Essential Terminology Connects NMR and qNMR Spectroscopy to Its Theoretical Foundation

Guido F. Pauli[1a,1b,2,$,*], G. Joseph Ray[1,2,5,$], Anton Bzhelyansky[3], Birgit U. Jaki[2], Charlotte Corbett[4], Christina Szabo[5], Christoph Steinbeck[6], Dan Sørensen[7], Damien Jeannerat[8], Daneel Ferreira[9], David C. Lankin[1,2], James B. McAlpine[1,2], Jean-Nicolas Dumez[10], Jonathan Bisson[1ab,2], Krish Krishnamurthy[11], Matthias Niemitz[12], Michael A. Nelson[13], Patrick Giraudeau[10], Samuli-Petrus Korhonen[12], Stefan Kuhn[14], Toru Miura[15], Lucy Botros[3]

[1a]Pharmacognosy Institute and [1b]Institute of Tuberculosis Research, College of Pharmacy, University of Illinois at Chicago, Chicago, IL 60612, United States

[2]Department of Pharmaceutical Sciences, College of Pharmacy, University of Illinois at Chicago, Chicago, IL 60612, United States

[3]The United States Pharmacopeial Convention (USP), Rockville, MD 20852, United States

[4]Drug Enforcement Administration, Dulles, VA, United States

[5]Baxter Healthcare, 25212 W IL Route 120, Round Lake, IL, 60073, United States

[6]Institute for Inorganic and Analytical Chemistry, Friedrich-Schiller University, Lessingstrasse 8, 07743 Jena, Germany

[7]Health Canada, Toronto, ON, Canada

[8]NMRprocess.ch, CH-1200, Geneva, Switzerland

[9]National Center for Natural Products Research, and Department of Biomolecular Sciences, Division of Pharmacognosy, School of Pharmacy, University of Mississippi, University, Mississippi 38677, United States

[10]Université de Nantes, CNRS, CEISAM UMR 6230, F-44000 Nantes, France

[11]Chempacker LLC, San Jose, CA, United States

[12]NMR Solutions Ltd, Puijonkatu 24 B 5, 70110 Kuopio, Finland

[13]The National Institute of Standards and Technology (NIST), Gaithersburg, MD 20899-8392, United States

[14]Cyber Technology Institute, De Montfort University, Leicester LE23WQ, UK

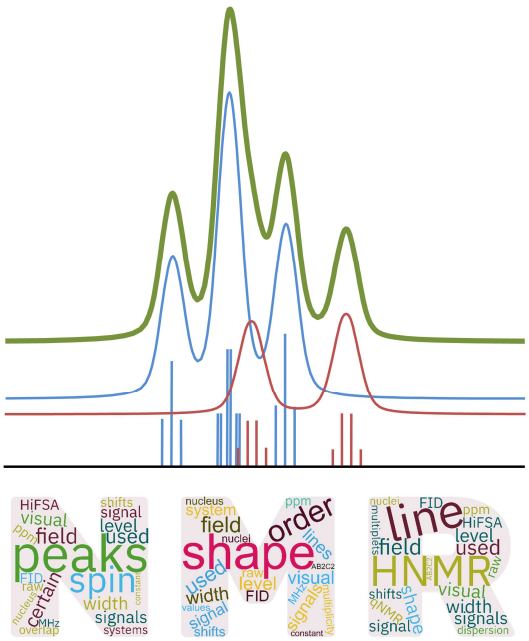
[12]FUJIFILM Wako Pure Chemical Corporation, Kawagoe 350-1101, Japan

[$]Equal contributors

Correspondence

Guido F. Pauli, gfp@uic.edu

ABSTRACT



Classical 1D ^{1}H NMR spectra are prototypic for NMR spectroscopy in that they represent a wealth of chemical information encoded into convoluted graphs or patterns that contain complex features (aka multiplets), even for seemingly simple molecules. Accordingly, the utility of NMR depends on the theoretical and visual skills required to extract all the physical parameters that represent usable structural and quantitative information. Moreover, it depends on the ability of the analyst to communicate them effectively and reproducibly. After decades of continuous development, NMR spectroscopy has reached a stage where its analytical capabilities have outgrown the typical level of detail of interpretation, especially of 1D NMR spectra. The quantum-mechanical (QM) foundation, history, evolution, and (in-)consistency of widely applied terminology calls for re-examination and recalibration. In order to develop new perspectives on solution-state NMR analysis, including the rapidly evolving quantitative NMR (qNMR), the present study draws on the well-established NMR model systems and molecules (AB_2C_2, strychnine, testosterone, α-santonin). Through well-documented key topics related to spectral acquisition and analysis, the study builds the foundation for a modular, coherent, and standardized nomenclature of NMR terminology. This is a necessary condition for a healthy research data lifecycle including their management and reuse. This work presents experimental evidence and connects with essential concepts of QM theory that clarify the distinct meaning of the primary terms: resonance, signal, pattern, peak, line, transition; as well as other widely used terms: splitting, multiplicity/multiplet, resolution, and dispersion. The proposed NMR terminology was built through a consensus-finding process that evolved from extended pharmacopoeial and research coordination efforts. It is supported by detailed figures and NMR data interpretation that employs QM-based full spin analysis.

KEYWORDS

NMR resonance, signal, peak, line, transition, resolution, dispersion, multiplets, multiplicity, splitting, nomenclature

1 | INTRODUCTION

NMR spectra are complex entities (**Figure 1**) that require a combination of theoretical knowledge, visual recognition, and well-conditioned analytical thinking for interpretation. In fact, the mother of all NMR experiments, i.e., the 1D ^{1}H NMR (HNMR) experiment, to which novice and experienced users alike are typically exposed from the onset of their studies, embodies a particularly high degree of complexity: the liquid (isotropic) HNMR spectra of most, including very small, molecules in solution give rise to highly convoluted spectra that make decoding the underlying molecular structure a less-than-straightforward exercise.

While the principles of quantum mechanical (QM) theory that explain how NMR works have long been understood, in particular the mechanism relating to spin-spin coupling that lead to doublet (d), triplet (t), and more complex multiplicities, the amount of information condensed in the HNMR spectrum tend to overwhelm analysts. We should keep in mind that humans are not capable of applying QM principles without the help of a computer and rely cognitively on a simplified understanding of the theory [1]. Accordingly, the designation of *multiplets*, commonly abbreviated as “m”, is sadly highly abundant if not ubiquitous in the scientific literature [2]. This is unfortunate, as all the underlying coupling and chemical information is willfully disregarded in the arbitrary *multiplet* designation.

The observed multiplicities and splittings are a function of the chemical and magnetic environment of the resonating ^{1}H nuclei, both within the analyte molecule and the solvent/solutes surrounding it. Therefore, multiplicity is a highly characteristic property of chemical structure. The plethora of structural information encoded within *multiplets* has both advantages and challenges: while higher degrees of signal multiplicity encode additional valuable structural information, the multiplets and NMR spectra as a whole do not have straightforward connections with structural motifs; unlike in IR and UV spectroscopy, NMR spectra are collections of rather unique *patterns* that tend to change in response to apparently subtle changes in the molecular structure and/or sample environment. Recent work indicates that patterns can be attributed to molecular classes in an automated manner [3].

The abundance of *multiplets* as a blanket term in publications, with the meaning of "unknown splitting pattern", reflects the likewise haphazard handling of other NMR terms that are used and accepted widely, yet may not possess a strictly agreed-upon meaning. As a result, the following terms are frequently used loosely and/or interchangeably: **resonance, signal, pattern, peak,**

line, and **transition**. Additional terms are associated with them and also used commonly, sometimes as modifiers or to impart favorable or unfavorable differences: **splitting**, **multiplicity/multiplet**, **resolution,** and **dispersion**. Following their definitions below, the present work uses all these terms in bolded font to highlight their essential nature in NMR nomenclature.

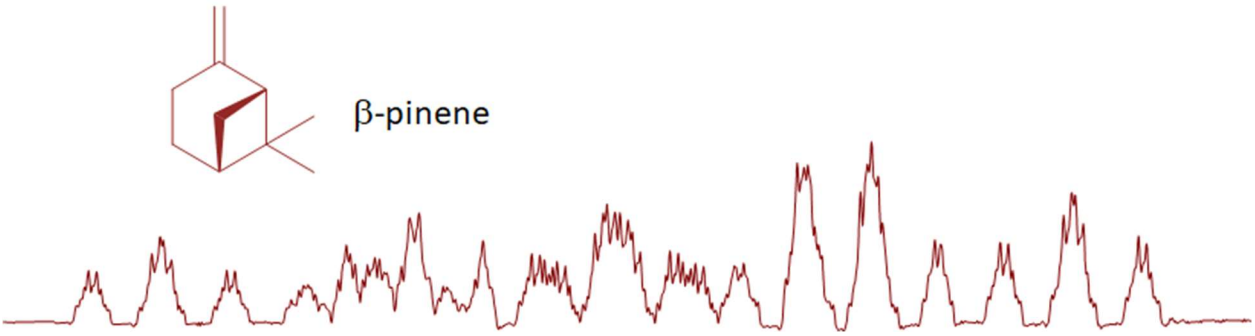


FIGURE 1 Plotted without an x-axis, this HNMR spectrum of β-pinene demonstrates the complexity of ^{1}H NMR spectra as well as the currently existing uncertainties of NMR terminology used when interpreting NMR spectra. The difficulty in formulating clear answers to the following questions exemplifies the import and breadth of this issue: Is this the entire spectrum, possibly taken at low magnetic field strength? Or is it a small segment of the spectrum? Or even a tiny portion arising from a single hydrogen, maybe acquired at ultra-high magnetic field strength? Does this plot represent a (single) resonance, (single) signal, or multiple peaks? How many peaks and how many lines does it contain? What is the meaning of the term *splitting* in this context? Is this an experimental or a calculated spectrum? For further inspiration regarding these questions and potential answers, see S1, Supplementary Material.

Depending on the magnetic field strength and the particular properties of the sample used, the same compound yields, often significantly, different NMR spectra. At the first sight, this discourages the simple visual comparison of NMR spectra and complicates the analysis of identical and structurally closely related molecules. This is due to the highly non-linear dependence between spectrum and the difference in the resonance frequencies ($\Delta\nu$) of the coupled nuclei ($\Delta\Delta\nu$) relative to their coupling (J) (see Section 3.4 for a typical example). The so-called higher-order effects appear when $|\Delta\Delta\nu/J|$ becomes sufficiently small, usually <10. This behavior makes the observed spectrum often highly sensitive to the differences between the frequencies of the resonances (*relative* resonance frequencies). While the *absolute* resonance frequencies, $\Delta\nu$ are dependent on the B_0 magnetic field (NMR macroenvironment), the ratio of the resonance frequency relative to a reference signal (typically TMS), namely $(\Delta\nu-\Delta\nu_{ref})/\Delta\nu_{ref}$, is independent of B_0 and often referred to as the *relative* chemical shifts, $\Delta\delta$, of the nuclei.

Solvent dependence, as well as other sample conditions like concentration, temperature, and pH (NMR microenvironment) often subtly change the resonance frequencies, further complicating the analysis of the NMR spectrum. This explains why identical *J* coupling networks of molecules produce NMR spectra of different overall shapes when evaluated under different solution conditions (solvents, temperature, pH) or at different magnetic fields - despite originating from (supposedly) identical molecular frameworks. However, the good news is that the *J* coupling network of a molecule is relatively highly conserved and much less variable than its chemical shift pattern.

This *Perspectives* article seeks to collect rationales for the importance of clearing up this situation. It builds evidence for a resolution, including by means of theoretical and experimental scenarios, lay out and clarify the terminology, and take the first step towards the development of a more coherent nomenclature. To make the connection between the NMR spectrum and its properties transparent, it is paramount to define all the involved terms precisely. This approach concurrently serves to explain why the QM-based analysis of NMR spectra is the ultimate solution to this apparent nomenclatural conundrum - as it applies in our current understanding of nuclear physics.

Developing coherent definitions or terms used to describe NMR spectra goes well beyond a rather formal clarification and nomenclatural exercise. As shown below, based on theoretical and experimental data, a wider adoption of fully determinant and more precise terminology has the ability to significantly advance NMR as a qualitative and quantitative (qNMR) analytical tool. For example, in the context of qNMR, the ability to clearly differentiate signals from peaks, and peaks from lines, is essential for analytical accuracy and further enables an entirely new level of opportunities for deriving quantitative values (quantitative measures [4]) from HNMR spectra. The more precise terminology can also benefit the field of NMR metabolomics, where results depend critically on how the users process and/or interpret peaks, signals, lines, multiplets, etc. Definitions of these terms are critical as this determines the size of the integration intervals/regions when bucketing and binning steps are performed prior to integration [5].

As the NMR literature pertaining to metabolomics lacks a standardization of this part of NMR terminology, key aspects of published data are user-dependent and make terminology a matter that is critical for research reproducibility. Furthermore, a cohesive framework of terms also foster the utility of modern NMR tools such as computer-aided structure elucidation (CASE) (see [6] and references therein) and automated, QM-based dereplication and interpretation of HNMR spectra [7]. Finally, a well-defined and community-agreed nomenclature is a necessary condition for

research data management (RDM) and data-reuse and is already common-place in fields such as biology [8]. With the national and continent-wide RDM infrastructures arising [9] or already existing [10], the need for research communities to develop controlled vocabularies and minimum information standards is growing.

The core topics of this study are organized into four main Sections 2-5 and followed by a Summary and Outlook in Section 6.

- Section 2 discusses the essential NMR terminology;
- Section 3 connects NMR data quality with terminology;
- Section 4 explains why terminology is a prerequisite for integrity,
- Section 5 presents detailed experimental evidence for the terminology and applies it to exemplary cases covering small molecules of up to ~2,000 amu.

2 | ESSENTIAL TERMS FOR THE INTERPRETATION OF NMR SPECTRA

2.1 | DISSECTING AN NMR SPECTRUM

The NMR instrument interacts with nuclei, such as those of hydrogen atoms, placed in a magnetic field, and measures their resonances in response. It makes them accessible as visual human-interpretable collections of more or less complex NMR time-domain data and their related spectra. While this *Perspectives* article focuses on the simplistic 1D ^{1}H NMR (HNMR) experiment in the (isotropic) liquid state, the same principles apply to heteronuclear spectra and, to an extent, to 2D/nD NMR spectra as well. This is true with the most widely practiced, pulsed Fourier Transformation NMR (FT-NMR). But time-domain based approaches such as the Bayesian statistics based CRAFT (Complete Reduction to Amplitude Frequency Table) approach [11,12], allow the analysis of raw time-domain data and are the foundation of any subsequent interpretation and constitute the true raw data [13,14], while, at least in liquid NMR, only the frequency-domain spectra are directly accessible to visual and thus cognitive human interpretation. The interpretation of frequency-domain NMR spectra is associated with a set of terms that take a certain place in the connection between the nuclear spins, as the probed matter

with certain physicochemical properties, and the NMR spectra as the corresponding human-readable output.

By critically assessing the properties of NMR spectra, the present study demonstrates that precisely defined terminology is essential for their understanding and interpretation. Among the key terms are **resonance, signal, spectrum, peak, pattern, line,** and **transition**, which form a modular set (**Figure 2**). In NMR, these terms bear strictly specific meanings occasionally disagreeing with their concurrently used definitions in other analytical disciplines, particularly chromatography.

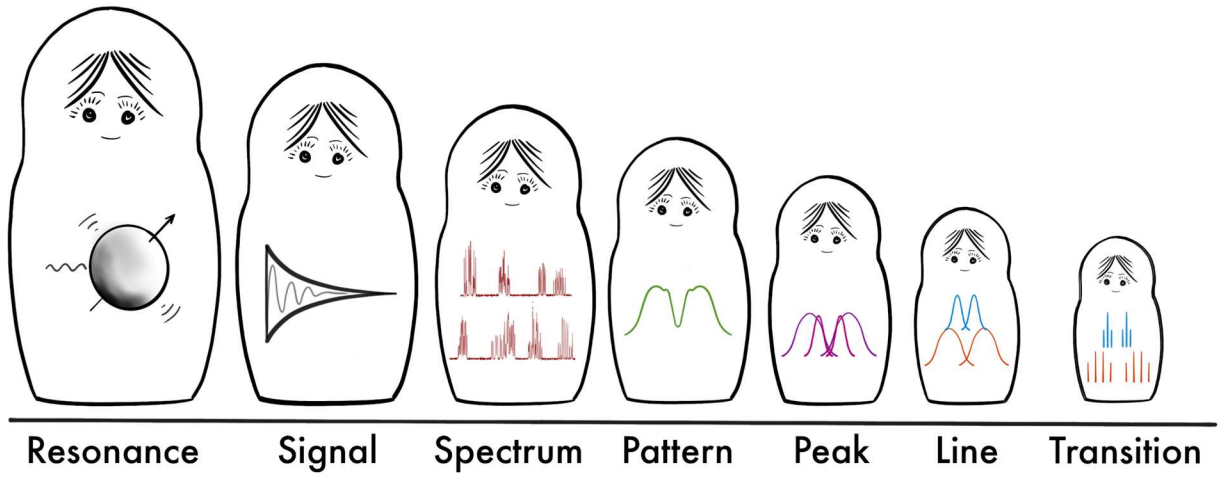


FIGURE 2 The seven key terms of NMR spectroscopy form a modular set that is reminiscent of the nested *Matryoshka* design. The definition of the seven key terms follows a logical flow that covers the generation and interpretation of NMR spectra (from left to right; not necessarily in order of granularity) and involves three types of entities: (i) **Resonance, Signal,** and **Spectrum** as *observables* of the physical phenomenon NMR; (ii) **Pattern** and **Peak** as visual ("phenotypical") descriptors that are to an extent heuristic and analysis dependent; (iii) **Line** and **Transition** as analytical concepts arising from QM-based analysis of the NMR phenomenon. This *Matryoshka* analogy is intended to emphasize the modularity and overall logic of the nomenclature. Conceptually—approaching infinite magnetic field strength—**Lines** will approximate **Transitions,** which represent the discrete frequencies that can be calculated according to a QM model of the magnetic resonance phenomenon. Considering that the QM concept of **spins** would be located below **transitions** to the right**,** but also is analogous to **resonance**, which would place them above **resonance** to the left, the *Matryoshka* modularity of NMR terminology fits strikingly with M.C. Escher's graphic art concepts (see mcescher.com).

2.2 | MODULAR AND CONSISTENT DEFINITIONS OF KEY NMR TERMS

This subsection provides a concise definition of key terms that are widely used in NMR spectroscopy. The definitions take into account the physical basis of NMR (the mathematical definitions are intentionally set aside), accepted analytical terminology, as well as the historic development and usage of the terms. Defining these terms is independent of the extent of the mathematical, physical, chemical, or biological interpretation that the NMR spectra are supposed to provide. Thus, the terms are equally important to all practitioners of NMR spectroscopy, including those that use NMR spectroscopy as part of a multi-disciplinary experimental design. While the quantitative (mathematical) foundation of NMR is well-established (the interested reader is referred to [15] and [16] as exemplary introductory texts), and whereas automated tools have become available that aid with full quantum mechanics-based spectral interpretation (see [7] and references therein), the present study remains mostly qualitative and focused on deriving a consistent and modular nomenclature that is as practical as possible.

The following seven terms establish a modular system (**Figure 2**) that covers all entities involved: starting with the physical nuclear resonance phenomenon, continuing with the visual elements of NMR spectra, and eventually leading to aspects involved in quantum mechanical interpretation of NMR spectra.

- **Resonance - Signal - Spectrum - Pattern - Peak - Line - Transition**

In support of the following definitions, **Figure 3** provides a graphical explanation of the four key terms.

Resonance: the physical phenomenon of radio frequency (measured in Hz) absorption/reemission by a given nucleus, modulated by its environment, which gives rise to its NMR **signal**.

- In the present context, **resonance** refers to the interaction between a given nucleus and the magnetic component of an external radio-frequency field and the resulting contribution to the experimental data.

- The **resonance** frequency for a particular nucleus in a molecule is a function of its Larmor frequency, which reflects the natural nuclear precession frequency for that isotope in a given magnetic field.
- The **resonance** of a particular nucleus in a molecule has a characteristic radio frequency that depends on the magnetic field to which it is subjected. This local magnetic field is impacted by the shielding arising from the specific chemical environment of that nucleus. The **resonance** radio frequency is commonly expressed as a **chemical shift:** the difference between the resonance frequency of a particular nucleus and the resonance frequency of a reference nucleus (TMS as per convention). Whereas the chemical shift is initially measured in Hz, it is usually expressed in the unit-less ppm deviation from the B_0 magnetic field by calculating the relative difference between the chemical shift and the B_0 magnetic field as follows: $10^{6*}(v\text{-}B_0)/B_0$.
- When using the term **Resonance**, it is important to specify whether it refers to an individual nucleus or to a collection of such as in a given molecule. While the frequency values of resonances of single nuclei are by definition exact, the values for an ensemble of nuclei can only be described as a statistical distribution.

Signal: the response to the electromagnetic **resonance** at a given time (in FT-NMR; single acquisition) or frequency (in CW-NMR; see [17] for recent trends), as detected in the coil of an NMR spectrometer. The signal is the combination of the frequencies of the individual energy transitions (akin to a musical polychord of tones and overtones).

Spectrum: the entirety of all detected **signals** in the Fourier Transform of the FID in the case of FT-NMR, or the entirety of all detected **signals** in CW-NMR. Decomposing the FID into individual signals using time-domain, wavelet, and other signal decomposition approaches can also produce spectra.

Pattern: the segment of a frequency-domain NMR spectrum that results from **one resonance** or from **multiple overlapping resonance(s)** and is commonly assigned to a certain **multiplicity**.

- A **pattern** may consist of a single peak or multiple **peaks**, which in turn are composed of a more or less complex collection of **lines** that belong to the **resonance(s)** (see explanation below).

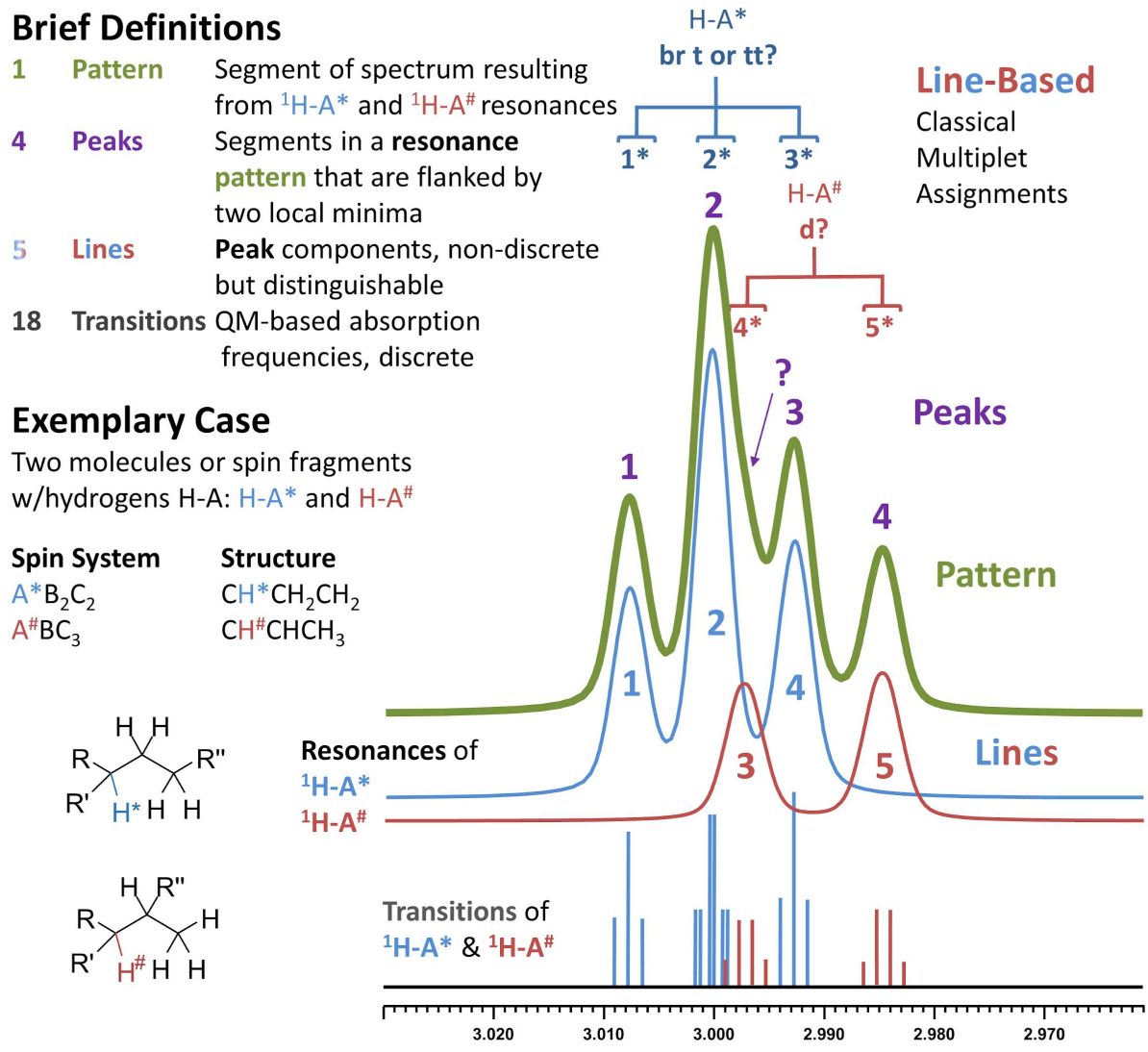


FIGURE 3 Graphical explanation of the NMR-specific terms **resonances, signals, patterns, peaks, lines, and transitions**, along with the example of a simple structural motif (see main text for details and S2., Supplementary Material, for a description of the spin matrix).

- **Multiplicity** combines the visual description of a **pattern** with the theoretical foundation of resonances by applying first-order approximation, such as the n+1 multiplets for spin-½ or 2n+1 multiplets for spin-1 nuclei.
- **Patterns**, historically often called *peak patterns* or *splitting patterns* (see definition of **peak** below), have been designated as singlets, doublets, triplets, quartets, etc. as well as doublets of doublets, doublet of triplets, etc. Despite their valid theoretical foundation (e.g., the n+1 rule for ^{1}H coupling patterns) and simplicity, the unthinking use of **multiplicity** terms holds a significant risk of oversimplification or even incorrectness due to unmet assumptions, as becomes evident from the present study.

Peak: the continuous segment of a **pattern**, which has a defined maximum and is bound by two flanking local minima. **Peaks** are composed of one or multiple **lines**.

- A **peak** *maximum* usually has a distinct location and intensity that can be determined by various means (in some cases their assessment is difficult); in NMR analysis, this process is typically called *peak picking*.
- The commonly used term **peak** *shape* is ignored in the **peak** *picking* process, as a **peak** can consist of more than one **line** each with individual **line** *shapes* (see definition of **line** below).
- In a frequency-domain spectrum, the width of a **peak** is typically the frequency range (Hz) between the two points at half height.

Line(s): the component(s) that form(s), or overlap to form, a **peak** in a frequency-domain **spectrum**. **Lines** are composed of one or more discrete **transitions**.

- **Lines** are non-discrete/continuous and, thus, have a certain width (in Hz; aka as **line** *width*) and a certain shape (aka **line** *shape*), both a result of various phenomena.
- **Lines** have Lorentzian and/or Gaussian **line** *shape* and are symmetric (in a homogeneous magnetic field, i.e., with proper shimming).
- **Lines** can contribute to multiple **peaks** (as stated above, it is still correct that **peaks** are composed of one or multiple **lines**).
- By default, **lines** arising from an individual nucleus in a molecule have the same **line** *shape* and **line** *width* (cross-correlation and scalar relaxation effects are known exceptions).
- **Line** positions are typically referred to in the frequency domain in Hz or ppm.

- **Lines** are related to the **resolution** but not to dispersion of NMR spectra (see separate discussion below).
- The number and relative location of **lines** are determined by quantum mechanical **transitions**, which ultimately determine the complexity of **peak patterns** or **splitting patterns**.

Transition: the nuclear absorption/reemission of a photon of a given *discrete* radio frequency, it is determined by the Spin Hamiltonian function in quantum mechanics (difference in Eigenvalues of the spin Hamiltonian for two Eigenstates that differ by Delta m = ± 1).

- As opposed to **lines**, which are continuous and have a certain **line** *shape*, **transitions** are the discrete energy differences between two different spin states.
- Thus, a transition does not have a certain width (in Hz), but does have a discrete frequency in Hz.

It is important to note the *discrete* nature of **transitions** vs. the *non-discrete* (aka continuous) nature of **lines**, **peaks**, and **patterns**. The latter are observable elements of the experimental NMR **spectra**, and the contemporary borderline of experimental resolution is between **peaks** and **lines**. Reflecting the underlying QM theory, **transitions** are not observed but can be computed and form the basis of calculated (aka theoretical) **spectra.** One intrinsic value of the QM model in NMR is that it allows the dissection of fully degenerate (not degenerated) **lines** in observed NMR **spectra**, i.e., **lines** that originate from different spin particles but still have the same frequency (isochronic) and, thus, are indistinguishable by classical (visual) analysis. This enables the accurate determination of the essential NMR parameters, even in the presence of severe **peak**/**line** overlap and higher-order effects.

While QM can fully describe the NMR phenomenon, it is usually necessary to use a simplified model to keep the computations tractable. By using a small number of representative species (usually called spins, forming a spin system), it is possible to expediently compute an approximate spectrum with extremely good agreement with the observation using a limited number of transitions. However, due to the approximation it is usually necessary to apply a line width and shape to the transitions to prevent artefacts when constructing lines. Calculated spectra can be used for iterative data interpretation processes such as ^{1}H iterative Full Spin Analysis (HiFSA), which systematically varies the parameters used to obtain the calculated spectrum until it precisely matches the experimental spectrum.

Now that the components of the spectrum have been defined, we can clearly define two related key terms: **Resolution** and **Dispersion**.

Resolution: Refers to the degree of distinguishability of peaks or lines (depending on the context) and reflects the relationship between peak or line width, shape, and relative location on the x-axis of the frequency-domain spectrum.

- The actual or achievable **resolution** of an NMR spectrum (sometimes called *spectral* **resolution**) is a result of many factors that work in concert: the proper operation of the spectrometer and careful sample preparation, including adequate field and pulse homogeneity, nuclear relaxation conditions, and the properties of the sample such as solvent, pH, viscosity, and homogeneity.
- In contrast, the term *digital* **resolution** refers to the density of the (digital) data points in the spectrum, which is a parameter controlled by the dwell time of the originally acquired time-domain data points (ignoring over-sampling for the sake of simplicity) and any points added to the end of the time-domain data (zero filling), or as part of special processing methods such as non-linear processing or when transforming non-uniform sampling (NUS) data.
- Importantly, while digital resolution has to be adequate for the achievable level of detail (modern NMR spectrometers can resolve **lines** as close as 0.2 Hz), it cannot replace actual **resolution** of the spectrum, which is a function of many parameters and optimized instrument settings. For a reasonable description of the line shape, a digital resolution of at least 10 data points per line width is recommended [2].
- Usage of the term **resolution** requires contextual specification, making reference to patterns, peaks, or lines when using FT; or segments when applying CRAFT transformation [11,12].

Dispersion: Refers to the degree of separation of **resonances** and **patterns** relative to each other, on the frequency-domain scale of the spectrum. **Dispersion** is a direct linear function of the static magnetic field strength. **Spectra** with increased **dispersion** have fewer higher order effects, making them less complex due to increased ratio between the difference in frequencies of coupled resonances when compared to the magnitude of their couplings.

- While greater **dispersion** increases the separation of **resonances** and **patterns**, it should not be confused with **resolution**.

2.3 | SPLITTING IS NOT COUPLING BUT COUPLING LEADS TO SPLITTING - SPLITTING HAIRS?

Splitting implies that an entity that was previously one or uniform gets divided into two or more pieces. Whereas *splitting wood* is readily understood, the unexpected complexity of *splitting* in the context of NMR spectra becomes evident when considering the relationship of **peaks** and **lines** (**Figure 3**): a *shoulder*, as is visible between peaks 2 and 3 in **Figure 3**, can be split by, e.g., applying a window function that enhances spectral resolution. In this process, it is difficult to say whether **peak** 2 or 3 or another unknown/unassigned **peak** is split. However, this situation resolves when we consider the underlying **lines**: as also shown in **Figure 3**, the intensities and frequencies/locations of **lines** 2, 3, and 4 explain why the pattern of the overlapping **resonances** of the two hydrogens contains the shoulder at the **peak** level and fully splits at the **line** level of interpretation. This shows that precise nomenclature does *not* represent a form of the proverbial *splitting of hairs*, but is rather essential for NMR data interpretation.

Going to the QM-based level of **transitions**, it is equally important to realize that **lines** and **line** *shapes* result from both the inherent properties of the resonating nucleus and the spin system it is embedded in, which determines how many **transitions** result, how the lines are located relative to each other, and what their relative intensities are. The overlap of all transitions produces an envelope with a certain resolution that can be observed depending on the quality of the spectrum. This explains why the splitting of **patterns** can occur at both the level of **peaks** *and* **lines**, which once more highlights the ambiguity of the term **splitting**. Its proper use, therefore, falls into its context.

In the current understanding of the authors, **splitting** of HNMR spectroscopic **patterns** cannot be observed experimentally at the level of **transitions**, because **transitions** are discrete entities (i.e., have zero width), whereas NMR **spectra** are non-discrete (i.e., exhibit natural **line** *width*) and lack sufficient resolution and sensitivity to observe the individual transitions. However, it should be noted, that in the case of a true first-order spin system and at an infinitely high magnetic field strength, a resonance pattern can be designated to be a triplet, with three **peaks** that consist of three **lines** but four **transitions** (two of which are degenerate). However, if the underlying two couplings are not identical, the two **transitions** are no longer degenerate and give rise to four **lines**, which, depending on the **line** *shape* of the measurement, can lead to a doublet of doublets rather than a triplet. However, it must be noted that, strictly speaking, all triplets are doublets of

doublets, but not vice versa. Thus, a first-order analysis based on a visual inspection of the spectrum alone is inherently ambiguous.

In 1961, Roberts and Salzberg noted that the "*...first order component of a multiplicity is field independent...*", that this applies "**...when the chemical shifts are large...**", and further emphasized that "*...the higher order effects are field dependent*" [18]. The theoretical situation of a true first-order spin system may serve as an approximation, but also as a reminder of the fact that, in practice, the number of **transitions** almost always exceeds the number of **lines**, just as the number of **lines** surpasses the number of **peaks** - often by an order of magnitude between both modular stages. For an experimental example, the reader is referred to the case of strychnine presented below. Regardless of how well resolved the **spectra** are, whether **peak** patterns are sufficiently split and/or single **lines** can be distinguished, the number, intensities, and frequency distribution of the QM-based **transitions** provide the ultimate explanation for the NMR spectrum.

It is commonly said that *J* coupling leads to **peak** *splitting*. Considering the first-order triplet as an example, **peak** *splitting* appears to be an accurate verbal description of the observed **pattern**: *J*-coupling to two magnetically equivalent nuclei explains why the resonance shows three peaks with a 1:2:1 intensity ratio, thereby leading to **peak** *splitting*. However, this widespread terminology causes a major confusion as it implies that **peak** *splitting* and coupling constants are the same - which is *not* generally the case. **In fact, it cannot be overstated that frequency differences amongst peaks *are not* coupling constants and vice versa**. In other words, while *J*-coupling leads to certain **splitting patterns** (or **peak** *splitting*), it is *not* the splitting *itself*. This again highlights the importance of specific terminology when connecting *J*-couplings with **peaks**, **splittings**, and **lines**.

Fitting into this context is one piece of existing NMR knowledge that makes things more complicated - yet consistent: the early NMR literature had already recognized the phenomenon of observed **peak splitting** that cannot be explained by any existing or perceivable coupling in the molecule through the first-order n+1 rule. In these not too infrequent cases, the interaction between nuclei can lead to **transitions** and, subsequently, appearance of additional **lines** (**peak** *splitting*) as a result of mixed spin states for which no pair of nuclei alone can be identified as coupling partners. Coincidentally, the **splitting** of the **signal pattern** in these cases is deceptively similar to what would be observed for actual ("real") *J*-couplings. This was likely the rationale for the introduction of the historic term, "virtual coupling", which has led to substantial confusion as coupling is not the cause of this type of additional **peak splitting**. The explanation of this

phenomenon is, again, rooted in the QM relationships of the underlying nuclei, particularly the relationship between the resonance frequencies and the magnitude of the *J*-coupling constant. Examples of "virtual coupling" will be discussed in Section 5.3. These cases will help demonstrate the importance of distinguishing between **splittings** that can be mapped directly to *J*-coupling constants via the first-order n+1 rule vs. **splittings** that do not represent a coupling, but are the incidental **pattern** produced by the **lines** (and underlying **transitions**) of a particular nuclear **resonance** due to non-first order effects.

Collectively, it is advisable that the term **splitting** be used in NMR spectroscopy only very carefully and always in combination with the specific terms **patterns**, **peaks**, and **lines**, in order to be clear about which entity is considered to be split.

3 | FUZZINESS VS. QUALITY - DATA VS. NOMENCLATURE

3.1 | PEAK OR LINE FREQUENCY DIFFERENCES *ARE NOT* COUPLING CONSTANTS

In NMR spectroscopy, *J* coupling is the scalar interaction between nuclei that results in the resonances being represented by multiple lines. When teaching the phenomenon of *J* coupling, chemistry curricula worldwide, from high-school to graduate programs, suggest that *J* couplings can be determined from the frequency differences of **peaks** or **lines** in the various levels of **multiplets** following the *n+1* rule for ^{1}H. A common term, *peak distance,* evolved from the time when spectra were plotted, and a ruler was used to measure the distance between peaks within a multiplet. This distance was then converted into frequency units. The term, (*peak) distance*, is obsolete and should not be used.

Even when using the correct term, *frequency difference*, a number of questions remain: Does this frequency difference refer to **peaks**, **lines**, or **transitions**? How precisely does a measured frequency difference reflect the coupling constant? Connected with the latter, and considering that NMR tables throughout the literature are filled with examples where the *J* value of a pair of coupled hydrogens is listed as having two (slightly) different values: is it possible that a coupling constant is not constant, i.e., differs depending on each nucleus? Can frequency difference for

the same *J* coupling differ between the **resonances** of coupled nuclei? And if so, how can this be explained? Finally: how can coupling constants be determined from **lines** that show no evidently useful frequency differences?

The simple statement is that the frequency difference between the **lines** in a **resonance** reflect the true coupling constant only if all the frequencies of the components of the spin system giving the **resonance** follow the first-order condition of the **resonance** *frequencies* being much larger than the coupling constants involved. In all other instances, which in practice is actually the majority, the difference in the frequencies between **lines** is ***not*** the true coupling constant as second- or higher-order effects take place. Already in second order spectra, not only is the coupling constant not equal to the *line* frequency difference, but the chemical shift is not the center of the **lines**. These considerations do not only apply to multiplets, but to many signals that appear to be of relatively simple multiplicity (t, dd, ddd, etc.).

In essence, these considerations also show that the fresh minds of young NMR spectroscopists get inoculated with rather fuzzy and essentially faulty terminology. This discrepancy between teaching, theory, and practice has been discussed at length by Szantay [1]. This can explain why, over five decades following the establishment of NMR spectroscopy as a routine analytical technique in chemistry, confusion about proper NMR terminology infects the scientific literature and community. Therefore, one outcome of the present study is the notion that **peak or line frequency differences *are not* coupling constants, but just peak or line frequency differences**. Rather, coupling constants are intrinsic physical properties of the resonating nuclei and, therefore, parameters for the QM calculations of spin systems, which do not necessarily follow the first-order n+1 rule (for ^{1}H). Importantly, while a relationship does exist between the coupling constants and the **resonances** of individual nuclei, **peak**/**line** frequency differences in observed NMR spectra may(!) resemble coupling constants more or less closely, but also may not.

A common misconception in this context is that strong coupling affects only signals of spin particles for which the shift difference is not much larger than their mutual coupling. Such a situation can indeed also affect other neighboring spin particles that have a large chemical shift difference to the two close ones. This is for example seen as a distortion in the expected doublet of the alpha proton in sugars when their β and γ hydrogens are very close. Therefore, it is important to realize that the HNMR spectra of the vast majority of organic molecules have

significant elements of higher-order spin systems. In fact, it is rather challenging to identify chemicals that do not contain such elements and give rise to pure first-order HNMR spectra [28].

In summary, it is reasonable to hypothesize that this semantic shortcoming in NMR terminology is a major impediment to the understanding, utility, and further development of the technique. For example, unravelling the full information potential of the simple HNMR spectrum will be hampered until practitioners distinguish clearly between **patterns**, **peaks**, and the underlying **lines** and **transitions**, and map them more transparently to existing conventions for first-order **multiplicity** assignments. Another implication of this hypothesis is that machine-readable and, therefore, more automated forms of spectroscopic data analysis require acknowledgement of the fact that visual vs. computational interpretation have distinct requirements that are both closely tied to specific terminology. It should be kept in mind that alternative processing approaches such as CRAFT [11,12] also necessitate a more specific terminology, such as presented here.

3.2 | REPRODUCIBILITY, QUALITY, AND TRENDS

As long as sample preparation and instrument measurement conditions are comparable for two samples, NMR spectroscopy is a highly reproducible method. The standard deviation of the chemical shifts derived from the HNMR spectra of two independently prepared samples of the sesquiterpenoid lactone, α-santonin, in $CDCl_3$ can serve as an example: measured independently on two different instruments, the deviation was less than 5 ppb (0.005 ppm), equivalent to 2.5 Hz at 500 MHz, despite different sample concentrations (~50 mM in DMSO-d_6 vs. 100 mM in $CDCl_3$). It is well known that chemical shifts can vary with the solvent, pH, temperature, and sample concentration, with differences ranging from <0.1 ppm to >1ppm. For example, the chemical shifts of parthenolide in DMSO-d_6 vs. $CDCl_3$ vary by up to 0.22 ppm with a standard deviation of 0.12 ppm (data not shown). Considerably larger effects can be observed, especially for nuclei close to proton donors and acceptors at different pH. This means that chemical shifts are highly reproducible under well-controlled conditions, while posing reproducibility challenges when the exact conditions are not reported and, therefore, cannot be reproduced.

In contrast, coupling constants are highly stable entities and tend to be nearly unaffected by the above experimental conditions unless they induce a change in the structure's conformational space, which is rare. For example, the standard deviation of all(!) the coupling constants derived from the two independent HNMR data sets of α-santonin in $CDCl_3$ was <0.01 Hz (0.02 ppb[!] at 500 MHz), and <0.3 Hz for parthenolide in DMSO-d_6 vs. $CDCl_3$. These variations are smaller than

the natural line width in the corresponding spectra. Deriving NMR parameters from experimental data with this precision requires QM-based total **line** *shape* fitting, e.g. via HiFSA analysis, which takes into account the complete, continuous line shape of the **resonances**.

The very favorable accuracy and precision of chemical shifts and, in particular, *J* coupling constants derived from samples measured under similar conditions explains why their highly accurate and precise determination via HiFSA enables the generation of highly selective ^{1}H NMR profiles that are unique identifiers of most organic structures, unless they are severely ^{1}H deficient. Interestingly, when compared across different solvents, the coupling constants of a given molecule are indeed rather constant. Notably, the corresponding **peak patterns** often vary substantially due to the impact of non-first order effects. While the impact of conformational changes in conformationally more flexible systems has to be acknowledged, *J* couplings generally show constant behavior as their name implies, which, for all practical purposes, makes them important focal points in the interpretation of HNMR spectra. Similarly, the observed chemical shifts of identical or analogous hydrogens in HNMR spectra is often highly variable, thereby also fitting the designation as being a shifting entity.

3.3 | THE SMALL LB EXPONENTIAL MULTIPLICATION HABIT

Looking at the development of NMR instrumentation and practice over the last seven decades, the basic HNMR experiment is broadly affected by a habit that might appear to be a minor detail, but could represent a significant (conceptual) impediment to high-resolution(!) NMR spectroscopy: the near-ubiquitous routine application of line broadening via exponential multiplication prior to FT. Representing the typical default setting in the software of most NMR spectrometers, it has not changed since the first FFT instrument and consists of exponential multiplication (EM) of the FID (time domain NMR spectrum, raw data) with a 0.2 to 0.3 Hz (sometimes more) line broadening (LB) factor applied. This long-term habit is actually counter to the development of modern spectrometer hardware and software capability and artificially serves to degrade resultant performance.

From the historical perspective, the blanket application of EM with small LB factors is a relic of the inherent sensitivity challenge of NMR spectrometers. While relatively high-field instruments have been available for about four decades, sensitivity specifications have always been important and an object of a certain degree of contempt (as reflected by the common use of thin-walled NMR sample tubes in instrument specification). In this context, EM processing with small LB

values represents a means of enhancing signal-to-noise ratio (SNR) specifications. Notably, such SNR-based specifications abstract the NMR spectrum from the FID as the actual raw data output of the spectrometer.

However, as many modern NMR spectrometers, including entry-level cryomagnetic instrumentation (300-400 MHz), exhibit significantly improved sensitivity specifications, especially when equipped with cryoprobes, the historic habit of blanket EM processing should be reconsidered. Considering the relationship between resolution and SNR, it is important to emphasize that high-SNR HNMR spectra are more readily susceptible to resolution enhancement processing (e.g., Lorentzian-Gaussian multiplication using a negative LB). This opens an opportunity for extracting more (precise) structural and quantitative information from the same raw NMR data (FID): the application of resolution-enhancing pre-FT window functions, or the use of non-FT methods such as CRAFT, offer great potential for generating a better understanding of HNMR spectra. The resolution power (see definition above) inherent to modern NMR spectrometers is generally superb and, depending on the specifics of the molecule (rigidity/flexibility, relaxation, exchange, and other dynamic properties) and proper operation can readily resolve peaks and lines that are apart by as little as 0.2 Hz. In this context, regular instrument performance checks with the classical $CHCl_3$ lineshape sample (1% $CHCl_3$ in acetone-d_6) is a very worthwhile and even necessary exercise.

3.4 | TRADING A FEW SECONDS AND SOME BYTES FOR BETTER RESOLUTION

From the authors' collective experience with reviewing shared raw NMR data (FID) from public databases, peer review, and collaborative projects, indications are that the majority of HNMR spectra acquired today still use insufficiently short acquisition times (AQ) and/or involve FID data point sizes that are too small. Especially when employing ultra-high magnetic field strengths, the acquired number of data points must be increased proportionally to obtain adequately resolved spectra. As the resolution of an FT NMR spectrum is defined as (1/AQ) and expressed in Hz, proper setting of AQ represents a critical parameter for obtaining good-quality NMR data. However, if the AQ is too long, that is beyond the point where the FID has decayed to zero, one is simply sampling noise that decreases the signal-to-noise ratio.

If sampling the data to the end of the decay results in too few data points for good peak definition, subsequent post-acquisition zero-filling of the time-domain data (TD) is also critically important for producing proper digital definition of the spectrum. This is especially true for HNMR data where the extraction of accurate locations of **peak** and **line** frequencies as well as of frequency differences are essential for data interpretation. For practical considerations, it shall be noted that setting AQ is the most direct way to determine the digital resolution of the acquired, raw NMR data. Unlike the number of data points used to acquire the FID (TD) does not have to be a multiple of 2 to be amenable to FT.

These reflections show the importance of the proper setting of basic acquisition and processing parameters in FT-NMR spectroscopy: AQ and TD, respectively. While the experimental “costs" of such choices are nil, they are the prerequisite for adequate data interpretation and recognition of the modular characteristics (**Figures 2** and **3**) of the terminology involved with HNMR spectra.

4 | TERMINOLOGY & INTEGRITY

4.1 | TERMINOLOGY IS ESSENTIAL FOR INTEGRITY

The modular and systematic nomenclature for entities related to the understanding and interpretation of NMR **spectra**, is essential for the integrity of research and applied NMR spectroscopy. As exemplified here for the basic 1D HNMR experiment, the proposed terminology employs terms that have existed since the inception of NMR spectroscopy. Its evolution over decades has advanced the technique substantially, and focus shifted from theoretical and physical concepts to more practical and chemical applications. As a result, a certain degree of fatigue settled in with regard to the consistency of terminology and what could be considered the institutional knowledge of NMR spectroscopy. This is reflected in the general sense by the present authors, who are all long-term NMR practitioners, that widely used and basic NMR terms such as **resonance**, **signal**, **pattern**, **peak**, **line**, **transition,** and **splitting**, as well as **resolution**, **dispersion**, and **multiplicity** need to be better defined moving forward. An anticipated corollary is that updating and re-consolidating essential NMR terminology will not only enhance the integrity, but also foster the future development of the methodology altogether, including both structural/qualitative and quantitative techniques (qNMR).

4.2 | INTEGRITY IS A MATTER OF QUALITY

The aspects mentioned in Section 3, as well as in the context of the examples in the following Section 5, emphasize the crucial importance of the quality of raw NMR data: recognition of details in the spectra and coverage of the various levels of granularity from **patterns** to **lines** (**Figure 2**) requires these entities to actually be captured by the measurement. The fact of these entities being embodied in the FID also precedes the use of specific nomenclature, as shown in the previous sections. For example, it is important to realize that any extra time spent on sample preparation, acquisition, and data processing pays back multiple times during data interpretation, documentation, and further downstream use. Regardless of how raw time-domain NMR data (FID) are processed to yield frequency domain spectra (e.g., FT, CRAFT; see above), every detail in acquiring the FID including but not limited to S/N, dwell time, and the dynamic range is constrained by the information contained in the FID.

As the chain of quality in NMR spectroscopy starts with the acquisition of the FID, the integrity of the structural and/or quantitative interpretation of NMR spectra is inherently rooted in FID quality. Moreover, as is evident from the examples and considerations in the sections above and below, interpretation integrity equally depends on the use of specific nomenclature, which involves the appropriate distinction especially of **patterns**, **peaks**, and **lines** as well as QM **transitions**. As such, nomenclature is an integral part of scientific integrity in NMR analysis.

Going beyond the relatively simple AB_2C_2 spin system discussed above (**Figure 3**), this study also gathers experimental evidence to explain why and how precise and coherent terminology is crucial to the interpretation of NMR spectra. For this purpose, the following section re-visits α-santonin, strychnine, and testosterone as cases of classical NMR standard molecules and demonstrates the use of the terminology in the specific context of select ^{1}H resonances of these molecules. In order to provide clear rationales for the terminology, in-depth evaluations of their spin systems are presented and include the establishment of fully QM-based interpretations of the HNMR spectra of these molecules. To expand the picture for broader adoption, the following also reviews a portfolio of already reported cases that cover the entire range of small molecules.

5 | FURTHER RATIONALES AND EXEMPLARY APPLICATIONS

5.1 | X-SMALL MOLECULES: OVERLAPPING RESONANCES OF TWO METHINE HYDROGENS

The introductory example shown in **Figure 3** employs a synthetic spectrum to explain and visualize the NMR-specific terms as described above. The figure uses the overlapping NMR **patterns** of two methine hydrogens (A^* vs. $A^\#$) from the two different minimalistic AB_2C_2 and ABC_3 spin systems. These patterns can only be understood by knowing the actual underlying spin systems, and by taking into account the relations between the five **lines** that give rise to the four readily visible **peaks**. Spin A^* of the more abundant AB_2C_2 component gives rise to a triplet **pattern** with a slight 'roof effect', where line 1 intensity is slightly less than 1, and line 3 intensity is slightly greater than 1. Spin $A^\#$ of the less abundant ABC_3 system appears as a doublet that also slightly leans to the right.

Such minor deviations from the first-order **multiplicity** rules are often referred to as 'higher-order' effects and are very common, even in **spectra** acquired at ultra-high magnetic fields. Higher-order effects occur when the Larmor frequencies of two coupled nuclei are not significantly greater than their *J*-coupling value. These deviations also mean that the intensity ratios in the **signals** of the coupled nuclei do not exactly match the theoretical values of 1:2:1 in the triplet and 1:1 in in the doublet. In fact, deviations are often substantial. This already exemplifies one main challenge of integration-based qNMR (int-qNMR): unless pure first-order **multiplicities** are achieved, the dissection of the intensities of overlapping **resonances** is hampered or even impossible.

Figure 3 also shows that the shapes of the **lines** include information about possible long-range *J*-couplings between A and C. Such **line** *shape* information can be highly diagnostic by indicating that long-range couplings exist, compared to situations where **lines** are narrow, even though the values for these coupling constants correlate with the respective **line** *width*. However, by taking these QM relations into account, overlapping **signals** such as those shown here can still be completely resolved and all related chemical shifts and relevant coupling constants can be determined. Moreover, QM-based analysis can provide the accurate molar ratio of two spin

systems, which is hardly accessible via integration, and still difficult and/or less accurate when applying peak-fitting methods (PF-qNMR) because there are more lines than peaks to be fitted.

In support of the above definitions of terms, **Figure 3** includes **signals**, **peaks**, **lines**, and QM **transitions** as color-coded elements. Considering how widespread the structural motif of two overlapping AB_2C_2 and ABC_2 systems is in organic molecules, this simplistic case shows that precise usage of the four terms is key to the proper interpretation and use of (q)NMR spectra.

5.2 | SMALL MOLECULES: UNDERSTANDING APPARENTLY SIMPLE RESONANCES

Upon closer inspection, even apparently simple ^{1}H **resonances** often reveal relatively complex **patterns** that cannot be derived from visual interpretation or measurement of **peak** frequency differences.

One such example is the **resonance pattern** of H-9a in the sesquiterpenoid lactone, α-santonin (**Figure 4**). This deceptively simple resonance exemplifies several of the key terms as defined above and also helps explain why precise terminology is of utmost importance.

- The H-9a **resonance** gives rise to a **signal** with an apparent **pattern** of a triplet of doublets (td), which is in fact a doublet of doublets of doublets (ddd), involving two nearly-identical large couplings, which is another reminder that only the consideration of the underlying spin system allows the correct determination of the **multiplicity**.
- While visual inspection identifies six **peaks**, consistent with the preliminary assignment of a td **pattern**, QM analysis via HiFSA shows that the true *J* coupling network involves four coupling constants representing a dddq (-13.66 Hz with H-9b, 13.16 Hz with H-8a, 4.63 Hz with H-8b, 0.67 Hz with H-14) with up to 32 expected **lines**.
- The QM analysis reveals that the **six peaks** or up to **32 expected lines** (2^5 due to five *J*-couplings) are in fact composed of **155 total non-degenerate transitions**. Depending on the line width and shape of the measurement, the transitions give rise to at least six (as in **Figure 4**) or up to even more than the expected 32 **lines**.
- While this level of detail appears to be hidden in the 0.93 Hz experimental **line width** of the H-9a resonance, HiFSA can resolve the complex **line** overlap and composition of the individuals **peaks** of the H-9a **resonance patterns** via the mathematical interdependence of the underlying spin parameters on the entire spin system of α-santonin.

- The subtle but significant difference in the exact location of the **transitions** and the resulting experimental **lines** and **peaks** vs. the first-order expression of the actual chemical shifts and coupling constants in the coupling tree, indicates the presence of a non-first order effect in the spin system. This is caused by the relative proximity of the **resonances** of the tightly *J*-coupled H-8 methylene hydrogens. The presence of higher-order coupling effects is also indicated by the difference between the *center of the experimental* **splitting pattern** (the minimum of the central "doublet") vs. the *actual* **resonance** *frequency* of H-9a, which is located at the center of the overall experimental intensity of this signal.

The extent to which the first-order expression of a splitting tree, e.g. as shown in **Figure 4**, and the actual **transitions** differ, depends on the spin system of the molecule and the magnetic field. The difference between the two can be substantial (tens of Hz), while it is still larger than zero for most ^{1}H resonances - a fact that cannot be overemphasized. Importantly, this difference is **not** an uncertainty of the NMR measurement, but a natural property that is predicated on the QM foundation of NMR spectroscopy. As such, even if the differences are in the order of a few Hz and might be considered as small, they are highly significant and can be determined with mHz accuracy.

Transitions are combined into lines, which have non-discrete, continuous character. Combined into **lines** with certain shapes and widths, this results in **peaks** that form part of the **patterns** and ultimately the NMR **spectrum**. From a generalized perspective, the number of **transitions** is always greater or equal to the number of **lines**, and the number of **lines** is always greater or equal to the number of **peaks**.

Collectively, the example of the axial H-9a in α-santonin explains two important aspects: (a) that a near-first-order resonance is still not fully first-order, but rather exhibits noticeable and precisely measurable deviations in the frequency locations of the **peaks**, **lines**, and entire **resonance**; (b) that the terms like **peak**, **pattern**, **line**, and **resonance** have to be clearly defined and differentiated in order to describe NMR **spectra** properly.

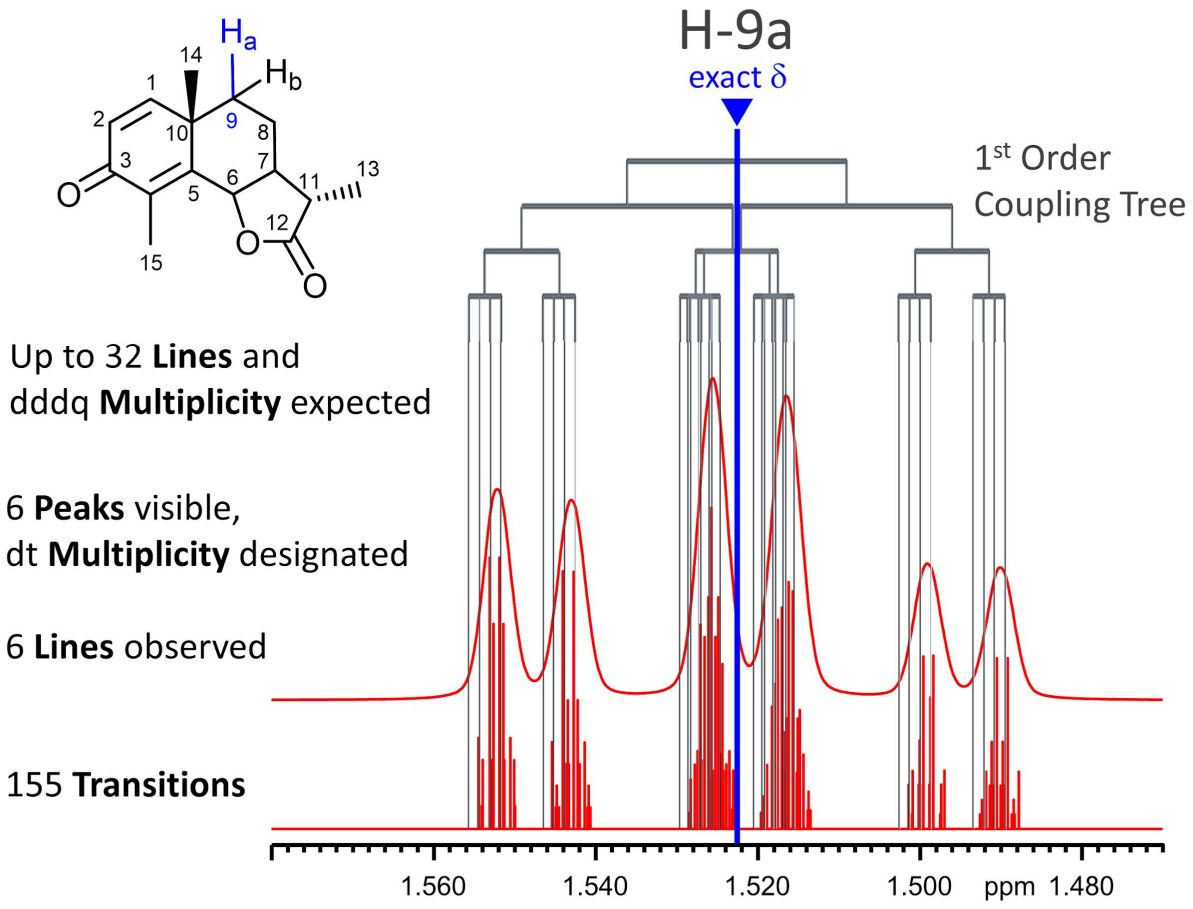


FIGURE 4 The **resonance** of the axial H-9a in α-santonin (500 MHz, $CDCl_3$) exemplifies how clear definitions of the terms **pattern**, **peaks**, **lines**, and **transitions** are key for the interpretation of the observed **signal**, and for all scientific communications of such matter. Refer to the main text for further explanations. In this example, the numbers of **peaks** vs. **lines** are identical (six), due to the very large number and very close proximity of the underlying 155 **transitions**. This myriad of transitions can be considered as producing 155 **lines** that are indistinguishable with current instrumentation. Note the differences in the exact positions of the individual **peaks** of the apparent dt/ddd **pattern** vs. those of the **lines** in the first-order coupling trees that represent the *J*-couplings with the geminal H-9_b,(13.66 Hz), the vicinal H-8a (4.63 Hz) and H-8b (13.16 Hz), and long-range with the angular Me-14 (0.76 Hz). This apparent mismatch in fact shows the subtle but important difference between the first-order assumption of a visual interpretation and the true spin parameters determined by QM-based full spin analysis. This case also exemplifies why non-QM-based fitting methods (peak deconvolution) and automated methods of multiplicity analysis will intrinsically yield inaccurate results, with the level of inaccuracy and deviation from truth depending on the particular spin system.

Comparison of an automated, first-order pattern analysis vs. a QM-based HiFSA shows that first-order assumptions in this case yield reasonably close values for the coupling constants. Automated multiplicity or multiplet analysis such as provided by contemporary NMR software yields an obviously wrong output such as the following: δ 1.52 (td, *J* = 13.2, and 4.5 Hz; 1H), whereas HiFSA analysis (CT, NMR Solutions Ltd, Kuopio, Finland) reveals the actual spin parameters as δ 1.52264 (dddq, *J* = -13.660, 13.158, 4.631, and -0.668 Hz; 1H). In fact, without QM analysis the coupling to the methyl group would remain undetected. However, this (over)simplified approach obscures important facts behind the data, as discussed above, which blurs conclusions and hides actual information, especially when imprecise nomenclature is used. Importantly, at the present stage of evolution of QM-based NMR tools, the computational effort for HiFSA is often lower than that of automated **pattern** analysis. Notably, the latter tends to fail as seen above and usually does not work at all for more complex **patterns** that only very experienced NMR practitioners are able to discern correctly. Usually, such **patterns** are ubiquitously labelled as multiplets (see also next section).

Considering that H-9a and its coupling partners (H-9b, H-8a, H-8b, and H-14) can indeed be classified as forming a near(!) first-order spin system and keeping in mind that the majority of organic molecules contain structural motifs that form non-first order/higher order spin systems, the observed mismatch between visual and proper QM-based spectral interpretation occurs in a wide range of molecules and deserves much wider attention in HNMR spectroscopic data interpretation.

5.3 | MEDIUM MOLECULES: PEAK PATTERN MISMATCH AND UNREAL COUPLING

Strychnine has been used widely as a test molecule for the establishment and validation of NMR experiments. In the present context, the HNMR spectrum of strychnine (**Figure 5**) demonstrates two frequently encountered properties that require careful terminological consideration for proper interpretation: visual **multiplicity** pattern mismatch and **peak** overlap. The **resonance** of H-18b gives rise to a **peak pattern** that apparently does not match the expected (first-order) **multiplicity** of the two strongly and one weakly coupled nucleus. The customary determination of the **peak** distances in this apparent yet illogical ddd-type **pattern** quickly reveals that they cannot be aligned with the geometries such as the relative dihedral angles as calculated via the Karplus relationships in this relatively constrained alicyclic molecule.

The QM-based solution encoded in the first-order coupling trees shown in **Figure 5** reveals that the frequency differences between individual constituting **peaks** within the resonance **pattern** clearly do not match the actual underlying coupling constants (see Section 3.1 regarding the paradigm “Peak Frequency Differences *Are Not* Coupling Constants”). This is due to the presence of a higher-order spin system. Note that, while the small long-range coupling of 0.53 Hz is not visible as **peak** separation (aka **peak splitting**), the noticeable 2-3 Hz **peak** separation does not have a corresponding (“matching”) coupling constant. This behavior has been designated in the early NMR literature as “virtual coupling” [19], reflecting the fact that peak splitting sometimes cannot be explained by first-order spin-spin coupling rules, and tends to go unnoticed even in prominent compound classes, such as in quinic acid derivatives [20]. The extent of **peak splitting** in such cases of depends on the higher-order level of the spin system: additional peak separation typically occurs in the 1-4 Hz range, thereby presenting a significant confusion potential for visual interpretation: the observed splittings are “real”, but are <u>not</u> due to actual *J*-coupling and are the result of the exact (“incidental”) frequency locations of the **lines** and underlying QM **transitions** of the observed **resonance**.

The overlapping **resonances** of H-14 and H-11b in strychnine show the importance of proper terminology: looking at the 2nd (from the left) quadrant of the apparent and overlapping dd pattern, one could wonder if this represents one **signal/resonance/peak/line** vs. two (or more) overlapping **signals/resonances/peaks/lines**. Following routine **peak-**picking methods, this quadrant would typically yield one picked frequency. However, the overall intensity and broad underlying component might raise concerns about the validity of this simplistic view.

The QM-based analysis encoded in the first-order coupling trees in **Figure 5** is consistent with H-14 having six coupling partners, forming an array of relatively closely matched coupling constants. This leads to a dddddd-type **pattern** consisting of 2^5=32 individual **lines**. The **line** frequencies are so close that, when combined with experimental or natural line width, a highly complex set of **peaks** results, which overlap to what would typically be called a broad singlet that coincides with the one **peak** of the H-14 dd **pattern** located in the 2nd quadrant. As QM-based full spin analysis considers the entire *J*-coupling network of the molecule, an accurate description of all **patterns** including that of H-14 is achieved (see HiFSA profile, Supplementary Material).

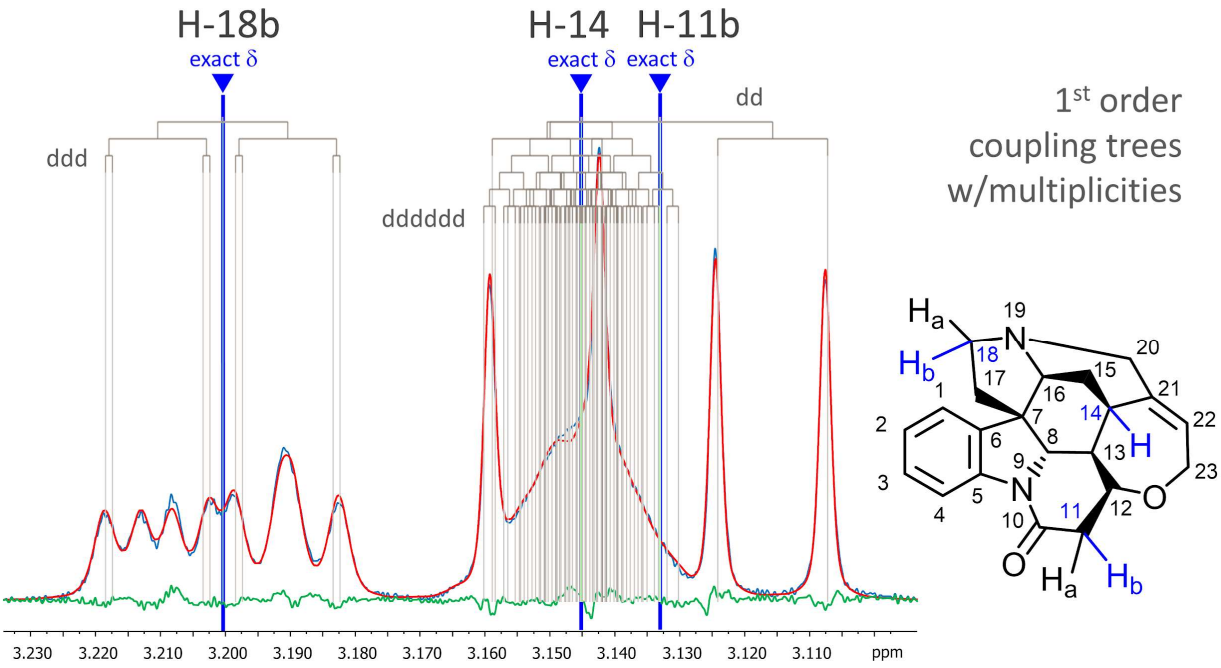


FIGURE 5 The **resonances** of H-18b (left) as well as H-14 and H-11b (right) in strychnine (500 MHz, $CDCl_3$) exemplify two seemingly straightforward yet challenging and frequently occurring situations in HNMR spectral interpretation: mismatch of visual multiplicity and spectral overlap. The **resonance pattern** of H-18b clearly does not match the first-order coupling tree of the true coupling constants that underlie this multiplet, as determined by QM analysis (observed spectrum in blue, calculated spectrum in red, residual in green; see also S3, Supplementary Material). For example, H-14 shows six *J*-couplings to H-15a, H-15b, H-16, H-20a, H-20b, and H-22. Collectively, this exemplifies the paradigm that peak frequency differences *are not* coupling constants (see Section 3.1 for explanation of this paradigm). Moreover, while H-11b can be readily assigned to the dd-type **multiplicity** at lower frequency, the broad, rather undefined resonance of H-14 eludes further interpretation. QM-based (S3, Supplementary Material) analysis reveals the origin of its generated **peak pattern**, which normally would be characterized as a broad singlet or multiplet, whereas HiFSA shows that it contains no less than 64 underlying **lines** (and far more **transitions**).

5.4 | MEDIUM-SIZE MOLECULES WITH CHRONIC PEAK OVERLAP: INTERPRETING NMR SPECTRA OF STEROIDS

The expression of *envelope protons* was coined in the early days of NMR analysis. It vividly describes how the **spectra** of most steroidal compounds look: their many aliphatic, non-oxygenated, unsubstituted methine and methylene hydrogens give rise to a myriad of

resonances between ca ~0.8 and ~2.6 ppm ("the envelope") that exhibit complex **splitting patterns** due to the numerous possible *J*-coupling relationships in these molecules. While some resonances can stand out as un- or less-overlapped, "the envelope" is mostly assigned as a collection of multiplets, which sometimes get partially assigned via 2D ^{1}H,^{13}C correlation spectra.

These assignment challenges vary by molecule. Interestingly, testosterone represents a relatively modest case of **resonance** overlap (**Figure 6A**), especially when the **spectrum** is acquired at 600 MHz or above. However, a more detailed evaluation reveals that the differentiation and specific nomenclature of **resonances** vs. **peaks** vs. **lines** is essential for both proper interpretation and qNMR work.

The integrals in the overview **spectrum** in **Figure 6A** show the high consistency of the intensities of the **resonances** in terms of their relative proportions in the molecule: the **resonance** at ~1.88 ppm is arbitrarily calibrated to 2.000, the integral value for the 22 H's resonating between 0.9 and 2.6 ppm is of 22.008, equivalent to 1.00036 for one H (0.04% deviation). The **resonances** of H-6a=α (2.3072 ppm) and H-2a=α (2.2893 ppm) overlap in such a way that it is possible to sort out the overlapping contributions from each hydrogen visually, using the non-overlapped halves at the high and low frequency ends of the **peak pattern** as guidance (**Figure 6B**). However, when breaking down the 1.997H integral into three segments, the middle part reflects the two half contribution from each hydrogen (1.002) - but this is deceptive, as it is not composed of equal 0.500 halves. The same deception is evident from the integrals of the higher and lower frequency segments of 0.572 and 0.423, respectively. Their significant deviation from the 0.500 half-value proves that both **resonances** are of higher order, which gives rise to noticeable *roof effects* (i.e., **peak patterns** leaning towards higher frequency for both H-6a=α and H-2a=α). In fact, the quantitative disbalance is significant ~15% and (0.423 [H-6a=α] and 0.572 [H-2a=α] observed vs. 0.500 theoretical first-order value), which raises major concerns for int-qNMR.

Performing HiFSA on the same dataset, using the Cosmic Truth (CT) software tool (NMR Solutions; ct.nmrsolutions.io) in automation, fully clarifies the assignments (**Figure 6C**; see also S5, Supplementary Material): **(a)** the first-order *J*-coupling forks confirm that the dd (H-6a=α) and ddd (H-2a=α) **splitting patterns** of the "outer halves" are to be duplicated, due to the geminal couplings; and that **(b)** the actual **resonance** frequencies of the two nuclei are slightly different from the visual centers of the **peak patterns**; **(c)** HiFSA detects the presence of a long-range ^{4}J = *W*-coupling of 0.97 Hz between H-2a=α and H-4. Collectively, HiFSA confirms that **(d)** both **resonances** can be considered as having a general dddd **multiplicity**, keeping in mind the

higher-order effects. Importantly, the full spin analysis shows that 32 **lines** underlie what appears to be 19 **peaks** with 2 additional shoulders. At 600 MHz, the 32 lines are composed of 32 quantum mechanical transitions, indicating that these two experimental **patterns** could be explained using first-order approximation. However, at 60 MHz, 464 transitions would be needed to explain the **resonance** arising from the same two spins. It should be kept in mind that numerous **transitions** within such large sets of **transitions** are often degenerate, which means they have the same or nearly the same frequency. To accelerate QM calculations, such degenerate or near-degenerate **transitions** can be compressed without compromising accuracy. For the purpose of concise nomenclature, however, it is still important to recognize that each visually identified **peak** typically comprises several **lines**, which are themselves typically composed of multiple **transitions**, even when using compression algorithms for (near) degenerate **transitions**.

This has a number of broader implications that reach well beyond the chosen steroidal example. In fact, the implications affect the vast majority of HNMR and qNMR analyses performed today:

- Whenever **resonances** of two or more nuclei overlap, analysis to the level of **lines** is required in order to understand the **peaks** and **peak patterns:** as shown in the testosterone case, **peaks** frequently represent degenerate **lines** that cannot be resolved by the measurement - even at relatively high magnetic field strengths.
- The validity of an int-qNMR measurement result depends closely on how accurately the integration range represents the **lines** that belong to the target molecule.
- Unless it is based on assigned **lines**, **peak** fitting-based analyses of NMR spectra are systemically flawed. This also means that Peak Fitting qNMR without QM-base validation is empirical and associated with an unknown, potentially significant systematic error.
- Higher-order effects, that are more common than generally perceived, even at high magnetic field strengths, are associated with significant disbalances in the relative intensities within the **resonances** of individual nuclei. Sorting out these situations requires full knowledge of the **lines** and, thus, **peak patterns**.
- Higher-order effects in NMR **spectra**, even if they are small, are the inevitable cause of substantial errors in int-qNMR - unless the **line** composition of the integrated **peaks** has been fully established. The resulting errors are well above the typical 1-2% accuracy/precision targets of most qNMR analysis (14% in the given testosterone example).

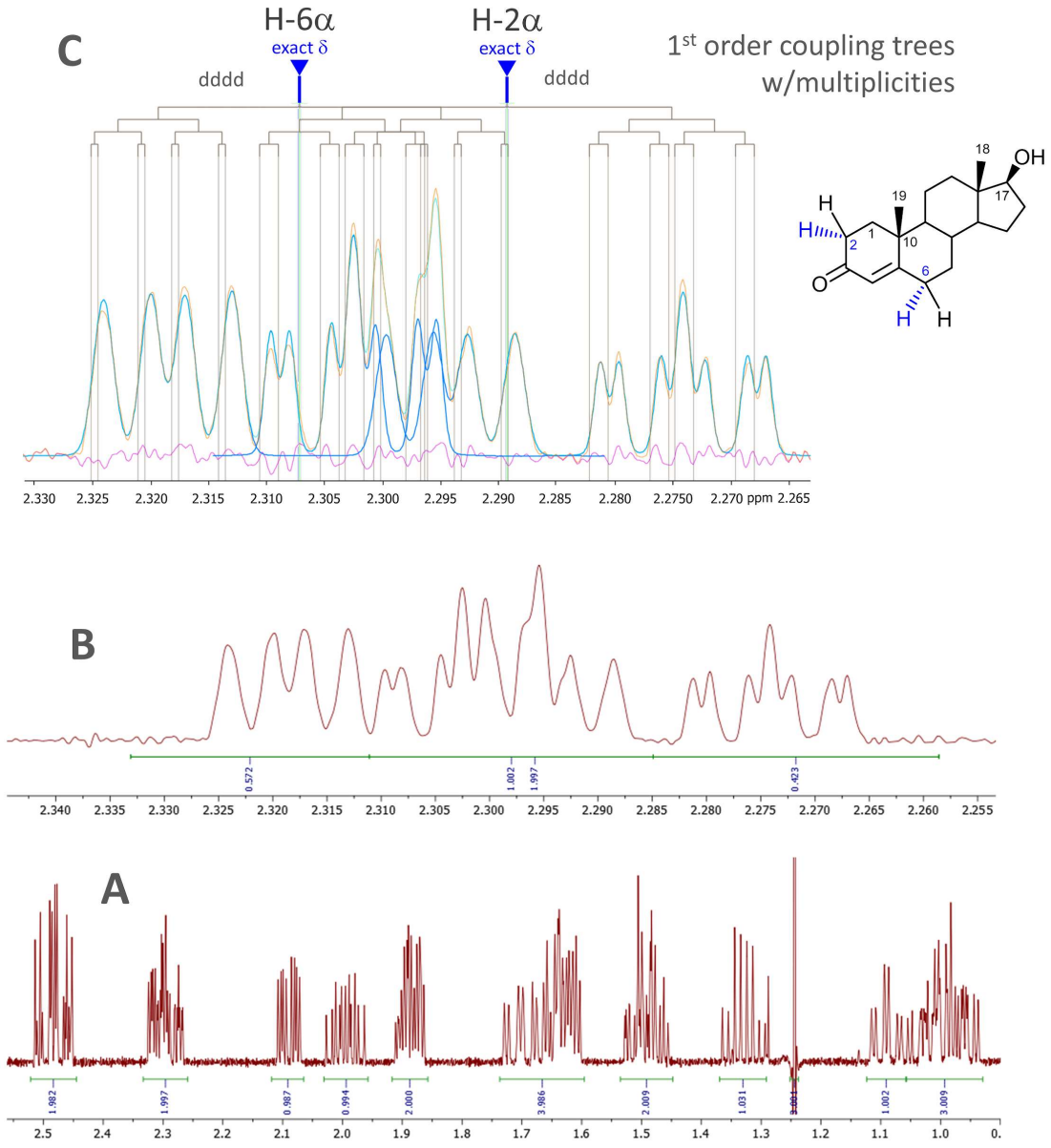


FIGURE 6 The 600 MHz HNMR spectrum of testosterone provides an illustrative example, why NMR nomenclature is essential for spectral interpretation and qNMR applications. The resonances of H-6a=α and H-2a=α around 2.3 ppm give rise to a **pattern** of 19 **peaks**, with additional two shoulders, that result from 32 **lines**. This **pattern** corresponds to two dddd-type multiplets that reflect the J-couplings of H-6a with H-6b, H-7a, H-7b, and H-4, and H-2a with H-2b, H-1a, H-1b, and H-4. Ignoring such specific information inevitably leads to misleading and/or erroneous assignments and quantitative evaluations when doing qNMR. The raw NMR data (FID) for this case study was kindly shared by the authors of [21] at DOI:10.7910/DVN/MOCHRD.

5.5 | LARGE MOLECULES: UNDERSTANDING HNMR SPECTRA OF 1,000+ AMU COMPOUNDS

At the current stage of methodology, the extraction of accurate chemical shifts and all relevant spin-spin coupling constants is possible for quite large molecules. Possibly due to the limitations of historic tools such as LAOCN3/LAOCOON (Least-squares Adjustment of Calculated on Observed NMR Spectra, see [22,23] and references therein), there is a general belief that this type of analysis is limited to 8-12 spins and not useful as a general application. Today, as we have much more powerful computers, several cases demonstrating the feasibility of analyzing molecules with a molecular weight in the range of 1,000 amu and above have been reported, as follows.

Flavonoid glycosides represent a large class of compounds that are ubiquitous and often relatively abundant in vascular plants. For biosynthetic reasons, flavonoids typically occur as conjugates with sugar moieties, which often amount to three or more sugars, thereby reaching the molar mass range ~800 g/mol to 1,100 g/mol. The close stereochemical relationships between the sugar moieties and the variable inter-glycosidic linkages explain the challenges associated with the structure elucidation of such compounds. For example, specific triglycosides of quercetin and kaempferol are important plant-organ specific quality markers of the medicinal plant, Ashwagandha (*Withania somnifera*). As the compounds contain robinose (α-L-Rha1$_1$→^{6}β-D-Gal) or rutinose (α-L-Rha1$_1$→^{6}β-D-Glc) and β-D-glucose, they give rise to many near-degenerate resonances, which can be fully understood by HiFSA [24]. While otherwise concealed by multiplet designation, the NMR **spectra** of these flavonoid glycosides contain highly structure-specific **peak patterns** with a myriad of underlying **lines**, all of which can be fully assigned to the individual H atoms. Importantly, the HiFSA-based understanding of the **spectra** allows the distinction and identification of close congeners, and even enables their analysis by benchtop NMR [24].

Also representing oligomers of congeneric building blocks, peptides are another class of higher molecular weight compounds that challenge HNMR spectroscopic data interpretation. For example, a recent systematic study of the spin systems of linear oligopeptides up to the size of the octapeptide, angiotensin II (1,046 amu) and the nonapeptide, oxytocin (1,067 amu), demonstrated the feasibility of HNMR spectroscopy to perform an NMR-based structural sequencing (HiFSA sequencing) of the peptides [25]. Approaching oxytocin via stepwise synthesis, the study showed how each individual amino acid and substituent chemical shift (s.c.s.) effects

contribute to the final spectrum. The outcomes also revealed that the characteristics of the resonances of the individual amino acid moieties show certain variation in terms of observed **patterns**, but almost no variation of the underlying spin couplings. This highlights the relevance of full spin system analysis in the understanding of HNMR spectra as well as the importance of distinguishing between **patterns**, **peaks**, and **lines** when performing visual and numerical interpretation.

Further examples of therapeutically relevant "larger" peptides are cyclopeptides, which have gained prominence as potential leads evolving from natural products-based drug discovery programs. Ecumicin is a tridecapeptide from a *Nonomuraea spec.* actinomycete with a promising bactericidal activity and new mechanism of action targeting *Mycobacterium tuberculosis* proteostasis via the ClpC1 chaperone protein. The HNMR spectrum of this 1,599 amu molecule was fully assigned to the level of **lines** (and underlying **transitions**) in 2016 [26]. This knowledge was used to deduce the structures of five new congeners (nor- and deoxy- analogues) via HNMR spectroscopy only, as samples were very limited at the time. These structures were subsequently verified by full 2D NMR using micro-cryoprobe NMR analysis at 750 MHz [27].

These examples have in common that QM-based spectroscopic data interpretation not only enables the full understanding of the NMR **spectra**, but also demonstrates the distinct nature and modularity of the spectral features (**Figures 2** and **3**). Collectively, these examples also highlight the feasibility of achieving this level of understanding with higher-amu compounds, thereby covering the important range up to ~2,000 amu of drug molecules and toxins.

5.6 | MYRIADS OF LINES VS. HANDFUL OF QM PARAMETERS

While QM-based analysis is required for the interpretation of non-first-order HNMR spectra, the approach can also resolve **signal** overlap that occurs nearly ubiquitously even in (near) first-order ultra-high field NMR **spectra**. Non-QM based **multiplet/multiplicity** analysis typically fails in such instances because computations must consider not just the **peaks**, but the entirety of the many more **lines**. Evidently, completely overlapping lines remain unknown unless QM calculations are performed. In contrast, the QM-based approach deals with a much lower number of parameters: rather than all lines (32 in the relatively simple dt/ddd pattern of the axial H-9a in α-santonin), QM only requires significantly fewer values, i.e., the chemical shifts and all coupling constants (1 and 4, resp., in the same example). In addition, **line** fitting requires matching of **line** intensity, width, and shape, which quadruples the number of parameters that need to be matched (128). In

contrast, intensity, width, and shape are only one parameter each in HiFSA, because they are uniform for a given spin particle. Therefore, QM requires only eight parameters to be determined, compared to 128 for the line-fitting methods.

Thus, QM-based HNMR analysis is more efficient and stable than non-QM based methods. Since this advantage is independent of the occurrence of higher-order effects, which are most noticeable in low-magnetic field spectra, but are still common in high-field spectra, QM analysis is not only an approach for particularly complex cases, but the one that ought to be widely applied generally for NMR spectral interpretation and is becoming increasingly automated.

5 | SUMMARY & OUTLOOK

The terms **pattern**, **peak**, **line**, and **transition** are located at the heart of the NMR vocabulary (**Figure 3**). Their appropriate and consistent use is required to reflect the modularity of the physical principles (**Figure 2**) encoded into NMR **spectra** and to maintain proper rigor for the use of (q)NMR spectroscopy as a primary analytical tool. The definitions provided in Section 2 clarify the distinct meaning of the terms and are presented as a consensus proposal for broader discussion and adoption.

The fact that the present study was able to utilize previously acquired data to develop a set of experimental rationales that offer a new perspective on NMR nomenclature has two important implications. First, it highlights the immense value of raw NMR data sharing: the additional value that can be gained from the re-analysis of existing primary data has been identified as a major incentive for not just the NMR, but the entire scientific community to engage in FAIR data sharing practices [13,14]. Second, it underscores the need for broader standardization of NMR methodologies and dissemination mechanisms, as reflected by the recent NMReDATA initiative [29,30], which aims at the extracted data such as chemical shift values, integrals, intensities, multiplicities, and scalar coupling constants. As shown throughout the article, the development of exact definitions of NMR terms is a prerequisite for the meaningful extraction of structural and quantitative information from NMR spectra and their subsequent use by the scientific community.

To this end, the present study contributes a major piece to the greater puzzle of qualitative and quantitative NMR spectroscopy and its increasing prominence in chemistry and the applied health sciences. Accordingly, the authors envision that the concerted, broad implementation of consolidated NMR terminology, raw NMR data (FID) sharing, and standardized digital reporting

of interpreted data into NMR practice and literature is key to the advancement of NMR methodology and Science in general.

ORCID

to be populated by MRC

CONFLICT OF INTEREST DECLARATION

MN and SPK are founders of NMR Solutions Ltd. DJ is founder of NMRprocess.ch and collaborates with Mestrelab (mestrelab.com). KK is founder of Chempacker LLC. DS is an inspector at Health Canada, but is not representing Health Canada in this publication. The remaining authors declare no conflict of interest.

DISCLAIMER

Certain equipment and software have been identified in order to adequately specify the work. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology or any of the authors or their institutions, nor is it intended to imply that the mentioned equipment or software is necessarily the best available for the purpose.

ACKNOWLEDGEMENT

The authors are grateful to Dr. Gary Martin for sharing raw NMR data of α-santonin and strychnine, and to Dr. Der-Lii Tzou for engaging in raw NMR data sharing and providing the spectrum of testosterone. GFP and SNC are grateful to Dr. Ashok Krishnaswami of Jeol and Jeol Resonance, USA, for their NMR advice and support.

SUPPLEMENTARY MATERIAL

The QM-based ^{1}H iterative Full Spin Analysis (HiFSA) profiles of the presented cases are included as Supplementary Material. Furthermore, the raw NMR data and related electronic files are made available via the Harvard Dataverse repository at DOI: 10.7910/DVN/FGJ6ZU.

REFERENCES

[1] C. Szantay, *Anthropic Awareness: The Human Aspects of Scientific Thinking in NMR Spectroscopy and Mass Spectrometry*, Elsevier Science Publishing, Philadelphia, PA, **2015**.

[2] G. F. Pauli, S.-N. Chen, D. C. Lankin, J. Bisson, R. J. Case, L. R. Chadwick, T. Gödecke, T. Inui, A. Krunic, B. U. Jaki, et al., *J. Nat. Prod.* **2014**, *77*, 1473.

[3] R. Reher, H. W. Kim, C. Zhang, H. H. Mao, M. Wang, L.-F. Nothias, A. M. Caraballo-Rodriguez, E. Glukhov, B. Teke, T. Leao, et al., *J. Am. Chem. Soc.* **2020**, *142*, 4114.

[4] R. S. Phansalkar, C. Simmler, J. Bisson, S.-N. Chen, D. C. Lankin, J. B. McAlpine, M. Niemitz, G. F. Pauli, *J. Nat. Prod.* **2017**, *80*, 634.

[5] O. Beckonert, H. C. Keun, T. M. D. Ebbels, J. Bundy, E. Holmes, J. C. Lindon, J. K. Nicholson, *Nature Protocols* **2007**, *2*, 2692.

[6] **2015**, DOI 10.1039/9781849735186.

[7] P. S. Achanta, B. U. Jaki, J. B. McAlpine, J. B. Friesen, M. Niemitz, S.-N. Chen, G. F. Pauli, *J. Pharm. Biomed. Anal.* **2021**, *192*, 113601.

[8] P. McQuilton, A. Gonzalez-Beltran, P. Rocca-Serra, M. Thurston, A. Lister, E. Maguire, S.-A. Sansone, *Database* **2016**, *2016*, DOI 10.1093/database/baw075.

[9] C. Steinbeck, O. Koepler, F. Bach, S. Herres-Pawlis, N. Jung, J. Liermann, S. Neumann, M. Razum, C. Baldauf, F. Biedermann, et al., *Res. Ideas Outcomes* **2020**, *6*, DOI 10.3897/rio.6.e55852.

[10] M. Barker, R. Wilkinson, A. Treloar, *Data Sci. J.* **2019**, *18*, DOI 10.5334/dsj-2019-044.

[11] K. Krishnamurthy, *Magn. Reson. Chem.* **2013**, *51*, 821.

[12] K. Krishnamurthy, *Magn. Reson. Chem.* **2021**, DOI 10.1002/mrc.5135.

[13] J. B. McAlpine, S.-N. Chen, A. Kutateladze, J. B. MacMillan, G. Appendino, A. Barison, M. A. Beniddir, M. W. Biavatti, S. Bluml, A. Boufridi, et al., *Nat. Prod. Rep.* **2019**, *36*, 35.

[14] J. Bisson, C. Simmler, S.-N. Chen, J. B. Friesen, D. C. Lankin, J. B. McAlpine, G. F. Pauli, *Nat. Prod. Rep.* **2016**, *33*, 1028.

[15] M. H. Levitt, *Spin Dynamics: Basics of Nuclear Magnetic Resonance*, Wiley, **2008**.

[16] J. Keeler, *Understanding NMR Spectroscopy*, John Wiley & Sons, **2011**.

[17] M. Newton, E. Breeds, R. Morris, *Electronics (Basel)* **2017**, *6*, 89.

[18] J. D. Roberts, H. W. Salzberg, *Journal of The Electrochemical Society* **1962**, *109*, 116C.

[19] J. I. Musher, E. J. Corey, *Tetrahedron* **1962**, *18*, 791.

[20] G. F. Pauli, U. Kuczkowiak, A. Nahrstedt, *Magnetic Resonance in Chemistry* **1999**, *37*, 827.

[21] D. Wu, K. J. Carillo, J.-J. Shie, S. S.-F. Yu, D.-L. M. Tzou, *Molecules* **2021**, *26*, DOI 10.3390/molecules26092643.

[22] A. A. Bothner-By, S. M. Castellano, *Computer Programs for Chemistry\textperiodcentered, de Tar, DF, Ed. ; WA Benjamin, Inc. : New York* **1968**, *1*.

[23] M. Clark, J. S. Thrasher, *J. Chem. Educ.* **1990**, *67*, 235.

[24] S. Bhavaraju, D. Taylor, M. Niemitz, D. C. Lankin, A. Bzhelyansky, G. I. Giancaspro, Y. Liu, G. F. Pauli, *J. Nat. Prod.* **2021**, *84*, 836.

[25] M. P. Choules, J. Bisson, W. Gao, D. C. Lankin, J. B. McAlpine, M. Niemitz, B. U. Jaki, S. G. Franzblau, G. F. Pauli, *J. Org. Chem.* **2019**, *84*, 3055.

[26] W. Gao, J. G. Napolitano, D. C. Lankin, J.-Y. Kim, Y.-Y. Jin, H. Lee, J.-W. Suh, S.-N. Chen, G. F. Pauli, *Magn. Reson. Chem.* **2017**, *55*, 239.

[27] W. Gao, J. B. McAlpine, M. P. Choules, J. G. Napolitano, D. C. Lankin, C. Simmler, N. A. Ho, H. Lee, J.-W. Suh, I. W. Burton, et al., *J. Nat. Prod.* **2017**, *80*, 2630.

[28] J. F. Araneda, T. Mendonça Barbosa, P. Hui, M. C. Leclerc, J. Ma, A. F. G. Maier, S. D. Riegel, *J. Chem. Educ.* **2021**, *98*, 1227.

[29] M. Pupier, J.-M. Nuzillard, J. Wist, N. E. Schlörer, S. Kuhn, M. Erdelyi, C. Steinbeck, A. J. Williams, C. Butts, T. D. W. Claridge, et al., *Magn. Reson. Chem.* **2018**, *56*, 703.

[30] S. Kuhn, L. H. E. Wieske, P. Trevorrow, D. Schober, N. E. Schlörer, J.-M. Nuzillard, P. Kessler, J. Junker, A. Herráez, C. Farès, et al., *Magn. Reson. Chem.* **2021**, DOI 10.1002/mrc.5146.

Essential Terminology Connects NMR and qNMR Spectroscopy to Its Theoretical Foundation

Guido F. Pauli[1a,1b,2,$,*], G. Joseph Ray[1,2,5,$], Anton Bzehlyansky[3], Birgit U. Jaki[2], Charlotte Corbett[4], Christina Szabo[5], Christoph Steinbeck[6], Dan Sørensen[7], Damien Jeannerat[8], Daneel Ferreira[9], David C. Lankin[1,2], James B. McAlpine[1,2], Jean-Nicolas Dumez[10], Jonathan Bisson[1ab,2], Krish Krishnamurthy[11], Matthias Niemitz[12], Michael A. Nelson[13], Patrick Giraudeau[10], Samuli-Petrus Korhonen[12], Stefan Kuhn[14], Toru Miura[15], Lucy Botros[3]

Supplementary Material

Table of Contents

S1. Questions and Answers Regarding the Axis-free 400 MHz HNMR Spectrum of β-Pinene Shown in Figure 1.

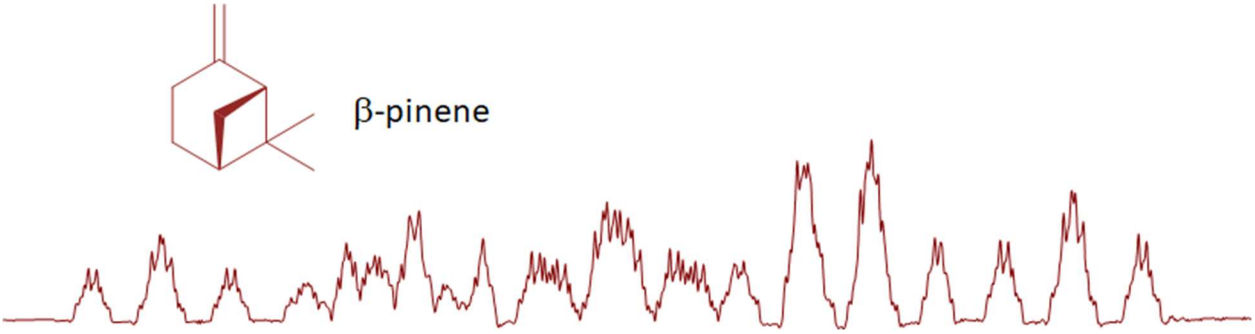


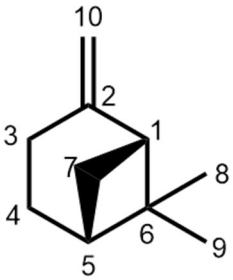
FIGURE 1 in the main text shows a small section of the 400 MHz HNMR spectrum of β-pinene.

The following provides some answers to the questions that were raised to exemplify the importance and breadth of the issue at hand related to specificity in NMR terminology.

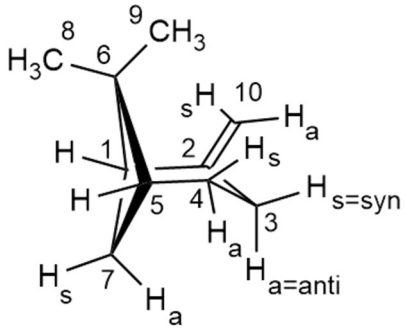
Q: Is this the entire spectrum, possibly taken at low magnetic field strength? Or is it a small segment of the spectrum? Or even a tiny portion arising from a single hydrogen, maybe acquired at ultra-high magnetic field strength?

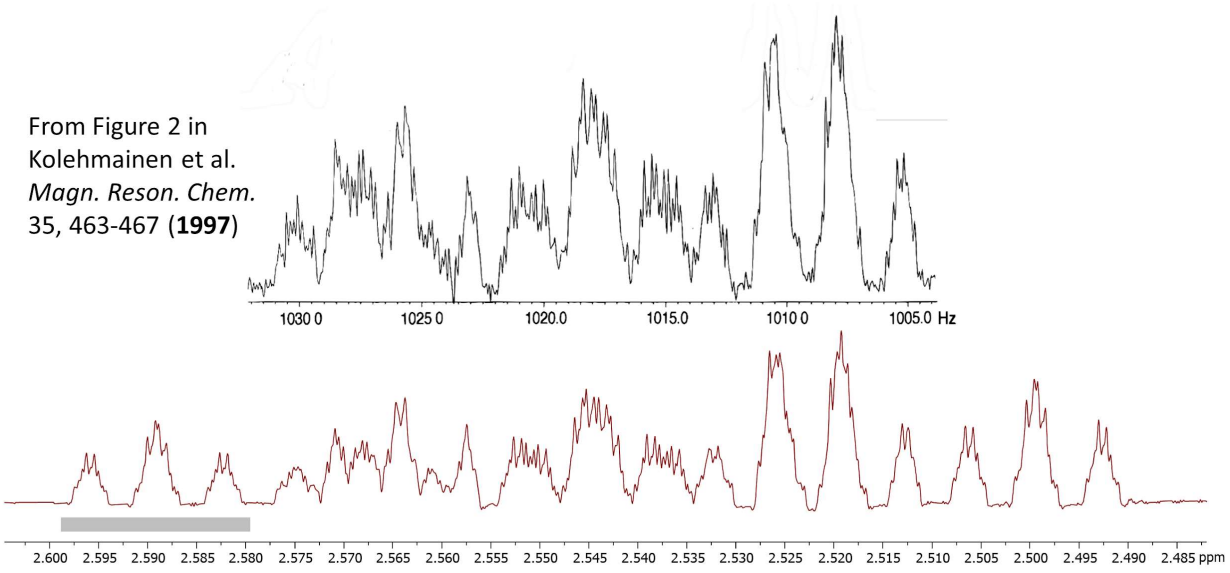
A: No, this is not the entire spectrum, but only the 2.482-2.604 ppm portion that shows the resonance of one hydrogen, H-3a (anti to the exocyclic methylene). It was acquired at 400 MHz (Jeol ECZ 400 spectrometer at UIC; operator: Shao-Nong Chen) and processed with a Lorentzian-Gaussian window function with -0.40 Hz exponential and 0.15 Hz Gaussian factors.

The following plot below shows the same section with the ppm axis.



β-pinene





This spectrum was acquired in December 2017. It is strikingly similar to and practically identical with the spectrum acquired 20 years earlier and shown in Figure 2 in the following publication: E Kolehmainen, K Laihia, R Laatikainen, J Vepsalainen, M Niemitz, R Suontamo, *Magn. Reson. Chem.* 35, 463-467 (**1997**), who first established the HiFSA profile of β-pinene and, in fact, the first published example of a full-spin analysis of an HNMR spectrum of this level of complexity.

Note that Figure 2 in the 1997 publication shows only part of the HNMR signal of H-3a; for comparison, the published spectrum is shown here as an insert. The gray bar in the 2017 spectrum indicates the zoom range from which the detail below was plotted to answer the question about *splitting*.

The high resolution and high level of detail in this spectrum is due to a combination of high magnetic homogeneity after careful shimming and the application of a Lorentzian-Gaussian window function. When plotted without an x-axis or frequency scale reference, the small segment of the entire spectrum shown in **Figure 1** could potentially give the false impression that it represents a bigger part of or full spectrum acquired at a much lower magnetic field strength, or a highly disperse segment of a spectrum acquired at much higher magnetic field strength.

Q: Does this plot represent a (single) resonance, (single) signal, or multiple peaks?
A: It represents the resonance of a single hydrogen, H-3a, within the β-pinene molecule.

Q: How many peaks and how many lines does it contain?

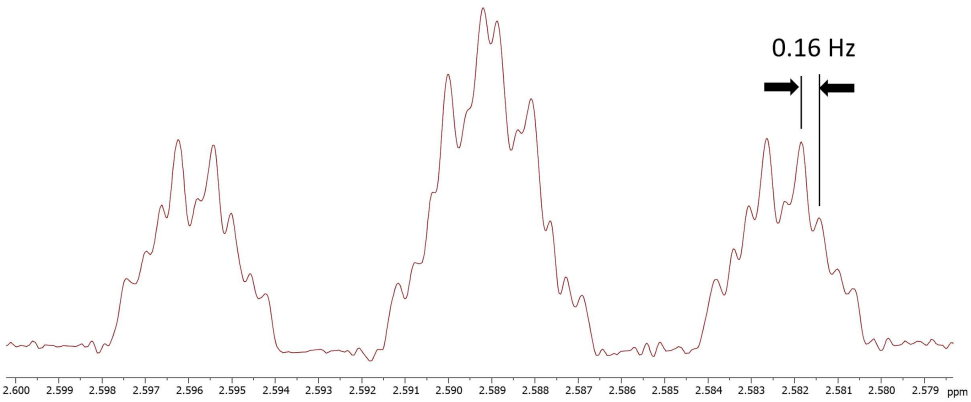
A: The answer to the first part of question ("how many **peaks**?") depends on the peak-picking algorithm. For the sake of this discussion, a manual peak picking exercise will identify >100 peaks and shoulders that are sufficiently split to count as peaks.

The answer to the second part of the question ("how many **lines**?") depends on the QM model applied to analyze the spin system, in particular how it handles the underlying small couplings and the natural dispersion of theoretically degenerate transitions.

Considering the very narrow **line** shape of this spectrum (0.10 to 0.20 Hz; see following answer), the **peaks** can also be considered as being the experimentally distinguishable lines for this processed version of the experimental spectrum.

Q: What is the meaning of the term *splitting* in this context?

A: First, It is important to realize that *splitting* can refer to either **peaks** or **lines** (see main text for explanation). As shown in the expanded plot below, the spectrum shows *splitting* at levels as low as 0.10 to 0.20 Hz. The 0.16 Hz frequency difference highlighted in the expansion below represents the **line** *width* of this spectrum; therefore, the entities for which this difference was determined are **lines** rather than **peaks**.



The high conformational rigidity of β-pinene leads to such a narrow natural **line** *width* below 0.3.-0.5 Hz, which otherwise represents the typical experimentally achievable range for most small molecules. The gray bar in the 2017 spectrum shown above indicates the zoom range from which this expansion was plotted.

Q: Is this an experimental or a calculated spectrum?

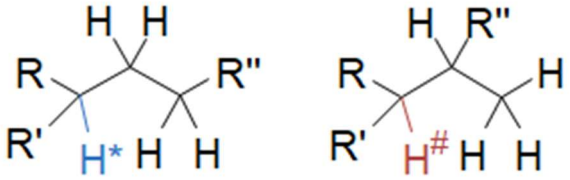
A: It is an experimental spectrum.

S2. Description of the Spin Matrix of the Example in Figure 3.

Spin System Structure

$A^*B_2C_2$ $CH^*CH_2CH_2$

$A^{\#}BC_3$ $CH^{\#}CHCH_3$



Calculated at 400 MHz with a line width of 1 Hz, a data-point resolution of 0.065 Hz, and a line shape with 1:1 Lorentzian/Gaussian contribution

Table of chemical shifts and coupling constants:

$CH^*CH_2^{**}CH_2^{***}$ Population = 75 %	Chemical Shifts (ppm)	$CH^{\#}CH^{\#\#}CH_3^{\#\#\#}$ Population = 25 %	Chemical Shifts (ppm)
H^*	3.000000	$H^{\#}$	2.990825
H_2^{**}	2.859600	$H^{\#\#}$	2.664951
H_2^{***}	2.535545	$H_2^{\#\#\#}$	2.506599
Nuclei	*J*-couplings (Hz)	Nuclei	*J*-couplings (Hz)
$H^* - H_2^{**}$	3.000	$H^{\#} - H^{\#\#}$	5.000
$H^* - H_2^{***}$	0.500	$H^{\#} - H_3^{\#\#\#}$	0.500

S3. The HiFSA Profile of α-Santonin.

Original NMR data: Biological Magnetic Resonance Data Bank (DOI: 10.13018/BMSE001233).

Note the relevance of five decimal places in the ppm values of the chemical shifts and three decimal places in the Hz values of the coupling constants (for further information, see Pauli et a. Journal of Natural Products 77: 1473-1487 (2014) DOI:10.1021/np5002384).

Chemical shifts			***J*-coupling constants**		
Atom	ppm		Atom	Atom	Hz
H1	6.70169		H1	H2	9.871
H2	6.25612		H6	H7	11.423
H6	4.80616		H6	H15	1.415
H7	1.82514		H7	H8a	3.633
H8a	2.03701		H7	H8b	12.298
H8b	1.70191		H7	H11	12.315
H9a	1.52264		H8a	H8b	-13.326
H9b	1.91038		H8a	H9a	4.631
H11	2.43113		H8a	H9b	2.267
H13	1.28356		H8b	H9a	13.158
H14	1.33481		H8b	H9b	4.007
H15	2.13708		H9a	H9b	-13.66
			H9a	H14	-0.668
			H9b	H14	-0.144
			H11	H13	6.929

S4. The HiFSA Profile of Strychnine.

500 MHz, 25 mg in 600 uL of $CDCl_3$.

Note the relevance of five decimal places in the ppm values of the chemical shifts and three decimal places in the Hz values of the coupling constants (for further information, see Pauli et a. Journal of Natural Products 77: 1473-1487 (2014) DOI:10.1021/np5002384).

Chemical shifts			J-coupling constants		
Atom	ppm		Atom	Atom	Hz
H1	7.16181		H11a	H11b	-17.382
H2	7.09806		H11a	H12	8.460
H3	7.25483		H11b	H12	3.319
H4	8.09586		H12	H13	3.347
H8	3.85871		H13	H14	3.111
H11a	3.13284		H8	H13	10.496
H11b	2.67255		H14	H15a	1.853
H12	4.28725		H14	H15b	4.774
H13	1.27444		H14	H16	0.879
H14	3.14499		H14	H20a	1.541
H15a	1.45672		H14	H20b	0.270
H15b	2.35906		H14	H22	-2.837
H16	3.94770		H15a	H15b	-14.367
H17a	1.89047		H16	H15a	2.160
H17b	1.89315		H16	H15b	3.901
H18a	3.20026		H17a	H18a	7.514
H18b	2.87306		H17a	H18b	12.692
H20a	3.70790		H17a	H17b	-12.519
H20b	2.73053		H17b	H18a	0.528
H22	5.90146		H17b	H18b	6.045
H23a	4.06760		H18a	H18b	-10.020
H23b	4.14671		H1	H2	7.555
			H20a	H20b	-14.791

H20a	H23a	1.653
H20a	H22	-1.518
H20b	H22	-0.672
H22	H23a	6.103
H22	H23b	6.977
H23a	H23b	-13.789
H1	H3	1.295
H2	H3	7.443
H1	H4	0.232
H2	H4	1.081
H3	H4	8.087

S5. The HiFSA Profile of Testosterone.

600 MHz spectrum from DOI: 10.7910/DVN/MOCHRD in CD_3OD.

Note the relevance of five decimal places in the ppm values of the chemical shifts and three decimal places in the Hz values of the coupling constants (for further information, see Pauli et a. Journal of Natural Products 77: 1473-1487 (2014) DOI:10.1021/np5002384).

Chemical shifts			J-coupling constants		
Atom	ppm		Atom	Atom	Hz
H1a	1.70267		H1a	H19	-0.636
H1b	2.0895		H1a	H1b	-13.480
H2a	2.28933		H1a	H2a	4.420
H2b	2.48183		H1a	H2b	14.827
H4	5.70901		H1b	H19	-0.130
H6a	2.30719		H1b	H2a	3.139
H6b	2.48307		H1b	H2b	5.182
H7a	1.01956		H2a	H2b	-17.045
H7b	1.89079		H2a	H4	0.970
H8	1.65098		H2b	H4	-0.036
H9	0.96013		H6a	H4	-0.356
H11a	1.62244		H6a	H6b	-14.614
H11b	1.4922		H6a	H7a	4.165
H12a	1.09085		H6a	H7b	2.424
H12b	1.88013		H6b	H4	-1.853
H14	0.99658		H6b	H7a	13.969
H15a	1.63044		H6b	H7b	5.423
H15b	1.3298		H7a	H7b	-12.843
H16a	1.99483		H7a	H8	11.712

H16b	1.4941		H7b	H8	3.573
H17	3.58128		H8	H14	10.909
H18	0.79576		H8	H9	10.693
H19	1.24437		H9	H11a	4.133
			H9	H11b	12.470
			H9	H19	-0.310
			H11a	H11b	-13.659
			H11a	H12a	4.203
			H11a	H12b	2.783
			H11b	H12a	13.242
			H11b	H12b	4.129
			H12a	H12b	-12.682
			H12a	H18	-0.623
			H12b	H18	-0.276
			H14	H15a	7.356
			H14	H15b	12.457
			H15a	H15b	-12.520
			H15a	H16a	9.636
			H15a	H16b	3.625
			H15b	H16a	5.968
			H15b	H16b	12.099
			H16a	H16b	-13.540
			H16a	H17	9.049
			H16b	H17	8.368