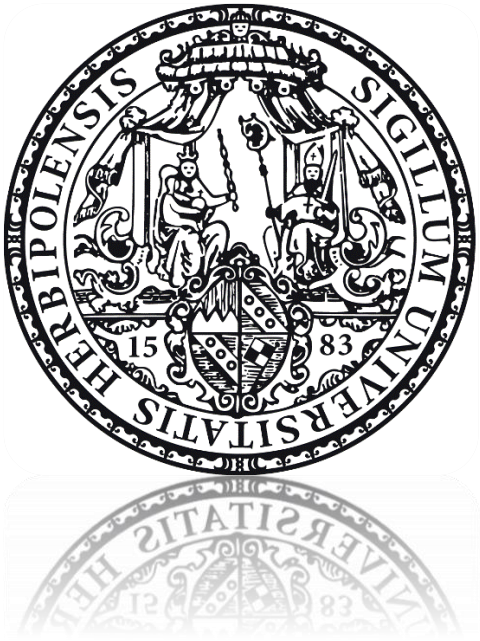
New applications for spectroscopic and chemometric studies of drugs

Dissertation zur Erlangung des naturwissenschaftlichen
Doktorgrades der Julius-Maximilians-Universität Würzburg

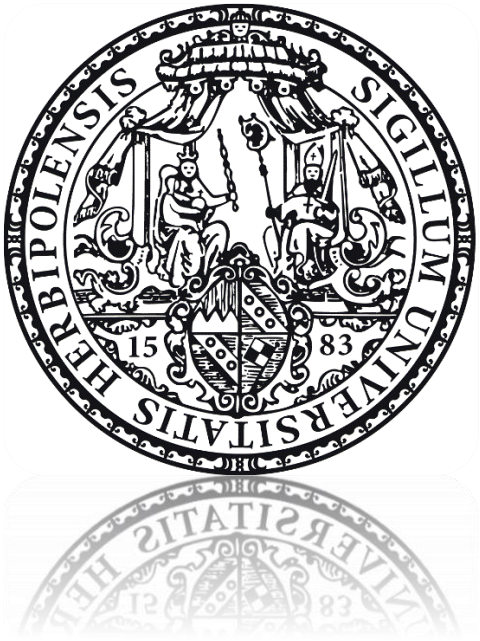


vorgelegt von
Alexander Ulrich Becht
aus
Heidelberg

Würzburg 2022

New applications for spectroscopic and chemometric studies of drugs

Dissertation zur Erlangung des naturwissenschaftlichen
Doktorgrades der Julius-Maximilians-Universität Würzburg



vorgelegt von
Alexander Ulrich Becht
aus
Heidelberg

Würzburg 2022

Eingereicht bei der Fakultät für Chemie und Pharmazie am

_____

Gutachter der schriftlichen Arbeit

1. Gutachter: _____

2. Gutachter: _____

Prüfer des öffentlichen Promotionskolloquiums

1. Prüfer: _____

2. Prüfer: _____

3. Prüfer: _____

Datum des öffentlichen Promotionskolloquiums

_____

Doktorurkunde ausgehändigt am

_____

“Nicht das Beginnen wird belohnt, sondern einzig und allein das Durchhalten.”

Katharina von Siena

~ Meiner Familie ~

Danksagung

Die hier vorliegende Arbeit wurde
am Institut für Pharmazie und Lebensmittelchemie
der Bayerischen Julius-Maximilians-Universität Würzburg angefertigt.

Sie entstand auf Anregung und unter der Anleitung von

Frau Prof. Dr. Ulrike Holzgrabe.

Bei Ihr möchte ich mich recht herzlich für die freundliche Aufnahme in den Arbeitskreis, die interessanten und abwechslungsreichen Projekte, das stets in mich gesetzte Vertrauen sowie die Möglichkeit selbstständig und eigenverantwortlich arbeiten zu können bedanken. Dabei konnte ich immer auf Ihre Unterstützung und fachkundigen Anregungen zählen.

Besonderer Dank gilt all meinen Kolleginnen und Kollegen des AK Holzgrabe, die die Promotionszeit zu so einer unvergesslichen gemacht haben.

Adrian, Anja, Anna, Antonio, Christiane, Christine E., Christine H., Daniela, Flo, Jens, Jogi, Jonas U., Jonas W., Joshi, Klaus, Laura, Liling, Lina, Lu, Lukas, Michi, Miri, Mohamed, Nelson, Niclas, Nicolas, Nils, Patrick K., Patrick N., Paul, Rasmus, Regina, Ruben, Sebastian S., Sylvia, Theresa L., sowie auch Frau Ebner, Frau Möhler und Frau Weidinger; Christoph, Karl, Matthias und Herr Walter.

Ein großer Dank gilt deswegen auch meinen ehemaligen Büro-Kollegen und den weiteren Kolleginnen und Kollegen der klinischen Pharmazie:

Bettina, Charlotte, Flo L., Jasmin, Linda, Markus Z., Max, Nina, Sebbi, Simon und Theresa H.

Weiterhin gilt ein besonderer Dank Prof. Dr. Yulia Monakhova, die mir stets bei Fragen zur Chemometrie zur Seite stand und viel Licht ins Dunkle gebracht hat, sowie Dr. Curd Schollmayer, der mir bei Fragen rund um das Thema NMR-Spektroskopie immer bereitwillig geholfen hat.

Das letzte und größte Dankeschön geht an meine Familie und Freunde, die mich während der Promotionszeit und darüber hinaus immer unterstützt und ermutigt haben und somit auch einen großen Anteil an der Fertigstellung dieser Arbeit haben!

Table of contents

Lists of abbreviations

API	Active pharmaceutical ingredient
CMO	Contract manufacturing organization
DOSY	Diffusion-ordered spectroscopy
FID	Free induction decay
FT-IR	Fourier-transformation infrared
HCA	Hierarchical cluster analysis
HCT	Hydrochlorothiazide
HPLC	High-performance liquid chromatography
IS	Internal standard
LDA	Linear discriminant analysis
MSC	Multiplicative scatter correction
MVA	Multivariate data analysis
NIPALS	Nonlinear iterative partial least squares
NIR	Near-infrared
NMR	Nuclear magnetic resonance
PAT	Process analytical technology
PCA	Principal component analysis
PGSE	Pulse gradient spin echo
qNMR	Quantitative NMR
S/N	Signal-to-noise
SNV	Standard normal variate
SVD	Singular value decomposition
TMS	Tetramethylsilane
TSP	3-(Trimethylsilyl)propionic acid sodium salt

List of variables

A_R	Pseudo absorption
B_0	Magnetic field
D	Diffusion coefficient
E	Residual matrix
F	Number of principal components
G	Gradient intensity
I	Spin quantum number
I_G	Signal intensity when using a gradient during a DOSY measurement
I_0	Signal intensity without a gradient during a DOSY measurement
I_r	Reflected radiation if the reflectance standard
I_s	Reflected radiation of the sample
m	Magnetic quantum number
M	Number of variables
M_0	Macroscopic magnetization
N	Number of possible spin orientations
n	Number of samples
P^T	Loadings matrix
r	Radius of a spherical particle
R	reflectance
T	Absolute Temperature
T_1	Longitudinal relaxation
X	Mean-centered data-matrix
γ	Gyromagnetic ratio
δ	Gradient duration

$\Delta t'$	Corrected diffusion time
η	Dynamic viscosity
$\underline{K_B}$	Boltzmann constant

1. Introduction

1.1 NMR spectroscopy

1.1.1 ^{1}H NMR spectroscopy

NMR spectroscopy examines the behavior of atomic nuclei, which are exposed to a homogenous, static magnetic field and excited to resonate by means of electromagnetic radiation. The first experiments that could detect the resonance signals date back to the late 1940s [1]. Since then, NMR spectroscopy has evolved significantly. This introduction aims to discuss the basic aspects and some special measurement techniques. However, the field of NMR spectroscopy offers many more possibilities and applications in the most diverse scientific fields.

Atomic nuclei, which are accessible to NMR spectroscopy, must have an odd number of protons or neutrons [2]. This results in a spin around its own nuclear axis, creating a magnetic moment. The spin of the atomic nuclei is also called angular momentum [3]. The magnetic moment and the angular momentum are proportional to each other and can be related by means of a proportionality factor, the gyromagnetic ratio γ. The gyromagnetic ratio is a natural constant that is different for each isotope and has a strong influence on the sensitivity of the NMR experiment [1].

When a nucleus is placed in the magnetic field, the nuclear spins align along the magnetic field B_0 [4]. The number of possible orientations N of the spins depends on the spin quantum number I and is equal to $2I + 1$. In the case of hydrogen with $I = ½$, this results in two different orientation possibilities and thus two energy levels characterized by the magnetic quantum number m. These energy levels are also referred to as the Zeeman levels. As can be seen in Figure 1, $m_1 = + ½$ is parallel (α state) and $m_2 = - ½$ is antiparallel (β state) to B_0 [1].

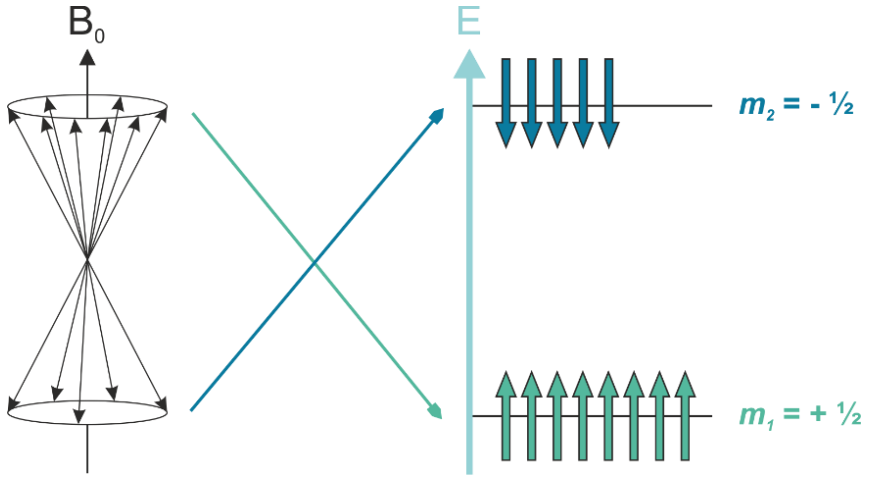


Figure 1: The energy difference of the two Zeeman levels results in the energetically more favorable α state (m_1; parallel to B_0) being slightly more occupied than the β state (m_2; antiparallel to B_0).

These orientations result in the *α* state being more energetically favorable compared to the *β* state, resulting in a slight occupation difference. The difference between these energy levels depends on B_0 and is responsible for the sensitivity of NMR measurements [5]. This, along with increasing signal resolution, is the reason for the drive towards more and more powerful magnets.

In the classical conception, the nuclei precess in a circular path statistically distributed on a double precession cone around the magnetic field. The frequency of this precession motion is called Larmor frequency. The macroscopic magnetization M_0 resulting from the occupation difference shows along B_0, because more nuclei are in the lower energy level. If the Larmor frequency of the corresponding nucleus is irradiated, resonance occurs. In this case, the nuclei in the *α* state absorb the energy and change to the higher energy state. This results in a compensation of the occupation difference and transverse magnetization occurs due to phase coherence of some spins [1].

When the radio frequency is allowed to act on the nuclei for a certain time, the macroscopic magnetization flips into the *x,y*-plane. When the radio frequency is turned off, the nuclei begin to relax and the initial magnetization M_0 in the direction of B_0 is restored. The resulting decrease in transverse magnetization can be measured using a transmitter coil located in the *x,y*-plane, which also serves as receiver coil. The so-called free induction decay (FID) is obtained. To obtain the usual NMR spectrum, the signal must be converted from the time domain to the frequency domain using the Fourier transformation [5]. Since the resonance frequency of a nucleus is influenced by its chemical environment, it is dependent on the molecular structure. The extent of the influence of the chemical environment on the resonance frequency is measured by the chemical shift [6].

1.1.2 Diffusion ordered spectroscopy

Diffusion ordered spectroscopy (DOSY) is a pseudo-2D experiment which combines the information of the self-diffusion coefficient of molecules based on the Brownian motion and their chemical shifts. This enables the analysis and separation of substance mixtures [7-11] by means of one measurement without carrying out a complex separation beforehand [11].

In contrast to the usual measurements where a static, homogenous magnetic field is desired, the pulsed field gradient (PFG) technique is used [12] with an inhomogeneous field [13]. For this purpose, linear field gradients are irradiated for a short time, which interfere with the static magnetic field and thus lead to different field strengths within the sample. This leads to a spatial resolution of the individual nuclei since their resonance frequency depends on their position in the magnetic field.

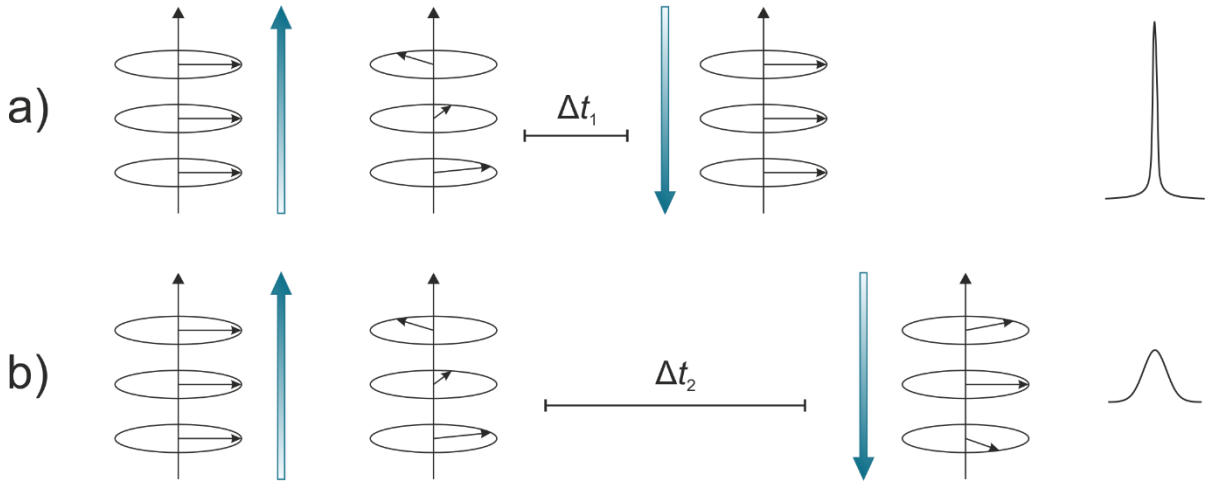


Figure 2 Influence of the time between two field gradients with opposite direction on the echo signal. In case a), no diffusion has taken place (Δt_1 is very small) which leads to a complete refocusing of the spins and a normal echo signal. b) describes the case of diffusion, which leads to incomplete refocusing and thus an attenuated signal.

A widespread technique is the pulse gradient spin echo (PGSE) experiment established by E.O. Stejskal and J.E. Tanner [14]. It is a slightly modified version of the spin echo experiment, in which the attenuation of the signals caused by self-diffusion is measured. A gradient pulse is generated along the z-axis after the 90° pulse, resulting in a different Larmor frequency of the spins within the sample. By applying a field gradient in the opposite direction within a very short time window it would be possible to refocus the spins and measure the echo signal after the 180° pulse. In order to be able to evaluate the diffusion coefficient the second gradient is applied after a predefined time, which gives the molecules time to diffuse. As a result, they are no longer in the same location as in the first gradient and can no longer be refocused. This leads to an attenuation of the echo signal (see Figure 2). This experiment is performed several times with varying gradient strengths and the diffusion coefficient is thus calculated from the attenuation of the signals [3]. For this, the well-known *Stejskal and Tanner* [15] equation can be used where I_G *and* I_0 represent the signal intensities with and without a gradient, respectively, γ the gyromagnetic ratio, δ the gradient duration, G the gradient intensity, $\Delta t'$ the corrected diffusion time and D the diffusion coefficient.

$$I_G = I_0 \, e^{-D\Delta t' \gamma^2 \delta^2 G^2}$$

By plotting the chemical shift against the diffusion coefficient, the typical DOSY diagram is obtained. Here, all signals of a substance can be found on a line corresponding to the diffusion coefficient [7]. Since the diffusion coefficient depends not only on the temperature but also on the size and shape of the molecule, the molecules can be distinguished. This fact is described in the *Stokes-Einstein* relation with D as diffusion coefficient, $\underline{k_B}$ as Boltzmann constant, T as absolute temperature, η as dynamic viscosity and r as radius of a spherical particle [16].

$$D = \frac{k_B T}{6\pi\eta r}$$

Since each slice of the DOSY diagram corresponds to a measurement, the corresponding ^{1}H spectra of the substances can also be obtained in this way. However, this method is limited in the case of strongly overlapping signals of very different intensities [17,15].

1.1.3 Quantitative NMR spectroscopy

Mainly ^{1}H NMR spectroscopy is used for structure elucidations in combination with ^{13}C NMR and special 2D NMR measurements [1,3]. But another use, which nowadays is widely applied, is the ability of quantitative NMR spectroscopy (qNMR). The intensity of a signal, being the area under the curve, is directly proportional to the number of nuclei that generate it [18]. Therefore, not only the ^{1}H nuclei can be used but also many others that are accessible to NMR spectroscopy, e.g., ^{31}P, ^{19}F, ^{35}Cl or ^{79}Br. Nevertheless, there are certain constraints that affect the measurements, such as the natural abundances of the nuclei [19].

In order to achieve accurate results, several points must be taken into account and the correct parameters for the acquisition and processing must be selected. A basic requirement is a pure signal [20]. Sometimes signals have a similar chemical shift and overlap, which does not allow proper integration and distorts the results. A major advantage of qNMR spectroscopy is that, unlike other methods such as high-performance liquid chromatography (HPLC), one molecule usually generates multiple signals and only one well-separated signal is needed for quantification. Even though there are ways to influence the chemical shift, such as solvents and their mixtures, pH value or auxiliary reagents, it is not possible to modify the separation of signals to the same extent as with HPLC [20].

One value that is highly influential to the measurement time is the longitudinal relaxation T_1 which determines the recycling delay. The aim is to give the nuclear spins time to relax almost completely so that a maximum signal can be obtained again. Since the T_1 time is different for each signal in the spectrum and in each solvent, it should be determined beforehand using the inversion-recovery experiment. By default, the recycling delay for a 90° pulse is set to at least five times T_1 of the slowest relaxing signal (corresponds to 99% recovery) [21,22].

Among other parameters that can be optimized, the signal-to-noise-ratio (*S/N*), the digital resolution of the signals and their integration ranges should be considered in any case. They all have a direct impact on subsequent integration. The *S/N* should have a ratio of $S/N > 250{:}1$ for ^{1}H nuclei [23], which is often difficult to achieve at low concentrations due to the inherently low sensitivity of NMR spectroscopy. This can lead to long measurement times, as the increase in *S/N* only improves with the square root of the number of scans [24]. With digital resolution, it should be ensured that the signals are described with at least five points above the half-height

to enable correct integration [22]. For this, the integration range should ideally be 64 times the half-width of the signal [21]. However, this is often not possible due to nearby signals, so that an intermediate solution and compromise must be found.

1.2 Infrared spectroscopy

Infrared (IR) spectroscopy is a very versatile tool that also has its place in pharmaceutical analytics. The basic principle is based on the measurement of the absorbed or emitted frequency range of the irradiated infrared radiation. The infrared radiation used covers the spectral range from 13,000 to 10 cm^{-1} and can be further subdivided into the near infrared (NIR) 13,000 – 4,000 cm^{-1}, the mid infrared (MIR) 4,000 – 400 cm^{-1}, and the far infrared with 400 – 10 cm^{-1} [25].

Due to the emitted electromagnetic radiation the chemical bonds are excited to vibrations. For a chemical bond to start oscillating, radiation must be emitted at the frequency at which it itself oscillates. However, only those bonds absorb light during whose oscillations the dipole moment changes. They are called IR active. All other vibrations, such as in symmetric molecules, are thus IR inactive. They are not accessible to IR spectroscopy and have to be investigated by other methods, e.g. Raman spectroscopy [26].

The vibrations in the molecules can generally be divided into two different modes. The stretching mode, which influence the bond length, and the bending mode, which lead to a change in the bond angle. Most stretching vibrations occur at larger wavenumbers above 1,500 cm^{-1}, since significantly more energy is required to excite them. Bending vibrations require less energy and are therefore mainly found at wave numbers below 1,500 cm^{-1}. In addition, framework vibrations (1,600 – 1,000 cm^{-1}), which are characteristic for the molecule, also occur in this region, which is why this region is also called the fingerprint region. It is therefore often used for identity testing [5].

The position of the signals (absorption bands) in the spectrum depends on the mass of the atoms and the strength of the bond. The smaller the mass and the stronger the bond, the higher the wavenumbers and the more energy is required to make them oscillate. The position of the signals is thus characteristic of the individual bonds and can be used to detect functional groups. Furthermore, the vibrating groups influence each other, which allows one to make a statement about their chemical environment.

1.2.1 Fourier transform infrared spectroscopy

Nowadays mainly Fourier transform infrared (FT-IR) spectroscopy is used because of its clear advantages [25,27]. Compared to conventional IR spectrometers, the entire infrared range is irradiated into the sample at once with the aid of an interferometer. This results in a complex

sum interferogram, in which the intensity is plotted against time. Only by a Fourier transformation one obtains the usual IR spectrum in the frequency domain.

The best-known interferometer is the Michelson interferometer, as can be seen in Figure 3. Here, the IR radiation is split into two beams by means of a beam splitter. One beam is deflected by 90° and reflected by a fixed mirror. The second beam passes straight through the beam splitter and then hits a movable mirror. The two beams are reunited and interfere, adding or canceling each other. The resulting beam is directed onto the sample. Here, each λ corresponds to a certain time of the mirror movement.

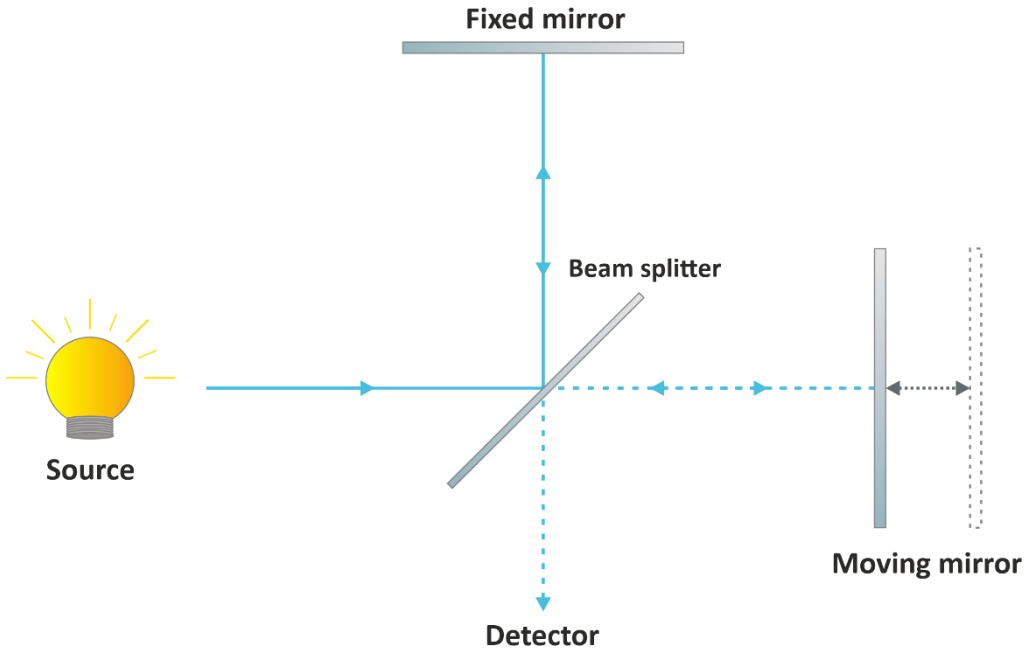


Figure 3 Schematic presentation of the Michelson interferometer

The main advantages of this method are shorter measurement times, since the entire IR spectrum is recorded at once, and high precision of wavenumber detection due to calibration by means of a laser.

1.2.2 NIR

The NIR range extends from 13,000 – 4,000 cm^{-1} (760 – 2,600 nm). In this range, mainly the overtone vibrations of the X-H valence and combination vibrations can be seen [28]. Accordingly, there are no characteristic and intense absorption bands as in MIR spectroscopy, which makes chemometric evaluation necessary [29]. For this purpose, reference databases must be recorded and pre-processed to be able to guarantee a sufficiently precise determination of the target quantity later on, such as the identity or the content of a drug.

The basic setup of a NIR spectrometer with light source, monochromator and detector is the same as that of a MIR instrument, but in addition to gratings, the monochromators also use interference filters. A special feature is the use of glass fiber probes, which enables in situ

analysis and the possibility to measure solid samples, solutions, liquids or suspensions. Three different measurement methods can be used for this purpose. The transmission, the diffuse reflection and the transflection [5]. Diffuse reflection is a combination of internal and external scattering and is mostly used for solids. Here, part of the radiation is reflected directly from the surface, while another part penetrates the sample and is partially absorbed. Since the radiation penetrates different substances to different depths, a reflection standard is necessary. The ratio of the reflected radiation from the sample (I_s) to the reflection standard (I_r) is called the reflectance (R). Generally, the spectra for diffuse reflection are given in the pseudo-absorption A_R.

$$R = \frac{I_s}{I_r} \qquad A_R = \log\frac{1}{R}$$

In summary, the major advantages of NIR spectroscopy are the very short measurement times, very simple sample preparation, which can be completely omitted, the precision and the possibility of measurements through glass vessels and on/in/at line. A prerequisite for this, however, is a rather complex method development which requires high qualified personal and adequate data sets [30].

1.3 Multivariate Data Analysis

Multivariate Data Analysis (MVA) or Chemometrics is the application of mathematical and statistical methods to chemical data [31] to optimize measurement procedures and experiments and to extract relevant chemical information [32,33]. In the field of spectroscopic analysis, it is used to reduce the data dimensionality of hundreds of data points and to display the relevant information contained by means of descriptive graphics. A method frequently used for this purpose is principal component analysis (PCA), which therefore will be discussed more detailed. Besides such methods that are used to recognize the relationships between the data and to classify them, there are also regression methods. They combine the spectra with a target variable and can be used to determine the content of active ingredients for example. One of the best known representatives of these methods is probably the partial least square regression [34].

1.3.1 Data pre-processing

Pre-processing of spectral data is a crucial part of chemometric analysis. The aim is to extract the important spectral information by reducing the noise or to prepare it for further processing and analysis [35]. Depending on the spectral measurement method, several of different methods are applicable. Standard Normal Variate (SNV) or Multiplicative Scatter Correction (MSC) transformations are often applied to IR data to eliminate scatter and additive effects.

Here, SNV scales and centers each individual spectrum, while MSC refers to the mean spectrum [36]. Further commonly used transformations are baseline corrections, transformation from transmission or reflectance spectra to absorption spectra or derivatives. Forming the derivatives can remove baseline effects and increase resolution [34]. During derivation, smoothing is also frequently applied. However, caution and trained personnel is required. If the smoothing is too strong, important spectral information can be lost.

For NMR data, the so-called bucketing plays an important role. First, it corrects small deviations in the chemical shift and second, it leads to a data reduction that enables further processing. For this purpose, the spectral range of the spectrum is divided into buckets of equal size, with the resulting value corresponding to the sum of intensities within the bucket (see Figure 4) [37].

Further pre-treatments that are applied before a multivariate method are centering and scaling. Commonly used mean-centering shifts the zero point to the middle of the center of the multivariate data, focusing the analysis on the portion of the data that varies and thus on the differences between the samples [35,38]. Scaling gives all variables the same influence. This is important to give a weighting to small signals in spectral data, otherwise information could be lost that might be relevant. Furthermore, it is used when different variables are used in the data set, such as by IR and NMR measurements and these are to be analyzed simultaneously [35,39].

Another valuable method is the variable selection to further improve the informative value of a method. With some prior knowledge of the sample, variables that only contain noise and not any information for the corresponding question can be excluded from further analysis.

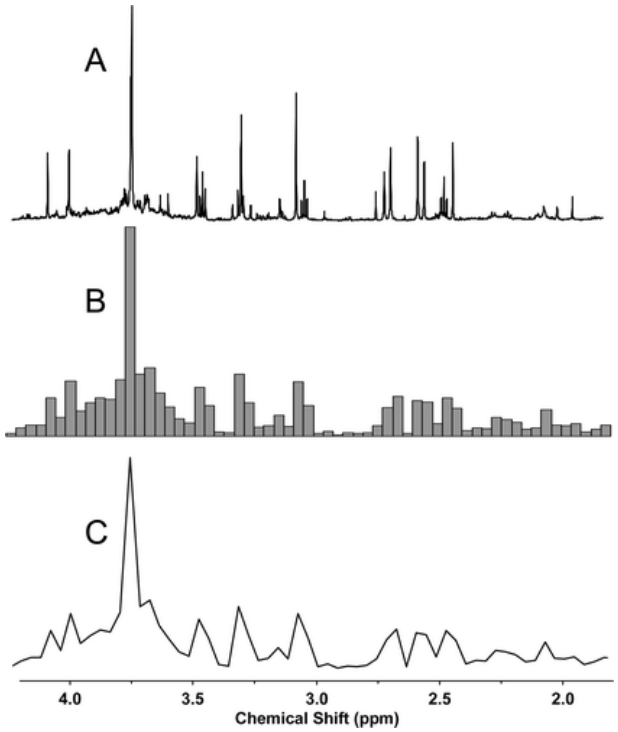


Figure 4 Illustration of data reduction through bucketing on the basis of a ^{1}H NMR spectrum (600 MHz) of rat urine (Reprinted (adapted) with permission from Cherney E et al. [40]. Copyright (2007) American Chemical Society.)

1.3.2 Principal Component Analysis

Principal Component Analysis (PCA) is one of the most commonly used multivariate methods for first exploratory data analysis. PCA helps to recognize correlations in the data, to identify outliers or the essential spectral range but most of all it helps to present the data in a clear and understandable way with suitable graphical plots [41].

The PCA calculates so called principal components (PC), which are a linear combination of the original variables [34]. In this way, the hundreds of variables that make up the spectra can be described with a few PCs, which also results in a significant data reduction. The PCs form a new coordinate system in multivariate space with the new center point at the center of the data. Here, the first PC points along the largest variance and the other PCs are each orthogonal to it. A simplified representation can be seen in Figure 5.

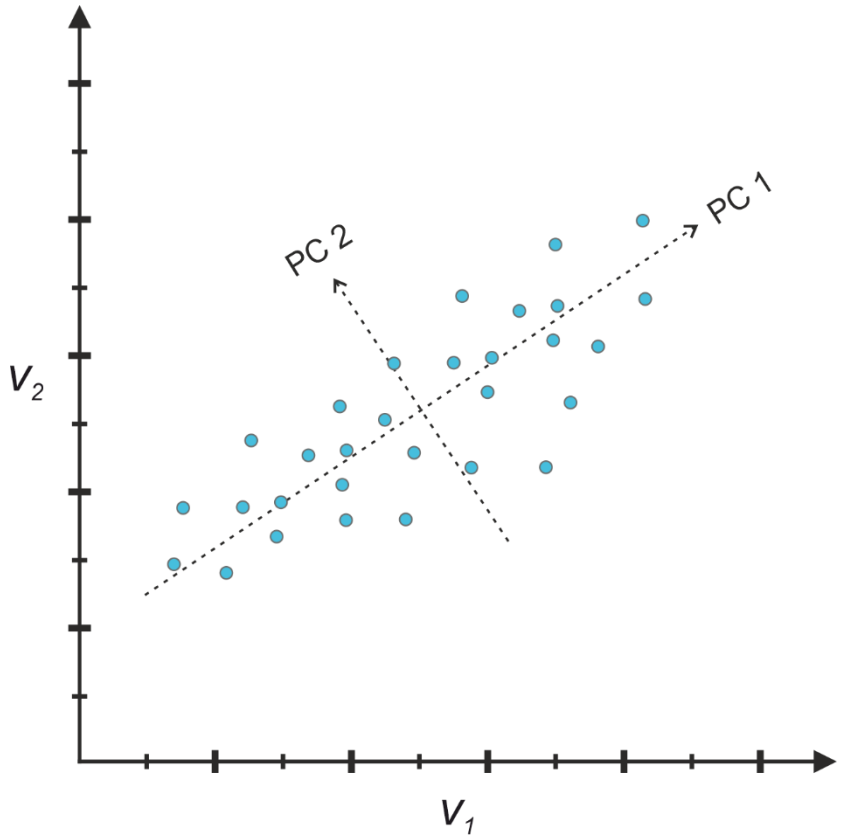


Figure 5 Schematic representation of data points of variables v_1 and v_2 and the corresponding coordinate system spanned by the first two principal components PC1 and PC2. Here PC1 points along the largest variance and the origin is in the middle of the data.

Various methods can be used to calculate the PCs. Usually, these are calculated using the mean-centered covariance matrix of the data matrix. The singular value decomposition (SVD) is based on the fact that a matrix can be represented by the product of three specific matrices and always determines all PCs. The second algorithm is an iterative procedure that can be terminated after the relevant PCs, the nonlinear iterative partial least squares (NIPALS) algorithm.

The PCA can be represented schematically with the equation presented in Figure 6. Here X corresponds to the mean-centered data matrix with n rows (samples) and M columns (variables). This is decomposed into the scores matrix T also with n rows and F principal components and the loadings matrix P^T with F principal components and M columns. The loadings indicate the influence of a variable on the principal component and the scores correspond to the projection of the original data on the principal component axes. The residual matrix E is the error or unexplained amount of the original data after a given number of principal components F. If the number of variables and principal components were equal, it would yield zero. However, since one wants to obtain a data reduction, usually $M \gg F$.

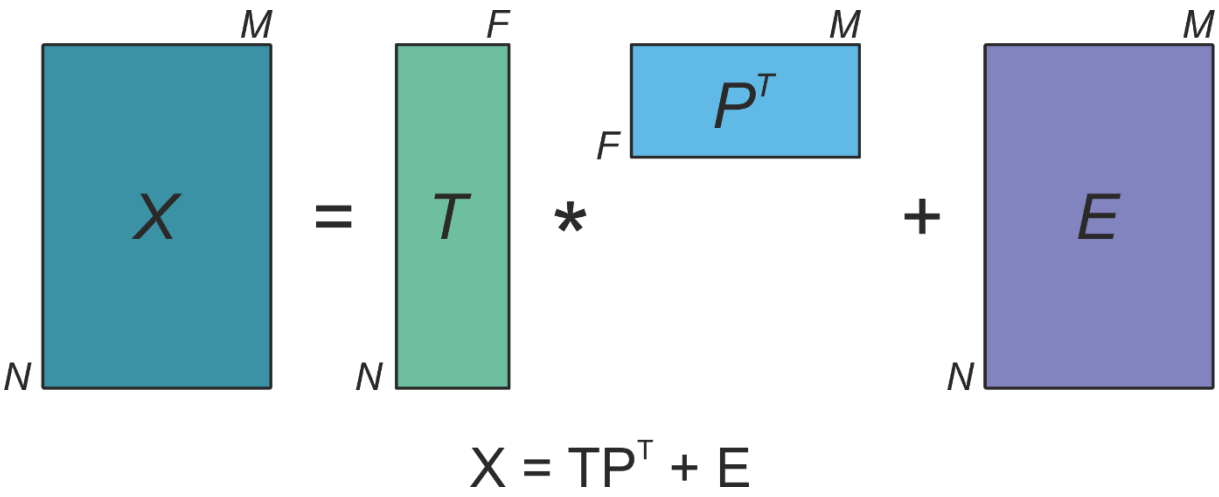


Figure 6 Schematic illustration of the principal component analysis with the mean-centered data matrix *X* with all the samples and variables, which are represented by the scores matrix *T*, the loadings matrix P^T and the residual matrix *E*.

1.3.3 Linear Discriminant Analysis

Linear discriminant analysis (LDA) is a supervised classification method in which the number of categories is determined in advance and each sample is assigned to one of these categories [42]. LDA uses this classification to divide the multivariate space into different regions. This is done using hyperplanes [41]. Thus, unlike PCA, the computed vector is not one that corresponds to the largest variance in the data, but one that can best discriminate between categories. Thus, the goal in setting the discriminant function is to minimize the differences within a class while maximizing them to the other classes. [43]

However, with spectral data, one often runs into the problem of having more variables than samples, which makes it impossible to compute the LDA. To avoid this, a PCA can be performed beforehand and then the calculation can be continued with the scores [41].

1.4 Application of MIR, NIR and NMR spectroscopic methods in the pharmaceutical field

Nowadays, spectroscopic methods are used in many different fields. In medicine for diagnostic purposes such as magnetic resonance imaging, in food control for identity or impurity analysis and in chemistry for structure elucidation or reaction monitoring, to name just a few examples. They have long been established in the pharmaceutical field, where they are classically used for drug quality control. The methods are well described in the European Pharmacopoeia but are found to varying degrees in the monographs. MIR and UV-Vis spectroscopy are very commonly used for identity and impurity testing, whereas NIR and NMR spectroscopy are not yet widely used. NIR is only mentioned in the determination of the water content of a few plasma proteins fractions, e.g., human coagulation factor VII. NMR methods are mostly used for identification purposes of drugs or excipients [19]. However, this does not mean that they

are of lesser importance. Outside of monographs, there are now a variety of applications and advancements that go beyond these common uses. For example, some papers deal with DOSY measurements as a screening method. Due to the separation of substances based on the diffusion coefficient, it is very well possible to describe and investigate mixtures. This allows the detection of impurities in drugs, the identification and investigation of drug falsifications, and the characterization of drug formulations [8,7,10,9]. The same applies to dietary supplements or herbal drugs [44]. In subchapters 3.4 through 3.5, a more detailed review of the capabilities of DOSY measurements in characterizing drug formulations is given.

Other methods were briefly presented by Holzgrabe in 2019 [45]. Amongst them are two methods of analysis for stable isotopes such as the internal site specific $^{12}C/^{13}C$ ratio, which can be used as a fingerprinting method for manufacturer identification, API traceability or counterfeit detection [46-48]. In addition, there have been technical developments. Assemat et al. [49,50] were able to show that benchtop NMRs are also capable of analyzing drugs by means of ^{1}H and DOSY measurements. This could give smaller companies and laboratories access to NMR measurements and can be beneficial to increasing the prevalence of such methods in compendial and routine analyses. MIR spectroscopy also has a variety of other applications and methods for pharmaceutical use, which are discussed in many research articles and reviews. Several different techniques are applied, including the FTIR attenuated total reflectance (FTIR-ATR) spectroscopy, FTIR photo acoustic spectroscopy (FTIR-PAS), nano-FTIR spectroscopy and FTIR imaging spectroscopy [27,25,51-53]. These techniques are then widely used to characterize crystalline and amorphous forms or polymers, identify counterfeits, study drug release and drug penetration in membranes, or analyze the distribution of active drugs and excipients.

1.4.1 Spectroscopic methods in combination with chemometrics

One tool that further improved some of these spectroscopic analysis techniques and even made others possible in the first place was chemometrics. Especially NIR spectroscopy benefited from this since the spectra obtained here are not as characteristic by mere examination as those obtained by MIR or NMR spectroscopy. On a broad scale, NIR spectroscopy did not become established in the pharmaceutical industry until the introduction of process analytical technology (PAT) in the early 2000s [54]. Nowadays it is an important part of process monitoring and quality control due to its variable use and the possibility to determine physical and chemical parameters [55]. Especially the nondestructive nature and possibility to measure directly in the samples due to fiber optics or the use as a handheld instrument that can measure samples through packaging makes NIR spectroscopy so valuable [56]. These features and the development of fast NIR instruments even allow 100% monitoring as well as continuous validation throughout the entire manufacturing process [57,58].

Applications vary from tests for content uniformity and active ingredient content, to determinations of pharmaceutical parameters such as particle size or dissolution rate as well as moisture determination, which is of interest for lyophilization and granulation processes [59-67]. Even though NIR spectroscopy plays a major role in PAT, other spectroscopic methods such as FTIR, Raman, and UV-Vis are also widely used, depending on the task at hand.

Besides the use of chemometrics in process controlling, it can also be used to counteract one major problem in the pharmaceutical field - drug counterfeiting [68]. These have a major impact on public health, as well as economic and social consequences [69]. The World Health Organization defined three different classes: 1) so-called substandard medicines, which do not meet the required specifications, 2) unregistered/unlicensed medical products, which have not undergone evaluation and/or approval by the appropriate regulatory authority, and 3) falsified medical products [70]. The type of counterfeiting can be quite different, e.g. it may concern the composition (no or wrong active ingredient, wrong content) or the packaging. The challenge is to detect and trace such counterfeits. Traceability in particular is a major problem, as the supply chains in the pharmaceutical industry are very complex [68]. The APIs and excipients can be produced, processed, packaged, repackaged and sold in different countries. Therefore, the possibility of tracing back drugs to their manufacturer is of paramount interest for safe medicines and is more detailed in subchapter 3.1.

However, it has been shown in several publications that spectroscopic methods, especially in combination with chemometrics, are very well suited for authenticity control [71-73]. A brief overview of the uses of NIR, MIR, and NMR methods and their purpose is given in Table 1.

Table 1 Examples for the use of IR and NMR spectroscopic methods for authenticity control

Spectroscopic method	Application	Reference
Infrared		
FTIR-ATR	Discrimination of genuine from counterfeit samples of Viagra® and Cialis®	[71]
IR-ATR/NIR	Quality control of Chinese traditional medicine	[74]
NIR	Detection of counterfeit antimalarial tablets	[75]
NIR	Authentication of fluconazole capsules	[76]
NMR		
DOSY	Mixture analysis of genuine and fake sildenafil formulations	[77]
Low field ^{1}H-NMR	Revealing the adulteration of dietary supplements by pharmaceutical compounds	[78]

References

[1] Friebolin H (2013) Ein- und Zweidimensionale NMR-Spektroskopie : Eine Einführung. 5 edn. John Wiley & Sons, Incorporated, Weinheim, GERMANY

[2] Mlynárik V Introduction to nuclear magnetic resonance. *Anal Biochem* (2017) 529, 4-9

[3] Günther H (2013) NMR Spectroscopy : Basic Principles, Concepts and Applications in Chemistry. John Wiley & Sons, Incorporated, Somerset, GERMANY

[4] Bruice PY (2011) Organische Chemie : Studieren kompakt. Pearson Deutschland GmbH,

[5] Neugebauer M, Scriba GKE, Hubert MA, Rücker G, Neugebauer M, Willems GG (2013) Rücker/Neugebauer/Willems Instrumentelle pharmazeutische Analytik : Lehrbuch zu spektroskopischen, chromatographischen, elektrochemischen und thermischen Analysenmethoden

[6] Holzgrabe U, Wawer I, Diehl B (1999) NMR Spectroscopy in Drug Development and Analysis. WILEY-VCH,

[7] Balayssac S, Gilard V, Delsuc M-A, Malet-Martino M DOSY NMR, a new tool for fake drug analyses. *Spectrosc Eur* (2009) 21, 10-14

[8] Balayssac S, Retailleau E, Bertrand G, Escot M-P, Martino R, Malet-Martino M, Gilard V Characterization of heroin samples by 1H NMR and 2D DOSY 1H NMR. *Forensic Sci Int* (2014) 234, 29-38

[9] Kellenbach E, Burgering M, Kaspersen F Using Pulse Field Gradient NMR-Based Diffusion Experiments To Identify Signals of Low-Molecular-Weight Impurities. *Org Process Res Dev* (1999) 3, 141-144

[10] Trefi S, Gilard V, Malet-Martino M, Martino R Generic ciprofloxacin tablets contain the stated amount of drug and different impurity profiles: A 19F, 1H and DOSY NMR analysis. *J Pharm Biomed Anal* (2007) 44, 743-754

[11] Antalek B Using pulsed gradient spin echo NMR for chemical mixture analysis: How to obtain optimum results. *Concepts Magn Reson* (2002) 14, 225-258

[12] Barjat H, Morris GA, Smart S, Swanson AG, Williams SCR High-Resolution Diffusion-Ordered 2D Spectroscopy (HR-DOSY) - A New Tool for the Analysis of Complex Mixtures. *J Magn Reson, Ser B* (1995) 108, 170-172

[13] Stilbs P Fourier transform pulsed-gradient spin-echo studies of molecular diffusion. *Prog Nucl Magn Reson Spectrosc* (1987) 19, 1-45

[14] Stejskal EO, Tanner JE Spin Diffusion Measurements: Spin Echoes in the Presence of a Time-Dependent Field Gradient. *J Chem Phys* (1965) 42, 288-292

[15] Pagès G, Bonny A, Gilard V, Malet-Martino M Pulsed Field Gradient NMR with Sigmoid Shape Gradient Sampling To Produce More Detailed Diffusion Ordered Spectroscopy Maps of Real Complex Mixtures: Examples with Medicine Analysis. *Anal Chem* (2016) 88, 3304-3309

[16] Li D, Kagan G, Hopson R, Williard PG Formula Weight Prediction by Internal Reference Diffusion-Ordered NMR Spectroscopy (DOSY). *Journal of the American Chemical Society* (2009) 131, 5627-5634

[17] Becht A, Schollmayer C, Wiest J, Heller D, Baumann W, Buschmann H, Holzgrabe U Diffusion ordered NMR spectroscopy measurements as screening method of potential reactions of API and excipients in drug formulations. *J Pharm Biomed Anal* (2019) 162, 41-46

[18] Malz F, Jancke H Validation of quantitative NMR. *J Pharm Biomed Anal* (2005) 38, 813-823

[19] Diehl B, Holzgrabe U, Monakhova Y, Schönberger T Quo Vadis qNMR? *J Pharm Biomed Anal* (2020) 177, 112847

[20] Beyer T, Schollmayer C, Holzgrabe U The role of solvents in the signal separation for quantitative 1H NMR spectroscopy. *J Pharm Biomed Anal* (2010) 52, 51-58

[21] Holzgrabe U Quantitative NMR spectroscopy in pharmaceutical applications. *Prog Nucl Magn Reson Spectrosc* (2010) 57, 229-240

[22] Szńtay C High-field NMR spectroscopy as an analytical tool for quantitative determinations: Pitfalls, limitations and possible solutions. *TrAC, Trends Anal Chem* (1992) 11, 332-344

[23] Martino R, Gilard V, Desmoulin F, Malet-Martino M Fluorine-19 or phosphorus-31 NMR spectroscopy: A suitable analytical technique for quantitative in vitro metabolic studies of fluorinated or phosphorylated drugs. *J Pharm Biomed Anal* (2005) 38, 871-891

[24] Maniara G, Rajamoorthi K, Rajan S, Stockton GW Method Performance and Validation for Quantitative Analysis by 1H and 31P NMR Spectroscopy. Applications to Analytical Standards and Agricultural Chemicals. *Anal Chem* (1998) 70, 4921-4928

[25] Bunaciu AA, Aboul-Enein HY, Fleschin S Application of Fourier Transform Infrared Spectrophotometry in Pharmaceutical Drugs Analysis. *Appl Spectrosc* (2010) 45, 206-219

[26] Schrader B (2017) Raman Spectrometers. In: Lindon JC, Tranter GE, Koppenaal DW (eds) Encyclopedia of Spectroscopy and Spectrometry (Third Edition). Academic Press, Oxford, p 900-905

[27] Dole MN, Patel PA, Sawant SD, Shedpure PS Advance applications of Fourier transform infrared spectroscopy. *Int J Pharm Sci Rev Res* (2011) 7, 159-166

[28] Siesler HW, Ozaki Y, Kawata S, Heise HM (2002) Near-Infrared Spectroscopy: Principles, Instruments, Applications. Wiley,

[29] Roggo Y, Chalus P, Maurer L, Lema-Martinez C, Edmond A, Jent N A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *J Pharm Biomed Anal* (2007) 44, 683-700

[30] Jamrógiewicz M Application of the near-infrared spectroscopy in the pharmaceutical technology. *J Pharm Biomed Anal* (2012) 66, 1-10

[31] Kowalski BR Chemometrics. *Anal Chem* (1980) 52, 112-122

[32] Otto M (2016) Chemometrics: Statistics and Computer Application in Analytical Chemistry. Third edn. Wiley-VCH,

[33] Wold S Chemometrics; what do we mean with it, and what do we want from it? *Chemometrics Intellig Lab Syst* (1995) 30, 109-115

[34] Kessler W (2007) Multivariate Datenanalyse: für die Pharma, Bio- und Prozessanalytik. WILEY-VCH Verlag GmbH & Co. KGaA,

[35] Brereton RG, Jansen J, Lopes J, Marini F, Pomerantsev A, Rodionova O, Roger JM, Walczak B, Tauler R Chemometrics in analytical chemistry—part I: history, experimental design and data analysis tools. *Anal Bioanal Chem* (2017) 409, 5891-5899

[36] Rinnan Å, Berg Fvd, Engelsen SB Review of the most common pre-processing techniques for near-infrared spectra. *TrAC, Trends Anal Chem* (2009) 28, 1201-1222

[37] Worley B, Powers R Multivariate Analysis in Metabolomics. *Curr Metabolomics* (2013) 1, 92-107

[38] van den Berg RA, Hoefsloot HCJ, Westerhuis JA, Smilde AK, van der Werf MJ Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* (2006) 7, 142

[39] Wold S, Esbensen K, Geladi P Principal component analysis. *Chemometrics Intellig Lab Syst* (1987) 2, 37-52

[40] Cherney DP, Ekman DR, Dix DJ, Collette TW Raman Spectroscopy-Based Metabolomics for Differentiating Exposures to Triazole Fungicides Using Rat Urine. *Anal Chem* (2007) 79, 7324-7332

[41] Biancolillo A, Marini F Chemometric Methods for Spectroscopy-Based Pharmaceutical Analysis. *Front Chem* (2018) 6

[42] Cozzolino D, Chree A, Scaife JR, Murray I Usefulness of Near-Infrared Reflectance (NIR) Spectroscopy and Chemometrics To Discriminate Fishmeal Batches Made with Different Fish Species. *Journal of Agricultural and Food Chemistry* (2005) 53, 4459-4463

[43] Martinez AM, Kak AC PCA versus LDA. *IEEE Trans Pattern Anal Mach Intell* (2001) 23, 228-233

[44] Balayssac S, Trefi S, Gilard V, Malet-Martino M, Martino R, Delsuc M-A 2D and 3D DOSY 1H NMR, a useful tool for analysis of complex mixtures: Application to herbal drugs or dietary supplements for erectile dysfunction. *J Pharm Biomed Anal* (2009) 50, 602-612

[45] Holzgrabe U (2019) Nuclear Magnetic Resonance Spectroscopy | Identification of Counterfeit Drugs and Herbal Medicine. In: Worsfold P, Poole C, Townshend A, Miró M (eds) Encyclopedia of Analytical Science (Third Edition). Academic Press, Oxford, pp 182-185

[46] Silvestre V, Mboula VM, Jouitteau C, Akoka S, Robins RJ, Remaud GS Isotopic 13C NMR spectrometry to assess counterfeiting of active pharmaceutical ingredients: Site-specific 13C content of aspirin and paracetamol. *J Pharm Biomed Anal* (2009) 50, 336-341

[47] Bussy U, Thibaudeau C, Thomas F, Desmurs J-R, Jamin E, Remaud GS, Silvestre V, Akoka S Isotopic finger-printing of active pharmaceutical ingredients by 13C NMR and polarization transfer techniques as a tool to fight against counterfeiting. *Talanta* (2011) 85, 1909-1914

[48] Remaud GS, Bussy U, Lees M, Thomas F, Desmurs J-R, Jamin E, Silvestre V, Akoka S NMR spectrometry isotopic fingerprinting: A tool for the manufacturer for tracking Active Pharmaceutical Ingredients from starting materials to final medicines. *Eur J Pharm Sci* (2013) 48, 464-473

[49] Assemat G, Balayssac S, Gerdova A, Gilard V, Caillet C, Williamson D, Malet-Martino M Benchtop low-field 1H Nuclear Magnetic Resonance for detecting falsified medicines. *Talanta* (2019) 196, 163-173

[50] Assemat G, Gouilleux B, Bouillaud D, Farjon J, Gilard V, Giraudeau P, Malet-Martino M Diffusion-ordered spectroscopy on a benchtop spectrometer for drug analysis. *J Pharm Biomed Anal* (2018) 160, 268-275

[51] Kowalczuk D, Pitucha M Application of FTIR Method for the Assessment of Immobilization of Active Substances in the Matrix of Biomedical Materials. *Materials* (2019) 12, 2972

[52] Van Eerdenbrugh B, Taylor LS Application of mid-IR spectroscopy for the characterization of pharmaceutical systems. *Int J Pharm* (2011) 417, 3-16

[53] Wartewig S, Neubert RHH Pharmaceutical applications of Mid-IR and Raman spectroscopy. *Adv Drug Del Rev* (2005) 57, 1144-1170

[54] Swarbrick B The Current State of near Infrared Spectroscopy Application in the Pharmaceutical Industry. *J Near Infrared Spectrosc* (2014) 22, 153-156

[55] Prajapati P, Solanki R, Modi V, Basuri T A brief review on NIR spectroscopy and its pharmaceutical applications. *Int J Pharm Chem Anal* (2016) 3, 117-123

[56] Martino R, Malet-Martino M, Gilard V, Balayssac S Counterfeit drugs: analytical techniques for their identification. *Anal Bioanal Chem* (2010) 398, 77-92

[57] Challa S, Potumarthi R Chemometrics-based process analytical technology (PAT) tools: applications and adaptation in pharmaceutical and biopharmaceutical industries. *Appl Biochem Biotechnol* (2013) 169, 66-76

[58] Kessler RW (2012) Strategien und Fallbeispiele aus der industriellen Praxis. In: Prozessanalytik. WILEY-VCH Verlag GmbH & Co. KGaA,

[59] Berntsson O, Zackrisson G, Östling G Determination of moisture in hard gelatin capsules using near-infrared spectroscopy: applications to at-line process control of pharmaceutics. *J Pharm Biomed Anal* (1997) 15, 895-900

[60] Blanco M, Bautista M, Alcalà M API Determination by NIR Spectroscopy Across Pharmaceutical Production Process. *AAPS PharmSciTech* (2008) 9, 1130-1135

[61] Blanco M, González Bañó R, Bertran E Monitoring powder blending in pharmaceutical processes by use of near infrared spectroscopy. *Talanta* (2002) 56, 203-212

[62] Chalus P, Roggo Y, Walter S, Ulmschneider M Near-infrared determination of active substance content in intact low-dosage tablets. *Talanta* (2005) 66, 1294-1302

[63] De Beer TRM, Vercruysse P, Burggraeve A, Quinten T, Ouyang J, Zhang X, Vervaet C, Remon JP, Baeyens WRG In-line and real-time process monitoring of a freeze drying process using Raman and NIR spectroscopy as complementary process analytical technology (PAT) tools. *J Pharm Sci* (2009) 98, 3430-3446

[64] Frake P, Greenhalgh D, Grierson SM, Hempenstall JM, Rudd DR Process control and end-point determination of a fluid bed granulation by application of near infra-red spectroscopy. *Int J Pharm* (1997) 151, 75-80

[65] Freitas MP, Sabadin A, Silva LM, Giannotti FM, do Couto DA, Tonhi E, Medeiros RS, Coco GL, Russo VFT, Martins JA Prediction of drug dissolution profiles from tablets using NIR diffuse reflectance spectroscopy: A rapid and nondestructive method. *J Pharm Biomed Anal* (2005) 39, 17-21

[66] Jones JA, Last IR, MacDonald BF, Prebble KA Development and transferability of near-infrared methods for determination of moisture in a freeze-dried injection product. *J Pharm Biomed Anal* (1993) 11, 1227-1231

[67] O'Neil AJ, Jee RD, Moffat AC Measurement of the percentage volume particle size distribution of powdered microcrystalline cellulose using reflectance near-infrared spectroscopy. *Analyst* (2003) 128, 1326-1330

[68] Dégardin K, Roggo Y, Margot P Understanding and fighting the medicine counterfeit market. *J Pharm Biomed Anal* (2014) 87, 167-175

[69] Medicine Io (2013) Countering the Problem of Falsified and Substandard Drugs. The National Academies Press, Washington, DC

[70] World Health Organization (2018) Substandard and falsified medical products. https://www.who.int/news-room/fact-sheets/detail/substandard-and-falsified-medical-products. Accessed 06.04.2021

[71] Custers D, Cauwenbergh T, Bothy JL, Courselle P, De Beer JO, Apers S, Deconinck E ATR-FTIR spectroscopy and chemometrics: An interesting tool to discriminate and characterize counterfeit medicines. *J Pharm Biomed Anal* (2015) 112, 181-189

[72] Krakowska B, Custers D, Deconinck E, Daszykowski M Chemometrics and the identification of counterfeit medicines—A review. *J Pharm Biomed Anal* (2016) 127, 112-122

[73] Sacré P-Y, Deconinck E, De Beer T, Courselle P, Vancauwenberghe R, Chiap P, Crommen J, De Beer JO Comparison and combination of spectroscopic techniques for the detection of counterfeit medicines. *J Pharm Biomed Anal* (2010) 53, 445-453

[74] C.K. Pezzei MW, V.A. Huck-Pezzei, C.B.S. Lau, Z. Zuo, P.C. Leung, C.W. Huck Infrared spectroscopic techniques for the non-invasive and rapid quality control of Chinese traditional medicine Si-Wu-Tang. *Spectrosc Eur* (2016) 28, 16-21

[75] Dowell FE, Maghirang EB, Fernandez FM, Newton PN, Green MD Detecting counterfeit antimalarial tablets by near-infrared spectroscopy. *J Pharm Biomed Anal* (2008) 48, 1011-1014

[76] Rodionova OY, Titova AV, Demkin NA, Balyklova KS, Pomerantsev AL Qualitative and quantitative analysis of counterfeit fluconazole capsules: A non-invasive approach using NIR spectroscopy and chemometrics. *Talanta* (2019) 195, 662-667

[77] Trefi S, Gilard V, Balayssac S, Malet-Martino M, Martino R The usefulness of 2D DOSY and 3D DOSY-COSY 1H NMR for mixture analysis: application to genuine and fake formulations of sildenafil (Viagra). *Magn Reson Chem* (2009) 47, S163-S173

[78] Wu N, Balayssac S, Danoun S, Malet-Martino M, Gilard V Chemometric Analysis of Low-field 1H NMR Spectra for Unveiling Adulteration of Slimming Dietary Supplements by Pharmaceutical Compounds. *Molecules* (2020) 25, 1193

2. Aims of the thesis

Spectroscopic methods have been used for a long time in a wide variety of fields for quite different tasks, including the pharmaceutical environment. In this thesis, new approaches of some spectroscopic methods have been investigated to show the enduring importance and possibilities in this field in the present time. This includes chemometric evaluation of spectral data, quantitative ^{1}H NMR spectroscopy using a benchtop instrument, as well as DOSY measurements for purity analysis or generation of fingerprint spectra. The individual topics that are intended to support the merits of spectral analysis methods are described more detailed below.

Chemometric evaluation of spectral data

In the introduction, some possibilities of the chemometric evaluation of spectral data have already been described, as well as their increased importance, for example, in the production of pharmaceuticals. In this work, further innovative approaches are to be pursued in this respect. Using paracetamol tablets from various pharmaceutical companies as a reference group, it is to be investigated whether it is possible to determine the origin of the individual preparations, since many marketing authorization holders no longer manufacture their drugs themselves, but rather at a contract manufacturing organization (CMO) which makes it considerably more difficult to identify the origin beyond doubt. For this purpose, analytical measurement procedures coupled with chemometric evaluation methods will be applied. In order to achieve these goals, the following points are to be worked on:

- Development of suitable sample preparations and measurement methods to obtain reproducible MIR, NIR and NMR spectra of the paracetamol tablets.
- Comparison of unsupervised chemometric methods (principal component analysis (PCA) and hierarchical cluster analysis (HCA)), as well as supervised methods (linear discriminant analysis (LDA)) to identify the origin of the tablets.
- Development of a NIR attachment to measure intact tablets.
- Use of a DOSY filter to obtain fingerprint spectra for different formulations.

Stability studies using DOSY

Under certain conditions, DOSY measurements can be used to distinguish large molecules from small molecules on the basis of the diffusion coefficient. Because of this fact, the possibilities of DOSY measurement shall be investigated to use it as a screening method in

stability studies of drugs. This will be tested based on possible reactions between APIs and excipients.

Quantitative ^{1}H NMR spectroscopy

NMR spectroscopy has developed continuously, not only in terms of measurement methods, but also in terms of the devices themselves. In the meantime, devices are available that merely occupy small spaces upon the analysis bench. The aim of this project is to investigate whether these benchtop instruments are capable of producing quantitative results comparable to those obtained with a high-field spectrometer.

3. Results

3.1 Tracing the origin of paracetamol tablets by near infrared, mid infrared and nuclear magnetic resonance spectroscopy using principal component analysis and linear discriminant analysis

Alexander Becht, Curd Schollmayer, Yulia Monakhova, Ulrike Holzgrabe

Copyright

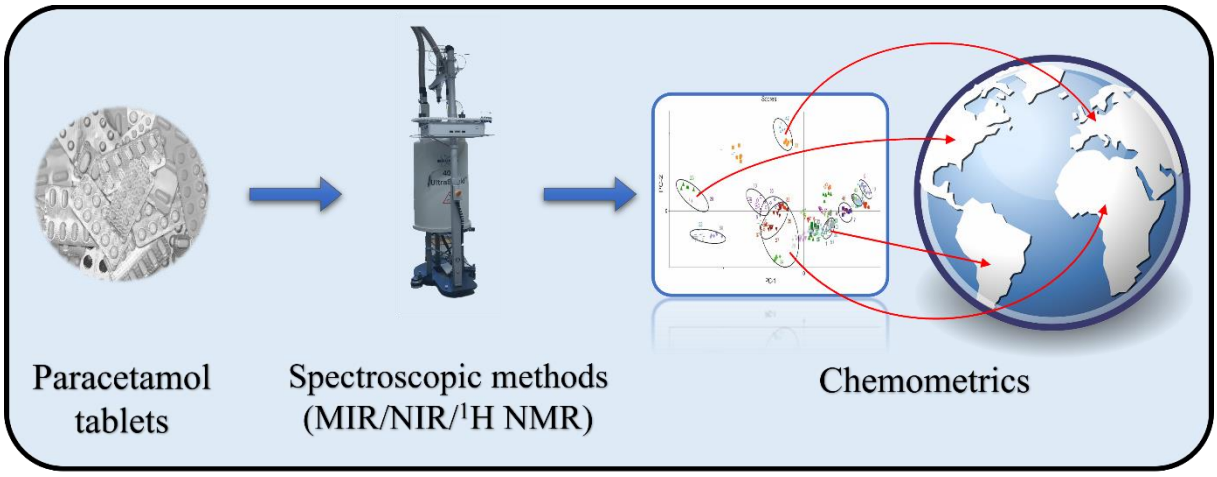
Anal. Bioanal. Chem. **2021**, *413*, 3107-3118.

https://doi.org/10.1007/s00216-021-03249-z

Abstract

Most drugs are no longer produced in their own countries by the pharmaceutical companies, but by contract manufacturers or at manufacturing sites in countries that can produce more cheaply. This not only makes it difficult to trace them back but also leaves room for criminal organizations to fake them unnoticed. For these reasons, it is becoming increasingly difficult to determine the exact origin of drugs. The goal of this work was to investigate how exactly this is possible by using different spectroscopic methods like nuclear magnetic resonance and near- and mid-infrared spectroscopy in combination with multivariate data analysis. As an example, 56 out of 64 different paracetamol preparations, collected from 19 countries around the world, were chosen to investigate whether it is possible to determine the pharmaceutical company, manufacturing site, or country of origin. By means of suitable pre-processing of the spectra and the different information contained in each method, principal component analysis was able to evaluate manufacturing relationships between individual companies and to differentiate between production sites or formulations. Linear discriminant analysis showed different results depending on the spectral method and purpose. For all spectroscopic methods, it was found that the classification of the preparations to their manufacturer achieves better results than the classification to their pharmaceutical company. The best results were obtained with nuclear magnetic resonance and near-infrared data, with 94.6%/99.6% and 98.7/100% of the spectra of the preparations correctly assigned to their pharmaceutical company or manufacturer.

Graphical Abstract



1. Introduction

Nowadays, it is very difficult to determine the origin of a drug based on the declaration on primary or secondary packaging. One reason for this is the fact that a large number of counterfeit drugs are in circulation, not only in developing countries but also worldwide [1]. In addition to the actual medication, the packaging can be copied down to the last detail. Even the holograms can be counterfeited so perfect that these packages look more trustworthy than the original [2]. Another reason is that often many different companies are involved in the manufacturing process and most of them are not named in publicly available documents. In most cases, rather than the specific manufacturer, the marketing authorization holder is mentioned on the primary and secondary packaging, respectively. Research to figure out the manufacturer often leads to debatable websites. Even in developed countries, the manufacturer is not always clearly defined. For instance, in Germany, the pharmaceutical companies have to name the manufacturer in the package insert. However, these manufacturers just have to execute the last step in the production chain, which is the certification of the final product release. The excipients, the active pharmaceutical ingredient, or even the finished product can come from different manufacturing plants and suppliers from all over the world without the need of declaration. This makes it extremely difficult to determine the origin of a drug, although the identification of the manufacturer has a safety aspect for the patient, and also a commercial aspect for the pharmaceutical companies in terms of counterfeiting.

Chemometric methods are a common way to analyze large, complex spectral data and have found their way into various fields [3-12]. The combination of chemometric methods with near-infrared (NIR), mid-infrared (MIR), nuclear magnetic resonance (NMR), and Raman spectroscopy has solved many analytical challenges. It can be used to differentiate between

organically and conventionally grown tomatoes in food chemistry, to evaluate complex samples in metabolomics studies, or to date documents in the forensic science [3-7], to name only a few applications. In the pharmaceutical sector, chemometric methods have also gained importance, for example, as a tool in process analytical technology [8, 9], counterfeit detection, and characterization of drug products or herbal medicines [10-12].

The aim of this study was to identify the origin of drugs and the country of their manufacturing plant, respectively, solely based on measured spectra and the use of principal component analysis (PCA) and linear discriminant analysis (LDA). Since paracetamol formulations are widespread and easy to acquire, they have been used as a model product. Sixty-four paracetamol drug samples were purchased from 52 pharmaceutical companies from all parts of the world. Of these, 56 preparations formulated as tablets were used to generate three data sets using NIR, MIR, and NMR spectroscopy. With the help of PCA, an unsupervised method, the data was screened for patterns, which allows a tracing of the tablets back to their origin. In a second step and with the information gathered from PCA, the spectroscopic data were examined by LDA. Each of the spectral data sets was analyzed individually.

2. Materials and Methods

2.1 Materials

Dimethylsulfoxide (D_6, 99.8%) containing 0.03% (v/v) tetramethylsilane (TMS) and the 507-HP-7 5 mm routine NMR tubes were purchased from Euriso-top (Saarbrücken, Germany).

2.2 Drug samples

All samples (see Table 1) were purchased in local pharmacies or in hospitals. Depending on the country, they were dispensed either in their original packaging or as single blister packs. Each tablet contained 500 mg of paracetamol and had different sizes, colors, or shapes. According to the labelling of some preparations, the types of excipients differed only slightly. The number of tablets contained in the marketed products varied between 4 and 30.

Table 1 Listing of all investigated paracetamol tablets. Samples marked with * were used for the LDA models to predict the manufacturer and land of production. Sample no. 2, 3, 45, 49, 54, 62 and 63 were deleted from the sample collection because of the large differences in the formulation (see "*Spectral experiments*"). (AUS: Australia; DEU: Germany; ITA: Italy; AUT: Austria; THA: Thailand; IDN: Indonesia; BGD: Bangladesh; TZA: Tanzania; CZE: Czech Republic; POL: Poland; GBR: Great Britain; ESP: Spain; RUS: Russia; MNG: Mongolia; CHN: China; HKG: Hong Kong; COL: Colombia; USA: United States of America; PRT: Portugal)

Sample No.	Name	Pharmaceutical Companie	Origin	Quantity
1*	Panadol	GlaxoSmithKline plc.	AUS	12
4*	Paracetamol Tablets	Chemists' Own	AUS	24
5*	Paracetamol	AFT Pharmaceuticals Ltd.	AUS	20
6*	Paracetamol	Priceline	AUS	20
7*	Paracetamol 500	1A Pharma GmbH	DEU	20
8*	Paracetamol ratiopharm	Ratiopharm GmbH	DEU	20
9*	Tachiprina	Angelini ACRAF SpA	ITA	30
10*	Paracetamolo Farmakopea	Farmakopea SpA	ITA	20
11*	Acetamol Adulti	Abiogen Pharma SpA	ITA	20
12*	Paracetamolo Sella	Laboratorio Chimico Farmaceutico "A. SELLA" S.r.l.	ITA	30
13*	Mexalen	Ratiopharm GmbH	AUT	10
14*	Paracetamol Genericon	Genericon Pharma GmbH	AUT	10
16	McXY Para	Millimed Co., Ltd.	THA	10
17*	Paracetamol 500	Kamol	THA	10
18*	Sanmol	P.t. Sanbe Farma	IDN	4
19	Paracetamol	P.t. Bernofarm	IDN	10
20	Paracetmaol	P.t. Phyto Kemo Agung Farma	IDN	10
21	Pamol	P.t. Interbat	IDN	4
22*	Panadol	GlaxoSmithKline plc.	IDN	10
23*	Dumin	P.t. Actavis Indonesia	IDN	10
24	Ace	Square Pharmaceuticals Ltd.	BGD	10
25	Napa	Beximco Pharmaceutials Ltd.	BGD	10
26	Paracetamol	Crescent Pharma Ltd.	BGD	10
27*	Vetocin	Nestor Pharmaceuticals Ltd.	TZA	10
28	Paracetamol	North China Pharmaceutical Co., Ltd.	TZA	10
29	Cetamol	Regal Pharmaceuticals Ltd.	TZA	10
30	Panadol Advance	GlaxoSmithKline plc.	TZA	10
31*	Asmol	Astra Lifecare (India) Pvt. Ltd.	TZA	10
32*	Dolomol	Lincoln Pharmaceuticals Ltd.	TZA	10
33*	Para-Denk 500	DENK PHARMA GmbH & Co. KG	TZA	10
34	Elymol	Elys Chemical Industries Ltd.	TZA	10
35*	Agomol	Agog Pharma Ltd.	TZA	10
36	Diodol	Keko Pharmaceutical Industries Ltd.	TZA	10
37*	Parakant	S Kant Healthcare Ltd.	TZA	10
38*	Paracetamol Dr.Max	Dr. Max Pharma Ltd.	CZE	30
39*	Paralen	Zentiva Group, a.s.	CZE	24
40*	Paracetamol Actavis	Actavis	POL	24
41	Paracetamol Polfa Lodz	Bio-Profil Polska Sp.z o.o / Laboratoria Polfa Łódź Sp. z o.o.	POL	10
42	Paracetamol	BIOFARM Sp. z o.o.	POL	20
43*	Paracetamol	STADA Arzneimittel AG	DEU	20
44*	Paracetamol AL 500	Aliud Pharma GmbH	DEU	20

46*	Paracetamol 500mg elac	Inter Pharm Arzneimittel GmbH	DEU	20
47*	ben-u-ron	bene-Arzneimittel GmbH	DEU	20
48*	Paracetamol Hexal	Hexal AG	DEU	20
50*	Paracetamol	Aspar pharmaceuticals Ltd.	GBR	16
51*	Panadol Advance	GlaxoSmithKline plc.	GBR	16
52*	Paracetamol Winthrop	sanofi-aventis, S.A.	ESP	20
53*	Antidol	Laboratorios Cinfa S.A.	ESP	20
55*	Paracetamol	Renewal JSC	RUS	20
56*	Paracetamol	pharmstandard JSC	RUS	20
57*	Paracetamol	Nakhia Impex LLC	MNG	20
58	Panadol	tskf Co., Ltd.	CHN	10
59*	Panadol ActiFast	GlaxoSmithKline plc.	HKG	16
60	Acetaminofen	Laproff S.A.	COL	10
61	Tylenol	Johnson & Johnson Services, Inc.	USA	10
64*	Paracetamol Farmoz	Farmoz - Sociedade Técnico Medicinal, S.A.	PRT	20

2.3 Sample preparation

For NIR and MIR measurements, the tablets were mortared and measured directly. For NMR experiments, an additional tablet per sample was mortared and placed in a falcon tube. Then, 6 mL of DMSO-d_6 containing 0.03% TMS (v/v) as reference standard was added. The samples were vortexed (1 min), sonicated (1 h), and centrifuged (20 min/6k U/min). Six aliquots of the supernatant were analyzed by NMR spectroscopy (600 µL each).

Due to the small number of available tablets of some samples, therefore usually only the availability of one batch and the need for a whole tablet to obtain reproducible NMR spectra, only one tablet per company could be measured. For these reasons, an additional batch of the German preparations was acquired, measured, and compared with the first batch. It could be seen that the difference between the batches is very small for most companies, and therefore, the choice of sample preparation is acceptable, even if the validation is thus only valid for repeat measurements.

2.4 Spectral experiments and analysis

The acquisition parameters for NMR, NIR, and MIR spectroscopy measurements were already reported in Belugina R. B. et al. [13].

2.4.1 NMR spectroscopy

All samples were analyzed with a Bruker Avance III 400 MHz spectrometer operating at 400.13 MHz with an inverse probehead. The ^{1}H NMR experiments were measured at 300.11 ± 0.03 K with a 90° flip angle, 64 scans, no rotation, and an acquisition time of 5.45 s followed by a relaxation delay of 12 s. The receiver gain was set to 14.04 and a line broadening factor of 0.3 Hz was applied. The resulting digital resolution was 0.183 Hz over a spectral width

of 30.04 ppm (time domain size 128k). Phasing and baseline correction were performed manually with TopSpin versions 3.5 and 4.0 (Bruker BioSpin GmbH, Rheinstetten, Germany). All signals were referred to the TMS signal. Each sample was measured six times.

2.4.2 NIR spectroscopy

Reflectance spectra were performed on a MicroNIR™ 1700 ES spectrometer with a windowed collar (VIAVI Solutions Deutschland GmbH, Eningen unter Achalm, Germany) covering a spectral range of 950–1650 nm. It works with two tungsten lamps and detection was performed with a photodiode array detector. The drug samples were measured by placing the glass vials on the windowed collar and rotating them after every measurement. Twelve spectra were recorded per sample with an average of 12 scans and an integration time of 12.2 s.

2.4.3 MIR spectroscopy

An FT/IR-6100 spectrometer (JASCO Deutschland GmbH, Pfungstadt, Germany) equipped with an attenuated total reflectance unit was used to acquire the MIR spectra. Twelve spectra of every drug sample were measured in a spectral range of 4000–550 cm^{-1} with 256 scans per spectrum and a resolution of 4 cm^{-1}.

2.5 Pre-processing and multivariate data analysis

For multivariate analysis, NMR spectra were reduced by bundling spectral regions of equal width of 0.04 ppm using Amix 3.9.15 (Bruker BioSpin GmbH, Rheinstetten, Germany). The spectral range from 0 to 11 ppm was used for further examination with PCA. The spectral regions of the residual water signal from 3.42 to 3.50 ppm, the residual dimethylsulfoxide signal from 2.34 to 2.70 ppm, and the TMS signal from −0.06 to 0.06 ppm were excluded. The final range used for PCA was 6.42–3.54 ppm, 3.38–2.74 ppm, 2.30–2.18 ppm, and 1.78–0.50 ppm.

Before further processing, the MIR-transmission spectra were transformed into absorption spectra and a baseline correction was applied. To remove scatter effects or compensate for additive effects from MIR and NIR data, an extended multiplicative scatter correction (EMSC) and a standard normal variate transformation (SNV) were applied, respectively. The first derivative was performed for both spectral data sets.

The final spectral range of interest was limited to 1175.401 to 861.0605 cm^{-1} for MIR and 1100.125–1242.595 nm and 1347.899–1570.896 nm for NIR spectra. All pre-processing methods and the individual analysis of the three spectral methods with PCA and LDA were performed with the Unscrambler X 10.4 (CAMO Software AS., Oslo, Norway). The permutation tests and data fusion analysis were performed with MATLAB 2016a (The MathWorks, Natick, MA). For the LDA, the prior probabilities were assumed to be equal and it was performed with the PCA scores due to the high number of variables [14]. The optimal number of scores was evaluated individually by comparing the results and accuracy of several LDA models, where

the scores were successively reduced. For the LDA models of the NIR and MIR data, six or seven components must be used to obtain sufficient accuracy. Depending on the goal to determine the manufacturer or the pharmaceutical company, four or three components were sufficient to create the model with the NMR data. For the examination with PCA and LDA, not all variables were used, but only those of the final ranges defined above. The prediction performance of the LDA models was tested with a custom cross-validation for repeated measurements.

3. Results and discussion

3.1 Sample information

For most of the drugs, the information like their manufacturer or composition was noted on the primary or secondary packaging or in the package insert. For the remaining drugs, the websites of the authorization holders or the relevant national authorities were screened for further information about the manufacturer and the samples itself. As far as possible, information was also collected on the legal and business relations between the companies.

3.2 Spectral experiments

Since paracetamol was the active pharmaceutical ingredient in each drug sample, the formulations differ mainly with regard to excipients and their amount. Because mainly tablets were available and the other formulations differed too much from them, only tablets with the same paracetamol content were considered (56 out of 64 samples). The other samples were discarded from the sample collection. However, the mass fraction of the API in the tablets ranged between 74 and 95% (w/w), which resulted in the paracetamol signals being the dominant part in the spectra. Exceptions were Paracetamol Polfa Lodz® (62%) and Panadol Acti Fast® (38%). The focus of the first steps was to identify the spectral fingerprints of each sample in every method.

For multivariate data analysis, it is mandatory to have a sample preparation that generates reproducible spectra. Due to the variety of excipients in the tablets, e.g., large organic molecules alongside small inorganic molecules, it was difficult to find an appropriate solvent for NMR spectroscopy. DMSO was chosen because it was able to dissolve most of the excipients. Nevertheless, a residue often remained, which was centrifuged off. The sample preparation was identical for every tablet. Due to reproducibility issues caused by the one-sided ratio between paracetamol and excipients, it was necessary to measure the whole tablet and not just an aliquot. With the help of reference spectra, it was possible to identify the signals of paracetamol and of the excipients. The tablets mainly contained on average small amounts of cellulose derivatives, a type of starch (mostly maize starch), silica, stearic acid,

povidone, and talcum. The spectral range in which the signals of the excipients appear was similar for all drug samples (0.5-6 ppm) and was therefore used for further PCA and LDA (reduced NMR spectra of tablets, see Figure 1).

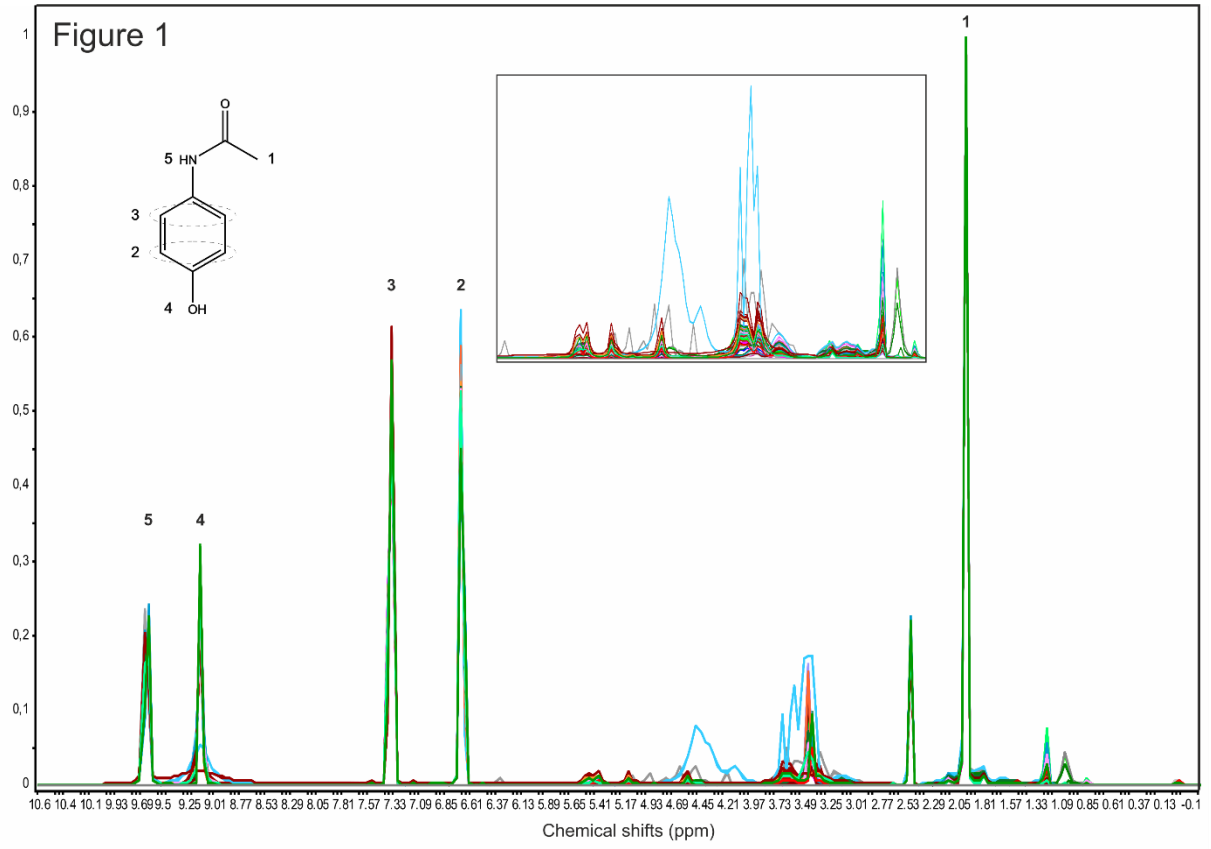


Figure 1 Reduced ^{1}H NMR spectra of all paracetamol tablets with enlarged spectral range of the excipients (without residue signals of water, dimethylsulfoxide, and tetramethylsilane). The corresponding paracetamol signals are additionally marked. The spectra are color-coded according to the pharmaceutical companies.

MIR and NIR spectra (see Figure 2 and 3) of the samples were very similar due to the high percentage of paracetamol. As expected, the only exceptions were Paracetamol Polfa Lodz (62% w/w) and Panadol ActiFast (38% w/w). This can be explained with their different formulations: they contain an additional amount of 170 mg sorbitol and 630 mg sodium hydroxycarbonate, respectively, which leads to the significant different spectra.

Because a specific device for the NIR instrument to directly measure the tablets was not available, the tablets had to be mortared. Physical information such as particle size or compression force can disappear as a result of this preparation step. Similar limitations hold true for the MIR spectra. However, after suitable pre-processing, a spectral range was found for both methods in which the spectra of the tablets differ.

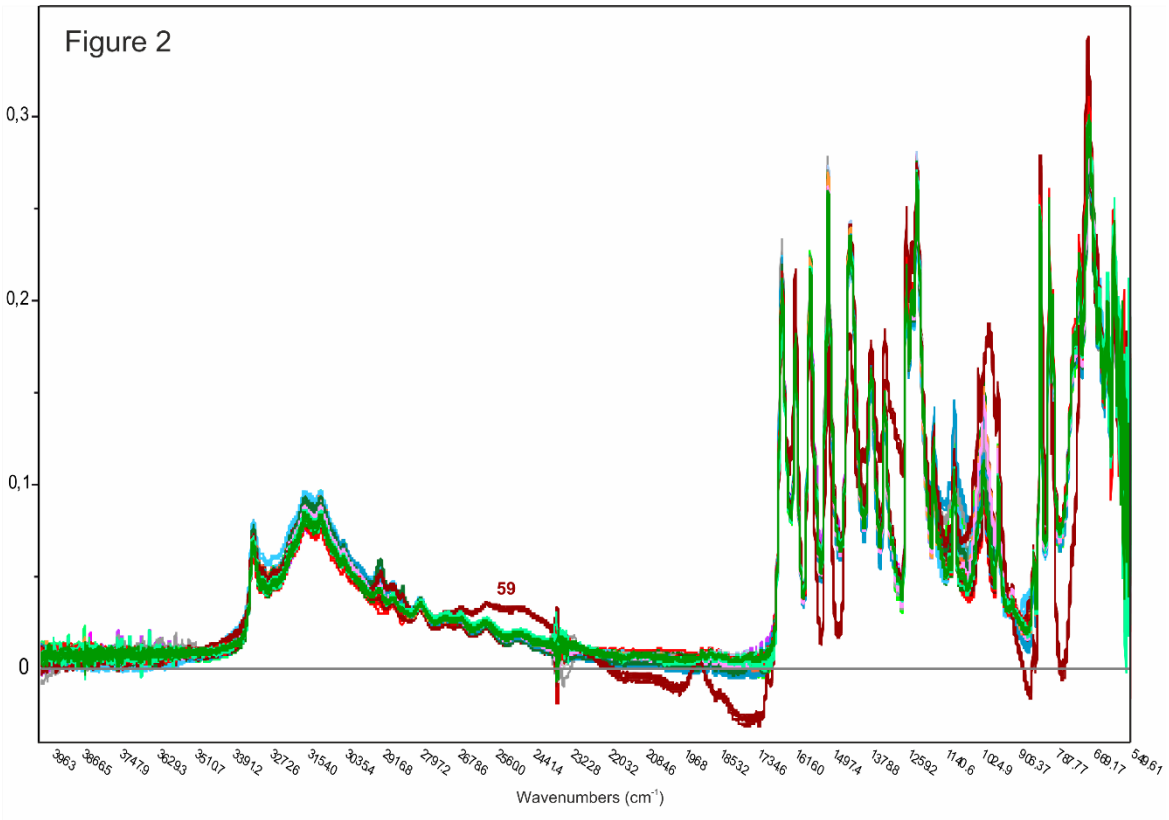


Figure 2 MIR spectra after transformation into absorbance spectra, baseline, and EMSC correction. Grouped by color according to the pharmaceutical companies. The marked spectrum is Panadol ActiFast (sample no. 59), which differs from the other spectra due to the additional high excipient content of sodium hydroxycarbonate.

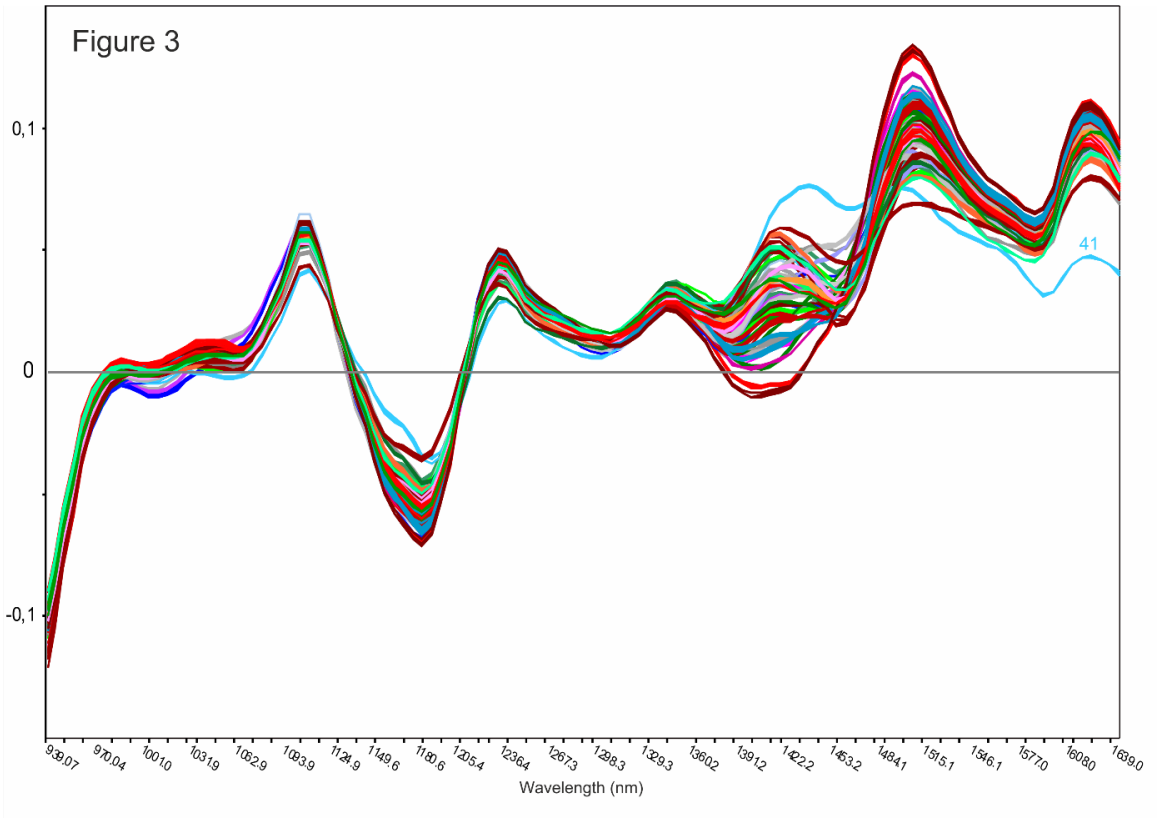


Figure 3 SNV-corrected NIR spectra (1st derivative). Grouped by color according to the pharmaceutical companies. The marked spectrum is Paracetamol Polfa Lodz (sample no. 41), which differs from the other spectra due to the additional high excipient content of sorbitol.

3.3 Principal component analysis

The main intention of multivariate data analysis is the extraction of useful information from the experimental data and revealing hidden relations as well as the reduction of the dimensionality of the spectral data to so-called principal components or latent variables [15]. One commonly used multivariate analysis technique is the principal component analysis (PCA). It is a projection method, which narrows the data dimensionality of several hundred or even thousand spectral values down to a few principal components and grants a better visualization of the data with appropriate plots. This allows for a better identification of the crucial spectral range, outliers, and clusters [16]. By means of the scores plot, in which the principal components are plotted against each other, the data sets were examined for clusters.

In a first review of the scores plots, two drug products were very noticeable. Depending on the spectroscopic method used, at least one of them was always clearly different from the other samples. These samples were Paracetamol Polfa Lodz and Panadol ActiFast. This was due to the high proportion of additional excipients, as already mentioned. For this reason, these samples had to be removed from the MIR data set to allow for a better evaluation of the other tablets, as they distorted the PCA too much.

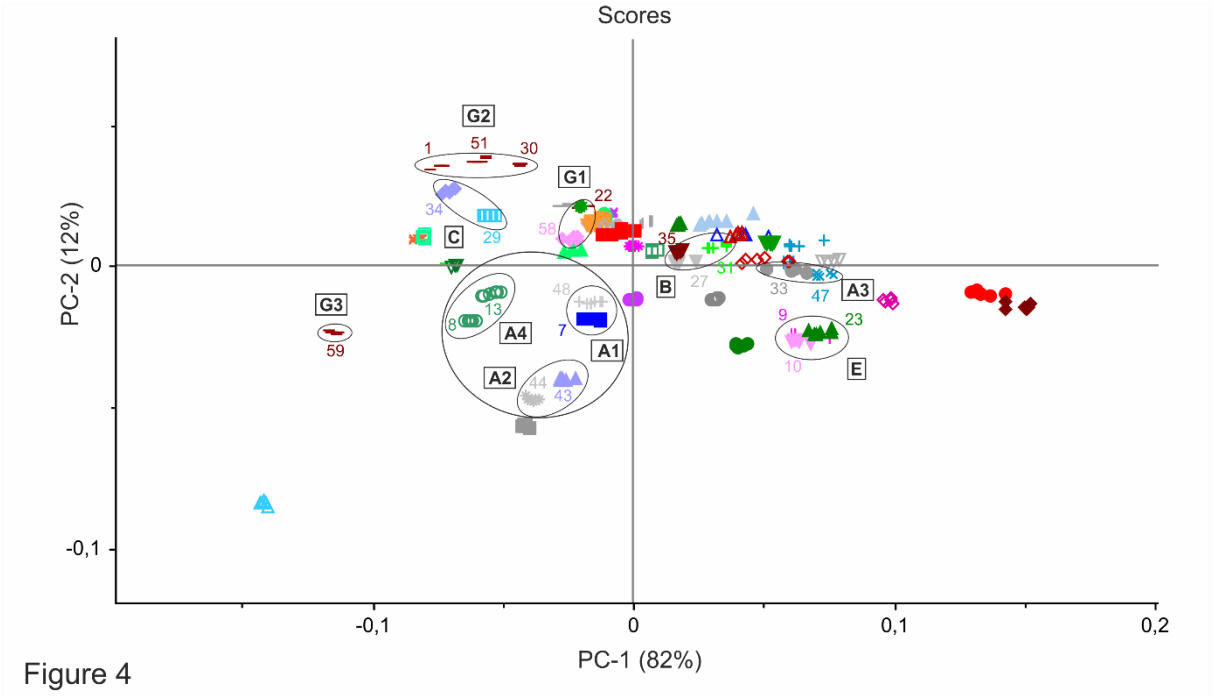


Figure 4 PCA of NIR spectra: 2D scatterplot with different colored symbols for each pharmaceutical company with the corresponding clusters as mentioned in "*Principal component analysis*" The colored numbers correspond to the sample numbers from Table 1.

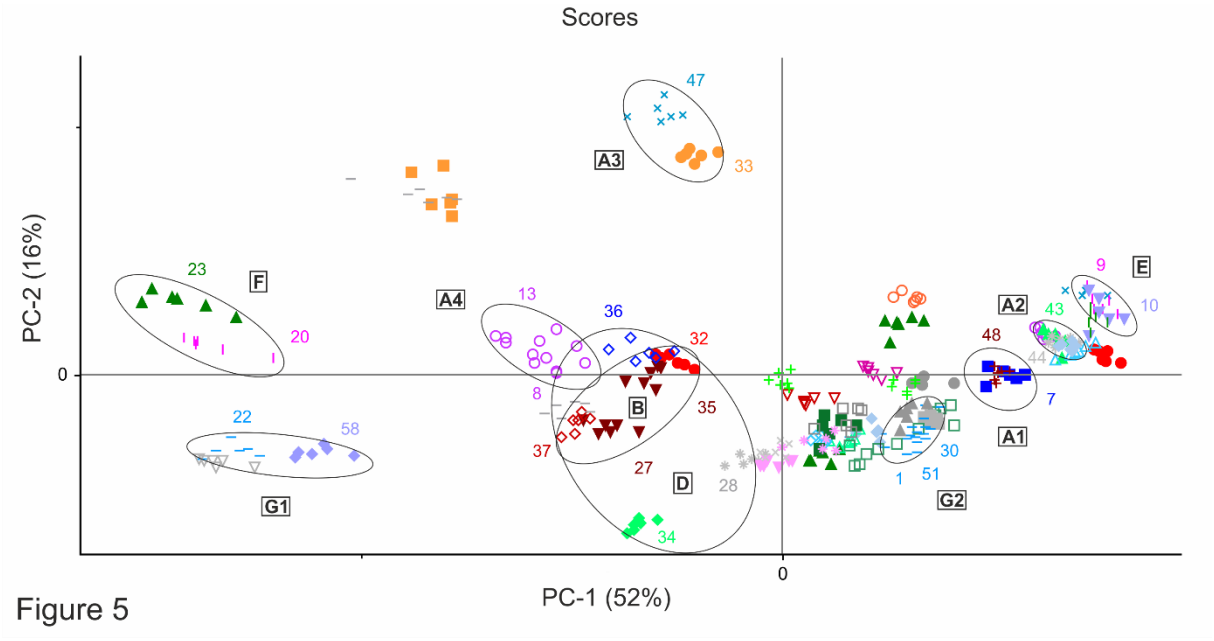


Figure 5

Figure 5 PCA of MIR spectra: 2D scatterplot with different colored symbols for each pharmaceutical company with the corresponding clusters as mentioned in "*Principal component analysis*" The colored numbers correspond to the sample numbers from Table 1.

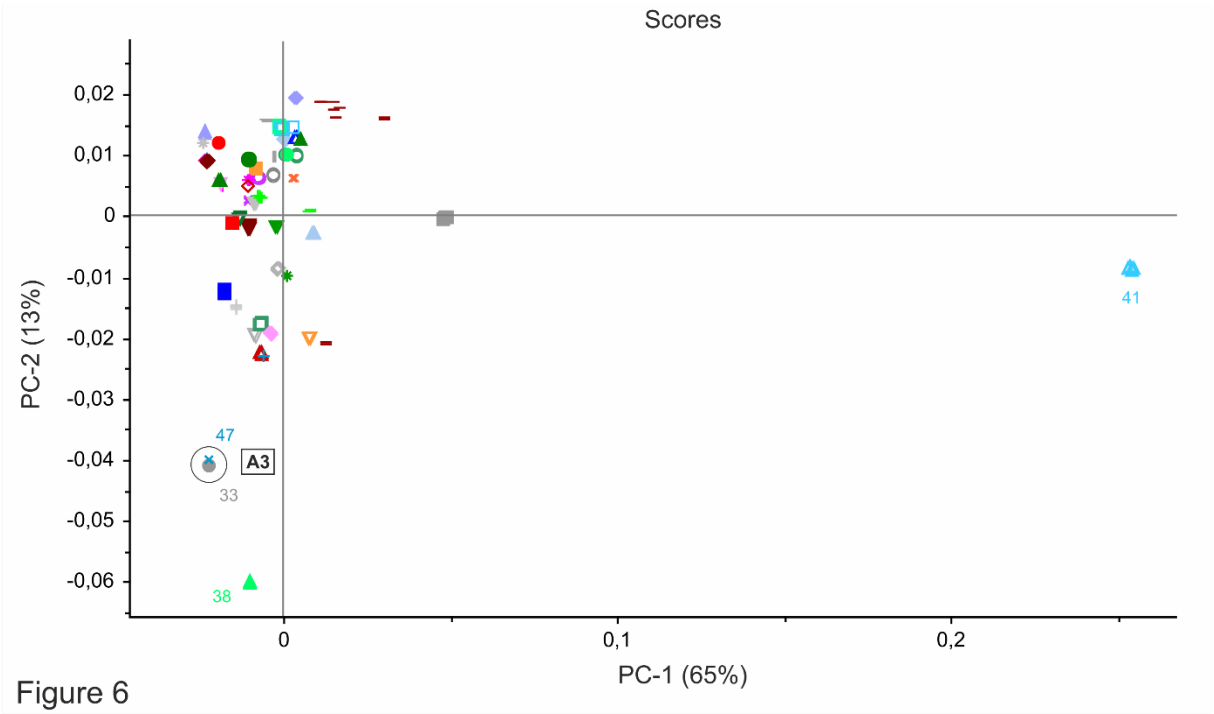


Figure 6

Figure 6 PCA of ^{1}H NMR spectra: 2D scatterplot with different colored symbols for each pharmaceutical company with the corresponding clusters as mentioned in "*Principal component analysis*" The colored numbers correspond to the sample numbers from Table 1.

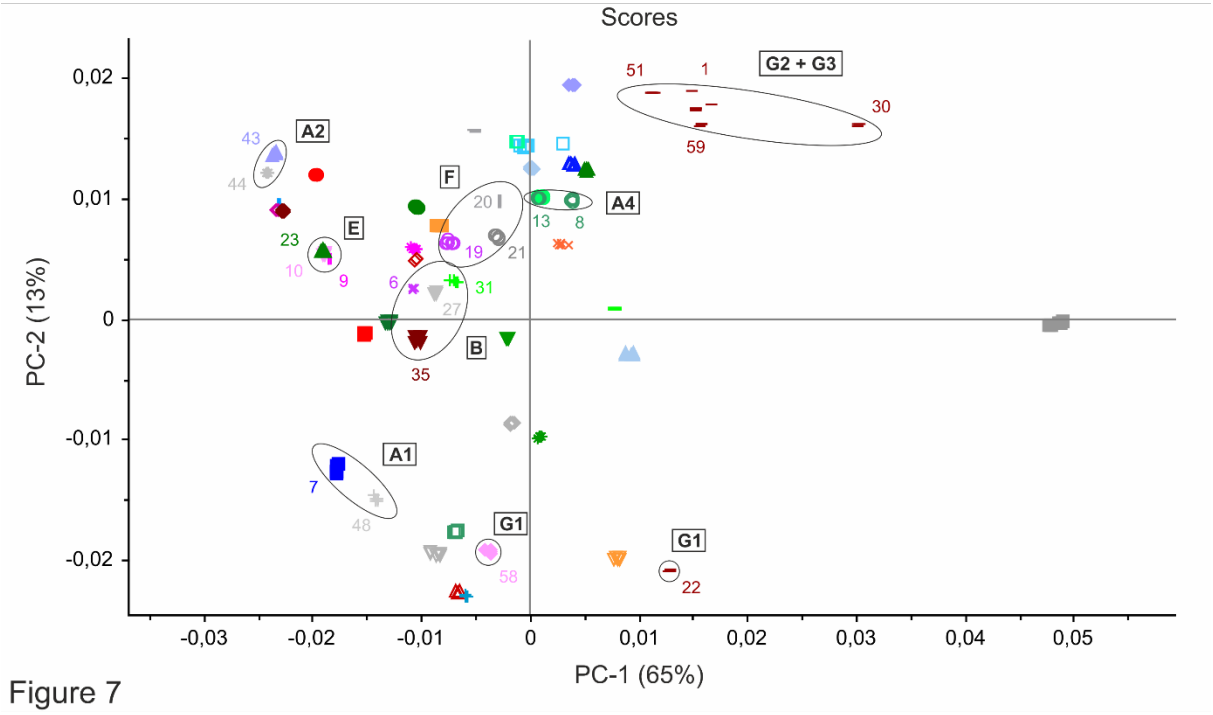


Figure 7

Figure 7 PCA of ^{1}H NMR spectra: detail of the 2D scatterplot of Figure 6 with the corresponding clusters as mentioned in "*Principal component analysis*" Each colored symbol represents a pharmaceutical company and the colored numbers indicate the corresponding sample number (see Table 1).

For the remaining samples, cluster formation was observed for all three spectral methods (Figure 4, 5, 6 and 7). The clusters were defined by the same manufacturer or country of origin and production, respectively. The analysis of the scores plots showed that some clusters were more differentiated from the other samples depending on the spectroscopic method. This was especially true for the MIR and NIR data. However, most of the clusters found in these plots were clearly separated from all other samples. Yet there was an accumulation of several samples, which could not be completely parted even in the higher PCs and therefore could not be assigned to specific clusters. Nevertheless, almost all of the determined clusters were found in all three scores plots, with the exception of two (clusters C and D, see below). These could only be seen using the NIR data (see Figure 4). The identified clusters are listed below with the corresponding sample numbers in parentheses:

- Clusters of samples produced by the same manufacturer (Figure 4, 5, 6 and 7): cluster A1: Hexal AG (48) and 1A Pharma GmbH (7)/cluster A2: STADA Arzneimittel AG (43) and Aliud Pharma GmbH (44)/cluster A3: bene-Arzneimittel GmbH (47) and Denk Pharma GmbH & Co. KG (33)/cluster A4: two drugs of Ratiopharm GmbH (from Germany (8) and Austria (13))
- Cluster B (bought in Tanzania but manufactured in India): Nestor Pharm. Lim. (27), Agog Pharma Ltd. (35), Lincoln Pharma (32), and S Kant Healthcare (37) (MIR; Figure 5); Nestor Pharm. Lim., Agog Pharma Ltd., Astra Lifecare (India) Pvt. Ltd. (31)

(NIR and NMR; Figure 4 and 7). In addition, the NMR spectra cluster included a Priceline preparation (6), which originates from Australia, and also was manufactured in India.

- Cluster C (bought in Tanzania but manufactured in Kenya): Elys Chemical Industries Ltd. (34) and Regal Pharm. Ltd. (29) (NIR; Figure 4)
- Cluster D (bought in Tanzania): Nestor Pharm. Lim., S Kant Healthcare, Agog Pharma Ltd., Lincoln Pharma, and Keko Pharm. Industries Ltd. (36). The samples from Elys Chemical Industries Ltd. and North China Pharm. (28) were a little further away, but still in the vicinity (MIR; Figure 5)
- Cluster E (bought and manufactured in Italy): Angelini ACRAF SpA (9) and Farmakopea SpA (10) (MIR; Figure 5). In the evaluation of the NIR and NMR spectra, PT. Actavis Indonesia (23) was additionally present (Figure 4 and 7). Further investigations have shown that Farmakopea SpA belongs to the Unifarm group (subsidiary E-Pharma), based in Italy [17]. This group in turn produces pharmaceutical products for many companies, including Actavis and Angelini [18].
- Cluster F (bought and manufactured in Indonesia): PT. Actavis Indonesia and P.t. Phyto Kemo Agung Farma (20) (MIR; Figure 5); PT. Interbat Pharmaceutical Industry (21), P.t. Phyto Kemo Agung Farma and Bernofarm (19) or PT. Interbat Pharmaceutical Industry and P.t. Sanbe Farma or PT. Actavis Indonesia and Bernofarm (NMR; depending on the PCs); PT. Interbat Pharmaceutical Industry and Bernofarm or PT. Interbat Pharmaceutical Industry and P.t. Sanbe Farma (PT. Actavis Indonesia and Bernofarm only in the vicinity) (NIR; higher PCs)
- Clusters of Panadol® preparations from GlaxoSmithKline (GSK): The sample set contained six different tablets of this brand, five from GSK and one from tskf. The latter is a joint venture between GSK and other pharmaceutical companies [19]. The preparations most likely came from four different manufacturers from Ireland (30, 51, 59), Australia (1), China (58), and Indonesia (22) and correspond to three different formulations: Panadol® (cluster G1; 22, 58), Panadol® Advance/Optizorb (cluster G2; 1, 30, 51), and Panadol® ActiFast (cluster G3; 59). These samples formed three clusters in the scores plot of the MIR and NIR spectra according to their formulation (Figure 3 and 7). The PCA of the NMR spectra mainly differentiated between the standard formulation and those with a modified drug formulation (Figure 7). Furthermore, it was also possible to distinguish between the manufacturing sites of the two standard preparations. A complete differentiation of all GSK samples according to their manufacturing site could be achieved by an individual PCA of the NIR and NMR spectra (for the corresponding scores plots, see Figures S1–S4 in the Supplementary

Information (ESM)). The information about the different tablet formulations was contained in the first principal components, whereas the information about the plant of manufacture was in the higher ones.

All in all, the three methods have produced valuable results, and in some cases, they have complemented each other. However, the main difference between them was the effort required for sample preparation. For the infrared spectra, the tablets had only to be mortared, whereas for the NMR spectra, first a suitable sample preparation and measurement method had to be established.

However, in the PCA of NMR spectra, it was possible to quickly and easily identify which excipients correlate significantly with the main components by means of the loadings (see Figure S5 in the ESM for the loadings plots). From this, it can be seen which excipients are characteristic for the individual preparations. Due to the high proportion of sorbitol in Paracetamol Polfa Lodz® compared to the other excipients in the remaining preparations, the loadings of PC1 correspond to the signals of sorbitol. This also explains why PC1 mainly describes Polfa Lodz®. The second PC is determined by magnesium stearate, which appears to be present in large quantities at Dr. Max Pharma Ltd. (38), bene-Arzneimittel GmbH (47), and DENK PHARMA GmbH & Co. KG (33). PC3 correlates mainly with the starch derivatives, but to a certain extent it has additionally a positive correlation with lactose and a negative one with sorbitol and hydroxypropyl cellulose or hydroxypropyl(methyl)cellulose. PC4, on the other hand, has a strong positive correlation with the cellulose derivatives mentioned, which are contained in higher amounts in the preparations P.t. Sanbe Farma (18), PT. Interbat Pharmaceutical Industry (21), and Actavis (40). In the last PC, povidone has the highest weight, whereas this excipient is contained in almost all preparations and therefore contributes less to the differentiation of the drug samples.

3.4 Linear discriminant analysis

In contrast to PCA, linear discriminant analysis (LDA) is a supervised method. The samples are first assigned to individual groups [20] before the parameters of the discriminant function are chosen in such a way that the differences within a group become minimal and maximal to the others [21]. The resulting model can then be used to make predictions about unknown samples and assign them to one of the previously defined groups.

First, different LDA models were created and compared to each other to test which categories are well predictable. Attempts were made to predict the pharmaceutical company, the manufacturer, or the country of production or origin, respectively. For the categories manufacturer and country of production, only those samples were used, for which this information could be verified by means of the blister or the package leaflet or via the website of the pharmaceutical company (s. Table 1). In addition, as in the PCA, the sample Panadol

ActiFast was removed from the MIR data set for the determination of the pharmaceutical companies because its formulation is very different in comparison to the other tablets. This led to a much better classification for the remaining companies.

To ensure that the models are suitable for their intended purpose, a custom cross-validation was carried out to get an impression of the performance of the model. Therefore, the data sets were divided into six different test sets and model-building sets. For MIR and NIR data, eight spectra were used for the model-building set and two for the validation set. To get six different sets, the two spectra for validation testing were switched with two other spectra from the model-building set. The selection was limited to 10 spectra, since PCA could identify one or two outliers in some samples. This concerned the following sample numbers: NIR: 16, 31, 39, 52; MIR: 11, 38, 41, 55 (52, 55 with two outliers, the rest with one). For NMR data, five spectra were used to build the model and one for the prediction, always using a different spectrum for the prediction. The average percentage of correctly assigned spectra for every model is listed in Table 2.

Table 2 LDA results (average number of correctly assigned spectra in percent after six LDAs) of different spectroscopic methods in relation to two different categories: manufacturers (35 different; 40 samples) and pharmaceutical companies (50 different; 56 samples). For MIR Pharmaceutical companies only 55 samples were used, as described in "*Linear discriminant analysis*"

	MIR	**NIR**	**^{1}H NMR**
Manufacturers (n = 40)	91 (LDA1)	100 (LDA3)	99 (LDA5)
Pharmaceutical Companies (n = 56)	89 (LDA2)	99 (LDA4)	99 (LDA6)

It was soon apparent that it is hardly possible to determine the country in which the preparations were acquired or in which they were produced. In some cases, far less than 60% correct assignments for MIR and NMR models were achieved. Only the NIR model was able to make between 60 and 70% correct classifications. This is not surprising, however, since the drugs are manufactured according to the companies' specifications and not those of the countries. The fact that in PCA within some countries some preparations of different manufacturers are nevertheless very similar could be due to the local suppliers of excipients and the low variability in the compositions. For this reason, we have focused more on the other two categories being pharmaceutical company and manufacturer.

For the manufacturer and the pharmaceutical company, very good results were achieved (see Table 2), and therefore, an additional permutation test was performed [22]. This is a randomization test to check whether the chosen descriptors, like wavenumbers, are truly correlated to the response variable and does not lead to a correct selection just by chance [23]. For this purpose, the assignment of manufacturers or pharmaceutical companies to the spectra was scrambled and the percentage of the correct classification was compared to the original assignment. As can be seen in Table 3, the original correct classification rate is significantly higher for all three data sets than the one after scrambling.

Table 3 Results of permutation tests for three data sets regarding manufacturers and pharmaceutical companies. The percentage of correct classification is shown.

	MIR	NIR	^{1}H NMR
Manufacturers	59	69	74
Pharmaceutical Companies	71	85	93

As can be seen in Table 2, the NMR and NIR methods were most capable of providing a correct classification. The LDA models based on the NIR data were able to correctly assign the spectra to both the manufacturer (LDA3) and the pharmaceutical companies (LDA4). Among the pharmaceutical companies, only Biofarm (42) had three wrong assignments for the 6 LDAs, as two wrong assignments for Hexal AG (48) and GlaxoSmithKline (1). The NMR models achieved very similarly results. Only the model for the determination of the pharmaceutical companies could not correctly determine three samples with any of the models. These were two samples from GlaxoSmithKline (22, 51) and one from Actavis (23). The MIR models (LDA1; LDA2) gave with around 90% correct classified spectra for both categories a little worse result for the classification. The biggest problems in the assignment of the pharmaceutical manufacturer (LDA2) were found in the spectra of GlaxoSmithKline (22), Farmoz (64), and Johnson&Johnson (61). Furthermore, the method was not able to distinguish between Aliud Pharma (44) and STADA (43), as well as Angelini (9) and Farmakopea (10). The latter also caused problems with the model for determining the manufacturer (LDA1), as did another sample from GlaxoSmithKline (51) and Farmoz (64).

A closer look revealed the reason of these incorrect assignments. The tablets of Aliud Pharma and STADA are both produced by STADA, which is why mix-ups occurred when determining the pharmaceutical company. Farmakopea and Angelini also seem to be connected via the Unifarm Group, as clarified in "*Principal component analysis*" – Cluster E. The problems of the GSK samples are due to the different formulations, which is why the models have problems assigning them to the same manufacturer. However, by further subdividing the pharmaceutical companies into manufacturers, the results can be improved and the models are better able to make classifications as seen for all spectral data sets. Overall LDA3 showed the best results. It was able to differentiate between all production sites of GSK, despite different formulations, and due to the clarification of the manufacturers, no more mix-ups occurred.

4. Conclusion

It has been shown that chemometric evaluation of mid-infrared (MIR), near-infrared (NIR), and nuclear magnetic resonance (NMR) spectra using principal component analysis (PCA) and linear discriminant analysis (LDA) can be very useful in characterizing drugs and determining their origin. It was possible to identify relationships between companies and suppliers and to detect major differences or similarities in formulations. In addition, most of the samples could

be assigned to their manufacturer or pharmaceutical company. Nevertheless, some points must be taken into account. This includes cooperations or mergers of companies, different production sites, or different formulations of the pharmaceutical companies. However, as is often the case with the manufacturer itself, this information is not listed or is difficult to retrieve and can lead to falsified correct or incorrect classifications.

The country of manufacture or country of origin could not be determined exactly by means of LDA, as this depends on too many factors, above all the fact that the drugs are of course not produced according to the specifications of the countries but of the respective companies. Added to this is the low variability in the formulations as well as the worldwide marketing of excipients that are rarely purchased locally or only by one company. However, the PCA shows that there may be some similarities between products from the same country. If an unknown sample is projected onto the PCA and it is inside the borders of a particular cluster (at a given probability), it can be assumed a new sample may also originate from that country.

When comparing the spectroscopic methods, NIR and NMR are preferred. With NIR, the sample preparation is very easy, the acquisition of the spectra is very fast, and the results are valid. With NMR, on the other hand, the sample preparation is more difficult but the classifications led to almost the same results, especially for the manufacturers. Furthermore, it is possible to obtain information about the composition of the different samples, allowing them to be characterized and then compared with other or unknown samples. However, it was shown that the information of all three methods can complement each other and that there is a benefit in using and analyzing with different spectral methods. Therefore, a further analysis using data fusion, where the spectral information of all methods is combined and analyzed simultaneously, would be beneficial. Our preliminary studies have shown that data fusion approach, namely, common components and specific weights analysis (CCSWA) [24, 25], can be used to differentiate paracetamol producer and marketing authorization holder (MAH). In this case, the percentage of correct classification varied between 93 and 96%. Similarly to the findings described in this study, NMR was proven to be the best method to detect paracetamol origin; the other data sets tend to worsen the model.

Since there is no complete disclosure of the pharmaceutical companies about the origin of the tablets or excipients, only assumptions can be made about some relations. If there were a better traceability, it should be possible to make even more precise statements about the origin with this method.

Funding

Open Access funding enabled and organized by Projekt DEAL. Y. Monakhova received support from the Russian Science Foundation (project 18-73-10009).

Conflict of interest

The authors declare no competing interests.

References

[1] Cockburn R, Newton PN, Agyarko EK, Akunyili D, White NJ The Global Threat of Counterfeit Drugs: Why Industry and Governments Must Communicate the Dangers. *PLoS Med* (2005) 2, e100

[2] Aldhous P Counterfeit pharmaceuticals: murder by medicine. *Nature* (2005) 434, 132-136

[3] Carolina S. Silva MFP, José Manuel Amigo, Carmen Garcia-Ruizc, Fernando Ortega-Ojedac Infrared spectroscopy and chemometrics to evaluate paper variability in document dating. *Spectrosc Eur* (2018) 30, 12 - 15

[4] Gu H, Pan Z, Xi B, Hainline BE, Shanaiah N, Asiago V, Gowda GAN, Raftery D 1H NMR metabolomics study of age profiling in children. *NMR Biomed* (2009) 22, 826-833

[5] Hohmann M, Monakhova Y, Erich S, Christoph N, Wachter H, Holzgrabe U Differentiation of Organically and Conventionally Grown Tomatoes by Chemometric Analysis of Combined Data from Proton Nuclear Magnetic Resonance and Mid-infrared Spectroscopy and Stable Isotope Analysis. *J Agric Food Chem* (2015) 63, 9666-9675

[6] Hohmann M, Christoph N, Wachter H, Holzgrabe U 1H NMR profiling as an approach to differentiate conventionally and organically grown tomatoes. *J Agric Food Chem* (2014) 62, 8530-8540

[7] Mao H, Wang H, Wang B, Liu X, Gao H, Xu M, Zhao H, Deng X, Lin D Systemic Metabolic Changes of Traumatic Critically Ill Patients Revealed by an NMR-Based Metabonomic Approach. *J Proteome Res* (2009) 8, 5423-5430

[8] Challa S, Potumarthi R Chemometrics-based process analytical technology (PAT) tools: applications and adaptation in pharmaceutical and biopharmaceutical industries. *Appl Biochem Biotechnol* (2013) 169, 66-76

[9] Matero S, van Den Berg F, Poutiainen S, Rantanen J, Pajander J Towards better process understanding: chemometrics and multivariate measurements in manufacturing of solid dosage forms. *J Pharm Sci* (2013) 102, 1385-1403

[10] Custers D, Cauwenbergh T, Bothy JL, Courselle P, De Beer JO, Apers S, Deconinck E ATR-FTIR spectroscopy and chemometrics: An interesting tool to discriminate and characterize counterfeit medicines. *J Pharm Biomed Anal* (2015) 112, 181-189

[11] Li L, Zang H, Li J, Chen D, Li T, Wang F Identification of anisodamine tablets by Raman and near-infrared spectroscopy with chemometrics. *Spectrochim Acta A Mol Biomol Spectrosc* (2014) 127, 91-97

[12] Said MM, Gibbons S, Moffat AC, Zloh M Near-infrared spectroscopy (NIRS) and chemometric analysis of Malaysian and UK paracetamol tablets: a spectral database study. *Int J Pharm* (2011) 415, 102-109

[13] Belugina RB, Monakhova YB, Rubtsova E, Becht A, Schollmayer C, Holzgrabe U, Legin AV, Kirsanov DO Distinguishing paracetamol formulations: Comparison of potentiometric “Electronic Tongue” with established analytical techniques. *J Pharm Biomed Anal* (2020) 188, 113457

[14] Bertrand D, Courcoux P, Autran J-C, Meritan R, Robert P Stepwise canonical discriminant analysis of continuous digitalized signals: Application to chromatograms of wheat proteins. *J Chemom* (1990) 4, 413-427

[15] Schönberger T, Monakhova, Y. B., Lachenmeier, D.W., Walch, S., Kuballa, T., ., et al. Guide to NMR Method Development and Validation – Part II: Multivariate data analysis. *Eurolab Technical Report No 01/2015* (2015)

[16] Biancolillo A, Marini F Chemometric Methods for Spectroscopy-Based Pharmaceutical Analysis. *Front Chem* (2018) 6

[17] Homepage of Unifarm S.p.A. https://www.unifarm.it/it/gruppo/. Accessed 15.04.2020

[18] Business partners of E-Pharma. https://www.e-pharma.com/en/health/partners. Accessed 15.04.2020

[19] Homepage of GlaxoSmithKline China. https://www.gsk-china.com/en-gb/about-us/gsk-in-china/. Accessed 21.04.2020

[20] Cozzolino D, Chree A, Scaife JR, Murray I Usefulness of Near-Infrared Reflectance (NIR) Spectroscopy and Chemometrics To Discriminate Fishmeal Batches Made with Different Fish Species. *J Agric Food Chem* (2005) 53, 4459-4463

[21] Martinez AM, Kak AC PCA versus LDA. *IEEE Trans Pattern Anal Mach Intell* (2001) 23, 228-233

[22] Lindgren F, Hansen B, Karcher W, Sjöström M, Eriksson L Model validation by permutation tests: Applications to variable selection. *J Chemom* (1996) 10, 521-532

[23] Rücker C, Rücker G, Meringer M y-Randomization and Its Variants in QSPR/QSAR. *J Chem Inf Model* (2007) 47, 2345-2357

[24] Monakhova YB, Hohmann M, Christoph N, Wachter H, Rutledge DN Improved classification of fused data: Synergetic effect of partial least squares discriminant analysis (PLS-DA) and common components and specific weights analysis (CCSWA) combination as applied to tomato profiles (NMR, IR and IRMS). *Chemometrics Intellig Lab Syst* (2016) 156, 1-6

[25] Qannari EM, Wakeling I, Courcoux P, MacFie HJH Defining the underlying sensory dimensions. *Food Qual Prefer* (2000) 11, 151-154

Supplementary Material

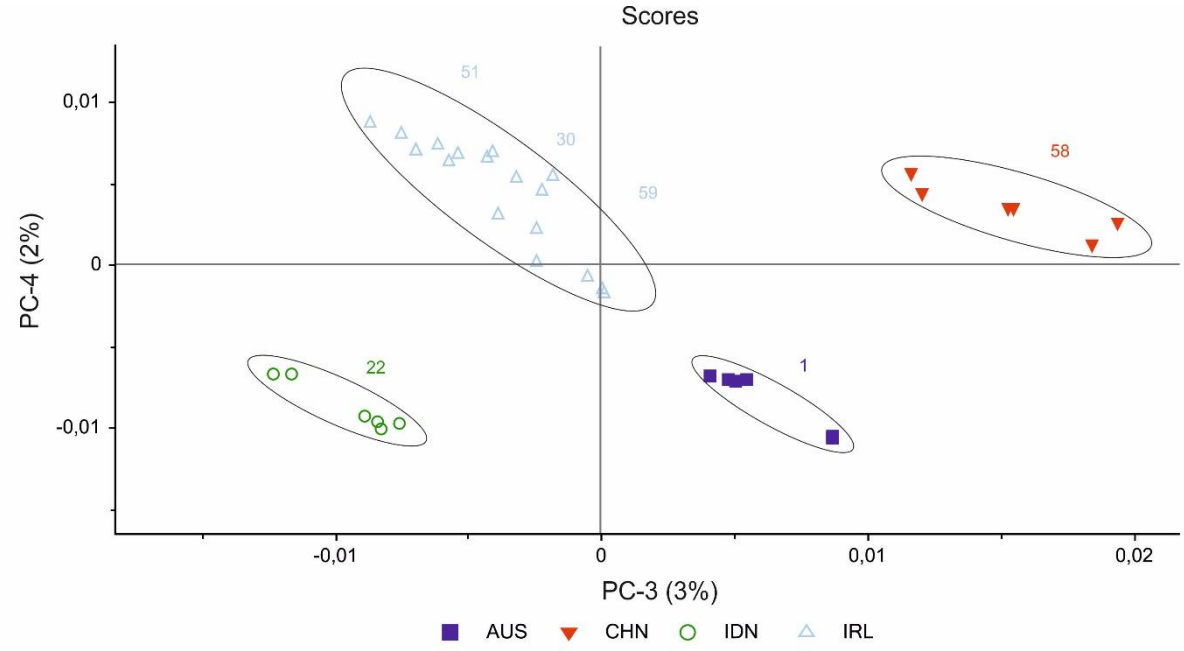


Figure S1 Scores plot of all GlaxoSmithKline samples clustered according to the country of their manufacturing site based on the NIR data. The colored numbers correspond to the sample numbers from Tab. 1 and in color to the manufacturing site in the legend (AUS: Australia; CHN: China; IDN: Indonesia; IRL: Ireland).

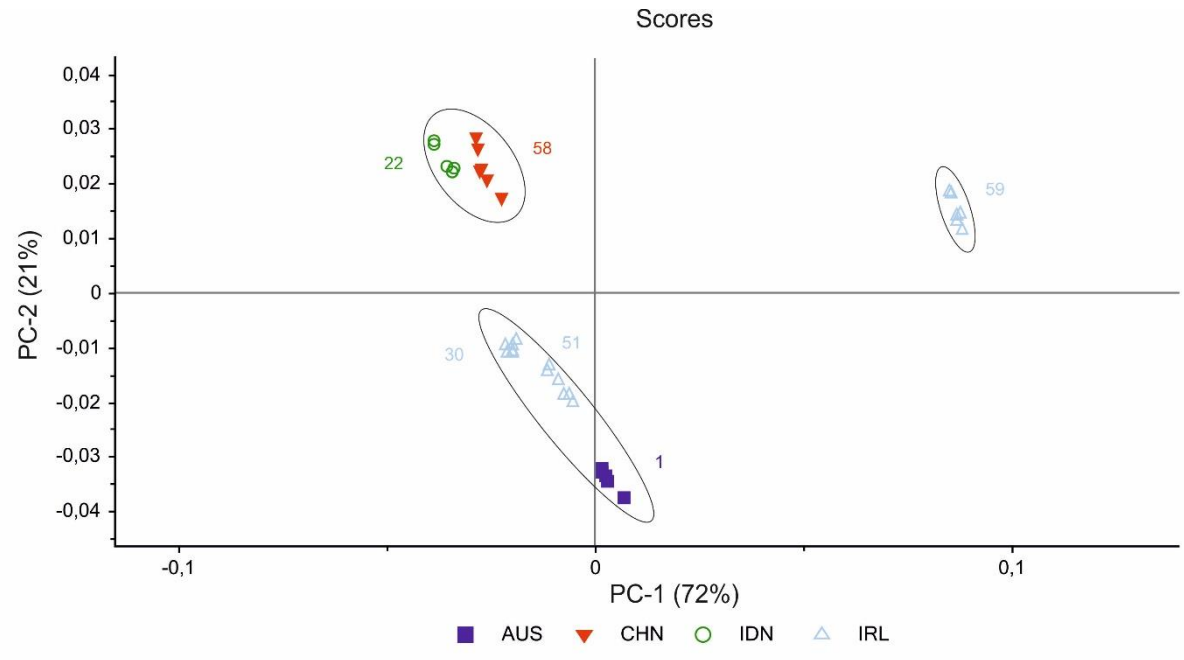


Figure S2 Scores plot of all GlaxoSmithKline samples clustered according to their formulation (Panadol®: 22, 58; Panadol® ActiFast: 59; Panadol® Advanced/Optizorb: 1, 30, 51) based on the NIR data. The colored numbers correspond to the sample numbers from Tab. 1 and in color to the manufacturing site in the legend (AUS: Australia; CHN: China; IDN: Indonesia; IRL: Ireland).

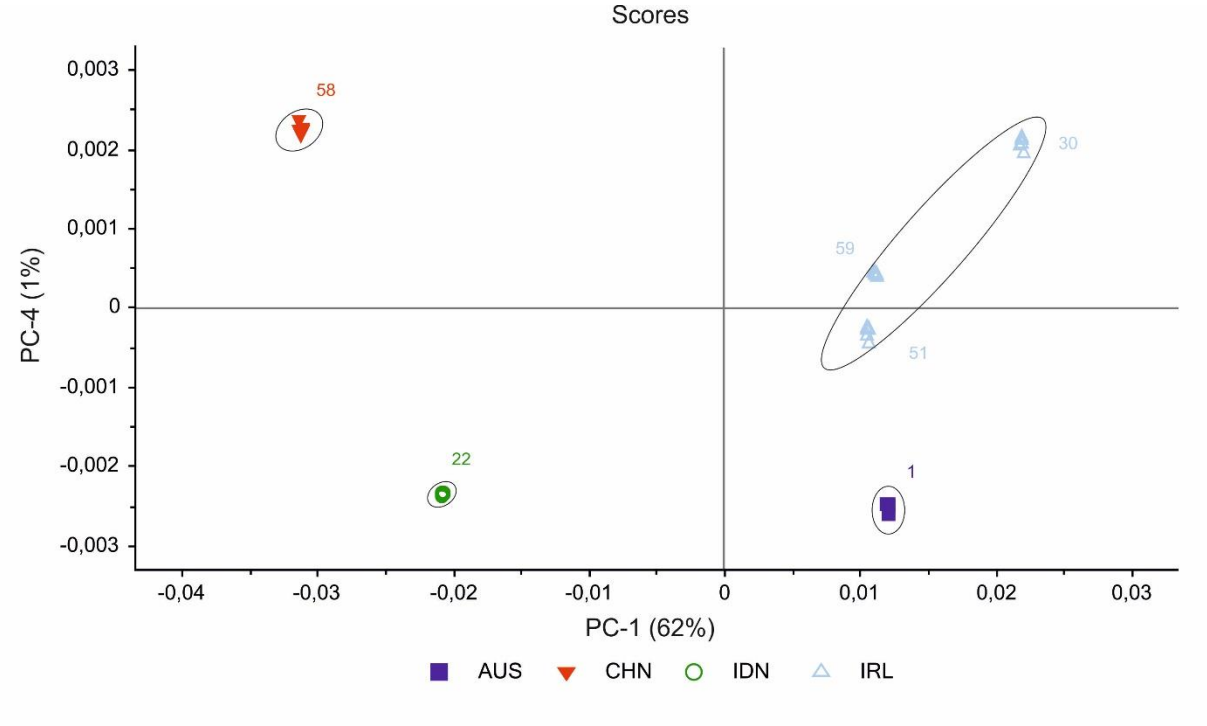


Figure S3 Scores plot of all GlaxoSmithKline samples clustered according to the country of their manufacturing site based on the NMR data. The colored numbers correspond to the sample numbers from Tab. 1 and in color to the manufacturing site in the legend (AUS: Australia; CHN: China; IDN: Indonesia; IRL: Ireland).

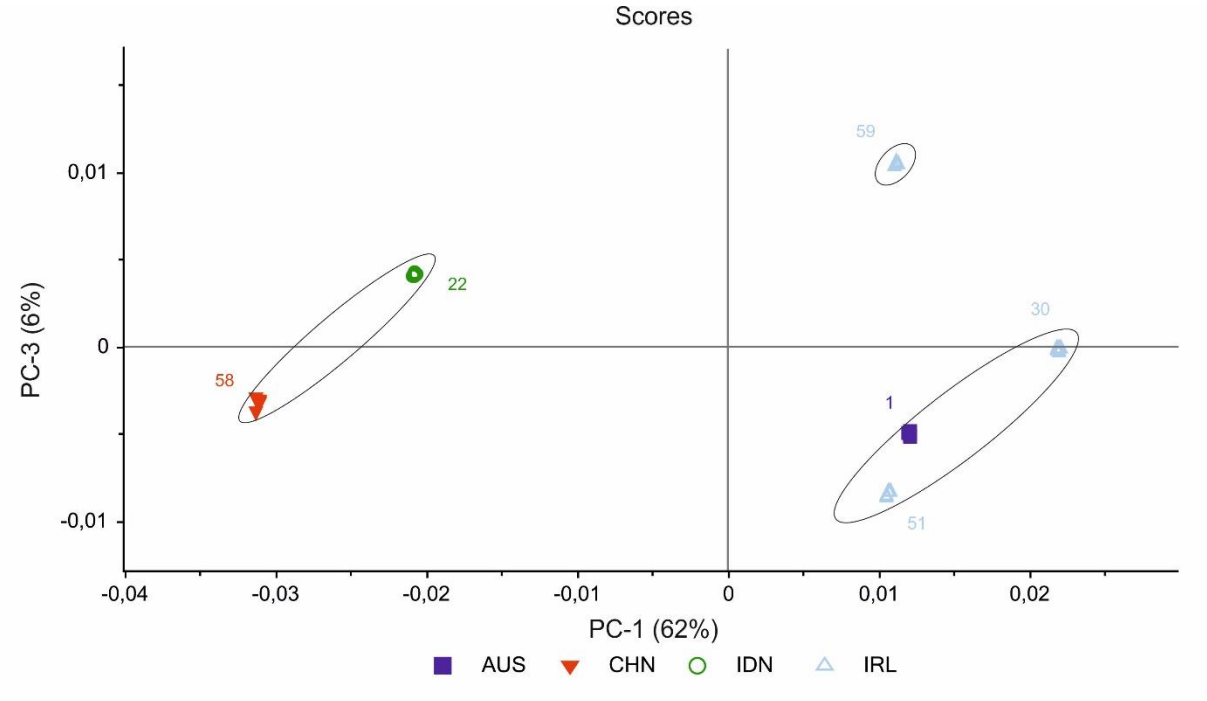


Figure S4 Scores plot of all GlaxoSmithKline samples clustered according to their formulation (Panadol®: 22, 58; Panadol® ActiFast: 59; Panadol® Advanced/Optizorb: 1, 30, 51) based on the NMR data. The colored numbers correspond to the sample numbers from Tab. 1 and in color to the manufacturing site in the legend (AUS: Australia; CHN: China; IDN: Indonesia; IRL: Ireland).

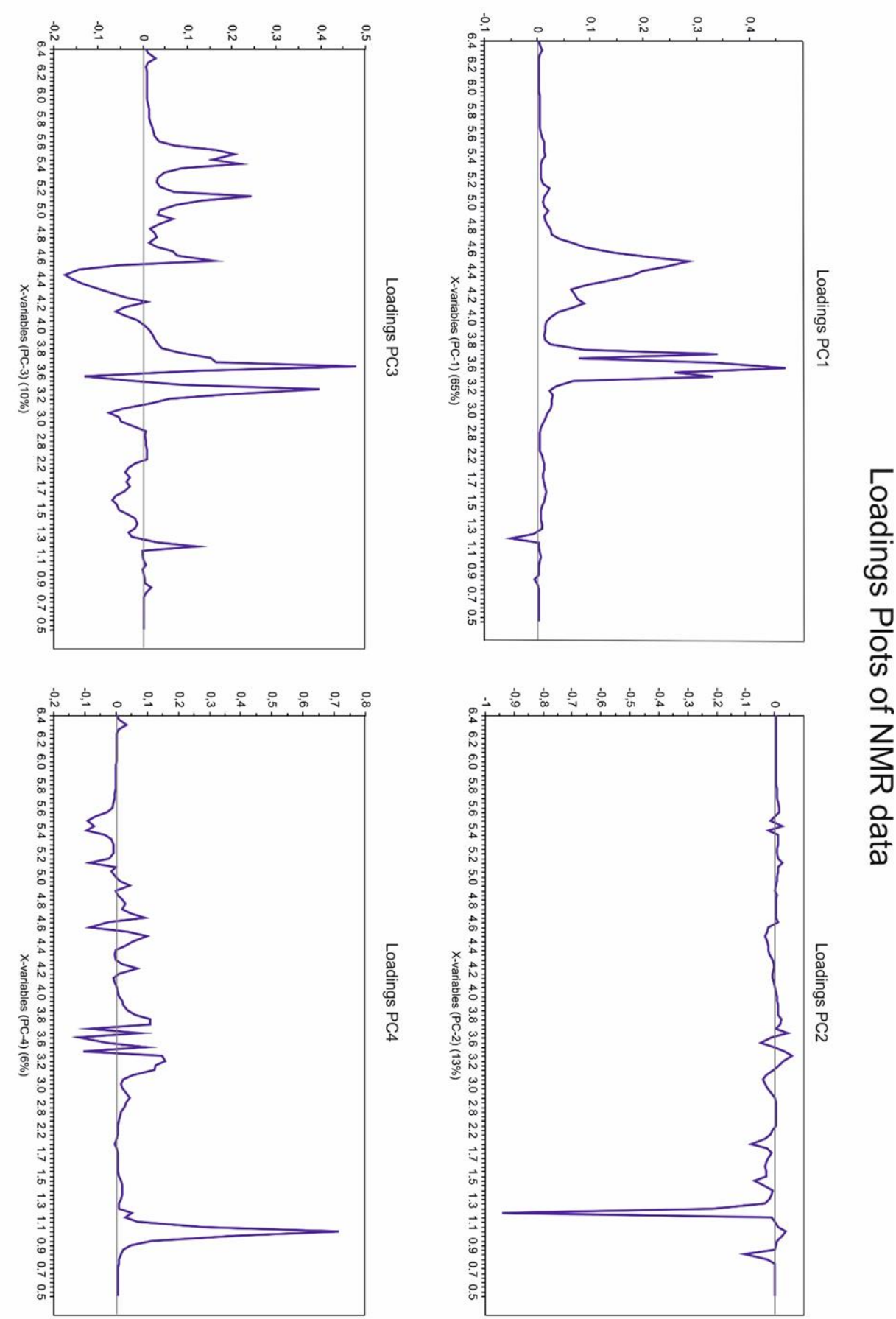


Figure S5 Loadings Plots of the first four principal components of the PCA of NMR data with the help of which the excipients could be identified, which correlate mainly with the respective principal component.

Table S1 Exemplary confusion matrix of one of the LDAs of the MIR spectra with the correct (blue) and wrong (red) predicted spectra in percent per pharmaceutical company.

Predicted \ Actual	1A Pharma	Abiogen Pharma	Actavis	AFIT Pharmaceuticals	Agog Pharma ltd.	Aliud Pharma	Angelini	Aspar pharm. Ltd.	Astra Lifecare	bene-Arzneimittel GmbH	Bernofarm	Beximco Pharma	Bio Profil	Biofarm	Chemists'Own	Crescent Pharma Ltd.	Denk Pharma	Dr. Max Pharma Ltd.	Elys	Farmakopea	Genericon Pharma	GlaxoSmithKline	Hexal AG	Inter Pharm Arzneimittel Gmbh	Interbat	Kamol	Keko Pharm. Industries ltd.	Laboratorios Cinfa	Lincoln Pharm. Ltd.	Millimed	nakhia (Haxия)	Nestor Pharm. Lim.	North China Pharm.	P.t. Phyto Kemo Agung Farma	P.t. Sanbe Farma	pharmstandard (phs)	Priceline	Ratiopharm	Regal Pharm. LTD.	Renewal	S Kant Healthcare ltd.	sanovi-aventis	Sella	Square	STADA	tskf	Zentiva (Sanofi)	Farmoz	Laproff	Johnson & Johnson
1A Pharma	87.5%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Abiogen Pharma	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actavis	0	0	81.25%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AFIT Pharmaceuticals	0	0	0	87.5%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Agog Pharma ltd.	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aliud Pharma	0	0	0	0	0	62.5%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Angelini	0	0	0	0	0	0	87.5%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aspar pharm. Ltd.	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Astra Lifecare	0	0	0	0	0	0	0	0	87.5%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
bene-Arzneimittel GmbH	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bernofarm	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Beximco Pharma	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bio Profil	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Biofarm	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Chemists'Own	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Crescent Pharma Ltd.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Denk Pharma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dr. Max Pharma Ltd.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Elys	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Farmakopea	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Genericon Pharma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GlaxoSmithKline	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hexal AG	12.5%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Inter Pharm Arzneimittel Gmbh	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Interbat	0	0	18.75%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kamol	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Keko Pharm. Industries ltd.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Laboratorios Cinfa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lincoln Pharm. Ltd.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Millimed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
nakhia (Haxия)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nestor Pharm. Lim.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
North China Pharm.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P.t. Phyto Kemo Agung Farma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P.t. Sanbe Farma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0
pharmstandard (phs)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0
Priceline	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0
Ratiopharm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0
Regal Pharm. LTD.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0
Renewal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0
S Kant Healthcare ltd.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0
sanovi-aventis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0
Sella	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0
Square	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0
STADA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0
tskf	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0
Zentiva (Sanofi)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0
Farmoz	0	0	0	12.5%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%
Laproff	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Johnson & Johnson	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Number of spectra per pharmaceutical company	8	8	16	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8

Renewal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	8
Regal Pharm. LTD.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	8
Ratiopharm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	16
Priceline	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	75%	0	0	0	0	0	0	0	0	0	0	0	0	0	8
Pharmstandard (phs)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
P.t. Sanbe Farma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
P.t. Phyto Kemo Agung Farma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
North China Pharm.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
Nestor Pharm. Lim.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
nakhia (naxwa)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
Millimed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12.5%	0	0	0	87.5%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
Lincoln Pharm. Ltd	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
Laboratorios Cinfa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
Keko Pharm. Industries ltd.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
Kamol	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
Interbat	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
Inter Pharm Arzneimittel Gmbh	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
Hexal AG	12.5%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	87.5%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8
GlaxoSmithKline	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	75%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	25%	0	0	0	0	0	0	0	0	32
Genericon Pharma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	87.5%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12.5%	0	0	0	0	0	0	0	0	0	0	0	0	0	8
Farmakopea	0	0	0	0	0	0	12.5%	0	0	0	0	0	0	0	0	0	0	0	0	87.5%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8

Johnson & Johnson	Laproff	Farmoz	Zentiva (Sanofi)	tskf	STADA	Square	Sella	Sanovi-avents	S Kant Healthcare ltd.
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
12.5%	12.5%	12.5%	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	12.5%	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
12.5%	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	12.5%	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	12.5%	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	100%
0	0	0	0	0	0	0	0	100%	0
0	0	0	0	0	0	0	100%	0	0
0	0	0	0	0	0	100%	0	0	0
0	0	0	0	0	87.5%	0	0	0	0
0	0	0	0	100%	0	0	0	0	0
0	0	0	100%	0	0	0	0	0	0
0	0	25%	0	0	0	0	0	0	0
12.5%	75%	37.5%	0	0	0	0	0	0	0
62.5%	12.5%	0	0	0	0	0	0	0	0
8	8	8	8	8	8	8	8	8	8

Table S2 Exemplary confusion matrix of one of the LDAs of the MIR spectra (manufacturer) with the correct (blue) and wrong (red) predicted spectra in percent per manufacturer.

Predicted \ Actual	Salutas Pharma	Abiogen Pharma	PT Actavis Indonesien	Balkanpharma	AFT Pharmaceuticals	Agog Pharma ltd.	STADA	Angelini	Aspar pharm. Ltd.	Astra Lifecare	bene-Arzneimittel GmbH	Biofarm
Salutas Pharma	100%	0	0	0	0	0	0	0	0	0	0	0
Abiogen Pharma	0	100%	0	0	0	0	0	0	0	0	0	0
PT Actavis Indonesien	0	0	100%	0	0	0	0	0	0	0	0	0
Balkanpharma	0	0	0	100%	0	0	0	0	0	0	0	0
AFT Pharmaceuticals	0	0	0	0	100%	0	0	0	0	0	0	0
Agog Pharma ltd.	0	0	0	0	0	100%	0	0	0	0	0	0
STADA	0	0	0	0	0	0	100%	0	0	0	0	0
Angelini	0	0	0	0	0	0	0	87.5%	0	0	0	0
Aspar pharm. Ltd.	0	0	0	0	0	0	0	0	100%	0	0	0
Astra Lifecare	0	0	0	0	0	0	0	0	0	100%	0	0
bene-Arzneimittel GmbH	0	0	0	0	0	0	0	0	0	0	100%	0
Biofarm	0	0	0	0	0	0	0	0	0	0	0	100%
ChemistsOwn	0	0	0	0	0	0	0	0	0	0	0	0
Genmed (Holland)/Medis International (CZ)	0	0	0	0	0	0	0	0	0	0	0	0
Francia Farmaceutici / Nova Argentia	0	0	0	0	0	0	0	12.5%	0	0	0	0
Genericon Pharma	0	0	0	0	0	0	0	0	0	0	0	0
gsk AUS	0	0	0	0	0	0	0	0	0	0	0	0
gsk IRL	0	0	0	0	0	0	0	0	0	0	0	0
gsk IDN	0	0	0	0	0	0	0	0	0	0	0	0
Inter Pharm Arzneimittel Gmbh	0	0	0	0	0	0	0	0	0	0	0	0
HK Pharmaceutical	0	0	0	0	0	0	0	0	0	0	0	0
Laboratorios Cinfa	0	0	0	0	0	0	0	0	0	0	0	0
Lincoln Pharm. Ltd.	0	0	0	0	0	0	0	0	0	0	0	0
nakhia (Наxиа)	0	0	0	0	0	0	0	0	0	0	0	0
Nestor Pharm. Lim.	0	0	0	0	0	0	0	0	0	0	0	0
PT Caprifarmindo Labs.	0	0	0	0	0	0	0	0	0	0	0	0
phs-Leksredstva JSC	0	0	0	0	0	0	0	0	0	0	0	0
Priceline	0	0	0	0	0	0	0	0	0	0	0	0
Merckle	0	0	0	0	0	0	0	0	0	0	0	0
Renewal	0	0	0	0	0	0	0	0	0	0	0	0
S Kant Healthcare ltd.	0	0	0	0	0	0	0	0	0	0	0	0
sanovi-aventis	0	0	0	0	0	0	0	0	0	0	0	0
Sella	0	0	0	0	0	0	0	0	0	0	0	0
Zentiva (Sanofi)	0	0	0	0	0	0	0	0	0	0	0	0
Farmalabor	0	0	0	0	0	0	0	0	0	0	0	0
Number of spectra per manufacturer	16	8	8	8	8	8	16	8	8	8	16	8

Predicted \ Actual	Agog Pharma ltd.	STADA	Angelini	Aspar pharm. Ltd.	Astra Lifecare	bene-Arzneimittel GmbH	Biofarm
Salutas Pharma	0	0	0	0	0	0	0
Abiogen Pharma	0	0	0	0	0	0	0
PT Actavis Indonesien	0	0	0	0	0	0	0
Balkanpharma	0	0	0	0	0	0	0
AFT Pharmaceuticals	0	0	0	0	0	0	0
Agog Pharma ltd.	100%	0	0	0	0	0	0
STADA	0	100%	0	0	0	0	0
Angelini	0	0	87.5%	0	0	0	0
Aspar pharm. Ltd.	0	0	0	100%	0	0	0
Astra Lifecare	0	0	0	0	100%	0	0
bene-Arzneimittel GmbH	0	0	0	0	0	100%	0
Biofarm	0	0	0	0	0	0	100%
ChemistsOwn	0	0	0	0	0	0	0
Genmed (Holland)/Medis International (CZ)	0	0	0	0	0	0	0
Francia Farmaceutici / Nova Argentia	0	0	12.5%	0	0	0	0
Genericon Pharma	0	0	0	0	0	0	0
gsk AUS	0	0	0	0	0	0	0
gsk IRL	0	0	0	0	0	0	0
gsk IDN	0	0	0	0	0	0	0
Inter Pharm Arzneimittel Gmbh	0	0	0	0	0	0	0
HK Pharmaceutical	0	0	0	0	0	0	0
Laboratorios Cinfa	0	0	0	0	0	0	0
Lincoln Pharm. Ltd.	0	0	0	0	0	0	0
nakhia (Наxиа)	0	0	0	0	0	0	0
Nestor Pharm. Lim.	0	0	0	0	0	0	0
PT Caprifarmindo Labs.	0	0	0	0	0	0	0
phs-Leksredstva JSC	0	0	0	0	0	0	0
Priceline	0	0	0	0	0	0	0
Merckle	0	0	0	0	0	0	0
Renewal	0	0	0	0	0	0	0
S Kant Healthcare ltd.	0	0	0	0	0	0	0
sanovi-aventis	0	0	0	0	0	0	0
Sella	0	0	0	0	0	0	0
Zentiva (Sanofi)	0	0	0	0	0	0	0
Farmalabor	0	0	0	0	0	0	0
Number of spectra per manufacturer	8	16	8	8	8	16	8

ChemistsOwn	Genmed (Holland)/Medis International (CZ)	Francia Farmaceutici / Nova Argentia	Genericon Pharma	gsk AUS	gsk IRL	gsk IDN	Inter Pharm Arzneimittel Gmbh	HK Pharmaceutical	Laboratorios Cinfa	Lincoln Pharm. Ltd.	nakhia (нахиа)	Nestor Pharm. Lim.	PT Caprifarmindo Labs.	phs-Leksredstva JSC	Priceline
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	12.5%	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	75%	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	12.5%	100%	0	0	0	0	0	0	0	0	0	0	0	25%
0	0	0	0	100%	43.75%	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	56.25%	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	87.5%	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	75%
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	12.5%	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	8	8	8	8	16	8	8	8	8	8	8	8	8	8	8

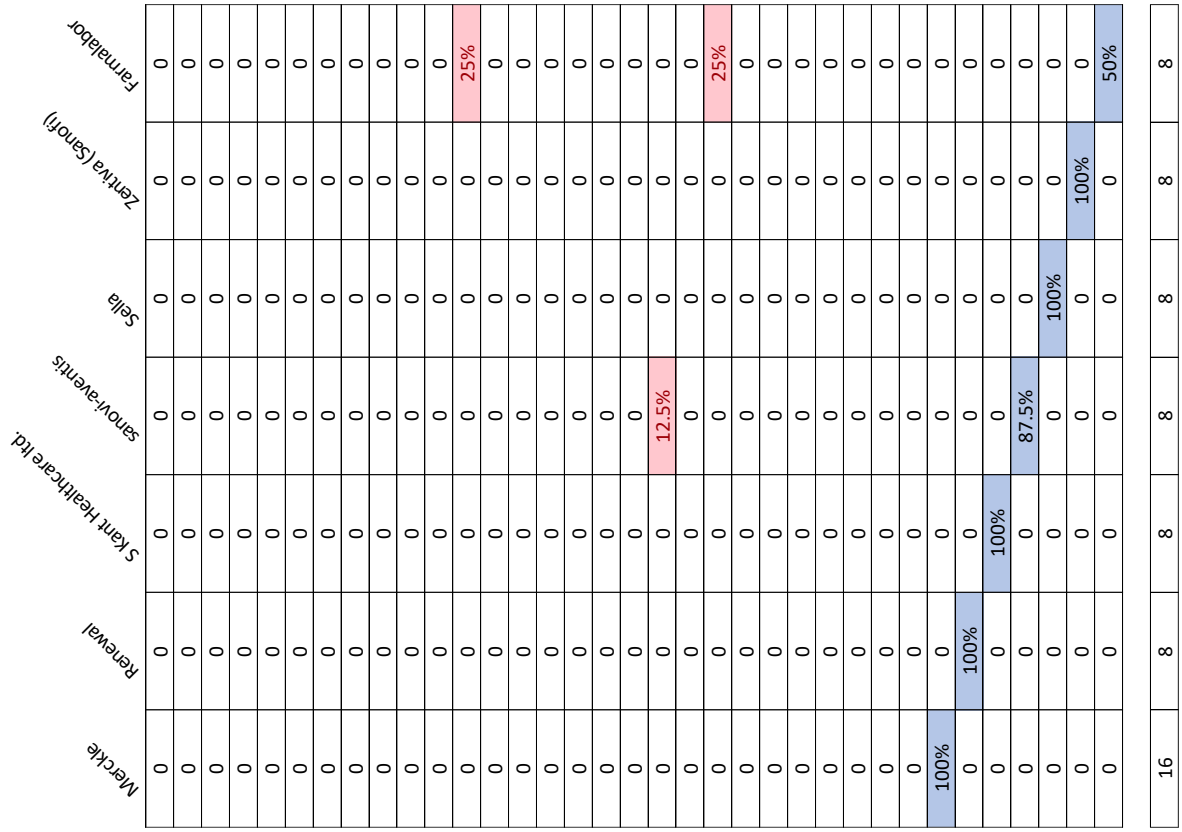

Merckle
Renewal
S Kant Healthcare ltd.
sanovi-aventis
Sella
Zentiva (Sanofi)
Farmalabor
0
25%
12.5%
100%
87.5%
50%
16
8

Table S3 Exemplary confusion matrix of one of the LDAs of the NIR spectra with the correct (blue) and wrong (red) predicted spectra in percent per pharmaceutical company.

Predicted \ Actual	1A Pharma	Abiogen Pharma	Actavis	AFIT Pharmaceuticals	Agog Pharma ltd.	Aliud Pharma	Angelini	Aspar pharm. Ltd.	Astra Lifecare	Bene-Arzneimittel GmbH	Berndorf	Beximco Pharma	Bio Profil	Biofarm	Chemist's Own	Crescent Pharma Ltd.	Denk Pharma	Dr. Max Pharma Ltd.	Elys
1A Pharma	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Abiogen Pharma	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actavis	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AFIT Pharmaceuticals	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Agog Pharma ltd.	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aliud Pharma	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0
Angelini	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0
Aspar pharm. Ltd.	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0
Astra Lifecare	0	0	0	0	0	0	0	0	100%	0	0	0	0	12.5%	0	0	0	0	0
bene-Arzneimittel GmbH	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0
Berndorf	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0
Beximco Pharma	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0
Bio Profil	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0
Biofarm	0	0	0	0	0	0	0	0	0	0	0	0	0	87.5%	0	0	0	0	0
Chemist's Own	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0
Crescent Pharma Ltd.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0
Denk Pharma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0
Dr. Max Pharma Ltd.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0
Elys	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%
Farmakopea	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Genericon Pharma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GlaxoSmithKline	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hexal AG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Inter Pharm Arzneimittel GmbH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Interbat	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kamol	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Keko Pharm. Industries ltd.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Laboratorios Cinfa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lincoln Pharm. Ltd.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Millimed	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
nakhia (Haxия)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nestor Pharm. Lim.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
North China Pharm.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P.t. Phyto Kemo Agung Farma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P.t. Sanbe Farma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pharmstandard (phs)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Priceline	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ratiopharm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Regal Pharm. LTD.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Renewal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S Kant Healthcare ltd.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
sanovi-aventis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sella	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Square	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
STADA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
tskf	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Zentiva (Sanofi)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Farmoz	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Laproff	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Johnson & Johnson	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Number of spectra per pharmaceutical company	8	8	16	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8

	Famakopea	Genericon Pharma	GlaxoSmithKline	Hexal AG	Inter Pharm	Interfarm GmbH	Interbat	Kanol	Keko Pharm. Industries ltd.	Laboratorios Cinfa	Lincoln Pharm. Ltd.	Milliméd	Nakina (Hexma)	Nestor Pharm. Ltn.	North China Pharm.	P.L. Phyto Kemo Agrus Farma	P.L. Sanbe Farma	Pharmstandard (phs)	Priceline	Rekopharm	Regal Pharm. LTD.	Renewal	
	8	8	40	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	16	8	8	
Famakopea	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Genericon Pharma	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
GlaxoSmithKline	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Hexal AG	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Inter Pharm	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Interfarm GmbH	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Interbat	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kanol	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Keko Pharm. Industries ltd.	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Laboratorios Cinfa	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lincoln Pharm. Ltd.	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Milliméd	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nakina (Hexma)	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nestor Pharm. Ltn.	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
North China Pharm.	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P.L. Phyto Kemo Agrus Farma	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P.L. Sanbe Farma	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Pharmstandard (phs)	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Priceline	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Rekopharm	16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Regal Pharm. LTD.	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Renewal	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

	S Kant Healthcare ltd.	sanovy aventis	Sella	Square	STADA	tskf	Zentiva (Sanofi)	Farmoz	Laporff	Johnson & Johnson	
	0	0	0	0	0	0	0	0	0	0	8
	0	0	0	0	0	0	0	0	0	0	8
	0	0	0	0	0	0	0	0	0	0	8
	0	0	0	0	0	0	0	0	0	0	8
	0	0	0	0	0	0	0	0	0	0	8
	0	0	0	0	0	0	0	0	0	0	8
	0	0	0	0	0	0	0	0	12.5%	0	8
	100%	0	0	0	0	0	0	0	0	0	8
	0	100%	0	0	0	0	0	0	0	0	8
	0	0	100%	0	0	0	0	0	0	0	8
	0	0	0	100%	0	0	0	0	0	0	8
	0	0	0	0	100%	0	0	0	0	0	8
	0	0	0	0	0	100%	0	0	0	0	8
	0	0	0	0	0	0	100%	0	0	0	8
	0	0	0	0	0	0	0	100%	0	0	8
	0	0	0	0	0	0	0	0	75%	0	8
	0	0	0	0	0	0	0	0	0	100%	8

Table S4 Exemplary confusion matrix of one of the LDAs of the NIR spectra with the correct (blue) and wrong (red) predicted spectra in percent per manufacturer.

Actual / Predicted	Salutas Pharma	Abiogen Pharma	PT Actavis Indonesien	Balkanpharma	AFT Pharmaceuticals	Agog Pharma ltd.	STADA	Angelini	Aspar pharm. Ltd.	Astra Lifecare	bene-Arzneimittel GmbH	Biofarm
Salutas Pharma	100%	0	0	0	0	0	0	0	0	0	0	0
Abiogen Pharma	0	100%	0	0	0	0	0	0	0	0	0	0
PT Actavis Indonesien	0	0	100%	0	0	0	0	0	0	0	0	0
Balkanpharma	0	0	0	100%	0	0	0	0	0	0	0	0
AFT Pharmaceuticals	0	0	0	0	100%	0	0	0	0	0	0	0
Agog Pharma ltd.	0	0	0	0	0	100%	0	0	0	0	0	0
STADA	0	0	0	0	0	0	100%	0	0	0	0	0
Angelini	0	0	0	0	0	0	0	100%	0	0	0	0
Aspar pharm. Ltd.	0	0	0	0	0	0	0	0	100%	0	0	0
Astra Lifecare	0	0	0	0	0	0	0	0	0	100%	0	0
bene-Arzneimittel GmbH	0	0	0	0	0	0	0	0	0	0	100%	0
Biofarm	0	0	0	0	0	0	0	0	0	0	0	100%
ChemistsOwn	0	0	0	0	0	0	0	0	0	0	0	0
Genmed (Holland)/Medis International (CZ)	0	0	0	0	0	0	0	0	0	0	0	0
Francia Farmaceutici / Nova Argentia	0	0	0	0	0	0	0	0	0	0	0	0
Genericon Pharma	0	0	0	0	0	0	0	0	0	0	0	0
gsk AUS	0	0	0	0	0	0	0	0	0	0	0	0
gsk IRL	0	0	0	0	0	0	0	0	0	0	0	0
gsk IDN	0	0	0	0	0	0	0	0	0	0	0	0
Inter Pharm Arzneimittel Gmbh	0	0	0	0	0	0	0	0	0	0	0	0
HK Pharmaceutical	0	0	0	0	0	0	0	0	0	0	0	0
Laboratorios Cinfa	0	0	0	0	0	0	0	0	0	0	0	0
Lincoln Pharm. Ltd.	0	0	0	0	0	0	0	0	0	0	0	0
nakhia (Нахиа)	0	0	0	0	0	0	0	0	0	0	0	0
Nestor Pharm. Lim.	0	0	0	0	0	0	0	0	0	0	0	0
PT Caprifarmindo Labs.	0	0	0	0	0	0	0	0	0	0	0	0
phs-Leksredstva JSC	0	0	0	0	0	0	0	0	0	0	0	0
Priceline	0	0	0	0	0	0	0	0	0	0	0	0
Merckle	0	0	0	0	0	0	0	0	0	0	0	0
Renewal	0	0	0	0	0	0	0	0	0	0	0	0
S Kant Healthcare ltd.	0	0	0	0	0	0	0	0	0	0	0	0
sanovi-aventis	0	0	0	0	0	0	0	0	0	0	0	0
Sella	0	0	0	0	0	0	0	0	0	0	0	0
Zentiva (Sanofi)	0	0	0	0	0	0	0	0	0	0	0	0
Farmalabor	0	0	0	0	0	0	0	0	0	0	0	0
Number of spectra per manufacturer	16	8	8	8	8	8	16	8	8	8	16	8

Actual / Predicted	Agog Pharma ltd.	STADA	Angelini	Aspar pharm. Ltd.	Astra Lifecare	bene-Arzneimittel GmbH	Biofarm
Salutas Pharma	0	0	0	0	0	0	0
Abiogen Pharma	0	0	0	0	0	0	0
PT Actavis Indonesien	0	0	0	0	0	0	0
Balkanpharma	0	0	0	0	0	0	0
AFT Pharmaceuticals	0	0	0	0	0	0	0
Agog Pharma ltd.	100%	0	0	0	0	0	0
STADA	0	100%	0	0	0	0	0
Angelini	0	0	100%	0	0	0	0
Aspar pharm. Ltd.	0	0	0	100%	0	0	0
Astra Lifecare	0	0	0	0	100%	0	0
bene-Arzneimittel GmbH	0	0	0	0	0	100%	0
Biofarm	0	0	0	0	0	0	100%
ChemistsOwn	0	0	0	0	0	0	0
Genmed (Holland)/Medis International (CZ)	0	0	0	0	0	0	0
Francia Farmaceutici / Nova Argentia	0	0	0	0	0	0	0
Genericon Pharma	0	0	0	0	0	0	0
gsk AUS	0	0	0	0	0	0	0
gsk IRL	0	0	0	0	0	0	0
gsk IDN	0	0	0	0	0	0	0
Inter Pharm Arzneimittel Gmbh	0	0	0	0	0	0	0
HK Pharmaceutical	0	0	0	0	0	0	0
Laboratorios Cinfa	0	0	0	0	0	0	0
Lincoln Pharm. Ltd.	0	0	0	0	0	0	0
nakhia (Нахиа)	0	0	0	0	0	0	0
Nestor Pharm. Lim.	0	0	0	0	0	0	0
PT Caprifarmindo Labs.	0	0	0	0	0	0	0
phs-Leksredstva JSC	0	0	0	0	0	0	0
Priceline	0	0	0	0	0	0	0
Merckle	0	0	0	0	0	0	0
Renewal	0	0	0	0	0	0	0
S Kant Healthcare ltd.	0	0	0	0	0	0	0
sanovi-aventis	0	0	0	0	0	0	0
Sella	0	0	0	0	0	0	0
Zentiva (Sanofi)	0	0	0	0	0	0	0
Farmalabor	0	0	0	0	0	0	0
Number of spectra per manufacturer	8	16	8	8	8	16	8

ChemistsOwn	Genmed (Holland)/Medis International (CZ)	Francia Farmaceutici / Nova Argentia	Genericon Pharma	gsk AUS	gsk IRL	gsk IDN	Inter Pharm Arzneimittel Gmbh	HK Pharmaceutical	Laboratorios Cinfa	Lincoln Pharm. Ltd.	nakhia (нахиа)	Nestor Pharm. Lim.	PT Caprifarmindo Labs.	phs-Leksredstva JSC
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	8	8	8	8	16	8	8	8	8	8	8	8	8	8

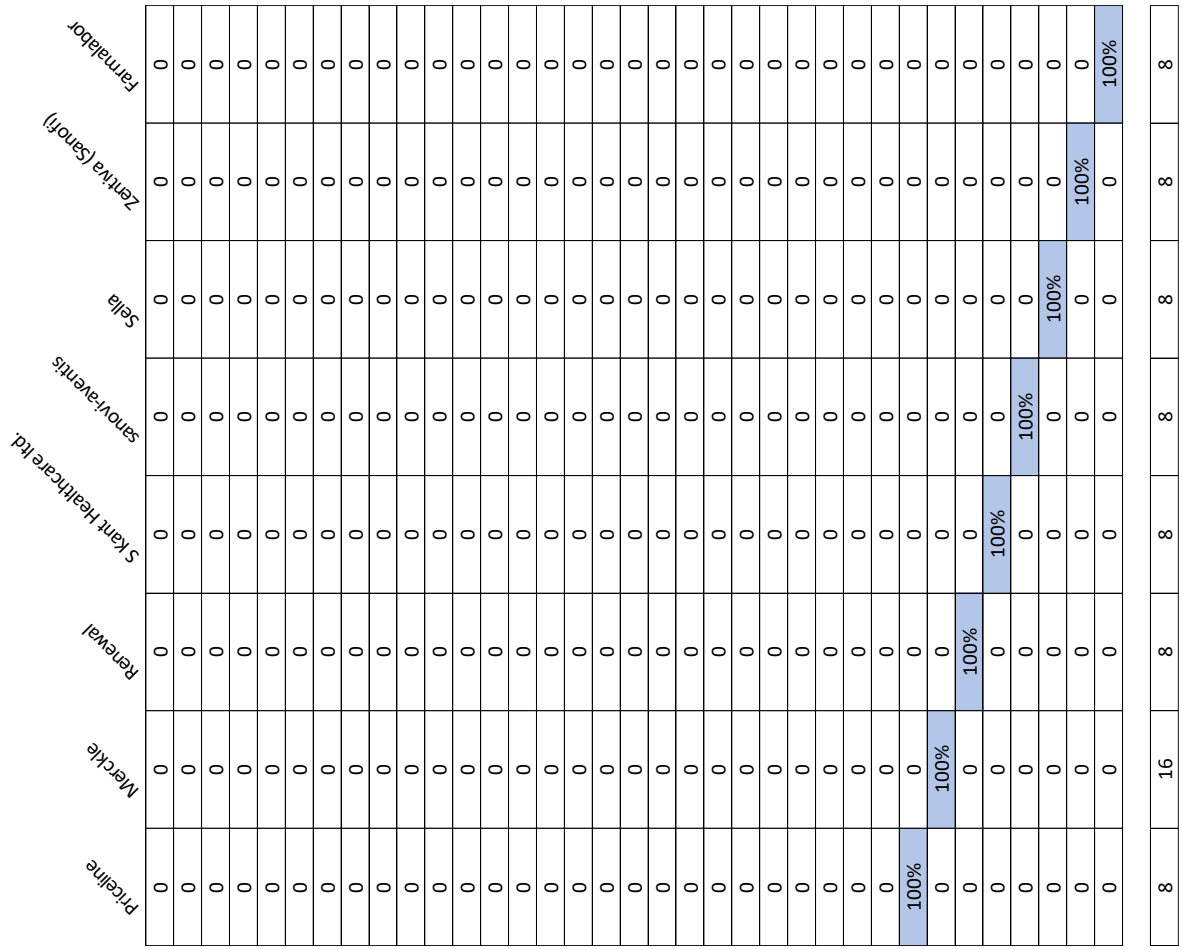

Pricelline
Merckle
Renewal
S Kant Healthcare ltd.
sanovi-aventis
Sella
Zentiva (Sanofi)
Farmalabor

Table S5 Exemplary confusion matrix of one of the LDAs of the NMR spectra with the correct (blue) and wrong (red) predicted spectra in percent per pharmaceutical company.

Predicted \ Actual	GlaxoSmithKline	Chemists' Own	AFT Pharmaceuticals	Priceline	1A Pharma	Ratiopharm	Angelini	Farmakopea	Abiogen Pharma	Sella	Genericon Pharma	Millimed	Kamol	P.t. Sanbe Farma	Bernofarm	P.t. Phyto Kemo Agung Farma	Interbat	Actavis	Square	Beximco Pharma	Crescent Pharma Ltd.	Nestor Pharm. Lim.	North China Pharm.	Regal Pharm. LTD.	Astra Lifecare	Lincoln Pharm. Ltd.	Denk Pharma	Elys	Agog Pharma ltd.	Keko Pharm. Industries ltd.	S.Kant Healthcare ltd.	Dr. Max Pharma Ltd.	Zentiva (Sanofi)	Bio Profil	Biofarm	STADA	Aliud Pharma	Inter Pharm Arzneimittel Gmbh	bene-Arzneimittel GmbH	Hexal AG	Aspar pharm. Ltd.	sanovi-aventis	Laboratorios Cinfa	Renewal	pharmstandard (phs)	nakhia (нахиа)	tskf	Laproff	Johnson & Johnson	Farmoz
GlaxoSmithKline	64%	0	0	0	0	16%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20%	0	0	0	0	0	0
Chemists' Own	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AFT Pharmaceuticals	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Priceline	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1A Pharma	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ratiopharm	16%	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Angelini	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Farmakopea	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0
Abiogen Pharma	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sella	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Genericon Pharma	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Millimed	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Kamol	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P.t. Sanbe Farma	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bernofarm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P.t. Phyto Kemo Agung Farma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Interbat	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Actavis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Square	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%
Beximco Pharma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Crescent Pharma Ltd.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nestor Pharm. Lim.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
North China Pharm.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Regal Pharm. LTD.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Astra Lifecare	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Lincoln Pharm. Ltd.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Denk Pharma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Elys	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Agog Pharma ltd.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Keko Pharm. Industries ltd.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S.Kant Healthcare ltd.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Dr. Max Pharma Ltd.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Zentiva (Sanofi)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Bio Profil	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Biofarm	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0
STADA	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0
Aliud Pharma	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0
Inter Pharm Arzneimittel Gmbh	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0
bene-Arzneimittel GmbH	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0
Hexal AG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0
Aspar pharm. Ltd.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0
sanovi-aventis	20%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0
Laboratorios Cinfa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0
Renewal	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0
pharmstandard (phs)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0
nakhia (нахиа)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0
tskf	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0
Laproff	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0
Johnson & Johnson	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%
Farmoz	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Number of spectra per pharmaceutical company	25	5	5	5	5	5	10	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	10	5	5	5	5	5	5

Bexinco Pharma	Crescent Pharma Ltd.	Nestor Pharm. Lim.	North China Pharm.	Regal Pharm. LTD.	Astra Lifecare	Lincoln Pharm. Ltd.	Denk Pharma	Elys	Agog Pharma ltd.	Keko Pharm. Industries ltd.	S Kant Healthcare ltd.	Dr. Max Pharma Ltd.	Zentiva (Sanofi)	Bio Profil	Biofarm	STADA	Allied Pharma	Inter Pharm Arzneimittel GmbH	bene-Arzneimittel GmbH	Hexal AG
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5

Aspar pharm. Ltd.	sanovi-aventis	Laboratorios Cinfa	Renewal	pharmstandard (phs)	nakhia (нахиа)	tskf	Laproff	Johnson & Johnson	Farmoz
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
100%	0	0	0	0	0	0	0	0	0
0	100%	0	0	0	0	0	0	0	0
0	0	100%	0	0	0	0	0	0	0
0	0	0	100%	0	0	0	0	0	0
0	0	0	0	100%	0	0	0	0	0
0	0	0	0	0	100%	0	0	0	0
0	0	0	0	0	0	100%	0	0	0
0	0	0	0	0	0	0	100%	0	0
0	0	0	0	0	0	0	0	100%	0
0	0	0	0	0	0	0	0	0	100%
5	5	5	5	5	5	5	5	5	5

Table S6 Exemplary confusion matrix of one of the LDAs of the NMR spectra with the correct (blue) and wrong (red) predicted spectra in percent per manufacturer.

Predicted \ Actual	gsk AUS	ChemistsOwn	AFT Pharmaceuticals	Priceline	Salutas Pharma	Merckle	Angelini	Francia Farmaceutici / Nova Argentia	Abiogen Pharma	Sella	Genericon Pharma
gsk AUS	100%	0	0	0	0	0	0	0	0	0	0
ChemistsOwn	0	100%	0	0	0	0	0	0	0	0	0
AFT Pharmaceuticals	0	0	100%	0	0	0	0	0	0	0	0
Priceline	0	0	0	100%	0	0	0	0	0	0	0
Salutas Pharma	0	0	0	0	100%	0	0	0	0	0	0
Merckle	0	0	0	0	0	100%	0	0	0	0	0
Angelini	0	0	0	0	0	0	100%	0	0	0	0
Francia Farmaceutici / Nova Argentia	0	0	0	0	0	0	0	100%	0	0	0
Abiogen Pharma	0	0	0	0	0	0	0	0	100%	0	0
Sella	0	0	0	0	0	0	0	0	0	100%	0
Genericon Pharma	0	0	0	0	0	0	0	0	0	0	100%
HK Pharmaceutical	0	0	0	0	0	0	0	0	0	0	0
PT Caprifarmindo Labs.	0	0	0	0	0	0	0	0	0	0	0
gsk IDN	0	0	0	0	0	0	0	0	0	0	0
PT Actavis Indonesien	0	0	0	0	0	0	0	0	0	0	0
Nestor Pharm. Lim.	0	0	0	0	0	0	0	0	0	0	0
Astra Lifecare	0	0	0	0	0	0	0	0	0	0	0
Lincoln Pharm. Ltd.	0	0	0	0	0	0	0	0	0	0	0
bene-Arzneimittel GmbH	0	0	0	0	0	0	0	0	0	0	0
Agog Pharma ltd.	0	0	0	0	0	0	0	0	0	0	0
S Kant Healthcare ltd.	0	0	0	0	0	0	0	0	0	0	0
Genmed (Holland)/Medis International (CZ)	0	0	0	0	0	0	0	0	0	0	0
Zentiva (Sanofi)	0	0	0	0	0	0	0	0	0	0	0
Balkanpharma	0	0	0	0	0	0	0	0	0	0	0
Biofarm	0	0	0	0	0	0	0	0	0	0	0
STADA	0	0	0	0	0	0	0	0	0	0	0
Inter Pharm Arzneimittel Gmbh	0	0	0	0	0	0	0	0	0	0	0
Aspar pharm. Ltd.	0	0	0	0	0	0	0	0	0	0	0
gsk IRL	0	0	0	0	0	0	0	0	0	0	0
sanovi-aventis	0	0	0	0	0	0	0	0	0	0	0
Laboratorios Cinfa	0	0	0	0	0	0	0	0	0	0	0
Renewal	0	0	0	0	0	0	0	0	0	0	0
phs-Leksredstva JSC	0	0	0	0	0	0	0	0	0	0	0
nakhia (Haxwa)	0	0	0	0	0	0	0	0	0	0	0
Farmalabor	0	0	0	0	0	0	0	0	0	0	0
Number of spectra per manufacturer	5	5	5	5	10	10	5	5	5	5	5

HK Pharmaceutical	PT Caprifarmindo Labs.	gsk IDN	PT Actavis Indonesien	Nestor Pharm. Lim.	Astra Lifecare	Lincoln Pharm. Ltd.	bene-Arzneimittel GmbH	Aegog Pharma ltd.	S Kant Healthcare ltd.	Genmed (Holland)/Medis International (CZ)	Zentiva (Sanofi)	Balkanpharma	Biofarm	STADA
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
100%	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	100%	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	100%	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	100%	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	100%	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	100%	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	100%	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	100%	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	100%	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	100%	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	100%	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	100%	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	100%	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	100%	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	5	5	5	5	5	5	10	5	5	5	5	5	5	10

Inter Pharm Arzneimittel Gmbh	Aspar pharm. Ltd.	Bsk IRL	Sanovi-aventis	Laboratorios Cinfa	Renewal	phs-Leksredstva JSC	nakhia (haxna)	Farmalabor
0	0	10%	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
100%	0	0	0	0	0	0	0	0
0	100%	0	0	0	0	0	0	0
0	0	90%	0	0	0	0	0	0
0	0	0	100%	0	0	0	0	0
0	0	0	0	100%	0	0	0	0
0	0	0	0	0	100%	0	0	0
0	0	0	0	0	0	100%	0	0
0	0	0	0	0	0	0	100%	0
0	0	0	0	0	0	0	0	100%
5	5	10	5	5	5	5	5	5

3.2 Hierarchical cluster analysis of spectral data of different paracetamol tablets

1. Introduction

Cluster analysis is another unsupervised tool for exploratory data analysis [1,2]. It aims to analyze similarities between samples and then use them to classify the samples into groups (clusters). Thereby, the properties within the groups should be homogeneous, whereas they should show a large heterogeneity towards the other clusters [2]. One of the many different clustering methods used today is hierarchical cluster analysis (HCA). It is characterized by the fact that it is not necessary to define a certain number of groups in advance and that the results are presented using a dendrogram. The obtained dendrogram can then be used to define different groups with the help of their relative distance [3].

The HCA is divided into two types that determine how the clusters are formed: agglomerative and divisive. In the divisive method, all samples belong to one large cluster at the beginning and are divided into smaller ones as they progress until each cluster consists of only one sample. In the agglomerative method, each sample starts as a separate cluster, which are then combined into larger ones until all samples belong to only one cluster [2]. Independently of the choice of clustering, the way of quantifying the similarity or dissimilarity of two objects must also be determined, e.g., by using the Euclidean distance [3].

The aim of this chapter was to compare the PCA results of subchapter 3.1 in relation to the manufacturer affiliations of the different paracetamol preparations and to see if similar clusters form or if other relationships can be identified.

2. Materials and methods

The spectroscopic data (NIR, MIR, NMR) used for the HCA were taken from the work in subchapter 3.1. There, the materials, sample preparation, spectral experiments, and data preprocessing are described in detail. For HCA, the MIR absorption spectra, as well as the bucketed NMR spectra (0.04 ppm) were used. Further data pre-processing used is described below.

2.1 Drug samples

Table 1 Clusters as observed in the PCA in subchapter 3.1 (*AUS*, Australia; *CHN*, China; *GBR*, Great Britain; *HKG*, Hong Kong; *IDN*, Indonesia; *TZA*, Tanzania)

Cluster	Definition	NIR	MIR	NMR
A	samples produced by the same manufacturer	Hexal AG/1A Pharma GmbH; STADA Arzneimittel AG/Aliud Pharma GmbH; bene-Arzneimittel GmbH/Denk Pharma GmbH & Co. KG; Ratiopharm GmbH (from DEU/AUT)	as for NIR	as for NIR
B	bought in Tanzania but manufactured in India	Nestor Pharm. Lim.; Agog Pharma Ltd.; Astra Lifecare (India) Pvt. Ltd.	Nestor Pharm. Lim.; Agog Pharma Ltd.; Lincoln Pharma; S Kant Healthcare	Nestor Pharm. Lim.; Agog Pharma Ltd.; Astra Lifecare (India) Pvt. Ltd.; Priceline (from AUS)
C	bought in Tanzania but manufactured in Kenya	Elys Chemical Industries Ltd.; Regal Pharm. Ltd.	-	-
D	bought in Tanzania	-	Nestor Pharm. Lim.; S Kant Healthcare; Agog Pharma Ltd.; Lincoln Pharma; Keko Pharm. Industries Ltd.; Elys Chemical Industries Ltd.; North China Pharm.	-
E	bought and manufactured in Italy	Angelini ACRAF SpA; Farmakopea SpA; PT. Actavis Indonesia	Angelini ACRAF SpA; Farmakopea SpA	as for NIR
F	bought and manufactured in Indonesia	PT. Interbat Pharmaceutical Industry and Bernofarm or PT. Interbat Pharmaceutical Industry and P.t. Sanbe Farma	PT. Actavis Indonesia; P.t. Phyto Kemo Agung Farma	PT. Interbat Pharmaceutical Industry; P.t. Phyto Kemo Agung Farma and Bernofarm or PT. Interbat Pharmaceutical Industry and P.t. Sanbe Farma or PT. Actavis Indonesia and Bernofarm (depending on PCs)
G	Panadol® preparations from GlaxoSmithKline	G1 (Panadol®): GSK IDN/tskf CHN; G2 (Panadol® Advance/Optizorb): GSK AUS/GBR/TZA; G3 (Panadol® ActiFast): GSK HKG	as for NIR	G2 + G3

The 56 drug samples used are listed more detailed in Table 1 in subchapter 3.1. For ease of comparison and understanding, the same definitions of clusters were used as for PCA and summarized in Table 1.

2.2 Pre-processing and hierarchical cluster analysis

For HCA, the MIR data were reduced by a factor of two by averaging before calculating the mean spectra of each tablet and forming their first derivative. Mean spectra were also formed for the NIR and NMR spectra, with an additional standard normal variate (SNV) transformation followed by formation of the first derivative for the NIR data. The final spectral ranges used for HCA remained the same for NMR (6.42–3.54 ppm, 3.38–2.74 ppm, 2.30–2.18 ppm, and 1.78–0.50 ppm), MIR (1175.401–861.0605 cm^{-1}), and NIR (1100.125–1242.595 nm and 1347.899–1570.896 nm) as for the PCA.

The HCA was performed using hierarchical complete-linkage as clustering method and absolute correlation as distance measure. The pre-processing methods and the HCA of all spectral methods were performed using the Unscrambler X 10.4 and 11.0 (CAMO Software AS., Oslo, Norway). The plots were created and edited with Unscrambler 11.0 and CorelDRAW 2019 v 21 (Corel Corporation, Ottawa, Canada).

3. Results and discussion

The absolute correlation measure was used because it can capture the spectral variable similarities between formulations better than a straight distance measure such as the Euclidean measure [4]. To test whether the chosen HCA method was suitable for the study of the samples, the first step was the execution of the HCA only with those samples for which the correlations were known beyond doubt, such as the preparations of cluster A or the three different GlaxoSmithKline formulations (Cluster G). As can be seen in Figure 1 and 2, the HCAs were able to form the correct clusters for all samples and formulations. Therefore, all samples were analyzed at once by HCA to see if clusters similar to the PCA could be observed.

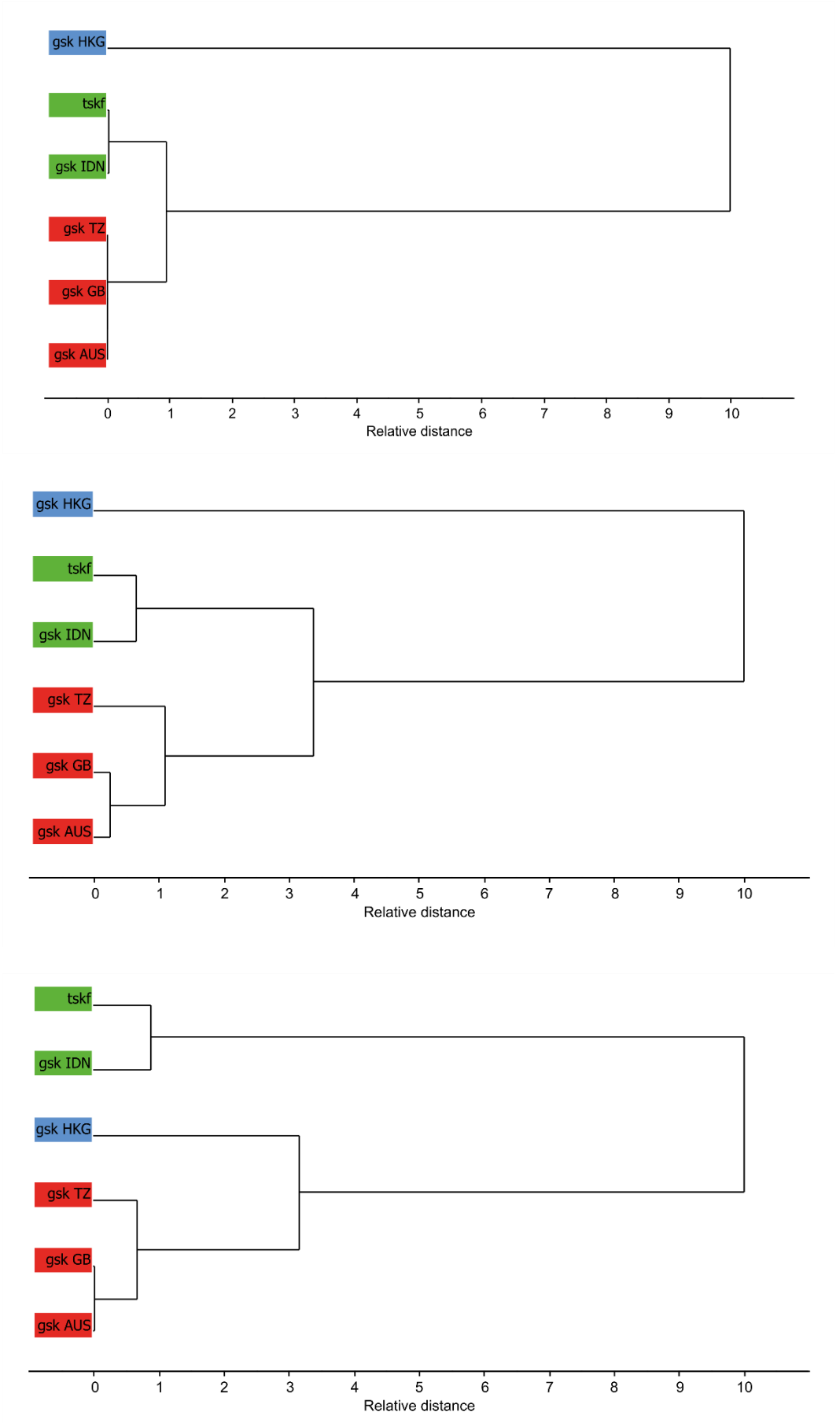


Figure 1 HCAs of the different GSK formulations: for the MIR (top), NIR (middle) and NMR (bottom) data (green, standard; blue, ActiFast; red, Advance/Optizorb).

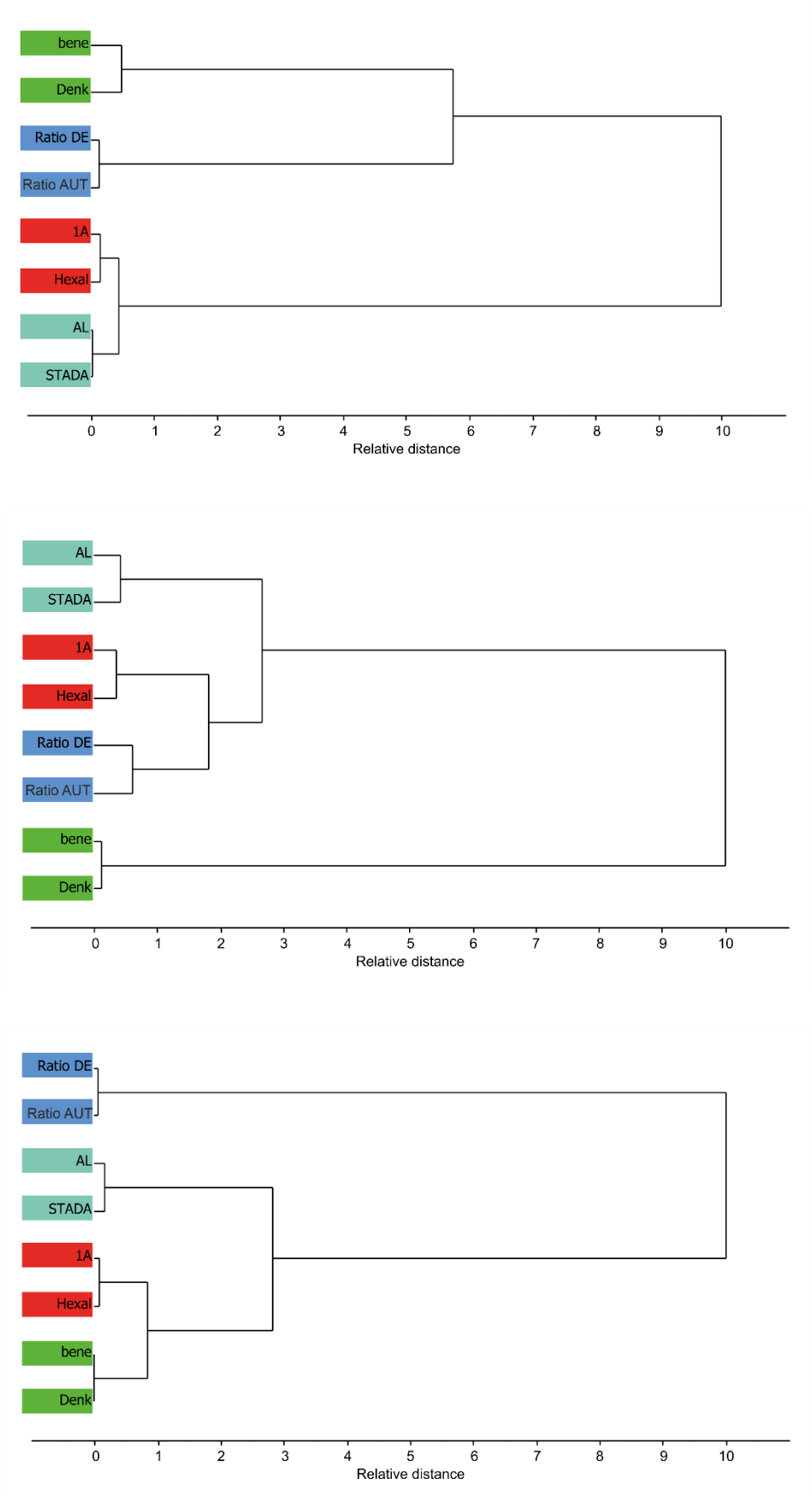


Figure 2 HCAs of the samples with same manufacturer: for MIR (top), NIR (middle) and NMR (bottom) data (blue, Merckle GmbH; turquoise, STADA Arzneimittel AG; red, Salutas Pharma GmbH; green, bene-Arzneimittel GmbH).

The samples from cluster A, which were also used to test the suitability of the HCA method, also formed clusters when all samples were evaluated. In the MIR and NMR data (see Figure 3 and 5), even a larger cluster formed with the four German formulations of bene-Arzneimittel GmbH, Denk Pharma GmbH, Hexal AG and 1A Pharma GmbH. In the HCA of NMR data (Figure 5), the preparation of Dr. Max Pharma Ltd. additionally belongs to this cluster. One reason for this could be the very similar composition to 1A and Hexal, as well as the increased proportion of stearic acid in the bene, Denk and Dr. Max preparations, which was already noted in subchapter 3.1. However, the larger cluster of preparations manufactured in Germany that was evident in the PCA scores plot of the NIR data could no longer be identified in the corresponding HCA.

The same samples of Cluster B, which were bought in Tanzania but manufactured in India, formed a cluster in the HCA of the MIR data (Figure 3) together with Keko Pharmaceutical Industries Ltd., pharmastandard JSC (from Russia) and Aspar pharmaceuticals Ltd. (from Great Britain). HCA of the NIR data also recovered the corresponding cluster from the PCA. Whereas Agog Pharma Ltd. and Astra Lifecare (India) Pvt. Ltd. formed an own cluster and Nestor Pharm. Lim. was in another one, which however has a very small relative distance to it. For the NMR data, cluster B could not be recovered. Of note is that all preparations that could belong to cluster B fall into a larger cluster at a relative distance of four, although the decisive criterion for this grouping is not clear.

Clusters C and D were each seen in only one of the spectral methods, NIR and MIR, respectively, and this was also the case with HCA. The two preparations of cluster C, Elys Chemical Industries Ltd. and Regal Pharm. Ltd. formed a cluster with Inter Pharm Arzneimittel GmbH, which was not the case with PCA (Figure 4). Similar behavior was observed for cluster D. The same samples formed a cluster in HCA (Figure 3), but had further samples added to the group. In addition to the two preparations from Ratiopharm GmbH, which were also present in close proximity in PCA, pharmastandard and Aspar pharmaceuticals Ltd. were now also included. The two samples from Elys Chemical Industries Ltd. and North China Pharm, which were only in the vicinity when analyzed by means of PCA, were no longer part of the cluster, indicating larger differences between the samples.

The Italian preparations (cluster E) clustered with the same behavior as in the PCA. This reinforces the connection between these preparations already found in subchapter 3.1, which could explain the similarities. The Unifarm Group, to which Farmakopea SpA belongs, also produces for PT. Actavis Indonesia and Angelini ACRAF SpA [5].

The cluster of Indonesian samples (cluster F) could no longer be observed in any of the three HCAs. The samples were distributed over all the clusters that formed. Only in the HCA of NMR data, PT. Interbat Pharmaceutical and P.t. Sanbe Farma formed a cluster with Actavis (Poland)

showing a large relative distance from the other samples. However, this is more likely to be due to the fact that all these samples contain a higher proportion of cellulose derivatives [5].

The results for the different Panadol® formulations were also quite similar to those of the PCA. In the HCA of the NIR und MIR data, the Advance/Optizorb formulations were separated as an own cluster. This was also true for the ActiFast formulation. The standard formulation did not form a cluster for itself in the NIR data but with several other samples. However, looking at the scores plot of the corresponding data (see Figure 4 in subchapter 3.1), this formulation could not form a separate cluster there either and overlapped with other samples. The HCA of MIR data, on the other hand, showed an own cluster for the standard formulation in addition with sanofi-aventis, S.A. (Spain), which also can be seen in the corresponding scores plot (Figure 5 in subchapter 3.1). Similar observations also applied to the NMR data. There, at a relative distance of six, all three formulations were in a separate cluster. However, only the Advance/Optizorb formulation formed an individual separate cluster with a very small relative distance, even if the preparation of Elys was included. The two standard formulations were even further apart here than in the scores plot.

Furthermore, one entirely new cluster was found in the HCA of the NIR data. This contained the two preparations Crescent and Square, both of which were purchased in Bangladesh.

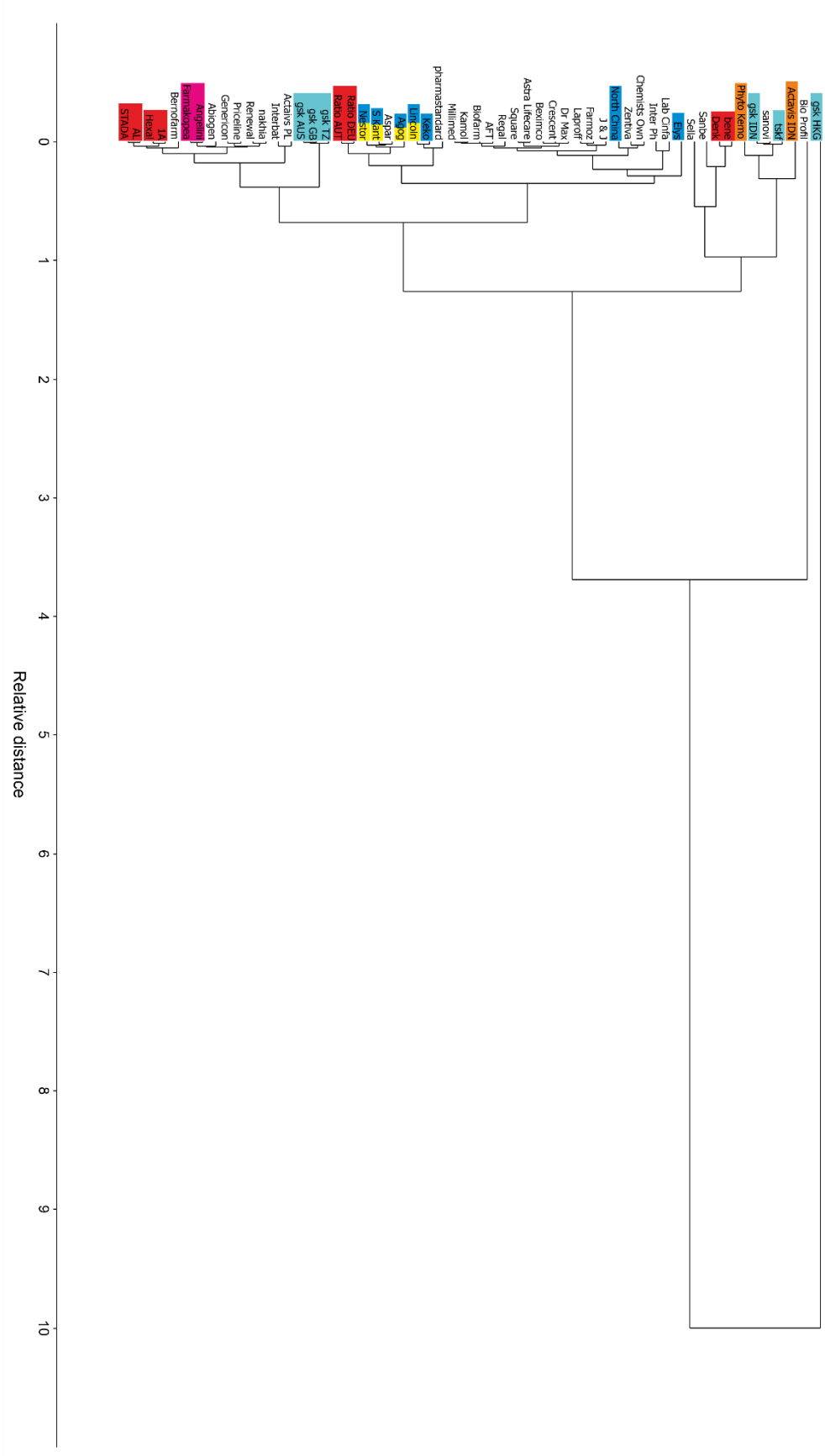


Figure 3 HCA of the MIR data with all samples. Samples were colored according to their membership in the clusters from PCA in subchapter 3.1 (red, Cluster A; yellow, Cluster B; blue, Cluster D; pink, Cluster E; orange, Cluster F; turquoise, Cluster G).

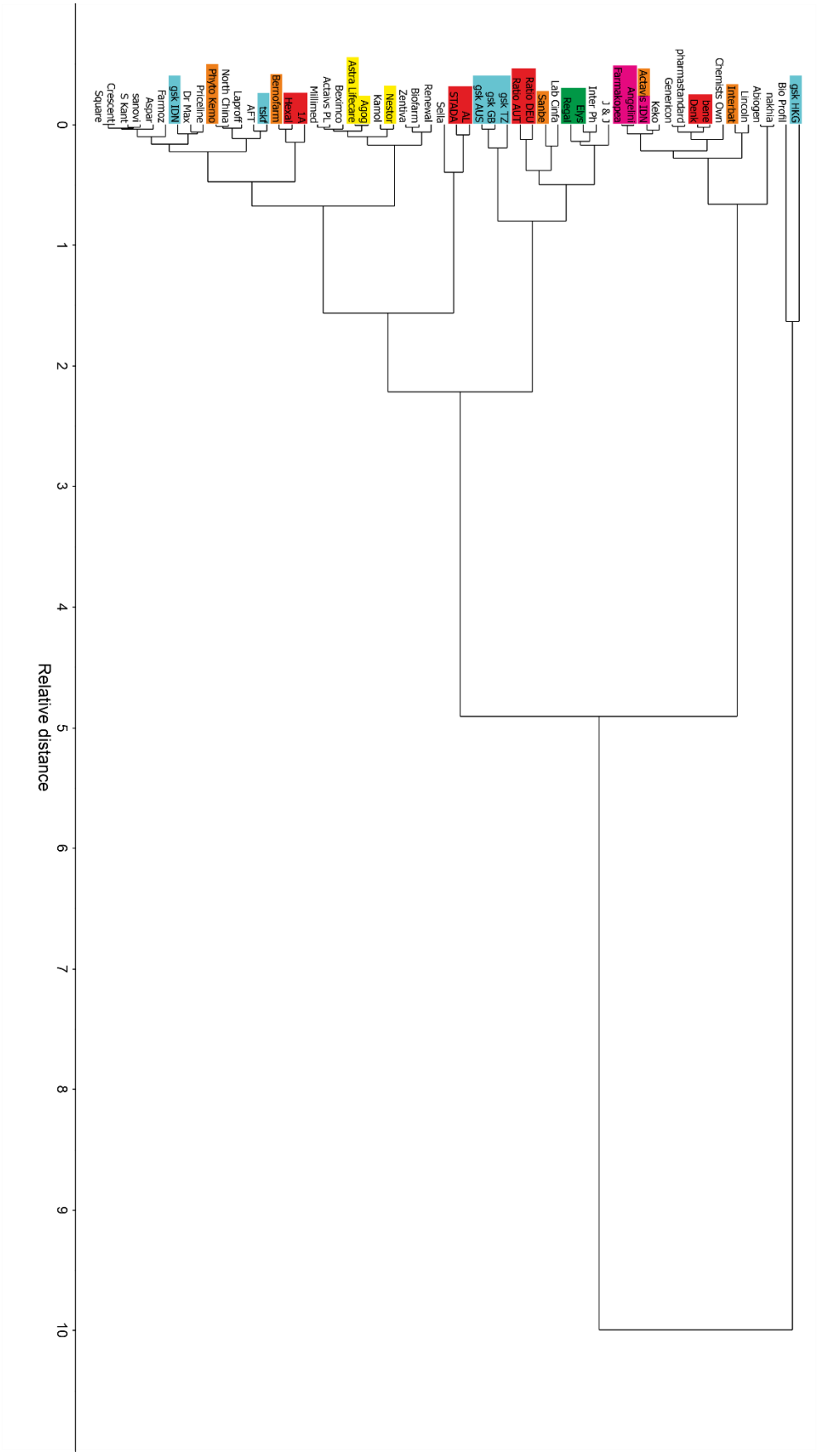


Figure 4 HCA of the NIR data with all samples. Samples were colored according to their membership in the clusters from PCA in subchapter 3.1 (red, Cluster A; yellow, Cluster B; green, Cluster C; pink, Cluster E; orange, Cluster F; turquoise, Cluster G).

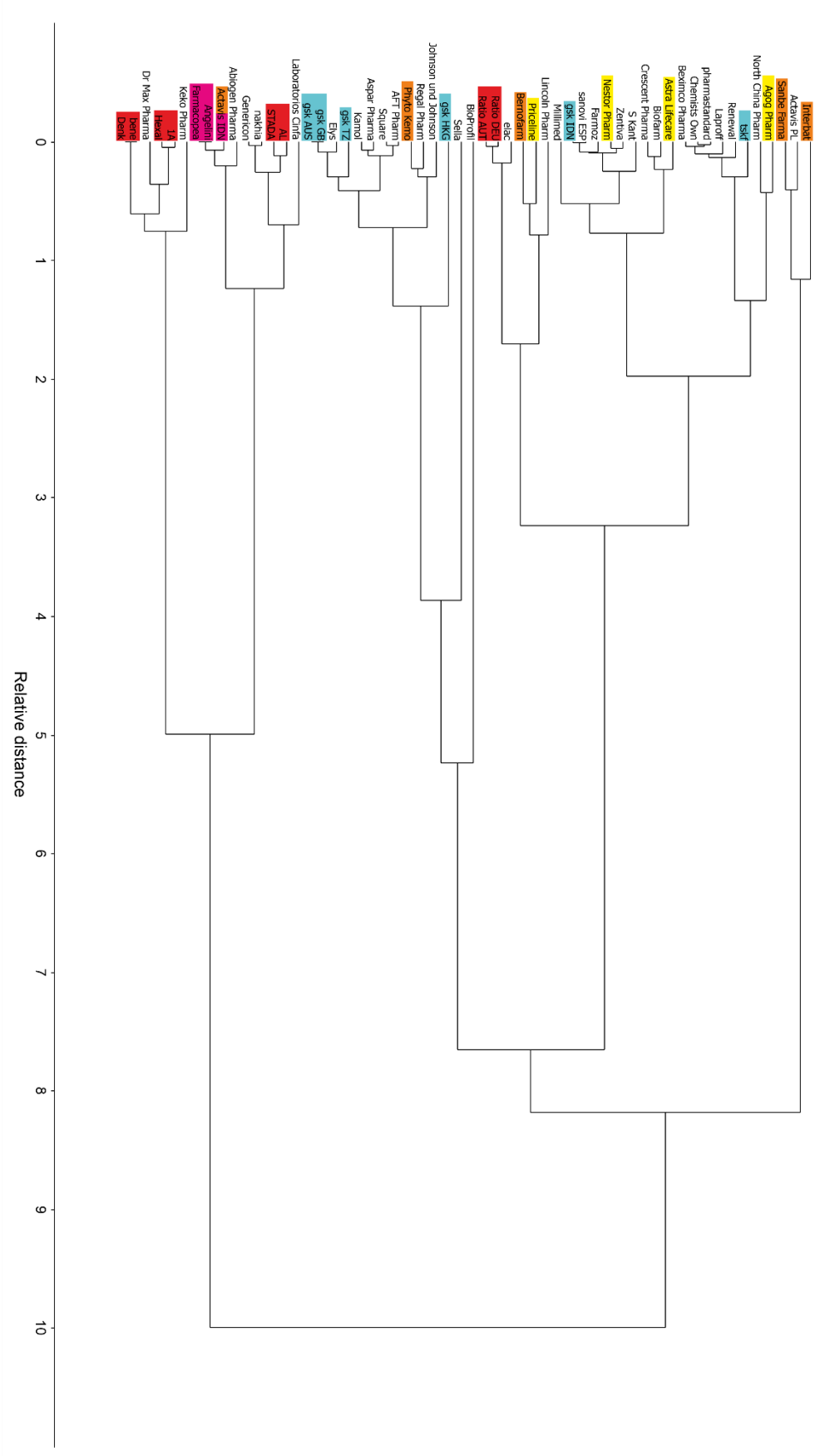


Figure 5 HCA of the NMR data with all samples. Samples were colored according to their membership in the clusters from PCA in subchapter 3.1 (red, Cluster A; yellow, Cluster B; pink, Cluster E; orange, Cluster F; turquoise, Cluster G).

4. Conclusion

The HCA of the different spectral data could mostly confirm the results of the PCA. All of the previously found clusters were found with this procedure, with only one exception. However, it could not always be plausibly explained why some preparations were additionally represented in some other clusters. In most cases, however, these preparations already showed a proximity to the clusters in question in the scores plots of the PCAs. Nevertheless, the HCA was also able to show new clusters and correlations, which could be partially explained by previously obtained information (e.g., with an increased proportion of stearic acid or cellulose derivatives). In addition, it shows new clusters whose reason for formation is still unknown and should be looked at more closely. Yet, these reasons are not directly apparent, since on the one hand the exact compositions of the preparations are not known, which is a very probable reason for the formation of some clusters (see NMR data) and on the other hand further possibilities for analysis, such as the loadings plot of the PCA, are not available.

References

[1] Reh E (2017) Chemometrie : Grundlagen der Statistik, Numerischen Mathematik und Software Anwendungen in der Chemie. Walter de Gruyter GmbH, Berlin/Boston, UNITED STATES

[2] Backhaus K, Erichson B, Plinke W, Weiber R (2015) Multivariate Analysemethoden Eine anwendungsorientierte Einführung

[3] Wentura D, Pspeschill M (2015) Multivariate Datenanalyse Eine kompakte Einführung

[4] Tutorial K: Clustering (2020). Unscrambler 11.0. *CAMO Software AS.*,

[5] Becht A, Schollmayer C, Monakhova Y, Holzgrabe U Tracing the origin of paracetamol tablets by near-infrared, mid-infrared, and nuclear magnetic resonance spectroscopy using principal component analysis and linear discriminant analysis. *Anal Bioanal Chem* (2021) 413, 3107-3118

3.3 Designing an NIR attachment using a 3D printer to measure tablets and distinguish batches from multiple pharmaceutical companies

1. Introduction

NIR spectroscopy has established itself in the pharmaceutical industry through process analytical technology (PAT) not only in the control of manufacturing processes or identity control of incoming raw materials like active ingredients and excipients, but also in final products, e.g. of tablets [1]. A major advance here was the miniaturization of the NIR devices, which enabled mobile use. Coupled with a tablet, analyses can thus be carried out directly on site [2]. However, such devices can be used not only to check the identity of starting materials in production, but also, for example, to detect counterfeit drugs [3]. Tablets can be measured without further sample preparation, as the spectra contain not only chemical information but also physical information such as compression pressure and particle size [4]. Even with the same composition, tablets can be distinguished from one another this way (see Figure 1).
Hence, NIR was used to test whether it is possible to distinguish individual batches of the same finished drug preparation. For this purpose, an attachment was developed that allows for direct measurement of tablets with the existing portable NIR device, which is initially built for the measurement of powders and liquids. Its applicability was tested using two different batches of paracetamol monopreparations from nine different pharmaceutical companies.

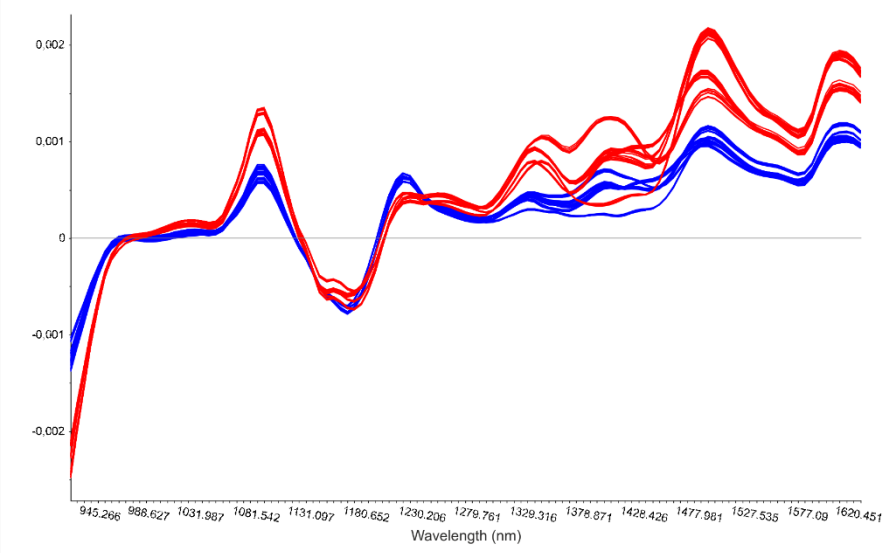


Figure 1 Influence of the dosage form on the NIR spectra. Red spectra are from tablets and blue spectra are from the corresponding powders (mortared). Samples 7, 8, 43, 48, 47 and 44 were used. For better visualization of the differences, SNV correction was performed and the 1St derivative (Savitzky-Golay) was used.

2. Experimental

2.1 3D printing of the attachment

The attachment was designed using an individually self-made sketch so that the various tablets can be measured in the best possible way. Therefore, various points were taken into account such as the accuracy of fit on the NIR device, different shapes of the tablets, light source and detector position on the NIR device, light protection and the reflection standard. A Prusa i3 MK2.5 (0.4 mm nozzle) 3D printer was then used with the following settings:

Printing process	: Fused Deposition Modeling (FDM)
Nozzle temperature	: 245 °C
Printing temperature	: 80 °C
Printing speed	: 80 mm/s (average value)
Filament	: PETG Black (eSUN / Prusament)

The finished attachment consists of a plate with an opening on which the tablets are placed, and a removable ring to serve as a light protection (Figure 2).

Figure 2 Individual parts of the attachment (left) and the setup during the measuring process (right)

2.2 Drug samples

Two batches of every investigated paracetamol formulation (from nine different pharmaceutical companies) were purchased from local pharmacies (Table 1). They were of different shapes and sizes and contained 500 mg of paracetamol, with the exception of Efferalgan (sample no. 15 & 15a) and Doliprane Tabs (sample no. 63 & 63a), which contained 1000 mg per tablet.

Table 1 Listing of all paracetamol samples which were used for batch comparison

Sample No.	Name	Pharmaceutical Company	batch
1	Panadol	GlaxoSmithKline plc.	135403
1a	Panadol	GlaxoSmithKline plc.	134745
7	Paracetamol 500	1A Pharma GmbH	FN6315
7a	Paracetamol 500	1A Pharma GmbH	HL0517
8	Paracetamol ratiopharm	Ratiopharm GmbH	P42467
8a	Paracetamol ratiopharm	Ratiopharm GmbH	S45886A
15	Efferalgan	UPSA SAS	R5400
15a	Efferalgan	UPSA SAS	T2821
43	Paracetamol	STADA Arzneimittel GmbH	53542
43a	Paracetamol	STADA Arzneimittel GmbH	71702
44	Paracetamol AL 500	Aliud Pharma GmbH	54449
44a	Paracetamol AL 500	Aliud Pharma GmbH	71706
47	ben-u-ron	bene-Arzneimittel GmbH	606B161
47a	ben-u-ron	bene-Arzneimittel GmbH	611G171
48	Paracetamol Hexal	Hexal AG	GC7080
48a	Paracetamol Hexal	Hexal AG	HC9249
63	Doliprane Tabs	sanofi-aventis france	7V111
63a	Doliprane Tabs	sanofi-aventis france	7V018

2.3 NIR spectroscopy

The samples were analyzed with a Micro NIR™ 1700 ES spectrometer (VIAVI Solutions Deutschland GmbH, Eningen unter Achalm, Germany) covering a spectral range from 950 – 1650 nm. The integration time was set to 22.7 ms and the scan count to 100. The temperature range in which the spectra were recorded was 38 °C to 40.7 °C.

Without further sample preparation, the tablets were placed on the gap of the attachment so that they cover it as completely as possible. It needs to be ensured that the score line always points upwards, as it otherwise serves as a source of deflection impairing the scattering and thus the measurement process. Twelve spectra of each tablet were recorded, with the tablet being oriented slightly different by means of rotation after each measurement. The dark current and reference scan (99% reference standard) were remeasured before each tablet.

2.4 Multivariate and statistical tests

Principal component analysis (PCA) is a multivariate tool which is capable of reducing the data dimensionality to a few principal components and enables a better visualization of the data [5]. It was used to better evaluate the measurement parameters, to get an overview of the data and finally to perform the statistical tests by means of the score values. For PCAs, the entire recorded spectral range was analyzed and the singular value decomposition algorithm was

used for calculation. The internal validity was checked by cross-validation. In order not to lose physical information such as particle size or pressing pressure, no pre-processing methods were performed. The PCAs and the statistical tests, such as the Kolmogorov-Smirnov test, the Student's *t*-test for equal and unequal variance (two-tailed) and the *F*-test, were performed with the Unscrambler X 10.4 (CAMO Software AS., Oslo, Norway) at 5% significance level. The free software R Studio 1.1.383 (RStudio Incorporation, Boston, USA) running R 3.6.3 (R Foundation for Statistical Computing, Vienna, Austria) was used for the Wilcoxon rank-sum test.

3. Results and discussion

3.1 3D print

Polyethylene terephthalate modified with glycol (PETG) was used as filament. This is mechanically more stable than the otherwise frequently used PLA (polylactide) and is nevertheless easier to print than other engineering plastics. Disadvantages include stringing (formation of threads) and unclean overhangs [5]. Furthermore, the orientation of the components during printing was chosen to make the bottom surface facing the NIR device as smooth as possible to avoid possible measurement errors.

3.2 NIR experiments

Before the final measurements of the tablets, some parameters were checked to obtain reproducible and comparable spectra. These include the alignment of the NIR instrument during the dark current scan, measurements with and without lid, the influence of temperature, the alignment of the attachment, and the reapplication of the tablet before each measurement. Additionally, the orientation of the tablet was tested as it can have different shapes (round, oval, planar, convex), engravings and score lines that have an influence on the reflection. For this purpose, tablets of the German samples, 1A Pharma, Aliud Pharma, Ratiopharm, STADA, Hexal and ben-u-ron were measured and the spectra as well as their PCAs were evaluated.

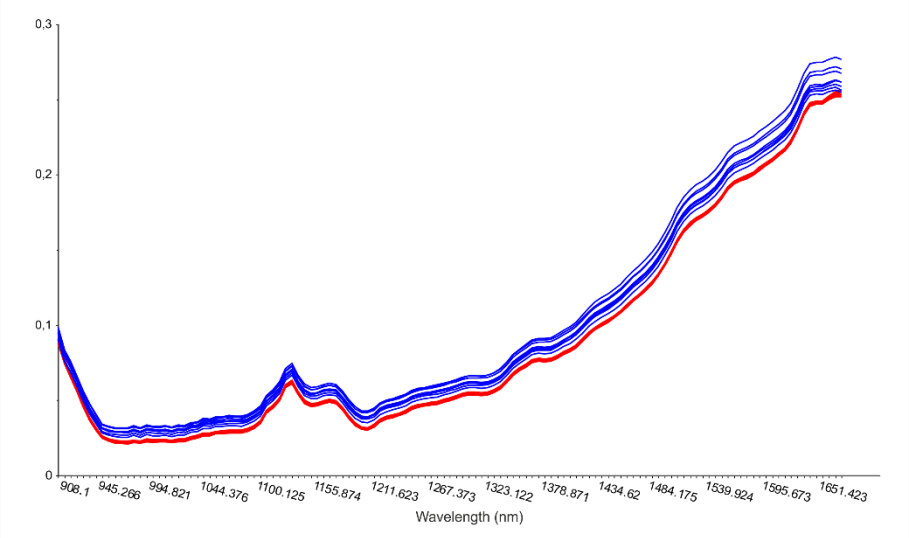


Figure 3 NIR spectra of STADA batch 53542 measured with the score line pointing upwards (red; n = 12) and downwards (blue; n = 12) with reapplication before each measurement. The shift of the spectra due to the score line is clearly visible.

Some parameters had only a minimal influence on the spectra, such as the lid or the shape of the tablet, provided it was placed correctly. This required the tablets to be placed on the measurement gap so that they cover it as completely as possible, with the score lines pointing upwards. This prevents further scattering of the infrared light at the score lines, which otherwise cause minimal shifts of the spectra (see Figure 3) and can be avoided this way. For the dark current scan, it proved to be optimal when the NIR instrument was pointed towards the floor, as was also indicated in the manual when using the NIR instrument normally without the attachment.

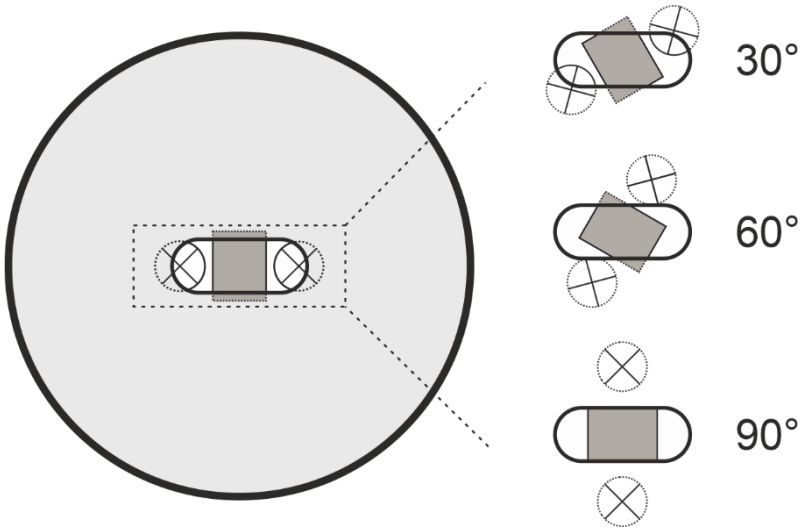


Figure 4 Schematic top view of the NIR attachment. Shown is how far the attachment covers the detector (grey box) and the lamps at different orientations. At an angle of 90°, the lamps are completely concealed.

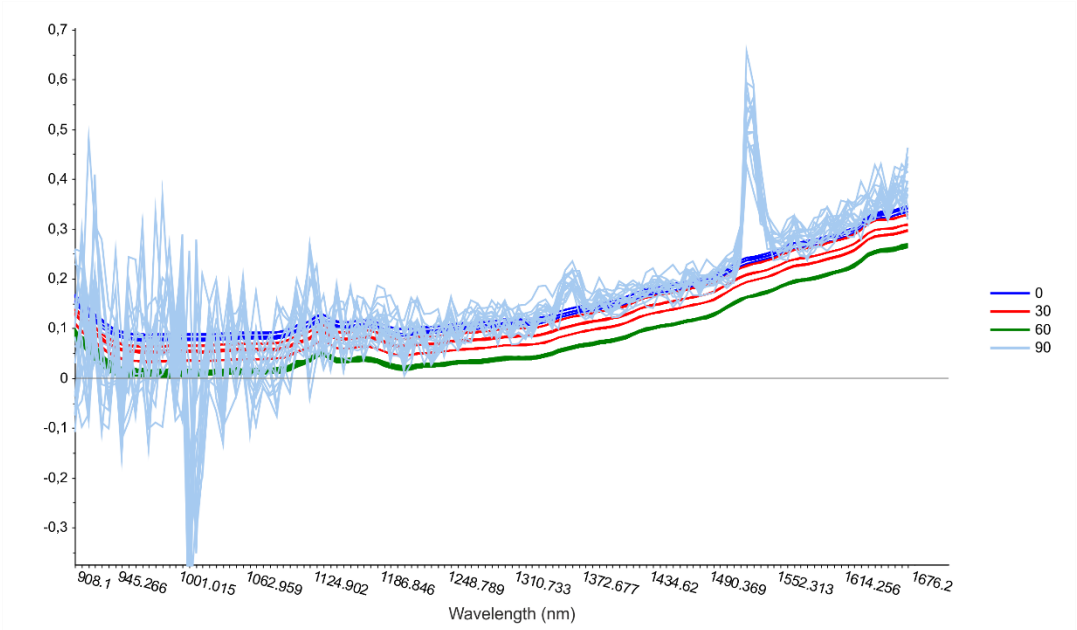


Figure 5 Influence of the orientation of the attachment on the NIR spectra of Ratiopharm. The shifts are stronger the more of the detector is covered by the attachment. The different orientations in degrees between the measurement gap and detector are highlighted in color (0°: dark blue; 30°: red; 60°: green; 90°: light blue).

The temperature of the NIR device, the orientation of the attachment and repeated measurements without repositioning the tablet had a stronger influence. Since the attachment only exposes a small measurement gap, exact care had to be taken to orient the attachment in parallel to the detector and lamps. As more of the detector and lamps were covered (Figure 4), the shifts in the spectra increased until only noise could be detected at an orientation of 90° (Figure 5). Also, different temperatures of the NIR instrument led to different shift intensities in the spectra (Figure 6), which also resulted from the fact that the instrument heated up more severely, when used with the attachment in comparison to the standard use. Therefore, before starting the series of measurements, the temperature was to settle between 38 °C and 40.7 °C. In direct successive measurements, without moving the tablet, a linear drift of the score values was always visible in the PCAs (see Figure 7). Although the score values of the spectra were closer to each other in the scores plot without repositioning, such drifts, which were caused by the same position of the tablet in each of the measurements, should be avoided. These score values do not reflect the variability of the NIR spectra between measurements obtained by reapplying the tablet. This is particularly important for the subsequent creation of models where, for example, unknown tablets shall be assigned to a specific batch. In these measurements, it would never be possible to place the tablets on the NIR attachment exactly as was done during method development. To exclude this effect, the tablets were repositioned before each measurement.

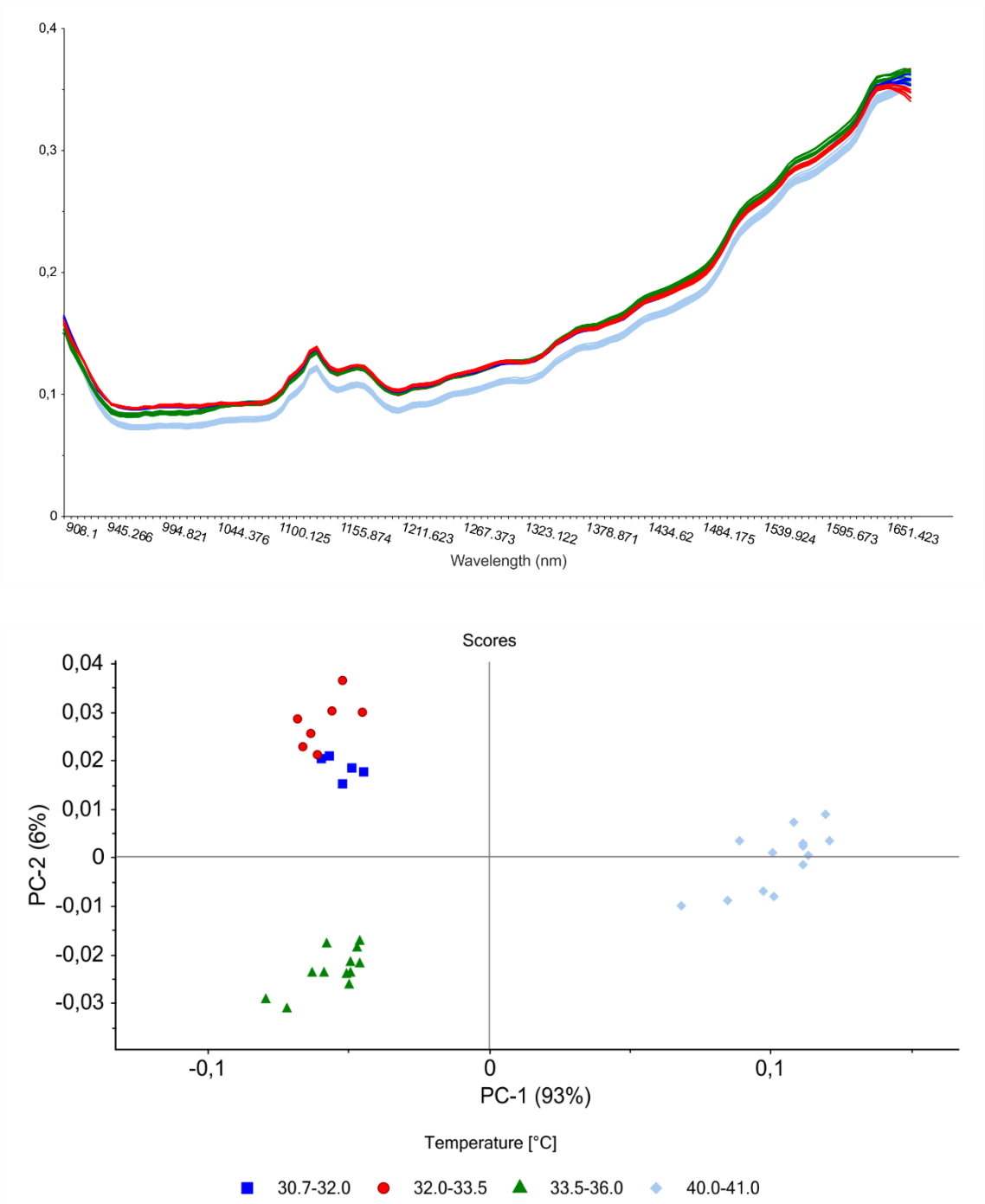


Figure 6 NIR spectra of the Hexal batch GC 7080 (top) and their corresponding PCA (bottom). The influence of temperature on the shift of the spectra is clearly visible. The same applies to the groupings for the PCA (dark blue: 30.7-32.0 °C; red: 32.0-33.5 °C; green 33.5-36.0 °C; light blue: 40.0-41.0 °C).

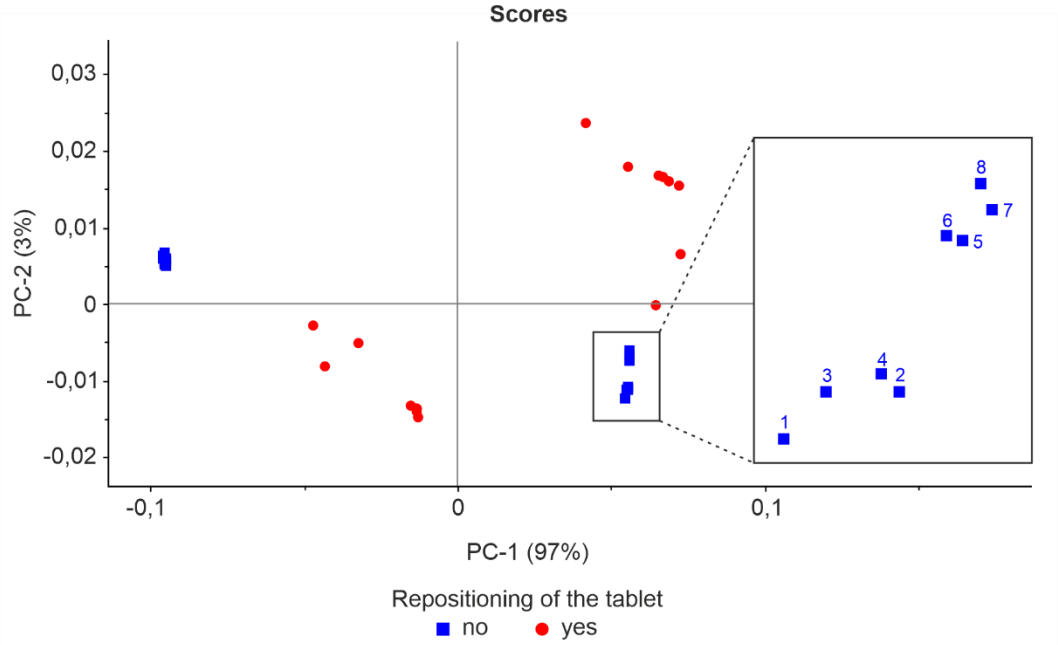


Figure 7 PCA of STADA batch 71702. Samples with negative correlation to PC1 were measured on the score line, with positive correlation on the other side of the tablet. The blue samples were measured without repositioning the tablet, resulting in a trend in the scores plot. This can be seen well in the enlargement (the numbers correspond to the measurement sequence).

3.3 Principal component analysis

To get an overview of the data, a PCA was first performed for all manufacturers. Since a separation between the batches could be seen in the scores plot (Figure 8) a PCA was subsequently created for the batches of each individual pharmaceutical company.

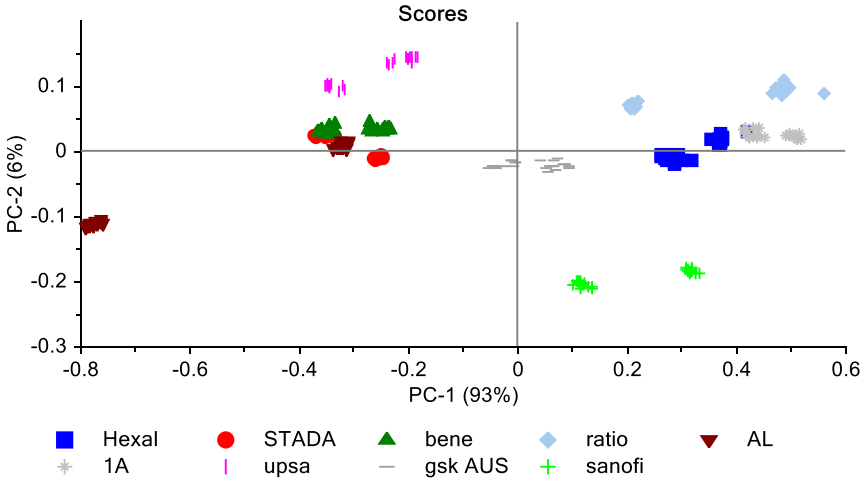


Figure 8 Scores plot with different colored symbols for all nine pharmaceutical companies. For each of the pharmaceutical companies, two groups can be identified, corresponding to the two different batches.

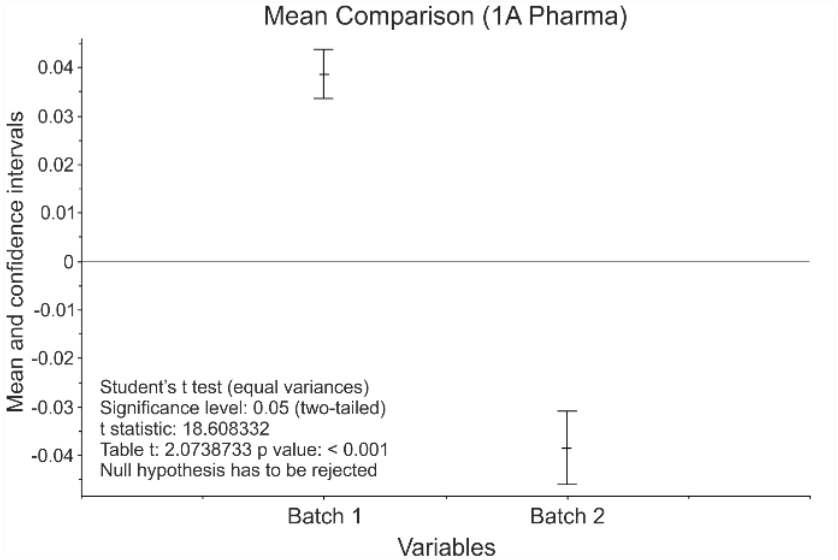
The scores plots were then examined to see if the batches could be distinguished based on a single principal component and if there were any outliers. This was the case for all nine pharmaceutical companies, although one spectrum for Ratiopharm batch P42467 and

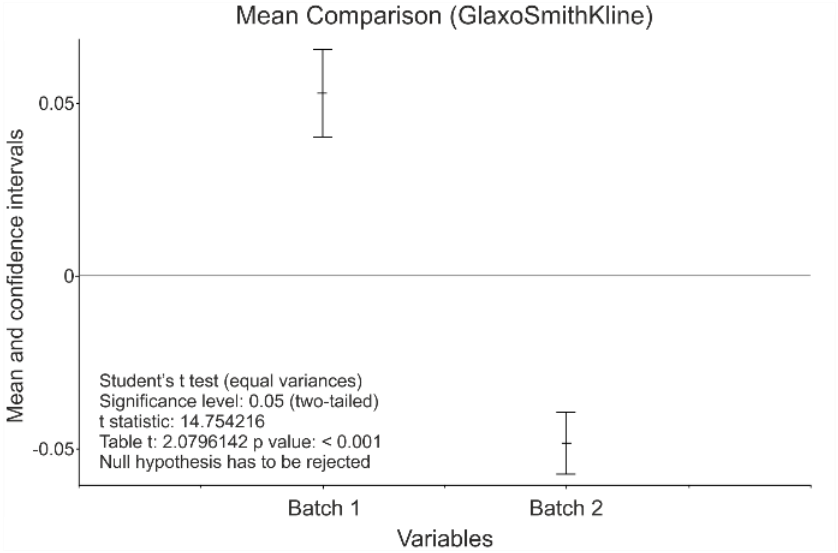
GlaxoSmithKline batch 135403 was excluded from the statistical evaluation because they showed a strong shift compared to the other spectra for unknown reasons (outliers).

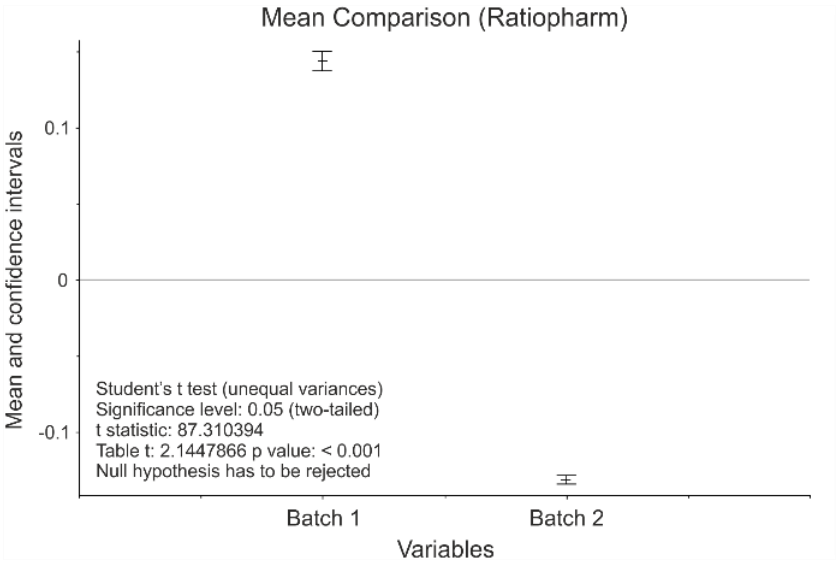
3.4 Statistics

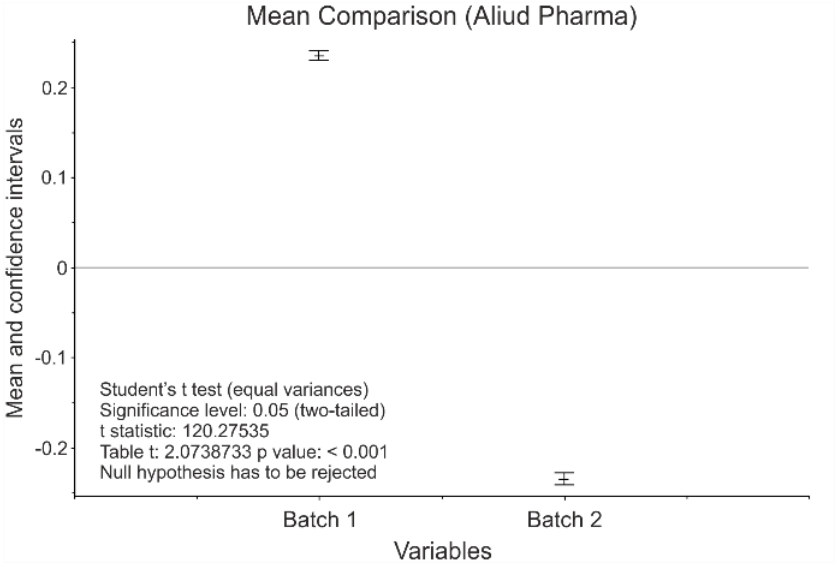
Besides the data reduction and the better visualization of the data, the score values of the first principal component can also be used for statistical purposes. The prerequisite for this is the complete differentiation of the batches by means of this main component which was shown in the scores plots of the PCAs in section 3.3.3.

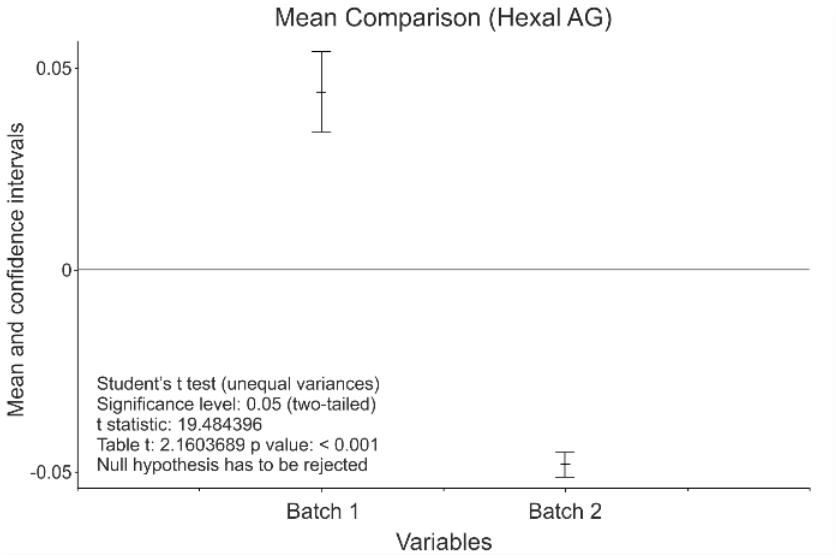
First, the score values of each batch were tested for normal distribution using a Kolmogorov-Smirnov test. This test compares the cumulative distribution function with the hypothetical cumulative distribution function of the values in a normal distribution [7]. Except for Efferalgan (15 & 15a) and ben-u-ron (47 & 47a), all data from the other pharmaceutical companies were normally distributed. With these, in the second step a *F*-test was performed to see if the two batches from each pharmaceutical company had comparable variances. This was necessary to choose the appropriate Student's *t*-test assuming either equal or different variance between batches. Since the variances of the Hexal, STADA and Ratiopharm batches could not be assumed to be equal, the Student's *t*-test for unequal variances was used for them and the Student's *t*-test for equal variances was used for Aliud Pharma, 1A Pharma, GlaxoSmithKline and sanofi aventis. However, the results of the corresponding *t*-tests were the same for all pharmaceutical companies. The p-values were smaller than the significance level (0.05) for every test and therefore the null hypothesis, that there is no difference in the means, had to be rejected. This means that the sets of score values are significantly different and so are the batches (see Figure 9).

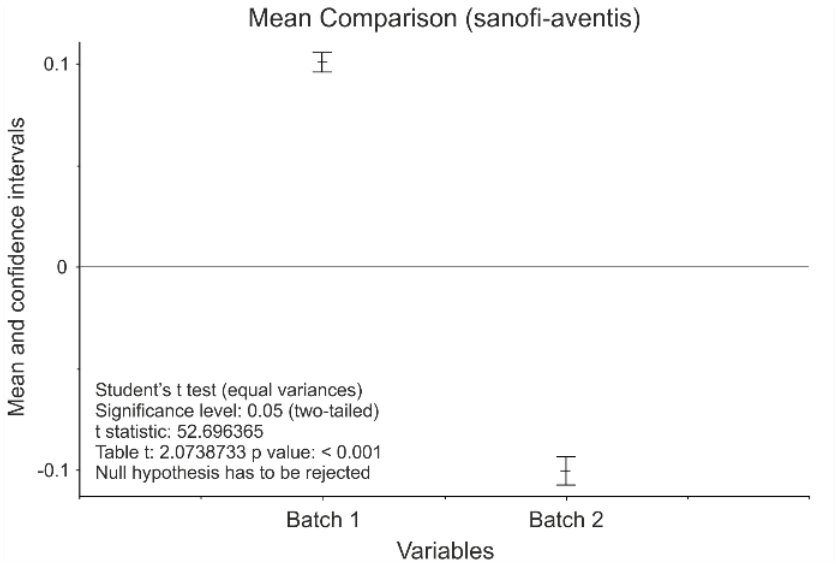

Mean Comparison (1A Pharma)
Mean and confidence intervals
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
Student's t test (equal variances)
Significance level: 0.05 (two-tailed)
t statistic: 18.608332
Table t: 2.0738733 p value: < 0.001
Null hypothesis has to be rejected
Batch 1
Batch 2
Variables


Mean Comparison (GlaxoSmithKline)
Mean and confidence intervals
0.05
0
-0.05
Student's t test (equal variances)
Significance level: 0.05 (two-tailed)
t statistic: 14.754216
Table t: 2.0796142 p value: < 0.001
Null hypothesis has to be rejected
Batch 1
Batch 2
Variables


Mean Comparison (Ratiopharm)
Mean and confidence intervals
0.1
0
-0.1
Student's t test (unequal variances)
Significance level: 0.05 (two-tailed)
t statistic: 87.310394
Table t: 2.1447866 p value: < 0.001
Null hypothesis has to be rejected
Batch 1
Batch 2
Variables


Mean Comparison (Aliud Pharma)
Mean and confidence intervals
0.2
0.1
0
-0.1
-0.2
Student's t test (equal variances)
Significance level: 0.05 (two-tailed)
t statistic: 120.27535
Table t: 2.0738733 p value: < 0.001
Null hypothesis has to be rejected
Batch 1
Batch 2
Variables


Mean Comparison (Hexal AG)
Mean and confidence intervals
0.05
0
-0.05
Student's t test (unequal variances)
Significance level: 0.05 (two-tailed)
t statistic: 19.484396
Table t: 2.1603689 p value: < 0.001
Null hypothesis has to be rejected
Batch 1
Batch 2
Variables


Mean Comparison (sanofi-aventis)
Mean and confidence intervals
0.1
0
-0.1
Student's t test (equal variances)
Significance level: 0.05 (two-tailed)
t statistic: 52.696365
Table t: 2.0738733 p value: < 0.001
Null hypothesis has to be rejected
Batch 1
Batch 2
Variables

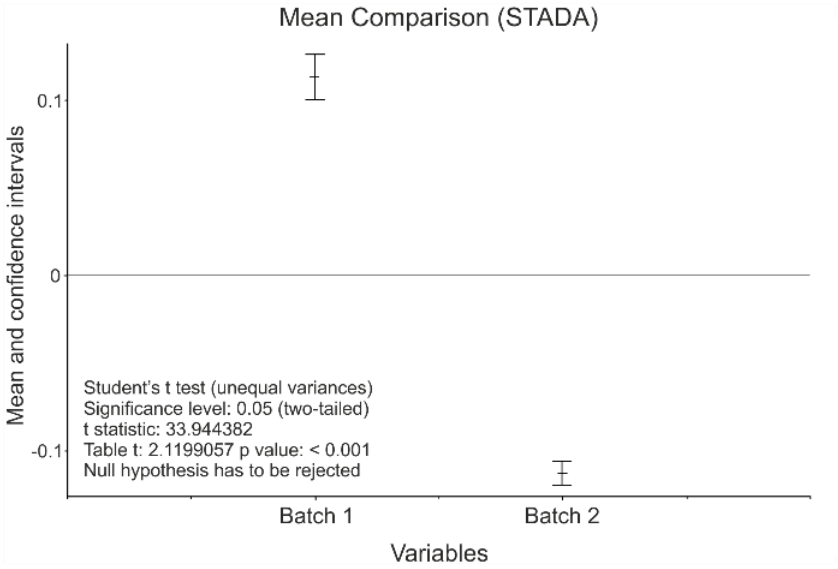


Figure 9 The results of the Student's t-tests are presented as mean comparison plots. It shows the mean value and its 95% confidence interval for each batch. It allows a visual comparison of the means of the scores for each batch and shows if the spread of the scores is the same. As can be seen for each pharmaceutical company, the batches differ significantly (no overlap of confidence intervals).

Since a normal distribution could not be assumed for Efferalgan and ben-u-ron, a Wilcoxon rank-sum test had to be applied [8]. The null hypothesis had to be rejected here as well, which means that the score values of one batch tended to be larger or smaller than those of the other batch. This result was also consistent with those of the other pharmaceutical companies, thus demonstrating a significant difference between batches.

4. Conclusion

For each pharmaceutical company, it was possible to distinguish the two batches by means of a single principal component without having to perform any sample preparation or special processing of the spectra beforehand. This result could also be supported with the statistical tests, as for each pharmaceutical company a significant difference between the batches could be demonstrated.

It should be noted that these measurement results could only be compared with each other if all measurements are performed consecutively and attention is paid to some critical measurement parameters. This concerns temperature fluctuations of the lamp which can be avoided by not switching off the lamp between measurements and waiting for a constant operating temperature. In addition, the tablet should be repositioned before each measurement to avoid trends. Finally, the series of experiments should be carried out in one piece, as the attachment cannot be locked in any particular position. As a result, the measuring gap assumes a slightly different position each time the attachment is removed and installed, which can lead to differences in the spectra.

With the present attachment and measuring method, it was possible to distinguish individual batches within one measuring process. However, to establish this in a larger scale such as in the pharmaceutical industry, more samples and measurements would have to be performed. In addition, the attachment has to be further optimized to obtain a truly robust and reliable method, as it is not optimally adapted to each tablet shape in the current variant.

Acknowledgements

Special thanks to Simon Luposchainsky, who designed the geometric model of the NIR attachment on the computer and then printed it, and to Bettina Gerner for help with the Wilcoxon rank sum test.

References

[1] Roggo Y, Chalus P, Maurer L, Lema-Martinez C, Edmond A, Jent N A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *J Pharm Biomed Anal* (2007) 44, 683-700

[2] Alcalà M, Blanco M, Moyano D, Broad NW, O'Brien N, Friedrich D, Pfeifer F, Siesler HW Qualitative and Quantitative Pharmaceutical Analysis with a Novel Hand-Held Miniature near Infrared Spectrometer. *J Near Infrared Spectrosc* (2013) 21, 445-457

[3] Scafi SHF, Pasquini C Identification of counterfeit drugs using near-infrared spectroscopy. *Analyst* (2001) 126, 2218-2224

[4] Matero S, van Den Berg F, Poutiainen S, Rantanen J, Pajander J Towards better process understanding: chemometrics and multivariate measurements in manufacturing of solid dosage forms. *J Pharm Sci* (2013) 102, 1385-1403

[5] Biancolillo A, Marini F Chemometric Methods for Spectroscopy-Based Pharmaceutical Analysis. *Front Chem* (2018) 6

[6] Prusa Knowledge Base | PETG. (2021) Prusa Research a.s. https://help.prusa3d.com/en/article/petg_2059. Accessed 13.07.2021

[7] Antoneli F, Passos FM, Lopes LR, Briones MRS A Kolmogorov-Smirnov test for the molecular clock based on Bayesian ensembles of phylogenies. *PLoS One* (2018) 13, e0190826-e0190826

[8] Rosner B, Glynn RJ, Ting Lee M-L Incorporation of Clustering Effects for the Wilcoxon Rank Sum Test: A Large-Sample Approach. *Biometrics* (2003) 59, 1089-1098

3.4 Diffusion ordered NMR spectroscopy measurements as screening method of potential reactions of API and excipients in drug formulations

Alexander Becht, Curd Schollmayer, Johannes Wiest, Detlef Heller, Wolfgang Baumann, Helmut Buschmann, Ulrike Holzgrabe

Copyright

https://doi.org/10.1016/j.jpba.2018.09.020

Abstract

In the development of new pharmaceutical formulations it is important to consider the possible interactions between the active pharmaceutical ingredient (API) and excipients which is a well-known problem. The objective of the work presented here was to investigate such reactions by means of diffusion ordered NMR spectroscopy (DOSY). The known reaction of 5-aminosalicylic acid (5-ASA) and the excipient citric acid was studied. Three reaction products have been verified by DOSY, ^{1}H NMR and HPLC measurements. Despite a poor separation in the DOSY diagram, the reaction products could be assigned due to the processing of thoughtful selected parts of the signals. The reaction of 5-ASA with formic acid and benzocaine with dibutyl phthalate was also studied by means of DOSY experiments.

1. Introduction

It is commonly known that reactions between active pharmaceutical ingredients (APIs) and excipients can occur in pharmaceutical formulations which are of importance especially in the development of new drugs to guarantee their efficacy and safety, and prevent further inconvenience after the approval. The reactions can be manifold, e.g. acid base reactions or Maillard browning. Some examples are given by Bharate et al. [1].

The aim of this study is to demonstrate the capabilities of routine NMR spectroscopy in the field of impurity profiling using several examples. The reaction products between 5-aminosalicylic acid (5ASA), an anti-inflammatory drug, and citric acid were studied. This combination was chosen because it is known to give several reaction products [2,3], and some formulations like Salofalk® Granu-Stix® still contain both substances. The reactions lead to an ester or an amide between the hydroxyl or the amino group of 5-ASA and the carboxylic group of citric acid (see Figure 1) as described by Larsen et al. [3]. Further the reaction of 5-ASA with formic acid and benzocaine, a local anesthetic, and dibutyl phthalate, a plasticizer in film coatings, were investigated [4].

DOSY can be used as a screening method [5] and is described for various applications [6–8]. It makes use of the translational diffusion of molecules. Small changes in structure affect the molecule size and shape and therefore the diffusion behavior. The resulting different diffusion coefficient can be used to separate the spectra of substances in a mixture. The method is based on the decay of signals depending on diffusion time and gradient parameters [9]. It enables the possibility to assign signals to a certain substance with its corresponding diffusion coefficient [10] and to check the purity by means of a reference spectrum. Hence, it is possible to measure the mixture of substances with just one measurement [10].

^{1}H-NMR measurements and DOSY experiments of stressed ASA-citric acid mixtures will be compared with the results of the HPLC method for related substances described in the monograph of mesalazine in the European Pharmacopoea [11].

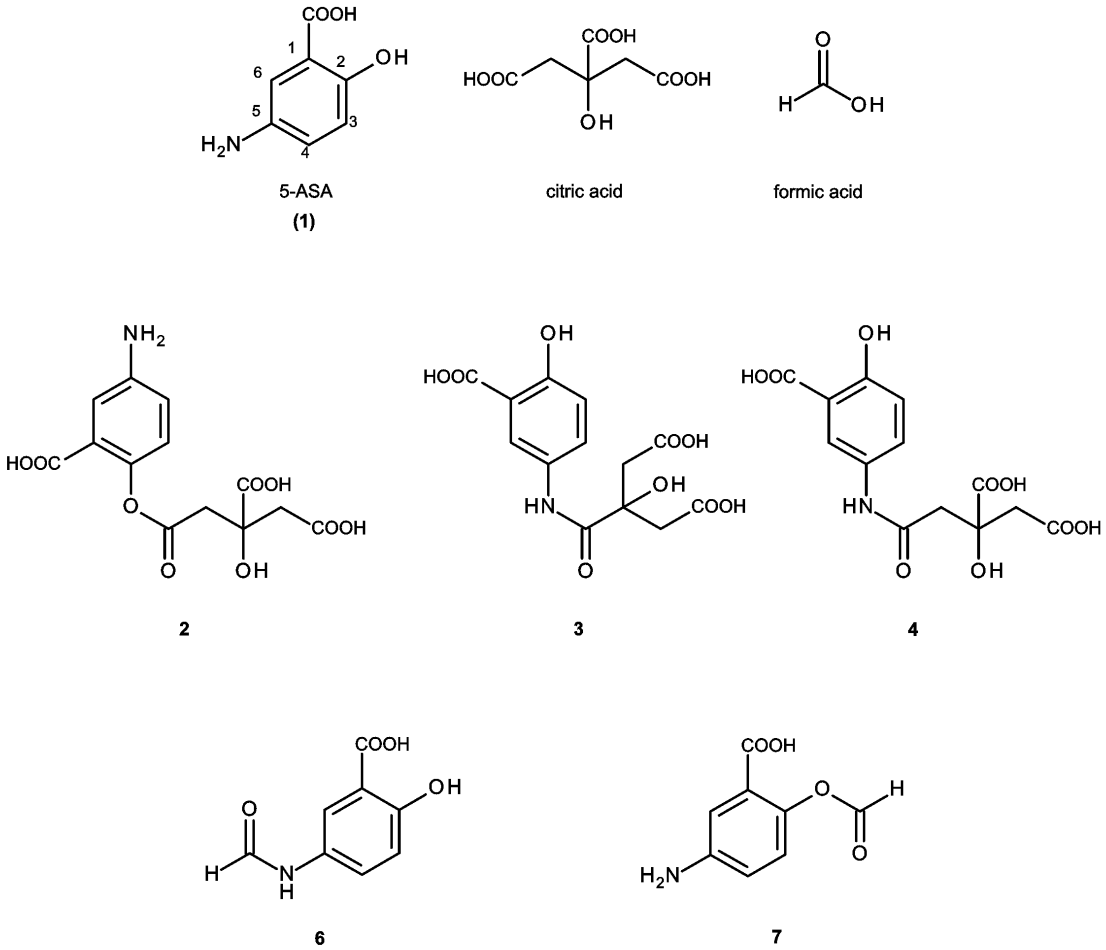


Figure 1 5-ASA, citric acid and their reaction products (**2**, **3**, and **4**) as postulated in [3] and formic acid with reaction products (**6** and **7**)

2. Experimental

2.1 Materials

Citric acid monohydrate, citric acid anhydrous, 5-aminosalicylic acid, maleic acid for quantitative NMR, dibutyl phthalate, formic acid, sodium hydroxide solution 50% for analysis and HPLC grade acetonitrile were purchased from Sigma-Aldrich Chemie GmbH (Steinheim, Germany). Benzocaine was purchased from Caelo (Hilden, Germany). 0.1 M Hydrochloric acid was obtained from VWR international (Darmstadt, Germany). Sodium dihydrogen phosphate monohydrate was purchased from Grüssing GmbH Analytika (Filsum, Germany), deuterium oxide and dimethylsulfoxide-d6 from Deutero GmbH (Kastellaun, Germany). Ultrapure water was produced by a water purification system from Merck Millipore (Schwalbach, Germany) specified at a resistivity of 18.2 MΩ-cm. All NMR samples were measured in 507-HP-7 5 mm routine NMR tubes from Euriso-top (Saarbrücken, Germany).

2.2 Sample preparation

100 mg of 5-ASA and 100 mg citric acid monohydrate or 200 mg of 5-ASA and 200 mg anhydrous citric acid were pestled together. Then, either 1 mL or 2 mL of deuterium oxide was added. 100 mg of 5-ASA and 25 µl of formic acid dissolved in 2 mL of deuterium oxide. The mixtures were equilibrated at 70 °C for three hours. Afterwards, the probes were sonicated for a certain time, centrifuged, and the supernatant was analyzed by NMR and/or HPLC.

In order to simulate the reaction between benzocaine and dibutyl phthalate, a mixture of 0.12 mmol of benzocaine, 0.12 mmol of dibutyl phthalate, and 0.0048 mmol of N-(4-ethoxycarbonylphenyl)phthalimide were dissolved in 2 mL of dimethylsulfoxide-d6 and sonicated for dissolution. The amount of N-(4-ethoxycarbonylphenyl)phthalimide was chosen to give a content of 4%. The sample was analyzed by NMR.

2.3 HPLC experiments and data analysis

The test for related substances described in the European Pharmacopeia (PhEur 8.0 [11]) was performed on an Agilent system from the 1100 series (Waldbronn, Germany) using a reversed-phase column packed with octadecylsilane phase particles (250 x 4.6 mm, 5 µm, ZORBAX Eclipse Plus, Agilent, Waldbronn, Germany). The system consisted of a vacuum degasser (G1322 A), a binary pump (G1312 A), an autosampler (G1313 A), a thermostated column oven (G1316 A), and a diode array detector(G1315B). Mobile phase A consisted of 6.9 g of sodium dihydrogen phosphate monohydrate dissolved in 1000 mL of water. The pH value of 6.2 was adjusted with sodium hydroxide solution. Mobile phase B consisted of acetonitrile and mobile phase A (40:60, v/v). The flow rate was set at 1.0 mL/min, the detection wavelength at 240 nm, and the temperature at 40 °C. Initially, the gradient started with 0% B until the 8th minute, then it was linerarily increased to 15% B from 8 to 20 min, to 75% from 20 to 40 min and to 100% B from 40 to 60 min.

The samples were prepared by diluting the supernatant of the stock solution to receive a sample concentration of about 0.1 mg/mL. To verify the suitability of the method, 5-ASA and the mixture thereof with citric acid were dissolved in 0.01 M hydrochloric acid to give a sample concentration of about 1 mg/mL. 20 µL of the samples were injected into the HPLC system. The chromatograms were processed using the ChemStation® software (Rev.B.03.02) from Agilent.

2.4 NMR experiments and data analysis

All NMR experiments were performed on a Bruker Avance III 400 MHz spectrometer operating at 400.13 MHz with an inverse probehead. The data were processed with TopSpin 3.5. pl7 (DOSY2D), Dynamic Center 2.5.3 (Bruker BioSpin GmbH, Rheinstetten) and General NMR Analysis Toolbox (Manchester NMR Methodology Group, University of Manchester).

For the ^{1}H NMR and DOSY experiments 600 µl of the supernatant was transferred into 5 mm tubes and were examined. The ^{1}H NMR experiments were performed at 300 K using 256 or 512 scans, resulting in a digital resolution of 0.251 Hz over a spectral width of 20.55 ppm (time domain size 64k), 30° flip angle, acquisition time of 3.98 s followed by a relaxation delay of 1 s and a line broadening factor of 0.3 Hz was applied. Baseline corrections were performed automatically though phasing was performed manually. For the COSY experiment, a standard Bruker pulse program was used with an increase of the scans to 64.

The DOSY experiments are performed using a stimulated echo sequence with bipolar gradients and LED (Longitudinal Eddy-current Delay). The diffusion delay (big delta) was set to 50 ms, the gradient length (little delta) to 3.2 ms (5-ASA with citric acid; D_2O), 3 ms (5-ASA with formic acid; D_2O) and 4.4 ms (benzocaine; DMSO-d6). An Eddy Current delay of 5 ms and a number of scans of 32, 256 or 512 were used. A linear gradient ramp with 32 steps (start value 2%, final value 98%) or 16 steps (for benzocaine and 5-ASA with formic acid) was applied. 90° pulse length was calibrated to ensure a perfect 180° pulse for gradient echo sequence. Signal decay was optimized for 6-H proton down to 3.5%, for 6'H down to 7% for the sample of 5-ASA with citric acid. For all other samples the signal decay was optimized down to 2% for the smallest signal of interest. Lock system was optimized for the gradient experiment. The gradient was calibrated to 4.97 G/mm at 100% (10 A).

3. Results and discussion

3.1 NMR experiments

According to ref. [3], the ratio of 5-ASA and citric acid was chosen 1:1 and the compounds were mixed in 1 mL of deuterium oxide. The 5-ASA/formic acid mixture was prepared, accordingly. For the benzocaine/dibutyl phthalimide reaction an artificial mixture consisting of benzocaine, dibutyl phthalate, and 4 percent of N-(4-ethoxycarbonylphenyl)phthalimide (as reaction product) was prepared.

3.1.1 5-ASA and citric acid

In accordance to Larsen et al. [3], the NMR signals could be assigned to 5-ASA and the expected reaction products: the NMR signals were assigned to compound **1** (7.70 ppm, 7.34 ppm, 6.92 ppm), **2** (7.62 ppm, 7.25 ppm, 6.96 ppm), and **3** and **4** (7.68 ppm, 7.36 ppm,6.83 ppm, 6.79 ppm) (see Figure 2). Discrimination between compounds **3** and **4** was not possible because Larsen et al. [3] used a different solvent. To assign the signals of **2**, **3**, and **4**, a COSY experiment was performed.

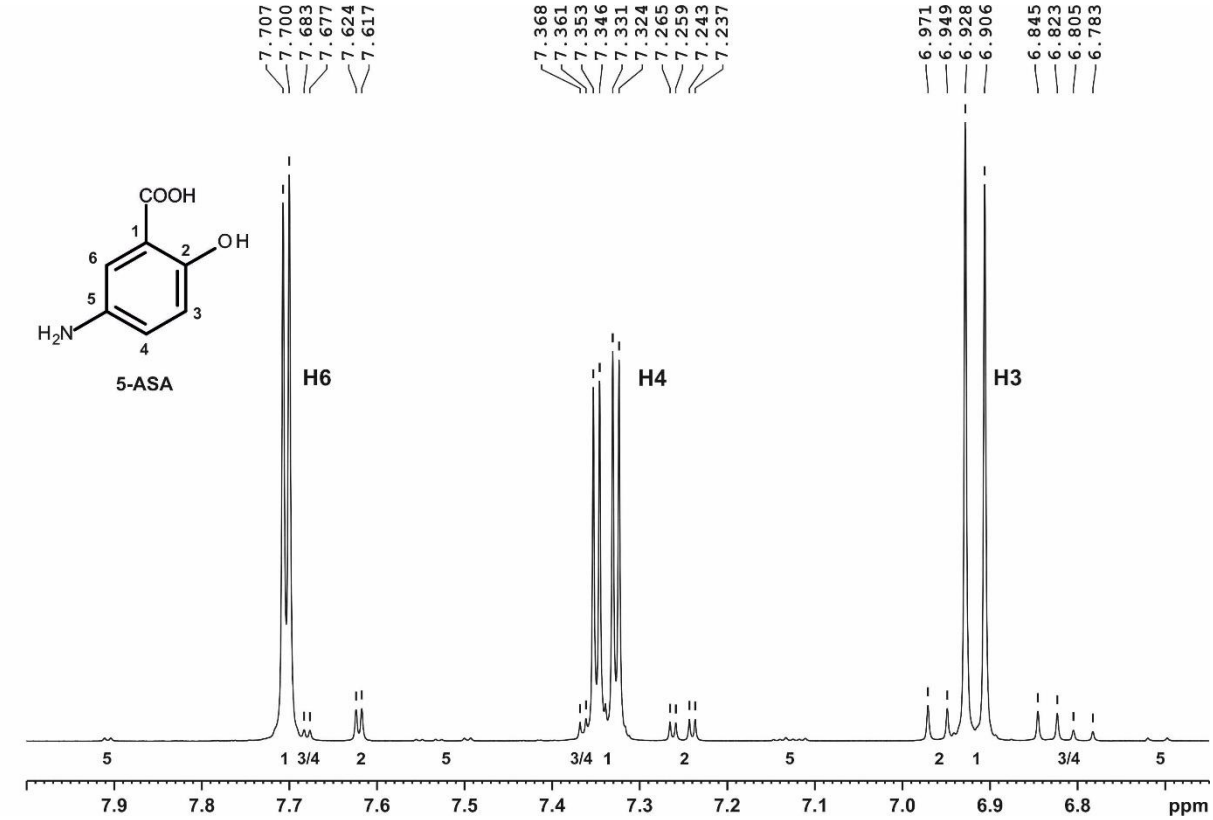


Figure 2 ^{1}H NMR spectrum (D_2O) with the corresponding aromatic signals of 5-ASA (1;*J* [Hz]: d,*8.9*; dd,*8.9*,*2.8*; d,*2.8*), the reaction products (ester (**2**), amides (**3** and **4**)) and ^{13}C satellites (**5**).

Since the sizes of the reaction products are larger in comparison to the starting products, the outer shape of the molecule has changed and along with this, the diffusion coefficient has decreased. Thus, the signals of corresponding spectra should be sorted due to their different diffusion coefficient in the 2D DOSY diagram.

In DOSY2D software all columns of the pseudo 2D spectrum are interpreted independently using either the integral or the peak height information for calculation. Since the differences were negligible, the integral information was used.

In the DOSY spectra the reaction product 2 could be clearly assigned by means of the ^{1}H spectrum (cf. 4). The assignment for the remaining two reaction products **3** and **4** was more difficult due to their proximity to the signals of 5-ASA (**1**). A closer examination of H6 and H4 signal of 5-ASA revealed the much smaller signals of **3** and **4** right next to them (see Figure 2). The decay behavior of the smaller signal is influenced by the bigger one and leads to a shift in the DOSY spectrum as can be seen in Figure 3. The reason is a lack of an individual baseline correction for each signal. Though, the distortion depends on percentage of the overlap with other signals and their decay behavior. This is a typical problem of small signals right next to larger ones, especially when they are not baseline separated. Nevertheless, the resulting signals of the H6 and H4 protons of **3** and **4** can be seen.

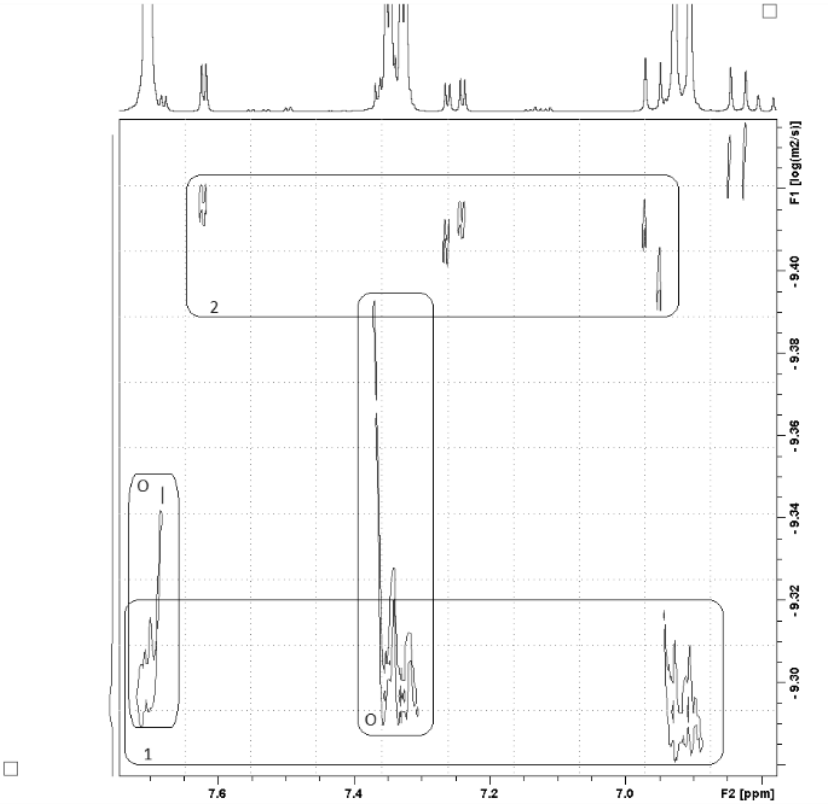


Figure 3 DOSY diagram (D_2O) of the supernatant zoomed in on the aromatic field with the signals of 5-ASA (**1**) and **2**. Due to the influence of the larger H4 and H5 signals of 5-ASA on the decay behavior of the smaller signals of **3** and **4**, a signal shift of the corresponding DOSY signals is seen (**O**). (Processing: Software: DOSY2D (TopSpin 3.5 pl 7, Bruker BioSpin), SI(F2) 64K, SI(F1) 512, exponential fit of signal decay (Levenberg-Marquardt algorithm), Imod was set to Integral)

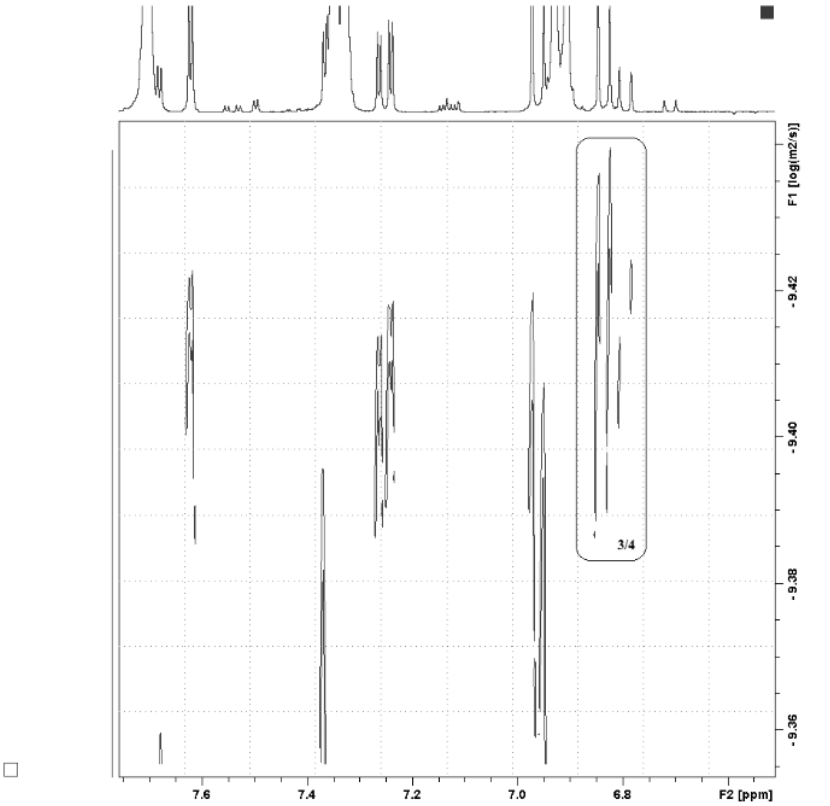


Figure 4 Assignment of the remaining DOSY cross peaks to protons of **3** and **4** by means of the corresponding H'3 signals (D_2O; Processing: see Figure 3).

Using the software of the Dynamic Center it is possible to process diffusion data in a more variable way. For this reason, it was used to see if it is possible to improve the DOSY processing (Figure 4).

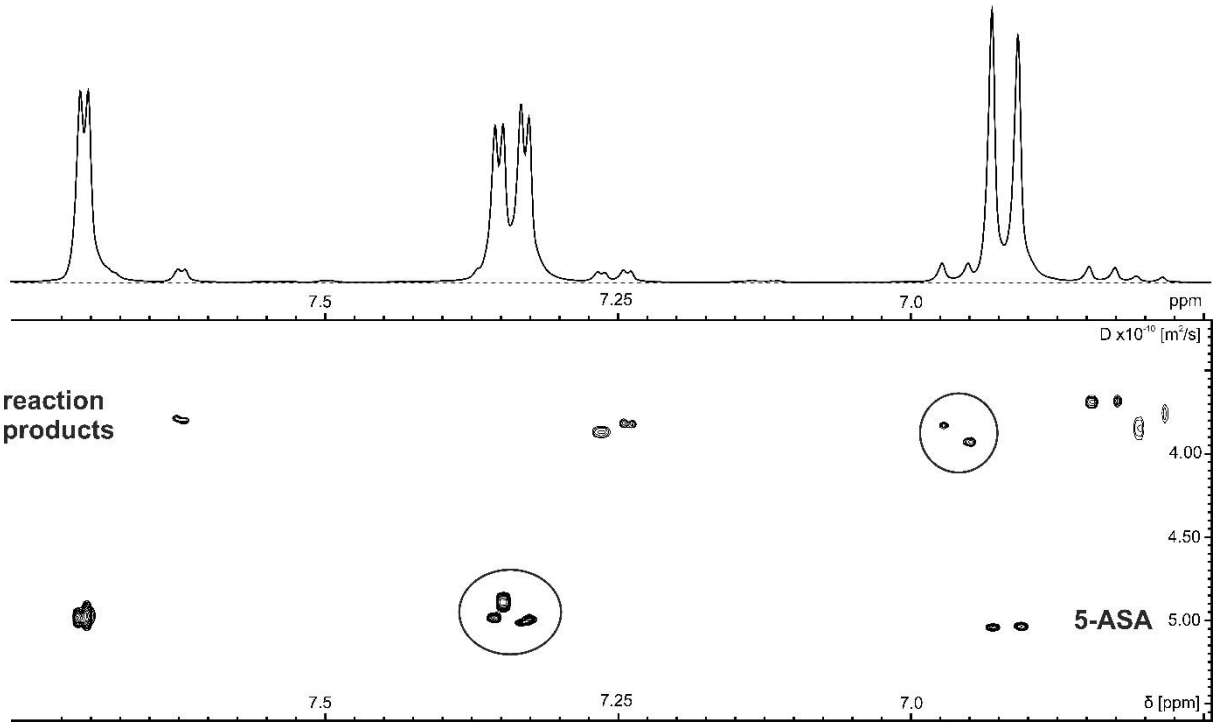


Figure 5 DOSY cross peaks of the reaction products (above) and 5-ASA. In the cycle, the problem of hidden signals and different processing is shown. For explanation see chapter 3.1.1. (D_2O; Processing: DC, number of points in F2: 64K, number of points in F1: 512, all peak heights were used).

The examination of signals split by direct spin-spin-coupling was examined as fragmented signals on peak height. With all peak heights selected, it allows the detection of hidden signals which lie under the analyzed ones as can be seen for the H4 signal of 5-ASA in Figure 5. Due to the hidden signal it comes to a shift in the DOSY signal. Then, the signals with the smallest influence were chosen to gain a better diffusion signal location of the signal in 2D-DOSYcorresponding to a more realistic diffusion coefficient and to obtain a clear separation of 5-ASA and the reaction products (see Figure 6). With these variations it was possible to fully analyze the mixture even with the overlapping signals.

These findings were confirmed by processing the DOSY data with a further tool, the General NMR Analysis Toolbox (GNAT), which provided the same results.

In order to cross check the results of the DOSY experiments of 5-ASA and citric acid the HPLC method of related substances for mesalazine from the PhEur [11] was applied even though not optimized for this application. In addition to the API peak three new signals with retention times of 3.06 min, 3.83 min and 5.19 min were observed which belong to the reaction products. The short retention time is due to the deprotonation at pH 6.2. This measurement confirms the assumption that three molecules were formed.

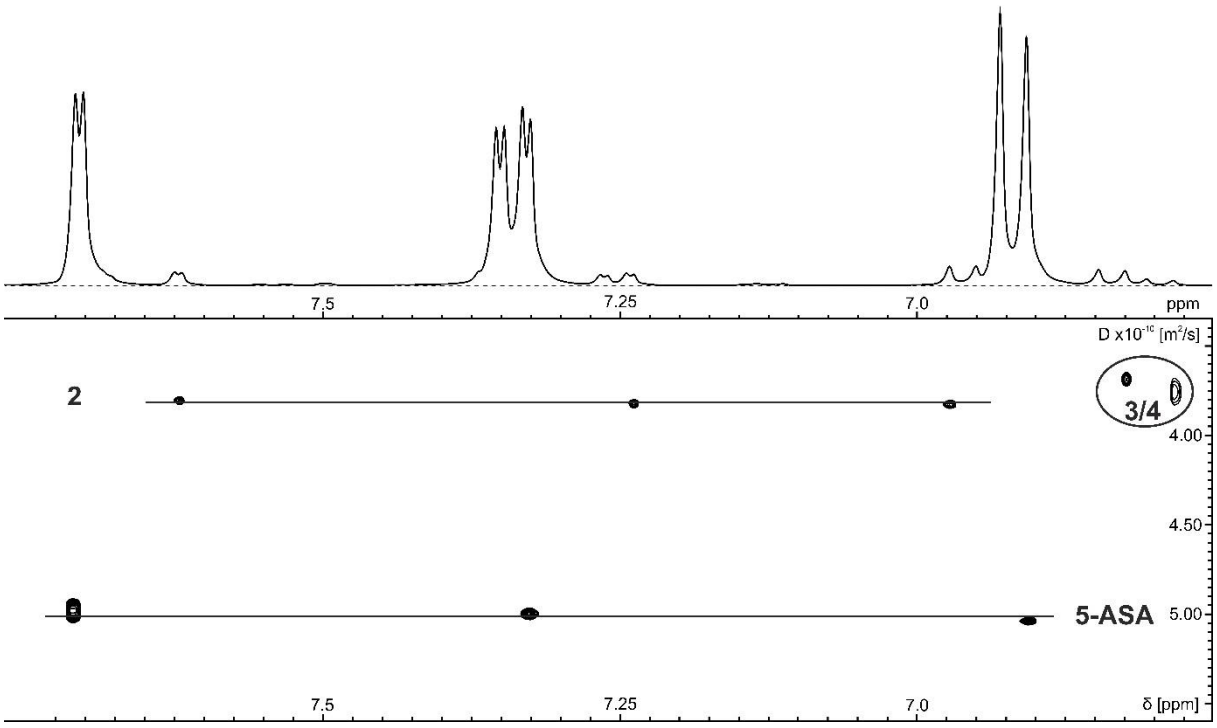


Figure 6 Distinct separation of spectra of 5-ASA and of the reaction products (D_2O; Processing: DC, number of points in F2: 64K, number of points in F1: 512, selected peak heights).

3.1.2 5-ASA and formic acid

By means of a ^{1}H and COSY measurement it was possible to assign the signals to 5-ASA (7.76 ppm, 7.38 ppm, 6.98 ppm), formic acid (8.13 ppm), and two reaction products, the ester **6** (8.37 pm, 7.82 ppm, 7.45 ppm, 6.89 ppm) and the amide **7** (7.57 ppm, 7.25 ppm, 6.89 ppm) (see Figure 7). The signals of **6** and **7** at 6.89 ppm overlap almost completely and can only be recognized due to a small shoulder at the bigger signal of **6**. The assignment of the signals of **6** and **7** to the ester or amide, respectively was due to the calculated chemical shifts according to increment tables and the chemical shifts of the ester and amide of 5-ASA and citric acid.

This example demonstrated the limits of DOSY measurements. Due to the overlap of signals like the signal of **7** at 7.57 ppm with ^{13}C satellites or the signals of **6** and **7** at 6.89 ppm and the proximity to larger signals seen at 8.13 ppm (formic acid), 7.82 ppm (**6**) and 7.45 ppm (**6**), most of the DOSY signals are influenced. As can be seen from Figure 7, cross peaks of **6** and **7**, respectively, appear at different diffusion coefficients which physically makes no sense. In contrast, the 5-ASA signals as well as the signal of **7** at 7.25 ppm can be clearly assigned. Hence, care must be taken when signals of a large intensity difference are close together. In this case, the diffusion coefficient read from the diagram is not realistic. Even with the GNAT, which delivered a clearer separation of the DOSY cross peaks than TopSpin, the results did not improve (see Figure S1 and S2).

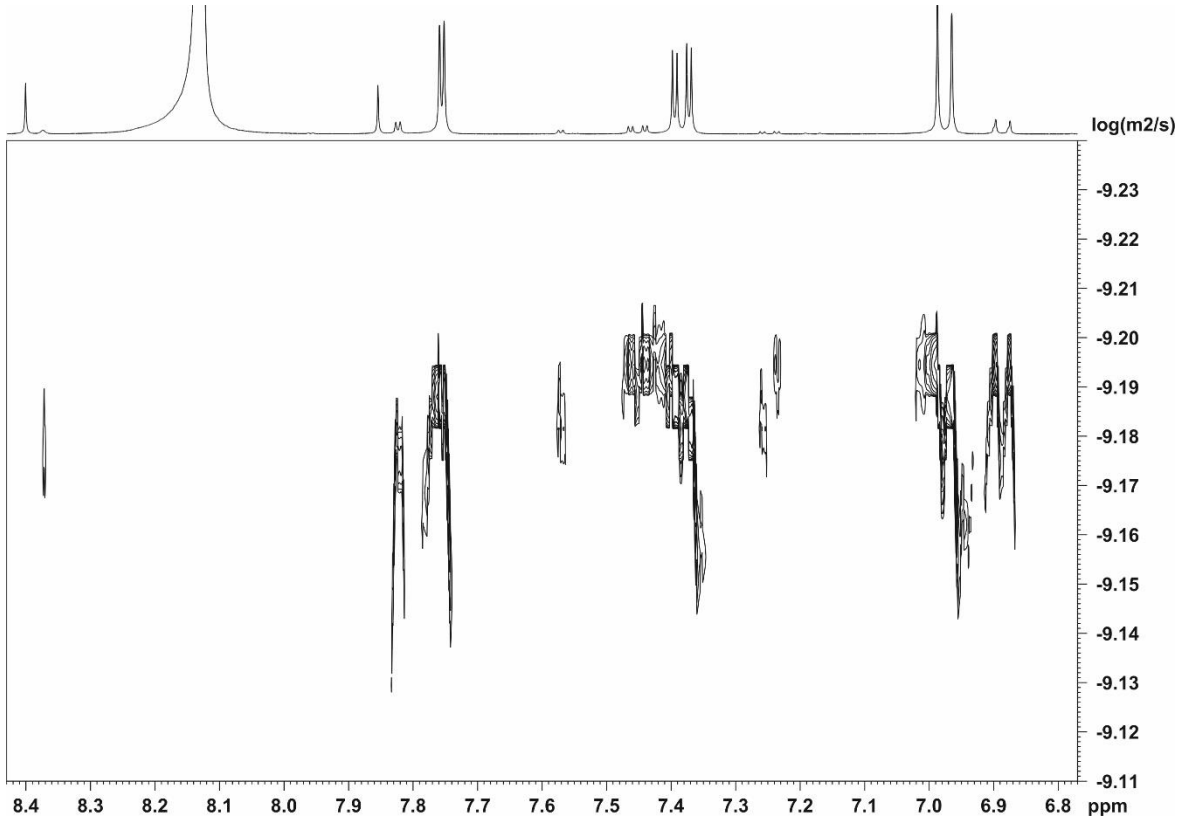


Figure S1 DOSY diagram of 5-ASA, formic acid and their reaction products processed with TopSpin (D_2O).

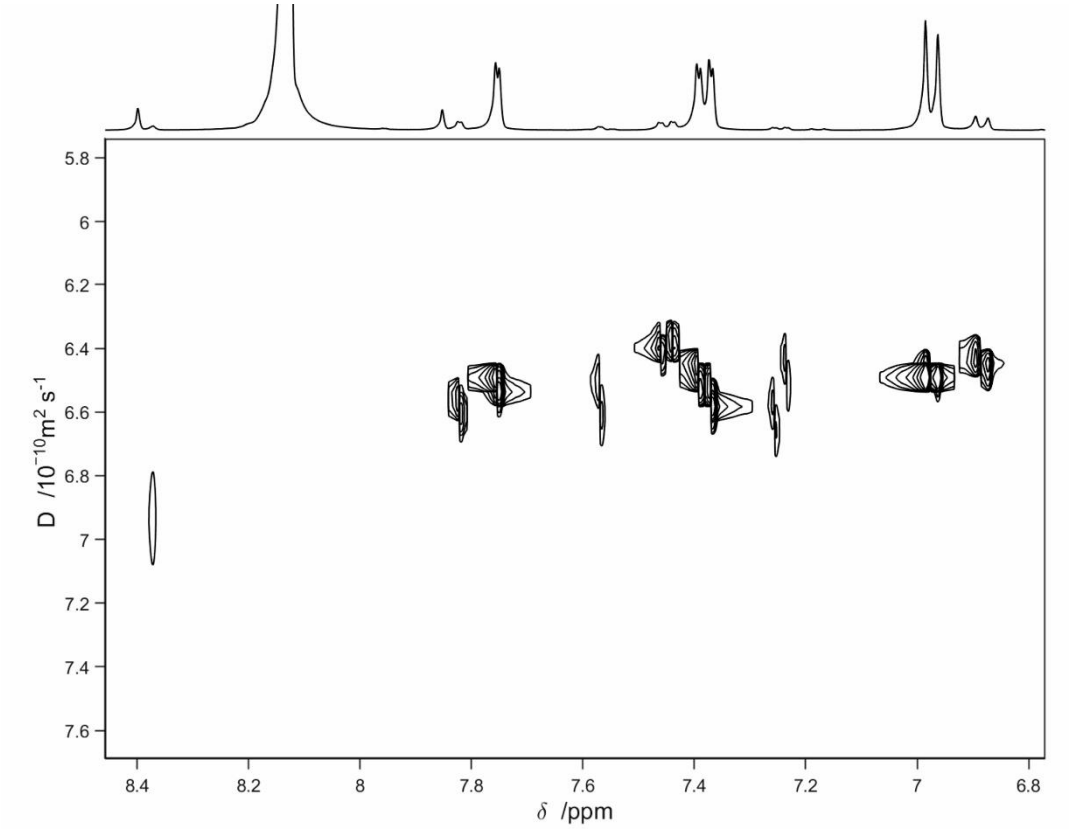


Figure S2 DOSY diagram of 5-ASA, formic acid and their reaction products processed with GNAT (D_2O).

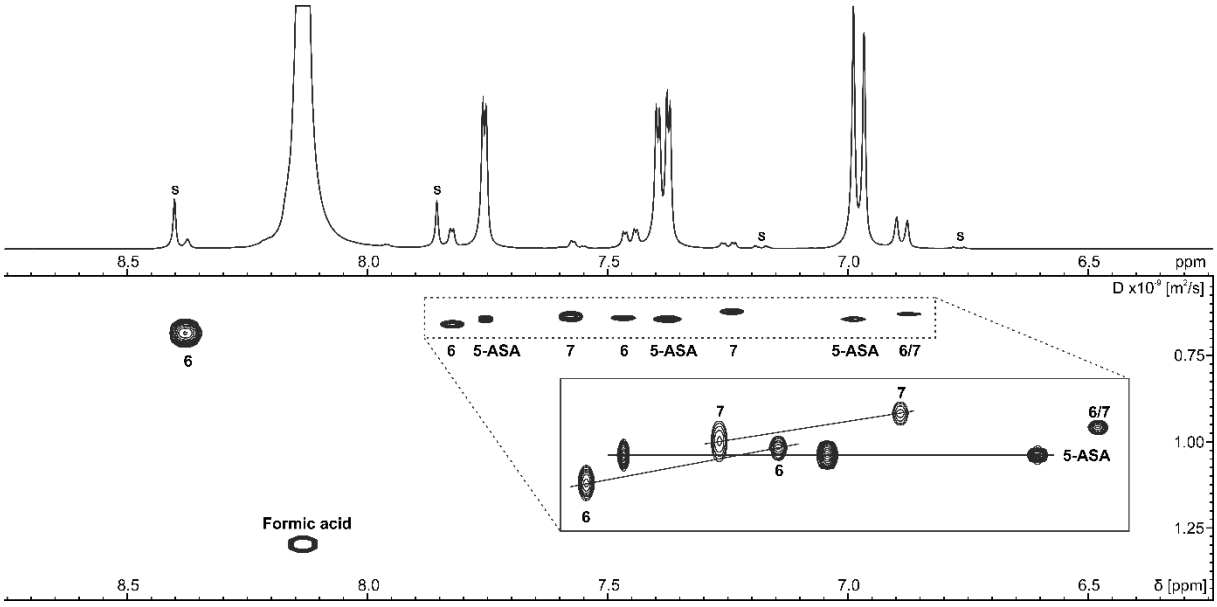
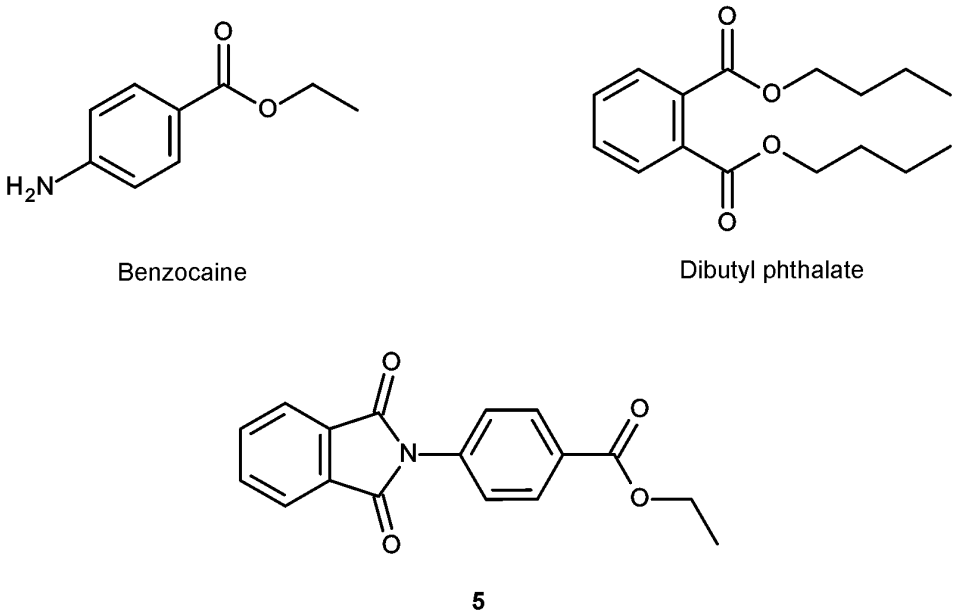


Figure 7 DOSY diagram of 5-ASA, formic acid, and their reaction products (amide (**6**), ester (**7**), ^{13}C satellites (**S**)). It can be seen that the diffusion coefficients within **6** and **7** differ clearly. (D_2O; Processing: DC, number of points in F2: 32K, number of points in F1: 512, selected peak heights)

3.1.3 Benzocaine and dibutyl phthalate



Benzocaine

Dibutyl phthalate

Figure 8 Structures of benzocaine, dibutyl phthalate and *N*-(4-ethoxycarbonylphenyl)phthalimide (**5**).

The signals of the ^1H NMR spectrum can be assigned to benzocaine (7.64 ppm, 6.57 ppm, 5.94 ppm, 4.20 ppm, 1.27 ppm), dibutyl phthalate (7.70 ppm, 4.23 ppm, 1.65 ppm, 1.38 ppm, 0.91 ppm), and the reaction product **5** (8.12 ppm, 7.96 ppm, 4.36 ppm) (Figure 8 and 9). The separated signals or the signals which were least affected by overlaps were chosen for the DOSY processing with Dynamic Center. Here, a distinct separation of the substance could be

achieved. Of note, the cross peaks of all components between δ = 4.2 and 4.4 ppm are located at the correct diffusion coefficient even though the intensity of the signals are very different. This might be due to the fact that the difference in size is large enough for separation of the compound spectra in the DOSY experiment. For this example GNAT confirmed the results of Dynamic Center and TopSpin as well.

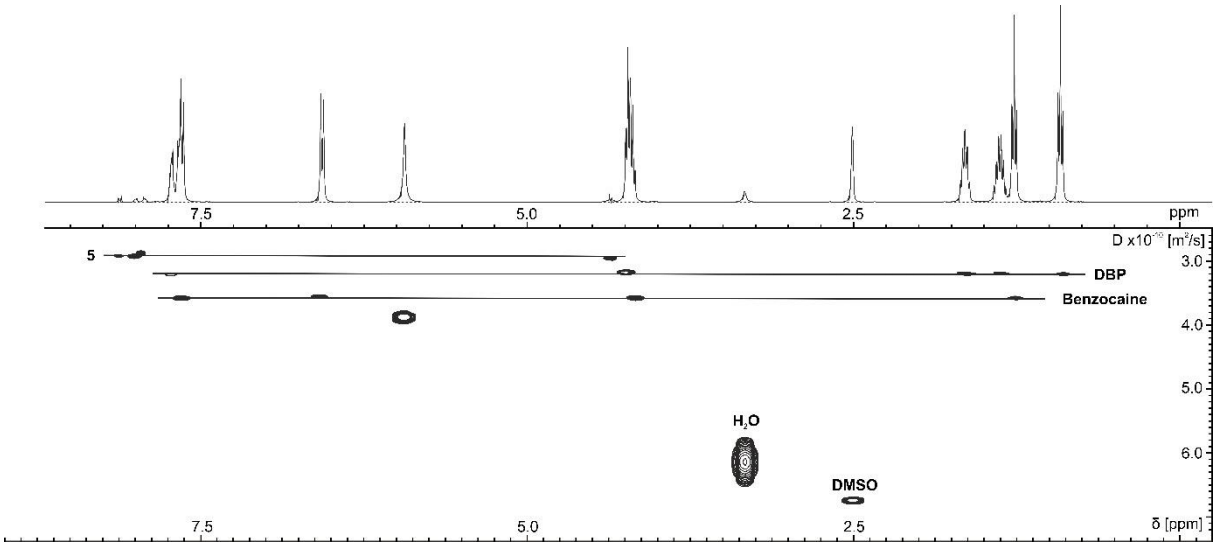


Figure 9 DOSY diagram with a clear separation of benzocaine, dibutyl phthalate and **5** based on the different diffusion coefficient. The signal of the amine of benzocaine at 5.94 ppm is shifted which might be due to proton exchange. (DMSO-d6; Processing: DC, number of points in F2: 32K, number of points in F1: 512, selected peak heights)

4. Conclusion

It has been shown that DOSY is capable of screening a mixture of API, excipient, and their reaction products based on their translational diffusion behavior. The measurement of the DOSY experiments is often less than 1 h, even when a simple 400 MHz spectrometer has been used. The time for the measurement can substantially be shortened using a 600 MHz machine equipped with a cryo probe. With the use of variable processing and previous knowledge of coupling patterns, an assignment was possible even of the much smaller signals next to the larger ones of the API. The processing of reasonable selected parts of a signal was valid and allows an insight into the mixture and a differentiation of slightly overlapping signals.

DOSY experiments come to their limit, when a small and a large signal of different components overlap and the difference in the diffusion coefficient is too small. In all other cases the DOSYs are valuable tool to characterize a mixture.

Acknowledgments

Thanks are due to Markus Zilker, Niclas Förtig, Klaus Schilling, and Florian Geyer (all University of Würzburg) for their assistance and worthwhile discussions.

References

[1] Bharate SS, Bharate SB, Bajaj AN Interactions and incompatibilities of pharmaceutical excipients with active pharmaceutical ingredients: a comprehensive review. *J Excipients Food Chem* (2016) 1, 3-26

[2] Larsen J, Cornett C, Jaroszewski JW, Hansen SH Reaction between drug substances and pharmaceutical excipients: Formation of citric acid esters and amides of carvedilol in the solid state. *J Pharm Biomed Anal* (2009) 49, 11-17

[3] Larsen J, Staerk D, Cornett C, Hansen SH, Jaroszewski JW Identification of reaction products between drug substances and excipients by HPLC–SPE–NMR: Ester and amide formation between citric acid and 5-aminosalicylic acid. *J Pharm Biomed Anal* (2009) 49, 839-842

[4] Yoshioka S, Stella VJ (2000) Stability of Drugs and Dosage Forms. Kluwer Academic / Plenum Publishers, New York, 33-34

[5] Balayssac S, Gilard V, Delsuc M-A, Malet-Martino M DOSY NMR, a new tool for fake drug analyses. *Spectrosc Eur* (2009) 21, 10-14

[6] Trefi S, Gilard V, Malet-Martino M, Martino R Generic ciprofloxacin tablets contain the stated amount of drug and different impurity profiles: A 19F, 1H and DOSY NMR analysis. *J Pharm Biomed Anal* (2007) 44, 743-754

[7] Kellenbach E, Burgering M, Kaspersen F Using Pulse Field Gradient NMR-Based Diffusion Experiments To Identify Signals of Low-Molecular-Weight Impurities. *Org Process Res Dev* (1999) 3, 141-144

[8] Balayssac S, Retailleau E, Bertrand G, Escot M-P, Martino R, Malet-Martino M, Gilard V Characterization of heroin samples by 1H NMR and 2D DOSY 1H NMR. *Forensic Sci Int* (2014) 234, 29-38

[9] Morris Gareth A (2009) Diffusion-Ordered Spectroscopy. In: Harris RK, Wasylishen RL (Eds) Encyclopedia of Magnetic Resonance.

[10] Antalek B Using pulsed gradient spin echo NMR for chemical mixture analysis: How to obtain optimum results. *Concepts Magn Reson* (2002) 14, 225-258

[11] Council of Europe Mesalazine Monograph No. 04/2013:1699. *European Pharmacopeia*, 8th ed. (2017) 2717-2720

3.5 Fingerprint spectra for formulations using a DOSY filter[1]

[1] This chapter was submitted for publication in a slightly modified form during the writing of this thesis. For details see chapter 7.1 List of publications.

Copyright

https://doi.org/10.1016/j.jpba.2022.114723

1. Introduction

In order to evaluate spectral data using multivariate data analysis and to obtain robust results, it is necessary not only to generate reproducible spectra, but also spectra that are characteristic and significant for the particular task (as a kind of fingerprint [1], e.g., like the fingerprint region in IR spectroscopy for identity testing). In addition to sample preparation, the measurement parameters are also an important element. For this, knowledge about the sample with regard to the goal of the analysis is also of great importance.

In the chapter *"Tracing the origin of paracetamol tablets by near infrared, mid infrared and nuclear magnetic resonance spectroscopy using principal component analysis and linear discriminant analysis"* [2], the excipients were the crucial part of the ^{1}H NMR spectra to distinguish the individual formulations and the manufacturers, respectively. For the later multivariate data analysis (MVA), the spectral regions of the paracetamol signals were excluded because they were present in each formulation and in equal amounts. To save this step and possibly make other excipient signals visible that are covered by the paracetamol signals, one can make use of the DOSY measurement principle. Here, the measurement acts as a kind of filter for unnecessary signals.

The DOSY experiment allows to determine the diffusion coefficient based on signal attenuations caused by self-diffusion of the molecules. For this purpose, a gradient pulse is irradiated which interferes with the homogenous static field and leads to locally different field strengths. After a predefined time, a field gradient is applied in the opposite direction to refocus the spins. Due to self-diffusion, the spins are no longer at the same location in the sample and thus cannot be refocused, resulting in an attenuated signal [3]. To obtain spectra without signals of the active pharmaceutical ingredient (API), in this case paracetamol, the idea was to measure a single ^{1}H spectrum and choose the gradient length, gradient level, and the diffusion delay in a way that the API signals would be completely attenuated and those of the excipients would still be visible due to slower diffusion. The goal was to obtain spectra that are

even more informative with respect to the different excipient formulations of the same API than a conventional ^{1}H spectrum. To verify these considerations and to have a better comparison, 1D DOSY measurements were performed for three of the German paracetamol formulations from subchapter 3.1 (same batches).

2. Experimental

2.1 Materials

The 507-HP-7 5-mm routine NMR tubes and the dimethylsulfoxide (DMSO-d_6, 99.8%) containing 0.03% tetramethylsilane (TMS) were purchased from Euriso-top (Saarbrücken, Germany). The paracetamol preparations from Ratiopharm GmbH, STADA Arzneimittel GmbH and Hexal AG contained 500 mg paracetamol and were purchased at local pharmacies.

2.2 Sample preparation

The samples were prepared as in Becht et al. [2]. The tablets were mortared and transferred into a falcon tube and 6 mL DMSO-d_6 containing 0.03% TMS (v/v) were added. Finally, the samples were vortexed (1 min), sonicated (1 h), and centrifuged (20 min/6k U/min). 600 µL of the resulting supernatant was then analyzed by means of NMR spectroscopy.

2.3 NMR measurements

The NMR experiments were carried out on a Bruker Avance III 400 MHz spectrometer operating at 400.13 MHz with an inverse probehead. Data processing was done with TopSpin 4.0.9 (Bruker BioSpin GmbH, Rheinstetten, Germany). ^{1}H NMR experiments were performed at 300 K and a scan number of 32 was used. The resulting digital resolution was 0.245 Hz across a spectral width of 20.03 ppm (time domain size 64k). Additionally, a 30° flip angle and an acquisition time of 4.09 s followed by a relaxation delay of 1 s was used. A line broadening factor of 0.3 Hz was applied. The 1D DOSY measurements were performed at 300 K using a stimulated echo sequence with bipolar gradients and longitudinal Eddy-current delay. Gradient length (little delta) was set to 3 ms and the diffusion delay (big delta) to 100 ms for all samples. 32 scans and an Eddy current delay of 5 ms were used. The DOSY measurement of the Ratiopharm preparation were performed using the same parameters as for the 1D DOSY measurements. Additionally, a linear gradient ramp with 8 steps (start value 2%, final value 98%) was applied. The gradient was calibrated to 4.97 G/mm at 100% (10 A).

3. Results and discussion

The performance of this approach was tested with three different paracetamol preparations from Ratiopharm GmbH, STADA Arzneimittel GmbH and Hexal AG (Table 1). They differed in the type of composition and the respective quantities of the excipients.

Table 1 List of measured paracetamol preparations with their excipients and the weight of the tablet.

	Ratiopharm	**STADA**	**Hexal**
Weight	645 mg	630 mg	595 mg
Excipients	Corn starch Povidone Magnesium stearate Talc Microcrystalline cellulose Silicon dioxide Croscarmellose sodium	Sodium carboxymethyl starch Povidone Magnesium stearate Microcrystalline cellulose	Corn starch Povidone Magnesium stearate Microcrystalline cellulose Silicon dioxide Stearic acid

Based on the experience from previous DOSY measurements of the Ratiopharm preparation, the gradient length was set to 3 ms, as this value has been shown to be effective in suppressing the paracetamol signals. The diffusion delay was initially set to 50 ms. This was sufficient to completely attenuate the residual signals of water and DMSO, but not the signals of paracetamol. Their intensity decreased only to about that of the excipient signals (Figure 1). Nevertheless, this is an improvement for further multivariate analysis because it simplifies the preprocessing process. Since the large difference in intensity between excipients and paracetamol is cancelled out, either area normalization can be omitted or the entire spectrum can be evaluated.

In the next step, the diffusion delay was increased to 100 ms, which completely attenuated the paracetamol signals except for that of the hydroxyl group (9.14 ppm; Figure 1). As a result, a new signal could be revealed, which previously overlapped with the methyl group of paracetamol, and belonged to povidone. This was also true for another one overlapping with the water signal. Based on the location of the signal in the DOSY measurement, it most likely belongs to the corn starch. The disadvantage, however, was that the signals of magnesium stearate were also completely attenuated due to the low molecular mass, and changing the gradient level to 50% did not improve it either. Here, the signals of stearic acid (1.23, 0.86 ppm) were still visible, but the signals of paracetamol were still too large.

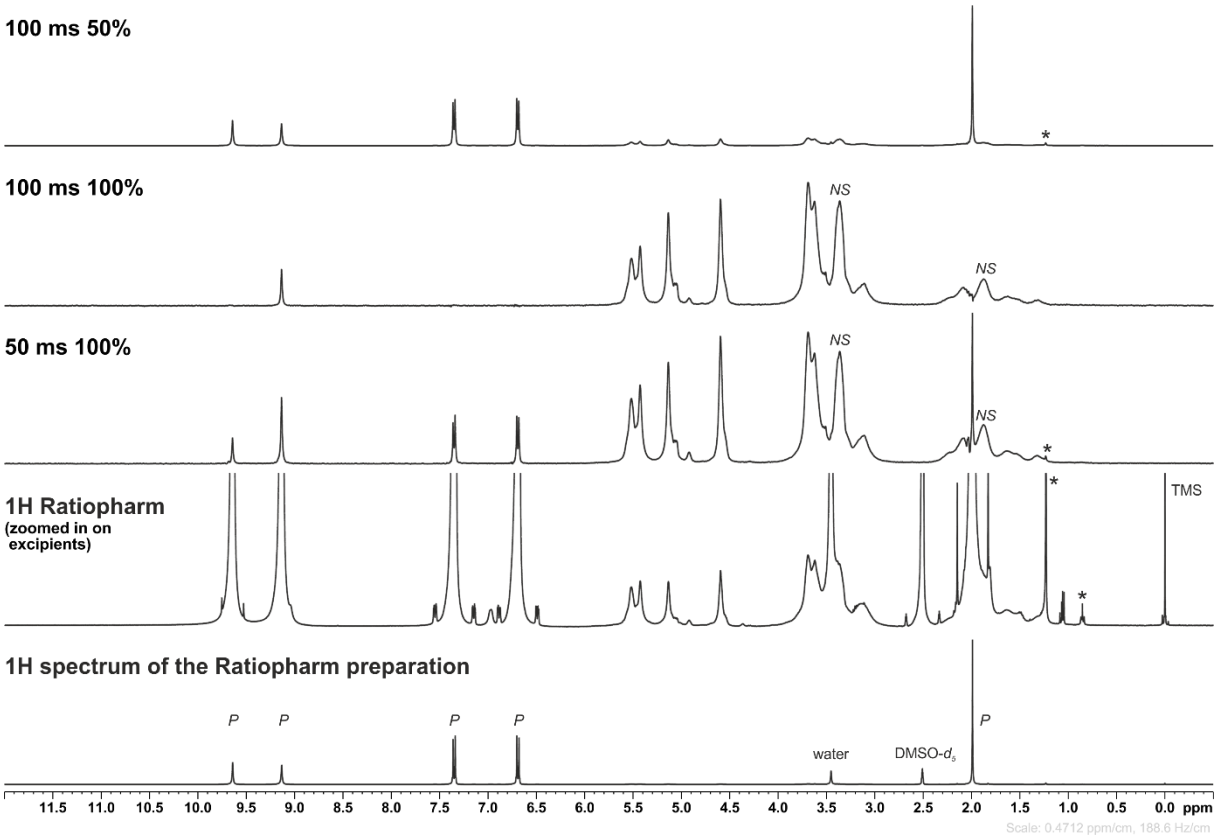


Figure 1 NMR spectra of the Ratiopharm preparation (scaled to the largest signal). Shown is the influence of the gradient on the signals of paracetamol and the excipients. The standard ^{1}H spectrum is shown for comparison, with the range of excipients additionally zoomed in (the bottom two). The diffusion delay is given in ms and the gradient level in percent. The asterisk marks the signals of magnesium stearate, *P* the signals of paracetamol and *NS* the new signals which were revealed.

The STADA preparation showed only very small signals of the excipients compared to that of Ratiopharm, which can be attributed to the overall slightly lower amount of excipients, but mainly to the number and type of excipients. Sodium carboxymethyl starch tended to be less soluble than the corn starch contained in Ratiopharm, which is only slightly soluble in DMSO [4], and therefore could not be identified in a reference spectrum or in the sample. Using a reference spectrum, some very small signals could be assigned to cellulose, although it is not soluble in most organic solvents [4,5]. Therefore, the spectrum only consisted of signals of povidone, magnesium stearate, very small signals of microcrystalline cellulose and two signals, which could not be assigned to any of the excipients (Figure 2). As could be expected from the Ratiopharm sample only a spectrum of povidone was left with a selected diffusion delay of 100 ms, gradient level of 100% and gradient length of 3 ms. Also the unknown signal was completely attenuated. The decrease of the gradient level to 50% showed the same results as for Ratiopharm. Even a gradient length of 2 ms did not bring any improvements.

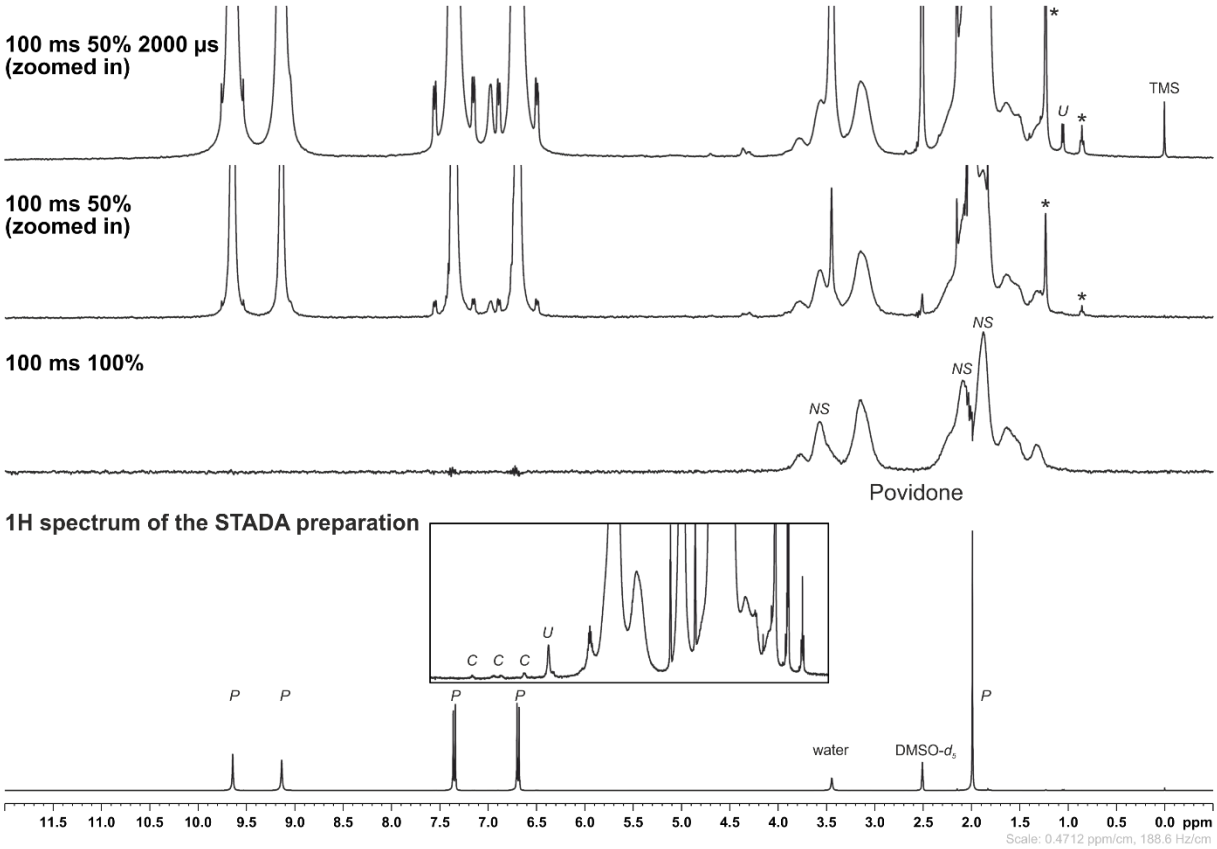


Figure 2 NMR spectra of the STADA preparation. Shown is the influence of the gradient on the signals of paracetamol and the excipients. The standard ^{1}H spectrum is shown for comparison, with the range of excipients (6 – 0.5 ppm) additionally zoomed in. The selected diffusion delay is given in ms, the gradient level in percent and the gradient length in µs. The asterisk marks the signals of magnesium stearate, *P* the signals of paracetamol, *U* the unassigned signals and *NS* the new signals which were revealed.

Due to the very similar composition of the tablet, the Hexal sample also showed a similar spectrum (Figure 3) as the Ratiopharm sample but contained approximately 50 mg less excipients. A diffusion delay of 100 ms and a gradient level of 100% already gave good results, as all paracetamol signals were attenuated, except for that of the hydroxyl group. Of the detectable excipients, only the signals of stearic acid were attenuated. Other settings such as a gradient level of 50% did not improve the results. As with the other samples, either the signals of paracetamol were not sufficiently attenuated and those from stearic acid were still seen or both were attenuated, except for the signals of povidone and corn starch or microcrystalline cellulose.

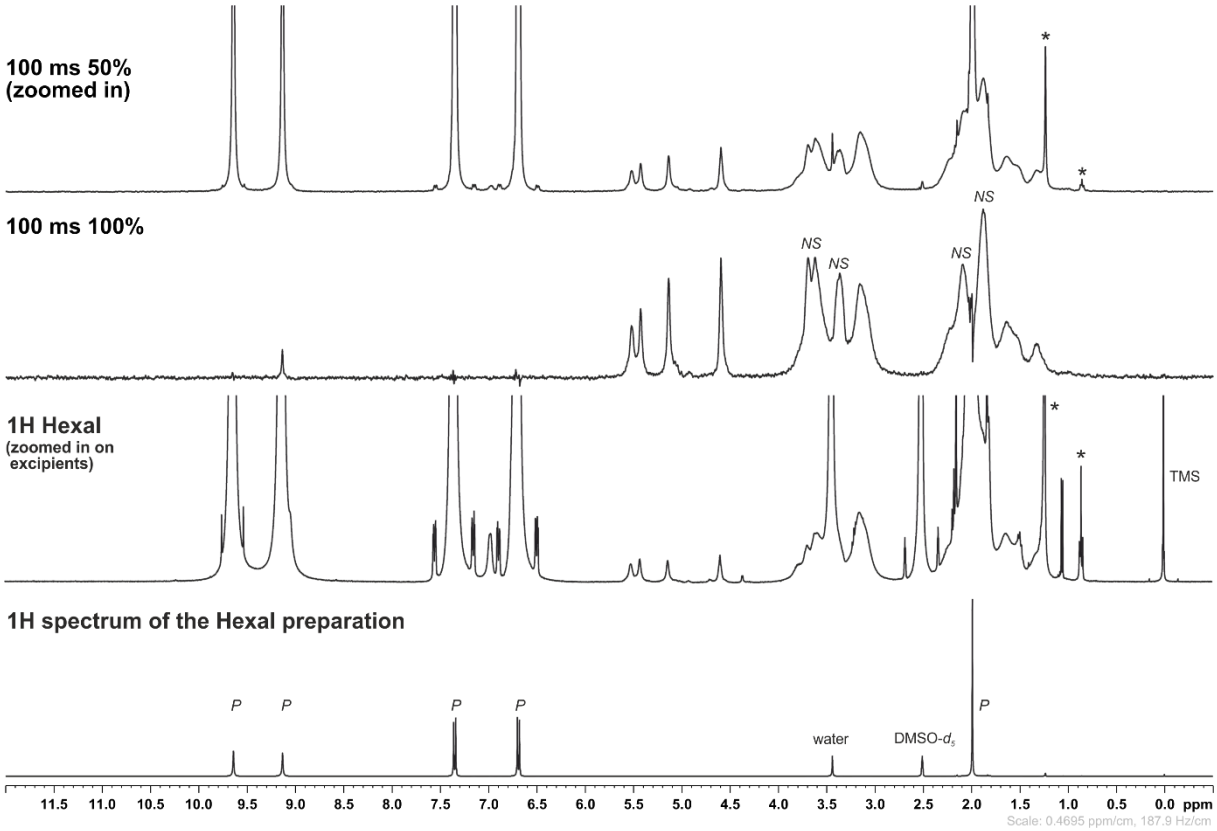


Figure 3 NMR spectra of the Hexal preparation. Shown is the influence of the gradient on the signals of paracetamol and the excipients. The standard ^{1}H spectrum is shown for comparison, with the range of excipients additionally zoomed in (the bottom two). The diffusion delay is given in ms and the gradient level in percent. The asterisk marks the signals of magnesium stearate, *P* the signals of paracetamol and *NS* the new signals which were revealed.

Although only a few signals remained after the 1D DOSY measurement, the formulations of all three MAHs could be easily distinguished based on the spectra (see Figure 4). This was achieved either by the hydroxyl group of the paracetamol, which was still visible for Ratiopharm and Hexal, or by the different amounts of excipients and the associated different signal intensities.

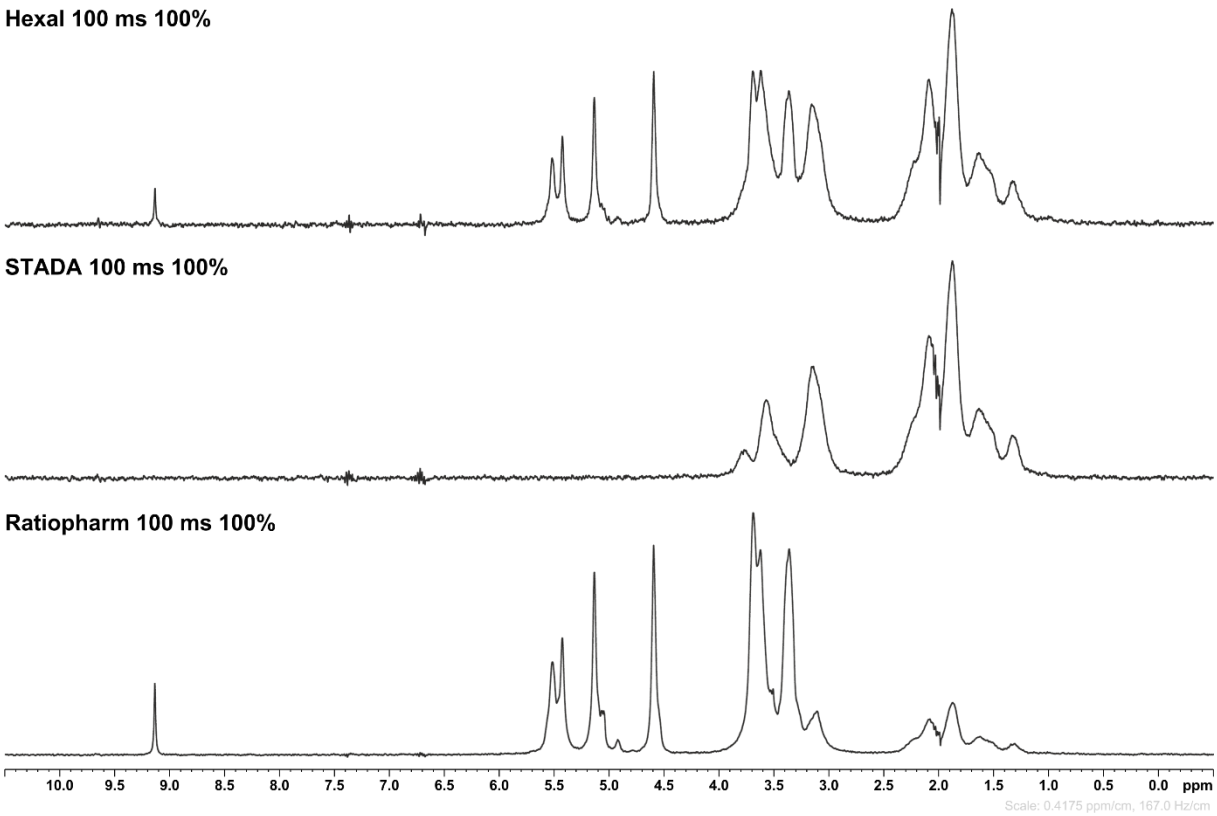


Figure 4 Comparison of the 1D DOSY spectra of the three different preparations. Based on the remaining excipient signals, as well as their intensities, the three formulations can be clearly distinguished.

4. Conclusion

It could be shown that it is possible to get more characteristic spectra of paracetamol tablets by eliminating the paracetamol signals by means of 1D DOSY measurements as filters. This simplifies the multivariate analysis of the NMR spectra, as it allows to remove the active ingredient signals altogether or to attenuate them to the point where a normalization procedure is no longer necessary to give more importance to the small excipient signals. It also revealed signals not seen before and removed some signal overlaps. This permitted the formulations to be distinguished only based on the spectra, even though the compositions were very similar and only two to three excipients were visible in the spectra.

From the results it can also be clearly deduced which conditions must be fulfilled if a formulation shall be investigated by NMR spectroscopy using this method. Ideally, formulations should contain many different excipients that are accessible to NMR spectroscopy. This provides more characteristic spectral fingerprints with which to distinguish formulations of the same API. The excipients should be readily soluble in the chosen solvent, although this will almost never be the case for all of them due to widely varying solubilities. In addition, they should have a significant difference in molar mass compared to the API so that their signals are not also completely attenuated by the gradient, as was the case with stearic acid and magnesium stearate, respectively.

References

[1] Biancolillo A, Marini F Chemometric Methods for Spectroscopy-Based Pharmaceutical Analysis. *Front Chem* (2018) 6

[2] Becht A, Schollmayer C, Monakhova Y, Holzgrabe U Tracing the origin of paracetamol tablets by near-infrared, mid-infrared, and nuclear magnetic resonance spectroscopy using principal component analysis and linear discriminant analysis. *Anal Bioanal Chem* (2021) 413, 3107-3118

[3] Günther H (2013) NMR Spectroscopy : Basic Principles, Concepts and Applications in Chemistry. John Wiley & Sons, Incorporated, Somerset, GERMANY

[4] Schmidt PC, Lang S (2013) Pharmazeutische Hilfsstoffe: Eigenschaften, Anwendung und Handelsprodukte

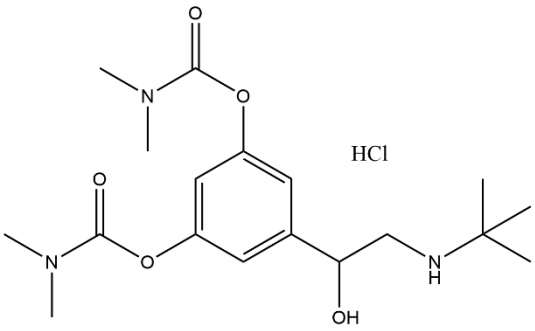
[5] Rowe RC, Sheskey PJ, Quinn ME (2009) Handbook of Pharmaceutical Excipients. Sixth edn. Pharmaceutical Press and American Pharmacists Association

3.6 Comparison of a low-field NMR spectrometer (80 MHz) with a high-field NMR spectrometer (400 MHz) in the quantitative analysis of drugs

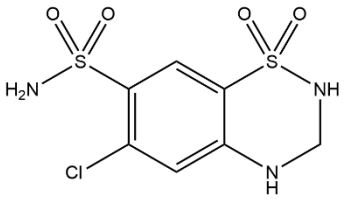
1. Introduction

Although NMR spectroscopy offers a wide range of analytical possibilities (qualitative as well as quantitative) in addition to classical structure elucidation, it is rarely used in routine drug quality control. Some reasons for this are probably the high acquisition and maintenance costs, as well as the size of the spectrometers. To achieve higher resolution for challenging analytes, more powerful and thus larger superconducting magnets have been developed. In contrast to this trend, however, small permanent magnets were reinvented and developed further [1], which enabled the construction of NMR spectrometers that fit on a laboratory bench. The difficulty here was to construct a permanent magnet with sufficient field homogeneity to resolve the ^{1}H shifts of molecules [2]. These benchtop spectrometers typically operate at a ^{1}H frequency between 40 – 80 MHz [3]. Because of the weaker magnetic field, which results in less frequency dispersion and thus more overlap of the signals, and the lower field homogeneity, benchtop instruments are usually used for small molecules [2,3]. However, these disadvantages are balanced by the much lower cost and space requirements. Therefore, benchtop devices are now used in a wide variety of areas such as process and reaction monitoring, teaching and drug analysis [4-6].

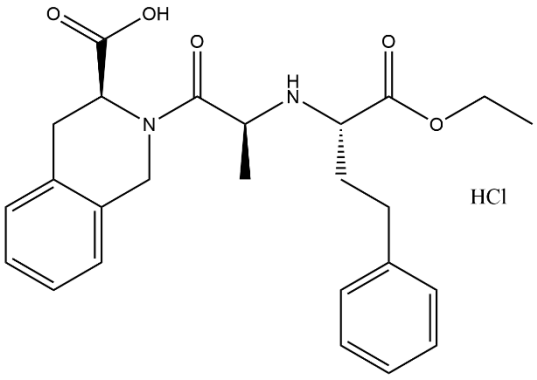
Nevertheless, the gold standard in the European Pharmacopoeia for purity and content assessments is the HPLC analysis. Yet, HPLC always require a chemical reference standard (CRS), which is usually very expensive and needed for each individual active pharmaceutical ingredient (API) to be tested. Here lies a decisive advantage of quantitative NMR (qNMR) analysis. A single reference substance can be used for many different analytes, as long as its signal does not overlap with others, such as those of the analyte or the solvent [7]. The aim of this exploratory investigation was to determine whether qNMR analysis would be a useful addition to routine analysis in pharmacies. For this purpose, two commercially available drugs (see Figure 1 and Table 1) were analyzed by means of qNMR and the results of a high-field spectrometer (400 MHz) were compared with those of a low-field spectrometer (80 MHz).



bambuterol hydrochloride



hydrochlorothiazide



quinapril hydrochloride

Figure 1 Structural formulas of the APIs contained in the drugs Bambec and Quinaplus AL.

2. Materials and methods

2.1 Chemicals and Materials

Dimethylsulfoxide-d_6 (DMSO-d_6) and deuterium oxide (D_2O) were purchased from Deutero GmbH (Kastellaun, Germany). 3-(Trimethylsilyl)propionic acid-d_4 sodium salt (TSP-d_4) and quinapril hydrochloride (european pharmacopoeia reference standard) were purchased from Sigma-Aldrich Chemie GmbH (Taufkirchen, Germany). Hydrochlorothiazide (HCT) was purchased from Schwarz Pharma AG (Monheim, Germany). The samples were measured in 507-HP-7 5 mm routine NMR tubes obtained from Euriso-top (Saarbrücken, Germany).

2.2 Drug samples

All drug samples (Table 1) were kindly provided by Dr. Andreas Winterfeld (Montanus Apotheke, Leichlingen).

Table 1 Information on the drugs used and the APIs contained therein**.**

Drug	API	M_R	batch	Pharmaceutical Company
Bambec	Bambuterol HCl (10 mg)	403.9	V337A	AstraZeneca GmbH
Quinaplus AL	Quinapril HCl (21,66 mg) / Hydrochlorothiazide (12,5 mg)	475.0 / 297.7	81003V	Aliud Pharma GmbH

2.3 Sample preparation

Regardless of whether a whole or half a tablet was measured, they were homogenized first, transferred to a microcentrifuge tube and 1 mL of the deuterated solvent was added (D_2O for Bambec and DMSO-d_6 for Quinaplus). Afterwards, the samples were vortexed (1 min), sonicated (10 min), and centrifuged (10 min/ 13k U/min). 600 µL of the supernatant was mixed with 100 µL of the corresponding internal standard (IS) solution and analyzed. For Bambec, the concentration of the IS solution was 25 mg/mL of TSP in D_2O, for Quinaplus, 5 mg/mL of TSP in DMSO-d_6. For the reference measurements of quinapril (5.38 mg) and the mix of quinapril and HCT (4.89 mg/2.83 mg), the respective substances were dissolved in 700 µL. 600 µL of these were mixed with 100 µL of an IS solution (1.48 mg/mL TSP in DMSO-d_6) and measured. For the reference measurement of HCT, 0.97 mg of TSP and 12.56 mg of HCT were dissolved in 1 ml DMSO-d_6 of which 600 µL were analyzed.

2.4 NMR spectroscopy

The high-field measurements were conducted on a Bruker Avance III 400 MHz spectrometer operation at 400.13 MHz with an inverse probehead operating at 300 K. The quantitative ^{1}H NMR experiments were performed using a flip angle of 90° without rotation. 16 Scans were used, resulting in a digital resolution of 0.183 Hz over a spectral width of 30.04 ppm (time domain size 128k). Acquisition time was 5.45 s followed by a relaxation delay of 30 s. A line broadening factor of 0.3 Hz was applied. Baseline correction (automatically) and phasing (manually) were performed with TopSpin 4.0 (Bruker BioSpin GmbH, Rheinstetten, Germany).

To ensure sufficient relaxation time, an inversion recovery experiment was performed for all samples. The relaxation delay was then chosen to be at least seven times the T1 time for the signal with the longest T1 time, including the acquisition time, which corresponds to a recovery of 99.9%.

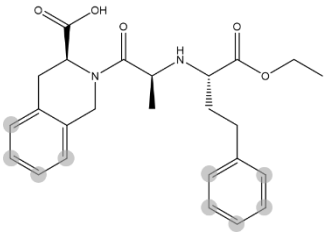
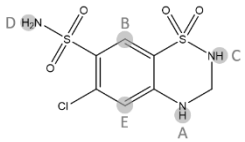
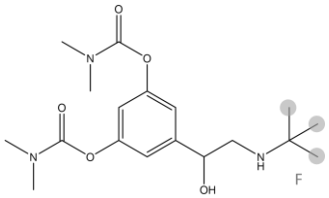
The comparative measurements with a benchtop spectrometer were carried out by Magritek GmbH (Aachen, Germany). For this purpose, the Spinsolve 80 instrument (80 MHz) was used with the same acquisition parameters as for the high-field spectrometer. Only the number of scans was increased to 64. Spectra processing was performed by Magritek itself, while integration was performed using TopSpin.

Resonance assignments were based on literature [8,9], a spectral database [10] and measured reference spectra of the corresponding APIs. The integration ranges are presented in Table 2. For Quinaplus, it was necessary to apply baseline correction to specific regions of the spectrum to provide adequate integration. The individual regions were 10.5 – 6.2 ppm and 0.6 – -0.6 ppm. The integral for TSP was calibrated to 100.

3. Results and discussion

The preliminary tests described below to develop a suitable method were all performed with the 400 MHz spectrometer (such as suitability of the internal standard and choice of signals for quantification). A new sample was then prepared, measured with the 400 MHz spectrometer, and sent to Magritek for a comparison measurement with the 80 MHz benchtop device.

Table 2 Resonance assignments and integration ranges of the protons used for quantification (400 MHz).

Structures	Resonance assignments in ppm	Integration range in ppm	
		API	**TSP**
	Quinapril 7.33 – 7.12 (m, 9H-Ar)	7.40 – 7.10	0.50 – -0.50
	HCT A: 8.27 (s, 1H) B: 8.03 (s, 1H) E: 7.06 (s, 1H)	8.37 – 8.20 8.08 – 7.99 7.10 – 7.01	0.50 – -0.50
	Bambuterol F: 1.38 (s, 9H, *tert*-Butyl)	1.55 – 1.20	0.50 – -0.50

First, the signals had to be correctly assigned (Table 2) to the corresponding APIs and distinguished from those of the excipients, especially in the case of Quinaplus, which contained two APIs. For this purpose, reference spectra of quinapril HCl and HCT were recorded (see Figure 2). The exact assignments to all protons were already described in the literature [8,10,11].

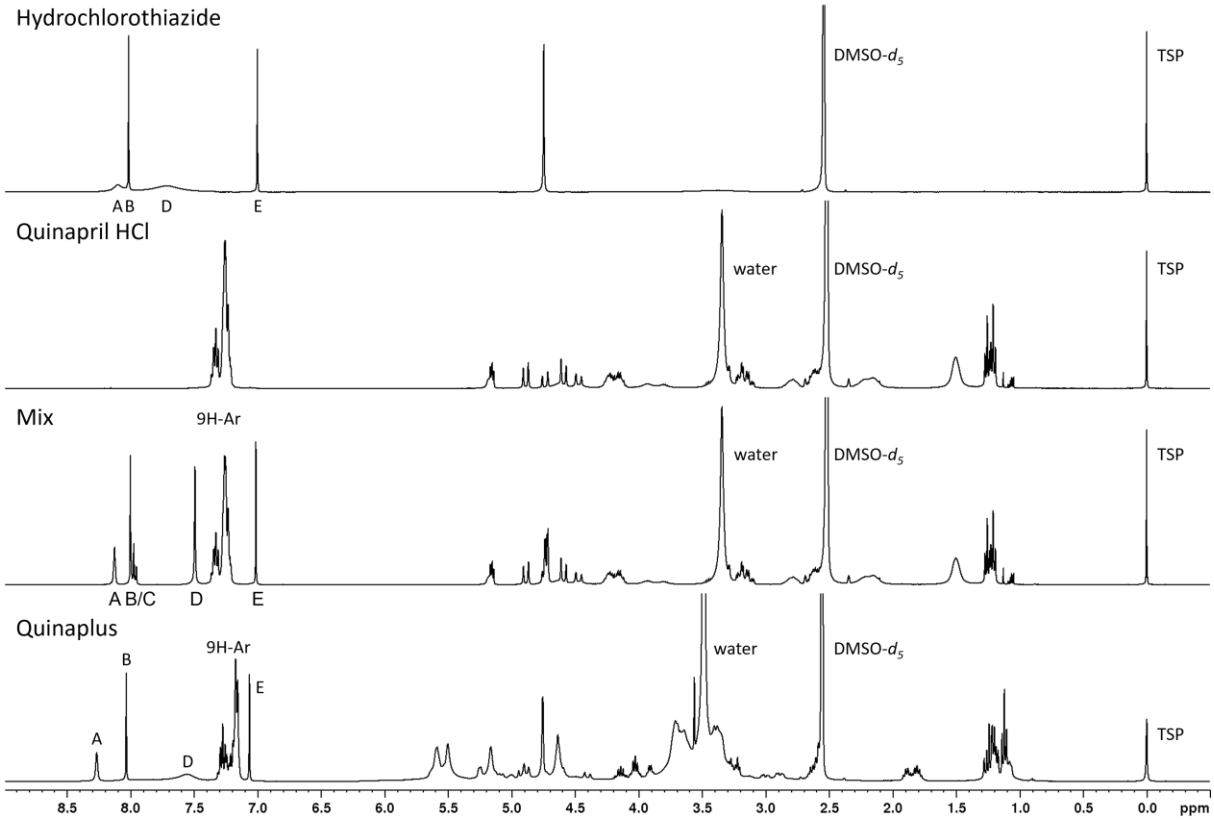


Figure 2 ^{1}H NMR spectra (400 MHz; DMSO-d_6) of Quinaplus, HCT, Quinapril HCl and the mix thereof. Most of the API signals can be assigned in the Quinaplus preparation using the reference spectra. The remaining signals belong to TSP (0.00 ppm), water (3.47 ppm), DMSO-d_5 (2.56 ppm) and excipients. The assignments correspond to those of the quinapril and HCT signals shown in Table 2.

For Bambec (bambuterol), most of the signals could be assigned to the individual protons (Figure 3). This was especially true for the aromatic protons (signals A and B), since there were no other signals in this region, and for those of the *tert*-butyl group (signal F), which appeared as a singlet in the aliphatic region due to the axial symmetry. The proton signals of the methylene group (signals D), the chiral center (signal C) and the methyl groups of the carbamates (signals E) could be identified using a ^{1}H NMR prediction web interface based on the work of Binev et al. [12] and the NMR studies of Abiramasundari et al. [9]. Here, the *N*-methyl groups appeared as two singlets due to their different orientations to the oxygen of the benzene ring or the carbonyl oxygen. Furthermore, since the protons of the methylene group are diastereotopic, they appeared as two doublets. For the same reason, the proton at the chiral center appeared as a doublet of doublets [9].

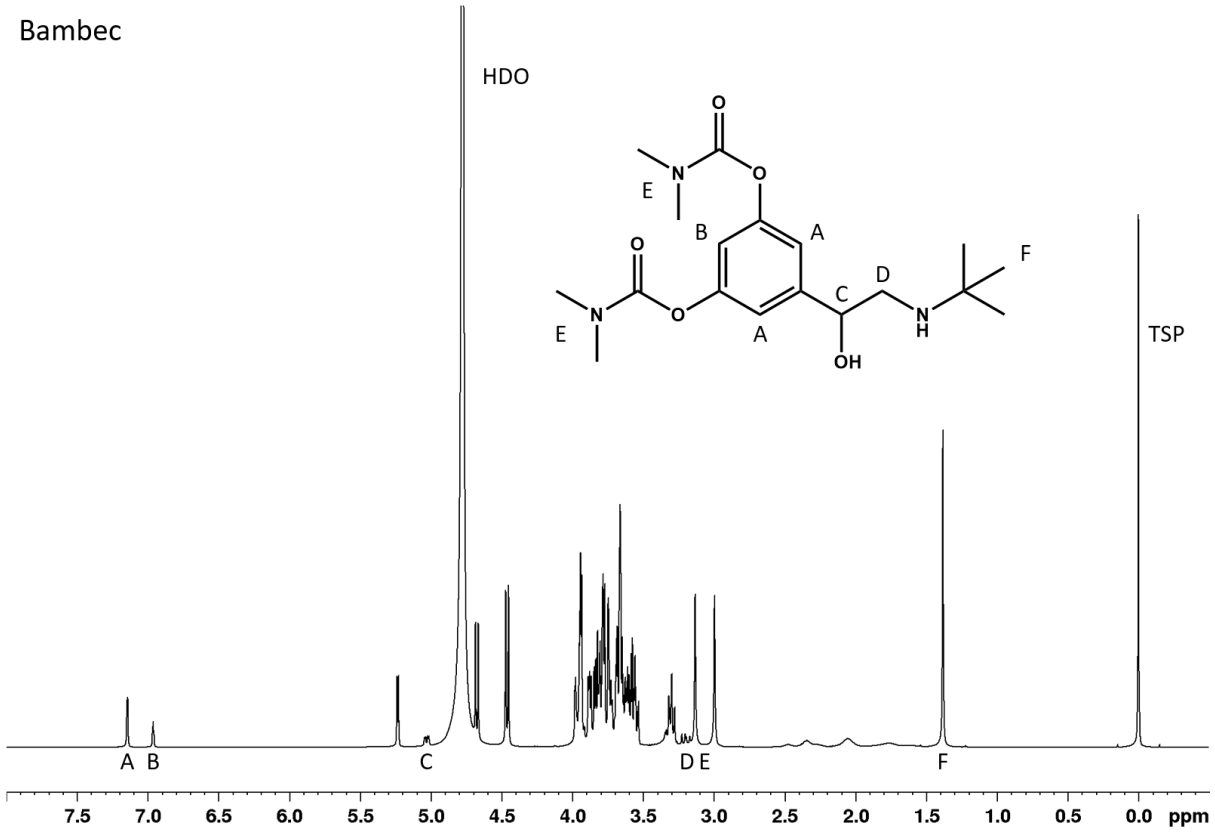


Figure 3 ^{1}H NMR spectrum (400 MHz; D_2O) of Bambec with the structural formula of bambuterol. The corresponding proton signals of bambuterol are indicated by different letters.

Next, the appropriate signals for quantification were to be selected. As can been seen in Figure 2 and 3, many of the API signals overlap with those of the excipients. Therefore, for Quinapril only the downfield-shifted signals were used. In the case of bambuterol, the protons of the *tert*-butyl group were used (see Table 2) because the high intensity singlet is ideal for quantification [13]. The content was then calculated using the following equation

$$P_A = \frac{I_A}{I_{IS}} \times \frac{N_{IS}}{N_A} \times \frac{M_A}{M_{IS}} \times \frac{m_{IS}}{m_A} \times P_{IS}$$

where *I* is the signal intensity, *N* the number of protons, *M* the molecular weight, *m* the weight in the sample and *P* the concentration of the analyte (*A*) and the internal standard (*IS*), respectively.

For bambuterol, a good result was already obtained with a simple sample preparation using half a tablet and adding a defined amount of TSP to the sample. The calculated content for the *tert*-butyl signal was 97.4%. Here, a good result was defined as a maximum deviation of 5% from the declared content, which is in accordance with the specifications stated by the EMA (European Medicines Agency) and BfArM (Federal Institute for Drugs and Medical Devices), where a maximum deviation of ±5% in the content is permitted for the finished product [14,15].

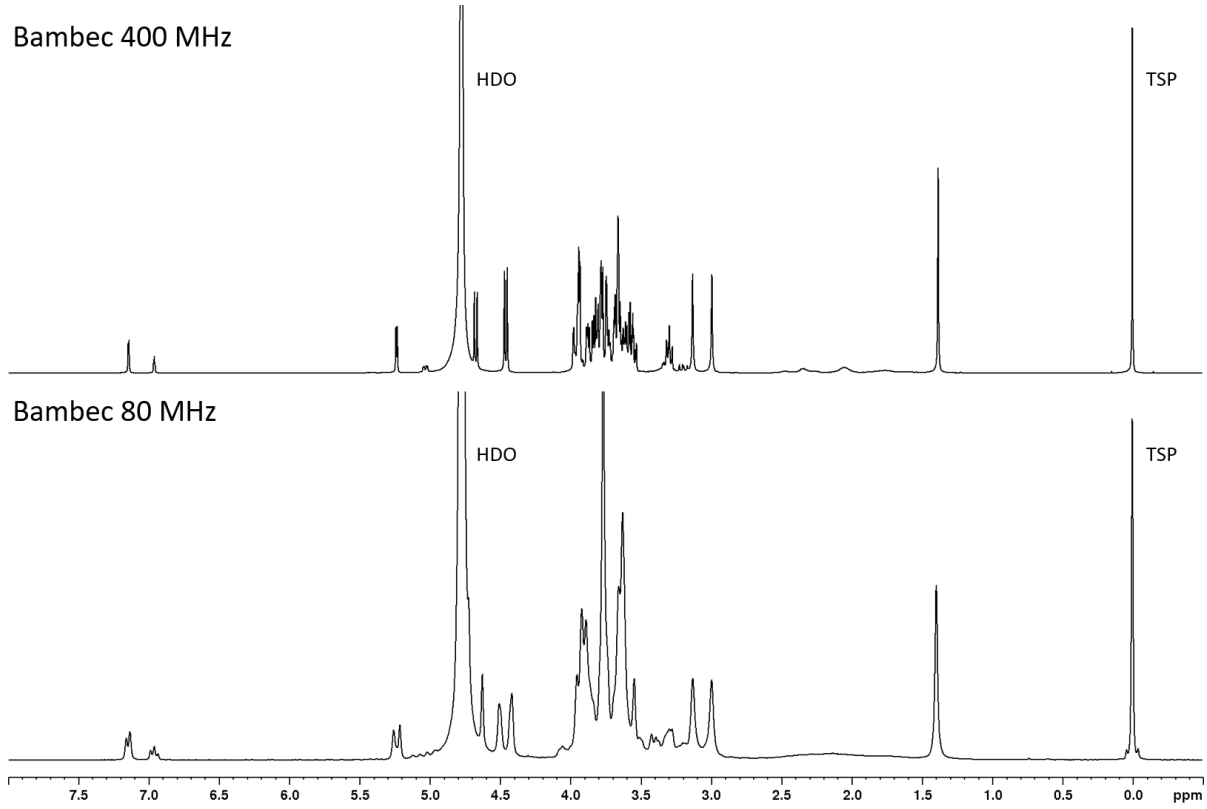


Figure 4 ^{1}H NMR spectra of Bambec recorded with the 400 MHz (D_2O; 16 Scans; top) and the 80 MHz (D_2O; 64 Scans; bottom) spectrometer.

In the final measurement, which was then also sent to Magritek (Figure 4), the result for the 400 MHz spectrometer were worse than in the preliminary tests and were partially outside the 5% limit. This was the case for both the 400 MHz spectrometer (9.20 mg (92.0%)) and the 80 MHz spectrometer (9.23 mg (92.3%)). Nevertheless, if the results from both spectrometers were compared, the contents were in good accordance and deviate only 0.3% from each other.

Since the signals in the aromatic region of the Quinaplus spectrum were very close to each other, a sample with a known content of quinapril and HCT was measured first to see if a quantification was possible. For this purpose, only the two active ingredients were dissolved in DMSO-d_6 without any excipients and measured. For this sample, a content of 102.8% was calculated for quinapril based on the aromatic protons (*9H-Ar*), and for HCT, contents between 98.8% and 102.6%, depending on the choice of signal. Since the signals *B* and *C* of HCT overlapped in this sample without excipients (see Figure 2), they were integrated and evaluated together. Due to these results, a tablet of Quinaplus was measured in the next step.

The spectrum of Quinaplus (400 MHz) showed that the excipients had a considerable influence (Figure 2). The signals of the excipients and APIs overlapped in the spectral range of 5.3 – 3.7 ppm and 1.3 – 1.0 ppm, and they led to signal shifts, causing the signals in the aromatic region to be even closer to each other. In addition, the signals of the sulphonamide protons of HCT (signal *D* and *C*) became very broad, with signal *C* no longer recognizable as such unless

the entire spectral region containing signals *A* to *D* was integrated. This affected the integration of other signals, especially that of the aromatic proton next to the sulphonamide (signal *B*), resulting in a large deviation of the content compared with the other two signals. Depending on the signal, a content (relative to the declared content) between 92.0% and 101.8% was obtained for HCT and 86.2% for quinapril (Table 3). Considering only the two proton signals for HCT, which were least affected by the excipients, a content of 92.7% (RNH-R; signal *A*) and 92.0% (RCH-R; signal *E*) was obtained. This is much lower than declared, even though the tablets were stored in a climate-controlled laboratory with light excluded. It stands to reason that the quantification method is not yet ideal, especially if the influences of the excipients are also taken into account.

Table 3 Calculated content of quinapril and HCT for different signals and spectrometers for the same sample (n = 1). The content in percent refers to the labeled content.

Quinaplus	**Quinapril**		**HCT**					
	9H-Ar		A		B		E	
400 MHz	18.67 mg	86.2%	11.59 mg	92.7%	12.72 mg	101.8%	11.50 mg	92.0%
80 MHz	18.34 mg	84.7%	11.77 mg	94.2%	13.72 mg	109.8%	13.20 mg	105.6%
Difference	0.33 mg	1.5%	0.18 mg	1.5%	1.00 mg	8.0%	1.70 mg	13.6%

Compared to the results of the 80 MHz spectrometer, the differences between the API contents were larger for Quinaplus than for Bambec. While the difference was only 1.5% for Quinapril, it varied between 1.5 and 13.6% for HCT. This was due to the lower resolution of the instrument, as well as the influence of the excipients. This was most prominent for the aromatic proton signal (signal *E*) with a difference of 13.6%. Here, the 80 MHz spectrometer was no longer able to completely resolve this signal from the signal of the aromatic protons of quinapril (*9H-Ar*) (Figure 5). In addition, the S/N ratios were also significantly lower again, with the signal of the amin (signal *A*) even dropping below 250 (85.6).

It was shown that the limits of the method for Quinaplus were reached here, since on the one hand no ideal signal could be found for quinapril and on the other hand the influence of the excipients in the tablet led to overlaps of the relevant signals. Due to this strong influence of the excipients on the HCT signals, it was unfortunately also not possible to make an optimal comparison here, even though it could be seen from signal A, which was the least affected, that it should be possible to obtain very similar results. Since these were only initial exploratory experiments, it should be possible to obtain better spectra and thus significantly better results with optimized processing or sample preparation. For example, it would be possible to shift signals by means of the measurement temperature in order to further separate signals from each other. In addition, during the sample preparation of Quinaplus, the addition of D_2O could

be used to eliminate the NH signals from HCT to allow better integration of the aromatic protons.

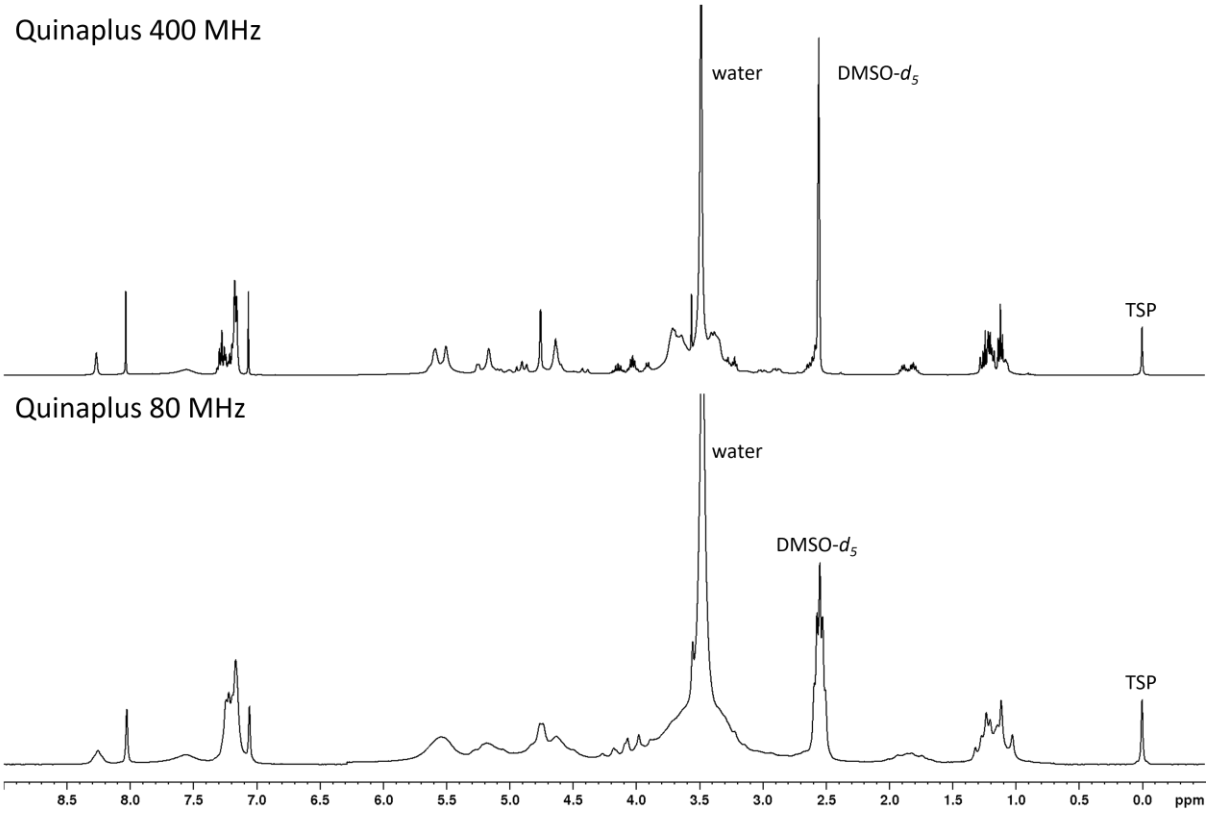


Figure 5 ^{1}H NMR spectra of Quinaplus recorded with the 400 MHz (DMSO-d_6; 16 Scans; top) and the 80 MHz (DMSO-d_6; 64 Scans; bottom) spectrometer.

4. Conclusion

With the measurements performed, the 400 MHz and 80 MHz spectrometer could be successfully compared and the problems, but also the possibilities, could be shown. Especially for Bambec with a deviation of 0.3% in bambuterol content it could be shown that it is possible to do proper quantifications of drug samples even with a benchtop spectrometer. In addition, the qNMR standards offer further advantages. For example, TSP is less expensive than the corresponding CRS for HPLC analysis and does not interfere with drugs and excipients, since no other signals are normally found in the spectral range of 0 ppm. However, with Quinaplus it became clear which problems are to be expected in the quantification of drugs by NMR spectroscopy, such as the overlapping and shifts of signals due to the excipients. These have an even greater impact on benchtop instruments due to their lower resolution. However, with a revised method, the benchtop device should be able to quantify the drug substances here as well, as has been shown in other cases [5]. With these capabilities, it is also a very good alternative for smaller companies, such as pharmacies working in unit dose, as it provides a fast and reliable measurement that also eliminates the need to have a reference substance for each individual API.

Acknowledgements

Special thanks to Dr. Winterfeld of Montanus pharmacies for providing the drugs and the quinapril reference standard, and to Dr. Sanel Suljić and Dr. Jürgen Kolz from Magritek for conducting the 80 MHz measurements.

References

[1] Danieli E, Perlo J, Blümich B, Casanova F Small Magnets for Portable NMR Spectrometers. *Angew Chem Int Ed* (2010) 49, 4133-4135

[2] Blümich B, Singh K Desktop NMR and Its Applications From Materials Science To Organic Chemistry. *Angew Chem Int Ed* (2018) 57, 6996-7010

[3] Assemat G, Gouilleux B, Bouillaud D, Farjon J, Gilard V, Giraudeau P, Malet-Martino M Diffusion-ordered spectroscopy on a benchtop spectrometer for drug analysis. *J Pharm Biomed Anal* (2018) 160, 268-275

[4] Dalitz F, Cudaj M, Maiwald M, Guthausen G Process and reaction monitoring by low-field NMR spectroscopy. *Prog Nucl Magn Reson Spectrosc* (2012) 60, 52-70

[5] Zivkovic A, Bandolik JJ, Skerhut AJ, Coesfeld C, Prascevic M, Zivkovic L, Stark H Quantitative Analysis of Multicomponent Mixtures of Over-the-Counter Pain Killer Drugs by Low-Field NMR Spectroscopy. *J Chem Educ* (2017) 94, 121-125

[6] Assemat G, Balayssac S, Gerdova A, Gilard V, Caillet C, Williamson D, Malet-Martino M Benchtop low-field 1H Nuclear Magnetic Resonance for detecting falsified medicines. *Talanta* (2019) 196, 163-173

[7] Beyer T, Schollmayer C, Holzgrabe U The role of solvents in the signal separation for quantitative 1H NMR spectroscopy. *J Pharm Biomed Anal* (2010) 52, 51-58

[8] Shen S, Yang X, Shi Y Application of quantitative NMR for purity determination of standard ACE inhibitors. *J Pharm Biomed Anal* (2015) 114, 190-199

[9] Abiramasundari A, Sudarsanam V, Vasu KK Characterization of the degradation products of bambuterol using LCMS-QTOF and NMR. *Analytical Methods* (2015) 7, 7659-7673

[10] Saito T, Yamaji T, Hayamizu K, Yanagisawa M, Yamamoto O SDBSWeb - Spectral database for organic compounds. National Institute of Advanced Industrial Science and Technology. https://sdbs.db.aist.go.jp. Accessed 14.01.2019

[11] Harmon PA, Yin W, Bowen WE, Tyrrell RJ, Reed RA Liquid chromatography–mass spectrometry and proton nuclear magnetic resonance characterization of trace level condensation products formed between lactose and the amine-containing diuretic hydrochlorothiazide. *J Pharm Sci* (2000) 89, 920-929

[12] Binev Y, Marques MMB, Aires-de-Sousa J Prediction of 1H NMR Coupling Constants with Associative Neural Networks Trained for Chemical Shifts. *J Chem Inf Model* (2007) 47, 2089-2097

[13] Holzgrabe U Quantitative NMR spectroscopy in pharmaceutical applications. *Prog Nucl Magn Reson Spectrosc* (2010) 57, 229-240

[14] (EMA) EMA (1991) Specifications and Control Tests on the Finished Product. 1.5.2 Maximum acceptable deviation in the content of active substances.

[15] BfArM (1996) Bekanntmachung über die Zulassung von Arzneimitteln durch das Bundesinstitut für Arzneimittel und Medizinprodukte.

4. Final Discussion

This thesis aimed at the analysis of APIs, excipients, and finished drug formulations with new approaches of established spectroscopic methods and chemometric evaluation procedures with regard to origin, stability and drug content to show the possibilities in the present time, as there has been an outstanding development within the recent years. These issues were addressed by various projects. For tracing the origin of different paracetamol tablets the combined information of MIR, NIR and NMR spectroscopy was used in combination with chemometric evaluation. In this context, a method for distinguishing different batches using NIR spectroscopy was successfully developed. DOSY measurements were demonstrated to be a useful tool as a screening method to investigate possible reactions between active ingredients and excipients on the one hand and as a filter for signals in order to obtain better fingerprint spectra (^{1}H NMR) on the other hand. In addition, the possibilities of a quantitative NMR method using a benchtop spectrometer were shown. The results of the projects presented in chapter 3 are now discussed in their entirety below, including their capabilities and limitations.

4.1 Spectroscopic analysis of different paracetamol formulations

Tracing the origin of paracetamol tablets by means of PCA, HCA and LDA

Since in many cases nowadays it is no longer possible to clearly trace the origin of individual drugs, an attempt was made to determine their origin by means of chemometric evaluation of spectral data. It could be demonstrated that this combination is a proper tool to determine the origin of drugs and also to characterize them. However, the determination of origin is limited to the pharmaceutical company and manufacturer, respectively, and not to geographical regions as it is possible for plants [1]. Unlike plants, whose composition of ingredients depends on regional conditions such as weather and nutrients in the soil, this is obviously not the case for drugs, where general requirements of the starting materials are defined by pharmacopoeial monographs.

Regarding the sample preparation for the spectral methods, it was shown that it is possible to obtain reproducible spectra allowing for a determination of the origin, despite the strongly varying ingredients. However, to obtain even more information from the formulations, two additional approaches could be pursued in further experiments. For NMR spectroscopy, the tablets could be dissolved in different solvents and measured separately. This would allow more types and larger quantities of the excipients to be dissolved. Subsequently, the spectra would need to be evaluated chemometrically and the information obtained combined. For MIR

and NIR spectroscopy it would be advantageous to measure the whole tablet without mortaring it. This would include other information such as compression pressure and particle size [2,3].

Besides the determination of the origin by means of the chemometric methods, the preparations can also be characterized very well by means of their NMR data. The NMR spectra can be used to compare the information in the package inserts, if available, and the excipients actually used. In addition, with an appropriately validated method, it would also be possible to check the content of the active ingredients. Further possibilities for characterization are provided by the PCA and HCA themselves. Based on the loadings in the PCA, it is usually very easy to see which excipients are characteristic of a formulation and contribute to the formation of clusters. This information could in turn be confirmed by the HCA, since mostly the same clusters formed there.

Generation of fingerprint spectra

The use of DOSY measurements can act as a kind of filter in order to obtain ^{1}H NMR spectra with only the necessary signals to distinguish and characterize the formulations. As shown, the characteristic part of the ^{1}H NMR spectra of the paracetamol tablets were the signals of the excipients. With the adequate choice of parameters, it was possible to simplify the spectra by eliminating unnecessary signals within a few minutes of measurement time. In particular the signals of the solvent, residual water and that of the API. In addition to that, more signals came to light, which allows the access to much more specific spectra. The downsides of this method are that not all excipients are accessible to NMR spectroscopy and that signals of molecules with no sufficient difference in the molecular mass to the API will distinguish, which is particularly problematic for drug formulations with few excipients.

Nevertheless, this is a valuable method to obtain clearer spectra and extract more information from them, allowing better characterization and chemometric evaluation. For the future, this method should be further tested by applying it to all the formulations and compare the chemometric results with the ones of the regular spectra.

NIR attachment for batch distinction

The goal of developing an attachment was to take advantage of the additional physical information, such as compression pressure and particle size, contained in the NIR spectra of whole tablets [2,3]. This was fully achieved by developing a proprietary 3D attachment, which is capable of measuring the tablet without any sample preparation while accurately adhering to a number of parameters. The relevant parameters are a constant operating temperature of the lamp, repositioning of the tablet before each measurement to avoid trends, and the execution of a series of measurements in one piece, since the attachment cannot be mounted and remounted in a defined, fixed position.

The suitability of the attachment was demonstrated by successfully distinguishing the batches of nine different paracetamol preparations using chemometric as well as statistical methods. This was done by means of PCA and statistical evaluation, for which the scores values of the first principal component of PCA were also used. The prerequisite is a clear differentiation of the batches by means of the first principal component.

Although the attachment is not perfectly engineered for every tablet shape, it can be used in accordance with the mentioned parameters not only for batch distinction, but also to assign tablets to their pharmaceutical companies. For this purpose, it was also successfully employed in the student's educatinal laboratory to introduce students to spectroscopy as well as chemometrics.

4.2 Analysis of stressed mixtures by DOSY

A new approach to the study of reactions between active ingredients and excipients was pursued using DOSY as a screening method for stressed mixtures. With this method the capabilities of NMR spectroscopy in the routine analytics of impurity profiling could be demonstrated. The APIs, excipients and reaction products of 5-ASA and citric acid, 5-ASA and formic acid, and benzocaine and dibutylphthalate were successfully detected and separated based on their translational diffusion behavior with measurement times of less than one hour. Limitations regarding large signals that can affect small signals in the immediate vicinity and lead to false cross peaks or only slightly overlapping signals have been overcome by careful processing. This was done by using Dynamic Center (Bruker BioSpin GmbH) which allows a fragmented analysis of signals split by direct spin-spin coupling for calculating the 2D spectra. This also shows that it can be advantageous to look at the same data with different processing software.

However, the method has its limitations when large and small signals of different substances overlap, thus strongly distorting the diffusion coefficient indication. Nevertheless, this method is very well suited for screening mixtures such as drug formulations and not only for calculating the diffusion coefficient.

4.3 Quantitative NMR spectroscopy of drugs using a high field and a low field spectrometer

Even though the study was an exploratory method, which still requires a certain degree of optimization, it was shown that it is possible to achieve comparable results with a compact 80 MHz spectrometer as with a 400 MHz spectrometer. In particular, for bambuterol and quinapril, the deviations between spectrometers were only 0.3% - 1.5% with the most suitable

signals for quantification. At the same time the problems of benchtop instruments and qNMR spectroscopy in general for drug formulations were highlighted. This was illustrated by the example of HCT, where no adequate comparison was possible. For HCT, content fluctuations between 1.5% and 13.6% were measured, which on the one hand is due to the excipients and on the other hand to the lower resolution of the benchtop instrument. The former is influential to the signal position and width, which can lead to signals overlapping and thus no longer allowing for an adequate integration. This leads to problems more quickly with the benchtop devices with their lower resolution.

Nevertheless, this possibility of quantification is an interesting alternative, especially for smaller companies, as it also includes the general advantages of NMR spectroscopy. Usually, only one internal standard is needed for different active ingredients or formulations, which can reduce costs compared to the CRS standards used in HPLC analysis. In addition, the spectrum contains more information, since all detectable substances are measured and thus, for example, new impurities can also be detected.

In the future, however, this measurement method would need to be further optimized before it can be validated and used. This concerns HCT in particular. In order to improve the signal location and width of some signals, additional solvents or solvent mixtures could be tested. For this purpose, a small amount of D_2O should be added in addition to the DMSO-d_6 during sample preparation to eliminate the NH signals. However, the influence on the shim and the lock must be considered. Alternatively, a DMSO-d_6 with a lower content of water can be used or the measurement temperature can be changed to influence the width and location of the signals.

4.4 Overall conclusion

Although spectroscopic methods have been in use for decades, this work has shown that there are still opportunities to find new approaches or to further develop old ones that represent added value in the pharmaceutical sector.

Especially the combination of chemometric methods and spectroscopic data is a great advantage in order to make comprehensive use of the many different types of information contained in the spectra, e.g., as described in the identity or origin determinations in this work. However, the fundamental prerequisite obtaining trustworthy results is a representative database consisting of authentic samples, as well as a profound comprehension of the possibilities and limitations of the pre-processing methods and individual models. If these circumstances are given, the chemometric evaluation is a very powerful tool that can be used for many purposes [4-6].

Yet individual spectroscopic measurements, without chemometric evaluation, can also provide important information. DOSY measurements proved to be an excellent screening method to analyze different mixtures. They can be used as proof of identity, since the DOSY spectrum can also be understood as a fingerprint as well as for purity or stability analysis, in order to detect new and unknown substances, as in the examples described. However, to use this method effectively, a comprehensive knowledge of the samples should be available as well as the awareness that not all substances can be detected by NMR.

The benchtop instruments, in turn, provide access to NMR spectroscopy to a wider audience, as they are more space-saving and cheaper to purchase, as well as to maintain. In addition, they are meanwhile capable of performing a wide range of different measurements with decent results. These include qualitative 1D measurements as well as the possibility of quantitative determinations or DOSY measurements [7-9]. Nevertheless, these systems can not be expected to be on par with highfield spectrometers in every respect.

Thus, it can be said that the results shown reinforce the importance of spectroscopic methods and demonstrate that they can and should be used on a larger scale in the pharmaceutical sector as well. This applies not only to large pharmaceutical companies, but also to smaller companies such as pharmacies.

References

[1] Wang P, Yu Z Species authentication and geographical origin discrimination of herbal medicines by near infrared spectroscopy: A review. *J Pharm Anal* (2015) 5, 277-284

[2] Sparén A, Hartman M, Fransson M, Johansson J, Svensson O Matrix Effects in Quantitative Assessment of Pharmaceutical Tablets Using Transmission Raman and Near-Infrared (NIR) Spectroscopy. *Appl Spectrosc* (2015) 69, 580-589

[3] Matero S, van Den Berg F, Poutiainen S, Rantanen J, Pajander J Towards better process understanding: chemometrics and multivariate measurements in manufacturing of solid dosage forms. *J Pharm Sci* (2013) 102, 1385-1403

[4] Hohmann M, Monakhova Y, Erich S, Christoph N, Wachter H, Holzgrabe U Differentiation of Organically and Conventionally Grown Tomatoes by Chemometric Analysis of Combined Data from Proton Nuclear Magnetic Resonance and Mid-infrared Spectroscopy and Stable Isotope Analysis. *J Agric Food Chem* (2015) 63, 9666-9675

[5] Custers D, Cauwenbergh T, Bothy JL, Courselle P, De Beer JO, Apers S, Deconinck E ATR-FTIR spectroscopy and chemometrics: An interesting tool to discriminate and characterize counterfeit medicines. *J Pharm Biomed Anal* (2015) 112, 181-189

[6] Challa S, Potumarthi R Chemometrics-based process analytical technology (PAT) tools: applications and adaptation in pharmaceutical and biopharmaceutical industries. *Appl Biochem Biotechnol* (2013) 169, 66-76

[7] Wu N, Balayssac S, Danoun S, Malet-Martino M, Gilard V Chemometric Analysis of Low-field 1H NMR Spectra for Unveiling Adulteration of Slimming Dietary Supplements by Pharmaceutical Compounds. *Molecules* (2020) 25, 1193

[8] Assemat G, Gouilleux B, Bouillaud D, Farjon J, Gilard V, Giraudeau P, Malet-Martino M Diffusion-ordered spectroscopy on a benchtop spectrometer for drug analysis. *J Pharm Biomed Anal* (2018) 160, 268-275

[9] Zivkovic A, Bandolik JJ, Skerhut AJ, Coesfeld C, Prascevic M, Zivkovic L, Stark H Quantitative Analysis of Multicomponent Mixtures of Over-the-Counter Pain Killer Drugs by Low-Field NMR Spectroscopy. *J Chem Educ* (2017) 94, 121-125

5. Summary

Spectroscopic methods were established decades ago in a wide variety of fields. This also applies to the pharmaceutical field, although they initially were mostly used for identity testing or structure elucidation only. Technical developments, such as miniaturization (NMR benchtop devices), Fourier transformations (for NMR, MIR spectroscopy) or the combination with chemometric evaluation (e.g., in Process Analytical Technology, PAT), have further increased their importance and opened up new applications. The aim of this work was to investigate further new approaches and to find new applications for already established methods and to show their benefits.

By means of MIR, NIR and NMR data and their chemometric evaluation (principal component analysis, PCA; hierarchical cluster analysis, HCA; linear discriminant analysis, LDA), possibilities were presented to successfully determine the manufacturer or the pharmaceutical company of various paracetamol preparations. In the course of this, various similarities and correlations between the preparations of individual companies could also be identified. For this purpose, a suitable sample preparation was developed for each spectroscopic method, and suitable measurement parameters in order to obtain reproducible spectra for the chemometric evaluation were determined. Furthermore, the results of the two unsupervised methods (HCA, PCA) were compared with each other. The HCA was able to confirm those of the PCA for the very most part. Additionally, through these methods it was possible to characterize many of the preparations based on clusters formed by comparable tablet compositions.

In order to be able to measure unmortared, whole tablets using the NIR spectrometer, an attachment was developed and manufactured using 3D printing. Its functionality was demonstrated by measuring and analyzing the tablets of two different batches of nine paracetamol preparations. The batches were clearly distinguished on the basis of a PCA and a significant difference was also demonstrated by means of statistical tests.

For NMR spectroscopy, a method was developed to obtain optimized "fingerprint" spectra of drug formulations. For this purpose, a 1D DOSY measurement was elaborated, in which the signals of the active ingredient could be filtered out by the appropriate choice of measurement parameters. The chemometric evaluation can thus focus on the remaining signals of the excipients, on the basis of which the preparations of the same API can be distinguished. Especially in the case of formulations that consist largely of active ingredient, data pre-processing of the spectra can thus be simplified and greater importance can be assigned to the originally very small excipient signals.

A quantitative ^{1}H NMR method was developed for the comparison of a high-field spectrometer (400 MHz) with a benchtop spectrometer (80 MHz) for two finished drugs. It was shown that it

is possible to obtain comparable results with both instruments, but that the influence of the excipients on the signals and the lower resolution of the benchtop instrument must be taken into account. Therefore, it was not possible to obtain comparable results without further optimization of the method for one of the active ingredients.

In the investigation of various reactions between APIs and excipients using DOSY, its usefulness as a screening method in stability testing was demonstrated. For this purpose, three different APIs and excipients were stressed together and the reaction mixtures were subsequently measured using DOSY. Based on the translational diffusion coefficient, the reaction products could be identified and distinguished from the active ingredients and the excipients used. The importance of thoughtful processing could also be demonstrated. If all peak heights are selected when evaluating signals split by direct spin-spin-coupling, this allows the detection of hidden signals as long as not all signals have the same diffusion coefficient. The selective selection of individual peak heights in the case of split signals also enables the evaluation of signals that overlap slightly. However, the limitations of this method were also shown when two signals overlap too much and differ too little in their diffusion coefficients.

Hence, it has been successfully demonstrated in the various projects that the new chemometric approaches, as well as the new applications of already established methods, enable in-depth findings and thus have a clear added value.

6. Zusammenfassung

Spektroskopische Methoden haben sich schon vor Jahrzehnten in den verschiedensten Bereichen etabliert. Dies betrifft auch den pharmazeutischen Bereich, auch wenn sie hier zunächst meist nur zur Identitätsprüfung oder Strukturaufklärung verwendet wurden. Durch technische Weiterentwicklungen, wie Miniaturisierungen (NMR benchtop Geräte), Fourier-Transformationen (NMR, MIR) oder die Kombination mit einer chemometrischen Auswertung (z. B. bei Process Analytical Technology, PAT), haben sie weiter an Bedeutung gewonnen, und es wurden neue Einsatzbereiche erschlossen. Das Ziel der vorliegenden Arbeit war es, weitere neue Ansätze zu untersuchen und neue Anwendungen für bereits etablierte Methoden zu finden und deren Mehrwert aufzuzeigen.

Es wurden Möglichkeiten aufgezeigt mittels MIR-, NIR- und NMR-Daten und deren chemometrischen Auswertungen (Hauptkomponentenanalyse, PCA; hierarchische Clusteranalyse, HCA; lineare Diskriminanzanalyse, LDA) erfolgreich den Hersteller bzw. das pharmazeutische Unternehmen verschiedener Paracetamol-Präparate zu bestimmen. In diesem Zuge konnten Ähnlichkeiten zwischen Präparaten unterschiedlicher Firmen identifiziert werden. Um dies zu erreichen, wurde für jede spektroskopische Methode eine geeignete Probenvorbereitung entwickelt sowie geeignete Messparameter festgelegt, um reproduzierbare Spektren für die chemometrische Auswertung zu erhalten. Weiterhin wurden die Ergebnisse der zwei unüberwachten Methoden (HCA, PCA) miteinander verglichen, wobei die HCA die der PCA zum allergrößten Teil bestätigen konnte. Zudem war es möglich durch diese Methoden viele der Präparate anhand von Clustern zu charakterisieren, die durch vergleichbare Tablettenzusammensetzungen gebildet wurden.

Um mit Hilfe des NIR-Spektrometers intakte Tabletten vermessen zu können, wurde ein Aufsatz entwickelt und mittels 3D-Druck hergestellt. Dessen Funktionalität wurde überprüft, indem Tabletten aus je zwei unterschiedlichen Chargen von neun Paracetamol-Präparaten vermessen und analysiert wurden. Dabei konnten die Batches anhand einer PCA eindeutig unterschieden und zudem mittels statistischer Tests ein signifikanter Unterschied nachgewiesen werden.

Für die NMR-Spektroskopie wurde eine Methode entwickelt, um optimierte „Fingerprint“ Spektren von Arzneimittelformulierungen zu erhalten. Dazu wurde eine 1D-DOSY-Messmethode erarbeitet, bei der durch die passende Wahl der Messparameter die Signale des Wirkstoffes herausgefiltert werden konnten. Die chemometrische Auswertung konnte sich somit auf die Signale der Hilfsstoffe beschränken, anhand derer die Präparate unterschieden werden können. Vor allem bei Formulierungen, die zum größten Teil aus

Wirkstoff bestehen, kann so eine Datenvorverarbeitung der Spektren vereinfacht und den ursprünglich sehr kleinen Hilfsstoffsignalen eine größere Bedeutung beigemessen werden.

Für den Vergleich eines Hochfeld-Spektrometers (400 MHz) mit einem „benchtop“-Spektrometer (80 MHz) wurde für zwei Fertigarzneimittel eine quantitative ^{1}H-NMR-Methode entwickelt. Es konnte gezeigt werden, dass es möglich ist, mit beiden Geräten vergleichbare Ergebnisse zu erzielen. Dabei ist jedoch der Einfluss der Hilfsstoffe auf die Signale sowie die geringere Auflösung des „benchtop“-Gerätes zu berücksichtigen. Aus diesen Gründen war es ohne eine weitere Optimierung der Methode für einen der Wirkstoffe nicht möglich vergleichbare Ergebnisse mit beiden Geräten zu erzielen.

Bei der Untersuchung verschiedener Reaktionen zwischen Wirk- und Hilfsstoffen mittels DOSY konnte dessen Nutzen als Screening-Methode bei Stabilitätstests gezeigt werden. Für diesen Zweck wurden drei verschiedene Wirk- und Hilfsstoffe gemeinsam gestresst und die Reaktionsgemische anschließend mittels DOSY vermessen. Anhand des translationalen Diffusionskoeffizienten konnten die Reaktionsprodukte identifiziert und von den eingesetzten Wirk- und Hilfsstoffen unterschieden werden. Ebenso konnte die Bedeutung einer sorgfältigen Prozessierung demonstriert werden. Werden bei der Auswertung von Signalen, die durch direkte Spin-Spin-Kopplung aufgespalten wurden, alle Peakhöhen ausgewählt, erlaubt dies die Detektion von versteckten Signalen, falls nicht alle Signale den gleichen Diffusionskoeffizienten besitzen. Die selektive Auswahl einzelner Peakhöhen bei aufgespaltenen Signalen ermöglicht zudem die Auswertung von leicht überlappenden Signalen. Es wurden jedoch auch die Grenzen dieser Methode aufgezeigt: wenn zwei Signale zu stark überlappen und sich dabei in ihrem Diffusionskoeffizienten zu wenig unterscheiden.

Somit konnte in den verschiedenen Projekten erfolgreich gezeigt werden, dass die neuen chemometrischen Ansätze, sowie die neuen Anwendungen bereits etablierter Methoden vertiefte Erkenntnisse ermöglichen und somit einen deutlichen Mehrwert besitzen.

7. Appendix

7.1 List of publications

Research papers

Becht, A.; Schollmayer, C.; Wiest J.; Heller D.; Baumann W.; Buschmann H.; Holzgrabe U. *Diffusion ordered NMR spectroscopy measurements as screening method of potential reactions of API and excipients in drug formulations,* J Pharm Biomed Anal **2019**, 162, 41-46.

Becht, A.; Schollmayer, C.; Monakhova, Y.; Holzgrabe U. *Tracing the origin of paracetamol tablets by near-infrared, mid-infrared, and nuclear magnetic resonance spectroscopy using principal component analysis and linear discriminant analysis,* Anal Bioanal Chem **2021**, 413, 3107-3118

Becht, A.; Schollmayer, C.; Holzgrabe U. *Fingerprint spectra for drug formulations using a DOSY filter: Pros and Cons,* J Pharm Biomed Anal **2022**, 214, 114723.

Belugina R.B.; Monakhova, Y.B.; Rubtsova, E.; Becht, A.; Schollmayer, C.; Holzgrabe, U.; Legin, A.V.; Kirsanov D.O. *Distinguishing paracetamol formulations: Comparison of potentiometric "Electronic Tongue" with established analytical techniques*, J Pharm Biomed Anal **2020**, 188, 113457

Schripsema, J.; Merlim, R.d.S.; Parvan, L.G.; Ribeiro, H.S.d.S.; Schripsema, L.D.; Aleixo, S.; Becht, A.; Holzgrabe, U.; Dagnino, D. *Towards a holistic view of tablet quality, an extensive study on paracetamol tablets with nuclear magnetic resonance using similarity calculations, differential NMR and hierarchical cluster analysis*, J Pharm Biomed Anal (submitted)

Poster contributions

DOSY measurements as screening method of potential reactions of API and excipients in drug formulations

qNMR Summit 2018

10. – 11. October, 2018, Würzburg, Germany

Alexander Becht, Curd Schollmayer, Johannes Wiest, Detlef Heller, Wolfgang Baumann, Helmut Buschmann, Ulrike Holzgrabe

7.2 Documentation of authorship

In this section, the individual contribution for each author to the publications reprinted in this thesis is specified. Unpublished manuscripts are handled, accordingly.

Erklärung zur Autorenschaft

Diffusion ordered NMR spectroscopy measurements as screening method of potential reactions of API and excipients in drug formulations

A. Becht; C. Schollmayer; J. Wiest; D. Heller; W. Baumann; H. Buschmann; U. Holzgrabe

Journal of Pharmaceutical and Biomedical Analysis 162 (2019) 41-46

Alexander Becht (AB), Curd Schollmayer (CS), Johannes Wiest (JW), Detlef Heller (DH), Wolfgang Baumann (WB), Helmut Buschmann (HB), Ulrike Holzgrabe (UH)

Autor	**AB**	**CS**	**JW**	**DH**	**WB**	**HB**	**UH**					**∑ in Prozent**
Studiendesign	3	2		1	1	1	2					10%
Experimentelle Arbeit	25	2.5										27.5%
Datenanalyse und Interpretation	22.5	5					2.5					30%
Verfassen der Veröffentlichung	15											15%
Korrektur der Veröffentlichung		1.5	1.5	1.5	1.5	1.5	3					10.5%
Koordination der Veröffentlichung							7					7%
Summe	65.5	11	1.5	2.5	2.5	2.5	14.5					100%

Die Mitautoren der in dieser (teil-)kumulativen Dissertation verwendeten Manuskripte sind sowohl über die Nutzung als auch über die angegebenen Eigenanteile informiert und stimmen dem zu.

Alexander Becht

Hauptautor/in

☒ Verweis: E-Mail hinterlegt

Curd Schollmayer

Koautor/in

☒ Verweis: E-Mail hinterlegt

Johannes Wiest

Koautor/in

☒ Verweis: E-Mail hinterlegt

Detlef Heller

Koautor/in

☒ Verweis: E-Mail hinterlegt

Wolfgang Baumann

Koautor/in

☒ Verweis: E-Mail hinterlegt

Helmut Buschmann

Koautor/in

☒ Verweis: E-Mail hinterlegt

Ulrike Holzgrabe

Korrespondenzautor/in

☒ Verweis: E-Mail hinterlegt

Würzburg, 20.03.2022

_____

Prof. Dr. Ulrike Holzgrabe

Erklärung zur Autorenschaft

Tracing the origin of paracetamol tablets by near-infrared, mid-infrared, and nuclear magnetic resonance spectroscopy using principal component analysis and linear discriminant analysis

A. Becht; C. Schollmayer; Y. Monakhova; U. Holzgrabe

Analytical and Bioanalytical Chemistry 413 (2021) 3107-3118

Alexander Becht (AB), Curd Schollmayer (CS), Yulia Monakhova (YM), Ulrike Holzgrabe (UH)

Autor	AB	CS	YM	UH								∑ in Prozent
Studiendesign	2.5			2.5								5%
Experimentelle Arbeit	17.5	2.5										20%
Multivariate Datenanalyse für die einzelnen Daten jeder Methode	20		2.5									22.5%
Multivariate Datenanalyse für kombinierte Daten aus mehreren Methoden			2.5									2.5%
Interpretation der Multivariaten Datenanalyse	15		2.5									17.5%
Verfassen der Veröffentlichung	15											15%
Korrektur der Veröffentlichung		2.5	5	5								12.5%
Koordination der Veröffentlichung				5								5%
Summe	70	5	12.5	12.5								100%

Die Mitautoren der in dieser (teil-)kumulativen Dissertation verwendeten Manuskripte sind sowohl über die Nutzung als auch über die angegebenen Eigenanteile informiert und stimmen dem zu.

Alexander Becht

Hauptautor/in

☒ Verweis: E-Mail hinterlegt

Curd Schollmayer

Koautor/in

☒ Verweis: E-Mail hinterlegt

Yulia Monakhova

Koautor/in

☒ Verweis: E-Mail hinterlegt

Ulrike Holzgrabe

Korrespondenzautor/in

☒ Verweis: E-Mail hinterlegt

Würzburg, 20.03.2022 _____

Prof. Dr. Ulrike Holzgrabe

Erklärung zur Autorenschaft

Fingerprint spectra for drug formulations using a DOSY filter: Pros and Cons

Journal of Pharmaceutical and Biomedical Analysis 214 (2022) 114723

A. Becht; C. Schollmayer; U. Holzgrabe

Alexander Becht (AB), Curd Schollmayer (CS), Ulrike Holzgrabe (UH)

Autor	AB	CS	UH	A4	A5	A6	A7	A8	A9	A10	A11	∑ in Prozent
Studiendesign	2.5	2.5	2.5									7.5%
Experimentelle Arbeit	30											30%
Datenanalyse und Interpretation	20	5	5									30%
Verfassen der Veröffentlichung	15		5									20%
Korrektur der Veröffentlichung		2.5	5									7.5%
Koordination der Veröffentlichung			5									5%
Summe	67.5	10	22.5									100%

Die Mitautoren der in dieser (teil-)kumulativen Dissertation verwendeten Manuskripte sind sowohl über die Nutzung als auch über die angegebenen Eigenanteile informiert und stimmen dem zu.

Alexander Becht	Curd Schollmayer	Ulrike Holzgrabe
Hauptautor/in	**Koautor/in**	**Korrespondenzautor/in**
☒ Verweis: E-Mail hinterlegt	☒ Verweis: E-Mail hinterlegt	☒ Verweis: E-Mail hinterlegt

Würzburg, 20.03.2022

Prof. Dr. Ulrike Holzgrabe