ORIGINAL ARTICLE

Metabonomics based NMR in Crohn's disease applying PLS-DA

Fariba Fathi[1], Afsaneh Arefi Oskouie[2], Mohsen Tafazzoli[1], Nosratollah Naderi[3], Kaveh Sohrabzedeh[4], Soraya Fathi[5], Mohsen Norouzinia[3], Mohammad Rostami Nejad[3]

[1]Department of Chemistry, Sharif University of Technology, Tehran, Iran

[2]Department of Basic Science, Faculty of Paramedical, Shahid Beheshti University of Medical Sciences, Tehran, Iran

[3]Gastroenterology and Liver Disease Research Center, Shahid Beheshti University of Medical Sciences, Tehran, Iran

[4]Department of electrical engineer, Golpayegan Payam University, Tehran, Iran

[5]Department of Mathematics, Teacher Training University, Tehran, Iran

ABSTRACT

Aim: The aim of this study was to search for metabolic biomarkers of Crohn's disease (CD).

Background: Crohn's disease (CD) is a type of inflammatory bowel disease that causing a wide variety of symptoms. CD can influence any part of the gastrointestinal tract from mouth to anus. CD is not easily diagnosed because monitoring tools are currently insufficient. Thus, the discovery of proper methods is needed for early diagnosis of CD.

Patients and methods: We utilized metabolic profiling using proton nuclear magnetic resonance spectroscopy (^{1}HNMR) to find the metabolites in serum. Classification of CD and healthy subject was done using partial least squares discriminant analysis (PLS-DA).

Results: According to PLS-DA model, we concluded that just using one descriptor CD and control groups could be classified separately. The level of lipid in blood serum of CD compared to healthy cohorts was decreased. For the external test set, the classification model showed a 94% correct classification of CD and healthy subject.

Conclusion: The result of classification model presents that NMR based metabonomics is key tool as well as insight into potential targets for disease therapy and prevention.

Keywords: ^{1}HNMR, Metabonomics, Crohn's disease, PLS-DA.

(Please cite as: **Fathi F, Arefi Oskouie A, Tafazzoli M, Naderi N, Sohrabzedeh K, Fathi S, et al. Metabonomics based NMR in Crohn's disease applying PLS-DA. Gastroenterol Hepatol Bed Bench 2013;6(Suppl.1):S82-S86).**

Introduction

Crohn's disease (CD) is a type of inflammatory bowel disease (IBD) that is found in the last part of the small intestine and the first part of the large intestine (1-3). Both the host genotype and environmental factors play a role in etiology of CD. Also presence of bacteria needs for disease induction.

In order to exact diagnosis of CD, several serological biomarkers have been proposed. But they are used in conjunction and as a supplement to endoscopy. Thus metabonomics as a monitoring tool are needed for early diagnosis (4). Metabonomics is defined as "the quantitative measurement of the dynamic multi-parametric response of living systems to patho physiological stimuli or genetic modification" (5). One of the most commonly applied for metabonomics is proton nuclear magnetic resonance (^{1}H NMR). This analytical technology has several advantages

as follows: It provides quantitative and reproducible information with little sample preparation, and hence it is widely used to build metabolic profiles in diverse metabolic studies. ^{1}HNMR spectroscopy of biofluid and tissue samples has been applied to the investigation and diagnosis of many diseases of gastrointestinal (6-8).

So far several study has investigated metabolite profiles in serum of patients with CD in the literature (9). In the study by Schicho et al. metabolites in serum and plasma of IBD subjects were analysed. They obtained regular one-dimensional proton NMR spectra using a standard pulse sequence (Bruker pulse program prnoesy1d) (9). In other survey Fathi et al. applied classification and regression tree (CART) to explore the metabolic biomarkers causes of CD compare to control group (10).

In this study, to determine the metabolites, we used ^{1}H NMR spectroscopy and performed quantitative analysis of metabolites in the serum of patients with active CD. PLS-DA was employed as a powerful classification method. Our aim was to search for metabolic biomarkers of CD to classify the control and CD groups.

Patients and Methods

Sample collection

Twenty-six adult patients with mean age (± standard deviation) of 33.6 ± 11.3 years diagnosed with Crohn's disease and twenty-nine healthy subjects with mean age (± standard deviation) 34.7±12.2 years were recruited from Gastroenterology and Liver Disease Research Center, Shahid Beheshti University of Medical Sciences. To avoid the affect of aging and gender influence on metabonomics, the healthy subject were matched with CD subjects (9). Experienced gastroenterologists made diagnosis of CD on the basis of radiographic, clinical findings and often colonoscopy criteria. Both CD and healthy cohort who entered to study had

not significant other past medical history including hypertension, diabetes mellitus or hyperlipidemia. Serum samples were collected in the morning after a 12 hour fast and store at −70 °C till measurement. ^{1}HNMR spectroscopy and data preprocessing were thoroughly explained in our previous study (11).

Statistical analysis

Partial least squares discriminate analysis (PLS-DA) was employed using PLS_Toolbox method (Version 2, Eigenvector Research Inc., Manson, WA) within MATLAB (version 6.5.1, The Mathworks, Cambridge, U.K.). This method is a regression technique adapted to a supervised classification task. Thus it is a frequently used for classification method. It is on the basis of the partial least squares (PLS) approach (12). The standard PLS algorithm can be utilized and class labels can be used for the dependent Y vector. Usually in the two-class case, the values of the Y are given 1 for one class and 0 or -1 for the other class. Using this supervised analysis technique; we can identify those metabolites which show a discrepancy between diagnostic groups (13). In this study, ^{1}HNMR data and class labels were used as x matrix and y matrix respectively. The dataset was divided into two parts training set and test set. Training set was used to build a model and identify the most relevant metabolites and in order test predictive ability of the classification model, test set was employed.

Results

In classification part, PLS-DA was used to classify CD and healthy samples. Based on a randomly choice, approximately 30% of samples are left out to form the test set. Consequently, the training set was included of 39 ^{1}H NMR spectra and the test set was composed of 16 spectra. By means of this procedure we can reduce the risk of over-fitting and avoid any possibility that the best classification

models selected had a chance correlation to peculiarities in the capacity of the test set(10). Figure 1 and Figure 2 present the score plots of the related PLS-DA of the ^{1}H NMR spectra of the serum for training and test sets respectively.
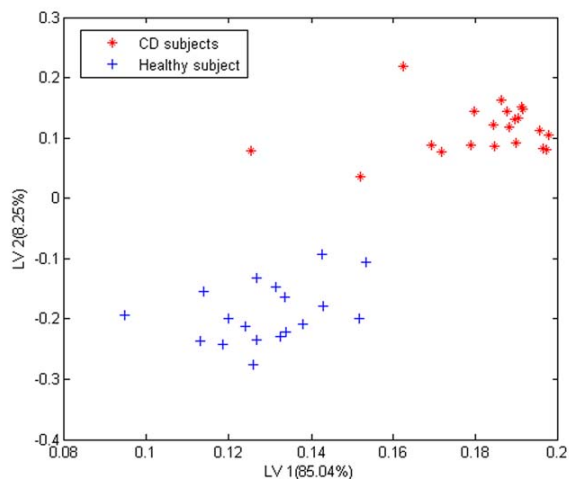

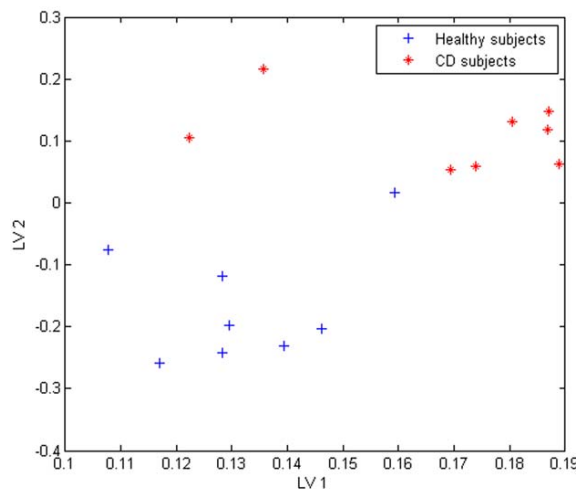**Figure 1.** PLS-DA sores plot of serum dataset: training set

Figure 2. PLS-DA sores plot of serum dataset: test set

In Figures 1 and 2, the first two latent variables LV1 versus LV2 map the representative points of the serum samples in the space were spanned. Also scores plot was elucidating a logical clustering appearing according to class (CD and healthy subject). To explore exactly which metabolites have caused the separation between CD and healthy subjects, the loadings plot of PLS-DA model is depicted in Figure 3. Table 1 reveals that the metabolite was the most prominent in serum separation with P-value < 0.00001.
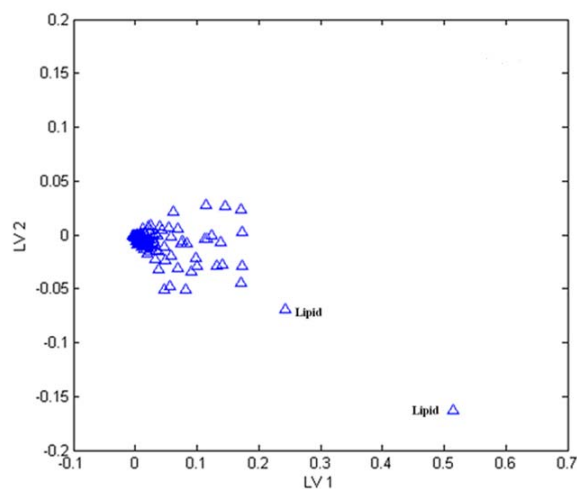

Figure 3. PLS-DA the corresponding loadings plot: training set

A confusion matrix (17) includes knowledge about the number of correct and incorrect predictions compared to the real outcomes by a classification model. Performance of such systems is commonly evaluated using the data in the matrix. The Table 2 shows the confusion matrix for a two class classifier. As it is clear from Table 2, PLS-DA model has an accuracy of 0.94 in detecting CD patients for external test set.

Table 1. Specifications of the selected PLS-DA descriptor

Descriptors	Assignment	[1]H chemical shift (ppm)
Lipid	$CH_3CH_2(CH_2)_n$	1.26

Table 2. Confusion matrix for training and test set

set	Observation	CD class	Healthy class
Training	CD class	18	0
	Healthy class	0	21
Test	CD class	7	1
	Healthy class	0	8

A summary of the classification parameters are shown in Table 3. These results show that PLS-DA classification model has great chance in the

diagnosis of CD. Area under ROC curve is often used as a measure of quality of the classification models. A random classifier has an area under the curve of 0.5, while AUC for a perfect classifier is equal to 1. In practice, most of the classification models have an AUC between 0.5 and 1. The obtained values of AUC for training and test set are 1 and 0.94, respectively. The high AUC score of the proposed model for the samples in the external test set is another evidence for capability of PLS-DA model in CD detection.

Table 3. The calculated error and non-error rates of the classification index and the classification performances of training and test sets

	Error rate	Non-error rate	specificity	sensitivity	accuracy
Training set	0	1	1	1	1
Test set	0.06	0.94	0.89	1	0.94

Discussion

Investigation of the selected variable revealed that the selected chemical shift correspond to the NMR spectrum of lipid (14-16). Consequently it can be stated that the discrimination of CD and control samples by the PLS_DA model, based on NMR data, is on the basis of different amounts of lipid in the two groups. This result shows the reduction of lipid level in blood serum of patients compared to healthy individuals. Using PLS_DA method, the most important metabolite lipid was identified. The results of Kuroki and Hrabovsky et al. for lipid level in CD are similar to our study. They found that the serum levels of total lipids and total cholesterol were reduced in patients with CD (18, 19).

Lipids are molecules that include fat-soluble vitamins (such as vitamins A, D, E and K), monoglycerides, diglycerides, phospholipids, and others. They are necessary elements which acting as structural components of cell membranes. Main

energy supply in cells and tissues is role of lipids. Also it has been proved that the serum vitamin E levels associates almost with serum lipid levels (20, 21). In the study by Fernandez-Banares et al. and Kuroki et al. (22, 23), serum concentrations of vitamin E have also been found to be lower in CD patient.

Fernandez-Banares et al. stated that the pathophysiological and clinical implications of the suboptimal vitamin status observed in acute CD are unknown. Kuroki and colleagues suggested that there is a variety of vitamin deficiencies in CD prior to treatment that may reflect the severity of the disease (23).

Vitamin E is as a fat-soluble antioxidant; it can be effective at scavenge free radicals at the cellular level and can be also prevented a great deal of the resultant scarring during CD. The result of inflammation in the gut mucosa and decreased oral intake can be cause risk for vitamin and mineral deficiencies in CD. Probably the reduce of lipid levels may relate to the remission of body level of Vitamin E. Classification model prediction procedure permits demonstrating that the classification obtained by the PLS-DA technique is good enough to execute classification of unknown samples.

In the present work, a ^{1}H NMR based metabonomics approach was gave an evidence for the existence of clear metabolic differentiation between two groups (CD and control group).

Since ^{1}H NMR based metabonomics is effective to monitor the progression of disease, and helpful to discover biomarkers of CD, we can suggest that, NMR based on metabonomics can provide the possibility for assisting in early of CD. Therefore, further investigations are required to establish its real usefulness in clinical practice.

Acknowledgements

References

1. Baumgart D, Carding S. Inflammatory bowel disease: cause and immunobiology. Lancet 2007;369:1627–40.

2. Baumgart D, Sandborn W. Inflammatory bowel disease: clinical aspects and established and evolving therapies. Lancet 2007;369:1641–57.

3. Xavier R, Podolsky D. Unravelling the pathogenesis of inflammatory bowel disease. Nature 2007;448:427–34.

4. German JB, Hammock BD, Watkins SM. Metabolomics: building on a century of biochemistry to guide human health. Metabolomics 2005;1:3-9.

5. Nicholson J, Lindon J, Holmes E. Metabonomics: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 1999;29:1181-89.

6. Hong Y-S, Hong KS, Park M-H, Ahn Y-T, Lee J-H, Huh C-S, et al. Metabonomic understanding of probiotic effects in humans with irritable bowel syndrome. J Clin Gastroenterol 2011:415–25.

7. Bezabeh T, Somorjai RL, Smith ICP. MR metabolomics of fecal extracts: applications in the study of bowel diseases. Magn Reson Chem 2009;47:54-61.

8. Gall GL, Noor SO, Ridgway K, Scovell L, Jamieson C, Johnson IT, et al. Metabolomics of Fecal extracts detects altered metabolic activity of gut microbiota in ulcerative colitis and irritable bowel syndrome. J Proteome Res 2011;10:4208–18.

9. Schicho R, Shaykhutdinov R, Ngo J, Nazyrova A, Schneider C, Panaccione R, et al. Quantitative metabolomic profiling of serum, plasma, and urine by ^{1}H NMR spectroscopy discriminates between patients with inflammatory bowel disease and healthy individuals. J Proteome Res 2012;11:3344−57.

10. Fathi F, Kyani A, Rostami Nejad M, Rezaye-Tavirani M, Naderi N, Zali MR, et al. A metabonomics study on Crohn's disease using nuclear magnetic resonance spectroscopy. HealthMED 2012;6:3577-84.

11. Fathi F, Tafazzoli M, Mehrpour M, Shahidi G. Study of metabolic profiling Parkinson's disease. HealthMED 2013; 7:204-10.

12. Barker M, Rayens W. Partial least squares for discrimination. J Chemometr 2003;17:166–73.

13. Ramadan Z, Jacobs D, Grigorov M, Kochhar S. Metabolic profiling using principal component analysis, discriminant partial least squares, and genetic algorithms. Talanta 2006;68:1683–91.

14. Nicholson JK, Foxall PJ, Spraul M, Farrant RD, Lindon JC. 750 MHz 1H and 1H-13C NMR spectroscopy of human blood plasma. Anal Chem 1995;67: 793-811.

15. Fan M. Metabolite profiling by one- and two-dimensional NMR analysis of complex mixtures. Prog Nucl Magn Reson Spectrosc 1996;28:161-219.

16. Lindon JC. NMR Spectroscopy on biofluids. Annu Rep NMR Spectrosc 1999;38:1-88.

17. Kohavi R, Provost F. Glossary of Terms. Machine Learning 1998;30:271-74.

18. Hrabovsky V, Zadak Z, Blaha V, Hyspler R, Ticha A, Karlik T. Lipid metabolism in active crohn's disease: pre-results. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub 2006;150:363–66.

19. Kuroki F, Iida M, Tominaga M, Matsumoto T, Kanamoto K, Fujishima M. Is vitamin E depleted in Crohn's disease at initial diagnosis? Dig Dis 1994;12:248-54.

20. Rubinstein HM, Dietz AA, Srinivasan R. Relation of vitaminE and serumlipids. Clin Chim Acta 1969;23:1–6.

21. Horwitt MK, Harvey CC, Dahm CHJ, Searcy MT. Relationship between tocopherol and serum lipid levels for determination of nutritional adequacy. Ann NY Acad Sci 1972;203:223-36.

22. Fernandez-Banares F, Abad-Lacruz A, Xiol X, Gine JJ, Dolz C, Cabre E, et al. Vitamin status in patients with inflammatory bowel disease. Am J Gastroenterol 1989;84:744-48.

23. Kuroki F, Iida M, Tominaga M, Matsumoto T, Hirakawa K, Sugiyama S, et al. Multiple vitamin status in Crohn's disease; correlation with disease activity. Dig Dis Sci 1993;38:1614-18.