MPhys Project

Using Solvated Density Functional Theory to Obtain NMR spectra for Azo Dyes

René C. Castberg

April 16, 2004

Contents

1 Introduction 1

2 Theory 1

2.1 Molecular calculations 1

2.2 Hartree-Fock 3

2.3 Møller-Plesset 3

2.4 Density Functional Theory 4

2.5 Solvated Models 5

2.6 Choice of algorithm 8

2.7 Computational details 8

3 Small molecules 9

3.1 NMR measurements 10

3.2 How the small molecules compare to dyes 11

3.3 Direct NMR comparison gas phase 12

3.4 NMR comparison solvated vs gas phase 15

3.5 NMR comparison with experimental 20

3.5.1 Direct comparison with Expt 20

3.5.2 Accuracy of the solvated model 34

3.6 Overview 38

4 Dyes 39

4.1 Experimental comparison/description 39

4.1.1 NMR Spectroscopy Details 39

4.1.2 Computation of dye molecules 41

4.2 Computed data: solvated NMR vs gas phase 43
4.3 Solvated NMR vs experimental 45
4.4 Accuracy of the solvated models vs gas phase 53

5 Conclusion 55

A Benchmarking 57

A.1 Introduction 57
A.2 Molecule specifications 57
A.3 6-31g(d,p) Basis Set 60
A.3.1 Optimisation Benchmark Results for Smaller Basis Set 60
A.3.2 Solvated Optimisation Benchmarks 63
A.3.3 Frequency Calculation 66
A.4 6-311++g(2d,2p) Basis Set 66
A.4.1 Optimisation Benchmark with DFT 6-311+G(2d,2p) . 66
A.4.2 Optimisation Benchmark with HF 6-311+G(2d,2p) . . 68
A.4.3 Large Basis set MP2 Optimisations 68
A.4.4 6-311+G(2d,2p) Solvated Optimisations 69

B Data 71

C Optimised stuctures 73

C.1 Gas Phase 73
C.2 Cyclohexane 79
C.3 Benzene 83

D Acknowledgements 88

E Bibliography 88

Abstract

Calculations based on Solvated NMR were performed on a number of smaller molecules as well as 3 larger dyes. The calculated data is compared to experimental results and from this it is hoped that calculated NMR spectra of azo dyes can be better understood. The solvated calculations for azo dyes are also investigated to see if they offer any improvement on gas phase calculations. These showed a close agreement when two isomers are compared with one another. For a number of molecules there was some improvement in the spectra while in general the gas phase calculation showed better agreement with experimental results

1 Introduction

Nuclear Magnetic Resonance (NMR) spectroscopy is an extremely useful and widely used technique in chemistry. Trying to understand the spectra for smaller molecules is usually relatively simple. Unfortunately larger molecules can have complicated spectra which change according to the solvent used. It has already been shown that the chemical shifts of experimental azo dyes are comparable to those from computational experiments[9]. These results did show slight discrepancies and it is hoped that they can be improved by performing solvated calculations. I will be performing solvated calculations for the same set of dyes, namely Disperse Red, Disperse Orange and Methyl Yellow using the Gaussian package.

2 Theory

2.1 Molecular calculations

Molecular calculations are usually split into three categories, Molecular mechanics, semi-empirical and ab-initio methods. Molecular mechanics (MM) calculations are based on the classical laws of physics and thus ignore all quantum effects. This is done by using force field methods, which are designed for a specific group of molecules. This means that if you end up using a force field which has been parameterised for molecules other than the one you are working on, you could end up with results that are completely incorrect. Due to the fact that MM has simplified the molecules to the extent of balls and springs it becomes possible to model 1000s of atoms. This is not the case for semi-empirical and ab-initio methods; where we are modelling the electronic effects of the molecules. This is done by solving the Schrödinger equation:

$$\frac{\hbar^2}{2m}\nabla^2\Psi_n(\boldsymbol{r}) + V(\boldsymbol{r})\Psi_n(\boldsymbol{r}) = E_n\Psi_n(\boldsymbol{r})$$

The Schrödinger equation cannot be calculated exactly due to the electron-electron interaction and for this reason it becomes computationally expensive. Calculations are usually simplified using a number of methods. The largest simplification is the semi-empirical calculation. This involves simplifying the Schrödinger equation by using experimental data to substitute computationally intensive calculations. This implies that you need experimental data about that system in question. If there is good data then semi-empirical calculations are quite accurate but will seldom be able to challenge ab-inito methods in terms of accuracy or diversity of molecules. Semi-empirical methods, on the other hand, give you a huge benefit in terms of running many times faster. Ab-initio methods calculate the full Schrödinger equation. This makes them inherently slow to calculate and you will often see a scaling proportional to N^4, this means that you typically won't be modelling more than 100 atoms in a particular system. Ab-initio methods have to be simplified in order for the calculation to be at all possible. The most common methods are Hartree-Fock (HF), Møller Plessent (MP) and Density Functional Theory (DFT).

2.2 Hartree-Fock

Hartree Fock makes a large approximation, by ignoring the electron interactions, essentially creating a static field for the electrons to move around in. This in turn allows the Schrödinger equation to be solved as if it were a single bodied problem. Unfortunately, this introduces a large error of the order of 14eV for a separation of 1Å, thus drastically altering the structure of the molecule. It includes the effects from the Pauli exclusion principle by using the anti symmetric form of the wave-function. The approximations in the HF equation can still give a reasonably accurate model, unfortunately the large simplification can change a molecule to such an extent as to slightly alter the outcome of an experiment. There are many models which take the standard HF equations and improve it by adding various perturbations, often vastly improving the accuracy.

2.3 Møller-Plesset

Møller-Plesset methods are based on a single closed shell Hartree Fock determinant. This information is then used to perturb the wave function. Essentially what is achieved is splitting the Hamiltonian into two: $H = H_0 + \lambda V$. Here H_0 can be solved exactly using HF theory and λV, which is a small correction in comparison to H_0, can be calculated using many body perturbation theories. If a second order perturbation is used the method is known as MP2. MP methods are very computationally demanding and for this reason will not be used.

2.4 Density Functional Theory

DFT is a method which is based on the knowledge that you can model the electron correlation on the electron density by using a general functional. In 1964 Hoenberg-Kohn [4] published a paper in which they demonstrated that it was possible to model the ground state energy directly by using the electron density. This information was not enough and it was not until a paper publish by Kohn-Sham in 1964 [6] that it became possible to model systems with DFT. Kohn-Sham approximated the functionals used in DFT by splitting it into several energy terms.

$$E = E^T + E^V + E^J + E^{XC}$$

E^T is the Kinetic energy,

E^V is the potential (n-e, n-n interaction),

E^J is the e-e repulsion energy,

E^{XC} is the exchange correlation of the electron.

The kinetic, potential and electron-electron repulsion energy are all functions of the electron density and can thus be calculated relatively easily. This makes up what we generally know as the classical energy. The remaining energy exchange-correlation (E^{XC}) was shown in the paper by Kohn-Sham to split into two parts, the exchange energy E^X and the correlation energy E^C. The exchange energy was initially approximated with the Local Density Approximation (LDA), unfortunately this had limitations and it was soon replaced with a slightly more complicated function by Becke. This is essentially a gradient corrected LDA. In the same way the correlation energy can also be calculated since it too is proportional to the electron density. A very well known correlation function is the Lee Yang Parr(LYP) functional, this is essentially the electron correlation with a gradient correction. By combining the Becke and LYP correlation functionals we get a functional pair called BLYP. This can be further improved by combining these functionals with others like the Hartree Fock. This lead to the 3 parameter functional, B3LYP:

$$E^{XC}_{B3LYP} = A * E^{X}_{Slater} + (1 - A) * E^{X}_{HF} + B * E^{X}_{Becke} + C(E^{C}_{VWN} + E^{C}_{LYP})[3]$$

This functional includes the Local Density, the Hartree Fock energies, the Becke exchange functional, LYP and Vosko Wilk Nusair (VWN) correlation functionals. The VWN correlation is only there to provide the excess local correlation that is not found in the LYP.

2.5 Solvated Models

Solvation models are based on the same principles as the (gas phase) models already discussed, except in this case the effects of the solvent must be taken into account. One method is to take the molecule we are studying and surround it with the solvent, unfortunately this would require enormous computational requirements because it is not known how far the interaction can reach. One way of solving this is to add layer upon layer of water molecules until the properties of the molecule in question remain constant. This again is not making efficient use of the computational resources. Using a method called Self Consistent Reaction Field (SCRF) it is possible to have a larger efficiency. This can be achieved by taking the solvent as a uniform polarizable medium with a dielectric constant ε. The solute is then placed in a cavity, usually created by several interlocking spheres within the solvent. Creating this cavity has a cost on the system energy because we are disturbing the continuum, leading to a destabilisation.

This is countered by the stabilisation caused by the interaction of the solute with the solvent (usually the van der Waals energy) and due to the charge distribution within the cavity, inducing a polarisation in the solvent which causes the solute to be stabilised. Addition of these three energies allows us to specify the Solvation energy:

$$\Delta G_{Solvation} = \Delta G_{Electrostatic} + \Delta G_{Dispersion} + \Delta G_{Cavity} \ [5]$$

There are a number of different ways in which the cavity can be specified. Gaussian 03 uses the United Atom model. This model takes all the non Hydrogen atoms and places spheres around them to construct a van der Waals surface. Any hydrogen atoms are placed in the same sphere as the parent atom. Using the van der Waals surface we can calculate the ΔG_{Cavity} since this is inaccessible to the liquid and will cause a loss in energy. Two additional surfaces have to be constructed to be able to calculate the free energy, the solvent excluding surface and the solvent accessible surface. The solvent excluding surface is generated by taking the van der Waals radius and multiplying this by a factor of 1.2 [10]. This is also often smoothed by shifting the centre of the van der Waals spheres, thus creating multiple surfaces thereby removing any crevices. The solvent accessible surface is created by tracing a simplified solvent molecule over the van der Waals spheres, and using the centre of the molecule to create a surface.

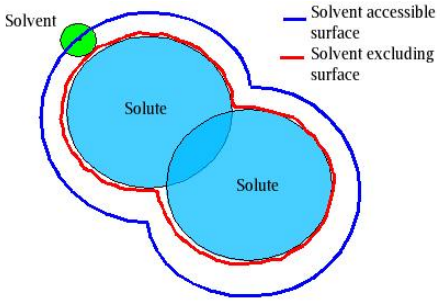

Figure 1: SCRF Cavity

Another example can be taken by considering H_2CO (Figure 2). In this case one would place a sphere of radius 2.34Å around the CH_2 (compared to a typical C radius of 2.04) and a sphere around the O with a radius of 1.692Å. These surfaces are then used for the solvation model.

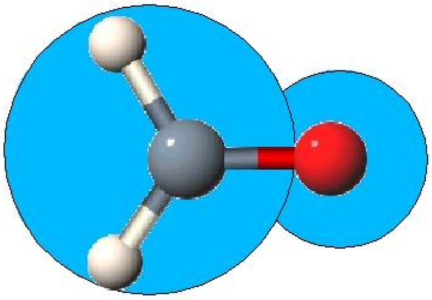

Figure 2: H_2CO with excluded volume

The $\Delta G_{Dispersion}$ is calculated using the solvent accessible surface while the Solvent excluding surface is used to calculate the $\Delta G_{Electrostatic}$ energy contribution.

2.6 Choice of algorithm

For this project the DFT algorithm was chosen. The main reason for this is that you have similar accuracy to more complex methods while only using the computational power equivalent to HF calculations. HF, while a true ab-initio method, does not provide as accurate results with the same number of basis function. As can be seen in Appendix A the difference between DFT and HF is minor in terms of computational time, but as you can see MP2 methods typically require 2-3 times more resources than DFT or HF. The basis set chosen was 6-311G+(2d,2p) for larger molecules. The basis set includes a disperse function which allows the lone pairs to be treated with reasonable accuracy. Having such a large basis set implies that we are restricted in how large the molecules could be. This however was not an issue with the dyes we were using as there were at most 41 atoms. For the initial testing we used a smaller basis set of 6-31g(d,p) to check that the molecule would optimise and give a reliable structure. For the solvated model we chose to use the standard Polarizable Continuum Model (PCM) supplied with Gaussian, this is the IEF-PCM method. Gaussian does provide more accurate models but these are more computationally demanding.

2.7 Computational details

Most of these calculations were performed on either the Edinburgh University Supercomputer, a Sun Fire 15K (52 processors) or on Red Hat 9.0 Linux PC's. All the solvated calculations were carried out on Gaussian 03 revision B.02 while some of the gas phase calculations were carried out on Gaussian 98, revision A.7. Where possible Gaussian was left to complete calculation, the exceptions to this were the disperse orange cis molecules. These molecules were not rigid enough to be optimised without intervention. In this case, Gaussian was left to optimise until there was little change in the forces on the molecule.

```
         Item               Value     Threshold  Converged?
 Maximum Force            0.000023     0.000450     YES
 RMS     Force            0.000006     0.000300     YES
 Maximum Displacement     0.003757     0.001800     NO
 RMS     Displacement     0.001060     0.001200     YES
--
         Item               Value     Threshold  Converged?
 Maximum Force            0.000042     0.000450     YES
 RMS     Force            0.000009     0.000300     YES
 Maximum Displacement     0.003756     0.001800     NO
 RMS     Displacement     0.001060     0.001200     YES
```

In this example we see there has been little change in the Maximum displacement, the only value which has not yet converged, and for this reason the optimisation was stopped. The next stage was started and the NMR spectra was calculated.

In experiments any NMR data obtained will usually be given relative to Tetramethylsilane, this reduces the ambiguity of the experimental conditions. The same is true of the computational data, the calculated data is the shielding so in order to obtain the chemical shifts the NMR spectra of TMS also has to be obtained. Subtracting the Isotropic NMR shift from the shift for TMS will give you the shift which is comparable to the experimental results.

3 Small molecules

In order to be able to make any conclusions from the dye data it was decided that it was necessary to do some calculations on smaller molecules. These molecules are to be representative of the dyes, each with specific features. The molecules that were chosen are: Benzene, Phenol, Aniline, Chloroaniline and Paraxylene. The calculations were performed in a number of solvents: Gas, Water, Benzene, Dimethyl Sulphoxide (DMSO) and Chloroform. The reasons for these choices were to cover the largest possible range of interaction between the solvent and molecule. Gas calculations were the first to be performed, these were quicker than the solvated models and gave a partially optimised structure which can be further optimised for a specific solvent. Water calculations were used since water is a very polar solvent and usually alters the NMR peaks drastically. Chloroform was used since the NMR spectra for most molecules are readily available, additionally chloroform is a common solvent for NMR experiments. DMSO is another common solvent, it will give a good indication on how a polar solvent would behave. This was used instead of water since not all the molecules listed will readily dissolve in water (benzene and paraxylene are completely insoluble). Benzene and Cyclohexane were used since these usually don't interact much with the molecule and should give a closer match with the gas phase data.

3.1 NMR measurements

The experimental NMR data for these molecules were taken on a Bruker 360DMX spectrometer operating at 360MHz. The data was analysed on Xwinnmr (version 2.0, Brukner UK limited) and MestRe-C (version 3.5.1b, www.mestrec.com)

3.2 How the small molecules compare to dyes

As I have mentioned the small molecules were chosen to have as many of the features of the dyes as possible. The only feature we were unable to mimic was that of the N=N bond. The only way we would have been able to achieve this was by working with larger molecules like azobenzene and this would have defied the point of doing the small molecules.

Benzene

The benzene molecule shows the effect of how the aromaticity can be affected by solvents, this will only give an indication since the dyes being studied will contain substitute groups and thus behave slightly differently.

Phenol

Phenol will give a good indication of what will happen to any hydroxyl groups on molecules. It should also provide information on any hydrogen bonding effects, if these are present.

Aniline

Aniline should provide information on how the amine group on disperse orange will behave.

Chloroaniline

Chloroaniline will give a useful comparison to Aniline. The addition of the chlorine will give an indication of the effect of the electron withdrawing group.

Paraxylene

This is useful when comparing with Methyl Yellow since this molecule also contains the same CH_3 group. It will also show how well Gaussian handles a simpler molecule that does not contain any electron withdrawing or donating groups.

3.3 Direct NMR comparison gas phase

The test molecules were all initially run in a gas phase environment, this optimised the molecule from an initial guess into an accurate structure, also enabling us to get an estimate of how the molecules were affected once they were removed from the vacuum and placed in a solution. These effects would be similar to those seen with the dye molecules.

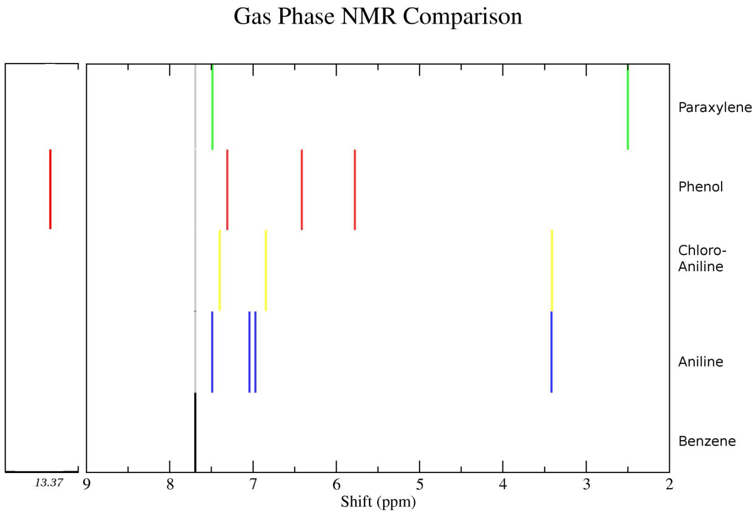

Figure 3: Comparison of Gas Phase NMR

In Figure 3 we see the NMR spectra of all the test molecules plotted with benzene used as a reference spectra that can be seen in the background. The phenol O-H peak is at 13.37ppm and for this reason not plotted on the same scale. It should also be noted that the average of the peaks was taken for equivalent protons. The reason that multiple peaks were obtained for equivalent protons is that the computational package can differentiate where a proton is located. The simplest example of this is by Phenol (Figure 4). In Figure 4 it is clear that hydrogen 8 and hydrogen 7 are in equivalent positions. It should also be obvious that if the computational package calculates the shift for the static molecule as it is displayed, proton 8 will have a charge associated with it because of the vicinity of the hydroxyl proton. On the other hand, proton 7 will hardly be affected. Protons 7 and 9 will be in a completely different environments and thus provide two different values for their chemical shift. Unfortunately, this difference is not possible to differentiate in experimental NMR data and thus cannot be analysed in more detail.

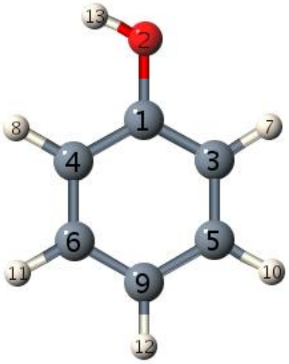

Figure 4: Phenol Symmetry

Benzene also showed two different values for the proton chemical shifts. The difference was between protons 2,12 and 7,8,10,11 (Numbered as in Figure 4). The reason for this was that Gaussian refused to optimise benzene in the C6 symmetry and instead choose C3 symmetry, thus giving two peaks. The benzene values displayed are the average of these two values.

Starting from the lowest chemical shift in Figure 3 we first come across the paraxylene CH_3 hydrogens that peak where we would expect to see them. Next see the NH_2 protons from Aniline and Chloroaniline. These are also in what is considered a typical region for the signals to appear. The computational package shows that the shift for the amine protons is approximately uniform and the chlorine has little effect on these chemical shifts. The rest of the peaks all show aromatic hydrogens. The phenol spectra seems to have spread out and shifted to a slightly lower ppm than would be expected from experimental data. This suggests that Phenol is a problematic molecule for calculating the computational spectra. Additionally the O-H peak is up at 13.37ppm which is much higher than would be expected for a typical O-H shift. A reason for this is that the hydrogen attached to the oxygen is heavily deshielded and in the gas phase calculations there is nothing for it to interact with. Generally in experiments we would expect the hydroxyl group to interact quite strongly with the solvent. The chlorine is also seen to shift the aromatic protons to the high field region indicating that the chlorine has caused a slight increase in the shielding on the aromatic protons.

Calculated GAS Phase Data

Molecule	Aromatic			N-H_2	O-H	C-H_3
	Ortho	Meta	Para			
Benzene	7.69					
Phenol	5.77	7.31	6.41		13.37	
Paraxylene	7.48					2.50
Chloroaniline	6.85	7.40	Cl	3.41		
Aniline	6.97	7.49	7.04	3.42		

3.4 NMR comparison solvated vs gas phase

From experimental studies it is clear that the effects of the solution on our target molecules have a large impact on the NMR spectra. For this reason calculations were performed for each of the molecules in a number of test solutions. For all the following spectra the placement of the hydrogen atom is the same as for the gas phase calculation unless specifically stated.

Legend:
A = Any Aromatic Hydrogen
P = Para Position
O = Ortho Position
M = Meta Position
N-H = Hydrogens Attached to N
O-H = Hydrogen Attached to O
C-H = Hydrogens Attached to Methyl groups

Aniline (Figure 5) The solvated spectra show very similar results to the aromatic hydrogens with the largest deviation apparent for water and DMSO. Benzene and Cyclohexane showed the least change in the spectra. Water and DMSO have a large dielectric value and will therefore shift the spectra more than benzene and cyclohexane. The effect of Benzene and cyclohexane is small, as would be expected for molecules with such a small dielectric value. The fact that there is little hydrogen bonding in these two solvents will also prevent there from being too large a shift. Chloroform seems to have slightly stronger interaction than benzene and cyclohexane and causes a slight deshielding. It is also apparent that the effect on the N-H_2 group is much larger with a shift close to 1ppm for water and DMSO. This shows the extent to which hydrogen bonds are treated as an explicit interaction.

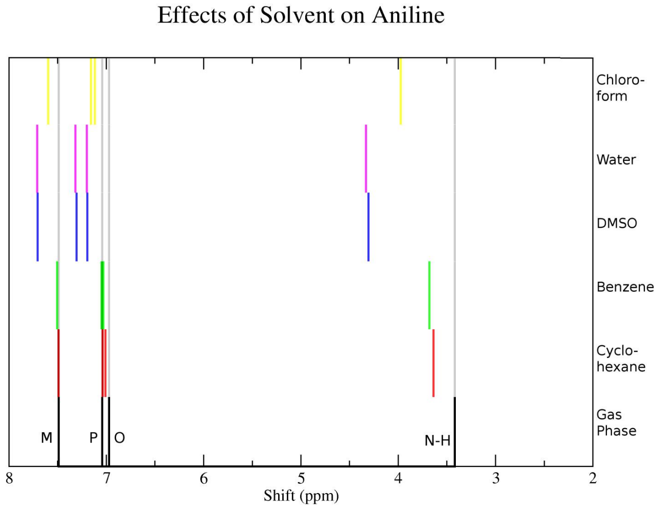

Figure 5: Aniline: Gas Phase vs Solvated Model

Benzene (Figure 6) Benzene showed very little variation in the spectra's, the largest shift with respect to the gas phase was approximately 0.3ppm. Here again Chloroform is placed about halfway between the benzene/cyclohexane and water/DMSO pairs.

Chloroaniline (Figure 7) This spectra is very similar to that of aniline showing shifts of approximately the same magnitude (around 1ppm for the N-H_2 group). When comparing the spectra to aniline it can be seen that the addition of chlorine has introduced some shielding to the Meta and Ortho positions and deshielded the N-H hydrogens. This is to be expected due to the chlorine shifting some of the charge density towards itself and therefore

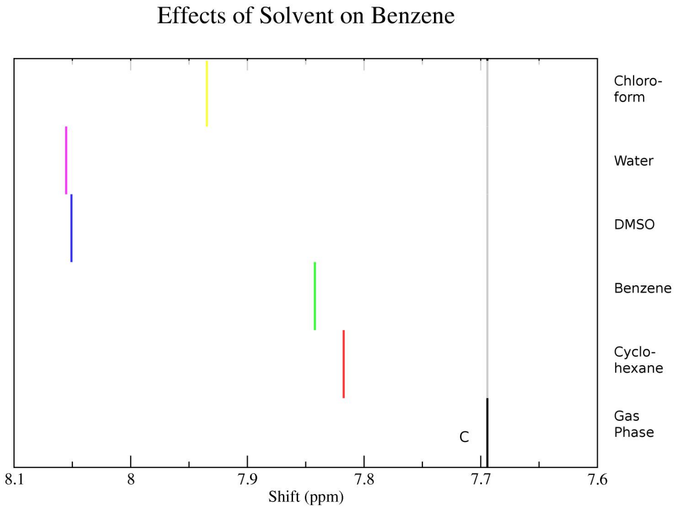

Figure 6: Benzene: Gas Phase vs Solvated Model

leaving a lower electron charge density on the nitrogen.

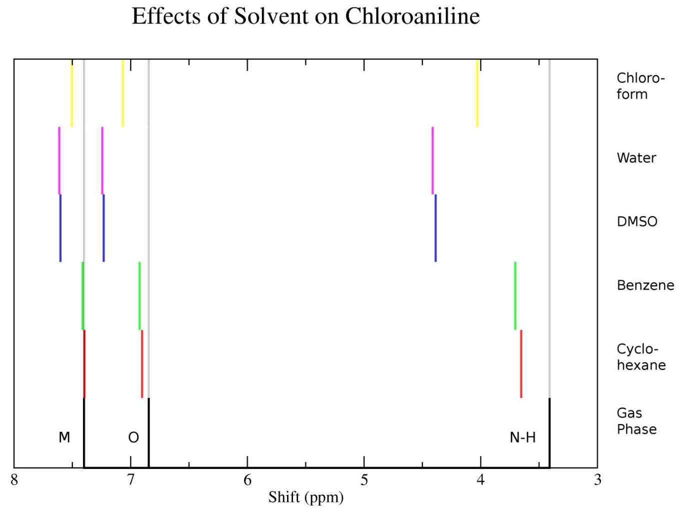

Figure 7: Chloroaniline: Gas Phase vs Solvated Model

Paraxylene (Figure 8) Paraxylene was the least effected calculation on the introduction of solvation with shifts of only 0.1ppm to 0.2ppm for cyclohexane and water respectively. The shifts for the aromatic hydrogens are similar to the other molecules that have been discussed. The C-H_3 group shows a smaller shift indicating that the solvents cause an increased shielding.

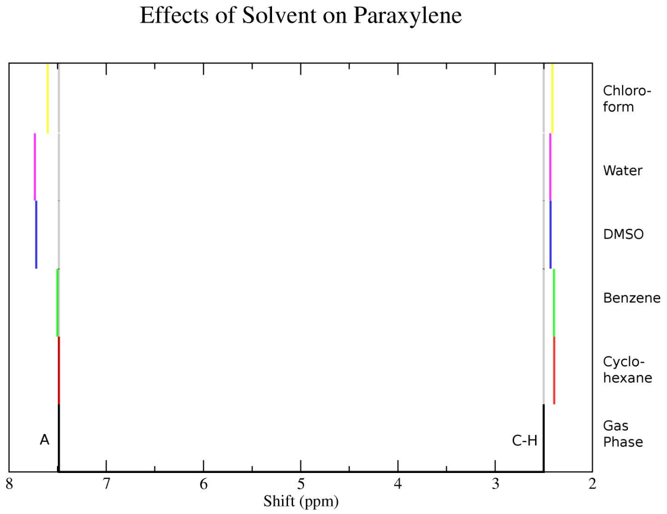

Figure 8: Paraxylene: Gas Phase vs Solvated Model

Phenol (Figure 9) Phenol showed the largest variations between gas phase and solvated models. The Hydroxyl group shifted wildly from 13.37 in the gas phase calculation to 4.83 in cyclohexane giving a difference of 8.54ppm. This is a huge shift in terms of ^1H spectroscopy and would not be expected in experiments. The aromatic protons showed a smaller shift of around 1.6-1.7ppm indicating as large an effect a solvent will have on phenol. By comparing only the solvated spectra the difference is much smaller with approximately

2.8ppm for the hydroxyl proton and 0.25ppm for the aromatic protons.

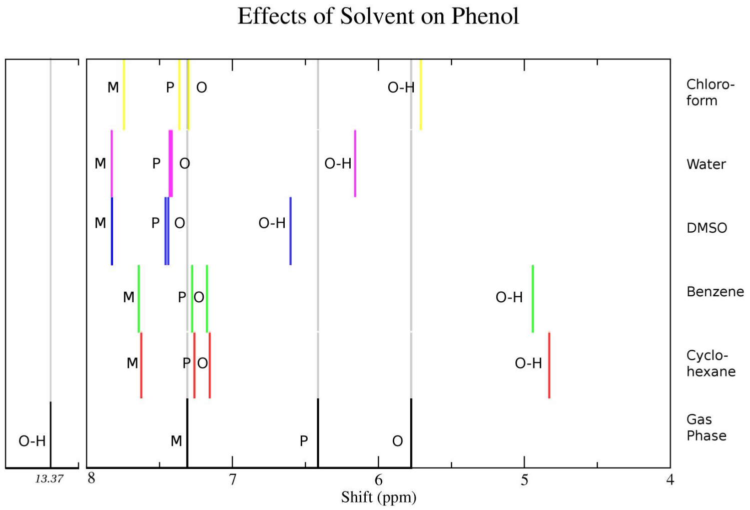

Figure 9: Phenol Gas Phase vs Solvated Model

It is clear from these graphs that two pairs of solutions show very similar properties. The two with the closest match were water and DMSO with about 1/100 of a ppm for all the molecules with the exception of the hydroxyl proton. As it will become clearer when comparing experimental results Gaussian obviously has a problem in modelling phenol. Another two with very similar results are Benzene and Cyclohexane where the largest differences are again observed in the hydroxyl proton on phenol.

3.5 NMR comparison with experimental

3.5.1 Direct comparison with Expt

In comparing the experimental shifts with the calculated it is immediately obvious that the shifts aromatic hydrogens have are all over estimated by the IEF-PCM method. They almost all have an increased shielding by approximately $\frac{1}{2}$ppm. The amine group on Aniline and Chloroaniline showed increased shielding with respect to the experimental results in all the spectra with exception to DMSO. Couple this to the large δ for phenol and it becomes clear that DMSO is a problematic solvent. Phenol on the other hand showed an increase in the shielding for all the spectra except for cyclohexane indicating a possible interaction between cyclohexane and phenol in the experimental results.

In all the plots the experimental results are placed above the computed results. Data for Spectra with Chloroform was plotted using information from the *Aldrich library of* ^{1}*3C and* ^{1}*H FT NMR spectra* [7]. I have only included the spectra which show some interesting information.

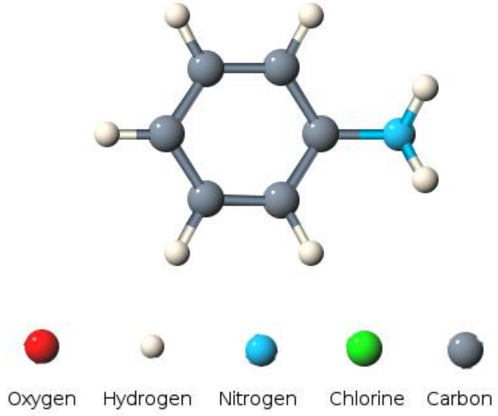

Figure 10: Aniline Molecule

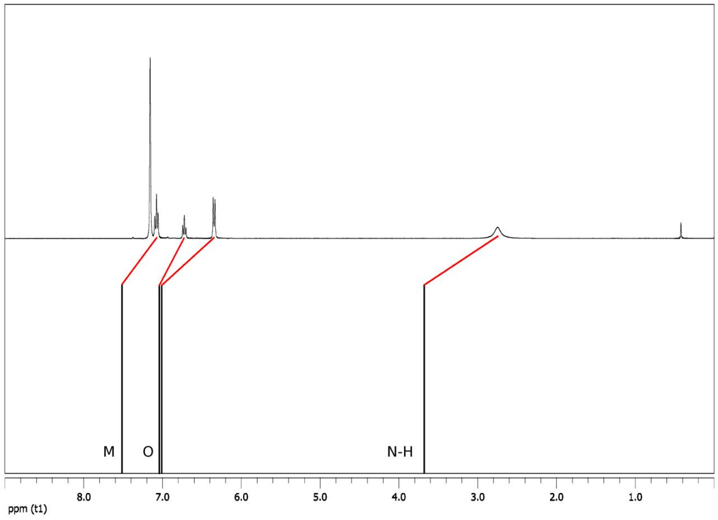

Figure 11: Aniline solvated in Benzene Expt vs Computational

Aniline Comparing the benzene experimental results with computed shows that the aromatic protons are shifted down field by about $\frac{1}{2}$ppm for the Ortho and Meta positions while the para is shifted by about $\frac{1}{4}$ppm. The amine group is shifted by about 1ppm.

Cyclohexane shows a similar spectra except here the amine group is only

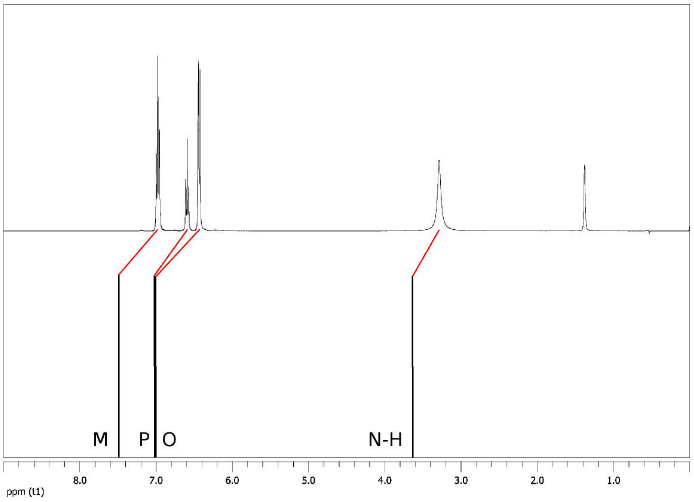

Figure 12: Aniline solvated in Cyclohexane Expt vs Computational

shifted upfield by about $\frac{1}{4}$ppm and there is a slightly smaller shift with respect to the ortho position.

The benzene and cyclohexane show a very similar computed spectra. Comparing the spectra cyclohexane shows a slight upfield shift. This indicates a reduced electron density on the protons when compared to the same situation on benzene. This would indicate that there is less interaction than with cyclohexane.

Chloroform also shows a similar spectra to Cyclohexane. There is a reasonably constant shift of around $\frac{1}{2}$ppm. Chloroform also shows an interchange of the positions of the para and ortho protons. The computational results are separated by 0.05ppm while the experimental separation is 0.13ppm thus in relatively close agreement. From the benzene and cyclohexane spectra it can be seen that the shift for the ortho positions is larger than the para position. In the chloroform spectra it shows that the peaks for the ortho

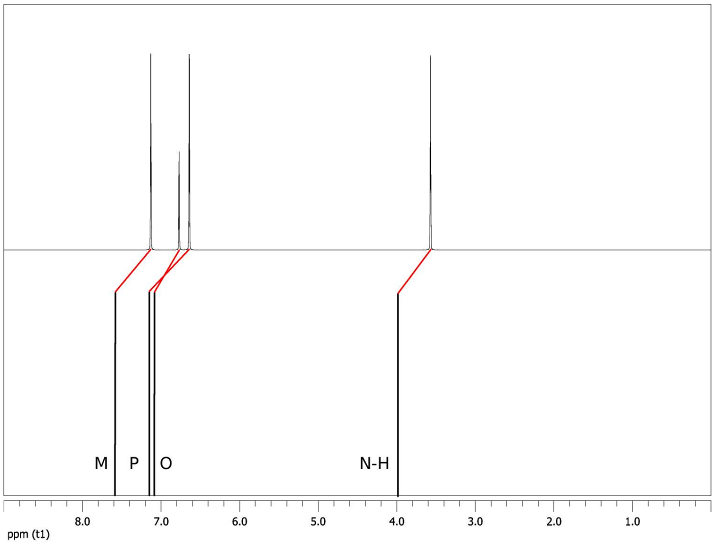

Figure 13: Aniline solvated in Chloroform Expt vs Computational

and para have interchanged is indicative that chloroform causes an increase in the electron density around the ortho position with respect to the para showing possible interaction of the chloroform with aniline.

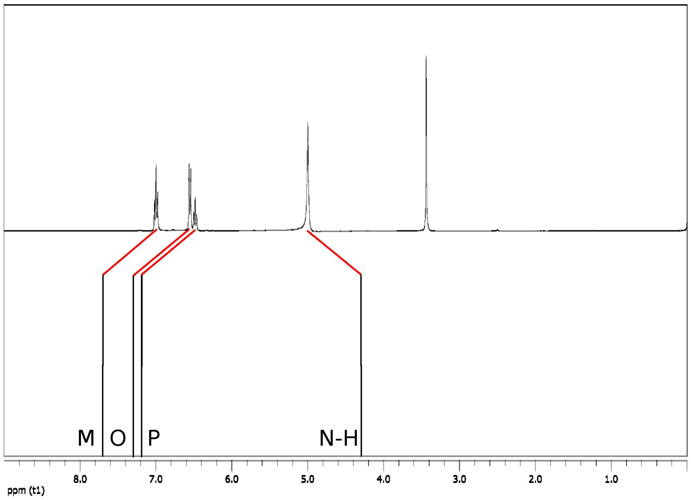

Figure 14: Aniline solvated in DMSO Expt vs Computational

DMSO showed a constant shift of approximately $\frac{3}{4}$pmm for the aromatic carbons. Unlike the other solvents DMSO showed a clear shift in the low-field direction for the amine group by $\frac{3}{4}$pmm. This points to a problem in determining the correct electron density in the presence of DMSO.

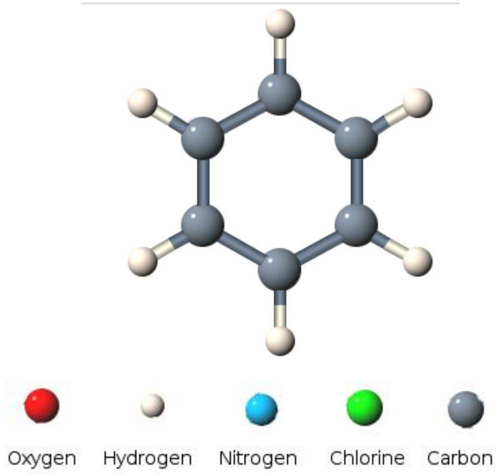

Figure 15: Benzene Molecule

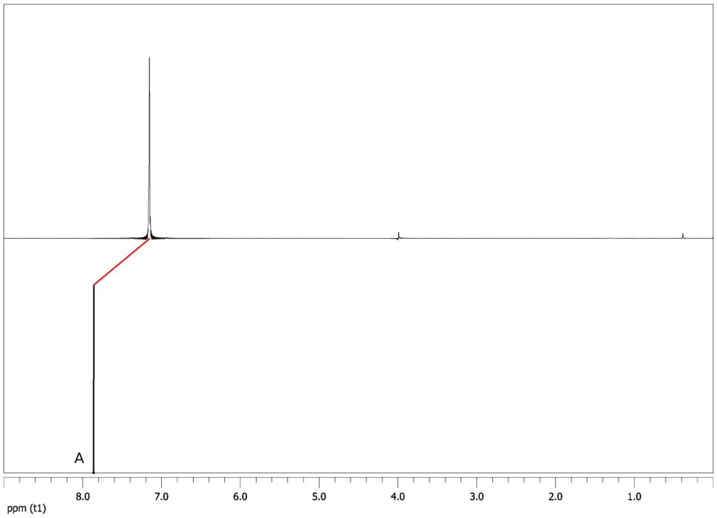

Figure 16: Benzene solvated in Benzene Expt vs Computational

Benzene Benzene solvated in benzene and cyclohexane show a similar trend to the aromatic hydrogens belonging to aniline. There was a general shift of approximately 0.6ppm for all the solvents used showing that there is a general issue with the under estimation of the electron density. This is not so much an issue with the solvation model but rather a general problem with

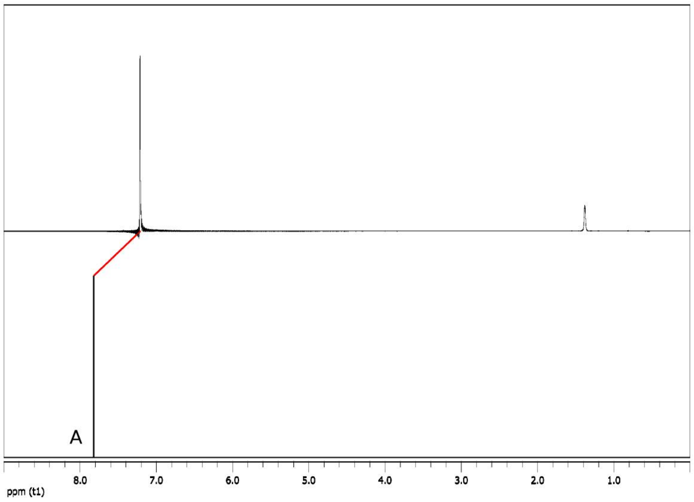

Figure 17: Benzene solvated in Cyclohexane Expt vs Computational

the model. Although for benzene and DMSO there was a slight increase of approximately 0.09ppm which would indicate the interaction isn't modelled as perfectly it should allow recognition of the spectra without too much trouble.

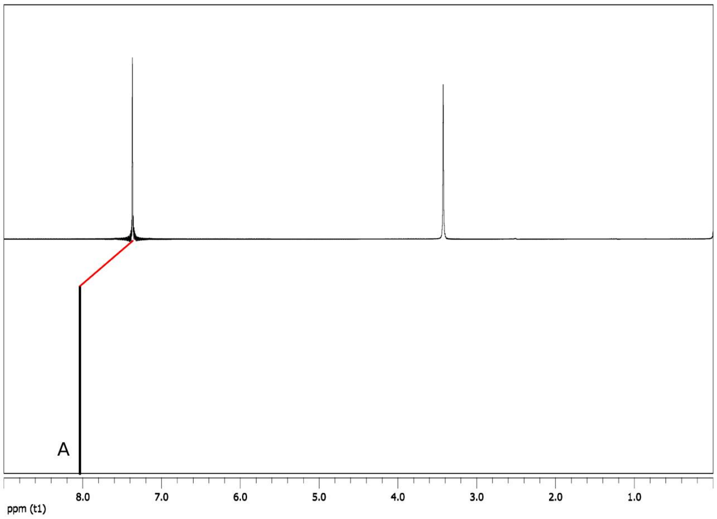

Figure 18: Benzene solvated in DMSO Expt vs Computational

In general, the benzene spectra were quite accurately predicted when comparing them with respect to each other which was to be expected since benzene is a relatively simple molecule.

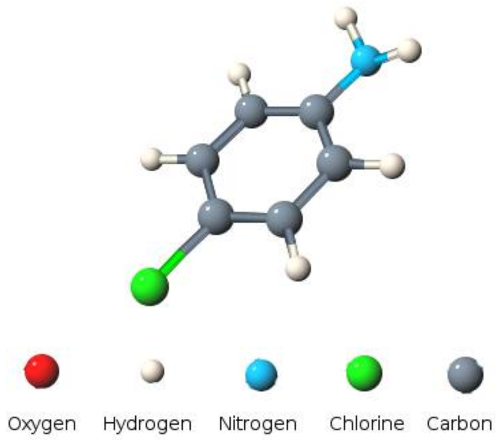

Figure 19: Chloroaniline Molecule

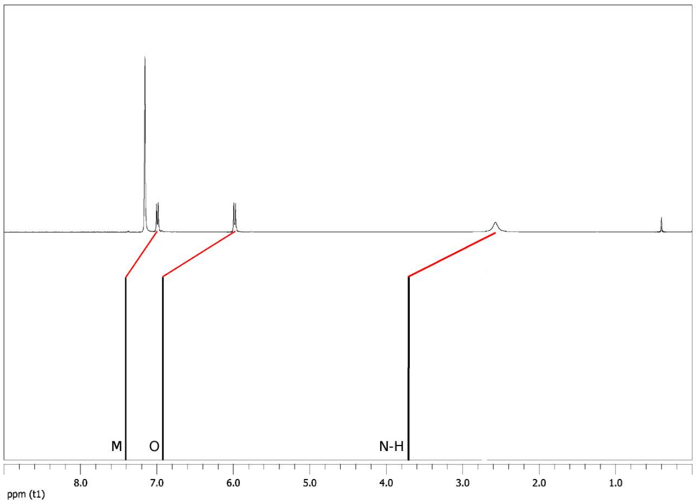

Figure 20: Chloroaniline solvated in Benzene Expt vs Computational

Chloroaniline The aromatic shift was very similar to those of aniline. With the exception of benzene there are all within about 0.1ppm of the aniline spectra. Using benzene as the solvent seems to have increased the shift of the ortho proton by 0.3ppm with respect to aniline indicating a possible interaction of benzene with the chlorine causing an increased electron density.

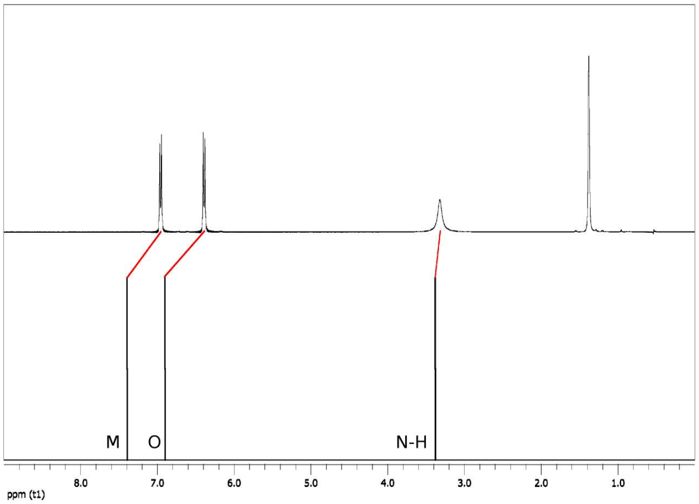

Figure 21: Chloroaniline solvated in Cyclohexane Expt vs Computational

When compared to a benzene molecule you will see a very large interaction difference between cyclohexane and benzene. Cyclohexane is shifted up by 0.8ppm indicating a substantial decrease in the electron density on the amine that may indicate a stabilisation by benzene.

Again with DMSO we see the down field shift for the amine group while the rest is shifted up field with respect to the computed. The shift for the aromatic protons are slightly greater than the other solvents but similar to those produced by aniline.

The chloroform shifts are consistently out by about 0.6ppm indicating a general problem in calculating the interaction.

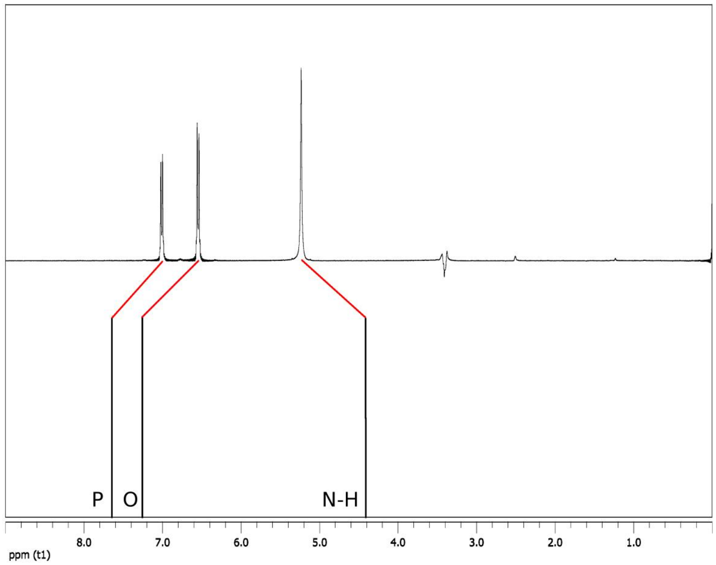

Figure 22: Chloroaniline solvated in DMSO Expt vs Computational

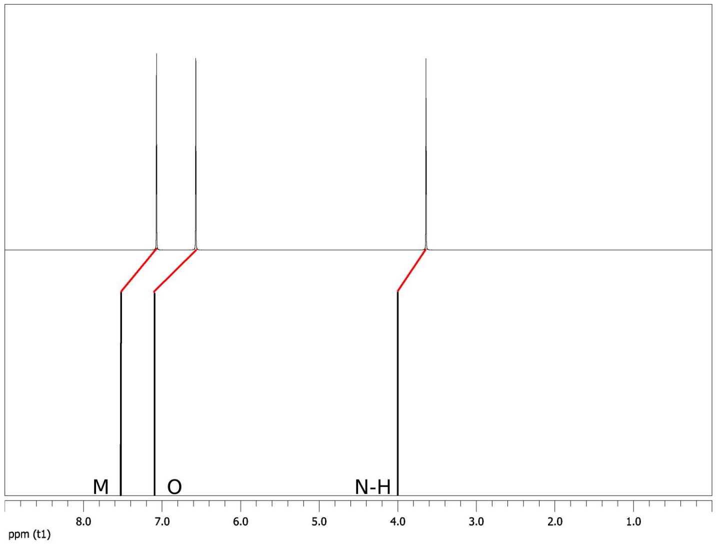

Figure 23: Chloroaniline solvated in Chloroform Expt vs Computational

Paraxylene Paraxylene shows the most consistent shift of all the molecules tested this is because paraxylene does not have any electron withdrawing or donating groups with large effects.

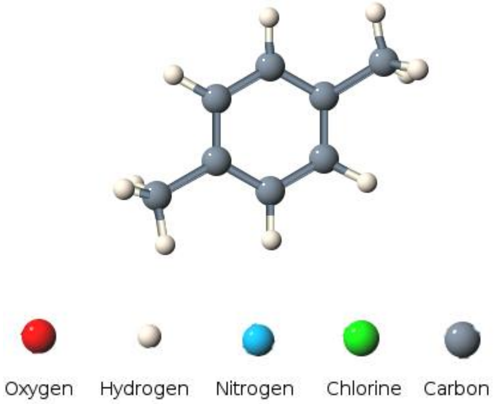

Figure 24: Paraxylene Molecule

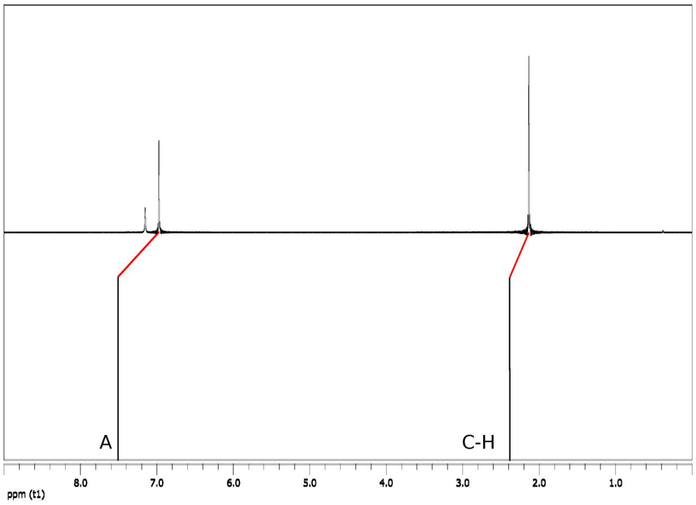

Figure 25: Paraxylene solvated in Benzene Expt vs Computational

DMSO showed the largest shift for the aromatic hydrogens shifting it by 0.67ppm that indicates a problem in accounting for the effects of DMSO.

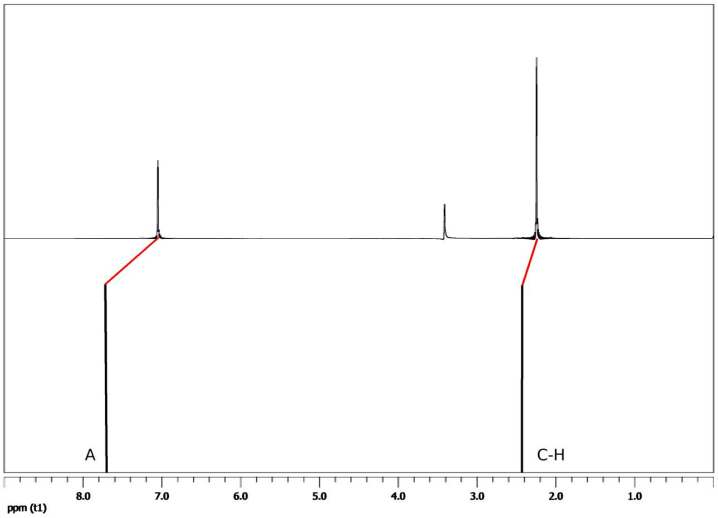

Figure 26: Para-xylene solvated in DMSO Expt vs Computational

Paraxylene shows a general trend; shifting the aromatic protons up field by about 0.5ppm and about 0.2ppm for the methyl group. This shows that for simple molecules the model works relatively accurately.

Phenol As in the previous molecules, Phenol's aromatic hydrogens are relatively similar in terms of their shift. The shift was in the region 0.43 to 0.69 aromatic hydrogens. The hydroxyl group shifted from 0.88pmm to -1.11ppm. This indicates there is a large problem in calculating the response of the hydroxyl group depending on the solvent it is in.

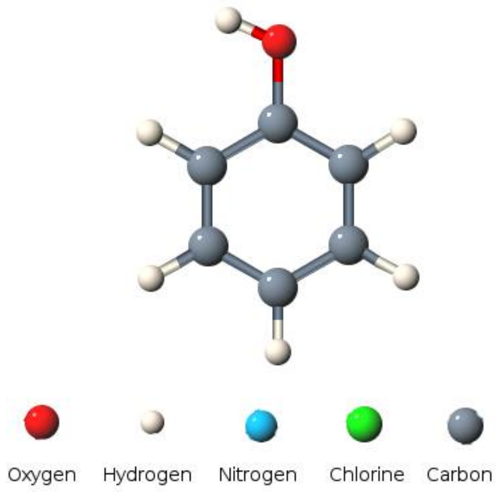

Figure 27: Phenol Molecule

Phenol solvated in Benzene shows a very similar computed spectra to the gas phase data, but when it came to the experimental data the hydroxyl proton was shifted by an almost equal amount in the opposite direction. This gives an indication of the sensitivity of the hydroxyl group on the environment.

DMSO caused an increase in the chemical shift when compared to the calculated results which is in contrast to amine groups which all showed a shift. This indicates that there is an increase the electron density for the amine groups while the hydroxyl group caused a decrease in electron density when compared to the calculated results.

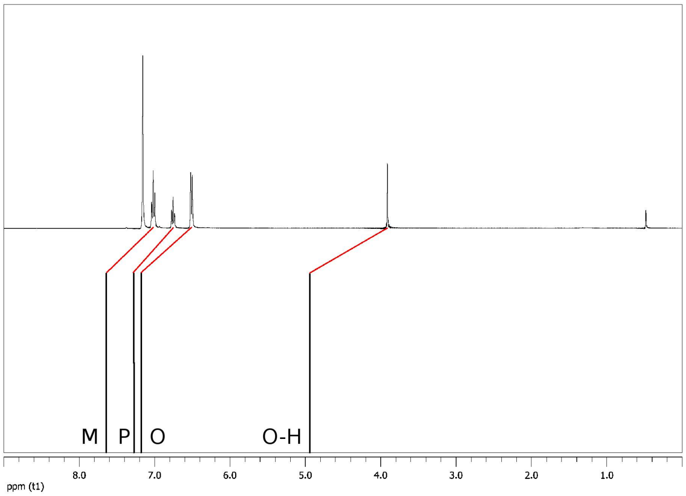

Figure 28: Phenol solvated in Benzene Expt vs Computational

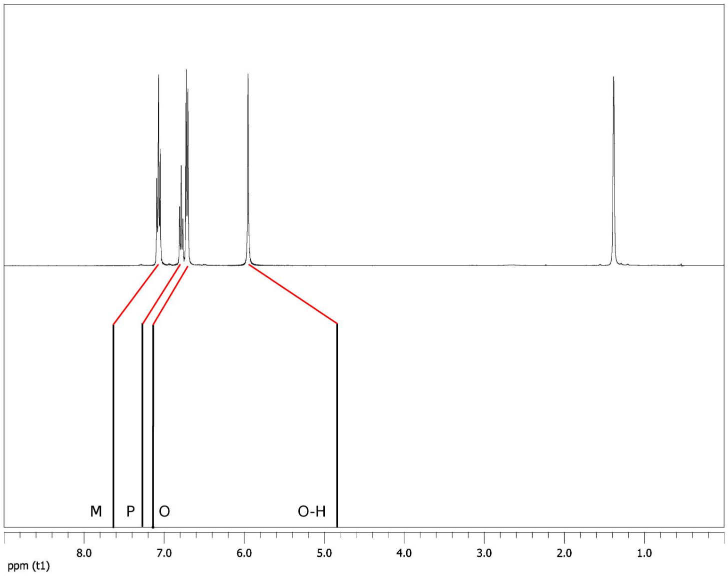

Figure 29: Phenol solvated in Cyclohexane Expt vs Computational

3.5.2 Accuracy of the solvated model

As we have seen the computational methods show a relatively close agreement for the aromatic protons. Although from these spectra it is still not clear if

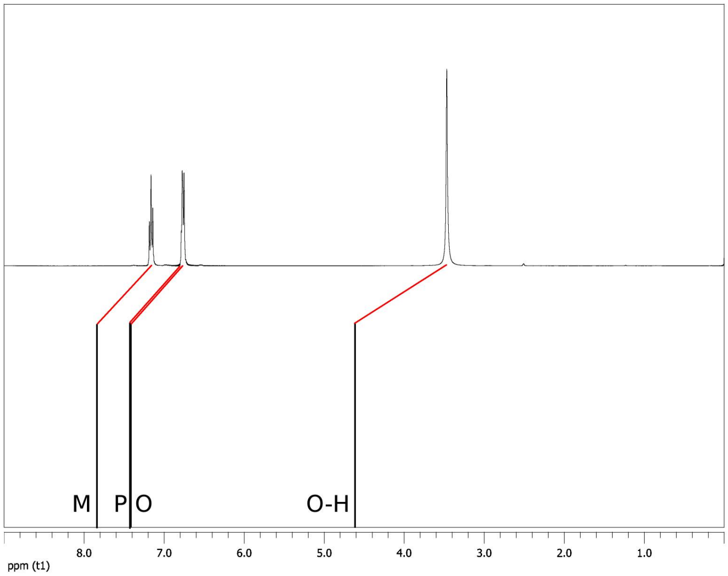

Figure 30: Phenol solvated in DMSO Expt vs Computational

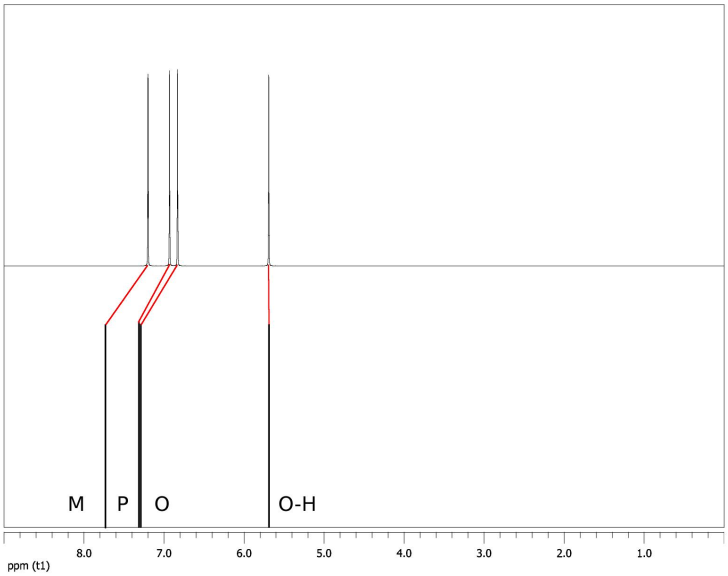

Figure 31: Phenol solvated in Chloroform Expt vs Computational

there had been an improvement in the ability an accurate spectra or they have just been shifted. In order to compare the spectra I have taken the difference between two protons. For aniline, Chloroaniline and phenol I have taken the difference between meta and the other sites the proton could occupy.

Comparing the experimental results with the solvated and gas phase results it becomes apparent that the solvated models don't always improve the spectra.

Benzene

Benzene solvated model showed a larger error than the gas phase calculation when compared to the experimental results. The only exception to this is in phenol where there was a vast improvement. In Phenol the error was down from 1.04ppm to 0.03ppm for the ortho proton. Para showed less of an improvement, from 0.65ppm to 0.11ppm. As can be expected the hydroxyl showed a tremendous improvement from -9.16ppm to 0.4ppm.

Cyclohexane Cyclohexane showed a varying improvement, It showed an improvement for the amine groups on aniline and Chloroaniline, while all the aromatic carbons where larger error, here again the exception is phenol where it predicted the peaks with a much larger accuracy.

Chloroform The solvated model for chloroform followed the same pattern as cyclohexane.

DMSO DMSO showed errors which were smaller than with either Chloroform or Cyclohexane, but again it was only able to predict the spectra for

Position	Benzene		Cyclohexane		Chloroform		DMSO		Gas Phase
	Expt	Calc	Expt	Calc	Expt	Calc	Expt	Calc	
Ortho	6.35	7.03	6.45	7.01	6.64	7.16	6.48	7.31	6.97
Meta	7.08	7.51	6.98	7.59	7.13	7.7	6.99	7.7	7.49
Para	6.72	7.05	6.59	7.04	6.77	7.2	6.55	7.2	7.04
Amine	2.75	3.68	3.29	3.64	3.57	4.31	4.99	4.31	3.42
Delta: Meta - Position									
Ortho	0.73	0.48	0.53	0.58	0.49	0.54	0.51	0.39	0.52
Para	0.36	0.46	0.39	0.55	0.36	0.5	0.44	0.5	0.45
Amine	4.33	3.83	3.69	3.95	3.56	3.39	2	3.39	4.07

Table 1: Delta values for Aniline

Position	Benzene		Cyclohexane		Chloroform		DMSO		Gas Phase
	Expt	Calc	Expt	Calc	Expt	Calc	Expt	Calc	
Ortho	5.98	6.92	6.39	6.90	6.57	7.07	6.55	7.23	6.85
Meta	6.99	7.41	6.96	7.40	7.07	7.5	7.01	7.6	7.40
Amine	2.57	3.70	3.32	3.65	3.64	3.02	5.23	4.39	3.41
Delta: Meta - Position									
Ortho	1.01	0.49	0.57	0.50	0.50	0.43	0.46	0.37	0.55
Amine	4.42	3.71	3.64	3.75	3.43	4.48	1.78	3.32	3.99

Table 2: Delta values for Chloroaniline

Position	Benzene		Cyclohexane		Chloroform		DMSO		Gas Phase
	Expt	Calc	Expt	Calc	Expt	Calc	Expt	Calc	
Aromatic	6.97	7.50	6.92	7.49	7.04	7.60	7.05	7.72	7.48
Methyl	2.13	2.49	2.22	2.39	2.29	2.41	2.25	2.43	2.50
Delta: Meta - Position									
Delta	4.84	5.01	4.70	5.10	4.75	5.19	4.80	5.29	4.98

Table 3: Delta values for Paraxylene

Position	Benzene		Cyclohexane		Chloroform		DMSO		Gas Phase
	Expt	Calc	Expt	Calc	Expt	Calc	Expt	Calc	
Ortho	6.51	7.17	6.71	7.16	6.80	7.30	6.76	7.43	5.77
Meta	7.01	7.64	7.07	7.62	7.20	7.74	7.16	7.85	7.31
Para	6.76	7.28	6.79	7.26	6.93	7.36	6.78	7.45	6.41
Hydroxyl	3.91	4.94	5.95	4.83	5.69	5.70	3.46	4.60	13.37
Delta: Meta - Position									
Ortho	0.5	0.47	0.36	0.46	0.4	0.44	0.4	0.42	1.54
Para	0.25	0.36	0.28	0.36	0.27	0.38	0.38	0.4	0.9
Hydroxyl	3.1	2.7	1.12	2.79	1.51	2.04	3.7	3.25	-6.06

Table 4: Delta values for Phenol

Phenol and the amine groups on aniline and Chloroaniline.

3.6 Overview

In comparing these NMR spectra it is clear that the computational models still require some refinement before they accurately predict the correct NMR peaks values when comparing to experiments. For almost all the shifts computed the gas phase calculation was closer to the experimental value. The only exception as phenol where the improvements were vast, in some cases a couple of hundred percent.

As I mentioned in the theory section the PCM model works by explicitly modelling the solvent. If we were to use a model where you calculate some implicit interactions between the solvent and molecule it should be possible to model hydrogen bonding more accurately. This would suggest that the peaks for the amines and phenol could be vastly improved.

4 Dyes

4.1 Experimental comparison/description

4.1.1 NMR Spectroscopy Details

The experimental data for these dyes were previously measured by Jones et Al. These were carried on a Bruker DMX 500 NMR Spectrometer. The cis isomers for Disperse Orange and Disperse RED had a half-life of around one second. For this reason it was necessary to irradiate the samples while the NMR measurements were taking place. For further details please refer to [8, 9].

The structures for the three dyes were as follows:

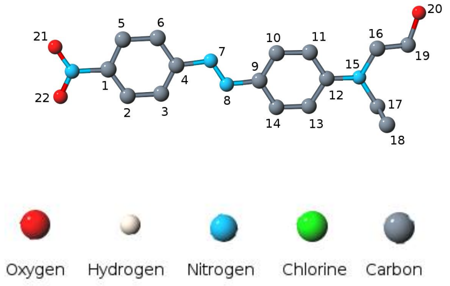

Figure 32: Disperse Red Numbering

The numbering of the protons is representative of the parent atom they are attached to. This means that each number can represent multiple protons. In Figure 32, the reference to 16/19 would represent the four protons in CH_2CH_2.

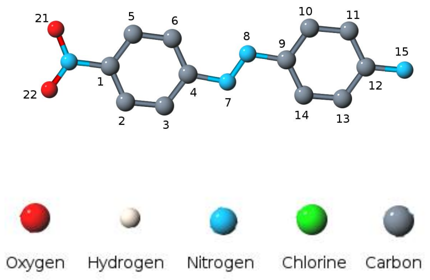

Figure 33: Disperse Orange Numbering

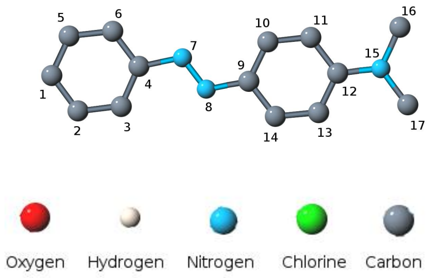

Figure 34: Methyl Yellow Numbering

4.1.2 Computation of dye molecules

The dyes were calculated in the same way as discussed in Section 2.7. The cis isomers of Disperse Orange, in gas phase and solvated had to be manually stopped because the molecule wasn't ridged enough to be fully optimised. Unfortunately the cis isomers of disperse red would not optimise, this is possibly due to a bug in the program. Disperse red solvated in cyclohexane would optimise for six iteration before it died with an arithmetic exception, on the other hand the benzene molecule would die without having completed any optimisations. Trying to start from a slightly different starting structure made no difference, for this reason only the computed data for the trans isomer are presented. For further details on the exact computational details for these molecules please refer to Appendix A.

Data for dye molecules Tables (5,6,7) show the calculated and experimental data for the dyes. The rows which are labelled with the position are the calculated values while the other rows give the experimental value. In order to be able to compare the computed data with experimental a delta (Δ) value for the difference between cis and trans isomers is given. Addition-

Position	Gas Phase				Cyclohexane				Benzene			
	Cis	Trans	Δ	$\delta\Delta$	Cis	Trans	Δ	$\delta\Delta$	Cis	Trans	Δ	$\delta\Delta$
2/5	8.73	8.87	0.14		8.83	8.96	0.13		8.87	8.97	0.10	
				-0.03	8.07	8.24	0.17	-0.04	7.62	7.89	0.27	-0.17
3/6	7.01	8.34	1.33		7.17	8.41	1.24		7.23	8.42	1.18	
				0.26	6.80	7.87	1.07	0.17	6.35	7.64	1.29	-0.11
10/14	7.32	7.71	0.39		7.44	7.71	0.27		7.62	7.73	0.10	
				-0.67	6.75	7.81	1.06	-0.79	6.77	7.96	1.19	-1.09
11/13	6.62	6.93	0.31		6.75	6.97	0.22		6.74	7.00	0.25	
				0.04	6.31	6.58	0.27	-0.05	5.82	6.12	0.30	-0.05
15	3.63	3.92	0.3		4.17	4.66	0.49		4.39	4.74	0.44	
									2.73	2.94	0.21	0.23

Table 5: NMR Data for Disperse Orange

Position	Gas Phase			Cyclohexane			Benzene		
	Cis	Trans	Δ	Cis	Trans	Δ	Cis	Trans	Δ
11/13	7.48	7.15	0.32		7.25			7.27	
				6.44	6.71	0.27	6.09	6.44	0.35
10/14	6.83	8.43	0.95		8.50			8.51	
				6.90	7.85	0.95	6.99	8.13	1.14
17	3.35	3.48	0.13		3.52			3.52	
				3.67	3.78	0.11	3.08	3.17	0.09
16/19	3.69	3.94	0.26		3.99			3.99	
				3.38	3.53	0.15	2.78	2.89	0.11
18	1.43	1.08	-0.35		1.10			1.11	
				1.12	1.22	0.1	0.66	0.75	0.09
20	0.25	0.60	0.35		1.44			1.54	
							2.78	2.89	0.11
3/6	6.98	8.37	1.39		8.44			8.45	
				6.82	7.87	1.05	6.47	7.74	1.27
2/5	8.68	8.87	0.19		8.95			8.96	
				8.10	8.22	0.12	7.71	7.94	0.23

Table 6: NMR Data for Disperse Red

Position	Gas Phase				Cyclohexane				Benzene			
	Cis	Trans	Δ	$\delta\Delta$	Cis	Trans	Δ	$\delta\Delta$	Cis	Trans	Δ	$\delta\Delta$
2/5	7.53	7.72	0.18		7.69	7.85	0.16		7.71	7.87	0.17	
				0.00	7.13	7.31	0.18	-0.02	6.98	7.24	0.26	-0.09
1	7.33	7.62	0.29		7.48	7.75	0.27		7.50	7.77	0.27	
				0.04	6.96	7.21	0.25	0.02	6.82	7.10	0.28	-0.01
3/6	7.10	8.39	1.29		7.21	8.44	1.23		7.24	8.46	1.21	
				0.20	6.72	7.81	1.09	0.14	6.89	8.17	1.28	-0.07
10/14	7.43	8.37	0.94		7.51	8.42	0.91		7.51	8.43	0.91	
				-0.06	6.83	7.83	1.05	-0.09	7.15	8.23	1.08	-0.17
11/13	6.63	6.82	0.19		6.70	6.92	0.22		6.73	6.03	0.20	
				-0.10	6.37	6.66	0.29	-0.07	6.10	6.46	0.36	-0.16
16/17	3.00	3.17	0.17		3.04	3.20	0.16		3.04	3.20	0.17	
				0.03	2.86	3.00	0.14	0.02	2.23	2.35	0.12	0.05

Table 7: NMR Data for Methyl Yellow

ally the difference between the experimental delta and the computed delta is given ($\delta\Delta$). This gives an indication on how well the data matches with the experimental measurements. A delta value is also given between gas phase calculations and experimental cyclohexane measurements.

4.2 Computed data: solvated NMR vs gas phase

The dyes followed some of the same patterns as the smaller molecules when the gas phase calculation was performed. They have been represented in the same way as the smaller molecules were.

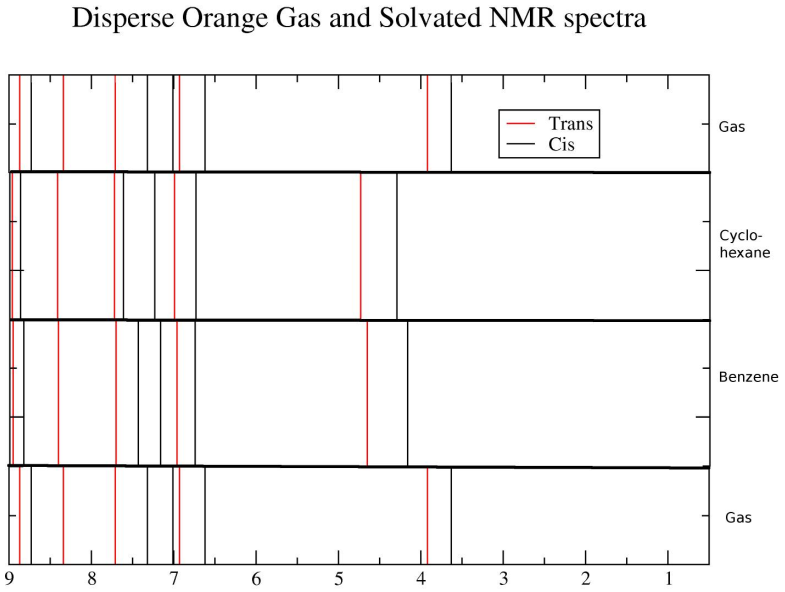

Figure 35: Disperse Orange: Comparison of computed NMR data

Figure 35 shows the shifts experienced by the disperse orange in benzene and cyclohexane. As with aniline the amine groups shift upon solvation in benzene and cyclohexane, although the shifts were notably larger. Aniline shifted by approximately $\frac{1}{4}$ppm while in disperse orange we see a shift just under 1ppm. The aromatic protons have shifted by $\frac{1}{4}$ppm, a much larger shift than seen by any of the smaller molecules.

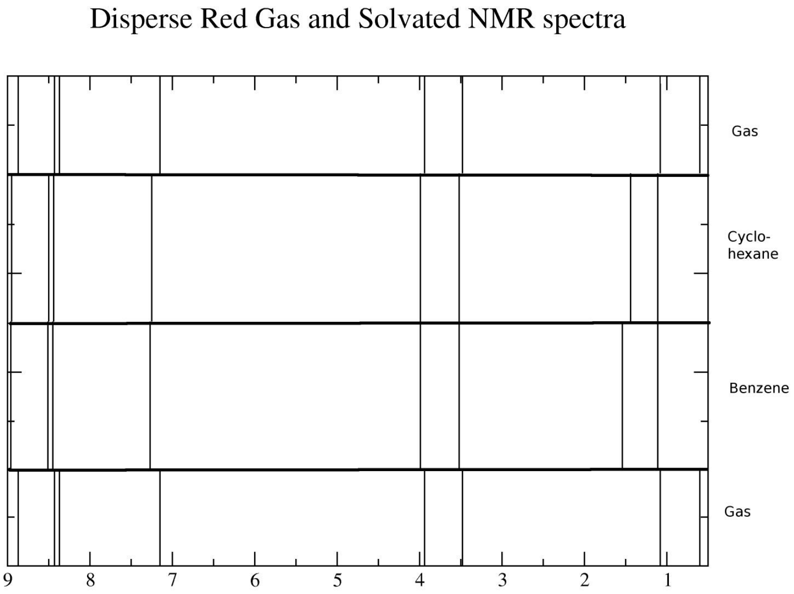

Figure 36: Disperse Red: Comparison of computed NMR data

The trans isomer in disperse red showed little sift in the aromatic protons, similar to what we saw when benzene was solvated. The largest shifts are apparent for the CH_2 and CH_3 shifts, this is obviously because there are quite strong interactions between the solvent and the disperse red dye.

Methyl yellow shows some of the features that we saw with paraxylene, the CH_3 groups showed little shift once the solvent was introduced. The aromatic protons showed a shift much larger than most of the small molecules, showing that some interaction resulting in deshielding of those protons.

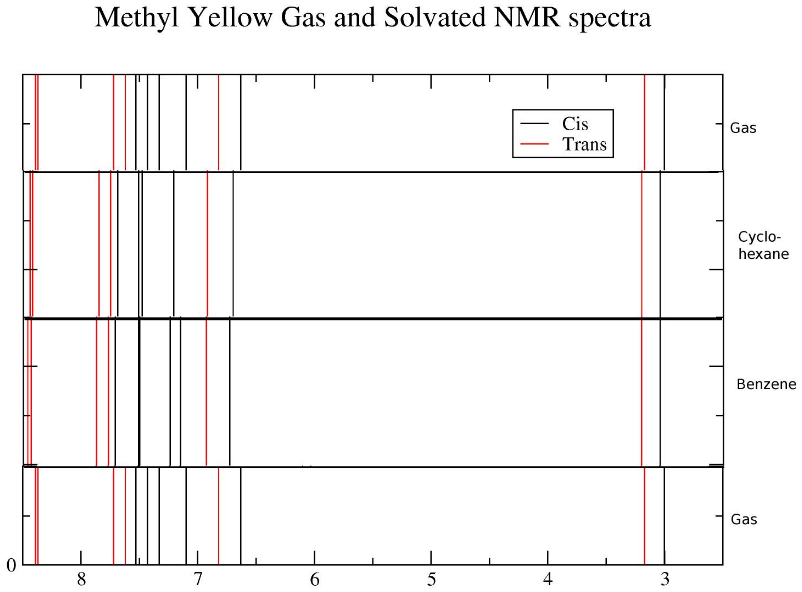

Figure 37: Methyl Yellow: Comparison of computed NMR data

4.3 Solvated NMR vs experimental

Since we were unable to obtain the precise shift for the amine and hydroxyl groups in cyclohexane i have opted to only show the spectra in the range 5.5-9ppm. This ensures a clear view of all the peaks. Where the spectra show interesting features outside this range i shall mention them.

Disperse Red It is interesting to note that Gaussian has problems calculating the positions of the peaks to the extent where peaks have been interchanged in both solvents, although benzene does show the worst simulation of the two.

In Figure 40 we can immediately see that the gradient of the lines connecting the experimental and calculated results are approximately similar.

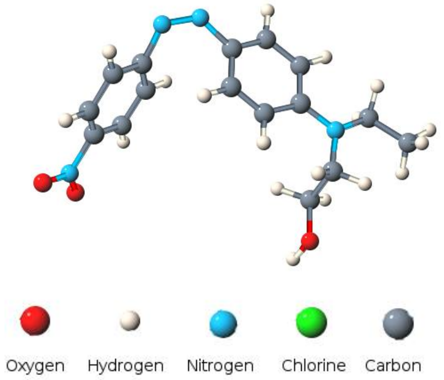

Figure 38: Optimised Cis-Disperse Red Molecule

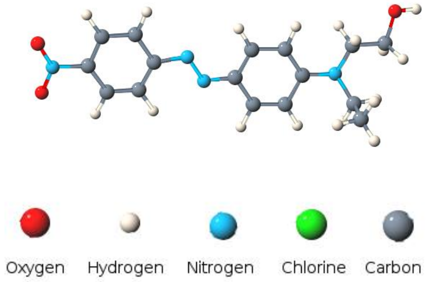

Figure 39: Optimised Trans-Disperse Red Molecule

This indicates that as with most of the aromatic protons in the smaller molecules they are all deshielded with respect to experimental. The interchanged positions show only a small error in the calculations as the peaks are so closely spaced, indicating a similar environment, due to the similarity in position and the fact that the aromatic ring will act as a buffer to the electron withdrawing effect of the NO_2.

Disperse red in benzene (Figure 41) doesn't show a similar trend as with cyclohexane, the gradients vary more and two positions are completely interchanged suggesting a large error in the calculation of the chemical shift. It clearly shows that there is an interaction which is not being taken into

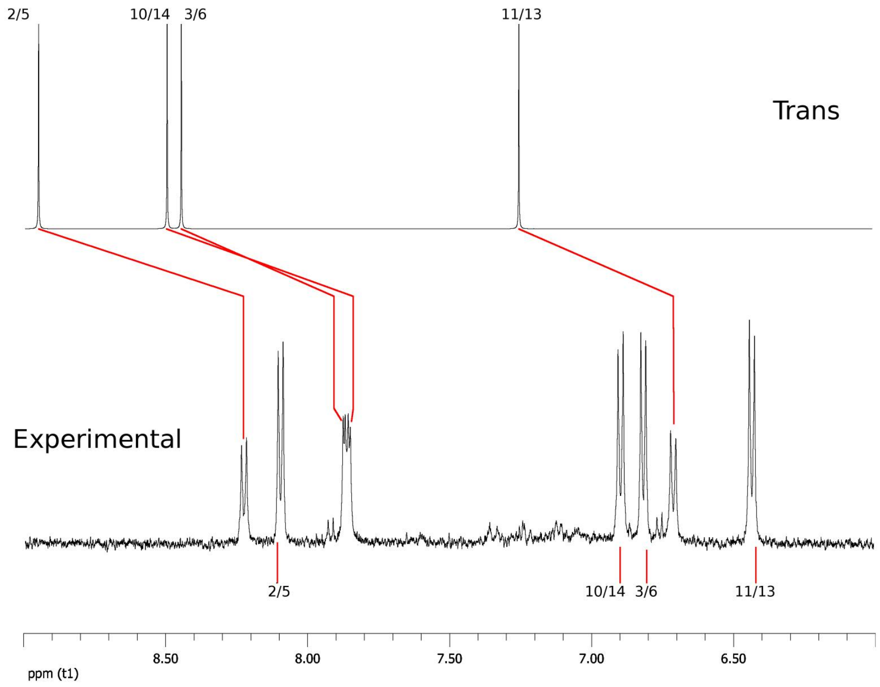

Figure 40: Disperse Red NMR Spectra in Cyclohexane

account. The 2/5 protons are much more deshielded in the calculation than any other pair. This cold indicate that the benzene manages to increase the electron density around the NO_2 portion of the molecules leading to a much lower ppm.

Disperse Orange Disperse orange shows a very constant, with the exception of the 10/14 protons , deshielding of about 0.6-0.7ppm out.

Disperse orange in cyclohexane (Figure 44) is reasonably straight forward with a similar error for most of the positions. Assuming as that disperse orange spectra like all the others is more deshielded than the experimental (which it is, except for protons 10/14) it is clear that Gaussian is overestimating the shielding. A reason for this is that there are two electron donating

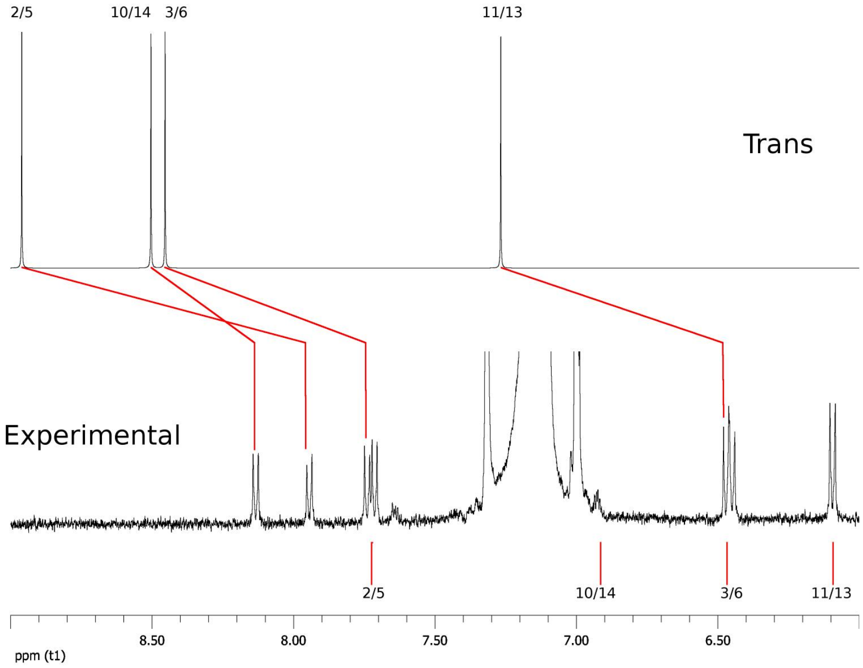

Figure 41: Disperse Red NMR Spectra in Benzene

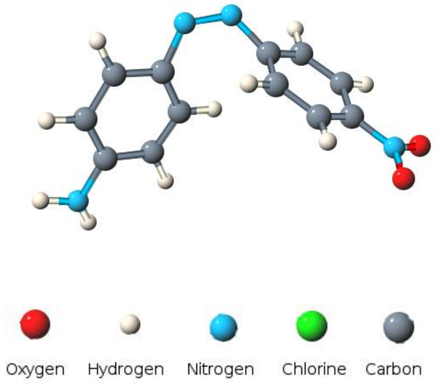

Figure 42: Optimised Cis-Disperse Orange Molecule

groups attached to the aromatic ring and that Gaussian is assuming an excess electron charge for this ring.

Benzene (Figure 45 on the other hand shows a very clear increase in the

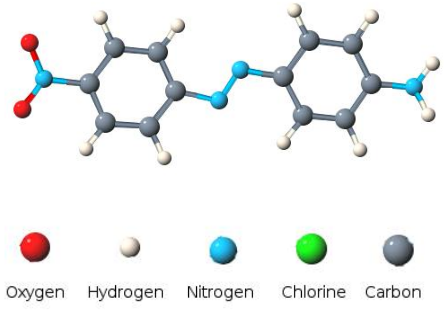

Figure 43: Optimised Trans-Disperse Orange Molecule

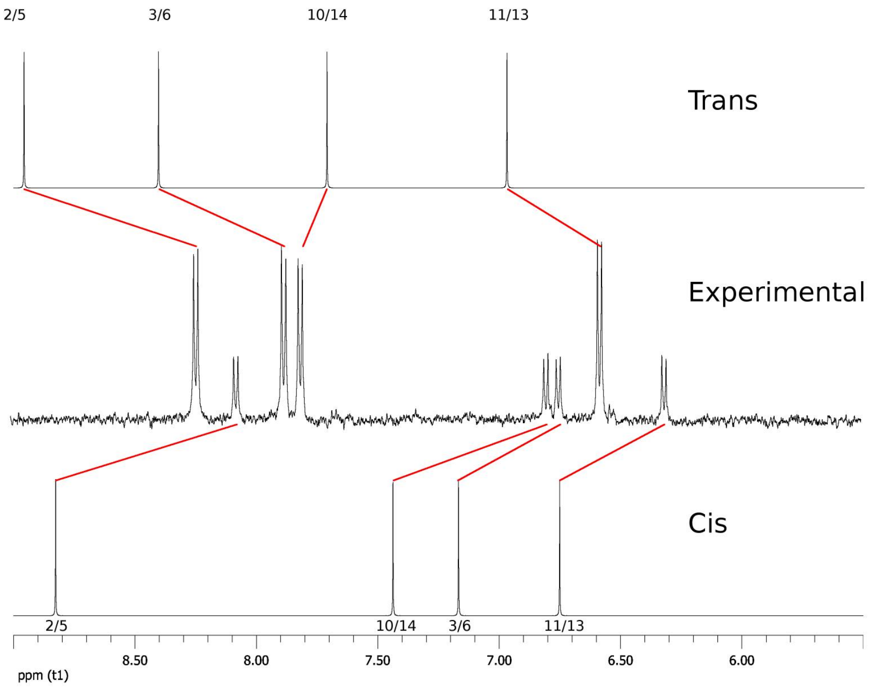

Figure 44: Disperse Orange NMR Spectra in Cyclohexane

difference between the experimental and computed shifts for protons 10/14. The cis isomers shows an almost constant shift, experimental are shielded by around 1ppm when compared to the computed.

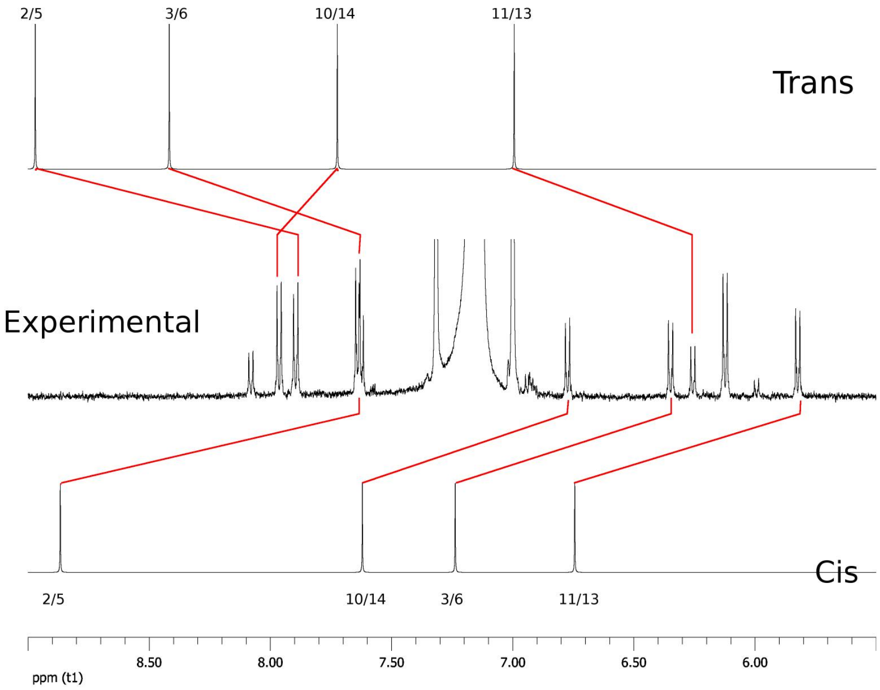

Figure 45: Disperse Orange NMR Spectra in Benzene

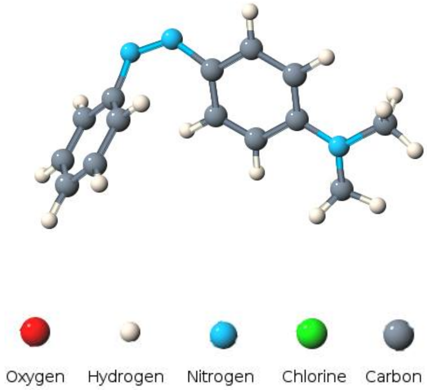

Figure 46: Optimised Cis-Methyl Yellow Molecule

Methyl Yellow Methyl Yellow showed a reasonably constant shift although for benzene there were some interchanged positions,

Methyl yellow in cyclohexane (Figure 48) showed the most constant shift

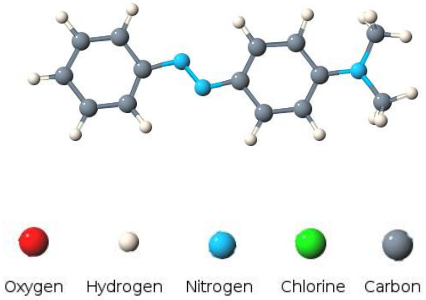

Figure 47: Optimised Trans-Methyl Yellow Molecule

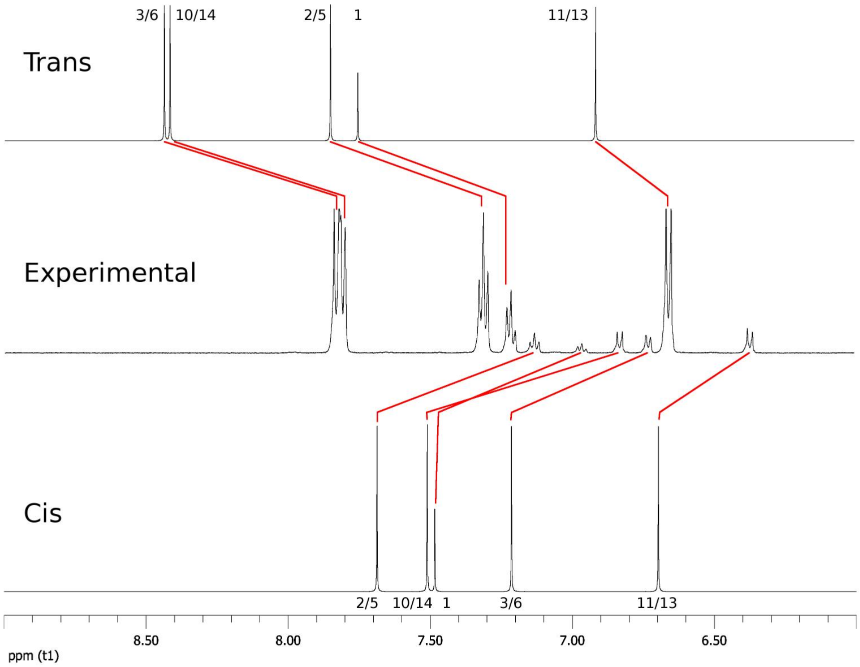

Figure 48: Methyl Yellow NMR Spectra in Cyclohexane

of all the dyes, varying between 0.28 and 0.68ppm, with the majority around 0.5ppm. This is due to the simple nature of this molecules there are no real hydrogen bonding sites, ant there are no electron withdrawing groups. For the cis isomer Gaussian had a slight problem in calculating the correct values for protons 10/14 and 1 (when compared to each other) The trans isomer is predicted correctly when the protons are observed with respect to one another.

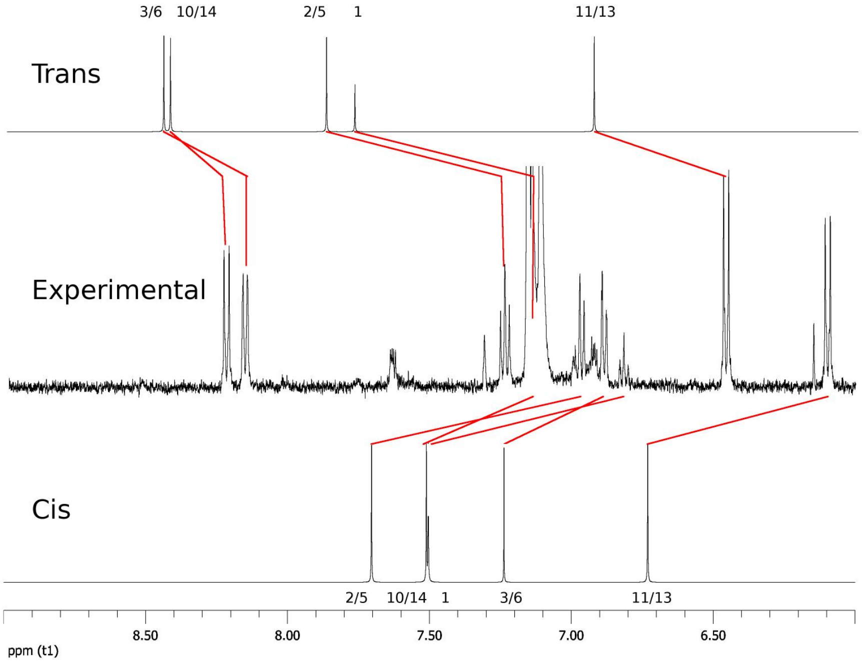

Figure 49: Methyl Yellow NMR Spectra in Benzene

Using benzene as a solvent (Figure 49) with the cis isomer proved to be a larger challenge, here the peaks aren't in the correct order. Proton 2/5 and 1 have been deshielded more than the other atoms on the molecule. This indicates that the benzene may be interacting with the methyl groups and protons 1/2/5. This possibly shows the methyl helping the charge from the methyl over to the other end of methyl yellow. The trans shows one interchanged pair, 3/6 and 10/14 but considering the proximity of the peaks it only requires a small error.

General effects on the dyes Overall we see, with the exception of protons 10/14 on disperse orange, a deshielding effect. This is consistent with other reposts on the calculations with DFT [11, 1] in that it generally underestimates the shielding provided by atoms. Looking at all the small molecules data and the dyes data it is with only few exceptions that DFT underestimates the shielding.

4.4 Accuracy of the solvated models vs gas phase

In tables (5,6,7) i have given values for what i described as $\delta\Delta$, these values should give an indication of the accuracy of the solvated model rather than the accuracy of density functional theory. Admittedly you need DFT for the solvated model, but in taking the difference we are avoiding the inaccuracy that i have mentioned in the previous section. having taken the average of all the $\delta\Delta$ values for each molecule gives a rough indication on any improvements:

	Gas-phase		Benzene	Cyclohexane
	Cyclohexane	Benzene		
Disperse Orange	-0.02	-0.16	-0.24	-0.04
Methyl Yellow	0.02	-0.05	-0.08	0.00

These values indicate that the gas phase still is more accurate than the solvated models. This gas phase provided results more accurate when compared to both the benzene data and the cyclohexane data. When the $\delta\Delta$ values for the amine group are examined this shows an improvement over the gas phase. The methyl group on methyl yellow also showed some improvement int he solvated calculations.

	Gas-phase	Benzene	Cyclohexane
Disperse Orange (N-H)	0.30	0.23	
Methyl Yellow (C-H)	0.02	0.05	0.02

5 Conclusion

PCM solvated calculations seemed to make some progress with regards to the spectra of N-H and O-H groups. Unfortunately it is still lacking in predicting the correct value of aromatic shifts, this is partially due to the limitations of DFT in that it underestimates the shielding. There have been several proposals to improve this limitation[**?**]. One such proposal is to replace the part of the generalised gradient function with a functional that is a second order derivative. This function would also be independent of eigen value and self interaction corrections. These functionals are also parameterised against experimental data which ensures accurate results. Using this method the shielding was improved from a mean error of 0.42 to around 0.35. The method that has been used throughout this report is the IEF-PCM method, this is a very simplified model of a solution, if higher accuracy was required it could be possible to use a combination of molecular mechanics (to model the solution) and ab-initio (to model the solute). This ensures that your computational requirements for the solvent remain low, and yet you gain the accuracy of the direct solvation effects (e.g. hydrogen bonding). This method works by taking a number of specified solvent models and calculating the interaction directly. It should also be noted that the effects of the solvation of the dyes wasn't tested to the extremes. Benzene and cyclohexane have a moderately low dielectric effect, thus if a solvent with a higher dielectric value had been used the solvated spectra would have been in closer agreement with experiment than gas phase. If the intention is only to confirm the experimental results it is possible to take the difference between two isomers or molecules and use this information to confirm the identity. This method showed very good agreement with the experimental results, this is in part due to the cancellation of the over estimating terms in DFT, and a general cancellation of any errors in the calculation and experiment. It is expected that the computational power will increase in the coming years and with it will come the ability to run ever more complicated models.

A Benchmarking

A.1 Introduction

The Gaussian 03 program allows you to model chemical species using various computational methods. Version 03 is a complete rewrite, thus adding many new features and also new parallelised code. To ensure that we were not wasting valuable computational time and money. During this benchmarking I am using one of the molecules for my project, namely Trans Dispersed Orange, this was put though various ab-initio optimisations, MP2, DFT and HF. Initially a large basis set was used, but this was found to be too large for benchmarking purposes and thus we resorted to a lower basis set.

A.2 Molecule specifications

The molecular formula for Dispersed Orange is $C_{12}H_{10}N_4O_2$ (Figure: 50), this implies that there are a total of 28 atoms that have to be taken into account when doing an ab-initio calculation.

Calculations were performed with two different basis sets: 6-31G(d,p)(320 basis fucntions) and the larger 6-311+G(2d,2p) (576 basis functions). The larger basis set was only calculated throughly for Density Functional theory (DFT). Hartree Fock (HF) scales approximately the same as DFT but it was decided that the lower accuracy provided meant that it wasn't worth concentrating on it. MP2 calculations do provide an alternative indication to DFT but are time and spaceconsuming. Additionally we were unable to do a 1 processor calculation because it would exceed the set limits for lomond.

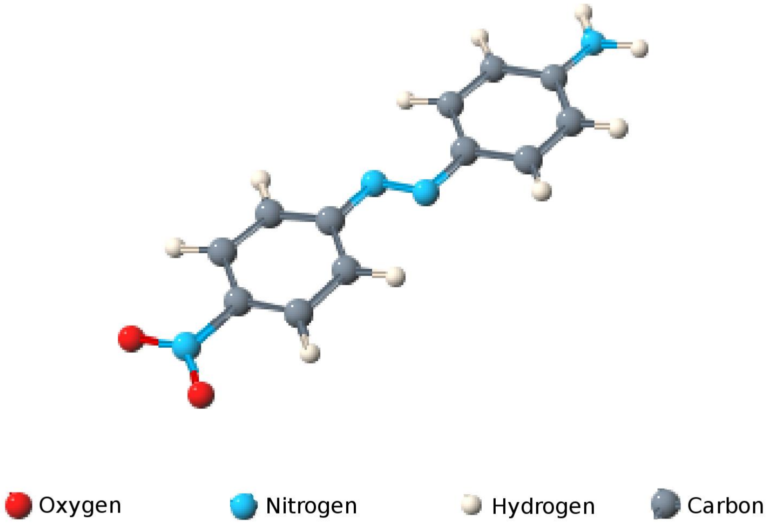

Figure 50: Dispersed Orange

The molecule was chosen to be in a partially optimised form, this was done to prevent a repetitive calculation and yet provide a reasonable accuracy. In this case it took a total of 6 iterations to optimise the geometry of the molecule.

Default parameters:

- Storage Space, All the jobs were run on Lomond's back-end on the physics scratch space.

  Gaussian was set up with the `#MaxDisk=10GB` parameter. This instructs Gaussian not to use more than 10GB of space. This is not very important with DFT or HF calculations since these typically only used up to 10GB of space, but it was more important for the MP2 calculations. This is due to Gaussian changing it defaults from Gaussian

98 (G98) to Gaussian 03 (G03). In G03 it is assumed that disk space is abundant and will use as much as it requires. Without this setting the files will increase in size until it runs out of storage space and then cause a segmentation fault. The larger the system you are trying to describe is the larger your MaxDisk value should be.

- Memory usage The jobs were set to use a default memory of 1024MB using `%mem=1024MB`. This instructs Gaussian that it can use 1024MB per process.
- Processors The number of processors (N) was set to the number that we wished to benchmark. This was done using the `%nproc=N` keyword.
- Checkpoint file We also generated a default checkpoint file, using the `%chk=filename` setting.
- Default Route Keyword The default route was specified according to the benchmark that was being carried out. For DFT this would be the b3lyp, HF would just be HF, and similarly for MP2 we used MP2.

To do the benchmarking, the jobs were run on lomond with a varying number of processors, ranging from 1 to 32, doubling the number for each step increase in processing power. It was necessary to make sure that the one processor job would be able to run within the time constraints of the queue system on lomond. Once all the data had been collected the speed up and efficiency for each benchmark was calculated.

Speed Up: $S = T_1/T_N$

S=Speed Up, T_1=Time for job on 1 Processor, T_N=Time for Job on N processors.

Efficiency: $E = S/N = T_1/(NT_N)$

E=Efficiency, S=Speed Up, T_1=Time on 1 processor, T_N=Time on N Processors.

This is then plotted to give an indications of speed and efficiency.

A.3 6-31g(d,p) Basis Set

A.3.1 Optimisation Benchmark Results for Smaller Basis Set

The default job setting was a basis set of 6-31g(d,p), which was set to optimise on 1,2,4,8,16,32 processors. Of importance for all the molecules:

- Set the maximum memory using the `%mem=` keyword

DFT Optimisation The DFT jobs all completed within 2 $\frac{1}{2}$ hours, which made them relatively quick in comparison to MP2. The efficiency decreases rapidly after 8 processors showing an efficiency greater than 0.5 for 16 but drops off quickly after this. This can also be seen by looking at the Figure: 51 where it clearly can be seen that the speedup levels off around 16 processors.

- DFT should be run with at most 8 processors, if it is required to run very quickly or for very large molecules where memory can be beneficial 16 could be considered as an option.
- You may want to set the `#MaxDisk=` keyword if you are short on space, but in general this will not pose a problem

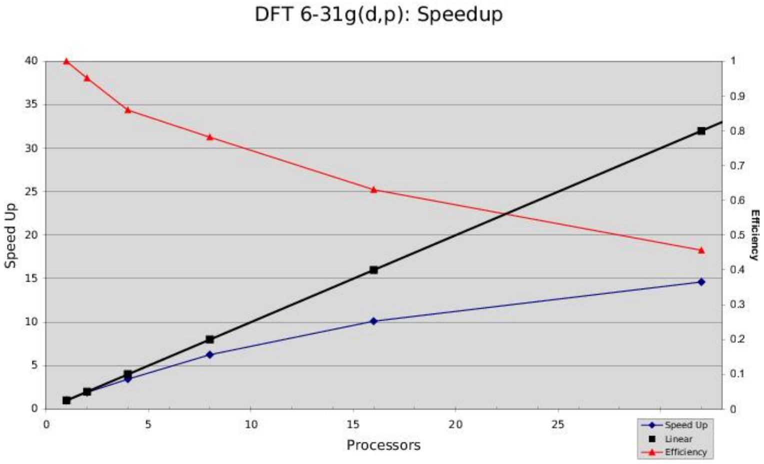

Figure 51: Plot of Speedup and Efficiency for DFT calculations

MP2 Optimisation The jobs submitted with MP2 optimisation had a far lower efficiency and ran for a lot longer, a single processor job for MP2 took close to 18hours, compare this to DFT taking only 2 1/2, it should be considered whether the gain from doing a MP2 calculation is worth the additionally computational time. The efficiency rapidly dropped after 4 processors and as can be seen form Figure: 52 the speed up is marginal after 8 processors. The reason it drops off so quickly probably has something to do with the way that the calculation is performed. You could possible get a slight gain in computational speed by increasing the scratch file size, this is done by increasing the MaxDisk parameter. During the whole of this calculation G03 was using almost all of the permitted 10GB. It should be noted that when the MaxDisk parameter was unspecified G03 crashed leaving a file close to 100GB in size.

- Use at most 8 processors, 4 is preferable.
- You should set the `#MaxDisk=nnGB` parameter to a reasonable value, for large systems try a value of 50GB as space is not a problem.

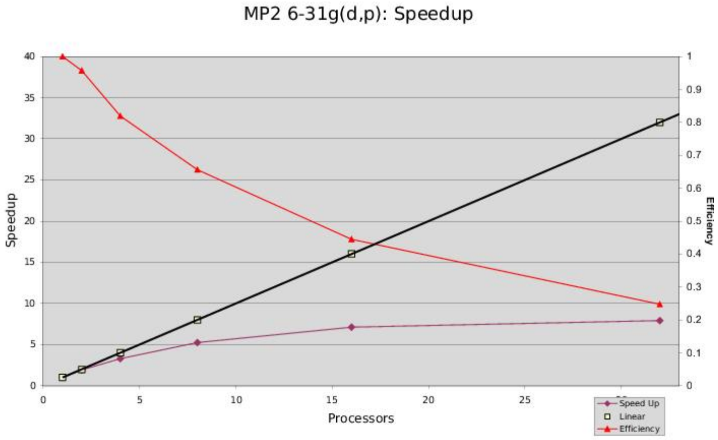

Figure 52: Plot of Speedup and Efficiency for MP2 Calculations

HF Optimisation The Hartree Fock calculations were some of the quickest to complete, they took under 2 hours to run with roughly the same scaling as with DFT. From Figure: 53 we can clearly see it levelling off after 8 processors with only a slight gain after this point. For eight processors it has an deficiency of close to 80%.

- Similarly as with DFT you should preferably run with no more than 8 processors unless time or additional memory are of importance.
- For HF the `MaxDisk` keyword doesn't have much effect since the calculations do not use much disk space

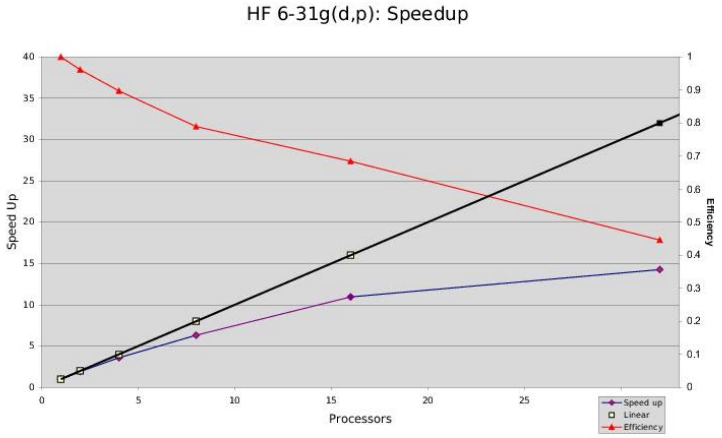

Figure 53: Plot of Speedup and Efficiency for HF calculations

A.3.2 Solvated Optimisation Benchmarks

The route for the solvated benchmarks looked like this:

```
#b3lyp/6-31g(d,p) OPT SCRF(PCM,Solvent=Cyclohexane) SCF=tight
```

the OPT parameter was changed to NMR for the NMR calculation.

Optimisation As can be seen from Figure: 54, this method shows a drop in speed up almost immediately, after 4 processors the curve is almost completely flat. With an efficiency of around 50% for 4 processors which continues dropping and giving about 11% fro the 32 processor run.

- Using more than 4 processors will give no additional benefit in terms of speed.
- In the same way as for normal DFT you may want to specify the `MaxDisk` keyword

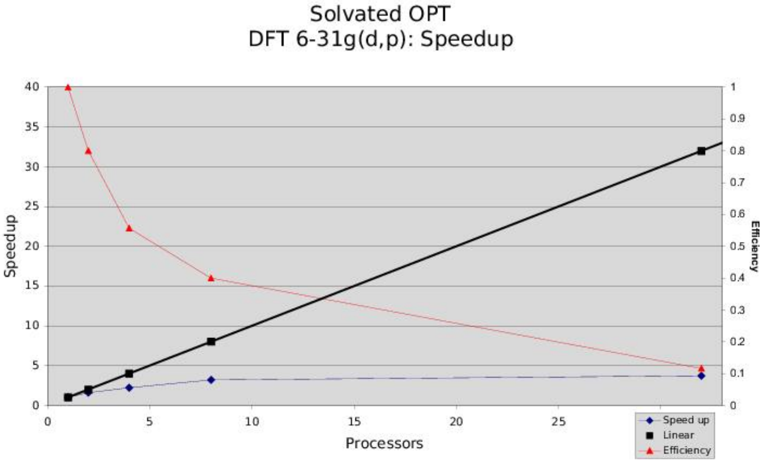

Figure 54: Plot of a Speedup and Efficiency for a solvated DFT optimisation

NMR Calculation Likewise for the NMR calculation it showed an immediate drop in the speedup, Figure: 55 is flat after 8 processors, with almost no gain by going to 32. The efficiency is below 10% for the 32 processor job.

- As for the “Solvated opt” there is almost no additional gain from using more than 4 processors.

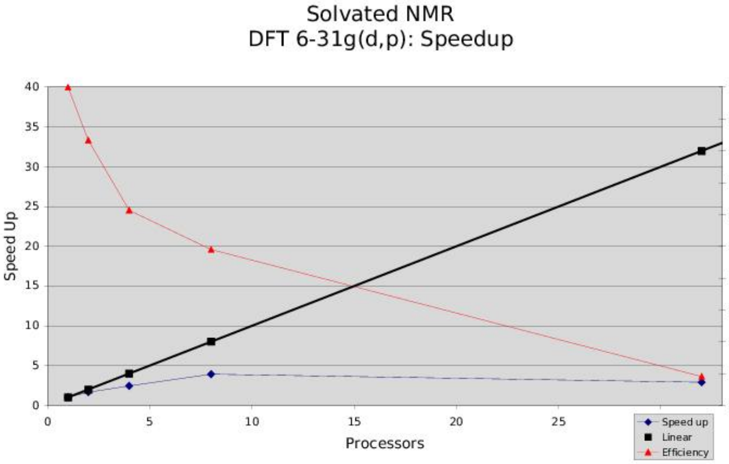

Figure 55: Plot of speedup and Efficiency for an NMR calculation

A.3.3 Frequency Calculation

The molecule was first optimised using the normal DFT optimisation, the checkpoint file was then used to do the frequency calculation, by adding the freq command instead of opt to the route line. The speedup flattened out after approximately 16 processors. This can be clearly seen when looking at the efficiency curve in Figure 56, where it drops quite rapidly after eight processors. The graph shows less than 50% efficiency at 32 processors.

- You should preferably use 8 processors but if the system is very large 16 processors could be considered.
- Comparing these results with those of the larger basis set calculations of DFT it may be possible to run these calculations with more processors depending on how well this calculation scales, more benchmarking would be required to determine this.

A.4 6-311++g(2d,2p) Basis Set

The large basis set was a 6-311+G(2d,2p), this is a very large system and thus the calculations take a long time to complete. A complete set was run for DFT while only 3 data points were done for HF, this was mainly done for a comparison with DFT.

A.4.1 Optimisation Benchmark with DFT 6-311+G(2d,2p)

For DFT we did all the points as we did with the smaller basis set. These results were very good, it showed that even with 32 processors we still had

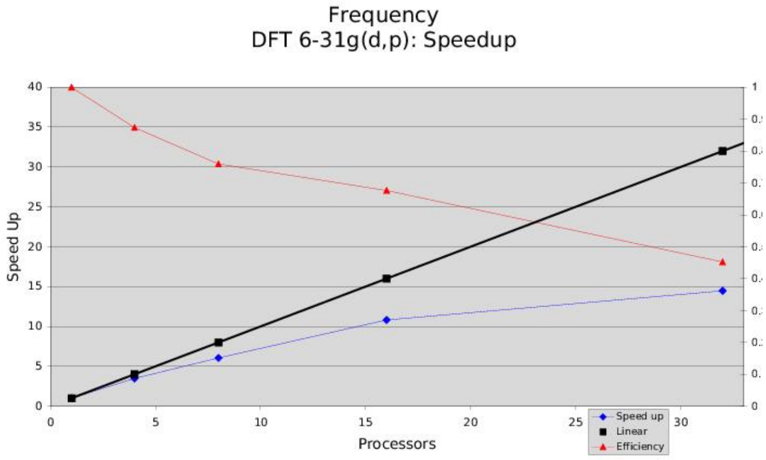

Figure 56: Plot of Speedup and Efficiency for a Frequency Calculation

an efficiency of over 80%. The speedup curve in Figure 57 seems to show that there is an almost linear scaling with the processors.

- For the larger basis sets efficiency seemed to be higher. For this reason you can run up to at least 24 processors and 32 is also possible if the calculations are required to end quicker.
- You should set the `MaxDisk` keyword for these calculations as the data files can get quite large.
- It is usually quicker to optimise in multiple steps with increasing size of the basis set. i.e. start with a 6-31(d,p), optimise and then optimise in the 6-311g+(d,p)

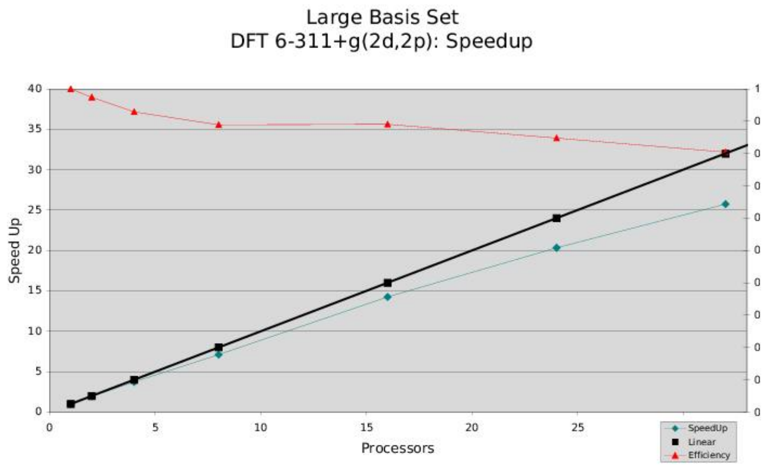

Figure 57: Plot of Speedup and Efficiency for a Large basis set DFT calculation

A.4.2 Optimisation Benchmark with HF 6-311+G(2d,2p)

HF didn't scale as well as the larger DFT did, the efficiency is still over 75% at 32 processors. The speedup curve in Figure 58 doesn't show as close to linear dependence as with DFT.

- HF calculations didn't scale as well as DFT, the calculations should preferably be done at 16, and possibly 24 processors.

A.4.3 Large Basis set MP2 Optimisations

By looking at the data it seems likely that you will see an increased speedup for MP2 with a larger basis set. To be sure of this benchmarks would have to be taken in the larger system.

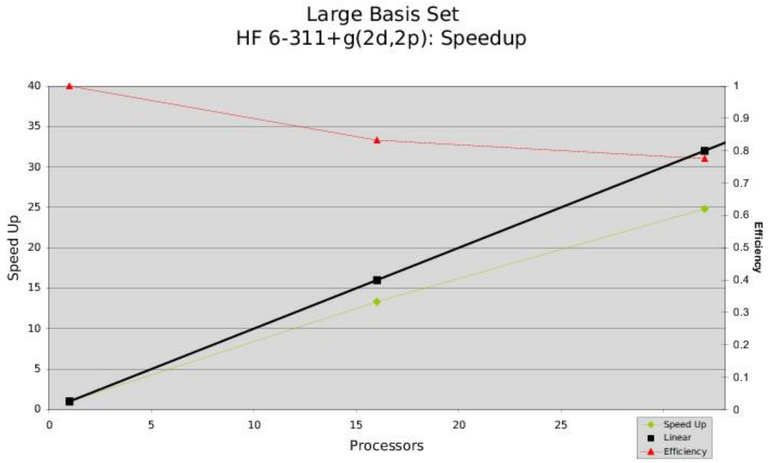

Figure 58: Plot of Speedup and Efficiency for a Large basis set HF calculation

A.4.4 6-311+G(2d,2p) Solvated Optimisations

Due to the large size of these basis sets we were unable to perform a benchmark for a one processor job due to the restrictions of lomond. Additionally there was not sufficient time for us to perform these benchmarks. These limitations meant that only the four and eight processor benchmarks were performed. This allows for a relative judgement on how well the calculations scale. Form earlier discussion we saw that scaling from a smaller basis set to a larger one was very efficient and often led to a large increase in efficiency. From the small basis set analysis it was understood that there was no gain in using more than 4 processors. By doubling the number of processors there was a gain of 1.43 and 1.59 for the optimisation and NMR calculation respectively. In contrast for the larger basis set these were 1.77 and 1.83 respectively. This scaling is roughly similar to the one we saw in the Hartree-Fock benchmarks for four to eight processors. This allows us to say that you use up to at least 8 processors for the solvated calculations.

- Due to increased efficiency it is possible to use up to 8 processors for Solvated calculations in large basis sets.

B Data

The following is the data for both experimental (Expt) and Calculated (Calc) data. All units are in ppm. The positions on the molecule are indicated in relation to either the amine or hydroxyl group.

NMR Data for Aniline (ppm)

	Benzene		Cyclohexane		Chloroform		DMSO		Water	Gas Phase
	Expt	Calc	Expt	Calc	Expt	Calc	Expt	Calc	Calc	Calc
Ortho	6.35	7.03	6.45	7.01	6.64	7.16	6.48	7.31	7.32	6.97
Meta	7.08	7.51	6.98	7.49	7.13	7.59	6.99	7.70	7.71	7.49
Para	6.72	7.05	6.59	7.04	6.77	7.11	6.55	7.20	7.20	7.04
Amine	2.75	3.68	3.29	3.64	3.57	3.98	4.99	4.31	4.33	3.42

Table 8: NMR Data for Aniline

NMR Data for Benzene (ppm)

	Benzene		Cyclohexane		Chloroform		DMSO		Water	Gas Phase
	Expt	Calc	Expt	Calc	Expt	Calc	Expt	Calc	Calc	Calc
Aromatic	7.15	7.84	7.21	7.81	7.33	7.93	7.37	8.05	8.05	7.69

Table 9: NMR Data for Benzene

NMR Data for Chrloroaniline (ppm)

	Benzene		Cyclohexane		Chloroform		DMSO		Water	Gas Phase
	Expt	Calc	Expt	Calc	Expt	Calc	Expt	Calc	Calc	Calc
Ortho	5.98	6.92	6.39	6.90	6.57	7.07	6.55	7.23	7.24	6.85
Meta	6.99	7.41	6.96	7.40	7.07	7.50	7.01	7.60	7.61	7.40
Para	Cl	Cl	Cl	Cl	Cl	Cl	Cl	Cl	Cl	Cl
Amine	2.57	3.70	3.32	3.65	3.64	3.02	5.23	4.39	4.41	3.41

Table 10: NMR Data for Chloroaniline

NMR Data for Phenol (ppm)

	Benzene		Cyclohexane		Chloroform		DMSO		Water	Gas Phase
	Expt	Calc	Expt	Calc	Expt	Calc	Expt	Calc	Calc	Calc
Ortho	6.51	7.17	6.71	7.16	6.80	7.30	6.76	7.43	7.42	5.77
Meta	7.01	7.64	7.07	7.62	7.20	7.74	7.16	7.85	7.83	7.31
Para	6.76	7.28	6.79	7.26	6.93	7.36	6.78	7.45	7.43	6.41
Hydroxyl	3.91	4.94	5.95	4.83	5.69	5.70	3.46	4.60	6.16	13.37

Table 11: NMR Data for Phenol

NMR Data for Para-xylene (ppm)

	Benzene		Cyclohexane		Chloroform		DMSO		Water	Gas Phase
	Expt	Calc	Expt	Calc	Expt	Calc	Expt	Calc	Calc	Calc
Aromatic	6.97	7.50	6.92	7.49	7.04	7.60	7.05	7.72	7.73	7.48
Methyl	2.13	2.39	2.22	2.39	2.29	2.41	2.25	2.43	2.43	2.50

Table 12: NMR Data for Para-xylene

C Optimised stuctures

The data provied is here is in the XYZ format, this can be imported directly into most computational packages.

C.1 Gas Phase

Gas: Dispersed orange Cis

C	0.682383	1.494037	0.022311
C	1.632933	1.542425	-1.003822
C	2.723766	0.691056	-0.983639
C	2.882061	-0.178921	0.090234
C	1.972321	-0.20905	1.142526
C	0.87001	0.625724	1.106379
N	-0.33095	2.49083	0.047984
N	-1.555684	2.281376	0.140655
C	-2.176368	1.011629	0.017226
C	-3.416331	0.902573	0.660009
C	-4.15965	-0.258412	0.588231
C	-3.713805	-1.338261	-0.190194
C	-2.500831	-1.205184	-0.888172
C	-1.740128	-0.057932	-0.779473
H	1.503652	2.250419	-1.809881
H	3.456863	0.697578	-1.774588
H	2.136308	-0.884401	1.967292
H	0.154855	0.616568	1.915624
H	-3.775511	1.75161	1.224491
H	-5.100877	-0.331678	1.117123
H	-2.15995	-2.01459	-1.520774
H	-0.823375	0.014744	-1.342299
N	-4.432902	-2.516105	-0.246356
N	4.040378	-1.07681	0.120214
H	-5.403467	-2.47644	0.014455
H	-4.2442	-3.135438	-1.015979
O	4.824329	-1.031983	-0.823266
O	4.159625	-1.825706	1.085438

Gas: Dispersed orange trans

C	-1.027045	-0.304023	-8.64E-4
C	-1.592406	0.980422	-0.003836
C	-2.965887	1.130671	-0.002984
C	-3.774092	-0.00458	7.97E-4
C	-3.237011	-1.285567	0.003496
C	-1.859319	-1.428242	0.002397
N	0.361509	-0.574006	-0.001909
N	1.096074	0.447584	0.001002
C	2.475161	0.214831	-0.001395
C	3.081051	-1.054131	-0.006496
C	4.452469	-1.169259	-0.007899
C	5.274329	-0.023483	-0.002503
C	4.667713	1.242119	0.001533
C	3.291151	1.352859	0.002421
H	-0.943031	1.841608	-0.006984
H	-3.424931	2.106588	-0.005365
H	-3.894863	-2.139777	0.006187
H	-1.403636	-2.407743	0.004073
H	2.45552	-1.933798	-0.008596
H	4.912588	-2.149228	-0.016191
H	5.284548	2.131136	9.84E-4
H	2.816505	2.324042	0.007285
N	6.646849	-0.150214	-0.052147
N	-5.234594	0.15784	0.001777
H	7.04272	-1.034448	0.216964
H	7.198117	0.646316	0.21743
O	-5.683076	1.299835	-7.96E-4
O	-5.92306	-0.857688	0.005189

Gas: Dispersed red cis

C	0.300339	-2.284226	-0.067142
C	0.292298	-1.064848	-0.763152
C	1.418681	-0.272436	-0.818215
C	2.62411	-0.639609	-0.167197
C	2.621	-1.876661	0.520703
C	1.500026	-2.682122	0.532448
N	3.753191	0.151443	-0.229165
N	-0.742547	-3.23665	-0.001079
N	-1.959604	-2.977016	-0.094914
C	-2.537232	-1.681455	-0.052476
C	-3.416061	-1.307655	-1.076381
C	-4.117373	-0.117723	-0.993546
C	-3.968179	0.67866	0.137647
C	-3.136192	0.30056	1.187499
C	-2.419016	-0.877516	1.090662
N	-4.710165	1.937673	0.231443
O	-4.556551	2.620065	1.24115
O	-5.441408	2.244207	-0.705886
C	4.952374	-0.275438	0.50877
C	6.221726	0.514247	0.213859
C	3.670646	1.51126	-0.747707
C	3.089258	2.512738	0.24968
O	3.076047	3.778397	-0.404836
H	-0.588201	-0.741029	-1.294453
H	1.360987	0.643101	-1.383719
H	3.499844	-2.22155	1.038283
H	1.527045	-3.641099	1.031115
H	-3.536834	-1.957906	-1.930908
H	-4.782559	0.198195	-1.781511
H	-3.060141	0.932775	2.057992
H	-1.770341	-1.18718	1.89688
H	4.756245	-0.252527	1.587409
H	5.144375	-1.316729	0.249681
H	7.043488	0.048923	0.757825
H	6.16367	1.55187	0.539397
H	6.474598	0.495433	-0.845923
H	4.661842	1.844243	-1.036407
H	3.078618	1.528874	-1.66012
H	2.081373	2.209363	0.544822
H	3.711691	2.543822	1.150444
H	2.645042	4.42525	0.159063

Gas: Dispersed red trans

C	2.618534	-1.52124	-0.413563
C	3.325136	-0.318856	-0.173626
C	1.23988	-1.550778	-0.411066
H	3.145451	-2.439981	-0.609655
C	2.545823	0.835838	0.103986
C	0.483895	-0.400153	-0.158329
N	4.706826	-0.267717	-0.222097
H	0.716098	-2.476703	-0.604481
C	1.169703	0.796024	0.104738
N	-0.902998	-0.55167	-0.176143
C	5.483377	-1.507017	-0.115534
C	5.384752	0.983761	0.117777
H	3.023418	1.77281	0.336552
N	-1.578816	0.487994	0.048432
C	6.821887	1.080217	-0.383798
C	5.58075	-2.066847	1.306454
H	0.603126	1.689068	0.321463
H	5.056225	-2.24799	-0.786357
H	6.47926	-1.313803	-0.50498
H	4.855822	1.805297	-0.357944
H	5.372339	1.174839	1.197251
C	-2.978838	0.297239	0.024562
O	7.196523	2.443064	-0.195025
H	7.492697	0.424673	0.17711
H	6.867324	0.803985	-1.442191
H	6.171087	-2.98404	1.310737
H	4.594547	-2.295933	1.70824
H	6.061884	-1.355929	1.97878
C	-3.748608	1.436574	0.284517
C	-3.615681	-0.926471	-0.235775
H	8.129866	2.544338	-0.396196
C	-5.131725	1.367894	0.289152
C	-4.995054	-1.002533	-0.233489
H	-3.239556	2.368855	0.481637
H	-3.015137	-1.799785	-0.435798
C	-5.739719	0.145913	0.029645
H	-5.740592	2.235207	0.488003
H	-5.507561	-1.930834	-0.430525
N	-7.205542	0.06211	0.032359
O	-7.717353	-1.028763	-0.200993
O	-7.837577	1.087387	0.268091

Gas: Methyl yellow cis

N	-1.016228	2.241864	-0.318491
N	-2.211261	1.886886	-0.346723
C	0.098714	1.364989	-0.20377
C	-2.673492	0.561358	-0.064277
C	0.208323	0.080624	-0.752383
C	1.241338	1.931239	0.369716
C	-3.564057	-0.024521	-0.965622
C	-2.422982	-0.057445	1.163881
C	1.398327	-0.618915	-0.687884
C	2.423411	1.224882	0.479672
H	-0.6296	-0.368817	-1.261508
H	1.180725	2.947659	0.733704
C	-4.148683	-1.249389	-0.668756
C	-3.035682	-1.266729	1.465118
C	2.534327	-0.084745	-0.037316
H	-3.78823	0.48862	-1.890692
H	-1.761826	0.414259	1.87664
H	1.444601	-1.591015	-1.15175
H	3.26953	1.703092	0.945926
C	-3.888822	-1.875317	0.547562
N	3.703029	-0.808622	0.077931
H	-4.822595	-1.706792	-1.3805
H	-2.846046	-1.73617	2.421096
C	4.906228	-0.158841	0.569608
C	3.853666	-2.058641	-0.647995
H	-4.359463	-2.818983	0.785815
H	5.70035	-0.895849	0.645774
H	4.74654	0.257266	1.564844
H	5.248167	0.648948	-0.087799
H	4.812441	-2.501296	-0.394085
H	3.815527	-1.92212	-1.734946
H	3.077402	-2.770998	-0.366852

Gas: Methyl Yellow trans

C	-5.290261	1.166213	3.68E-4
C	-5.847355	-0.108417	2.26E-4
C	-5.014892	-1.229169	-1.6E-5
C	-3.636824	-1.083234	-1.07E-4
C	-3.073787	0.200385	4.1E-5
C	-3.909217	1.32007	2.66E-4
N	-1.683114	0.474166	-3.4E-5
N	-0.939685	-0.538512	-2.4E-5
C	0.440961	-0.287327	-2.42E-4
C	1.27571	-1.40842	-1.68E-4
C	2.652141	-1.28357	-3.19E-4
C	3.260092	-0.008445	-6.76E-4
C	2.405202	1.123363	-5.74E-4
C	1.034497	0.98443	-4.12E-4
N	4.626926	0.133127	-0.001182
C	5.230143	1.45435	5.98E-4
C	5.481979	-1.040397	0.00117
H	-5.929927	2.038203	5.53E-4
H	-6.921746	-0.231953	2.96E-4
H	-5.448466	-2.220395	-1.37E-4
H	-2.983737	-1.942516	-2.93E-4
H	-3.45395	2.300653	3.66E-4
H	0.817715	-2.388016	6.7E-5
H	3.255848	-2.176775	-1.74E-4
H	2.824285	2.117173	-5.84E-4
H	0.399425	1.857624	-3.45E-4
H	6.31088	1.351724	-7.7E-4
H	4.947881	2.029523	0.886848
H	4.946096	2.032289	-0.883193
H	6.520656	-0.724948	1.65E-4
H	5.317665	-1.66327	-0.882528
H	5.318143	-1.659785	0.887478

C.2 Cyclohexane

Cyclohexane: Disperse orange cis

C	0.686328	1.487778	0.017046
C	1.618276	1.515299	-1.027609
C	2.713686	0.670706	-1.005461
C	2.894941	-0.173549	0.086719
C	2.002202	-0.18291	1.155727
C	0.896499	0.646207	1.119104
N	-0.330336	2.478477	0.042305
N	-1.557357	2.267636	0.138178
C	-2.182707	1.005715	0.015305
C	-3.448327	0.923759	0.615509
C	-4.199405	-0.231183	0.549894
C	-3.733853	-1.339346	-0.179729
C	-2.492774	-1.235138	-0.837708
C	-1.727367	-0.092269	-0.734859
H	1.470945	2.200587	-1.851606
H	3.430053	0.665254	-1.812908
H	2.178694	-0.836599	1.996457
H	0.196883	0.654169	1.94335
H	-3.820337	1.791048	1.145262
H	-5.161572	-0.281827	1.044405
H	-2.13818	-2.06557	-1.436112
H	-0.79336	-0.044163	-1.271144
N	-4.457914	-2.506593	-0.230613
N	4.053451	-1.062471	0.117737
H	-5.43458	-2.467259	0.015629
H	-4.237836	-3.167687	-0.958991
O	4.82577	-1.040888	-0.838684
O	4.197904	-1.791325	1.097283

Cyclohexane: Dispersed orange trans

C	-1.040012	-0.264764	0.0
C	-0.851295	-1.656844	0.0
C	-1.942507	-2.50337	0.0
C	-3.227094	-1.959723	0.0
C	-3.438677	-0.585294	0.0
C	-2.339823	0.256378	0.0
N	0.0	0.690261	0.0
N	1.163689	0.20084	0.0
C	2.219794	1.107499	0.0
C	2.087529	2.510487	0.0
C	3.201233	3.315435	0.0
C	4.500781	2.756008	0.0
C	4.631415	1.354117	0.0
C	3.509338	0.553184	0.0
H	0.151685	-2.055529	0.0
H	-1.816807	-3.575467	0.0
H	-4.444813	-0.19511	0.0
H	-2.467497	1.330308	0.0
H	1.099076	2.946347	0.0
H	3.091794	4.393375	0.0
H	5.619122	0.910275	0.0
H	3.604458	-0.524871	0.0
N	5.598887	3.562087	0.0
N	-4.381275	-2.858616	0.0
H	5.508421	4.563084	0.0
H	6.526723	3.175508	0.0
O	-4.171429	-4.069753	0.0
O	-5.504241	-2.358822	0.0

Cyclohexane: Disperse red trans

C	2.615114	-1.520397	-0.389291
C	3.324371	-0.315414	-0.157518
C	1.237518	-1.548209	-0.384849
H	3.139908	-2.442044	-0.579578
C	2.54439	0.842327	0.118153
C	0.480574	-0.394409	-0.138149
N	4.700605	-0.263852	-0.209274
H	0.713986	-2.476804	-0.572573
C	1.169568	0.803457	0.120307
N	-0.902387	-0.544646	-0.157615
C	5.483919	-1.503338	-0.154225
C	5.386005	0.985568	0.129694
H	3.021662	1.781013	0.346536
N	-1.584122	0.49712	0.056028
C	6.816184	1.084676	-0.391121
C	5.621414	-2.093015	1.251967
H	0.60704	1.700679	0.334506
H	5.040519	-2.230067	-0.830151
H	6.468128	-1.297081	-0.565382
H	4.850461	1.811975	-0.330268
H	5.387901	1.163131	1.211372
C	-2.981478	0.301312	0.029134
O	7.198779	2.444423	-0.191249
H	7.493876	0.421531	0.15185
H	6.847256	0.822308	-1.453605
H	6.217797	-3.005675	1.218695
H	4.648054	-2.338884	1.675167
H	6.11622	-1.393684	1.926571
C	-3.757093	1.441697	0.27355
C	-3.615578	-0.927379	-0.22041
H	8.135988	2.543291	-0.399043
C	-5.13925	1.369726	0.273308
C	-4.994244	-1.007364	-0.222926
H	-3.253426	2.379851	0.462877
H	-3.015384	-1.804375	-0.409114
C	-5.744303	0.142466	0.024593
H	-5.746652	2.241691	0.461034
H	-5.498628	-1.942726	-0.412541
N	-7.204276	0.055301	0.022288
O	-7.718721	-1.038483	-0.202112
O	-7.845825	1.080081	0.244655

Cyclohexane: Methyl yellow cis

N	-1.008874	2.241141	-0.298574
N	-2.208547	1.893065	-0.330236
C	-2.678417	0.567069	-0.060713
C	-3.553213	-0.017904	-0.978002
C	-4.142342	-1.243816	-0.691398
C	-3.900937	-1.871205	0.528299
C	-3.062485	-1.263253	1.460521
C	-2.446384	-0.052442	1.170948
C	0.1037	1.364867	-0.194997
C	0.202491	0.066554	-0.717
C	1.392328	-0.632038	-0.660213
C	2.543384	-0.083929	-0.044091
C	2.44309	1.238402	0.446551
C	1.259549	1.942009	0.344199
N	3.710389	-0.802371	0.059928
C	4.919116	-0.156643	0.5466
C	3.838071	-2.087619	-0.608853
H	-3.76126	0.493947	-1.908879
H	-4.80469	-1.701253	-1.41534
H	-4.374774	-2.816526	0.757974
H	-2.887296	-1.734207	2.419596
H	-1.798145	0.419694	1.896714
H	-0.64305	-0.394434	-1.202429
H	1.428486	-1.61438	-1.103879
H	3.297961	1.72668	0.886555
H	1.207507	2.966404	0.690198
H	5.71182	-0.895544	0.614831
H	4.766997	0.259515	1.543375
H	5.256742	0.649967	-0.113298
H	4.8051	-2.517542	-0.365812
H	3.765323	-2.000477	-1.698414
H	3.070781	-2.784994	-0.270121

Cyclohexane: Methyl yellow trans

C	-5.295653	1.163044	0.00127
C	-5.850993	-0.112737	-6.9E-5
C	-5.015921	-1.231886	-0.001371
C	-3.637652	-1.083267	-0.001332
C	-3.075876	0.201317	1.3E-5
C	-3.914551	1.319314	0.001273
N	-1.685619	0.478495	1.89E-4
N	-0.937548	-0.532826	4.79E-4
C	0.441561	-0.282904	7.0E-5
C	1.276281	-1.405731	5.88E-4
C	2.652346	-1.28367	3.27E-4
C	3.263767	-0.008565	-6.36E-4
C	2.409096	1.125333	-9.8E-4
C	1.038654	0.988463	-6.73E-4
N	4.627274	0.13061	-0.001278
C	5.234491	1.451787	-2.64E-4
C	5.482443	-1.044827	0.001551
H	-5.936873	2.03511	0.002278
H	-6.9261	-0.238398	-1.21E-4
H	-5.447315	-2.225106	-0.002443
H	-2.986028	-1.944557	-0.002343
H	-3.463056	2.302843	0.002265
H	0.818379	-2.386614	0.001285
H	3.253579	-2.178938	8.41E-4
H	2.828587	2.119306	-0.001471
H	0.408434	1.866118	-9.45E-4
H	6.314751	1.345583	-0.001706
H	4.953055	2.026803	0.885974
H	4.951059	2.028656	-0.88458
H	6.520978	-0.729363	2.18E-4
H	5.316544	-1.66692	-0.882029
H	5.317294	-1.662497	0.888445

C.3 Benzene

!

Benzene: Disperse orange trans

Benzene: Disperse orange trans

C	-1.039914	-0.264941	0.0
C	-0.851929	-1.657278	0.0
C	-1.94354	-2.503244	0.0
C	-3.22768	-1.958529	0.0
C	-3.438956	-0.58405	0.0
C	-2.33966	0.256973	0.0
N	0.0	0.689726	0.0
N	1.164468	0.200919	0.0
C	2.220155	1.107021	0.0
C	2.088576	2.510469	0.0
C	3.202431	3.314816	0.0
C	4.502125	2.754735	0.0
C	4.632152	1.35245	0.0
C	3.509884	0.552198	0.0
H	0.150961	-2.056433	0.0
H	-1.818161	-3.575474	0.0
H	-4.444892	-0.193095	0.0
H	-2.466772	1.331089	0.0
H	1.100211	2.94679	0.0
H	3.093822	4.392921	0.0
H	5.619921	0.908545	0.0
H	3.604655	-0.52601	0.0
N	5.599978	3.559911	0.0
N	-4.38236	-2.856795	0.0
H	5.50998	4.561356	0.0
H	6.528036	3.172813	0.0
O	-4.173765	-4.068343	0.0
O	-5.505514	-2.356976	0.0

Benzene: Disperse red trans

C	2.616203	-1.522187	-0.381563
C	3.324408	-0.315846	-0.151693
C	1.238836	-1.55164	-0.37543
H	3.14181	-2.443388	-0.571963
C	2.543082	0.841084	0.125525
C	0.480427	-0.398456	-0.128417
N	4.699682	-0.262119	-0.206772
H	0.716325	-2.481148	-0.56234
C	1.168568	0.800506	0.129533
N	-0.901766	-0.549814	-0.148093
C	5.48589	-1.500313	-0.162085
C	5.384946	0.988441	0.129187
H	3.019347	1.780569	0.353085
N	-1.584673	0.491831	0.065313
C	6.812468	1.090062	-0.398415
C	5.632714	-2.095862	1.240617
H	0.605577	1.697438	0.344297
H	5.039917	-2.224863	-0.838718
H	6.467234	-1.290091	-0.577989
H	4.845762	1.814044	-0.327987
H	5.39142	1.16531	1.210923
C	-2.98155	0.297425	0.035337
O	7.193343	2.450928	-0.20205
H	7.494237	0.428985	0.141914
H	6.839224	0.826848	-1.460803
H	6.230866	-3.007035	1.199631
H	4.6624	-2.34586	1.668436
H	6.130017	-1.398511	1.915447
C	-3.75627	1.438727	0.279999
C	-3.617055	-0.929953	-0.218178
H	8.129986	2.551294	-0.414054
C	-5.138335	1.368972	0.275539
C	-4.995717	-1.007629	-0.225024
H	-3.251855	2.37599	0.472534
H	-3.018091	-1.807899	-0.407046
C	-5.744923	0.143087	0.022241
H	-5.744576	2.241959	0.463121
H	-5.50077	-1.942104	-0.418329
N	-7.204237	0.058647	0.014045
O	-7.720754	-1.03313	-0.216635
O	-7.845592	1.08347	0.238103

Benzene: Methyl yellow cis

N	-1.0101	2.241035	-0.304456
N	-2.209594	1.891389	-0.336411
C	0.102553	1.365679	-0.197088
C	-2.677771	0.565371	-0.062406
C	0.20386	0.070458	-0.72598
C	1.255488	1.939706	0.351201
C	-3.557931	-0.019479	-0.974618
C	-2.438907	-0.053765	1.168115
C	1.393005	-0.629065	-0.665802
C	2.437839	1.234413	0.457234
H	-0.638753	-0.386547	-1.220508
H	1.203024	2.962617	0.701805
C	-4.145748	-1.245239	-0.684571
C	-3.053921	-1.264291	1.461461
C	2.540285	-0.085963	-0.038214
H	-3.771124	0.492194	-1.904531
H	-1.786824	0.418185	1.890608
H	1.431179	-1.608309	-1.116117
H	3.290262	1.72069	0.904179
C	-3.897732	-1.872238	0.534031
N	3.705254	-0.807873	0.073733
H	-4.81232	-1.702712	-1.404718
H	-2.873642	-1.734952	2.419824
C	4.916616	-0.157209	0.549084
C	3.844503	-2.076984	-0.624356
H	-4.370549	-2.817454	0.76666
H	5.705066	-0.898909	0.635116
H	4.763471	0.279029	1.536544
H	5.259639	0.63439	-0.126368
H	4.802138	-2.518467	-0.365046
H	3.801533	-1.962675	-1.71318
H	3.065153	-2.777303	-0.32288

Benzene: Methyl yellow trans

C	-5.296707	1.162497	4.58E-4
C	-5.85173	-0.113473	4.74E-4
C	-5.016211	-1.232379	2.3E-4
C	-3.637901	-1.083285	-1.9E-5
C	-3.076381	0.201483	-2.3E-5
C	-3.915555	1.319191	1.98E-4
N	-1.685982	0.479211	-3.46E-4
N	-0.937291	-0.53186	-3.43E-4
C	0.441775	-0.282067	-4.3E-4
C	1.276344	-1.40523	-3.18E-4
C	2.652421	-1.283658	-3.87E-4
C	3.26441	-0.008563	-6.55E-4
C	2.409842	1.125749	-5.89E-4
C	1.039374	0.989206	-5.35E-4
N	4.627486	0.130149	-0.001036
C	5.235454	1.451163	8.12E-4
C	5.482591	-1.045549	0.001289
H	-5.938248	2.034568	6.39E-4
H	-6.926996	-0.239507	6.62E-4
H	-5.447319	-2.225929	2.24E-4
H	-2.986552	-1.944993	-2.27E-4
H	-3.464778	2.303298	1.62E-4
H	0.8185	-2.38637	-1.41E-4
H	3.253141	-2.179341	-2.55E-4
H	2.829334	2.119786	-5.61E-4
H	0.410017	1.867694	-5.06E-4
H	6.315656	1.344359	-5.12E-4
H	4.954095	2.025671	0.887364
H	4.95239	2.028522	-0.883262
H	6.521162	-0.730201	2.23E-4
H	5.316584	-1.66704	-0.882654
H	5.317155	-1.663517	0.887882

D Acknowledgements

Thanks for the advice from my supervisors:
Dr. S. Bates
Dr. A. Jones

D. Gibson for the experimental data for small molecules as well as understanding the spectra.
K. Tait for advice and the NMR data for the dyes.

E Bibliography

References

[1] M. J. Allen, T. W. Keal, and D. J. Tozer. Improved nmr chemical shifts in density functional theory. *Chemical Physics letters*, 2003.

[2] J. B. Foresman and leen Frish. *Exploring Chemistyr with electronic Stucture methods*. Gaussian Inc. Pittsburgh, 1996.

[3] M. J. Frisch, Æ. Frisch, and G. W. Trucks. *Gaussian 03 User's Refference*. Gassian Inc, 2003.

[4] P. Hohenberg and W. Kohn. Self-consistent equations including exchange and correlation effects. *Physical Review B*, 136:864–871, 1965.

[5] F. Jensen. *Introduction to computational chemistry*. John Wiley and Sons Ltd, 1999.

[6] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review A*, 140:1133–1138, 1965.

[7] C. J. Pouchert and J. Behnke. *The Aldrich Library of* ^{1}*3C and* ^{1}*H FT NMR spectra*. The Aldrich Chemical Company, 1993.

[8] K. M. Tait, J. A. Parkinson, S. P. Bates, W. J. Ebenezer, and A. C. Jones. The novel use of NMR spectroscopy with in situ laser irradiation to study azo photoisomerisation. *Journal of Photochemistry and Photobiology A: Chemistry*, 154:179–188, 2003.

[9] K. M. Tait, J. A. Parkinson, A. C. Jones, W. J. Ebenezer, and S. P. Bates. Comparison of experimental and calculated ^{1}H NMR chemical shifts of geometric photoisomers of azo dyes. Not in print yet, 2003.

[10] J. Tomasi, R. Cammi, B. Mennucci, C. Cappelli, and S. Corni. Molecular properties in solution described with a continuum solvation model. *Physical Chemistry and Chemical Physics*, 4:5697–5712, 2002.

[11] P. J. Wilson and D. J. Tozer. Varying the fraction of orbital exchange in density functional theory: Influence on nuclear magnetic resonance shielding constants. *Journal of Chemical Physics*, 2002.