Supelco®
Analytical Products

White Paper

Digital Reference Materials

Sustainability by savings

Digital Reference Materials for NMR Spectroscopy

Creating new sustainable and cost-efficient reference materials for NMR

Reference materials have existed in the market for decades and provide traceability to metrological standards. The use of reference materials offers laboratories certainty about the identity of analytes and confidence in measurement quality. However, their limited shelf-life and environmental impact associated with production, testing, distribution, and disposal of chemicals makes alternatives desirable for various reasons. A library of reference data can, in many instances, replace a true measurement, as measuring the same analyte under the exact same conditions must give the same data. Data does not have an expiry date and is readily distributed and accessed. But creating an extensive library requires technical expertise for data interpretation and IT resources for database creation and maintenance. Herein, we describe how MilliporeSigma creates a new portfolio of digital reference materials (dRMs) for Nuclear Magnetic Resonance (NMR) spectroscopy available on our AI supported ChemisTwin™ portal. The portal can identify, verify, and quantify analytes, where the extensive database of dRMs is created from human assigned spectra, quality controlled, and reviewed to guarantee the correctness of NMR assignments.

Introduction

MilliporeSigma has been an industry leader for Certified Reference Materials (CRMs) for decades. The production and regulatory compliance of CRMs are governed by **ISO 17025** & **ISO 17034** standards for accredited test laboratories and the requirements for competence of reference material producers, respectively. The new portfolio of Digital References Materials (dRMs) is the logical extension of CRMs for the 21st century.

Author and contact Information

Dr. Johannes. F.P. Colell

Merck KGaA, 9470 Buchs, Switzerland
Industriestrasse 25

The dRMs are produced using our CRMs and Analytical Standards (AS), but the dRM portfolio, which is growing presently, is being extended to include synthesis and technical chemicals. Access to the dRMs is provided by our ChemisTwin™ portal [1], where customers may request a new compound as a dRM. In the portal, spectra can be browsed, compounds localized via metadata search (CAS number, Our product number or analyte name) and frequently used compounds can be set as favorites to facilitate comparing against our library. Customers may use the portal as they please, for example for quality control in a targeted search (known analyte) or to identify a compound using an untargeted search (search of entire library), where a spectrum is compared against every entry in the library. **Figure 1** shows the detailed view of a compound in ChemisTwin™ portal. Here, the product metadata provides traceability to the batch used in its production, molecular structure (SMILES) aids identification, and the CAS Number aids localization. The molar mass is a property of the molecule, but is also used for content quantification.

MilliporeSigma is the U.S. and Canada Life Science business of Merck KGaA, Darmstadt, Germany.

MilliporeSigma

The digital reference material is constituted by experimental ^{1}H NMR reference data with resonances assigned to the exact molecular structure. In contrast to software based auto-assign algorithms, Our spectral assignments are performed by a team of expert NMR spectroscopists and, if necessary, confirmed with 2D NMR. The full content of the dRM is reviewed by quality control and quality assurance personnel to guarantee our high standards.

In the following, we will describe in detail how a digital reference material is produced. We will explain which materials were chosen for digital reference production, and also elaborate on the process ensuring that our customers get only the best reference material quality. Furthermore, we will provide cautionary notes on when MilliporeSigma recommends continued use of physical reference materials.

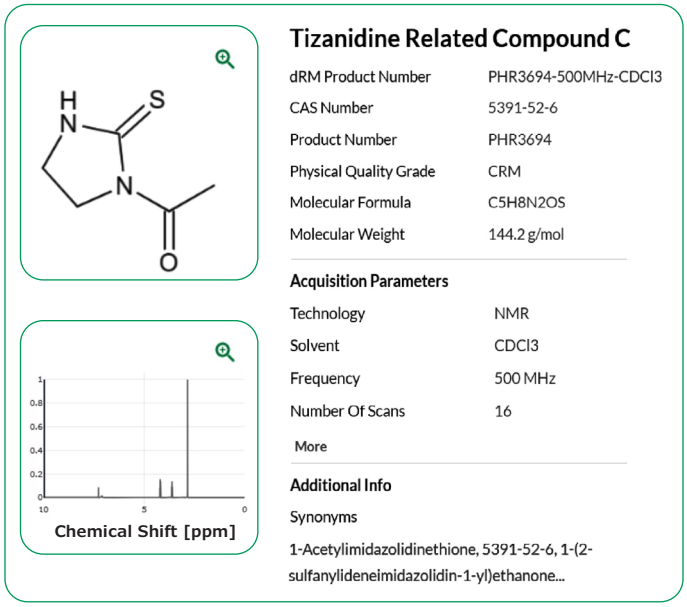

Figure 1: Detailed view of a dRM in the ChemisTwin™ interface with metadata (molecular formula, molecular weight, dRM product number, CAS, and product numbers) and ^{1}H NMR spectrum of the compound with corresponding acquisition parameters (NMR frequency, solvent, relaxation time, receiver gain).

The production of a dRM – Selection criteria for physical materials

A digital Reference Material is created only when molecules give a distinct and characteristic ^{1}H-NMR spectrum. But when and why does that definition apply?

The trivial example is when compounds containing no ^{1}H nuclei, that cannot be analyzed in a meaningful way with ^{1}H NMR spectroscopy. Molecules containing only exchangeable protons are disregarded for dRM production, as chemical shifts and linewidths of resonances depend strongly on choice of solvent, measurement temperature, and sample concentration. This means that spectra will depend strongly on measurement conditions, and we recommend using physical materials and calibration curves. And last, we waive mixtures of materials as dRMs, as not the spectrum of mixture is the reference, but the component spectra, which fulfil the above definition of distinct.

NMR cannot distinguish between enantiomers so compounds with one chiral center are, from an NMR viewpoint, pure. For example, *L* and *R*-alanine give the same NMR spectrum as a mixture of *L* and *R*-alanine in any ratio. This means that compounds with one chiral center are always suitable for dRM production.

In contrast to enantiomers, diastereomers (often) give different NMR spectra. In a mixture of diastereomers, the overall number of components (and spectra) depends on the number of chiral centers *n* in a molecule. With two possibilities per chiral center (*R, S*) the number of diastereomers is 2^{n}. However, every diastereomer has a mirror image (enantiomer), so that the number of component spectra becomes $2^{n}/2$. For dRM production this means that (in most cases) all chiral positions need to be defined. The spectrum is a reference only for this exact stereochemical configuration (and its exact mirror image).

Some molecules may exhibit tautomery, which may result in different structural isomers each associated with their own distinct spectrum coexisting in solution. Due to the small energy difference between tautomers, the equilibrium composition can be very sensitive to small changes in temperature or polarity of the solvent. Spectra of compounds exhibiting tautomery do not have reference material character and are unsuitable to produce dRMs. In these cases, we strongly recommend the use of physical reference materials.

Quality criteria for the spectral data of dRMs

In this section we will discuss the quality criteria applied to the NMR data. NMR data is affected by parameters in the hand of the operator (sample concentration, solvent, measurement parameters) and properties of the sample (e.g., solubility in a solvent, exchange rates).

- Relevant metrics for data quality are:
- Signal to noise ratio (SNR) > 100
- Linewidth < 2 Hz
- Lineshape (Lorentzian)
- Absence of artifacts (e.g., shim drift)
- No overlap of impurities/solvent residual with analyte resonances
- Integrals (correspond to the number of nuclei in a chemical moiety)

In the following part, we will briefly discuss these metrics touching on the theoretical background of NMR spectroscopy and experimental realities.

We set a minimum SNR of 100 for data suitable for dRM production. Uncorrelated random fluctuations in the form of noise are unavoidable, limiting the ratio of information to randomness. In spectroscopy, that ratio is given by the signal to noise ratio (SNR), where an NMR signal and one resonance from a spectrum with good signal to noise are shown in **Figure 2** (top right). The Signal to Noise in NMR is proportional to the sample concentration, the square root of the number of measurement repetitions and many other factors related to experimental hardware.

We set a relatively lax criterion for the linewidth $\Delta\nu < 2$ Hz (for ^{1}H nuclei not affected by chemical dynamics). The linewidth expresses how well defined a frequency is (see **Figure 2**, top), which is related to the exponential decay of the signal. The decay is exponential (shown as the green hull curve in **Figure 2** (top right, insert) and $\Delta\nu = 1/\pi T_2^*$. Note that mobility and dynamics affect linewidths and exchangeable protons (OH or NH) or nitrogen sites undergoing amine inversion (e.g. piperazines) often have larger linewidth than other resonances in a spectrum.

Artifacts, in NMR language, refers to the undesired features in a spectrum caused by experimental flaws or hardware imperfections. The appearance of artifacts is familiar to experienced spectroscopists, and several examples are shown in **Figure 2**. A familiar artifact is caused by mechanical vibrations, for example from ongoing construction efforts or measuring very shortly after a Helium fill, can give rise to effects reminiscent of a *J* coupling (see **Figure 2**, middle). Another well-known artifact is the lineshape deviating from the expected Lorentzian shape (shim algorithm failing to reach the desired result in the allotted maximum time/ sample movement). If a sample is not sufficiently fixed in its spinner or if the shim is insufficient, the lines are broad or resonances "tail" (see **Figure 2**, bottom). Artifacts are obvious, and affected data is removed for dRM production by our team.

dRMs are created using real measurements of physical materials. A material and the deuterated solvent may contain impurities (for solvents often H_2O) giving additional NMR signals. For the spectra to be considered 'reference' quality, impurities must be very small and they should not overlap with relevant signals originating from the compound of interest.

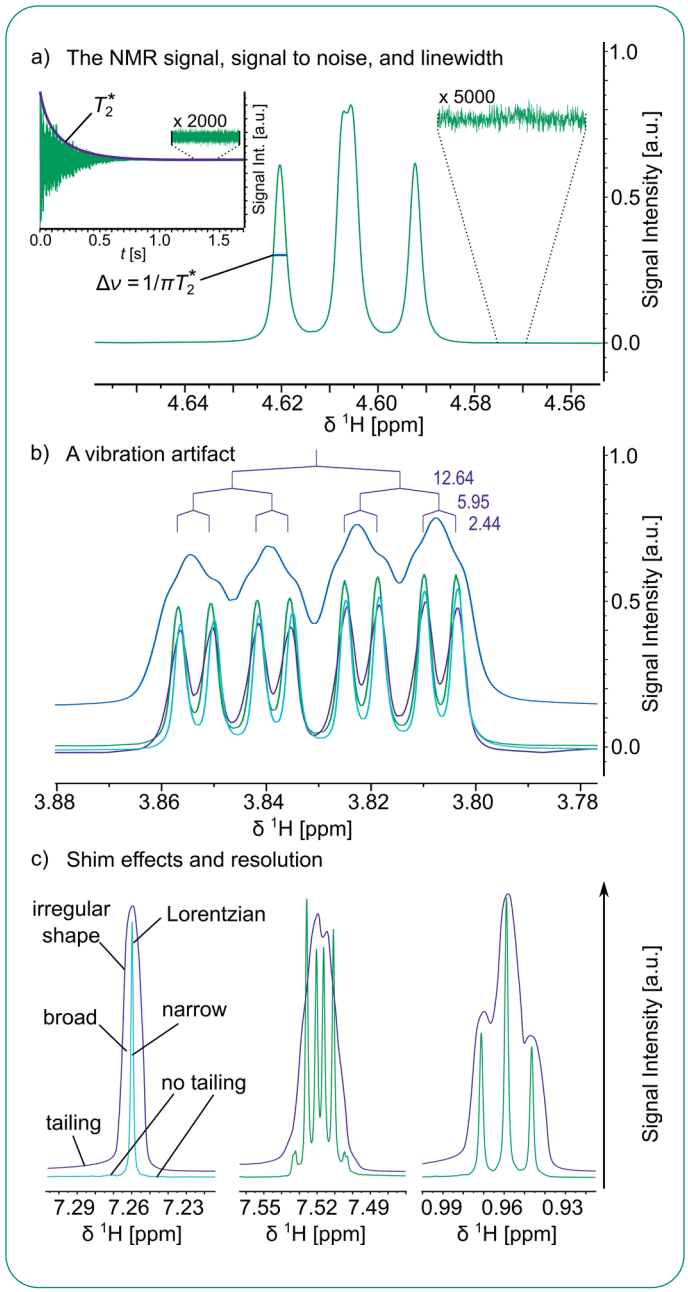

Figure 2: **a**) Example of signal to noise in an experiment. The NMR signal (**left insert**) is an oscillating voltage which exponentially decays with a time constant T_2^* (green hull curve). White noise (magnified x 2000) is caused by thermal motion of electrons leading to random voltage fluctuations. Top right: the Fourier transformed NMR spectrum with linewidth $1/\pi T_2^*$ and limited signal to noise ratio (noise shown x5000). **b**) Example of a low frequency vibration artifact in a spectrum. Three repetitions of a glycidol spectrum (cyan, green, purple) show the same doublet of doublets of doublets (ddd, J = 12.94, 5.95, 2.44 Hz) for the compound. The black trace is affected by low frequency vibration, which leads to an additional splitting (here 0.8 Hz). **c**) Effect of shim on resolution and lineshape. The spectrum of dibutyl phthalate in $CDCl_3$ exemplifies bad shim/sample position (red) and good shim (green). The solvent residual (left, 7.26 ppm, s) informs about spectrum quality. Analyte resonances (middle, right) exhibit the same lineshape, linewidth, and tailing effects as the solvent residual.

The integral under all peaks in an NMR spectrum (normalized to the full number of ^{1}H-nuclei) splits in such a way, that the integral under each resonance is the number of protons responsible for this part of the signal. However, we must define a metric for acceptable data quality with respect to integral values and how to deal with exchangeable protons. We make two allowances to the integral for chemical and technical reasons:

1. Exchangeable protons may be left unassigned. Protic deuterated solvents (e.g., methanol-d4) allow for OD/OH exchange and resonances vanish. In addition, chemical shifts of exchangeable protons are sensitive to experimental conditions (solvent, temperature, pH). Additionally, resonances may be extremely broad. This makes calculating correct and distinct integrals or predicting chemical shifts and multiplicities for exchanging moieties difficult. If exchanging protons are left unassigned, the integral normalization is adapted accordingly.
2. Qualitative and quantitative NMR spectroscopy involve different choice of methodology. In qualitative NMR, one wishes to be time efficient to maximize the value of a spectrometer. Thus, spectra are acquired with small excitation angles and small delays, but integrals can deviate from the expected value of exactly one per proton. We allow a ±22.5 % deviation for qualitative use dRMs, where a single proton site (e.g., CH, OH) may have an integral of 0.78 – 1.22. Quantitative NMR spectroscopy is not concerned with conserving experimental time, but with accuracy. To acquire quantitative spectra, measurement parameters are chosen to fulfill the expectation of a (normalized) integral of exactly one per proton, as this is the basis of quantification.

Assigning spectra to create the digital Reference Material

After the data has passed the initial screening for quality, all resonances in the 1D ^{1}H NMR spectrum are assigned to the molecular structure. The basis of assignment are integrals, chemical shifts, multiplicities, and values of coupling constants. If multiplicity, chemical shift, and integrals provide insufficient information, additional insight may be gained using established two-dimensional NMR methods.

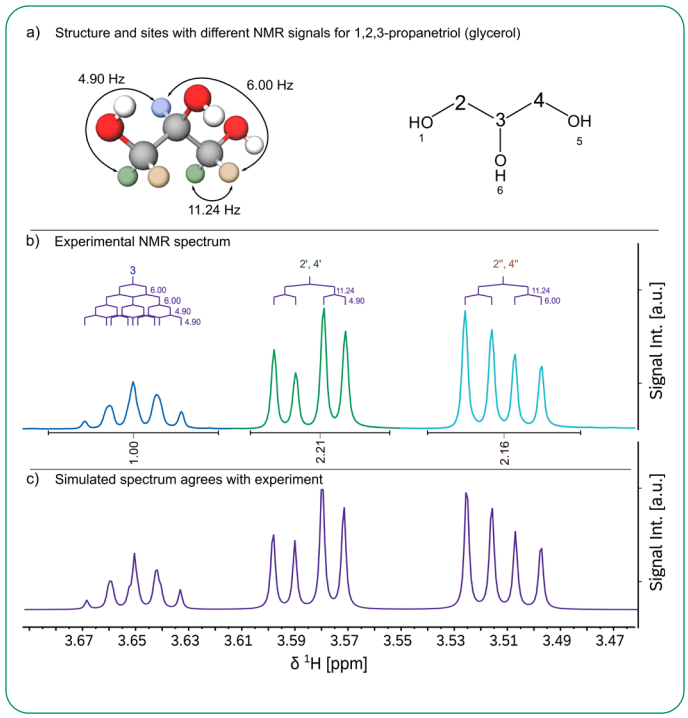

Figure 3: **a**) A conformer of 1, 2, 3-propanetriol as 3D structure with the three different symmetry environments of ^{1}H (blue site, green sites, cyan sites) and values of the *J* couplings acting between the sites. Each site has a slightly different electronic environment and thus a distinct chemical shift and characteristic *J*-couplings to other nuclei. **b**) The experimental spectrum in methanol-d4 shows only the CH-bound ^{1}H nuclei, OH is invisible due to exchange with the deuterated solvent. The color coding of resonances in the experimental spectrum corresponds to the sites shown in the 3D structure, the integral corresponds to the number of nuclei, the *J*-coupling network is shown as *J*-tree diagram. The integral (black number below the resonance) identifies the number of nuclei. **c**) The simulated NMR spectrum corresponds to the experimental data.

Figure 3 exemplifies how identification of the (primary) structure of 1, 2, 3-propanetriol (glycerin) using 1D NMR is performed. Here, the OH groups do not give a signal due to chemical exchange with the deuterated methanol and we only need to consider carbon bound ^{1}H-nuclei. In **Figure 3a**, both 2D and 3D structures (of one conformer) of glycerin are shown, which allows to recognize sites with different symmetry environments in the glycerol molecule.

In ^{1}H NMR, typically only H-H couplings over two and three bonds are observable and couplings between nuclei in same symmetry environments are invisible. Here, each “green” proton sees one “brown” proton with a geminal coupling (2-bond coupling ^{2}J) of 11.24 Hz and one “blue” proton with a vicinal coupling (3-bond coupling ^{3}J) of 4.90 Hz. Couplings acts between two nuclei and split both resonances so that two couplings give a doublet of doublets for the “green” proton The “brown” protons couple to “green” protons with the same value of 11.24 Hz. However, their coupling to “blue” is different.

The resulting doublet of doublets has the familiar *J*-coupling of 11.24 Hz and one new, and different, coupling constant (6.00 Hz). The "blue" proton couples to two "greens" and two "browns" with the familiar coupling constants of 4.90 Hz and 6.00 Hz. Coupling to two spins which each coupling results in a triplet of triplets. Generally, the values taken by *J*-couplings are indicative of coupling types (e.g., vicinal or geminal), distinguish types of coupled nuclei (e.g. ^{1}H-^{19}F) and sometimes differentiate isomers (e.g. *cis-olefin* 10 Hz, *trans*-olefin 17 Hz).

The chemical shift of resonances in **Figure 3** is indicative of presence of an alcohol group on all the carbon nuclei (e.g., the CH3-group of methanol has a chemical shift of 3.34 ppm in methanol-d4). The integral ratio of 1:2:2 (**Figure 3**, middle, black numbers), presence of three different symmetry environments, values of the coupling constant and chemical shifts allows to identify the glycerol.

More complex compounds routinely require 2D NMR. We will in the following exemplify how 2D methods, such as Correlation Spectroscopy (COSY), Heteronuclear Single Quantum Coherence (HSQC), and Heteronuclear Multiple Bond Correlation (HMBC) experiments are used to elucidate structure.

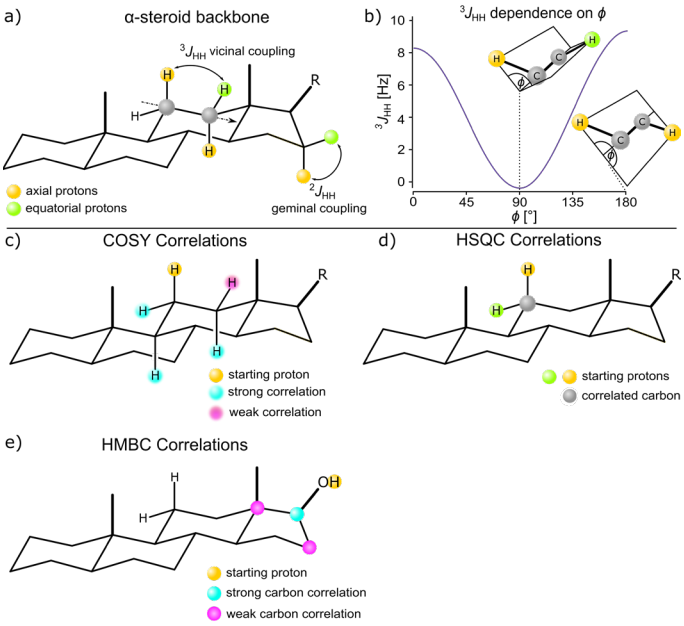

Figure 4: **a**) A generalized steroid backbone with axial (yellow) and equatorial (green) positions highlighted and one geminal and vicinal coupling shown. The arrow denotes the carbon carbon axis in b. **b**) The Karplus relation (curve uses standard parameters) correlates the expected values of the vicinal 3*J*-coupling and the dihedral angle Φ between two ^{1}H. **c**) Correlation of the ^{1}H highlighted in yellow. As evident from the Karplus relation, correlation to protons highlighted in cyan is strong, correlation to magenta very weak. **d**) Both starting protons would be correlated to the same carbon shift and identified as ^{1}H in a CH_2 group. This allows to distinguish geminal and vicinal couplings in the COSY. **e**) HMBC correlates ^{1}H and ^{13}C positions multiple (usually 2-4) bonds away and allows to use known rest (OH, $-OCH_3$, $-NH_2$) to start or continue assignments. Here, OH would be identified by having no HSQC correlation. The carbon highlighted in blue would be identified by a strong HMBC correlation. Combining all methods, this position is encoded as a CH in HSQC and the nearby positions highlighted in magenta are distinguished by HSQC. The highlighted quaternary ^{13}C has no attached ^{1}H (no HSQC signal), the other highlighted carbon shows HSQC to both attached ^{1}H nuclei identified as CH_2 group ^{1}H-nuclei. COSY allows to continue the assignment down the chain.

Compounds that often require multidimensional NMR are unsaturated (poly) cyclic structures, which can form various conformers and have substituents in equatorial and axial positions with different chemical shifts. **Figure 4a** shows a simplified projection of a generalized α steroid (several ^{1}H omitted, Rest R not shown).

With respect to COSY, the mechanism correlating nuclei is the *J*-coupling. COSY identifies which nucleus at which chemical shift is correlated with which other chemical shift and thus identifies neighbors. In NMR couplings between inequivalent nuclei are observable and axial and equatorial protons (marked in yellow and green in **Figure 4a**) are inequivalent. Geminal couplings between ^{1}H range from -45 to +23 Hz, whereas vicinal couplings depend on the conformation and dihedral angle (see **Figure 4b**). [2] This dependency on the angle and the vicinal coupling passing through 0 Hz can make COSY "blind" to a neighbor. In **Figure 4c** the ^{1}H nucleus marked in yellow and the ^{1}H marked in purple have a 90° dihedral angle realizing the $^{3}J \approx 0$ Hz situation and leading to a weak (or absent) correlation in COSY. This, falsely, suggests that the ^{1}H marked in purple is not nearby, which would be inconsistent with information provided by other highlighted nuclei.

HSQC compensates for the weaknesses of COSY: It correlates ^{1}H and ^{13}C chemical shifts and informs which ^{1}H is attached to which carbon and is usually carried out to distinguish CH/CH_3 groups from CH_2. By combining COSY and HSQC all highlighted nuclei in **Figure 4c** can be assigned from any arbitrary starting point.

HMBC is valuable to identify starting points for an assignment and to continue assignments over quaternary carbons. In **Figure 4e** we assume the organic rest R = OH. This OH position would not show a carbon correlation in HSQC, as ^{1}H is not bound to a carbon thus informing the operator that this resonance in an OH group. HMBC shows ^{13}C nuclei farther away (two to four bonds) from ^{1}H. This reveals the position marked in blue with a strong correlation and adjacent positions (purple) with a weaker correlation.

A combination of ^{1}H NMR, COSY, HSQC and HMBC is typically sufficient to fully clarify the structure of compounds with a molar weight of 1000 g/mol or less and a ratio of hydrogen to other nuclei of 1:2 (or more), but in rare cases the information about distances in 3D space provided by Nuclear Overhauser Enhancement Spectroscopy (NOESY) may be needed for assignment.

For example, the cyclobenzaprine molecule shown in **Figure 5** poses a challenge. The symmetry distortion induced by the rests R_1 and R_2 (H, CH_2-CH_2-$N(CH_3)_2$) is minute and, while sites are associated with certain spectral features, (e.g., magenta: doublet, yellow: doublet with additional small couplings, green: triplet with small coupling), a definite assignment is not feasible. Utilizing the HH-through space information (arrow in **Figure 5**) allows to identify a starting point for sequence-based assignment.

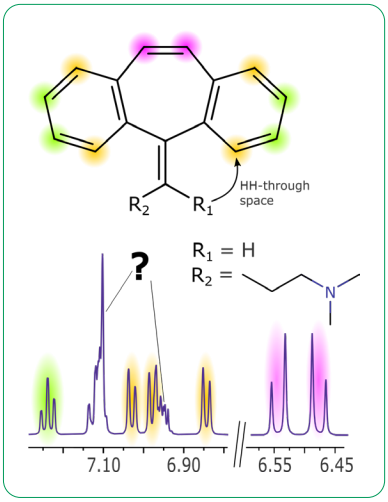

Figure 5: Sequential analysis breaks down for high symmetry compounds with slight distortions such as Cyclobenzaprine. While ^{1}H highlighted in yellow, magenta, and green have the same spectral pattern (e.g., magenta: one neighbor, doublet), the difference between R_1 and R_2 induced a relevant chemical shift difference between phenylic and olefinic protons. A definite assignment requires utilizing H-H through space coupling (arrow). Using the NOESY correlation between R_1 and one phenylic proton one nucleus is identified with certainty and sequential assignment is possible.

The Digital Reference Material & ChemisTwin™

In this article we have described how our team of experts uses ^{1}H NMR spectra of certified reference materials and other well characterized compounds to create a new library of digital references offering superior functionality to customers.

Our process is dedicated to quality spectra and materials, where we sort out bad data and ensure that compounds give definite and unique spectra to have reference character. We have illustrated, how chemical shift, integrals and multiplicities are used to obtain information about molecular structure and how uncertainties are removed with multidimensional NMR. With our three-step process involving different individuals preparing and evaluating the spectra and two review and control steps, we ensure a fresh and neutral view at spectra and metadata (CAS, product number, molecular weight) to guarantee correctness.

While this article is concerned with the production of digital Reference Materials (dRMs), the materials are stored and utilized on our ChemisTwin™ online platform.

With ChemisTwin™ portal, a user can search for reference compounds and their spectra, as well as the associated measurement parameters, which can be used to obtain data at identical conditions for optimal comparability and performance. ChemisTwin™ portal can compare user data to one or multiple references and quantify compositions. It offers various scripts for routine situations in laboratories.

The verification feature is meant for QC and identity analysis. Here, a user knows the compound and data evaluation is handled by ChemisTwin™ portal. Drag and drop raw data to evaluate similarity of your spectrum and reference data and get the identity verified and similarity expressed as “at a glance” numeric value. If the identity can be verified, the detailed analysis results include a full assignment of your spectrum, our values, predicted values by our AI algorithm and comparison of spectra. Additionally, the detailed analysis is available for download.

An untargeted search (identification) can be used for identification of the primary compound present in a sample, but more creative uses are intended. The possibility to compare your own assignment against our verified assignments allows you to gain the practical experience to utilize NMR to its maximum efficiency. The varied nature of components and structural motifs present in our portfolio provides access to a well-stocked source of empirical data in one single place. This is very useful for analysis of new and unknown compounds made in your lab. Comparing an unknown spectrum against the full library will preselect structure motifs that give parts of the spectrum, which aids significantly in evaluating spectra and avoids tedious manual browsing of similar compounds.

References

1. Albert Farre-Perez, J. F. P. Colell, C. Leonard, ChemisTwin - Your digital Twin for solving routine analytic problems - Scientific development and assessment of ChemisTwin Performance, MilliporeSigma White Paper, 2024.
2. M. Karplus, *J. Am. Chem. Soc.*, 85, **18**, 2870–2871 (1963).

Additional sources

For more information, please visit

SigmaAldrich.com/chemistwin

https://www.merckgroup.com/de/news/digital-reference-materials-chemistwin-14-11-2023.html

https://chemistwin.com

MilliporeSigma
400 Summit Drive
Burlington, MA 01803

To place an order or receive technical assistance

Order/Customer Service: **SigmaAldrich.com/order**
Technical Service: **SigmaAldrich.com/techservice**

SigmaAldrich.com

MS_WP13739EN Ver. 1.0
56560
07/2024