Biofuels and Environmental Biotechnology **Biotechnology and Bioengineering**
DOI 10.1002/bit.24741

Rapid determination of sugar content in biomass hydrolysates using nuclear magnetic resonance spectroscopy[†]

Erica Gjersing*, Renee M. Happs, Robert W. Sykes, Crissa Doeppke, and Mark F. Davis

National Bioenergy Center, National Renewable Energy Laboratory, 1617 Cole Blvd., Golden, CO 80401

*Address correspondence to: Erica.Gjersing@nrel.gov; phone: 303-384-7984; fax: 303-384-6363

Key Words: hydrolysate, Partial Least Squares, 1H NMR, PLS regression

Received May 21, 2012; Revision Received July 27, 2012; Accepted September 24, 2012

Abstract

Large populations of potential cellulosic biomass feedstocks are currently being screened for fuel and chemical applications. The monomeric sugar content, released through hydrolysis, is of particular importance and is currently measured with time-consuming HPLC methods. A method for sugar detection is presented here that employs ^{1}H NMR spectra regressed against primary HPLC sugar concentration data to build Partial Least Squares models. The PLS2 model is able to predict concentrations of both major sugar components, like glucose and xylose, and minor sugars, such as arabinose and mannose, in biomass hydrolysates. The model was built with 65 samples from a variety of different biomass species and covers a wide range of sugar concentrations. Model predictions were validated with a set of 15 samples which were all within error of both HPLC and NMR integration measurements. The data collection time for these NMR measurements is less than twenty minutes, offering a significant improvement to the one hour acquisition time that is required for HPLC.

Introduction

The conversion of cellulosic, non-food-source biomass into biofuels is being optimized as a replacement for fossil fuels (Wyman 2007). One of the primary steps in the biochemical conversion process involves using pretreatment and enzymatic hydrolysis to release monomeric sugars, which can then be further fermented to products, chemicals, or fuels. In most feedstock species, the most abundant form of monomeric sugar is glucose, which is formed from the depolymerization of cellulose. The other major source of sugars is hemicellulose, which can include xylose, mannose, galactose, rhamnose, and arabinose monomers. The ease by which biomass can be broken down depends on a variety of factors including lignin content and composition, lignin carbohydrate interactions, and plant anatomy (Himmel 2008). Thus understanding, improving, and quantifying biomass recalcitrance is necessary for the large scale implementation of a biomass-to-biofuels industry (Lynd et al. 2008).

The development of new biomass feedstocks with reduced recalcitrance requires detailed characterization of both plant cell wall chemistry and anatomy. High throughput (HTP) methods for characterizing carbohydrate components are necessary to screen large populations for natural variations in cell wall chemistry traits and identify changes due to transformation of specific genes believed to be able to influence biomass recalcitrance. Recently, higher throughput genetic transformation pipelines have been developed to generate large numbers of modified switchgrass and poplar feedstocks that can produce hundreds of plants for a single gene of interest. Biomass cell wall chemistry measurements of lignin and carbohydrate content are obtained through a wet chemistry process of hydrolysis followed by HPLC and gravimetric determination, which is both time and labor intensive (Sluiter et al. 2010). The traditional HPLC analysis described in ASTM and NREL standard procedures for analysis of the liquid hydrolysate can take more than an hour for data collection and processing of a single sample. Recent focus has turned to high throughput

methods, which can screen 1000s of potential biomass feedstock samples per month and require small sample volumes (Decker et al 2009; Sykes et al. 2009). Pyrolysis molecular beam mass spectometry (PyMBMS) is currently employed to determine lignin contents and lignin S/G ratios in biomass samples but does not provide quantitative information on the sugar content (Sykes et al. 2009; Sykes et al 2008). Automated compositional analysis allows for preparation of samples in 96-well plates but still requires time-consuming HPLC methods for accurate sugar determination (DeMartini et al 2011; Studer et al. 2010). Faster methods for determining sugar concentrations in the liquid hydrolysates prepared through these methods are needed to make the entire process truly high throughput.

Simple ^{1}H NMR screening methods are able to detect the sugar components in liquid hydrolysates, making this an attractive method for high throughput screening. There is a wealth of information published about mixture analysis with NMR in applications including drug discovery (Powers 2009) and quality control of wine (da Silva Neto et al 2009) and fruit juices (Spraul 2009). In biomass hydrolysates, sugar concentrations have been determined with ^{1}H NMR by relying on the anomeric proton region between 4.4 and 5.4 ppm where the signals from the individual sugars are integrated against a known reference compound that has been added to the sample (Kiemle et al 2004; Mittal et al 2009). In this anomeric region of the NMR spectra, two signals exist for each sugar, one for each of the α and β anomers, as shown in Figure 1. Although this region of a hydrolysate spectra is not particularly complicated, overlap still exists for some peaks above 5 ppm and for the galactose and xylose peaks around 4.55 ppm. Another drawback is the fact that this region overlaps with the large water peak at 4.8 ppm, which is always present in hydrolysates produced using ASTM procedures (ASTM E1758-01 2003). One way to overcome the water peak overlap is with the addition of acid, which shifts the water peak

to higher ppm. However, probe tuning and signal to noise ratio suffer due to the high acid concentrations, and longer experimental times are required. These effects are magnified further with the use of a cryoprobe, where the acidity increases the resistance in the NMR coil (Kelley et al. 2002). Another option for shifting the water peak is to use temperature, where a small shift of ~0.1ppm per 10°C is observed. In order to shift the water peak to 5.3ppm, equivalent to a 4% H_2SO_4 solution (Kiemle et al 2004), a temperature below the freezing point of water would be required, rendering this approach unmanageable. These issues have hampered widespread adoption of this ^{1}H NMR technique for high throughput analysis.

This study uses a different approach for hydrolysate sugar determination that focuses on the sugar signals in the region of 3–4 ppm. The peaks from the different sugars overlap considerably in this region, as shown in Figure 2, making simple integration techniques impossible. Quantitative analysis is further complicated by the order of magnitude differences in component concentrations, with glucose and xylose on the order of ~1 mg/mL and galactose, arabinose, and mannose on the order of ~0.1 mg/mL. Therefore, the ^{1}H NMR spectra are combined with HPLC-measured concentration information and used to construct Partial Least Squares (PLS) models of the 3–4 ppm region for determining monomeric sugar concentrations. In addition, Principal Component Analysis (PCA) can be applied as a means of understanding the variation within sample sets before PLS modeling is attempted.

Materials and Methods

Samples

Samples for collecting NMR spectra of each individual sugar were prepared by dissolving ~10mg sugar in 1mL of D_2O and allowing the solution to equilibrate overnight at

room temperature. D-glucose, D-xylose, D-galactose, L-arabinose, and D-mannose were purchased from Sigma Aldrich (St. Louis, MO, USA).

Biomass samples were prepared with sequential water and ethanol extraction processes prior to hydrolysis. The wet chemical analysis was performed on these samples using the NREL Laboratory Analytical Procedures "Preparation of Samples for Compositional Analysis" (Hames et al 2008), "Determination of Structural Carbohydrates and Lignin from Biomass" (Sluiter et al. 2008).

Hydrolysis

All biomass samples were prepared following the NREL Laboratory Analytical Procedure "Preparation of Samples for Compositional Analysis" (Hames et al 2008). A scaled-down version of the NREL Laboratory Analytical Procedure "Determination of Structural Carbohydrates and Lignin from Biomass" (Sluiter et al. 2008a) was used for preparing hydrolysates. Autoclave pressure tubes (2-5mL) were loaded with 7.5 mg ($\pm$0.5 mg) milled and extracted biomass. The samples were then mixed with 75 µL of 72% H_2SO_4 and placed in a 30 °C water bath for 1 hour. After the samples were removed from the water bath, 2.1 mL of water was added to each tube and the tubes were autoclaved at 121 °C for 1 hour. The samples were allowed to cool to room temperature and the liquid hydrolysate fraction was then decanted into 15 mL conical tubes and neutralized with $CaCO_3$ to a pH of 7. The neutralized samples were spun at 5000 rpm for 10 minutes; the liquid hydrolysate fraction was filtered at 2 microns to ensure all solids were removed from the solution. Filtered hydrolysates were stored in 1.5 mL microfuge tubes until further preparation for quantitative NMR analysis. An additional set of hydrolysate samples was prepared in the same manner as above, but with D_2O and deuterated

sulfuric acid so that integration of the anomeric proton region could be performed and compared with the results from the PLS modeling.

HPLC

HPLC was performed as specified in the NREL Laboratory Analytical Procedure "Determination of Structural Carbohydrates and Lignin from Biomass" (Sluiter et al. 2008a). An Agilent Infinity 1220 Series HPLC system with a Bio-Rad HPX-87P column was used. The injection volume was 10–50 µL for each sample, with a mobile phase of HPLC-grade water at a flow rate of 0.6 mL/min. The column temperature was 80–85 °C with a run time of 35 minutes.

Nuclear Magnetic Resonance

Hydrolsates in 100% H_2O were prepared for NMR by adding 50µL D_2O, for NMR locking, to 450µL of hydrolysate. For the samples prepared in deuterated solvents, 0.5mL of hydrolysate was placed in the NMR tube with no additional solvents added. ^{1}H NMR spectra were collected on a Bruker Avance III spectrometer at 14.1 T and 600.16 MHz and a Bruker 5 mm BBO probe. Automatic shimming was performed prior to acquisition of every sample via Bruker Topspin 3.0 TopShim. A NOESY-1D with presaturation water suppression sequence was employed for samples hydrolyzed in H_2O with a 90° pulse of 15 µs and a recycle delay of 10 s. Spectra for the individual sugar samples, shown in Figures 1 and 2, and samples prepared in D_2O were run with a Bruker zg pulse sequence (no water suppression) with a 90°=15 µs and 10 s recycle delay. All NMR spectra were acquired at 25°C and automatic phase correction, baseline correction, and chemical shift referencing to the furfural peak at 7.5ppm were applied to every spectrum.

Data Analysis

Concentrations determined by integration of anomeric peaks was performed as described previously (Kiemle et al 2004; Mittal et al 2009). Briefly, the β anomer peaks of glucose, xylose, galactose and mannose and the α anomer peak of arabinose was integrated manually in Topspin 2.1 for each hydrolysate spectra. The integrated values were normalized to a sealed, external standard of DSS of 0.5mM concentration. Ratios of α/β anomeric peaks at equilibrium were calculated from spectra of pure individual sugars, shown in Figure 1, which matched well with published values (Franks 1987). These ratios were then used to determine the total concentration of each sugar in the hydrolsate based on the known standard concentration.

Principal Component Analysis (PCA) was performed in Bruker's AMIX software, where each spectrum was divided into 0.05 ppm buckets in the region of 3.2–4.15 ppm for a total of 187 buckets. The Partial Least Squares modeling was performed in The Unscrambler v. 9.7 (CAMO A/S, Trondheim, Norway). Methanol is formed during the acid hydrolysis procedure and has an NMR signal in the region of 3.33–3.35 ppm; this peak has been excluded (set to zero) for all of the samples so that only peaks for the sugars will be included in the PCA and PLS modeling.

Results and Discussion

The main goal of this work was to provide a rapid method of determining sugar concentrations in hydrolysates. To minimize NMR experiment time, neutral pH samples were employed because signal to noise is decreased in acidic samples due to increased conductivity. The use of neutral pH samples will allow for the implementation of this experimental procedure on a cryoprobe which will further reduce the NMR experiment time to less than 5 minutes.

Previous work (Kiemle et al 2004; Mittal et al 2009) has shown that integration of the anomeric proton region is a reliable method for predicting sugar concentration. However, in neutral pH samples the ^{1}H signal from water appears at 4.8 ppm and overlaps with the anomeric proton region for sugars. Therefore, instead of integration of the anomeric proton region, which is unobservable under neutral pH conditions, we focused on the region between 3.2 and 4 ppm and utilized PLS models regressed from HPLC measured sugar concentrations to determine the concentrations of monomeric sugars.

Principle Component Analysis (PCA) is a technique that is extremely useful for visualizing variation in large, high-throughput datasets. Two types of plots are generated from the analysis: 1) the PCA scores plot groups similar samples together based on the input data and 2) PCA loadings which indicate which areas of the spectra are contributing to the variation between the groups. The PCA scores and loadings plots for the NMR spectra collected in sets of 10-20 samples over the course of six months for 80 samples that span nine potential biomass feedstock species are shown in Figure 3. The PCA scores plot (3a) indicates that the samples group into three categories of monocot angiosperms, dicot angiosperms, and gymnosperms. PC1 accounts for 56.3% of the explained variance in the data while PC2 accounts for an additional 16% of the variation. In order to determine which areas of the NMR spectra vary most between the different groupings, the loadings plots are investigated. Comparison of the PC1 loadings plot in Figure 3b with the NMR spectra for the individual sugars indicates that PC1 can be attributed to xylose. This is most evident in the triplet at 3.3ppm, which is seen in the NMR spectra of pure xylose (Figure 2), and is inverted in the PC1 loadings plot (Figure 3b). The fact that the peak is inverted in the loadings indicates that samples having negative scores associated with higher xylose concentrations. Therefore, based on their placement along PC1, the pine

samples have the lowest xylose concentrations while the monocot angiosperm samples have the highest xylose concentrations and the dicot angiosperm samples fall somewhere in the middle. The interpretation of the loadings of PC2 in Figure 3c is slightly more complicated because the positive loadings are indicative of glucose while the negative loadings correspond to both mannose and arabinose concentrations.

A PLS-2 model was constructed using 65 samples, in duplicate, along with the sugar concentration data obtained from HPLC. The remaining fifteen samples were used for predictive validation. The concentrations for all five sugars were input into a single PLS2 model, and the fully cross-validated models along with their regression coefficients are displayed in Figures 4 and 5. The PLS2 models for the major sugar components in hydrolysates, glucose and xylose, are shown in Figure 4, and the models for the minor sugars galactose, arabinose, and mannose are displayed in Figure 5. The glucose, xylose and mannose models have R^2 values of 0.821, 0.93 and 0.944, respectively, indicating good correlation between the HPLC measured and NMR PLS predicted concentrations. The galactose and arabinose models, with R^2 = 0.66 and 0.707, respectively, have lower correlations due to the lack of a wide range of concentrations for these minor sugars. In addition, the concentrations of these two sugars are close to the lower detection limit of the HPLC method, 0.05mg/mL, making any values below this value unreliable and susceptible to higher error. Future models may be improved as biomass samples with higher galactose and arabinose concentrations are identified and added to the sample pool.

The NMR spectra for each of the modeled sugars are shown above the regression coefficients in Figures 4 and 5. High value, positive regression coefficients indicate areas of the spectra that are relevant to the regression. Comparison of the regression coefficients with the NMR spectra demonstrates that the model is using areas of the spectra assigned to each sugar to

predict concentrations for that sugar. For example, in Figure 4 the regression coefficients (d) is highest in the region around 3.4ppm and 3.2ppm which correspond to a multiplet and a triplet in the glucose NMR spectra (c). A similar pattern is observed for the xylose regression coefficients in Figure 4 (f) where the areas around 3.4, 3.3, and 3.2ppm are the most relevant to the regression and correspond to regions observed in the xylose NMR spectra (e). From PLS2 modeling of the two major sugars, glucose and xylose, it would appear that the most significant regions of the spectra are between 3.2-3.5 ppm and that the rest of the region, between 3.5-4.2 ppm might not be necessary. However, in a PLS2 model all five sugars are regressed simultaneously and the region between 3.5-4.2 ppm is needed for minor sugar components, as shown in Figure 4. For the arabinose model the multiplets around 3.65 ppm and the broad peak at 3.93 ppm in the NMR spectra Figure 4 (d) are the most intense areas in the regression coefficients (e). The important areas for the mannose regression are around 3.65, 3.8-3.9, and 3.9 ppm Figure 4 (i) and correspond well to the mannose NMR spectra (h). The regression coefficients for galactose in Figure 4 (g) indicate that the most important region for predicting galactose concentrations is around 3.47 ppm, which corresponds to a multiplet in the NMR spectra (f).

In order to demonstrate the validity of the PLS2 model, concentrations of all five sugars in 15 biomass samples were predicted and the results are compared with HPLC measured values in Figure 6. Samples were also prepared in D_2O so that the previous method developed for measuring sugar concentrations in biomass hydrolysates based on integrations of the anomeric proton region (Kiemle et al 2004; Mittal et al 2009) could be compared with the PLS modeling. All three methods have been weight corrected to the total amount of biomass that was hydrolyzed so that the different preparation procedures could be compared. All three methods

agree, within their respective error, where the NMR PLS error bars indicate the root mean squared error of prediction (RMSEP) and the NMR integration and HPLC error bars are the standard deviation of the replicate samples. The wide range of sugar concentrations found in the variety of feedstocks allows sugar concentrations outside of the normal range of variability within a single species to be accurately predicted. In addition, this PLS2 model is able to predict sugar concentrations for species not contained in the model. For example, the model did not contain any bagasse samples but is able to accurately predict sugar concentrations from this species as shown in Figure 6. This flexibility of the model predications is particularly important in transgenic plant studies and for studies involving samples from a variety of locations, which may have vastly different sugar release properties.

For future studies, the implementation of a high sensitivity NMR cryoprobe will allow for the NMR spectra to be collected at even faster rates, allowing for a truly high-throughput method. Preliminary work with a Bruker 5mm TCI cryoprobe has shown that adequate NMR spectra can be collected with 8 scans in 1.5 minutes. With sample change time the full measurement time per sample is 4 minutes offering a 15-fold increase in the number of samples that can be run in an hour compared to HPLC.

Conclusions

The method presented here employing Partial Least Squares prediction of ^{1}H NMR spectra regressed against primary HPLC sugar concentration data is able to predict concentrations of both major sugar components, like glucose and xylose, and minor sugars, such as arabinose and mannose, in biomass hydrolysates. The data collection time for these NMR measurements is less than twenty minutes, offering a significant improvement to the one hour

acquisition time that is required for HPLC. Concentrations predicted from the PLS2 models are within error of both HPLC and NMR integration techniques for a set of 15 validation samples. The robust PLS2 model, built from a variety of different biomass species, covers a wide range of sugar concentration and is able to provide accurate concentrations even for species that are not included in the model.

Acknowledgements

This work was conducted as part of the BioEnergy Science Center (BESC). The BESC is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. This work was supported by the U.S. Department of Energy under Contract No. DE-AC36-08-GO28308 with the National Renewable Energy Laboratory.

References

ASTM E1758-01. 2003. Standard Test Method for Determination of Carbohydrates in Biomass by High Performance Liquid Chromatography. In: 2003 Annual Book of ASTM Standards, Vol. 11.05. West Conshohocken, PA: American Society for Testing and Materials, Int.

da Silva Neto HG, da Silva JBP, Pereira GE, Hallwass F. 2009. Determination of metabolite profiles in tropical wines by ^{1}H NMR spectroscopy and chemometrics. Magn Reson Chem 47(S1):S127–S129.

Decker SR, Bruenecky R, Tucker MP, Himmel ME, Selig M J. 2009. High-throughput screening techniques for biomass conversion. Bioenergy Res 2:179–192.

DeMartini JD, Studer MH, Wyman CE. 2011. Small-scale and automatable high-throughput compositional analysis of biomass. Biotechnol Bioeng 108(2):306–312.

Franks F. 1987. Physical chemistry of small carbohydrates – equilibrium solution properties. Pure Appl Chem 59(9):1189–1202.

Hames B, Ruiz R, Scarlata C, Sluiter A, Sluiter J, Templeton D. 2008. Preparation of Samples for Compositional Analysi. Golden, CO: National Renewable Energy Laboratory.

Himmel, M. 2008. Biomass Recalcitrance: Deconstructing the Plant Cell Wall for Bioenergy. New York: John Wiley & Sons.

Kiemle DJ, Stipanovic AJ, Mayo KE. 2004. Proton NMR methods in the compositional characterization of polysaccharides. ACS Symp Ser 864:122–139.

Lynd LR, Laser MS, Bransby D, Dale BE, Davison B, Hamilton R, Himmel M, Keller M, McMillan JD, Sheehan J, Wyman CE. 2008. How biotech can transform biofuels. Nat Biotechnol 26(2):169–172.

Mittal A, Scott GM, Amidon TE, Kiemle DJ, Stipanovic AJ. 2009. Quantatavie analysis of sugars in wood hydrolysates with ^{1}H NMR during the autohydrolysis of hardwoods. Bioresour Technol 100:6398–6406.

Powers R. 2009. NMR metabolomics and drug discovery. Magn Reson Chem 47(S1): S2–S11.

Sluiter A, Hames B, Ruiz R, Scarlata C, Sluiter J, Templeton D, Crocker D. 2008. Determination of Structural Carbohydrates and Lignin in Biomass. Golden, CO: National Renewable Energy Laboratory.

Sluiter JB, Ruiz RO, Scarlata CJ, Sluiter AD, Templeton DW. 2010. Compositional analysis of lignocellulosic feedstocks. 1. Review and description of methods. J Agric Food Chem 58(16):9043–9053.

Spraul M, Schütz B, Humpfer E, Mörtter M, Schäfer H, Koswig S, Rinke P. 2009. Mixture analysis by NMR as applied to fruit juice quality control. Magn Reson Chem 47(S1):S130–S137.

Studer MH, DeMartini JD, Brethauer S, McKenzie HL, Wyman CE. 2010. Engineering of a high-throughput screening system to identify cellulosic biomass, pretreatments, and enzyme formulations that enhance sugar release. Biotechnol Bioeng 105(2):231–238.

Sykes R, Yung M, Novaes E, Kirst M, Peter G, Davis M. 2009. High-throughput screening of platn cell-wall composition using pyrolysis molecular beam mass spectroscopy. In: Mielenz J, editor. Biofuels: Methods and Protocols, Vol. 581. New York: Humana Press, Inc. p 169–183.

Sykes R, Kodrzycki B, Tuskan G, Foutz K, Davis M. 2008. Within tree variability of lignin composition in *Poplus*.Wood Sci Technol 42(8):649–661.

Wyman C. 2007. What is (and is not) vital to advancing cellulosic ethanol. Trends Biotechnol 25(4):153–157.

List of Figures

Figure 1. Anomeric proton region of the ^{1}H NMR spectra for sugars dissolved in D_2O compared to a representative poplar biomass hydrolysate that was prepared in 100% D_2O and neutralized. In the region of the NMR spectra integration can be employed to determine concentrations (Kiemle et al 2004; Mittal et al 2009).

Figure 2. The 3.2–4.2 ppm region of the ^{1}H NMR spectra that can be used for PLS modeling for a poplar hydrolysate and pure sugars.

Figure 3. (a) PCA scores and (b) PC1 and (c) PC2 loadings plots for hydrolysates prepared in duplicate from nine different biomass feedstocks for the multiple feedstock species study.

Figure 4. PLS models for the two major sugars, glucose and xylose, for which predictive models were built from the multiple feedstock species dataset. (a,b) Model along with R^2 of validation, number of factors used to create the model, Root Mean Squared Error of Calibration (RMSEC), and Root Mean Squared Error of Validation (RMSEV). (c,e) NMR spectra for the modeled sugar for comparison and (d,f) the regression coefficients for the models.

Figure 5. PLS models for the minor sugars galactose, arabinose, and mannose. (a,b,c) Model along with R^2 of validation, number of factors used to create the model, Root Mean Squared Error of Calibration (RMSEC), and Root Mean Squared Error of Validation (RMSEV). (d,f,h) NMR spectra for the modeled sugar for comparison and (e,g,i) the regression coefficients for the models.

Figure 6. NMR PLS predictions for (a) glucose, (b) xylose, (c) arabinose, (d) galactose, and (e) mannose for samples prepared in H_2O compared to HPLC data and to concentrations determined using the previously published method of integration of the anomeric protons in samples that

were prepared in D_2O (Kiemle et al 2004; Mittal et al 2009). All concentrations are weight corrected and given in mg sugar/mg biomass.

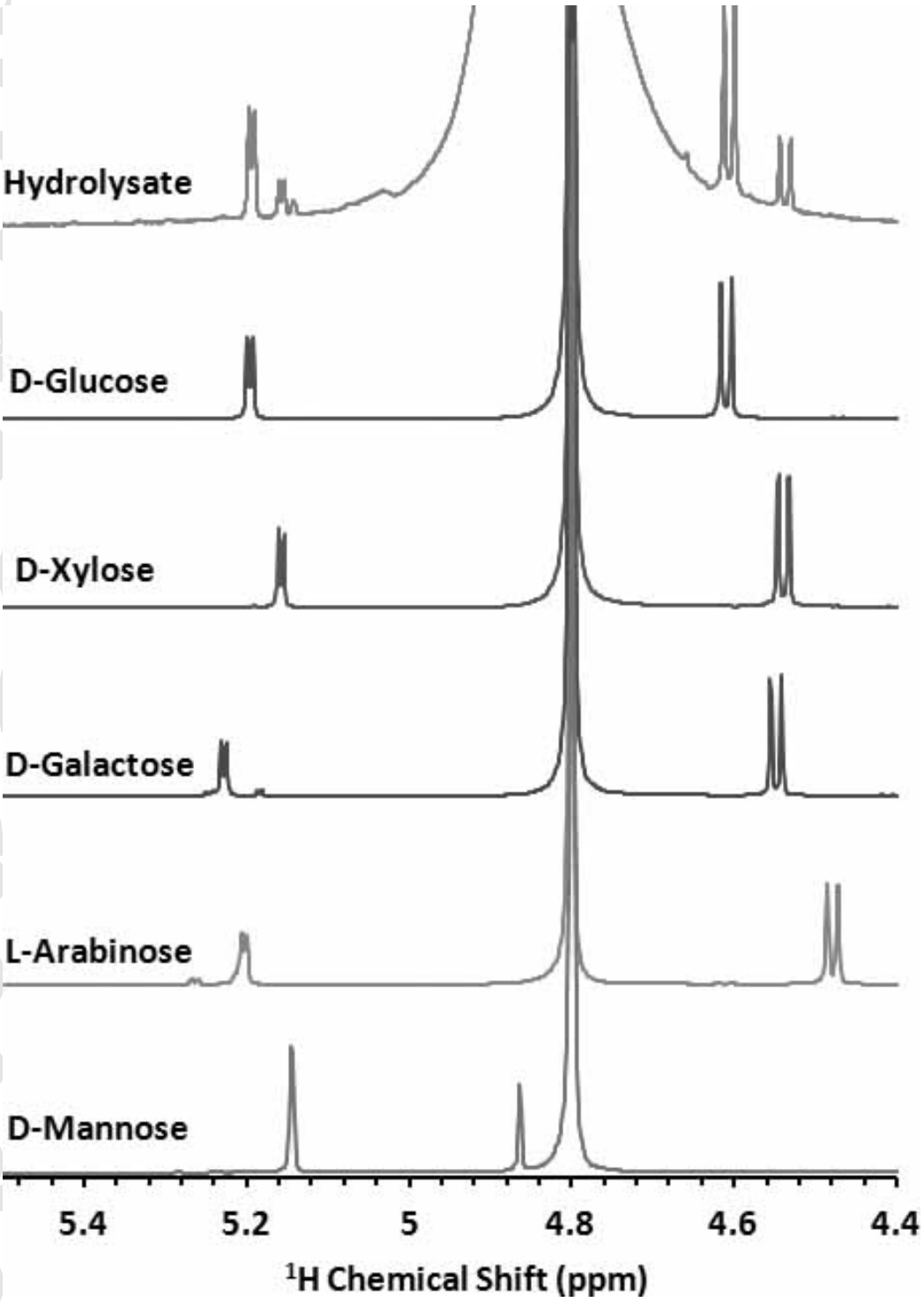

Figure 1

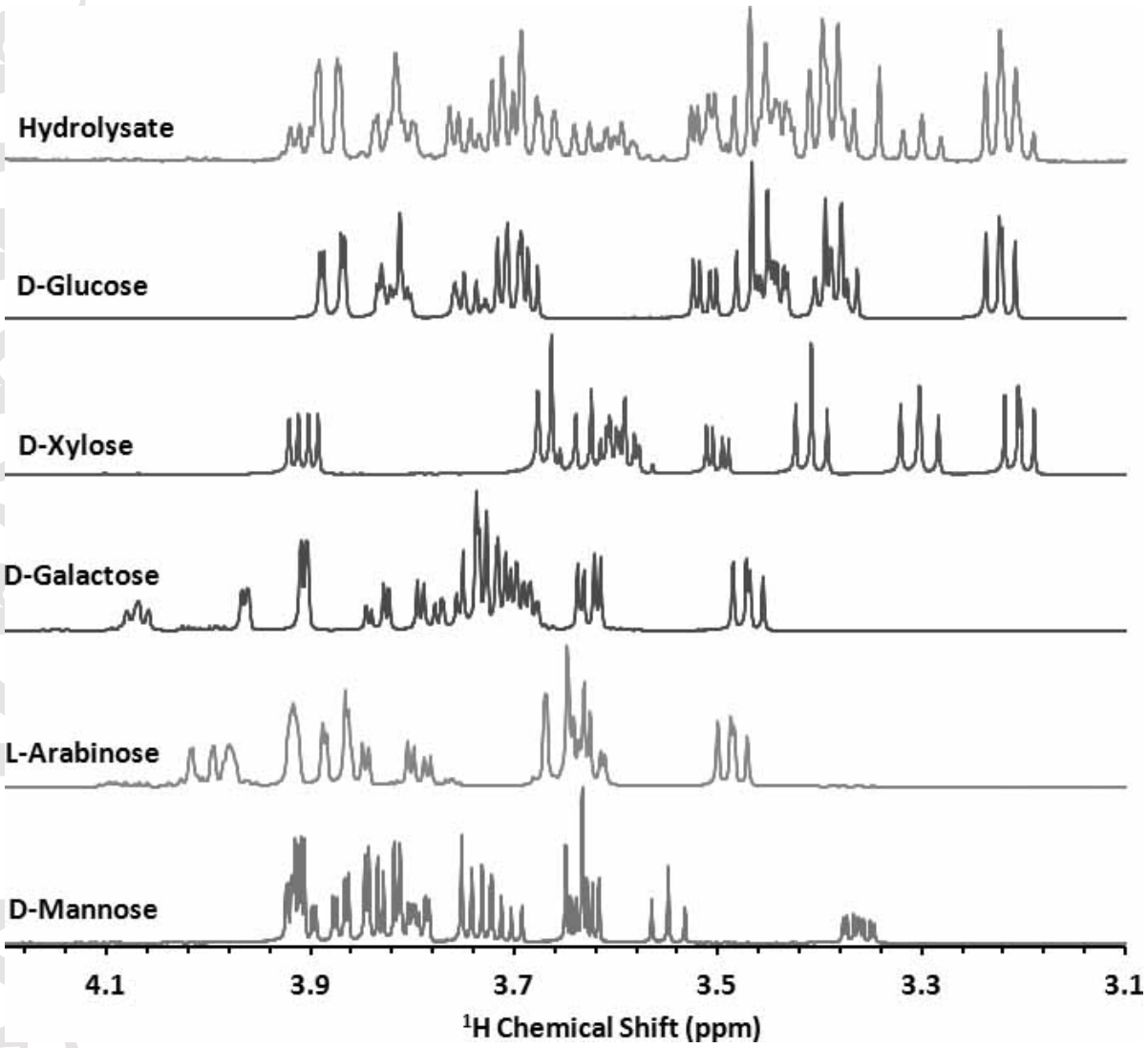

Figure 2

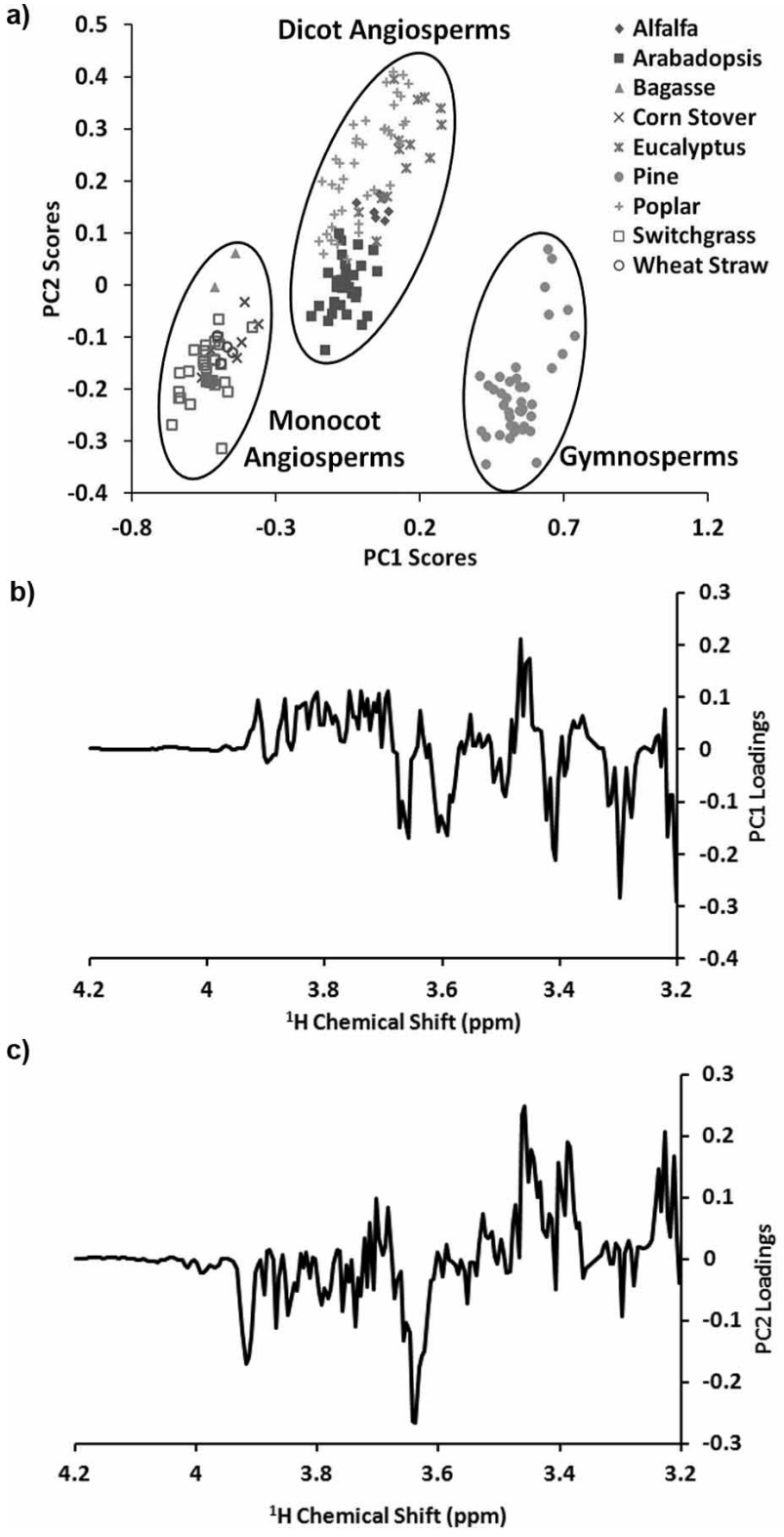

Figure 3

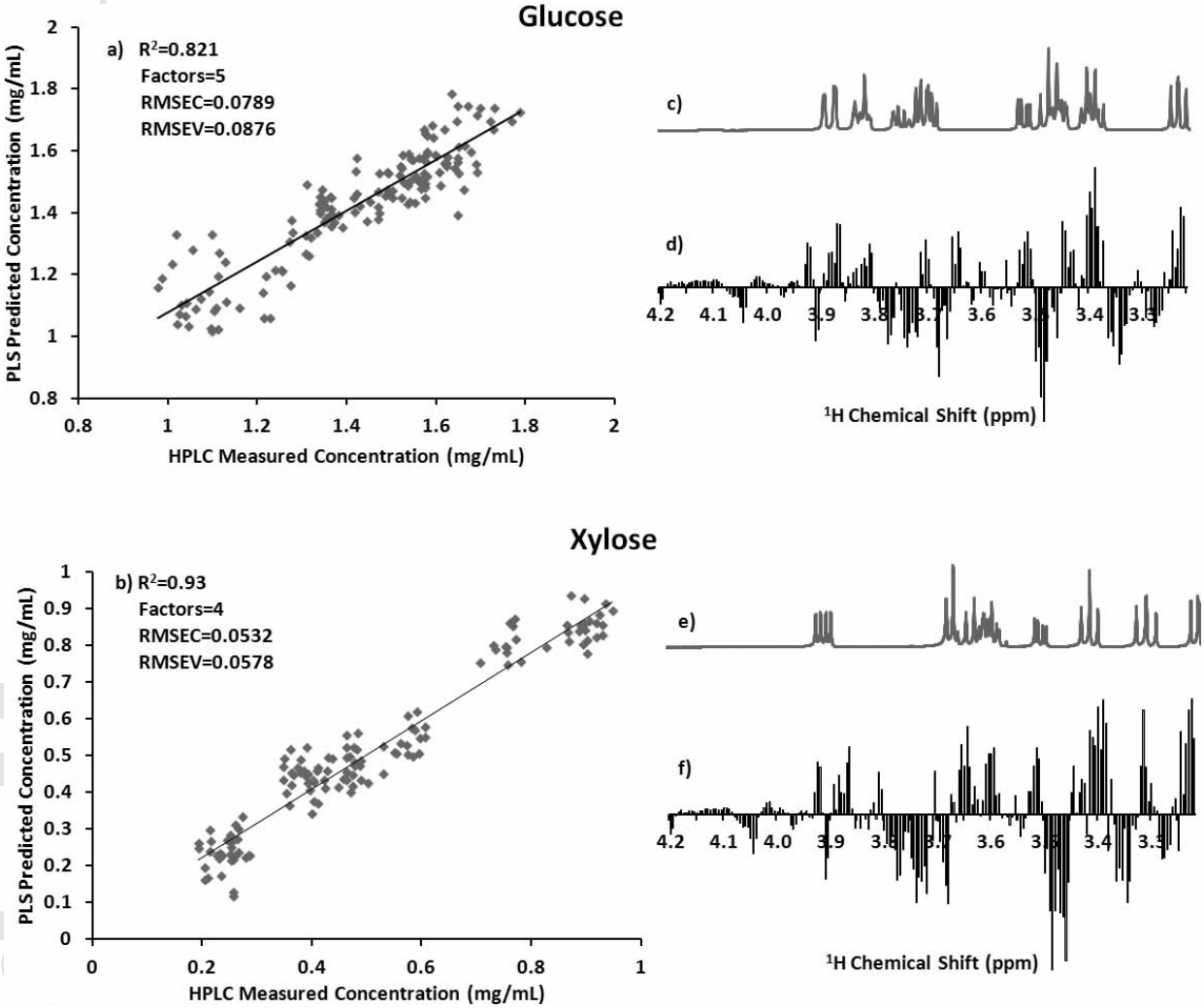

Figure 4

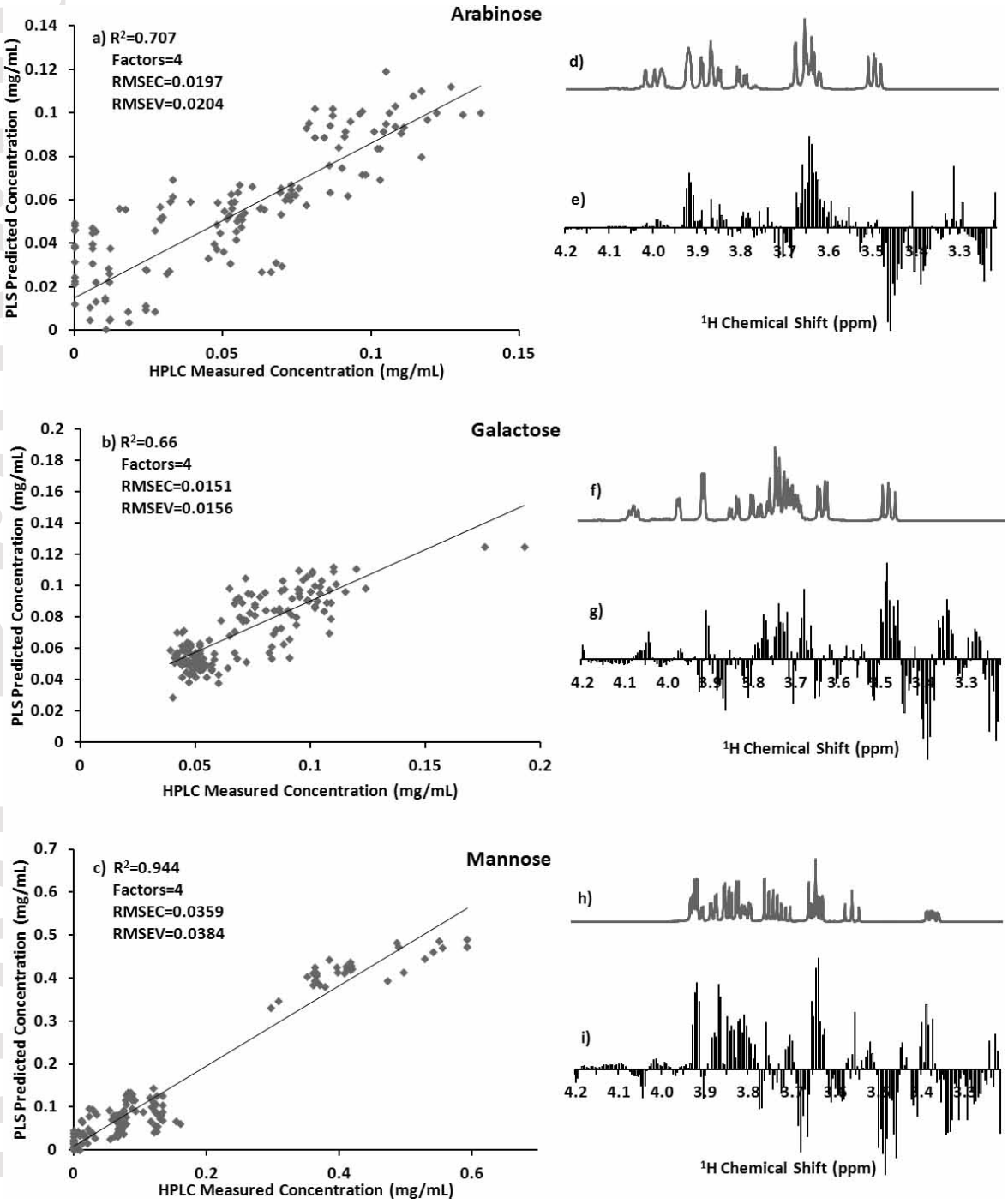

Figure 5

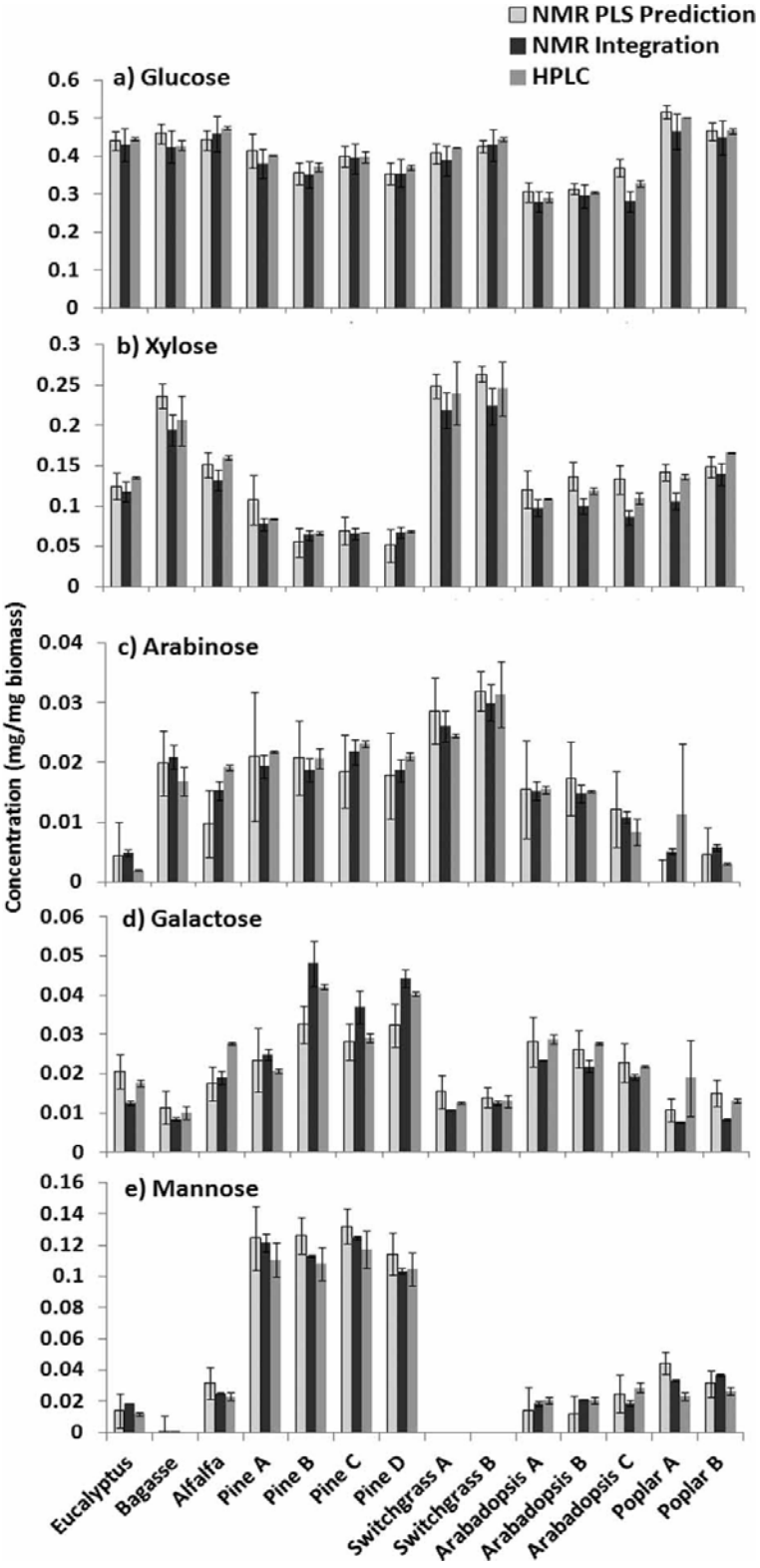

Figure 6