Experimental and Computational NMR Studies of Large Alkaloids Exemplified with Vindoline Trimer: Advantages and Limitations

Valentin A. Semenov, Sergey V. Zinchenko, G. Massiot, Leonid B. Krivdin

Experimental and Computational NMR Studies of Large Alkaloids Exemplified with Vindoline Trimer: Advantages and Limitations

Valentin A. Semenov,[1,*] Sergey V. Zinchenko,[1] Georges Massiot[2] and Leonid B. Krivdin[1]

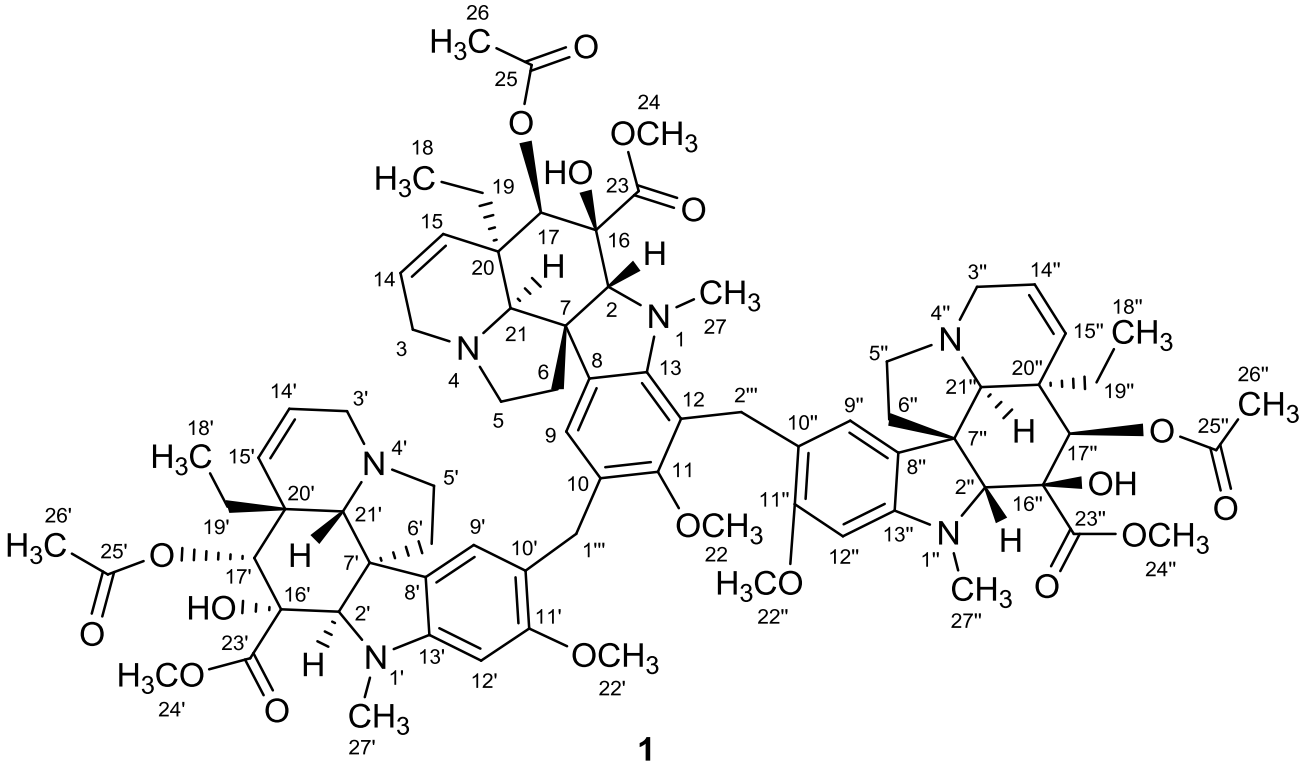
[1] A. E. Favorsky Irkutsk Institute of Chemistry, Siberian Branch of the Russian Academy of Sciences, Favorsky st. 1, 664033 Irkutsk, Russia.
[2] Institut de Chimie Moléculaire de Reims, UMR CNRS 7312, Université Reims-Champagne-Ardenne, UFR Sciences, BP 1039, CEDEX 2, 51687, Reims, France
***** Correspondence: semenov@irioch.irk.ru

Abstract: The complete ^{1}H and ^{13}C NMR assignments of a trimeric vindoline together with its individual components, dimeric vindolicine and monomeric vindoline, are performed based on a thorough analysis of the ROESY, COSY, HSQC, and HMBC spectra in combination with the state-of-the-art quantum-chemical calculations. A spatial structure of vindoline trimer is determined by means of computational conformational analysis in combination with the probability distribution map of its basic conformers. On the example of monoterpene indole alkaloid, the trimer vindoline, the present study reveals the power of modern computational NMR to perform identification and stereochemical studies of large natural compounds with some limitations, that may arise in the quantum chemical computing workflow.

Keywords: vindoline trimer; stereochemistry; ^{1}H and ^{13}C NMR spectra; computational NMR; DFT

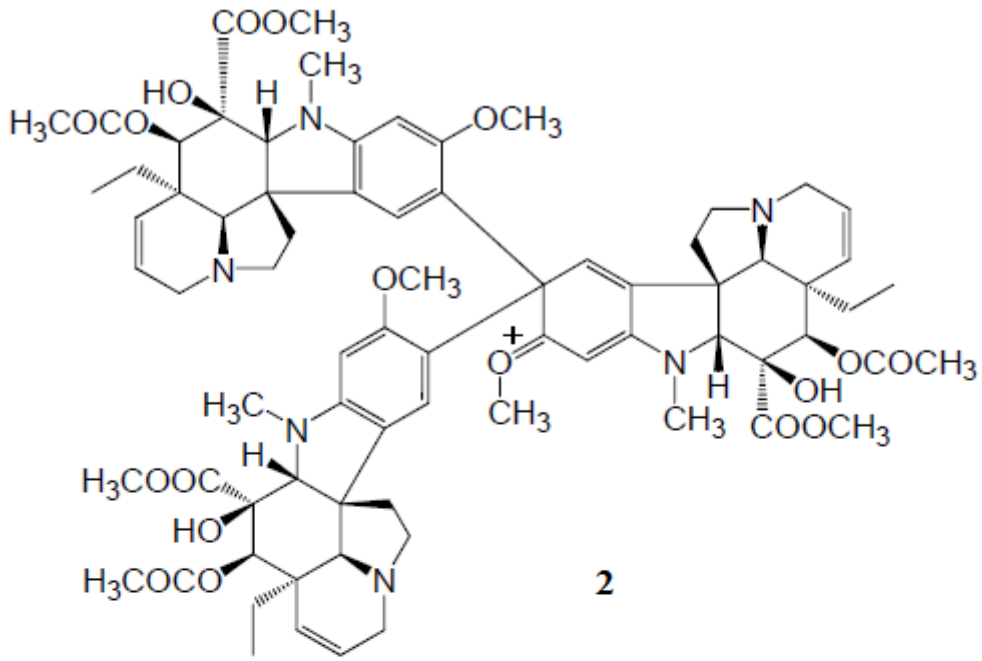
INTRODUCTION

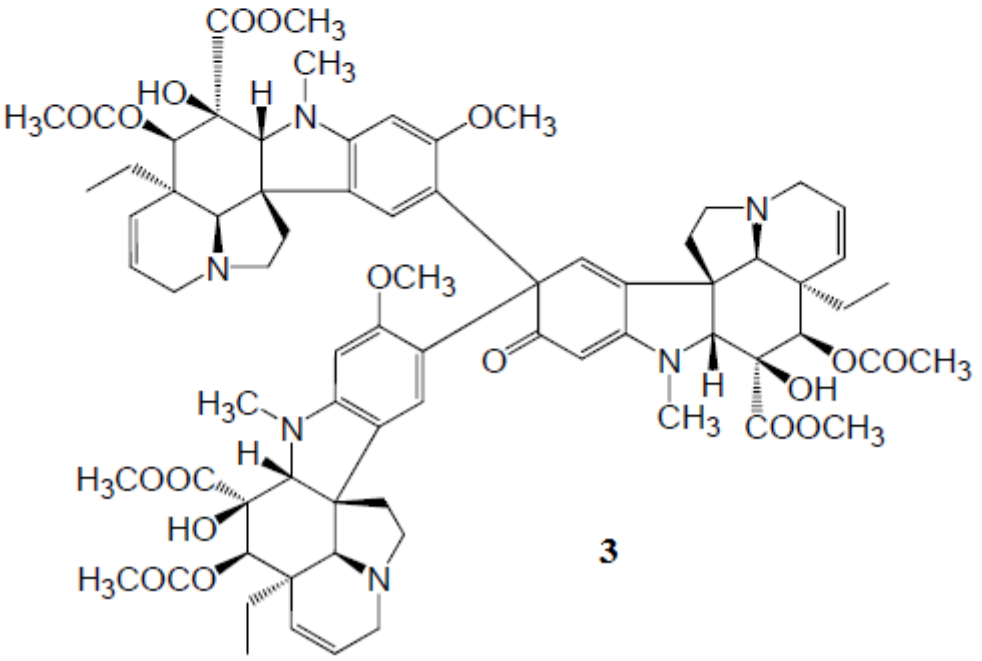
Monoterpene indole alkaloids (MIA) constitute an extensively studied class of compounds with over 3000 representatives found in the Apocynaceae, Loganiaceae and Rubiaceae plant families.[1] They owe their success to a wide variety of biological properties, and to the complexity of their structures, progresses of which went along with fundamental advances in organic synthesis. Among these substances, one can cite strychnine, yohimbine of a former wide use, and camptothecin and vinblastine, which found place on the market as anticancer drugs. Most of the MIA exist as "monomers", which consist of a tryptamine unit and of a terpenic part derived from secologanosic acid. A small percentage of the MIA are found as "dimers" formed by the coupling of two "monomers" with a wide variety of modes of attachment. There should be no reason why "trimers" should be absent, and a recent publication lists *ca* ten of these "trimers" with not including the oligomeric *Psychotria* alkaloids.[2] Solving the structures of so large molecular architectures and assigning of their NMR spectra is quite challenging, especially when the trimers are made of similar units. The synthesis of several of these large molecules was devised by one of us (GM) and by a Hungarian group as a means to find new uses for vindoline the readily available alkaloid from *C. roseus*.[3,4,5,6,7] Scheme 1 shows the structure of one of the newly synthesized vindoline trimers.



Scheme 1. The structure of vindoline trimer (**1**) reported by Asia *et al*.[4]

Compounds of similar molecular weight range were also obtained by Keglevich and coauthors[7] in continuation of a number of publications related to the reactivity of vindoline, Scheme 2.[5,6] Thus trimers **2** and **3** were surprisingly obtained during an attempted oxidation of vindoline being characterized by their ^{1}H and ^{13}C NMR spectra.





Scheme 2. Two vindoline trimers, extracted and described by Keglevich *et al*.[7]

Nowadays we are witnessing a marked progress in the calculation of NMR chemical shifts of a variety of different chemical and biological molecules based on the modern advanced levels of theory.[8,9] Computation of NMR parameters provides a new guide in the understanding of the fundamental factors controlling stereochemistry

and chemical reactivity of a variety of bioorganic molecules, being on the cutting edge of modern computational chemistry.

Calculations of NMR chemical shifts of natural products like alkaloids with multiple asymmetric centers are performed nowadays within the DFT framework, in contrast to the non-empirical computations applied to much smaller molecules. This is not surprising in view of the fact that at the DFT level electron correlation is involved in an implicit way, so that such calculations are much more economic, as compared to the non-empirical methods, the latter taking into account electron correlation explicitly. From this point of view, to date, many different approaches and add-ons to the existing ones have been developed to improve the accuracy of calculations within the DFT formalism. A number of methods based on determining the probability distribution of a set of stereoisomers like DP4,[10,11,12] DP4+,[13] (ML)DP4-*J*,[14,15] DP5[16] are widely employed in stereochemical analysis of complex organic molecules like alkaloids and other natural products being much more robust as compared to classical statistical methods. All of them are derived from Smith and Goodman's CP3 approach, which is based on Bayes' theorem.[17] Among other approaches, machine learning methods are also extensively used in routine quantum chemical calculations of magnetic resonance parameters.[15,18,19,20,21,22,23,24]

However, it should be kept in mind that results of the DFT calculations drastically depend on the choice of a particular functional. In this study, all calculations of ^{1}H and ^{13}C NMR isotropic magnetic shielding constants (and, accordingly, chemical shifts) were carried out at the DFT level in liquid phase with the most reliable for this purpose PBE0 functional being compared to available experiment. In doing so, we used our recently proposed pecS-2 basis set, which was specially developed for calculating shielding constants in large synthetic compounds and natural products.[25,26] This basis set makes it possible to achieve better agreement with experimental data at acceptable computational costs.

Computational NMR rapidly develops in parallel with a marked progress in general structural studies of bioorganic and biological molecules, as well as natural products. Total synthesis of natural products is one of the most challenging and

exciting areas of bioorganic chemistry, imitating nature's ability to create diverse molecular structures. At that, the development of complex multistage syntheses is associated with significant material and time costs. In this line, of special attention is a recent renaissance of computational NMR of those species.[27,28]

In this paper, we performed a complete ^{1}H and ^{13}C NMR assignment of huge trimeric vindoline alkaloid based on a thorough analysis of its ROESY, COSY, HSQC, and HMBC spectra in combination with a state-of-the-art quantum-chemical calculations required to construct its probable spatial structure. Spectral assignments of ^{1}H and ^{13}C NMR spectra were performed based on the established ROESY, COSY, HSQC, and ^{1}H-^{13}C HMBC correlations presented in Figures 1-3.

RESULTS AND DISCUSSION

Vindoline trimer (**1**) provides a particular challenge for structural elucidation in terms of interpreting its ^{1}H and ^{13}C NMR spectra. This is primarily due to its difficult molecular structure, which has a large number of the degrees of freedom of its rotational isomers. The molecule of vindoline trimer consists of three similar molecular subunits being in different stereochemical environments, and containing a number of asymmetric centers. As a result, three groups of signals are observed in the NMR spectra of **1** with some peaks being pretty close, so that it was extremely difficult to perform their spectral assignments. Taken together, all of the above peculiarities make it practically impossible to establish spatial structure of vindoline trimer, as well as to perform unambiguous assignment of its ^{1}H and ^{13}C NMR signals. Under such circumstances, theoretical calculations of NMR chemical shifts of vindoline trimer performed in the present study, generally appeared to be essential providing unambiguous solution of those problems.

1. Experimental structural elucidation

All experimental ^{1}H and ^{13}C chemical shifts of **1** given in Table 1 were assigned based on ROESY, HSQC, and HMBC correlations illustrated in Figures 2 and 3. The details of those correlations are provided in Supporting Information in Table S1.

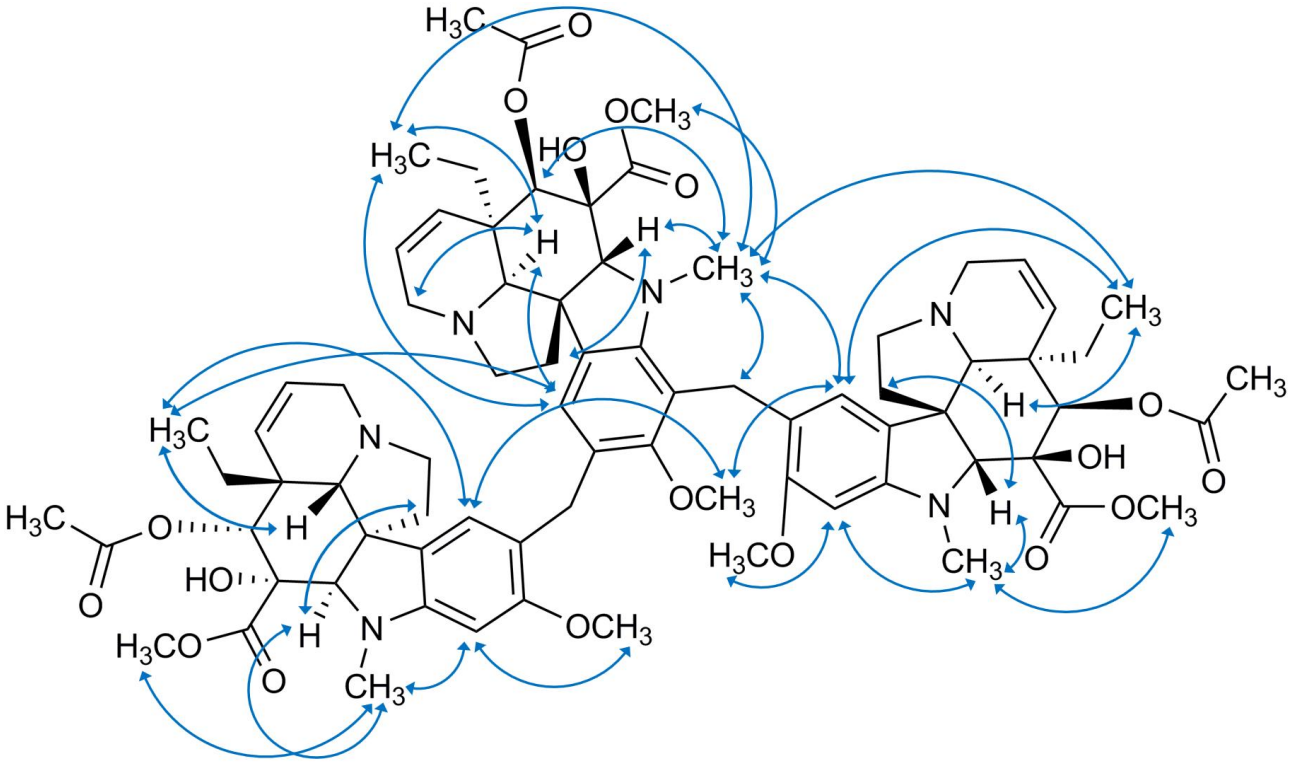


Figure 1. Principal ROESY correlations of vindoline trimer (**1**).

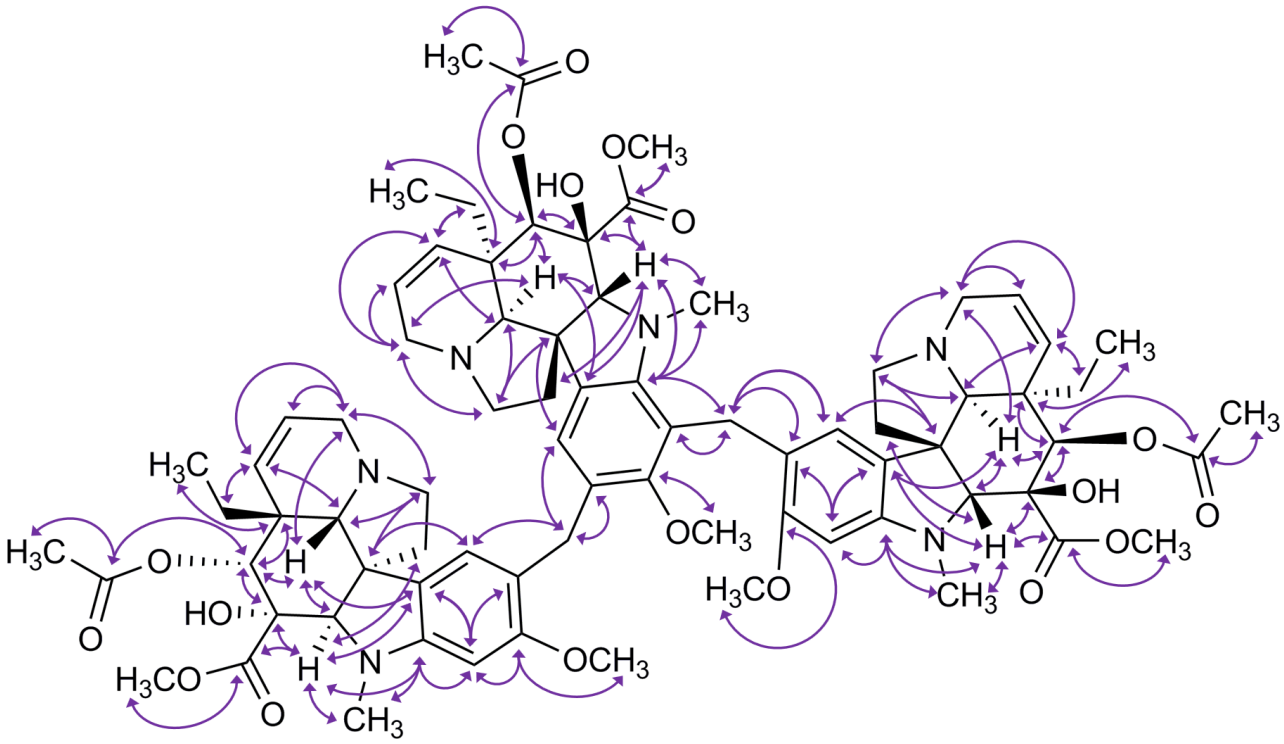


Figure 2. Principal ^{1}H-^{13}C HMBC correlations of vindoline trimer (**1**).

Table 1. Experimental ^{1}H and ^{13}C NMR data of vindoline trimer (**1**) ($CDCl_3$, rt).

	Unit 1			Unit 2			Unit 3	
	δ ^{1}H, ppm (m, J_{H-H}, Hz)	δ ^{13}C, ppm		δ ^{1}H, ppm (m, J_{H-H}, Hz)	δ ^{13}C, ppm		δ ^{1}H, ppm (m, J_{H-H}, Hz)	δ ^{13}C, ppm
2	3.56, s	84.69	**2'**	3.68, s	83.57	**2''**	3.66, s	83.35
3a	3.40, ddd, 16.5, 5.0, 1.5	50.99	**3'a**	3.47, ddd, 16.0, 4.8, 1.7	51.19	**3''a**	3.38, m	50.66
3b	2.74, ddd, 16.5		**3'b**	2.76, m		**3''b**	2.80, ddd, 16.0	
5a	2.45, m	51.27	**5'a**	2.43, m	52.00	**5''a**	3.33, ddd, 9.2, 5.2, 5.2	50.70
5b	3.38, m		**5'b**	3.37, m		**5''b**	2.54, m	
6a	2.16, m	43.63	**6'a**	2.29, m	44.11	**6''a**	2.16, m	44.25
6b	2.14, m		**6'b**	2.22, m		**6''b**	2.19, m	
7		53.17	**7'**		53.16	**7''**		53.55
8		129.06	**8'**		123.85	**8''**		122.83
9	6.61, s	121.85	**9'**	6.74, s	124.24	**9''**	6.49, s	122.54
10		127.41	**10'**		120.57	**10''**		121.11
11		159.01	**11'**		158.79	**11''**		158
12		119.23	**12'**	6.06, s	93.89	**12''**	6.09, s	93.18
13		151.48	**13'**		152.16	**13''**		151.55
14	5.83, ddd, 10.2, 4.9, 1.7	124.41	**14'**	5.83, ddd, 10.2, 4.9, 1.7	124.28	**14''**	5.81, ddd, 10.2, 5.0, 1.6	124.28
15	5.26, ddd, 10.2, 2.5, 1.8	130.77	**15'**	5.22, ddd, 10.2, 2.5, 1.7	130.6	**15''**	5.22, ddd, 10.2, 2.6, 1.6	130.86
16		80.34	**16'**		79.72	**16''**		80.11
17	5.30, s	76.19	**17'**	5.44, s	76.40	**17''**	5.34, s	76.61
18	0.39, t, 7.4	7.87	**18'**	0.44, t, 7.4	8.30	**18''**	0.24, t, 7.4	7.55
19a	1.49, dq, 7.4	30.68	**19'a**	1.10, dq, 7.4	30.93	**19''a**	1.01, dq, 7.4	30.37
19b	0.91, , dq, 7.4		**19'b**	1.62, dq, 7.4		**19''b**	1.46, dq, 7.4	
20		42.78	**20'**		43.05	**20''**		42.83
21	2.57, s	66.54	**21'**	2.58, s	66.97	**21''**	2.64, s	65.37
22	3.52, s	61.09	**22'**	3.73, s	55.64	**22''**	3.86, s	55.55
23		171.33	**23'**		172.03	**23''**		171.85
24	3.71, s	52.10	**24'**	3.78, s	52.35	**24''**	3.76, s	52.22
25		170.58	**25'**		170.92	**25''**		170.87
26	2.03, s	21.25	**26'**	2.06, s	21.18	**26''**	2.04, s	21.24
27	2.53, s	44.06	**27'**	2.65, s	38.93	**27''**	2.69, s	39.19
1'''a	3.70, d, 15.7	30.02						
1'''b	3.79, d, 15.7							
2'''a	3.87, d, 16.5	25.45						
2'''b	3.75, d, 16.5							

At the initial stage, to perform the assignment of ^{1}H and ^{13}C NMR spectra of **1**, it was important to identify the signals of protons and carbons at the positions 9 and 12.

Obviously, the position of CH-carbon C-9" is determined unambiguously by HMBC, since only the H-2" proton is correlated to C-9", while the H-1'" protons have cross-peaks with both carbons, C-9 and C-9'. The H-9' proton is uniquely determined by its long-range (across four bonds) correlation with C-12', and a similar cross-peak exists between H-9" and C-12", see Figure 3a. Further, having the assignment for atoms at positions 1'", 2'" together with that for H-9, H-12, C-9 and C-12 in all three tris-vindoline subunits, no problem arises with identifying both proton and carbon signals of aromatic fragments at C-8, C-10, C-11, and C-13.

The next step was the determination of protons and carbons in positions 2 and 21 by means of HMBC. The cross-peaks H-2–C-8 and H-2–C-13 as well as H-21–C-8, allowed performing this unambiguously. It is worth noting the almost complete coincidence of H-21 and H-21' peaks: for their assignment, a very good resolution in the proton spectrum (as well as in all 2D spectra) was required. In addition, the signals of C-7 and C-7' are almost identical, so that they could hardly be used for the assignment of H-2 and H-21, see Figure 3b. HMBC interactions of carbons C-14 and C-15 with protons H-3 is given in Figure 3c.

Certain difficulties aroused when assigning the signals of CH_2 groups in positions 3, 5 and 6 in all three subunits. This was due to the nature of the observed multiplets (all methylene protons are nonequivalent and highly splitted) and due to severe signal overlaps. The positions of all three H-3 signals were determined by their correlation with carbon C-15. Performed assignment of C-6 was based on the correlations of C-6 with H-2 and H-6 with C-8. The assignment of protons was further refined using the band-selective HSQC spectrum with f1 resolution of 2 Hz/pts and, independently, based on the H-21–H-6 correlations in the ROESY spectrum.

An interesting task was to determine the protons and carbons in position 5. The C-5' was easily identified through the correlation with a separately standing H-6' proton. Knowing the positions of all signals provided by C-6, all the high-field signals from H-5 were unambiguously determined by means of HMBC. After that, using the band selective HSQC spectrum, all carbons in the 5-th position were assigned following by the identification of the remaining low-field protons in the 5th position

with the two multiplets almost completely overlapped, see Figure 3d. The assignments of the O-CH_3 and N-CH_3 signals were just trivial.

The confirmation of the molecular configuration by means of the ROESY spectra of **1** was based on the presence of interactions H-2–H-6, H-2–H-27, H-21–H-18 and H-21–H-19, H-21–H-17 and the absence of such between H-2–H-21, H-2–H-18 and H-2–H-19. Fundamentally important for the evaluation of stereochemical structure and relative arrangement of three molecular subunits of **1** is the interaction of protons belonging to different subunits. Thus, the N-CH_3 protons "see" (meaning "interact with") the H-9" proton in the ROESY spectrum, indicating on the relative position of the V and V" subunits, see Figure 4a.

The characteristic through-space interaction of the methyl protons H-18 with H-9 and H-18 with H-18' provides the equal intensity cross-peaks with both H-9 and H-9', while H-18" interacts only with H-9", see Figure 4b. It is also worth noting the interaction of H-18" with the *N*-methyl group H-27. The H-9' and H-9" interact with the H-22 methyl group in the "foreign" subunit, while the H-9 proton provides no cross-peak with any of the H-22 protons, see Figure 4c.

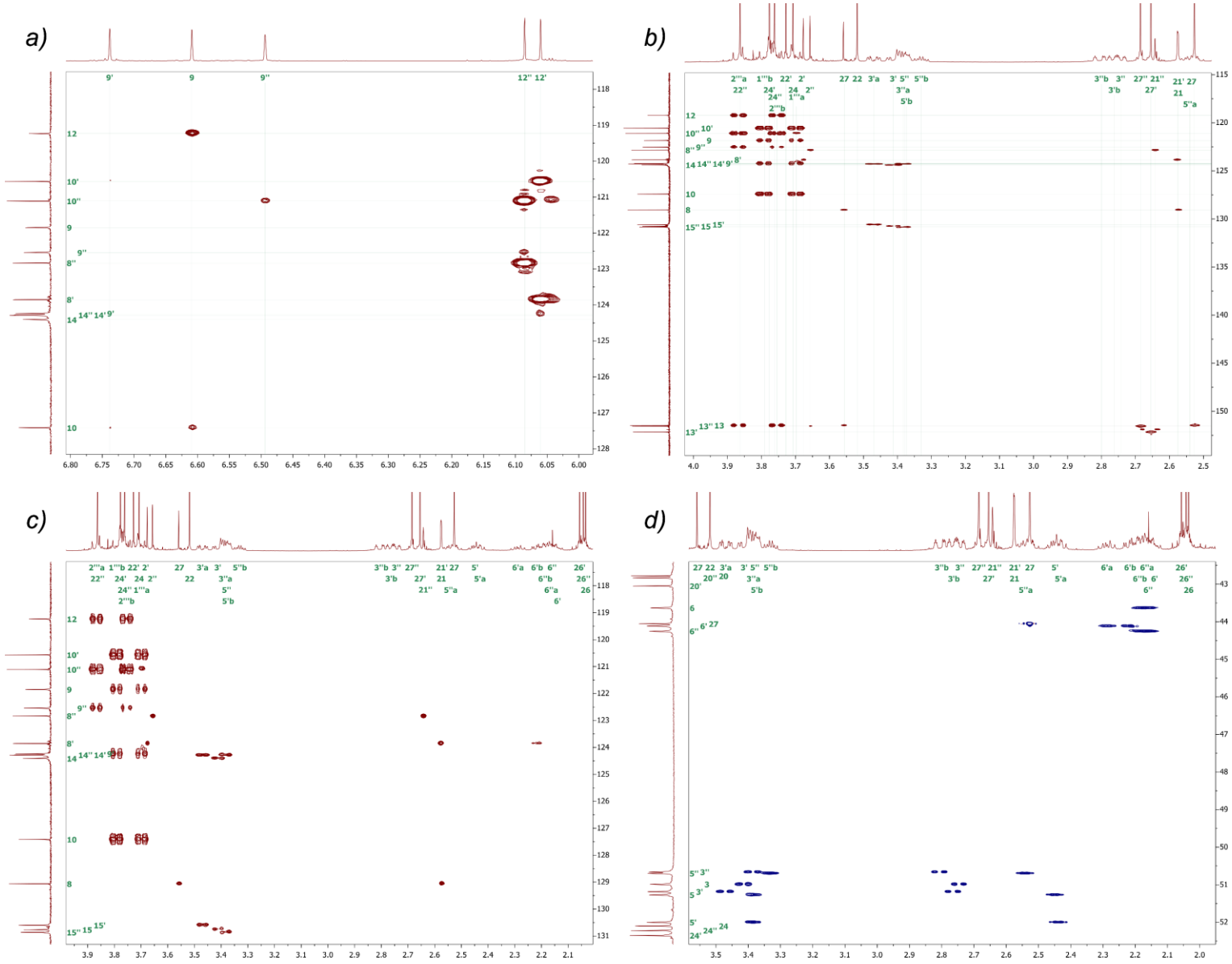


Figure 3. 2D HMBC correlations of **1** between: (*a*) - carbons C-9 and protons H-2; C-12 and H-9; (*b*) - carbons C-8 and C-13 with protons H-2 and H-21; (*c*) - carbons C-14 and C-15 with protons H-3; (*d*) - the region of 2D HSQC spectrum with correlation for carbons C-3, C-5 and C-6.

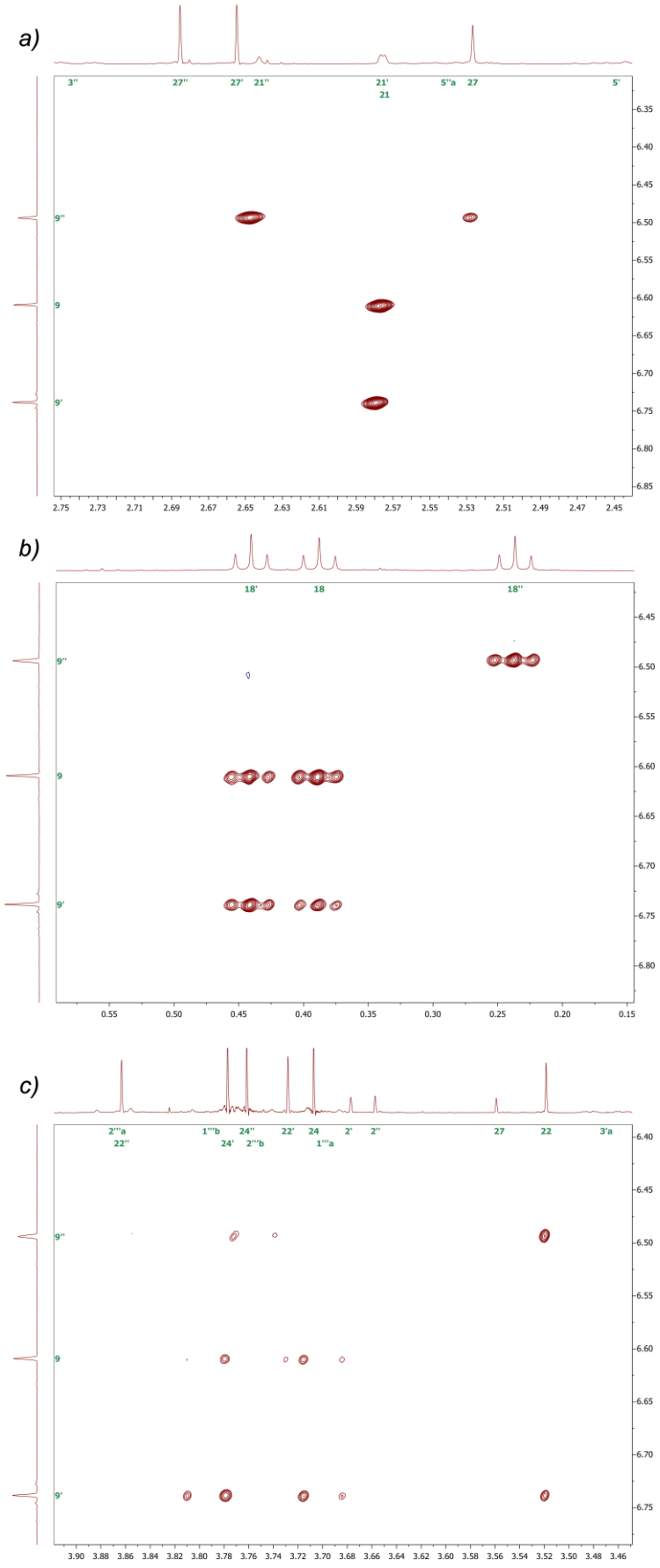


Figure 4. 2D ROESY correlations of **1** between protons: (*a*) - H-9 and H-27; (*b*) - H-18 and H-9; (*c*) - H-9 and H-22.

2. Computational stereochemical study

Conformational analisys. At the first stage of this study, to establish the three-dimensional structure of vindoline trimer (**1**), a preliminary conformational search was initially carried out. The structural features of the location of three subunits V, V' and V" relative to each other were carefully established. When constructing stereochemical model, the available ROESY correlations presented in Figure 1 were taken into account whenever possible. Thus, the conformational search was mainly reduced, on the one hand, to the detection of exact rotational angles along the C-10–C-1'", C-10'–C-1'", C-12–C-2'", and C-10"–C-2'" and on the other hand – to the search for the accurate conformations of methoxycarbonyl and acetoxy fragments and their positions relative to the fused ring systems in the corresponding subunits. Thus, in the process of a conformational search in the energy range of 0-15 kcal/mol relative to the minimum, several groups of conformers were identified that properly corresponded to the stereochemical structure of the vindoline trimer, as was initially established using a series of NMR experiments (see Part 1).

As a result of molecular modeling, four groups of conformers **1a**-**1d** of the vindoline trimer were selected from all established groups, and they are shown in Figure 5. For these conformers, the following distribution of their relative free energies from **1a** to **1d** was evaluated: 0.0, 0.3, 0.6, and 1.3 kcal/mol, respectively. Next, geometric parameters of these four low-energy conformers were optimized at the M06-2X/cc-pVTZ//aug-cc-pVTZ level of theory in the liquid phase of chloroform-*d*. For the optimized structures, all shielding constants together with corresponding ^{1}H and ^{13}C NMR chemical shifts were calculated at the PBE0/pecS-2 level.

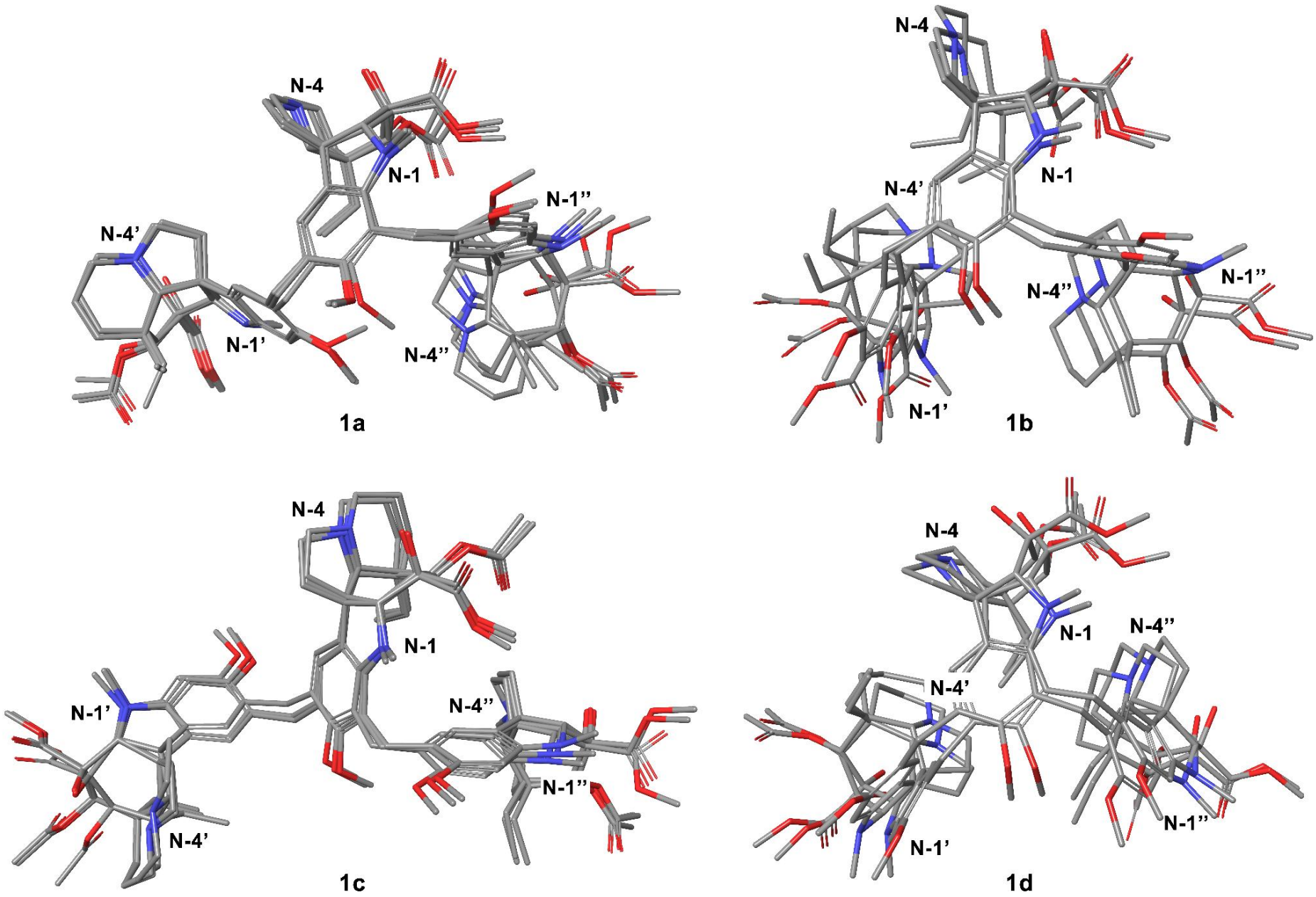


Figure 5. Three-dimensional structures of the main low-energy groups of conformers **1a**-**1d** of vindoline trimer, optimized at the M06-2X/cc-pVTZ//aug-cc-pVTZ level. Hydrogen atoms are not shown for clarity.

Calculation of NMR chemical shifts. At the next step, all calculated ^{1}H and ^{13}C NMR shielding constants were subjected to correlation analysis based on linear regression of the unscaled theoretical shielding constants and the corresponding experimental chemical shifts. For the scaled NMR chemical shifts calculated using equation (1), corrected mean absolute errors (CMAE) were estimated from equation (2). The most accurate correlation with experimental data was achieved for optimized structure of **1c**. In this case, we attempted to reconstruct a reliable stereochemical structure of the vindoline trimer in such a way that it would correlate as closely as possible with the experimental ROESY data. These main principal NOE correlations of the optimized conformer **1c** are presented in the Figure 6.

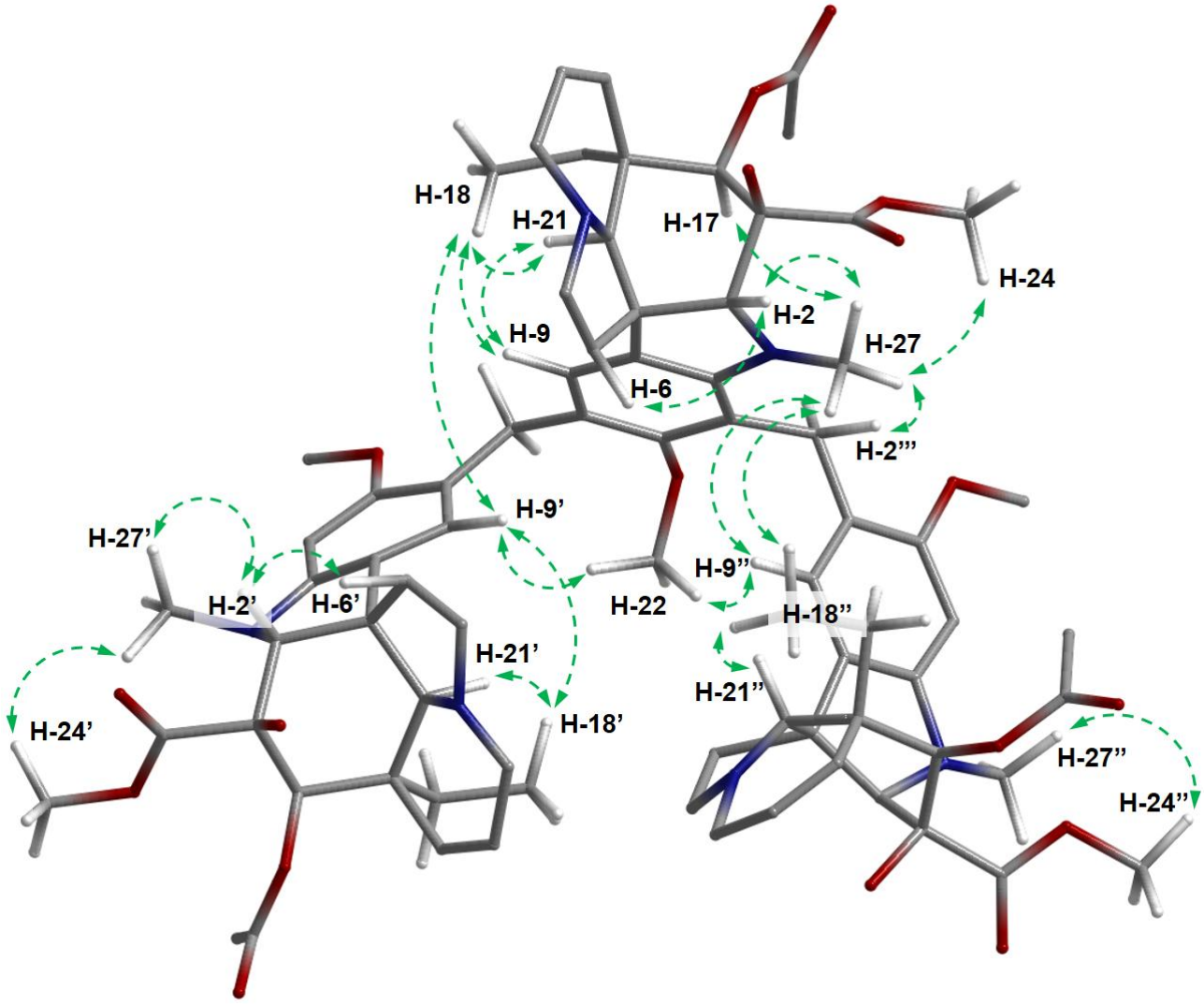
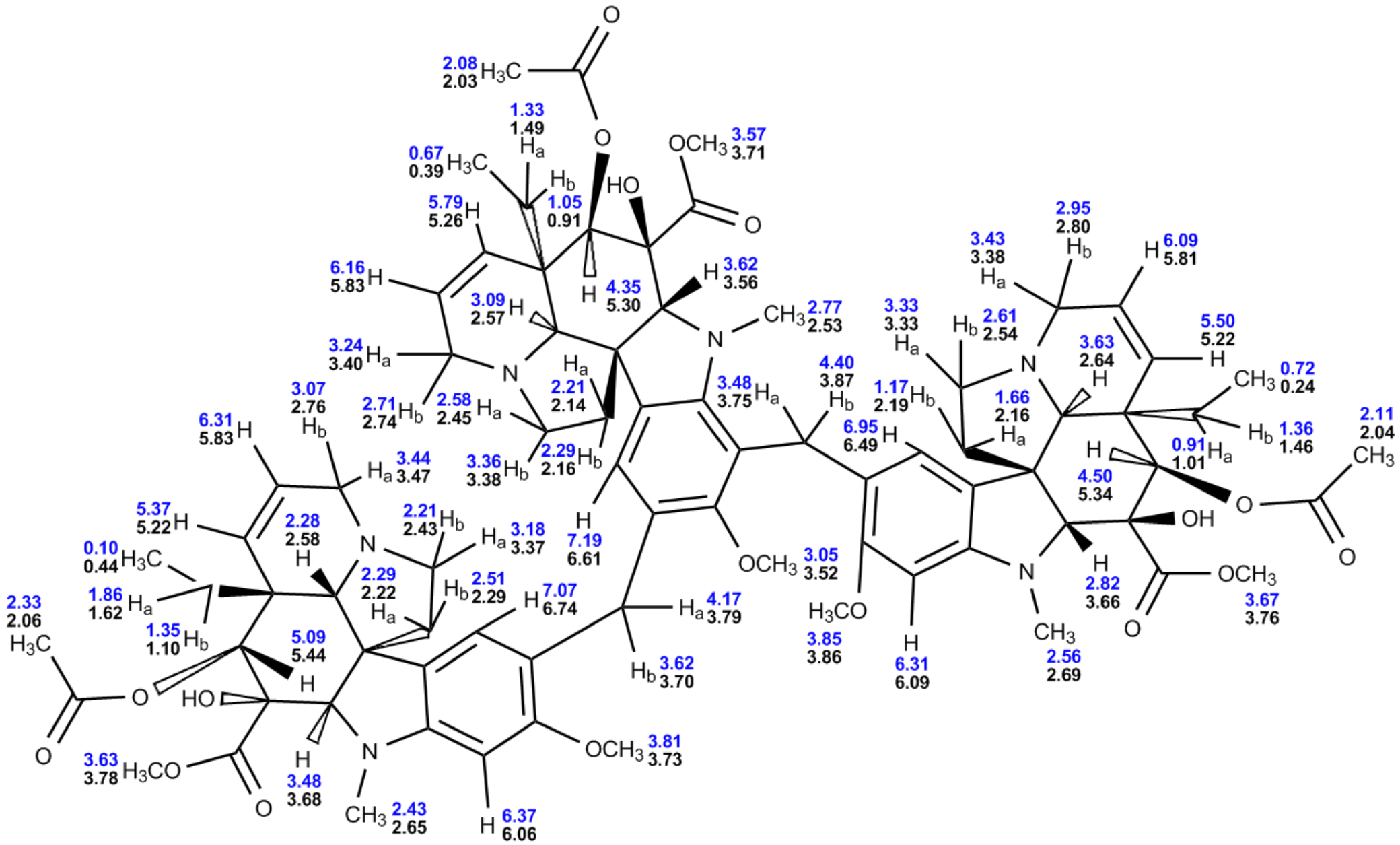
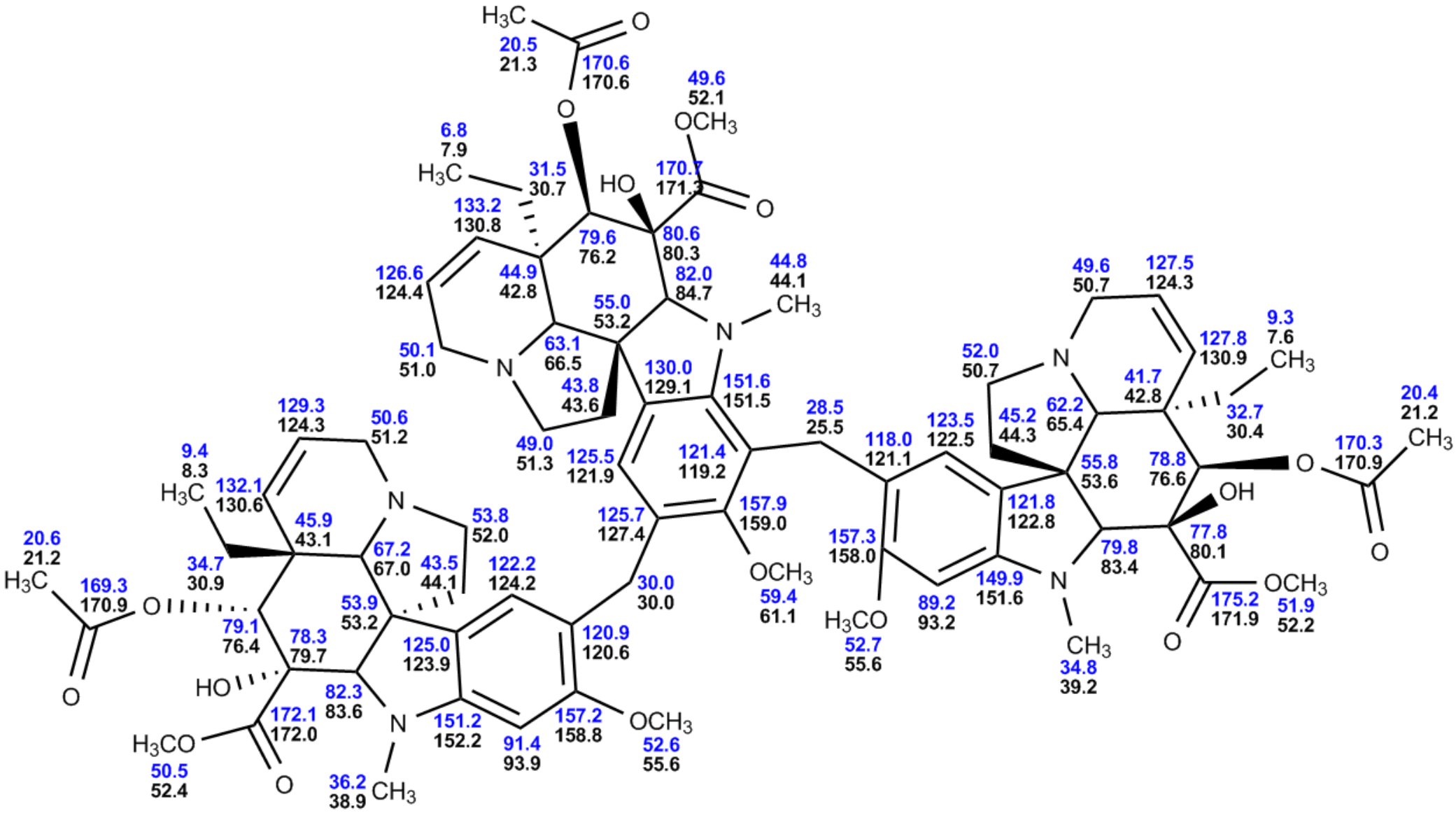


Figure 6. Key ROESY correlations of the optimized conformer **1c** of vindoline trimer.

Generally, a good agreement of calculated ^{1}H and ^{13}C NMR chemical shifts with experiment was established, which is seen in Schemes 3 and 4. Performed calculations as compared to experiment were characterized by the ranges of deviations of -0.99 to +1.02 and -5.03 to +4.43 ppm for the ^{1}H and ^{13}C NMR chemical shifts, accordingly, and this is graphically presented in Figure 7. In addition, for ^{1}H NMR chemical shifts, the root-mean-square deviations (RMSD) was found to be 0.33 ppm, while in the case of ^{13}C NMR, the RMSD was 2.12 ppm for the range of about 170 ppm.



Scheme 3. Calculated (blue) and experimental (black) ^{1}H NMR chemical shifts of vindoline trimer (**1**).



Scheme 4. Calculated (blue) and experimental (black) ^{13}C NMR chemical shifts of vindoline trimer (**1**).

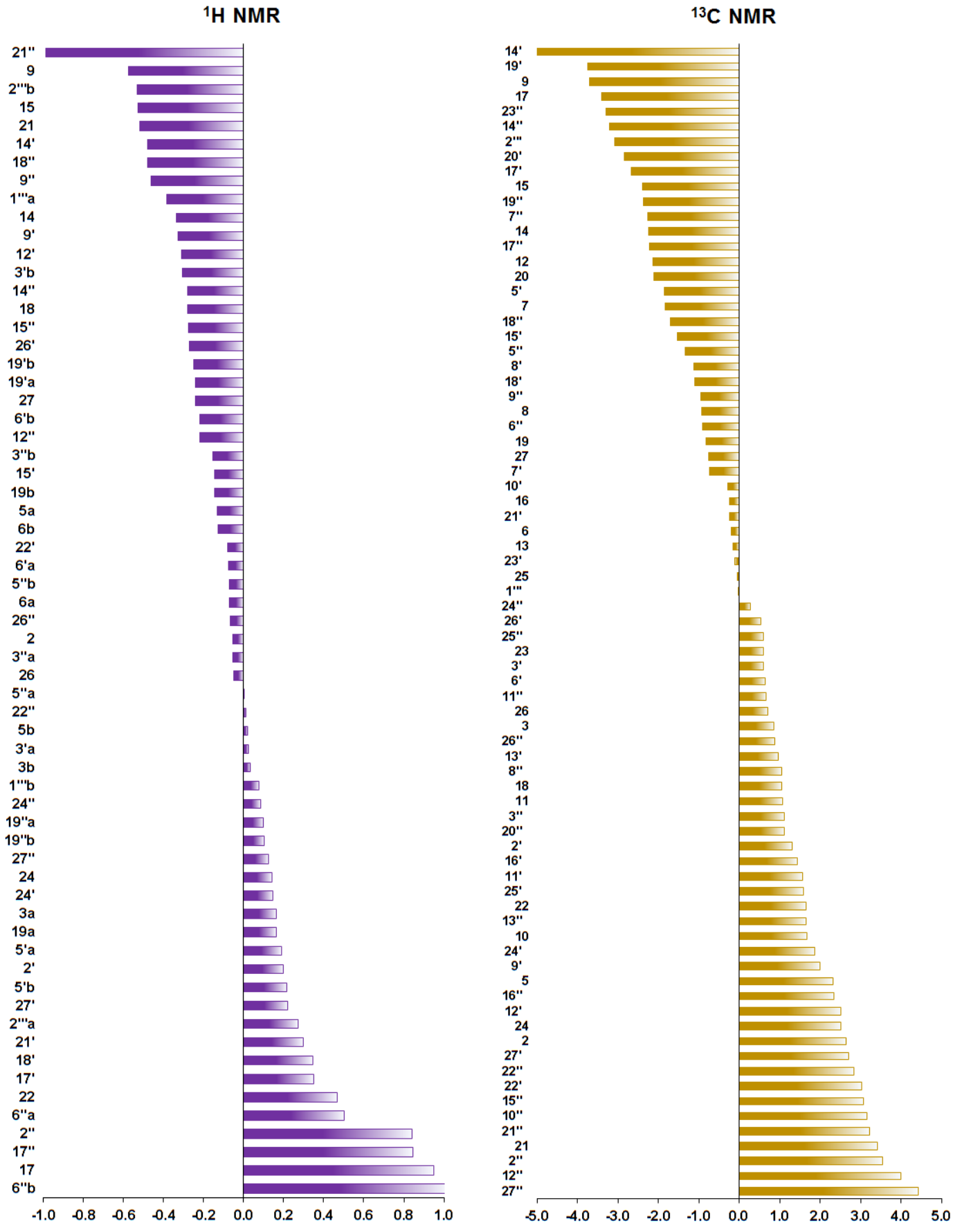


Figure 7. Bar graphs of the distribution of errors of calculated scaled ^{1}H and ^{13}C NMR chemical shifts of the conformer **1c** of vindoline trimer (**1**).

It followed that generally, calculated ^{1}H and ^{13}C NMR chemical shifts of this large alkaloid, considered here with taking into account its stereochemical structure, demonstrated a perfect agreement with experimental spectra, with more reliable correlation observed for carbons as could be seen in Figure 8. However, it should be noted that there are several points of disagreement for the ^{1}H NMR chemical shifts (which are marked in Figure 8), namely, in the case of H-17, H-18', Hb-6'', H-16'', H-17'', and H-21''.

We believe that there are two main reasons of the mentioned discrepancies. The first one is associated with the possible underestimation of complex intramolecular stereoelectronic interactions in theoretical calculations of molecular properties at the DFT level. There is an assumption that applying of methods of a higher level in terms of accounting for electronic correlation effects, such as those of coupled clusters theory, would result in a more adequate description of stereoelectronic states of the specified molecular moieties. However, the study of such large molecules as the vindoline trimer at the CCSD level is currently an extremely resource-demanding task being unlikely feasible in practice.

The second reason for the mentioned ambiguity may be connected with the incorrect construction of the spatial model for the **1c** conformer, since one of the major source of error in the ^{1}H NMR calculations lies in the fact that experimental spectrum is actually a superposition of several rotational isomers. Unfortunately, practically it was not possible to optimize and calculate NMR chemical shifts of dozens (if not hundreds) of probable vindoline trimer conformers. Even the use of the models based on the probabilistic distribution of diastereomers (in this case, conformers) does not help in resolving this contradiction. Thus, the calculation of chemical shifts at the mPW1PW91/6-311+G(*d*,*p*) level within the framework of the DP4+ approach[13] of the four main conformers **1a**-**1d** assumes the absolute predominance of conformer **1a** in a liquid phase, see SI - Figure S11. However, that particular conformer does not correspond to the experimental ROESY data and thus cannot be used for a further higher-level geometry optimization and correlation analysis of NMR chemical shifts.

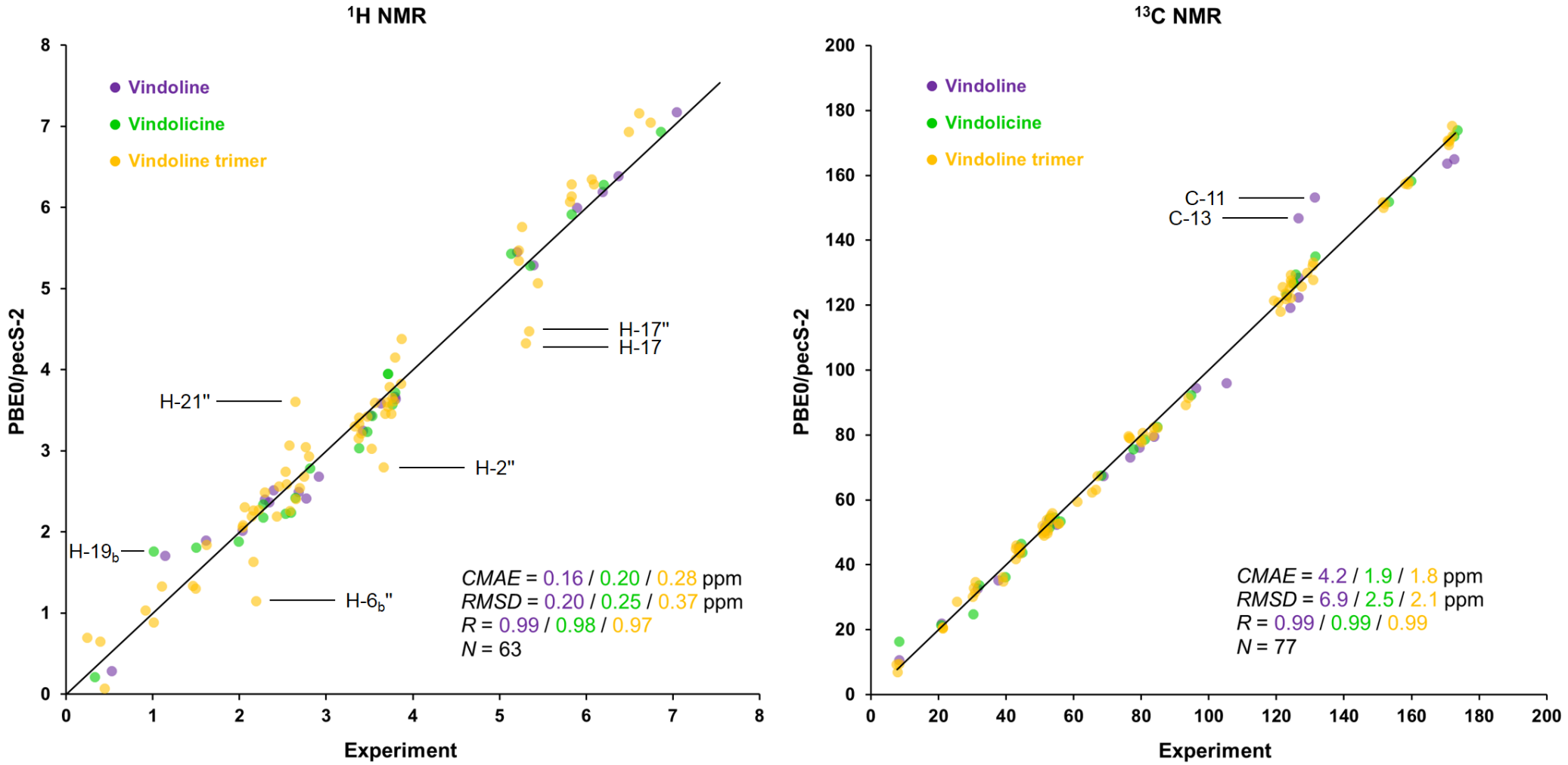


Figure 8. Correlation plots of calculated at the PBE0/pecS-2 level ^{1}H (left) and ^{13}C (right) NMR chemical shifts (ppm) of **1a**, **4** and **5** *vs*. experiment.

For a deeper understanding of a second reason and to find out at what stage may the discrepancies arise between experimental and calculated NMR chemical shifts, we carried out correlation analysis of these data in the series of monomeric vindoline (**4**) – dimeric vindolicine (**5**) – vindoline trimer (**1**), see Figure 9. Since vindolicine (**5**) consists of two vindoline subunits, mirrored located relative to their center of symmetry, so that only one set of NMR signals is observed. As a result, for monomeric vindoline (**4**) and dimeric vindolicene (**5**), no difficulties arise in reproducing experimental ^{1}H NMR data by the performed calculations. However, when going to the space-dependent trimer of vindoline (**1**), the deviations of calculated ^{1}H NMR chemical shifts from experiment do arize, which are marked by red areas in Figure 9. The steric hindrance of the interaction of the three subunits can thus "leave their imprint" on theoretical values of NMR chemical shifts.

As for the calculated ^{13}C NMR chemical shifts, significant deviations are observed even in monomeric vindoline (**4**) for the atoms C-11 and C-13. In view of the fact that experimental NMR parameters are a superposition of a number of molecule thermodynamic states, they are fundamentally impossible to being reproduced within a single calculation. This is one of the main limitations of calculated NMR chemical shifts of large molecules.

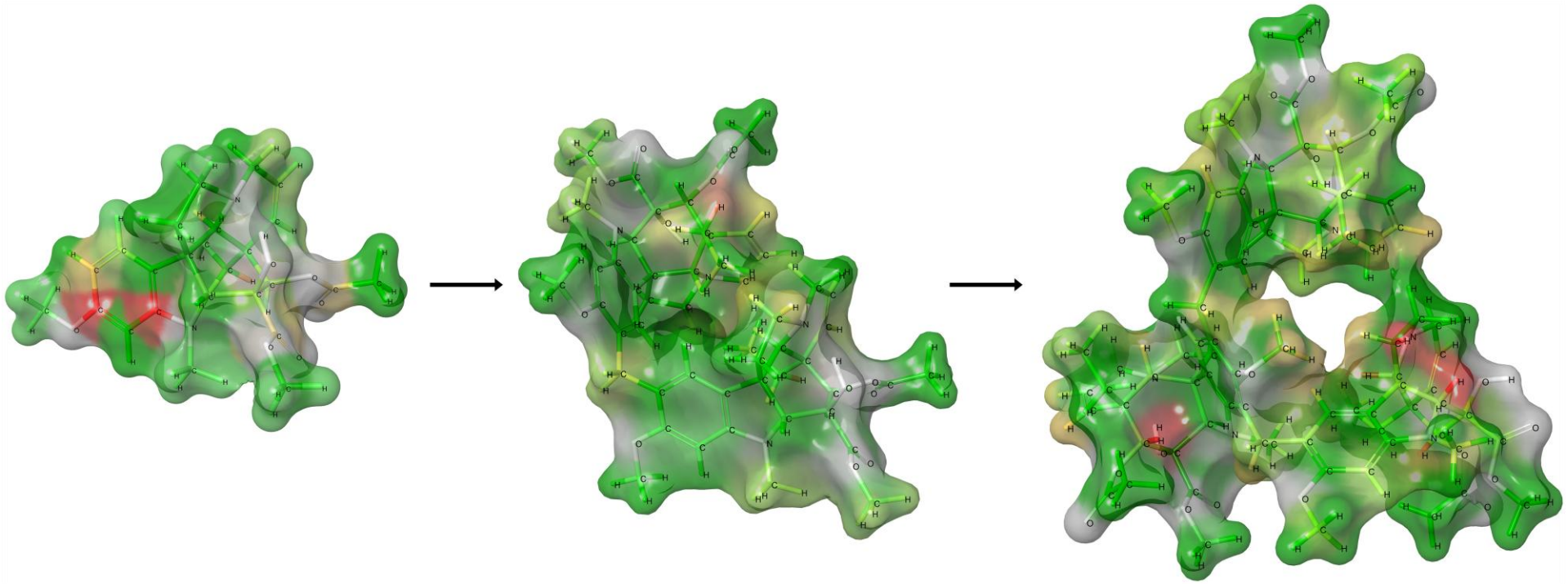


Figure 9. Distribution of errors for the calculation of ^{1}H and ^{13}C NMR chemical shifts in the series vindoline (**4**) – vindolicine (**5**) – vindoline trimer (**1**). The color gradation from green to red corresponds to the error increase.

EXPERIMENTAL AND COMPUTATIONAL DETAILS

NMR experiments

The ^{1}H and ^{13}C NMR spectra were recorded in $CDCl_3$ solutions at ambient temperature on Bruker Neo spectrometer operating at 500.28 and 125.81 MHz or 600.16 and 150.91 MHz, respectively. ^{1}H and ^{13}C chemical shifts (δ in ppm) were measured with accuracy of 0.01 and 0.02 ppm, respectively, and referred to the solvent residual peak. Coupling constants (J in Hz) values approaches to 0.1 Hz. The assignment of ^{1}H and ^{13}C signals in spectra was performed using 2D inverse proton-detected heteronuclear correlation HMBCETGPL3ND and HSQCEDETGPSISP ^{1}H-^{13}C methods. All 2D NMR spectra were recorded by using standard gradient Bruker pulse programs. The HMBC spectra were recorded with an acquisition time of 0.28 s, 28 kHz spectral width for ^{13}C dimension, 4096 points in the ^{1}H dimension, a spectral width of 7100 Hz, 1088 increments, and aiming at long-range coupling constant of 10 Hz for ^{13}C, with a relaxation delay of 1.5 s. The COSY sequence (COSYGPPPQF) used gradient pulses for selection and trim pulses before relaxation delay (30 ms). Two-dimensional ^{1}H–^{1}H ROESY spectra were used to determine the configurational structure of the studied compounds (mixing time D8=0.6s). The spin lock consists of two pulses of 216 microseconds each at a power of 0.2 watt

Conformational search and geometry optimization

The initial conformational search of studied compounds **1**, **4**, **5** was carried out using the OPLS3 force field in the liquid phase of a specific solvent, employing the MacroModel module implemented in the Schrödinger Maestro 11.5 package.[29] The first-step geometries of the lowest-energy conformers were then reoptimized at the DFT level of theory with the Minnesota M06-2X exchange-correlation functional[30] using the cc-pVTZ basis set both for hydrogen and carbon atoms[31] and aug-cc-pVTZ[32] for nitrogen and oxygen to describe properly their lone pairs with using the Gaussian 09 program.[33] All optimizations were performed with taking into

account the solvent effect of chloroform-*d* by using the integral equation formalism polarizable continuum model (IEF-PCM), [34,35] parametrized for this solvent. The equilibrium geometries of the conformers **1a**-**1d** of vindoline trimer are given in the Supporting Information.

Calculation of chemical shifts

Calculations of ^{1}H and ^{13}C NMR isotropic magnetic shielding constants and corresponding chemical shifts were carried out at the DFT level in the liquid phase of chloroform-*d* using the Gaussian 09 package. In these calculations, we employed the one-parameter hybrid functional PBE0,[36] which was used in combination with the original pecS-2 basis set, the latter being a property-energy consistent basis set of the second level.[25,26] TMS was used as a reference compound for both ^{1}H and ^{13}C NMR calculations, which was calculated at the same level of theory as the compounds under study.

To take into account systematic errors of calculated chemical shifts, we have established correlations between their experimental chemical shifts (x) and isotropic magnetic shielding constants (y) values, which were further used to find the linear regression equation of the $y = ax + b$ type. The parameters a and b were then used for the calculation of scaled chemical shifts using the equation

$$\delta_{calc} = (\sigma_{calc} - b)/a \quad (1)$$

Corrected Mean Absolute Errors (CMAE) were evaluated as

$$CMAE = \frac{\sum_{i=1}^{n} \left| \delta_{exp} - \left(\frac{\sigma_{calc} - b}{a} \right) \right|}{n} \quad (2)$$

where σ_{calc} are the unscaled shielding constants for each of the n nuclei, while a and b are the slope and intercept of the linear regression (1).

The Root Mean Square Deviations (RMSD) were calculated as

$$RMSD = \sqrt{\frac{\sum_{i=1}^{n}(\delta_{exp} - \delta_{calc})^2}{n}} \quad (3)$$

CONCLUSIONS

A complete ^{1}H and ^{13}C NMR assignment of a large trimeric vindoline alkaloid based on a thorough analysis of its ROESY, COSY, HSQC, and ^{1}H-^{13}C HMBC correlations in combination with the state-of-the-art quantum-chemical calculations are carried out. It is worth noting that the presence of a three time repeated unit (vindoline) allows to separate the contributions from the sub-units from those coming from longer distances interactions. While the first ones are of larger amplitude, the others are at the origin of subtle chemical shift deviations, which remain difficult to take into account. Based on the results of the correlation analysis of chemical shifts, the main "bottlenecks" in the calculation of magnetic resonance parameters are outlined. In general, the perspectives of the quantum chemical methods in the interpretation of experimental NMR data together with the overall limitations in carrying out such surveys are outlined and discussed.

Supporting information: Experimental ^{1}H, ^{13}C, ROESY, COSY, HSQC and HMBC NMR spectra of vindoline trimer. Cartesian coordinates (Angstroms) of the main low-energy conformers **1a-1d** of vindoline trimer optimized at the M06-2X/cc-pVTZ//aug-cc-pVTZ level in the liquid phase of chloroform-*d* within the IEF-PCM scheme.

Author Contributions: Conceptualization, methodology, software, investigation, writing – V.A.S.; validation, formal analysis, data curation, investigation, writing – S.V.Z.; resources, writing – G.M.; supervision, project administration, writing – L.B.K. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: All calculations were performed at Irkutsk Supercomputer Center of the Siberian Branch of the Russian Academy of Sciences using the HPC cluster “Academician V.M. Matrosov” (http://hpc.icc.ru, accessed on 02 October 2024) and at A.E. Favorsky Irkutsk Institute of Chemistry using the facilities of Baikal Analytical Center (http://ckp-rf.ru/ckp/3050, accessed on 02 October 2024).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations list

CCSD	Coupled Cluster Singles and Doubles
CMAE	Corrected Mean Absolute Error
COSY	COrrelated SpectroscopY
DFT	Density Functional Theory
HMBC	Heteronuclear Multiple Bond Correlation
HSQC	Heteronuclear Single-Quantum Correlation spectroscopy
IEF-PCM	Equation Formalism Polarizable Continuum Model
IUPAC	International Union of Pure and Applied Chemistry
M06-2X	Minnesota exchange-correlation hybrid functional with 54% HF exchange
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
OPLS	Optimized Potentials for Liquid Simulations
PBE	Perdew–Burke–Ernzerhof exchange-correlation functional
pecS-2	property-energy consistent basis set of the second level purposed for the chemical Shifts calculations
PES	Potential Energy Surface
RMSD	Root-Mean-Square Deviation
ROESY	Rotating-frame nuclear Overhauser effect SpectroscopY

References

1. Szabó, L.F. Rigorous Biogenetic Network for a Group of Indole Alkaloids Derived from Strictosidine. *Molecules* **2008**, *13*, 1875-1896.
2. Szwarc, S.; Jagora, A.; Derbré, S.; Leblanc, K.; Rharrabti, S.; Said-Hassane, C.; El Kalamouni, C.; Gallard, J.-F.; Le Pogam, P.; Beniddir. M.A. Combination of Machine Learning and Empirical Computation for the Structural Validation of Trirosaline, a Natural Trimeric Monoterpene Indole Alkaloid from *Catharanthus roseus*. *Org. Lett.* **2024**, *26*, 274–279.
3. Ali, S.; Hénon, E.; Leroy, R.; Massiot, G. Addition of Vindoline to *p*-Benzoquinone: Regiochemistry, Stereochemistry and Symmetry Considerations. *Molecules* **2021**, *26*, 6395.
4. Asia; Sammer, Y.; Vendier, L.; Massiot G. Structure and synthesis of vindolicine and derivatives. *Chem. Biodiversity*, **2024**, e202301928.
5. Keglevich, P.; Hazai, L.; Dubrovay, Zs.; Dékány, M.; Szántay, Cs., Jr.; Kalaus, Gy.; Szántay, Cs. Bisindole alkaloids condensed with a cyclopropane ring. Part 1. 14,15-Cyclopropanovinblastine and -vincristine. *Heterocycles* **2014**, *89*, 653–668.
6. Keglevich, P.; Hazai, L.; Dubrovay, Zs.; Sánta, Zs.; Dékány, M.; Szántay, Cs., Jr.; Kalaus, Gy.; Szántay, Cs. Bisindole alkaloids condensed with a cyclopropane ring. Part 2. Cyclopropano-vinorelbine and its derivatives. *Heterocycles* **2015**, *90*, 316–326.
7. Keglevich, A.; Mayer, S.; Pápai, R.; Szigetvári, A.; Sánta, Z.; Dékány, M.; Szántay, C., Jr.; Keglevich, P.; Hazai, L. Attempted synthesis of *Vinca* alkaloids condensed with three-membered rings. *Molecules* **2018**, *23*, 2574.
8. Jensen, F. *Introduction to Computational Chemistry*, John Wiley and Sons Ltd, Chichester, P019 8SQ, England, **2007**, 624 p.
9. Sauer, S.P.A. *Molecular Electromagnetism. A Computational Chemistry Approach*, University Press, Oxford, **2012**, 322 p.

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

11. Ermanis, K.; Parkes, K.E.B.; Agback, T.; Goodman, J.M. Doubling the power of DP4 for computational structure elucidation. *Org. Biomol. Chem.* **2017**, *15*, 8998–9007.

12. Ermanis, K.; Parkes, K.E.B.; Agback, T.; Goodman, J.M. The optimal DFT approach in DP4 NMR structure analysis-pushing the limits of relative configuration elucidation. *Org. Biomol. Chem.* **2019**, *17*, 5886–5890.

13. Grimblat, N.; Zanardi, M.M.; Sarotti, A.M. Beyond DP4: an improved probability for the stereochemical assignment of isomeric compounds using quantum chemical calculations of NMR shifts. *J. Org. Chem*. **2015**, *80*, 12526−12534.

14. Grimblat, N.; Gavín, J.A.; Daranas, A.H.; Sarotti, A.M. Combining the power of *J* coupling and DP4 analysis on stereochemical assignments: The *J*‐DP4 methods. *Org. Lett.* **2019**, *21*, 4003−4007.

15. Tsai, Y-H.; Amichetti, M.; Zanardi, M.M.; Grimson, R.; Daranas, A.H.; Sarotti, A.M. ML-*J*-DP4: An integrated quantum mechanics-machine learning approach for ultrafast NMR structural elucidation. *Org. Lett.* **2022**, *24*, 7487–7491.

16 . Howarth, A.; Goodman, J.M. The DP5 probability, quantification and visualisation of structural uncertainty in single molecules. *Chem. Sci.* **2022**, *13*, 3507–3518.

17 . Smith, S.G.; Goodman, J.M. Assigning the stereochemistry of pairs of diastereoisomers using GIAO NMR shift calculation. *J. Org. Chem.* **2009**, *74*, 4597–4607.

18. Kutateladze, A.G.; Reddy, D.S. High-throughput *in silico* structure validation and revision of halogenated natural products is enabled by parametric corrections to DFT-computed ^{13}C NMR chemical shifts and spin-spin coupling constants. *J. Org. Chem.* **2017**, *82*, 3368–3381.

19. Novitskiy, I.M.; Kutateladze, A.G. DU8ML: Machine learning-augmented density functional theory nuclear magnetic resonance computations for high-throughput *in silico* solution structure validation and revision of complex alkaloids. *J. Org. Chem.* **2022**, *87*, 4818–4828.
20. Novitskiy, I.M.; Kutateladze, A.G. Peculiar reaction products and mechanisms revisited with machine learning-augmented computational NMR. *J. Org. Chem.* **2022**, *87*, 8589–8598.
21. Sarotti, A.M. Successful combination of computationally inexpensive GIAO ^{13}C NMR calculations and artificial neural network pattern recognition: A new strategy for simple and rapid detection of structural misassignments. *Org. Biomol. Chem.* **2013**, *11*, 4847–4859.
22. Zanardi, M.M.; Sarotti, A.M. GIAO C-H COSY simulations merged with artificial neural networks pattern recognition analysis. Pushing the structural validation a step forward. *J. Org. Chem.* **2015**, *80*, 9371–9378.
23. Howarth, A.; Ermanis, K.; Goodman, J.M. DP4-AI automated NMR data analysis: Straight from spectrometer to structure. *Chem. Sci.* **2020**, *11*, 4351–4359.
24. Gerrard, W.; Bratholm, L.A.; Packer, M.J.; Mulholland, A.J.; Glowacki, D.R.; Butts, C.P. IMPRESSION-prediction of NMR parameters for 3-dimensional chemical structures using machine learning with near quantum chemical accuracy. *Chem. Sci.* **2020**, *11*, 508–515.
25. Rusakov, Y.; Rusakova, I.L. New pecS-*n* (*n* = 1, 2) basis sets for quantum chemical calculations of the NMR chemical shifts of H, C, N, and O nuclei. *J. Chem. Phys.* **2022**, *156*, 244112.
26. Rusakov, Y.Y.; Semenov, V.A.; Rusakova, I.L. On the efficiency of the Density Functional Theory (DFT)-based computational protocol for ^{1}H and ^{13}C Nuclear Magnetic Resonance (NMR) chemical shifts of natural products: studying the accuracy of the pecS-*n* (*n* = 1, 2) Basis Sets. *Int. J. Mol. Sci.* **2023**, *24*, 14623.

27. Krivdin, L.B. Computational protocols for ^{13}C NMR chemical shifts. *Prog. NMR Spectrosc.* **2019**, *112−113*, 103−156.
28. Krivdin, L.B. Computational ^{1}H and ^{13}C NMR in structural and stereochemical studies. *Magn. Reson. Chem*. **2022**, *60,* 733-828.
29. *Schrödinger Release 2018-1: Maestro*; Schrödinger, LLC.: New York, NY, USA, 2018. Available online: https://www.schrodinger.com/freemaestro (accessed on 09 May 2024).
30. Zhao, Y.; Truhlar, D.G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241.
31. Dunning, T.H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023.
32. Kendall, R.A.; Dunning, T.H.; Harrison, R.J. Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. *J. Chem. Phys.* **1992**, *96*, 6796–6806.
33. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Mennucci, B. *GAUSSIAN 09, Revision, C.01*; Gaussian, Inc.: Wallingford, CT, USA, 2009. http://www.gaussian.com, accessed on 09 May 2024.
34. Tomasi, J.; Mennucci, B.; Cancès, E. The IEF version of the PCM solvation method: An overview of a new method addressed to study molecular solutes at the QM ab initio level. *THEOCHEM* **1999**, *464*, 211–226.
35. Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev*. **2005**, *105*, 2999–3093.

36. Adamo, C.; Barone, V. Toward chemical accuracy in the computation of NMR shieldings: the PBE0 model. *Chem. Phys. Lett.* **1998**, *298*, 113-119.