Validation documentation
Hepart raw materials, liquid

HiperScan GmbH

March 22, 2021

Validation documentation of the chemometrics database of the near-infrared spectrometer *Apo-Ident*.
Report number: **61-2.3-2021-01**
info@hiperscan.com, www.hiperscan.com

Contents

Introduction

The unequivocal proof of the identity of pharmaceutical raw materials based on a monograph or traditional alternative methods is work-intense, time-intense and economically often no longer makes sense. Near-infrared spectroscopy (NIR) provides a new means here. It enables relatively easy, fast and nonetheless safe identity testing via the preparation and evaluation of spectra.

The analysis system *Apo-Ident* was developed specially for pharmacy use. Pharmacists have the duty to check the identity of all raw materials for extemporaneous products in their pharmacies. This normally takes place based on the monographs for the respective substances in the European Pharmacopoeia. But also NIR spectroscopy is described in the European Pharmacopoeia as an identification method which, as opposed to the methods incorporated in the respective monographs, is approved for testing [1] (quotation translated)

> *with the prerequisite that the same results ("namely the determination of the identity" [2]),* (quotation translated) *are achieved as with the described methods and instruments.*

The *Apo-Ident* analysis system serves to identify raw materials for prescriptions in the defined manner according to *ApBetrO* [Pharmacies Rules and Regulations] §§ 6 and 11 at pharmacies (NIR spectroscopy as an alternative testing method). *Apo-Ident* consists of three components:

- An *NIR spectrometer*, which records the spectra of non-preprocessed raw materials in a measuring glass in diffuse reflection or transflection.
- The *QuickStep* spectroscopy software controls the instrument and records the spectra and user inputs via a pharmacy-specific software plug-in. It also generates the test protocol for documentation of the testing and storage of the printout to be signed in the pharmacy.
- The software module *IdentModule* incorporates *reference databases*. The spectra from the *QuickStep* software are presented to it for evaluation.

NIR spectroscopy is a very powerful analytical method. It is also able to establish the identity of several chemical compounds and mixes in as far as an appropriate database (technically correct: a chemometric model) was created. Identity testing with *Apo-Ident* is a very safe, very fast and easy to operate analytical method for testing a large number of raw materials.

Context of this document

The suitability of the instrument, method and database is proven as follows:

- *NIR spectroscopy as a method for identity testing:* The *Ph. Eur.* [3] describes NIR spectroscopy in *Section 2.2.40* as an analytical method which is also suitable for the identification of raw materials. Therefore, validation of the method as such is not necessary.
- *Performance of the instrument:* The *Ph. Eur.* [3] furthermore describes the apparatus and the testing of its performance in *Section 2.2.40*. The document *Erfüllung von 2.2.40 Ph. Eur. durch Apo-Ident* [4] compares the implementation by *Apo-Ident* with this monograph in order to prove that *Apo-Ident* meets the specifications of the Pharmacopoeia. Each individual instrument delivered to a pharmacy is qualified in accordance with the tests described in *"Control of Instrument Performance"*. In this test, the unit consisting of analysis instrument hardware and the *QuickStep* spectroscopy software is assessed. The result is documented in a test protocol which is kept at the pharmacy.
- *Validation of the database* is documented separately for each substance class. The report at hand documents the substance class *Hepart raw materials, liquid*.

The *Arbeitsgemeinschaft der Pharmazieräte Deutschlands (APD)* [Working Group of German Pharmacy Inspectors] has clarified the following in its resolution dated October 16, 2013 ([5], quotation translated):

> *NIR is a testing method incorporated in the Pharmacopoeia. The testing quality depends on the quality of the database stored. The* APD *views the use of NIR instruments in case of ensured validation of the databases used in conjunction with it as one of several options for identity testing.*

The *APD* defined more precisely ([6], quotation translated) on October 1, 2014:

> *The use of near-infrared is a recognised testing method according to* Ph. Eur. 8. *For the use of NIR instruments in pharmacies for testing the identity of raw materials, sufficient and verifiable validation of the instrument used is required. The quality of the database stored by the instrument manufacturer is decisive for quality. Batch-specific differences with the same original substances must be taken into account if present.*

So NIR is basically suitable. The validity of the reference database is proven with the existing validation documentation.

Criteria for the inclusion of substances

This validation documentation describes the results of the validation of the reference database for the substance class *Hepart raw materials, liquid*. Validation documentation is created for each published version of the reference database for all substance classes incorporated.

The reference database is incorporated in the software module *IdentModule*. During identification testing with *Apo-Ident*, spectra which are used for evaluation purposes are presented to it by the *QuickStep* software. In the same manner, the *IdentModule* is presented all validation spectra successively during the validation runs for evaluation purposes. The *IdentModule* responds respectively (without taking the initial assumption into account) with the identified substance or rejects it as unknown. The correctness of this response is checked for each possible initial assumption and counted.

The results are summarised for each substance and reproduced in this document. The core statement of this validation report is that the following criteria must be fulfilled for each database entry, so that *Apo-Ident* offers verification of identity by means of NIR for the relevant substance/substance group:

- The database is exclusively generated from spectra which have been recorded by *HiperScan GmbH* on traceable samples in pharmaceutical quality.
  - The samples are procured via typical pharmacy sources (*DAC III.2.: Bezugsquellennachweis für Rezepturbestandteile [reference source for prescription components]* [7]).
  - A valid manufacturer's certificate exists (content, purity and identity of the batch).
  - The identity was confirmed by a certified test laboratory or *HiperScan GmbH*.
- Each version of the reference database (every update) is validated in-full.
  - Calibration spectra (*Type A*), other spectra recorded under the control of *HiperScan GmbH* (*Type B*), and spectra from the field (*Type C*) are presented to the *IdentModule* for evaluation in three separately evaluated validation runs.
  - Here, no single *false positive* result may arise.
  - Here, the various substance classes are also tested for reciprocal rejection, where this is objectively justified (see *Summary* section).
- In the validation with spectra recorded under the control of *HiperScan GmbH*, spectra of at least one independent sample must be considered, i.e. spectra from a batch whose spectra have not been used for the generation of the database. In addition, the set of *Type A* and *Type B* spectra must originate from at least three different batches.
- Spectra of additional substances may be used for the generation of the database even though they will not be offered for identification by this database. The purpose is the reliable distinction from these substances.

- Any positive result of *Apo-Ident* confirms the identity of the substance/substance group and distinguishes it from all other substances of the database. In the case of substance groups, the result is ambiguous: Distinction from all substances not belonging to the group is proven. The substance is identified as a member of this group. However, within the substance group, it is not possible to reliably classify which substance has been tested.

- The criteria for clear identifiability are a specificity of 100 % (true negative rate) and a minimum distance in the distance matrix. See 2. d) under Model creation procedure and validation runs.

Validation concept

Chemometrics is a statistical technique for the extraction of relevant chemical information from spectra. In mathematics, this method is described as *multivariate data analysis*. Chemometrics proceeds here as follows:

1. Collection of spectra for the *calibration sample*. The results (identities) of the calibration sample must be known. The calibration samples must be representative for the samples which are to be evaluated later. Therefore, they must take the various possible (physical) compositions into account. (Therefore, sourcing calibration samples for NIR from the specialist trade is superior to the use of CRS reference substances.)

2. The first mathematical step is *calibration*. Here, the chemometric model is calculated from the *calibration sample spectra* (reference spectra) and limits as well as some parameters are stipulated. The chemometric model is used later to calculate the analysis result (*prediction*).

3. Collection of further spectra for the *validation sample* which should be independent of the *calibration sample*. The results (identities) of the *validation sample* must also be known. The textbook suggests a random sample with a normal scope of 25 % to 50 % of the *calibration sample* [8].

4. The second technical data step is *validation*. Here, the chemometric model created is evaluated based on the spectra of the *validation samples*. As validation parameters for the identification, the *Ph. Eur. Section 2.2.40* [3] specifies the specificity and robustness.

The validation step according to the textbook has the target of estimating the performance capability of the model created based on a random sample. In order to achieve the best possible precision, attention is paid to the calibration sample. In the field of pharmaceuticals, the safety of the method has priority. In order to be able to *validate* the model within the regulatory scope, the validation step must include probative force. For this purpose, the validation sample must be *representative and complete* in order to enable the testing of all cases.

A *sufficient number of batches* must be secured for validation because validation finally proves whether the number of batches in calibration suffices.

Each substance is validated individually. The validation results are documented per substance in this document. Moreover, the documents show how many and which batches have been used for creating the model or model validation.

At least one certificate is taken in for each substance from an accredited test laboratory for the independent testing of identity of the sample. The identification number of the corresponding test certificate is listed in the report, enabling traceability of a substance tested according to the monographs in the Pharmacopoeia.

Model creation procedure and validation runs

The safety of the chemometric models is guaranteed by several measures during model creation, of which the validation step is the final one. Normally, the procedure is as follows. It is in particular valid for the active pharmaceutical ingredients (APIs) *solid API excipients*, *liquid/semi-solid API excipients (with a test certificate)*, *narcotics - solid medicinal substances* and *drugs*. If, for individual substance classes, variations are required, they are depicted in the section *Particularities of individual substance classes*.

1. Collecting the reference spectra (calibration sample)

a) Procurement of the samples from the same sources from which pharmacies source their raw material for compounding (Caelo, Fagron, Euro-OTC, . . . , see also *DAC III.2. Bezugsquellennachweis für Rezepturbestandteile* [Sources of supply for compounding] [7]).

b) Testing the suitability according to *ApBetrO* [Pharmacies Rules and Regulations] §§ 6, 11, that is to say the availability of a valid manufacturer certificates via identity, purity and contents of the batch.

c) Recording standard 40 spectra of the sample in different positions, as a standard on four instruments. Here, handling and presentation of the samples as later in the pharmacy.

d) Visual checking for anomalies in the spectra. In case of indications of measurement errors, measurement must be repeated. If a signature is missing in the spectrum, the substance may be excluded from the start as not promising (the spectra are nonetheless entered in the database validation as independent *Type B* spectra).

e) Testing identity. For each substance, a certificate of correct identity is obtained from an independent GMP-certified test laboratory. If the identity of the sample can be provided using NIR on an independently tested reference sample, in the respective following substance page of this validation documentation the *Mahalanobis distance* to this reference sample is specified as well as the *Mahalanobis distance* to the next non-identical substance. Such samples underline the statistical spread of the original reference substance, but cannot add any new characteristics of the substance.

   HiperScan GmbH cooperates with some suppliers as follows: the raw materials supplier takes a sufficiently large sample in his incoming goods area so that part of it can be used for recording the NIR spectra. The remainder of the sample goes to analytics for market approval. The manufacturer's batch certificate emerges from these identity, contents and purity tests, which consequently also prove the correct identity of the NIR reference sample. Therefore, the NIR spectra are suitable for structuring the database *(Type A)* and can also optionally be used for validation *(Type B)*. The samples which this applies to are marked in the validation report with a footnote.

f) If the identity of the new sample is proven, it is declared as a reference sample and the spectra are approved for structuring the database.

2. Generating the chemometric models (calibration)

a) Determination of the transformation matrix from the reference spectra using variance maximisation [8, 9]. (*All* reference spectra are always included, even if only a few spectra are added for an update.) All reference spectra receive the same data pre-treatment, which is also later applied to all measurement spectra in the field (in the pharmacies).

b) Checking that the number of principal components used is still sufficient.

c) Calculating the limits for each substance from the spread of the reference spectra. The calculation regulation is identical for each substance in a substance class.

d) Überprüfen der Abstände zwischen den Grenzen der trennbaren Substanzen: Die Distanzmatrix enthält die *Mahalanobis-Abstände* von jeder Substanz zu jeder anderen. Die Werte hin und zurück sind jeweils unterschiedlich, weil die Streuung der Ausgangssubstanz eingeht. Ist eine Distanz kleiner als der Mindestabstand, so gelten die Substanzen als nicht sicher trennbar. Der Mindestabstand ist auf 9 festgelegt. Der Entwickler des Modells darf einen größeren Mindestabstand festlegen (ein Wert für das gesamte *chemometrische Modell*), um die Trennschärfe zu erhöhen.

e) Überprüfung des Modells anhand der Referenzspektren. Es sind keine *falsch-positiven* Ergebnisse erlaubt.

f) Wird eines der Kriterien verletzt (d) *Unterschrittener Mindestabstand zwischen zwei Substanzen* oder (e) *Eine Substanz wird als eine andere identifiziert*, entscheidet der Entwickler der Datenbank, welche der folgenden Optionen er anwendet:

- Er nimmt beide Substanzen aus der Datenbank. (Die Spektren bleiben in der Validierung und dürfen auch in den Aufbau eingehen. Sie werden aber nicht zur Prüfung angeboten.)
- Er bildet eine Substanzgruppe mehrerer nicht sicher trennbarer Substanzen. Dann ist das Ergebnis mehrdeutig: Das chemometrische Modell stellt fest, dass es sich bei der Probe um eine der Substanzen aus der Gruppe handelt und dass es sich um keine andere Substanz handelt. Es kann aber nicht sagen, um welche der Substanzen es sich handelt. Um die eindeutige Identität festzustellen, muss der Anwender eine geeignete ergänzende Prüfung durchführen.
- Er erstellt ein weiteres *chemometrisches Modell* mit geringerem Umfang, in das mindestens alle Substanzen der nicht sicher trennbaren Substanzgruppe eingehen (Zweite-Stufe-Modell). Zweite-Stufe-Modelle werden nur aufgerufen, wenn die erste Stufe festgestellt hat, dass es sich nur um eine der Substanzen handeln kann, die in den Aufbau der Zweiten Stufe eingegangen ist.

3. Set of validation spectra (validation samples)

The following is provided for validation:

a) *Type A*: The reference spectra = calibration spectra from which the database was generated. These also include spectra from substances which the *chemometric model* should not identify, but were also recorded during generation in order to increase selectivity. (As a result, the model "learns" to differentiate from other substances which are actually unknown to it.)

b) *Type B*: Spectra recorded under the control of *HiperScan GmbH* and not used for the generation of the database. These also include reference spectra of other substance classes, and spectra that are not used as reference spectra. Samples are considered to be independent, if they originate from a batch, of which no spectra have been used for the generation of the database. (Up to *IdentModule 2018-01*, samples were still considered independent if the sampling was done independently, i.e. if they originate from another sales container.)

c) *Type C*: Spectra from the field, which have not been recorded under the control of *HiperScan GmbH* and have not been used for the generation of the database. The spectra include not only substances of the substance class to be tested, but also substances from other classes.

All manufacturers' batches from which spectra are used for the validation are listed by substance in this document: for substances included in the substance class *Hepart raw materials, liquid* in the respective validation reports; otherwise in attachments *A*, *B* and *C*.

Furthermore remains valid: validation spectra may only be removed if a error in the spectrum can be proven. Here, the spectra are not deleted, but instead placed on a *blacklist* incorporating the reason, date and initials in the commentary.

The section *Particularities of individual substance classes* treats the other substance classes from which *Type B* and *Type C* spectra are cited for validation purposes.

4. Validation runs and approval

a) Validation spectra are transferred holistically to the *IdentModule* for evaluation in the same way as the spectroscopy software *QuickStep* transfers measured spectra.

b) Following the provision of each spectrum, the *IdentModule* responds as to whether it has recognised a substance and which substance was recognised.

c) The correctness of this response is checked for each possible initial assumption (each measurable substance with the substance class) and counted according to *true negative*, *false negative*, *true positive* and *false positive*. These figures are provided for each substance and additionally in the section *Summary*, separated according to types *A*, *B* and *C* .

d) No *false positive* results whatsoever are permissible.

e) If the criterion is also met for all substance classes, the *IdentModule* is approved.

Particularities of individual substance classes

Basically, *HiperScan GmbH* procures and tests the manufacturer's certificate for the batch, commissions external testing of the identity of the sample or carries it out independently and stores the certificates. As described, this process is established for the Pharmacopoeia substances, that is to say for substance classes **APIs & excipients, solid**, **APIs & excipients, liquid/semi-solid (with analysis certificate)**, **Narcotic substances, liquid/semi-solid** and **Drugs**. Therefore, *HiperScan GmbH* is able to furnish proof of the identity of the reference samples. In case of manufacturer-specific substance classes and others, individual steps are organised differently in-part:

The substance class **APIs & excipients, liquid/semi-solid (other)** (often described as cosmetics) incorporates substances for which no specification of the requirements of the pharmaceutical quality is determined, neither in a Pharmacopoeia monograph, a DAC/NRF monograph nor via a manufacturer's specification. Consequently, neither the identity nor contents can be tested independently. No certificates whatsoever exist for the reference samples. So here, merely the matching of the sample with former samples of this product is established and confusion with the other substances is ruled-out. (If the manufacturer of such a substance prepares a specification, determines testing methods and provides manufacturer's certificates in accordance with *ApBetrO* [Pharmacies Rules and Regulations] §§ 6, 11, *HiperScan GmbH* can assign the substance to the substance class *APIs & excipients, liquid/semi-solid (with analysis certificate)* again in the future).

Substance class **HCK – nutritional supplements (Hepart)** contains the HCK micro-nutrients from the Swiss company *Hepart AG*. *HiperScan GmbH* receives the reference samples directly from the manufacturer. For each reference sample, *HiperScan GmbH* also receives manufacturer's certificates and keeps these. New checking of the identity of the reference sample is not carried out by *HiperScan GmbH*. The identity of the reference samples is therefore documented by *Hepart AG*. The spectra of all batches provided by *Hepart AG* are recorded by *HiperScan GmbH* and entered in the database.

All the manufacturer's batches are used for the generation and validation of the substance class *HCK*. The expected variation is also represented in the generation and validation if there are less than three batches.

Also, for the substance class **PhytoComm** (TCM-Granulated herbal extracts of the manufacturer *PhytoComm*) spectra for all useable batches are recorded by *HiperScan GmbH* and entered in the database. The supplier organises the respective tests themself and keeps the test certificates.

A new evaluation option was created for the class *PhytoComm* with the update 2016-01. As the risks are considerably fewer than those from chemical agents, the pharmacist can specify a reasonable criterion for the *specificity* in accordance with internal risk estimation. The database for this is created without taking safety distances into account and no criterion is determined in advance for the *specificity*. Instead, the *specificity* for testing the identity with this concrete substance is calculated in the validation for each substance and provided with the measurement result. The pharmacist then judges himself whether this safety is reasonable with regard to the risk of the substance.

Additionally, a statistical forecast is provided for the *specificity* which is determined according to the *Rule of Three* [10, 11]. For this forecast, it is assumed that there would have been three wrong results more and is provided with a lower limit for *specificity*. This value has a special meaning if a *specificity* of 100 % is achieved for a substance during validation. In this case, the lower limit allows conclusions regarding the scale of existing safety for which with an endless number of validation spectra a value of less than 100 % is to be assumed.

If, for example 14 000 spectra not belonging to the substance are presented and no *false positive* classification is made, a hypothetical number of three *false positive* results is assumed (*Rule of Three* [10, 11]) and the *specificity* is defined with 100.0000 % (> 99.9786 %). Here, it applies that the higher the number of validation spectra which form the statistical basis, the better the *specificity* calculated via the lower *specificity* limit will be approximated.

The positive result of the identity test using *Apo-Ident* establishes that the sample spectrum is in accordance with a batch of the specified granulate from the supplier *PhytoComm*, whereby all useable batches from the supplier are known.

The *PhytoComm* class can only confirm the identity of batches that have been used for the generation of the database. As a consequence there cannot be any validation spectra of other batches. Therefore, the criterion reads that two samples (from different sales containers) from each batch must exist, one for the structure of the database (*Type A*) and one for the validation (*Type B*).

Significance of testing with *Apo-Ident*

The analysis result is determined using sophisticated statistical methods according to state-of-the-art science and technology. Chemical and pharmaceutical knowledge is applied for the selection of the samples from which the calibration spectra and validation spectra are recorded. Otherwise it does not influence the further steps of model creation.

Verbally, the statement of the analysis result can be expressed as follows. Here *"the spectra match"* means that the criteria *Mahalanobis distance*, *outlier analysis* and *correlation* are met as shown in *Erfüllung von 2.2.40 Ph. Eur. durch Apo-Ident* [4]. "The spectra do not match", on the other hand, means that at least the criterion *Mahalanobis distance* is not met.

The positive analysis result *"was identified as ..."* is very meaningful because both the quantity of substances to be taken into account and the number of underlying samples is very comprehensive.

1. The spectrum of the sample measured matches spectra of the defined substance.
2. The spectrum of the sample measured does not match any spectrum of any other substance in this substance class. Therefore, all other substances can be clearly ruled-out.
3. As the spectra from other substance classes were used for validation, it is proven that no spectrum of one of these other substances matches the defined substance. (All substance classes with which a spectrum comparison is possible and makes sense are used for validation. This is documented for each substance class in the section *Summary*.)
4. If the defined substance belongs to a *substance group* which in itself is not clearly separable with *Apo-Ident*, matching with the spectra of one or several substances in this group is confirmed. Which of these substances it actually is cannot be determined clearly. All other substances are excluded analogous to 2 and 3.

On the other hand, a negative analysis result *"was not identified as ..."* means:

1. The substance offered could not be recognised based on the spectrum of this sample.
2. The identity of this sample is not confirmed.
3. Testing must be repeated in accordance with the specifications of the Pharmacopoeia.

Conclusion

NIR spectroscopy is a testing method incorporated in the Pharmacopoeia. In case of successful database validation, it is a possible method for identity testing [5]. *Apo-Ident* meets the criteria of the *European Pharmacopeia* as a near-infrared spectrometer and proves the validity of the reference database with the existing validation documentation. This means that *Apo-Ident* can be used as an alternative testing method for testing raw materials at pharmacies.

Explanation of terminology

The following section serves to explain or define specialist terminology which is required in order to understand this document. If necessary, definitions for the analysis system *Apo-Ident* are defined more precisely.

The term database is used in this document exactly as in the *Ph. Eur. Section 2.2.40* [3] synonymous with chemometric model. In order to differentiate the databases which are relatively independent of each other, *HiperScan GmbH* frequently also uses the term substance class (primarily in the plural). On the other hand, the spectra used to structure the database are termed spectrum collection and not database.

Substance classes are units of the organisational structure of the *IdentModule*. The substance classes are substance databases which are also broadly independently subscribable. On the one hand, the liquid and semi-solid substances are separated from the solid powders because they are measured against different references and therefore the spectra cannot be compared. On the other hand, for example the Pharmacopoeia substances are kept separated from the manufacturer-specific database *PhytoComm* for TCM (traditional Chinese medicine) raw materials.

The individual substance classes need only be limited against each other in-part. Often, no risk of confusion exists because they can only be procured from different sources. On the other hand, in several cases we handle substances which need not be distinguished. For example, em Huang Qi granulate from the company *PhytoComm* neither needs to be delimited from *Huang Qi* granulate from the company *HerbaSinica* nor is matching required. Respectively one single chemometric model is behind a substance class. (Even if several reciprocally secured chemometric models would be permissible.) The terms *substance class*, *chemometric model* and *databases* are mostly used here as synonyms.

A substance group respectively summarises all the substances within a substance class which cannot safely be distinguished from one another based on their NIR spectra. However, all the other substances in the database can be excluded.

The formation of subgroups is mentioned in the *Ph. Eur. Section 2.2.40* [3]. In this manner, technical restrictions in case of extensive databases can be avoided and it is possible to prepare individual subgroups with different spectrum pre-treatment. Validation of the subgroups against each other is required. *HiperScan GmbH* has solved these technical restrictions and doesn't use any subgroups within a substance class any longer.

Principal component analysis (PCA) [8, 9] is a multivariate statistics process or multivariate data analysis. It serves to structure, simplify and illustrate comprehensive data records by describing a large number of statistical variables by describing a lower number of linear combinations (the *principal components*) which are as significant as possible. In the *Apo-Ident IdentModule*, *PCA* is used to evaluate the recorded spectrum data (corresponding with *Ph. Eur. 2.2.40* [3]).

The term validation is defined in both relevant contexts here with different (even if related) meanings.

Within the sense of the expert discipline of *chemometrics*, validation is a process step when creating a chemometric model: after a transformation matrix, limits and various parameters have been calculated or determined from a set of reference spectra during the course of the calibration step [8, 9], the validation step determines the performance capability of the model (selectivity, precision, ...) based on the validation spectra. Normally, random sampling is planned here. In order for the validation to gain strength of proof, the validation spectrum set must be selected with an appropriately wide scope (*representative* and *complete*). The terms *validation run* and *validation step* always actually mean the process step in this sense.

In the regulatory sense (of pharmaceutical production), validation is the documented proof that a process or system meets the previously specified requirements reproducibly when applied practi-

cally. In this sense, the *Apo-Ident* databases only become validated databases with the validation documentation, which this document is part of.

The *European Pharmacopeia* uses the term validation in *Section 2.2.40* within the sense of the specialist discipline of *chemometrics* [3].

The robustness of a process is the property of only being influenced by environmental fluctuations (e.g. temperature or humidity) a little. A method is robust if the environmental conditions do not or hardly falsify the final result.

The specificity of a classification (of a chemometric model) is the true negative rate.

The recognition rate (also sensitivity) is the true positive rate. It defines in how many percent of cases a correctly set up substance is actually confirmed.

The true negative rate describes the share of spectra correctly classified as non-identity during validation. This is equivalent to correct classification. It means that a substance A within identity checking as substance B is judged as *"not identified"*. The *true negative rate* is equivalent to the conditional frequency

$$h(\text{rejected}|\text{genuinely no identity}) = \frac{r_n}{r_n + f_p}$$

with r_n as the total number of *true negative* classifications and f_p as the total number of *false positive* classifications. For successful validation of an *IdentModule*, all spectra presented belonging to this category must be classified as *not in accordance*.

In order to ensure that each substance is entered with the same weight irrespective of the number of spectra available, each spectrum is weighted with the reciprocal value of this number. The weight of each spectrum of a substance/substance group i therefore results as

$$w_i = \frac{1}{n_i}$$

with n_i number of spectra in this substance/substance group. This weighting ensures that the overall result cannot be enhanced by adding especially large numbers of spectra from easily separable substances.

The true positive rate describes the share of spectra correctly classified as identity during validation. This is equivalent to correct classification. It means that a substance A within identity checking as substance A is judged as *"identified"*. The *true positive rate* is equivalent to the conditional frequency

$$h(\text{identified}|\text{genuine identity}) = \frac{r_p}{r_p + f_n}$$

with r_p as the total number of *true positive* classifications and f_n as the total number of *false negative* classifications. The *true positive* rate is a measure for the recognition rate of the validated *Apo-Ident IdentModule*.
In order to ensure that each substance is received with the same weight, the spectra are weighted as described for the true negative rate.

The true negative result describes a spectrum correctly classified as non-identity during validation. It is equivalent to correct classification. It means that a substance A within identity checking as substance B is judged as *"not identified"*.

The false positive result describes a spectrum falsely classified as non-identity during validation. This is the most critical type of possible false classification. It means that a substance A within identity checking as substance B is judged as *"identified"*. For successful validation of an *IdentModule*, a number of false positive events of zero are demanded for all spectra entering the validation. The exception to this restriction is the class of TCM granulates from the company *PhytoComm* as described under *Particularities of individual substance classes*.

The true positive result describes a spectrum correctly classified as identity during validation. It is equivalent to correct classification. It means that a substance A within identity checking as substance A is judged as *"identified"*.

The false negative result describes a spectrum falsely classified as non-identity during validation. It is equivalent to false classification. It means that a substance A within identity checking as substance A is judged as *"not identified"*.

The 'Rule of Three' says that with a probability of 95 % the next random sample of the same size no more than three false results are to be expected if no false result existed in the existing random sample [10, 11].

The *specificity* and *recognition rate* are determined both globally and from the validation runs for all substances. The information is supplemented with the hypothetical value if there had been three false results more. The percent information is provided in parentheses with the "greater than" symbol '>', e.g. *specificity* 100.000 % (>99.983 %) if 17 567 false spectra have been presented without one single *false positive* result. The larger the statistical basis, the lower the influence of the hypothetical false results.

The Mahalanobis distance is a distance measure between two points in n-dimensional vector space. Here, the respective direction component of the distance to *standard deviation* [12] of an n-dimensional distribution is standardised. In case of the *principal component analysis* [8, 9] this standardisation relates to the distribution of the respective calibration data set for a classification (substance/substance group) in the *principal component space* [8]. The *Mahalanobis distance* of a point (mapping of a spectrum) $\vec{y}$ in the n-dimensional principal component space to the expected value of an n-dimensional distribution $\mathbf{X}$ then results as

$$d(\mathbf{X}, \vec{y}) = \sqrt{(\bar{\mathbf{X}} - \vec{y})^T \mathbf{S}^{-1} (\bar{\mathbf{X}} - \vec{y})} \qquad \text{with} \qquad \mathbf{X} \in \mathbb{R}^{m \times n}, \vec{y} \in \mathbb{R}^m$$

[13]. Here, m is equivalent to the number of principal components used (dimension of the principal component space) and n the number of measurements existing in the calibration data set (spectra). $\bar{\mathbf{X}}$ is the expected value of the resulting distribution for the calibration data set (the average value of n measurements received). $\mathbf{S}^{-1}$ is the inverse covariance matrix [12] for distribution $\mathbf{X}$.

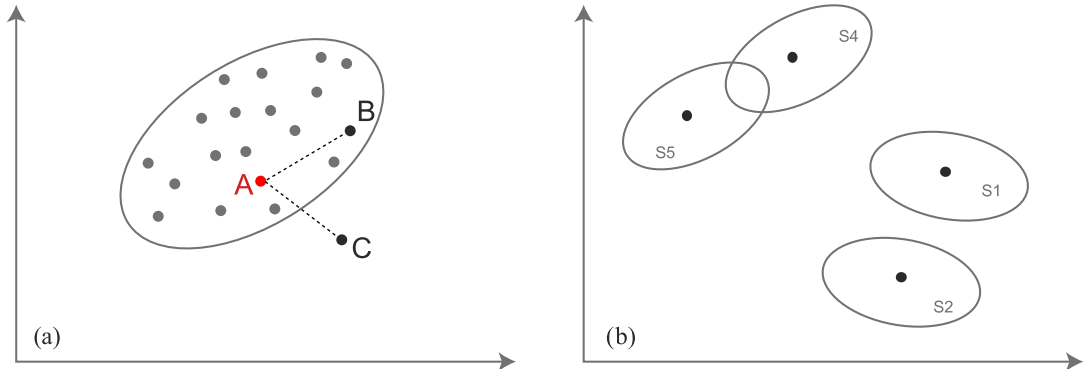

Figure 1: (a) The *Mahalanobis distance* from A to B is less than from A to C. However, the Euclidean distances are identical. (b) The *Mahalanobis distance* between the two measurement series *S4* and *S5* is smaller than between *S1* and *S2*. However, the Euclidean distances are identical.

The *Mahalanobis distance* offers advantages compared to the *Euclidean distance*: For the calculation of the distance it takes the statistical properties of a data point distribution (measurement series), i.e. average value, variance and covariance of the data points [14] into account. The *Mahalanobis distance* is applied while creating the reference database for evaluating the spectra from different samples of a substance.

A chemometric model is a classifier based on statistical methods [8, 9]. Through the respective algorithm used (e.g. *Principal Component Analysis*, *Cluster Analysis*), a maximum of chemical information is extracted from measurement data. Here, systematical or physical disturbances are eliminated using appropriate data pre-processing [15, 16].

At several places in this document, in order to simplify understanding, the term database is used instead of *chemometric model* – in the same manner as in the *Ph. Eur. Section 2.2.40* [3].

A sample (with its own sample ID) refers to substance in a sales container. Repeated sampling from the same sales container is listed under the same sample ID. (The suffix "SI" is not part of the sample ID.) Several samples may originate from the same batch. Samples are called "independent", if they originate from a batch, of which no spectra have been used for the generation of the database. (Up to *IdentModule 2018-01*, samples from different sales packages were considered to be independent.) The information above the list of validation spectra now includes also the number of batches that deliver independent samples for the validation (for both *Type B* and *Type C*).

In case a supplier takes a sample for testing from its incoming goods and splits it to multiple laboratory containers, the substance in all laboratory containers will still be ascribed to the same sample. *HiperScan GmbH* only uses one of the subsamples.

Reference samples are used to structure the database. The *reference spectra* originate from these samples. In chemometric technical jargon you would normally say: For *calibration*, a *chemometric model* is generated from the *calibration spectra* recorded from the *calibration samples*, whose quality is subsequently assessed in validation.

Reference samples are procured via typical pharmacy sources. Their identity is tested. The *reference spectra* are recorded by *HiperScan GmbH*. The documentation also includes the manufacturer's name and batch number.

Reference samples are clearly identified by a sample ID. Samples without sample ID may not be used as *reference samples*.

Summary

A total of 1880 spectra from 44 different batches for a total of 18 substances were used to validate the substance class *Hepart raw materials, liquid*.

Validation samples

The validation samples can be categorised as follows:

Typ A Calibration spectra. These are the spectra used to generate the chemometric model. They were recorded by *HiperScan GmbH*. Detailed information regarding the batches or samples can be found in the following validation reports under *calibration samples* and under *Type A*. Further information is listed in *Appendix A*.

Substance class	Substances	Batches	Spectra
Hepart raw materials, liquid	13	21	960

From category *A* a total of 960 spectra from 21 batches for a total of 13 substances were taken into account for validation.

Typ B Spectra from independent samples which are not included in database generation. These spectra were recorded by *HiperScan GmbH*. Detailed information regarding the batches or samples can be found in the following validation reports in the section *Type B* or in *Appendix B*.

Substance class	Substances	Batches	Spectra
Hepart raw materials, liquid	17	23	920

From category *B* a total of 920 spectra from 23 batches for a total of 17 substances were taken into account for validation.

Typ C Spectra from independent samples which are not included in database generation. *Apo-Ident* customers carried out the measurements. Detailed information regarding the batches or samples can be found in the following validation reports in the section *Type C* or in *Appendix C*.

Substance class	Substances	Batches	Spectra

From category *C* a total of 0 spectra from 0 batches for a total of 0 substances were taken into account for validation.

Validation results

The validation runs checked whether all substances/substance groups in the substance class *Hepart raw materials, liquid* can clearly be distinguished from all other substances using NIR spectroscopy with *Apo-Ident*. For this purpose, the matching of all relevant spectra of various substances with the substances/substance groups in the substance class *Hepart raw materials, liquid* was checked and the correctness of the results was evaluated. The following table breaks down the numbers of correct and incorrect results according to the expected result (*positive*/*negative*) and the validation spectrum type (*A*/*B*/*C*).

	False positive	True positive	False negative	True negative
Type A	0	960	0	3840
Type B	0	656	4	3785
Type C	0	0	0	0

All substances/substance groups in the substance class *Hepart raw materials, liquid* can be clearly distinguished from all other substances using NIR spectroscopy with *Apo-Ident*. In order to make these figures comparable, the weighted *true negative rate* (*specificity*) and the weighted *true positive rate* (*recognition rate*) are determined:

	Specificity	Recognition rate
Typ A	100.000 00 % (> 99.693 17 %)	100.000 00 % (> 98.772 70 %)
Typ B	100.000 00 % (> 99.684 18 %)	99.736 84 % (> 98.656 25 %)
Typ C	n/a (n/a)	n/a (n/a)

In order to ensure that each substance is entered with the same weight irrespective of the number of spectra available, each spectrum is weighted with the reciprocal value of its number. (If several new spectra of a substance with a very distinctive spectrum are added, the values for *specificity* and *recognition rate* are safer, but are by no means idealised.)

In order to illustrate the impact of the number of spectra, a comparison is made in parentheses as to how the *specificity* or *recognition rate* would worsen if three additional incorrect results were to exist amongst the spectra (*Rule of Three* [10, 11]). The larger the number of spectra, the lower the deterioration if the three hypothetical *incorrect results* were added.
Where the number of spectra does not exceed 20, no *recognition rate* is provided.

VALIDATION REPORT

IdentModule 2.3-2021-01

Validated substance/substance group	**Aloe vera gel 1:1 (R0110)**
Substance class	Hepart raw materials, liquid
Report date	22/03/2021
Report number	85030-2021-03-22
Executing company	HiperScan GmbH Weißeritzstraße 3 01067 Dresden Germany

Relevant substance names

Aloe vera gel 1:1 (R0110)

Special notes

When selecting the *Aloe vera gel 1:1 (R0110)* substance/substance group, the following information is displayed to the user:

no information

Applicable documents

978-3-7692-7517-9	*European Pharmacopoeia 10th Edition, Basic Version 2020* [3]
Komm2.2.40	*Erfüllung von* Ph. Eur. 2.2.40 *durch* Apo-Ident [4]
AA004	*Erstellung und Validierung eines IdentModul-Updates*

Validation method

Validation is performed after every change to the chemometric model (also "database") in three steps:

1. The chemometric model is calculated from the calibration spectra using a *PCA algorithm*. The calibration spectra originate from the calibration samples of all substances in this class.
2. In the *chemometric model* generated, the distances between all separable substances are checked for compliance with the specified safety distances.
3. All spectra are presented to the generated chemometric model for evaluation. In three runs, the reference spectra (*Type A*), spectra from independent samples (*Type B*), and spectra from the field (*Type C*) are presented successively. Here, no single *false positive* result is permitted.

Finally, a report is generated from the validation run results and is archived together with the parameters of the model generation in audit-compliant manner.

Number of independent samples (batches) in calibration and validation

A sample is considered independent if no sample from the same batch has been included in the calibration of the chemometric model.

Substance	Type A	Type B	Type C
Aloe vera gel 1:1 (R0110)	1	0	0

Second-stage model

For differentiation of the substance/substance group *Aloe vera gel 1:1 (R0110)* the following second-stage model is used:

no second-stage model

Calibration samples

Only spectra which have been recorded by *HiperScan GmbH* from traceable samples are used for the generation of the chemometric models. The following samples have been obtained from the substance/substance group *Aloe vera gel 1:1 (R0110)*:

Supplier	Substance	Batch	Sample ID	Spectra	Certificate
Hepart	Aloe vera gel 1:...	12210902	85090	40	not required

Validation samples

A total of 1880 spectra were provided for validation. The results were evaluated separately according to the following sample categories:

Type A All calibration spectra.

- 40 spectra of 1 reference samples from the substance/substance group *Aloe vera gel 1:1 (R0110)*. These samples are listed above in the *calibration samples* section. The reference samples come from 1 different batches.

- 920 spectra from a total of 20 batches from further 12 substances. These spectra were recorded by *HiperScan GmbH*. *Type A* samples are documented within this entire report. If the substance can be identified with this chemometric model, the samples are listed in the *Type A* section of each substance page. If the substance cannot be identified by this model, the samples are listed in *Appendix A*. The samples in *Appendix A* were added for model generation not with the intention to have the model identify these substances, but with the intention to make the model reject them despite their similarity to substances identifiable.

Type B Spectra from independent samples not included in the database structure. Measurement is carried out by *HiperScan GmbH*.

Samples from batches of which no spectra have been included in the database structure are considered as independent samples. The number of batches from which independent samples supplied *Type B* spectra for validation is shown below, representing the number of independent *Type B* samples. Samples, of which some of the spectra have been included in the database structure and other spectra in the validation, are marked with a †. The following applies to the remaining unmarked samples: In the same batch, there was at least one additional sample (other sales container, other sample ID) from which reference spectra (*Type A*) were included in the database structure.

- 0 spectra of 0 reference samples from the substance/substance group *Aloe vera gel 1:1 (R0110)*.

- Among them are spectra of independent samples from 0 batches from which no spectra have been included in the database structure. They are sorted upwards in the following table and separated from the additional samples by a line.

- 920 spectra from a total of 23 batches from further 17 substances. These spectra were recorded by *HiperScan GmbH*. *Type B* samples are documented within this entire report. If the substance can be identified with this chemometric model, the samples are listed in the *Type B* section of each substance page. If the substance cannot be identified by this model, the samples are listed in *Appendix B*.

Type C Spectra of independent samples not included in the database structure. *Apo-Ident* customers carry out the measurements. *HiperScan GmbH* generally does not verify the information provided by the customer regarding manufacturer and batch number.

- 0 spectra from 0 *Apo-Ident* customers from 0 batches from the substance/substance group *Aloe vera gel 1:1 (R0110)*.
- Among them are spectra of independent samples from 0 batches from which no spectra have been included in the database structure. These are sorted upwards in the following table.
- 0 spectra from 0 *Apo-Ident* customers from a total of 0 batches from a further 0 substances. These spectra were recorded by *Apo-Ident* customers. *Type C* samples are documented within this entire report. If the substance can be identified with this chemometric model, the samples are listed in the *Type C* section of each substance page. If the substance cannot be identified by this model, the samples are listed in *Appendix C*.

Validation results

The validation runs checked whether the substance/substance group *Aloe vera gel 1:1 (R0110)* can clearly be distinguished from all other substances using NIR spectroscopy with *Apo-Ident*. For this purpose, all relevant spectra of the various substances were compared with *Aloe vera gel 1:1 (R0110)* and it was evaluated how many matches (positive) and rejections (negative) were correct or incorrect. The following table breaks down the numbers of correct and incorrect results according to the expected result (*positive*/*negative*) and the validation spectrum type (*A*/*B*/*C*).

	False positive	True positive	False negative	True negative
Type A	0	40	0	920
Type B	0	0	0	920
Type C	0	0	0	0

The substance/substance group *Aloe vera gel 1:1 (R0110)* can be clearly distinguished from all other substances. In order to make these figures comparable, the weighted *true negative rate* (*specificity*) and the weighted *true positive rate* (*recognition rate*) are determined:

	Specificity	Recognition rate
Type A	100.0000 % (> 99.0198 %)	100.0000 % (> 85.0000 %)
Type B	100.0000 % (> 98.6472 %)	n/a (n/a)
Type C	n/a (n/a)	n/a (n/a)

In order to ensure that each substance is entered with the same weight irrespective of the number of spectra available, each spectrum is weighted with the reciprocal value of its number. (If several new spectra of a substance with a very distinctive spectrum are added, the values for *specificity* and *recognition rate* are safer, but are by no means idealised.)

In order to illustrate the impact of the number of spectra, a comparison is made in parentheses as to how the *specificity* or *recognition rate* would worsen if three additional incorrect results were to exist amongst the spectra (*Rule of Three* [10, 11]). The larger the number of spectra, the lower the deterioration if the three hypothetical *incorrect results* were added.
Where the number of spectra does not exceed 20, no *recognition rate* is provided.

Nearest chemometric neighbours

The following table lists the closest chemometric neighbours of the substance/substance group *Aloe vera gel 1:1 (R0110)* in the substance class *Hepart raw materials, liquid*. Furthermore, their *Mahalanobis distance* is specified within the main model and, where appropriate, within the second-stage model.

Substance	Distance in main model	Distance in second-stage model
Cabbage juice concentrate (ca. 60 brix…	20.08	–
Base cream DAC (R0109)	23.90	–

continued on the next page

continued from previous page

Substanz	Distance in main model	Distance in second-stage model
Acerola concentrated juice (R0266)	24.12	–
Cranberry concentrated juice (R0265)	25.92	–
Beetroot concentrated juice (R0253)	27.31	–
Seabuckthorn concentrated juice (R0256)	28.11	–
Chokeberry concentrated juice (R0267)	41.99	–
Blueberry concentrated juice (R0270)	51.11	–

The list stops after the first substance with a *Mahalanobis distance* greater than 50. If the substance/substance group *Aloe vera gel 1:1 (R0110)* is separated from critical neighbours in a second-stage model, all the remaining substances in the second-stage model follow.

Identity according to identity test using NIR spectroscopy

For each substance, a certificate of the correct identity is obtained from an independent GMP-certified test laboratory. If the identity of the sample can be provided using NIR on an independently tested reference sample, in the following table the *Mahalanobis distance* to the reference sample is specified as well as the *Mahalanobis distance* to the next non-identical substance:

Sample ID	Reference sample ID	Distance to reference sample	Distance to next foreign sample
85090	85090	0.00	20.08

The identity of a sample will be confirmed by NIR if the distance to the next foreign sample is at least 50 % greater than the distance to a reference sample whose identity has been established by laboratory testing. This criterion is always considered in the chemometric model, which contains all substances of the substance class, even if a second-stage dissolves a subgroup of similar substances, thereby increasing the distances between them. The samples confirmed by NIR support the statistical variance of the original reference substance, but cannot add new aspects or forms of the substance.

VALIDATION REPORT

IdentModule 2.3-2021-01

Validated substance/substance group	**Base cream DAC (R0109)**
Substance class	Hepart raw materials, liquid
Report date	22/03/2021
Report number	85029-2021-03-22
Executing company	HiperScan GmbH Weißeritzstraße 3 01067 Dresden Germany

Relevant substance names

Base cream DAC (R0109)

Special notes

When selecting the *Base cream DAC (R0109)* substance/substance group, the following information is displayed to the user:

no information

Applicable documents

978-3-7692-7517-9	*European Pharmacopoeia 10th Edition, Basic Version 2020* [3]
Komm2.2.40	*Erfüllung von* Ph. Eur. 2.2.40 *durch* Apo-Ident [4]
AA004	*Erstellung und Validierung eines IdentModul-Updates*

Validation method

Validation is performed after every change to the chemometric model (also "database") in three steps:

1. The chemometric model is calculated from the calibration spectra using a *PCA algorithm*. The calibration spectra originate from the calibration samples of all substances in this class.
2. In the *chemometric model* generated, the distances between all separable substances are checked for compliance with the specified safety distances.
3. All spectra are presented to the generated chemometric model for evaluation. In three runs, the reference spectra (*Type A*), spectra from independent samples (*Type B*), and spectra from the field (*Type C*) are presented successively. Here, no single *false positive* result is permitted.

Finally, a report is generated from the validation run results and is archived together with the parameters of the model generation in audit-compliant manner.

Number of independent samples (batches) in calibration and validation

A sample is considered independent if no sample from the same batch has been included in the calibration of the chemometric model.

Substance	Type A	Type B	Type C
Base cream DAC (R0109)	2	1	0

Second-stage model

For differentiation of the substance/substance group *Base cream DAC (R0109)* the following second-stage model is used:

no second-stage model

Calibration samples

Only spectra which have been recorded by *HiperScan GmbH* from traceable samples are used for the generation of the chemometric models. The following samples have been obtained from the substance/substance group *Base cream DAC (R0109)*:

Supplier	Substance	Batch	Sample ID	Spectra	Certificate
Hepart	Base cream DAC (...	12303901	85089	40	not required
Hepart	Base cream DAC (...	15001286	85315	60	not required

Validation samples

A total of 1880 spectra were provided for validation. The results were evaluated separately according to the following sample categories:

Type A All calibration spectra.

- 100 spectra of 2 reference samples from the substance/substance group *Base cream DAC (R0109)*. These samples are listed above in the *calibration samples* section. The reference samples come from 2 different batches.
- 860 spectra from a total of 19 batches from further 12 substances. These spectra were recorded by *HiperScan GmbH*. *Type A* samples are documented within this entire report. If the substance can be identified with this chemometric model, the samples are listed in the *Type A* section of each substance page. If the substance cannot be identified by this model, the samples are listed in *Appendix A*. The samples in *Appendix A* were added for model generation not with the intention to have the model identify these substances, but with the intention to make the model reject them despite their similarity to substances identifiable.

Type B Spectra from independent samples not included in the database structure. Measurement is carried out by *HiperScan GmbH*.

Samples from batches of which no spectra have been included in the database structure are considered as independent samples. The number of batches from which independent samples supplied *Type B* spectra for validation is shown below, representing the number of independent *Type B* samples. Samples, of which some of the spectra have been included in the database structure and other spectra in the validation, are marked with a †. The following applies to the remaining unmarked samples: In the same batch, there was at least one additional sample (other sales container, other sample ID) from which reference spectra (*Type A*) were included in the database structure.

- 40 spectra of 1 reference samples from the substance/substance group *Base cream DAC (R0109)*.
- Among them are spectra of independent samples from 1 batches from which no spectra have been included in the database structure. They are sorted upwards in the following table and separated from the additional samples by a line.

Supplier	Substance	Batch	Sample ID	Spectra
Hepart	Base cream DAC (R0109)	12245201	85088	40

- 880 spectra from a total of 22 batches from further 16 substances. These spectra were recorded by *HiperScan GmbH*. *Type B* samples are documented within this entire report. If the substance can be identified with this chemometric model, the samples are listed in the *Type B* section of each substance page. If the substance cannot be identified by this model, the samples are listed in *Appendix B*.

Type C Spectra of independent samples not included in the database structure. *Apo-Ident* customers carry out the measurements. *HiperScan GmbH* generally does not verify the information provided by the customer regarding manufacturer and batch number.

- 0 spectra from 0 *Apo-Ident* customers from 0 batches from the substance/substance group *Base cream DAC (R0109)*.
- Among them are spectra of independent samples from 0 batches from which no spectra have been included in the database structure. These are sorted upwards in the following table.
- 0 spectra from 0 *Apo-Ident* customers from a total of 0 batches from a further 0 substances. These spectra were recorded by *Apo-Ident* customers. *Type C* samples are documented within this entire report. If the substance can be identified with this chemometric model, the samples are listed in the *Type C* section of each substance page. If the substance cannot be identified by this model, the samples are listed in *Appendix C*.

Validation results

The validation runs checked whether the substance/substance group *Base cream DAC (R0109)* can clearly be distinguished from all other substances using NIR spectroscopy with *Apo-Ident*. For this purpose, all relevant spectra of the various substances were compared with *Base cream DAC (R0109)* and it was evaluated how many matches (positive) and rejections (negative) were correct or incorrect. The following table breaks down the numbers of correct and incorrect results according to the expected result (*positive*/*negative*) and the validation spectrum type (*A*/*B*/*C*).

	False positive	True positive	False negative	True negative
Type A	0	100	0	860
Type B	0	40	0	880
Type C	0	0	0	0

The substance/substance group *Base cream DAC (R0109)* can be clearly distinguished from all other substances. In order to make these figures comparable, the weighted *true negative rate* (*specificity*) and the weighted *true positive rate* (*recognition rate*) are determined:

	Specificity	Recognition rate
Type A	100.0000 % (> 98.4573 %)	100.0000 % (> 94.0000 %)
Type B	100.0000 % (> 98.5222 %)	100.0000 % (> 85.0000 %)
Type C	n/a (n/a)	n/a (n/a)

In order to ensure that each substance is entered with the same weight irrespective of the number of spectra available, each spectrum is weighted with the reciprocal value of its number. (If several new spectra of a substance with a very distinctive spectrum are added, the values for *specificity* and *recognition rate* are safer, but are by no means idealised.)

In order to illustrate the impact of the number of spectra, a comparison is made in parentheses as to how the *specificity* or *recognition rate* would worsen if three additional incorrect results were to exist amongst the spectra (*Rule of Three* [10, 11]). The larger the number of spectra, the lower the deterioration if the three hypothetical *incorrect results* were added.
Where the number of spectra does not exceed 20, no *recognition rate* is provided.

Nearest chemometric neighbours

The following table lists the closest chemometric neighbours of the substance/substance group *Base cream DAC (R0109)* in the substance class *Hepart raw materials, liquid*. Furthermore, their *Mahalanobis distance* is specified within the main model and, where appropriate, within the second-stage model.

Substance	Distance in main model	Distance in second-stage model
Aloe vera gel 1:1 (R0110)	101.22	–

The list stops after the first substance with a *Mahalanobis distance* greater than 50. If the substance/substance group *Base cream DAC (R0109)* is separated from critical neighbours in a second-stage model, all the remaining substances in the second-stage model follow.

Identity according to identity test using NIR spectroscopy

For each substance, a certificate of the correct identity is obtained from an independent GMP-certified test laboratory. If the identity of the sample can be provided using NIR on an independently tested reference sample, in the following table the *Mahalanobis distance* to the reference sample is specified as well as the *Mahalanobis distance* to the next non-identical substance:

Sample ID	Reference sample ID	Distance to reference sample	Distance to next foreign sample
85089	85089	0.00	123.68
85315	85315	0.00	101.22

The identity of a sample will be confirmed by NIR if the distance to the next foreign sample is at least 50 % greater than the distance to a reference sample whose identity has been established by laboratory testing. This criterion is always considered in the chemometric model, which contains all substances of the substance class, even if a second-stage dissolves a subgroup of similar substances, thereby increasing the distances between them. The samples confirmed by NIR support the statistical variance of the original reference substance, but cannot add new aspects or forms of the substance.

VALIDATION REPORT

IdentModule 2.3-2021-01

Validated substance/substance group	**Concentrated juices**
Substance class	Hepart raw materials, liquid
Report date	22/03/2021
Report number	85325-2021-03-22
Executing company	HiperScan GmbH Weißeritzstraße 3 01067 Dresden Germany

Relevant substance names

Concentrated juices; Acerola concentrated juice (R0266); Beetroot concentrated juice (R0253); Blueberry concentrated juice (R0270); Broccoli concentrated juice (R0268); Cabbage juice concentrate (ca. 60 brix) (R0229); Chokeberry concentrated juice (R0267); Cranberry concentrated juice (R0265); Pomegranate concentrated juice (R0262); Seabuckthorn concentrated juice (R0256)

Special notes

When selecting the *Concentrated juices* substance/substance group, the following information is displayed to the user:

no information

Applicable documents

978-3-7692-7517-9	*European Pharmacopoeia 10th Edition, Basic Version 2020* [3]
Komm2.2.40	*Erfüllung von* Ph. Eur. 2.2.40 *durch* Apo-Ident [4]
AA004	*Erstellung und Validierung eines IdentModul-Updates*

Validation method

Validation is performed after every change to the chemometric model (also "database") in three steps:

1. The chemometric model is calculated from the calibration spectra using a *PCA algorithm*. The calibration spectra originate from the calibration samples of all substances in this class.
2. In the *chemometric model* generated, the distances between all separable substances are checked for compliance with the specified safety distances.
3. All spectra are presented to the generated chemometric model for evaluation. In three runs, the reference spectra (*Type A*), spectra from independent samples (*Type B*), and spectra from the field (*Type C*) are presented successively. Here, no single *false positive* result is permitted.

Finally, a report is generated from the validation run results and is archived together with the parameters of the model generation in audit-compliant manner.

Number of independent samples (batches) in calibration and validation

A sample is considered independent if no sample from the same batch has been included in the calibration of the chemometric model.

Substance	Type A	Type B	Type C
Acerola concentrated juice (R0266)	1	1	0
Beetroot concentrated juice (R0253)	2	1	0
Blueberry concentrated juice (R0270)	1	1	0

continued on the next page

continued from previous page

Substance	Type A	Type B	Type C
Broccoli concentrated juice (R0268)	1	1	0
Cabbage juice concentrate (ca. 60 brix) (R0229)	3	1	0
Chokeberry concentrated juice (R0267)	1	1	0
Cranberry concentrated juice (R0265)	1	1	0
Pomegranate concentrated juice (R0262)	1	1	0
Seabuckthorn concentrated juice (R0256)	1	1	0

Second-stage model

For differentiation of the substance/substance group *Concentrated juices* the following second-stage model is used:

no second-stage model

Calibration samples

Only spectra which have been recorded by *HiperScan GmbH* from traceable samples are used for the generation of the chemometric models. The following samples have been obtained from the substance/substance group *Concentrated juices*:

Supplier	Substance	Batch	Sample ID	Spectra	Certificate
Hepart	Acerola concentr...	16001342	85562	40	not required
Hepart	Beetroot concent...	18001279	85565	40	not required
Hepart	Beetroot concent...	19002649/0	85833	40	not required
Hepart	Blueberry concen...	18001277	85560	40	not required
Hepart	Broccoli concent...	18001856/0	85592	40	not required
Hepart	Cabbage juice co...	15001471	85325	60	not required
Hepart	Cabbage juice co...	15001472	85328	60	not required
Hepart	Cabbage juice co...	16000417	85359	60	not required
Hepart	Chokeberry conce...	18001276	85567	40	not required
Hepart	Cranberry concen...	18001236	85564	40	not required
Hepart	Pomegranate conc...	18001274	85568	40	not required
Hepart	Seabuckthorn con...	18001235	85559	40	not required

Validation samples

A total of 1880 spectra were provided for validation. The results were evaluated separately according to the following sample categories:

Type A All calibration spectra.

- 540 spectra of 12 reference samples from the substance/substance group *Concentrated juices*. These samples are listed above in the *calibration samples* section. The reference samples come from 12 different batches.

- 420 spectra from a total of 9 batches from further 4 substances. These spectra were recorded by *HiperScan GmbH*. *Type A* samples are documented within this entire report. If the substance can be identified with this chemometric model, the samples are listed in the *Type A* section of each substance page. If the substance cannot be identified by this model, the samples are listed in *Appendix A*. The samples in *Appendix A* were added for model generation not with the intention to have the model identify these substances, but with the intention to make the model reject them despite their similarity to substances identifiable.

Type B Spectra from independent samples not included in the database structure. Measurement is carried out by *HiperScan GmbH*.

Samples from batches of which no spectra have been included in the database structure are considered as independent samples. The number of batches from which independent samples supplied *Type B* spectra for validation is shown below, representing the number of independent *Type B* samples. Samples, of which some of the spectra have been included in the database structure and other spectra in the validation, are marked with a †. The following applies to the remaining unmarked samples: In the same batch, there was at least one additional sample (other sales container, other sample ID) from which reference spectra (*Type A*) were included in the database structure.

- 380 spectra of 9 reference samples from the substance/substance group *Concentrated juices*.
- Among them are spectra of independent samples from 9 batches from which no spectra have been included in the database structure. They are sorted upwards in the following table and separated from the additional samples by a line.

Supplier	Substance	Batch	Sample ID	Spectra
Hepart	Acerola concentrated juice (...	18001295	85563	40
Hepart	Beetroot concentrated juice ...	20001112/0	85853	40
Hepart	Blueberry concentrated juice...	20001117	85852	40
Hepart	Broccoli concentrated juice ...	20000991/0	85847	40
Hepart	Cabbage juice concentrate (c...	16000474	85360	60
Hepart	Chokeberry concentrated juic...	20001116/0	85849	40
Hepart	Cranberry concentrated juice...	20000099/0	85850	40
Hepart	Pomegranate concentrated jui...	20000114	85864	40
Hepart	Seabuckthorn concentrated ju...	19001388	85606	40

- 540 spectra from a total of 14 batches from further 8 substances. These spectra were recorded by *HiperScan GmbH*. *Type B* samples are documented within this entire report. If the substance can be identified with this chemometric model, the samples are listed in the *Type B* section of each substance page. If the substance cannot be identified by this model, the samples are listed in *Appendix B*.

Type C Spectra of independent samples not included in the database structure. *Apo-Ident* customers carry out the measurements. *HiperScan GmbH* generally does not verify the information provided by the customer regarding manufacturer and batch number.

- 0 spectra from 0 *Apo-Ident* customers from 0 batches from the substance/substance group *Concentrated juices*.
- Among them are spectra of independent samples from 0 batches from which no spectra have been included in the database structure. These are sorted upwards in the following table.
- 0 spectra from 0 *Apo-Ident* customers from a total of 0 batches from a further 0 substances. These spectra were recorded by *Apo-Ident* customers. *Type C* samples are documented within this entire report. If the substance can be identified with this chemometric model, the samples are listed in the *Type C* section of each substance page. If the substance cannot be identified by this model, the samples are listed in *Appendix C*.

Validation results

The validation runs checked whether the substance/substance group *Concentrated juices* can clearly be distinguished from all other substances using NIR spectroscopy with *Apo-Ident*. For this purpose, all relevant spectra of the various substances were compared with *Concentrated juices* and it was evaluated how many matches (positive) and rejections (negative) were correct or incorrect. The following table breaks down the numbers of correct and incorrect results according to the expected result (*positive*/*negative*) and the validation spectrum type (*A*/*B*/*C*).

	False positive	True positive	False negative	True negative
Type A	0	540	0	420
Type B	0	376	4	385
Type C	0	0	0	0

The substance/substance group *Concentrated juices* can be clearly distinguished from all other substances. In order to make these figures comparable, the weighted *true negative rate* (*specificity*) and the weighted *true positive rate* (*recognition rate*) are determined:

	Specificity	Recognition rate
Type A	100.0000 % (> 98.1518 %)	100.0000 % (> 98.8889 %)
Type B	100.0000 % (> 98.0500 %)	98.9474 % (> 98.1579 %)
Type C	n/a (n/a)	n/a (n/a)

In order to ensure that each substance is entered with the same weight irrespective of the number of spectra available, each spectrum is weighted with the reciprocal value of its number. (If several new spectra of a substance with a very distinctive spectrum are added, the values for *specificity* and *recognition rate* are safer, but are by no means idealised.)

In order to illustrate the impact of the number of spectra, a comparison is made in parentheses as to how the *specificity* or *recognition rate* would worsen if three additional incorrect results were to exist amongst the spectra (*Rule of Three* [10, 11]). The larger the number of spectra, the lower the deterioration if the three hypothetical *incorrect results* were added.
Where the number of spectra does not exceed 20, no *recognition rate* is provided.

Nearest chemometric neighbours

The following table lists the closest chemometric neighbours of the substance/substance group *Concentrated juices* in the substance class *Hepart raw materials, liquid*. Furthermore, their *Mahalanobis distance* is specified within the main model and, where appropriate, within the second-stage model.

Substance	Distance in main model	Distance in second-stage model
Aloe vera gel 1:1 (R0110)	27.84	–
Base cream DAC (R0109)	30.28	–

The list stops after the first substance with a *Mahalanobis distance* greater than 50. If the substance/substance group *Concentrated juices* is separated from critical neighbours in a second-stage model, all the remaining substances in the second-stage model follow.

Identity according to identity test using NIR spectroscopy

For each substance, a certificate of the correct identity is obtained from an independent GMP-certified test laboratory. If the identity of the sample can be provided using NIR on an independently tested reference sample, in the following table the *Mahalanobis distance* to the reference sample is specified as well as the *Mahalanobis distance* to the next non-identical substance:

Sample ID	Reference sample ID	Distance to reference sample	Distance to next foreign sample
85833	85833	0.00	68.18
85325	85325	0.00	36.36
85565	85565	0.00	49.06
85567	85567	0.00	46.24
85562	85562	0.00	29.10
85359	85359	0.00	41.77
85592	85592	0.00	41.73
85564	85564	0.00	28.06

continued on the next page

continued from previous page

Sample ID	Reference sample ID	Distance to reference sample	Distance to next foreign sample
85560	85560	0.00	38.85
85559	85559	0.00	27.84
85328	85328	0.00	40.35
85568	85568	0.00	42.61

The identity of a sample will be confirmed by NIR if the distance to the next foreign sample is at least 50 % greater than the distance to a reference sample whose identity has been established by laboratory testing. This criterion is always considered in the chemometric model, which contains all substances of the substance class, even if a second-stage dissolves a subgroup of similar substances, thereby increasing the distances between them. The samples confirmed by NIR support the statistical variance of the original reference substance, but cannot add new aspects or forms of the substance.

VALIDATION REPORT

IdentModule 2.3-2021-01

Validated substance/substance group	**D-alpha tocopheryl acetate (R0082)**
Substance class	Hepart raw materials, liquid
Report date	22/03/2021
Report number	85031-2021-03-22
Executing company	HiperScan GmbH Weißeritzstraße 3 01067 Dresden Germany

Relevant substance names

D-alpha tocopheryl acetate (R0082)

Special notes

When selecting the D-*alpha tocopheryl acetate (R0082)* substance/substance group, the following information is displayed to the user:

no information

Applicable documents

978-3-7692-7517-9	*European Pharmacopoeia 10th Edition, Basic Version 2020* [3]
Komm2.2.40	*Erfüllung von* Ph. Eur. 2.2.40 *durch* Apo-Ident [4]
AA004	*Erstellung und Validierung eines IdentModul-Updates*

Validation method

Validation is performed after every change to the chemometric model (also "database") in three steps:

1. The chemometric model is calculated from the calibration spectra using a *PCA algorithm*. The calibration spectra originate from the calibration samples of all substances in this class.
2. In the *chemometric model* generated, the distances between all separable substances are checked for compliance with the specified safety distances.
3. All spectra are presented to the generated chemometric model for evaluation. In three runs, the reference spectra (*Type A*), spectra from independent samples (*Type B*), and spectra from the field (*Type C*) are presented successively. Here, no single *false positive* result is permitted.

Finally, a report is generated from the validation run results and is archived together with the parameters of the model generation in audit-compliant manner.

Number of independent samples (batches) in calibration and validation

A sample is considered independent if no sample from the same batch has been included in the calibration of the chemometric model.

Substance	Type A	Type B	Type C
D-alpha tocopheryl acetate (R0082)	3	1	0

Second-stage model

For differentiation of the substance/substance group D-*alpha tocopheryl acetate (R0082)* the following second-stage model is used:

no second-stage model

Calibration samples

Only spectra which have been recorded by *HiperScan GmbH* from traceable samples are used for the generation of the chemometric models. The following samples have been obtained from the substance/substance group D-*alpha tocopheryl acetate (R0082)*:

Supplier	Substance	Batch	Sample ID	Spectra	Certificate
Hepart	D-alpha tocopher…	1209024	85060	40	not required
Hepart	D-alpha tocopher…	16000191	85361	60	not required
Hepart	D-alpha tocopher…	17001256	85508	40	not required

Validation samples

A total of 1880 spectra were provided for validation. The results were evaluated separately according to the following sample categories:

Type A All calibration spectra.

- 140 spectra of 3 reference samples from the substance/substance group D-*alpha tocopheryl acetate (R0082)*. These samples are listed above in the *calibration samples* section. The reference samples come from 3 different batches.

- 820 spectra from a total of 18 batches from further 12 substances. These spectra were recorded by *HiperScan GmbH*. *Type A* samples are documented within this entire report. If the substance can be identified with this chemometric model, the samples are listed in the *Type A* section of each substance page. If the substance cannot be identified by this model, the samples are listed in *Appendix A*. The samples in *Appendix A* were added for model generation not with the intention to have the model identify these substances, but with the intention to make the model reject them despite their similarity to substances identifiable.

Type B Spectra from independent samples not included in the database structure. Measurement is carried out by *HiperScan GmbH*.

Samples from batches of which no spectra have been included in the database structure are considered as independent samples. The number of batches from which independent samples supplied *Type B* spectra for validation is shown below, representing the number of independent *Type B* samples. Samples, of which some of the spectra have been included in the database structure and other spectra in the validation, are marked with a †. The following applies to the remaining unmarked samples: In the same batch, there was at least one additional sample (other sales container, other sample ID) from which reference spectra (*Type A*) were included in the database structure.

- 40 spectra of 1 reference samples from the substance/substance group D-*alpha tocopheryl - acetate (R0082)*.

- Among them are spectra of independent samples from 1 batches from which no spectra have been included in the database structure. They are sorted upwards in the following table and separated from the additional samples by a line.

Supplier	Substance	Batch	Sample ID	Spectra
Hepart	D-alpha tocopheryl acetate (…	17001692	85509	40

- 880 spectra from a total of 22 batches from further 16 substances. These spectra were recorded by *HiperScan GmbH*. *Type B* samples are documented within this entire report. If the substance can be identified with this chemometric model, the samples are listed in the *Type B* section of each substance page. If the substance cannot be identified by this model, the samples are listed in *Appendix B*.

Type C Spectra of independent samples not included in the database structure. *Apo-Ident* customers carry out the measurements. *HiperScan GmbH* generally does not verify the information provided by the customer regarding manufacturer and batch number.

- 0 spectra from 0 *Apo-Ident* customers from 0 batches from the substance/substance group D-*alpha tocopheryl acetate (R0082)*.

- Among them are spectra of independent samples from 0 batches from which no spectra have been included in the database structure. These are sorted upwards in the following table.

- 0 spectra from 0 *Apo-Ident* customers from a total of 0 batches from a further 0 substances. These spectra were recorded by *Apo-Ident* customers. *Type C* samples are documented within this entire report. If the substance can be identified with this chemometric model, the samples are listed in the *Type C* section of each substance page. If the substance cannot be identified by this model, the samples are listed in *Appendix C*.

Validation results

The validation runs checked whether the substance/substance group D-*alpha tocopheryl acetate (R0082)* can clearly be distinguished from all other substances using NIR spectroscopy with *Apo-Ident*. For this purpose, all relevant spectra of the various substances were compared with D-*alpha tocopheryl acetate (R0082)* and it was evaluated how many matches (positive) and rejections (negative) were correct or incorrect. The following table breaks down the numbers of correct and incorrect results according to the expected result (*positive*/*negative*) and the validation spectrum type (*A*/*B*/*C*).

	False positive	True positive	False negative	True negative
Type A	0	140	0	820
Type B	0	40	0	880
Type C	0	0	0	0

The substance/substance group D-*alpha tocopheryl acetate (R0082)* can be clearly distinguished from all other substances. In order to make these figures comparable, the weighted *true negative rate* (*specificity*) and the weighted *true positive rate* (*recognition rate*) are determined:

	Specificity	Recognition rate
Type A	100.0000 % (> 98.3502 %)	100.0000 % (> 95.7143 %)
Type B	100.0000 % (> 98.5222 %)	100.0000 % (> 85.0000 %)
Type C	n/a (n/a)	n/a (n/a)

In order to ensure that each substance is entered with the same weight irrespective of the number of spectra available, each spectrum is weighted with the reciprocal value of its number. (If several new spectra of a substance with a very distinctive spectrum are added, the values for *specificity* and *recognition rate* are safer, but are by no means idealised.)

In order to illustrate the impact of the number of spectra, a comparison is made in parentheses as to how the *specificity* or *recognition rate* would worsen if three additional incorrect results were to exist amongst the spectra (*Rule of Three* [10, 11]). The larger the number of spectra, the lower the deterioration if the three hypothetical *incorrect results* were added.
Where the number of spectra does not exceed 20, no *recognition rate* is provided.

Nearest chemometric neighbours

The following table lists the closest chemometric neighbours of the substance/substance group D-*alpha tocopheryl acetate (R0082)* in the substance class *Hepart raw materials, liquid*. Furthermore, their *Mahalanobis distance* is specified within the main model and, where appropriate, within the second-stage model.

Substance	Distance in main model	Distance in second-stage model
D-tocopherols (70%G mix) (R0083)	42.15	–

The list stops after the first substance with a *Mahalanobis distance* greater than 50. If the substance/substance group D-*alpha tocopheryl acetate (R0082)* is separated from critical neighbours in a second-stage model, all the remaining substances in the second-stage model follow.

Identity according to identity test using NIR spectroscopy

For each substance, a certificate of the correct identity is obtained from an independent GMP-certified test laboratory. If the identity of the sample can be provided using NIR on an independently tested reference sample, in the following table the *Mahalanobis distance* to the reference sample is specified as well as the *Mahalanobis distance* to the next non-identical substance:

Sample ID	Reference sample ID	Distance to reference sample	Distance to next foreign sample
85060	85060	0.00	44.75
85508	85508	0.00	42.79
85361	85361	0.00	42.15

The identity of a sample will be confirmed by NIR if the distance to the next foreign sample is at least 50 % greater than the distance to a reference sample whose identity has been established by laboratory testing. This criterion is always considered in the chemometric model, which contains all substances of the substance class, even if a second-stage dissolves a subgroup of similar substances, thereby increasing the distances between them. The samples confirmed by NIR support the statistical variance of the original reference substance, but cannot add new aspects or forms of the substance.

VALIDATION REPORT

IdentModule 2.3-2021-01

Validated substance/substance group	**D-tocopherols (70%G mix) (R0083)**
Substance class	Hepart raw materials, liquid
Report date	22/03/2021
Report number	85058-2021-03-22
Executing company	HiperScan GmbH Weißeritzstraße 3 01067 Dresden Germany

Relevant substance names

D-tocopherols (70%G mix) (R0083)

Special notes

When selecting the *D-tocopherols (70%G mix) (R0083)* substance/substance group, the following information is displayed to the user:

no information

Applicable documents

978-3-7692-7517-9	*European Pharmacopoeia 10th Edition, Basic Version 2020* [3]
Komm2.2.40	*Erfüllung von* Ph. Eur. 2.2.40 *durch* Apo-Ident [4]
AA004	*Erstellung und Validierung eines IdentModul-Updates*

Validation method

Validation is performed after every change to the chemometric model (also "database") in three steps:

1. The chemometric model is calculated from the calibration spectra using a *PCA algorithm*. The calibration spectra originate from the calibration samples of all substances in this class.
2. In the *chemometric model* generated, the distances between all separable substances are checked for compliance with the specified safety distances.
3. All spectra are presented to the generated chemometric model for evaluation. In three runs, the reference spectra (*Type A*), spectra from independent samples (*Type B*), and spectra from the field (*Type C*) are presented successively. Here, no single *false positive* result is permitted.

Finally, a report is generated from the validation run results and is archived together with the parameters of the model generation in audit-compliant manner.

Number of independent samples (batches) in calibration and validation

A sample is considered independent if no sample from the same batch has been included in the calibration of the chemometric model.

Substance	Type A	Type B	Type C
D-tocopherols (70%G mix) (R0083)	3	5	0

Second-stage model

For differentiation of the substance/substance group D-*tocopherols (70%G mix) (R0083)* the following second-stage model is used:

no second-stage model

Calibration samples

Only spectra which have been recorded by *HiperScan GmbH* from traceable samples are used for the generation of the chemometric models. The following samples have been obtained from the substance/substance group D-*tocopherols (70%G mix) (R0083)*:

Supplier	Substance	Batch	Sample ID	Spectra	Certificate
Hepart	D-tocopherols (7…	1209023	85058	40	not required
Hepart	D-tocopherols (7…	130926500	85152	40	not required
Hepart	D-tocopherols (7…	15000951	85316	60	not required

Validation samples

A total of 1880 spectra were provided for validation. The results were evaluated separately according to the following sample categories:

Type A All calibration spectra.

- 140 spectra of 3 reference samples from the substance/substance group D-*tocopherols (70%G mix) (R0083)*. These samples are listed above in the *calibration samples* section. The reference samples come from 3 different batches.
- 820 spectra from a total of 18 batches from further 12 substances. These spectra were recorded by *HiperScan GmbH*. *Type A* samples are documented within this entire report. If the substance can be identified with this chemometric model, the samples are listed in the *Type A* section of each substance page. If the substance cannot be identified by this model, the samples are listed in *Appendix A*. The samples in *Appendix A* were added for model generation not with the intention to have the model identify these substances, but with the intention to make the model reject them despite their similarity to substances identifiable.

Type B Spectra from independent samples not included in the database structure. Measurement is carried out by *HiperScan GmbH*.

Samples from batches of which no spectra have been included in the database structure are considered as independent samples. The number of batches from which independent samples supplied *Type B* spectra for validation is shown below, representing the number of independent *Type B* samples. Samples, of which some of the spectra have been included in the database structure and other spectra in the validation, are marked with a †. The following applies to the remaining unmarked samples: In the same batch, there was at least one additional sample (other sales container, other sample ID) from which reference spectra (*Type A*) were included in the database structure.

- 200 spectra of 5 reference samples from the substance/substance group D-*tocopherols (70%G mix) (R0083)*.
- Among them are spectra of independent samples from 5 batches from which no spectra have been included in the database structure. They are sorted upwards in the following table and separated from the additional samples by a line.

Supplier	Substance	Batch	Sample ID	Spectra
Hepart	D-tocopherols (70%G mix) (R0…	131111502	85153	40

continued on the next page

continued from previous page

Supplier	Substance	Batch	Sample ID	Spectra
Hepart	D-tocopherols (70%G mix) (R0…	16001165	85431	40
Hepart	D-tocopherols (70%G mix) (R0…	18001105	85557	40
Hepart	D-tocopherols (70%G mix) (R0…	19001604	85607	40
Hepart	D-tocopherols (70%G mix) (R0…	20000258	85840	40

- 720 spectra from a total of 18 batches from further 16 substances. These spectra were recorded by *HiperScan GmbH*. *Type B* samples are documented within this entire report. If the substance can be identified with this chemometric model, the samples are listed in the *Type B* section of each substance page. If the substance cannot be identified by this model, the samples are listed in *Appendix B*.

Type C Spectra of independent samples not included in the database structure. *Apo-Ident* customers carry out the measurements. *HiperScan GmbH* generally does not verify the information provided by the customer regarding manufacturer and batch number.

- 0 spectra from 0 *Apo-Ident* customers from 0 batches from the substance/substance group D-*tocopherols (70%G mix) (R0083)*.
- Among them are spectra of independent samples from 0 batches from which no spectra have been included in the database structure. These are sorted upwards in the following table.
- 0 spectra from 0 *Apo-Ident* customers from a total of 0 batches from a further 0 substances. These spectra were recorded by *Apo-Ident* customers. *Type C* samples are documented within this entire report. If the substance can be identified with this chemometric model, the samples are listed in the *Type C* section of each substance page. If the substance cannot be identified by this model, the samples are listed in *Appendix C*.

Validation results

The validation runs checked whether the substance/substance group D-*tocopherols (70%G mix) (R0083)* can clearly be distinguished from all other substances using NIR spectroscopy with *Apo-Ident*. For this purpose, all relevant spectra of the various substances were compared with D-*tocopherols (70%G mix) (R0083)* and it was evaluated how many matches (positive) and rejections (negative) were correct or incorrect. The following table breaks down the numbers of correct and incorrect results according to the expected result (*positive*/*negative*) and the validation spectrum type (*A*/*B*/*C*).

	False positive	True positive	False negative	True negative
Type A	0	140	0	820
Type B	0	200	0	720
Type C	0	0	0	0

The substance/substance group D-*tocopherols (70%G mix) (R0083)* can be clearly distinguished from all other substances. In order to make these figures comparable, the weighted *true negative rate* (*specificity*) and the weighted *true positive rate* (*recognition rate*) are determined:

	Specificity	Recognition rate
Type A	100.0000 % (> 98.3502 %)	100.0000 % (> 95.7143 %)
Type B	100.0000 % (> 98.3347 %)	100.0000 % (> 97.0000 %)
Type C	n/a (n/a)	n/a (n/a)

In order to ensure that each substance is entered with the same weight irrespective of the number of spectra available, each spectrum is weighted with the reciprocal value of its number. (If several new spectra of a substance with a very distinctive spectrum are added, the values for *specificity* and *recognition rate* are safer, but are by no means idealised.)

In order to illustrate the impact of the number of spectra, a comparison is made in parentheses as to how the *specificity* or *recognition rate* would worsen if three additional incorrect results were to exist amongst the spectra (*Rule of Three* [10, 11]). The larger the number of spectra, the lower the deterioration if the three hypothetical *incorrect results* were added.
Where the number of spectra does not exceed 20, no *recognition rate* is provided.

Nearest chemometric neighbours

The following table lists the closest chemometric neighbours of the substance/substance group D-*tocopherols (70%G mix) (R0083)* in the substance class *Hepart raw materials, liquid*. Furthermore, their *Mahalanobis distance* is specified within the main model and, where appropriate, within the second-stage model.

Substance	Distance in main model	Distance in second-stage model
D-alpha tocopheryl acetate (R0082)	79.10	–

The list stops after the first substance with a *Mahalanobis distance* greater than 50. If the substance/substance group D-*tocopherols (70%G mix) (R0083)* is separated from critical neighbours in a second-stage model, all the remaining substances in the second-stage model follow.

Identity according to identity test using NIR spectroscopy

For each substance, a certificate of the correct identity is obtained from an independent GMP-certified test laboratory. If the identity of the sample can be provided using NIR on an independently tested reference sample, in the following table the *Mahalanobis distance* to the reference sample is specified as well as the *Mahalanobis distance* to the next non-identical substance:

Sample ID	Reference sample ID	Distance to reference sample	Distance to next foreign sample
85058	85058	0.00	79.10
85152	85152	0.00	81.23
85316	85316	0.00	84.54

The identity of a sample will be confirmed by NIR if the distance to the next foreign sample is at least 50 % greater than the distance to a reference sample whose identity has been established by laboratory testing. This criterion is always considered in the chemometric model, which contains all substances of the substance class, even if a second-stage dissolves a subgroup of similar substances, thereby increasing the distances between them. The samples confirmed by NIR support the statistical variance of the original reference substance, but cannot add new aspects or forms of the substance.

Appendix A: Additional calibration samples (*Type A*)

Not required.

Appendix B: Additional validation samples (*Type B*)

It is necessary that spectra also enter the validation which cannot be identified with this model. In this manner, it is verified that the model also rejects unknown substances. The spectra for these samples were recorded by *HiperScan GmbH*. They are allocated to the *Type B*. They also include the calibration spectra for other models.

The samples originate from 7 batches. From these 260 spectra were recorded. The spectra recorded on independent samples of substances which can be identified with the model are listed respectively in the section *Type B* for the individual substances and do not appear again elsewhere in this list.

Supplier	Substance	Batch	Spectra	Certificate
Hepart	Beta carotene 20% (R0009)	13000006	20	not required
Hepart	Carrot concentrated juice (R0252)	18001273	40	not required
Hepart	Raspberry concentrated juice (R0271)	18001278	40	not required
Hepart	Raspberry concentrated juice (R0271)	20001118	40	not required
Hepart	Retinyl acetate (R0072)	077901	40	not required
Hepart	Rosehip concentrated juice (R0269)	18001275	40	not required
Hepart	Rosehip concentrated juice (R0269)	20001115	40	not required

Appendix C: Additional validation samples (*Type C*)

Not required.

Appendix D: Requirements of validation

In order to ensure adherence to the safe scientific status, the individual methods for manufacturing and testing must be validated under certain circumstances (compare § 34 para. 1 no. 3, § 35 para. 1 no. 4 and para. 4 sentence 1 no. 2 b, para. 6 sentence 3 *ApoBetrO* [Pharmacies Rules and Regulations]). The *ApoBetrO* [Pharmacies Rules and Regulations] incorporates a legal definition in § 1 a para. 16 (quotation translated):

> *"Validation is the provision of documented proof which with a high degree of safety documents that, via a specific process or standard work process, a medicinal product is manufactured and tested, which is in accordance with previously determined quality features."*

Validation documentation can be used to prove that methods or devices which are not described in the Pharmacopoeia within the meaning of § 6 para. 1 sentence 3 *ApBetrO* [Pharmacies Rules and Regulations] achieve the same results as those in the Pharmacopoeia. On the other hand, with the requirements of the demanded validation it must be observed whether the respective testing method is already incorporated in the Pharmacopoeia.

NIR spectroscopy as a general testing method need not be validated in accordance with the express ruling in the *Ph. Eur. Section 1.1* [3], as it is already described in *Section 2.2.40* of the *Ph. Eur.* as an area of application for the identification of raw materials.

However, a special validation requirement exists for the reference database. This requirement is met with the existing document. Further requirements or rules as to how this proof must be furnished do not exist. It is required that the processes guarantee the same results as the methods and devices in the Pharmacopoeia [17].

Carrying out identity tests with *Apo-Ident* is therefore also possible if the NIR spectroscopy process is not required in the Pharmacopoeia monograph of the substance for identity testing. All NIR analyses with *Apo-Ident* prove several, often all molecule groups and are therefore comparable with a series of individual, targeted chemical proofs [4]. Therefore, the identity proof with *Apo-Ident* replaces the monograph test series (with two or more test combinations).

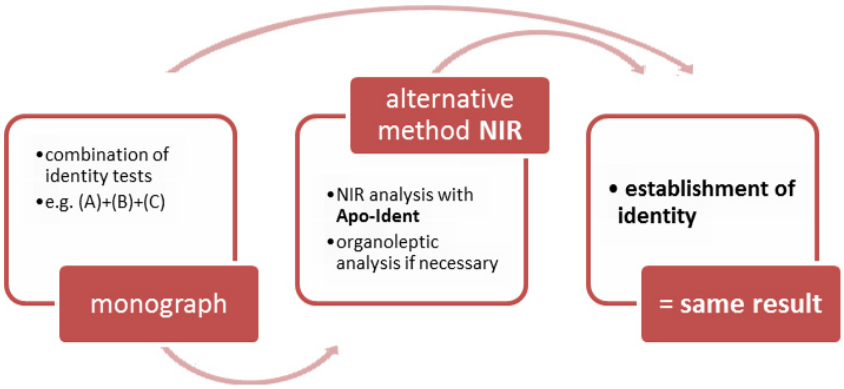

Figure 2: The combination of tests of the monograph is replaced by the alternative method NIR spectroscopy using *Apo-Ident*. This is permissible because both test procedures result in the establishment of the identity of the raw material.

With this validation documentation, proof is furnished that identical results are achieved with *Apo-Ident* and Pharmacopoeia methods, i.e. confirmation of the identity of the raw material [2].

Appendix E: Conformity of *Apo-Ident* with the *European pharmacopeia*

According to *Ph. Eur. Section 2.2.40*, NIR spectroscopy is basically suitable for: "Identification of agents, excipients, dosage forms, intermediate manufacturing products, chemical raw materials and packaging materials" ([3], quotation translated).

The fact that *Apo-Ident* meets the further criteria of the European Pharmacopoeia under the headings in *Section 2.2.40*

- Apparatus
- Measurement methods
- Sample preparation and presentation
- Testing the functionality of the instrument
- Identification and characterisation (qualitative analysis)
- Quantitative analysis
- Ongoing model evaluation
- Transfer of databases
- Data storage

can be proven based on the *HiperScan GmbH* documentation of "Meeting *2.2.40 Ph. Eur.* by *Apo-Ident*" [4].

References

[1] ABDA – Bundesvereinigung Deutscher Apothekerverbände: Verordnung über den Betrieb von Apotheken (Apothekenbetriebsordnung – ApBetrO), 2012

[2] Reimann, B. ; Regierungspräsidium Darmstadt: Hinweise zur ordnungsgemäßen Prüfung von Arzneimitteln und Ausgangsstoffen (§§ 6 und 11 *ApBetrO*), 2007

[3] *Europäisches Arzneibuch, Grundwerk 2020*. 10. Ausgabe. Deutscher Apotheker Verlag (978-3-7692-7517-9)

[4] HiperScan GmbH: Erfüllung von 2.2.40 Ph. Eur. durch Apo-Ident, 2013

[5] Arbeitsgemeinschaft der Pharmazieräte Deutschlands (APD): Resolution in Verbindung mit §§ 6 und 11 *ApBetrO* Verwendung eines Nah-Infrarot-Spektrometers (NIR) zur Identitätsprüfung, 16. 10. 2013, DAZ 48, November 2013

[6] Arbeitsgemeinschaft der Pharmazieräte Deutschlands (APD): Resolution 2014, Arbeitsgemeinschaft der Pharmazieräte Deutschlands (APD), Oktober 2014

[7] *DAC/NRF*. Govi-Verlag (978-3-7741-0044-2)

[8] Kessler, W.: *Multivariate Datenanalyse*. WILEY-VCH Verlag, 2007 (978-3-527-31262-7)

[9] Næs, T. ; Isaksson, T. ; Fearn, T. ; Davies, T.: *Multivariate Calibration and Classification*. NIR Publications, 2002 (978 0 9528666 2 6)

[10] Hanley, J. A. ; Lippman-Hand, A.: If nothing goes wrong, is everything all right? In: *Journal of the American Medical Association* 249 (1983), S. 1743–1745

[11] Jovanovic, B. D. ; Levy, P. S.: A Look at the Rule of Three. In: *The American Statistician* 51 (1997), S. 137–139

[12] Bronstein, I. N. ; Semendjajew, K. A. ; Musiol, G. ; Mühlig, H.: *Taschenbuch der Mathematik*. 5. überarbeitete und erweiterte Auflage. Verlag Harri Deutsch, 2000 (3-8171-2015-2)

[13] Mahalanobis, P.: On the generalized distance in statistics. In: *Proc. Nat. Inst. Sci. India (Calcutta)* 2 (1936), S. 49–55

[14] Yambor, B. ; Draper, W. ; Beveridge, R.: Analyzing PCA-based face recognition algorithms: Eigenvector selection and distance measures. In: *Second Workshop Empirical Evaluation in Computer Vision* (2000)

[15] HiperScan GmbH: Identifikationsmethodik Apo-Ident, 2012

[16] HiperScan GmbH: Datenvorbehandlung des Identifikationssystems Apo-Ident, 2012

[17] Cyran ; Rotta: Apothekenbetriebsordnung, Kommentar § 6 Rn. 10, 2010

Index