Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Pharmaceutical and Biomedical Analysis 38 (2005) 833–839
JOURNAL OF PHARMACEUTICAL AND BIOMEDICAL ANALYSIS
www.elsevier.com/locate/jpba

Multivariate analysis of nuclear magnetic resonance data—characterization of critical drug substance quality of gentamicin sulfate

Werner Winter[a,*], Ralph Deubner[b,1], Ulrike Holzgrabe[b]

[a] *Birkenstr. 5, 52078 Aachen, Germany*
[b] *Institute of Pharmacy and Food Chemistry, University of Würzburg, Am Hubland, 97074 Würzburg, Germany*

Received 2 September 2004; received in revised form 29 January 2005; accepted 31 January 2005
Available online 26 April 2005

Abstract

Fourteen gentamicin sulfate lots collected from international markets showed high quantities of impurities (∼30% of studied lots). ^{1}H NMR spectroscopy as a primary analytical method was applied in order to validate the quantification results obtained from micellar electrokinetic chromatography method (MEKC). In this study, ^{1}H NMR data of 46 gentamicin sulfate drug substance lots were used to classify the lots by means of principal component analysis (PCA) of 14 ^{1}H NMR-signals in the 5.0–6.0 ppm region. Three main groups could be classified: high purity (3 lots), average quality (28 lots) and low purity (14 lots); one lot proved to be atypical. The 14 normalized signal heights in the 5.0–6.0 ppm region are predictive for purity quality according to a partial least squares (PLS)-model with sum of all impurities as *Y*-variable (measured by MEKC).

Keywords: Gentamicin components and impurities; Principal component analysis (PCA); Partial least squares (PLS); NMR spectroscopy

1. Introduction

Biopharmaceuticals, especially the aminoglycoside antibiotics, makrolide antibiotics, e.g. erythromycin, and peptide antibiotics, e.g. bacitracin, are typically composed of a couple of main components accompanied by minor components and impurities of a level often higher than 0.1% being the limit allowed for a small-molecule drugs. In all cases, the composition of the antibiotics depends sensitively on the fermentation conditions and the subsequent purification applied.

Gentamicin sulfate consists of four major components, i.e. gentamicin C1, C1a, C2 and C2a, and some minor compounds such as gentamicin C2b, 2-deoxystreptamine (DSA), garamine (GA), sisomicin and netilmicin, the latter two being antibiotics on their own (see Fig. 1).

The spectra of all components of gentamicin were assigned in a previous study [1]. Having this information in hands ^{1}H NMR spectra can be used to quantify the components in the mixture and quantification results could be confirmed recently by an independent micellar electrokinetic capillary chromatography (MEKC) method [2,3].

The aim of the present study is the evaluation of the ^{1}H NMR spectra pattern by means of principal component analysis (PCA), the quality classification of gentamicin sulfate lots collected from international markets and the search for quality-marker signals in the ^{1}H NMR spectra to be used for a calibration model (i.e. correlation of selected NMR-signals with MEKC-results).

* Corresponding author.
E-mail address: Winter.w@t-online.de (W. Winter).
[1] Present address: ALTANA Pharma AG, Robert-Bosch-Str. 8, 78224 Singen, Germany.

0731-7085/$ – see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.jpba.2005.01.051

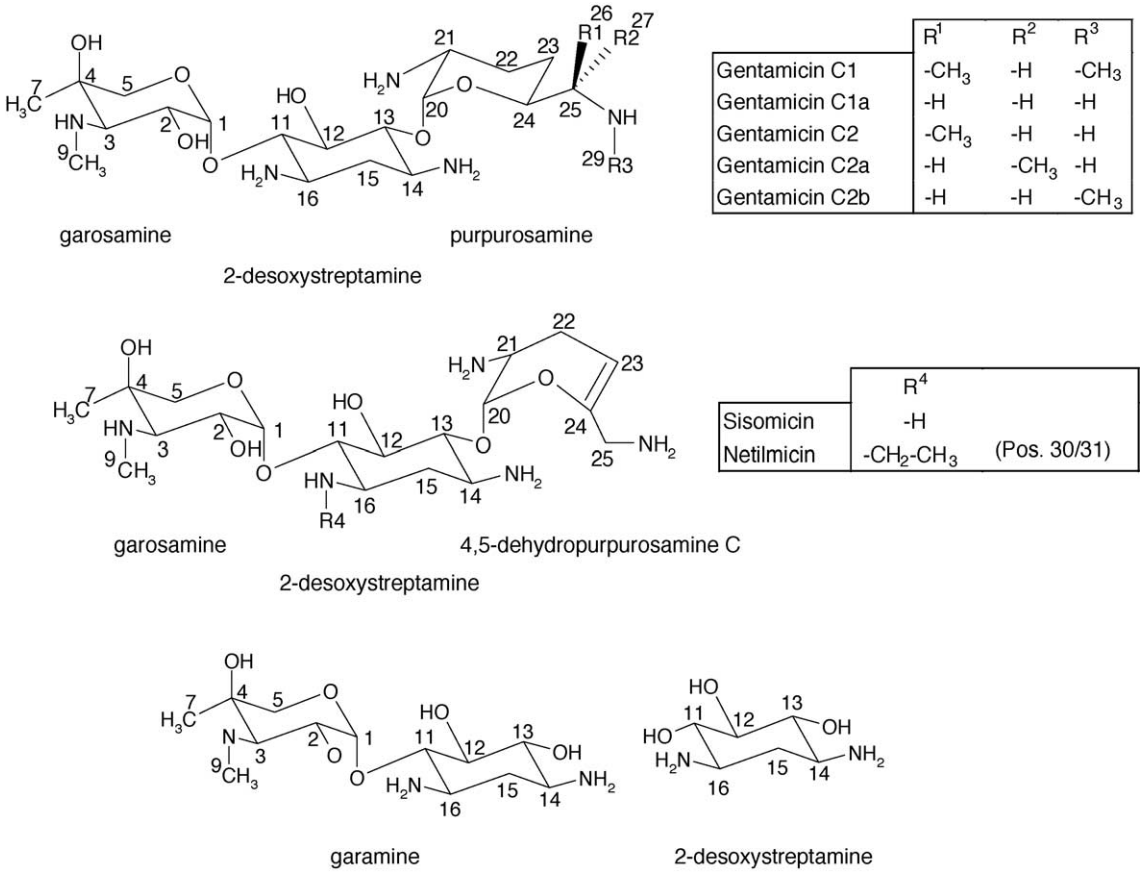

Fig. 1. Structural formulae of the gentamicin components and respective impurities.

2. Experimental

2.1. Gentamicin sulfate samples

Forty-six gentamicin sulfate samples were kindly provided by the Federal Institute of Drugs and Medical Devices (Bundesinstitut für Arzneimittel und Medizinprodukte, BfArM, Bonn, Germany).

2.2. NMR experiments and data processing

The first set of experiments was performed on a Bruker Avance 400 MHz NMR spectrometer, operating at 400.13 MHz (^{1}H), equipped with XWIN-NMR software (Version 3.0, Bruker Analytik GmbH, Rheinstetten, Germany) running on Microsoft Windows PC. For ^{1}H NMR spectra, about 15 mg of the gentamicin as mixture of components was dissolved in 650 μl deuteriumoxide. The spectral width was 8278 Hz, the transmitter offset at 6.15 ppm, the flip angle was 30°. Using an acquisition time of 3.96 s and an additional delay of 1 s a pulse repetition period of about 4.96 s results. Prior to Fourier transformation no zero filling was performed but an appropriate window function (exponential multiplication with a line broadening factor of 0.30 Hz) and manual phasing was applied.

2.3. Multivariate analysis

Calculations were performed with the multivariate software The Unscrambler™ [4]. Spectral data import of 46 samples was achieved in JCAMP-DX format (16,384 data points per sample); after reduction of the data points by a factor of 4 and removing the data segment of the HDO-signal (5.00–4.25 ppm), the remaining 3581 data points per sample were normalized. The data reduction by a factor of 4 was selected in view of the preserved resolution of all NMR-signals; four consecutive data points were replaced by their mean value. As normalization procedure the mean-normalization option of The Unscrambler™ was selected; the sum of all 3581 data points is then the same for all samples. This data matrix was used for the first PCA overview.

The second data set was built up by graphical display of the 5.0–6.0 ppm region for each sample spectrum and measuring the peak heights of 14 signals (see below; *X*-data matrix: 46 rows = samples × 14 columns = signal heights). An expansion of the second data set by an additional column with %

total impurities (measured by MEKC) as *Y*-variable was used for PLS-model building.

3. Results and discussion

3.1. PCA overview of the ^{1}H NMR spectra

Overlaid line plots of all samples ^{1}H NMR spectra are shown in Fig. 2 (solvent signal region 4.25–5.00 ppm removed and normalized).

The spectra of the 46 gentamicin sulfate samples in Fig. 2 may be imagined as a swarm of 46 data points in a 3581-dimensional space, which can be reduced to a few artificial latent variables (principal components, PC). The first PCA-result with the full (mean-centered) data set of Fig. 2 resulted in two orthogonal PCs, which explained 84% of the variance in the 3581-dimensional space; the new sample coordinates in the new PC-plane are called scores (*t*) and the corresponding score plot revealed no grouping of the samples, except for the outlying sample G46 (see Fig. 3).

The loading plot of the first PC (58% explained variance) is shown in Fig. 4 and the loadings describe the influence of the 3581 NMR-absorption data on the new PC coordinate system; the loading plot of the second PC (additional 26% explained variance) looks similar to Fig. 4.

Fig. 4 demonstrates that PCA mainly detects the small signal shifts of the dominating ^{1}H NMR-signals from the methyl-group H-7, the amino-methyl-groups H-29 and H-9 and the anomeric H-1. Small shifts due to little differences in the experimental conditions like concentration and pH during the NMR-measurements are responsible for this loading plot pattern. All attempts to correct these shifts by means of reducing the resolution were not successful.

3.2. PCA of selected ^{1}H NMR-signals (5.0–6.0 ppm region)

The problem of signal shifts with very similar ^{1}H NMR spectra (like in Fig. 2) and PCA is well known. Since the dominating signals in Figs. 2 and 4 are hardly discriminating between the gentamicin main components and the impurities, no attempt was made to correct the signal shifts by recently proposed methods [5–8].

Therefore, the most interesting region of the normalized ^{1}H NMR spectra, namely 5.0–6.0 ppm with signals of the anomeric H20 of the gentamicin components and the signals for the main impurities (see Fig. 5), was used for the assignment of 14 peak heights per sample spectrum, i.e. 8 peak heights for the anomeric H20 group, 2 for the JI-20B doublet, 2 for the sisomicin as indicated by the arrows in Fig. 5 and 2 for the garamin doublet.

The rows in the data matrix were mean-normalized and the columns mean-centered. Subsequent PCA with the new 46×14 data matrix resulted in a model with three significant PCs and 95% explained variance. The corresponding three-dimensional score plot clearly showed the separation of four main "quality groups" 1–4 as indicated in Fig. 6.

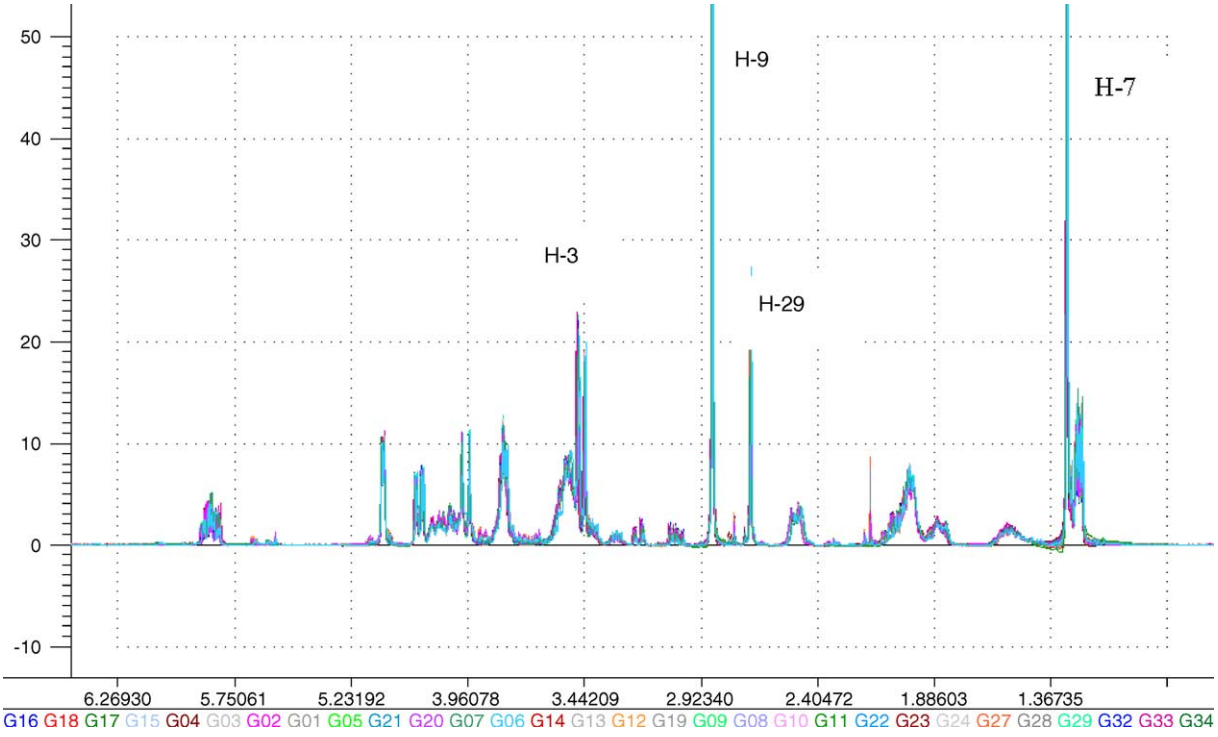

Fig. 2. Overlaid and normalized ^{1}H NMR spectra of all gentamicin sulfate samples (HOD region between 4.25 and 5.00 ppm removed); numbering scheme of the highest signals according to Fig. 1.

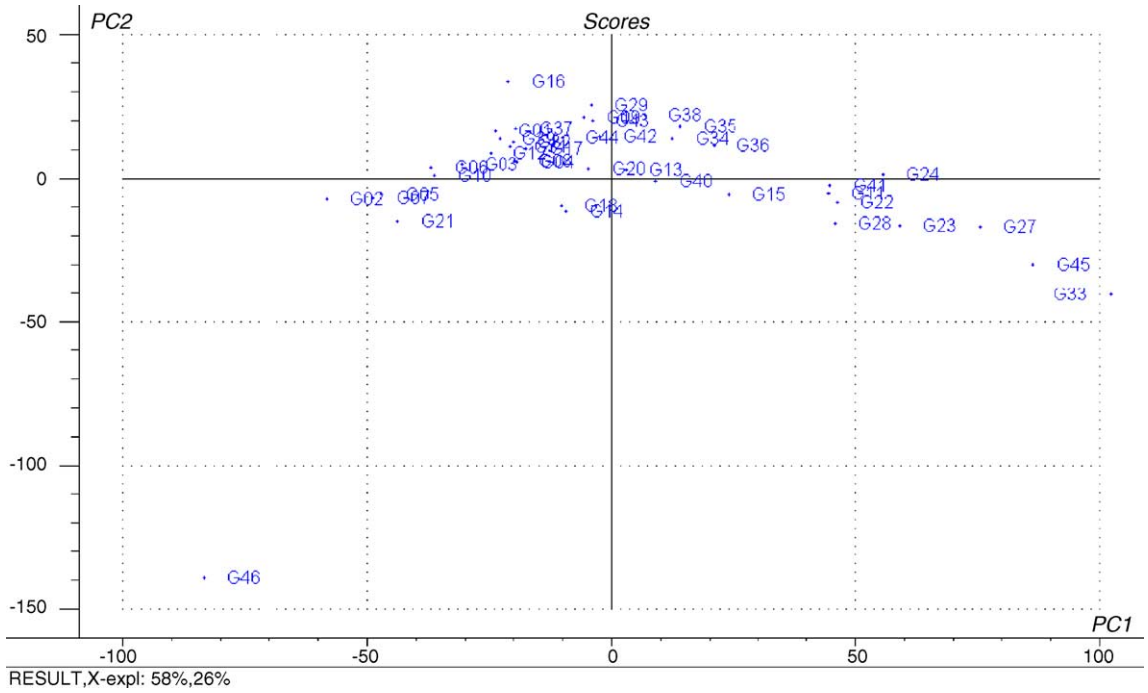

Fig. 3. *t*1/*t*2-scores scatter plot of the first PCA-model using the complete ^{1}H NMR spectra data matrix (overview of the samples space).

A two-dimensional bi-plot with 86% explained variance, on which the scores and *X*-loadings are displayed on the same plot, shows the impact of the various signal heights on the sample grouping (Fig. 7): those sample points – with score coordinates designated as G01–G46 – which are located near specific signal heights and loading coordinates designated as C2-1, C2-2 or Siso-1, Siso-2, for example, are positively correlated. Inversely, score points on opposite sides to the loading points in Fig. 7 are negatively correlated: for example, sample G20 shows low signal heights for the gentamicin component C2 (C2-1 and C2-2 in Fig. 7).

From Fig. 7, a correlation between the gentamicin component signals and the impurity signals is also obvious. In particular, a high proportion of gentamicin component C2 is

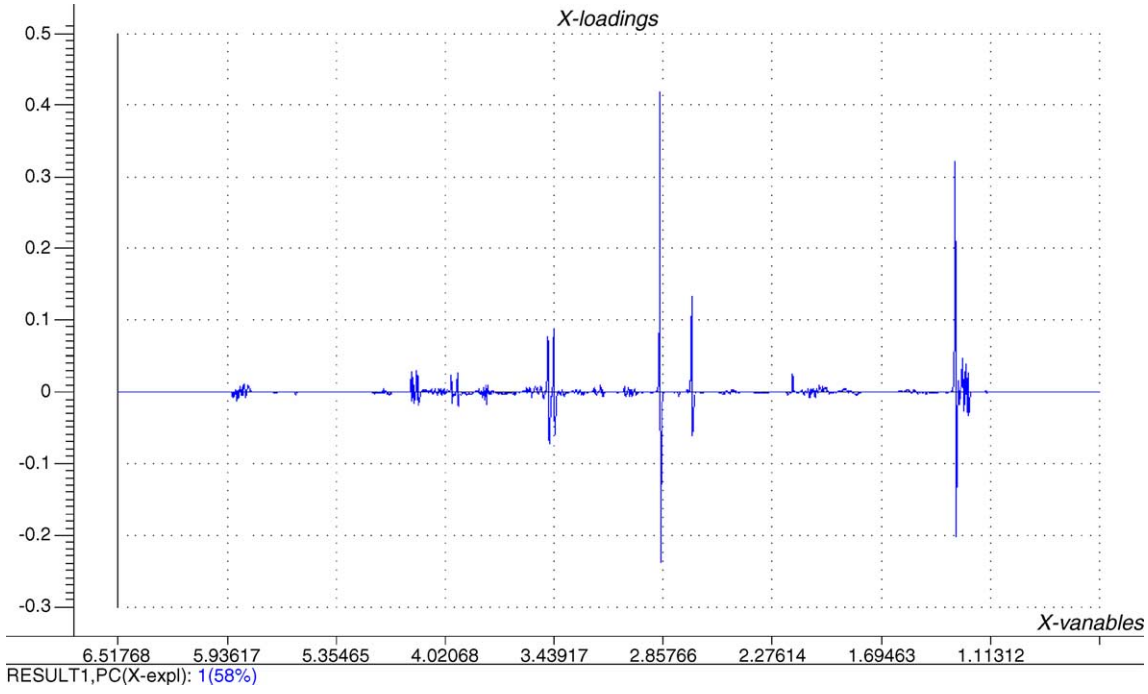

Fig. 4. Loading line plot of the first PC from the first PCA-model using the complete ^{1}H NMR spectra data matrix (overview of the ^{1}H NMR-signals importance).

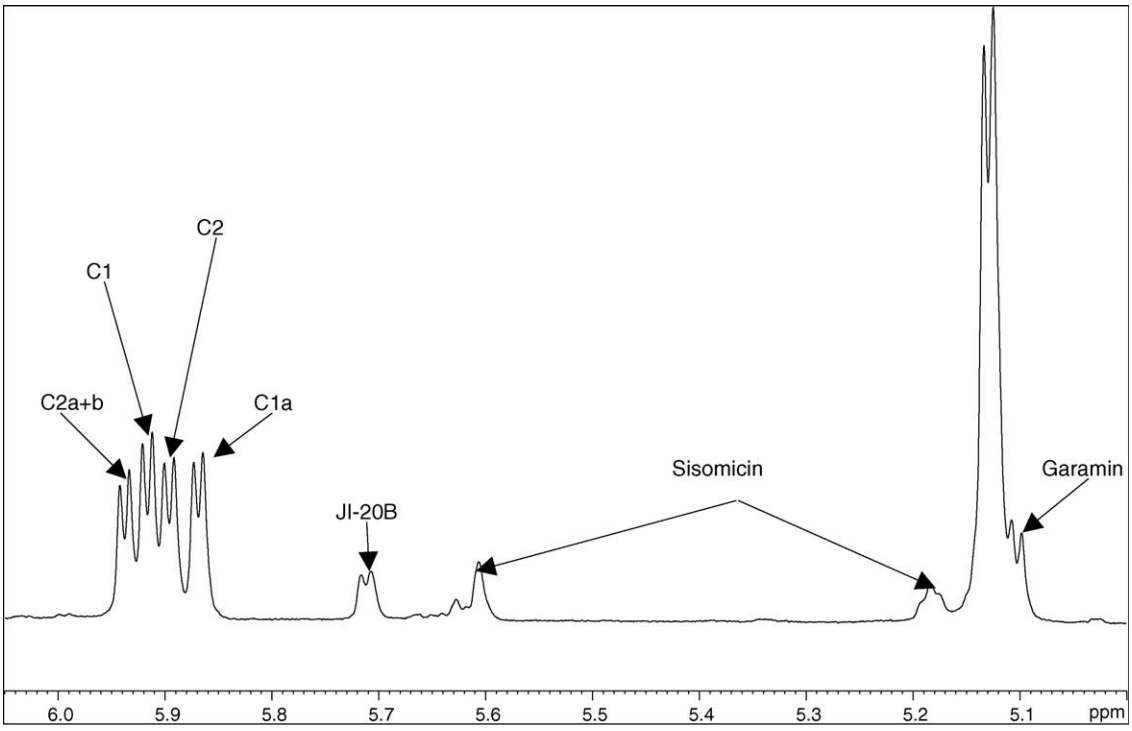

Fig. 5. ^{1}H NMR spectral zoom-out of an impure gentamicin sulfate sample: selected peak heights for second PCA-model (four doublets from anomeric H-20 of gentamicin C2a + b, C1, C2 and C1a; one doublet from JI-20B (=dihydroxy-gentamicin C2a), H-20 and H-23 from sisomicin and doublet from garamin).

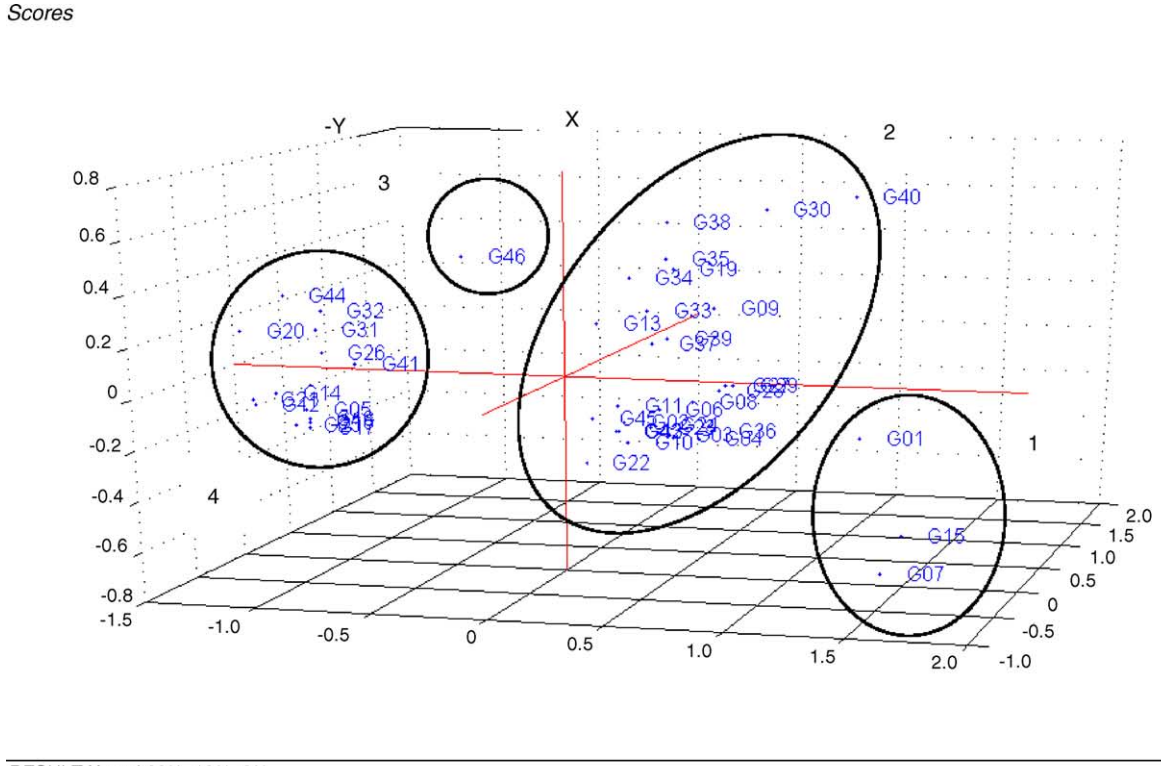

Fig. 6. Three-dimensional score plot of the second PCA-model with marked quality groups 1–4.

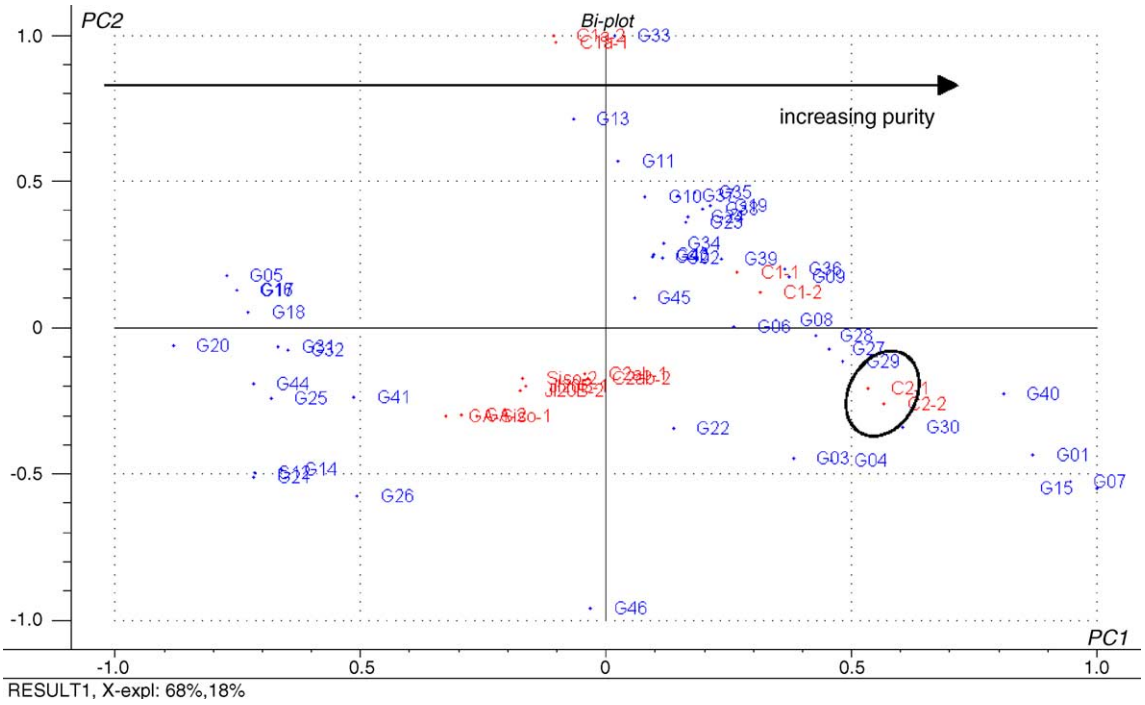

Fig. 7. Two-dimensional bi-plot of the second PCA-model showing the correlation between variables loadings (i.e. NMR-signal heights) and sample scores. The marked gentamicin C2 variables are negatively correlated with the impurity variables in the lower left quadrant.

generally indicative for low amounts of impurities; see loadings C2-1/C2-2 in the right lower quadrant of Fig. 7 and loadings for impurity signals GA, Siso and JI-20B in the left lower quadrant. The atypical sample G46 is characterized by a low gentamicin C1a content, because the scores of G46 and the loadings of C1a-1/C1a-2 are located on opposite sides of the PC2-axis in Fig. 7. The 46 gentamicin sulfate samples may be classified in four main groups (see Fig. 6):

- high purity and high gentamicin C2 group 1 (3 samples);
- average purity group 2 (28 samples);

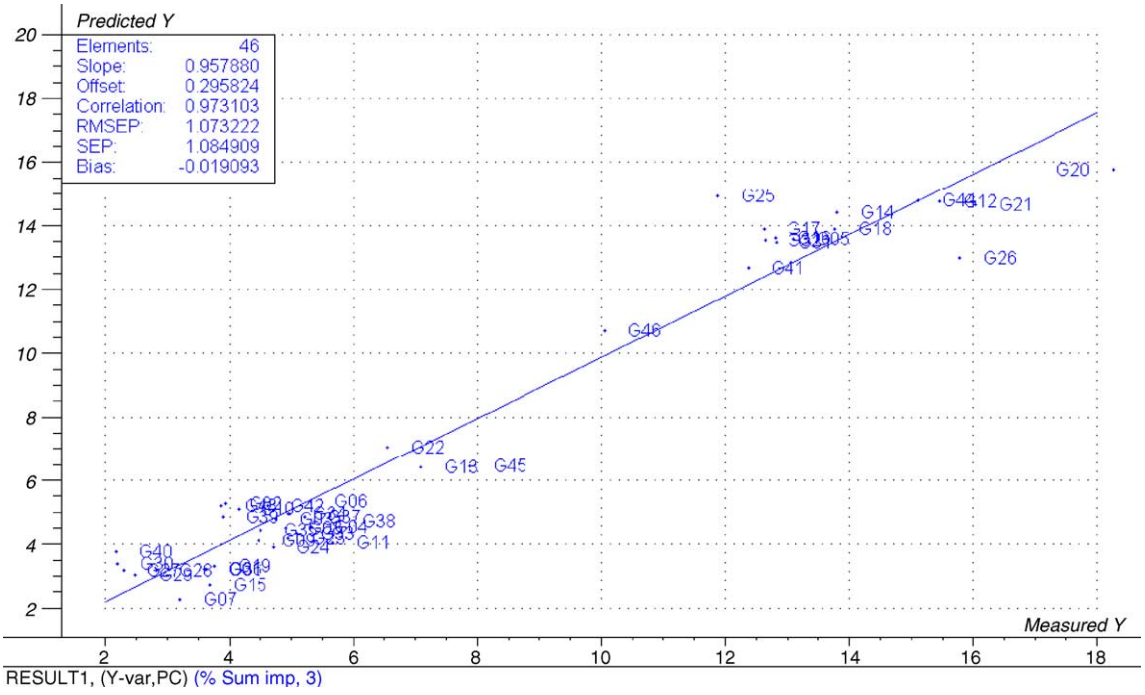

Fig. 8. Measured total impurities by MEKC (in %, *x*-axis) vs. predicted total impurities (in %, *y*-axis).

- atypical group 3 (1 sample with atypical low content of gentamicin C1a);
- low purity group 4 (14 samples).

3.3. PLS calibration model

In order to test the hypothesis that selected signal heights in the 5.0–6.0 ppm region of the ^{1}H NMR spectrum might be predictive for the purity of the samples, a multivariate partial least squares regression model (PLS-model) was established with 14 mean-normalized signal heights of Fig. 5 as *X*-variables and the sum of all impurities as *Y*-variable (mean-centered), measured by the recently published MEKC method [2,3]. PLS models the structure of the *X*-variables and the relationship to the *Y*-variables.

The satisfactory prediction property of the PLS-model is shown in Fig. 8.

The excellent model with three PLS-components is characterized by 89% explained and 81% predictive *X*-variance by leave-one-out cross-validation as well as 96% explained and 95% predictive *Y*-variance by cross-validation. The model allows the prediction of the percent total impurities in gentamicin sulfate samples with a root mean square error of prediction (RMSEP) of 1.07% (see statistics in Fig. 8). Calibration models with more than three PLS-components did not improve the RMSEPs.

From Fig. 8, it is also evident that a large portion of the 46 gentamicin sulfate samples is out-of-specification in terms of the current Ph. Eur. limit of 10.0% for total related substances [9] and that a few normalized ^{1}H NMR-signal heights may be used as rapid classification [10] and purity-estimate of gentamicin sulfate samples.

4. Conclusion

The information content in complete ^{1}H NMR spectra of gentamicin sulfate proved to be unsuitable for PCA due to signal shifts of un-predictive signals. However, the low-field signals of the anomeric H-20 atoms and some characteristic impurity signals in the 5.0–6.0 ppm region allow the satisfactory classification with PCA in terms of gentamicin sulfate main components and impurities.

The selected signal heights from the ^{1}H NMR spectra can be PLS calibrated against the independently measured percent total impurities by MEKC and may be used for a rapid quality classification of new samples.

Acknowledgements

Thanks are due to the Federal Institute of Drugs and Medical Devices, Bonn, Germany, and the Fonds der Chemischen Industrie for financial support.

References

[1] R. Deubner, C. Schollmayer, F. Wienen, U. Holzgrabe, Magn. Res. Chem. 41 (2003) 589–598.
[2] R. Deubner, U. Holzgrabe, J. Pharm. Biomed. Anal. 35 (2004) 459–467.
[3] F. Wienen, U. Holzgrabe, Elektrophoresis 24 (2003) 2948–2957.
[4] The Unscrambler™ Version 8.0, CAMO Process AS, N-0158 Oslo, Norway.
[5] J.T.W.E. Vogels, A.C. Tas, J. Venekamp, J. van der Greef, J. Chemometrics 10 (1996) 425–438.
[6] R. Stoyanova, T.R. Brown, NMR Biomed. 14 (2001) 271–277.
[7] M. Defernez, I.J. Colquhoun, Phytochemistry 62 (2003) 1009–1017.
[8] R.J.O. Torgrip, M. Aberg, B. Karlberg, S.P. Jacobsson, J. Chemometrics 17 (2003) 573–582.
[9] European Pharmacopoeia, fifth edition, European Directorate for Quality of Medicines, Strausbourg, France, 2005.
[10] F. Wienen, R. Deubner, U. Holzgrabe, Pharmeuropa 15 (2003) 273–279.