Ultra-filtration of human serum for improved quantitative analysis of low molecular weight biomarkers using ATR-IR spectroscopy

Franck Bonnier[*,1], Hélène Blasco[2,3], Clément Wasselet[1], Guillaume Brachet[4], Renaud Respaud[5,], Luis Felipe CS. Carvalho[6], Dominique Bertrand[7], Matthew J. Baker[8], Hugh J. Byrne[9], Igor Chourpa[1]

[1] Université François-Rabelais de Tours, Faculté de Pharmacie, EA 6295 Nanomédicaments et Nanosondes, 31 avenue Monge, 37200 Tours, France.
[2] CHRU de Tours, Laboratoire de Biochimie et de Biologie Moléculaire, Tours, France.
[3] INSERM, UMR U930 "Imagerie et Cerveau", Université François Rabelais, Tours, France.
[4] Université François Rabelais de Tours, UMR CNRS 7292 Génétique, Immunothérapie, Chimie et Cancer, Faculté de Médecine, 10 Bd Tonnellé, 37032 Tours Cedex
[5] Université François-Rabelais de Tours, UMR 1100, CHRU de Tours, Service de Pharmacie, F-37032 Tours, France
[6] Universidade do Vale do Paraíba, Laboratory of Biomedical Vibrational Spectroscopy, Sao José dos Campos, Brazil
[7] Data Frame, Nantes, France
[8] WestCHEM, Technology and Innovation Centre, Department of Pure and Applied Chemistry, University of Strathclyde, 295 Cathedral Street, Glasgow G1 1XL, UK
[9] FOCAS Research Institute, Dublin Institute of Technology (DIT), Camden Row, Dublin 8, Ireland

Supplementary information

S.1 Data pre-processing: Baseline correction and vector normalisation

Figure S.1 illustrates examples of ATR-IR spectra baseline correction and vector normalisation. The data collected from mixed pool human serum spiked with increasing concentrations of glucose exhibit similar backgrounds. The ATR minimises strong baseline distortion related to Mie scattering which leads to simplified data pre-processing requirements (Figure S.1 A). For instance, although the baseline correction is based on a rubber band algorithm, only 2 nodes at 1800 and 900 cm^{-1} are required in order to compensate for a slight offset that can be observed in the data (Figure S.1 B). Finally, a standard vector normalisation is applied to scale up of the spectra for direct comparison and analysis by means of PLSR.
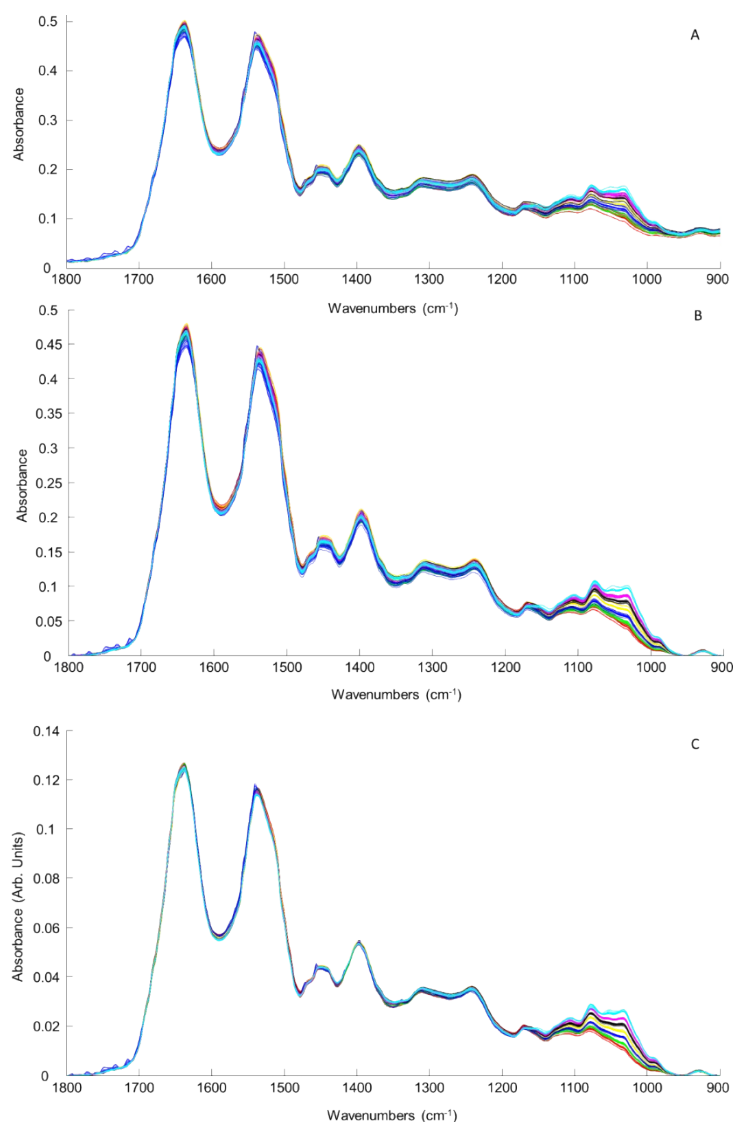
Figure S.1: Example of data processing. **A:** Raw infrared spectra collected from human serum spiked with different glucose concentrations. **B:** Same data set after baseline correction using the rubber band algorithm. **C:** Same data after vector normalisation. Spectra are colour coded as follows; *Red:* Human serum stock solution, *Green:* serum spiked with 20 mg.dL^{-1}, *Blue:* serum spiked with 60 mg.dL^{-1}, *Yellow:* serum spiked with 100 mg.dL^{-1}, *Black:* serum spiked with 140 mg.dL^{-1}, *Magenta:* serum spiked with 180 mg.dL^{-1} and *Cyan:* serum spiked with 220 mg.dL^{-1}.

S.2 Pure Glucose

S.2.1. Minimum and maximum concentrations

In such studies, it is also interesting to address the question of the limit of detection for a given molecule, for instance glucose. While the PLSR model delivers information related to quantitative precision, working with air dried samples can lead to questions of the maximum and minimum of glucose concentrations which can be analysed. As documented previously, the behaviour of the spectral changes is not similar for low and high concentrations and thus the limitations need to be considered separately[24]. Figures S.2 A and B display correlations of the amount of glucose deposited on the crystal and the measured absorbance in the Area Under the Curve (AUC) for the glucose bands in the region 1180-955 cm^{-1}. The Table in Figure S.3 gives an overview of the maximum and minimum concentrations of glucose measurable with ATR-IR, depending on the drop size use for the recording. It is notable that, with a limit of detection at 0.015 μg, even the use of 2 μL drops is limited to a minimum concentration of 0.75 mg/dL.
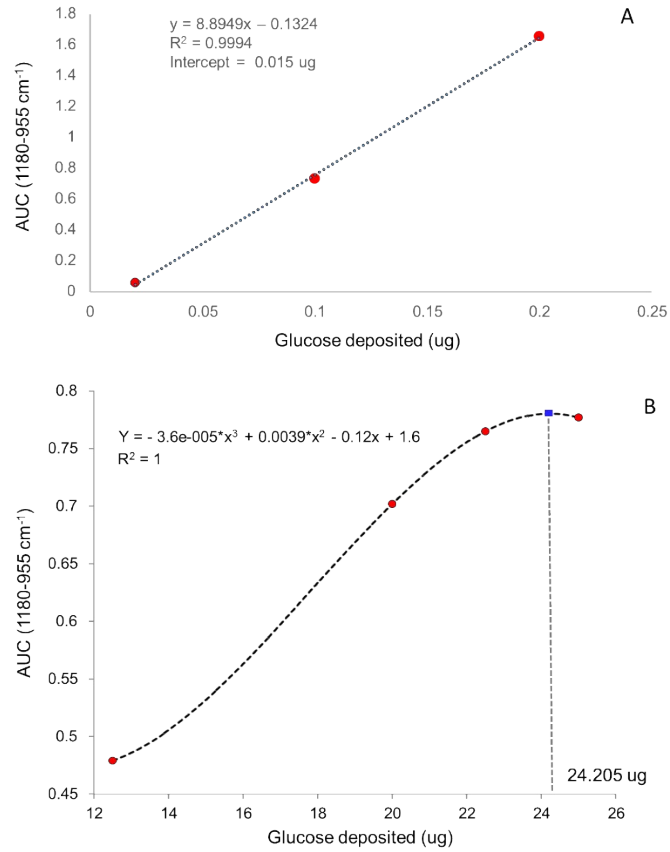
Figure S.2: Estimation of the limit of detection for ATR-IR analysis of glucose solutions. **A:** linear fitting for amount of glucose below 0.2 µg. **B:** Polynomial fitting for amount of glucose deposited above 12.5 µg.

V (µL)	C_{min} (mg/dL)	C_{max} (mg/dL)
0,1	15	24 205
0,2	7,5	12 102
0,5	3	4 841
1	1,5	2 420
2	0,75	1 210

Figure S.3: Summary minimum and maximum concentrations hat can be analysed depending on the drop size deposited on the ATR crystal. Concentrations expressed in mg/dL for comparison with clinical results.

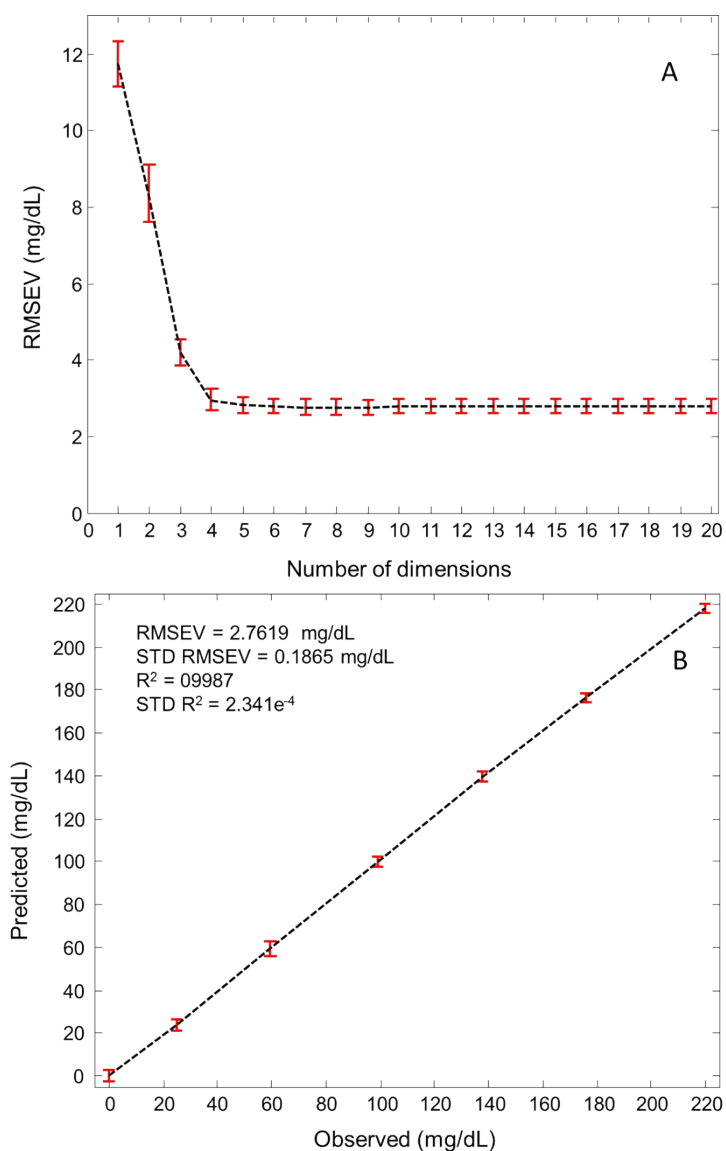## S.2.2 PLSR of aqueous glucose solutions

Figure S.4: PLSR model constructed from IR spectra collected from aqueous glucose solutions. Similar to patient samples, 2 µL drops air dried before recording have been analysed. **A:** Evolution of the root mean square error on the validation set (RMSEV) according to the number of dimensions selected in the PLSR model. Values are averages calculated from the 20 iterations of the cross validation associated with the corresponding error bar illustrating the standard deviation. B: Predictive model built from the PLSR analysis. For each concentration, the value displayed is an average of the concentration predicted with the corresponding standard deviation calculated from the 20 iterations of the cross validation. Mean RMSEV and R^2 values are given on the plot both also with their respective standard deviation.