Tetrahedron Chem 14 (2025) 100125

Contents lists available at ScienceDirect

Tetrahedron Chem

journal homepage: www.journals.elsevier.com/tetrahedron-chem

Efficient workflow for structural elucidation applied to solidagenone derivative: Combining the speed of ML-*J*-DP4 screening with the precision of DP4+

Iván Cortés [a,1], María Luz Tibaldi-Bollati [b,1], Viviana E. Nicotra [b], Manuela E. García [b,*], Ariel M. Sarotti [a,**]

[a] *Instituto de Química Rosario (CONICET), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario, Argentina*
[b] *Instituto Multidisciplinario de Biología Vegetal (CONICET), Facultad de Ciencias Químicas, Universidad Nacional de Córdoba, Haya de la Torre y Medina Allende, Ciudad Universitaria, CP X5000HUA, Córdoba, Argentina*

ARTICLE INFO

Keywords:
Solidagenone derivatives
DFT-Calculations
NMR shifts

ABSTRACT

The derivatization of solidagenone through ring expansion, using the "Complexity to Diversity" synthetic strategy to obtain polyketones, yielded two products upon reaction of an intermediate with *in situ*-generated ruthenium tetroxide. As might happen, the spectroscopic and spectrometric data were inconclusive for the structural determination of one of these products. To address the issue, a ML-*J*-DP4//DP4+ workflow was studied, which allowed simultaneous determination of connectivity and relative configuration with high accuracy, while saving computational resources. Furthermore, during the calculation of NMR data at the DP4+ level, abnormally large errors were observed for the beta carbon of a polyhydroxylated conjugated enone. A literature analysis revealed that this behavior could be generalized, and after careful examination, it was determined to be the result of small geometric distortions caused by an overestimated network of intramolecular hydrogen bonds. To mitigate this common source of discrepancy, a heuristic solution was introduced as a proof of concept by *in silico* blocking of hydroxyl groups.

1. Introduction

Natural products (NPs) exhibit a unique and extensive chemical diversity [1,2]. As a result of evolutionary selection processes, secondary metabolites possess significant biological activities, leading to their therapeutic applications and serving as models for structural modifications aimed at developing new therapeutic agents [3,4].

However, obtaining NPs in sufficient quantities can be challenging, often involving lengthy isolation and purification processes or multistep total synthesis. This limited access stimulated the development of novel methods for efficiently generating analogous compounds.

To meet the demands for molecular diversity, inspired by the variations produced in nature from a single precursor, one of the most promising approaches is Complexity to Diversity (CtD) [5]. This methodology involves systematic distortions of ring systems and structural rearrangements (including cleavage, expansion, fusion, rearrangement, aromatization, and ring combination) to generate collections of unique compounds with high diversity and complexity in their skeletons in just a few steps [6]. In this regard, the implementation of a rapid and reliable structural elucidation framework streamlines the overall process, particularly considering that unexpected structures may emerge during the workflow. In this context, NMR calculations represent a simple, useful, and convenient strategy to facilitate the assignment of complex molecules. These tools have been consolidated over time to the point that, today, they are routinely used in structural elucidation, complementing experimental NMR spectroscopy in an integrated manner. This is particularly evident in the field of NPs, where constitutional, conformational, and configurational complexities make these tools essential for unequivocal structural determination. To date, there is a wide variety of methods available, significantly differing in computational demand and accuracy [7–14]. Generally, methods based on DFT-level optimizations (such as DP4+ [8]) tend to be considerably

* Corresponding author.
** Corresponding author.
E-mail addresses: manuelagarcia@unc.edu.ar (M.E. García), sarotti@iquir-conicet.gov.ar (A.M. Sarotti).
[1] These authors have contributed equally to this work.

https://doi.org/10.1016/j.tchem.2025.100125
Received 27 December 2024; Received in revised form 25 February 2025; Accepted 5 March 2025
Available online 6 March 2025

more accurate but require high computational resources. In contrast, molecular mechanics-based methods are significantly faster but sacrifice some precision. In this context, our recent ML-*J*-DP4 [14]stands out as a machine learning-accelerated version of *J*-DP4 [7], enabling reasonably good predictions in seconds or a few minutes.

Our research group has extensively focused on the isolation and characterization of various natural compounds, primarily terpenoids [15]. In particular, this work highlights the diterpene solidagenone (**1**) [16], which is isolated as the major metabolite from *Solidago chilensis* Meyen (Fam. Asteraceae) [17], a plant species widely used in Argentine and regional folk medicine. This diterpene (**1**, Scheme 1) is particularly intriguing due to the functionality of its structure, which includes an α,β-unsaturated carbonyl group, a furan, and a hydroxyl group; these characteristics make it an excellent model for exploring innovative transformations [18].

In this work, we wish to highlight the synergistic utility of the DP4+ and ML-*J*-DP4 methodologies, emphasizing the importance of combining multiple tools for making decisions regarding the structure of complex molecules, particularly when experimental spectroscopic data are consistent with more than one chemical structure. Furthermore, other aspects of the implementation of NMR calculations on DFT-optimized structures were explored, emphasizing the importance of a thorough analysis in computational studies.

2. Results and discussion

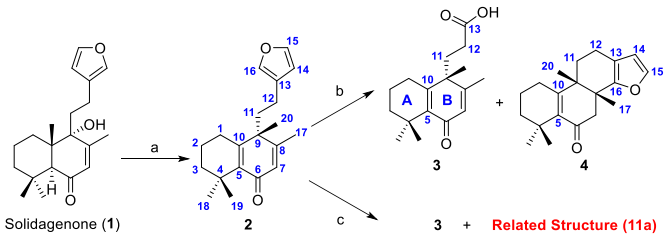
To achieve ring distortions aimed at ring expansion, intermediate **2** was prepared from solidagenone (natural precursor **1**) under acidic conditions via a Wagner-Meerwein rearrangement, as previously documented (Scheme 1) [15,19,20]. This synthetic proposal was inspired by the work of Ignatenko and collaborators [21], who remodeled bryonolic acid (a triterpenoid structurally related to lanosterol) to obtain multiple molecular skeletons through oxidative cleavages that produce transannular polyketones and subsequent aldol additions. The use of catalytic $NaIO_4$ and $KMnO_4$ for the oxidative cleavage of olefins led to the formation of **3**, an oxidative decarboxylation product of furan (not previously reported), as well as a ring-fused derivative described by other authors (Scheme 1, **pathway b**) [19,20]. The ^1H and ^{13}C NMR data of compound **3** closely resembled those of compound **2**, differing primarily in the side chain at C-9. Notable differences included the disappearance of signals corresponding to the furan ring, the downfield shift of the C-12 resonance from δ_C 19.9 in **2** to δ_C 30.8 in **3**, and the presence of a new carboxy signal at δ_C 177.0. This carbon displayed a cross-peak in the HMBC spectrum with H_2-12 (δ_H 1.81) and H-11a (δ_H 2.01), confirming the proposed structure of compound **3**. Furthermore, compound **3** contains fewer than 20 carbons in its structure, consistent with a possible oxidative decarboxylation (see the full spectral assignment in the Supporting Information). When employing *in situ* ruthenium tetroxide, compound **3** was once again observed, along with the formation of a by-product (Scheme 1, pathway c).

Although the desired polyketone was unfortunately not obtained, it was challenging to determine the structure of the new derivative.

In this regard, the analysis of one- and two-dimensional NMR spectra of the related compound (later identified as **11a**) reveals significant similarities with compound **3**, particularly in the chemical shifts corresponding to ring A and the side chain at C-9. For ring B, the disappearance of the olefin signals at C7–C8 was noted, replaced by two carbon signals: a quaternary carbon at δ_C 77.7 and a methine at δ_C 76.1/δ_H 4.06 s. The positions of the hydroxyl groups at C7–C8 were suggested based on the analysis of correlations between H-7 and C-6 at δ_C 197.7, as well as C-17 at δ_C 21.2, and between H_3-17 at δ_H 1.28 and carbons C-7, C-8, and C-9 (δ_C-9 46.9) in the HMBC spectrum. The presence of the hydroxyl groups at these positions and a carboxy group at C-13 suggests the potential for C-8/C-13 lactonization to form a new heterocycle. However, no conclusive correlations were observed in either the HMBC or NOESY spectra to confirm or rule out this possibility. Regarding the HRMS spectrum, a peak at $m/z = 333.1694$ is observed, corresponding to the $[M+Na]^+$ ion, which matches the molecular formula $C_{17}H_{26}O_5$. However, considering the base peak at $m/z = 293.1776$ and its mass difference, it could be suggested that the compound underwent dehydration under the experimental conditions of the spectrometer, leading to the formation of the $[M + H–H_2O]^+$ ion, or that the compound has an real molecular formula of $C_{17}H_{24}O_4$ and that peak corresponds to $[M+H]^+$. While the observation of a dehydration peak is expected, there are cases where water adduct artifacts have complicated the elucidation process, such as in the recent case of dysiherbol A [22]. Spectroscopic evidence and chemical knowledge suggested that hydroxylation occurred at C7–C8. However, it is important to note that, in the context of CtD, the formation of rearrangement products, condensates, or simple artifacts is to be expected [6]. This is where one of the main strengths of the proposed workflow becomes evident: its ability to rapidly evaluate a wide range of potential structures, efficiently ruling out unlikely candidates within minutes. Therefore, other options were considered based on chemical shifts and fitting to correlations found in HMBC spectra (compounds **5**–**12,** Fig. 1).

It is crucial to emphasize that, in rigid systems, configurational changes can significantly affect chemical shifts. Put differently, a correct connectivity with an incorrect configuration can lead to errors as severe as an incorrect connectivity [22]. Therefore, it is essential to approach elucidation in an integrated manner, considering both connectivity and configuration. As the configuration at C-9 is known due to the known precursor **2**, and considering possible skeletal rearrangements at C-8 and C-10, each structure depicted in Fig. 1 can exist as four possible diastereoisomers. Therefore, the total number of potential candidates rises to 32, a fairly large number given the conformational flexibility of certain members (the total number of unique conformations found at the MMFF (Merck molecular force field) level was 740 for the whole test set). Therefore, a workflow, shown in Fig. 2, was implemented to identify the correct structure while minimizing computational cost.

For this purpose, an initial screening was performed using the ML-*J*-DP4 method, which has proven to be highly reliable in terms of prediction accuracy and computational efficiency. This approach involves MMFF-level optimization, NMR calculations at the RHF/STO-3G level enhanced by ML, and statistical methods to identify the most likely candidate [14]. To ensure unbiased analysis based on the experimental assignment of the resonances of the candidate molecules, a data



Scheme 1. *Reagents and Conditions*: **a)** HCl, MeOH, reflux, 1 h, (**2** 39 %). **b)** $NaIO_4/KMnO_4$, *i*-PrOH, Na_2CO_3, reflux, 4h (**3** 18 %, **4** 14 %). **c)** $NaIO_4$ $RuCl_3$, MeCN, CCl_4, RT., 20 h (**3** 10 %, **11a** 18 %).

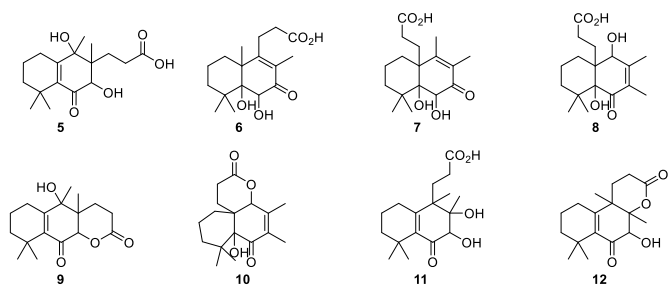


Fig. 1. Possible structures proposed for the unknown structure.

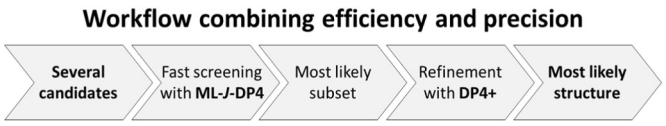


Fig. 2. Workflow illustrating the structural elucidation process used in this work, combining high efficiency (ML-*J*-DP4) and precision (DP4+).

correlation was performed in an inverse (upside-down) fashion.

The 32 structures were ranked according to ML-*J*-DP4 probability, with the four most likely structures selected for further refinement at the recommended DP4+ level (PCM/mPW1PW91/6-31 + G**//B3LYP/6-31G*). Notably, while ML-*J*-DP4 has demonstrated high accuracy, DP4+ remains the superior standard. This screening process was highly efficient, requiring only about 90 min on a single standard CPU. This is considerably less than the 12 days it would have taken to perform a full DP4+ level study using the same computational resources. An additional advantage of this approach is its capability to swiftly eliminate both incorrect connectivity and stereochemistry in a single step. In this regard, Fig. 3 presents the most likely relative configuration for each structural proposal, together with the corrected mean absolute errors (CMAE) obtained for each, as well as the global probability within the set of 32 candidates. This evaluation allows for the rapid elimination of structures **6**, **7**, **8**, **9**, **10**, and **12**, as they exhibit high errors in either carbon and/or proton data, resulting in low probabilities. Notably, this rapid assessment ruled out the lactonized structures (**9**, **10**, and **12**) by revealing significant errors at the cyclization sites (up to 12.4 ppm in ^{13}C NMR). Analysis of the ML-*J*-DP4 results reveals that structure **5a** achieved the highest score, while **11a** also demonstrated remarkable probability (Table 1A). Notably, both structures have the same groups exchanged between C-8 and C-9 positions, even preserving the orientation of the substituents. Although the probability percentages suggest a slight preference towards **5a** (76.04 % for **5a** versus 14.09 % for **11a**), a thorough evaluation shows that both structures exhibit similar average chemical shift errors. Specifically, the error values are 1.6 ppm for C-CMAE and 0.08 ppm for H-CMAE for **5a**, and 1.8 ppm and 0.06 ppm, respectively, for **11a**. Given the structural similarity between these candidates, it is essential to prioritize error evaluation over probability percentages alone. Therefore, **11a** was included in further analysis. The other candidates showed remarkably lower probabilities, mainly due to poor ^{13}C predictions, with C-CMAE ranging from 2.23 to 3.72 ppm (see SI for further details).

Following the workflow, the four most likely structures were selected for DP4+ analysis. Although most of the global probability is concentrated on structures **5a** and **11a**, the candidate set was expanded to

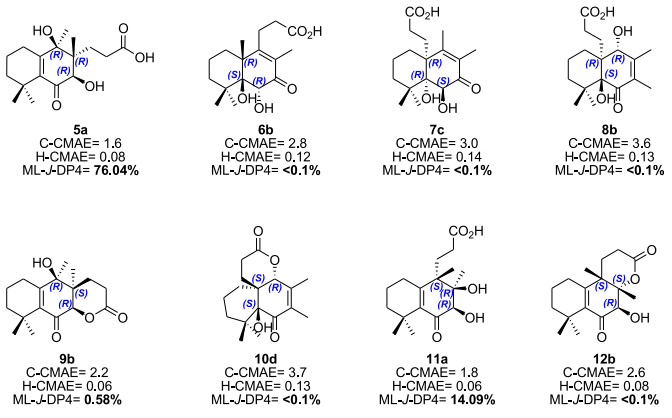


Fig. 3. Most probable relative configuration for each of the connectivities shown in Fig. 1, along with CMAE values using carbon and proton data (in ppm). The ML-*J*-DP4 values represent the overall probability calculated using the complete set of 32 candidates.

Table 1
C-CMAE, H-CMAE, C-CMAX (corrected maximum ^{13}C error), H-CMAX (corrected maximum ^{1}H error), and probability values obtained using ML-*J*-DP4 (**A**) and DP4+ (**B**) for the most promising structures. The data correlation was performed in an inverted (upside-down) manner. Note: compounds **5c** and **5d** are the 7*R*,8*S*,9*R* and 7*S*,8*S*,9*R* diastereoisomers of **5a**, respectively. The structures are shown in the Supporting Information (Fig. S19).

A)

ML-*J*-DP4	5a	5c	5d	11a
C-CMAE	1.6	1.7	1.6	1.8
H-CMAE	0.08	0.1	0.12	0.06
C-CMAX	5.1	3.4	3.6	5.9
H-CMAX	0.15	0.24	0.23	0.13
H data	5 %	0.56 %	0.11 %	18.33 %
C data	25.67 %	12.53 %	47.48 %	1.30 %
All data	**76.04 %**	4.17 %	3.10 %	**14.09 %**

B)

DP4+	**5a**	**5c**	**5d**	**11a**
C-CMAE	2.6	2.6	1.6	1.6
H-CMAE	0.16	0.15	0.12	0.05
C-CMAX	9.2	9.1	3.8	7.8
H-CMAX	0.45	0.29	0.35	0.12
H data	<0.01 %	<0.01 %	<0.01 %	**100.00 %**
C data	<0.01 %	<0.01 %	0.13 %	**99.87 %**
All data	<0.01 %	<0.01 %	<0.01 %	**100.00 %**

minimize potential biases. To ensure, with a high level of confidence, that the actual structure was included in the restricted set, an expansion scheme was implemented that guarantees a cumulative probability greater than 97 %. Therefore, in addition to the most probable structures, **5a** (76.04 %) and **11a** (14.09 %), **5c** (4.17 %) and **5d** (3.10 %) were also included. Accordingly, the NMR chemical shifts were recomputed at the recommended PCM/mPW1PW91/6-31 + G**//B3LYP/6-31G* level, and correlated with experimental information using the DP4+App, a Python program that automates the data extraction and probability calculation. The DP4+ analysis strongly and unambiguously indicated that structure **11a** was the most probable, with an associated probability >99 %. Furthermore, the C-CMAE and H-CMAE were the lowest observed among all evaluated structures (Table 1B), with values of 1.6 ppm and 0.05 ppm for carbon and protons, respectively, which are entirely consistent with the expected results for the level of theory used, thereby validating the finding.

Although structure **11a** stands out as the most likely candidate with high certainty (>99 %), these tools are known to operate on a relative scale, aiming to identify the most probable structure among a set of candidates. Therefore, a close examination of calculated values is essential to validate the structural proposal. In this context, while all nuclei showed excellent agreement between experimental and calculated values, an unusually high error was noted for C-10 (Δδ 7.80 ppm). Although an error of this magnitude is not sufficient to question the suggested connectivity, it cannot be dismissed either. Even more strikingly, this carbon was predicted with high accuracy at the ML-*J*-DP4 level (Δδ 0.2 ppm), which employs MMFF geometries (Fig. 4). This finding is counterintuitive, as a higher theoretical level would typically reduce discrepancies rather than cause a deviation of over 7 ppm. In this context, it is worth noting that the ^{13}C signal for C-13 also exhibited significant errors at both levels. However, it is well known that the experimental ^{13}C chemical shifts of carboxylic acids can show larger discrepancies relative to calculated values due to factors such as pH, dimerization, and other effects.

Analysis of the main contributing conformers of **11a** reveals the presence of an intramolecular hydrogen bond (IHB) between the hydroxyl group at C-7 and the adjacent carbonyl oxygen. Fig. 5 shows the two most stable conformations found at ML-*J*-DP4 and DP4+ levels of theory. Interestingly, the strength of this interaction appears

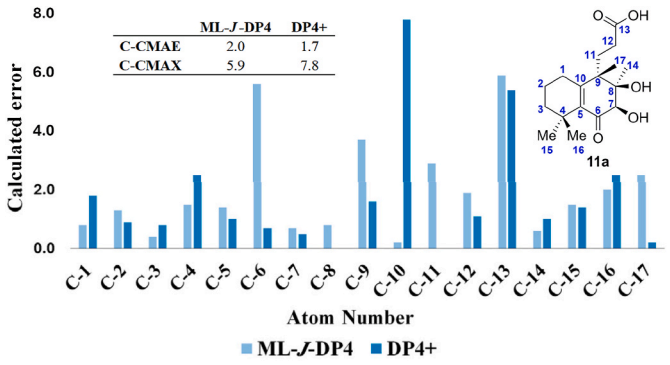


Fig. 4. Calculated absolute errors obtained at the ML-*J*-DP4 and DP4+ levels, respectively, for ^{13}C. CMAE and CMAX values for ^{13}C and ^{1}H of compound **11a**. The data correlation was performed by assigning each calculated value to its corresponding experimental value.

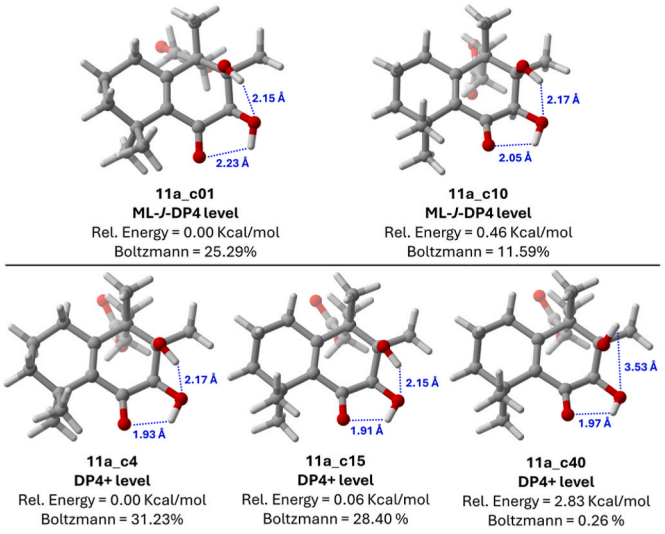


Fig. 5. Structures of relevant conformations for **11a** at the ML-*J*-DP4 (top) and DP4+ (bottom) levels; interatomic distances indicated with blue dashed lines (in angstroms, Å). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

significantly higher in geometries optimized at the B3LYP/6-31G* level, as indicated by the shorter bond distances (1.93 and 1.91 Å) compared to those observed with MMFF (2.23 and 2.05 Å, respectively). Furthermore, a second IHB is observed between the C-8 and C-7 hydroxyl groups, although no significant differences in bond distances were noted between MMFF and B3LYP/6-31G* geometries. A closer examination reveals an appreciable geometric change, shifting from ^{8}H half-chair conformations in MMFF to distorted H_7 shapes at the B3LYP/6-31G* level. This change maximizes the IHB between C7–OH and the C═O groups, leading to pronounced changes in bond angles and dihedral angles in ring B (averaging 2.4° and 19.9°, respectively). To determine whether this behavior is a limitation of the B3LYP/6-31G* level, the geometries were recalculated at various theoretical levels, including M06-2X, M06-2X-GD3, B3LYP-GD3, and ωB97XD functionals and the def2TZVP basis set. In all cases, similar structural features were observed, suggesting that this trend is consistent across different DFT methods.

Initially, the predictive discrepancy could be attributed to a simple shift in conformational preference. However, the effect is more profound, as there are other conformations, like c40, that are virtually identical to c15 but lack the hydrogen bond between C8–OH and C7–OH. These differences are subtle (Δ angle = 1°, Δ dihedral = 3.2°), yet they significantly reduce the error for C-10 to only 0.2 ppm. Although it is well-known that minor structural changes can produce major changes in the theoretical chemical shifts [23], this behavior requires attention. It was hypothesized that the difference could also be due to the strength of the IHB interaction. In c15, the IHB between C-7 and C═O is stronger than c40 (1.91 Å vs 1.97 Å), likely due to the assistance of C8–OH. This intensifies the electron-withdrawing effect (EWG) of the carbonyl group, which influences the conjugated system. NBO calculations (at B3LYP/6-31G* level) revealed an interesting insight: the π (C═C) → π* (C═O) donation is 24.35 kcal/mol for c15, compared to only 21.69 for c40 (Fig. 6). Consequently, the terminal carbon in the conjugated system is the most deshielded, consistent with computational observations.

However, this conformation is high in energy (>2.8 kcal/mol from the global minimum), and thus does not impact the weighted chemical shift. In fact, all conformations within 2.8 kcal/mol present the characteristic C8–OH→C7–OH→C═O IHB sequence, yielding substantial errors in C10 (Δδ from 4.4 ppm to 14.7 ppm). The large gap complicates accurate NMR predictions, as bridging it by simply changing the theory level is challenging. To mitigate potential energetic bias, we recalculated the DP4+ probabilities using MESSI (Multi-Ensemble Strategy for Structural Identification), a recent method inspired by the 'wisdom of the crowds' theory, which has proven highly effective in the stereochemical assignment of polyhydroxylated molecules [24]. Following the reported procedure (see Computational Details), isomer **11a** once again emerged as the most probable. After resolving the structure of **11a**, we sought to investigate whether the high error in C10 was merely a peculiar feature of this specific molecule or indicative of a broader phenomenon. To this end, similar systems were explored in the literature.

Smardaesidin F (**13**, Fig. 7), a 20-nor-isopimarane diterpene isolated from an endophytic fungal strain, *Smardaea* sp. AZ0432, occurring in the living photosynthetic tissue of the moss *Ceratodon purpureus* [25], exhibits significant structural similarity to **11a**.

In this case, an acceptable error was observed for the C-10 chemical shift using the ML-*J*-DP4 method (1.8 ppm), whereas the DP4+ level produced a significantly higher error (7.8 ppm). This observation again suggests that hydrogen-bond interactions may be overestimated at the DFT level, as evidenced by comparing the distances between the carbonyl oxygen at C-6 and the hydroxyl group at C-7, measured as 1.99 Å (DP4+) versus 2.18 Å (ML-*J*-DP4). The influence of hydrogen bonding between OH groups and C═O groups has been observed in previous works [26]. To further support our hypothesis, a literature review was performed looking for other structures with similarities to compounds **11a** and **13**. Accordingly, structures **15** [27] and **16** [28], as well as **3** and **14** [29] (Fig. 8), were evaluated using the DP4+ method. The findings reinforced our previous observations: structures **15** and **16** exhibited significantly high unscaled errors for C-β (7.2 and 8.4 ppm, respectively). Consistent with this, it was possible to confirm that in the absence of a hydroxyl group adjacent to a carbonyl within an α,β-unsaturated ring-fused system, no substantial errors were observed for the C-β. Structures **3** and **14** showed acceptable unscaled errors,

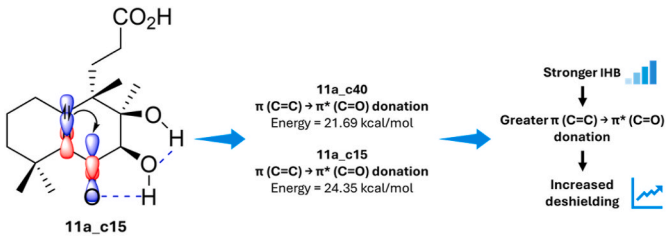


Fig. 6. Natural bond orbital (NBO) second-order perturbation theory analysis illustrating the donor-acceptor electronic density interactions. The stabilizing interaction observed is π (C═C) → π* (C═O), with an energy of 24.35 and 21.69 kcal/mol for c15 and c40, respectively.

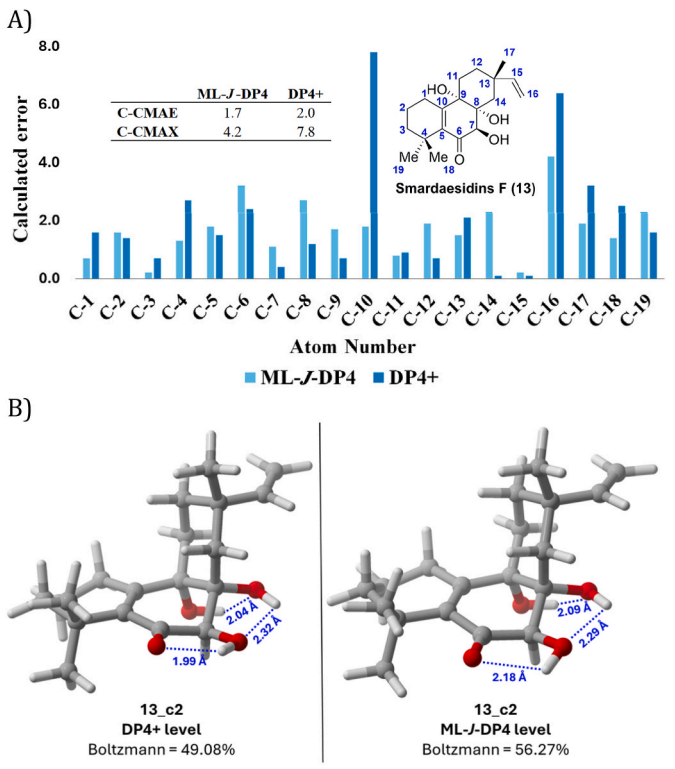


Fig. 7. A) Experimental values and calculated errors (ML-*J*-DP4 and DP4+) for ^{13}C and CMAE and CMAX values for ^{13}C and ^{1}H in Smardaesidins F (**13**). B) Structures of the most relevant conformations for **13** at the DP4+ and ML-*J*-DP4 levels with atom distances marked by blue dashed lines (in angstroms, Å). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

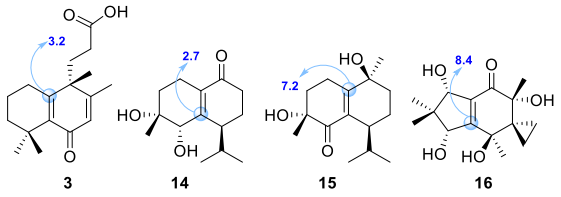


Fig. 8. Structures **3**, **14**, **15**, and **16**. The unscaled error (in ppm) between the experimental and calculated values at the DP4+ level is shown in blue for each structure. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

ranging from 2.7 to 3.2 ppm. A closer examination revealed that, despite the similarities between structures **14** and **15**, the presence of a hydroxyl group adjacent to the carbonyl in structure **15** results in a significantly larger error of 7.2 ppm.

In previous publications, we demonstrated that, in those cases where IHB generate conflicts that cannot be resolved by changing the level of theory, the manual removal of spurious conformations has proven to be effective [30,31]. Specifically, for compound **11a**, removing the conformers with the C8–OH→C7–OH→C═O IHB sequence significantly reduced the error at C-10 ($\Delta\delta \leq 3$ ppm). However, this approach introduces a considerable bias in the conformational landscape. The incorporation of explicit solvation could partially mitigate these issues, although it is challenging to implement for structural elucidation purposes [32]. In a recent publication, Hashimoto and collaborators solved this limitation through exhaustive acetylation. While this is an elegant solution, it requires additional experimentation, which is not always feasible when the amount of natural compounds material is limited [33].

Here, we propose that the virtual manipulation of hydroxy groups in similar entities, preventing IHB formation, could help address these problems associated with such interactions. In this context, the methoxy group emerged as an ideal alternative due to its low steric demand compared to other possible substituents. While this modification is expected to influence the chemical shifts of carbons directly bonded to oxygen, it is reasonable to assume that more distant signals should not be significantly affected. To examine this, compounds **11a-OMe** and **13-OMe** were *in silico* generated by replacing hydroxyl groups with methoxy groups. An exhaustive conformational search was then conducted at the MMFF level, followed by reoptimization at the B3LYP/6-31G* level and NMR calculations at the PCM/mPW1PW91/6-31 + G** level using a standard Boltzmann averaging approach. Consistent with the initial assumption, the errors in the C-10 carbon chemical shifts decreased significantly, from 7.8 to 4.5 ppm for **11a-OMe** and from 7.8 to 4.1 ppm for 13-OMe (Fig. 9).

Notably, the most stable conformation of **11a-OMe** closely aligned with **11a_c40**, which, despite its high energy, yielded the most accurate prediction at C-10 (Fig. 10). Furthermore, DP4+ calculations conducted on the 4 diastereoisomers of **11-OMe** yielded **11a-OMe** compound as the most likely one (>99 %), in perfect agreement with previous results. This demonstrates that the *in silico* methylation strategy opens new avenues for spectroscopic studies of molecules that might be influenced by IHB.

As a final attempt to validate our workflow, all structures under study were recalculated using the DP4+ method, with **11a** emerging as the most probable structure (see all calculations in the Supporting Information). Additionally, while the NOESY data obtained are not conclusive, weak correlations were observed between H-7/H-11, H-7/H-12, and H-11/H-14, lending support to the structure suggested by the calculations. The programs used in this work, ML-*J*-DP4 and DP4+App [34], are freely available at https://github.com/Sarotti-Lab, along with tutorials and detailed instructions.

3. Conclusions

In the search for ring-distortion derivatives of solidagenone, *in situ* ruthenium tetroxide was used for the oxidative cleavage of the olefin resulting in the formation of an oxidation product (**3**, which had not been previously reported) together with an unknown compound. In order to determine the chemical structure, one- and two-dimensional NMR measurements were performed, which allowed several structures to be proposed that were consistent with the spectroscopic data. To discover its identity, the computational methods commonly employed were utilized, successfully combining the speed of ML-*J*-DP4 to eliminate fewer promising structures with the accuracy of DP4+ to assign the most likely structure for the unknown compound with high confidence. Furthermore, the importance of evaluating the calculated and experimental differences for each atom was highlighted. In this context, a detailed analysis of the β-carbon in the α,β-unsaturated ring-fused system was conducted, which displayed an abnormally high error using the DP4+ approach but not when applying ML-*J*-DP4. It was proposed that this discrepancy may arise because IHB are overestimated at the DFT level. As a heuristic solution, *in silico* methylation of hydroxyl groups

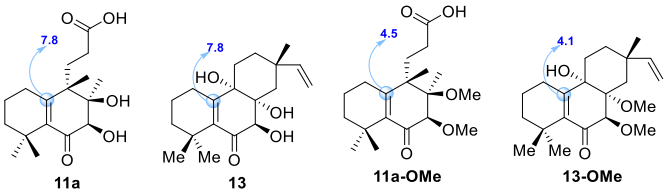
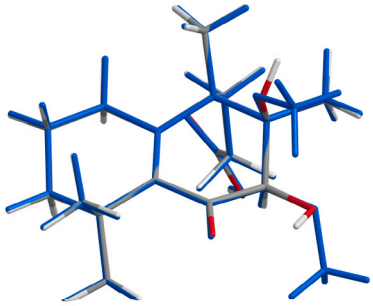


Fig. 9. Structures of **11a**, **13**, **11a-OMe** and **13-OMe**. The errors (in ppm) between the experimental and calculated values at the DP4+ level are shown in blue for each structure. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)



Structure	Boltzmann (%)	Exp	Calc.	Δ
11a _c40	0.26	159.1	159.3	0.2
11a-OMe_c12	20.23	159.1	157.4	1.7

Fig. 10. Overlaid structures of conformations **11a_c40** and **11-OMe_c12** (colored in blue). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

was tested, which led to a substantial improvement in the observed errors. Finally, previously reported structures in the literature similar to the conflicting structure of the present work were compared and the same magnitude of error was noted for the β-carbon when using DP4+.

CRediT authorship contribution statement

Iván Cortés: Investigation, Formal analysis. **María Luz Tibaldi-Bollati:** Investigation, Formal analysis. **Viviana E. Nicotra:** Writing – review & editing, Writing – original draft, Investigation. **Manuela E. García:** Writing – review & editing, Writing – original draft, Investigation, Conceptualization. **Ariel M. Sarotti:** Writing – review & editing, Writing – original draft, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This research was funded by ANPCyT (PICT-2019-4052, PICT-2021-CAT-II-00033, and PICT-2020- SERIEA-02162), CONICET (PIP 11220200102205CO and PIP 11220200101065CO), ONR Global (Grant N62909-21-1-2052, USA), and SeCyT-UNC (2023-CONSOLIDAR-33620230100785CB01). IC and MLT-B thank CONICET for their post-doctoral and doctoral fellowships, respectively.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tchem.2025.100125.

Data availability

Data provided in the Supporting Information

References

[1] S. Majhi, D. Das, Chemical derivatization of natural products: semisynthesis and pharmacological aspects-A decade update, Tetrahedron 78 (2021) 131801, https://doi.org/10.1016/j.tet.2020.131801.
[2] J. Kim, H. Kim, S.B. Park, Privileged structures: efficient chemical "navigators" toward unexplored biologically relevant chemical spaces, J. Am. Chem. Soc. 136 (2014) 14629–14638, https://doi.org/10.1021/ja508343a.
[3] A.G. Atanasov, S.B. Zotchev, V.M. Dirsch, I.E. Orhan, M. Banach, J.M. Rollinger, D. Barreca, W. Weckwerth, R. Bauer, E.A. Bayer, M. Majeed, A. Bishayee, V. Bochkov, G.K. Bonn, N. Braidy, F. Bucar, A. Cifuentes, G. D'Onofrio, M. Bodkin, M. Diederich, A.T. Dinkova-Kostova, T. Efferth, K. El Bairi, N. Arkells, T.-P. Fan, B. L. Fiebich, M. Freissmuth, M.I. Georgiev, S. Gibbons, K.M. Godfrey, C.W. Gruber, J. Heer, L.A. Huber, E. Ibanez, A. Kijjoa, A.K. Kiss, A. Lu, F.A. Macias, M.J.S. Miller, A. Mocan, R. Müller, F. Nicoletti, G. Perry, V. Pittalà, L. Rastrelli, M. Ristow, G. L. Russo, A.S. Silva, D. Schuster, H. Sheridan, K. Skalicka-Woźniak, L. Skaltsounis, E. Sobarzo-Sánchez, D.S. Bredt, H. Stuppner, A. Sureda, N.T. Tzvetkov, R.A. Vacca, B.B. Aggarwal, M. Battino, F. Giampieri, M. Wink, J.-L. Wolfender, J. Xiao, A.W. K. Yeung, G. Lizard, M.A. Popp, M. Heinrich, I. Berindan-Neagoe, M. Stadler, M. Daglia, R. Verpoorte, C.T. Supuran, Natural products in drug discovery: advances and opportunities, Nat. Rev. Drug Discov. 20 (2021) 200–216, https://doi.org/10.1038/s41573-020-00114-z.
[4] H. Lachance, S. Wetzel, K. Kumar, H. Waldmann, Charting, navigating, and populating natural product chemical space for drug discovery, J. Med. Chem. 55 (2012) 5989–6001, https://doi.org/10.1021/jm300288g.
[5] R.W. Huigens III, K.C. Morrison, R.W. Hicklin, T.A. Flood Jr., M.F. Richter, P. J. Hergenrother, A ring-distortion strategy to construct stereochemically complex and structurally diverse compounds from natural products, Nat. Chem. 5 (2013) 195–202, https://doi.org/10.1038/nchem.1549.
[6] S.E. Motika, P.J. Hergenrother, Re-engineering natural products to engage new biological targets, Nat. Prod. Rep. 37 (2020) 1395–1403, https://doi.org/10.1039/D0NP00059K.
[7] N. Grimblat, J.A. Gavín, A. Hernández Daranas, A.M. Sarotti, Combining the power of *J* coupling and DP4 analysis on stereochemical assignments: the *J*-DP4 methods, Org. Lett. 21 (2019) 4003–4007, https://doi.org/10.1021/acs.orglett.9b01193.
[8] N. Grimblat, M.M. Zanardi, 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. 80 (2015) 12526–12534, https://doi.org/10.1021/acs.joc.5b02396.
[9] A. Howarth, J.M. Goodman, The DP5 probability, quantification and visualisation of structural uncertainty in single molecules, Chem. Sci. 13 (2022) 3507–3518, https://doi.org/10.1039/D1SC04406K.
[10] H. Rull, M. Fischer, S. Kuhn, NMR shift prediction from small data quantities, J. Cheminf. 15 (2023) 114, https://doi.org/10.1186/s13321-023-00785-x.
[11] Y. Guan, S.V. Shree Sowndarya, L.C. Gallegos, P.C. St John, R.S. Paton, Real-time prediction of 1 H and 13 C chemical shifts with DFT accuracy using a 3D graph neural network, Chem. Sci. 12 (2021) 12012–12026, https://doi.org/10.1039/D1SC03343C.
[12] I.M. Novitskiy, A.G. Kutateladze, DU8+ computations reveal a common challenge in the structure assignment of natural products containing a carboxylic anhydride moiety, J. Org. Chem. 86 (2021) 17511–17515, https://doi.org/10.1021/acs.joc.1c02291.
[13] S.G. Smith, J.M. Goodman, Assigning stereochemistry to single diastereoisomers by GIAO NMR calculation: the DP4 probability, J. Am. Chem. Soc. 132 (2010) 12946–12959, https://doi.org/10.1021/ja105035r.
[14] Y.-H. Tsai, M. Amichetti, M.M. Zanardi, R. Grimson, A.H. Daranas, A.M. Sarotti, ML-*J*-DP4: an integrated quantum mechanics-machine learning approach for ultrafast NMR structural elucidation, Org. Lett. 24 (2022) 7487–7491, https://doi.org/10.1021/acs.orglett.2c01251.
[15] M. Luz Tibaldi-Bollati, V. Nicotra, G. Oksdath-Mansilla, M.E. García, Expanding diterpene complexity and diversity via photoinduced ring distortions, Chempluschem 89 (2024), https://doi.org/10.1002/cplu.202300537.
[16] T. Anthonsen, P.H. McCabe, R. McCrindle, R.D.H. Murray, Constituents of *Solidago species*—I, Tetrahedron 25 (1969) 2233–2239, https://doi.org/10.1016/S0040-4020(01)82771-X.
[17] D.M.F. de Souza, R.D. Sá, E.L. Araújo, K.P. Randau, Anatomical, phytochemical and histochemical study of *Solidago chilensis* Meyen, An. Acad. Bras. Ciênc. 90 (2018) 2107–2120, https://doi.org/10.1590/0001-3765201720160280.
[18] T. Anthonsen, P.H. McCabe, R. McCrindle, R.D.H. Murray, The structure of solidagenone, Chem. Commun. (1966) 740, https://doi.org/10.1039/c19660000740.
[19] T. Anthonsen, P.H. McCabe, R. McCrindle, R.D.H. Murray, G.A.R. Young, Constituents of solidago species—II, Tetrahedron 26 (1970) 3091–3097, https://doi.org/10.1016/S0040-4020(01)92892-3.
[20] I. Cuadrado, Á. Amesty, J. Cedrón, J. Oberti, A. Estévez-Braun, S. Hortelano, B. De las Heras, Semisynthesis and inhibitory effects of solidagenone derivatives on TLR-mediated inflammatory responses, Molecules 23 (2018) 3197, https://doi.org/10.3390/molecules23123197.
[21] V.A. Ignatenko, Y. Han, G.P. Tochtrop, Molecular library synthesis using complex substrates: expanding the framework of triterpenoids, J. Org. Chem. 78 (2013) 410–418, https://doi.org/10.1021/jo302211f.
[22] I. Cortés, A.M. Sarotti, Road map toward computer-guided total synthesis of natural products. The dysiherbol A case study: what if serendipity hadn't intervened? J. Org. Chem. 88 (2023) 14156–14164, https://doi.org/10.1021/acs.joc.3c01738.
[23] C. Cuadrado, A.H. Daranas, 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 20 (2022) 699, https://doi.org/10.3390/md20110699.
[24] M.O. Marcarino, L. Passaglia, M.M. Zanardi, A.M. Sarotti, Breaking the DFT energy bias caused by intramolecular hydrogen-bonding interactions with MESSI, A structural elucidation method inspired by wisdom of crowd theory, Chem. Eur J. 29 (2023), https://doi.org/10.1002/chem.202300420.
[25] X.-N. Wang, B.P. Bashyal, E.M.K. Wijeratne, J.M. U'Ren, M.X. Liu, M. K. Gunatilaka, A.E. Arnold, A.A.L. Gunatilaka, A.-G. Smardaesidins, Isopimarane and 20-nor-Isopimarane diterpenoids from *Smardaea* sp., a fungal endophyte of the moss *Ceratodon purpureus*, J. Nat. Prod. 74 (2011) 2052–2061, https://doi.org/10.1021/np2000864.

[26] C. Li, A.M. Sarotti, X. Wu, B. Yang, J. Turkson, Y. Chen, Q. Liu, S. Cao, An unusual benzoisoquinoline-9-one derivative and other related compounds with antiproliferative activity from Hawaiian endophytic fungus *Peyronellaea* sp, FT431, Molecules 24 (2019) 196, https://doi.org/10.3390/molecules24010196.
[27] C. Chyu, M. Ke, Y. Chang, S. Chien, Y. Kuo, New cadinane-type sesquiterpenes from the roots of *Taiwania cryptomerioides* hayata, Helv. Chim. Acta 90 (2007) 1514–1521, https://doi.org/10.1002/hlca.200790158.
[28] M.L. Burgess, Y.L. Zhang, K.D. Barrow, Characterization of new illudanes, illudins F, G, and H from the basidiomycete *Omphalotus nidiformis*, J. Nat. Prod. 62 (1999) 1542–1544, https://doi.org/10.1021/np990247d.
[29] O. Muraoka, M. Fujimoto, G. Tanabe, M. Kubo, T. Minematsu, H. Matsuda, T. Morikawa, I. Toguchida, M. Yoshikawa, Absolute stereostructures of novel norcadinane- and trinoreudesmane-type sesquiterpenes with nitric oxide production inhibitory activity from *Alpinia oxyphylla*, Bioorg. Med. Chem. Lett. 11 (2001) 2217–2220, https://doi.org/10.1016/S0960-894X(01)00413-9.
[30] M.M. Zanardi, M.A. Sortino, A.M. Sarotti, On the effect of intramolecular H-bonding in the configurational assessment of polyhydroxylated compounds with computational methods. The hyacinthacines case, Carbohydr. Res. 474 (2019) 72–79, https://doi.org/10.1016/j.carres.2019.01.011.
[31] M.M. Zanardi, M.O. Marcarino, A.M. Sarotti, Redefining the impact of Boltzmann analysis in the stereochemical assignment of polar and flexible molecules by NMR calculations, Org. Lett. 22 (2020) 52–56, https://doi.org/10.1021/acs.orglett.9b03866.
[32] A. Bagno, F. Rastrelli, G. Saielli, Prediction of the 1H and 13C NMR spectra of α-D-glucose in water by DFT methods and MD simulations, J. Org. Chem. 72 (2007) 7373–7381, https://doi.org/10.1021/jo071129v.
[33] R. Kanehara, Y. Oinuma, H. Maeda, K. Tanaka, M. Hashimoto, Triantaspirols A–C and paraphaeolactone Cs from *paraphaeosphaeria* sp. KT4192: sensitivity of CP3 in distinguishing close NMR signals, J. Nat. Prod. 87 (2024) 2487–2498, https://doi.org/10.1021/acs.jnatprod.4c00935.
[34] B.A. Franco, E.R. Luciano, A.M. Sarotti, 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. 86 (2023) 2360–2367, https://doi.org/10.1021/acs.jnatprod.3c00566.