JOC The Journal of Organic Chemistry

Featured Article

pubs.acs.org/joc

GIAO C−H COSY Simulations Merged with Artificial Neural Networks Pattern Recognition Analysis. Pushing the Structural Validation a Step Forward

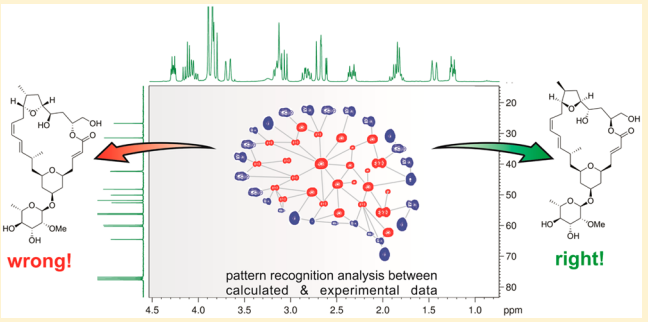
María M. Zanardi[†,‡] and Ariel M. Sarotti*[,†]

[†]Instituto de Química Rosario (CONICET), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Suipacha 531, Rosario 2000, Argentina

[‡]Facultad de Química e Ingeniería "Fray Rogelio Bacón", Pontificia Universidad Católica Argentina, Av. Pellegrini 3314, Rosario 2000, Argentina

Supporting Information

ABSTRACT: The structural validation problem using quantum chemistry approaches (confirm or reject a candidate structure) has been tackled with artificial neural network (ANN) mediated multidimensional pattern recognition from experimental and calculated 2D C−H COSY. In order to identify subtle errors (such as regio- or stereochemical), more than 400 ANNs have been built and trained, and the most efficient in terms of classification ability were successfully validated in challenging real examples of natural product misassignments.



INTRODUCTION

Nuclear magnetic resonance spectroscopy has revolutionized structural elucidation of natural products since the last half of the past century. An extraordinary improvement in NMR hardware and methodology has occurred over time,[1] though the structural or stereochemical misassignments were not completely inhibited by such evolution. This is evidenced by the hundreds of structural revisions published in the last two decades, many of them detected after the total synthesis of the originally proposed (wrong) structure.[2] The contribution of computational chemistry to prevent these painful and frustrating situations has been substantial, as accurate predictions of NMR parameters can be made with most quantum chemistry packages.[3] Good correlation between experimental and calculated NMR data provides confidence in the structural assignment. This has been facilitated and improved by sophisticated statistical methods introduced by Smith and Goodman to assign two sets of experimental data to two possible candidates (CP3)[4] or one set of experimental data to two or more plausible structures (DP4).[5]

All of the current approaches are "comparison-based" as they share a basic underlying principle: the correct structure shows the best correlation between calculated and experimental data among the candidates taken into consideration. While their importance in the structural elucidation field is unquestionable, it is essential to recognize the weak points. First, they are intended to point out the structure that, among the chosen candidates, correlates better with the experimental data, which could bring to a successful conclusion only if the correct structure was included as candidate. Second, they cannot be used in structural validation problems (confirmation of a putative structure), as the intrinsic absence of a second candidate structure makes impossible any comparison.

To sort these limitations, we recently demonstrated that NMR shift calculations in conjunction with artificial neural networks (ANNs) provided an efficient method for the detection of structural mistakes using one set of experimental and calculated data.[6] The proof-of-principle of this new approach, developed exclusively on the pattern recognition of ^{13}C data, showed excellent results mainly in the identification of connectivity errors. However, it tends to fail (vide infra) where the source of error is more subtle (i.e., stereochemical).

In order to take this approach a step further, we envisaged the development of powerful ANNs with enhanced classification ability refined enough to detect slight errors typically found in regio- or stereochemical missasignments. We considered that adding extra dimensions of data would increase the network ability to differentiate between right and wrong structures. Thus, 2D C−H COSY (commonly known as HSQC) experiments were introduced as templates for multilevel pattern recognition analysis. The inclusion of one-dimensional ^{1}H NMR (and its corresponding correlation with ^{13}C NMR) was thought of particular importance as it was found that proton data makes the most decisive contribution in structural elucidation problems.[7]

Received: July 1, 2015
Published: September 4, 2015

DOI: 10.1021/acs.joc.5b01663
J. Org. Chem. 2015, 80, 9371−9378

RESULTS AND DISCUSSION

Briefly, ANNs are mathematical models in which interconnected artificial neurons emulate a biological brain. Among several interesting properties, the ability of ANNs to learn from the data makes them convenient tools in pattern recognition, classification, and clustering analysis.[8] As in our previous study, herein we used two-layered feed-forward networks. As depicted in Figure 1, the input, hidden, and output layers (the three main components) are fully connected, meaning that each neuron is linked to every neuron in the precedent layer. In analogy with the biological neuron, the synapse (strength of the connection between two neurons) is measured by a weight number (w). Each connection carries an assigned weight, and activation functions (also known as transfer functions, f) control the output value. In this work, we used sigmoid transfer functions both in the hidden and output layers, as this architecture can classify vectors arbitrarily well.[9] A more detailed mathematical explanation of the net architecture is given in the Supporting Information.

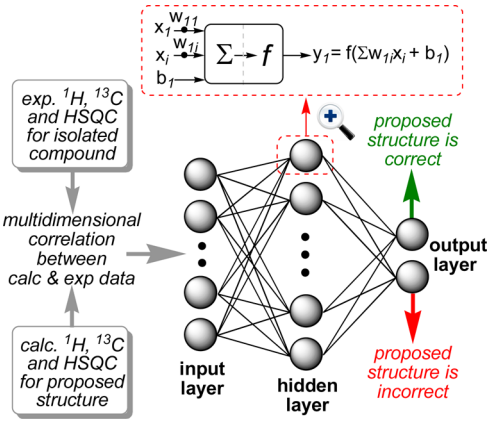


Figure 1. Two-layer feed-forward ANN.

Under supervised network training, a large number of data are given (input and output values) to train the ANN. A scaled conjugate gradient back-propagation algorithm was used to set the optimal weights and bias values for each connection.[8] Once trained, the network can then be used to make further predictions. The set of data used during the training step are statistical parameters of correlation between experimental and calculated NMR shifts from correct and incorrect molecules. For the correct test set, 100 known compounds (for which the ^{13}C and HSQC (heteronuclear single quantum correlation) NMR spectra are confidently assigned)[10] were taken to ensure functional groups diversity and molecular complexity. The set of incorrect structures was built by introducing slight modifications to some of the compounds used in the correct set by inverting a stereocenter or changing the position of few atoms. The calculated NMR shifts of the resulting stereo- and regioisomers (82 examples) were correlated with the experimental data corresponding to its precursor structure (Figure 2 shows some representative examples; for the complete set of structures see the Supporting Information). In addition, because some of the compounds of the correct set are diastereoisomers we could make 26 additional correlations between there experimental data computed for one isomer with the calculated data for the other isomer.[11] The resulting 182 structures were optimized at the B3LYP/6-31G* level,[12] and the NMR shielding tensors were calculated using the GIAO (gauge including atomic orbitals) method[13] at the mPW1PW91/6-31G* (gas phase)[14] and mPW1PW91/6-31G** (solution) levels of theory.

Once the NMR shifts were calculated at the two levels of theory, and using two different reference standards (TMS and MSTD, see Computational Methods), we proceeded to compute the statistical parameters that are employed to set the goodness of fit between experimental and calculated shifts that in turn serve as input layers to train the ANNs. There are two main procedures to do so: using assigned or unassigned data.

The first case entails knowing which simulated shift is linked to which experimental resonance. However, even after exhaustive NMR experiments, the misassignment of at least some signals can often occur, giving potential problems in the data correlation. Moreover, if the proposed structure is wrong any assignment of the experimental signals would be probably incorrect. For those reasons, we used unassigned data; that is, the experimental (δ_{exp}) and calculated (δ_{calc}) chemical shifts were arranged in descending order of size for both ^{1}H and ^{13}C signals. The systematic errors from the NMR calculations were eliminated by an empirical scaling procedure as $\delta_{scaled} = (\delta_{calc} - b)/m$, where m and b are the slope and the intercept, respectively, resulting from a linear regression calculation on a plot of δ_{calc} against δ_{exp}.[3] Apart from m, b, and R^2 (the correlation coefficient), we computed the mean absolute error (MAE, defined as $\Sigma_n|\delta_{calc} - \delta_{exp}|/n$), the corrected mean

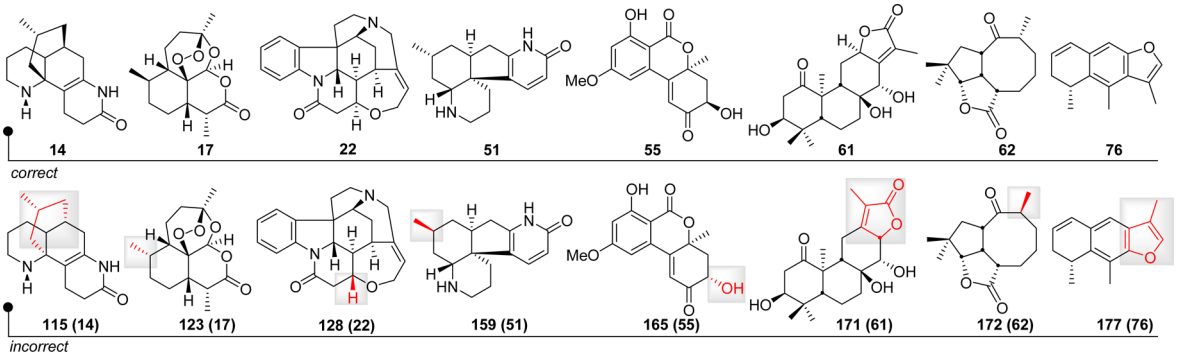


Figure 2. Eight representative examples of the molecules used in the correct test set (top) and the incorrect test set (bottom, data for precursor compound whose NMR was employed is shown in parentheses). The structural and/or stereochemical modification is framed in gray and highlighted in red.

DOI: 10.1021/acs.joc.5b01663
J. Org. Chem. 2015, 80, 9371−9378

absolute error (CMAE, defined as $\Sigma_n|\delta_{scaled} - \delta_{exp}|/n$), the standard deviation (σ, defined as $[\Sigma_n(|\delta_{calc} - \delta_{exp}| - MAE)^2/(n - 1)]^{1/2}$), the corrected standard deviation ($C\sigma$, defined as $[\Sigma_n(|\delta_{scaled} - \delta_{exp}| - CMAE)^2/(n - 1)]^{1/2}$), the maximum error (MaxErr, defined as $\max|\delta_{calc} - \delta_{exp}|$), and the corrected maximum error (CMaxErr, defined as $\max|\delta_{scaled} - \delta_{exp}|$). Furthermore, 18 extra parameters were computed from the correlation between experimental and calculated HSQC data. It is important to point out that the lack of importance of knowing which shift corresponds to which nucleus (unassigned data) should not be extrapolated into the C−H correlation. In the two-dimensional experiment, each cross-peak indicates directly bonded carbon−proton nuclei, while in the calculated analogue this information is obtained by considering the atom labels during the shift calculation procedure.[11] As depicted in Figure 3A, each error was defined as a distance from center to

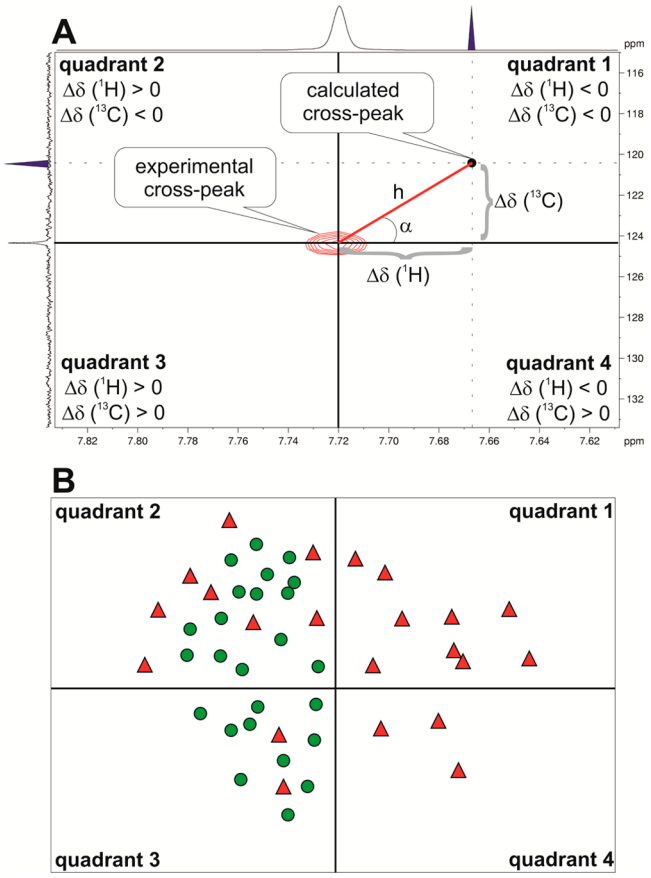


Figure 3. (A) Schematic representation of the correlation between experimental and calculated C−H cross-peaks. (B) Distribution of cross-peaks of compound **22** (correct, green dots) and its incorrect analogue (**128**, red triangles), computed using scaled chemical shifts at the mPW1PW91/6-31G*//B3LYP/6-31G* level with MSTD as the reference standard. All the experimental cross-peaks are merged in the center of coordinates.

center of the experimental and calculated cross-peaks ($h = [\Delta\delta(^{13}C)^2 + \Delta\delta(^{1}H)^2]^{1/2}$) and the corresponding angle α, defined as $\alpha = \arctan[\Delta\delta(^{13}C)/\Delta\delta(^{1}H)]$. The averaged values of h and α (M-h and M-α, respectively), the maximum value and standard deviation of h (Max-h and σ-h, respectively), along with the corresponding corrected values (CM-h, CM-α, CMax-h, and Cσ-h) represent 8 C−H terms. We also considered the quadrant distribution (Q1, Q2, Q3, and Q4) as a potential measure of the randomness of the correlations. This can be done by (hypothetically) placing all the experimental cross-peaks in a common origin of coordinates and computing the percentage of calculated cross-peaks that fall in each quadrant. Figure 3B illustrates this by a representative example of the distribution of calculated cross-peaks of **22** (correct, green dots) and **128** (incorrect, red triangles) using scaled chemical shifts computed at the mPW1PW91/6-31G*//B3LYP/6-31G* with MSTD as reference standard. Note that in this particular case the Q1, Q2, Q3, and Q4 values of **22** are 0%, 59%, 41%, and 0%, respectively, while for **128** the distribution changes to 41%, 36%, 9%, and 14%, respectively. The Q1-Q4 terms, along with the averaged value (MQ) can be computed from scaled and unscaled shifts, leading to 10 additional parameters.

In summary, for each reference standard (TMS and MSTD), 18 parameters are taken from the 1D correlations (9 from ^{1}H and 9 from ^{13}C) and 18 from 2D correlations to give a full matrix of 72 elements.

With the data in hand, we next explored the optimal combination of input and hidden layers to afford the best classification with the 208 examples of the training set. Regarding the size of the input layer, we used diverse arrays of statistical descriptors: the full matrix of 72 elements, the half matrices of 36 parameters from TMS or MSTD, and other submatrices (such as those containing only 1D data, 2D data, ^{1}H data, ^{13}C data, using TMS, MSTD, or both reference standards). For each input layer, different sizes of the hidden layer were investigated, ranging from 10 to 100 neurons. After several trials (in average, we generated and trained more than 400 different ANNs), we identified two networks that performed particularly well after the training, namely: ANN-TMS_{vac} and ANN-$MSTD_{sol}$. The first one was built using the 36 parameters computed at the mPW1PW91/6-31G* in gas phase with TMS as reference standard and 10 neurons in the hidden layer, and the second one was built with the 36 parameters computed at the mPW1PW91/6-31G** in solution with MSTD as standard and 20 neurons in the hidden layer. The percentage of correct classification achieved by both networks after the training was high (97% and 92%, respectively). From the collected results two main conclusions could be drawn:

(a) The best classification was achieved with all the data provided by the ^{1}H, ^{13}C, and C−H correlations, and removing any subset of statistical parameters resulted in a decrease of the pattern recognition ability of the network.

(b) Mixing the 36 parameters from TMS with the 36 parameters from MSTD (both in gas phase or in solution) did not improve the results (in fact, it declined the classification capacity of the ANNs). We speculated that this could be due, at least in part, to misleading data. In general, MSTD performs better than TMS when computing unscaled chemical shifts, though TMS can afford better correlation after scaling.[15] As the improvement of MSTD over TMS is much higher when including the solvent effects (in particular, for ^{1}H NMR), it is clearly the best performance of TMS-derived parameters in gas-phase calculations and MSTD-derived parameters computed in solution.

At this point, we speculated that mixing the input layers of ANN-TMS_{vac} and ANN-$MSTD_{sol}$ would allow to take the best of both worlds. In fact, the new network (called ANN-mix)

DOI: 10.1021/acs.joc.5b01663
J. Org. Chem. 2015, 80, 9371−9378

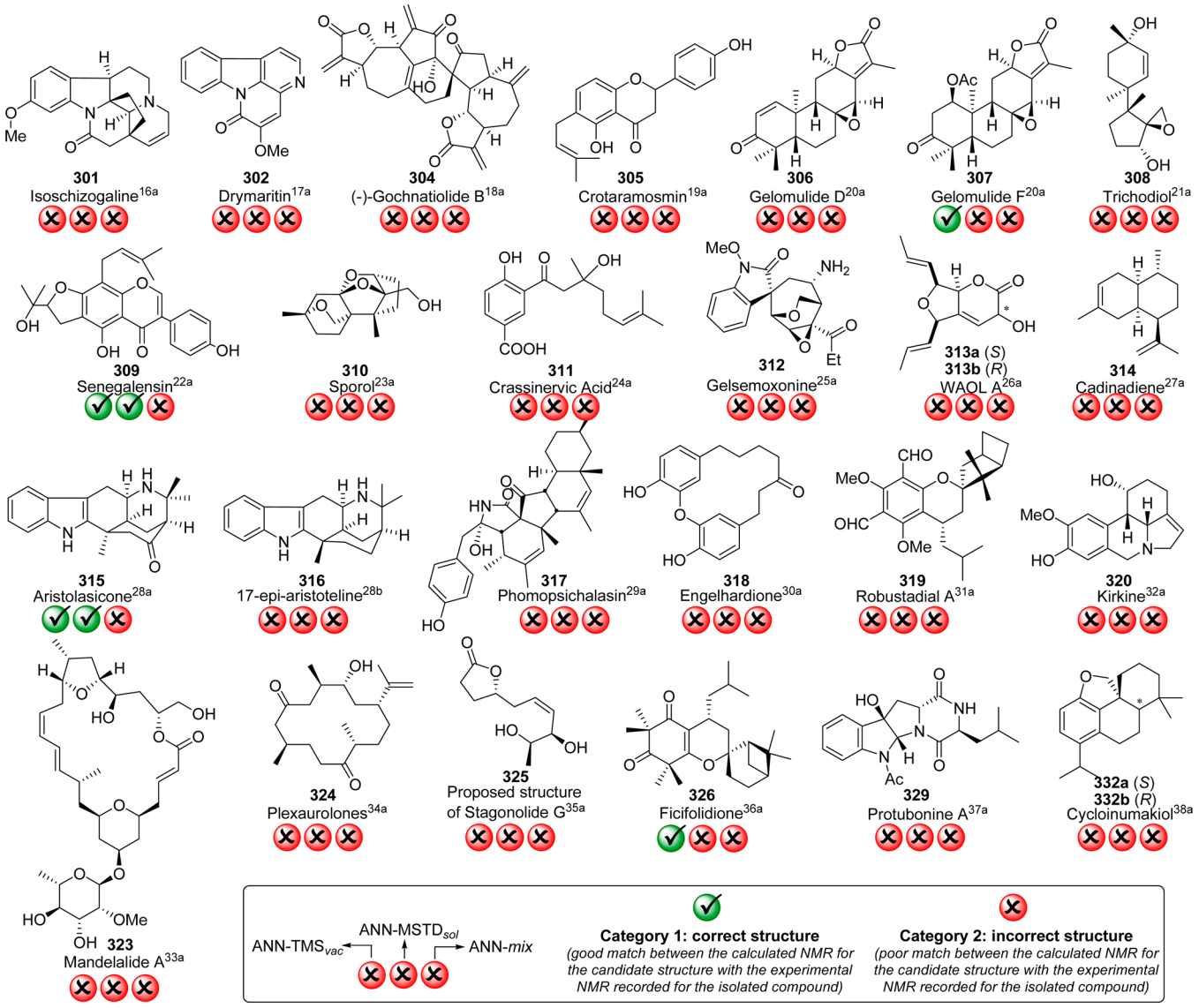


Figure 4. Natural products originally misassigned and classification results of the three optimal ANNs.

constructed with an input layer of 72 elements (36 parameters derived from gas phase calculations using TMS as standard and 36 parameters derived from solution calculations using MSTD as standard) and 14 neurons in the hidden layer performed very good after the training (94% of correct classification).

To examine whether adding extra dimensions of information indeed resulted in a significant improvement of the ANNs, we next evaluated the classification ability of the three selected trained networks (ANN-TMS$_{vac}$, ANN-MSTD$_{sol}$ and ANN-mix) in 25 challenging real cases of structural misassignments (Figure 4). The differences between the originally proposed and the revised structures (Figure 5) are mainly regio- and/or stereochemical, though we include some difficult examples of constitutional isomerism.

After extensive conformational searches, the shielding tensors of all structures shown in Figure 4 were computed at the mPW1PW91/6-31G*//B3LYP/6-31G* and PCM/mPW1PW91/6-31G**//B3LYP/6-31G* levels of theory, and the resulting chemical shifts were correlated with the experimental NMR values originally reported for those compounds using our trained networks.

To our delight, ANN-TMS$_{vac}$, ANN-MSTD$_{sol}$, and ANN-mix successfully detected the misassignments of the originally proposed structures in 85%, 94%, and 100% of the cases, respectively (Figure 4).[11] Such classification performance gains additional importance when considering the 25% correct identification achieved by of our best previously reported ANN-HF-18 (that was conceived to find patterns only in the ^{13}C NMR data).[6]

On the other hand, when dealing with the revised (correct) structures (Figure 5), ANN-TMS$_{vac}$ and ANN-MSTD$_{sol}$ failed in only one and two cases, respectively, whereas ANN-mix correctly classified all examples, clearly indicating its superior pattern recognition ability.[11]

In order to further validate our networks, we selected the three diastereoisomeric pairs **221**−**222**, **227**−**228**, and **230**−**231** (Table 1) and created "ex professo" errors by correlating the experimental data of one isomer with the calculated data for the other isomer (for example, **228**$_{calc}$−**227**$_{exp}$ and **227**$_{calc}$−**228**$_{exp}$). Again, the results were highly satisfactory in terms of pattern recognition ability.

DOI: 10.1021/acs.joc.5b01663
J. Org. Chem. 2015, 80, 9371−9378

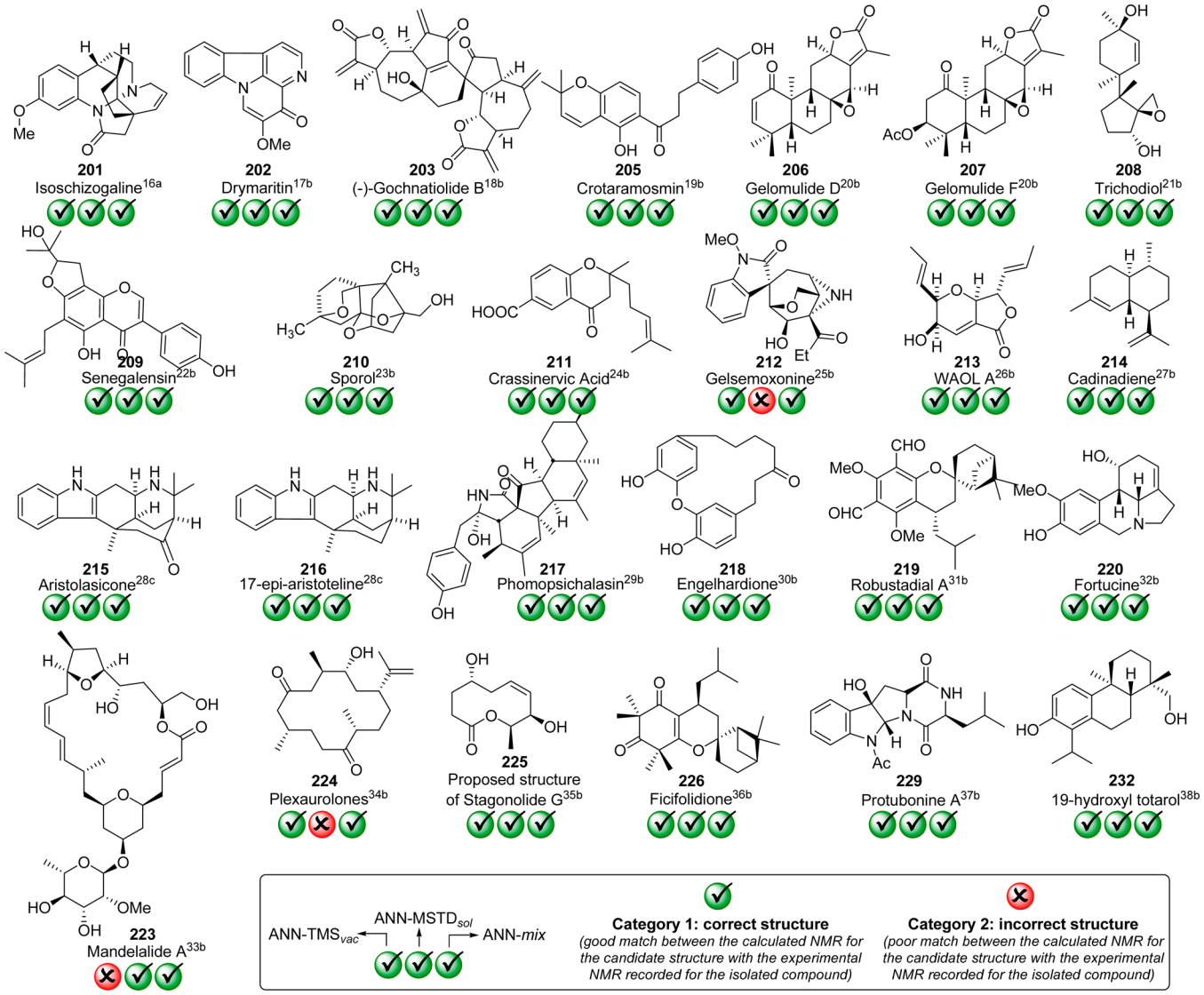
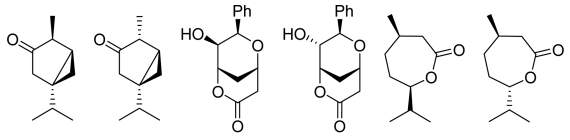


Figure 5. Revised structures of the natural products shown in Figure 4 and classification results of the three optimal ANNs.

Case Study. To illustrate the usefulness of our methodology, the recent case of the structural revision of mandelalide A (original structure: **323**, Figure 4; revised structure: **223**, Figure 5) is discussed. This complex glycosylated macrolide was isolated in 2012 from new species of *Lissoclinum* ascidian (collected from South Africa), and its planar structure was elucidated by extensive 1D and 2D NMR data, whereas the relative configuration was assigned from homonuclear and heteronuclear coupling constants analysis and ROESY experiments.[33a] Hydrolysis and GC−MS identification of the resulting monosaccharide (2-*O*-methyl-α-L-rhamnose) allowed the unequivocal determination of the absolute configuration of the southern region of the molecule and, by extrapolation, of the northern region as well. The initial screenings suggested remarkable cytotoxic activity against human NCI-H460 lung cancer (IC_{50} = 12 nM) and mouse Neuro-2A neuroblastoma cell lines (IC_{50} = 44 nM), though further investigation had to be suspended due to the limited amount of isolated sample (0.8 mg).[33a] This motivated synthetic organic groups to pursue the total synthesis of mandelalide A, and it was not long before the first results appeared.[33b,41] In 2014, Willwacher and Fürstner accomplished the first total synthesis of the putative structure of mandelalide A (**323**) and found that the NMR data of the synthesized compound showed small by indisputable differences with the natural product.[41a] The preparation of the C-11 epimer did not solve the correct architecture of mandelalide A, which was finally unraveled the same year by Xu and Ye (and co-workers) after total synthesis of the proposed structure and two more stereoisomers (one of them, the revised mandelalide A, **223**).[33b] The origin of the mistake was the stereochemistry of the upper side of the molecule, with all the corresponding stereocenters inverted, that had been originally determined on the basis or the relative stereochemistry between the tetrahydrofuran and tetrahydropyran fragments, though the assignment of two separated stereoclusters can be challenging.[3e] To analyze how the present methodology could have been useful in preventing such frustrating situation (as stated by Fürstner),[41a] it is important to begin pointing out that computed chemical shifts of compound **323** showed good agreement with the experimental data. For example, at the mPW1PW91/6-31G* level the CMAE was 1.6 ppm (^{13}C) and 0.15 ppm (^{1}H), with maximum outliers (CMaxErr) of 6.8 ppm (^{13}C) and 0.50 ppm (^{1}H).[42] These representative values do not indicate the presence of any structural error as they fall in the

DOI: 10.1021/acs.joc.5b01663
J. Org. Chem. 2015, 80, 9371−9378

Table 1. Performance of the Trained ANNs in Additional Validation Examples



Experimental NMR from	Calculated NMR for	Classification
221	222	✗✗✗
222	221	✗✗✗
221	221	✓✓✓
222	222	✓✓✓
227	228	✗✗✗
228	227	✗✗✗
227	227	✓✓✓
228	228	✓✓✓
230	231	✗✗✗
231	230	✓✗✗
230	230	✓✓✓
231	231	✓✓✓

typical range of correct structures (CMAE ^{13}C: 0.2−2.2 ppm; CMAE ^{1}H: 0.03−0.23 ppm; CMaxErr ^{13}C: 0.4−9.5 ppm; CMaxErr ^{1}H: 0.05−0.72 ppm). Interestingly, it is the combination of these parameters (and many other as well) which defines the pattern that allows our trained ANNs to correctly classify the originally proposed structure (**323**) as incorrect (Figure 4). This alarm could have triggered a more thorough revision (for example, considering all plausible isomers and computing the DP4 probability,[5] additional NMR experiments, etc.).

Quoting Fürsner's concluding remark after his total synthesis of the putative mandelalide A, "*somewhat ironically, they remind us that contemporary natural product total synthesis does not only serve the supply management alluded to in the introduction; all too often it is needed, even in the age of ever more sophisticated spectroscopy, to decide on structural issues*".[41a] We hope that the method herein presented will provide a helpful tool in making structural decisions faster, simpler, and cheaper than by total synthesis of the incorrectly assigned compound.

CONCLUSION

In summary, we have shown that using ANN-mediated multidimensional pattern recognition from experimental and calculated 2D C−H COSY allows the identification of subtle structural misassignments, such as regio- or stereochemical. The best results were obtained by mixing the statistical descriptors computed at the mPW1PW91/6-31G* level (in gas phase) with TMS as reference standard with those obtained at the mPW1PW91/6-31G** level (in solution) with the MSTD approach, and this approach is recommended to obtain the most reliable results. Moreover, in an effort to bring this methodology to the organic chemistry community, an Excel file that facilitates the calculation procedure is available from the authors in the Supporting Information.

One final reflection should be discussed. When analyzing the right and wrong pair of a given molecule (for instance, those shown in Figures 4 and 5 and Table 1), comparing the calculated data for both with the experimental values might be the instinctive reflex to determine which is the incorrect example. However, such an impulse must be dismissed as the decision making should not lie on any comparison; the challenge here is to decide if a given (only one) putative structure is correct based on the experimental data and the computed NMRs for that candidate. Having clarified this issue, the fact that the trained ANNs could actually identify structural errors as subtle as the inversion of one stereocenter can be represents a remarkable feature.

EXPERIMENTAL SECTION

Computational Methods. All of the quantum mechanical calculations were performed using Gaussian 09.[43] In the case of conformationally flexible compounds, the conformational search was done in the gas phase using the MM+ force field[44] (implemented in Hyperchem),[45] with the number of steps large enough to find all low-energy conformers at least 10 times. All conformers within 5 kcal/mol of the lowest energy conformer were subjected to further reoptimization at the B3LYP/6-31G* level of theory. All conformers within 2 kcal/mol from the B3LYP/6-31G* global minima were subjected to further NMR calculations. The magnetic shielding constants (σ) were computed using the gauge including atomic orbitals (GIAO) method, the method of choice among the different approaches to solve the gauge origin problem, with the mPW1PW91 functional (one of the most reliable DFT functionals for NMR calculations).[3] Single-point NMR calculations were carried out in the gas phase (with the 6-31G* basis set) and in solution (with the 6-31G** basis set) using the polarizable continuum model (PCM) with chloroform as solvent.[46] The NMR shielding constants were subjected to Boltzmann averaging over all conformers according to eq 1

$$\sigma^x = \frac{\sum_i \sigma^x{}_i \exp(-E_i/RT)}{\sum_i \exp(-E_i/RT)} \quad (1)$$

where σ^x is the Boltzmann-averaged shielding constant for nucleus x, $\sigma^x{}_i$ is the shielding constant for nucleus x in conformer i, R is the molar gas constant (8.3145 J K^{-1} mol^{-1}), T is the temperature (298 K), and E_i is the energy of conformer i (relative to the lowest energy conormer) obtained from the single-point NMR calculations (mPW1PW91/6-31G* in gas phase or mPW1PW91/6-31G** in solution). The chemical shifts were calculated according to eq 2

$$\delta^x{}_{\text{calc}} = \sigma_{\text{ref}} - \sigma^x + \delta_{\text{ref}} \quad (2)$$

where σ_{ref} is the NMR isotropic magnetic shielding values for the reference compound and δ_{ref} is the experimental chemical shift of the reference compound in deuterated chloroform. In this study, two different methods to calculate the NMR chemical shifts were used, namely TMS and MSTD.[15] In the TMS method, all chemical shifts are calculated using tetramethylsilane (TMS) as reference standard (δ_{ref} = 0.00 ppm), while in the multistandard approach (MSTD), methanol (δ_{ref} = 50.41 ppm for ^{13}C and δ_{ref} = 3.49 ppm for ^{1}H) and benzene (δ_{ref} = 128.37 for ^{13}C and δ_{ref} = 7.36 ppm for ^{1}H) were used as reference standards for sp^3 and sp−sp^2-hybridized carbon atoms (or proton attached to sp^3- and sp−sp^2-hybridized carbon atoms), respectively. Sarotti and Pellegrinet have recently found that this simple modification allowed much better accuracy and lower

DOI: 10.1021/acs.joc.5b01663
J. Org. Chem. 2015, 80, 9371−9378

dependence on the theory level employed, both for ^{13}C and ^{1}H NMR shift calculation procedures.[15]

The ANN training was done using the Neural Network Toolbox incorporated in MATLAB 7.0.22.[9]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.joc.5b01663.

Instructions for using the Excel file; full list of compounds used in the training set; experimental chemical shifts, GIAO isotropic magnetic shielding values, and statistical descriptors for all structures; weight and bias matrices for all trained ANNs (PDF)

Excel file for automatic chemical shift and ANN calculations (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: sarotti@iquir-conicet.gov.ar.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This research was supported by UNR (BIO 316), ANPCyT (PICT 2011-0255 and PICT-2012-0970), and CONICET (PIP 11220130100660CO). M.M.Z. thanks CONICET for the award of a fellowship (Exp. N° 3759/13).

■ REFERENCES

(1) (a) Gil, R. R. *Angew. Chem., Int. Ed.* **2011**, *50*, 7222. (b) Breton, R. C.; Reynolds, W. F. *Nat. Prod. Rep.* **2013**, *30*, 501.

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

(3) For recent reviews, see: (a) Lodewyk, M. W.; Siebert, M. R.; Tantillo, D. J. *Chem. Rev.* **2012**, *112*, 1839. (b) Bagno, A.; Saielli, G. *Wiley Interdisciplinary Reviews: Computational Molecules Science* **2015**, *5*, 228. (c) Tantillo, D. J. *Nat. Prod. Rep.* **2013**, *30*, 1079. (d) Bifulco, G.; Dambruoso, P.; Gomez-Paloma, L.; Riccio, R. *Chem. Rev.* **2007**, *107*, 3744−3779. For recent references, see: (e) Cen-Pacheco, F.; Rodríguez, J.; Norte, M.; Fernández, J. J.; Hernández Daranas, A. *Chem. - Eur. J.* **2013**, *19*, 8525. (f) Lodewyk, M. W.; Soldi, C.; Jones, P. B.; Olmstead, M. M.; Rita, J.; Shaw, J. T.; Tantillo, D. J. *J. Am. Chem. Soc.* **2012**, *134*, 18550. (g) Quasdorf, K. W.; Huters, A. D.; Lodewyk, M. W.; Tantillo, D. J.; Garg, N. K. *J. Am. Chem. Soc.* **2012**, *134*, 1396. (h) Saielli, G.; Nicolaou, K. C.; Ortiz, A.; Zhang, H.; Bagno, A. *J. Am. Chem. Soc.* **2011**, *133*, 6072. (i) Lodewyk, M. W.; Tantillo, D. J. *J. Nat. Prod.* **2011**, *74*, 1339.

(4) Smith, S. G.; Goodman, J. M. *J. Org. Chem.* **2009**, *74*, 4597.

(5) Smith, S. G.; Goodman, J. M. *J. Am. Chem. Soc.* **2010**, *132*, 12946.

(6) Sarotti, A. M. *Org. Biomol. Chem.* **2013**, *11*, 4847.

(7) (a) Chini, M. G.; Riccio, R.; Bifulco, G. *Eur. J. Org. Chem.* **2015**, *2015*, 1320. (b) Marell, D. J.; Emond, S. J.; Kulshrestha, A.; Hoye, T. R. *J. Org. Chem.* **2014**, *79*, 752.

(8) Zupan, J.; Gasteiger, J. *Neural Networks in Chemistry and Drug Design*; Wiley VCH: Weinheim, 1999.

(9) *MATLAB*; MathWorks: Natick, MA, 2007.

(10) The complete set of structures and the full references of the original papers are provided in the Supporting Information.

(11) For further details on this issue, see the Supporting Information.

(12) In our previous work, we used MM+, AM1, and HF/3-21G geometries for further mPW1PW91/6-31G* ^{13}C NMR chemical shift calculations. However, in preliminary stages of this study we noted that such geometries often led to poor predictions mainly in the ^{1}H NMR chemical shifts. A full account of this observation is beyond the scope of the present paper, though it is well-known that small errors in the starting geometry can lead to significant errors in the NMR calculations. For that reason, we chose to use the standard and well-known B3LYP/6-31G* method for the geometry optimization step.

(13) (a) Ditchfield, R. *J. Chem. Phys.* **1972**, *56*, 5688. (b) Ditchfield, R. *Mol. Phys.* **1974**, *27*, 789. (c) McMichael Rohlfing, C.; Allen, L. C.; Ditchfield, R. *Chem. Phys.* **1984**, *87*, 9. (d) Wolinski, K.; Hinton, J. F.; Pulay, P. *J. Am. Chem. Soc.* **1990**, *112*, 8251.

(14) Adamo, C.; Barone, V. *J. Chem. Phys.* **1998**, *108*, 664.

(15) (a) Sarotti, A. M.; Pellegrinet, S. C. *J. Org. Chem.* **2009**, *74*, 7254. (b) Sarotti, A. M.; Pellegrinet, S. C. *J. Org. Chem.* **2012**, *77*, 6059.

(16) (a) Renner, U. *Loydia* **1964**, *27*, 406. (b) Kariba, R. M.; Houghton, P. J.; Yenesew, A. *J. Nat. Prod.* **2002**, *65*, 566−569.

(17) (a) Hsieh, P.-W.; Chang, F.-R.; Lee, K.-H.; Hwang, T.-L.; Chang, S.-M.; Wu, Y. C. *J. Nat. Prod.* **2004**, *67*, 1175. (b) Wetzel, I.; Allmendinger, L.; Bracher, F. *J. Nat. Prod.* **2009**, *72*, 1908.

(18) (a) Bohlmann, F.; Zdero, C.; Schmeda-Hirschmann, G.; Jakupovic, J.; Dominguez, X. A.; King, R. M.; Robinson, H. *Phytochemistry* **1986**, *25*, 1175. (b) Li, C.; Dian, L.; Zhang, W.; Lei, X. *J. Am. Chem. Soc.* **2012**, *134*, 12414−12417.

(19) (a) Khalilullah, M. D.; Sharma, V. M.; Rao, P. S.; Raju, K. R. *J. Nat. Prod.* **1992**, *55*, 229. (b) Rao, M. S.; Rao, P. S.; Tóth, G.; Balázs, B.; Duddeck, H. *J. Nat. Prod.* **1998**, *61*, 1148.

(20) (a) Talapatra, S. K.; Das, G.; Talapatra, B. *Phytochemistry* **1989**, *28*, 1181. (b) Choudhary, M. I.; Gondal, H. Y.; Abbaskhan, A.; Jahan, I. A.; Parvez, M.; Nahar, N. *Tetrahedron* **2004**, *60*, 7933.

(21) (a) Nozoe, S.; Machida, Y. *Tetrahedron* **1972**, *28*, 5105. (b) Hesketh, A. R.; Bycroft, B. W.; Dewick, P. M.; Gilbert, J. *Phytochemistry* **1992**, *32*, 105.

(22) (a) Wandji, J.; Nkengfack, A. E.; Fomum, Z. T.; Ubillas, R.; Killday, K. B.; Tempesta, M. S. *J. Nat. Prod.* **1990**, *53*, 1425. (b) Tanaka, H.; Doi, M.; Etoh, H.; Watanabe, N.; Shimizu, H.; Hirata, M.; Khan, M. R. *J. Nat. Prod.* **2001**, *64*, 1336.

(23) (a) Corley, D. G.; Rottinghaus, G. E.; Tempesta, M. S. *Tetrahedron Lett.* **1986**, *27*, 427. (b) Ziegler, F. E.; Nangia, A.; Tempesta, M. S. *Tetrahedron Lett.* **1988**, *29*, 1665.

(24) (a) Lago, J. H. G.; Ramos, C. S.; Casanova, D. C. C.; Morandim, A. D. A.; Bergamo, D. C. B.; Cavalheiro, A. J.; Kato, M. J. *J. Nat. Prod.* **2004**, *67*, 1783. (b) Chakor, J. N.; Merlini, L.; Dallavalle, S. *Tetrahedron* **2011**, *67*, 6300.

(25) (a) Lin, L. Z.; Cordell, G. A.; Ni, C. Z.; Clardy, J. *Phytochemistry* **1991**, *30*, 1311. (b) Kitajima, M.; Kogure, N.; Yamaguchi, K.; Takayama, H.; Aimi, N. *Org. Lett.* **2003**, *5*, 2075.

(26) (a) Nozawa, O.; Okazaki, T.; Sakai, N.; Komurasaki, T.; Hanada, K.; Morimoto, S.; Mizoue, K. *J. Antibiot.* **1995**, *48*, 113. (b) Gao, X.; Nakadai, M.; Snider, B. B. *Org. Lett.* **2003**, *5*, 451.

(27) (a) Brown, G. D.; Shill, J. *Planta Med.* **1994**, *60*, 495. (b) Ngo, K. S.; Brown, G. D. *Tetrahedron* **1999**, *55*, 15099.

(28) (a) Kan-Fan, C.; Quirion, J. C.; Bick, I. R. C.; Husson, H. P. *Tetrahedron* **1988**, *44*, 1651. (b) Kyburz, R.; Schopp, E.; Bick, I.R. C.; Hesse, M. *Helv. Chim. Acta* **1981**, *64*, 2555. (c) Quirion, J. C.; Husson, H. P.; Kan, C.; Laprevote, O.; Chiaroni, A.; Riche, C.; Bick, I. R. C. *J. Org. Chem.* **1992**, *57*, 5848.

(29) (a) Horn, W. S.; Simmonds, M. S. J.; Schwartz, R. E.; Blaney, W. M. *Tetrahedron* **1995**, *51*, 3969. (b) Brown, S. G.; Jansma, M. J.; Hoye, T. R. *J. Nat. Prod.* **2012**, *75*, 1326.

(30) (a) Lin, W. Y.; Peng, C. F.; Tsai, I. L.; Chen, J. J.; Cheng, M. J.; Chen, I. S. *Planta Med.* **2005**, *71*, 171. (b) Shen, L.; Sun, D. *Tetrahedron Lett.* **2011**, *52*, 4570.

(31) (a) Xu, R.; Snyder, J. K.; Nakanishi, K. *J. Am. Chem. Soc.* **1984**, *106*, 734. (b) Cheng, Q.; Snyder, J. K. *J. Org. Chem.* **1988**, *53*, 4562.

(32) (a) Bastida, J.; Codina, C.; Peeters, P.; Rubiralta, M.; Orozco, M.; Luque, F. J.; Chhabra, S. C. *Phytochemistry* **1995**, *40*, 1291. (b) Biechy, A.; Hachisu, S.; Quiclet-Sire, B.; Ricard, L.; Zard, S. Z. *Tetrahedron* **2009**, *65*, 6730.

(33) (a) Sikorska, J.; Hau, A. M.; Anklin, C.; Parker-Nance, S.; Davies-Coleman, M. T.; Ishmael, J. E.; McPhail, K. L. *J. Org. Chem.*

DOI: 10.1021/acs.joc.5b01663
J. Org. Chem. 2015, 80, 9371−9378

2012, *77*, 6066. (b) Lei, H.; Yan, J.; Yu, J.; Liu, Y.; Wang, Z.; Xu, Z.; Ye, T. *Angew. Chem., Int. Ed.* **2014**, *53*, 6533.
(34) (a) Ealick, S. E.; Van der Helm, D.; Gross, R. A.; Weinheimer, A. J.; Ciereszko, L. S.; Middlebrook, R. E. *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.* **1980**, *36*, 1901. (b) Tello, E.; Castellanos, L.; Arevalo-Ferro, C.; Rodríguez, J.; Jiménez, C.; Duque, C. *Tetrahedron* **2011**, *67*, 9112.
(35) (a) Mikhailova, N. F.; Tarasov, À. V. *Bot. J.* **1989**, *74*, 509. (b) Rajendra Prasad, K.; Venkanna, A.; Babu, K. S.; Prasad, A. R.; Rao, J. M. *Tetrahedron Lett.* **2014**, *55*, 616.
(36) (a) Cheng, Q.; Snyder, J. K. *J. Org. Chem.* **1988**, *53*, 4562. (b) Nishiwaki, H.; Fujiwara, S.; Wukirsari, T.; Iwamoto, H.; Mori, S.; Nishi, K.; Shuto, Y. *J. Nat. Prod.* **2015**, *78*, 43.
(37) (a) Lee, S. U.; Asami, Y.; Lee, D.; Jang, J. H.; Ahn, J. S.; Oh, H. *J. Nat. Prod.* **2011**, *74*, 1284. (b) Lorenzo, P.; Álvarez, R.; de Lera, Á. R. *Eur. J. Org. Chem.* **2014**, *2014* (12), 2557.
(38) (a) Devkota, K. P.; Ratnayake, R.; Colburn, N. H.; Wilson, J. A.; Hendrich, C. J.; McMahon, J. B.; Beutler, J. A. *J. Nat. Prod.* **2011**, *74*, 374. (b) Xu, T.; Dong, G. *Angew. Chem.* **2014**, *126*, 10909.
(39) Berger, S.; Sicker, D. *Classics in spectroscopy*; Wiley-VCH: Weinheim, 2009; p 406.
(40) Liou, J. R.; Wu, T. Y.; Thang, T. D.; Hwang, T. L.; Wu, C. C.; Cheng, Y. B.; Wu, Y. C. *J. Nat. Prod.* **2014**, *77*, 2626−2632.
(41) (a) Willwacher, J.; Fürstner, A. *Angew. Chem., Int. Ed.* **2014**, *53*, 4217. (b) For the synthesis of the putative aglycon, see: Reddy, K. M.; Yamini, V.; Singarapu, K. K.; Ghosh, S. *Org. Lett.* **2014**, *16*, 2658. (c) For the second total synthesis of the revised mandelalide A, see: Willwacher, J.; Heggen, B.; Wirtz, C.; Thiel, W.; Fürstner, A. *Chem. - Eur. J.* **2015**, *21*, 10416.
(42) The same results were obtained at the PCM/mPW1PW91/6-31G** level using MSTD: CMAE = 1.9 ppm (^{13}C) and 0.13 ppm (^{1}H); CMaxErr = 5.4 ppm (^{13}C) and 0.50 ppm (^{1}H).
(43) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, O.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, 2009.
(44) Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127.
(45) *Hyperchem Professional Release 7.52*; Hypercube, Inc.: Gainesville, 2005.
(46) For a review on continuum solvation models, see: Tomasi, J.; Mennucci, B.; Cammi, R. *Chem. Rev.* **2005**, *105*, 2999.

DOI: 10.1021/acs.joc.5b01663
J. Org. Chem. 2015, 80, 9371−9378