Volume 8 | Number 9 | 7 May 2021



ORGANIC CHEMISTRY

FRONTIERS

CCS 中国化学会 CHINESE CHEMICAL SOCIETY
SIOC 1950
ROYAL SOCIETY OF CHEMISTRY
rsc.li/frontiers-organic

ORGANIC CHEMISTRY FRONTIERS

REVIEW

View Article Online
View Journal | View Issue

Check for updates

Cite this: *Org. Chem. Front.*, 2021, **8**, 2019

Structural characterisation of natural products by means of quantum chemical calculations of NMR parameters: new insights

Fabio L. P. Costa,[a] Ana C. F. de Albuquerque,[b] Rodolfo G. Fiorot,[b] Luciano M. Lião,[c] Lucas H. Martorano,[b] Gunar V. S. Mota,[d] Alessandra L. Valverde,[b] José W. M. Carneiro[e] and Fernando M. dos Santos Junior*[b]

In the early 2000s, the first articles regarding the calculation of NMR parameters for natural products appeared in the literature. Since then, modelling ^{1}H and ^{13}C chemical shifts and spin–spin coupling constants for this class of compounds has experienced a remarkable increase in precision, accessibility, and application, leading to considerable advances in the field. More recently, significant contributions from several authors have led to continuous growth in this research field, updating and broadening the simulation of NMR parameters, in particular with the application of new techniques for data treatment. Nowadays, such studies are routinely found in the high impact literature. In this review, we intend to cover the general guidelines and the main advances in NMR calculations of natural products published since 2012. We intend to address the bottlenecks of quantum chemical calculations of NMR parameters, including mathematical definitions, updates, and a discussion of relevant examples, and to highlight novel tools, for example DU8+, CP3, DP4, DP4+ and *J*-DP4. We will cover all aspects of NMR simulation focusing on natural products, from the fundamentals to the new computational toolboxes available, combining advanced quantum chemical calculations with complex upstream data processing and machine learning.

Received 7th January 2021,
Accepted 20th February 2021

DOI: 10.1039/d1qo00034a

rsc.li/frontiers-organic

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

I. Introduction

According to Pye and co-workers, in a modern definition, natural products are small molecules derived from the environment, also known as secondary metabolites, that are not involved in primary metabolism.[1] A key feature of natural products is their extensive structural and chemical diversity, presenting molecules with unique molecular frameworks and generating intricate molecular structures.[2] Historically, natural products have had unquestionable importance in the evolution of organic chemistry and related areas. This advance can be seen in the development of synthetic methodologies, frontier orbital theory, biotechnology and, mainly, in the discovery of new drugs.[3]

These natural compounds are widely used in agriculture,[4,5] in the cosmetic and food industries,[6,7] and especially in medicines; they are obtained from terrestrial and marine organisms, demonstrating the economic importance of these ecosystems.[8,9] The history of medicine is told practically since the existence of the first human civilisations, and most new drugs have been developed from natural products or their derivatives.[10] In the last 40 years, as natural products have represented an unparalleled source of structural information, more than 33.5% of approved small molecule drugs can trace their origins to natural products or synthetically designed derivatives. This number increases to 64.9% if synthetic drugs containing pharmacophore groups or mimics of natural products are included,[8] indicating that researchers continue to discover appreciable numbers of natural products with no structural precedent.[1] However, their biological activities are governed by their three-dimensional structures, which is related to their functional groups.[11,12] Consequently, the correct determination of the structure of these substances, usually a difficult task, is of unquestionable importance. Failure to establish the correct structure can dramatically affect the drug development process.[13] In new prototypes, as an example, the synthesis development and the recognition of structural characteristics that correlate with chemical and bio-

[a]*Instituto de Química, Universidade Federal de Jataí, 75801615 Jataí-GO, Brazil. E-mail: fernando_martins@id.uff.br*
[b]*Departamento de Química Orgânica, Instituto de Química, Universidade Federal Fluminense, 24020-141 Niterói-RJ, Brazil*
[c]*Instituto de Química, Universidade Federal de Goiás, 74690-900 Goiânia-GO, Brazil*
[d]*Faculdade de Ciências Naturais/Instituto de Ciências Exatas e Naturais, Universidade Federal do Pará, 66075-110 Belém-PA, Brazil*
[e]*Departamento de Química Inorgânica, Instituto de Química, Universidade Federal Fluminense, 24020-141 Niterói-RJ, Brazil*

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

logical reactivity are directly affected by an incorrect structure elucidation.

In the past, during the 19th and the early half of the 20th century, structure determination of natural products was an arduous process depending almost entirely on chemical synthesis, mainly by derivatisation and degradation processes, taking years of effort. However, in most cases, the assignment of an absolute or relative configuration was practically out of the question.[14]

During the following decades, classical chemical synthetic approaches were gradually replaced by more accurate spectroscopic methods, such as ultraviolet (UV), infrared (IR), circular dichroism (CD) and mass spectrometry (MS). Nevertheless, in the late 1960s, NMR spectroscopy established itself as an indispensable tool in structure determination procedures, accelerating the pace at which isolated compounds can be identified, thus leading to a new era in the chemistry of natural products and drug discovery.[13,15]

Since then, NMR methods have grown in both sensitivity and signal dispersion by the introduction of spectrometers with a stronger magnetic field, associated with inverse detection micro cryoprobes and advanced electronics. This instrumental progress has allowed the rapid acquisition of spectra with a much smaller sample requirement. Another important breakthrough was the development of two-dimensional (2D) NMR spectroscopy. 2D NMR experiments provide information on the correlation of nuclei *via* scalar or dipolar couplings, making it possible to obtain a complete structural assignment relying entirely on NMR spectroscopic techniques.[16] The 2D NMR advance has resulted, for example, in long-range correlation experiments, which allow for observing even six-bond ^{n}JCH heteronuclear couplings, useful for the structure elucidation of proton-deficient molecules.[17] However, despite the extraordinary ongoing advances in spectroscopy, the mischaracterisation of natural products has been and remains a recurrent problem, especially in the presence of several chiral centres. The misinterpretation of NMR data has resulted in frequent reports addressing the issue of structural reassignment.[18]

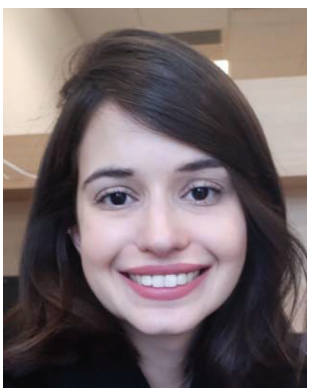
In this context, considerable effort has been devoted to the development of quantum chemical calculations to predict NMR parameters, and thus achieve a more accurate spectral interpretation. The first articles regarding the calculation of NMR parameters for natural products appeared in the literature in the early 2000s, where the groups leaded by Bifulco, Bagno, Sebag and Köck pioneered this application of the technique.[19–22] In 2012, a comprehensive revision paper[23] covered the state-of-the-art of quantum chemical calculation of NMR parameters in natural product chemistry. Since these seminal papers, the computational modelling of ^{1}H and ^{13}C chemical shifts and spin–spin coupling constants has undergone a remarkable increase in precision, accessibility and application, generating considerable advances in the field of natural products.[23–26] To date, with the possibility of applying user-friendly computational protocols to reproduce with high and adequate precision the experimental NMR parameters, in addition to aiding the structure elucidation of new natural products, computational methods have found great utility in helping to uncover and ultimately revise the structure of previously reported compounds.[13] This can be envisaged by examining the statistics on the use of computational protocols in the structure elucidation of natural products. From 2012 to 2020 we found in the literature 199 papers reporting on the isolation of natural products, presenting quantum chemical calculations of NMR parameters in their structure elucidation procedures, and others 30 that employed this protocol to correct errors in the molecular structures of natural products.

This review will cover the general guidelines from the authors' perspective, addressing the bottlenecks of quantum chemical calculations of NMR parameters in natural product



Fabio L. P. Costa

Fabio L. P. Costa has PhD in Chemistry of Natural Products from the Federal University of Rio de Janeiro (2011). Post-doctorate at the Fluminense Federal University (2012) working on organic synthesis of chaconnes. Experience in the field of Chemistry, with emphasis in Theoretical Chemistry and synthesis of natural and analogous products, working mainly in the following subjects: Ab initio *calculations and molecular modeling, with emphasis on the calculation of spectroscopic properties, electronic structure and properties of natural products. His research is focused on in electronic structure and spectroscopic properties of natural products and organic molecules.*



Ana C. F. de Albuquerque

Ana Carolina Albuquerque currently works as a postdoctoral fellow at Fluminense Federal University – Brazil, under the supervision of professor Fernando Martins. She obtained her PhD (2019) in Natural Products Chemistry at Federal University of Rio de Janeiro – Brazil. While she was pursuing her doctoral degree, she worked as a visiting scholar at University of St Andrews, in the research group of professor Michael Bühl. Ana's work includes DFT calculations of NMR chemical shifts and coupling constants in order to study structural parameters of natural products and biomolecules.

This journal is © the Partner Organisations 2021

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

chemistry, including mathematical definitions, updates and a discussion of relevant examples. It will also promote and present the novel tools made available in the last few years to assist in the NMR elucidation process. This new computational toolbox includes DU8, CP3, DP4, DP4+ and *J*-DP4, which combine quantum chemical calculations of NMR parameters with complex upstream data processing and has become the most accurate and important alternatives to solve structural and stereochemical problems in natural product chemistry.[27]

II. ^{13}C and ^{1}H NMR simulation

II.1: Parameters

Shielding tensors. We know that, in general, the magnetic moment of the electron is primarily responsible for the magnetic response of materials. However, atomic nuclei, composed of protons and neutrons, also show magnetic moment, due to the resulting nuclear spin. The electron magnetic moment (m_s) is usually three orders of magnitude greater than the nuclear magnetic moment. Thus, the application of an external magnetic field produces a pronounced effect on unpaired electrons, which tend to align antiparallel to the field, and a smaller, although detectable, effect on the nucleus.

An electron has an angular moment due to its intrinsic spin, with the z component equal to $\pm\hbar/2$, or spin of $\pm\frac{1}{2}$.[28] Two spin functions may be defined, $\alpha(\sigma)$ and $\beta(\sigma)$, where σ is the spin variable which satisfies the eigenvalue equations. As the electron is a charged particle, its intrinsic spin acquires the property of a magnetic dipole. In other words, because of its spin, an electron acts like a magnet when it is under the action of a magnetic field.

However, not only the electron has this property. The nucleus also has an angular moment due to its intrinsic spin (which we call $\boldsymbol{I}$) and an associated magnetic dipole. Unlike electrons, the spin of the nucleus is not restricted to $\frac{1}{2}$. For example, the ^{1}H and ^{2}H nuclei have spins of $\frac{1}{2}$ and 1, respectively. In many organic compounds, hydrogen and carbon are predominant, and they have spins $\frac{1}{2}$.

To investigate the chemical shift, we expose a sample to an external magnetic field, $\boldsymbol{B}_{\text{ext}}$, producing an induced current due to the electronic response to this field and generating an inhomogeneous magnetic induced field, $\boldsymbol{B}_{\text{ind}}$. The induced magnetic field can be of two forms: either the applied static field induces an interaction with the nucleus due to modified fields or the nucleus interacts with the induced and static fields (in this case, the Hamiltonian produces a Zeeman and chemical shift contribution). Here, the main focus is the isotropic chemical shifts produced by a static field, where $\boldsymbol{B}_{\text{ind}} = -\hat{\sigma}\cdot\boldsymbol{B}_0$, which is induced by the surrounding electrons, $\boldsymbol{\sigma}$ is the chemical shift tensor and $\boldsymbol{B}_0$ is the applied field strength vector (spectrometer field). The energy from this induced field is based on the interaction energy between a magnetic field and the magnetic moment (from Zeeman interactions):

$$E_{\text{CShift}} = -\mu\cdot B_{\text{ind}} = \mu\cdot\hat{\sigma}\cdot B_0 = \hbar\gamma I\cdot B_0 \tag{1}$$

where $\gamma\hbar\boldsymbol{I}$ is the quantum mechanics equivalent for the classical μ. The Zeeman Hamiltonian for all spins can be written as:

$$\begin{aligned} H_{\text{Zeeman}} &= -\hbar\sum_{i}^{\text{spins}}\gamma_i I(i)\cdot B_0 = -\hbar B_0\sum_{i}^{\text{spins}}\gamma_i I(i)\cdot 1\cdot B_n \\ &= -\hbar B_0\sum_{i}\gamma_i[I_x(i)B_x + I_y(i)B_y + I_z(i)B_z] \end{aligned} \tag{2}$$

If we consider the field along the z-axis coordinate ($B_x = B_y = 0$, $B_z = 1$), $\boldsymbol{B}_n = \boldsymbol{B}_0/B_0$ ($n = x$, y or z) is the vector and $\boldsymbol{1}$ is the identity matrix. However, we can write the Hamiltonian for the chemical shift (H_{CShift}) in terms of the tensors; in this case,



Rodolfo G. Fiorot

Rodolfo G. Fiorot is professor in the Organic Chemistry Department at Universidade Federal Fluminense-UFF (Niterói) since 2020. He was professor at IFRJ (2016–2019). He earned his BSc (2013) and his MSc (2015) in Chemistry working with organic synthesis of bioactive compounds at the Universidade Federal do Espírito Santo. In 2019, he received his PhD in Chemistry (UFF) under the supervision of Prof. Walkimar Carneiro exploring potential energy surfaces by molecular modeling for systems of environmental interest and organic reactions. His research interests include the computational study of the structure and reactivity of organic compounds and the rational design of molecular materials.



Luciano M. Lião

Luciano Morais Lião is Full Professor at Federal University of Goias, where works since 1996. Graduated in chemistry from Federal University of Goias (1991), with MSc and Ph.D. in organic chemistry from Federal University of São Carlos. He was a visiting Professor at the Universities of Toronto (2019) and Federal of São Carlos (2007) working with nuclear magnetic resonance. Liao's research specializes in structure elucidation of natural products and other chemical compounds classes by means of NMR. He is the coordinator of the NMR Laboratory in UFG and has more than 150 scientific articles in indexed journals.

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

the identity matrix will be changed by the chemical shift tensor (3 × 3 matrix) in eqn (2).

$$H_{\mathrm{CShift}} = \hbar\gamma_i B_0 [I_x(i) I_y(i) I_z(i)] \cdot \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{bmatrix} \cdot \begin{bmatrix} B_x \\ B_y \\ B_z \end{bmatrix} \quad (3)$$

The matrix components σ_{ij} represent the component of chemical shift in the *i*-direction, where the external field is applied in the *j*-direction. Therefore, we can assume that the chemical shift tensor is symmetric for the principal diagonal elements, where the isotropic tensor is $\sigma_{\mathrm{iso}} = 1/3\ (\sigma_{xx} + \sigma_{yy} + \sigma_{zz}) = (1/3)\mathbf{Tr}\{\boldsymbol{\sigma}\}$.

Using the perturbation theory to calculate the chemical shielding for a system under an applied external magnetic field, the chemical shifts derived from the orbital magnetisation induced by the magnetic dipole is given by:

$$\sigma_n = -\begin{bmatrix} \frac{\partial B_x^{\mathrm{ind}}}{\partial B_x^{\mathrm{ext}}} & \frac{\partial B_x^{\mathrm{ind}}}{\partial B_y^{\mathrm{ext}}} & \frac{\partial B_x^{\mathrm{ind}}}{\partial B_z^{\mathrm{ext}}} \\ \frac{\partial B_y^{\mathrm{ind}}}{\partial B_x^{\mathrm{ext}}} & \frac{\partial B_y^{\mathrm{ind}}}{\partial B_y^{\mathrm{ext}}} & \frac{\partial B_y^{\mathrm{ind}}}{\partial B_z^{\mathrm{ext}}} \\ \frac{\partial B_z^{\mathrm{ind}}}{\partial B_x^{\mathrm{ext}}} & \frac{\partial B_z^{\mathrm{ind}}}{\partial B_y^{\mathrm{ext}}} & \frac{\partial B_z^{\mathrm{ind}}}{\partial B_z^{\mathrm{ext}}} \end{bmatrix} \quad (4)$$

The shielding is the tensor that describes the relative change in the local magnetic field at the nucleus position relative to the external magnetic field, which originates in the interaction of the electron cloud with the external magnetic field. It can produce shielding (local field is increased) or deshielding (local field is decreased) of the nucleus.

For the study of nuclear spins, it is necessary to submit a sample to high fields for appreciable magnetisation to occur, which will produce a transition between the aligned state and the anti-aligned state in relation to the externally applied magnetostatic field vector (megahertz). The magnetic moment value associated with the nuclear spin depends not only on the type of atom but also on the electron density around the nuclei, since the magnetic field produced by the alignment of the electronic magnetic dipoles modifies the magnetic field value effectively perceived by the nucleus. The application of a magnetic field affects the levels of spin energy, allowing the observation (in resonance frequencies) of the spectra resulting from the transitions between these levels. The resonance frequencies are proportional to the strength of the resulting magnetic field generated by the spectrometer. Due to the use of different types of spectrometers, standardisation is necessary for the resonance frequency, ν (almost always, tetramethylsilane, TMS, is used as the standard). This procedure is defined by the chemical shift (δ_{k}),

$$\delta_{\mathrm{k}} = \frac{\text{resonance frequency of nucleus } k}{\text{spectrometer frequency}} \times 10^6$$
$$\delta_{\mathrm{k}} = \left(\frac{\nu_{\mathrm{k}} - \nu_{\mathrm{k,ref}}}{\nu_{\mathrm{spectrometer}}}\right) \times 10^6\ [\mathrm{ppm}] \quad (5)$$

where ν_{k} and $\nu_{\mathrm{k,ref}}$ are the resonance frequencies associated with the sample and the standard, respectively. In some cases, the chemical shift is presented as:

$$\delta_{\mathrm{k}} = \left(\frac{\sigma_{\mathrm{k,ref}} - \sigma_{\mathrm{k}}}{1 - \sigma_{\mathrm{k,ref}}}\right) \times 10^6 \approx \sigma_{\mathrm{k,ref}} - \sigma_{\mathrm{k}} \times 10^6\ [ppm] \quad (6)$$

This procedure removes the dependence on resonance frequency by the relation $\nu = \frac{\gamma B_0}{2\pi}(1 - \sigma)$. The approach considers $|\sigma_{\mathrm{k,ref}}| \ll 1$ in the denominator of the relation, which is justified by the ppm scale.

The constant γ is called the gyromagnetic factor and is a characteristic of each nucleus; B_0 is the magnitude of the external magnetic field, and σ_{ref} and σ_{k} are the shielding constants of the reference nucleus and the sample (in ppm), respectively. For large values of σ_{ref}, as in the case of heavy atoms (for example, ^{119}Sn), where σ_{ref} is in the order of 2000 ppm, there is an error of 0.2%; in general, error = $|\sigma_{\mathrm{ref}}| \times 10^{-4}$, with σ_{ref} in ppm, if we use eqn (2) instead of eqn (1).[29] In addition, for large values of chemical shifts, the sensitivity of this parameter increases in relation to structural and chemical changes, due



Lucas H. Martorano

Lucas Haidar Martorano has a degree in Chemistry from Universidade Federal Fluminense – Brazil. Currently, Lucas is a graduate student at Institute of Chemistry of Fluminense Federal University – Brazil, under supervision of professor Fernando Martins and Alessandra Valverde. His work focuses mainly on computational spectroscopy applied to the structural ressignment of chiral molecules.



Gunar V. S. Mota

Gunar Mota is an associated professor in the Institute of Exact and Natural Sciences at the Federal University of Pará, Brazil. He received his B.S. (1999) in Physics at the Federal University of Pará and PhD (2006) in physical chemistry at the Federal University of Rio de Janeiro. He then carried out post-doctoral research with Prof. Lothar Frommhold, in the University of Texas at Austin focusing on Collision-induced absorption in gases. He is acting on the physics of atoms and molecules, computational simulation and molecular spectroscopy.

This journal is © the Partner Organisations 2021

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

to the current induced by the magnetic field, mixing the ground state with excited states.

Spin–spin coupling constants. Since its beginnings, nuclear magnetic resonance (NMR) spectroscopy has been considered a technique with enormous potential for determining the relative[30] and absolute stereochemistry of organic compounds.[31] In addition to the chemical shift, another important NMR parameter is the indirect scalar spin–spin coupling constant, as this is very sensitive to the local effects of electronic distribution, angle of attachment and effects of substituents. Indirect spin–spin coupling constants are transmitted through the electrons in the bonds and are called electron coupling interactions.[32–40] This indirect coupling of nuclear spins normally leads to an unfolding of the NMR signal of each nucleus (or group of equivalent nuclei), since the resonance condition for that nucleus will be reached at a frequency that varies depending on the spin states of other nuclei in the molecule.[41]

Recent developments in theoretical and computational approaches have enabled studies with quantum chemistry methods for a wide variety of molecular properties. In particular, properties resulting from a non-linear response of the molecule to a combination of electric and magnetic fields can be calculated. Although the theory to describe these properties has been known for many years,[42–44] the calculation of magnetic properties is hampered by strong dependence in relation to choice of the gauge origin, the choice of the set of base functions, and the description of the electronic correlation. In theoretical methods as GIAO (Gauge-Independent Atomic Orbitals), the problem of the origin of the gauge can be solved with the introduction of London atomic orbitals or atomic orbitals that include the gauge.[45,46] Therefore, the application of these methods has become widely common in computational methods for the determination of chemical shifts.

The significant increase in the number of published papers where natural *J*-coupling was studied within the universe of density functional theory (DFT) provides interesting insight into how spin–spin couplings are affected by different types of interactions in molecular systems. Ramsey[47] described the non-relativistic calculation of spin–spin couplings based on four interactions between electrons and magnetic nuclei. According to Ramsey's theory, the second-order contribution involving the diamagnetic spin–orbit contribution (DSO) also consists of the indirect nuclear spin-coupling tensor, given by:

$$H_{\mathrm{DSO}} = \frac{\mu_0^2 e\hbar\mu_{N'}}{16\pi^2}\sum_{N,N'}\gamma_N\gamma_{N'} \sum_{\mathrm{k}}\frac{(I_N\cdot I_{N'})(r_{\mathrm{k}N}\cdot r_{\mathrm{k}N'}) - (I_N\cdot r_{\mathrm{k}N'})(I_{N'}\cdot r_{\mathrm{k}N})}{r_{\mathrm{k}N}^3 r_{\mathrm{k}N'}^3} \quad (7)$$

where $\boldsymbol{r}$ is the vector from the nucleus N and N' to the electron k, $\boldsymbol{I}$ is the angular moment, μ is the magnetic moment and γ is the gyromagnetic factor. The DSO contribution is significant for certain coupling constants and involves a product of two gyromagnetic ratios of the pair of two nuclei that can be larger for lighter atoms (resulting in a large overall DSO contribution). Note that the criterion to estimate the order of magnitude of the individual terms is based on an electron in a Bohr orbit; on some occasions, this estimate may not give a good indication of the actual magnitude.

The Fermi Contact (FC) term describes the interactions of nuclear spins with the electronic spin; it is the most important of the terms, and can be affected by hyperconjugative interactions such as, n → σ*, σ → σ*, π → σ* and σ → π*.[48] The FC term is given by:

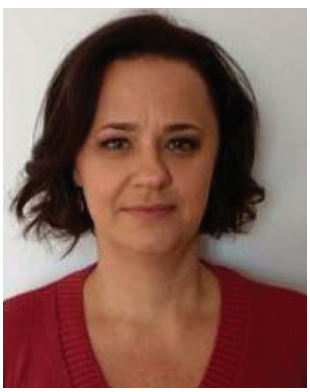
$$H_{\mathrm{FC}} = \frac{4\mu_0\hbar\mu_{N'}}{3}\sum_{N}\gamma_N\sum_{\mathrm{k}}\delta(r_{\mathrm{k}N})\frac{(S_{\mathrm{k}}\cdot I_N)}{r_{\mathrm{k}N}^3} \quad (8)$$

where S_{k} is the spin of the k^{th}-particle and $\delta(\boldsymbol{r}_{\mathrm{k}N})$ is the Dirac delta function. Many authors adopt the point of view that



José W. M. Carneiro

J. W. de M. Carneiro is graduated in Industrial Chemistry from the Federal University of Maranhão (1982), with a doctorate in Sciences, Chemistry modality, from the Military Institute of Engineering-RJ (1990). He joined the Fluminense Federal University in 1992, currently occupying the position of Full Professor. He has experience in computational chemistry, with an emphasis on organic and inorganic physical chemistry, working on reaction mechanism, molecular properties and structures, DFT calculations, QSAR and homogeneous and heterogeneous catalysis. He has published more than 150 scientific articles in indexed journals.



Alessandra L. Valverde

Alessandra Leda Valverde is an Associate Professor in the Institute of Chemistry at Fluminense Federal University, Brazil. She received a BS in Industrial Chemistry (1993) and MSc in Organic Chemistry (1996) from Fluminense. Federal University, Brazil. Her PhD (1999) degrees in Science at Rio de Janeiro Federal University. Over the years, she also worked as P&D Researcher at FIOCRUZ (1999–2004) and as a postdoc at Department of Chemistry at São Carlos Federal University in 2017. She supervises postdoctoral, doctoral, master's and undergraduate students. Her research interests involve isolation and stereostructural determination of bioactive natural products.

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

either FC is a relativistic effect (derived from the Dirac equation) or FC is not a relativistic effect (derived from classical electrodynamics).

The mechanism that produces an electronic polarisation of the spins originates from dipolar interactions between the magnetic moment and the electronic spins of one nucleus. This interaction, that is not annihilated by the magnetic field, is known as spin–dipolar (SD) interaction, and is represented by:

$$H_{\mathrm{DSO}} = \frac{\mu_0 \hbar \mu_{N'}}{16\pi^2} \sum_N \gamma_N \sum_{\mathrm{k}} \frac{3(S_{\mathrm{k}} \cdot r_{\mathrm{k}N})(I_N \cdot r_{\mathrm{k}N'})}{r_{\mathrm{k}N}^5} - \frac{(S_k \cdot I_N)}{r_{\mathrm{k}N}^3} \quad (9)$$

In the isotropic phase, NMR measurements are carried out from four isotropic contributions to the scalar-coupling constant between the M and N nuclei, given by:

$$J_{M,N} = J^{\mathrm{DSO}}(M,N) + J^{\mathrm{PSO}}(M,N) + J^{\mathrm{SD}}(M,N) + J^{\mathrm{FC}}(M,N) \quad (10)$$

Vaara *et al.*[49] published a review on spin–spin coupling tensors that can be determined by either quantum chemical calculations or experimental measurements. These methods are now a standard part of quantum chemical NMR chemical shift packages as a practical aspect for numerical calculations. *J*-Coupling is a phenomenon of quantum mechanics, unlike other NMR properties, but can be understood from semi-classical physics.

A practical approach is the analysis of natural *J*-coupling (NJC) based on natural bond orbital (NBO) methods.[50] The NJC approach differs from the others by computing *J*-coupling based on valence bond[51] or charge density.[52] Thus, the NJC contributions could be expressed in terms of natural localised molecular orbitals (NLMOs), intrinsic natural atomic orbitals (NAOs) and natural bond orbitals (NBOs).[53] It provides an idea of how much the interaction can influence the total FC term and the NJC contribution. The information is transferred over the electronic structure to other nuclei by orientation of the nuclear spin, allowing that *J*-coupling can be determined by the nature of the chemical bond and providing some correspondence with the Lewis structure. In particular, the covalent character of hydrogen bonds can provide an understanding about hydrogen bonding in biological systems.[54–57]

Both NBOs and NLMOs have populations (or occupancies) of their orbitals highly condensed in Lewis-like bonds and lone pairs of the Lewis structure. Instead, the transfer of spin density (or *J*-coupling) is related to wave function antisymmetrisation (see ref. 23 for more details) and to spin hyperconjugative delocalisation. Simply, the NJC approach can predict *J*-coupling interactions and understand the signs and magnitudes of couplings related to molecular conformation. It allows contributions to be satisfactorily treated using computational methods, by the decomposition of the localised molecular orbital contributions from electronic unrestricted density.[58]

From a theoretical point of view, the FC contributions are sensitive to the inclusion of electron correlation effects[59] and are most sensitive to geometry changes. In this way, it is recommended to use DFT, in which electron correlation is accurately treated, unlike *ab initio* calculations based on the usual many-body techniques.

The NJC is based on finite perturbation theory (in this case, the coupled Hartree–Fock theory has been used for the calculation of nuclear spin–spin coupling constants) for the FC term. The perturbation theory of the electronic system has a Hamiltonian describing the nuclear spin operators μ_{M} and μ_{N} for nuclei M and N, respectively. Here, we will show the calculation of the FC term related to the dependence of the polarisation (or distortion) of the electron clouds from external perturbations (as an example, NMR chemical shift in molecules under an external magnetic field). The Hamiltonian is of the form:

$$H = H_0 + A_{\mathrm{M}}\mu_{\mathrm{M}} + A_{\mathrm{N}}\mu_{\mathrm{N}} \quad (11)$$

where A_{M} and A_{N} are the FC terms for nuclei N and M, H_0 is the unperturbed Hamiltonian and the nuclear spin operators are μ_{M} and μ_{N}. The FC term for nuclei N and M can be described by:

$$A_{\mathrm{nuclei}} = \left(\frac{16\pi\mu_{\mathrm{b}}}{3}\right) \sum_{\mu} \delta(r_{\mu\mathrm{nuclei}}) S_{\mu} \quad (12)$$

where $\delta(r_{\mu\mathrm{nuclei}})$ is the Dirac delta function, μ_{b} is the Bohr magneton and S_{μ} is the electron spin operator for the μ^{th} electron. The FC perturbation is given by $\lambda\langle\phi_{\mu}|\delta(r_{\mathrm{nuclei}})|\phi_{\nu}\rangle$, from matrix elements of the Hamiltonian $H^{\alpha} = H_0{}^{\alpha} + \lambda\langle\phi_{\mu}{}^{\alpha}|\delta(r_{\mathrm{nuclei}})|\phi_{\nu}{}^{\alpha}\rangle$ for spin α and $H^{\beta} = H_0{}^{\beta} + \lambda\langle\phi_{\mu}{}^{\beta}|\delta(r_{\mathrm{nuclei}})|\phi_{\nu}{}^{\beta}\rangle$ for spin β, where λ is a perturbation parameter and Φ is the atomic orbital at the



Fernando M. dos Santos Junior

Fernando Martins obtained his PhD (2016) from Natural Products Research Institute at Federal University of Rio de Janeiro – Brazil, under the supervision of Professor Mauro Amorim. He worked as a visiting scholar at Vanderbilt University (2015) in Professor Prasad Polavarapu's group. Following post-doctoral work (2017) at São Carlos Federal University – Brazil under the supervision of Professor Quezia Cass. In 2018, he joined the Institute of Chemistry at Fluminense Federal University – Brazil, where he currently is an adjunct Professor. Fernando's research specializes in structure elucidation of natural products by means of computational spectroscopy, mainly NMR and electronic and vibrational circular dichroism.

This journal is © the Partner Organisations 2021

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

nucleus position.33 For example, the FC term for nucleus N is given by:

$$J_{MN} = \left(\frac{\hbar}{2\pi}\right)\left(\frac{8\pi\mu_b}{3}\right)^2 \gamma_M\gamma_N\lambda^{-1}\sum_{\mu\nu}\rho^N_{\mu\nu}\langle\phi_\mu|\delta(r_N)|\phi_\nu\rangle \quad (13)$$

where γ_N and γ_M are the gyromagnetic ratio for nuclei N and M, and $\rho^M_{\mu\nu}$ is the μth element one-electron spin density matrix related to nucleus N. In terms of the NLMOs, the element of the α-spin matrix elements is subtracted from the corresponding matrix elements of β-spin, and eqn (13) can be easily modified to:

$$J_{MN} = \left(\frac{\hbar}{2\pi}\right)\left(\frac{8\pi\mu_b}{3}\right)^2 \gamma_M\gamma_N\lambda^{-1} \sum_{\mu\nu}\rho^N_{\mu\nu}[\langle\phi^\alpha_\mu|\delta(r_N)|\phi^\alpha_\nu\rangle - \langle\phi^\beta_\mu|\delta(r_N)|\phi^\beta_\nu\rangle] \quad (14)$$

where ϕ^α and ϕ^β are the atomic orbitals for spin α and spin β, respectively, and $\delta(r_N)$ is in the coordinate of the atom being observed. This results in unpaired electron spin, which propagates through the molecule ($\rho^N_{\mu\nu}$).

II.2. Computational aspects

Molecules with multiple conformations: conformational sampling. The NMR spectroscopy timescale is relatively long. Conformational changes in the timescale of picoseconds, such as those arising from fast rotation around unhindered single bonds, are too fast to be detected in common NMR experiments.[60] Consequently, NMR spectra usually represent a Boltzmann-weighted average of the accessible conformers, dictated by the difference in the free energies among these conformers at the specific temperature set for the experiment. On the other hand, the calculation of NMR parameters is commonly performed on a single (and static) structure, defined by a specific nuclear configuration. Thus, in most cases, performing an exhaustive exploration of the conformational space might be recommended, which can often be an overwhelming task. Before starting sampling conformations, however, one should be aware of a few points to avoid unnecessary work. Knowing whether the NMR parameters are significantly sensitive to conformational changes, what conformers are accessible at the given temperature and the most efficient procedure to perform the conformational analysis will help save time and effort.[23]

For molecules with few rotatable bonds, conformational sampling can be accomplished by simple inspection of the minimum energy structures; that is, by fully optimising their geometries with a reliable quantum chemical method and using the corresponding relative energies to compute the Boltzmann distribution. Alternatively, one can perform a systematic conformational analysis, by means of a regular and controlled sampling of selected dihedral angles for each candidate structure, followed by geometry optimisation and the computation of Hessian matrices of the most stable conformers with the recommended theoretical model. The relative population of each conformer is then computed from the Boltzmann distribution.

For molecules with many rotatable bonds, automated searches are indicated. For a molecule with N rotatable bonds in an aliphatic or acyclic chain, 3^N conformers are possible. Even for a small molecule with only three rotatable bonds, up to 27 conformers may be expected. Identification and computation of all the inequivalent conformers may represent a complex task. For natural products, the existence of more than three rotatable bonds is the rule; therefore, an automated conformational search is recommended. The most frequent automated conformational search techniques are the stochastic Monte Carlo (MC) method[61,62] (commonly based on the Metropolis algorithm) and molecular dynamics (MD) simulations. They typically rely on semiempirical or molecular mechanics methods and are able to quickly sample a large subsection of the conformational space and find local minimum-energy candidate structures. MC simulations perform random rotations around single bonds to generate a large set of structures, which are further optimised, usually with a fast computational method such as molecular mechanics. The result is a set of energy-minimised structures. Several studies on natural product structures, especially for the challenging task of determining their absolute configuration, have adopted the MC method to explore the conformational space of the molecule in question.[63–66] When employing MD simulations, the user must define the time span of the simulation and the size of the box, as well as the physical conditions to which the structure is subjected. This methodology is particularly useful when explicit solvation needs to be considered. Plotting the dihedral angle distribution may inform about the efficiency of the conformational sampling throughout the simulation. The outcome of MD simulations is a trajectory file containing information about the time-evolution of the conformations. These conformations can be extracted with or without a given number of solvent molecules (selected within a defined cut-off radius) for a more refined optimisation. MD has been employed in several studies to explore the conformational space of structurally complex organic molecules.[67–69] Alternative automated conformational search strategies, such as genetic algorithms,[70–72] simulated annealing[73,74] and distance geometry methods[75,76] have been used to locate possible conformers. Reviews have been published concerning automated conformational search methods.[23,26,77,78]

Once the full set of conformers is sampled, the ones within a suitable energy window (those in an energy range of 5 to 10 kcal mol^{-1} above the lowest energy conformation should be enough to encompass all the structures contributing to the experimental NMR parameters) are further fully optimised with a more refined method, usually one based on quantum mechanics.

Refined geometry optimisation. This is a fundamental step to compute reliable molecular geometries and their respective energies. Both correct geometry and reliable relative energies are requirements to accurately compute NMR parameters. Usually, this is done using either wavefunction post-Hartree–Fock methods or DFT. The latter is the most frequently used,

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

as it provides good results at a reasonable computational cost. From a practical perspective, the averaged NMR chemical shifts should take into account all the conformers within an energy range of 3 kcal mol^{-1} above the lowest energy conformation, enough to encompass >99% of all conformers at 25 °C. The final weighted NMR chemical shifts are computed considering the population of each conformer obtained from a Boltzmann distribution based on their relative energies.

Regarding the specific functional to refine the geometry optimisation, B3LYP is still the most frequent, although there are severe criticisms of the performance of this functional. We address the reader to some comprehensive works on the performance of different functionals based on kinetic and thermodynamic property simulations.[79–84] B3LYP has been shown to fail in properly describing energy changes in chemical reactions,[85–88] bond dissociation enthalpies[89] and formation enthalpies.[90,91] For instance, Schwabe and Grimme assessed the thermodynamic relative energies of more than 30 isomerisation reactions of organic molecules and found a mean average deviation (MAD) of 2.29 kcal mol^{-1} using the B3LYP/TZV(2df,2pd) method.[80] A MAD of this magnitude could lead to a misinterpretation in the Boltzmann distribution to the conformational analysis. However, many of the effects causing errors are of a systematic nature. For this reason, by comparing conformers, there is an error cancelation.[23] Some reports have shown that the B3LYP functional employed with at least a double-ζ basis set with diffuse polarisation functions (*e.g.*, 6-31+G**) can be used successfully to calculate refined geometries and their free energies, validated by the experimental relative ratio of the conformers.[92,93] Alternatively, Minnesota functionals (*e.g.*, M06 and M06-2X), combined with the same type of basis set, have also been found to be adequate in the refined optimisation step before chemical shifts.[94–96] Finally, a Boltzmann distribution analysis is performed to account for the relative contribution of each conformer to the overall NMR simulated spectrum. This is the point at which the accurate computation of relative Gibbs free energies for each conformer is fundamental. Eqn (15) uses the calculated values of free energies for each conformer to evaluate the abundance of conformer *i* relative to conformer *j*:

$$\frac{P_i}{P_j} = e^{\frac{(G_i - G_j)}{k_B T}} \quad (15)$$

where T is the absolute temperature and k_B is the Boltzmann constant. Eqn (16) accounts for the overall computed chemical shift (δ) of a particular nucleus using the chemical shift of that specific nucleus calculated for each conformer weighted by its relative contribution (δ_i).

$$\delta = \sum_i \frac{\delta_i e^{-\frac{G_i}{RT}}}{\sum_j e^{-\frac{G_j}{RT}}} \quad (16)$$

Any of the eqn (15) and (16) can be fed in G by molar or per-particle energy values. This will dictate the use of the appropriate constant (molar gas or Boltzmann, respectively). Whilst eqn (16) provides the actual calculation of the Boltzmann-averaged chemical shift, eqn (15) allows rapid determination whether the conformer in question will or will not contribute to the chemical shift. Usually, a free energy window of 3 kcal mol^{-1} above the lowest energy conformer must account for all conformers that relevantly contribute to the overall NMR chemical shifts.

Computing NMR parameters: chemical shifts and coupling constants. The final goal is the calculation of the NMR parameters, whether to predict, to confirm or to revise a chemical structure. In the case of confirmation/revision, the simulated data are compared to the experimental chemical shifts by statistical treatment. Section II.4 is dedicated to presenting some possibilities, applications, and limitations. For the particular influence of computational quantum-chemical methods (such as semiempirical, DFT or other quantum-chemical methods based on wavefunction) on the quality of the NMR data simulation, there are comprehensive reviews that we address the interested reader to look for ref. 23, 97 and 98. Our aim in this section is to highlight some of the most relevant and frequently used approaches for this purpose.

One of the simplest approaches to estimating chemical shifts is the increment-based method, also known as fragment-based model. In such an approach, libraries of molecules with known chemical shifts are used to compare with the molecule of interest based on their structural similarities. This methodology of analysis has been used for decades, *e.g.* by using the tables found in many undergraduate textbooks describing the spectrometric identification of chemical compounds.[99,100] Currently, several platforms allow for automated estimates of ^{1}H and ^{13}C NMR chemical shifts. Featured examples are ChemOffice,[101] ACD[102] and NMRPredict.[103] Although some works highlight the utility of increment-based methods, which can be as good as or even better than NMR computed by quantum chemical methods (especially for ^{13}C chemical shifts),[104,105] there are several significant limitations to this approach. One is that the breadth of the library queried limits the accuracy of prediction. If the molecule of interest has nuclei that are not well-represented in the library, then the discrepancies could be significant, leading to a misassignment.[23,106] Furthermore, these methods may not be capable of correctly distinguishing stereoisomers.[104] Moreover, if the case comprises molecules with high conformational mobility, then quantum-chemical methods are of utmost importance to access reliable data.

DFT is the prevalent method to compute both nuclear chemical shifts and coupling constants, as it can provide high-quality results associated with a small computational cost. There are several reviews showing successful applications of different functionals in predicting, revising or confirming a chemical structure.[13,23,26,77,97,107,108] Herein, we highlight popular functionals that have been effectively used in NMR chemical shift calculations. Among the great zoo of functionals available, the widespread B3LYP is usually suitable enough to produce reliable results, especially if scaled for both ^{1}H and ^{13}C chemical shifts. Jain and co-workers[109] compared

 This journal is © the Partner Organisations 2021

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

the efficiency of reproducing ^{1}H chemical shifts for a set of 80 small organic molecules and showed that scaled TPSS and B3LYP functionals, with the aug-cc-pVDZ basis set and computing with the continuous set of gauge transformations (CSGT), provided the smallest errors (root mean-square error of 0.132 and 0.127 ppm, respectively). Besides the hybrid functional B3LYP, other traditionally used functionals to compute both ^{1}H and ^{13}C NMR chemical shifts are the hybrid functionals mPW1PW91, PBE0 and PBE1.[23]

In 2006, Cramer and co-workers[92] developed two empirically parametrised functionals, WC04 and WP04, to respectively improve ^{1}H and ^{13}C chemical shift simulations by removing systematic error. These functionals are derived from B3LYP by systematically adapting the weighting parameters that account for the exchange and correlation energies. To evaluate their efficiency, the author simulated the chemical shifts of more than 40 organic molecules and compared with their experimental values obtained in deuterated chloroform. The chemical shifts computed with WP04 and WC04 were more accurate than the values obtained with the original B3LYP and other functionals commonly used for this purpose, such as mPW1PW91 and PBE1. In this work, they used the 6-311+G (2d,p) basis set, considered implicit solvation by the IEFPCM method in chloroform and chemical shifts were computed by the GIAO method. However, they also noticed that, by applying scaling factors (discussed in section II.3), the regular B3LYP, mPW1PW91 and PBE1 functionals were able to reproduce ^{1}H NMR chemical shifts as well as WP04 and even better than WC04 for ^{13}C.

In 2017, Iron[110] called attention to the fact that the usage of long-range corrected functionals for calculating ^{13}C NMR chemical shifts outperformed the other classes of functionals. In a thorough study, the author evaluated over 70 exchange–correlation functionals, as well as the implicit solvation model, basis set effect and the NMR method for computing the shielding tensors. The recommendation was to use the COSMO-CSGT-LC-TPSSTPSS/cc-pVTZ level of theory in the chemical shift calculation. This level was then applied successfully in four case studies comprising problems in natural product identification. The author also claimed that, although this method was computationally more expensive than 'conventionally' used functionals (*i.e.*, B3LYP, PBE0, PBE1 and mPW1PW91), the cost is not constraining.

If one desires higher accuracy and is willing to pay the cost of using more computationally demanding methods, post-Hartree–Fock (post-HF) *ab initio* methods can be used. However, it is worthy to highlight that these methods are usually limited to small molecules and are unfeasible for more complex natural product structures. The second order Møller–Plesset perturbation theory, MP2, and the coupled cluster single-double-triple level of theory, CCSD(T), are examples of such post-HF methods. Auer and Gauss[111] reported a benchmarking study of ^{13}C nuclear magnetic shielding constants in a set of small organic molecules, with low or no conformational mobility, showing that near-quantitative accuracy (deviation below 2 ppm from experiment) was achieved by using CCSD (T). The ^{13}C NMR chemical shifts were computed at accurate equilibrium geometries, calculated with CCSD(T) using the cc-pVTZ or cc-PVQZ basis set. MP2 also showed good results, followed by DFT (unscaled B3LYP and BP86) and the Hartree–Fock self-consistent field. In 2014, Flaig and co-workers[112] presented an extensive study calculating 48 hydrogen and 40 carbon nuclei chemical shifts in 30 molecules capable of covering a broad range of the NMR scale relevant to organic chemistry. They compared the ability of the HF, MP2 and DFT methods to reproduce high-quality results (provided by the gold standard method, CCSD(T)/cc-pVQZ). Among the functionals, the considered hybrid-GGA functional B97-2 reached the lowest standard deviations; however, MP2 provided the result closest to CCSD(T).

Concerning spin–spin coupling constants (SSCC), where the proton–proton coupling constant (J_{H-H}) is one of the most important, especially when dealing with organic molecules, the protocol established by Bally and Rablen[113] to simulate J_{H-H} values has been shown to result in accurate data. They focused exclusively on evaluating the capability of DFT-based procedures to reproduce well-established experimental J_{H-H} values (a range of 165 coupling constants from 65 molecules, representing a good variety of the chemical space). After assessing 250 different computational approaches (combination of methods, density functionals and basis sets), they concluded that calculating only the Fermi contact (FC) term provided more accurate data instead of computing all the contributory terms accounting for J_{H-H}, if one allows for linear scaling of raw results and uses core-augmented basis sets. They recommend using B3LYP/6-31G (d,p) u + 1s as the best and one of the most economical combination for predicting the coupling constants, with Root Mean Square Deviation (RMSD) values below 0.5 Hz compared to the experimental data.

A helpful guide in the choice of the computational method to perform both refined geometry optimisations and further NMR parameters, as well as scaling factors for both ^{1}H and ^{13}C and solvation suggestions, can be found at the CHEmical SHIft REpository with Coupling Constants Added Too (CHESHIRE CCAT) developed by Tantillo.[114]

II.3. Methods for reducing error

Use of a specific reference compound. Computed isotropic magnetic shielding (σ) values can be converted to chemical shifts (δ) by the use of a standard reference compound. In this procedure, the calculated chemical shift of the molecule of interest (δ^{x}_{calc}) is obtained by subtraction of the calculated isotropic magnetic shielding of the reference compound (σ^{ref}_{calc}) and of the molecule of interest (σ^{x}_{calc}) (see eqn (17)).[108]

$$\delta^{x}_{calc} = \sigma^{ref}_{calc} - \sigma^{x}_{calc} \quad (17)$$

Usually, for natural product calculations, tetramethylsilane (TMS) is chosen as the reference standard, since it is the one usually employed for experimental NMR spectroscopy. However, any molecule can be used as an internal reference,

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

giving that its experimental chemical shifts (δ_{exp}^{ref}) are known. Thus, calculated chemical shifts can be obtained according to eqn (18).[23]

$$\delta_{exp}^{x} = \sigma_{calc}^{ref} - \sigma_{calc}^{x} + \delta_{exp}^{ref} \quad (18)$$

The main advantage of using a reference compound is the systematic error cancellation of the calculated parameters. Considering that the same level of theory is used for calculating the molecule of interest and the reference standard, errors in both calculations are proportional and, thus, can be cancelled. In addition, if nuclei in both molecules experience the same chemical environment, the error cancellation is more significant.[115]

Based on these ideas, Sarotti and Pellegrinet proposed a 'multi-standard approach' (MSTD). In their work, several organic molecules were tested as different standards for ^{13}C chemical shift calculations. They concluded that the use of methanol as a reference for sp^3 carbons and benzene for sp^2 carbons provides theoretical chemical shifts with much higher accuracy than when TMS is employed.[116] In 2013, Andrews and Spivey expanded this approach by using fragments with similar chemical environments present in the molecules of interest.[117]

Another alternative to the use of reference compounds is to employ a linear regression method. In this approach, isotropic magnetic shielding values are calculated for several molecules and then plotted against experimental chemical shifts. The *Y*-intercept value of the best fit line can be used to obtain chemical shifts for the molecule of interest, as seen in eqn (19).[118] In the next subsection, linear regression approaches will be addressed in more detail.

$$\delta_{calc}^{x} = Y\text{-intercept} - \sigma_{calc}^{x}. \quad (19)$$

Linear regression: empirical scaling methods. As mentioned in the previous subsection, linear regression approaches can be used to generate a *Y*-intercept value, employed as an alternative to reference compounds in chemical shift calculations.[118] However, the main application of these procedures is to cancel systematic errors in calculations, using a method called empirical scaling.[119]

There are mainly two types of empirical scaling methods: internal and generalised scaling. In the former, a linear regression is generated from computed and experimental data of the molecule of interest. In the latter, the linear regression is generated from computed and experimental data of a set of molecules and then the scaling factor is applied to the molecule of interest. Generalised scaling factors are more frequently applied in natural product calculations.

Generalised scaling factors are generated through linear regressions between experimental chemical shifts and calculated parameters (either computed chemical shifts or magnetic shielding tensors can be used) for a set of organic molecules. From these regressions, slope (a) and intercept (b) values of the best-fit lines are used to correct calculated parameters (δ_{calc}) of the molecule of interest, obtaining scaled chemical shifts (δ_{scal}) (see eqn (20)).[120,121]

$$\delta_{scal} = a\delta_{calc} + b \quad (20)$$

Several studies indicate that empirical scaling methods can reduce errors from several different sources, such as the neglect of rovibrational effects, the treatment of solute–solvent interactions and inaccuracies associated with the level of theory. Since these errors appear to be systematic in many cases, they can be cancelled.[25,122,123]

The use of scaling factors started with the pioneering work of Forsyth and Sebag, in 1997. Based on a set of simple organic molecules, they generated an equation to predict chemical shifts from calculated magnetic shielding tensors. From geometries optimised with the MM3 force field and tensors calculated with B3LYP, along with the 3-21G basis set for hydrogen atoms and 6-31+G* for carbon atoms, it was possible to obtain RMSD values close to 3 ppm for ^{13}C nuclei.[22]

One of the first uses of scaling factors in natural product calculations was published by Barone's group. Using an approach similar to that of Forsyth and Sebag, a linear regression was obtained from a set of 14 organic molecules with complex structures. The scaling factor was applied to the calculated chemical shifts of three natural products, which had their structures revised previously in the literature. The results showed that the difference between scaled and experimental chemical shifts were much smaller for the right structures than for the wrong ones, for all three compounds.[19]

Empirical scaling procedures can be applied to NMR calculations of any nucleus, using any level of theory. It is recommended that the set of molecules used for linear regression and the molecule of interest be calculated with the same level of theory. In this way, similar systematic errors are produced in both cases and error cancellation should be more effective. Similarly, it is preferable that all calculations are performed in the gas phase or in the same solvent.[23]

Additionally, special attention should be paid to the set of molecules chosen for the linear regression procedure. There are two possibilities available: a set with several molecules, considering a wide structural variety or, alternatively, a smaller set with structural similarity with the molecule of interest.

In 2016, de Albuquerque *et al.* analysed the effect of using a scaling factor parameterised with a set of molecules with structural similarity to the molecule of interest. In this study, they used a set of 10 sesquiterpene molecules to create a scaling factor for ^{13}C chemical shift calculations. When the equation was applied to other sesquiterpene structures, it resulted in computed chemical shifts with much higher accuracy than when a generic scaling factor was used. Therefore, it was concluded that the use of molecules with similar structures implicates error cancellation of a higher degree.[124]

In summary, we strongly recommend the use of scaling factors in chemical shift calculations. In many cases, the amount of error cancellation in computed parameters is sufficient to provide accurate results, without the need to apply

 This journal is © the Partner Organisations 2021

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

more expensive computational methods. Currently, there are several papers reporting on the creation of scaling factors for any kind of organic molecule. As mentioned before, a helpful guide in the choice of the scaling factors for both ^{1}H and ^{13}C chemical shifts can be found in CHESHIRE CCAT, developed by Tantillo's group, which can be accessed through the website http://www.cheshirenmr.info.

Solvation modeling. Most quantum chemical calculations, including those of NMR parameters, consider molecules in an isolated state. However, experiments are usually performed in solution, where molecules can interact with each other. The presence of solvent molecules can influence the system in two ways: through the induction of structural changes (indirect effect) or by modifications in electronic density (direct effect). NMR parameters are sensitive to both effects. This way, to obtain a satisfactory reproduction of experimental data, in some cases it is necessary to simulate solute–solvent interactions.[125,126]

There are two kinds of methods to simulate solvent effects in quantum chemical calculations: those that describe molecules in an atomistic way (discrete models) and those that treat solvent molecules as a continuum (implicit models). However, frequently, combinations of the two models are employed: the first solvation layer is simulated explicitly, whereas the remaining solvation layers are considered as a continuum.[115]

Generally, discrete models generate a more precise description of solvent effects, since even specific interactions with the solute can be taken into account. Meanwhile, explicit treatment of solvent molecules implicates in a significant increase in computational cost.[127] This problem can be addressed using methods such as QM/MM, in which part of the system (usually, the molecule of interest) is calculated using quantum mechanical methods, while the rest of the system (solvent molecules) is treated with a molecular mechanics force field.[128,129] Recently, several fragmentation methods (some of them specific to NMR calculations) have been used as a tool to reduce the demand on computational resources by partitioning the system into several small fragments that can be computed in parallel.[130]

Nevertheless, most of the interactions between organic molecules, such as natural products, and solvents commonly used in NMR experiments (like deuterated chloroform and dichloromethane) are non-specific and small.[106] Therefore, one alternative to discrete models, which usually generate accurate calculated parameters, is implicit models. As opposed to the explicit method, solvent molecules are treated as a continuum and they are characterised by their dielectric constant. Thus, the solvated system can be simulated without increasing the number of explicit electrons; consequently, this model requires less computational cost.[131] For this reason, implicit methods such as the integral equation formalism-polarisable continuum model (IEF-PCM) and conductor-like polarisable continuum model (CPCM) are the most commonly used tools to simulate solvent effects in NMR calculations.[126]

The standard and most recommended approach for the calculations of NMR parameters in natural product compounds is to use continuum models, like IEF-PCM, in calculations of chemical shifts and coupling constants. Since solvent effects can induce modifications in molecular geometry, it is recommended to include these methods in the geometry optimisation step as well. Finally, implicit models should also be implemented in the calculations of the reference compound, such as TMS.

Parametric corrections. Parametric corrections are the most recent tools used for reducing errors in the calculation of ^{13}C NMR δ and in the calculation of spin–spin coupling constants (SSCC). The Kutateladze research group has made seminal contributions to this field.[121,132–134] As previously stated, one of the greatest challenges in calculating the magnetic properties of natural products to determine or reassign their structures is to obtain protocols for fast and accurate calculations of SSCC. Based on the fact that the Fermi contact (FC) mechanism dominates scalar nuclear spin couplings, in 2011 Bally and Rablen proposed a practical methodology for calculating and correcting linear FC to estimate SSCCs.[113] This methodology involved an uncontracted basis set defined for hydrogen enlarged atoms with a function of type 1s so restricted that they termed it (u + 1s).[113] This approach can be used to obtain reasonably accurate estimates of proton–proton SSCCs.[23] Recently, the Kutateladze group has developed a multiparametric approach to scaling of FC based on a representative set of SSCC 'types' defined by connectivity and hybridisation.[132] Kutateladze termed this relativistic force field (rff) method DU4. This first rff DU4 is a method for calculating proton SSCCs *via* a hybridisation-discriminated multiparameter scaling scheme for FC computed with a new DU4 basis set for hydrogen atoms and an inexpensive 4-31G basis set for carbons. The second version of the rff method published by the Kutateladze group is DU8, a new method using a new basis set; it is an NBO-aided scaling of proton–proton SSCCs.[133] Thus, DU8 was developed for the scaling of FC to compute accurate proton SSCC. It is a powerful and easy to apply method that is able to accurately reach RMSD ~0.2–0.4 Hz in predicting SSCCs.[134] DU8 utilises a smaller parameter set and NBO hybridisation for parametrisation (NBO hybridisation coefficients) of FCs, which involves only four general SSCC types, scaled using both empirical scaling parameters. The DU8 basis set for heavy atoms is the inexpensive 4-21G basis set for carbons; 3-21G* for Si, S, P, Cl, Se, Br and I; and finally, 6-31G(d) for Li, Be, B, N, O and F. The third generation rff parametrisation method DU8c has been developed for scaling of FCs to determine nuclear SSCC (C–H, C–C, H–H) with small errors.[133] DU8c employs hybridisation correction based on NBO coefficients with small empirical scaling set parameters for C–H and H–H SSCCs. The DU8c method is able to accurately reach RMSD ~0.2–0.4 Hz in predicting carbon-proton SSCCs (2.3 Hz for 1J).[134] The information provided by DU8c plus sophisticated NMR experiments capable of producing accurate SSCC can be a powerful tool for determining the stereochemistry of natural products. With DU8c, the Kutateladze research group has made some important changes to keep both ^{1}H–^{1}H and ^{13}C–^{1}H parametric schemes uniform. Unfortunately, these changes did not produce any significant

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

improvement in the performance of DU8 when predicting ^{1}H–^{1}H SSCCs.[134]

Thanks to the parametric studies performed by the Kutateladze group, the validation of how to use parametric equations (in a simple way) to take into account the contributions of spin–orbit (SO) in terms of the type of hybridisation and structure allowed the application of rff parametrisation methods in diverse scenarios.[135–138] Spin polarisation induced by heavy-atom SO coupling interacts with nuclear magnetic moments of the system and are principally guided by the same FC mechanism.[139] In other words, this 'phenomenon' is mandatory to rationalise how the effect of heavy atoms (for example, halogens) propagandise through the molecule; it is analogous to the known mechanisms for indirect FC nuclear spin–spin coupling.[121] Finally, DU8+ is a protocol that combines low computational cost SSCC and ^{13}C NMR δ DFT calculations for more reliable organic compound structure elucidation.[121] DU8+ was mainly applied to the structure determination and stereochemistry of compounds containing heavy atoms, which is a real challenge. Although for the structural assignment of natural products, ^{13}C δ has been employed most frequently and has been a successful tool, one important exception to this scenario is to calculate the ^{13}C NMR δ when heavy elements, for example halogens, are connected or in the neighbourhood of carbon atoms. Kutateladze and Reddy validated the isotropic magnetic shielding (σ) of DU8+ on a large set of halogenated natural products.[121] To save computational time, Kutateladze used the gas phase GIAO-ωB97x-D/6-31G(d) level of theory to calculate the ^{13}C NMR δ for a set of halogenated and non-halogenated test compounds (experimental data were in $CDCl_3$). For carbons bearing heavy atoms, σ was scaled using a binomial formula:

$$\sigma_{corr} = a\sigma^2 + b\sigma + c \quad (21)$$

For the training molecules, the RMSD values for σ_{corr} (σ corrected) varied from 1.14 to 2.73, with CCl_3 and Cl, respectively.[121] The corrected σ_{corr} were then converted into δ, in ppm, using a global linear correction (*i.e.*, the same linear correction used for σ values calculated for any carbon atom with or without a heavy atom).

$$\delta\ (\text{ppm}) = 201.8\text{–}1.044\sigma_{corr} \quad (22)$$

This level of theory for the test compounds revealed systematic errors for a limited number of hybridisation/structural types. Consequently, they identified several structural types producing such systematic errors and introduced an additional empirical correction.[121] For solvents other than chloroform, the calculated δ could be linearly scaled to best match the experimental data. It worked well for a test set of ^{13}C δ for atoms not bearing heavy atoms, with RMSD = 1.49 ppm (R^2 = 0.999). With the DU8+ method, Kutateladze and Reddy showed that parametric corrections to DFT-computed δ in conjunction with rff-computed SSCC allowed for the fast and reliable screening of a large number of reported halogenated natural products, resulting in expedient structure validation or revision.[121] They applied DU8+ to a test set of more than 100 halogenated sesquiterpenes and other natural products with a rather high rate of misassignment (~13%). In the original DU8+ paper, 16 structures of halogenated natural products were revised. In one recent paper, they applied DU8+ to revisit briarellins, as cited before.[140]

II.4: Procedures to correlate experimental and computational data

Statistical parameters: fit between computed and experimental shifts using R^2, MAE and RMSD. After the simulation of chemical shift values, some fundamental questions arise. How should the fit between the computed and experimental shifts be evaluated? How do we know whether the compound we simulated is the correct one? The answers to these questions are based on the same basic premises that the correct structure will show better agreement between experimental and computed NMR data than the wrong ones. The simplest way to evaluate the agreement between experimental and computed NMR data is through the common statistical parameters of correlations used to measure the accuracy of continuous variables, such as the linear correlation coefficient (R^2), the mean absolute error (MAE) and the root mean square deviation (RMSD).

$$\text{MAE} = \frac{\sum_n \left|\delta_{calc} - \delta_{exp}\right|}{n} \quad (23)$$

$$\text{RMSD} = \sqrt{\frac{\sum_n \left|\delta_{calc} - \delta_{exp}\right|^2}{n}} \quad (24)$$

R^2 measures the strength and direction of a linear relationship between two variables; in this case, between the computed data and the experimental NMR chemical shift values. MAE measures the average magnitude of the absolute error. RMSD also measures the average magnitude of the error between the simulated and the experimental NMR chemical shift values, but using a quadratic score. These basic parameters are easy to calculate and make it possible to evaluate discrepancies between calculated and experimental NMR chemical shift values.

Therefore, if the NMR calculation of the compound is expected to support a given assignment, any outside value of those parameters indicates an unreliable proposal. In this way, through statistical analysis, it is possible to indicate the correct structure among possible candidates. In this case, the right candidate is expected to yield an R^2 value close to 1, in addition to small values for both MAE and RMSD, when compared with the corresponding statistical parameters for the wrong candidates.[27] This kind of analysis was used by Rychnovsky[141] in a seminal paper to revise the proposed structure of hexaciclinol, a complex natural product presenting a highly oxygenated structure isolated in 2002 from *Panus rudis*.[142] The originally proposed structure (Fig. 1a) presented an endoperoxide moiety, which was claimed to be synthesised later in 2006.[143] However, by using quantum chemical simu-

 This journal is © the Partner Organisations 2021

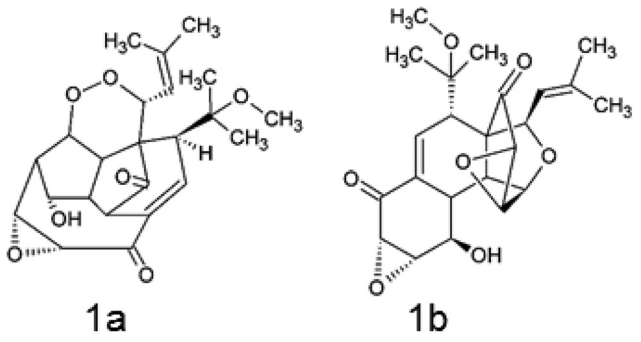


Fig. 1 Originally proposed (**1a**) and the correct revised (**1b**) structures of hexacyclinol.

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

lation of ^{13}C NMR chemical shifts, Rychnovsky argued that a synthetic alteration of the substance isolated in the same year from another strain of *Panus rudis* (Fig. 1b) was a more feasible structure for the isolated natural product. For the original proposed structure, Rychnovsky obtained an MAE of 6.8 ppm with a maximum error of 22 ppm, while the correct revised structure displayed a MAE of 1.8 ppm with a maximum error of 5.8 ppm.[141] In the same year, Rychnovsky joined an experimental group and through the total synthesis and X-ray crystallography, ultimately demonstrated that the revised structure proposed by the NMR simulations was the correct one.[144] In 2009, Bagno and Saielli extended the analysis of hexaciclinol with computation of ^{1}H NMR chemical shifts and scalar *J* couplings and, as expected, obtained a much better fit for the revised structure.[145]

With the use of these basic statistical parameters of correlation, several structural questions were solved in natural product chemistry in recent decades (see ref. 146–153), representing a powerful tool that provides helpful assistance in the structure elucidation of new natural products and helping to uncover and ultimately revise the structure of previously reported compounds.[13] However, it is important to highlight that, in some cases, the simulated chemical shifts for two or more isomers may be very similar, leading to similar statistical parameters, preventing an unequivocal assignment towards a particular structure. This can often occur in stereochemical assignments where several diastereoisomers need to be evaluated.[154,155] In these cases, more sophisticated statistical procedures to correlate experimental and computational values must be employed.

Beyond statistical parameters: CP*n*, DP4, Py-DP4, DP4+, DP4.2, DICE AND *J*-DP4. Considering the cases where the simulated chemical shifts for two or more isomers may be very similar, preventing an unequivocal assignment towards a particular structure, more sophisticated statistical procedures have recently been developed to help correlate experimental and computational data. These new parameters combine the quantum chemical calculations of NMR shifts with upstream data processing, generating a higher level of confidence than those expected from simple aforementioned statistical parameters.[27] These new procedures, called CPn and DP4 (and its derivatives py-DP4, CP4+, DP4.2, DiCE and *J*-DP4), are currently seen as the most accurate and important alternative tools to solve complex structural and stereochemical problems from a theoretical point of view.

CPn. The CPn method, first introduced in 2009 by Smith and Goodman,[64] was developed to address the question of assigning two sets of experimental data to two possible candidates. This is a common situation in synthetic organic chemistry, in which many chemical transformations might afford two isomers that each need to be reliably confirmed. This comparison method is based on the fact that the differences between NMR chemical shifts in similar carbons should be calculated in a more accurate way than the NMR chemical shifts themselves due to a cancellation of systematic errors.[64] With this in mind, it is easy to accept that by having two sets of experimental values (namely, *A* and *B*) and two sets of calculated values (namely, *a* and *b*), the differences between the chemical shifts obtained by the combination $A = a$, $B = b$ (option 1) will be completely different from the combination $A = b$, $B = a$ (option 2). In other words, the differences in calculated chemical shifts between the two isomers compared with the corresponding differences in experimental shifts are used for structure assignment in CPn parameter. For the correct assignment (option 1: $A = a$, $B = b$) the experimental differences are expected to be correctly reproduced by the calculated data; however, for the incorrect assignment (option 2: $A = b$, $B = a$), the agreement is expected to be much less good because of the mismatch between the NMR chemical shifts. The information obtained by these comparisons can be combined into a single number by multiplying the experimental and calculated differences ($\Delta_{exp}\,\Delta_{calc}$) and summing the products:

$$\sum \Delta_{exp}\,\Delta_{calc} \tag{25}$$

where $\Delta_{exp} = (\delta_A{}^i - \delta_B{}^i)$ is the difference in the experimental chemical shift of the i^{th} nuclei between *A* and *B*, and $\Delta_{calc} = (\delta_a{}^i - \delta_b{}^i)$ is the difference in the calculated chemical shift of the i^{th} nuclei between *a* and *b*. Doing this, it is expected for the first option ($A = a$, $B = b$) that certain nuclei present opposite values when comparing with the same nuclei in the second option ($A = b$, $B = a$). Dividing this by a normalisation term, $\sum(\Delta_{exp})^2$, which represents the value of the above overall sum that would be obtained if all the differences in experimental shift were perfectly reproduced by the calculation, affords the CP1 parameter:

$$\mathrm{CP1} = \frac{\sum \Delta_{exp}\Delta_{calc}}{\sum \left(\Delta_{exp}\right)^2} \tag{26}$$

A large positive value indicates good agreement (first option: $A = a$, $B = b$) and the assignment likely to be correct, whereas a large negative value indicates poor agreement (second option: $A = b$, $B = a$).

However, the calculations sometimes do not reproduce the experimental trend perfectly. Very often the calculations generate a Δ_{calc} value with the same sign but smaller magnitude than Δ_{exp}, seen as a reduction in the CP1 value when comparing with a case where all the NMR shifts were perfectly repro-

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

duced by the calculations. This is reasonable, since the agreement is worse. On the other hand, in cases where the Δ_{calc} displays the same sign but larger magnitude than Δ_{exp}, the value of CP1 is increased, even though the agreement is also worse. In other words, the CP1 method lacks an upper threshold. To circumvent this problem, Smith and Goodman generated two new comparison parameters by replacing the term $(\Delta_{exp}\Delta_{calc})$ for $[(\Delta_{exp})^3/\Delta_{calc}]$ whenever the conflict situation $\Delta_{calc} > \Delta_{exp}$ is met, affording the CP2 parameter.

$$\mathrm{CP2} = \frac{\sum f_2(\Delta_{exp}\Delta_{calc})}{\sum (\Delta_{exp})^2} \tag{27}$$

where:

$$f_2(\Delta_{exp}\Delta_{calc}) = \left\{ \begin{array}{c} (\Delta_{exp})^3/\Delta_{calc} \text{ if } |\Delta_{calc}| > |\Delta_{exp}| \\ \text{or } \Delta_{exp}\Delta_{calc} \text{ otherwise} \end{array} \right\} \tag{28}$$

CP2 has the advantage of generating values in the range between +1, for a perfect agreement, and −1, for a perfect disagreement, where the candidates have been assigned the wrong way around.

Additionally, the authors have also introduced the CP3 parameter, which differs from CP2 in the fact that the correction is applied only if Δ_{calc} is larger in magnitude than Δ_{exp} and has the correct sign. In other words, it includes an extra requirement that the correction be applied only when both Δ_{calc} and Δ_{exp} present the same sign.

$$\mathrm{CP3} = \frac{\sum f_3(\Delta_{exp}\Delta_{calc})}{\sum (\Delta_{exp})^2} \tag{29}$$

where:

$$f_3(\Delta_{exp}\Delta_{calc}) = \left\{ \begin{array}{c} (\Delta_{exp})^3/\Delta_{calc} \text{ if } \Delta_{calc} > \Delta_{exp} \\ \text{or } \Delta_{exp}\Delta_{calc} \text{ otherwise} \end{array} \right\} \tag{30}$$

Smith and Goodman tested the robustness of the CP*n* methods in their paper[64] and concluded that CP3 afforded better results when compared to the CP1 and CP2 methods. The CP3 parameter also provides assignments with much greater confidence in comparison with the more basic statistic conventional parameters, such as R^2, MAE and RMSD. Moreover, the authors provided a quantification of the probability that the assignment made by CP3 method is the right one using Bayes' theorem:

$$P(\mathrm{AC}_1|R_1,R_2) = \frac{P(R_1|\mathrm{AC}_1)P(R_2|\mathrm{AC}_1)P(\mathrm{AC}_1)}{\sum_i P(R_1|\mathrm{AC}_i)P(R_2|\mathrm{AC}_i)P(\mathrm{AC}_i)} \tag{31}$$

where $P(\mathrm{AC}_1|R_1,R_2)$ is the probability that the proposed assignment ($A = a$ & $B = b$) is correct. AC_1, R_1 and R_2 are the parameter values (R^2, MAE, CP3, *etc.*) used for each assignment (AC_1 and AC_2), and $P(R_i(\mathrm{AC}_i)$ is the probability of getting a given parameter value assuming the assignment is correct (or incorrect) and that the terms R_i are assumed as independent variables that have a normal distribution (see Smith and Goodman 2009 for more detail[64]). Thus, Bayes' theorem increases the certainty of the decision made by CP3. The authors developed an applet that is available freely at http://www-jmg.ch.cam.ac.uk/tools/nmr/ for the CP3 method. Recent examples of the application of the CP3 procedure in natural product chemistry are available.[156–161]

DP4. Despite the great success of CP3, it requires two sets of experimental data to be assigned to two possible structures. However, in natural product chemistry, we often deal with the isolated compound as a single stereoisomer, and assigning their relative stereochemistry is not an easy task. In other words, we are dealing with the problem of having only one set of experimental data to assign to several possible candidates; in this situation, the CP3 method cannot be employed. With this in mind, the same authors of CP3, *i.e.* Smith and Goodman, developed a methodology that is considered a breakthrough and improved the research field, called the DP4 method. They brought the advantages of CP3, that is, the ability to make clear assignments with a high and quantifiable level of confidence, which could be applied to the situation in which only one set of experimental data is available.[162] For this, the authors had to deal with two crucial questions. What is the best way to determine the candidate presenting the NMR calculated shifts that provide the best fit to the experimental data? Moreover, once achieving a conclusion, can we assign a numerical probability to each candidate structure? Although the agreement between calculated and experimental data for a given assignment can be evaluated by the basic statistical parameters, such as R^2, MAE and RMSD, the DP4 probability tackles the problem in a different way, giving a direct probability without using additional parameters.[162]

DP4 is based on the errors e between experimental chemical shifts (δ_{exp}) and the calculated shifts on a scaled form (δ_{scal}) *i.e.*, $\delta_{scal} - \delta_{exp}$. To obtain the δ_{scal}, it is necessary to plot the calculated chemical shifts (δ_{calc}) against the δ_{exp} for the data being compared. After obtaining the slope and the intercept of the regression line, δ_{scal} is obtained from the ratio (δ_{calc} − intercept)/slope. Then, it is assumed that the error e in the shift of each atom in the molecule is an independent random variable obeying a t distribution with mean μ (as consequence of the linear scaling procedure, $\mu = 0$ is considered), standard deviation σ and degrees of freedom ν. With these assumptions, the probability of each observed error between δ_{exp} and δ_{scal} is calculated. For a molecule with N nuclei, the probability associated to each i^{th} error is calculated for each candidate structure by the cumulative probability distribution function defined by the terms σ and ν. This way, the multiplication of all individual i^{th} probabilities gives the overall probability for that molecule.[27]

$$P(i|\delta_1,\delta_2,\ldots,\delta_N) = \frac{\prod_{k=1}^{N}\left(1 - T^{\nu}\left(\left|\left(\delta^i_{\mathrm{scal},k} - \delta_{\mathrm{exp},k}\right) - \mu\right|/\sigma\right)\right)}{\sum_{j=1}^{m}\left[\prod_{k=1}^{N}\left(1 - T^{\nu}\left(\left|\left(\delta^j_{\mathrm{scal},k} - \delta_{\mathrm{exp},k}\right) - \mu\right|/\sigma\right)\right)\right]} \tag{32}$$

The DP4 equation provides the probability that a candidate structure i (out of m possible candidates) is the correct one from the NMR experimental shifts δ_1, δ_2, ..., δ_N. T^{ν} is the cumu-

 This journal is © the Partner Organisations 2021

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

lative t distribution function with ν degrees of freedom. $\delta_{exp,k}$ is the experimental shift for nucleus k (where k goes from 1 to N) and $\delta^{j}_{scaled,k}$ is the scaled calculated shift of nucleus k in candidate structure j.[162] Finally, in the same way as in CP3, Bayes's theorem is used to convert the resulting probabilities obtained in DP4 to a set of probabilities of correctness for each candidate structure.[162]

To obtain the key DP4 terms, σ and ν, the authors computed 1717 ^{13}C and 1794 ^{1}H chemical shifts using 117 known organic molecules at the B3LYP/6-31G(d,p)//MMFF level, which was the same used in the CP3 method. The resulting errors between the experimental and calculated shifts were then fitted to t distributions, obtaining σ = 2.306 ppm for ^{13}C shifts and 0.185 ppm for ^{1}H, and ν = 11.38 for ^{13}C and 14.18 for ^{1}H. However, after the publication of the original DP4 paper,[162] the authors increased their database by adding more compounds and providing improved values for σ and ν.[27] As was done for CP3, to simplify the calculation process, the authors developed a Java applet freely available at http://www-jmg.ch.cam.ac.uk/tools/nmr/DP4/. Recent examples of the application of the DP4 procedure in natural product chemistry are available.[163–171]

A few years later, MacGregor and coworkers successfully extended the DP4 method to analyse natural products where the structure displayed such a high degree of complexity that it can be treated in the form of several well-separated virtual fragments, to which the DP4 method can be applied separately. As this analysis represents an extension of the applicability of the DP4 method, the authors decided to distinguish it from the original DP4 method by using the descriptor DP4*f*, since it is a fragment-based method.[172]

After the great success of DP4 in natural product chemistry, Goodman and coworkers decided to test the applicability of the method to drug compounds, which is another class of compounds that also exhibits diverse stereochemistry. Although the fact that the DP4 method has been developed for natural products and their fragments, the authors showed that DP4 can be also widely used for drug compounds. Besides that, the authors have also created an automated process in the form of a Python script, to be used in open-source software as alternatives to MacroModel and Gaussian, which they called PyDP4. In this script, the only required user interaction is the creation of the 3D geometry of the candidate structures in the form of an SDF file. After this interative step, PyDP4 automates the rest of the process. From the SDF files, the script prepares the input files for conformational search, which can be done by open-source software such as TINKER. After completing the conformational analysis step, low energy conformations are selected and input files (for Gaussian, Jaguar or the open source NWChem) are prepared to run. Finally, it extracts the predicted NMR chemical shifts, performs the Boltzmann averaging, combines the computational with experimental NMR chemical shift data, and applies it directly to the DP4 analysis. The authors emphasise that this fully automated open-source version of the DP4 workflow should significantly improve the accessibility of DP4 analysis to occasional users and chemists in all areas of organic chemistry.[173] The PyDP4 script is freely available on the author's group website (http://www.jmg.ch.cam.ac.uk/tools/nmr) and also at the GitHub webpage (https://github.com/KristapsE/PyDP4).

Modifications to the original DP4 method: DP4 +, DP4.2, DICE AND *J*-DP4

DP4 + :. Since the launching of DP4, it has been extensively used in natural product chemistry, either to confirm or to correct the structural identification of several complex metabolites. It has been considered one of the best methods to solve stereochemical problems, which clearly is an important and difficult task. Nevertheless, in many cases where closely related isomeric compounds are possible candidates, DP4 showed modest performance, sometimes pointing out the incorrect isomer or affording inconsistent or unreliable results. In fact, in the original paper itself,[162] the authors included some examples of cases where the DP4 method was not successfully applied. Therefore, Grimblat and coworkers focused their attention on building an improved probability based on DP4 foundations.[174] After a careful evaluation of the method, Grimblat and coworkers identified two potential drawbacks in the DP4 architecture: the level of theory applied and the exclusive use of scaled shifts.[174]

Regarding the level of theory, Smith and Goodman used the B3LYP/6-31G(d,p)//MMFF level in the original DP4 formulation. This level of theory was chosen because it provides good results in calculating the NMR shifts with a low computational cost, avoiding expensive *ab initio* or DFT treatments for the geometry optimisation step. However, it is precisely this step of geometry optimisation at a molecular mechanics level in the original DP4 that would sometimes generate modest performance of the method.[174,175] Thus, obtaining more refined geometries and applying a better treatment for the GIAO NMR calculations at higher levels, like the inclusion of solvent effects and more refined DFT functionals and basis sets, would significantly improve the accuracy of the overall NMR prediction step and, therefore, would generate more reliable DP4 probabilities.[174]

Regarding the exclusive use of NMR shifts under the scaled form, this is a common approach to remove systematic errors, since when using linear scaling the corrected shifts become closer to the experimental values. However, doing this, the magnitude of error is now independent of the chemical environment (hybridisation), which normally does not occur. Moreover, in some cases, linear scaling may increase the risk of false positives, since it can happen that a wrong candidate displays an unforeseen better fit with the experimental data than the correct candidate. In other words, sometimes an incorrect structure may randomly provide a better fit in the scaled form. Thus, adding the NMR shifts under the 'pure' unscaled form to the DP4 equation would emphasise the environmental contrast between all the candidates and might reduce these potential problems.[174]

Taking these issues into account, Grimblat and coworkers developed a modified DP4 version, called DP4+, in which the probability of candidate i (out of m possible candidates) being

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

the correct one, is defined by ($P(i)$), which is a function of the corresponding probabilities computed using scaled (sDP4+) and unscaled (uDP4+) NMR shift forms, where each term can be computed using the original DP4 formalism:

$$P(i) = \frac{\prod_{k=1}^{N}\left[1 - T^{v}_{\text{scal}}\left(\left(\delta^{i}_{\text{scal},k} - \delta_{\text{exp},k}\right)/\sigma_{\text{scal}}\right)\right]\left[1 - T^{v}_{\text{u-spx}}\left(\left((\delta^{i}_{u,k} - \delta_{\text{exp},k}) - \mu_{\text{u-spx}}\right)/\sigma_{\text{u-spx}}\right)\right]}{\sum_{j=1}^{m}\prod_{k=1}^{N}\left[1 - T^{v}_{\text{scal}}\left((\delta^{j}_{\text{scal},k} - \delta_{\text{exp},k})/\sigma_{\text{scal}}\right)\right]\left[1 - T^{v}_{\text{u-spx}}\left(\left((\delta^{j}_{u,k} - \delta_{\text{exp},k})\right) - \mu_{\text{u-spx}})/\sigma_{\text{u-spx}}\right)\right]} \quad (33)$$

where T^{v}, σ, μ and δ were calculated from B3LYP/6-31G(d) geometries.

The main equation of DP4+ can be decomposed into two terms; the first one, sDP4+, presents the scaling shifts contribution:

$$\text{sDP4} : P(i) = \frac{\prod_{k=1}^{N}\left[1 - T^{v}_{\text{scal}}\left(\delta^{i}_{\text{scal},k} - \delta_{\text{exp},k})/\sigma_{\text{scal}}\right)\right]}{\sum_{j=1}^{m}\prod_{k=1}^{N}\left[1 - T^{v}_{\text{scal}}\left(\delta^{j}_{\text{scal},k} - \delta_{\text{exp},k})/\sigma_{\text{scal}}\right)\right]} \quad (34)$$

while the second term, uDP4+, presents the unscaling shifts contribution:

$$\text{uDP4+} : P(i) = \frac{\prod_{k=1}^{N}\left[1 - T^{v}_{\text{u-spx}}\left(\left((\delta^{i}_{u,k} - \delta_{\text{exp},k}) - \mu_{\text{u-spx}}\right)/\sigma_{\text{u-spx}}\right)\right]}{\sum_{j=1}^{m}\prod_{k=1}^{N}\left[1 - T^{v}_{\text{u-spx}}\left(\left((\delta^{j}_{u,k} - \delta_{\text{exp},k}) - \mu_{\text{u-spx}}\right)/\sigma_{\text{u-spx}}\right)\right]} \quad (35)$$

It is important to highlight, however, that the second term (uDP4+) presents two important modifications from the originally proposed DP4. First, as we are no longer dealing with a linear scaling procedure, the centre of the distribution μ is no longer centred on zero as in DP4. Second, it was found that the full set of unscaled errors did not follow a *t*-distributed series; however, each sp^2 and sp^3 series did. Therefore, the uDP4+ term must be decomposed regarding the hybridisation of the nucleus in question.[174] This is why $\sigma_{\text{u-spx}}$, $\nu_{\text{u-spx}}$ and $\mu_{\text{u-spx}}$ are used in the form of u-spx. Hence, to build the DP4+ probability, Grimblat and coworkers defined 16 parameters for each level of theory used: eight for the ^{13}C distributions (ν_{scal}, σ_{scal}, $\nu_{\text{u-sp}^2}$, $\mu_{\text{u-sp}^2}$, $\sigma_{\text{u-sp}^2}$, $\nu_{\text{u-sp}^3}$, $\mu_{\text{u-sp}^3}$, and $\sigma_{\text{u-sp}^3}$) followed by their corresponding eight parameters for the ^{1}H series.

Besides this, Grimblat and coworkers realised that, in most of cases where the original DP4 was applied, a high diversity of different methods was used. The author showed that in only 27% of the cases the DP4 procedure was conducted at the correct level of theory at which DP4 was formulated (B3LYP/6-31G(d,p)//MMFF). However, it should be emphasised that the DP4 probability values are directly dependent on the obtained parameters σ and ν, which presents a sharp relationship with the level of theory employed to compute the NMR shifts. Thus, the original DP4 method must be computed at the originally proposed theoretical level unless the σ and ν parameters are deduced for any other alternative method.[27] Although this fact in no way excludes the conclusions of all studies in which more accurate NMR procedures were employed, the obtained DP4 values might have been different if the correct theoretical level was applied. Because of this, Grimblat and coworkers, to provide more freedom to users to choose a method, decided to include different levels of theory in the development of the DP4+ procedure. At the geometry optimisation step, the B3LYP/6-31G(d) level was used. For the GIAO single point NMR calculations, on the other hand, 24 different levels of theory were used, combining two functionals (B3LYP and mPW1PW91) and six basis sets [6-31G(d), 6-31G(d,p), 6-31+G(d,p), 6-311G(d), 6-311G(d,p), and 6-311+G(d,p)] that were also computed in the gas phase and in solution (with the PCM formalism, using chloroform as the solvent).

To demonstrate the improved performance of DP4+, the authors applied it to a set of 48 challenging compounds for which DP4 afforded modest performance. The DP4+ procedure was evaluated using all 24 different levels of theory. Impressively, in all cases, the DP4+ procedure performed better in relation to the original DP4, with the PCM/mPW1PW91/6-31+G(d,p)//B3LYP/6-31G(d) level of theory being the most recommended one.

In the same way that the authors of CP3 and DP4 did, to facilitate the overall process, Grimblat and coworkers created an Excel spreadsheet, which is freely available at sarotti-NMR.weebly.com. Recent examples of the application of the DP4+ procedure to natural products are available.[176–181]

DP4.2. Following the success of DP4+, other groups also reported modifications of the original DP4 model. Ermanis and coworkers reported a DP4 improvement under the pyDP4 workflow foundation, called DP4.2. In this improved method, the NMR chemical shifts are calculated at a higher level of theory, *i.e.* mPW1PW91/6-311G(d). However, the relative energies of the conformers used in the Boltzmann's analysis were obtained at the M06-2X/6-31G(d,p) level calculated directly from the geometries obtained in the conformational analysis step at the MMFF level. This new method has been tested on a significant number of structures in which the original DP4 did not show good results. In all cases, the new DP4.2 method was able to identify the correct structure with a realistic level of confidence and showed excellent performance in all cases where the original DP4 was successful. This fact indicates that DP4.2 is a more general method than the original DP4. However, it is important to highlight that most of the improved molecules by DP4.2 are of medicinal origin. Thus, the authors believe that DP4.2 displayed an even wider application in the drug discovery field.[182] Later in 2019, the same

 This journal is © the Partner Organisations 2021

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

authors suggested that DP4.2, which uses MMFF geometries with mPW1PW91 for NMR chemical shift calculation and M06-2X for conformer energies, could be improved by adding a geometry optimisation step at the B3LYP/6-31G(d,p) level, thus obtaining the best method: B3LYP for geometry optimisation, mPW1PW91 for NMR chemical shifts calculations and M06-2X for conformation energies. The authors added this improvement to the automated NMR calculation workflow PyDP4.[183] The latest version of the PyDP4 script can be obtained from the author's website (http://www.jmg.ch.cam.ac.uk/tools/nmr) and also at GitHub (https://github.com/KristapsE/PyDP4).

Diastereomeric *in silico* chiral elucidation (DiCE). Another interesting modification to the original DP4 method was proposed by Xin and coworkers in 2018. To improve the original method, the authors developed a probability theory process called diastereomeric *in silico* chiral elucidation (DiCE). This new method performs the calculation of NMR chemical shifts at the B3LYP/cc-pVDZ level of theory, considering the PCM method to include solvent effects in the geometry optimisation step for each selected conformer and by applying the same probability theory approach based on statistical analysis of calculation errors, just like the DP4 method. Two factors distinguish DiCE from DP4 and DP4+ methods. The first is the use of generalised scaling factors in the conversion of shield constants into chemical shifts in comparison to the internal linear scale within each molecule, as described in DP4 and DP4+. The second difference involves the use of probability factors in DiCE calculations obtained for ^{15}N nuclei. Since ^{15}N is typically not as prevalent in small organic molecules (but can be in natural products), the use of internal scaling factors can compromise the outcome of the predicted chemical shifts. Therefore, the use of generalised scaling factors is important to improve the accuracy of ^{15}N NMR chemical shifts and associated calculated probabilities. The authors tested the method using a set of structurally complex natural products and synthetic intermediates that were chosen due to their unsatisfactory DP4 results. Most of the selected structures were also highlighted in the DP4+ study. It is interesting to highlight that the final results using DiCE were better than the original DP4 method and showed overall performance comparable to the DP4+ method.[184]

J-DP4. The improvements made to the DP4 foundations by Grimblat and coworkers (DP4+), Ermanis and coworkers (DP4.2) and Xin and coworkers (DiCE) definitely outperformed the original method. However, as previously mentioned, these improvements were accompanied by a higher computational cost. Consequently, Grimblat and co-workers started thinking about how to develop a new method that could also offer better results than the original DP4 method, while keeping the associated computational cost to a minimum.

Vicinal couplings ($^3J_{HH}$) can be easily obtained from a ^{1}H NMR spectrum, and the information provided are extremely valuable for conformational and stereochemical analysis. Therefore, considerable effort was made in quantum chemical simulations of J couplings, which generated great advances in configurational determination of natural products.[113,132] Based on that, Grimblat and coworkers recently decided to explore the possibility of incorporating vicinal couplings ($^3J_{HH}$) into DP4 formalism, something that had never been done before.[185]

To this end, the authors thought of two possibilities for incorporating information from $^3J_{HH}$ into the standard DP4 formalism, called direct (dJ-DP4) and indirect (iJ-DP4). In the first one, dJ-DP4, besides calculating the NMR chemical shifts (^{1}H and ^{13}C), one also needs to calculate $^3J_{HH}$ at the DFT level to be later correlated with experimental values through a modified DP4 equation, which must be able to provide the probability of the difference between the calculated and experimental $^3J_{HH}$ values. Alternatively, in the second one, iJ-DP4, experimental $^3J_{HH}$ values are used only to introduce geometric restrictions into the conformational analysis step. In this way, only compatible conformations will be selected for the NMR chemical shift calculation step, which will be further evaluated using the standard DP4 approach.[185]

In both cases, the error in the original DP4 analysis related to the capabilities of the theory level employed to adequately represent the real conformational distribution of the target molecule will be smaller, since it reduces the dependence of the theoretically calculated energy values used to obtain the relative populations of the possible conformations.

In the dJ-DP4 approach, the probability of candidate i (out of m possible candidates) being the correct one involves a modification in the original DP4 formalism (left product) by including the new probability term, given by $^3J_{HH}$ (right product) (see eqn (36)).

$$\mathrm{d}J\text{-DP4}_{(i)} = \frac{\prod_{k=1}^{N}\left(1 - T^{\nu,\delta}(\delta^{i}_{\mathrm{scal},k} - \delta_{\mathrm{exp},k})/\sigma_{\delta,\mathrm{scal}}\right)\prod_{k=1}^{0}\left(1 - T^{\nu,J}(J^{i}_{\mathrm{scal},k} - J_{\mathrm{exp},k})/\sigma_{J,\mathrm{scal}}\right)}{\sum_{j=1}^{m}\left[\prod_{k=1}^{N}\left(1 - T^{\nu,\delta}(\delta^{i}_{\mathrm{scal},k} - \delta_{\mathrm{exp},k})/\sigma_{\delta,\mathrm{scal}}\right)\prod_{k=1}^{0}\left(1 - T^{\nu,J}(J^{i}_{\mathrm{scal},k} - J_{\mathrm{exp},k})/\sigma_{J,\mathrm{scal}}\right)\right]} \quad (36)$$

The dJ-DP4 probability can be determined by the degrees of freedom (ν) and standard deviation (σ) of the NMR chemical shift and $^3J_{HH}$ terms at a given level of theory. The ν and σ terms corresponding to NMR ^{1}H and ^{13}C were the same used in the original DP4 formalism, while the $^3J_{HH}$ ν and σ values were estimated as 0.992 and 3.06, respectively, by correlating the experimental and calculated J values using a data set. To follow a t-distributed series, the calculated $^3J_{HH}$ values were used under the scaled form according to $J_{scal} = (J_{calc} - b)/m$. However, it is important to highlight that the authors chose a generalised linear scaling procedure proposed by Xin and coworkers during the development of DiCE, which uses fixed m and b values resulting from a large data set, instead of

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

the internal scaling procedure of the original DP4 method.[184,185]

In the indirect approach, i*J*-DP4 will only be considered in the conformational analysis step on the subset of candidates compatible with the experimental $^3J_{HH}$ values. Since the experimental $^3J_{HH}$ is the result of averaging the contributions made by different conformers, according to the Karplus equation, a large $^3J_{HH}$ value (*i.e.*, >8–9 Hz) means that the protons are almost coplanar, while a low $^3J_{HH}$ value (*i.e.*, <4 Hz) means that the protons are almost perpendicular. This information can be used to constrain the conformational sampling. However, an intermediate $^3J_{HH}$ value (between 4 and 8 Hz) makes it impossible to use it as a guide at the conformational search step, since it results from the average of large and small $^3J_{HH}$ values. This is the main practical limitation of the i*J*-DP4 approach and the reason why some systems could not be evaluated by this method.[185]

To explore the effectiveness of both approaches, the authors applied it to a set of 69 selected examples, featuring broad structural variety. To compare it with the original DP4 procedure, all simulations (NMR ^{1}H, ^{13}C and $^3J_{HH}$) were carried out at the same proposed theoretical level (B3LYP/6-31G(d,p)//MMFF) where $^3J_{HH}$ values were calculated, considering just the FC term, as proposed by Bally and Rablen in 2011.[113] For this selected data set, the original DP4 method provided 75% accuracy, while the probability of correct answers obtained by the d*J*-DP4 method reached 96%, a significant enhancement. In addition, the original DP4 method generated probabilities of success with higher confidence (>95%) in only 39% of cases, while the d*J*-DP4 reached 70%, a clearly higher value. Moreover, in some examples where the original DP4 generated an incorrect assignment, the d*J*-DP4 was able to reverse the assignment. On the other hand, in only three cases within the dataset did d*J*-DP4 provide more modest results than DP4, although the right isomers were still identified as the most feasible candidate.[185] The indirect approach (i*J*-DP4) was also evaluated. However, only 61 of the 69 selected examples were compatible to guide the conformational search with experimental $^3J_{HH}$ values. In those compatibles cases, the authors demonstrated that the use of only one experimental $^3J_{HH}$ value is enough to afford good results in terms of DP4 performance. The use of a second $^3J_{HH}$ value afforded an even more pronounced effect. In summary, i*J*-DP4 afforded generally better results than the original DP4 with a significantly reduced computational cost. This could be helpful to study large molecules in which a higher theoretical level would be a problem.[185]

Although the d*J*-DP4 approach offers better performance than the i*J*-DP4 approach, i*J*-DP4 has the advantage of being much more accessible in terms of computational cost. Thus, besides everything that has already been pointed out, the authors have already anticipated that mixing both approaches would be an interesting way to afford better results at a low computational cost. Thus, a third alternative was created, called the i*J*/d*J*-DP4 approach, where relevant $^3J_{HH}$ couplings can be used to guide the conformational analysis (i*J*-DP4) and the resulting selected conformations are employed to calculate all the NMR parameters (^{1}H, ^{13}C and $^3J_{HH}$) to be used in the d*J*-DP4 approach. To avoid redundancy, $^3J_{HH}$ values that were used to constrain the conformational search were neglected in the d*J*-DP4 calculation step. This hybrid approach afforded excellent results, keeping the better performance of d*J*-DP4 at a fraction of the computational cost, as in i*J*-DP4. To conclude, using only one $^3J_{HH}$ value to restrict conformational sampling, the hybrid i*J*/d*J*-DP4 approach displayed better results than the original DP4 method at about the half of the computational cost. By using two $^3J_{HH}$ values in the conformational sampling step, an even higher performance was obtained with an impressively lower computational cost than the original DP4 method.[185]

As usual, the authors created and made available free of charge an Excel file as an effort to facilitate *J*-DP4 calculations at sarotti-NMR.weebly.com, iquir-conicet.gov.ar/Sarotti-NMR and the digital CSIC repository. Recent examples of the application of the *J*-DP4 procedure to natural product chemistry are available.[186]

Other tools

ANN-PRA. So far, different methods for the stereochemical assignment of natural products that revolutionised the NMR research field have been discussed. These methodologies are based on the use of sophisticated statistical procedures in the form of 'comparison-based' approaches, where the analysis is entirely based on the degree of correlation between experimental data and two or more possible theoretical candidates. However, another tool, called ANN-PRA, which is based on a completely different protocol, deserves to be highlighted. This method can also be used to minimise the risk of errors in elucidating the structure of natural products.

In 2013, Ariel Sarotti realised that the results obtained by comparison-based statistical methods, like CP3 and DP4 among others, are relative and not absolute. That is, the information provided is that one candidate is more likely than another one and not that a candidate is right or wrong. Therefore, in situations where the correct structure is not present in the list of possible candidates, the 'comparison-based' methods will provide a false positive.[174] With that in mind, Sarotti developed a procedure that could be used directly as a structural validation, classifying the proposed structure as correct or incorrect. This is extremely important in detecting structural misassignment in natural products. Conceptually, the methodology proposed by Sarotti is based on pattern recognition analysis (PRA) using artificial neural networks (ANNs). An ANN is a mathematical model that simulates the behaviour of a biological brain with several interconnected artificial neurons. ANNs have the ability to learn from the data and then can be used to recognise patterns; this has become an interesting approach to structural validation using NMR chemical shift calculations. The methodology uses a set of different statistical parameters (like MAE, standard deviation, maximum error, slope, intercept and R^2, among others) obtained from the correlation between experimental and calculated ^{13}C NMR data for the structure to be tested at the MM+,

 This journal is © the Partner Organisations 2021

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

AM1 or HF/3-21G level. This information is taken to an ANN previously trained to identify whether the structure is correct or incorrect. For more detailed information about the functioning of the ANN, the performed training process and the math behind the methodology, see the original reference.[187] In this article, Sarotti evaluated the methodology in 26 cases of reassigned natural products, testing the original proposed structure with its corresponding revised structures and obtained a success rate of 100% at the HF level. It is interesting to highlight that the methodology presents a very low computational cost, since it uses structures optimised at low levels of theory.[187]

The limitation of ANN-PRA is the fact that it is restricted to ^{13}C NMR data. When compared to ^{1}H data, ^{13}C chemical shifts present some advantages that facilitate the use of ANNs, *i.e.*, a spread over a wide range, relative insensitivity to solvent changes, a strong influence on steric and electronic factors in the structure and a low probability of signal overlap. These properties considerably simplify the entire procedure of ANN-PRA. Another issue that deserves attention is natural products containing heavy atoms (third row or heavier elements) in the structure. In these cases, the errors associated with the ^{13}C NMR calculation of the carbon directly attached are above average. This well-known effect, called the heavy atom effect, is caused by neglecting spin–orbit contributions from relativistic effects.[23] Since ANNs were not trained to recognise this type of effect, any carbon atom attached to a heavy atom must be excluded to avoid incorrect classification.[187]

The original ANN-PRA method, developed only for ^{13}C NMR data, provided good results in the detection of connectivity and regiochemical structural errors. However, the method demonstrated low performance when dealing with stereochemical isomerism, where the source of error is considerably more subtle. Therefore, following a conceptually similar strategy, in 2015 Zanardi and Sarotti expanded the original ANN-PRA method by including ^{1}H and ^{1}H–^{13}C HSQC experiments. In addition to adding several extra parameters to explain the fit between ^{1}H NMR and its corresponding correlation with ^{13}C NMR data, the NMR calculations were now performed at a higher level, at mPW1PW91/6-31G(d) (gas phase) and mPW1PW91/6-31G(d,p) (PCM), with geometries obtained at B3LYP/6-31G(d) level. This higher computational cost to simulating the NMR data led to an improvement in the accuracy of the analysis.[188] With these improvements, the authors demonstrated that the modified ANNs could detect structural errors at a stereochemical level, affording more than 92% of correct classification after the ANN training process. The authors evaluated the performance of the improved ANN-PRA method in challenging real cases of misassignments and obtained a rate of up to 100% correct prediction. However, the authors recognized that even having a right and wrong pair of a given molecule, ANN-PRA should not be used to compare the calculated data for both candidates with the experimental values, because doing this, the decision will be based on comparison. Conceptually, the ANN-PRA method was created only to decide whether a given putative structure (only one) is correct based on experimental data and the calculated NMR for that candidate. With that in mind, the improved ANN-PRA method can be considered an extremely useful tool for making structural decisions faster, simpler and cheaper when compared to total synthesis of the incorrectly assigned compound, which is the best practice today to solve misassignment in natural product chemistry.[188]

The math behind ANN-PRA is more complicated than in other comparison-based statistical methods, like CP3 and DP4, making calculation a difficult task. Thus, the authors incorporated all the mathematical equations that define trained ANNs in an Excel spreadsheet to simplify the whole process and made it available free of charge in the ESI of Sarotti 2013 (original ANN-PRA) and Zanardi and Sarotti 2015 (improved ANN-PRA) papers. To use the ANN-PRA procedure (original and improved version), one just needs to fill in the corresponding cells with the experimental NMR chemical shifts and with the calculated shielding tensors. After indicating which carbon (or proton attached to) is sp–sp^2 hybridised, the spreadsheet automatically calculates the unscaled and scaled chemical shifts using TMS or the multi-standard approach (methanol and benzene for sp^3 and sp–sp^2, respectively) as reference standards. Then, it computes all the necessary statistical parameters and directly introduces the results into the trained ANN to get the output value (right or wrong).[174,187,188] Recent examples of the application of the ANN-PRA procedure to natural product chemistry are available.[189,190]

Confidence levels and choice of the method. To complete this topic, it is worth mentioning that in the methods discussed above the computed probability for a given assignment reflects the certainty in relation to the chosen candidate. Thus, although there is no strictly defined cut-off, high calculated values of probability (over 90%, preferably close to 100%) or very low (close to 0%, up to 10%) are results that clearly indicate, with high confidence, that the assignment is correct or incorrect, respectively. However, in less defined situations, the certainty is reduced as the probability values move away from those limits, reaching a point with probability values around 50%, where no assignment can be made. Therefore, one has to use these methodologies as a theoretical guide to identify the most likely assignment, rather than giving them a strict mathematical meaning as a true probabilistic measure. In our opinion, results below 90% in all the methods mentioned above should be considered with caution. The ideal scenario would be to evaluate the proposed candidate in the ANN-PRA protocol to obtain a more reliable answer or take into account other NMR observables, when available.

Even so, this combination of quantum chemical calculations with sophisticated data processing has provided results that allowed stereochemical assignments of complex natural products with remarkable levels of confidence when compared to the simple aforementioned statistical parameters (R^2, MAE and RMSD). For this reason, these methods have emerged as indisputable complements to the experimental NMR spectroscopy, revolutionising and renewing the field of natural product chemistry. In conclusion, all alternatives cited herein

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

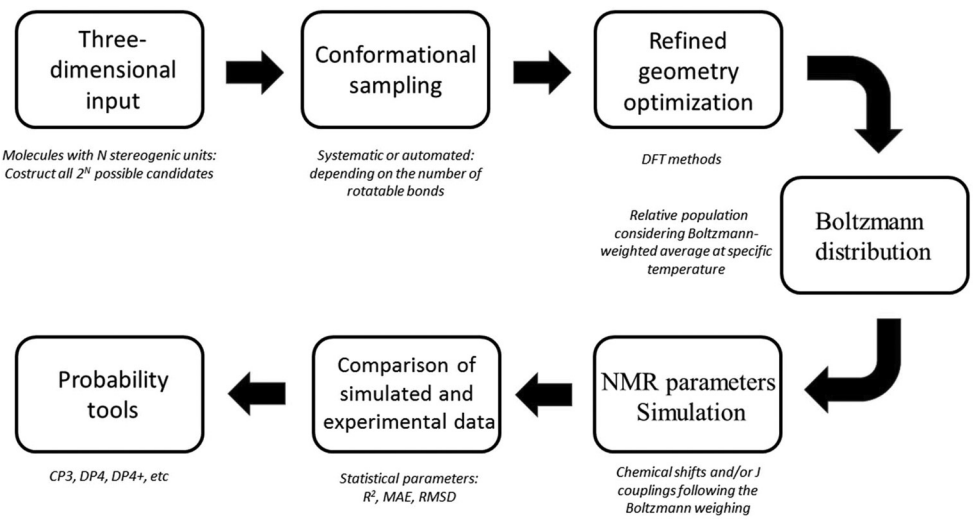
are attractive from the natural product point of view, leaving to each researcher the choice of which method is most suitable for the specific problem at hand.

II.5. Tutorial and recommendations

In 2014. Hoye's group published an original and relevant article[191] presenting a step-by-step protocol for simulating NMR parameters in order to provide the necessary guidance to a non-specialist to start such a computational study.[191] Herein, we present our suggestions for helping a non-specialist to start this type of computational study through a didactic workflow. The simulation of NMR parameters of natural products usually involves the following steps: (1) construction of an input containing information about the three-dimensional structure; (2) conformational sampling at a low-level of calculation; (3) refined geometry optimisation; (4) computation of the Boltzmann distribution; (5) NMR parameter simulations for each relevant conformer weighted by the Boltzmann distribution; (6) comparison of simulated and experimental data; (7) use of a probability tool (Scheme 1).

The initial step for NMR data simulation is to construct the input with the three-dimensional structure information for all candidates. A molecular editor software with a graphical interface may be used to draw and manipulate the chemical structures, such as ArgusLab,[192] Avogadro,[193] ChemSketch,[194] Chem3D,[195] HyperChem[196] and GaussView,[197] among others. Subsequently, conformational sampling can be performed using different types of methods, which will depend on the degree of freedom and the number of rotatable bonds for each molecule. For structures with few rotatable bonds, a systematic conformational search is indicated. However, for molecules with higher conformational mobility, an automated conformational search is recommended. The most commonly used method for NMR calculations of natural product structures is the Monte Carlo algorithm combined with molecular mechanics force fields, like MMFF and MM+, among others, incorporated into several conformational search software such as Spartan,[198] HyperChem,[196] MacroModel[199] and Conflex,[200] among others. The calculated conformers are filtered according to an energy criterion. Around 5–10 kcal mol^{-1} above the minimum energy conformer is usually enough to account for all candidates that might be relevant to the NMR data simulation. Another possibility of filtering the conformers, although less used, are based on the compatibility of the simulated and the experimental NMR data. These selected conformers are afterwards subjected to a refined geometry optimisation step using quantum mechanical methods, such as HF, DFT or a higher-level method, in which DFT at the B3LYP/6-31G(d) level is one of the most commonly used method. Next, vibrational normal modes are computed with the same theoretical method used in the latter step to ensure that the optimised structures are true minimum energy points on the PES.

After obtaining refined geometries and energies values with the higher level of theory, a new cut-off based on the Boltzmann distribution is applied to account for more than 90% of the accessible conformers that contribute to the experimental spectrum at the temperature of interest. In practice, this refers to all the conformers that coexist within a 2.5–3.0 kcal mol^{-1} free energy window relative to the most stable conformer. Finally, NMR parameters are calculated for all of them. Nowadays, a gauge-independent atomic orbitals-hybrid density functional theory (GIAO-HDFT) calculation procedure is commonly performed at different theoretical methods, in which one can use appropriate generalised functionals like B3LYP, mPW1PW91, PBE0 or M06-2X, or more specialised ones such as WC04 and WP04, created especially for ^{13}C and ^{1}H NMR chemical shifts, respectively. Standard valence basis sets of at least double-ζ quality with the addition of polarisation and diffuse functions like 6-31+G(d) or 6-311+G(d,p) are the common choices. Currently, commercial or open-source packages are available for the DFT calculations such as Gaussian,[201] TURBOMOLE,[202] Tinker[203] and GAMESS,[204,205]



Scheme 1 Flowchart of recommended steps for the simulation of the NMR parameters of natural products.

 This journal is © the Partner Organisations 2021

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

among others. After the calculation of the absolute isotropic magnetic shielding constants for all selected conformers, an internal reference should be calculated at the same level of theory. For calculations of natural product structures, tetramethylsilane (TMS) is commonly used. The predicted NMR chemical shifts are then obtained by subtracting the isotropic magnetic shielding of the molecule of interest from that of TMS. After the NMR chemical shifts of each contributing conformer have been obtained, they are weighted by their fractional Boltzmann population and summed to produce the final NMR chemical shifts for that candidate. Once calculated, the NMR chemical shifts can be directly compared to experimental data, or generalised linear scaling factors can be used to improve the simulated NMR data. The CHESHIRE webpage (http://cheshirenmr.info), created and maintained by Tantillo's group, holds a comprehensive collection of recommended linear scaling factors, parameterised at different theoretical levels in the gas phase and including solvation models for different classes of organic compounds, including natural products.

Simulation of nuclear spin–spin coupling constants (^nJ) is a much more expensive procedure than chemical shift calculations, so it is recommended to calculate this considering just the Fermi contact term. Benchmark studies suggest that hybrid functionals like B3LYP, combined with a double-ζ basis set, along with a polarised function, such as 6-31G(d,p), provide the most accurate results. Once calculated, it can be directly compared to experimental data, or linear scaling factors can also be included. Currently, DU4, DU8, DU8C and DU8+ are the best performing methods for scaling calculated ^nJ. Finally, after obtaining the scaled chemical shifts and/or ^nJ, the basic statistical parameters (R^2, MAE and RMDS) can be obtained by comparison with the experimental data. To improve the confidence of a given assignment made with these statistic descriptors, one can choose one of the probability tools available, such as DP4, DP4+ or *J*-DP4, among others. Results with a probability value over 90% (preferably closer to 100%) clearly indicate, with high confidence, that the candidate chosen at the assignment is correct. Then, it is necessary to further confirm the assignment, a step where the ANN-PRA protocol can be applied.

III. Applications

In this section, we present a recent comprehensive list of natural compounds in which their structure elucidation was subjected to the simulation of NMR parameters in combination with the new tools available. We believe that new researchers, as well as established ones, will benefit from these periodic reviews of the state-of-the-art, showing recent advances in the research field. As natural products represent an extremely broad topic, we found it difficult to classify the studies reported here into limited categories in relation to each class. For this reason, we carefully chose the examples discussed below and placed them in sections according to the relevance of each study, dividing them into new natural products, new skeletons, complex cases, marine natural products, structural reassignment, and studies that used parametric corrections.

III.1 New natural products

Flabellipparicine: DP4+[206]

In 2018, Cai and co-workers performed a chemical investigation on the stems of *Tabernaemontana divaricate*, belonging to the Apocynaceae family, which led to the isolation of four new monoterpenoid bisindole alkaloids, together with 10 known alkaloids. Among them, one of the new monoterpenoid bisindole alkaloids, called flabellipparicine, stands out for presenting an interesting flabelliformide-apparicine-type skeleton. By the evaluation of HR-ESIMS, IR and UV spectra, and ^{1}H NMR data, it was possible to affirm that flabellipparicine was an asymmetric dimeric indole alkaloid with two unsubstituted aromatic rings, indicating an uncommon attachment between the two units, which is different from the usual vobasinyl-ibogan-type bisindole alkaloids. After an extensive analysis of 1D and 2D NMR spectra, based on ^{1}H, ^{13}C HSQC, COSY, ^{1}H–^{15}N HMBC and NOESY data, the structures of each monomeric units I and II and their connectivities were elucidated. Assuming monomeric unit I had the (3*S**,15*R**,20*S**) relative configuration, there should be four possible structures for flabellipparicine: (3*S**,15*R**,20*S**,15′*R**,16′*R**), (3*S**,15*R**,20*S**,15′*S**,16′*R**), (3*S**,15*R**,20*S**,15′*S**,16′*S**), and (3*S**,15*R**,20*S**,15′*R**,16′*S**). To define the absolute configuration, the authors decided to apply the TD-DFT ECD method on the four possible diastereomers. After a preliminary conformational search at the MMFF level, the compound indicated a relatively low flexibility, with only 6–18 low-energy initial conformers. These conformers were reoptimised at the B3LYP/6-31G(d) level *in vacuo*. The ECD spectra were simulated for each set of conformers of all candidates at the CAM-B3LYP/TZVP level PCM (MeOH). The simulated spectra for diastereoisomers (3*S*,15*R*,20*S*,15′*R*,16′*R*) and (3*S*,15*R*,20*S*,15′*S*,16′*R*) yielded mirror-image agreement with the experimental ECD spectra of the isolated compound, while those of diastereomers (3*S*,15*R*,20*S*,15′*R*,16′*S*) and (3*S*,15*R*,20*S*,15′*S*,16′*S*) were similar to the experimental one. This fact clearly suggests that the ECD spectra are governed mainly by the newly formed C-16′ chirality centre, which determines the relative orientation of the two indole chromophores. In view of this, ECD cannot be used to discriminate between possible diastereoisomers. However, this analysis by ECD allowed the authors to state that the absolute configuration of C-16′ should be *S*. To determine which possible diastereoisomer is the correct one, the authors analysed the interatomic distances of all low-energy conformers found for each diastereoisomer to account for the experimentally observed NOE correlations. With this analysis, only the diastereoisomer (3*S*,15*R*,20*S*,15′*R*,16′*R*) matched well with the NOE correlations, with all key interatomic distances displaying less than 3.0 Å. To confirm this, the NMR calculations of the four candidate diastereoisomers were carried out at the PCM/mPW1PW91/6-31+G(d,p)//B3LYP/6-31G(d) level of theory.

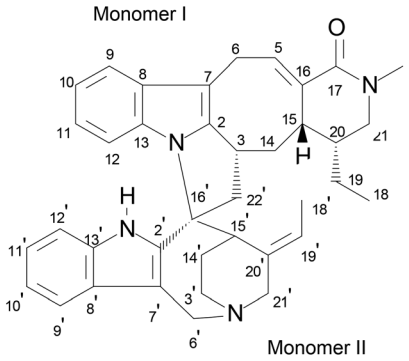


Fig. 2 Structure of flabellipparicine.

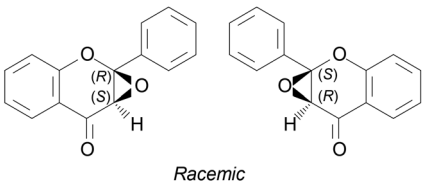


Fig. 3 Enantiomeric forms of the isolated epoxyflavone.

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

As expected, the isomer featuring the (3*S*,15*R*,20*S*,15′*R*,16′*R*) relative configuration displayed the closest match between experimental and computed NMR shifts, indicated by the smallest MAE value of 2.2 ppm for ^{13}C NMR and 0.13 ppm for ^{1}H NMR data. However, as all the MAE values for the remaining diastereoisomers were close to each other (2.3–2.6 and 0.19–0.21 ppm for ^{13}C and ^{1}H NMR, respectively), the author decided to obtain confirmation using the DP4+ method. The probability value obtained strongly supported the most likely relative configuration of flabellipparicine as (3*S*,15*R*,20*S*,15′*R*,16′*R*) with a high degree of confidence (>99.9%). Therefore, based on the relative configuration proposed by DP4+ and NOE data, and in view of the information given by the ECD data that the chiral centre C-16′ must present an *S* configuration, the absolute configuration of flabellipparicine was defined as the enantiomer (3*R*,15*S*,20*R*,15′*S*,16′*S*) (Fig. 2).

Epoxyflavone: DP4+[207]

In 2020, Alilou and co-workers carried out a phytochemical investigation of the aerial parts of *D. diapensifolia* Bioss., which led to the isolation of 24 phenolic compounds and one sesquiterpenoid; one of the phenolic compounds, with an epoxyflavone skeleton, is a new natural product. After an extensive HRESIMS and UV spectral analysis, a flavonoid scaffold was identified. The ^{1}H, HSQC and HMBC NMR spectra confirmed the epoxyflavone moiety. However, optical rotation measurements (zero rotation) suggested the presence of a racemate. As the obtained NOESY spectrum was not conclusive to determine the relative configuration of the chiral centres, the authors employed the DP4+ procedure using ^{13}C and ^{1}H NMR data in the unscaled and scaled form (^{13}C, ^{1}H and all data; uDP4+ and sDP4+) at mPW1PW91/6-311+(d,p) in CPCM (chloroform) for both possible diastereoisomers. The obtained DP4+ values enabled the assignment of diastereoisomer 1 (C-2*R** and C-3*S**) as the most likely one, with an overall confidence level of 100% (^{1}H on the sDP4+ form), 99.9% (^{13}C on the sDP4+ form), 100% (all data on the sDP4+ form), 80% (^{1}H on the uDP4+ form), 100% (^{13}C on the uDP4+ form) and 100% (all data on the uDP4+ form). With confidence on the relative configuration, the authors evaluated the anti-inflammatory activity of the isolated compound (Fig. 3).

III.2 Unprecedented skeletons

Isopenicins A–C: ANN-PRA protocol[190]

In 2019, Tang and co-workers isolated three meroterpenoids from the plant endophytic fungus *Penicillium* sp. sh18, called isopenicins A–C, with antitumor activity. The isolated compounds present two types of unprecedented terpenoid–polyketide hybrid skeletons, in which the complex and intriguing structures are related to a broad spectrum of biological properties.

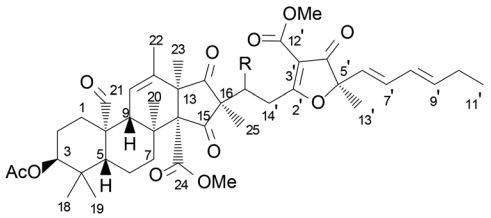
The planar structure of isopenicin A was proposed based on extensive analysis of HRESIMS, 1D and 2D-NMR methods. However, to unambiguously determine its structure and the stereochemistry, a single crystal was obtained using vapor exchange of methanol and water in a closed tube. Subsequently, through a single-crystal X-ray diffraction experiment, a conclusive confirmation of the structure and its absolute configuration as 3*S*,5*R*,8*S*,9*R*,10*S*,13*R*,14*R*,16*R*,5′*R*,15′*R* was obtained. However, for isopenicins B and C, it was not possible to obtain suitable crystals for X-ray diffraction experiments. Based on the similarity of the HRESIMS and 1D and 2D-NMR data, the authors assumed that isopenicin B might have the same planar structure as isopenicin A, differing only in the relative configuration. After extensive ^{1}H–^{1}H COSY and HMBC data acquisition, it was proposed that the only difference in configuration should be C-15′, establishing the absolute configuration as 3*S*,5*R*,8*S*,9*R*,10*S*,13*R*,14*R*,16*R*,5′*R*,15′*S*. In the absence of an ideal crystal, the authors decided to apply the ANN-PRA protocol to validate the correctness of the proposed structure. The calculations were carried out following the instructions of ANN-PRA, *i.e.* geometry optimisation at the B3LYP/6-31G(d) level (gas phase) followed by NMR shielding tensor calculation at the mPW1PW91/6-31G(d) and mPW1PW91/6-31G(d,p) level in the gas phase and with the PCM model (chloroform), respectively. Using two reference standards (TMS and MSTD), the ANN-PRA protocol gave rise to the conclusion that the structure of isopenicin B was correct. In addition, to confirm its absolute configuration, the authors obtained a ECD spectrum associated with TDDFT calculations. For isopenicin C, the 1D and 2D NMR data suggested the same subunit as in isopenicin A and B, with a different side chain. As for compound 2, the authors used the ANN-PRA protocol with TDDFT calculations of the ECD spectrum to validate the proposed structure and to confirm the absolute configuration of isopenicin C as 3*S*,5*R*,8*S*,9*R*,10*S*,13*R*,14*R*,16*R* (Fig. 4).

Premnafulvol A: DP4+ and ANN-PRA[189]

In 2018, Pu and co-workers isolated from the aerial parts of *Premna fulva* a unique diterpenoid featuring a complex 6/5/7/

 This journal is © the Partner Organisations 2021

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.



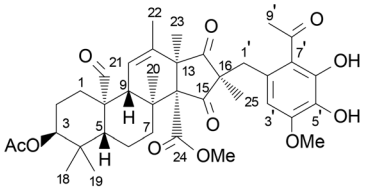


Fig. 4 Structures of isopenicin A–C.

3-fused tetracyclic carbon skeleton, called premnafulvol A. The structure was proposed based on HR-ESI-MS and 1D and 2D-NMR methods. After extensive evaluation of ^{1}H–^{1}H COSY and HMBC data, a tetracyclic 6/5/7/3 ring system was proposed. A partial relative configuration was assigned based on ROESY experiments. To elucidate the overall relative configuration, ^{1}H and ^{13}C NMR chemical shift calculations were performed for two possible diastereoisomers at the mPW1PW91/6-31G(d,p) level using PCM to simulate solvent effects (pyridine). Subsequently, the experimental and calculated NMR chemical shifts data were further analysed by the DP4+ method. The DP4+ value obtained using both the ^{13}C and ^{1}H NMR data in the unscaled form (^{13}C, ^{1}H and all data, uDP4+) enabled the assignment of diastereoisomer 1a as the most likely one, with an overall confidence level of 100%. In addition, the NMR data were also calculated at the mPW1PW91/6-31G(d) level (gas phase) to be used in the ANN-PRA protocol. Using TMS and MSTD as reference standards, the ANN-PRA protocol was able to assign the diastereoisomer 1a as the correct one while it gave the opposite results (an incorrect assignment) for diastereoisomer 1b. Experimental and calculated ECD spectra were employed to confidently establish the absolute configuration (Fig. 5).

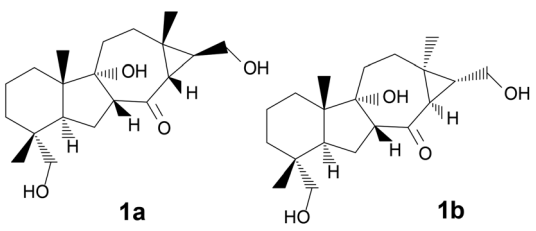


Fig. 5 Two possible diastereoisomers of premnafulvol A: 1A, left: correct and 1B, right: incorrect.

III.3 Complex cases

Cubebin: MAE, RMSD, DP4+, X-ray[154]

In 2020, Macedo and co-workers isolated (−)-cubebin, a bioactive natural dibenzylbutyrolactolic lignan found in the genus *Piper*. This compound presents several biological properties, such as analgesic, anti-inflammatory and antimicrobial, among others, which are intrinsically related to its configuration. Despite its relatively simple structure, cubebin presents three stereogenic centres, in which the configuration of the anomeric centre remained uncertain for many years, as seen in other butyrolactolic lignans. Based on the extensive analyses of CG-MS and 1D and 2D-NMR, it was possible to assign the basic structure of cubebin.

As cubebin presents three stereogenic centres, eight stereoisomers are possible, comprising four pairs of enantiomers. As each pair of enantiomers displays the same NMR chemical shift in isotropic media, to determine the correct relative configuration, the authors simulated the ^{13}C and ^{1}H NMR chemical shifts for the four possible diastereoisomers of cubebin at the mPW1PW91/6-31G(d)//mPW1PW91/6–31G(d) level, which was subsequently compared with those obtained experimentally for the isolated compound. Through these comparisons, the ^{13}C and ^{1}H NMR MAE and RMSD obtained for the four possible diastereoisomers were extremely similar, which did not allow a definitive determination of the relative configuration. This fact clearly demonstrates that although cubebin presents a relatively simple structure, determining its relative configuration is challenging based only on experimental NMR methods. Trying for a more reliable discrimination between the possible diastereoisomers, the authors employed the DP4+ procedure using both the ^{13}C and ^{1}H NMR data in the scaled form (all data, sDP4+). The obtained DP4+ values enabled the assignment of diastereoisomer 2 as the most likely one, with an overall confidence level of 97.9%, against 1.15%, 0.87% and 0.08% for diastereoisomers 1, 3 and 4, respectively. For further verification, the authors also carried out an X-ray diffraction study of the isolated compound, which confirmed the molecular structure of cubebin as diastereoisomer 2. With confidence on the relative configuration, the absolute configuration of cubebin was determined through comparison of the experimental ECD spectrum with the corresponding one computed by quantum chemical methods (Fig. 6).

Dichrocephones A and B: Synthesis and DP4.2[208]

In 2013, Ding and Zou isolated two new cytotoxic propellane triquinanes from *dichrocephala benthamii*, called dichrocephones A and B. These compounds present a dense array of stereogenic centres, featuring three contiguous quaternary carbon centres embedded in a fully substituted cyclopentanone fused with two cyclopentanes. Furthermore, to date, it is the highest oxidised known modhephane.[209]

For dichrocephone A, the connectivity and relative configuration were determined by extensive HMBC and ROESY experiments and the absolute configuration was assigned based on the comparison of experimental and calculated ECD spectra. However, in 2017, Kutateladze and Kuznetsov, using the para-

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

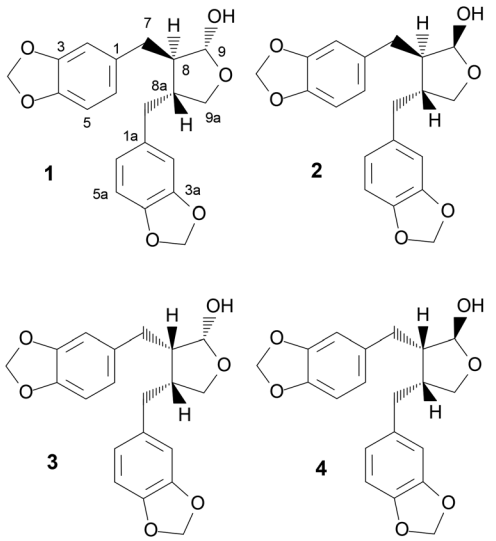


Fig. 6 Four possible diastereoisomers of cubebin.

metric/DFT hybrid computational method DU8+, demonstrated that the structure of dichrocephone A should be revised to the C-2 epimer.[210]

In 2018, motivated by the unprecedented functionalised tetracyclic core, Schmiedel and co-workers embarked on the total synthesis of dichrocephone B. To do that, the authors envisioned that the oxetane motif in dichrocephone B could be formed through dehydrative etherification of dichrocephone A, in a biomimetic fashion.[208] After an elegant synthetic approach, the author successfully reached the desired product. The key features of the synthesis included the application of a catalytic asymmetric Wittig reaction, followed by stereoselective functionalisation of the propellane core into a pentacyclic intermediate and double reductive ring cleavage. The structure of the synthetic product was confirmed by X-ray crystallography. However, the ^{1}H and ^{13}C NMR chemical shifts of the synthetic compound did not match the NMR reported values of the isolated compound. Thus, the synthetic compound could not be the correct structure of the isolated dichrocephone A. Similarities obtained between the HMBC spectra of the synthetic product with the isolated dichrocephone A suggested that they should be diastereoisomers. The authors realised that the biosynthetic proposal of dichrocephone A would be the precursor of dichrocephone B, through a dehydrative etherification, but would only be possible if the structure of dichrocephone A was the same as that of the synthetic molecule or its epimer at C-8. All the other possible diastereoisomers would prevent intramolecular ring closure. Thus, to obtain a more reliable discrimination between the possible diastereoisomers, the authors decided to use the DP4.2 method. To do that, all diastereoisomers were calculated at the mPW1PW91/6-311+G(2d,p)//B3LYP/631+G(d,p) level with PCM to simulate solvent effects (acetone). The DP4.2 probabilities (all data in scaled form) obtained through comparison of the simulated NMR chemical shifts with the experimental data of the synthetic compound confirmed the structure proposed by X-ray analysis with a high level of confidence (100%). On the other hand, upon performing the same DP4.2 analysis with the NMR data reported for the isolated compound, the epimer at C8 was predicted with a high level of confidence (99.8%) using the DP4.2 method. These results revised the recent reassignment reported by the DU8+ method. The authors further justified that the error in the previous DU8+ study probably occurred due to the way in which the contributions of certain conformers were weighed. To support this proposal, the authors followed a synthetic route to reverse the configuration at C-8. Doing so, the NMR data of the inverted synthetic C-8 product material matched those of the isolated natural product, thus confirming the biosynthetic proposal and the corrected structure of dichrocephone A (Fig. 7).[208]

Lobophytolins A and B: *j*-DP4[186]

In 2020, Li and co-worker isolated two new cembranoids, namely lobophytolins A and B, from the Chinese soft coral *Lobophytum* sp. The structure of these bioactive marine natural products was elucidated by NMR spectroscopy, quantum chemical calculations and X-ray diffraction analysis. In addition, the authors also reported the difficulties encountered in determining their full structures.

The planar structure of lobophytolins A and B was determined based on extensive analysis of HR-ESIMS, ^{1}H and ^{13}C NMR data using the COSY and HMBC 2D methods. Once the planar structures of lobophytolins A and B were elucidated, the authors focused their attention on the determination of their stereochemical features. However, this was not an easy task. The overlapping chemical shift of H-1 and H-2 prevented the measurements of experimental $^{3}J_{HH}$ as well as the unambiguous analysis of the NOESY experiment. Another problem

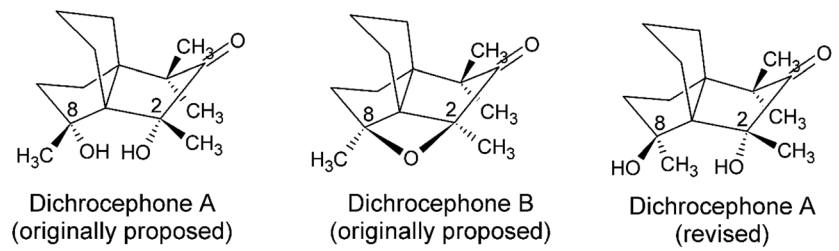
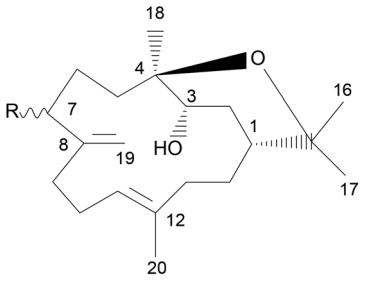


Fig. 7 Structure of dichrocephone A and B before and after reassignment.

 This journal is © the Partner Organisations 2021

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

was to connect the relative configuration of the oxane ring, since it is located on a remote position at C-7, separated by the non-protonated atom C-4. Thus, the authors decided to face this problem computationally, by calculating NMR spectroscopic properties. For that, eight possible candidate disastereoisomers were evaluated. First, conformational searches were performed. To keep the computational cost to a minimum, the authors decided to use only a 12 kJ mol^{-1} energy cut-off value. Afterward, geometrical optimisation at the DFT level was undertaken at the B3LYP/6-31G(d) level, followed by the NMR calculations at the PCM/mPW1PW91/6-31+G(d,p) level, as recommended for the DP4+ method. The NMR shielding constants were also calculated and averaged over the Boltzmann distribution obtained for each stereoisomer and correlated with the experimental data. The shielding constant simulations provided conclusive guidance about the relative configuration at C-1, C-3 and C-4. However, the configuration at C-7 could not be irrefutably determined. This result was not unexpected, given the separation of the two stereoclusters, in addition to the choice of using a low-energy cut-off value during the conformational search, which could have caused the loss of potentially relevant conformations, leading to an inappropriate representation of the real conformational distribution of the studied molecule. Therefore, the authors decided to use the i*J*/d*J*-DP4 protocol, which includes the JHH values to constrain the conformational sampling. To do that, the NMR calculations were performed as recommended, at the B3LYP/6-31G(d,p)//MMFF level, and $^{3}J_{HH}$ values were computed considering only the Fermi contact term. Once the stereocentre of interest was located around C-7, only H-7 coupling constants were used to constrain the conformational sampling, creating a subset of candidates compatible with experimental measurements (iJDP4 step). A cut-off value of 2.5 Hz from the experimental value was used as the restraining factor. Thus, using only the selected compatible conformations, the authors proceeded with the i*J*/d*J*-DP4 analysis, making use of all other $^{3}J_{HH}$ in combination with the chemical shift (dJ-DP4 step). Finally, the experimentally observed NMR data for lobophytolin A gave the best match, with a high level of confidence of >99%, with the candidate of configuration 1*R*,3*S*,4*R*,7*R*. On the other hand, the comparison using the NMR experimental data for lobophytolin B gave its best match, with a high level of confidence of >99%, for the epimer 1*R*,3*S*,4*R*,7*S*. Even though most $^{3}J_{HH}$ experimental values were similar, the different coupling constant values measured for H-7 in lobophytolin A (dd, J = 6.0, 4.0 Hz) and B (dd, J = 10.3, 4.0 Hz) were decisive for discriminating them. To confirm the previous proposal, X-ray diffraction analysis was used. The authors also discuss that the two obtained crystallographic structures were different from the conformations obtained using the restrained i*J*/d*J*-DP4 method, to the point that the DP4+ analysis performed using only the X-ray structures pointed to the 7*R* isomer for both lobophytolins A and B. This fact clearly suggests that the solution and solid-state conformations of lobophytolins A and B are different, demonstrating the great difficulty in determining the structure of these molecules (Fig. 8).



Lobophytolins A - R = α-H
Lobophytolins B - R = β-H

Fig. 8 Structure of lobophytolins A and B.

III.4 Marine natural products

Donghaecyclinones A–C: CP3 and DP4[211]

In 2020, Bae and co-workers, working on the chemical profiling of the *Streptomyces* sp. strain SUD119 from a marine sediment sample collected from a volcanic island in Korea, isolated three new metabolites called donghaecyclinones A–C (1–3, Fig. 9).

The planar structure of donghaecyclinones A–C (1–3) was proposed based on extensive analysis of high resolution-fast atom bombardment mass spectrometry (HR-FABMS), and the combination of ^{1}H and ^{13}C with HSQC, HMBC and COSY NMR data. For donghaecyclinones A (1), this analysis led to a 6/6/6/5/6 pentacyclic skeleton. On the other hand, after a comprehensive analysis of the 1D and 2D NMR data, it was revealed that donghaecyclinone B (2) and C (3) present the same planar structure. However, the optical rotations of donghaecyclinone B (2) and C (3) were different, yielding opposite signs with different absolute values (+125.6 and −156.7, respectively). Thus, the authors concluded that these two compounds were diastereomers, which requires a rigorous stereochemical analysis. With the planar structure of donghaecyclinones A–C (1–3) determined, the authors focused their attention on the relative configuration determination. For donghaecyclinone A (1), the relative configuration of ring A was estab-

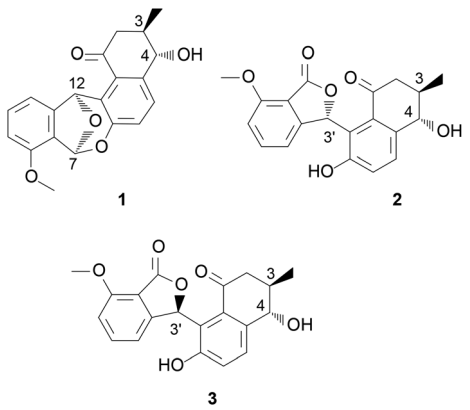


Fig. 9 Structures of donghaecyclinones A–C.

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

lished based on the $^{3}J_{HH}$ values, ROESY correlations and Mosher's method. However, the lack of ROESY correlations at C-7 and C-12 prevented the assignment of the relative configuration at the junction of rings Ca and Cb. To overcome this issue, the author decided to use the DP4 method. The two possible diastereomers with configuration (3*R*,4*S*,7*S*, and 12*S*) and (3*R*,4*S*,7*R*, and 12*R*) were proposed. After performing a conformational search using the MMFF force field at the MacroModel program, 12 low energy conformers within 10 kJ mol^{-1} were selected for both possible diastereoisomers. Then, the ^{1}H and ^{13}C NMR chemical shifts were calculated for both diastereoisomers at the B3LYP/6-31G(d,p) level in the gas phase. The DP4 method afforded 98.3% probability for the (3*R*,4*S*,7*S*, and 12*S*) diastereoisomer. For donghaecyclinone B (2) and C (3), it was possible to establish the relative configuration at C-3 and C-4 as 3*R** and 4*S** based on the $^{3}J_{HH}$ values and ROESY correlations. However, as they were expected to be diastereomers, these compounds must have opposite configurations at C-3′, which is the only remaining chiral centre with an undetermined relative configuration. To solve this, as the situation presented two possible experimental NMR values (isolated donghaecyclinone B and C) and two possible diastereoisomers (3*R**,4*S** and 3′*S**) and (3*R**,4*S** and 3′*R**), the authors cleverly decided to apply the CP3 method instead of DP4. The CP3 method showed, with a high level of confidence (100%), that the relative configurations of donghaecyclinone B (2) and C (3) were 3*R**,4*S**, and 3′*S** and 3*R**,4*S** and 3′*R**, respectively. To determine the absolute configurations of donghaecyclinone B (2) and C (3), the authors also applied Mosher's method, as applied for donghaecyclinone A (1). However, during the MTPA derivatisation, donghaecyclinones B (2) and C (3) underwent isomerisation at C-3′ at the furanone moiety, which prevented its assignment. This problem was circumvented through the comparison of experimental ECD spectra of donghaecyclinone B (2) and C (3) with the TD-DFT simulated spectra at the B3LYP/def-SVP//B3LYP/def-SVP level for both enantiomers. With these analyses, it was possible to assign donghaecyclinone B (2) as having an absolute configuration of 3*R*,4*S* and 3'*S* and donghaecyclinone C (3) as having an absolute configuration of 3*R*,4*S* and 3′*R*.

Waikikiamides A–C: DP4+[212]

In 2020, Wang and co-workers worked with the marine fungal strain of *Aspergillus* sp. FM242, isolated from a sample collected at Waikiki beach in Oahu, Honolulu, Hawaii. They isolated the structurally complex diketopiperazine derivatives called waikikiamides A–C, along with its putative biogenic precursors, called (+)-semivioxanthin, notoamide F and (−)-notoamide A. Waikikiamides A and B bear an unprecedented hendecacyclic ring system, whereas waikikiamide C was the first prenylated indole diketopiperazine heterodimer presenting an N–O–C bridge.

Waikikiamide A was isolated as an optically active powder with a molecular formula of $C_{41}H_{39}N_3O_9$, as established by high-resolution electrospray ionisation mass spectrometry (HRESIMS). Its structural analysis based on NMR methods was challenging due to the low H/C ratio. After extensive COSY and HMBC analysis with different solvents, dividing the structure in two fragments, it was possible to determine that the structure of fragment I of waikikiamide A is similar to that of (+)-avrainvillamide, except for positions 20 and 21, while fragment II presented a highly aromatic character similar to (+)-semivioxanthin, except for positions 8′ and 9′, which was also isolated in the same study. However, according to the molecular formula of waikikiamide A, one additional ring and one oxygen atom had to be assigned. Because only the valences at C-20 and C-9′ remained unsatisfied, the authors realised that the additional ring and oxygen could only result from an ether linkage between C-20 and C-9′, determining the planar structure of waikikiamide A. The NOESY spectrum indicates that H-6 and 26-NH had the same orientation, whereas H-21 and the C20–C19 bond had opposite orientations. However, to further determine the relative configuration of fragment I of waikikiamide A, the authors decided to use the quantum chemistry calculations of NMR shifts for correlation with experimental data, followed by the analysis of the DP4+ probability. For that, eight possible diastereoisomers were generated, resulting from varying the configurations at C-6, C-20 and C-21. The eight possible diastereoisomers (called 1a to 1h) were submitted to conformational sampling at the MMFF level. After that, the NMR chemical shifts of all conformers found with a 10 kcal mol^{-1} cut-off value for each configuration were computed at the PCM/mPW1PW91/6-31+G(d,p)//B3LYP/6-31G(d) level. Among the possible candidates, one diastereoisomer (1h) stood out with MAE values of 1.2 and 0.09 ppm for the carbon and proton data, respectively. After the application of DP4+ analysis, the same candidate yielded an overall probability value over 99.9% (all data, ^{13}C + ^{1}H), which represented a good agreement with the NOESY correlations observed for waikikiamide A. Despite that, determine the relative configuration of the connected fragments (I and II) was a challenging task, mainly for the C-3′ position, given the large separation of both stereoclusters by a planar aromatic core. As it was not possible to obtain suitable crystals for X-ray analysis, the authors decided to compute the NMR shifts for the 3′ epimer (3′*epi*-1h), in an attempt to suggest a full stereochemical assignment. As expected, the calculated NMR chemical shifts of both epimers were similar, with slightly larger MAE values for 3′*epi*-1h (1.26 and 0.093 ppm) *vs.* 1.21 and 0.087 ppm for 1h. As the MAE values were not enough to discriminate between the two epimers, DP4+ was also applied, favouring candidate 1 h with a moderate probability value of 80.7%. Although this is not an ideal DP4+ value (preferably close to 100%), it was enough for the authors to make a structural proposal for waikikiamide A. With confidence on the relative configuration, the absolute configuration of waikikiamide A was determined through the comparison of experimental ECD spectra with the corresponding computed spectra of both epimers (1h and 3′*epi*-1h) using quantum chemical methods. The Boltzmann-weighted ECD spectra of both candidates matched well with the experimental ECD spectrum, suggesting that the ECD spectrum seems to be governed mainly by the handedness of fragment I, allowing the authors to assign the absolute configuration of waikikiamide A.

 This journal is © the Partner Organisations 2021

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

For waikikiamide B, the molecular formula of $C_{41}H_{39}N_3O_8$ suggests one less oxygen than that of waikikiamide A. The extensive NMR analysis revealed that both compounds present almost the same structure, except for the oxygen at position 9, which is missing in waikikiamide B. The NOESY correlations and CD spectrum were almost the same as those of waikikiamide A. Hence, based on that data, the authors proposed the structure of waikikiamide B.

For waikikiamide C, the HRESIMS data revealed a molecular formula of $C_{54}H_{62}N_6O_{11}$. After an extensive NMR analyses, the authors concluded that waikikiamide C was also composed of two fragments (I and III), where fragment I is the same as found in waikikiamide A and B, whereas fragment III is 12,13-dihydroxy-13-hydroxy-12-methoxy-notoamide G. However, the lack of HMBC correlations between these two fragments prevented the determination of the connectivity between them. To solve this structural problem, after repeated recrystallisation with acetonitrile at room temperature, the authors succeeded in obtaining a suitable crystal for an X-ray crystallographic study. Thus, the final structure of waikikiamide C, including the connectivity of fragments I and III and the absolute configuration, were unambiguously determined (Fig. 10).

III.5 Reassignments

Curcusones I and J: DP4+, DP4, statistical parameters[213]

In 2013, Liu and co-workers isolated several rhamnofolane diterpenoids from the roots of *Jatropha curcas*. Among them, curcusone I and J stood out for featuring an unusual oxa-bridge, as suggested by detailed 1D and 2D NMR experiments.[214] Their challenging molecular structures, coupled with the scarcity of synthetic efforts towards rhamnofolane-derived natural products, led Li and Dai to pursue the total synthesis of the two isomers through an elegant tandem gold catalysis approach.[215] However, the authors demonstrated that the NMR data of the synthetic samples did not match those obtained for the isolated natural products. Thus, there was a need to review the original proposed structures (Fig. 11).[215]

Later in 2018, after a detailed analysis of the NMR data of synthetic and natural curcusones I and J, Sarotti concluded

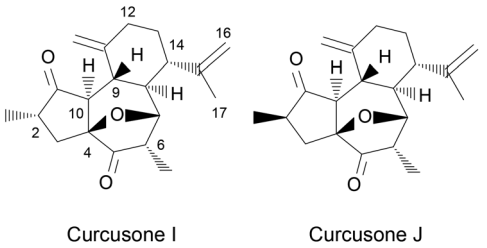


Fig. 11 Original proposed structures for curcusone I (left) and J (right).

that the source of the structural error should be stereochemical in nature due to the modest correlation between them. With this assumption in mind, as it was not possible to perform a more detailed analysis by accessing the authentic samples of the natural products or the original FID files, Sarotti decided to approach this structural problem computationally. Thus, the NMR chemical shifts of all possible 64 diastereoisomers were computed at one of the recommended theoretical levels for DP4+ analysis, the mPW1PW91/6-31+G(d,p)//B3LYP/6-31G(d) (using chloroform as the solvent in the PCM method). The author decided to build the 64 possible diastereoisomers by keeping the configurations arbitrarily fixed at the oxa-bridge and making all possible permutations in the remaining stereocentres. After a conformational analysis using the MMFF force field, all geometries within 10 kcal mol^{-1} range for each candidate were then fully optimised at the B3LYP/6-31G(d) level and then taken for further GIAO NMR chemical shift calculations. After all the calculations were done, the performance of the ^{13}C and ^{1}H DP4+ method was evaluated. First, the author validated the DP4+ method by correlating the experimental NMR data obtained for the synthetic compounds with all possible diastereoisomers, where the suitable diastereoisomer was correctly pointed out in the DP4+ method with high confidence (>99.9%). Next, the DP4+ method was evaluated for the real structural problem, correlating the experimental NMR data of the isolated natural products (curcusone I and J) with all possible diastereoisomers. The diastereoisomer corresponding to the original proposed

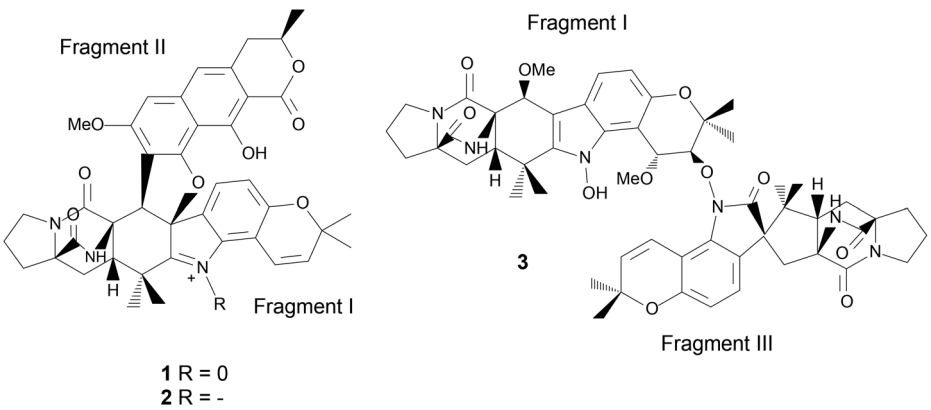


Fig. 10 Chemical structure of waikikiamides A–C.

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

structure for curcusone I presented just a modest match in correlation with the experimental data, taking the 32nd position after arranging the 64 candidates in descending order of DP4+ probability, clearly indicating that the putative original structure was highly unlikely. On the other hand, the H-DP4+, C-DP4+ and all data DP4+ values identified with high confidence (above 99.9%) a candidate displaying an inversion of the configuration at C-2, C-4 and C-7 for curcusone I. It is interesting to highlight that this selected candidate also demonstrated the lowest values for the basic statistical parameters, like CMAE and CMaxErr, corroborating the revised proposal. For further verification of the reassignment, the *J* coupling constants of all candidates were computed using only the Fermi contact term at the affordable B3LYP/6-31G(d,p)//B3LYP/6-31G (d) level of theory. For curcusone I, an excellent correlation between the ROESY spectrum of the isolated compound with the distances obtained from the calculated *J* couplings for the DP4+ selected candidate served as an additional argument favouring the reassignment.

With the structure of curcusone I solved, the author focused his attention on curcusone J. Likewise, after arranging all 64 candidates in descending order of DP4+ probability, the diastereoisomer corresponding to the original proposed structure was placed only in the 29th position, once again indicating that the original structure was highly unlikely. On the other hand, in this case, DP4+ obtained for proton and carbon data pointed in different directions: whereas C-DP4+ identified a most likely candidate with 92.5% probability, H-DP4+ suggested another candidate as the most likely one with more than 99.9% probability. However, the author realised that the overall assignment made by H-DP4+ was more meaningful, generating an all data DP4+ probability value of 95.9% toward the selected candidate suggested for H-DP4+. Furthermore, the structure of the selected candidate for curcusone J was fully consistent with the experimental chemical shift and multiplicity of the H-7 signal observed experimentally, suggesting that the C-6/C-7 relative configuration should be syn, as in the selected candidate, and not anti as in the putative original structure. Once again, the calculated *J* coupling for the selected candidate also displayed interatomic distances consistent with the ROESY correlations reported for the isolated curcusone J, favouring reassignment.

It is interesting to point out that the author also used the original DP4 method at the B3LYP/6-31G(d,p)//MMFF level of theory for the stereoassignment of curcusones I and J, reaching the same conclusions and demonstrating that, in this case, both methods (DP4 and DP4+) worked properly (Fig. 12).

Rubriflordilactone B: DP4+[216]

In 2006, Xiao and co-workers isolated two highly unsaturated bisnortriterpenoids from *Schisandra rubriflora*, called rubriflordilactones A and B. Both isolated compounds presented polysubstituted central aromatic rings in the proposed structure elucidated based on NMR data and X-ray crystallography.[217] Recently, the total synthesis of rubriflordilactone B was accomplished by Li and co-workers.[218] However, the authors realised that although the X-ray analysis of the total synthesis

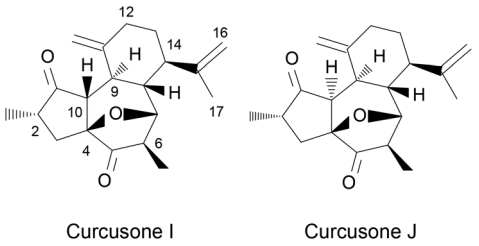


Fig. 12 Revised structure of curcusone I (left) and J (right).

product clearly confirmed its identity as the proposed structure, its ^{1}H and ^{13}C NMR spectra were surprisingly unlike those of the isolated natural product. As the observed differences could not be attributed to factors like solvent, temperature or concentration, the authors considered the possibility that the originally isolated compound may have contained two compounds with different amounts and solubility. The crystalised compound, which is the minor amount, afforded the reported X-ray structure which agrees with the synthetised compound, whereas the major component probably remained in solution but was responsible for the observed bioactivities and published NMR spectroscopic data. The authors also suggested that this major compound is probably an isomer of rubriflordilactone B, called pseudorubriflordilactone. Considering the remote possibility of unravelling the structure of pseudorubriflordilactone B by reisolation of the natural product, Grimblat and co-workers decided to examine this structural problem theoretically. As the most of the ^{13}C chemical shifts between the isolated and synthetic compound were in close agreement, it seemed plausible for the authors to postulate that the misassignment problem was stereochemical in nature, as suggested by Li and co-workers. However, as the proposed structure presents eight stereogenic centres, 128 different diastereoisomers are plausible for that hypothesis, called PRB1 to PRB128 by Grimblat and coworkers. To narrow down the possible candidates, as minor differences in the ^{13}C NMR spectra of both compounds were noted for carbons C1 to C10, the authors assumed that the relative configuration of the ABC ring system of the isolated and synthetic compound might be the same. With this assumption, 32 candidates bearing the 1*R*,5*S*,10*R* configuration remained plausible (PRB33–PRB64).

After an exhaustive conformational search using the MMFF force field, the authors realised that these molecules are fairly rigid, so unique conformations were located for each candidate. These were optimised at the B3LYP/6-31G(d) level of theory for further calculations of the NMR chemical shifts at the mPW1PW91/6-31+G(d,p) level using the PCM (pyridine) model, for subsequent application of the DP4+ method.

To validate the method, the authors first correlated the NMR data computed for candidates PRB33–PRB64 with the experimental NMR values reported for the synthetic compound, whose structure was unambiguously determined by X-ray analysis. The DP4+ method yielded a probability value of

 This journal is © the Partner Organisations 2021

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

92.6% for the candidate called PRB48, which was indeed the correct structure of rubriflordilactones B. However, during the validation of the method, the authors noted an interesting fact. While the C-DP4+ probability value was high (94.0%) for candidate PRB48, the obtained H-DP4+ value for candidate PRB49 was higher than the one obtained for candidate PRB48 (58.8% *vs.* 6.1%). The fact that the overall obtained DP4+ value indicates candidate PRB48 with a high level of confidence even with unfavourable H-DP4+ values strengthens the need for using both types of data whenever possible.

With the DP4+ methodology validated for this case, the authors focused their analysis on the correlation with the experimental data reported for pseudorubriflordilactone B. In this case, the DP4+ method yielded an extremely low confidence (<0.01%) for the candidate PRB48, representing the originally proposed structure. On the other hand, PRB60, with configuration 1*R*,5*S*,10*R*,16*S*,17*R*,20*S*,22*S*,23*S*, was identified as the most likely candidate with an overall probability value of 99.5%. These candidates (PRB48 and PRB60) differ only in the C-16 and C-17 configurations, describing the E/F ring fusion. However, the authors noted that, in this case, the ^{1}H NMR data conclusively favoured the most likely candidate with a probability value of 99.9%, while the carbon data, on the other hand, provided a weak argument toward PRB60, yielding a C-DP4+ value of only 0.5%. The authors argued that although it is not uncommon that carbon and proton data point toward different directions in the DP4+ method, the large difference obtained between H-DP4+ and C-DP4+ caught the authors' attention. With a close inspection of the experimental and calculated ^{13}C NMR data, a large discrepancy was noted in the chemical shift assigned for C-15 methylene. After an extensive detailed review of the 1D and 2D NMR data reported in the ESI of the isolation paper, the authors considered that a misinterpretation of this signal probably occurred due to an impurity present in the sample. In fact, after the calculated and experimental chemical shifts for C-15 were removed from all candidates, the calculated C-DP4+ probability changed dramatically, clearly indicating that the low C-DP4+ value obtained initially for the candidate PRB60 was strongly related to the C-15 signal. An additional argument favouring for candidate PRB60 was obtained from the detailed examination of the NOESY spectrum of the isolation sample, where a correlation between H-1 and H-27 was noticed. Among all the 32 candidates, the only one in which this NOESY interaction is clearly compatible with a low energy conformation is the candidate PRB60. Nevertheless, to rule out any possibility that another stereoisomer candidate besides the 32 considered thus far by the authors could be the true structure of pseudorubriflordilactone B, the authors decided to calculate the chemical shifts of the remaining 96 possible candidates. Then, the DP4+ method was recomputed considering the entire set of all 128 possible candidates. Fortunately, even considering all the possibilities, the obtained DP4+ values were practically unchanged, and once again the candidate PRB60 was identified as the most likely with a high overall level of confidence 99.5%. This study clearly highlights the potential of the DP4+ method as a key tool for evaluating the structure of complex natural products, in addition to providing a powerful guide to the additional synthetic work needed to finally elucidate the structure of rubriflordylactone B (Fig. 13).

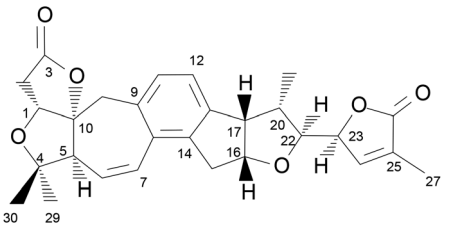


Fig. 13 Structure of rubriflordylactone B.

III.6 Parametric corrections

Xylogranatin F: DU8+[219]

In 2008, Wu and co-workers isolated 13 limonoid compounds from the seeds of a Chinese mangrove, *Xylocarpus granatum.* These compounds, called xylogranatins F–R, presented a new carbon skeleton, for which structures were elucidated on the basis of spectroscopic data and chemical methods.[220] However, in 2019, Schuppe and co-workers working on the total synthesis of some bislactone limonoid alkaloids, determined that xylogranatin F was originally structural mischaracterised. In the original isolation paper, the authors proposed the relative configuration for C-3, C-5, and C-10 of xylogranatin F based on NOESY data. Schuppe and co-workers realised that although the NOE signal between the protons at C-5 and C-19 was unequivocal and suggested the *cis* fusion between rings A and F, the proposed NOE signals between the protons at C-3 and C-5, as well as between C-3 and C-19, were very weak and ambiguous. Thus, the authors suggested that a misassignment might have occurred. The signal that was interpreted as an authentic NOE signal could be attributed to background noise. In this way, suspicious of this structural misassignment, Liu and co-workers decided to investigate this issue based on NMR calculations. The authors considered four different possible structures by varying the stereochemistry at positions C-3, C-5 and C-10 while maintaining the *cis* fusion between rings, which was reliably supported by NOESY data.

First, a conformational search was performed for the four possible diastereoisomer candidates using the MMFF force field with the Macromodel software. Then, each conformer was subjected to geometry optimisation at the B3LYP/6-31+G (d,p) level of theory in the gas phase. NMR single point calculations using the GIAO method were performed for all geometrically optimised structures using mPW1PW91/6-311+G(2d,p) in chloroform with the SMD solvation method. After the calculations, the simulated results were also subjected to statistical analysis using the DP4 method. The ^{13}C NMR calculation yielded chemical shifts for the originally proposed structure (3*R*,5*S*,10*S*) with a MAE deviation of 1.5 and RMSD of 2.0 ppm. Furthermore, the chemical shifts of several carbon atoms on

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

rings A and F, around the C3 stereogenic centre, yielded errors over 3.0 ppm. On the other hand, the candidate consistent with the reassignment of xylogranatin F (3*S*,5*S*,10*S*) presented the lowest average chemical shift deviation of 1.2 and RMSD of 1.5 ppm. Also, the deviations around the C-3 stereogenic centre were all within 3.0 ppm. In addition, this candidate yielded a C-DP4 probability value of 81.8%. The ^{1}H NMR calculation results also suggested a structural misassignment, where the chemical shifts of the originally proposed structure (3*R*,5*S*,10*S*) yielded an MAE deviation of 0.18 and RMSD of 0.22 ppm. In contrast, the candidate (3*S*,5*S*,10*S*) presented the lowest MAE deviation of 0.14 with an RMSD of 0.17 ppm. Additionally, the H-DP4 probability value obtained was 100%, which led to a total DP4 (^{1}H + ^{13}C) probability value of 100%.

The authors also decided to apply the DU8+ parametric calculations, at the ωB97xD/6-31G(d) level in chloroform with the default PCM solvation method, where all the optimised structures were subjected to the DU8+ empirical corrections for ^{13}C. The results of the DU8+ calculations confirmed the structural misassignment of xylogranatin F. For the originally proposed structure, the discrepancy between the experimental ^{13}C NMR chemical shift and the calculated data obtained was over 7 ppm for carbons around the C-3 stereogenic centre. On the other hand, interestingly, the calculated chemical shifts of the correct candidate 3*S*,5*S*,10*S* were in an excellent agreement with the experimental data, yielding an MAE of 0.6 and an RMSD of 0.8 ppm. However, although the DU8+ method presented this excellent result for the correct candidate, the difference in ^{13}C NMR chemical shifts obtained for the other two possible diastereomers candidates becomes less pronounced, in a way that it was not possible to discriminate them once the obtained differences dropped below the accuracy of the computational method. The authors concluded that the hybrid DFT-parametric DU8+ method demonstrated superior accuracy using a fraction of the computation cost, yielding an excellent MAE value of only 0.6 ppm for the revised structure, significantly better than the MAE of 1.2 ppm achieved with the more standard GIAO method at the mPW1PW91/6-311+G(2d,p)//B3LYP/6-31+G(d,p) level of theory. However, although the accuracy was lower for the standard GIAO method, it correctly predicted the right structure with larger differences predicted among the multiple candidates, which did not occur with the DU8+ method. In addition, the results obtained with the standard GIAO method can be further analysed by powerful statistical methods, like DP4 for example.

Briarellins: DU8+[140]

In 1995, a subset of C2–C11 cyclised cembranoids, called briarellins, were isolated from gorgonian octocorals. To date, the briarellin family comprises 20 compounds, called briarellins A–T. Several compounds from this family exhibit interesting pharmacologic properties, creating a broad interest in the synthesis of these complex structural compounds. In 2009, Corminboeuf and co-workers succeed in synthesising briarellins E and F, confirming the ethereal C3–C14 link in the structure of briarellins.[221] However, when Crimmins and co-workers succeed in synthesising briarellin J, they argued that the ^{1}H and ^{13}C NMR spectra of their synthetic product, which had its structure determined by X-ray crystallography, are perceptively different from the NMR data obtained for the isolated natural compound. The authors suggested, based on their synthetic efforts and after an extensive analysis of Faulkner's biosynthetic proposal and re-examination of the original isolation data, that the true structure of briarellin J contains the stereochemistry at C-11 in opposition to that originally reported.[222]

As the structure of briarellins was elucidated mainly by NMR, and after the reassignment made by Crimmins group, Holt and co-workers decided to investigate the whole family of briarellin compounds with the help of the hybrid DFT/parametric computational DU8+ method. After this study, the authors reported that most of them required structural revision. The authors showed that, with the DU8+ analysis of all reported briarellins, it was possible to suggest that structures containing the C3–C14 ethyleneoxy bridge, such as briarellins E, F, I, S and T, were originally assigned correctly. However, for the C3–C14 bridged lactones, such as briarellins A–D, H, and J–R, the original structure was misassigned. The authors' analysis of briarellin J revealed two most severe deviations between the calculated and experimental ^{13}C NMR data. The first was the chemical shift for Me17, calculated at 11.1 ppm, which presented an error greater than 6 ppm from the experimental value of 17.7 ppm. Similarly, the calculated value of 23.7 ppm for C13 presented an error of almost 7 ppm from the experimental value of 16.8 ppm. The authors also realised that all CH_2 carbons in briarellin J were in fact incompatible with the originally proposed structure, since the simulated chemical shift values were superior to 23 ppm, while the corresponding experimental CH_2 signal was around 16.8 ppm.

As the proton spectrum of the experimental data yielded multiple overlapping signals, scalar coupling data were not helpful in this case. So, after trial and error, the authors succeeded in a revision that matched the ^{13}C NMR experimental data. This revision required a lactone bridge connecting carbons C-14 and C-11, while the C-11-acetate was moved to position C-3. The calculated NMR data for the proposed revised structure for briarellin J were closely comparable to the experimental NMR chemical shift data, yielding an RMSD value of only 1.06 ppm. After the DU8+ analysis for briarellins, the authors also proposed that the ^{13}C chemical shift of Me17 could be used as a differentiation criterion to discriminate between structures containing the original triatomic C3–C14 bridge and the structures possessing the new structural motif, the C11–C14 bridge. All compounds presenting the high field ^{13}C chemical shift values for Me-17 of 10–12 ppm range could be reliably determined as presenting the original proposed C3–C14 bridge, while compounds presenting the Me-17 peak in the range of 17.4–17.9 ppm should be revised into the structures containing the C11–C14 bridge. However, briarellin H was the only member of this lactone family that yielded an experimental Me-17 value slightly outside of the 17.4–17.9 ppm range and the C11–C14 candidate structure showed a poor fit. After a careful analysis of the DU8+ NMR computed and experimental data, the authors revised the structure to yet another structural type, a C12–C14 bridged

 This journal is © the Partner Organisations 2021

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

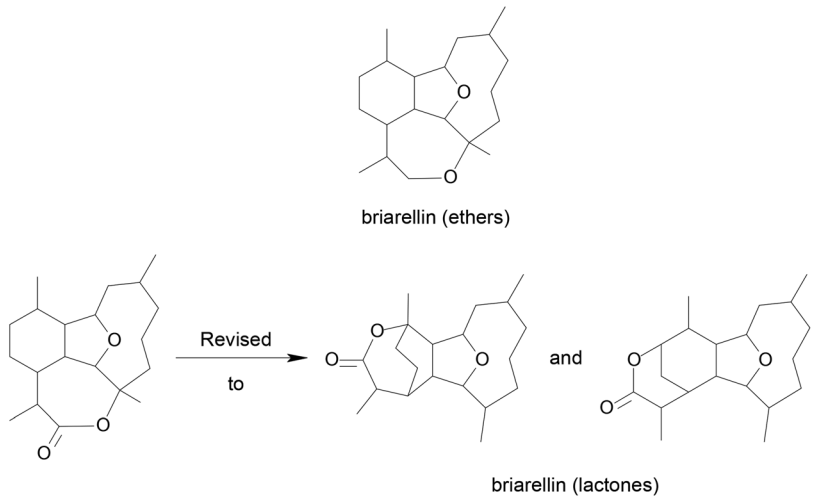


Fig. 14 Correct originally proposed structure of briarellin ethers (top), and the revised structural types of briarellin lactones (bottom).

lactone. In conclusion, after DU8+ calculations, the authors suggested that briarellins A–D, H, and J–R should be revised to two new structural types, possessing a C11–C14 or C12–C14 lactone bridge (Fig. 14).[140]

IV Conclusions and future perspectives

From our viewpoint, the new statistical procedures that are being employed to help rationalise assignments of experimental and theoretical chemical shifts represent valuable tools to reduce the difficulties associated with the determination of chemical structures by NMR spectroscopy. To date, the accuracy of simulation methods complemented with the new statistical tools enables the discrimination of closely related structures, like diastereosiomers, with a high level of confidence. The seminal contributions of several research groups, mainly those led by Tantillo, Sarotti, Goodman and Kutateladze, have induced overwhelming growth in the field, updating and broadening research activity in the NMR simulation area, to the point that nowadays such studies are routinely found in the high impact literature. Nevertheless, this research field is still far from stagnant and has much to advance. We believe that growth will proceed into two main directions. The first is related to artificial intelligence. Sarotti is the pioneer in this field, implementing the use of artificial neural network pattern recognition analysis (the method called ANN-PRA) to the simple and rapid detection of structural misassignments. However, ANN are just one of the algorithms available in what is now called machine learning. We believe that, in the near future, new methodologies will emerge using new algorithms in a combined approach, further improving this type of methodology pioneered by Sarotti.

The second direction relates to the automation process. It is well-known that managing and processing multiple files to perform an *in silico* structural assignment, *i.e.*, by dealing with a large number of isomers and/or molecules with expressive conformational flexibility, can be a discouraging and tedious work. In this sense, there is a search for the development of protocols for the automation of the whole process. In recent years, we have witnessed the significant development of the DP4-AI method by the Goodman group, a robust system for automatic processing and assignment of raw ^{13}C and ^{1}H NMR data, under development to be integrated into a computational organic molecule structure elucidation automated workflow.[223] Although it may be considered a game changer in the field, in the same year, Sarotti's group published a paper exploring the repercussions of energy mismatch when automated protocols are used.[224] The results obtained by these authors make clear the risks of automation, as energy calculation errors might easily lead to incorrect structural assignments without 'human' validation. We believe that an automation process able to overcome this energy issue is close to be developed and will be extremely relevant to the field.

Author contributions

Fabio L. P. Costa: Conceptualization, investigation, writing – original draft and writing – review & editing. Ana C. F. de Albuquerque: Investigation, writing – original draft and writing – review & editing. Rodolfo G. Fiorot: Investigation, writing – original draft and writing – review & editing. Luciano M. Lião: Investigation, funding acquisition, writing – original draft and writing – review & editing. Lucas H. Martorano: Investigation, writing – original draft and writing – review & editing. Gunar V. S. Mota: Investigation, writing – original draft and writing – review & editing. Alessandra L. Valverde: Investigation, writing – original draft and writing – review & editing. José W. M. Carneiro: Visualization, writing – original draft and writing – review & editing. Fernando M. dos Santos Junior: Conceptualization, investigation, project administration, supervision, visualization, funding acquisition, writing – original draft and writing – review & editing.

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We wish to thank CNPq, CAPES (Code 001) and FAPERJ (grant 211.319- 2019), for financial support. We dedicate this paper to memory of Professor Mauro Barbosa de Amorim (1954–2019), a masterful teacher and mentor, and more than that, a father, a husband and a friend.

References

1 C. R. Pye, M. J. Bertin, R. S. Lokey, W. H. Gerwick and R. G. Linington, Retrospective analysis of natural products provides insights for future discovery trends, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 5601–5606.
2 M. L. Campos, L. M. Lião, E. S. F. Alves, L. Migliolo, S. C. Dias and O. L. Franco, A structural perspective of plant antimicrobial peptides, *Biochem. J.*, 2018, **475**, 3359–3375.
3 T. F. Molinski, All natural: The renaissance of natural products chemistry, *Org. Lett.*, 2014, **16**, 3849–3855.
4 I. W. Keesey, N. Jiang, J. Weißflog, R. Winz, A. Svatoš, C. Z. Wang, B. S. Hansson and M. Knaden, Plant-Based Natural Product Chemistry for Integrated Pest Management of Drosophila suzukii, *J. Chem. Ecol.*, 2019, **45**, 626–637.
5 B. A. Lorsbach, T. C. Sparks, R. M. Cicchillo, N. V. Garizi, D. R. Hahn and K. G. Meyer, Natural products: a strategic lead generation approach in crop protection discovery, *Pest Manage. Sci.*, 2019, **75**, 2301–2309.
6 A. Leung and S. Foster, *Encyclopedia of common natural ingredients*, Wiley, New Jersey, 1982.
7 J. B. Seibert, J. P. Bautista-Silva, T. R. Amparo, A. Petit, P. Pervier, J. C. dos Santos Almeida, M. C. Azevedo, B. M. Silveira, G. C. Brandão, G. H. B. de Souza, L. F. de Medeiros Teixeira and O. D. H. dos Santos, Development of propolis nanoemulsion with antioxidant and antimicrobial activity for use as a potential natural preservative, *Food Chem.*, 2019, **287**, 61–67.
8 D. J. Newman and G. M. Cragg, Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019, *J. Nat. Prod.*, 2020, **83**, 770–803.
9 A. R. Carroll, B. R. Copp, R. A. Davis, R. A. Keyzers and M. R. Prinsep, Marine natural products, *Nat. Prod. Rep.*, 2020, **37**, 175–223.
10 M. Lahlou, The Success of Natural Products in Drug Discovery, *Pharmacol. Pharm.*, 2013, **4**, 17–31.
11 K. Mori, Bioactive natural products and chirality, *Chirality*, 2011, **23**, 449–462.
12 T. Rodrigues, D. Reker, P. Schneider and G. Schneider, Counting on natural products for drug design, *Nat. Chem.*, 2016, **8**, 531–541.
13 B. K. Chhetri, S. Lavoie, A. M. Sweeney-Jones and J. Kubanek, Recent trends in the structural revision of natural products, *Nat. Prod. Rep.*, 2018, **35**, 514.
14 K. C. Nicolaou and M. O. Frederick, On the structure of maitotoxin, *Angew. Chem., Int. Ed.*, 2007, **46**, 5278–5282.
15 K. C. Nicolaou and S. A. Snyder, Chasing molecules that were never there: Misassigned natural products and the role of chemical synthesis in modern structure elucidation, *Angew. Chem., Int. Ed.*, 2005, **44**, 1012–1044.
16 N. Bross-Walch, T. Kühn, D. Moskau and O. Zerbe, Strategies and tools for structure determination of natural products using modern methods of NMR spectroscopy, *Chem. Biodivers.*, 2005, **44**, 1012–1044.
17 R. T. Williamson, A. V. Buevich, G. E. Martin and T. Parella, LR-HSQMBC: A sensitive NMR technique to probe very long-range heteronuclear coupling pathways, *J. Org. Chem.*, 2014, **79**, 3887–3894.
18 M. O. Marcarino, M. M. Zanardi, S. Cicetti and A. M. Sarotti, NMR Calculations with Quantum Methods: Development of New Tools for Structural Elucidation and beyond, *Acc. Chem. Res.*, 2020, **53**, 1922–1932.
19 G. Barone, L. Gomez-Paloma, D. Duca, A. Silvestri, R. Riccio and G. Bifulco, Structure validation of natural products by quantum-mechanical GIAO calculations of ^{13}C NMR chemical shifts, *Chemistry*, 2002, **8**, 3233–3239.
20 J. Meiler, E. Sanli, J. Junker, R. Meusinger, T. Lindel, M. Will, W. Maier and M. Köck, Validation of Structural Proposals by Substructure Analysis and ^{13}C NMR Chemical Shift Prediction, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 241–248.
21 A. Bagno, Complete Prediction of the ^{1}H NMR Spectrum of Organic Molecules by DFT Calculations of Chemical Shifts and Spin – Spin Coupling Constants, *Chem. – Eur. J.*, 2001, **7**, 1652–1661.
22 D. A. Forsyth and A. B. Sebag, Computed ^{13}C NMR Chemical Shifts via Empirically Scaled GIAO Shieldings and Molecular Mechanics Geometries. Conformation and Configuration from 13C Shifts, *J. Am. Chem. Soc.*, 1997, **119**, 9483–9494.
23 M. W. Lodewyk, M. R. Siebert and D. J. Tantillo, Computational prediction of 1H and 13C chemical shifts: A useful tool for natural product, mechanistic, and synthetic organic chemistry, *Chem. Rev.*, 2012, **112**, 1839–1862.
24 M. Bühl and T. van Mourik, NMR spectroscopy: Quantum-chemical calculations, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2011, **1**, 634–647.
25 V. A. Semenov and L. B. Krivdin, DFT computational schemes for 1H and ^{13}C NMR chemical shifts of natural products, exemplified by strychnine, *Magn. Reson. Chem.*, 2020, **58**, 1–9.
26 A. Bagno and G. Saielli, Addressing the stereochemistry of complex organic molecules by density functional theory-NMR, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2015, **5**, 228–240.
27 N. Grimblat and A. M. Sarotti, Computational Chemistry to the Rescue: Modern Toolboxes for the Assignment of

 This journal is © the Partner Organisations 2021

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

Complex Molecules by GIAO NMR Calculations, *Chem. – Eur. J.*, 2016, **22**, 1–17.
28 D. A. McQuarrie and J. D. Simon, in *Physical Chemistry : A molecular approach*, University Sciece Books, 1997.
29 P. W. Atkins and J. De Paula, in *Atkins' Physical chemistry*, Chemistry (Easton), 2009.
30 A. D. Buckingham, Permanent and Induced Molecular Moments and Long-Range Intermolecular Forces, *Adv. Chem. Phys.*, 1967, **12**, 107–142.
31 A. D. Buckingham and B. J. Orr, Molecular hyperpolarisabilities, *Q. Rev., Chem. Soc.*, 1967, **21**, 195–212.
32 W. T. Raynes, in *The Encyclopedia of Nuclear Magnetic Resonance*, Wiley, Chichester., 1996.
33 F. London, Théorie quantique des courants interatomiques dans les combinaisons aromatiques, *J. Phys. Radium.*, 1937, **8**, 397–409.
34 P. Pulay and J. F. Hinton, in *Shielding Theory: GIAO Method, Encycl. Magn. Reson. Wasylishen*, 2007.
35 J. S. Hu, S. Grzesiek and A. Bax, Two-dimensional NMR methods for determining χ1 angles of aromatic residues in proteins from three-bond J(C'Cγ) and J(NCγ) couplings, *J. Am. Chem. Soc.*, 1997, **119**, 1803–1804.
36 F. Lohr and H. Ruterjans, A sensitive method for the measurement of three-bond C′,H(α) J couplings in uniformly 13C- and 15N-enriched proteins, *J. Am. Chem. Soc.*, 1997, **119**, 1468–1469.
37 A. Pardi, Multidimensional heteronuclear NMR experiments for structure determination of isotopically labeled RNA, *Methods Enzymol.*, 1995, **261**, 350–380.
38 S. S. Wijimenga, M. M. Mooren and C. W. Hilbers, in *N.M. R. Of Macromolecules: A Practical Approach (Practical Approach Series)*, Oxford University Press, Oxford., 1993.
39 A. C. Wang and A. Bax, Determination of the backbone dihedral angles Φ in human ubiquitin from reparametrized empirical Karplus equations, *J. Am. Chem. Soc.*, 1996, **118**, 2483–2494.
40 M. J. Minch, Orientational dependence of vicinal proton–proton NMR coupling constants: The Karplus relationship, *Concepts Magn. Reson.*, 1994, **6**, 41–56.
41 V. M. S. Gil and C. F. C. Geraldes, in *Ressonância Magnética Nuclear: Fundamentos, Métodos e Aplicações*, Coimbra, 1987.
42 G. Kummerlöwe and B. Luy, Residual Dipolar Couplings for the Configurational and Conformational Analysis of Organic Molecules, *Annu. Rep. NMR Spectrosc.*, 2009, **68**, 193–232.
43 J. M. Seco, E. Quiñoá and R. Riguera, The Assignment of Absolute Configuration by NMR, *Chem. Rev.*, 2004, **104**, 17–118.
44 R. H. Contreras and J. E. Peralta, Angular dependence of spin-spin coupling constants, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2000, **37**, 321–425.
45 J. R. Tolman and J. H. Prestegard, Measurement of amide15N-1H one-bond couplings in proteins using accordion heteronuclear-shift-correlation experiments, *J. Magn. Reson., Ser. B*, 1996, **112**, 245–252.
46 G. W. Vuister, T. Yamazaki, D. A. Torchia and A. Bax, Measurement of two- and three-bond 13C-1H J couplings to the Cδ carbons of leucine residues in staphylococcal nuclease, *J. Biomol. NMR*, 1993, **3**, 297–306.
47 N. F. Ramsey, Electron coupled interactions between nuclear spins in molecules, *Phys. Rev.*, 1953, **91**, 303–307.
48 G. Cuevas and E. Juaristi, Manifestation of stereoelectronic effects on the calculated carbon-hydrogen bond lengths and one bond 1JC-H NMR coupling constants in cyclohexane, six-membered heterocycles, and cyclohexanone derivatives, *J. Am. Chem. Soc.*, 2002, **124**, 13088–13096.
49 J. Vaara, J. Jokisaari, R. E. Wasylishen and D. L. Bryce, Spin-spin coupling tensors as determined by experiment and computational chemistry, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2002, **41**, 233–304.
50 A. E. Reed, L. A. Curtiss and F. Weinhold, Intermolecular Interactions from a Natural Bond Orbital, Donor—Acceptor Viewpoint, *Chem. Rev.*, 1988, **88**, 899–926.
51 H. Benedict, I. G. Shenderovich, O. L. Malkina, V. G. Malkin, G. S. Denisov, N. S. Golubev and H. H. Limbach, Nuclear scalar spin-spin couplings and geometries of hydrogen bonds, *J. Am. Chem. Soc.*, 2000, **122**, 1979–1988.
52 W. D. Arnold, J. Mao, H. Sun and E. Oldfield, Computation of through-space 19F-19F scalar couplings via density functional theory, *J. Am. Chem. Soc.*, 2000, **122**, 12164–12168.
53 S. J. Wilkens, W. M. Westler, J. L. Markley and F. Weinhold, Natural J-coupling analysis: Interpretation of scalar J-couplings in terms of natural bond orbitals, *J. Am. Chem. Soc.*, 2001, **123**, 12026–12036.
54 A. J. Dingley, J. E. Masse, J. Feigon and S. Grzesiek, Characterization of the hydrogen bond network in guanosine quartets by internucleotide (3 h)J(NC'), and (2 h)J (NN) scalar couplings, *J. Biomol. NMR*, 2000, **16**, 279–289.
55 A. Liu, W. Hu, S. Qamar and A. Majumdar, Sensitivity enhanced NMR spectroscopy by quenching scalar coupling mediated relaxation: Application to the direct observation of hydrogen bonds in 13C/15N-labeled proteins, *J. Biomol. NMR*, 2000, **17**, 55–61.
56 A. Meissner and O. W. Sørensen, New Techniques for the Measurement of C′N and C′H N J Coupling Constants across Hydrogen Bonds in Proteins, *J. Magn. Reson.*, 2000, **143**, 387–390.
57 K. Pervushin, C. Fernández, R. Riek, A. Ono, M. Kainosho and K. Wüthrich, Determination of h2J(NN) and h1j(HN) coupling constants across Watson- Crick base pairs in the Antennapedia homeodomain-DNA complex using TROSY, *J. Biomol. NMR*, 2000, **16**, 39–46.
58 D. G. Zaccari, J. P. Snyder, J. E. Peralta, O. E. Taurian, R. H. Conteras and V. Barone, Natural J coupling (NJC) analysis of the electron lone pair effect on NMR couplings: 2. The anomeric effects on 1J(C,H) couplings and its dependence on solvent, *Mol. Phys.*, 2002, **100**, 705–715.
59 V. G. Malkin, O. L. Malkina and D. R. Salahub, Calculation of spin-spin coupling constants using density functional theory, *Chem. Phys. Lett.*, 1994, **221**, 91–99.

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

60 T. D. W. Claridge, in *High-Resolution NMR Techniques in Organic Chemistry: Third Edition*, 2016.
61 R. Y. Rubinstein and D. P. Kroese, in *Simulation and the Monte Carlo Method: Second Edition*, 2007.
62 N. Metropolis and S. Ulam, The Monte Carlo Method, *J. Am. Stat. Assoc.*, 1949, **44**, 335–341.
63 B. Esquivel, E. Burgueño-Tapia, C. Bustos-Brito, N. Pérez-Hernández, L. Quijano and P. Joseph-Nathan, Absolute configuration of the diterpenoids icetexone and conacytone from Salvia ballotaeflora, *Chirality*, 2018, **30**, 177–188.
64 S. G. Smith and J. M. Goodman, Assigning the stereochemistry of pairs of diastereoisomers using GIAO NMR shift calculation, *J. Org. Chem.*, 2009, **74**, 4597–4607.
65 J. H. Ke, L. S. Zhang, S. X. Chen, S. N. Shen, T. Zhang, C. X. Zhou, J. X. Mo, L. G. Lin and L. S. Gan, Benzofurans from Eupatorium chinense enhance insulin-stimulated glucose uptake in C2C12 myotubes and suppress inflammatory response in RAW264.7 macrophages, *Fitoterapia*, 2019, **134**, 346–354.
66 G. V. Ccana-Ccapatinta, B. L. Sampaio, F. M. dos Santos, J. M. Batista and F. B. Da Costa, Absolute configuration assignment of caffeic acid ester derivatives from Tithonia diversifolia by vibrational circular dichroism: the pitfalls of deuteration, *Tetrahedron: Asymmetry*, 2017, **28**, 1823–1828.
67 Q. N. N. Nguyen, J. Schwochert, D. J. Tantillo and R. S. Lokey, Using 1H and ^{13}C NMR chemical shifts to determine cyclic peptide conformations: A combined molecular dynamics and quantum mechanics approach, *Phys. Chem. Chem. Phys.*, 2018, **20**, 14003–14012.
68 M. Dračínský and P. Hodgkinson, A molecular dynamics study of the effects of fast molecular motions on solid-state NMR parameters, *CrystEngComm*, 2013, **15**, 8705–8712.
69 Y. Yesiltepe, J. R. Nuñez, S. M. Colby, D. G. Thomas, M. I. Borkum, P. N. Reardon, N. M. Washton, T. O. Metz, J. G. Teeguarden, N. Govind and R. S. Renslow, An automated framework for NMR chemical shift calculations of small organic molecules, *J. Cheminform.*, 2018, **10**, 52.
70 A. Supady, V. Blum and C. Baldauf, First-Principles Molecular Structure Search with a Genetic Algorithm, *J. Chem. Inf. Model.*, 2015, **55**, 2338–2348.
71 N. Nair and J. M. Goodman, Genetic algorithms in conformational analysis, *J. Chem. Inf. Comput. Sci.*, 1998, **38**, 317–320.
72 T. Pavlov, M. Todorov, G. Stoyanova, P. Schmieder, H. Aladjov, R. Serafimova and O. Mekenyan, Conformational coverage by a genetic algorithm: Saturation of conformational space, *J. Chem. Inf. Model.*, 2007, **47**, 851–863.
73 S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, Optimization by simulated annealing, *Science*, 1983, **220**, 671–680.
74 S. R. Wilson, W. Cui, J. W. Moskowitz and K. E. Schmidt, Applications of simulated annealing to the conformational analysis of flexible molecules, *J. Comput. Chem.*, 1991, **12**, 342–349.
75 T. E. Malliavin, A. Mucherino, C. Lavor and L. Liberti, Systematic Exploration of Protein Conformational Space Using a Distance Geometry Approach, *J. Chem. Inf. Model.*, 2019, **59**, 4486–4503.
76 S. M. Stow, C. R. Goodwin, M. Kliman, B. O. Bachmann, J. A. McLean and T. P. Lybrand, Distance geometry protocol to generate conformations of natural products to structurally interpret ion mobility-mass spectrometry collision cross sections, *J. Phys. Chem. B*, 2014, **118**, 13812–13820.
77 U. Azzena, M. Carraro and L. Pisano, Addressing Stereochemistry of Heterocyclic Compounds by DFT NMR Calculations, *Chem. Heterocycl. Compd.*, 2018, **54**, 380–388.
78 A. G. Petrovic, A. Navarro-Vazquez and J. Lorenzo Alonso-Gomez, From Relative to Absolute Configuration of Complex Natural Products: Interplay Between NMR, ECD, VCD, and ORD Assisted by ab initio Calculations, *Curr. Org. Chem.*, 2010, **14**, 1612–1628.
79 N. Mardirossian and M. Head-Gordon, Thirty years of density functional theory in computational chemistry: An overview and extensive assessment of 200 density functionals, *Mol. Phys.*, 2017, **115**, 2315–2372.
80 T. Schwabe and S. Grimme, Theoretical thermodynamics for large molecules: Walking the thin line between accuracy and computational cost, *Acc. Chem. Res.*, 2008, **41**, 549.
81 Y. Zhao, N. E. Schultz and D. G. Truhlar, Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions, *J. Chem. Theory Comput.*, 2006, **2**, 364–382.
82 Y. Zhao, N. González-Garda and D. G. Truhlar, Benchmark database of barrier heights for heavy atom transfer, nucleophilic substitution, association, and unimolecular reactions and its use to test theoretical methods, *J. Phys. Chem. A*, 2005, **109**, 2012–2018.
83 Y. Zhao and D. G. Truhlar, Design of density functionals that are broadly accurate for thermochemistry, thermochemical kinetics, and nonbonded interactions, *J. Phys. Chem. A*, 2005, **109**, 5656–5667.
84 A. D. Boese and J. M. L. Martin, Development of density functional for thermochemical kinetics, *J. Chem. Phys.*, 2004, **121**, 3405–3416.
85 N. Chéron, D. Jacquemin and P. Fleurat-Lessard, A qualitative failure of B3LYP for textbook organic reactions, *Phys. Chem. Chem. Phys.*, 2012, **14**, 7170–7175.
86 S. N. Pieniazek and K. N. Houk, The origin of the halogen effect on reactivity and reversibility of Diels-Alder cycloadditions involving furan, *Angew. Chem., Int. Ed.*, 2006, **45**, 1442–1445.
87 C. E. Check and T. M. Gilbert, Progressive systematic underestimation of reaction energies by the B3LYP model as the number of C-C bonds increases: Why organic chemists should use multiple DFT models for calculations involving polycarbon hydrocarbons, *J. Org. Chem.*, 2005, **70**, 9828–9834.

 This journal is © the Partner Organisations 2021

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

88 H. Kruse, L. Goerigk and S. Grimme, Why the standard B3LYP/6-31G* model chemistry should not be used in DFT calculations of molecular thermochemistry: Understanding and correcting the problem, *J. Org. Chem.*, 2012, **77**, 10824–10834.
89 X. Q. Yao, X. J. Hou, G. S. Wu, Y. Y. Xu, H. W. Xiang, H. Jiao and Y. W. Li, Estimation of C-C bond dissociation enthalpies of large aromatic hydrocarbon compounds using DFT methods, *J. Phys. Chem. A*, 2002, **107**, 9991–9996.
90 L. A. Curtiss, K. Raghavachari, P. C. Redfern and J. A. Pople, Assessment of Gaussian-3 and density functional theories for a larger experimental test set, *J. Chem. Phys.*, 2000, **112**, 7374–7383.
91 P. C. Redfern, P. Zapol, L. A. Curtiss and K. Raghavachari, Assessment of gaussian-3 and density functional theories for enthalpies of formation of C1-C16 alkanes, *J. Phys. Chem. A*, 2000, **104**, 5850–5854.
92 K. W. Wiitala, T. R. Hoye and C. J. Cramer, Hybrid density functional methods empirically optimized for the computation of 13C and 1H chemical shifts in chloroform solution, *J. Chem. Theory Comput.*, 2006, **2**, 1085–1092.
93 M. Branca, V. Alezra, C. Kouklovsky and P. Archirel, Accurate conformation analysis in solution: NMR and DFT/PCM study of the S-3-(1-naphthoyl)-4-isopropyl-2,2-dimethyloxazolidin-5-one in CDCl3, *Tetrahedron*, 2008, **64**, 1743–1752.
94 Q. Yang, K. Hu, B. C. Yan, M. Liu, X. N. Li, H. D. Sun and P. T. Puno, Maoeriocalysins A-D four novel ent -kaurane diterpenoids from Isodon eriocalyx and their structure determination utilizing quantum chemical calculation in conjunction with quantitative interproton distance analysis, *Org. Chem. Front.*, 2019, **6**, 45–53.
95 Y. Yi, B. Adrjan, J. Wlodarz, J. Li, K. Jackowski and S. Roszak, NMR measurements and DFT studies of nuclear magnetic shielding in emodin and chuanxiongzine molecules, *J. Mol. Struct.*, 2018, **1166**, 304–310.
96 A. Baranowska-Łączkowska, M. Kozak, K. Z. Łączkowski and B. Fernández, Theoretical calculation of NMR shifts in newly developed antibacterial 4-formylbenzoic acid-based thiazoles, *Theor. Chem. Acc.*, 2018, **137**, 46.
97 L. B. Krivdin, Computational ^{1}H NMR: Part 1. Theoretical background, *Magn. Reson. Chem.*, 2019, **57**, 894.
98 J. C. Facelli, Chemical shift tensors: Theory and application to molecular structural problems, *Prog. Nucl. Magn. Reson. Spectrosc.*, 2011, **58**, 176–201.
99 D. L. Pavia, G. M. Lampman and G. S. Kriz, *Introduction to Spectroscopy third edition*, Thomson Learn. Inc., 2001.
100 R. M. Silverstein, F. X. Webster and D. Kiemle, *Spectrometric Identification of Organic Compounds*, Wiley, 7th edn, 2005.
101 *Chemoffice Professional 18.0*, CambridgeSotf Corp., Cambridge, 2018.
102 *ACD/NMR Predictors, version 2020*, Advanced Chemistry Development Inc., Toronto, ON, Canada, 2020.
103 A. J. Brandolini, NMRPredict Modgraph Consultants, *J. Am. Chem. Soc.*, 2006, **128**, 13313.
104 M. Pérez, T. M. Peakman, A. Alex, P. D. Higginson, J. C. Mitchell, M. J. Snowden and I. Morao, Accuracy vs time dilemma on the prediction of NMR chemical shifts: A case study (chloropyrimidines), *J. Org. Chem.*, 2006, **71**, 3103–3110.
105 S. Thomas, I. Brühl, D. Heilmann and E. Kleinpeter, ^{13}C NMR chemical shift calculations for some substituted pyridines: A comparative consideration, *J. Chem. Inf. Comput. Sci.*, 1997, **37**, 726.
106 D. J. Giesen and N. Zumbulyadis, A hybrid quantum mechanical and empirical model for the prediction of isotropic 13C shielding constants of organic molecules, *Phys. Chem. Chem. Phys.*, 2002, **4**, 5498–5507.
107 A. Bagno, F. Rastrelli and G. Saielli, Toward the complete prediction of the 1H and ^{13}C NMR spectra of complex organic molecules by DFT methods: Application to natural substances, *Chem. – Eur. J.*, 2006, **12**, 5514–5525.
108 G. J. O. Beran, Calculating nuclear magnetic resonance chemical shifts from density functional theory: A primer, *EMagRes*, 2019, **8**, 215.
109 R. Jain, T. Bally and P. R. Rablen, Calculating accurate proton chemical shifts of organic molecules with density functional methods and modest basis sets, *J. Org. Chem.*, 2009, **74**, 4017–4023.
110 M. A. Iron, Evaluation of the Factors Impacting the Accuracy of ^{13}C NMR Chemical Shift Predictions using Density Functional Theory - The Advantage of Long-Range Corrected Functionals, *J. Chem. Theory Comput.*, 2017, **13**, 5798–5819.
111 A. Auer, J. Gauss and J. F. Stanton, Quantitative prediction of gas-phase 13C nuclear magnetic shielding constants, *J. Chem. Phys.*, 2003, **118**, 10407–10417.
112 D. Flaig, M. Maurer, M. Hanni, K. Braunger, L. Kick, M. Thubauville and C. Ochsenfeld, Benchmarking Hydrogen and Carbon NMR Chemical Shifts at HF, DFT, and MP2 Levels, *J. Chem. Theory Comput.*, 2014, **10**, 572–578.
113 T. Bally and P. R. Rablen, Quantum-chemical simulation of ^{1}H NMR spectra. 2. Comparison of DFT-based procedures for computing proton-proton coupling constants in organic molecules, *J. Org. Chem.*, 2011, **76**, 4818–4830.
114 D. J. Tantillo, *CHEmical SHIft REpository with Coupling Constants Added Too*. http://cheshirenmr.info/index.htm. [online]., 2009.
115 L. B. Casabianca, Calculating nuclear magnetic resonance chemical shifts in solvated systems, *Magn. Reson. Chem.*, 2020, **58**, 611–624.
116 A. M. Sarotti and S. C. Pellegrinet, A multi-standard approach for GIAO ^{13}C NMR calculations, *J. Org. Chem.*, 2009, **74**, 7254.
117 K. G. Andrews and A. C. Spivey, Improving the accuracy of computed ^{13}C NMR shift predictions by specific environment error correction: Fragment referencing, *J. Org. Chem.*, 2013, **78**, 11302.
118 G. K. Pierens, 1H and ^{13}C NMR scaling factors for the calculation of chemical shifts in commonly used solvents

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

using density functional theory, *J. Comput. Chem.*, 2014, **35**, 1388.

119 A. E. Aliev, D. Courtier-Murias and S. Zhou, Scaling factors for carbon NMR chemical shifts obtained from DFT B3LYP calculations, *J. Mol. Struct.: THEOCHEM.*, 2009, **893**, 1–5.

120 F. L. P. Costa, A. C. F. De Albuquerque, F. Martins Dos Santos and M. B. De Amorim, GIAO-HDFT scaling factor for ^{13}C NMR chemical shifts calculation, *J. Phys. Org. Chem.*, 2010, **23**, 972–977.

121 A. G. Kutateladze and D. S. Reddy, High-Throughput in Silico Structure Validation and Revision of Halogenated Natural Products Is Enabled by Parametric Corrections to DFT-Computed ^{13}C NMR Chemical Shifts and Spin-Spin Coupling Constants, *J. Org. Chem.*, 2017, **82**, 3368.

122 W. Hehre, P. Klunzinger, B. Deppmeier, A. Driessen, N. Uchida, M. Hashimoto, E. Fukushi and Y. Takata, Efficient Protocol for Accurately Calculating 13C Chemical Shifts of Conformationally Flexible Natural Products: Scope, Assessment, and Limitations, *J. Nat. Prod.*, 2019, **82**, 2299.

123 A. T. Merrill and D. J. Tantillo, Solvent optimization and conformational flexibility effects on 1H and ^{13}C NMR scaling factors, *Magn. Reson. Chem.*, 2020, **58**, 576.

124 A. C. F. de Albuquerque, D. J. Ribeiro and M. B. de Amorim, Structural determination of complex natural products by quantum mechanical calculations of ^{13}C NMR chemical shifts: development of a parameterized protocol for terpenes, *J. Mol. Model.*, 2016, **22**, 183.

125 M. Kaupp, M. Bühl and V. G. Malkin, *Calculation of NMR and EPR Parameters, Theory and Applications*, Wiley, 1 edn, 2004.

126 R. Frach and S. M. Kast, Solvation effects on chemical shifts by embedded cluster integral equation theory, *J. Phys. Chem. A*, 2014, **118**, 11620.

127 C. C. Roggatz, M. Lorch and D. M. Benoit, Influence of Solvent Representation on Nuclear Shielding Calculations of Protonation States of Small Biological Molecules, *J. Chem. Theory Comput.*, 2018, **14**, 2684.

128 A. Warshel and M. Levitt, Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme, *J. Mol. Biol.*, 1976, **103**, 227–249.

129 G. Colherinhas, L. B. A. Oliveira, M. A. Castro, T. L. Fonseca, K. Coutinho and S. Canuto, On the calculation of magnetic properties of nucleic acids in liquid water with the sequential QM/MM method, *J. Mol. Liq.*, 2019, **294**, 111611.

130 M. S. Gordon, D. G. Fedorov, S. R. Pruitt and L. V. Slipchenko, Fragmentation methods: A route to accurate calculations on large systems, *Chem. Rev.*, 2012, **112**, 632–672.

131 B. Mennucci, Polarizable continuum model, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 386–404.

132 A. G. Kutateladze and O. A. Mukhina, Relativistic force field: Parametric computations of proton-proton coupling constants in ^{1}H NMR spectra, *J. Org. Chem.*, 2014, **79**, 8397–8406.

133 A. G. Kutateladze and O. A. Mukhina, Relativistic Force Field: Parametrization of 13C-1H Nuclear Spin-Spin Coupling Constants, *J. Org. Chem.*, 2015, **80**, 10838–10848.

134 A. G. Kutateladze and O. A. Mukhina, Minimalist Relativistic Force Field: Prediction of Proton-Proton Coupling Constants in ^{1}H NMR Spectra Is Perfected with NBO Hybridization Parameters, *J. Org. Chem.*, 2015, **80**, 5218–5225.

135 L. B. Krivdin, Computational ^{1}H NMR: Part 3. Biochemical studies, *Magn. Reson. Chem.*, 2020, **58**, 15–30.

136 D. S. Reddy and A. G. Kutateladze, Structure Revision of an Acorane Sesquiterpene Cordycepol A, *Org. Lett.*, 2016, **18**, 4860–4863.

137 C. P. Sun, A. G. Kutateladze, F. Zhao, L. X. Chen and F. Qiu, A novel withanolide with an unprecedented carbon skeleton from Physalis angulata, *Org. Biomol. Chem.*, 2017, **15**, 1110–1114.

138 S. De Vita, S. Terracciano, I. Bruno and M. G. Chini, From Natural Compounds to Bioactive Molecules through NMR and In Silico Methodologies, *Eur. J. Org. Chem.*, 2020, 6297–6317.

139 O. L. Malkina, B. Schimmelpfennig, M. Kaupp, B. A. Hess, P. Chandra, U. Wahlgren and V. G. Malkin, Spin-orbit corrections to NMR shielding constants from density functional theory. How important are the two-electron terms?, *Chem. Phys. Lett.*, 1998, **296**, 93–104.

140 T. A. Holt, D. S. Reddy, D. B. Huple, L. M. West, A. D. Rodríguez, M. T. Crimmins and A. G. Kutateladze, The Discreet Structural Diversity of Briarellins: DU8+ Guided Multiple Structure Revisions Yielded Two Unknown Structural Types, *J. Org. Chem.*, 2020, **85**, 6201–6205.

141 S. D. Rychnovsky, Predicting NMR spectra by computational methods: Structure revision of hexacyclinol, *Org. Lett.*, 2006, **13**, 2895–2898.

142 B. Schlegel, A. Härtl, H. M. Dahse, F. A. Gollmick, U. Gräfe, H. Dörfelt and B. Kappes, Hexacyclinol, a new antiproliferative metabolite of Panus rudis HKI 0254, *J. Antibiot.*, 2002, **55**, 814–817.

143 J. J. La Clair, Total syntheses of hexacyclinol, 5-epi-hexacyclinol, and desoxohexacyclinol unveil an antimalarial prodrug motif, *Angew. Chem., Int. Ed.*, 2006, **45**, 2769–2773.

144 J. A. Porco, S. Su, X. Lei, S. Bardhan and S. D. Rychnovsky, Total synthesis and structure assignment of (+)-hexacyclinol, *Angew. Chem., Int. Ed.*, 2006, **45**, 5790–5792.

145 G. Saielli and A. Bagno, Can two molecules have the same NMR spectrum? Hexacyclinol revisited, *Org. Lett.*, 2009, **11**, 1409–1412.

146 A. Aiello, E. Fattorusso, P. Luciano, A. Mangoni and M. Menna, Isolation and structure determination of aplidinones A-C from the Mediterranean ascidian Aplidium conicum: A successful regiochemistry assignment by quantum mechanical ^{13}C NMR chemical shift calculations, *Eur. J. Org. Chem.*, 2005, 5024–5030.

 This journal is © the Partner Organisations 2021

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

147 C. Bassarello, G. Bifulco, P. Montoro, A. Skhirtladze, E. Kemertelidze, C. Pizza and S. Piacente, Gloriosaols A and B, two novel phenolics from Yucca gloriosa: structural characterization and configurational assignment by a combined NMR-quantum mechanical strategy, *Tetrahedron*, 2007, **63**, 148–154.

148 D. C. Braddock and H. S. Rzepa, Structural reassignment of obtusallenes V, VI, and VII by GIAO-based density functional prediction, *J. Nat. Prod.*, 2008, **71**, 728–730.

149 K. N. White, T. Amagata, A. G. Oliver, K. Tenney, P. J. Wenzel and P. Crews, Structure revision of spiroleucettadine, a sponge alkaloid with a bicyclic core meager in H-atoms, *J. Org. Chem.*, 2008, **73**, 8719–8722.

150 C. Fattorusso, E. Stendardo, G. Appendino, E. Fattorusso, P. Luciano, A. Romano and O. Taglialatela-Scafati, Artarborol, a nor-caryophyllane sesquiterpene alcohol from artemisia arborescens. Stereostructure assignment through concurrence of NMR data and computational analysis, *Org. Lett.*, 2007, **9**, 2377–2380.

151 F. M. Fernando, L. S. M. Velozo, E. M. De Carvalho, A. M. Marques, R. M. Borges, A. P. F. Trindade, M. I. S. Dos Santos, A. C. F. De Albuquerque, F. L. P. Costa, M. A. C. Kaplan and M. B. De Amorim, 3-Ishwarone, a rare ishwarane sesquiterpene from peperomia scandens ruiz & pavon: Structural elucidation through a joint experimental and theoretical study, *Molecules*, 2013, **18**, 13520–13529.

152 J. X. Pu, S. X. Huang, J. Ren, W. L. Xiao, L. M. Li, R. T. Li, A. B. Li, T. G. Liao, L. G. Lou, H. J. Zhu and H. D. Sun, Isolation and structure elucidation of kadlongilactones C-F from Kadsura longipedunculata by NMR spectroscopy and DFT computational methods, *J. Nat. Prod.*, 2007, **70**, 1707–1711.

153 S. G. Smith, R. S. Paton, J. W. Burton and J. M. Goodman, Stereostructure assignment of flexible five-membered rings by GIAO ^{13}C NMR calculations: Prediction of the stereochemistry of elatenyne, *J. Org. Chem.*, 2008, **73**, 4053–4062.

154 A. Macedo, L. Martorano, A. C. de Albuquerque, R. Fiorot, J. Carneiro, V. Campos, T. Vasconcelos, A. Valverde, D. Moreira and F. dos Santos Jr., Absolute Configuration of (−)-Cubebin, a Classical Lignan with Pharmacological Potential, Defined by Means of Chiroptical Spectroscopy, *J. Braz. Chem. Soc.*, 2020, **31**, 2030–2037.

155 D. J. Shepherd, P. A. Broadwith, B. S. Dyson, R. S. Paton and J. W. Burton, Structure reassignment of laurefurenynes A and B by computation and total synthesis, *Chem. – Eur. J.*, 2013, **19**, 12644.

156 D. F. Llompart, A. M. Sarotti, V. Corne, A. G. Suárez, R. A. Spanevello, G. A. Echeverría, O. E. Piro and E. E. Castellano, Asymmetric construction of substituted pyrrolidines via 1,3-dipolar cycloaddition of azomethine ylides and chiral acrylates derived from biomass, *Tetrahedron Lett.*, 2014, **55**, 2394.

157 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 and A. H. Daranas, Oxasqualenoids from Laurencia viridis: Combined spectroscopic-computational analysis and antifouling potential, *J. Nat. Prod.*, 2015, **78**, 712.

158 F. M. S. Junior, C. L. Covington, A. C. F. De Albuquerque, J. F. R. Lobo, R. M. Borges, M. B. De Amorim and P. L. Polavarapu, Absolute Configuration of (-)-Centratherin, a Sesquiterpenoid Lactone, Defined by Means of Chiroptical Spectroscopy, *J. Nat. Prod.*, 2015, **78**, 2617–2623.

159 I. H. Hwang, J. Oh, W. Zhou, S. Park, J. H. Kim, A. G. Chittiboyina, D. Ferreira, G. Y. Song, S. Oh, M. Na and M. T. Hamann, Cytotoxic activity of rearranged drimane meroterpenoids against colon cancer cells via down-regulation of β-catenin expression, *J. Nat. Prod.*, 2015, **78**, 453.

160 S. G. Smith, J. A. Channon, I. Paterson and J. M. Goodman, The stereochemical assignment of acyclic polyols: A computational study of the NMR data of a library of stereopentad sequences from polyketide natural products, *Tetrahedron.*, 2010, **66**, 6437.

161 T. D. Tran, N. B. Pham and R. J. Quinn, Structure determination of pentacyclic pyridoacridine alkaloids from the australian marine organisms ancorina geodides and cnemidocarpa stolonifera, *Eur. J. Org. Chem.*, 2014, 4805.

162 S. G. Smith and J. M. Goodman, Assigning stereochemistry to single diastereoisomers by GIAO NMR calculation: The DP4 probability, *J. Am. Chem. Soc.*, 2010, **132**, 12946–12959.

163 E. Cairns, M. A. Hashmi, A. J. Singh, G. Eakins, M. Lein and R. Keyzers, Structure of Echivulgarine, a Pyrrolizidine alkaloid isolated from the pollen of Echium vulgare, *J. Agric. Food Chem.*, 2015, **63**, 7421.

164 F. Cen-Pacheco, M. Norte, J. J. Fernández and A. H. Daranas, Zoaramine, a zoanthamine-like alkaloid with a new skeleton, *Org. Lett.*, 2014, **16**, 2880.

165 M. J. Riveira, C. Gayathri, A. Navarro-Vázquez, N. V. Tsarevsky, R. R. Gil and M. P. Mischne, Unprecedented stereoselective synthesis of cyclopenta[b] benzofuran derivatives and their characterisation assisted by aligned media NMR and 13C chemical shift ab initio predictions, *Org. Biomol. Chem.*, 2011, **9**, 3170–3175.

166 M. W. Lodewyk and D. J. Tantillo, Prediction of the structure of nobilisitine A using computed NMR chemical shifts, *J. Nat. Prod.*, 2011, **74**, 1339–1343.

167 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 and A. E. Wright, Leiodermatolide, a potent antimitotic macrolide from the marine sponge leiodermatium sp., *Angew. Chemie*, 2011, **123**, 3277–3281.

168 S. G. Brown, M. J. Jansma and T. R. Hoye, Case study of empirical and computational chemical shift analyses: Reassignment of the relative configuration of phomopsichalasin to that of Diaporthichalasin, *J. Nat. Prod.*, 2012, **75**, 1326–1331.

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

169 S. K. Amini, Assignment of absolute configuration of haladyson by computed NMR chemical shifts, *J. Iran. Chem. Soc.*, 2014, **11**, 179.
170 T. P. Wyche, Y. Hou, E. Vazquez-Rivera, D. Braun and T. S. Bugni, Peptidolipins B-F, antibacterial lipopeptides from an ascidian-derived Nocardia sp., *J. Nat. Prod.*, 2012, **75**, 735–740.
171 V. L. Challinor, R. C. Johnston, P. V. Bernhardt, R. P. Lehmann, E. H. Krenske and J. J. De Voss, Biosynthetic insights provided by unusual sesterterpenes from the medicinal herb Aletris farinosa, *Chem. Sci.*, 2015, **6**, 5740.
172 C. I. MacGregor, B. Y. Han, J. M. Goodman and I. Paterson, Toward the stereochemical assignment and synthesis of hemicalide: DP4f GIAO-NMR analysis and synthesis of a reassigned C16-C28 subunit, *Chem. Commun.*, 2016, **52**, 4632–4635.
173 K. Ermanis, K. E. B. Parkes, T. Agback and J. M. Goodman, Expanding DP4: Application to drug compounds and automation, *Org. Biomol. Chem.*, 2016, **14**, 3943–3949.
174 N. Grimblat, M. M. Zanardi and A. M. Sarotti, Beyond DP4: An Improved Probability for the Stereochemical Assignment of Isomeric Compounds using Quantum Chemical Calculations of NMR Shifts, *J. Org. Chem.*, 2015, **80**, 12526–12534.
175 R. B. Nazarski, B. Pasternak and S. Leśniak, Three-component conformational equilibria of some flexible pyrrolidin-2-(thi)ones in solution by NMR data (δC, δh, and nJHH) and their DFT predictions: A confrontation of different approaches, *Tetrahedron*, 2011, **67**, 6901–6916.
176 A. N. L. Batista, F. M. Dos Santos, A. L. Valverde and J. M. Batista, Stereochemistry of spongosoritins: Beyond optical rotation, *Org. Biomol. Chem.*, 2019, **17**, 9772.
177 D. Dardić, G. Lauro, G. Bifulco, P. Laboudie, P. Sakhaii, A. Bauer, A. Vilcinskas, P. E. Hammann and A. Plaza, Svetamycins A-G, Unusual Piperazic Acid-Containing Peptides from Streptomyces sp., *J. Org. Chem.*, 2017, **82**, 6032–6043.
178 L. H. Martorano, A. L. Valverde, C. M. R. Ribeiro, A. C. F. De Albuquerque, J. W. M. Carneiro, R. G. Fiorot and F. M. Dos Santos Junior, Unraveling the helianane family: A complementary quantum mechanical study, *New J. Chem.*, 2020, **44**, 8055.
179 M. M. Zanardi, A. G. Suárez and A. M. Sarotti, Determination of the Relative Configuration of Terminal and Spiroepoxides by Computational Methods. Advantages of the Inclusion of Unscaled Data, *J. Org. Chem.*, 2017, **82**, 1873–1879.
180 P. Sun, Q. Yu, J. Li, R. Riccio, G. Lauro, G. Bifulco, T. Kurtán, A. Mándi, H. Tang, T. J. Li, C. L. Zhuang, W. H. Gerwick and W. Zhang, Bissubvilides A and B, Cembrane-capnosane heterodimers from the soft coral sarcophyton subviride, *J. Nat. Prod.*, 2016, **79**, 2552–2558.
181 Y. Y. Fan, L. S. Gan, H. C. Liu, H. Li, C. H. Xu, J. P. Zuo, J. Ding, J. M. Yue and A. Phainanolide, Highly Modified and Oxygenated Triterpenoid from Phyllanthus hainanensis, *Org. Lett.*, 2017, **19**, 4580–4583.
182 K. Ermanis, K. E. B. Parkes, T. Agback and J. M. Goodman, Doubling the power of DP4 for computational structure elucidation, *Org. Biomol. Chem.*, 2017, **15**, 8998–9007.
183 K. Ermanis, K. E. B. Parkes, T. Agback and J. M. Goodman, The optimal DFT approach in DP4 NMR structure analysis-pushing the limits of relative configuration elucidation, *Org. Biomol. Chem.*, 2019, **17**, 5886.
184 D. Xin, P. J. Jones and N. C. Gonnella, D iCE: Diastereomeric in Silico Chiral Elucidation, Expanded DP4 Probability Theory Method for Diastereomer and Structural Assignment, *J. Org. Chem.*, 2018, **83**, 5035–5043.
185 N. Grimblat, J. A. Gavín, A. H. Daranas and A. M. Sarotti, Combining the Power of J Coupling and DP4 Analysis on Stereochemical Assignments: The J-DP4 Methods, *Org. Lett.*, 2019, **41**, 4003–4007.
186 S. W. Li, C. Cuadrado, L. G. Yao, A. H. Daranas and Y. W. Guo, Quantum Mechanical-NMR-Aided Configuration and Conformation of Two Unreported Macrocycles Isolated from the Soft Coral Lobophytum sp.: Energy Calculations versus Coupling Constants, *Org. Lett.*, 2020, **22**, 4093–4096.
187 A. M. Sarotti, Successful combination of computationally inexpensive GIAO ^{13}C NMR calculations and artificial neural network pattern recognition: A new strategy for simple and rapid detection of structural misassignments, *Org. Biomol. Chem.*, 2013, **11**, 4847.
188 M. M. Zanardi and A. M. Sarotti, GIAO C-H COSY Simulations Merged with Artificial Neural Networks Pattern Recognition Analysis. Pushing the Structural Validation a Step Forward, *J. Org. Chem.*, 2015, **80**, 9371–9378.
189 D. B. Pu, B. W. Du, W. Chen, J. B. Gao, K. Hu, N. Shi, Y. M. Li, X. J. Zhang, R. H. Zhang, X. N. Li, H. B. Zhang, F. Wang and W. L. Xiao, Premnafulvol A: A Diterpenoid with a 6/5/7/3-Fused Tetracyclic Core and Its Biosynthetically Related Analogues from Premna fulva, *Org. Lett.*, 2018, **20**, 6314–6317.
190 J. W. Tang, L. M. Kong, W. Y. Zu, K. Hu, X. N. Li, B. C. Yan, W. G. Wang, H. D. Sun, Y. Li and P. T. Puno, Isopenicins A-C: Two Types of Antitumor Meroterpenoids from the Plant Endophytic Fungus Penicillium sp. sh18, *Org. Lett.*, 2019, **21**, 771–775.
191 P. H. Willoughby, M. J. Jansma and T. R. Hoye, A guide to small-molecule structure assignment through computation of (1H and 13C) NMR chemical shifts, *Nat. Protoc.*, 2014, **9**, 643–660.
192 M. A. Thompson, Molecular docking using ArgusLab, an efficient shape-based search algorithm and the AScore scoring function, in *ACS Meet.*, 2004.
193 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeerschd, E. Zurek and G. R. Hutchison, Avogadro: An advanced semantic chemical editor, visualization, and analysis platform, *J. Cheminform.*, 2012, **4**, 17.

 This journal is © the Partner Organisations 2021

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

194 *ACD/ChemSketch, version 2020.1.2*, Advanced Chemistry Development, Inc., Toronto, ON, Canada, http://www.acdlabs.com, 2020.
195 K. R. Cousins, *ChemDraw Ultra 9.0. CambridgeSoft*, 100 CambridgePark Drive, Cambridge, MA 02140. http://www.cambridgesoft.com. See Web site for pricing options, *J. Am. Chem. Soc.*, 2005, **127**, 4115–4116.
196 M. Froimowitz, HyperChem(TM): A software package for computational chemistry and molecular modeling, *Biotechniques*, 1993, **14**, 1010–1013.
197 R. Dennington, T. A. Keith and J. M. Millam, *GaussView, Version 6.1*, Semichem Inc., Shawnee Mission, KS, 2016.
198 *Spartan '10. Version 1.1.0*, Wavefunction Inc., Irvine, CA, 2011.
199 *MacroModel*, Schrödinger, LLC, New York, NY, 2020.
200 *Conflex Program*, CONFLEX Corporation, Tokyo, Japan, http://www.conflex.us.
201 G. M. J. Frisch, W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, G. Cheeseman, J. R. Scalmani, V. Barone, B. Mennucci, H. Petersson, G. A. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, K. Hada, M. Ehara, M. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. E. Montgomery, J. A. Peralta Jr., F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, R. Staroverov, V. N. Kobayashi, K. Normand, J. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, S. Dannenberg, J. J. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09, Revision E. 01*, Gaussian, 2009.
202 F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka and F. Weigend, Turbomole, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2014, **4**, 91–100.
203 J. A. Rackers, Z. Wang, C. Lu, M. L. Laury, L. Lagardère, M. J. Schnieders, J. P. Piquemal, P. Ren and J. W. Ponder, Tinker 8: Software Tools for Molecular Design, *J. Chem. Theory Comput.*, 2018, **14**, 5273–5289.
204 M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis and J. A. Montgomery, General atomic and molecular electronic structure system, *J. Comput. Chem.*, 1993, **14**, 1347–1363.
205 M. S. Gordon and M. W. Schmidt, Advances in electronic structure theory: GAMESS a decade later, *Theory Appl. Comput. Chem.*, 2005, 1167–1189.
206 Y. S. Cai, A. M. Sarotti, T. L. Zhou, R. Huang, G. Qiu, C. Tian, Z. H. Miao, A. Mándi, T. Kurtán, S. Cao and S. P. Yang, Flabellipparicine, a Flabelliformide-Apparicine-Type Bisindole Alkaloid from Tabernaemontana divaricata, *J. Nat. Prod.*, 2018, **81**, 1976–1983.
207 M. Alilou, S. Marzocco, H. Batooli, J. Troppmair, S. Schwaiger and H. Stuppner, Unusual secondary metabolites of the aerial parts of dionysia diapensifolia bioss. (primulaceae) and their anti-inflammatory activity, *Biomolecules*, 2020, **10**, 438.
208 V. M. Schmiedel, Y. J. Hong, D. Lentz, D. J. Tantillo and M. Christmann, Synthesis and Structure Revision of Dichrocephones A and B, *Angew. Chem., Int. Ed.*, 2018, **57**, 2419–2422.
209 X. Tian, L. Li, Y. Hu, H. Zhang, Y. Liu, H. Chen, G. Ding and Z. Zou, Dichrocephones A and B, two cytotoxic sesquiterpenoids with the unique [3.3.3] propellane nucleus skeleton from Dichrocephala benthamii, *RSC Adv.*, 2013, **3**, 7880–7883.
210 A. G. Kutateladze and D. M. Kuznetsov, Triquinanes and Related Sesquiterpenes Revisited Computationally: Structure Corrections of Hirsutanols B and D, Hirsutenol E, Cucumin B, Antrodins C-E, Chondroterpenes A and H, Chondrosterins C and E, Dichrocephone A, and Pethybrene, *J. Org. Chem.*, 2017, **82**, 10795–10802.
211 M. Bae, J. S. An, S. H. Hong, E. S. Bae, B. Chung, Y. Kwon, S. Hong, K. B. Oh, J. Shin, S. K. Lee and D. C. Oh, Donghaecyclinones A-C: New cytotoxic rearranged angucyclinones from a volcanic island-derived marine Streptomyces sp., *Mar. Drugs*, 2020, **18**, 121.
212 F. Wang, F. Wang, A. M. Sarotti, G. Jiang, J. C. Huguet-Tapia, S. L. Zheng, X. Wu, C. Li, Y. Ding, S. Cao and S. Cao, Waikikiamides A-C: Complex Diketopiperazine Dimer and Diketopiperazine-Polyketide Hybrids from a Hawaiian Marine Fungal Strain Aspergillus sp. FM242, *Org. Lett.*, 2020, **22**, 4408–4412.
213 A. M. Sarotti, Structural revision of two unusual rhamnofolane diterpenes, curcusones I and J, by means of DFT calculations of NMR shifts and coupling constants, *Org. Biomol. Chem.*, 2018, **16**, 944–950.
214 J. Q. Liu, Y. F. Yang, X. Y. Li, E. Q. Liu, Z. R. Li, L. Zhou, Y. Li and M. H. Qiu, Cytotoxicity of naturally occurring rhamnofolane diterpenes from Jatropha curcas, *Phytochemistry*, 2013, **96**, 265–272.
215 Y. Li and M. Dai, Total Syntheses of the Reported Structures of Curcusones I and J through Tandem Gold Catalysis, *Angew. Chem., Int. Ed.*, 2017, **56**, 11624–11627.
216 N. Grimblat, T. S. Kaufman and A. M. Sarotti, Computational Chemistry Driven Solution to Rubriflordilactone B, *Org. Lett.*, 2016, **18**, 6420–6423.
217 W. L. Xiao, L. M. Yang, N. B. Gong, L. Wu, R. R. Wang, J. X. Pu, X. L. Li, S. X. Huang, Y. T. Zheng, R. T. Li, Y. Lu, Q. T. Zheng and H. D. Sun, Rubriflordilactones A and B, two novel bisnortriterpenoids from Schisandra rubriflora and their biological activities, *Org. Lett.*, 2006, **8**, 991–994.

218 P. Yang, M. Yao, J. Li, Y. Li and A. Li, Total Synthesis of Rubriflordilactone B, *Angew. Chem., Int. Ed.*, 2016, **55**, 6964.
219 Y. Liu, T. A. Holt, A. Kutateladze and T. R. Newhouse, Stereochemical revision of xylogranatin F by GIAO and DU8+NMR calculations, *Chirality*, 2020, **32**, 515–523.
220 J. Wu, S. Zhang, T. Bruhn, Q. Xiao, H. Ding and G. Bringmann, Xylogranatins F-R: Antifeedants from the Chinese mangrove, Xylocarpus granatum, a new biogenetic pathway to tetranortriterpenoids, *Chem. – Eur. J.*, 2008, **14**, 1129–1144.
221 O. Corminboeuf, L. E. Overman and L. D. Pennington, A unified strategy for enantioselective total synthesis of cladiellin and briarellin diterpenes: Total synthesis of briarellins E and F and the putative structure of alcyonin and revision of its structure assignment, *J. Org. Chem.*, 2009, **74**, 5458–5470.
222 M. T. Crimmins, M. C. Mans and A. D. Rodríguez, Total synthesis of the proposed structure of briarellin, *Org. Lett.*, 2010, **12**, 5028–5031.
223 A. Howarth, K. Ermanis and J. M. Goodman, DP4-AI automated NMR data analysis: straight from spectrometer to structure, *Chem. Sci.*, 2020, **11**, 4351–4359.
224 M. O. Marcarino, M. M. Zanardi and A. M. Sarotti, The Risks of Automation: A Study on DFT Energy Miscalculations and Its Consequences in NMR-based Structural Elucidation, *Org. Lett.*, 2020, **22**, 3561–3565.

Published on 22 February 2021. Downloaded on 5/15/2021 8:32:12 AM.

This journal is © the Partner Organisations 2021