Chemical Science

EDGE ARTICLE

Doubling the Power of DP4 for Computational Structure Elucidation

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

K. Ermanis,[a] K. E. B. Parkes[b], T. Agback[b] and J. M. Goodman[c]

A large-scale optimisation of density functional theory (DFT) conditions for computational NMR structure elucidation has been conducted by systematically screening the DFT functionals and statistical models. The extended PyDP4 workflow was tested on a diverse and challenging set of 42 biologically-active and stereochemically rich compounds, including highly flexible molecules. MMFF/mPW1PW91/M06-2X in combination with 2 Gaussian, 1 region statistical model was capable of identifying the correct diastereomer among up to 32 potential diastereomer upper limit. Overall a 2-fold reduction in structural uncertainty and 7-fold reduction in model overconfidence has been achieved. Tools for rapid set-up and analysis of computational and experimental results, as well as for the statistical model generation have been developed and are provided. All of this should facilitate rapid and reliable computational NMR structure elucidation, which has become a valuable tool to natural product chemists and synthetic chemists alike.

Introduction

Ever since their development, the methods for the computational prediction of NMR spectra have been an invaluable tool in the structure elucidation.[1] A particular area where simpler increment based methods cannot provide an answer and where density functional theory (DFT) methods excel is the determination of the relative stereochemistry of natural and synthetic compounds.[1-3] A key part of this process is the final decision of which candidate structure matches the experimental data the best. Several measures have been developed for this purpose, including mean absolute error, corrected mean absolute error and correlation coefficient. In addition to these, CP3 and DP4 statistical parameters have been developed to help choosing the correct structure when several sets or just one set of experimental NMR data are available, respectively.[2,3] Both CP3 and DP4 tend to give higher confidence in the correct structure than other parameters. This has led to DP4 being widely used in structure elucidation of many complex natural products[4] and also synthetic compounds.[5] Modified DP4 models have also been reported by other groups.[6] In addition to our contributions, several other groups have reported advancements in the field. These include tailored statistical models and basis sets for NMR coupling constant calculation[7], and application of neural networks to the interpretation of 1D[8] and 2D NMR data.[9] Computer Assisted Structural Elucidation (CASE) methods[10] have also been shown to be useful in resolution of structural ambiguities.

The typical computational process for NMR shift prediction has three stages. Process starts with a conformational search at molecular mechanics level and generates a number of conformers. DFT geometry optimisation is sometimes conducted at this stage on important conformers. Gauge-Independent Atomic Orbital (GIAO) NMR shift calculation is done at the DFT level on the low-lying conformers within a chosen energy window. The NMR data from all conformers is combined using Boltzmann weighting and DP4 analysis is then used to decide which is the most likely structure from the candidates.

DP4 achieves this by first applying empirical linear correction to the calculated shifts. Next, assuming that the errors between calculated and experimental data follow normal or Student's *t* distribution, probabilities are assigned to every NMR shift error. All probabilities for a particular candidate are multiplied which gives the absolute probability that this structure is the correct one. Finally, relative probabilities are derived by dividing each candidate probability by the sum of all absolute probabilities.

DP4 was originally developed for the elucidation of the relative stereochemistry of natural products. Despite that we recently tested this method on a set of drug compounds with encouraging results, which shows the generality of both the overall approach and of the particular statistical model.[11] In our studies several particularly challenging compounds were also identified. We set out to determine whether the performance of DP4 could be improved even further by optimizing the DFT conditions and statistical model used.

[a] Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. E-mail: ke291@cam.ac.uk
[b] Medivir AB, PO Box 1086, SE-141 22 Huddinge, Sweden
[c] Centre for Molecular Science Informatics, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. E-mail: jmg11@cam.ac.uk; Tel: +44 (0)1223 336434

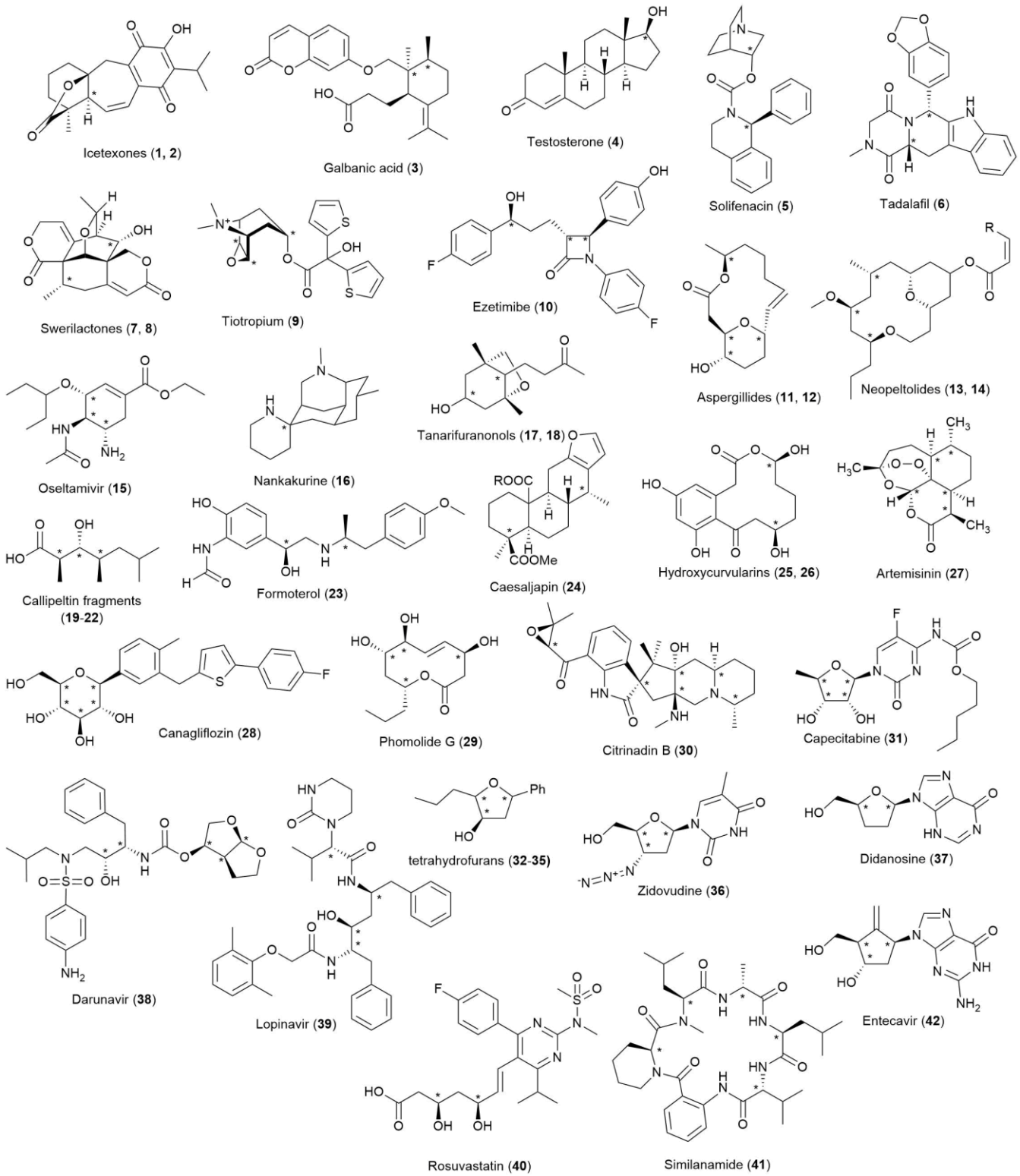

Figure 1 Molecules studied. All diastereomers were considered, by varying the configuration of the stereocentres marked with an asterisk

Results and discussion

Molecules studied

In our recent investigations in the computational elucidation of NMR structure we have paid particular attention to medicinal compounds, including peptidomimetics and nucleoside analogues (**32**-**42**, Figure 1). Prediction error distribution plays a central part in the DP4 process. We noticed these particular compound classes tended to produce larger prediction errors when compared to the original DP4 database (Figure 2). This could be explained by several factors. Some of these compounds are significantly more flexible and can have larger

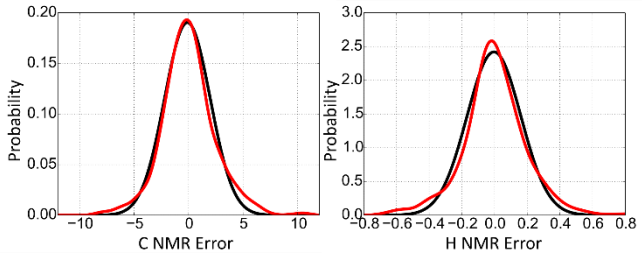

Figure 2 Comparison of fitted Gaussian distribution (black) and estimated true distribution (red) of NMR prediction errors for carbon and proton NMR

errors because of more challenging conformational searches, as well as the energetic ordering of the conformers. Others have increased potential for tautomerism which also makes accurate shift prediction more difficult.

While the original DP4 still performs much better than competing measures in these more challenging compound classes, the high prevalence of larger errors mean that the original DP4 is not an optimal model for these classes. To ensure that our investigations are applicable to as many molecules as possible, we sought to make the compound database more diverse and also challenging. Therefore a new database was composed (Figure 1) containing examples of both natural products and medicinal compounds.[12] Some of the compounds were carried over from the previous DP4 database and also from our previous study on medicinal compounds.[11] From natural products there are examples of polyketides, macrolides, alkaloids, terpenes and peptides. From medicinal compounds there is just as large structural variety, including hormones, cholesterol lowering drugs, antiviral, antimalarial, anti-asthmatic, antidiabetic and chemotherapeutic agents. Unless otherwise stated, we considered all diastereoisomers generated by varying the stereocenters marked with an asterisk in Figure 1. Initially, the calculation of the NMR spectra was done in the same manner as in previous studies and the data was used as a starting point for further investigations.

Statistical models for NMR structure elucidation

In our original report about DP4 we noted that normal distribution is not the true distribution of the prediction errors. In particular, the tail regions of the error distribution were significantly more pronounced than in the normal distribution.

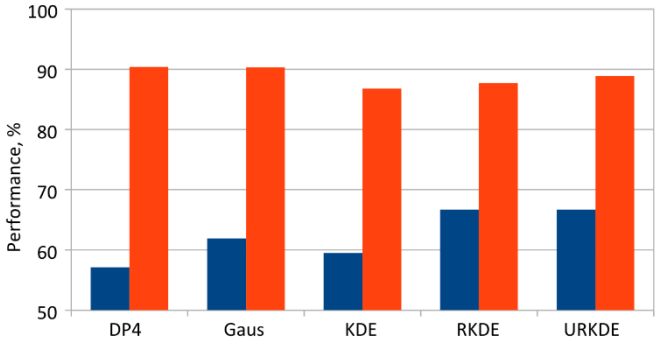

Figure 3 Initially tested statistical model actual performance (blue) and average confidence (red). NMR shifts calculated using B3LYP functional and 6-31G** basis set.

This discrepancy can lead to the overconfidence displayed by DP4 in some cases. Ideally, the average confidence of DP4 would match its average performance. By average confidence we mean the average of relative DP4 probabilities assigned to the most likely (not necessarily correct) structures in a set of compounds. By performance here and in the rest of the paper we mean percentage of correct structures assigned the highest probability among the diastereomers considered by the DP4. Close agreement of these two parameters would mean that the method not only indicates the statistically most likely structure, but also accurately notes how much confidence should be assigned to the result.

To model the distribution of errors in real compounds better, we decided to test more flexible empirical distributions. A quite popular method for the construction of empirical distributions is kernel density estimation (KDE).[13] This approach can in principle model distributions of any shape. KDE works by placing a Gaussian at each experimental point and the empirical distribution then is the sum of all the constituent normal distributions. This model was added to PyDP4 and tested on the new database. A few selected results are shown in Figure 3. As can be seen, KDE only slightly changes the rate of correct identification, however, the average confidence in the top result is now closer to the actual performance, thus providing better indication of the quality of a particular decision.

It has long been recognized that the accuracy of NMR prediction by DFT depends on the environment of the particular nucleus.[14,15] The systematic errors are different for sp^2 and sp^3 carbon atoms, and the same is true for protons attached to these carbon atoms. Several approaches have been previously used to deal with this, including using multiple computational references[14] or internal scaling.[16] The latter approach was chosen for DP4. These systematic errors were also later leveraged for the development of DP4+ method.[6] In the hopes of using this systematic error information for the development of better statistical models, we investigated the chemical shift dependence of the errors (Figure 6a). When using B3LYP for carbon NMR shift calculation, the errors are clearly clustered around 2 ppm in the sp^3 region and around -3 ppm in the sp^2 region. A more extreme example of the variation in the systematic error can be seen when M06-2X functional is used instead. Here the sp^3 region systematic error is still around 2 ppm, however the sp^2 region systematic error is of opposite sign and much larger – around 10 ppm. Another important feature in both cases is the variable dispersion of the errors around the mean, depending on the region of the spectrum. This suggests that if regional models would be developed, more than two separate distributions should be used to model the errors over all of the NMR spectrum accurately. Based on this we modified our KDE distribution statistical model into a regional model. A separate error distribution was constructed for each of 4 regions of both carbon and proton spectra. Region endpoints were chosen so that the error distributions would capture the variable mean and distribution of the errors as well as possible. The carbon spectra were divided in regions <50 ppm, 50 – 106 ppm, 106 – 148 ppm and >148 ppm. The proton spectra were divided in regions <3.0 ppm, 3.0 – 5.0 ppm, 5.0 – 7.0 ppm and

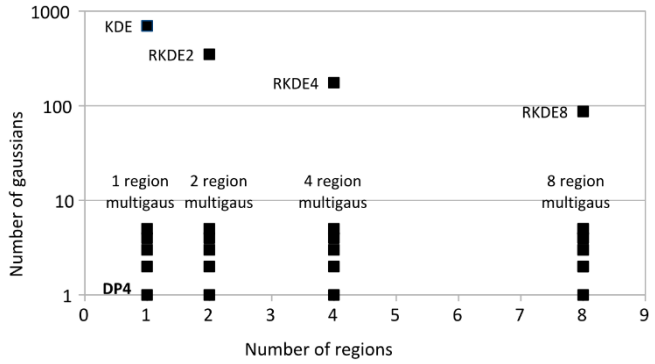

Figure 4 Statistical model space explored

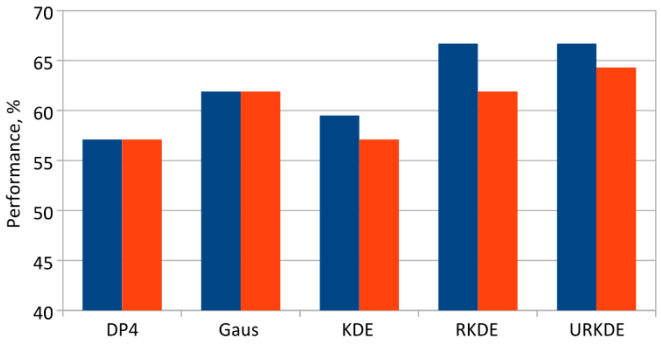

Figure 5 Statistical model in-sample (blue) and cross-validated performance (red)

>7.0 ppm. Workflows both with and without internal scaling were tested and the corresponding statistical models – called regional kernel density estimation (RKDE) and unscaled regional kernel density estimation (URKDE). The results of the initial tests are shown in Figure 3. These statistical models show improved performance with reduced overconfidence in the result.

In these two approaches we changed two meta-parameters – number of regions used to cover whole NMR spectrum for a particular nucleus, and the number of Gaussian functions used to describe the error distribution. By combining these two meta-parameters, one can plot a space of the possible statistical models (Figure 4). Models discussed so far have explored the extremes of the y axis, which is the number of Gaussian functions used in a model. The standard DP4 is a single distribution and uses a single region for the whole spectrum. KDE models use 800 Gaussians to describe the same distribution and could likely be simplified. Regional KDE models in the first instance contained 4 regions and on average 200 Gaussians per distribution. At this point we were eager to find answers to several questions: 1) Is there any redundancy in the KDE models and could they be simplified while retaining their superior performance? 2) What is the optimal number of regions used to describe errors across the chemical shift range?

To find out if there is any redundancy or even over-fitting in the rather large KDE models, a leave-one-out (LOO) cross-validation was performed. The LOO process is as follows: a compound is removed from the training set, statistical models are created from the remaining set and then DP4 probability is calculated for the removed compound. This is repeated for every compound in the database and the result is the overall success rate that was achieved without using the test compounds in the training set. The percentage of correctly identified compounds is then considered the overall out-of-sample performance.

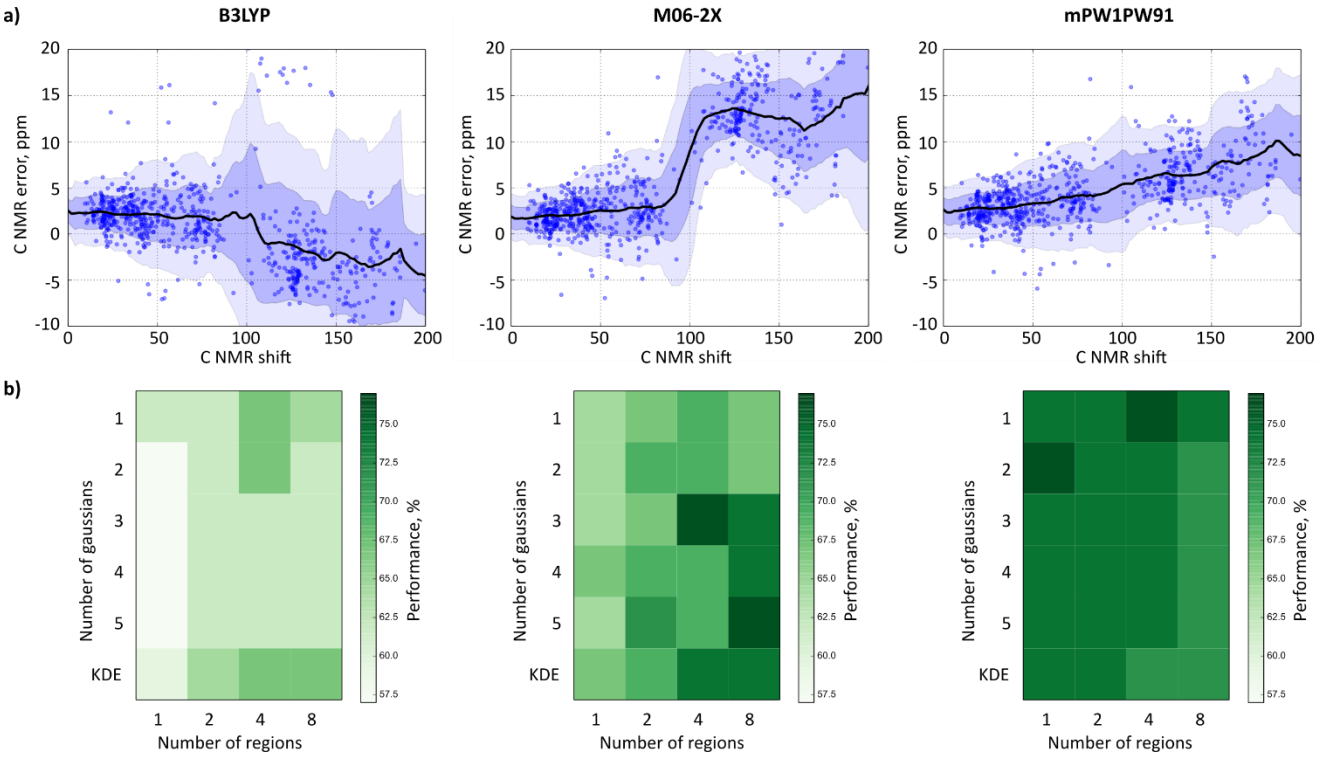

Figure 6 a) Comparison of C NMR prediction systematic errors when using B3LYP, M06-2X and mPW1PW91; b) Results from statistical model space exploration using predicted NMR shifts from B3LYP, M06-2X and mPW1PW91 functionals.

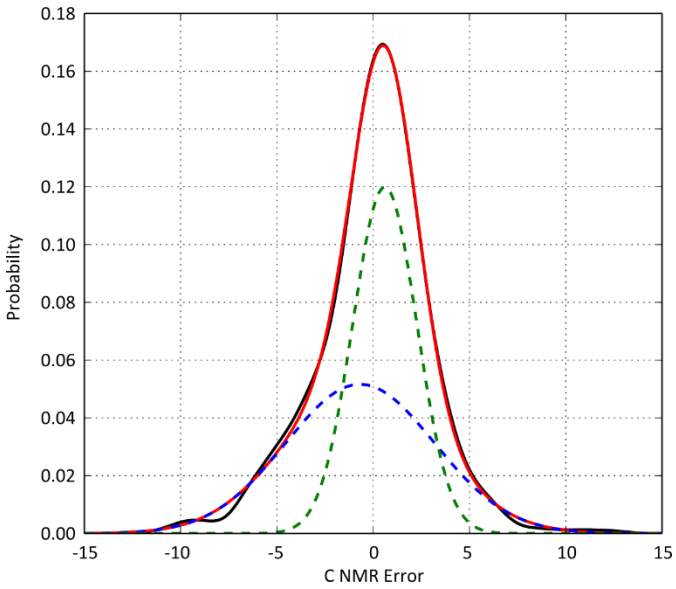

Figure 7 Use of a sum of multiple Gaussian functions (red) to capture essential features of a complex empirical distribution (black). The dashed lines represent the constituent Gaussian functions, sum of which forms the overall multiple-Gaussian distribution (red)

Comparison of in-sample and out-of-sample performance for the various statistical models is shown in Figure 5. DP4 and refitted Gaussian models appear to be robust as their performance is the same in the cross-validation. Scaled and unscaled regional KDE models performed significantly better in-sample, but in cross-validation the results are lower, while still at least as good as the simpler models. This indicates that there is some over-fitting taking place and a reduction of the number of parameters should be attempted. Global KDE also showed signs of over-fitting and performed worse than refitted single-Gaussian model both in-sample and in cross-validation.

The robust models contain one Gaussian function, the higher performing, but less robust models contain hundreds. To find the best combination of robust performance and more accurate probabilities, we tested distributions that were a sum of increasing number of Gaussian functions. These were created by taking the KDE distributions and then fitting the chosen number of Gaussian functions so as to minimize the differences between the two – and example of this is shown in Figure 7. It was found that improvements in fit in most cases diminished beyond 3 Gaussian functions. Also, these multi-Gaussian models proved to be robust in cross-validation, when 3 or less Gaussian functions were used.

With the simplification of the KDE models we had explored another part of the model space and found that there is an upper limit in the desirable model complexity. Extending the multi-Gaussian approach to regional models would complete the sampling of the model space and provide answer as to what is the best model for NMR structure elucidation. The highest number of regions tested previously has been two – both in the case of DP4+ and, effectively, in the multi-reference scaling of the NMR shifts. We varied the number of regions between 2 and 8 and the number of Gaussian functions to model each region between 1 and 5. The results are shown in Figure 6b. It was found that even in-sample the question about the best statistical model is not straightforward and very much depends on the computational method used for the calculation of the NMR shifts. So for a well behaved functional like mPW1PW91 the regional statistical models gave no additional benefit in performance and 1-region, 2-Gaussian statistical model appears to be optimal. In contrast, the versatile M06-2X functional exhibits a very non-linear systematic error and therefore benefits from more sophisticated statistical models. 4-region, 3-Gaussian and 8-region, 5-Gaussian models both looked promising and warranted repeated testing (see below).

Another important factor in the statistical model generation is the training set. In this study we strove to compile a diverse training set and thus achieve highly general statistical models. However, chemical shift calculations produce results of varying precision for different classes of compounds and different molecular features. Therefore it can be beneficial to develop tailored models when analyzing a focused library of similar compounds. While this can provide better rate of correct of identification, the largest impact of custom statistical models would be on the relative probabilities produced by DP4. This means that custom models would be able to better estimate their confidence in a particular guess.

Optimization of computational conditions

While the B3LYP functional and the double zeta basis set typically used in our DFT calculations generally performs quite well, it was hoped that a careful optimization of the chosen DFT conditions might improve the performance of DFT even further. Four different functionals were chosen for this study, including B3LYP,[17] mPW1PW91,[18] WP04[19] and M06-2X.[20] B3LYP has proved to be a very versatile general-purpose functional and has also been extensively used for chemical shift calculation.[2,3] mPW1PW91 is another general purpose functional and has been previously shown to give very good results for both proton and carbon chemical shifts.[21] WP04 is a functional designed to reproduce proton chemical shifts and also gives improved results for carbon. Finally, M06-2X is a general-purpose functional with improved handling of non-covalent interactions among other advantages.

Three different basis sets of comparable size were also chosen. Double-zeta 6-31G**, triple-zeta 6-311G*[22] and pcS-1,[23] which is a polarization-consistent basis set optimized for chemical shift calculation. The four functionals and 3 basis sets gave 12 possible DFT conditions. Calculations were repeated for every compound in the database for every DFT condition, using the same MMFF level geometries as a starting point. Each of the computational conditions was tested with 5 different statistical models – DP4, refitted Gaussian, 1-region KDE and 4-region scaled and unscaled KDE models (RKDE and URKDE).

Table 1 Mean absolute errors in ppm for carbon and proton NMR prediction for various functional and basis set combinations.

	B3LYP			mPW1PW91		
	6-31G**	6-311G*	pcS-1	6-31G**	6-311G*	pcS-1
C MAE	1.80	1.77	1.62	1.59	1.60	1.57
H MAE	0.14	0.14	0.15	0.14	0.14	0.15

	WP04			M06-2X		
	6-31G**	6-311G*	pcS-1	6-31G**	6-311G*	pcS-1
C MAE	2.21	2.00	1.80	2.16	2.35	2.49
H MAE	0.14	0.13	0.14	0.16	0.16	0.18

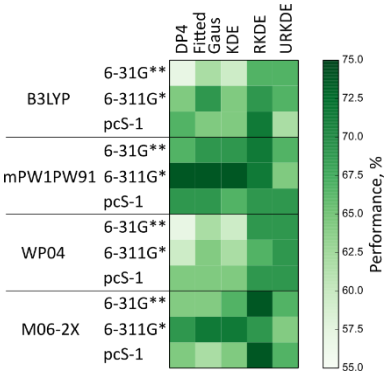

Figure 8 In-sample performance of various functional, basis set and statistical model combinations

The resulting mean absolute errors for carbon and proton NMR prediction are shown in Table 1. There is very little variation in accuracy for proton shift predictions for various computational conditions. Part of the reason might be in the way experimental data is described in the literature. Experimental proton spectra often contains several overlapping signals which are then reported as a broad range, rather than accurate chemical shift. In DP4 these broad ranges are converted to the appropriate number of signals at the mid-point of the range. This fundamentally limits the accuracy to which these shifts can be computationally predicted, since for most of the overlapping proton signals the mid-point of the range will not correspond to the actual experimental chemical shift. For carbon NMR shifts mPW1PW91 functional appears to give the best accuracy. For all but one functional, pcS-1 basis set gives the best carbon accuracy. There is no clear preference between 6-31G** and 6-311G* basis sets.

There is also marked difference in the systematic errors exhibited by different basis sets and functionals (Figure 6a). All of the DFT conditions exhibited a systematic error in carbon chemical shifts, and this error was larger in the sp^2 region of the spectra. The severity of this systematic error appears to be mostly dependent on the functional used, with mPW1PW91 having the smallest and M06-2X having the largest. The sign of this systematic error, however, appears to be mostly dependent on the basis set used. When using 6-31g**, the sp^2 region featured significant negative systematic errors, while with 6-311g* and pcS-1, the same errors were positive.

The resulting performance in diastereomer elucidation is shown in Figure 8. mPW1PW91 appears to perform the best overall, in combination with either DP4, refitted Gaussian or KDE model. The choice of basis set seems to be less important, but generally 6-311G* performs better than the other two tested here. It is worth noting that the best computational conditions for diastereomer identification are not always the ones giving the most accurate absolute prediction of chemical shifts. For example, with both B3LYP and mPW1PW91 the pcS-1 basis set gives the most accurate shift prediction, but for diastereomer identification 6-311G* works better with these functionals (Table 1 and Figure 8). The likely reason for this is that for diastereomer identification absolute chemical shift accuracy is less important than the ability to effectively predict differences between the diastereomeric candidate structure NMR spectra.

Optimization of computational conditions for energy calculation

Most DFT functionals are developed with a certain goal in mind. B3LYP and M06-2X are general purpose functionals and are optimized to provide accurate geometries and energies. WP04, on the other hand, was developed for reproduction of proton chemical shifts in GIAO calculations. The current DP4 process calculates both the energy and the NMR shifts of the molecule at the same level of theory. However, there is no reason to

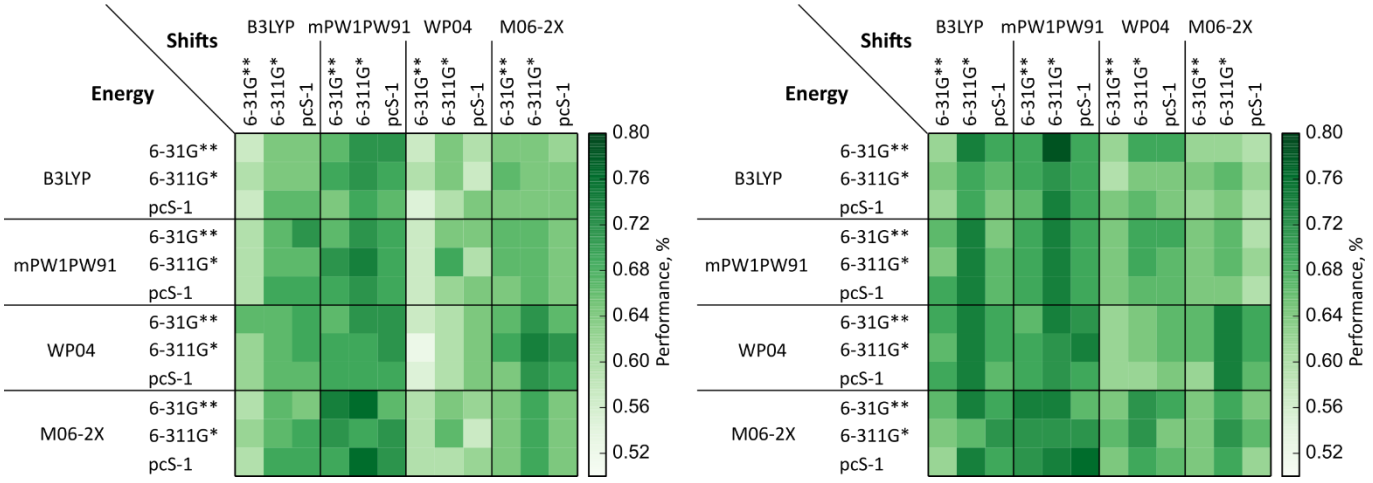

Figure 9 Performance of various functional combinations for NMR shift and energy calculations using DP4 (left) and refitted Gaussian (right) statistical model.

believe that the methods ideal for NMR shift calculation and the energy calculation would be the same. It is obvious how a poor method for NMR shift calculation would adversely impact DP4 performance. The effect of incorrect energies is more subtle. Energies are only important for the combination of NMR shifts of different conformers by Boltzmann weighting. Therefore incorrect DFT energies will only introduce significant errors in the calculated NMR shifts of flexible molecules.

To test the effect of different DFT conditions used for energy calculation, we used the data from the previous section and combined the NMR shifts calculated in one DFT conditions with energies calculated in different conditions. All possible combinations of the 12 DFT conditions were tested for the calculation of NMR shifts and energy, giving 144 combinations. In addition, 5 statistical models were tested with each of the combined DFT conditions, giving 720 workflows and 30 240 probability evaluations overall. Optimization results when using standard DP4 statistical model and refitted 1 Gaussian, 1 region model is shown in Figure 9.

The best results were obtained when mPW1PW91 functional was used for the NMR shift calculation and the M06-2X functional was used for the energy calculation. This is not surprising, as M06-2X is one of the best current general purpose functionals, especially for non-covalent interactions, which are important for accurate energetic ordering of conformers for the molecules in our test set. Similarly, as we and others have shown,[1c] mPW1PW91 gives superior accuracy in NMR shift calculations.

The results from this and previous section also highlights the complex relationship between the DFT method used and the corresponding optimal statistical model. For functionals that feature large systematic errors that vary non-linearly with regards to the chemical shift, regional models give superior results, as they are better able to deal with these complex conditions. On the other hand, mPW1PW91 performs best with a simple single region, single Gaussian statistical model, because it is very well behaved with mild and linear systematic error dependence on chemical shift.

Final testing and cross-validation

The few best performing combinations of functionals and statistical models were then retested and cross-validated using the same 42 compound dataset as previously (Table 2, Figure 10).

The most desirable features of a method is out-of-sample performance in combination with realistic estimation of confidence in the result. Promising candidates included both single functional and mixed functional workflows. mPW1PW91 functional was among the top performers. It was found that this could be further improved if the energies were calculated with M06-2X functional rather than mPW1PW91. B3LYP functional also seemed promising for conformer energy calculation, with mPW1PW91/B3LYP workflow having similar performance to mPW1PW91/M06-2X workflow. In cross-validation this performance was slightly reduced, while mPW1PW91/M06-2X performance proved to be robust both in-sample and out-of-sample.

Table 2 Summary of the best performing combinations of computational conditions and statistical models. The performance of original DP4 conditions is also included for comparison

		1	2	3	4
Shifts	Functional	B3LYP	mPW1PW91	mPW1PW91	**mPW1PW91**
	Basis set	6-31G**	6-311G*	6-311G*	**6-311G***
Energy	Functional	B3LYP	mPW1PW91	B3LYP	**M06-2X**
	Basis set	6-31G**	6-311G*	6-31G**	**6-31G****
Stat. model	N. of regions	1	1	1	**1**
	N. of Gaussians	1	2	1	**2**
Cross-validated perf. (%)		57	76	74	**79**
Average confidence (%)		90	89	89	**84**
Uncertainty (%)		43	24	26	**21**
Overconfidence (%)		33	13	15	**5**

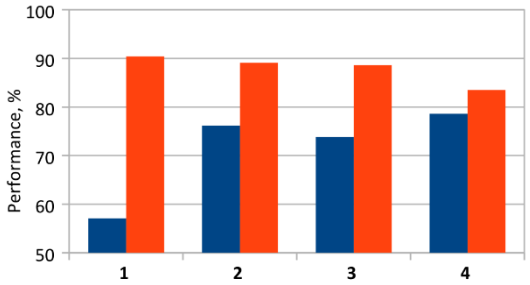

Figure 10 Performance of the original DP4 workflow and the best performing workflows from this study. Cross-validated performance is shown in blue, average confidence in the most likely structure in red.

The outcome of this final round of testing is shown in Figure 10 and Table 2. It was found that overall best results were obtained when functional mPW1PW91 was used for shift calculation, and M06-2X functional was used for energy calculation, in combination with 1 region, 2 Gaussian model. We define structural uncertainty as the gap between certainty and the DP4 performance, so a 90 % performance in a set means the method has 10 % structural uncertainty. This workflow (DP4.2) reduces the average structural uncertainty in our new dataset from 43%, with the original DP4, to 21%.This final workflow also represents a significant improvement in accurate estimation of confidence in the result, reducing the overconfidence by almost seven-fold. We calculate the overconfidence as the difference between the average confidence in the most likely structure and the percentage of correctly identified compounds in the set. As mentioned previously, we count a compound correctly identified if it is assigned the highest probability among the diastereomers.

Significant number of structures that were intractable with the original DP4 workflow, could now be computationally elucidated. Examples of these are shown in Figure 11. In all of these cases, the new DP4.2 workflow (column 4, Table 2) was able to identify the correct structure with high, but realistic confidence level. Importantly, DP4.2 is still excellent in all of the cases where DP4 was successful, indicating that we have achieved a more general method, rather than a specialized

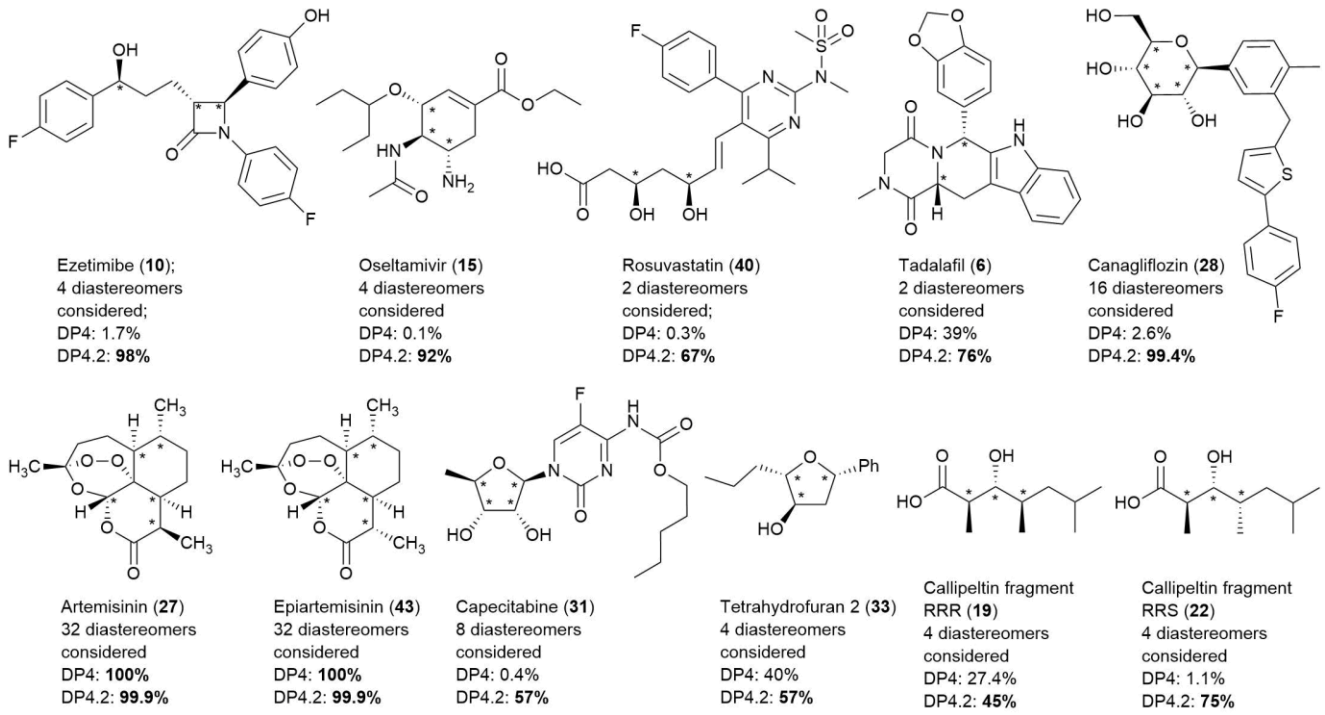

Figure 11 Examples of successfully elucidated relative stereochemistry enabled by the new version of workflow. The probabilities are combined probabilities from carbon and proton NMR data given by DP4 and the new DP4.2, respectively. Probabilities in bold denote that this structure was assigned the highest probability among the diastereomers considered.

workflow with a different focus. It is also interesting to note that most of the improved molecules are of medicinal origin, and we hope that this will encourage even wider application in the drug discovery field.

The relative stereochemistry of even such challenging molecules as canagliflozin, artemisinin and epiartemisinin could be determined with high confidence (Figure 11). Overall, the mPW1PW91/M06-2X with 1-region, 2-Gaussian statistical model provides high accuracy in diastereomer identification, high generality as shown by robustness in cross-validation and performance in a diverse dataset, and accurate estimation of the confidence in the result.

Computational methods

All molecules were first submitted to conformational search using MacroModel[24] and MMFF force field.[25] The conformational searches were done in the gas phase. Compounds containing a saturated five-membered ring were submitted to the conformational search twice, once for each of the ring-flip conformers. To ensure a thorough search of the conformational space, step count for MacroModel was adjusted so that all low-energy conformers were found at least 5 times.

Quantum mechanical calculations were carried out using Gaussian '09 software package[26] and functionals and basis sets as indicated. NMR shielding constant calculations used the GIAO method[27] and were done on the MMFF geometries from the conformational search without further optimization. Only conformers with energies within 10 kJ/mol from the global minimum were submitted to the GIAO calculation. In the rare cases where this energy-based pruning still gave more than 100 conformers for a structure, additional RMSD pruning was performed. PCM solvent models[28] were used for both DFT energy and shielding constant calculations. Calculation setup, data extraction and DP4 analysis were done using the PyDP4 script written in Python 2.7.7 The script along with further scripts for statistical model generation and testing are available on the group website (http://www-jmg.ch.cam.ac.uk/tools/nmr), as well as on GitHub (https://github.com/KristapsE/PyDP4) and in the ESI under the MIT license. All of the structure files, MacroModel and Gaussian input and output files are available from the University of Cambridge repository (https://doi.org/10.17863/CAM.13222). Similarily, parameters for selected statistical models, the calculated NMR shifts and probabilities assigned to the correct structures and other information are all available in the SI.

Conclusions

We have conducted a large-scale systematic study on the best computational methods for the elucidation of relative stereochemistry using DFT. The statistical model space was thoroughly sampled and the complex relationship between the statistical models and DFT conditions was explored. Optimal DFT conditions and the corresponding best statistical models for each have been identified. A combination of the best computational conditions with an optimized statistical model gave almost 40% improvement in correct elucidation of the relative stereochemistry in a diverse and challenging test set of biologically-active molecules. It reduced the structural uncertainty two-fold, and provided accurate estimation of the

confidence in the result by reducing the overconfidence almost seven-fold.

From a practical point of view, we consider the best DFT conditions for the elucidation of relative stereochemistry to be mPW1PW91 functional for the chemical shift calculation and the M06-2X functional for the energy calculation, in combination with a 1-region, 2-gaussian statistical model. We recommend the use of statistical models that have been trained on calculations using matching DFT conditions as this generally provides the best results. Various statistical models for several of the best DFT conditions in this study are provided in the SI and also from the group website. Also, scripts for tailored statistical model generation are provided on the group website.

This study has conducted a thorough and very large scale investigation of the various parameters involved in the computational NMR structure elucidation and it would not have been possible without a highly automated workflow PyDP4.[11] The latest version of the PyDP4 and additional scripts for custom statistical model generation can be obtained from the group website (http://www-jmg.ch.cam.ac.uk/tools/nmr), as well as from GitHub. (https://github.com/ KristapsE/PyDP4).

Acknowledgements

The authors wish to thank Medivir for the generous financial support. Part of this work was performed using the Darwin Supercomputer of the University of Cambridge High Performance Computing Service (http://www.hpc.cam.ac.uk/), provided by Dell Inc. using Strategic Research Infrastructure Funding from the Higher Education Funding Council for England and funding from the Science and Technology Facilities Council.

Conflicts of Interest

There are no conflicts of interest to declare.

Notes and references

1 For reviews in the area see: (a) M. W. Lodewyk, M. R. Siebert, D. J. Tantillo, *Chem. Rev.*, 2012, **112**, 1839; (b) D. J. Tantillo, *Nat. Prod. Rep.*, 2013, **30**, 1079. (c) N. Grimblat, A. M. Sarotti, *Chem. Eur. J.* 2016, **22**, 12246.
2 S. G. Smith, J. M. Goodman, *J. Org. Chem.*, 2009, **74**, 4597.
3 S. G. Smith, J. M. Goodman, *J. Am. Chem. Soc.*, 2010, **132**, 12946.
4 Some examples of DP4 use in natural product structure elucidation and confirmation: (a) K. M. Snyder, J. Sikorska, T. Ye, L. Fang, W. Su, R. G. Carter, K. L. McPhail, P. H.-Y. Cheong, *Org. Biomol. Chem.*, 2016, **14**, 5826 (b) L.-B. Dong, X.-D. Wu, X. Shi, Z.-J. Zhang, J. Yang, Q.-S. Zhao, *Org. Lett.* 2016, **18**, 4498 (c) T. P. Wyche, J. S. Piotrowski, Y. Hou, D. Braun, R. Deshpande, S. McIlwain, I. M. Ong, C. L. Myers, I. A. Guzei, W. M. Wrestler, D. R. Andes, T. S. Bugni, *Angew. Chem., Int. Ed.*, 2014, **126**, 11767; (d) I. Paterson, S. M. Dalby, J. C. Roberts, G. I. Naylor, E. A. Guzmán, R. Isbrucker, T. P. Pitts, P. Linley, D. Divlianska, J. K. Reed, A. E. Wright, *Angew. Chem., Int. Ed.*, 2011, **50**, 3219;
5 (a) J. Cho, S. Lee, S. Hwang, S. H. Kim, J. S. Kim, S. Kim, *Eur. J. Org. Chem.*, 2013, 4614; M. J. Bartlett, P. T. Northcote, M. Lein J. E. Harvey, *J. Org. Chem.*, 2014, **79**, 5521; (b) I. Paterson, M. Xuan, S. M. Dalby, *Angew. Chem., Int. Ed.*, 2014, **53**, 7286.
6 N. Grimblat, M. M. Zanardi, A. M. Sarotti, *J. Org. Chem.*, 2015, **80**, 12526.
7 (a) A. G. Kutateladze, O. A. Mukhina , *J. Org. Chem.*, 2014, **79**, 8397; (b) A. G. Kutateladze, O. A. Mukhina , *J. Org. Chem.*, 2015, **80**, 5218; (c) A. G. Kutateladze, O. A. Mukhina , *J. Org. Chem.*, 2015, **80**, 10838;
8 A. M. Sarotti *Org. Biomol. Chem.* 2013, **11**, 4847
9 M. M. Zanardi, A. M. Sarotti *J. Org. Chem.*, 2015, **80**, 9371;
10 A. V. Buevich, M. E. Elyashberg *Magn. Reson. Chem.* 2017, 1
11 K. Ermanis, K. E. B. Parkes, T. Agback, J. M. Goodman, *Org. Biomol. Chem.* 2016, **14**, 3943
12 Full references for the sources of all compounds spectral data can be found in the Supporting Information
13 B. W. Silverman, Density Estimation for Statistics and Data Analysis; Chapman and Hall: Bath, U.K., 1986
14 A. M. Sarotti, S. C. Pellegrinet, *J. Org. Chem.* 2009, **74**, 7254
15 (a) P. Wipf, A. Kerekes, *J. Nat. Prod.* 2003, **66**, 716. (b) C. Timmons, P. Wipf, *J. Org. Chem.* 2008, **73**, 9168.
16 G. Barone, L. Gomez-Paloma, D. Duca, A. Silvestri, R. Riccio, G. Bifulco, *Chem. Eur. J.* 2002, **8**, 3233
17 (a) A. D. Becke, *Phys. Rev. A* 1988, **38**, 3098. (b) C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* 1988, **37**, 785. (c) A. D. Becke, *J. Chem. Phys.* 1993, **98**, 5648. (d) P. J. Stephens, F. J. Devlin, C. F. Chabalowski, M. J. Frisch, *J. Phys. Chem.* 1994, **98**, 11623
18 C. Adamo, V. Barone, *J. Chem. Phys.* 1998; **108**, 664
19 K. W. Wiitala, T. R. Hoye, C. J. Cramer, *J. Chem. Theory Comput.* 2006, **2**, 1085
20 Y. Zhao, D. Truhlar, *Theor. Chem. Acc.* 2008, **120**, 215.
21 P. Cimino, L. Gomez-Paloma, D. Duca, R. Riccio, G. Bifulco, *Magn. Reson. Chem.* 2004, **42**, S26.
22 W. J. Hehre, L. Radom, P. v. R. Schleyer, J. A. Pople, Ab Initio Molecular Orbital Theory; Wiley: New York, 1986.
23 F. Jensen, *J. Chem. Theory Comput.* 2008, ***4***, 719
24 MacroModel, version 9.9, Schrödinger, LLC, New York, NY, 2009.
25 (a) T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 490; (b) T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 520; (c) T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 553; (d) T. A. Halgren, R. B. Nachbar, *J. Comput. Chem.*, 1996, **17**, 587; (e) T. A. Halgren, *J. Comput. Chem.*, 1996, **17**, 616; (f) T. A. Halgren, *J. Comput. Chem.*, 1999, **20**, 720; (g) T. A. Halgren, *J. Comput. Chem.*, 1999, **20**, 730.
26 Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J. ; Fox, D. J. Gaussian 09, Revision D.01, Gaussian, Inc., Wallingford CT, 2009.
27 (a) F. London, *J. Phys. Radium* 1937, **8**, 397. (b) R. Ditchfield, *J. Chem. Phys.* 1972, **56**, 5688. (c) K. Wolinski, J. F. Hinton, P. Pulay, *J. Am. Chem. Soc.* 1990, **112**, 8251.
28 B. Mennucci, J. Tomasi, *J. Chem. Phys.* 1997, **106**, 5151.