Organic & Biomolecular Chemistry

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/d4ob01959h

Overcoming the ring tension: computational approaches to stereochemical assignments and geometrical insights in small heterocycles†

Ezequiel R. Luciano,[a] Milagros Amichetti,[a,b] Ariel M. Sarotti *[b] and Maria M. Zanardi*[a]

Received 3rd December 2024,
Accepted 27th January 2025

DOI: 10.1039/d4ob01959h

rsc.li/obc

In this study, we evaluate the performance of the DP4+ and MM-DP4+ methods on molecules featuring small heterocyclic rings. A dataset of 71 molecules containing three- and four-membered heterocycles, known for their stereochemical assignment challenges, was analyzed. We compared molecular geometries optimized at different computational levels, including MMFF and B3LYP/6-31G*, to assess deviations in key geometric parameters relative to the heterocycle structures. Furthermore, the geometric properties of these molecules were investigated using various force fields to evaluate their differences. Our aim was to assess the reliability of B3LYP/6-31G* in comparison with more accurate methods and to elucidate how different force fields influence in geometric precision.

Introduction

The chemistry of strained heterocycles, especially the three- and four membered rings like epoxides, aziridines, azetidines and oxetanes, has played an important role in the development of modern organic chemistry. Their high reactivity makes them versatile synthons for organic synthesis, as well as valuable synthetic targets.[1] The presence of an heteroatom in the strained ring imparts a dipole moment to the molecule which increase its reactivity. Due to the compressed bond angles (around 60° in three-membered rings and 90° in four-membered rings), their synthesis and structural elucidation are challenging.

One of the most recognized characteristics of small heterocycles is their tendency to undergo ring-opening reactions, involving cleavage of a carbon–heteroatom bond and formation of a new bond with a nucleophile. Significant progress has been made in the photochemistry, regioselectivity, and stereoselectivity of reactions involving these three- and four-membered rings. Moreover, there is increasing interest in natural products containing these functionalities, along with studies on their biological properties and polymerization behavior.[1]

The structural elucidation of strained cyclic systems presents significant challenges due to their unique electronic and steric properties. These difficulties arise from overlapping NMR signals, ambiguous spin–spin coupling constants (SSCCs), and complex NOE patterns, all of which hinder accurate stereochemical determinations.[2–5]

Three-membered rings like epoxides and aziridines exemplify these issues. Epoxides, ubiquitous in natural and synthetic compounds, are frequently misassigned due to the minimal pyramidalization of their carbons and the similarity of SSCC ranges for *syn*- and *anti*-configurations. Aziridines, though less common, have gained prominence for their high reactivity and biological relevance, yet their strained nature and synthetic accessibility pose additional elucidation challenges.[6] Four-membered rings, while less explored, also demand attention. Oxetanes are difficult to detect due to overlapping ^{13}C NMR signals with other ethers or alcohols, whereas azetidines, despite their applications in medicinal chemistry, are understudied due to synthetic hurdles. Accurate stereochemical assignment remains essential, especially for derivatives like β-lactams, critical in pharmacology and peptidomimetics.[7]

This background underscores the persistent challenges in structural characterization and reinforces the need for advanced methodologies. While NMR spectroscopy is a valuable tool for determining conformational and configurational patterns in solution, chemical shifts, *J*-couplings, and nuclear

[a]*Instituto de Investigaciones en Ingeniería Ambiental, Química y Biotecnología Aplicada (INGEBIO), Facultad de Química e Ingeniería del Rosario, Pontificia Universidad Católica Argentina, Av. Pellegrini 3314, Rosario 2000, Argentina. E-mail: zanardi@rosario-conicet.gov.ar*

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

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4ob01959h

Overhauser effects are sometimes insufficient for full elucidation.[4,5,8] Nevertheless, density functional theory (DFT) calculations of NMR parameters, have become a very popular complement for synthetic and natural product chemistry. Achieving a strong correlation between calculated and experimental data serves as an excellent tool to support structural analysis. In this regard, several computer-assisted structural elucidation (CASE) methods have been developed, combining theoretical and experimental data. These methods provide powerful alternatives to X-ray crystallography when crystallographic data is unavailable. Since the introduction of the CP3[9] and DP4[10] from Goodman group, a growing number of sophisticated approaches with variable confidence, computational cost and complexity, have been published.[11–13] Among them can be mentioned DP4,[10] DP4+,[14] *J*-DP4,[15] ML-*J*-DP4,[16] MM-DP4+,[17] DICE,[18] CASE-3D,[19] and DU8+,[20] among others.

Over the years, DP4+ has become one of the most widely adopted methods within the scientific community, because of its user-friendliness and consistently reliable performance.[21] In short, DP4+ is an adjusted version of DP4 probability,[14] in which $P(i)$ represents the likelihood that candidate i (from a set of m isomers) is correct.[10] The method is grounded in the idea that the errors are independent and follow a *t*-distribution, defined by the mean (μ), standard deviation (σ), and degrees of freedom (ν). With this framework, the probability for each error is computed, and Bayes' theorem is used to calculate the overall probability for each molecule. In DP4+, the chemical shifts are computed from B3LYP/6-31G* optimized geometries, with the PCM/mPW1PW91/6-31+G** level being the recommended one based on CPU time and overall prediction accuracy.[14] Another key difference with DP4 is that probabilities are calculated using both scaled and unscaled chemical shifts, referred to as *s*DP4+ and *u*DP4+, respectively, which significantly enhances the assignment reliability.[14] The scaling procedure is done to remove systematic errors during DFT calculations (see Computational details), leading to chemical shifts that are closer to the experimental values.

Additionally, those probabilities can be computed using only ^{1}H NMR data, ^{13}C NMR data, or a combination of both (full data). The inclusion of unscaled data, as well as the higher level of calculation used for NMR, positively impacts the method's performance. However, it increases the computational cost, mainly due to the expensive DFT-level optimizations. To address this limitation, we recently developed MM-DP4+, a conceptually similar approach that uses geometries optimized at the MMFF level, resulting in an average time savings of 75%. After an extensive exploration of theory levels, it was concluded that SMD/ωB97XD/6-31+G** is optimal for MM-DP4+, achieving 90.5% accuracy relative to DP4+.[17]

A frequently overlooked but crucial step in these methods is the molecular mechanics (MM) conformational search, which generates conformer ensembles for each candidate structure.[22] Over the years, molecular mechanics force fields (FF) have evolved significantly, and the performance of different force fields varies depending on the system.[23] The MMFF force field is well-established for organic molecules and has been used in conformational searches for common structural prediction metrics like DP4, *J*-DP4, and MM-DP4+.[10,15,17] The choice of force field affects the conformational landscape, influencing the number and types of conformers identified within a given cutoff. This variation can be critical, as missing key conformations may lead to incorrect results. Additionally, the impact of subtle geometric differences on predicted NMR parameters remains uncertain but potentially significant.[24]

In a previous study, we evaluated the performance of DP4+ for the stereochemical assignment of 32 spiroepoxides.[5] Spirooxiranes, commonly formed *via* epoxidation of exocyclic double bonds or Corey–Chaykovsky epoxidation of carbonyl groups, present significant challenges in stereochemical assignment. This comprehensive analysis, utilizing DFT-optimized geometries, delivered excellent results, with accurate stereochemical assignment of all the compounds studied.[5]

Following the same framework, we extended our investigation to include a broader set of constrained heterocycles. In this work, we explore the scope and limitations of the Bayesian methods DP4+ and MM-DP4+ for stereochemical assignment of these challenging molecules: epoxides, aziridines, azetidines and oxetanes (Fig. 1), aiming to identify the optimal balance between accuracy and computational cost. Additionally, we analyze variations in geometry optimization using different force fields (FF) and DFT levels, assessing their impact on the overall quality of the geometries. This study aims to provide a systematic evaluation of DP4+ and MM-DP4+ methodologies, offering practitioners a robust framework to enhance the reliability of stereochemical assignments across diverse molecular scenarios.

Results and discussion

To investigate the impact of geometric quality on NMR calculations and, consequently, on the *in silico* stereochemical assignment of these constrained rings, we analyzed 39 new examples—8 epoxides, 9 aziridines, 9 oxetanes, and 13 azetidines—as shown in Fig. 2, along with 32 previously studied spiroepoxides.[5] This resulted in a comprehensive dataset of 71 examples, examined using both the DP4+ method and the recent MM-DP4+ approach. The selection of molecules in this

Fig. 1 Functional groups explored in this work.

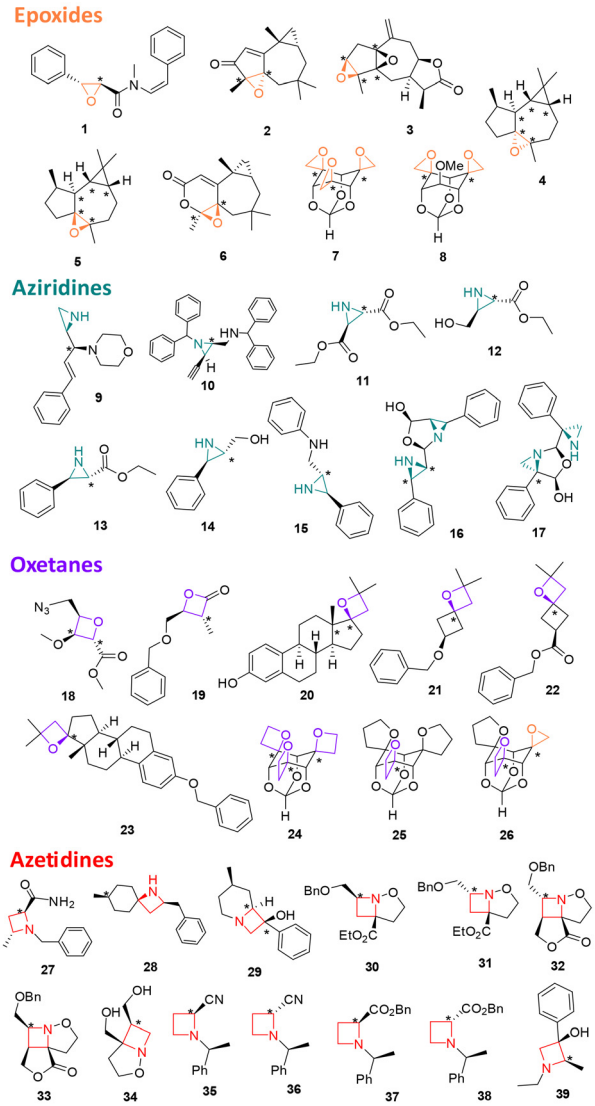

Fig. 2 Set of new compounds analysed by DP4+ and MM-DP4+. The different diastereoisomers were generated by varying the configurations at the carbons marked with an asterisk.

study was made to cover a broad spectrum of chemical scenarios, providing a comprehensive and unbiased evaluation of DP4+ and MM-DP4+, and allowing for a more generalizable understanding of their performance. Following the MM-DP4+ procedure, the chemical shifts of the corresponding correct isomers along with all possible diastereoisomers were computed at the SMD/ωB97XD/6-31+G**//MMFF level of theory using the GIAO method. All the MMFF geometries were subsequently re-optimized at B3LYP/6-31G* for further NMR calculations at the PCM/mPW1PW91/6-31+G** level. At this point, it is important to clarify that different levels were used because previous studies demonstrated them to be optimal for each method (SMD/ωB97XD/6-31+G** for MM-DP4+ and PCM/mPW1PW91/6-31+G** for DP4+, respectively). However, to better quantify the impact of the geometric change, the MM-DP4+ results were recalculated at the PCM/mPW1PW91/6-31+G**//MMFF level. The whole procedure was automated using our recently developed DP4+ App application, available at **https://github.com/Sarotti-Lab**.

The collected results are shown in Fig. 3, including the unscaled, scaled and full probabilities calculated using ^{1}H NMR data, ^{13}C NMR data, or both. To enhance the visual analysis of the results, a color scale is utilized: green indicates correctly assigned compounds with high probability, while red highlights misassigned compounds, meaning those associated with low probability. The exact values obtained in each case are provided in the ESI.†

Upon analyzing the data in Fig. 3, it is evident that DP4+ demonstrates outstanding performance in stereochemical assignment of strained heterocycles, with all compounds correctly identified. Moreover, the probabilities associated with the correct isomers are notably high (>90%), except for compound **62** (see ESI†), where it is slightly lower (70%), still reflecting a high level of confidence. When comparing the results from *u*DP4+ and *s*DP4+, it can be seen that, while both methods generally lead to the same conclusion, *u*DP4+ exhibits superior performance, with an average probability of 96.3% and an accuracy of 98.6% (defined as the % of examples correctly assigned), compared to *s*DP4+'s 92.7% average probability and 94.4% accuracy. This indicates that incorporating unscaled chemical shifts enhances stereochemical discrimination, consistent with earlier observations.[5,14,21,25] Regarding the impact of data type, our results indicate that proton data generally provide better discrimination than carbon data. This

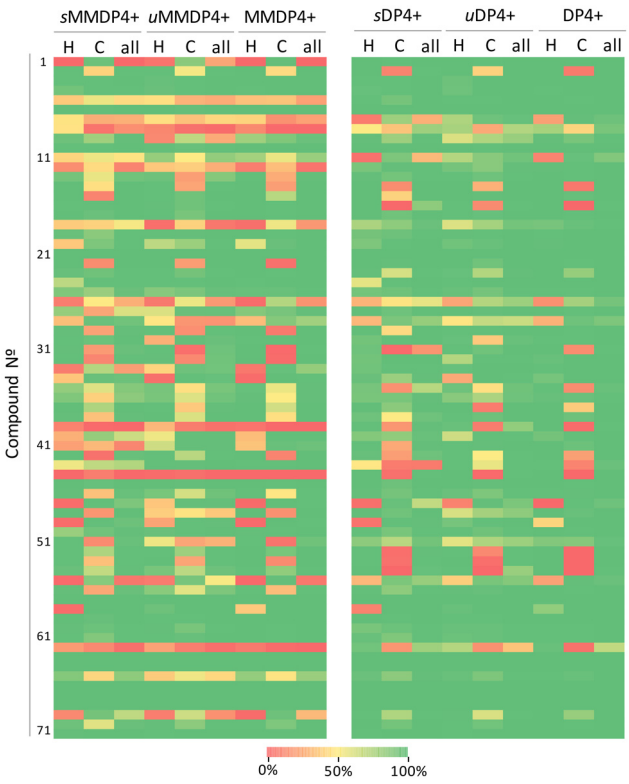

Fig. 3 Overall performance of MM-DP4+ (left) and DP4+ (right).

finding is new, as previous studies typically showed greater parity.[5] Nevertheless, we observed a compensatory effect in more sensitive systems, such as compounds **2**, **3**, **20**, **23**, **31**, **35**, and **39**. In these cases, a less reliable assignment based on ^{1}H data is balanced by ^{13}C data, or *vice versa*, highlighting the importance of using both. Therefore, DP4+ delivered outstanding results in determining the configuration of strained heterocycles. However, it is essential to acknowledge that DP4+ relies on geometry optimizations at the DFT level, which raises its computational expense. Therefore, we investigated the performance of the fastest MM-DP4+. As shown in Fig. 3, the data dispersion is significantly greater than that observed for DP4+, suggesting that these structural motifs pose considerable challenges for elucidation. This leads to a reduction in predictive capability, with 59 compounds correctly assigned, representing 83.1% of the total. A breakdown of the data types reveals several interesting findings. On one hand, *u*MM-DP4+ offers slightly lower predictive accuracy than *s*MM-DP4+ (77.5% *vs.* 81.7%, respectively), in contrast to the observations made for DP4+. The results obtained with *s*MM-DP4+ are very similar to those obtained with MM-DP4+. However, a detailed analysis reveals interesting trends. In 85% of the cases, both *u*MM-DP4+ and *s*MM-DP4+ indicate the same direction, either in favor of the correct isomer (51 cases) or against it (9 cases). In the remaining 11 examples (15% of the total), the assignments differ in direction, with 7 of these cases being worsened by *u*MM-DP4+ and 4 by *s*MM-DP4+. This suggests that, in this instance, the inclusion of unscaled data is not statistically significant, unlike what is observed for a broader set of molecules.[17] Regarding the effect of the type of nucleus, it is again observed that protons are more decisive than carbons, but in a negative sense. Specifically, of the twelve examples analyzed using MM-DP4+, all incorrectly assigned cases showed errors in the ^{1}H data, while only two exhibited incorrect assignments in the ^{13}C data (Fig. 4).

When the MM-DP4+ results were recalculated using the same level as DP4+ (mPW1PW91/6-31+G**), a slightly lower classification performance was observed (78.9% *vs.* 83.1%), compared to the optimal MM-DP4+ level (SMD/ωB97XD/6-31+G**). This reinforces the choice of SMD/ωB97XD/6-31+G** for MM-DP4+ calculations. In addition, given that the mathematical formalism of both methods is the same and that the NMR calculations are performed at comparable levels (PCM/mPW1PW91/6-31+G** in DP4+ and SMD/ωB97XD/6-31+G** in MM-DP4+), it can be concluded that the 17% reduction in the predictive capacity of the latter is attributed to geometric changes arising from the geometry optimization. In this context, it is noteworthy that oxygenated compounds tend to fail more frequently (~20%) than nitrogenated ones (~10%), with no significant differences observed regarding ring size (Fig. 5).

Thereby, while MM-DP4+ exhibited strong performance given its low computational cost, its results were not as accurate as those of DP4+. It is well established that small structural nuances can have a significant impact on chemical shifts, thus influencing the assignment process.[24] To investigate the discrepancies, a detailed examination was conducted on specific cases where MM-DP4+ failed to provide accurate predictions, aiming to identify potential sources of error. It is important to emphasize that the relationship between probability magnitude and differences in the computed chemical shifts of candidates is not straightforward; both substantial differences in a single nucleus and small variations across multiple nuclei can result in high probabilities.[21] Nevertheless, analyzing the magnitude of the errors can provide valuable insights into the reasons behind an incorrect assignment.

Upon comparing the calculated values with those of the correct isomer, the global errors in MM-DP4+ were found to be larger than in DP4+ (2.0 *vs.* 1.4 ppm for ^{13}C, and 0.15 *vs.* 0.09 ppm for ^{1}H), as expected due to the lower quality of MMFF geometries. However, when focusing exclusively on the errors associated with the heterocyclic nuclei, the discrepancies in MM-DP4+ were notably higher than the overall average (2.2 ppm and 0.16 ppm), whereas in DP4+ they were lower than average (1.2 and 0.08 ppm). These observations were consistent when using scaled chemical shifts, and qualitatively similar results were obtained with unscaled shifts (see ESI†), exposing a deficiency in MM-DP4+ regarding these specific cases. The poor prediction of the heterocyclic region had such a significant effect that, in most of the misassigned cases, removing the signals from this region led to a considerable

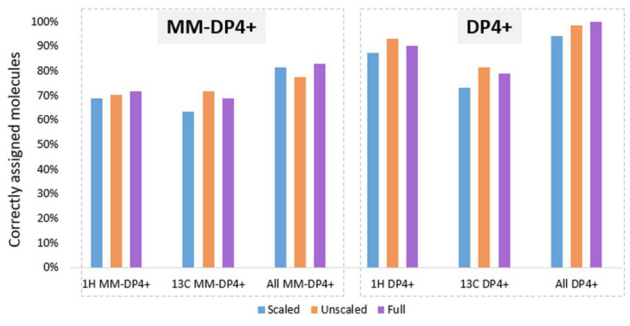

Fig. 4 MM-DP4+ and DP4+ scaled, unscaled and full performance.

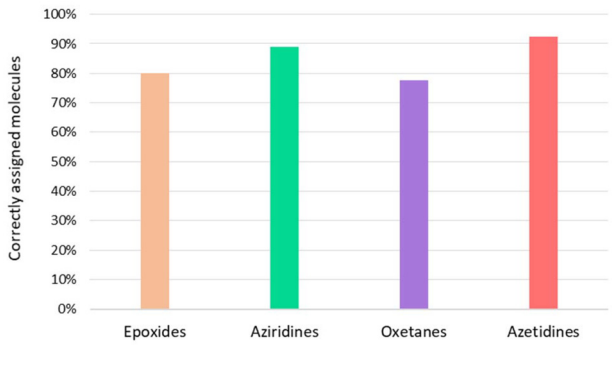

Fig. 5 Classification of MM-DP4+ performance based on constrained ring systems.

improvement in the assignment. This outcome was entirely unexpected, as the heterocyclic region plays a central role in stereochemical definition across all examples.

Given the similarity in the levels of theory used for NMR calculations, this discrepancy can only be attributed to geometric factors. This prompted a thorough investigation of the geometries provided by both levels of theory and their impact on NMR predictions. Initially, twelve molecules were selected (compounds **1**, **4**, **5**, **9**, **12**, **16**, **18**, **21**, **24**, **27**, **36**, and **39**), and the geometries obtained from MMFF and B3LYP/6-31G* were compared, focusing on bond lengths (d), angles (α), and dihedral angles (ϕ) of the strained heterocyclic moiety. The selection was made to ensure an appropriate distribution of heterocyclic groups (3 from each type), conformational flexibility, and MM-DP4+ performance (including both successfully assigned and misassigned cases). Regarding bond lengths, MMFF tends to overestimate the C–C bond distance in the heterocycle, particularly in more strained systems (averaged $\Delta d \sim 0.047$ Å for both epoxides and aziridines), representing relative deviations of 3.2% and 1.6%, respectively. On the other hand, C–X bond distances are slightly longer in B3LYP/6-31G*, with average deviations of less than 1%. This naturally affects the bond angles, predominantly in strained systems, with average $\Delta\alpha$ of 1.6° and 1.1° for epoxides and aziridines, respectively, and 0.3° and 0.9° for oxetanes and azetidines, respectively. We also observed important changes in the dihedral angles, with average $\Delta\phi$ of 3.6°, 2.7°, 3.8°, and 6.2° for epoxides, aziridines, oxetanes, and azetidines, respectively, with differences reaching up to 9.6° in some cases. Consequently, no consistent pattern was found in the dihedral angles that would allow for a generalized behavior across these systems.

Next, we investigated whether the observed discrepancies are primarily due to intrinsic limitations of MMFF, or if they also stem from potential shortcomings of B3LYP/6-31G*, a method that has faced significant criticism and is sometimes considered obsolete by some authors.[22] To this end, representative molecules were selected, and their geometries were re-optimized at different levels of theory, including M06-2X/6-31G*, B3LYP-D3/6-311++G(3df,2pd), M06-2X-D3/6-311++G (3df,2pd), and MP2/6-31+G**. Since the RMSD provides a global metric of fit, we define a parameter that allows us to determine the relative fit of geometries, focusing specifically on the heterocyclic system. Thus, we define the NMAD (Normalized Mean Absolute Difference), which is given by:

$$\mathrm{NMAD} = 100\frac{1}{n}\sum_{1}^{n}\left[\frac{x_i - x_{\mathrm{B3LYP}}}{x_{\mathrm{B3LYP}}}\right]$$

where x_i refers to a bond distance or angle measured from the geometry optimized at a given level, and x_{B3LYP} is the corresponding value extracted from the B3LYP/6-31G* optimized geometry, which is the standard for DP4+ calculations. The factor of 100 is included to bring the NMAD to a more practical and easily discussed scale.

The averaged NMAD value computed for the MMFF geometries of the selected twelve molecules was 2.00, significantly higher than those obtained for the other methods: 0.41 for B3LYP-D3/6-311++G(3df,2pd), 0.58 for M06-2X/6-31G*, 0.70 for M06-2X-D3/6-311++G(3df,2pd), and 0.80 for MP2/6-31G*. To grasp the significance of these values, they represent differences of up to 0.02 Å in distance and 0.9° in angle, suggesting a strong geometric agreement across the DFT methods. Hence, it can be concluded that the observed discrepancies are primarily due to an intrinsic limitation of the MMFF method in accurately describing the geometric characteristics of strained heterocycles. To strengthen this assertion, the X-ray structures of 4 epoxides were compared with the geometries optimized at the MMFF, B3LYP/6-31G*, B3LYP-D3/6-31G*, and M06-2X-D3/6-311++G(3df,2dp) levels. The corresponding average NMAD values (now relative to the X-ray geometries) are 2.97, 0.66, 0.67, and 1.5, respectively, further suggesting that the "modest" B3LYP/6-31G* provides an excellent geometric representation of these strained systems. These results indicate that the inclusion of dispersion has a negligible effect on the resulting geometries.

Force fields are mathematical representations of the interactions between atoms in a molecule, which allow modeling the structures and properties of different molecules. These interactions include covalent and ionic bonds, van der Waals forces, and electrostatic repulsions. In organic compounds containing constrained rings, force fields are crucial for the geometry optimization process because the ring constraint is an internal force that tends to distort the ideal geometry of the ring. Depending on the force field parametrization, this representation of the ring tension could affect the distances between atoms, bond angles, and dihedral angles. Additionally, the tension exerted by these three- and four-membered rings restricts rotation around the bonds within the ring, resulting in a limited number of stable conformations. Therefore, the key to the success of molecular simulation studies will be in the quality of the molecular mechanics force field employed for the calculations.[26]

To assess the influence of force fields on the geometry of molecules containing small heterocycles, a set of compounds with unique conformations (Fig. 6) was analyzed using various methods: MMFF, AMBER, MM2, MM3, OPLS4, OPLS2005 and B3LYP-D3/6-311++G(3df,2pd). The results were compared against B3LYP/6-31G* geometries.

Fig. 7 shows the overlaid global minima of epoxide **72** optimized at the 7 levels of theory discussed. All the structures exhibit close similarity, though minor differences are observed, mainly in bond lengths and torsion angles. The NMAD values are as follows: 3.73 for MMFF, 2.90 for OPLS2005, 4.31 for AMBER, 2.96 for MM2, 4.52 for MM3, and

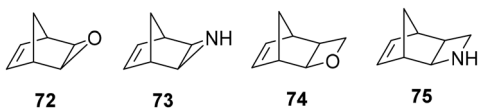

Fig. 6 Molecules chosen for Force Field analysis.

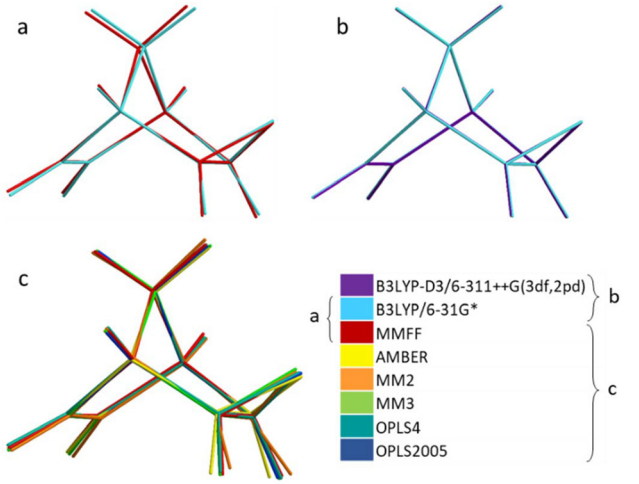

Fig. 7 Overlay of the global minima structures of epoxide 72 obtained at different levels of theory (using color coding for clarify). (a) Comparison between MMFF and B3LYP/6-31G* geometries. (b) Comparison between B3LYP/6-31G* and B3LYP-D3/6-311++G(3df,2dp) geometries. (c) Comparison between MMFF, AMBER, MM2, MM3, OPLS4, and OPLS2005 geometries.

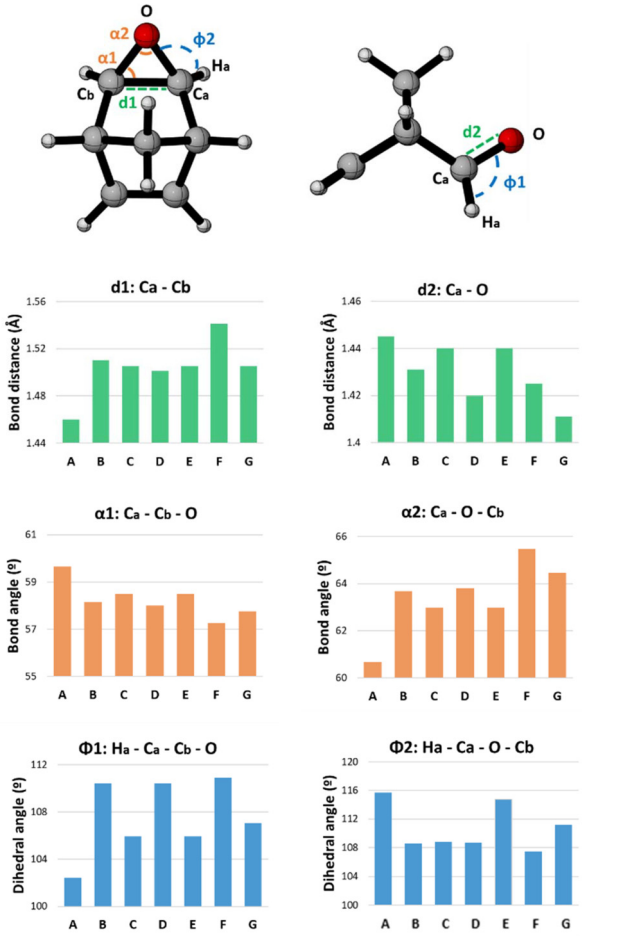

Fig. 8 Effect of the level of theory in the bond distances (d), angles (α), and dihedral angles (ϕ_1) of the oxirane moiety of compound **72**. A: B3LYP/6-31G*, B: MMFF, C: OPLS2005, D: AMBER, E: MM2, F: MM3, G: OPLS4.

3.17 for OPLS4. Among the geometries corresponding to NMR correlation methods (MMFF and B3LYP), some observable differences are present. However, the NMAD for the optimized structure at B3LYP-D3/6-311++G(3df,2pd) is 0.21, which aligns with Fig. 7b, where both DFT geometries are nearly identical, confirming the quality of the B3LYP/6-31G* method.

Consistent with the previous observations, MM methods tend to overestimate bond distances, ranging from 1.50 Å for OPLS4 to 1.54 Å for MM3, compared to DFT (1.46 Å). In contrast, MM typically yields slightly shorter C–O bond lengths (1.41–1.44 Å) relative to DFT (1.44 Å). Consequently, bond angles are somewhat distorted, as shown in Fig. 8. Additionally, significant variations were observed in the dihedral angles between the geometries. The largest discrepancy was found in the ϕ_1 dihedral angle, which reflects the inclination of the oxirane hydrogen with respect to the heterocycle plane. In the DFT geometries, the dihedral angle was 102.5°, while for the different force fields, it ranged from 105.9° to 110.9°. Notably, the MMFF value was 110.4°, showing a difference of 7.9° compared to DFT.

To investigate the impact of these changes in the chemical shift, the isotropic shielding constants were calculated at the PCM/mPW1PW91/6-31+G** level using each of the aforementioned geometries as a starting point. The results showed a strong dependence in the nuclei of oxirane system (C and H), with errors calculated using MM-derived geometries being larger (1.7–4.2 ppm for ^{13}C and 0.06–0.44 ppm for ^{1}H) compared to those obtained for DFT-optimized structures (1.2 ppm and 0.04 ppm, respectively). To validate this geometric dependence, the structure was recalculated at the B3LYP-D3/6-311++G(3df,2pd) level, showing minimal differences from those obtained at the B3LYP/6-31G* level (see ESI†). Accordingly, the errors obtained were very similar (0.2 and 0.03 ppm, for ^{13}C and ^{1}H respectively). Interestingly, MMFF is one of the MM methods that produces the largest errors (along with AMBER & MM3), which is consistent with its greater deviation from DFT geometries. This explains why MM-DP4+ tends to perform poorly in these types of strained systems. These values show that, in general, all force fields generate a considerable deviation of the geometry compared to the DFT, which is partly responsible for the errors in the NMR estimates.

In a previous study conducted by the group,[27] it was investigated how little differences in C–halogen (Cl and Br) bond distances between MMFF and DFT methods led big changes in the assignment of the DP4+ and MM-DP4+ methods. This motivated us to carry out a similar study on compounds with heterocycles. Starting with geometries optimized of molecules **72** and **73** at the B3LYP/6-31G* level, a scan of C–C, C–O, C–N distances, as well as dihedral angles, was performed, and their influence on the calculated chemical shift was analyzed. The angles were not analyzed, as modifying the dis-

tances automatically alters the angles. For this, the optimized geometry at the B3LYP/6-31G* level was used as the starting point, and the parameters under study were modified while keeping the rest of the atoms frozen. Each resulting geometry was calculated using PCM/mPW1PW91/6-31+G** GIAO NMR calculations. The results for the epoxide **72** are shown in Fig. 9.

For ^{13}C calculated shifts, the sensitivity is 34.3 ppm Å^{-1} for the C–C distance (d_1) and 121.3 ppm Å^{-1} for the C–O distance (d_2). In both cases, the series show a positive slope, suggesting that the chemical shift increases with distance. However, the C–O distance is much more conserved than the C–C distance, making the latter the primary source of difference. On the other hand, the dependence on dihedral angles is smaller (0.04 ppm Å^{-1} for dihedral ϕ_1 and 0.28 ppm Å^{-1} for dihedral ϕ_2). Regarding protons, the sensitivities are 0.54 ppm Å^{-1}, 5.49 ppm Å^{-1}, 0 ppm Å^{-1}, and 0.05 ppm Å^{-1} for d_1, d_2, ϕ_1, and ϕ_2, respectively. Based on these trends, it was possible to segment the chemical shift change between MMFF and B3LYP/6-31G* using the four geometric parameters calculated for both structures. The $\Delta\delta^{13}$C values obtained were 1.73 ppm, 1.70 ppm, 0.32 ppm, and 1.98 ppm for Δd_1, Δd_2, $\Delta\phi_1$, and $\Delta\phi_2$, respectively, suggesting that ϕ_2, along with both bond distances, plays a significant role. For the ^{1}H data, the $\Delta\delta$ values were 0.03 ppm for Δd_1, 0.08 ppm for Δd_2, <0.01 ppm for $\Delta\phi_1$, and 0.35 ppm for $\Delta\phi_2$, indicating that, in this case, ϕ_2 is the most influential factor.

The same analysis was conducted on an aziridine **73**. Upon examining the bond lengths and angles within the heterocyclic ring, similar parameters to the previous molecule were observed. Fig. 10 shows the overlay of the geometries optimized at different levels of theory.

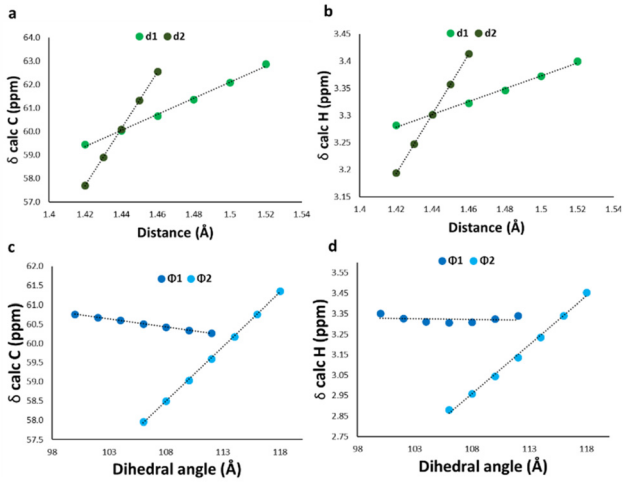

Fig. 9 Influence of changes in geometric elements on the calculated ^{13}C and ^{1}H chemical shifts: Sensitivity analysis. (a) shows the dependence of C–C (d_1) and C–O (d_2) distance in ^{13}C shift, (b) indicates the d_1 and d_2 impacts in ^{1}H shift, (c) illustrates the variation of ^{13}C shift with dihedrals ϕ_1 and ϕ_2, and (d) displays the changes of ^{1}H shift with ϕ_1 and ϕ_2.

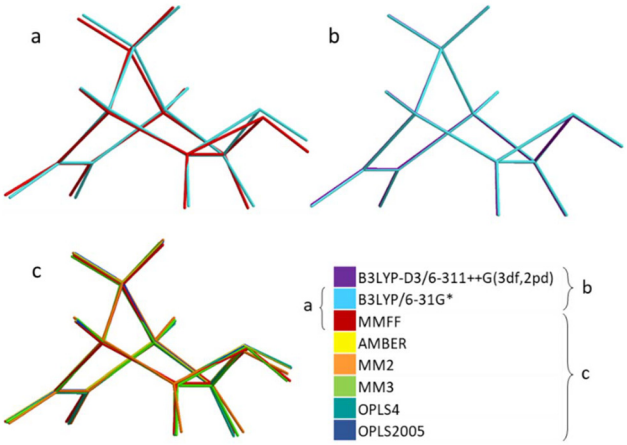

Fig. 10 Overlay of the global minima structures of aziridine 73 obtained at different levels of theory (using color coding for clarify). (a) Comparison between MMFF and B3LYP/6-31G* geometries. (b) Comparison between B3LYP/6-31G* and B3LYP-D3/6-311++G(3df,2dp) geometries. (c) Comparison between MMFF, AMBER, MM2, MM3, OPLS4, and OPLS2005 geometries.

When analyzing the geometric values, certain trends similar to those previously discussed are observed, with shorter C–C distances according to DFT, and longer C–N distances. However, unlike the previous compound, the differences in dihedral values involving the hydrogen atom with the three-membered ring across MMFF and DFT geometries are not particularly significant (except for MM2 and MM3). The NMAD value is 3.01 for MMFF, moderately lower than that for the epoxide **72** (3.73). The smaller geometrical differences between MMFF and B3LYP/6-31G* is in agreement with the lower percentage of aziridines misassigned by MM-DP4+ (18% for epoxides and 10% for aziridines). It is also worth noting the heterogeneity in results across different force fields, particularly MM2 (NMAD of 7.97), which predicts highly distorted heterocyclic systems, showing abnormally long and short C–C and C–N bond lengths, respectively, as well as significantly deviated dihedral angles. Similar to the previous example, the geometries optimized at B3LYP/6-31G* and B3LYP-D3/6-311++G(3df,2pd) are nearly identical.

A scan of C–C and C–N distances, as well as dihedral angles, was performed with aziridine **73**, to study their influence on the calculated chemical shift. The sensitivity of each geometric element to the calculated ^{13}C and ^{1}H chemical shifts was lower than that observed for the epoxide, indicating a reduced dependence of the NMR calculation on geometry changes in nitrogen-containing compounds. For example, the rate of change in C–C was 27.1 ppm Å^{-1} (*vs.* 34.3 ppm Å^{-1} in epoxide), and the sensibility for C–N distance was 97.1 ppm Å^{-1} (*vs.* 121.3 ppm Å^{-1} in C–O). The dependence on dihedral angles was similar to that observed for compound **72** (see ESI for further details on this topic†).

The analysis with oxetane **74**, in contrast, yielded less divergent results. Relative deviations in C–C and C–O bond lengths between the force fields and the B3LYP/6-31G* geometry

ranged from 0.19% to 1.55%, notably lower than those observed in the previous compounds. Dihedral angle values showed only slight variation across force fields, as can be seen in Fig. 11. The NMAD value for MMFF was 1.71, significantly lower than for epoxide **72** (3.73) and aziridine **73** (3.01). Once again, B3LYP/6-31G* and B3LYP-D3/6-311++G(3df,2pd) geometries were found to be practically identical.

The analysis of azetidine **75** under different optimizations follows the trend observed in previous cases: no significant variations are found between the DFT and molecular mechanics geometries (Fig. 12). The NMAD value for MMFF was 1.08, slightly lower than for oxetane **74**. Notably, unlike prior cases, the dihedral angle values for azetidine do not exhibit marked differences, with both DFT geometries falling within an intermediate range among the various force fields. This consistency aligns with the improved assignment accuracy of azetidines by MM-DP4+.

Finally, we decided to investigate xTB, a semiempirical method recently designed to bridge the gap between quantum mechanical (QM) and molecular mechanics (MM) methods.[28] By re-optimizing the geometries of the 12 molecules previously discussed (compounds **1**, **4**, **5**, **9**, **12**, **16**, **18**, **21**, **24**, **27**, **36**, and **39**), along with the norbornene derivatives **72–75**, the results show that xTB provides geometries that lie between those obtained with DFT and MM, with NMAD values ranging from 0.62 to 2.21. As shown in Fig. 13, while the agreement is remarkably accurate for rigid systems (*e.g.*, **5**, **24**, **72**, and **75**), slightly larger discrepancies are observed for more flexible systems. The improved geometric accuracy (relative to DFT) positively impacts NMR predictions at the PCM/mPW1PW91/6-31+G** level, with CMAE values of 1.9 ppm (MMFF), 1.4 ppm (xTB), and 1.3 ppm (B3LYP/6-31G*) for carbon data, and 0.19 ppm (MMFF), 0.17 ppm (xTB), and 0.14 ppm (B3LYP/6-31G*) for proton data. Although developing a new probability model based on xTB geometries is beyond the scope of this study—since it would require an extensive re-parameterization—we expect to address this in the near future, given the favorable cost-benefit ratio offered by this method.

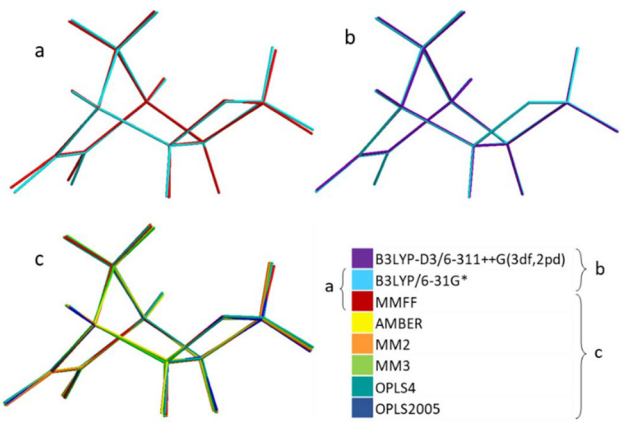

Fig. 11 Overlay of the global minima structures of oxetane 74 obtained at different levels of theory (using color coding for clarify). (a) Comparison between MMFF and B3LYP/6-31G* geometries. (b) Comparison between B3LYP/6-31G* and B3LYP-D3/6-311++G(3df,2dp) geometries. (c) Comparison between MMFF, AMBER, MM2, MM3, OPLS4, and OPLS2005 geometries.

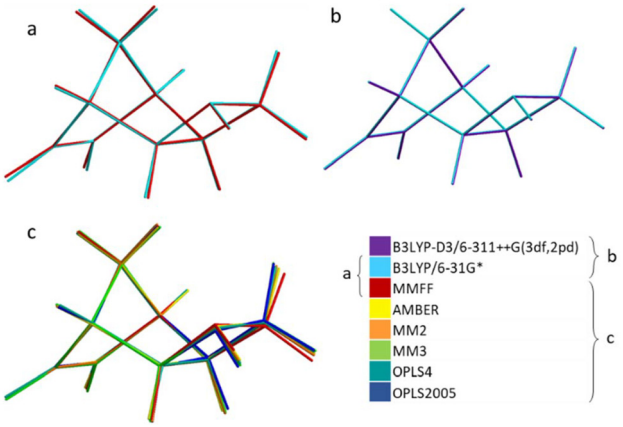

Fig. 12 Overlay of the global minima structures of azetidine 75 obtained at different levels of theory (using color coding for clarify). (a) Comparison between MMFF and B3LYP/6-31G* geometries. (b) Comparison between B3LYP/6-31G* and B3LYP-D3/6-311++G(3df,2dp) geometries. (c) Comparison between MMFF, AMBER, MM2, MM3, OPLS4, and OPLS2005 geometries.

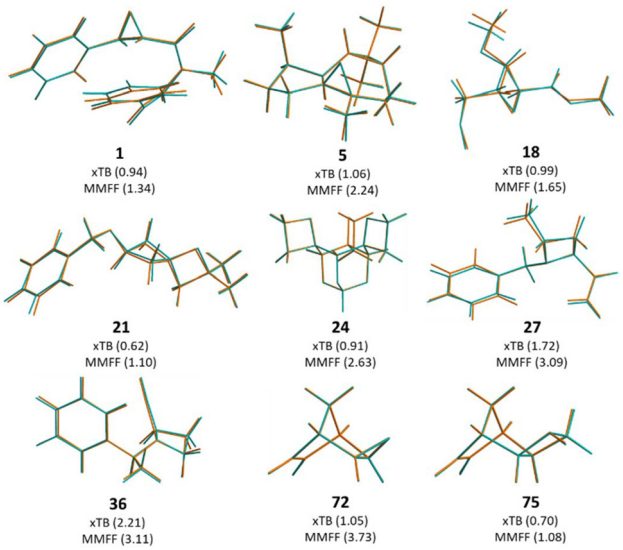

Fig. 13 Overlay of the global minima structures of nine selected molecules optimized at B3LYP/6-31G* (light blue) and xTB (orange). The NMAD computed for xTB and MMFF geometries are given in parenthesis.

Conclusions

To sum up, we thoroughly explored the nature of the strained heterocycle in the DP4+ and MM-DP4+ architectures. Our results showed that DP4+ afforded an excellent classification

performance (100%), while MM-DP4+ achieved an 83% score. Although this outcome is very favorable, especially considering the remarkable speed of MM-DP4+, we aimed to assess how geometric quality influences assignment accuracy. Upon examining the geometric parameters of molecules optimized at the MMFF and B3LYP/6-31G* levels (corresponding to MM-DP4+ and DP4+, respectively), substantial deviations in bond distances and dihedral angles were observed in the examples analyzed, mainly in the more strained heterocycles. In comparison with high-quality DFT levels and other force fields, we conclude that MMFF has limitations in providing accurate geometries, further reaffirming the simple and affordable B3LYP/6-31G* as a reliable choice for this purpose. For single-conformation molecules, the calculated chemical shifts of ring nuclei were found to be highly sensitive to bond distances, particularly to C–O, followed by C–N distances. This may account for the lower performance of oxygen-containing examples analyzed with MM-DP4+, compared to nitrogen-containing ones. Based on this study, we generally recommend using MM-DP4+ for a quick and reliable initial screening. However, if the system under evaluation contains a strained heterocycle, we suggest validating the results with DP4+ for greater accuracy.

Computational details

To achieve our goals, we selected 39 molecules containing constrained ring from the literature, featuring different complex architectures **1–39**.[4,29,30–37,38–40] Each compound and their respective diastereoisomers were modelled, and a systematic conformational samplings at the MMFF force field implemented in Spartan[41] were done. An energy cutoff window of 5 kcal mol^{-1} was selected to ensure good conformational diversity. For MM-DP4+,[17] the MMFF geometries were used as inputs for GIAO NMR calculations at the SMD/ωB97XD/6-31+G** level using chloroform as solvent. For DP4+ calculations,[14] the MMFF geometries were re-optimized at the B3LYP/6-31G* level of theory, prior to the GIAO NMR calculations at the PCM/mPW1PW91/6-31+G** level of theory with chloroform as solvent using Gaussian16.[42] The levels chosen (SMD/ωB97XD/6-31+G**//MMFF and PCM/mPW1PW91/6-31+G**//B3LYP/6-31G*) are the recommended ones for MM-DP4+ and DP4+ calculations, respectively. In all cases, the unscaled chemical shifts (δ_u) were computed using TMS as reference standard according to $\delta_u = \sigma_0 - \sigma_x$, where σ_x is the Boltzmann averaged isotropic shielding constant (over all significantly populated conformations) and σ_0 is the isotropic shielding constant of TMS computed at the same level of theory. The Boltzmann averaging was done at 298 K using the relative SCF energies obtained at the corresponding level of theory, (SMD/ωB97XD/6-31+G**//MMFF, and PCM/mPW1PW91/6-31+G**//B3LYP/6-31G* according the employed method). The scaled chemical shifts (δ_s) were computed as $\delta_s = (\delta_u - b)/m$, where m and b are the slope and intercept, respectively, resulting from a linear regression calculation on a plot of δ_u against δ_{exp}. The whole procedure was automated using DP4+ App application, available at **https://github.com/Sarotti-Lab/DP4plus-App**. The probabilities for MM-DP4+ and DP4+ are reported with one decimal place, following the standard convention.

For molecules **1**, **5**, **12**, **72**, **73**, **74** and **75** MMFF geometries of the more stable conformer were re-optimized at B3LYP-GD3/6-311(3df,2pd), M062X/6-31G*, M062X-GD3/6-311(3df,2pd), and MP2/6-31G* levels.

In order to achieve the force field analysis, molecules **72**, **73**, **74** and **75** were modelled on Maestro, and geometry optimizations were carried out with MacroModel[43] at MMFF, OPLS2005, AMBER, MM2, MM3, OPLS4 force fields. The parameters of all optimization were: none solvent, cutoff normal, minimization method: PRCG with maximum iterarions of 2500, converge on gradient with a threshold of 0.05. The xTB calculations were done using AQME, a free and open-source Python package for cheminformatics and quantum chemistry.[44]

Author contributions

E. R. L.: methodology, validation, data analysis, writing. M. A.: methodology, validation. M. M. Z.: conceptualization, writing, data analysis. A. M. S.: conceptualization, writing, data analysis.

Data availability

The data supporting this article are provided in two files. ESI I contains the experimental data, NMR calculated tensors, DP4+ and MM-DP4+ results, geometrical parameters, NMAD values, and additional complementary graphs. ESI II includes the coordinate matrices for each compound, isomer, and conformation, along with their SCF energies.†

The data supporting this article have been included as part of the ESI.†

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This research was funded by ANPCyT (PICT-2019-4052 and PICT-2021-CAT-II-00033), and CONICET (PIP 11220200102205CO and PIBAA 28720210101005CO). E. R. L and M. A. thank CONICET for the award of a fellowship.

References

1 B. H. Rotstein, S. Zaretsky, V. Rai and A. K. Yudin, Small heterocycles in multicomponent reactions, *Chem. Rev.*, 2014, **114**(16), 8323–8359.

2 U. Azzena, M. Carraro and L. Pisano, Addressing Stereochemistry of Heterocyclic Compounds by DFT NMR Calculations, *Chem. Heterocycl. Compd.*, 2018, **54**(4), 380–388.
3 L. Pisano, L. Degennaro, M. Carraro, U. Azzena, F. Fanelli, P. Mastrorilli, *et al.*, Computational NMR as Useful Tool for Predicting Structure and Stereochemistry of Four-Membered Sulfur Heterocycles, *Eur. J. Org. Chem.*, 2016, **2016**(19), 3252–3258.
4 A. G. Kutateladze, D. M. Kuznetsov, A. A. Beloglazkina and T. Holt, Addressing the Challenges of Structure Elucidation in Natural Products Possessing the Oxirane Moiety, *J. Org. Chem.*, 2018, **83**(15), 8341–8352.
5 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**(4), 1873–1879.
6 B. Ritzen, M. C. M. Van Oers, F. L. Van Delft and F. P. J. T. Rutjes, Enantioselective chemoenzymatic synthesis of trans-aziridines, *J. Org. Chem.*, 2009, **74**(19), 7548–7551.
7 A. Brandi, S. Cicchi and F. M. Cordero, Novel syntheses of azetidines and azetidinones, *Chem. Rev.*, 2008, **108**(9), 3988–4035.
8 A. G. Kutateladze, T. Holt and D. S. Reddy, Natural Products Containing the Oxetane and Related Moieties Present Additional Challenges for Structure Elucidation: A DU8+ Computational Case Study, *J. Org. Chem.*, 2019, **84**(12), 7575–7586.
9 S. G. Smith and J. M. Goodman, Assigning the stereochemistry of pairs of diastereoisomers using GIAO NMR shift calculation, *J. Org. Chem.*, 2009, **74**(12), 4597–4607.
10 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**(37), 12946–12959.
11 G. Lauro and G. Bifulco, Elucidating the Relative and Absolute Configuration of Organic Compounds by Quantum Mechanical Approaches, *Eur. J. Org. Chem.*, 2020, 1–14.
12 F. L. P. Costa, A. C. F. de Albuquerque, R. G. Fiorot, L. M. Lião, L. H. Martorano, G. V. S. Mota, *et al.*, Structural characterisation of natural products by means of quantum chemical calculations of NMR parameters: new insights, *Org. Chem. Front.*, 2021, **8**(9), 2019–2058.
13 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**(9), 1922–1932.
14 N. Grimblat, M. M. M. M. Zanardi and A. M. 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**(24), 12526–12534.
15 N. Grimblat, J. A. Gavín, A. Hernández Daranas and A. M. Sarotti, Combining the Power of J Coupling and DP4 Analysis on Stereochemical Assignments: The J-DP4 Methods, *Org. Lett.*, 2019, **21**(11), 4003–4007.
16 Y. H. Tsai, M. Amichetti, M. M. Zanardi, R. Grimson, A. H. Daranas and A. M. Sarotti, ML-J-DP4: An Integrated Quantum Mechanics-Machine Learning Approach for Ultrafast NMR Structural Elucidation, *Org. Lett.*, 2022, **24**(41), 7487–7491.
17 B. A. Franco, E. R. Luciano, A. M. Sarotti and M. M. Zanardi, DP4+ App: Finding the Best Balance between Computational Cost and Predictive Capacity in the Structure Elucidation Process by DP4+. Factors Analysis and Automation, *J. Nat. Prod.*, 2023, **83**(11), 2360–2367.
18 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**(9), 5035–5043.
19 A. Navarro-Vázquez, R. R. Gil and K. Blinov, Computer-Assisted 3D Structure Elucidation (CASE-3D) of Natural Products Combining Isotropic and Anisotropic NMR Parameters, *J. Nat. Prod.*, 2018, **81**(1), 203–210.
20 I. M. Novitskiy and A. G. Kutateladze, DU8+ Computations Reveal a Common Challenge in the Structure Assignment of Natural Products Containing a Carboxylic Anhydride Moiety, *J. Org. Chem.*, 2021, **86**(23), 17511–17515.
21 M. O. Marcarino, S. Cicetti, M. M. Zanardi and A. M. Sarotti, A critical review on the use of DP4+ in the structural elucidation of natural products: the good, the bad and the ugly. A practical guide, *Nat. Prod. Rep.*, 2022, **39**(1), 58–76.
22 M. Bursch, J. M. Mewes, A. Hansen and S. Grimme, Best-Practice DFT Protocols for Basic Molecular Computational Chemistry, *Angew. Chem., Int. Ed.*, 2022, **61**, e202205735.
23 T. Lewis-Atwell, P. A. Townsend and M. N. Grayson, Comparisons of different force fields in conformational analysis and searching of organic molecules: A review, *Tetrahedron*, 2021, **79**, 131865.
24 C. Cuadrado, A. H. Daranas and A. M. Sarotti, May the force (Field) be with you: on the importance of conformational searches in the prediction of NMR chemical shifts, *Mar. Drugs*, 2022, **20**(11), 699.
25 M. M. M. M. Zanardi, F. A. Biglione, M. A. M. A. Sortino and A. M. A. M. Sarotti, General Quantum-Based NMR Method for the Assignment of Absolute Configuration by Single or Double Derivatization: Scope and Limitations, *J. Org. Chem.*, 2018, **83**(19), 11839–11849.
26 R. S. Paton and J. M. Goodman, Hydrogen bonding and π-stacking: how reliable are force fields? A critical evaluation of force field descriptions of nonbonded interactions, *J. Chem. Inf. Model.*, 2009, **49**(4), 944–955.
27 L. Passaglia, M. M. Zanardi and A. M. Sarotti, Study of heavy atom influence on poly-halogenated compounds using DP4/MM-DP4+/DP4+: insights and trends, *Org. Biomol. Chem.*, 2024, **22**(12), 2435–2442.
28 S. Grimme, C. Bannwarth and P. Shushkov, A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent

Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1–86), *J. Chem. Theory Comput.*, 2017, **13**(5), 1989–2009.

29 J. Orjala and W. H. Gerwick, Barbamide, a chlorinated metabolite with molluscicidal activity from the Caribbean cyanobacterium Lyngbya majuscula, *J. Nat. Prod.*, 1996, **59**(4), 427–430.

30 L. A. Paquette, C. S. Ra, J. C. Gallucci, H. J. Kang, N. Ohmori, M. P. Arrington, W. David and J. S. Brodbert, Constraining of small-ring cyclic ether triads by stereodefined spiroannulation to an inositol orthoformate platform. Solution- and gas-phase alkali metal binding affinities for three- to five-membered ring structural combinations, *J. Org. Chem.*, 2001, **66**(25), 8629–8639.

31 F. B. Dyer, C. M. Park, R. Joseph and P. Garner, Aziridine-mediated ligation and site-specific modification of unprotected peptides, *J. Am. Chem. Soc.*, 2011, **133**(50), 20033–20035.

32 J. M. Srinivasan, P. A. Mathew, A. L. Williams, J. C. Huffman and J. N. Johnston, Stereoselective synthesis of complex polycyclic aziridines: Use of the Brønsted acid-catalyzed aza-Darzens reaction to prepare an orthogonally protected mitomycin C intermediate with maximal convergency, *Chem. Commun.*, 2011, **47**(13), 3975–3977.

33 R. Hili and A. K. Yudin, Readily available unprotected amino aldehydes, *J. Am. Chem. Soc.*, 2006, **128**(46), 14772–14773.

34 S. D. Lucas, A. P. Rauter and H. P. Wessel, Synthesis of 3-methoxyoxetane δ-amino acids with D-lyxo, D-ribo, and D-arabino configurations, *J. Carbohydr. Chem.*, 2008, **27**(3), 172–187.

35 C. A. Malapit, S. M. Chitale, M. S. Thakur, R. Taboada and A. R. Howell, Pt-Catalyzed Rearrangement of Oxaspirohexanes to 3-Methylenetetrahydrofurans: Scope and Mechanism, *J. Org. Chem.*, 2015, **80**(10), 5196–5209.

36 A. Osato, T. Fujihara and H. Shigehisa, Constructing Four-Membered Heterocycles by Cycloisomerization, *ACS Catal.*, 2023, **13**(7), 4101–4110.

37 D. H. Leng, D. X. Wang, J. Pan, Z. T. Huang and M. X. Wang, Highly efficient and enantioselective biotransformations of racemic azetidine-2-carbonitriles and their synthetic applications, *J. Org. Chem.*, 2009, **74**(16), 6077–6082.

38 N. E. Behnke, K. Lovato, M. Yousufuddin and L. Kürti, Titanium–Mediated Synthesis of Spirocyclic NH –Azetidines from Oxime Ethers, *Angew. Chem.*, 2019, **131**(40), 14357–14361.

39 K. A. Rykaczewski, M. R. Becker, M. J. Anantpur, R. C. Sausa, E. C. Johnson, J. A. Orlicki, *et al.,* Photochemical Strategies Enable the Synthesis of Tunable Azetidine-Based Energetic Materials, *J. Am. Chem. Soc.*, 2022, **144**(41), 19089–19096.

40 B. Alcaide, P. Almendros, C. Aragoncillo, R. Callejo, M. P. Ruiz and M. R. Torres, Rhodium-catalyzed synthesis of 3-hydroxy-β-lactams via oxonium ylide generation: Three-component reaction between azetidine-2,3-diones, ethyl diazoacetate, and alcohols, *J. Org. Chem.*, 2009, **74**(21), 8421–8424.

41 Y. Guan, S. V. Shree Sowndarya, L. C. Gallegos, P. C. St John and R. S. Paton, Real-time prediction of 1H and 13C chemical shifts with DFT accuracy using a 3D graph neural network, *Chem. Sci.*, 2021, **12**(36), 12012–12026.

42 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, *et al*., *Gaussian 16, Revision C.01*. 2016, Gaussian, Inc., Wallingford CT.

43 F. Mohamadi, N. G. J. Richards, W. C. Guida, R. Liskamp, M. Lipton, C. Caufield, *et al.,* Macromodel—an integrated software system for modeling organic, bioorganic molecules using molecular mechanics, *J. Comput. Chem.*, 1990, **11**(4), 440–467.

44 J. V. Alegre-Requena, S. V. S. Sowndarya, R. Pérez-Soto, T. M. Alturaifi and R. S. Paton, AQME: Automated quantum mechanical environments for researchers and educators, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2023, **13**(5), e1663.