Jurnal Teknologi

Full Paper

DISCRIMINATION OF DENGUE DISEASE FROM HEALTHY BASED ON THE CHEMOMETRY OF 1H NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY

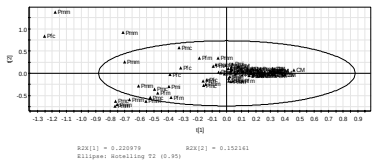
Nurul Shahfiza[a], Maulidiani[b], Hasnah Osman[c], Tang T. Hock[a], Khozirah Shaari[d], Baharudin Ibrahim[e], Abdel-Hamid Z. Abdel-Hamid[f,a*]

[a]Advanced Medical & Dental Institute, Universiti Sains Malaysia, Penang, Malaysia
[b]Laboratory of Natural Products, Institute of Bioscience, Universiti Putra Malaysia, Selangor, Malaysia
[c]School of Chemical Sciences, Universiti Sains Malaysia, Penang, Malaysia
[d]Department of Chemistry, Science Faculty, Universiti Putra Malaysia, Selangor, Malaysia
[e]School of Pharmaceutical Sciences, Universiti Sains Malaysia, Penang, Malaysia
[f]Therapeutic Chemistry Department, National Research Centre, Cairo, Egypt

Article history
Received
15 July 2015
Received in revised form
1 October 2015
Accepted
25 October 2015

*Corresponding author
abdelhamidzaki@hotmail.com

Graphical abstract



Abstract

Dengue is the most important human viral disease transmitted by arthropod vectors and over half of the world's populations live in areas at risk of infection. The severity of the infection depends on the form of the disease, which can be symptomatic or asymptomatic. Currently there is neither specific treatment nor vaccine to tackle this emerging disease. Metabolomics applied in this study, aimed to provide a global snapshot of all small-molecule metabolites in urine as biological sample of choice to more focused studies of metabolism to distinguish between healthy and dengue infected subjects. Fifty-two patients diagnosed with dengue fever at Penang General Hospital and fourty-three healthy individuals were recruited in this study. ^{1}H-nuclear magnetic resonance (NMR) spectroscopy combined with multivariate analysis (MVA) methods such as principal component analysis (PCA), partial least square discriminant analysis (PLS-DA) and orthogonal PLS-DA (OPLS-DA) were employed for statistical data exploration. The model score plot results showed that all three MVAs showed very good spatial distributions with clear clusters/grouping between healthy individuals and dengue infected individuals. Also, statistically, the PLS-DA and OPLS-DA models had high reproducibility and predictivity values, > 0.5. In conclusion, this study established the potential of using a combination of ^{1}H NMR spectroscopy and multivariate data analyses in differentiating healthy and non-healthy individuals, based on obtained score plots reflecting the metabolites pertubation, where spectral features contributing most to variation or separation are identified for further analysis.

Keywords: Dengue; NMR spectroscopy; multivariate analysis; metabolomics

Abstrak

Denggi adalah penyakit virus manusia yang paling penting disebarkan oleh vektor Artropod dan lebih separuh daripada penduduk dunia tinggal di kawasan berisiko dijangkiti. Keterukan jangkitan bergantung kepada bentuk penyakit, yang boleh menjadi gejala atau asimptomatik. Pada masa ini tiada rawatan khusus atau vaksin untuk menangani penyakit ini. Metabolomik digunakan dalam kajian ini, bertujuan untuk memberi gambaran global semua metabolit molekul kecil dalam air kencing sebagai sampel biologi pilihan untuk kajian yang lebih terperinci terhadap metabolisme untuk membezakan antara subjek yang sihat dan dijangkiti denggi. Lima puluh dua pesakit disahkan menghidap demam denggi di Hospital Besar Pulau Pinang dan empat puluh tiga individu yang sihat telah diambil dalam kajian ini. 1H-nuklear resonans magnetik (NMR) spektroskopi digabungkan dengan analisis multivariat (MVA) seperti 'principal component analysis (PCA)', 'partial least square discriminant analysis (PLS-DA)' dan 'orthogonal PLS-DA (OPLS-DA)' telah digunakan untuk statistik penerokaan data. Keputusan plot skor model menunjukkan bahawa ketiga-tiga MVA menunjukkan taburan spatial yang sangat baik dengan kelompok jelas / kumpulan antara individu yang sihat dan dijangkiti denggi. Juga, statistik, PLS-DA dan OPLS-DA model mempunyai kebolehulangan yang tinggi dan nilai-nilai predictivity,> 0.5. Kesimpulannya, gabungan 1H NMR spektroskopi dan multivariat analisis data dalam kajian ini berpotensi untuk membezakan individu yang sihat dan tidak sihat, berdasarkan skor plot yang diperoleh, mencerminkan perubahan pada metabolit, di mana ciri-ciri spektrum menyumbang paling banyak kepada perubahan atau pemisahan dikenal pasti untuk analisis selanjutnya.

Kata kunci: Denggi; NMR spektroskopi; analisis multivariat; metabolomik

© 2015 Penerbit UTM Press. All rights reserved

1.0 INTRODUCTION

In Malaysia, dengue was first documented in 1902 and made reportable in 1971 [1]. After the first outbreak of dengue hemorrhagic fever (DHF) in 1962 [2], major dengue outbreaks occurred every four years until 1992; since then the disease became endemic with yearly and frequent outbreaks [3]. Malaysia's reported incidence of dengue has remained high, with an average of 125 to 150 per 100,000 people annually from 2002 to 2006. In 2010, the number of reported dengue cases rose by 12 percent to 45,901 and the number of recorded dengue fever show an increment of 54 percent from 2009 made the total number of fatalities to 134 cases [4]. However this is probably an underestimate since notification is not compulsory and due to lack of awareness. The histories of the disease are traced over the years and changes of clinical presentation have been noticed, means the diagnosis is often not confirmed or delayed. Therefore, the actual magnitude of dengue infection in Malaysia might be larger than expected [5] has urge for an early identification for dengue is need.

Taking into consideration the steadily increasing rate of dengue infection in Malaysia, metabolomics technologies are expected to be a powerful tool for identifying any disturbances of metabolic processes caused by dengue infection that can reveal a variety of health and disease traits than either genetic or proteomics information. The perturbation of the biological pathway that vary according to the physiological, developmental or pathologic state of the cell, tissue, organ or organism [6] can be used to elucidate changes of the metabotype associated to disease-related biochemical reactions to improve diagnostic, prognostication and therapy of dengue disease.

NMR metabolomics is currently being used to search for disease biomarkers for infectious diseases like tuberculosis dengue [7; 8], tuberculosis [9], pneumonia [10], malaria [11] and numerous other human diseases. NMR metabolomics is being used to understand the underlying causes of these diseases, and to identify chemical markers to quickly and readily diagnose the disease. Recently, mass spectrometry (MS), nuclear magnetic resonance spectroscopy (NMR), particularly, and multivariate statistical techniques have been incorporated into a multidisciplinary approach to profile changes in small molecules associated with the onset and progression of human diseases [12]. Moreover, metabolomics data are essentially multivariate [13].

The chemometric multivariate analysis techniques are used for analyzing and interpreting the complex spectral data sets and able to provide a clear interpretation of the global alteration in the metabolome [12]. Multivariate statistical methods are categorized as either supervised or unsupervised. Unsupervised methods like a Principle Component Analysis (PCA) are based strictly on inherent variations in the data while supervised methods including Partial Least Square Discriminant Analysis (PLS-DA) and Orthogonal Partial Least Square Discriminant Analysis (OPLS-DA) are used to introduce sample classes. OPLS-DA has been a popular choice among researchers as an approach to interpret NMR spectra in metabolomics investigation.

Urine is a favourable biological fluid among metabolomics researchers. Urine is sterile, easy-to-obtain in large volumes, largely free from interfering proteins or lipids and chemically complex [14]. As a biological waste material, urine typically contains metabolic breakdown products resulted from the various metabolism including from the metabolism of the infected cells. These changes would appear quickly, much faster than antibodies [15] as the metabolome is dynamic, able to change from second to second.

In the present study, we applied a combination of NMR spectrometry and multivariate chemometric analysis as a proof of principle to distinguish subjects with different healthy state.

2.0 EXPERIMENTAL

In our study, we have analyzed urine samples from healthy and dengue-infected subjects using proton NMR. The samples were collected from Penang General Hospital under non-controlled conditions to ensure the samples collected were fit with the actual practices. The urines were centrifuged at 1500 × g for 10 min to remove any cellular debris and the aliquots of the supernatants were kept deep frozen. A volume of 0.9 ml of thawed urine samples was added with 0.1 ml of potassium phosphate buffer (pH7.4) in D_2O containing 0.1 % of TSP and 0.6 ml of the mixture was transferred to 5 mm NMR tube. All one-dimensional ^1H NMR spectra of the urine samples was acquired on AVANCE III 500 MHz Bruker spectrometer with BBO broadband probe using TSP (δ 0.00ppm) as an internal standard and D_2O as the frequency lock at 300 K. The pulse sequence used included an excitation sculpting routine for the suppression of the water signal [16]. The resulting spectra were manually phased and baseline corrected and reduced to ASCII file using Chenomx software (version 5.1, Alberta, Canada). For each spectrum, the spectral region δ 0.52-10.00 was binned into regions of 0.04 ppm width giving a total of 238 integrated regions per NMR spectrum. The signals of δ 4.69 ppm- 4.97 ppm were excised from the analysis, mainly to eliminate variation in water suppression efficiency peaks. The averaged signals of binned ^1H NMR data from each sub-sample group were subjected to Principle Component Analysis (PCA), Partial Least Square Discriminant Analysis (PLS-DA) and Orthogonal PLS-DA (OPLS-DA), performed by SIMCA-P+ version 12.0.1.0 (Umetrics AB, Umeå, Sweden.

3.0 RESULTS AND DISCUSSION

3.1 ^1H NMR Spectra

Proton NMR spectra of urine samples from the subjects were recorded followed by chemometric multivariate analysis. These spectra were extracted from δ 0.00 to δ 10.00 and expanded to aliphatic region from δ 0.00 to δ 4.70 as most of the metabolites in this study were found within these regions (Figure 1). Collectively, normalized data matrix contained n=102 (healthy individuals, n=50 and dengue patients, n=52) were applied for data clustering in Principal Component Analysis (PCA). Consequently, out of fifty urine samples obtained from healthy individuals, only forty-three urine samples were valid to be used for further analysis. There were precipitations and turbidity observed in the other seven urine which might be caused by excessive cellular material present in the urine. This was the reason of the exclusion of seven urine samples to avoid any other unnoticed diseases that might mask the urine metabolome of the healthy persons.

3.2 Multivariate Analysis

The NMR spectral were binned into 0.04 ppm, exported and used in SIMCA-P+ as the variables for the multivariate analysis and series of pattern recognition analysis were applied. Prior to bulk analysis of the data, the Pareto scaling method was used to assess the optimum scaling method in NMR spectral.

Initially, PCA was performed to the spectral data to visualize inherent clustering between healthy control and dengue patient groups [17]. The score plot was obtained with the first two PCs presenting 15 and 22% variance, respectively (Figure 2). On the contrary, earlier data analysis by PCA of the ^1H NMR spectra from dengue infected EA.hy 926 cell line showed an inherent segregation between the two classes; dengue-infected and non-dengue infected cell line [8]. This might be due to the influences of extraneous factors in the subjects compared to the study done by Birungi's group which is done in a controlled environment.

Due to the inconsistency and poor separation obtained by unsupervised PCA, supervised analysis were then used, including Partial Least Squares Discriminant Analysis (PLS-DA) and Orthogonal Partial Least Square Discriminant Analysis (OPLS-DA) which can maximize differences among groups.

On the basis of the PLS-DA model, dengue patients and control subjects were discriminated with R^2X = 0.56, R^2Y = 0.81, and Q^2 = 0.69. The validation plot clearly demonstrated that the PLS-DA model was efficient, as the Q^2 regression line was < 0.05 and all permuted R^2 values on the left were < 0.3 (Figure 3).

The OPLS-DA was built to visualize the class specific segregation and to obtain the significant bins contributing to the variation across the classes, *i.e* dengue infected patient versus healthy individuals. The application of this chemometric model shows a clear separation between the samples of the two classes as shown in Figure 4.

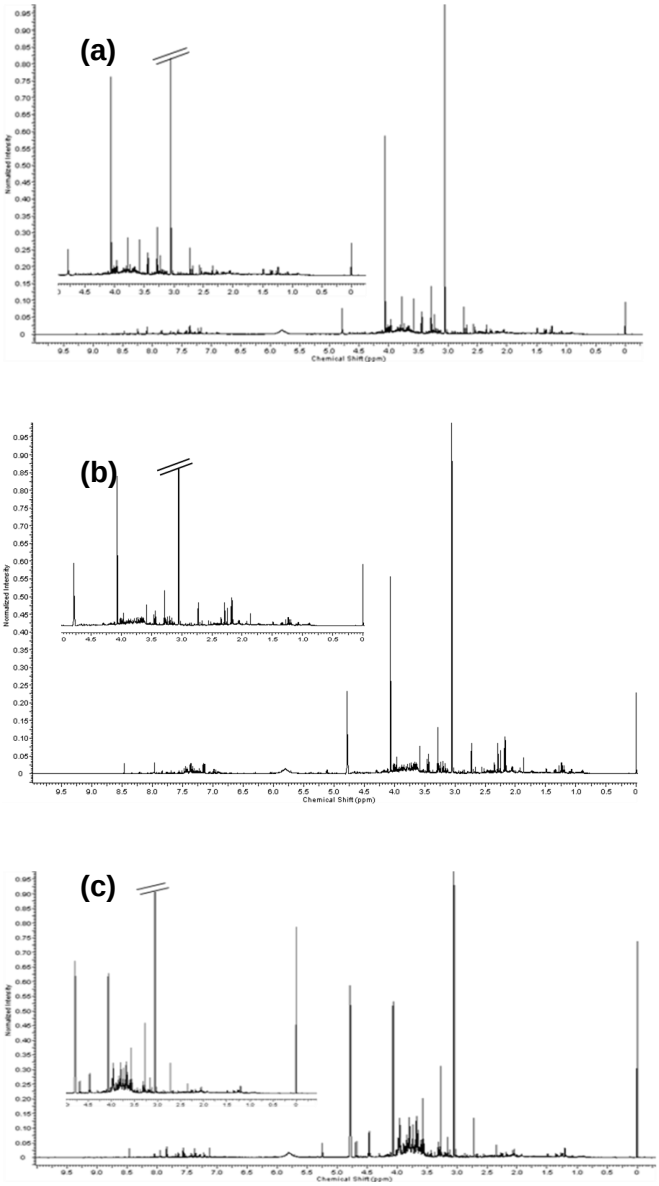


Figure 1 Aliphatic region of 1D ^{1}H NMR spectra of the urine samples (a) healthy individual (b) male patient infected with dengue disease (c) female patient infected with dengue disease

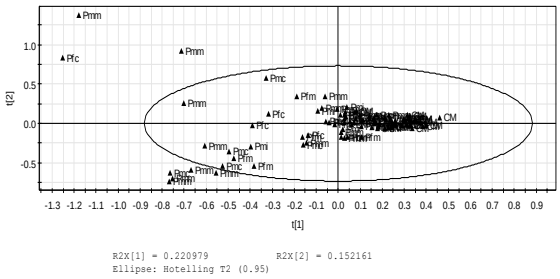


Figure 2 PCA score plot showing the variation between urinary metabolic profiles of dengue-infected individuals and healthy individuals. This PCA model was constructed from ^{1}H NMR of fifty-two patients with dengue infection and fourty-three healthy individuals, regardless their gender differences. R^2X for the model = 0.3731.

This model showed segregation of classes with R^2X= 0.9081 and Q^2 (cum) = 0.7880. Very narrow grouping are observed for the healthy (control) samples. Whereas, another group classify with dengue disease was spreading in the score plot highlighting their inhomogeneity of their metabolic profiles. Moreover, the spread is typical in human samples since there is enormous variation in genetic backgrounds as well as influences from the extraneous environment that may perturb the metabolome. In spite of all these factors, the Q^2 >0.5, was found to be significantly high, implicated a considerable differences in the urinary metabolic profiles of dengue patients and healthy individuals. In the OPLS-DA score plots, a significant biochemical distinction between the dengue cases and healthy control will be further evaluated.

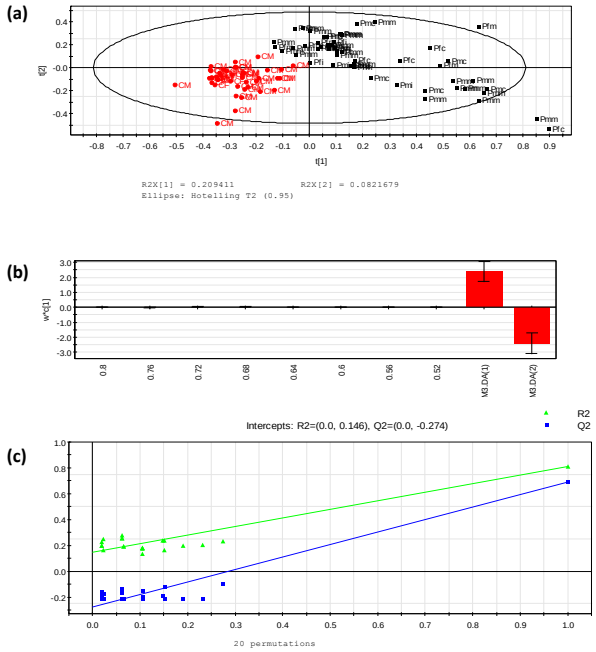


Figure 3 (a) PLS-DA scores plot showing the variation between urinary metabolic profiles of the dengue-infected patients and healthy individuals. This model was constructed from ^{1}H NMR data of urine from fifty-two patients with dengue infection and fourty-three healthy individuals, regardless their gender differences. Black square = dengue infected individuals and red circle = healthy individuals. The statistical parameters of the model were as follows R^2Y= 0.8119 and Q^2 (cum) = 0.920. The ellipse is a 95% Hotelling's T^2 ellipse. (b) PLS-DA loading column plot of patient-infected with dengue versus healthy control. (c) Plots of permutation tests of PLS-DA for urine profiles of dengue (black ■) and control (red dot ●) groups. It demonstrated clear metabolic difference between the two groups. Q^2 regression line was < 0.05 and permuted R^2 values on the left were < 0.3.

Based on the data documented in Table 1, the PCA model showed the lowest reproducibility (R^2) and predictivity (Q^2) values in comparison with PLS-DA and OPLS-DA. The R^2 and Q^2 values were

especially poor (<0.5) in the PCA model. OPLS-DA showed to be the best MVA type with Q^2 was 0.79, and the R^2 value was just nearly 1 in comparison with PLS-DA, implicated a considerable differences in the urinary metabolic profiles of dengue patients and healthy individuals.

Table 1 Statistical comparison of different MVA types

Scaling	MVA type	R^2X	R^2Y	Q2
Pareto	PCA	0.37	-	0.25
	PLS-DA	0.56	0.81	0.69
	OPLS-DA	0.55	0.91	0.79

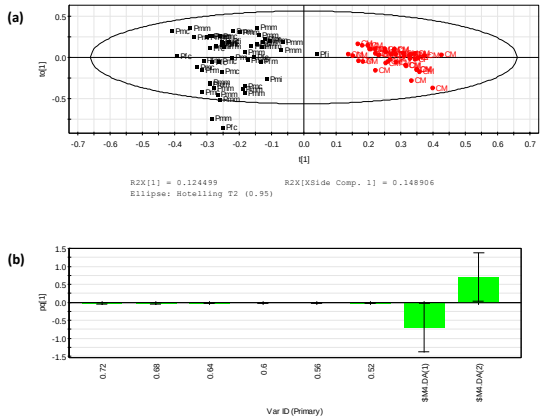


Figure 4 (a) OPLS-DA scores plot showing the variation between urinary metabolic profiles of the dengue-infected patients and healthy individuals. This model was constructed from ^{1}H NMR data of urine from fifty-two patients with dengue infection and fourty-three healthy individuals, regardless their gender differences. Black square = dengue infected individuals and red circle = healthy individuals. The statistical parameters of the model were as follows R^2Y= 0.9081 and Q^2 (cum) = 0.7880. The ellipse is a 95% Hotelling's T^2 ellipse. (b) OPLS-DA loading column plot of patient-infected with dengue versus healthy control.

4.0 CONCLUSION

The results from this study illustrated the successful application of ^{1}H NMR spectroscopy integrated with MVAs in discriminating the dengue patients and healthy control via metabolomics urinalysis. We suggest that using urine for dengue determination is a better prescreen to other forms of more invasive or uncomfortable screening. The data in our test set was small, but primary aim was to develop a good model to examine whether urinalysis with NMR spectroscopic integrated with chemometrics techniques could be used to discriminate subjects of interest. As the model we develop here was not a definitive model, a larger prospective cohort is needed to develop a more accurate model in the future.

Acknowledgement

We thank Bruker application scientists Dr. Teh Chin Hoe, Dr. Hicks and Dr. Fang Fang for their helpful discussions and comments. This work was supported by grants from AMDI Student Research Fund and RUT/1001/PKIMIA/855006. We are grateful for the Skim Latihan Akademik Bumiputra (SLAB) by Kementerian Pendidikan Malaysia to the author no. 1.

References

[1] Fang, R., Lo, E. and Lim, T. W. 1984. The 1982 Dengue Epidemic In Malaysia: Epidemiological, Serological And Virological Aspects. *Southeast Asian Journal of Tropic Medical Public Health*. 15: 51–57.

[2] Rudnick, A. Tan, E. E., Lucas, J. K. and Omar, M. B. 1965. Mosquito-Borne Haemorrhagic Fever In Malaya. *Britain Medical Journal*. 1: 1269–1272.

[3] Poovaneswari, S. 1993. Dengue Situation In Malaysia. *Malaysia Journal Of Pathology*. 15: 3-7.

[4] Ahmad Nizal, M. G., Rozita, H., Mazrura, S., Zainudin, M. A., Hidayatulfathi, O., Faridah, M. A., Noor Artika, I. and Er, A. C. 2012. Dengue Infections And Circulating Serotypes In Negeri Sembilan, Malaysia. *Malaysian Journal of Public Health Medicine*.12: 21-30.

[5] Muhammad Azami, N. A., Salleh, S. A. , Neoh, H. M., Syed Zakaria, S. Z. and Jamal, R. 2011. Dengue Epidemic In Malaysia: Not A Predominantly Urban Disease Anymore. *BMC Research Notes*. 4: 216.

[6] Fiehn, O. 2002. Metabolomics – The Link Between Genotypes And Phenotypes. *Plant Molecular Biology*. 48(1-2): 155-171.

[7] Shahfiza, N. *et al*. 2015. Metabolomics For Characterization Of Gender Differences In Patients Infected With Dengue Virus. *Asian Pacific Journal of Tropical Medicine*. http://dx.doi.org/10.1016/j.apjtm.2015.05.012.

[8] Birungi, G., Chen, S. M., Loy, B. P., Ng, M. L. and Li, S. F. Y. 2010. Metabolomics Approach For Investigation Of Effects Of Dengue Virus Infection Using The EA.hy926 Cell Line. *Journal of Proteome Research*. 9: 6523-6634.

[9] Shin, J. H., Yang, J. Y., Jeon, B. Y., Yoon, Y. J., Cho, S. N., Kang, Y. H., Ryu, D. H., Hwang, G. S. 2011. ^{1}H NMR-Based Metabolomic Profiling In Mice Infected With Mycobacterium Tuberculosis. *Journal of Proteome Research*. 10(5): 2238-2247.

[10] Slupsky, C. M. 2010. NMR-Based Analysis Of Metabolites In Urine Provides Rapid Diagnosis And Etiology Of Pneumonia. *Biomarkers Medicine*. 4(2): 195-197.

[11] Lakshmanan, V., Rhee, K. Y., Wang, W., Yu, Y., Khafizov, K., Fiser, A., Wu, P., Ndir, O., Mboup, S., Ndiaye, D. and Daily, J. P. 2012. Metabolomic Analysis Of Patient Plasma Yields Evidence Of Plant-Like A-Linolenic Acid Metabolism In *Plasmodium Falciparum*. *Journal of Infectious Disease*. 206(2): 238-248.

[12] Gebregiworgis, T and Powers, R. 2012. Application of NMR Metabolomics To Search For Human Disease Biomarkers. *Combinatorial Chemistry & High Throughput Screening*. 15: 595-610.

[13] Goodacre, R., Broadhurst, D., Smilde, A. K., Kristal, B. S., Baker, J. D., Beger, R., Bessant, C., Connor, S., Capuani, G., Craig, A., Ebbels, T., Kell, D. B., Manetti, C., Newton, J., Paternostro, G., Somorjai, R., Sjöström, M., Trygg, J. and Wulfert, F. 2007. Proposed Minimum Reporting Standards For Data Analysis In Metabolomics. *Metabolomics*. 3: 231-241.

[14] Bouatra, S., Aziat, F., Mandal, R., Guo, A. C., Wilson, M. R. *et al.* 2013. The Human Urine Metabolome. *PLoS ONE*. 8: e73076.

[15] Bernardo, C. D. N., Nogueira, M. C. O., D. Aquino, É. P., Schmidtke, S., D. Azeredo, E. L. *et al*. 2012. With NMR Towards New Diagnostic Methods For Dengue. *Journal of Analytical Bioanalysis Techniques*. 3: 140.
[16] Hicks, J., Sivakolundu, S. and Colson, K. 2009. *Metabolomics Guide User Manual Version 005*. Bruker Biospin. Billerica, USA.
[17] Zhou, A., Ni, J., Xu, Z., Wang, Lu, S., Sha, W., Karakousis, P. C. and Yao, Y. F. 2013. Application Of ^{1}H NMR Spectroscopy-Based Metabolomics To Sera Of Tuberculosis Patients. *Journal of Proteome Research*. 12: 4642-4649