

Implementation and Optimization of DFT-D/COSab with Respect to Basis Set and Functional: Application to Polar Processes of Furfural Derivatives in Solution

Roberto Peverati and Kim K. Baldridge*

University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received July 15, 2009

Abstract: The implementation, optimization, and performance of DFT-D, including the effects of solvation, has been tested on applications of polar processes in solution, where dispersion and hydrogen bonding is known to be involved. Solvent effects are included using our *ab initio* continuum solvation strategy, COSab, a conductor-like continuum solvation model, modified for *ab initio* in the quantum chemistry program GAMESS. Structure and properties are investigated across various functionals to evaluate their ability to properly model dispersion and solvation effects. The commonly used S22 set with accurate interaction energies of organic complexes has been used for parametrization studies of dispersion parameters and relevant solvation parameters. Dunning's correlation consistent basis sets, cc-pV_nZ ($n = D, T$), are used in the optimization, together with the Grimme B97-D exchange-correlation functional. Both water ($\epsilon = 78.4$) and ether ($\epsilon = 4.33$) environments are considered. Optimized semiempirical dispersion correction parameters and solvent extent radii are proposed for several functionals. We find that special parametrization of the semiempirical dispersion correction when used together in the DFT-D/COSab approach is not necessary. The global performance is quite acceptable in terms of chemical accuracy and suggests that this approach is a reliable as well as economical method for evaluation of solvent effects in systems with dispersive interactions. The resulting theory is applied to a group of push–pull pyrrole systems to illustrate the effects of donor/acceptor and solvation on their conformational and energetic properties.

Introduction

The treatment of van der Waals (vdW, dispersive) interactions is an active field of research as of late, particularly in the DFT community.^{1–25} The importance of vdW interactions is clear when considering molecular systems whose structures are largely influenced, and in some cases totally determined, by dispersion interactions. Moreover, the stability of a given molecular charge distribution may also be affected by its environment, specifically by the polarity of the medium or solvent. In particular, aromatic or π -conjugated molecules display a special sensitivity toward the redistribution of electron density as a function of their nuclear position, photonic excitation, and polar solvation. In general, chemical

recognition, whether in materials or biological systems, depends on physio-chemical modulation that may be determined either qualitatively or quantitatively. Molecules in close proximity create competition and selection based on relative free energies of reaction (complexation/association), excitation (change in electronic state), and/or product formation (new chemical bond). Proper treatment of dispersion and solvent effects, both important effects on their own, in combination still leave open several important questions for reliable treatment.

Given that dispersion energies are pure electron correlation effects,^{26,27} accurate computations via *ab initio* based wave functions, including at least single and double substitutions, can provide good accuracy with adequate basis sets (*e.g.*, CCSD(T) method),²⁸ albeit with a significant computational cost. On the cheaper side of these methods, Møller–Plesset

* Corresponding author phone: +41 44 635 4201; fax: +41 44 635 6888; e-mail:kimb@oci.uzh.ch.

perturbation (MP2, MP4) methods,²⁹ while an improvement over most density functional approaches, tend to be largely over bound for many systems.^{30,31} In general, neither computational solution is economical, particularly for large molecular systems and complexes. Recent literature including work of our own, on the other hand, illustrates the effectiveness of dispersion corrected DFT approaches.^{7,14,15,32–34} Such methods, coupled to effective strategies for treatment of solvent,^{35–42} promise to provide an economical yet accurate method where both solvent and dispersion effects dominate. One approach has already been shown in a recent paper by Riley et al.⁶ where a slightly modified version of the semiempirical correction of Grimme¹⁴ applied to the TPSS density functional⁴³ was coupled with the IEF-PCM approach for continuum solvation.⁴⁴ However, still not well understood is a more general view of such DFT-D functionals together with continuum solvent representations, including optimal representations of the dispersion correction, effects of basis set, and representations of the solvent.

Our goals in this work include a) contribution to the development of semiempirically corrected density functionals, emphasizing functionals and basis sets that are necessary for investigations of structure and properties in solution, b) implementation, optimization, and performance analysis of the semiempirically augmented density functional theory (DFT-D) in conjunction with our *ab initio* implementation of the conductor-like screening model, COSab, within the GAMESS software,⁴⁵ and c) illustration of the combined DFT-D/COSab method on systems with known dispersion and solvation phenomenon, presented with an expanded set of basis sets, functionals, and solvent dielectrics, with consideration of both structure and properties.

Computational Methods

All calculations reported here were carried out using a locally modified version of the GAMESS electronic structure program. In the present work, we consider the semiempirically corrected functional, B97-D,¹⁴ and two conventional hybrid functionals, B3PW91,^{46,47} and B3LYP.^{46,48} Additionally, comparisons are made with the second order Møller-Plesset perturbation theory (MP2)²⁹ and CCSD(T).²⁸ Dispersion corrections were recently implemented and tested in the GAMESS software, using the ansatz proposed by Grimme (2006).¹⁶ The B97-D functional is a special reparameterization of the original B97 hybrid functional of Becke.²² The new parameters in the functional form make B97-D less susceptible to spurious dispersion contamination in the exchange component than the original functional of Becke. The dispersion energy in B97-D is entirely handled by the semiempirical term, leading to smaller errors. For semiempirically corrected functionals, given the importance and dependence of the associated parameters on the choice of functional and basis set, we have carried out parameter optimization for several basis sets, including TZV2P,⁴⁹ with (2d,2p), and Dunning's correlation consistent basis sets,⁵⁰ denoted cc-pVnZ, where $n = D$ for double with [3s2p1d] contraction, T for triple with [4s3p2d1f] contraction, and Q for quadruple with [5s4p3d2f1g] contraction. The S22

reference set of data provided by Jurecka et al.⁵¹ is used as a validation test set, in addition to several test systems.

Solvation was taken into account using the most recent implementation of our COSab solvation method, based on the original COSMO theory of Klamt and modified for *ab initio* theory.^{52–54} Dielectric permittivities of water ($\epsilon = 78.4$) and ether ($\epsilon = 4.335$) are used for parametrization studies. An expanded set of dielectrics was used to investigate a set of substituted pyrroles, for illustration of the method, including toluene ($\epsilon = 2.38$), chloroform ($\epsilon = 5.0$), t-butanol ($\epsilon = 12.0$), acetone ($\epsilon = 20.7$), methanol ($\epsilon = 32.6$), and water ($\epsilon = 78.4$). The parameters of the cavity construction are 1082 points for the basic grid and 92 segments on the complete sphere. The outlying charge error was taken into account via the double cavity approach.^{53–55} The solvent radial extent was optimized in the parameter optimization studies and taken as 1.3 for the application studies. Solvent atomic radii were taken from Bondi⁵⁶ or from Klamt.⁵⁷

Theoretical Approach and Discussion

DFT Dispersion Corrections. Much work has been carried out to revise and enhance density functionals for increased accuracy, the most important component of late being the addition of corrections for dispersion. A particularly effective and consistent strategy to correct failures due to lack of dispersion has involved addition of an empirical potential^{19,58} to the final DFT energy. This empirical potential is typically of the form $C_6 R^{-6}$, where R are interatomic distances and C_6 dispersion coefficients. Such a strategy has been formulated and well established by several research groups.^{14,15,17,59} In particular, the method of Grimme establishes a semiempirical dispersion correction that can be applied to the final result of any mean field calculation that lacks a sufficient description of dispersion energy.^{60,61} Typically, this procedure is used in conjunction with density functional theory (leading to various DFT-D methods)

$$E_{\text{disp}} = -s_6 \sum_{i=1}^{N_{\text{nat}}} \sum_{j=i+1}^{N_{\text{nat}}} \frac{C_6^{ij}}{R_{ij}^6} f_{\text{dmp}}(R_{ij}) \quad (1)$$

where the dispersion coefficients C_6 are calculated from the atomic polarizabilities, and the f_{dmp} is a damping function with the form

$$f_{\text{dmp}}(R_{ij}) = \frac{1}{1 + \exp \left[-d \left(\frac{R_{ij}}{s_R R_{ij}^0} - 1 \right) \right]} \quad (2)$$

This semiempirical correction does not depend on the electronic structure of the system, and, therefore, optimization of the relevant parameters must be considered, in particular the s_6 prefactor to the C_6 atomic coefficients, the d damping factor, and the s_R prefactor to the vDW radii. The atomic polarizabilities (which lead to the C_6 coefficients), the vDW radii, and all semiempirical parameters were optimized in the original paper of Grimme,³² and further work was carried out by our own group more recently.³⁴ Such parameters have to compensate for differences in the electronic descriptions due to basis sets, functionals, and, in the present work, effects

of solvent. The nonlinear parameter, d , is considered optimal and fixed to the optimized value of Grimme, $d = 20$.¹⁴ Differences in implementations are mainly in the values of these parameters. The final mean-field (MF) DFT-D energy is the result of the direct addition of this dispersion energy to the computed DFT energy

$$E_{\text{MF-D}} = E_{\text{MF}} + E_{\text{disp}} \quad (3)$$

First Principles Continuum Solvation Theory. We have developed a rigorous self-consistent approach for the inclusion of electrostatic solvation effects in conventional *ab initio* gas phase computations in GAMESS.^{35,52} This model is implemented at the RHF (ROHF, UHF), DFT, HDFT, and MP2 (UMP2) levels, including first and second derivatives. Our approach to continuum solvent modeling (COSab) arises from the concepts of screening in conductors, a modification of the original COSMO theory of Klamt.⁶² Significant model features of this approach include a) molecular-shaped cavity construction, b) efficient solution to energy and derivative quantities, c) efficient ways of including dynamic correlation, d) inclusion of an alternate approach using multipoles up to hexadecapoles, e) assessment of outlying charge effects using a choice of two strategies, f) general electrostatic solvation accuracy within 2–6 kJ/mol, and g) several strategies for inclusion of first solvation shell effects.

We refer the details of the underlying theory of the COSab method to the original paper⁵² and instead only point out the main concepts related to the coupling of this theory with the semiempirical dispersion corrections for density functional treatments. The treatment of solvent effects within the continuum model involves the construction of a cavity around a solute system, with a dielectric continuum representing the solvent outside the cavity. The molecular electric field arising from the nuclei and the electronic distribution is screened by the polarization of this dielectric continuum. The iterated solute/solvent interaction can be uniquely represented by the resulting surface charge distribution at the interface to the continuum. Construction of the molecular cavity relies on van der Waals radii of Bondi⁵⁶ scaled by a solvent factor (\cos_{rad}) in the range of 1.2–1.3 that represents the area of approach of solvent molecules to a solute.⁵⁷ These radii in fact ensure that the cavity volumes are approximately equal to the molar volumes of the compounds and therefore are not really free parameters^{57,63} and demonstrated the physically correct screening energy decay behavior.⁵²

While authors of other solvent models may recommend different sets of vdW radii,^{64–66} there is a general agreement that the cavity size is closely related to the vdW surface. Recently, Truhlar et al.⁶⁷ provided a systematic study showing variation across four popularly used solvent models, including a discussion of what radii provide the most physical results and also showing that scaling factors greater than ~1.4 produce unphysical cavities. These important insights in fact imply that, for some molecules (e.g., those with diffuse electronic structure), a small but significant portion of the solute electron density can extend outside the molecular cavity. This leads to error in the prediction of solvent phenomenon, since the solute is not fully represented in the proper way in its interaction with the solvent. Such errors

are termed *outlying charge errors* and are most problematic for systems with diffuse charge localized or anionic systems and can lead to errors that are substantial (e.g., 40–80 kJ/mol error). Our present solvation implementation has two effective strategies for the treatment of outlying charge error.^{53,68}

It is important to note that the vdW radii involved in the damping function in eq 2 are different than those used in the continuum solvation method, as both are optimized independently using different approaches and for different properties. As such, one cannot use the same set of vdW radii values for both contributions. The model for solvation is in fact independent to that of the semiempirical dispersion model. Any eventual dependencies of the two effects would nevertheless be accounted for through the parametrization of the semiempirical scaling factors within both methods.

Together, the DFT-D+COSab method has a total of 4 parameters to consider for any interdependencies and for optimization. Three parameters, scaling factor s_R and s_6 (vdW radii and dispersion coefficients, respectively), and damping function exponent, d , arise from the DFT-D method. The fourth parameter, \cos_{rad} , is related to the creation of the solvent cavity surface. Following the approach used in our previous work (as well as others),^{14,34} the value of d is kept fixed at the value of 20. This leaves investigation into the behavior of only three parameters, s_6 , s_R , and solvent extent radii (\cos_{rad}), simultaneously with respect to basis set and functional.

Parameter Optimization. To perform simultaneous optimization of the three chosen parameters, a suitable database of molecules with accurate solvated interaction energies is required. The S22 set, proposed by Hobza et al.,⁵¹ is the standard database used in all previous considerations of the DFT-D method. However, this database provides accurate interaction energies only in vacuum. To obtain solvated interaction energies, the following term, first proposed by Riley et al.,⁶ is used to correct the vacuum energies

$$\Delta E_{\text{solv}}^{\text{CP}} = \Delta E_{\text{vac}}^{\text{CP}} + \Delta \Delta E_{\text{solv}} \quad (4)$$

In this way, a value for $\Delta \Delta E_{\text{solv}}$ can be calculated, using a suitable solvation method, to give

$$\Delta \Delta E_{\text{solv}} = \Delta E_{\text{solv}}^{\text{no-CP}} - \Delta E_{\text{vac}}^{\text{no-CP}} \quad (5)$$

Using this approach, one can propose a way forward for providing chemical accuracy in solution for a reasonable costs and including important effects of dispersion and solvation. In addition, this method provides a reasonable way to take basis set superposition error (BSSE) into account in solution. This is very important given that continuum solvation methods do not provide a clear and unique way to take into account BSSE in the calculations, since there is no analogue to the counterpoise CP corrections as used in gas phase. Alternatively, since the optimization of the s_6 parameter for the B97-D functional including counterpoise correction shows a minimum at $s_6 = 1.4$ regardless of basis set, one could carry out CP corrected solvation energies using $s_6 = 1.4$.

Computational solvent models are fundamentally different at the algorithmic level.⁴⁰ This results in different accounting of the various aspects of solvation representation within the quantum mechanics ansatz. For example, models have different (or missing) representations for cavity descriptions, solvent radii, outlying charge effects, nonelectrostatic effects, or other solvent related phenomenon. In particular, comparison with experimental data has shown substantial difference in predictability with different solvent cavity descriptions, which are a function of solvent radii parameters.⁶⁹ The radii used to generate the spheres associated with the cavity ultimately govern the volume of the cavity and the distance between the atoms and the solvent surface. Computed properties depend heavily on the choice of atomic radii.

The specific formalism used by Riley et al.⁶ involved the use of the Integral-Equation-Formalism Polarized Continuum Model, IEF-PCM, together with the use of united atom, UA0, solvent radii. The general PCM method uses a cavity of interlocking spheres approximately 20% larger than the vdW radius. The electrostatics are treated using the Poisson equation, with Green's functions used to define the integral operators determining the apparent charge. Nonelectrostatic effects are evaluated empirically, using a solvent accessible surface (SAS). In this way, the surface charge depends only on the potential and not on the vector electric field, thereby being less sensitive to numerical instabilities. The specification of the solute cavity, defined using the united atom topological model (UA0), is obtained from spheres centered on non-hydrogen atoms. The radius of each sphere, which is dependent on the atom type, connectivity, and number of hydrogen atoms attached, is based on the Universal Force Field (UFF).⁷⁰ This particular choice of radii used for cavity construction (default for cavity construction in other quantum chemistry software) has been found to lead to relatively high mean unsigned errors for solvation phenomena^{69,71} and is typically not recommended for quantitative calculations of solvation free energies.^{72–74}

Given the known sensitivities to details of solvation model, it becomes necessary to recalculate solvation reference corrections for the S22 set of molecules for this work. Here, we have calibrated a set of corrections for both high and low dielectric, water ($\epsilon = 78.4$) and ether ($\epsilon = 4.33$), respectively. Calculations are performed at the MP2/aug-cc-pVDZ level of theory similar to Riley et al.⁶ for comparative purposes. Results are reported in Table 1.

Optimization of the Solvent Extent Parameter (\cos_{rad}).

The algorithm for cavity construction in the COSab-GAMESS method involves construction of a basic grid representing the molecular cavity generated from the vdW surface around the molecule. This is followed by a series of projections onto the solvent surface, extending the cavity to approximately a distance of van der Waals + 20–30%.^{56,57} The cavity segmentation is small enough as to assume homogeneous charge distribution on each segment. The standard value for the parameter, \cos_{rad} , is typically taken as 1.2 or 1.3. Here, we actually investigate the response in the range from 1.1 to 1.5 to better understand the influence of this parameter on the prediction of solvation.

Table 1. MP2/aug-cc-pVDZ Solvent Reference Energy Corrections (kJ/mol) for the S22 Set of Molecules

S22 complex reference number	$\Delta\Delta E_{\text{solv}}$ MP2/COSab (this work)		$\Delta\Delta E_{\text{solv}}$ MP2/ IEF-PCM ^a		
	vacuum	ether $\epsilon = 4.33$	water $\epsilon = 78.4$	ether $\epsilon = 4.33$	
				water $\epsilon = 78.4$	
1	-13.26	-3.347	2.301	-4.48	-0.54
2	-21.00	-14.81	-11.21	-14.35	-9.87
3	-77.86	-50.58	-36.65	-32.93	-10.42
4	-66.78	-37.53	-20.67	-33.30	-17.40
5	-86.40	-50.71	-30.79	-39.50	-16.44
6	-69.92	-43.30	-28.83	-36.36	-20.88
7	-68.49	-43.76	-31.25	-36.32	-21.67
8	-2.22	-1.97	-1.84	-2.13	-2.13
9	-6.32	-4.35	-3.39	-4.39	-3.68
10	-6.28	-5.10	-4.48	-5.36	-4.64
11	-11.42	-7.11	-5.10	-7.36	-5.23
12	-18.49	-14.39	-11.89	NA ^b	NA ^b
13	-42.34	-23.81	-12.76	-16.40	-7.95
14	-21.84	-14.64	-11.30	-14.85	-12.09
15	-51.17	-34.31	-24.31	-32.09	-24.90
16	-6.402	-2.720	-0.92	-2.55	-0.50
17	-13.72	-7.24	-3.64	0.38	7.03
18	-9.832	-6.99	-5.44	-7.20	-0.08
19	-18.66	-9.83	-4.60	-10.96	-5.56
20	-11.46	-9.54	-8.58	NA ^b	NA ^b
21	-23.97	-16.11	-11.67	NA ^b	NA ^b
22	-29.50	-21.05	-16.78	-15.61	-7.45

^a Reference 6. ^b Riley et al.⁶ did not report this value because of problems in the convergence of the energies.

A first set of calculations was performed at the B97-D+CO Sab/cc-pVDZ level of theory, with the dispersion parameters fixed at $d = 20$, $s_R = 1.1$, and $s_6 = 1.0$. Further optimization was undertaken of the more critical s_6 parameter for each value of \cos_{rad} to confirm results. Figure 1 and Table 2 summarize these results.

The variation in the mean absolute deviation of the S22 set for the \cos_{rad} parameter in the range 1.1–1.5 is within 0.45 kJ/mol for water and 0.13 kJ/mol for ether. An optimal value of $\cos_{rad} = 1.30$ is suggested for two reasons. First, the performance is slightly better for the simultaneous optimization of \cos_{rad} and s_6 (see black dots on the graphic). Second, 1.30 is one of the possible values suggested for the CO Sab method (typical values are 1.20 and 1.30). The general message, however, is that values in the range 1.3 ± 0.2 do not affect the general behavior of the method, which could be surprising given the known sensitivity of electrostatic solvation properties to variation in atomic radii, as discussed above.

Optimization of DFT-D Parameters (s_6 and s_R). As in our previous work,³⁴ we investigate the combined behavior of the two DFT-D parameters, s_6 and s_R , using the optimal value determined for $\cos_{rad} = 1.3$. The simultaneous optimization of these two parameters is presented in Figure 2 as a surface function for three different environments, vacuum, water, and ether, was carried out. Results were investigated for the B97-D functional with both the cc-pVDZ and cc-pVTZ basis sets and are illustrated in Figure 2.

The topology of the curves for the three environments (vacuum, ether, H₂O) appear very similar to each other but with different relative height on the MAD axis. This similarity in topology suggests that behavior in solvated environments for these complexes with variation in disper-

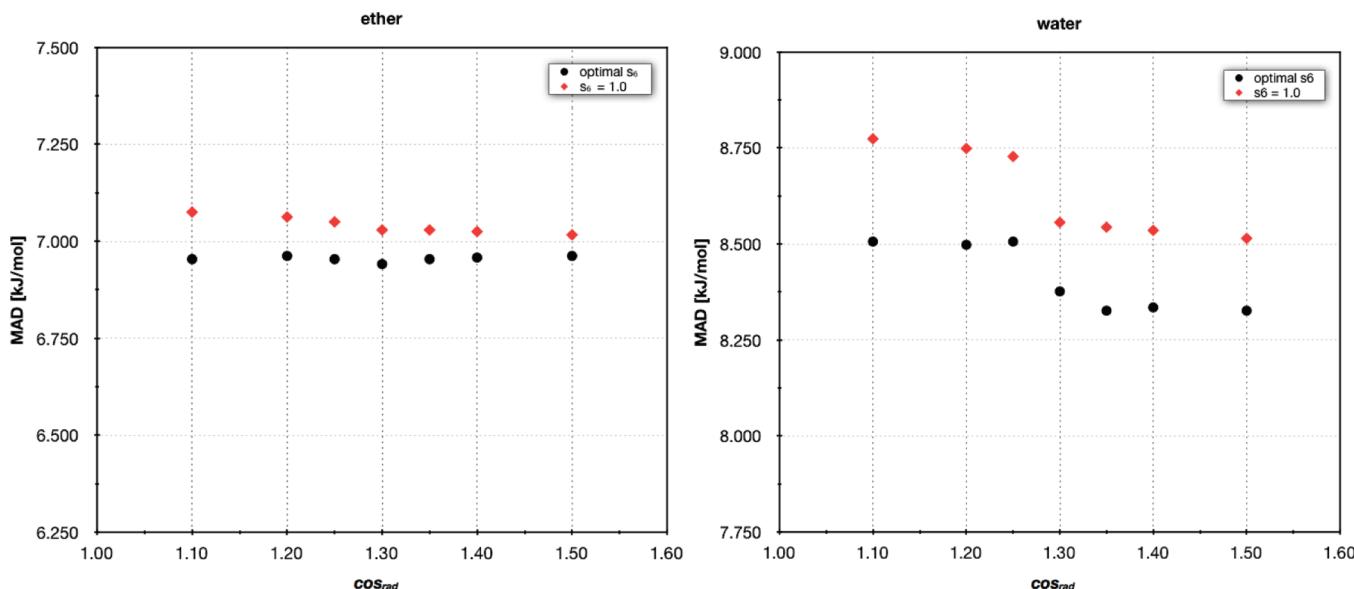


Figure 1. Mean absolute deviation of S22 set of molecules as a function of solvent radii extent value (\cos_{rad}) at $s_6 = 1.0$ (red), and at the optimal s_6 for the respective \cos_{rad} (black), at the B97-D+COSab/cc-pVDZ level of theory.

Table 2. Summary of Results of Mean Absolute Deviation (MAD, kJ/mol) for the S22 Set Showing Variance in Solvent Radii Extent Parameter, \cos_{rad} , as a Function of s_6 , at the B97-D+COSab/cc-pVDZ Level of Theory

solvent extent parameter, \cos_{rad}	water			ether		
	MAD with $s_6 = 1.0$	corresponding optimal s_6	MAD with optimal s_6	MAD with $s_6 = 1.0$	corresponding optimal s_6	MAD with optimal s_6
1.1	8.774	0.93	8.506	7.075	0.97	6.954
1.2	8.749	0.94	8.498	7.063	0.97	6.962
1.25	8.728	0.95	8.506	7.050	0.97	6.954
1.3	8.556	0.96	8.376	7.029	0.98	6.941
1.35	8.544	0.95	8.326	7.029	0.98	6.954
1.4	8.535	0.95	8.334	7.025	0.98	6.958
1.5	8.514	0.96	8.326	7.017	0.98	6.962
variance	0.01326	-	0.00784	0.00048	-	0.00005
std. error	0.04352	-	0.03346	0.00824	-	0.00270

sion functional parameters is qualitatively the same. The differences regarding the MAD from that of the gas phase are expected due to the fact that these were determined using MP2/aug-cc-pVDZ in the respective environment, rather than at CCSD(T)/CBS. However, because we are interested here in the optimal values of the s_6 and s_R , if the overall topology in different environments is truly within a small tolerance of that observed for the gas phase environment, this difference in height along the MAD axis can be tolerated.

To compare more closely the variance in surface topology in different medium, we can project one curve onto either of the other two and determine any significant point-to-point differences. These differences are plotted as a surface function of the two DFT-D parameters in Figure 3, with a color density map projected on the bottom of each diagram, for all three possibilities: water vs vacuum, ether vs vacuum, and water vs ether. Only the B97-D+COSab/cc-pVDZ case is shown here, since similar results are obtained at the B97-D+COSab/cc-pVTZ level of theory. In these surface representations (Figure 3), blue (negative) means that the solvated curve was lower than the vacuum (ether) curve in Figure 2, while red (positive) means that the solvated curve is higher than the vacuum (ether) curve in Figure 2.

The deviations from zero (red or blue, Figure 3) are in all cases smaller than 4.2 kJ/mol and in general very small for both basis sets, confirming the fact that the curves overlap almost perfectly. Moreover, the intersection of the optimized s_6/s_R parameters is always in a minimum deviation region. From these results, we can conclude that the method behaves in the same way for both solvents, when the optimization of the two DFT-D parameters is conducted simultaneously. When solvated results are compared to the vacuum case, this general conclusion does not change.

A more detailed slice of the surface functions is shown in Figure 4 for $s_R = 1.1$. Numerical values for the corresponding s_6 optimizations are reported in Table 3. As can be seen, the values of the optimal s_6 parameter for the solvated cases differ only slightly from that used in the vacuum, for each of the basis sets considered. Additionally, these results also include indirectly, consideration of BSSE, as shown above. Therefore, our overall conclusions made in our previous article, including discussion of BSSE, more extensive optimization, and summary of values proposed as a function of basis set, can be extended with good accuracy to the DFT-D + COSab procedure, for calculation of interaction energies in solution. As such, these results, together with

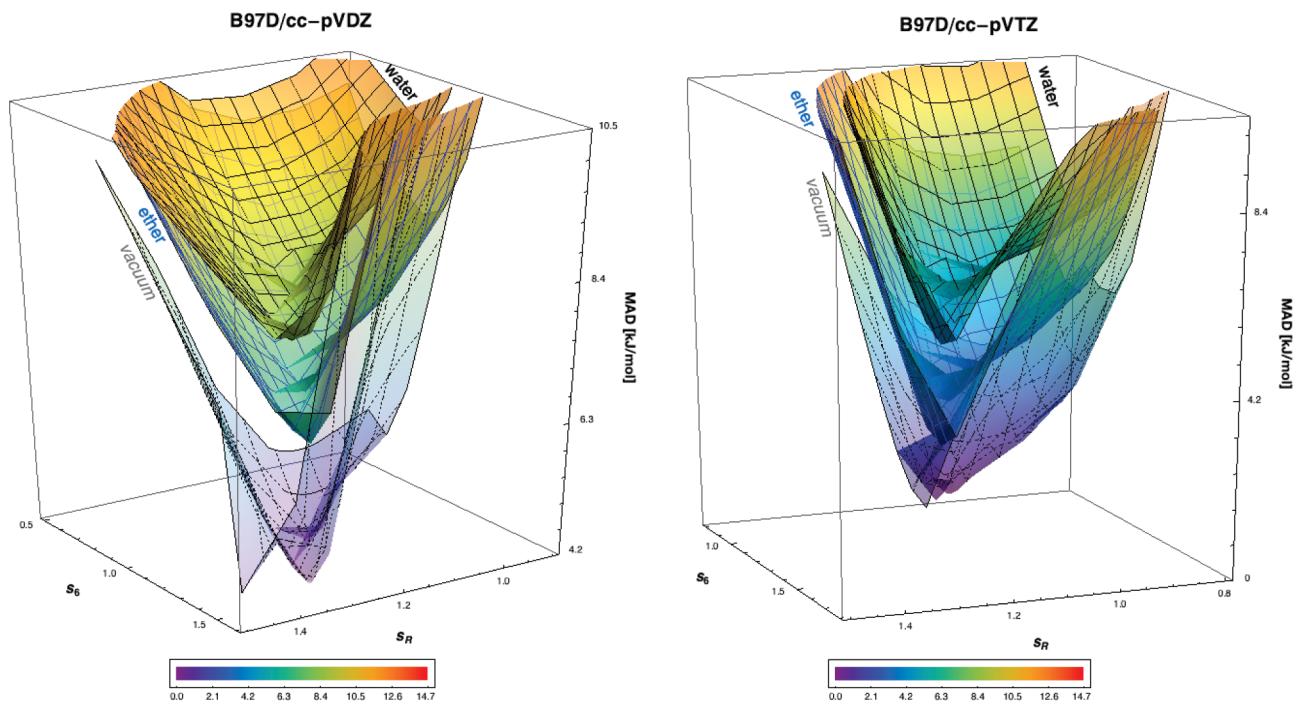


Figure 2. Optimized s_6 and s_R empirical dispersion function parameters for the B97-D functional, with cc-pVDZ and cc-pVTZ basis sets, respectively. Plots are given on the same relative scale.

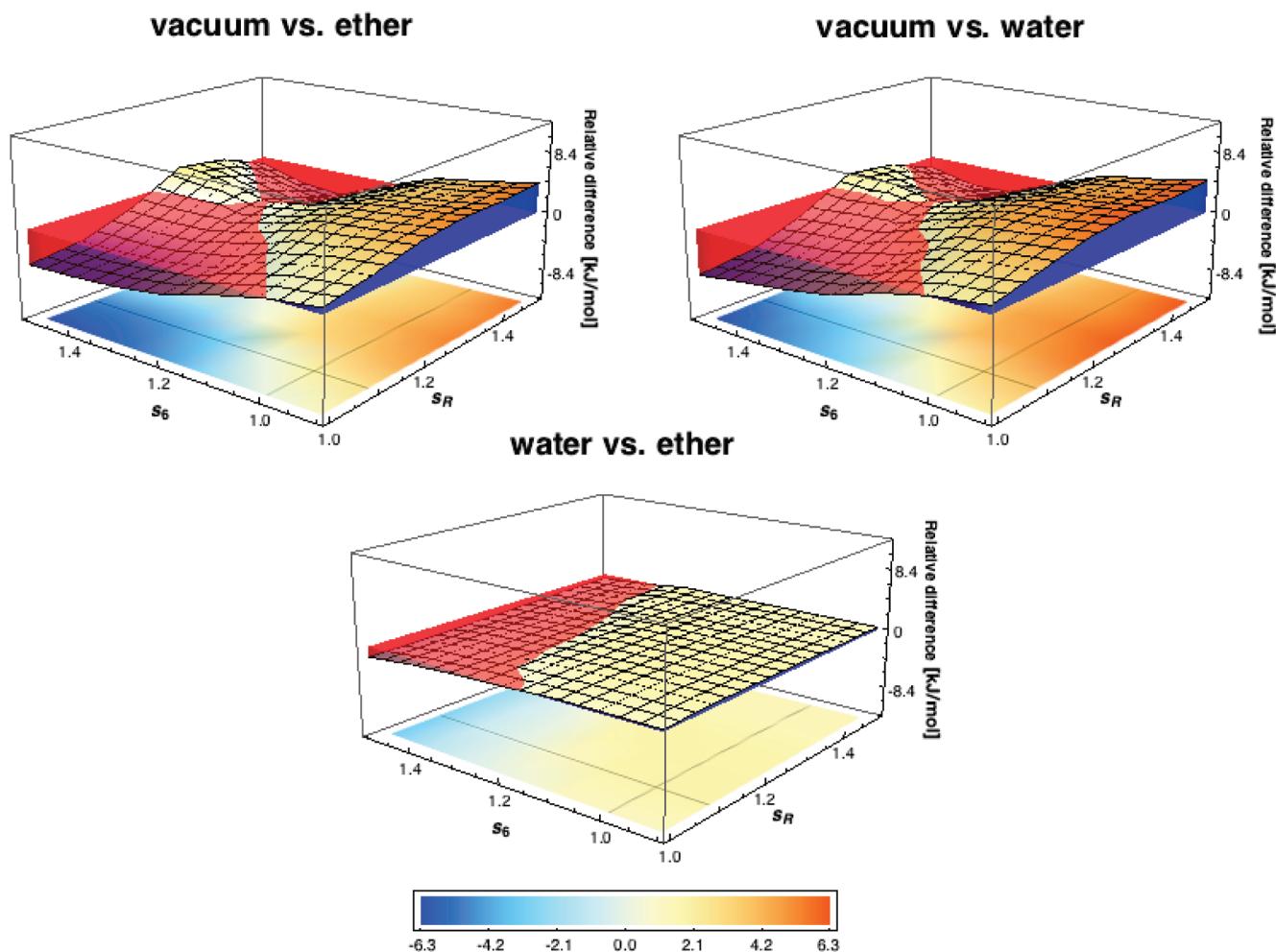


Figure 3. Water vs vacuum, ether vs vacuum, and water vs ether projection plots, at the B97-D+CO Sab/cc-pVDZ level of theory.

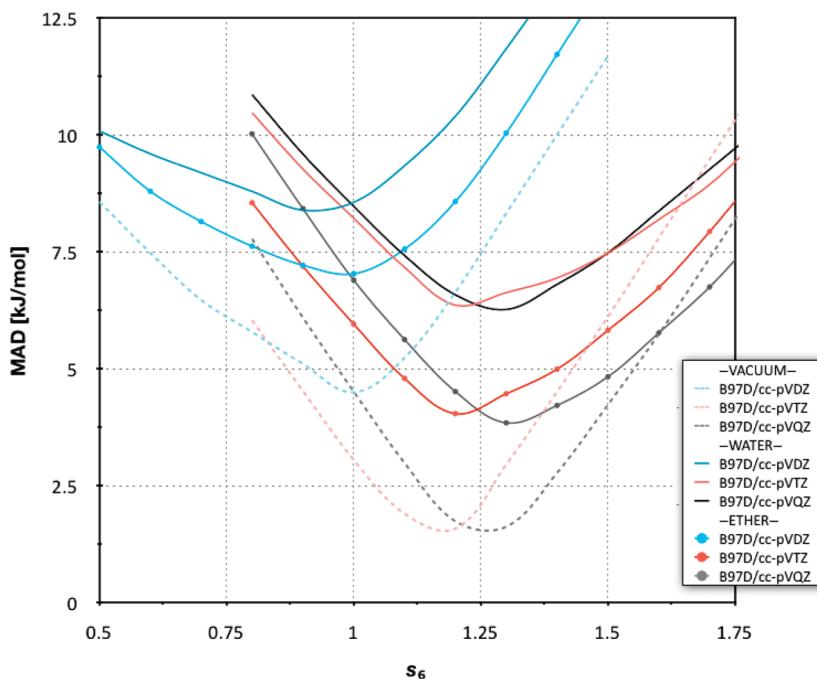


Figure 4. Water vs vacuum, ether vs vacuum, and water vs ether projection plots at the B97-D+COSab/cc-pVDZ level of theory, represented as a slice through the surface functions from Figure 3.

Table 3. Summary of Results of Mean Absolute Deviation (MAD, kJ/mol) for the S22 Set Showing Variance in s_6 with Solvent and Basis Set^a

solvent	basis-set	optimized s_6	MAD [kJ/mol]
vacuum	cc-pVDZ	1.00	4.498
	cc-pVTZ	1.18	1.410
	cc-pVQZ	1.26	1.381
water	cc-pVDZ	0.96	8.376
	cc-pVTZ	1.21	6.356
	cc-pVQZ	1.28	6.192
ether	cc-pVDZ	0.98	6.941
	cc-pVTZ	1.19	4.038
	cc-pVQZ	1.28	3.820

^a $\cos_{rad} = 1.3$, $s_R = 1.1$, $d = 20$.

those of Riley et al.⁶ using a different ansatz for including solvation effects together with semiempirically corrected DFT, support that in general, for any treatment of dispersion-governed systems in solution, no additional considerations need be made nor specialized optimizations undertaken.

Geometry Optimization. The DFT-D method is typically parametrized using single point energy calculations on an accurate gas phase geometry, for each complex in the S22 database. This procedure opens up questions regarding any variations in methodology and subsequent prediction of results, when full optimization of geometry is considered instead of a single point energy calculation. In the case of a vacuum environment, these questions are not considered significant due to the fact that the starting geometry is calculated at a very high level of theory. More specifically, the optimized geometry for most of the complexes in the S22 set has been determined with CCSD(T)/QZ, with a few cases optimized with CCSD(T)/TZ or MP2/TZ-CP. All final single point energies are then calculated at the CCSD(T)/CBS level. Explicit details of the optimizations used in each case can be found in the original paper.⁵¹

In contrast, it is not readily clear that the gas phase optimized geometries are the most appropriate when one is considering a solvent environment, as one can expect changes in the structures of the complexes due to their interactions with the solvent. In particular, one expects a change in distance between the two monomers in the complex. While the main purpose of this article is not to search for better procedures of parametrization of the DFT-D method, one still must consider possible sources of error due to changes in geometry and the consequences for the DFT-D + COSab procedure.

In order to carry out such an analysis, we have calculated the gradient for all molecular systems in the S22 set with the B97-D functional across a variety of basis sets, in a water environment. The root-mean-square deviation (RMSD) and the largest component (MAX) of the gradient are reported in Table 4, as a measure of the differences between the geometries as given in the S22 set and their respective optimized equilibrium geometries.

These summarized values for rmsd and MAX gradient indicate that, provided one uses a large enough basis set, a large majority of the complexes are not far from their gas phase optimized geometries. As expected, an incremental increase in basis set size provides a more accurate picture of just how close the solution phase complexes are to the optimized gas phase geometry. Cases where there are the greatest rmsd values, an indication of when solvation changes are expected to be significant, are highlighted in boldface in Table 4. For example, complexes that have bond length differences in the picometer range represent significant differences when one is trying to obtain chemical accuracy in structure and associated properties.

To further analyze these results, complexes in the S22 set were divided into three subgroups according to the nature of the primary interactions between the molecules. The three

Table 4. Summary of Results of RMSD and Maximum (MAX) Gradient Values for the S22 Set As a Function of Basis Set

molecule	cc-pVDZ		cc-pVTZ		cc-pVQZ	
	RMSD gradient	MAX gradient	RMSD gradient	MAX gradient	RMSD gradient	RMAX gradient
1	0.0077984	0.0174154	0.0026740	0.0053823	0.0022662	0.0041095
2	0.0105507	0.0191442	0.0056189	0.0124492	0.0051123	0.0120283
3	0.0083584	0.017122	0.0046930	0.0097907	0.0040162	0.0090413
4	0.0090051	0.0234469	0.0062381	0.0166286	0.0058922	0.0159450
5	0.0078098	0.0249773	0.0055591	0.0178010	0.0053169	0.0165761
6	0.0058952	0.0203691	0.0036121	0.0149223	0.0035065	0.0139418
7	0.0068203	0.0221163	0.0044578	0.0146959	0.0042735	0.0130864
8	0.0055312	0.0107665	0.0020459	0.0040859	0.0018506	0.0037130
9	0.0054494	0.0098256	0.0025617	0.0035783	0.0024370	0.0034128
10	0.0052723	0.0124369	0.0022704	0.0058819	0.0021843	0.0056609
11	0.0044198	0.0094194	0.0016463	0.0035123	0.0015786	0.0032221
12	0.0047943	0.0101749	0.0019266	0.0039867	0.0021137	0.0040012
13	0.0075948	0.0222007	0.0055311	0.0150658	0.0052161	0.0132548
14	0.0042266	0.0099944	0.0017430	0.0037012	0.0016530	0.0037704
15	0.0061554	0.0210383	0.0038124	0.0139875	0.0036075	0.0122721
16	0.0066904	0.0164377	0.0025598	0.0047258	0.0026196	0.0057154
17	0.0054453	0.0134867	0.0023467	0.0076936	0.0021896	0.0070529
18	0.0057686	0.0179676	0.0022156	0.0057129	0.0020522	0.0045684
19	0.0049909	0.0141728	0.0063532	0.0313533	0.0068787	0.0340759
20	0.0044762	0.0109458	0.0017516	0.0041108	0.0016882	0.0038493
21	0.0043044	0.0105029	0.0018935	0.0044323	0.0018214	0.0041939
22	0.0049381	0.0122261	0.0022128	0.0071765	0.0021013	0.0065132
average	0.0061953	0.0157358	0.0033511	0.0095761	0.0031989	0.0090911
maximum value	0.0105507	0.0249773	0.0063532	0.0313533	0.0068787	0.0340759

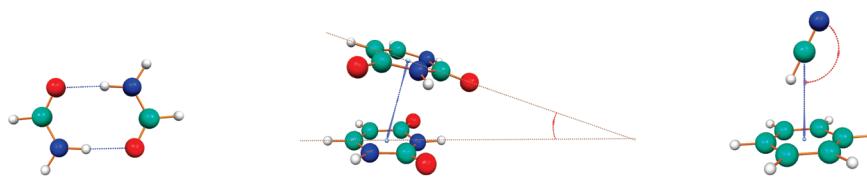
categories of complexes are A) hydrogen-bonded complexes (1–7), B) predominant dispersion interactions, complexes (8–15), and C) mixed hydrogen bonding and dispersion, complexes (16–22). Complexes with the highest rmsd and MAX gradient within each group are indicated in boldface in Table 4. In each category, we have chosen one complex to illustrate results of a full optimization in solution at the B97-D/cc-pVnZ ($n = D,T,Q$) level of theory, starting from the accurate gas-phase geometry. In category A, H-bond complexes, formamide dimer (#4) is chosen, with the highest rmsd gradient at cc-pVTZ and cc-pVQZ, and the second highest at cc-pVDZ. For category B, the dispersion complexes, the uracil dimer stack (#13) is the complex with highest rmsd and MAX gradient with all considered basis sets. Within category C, the mixed complexes, benzene-HCN (#19) has the highest rmsd and MAX gradient with cc-pVTZ and cc-pVQZ and also for cc-pVDZ. Data for the fully optimized solvent (water) structures and energetics for these three complexes as compared to gas phase are illustrated and summarized in Table 5.

The most sensitive geometry parameters in these complexes are the intermonomer distances and the angular relationship between the two monomers. Table 5 summarizes these values as a function of basis set and environment for the three complexes considered. The geometry taken from the S22 set has been optimized in the gas phase and in water for the basis sets shown. In the gas phase, each complex shows a different behavior in the variance of the interatomic distance with respect to increase in basis set. The intermonomer distance in the formamide dimer increases significantly with increase in basis set, with a convergence to the S22 reference value ($\Delta\text{Reference} = 0.032 \text{ \AA}$, 0.001 \AA , and 0.005 \AA , respectively). Complex 13, uracil dimer, on the other hand, shows a decrease in intermonomer distance with increase

in basis set, contrary to the reference limit ($\Delta\text{Reference} = 0.008 \text{ \AA}$, 0.021 \AA , and 0.086 \AA , respectively). Addition of diffuse functions does not change this trend. The angular relationship between the two monomers is, however, similar to that predicted at the reference level. Complex 19 shows an increase in intermonomer distance with basis set but away from the reference value of 3.387 \AA ($\Delta\text{Reference} = 0.031 \text{ \AA}$, 0.024 \AA , and 0.043 \AA , respectively). The angular dependence in the benzene-HCN system is relatively the same across basis sets, with a total variance of $\sim 5^\circ$.

Moving from gas phase to solution phase results in a significant change in geometry but again different trends for the three different cases. In the case of the formamide dimer, the hydrogen bond distance increases by almost 6 pm in going from the gas to the solution phase. It is known that high dielectric solvents have a dramatic effect in the dimerization of formamide.^{75–77} The increase in the polarity of the solvent is thought to reduce the dimerization, likely due to competition from the surrounding medium. This then causes the opportunity for other bonded forms, from the rather well-defined double hydrogen bonded form in the gas phase, to single hydrogen-bond dimers (medium dielectric) and less well-defined structures in water.

In the second class of complexes, the uracil dimer, the change in environment from gas phase to solution phase shows an even more dramatic increase in the distance between the two monomers (e.g., $\sim 15 \text{ pm}$ for cc-pVQZ). The uracil dimer can exist in both hydrogen-bonded configurations as well as various stacked arrangements.⁷⁸ The stacked and T-shaped conformations, with a large dispersive energetic component, are less stable than the hydrogen bonded configurations, albeit less so in the gas phase than in the solution phase.

Table 5. Comparison of Optimized Vacuum, Optimized Solvent (Water), and Reference S22 Gas Phase, Geometry Parameters for Three Complexes of the S22 Set Using Different Basis Sets^a


		vacuum			water		
bonds	S22 reference CCSD(T)/CBS	cc-pVDZ	cc-pVTZ	cc-pVQZ	cc-pVDZ	cc-pVTZ	cc-pVQZ
#4	1.841	1.809	1.842	1.846	1.860	1.901	1.905
#13	3.365	3.357	3.344	3.279	3.516	3.459	3.434
#19	3.387	3.418	3.412	3.430	3.279	3.226	3.180
		vacuum			water		
angles	S22 reference CCSD(T)/CBS	cc-pVDZ	cc-pVTZ	cc-pVQZ	cc-pVDZ	cc-pVTZ	cc-pVQZ
#4	-	-	-	-	-	-	-
#13	13.8	18.4	18.0	14.0	19.9	13.6	12.5
#19	174.6	175.7	177.0	179.6	154.3	153.7	149.4

respective optimized vacuum geometry vs optimized solvated (water) geometry									
initial solvated rms (au) from gas phase geometry			ΔE (kJ/mol) _{vacuum-water}			%max geometry change (vacuum-water)			
	cc-pVDZ	cc-pVTZ	cc-pVQZ	cc-pVDZ	cc-pVTZ	cc-pVQZ	cc-pVDZ	cc-pVTZ	
#4	0.00905	0.00624	0.00589	1.46	4.595	3.93	1.1	3.8	3.2
#13	0.00760	0.00553	0.00522	10.70	4.51	2.50	5.9	1.6	0.3
#19	0.00499	0.00635	0.00688	16.74	19.30	18.28	8.0	16.5	3.4

^a The graphic shows intermolecular distances (blue) and angles (red), as reported in the table.

In the last case, benzene-HCN system, one sees a very large decrease in the intermonomer distance in the high dielectric water environment, changing from 3.43 Å to 3.18 Å, at the B97-D/cc-pVQZ level of theory. In addition, the angular relationship between the benzene and the HCN changes significantly, from perpendicular in the gas phase, to canted in the solution environment. Other studies have also shown that this interaction weakens with stronger dielectric, with the interaction energy in water being less than half as strong as in the gas phase (~15 kJ/mol vs ~6.5 kJ/mol).⁶

In all of these cases, taking the optimized gas phase geometry as the starting point of the solution phase optimization, one sees that the gradient of the first step is quite high (on average 0.006 au), an indication that the solution phase structure is still far from the gas phase structure. For the three molecules considered here, one sees maximum geometry changes as high as 16% from that of the gas phase, depending on the system and level of theory. Such differences between gas and solution phase can constitute significant differences in energetics, for example, if one is trying to determine interaction energies in solution environment with a high level of accuracy but is only using gas phase geometries. In just these three examples, we see over 17 kJ/mol in energy difference (at the cc-pVQZ level of theory) between gas and solution phase. In general, one must carefully consider the implications of geometry optimization

in a solution environment depending on the system and associated properties of interest.

Illustrative Example: Ground State Conformational Dynamics of Polar Processes

Despite the general tendency for molecules to adopt electronic configurations with minimal charge separation (e.g., electroneutrality principle), many important structural and dynamical features are mediated by charge-separated states. In particular, the conformational analysis of push–pull conjugated π systems is strongly influenced by contributions from zwitterionic states. In donor–acceptor systems, one can identify a dual-resonance-form model, where the major contributor is determined by the degree of stabilization of the charge-separated state. Typically, one of the two resonance forms contributes local double-bond character, manifesting a higher rotational barrier. Depending on the nature of the donor (D)/acceptor (A) pair and the effect of solvent, the barrier to rotation can span a fairly large range. In addition to kinetic effects, it is often found that the dominant equilibrium conformer varies with the medium, with the conformer of higher dipole moment generally more favored in media of high dielectric constant.

A particular case of longstanding interest is the barrier to rotation in push–pull furfural systems, especially due to their

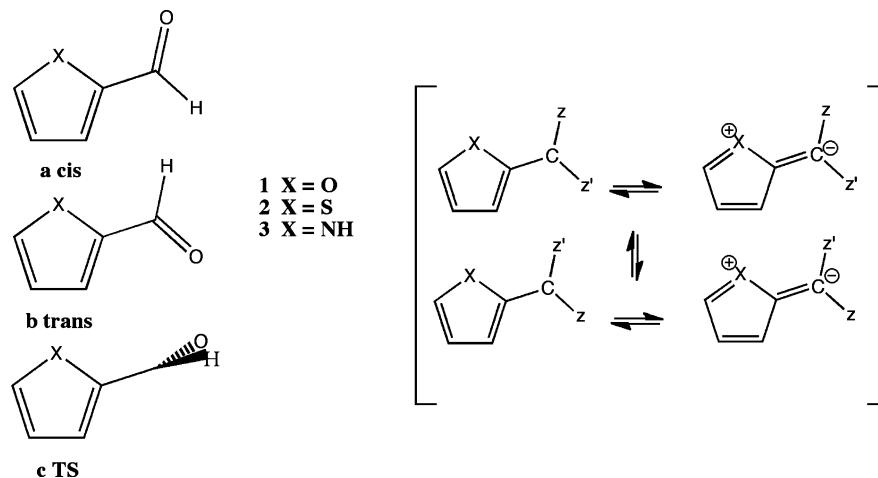


Figure 5. Conformations of furan-2-carbaldehyde ($X = O$), thiophen-2-carbaldehyde ($X = S$), and pyrrole-2-carbaldehyde ($X = NH$).

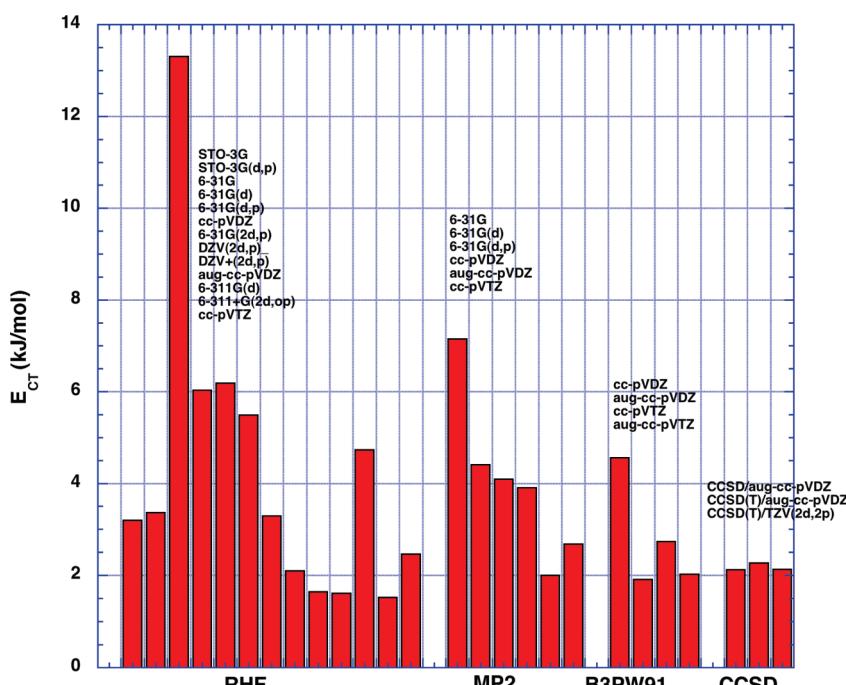


Figure 6. Rotational equilibrium of *trans*- vs *cis*-2-furanaldehyde as a function of wave function type and basis set, relative to the *trans* isomer.

importance as components in many biological redox centers and synthetic molecular devices.^{79–84} Moreover, these systems have associated experimental as well as computational challenges.⁸⁵ For example, the very small energy difference between the *cis* and *trans* isomers, the possibility for secondary interactions in some of the analogues as well as the strong solvent dependencies across a small range of dielectric (particularly for the parent furan-2-carbaldehyde system), creates particular challenges for both experimental as well as computational predictions. Unambiguous assignments of experimental spectral results have made difficult the predictions of the *cis-trans* energy difference, ΔE_{CT} , as well as the barrier to interconversion, E_a , with a wide range in experimental predictions in both (see, e.g., the Supporting Information).⁸⁶ Similar details pose challenges for computational solvent methods to accurately predict the small energy differences between isomers and pick up the solvent

detail that reproduces the conformational change as a function of dielectric. Solvent effects can actually be large enough to reverse the order of stability between the *cis* isomer and the *trans* isomer. As such, this class of system represents a particularly good test for the present method. In particular, here we investigate the parent furfural system and two analogues, as illustrated in Figure 5.

In our 2000 work, together with the experimental group of Bain,⁸⁷ we investigated a full set of push–pull conjugated pyrrole cognates. Bain and Hazendonk⁸⁷ used a combination of three NMR experiments to obtain rate data over 6 orders of magnitude representing an approximately 150 K temperature range, giving very reliable experimental numbers for the furfural system. Experimental activation parameters were obtained with errors less than 1 kJ/mol and 6 J/(mol K) for ΔH^\ddagger and ΔS^\ddagger , respectively. The ΔH^\ddagger for toluene, acetone, and methanol do not obey a simple relationship with ϵ ,

Table 6. Rotational Equilibrium (kJ/mol) of *trans*- vs *cis*-2-Furanaldehyde As a Function of Wave Function Type and Basis Set, for Gas Phase and 6 Different Solvents^{a,d}

ϵ	cc-pVDZ				aug-cc-pVDZ			
	MP2	B97-D	B3LYP	B3PW91	MP2	B97-D	B3LYP	tB3PW91
1 ^b	3.72	4.56 (4.61)	4.82 (4.93)	4.65 (4.75)	2.02	1.92 (2.31)	2.03 (2.50)	1.94 (2.40)
2.38	2.29	3.19 (3.35)	3.15 (3.37)	2.98 (3.13)	0.03	-0.10 (1.28)	0.49 (1.07)	-0.13 (0.38)
5	1.29	1.96 (2.09)	1.60 (1.86)	2.17 (2.37)	-1.47	-1.56 (-1.16)	-1.78 (-1.17)	-1.73 (-1.13)
12	0.55	1.08 (1.16)	1.17 (1.43)	1.01 (1.21)	-2.57	-2.72 (-1.78)	-2.22 (-1.57)	-3.47 (-2.87)
20.7	0.29	1.00 (1.10)	0.64 (0.83)	1.65 (1.92)	-3.06	-3.06 (-1.77)	-3.41 (-2.76)	-3.87 (-2.03)
32.6	0.15	0.88 (0.98)	0.66 (0.86)	0.65 (0.86)	-3.23	-3.44 (-2.12)	-3.83 (-3.15)	-3.94 (-3.36)
78.4	0.02	0.56 (0.65)	0.88 (1.09)	0.46 (0.69)	-3.45	-3.74 (-2.19)	-3.09 (-2.39)	-3.98 (-3.35)
average unsigned difference with respect to MP2	-	-0.70 (-0.80)	-0.66 (-0.87)	-0.75 (-0.94)	-	-0.14 (-0.90)	-0.35 (-0.61)	-0.36 (-0.43)
difference with respect to CCSD(T) ^c for $\epsilon = 1$	1.59	2.43 (2.48)	2.69 (2.80)	2.52 (2.62)	-0.11	-0.21 (0.18)	-0.10 (0.37)	-0.19 (0.27)
difference with respect to experiment for $\epsilon = 1$	0.30	1.14 (1.19)	1.40 (1.51)	1.23 (1.33)	-1.40	-1.50 (-1.11)	-1.39 (-0.92)	-1.48 (-1.02)

^a ΔE_{CT} relative to the *trans* isomer **1a** in kJ/mol. ^b Experimental gas phase value reported as 3.42 ± 0.38 .⁸⁸ ^c CCSD(T)/DZV(2d,p)//MP2/DZV(2d,p) [$\Delta E_{CT} = 2.99$, $E_a = 42.91$ kJ/mol], CCSD(T)/TZV(2d,2p)//MP2/TZV(2d,2p) [$\Delta E_{CT} = 2.13$, $E_a = 39.16$ kJ/mol], and CCSD(T)/aug-cc-pVDZ//B3PW91/aug-cc-pVDZ [$\Delta E_{CT} = 2.26$, $E_a = 41.10$ kJ/mol].⁸⁶ ^d Zero point corrected values are given in parentheses.

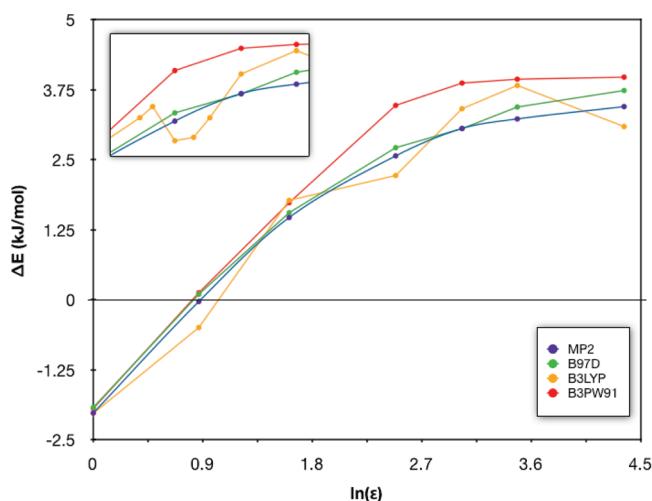
Table 7. B97-D/aug-cc-pVDZ Dipole Moments of *trans*- and *cis*-Conformers for 2-Furanaldehyde, **1**, in Gas Phase and 6 Different Solvents^b

ϵ	2-furanaldehyde			
	dipole (D)			
	cis	trans	TS	ΔE_{CT}^a , kJ/mol
1	4.46	3.85	2.82	-1.92 (-2.31)
2.38	5.29	4.57	3.24	0.10 (-1.28)
5	5.87	5.07	3.52	1.56 (1.16)
12	6.31	5.45	3.72	2.72 (1.78)
20.7	6.47	5.59	3.79	3.06 (1.77)
32.6	6.55	5.66	3.82	3.44 (2.12)
78.4	6.64	5.74	3.86	3.74 (2.19)
$\Delta(\text{gas}-\text{water})$	2.18	1.89	1.04	5.66 (4.50)

^a ΔE_{CT} relative to the *trans* isomer **1a** in kJ/mol. ^b Zero point corrected values are given in parentheses.

suggesting that the solvent effect is likely more complex than just the effect of solvent reaction field. Large ΔS^\ddagger values support this and also implied that equating ΔG^\ddagger and ΔH^\ddagger , as is often done, is not justified even for aprotic solvents. Using theoretical methods, we previously carried out a full investigation⁸⁶ of the effects of method/basis set on prediction of both ΔE_{CT} and barrier to interconversion. A summary of these results is shown in Figure 6. Results enable comparison across basis set as well as wave function type. Basis sets of at least double- ζ quality with polarization on both heavy and light atoms, and preferably diffuse functions, is shown to be optimal. Our best estimate for ΔE_{CT} (relative to the lower energy *trans* form) and barrier was determined with CCSD(T)/TZV(2d,2p)//MP2/TZV(2d,2p) [$\Delta E_{CT} = 2.13$, $\Delta E_{CT-ZPE} = 2.48$, $E_{a-ZPE} = 37.62$ kJ/mol]. This provided benchmark values, given the historical controversy regarding the experimental information. However, less expensive but with overall good predictability is the MP2/aug-cc-pVDZ level. Theoretical predictions can be compared to the 1989 IR and Raman spectroscopic study reporting $\Delta E_{CT} = 3.42 \pm 0.28$ kJ/mol and $E_a = 38.96 \pm 0.25$ kJ/mol.⁸⁸

We now compare the earlier, more computationally expensive CCSD(T) and MP2 results, with the present dispersion enabled functional, B97-D, for analogues **1–3** in

**Figure 7.** Comparison of calculated ΔE_{CT} as a function of dielectric for various wave function types using the aug-cc-pVDZ basis set. The insert figure shows an enhanced view with additional calculated points of the x axis region from $\ln(\epsilon) = 2.5$ –3.5.

both gas and solution phase. We first consider the parent, furan-2-carbaldehyde system (furfural), **1**. The energy difference between *trans*, **1a**, and *cis*, **1b**, forms of furfural in gas phase shows the *trans* conformation, **1a**, to be more stable by ~ 4 kJ/mol, with the exact value dependent on the level of theory, as previously found. In solution, the larger dipole moment of the *cis* isomer, **1b**, can preferentially stabilize this form over the *trans* conformation. In Table 6, results with B97-D/cc-pVDZ and B97-D/aug-cc-pVDZ are compared to other functionals as well as our previous MP2 and reference CCSD(T) results, across a wide range of dielectrics. Gas phase ΔE_{CT} values are overestimated with cc-pVDZ with all methods with respect to the benchmark CCSD(T) values. However, the aug-cc-pVDZ basis provides sufficiently accurate results for all considered functionals. The average unsigned difference with respect to MP2 shows that B97-D provides better results than B3LYP or B3PW91, at an overall lower computational cost than for hybrid functionals, MP2, or CCSD(T).

Table 8. B97-D/aug-cc-pVDZ Calculated Gas Phase Energetics and Dipole Moments for **1–3**

X	ΔE_{CT}	energetics ^a				dipole			
		expt	preferred tautomer in gas phase	E_a	expt	cis	trans	TS	preferred tautomer in solution phase
O	1.9	3.42 ± 0.28^b	<i>trans</i>	55.2	38.96 ± 0.25^b	4.5	3.8	2.8	<i>cis</i>
S	-3.85	4.1 ± 0.4	<i>cis</i>	45.5	-	4.1	3.7	2.8	<i>cis</i>
NH	-14.7	-	<i>cis</i>	57.1	-	3.1	4.6	2.7	<i>trans</i>

^a ΔE_{CT} relative to the *trans* isomer, **1a**, in kJ/mol. ^b Reference 88. ^c *cis* and *trans* from ref 89. TS value calculated from bond dipoles, ref 90. ^d Reference 91. ^e Reference 92.

Table 9. B97-D/aug-cc-pVDZ Dipole Moments of *trans*- and *cis*-Conformers for Thiophen-2-carbaldehyde, **2** (X = S), and Pyrrole-2-carbaldehyde, **3** (X = NH), in Gas Phase and 6 Different Solvents^a

ϵ	S						NH					
	cis	trans	TS	preference wrt dipole in solution	ΔE_{CT}	cis	trans	TS	preference wrt dipole in solution	ΔE_{CT}		
1	4.08	3.68	2.82	<i>cis</i>	3.85 (4.41)	3.07	4.64	2.74	<i>trans</i>	14.7 (14.0)		
2.38	4.88	4.39	3.26	<i>cis</i>	4.28 (4.96)	3.72	5.54	3.19	<i>trans</i>	11.7 (11.5)		
5	5.44	4.89	3.56	<i>cis</i>	4.66 (5.35)	4.18	6.19	3.49	<i>trans</i>	9.36 (9.17)		
12	5.87	5.27	3.78	<i>cis</i>	5.10 (5.82)	4.53	6.67	3.72	<i>trans</i>	7.28 (7.21)		
20.7	6.02	5.41	3.86	<i>cis</i>	5.06 (5.90)	4.66	6.84	3.79	<i>trans</i>	5.37 (5.40)		
32.6	6.10	5.48	3.90	<i>cis</i>	5.19 (5.97)	4.72	6.94	3.83	<i>trans</i>	6.26		
78.4	6.20	5.57	3.95	<i>cis</i>	4.84 (5.73)	4.80	7.04	3.87	<i>trans</i>	4.48 (4.33)		

^a ΔE_{CT} relative to the *cis* isomer, **1b**, in kJ/mol.

The general trend across the range of dielectrics (natural log scale), from toluene ($\epsilon = 2.38$) to water ($\epsilon = 78.4$) for the parent system, for a variety of wave function types, is shown in Figure 7. Notably, the B97-D and MP2 curves are very similar in predictability, while the two hybrid DFT methods show considerable variance, particularly the B3LYP method. We have carried out additional computations for the B3LYP curve around the region of $\epsilon = 5.0$, and, in fact, the prediction is shown to be continuous around these points, rather than exceptions (see, e.g., Figure 7 insert graph). This region corresponds to the cross over between preferences for *cis* over *trans* conformation. The overall dipole moment change over the span of dielectrics is $\Delta\mu = 1.9$ D for a corresponding change in $\Delta\Delta E_{CT} = 4.50$ kJ/mol. The slope of the curve on the $\ln(\epsilon)$ plot is relatively steep illustrating the strong change preference toward the *cis*-2-furanaldehyde form over the *trans*-2-furanaldehyde form in higher dielectric. While it is true that the *trans*-2-furanaldehyde form relieves the lone-pair/lone-pair oxygen repulsion that is found in the *cis*-2-furanaldehyde form, in higher dielectric the higher dipole moment of the *cis* form relative to the *trans* form will eventually dominate, changing that preference.

To further illustrate donor/acceptor properties in these systems, we have investigated thiophen-2-carbaldehyde (X = S) and pyrrole-2-carbaldehyde (X = NH). Table 8 summarizes the gas phase results for the three analogues. Unlike the parent system, thiophene-2-carbaldehyde, **2**, has a more polar S-*cis* conformation and is therefore more stable in the gas phase than is the S-*trans* conformation. Although the dipole difference between S-*cis* and S-*trans* is smaller than in **1**, one still expects preferential stabilization of the former conformation in solvent. Similarly, pyrrole-2-carbaldehyde, **3**, is also found to have a more stable *cis* conformation in the gas phase. Importantly, the less polar N(H)-*cis* form gains stabilization due to the possibility for

intramolecular interaction between the hydrogen of the nitrogen and the oxygen of the furan ring. In this case, however, unlike either **1** or **2**, the dipole moment would suggest a substantial preferential stabilization of the *trans* conformation in solution environment, due to the significantly larger dipole of the NH-*trans* structure, so the two effects, weak interactions in the *cis* form and strong dipole stabilization in the *trans* form, are competitive.

The tautomeric equilibrium in solution phase is in general found to be dependent on both ΔE_{CT} and $\Delta\mu$. Table 9 summarizes the effects of dielectric variance for analogues, **2**, S, and **3**, N(H). One sees a consistent preference for the *cis* isomer in both the calculated ΔE_{CT} as well as the predicted dipole across the full span of dielectrics. In **3**, N(H), as observed in the gas phase results, we see the competition between the hydrogen bond stabilization present in the *cis* isomer, and the stronger dipole component present in the *trans* isomer. The result is a preference for the *cis* isomer across the span of dielectric, since the weak hydrogen bond interaction will only partially compensate for the more dramatic solvent dipole effect in the N(H)-*trans* conformation.

A subtle feature of the solvation model capability is observed with system **2**, S, due to the second row element, sulfur. In particular, the detailed nature of the cavity construction algorithm is an inherent part of the quality of the solvation model. Different regions of the molecular cavity can have different sensitivity to surface discretization, depending on how rapidly varying the surface charge is, and may require a higher level of discretization to track this change properly. In the case of thiophen-2-carbaldehyde, one can observe this phenomenon. Figure 8 shows a comparison of ΔE_{CT} as a function of $\ln(\epsilon)$ for a normal cavity discretization and a highly refined cavity discretization. Checking also the effects of outlying charge and zero point energy, one sees that each of these have a consistent

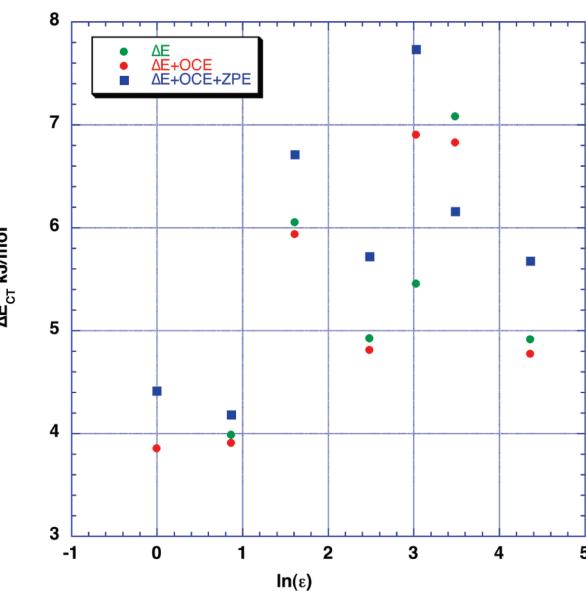
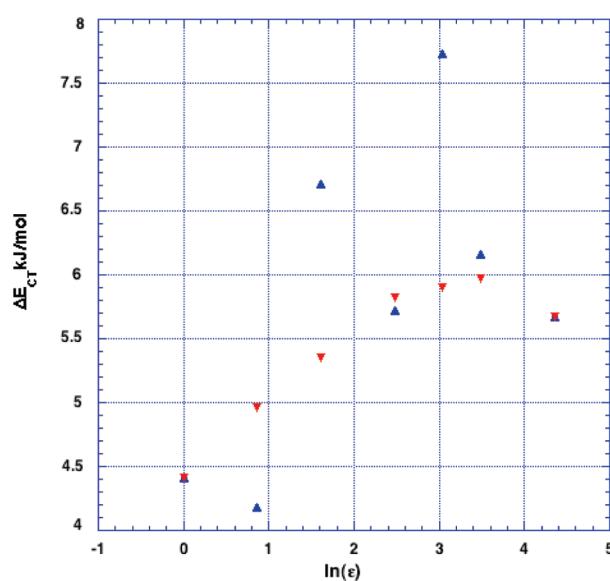


Figure 8. Variation in ΔE_{CT} as a function of dielectric for two levels of cavity discretization (a) and as a function of outlying charge and zero point energy (b) for thiophen-2-carbaldehyde, **2**.

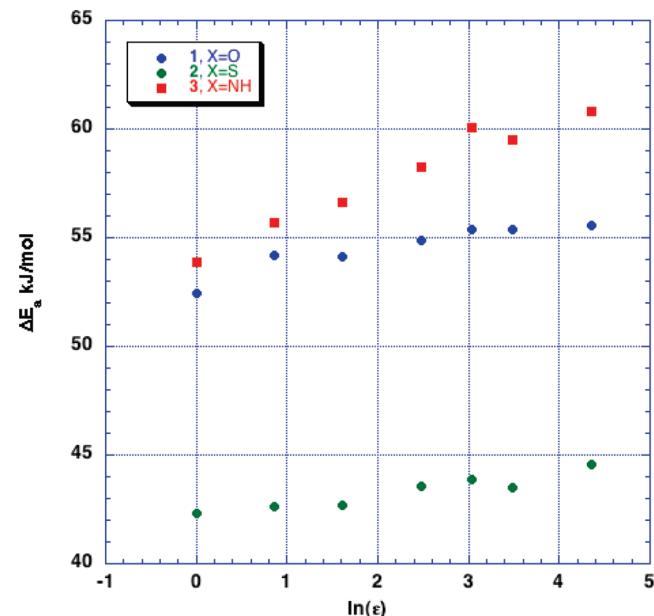
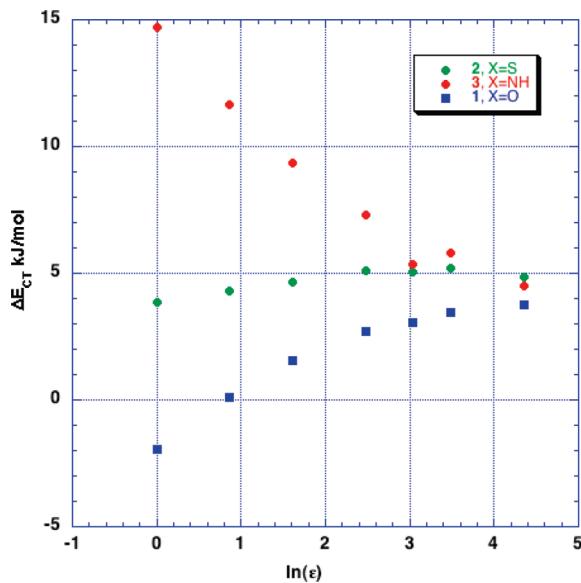


Figure 9. Comparison of calculated ΔE_{CT} as a function of cavity discretization (a) and as a function of outlying charge and zero point energy (b) for thiophen-2-carbaldehyde, **2**.

effect across dielectric and are not responsible for the inconsistent trend in the data.

Figure 9 shows comparison of solvation trends for all three analogues at the B97-D/aug-cc-pVDZ level of theory. The curves illustrate the dipole versus electrostatic trends across the span of dielectrics for the different types of functionalities. In particular, pyrrole-2-carbaldehyde **3**, shows a steep slope in the opposite direction as that of the parent furfural system, for the reasons explained above. In the sulfur substitution, the smaller change in dipole over the span of dielectrics, results in only a small increase in stabilization energy across the span of dielectrics.

Comparison of barrier height with available experiment is difficult due to the lack of experimental values for the substituted systems. For **2**, the experimentally available value of 42.27 ± 0.63 kJ/mol in CH_2Cl_2 solution^{93,94} and $43.11 \pm$

Figure 10. Comparison of calculated E_a (*cis* \rightarrow TS) as a function of dielectric ($\ln(\epsilon)$).

1.26 kJ/mol in the pure liquid⁹⁵ agrees well with the calculated barrier of 43.60 kJ/mol for $\epsilon = 12.0$ and 44.57 kJ/mol in water. Figure 10 shows a general trend of calculated E_a for the three systems studied. The increase in barrier to interconversion to the *trans* form shows steady increase for both **3** and **2**, corresponding to the stronger preference for the *cis* isomer with increasing dielectric, somewhat attenuated in the parent system as observed above. The sulfur system barrier is considerably less than the other two, again reflecting the relatively small increase in stabilization with increase in dielectric.

Conclusions

Detailed chemical treatments of molecular and electronic structure, including the effects of the environment, can offer

significant challenges unless highly accurate methods are employed. We present here an accurate and yet cost efficient way of including solvent effects with dispersion enabled density functional theory, such that a self-consistency is maintained with respect to the solvent charges and the interaction potential. Numerical values for corresponding s_6 semiempirical dispersion parameters are reported. As shown, values of optimal s_6 parameters for solvation theory differ only slightly from that used in the vacuum, for each of the basis sets considered. Results include indirectly, consideration of BSSE. The resulting DFT-D + COSab procedure can be extended with good accuracy for prediction of interaction energies in solution.

We have applied the DFT-D + COSab method to one category of push–pull conjugated heterocycles, with known challenges associated with prediction of isomer preference and barrier to interconversion in solution environment. The behavior of changing donor/acceptor on both gas and solution phase, conformational preferences, and internal rotational barrier show the importance of dispersion, hydrogen bonding, and solvation in the computational model. The dispersion corrected DFT model with appropriate basis set reproduces the trends in the known experimental data and projects trends across a large span of dielectric.

Acknowledgment. This work was supported by the Swiss National Science Foundation.

Supporting Information Available: Table of experimental relative energies of *cis*-, *trans*-, and transition state of furfural in different solvents. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References

- Zhao, Y.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 157–167.
- Benighaus, T.; DiStasio, R. A., Jr.; Lochan, R. C.; Chai, J.-D.; Head-Gordon, M. *J. Phys. Chem. A* **2008**, *112*, 2702–2712.
- Hill, J. G.; Platts, J. A.; Werner, H.-J. *Phys. Chem. Chem. Phys.* **2008**, *8*, 4072–4078.
- Silvestrelli, P. L. *Phys. Rev. Lett.* **2008**, *100*, 053002.
- Tarnopolsky, A.; Karton, A.; Sertchook, R.; Vuzman, D.; Martin, J. M. L. *Phys. Chem. A* **2008**, *112*, 3–8.
- Riley, K. E.; Vondrasek, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5555–5560.
- Schwabe, T.; Grimme, S. *Phys. Chem. Chem. Phys.* **2007**, *9*, 3397–3406.
- Grimme, S.; Antony, J.; Schwabe, T.; Mück-Lichtenfeld, C. *Org. Biomol. Chem.* **2007**, *5*, 741–758.
- Grimme, S.; Mück-Lichtenfeld, C.; Antony, J. *J. Phys. Chem. C* **2007**, *111*, 11199–11207.
- Neese, F.; Schwabe, T.; Grimme, S. *J. Chem. Phys.* **2007**, *126*, 124115.
- Grimme, S.; Neese, F. *J. Chem. Phys.* **2007**, *127*–154116.
- Grimme, S.; Steinmetz, M.; Korth, M. *J. Chem. Theory Comput.* **2007**, *3*, 42–45.
- Zhao, Y.; Truhlar, D. G. *J. Chem. Theory Comput.* **2006**, *2*, 1009–1018.
- Grimme, S. *J. Comput. Chem.* **2006**, *27*, 1787–1799.
- Jurecka, P.; Cerny, J.; Hobza, P.; Salahub, D. R. *J. Comput. Chem.* **2006**, *28*, 555–569.
- Grimme, S. *J. Chem. Phys.* **2006**, *124*, 034108–034115.
- Grimme, S. *J. Comput. Chem.* **2004**, *25*, 1463–1473.
- von Lilienfeld, O. A.; Tavernelli, L.; Rothlisberger, U. *Phys. Rev. Lett.* **2004**, *93*, 153004.
- Wu, Q.; Yang, W. *J. Chem. Phys.* **2002**, *116*, 515–524.
- Wu, X.; Vargas, M. C.; Nayak, S.; Lotrich, V.; Scoles, G. *J. Chem. Phys.* **2001**, *115*, 8748.
- Mooij, W. T. M.; van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Eijck, B. P. *J. Phys. Chem. A* **1999**, *103*, 9872.
- Becke, A. D. *J. Chem. Phys.* **1997**, *107*, 8554–8560.
- Meijer, E. J.; Sprik, M. *J. Chem. Phys.* **1996**, *105*, 8684.
- Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.
- Hobza, P.; Zahradník, R. *Chem. Rev.* **1988**, *88*, 871.
- Kutzelnigg, W. Pair correlation theories. In *Modern Theoretical Chemistry*; Schaefer, H. F., Ed.; Plenum Press: New York, London, 1977; pp 129–188.
- Kutzelnigg, W. *J. Mol. Struct. (Theochem)* **1988**, *181*, 33.
- Helgaker, T.; Jørgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; J. Wiley: New York, 2000.
- Møller, C.; Plesset, M. S. *Phys. Rev.* **1934**, *46*, 618–622.
- Bates, D. M.; Anderson, J. A.; Oloyede, P.; Tschumper, G. S. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2775–2779.
- Bachorz, R. A.; Bischoff, F. A.; Hoefener, S.; Klopper, W.; Ottiger, P.; Leist, R.; Frey, J. A.; Leutwyler, S. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2758–2766.
- Antony, J.; Grimme, S. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5287–5293.
- Cerny, J.; Jurecka, P.; Hobza, P.; Valdes, H. *J. Phys. Chem. A* **2007**, *111*, 1146–1154.
- Peverati, R.; Baldridge, K. K. *J. Chem. Theory Comput.* **2008**, *4*, 2030–2048.
- Baldridge, K. K.; Jonas, V. *J. Chem. Phys.* **2000**, *113*, 7511.
- Cammi, R.; Frediani, L.; Mennucci, B.; Tomasi, J.; Ruud, K.; Mikkelsen, K. V. *J. Chem. Phys.* **2002**, *117*, 13.
- Chen, W.; Gordon, M. S. *J. Chem. Phys.* **1996**, *105*, 11081–11090.
- Christiansen, O.; Mikkelsen, K. V. *J. Chem. Phys.* **1999**, *110*, 8348.
- Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Chem. Phys.* **2001**, *114*, 5691–5701.
- Cramer, C. J.; Truhlar, D. G. *Continuum Solvation Models: Classical and Quantum Mechanical Implementation*; VCH Publishers: New York, 1995; Vol. 6.
- Giesen, D. J.; Gu, M. Z.; Cramer, C. J.; Truhlar, D. G. *J. Org. Chem.* **1996**, *61*, 8720.
- Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999.
- Tao, J. M.; Perdew, J. P.; Staroverov, N.; Scuseria, G. E. *Phys. Rev. Lett.* **2003**, *91*, 146401.

- (44) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem. Phys.* **1981**, *55*, 117–129.
- (45) Schmidt, M.; Baldridge, K. K.; Boatz, J. A.; Elbert, S.; Gordon, M.; Jenson, J. H.; Koeski, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
- (46) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
- (47) Perdew, J. P. *Electronic Structure of Solids*; Akademie Verlag: Berlin, 1991.
- (48) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
- (49) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829–5835.
- (50) Dunning, T. H. *J. Chem. Phys.* **1989**, *90*, 1007.
- (51) Jurecka, P.; Sponer, J.; Černý, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8*, 1985–1993.
- (52) Baldridge, K.; Klamt, A. *J. Chem. Phys.* **1997**, *106*, 6622–6633.
- (53) Gregerson, L. N.; Baldridge, K. K. *Helv. Chim. Acta* **2003**, *86*, 4112–4132.
- (54) Baldridge, K. K.; Jonas, V. *J. Chem. Phys.* **2000**, *113*, 7511.
- (55) Klamt, A.; Jonas, V. *J. Chem. Phys.* **1996**, *105*, 9972.
- (56) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.
- (57) Klamt, A.; Jonas, V.; Bürger, T.; Lohrenz, C. W. *J. Phys. Chem.* **1998**, *102*, 5074–5085.
- (58) Zimmerli, U.; Parrinello, M.; Koumotsakos, P. *J. Chem. Phys.* **2004**, *120*, 2693.
- (59) Kubar, T.; Jurecka, P.; Černý, J.; Rezáć, J.; Otyepka, M.; Valdés, H.; Hobza, P. *J. Phys. Chem. A* **2007**, *111*, 5642–5647.
- (60) Ahlrichs, R.; Penco, R.; Scoles, G. *Chem. Phys.* **1977**, *19*, 119.
- (61) Hepburn, J.; Scoles, G. *Chem. Phys. Lett.* **1975**, *36*, 451.
- (62) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799–805.
- (63) Klamt, A.; Eckert, F.; Hornig, M. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 355–365.
- (64) Fortunelli, A.; Tomasi, J. *Chem. Phys. Lett.* **1994**, *231*, 34.
- (65) Luque, F. J.; Bachs, M.; Orozco, M. *J. Comput. Chem.* **1994**, *15*, 847.
- (66) Stefanovich, E. V.; Truong, T. N. *Chem. Phys. Lett.* **1995**, *244*, 65.
- (67) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 877–887.
- (68) Gregerson, L. N. Outlying Charge, Stability, Efficiency, and Extensibility in Quantum Mechanical Solvation Methods: the COSab-GAMESS Method, Ph.D. Thesis, University of California, San Diego, San Diego, 2004.
- (69) Chiorescu, I.; Duebel, D. V.; Arion, V. B.; Keppler, B. K. *J. Chem. Theory Comput.* **2008**, *4*, 499–506.
- (70) Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard III, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.
- (71) Mu, W.-H.; Chasse, G. A.; Fang, D.-C. *Int. J. Quantum Chem.* **2008**, *108*, 1422–1434.
- (72) Cramer, C. J.; Truhlar, D. G. *Acc. Chem. Res.* **2008**, *41*, 760–768.
- (73) Klamt, A.; Mennucci, B.; Tomasi, J.; Barone, V.; Curutchet, C.; Orozco, M.; Luque, F. J. *Theor. Chem. Acc.* **2009**, *42*, 489–492.
- (74) Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **2009**, *42*, 493–497.
- (75) Colominas, C.; Luque, F. J.; Orozco, M. *J. Phys. Chem. A* **1999**, *103*, 6200–6208.
- (76) Podolyan, Y.; Gorb, L.; Leszczynski, J. *J. Phys. Chem. A* **2002**, *106*, 12103–12109.
- (77) Bende, A.; Suhai, S. *Int. J. Quantum Chem.* **2005**, *103*, 841–853.
- (78) Kratochvíl, M.; Engkvist, O.; Sponer, J.; Jungwirth, P.; Hobza, P. *J. Phys. Chem. A* **1998**, *102*, 6921–6926.
- (79) Benassi, R.; Bertarini, C.; Hilfert, L.; Kempter, G.; Kleinpeter, E.; Spindler, J.; Taddei, F.; Thomas, S. *J. Mol. Struct. (Theochem)* **2000**, *520*, 273–294.
- (80) Benassi, R.; Bertarini, C.; Kleinpeter, E.; Taddei, F.; Thomas, S. *J. Mol. Struct. (Theochem)* **2000**, *498*, 201–215.
- (81) Kleinpeter, E.; Schroth, W.; Pihlaja, K. *Magn. Reson. Chem.* **2005**, *29*, 223–230.
- (82) Neuvonen, H.; Fülöp, F.; Neuvonen, K.; Koch, A.; Kleinpeter, E. *J. Phys. Org. Chem.* **2008**, *21*, 173–184.
- (83) Hinchliffe, A.; Mkadmh, A.; Nikolaidi, B.; Sosc'un, H. J.; Abu-Awwad, F. M. *Cent. Eur. J. Chem.* **2006**, *4*, 745–759.
- (84) Puterová, Z.; Krutošíková, A.; Lyčka, A.; Ďurčeková, T. *Molecules* **2004**, *9*, 241–255.
- (85) Dahlqvist, K.-I.; Forsen, S. *J. Phys. Chem.* **1965**, *69*, 4062–4071.
- (86) Baldridge, K. K.; Jonas, V.; Bain, A. *J. Chem. Phys.* **2000**, *113*, 7519.
- (87) Bain, A. D.; Hazendonk, P. *J. Phys. Chem. A* **1997**, *101*, 7182–7188.
- (88) Little, T. S.; Qiu, J.; Durig, J. R. *Spectrochim. Acta* **1989**, *45A*, 789.
- (89) Mönnig, F.; Dreizler, H.; Rudolph, H. D. *Z. Naturforsch., A: Phys Sci.* **1965**, *20*, 1323.
- (90) Abraham, R. J.; Sivers, T. M. *Tetrahedron* **1972**, *28*, 3015–3023.
- (91) Bertran, J. F.; Ortiz, E.; Ballester, L. *J. Mol. Struct.* **1973**, *17*, 161.
- (92) Marstokk, K.-M.; Mollendal, H. *J. Mol. Struct.* **1974**, *23*, 93.
- (93) Casarini, D.; Lunazzi, L.; Macciantelli, D. *J. Chem. Soc., Perkin Trans. 2* **1985**, 1839–1844.
- (94) Lunazzi, L.; Placucci, G.; Chatgilialoglu, C.; Macciantelli, D. *J. Chem. Soc., Perkin Trans. 2* **1984**, 819.
- (95) Pethrick, R. A.; Wyn-Jones, E. *J. Chem. Soc. A* **1969**, 713.