ChemPubSoc Europe

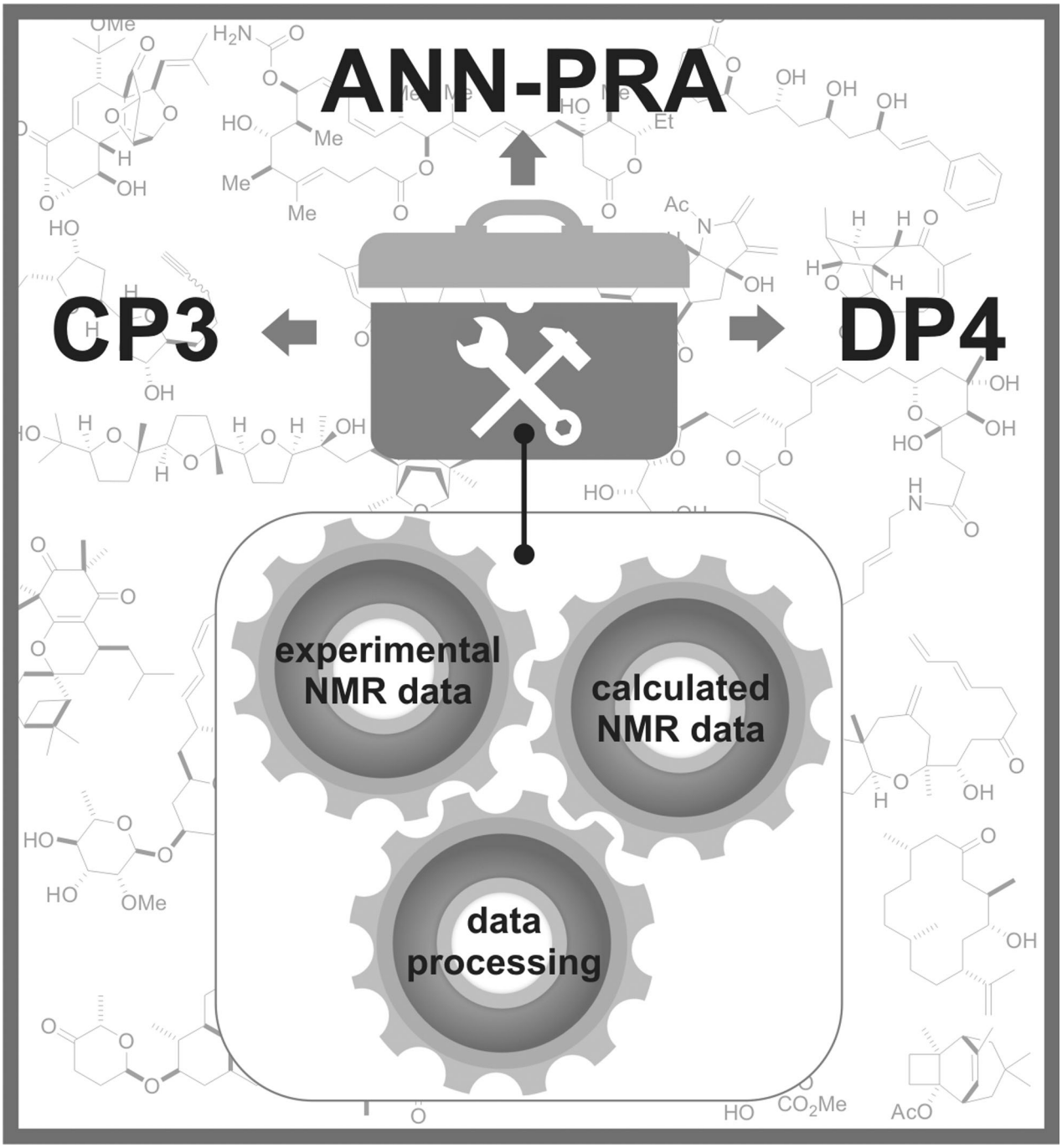
DOI: 10.1002/chem.201601150

CHEMISTRY
A European Journal
Minireview

Computational Chemistry

Computational Chemistry to the Rescue: Modern Toolboxes for the Assignment of Complex Molecules by GIAO NMR Calculations**

Nicolas Grimblat and Ariel M. Sarotti*[a]



Abstract: The calculations of NMR properties of molecules using quantum chemical methods have deeply impacted several branches of organic chemistry. They are particularly important in structural or stereochemical assignments of organic compounds, with implications in total synthesis, stereoselective reactions, and natural products chemistry. In studying the evolution of the strategies developed to support (or reject) a structural proposal, it becomes clear that the most effective and accurate ones involve sophisticated procedures to correlate experimental and computational data. Owing to their relatively high mathematical complexity, such calculations (CP3, DP4, ANN-PRA) are often carried out using additional computational resources provided by the authors (such as applets or Excel files). This Minireview will cover the state-of-the-art of these toolboxes in the assignment of organic molecules, including mathematical definitions, updates, and discussion of relevant examples.

Introduction

The calculation of magnetic properties of molecules with quantum chemical approaches has been extensively used in different branches of organic chemistry since the beginning of the 21st century.[1–3] The development of reliable methods to perform accurate NMR calculations, coupled with the impressive processing capabilities of modern computers and the user-friendly environment of most computational chemistry software packages, are the main responsible to the growing interest in the field. More than 2000 articles published in the recent past demonstrate that quantum chemical calculations of NMR shifts or coupling constants attains a privileged status in modern organic chemistry.[1–3]

The impact of this discipline has been felt in different areas—the structural assignment of natural products has had the most benefit. The importance of using computed NMR data in supporting (or rejecting) a structural proposal is magnified considering the alarmingly large number of natural products that have been incorrectly assigned in the recent past.[4,5] This has an adverse impact in different branches of the scientific community, from the isolation team to those who have been attracted by promising biological activities and beautiful structures and venture into the synthesis of erroneous target compounds. Ironically, it is precisely the total synthesis of the putative structure that often plays a decisive role in solving structural elucidation problems. During the last decade, the structures of more than 300 natural products have been revised by total synthesis.[6]

The high accuracy offered by quantum chemically computed NMR properties offers a valuable and helpful alternative to prevent structural or stereochemical misassignments. In this respect, a wide variety of strategies to manage the decision making process (to decide if a putative structure is correct, to select the most likely isomer among several candidates, etc.) have been used. In spite of the fact that they all share a main premise (the correct structure will show good or better agreement between experimental and computational NMR data), the principal differences arise from how such agreement is computed.

On one side, the entire decision making process lies mainly in the value of basic and easy-to-calculate statistical parameters of correlation, such as R^2, the mean absolute error (MAE) or corrected mean absolute error (CMAE). Then, if the NMR calculation is expected to support a given assignment, any outside value in those parameters indicates unsound proposals. For example, Rychnovsky[3i] and Tantillo[3c] used this approach to challenge the originally assigned structures of hexacyclinol and aquatolide, respectively (Figure 1).

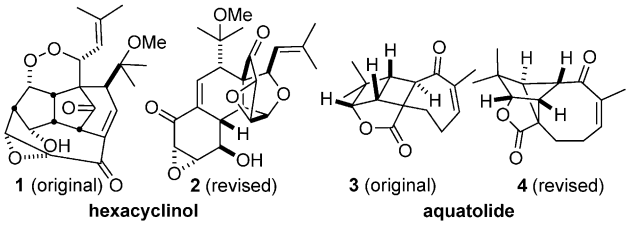


Figure 1. Originally proposed (left) and revised (right) structures of hexacyclinol and aquatolide.

Conversely, the NMR calculations can be used to determine the correct structure among several (at least two) candidates. In this case, the NMR values computed for the right structure are expected to display a closer match (higher R^2, lower MAE and CMAE, etc.) with the experimental data than those computed for the other candidates. Following this procedure, the alternative structural proposals for hexacyclinol and aquatlide were made (and further validated experimentally).

The use of simple statistical parameters of correlation have been extensively used in the recent past to settle structural issues of several natural products, including artarborol,[3o] obtusallene,[3p] samoquasine A,[3q] spiroleucettadine,[3r] kadlongilactones,[3s] ketopelenolides,[3t] and vannusal B,[3e] providing helpful assistance to identify the correct isomer among several candidates. However, in some cases the chemical shifts (and therefore, the MAE, R^2, etc.) computed for two or more isomers are similar enough to prevent the unequivocal assignment towards a particular structure.

Recently, more sophisticated procedures to correlate experimental and computational data have been developed, provid-

[a] *Dr. N. Grimblat, Dr. A. M. Sarotti*
Instituto de Química Rosario CONICET Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario
Suipacha 531, Rosario (2000) (Argentina)
E-mail: sarotti@iquir-conicet.gov.ar

[**] *GIAO = Gauge independent atomic orbitals.*

ing higher levels of confidence than those expected from simple statistical parameters. From their higher mathematical complexity, additional computational resources are often needed to simplify the calculations.

These computational toolboxes, which combine quantum chemical calculations of NMR shifts with complex upstream data processing, are among the most accurate and important alternatives to solve structural and stereochemical problems from a theoretical point of view. Three main methods have been developed, namely CP3,[7] DP4,[8] and ANN-PRA,[9] differing (among other things) on the size of raw data needed in their formulations (Figure 2). Thus, CP3 was developed to assign two sets of experimental data to two possible structures,[7] whereas DP4 was built to assign the most likely structure that, among two or more candidates, best correlates with only one set of experimental data.[8] Finally, the ANN-PRA (artificial neural network pattern recognition analysis) was introduced to determine the correctness of a structural proposal using only one set of experimental and computational data.[9]

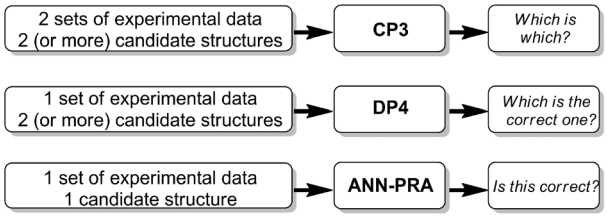


Figure 2. Schematic representation of CP3, DP4, and ANN-PRA.

This Minireview will cover the state-of-the-art of these toolboxes in the assignment of organic molecules, including mathematical definitions, updates and discussion of relevant examples. Because the oldest method (CP3) was published in 2009, the literature starting from that year to date was carefully and systematically covered. Therefore, earlier publications have not been taken into account, but some have been included in this article for different reasons.

While the main purpose of this Minireview is to promote and present all the quantum chemistry tools available to assist in the NMR elucidation process, there are other useful approaches that deserve to be mentioned. One of them is the calculation of coupling constants (*J*) using quantum chemistry methods,[3k,u–w] which have proven to be extremely useful to explore the conformational and stereochemical space of organic molecules.[1] Other important methods that must be highlighted are the non-quantum tools for structural assignment. The most prominent work in this area has been made and it is constantly under development mainly by Prof. Elyashberg from ACD Labs, for which in-depth analysis and revision has already been made. The expert systems developed are the result of over four decades of investigation in the development of computer-aided structure elucidation (CASE) methods.[10]

CP3

The CP3 parameter was introduced in 2009 by Smith and Goodman (from Cambridge University) to tackle the particular problem of assigning two sets of experimental NMR data to two possible structures.[7] This situation is very common in synthetic organic chemistry, in which many chemical transformations might afford two (or more) isomers that even after extensive NMR experiments cannot be indisputably assigned.

Inspired by the proposal of Belostotskii and Rodríguez that the cancellation of systematic errors during the NMR calculation procedure increases the accuracy in the calculation of the difference between the chemical shifts of related carbon atoms (rather than the calculations of the chemical shifts themselves),[3h,11] the Goodman group formulated three different comparison parameters (CP), see Equation (1), in which

$$CPn = \frac{\sum_i f_n}{\sum_i (\Delta_{exptl})^2} \quad \begin{cases} n=1;\ f_1 = \Delta_{exptl}\Delta_{calcd} \\ n=2;\ f_2 = \begin{cases} \Delta^3_{exptl}/\Delta_{calcd}\ if\ |\Delta_{calcd}| > |\Delta_{exptl}| \\ \Delta_{exptl}\Delta_{calcd}\ otherwise \end{cases} \\ n=3;\ f_3 = \begin{cases} \Delta^3_{exptl}/\Delta_{calcd}\ if\ \Delta_{calcd} > \Delta_{exptl} \\ \Delta_{exptl}\Delta_{calcd}\ otherwise \end{cases} \end{cases} \quad (1)$$

$\Delta^i_{exptl} = \delta^i_A - \delta^i_B$ is the difference in the chemical shift of the *i*th nuclei between the two sets of experimental data (namely, **A** and **B**, Figure 3), and $\Delta^i_{calcd} = \delta^i_a - \delta^i_b$ is the difference in the calculated shifts of the *i*th nuclei between candidate structures **a** and **b**, Figure 3.

Nicolas Grimblat graduated in Chemistry at the National University of Rosario (Argentina) in 2010. In 2015 he completed his Ph.D. in Organic and Computational chemistry under the supervision of Dr. Silvina C. Pellegrinet at the Instituto de Química Rosario (IQUIR, Argentina). He is currently carrying out postdoctoral research under the supervision of Dr. Ariel M. Sarotti at IQUIR, working on the development of more robust quantum NMR calculation methodologies and in silico rational design.

Ariel M. Sarotti was born in Rosario (Argentina) in 1978. He completed his Ph.D. in synthetic organic chemistry at the Instituto de Química Rosario (IQUIR, Argentina) in 2007, under supervision of Dr. Alejandra G. Suárez. He then carried out postdoctoral research in computational chemistry with Dr. Silvina C. Pellegrinet at IQUIR. He is now a researcher of the Argentine National Council Research (CONICET) and Professor of Organic Chemistry at the Universidad Nacional de Rosario (UNR). His research interests center on green and sustainable chemistry, asymmetric synthesis, and computational chemistry

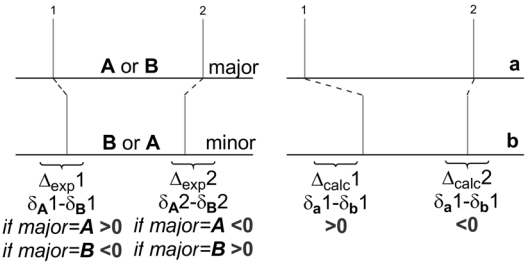


Figure 3. Pictorial and simplified example to illustrate the basic functioning underlying all the CP*n* methods. Left: experimental data (**A** and **B**); Right: calculated data (**a** and **b**).

Assuming that the experimental data has been fully assigned, that is, all resonances assigned to all the corresponding nuclei in the proposed structure (which might not be necessarily the case), the problem is simply a matter of selecting the best match among two options: **A** = **a**, **B** = **b** (option 1) or **A** = **b**, **B** = **a** (option 2). From the simplified example provided in Figure 3, in the first option the Δ_{exptl} computed for carbon 1 is positive ($\Delta^1_{exptl} > 0$), whereas the corresponding difference for carbon 2 is negative ($\Delta^2_{exptl} < 0$). The opposite is observed in the case of the second option. On the other hand, the calculated differences between **a** and **b** (Δ_{calcd}) for carbons 1 and 2 are positive and negative, respectively. The multiplication between experimental and calculated differences ($\Delta_{exptl}\Delta_{calcd}$) for both carbons 1 and 2 are positive (and therefore, their sum) for the first option, and negative for the second one. Dividing by a normalization term, $\Sigma(\Delta_{exptl})^2$, which represents the value of the overall sum that would be obtained if all the experimental differences would be perfectly reproduced by the calculations ($\Delta_{exptl} = \Delta_{calcd}$), affords the CP1 parameter. A large and positive CP1 value computed for the first option (**A** = **a**, **B** = **b**) indicates that the assignment is likely to be correct. On the other hand, the negative CP1 value obtained in option 2 means that the agreement is poor, and the assignment is probably incorrect.

However, very often the calculations do not perfectly reproduce the trend experimentally observed for close nuclei (Δ_{calcd} and Δ_{exptl} with different signs when $\delta_a \approx \delta_b$) even for the correct match, affording negative terms in the numerator of the equations and lowering the CP1 value for the correct assignment. The same occurs when Δ_{calcd} has the same sign but smaller magnitude than Δ_{exptl}. In both cases the CP1 value is reduced from the expected value when the experimental differences are perfectly reproduced by the calculations ($\Delta_{calcd} = \Delta_{exptl}$). In fact, such reduction is desirable, as it indicates a less good agreement. The main problem of CP1 is the lack of an upper threshold. For instance, if the calculation overstates the real differentiation between two nuclei ($\Delta_{calcd} \gg \Delta_{exptl}$), the CP1 value would be higher than that expected for a perfect match ($\Delta_{calcd} = \Delta_{exptl}$), which is totally counterintuitive. The problem was solved by replacing the term $\Delta_{calcd}\Delta_{exptl}$ with $(\Delta_{exptl})^3/\Delta_{calcd}$ whenever the conflict situation ($\Delta_{calcd} > \Delta_{exptl}$) is met, affording the CP2 parameter. Additionally, CP3 was also introduced to correct the CP2 underestimation in highly mismatched situations ($\Delta_{calcd} > -\Delta_{exptl}$) by including the extra requirement that the correction is applied only when both Δ_{calcd} and Δ_{exptl} have the same sign. It is important to point out that only CP2 has a lower limit, −1, corresponding to a perfect disagreement (meaning that the structures have been incorrectly assigned and $\Delta_{calcd} = -\Delta_{exptl}$ for all nuclei under consideration).

The robustness of the CP*n* methods was evaluated by using a set of 28 isomeric pairs, some of which are depicted in Figure 4, using the B3LYP/6-31G(d,p)//MMFF level of theory for

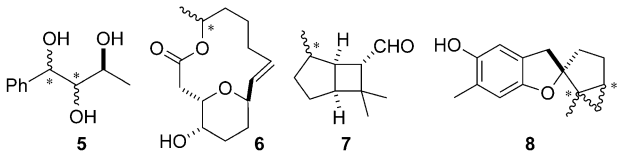


Figure 4. Representative set of compounds used in the development of CP3. Carbon atoms with varied stereochemistry are marked with an asterisk.

the GIAO (gauge independent atomic orbitals) NMR calculation procedure. The use of MMFF geometries for the geometry optimization step was chosen to provide good results in the NMR shift calculations at minimal computational cost (avoiding expensive ab initio or DFT methods, particularly demanding in the case of highly flexible compounds). On the basis of the percentage of correct assignment made by each parameter, the authors concluded that CP3 afforded better results than CP1 and CP2, and recommended the former for structural assignment in place of other conventional parameters such as R^2, MAE, and CMAE.[7]

Moreover, the new CP*n* parameters stood out on the high levels of confidence with which the assignments were made. For instance, both CP3 and R^2 performed equally well in correctly assigning the 28 isomeric pairs used in the testing set when all data (^{13}C and ^{1}H) were employed. However, whereas 26 cases were identified in high probability (>99%) with CP3, the same level of certainty was achieved in only 2 cases by using R^2.

The calculation of the probability associated to a given assignment is done from conditional probability elements and Bayes's theorem as given by Equation (2), in which $P(AC_1 \mid R_1,R_2)$ is the probability that the proposed assignment is correct (AC_1), R_1 and R_2 are the parameter values (R^2, MAE, CP3, etc.) computed for each assignment (AC_1 and AC_2), and $P(R_i(AC_j))$ is the probability of getting a given parameter value assuming the assignment is correct (or incorrect) and that the terms R_i are normally distributed independent random variables. For a more comprehensive discussion of this equation, see the original reference.[7]

$$P(AC_1|R_1,R_2) = \frac{P(R_1|AC_1)P(R_2|AC_1)P(AC_1)}{\sum_i P(R_1|AC_i)P(R_2|AC_i)P(AC_i)} \quad (2)$$

"Toolbox"

Despite the CP3 parameter being fairly easy to compute "by hand", the probability calculation is slightly more complicated

and requires the knowledge of the normal distribution descriptors (expectation value and standard deviation) corresponding to correct and incorrect assignments, and also to the type of nuclei from which the experimental data was taken (^{1}H, ^{13}C or both). In an effort to bring this useful methodology to the organic community, the authors developed an applet that is freely to use online at http://www-jmg.ch.cam.ac.uk/tools/nmr/ The user is asked to introduce the data (experimental and calculated), that can or cannot be assigned (this is an optional step), and the applet computes which experimental data belongs to which isomer, providing the confidence of the assignment as well.

Applications

Chemical reactions affording at least two isomers that cannot be unequivocally assigned with the experimental information available represent examples that might benefit from CP3 calculations. However, this might occur at early stages of a research program, when the preliminary conclusions afforded by CP3 can be overshadowed by further experimental evidence obtained in later stages (such as X-ray analysis of an advanced intermediate, etc.). Therefore, even though all the CP3-related details might be excluded from the final publication, the importance of this magnificent computational tool should not be diminished. This being said, we found that the CP3 parameter was used in solving structural problems in several studies.[12] Interestingly, only in 33% of the cases were the NMR shielding tensors computed at the level of theory reported in the seminal publication (B3LYP/6-31G(d,p)//MMFF).[12a–d] In the remaining 67%, the geometry optimization steps was conducted at higher levels, such as B3LYP or mPW1PW91 DFT functionals, coupled with double- or triple-ζ basis sets.[12e–k] Hereafter, three representative examples of the use of CP3 in solving different types of structural problems are given.

Stereochemical assignment

In 2010, one year after the publication of CP3, the Goodman group in collaboration with the Paterson group (both from Cambridge University), reported the synthesis of the 16 possible stereopentads **9** (Figure 5), bearing a common substitution pattern found in many natural products.[12d] After an in depth analysis of the similarity of the experimental NMR spectra of all isomers, the corresponding GIAO NMR calculations were carried out at the B3LYP/6-31G(d,p)//MMFF level. With the complete set of experimental and computational data, all possible 120 pairs were matched and the corresponding CP3 parameters were computed. Notably, the CP3 performed very well, correctly assigning 99 pairs in high confidence, and failing only in 2 cases (by selecting the wrong match in high probability).[12d] To date, this work is one of the most remarkable examples of the robustness of a computational methodology to help in the stereoassignment of closely related molecules.

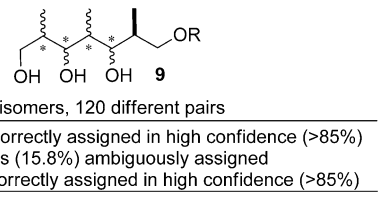


Figure 5. Sixteen isomeric stereopentads used to evaluate the CP3 performance. Carbon atoms with varied stereochemistry are marked with an asterisk. R = Bn or TBS.

Structural assignment

Hodgson et al. reported the intramolecular cycloaddition of diazoketodiester **10** under Rh^{II} catalysis to afford the bicyclic derivative **11** in good yields. Upon acidic treatment, **11** rearranged to **12**, featuring the 2,8-dioxabicyclo[3.2.1]octante motif present in zaragozic acid and other squalestatins (Figure 6). However, with the information available the authors

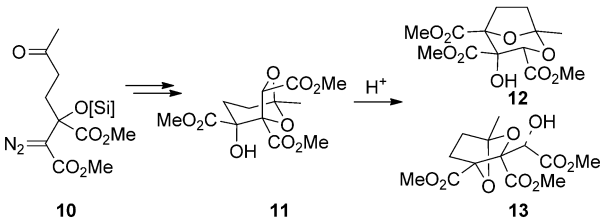


Figure 6. Synthesis of zaragozic acid intermediate.

could not unambiguously ruled out the formation of 5,7-dioxabicyclo[2.2.1]heptanetriolate **13**, the other plausible rearranged product. To settle this issue, the authors performed NMR calculations of **11**, **12** and **13** at the B3LYP/6-31G(d,p)//OPLSAA level of theory. The CP3 calculations showed that **11** and **12** had been correctly assigned with over 99.9% probability (0.58 for the correct assignment, −1.19 for the incorrect). Replacing **11** or **12** by **13** afforded negative CP3 values in all cases, indicating poor fit between any of the two sets of experimental values with those computed for **13** and reaffirming the assignment made for **12**.[12b]

Regiochemical assignment

In 1996, Miftakhov et al. reported an unusual *meta* selectivity for the the Diels–Alder reaction between levoglucosenone (**14**) and isoprene, indicating **15** as the major isolated adduct, both under thermal and Lewis acid promoted conditions.[13] More recently, from the basis of DFT calculations (conceptual DFT, transition state modeling, etc.), Sarotti et al. questioned the original assignment, suggesting that the *para* adduct **16** should be the preferred product of the reaction (Figure 7). The CP3 values computed at the mPW1PW91/6-31G(d) level of theory played a significant role for providing further evidence in favor of the revision, since the original arrangement ("major-**15**, minor-**16**") afforded a negative CP3 value (−0.79), whereas the corresponding parameter for the inverse assignment ("major-**16**, minor-**15**") was positive (0.76). Further experimental syn-

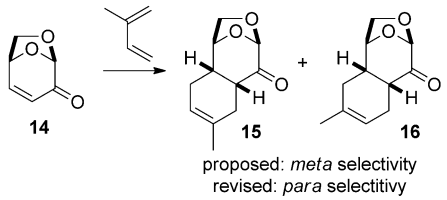


Figure 7. Diels–Alder reaction between levoglucosenone and isoprene.

thesis, isolation, and full characterization of the major adduct (**16**) validated the computational predictions.[12h]

DP4

As discussed in the previous section, the CP3 parameter was designed to match two sets of experimental data with two plausible structures. Therefore, it is useless in the case of isomerically pure natural products or chemical reactions with perfect levels of selectivity. To deal with the more difficult and challenging task of identifying the correct structure among several candidates when only one set of experimental data is available, in 2010 Smith and Goodman (the same developers of CP3) made a major breakthrough by introducing the DP4 probability.[8]

The basic formulation of the DP4 probability relies on two main assumptions regarding the errors *e* between experimental, δ_{exptl}, and calculated (scaled) chemical shifts, δ_s, ($e = \delta_s - \delta_{exptl}$), namely, that they are normally distributed and independent random variables. Regarding the first issue, the authors proved that the errors fit better to a *t* distribution, characterized with mean $\mu = 0$ (as consequence of the linear scaling procedure), standard deviation σ, and degrees of freedom ν. Then, for a molecule with *N* nuclei, the probability associated to each *i*th error can be calculated from the cumulative probability distribution function defined by the terms σ and ν. The key step for the DP4 formulation is that, under the assumption of error independence and randomness, multiplication of all individual *i*th probabilities gives the overall probability for that molecule. Finally, the percentage probabilities for all candidates are obtained using Bayes's theorem. Mathematically, the DP4 probability is defined as shown in Equation (3).

$$P(i) = \frac{\prod_{k=1}^{N}[1-T^{\nu}(e_k^i/\sigma)]}{\sum_{j=1}^{m}\prod_{k=1}^{N}[1-T^{\nu}(e_k^j/\sigma)]} \quad (3)$$

As shown in Figure 8, as each individual error increases in size, the associated probability (from standard cumulative *t*-distributed functions) rapidly decreases, causing a sudden drop in the combined probabilities for that *i*th candidate (that is, in the product of all individual probabilities). This severe penalization of large outliers made by DP4 is responsible for the known tendency to overstate the final probability in some cases.

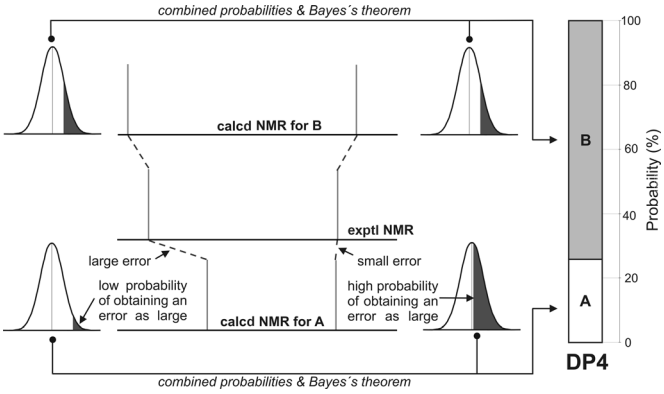


Figure 8. Schematic illustration of the DP4 probability.

The key σ and ν terms were estimated from 1717 ^{13}C shifts and 1794 ^{1}H shifts computed using 117 known organic molecules (some representative examples shown in Figure 9), at the B3LYP/6-31G(d,p)//MMFF level of theory (the same one used in the CP3 development, and for the same reasons). The resulting errors (differences between experimental and calculated shifts) were fitted to *t* distributions using the R statistical program, obtaining $\sigma = 2.306$ ppm (^{13}C) or 0.185 ppm (^{1}H), and $\nu = 11.38$ (^{13}C) or 14.18 (^{1}H) values for the carbon and proton series of errors, respectively.

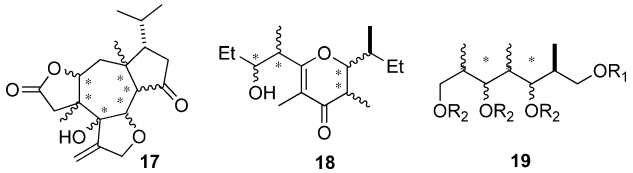


Figure 9. Representative set of compounds used in the development of DP4. Carbon atoms with varied stereochemistry are marked with an asterisk.

"Toolbox"

To simplify the calculation process, the authors created a Java Applet freely available at http://www-jmg.ch.cam.ac.uk/tools/nmr/DP4/. The user is first asked to select the database version, since the [σ,ν] values that define the probability distribution and therefore the values obtained from Equation (3), depend not only on the levels of theory employed for the NMR calculations, but also on the particular set of examples used to estimate them. The "DP4-original" database contains the [σ,ν] values indicated above [2.306,11.38] and [0.185,14.18] for ^{13}C and ^{1}H series, respectively. However, after the publication of the paper the authors incremented the number of examples in their database, providing more "realistic" values of [σ,ν]. Additionally, the user can select to compute the DP4 probabilities assuming either a normal or a *t* distribution for the individual errors. The authors recommend the second option to be less susceptible to overstating the probability values. Finally, the computational and experimental shifts must be introduced. The user can choose to compute the DP4 probability using only ^{13}C data, ^{1}H data or all data (^{13}C and ^{1}H), the last option

being the most recommended. The computational data must be labeled, whereas the experimental data might not. The applet can also help to perform the assignment (determine which experimental shift belongs to which nuclei of the molecule). Finally, the program calculates the probabilities that any proposal is the correct assignment.

Applications

The performance of DP4 in the assignment of the stereostructure of complex organic molecules was evaluated and compared with standard statistical parameters of correlation (MAE, CMAE, etc.), and the former was shown to provide much more successful and confident results.[8] Since then, it has been used in more than 50 assignments or revisions of natural and unnatural products, a remarkable fact considering the youth of the method. For that reason, DP4 emerges as one of the most popular and sophisticated computational methods for the structural or stereochemical determinations of organic compounds.[12c,i,14]

After thorough analysis of the recent literature on this subject, one of the major issues that caught our attention was the high diversity of methods that have been used to compute the NMR shifts for further DP4 calculations. Only in 27% of the cases this procedure was conducted at the level of theory at which DP4 was formulated (B3LYP/6-31G(d,p)//MMFF).[12c,14al–ay] In the remaining 73%, both geometry optimization and/or NMR calculation steps were carried out at different levels.[12i,14a–ak] The general observed trend was the performance of both stages using standard DFT functionals (B3LYP, mPW1PW91, M06-2X, wB97XD, OPBE, etc.), coupled with a variety of basis sets (Pople and Dunning-type double- and triple-ζ basis sets, pcS-2, etc.), including also the solvent effect in some cases. At this point it should be noted that the DP4 probability value depends heavily on the $[\sigma,\nu]$ values, that in turn, exhibit a sharp relationship with the level of theory employed to compute the NMR shifts. For instance, the $[\sigma,\nu]$ values for ^{1}H NMR series are [2.306; 11.38][8] and [1.27; 4.19][15] at the B3LYP/6-31G(d,p)//MMFF and mPW1PW91/6-31G(d)//B3LYP/6-31G(d) levels, respectively. Thus, strictly speaking, the DP4 values must be computed at the originally proposed method, unless the $[\sigma,\nu]$ are known for any other alternative method (to the best of our knowledge, this was only done by Paton and co-workers in their assignment of laurefurenynes A and B, vide infra).[14af] Naturally, this does not preclude the conclusions arrived in all studies in which more accurate NMR procedures were employed, though the DP4 values might have been different.

On the other hand, in 69% of the cases the DP4 probabilities were based on both ^{1}H and ^{13}C NMR data (as recommended by the DP4 developers).[12c,14a–g,j,l–n,p,u,v,y–ac,ae,af,ah–an,ap,aq,at–ay] In the remaining cases, the ^{13}C NMR data was strongly preferred to compute the DP4 probability (27%), finding only 2 examples (4%) that were built using ^{1}H NMR data.[14h,ao]

Despite the fact that DP4 can be performed with as many candidates as possible, it is a common practice to discard alternatives from relevant experimental NMR data. By doing this, the computational cost is reduced and the probability to get the correct isomer is increased. In fact, 71% of papers published that reported the use of DP4 have been carried out following this approach.[12c,i,14a–y,ao–ay] Another interesting fact is that in 28% of the cases the predictions made by DP4 were actually validated experimentally (mainly by total synthesis of the most likely candidate),[14z,aa,ac,af,ag,al,am,ao,ap,aw,ay,16] indicating a clear trend for the scientific community to accept stereochemical assignments based on theoretical calculations.

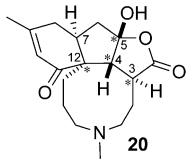
To sum up, we have found wide dispersion regarding the cases in which DP4 was used, both considering the molecular complexity, level of theory, and strategy approach.

All isomers

Including all isomers in the DP4 calculation procedure has pros and cons. The most obvious disadvantage is the increase in the CPU time and data processing, which are linearly dependent on the number of selected candidate structures. On the other hand, this minimizes the risk of removing the correct structure from the candidate list by misinterpretation of the experimental data (which is precisely the main source of structural misassignments). Therefore, unless the experimental data available is conclusive to unequivocally remove some structures, computing the DP4 probability with all candidates is the most recommended procedure. Fourrepresentative examples of natural products that have been assigned following this approach are given below.

Cernupalhine A

Yang, Zhao, and co-workers have devoted their recent research to the isolation and synthesis of *Lycopodium* alkaloids, mainly due to the interesting molecular architectures and biological activities. In 2014, they isolated from *Palhinhaea cernua L.*, a new alkaloid bearing a $C_{17}N$ skeleton with an unusual hydroxydihidrofuranone motif, and named it cernupalhine A.[14aa] Standard NMR experiments allowed the determination of the planar structure. On the basis of the proposed biosynthesis and ROESY experiments the authors suggested a plausible stereochemistry at C7 and C12, though could not unequivocally extend assignments to the rest of the molecule due to signal overlapping problems. Moreover, mainly due to the low amount of sample (0.7 mg), single crystals for X-ray experiments could not be obtained. Asymmetric total synthesis of cernupalhine A was considered to provide material for further biological essays and to unequivocally determine its structure. To decide which isomer to synthesize first, they carried out NMR calculations of all plausible isomers (16) at the PCM-mPW1PW91/6-311+G(2d,p)//B3LYP/6-31+G(d,p) level of theory. Once the shifts were computed and scaled for all candidates, DP4 analysis pointed to the 3*S**,4*R**,5*S**,7*S**,12*S** structure (**20**, Figure 10) as the most likely, with probabilities of 99.98%, 100%, and 100% for the ^{1}H, ^{13}C and the combination of both data, respectively. Starting from an advanced intermediate, and after 10 synthetic steps, they obtained the target 3*S*,4*R*,5*S*,7*S*,12*S* isomer **20** in an enatiomerically pure form. All spectroscopic data was identical to those obtained for the nat-



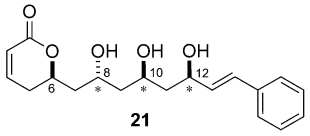
cernupalhine A
16 candidates
DP4: >99% (all data)
verified by total synthesis

Figure 10. Structure found for cernupalhine A. The carbon atoms that were varied to generate the candidate isomers for DP4 calculations are marked with an asterisk. The NMR calculations were carried out at the PCM/mPW1PW91/6-311+G(2d,p)//B3LYP/6-31+G(d,p) level.

ural product, therefore confirming the tridimensional structure.[14aa]

Cryptomoscatone E3

Cryptomoscatone E3 was isolated by Cavalheiro and Yoshida from the bark of the Brazilian tree *Cryptocarya mandiocanna*.[17] From circular dichroism experiments, the configuration at C6 was set as *R*, though the remaining three stereocenters could not be unequivocally assessed. In an effort to solve this stereochemical issue, Pilli and co-workers carried out an in silico guided, total synthesis approach. Thus, the DP4 probabilities of the resulting eight possible isomers were computed at the B3LYP/6-31G(d,p)//MMFF. The combined probabilities (^{1}H and ^{13}C data) suggested that 6*R*,8*R*,10*R*,12*R* (74%) or 6*R*,8*S*,10*S*,12*S* (24%) should be the correct isomer. Interestingly, both feature the opposite stereochemistry at C8, C10, and C12, indicating that an *anti–syn* arrangement in the stereotriad. To increase the confidence in the isomer to be synthesized, the chemical shifts calculations were refined at the B3LYP/6-31G(d,p)//B3LYP/6-31G(d) level, and the resulting DP4 probabilities strongly supported the all-*R* isomer (**21**) with high confidence (99%). With this guidance, the selected stereoisomer (compound **21**, Figure 11) was finally prepared in 14 steps and 9% overall yield. Interestingly, the spectroscopic data of the synthetic sample nicely matched those reported for the natural cryptomoscatone E3, highlighting the success of the computational predictions.[14al]

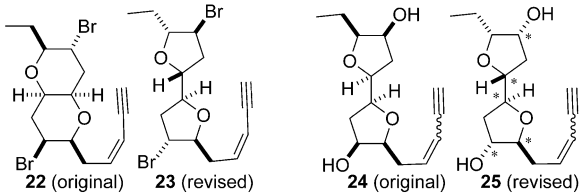


cryptomoscatone E3
8 candidates
DP4: 74% (all data, method A)
DP4: 99% (all data, method B)
verified by total synthesis

Figure 11. Structure found for cryptomoscatone E3. The carbon atoms that were varied to generate the candidate isomers for DP4 calculations are marked with an asterisk. The NMR calculations were carried out at the B3LYP/6-31G(d,p)//MMFF (method A) and B3LYP/6-31G(d,p)//B3LYP/6-31G(d) (method B) levels.

Laurefurenynes A and B

2,2′-Bifuranyl-containing natural products have been widely studied recently, both synthetically and computationally. The case of elatenyne is particularly interesting, as for long time it



32 candidates
DP4: 54% (C data)
DP4: 52% (H data)
verified by total synthesis

Figure 12. Original and revised structures of elatenyne and laurefurenynes A and B. The carbon atoms that were varied to generate the candidate isomers for DP4 calculations are marked with an asterisk. The NMR calculations were carried out at the mPW1PW91/6-311G(d,p)//wB97XD/6-31G(d) level.

was thought to contain a pyrano[3,2-*b*]pyran system (compound **22**, Figure 12).[18] In 2006, the Burton group proved by total synthesis that such arrangement did not correspond to the natural product, suggesting the 2,2′-bifuranyl alternative present in **23**.[19] Moreover, two years later and in collaboration with the Goodman group, they predicted the most likely stereochemistry of elatenyne (compound **23**) from extensive computational studies (perhaps one of the most important studies in the pre-DP4 era).[20] Finally, in 2012 the groups of Burton and Kim accomplished the total synthesis of **23**, confirming the computer-predicted stereostructure for elatenyne,[21] the absolute configuration of which was recently determined by the crystalline sponge method.[22]

In 2013, the groups of Paton and Burton undertook a similar approach on laurefurenynes A and B,[14af] isolated three years earlier by Jaspars and co-workers from *Laurencia* spp.[23] After preparation of three model 2,2′-bifuranyls and comparison of the corresponding ^{13}C NMR spectra with those of the natural product, the authors questioned the relative stereochemistry proposed by the isolation team on the basis of NOESY experiments (compound **24**, Figure 12). In order to determine the most likely stereostructure, Paton and co-workers undertook a DP4 analysis including all the plausible stereoisomers (32) at the mPW1PW91/6-311G(d,p)//wB97XD/6-31G(d) level. In an interesting discussion, the authors analyze the advantages of optimizing the geometries at higher levels, against the use of MMFF, and suggesting than whenever possible the optimization at higher levels is recommended, since more precise results may be obtained. It is also worth noting that in this case the authors realized that the key [σ,ν] DP4 terms might depend upon the level of theory used to perform the GIAO calculations, and consequently they recomputed those terms at the selected level using a database composed of 113 rigid molecules. All in all, the most likely candidate for laurefurenyne B (compound **25**) was efficiently synthesized, and its NMR data nicely agreed with the original data of the natural product.[14af]

Madeirolide A

The madeirolide family of macrolides was first reported by Wright and Winder.[14au] Among several analogues, madeiroli-

de A was shown to be a potent inhibitor of the fungal pathogen *Candida albicans*. Interested in the promising biological activities and molecular complexity, in 2013 the Paterson group reported the stereocontrolled synthesis of a fully elaborated C1–C11 subunit (western fragment) of madeirolide A conveniently functionalized to couple with the eastern part (not yet synthesized) to install the 24-membered macrolactone.[14au] Considering the synthetic challenge of this endeavor, the authors prudently took special care on the structural proposal before starting the synthetic planning. Assuming that the relative stereochemistry of each six-membered ring was right (from which NOE experiments are usually accurate), thus considering each ring as a single variable, 128 plausible diastereoisomers were taken into consideration. After performing the DP4 analysis at the B3LYP/6-31G(d,p)//MMFF level of theory, they found a strong preference (99.6%) towards the 9-*epi* isomer (compound **27**, Figure 13) of the initially proposed

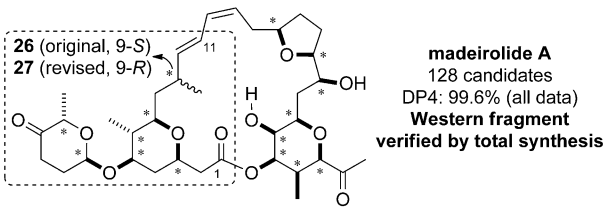


Figure 13. Original and revised structures of madeirolide A. The carbon atoms that were varied to generate the candidate isomers for DP4 calculations are marked with an asterisk. The NMR calculations were carried out at the B3LYP/6-31G(d,p)//MMFF level.

structure (compound **26**, Figure 13). Alarmed by this outcome, Paterson contacted Wright and Winder, who recognized having correctly assigned maiderolide A at C9 (as 9-*R*), but accidentally inverted to 9-*S* in the 2D-drawing. With the stereochemistry of the re-assigned madeirolide A, Paterson et al. successfully achieved the synthesis of the western fragment, finding a highly convincing homology for most of the ^{1}H and ^{13}C data between the synthesized molecule and the corresponding portion of the isolated natural product.

Pre-selected isomers

When the experimental data available (NOE, *J* couplings, etc.) provides conclusive evidence in favor to some particular stereoarrangement, the remaining candidates can be safely removed from the DP4 analysis. By doing this, the overall computational cost is reduced, increasing also the probability to identify the correct isomer. Actually, this is the most employed strategy (71% of the cases) in the stereoassigment of organic molecules with DP4.[12c,i, 14a–y,ao–ay] Six representative examples are given below of natural products the structures of which have been proposed or revised following this strategy.

Within this category is worth noting to mention the case of molecules containing stereoclusters. When two stereoarrangements are separated (and, therefore, contain weak interactions), the determination of the relative configuration between them is a challenging task. Often, the experimental data is not enough to unequivocally link one stereocluster to another. In these cases, the DP4 calculations can be performed on each separated regions using simplified computational models (recently defined as DP4f by Goodman, Paterson and coworkers),[14ay] or directly modeling the entire compound.

Gambierone

Gambierone, a ladder-shaped polyether from the dinoflagellate *Gambierdiscus belizeanus*, was recently isolated by Thomas, Botana, and co-workers. The unique and complex planar structure was proposed after extensive NMR analysis, though the ROESY experiments were not conclusive to fully assign the relative stereochemistry at C2, C4, and C38 (Figure 14). Moreover,

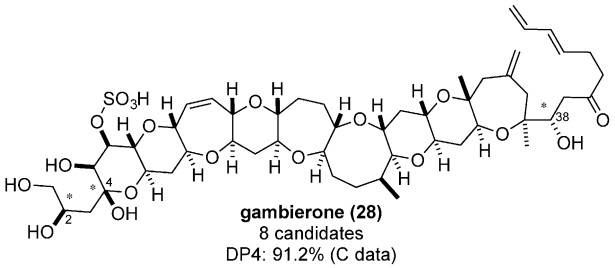


Figure 14. Proposed structure for gambierone. The carbon atoms that were varied to generate the candidate isomers for DP4 calculations are marked with an asterisk. The NMR calculations were carried out at the B3LYP/6-31G(d)//B3LYP/STO-3G level.

the low amount of sample precluded further chemical derivatization experiments to settle this issue. For that reason, the authors carried out NMR calculations of the eight resulting candidates at the B3LYP/6-31G(d)//B3LYP/STO-3G level. The DP4 probability computed from ^{13}C data strongly supported the 2*R**,4*R**,38*S** stereochemistry present in **28** (91.2%) that was finally proposed for this complex natural product.[14q]

Nobilisitine A

In 1999, Evidente and co-workers reported the isolation and characterization of nobilisitine A, an alkaloid from *Clivia nobilis*, and prosposed the structure **29** on the basis of standard NMR experiments.[24] In 2010, Banwell and co-workers concluded the total synthesis of *ent*-**29** and found significant deviations in the NMR data with those reported for the natural product.[25] Since the synthesized structure was confirmed by X-ray crystallography analysis, it seemed evident that the originally proposed structure was wrong. The close similarity of most the ^{1}H and ^{13}C NMR shifts reported by Evidente with those observed for the synthesized compound led the Banwell team to proposed that the source of error was stereochemical.[25] Nevertheless, the correct stereostructure of nobilisitine A remained unsolved until Tantillo and co-workers used NMR shift calculations to settle the controversy.[14l] Therefore, after assuming a *cis* ring fusion for the C/D rings, they computed the NMR shifts of all

remaining eight plausible isomers at the mPW1PW91/6-311+G(2d,p)//B3LYP/6-31+G(d,p) level, previously validated in the same work with known analogues. Based on standard statistics parameters of correlation (MAE, maxerr, etc.) they identified **30** (Figure 15) as the most likely structure. The assignment was strengthened by DP4 calculations, affording **30** as the most likely structure in high probability. Following Tantillo's proposal, the same year Banwell and co-workers synthesized *ent*-**30** in three steps from (+)-clividine, and found excellent match in the spectral data with those reported for the natural compound.[16c]

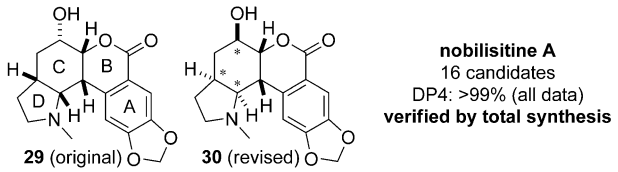


Figure 15. Originally proposed and revised structures of nobilisitine A. The carbon atoms that were varied to generate the candidate isomers for DP4 calculations are marked with an asterisk. The NMR calculations were carried out at the mPW1PW91/6-311+G(2d,p)//B3LYP/6-31+G(d,p) level.

Leiodermatolide

Leiodermatolide was isolated in 2011 by Paterson et al., from the marine sponge *Leiodematium sp.*, and showed a potent and selective antimitotic activity (IC50 <10 nM) againt a range of human cancer cell lines by inducing G2M cell arrest.[14am] After standard NMR experiments, the unpredecented 16-membered macrolide skeleton of this promising anticancer drug was revealed. From a *J*-based configurational analysis (JBA) and extensive NOE experiments, the relative stereochemistry of each main stereocluster (C1-C16 macrocycle and C20–C33 δ-lactone) was separately established. However, the experimental data was insufficient to unambiguously define the relative configuration between both stereoclusters, which was arbitrarily assigned as **31** (Figure 16). To reinforce the stereochemical analysis, a DP4 analysis was carried out at the originally proposed level of theory (B3LYP/6-31G(d,p)//MMFF) considering all plausible candidates (32 for the macrocycle and four for the six-membered lactone). Interestingly, while the obtained values strongly supported the proposed tridimensional structure for each region (>99% probability), the relative stereochemistry within the stereoclusters could not be unraveled.[14am]

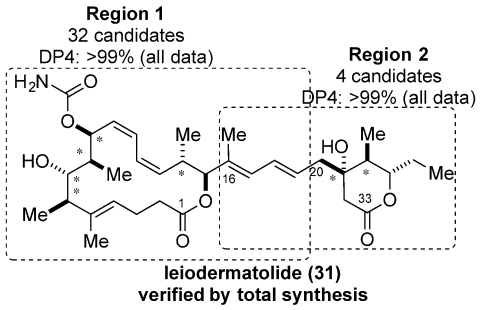


Figure 16. Proposed structure for leiodermatolide. The carbon atoms that were varied to generate the candidate isomers for DP4 calculations are marked with an asterisk. The NMR calculations were carried out at the B3LYP/6-31G(d,p)//MMFF level.

Finally, in 2012 the Fürstner group accomplished the first total synthesis of leiodermatolide by using an elegant ring-closing alkyne metathesis approach.[16d] To indisputably assign the correct configuration of the natural product, the authors synthesize the two possible stereoclusters (compound **31** and its corresponding isomer with all stereochemistries inverted in the δ-lactone moiety). Interestingly, they found that both diastereoisomers were almost indistinguishable from their ^{13}C NMR data, and only subtle differences in the ^{1}H NMR spectra could be observed to indisputably assign **31** as the correct leiodermatolide.[16d] In the meantime, the Paterson group reported their highly stereocontrolled total synthesis of leioermatolide employing a convergent palladium-mediated fragment assembly and macrolactonization sequence, providing further experimental confirmation of the natural product structure.[16a]

Belizentrin

Belizentrin was isolated in 2014 by Fernández, Daranas, and co-workers from the methanol extract of cell pellets of *Prorocentrum belizeanum*, a marine dinoflagellate originally collected from a coral reef of La Reunion Island (Indian Ocean, France).[14aq] This compound showed interesting pharmacological activity with cerebellar cells, causing important changes in the neuronal network integrity at nanomolar concentrations. Once the planar structure was determined by standard 1D and 2D NMR experiments, the key stereochemical features of Belizentrin (including the geometries of the double bonds, and the relative stereochemistry of all stereocenters at rings A, B and C) were determined from ROE and ROESY correlations. The next stage was to settle the configuration of the 25-membered macrocycle by connecting the relative stereochemistry of ring C with the remote C19 stereocenter. To solve this task, they carried out NMR calculations on a simplified C1–C15 truncated model, fixing the stereochemistry at ring C as 25*R**,27*S**,28*R**,29*S** and modeling the two possible stereoisomers with different configuration at C19: 19*R** and 19*S**. The most significant conformers were located using a conformational search sampling using 5000 cycles of molecular dynamics (using the OPLS05 force field) annelating at 1000 K, followed by 5000 cycles of large-scale low mode search steps, and the most stable rotamers were further optimized at the B3LYP/6-31+G(d,p) level. The most relevant geometries of each set were correlated with the ROE and *J* values, and the computed structure for the 19*S** isomer nicely matched with the experimental data. This was also supported by DP4 analysis at the B3LYP/6-31+G(d,p)//MMFF level, where the 19*S**,25*R**,27*S**,28*R**,29*S** isomer was predicted as the correct one with >99% probability over the 19*R**,25*R**,27*S**,28*R**,29*S** one. Determination of the stereochemistry of the side chain (C1–C18) was more challenging. The relative configuration of C3–C7 and C12–C15 had been determined on the basis of ex-

perimental NMR data, therefore the final goal was to connect both via the four-carbon acyclic tether (C8–C11), which also includes three contiguous stereocenters. From the Kishi universal database analysis they proposed an all-*syn* or all-*anti* relationships between the oxygenated substituents at this part of the molecule, though could not entirely settle the problem. Thus, the 16 possible stereoisomers (with the known relative stereochemistries of rings A and B) were computed at the HF/3-21G and the NMRs computed at the mPW1PW91/6-31+G(d,p) level. DP4 analysis showed that the 3*R**,4*S**,5*R**,6*R**,7*R**,9*S**,10*R**,11*S**,12*S**,14*S**,15*R** isomer was the correct one with high confidence (90.1%), with an all-*syn* relationship between C9–C12 oxygenated substituents, in good agreement with Kishi's analysis. Finally, the relative configurations of the macrolactam and the side chain stereoclusters were arbitrarily arbitrarily settled as **32** (Figure 17).

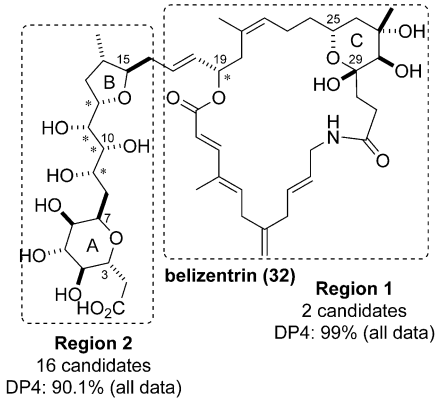


Figure 17. Proposed structure for belizentrin. The carbon atoms that were varied to generate the candidate isomers for DP4 calculations are marked with an asterisk. The NMR calculations were carried out at the B3LYP/6-31+G(d,p)//MMFF level (Region 1) and B3LYP/6-31+G(d,p)//HF/3-31G (Region 2) levels of theory.

Nivariol

Regarding the stereoclusters problem, perhaps one of the most significant examples was presented in 2011 by Daranas and co-workers in the isolation and assignment of nivariol, a new oxasqualenoid from a Macaronesian endemic species of *Laurencia viridis*.[14e] This interesting compound presents two well-separated stereoclusters (C1–C15 and C18–C24) and within the former, three aditional stereoclusters: C3–C6, C7–C10, and C11–C14, Figure 18. The elucidation of this complex molecule was not easy. Initially by using a 2D NMR analysis the authors unraveled the planar structure. The bicyclic C18–C24 stereocluster was confidently assigned as 18*R**,19*S**,22*R** from ROESY experiments, though the relative configurations within the C1–C15 region could not be unequivocally established. Therefore, mixing JBA with the information provided by synthesized oxasqualenoid fragments, the relative configurations for C3–C10 and C11–C15 were proposed. Finally, the connection between them with the C18–C24 portion was done computationally using DP4 analysis. In this fashion, four different possibilities were proposed and the corresponding NMR shifts

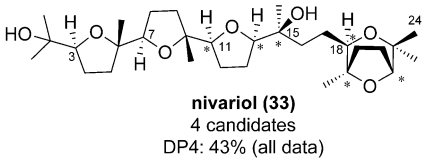


Figure 18. Proposed structure of nivariol. The carbon atoms that were varied to generate the candidate isomers for DP4 calculations are marked with an asterisk. The NMR calculations were carried out at the mPW1PW91/6-31G(d,p)//mPW1PW91/6-31G(d) level.

were computed at the mPW1PW91/6-31G(d,p)//mPW1PW91/6-31G(d) level of theory in $CHCl_3$. While the relative stereochemistry of the C3–C11 stereocluster could be established (since over 80% probability was found for the two candidates with the same configurations at those centers), the linking with the C18–C24 fragment required a deeper analysis. Finally, combining JBA, conformational analysis from DFT calculations, NOE correlations, and biosinthetic studies, the authors proposed the 3*S*,6*S*,7*R*,10*R*,11*S*,14*R*,15*S*,18*R*,19*S*,22*R* stereochemistry shown for compound **33** (Figure 18).

Mandelalide A

Mandelalide A represents not only a relevant and illustrative case of the role of total synthesis in solving structural problems, but also an emblematic example of the modest performance made by DP4 in some difficult scenarios. This complex natural product was isolated in 2012 from new species of *Lissoclin* by McPhail and co-workers, which after detailed and extensive NMR experiments, along with chemical derivatization analysis, proposed the structure **34** depicted in Figure 19.[26] Enticed by the remarkable anticancer activity and molecular complexity, several synthetic groups focused their efforts towards the total synthesis of mandelalide A—the group of Fürstner was the first to achieve this challenging goal in 2014. Regrettably, the NMR spectra of synthetic and natural mandelalide A did not perfectly match, demonstrating that, once again, a natural product had been incorrectly assigned.[27]

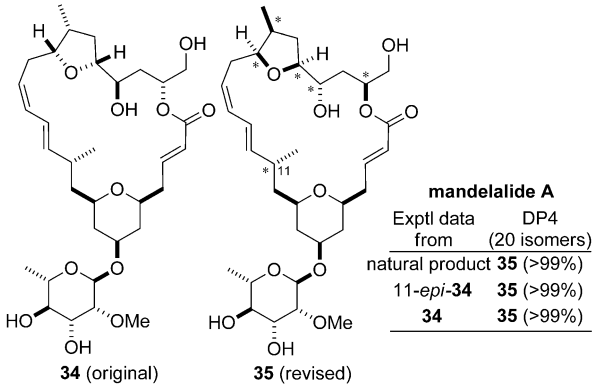


Figure 19. Originally proposed and revised structures of mandelalide A. The carbon atoms that were varied to generate the candidate isomers for DP4 calculations are marked with an asterisk. The NMR calculations were carried out at the B3LYP/6-31G(d,p)//CHARMM level.

Beyond frustration *"…although frustrating per se, these findings showcase that total synthesis remains an indispensable branch of natural product chemistry…"*, Fürstner and co-workers decided to find the source of error, suspecting that it was on the northern side of the molecule. To narrow down the number of possible candidates to synthesize, the authors computed the NMR shifts of 20 conceivable isomers at the B3LYP/6-31G(d,p)//CHARMM level, and correlated them by means of DP4 calculations with the experimental data corresponding to the natural, and two synthetic isomers (**34** and 11-*epi*-**34**). Interestingly, in all cases DP4 strongly supported isomer **35** with high probabilities (ca. 100%), which was wrong for the last two correlations (that were precisely included to provide confidence on the DP4 assignment).[14aw] The fact that in the end **35** turned out to represent the natural product is anecdotal, the inconsistency of the results left the authors no other choice than sorting out the mandelalide A issue by standard trial and error procedures. Then, after the total synthesis of three isomers, they found that the NMR data of **35** agreed satisfactorily well with the natural product.[14aw] In the meantime, the Xu, Ye, and co-workers, independently reached the same conclusion by finding that the upper side of the molecule was the source of the problem, with all the corresponding stereocenters inverted.[28]

DP4+

Considering the enormous challenge of correlating only one set of experimental data to closely related isomeric compounds, it should not in any way be surprising the modest performance exhibited by DP4 in some cases by pointing towards the incorrect isomer or affording inconsistent or unreliable results (see the mandelalide A case). In fact, in the seminal publication the authors included several examples that were not successfully predicted by DP4.[8]

In order to improve the performance of the method, Sarotti and co-workers recently introduced a modified probability (DP4+) by challenging two main arguments of the original formulation: the level of theory at which the method was developed and the exclusive use of scaled shifts.[15] Regarding the first issue, the developers of DP4 used B3LYP/6-31G(d,p)//MMFF as a useful alternative to provide reasonably good chemical shifts predictions at low computational cost (mainly, avoiding expensive ab initio or DFT methods in the geometry optimization step). However, as it often occurs in many other scientific disciplines, faster is not generally better. Using more refined geometries (for instance, B3LYP/6-31G(d)) and better treatments for the GIAO NMR calculations (inclusion of solvent, use of triple-ζ basis sets, etc.) can significantly improve the accuracy of the overall NMR prediction step. On the other hand, scaling the calculated shifts is commonly used to remove systematic errors. Therefore, the scaled shifts (δ_s) are closer to the experimental values, and the resulting errors follow a normally distributed series (a cardinal feature of DP4). In other words, the magnitude of an error becomes independent from the chemical environment (for instance, hybridization). Not only is this is not generally the case, but linear scaling can increase the risk of false positives by allowing incorrect structures to afford a fortuitously better fitting with the experimental data than the correct structure. Hence, adding unscaled shifts to the DP4 equation might reduce this potential problem, emphasizing the environmental contrast between the candidates.

Taking into account these issues, a modified DP4 probability (DP4+) was defined as Equation (4), in which the probability that candidate *i* (out of *m* isomers) is correct ($P(i)$) is a function of the corresponding probabilities computed using scaled and unscaled shifts. Each term (*s*DP4+ and *u*DP4+, respectively) follows the mathematical logic of DP4 [see Eq. (3)].

$$P(i) = \frac{\prod_{k=1}^{N}[1-T^{\nu}_{s}(e^{i}_{s,k}/\sigma_s)][1-T^{\nu}_{u\text{-}spx}((e^{i}_{u,k}-\mu_{u\text{-}spx})/\sigma_{u\text{-}spx})]}{\sum_{j=1}^{m}\prod_{k=1}^{N}[1-T^{\nu}_{s}(e^{j}_{s,k}/\sigma_s)][1-T^{\nu}_{u\text{-}spx}((e^{j}_{u,k}-\mu_{u\text{-}spx})/\sigma_{u\text{-}spx})]} \quad (4)$$

DP4+

*s*DP4+ (scaled shifts contribution): $\frac{\prod_{k=1}^{N}[1-T^{\nu}_{s}(e^{i}_{s,k}/\sigma_s)]}{\sum_{j=1}^{m}\prod_{k=1}^{N}[1-T^{\nu}_{s}(e^{j}_{s,k}/\sigma_s)]}$

*u*DP4+ (unscaled shifts contribution): $\frac{\prod_{k=1}^{N}[1-T^{\nu}_{u\text{-}spx}((e^{i}_{u,k}-\mu_{u\text{-}spx})/\sigma_{u\text{-}spx})]}{\sum_{j=1}^{m}\prod_{k=1}^{N}[1-T^{\nu}_{u\text{-}spx}((e^{j}_{u,k}-\mu_{u\text{-}spx})/\sigma_{u\text{-}spx})]}$

However, the second term bears two important modifications. First, the center of the distribution is no longer centered on zero ($\mu \neq 0$), and second, the series must be decomposed regarding the hybridization of the nuclei in question. This is because that it was found that even though the full set of unscaled errors did not follow a *t*-distributed series, each sp^2 and sp^3 series actually did. Therefore, the DP4+ probability was defined by 16 parameters (ν_s, σ_s, ν_{u-sp2}, μ_{u-sp2}, σ_{u-sp2}, ν_{u-sp3}, μ_{u-sp3} and σ_{u-sp3}) for the ^{13}C distributions, and the corresponding eight parameters for the ^{1}H series, which in turn depend upon the level of theory used to compute the NMR shifts.

Performance

By using a set of 48 challenging compounds from which DP4 afforded modest performance, the DP4+ was evaluated at 24 different levels of theory. In all cases, DP4+ performed better (up to 2.4 times) than DP4, with the PCM/mPW1PW91/6-31+G(d,p)//B3LYP/6-31G(d) level being the most recommended one. Moreover, it was found that better results were observed with triple-ζ or double-ζ polarized basis sets including diffuse functions, and including the solvent effect using PCM. Remarkably, such levels were found to afford a sharper *t* (lower σ) series for the ^{1}H error distributions, but not necessarily for the ^{13}C series, indicating a clear relationship between the accuracy of proton NMR prediction with the DP4+ performance. Another improvement of DP4+ was to reduce the known tendency of DP4 to overstate the probability when making incorrect assignments in high confidence.[8] In other words, unless the experimental and calculated data are conclusive towards a certain assignment, DP4+ does not advocate any specific option. These results clearly demonstrated that the level of theory employed to compute the NMR shifts play a key role in the probability distributions.[15] An insightful discussion of the role of

scaled and unscaled shifts, along with the type of nuclei (^{1}H or ^{13}C), in the DP4+ probability calculations was also made.[15]

The case of cryptomoscatone D1 and D2 (compounds **36** and **37**, isolated by Cavalheiro and Yoshida in 2002,[17] and revised by the Pilli group in 2014[29]) represents an illustrative example of the better performance of DP4+ over DP4. Interestingly, DP4+ correctly predicted the stereochemistry of cryptomoscatones D1 and D2 (Figure 20) in high confidence, while the original DP4 formulation significantly failed in both cases.

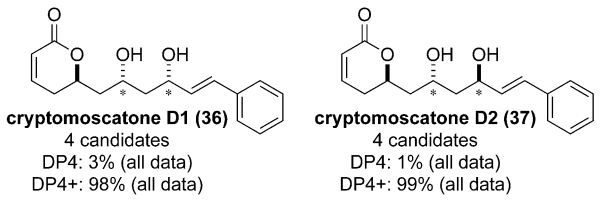


Figure 20. Revised structures of cryptomoscatones D1 and D2. The carbon atoms that were varied to generate the candidate isomers for DP4 calculations are marked with an asterisk. The NMR calculations were carried out at the B3LYP/6-31G(d,p)//MMFF (DP4) and PCM/mPW1PW91/6-31+G(d)//B3LYP/6-31G(d) (DP4+) levels.

"Toolbox"

To compute the DP4+ probability, an Excel file is available from the authors at sarotti-nmr.weebly.com. The user can select, from drop down lists, the level of theory at which the NMR data was computed (24 different options). Entering the experimental data and computed shielding tensors, the spreadsheet gives the DP4+ probability computed for each candidate. In a second sheet, this data can be also found disaggregated depending on the type of data (^{1}H, ^{13}C or all data; unscaled shifts and/or scaled shifts).

Confidence levels and probabilities—CP3, DP4 and DP4+

In all the methods discussed above (CP3, DP4, and DP4+) the probability computed for a given assignment reflects the certainty of such conclusion. Thus, despite the fact that there are no strictly defined cutoff limits, different general situations can be postulated according to the computed probability values. When such values are high (near 100%) or low (near 0%), these results clearly indicate that the data were probably well or incorrectly assigned, respectively. In less extreme situations, the confidence on the assignment lowers as the probability values move away from those limits. In fact, no firm statements can be done with probability values around 50%.

Finally, it should be also noted that several assumptions and simplifications are made to compute the probabilities (for example, that the errors are random and independent variables, and that the expectation values and standard deviations can be well approximated from a relatively small set of compounds). For these reasons, it should be wise to take the probability values as a rough guide to identify the most likely assignment, rather than giving them a more strict mathematical meaning.

ANN-PRA

Both CP3 and DP4 are "comparison-based" approaches, meaning that the decision making relies on the goodness of fit between experimental and calculated NMR data from a set of two or more candidates. The results are relative (candidate *i* is more likely to be correct than candidate *j*) and not absolute (candidate *i* is correct, candidate *j* is incorrect). This creates a circumstantial problem: the correct structure must be included in the candidate list, otherwise the methods will invariably point towards a false result. Moreover, they cannot be used to decide whether a given structural proposal is wrong or right, as the absence of any second alternative makes the comparison impossible to make. In 2013, Sarotti suggested a conceptually novel way to solve such structural validation problems on the basis of pattern recognition analysis (PRA) using artificial neural networks (ANNs), Figure 21.[9a] These mathematical

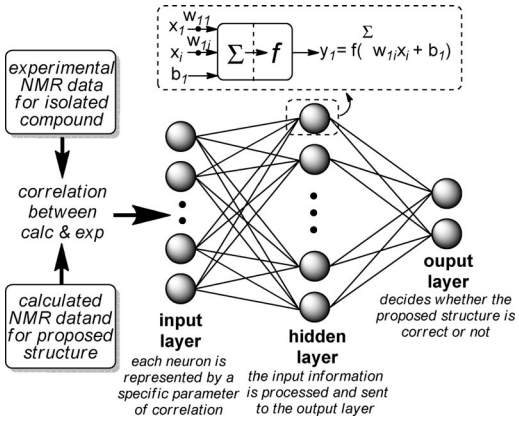


Figure 21. Schematic representation of a two-layer feed-forward ANN.

models are systems of data processing that were conceived on the grounds of biological brains, representing emblematic examples of artificial intelligence techniques.[30] The initial approach was conceived from the exclusive analysis of ^{13}C data. Thus, a set of statistical parameters computed from the correlation between the experimental and calculated NMR (using fast methods to perform the geometry optimization step, such as MM+, AM1 and HF/3-21G) for the proposed structure creates a footprint that is finally translated with a trained artificial neural network (ANN) to provide the solution (the structure is right or wrong). Using two-layer feed-forward network architectures with sigmoid transfer functions and scaled conjugate gradient back-propagation algorithms, the learning ability of the resulting ANNs was trained with 200 representative examples of correct and incorrect structures. Throughout the process, the weights between neurons synapses *w* (a measure of the strength of the connection) and the biases *b* values are determined iteratively to afford the optimal classification performance of the training set. In that study, 18 statistical de-

scriptors (that in turn defined the input layer) were evaluated: MAE, CMAE, σ, Cσ, MaxErr, CMaxErr, m, b, and R^2, both using TMS and MSTD[3m,n] as standard references. Several ANNs were built and trained using different sizes of input and hidden layers, and it was found that in all cases the best classification was achieved with the full set of descriptors (9 from TMS, and 9 from MSTD). The resulting trained ANNs from MM+, AM1 and HF/3-21G geometries (MM-18, AM1-18 and HF-18, respectively) were evaluated for 26 cases of natural product mis-assignments along with the corresponding revised structures. The high prediction capacity observed (100% with the best ANN-HF-18), along with the low amount of time required to perform the overall procedure, make this model a suitable test for rapid identification of structural misassignments. For example, the errors in the originally proposed structures of hexacyclinol and aquatolide (Figure 1), and the validation to support the revised structures could be successfully determined with trained ANNs, requiring only few minutes to perform the NMR calculations using standard CPUs.

The suitability of the ANN-PRA methodology in natural product chemistry can be well illustrated with the cases of dichomitol and paesslerin A (Figure 22). These natural products were

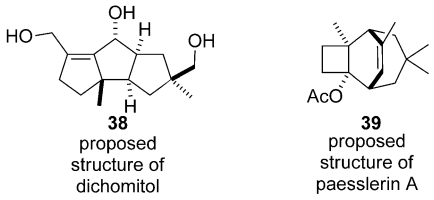


Figure 22. Proposed structures of dichomitol and paesslerin A, that were found incorrect by total synthesis.

isolated by the groups of Wei[31] and Palermo,[32] respectively, and they proposed the structures **38** and **39**, respectively, on the basis of extensive 1D and 2D NMR experiments. After total synthesis of the putative structures by Mehta[33] and Ihara,[34] respectively, significant differences in the NMR data between synthetic and natural products were observed, indicating that the originally proposed structures were wrong (in fact, up to date the real structures of both natural products remain unknown). In the eyes of the isolation teams, the structures **38** and **39** nicely accounted for the experimental data, meaning that there were not any other structural proposals (at least, not according to the original papers). Therefore, none of the comparison-based methods could have been helpful in preventing the publication of such structures. On the other hand, the ANN-PRA approach could have successfully detected an alarming mismatch between experimental and calculated data for each pair, indicating a failure in the structural elucidation process.

The proof-of-principle ANN-PRA approach, developed for ^{13}C NMR data, afforded very good results in detecting connectivity and regiochemical errors. However, low performance was achieved when dealing with stereochemical isomerism, much more subtle in terms of spectroscopic differentiation. To tackle this issue, in 2015 the same group expanded the ANN-PRA methodology to include ^{1}H and ^{1}H-^{13}C HSQC experiments. Following a conceptually similar strategy, several parameters of correlation between experimental and calculated monodimensional ^{1}H and ^{13}C data (MAE, CMAE, etc.) data were computed. In addition, 18 extra parameters were included to account for the goodness of fit between the experimental and predicted HSQC data. The NMR calculations were carried out at the mPW1PW91/6-31G(d) (gas phase) and mPW1PW91/6-31G(d,p) (PCM) levels, from B3LYP/6-31G(d) geometries. The use of a more demanding computational method for the geometry optimization step was justified on the basis of the concomitant increase in the accuracy of the predictions. Again, two different reference systems (TMS and MSTD) were used to extract the chemical shifts, leading to a total number of 72 descriptors. Using a training set of 208 examples (composed of known compounds with confidently assigned ^{13}C and HSQC spectra, and the corresponding incorrect structures built by introducing slight modifications to the former by inverting a stereocenter of changing the positions of few atoms), more than 400 ANNs were built and trained using different sizes of input and hidden layers. Three trained ANNs showed optimal results in terms of classification capacity: ANN-TMS$_{vac}$ (built using 36 parameters computed at the PCM/mPW1PW91/6-31G(d,p) level and TMS as reference standard), ANN-MSTD$_{sol}$ (built using 36 parameters computed at the mPW1PW91/6-31G(d,p) level and MSTD as reference standard), and ANN-mix (built using the 72 parameters described above). With these modifications, the new ANNs could detect errors in the stereochemical limit, affording very high (>92%) percentage of correct classification after the training. Finally, the performance of the method was successfully evaluated (up to 100% of correct prediction) in challenging real cases of misassignments (some prepresentative examples are given in Figure 23).[35] Interestingly, these trained ANNs could successfully detect the stereochemical mistake of putative mandelalide A (compound **34**, Figure 19). The best results were obtained with the ANN-mix network, which is recommended to obtain the most reliable outcomes.

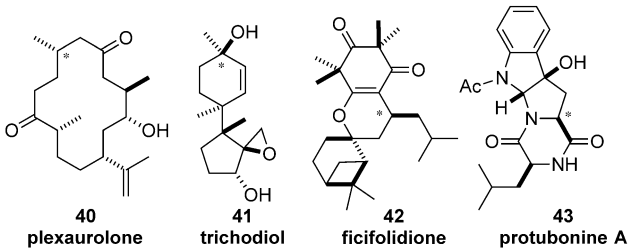


Figure 23. Revised structures of plexaurolone, trichodiol, ficifolidione and protubonine A. In the originally proposed structures, the carbon atoms marked with an asterisk had been assigned with the opposite configuration.

"Toolbox"

The ANN-PRA calculations are extremely tedious and difficult to compute "by hand". For that reason, unlike CP3 and DP4, in this case an additional software tool is essential. This was

solved by embodying all the mathematical equations that define the trained ANNs in simple Excel spreadsheets. Then, the user is asked to introduce the experimental data of the sample, and the computed shielding tensors of the candidate structure. After indicating which carbon (or proton attached to) is sp–sp^2 hybridized, the file computes the unscaled and scaled chemical shifts (using TMS and MSTD as reference standards), along with the corresponding statistical parameters of correlation, that in turn are used to feed the corresponding trained ANN to finally obtain a result (the candidate structure is right or wrong). It is important to point out that these ANNs have been developed using unassigned data (that is, it is not important to know which experimental NMR signal corresponds to which nuclei in the proposed structure), representing an additional simplification of the overall process.

Summary

As demonstrated above, the computational methods (CP3, DP4, and ANN-PRA) that combine affordable quantum chemical calculations with sophisticated data processing allow the structural or stereochemical assignment of complex organic compounds with remarkable levels of confidence. For that reason, they have already emerged as indisputable complements to the experimental NMR spectroscopy. In understanding the key role that these marvelous toolboxes have had in settling structural issues (many of which could only be solved by total synthesis), it could be predicted an even wider use in the near future.

Keywords: computer chemistry · NMR calculations · NMR spectroscopy · quantum chemistry · structure elucidation

[1] For previous reviews, see: a) A. Bagno, G. Saielli, *Wiley Interdiscip. Rev. Comput. Mols Sci.* **2015**, *5*, 228; b) D. J. Tantillo, *Nat. Prod. Rep.* **2013**, *30*, 1079; c) G. Bifulco, P. Dambruoso, L. Gomez-Paloma, R. Riccio, *Chem. Rev.* **2007**, *107*, 3744; d) P. H. Willoughby, M. J. Jansma, T. R. Hoye, *Nat. Protoc.* **2014**, *9*, 643; e) M. W. Lodewyk, M. R. Siebert, D. J. Tantillo, *Chem. Rev.* **2012**, *112*, 1839.

[2] For seminal references, see: a) A. Bagno, *Chem. Eur. J.* **2001**, *7*, 1652; b) A. Bagno, F. Rastrelli, G. Saielli, *Chem. Eur. J.* **2006**, *12*, 5514; c) A. Bagno, F. Rastrelli, G. Saielli, *J. Phys. Chem. A* **2003**, *107*, 9964; d) G. Barone, D. Duca, A. Silvestri, L. Gomez-Paloma, R. Riccio, G. Bifulco, *Chem. Eur. J.* **2002**, *8*, 3240; e) G. Barone, L. Gomez-Paloma, D. Duca, A. Silvestri, R. Riccio, G. Bifulco, *Chem. Eur. J.* **2002**, *8*, 3233.

[3] For leading references, see: a) Y. Li, *RSC Adv.* **2015**, *5*, 36858; b) E. V. Mercado-Marin, P. Garcia-Reynagal, S. Romminger, E. F. Pimenta, D. K. Rommey, M. .W. Lodewyk, D. E. Williams, R. J. Andersen, S. J. Miller, D. J. Tantillo, R. G. S. Berlinck, R. Sarpong, *Nature* **2014**, *509*, 318 ; c) M. W. Lodewyk, C. Soldi, P. B. Jones, M. M. Olmstead, J. Rita, J. T. Shaw, D. J. Tantillo, *J. Am. Chem. Soc.* **2012**, *134*, 18550; d) K. W. Quasdorf, A. D. Huters, M. W. Lodewyk, D. J. Tantillo, N. K. Garg, *J. Am. Chem. Soc.* **2012**, *134*, 1396; e) G. Saielli, K. C. Nicolaou, A. Ortiz, H. Zhang, A. Bagno, *J. Am. Chem. Soc.* **2011**, *133*, 6072; f) S. M. Koskowich, W. C. Johnson, R. S. Paley, P. R. Rablen, *J. Org. Chem.* **2008**, *73*, 3492; g) G. Hu, K. Liu, L. Williams, *J. Org. Lett.* **2008**, *10*, 5493; h) A. M. Belostotskii, *J. Org. Chem.* **2008**, *73*, 5723; i) S. D. Rychnovsky, *Org. Lett.* **2006**, *8*, 2895; j) R. Jain, T. Bally, P. R. Rablen, *J. Org. Chem.* **2009**, *74*, 4017; k) T. Bally, P. R. Rablen, *J. Org. Chem.* **2011**, *76*, 4818; l) K. G. Andrews, A. C. Spivey, *J. Org. Chem.* **2013**, *78*, 11302; m) A. M. Sarotti, S. C. Pellegrinet, *J. Org. Chem.* **2012**, *77*, 6059; n) A. M. Sarotti, S. C. Pellegrinet, *J. Org. Chem.* **2009**, *74*, 7254; o) C. Fattorusso, E. Stendardo, G. Appendino, E. Fattorusso, P. Luciano, A. Romano, O. Taglialatela-Scafati, *Org. Lett.* **2007**, *9*, 2377; p) D. C. Braddock, H. S. Rzepa, *J. Nat. Prod.* **2008**, *71*, 728; q) C. Timmons, P. Wipf, *J. Org. Chem.* **2008**, *73*, 9168; r) K. N. White, T. Amagata, A. G. Oliver, K. Tenney, P. J. Wenzel, P. Crews, *J. Org. Chem.* **2008**, *73*, 8719; s) J.-X. Pu, S.-X. Huang, J. Ren, W.-L. Xiao, L.-M. Li, R.-T. Li, L.-B. Li, T.-G. Liao, L.-G. Lou, H.-J. Zhu, H.-D. Sun, *J. Nat. Prod.* **2007**, *70*, 1706; t) E. Fattorusso, P. Luciano, A. Romano, O. Taglialatela-Scafati, G. Appendino, M. Borriello, C. Fattorusso, *J. Nat. Prod.* **2008**, *71*, 1988; u) A. G. Kutateladze, O. A. Mukhina, *J. Org. Chem.* **2014**, *79*, 8397; v) A. G. Kutateladze, O. A. Mukhina, *J. Org. Chem.* **2015**, *80*, 5218; w) A. G. Kutateladze, O. A. Mukhina, *J. Org. Chem.* **2015**, *80*, 10838.

[4] a) K. C. Nicolaou, S. A. Snyder, *Angew. Chem. Int. Ed.* **2005**, *44*, 1012; *Angew. Chem.* **2005**, *117*, 1036; b) T. L. Suyama, W. H. Gerwick, K. L. McPhail, *Bioorg. Med. Chem.* **2011**, *19*, 6675; c) M. E. Maier, *Nat. Prod. Rep.* **2009**, *26*, 1105.

[5] a) K. C. Nicolaou, A. A. Shah, H. Korman, T. Khan, L. Shi, W. Worawalai, E. A. Theodorakis, *Angew. Chem. Int. Ed.* **2015**, *54*, 9203; *Angew. Chem.* **2015**, *127*, 9335; b) L. Zhu, Y. Liu, R. Ma, R. Tong, *Angew. Chem. Int. Ed.* **2015**, *54*, 627; *Angew. Chem.* **2015**, *127*, 637; c) Q. Xiao, K. Young, K. Zakarian, *J. Am. Chem. Soc.* **2015**, *137*, 5907; d) K. C. Nicolaou, C. R. H. Hale, C. Nilewski, H. A. Ioannidou, A. ElMarrouni, L. G. Nilewski, K. Beabout, T. T. Wang, Y. Shamoo, *J. Am. Chem. Soc.* **2014**, *136*, 12137.

[6] The search was conducted using the Scopus database (date of search: 28/02/2016).

[7] S. G. Smith, J. M. Goodman, *J. Org. Chem.* **2009**, *74*, 4597.

[8] S. G. Smith, J. M. Goodman, *J. Am. Chem. Soc.* **2010**, *132*, 12946.

[9] a) A. M. Sarotti, *Org. Biomol. Chem.* **2013**, *11*, 4847; b) M. M. Zanardi, A. M. Sarotti, *J. Org. Chem.* **2015**, *80*, 9371.

[10] a) M. E. Elyashberg, K. A. Blinov, A. J. Williams, S. G. Molodtsov, G. E. Martin, E. R. Martirosian, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 771; b) M. E. Elyashberg, K. A. Blinov, A. J. Williams, E. R. Martirosian, S. G. Molodtsov, *J. Nat. Prod.* **2002**, *65*, 693; c) M. Elyashberg, A. J. Williams, K. Blinov, *Nat. Prod. Rep.* **2010**, *27*, 1296; d) M. Elyashberg, A. J. Williams, *Computer-Based Structure Elucidation from Spectral Data*, Springer, Berlin, **2015**; *Lect. Notes Chem.* **2015**, *89*; e) M. Elyashberg, A. Williams, G. Martin, *Prog. Nucl. Magn. Reson. Spectrosc.* **2008**, *53*, 1; f) M. Elyashberg, A. Williams, K. Blinov, *Contemporary Computer-Assisted Approaches to Molecular Structure Elucidation*, Royal Society of Chemistry, Cambridge, **2011**; g) M. Elyashberg, K. Blinov, S. Molodtsov, A. J. Williams, *J. Nat. Prod.* **2013**, *76*, 113; h) K. a. Blinov, D. Carlson, M. E. Elyashberg, G. E. Martin, E. R. Martirosian, S. Molodtsov, A. J. Williams, *Magn. Reson. Chem.* **2003**, *41*, 359.

[11] J. J. Poza, C. Jiménez, J. Rodríguez, *Eur. J. Org. Chem.* **2008**, 3960.

[12] a) F. Cen-Pacheco, A. J. Santiago-Benítez, C. García, S. J. Álvarez-Méndez, A. J. Martín-Rodríguez, M. Norte, V. S. Martín, J. a. Gavín, J. J. Fernández, A. H. Daranas, *J. Nat. Prod.* **2015**, *78*, 712; b) D. M. Hodgson, C. Villalonga-Barber, J. M. Goodman, S. C. Pellegrinet, *Org. Biomol. Chem.* **2010**, *8*, 3975; c) I. H. Hwang, J. Oh, W. Zhou, S. Park, J.-H. Kim, A. G. Chittiboyina, D. Ferreira, G. Y. Song, S. Oh, M. Na, *J. Nat. Prod.* **2015**, *78*, 453; d) S. G. Smith, J. A. Channon, I. Paterson, J. M. Goodman, *Tetrahedron* **2010**, *66*, 6437; e) E. De Gussem, W. Herrebout, S. Specklin, C. Meyer, J. Cossy, P. Bultinck, *Chem. Eur. J.* **2014**, *20*, 17385; f) D. F. Llompart, A. M. Sarotti, V. Corne, A. G. Suarez, R. A. Spanevello, G. A. Echeverria, O. E. Piro, E. E. Castellano, *Tetrahedron Lett.* **2014**, *55*, 2394; g) S. Qiu, E. De Gussem, K. Abbaspour Tehrani, S. Sergeyev, P. Bultinck, W. Herrebout, *J. Med. Chem.* **2013**, *56*, 8903; h) A. M. Sarotti, A. G. Suárez, R. A. Spanevello, *Tetrahedron Lett.* **2011**, *52*, 3116; i) T. D. Tran, N. B. Pham, R. J. Quinn, *Eur. J. Org. Chem.* **2014**, 4805; j) J. Zhang, C. Taylor, E. Bowman, L. Savage-Low, M. W. Lodewyk, L. Hanne, G. Wu, *Tetrahedron Lett.* **2013**, *54*, 6298; k) C. Talotta, C. Gaeta, M. De Rosa, J. R. Ascenso, P. M. Marcos, P. Neri, *Eur. J. Org. Chem.* **2016**, 158.

[13] M. S. Miftakhov, I. N. Gaisina, F. A. Valeev, *Russ. Chem. Bull.* **1996**, *45*, 1942.

[14] a) S. K. Amini, *Magn. Reson. Chem.* **2013**, *51*, 328; b) S. K. Amini, *J. Iran. Chem. Soc.* **2014**, *11*, 179; c) E. Cairns, M. A. Hashmi, A. J. Singh, G. Eakins, M. Lein, R. Keyzers, *J. Agric. Food Chem.* **2015**, *63*, 7421; d) F. Cen-Pacheco, M. Norte, J. J. Fernández, A. H. Daranas, *Org. Lett.* **2014**, *16*, 2880; e) F. Cen-Pacheco, J. Rodríguez, M. Norte, J. J. Fernández, A. H. Daranas, *Chem. Eur. J.* **2013**, *19*, 8525; f) V. L. Challinor, R. C. Johnston, P. V. Bernhardt, R. P. Lehmann, E. H. Krenske, J. J. De Voss, *Chem. Sci.* **2015**, *6*, 5740; g) G. Chianese, B.-B. Gu, F. Yang, W.-H. Jiao, Y.-W. Guo, H.-

W. Lin, O. Taglialatela-Scafati, *RSC Adv.* **2015**, *5*, 63372; h) J. Cho, S. Lee, S. Hwang, S. H. Kim, J. S. Kim, S. Kim, *Eur. J. Org. Chem.* **2013**, 4614; i) S. Di Micco, A. Zampella, M. V. D'Auria, C. Festa, S. De Marino, R. Riccio, C. P. Butts, G. Bifulco, *Beilstein J. Org. Chem.* **2013**, *9*, 2940; j) T. Iwai, T. Kubota, J. i. Kobayashi, *J. Nat. Prod.* **2014**, *77*, 1541; k) Q. Li, Y.-S. Xu, G. A. Ellis, T. S. Bugni, Y. Tang, R. P. Hsung, *Tetrahedron Lett.* **2013**, *54*, 5567; l) M. W. Lodewyk, D. J. Tantillo, *J. Nat. Prod.* **2011**, *74*, 1339; m) Y. Matsuo, K. Okuda, H. Morikawa, R. Oowatashi, Y. Saito, T. Tanaka, *J. Nat. Prod.* **2016**, *79*, 189; n) M. Menna, A. Aiello, F. D'Aniello, C. Imperatore, P. Luciano, R. Vitalone, C. Irace, R. Santamaria, *Eur. J. Org. Chem.* **2013**, 3241; o) A. E. Nugroho, M. Okuda, Y. Yamamoto, Y. Hirasawa, C.-P. Wong, T. Kaneda, O. Shirota, A. H. A. Hadi, H. Morita, *Tetrahedron* **2013**, *69*, 4139; p) M. Omar, Y. Matsuo, H. Maeda, Y. Saito, T. Tanaka, *Org. Lett.* **2014**, *16*, 1378; q) I. s. Rodríguez, G. g. Genta-Jouve, C. Alfonso, K. Calabro, E. Alonso, J. A. Sánchez, A. Alfonso, O. P. Thomas, L. M. Botana, *Org. Lett.* **2015**, *17*, 2392; r) J. Rodríguez, R. M. Nieto, M. Blanco, F. A. Valeriote, C. Jimenez, P. Crews, *Org. Lett.* **2014**, *16*, 464; s) N. Tanaka, M. Asai, T. Kusama, J. Fromont, J. i. Kobayashi, *Tetrahedron Lett.* **2015**, *56*, 1388; t) N. Tanaka, M. Takekata, S.-i. Kurimoto, K. Kawazoe, K. Murakami, D. Damdinjav, E. Dorjbal, Y. Kashiwada, *Tetrahedron Lett.* **2015**, *56*, 817; u) A. M. White, G. K. Pierens, L. C. Forster, A. E. Winters, K. L. Cheney, M. J. Garson, *J. Nat. Prod.* **2016**, *79*, 477; v) T. P. Wyche, Y. Hou, D. Braun, H. C. Cohen, M. P. Xiong, T. S. Bugni, *J. Org. Chem.* **2011**, *76*, 6542; w) T. P. Wyche, Y. Hou, E. Vazquez-Rivera, D. Braun, T. S. Bugni, *J. Nat. Prod.* **2012**, *75*, 735; x) T. P. Wyche, J. S. Piotrowski, Y. Hou, D. Braun, R. Deshpande, S. McIlwain, I. M. One, C. L. Myers, I. A. Guzei, W. M. Westler, D. R. Andes, T. S. Bugni, *Angew. Chem. Int. Ed.* **2014**, *53*, 11583; *Angew. Chem.* **2014**, *126*, 11767; y) L.-L. Zhu, W.-W. Fu, S. Watanabe, Y.-N. Shao, H.-S. Tan, H. Zhang, C.-H. Tan, Y.-F. Xiu, H. Norimoto, H.-X. Xu, *Planta Med.* **2014**, *80*, 1721; z) K. G. Andrews, C. S. Frampton, A. C. Spivey, *Acta Crystallogr. Sect. C* **2013**, *69*, 1207; aa) L.-B. Dong, Y.-N. Wu, S.-Z. Jiang, X.-D. Wu, J. He, Y.-R. Yang, Q.-S. Zhao, *Org. Lett.* **2014**, *16*, 2700; ab) S. Khokhar, Y. Feng, M. R. Campitelli, R. J. Quinn, J. N. A. Hooper, M. G. Ekins, R. A. Davis, *J. Nat. Prod.* **2013**, *76*, 2100; ac) M. Mirion, L. Andernach, C. Stobe, J. Barjau, D. Schollmeyer, T. Opatz, A. Lützen, S. R. Waldvogel, *Eur. J. Org. Chem.* **2015**, 4876; ad) M. J. Riveira, C. Gayathri, A. Navarro-Vázquez, N. V. Tsarevsky, R. R. Gil, M. P. Mischne, *Org. Biomol. Chem.* **2011**, *9*, 3170; ae) M. J. Riveira, P. Trigo-Mouriño, E. Troche-Pesqueira, G. E. Martin, A. Navarro-Vázquez, M. P. Mischne, R. R. Gil, *J. Org. Chem.* **2015**, *80*, 7396; af) D. J. Shepherd, P. A. Broadwith, B. S. Dyson, R. S. Paton, J. W. Burton, *Chem. Eur. J.* **2013**, *19*, 12644; ag) T. D. Tran, N. B. Pham, M. Ekins, J. N. A. Hooper, R. J. Quinn, *Mar. Drugs* **2015**, *13*, 4556; ah) M. M. Vaughan, Q. Wang, F. X. Webster, D. Kiemle, Y. J. Hong, D. J. Tantillo, R. M. Coates, A. T. Wray, W. Askew, C. O'Donnell, *Plant Cell* **2013**, *25*, 1108; ai) X. Wang, B. M. Duggan, T. F. Molinski, *J. Am. Chem. Soc.* **2015**, *137*, 12343; aj) L. Ye, B. Hu, F. El-Badri, B. M. Hudson, P.-W. Phuan, A. S. S. Verkman, D. J. Tantillo, M. J. Kurth, *Bioorg. Med. Chem. Lett.* **2014**, *24*, 5840; ak) C.-M. Yu, L. A. Calhoun, R. M. Konder, A. S. Grant, *Can. J. Chem.* **2014**, *92*, 406; al) L. F. T. Novaes, A. M. Sarotti, R. A. Pilli, *J. Org. Chem.* **2015**, *80*, 12027; am) I. Paterson, S. M. Dalby, J. C. Roberts, G. J. 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; *Angew. Chem.* **2011**, *123*, 3277; an) G. C. Resende, E. S. Alvarenga, P. H. Willoughby, *J. Mol. Struct.* **2015**, *1101*, 212; ao) S. G. Brown, M. J. Jansma, T. R. Hoye, *J. Nat. Prod.* **2012**, *75*, 1326; ap) H. J. Domínguez, G. D. Crespín, A. J. Santiago-Benítez, J. A. Gavín, M. Norte, J. J. Fernández, A. H. Daranas, *Mar. Drugs* **2014**, *12*, 176; aq) H. J. Domínguez, J. G. J. G. Napolitano, M. T. Fernández-Sánchez, D. Cabrera-García, A. Novelli, M. Norte, J. J. Fernández, A. H. Daranas, *Org. Lett.* **2014**, *16*, 4546; ar) A. Gutiérrez-Cepeda, A. H. Daranas, J. J. Fernández, M. Norte, M. L. Souto, *Mar. Drugs* **2014**, *12*, 4031; as) E. Mevers, F. P. J. Haeckl, P. D. Boudreau, T. Byrum, P. C. Dorrestein, F. A. Valeriote, W. H. Gerwick, *J. Nat. Prod.* **2014**, *77*, 969; at) R. B. Nazarski, B. Pasternak, S. Leśniak, *Tetrahedron* **2011**, *67*, 6901; au) I. Paterson, G. W. Haslett, *Org. Lett.* **2013**, *15*, 1338; av) A. L. Waters, J. Oh, A. R. Place, M. T. Hamann, *Angew. Chem. Int. Ed.* **2015**, *54*, 15705; aw) J. Willwacher, B. Heggen, C. Wirtz, W. Thiel, A. Fürstner, *Chem. Eur. J.* **2015**, *21*, 10416; ax) W. Zhou, J. Oh, W. Lee, S. Kwak, W. Li, A. G. Chittiboyina, D. Ferreira, M. T. Hamann, S. H. Lee, J.-S. Bae, *Biochim. Biophys. Acta Gen. Subj.* **2014**, *1840*, 2042; ay) C. I. MacGregor, B. Y. Han, J. M. Goodman, I. Paterson, *Chem. Commun.* **2016**, *52*, 4632.

[15] N. Grimblat, M. M. Zanardi, A. M. Sarotti, *J. Org. Chem.* **2015**, *80*, 12526.

[16] a) I. Paterson, K. K. H. Ng, S. Williams, D. C. Millican, S. M. Dalby, *Angew. Chem. Int. Ed.* **2014**, *53*, 2692; *Angew. Chem.* **2014**, *126*, 2730; b) M. R. Vippila, P. K. Ly, G. D. Cuny, *J. Nat. Prod.* **2015**, *78*, 2398; c) B. D. Schwartz, L. V. White, M. G. Banwell, A. C. Willis, *J. Org. Chem.* **2011**, *76*, 8560; d) J. Willwacher, N. Kausch-Busies, A. Fürstner, *Angew. Chem. Int. Ed.* **2012**, *51*, 12041; *Angew. Chem.* **2012**, *124*, 12207.

[17] A. J. Cavalheiro, M. Yoshida, *Phytochemistry* **2000**, *53*, 811.

[18] J. G. Hall, J. A. Reiss, *Aust. J. Chem.* **1986**, *39*, 1401.

[19] H. M. Sheldrake, C. Jamieson, J. W. Burton, *Angew. Chem. Int. Ed.* **2006**, *45*, 7199; *Angew. Chem.* **2006**, *118*, 7357.

[20] S. G. Smith, R. S. Paton, J. W. Burton, J. M. Goodman, *J. Org. Chem.* **2008**, *73*, 4053.

[21] B. S. Dyson, J. W. Burton, T.-i. Sohn, B. Kim, H. Bae, D. Kim, *J. Am. Chem. Soc.* **2012**, *134*, 11781.

[22] S. Urban, R. Brkljača, M. Hoshino, S. Lee, M. Fujita, *Angew. Chem. Int. Ed.* **2016**, *55*, 2678; *Angew. Chem.* **2016**, *128*, 2728.

[23] W. M. Abdel-Mageed, R. Ebel, F. A. Valeriote, M. Jaspars, *Tetrahedron* **2010**, *66*, 2855.

[24] A. A.-D. Evidente, A. H. Abou-Donia, F. A. Darwish, M. E. Amer, F. F. Kassem, H. A. Hammoda, A. Motta, *Phytochemistry* **1999**, *51*, 1151.

[25] B. D. Schwartz, M. T. Jones, M. G. Banwell, I. A. Cade, *Org. Lett.* **2010**, *12*, 5210.

[26] J. Sikorska, A. M. Hau, C. Anklin, S. Parker-Nance, M. T. Davies-Coleman, J. E. Ishmael, K. L. McPhail, *J. Org. Chem.* **2012**, *77*, 6066.

[27] J. Willwacher, A. Fürstner, *Angew. Chem. Int. Ed.* **2014**, *53*, 4217; *Angew. Chem.* **2014**, *126*, 4301.

[28] H. Lei, J. Yan, J. Yu, Y. Liu, Z. Wang, Z. Xu, T. Ye, *Angew. Chem. Int. Ed.* **2014**, *53*, 6533; *Angew. Chem.* **2014**, *126*, 6651.

[29] L. F. Toneto Novaes, R. L. Drekener, C. M. Avila, R. A. Pilli, *Tetrahedron* **2014**, *70*, 6467.

[30] J. Zupan, J. Gasteiger, *Neural Networks in Chemistry and Drug Design*, Wiley-VCH, Weinheim, **1999**.

[31] Z. Huang, Y. Dan, Y. Huang, L. Lin, T. Li, W. Ye, X. Wei, *J. Nat. Prod.* **2004**, *67*, 2121.

[32] M. F. Rodríguez Brasco, A. M. Seldes, J. A. Palermo, *Org. Lett.* **2001**, *3*, 1415.

[33] G. Mehta, K. Pallavi, *Tetrahedron Lett.* **2006**, *47*, 8355.

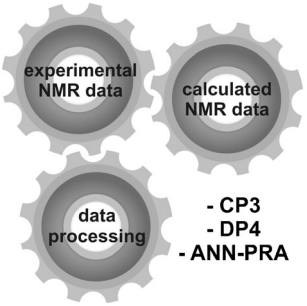
[34] K. Inanaga, K. Takasu, M. Ihara, *J. Am. Chem. Soc.* **2004**, *126*, 1352.

[35] Plexaurolone original: a) S. E. Ealick, D. Van der Helm, R. A. Gross, A. J. Weinheimer, L. S. Ciereszko, R. E. Middlebrook, *Acta Crystallogr. Sect. B* **1980**, *36*, 1901; plexaurolone revised: b) E. Tello, L. Castellanos, C. Arevalo-Ferro, J. Rodríguez, C. Jiménez, C. Duque, *Tetrahedron* **2011**, *67*, 9112; trichodiol original: c) S. Nozoe, Y. Machida, *Tetrahedron* **1972**, *28*, 5105; trichodiol revised: d) A. R. Hesketh, B. W. Bycroft, P. M. Dewick, J. Gilbert, *Phytochemistry* **1992**, *32*, 105; ficifolidione original: e) Q. Cheng, J. K. Snyder, *J. Org. Chem.* **1988**, *53*, 4562; ficifolidione revised: f) H. Nishiwaki, S. Fujiwara, T. Wukirsari, H. Iwamoto, S. Mori, K. Nishi, Y. Shuto, *J. Nat. Prod.* **2015**, *78*, 43; protubonine A original: g) S. U. Lee, Y. Asami, D. Lee, J. H. Jang, J. S. Ahn, H. Oh, *J. Nat. Prod.* **2011**, *74*, 1284; protubonine A revised: h) P. Lorenzo, R. Álvarez, Á. R. de Lera, *Eur. J. Org. Chem.* **2014**, 2557.

Received: March 10, 2016
Published online on ■■ ■■, 0000

MINIREVIEW

Help! The structural and stereochemical assignment of organic compounds is often a hard and difficult task. The calculation of NMR properties of molecules using quantum chemical methods has been extensively used in the recent past to settle the tridimensional structures of complex natural and unnatural products. In this Minireview, some recent advances in this area are presented and discussed.

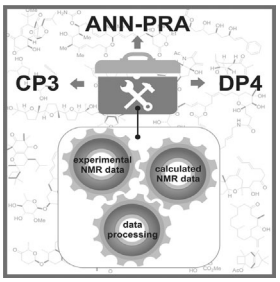


Computational Chemistry

*N. Grimblat, A. M. Sarotti**

■■ – ■■

Computational Chemistry to the Rescue: Modern Toolboxes for the Assignment of Complex Molecules by GIAO NMR Calculations



Computational Chemistry

In their Minireview on page ■■ ff., N. Grimblat and A. M. Sarotti present computational methods that combine affordable quantum chemical calculations with sophisticated data processing to allow the structural or stereochemical assignment of complex organic compounds with remarkable levels of confidence. Such methods have emerged as indisputable complements to experimental NMR spectroscopy.