Description of the The Human Serum and Urine databases

From Féraud et al. *(2015) and Martin* et al. *(2018)*

This article presents the Human Serum and Human Urine databases originally published in Féraud *et al.* (2015) and Martin *et al.* (2018). These databases were designed with spectroscopists from the University of Liège and the University of Nantes (France).

Motivations for creating this database

Samples from two well-known human biofluid (serum and urine) are explicitly designed to evaluate the preprocessing methodology performances in Chapter 2. These datasets were both designed to collect multiple measures from the same experimental unit, the donor, and capture other technical sources of variation, allowing the comparison of inter- and intra-units variability.

Design, sample preparation and data acquisition

Human Serum

Design

In the Human Serum dataset, a blood sample was collected from 4 different donors. For each sample, 8 sub-samples were measured across 8 days with one sub-sample from each donor per day and permutations according to a latin hypercube sampling method (Iman, 2014) in order to avoid confusion between donors and times of analysis. Data have been acquired at different moments of the day, thus creating different delays before spectral measurement. The total number of available 1D FID signals is then $4 \times 8 = 32$.

Data preparation

This experimental design is based on human serum. Peripheral blood was collected in serum separating tubes (Greiner). Sera were distributed into 0.5 ml aliquots and stored at -80°C just after sampling. Four blood donors were engaged for the study. For each collected sample, 500 μl of serum, defrosted just before the analysis, were supplemented with 200 μl of deuterated phosphate buffer (0.1 M, pH = 7.4) containing 1 % of sodium azide and 30 μl of TMSP (10 mg/ml). The solutions were then transferred into 5 mm NMR tubes before NMR measurement.

Data acquisition

1D spectra were recorded at 298 K on a Bruker Avance spectrometer operating at 500.13 MHz for the proton signal acquisition. The instrument was equipped with 3 channels and with a 5mm triple resonance (HCN) cryoprobe with a Z-gradient and automatic tuning and matching system (ATMA). All samples were locked using deuterated water, shims were tuned using auto-tune (Topshims) and samples are measured manually without spinning. Due to the nature of the samples, a presaturation sequence was used in all the experiments in order to minimise the water signal. All data were referenced to internal sodium 3-trimethylsilyl-2,2,3,3-d4-propionate (TMSP) at 0.00 ppm chemical shift (all spectra are calibrated with regard to TMSP).

The ^{1}H NMR spectra were acquired using a CPMG relaxation-editing sequence with presaturation (human sera). Upon the presence of proteins in serum, the use of a sequence with a T2 filter (CPMG) greatly

improves the baseline. The CPMG experiment (pulse sequence cpmgpr1d supplied by Bruker) used a RD-90°-(t-180° - t_n)-sequence with a relaxation delay (RD) of 2s, a spin echo delay (t) of 400 μs and the number of loops (n) equal to 80. The water suppression pulse was placed during the relaxation delay (RD) and a spectral window of 10245 Hz. The number of transients was typically 32. The acquisition time was fixed to 3.982555 s and a quantity of four dummy scans was chosen. FIDs were collected in 64 K time data points.

Human Urine

This dataset is resulting from an inter-laboratories collaboration involving Patrick Giraudeau and Estelle Martineau from the Université de Nantes (France), and Pascal de Tullio and Justine Leenders from ULiège.

Design

In this second dataset, morning urine from 3 different donors was collected. Each donor sample was divided into 2 sub-samples: one was kept pure and the other 25 % diluted. The 4 aliquotes of each dilution were analysed on 4 different days. For each day, the order of measurement was held constant across the 6 sub-samples. The total number of collected FID signals is then: $3 \times 2 \times 4 = 24$.

Data preparation

In order to conduct this experiment and to design the collection of urine samples, the morning urine of three different fasting donors was collected. For each donor, four aliquots of 400 μl and four aliquots of 320 μl were placed at -80°C. Then, on each consecutive day (four days), six aliquots were thawed (3 donors x 2 quantities) and routinely prepared as follows. Urine samples of 400 μl were supplemented with 300 μl of deuterated phosphate buffer (DPB, pH 7.4), while 320 μl urine aliquots were supplemented with 380 μl of the same buffer. 10 μl of a 10 mg/ml TMSP solution was then added to all aliquots. The four aliquots of each dilution were put in 5mm NMR tube for NMR acquisition and analysed. For each day, the order of measurement was held constant across the six sub-samples. A total of 24 1D collected signals are finally available.

Data acquisition

All the samples were acquired with a Bruker Avance III NMR spectrometer at 700.28 MHz at Ceisam, Nantes.

The final databases

The final databases are composed of ^{1}H NMR raw Bruker FID signals (in the *fids* folder) and a design .RData file.

The Human Serum spectra are labelled as: $JxDy$ where x is the day of measurement and y is the donor label. The design file contains the *HS_design* data.frame with variables *ID*, *donor* and *day*.

The Human Urine, spectra are labelled as: $SxDy$ where x is the donor label and y is the dilution (0: no dilution; 1: 25 % diluted). The design file contains the *HU_design* data.frame with variables *ID*, *donor* and *dilution*.

Manual pre-processing

Both datasets were manually pre-processed with a combination of Topspin 3.1 and AMIX 3.9.14. at the resolution of 500 buckets. The manual pre-processing included the following sequence: exponential apodization, Fourier transform, first and zero order phase corrections, internal calibration, polynomial baseline correction, bucketing and spectral region removal.

The manually pre-processed spectra are available under the form of a .RData file for Human Serum *HumanSerumManualPretreat* and as a .csv file *Nantes_Manual_B500_Spectra_data* for Human Urine.

References

Féraud, B., Govaerts, B., Verleysen, M., & de Tullio, P. 2015. Statistical treatment of 2D NMR COSY spectra in metabolomics: data preparation, clustering-based evaluation of the Metabolomic Informative Content and comparison with 1H-NMR. *Metabolomics*, **11**(6), 1756–1768.

Iman, R. L. 2014. Latin Hypercube Sampling. *Wiley StatsRef: Statistics Reference Online*.

Martin, M., Legat, B., Leenders, J., Vanwinsberghe, J., Rousseau, R., Boulanger, B., Eilers, P. H.C., Tullio, P. De, & Govaerts, B. 2018. PepsNMR for ^{1}H NMR metabolomic data pre-processing. *Analytica Chimica Acta*, **1019**, 1 – 13.