

Hydrogen-bond dependent conformational switching: a computational challenge from experimental thermochemistry

*James Luccarelli, Robert S. Paton**

Chemistry Research Laboratory, University of Oxford, 12 Mansfield Road, Oxford OX1 3TA, U.K.

Abstract. We have constructed an experimental dataset (SWITCH10) of equilibrium constants for a series of hydrogen-bond dependent conformational switches. These organic molecules possess common functionalities and are representative in terms of size and composition of systems routinely studied computationally. They exist as two well-defined conformations which serve as a useful tool to benchmark computational estimates of experimental Gibbs energy differences. We examine the performance of a variety of density functionals (B3LYP, B3LYP-D3, CAM-B3LYP, ω B97X-D, M06-2X) against these experimental benchmarks. Surprisingly, despite a strong similarity between the two switch conformations, the average errors (0.4 – 1.7 kcal·mol⁻¹) obtained across the dataset for all methods are larger than obtained with HF calculations. B3LYP was found to outperform implicitly and explicitly-dispersion corrected functionals, with an average error smaller by 1 kcal·mol⁻¹. Unsystematic errors in the optimized structures were found to contribute to the relatively poor performance obtained, while quasi-rigid

rotor harmonic oscillator thermal contributions are important in improving the accuracy of computed Gibbs energy differences. These results emphasize the challenge of quantitative computational thermochemistry and caution against the often used (but unstated) assumption of favorable error cancellation in comparing conformers or stereoisomers. While *ab initio* benchmark data are restricted to smaller molecules, the SWITCH10 dataset provides experimental benchmarks for realistic organic systems which are representative of those studied routinely using density functional theory.

Introduction

Precise control over hydrogen bonding interactions is essential in designing synthetic systems including supramolecular complexes¹ and organocatalysts.² We recently reported the synthesis and characterization of a series of benzamido-diphenylacetylene (DPA) molecular switches in which the relative hydrogen bonding capability of two amide groups is the critical determinant of conformation.^{3–5} In this system, two distinct hydrogen-bonded conformations can readily interconvert by a 180° rotation about the acetylene axis. Different substituents affect the electronics of the amides and thus change the relative strengths of the two intramolecular hydrogen bonds. By varying the attached functionalities (at positions R₁ and R₂) the conformational preference of the system is changed (Figure 1).

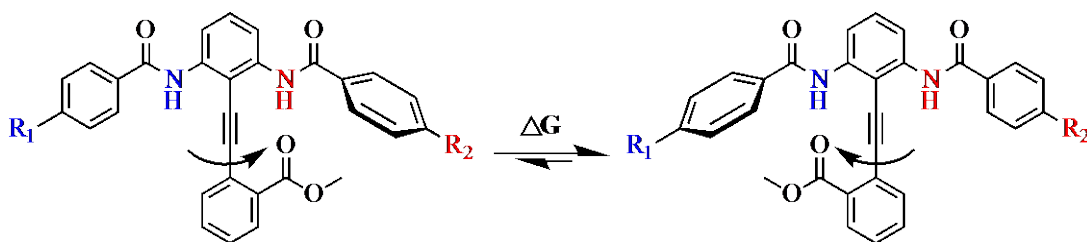
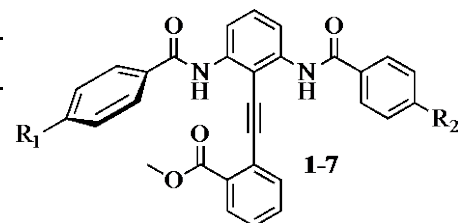


Figure 1. Two rotamers of the DPA-based molecular switch defined by a 180° rotation about the DPA axis. Variation of substituents R_1 and R_2 controls the equilibrium population of each conformation. Due to steric clashing with the methyl ester, the proximal benzamide ring rotates out of the plane.

Ten molecules of this class have been synthesized, incorporating electron withdrawing groups (*para*-nitro **5**, *para*-chloro **3**, 4-pyridyl **8**), electron donating groups (*para*-methoxy **7**, *para*-dimethylamino **6**), mixed cases (*para*-methoxy on one benzamide and *para*-chloro **2** or *para*-nitro **4** on the other), as well as 2-pyridyl substitution **9** and molecule **10** which substitutes a urea in place of one of the benzamides (Table 1).^{3–6} The equilibrium conformational ratio is measured by comparing the NMR shifts of the amide NHs in chloroform with those of control molecules which are incapable of forming hydrogen bonds. Using this assay, the equilibrium constants of ten molecules have been experimentally determined at room temperature, and were found to encompass relative hydrogen bond strengths spanning a range of approximately 2 kcal·mol⁻¹, with most of the molecules clustered in a 1 kcal·mol⁻¹ range.

Table 1. Experimental Gibbs energy difference between DPA molecular switch conformations. Molecules marked with a star are not phenyl substituents, but replacements of the ring with the marked group (e.g. molecule **8** juxtaposes a *para*-pyridine ring and a phenyl ring). Molecule **10** is of a different class, and features acetamide and phenylurea rather than the benzamides of the other molecules.

Molecule	R^1	R^2	ΔG (kcal·mol ⁻¹)
			

1	H	H	0.00
2	Cl	MeO	-0.21
3	Cl	H	-0.17
4	NO ₂	MeO	-0.44
5	NO ₂	H	-0.37
6	NMe ₂	H	0.17
7	MeO	H	0.03
8	4-pyridyl*	H	-0.38
9	2-pyridyl*	H	0.63
10	NHPh*	Me*	-1.37

Much recent computational work has highlighted the importance of accurately accounting for non-bonding interactions in the development of electronic structure theory.⁷⁻¹¹ Large benchmark datasets at the gold-standard¹² CCSD(T)/CBS level of theory, including the S22 and S66 sets of Hobza and co-workers for small complexes bound by hydrogen-bonding and dispersion interactions,^{13,14} the composite GMTKN30 set of Goerigk and Grimme,¹⁵ and a large composite database constructed by Friesner and co-workers,¹⁶ provide clear standards against which computational methods can be evaluated. More recently, datasets of much larger non-covalently bound complexes, such as the S12L,¹⁷ S30L¹⁸ and L7¹⁹ sets have increased the molecular size of the available benchmarks. While high-level methods such as Coupled Cluster provide accurate benchmark values,¹¹ they impose computational demands incompatible with the study of many systems relevant to organic chemistry. Instead, more affordable density functional theory (DFT) methods are used to model compounds of interest. The comparison of DFT performance against experimental measurements is critical to understand deficiencies and inform future usage.

Using these and other benchmarking sets, improved descriptions of non-bonding interactions have been implemented into semi-empirical DFT methods, such as hybrid meta-GGA functionals (e.g. the Minnesota family),²⁰ explicit dispersion corrections for London dispersion such as D3,²¹ and *van der Waals* functionals.^{22,23} Much specific work has been directed towards understanding hydrogen bonding, with recent reports investigating DFT's ability to reproduce the distances²⁴ and angles²⁵ of small H-bonded complexes. While molecular mechanics force fields systematically underestimate the stability of such complexes,²⁶ accurate potential energy curves can be calculated with appropriately-chosen DFT methods.^{27,28}

These molecular switches provide a difficult test for computation. They are too large (our molecules range from 35 – 41 heavy atoms) to be easily optimized with correlated *ab initio* methods, and so DFT is more appropriate. Furthermore, the conformation is dictated by the relative magnitude of hydrogen bonding interactions along with torsional rotational potentials and other non-bonding contacts. The ability of computational methods to accurately account for the observed behavior spanning only a few kcal·mol⁻¹ is of importance, since functional group effects of similar magnitude are fundamental to organic chemistry. Computational predictions of thermodynamic or kinetic selectivity in organic reactions rely on the ability to calculate energy differences in this range.²⁹ Given the precision of the NMR assay in measuring the equilibrium constants for the DPA switches, this system serves as a novel experimental dataset measured at consistent temperature (298 K) and solvent (CDCl₃). Moreover, although high level *ab initio* calculations provide useful “gold standards”, it is essential to compare against quantitative experimental data whenever possible.

Methods

All Hartree-Fock (HF) and DFT calculations were performed with *Gaussian09*, rev B.01.³⁰ Optimizations were performed using HF, the B3LYP hybrid generalized gradient approximation (GGA) functional,^{31,32} the long-range corrected CAM-B3LYP which increases the amount of HF exchange at long-range,³³ the M06-2X²⁰ hybrid meta-GGA functional, and the long range dispersion-corrected ω B97X-D³⁴ functional in combination with a Pople 6-31+G(d,p) basis set for all elements.^{35–37} Standard convergence criteria were used along with a fine grid for all numerical integrations. Stationary points on the potential energy surface were confirmed as minima by the presence of no imaginary harmonic vibrational frequencies. Free energies include unscaled zero point vibrational energies. Where indicated, low frequencies ($< 100 \text{ cm}^{-1}$) are corrected in the vibrational component of the entropy using a free rotor approximation according to the method of Grimme *et al*, since entropy associated with these loose vibrational modes is the most prone to computational error.^{17,38–41} Optimizations were performed in the gas phase or with the implicit conductor-like polarizable continuum model (CPCM) of the NMR solvent, chloroform.⁴² Single point calculations were performed at stationary points using Dunning's quadruple- ζ cc-pVQZ basis set.⁴³ To verify that the cc-pVQZ is sufficiently large to assure basis set convergence, extrapolation using the two-point formula of Halkier *et. al*.⁴⁴ was conducted:

$$E_{CBS} \approx \frac{E_X X^3 - E_Y Y^3}{X^3 - Y^3}$$

where X and Y are the cardinal numbers of the Dunning basis sets used for the extrapolation. Based on extrapolations from cc-pVTZ and cc-pVQZ optimizations in solvent using B3LYP and ω B97X-D, basis sets beyond the quadruple- ζ level would change the calculated energy differences of the switches by $< 0.02 \text{ kcal}\cdot\text{mol}^{-1}$ (Tables S1-S2). Where indicated, dispersion correction to energies were applied using the DFT-D3 correction with zero-damping and the three-body term included.²¹ Unless otherwise stated, all energies are quoted in $\text{kcal}\cdot\text{mol}^{-1}$ and

Gibbs free energies (incorporating ZPE, thermal, and entropic contributions obtained with the 6-31+G(d,p) basis set) are at 298.15 K and a standard state of 1 mol·L⁻¹.

Statistical Analysis

The agreement between predicted and experimental free energy differences of the rotamers was assessed using mean unsigned errors (MUEs), root-mean-square deviations (RMSDs), linear coefficient of determination (r^2), and Pearlman and Charifson's predictive index (PI).^{45,46} The latter quantifies the ability of the calculations to appropriately order the molecules in energy difference, with a value of +1 indicating perfect ranking and -1 indicating perfectly anticorrelated predictions. Due to the limited energy range of the systems (which is nonetheless representative of those encountered in organic chemistry), a calculation that predicted every switch to have no bias ($\Delta G = 0$) would have an RMSD of 0.56 kcal·mol⁻¹ and an MUE of 0.4 kcal·mol⁻¹. Despite these low errors, such a calculation would be unable to rank the compounds (PI undefined) and would show no linear correlation with the data.

Results

Gas Phase Optimizations. In the first instance, structures were optimized in the gas phase and energies determined by a single point calculation in CPCM chloroform. Comparison with the experimental data indicates relatively poor performance for several of the DFT methods (Table 2; Table S3). In particular, while B3LYP and the related CAM-B3LYP showed the lowest errors and greatest correlation with experiment, the more recently developed ω B97X-D and M06-2X functionals fared substantially worse, with errors more than double those of B3LYP. The D3 dispersion correction of Grimme added to the B3LYP geometries had the overall effect of

increasing the calculated energy differences, resulting in higher errors but better PI. HF had errors nearly as low as calculated by B3LYP, but the low coefficient of determination and PI reveal that HF cannot reliably distinguish between the molecules; the low errors are a serendipitous result of the small data range and calculations predicting all molecules to be roughly equal (Figure 2).

Table 2. Comparison of gas phase optimizations vs. experimental ΔG for switches **1-10**. Structures were optimized at various levels of theory with the 6-31+G(d,p) basis set. Single point electronic energies were calculated in CPCM chloroform at the same level of theory with the cc-pVQZ basis set.

Functional	MUE	RMSD	r^2	PI
HF	0.39	0.54	0.28	0.32
B3LYP	0.33	0.40	0.82	0.84
B3LYP-D3	1.13	1.32	0.83	0.95
CAM-B3LYP	0.37	0.49	0.97	0.94
ω B97X-D	1.20	1.44	0.61	0.80
M06-2X	0.76	0.91	0.79	0.91

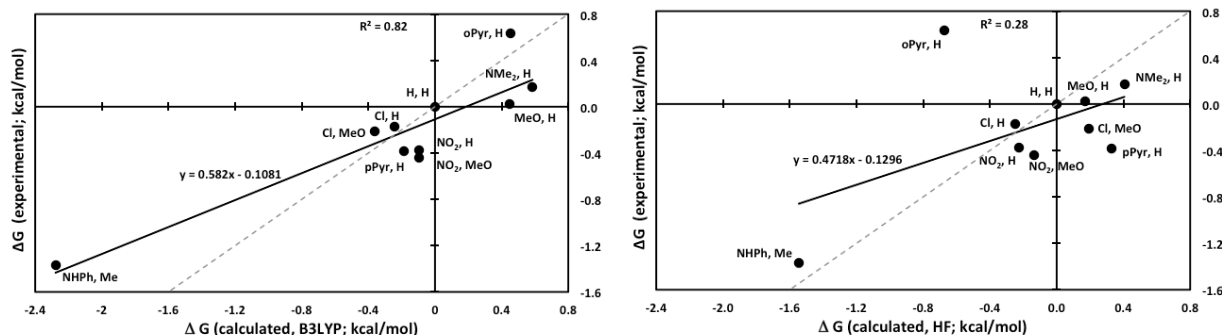


Figure 2: Experimental vs. calculated free energy differences ($\text{kcal}\cdot\text{mol}^{-1}$) differences using B3LYP (left) and HF (right). Structures were optimized using the two theories with the 6-31+G(d,p) basis set; single point electronic energies were calculated in CPCM chloroform at the same level of theory with the cc-pVQZ basis set. The grey dashed line is an ideal $y = x$ plot; the black line, whose equation is listed, is a linear fit to the data. These plots show that while the errors calculated using B3LYP and HF are low, there

is relatively poor differentiation between the molecules, resulting in poor predictive power, especially for HF.

Effect of Solvent Model. In order to investigate the effects of different implicit solvent models on the energies, single point calculations were repeated with the polarizable continuum model (PCM)⁴⁷ and the universal solvation model (SMD)⁴⁸ of chloroform at the B3LYP/cc-pVQZ level. While SMD solvation was consistently worse than CPCM solvation, PCM was very similar, with slightly improved PI and r^2 but slightly worse errors (Table 3; Table S4). Based on these results, the CPCM solvent model was used for all future calculations.

Table 3. Effects of implicit solvation model upon B3LYP/cc-pVQZ//B3LYP/6-31+G(d) single point calculations in chloroform.

Solvent Model	MUE	RMSD	r^2	PI
PCM	0.35	0.48	0.86	0.88
SMD	0.50	0.57	0.53	0.50
CPCM	0.33	0.40	0.82	0.84

Solution Phase Optimizations. To further investigate the role of solvation, the optimizations were repeated in CPCM chloroform, with the electronic energy determined by a single point calculation, as before. This leads to an overall improvement in the linear correlations between theory and experiment, but in general does not reduce the actual errors (Table 4; Table S5). This can be explained by many of the methods, in particular ω B97X-D, systematically overestimating the energy differences of the switches. This leads to an excellent rank-ordering of the compounds (a PI of 0.96), but relatively large MUE and RMSD as the linear correlation has a low slope of 0.23 relative to the ideal 1.0. The performance of HF is most improved, with the lowest errors of any of the methods and a PI increasing from 0.32 for the gas phase calculations to 0.87 with the

addition of solvent. This large improvement in predictive power results from only modest changes in energy for each molecule—on average $0.5 \text{ kcal}\cdot\text{mol}^{-1}$ —but given the small energy differences between the switches this is sufficient to drastically affect the results.

Table 4. Comparison of solution phase CPCM optimizations vs. experimental ΔG for switches **1-10**. These solution phase calculations improve the linear correlation of most methods, but at a cost of higher absolute errors for some.

Functional	MUE	RMSD	r^2	PI
HF	0.25	0.36	0.85	0.87
B3LYP	0.42	0.56	0.88	0.98
B3LYP-D3	1.35	1.57	0.77	0.98
CAM-B3LYP	0.64	0.91	0.75	0.89
ω B97X-D	1.52	1.75	0.74	0.96
M06-2X	0.77	1.04	0.81	0.88

Considering the electronic energy from the cc-pVQZ single point calculations alone ignores the zero point energy corrections and entropy terms from free energy calculations. Comparing these energy differences to the measured free energy differences of the switches results in large errors, but excellent predictive power and correlation (Table S6). This is largely because the consideration of thermal effects serves to reduce the energy differences between the switches, meaning the energies alone are along an erroneously large range.

Thermal Corrections to the Gibbs energy. The energy calculations clearly indicate the importance of accurately considering entropic effects in calculating the free energy differences of the switches. Perhaps the most commonly used approach in computations of molecular

thermochemistry is the Rigid Rotor-Harmonic Oscillator (RRHO) approximation. From the harmonic frequencies computed at the equilibrium geometry, for a frequency ω , the vibrational entropy is:

$$S_{vib} = R \left[\frac{h\omega}{k(e^{h\omega/kT} - 1)} - \ln(1 - e^{-h\omega/kT}) \right] \quad (1)$$

where k is Boltzmann's constant, h is Planck's constant, T is the temperature, and R is the universal gas constant. As the frequency tends towards zero, the logarithmic term in this equation diverges, meaning that numerical noise in calculating low vibrational modes can greatly influence the computed entropy values.⁴⁹ Furthermore, low frequency free or hindered internal rotations are inherently poorly described by a harmonic approximation. The adoption of quasi-RRHO approaches, in which entropic terms associated with low frequencies are computed outside of the RRHO paradigm, have recently been suggested for improved estimates of thermochemical quantities by Truhlar,⁴⁸ Grimme,¹⁶ and Head-Gordon.⁴⁹ Quasi-RRHO corrections are typically a few percent of the total thermal correction, and so are larger in absolute terms (i.e. several kcal·mol⁻¹) for bigger molecules.⁵⁰

Using the quasi-RRHO model of Grimme *et al.*,¹⁷ low frequencies are treated as free rotational modes, for which the rotational entropy is computed instead of a harmonic vibrational entropy. The moment of inertia μ of a free-rotor is given by:

$$\mu = \frac{h}{8\pi^2\omega} \quad (2)$$

In the limit of small ω , the moment of inertia becomes large, so it is used in a reduced form:

$$\mu' = \frac{\mu B_{av}}{\mu + B_{av}} \quad (3)$$

where B_{av} is the average molecular moment of inertia, taken to be 10^{-44} kg·m² (changing this value by two orders of magnitude did not affect the resulting entropy values). For each such mode this leads to a rotational entropy of:

$$S_{Rot} = R \left[1/2 + \ln \left\{ \left(\frac{8\pi^3 \mu' kT}{h^2} \right)^{1/2} \right\} \right] \quad (4)$$

A damping function is used to interpolate smoothly between harmonic entropies for large frequencies and free rotor entropies for frequencies smaller than a chosen cut-off value. The choice of the cut-off frequency is somewhat arbitrary; the results described below are with a cut-off of 100 cm⁻¹, but using cut-off energies of 50 cm⁻¹ or 200 cm⁻¹ only changes the results by an average of < 0.05 kcal·mol⁻¹ per switch (Table 5; Tables S7-8). There is a small reduction in average error of less than 0.1 kcal·mol⁻¹ for B3LYP ΔG values vs. experiment as the cut-off increases from 50 to 100 cm⁻¹, while a further increase has negligible (< 0.02 kcal·mol⁻¹) effect. We found uniformly better performance using this quasi-RRHO treatment than from simply transposing low frequencies to higher wavenumbers as also used in quasi-harmonic corrections in the literature.⁴⁹ This method is seen to be universally worse when compared to Grimme as it depends largely upon the number of low frequency modes for each conformer, and deteriorates for larger cut-off values.

Table 5. Comparison of the effects of vibrational entropy treatment on B3LYP results for all switches, and choice of cut-off frequency in quasi-RRHO/quasi-harmonic approach.

Cut-off freq.	<i>Quasi-RRHO (Grimme)</i>			<i>Quasi-Harmonic (Truhlar)</i>		
	MUE	RMSD	R ²	MUE	RMSD	R ²
0	0.33	0.40	0.82			
50	0.48	0.63	0.99	0.59	0.77	0.98

100	0.42	0.55	0.98		0.67	0.84	0.98
200	0.40	0.54	0.97		0.70	0.87	0.98

Grimme’s quasi-RRHO treatment was applied to the gas and solvent phase optimizations using an in-house Python script,⁵¹ resulting in a universal improvement in coefficient of determination and PI (Table 6). HF benefited the most from this entropic correction, with the gas phase r^2 improving from 0.28 to 0.82 with a corresponding increase in PI from 0.32 to 0.86. The largest correction was to the gas phase HF-calculated 2-pyridyl switch **9**, whose ΔG changed by 0.9 kcal·mol⁻¹. While both the “on” and “off” forms of that molecule were calculated to have 12 low vibrational frequencies, the lowest mode of the “on” form was 6.0 cm⁻¹, while for the “off” form it was 12.3 cm⁻¹. Without this correction, the RRHO approach incorrectly calculates the “on” form to be more stable than experiment (due to an overly favourable TΔS term), an error which is rectified by the quasi-RRHO correction.

For most other functionals and switches, the change in ΔG is quite small, generally < 0.2 kcal·mol⁻¹, and very rarely > 0.5 kcal·mol⁻¹. This is unsurprising as the vibrational differences between the “on” and “off” states of the system should be quite small, and thus errors generally cancel even without explicit consideration of the low modes. While the magnitude of the RRHO corrections for solution-phase optimizations is similar to the gas phase (ca. 7 – 9 kcal·mol⁻¹ in each case), the corrections generally cancel for the solvent-optimized molecules, resulting in smaller net changes to ΔG .

Table 6. Statistical parameters calculated using Gibbs energies corrected with a quasi-RRHO model for vibrational modes < 100 cm⁻¹. This modified treatment of entropy improves the coefficient of determination and PI for all of the functionals, but increases RMSD and MUE for some.

	gas phase optimization				solution phase optimization			
Functional	MUE	RMSD	r ²	PI	MUE	RMSD	r ²	PI
HF	0.24	0.31	0.82	0.86	0.24	0.33	0.95	0.97
B3LYP	0.42	0.55	0.98	0.97	0.43	0.49	0.90	0.98
B3LYP-D3	1.44	1.72	0.81	0.99	1.37	1.59	0.73	0.98
CAM-B3LYP	0.62	0.77	0.98	1.00	0.68	0.95	0.92	0.96
ω B97X-D	1.37	1.69	0.69	0.95	1.73	2.01	0.79	0.98
M06-2X	0.72	0.91	0.91	0.91	0.75	1.04	0.89	0.91

Variation in Optimized Geometries. While solution-phase optimizations coupled with a quasi-RRHO thermal correction result in excellent rank ordering of the switches at all levels of theory tested, errors exceed the true energy differences of the system. As a result, the optimized geometries from the calculations were investigated more closely. Crystal structures have been reported for the *para*-nitro **5** and *para*-dimethylamino **6** switches.^{3,5} Both show an approximately 10° dihedral angle for the central diphenyl moiety, with an approximately 50° out of plane twist of the phenyl ring adjacent to the H-bond caused by steric clashing with the methyl ester H-bond acceptor.

Histograms of the DPA dihedral angles were plotted from the optimized geometries for each method: significant disagreement within and among the functionals was found as to the preferred conformation (Figure 3). It is surprising that for a given level of theory, qualitatively different conformations are predicted for closely related switches. For example, in the gas phase optimizations, each of the methods results in a multimodal distribution of geometries, with the plurality of optimizations resulting in angles of less than 30° but a significant minority at angles of 50° or greater. Optimizations in solvent gave less right-skewed histograms (i.e. the

diphenylacetylene is closer to planarity), but all methods still calculated some rotamers as roughly planar and others as significantly twisted out of plane. The most consistent results were with HF solvent optimizations, which calculated 40° - 50° dihedral angles for all but two molecules. While this is inconsistent with the solid-state data and the other methods, these calculations were also the most accurate at reproducing the experimental free energies. This observation suggests that some of the discrepancy between theory and experiment is accountable to unsystematic variations in geometry between the two conformations of each switch.

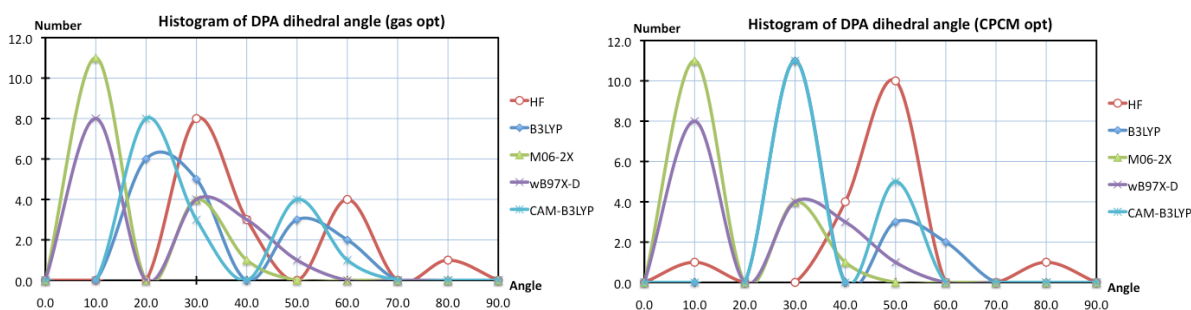


Figure 3. Histograms of the calculated DPA dihedral angle for each functional for optimizations performed in gas phase (top) or in CPCM chloroform (bottom). The addition of solvent reduces the skewing towards higher dihedral angles, but does not eliminate the multimodal profiles for most functionals. For the CPCM optimization, the highest peaks for B3LYP and CAM-B3LYP overlay at 30° .

Due to the relative flexibility of the system, several possible DPA dihedral angles can accommodate an $\text{NH}\cdots\text{O}$ hydrogen bond by twisting the amide and ester groups relative to their phenyl rings. Overlays with the crystal structures reveal this effect. For instance, the solvent optimization of **6** with M06-2X results in a DPA dihedral of 24.5° compared to 11.5° in the crystal structure (Figure 4a). Due to this 15° shift, the benzamide rings of the calculated structure are both twisted approximately 30° relative to the DPA. This results in a $\text{CO}\cdots\text{N}$ distance of 2.98

Å and an NHO angle of 153°. In the crystal structure, however, the phenyl ring adjacent to the amide is forced 50° out of plane by steric repulsion from the methyl ester, resulting in an H-bond distance of 3.10 Å and an angle of 145°. Likewise for the *para*-nitro substituted **5**, ωB97X-D calculates a DPA dihedral of 0.3°. Instead of this resulting in one phenyl ring rotated significantly out of plane, however, the calculated structure shows equal 30° rotations of both benzamide rings and a 3.07 Å hydrogen bond. The crystal structure shows a hydrogen bond that is very similar in length (3.09 Å), but this is due to amide and ester carbonyls oriented in very different directions from those in the calculations (Figure 4b).

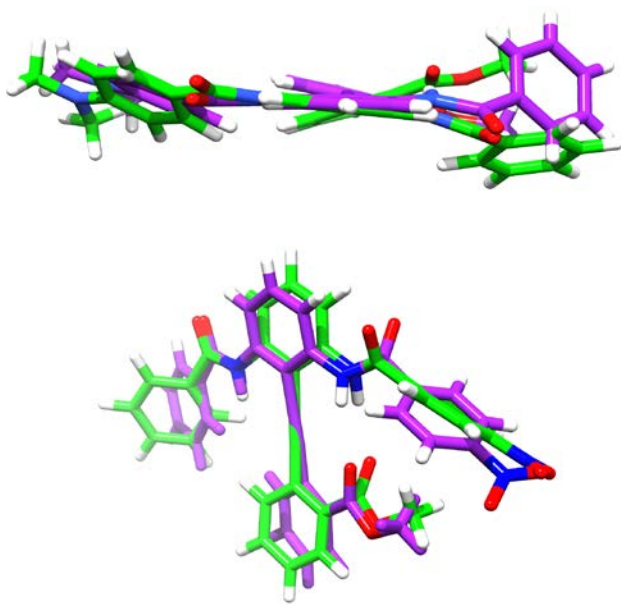


Figure 4. (a) Overlay of crystal structure of *para*-dimethylamino switch **6** (purple) with M06-2X (green). The crystal structure has a more planar DPA, forcing the phenyl ring out of plane. The greater DPA twist in the calculation overcomes this strain, resulting in a different benzamide conformation; (b) Overlay of crystal structure of *para*-nitro switch **5** (purple) with ωB97X-D (green). In the crystal structure, the methyl ester H-bond acceptor is in the plane of the DPA, while the amide carbonyl rotates out of

conjugation to position the N-H. The optimized structure predicts both the amide and ester to rotate out of conjugation with the DPA, resulting in a nearly identical H-bond length but a different conformation.

Constrained Geometry Optimizations. The inconsistent geometries obtained from unconstrained calculations led to evaluation of the energies calculated with the system constrained to defined conformations. Optimizations were conducted using HF and B3LYP in which the central diphenyl was constrained to planarity, 30°, or 50°. These solution-phase optimized structures are not true minima, and so only their energies are compared to the experimental free energies of the switches (Tables S9-11). For both functionals, the 50° dihedral angle results in the lowest errors calculated so far, within 0.2 kcal·mol⁻¹ for HF and 0.4 kcal·mol⁻¹ for B3LYP. The coefficient of determination and PI for both methods are similarly excellent, with both methods perfectly ordering the compounds at 0° and B3LYP doing so at 50° as well.

While the energy differences calculated at 0° and 50° dihedral angles were very close, when constrained to 30° the B3LYP calculations showed much greater spread in energies, with correspondingly larger errors. For the *para*-nitro molecule **5** for instance, comparison between the constrained optimizations, the unconstrained solution phase optimization, and experiment reveal discrepancies in angle and hydrogen bond length (Table S12). All optimizations overestimate the H-bond distance, by up to 0.2 Å for the 50° optimization. The conformational difference is likewise overestimated, but the consideration of entropic effects may reduce this somewhat (e.g. in the unconstrained case, the calculated Gibbs energy difference is just -0.22 kcal·mol⁻¹).

Discussion

The molecular switches of Hamilton *et al.* are an example of an organic system for which accurate quantum chemical calculations would be extremely useful. While synthesis of the molecules utilizes proven chemistry, each requires ca. eight steps to create, and two additional control molecules of similar complexity are needed to determine the free energy difference of the system. Calculations which could accurately predict the energies of novel switches could save considerable synthetic effort and allow efforts to be directed towards the most promising molecules. Once validated on this system, these computational methods could also be applied with some confidence to other hydrogen-bonded molecules.

While it is universally acknowledged that DFT estimates of absolute barrier heights and reaction energies may suffer from errors of several kcal·mol⁻¹, it is also often assumed that energetic comparisons between closely-related structures (such as stereoisomers) are more accurate due to favorable error cancellation. Such an assumption is necessary for DFT calculations of stereoselectivity, for example, since these are arguably predicated on achieving an accuracy greater than 1 kcal·mol⁻¹. In contrast, our results indicate that an accurate description of experimental Gibbs energy differences between two closely related conformations poses a significant challenge for standard “black box” DFT calculations with a variety of functionals. A correct ranking of the switches’ conformational preferences was observed for all functionals considered here (B3LYP, CAM-B3LYP, ω B97X-D and M06-2X), giving a predictive index (PI) above 0.9 in each case. However, the average computed errors in ΔG range from 0.4 – 1.7 kcal·mol⁻¹, and are outperformed by HF calculations. The incorporation of an explicit dispersion correction (B3LYP-D3 and ω B97X-D) gave the worse performance by some margin. The molecules considered are by no means “exotic”, and so these results suggest that extreme caution should be exercised in the calculation of Gibbs energy differences between similar structures in

the absence of experimental or other benchmark data. However, based on our findings, computing experimental *trends*, in contrast to *one-off* predictions, is a tangible goal for DFT calculations.

Thorough investigation of the calculations allowed some of the sources of error to be localized. One is spurious entropy components associated with anharmonic low vibrational modes obtained from routinely employed RRHO thermal corrections. In systems that are separated by small free energy differences, these entropic terms can lead to sizable relative errors. A quasi-RRHO correction to entropy greatly reduces these errors and adds essentially no computational cost to the calculations, so these results suggest that this correction should become standard. The largest entropic corrections are applied to the very smallest frequencies, so the cut-off frequency at which the RRHO treatment is applied does not appear critical. Additionally, results were improved when optimizations were performed in an implicit solvent model. This adds some cost to the calculations, but in general it is small and is more than compensated for by the significant improvement in linear correlation and PI.

For the DPA-based switches, the most accurate results were ultimately achieved using the relatively simple functional B3LYP in combination with solution-phase optimizations, RRHO treatment of entropies, and large electronic single point calculations. Using these methods, RMSDs of $< 0.50 \text{ kcal}\cdot\text{mol}^{-1}$ and a PI of 0.98 could be achieved, which are sufficient for guiding the development of related molecules. The performance of the more recent functionals ω B97X-D and M06-2X was comparatively worse than that of B3LYP. These methods are superior in their description of non-bonded interactions due to medium and long-range correlation, but nonetheless do not perform very well for this system. It is apt to mention the comparative

success of B3LYP/6-31G(d) calculations in many studies of organic stereoselectivity, despite its well-publicised failings.⁵²

These calculations, while still not perfect in modelling the switches, provide possible insight into the behaviour of these molecules. Principally, the constrained geometry optimizations indicate that the system is quite flexible. In contrast to the crystal structures which show relatively planar structures, energies of the molecules were found to be roughly equivalent both when planar and when the DPA is twisted up to 50°. This is possible because in the calculations, the H-bonding amide and accepting carbonyl can themselves twist to accommodate an optimal H-bond length at a range of DPA dihedrals. This could represent a failure of DFT, since these results are inconsistent with the X-ray data. Alternatively, the crystal structures may not reflect the true flatness of the potential energy surfaces. Instead, forces such as crystal packing effects or intermolecular hydrogen bonding may predominate in the solid state.⁵³

The constrained geometry calculations raise further questions regarding the conformational equilibria. In the crystal structures, a pronounced twist in the phenyl ring attached to the donating amide is present due to steric clashing with the methyl ester acceptor. In the calculations, this is not present because the DPA twists to avoid this unfavourable interaction. Individual molecules twist to different extents, resulting in the multimodal histogram of DPA dihedral seen in the unconstrained cases. Restricting all the molecules to a single angle generally results in poorer correlation with experiment, unless the constrained angle was 50°. It is unclear whether this is truly the optimal equilibrium conformation, or if this angle allows for fortuitous error cancellation, but it does provide the best correlation with experiment of any of the methods tested. These constrained calculations, however, do not allow for vibrations to be accurately calculated, and so this analysis does not include the entropic considerations that were found to be

critical in the unconstrained cases. Nonetheless, they highlight how changes in the optimized geometry can greatly affect calculated energies. DFT-derived geometries are used for many composite methods such as CBS-QB3⁵⁴ and G3-B3⁵⁵ that use very large post-HF electronic energy calculations. While these methods have proven their accuracy for small systems, for larger molecules the DFT geometry may limit accuracy more than the electronic calculations themselves.

Conclusions

Quantum chemical calculations were applied to a series of DPA molecular switches in which intramolecular hydrogen bonds template conformation. Poor correlation with experiment was found with standard gas phase optimizations with a variety of DFT methods, but a new protocol incorporating solution-phase optimizations, large basis sets, and modified RRHO treatment of entropy gave much more accurate results. Using these methods, the switches could be ranked successfully in relative order of hydrogen bond strength, although errors remained between calculation and experiment for the absolute free energy differences of the molecules. Based on constrained optimizations, the DPA dihedral was not found to be essential to the energies of the system, in contrast to the clear preference seen in crystal structures. Together, these results benchmark the accuracy of DFT methods for medium-sized organic systems, and highlight the importance of the accurate treatment of solvation and entropy in quantum chemical calculations. The modified DFT procedure present here may be broadly applicable, and could result in greater accuracy in modelling systems for which non-covalent interactions are critical.

Additional benchmarking of the solvation and entropic corrections using large CCSD(T)/CBS databases could help confirm the utility of these methods. New functionals could then be

developed which may improve on the poor performance of existing DFT methods. Future work should apply this modified DFT procedure to other systems of interest. For instance, the conformational preference of novel H-bonded peptidomimetics could be assessed *in silico* with greater accuracy than is possible with semi-empirical methods. This could allow for a more precise understanding of the conformational preferences of these molecules, which may aid future design efforts. Additionally, improved DFT modelling could aid the understanding of organocatalysis and allow greater confidence in the calculated structures of complexes. In any case, this study provides further evidence that “black box” modelling of structures cannot be considered truly reliable, especially in instances where the energy differences involved are slight.

ASSOCIATED CONTENT

Supporting Information.

The following files are available free of charge: Coordinates and energies of all optimized structures (PDF); Energies for single point calculations (PDF)

AUTHOR INFORMATION

Corresponding Author

*andrew.hamilton@nyu.edu *robert.paton@chem.ox.ac.uk

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

ACKNOWLEDGMENT

This work utilized the computational resources of the NIH HPC Biowulf cluster. (<http://hpc.nih.gov>). We thank Dr. Andrew Hamilton of NYU for helpful discussions with regard to the solid state conformations of the molecules. We thank Dr. Kelvin Jackson of the University of Oxford for advice with scripts for quasi-Harmonic and quasi-RRHO calculations.

ABBREVIATIONS

CBS, complete basis set; CCSD(T), coupled cluster with full treatment of singles and doubles and perturbative treatment of triples; CPCM, conductor-like polarizable continuum model; DFT, density functional theory; DPA, diphenylacetylene; HF, Hartree-Fock; MUE, mean unsigned error; NMR, nuclear magnetic resonance; PI, predictive index; RRHO, rigid rotor-harmonic oscillator; RMSD, root-mean-square deviation; ZPE, zero point energy.

REFERENCES

- (1) Fan, E.; Van Arman, S. A.; Kincaid, S.; Hamilton, A. D. *J. Am. Chem. Soc.* **1993**, *115* (1), 369–370.
- (2) Taylor, M. S.; Jacobsen, E. N. *Angew. Chem. Int. Ed.* **2006**, *45* (10), 1520–1543.
- (3) Jones, I. M.; Hamilton, A. D. *Org. Lett.* **2010**, *12* (16), 3651–3653.
- (4) Jones, I. M.; Hamilton, A. D. *Angew. Chem. Int. Ed.* **2011**, *50* (20), 4597–4600.
- (5) Jones, I. M.; Lingard, H.; Hamilton, A. D. *Angew. Chem. Int. Ed.* **2011**, *50* (52), 12569–12571.
- (6) Luccarelli, J.; M. Jones, I.; Thompson, S.; D. Hamilton, A. *Org. Biomol. Chem.* **2017**, *15* (43), 9156–9163.
- (7) Riley, K. E.; Pitoňák, M.; Jurečka, P.; Hobza, P. *Chem. Rev.* **2010**, *110* (9), 5023–5063.
- (8) Burns, L. A.; Mayagoitia, Á. V.-; Sumpter, B. G.; Sherrill, C. D. *J. Chem. Phys.* **2011**, *134* (8), 084107–084132.
- (9) Klimeš, J.; Michaelides, A. *J. Chem. Phys.* **2012**, *137* (12), 120901–120901–120912.
- (10) Antony, J.; Sure, R.; Grimme, S. *Chem Commun* **2015**, *51* (10), 1764–1774.
- (11) Řezáč, J.; Hobza, P. *Chem. Rev.* **2016**, *116* (9), 5038–5071.
- (12) Řezáč, J.; Hobza, P. *J. Chem. Theory Comput.* **2013**, *9* (5), 2151–2155.
- (13) Jurečka, P.; Sponer, J.; Cerný, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2006**, *8* (17), 1985–1993.
- (14) Řezáč, J.; Riley, K. E.; Hobza, P. *J. Chem. Theory Comput.* **2011**, *7* (8), 2427–2438.
- (15) Goerigk, L.; Grimme, S. *J. Chem. Theory Comput.* **2011**, *7* (2), 291–309.
- (16) Schneebeli, S. T.; Bochevarov, A. D.; Friesner, R. A. *J. Chem. Theory Comput.* **2011**, *7* (3), 658–668.
- (17) Grimme, S. *Chem. – Eur. J.* **2012**, *18* (32), 9955–9964.

- (18) Sure, R.; Grimme, S. *J. Chem. Theory Comput.* **2015**, *11* (8), 3785–3801.
- (19) Sedlak, R.; Janowski, T.; Pitoňák, M.; Řezáč, J.; Pulay, P.; Hobza, P. *J. Chem. Theory Comput.* **2013**, *9* (8), 3364–3374.
- (20) Zhao, Y.; Truhlar, D. G. *Theor. Chem. Acc.* **2008**, *120* (1–3), 215–241.
- (21) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. *J. Chem. Phys.* **2010**, *132* (15), 154104.
- (22) Dion, M.; Rydberg, H.; Schröder, E.; Langreth, D. C.; Lundqvist, B. I. *Phys. Rev. Lett.* **2004**, *92* (24), 246401.
- (23) Vydrov, O. A.; Wu, Q.; Van Voorhis, T. *J. Chem. Phys.* **2008**, *129* (1), 014106.
- (24) Thanthiriwatte, K. S.; Hohenstein, E. G.; Burns, L. A.; Sherrill, C. D. *J. Chem. Theory Comput.* **2011**, *7* (1), 88–96.
- (25) Ireta, J.; Neugebauer, J.; Scheffler, M. *J. Phys. Chem. A* **2004**, *108* (26), 5692–5698.
- (26) Paton, R. S.; Goodman, J. M. *J. Chem. Inf. Model.* **2009**, *49* (4), 944–955.
- (27) Riley, K. E.; Pitoňák, M.; Černý, J.; Hobza, P. *J. Chem. Theory Comput.* **2010**, *6* (1), 66–80.
- (28) Hujo, W.; Grimme, S. *Phys. Chem. Chem. Phys.* **2011**, *13* (31), 13942.
- (29) Dennington, R.; Keith, T.; Millam, J. *GaussView*; Semichem Inc.: Shawnee Mission KS, 2009.
- (30) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H.; Izmaylov, A.; Bloino, J.; Zheng, G.; Sonnenberg, J.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J.; Peralta, J.; Ogliaro, F.; Bearpark, M.; Heyd, J.; Brothers, E.; Kudin, K.; Staroverov, V.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J.; Iyengar, S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J.; Klene, M.; Knox, J.; Cross, J.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R.; Yazyev, O.; Austin, A.; Cammi, R.; Pomelli, C.; Ochterski, J.; Martin, R.; Morokuma, K.; Zakrzewski, V.; Voth, G.; Salvador, P.; Dannenberg, J.; Dapprich, S.; Daniels, A.; Farkas; Foresman, J.; Ortiz, J.; Cioslowski, J.; Fox, D. 2009.
- (31) Becke, A. D. *J. Chem. Phys.* **1993**, *98* (7), 5648–5652.
- (32) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98* (45), 11623–11627.
- (33) Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393* (1–3), 51–57.
- (34) Chai, J.-D.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2008**, *10* (44), 6615–6620.
- (35) Hehre, W. J.; Ditchfield, R.; Pople, J. A. *J. Chem. Phys.* **1972**, *56* (5), 2257–2261.
- (36) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28* (3), 213–222.
- (37) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. V. R. *J. Comput. Chem.* **1983**, *4* (3), 294–301.
- (38) Choi, H.; Min, M.; Peng, Q.; Kang, D.; Paton, R. S.; Hong, S. *Chem. Sci.* **2016**, *7* (6), 3900–3909.
- (39) Simón, L.; Paton, R. S. *Org. Biomol. Chem.* **2016**, *14* (11), 3031–3039.
- (40) Cortopassi, W. A.; Simion, R.; Honsby, C. E.; França, T. C. C.; Paton, R. S. *Chem. – Eur. J.* **2015**, *21* (52), 18983–18992.
- (41) Jackson, K. E.; Mortimer, C. L.; Odell, B.; McKenna, J. M.; Claridge, T. D. W.; Paton, R. S.; Hodgson, D. M. *J. Org. Chem.* **2015**, *80* (20), 9838–9846.
- (42) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *J. Comput. Chem.* **2003**, *24* (6), 669–681.
- (43) Dunning, T. H. *J. Chem. Phys.* **1989**, *90* (2), 1007–1023.

- (44) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. *Chem. Phys. Lett.* **1998**, 286 (3–4), 243–252.
- (45) Pearlman, D. A.; Charifson, P. S. *J. Med. Chem.* **2001**, 44 (21), 3417–3423.
- (46) Luccarelli, J.; Michel, J.; Tirado-Rives, J.; Jorgensen, W. L. *J. Chem. Theory Comput.* **2010**, 6 (12), 3850–3856.
- (47) Miertz, S. *Chem. Phys.* **1981**, 55, 117–129.
- (48) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2009**, 113 (18), 6378–6396.
- (49) Ribeiro, R. F.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2011**, 115 (49), 14556–14562.