larger than the radii collected by Bondi (see also in ESI†, Table S11).

Molecules in the conductor state are electrostatically optimally embedded by conductor screening charges and self-consistently polarized with respect to this embedding, *i.e.* their state is much more similar to the real situation in polar environments than the vacuum reference is. In addition, two molecules forming a cluster within the conductor embedding will not gain any attractive electrostatic interaction energy, because they already were electrostatically perfectly embedded individually. A repulsive energy contribution that typically is much smaller, the so called misfit energy as introduced within COSMO-RS, may occur if the electrostatic contact in the cluster is less favorable than the conductor embedding. Hence in HB energy calculations in the conductor state we can be sure that the long-range electrostatics does not contribute to the attractive HB energy. Because the conductor embedding does not mimic the dispersive interactions resulting from such a realistic condensed phase environment, the dispersive energy gain in conductor reference state HB energy calculations will be similar as *in vacuo*. Nevertheless, it would be more realistic to assume that the conductor acts like an average molecular dispersion partner. In that case the net dispersive energy gain in HB cluster formation would be very small, most likely negligible. Indeed, this situation can be reasonably mimicked by DFT calculations without empirical dispersion correction, since DFT itself does not include the dispersion energy. Hence we will consider the DFT/COSMO (DFTC) calculations without D3 correction as the best level for conductor reference state HB energy calculations.

In Table 2 and Fig. 1 DFTC HB-cluster results are displayed for the 14 clusters previously considered *in vacuo*.

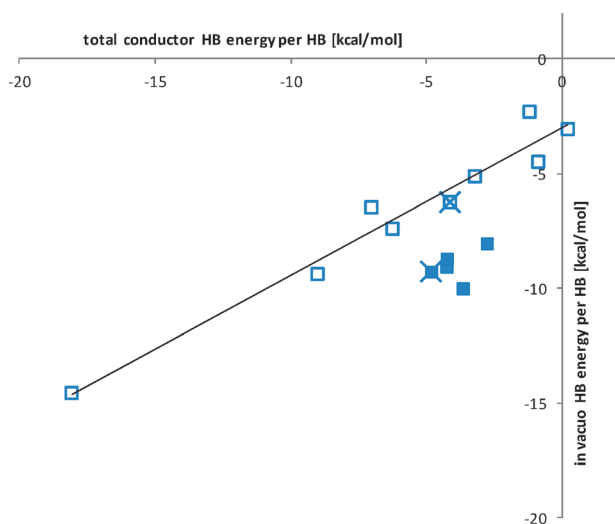


Fig. 2 Total HB energies per HB *in vacuo* vs. total conductor HB energies per HB. The open symbols are for single HB clusters, the full symbols are for clusters with double HBs. The two crosses mark two formic acid dimers with one and two HBs, respectively.

As expected, within the conductor reference state the total HB energy is almost entirely covalent. The non-covalent part is essentially zero, if dispersion is included, and even slightly positive if we exclude dispersion. This reflects the fact that in the vdW-distance complex the electrostatics is less favorable compared to a complete conductor embedding of both educts.

In Fig. 2 we show a comparison of the total HB energies *in vacuo* vs. the total conductor HB energies. In this plot we have normalized the HB energies by the number of HB of the complex. Despite the different partitioning into covalent and non-covalent contributions the *in vacuo* HB energies and the conductor HB energies correlate quite well with each other for the single HB complexes, with a slope of $\sim 2/3$. The stronger dependence of the conductor HB energies on the donor and acceptor polarity arises from the stronger polarization in the conductor environment. The five double HB complexes clearly behave very differently. In these cases the HB energy gain is much larger *in vacuo* than it is in the conductor state. This can be demonstrated quite well by the example of the mono and double HB formic acid dimers marked by a cross in Fig. 2. While in conductor the HB energy per HB of the double HB complex in conductor is only 17% larger, it increases about 50% *in vacuo*, *i.e.*, the cooperative effect of multiple HBs is much more pronounced *in vacuo* than it is in conductor. This is because parts of the cooperative polarization effect are already covered by the embedding conductor.

Predictive quantification of hydrogen bond energies of single hydrogen bonds based on conductor polarization charge densities σ

While the previous sections provide important preparative considerations, the main goal of our project was the investigation of the dependence of liquid phase HB energies on the product of the local conductor polarization charge

densities σ_{don} and σ_{acc} . Based on plausibility arguments, a linear dependence of the HB free energies was successfully used within the COSMO-RS solvation theory since about 14 years, but a theoretical or quantum chemically corroboration is still lacking. Since individual hydrogen bond free energies can neither experimentally separated from the other molecular interaction, nor can they be rigorously calculated by quantum chemistry, we decided to analyze the QC derived HB interaction energy. Since the starting point of COSMO-RS anyway is the conductor state, and since according to the previous considerations we found the L2 level to be most suitable for HB energy calculations in the condensed state, we choose DFTC calculations on the L2 level as our QC reference for a large scale investigation. Being aware of the complications which may arise from sterical hindrance, steric constraints due to multiple hydrogen bonding, and from conformational changes which may be induced by hydrogen bond formation, we have constructed and generated a large set of hydrogen bond clusters by the following procedure:

For acceptor or donor atoms (AoD), identified by a minimum of the screening charge density $\text{abs}(\sigma) > 1e \text{ nm}^{-2}$ present in our COSMObase database of common chemicals, which was used as reservoir of test molecules, and for all donor or acceptor (DoA) atoms of a set of small hydrogen bond probe molecules (donors: HF, H₂O, HCN, HCP, H₂S; acceptors: NH₃, PH₃, H₂O, H₂S, HCN):

- Find the σ -hotspot (positive for acceptors, negative for donors) on the DFTC surfaces of AoD and DoA, respectively, and merge the two molecules to an initial cluster by placing these hotspots on top of each other.
- Do a DFTC (DFTC) geometry optimization from this start geometry.
- Check, whether the final geometry has a hydrogen bond between the initially preconditioned AoD and DoA atoms. Otherwise skip.
- Check, whether multiple hydrogen bonds are formed. If yes, skip.
- Check whether steric contacts between other atoms or probe and test molecules are present. If yes, skip.
- Check for dissociation, *i.e.* whether the donor hydrogen atom finally is closer to the acceptor atom than to its initial bond partner. If yes, skip.
- Check, whether an equivalent HB-cluster was already generated by a previous combination of AoD and DoA. If yes, skip.
- Do a final single-point DFTC calculation with the def2-TZVPD basis set and calculate the DFTC HB energy as difference of the cluster DFTC energy and the sum of the donor and acceptor molecule DFTC energies.
- Store the DFTC HB energy together with the polarization charge densities σ_{acc} and σ_{don} and the elements of the acceptor atom and of the bond neighbor of the hydrogen donating atom.

This procedure is illustrated in Fig. 3. In this way we ended up with 2465 DFTC HB clusters covering a wide range of chemically different donors and acceptors and the entire range of hydrogen bond strengths, limited by dissociation at the upper end. The complete data set is given in the ESI† as Table SI2.

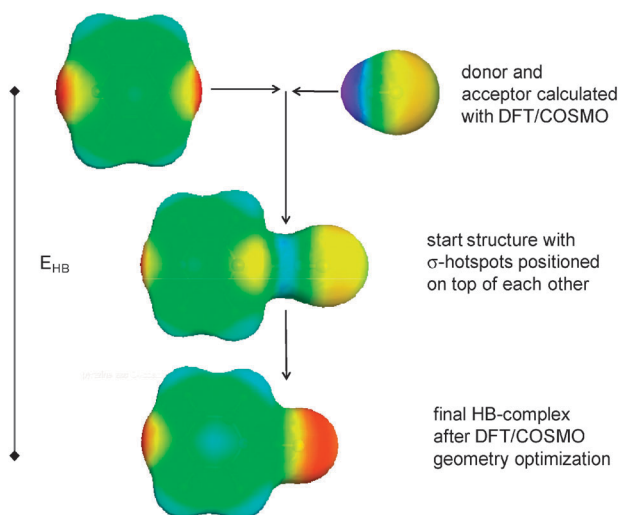


Fig. 3 Schematic visualization of the DFTC HB energy calculations for the case of a pyrazine hydrogen fluoride complex: starting point are the individual DFTC structures of donor and acceptor, with their conductor polarization charge densities σ , as visualized on their surfaces. Donor and acceptor σ -hotspots are placed on top of each other, and a full DFTC optimization leads to the final hydrogen-bonded complex. The color change from yellow to red shows the strong polarization of the HF fluorine atom during HB formation.

In order to investigate the dependence of hydrogen bond energies on the donor and acceptor polarization charged densities we plotted the DFTC hydrogen bonds for subsets with constant donor and acceptor probes *vs.* the partner polarization charge densities σ_{acc} and σ_{don} , respectively. Fig. 4 shows the DFT HB energies for the two subsets of

complexes with a strong donor (HF) and a weak donor (HCN), plotted *versus* the polarity σ_{acc} of the HB acceptor. Apparently, a single linear function of the DFTC HB energy with respect to σ_{acc} correlates the data for each probe donor to a surprising degree of accuracy across the complete range of HB strengths, for both weak and strong HB donor partners. On more detailed examination, it appears that the oxygen acceptors show a slightly lower slope with respect to σ_{acc} than the other acceptors. Not surprisingly the HB energies of the weak probe donor HCN are much less negative than those of the strong donor HF. They even reach up to $+2 \text{ kcal mol}^{-1}$, which means that the hydrogen bond complex only is a local minimum which is less stable than the two individually conductor embedded educts. The degree of linearity which is evident in both curves of Fig. 3 over the entire range of HB strengths supports the heuristic COSMO-RS assumption of a bilinear dependence of the HB energy on the σ_{acc} and σ_{don} to a surprising degree.

The inset of Fig. 4 shows a plot of the HB energies of HF-donor complexes, *i.e.* the same data as the lower curve in Fig. 2, but now plotted *versus* a commonly used polarity descriptor for HBs, the molecular electrostatic potential (MEP) at the acceptor position. The poor correlation with respect to MEP_{acc} shows that the good correlation achieved with respect to σ_{acc} is not a triviality and proves that σ is the much better descriptor for local molecular interactions than the widely used MEP.

Fig. 5 shows an analogous plot to Fig. 4, now for a fixed strong acceptor (NH_3) and a weak acceptor (HCN), respectively, for a variety of HB donors. Within each class of donors, that is for each hydrogen donating element, the HB energy appears to be linear with respect to σ_{don} , but for each class of donors a different regression is required.

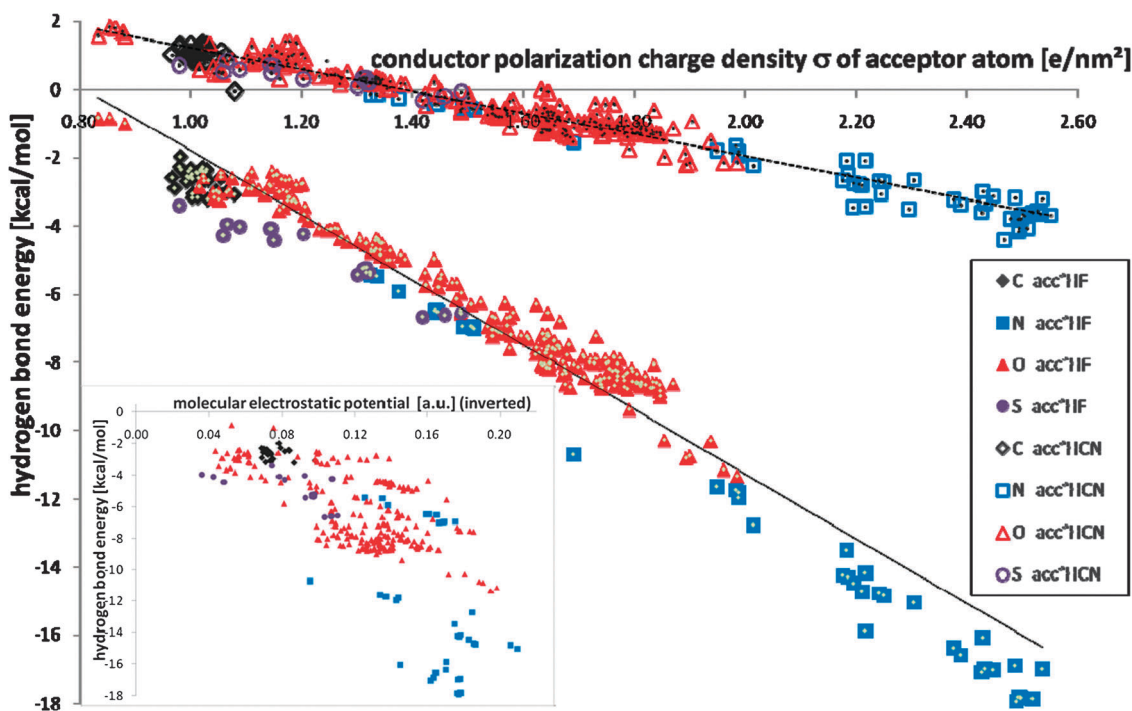


Fig. 4 DFTC HB energy plotted *vs.* the acceptor polarization charge density σ_{acc} of the acceptor atom for the very strong donor hydrogen fluoride (HF) and the weaker donor hydrogen cyanide (HCN). A linear correlation is apparent. The inset shows the same HF data plotted *vs.* the corresponding (negative) molecular electrostatic potential.

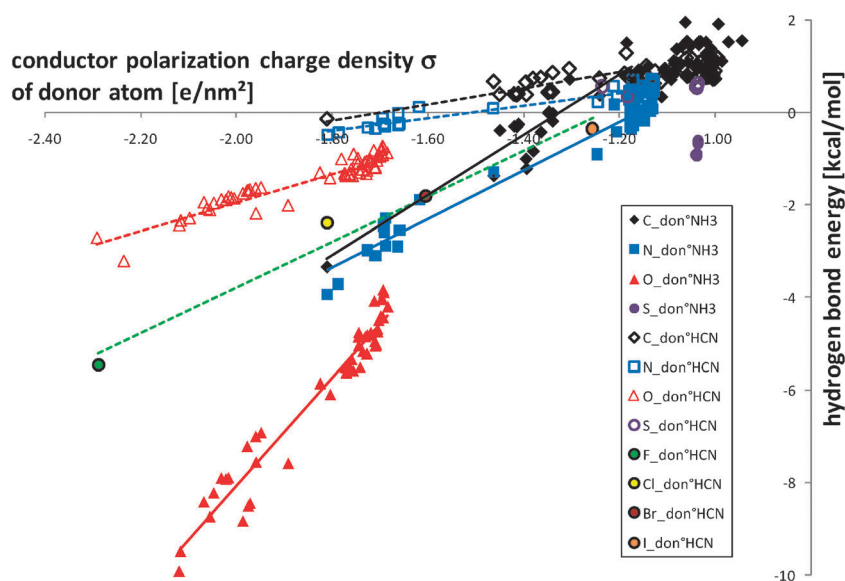


Fig. 5 DFTC HB energy plotted vs. the polarization charge density σ of the donor atom for the very strong acceptor ammonia (NH₃) and the weak acceptor hydrogen cyanide (HCN). A linear correlation is apparent within each donor type.

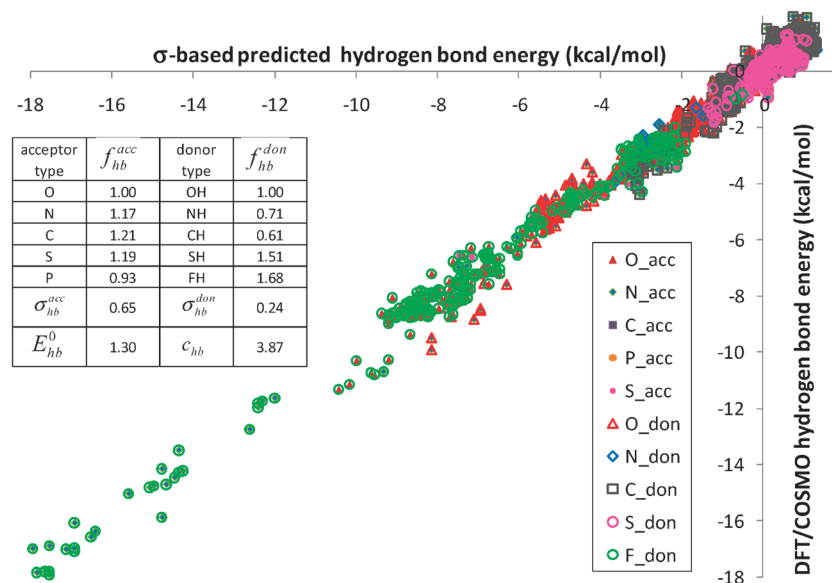


Fig. 6 Scatter plot of the DFTC HB energy vs. the predicted HB energy based on eqn (1). The fitted parameters are given in the inset.

The OH-donor regression lines show a much higher slope of the NH and CH donor types. The four halogen donor types appear to fall on one line with slightly higher slope than that of the oxygen donors for the HCN acceptor, while all halogen donors suffered from deprotonation with NH₃. In good agreement with the ratio of the acceptor polarization charge densities, the regression slopes for the weak acceptor HCN are roughly 60% of the corresponding slopes for the strong acceptor NH₃.

Based on these findings we fitted all 2465 HB energies to the following model equation:

$$E_{hb}(\sigma_{don}, e_{don}, \sigma_{acc}, e_{acc}) = E_{hb}^0 + c_{hb} + f_{hb}^{don}(e_{don})(\sigma_{don} + \sigma_{hb}^{don}) + f_{hb}^{acc}(e_{acc})(\sigma_{acc} - \sigma_{hb}^{acc}) \quad (1)$$

There are two parameters for each element e , one σ -scaling factor for donor and acceptor, respectively, plus four general constants. The scaling factors of oxygen are fixed to 1 in order to avoid parameter interdependence. The fit yields a RMSD of 0.36 kcal mol⁻¹ (Fig. 6).

Conclusions

QC calculations on DFT level with sufficiently large basis sets and with empirical correction for the DFT dispersion lack have been shown to reproduce high-level reference QC reference calculations of *in vacuo* HB energies within 0.7 kcal mol⁻¹ error (RMSD), which most likely is within the error bars of the reference calculations themselves. As a result larger scale studies on hydrogen bonding can be done on such computationally moderately expensive level.

In detail we have established here the BP/def-TZVP//BP-D3/def2-TZVPD as a sufficiently accurate DFT level for studying HB energies *in vacuo* for our purposes.

By introduction of a virtual vdW-HB contact state we have demonstrated that *in vacuo* the covalent, *i.e.* the really quantum chemical part of HB energies only accounts for 0%–60% for weak to strong hydrogen bonds, respectively, while a large part of the total energy gain of hydrogen bonded complexes *in vacuo* comes from long range electrostatics, mutual polarization and mutual dispersive interactions of the hydrogen bond partners.

We have introduced the conductor reference state, or conductor state, as a suitable reference state for hydrogen bonding in condensed phases, because the conductor embedding well mimics the fact that in most condensed phases the donor and acceptor molecules are already electrostatically well embedded and as a result pre-polarized. Starting from this situation they form stronger hydrogen bonds than *in vacuo*. In the conductor reference the hydrogen bond energy is entirely of covalent nature, without contamination arising from long range electrostatics contribution. For single hydrogen bonds a surprisingly good correlation between vacuum reference HB energy and conductor reference HB energy was found, which originates from the fact that the long-range electrostatic HB energy part, the covalent HB energy part, and the increase of the covalent HB energy due to pre-polarization to first order all scale with the product of the polarities of donor and acceptor.

The local conductor polarization charge density hotspot values σ_{don} and σ_{acc} of the donor and acceptor atoms have been shown to be excellent descriptors for a predictive quantification of conductor HB energy, and a simple linear dependence, as empirically postulated within the COSMO-RS theory, seems to fit the DFTC HB energies within their error bars. The molecular electrostatic potential MEP has been demonstrated to be a less suitable descriptor for HB energies than the conductor polarization charge density σ .

Discussion

We consider this HB energy model to be an excellent predictive model which allows the estimation of sterically non-hindered HB energies of a broad variety of neutral donor–acceptor pairs in solution. Our model has a number of advantages over other approaches for quantifying HB energy. To the best of our knowledge, no other descriptor based HB model achieves a correlation close to the one presented here. Due to the virtual conductor environment and reference state, the HB energies calculated here are not contaminated with long-range electrostatic components and moreover they are closer to the aqueous or polar solvent environments in which hydrogen bonding usually is important, especially for biochemical applications. The prediction is based only on individual DFTC information for donors and acceptors, which can be performed with moderate computational expense for molecules up to ~ 100 atoms. For larger molecules, advantage can be taken of the locality of the conductor polarization charge density, which allows for fragment-wise evaluation the σ -descriptors.

The presented model for the efficient predictive pair-wise quantification of HB energies should be of great value for

many more empirical simulation methods in which elaborate DFT calculations for each possible donor–acceptor pair are not feasible, *e.g.* for atomistic force-field simulations, association models and equation-of-state models used in chemical engineering, scoring functions in drug modeling, and in COSMO-RS fluid phase thermodynamics. In order to avoid double counting, in such models it is important to have expressions for the additional, *i.e.* covalent HB energy, because electrostatic and dispersion effects usually are accounted for by separate expression. In most force-fields the hydrogen bond energy is currently expressed based on partial charges and often merged into the electrostatic and effective vdW-parameters. A more specific and more accurate expression for hydrogen bonds should be beneficial for the force-field accuracy. Combining the σ -based prediction of the maximum HB energy gain of a donor acceptor pair with empirical or QC derived rules for the distance and directional dependence of HB energies should allow for the development of more accurate HB force fields. Chemical engineering thermodynamics models as association or equation-of-state models often describe hydrogen bonding by association sites, but usually they are lacking descriptors in order to assign quantitative values to the different pairs of association sites. The presented σ -based HB energy model provides a means in order to predictively assign such values and thus reduces the need for parameter fitting in such simulations. Also scoring functions in drug design, which are estimating the binding energies of drug candidate molecules to enzyme receptor pockets, could be put on a more rational and potentially more accurate basis by using the proposed σ -based HB energy model.

For the COSMO-RS liquid phase thermodynamics model an expression very similar to eqn (1) has been used for the hydrogen bond free energy since 12 years. The results of the current study confirm the intuitively assumed bi-linear dependence on the polarization charge densities of donors and acceptors, although it had been a motivation for the present study to find a potentially different functional form which would have helped to improve the currently used HB energy expressions of COSMO-RS. Despite the confirmed bi-linearity, our study shed some clearer light on the donor and acceptor element specificity, and this may help to improve future COSMO-RS parameterizations.

In all applications of the presented σ -based HB energy model it must be kept in mind that this is a model for the HB energy, excluding any zero-point vibrational enthalpy corrections and not taking into account the considerable entropy loss of the donor hydrogen atom due to the narrow HB energy minimum compared to the much wider minima resulting from electrostatic and dispersive forces at typical vdW intermolecular distances. Experimentally it is well known that the HB entropy loss at room temperature may cause a free energy increase in the order of 50% of the HB enthalpy gain. Hence it is crucial to take into account this contribution in any application which aims for free energies, as it is the case for chemical engineering models, scoring functions and COSMO-RS, while in force field based molecular dynamics or Monte Carlo the entropy loss may be taken into account implicitly by the sampling of molecular configurations, if the force-field does reflect the more narrow HB minimum. Empirically we may assume that the HB entropy loss can be described as a linear functional of the

HB energy, *i.e.* by a constant and a term proportional to HB energy. With this assumption the HB free energy still would be a bi-linear functional of the donor and acceptor polarization charge densities, and only two additional parameters need to be fitted to experimental data.

Outlook

In this study we restricted ourselves to the HB energy of single-hydrogen bonds of neutral compounds. Since hydrogen bonding is also very important for ions, we will consider neutral to ion and ion to ion HB complexes in a forthcoming study. Initial work in this direction seems to indicate very similar relationships for conductor reference HB energies, while any vacuum reference HB energy study would suffer from the overwhelming electrostatic contribution as soon as ions come into play. Another direction of future studies will be the cross influence of multiple hydrogen bonds, *i.e.* cooperative effects. A third issue of further investigations will be the HB entropy loss which to the best of our knowledge has not been considered systematically in the literature so far.

References

- 1 G. C. Pimentel and A. L. McClellan, *The Hydrogen Bond*, W. H. Freeman and Co., San Francisco, 1960.
- 2 G. Jeffrey and W. Saenger, *Hydrogen Bonding in Biochemical Structures*, Springer, Heidelberg, 1991.
- 3 E. Arunan, G. R. Desiraju, R. A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, H. G. Kjaergaard, A. C. Legon, B. Mennucci and D. J. Nesbitt, *Pure Appl. Chem.*, 2011, **83**, 1619–1636.
- 4 K. Wendler, J. Thar, S. Zahn and B. Kirchner, *J. Phys. Chem. A*, 2010, **35**, 9529–9536.
- 5 P. Jurečka, J. Šponer, J. Černý and P. Hobza, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1985–1993.
- 6 A. Klamt and G. Schüürmann, *J. Chem. Soc., Perkin Trans. 2*, 1993, 799–805.
- 7 A. Klamt, *J. Phys. Chem.*, 1995, **99**, 2224–2235.
- 8 A. Klamt, *COSMO-RS: From quantum Chemistry to Fluid Phase Thermodynamics and Drug Design*, Elsevier, Amsterdam, 2005.
- 9 A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100; J. P. Perdew, *Phys. Rev. B: Condens. Matter*, 1986, **33**, 8822–8824.
- 10 J. P. Perdew, *Phys. Rev. B: Condens. Matter*, 1986, **33**, 8822–8824.
- 11 A. Schäfer, C. Huber and R. Ahlrichs, *J. Chem. Phys.*, 1994, **100**, 5829–5835.
- 12 A. Schäfer, H. Horn and R. Ahlrichs, *J. Chem. Phys.*, 1992, **97**, 2571–2577.
- 13 D. Rappoport and F. Furche, *J. Chem. Phys.*, 2010, **133**, 134105.
- 14 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 15 TURBOMOLE V6.3 2011, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007. <http://www.turbomole.com>.
- 16 R. S. Paton and J. M. Goodman, *J. Chem. Inf. Model.*, 2009, **49**, 944–955.
- 17 J. Tomasi, B. Mennucci and R. Cammi, *Chem. Rev.*, 2005, **105**, 2999–3094.
- 18 A. Bondi, *J. Phys. Chem.*, 1964, **68**, 441–451.