Multivariate analysis application for a metabonomics study

Introduction

Onion intake is associated to health benefits with respect to properties such as anti-oxidation, anti-carcinogenic, anti-asthmatic. A large amount of waste from onion production is produced worldwide and not used as it can be toxic for some animals. The stabilization of the onion by-product could be used as dietary supplement if they contain the active molecules that have proven beneficial. This study by 1H NMR spectroscopy of onion-fed rat urines aims at localizing onion biomarkers.

Challenge

An exploratory NMR nutri-metabonomic investigation to identify dietary biomarkers from onion intake.

Abstract

Metabonomics is an extension of genomics and proteomics which deals with quantitative measurement of metabolomes, or molecular profiles, and their response to external factors. Metabonomics can lead to more efficient drug discovery, specialized treatments, and targeted diets. Multivariate data analysis is essential in metabonomic studies. The data presented here originates from 1H NMR analyses of urine from thirty-two rats fed a diet containing one of the two derived onion by-product fractions: an ethanol extract and the residue. This study is based on the following publication: An exploratory NMR nutri-metabonomic investigation reveals dimethyl sulfone as a dietary biomarker for onion intake,[1]. A 24-hour urine sample was analyzed using 1H NMR spectroscopy in order to investigate the effects of onion intake on the rat metabolism.

Application of variable selection by the uncertainty test in PLS regression proved to be able to identify two dietary biomarkers for onion intake. These were identified as dimethyl sulfone and 3-hydroxyphenylacetic acid.

> **Being able to detect specific dietary biomarkers is highly beneficial in the control of nutritionally enhanced functional foods.**

Conclusions

This study reveals two biomarkers for onion intake. They were found by PLS regression on different chemical areas of the NMR spectrum followed by variable selection based on an uncertainty test. Those markers were confirmed to be able to predict the onion content in the rats' diet using a test set of 4 samples.

This example of metabonomics study shows how multivariate analysis can be used to find signals of interest in a larger signal, and that one can develop a predictive model for the onion content in the diet from the NMR spectrum of urine.

Materials and Methods

A by-product of onion was collected from the production of a pasteurized onion paste and incorporated in the rats' diet. 31 male rats were divided into four groups and fed four weeks with different amount of onion by-product (0, 3, 7, 10%). The urine samples were collected over a period of 24h and frozen with NaN3. The sample preparation included a centrifugation and the addition of a buffer to reduce the pH shift effect and the addition of TSP-d4 to get a peak reference in the NMR. The spectra were acquired after 24 hours on a 400 MHZ Bruker spectrometer with a 32k resolution. The pulse sequence used a pre-saturation of the water peak. The spectra were aligned on the TSP peak and normalized on the spectral area. The area containing the water peak was removed (4.50 to 5.00 ppm). The data set was reduced to a smaller data set by reducing the resolution by averaging over 10 data points.

The data set after pretreatment contains:

- X: 31 x 29001 (NMR spectra in the region between 9.6 and 0.3 ppm)
- Y: amount of onion by-product in the diet in percentage

Analysis methods

A partial least squares regression (PLS) was conducted on the 4 separated regions and on the training set to find the peaks that act as predictors of the onion content in the urine. PLS allows for development of quantitative models based on the spectral data.

A variable-selection procedure based on the uncertainty test as described by Martens & Martens [2]. When the selection procedure was finished the segments found in the different zones were used to develop a predictive model. This model was tested on 4 samples. A full validation with a larger test set is recommended.

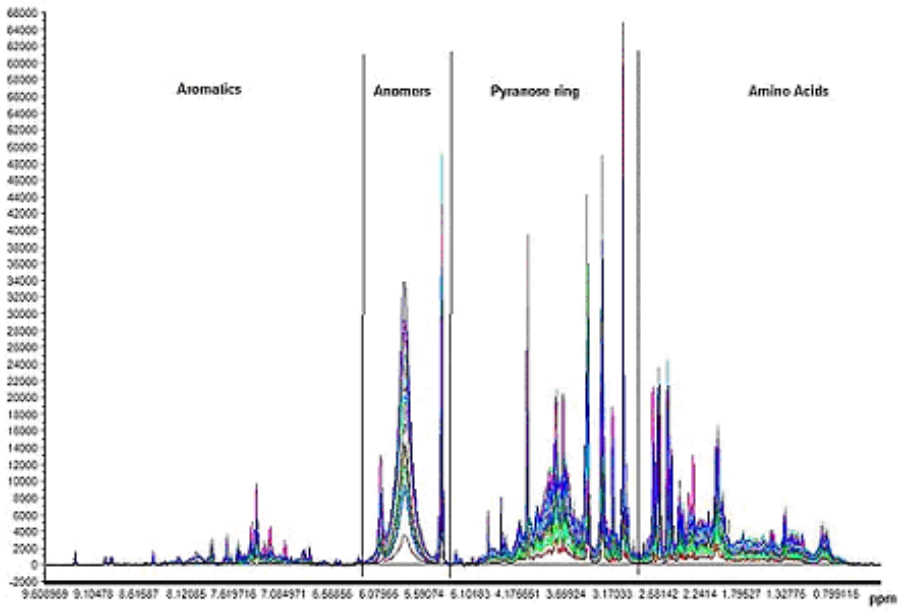

Figure 1: *Rats urine spectra divided into 4 chemical shift regions after pretreatment*

Results and discussion

Selection step

Four PLS models were made on the 4 chemical ranges. The results are presented in the table below.

Area	Optimal number of factors	R-square in calibration	R-square in validation	RMSECV
All spectra	9	0.985	0.746	2.03
Aromatics	4	0.925	0.828	1.68
Anomers	none; factor 1	0.193	NA	4.216
Pyranose rings	5	0.924	0.765	1.926
Amino acids	6	0.854	0.375	3.11

Table: PLS results on different spectral area

Two regions give better results than the full model with respect to RMSECV:

- aromatics and
- pyranose rings.

Those models were optimized by selecting only the variables that had a stable regression coefficient in the PLS regression model according to Marten's uncertainty test [2]. The results of this selection gave one peak in each region.

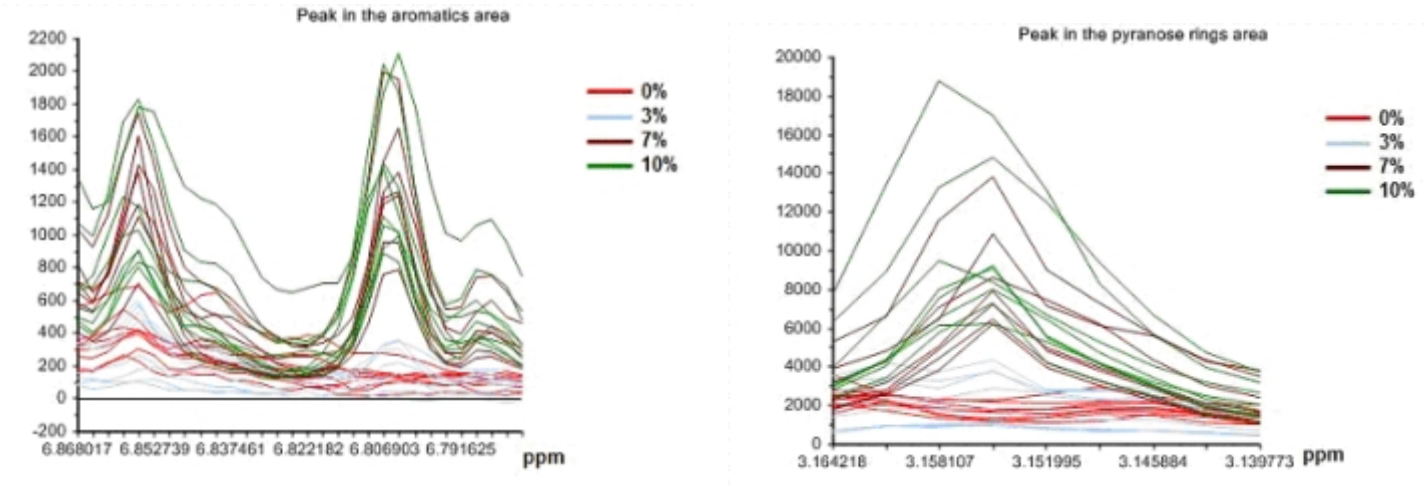

Figure 2: *Selected peaks in the spectra as biomarkers for onion by-product intake*

Model on aggregated peaks

As the two peaks do not have the same intensity range but all contain useful information it is possible to use a scaling function such as a normalization (1/(standard deviation)).

The model performs less well than a model on the total spectra which may be due to a reduction of information in going from 29001 variables in the original data to 39.

However the R-squares are 0.89 in calibration and 0.79 in validation. The error to be expected when predicting (RMSEP) is about 1.82 which is about 20% error.

The two isolated peaks have proven to be valuable to predict the onion by-products content in the diet at less than 20% error. These two peaks can then be identified as markers.

The structure shows a very good discrimination between the different levels of onion by-product and also shows the variability in each level, also seen in the raw data. Samples from the calibration and cross-validation are displayed in the plot.

Prediction

The model is tested on 4 samples that were not part of the calibration (one in each category). The prediction results show a good attribution to each of the 4 groups of concentration. The maximum error made on this prediction is for the first sample: 1.45 which is expected from the calibration model performance.

The R-square of the predicted vs. reference value is 0.95.

Attribution

From the literature and the composition of the onion by-products, the two selected peaks can be attributed to dimethyl sulfone and 3-hydroxyphenylacetic acid.

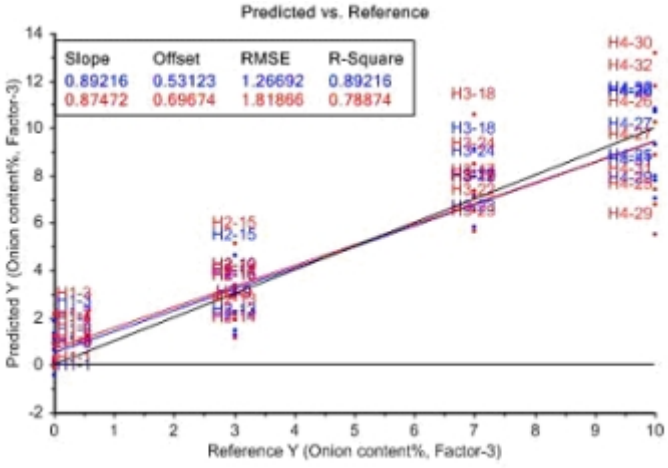

Figure 3: *Predicted vs. reference of the calibration model with 3 factors*

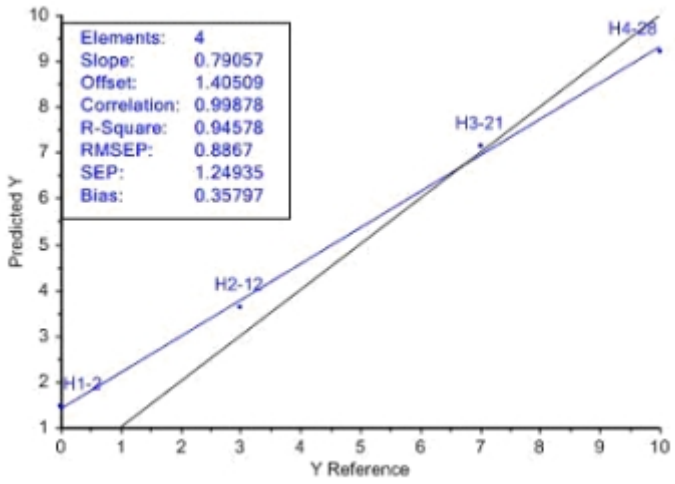

Figure 4: *Predicted vs. reference values of the diet content*

Conclusions

This study reveals two biomarkers for onion intake. They were found by PLS regression on different chemical areas of the NMR spectrum followed by variable selection based on an uncertainty test. Those markers were confirmed to be able to predict the onion content in the rats' diet using a test set of 4 samples.

This example of metabonomics study shows how multivariate analysis can be used to find signals of interest in a larger signal, and that one can develop a predictive model for the onion content in the diet from the NMR spectrum of urine.

Application note overview

Software	The Unscrambler® X
Methods	PLS and uncertainty test
Data type	1H NMR spectroscopy data
Industry	Biology, Metabonomics, Medicine
Added Value	PLS regression of 1H-NMR data was used to identify two dietary biomarkers for onion intake: dimethyl sulfone and 3-hydroxyphenylacetic acid.

For complete details of this work please see:

[1] H. Winning, E. Roldán-Marín, L.O. Dragsted, N. Viereck, M. Poulsen, C. Sánchez-Moreno, M.P. Cano, S.B. Engelsen, An exploratory NMR nutri metabonomic investigation reveals dimethyl sulfone as a dietary biomarker for onion intake, Analyst 2009, 134 (11) 2344-2351.

[2] H. Martens, M. Martens, - 1999 -, Modified Jackknife Estimation of Parameter Uncertainty in Bilinear Modelling (PLSR), Food Quality and Preference.