Supelco®
Analytical Products

Application Note

ChemisTwin™ - Your Digital Twin for Solving Routine Analytic Problems

Scientific Development and Assessment of ChemisTwin™ Performance

29 March 2024

Dr Albert Farre Perez

Principal NMR scientist for ChemisTwin™ portal

Merck KGaA NMR center of excellence, Buchs, Switzerland

Dr Johannes F. P. Colell

NMR expert, quality lead for ChemisTwin™ portal

Merck KGaA NMR center of excellence, Buchs, Switzerland

Dr Coralie Leonard

Digital and Chemistry R&D scientist, ChemisTwin™ product lead,

Merck KGaA distribution center, Lyon, France

The Life Science business of Merck KGaA, Darmstadt, Germany has been manufacturing and distributing Certified Reference Materials (CRM) for over 40 years. Based on our previous experience in producing high quality CRMs, we introduce our new portfolio of digital Reference Materials (dRMs). dRMs are created following our quality management processes dedicated to rigorously correct assignment of resonances to molecular structures using high quality one and multidimensional NMR spectra.

ChemisTwin™ portal provides cloud-based access to our digital Reference Materials (dRMs), where you can search molecular structures and their corresponding spectra or drag and drop your own measurement data for comparison with our dRMs database (qualitatively and quantitatively). Our smart algorithm can be used as part of quality control verifying spectra against our constantly growing portfolio of substances and their digital Reference Materials or to find commonalities between your spectra and reference spectra to aid in analyzing spectra of new compounds – ChemisTwin™ portal is your digital twin for routine and complex NMR spectra interpretation.

1. Introduction to ChemisTwin™ Portal

ChemisTwin™ portal aims to simplify qualitative and quantitative analysis of ^{1}H NMR data and accelerates quality related tasks. Your analytical data is compared against our database, resonances assigned, differences identified, and results are presented in a comprehensive way. This saves time on literature search or recording your own reference spectra, aids in standardizing the NMR spectra interpretation and offers an investigator increased certainty in their results.

ChemisTwin™ portal offers two possible workflows for both qualitative and quantitative analysis of single component analysis via NMR.

Qualitative Analysis

- **Identification of your compound via a Library search:** Simply drag and drop your own raw data and our untargeted analysis will process your data and browse ChemisTwin™ portal'sextensive library of digital Reference Materials for the best match to your data and proposes structures without needing to know the structure you are looking for.
- **Verification of your sample:** Verify your sample identity and get all resonances identified and assigned using our carefully evaluated experimental data of that same compound.

Quantitative Analysis

- **Targeted:** In a targeted search you know the molecule you are interested in. ChemisTwin™ portal provides instructions on how to prepare and measure your own data for accurate results. Our smart algorithm compares your NMR data and input structure with reference data, calculates a match factor (confidence), and determines the sample concentration.
- **Non-targeted:** In a non-targeted analysis you just provide measured data – ChemisTwin™ post-processes the spectra, and you select which peaks are to be quantified to calculate your sample concentration.

We are at the beginning of our journey with digital Reference Materials and the ChemisTwin™ platform, and we are constantly developing this novel tool for:

- Expert working group to investigate the certification of digital references using **ISO17034** framework
- Mixture analysis
- Custom digital references
- 13C database, multidimensional spectra,
- GMP version

Furthermore, additional technologies will be launched starting with an IR module for the dRM database mid-2024.

1.1. Choice of software

Our proof-of-concept study showcased several existing solutions available for NMR spectra comparison algorithms. Therefore, all viable solutions were evaluated with respect to performance in the envisioned qualitative (automated peak identification, structure verification, match factor) and quantitative (content determination and calculation of mass fractions) analyses with focus on ^{1}H-NMR.

ACD Labs was selected as the best solution to power the ChemisTwin™ application NMR comparison module. ChemisTwin™ portal's customized version of ACD Labs software performs the analysis of 1D-NMR spectra with focus on ^{1}H nuclei. This NMR predictor software provides two Neural Network algorithms enabling our automatic interpretation of spectra. A customized version of these algorithms is embedded into ChemisTwin™ portal and are continuously further trained by our digital reference materials database.

Besides the fast and simple full spectral analysis of chemical compounds, ACD Labs can store analytical data in an internal library, which we use to create our digital reference materials.

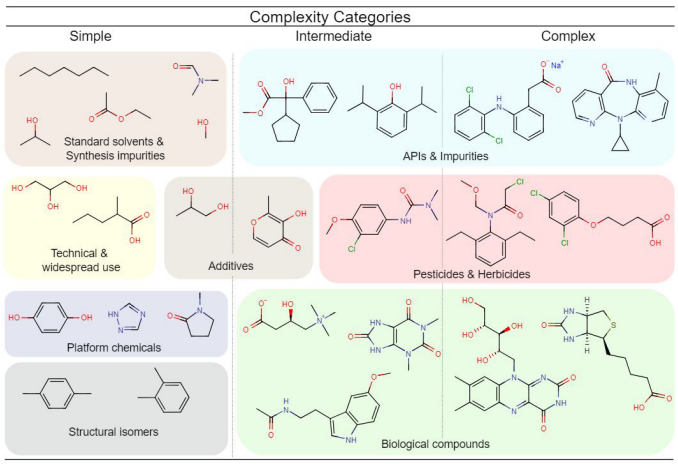


Figure 1: The dRM database contains molecules from all parts of our portfolio.

The dRM database includes structurally **simple molecules** such as solvents, technical compounds, synthesis building blocks, and platform chemicals but also **Active Pharmaceutical Ingredients (APIs)** and their **impurities/related compounds** as well as **biological structures, pesticides & herbicides**, enzymatic transformation products and **food and cosmetics additives** (see **Figure 1**).

1.2. Match Factor

The Match Factor (MF) describes the similarity between the input spectra and the digital Reference Material. The match factor provides an "at a glance" identifier for the similarity of your experimental data with a previously analyzed dataset, but also accounts for changes of a spectrum, for example a chemical shift changing as a results of concentration differences.

The digital reference materials use the mathematical expression of a spectrum, which takes the form of a sum of many peak functions with well-known analytical shape (Voigt functions). These are defined by their center frequency, intensity, and linewidth. A list of these parameters fully quantifies a spectrum. Thus, a customer's spectrum, and our reference, can be compared in terms of their mathematical descriptions and their similarity expressed as a match factor using ACD Labs sophisticated algorithm.

To ensure a pleasant customer performance, we performed an extensive validation study which has provided a large set of empirical data. Our experiments aimed at defining a match factor threshold that would have no false positive result without explicit warning for the "Satisfactory (***)" match, and at the same time would limit the likelihood of "Questionable (**)", or worse, "Unsatisfactory (%, *)" result when matching a sample with the correct dRM.

We find that the match factor exhibits the following properties:

- A Match Factor ≥ 93% indicates a sound likelihood that the target analyte is matched. These results are indicated in the portal as "Satisfactory (***)".

***** Satisfactory result: > 92%**

- A match factor ranging from 92 – 50% indicates compounds structurally closely related to the target molecule. These results will be labelled "Questionable (MF 75 – 92%, **)" or "Unsatisfactory (MF 50 - 74%, *)" by the software and require further investigation by the user.

**** Questionable result: > 92% - 75%**

*** Unsatisfactory result: > 74% - 50%**

- A Match Factor lower than 50% indicates a mismatch of molecular identities, where the molecular structure of the analyte is not the same as the dRM it was compared to. In this case, results are indicated as "No Match" (no stars).

No Match Found!

Glycine (**Cas No. 56-40-6**) has a match factor below 50%

Based on the empirical data, questionable and satisfactory (match factor > 74%) are considered positive, unsatisfactory and no match are considered negative results.

For our experiments we assessed the Match Factor thresholds based on the following possible results:

A **true positive** occurs when a spectrum is matched with the correct molecule, a **false positive** is a match with an incorrect molecule.

A **true negative** occurs when experimental spectra get "no match" or "Unsatisfactory" with an incorrect molecule. **False negatives** occur when spectra result in no match despite the dRM corresponding to the structure being present in the database.

Table 1: Summary of match factor ranges their meaning, match quality and result.

Match Factor	Test		Result
< 0.5	True/False negative	No match	No Match
0.5 - 0.74	True/False positive	*	Unsatisfactory
0.75 – 0.92		**	Questionable
≥ 0.93		***	Satisfactory

2. Performance Assessment

It is important to note, that it is not the well-established NMR spectroscopy (see ref 3 & 4) as an analytical method that is validated or assessed herein, but the automated processing, quantification, and interpretation of raw NMR data with our AI powered digital solution.

Using the ChemisTwin™ portal, you can compare your NMR raw data against the its dRM database to automate your analysis of a compound of interests' identity and quantity. We used our extensive knowledge as Certified Reference Materials manufacturer (**ISO 17034**) to define which performance parameters needed to be evaluated and performed measurements in the accredited measurement laboratories (**ISO 17025**).

2.1. Performance markers evaluated

The performance markers for each of the applications offered by the ChemisTwin™ portal are summarized in **Table 2.** It should be noted that not all performance markers are applicable to each of the applications.

Particularly noteworthy to performance are specificity, selectivity, and robustness, which are immediately related to a positive user experience and to providing you with a trustworthy scientific solution. For selectivity the algorithm needs to be capable of preferentially selecting the correct molecule from a set of similar compounds and specificity is given when ChemisTwin™ portal can confirm a spectrum in a direct comparison of identical compounds.

The second important consideration is robustness, where ChemisTwin™ portal needs to be flexible with respect to typical variables, such as solvent choice, NMR frequency and sample concentration. This is a must for your spectra to be conveniently comparable to the dRM's measurement without having to closely reproduce our experimental condition (solvent, instrument frequency or manufacturer). Some laboratories may only be interested in specific solvents, for example a laboratory investigating environmental pollutants may be especially interested in spectra in water (D_2O), whereas our dRM can be in a different, more commonly used, solvent for that compound.

Table 2: Types of analysis (qualitative, quantitative, library search) and applicable performance markers for ChemisTwin™.

Specification	Qualitative		Quantification
	Verification	Identification	targeted/non-targeted
Specificity	Yes	Yes	N.A.[a)]
Selectivity	Yes	Yes	N.A.[a)]
Linearity	N.A.	N.A.	Yes
Precision	Yes	Yes	Yes
Robustness	Yes	Yes	Yes
Limit of detection	Yes	Yes	N.A.[b)]
Limit of decision	Yes	Yes	N.A.[b)]
Limit of quantification	N.A.	N.A.	Yes
Accuracy	N.A.	N.A.	Yes

a) Specificity and Selectivity are not applied for Quantification analysis scenarios because these are solely focused on the calculation of the concentration. **b**) performing quantification close to the limits of decision or detection (LoDs) is not recommended, as signal noise ratio approaching 1:1 results in significant errors and is therefore considered not applicable.

Robustness is related to repeatability, reproducibility, and linearity. While these metrics inherent to the method NMR, they require evaluation for the portal. Testing repeatability and reproducibility is crucial to establish independence of the results from input data formats, spectrometer type/generation/manufacturer, operator, laboratory, and slight differences in linewidths between experiments.

And last, the limits of the algorithm defined by the three limits of detection, decision, and quantification. Using the above example of aqueous samples, a pollutant may have small solubility in water and the quadratic scaling of the signal to noise ratio with the number of scans allows a customer to select the correct measurement parameters.

In the qualitative analysis workflows, quantitative metrics such as limit of quantification and accuracy are inapplicable. Relevant and applicable metrics are selectivity and specificity as well as data quality, for example signal to noise ratio or linewidth, affect results so that limits of detection/decision remain relevant performance metrics. Last, tolerance to different experimental condition (robustness to solvent, NMR frequency) are relevant metrics.

For quantitative analysis limits of detection and decision are not applicable, as the error is proportional to the Signal to Noise Ratio (SNR) and sample concentration and measurement parameters must be chosen to attain large SNR. Likewise, specificity and selectivity are not applicable to quantification, as the NMR signal is proportional to the number of spins and any component may be used to reference any other if relevant parameters are known.

Table 3: Key constituents of reliability for ChemisTwin™ portal, and the investigation methodology for each analysis type.

		Investigation methodology		
		Qualitative		Quantitative
	Specification	Verification	Identification	targeted/non-targeted
1	Specificity	Main component of a sample (impurities allowed) can be verified	Main component of a sample (impurities allowed) can be identified against the whole library	N.A.
2	Selectivity	Incorrect compound is discriminated	Main component of a sample (impurities allowed) can be identified against the whole library	N.A.
3	Linearity	N.A.	N.A.	Linear correlation between experimental and theoretical concentrations
4	Precision	Evaluation of Repeatability and Reproducibility	Evaluation of Repeatability and Reproducibility	Evaluation of Repeatability and Reproducibility
4a	Repeatability	Processing data from multiple measurement of the same sample on the same instrument affords the same match factor	Processing data from multiple measurement of the same sample on the same instrument provides the same match factor	Processing data from multiple measurement of the same sample on the same instrument provides the same concentration
4b	Reproducibility	Processing spectra under identical conditions when measured by different individuals, on different instruments, or in different laboratories afford the same results.	Processing spectra under identical conditions when measured by different individuals, on different instruments, or in different laboratories afford the same results	Processing spectra under identical conditions when measured by different individuals, on different instruments, or in different laboratories afford the same results
5	Robustness	Varied solvents and magnetic fields • Solvents (DMSO, water, methanol, chloroform, acetonitrile) • Magnetic fields (400-600 MHz)	Varied solvents and magnetic fields • Solvents (DMSO, water, methanol, chloroform, acetonitrile) • Magnetic fields (400-600MHz)	Varied tube diameter: • 5 mm • 3 mm
6	Limit of detection	Threshold of sample dilution to verify the main compound	Threshold of sample dilution to identify main compound	N.A.
7	Limit of decision	Threshold of sample dilution to verify the main compound with satisfactory results	Threshold of sample dilution to identify the main compound with satisfactory results	N.A.
8	Limit of quantification	N.A.	N.A.	Threshold of sample dilution to quantify the main component

2.2. Test molecules

For testing the verification workflow (targeted qualitative analysis), we used a restricted dRM library of 700 molecules and selected a subset of 65 molecules to attain a good balance between calculational time and keeping a statistically representative sample size for our selectivity and specificity assessment. With this sample size we have 45,500 tests run by the software as illustrated in **Figure 2**.

For testing the library search (untargeted qualitative analysis) the number of overall compound present in the library increases the difficulty in finding a correct match, hence it is not sensible to restrict the library size and test set was tested against the full library of presently 2065 compounds.

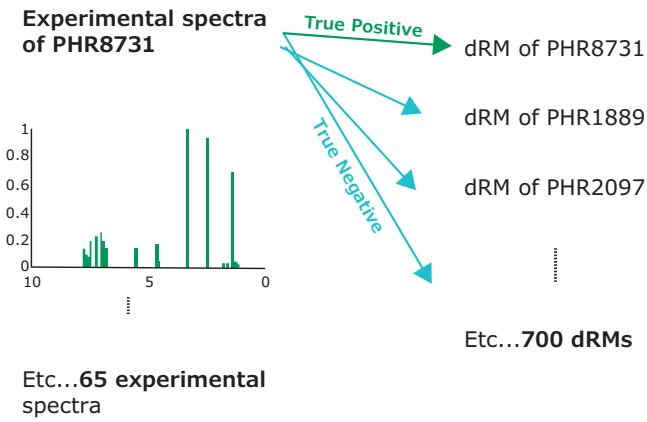


Figure 2: Methodology to assess selectivity and specificity for the qualitative analysis in ChemisTwin™ portal.

In the subset of 65 molecules 20 – 25 molecules are chosen from each of the following categories:

Simple: Molecules containing less than 18 protons and/or molecular weight lower than 120 g/mol

Intermediate: Molecules with 18 to 30 protons and/or molecular weight lower than 240 g/mol

Complex: Molecules with more than 30 protons and/or molecular weight exceeding 240 g/mol

To assess specificity of the ChemisTwin™ algorithm, it is relevant to create a test scenario, where different molecules are closely structurally related, and their spectra share many similarities. The property of interest is an algorithms' ability to select the correct match from a small subset of similar spectra.

To that end we chose derivate structures of drug backbones, documented in pharmacopeia entries [**2**] and are referred to as "Related Compounds" (RC), or Impurities, in the US and European Pharmacopoeia entries of drugs, respectively. These impurities can be isomers, homologues, analogues, and derivative compounds. We chose three APIs of different complexity (Propofol **PHR1663**, Rizatriptan Benzoate **PHR1889**, and Candesartan Cilexetil **PHR1854**) (see **Figure 2** and **Table A1**) for our test set of molecules.

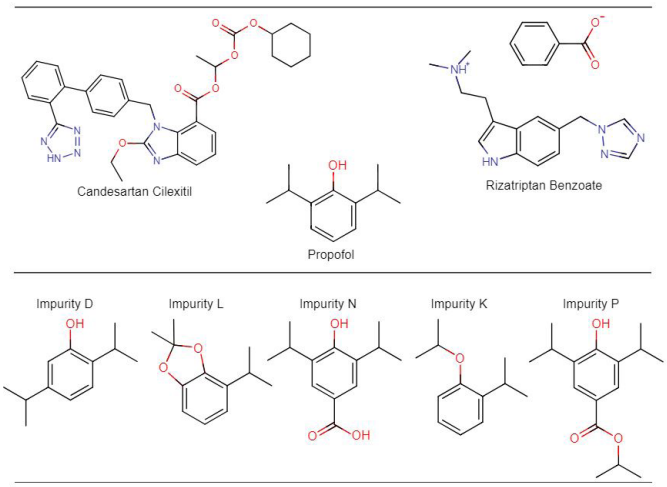


Figure 3: Parent structures of the drug APIs Candesartan Cilexetil, Propofol and Rizatriptan Benzoate used for selectivity evaluation and their product numbers (top) and Propofol Impurities according Ph. Eur. (bottom).

3. Results and discussion

3.1. Qualitative analysis

3.1.1. Verification of compound identity: qualitative targeted

3.1.1.1. Performance Goal

A good performance of the ChemisTwin™ portal using the verification workflow must return a "satisfactory" or, in few cases "questionable", result when comparing an experimental spectrum with the dRM of the corresponding compound.

3.1.1.2. Specificity

The specificity was evaluated by the ability to correctly confirm the identity of a compound. This was tested by uploading the experimental spectrum of a compound and selecting the corresponding dRM (the molecule itself) for comparison.

63 of the 65 tested molecules (97%) gave a true positive (successful result). Out of these, 80% fell into the "satisfactory" range (MF > 93%, ***) and 17% fell into the "questionable" range (MF 92% – 75%,**). 2 out of the 65 tested molecules failed. One molecule fell into the "unsatisfactory" range (MF 74 – 50%, *), which was caused by a peak of residual water in close proximity to a peak from the compound. The software erroneously assigned the water peak instead of the molecule peak, leading to a lower integral than expected and significantly impacting the match factor. The other molecule did not find a match (1.5%).

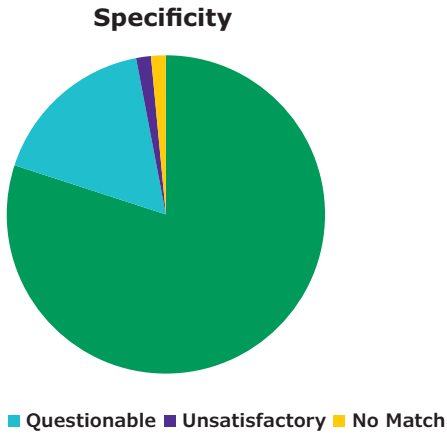


Figure 4: Distribution of match factors by confidence category for the verification workflow.

Table 4: Summary of the specificity results.

Satisfactory - True Positive 0.93-1.00 (***)	52	80.0%
Questionable - True Positive 0.75 and 0.92 (**)	11	17.0%
Unsatisfactory - False Negative 0.5 and 0.74 (*)	1	1.5%
No Match - False Negative < 0.5	1	1.5%
	65	**of 65**

Satisfactory results provide an immediately positive identification. While questionable results require an additional check by the operator, they provide good guidance. ChemisTwin™ porta was capable of successfully confirming a compound in 97% of the cases.

For the tests where ChemisTwin™ portal failed to automatically confirm the compound identity, we are continuously working on improving the performance, but it is to be noted that ChemisTwin™ portal cannot overcome limitations intrinsic to the 1D NMR technique itself and various structures exist where molecular identity can only be confirmed using 2D NMR or even other analytical techniques (e.g. distinguishing enantiomers requires optical rotation/ X-Ray diffractometry).

3.1.1.3. Selectivity

In the context of analytical chemistry, selectivity refers to the ability to choose a specific compound in the presence of impurities This definition aligns with the portal, where we consider the dRMs of other substances in the database "impurities". Selectivity in the verification workflow is the capacity to discriminate spectra from other molecular identities, in particular from structurally closely related compounds.

As the 65 target compounds are a subset of the 700 test compounds, there is only one true positive per compound. With respect to the capacity to discriminate molecules from the wrong identities, we find that 96.6% of the tests resulted in a true negative (93.9% no match, 2.7% MF 0.5 - 0.74 (unsatisfactory, *)). This means, that most compounds are identified as "impurities" and immediately sorted out.

False positives (FPs) accounted for 3.4% of the remaining tests (2.8% MF 0.75 – 0.92 (**), 0.6% MF ≥ 0.93 (***)). Results with MF 0.75 – 0.92 are uncritical, as these MFs indicate necessity for manual control by the operator.

The false positive analyses with MF > 0.93 (***) result from allowing for impurities to be present in the experimental spectra. Allowing for impurities means that it is possible for the software to identify a part of the spectrum as a match for a different compound and treat the rest as impurities. Thus, a high match factor can be calculated if molecules share many structural features and spectral patterns.

Figure 5: Distribution of match factors by confidence category for the verification workflow. Note that perfect selectivity corresponds to 65 satisfactory and 635 no match category results.

Table 5: Summary of the specificity results.

No Match – True Negative (0.00-0.50)	42674	93.9%
Unsatisfactory result – True Negative 0.50-0.74 (*)	1214	2.7%
Questionable - False Positive 0.75-0.92 (**)	1277	2.8%
Satisfactory - False Positive 0.93-1.00 (***)	270	0.6%
	45435	**of 45435**

The ability to distinguish a target compound from the related compounds was assessed by using the spectra of the three standard compounds. These compounds were tested against their respective dRMs and several related compounds (see **Figure 2** and **Table A1)**.

The structurally simplest of the test molecules (**PHR1663**) was discriminated from its impurities, with one true positive (MF = 100%) and 4 true negatives (MF < 50%). The more complex components **PHR1889** and **PHR1854** exhibited the largest match factor for the target (MF 94% and 95%). Furthermore, ChemisTwin™ portal provided warnings about integral mismatches or missing groups for related compounds. It should be noted that differences in calculated match factors become less significant with increasing complexity of the molecules, as the signal from one proton is a smaller fraction when compared to 50 protons than when compared to 10 protons.

Table 6: Target and candidate compounds for assessing selectivity among related compounds.

Target Molecule	Candidates (MF)
PHR1663	PHR1663 (100%)
	PHR9075 (< 50%)
	PHR9074 (< 50%)
	PHR9049 (< 50%)
	PHR9062 (< 50%)
PHR1889	PHR1889 (94%)
	PHR9148 (81%)
	PHR9151 (< 50%)
PHR1854	PHR1854 (95%)
	PHR2017 (87%)
	PHR2018 (93%)
	PHR2021 (0%)

3.1.1.4. Repeatability, Reproducibility, and Robustness

Repeatability, Reproducibility and Robustness for the verification workflow were evaluated with the test set of APIs (**PHR1663, PHR1889** and **PHR1854**).

Experimental repeatability was tested by repetition of the fully automated analysis of data from 3 independent measurements per substance. The resulting spectra were analyzed against the respective dRM. The resulting match factor must be within the same confidence range. Success 100%.

Software Repeatability was evaluated using the same data file of a measurement of one sample with the verification workflow against the respective dRM. The resulting match factor must be identical. Success 100%.

Table 7. Results compilation of the different parameters tested in ChemisTwin™ portal using the verification scenario.

Parameters	Subclass	N° of experiments	Successful analyses
Repeatability	Experimental	9	9 (100%)
	Software	9	9 (100%)
Reproducibility	Laboratory	12	12 (100%)
Robustness	Frequency	9	9 (100%)

Reproducibility was evaluated by assessing the ability to verify the identity of a compound when the sample was prepared and measured in different laboratories. This was accomplished by measuring the three exemplary substances on four different instruments across multiple labs (see **Table 7**). The resulting match factor for each analysis must be on the same range of confidence. All the tests were found a satisfactory match factor demonstrating the reproducibility of the portal.

The robustness was evaluated by assessing the capacity to correctly verify the identity of a compound at different magnetic fields and in different solvents.

Samples of the three example compounds were measured at the most common spectrometer frequencies of 400, 500 and 600 MHz to evaluate the effect of frequency variation on the ChemisTwin™ performance. The resulting match factor for each analysis must be within the satisfactory and questionable range. All the experiments gave a successful answer (**table 7**).

Robustness to solvent variation was investigated by analyzing a spectrum in DMSO-d6 to create the dRM. For the evaluation the three test molecules were additionally measured in four widely used NMR solvents (D_2O, $CDCl_3$, MeOH-d4, and CH_3CN-d3) to assess the effect of the solvent variation on the ChemisTwin™ portal's performance. The test resulted in 13/13 successful experiments with a MF > 0.74 (**Figure 6**). Compounds Propofol (**PHR1663**) and Candesartan Cilexetil (**PHR1854**) were not tested in D_2O due to low solubility.

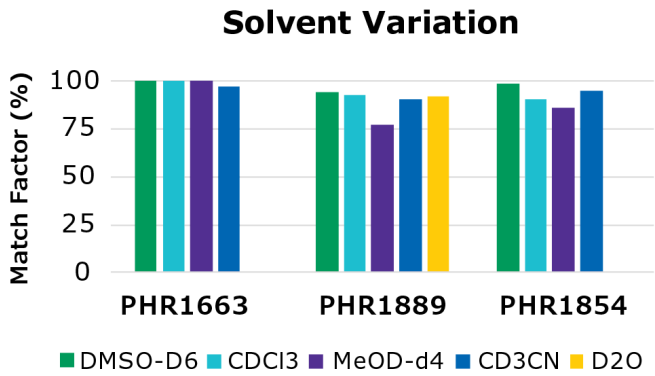


Figure 6: Match factor distribution for robustness with respect to solvent variation using the verification workflow.

3.1.1.5. Limit of Detection and Decision

The limits of detection and decision were evaluated by using 18 molecules measured at 3 or 4 different concentrations (around 0.005, 0.015, 0.05 and 0.1 mmol/g = mmol per gram of solution). The resulting match factor for each analysis must be within the satisfactory or, at worst, questionable range.

Out of 18 molecules, one molecule at the lowest concentration (around 0.005 mmol/g) resulted in a failure (no match). All the other measurements for all the molecules with concentrations ranging from 0.005 to 0.1 mmol/g resulted into a true positive. The molecule that failed at the lowest concentration was related to a difficulty with low signal to noise ratio for peaks split by multiple J-couplings leading to signal to noise below the threshold affecting two ^1H nuclei in a cyclohexyl-group and a low match factor. Molecules containing exchangeable protons may result in a lower match factor due to a current bug in the software. This bug attempts to assign the exchangeable protons to the peaks of the spectra. At low concentrations, the peaks of the exchangeable protons are not visible, and consequently, they are incorrectly assigned to other peaks, negatively affecting the match factor resulting in a false negative. The limits of detection and decision were extrapolated to 0.017 mmol/g for molecules containing exchangeable (NH/OH) protons and 0.006 mmol/g for molecules without exchangeable protons, respectively.

3.1.1.6. Conclusion

We have evaluated the performance of the verification workflow, demonstrating its success in verifying the correct molecule in most of the cases (97%). The qualitative targeted workflow exhibits selectivity and even effectively discriminates compounds from closely related compounds. We have established that the software performs reliably and consistently, and the process is robust. We found that spectra measured at different magnetic fields or different solvents could be reliably used in the identity verification workflow. We do. However, recommend using the same solvent for optimal verification workflow performance, as spectral differences in protic and aprotic solvents can be significant and affect reliability.

3.1.2. Identification of unknown sample by Library search

The identification workflow operates in 2 steps. The first step is an initial screening of the entire dRM library. In step 2 spectra are predicted using your sample experimental conditions (including solvent and frequency) and a match factor is calculated for all the dRM candidates identified in step 1 with match factor > 50%. In contrast to verification of a sample, testing against the full-sized library is required in all assessments, as here the number of potential mismatches scales directly with the number of dRMs.

3.1.2.1. Performance goal

An identification is successful if the correct compound has a match factor within the Satisfactory and Questionable range and there are no other compounds in a higher confidence category than the correct compound. For example, even if the correct compound (**true positive**) is in the Questionable range and there are no false positive matches in the Satisfactory range, the identification is considered successful.

Identification fails when a molecule does not find a match/falls in the Unsatisfactory match factor range, which we categorize as low severity failures. Furthermore, identification fails if a wrong compound falls into a higher confidence range, which constitutes a critical failure. For example, if the correct compound (**true positive**) is in the Questionable range, but there is a closely related compound shown in the Satisfactory range, the identification failed.

3.1.2.2. Specificity

The specificity was evaluated by the ability to identify the target compound in a library search, here against the full library (2065 structures). Note, that while the 65 test compounds have been measured at the same experimental conditions (solvent, and 600 MHz), the dRM database contains entries from various solvents and measured at various frequencies.

Out of 65 tested molecules, the correct match was found in 59 (91%) cases, out of which 79% fell into the satisfactory (MF > 93%, ***) and 12% into the questionable (MF 92% – 75%, **) confidence ranges. The portal failed to identity 6 molecules (9%), where 2 of those 6 molecules, fell into the unsatisfactory range (MF 74 – 50%, *), and four molecules found no match (6%).

Figure 7: Diagram of match factor confidence distribution for the specificity assessment of the identification workflow.

Table 8. Summary of the specificity results using identification workflow.

Satisfactory – True Positive 0.93-1.00 (***)	51	79%
Questionable – True Positive 0.75-0.92 (**)	8	12%
Unsatisfactory – False Negative 0.50-0.74 (*)	2	3%
No Match – False Negative 0.00-0.50	4	6%
	65	**of 65**

3.1.2.3. Selectivity

The selectivity of the identification tool was assessed by the capacity to discriminate spectra from the wrong identities. This was evaluated by application of the success and failure criteria outlined in section 3.1.2.1 an requires further analysis of results from 3.1.2.2.

58 out of the 59 tests that found the correct compound successfully identified the target compound with no other compound in the same or higher confidence range. One test resulted in a critical failure (incorrect compound in higher confidence range).

Out of the 6 tests that failed the specificity assessment (section 3.1.2.2), 5 are low severity failures, where no wrong candidate was found in a higher confidence range than the correct compound. The remaining test was a critical failure, as the correct compound found no match, and an incorrect compound was categorized in the questionable range. The critical failure is the result of a software limitation, where the issue is related to exchanging protons and spectral differences in protic and aprotic solvents and currently being addressed.

Table 9. Summary of the selectivity results using the identification workflow.

Identified	58	89%
low severity failure	5	8%
critical failure	2	3%
	65	**of 65**

3.1.2.4. Repeatability, Reproducibility, and Robustness

Repeatability, reproducibility, and robustness were tested in full analogy to 3.1.1.4 using the test APIs shown as parent structures in **Fig. 2** as test substrates.

The tests for experimental and software repeatability, reproducibility, and robustness to frequency variation were successful in all cases. However, library search proved to be much less robust to solvent change than verification. Here, only 5 out of 13 experiments were successful and the remaining 8 failed due to differences between spectra in different solvents. Results are summarized in **Table 10**.

Table 10. Compilation of results for the different parameters tested in the assessment of ChemisTwin™ portal identification workflow.

Parameters	Subclass	Expts.	Successful
Repeatability	Experimental	9	9 (100%)
	Software	9	9 (100%)
Reproducibility	Laboratory	12	12 (100%)
Robustness	Frequency	9	9 (100%)
	9	13	5 (39%)

It should be pointed out, that achieving robustness to solvent variation for an algorithm in the identification workflow is extremely difficult. The difficulties arise from the significant spectral differences between solvents, which include coupling over heteronuclei being visible/invisible, chemical shift changes for exchanging moieties of up to several ppm, and resonances vanishing by H/D exchange in protic solvents. Hence, a-priority knowledge is required even for human operators.

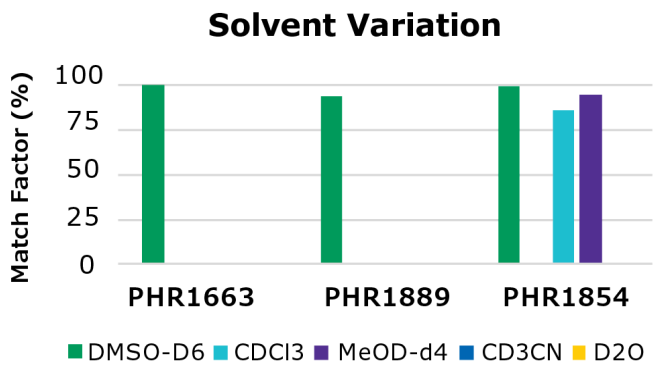


Figure 8: Robustness to solvent variation using the identification workflow.

3.1.2.5. Limit of Detection and Decision

The Limit of decision was evaluated in a similar way to 3.1.2.5 using the same 18 molecules and concentrations (0.005, 0.015, 0.05 and 0.1 mmol/g) as previously.

Out of all the measurements only three of the 18 molecules in their lowest concentration samples (around 0.005 mmol/g) found no match. All other measurements with concentrations ranging from 0.005 to 0.1 mmol/g resulted into a true positive. The molecules that failed at the lowest concentration were related to the difficulties involving exchangeable protons mentioned in 3.1.1.5. The limits of detection and decision were extrapolated to 0.017 mmol/g for molecules containing exchangeable (NH/OH) protons and 0.006 mmol/g for molecules without exchangeable protons, respectively.

3.1.2.6. Conclusion

We have evaluated the performance of the verification workflow, demonstrating its success in identifying the correct molecule in most of the cases (91%). The workflow exhibits selectivity and effectively discriminates against incorrect compounds and even closely related compounds (89%). We have established that the identification workflow is robust to different magnetic fields, but caution users that use of different solvents leads to a high chance of no match, as this workflow is susceptible to spectral changes.

3.2. Quantitative

3.2.1. Quantitative Targeted analysis

The quantitative targeted workflow requires that the sample you want to quantify must exist in the dRM library, as this workflow operates in two steps. The first step involves verifying the sample spectra with the corresponding dRM, where the software assigns all peaks to specific protons and subsequently determines the equivalent resonances for your spectrum and the reference and calculates a match factor. This step is necessary, because the NMR signal is proportional to the number of spins, and the stochiometric coefficients need to be known to quantify a concentration, which means it is relevant whether a resonance originated from e.g. a CH_2-group or CH_3-group. If the compound is plausible, interpreted by the software by a match factor of over 50%, the calculation of the concentration in solution occurs. When the customer provides the weights of the sample and the solvent, an expected concentration can be calculated by the portal. The division between the experimental concentration calculated by the portal and the expected concentration yields the purity of the sample. This purity is represented as a mass fraction which is expressed in g of sample/g total.

The quantitative targeted workflow was tested using 18 molecules prepared in samples of concentration close to certain target concentrations (0.005, 0.015, 0.05 and 0.1 mmol/g) but unknown exact concentration. A total of 193 measurements were acquired. An expectation value for the concentration for each measurement (expected concentration) was calculated by referencing to an external standard of known concentration.

Subsequently the data was uploaded to the portal and the quantitative analysis routine was used to determine concentrations, now using the digital reference material as digital external standard for quantification. The concentrations given by the portal were compared with the expected concentration this comparison is given as a mass fraction.

3.2.1.1. Repeatability

Experimental repeatability was tested by repetition of the fully automated analysis of data from 3 independent measurements per concentration per substance. The resulting spectra were analyzed against the respective dRM to assign the peaks to chemical moieties. The resulting mass fraction must be equal to 1 ±0.05 g_{sample} / g_{total} range. All the concentrations given by the portal were inside of the acceptance criteria. Success 100%.

Software Repeatability was evaluated using the same data file of a measurement of one sample at a defined concentration. The resulting mass fraction must be identical for all the repetitions. We found the calculated mass fractions given by the portal were to be identical for all repetitions. Success 100%.

3.2.1.2. Limit of Quantification

The limit of Quantification (LoQ) was evaluated by using 18 molecules measured at 3 or 4 targeted concentrations. The test failed for two out of 193 tests (3%) for compounds in their lowest concentration samples (c ≈ 0.005 mmol/g). The failure was related to the 1st verification step. For one molecule with a MF < 0.5 (no match) no concentration was calculated (see 3.1.1.5), for the second compound signal to noise of exchangeable protons was below the threshold resulting in misassignment of resonances due to mismatch of stoichiometry and integral. All other measurements (187 of 193) with concentrations ranging from 0.005 to 0.1 mmol/g fell into the specified mass fraction range of 1 ±0.05 g/g. The limit of quantification for targeted analysis was extrapolated to 0.017 mmol/g.

3.2.1.3. Linearity

The linearity was evaluated by comparing the experimental concentration provided by the portal with the theoretical concentration calculated using the external standard. To calculate the linearity between the experimental and the theoretical concentrations, at least 3 different experimental concentrations must fall within the range of mass fraction = 1 ±0.05 g/g. 17/18 molecules showed a good linear correlation between the concentrations, with an R^2 exceeding 0.999. For the remaining molecule, the linearity could not be calculated, as only 2 out of the 3 different concentrations met the acceptance criteria. We considered that having 17 out of the 18 molecules is sufficient to demonstrate the performance of the portal.

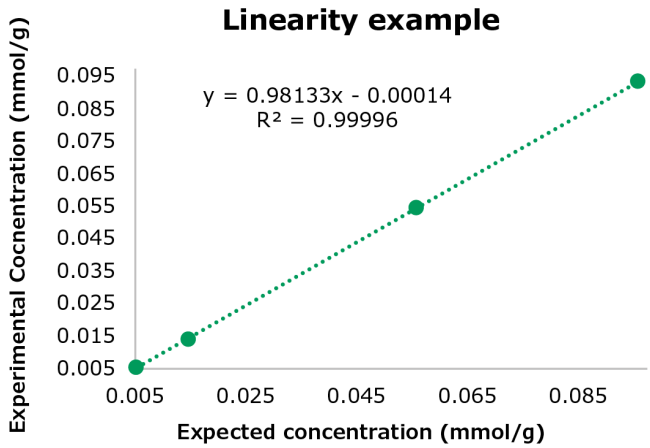


Figure 9: Plot of expected concentration against experimental concentration calculated by the portal shows the expected linear correlation and slope of 1 (see text).

3.2.1.4. Conclusion

We have assessed the performance of the quantitative targeted workflow, showcasing its efficacy in quantifying the target molecule in 97% the of cases. The workflow exhibits robust experimental and software repeatability. These results outline the determined quantification capabilities of this workflow which can operates from low to relatively high concentrations. This workflow has shown that molecules containing exchangeable protons (such as NH, OH) at lower concentrations may result in reduced portal performance due to signal to noise limitations for broad peaks.

3.2.2. Non-targeted Quantification

The quantitative non-targeted workflow, in contrast to other quantitative workflows, does not require the sample you want to quantify to exist in the dRM library. Once the sample spectrum is uploaded and interactive post-processing of the spectra is performed, a table appears on the analysis webpage containing all identified peaks with corresponding chemical shifts, multiplicities, relative integrals, and absolute integrals. You can then select the peaks you want to quantify along with the number of protons for each selected peak. Subsequently, the calculation of concentration and mass fraction takes place.

3.2.2.1. Repeatability

The repeatability was evaluated tested in full analogy to 3.2.1.1 using the subset of 18 molecules test substrates, but here with the quantitative non-targeted workflow. All the concentrations from the repetitions were inside of the acceptance criteria of mass fraction = 1 ±0.05 g/g.

3.2.2.2. Reproducibility

The reproducibility was assessed by measuring spectra of the subset of 18 molecules in different concentrations on multiple instruments at multiple sites. The resulting spectra are analyzed by the quantitative non-targeted workflow. All the experiments were inside the acceptance criteria of mass fraction = 1 ±0.05 g/g demonstrating the reproducibility of the non-targeted quantification workflow in the portal.

3.2.2.3. Robustness

The robustness of the quantitative non-targeted workflow was assessed by variating the NMR tube diameter. The same molecules at the same concentration were recorded in 5 mm tubes and the 3 mm tubes. All the samples recorded using a 3 mm tubes successfully passed giving a mass fraction within the range of 1 ±0.05 g/g.

3.2.2.4. Limit of quantification

The Limit of Quantification was tested using 18 molecules at 3-4 different concentrations (0.005, 0.015, 0.005, and 0.01 mmol/g). All the experiments yielded a mass fraction of 1 ± 0.05 g/g, indicating that the Limit of Quantification (LoQ) was determined to be above 0.005 mmol/g. The quantitative non-targeted workflow has a lower LoQ than the targeted, as the non-targeted approach does not depend on a verification step prior the content determination calculation.

3.2.2.5. Linearity

The linearity was evaluated in full analogy to 3.2.1.3 using the same subset of molecules. All the molecules showed a good linear correlation between the concentrations, with an R^2 exceeding 0.999 demonstrating the linearity of the portal. 18/18 molecules showed linearity, with an R^2 exceeding 0.999.

3.2.2.6. Conclusion

We have evaluated the performance of the quantitative non-targeted workflow, outstanding its efficacy in quantifying all the of cases. The workflow demonstrates great reproducibility as well as repeatability. These results outline the determined quantification capabilities of this workflow which can operates from low to relatively high concentrations.

3.3. Experimental

The experimental data to create the dRMs was gathered at our sites in Buchs/Switzerland, Darmstadt/Germany, and the US sites in Round Rock and Laramie using 400, 500, and 600 MHz NMR spectrometers (Bruker, JEOL) for data acquisition.

Qualitative ^{1}H NMR spectra were acquired using small pulse angle experiments with rapid repetition for maximized instrument-time utilization using the following characteristic parameters:

- Set number of scans $n_S \geq 8$
- Set interscan delay (D_1) to 1 s
- Set acquisition time to at least 2.5 s
- Set dummy scans to 2
- pulse angle 30°

Quantitative ^{1}H NMR spectra were acquired using well calibrated 90° pulse angles adapted to solvent susceptibility and concentration of the sample and calibrated to the frequency of the resonance to be used for quantification.

- Set number of scans $n_S \geq 16$
- Set interscan delay (D^1) to $\geq 5\ T_1$
- Set acquisition time to 10 s
- Set dummy scans to 2
- Set a Receiver Gain in the middle of the allowed region and verify that automatic gain adjust is not activated in the experiment method)
- Set pulse angle to 90°
- Set excitation pulse frequency on-resonant with to the be quantified resonance

An ampoule of Dimethyl terephthalate solution in DMSO-d6 (**Our Prod. Number 39387**) is opened and added into an NMR tube. Subsequent quantitative ^{1}H NMR measurement is conducted with the same settings mentioned at the previous paragraph. The Dimethyl terephthalate is used as external standard to determine the nominal concentration of the samples. All the concentrations used in the performance test are stated as mmol/g.

The raw ^{1}H NMR data of the samples are converted into .jdx files using Topspin (Bruker). The acquired spectra are uploaded into the portal for qualitative or quantitative evaluation by selection of the corresponding dRM for identification, in the case of quantification followed by the selection of the digital external standard. The quantification is performed by selecting the dRM that corresponds to the sample identity, which is used exclusively as digital external standard. The quantification is performed in the ChemisTwin™ portal which will conduct the content determination (concentration and mass fraction calculation). The results of the analysis were evaluated versus the our procedure for external quantification (**PULCON**).[3]

Prior to perform quantitative analysis in ChemisTwin™ portal, a single calibration of the instrument with each tubes type must be performed – all steps to perform this calibration are described in ChemisTwin™ portal, it includes recording the same calibration sample with six different receiver gains. Once this initial calibration has been performed in the portal, we only request a monthly “health check” with a single measurement to ensure the calibration is still valid for another month

4. Conclusion & next steps

The ChemisTwin™ portal has exhibited outstanding performance in verifying, identifying, and quantifying a target compound, demonstrating its suitability for the intended purpose.

It can automatically verify substances in 97% of all cases and successfully performs the much more challenging identification of compounds against a large set of other substances in 91% of the cases. Most of the false negatives were attributed to a software bug, which we are currently working to improve to enhance the accuracy of the results.

ChemisTwin™ portal can automatically detect without further check the incorrect substance in 96% of the cases. In 0.6% of the cases ChemisTwin™ algorithm gives a false positive using the verification workflow while testing against random dRMs. The identification workflow gives a 3% of false positives. Most of those false negatives occur because the software allows impurities in the sample spectra. This can cause the software to mistakenly identify a portion of the spectrum as matching a different compound, while treating the remainder as impurities. In most cases, false positives can be distinguished through a visual comparison between the sample and the dRM spectra. Additionally, the software issues a warning message indicating an excess of protons in the sample, prompting a double-check of the results.

The portal is capable of quantifying all the tested molecules using the targeted quantification workflow, spanning from low to relatively high concentrations. The non-targeted approach demonstrated that it can quantify even lower concentrations since it does not depend on the verification step.

References

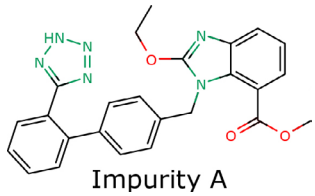
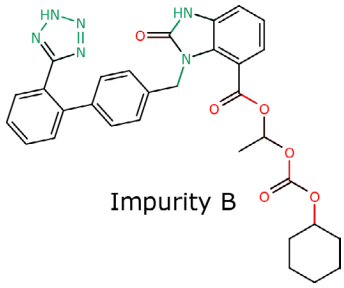
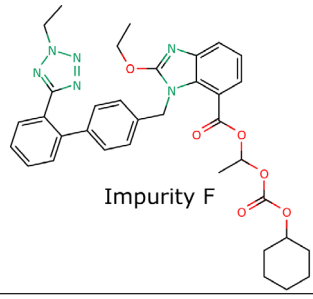
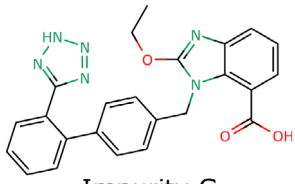
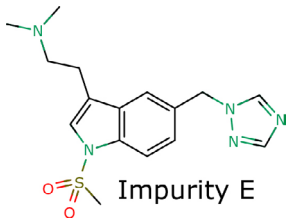
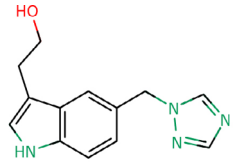
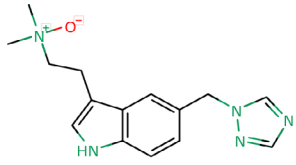
1. European Pharmacopeia 11.4 (https://pheur.edqm.eu/app/11-4/search/)
2. Wider G, Dreier L, J. Am. Chem. Soc., 2006, 128, 2571-2576
3. Anatole Abragam, Principles of Nuclear Magnetism, Clarendon Press, Oxford 1961, ISBN 0-198-520-14X
4. James Keeler, Understanding NMR spectroscopy, John Wiley and Sons Inc, 2010, ISBN 978-0-470-74608-0

Annex

Acknowledgments

We gratefully acknowledge the University of California Los Angeles (UCLA, CA, USA) and Orion for their participation in our validation study.

Table A1: Product numbers, chemical structures, and Ph. Eur identifier labels of Candesartan Cilexitil and Rizatriptan Benzoate impurities.

Product No.	Structure and Identifier
PHR2017	 Impurity A
PHR2018	 Impurity B
PHR2020	 Impurity F
PHR2021	 Impurity G
PHR9202	 Impurity E
PHR9148	 Impurity F
PHR9151	 Impurity H

4. Parameter definitions

4.1. Specificity

Specificity is the ability to assess unequivocally the analyte in the presence of components which may be expected to be present. Typically, these might include impurities, degradants, matrix, etc. [1998 Eurachem / ICH Q2A, CPMP/ICH/381/95]. For ChemisTwin™, Specificity is interpreted by the ability to correctly identify the main component of a sample in the qualitative workflow. The specificity in ChemisTwin™ was assessed by the identification/verification of a main component of a sample.

4.2. Selectivity

Selectivity is the ability to differentiate the components from each other. For ChemisTwin™, Selectivity is interpreted by the ability to correctly identify the main component in the qualitative workflow of a sample in comparison to two or more similar molecules.

4.3. Repeatability

Closeness of the agreement between the result of successive measurements of the same measurand carried out under the same conditions of measurements [*International Vocabulary of Basic and General Terms in Metrology, second edition, 1993 (VIM)*]. The repeatability in ChemisTwin™ is evaluated through the software and sample repeatability.

4.4. Reproducibility

Closeness of the agreement between the result of successive measurements of the same measurand carried out under the same conditions of measurements by a different researcher and instrument (of same performance). The reproducibility is evaluated by analysis under identical conditions same test materials different persons, instruments, and laboratories.

4.5. Robustness

The robustness of an analytical procedure is a measure of its capacity to remain unaffected by small, but deliberate variations in method parameters and provides an indication of its reliability during normal usage. [1998 Eurachem /ICH Q2A, CPMP/ICH/381/95]. The Robustness is evaluated in ChemisTwin™ in qualitative and quantitative by performing analysis varying experimental / environmental conditions such as solvents, Instrument's frequency or NMR tubes diameter.

4.6. Linearity

Defines the ability of the method to obtain test results proportional to the concentration of the analyte. [1998 Eurachem]. The Linearity is assessed in ChemisTwin™ using quantitative workflow. This evaluates the linearity between the theoretical concentration which has been calculated using an external standard and compared to the experimental concentration given by the portal.

4.7. Limit of Detection

The lowest content that can be measured with reasonable statistical certainty [1998 Eurachem/ AOAC-PVMC]. The LoD is evaluated by testing different concentration in qualitative workflows.

4.8. Limit of Decision

CCα means the limit at and above which it can be concluded with an error probability of α that a sample is non-compliant. The LoD is evaluated by testing different concentration in qualitative workflows.

4.9. Limit of Quantification

The accuracy of a quantitative NMR analytical procedure should be determined across the required analytical range [**USP 761**].

Merck KGaA
Frankfurter Strasse 250
64293 Darmstadt, Germany

MK_WP13836EN Ver. 1.0
56686
07/2024