Journal of Pharmaceutical and Biomedical Analysis 153 (2018) 168–174

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

journal homepage: www.elsevier.com/locate/jpba

# Improving reliability of chemometric models for authentication of species origin of heparin by switching from 1D to 2D NMR experiments

Yulia B. Monakhova[a,b,c,*], Jawed Fareed[d], Yiming Yao[e], Bernd W.K. Diehl[a]

[a] Spectral Service AG, Emil-Hoffmann-Straße 33, 50996 Cologne, Germany
[b] Institute of Chemistry, Saratov State University,Astrakhanskaya Street 83, 410012 Saratov, Russia
[c] Institute of Chemistry, Saint Petersburg State University, 13 B Universitetskaya Emb., St. Petersburg 199034, Russia
[d] Hemostasis and Thrombosis, Department of Pathology, Loyola University Medical Center, Maywood, IL, United States
[e] Ronnsi Pharma Co., LTD, Suite 328, Bldg. A2, bioBay 218 Xinghu, Rd/SIP, JiangSu 215123, China

ARTICLE INFO

*Article history:*
Received 12 January 2018
Received in revised form 20 February 2018
Accepted 20 February 2018

*Keywords:*
NMR spectroscopy
HMQC
DOSY
Heparin
Animal origin
Classification

ABSTRACT

Nuclear magnetic resonance (NMR) spectroscopy is regarded as one of the most powerful and versatile analytical approaches to assure the quality of heparin preparations. In particular, it was recently demonstrated that by using $^{1}$H NMR coupled with chemometrics heparin and low molecular weight heparin (LMWH) samples derived from three major animal species (porcine, ovine and bovine) can be differentiated [Y.B. Monakhova et al. J. Pharm. Anal. 149 (2018) 114–119].

In this study, significant improvement of existing chemometric models was achieved by switching to 2D NMR experiments (heteronuclear multiple-quantum correlation (HMQC) and diffusion-ordered spectroscopy (DOSY)). Two representative data sets (sixty-nine heparin and twenty-two LMWH) belonged to different batches and distributed by different commercial companies were investigated. A trend for animal species differentiation was observed in the principal component analysis (PCA) score plot built based on the DOSY data. A superior model was constructed using HMQC experiments, where individual heparin (LMWH) clusters as well as their blends were clearly differentiated. The predictive power of different classification methods as well as unsupervised techniques (independent components analysis, ICA) clearly proved applicability of the model for routine heparin and LMWH analysis.

The switch from 1D to 2D NMR techniques provides a wealth of additional information, which is beneficial for multivariate modeling of NMR spectroscopic data for heparin preparations.



## 1. Introduction

Heparin is a heterogeneous polymer, which belongs to the glycosaminoglycan (GAG) family [1]. Along with low molecular weight heparins (LMWHs), which are produced by depolymerization of heparin starting material, it is a widely used anticoagulant in medical and surgical indications [1].

Among instrumental methods, $^{1}$H NMR spectroscopy is recognized as one of the most powerful tools used for the structural elucidation of heparin as well as for quantitative assessment of contaminant levels and qualitative features such as animal origin or producer [2,3]. However, for large carbohydrate polymer mixtures like heparin, spectral overlap in 1D NMR impedes excellent data evaluation even using chemometric techniques.

This problem could be overcome to some extent by using multidimensional NMR experiments. Indeed, 2D NMR spectra, which have two frequency dimensions, under certain circumstances encode a larger amount of information then 1D data. Previously, a number of 2D NMR methods have been used to characterize the composition and structure of heparin active pharmaceutical ingredients (APIs) [4–6]. It should be mentioned that Heteronuclear Single Quantum Coherence (HSQC) and a similar approach Heteronuclear Multiple-Quantum Correlation (HMQC), which both reflect direct $^{13}$C–$^{1}$H coupling, are the most frequently applied 2D NMR experiments for heparin surveillance [7–10]. In the case of contaminated heparin found on the market in 2007 and 2008, HSQC was crucial for the identification of the contaminants [7]. This type of spectroscopy has been also utilized to determine variously substituted monosaccharide components obtained by

* Corresponding author at: Spectral Service AG, Emil-Hoffmann-Straße 33, 50996 Köln, Germany.
*E-mail address:* monakhova@spectralservice.de (Y.B. Monakhova).

https://doi.org/10.1016/j.jpba.2018.02.041
0731-7085/

chemical modifications of the *Escherichia coli* polysaccharide, which is structurally related to heparin [8]. Moreover, HSQC NMR revealed differences in the composition of four LMWH biosimilar brands produced by different manufacturers [9].

Quantitative HSQC was also applied to the study of heparin and LMWH [8,10]. For example, several commercially available LMWHs (enoxaparin, dalteparin, and tinzaparin) were investigated by HSQC focusing on the quantification of the reducing and non-reducing residuals [9]. To further develop and validate this approach, the HSQC technique was applied to a larger set of samples obtained from multiple sources [10]. The data showed that the HSQC assay was robust to small variations in signal to noise and relaxation effects [10].

Another type of 2D NMR spectroscopy, diffusion-ordered spectroscopy (DOSY), was also applied for the screening of heparin samples [11–13]. DOSY combines information, which is specific to $^1$H NMR chemical shifts and is related to molecular weight distribution for a given species [11,12]. Therefore, this type of 2D NMR enabled the investigation of unfractionated and depolymerized heparins for the presence of contamination (e.g., oversulfated chondroitin sulfate) and process related impurities (e.g., dermatan sulfate) as well as their polydisperse degradation products [12]. It was demonstrated that DOSY NMR could be used to follow the course of the enzymatic reaction non-invasively and at the desired time point [11]. Recently, the DOSY technique was employed for the calculation of average molecular weight of heparin products as a complementary measurement to standard 1D NMR quality control [13].

Thus, the use of 2D NMR experiments permits a detailed structural and quantitative analysis of the monosaccharide features of heparins and LMWHs with an acceptable error [9–13]. Therefore, the switch from 1D to 2D NMR could also provide a wealth of additional information for multivariate modeling. Despite the fact that several attempts were made to employ chemometric approaches to $^1$H NMR data of heparin [14,15], none of the existing applications of 2D NMR techniques utilized multivariate analysis to provide quality control of heparin regarding its animal origin. Therefore, in this study HMQC and DOSY NMR methods combined with chemometrics were evaluated for their ability to discriminate the animal origin of heparin samples. The method was validated on a representative dataset of authentic heparin and LMWH samples from various origin (porcine, bovine, ovine).

## 2. Materials and methods

### 2.1. Samples and chemicals

A total of fifty-seven heparin (30 bovine, 9 ovine, and 18 porcine) and twenty-two LMWH (10 ovine and 12 porcine) samples were investigated. All investigated samples were commercial active pharmaceutical ingredients (APIs) distributed by different commercial companies (all samples belonged to different batches), which encompasses at least eight various producers. This is the prerequisite to prove the heterogeneity of our reference library regarding samples structure (e.g., sulfation and acetylation profiles as well as the linkage region content). Deuterated water of 99.8% purity containing 0.1% trimethylsilyl propanoic acid (TSP) as internal standard was purchased from Euriso-top (Saarbrücken, Germany).

### 2.2. Sample preparation

For sample preparation, 100 mg of a heparin (LMWH) sample was mixed with 0.7 mL $D_2O$. A series of eighteen blends with a total weight of 100 mg was prepared from selected heparin and LMWH samples of different origin. The blends were prepared from the heparin (LMWH) batches not included in the reference database. The amount of ovine or bovine species in the mixtures with porcine material was 5%, 10%, 15%, 20%, 50% and 75% (w/w).

### 2.3. NMR measurements and processing

NMR measurements were performed on Bruker Avance III 600 MHz spectrometer (Bruker Biospin, Rheinstetten, Germany) with BBO cryo probe equipped with Bruker Automatic Sample Changer (B-ACS 120) at 297 K $^1$H NMR spectra were recorded with standard pulse program (zg30 in Bruker language) using 16 scans (NS) and 2 prior dummy scans (DS). The data of 132k points (TD) were acquired with a spectral width (SW) of 24.0155 ppm, a receiver gain (RG) of 72, an acquisition time (AQ) of 4.5438 s.

2D HMQC spectra were recorded with the following parameters: NS = 64, DS = 16, TD = 2048, RG = 2050, AQ = 0.3047 (F2) and 0.003852 (F1), SW = 5.60 ppm (F2) and 110 ppm (F1). The transmitter offset was set to 4.5 for $^1$H and 65 ppm for $^{13}$C.

Several regions (rectangles) in HMQC data were tested to choose the optimal one for the multivariate statistics for heparin data set: 5.8–1.8 & 110–15 ppm; 5.8–4.9 & 110–90 ppm; 5.8–4.3 & 110–90 ppm; and 4.8–3.0 & 85–50 ppm. Bucketing with 0.01 ppm ($^1$H NMR axis) and 0.1 ppm ($^{13}$C NMR axis) width as well as unfolding were successively applied to the data in order to align the spectral data. To normalize the intensities in different samples, buckets were scaled to total intensity.

2D DOSY (diffusion ordered spectroscopy) experiments were performed using standard DOSY pulse sequence with longitudinal eddy current delay (LED) with bipolar gradient pulse pair and 2 spoil gradients. The length of the gradient pulse (δ) was set to 1400 μs and diffusion time (Δ) was set to 0.05 s. 2 scans provided enough sensitivity for heparin measurements.

For the processing of DOSY spectra the following diffusion fit function was used:

$f(x) = I_0 + e^{\left(-y^2 g^2 \delta^2 \left(\Delta - \frac{\delta}{3}\right) D\right)}$, where $D$ is the diffusion coefficient, $g$ is the gradient strength and $y$ is the gyromagnetic ratio. $I_0$ and $I$ represent the maximum and observed signal intensity. The 2D plots show diffusion coefficient values $D$ in [m$^2$/s].

The DOSY spectra were normalized to TSP signal at −9.3 m$^2$/s and 0.0 ppm. The free induction decay to an exponential multiplication associated with line broadening was set to 1 Hz in the $^1$H direction. The data points from spectral regions (5.10–5.70 ppm, and 1.80–4.50 ppm for heparin) were then pre-processed by bucketing with 0.01 ppm width. Each resultant matrix was unfolded to an array (1 × 1500), so that traditional chemometric methods could be applied.

The data were recorded automatically under the control of ICON-NMR (Bruker Biospin, Rheinstetten, Germany). All NMR spectra were manually phased and baseline-corrected using Topspin 3.2 (Bruker Biospin, Rheinstetten, Germany). Dynamic Center v. 2.2 (Bruker Biospin, Rheinstetten, Germany) was used for the treatment of DOSY raw data.

### 2.4. Chemometric modelling and validation

Matlab 2015a (The Math Works, Natick, MA, USA) and SAISIR package for MATLAB was used for statistical calculations [16].

Principal component analysis (PCA), factor discriminant analysis (FDA), partial least squares – discriminant analysis (PLS-DA), and linear discriminant analysis (LDA) were utilized as chemometric approaches.

Validation of the discriminant analysis (DA) models was performed using leave-out-one cross validation (LOOCV) as well as independent test set. The calibration set for the heparin model con-