

Advanced NMR-Based Structural Investigation of Glucosinolates and Desulfoglucosinolates

Nada Ibrahim, Ingrid Allart-Simon, Gina de Nicola, Renato Iori, Jean-Hugues Renault, Patrick Rollin, Jean-Marc Nuzillard

▶ **To cite this version:**

Nada Ibrahim, Ingrid Allart-Simon, Gina de Nicola, Renato Iori, Jean-Hugues Renault, et al.. Advanced NMR-Based Structural Investigation of Glucosinolates and Desulfoglucosinolates. Journal of Natural Products, 2018, 81 (2), pp.323 - 334. 10.1021/acs.jnatprod.7b00776 . hal-01904948

HAL Id: hal-01904948
https://hal.univ-reims.fr/hal-01904948

Submitted on 9 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An Advanced NMR-based Structural Investigation of Glucosinolates and Desulfoglucosinolates

Nada Ibrahim,[†] Ingrid Allart-Simon,[†] Gina R. De Nicola,[‡] Renato Iori,[‡] Jean-Hugues Renault,[†] Patrick Rollin,[§] and Jean-Marc Nuzillard.[*,†]

† Université de Reims Champagne-Ardenne and CNRS, Institut de Chimie Moléculaire de Reims, UMR 7312, SFR CAP'SANTE, F-51687 Reims, France.

‡ Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, Agricoltura e Ambiente (CREA-AA), Via di Corticella 133, 40128 Bologna, Italy

§ Université d'Orléans and CNRS, Institut de Chimie Organique et Analytique, UMR 7311, BP 6759, F-45067 Orléans, France.

ABSTRACT:

Glucosinolates (GLs) constitute a class of plant secondary metabolites that are characteristic of the order Brassicales. They contain a common hydrophilic moiety connected to a mostly hydrophobic sidechain whose constitution is the most frequent structural variant. Their transformation by myrosinases leads to intensively studied and highly reactive compounds of biological relevancy. In other respects, the enzymatic desulfation of GLs produces derivatives (DS-GLs) that are useful for their analysis. A collection of 31 compounds representing 17 different sidechains either in GL or DS-GL form was established in order to report accurate descriptions of their ^{1}H, ^{13}C, and ^{15}N NMR parameters. The description of the ^{1}H NMR spectra was achieved using the PERCH software, which accurately analyzes the complex coupling patterns that arise from strongly coupled nuclei. The chemical shift assignments were supported by 2D COSY, HSQC, and HMBC spectra. The impact of desulfation and the influence of the sidechain nature on chemical shift values are discussed. The result of the spectroscopic analysis and the 3D chemical structure model of each studied molecule were grouped in Structure and Data Format (SDF) files. The NMR parameters were also collected in a simple text file, a spreadsheet file, and a relational database.

Glucosinolates invariably comprise a hydrophilic *β*-D-glucopyranosyl unit, an *O*-sulfated anomeric (*Z*)-thiohydroximate function (Chart 1a), connected to a rather hydrophobic sidechain whose constitution, depending on plant species, is the most frequent structural variant and derives from α-amino acid biosynthetic precursors (Chart 1b).[1,2] A small number of GLs have their carbohydrate unit esterified by aromatic acids such as benzoic, *p*-coumaric, or isoferulic acid.[2] Several methionine derived GLs, which constitute the largest group of GLs, bear in their sidechain an additional sulfur atom at different oxidation states (sulfide, sulfoxide, or sulfone). One of the most studied GLs is glucoraphanin, a compound found in the Brassicaceae, notably in broccoli and Tuscan black kale.[3] To date, about 130 different GLs have been identified,[2] and these compounds are associated in plants to myrosinase (E.C.3.2.1.147), an atypical glucohydrolase able to convert GLs into the corresponding isothiocyanates (ITCs). Thus far, ITCs are the most intensively studied GL breakdown products and have recently attracted intense research interest regarding their protective effect against cancer and neurodegenerative disorders. A recent study estimated a possible neuroprotective effect of glucomoringin, 4-(*α*-L-rhamnopyranosyloxy)benzyl glucosinolate, bioactivated with the enzyme myrosinase to form the corresponding ITC moringin, 4-(*α*-L-rhamnopyranosyloxy)benzyl ITC, in the treatment or prevention of Parkinson's disease.[4] The possible pest control by ITCs from GLs for agricultural crop protection has been also reported.[5,6] The relationships of the GL-myrosinase system with insects have therefore been thoroughly studied.[7,8]

The analysis of GLs by HPLC is a difficult task that is often carried out on desulfated GLs (DS-GLs) which result from the reaction of GLs with sulfatases.[9] DS-GLs being standards for GL identification, this study covered both native and desulfated GLs. Chart 1 shows the invariant part of GLs and the variable parts of the compounds involved in the present study. Each of the

GLs studied here has a trivial name, most often given according to the plant from which it has been isolated for the first time, and a "code", which is an abbreviated form of the name. The codes are by no means officially assigned to the compounds but are of common use among GL specialists.

The compounds studied were related to 17 sidechains (Table 1) and in 14 cases both GL and DS-GL forms were available. Only GRE, DS-GCL, and DS-GSY had no counterpart in our collection, thus leading to the analysis of 31 compounds, designated hereafter as the (DS)GL collection.

The present work was first intended to provide reference NMR data on selected GLs and DS-GLs for identification purposes. The description of the ^{1}H NMR spectra in terms of NMR parameters, considered here as chemical shifts and coupling constants, was hampered by the presence of strongly coupled spin systems, even in spectra recorded at 600 MHz. The PERCH software was consequently used for accurate spin network analysis.[10] In brief, the PERCH software determines a starting set of NMR parameters from a molecular structure and refines them so that the corresponding reconstructed ^{1}H NMR spectrum fits best with the experimental one. The necessity of reporting analysis results in a way that could fit with their dissemination *via* computerized databases has led to the design of a human- and machine-readable data format.[11] The influence of desulfation and of the sidechain constitution on chemical shifts is also discussed, in relation with molecular 3D structures.

RESULTS AND DISCUSSION

Compounds, Public Databases, and Study Outcome. An attempt to find Chemical Abstracts Service[12] (CAS) Registry Numbers of the compounds of the (DS)GL collection failed for DS-

GSY and DS-GCL. The DS compounds unknown to CAS (as well as the others) were prepared as analytical standards by action of sulfatase on intact GLs. All GLs are known to include a *Z*-configured imino double bond, as documented in section "Lossen and Beckmann Rearrangements" of a review article by Walter and Schaumann.[13] Since the configuration of the imino double bond does not always appear explicitly in the publication from which CAS collected the registered structures, this configuration is often wrong or left "unspecified" in the CAS database. Other databases such as ChemSpider or PubChem contain 2D and 3D GL structures that are incorrect. The side chains of GIB, GRA, and GRE contain a methylsulfinyl group, whose *R* absolute configuration of the sulfur atom is rarely specified.[14] The first sample of DS-GRA that we analyzed was indeed a mixture of diastereoisomers obtained by non-stereoselective oxidation of DS-GER.[15] A sample of DS-GRA from desulfation of GRA was also analyzed (*vide infra*).

The publicly available outcome of the present study consists of, i) a collection of SDF[16] files which comprise 3D structures, atom naming information, ^{1}H, ^{13}C, and ^{15}N chemical shifts, and ^{1}H-^{1}H scalar coupling constants determined by PERCH; ii) a collection of NMR files that include 1D and 2D time domain data, ^{1}H, ^{13}C, ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC, and ^{1}H-^{15}N HMBC spectra, and ^{1}H spectra reconstructed by PERCH; iii) a spreadsheet file that summarizes all the NMR-related information and a simple text file version of it; and iv) a relational database file that includes the content of the spreadsheet file as well as structural information that would enable the rebuild of the 3D molecular structures from scratch.

The 3D models of the (DS)GL collection members were built initially for the ^{1}H NMR parameter determination assisted by the PERCH software. The structure files in MMS PERCH

format were exported in the MOL[16] format in order to prepare the assembly of the SDF files. The assembly process also involved a re-indexing and a naming of the atoms.

Each (DS)GL is associated to a human-readable SDF file that contains, i) atom data, with one line per atom including 3D coordinates, the first line of atom data being related to atom indexed 1; ii) bond data that refer to atom indexes as previously defined; iii) atom naming data correlating numerical atom indices and textual atom names; and iv) property value pairs, including NMR parameter names derived from atom names and their associated numerical values. It must be emphasized that not only chemical shift values are associated to atom indexes (as usual for resonance assignment) but that, less commonly, coupling constant values are systematically related to index pairs of coupling atoms (for coupling assignment).

Atom re-indexing was carried out according to Chart 1a. Heavy (*i. e.* non-hydrogen) atoms in the DS-GL substructure that is common to all structures are numbered from 1 to 16, the sidechain heavy atoms are then numbered according to Chart 1b, always starting at index 17. In the sulfate group of an intact GL, the sulfur atom receives the next available index and the three bound oxygen atoms receive the next three available indexes. Hydrogen atoms are then indexed in the order of the heavy atoms they are attached to. Heavy atom naming is straightforward: atom 1 is named C1, the name being built from the atomic symbol followed by the atom index. The rhamnose carbon atoms in the GMG sidechain, indexed from 24 to 29, were exceptionally named from C1' to C6’ instead of C24 to C29 and their attached hydrogen atoms named from H1' to H6'. Diastereotopic methylene hydrogen atoms receive a suffix in their names, so that H6R and H6S designate the *pro-R* and *pro-S* hydrogen atoms bound to C-6. Identically, suffixes Z and E in terminal methylene groups are appended to the names of *pro-Z* and *pro-E* hydrogen atoms.

Suffixes A, B, and C are appended to the names of the equivalent hydrogen atoms in methyl groups.

NMR parameter names are used as property names in the data section of the SDF files. For example D_C1 is the name associated to the value of the chemical shift of C-1, D denoting delta, the usual symbol of chemical shift. In the same way, J_H1_H2 and J_H2_H1 are the names of the coupling constant value shared by H-1 and H-2, so that a textual search for "J_H2" in an SDF file easily highlights all the couplings of H-2 without missing its coupling with H-1.

Once all information pieces about each (DS)GL have been integrated in an SDF file, the atom 3D coordinates from PERCH structure exportation were re-calculated by conformational analysis. Conformers were generated and optimized using a classical force-field molecular mechanics approach. The coordinates of lowest energy conformer were reported in the final, publicly released SDF files (Supporting Information).

At this stage, it was possible to sequentially read all the SDF files in order to produce a single textual data table, for which Table 2 presents the sections for sinigrin and DS-sinigrin. A simple importation into a spreadsheet of the whole text, using the tabulation character as separator, provides an easy way to read it and to save it as an Excel file, for example.

In a more database-oriented perspective, we also provide the (DS)GL NMR data as an SQLite file as well. SQLite is a lightweight, public-domain, zero-configuration relational database system that is embedded in many software offerings, which one uses every day without being aware of it.[17] The GLs.db database file that was created from the SDF files contains seven tables. The NAMES table associates compound codes (like DS-SIN) to compound names (DS-sinigrin). The database tables named HNMR, CNMR, and NNMR associate a compound code, a position

name (1 for C-1, 6 for H-6R and H-6S), an atom label, and the corresponding chemical shift value. The JHH table associates a compound code, a position name, and an atom label for two hydrogen atoms, and the corresponding coupling constant value. Two tables were dedicated to structural information: ATOMS associates a compound name, an atom label, a chemical symbol, an electric charge, and a chirality flag (R/S for stereogenic centers, including those at sulfur in sulfoxides) and BONDS that associates a compound code, two atom labels, the multiplicity of the bond between the two atoms, and a configuration flag (Z/E). The ATOMS and BONDS tables content is sufficient in order to build a 3D model of the molecules. The GLs database was created in order to be able to extract data from the tables in any possible way using an SQL query like

"select Code, J from JHH where Atomname1 = "H1" AND Code NOT LIKE "DS%"

for the collection of the *J*(H-1, H-2) coupling constants among all intact GLs.

Determination, Assignment, and Collection of NMR Parameters. The determination and assignment of ^{1}H, ^{13}C, and ^{15}N chemical shifts of each compound in the (DS)GL set started with the recording of 1D and 2D spectra. The basic rules of ^{1}H and ^{13}C chemical shift interpretation[18] and the analysis of 2D ^{1}H-^{1}H COSY, ^{1}H-^{13}C HSQC, and ^{1}H-^{13}C HMBC[19] spectra easily produced ^{1}H and ^{13}C NMR chemical shift assignments due to the structural simplicity of the studied compounds. The assignment process is exemplified hereafter by means of the spectra of glucolimnanthin (GLI) in Figure 1.

Like in all (DS)GLs, C-7 plays a central role, as it correlates in the HMBC spectrum with H-1 in the sugar unit and with H-8 in the sidechain. The signal of H-1 is well separated and generally close to the HOD solvent signal; it correlates in the COSY spectrum to H-2. The HSQC

spectrum, when recorded with J-modulation (CH_2 signal in opposite phase with those of CH and CH_3 groups), readily gives the position of both the C-6 protons. In the COSY spectrum, the latter correlates with H-5, H-5 with H-4, and H-4 with H-3. In the present case, the near complete superimposition of H-2 and H-3 resonances complicates their differenciation by means of the COSY spectrum. The assignments of H-1 to H-6 and C-1 to C-6 may be done by the observation of their HMBC correlations and by the observance of the chemical shift order of the carbon atoms (see in section "Relationships between Structures and NMR Parameters"). The assignment of the resonances from the sidechain clearly depends on the nature of this sidechain. Because H-8 is identifiable by means of the HMBC spectrum and C-8 by the HSQC spectrum from H-8, the resonances arising from the aromatic and the methyl positions remain to be assigned. The intense singlet resonance of the Me-24 atoms correlates with the signal of the aromatic C-21, thus giving their respective assignments. C-17 is readily identified because it is the only unassigned quaternary carbon. The sharp signal of H-22 is readily assigned because H-22 is only subject to long-range couplings. The H-19 signal is readily located because it shows two aromatic vicinal couplings and appears as a broad pseudo-triplet. The HMBC correlation between C-8 and H-18, as well as the one between H-8 and C-18, provide the chemical shifts of H-18 and C-18 and, by elimination, those of H-20 and C-20, thus completing the resonance assignment task. Interestingly, the ^{1}H NMR spectra of GLI and DS-GLI are nearly superimposable: the most important variation is at one of the C-8 methylene protons, shielded by 0.11 ppm in DS-GLI. It is therefore almost impossible to distinguish GLI and DS-GLI by sole inspection of their ^{1}H NMR spectra. The ^{13}C NMR spectra of GLI and DS-GLI are highly similar, the most noticeable difference being the shielding of C-7 by 8.2 ppm in DS-GLI.

All NMR spectra of the GLs were recorded in D_2O, a solvent that mimics at best the natural environment of these compounds. NMR in D_2O precludes observation of the exchangeable protons, mainly those from the hydroxy groups of the glucose unit, a loss that is not detrimental for the structural analysis of (DS)GLs. Sodium 3-(trimethylsilyl)propionate-2,2,3,3-d_4 (TSP-*d4*) was used as internal reference compound for ^1H and ^{13}C NMR spectra. This gave the opportunity to evaluate the possibility of referencing ^{13}C chemical shifts from the ^1H shifts.[20] The study was carried out on a set of 28 samples for which ^1H and ^{13}C NMR spectra were recorded. The resonance frequency of the methyl groups of TSP-*d4* was calculated by adding their offset frequency to the spectrometer base frequency (BF1) and setting the Spectrum Reference (SR) calibration parameter to 0. The average apparent chemical shift of TMS in ^1H NMR spectra (ν_{TMS} – BF1) was -0.093 ppm (-55.8 Hz), thus meaning that if SR equals 0 then 0.093 ppm must be added to the chemical shift values read on spectra. The standard deviation of this calibration parameter was 6.5 ppb (3.9 Hz), ensuring that calibration without any added reference compound may be safely undertaken. The average apparent chemical shift of TMS in ^{13}C NMR spectra (SR(^{13}C) = 0) was -2.92 ppm (-440.5 Hz) with a 28 ppb (4.2 Hz) standard deviation. A ^{13}C NMR spectrum calibration with a null SR and without TSP-*d4* internal reference results in a systematic error of about 3 ppm on chemical shifts. The calculation of SR(^{13}C) from SR(^1H) using TSP-*d4* as reference in D_2O at 298K can be achieved using $\Xi = \nu_{TMS}(^{13}C)/ \nu_{TMS}(^1H)$. The average $10^9.\Xi$ value is 251449491 with a standard deviation of 9. The application of this mean value to the set of $\nu_{TMS}(^1H)$ leads to an average error of -0.01 ppm on ^{13}C chemical shifts with a 0.03 ppm standard deviation, which is compatible which the order of magnitude of chemical shift accuracy that is required by chemistry journals. Using the Ξ value from reference for TSP at 295K would result in a 0.2 ppm glitch of the ^{13}C chemical shift scale.[21]

The choice of TSP-*d4* as a water-soluble NMR reference compound is common albeit questionable. As salt of a strong base and a strong acid, DSS (or 4,4-dimethyl-4-silapentane-1-sulfonic acid sodium salt) would in theory be a better choice since the protonation state of the sulfonate group is not expected to be modified by one of the analyzed compounds. Conversely, the carboxylate anion in TSP may undergo protonation at low (<5) pH. GL samples were available as potassium salts, *i. e.* salts of a strong base and a strong acid. The pH value of their solution is close to 7 so that any possible change in the protonation state of TSP should not induce a significant variation of the methyl group chemical shifts. Another water-soluble standard, DSA (4,4-dimethyl-4-silapentane-1-ammonium trifluoroacetate), may also show some chemical shift dependence on pH variation due to the presence of an ammonium group.[22]

The ^{1}H-^{15}N HMBC spectra were used to indirectly determine the ^{15}N chemical shift of the nitrogen atoms, exploiting the correlation of the H-8 signals with those of N-10. For each studied compound, a spectrum was first recorded with a 100 ppm window in F_1 and a small number (16) of t_1 increments in order to obtain a quick (3 min recording time) estimation of the ^{15}N chemical shift. A 20 ppm spectral width and 128 increments (11 min recording time) were then used to obtain a better resolved spectrum with a better signal-to-noise ratio. The collected ^{15}N chemical shifts were stored uncorrected (SR = 0, BF = 60.813685 MHz) in the NMR parameter files. This study constitutes the first determination of ^{15}N NMR resonances in glucosinolates.

Handling ^{1}H NMR Data. The ^{1}H NMR spectrum of sinalbin (SNB) was among the first recorded ones; the first order analysis of its glucose signals was found to be inappropriate, as immediately visible by the inspection of the anomeric signal: while a doublet was expected, a set of 10 lines was observed, as reported in Fig. 2a. This high complexity of a signal is rationalized by the presence of a virtual coupling between H-1 and H-3 caused by the nearly equal resonance

frequencies of H-2 and H-3. In our experience, this situation is likely to happen with benzylic sidechains (DS)GLs and was observed for glucolimnanthin (GLI), glucomoringin (GMG), or glucotropaeolin (GTL), as well as their desulfated homologues. The complex aspect of the H-1 signal in benzylic (DS)GLs has most probably been observed but never reported, except for a recent publication by us.[23] The AA'XX' ($\Delta\nu$ = 217 Hz at 600 MHz) spin system of aromatic ^{1}H nuclei in SNB also showed two patterns, each featuring six peaks (Figures 2b and 2c) that cannot be analyzed at first-order. Therefore, use of iterative NMR parameter refinement software was needed to provide spectra interpretation. The goal of using such software is to extract the coupling constant and chemical shift values for which the corresponding simulated spectrum fits best with the experimental one. It is worth noting that a coupling constant value is associated to the pair of nuclei it originates from, information that is rarely reported. The PERCH software makes use of a hypothesized structure for the molecule under investigation in order to propose a realistic guess of the NMR parameters that are subsequently refined using the integral transform based method.[10] The matching between the refined and the initially predicted parameters, as well as the matching between the predicted and the experimental spectra contribute to the rating of the supposed structure as being the one of the analyzed compound. PERCH can therefore be recognized as a structure validation tool, even though we used it without doubting that the structure of the analyzed compounds were correct (with one exception, see below). We only used PERCH for its ability to extract NMR parameters even from the spectra of highly complex strongly coupled spin systems. Such an approach has been put into practice in the field of natural product chemistry.[24] The advantage of our approach, offering free accessibility to raw, Fourier transformed and interpreted NMR data, lies in the possibility for any interested reader to check and re-analyze data according to personal current interest. The analysis of the spectrum of

glucolimnanthin (GLI) by PERCH is exemplified in Figure 3 and shows how complex signals that would be reported as multiplets can be fully interpreted.

One of the problems inherent in the assignment of ^{1}H NMR chemical shifts to hydrogen atoms has been the handling of diastereotopic hydrogen atoms. In the first place, we paid attention to chemical shift assignments of the *pro-R* and *pro-S* C-6 methylene protons and to their coupling constants with H-5 in the glucose units because it is a common part of all (DS)GL compounds. Taking into account the detailed analysis of the ^{1}H NMR spectrum of methyl *β*-glucopyranoside in D_2O,[25] it appeared that the most deshielded H-6 should be the one with the smallest coupling constant with H-5, regardless of the diastereotopic assignment. Considering (DS)GL ^{1}H NMR spectra, it also appeared that the most deshielded H-6 has the small coupling constant in all cases, except for compounds GLI and GMG. For GLI, PERCH converged towards two identical values of δ[H-6(*pro-R*)] and δ[H-6(*pro-S*)], thus making the result of chemical shift versus coupling constant consistency test undetermined. For GMG, the chemical shift difference was 0.01 ppm and the coupling constant difference was lower than 0.01 Hz, thus making the consistency test unreliable. Following the conclusions of Nishida et al. we imposed the assignment that makes H-6(*pro-S*) more deshielded than H-6(*pro-R*). NMR by itself cannot provide directly any argument to decide which assignment is the correct one. In PERCH, NMR parameters are predicted on the basis of a set of conformers that is generated during a short molecular dynamics run.[26,27] Considering 31 compounds for which prediction was carried out, the inequality δ[H-6(*pro-S*)] > δ[H-6(*pro-R*)] was wrong in nine cases and J(H6S/H5) < J(H6R/H5) was wrong in 14 cases. The consistency between predicted orders of chemical shifts and coupling constants was wrong in seven cases. Parameter optimization thus reduced the number of inconsistencies from seven to two (among which an undecided and a borderline case

for the latter). From these observations, it appeared that PERCH predictions are not necessarily accurate regarding its ability to provide correct assignments of diastereotopic protons in methylene groups, at least in our standard usage conditions. Moreover, the accuracy of the initial NMR parameter prediction determines the relevance of the adjusted parameters because an inversion in the initial order of H-6(*pro-R*) and H-6(*pro-S*) chemical shifts may persist during parameter adjustment. The assignment inversion of *pro-R* and *pro-S* H-6 that occurred in nine cases for the adjusted parameters was corrected by swapping δ[H-6(*pro-R*)] and δ[H-6(*pro-S*)] values as well as the corresponding *J*(H-6(*pro-R*), H-5) and *J*(H-6(*pro-S*), H-5)values.

Proton NMR parameter evaluation of a compound by PERCH is possible only if its 3D structure has been established by means of the integrated MMS molecular modelling system. During interactive molecule building, each atom receives a label that does not correspond to the standardized ones in Chart 1. Even though MMS is able to assign *pro-R* and *pro-S* status to diastereotopic hydrogen atoms in methylene groups and to assign absolute configurations to stereogenic centers (except the sulfur-centered ones), atom labels do not include *pro-R*/*pro-S* or *pro-E*/*pro-Z* qualifiers when MMS files are converted to the MOL files from which SDF files were produced. Therefore, an equivalence table between MMS A/B and *pro-R*/*pro-S* or *pro-Z*/*pro-E* qualifiers was built by hand and exploited at the time the MMS-generated MOL files were converted into SDF files. For each compound, a so-called "master file" was created, containing all different pieces of information needed to build the SDF files, such as i) the method to replace MMS labels by standardized labels in the variable part of the GL or DS-GL structures; ii) the interpretation of A/B MMS qualifiers; iii) the location of PERCH-generated NMR parameter files; iv) the location of the ^{13}C NMR peak list files; v) the ^{1}H and ^{13}C calibration parameters (original 1D NMR files were not calibrated, meaning that the spectrum referencing

parameter was left to zero); vi) the list of MMS labels of the carbon atoms in the decreasing order of their chemical shifts; and vii) the uncorrected ^{15}N NMR chemical shift value of the imino nitrogen atom.

The automation of the ^{1}H NMR data handling was made possible by the processing of the ACA files (Automated Consistency Analysis) that contained the result of the NMR parameter determinations carried out by PERCH. An ACA file is indeed an XML file (eXtensible Markup Language). XML is a data description language with a structured and human-readable syntax similar to the one of HTML, for which efficient data look-up software libraries are available.[28] The search for the relationship between MMS atom labels and chemical shift values in an ACA file is therefore straightforward, as well as the search for the relationship between pairs of MMS atom labels and coupling constant values; this automated search step avoided the manual extraction of data from the PERCH graphical user interface. For ^{13}C chemical shifts, a peak list in XML format was created by the TopSpin software. This list was easily read and the chemical shift values were corrected by addition of the calibration parameter.

Relationships between Structures and NMR Parameters. A superficial observation of the 2D structure of GLs might lead to conclude that the D-glucopyranose unit just ignores the remote sidechain. A convincing proof of the opposite was shown by the ^{1}H and ^{13}C NMR spectra of semisynthetic (*RS*)-DS-GRA,[15] a mixture of epimers that differ by the absolute configuration of the sulfoxide group. PERCH analysis of the ^{1}H NMR spectrum of DS-GRA recorded at 600 MHz was initially performed at a time the sample was thought to be a single diastereoisomeric form. PERCH showed a poor fit for the H-6(*pro-R*) resonance, all other protons being correctly processed. The expected signal should be a doublet of doublet (*dd*), corresponding to the coupling of H-6(*pro-R*) with H-5 and H-6(*pro-S*). However, the observed signal resembled a *ddd*

that did not fit well with a *dd* (top 1D trace in Figure 4b). A ^{1}H, non-tilted, 2D *J*-resolved spectrum was recorded in order to obtain a clue to the origin of the extra splitting (Figure 4). It showed that the *ddd* was indeed a pair of *dd*s, the narrowest splitting, of about 3 ppb or 1.8 Hz, corresponding to the difference of resonance frequencies of the H-6(*pro-R*) in what was revealed to be a mixture of two epimeric sulfoxides. Figure 4a displays the 1D ^{1}H NMR spectrum of (*R*)-DS-GRA obtained by desulfation of natural GRA and shows that H-6(*pro-R*) is more deshielded in (*R*)-DS-GRA than in (*S*)-DS-GRA. Inspection of the ^{13}C NMR spectrum of (*RS*)-DS-GRA permitted measurement of peak splittings for C-7 ($\Delta\delta$ = 15 ppb), C-1 ($\Delta\delta$ = 14 ppb), C-4 ($\Delta\delta$ = 11 ppb), C-19 ($\Delta\delta$ = 20 ppb), and C-18 ($\Delta\delta$ = 42 ppb) thus revealing the influence of the sulfoxide configuration on the sidechain and on the sugar part of DS-GRA. Such a long distance influence between the end-of-chain sulfoxide and the sugar moiety would not have been easy to anticipate if not experimentally proven. A key point about ^{13}C NMR chemical shift values in the D-glucopyranose unit is the invariability of the chemical shift order: δ(C-1) > δ(C-5) > δ(C-3) > δ(C-2) > δ(C-4) > δ(C-6). The order of ^{1}H NMR shift values for H-2, H-3, H-4 and H-5 is more subject to changes, depending on the nature of the sidechain, even though δ(H-1) > δ(H-6), the latter being greater than all other ones. Assuming the given order for ^{13}C, the determination of ^{1}H chemical shifts can be secured by the ^{1}H-^{13}C HSQC spectrum, the particular case of H-6(*pro-R*) and H-6(*pro-S*) having been considered in the previous section. When the sidechain is of the benzylic type, the order δ(H-4) > δ(H-3) ≈ δ(H-2) > δ(H-5) prevails, the similarity of δ(H-3) and δ(H-2) resulting in the virtual coupling of H-1 and H-3 as shown in Figure 2a. In the case of non-benzylic sidechains, the observed order is mostly δ(H-3) > δ(H-5) > δ(H-4) > δ(H-2) with δ(H-3) close to δ(H-5) and δ(H-4) close to δ(H-2). The difference in ^{1}H chemical shifts in the D-

glucopyranose moiety between non-benzylic and benzylic (DS)GL is illustrated in Figure 5 by portions of the HSQC spectra of DS-GCL and GLI.

The search for low energy conformers of the (DS)GLs was partly motivated by an attempt to compare the structures of GLs with those of their corresponding DS-GLs and by the search for a possible source of interaction between the sugar moiety and the sidechain. The main result, exemplified by SNB and GRA in Figures 6a and 6c, is the presence of a hydrogen bond between an oxygen atom of the sulfate group and the hydrogen of HO-2 in all the GLs. The desulfated molecules do not show such a structural feature as demonstrated in Figures 6b and 6d in the case of DS-SNB and DS-GRA. Desulfation, therefore, induces a clear conformational change, even though no particular evidence of it is evident in the NOESY spectra. The hydrogen bond in intact GLs keeps C-8 close to the sugar moiety, thus setting the molecule in a folded conformation, with S-11 located at the turning point.

The main effect of desulfation on chemical shift values is most probably due to the electron withdrawing effect of the sulfate group: the ^{15}N nucleus N-10 is shielded by 8.0 ppm on average by desulfation and the adjacent imino C-7 position is shielded by 8.1 ppm on average. Nearly no effect can be detected for H-1. The example given in Table 2 illustrates the low impact of desulfation on the chemical shifts of sinigrin (SIN). Figure 7 also shows how desulfation has a negligible influence on ^{1}H chemical shifts of gluconasturtiin (GST) and only affects noticeably C-7 of this compound.

Beyond Glucosinolates, NMR Parameter Reporting. Reporting correct structures and NMR parameter assignments is an important part of the activity of organic chemists as the corresponding data constitute a basis for the identification of known molecules during

dereplication[29] and of new molecules during structure elucidation.[30] At the beginning of the application of NMR to organic molecules, the attention was focused on the reporting of 1D NMR parameters, which were the only ones available back then. The advent of 2D NMR made the 1D spectra interpretation more reliable when dealing with compounds of constantly increasing molecular complexity. In the context of a digital world, there is presently no incentive to publish NMR parameters in a format that would be readily understandable by computers with a high level of reliability.[11] The traditional way of reporting NMR parameters, *i. e.*, the one used by scientific journals in general and by the Journal of Natural Products (JNP) in particular is hardly usable by computers. The "line notation" or the "table notation" as reported on page 4 of the "Author Guidelines" of JNP constitute, therefore, dead-ends for significant findings.[31] One of the reasons is that line headers in tables are position numbers in molecular structures whose drawing is separated from NMR data across article pages and in which the numbering is often incompletely or implicitly defined. The back-conversion of a printed structure drawing into a MOL or SDF file is a complex problem for which the best solution is to avoid having to solve it. The SDF files we created for each analyzed (DS)GL gather at the same place the molecular constitution, the atom numbering scheme, and the NMR parameters, all reported in a human-readable/writable and a computer-readable/writable format. The NMR time-domain (raw NMR data) and frequency-domain data (the NMR spectra) are grouped together in a way that is understandable by most, if not all, NMR software and in a way specified by the presently most representative NMR instrumentation actor worldwide. The link between NMR data files and SDF files is ensured in the present case by the compound code; this simple option might have been replaced by a more general, but probably less readable, naming scheme.

The idea of aggregating NMR data, NMR parameters, and molecular structures has already been put into practice in the still developing NMR Mark-up Language project.[32] As indicated by its name, this project relies on an XML-based approach and provides data files that are computer-friendly but poorly adapted to the reading and writing by humans. In an even more general approach, the Allotrope Foundation attempts to develop a technology aiming to rule in a general way the production, exchange, exploration and storage of chemical and biological laboratory data.[33] Besides these heavyweight actors of analytical data handling, discussions are underway for the definition of an NMR parameter exchange format based on SDF files, whose details will soon be released and follow the guidelines proposed in references 11 and 34.

The shifting from a paper-based to a computer-based dissemination of chemical knowledge will have to proceed through a change in data publication habits. This change is motivated by a search for more efficiency in the rapid identification (dereplication) of known natural products, on elaboration of better NMR parameter prediction tools based on error-free databases, and therefore on availability of better tools for unknown structure identification.

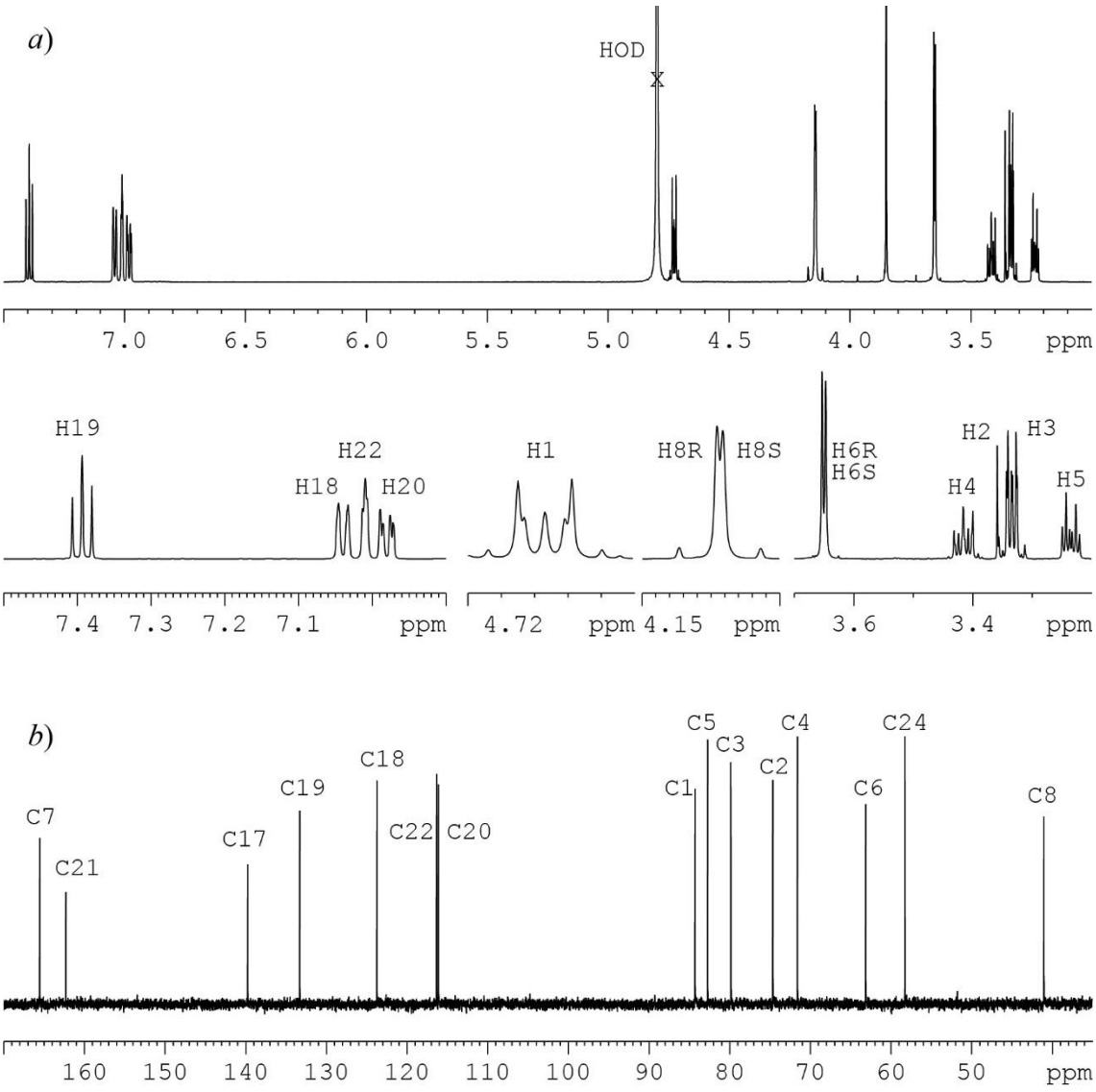
In conclusion, a collection of 16 glucosinolates and 15 desulfated glucosinolates, including 14 pairs of compounds with identical sidechains, has been studied by NMR, using D_2O consistently as solvent. The accurate interpretation of the ^{1}H NMR spectra initially required the assignment of the ^{1}H and ^{13}C resonances using standard 2D HSQC, COSY, and HMBC techniques. The extraction of ^{1}H-^{1}H coupling constants from the ^{1}H NMR spectra was achieved by computer-assisted iterative refinement of chemical shifts and coupling constants using the PERCH software. This approach to spectrum interpretation goes beyond the commonly undertaken first-order analysis and also assigns to each coupling constant value the pair of atoms it originates from. The question of the assignment of chemical shifts of diastereotopic proton pairs remains a

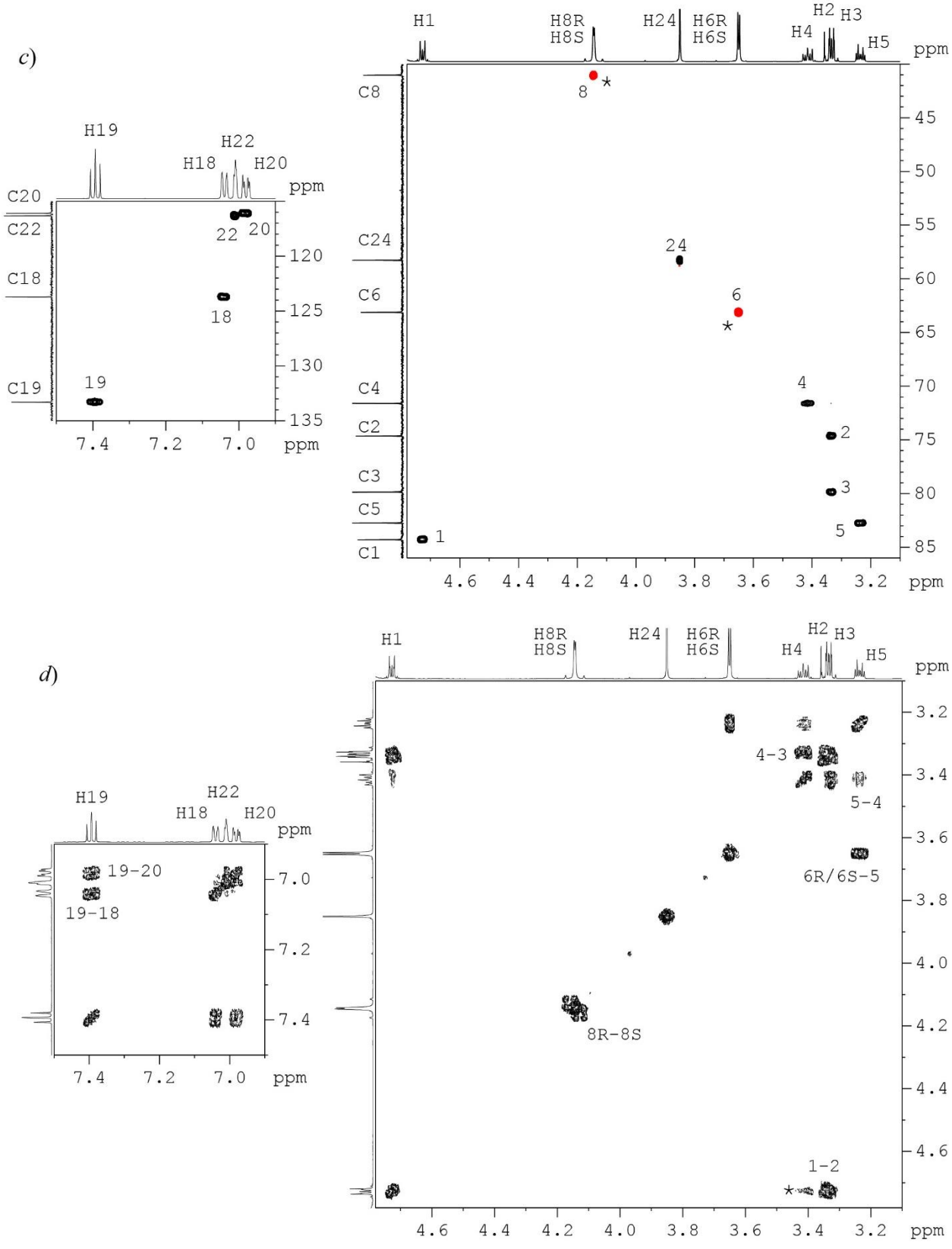
challenge, which has been discussed for the hydrogen atoms in the hydroxymethyl group of glucose units. The ^{15}N NMR data of the studied molecules are reported for the first time. All the NMR parameters were collected as text in a computer file and organized in a structured database for a possible exploitation by an automated process. The most visible effect of desulfation on chemical shifts is a shielding of the carbon and nitrogen atoms of the imino functional group, thus resulting in only small modifications of the ^{1}H NMR spectra. The presence of a benzylic sidechain, both in native and desulfated compounds, induces a strong similarity of chemical shift values of the hydrogens at positions C-2 and C-3 in the glucose unit and creates therefore strong coupling patterns in the ^{1}H NMR spectra of these compounds, clearly visible on the signal of the anomeric proton. A study of the 3D structure of the compounds in the collection was undertaken by force-field molecular modelling in order to attempt to relate the nature of sidechain, the sulfation state, the 3D structure, and NMR parameters. This study showed the presence of an intramolecular hydrogen bond between the sulfate group of native compounds and the sugar unit, a bond that induces a spatial proximity between the sidechain and the face of the sugar ring that contains axial hydrogens at C-2 and C-4. Even though it was not possible to rationalize the particularity of benzylic compounds by considering their geometry, the calculated 3D structures were aggregated with their corresponding NMR parameters in Structure Data Formatted files, in a way that is currently considered as a future standard for NMR data dissemination.

EXPERIMENTAL SECTION. All samples originate from the Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria, Agricoltura e Ambiente (CREA-AA), Via di Corticella 133, 40128 Bologna, Italy (previously Industrial Crop Research Centre, Agricultural Research Council). They were provided in air-free sealed vials due to their possible hygroscopic nature. Intact GLs were isolated as their potassium salts.

NMR spectra were recorded in D_2O at 298 K with TSP-*d4* as internal reference on i) a 500 MHz Bruker Avance III spectrometer equipped with a 5 mm BBFO+ probe; and ii) a 600 MHz Bruker Avance III spectrometer equipped with a 5 mm TCI cryoprobe, both driven by the TopSpin NMR software (v2.1), which was also used for data processing. Spectra were recorded by means of standard Bruker pulse sequences, namely cosygppr for COSY, hsqcedetgpsisp2.2 for HSQC, hmbcgplpndqf for ^{1}H-^{13}C HMBC and hmbcf3gpndqf for ^{1}H-^{15}N HMBC

Conformational analysis was carried out by means of the Macromodel software, with the MM3 force-field and water as implicit solvent, using default parameters for starting geometry generation and for atom coordinate optimization.[35] The handling of MOL and SD files, atom labelling, NMR parameter storage, and the determination of atom and bond configuration were carried out using the OEChem ToolKit.[36]


a)
HOD
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
ppm
H19
H22
H18
H20
H1
H8R
H8S
H6R
H6S
H2
H3
H4
H5
7.4
7.3
7.2
7.1
ppm
4.72
ppm
4.15
ppm
3.6
3.4
ppm
b)
C7
C21
C17
C19
C18
C22
C20
C5
C3
C4
C2
C1
C24
C6
C8
160
150
140
130
120
110
100
90
80
70
60
50
ppm


c)
H19
H22
H18
H20
ppm
C20
C22
C18
C19
22
20
18
19
120
125
130
135
7.4
7.2
7.0
ppm
H1
H8R
H8S
H24
H6R
H6S
H4
H2
H3
H5
ppm
C8
C24
C6
C4
C2
C3
C5
C1
8
24
6
4
2
3
5
1
45
50
55
60
65
70
75
80
85
4.6
4.4
4.2
4.0
3.8
3.6
3.4
3.2
ppm
d)
H19
H22
H18
H20
ppm
19-20
19-18
7.0
7.2
7.4
7.4
7.2
7.0
ppm
H1
H8R
H8S
H24
H6R
H6S
H4
H2
H3
H5
ppm
4-3
5-4
6R/6S-5
8R-8S
1-2
3.2
3.4
3.6
3.8
4.0
4.2
4.4
4.6
4.6
4.4
4.2
4.0
3.8
3.6
3.4
3.2
ppm

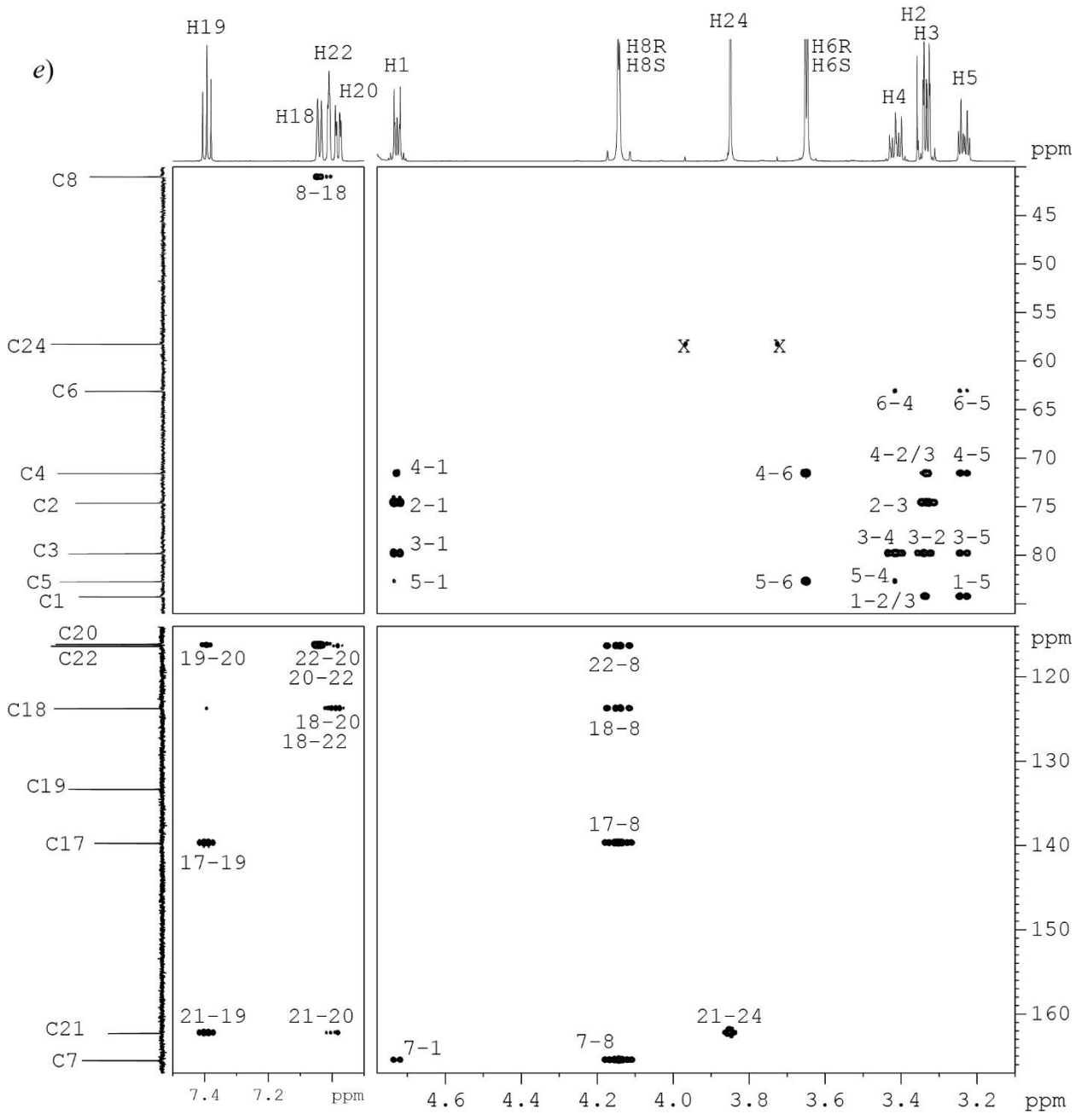


Figure 1. a) ^{1}H (top: overview, bottom: details), b) ^{13}C, c) ^{1}H-^{13}C HSQC, d) ^{1}H-^{1}H COSY, and e) ^{1}H-^{13}C HMBC NMR spectra of glucolimnanthin (GLI) recorded at 600 MHz for ^{1}H and 150 MHz for ^{13}C. Atoms are numbered according to Chart 1. The * symbols in the HSQC spectrum (c) indicate the correlations inside of the CH_2 groups, at C-6 and C-8. The X symbols in the HMBC spectrum (e) indicate residual direct coupling signals.

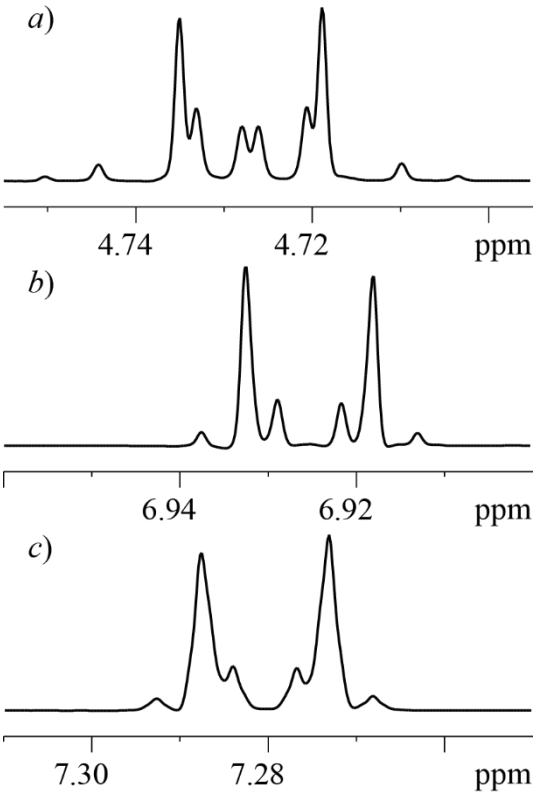


Figure 2. Expansions of the ^{1}H NMR spectrum of sinalbin, (SNB) recorded at 600 MHz, showing the effect of strong coupling on a) H-1, b) on chemically equivalent H-19 and H-21, c) on chemically equivalent H-18 and H-22.

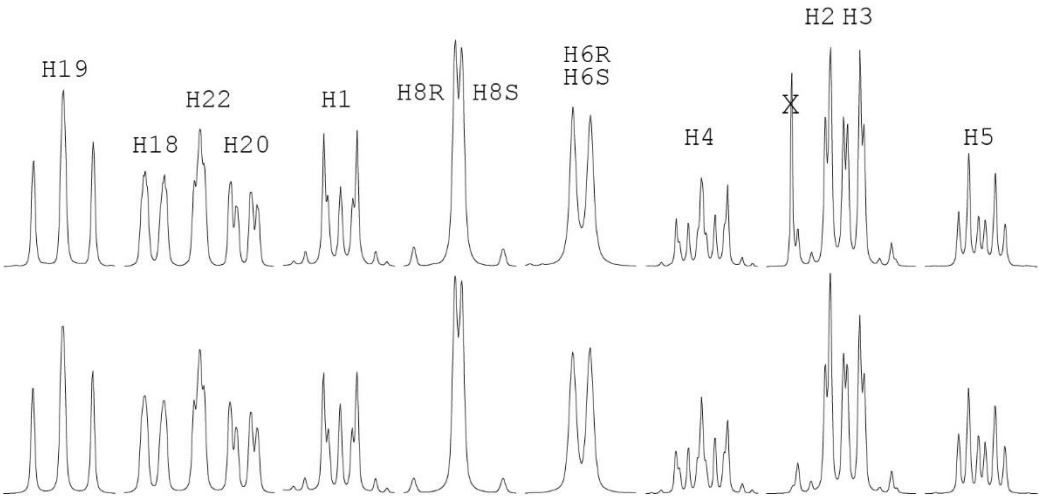


Figure 3. Top: ^{1}H NMR signals of glucolimnanthin (GLI), recorded at 600 MHz, and their assignment, according to numbering in Chart 1. Chemical shifts have not been reported for clarity but can be read in Figure 1. The signal marked with X arises from residual methanol. Bottom: signals reconstructed by PERCH, according to the chemical shift and coupling values accessible in file GLs.txt through the web link provided in the Supplementary Information file.

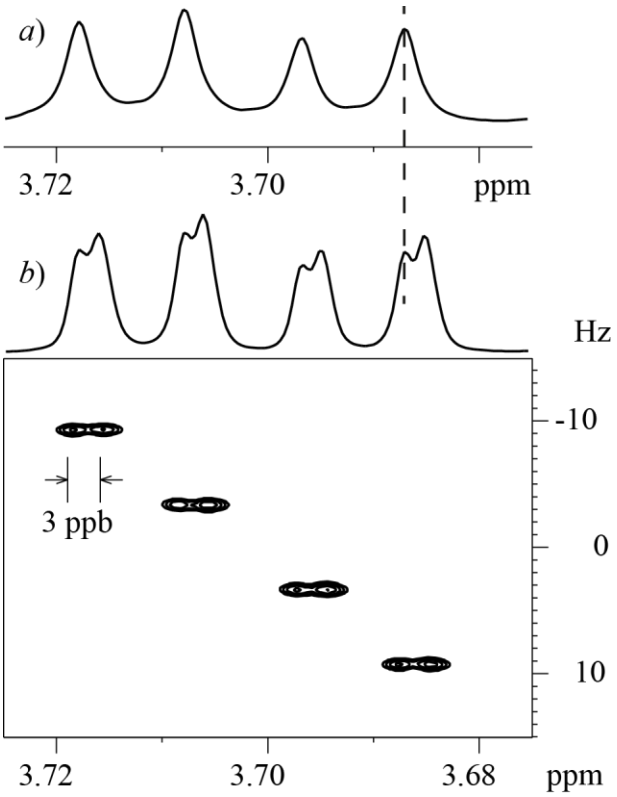


Figure 4. a) ^{1}H NMR signal of H-6(*pro-R*) in desulfo-glucoraphanin (DS-GRA). b) 2D *J*-resolved spectrum, non-tilted, of (*RS*)-DS-GRA showing that H-6(*pro-R*) has different chemical shifts in the (*R*)-DS-GRA and (*S*)-DS-GRA epimers.

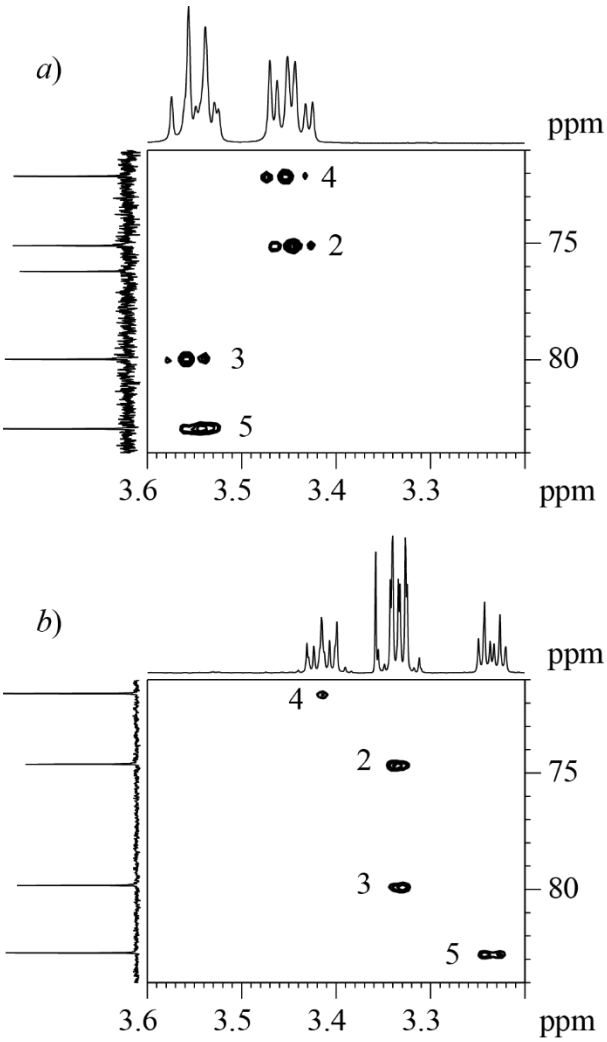


Figure 5. Parts of the HSQC spectrum of a) desulfoglucocleomin (DS-GCL) and b) of glucolimnanthin (GLI) that show typical patterns of ^{1}H NMR chemical shift distributions for H-2, H-3, H-4 and H-5 in (DS)GLs presenting non-benzylic (like in DS-GCL) and benzylic (like in GLI) sidechains.

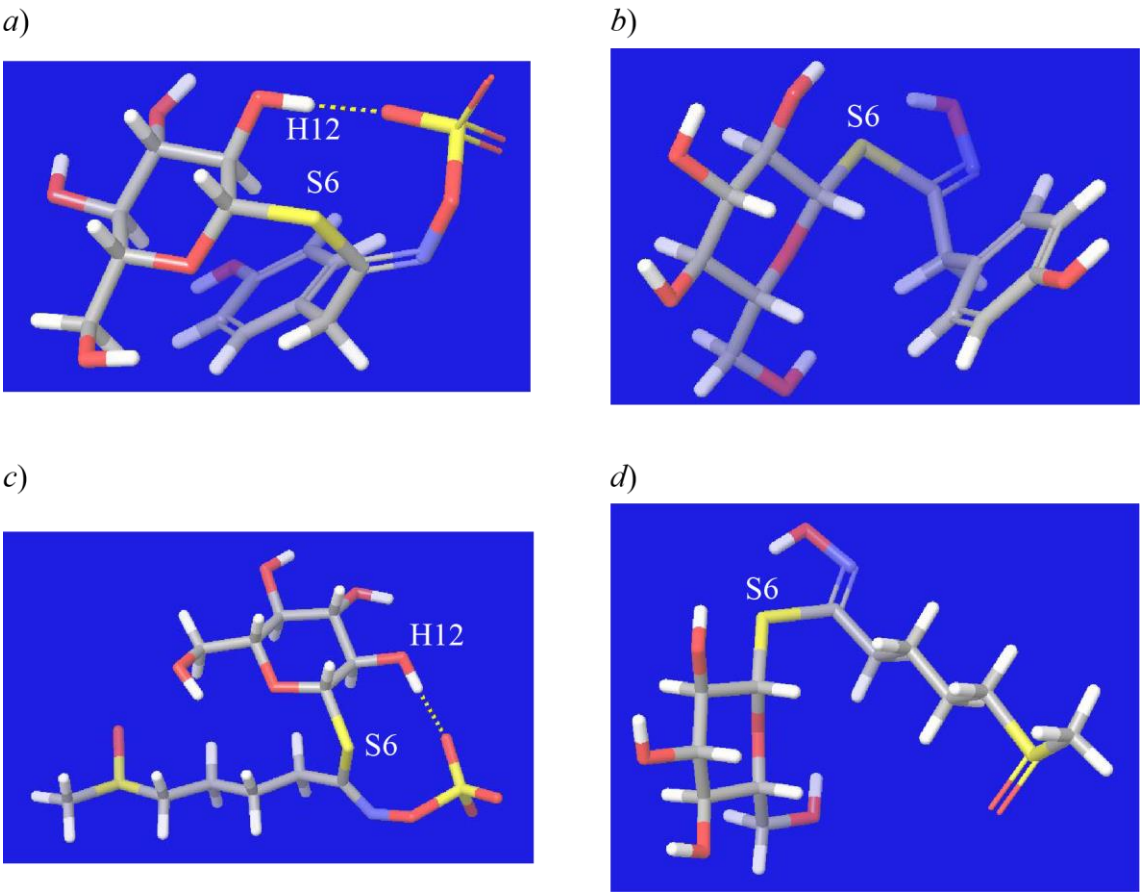


Figure 6. Lowest-energy conformers of a) sinalbin (SNB) and b) desulfosinalbin (DS-SNB) showing the conformational change induced by loss of the sulfate group, c) and d) as a) and b) for glucoraphanin (GRA) and desulfoglucoraphanin (DS-GRA). The dashed yellow lines indicate the hydrogen bonds between the sulfate group and H-12 (from 2-OH group) in SNB and GRA. The hydrogen bonds impose a particular turn around S-11 that is not present in DS-GLs.

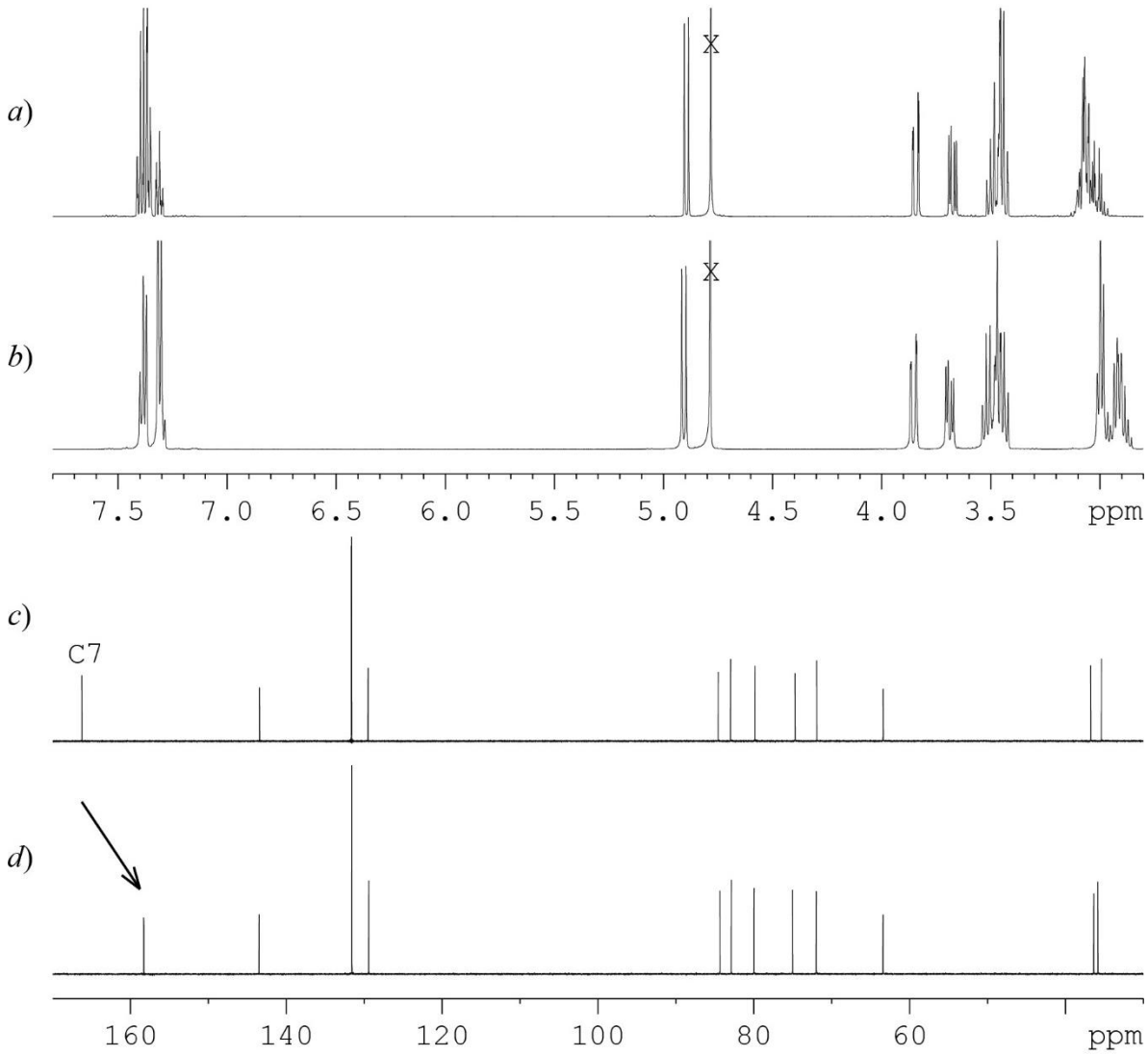
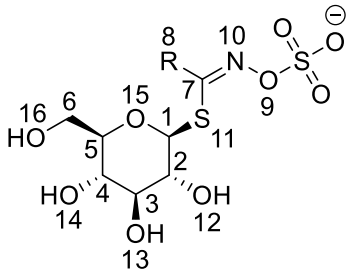


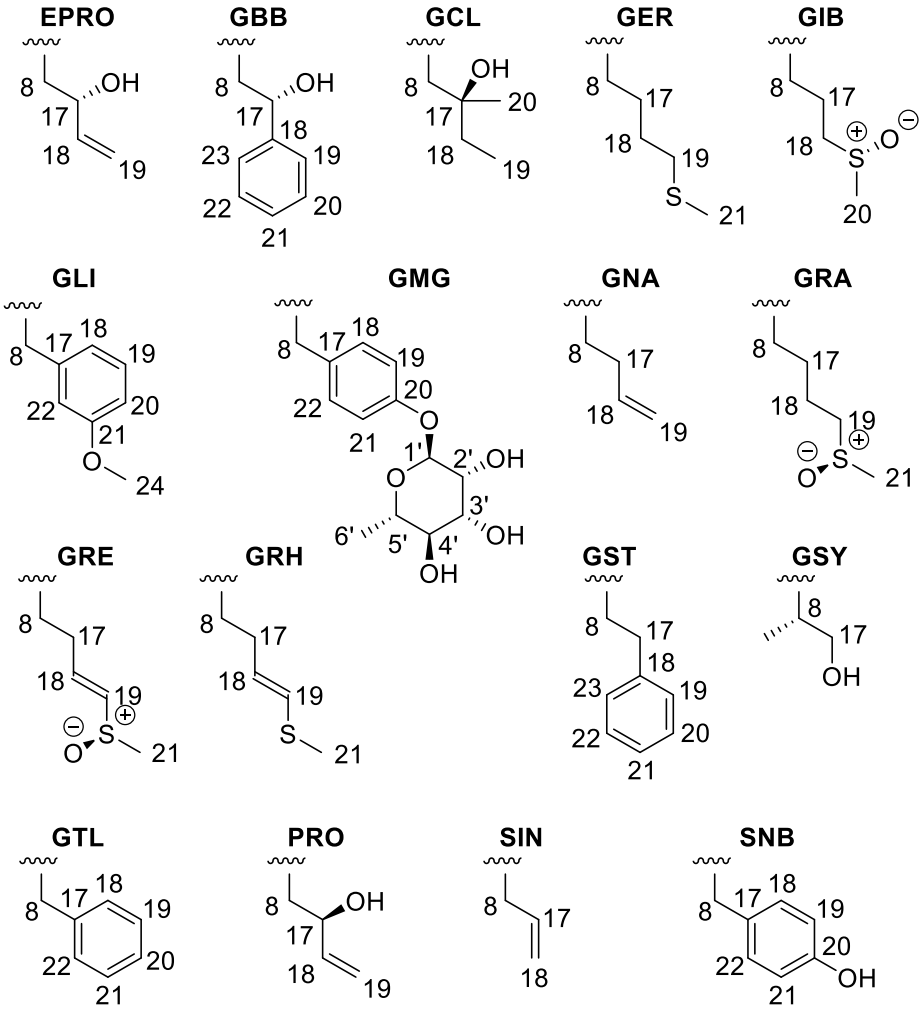
Figure 7. 1D NMR spectra of gluconasturtiin (GST) and its desulfated analogue (DS-GST) in D_2O. a) and b) comparison of the ^{1}H NMR spectra of GST and DS-GST (residual HOD signals marked with X), c) and d) comparison of the ^{13}C NMR spectra of GST and DS-GST.

Chart 1. a) Common structural part of glucosinolates and b) sidechains (R group) of the studied compounds, referenced by the code names. See text for an explanation about atom numbering.

a)



b)



Code	Name	GL	DS-GL	Botanical origin
EPRO	epi-progoitrin	Yes	Yes	*Crambe abyssinica*
GBB	glucobarbarin	Yes	Yes	*Barbarea vulgaris*
GCL	glucocleomin	No	Yes	*Cleome hassleriana*
GER	glucoerucin	Yes	Yes	*Eruca sativa*
GIB	glucoiberin	Yes	Yes	*Iberis amara*
GLI	glucolimnanthin	Yes	Yes	*Limnanthes alba*
GMG	glucomoringin	Yes	Yes	*Moringa oleifera*
GNA	gluconapin	Yes	Yes	*Brassica rapa*
GRA	glucoraphanin	Yes	Yes	*Brassica oleracea* var. *acephala*
GRE	glucoraphenin	Yes	No	*Raphanus sativus*
GRH	glucoraphasatin	Yes	Yes	*Raphanus sativus* var. *longipinnatus*
GST	gluconasturtiin	Yes	Yes	*Barbarea verna*
GSY	glucosisymbrin	Yes	No	*Sisymbrium loeselii*
GTL	glucotropaeolin	Yes	Yes	*Lepidium sativum*
PRO	progoitrin	Yes	Yes	*Brassica napus*
SIN	sinigrin	Yes	Yes	*Brassica jun*cea
SNB	sinalbin	Yes	Yes	*Sinapis alba*

Table 1. List of the Studied GLs and DS-GLs. Each Sidechain (Chart 1b) Refers to a Code and to a GL Usage Name. The “No” Entries Correspond to Compounds Unavailable to Us.

Table 2. Excerpt of the GLs.txt NMR Data File Fhowing the Similarity and Differences (Highlighted with Bold Face Text) of Chemical Shifts in Sinigrin (SIN) and Desulfosinigrin (DS-SIN).

SIN

Position	Chemical shift	H	Chemical shift	Couplings...			
C-1	84.5	H-1	5.0441	J(H-1,H-2) = 9.9432			
C-2	74.8869	H-2	3.4329	J(H-2,H-1) = 9.9432	J(H-2,H-3) = 9.0086		
C-3	80.0026	H-3	3.5208	J(H-3,H-2) = 9.0086	J(H-3,H-4) = 9.2094		
C-4	72.0782	H-4	3.4479	J(H-4,H-3) = 9.2094	J(H-4,H-5) = 9.9274		
C-5	82.9627	H-5	3.5333	J(H-5,H-4) = 9.9274	J(H-5,H-6S) = 2.2658	J(H-5,H-6R) = 5.8946	
C-6	63.5847	H-6R	3.7049	J(H-6R,H-5) = 5.8946	J(H-6R,H-6S) = -12.6271		
		H-6S	3.8992	J(H-6S,H-5) = 2.2658	J(H-6S,H-6R) = -12.6271		
C-7	**165.9459**						
C-8	39.0654	H-8R	3.5207	J(H-8R,H-8S) = -16.6555	J(H-8R,H-17) = 6.5648	J(H-8R,H-18E) = -1.3529	J(H-8R,H-18Z) = -1.4767
		H-8S	3.4906	J(H-8S,H-8R) = -16.6555	J(H-8S,H-17) = 5.211	J(H-8S,H-18E) = -1.822	J(H-8S,H-18Z) = -1.8945
N-10	**345.65**						
C-17	135.0021	H-17	6.0193	J(H-17,H-8S) = 5.211	J(H-17,H-8R) = 6.5648	J(H-17,H-18E) = 10.3333	J(H-17,H-18Z) = 17.2712
C-18	121.355	H-18Z	5.3119	J(H-18Z,H-8S) = -1.8945	J(H-18Z,H-8R) = -1.4767	J(H-18Z,H-17) = 17.2712	J(H-18Z,H-18E) = 1.1731
		H-18E	5.2685	J(H-18E,H-8S) = -1.822	J(H-18E,H-8R) = -1.3529	J(H-18E,H-17) = 10.3333	J(H-18E,H-18Z) = 1.1731

DS-SIN

Position	Chemical shift	H	Chemical shift	Couplings...			
C-1	84.0931	H-1	5.0242	J(H-1,H-2) = 9.9329			
C-2	74.9996	H-2	3.4337	J(H-2,H-1) = 9.9329	J(H-2,H-3) = 8.9363		
C-3	79.9785	H-3	3.5357	J(H-3,H-2) = 8.9363	J(H-3,H-4) = 9.189		
C-4	72.0432	H-4	3.4614	J(H-4,H-3) = 9.189	J(H-4,H-5) = 9.963		
C-5	82.7736	H-5	3.54	J(H-5,H-4) = 9.963	J(H-5,H-6S) = 2.2567	J(H-5,H-6R) = 5.9142	
C-6	63.4977	H-6R	3.7189	J(H-6R,H-5) = 5.9142	J(H-6R,H-6S) = -12.5835		
		H-6S	3.9125	J(H-6S,H-5) = 2.2567	J(H-6S,H-6R) = -12.5835		
C-7	**157.7195**						
C-8	38.6915	H-8R	3.3972	J(H-8R,H-8S) = -17.0994	J(H-8R,H-17) = 5.2857	J(H-8R,H-18E) = -1.8503	J(H-8R,H-18Z) = -1.9344
		H-8S	3.4244	J(H-8S,H-8R) = -17.0994	J(H-8S,H-17) = 6.4746	J(H-8S,H-18E) = -1.2706	J(H-8S,H-18Z) = -1.5677
N-10	**337.504**						
C-17	135.7516	H-17	5.9994	J(H-17,H-8S) = 6.4746	J(H-17,H-8R) = 5.2857	J(H-17,H-18E) = 10.2813	J(H-17,H-18Z) = 17.2954
C-18	120.6635	H-18Z	5.2705	J(H-18Z,H-8S) = -1.5677	J(H-18Z,H-8R) = -1.9344	J(H-18Z,H-17) = 17.2954	J(H-18Z,H-18E) = 1.2666
		H-18E	5.2419	J(H-18E,H-8S) = -1.2706	J(H-18E,H-8R) = -1.8503	J(H-18E,H-17) = 10.2813	J(H-18E,H-18Z) = 1.2666

ASSOCIATED CONTENT

Supporting Information. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jnatprod.xxxxxxx. It describes the contents of the files that can be freely downloaded from http://doi.org/10.5281/zenodo.1069439, namely the NMR acquisition files, the SDF files and the NMR parameter files in text, Excel, and SQLite formats (PDF).

AUTHOR INFORMATION

Corresponding Author

*jm.nuzillard@univ-reims.fr

ORCID: 0000-0002-5120-2556

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENT

Financial support by CNRS, Conseil Regional Champagne Ardenne, Conseil General de la Marne, Ministry of Higher Education and Research (MESR) and EU-programme FEDER to the PlAneT CPER project is gratefully acknowledged. We thank Dr. Matthias Niemitz, PERCH Solutions Ltd., for his advices on the use of the PERCH software.

REFERENCES

(1) Blažević, I.; Montaut, S.; Burčul, F.; Rollin, P. Glucosinolates: Novel Sources and Biological Potential. In *Reference Series in Phytochemistry. Glucosinolates*, 1st edition, 2017, K. G. Ramawat & J.-M. Mérillon Eds. / Springer International Publishing, Cham (CH), pp. 3-60.

(2) Agerbirk, N.; Olsen, C. E. *Phytochemistry* **2012**, *77*, 16-45.

(3) Giacoppo, S.; Galuppo, M.; Iori, R.; De Nicola, G. R.; Bramanti, P.; Mazzon, E. *Eur. Rev. Med. Pharmacol. Sci.* **2014**, *18*, 194–204.

(4) Giacoppo, S.; Rajan, T. S.; De Nicola, G. R., Iori, R.; Rollin, P.; Bramanti, P.; Mazzon, E. *Rejuvenation Res.* **2017**, *20*, 50–63.

(5) Mocali, S.; Landi, S.; Curto, G.; Dallavalle, E.; Infantino, A.; Colzi, C.; d'Errico, G.; Roversi, P. F.; D'Avino, L.; Lazzeri, L. *Ind. Crops Prod.* **2015**, *75*(Part_A), 79–90.

(6) Brown, P. D.; Morra, M. J. *Adv. Agron.* **1997**, *61*, 167–231.

(7) Kliebenstein, D. J.; Kroymann, J.; Mitchell-Olds, T. *Curr. Opin. Plant Biol.* **2005**, *8*, 264–71.

(8) Jeschke, V.; Kearney, E. E; Schramm, K.; Kunert, G.; Shekhov, A.; Gershenzon, J.; Daniel G. Vassão, D. G. *Front. Plant Sci.* **2017**, *8*, 1995.

(9) Wathelet, J-P.; Iori, R.; Leoni, O.; Rollin, P.; Quinsac, A.; Palmieri, S. *Agroindustria* **2004**, 3, 257−266.

(10) Laatikainen, R.; Niemitz, M.; Weber, U.; Sundelin, J.; Hassinen, T.; Vepsäläinen, J. *J. Magn. Reson. A* **1996**, *120*, 1–10.

(11) Jeannerat, D. *Magn. Reson. Chem.* **2017**, *55*, 7–14.

(12) https://www.cas.org/products/scifinder

(13) Walter, W.; Schaumann, E. *Synthesis* **1971**, 111–130.

(14) Vergara, F.; Wenzler, M.; Hansen, B. G.; Kliebenstein, D. J.; Halkier B. A.; Gershenzon, J.; Schneider, B.; *Phytochemistry* **2008**, *69*, 2737–2742.

(15) Iori, R.; Bernardi, R.; Gueyrard, D. ; Rollin, P. ; Palmieri, S. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 1047-1048.

(16) Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I; Grier, D. L; Leland, B. A.; Laufer, J. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 244–255.

(17) https://www.sqlite.org/

(18) Silverstein, R. M.; Webster, F. X.; Kiemle, D. J.; Bryce D. L. *Spectrometric Identification of Organic Compounds*; John Wiley and Sons: 2014.

(19) Bross-Walch, N.; Kühn, T.; Moskau, D.; Zerbe, O. *Chem. Biodiversity* **2005**, *2*, 147–177.

(20) Harris, R. K.; Becker, E. D.; Cabral de Menezes, S. M.; Granger, P.; Hoffman, R. E.; Zilm, K. W. *Pure Appl. Chem.* **2008**, *80*, 59–84.

(21) Bax, A.; Subramanian, S. *J. Magn. Reson.* **1986**, *67*, 565–569.

(22) Alum, M. F.; Shaw, P. A.; Sweatman, B. C.; Ubhi, B. K.; Haselden, J. N; Connor, S. C. *Metabolomics* **2008**, *4*, 122–127.

(23) Montaut, S.; Zhang, W.-D.; Nuzillard, J.-M.; De Nicola, G. R.; Rollin P. *J. Nat. Prod.* **2015**, *78*, 2001−2006

(24) (a) Mihaleva, V. V.; te Beek, T.A.; van Zimmeren, F.; Moco, S.; Laatikainen, R.; Niemitz, M.; Korhonen, S. P.; van Driel, M. A.; Vervoort, J. *Anal. Chem.* **2013**, *85*, 8700-8707. (b) Pauli G. F.; Niemitz, M.; Bisson, J.; Lodewyk, M. W.; Soldi, C.; Shaw, J. T.; Tantillo, D. J.; Saya, J. M.; Vos, K.; Kleinnijenhuis, R. A.; Hiemstra, H.; Chen, S.-N.; McAlpine, J. B.; Lankin, D. C; Brent Friesen, J. *J. Org. Chem.* **2016**, *81*, 878–889. (c) Napolitano, J. G.; Gödecke, T.; Rodríguez-Brasco, M. F.; Jaki, B. U.; Chen, S. N.; Lankin, D. C.; Pauli, G. F. *J. Nat. Prod.* **2012**, *75*, 238–248. (d) Melchor-Martínez, E. M.; Silva-Mares, D. A.; Torres-López, E.; Waksman-Minsky, N.;, Pauli, G. F.; Chen, S. N.; Niemitz, M.; Sánchez-Castellanos, M.; Toscano, A.; Cuevas, G.; Rivas-Galindo, V. M. *J. Nat. Prod.* **2017**, *80*, 2252–2262.

(25) Nishida, Y.; Hori, H.; Ohrui, H.; Meguro, H. *J. Carbohydr. Chem.* **1998**, *7*, 239–250.

(26) Laatikainen, R.; Hassinen, T.; Lehtivarjo, J.; Tiainen, M.; Jungman, J.; Tynkkynen, T.; Korhonen, S.-P.; Niemitz, M.; Poutiainen, P.; Jääskeläinen, O.; Väisänen, T.; Weisell, J.; Soininen, P.; Laatikainen, P.; Martonen, H.; Tuppurainen, K. *J. Chem. Inf. Model.* **2014**, *54*, 419–430.

(27) Lehtivarjo, J.; Niemitz, M.; Korhonen, S.-P. *J. Chem. Inf. Model.* **2014**, *54*, 810–817.

(28) https://www.w3.org/XML/

(29) Hubert, J. ; Nuzillard, J.-M.; Renault J.-H. *Phytochem. Rev.* **2017**, *16*, 55–95.

(30) Nuzillard, J.-M. *eMagRes* **2014**, *3*, 1–7.

(31) http://pubs.acs.org/paragonplus/submission/jnprdf/jnprdf_authguide.pdf

(32) http://nmrml.org/

(33) http://www.allotrope.org/

(34) http://nmredata.org/

(35) Schrödinger Release 2015-2: MacroModel, Schrödinger, LLC, New York, NY, 2015. https://www.schrodinger.com/

[36) OEChem, v2015.Jun.5, OpenEye Scientific Software, Inc.: Santa Fe, NM, USA. https://www.eyesopen.com/

Graphical TOC

