Error bounds on goodness of fit indicators in vibrational circular dichroism spectroscopy

Error Bounds on Goodness of Fit Indicators in Vibrational Circular Dichroism Spectroscopy

Guglielmo Monaco[a,*], Giovanni Procida[a], Antonia Di Mola[a], Wouter Herrebout[b], Antonio Massa[a]

[a] *Dipartimento di Chimica e Biologia "A. Zambelli", Università di Salerno, Fisciano 84084, Italy*
[b] *Department of Chemistry, University of Antwerp, Antwerp, Belgium*

Abstract

The bootstrap method has been applied to estimate error bounds on five goodness-of-fit indicators in vibrational spectroscopy. The use of these indicators has been first tested on the known vibrational circular dichroism spectrum of 3-chloro-1-butyne, and then it has been adopted for the assignment of the absolute configuration of 3-methyl-3-nitromethyl-isoindolinone. The (+) stereoisomer turns out to have the (*S*) configuration.

1. Introduction

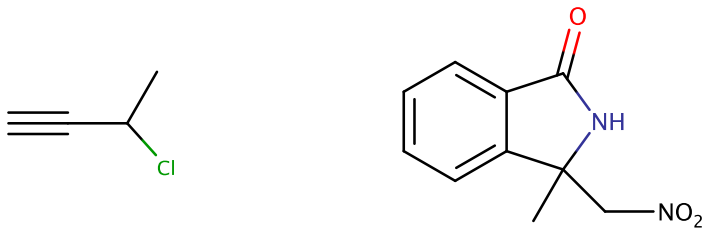
When prior information grants that there are few possible alternatives for the correct model, it can be sufficient to select the best of them, rather than entering the much more complicate endeavour of devising a model that is fully consistent with the data within the limits of experimental accuracy. The presence of this prior information is a standard in the assignment of the absolute configuration (AC) by vibrational circular dichroism (VCD) [1–3]. Typically, the spectra of the stereoisomers compatible with prior chemical information are computed by quantum-mechanical methods, and the AC is assigned to the stereoisomer whose spectrum gives the best visual match with the experimental one. Although this approach can nowadays be considered routine in many cases, the presence of more than a stereocenter, the occurrence of more conformations, or the difficulty in recording high signal-to-noise ratio spectra can considerably complicate the task.

*gmonaco@unisa.it

The combination of several experimental techniques has often been used in these cases [4], especially when relative configuration is also to be assessed [5]. However, additional techniques can be useless, as in the case of low optical rotations, which are very difficult to compute [6]. Sticking on VCD spectra, efforts have been done to develop techniques and software to cope with non-trivial situations. A first suggestion was to exclude from the comparison those peaks, called non-robust, which can change sign by small perturbations in experiment or computation [7–11]. A different and possibly complementary solution requires the introduction of numerical comparisons of computed and measured spectra. To this end several Goodness-Of-Fit Indicators (GOFIs) have been proposed. Rather than the usual mean squared error (MSE), which is notoriously biased by outliers, robust GOFIs have been suggested, like similarity indices [9, 12, 13], which for point-wise un-broadened spectra coincide with the cosine similarity (COSI) [14]. More recently, it has been proposed to estimate the uncertainties of the computed spectra to generate model-averaged (MA) spectra that should be comparable with experiment even using the non-robust MSE [15]. These approaches are promising to offer an ever more automated procedure for selecting the correct AC, which is a relevant target for pharmaceutical industry [16]. Yet, the final decision of considering meaningful the difference of the GOFIs computed for two stereoisomers can be delicate, particularly for small differences. To this end two tools have already been devised: i) a database of assignments has been assembled and an indicator of confidence level for new assignments has been derived there from [13], and ii) the robustness of a similarity index has been assessed via an approach conceptually based on randomization tests [17].
In this paper we will introduce estimates of the uncertainties of several GOFIs by the bootstrap method [18, 19], and we will apply them in conjunction with the MA approach. After a short description of the MA method and the bootstrap calculations, we will first apply them to 3-chloro-1-butyne **1**, which is a small rigid molecule of known AC, and we will then give results for the recently synthesized [20] (+)-3-methyl-3-nitromethyl-isoindolinone **2**, whose AC is reported here for the first time.

2. Theory

Our goal is the selection of the best stereochemical model $\mathcal{M}_{\mathrm{b}est}$ out of the possible stereoisomers. Each stereoisomer $\mathcal{M}$ will come with a computed spectrum $y_{i,\mathcal{M}}(\tilde{\nu}_i;\mathbf{P})$ which is a function of wavenumbers $\tilde{\nu}_i$, where



1 **2**

Scheme 1:

$i = 1, \ldots n$, and a vector of parameters $\mathbf{P}$, whose elements will be signed areas, central frequencies, and bandwidths of Lorentzian peaks, and, in case of more conformers, temperature and relative energies of conformers. The selection is to be done by comparison with the experimental spectrum $y_i(\tilde{\nu}_i)$. A typical GOFI for this procedure is the reduced chi-squared [21]

$$\chi^2_q := \frac{1}{q}\sum_{i=1}^{n}\frac{(y_i(\tilde{\nu}_i) - y_{i,\mathcal{M}}(\tilde{\nu}_i;\mathbf{P}))^2}{\sigma_i^2} = \frac{1}{q}\sum_{i=1}^{n}|z_i(\tilde{\nu}_i) - z_{i,\mathcal{M}}(\tilde{\nu}_i;\mathbf{P}))|^2 =: \text{MSE}, \tag{1}$$

which coincides with the mean squared error MSE [22] in terms of error-normalized dimensionless values $z_i = y_i/\sigma_i$ and $z_{i,\mathcal{M}} = y_{i,\mathcal{M}}/\sigma_i$, where σ_i is the standard deviation of the $i-$th difference $y_i(\tilde{\nu}_i) - y_{i,\mathcal{M}}(\tilde{\nu}_i;\mathbf{P})$. In eqn (1), q is the number of degrees of freedom. For models which are linear in p independent variables, with the p slopes determined from the experimental data by least squares, $q = n - p$ (in the limit case of a parameter-free model, $q = n$), and the expected value for MSE is 1. Values of MSE either significantly larger or smaller than 1 are considered unacceptable, indicating a bad model and/or the adoption of incorrect variances [21].

Although model selection via MSE is a well-established statistical procedure, this requires to deal with good models (MSE$\simeq$1), which is not the case for computed VCD spectra due to the impact of the basis set choice, functional and more generally all parameters $\mathbf{P}$ that influence the generation of the model spectrum. The selection among bad models (MSE$>>$1) risks to be meaningless, because it can be heavily affected by few outliers.

In our previous paper [15] we have discussed two workarounds to this prob-

lem.
First, one can consider the variance of the $i-$th difference $y_i(\tilde{\nu}_i) - y_{i,\mathcal{M}}(\tilde{\nu}_i;\mathbf{P})$ to have a contribution stemming from the calculation only, independently from the experiment. Following this line of thought, the model has been written as a function of a vector $\mathbf{p}$, composed by frequencies ($\tilde{\nu_0}$), magnetic and electric transition dipole moments (m and μ), angles between the magnetic and electric transition dipole moments (ξ), relative energies (enthalpies or free enthalpies) in case of more than one conformer), and a normal distribution has been assumed for each of the parameters, i.e.

$$f_j(p_j) = \frac{1}{\sqrt{2\pi}\sigma_{p_j}} \exp\left[-\frac{(p_j - p_{0,j})^2}{2\sigma_{p_j}^2}\right]. \tag{2}$$

It can be noted that the introduction of a distribution for the the central wavenumber $\nu_{0,k}$ corresponds to considering a family of peaks close to the peak of the plain calculation, and thus leads to broadened spectra, just as happens in the neighborhood similarity approach.[23, 24]
The variances on the parameters have been estimated from 8 computations ad different level of theory, used in previous successful works on VCD spectroscopy of small organic molecules: B3LYP/TZ2P,[25] B3LYP/cc-PVTZ,[13, 25, 26] B3LYP/6-31G*,[27] B3PW91/TZ2P,[25] B3PW91/cc-PVTZ,[25, 26] B97D/TZ2P,[28] ω-B97XD/6-31G*,[27] ω-B97XD/6-311++G**,[29] using in all cases the PCM model,[30] to model the effect of chloroform. Final standard deviations proposed were $\sigma(\tilde{\nu_0}) = 10$ cm^{-1}, $\sigma(\mu) = 2 \cdot 10^{-20}$ esu cm, $\sigma(m) = 1.5 \cdot 10^{-24}$ esu cm, $\sigma(\xi) = 10°$. Standard deviations of energies, enthalpies and free enthalpies were 0.05, 0.10 and 0.5 kcal mol^{-1}. Upon assumption of independent variation of the parameters, a model-averaged (MA) spectrum is computed as

$$y_{i,\mathrm{MA}-\mathcal{M}}(\tilde{\nu}_i;\mathbf{k}) = \int_{p_1,p_2\ldots p_p} y_{i,\mathcal{M}}(\tilde{\nu}_i,\mathbf{p};\mathbf{k}) \prod_{j=1}^{p} f_j(p_j)dp_j, \tag{3}$$

where a vector $\mathbf{k}$ of constants (temperature, average full-width at half maximum, scaling factors for harmonic frequencies and experimental absorbance) has been introduced. The variance of the computed spectrum is then

$$\sigma^2_{i,\mathrm{MA}-\mathcal{M}}(\tilde{\nu}_i;\mathbf{k}) = \int_{p_1,p_2\ldots p_p} [y_{i,\mathcal{M}}(\tilde{\nu}_i,\mathbf{p};\mathbf{k}) - y_{i,\mathrm{MA}-\mathcal{M}}(\tilde{\nu}_i;\mathbf{k})]^2 \prod_{j=1}^{p} f_j(p_j)dp_j. \tag{4}$$

Eq. 3 does not avoid the burden of performing more computations, as typically done in standardization of any novel computational protocol. However, if a parameters vector p_1^*, p_2^*, p_p^* is not formed by outliers within the set of parameters vector obtained by the computational methods examined, we can approximate Eq. 3 as

$$y_{i,\mathrm{MA}-\mathcal{M}}(\tilde{\nu}_i;\mathbf{k}) \simeq y_{i,\mathcal{M}}(\tilde{\nu}_i,\mathbf{p}^*;\mathbf{k}), \tag{5}$$

avoiding the need to perform the 8 computation, and resting on the need to have good estimates of the variances of the parameters to compute Eq. 4.

Eventually, we have considered the independence of model-averaging from the experimental measure to rewrite the variances in eq. 1 as

$$\sigma_i^2 = \sigma_{i,\mathrm{exp}}^2 + \sigma_{i,\mathrm{MA}-\mathcal{M}}^2, \tag{6}$$

The larger variances have been shown to lower the MSE values considerably, bringing them close to 1.

A second solution, that can be combined with the first, is the adoption of robust GOFIs, which are less sensible to outliers, as compared to the MSE. In ref. [15] we considered the mean absolute error

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|z_i - z_{i,\mathcal{M}}|, \tag{7}$$

and the cosine-similarity

$$\mathrm{COSI} = \frac{\mathbf{z}\cdot\mathbf{z}_{\mathcal{M}}}{||\mathbf{z}||\,||\mathbf{z}_{\mathcal{M}}||}. \tag{8}$$

In this paper we will also consider two further GOFIs. The first is the median of absolute deviations [19],

$$\mathrm{MAD} = \mathrm{Med}(|\mathbf{z} - \mathrm{Med}(\mathbf{z})\mathbf{1}|), \tag{9}$$

where Med indicates the median and $\mathbf{1}$ is a vector of ones the same size as $\mathbf{z}$. The second is the ratio of the mean absolute error over the mean of the absolute values of the $\mathbf{z}$ vector:

$$\mathrm{MMAR} = \frac{\mathrm{MAE}}{\mathrm{Mean}(|\mathbf{z}|)}, \tag{10}$$

If the elements of **z** are all non-negative, this GOFI coincides with the MAD/Mean ratio recently proposed in forecasting, and also with the discrepancy index R of widespread use in crystallography [31]. As we are interested in dichroic signals, and, especially in presence of strong couplets, the average z values can approach zero, we have modified those GOFI using the absolute value in the denominator, to get eq. 10.
In summary, we will consider these 5 GOFIs: RMSE, MAE, COSI, MAD and MMAR. As compared with competing models, the correct model should lead to smaller value for RMSE, MAE, MAD and MMAR and to larger values of COSI.
Although a concordant model selection using the different GOFIs is already an indication of a correct model selection, it is desirable to have error bounds on these GOFIs. To this end, we will use the bootstrap method [18, 19]. As applied to our problem, the method runs as follows. Any of the GOFIs is a function of random variables Z_1, Z_2,... Z_n, coming out of an unknown distribution F. The observed dataset $Z_1 = z_1$, $Z_2 = z_2$,... $Z_n = z_n$ is used to compute an estimate $\hat{F}$ for the unknown distribution function F. To do this, we give equal weight to all observed values, and we draw several n-elements samples from $\hat{F}$, such as $Z_1^* = z_1^*$, $Z_2^* = z_2^*$,... $Z_n^* = z_n^*$. As the elements of this bootstrap sample are selected with replacement, they will generally contain duplicates and will miss some of the data of the initial sample z_1, z_2,... z_n. Then, we replace the random sample Z_1, Z_2... Z_n from F by a random sample Z_1^*,Z_2^*, ... Z_n^* from $\hat{F}$, and approximate the probability distribution of GOFI(Z_1,Z_2, ... Z_n) by that of GOFI(Z_1^*,Z_2^*, ... Z_n^*). Eventually, drawing a large number of bootstrap samples, we can compute the bootstrap mean and standard deviation of the GOFI.

3. Materials and Methods

(*R*)-(+)-3-chloro-1-butyne **1** was synthesized from (*S*)-(−)-3-butyn-2-ol as described in literature [32, 33].
2 has been obtained by organocatalytic asymmetric nitro-aldol initiated cascade reactions of 2-acetylbenzonitrile with nitromethane ($[\alpha]_{20}^{D} = +21.7$) [20]. However, the enantioselectivity reached in this reaction was only moderate (up to 45% ee) and only after a reverse crystallization process (the crystallization as racemate is favored), we were able to improve the ee up to 87%. Despite the efficiency of this process, we performed several manipulations on this sample in order to further improve the enantiopurity and at the same

time to have suitable single crystals for x-ray structural analysis for the determination of AC. After these unsuccessful attempts, we observed a decrement of the amount of the enantioenriched sample. Instead of performing a new synthesis and crystallization, we challenged VCD analysis on the resulted tiny amount.

Vibrational Absorption (VA) and VCD spectra were recorded on a BioTools dual-PEM ChiralIR-2X spectrometer at room temperature. The PEMs were optimized for 1400 cm^{-1}, and a resolution of 4 cm^{-1} was used throughout. For measurements on **2**, solutions were prepared by dissolving 0.9 mg in 100 μL of $CDCl_3$. All spectra were recorded using BaF_2 windows and a spacer of 100 μm. The solution spectra were averaged over 100 000 scans, whose acquisition lasted 36 hours. Baseline corrections were introduced by subtracting the spectra of racemate and solvent.

Ab initio calculations have been performed by Gaussian 09 [34]. Conformers of **2** have been first guessed with Confab [35], and then optimized at the B3LYP/6-31G* level using the PCM method [30] to model the effect of chloroform.

4. Results and Discussion

The experimental VCD spectrum of 3-chloro-1-butyne is shown in Fig. 1, together with a fit with 6 Lorentzian lines. Observed peaks are well consistent with those reported in ref. [32], despite the change of solvent (chloroform instead of CCl_4). The fit has been used to estimate the experimental error σ_{exp}. In Fig. 2 the experimental spectrum of **1** is compared with its DFT computation at the B3LYP/6-31G* level for *R*-**1** (computed frequencies have been multiplied by 0.972 which was obtained from the absorption spectrum and is not very far from the value 0.9613 used in ref. [32]; full-widths at half maximum have been set at 7 cm^{-1}). For this simple molecule, inspection gives immediate confidence of the correctness of the assignment of the AC. Yet, the plain DFT model of the *R* configuration is not quantitatively correct as can be seen by the very large RMSE which is three orders of magnitude higher than its expected value 1 (Table 4). MA is also insufficient to have a quantitative agreement, although the RMSE becomes 4.10, and is thus much closer to the value expected for a correct stereochemical model. What we are here interested in is whether the difference of the GOFIs for the two models is significant. Notably for all the GOFIs the difference ΔGOFI=GOFI$_R$-

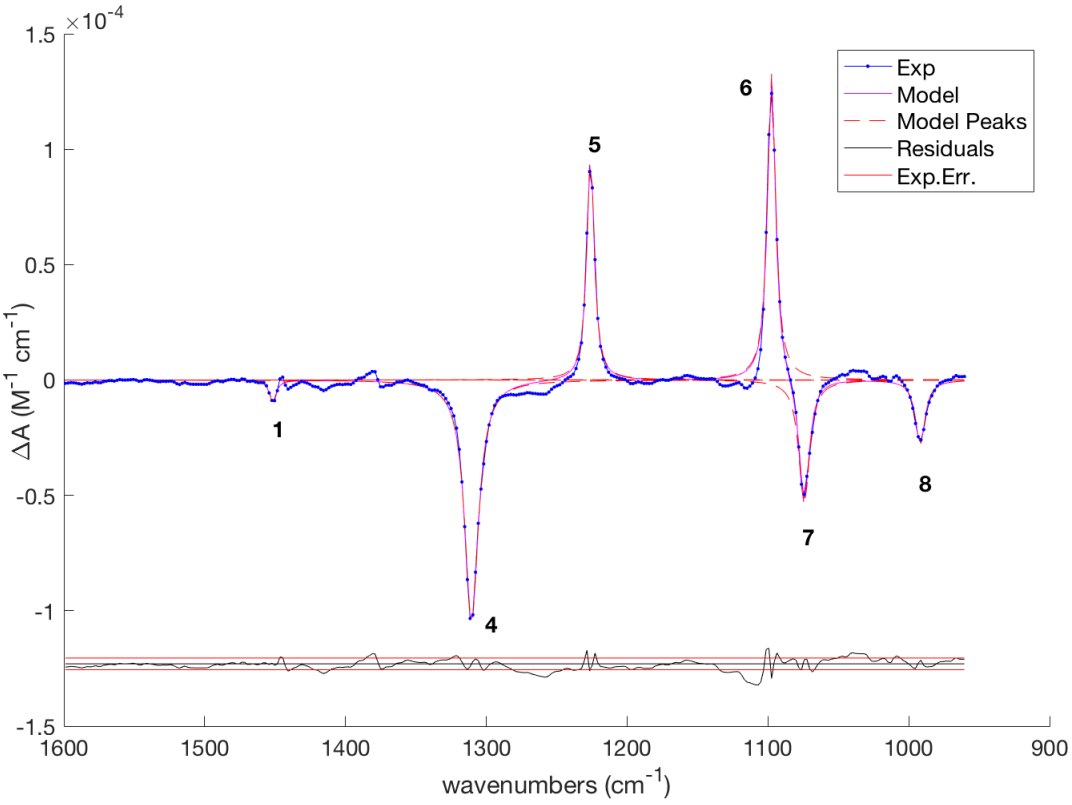


Figure 1: Experimental spectrum of **1** and its fit with 6 Lorentzian peaks. Numbering of peaks as in ref. [32]. Residuals are used to estimate the experimental standard deviation σ_{exp}.

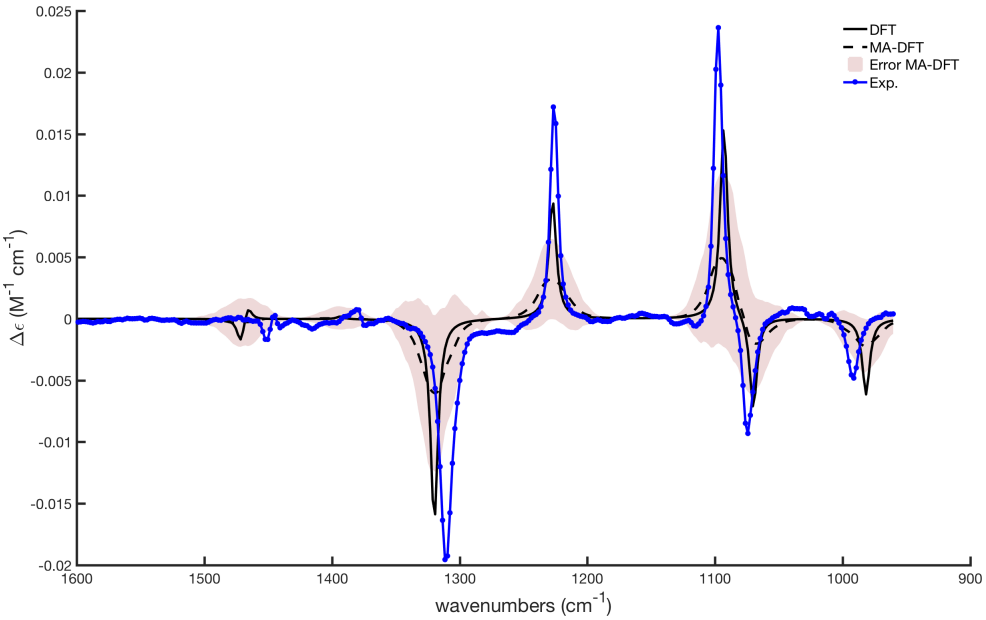


Figure 2: Experimental spectrum of (+)-**1** compared with its (*R*)-PCM-B3LYP/6-31G* model and its model-averaged version, which comes with an error estimate shown by the shading.

GOFI_S is significant, in the sense that the bootstrap estimate of its standard deviation is lower than the bootstrap mean. The single exception of the ΔMAD, which is found not significant ($s_b(\Delta\mathrm{MAD}) > \langle\mathrm{MAD}\rangle_b$) at the plain DFT level, is cured at the MA-DFT level. Thus, in this very simple case all GOFIs are able to select the correct AC of **1** at the MA-DFT level.

The VA and VCD spectra of **2** and their fit with 26 and 22 Lorentzian peaks, respectively, are shown in Fig. 3. Strong VA signals of the C=O and N=O stretching occur at 1712 and 1558 cm^{-1}. The VCD signal is very noisy, due to the low concentration. The Lorentzian guess has been built from the VA frequencies. Although few peaks have been removed from the fit because of their low intensities or large bandwidth, overfitting is unavoidable for such noisy spectra. Then, following our previous experience on poor experimental spectra [36], we have investigated on whether the frequencies of the strongest VCD peaks were also recognizable in the VA spectrum. This is certainly the case of the N=O stretching, which gives rise to a very strong dichroic signal. But also the second and third Lorentzian peaks in the strength-sorted list have a reasonable correspondence in the VA spectrum (Table 2), and they

	$\langle\cdot\rangle_b$	$s_b(\cdot)$	$\langle\cdot\rangle_b$	$s_b(\cdot)$	$\langle\Delta\cdot\rangle_b$	$s_b(\Delta\cdot)$
plain DFT	*S*		*R*		*S*-*R*	
RMSE	2068	214	1030	116	1038	243
MAE	872	108	459	51	413	119
MAD	144	14	136	14	8	20
MMAR	2.56	0.18	1.34	0.15	1.2	0.2
COSI	-0.66	0.05	0.66	0.04	-1.32	0.06
MA-DFT	*S*		*R*		*S*-*R*	
RMSE	8.17	0.06	4.10	0.03	4.07	0.07
MAE	2.92	0.02	1.542	0.011	1.37	0.03
MAD	0.389	0.002	0.372	0.003	0.017	0.004
MMAR	2.538	0.011	1.342	0.009	1.196	0.015
COSI	-0.652	0.003	0.652	0.003	-1.304	0.005

Table 1: Bootstrap values of means, $\langle\cdot\rangle_b$, and standard deviations, $s_b(\cdot)$, of the five GOFIs studied in the paper for the two possible AC of **1**, computed either by a plain PCM-B3LYP/6-31G* calculation or by its MA version. The last two columns of the table report the differences of the GOFIs computed for the *S* and *R* configuration, and an estimate of the error on the difference computed by standard error propagation: $s_b^2(\Delta\cdot) = s_b^2(\Delta\cdot_R) + s_b^2(\Delta\cdot_S)$.

can be used for AC assignment.

VA_{exp}		VCD_{exp}			Calculation			
$\tilde{\nu}$	*D*	$\tilde{\nu}$	*R*	$\tilde{\nu}$	$\tilde{\nu}$	*D*	*R*	#
1558	992	1557	-152	1633	1647	692	-102	56
1309	374	1313	-141	1372	1358	95	-50	46
1177	76	1172	131	1234	1231	45	11	42

Table 2: Experimental frequencies, dipole strengths, and rotational strength for **2** and their comparison with values calculated for *S*-**2-1** at the PCM-B3LYP/6-31G* level. Frequencies in cm^{-1}, dipole and rotational strengths in 10^{-40} and 10^{-44} esu^2 cm^2. The number of the normal mode (#) is also reported.

Conformational analysis of *S*-**2** gave the three conformers shown in Scheme 2, mainly differing for the dihedral angle **C**H_3-**C**(3)-**C**H_2-**N**O_2, where C(3) is the quaternary carbon on the isoindolinone. The dihedral angle **C**(3)-**C**H_2-**N**-**O** is close to 90 degrees in **2-2** and **2-3**, but deviates significantly in **2-1**, where a H-bond occurs between the N-H and the nitro groups (2.35 Å). Accordingly, **2-1** is the lowest energy conformer, **2-2** and **2-3** are computed

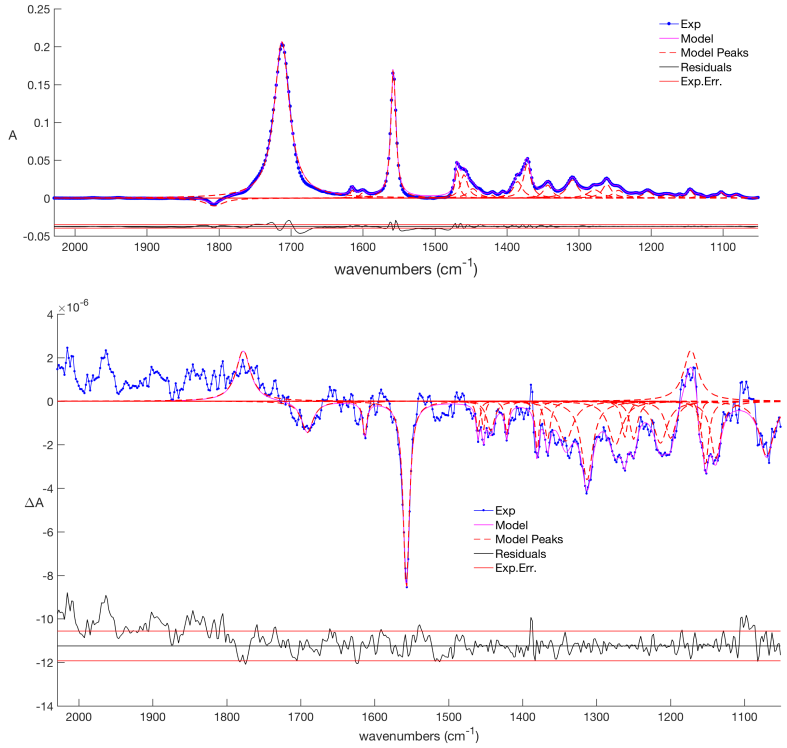
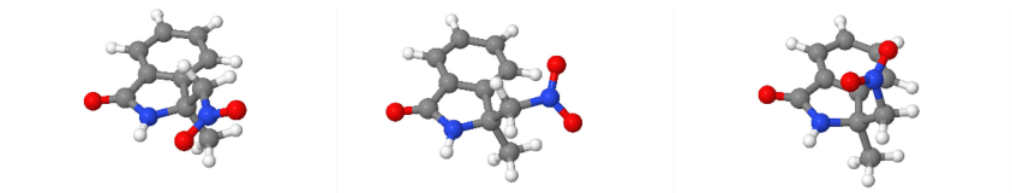


Figure 3: Experimental VA and VCD spectra of **2** and their fit with 26 and 22 Lorentzian peaks, respectively. Residuals are used to estimate the experimental standard deviation $\sigma_{exp}(A) = 2.3\ 10^{-3}$ and $\sigma_{exp}(\Delta A) = 6.8\ 10^{-7}$.

2.01 and 1.88 kcal mol^{-1} higher in energy. With such energy differences the **2-1** is the only relevant conformer; its enthalpy-based fractional abundance is 0.927. Optimization of cosine similarity of the experimental VA spectrum and the computed Boltzmann-weighted molar absorptivity led to a scale factor $\lambda = 0.954$, as a correction of the computed harmonic frequencies, and a scale factor $\kappa = 1.3\ 10^{-4}$ for the height of the computed spectrum. The computed VA and VCD spectra of **2-1** have three bands which reasonably match with the experimental values reported in Table 2. The match of signs of the three rotational strengths indicate that the AC is S. In order to have a more quantitative indication of the robustness of the assignment, we have computed the bootstrap values of the GOFIs discussed above together with their errors, for both the plain ab initio calculation and its MA version. Results are gathered in Table 4, while the MA spectra are compared with the experimental one in Figure 3. For the plain DFT calculation, the only meaningful GOFI is the COSI: only in this case the estimate of the error of the GOFIs and their differences are lower than the bootstrap expectation value. However, upon use of MA, all GOFIs are meaningful, and coherently indicate the S configuration.

Spurred by one of the reviewers, we have afforded a new synthesis of **2**, to obtain quantities sufficient to record of a VCD spectrum with a higher signal-to-noise ratio (Fig. 5). When subjected to non-linear least squares fitting in terms of sum of Lorentzians, the new spectra show 8 peaks with central frequencies in VA and VCD differing by less than 2 cm^{-1}. Of these 8 peaks, 7 have a reasonable correspondence in central frequency with the ω-B97XD/6-311++G** calculation and their rotational strengths is always consistent with the S stereochemistry (Table 4). Applications of MA approach to these spectra also confirms the assignment; in addition to the MA centered on the inexpensive SCRF-B3LYP/6-31G* calculation (as used above), we also considered centering on the original model-averaged parameters, Eq. 3. The numerical agreement is improved with respect to the previous spectrum. In all cases the assignment is firmly established as S, independently from the many features of the spectrum that are not well reproduced and should require computations at higher levels of theory. Among such subtle points, we not that the carbonyl region could probably benefit of the inclusion of explicit solvent in the calculation.[37]



Scheme 2:

	$\langle\cdot\rangle_b$	$s_b(\cdot)$	$\langle\cdot\rangle_b$	$s_b(\cdot)$	$\langle\Delta\cdot\rangle_b$	$s_b(\Delta\cdot)$
plain DFT	R		S		R-S	
RMSE	18697	962	17917	674	780	1175
MAE	14628	557	14428	504	200	751
MAD	12464	725	12136	615	329	951
MMAR	5.03	0.42	5.00	0.41	0.03	0.59
COSI	-0.07	0.04	0.07	0.04	-0.14	0.06
MA-DFT	R		S		R-S	
RMSE	2.072	0.005	1.990	0.004	0.082	0.006
MAE	1.546	0.003	1.525	0.003	0.021	0.004
MAD	1.204	0.004	1.169	0.004	0.035	0.006
MMAR	4.987	0.018	4.919	0.019	0.068	0.026
COSI	-0.074	0.002	0.074	0.002	-0.148	0.003

Table 3: Bootstrap values of means, $\langle\cdot\rangle_b$, and standard deviations, $s_b(\cdot)$, of the five GOFIs studied in the paper for the two possible AC of **2**, computed either by a plain PCM-B3LYP/6-31G* calculation or by its MA version. The last two columns of the table report the differences of the GOFIs computed for the R and S configuration, and an estimate of the error on the difference computed by standard error propagation: $s_b^2(\Delta\cdot) = s_b^2(\Delta\cdot_R) + s_b^2(\Delta\cdot_S)$.

ν_0^{VCD}	ν_0^{VA}	R_{exp}	D_{exp}	$\lambda\nu_0^{\mathrm{calc}}$	ν_0^{calc}	R_{calc}	D_{calc}
1102.4	1103.4	67	250	1110.2	1168.6	32	48
1130.2	1131.4	29	457	1131.2	1190.7	10	26
1161.1	1160.2	20	24				
1282.6	1280.7	19	154	1274.0	1341.0	15	370
1371.5	1371.7	47	604	1365.6	1437.5	40	430
1421.4	1419.7	-26	148	1406.3	1480.3	-4	310
1558.7	1557.4	-140	726	1574.0	1656.8	-94	1485
1705.6	1705.0	151	2218	1700.9	1790.4	25	2327

Table 4: The 8 peaks with a difference of fitted central frequency smaller than 2 cm^{-1}. The calculations are performed at the ω-B97XD/6-311++G** level. The scaling factor for the harmonic frequencies is $\lambda = 0.95$.

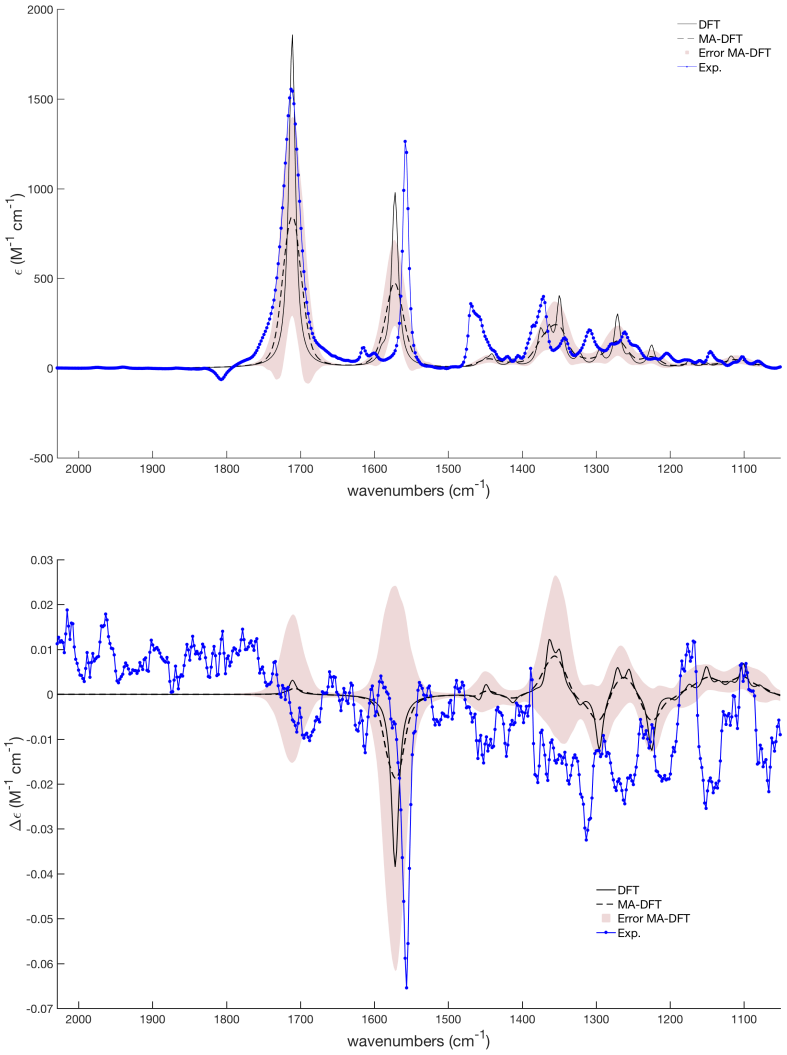


Figure 4: Experimental VA and VCD spectra of **2** compared with the plain PCM-B3LYP/6-31G* calculation and with their MA versions.

	$\langle\cdot\rangle_b$	$s_b(\cdot)$	$\langle\cdot\rangle_b$	$s_b(\cdot)$	$\langle\Delta\cdot\rangle_b$	$s_b(\Delta\cdot)$
plain DFT	*S*		*R*		*R-S*	
RMSE	9040	577	12000	840	2931	1019
MAE	5888	321	7618	440	1715	545
MAD	2816	349	3982	366	1149	506
MMAR	2.07	0.16	2.66	0.18	0.60	0.2
COSI	0.29	0.03	-0.29	0.03	-0.57	0.05
plain DFT	*S*		*R*		*R-S*	
RMSE	2.865	0.011	3.79	0.02	0.93	0.02
MAE	1.607	0.005	2.075	0.007	0.468	0.008
MAD	0.614	0.003	0.879	0.004	0.265	0.005
MMAR	2.052	0.008	2.649	0.010	0.597	0.013
COSI	0.286	0.002	-0.286	0.002	-0.572	0.003
plain DFT	*S*		*R*		*R-S*	
RMSE	7268	360	12500	992	5230	1055
MAE	4976	253	7928	443	2952	510
MAD	2416	204	4504	346	2088	402
MMAR	1.90	0.16	3.00	0.15	1.1	0.2
COSI	0.56	0.05	-0.56	0.05	-1.12	0.07
plain DFT	*S*		*R*		*R-S*	
RMSE	2.031	0.007	3.469	0.018	1.44	0.02
MAE	1.158	0.004	1.831	0.006	0.673	0.007
MAD	0.450	0.002	0.745	0.004	0.295	0.004
MMAR	1.889	0.008	2.984	0.009	1.096	0.012
COSI	0.561	0.003	-0.561	0.003	-1.122	0.004

Table 5: Bootstrap values of means, $\langle\cdot\rangle_b$, and standard deviations, $s_b(\cdot)$, of the five GOFIs studied in the paper for the two possible AC of **2**, computed either by a plain PCM-B3LYP/6-31G* calculation or by its MA version. The last two columns of the table report the differences of the GOFIs computed for the R and S configuration, and an estimate of the error on the difference computed by standard error propagation: $s_b^2(\Delta\cdot) = s_b^2(\Delta\cdot_R) + s_b^2(\Delta\cdot_S)$. The experimental spectrum with improved signal-to-noise ratio is the one of the newly synthesized sample (Fig. 5).

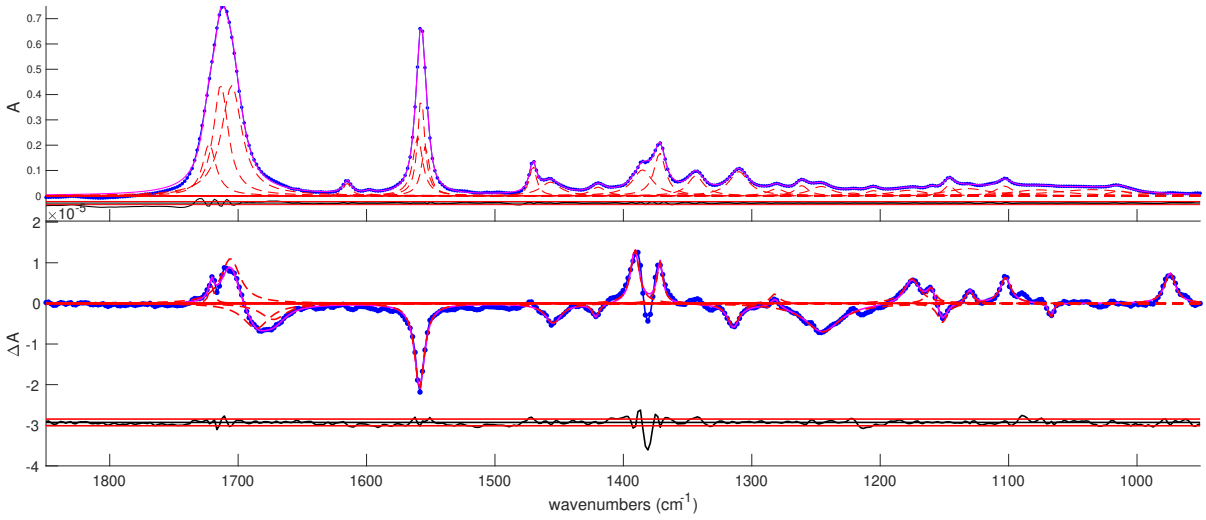


Figure 5: Experimental VA and VCD spectra of **2** compared with the plain PCM-B3LYP/6-31G* calculation and with their MA versions.

5. Conclusions

A goodness-of-fit indicator (GOFI) is generally needed to assess whether a computed VCD spectrum is better than a competitor spectrum to describe the experimental data. These GOFIs are generally used with no error bound in literature. This can be a limitation for poor spectra, such as the noisy spectrum of **2** here reported. In addition to the usual approach to scrutiny the spectrum for VCD resonances with a close match in the absorption spectrum, which pointed to the *S* configuration, we have also applied five different GOFIs with and without the use of the MA method. The error bounds on the GOFIs give additional support to the assignment. In this application, the COSI turned out the best GOFI, as it supports the assignment even in absence of the MA method. A second spectrum with a higher signal-to-noise ratio has fully confirmed the assignment of (S) confihuration for (+)3-methyl-3-nitromethyl-isoindolinone **2**.

Acknowledgments

Financial support from the MIUR is gratefully acknowledged.
GM is thankful to Prof. R. Zanasi for discussion.

[1] L. A. Nafie, Vibrational Optical Activity: Principles and Applications, John Wiley & Sons Inc., 2011.

[2] P. J. Stephens, F. J. Devlin, J. R. Cheeseman, VCD spectroscopy for organic chemists, CRC Press, Boca Raton, 2012.

[3] P. L. Polavarapu, Chiroptical Spectroscopy: Fundamentals and Applications, CRC Pr Inc, 2015.

[4] P. Polavarapu, Determination of the absolute configurations of chiral drugs using chiroptical spectroscopy, Molecules 21 (8) (2016) 1056. doi:10.3390/molecules21081056.

[5] C. Merten, T. P. Golub, N. M. Kreienborg, Absolute configurations of synthetic molecular scaffolds from vibrational CD spectroscopy, J. Org. Chem. 84 (14) (2019) 8797–8814. doi:10.1021/acs.joc.9b00466.

[6] P. Stephens, D. McCann, J. Cheeseman, M. Frisch, Determination of absolute configurations of chiral molecules using ab initio time-dependent density functional theory calculations of optical rotation: How reliable are absolute configurations obtained for molecules with small rotations?, Chirality 17 (S1) (2005) S52–S64. doi:10.1002/chir.20109.

[7] V. P. Nicu, E. J. Baerends, Phys. Chem. Chem. Phys. 11 (29) (2009) 6107.

[8] S. Góbi, G. Magyarfalvi, Phys. Chem. Chem. Phys. 13 (36) (2011) 16130.

[9] C. L. Covington, P. L. Polavarapu, J. Phys. Chem. A 117 (16) (2013) 3377–3386.

[10] G. Longhi, M. Tommasini, S. Abbate, P. L. Polavarapu, Chem. Phys. Lett. 639 (2015) 320–325.

[11] V. P. Nicu, Phys. Chem. Chem. Phys. 18 (31) (2016) 21213–21225.

[12] T. Kuppens, K. Vandyck, J. van der Eycken, W. Herrebout, B. van der Veken, P. Bultinck, Spectrochim. Acta, Part A 67 (2) (2007) 402–411.

[13] E. Debie, E. De Gussem, R. K. Dukor, W. Herrebout, L. A. Nafie, P. Bultinck, ChemPhysChem 12 (8) (2011) 1542–1549.

[14] A. Singhal, IEEE Data Eng. Bull. 24 (4) (2001) 35–43.

[15] G. Monaco, F. Aquino, R. Zanasi, W. Herrebout, P. Bultinck, A. Massa, Phys. Chem. Chem. Phys. 19 (41) (2017) 28028–28036.

[16] H.-U. Blaser, G. Hoge, B. Pugin, F. Spindler, in "Handbook of Green Chemistry" Vol I cap. 7, (2009) P. T. Anastas Ed. Wiley.

[17] J. Vandenbussche, P. Bultinck, A. K. Przybył, W. A. Herrebout, J. Chem. Theory Comput. 9 (12) (2013) 5504–5512.

[18] B. Efron, Ann. Stat. 7 (1) (1979) 1–26.

[19] F. M. Dekking, C. Kraaikamp, H. P. Lopuhaa, L. E. Meester, A Modern Introduction to Probability and Statistics: Understanding Why and How, Springer Nature, 2007.

[20] F. Romano, A. D. Mola, L. Palombi, M. Tiffner, M. Waser, A. Massa, Catalysts 9 (4) (2019) 327.

[21] P. Bevington, D. K. Robinson, Data Reduction and Error Analysis for the Physical Sciences, McGraw Hill Book Co, 2002.

[22] W. Dennis O. Wackerly, R. L. S. III Mendenhall, Mathematical Statistics with Applications, Cengage Learning, Inc, 2007.

[23] R. de Gelder, R. Wehrens, J. A. Hageman, J. Comput. Chem. 22 (3) (2001) 273–289.

[24] T. Kuppens, K. Vandyck, J. Van der Eycken, W. Herrebout, B. J. van der Veken, P. Bultinck, J. Org. Chem. 70 (23) (2005) 9103–9114, 00036.

[25] P. J. Stephens, F. J. Devlin, J.-J. Pan, Chirality 20 (5) (2008) 643–663.

[26] A. Krief, M. Dunkle, M. Bahar, P. Bultinck, W. Herrebout, P. Sandra, J. Sep. Sci. 38 (14) (2015) 2545–2550.

[27] S. Qiu, E. De Gussem, K. Abbaspour Tehrani, S. Sergeyev, P. Bultinck, W. Herrebout, J. Med. Chem. 56 (21) (2013) 8903–8914.

[28] A. Massa, P. Rizzo, G. Monaco, R. Zanasi, Tetrahedron Lett. 54 (46) (2013) 6242–6246.

[29] F. Cherblanc, Y.-P. Lo, E. De Gussem, L. Alcazar-Fuoli, E. Bignell, Y. He, N. Chapman-Rothe, P. Bultinck, W. A. Herrebout, R. Brown, H. S. Rzepa, M. J. Fuchter, Chem. Eur. J. 17 (42) (2011) 11868–11875.

[30] J. Tomasi, B. Mennucci, R. Cammi, Chem. Rev. 105 (8) (2005) 2999–3094.

[31] International tables for X-Ray Crystallography. Volume II - Mathematical Tables. Kasper, J. S. and Lonsdale, K.; 2d Ed. The Kynoch Press, 1967, Birmingham, 1967, p. 332.

[32] J. He, A. Petrovich, P. L. Polavarapu, J. Phys. Chem. A 108 (10) (2004) 1671–1680.

[33] F. Munyemana, A.-M. Frisque-Hesbain, A. Devos, L. Ghosez, Tetrahedron Lett. 30 (23) (1989) 3077–3080.

[34] Frisch, M. J. et al. Gaussian 09, Revision D.01, 2013, Wallingford CT, Gaussian, Inc.

[35] N. M. OBoyle, T. Vandermeersch, C. J. Flynn, A. R. Maguire, G. R. Hutchison, J. Cheminf. 3 (1).

[36] A. Massa, P. Rizzo, F. Scorzelli, G. Monaco, R. Zanasi, J. Pharm. Biomed. Anal. 144 (2017) 52–58.

[37] E. Debie, P. Bultinck, W. Herrebout, B. van der Veken, Solvent effects on IR and VCD spectra of natural products: an experimental and theoretical VCD study of pulegone, Physical Chemistry Chemical Physics 10 (24) (2008) 3498. doi:10.1039/b801313f.