Academic Sciences International Journal of Pharmacy and Pharmaceutical Sciences

ISSN- 0975-1491 Vol 4, Issue 4, 2012

Research Article

PRINCIPAL COMPONENT AND PARTIAL LEAST SQUARE DISCRIMINANT BASED ANALYSIS OF METHANOL EXTRACTS OF BARK AND RE-GENERATED BARK OF *SARACA ASOCA*

ANJUM GAHLAUT[1], PURVA TANEJA[1], AMEY SHIROLKAR[2], AMIT NALE[2], VIKAS HOODA[1] AND RAJESH DABUR[3]*

[1]Centre for Biotechnology, Maharishi Dayanand University, Rohtak, Haryana-12400, [2]National Research Institute of Basic Ayurvedic Sciences, (Central Council for Research in Ayurvedic Sciences), Nehru Garden, Kothrud, Pune-411038, [3]Deparment of Biochemistry, Maharishi Dayanand University, Rohtak, Haryana-124001. Email: rajeshdabur@yahoo.com.

Received: 1 Aug 2012, Revised and Accepted: 06 Sep 2012

ABSTRACT

Saraca asoca is used worldwide to treat various kinds of human disorders. Increasing demand and destructive harvesting of bark has led to the depletion of this valuable medicinal tree. Scarcity of this plant leads toward the use of various substitutes/adulterants. Hence, it is very important to develop methods for the authentication of bark samples and to compare the contents of bark and regenerated bark to define their medicinal properties. Therefore, the metabolomic profiles of methanol extracts of harvested bark samples of *S. asoca* were examined by liquid chromatography interfaced with quadruple time of flight tendom mass spectrometer. An automated feature extraction algorithm was employed for processing of MS/MS data. Principal component analysis and partial least square discriminant analysis (PLS-DA) were used to explore the structure of processed data and to prepare classification models. The accuracy of developed PLS-DA model for sample validation was found to be 100%. A total number of 206 molecular features (MFs) were recorded to be differentially expressed across samples at a threshold of 5000 cpu. The PLS-DA models can be used as analytical model for authentication and showed content variations in the bark samples of *S. asoca*. Hence it important to assess the medicinal properties of bark and regenerated bark to ensure the proper collection of bark samples.

Keywords: Medicinal Plant, *Saraca asoca*, Quadrupole Time of Flight Mass Spectrometry, Phyto-chemistry, Multivariate data analysis.

INTRODUCTION

S. asoca (Roxb.), De. Wild (Indian name; Ashoka) belongs to family *Caesalpinaceae*, mentioned in the famous Indian treatise *Charaka Samhita* (100 A.D.). The plant has been recommended in formulations for the various gynaecological disorders as anodynes. In another treatise i.e. Bhavprakasha Nighantu, this plant has been referred as a uterine tonic for regularizing the menstrual disorders. Its bark has a stimulating effect on endometrium & ovarian tissues and is useful in menorrhagia during uterine fibroids [1]. Scanty literature is available on the antimicrobial effect of *S. asoca* extracts [2, 3].

The principal constituents of *S. asoca* includes steroids and calcium salts. Bark is the most utilized plant part and is used as a major constituent for the preparation of various formulations. Leaves, seeds, pods and flower are reported to contain number of compounds including catechin, epicatechin, epigallocathechin, gallic acid which are well-known flavonoids, and utilized as substitute of bark for the symptomatic treatment of several gastrointestinal, respiratory and vascular diseases [4-11]. Since the medicinal properties of *S. asoca* are being commercially exploited throughout the world to treat gynaecological disorders and diabetes [12]. To strengthen faith in herbal drugs, it is necessary to develop discriminative analytical models for the authentication and quality control of raw as well as processed herbal drugs and to identify substitutes/adulterants. Moreover, the medicine prepared from the plant should have quality standards parameters to get confidence in the plant based drugs.

Being a complex mixture of chemical entities and unknown mechanism of action, quality control of herbal drugs is a difficult task. Factors like collection time of plant materials, geographical variations and different processing methods, leads to the chemical variations in the herbal drugs putting another challenge. Number of techniques has been reported to monitor the quality parameters includes thin layer chromatography, high performance thin layer chromatography, gas chromatography, high performance liquid chromatography and mass spectrometry [13-15]. Q-TOFMS is excellent technique to analyze multi-components in the complex herbal extracts due to accurate mass measurement, high resolution and ion separation due to Time of Flight (TOF) [14]. Rapid data mining procedures and aligning algorithms tools been used to process huge raw data generated from metabolome analyses [16-18]. These processed data have been used successfully in various pharmaco-physiological studies such as disease diagnostics, drug discovery [19] and human nutritional science [20, 21].

In the present study, high performance liquid chromatography (HPLC) coupled with quadrupole time-of-flight mass spectrometry (Q-TOFMS) has been used to generate auto MS/MS data of *S. asoca* bark samples. Auto MS/MS data was processed using Mass Hunter and Mass Profiler Professional softwares for the extraction of input variables, alignment of retention times/mass-to-charge ratios, principal component analysis (PCA) and partial least square discriminant analysis (PLS-DA), and an informative and predictive model was prepared.

MATERIALS AND METHODS

Chemicals

Standard compounds lidocaine, D-camphor, 5-7-isoflavone and solvents i.e. acetonitrile, tri-fluroacetic acid and water of LCMS grade were purchased from Sigma-Aldrich.

Plant Material

Bark and regenerated bark of *S. asoca* were collected in February, 2012 from Botanical Garden of National Research Institute of Basic Ayurvedic Sciences, CCRAS, (Dept of AYUSH), Nehru Garden, Kothrud, Pune. The collected plant materials were identified and voucher specimens (No. 207) kept at the medicinal plant museum of the Institute.

Extraction

Fresh plant materials (20 gm each) were extracted overnight (at 25 and 70^{o}C) with water and subsequently with methanol (1:1 w/v). Extraction steps were repeated three times to ensure complete recovery of metabolites. Samples were filtered through 0.22 µ filters (Hi-media), lyophilized using a lyophilizer (Freezone 4.5 Labconco, CA, USA) and stored at -80^{o}C till further use. The methanolic plant extracts were reconstituted in LC/MS grade acetonitrile (5.0 mg/ml) for further analytical study.

Auto Q-TOF MS/MS

MS/MS experiments were performed on Agilent 1290 Infinity Series RRLC-MS interfaced with an Agilent 6538 Accurate-Mass Q-TOFMS. Sample volume of 20 µl was injected by auto-sampler into ZORBAX 300SB reversed phase column (C18, 4.5 mm x 250 mm, 5 µ particle size). The column temperature was maintained at 40^{o}C. Mobile phase comprising of solvent A (water containing 0.1 % formic acid)

and solvent B (acetonitrile containing 0.1 % formic acid) were used in gradient mode {concentration/time (%/min) 5% / 8; 10% / 15; 45% /22; 65% /30; 90% /35; 5% /40}. Mobile phase flow of 0.4 ml/min was maintained. Q-TOFMS was operated in positive ion polarity mode and extended dynamic range (1700 m/z, 2GHz) with following parameters: gas temperature 350°C, nebulizer 50 Psi, gas flow 11 L/Min, capillary voltage 3500V, nozzle 500 V, skimmer voltage 65 V, octapole RF 250 V, octapole DC1 48 V and fragmentor voltage 175V. The instrument was calibrated and tuned as recommended by the manufacturer to get accuracy less than 5 ppm. Accurate MS/MS spectra were acquired in the range 100-1100 m/z with acquisition rate 3 spectra s^{-1}. To assure the mass accuracy of recorded data, standards of lidocaine (234.3m/z) and 5, 7-isoflavone (284.3 m/z) were infused with samples along with continuous internal calibration with the use of signals at a range of m/z 121.05 to m/z 922.0098 (as per instrument standards).

Data Processing and Analysis

Initial processing of LC-Q-TOF//MS/MS raw data i.e. baseline correction, noise reduction, and removal of background contaminants, to extract MS/MS and MS data, using Mass Hunter Qualitative Software, Version 3.1 (Agilent Technologies). The molecular features (MF) of the spectra were extracted using molecular feature extraction (MFE) tool of Mass Hunter software. Mass Hunter was used to generate molecular formulas, searched in a specific generated in house library. The ions with identical elution profile and related m/z value were extracted as single molecular feature (MF), with in the algorithm employed for full MS/MS and MS data. MFs were characterized by retention time, intensity in the apex chromatographic peak and accurate mass. Various intensity thresholds i.e. 1000, 5000 and 10000 counts per second (cpu) were tested for MFs extraction in the retention time range from 1 to 45 minutes. MS/MS profiles of blank samples (LCMS grade water) were subtracted from the samples to remove the background. Background subtracted data were converted into compound exchange (cef) file for further use in Mass Profiler Professional (MPP). MPP (Agilent, version B 02.02) was used for statistical evaluation of technical reproducibility and comparison of bark samples. In MPP, the retention time and m/z alignment across the sample sets was performed using a tolerance window of 0.2 min and 20 mDa. The MFs were reduced stepwise based on frequency of occurrence, abundance of respective MFs in classes and one way analysis variance (ANOVA). The PCA and PLS-DA were performed using MPP software, working both under the correlation and covariance methods. Oneway ANOVA with Benjamini-Hocchberg correction was used to remove common mass features shared by all of the samples that could produce a identical pattern of non-informative signals [19]. A probability level of $p<0.05$ was applied to reduce nonsignificant MFs. Compounds that satisfied fold change cut-off 2.0 in atleast one condition pair were selected for further analysis and differentiation. The extracted MFs were mean centred and logrithmic transformed in order to reduce the relative large differences in the respective aduntant MFs.

RESULTS

Method Development

In the current study, non targeted analysis of *S. asoca* methanol extracts of bark and regenerated bark were performed employing reverse phase chromatographic system without considering any specific group of metabolites. Therefore, LC-Q-TOFMS conditions were optimized in order to obtain maximum number of metabolites. The concentration of sample solutions were optimized to 5.0 mg/ml. Aqueous mobile containing 0.1% formic acid was used with acetonitrile for the elution of metabolites. No pretreatment was given to the sample to avoid discrimination and to get maximum number of metabolites. Q-ToF provides accurate MS/MS spectra due to internal mass calibration during acquisition and mass drift compensation. In the present study, mass accuracy less than 5 ppm was obtained with internal and external standards. Figure 1 represents the total ion current (TIC) of bark and regenerated bark because. The solvent system was optimized to elute all the detectable metabolites within 35 min. The elution was continued for additional 10 min to ensure the complete removal of sample from the column and to get reproducible results. Retention time (RT) variability across the samples was found to be 4 sec and a relative standard deviation of less than 2% was observed. Maximum variations in TIC were observed in RT window of 35.5 to 39 minutes where regenerated bark showed maximum differential peaks as compared to bark.

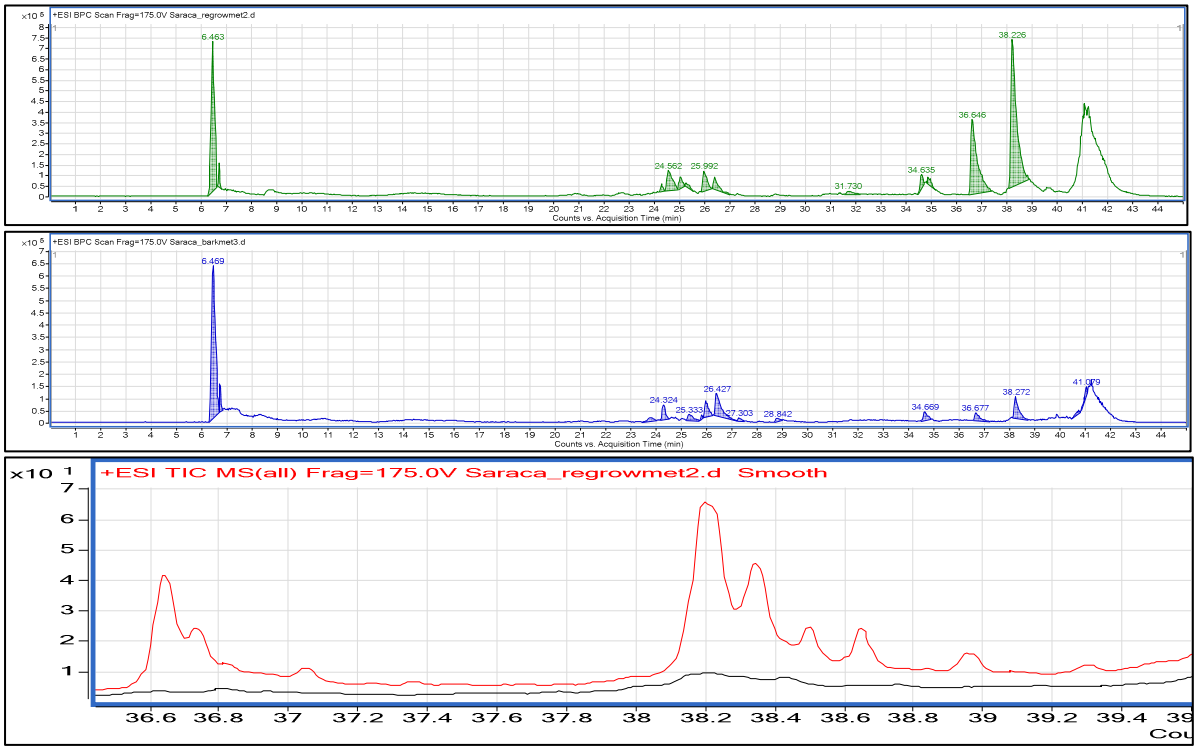

Fig. 1: It shows TIC chromatograms of (A) Methanol of bark with blank and (B) Methanol of regenerated bark with blank in positive ion mode.

Data mining

Auto MS/MS data was processed by using qualitative Mass Hunter (B.04.00 Version) and MPP. A total number of 2407 metabolites at 5000 cps threshold were extracted to avoid false positives. Molecular formulae of the compounds were generated using the molecular formula generator in the same software. An in-house generated library of compounds specific for *S. asoca* was used to identify the metabolites present in the different extracts. Considering the complexity of the data obtained by LC–MS measurements (Table 1), algorithm enabling automated extraction of ions corresponding to compounds present in the different extracts was employed. Positive ions with accurate m/z values and with a difference corresponding to adduct isotopes or multiply charged species were merged into MFs as a single variable. MFs absent in at least 75% samples of one group were removed to reduce the dimensionality of the data sets prior to PCA and PLS-DA. Furthermore, MFs were filtered on the basis of p-values ($p < 0.05$) calculated for each MF by one-way ANOVA. This ensures the filtration of MFs which differed in the respective varieties with statistical significance (95% in this particular case). In the final step, MFs filtration was performed using fold change (FC) analysis to find MFs with high abundance ratios among the samples (Table 1).

A 11.6 fold decrease in MFs was observed after selecting the metabolite with FC≥2 and high abundance filtration (Table 1). Final groups of MFs were re-extracted from the raw data files of all the samples using recursive feature extraction. Table 2 shows minor effect on the number of MFs by applying different fold change values confirm the consistency of data.

Table 1: shows molecular feature number extracted in +ESI mode, at various intensity threshold settings and applied filtering steps.

Counts	Initial Entities Aligned	Filter by Frequency 75%	One-way ANOVA (P < 0.05)	Filtering by fold change (≥2)	Filtered by Abundance
1000	2413	740	379	370	370
5000	2402	459	226	212	206
10000	2233	414	165	160	155

Table 2: shows tendency of molecular features with respect to 'P' values applied in the T-test with Benjamini – Hochberg correction and fold change.

	P All	P<0.05	P<0.02	P<0.01	P<0.005	P<0.001
FC All	206	153	153	153	153	153
FC>1.1	171	147	147	147	147	147
FC>1.5	122	121	121	121	121	121
FC>2.0	122	121	121	121	121	121
FC>3.0	121	121	121	121	121	121

Chemo-metric Analysis

PCA was performed to reduce data dimensionality by performing covariance analysis of bark and regenerated bark samples. All metabolites that could easily be sorted out from extracts under study are given in Figure 2. It shows more variability among MFs from bark and regenerated bark. For the chemo-metric analysis, PCA results were used with 5000 cps threshold data because the results with 1000, 5000 and 10000 cps thresholds were almost equivalent (Table 2). Demonstration of the discrimination potential offered by the above data, PLS-DA a widely used supervised pattern recognition method capable of sample class prediction, was used to construct and validate a statistical model for sample classification. The results of sample classification (Table 3) are presented in terms of recognition abilities, representing the percentage of the samples correctly classified during model training and cross-validation. The recognition and prediction ability of the model was found to be 100%.

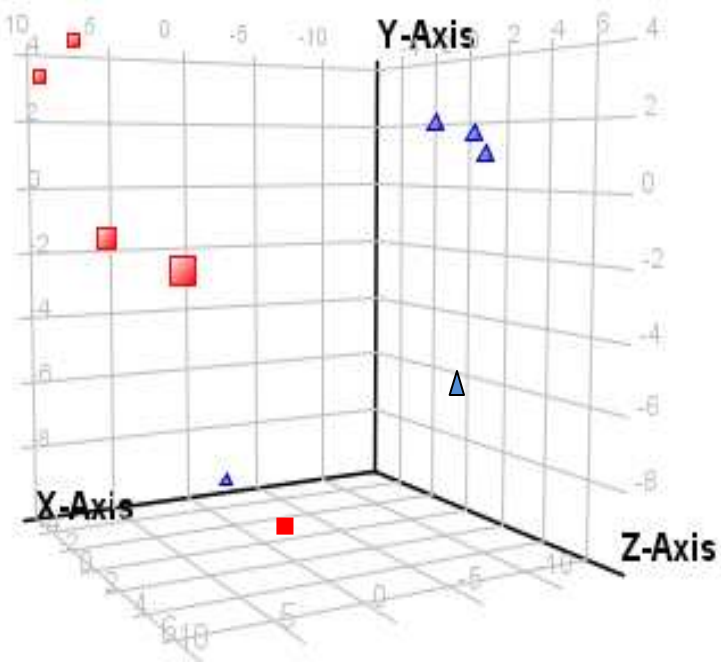

Fig. 2: It shows PCA plots shows spatial relation between the metabolites extracted from bark ( ) and regenerated barks ( ) of *S. asoca*.

Thus PLS-DA model provides excellent separation among the sample varieties (Table 3). Though PLS-DA model itself is sufficient for sample discrimination and authentication of any sample, elemental formulae were generated with the idea to find plant part specific biomarkers. Mass Hunter software equipped with molecular formula generator and library search features was used in formulae generation and to search in a small in-house library generated specifically for the secondary metabolites of the *S. asoca* (Table 4). Table 4 shows the presence of uracil and two other compounds specifically present in the regenerated bark.

Table 3: shows sample classification using PLS-DA model of *S. asoca*

	Bark Methanol Extract	Re-Generated Bark Methanol Extract	Accuracy %
Bark Methanol Extract	4	0	100
Re-Generated Bark Methanol Extract	0	4	100
Overall Accuracy			100
Model Cross Validation			
Bark Methanol Extract	4	0	100
Re-Generated Bark Methanol Extract	0	4	100
Overall Prediction Ability			100

Table 4: shows compounds Indentified from the Methanol extracts of Bark.

S. No.	Name	Abundance		Formula	Mass	RT
		Bark	Regenerated Bark			
1.	(Z)-2-Methylbutanal oxime	Present	Present	C5H11N1O1	101.0846	42.464
2.	1,2-Bis-O-sinapoyl-beta-D-glucoside	-	Present	C28H32O14	592.1787	38.435
3.	16-Methoxytabersonine	Present	-	C22H27N2O3	367.2008	6.574
4.	3-Butenylglucosinolate	-	Present	C11H18N1O9S2	372.0432	38.274
5.	3-Glc, 28-ara-rha-xyl medicagenic acid	Present	-	C52H82O23	1074.534	38.545
6.	Acetoacetyl-CoA	Present	-	C25H36N7O18P3S1	847.1064	25.757
7.	CDP	Present	Present	C9H12N3O11P2	399.9956	41.152
8.	Dihydroxyferuloyl-sinapoyl spermidine	Present	-	C38H45N3O12	735.2989	6.469
9.	D-myo-inositol (1,2,5,6) tetrakisphosphate	Present	-	C6H9O18P4	492.8748	41.152
10.	Indolylmethyl glucosinolate aglycone	Present	Present	C10H8N2O4S2	283.9913	38.11
11.	Nonaprenyl-4-hydroxybenzoate	Present	-	C52H77O3	749.5837	41.832
12.	OPC8-CoA	Present	-	C39H60N7O18P3S1	1039.29	40.692
13.	o-Succinylbenzoyl-CoA	Present	-	C32H39N7O20P3S1	966.1182	22.935
14.	Phenylethylamine	Present	Present	C8H12N1	122.0965	22.902
15.	Uracil	-	Present	C4H4N2O2	112.0276	8.937
16.	Violaxanthin	Present	Present	C40H56O4	600.4178	24.237
17.	Caffeoyl-CoA	Present	Present	C30H42N7O19P3S	929.1509	30.609
18.	Catechol	Present	Present	C6H6O2	110.0366	16.241
19.	Feruloyl-CoA	Present	Present	C31H44N7O19P3S	943.1583	26.489
20.	Lyoniside	Present	Present	C35H60O6	576.4411	23.587
21.	Neohesperidin	Present	-	C28H34O15	610.1904	7.673
22.	Nudiposide	Present	-	C27H36O12	552.2204	38.414

DISCUSSION

Herbal medicines are very complex by nature, therefore, metabolites were separated using reverse phase chromatography suitable for the separation of polar compounds. Fold change analysis of abundant metabolites showed high level of variations in the MFs of bark and regenerated bark (Fig 1 and 2). The variations in the described samples are clearly visible after the careful inspection of their TICs (Fig 1). This may be due to up regulated levels of metabolite required to boost up the healing process and anti-infective metabolites during bark regeneration. Flavanoids and cutin protect the plant from various infections and might be elevated in response to infections and to avoid moisture loss after the bark damage[22]. Automatic data mining and data processing of crude extracts was used to identify the metabolites present in the samples (Table 4). Presence of higher amounts of uracil in the regenerated bark may attribute to the active processing of RNA required for protein synthesis. The study explored the potential of LC-Q-TOFMS for accurate mass measurement of unknown compounds and to determine authenticity of variety of sample from the same origin [11]. Eigen vectors and Eigen values relevant to data using covariance matrix were explored by PCA. Eigen values give quantitative assessment of the data represented by a component. Eigen values also represent the level of covariance as percentage of total variance. Therefore, PCA reduced the data dimensionality and covariance was measured between x (bark) and y (regenerated bark), y and z, x and z. Fig 2 shows the percent covariance of the sample which is clearly visible. A partial least square discriminant model developed from highly variable data sets was used to identify bark samples. PLS-DA finds a linear regression model by projecting the prediction variables and observable variables to new space. It finds fundamental relations between two matrices. In this study, PLS-DA based predictive model showed 100% accuracy; therefore, it can be used for authentication and identification of raw as well as processed drugs. At the same time it shows the high variation in the samples. The issue of quality control and standardization of herbal drugs are the major challenges before the pharmaceutical sector, which can be addressed with the use of PLS-DA and PCA models to monitor the variation in the constituents and use of substitutes/adulterants.

In addition, aaccurate MS/MS spectra of marker compounds and their quantization (metabolites) will provide a high level of confidence for the identification process of raw herbal drugs, which is not possible with other technologies [19]. Furthermore, identification of metabolites by comparing the MS/MS fragment pattern from the data sources such as pubchem or in house/national repositories will be helpful to study metabolomics and role of different metabolite. The present study shows the variations in the contents of bark and regenerated (immature) bark; thus, their

medicinal properties might be different. As contents of bark and regenerated bark are very different one should avoid the collection of immature bark or growing bark which might lead toward different medicinal properties.

CONCLUSION

PCA and PLS-DA models generated by LC-Q-TOFMS data were used successfully for metabolome based discrimination and identification of the samples used in the study. This demonstrates the potential of the models used in the study to discriminate the samples without using standard biomarkers. However if standard markers can be quantified and standardised will be an additional advantage in the quality control. Therefore, the application of LC-Q-ToFMS in authentication of the medicinal plants and drugs prepared from them by developing partial least square discriminant models seems to be very promising. It is also advisable to avoid collection of immature or growing bark as it has different constitution which may lead to different medicinal properties.

REFERENCES:

1. Satyavati GV, Prasad DN, Sen SP, Das PK. Further studies on the uterine activity of Saraca indica Linn. Ind J Med 1970; 58: 947-960.
2. Sainath RS, Prathiba J, Malathi R. Antimicrobial properties of the stem bark of *Saraca indica* (Caesalpiniaceae). Eur Rev Med and Pharmacol Sci 2009; 13: 371-374.
3. Singh A, Singh S, Sarma B, Singh U, Srivastava R, Singh K. Phenolic acid analysis and biological activity of methanolic extracts of some medicinal plants against some phytopathogenic fungi. The Internet J Alternative Med 2009; 6: 2-11.
4. Maruthappan V, Shree SK. Antiulcer activity of aqueous suspension of *Saraca indica* flower against gastric ulcers in albino rats. J of Pharm Res 2010; 3: 17-20.
5. Prerana S, Krishnamoorthy M, Vijayanarayana K. Antidepressant activity of bark of *Saraca indica* linn. Asian J Chem 2008; 20: 1075-1080.
6. Preethi F, Krishnakumar K Anti-inflamatory activity of the bark of *Saraca indica Linn*. Pharmacologyonline 2011; 2: 657-662.
7. Pal CS, Maiti PA, Chatterjee PB. Antibacterial activity of flower and flower buds of *Saraca indica.* Indian J Med Res 1985; 82:188-189.
8. Pradhan P, Joseph L, Gupta V, Chulet R, Arya H, Verma R, Bajpai A. A review on Ashoka. J Chem and Pharma Res 2009; 1: 62-71.
9. Sandhu KS, Khatun A, Phattanawasin P. Lignan glycosides and flavonoids from *Saraca asoca* with antioxidant activity. J Nat Med 2007; 61: 480-482.
10. Jayita S, Taniya M, Kamala G, Sumona M. Phytoconstituents and HPTCL analysis in *Saraca asoca* (ROXB.) wilde. Inter J Pharm And Pharmaceutical Sci 2012; 4:96-99.
11. Saha J, Mitra T, Gupta K, Mukherjee S. Phytoconstituents and HPTLC Analysis in *Saraca asoca* (Roxb.)Wilde. *Int J Pharm Pharm Sci*. 2012; 4(1): 96-99.
12. Shivashankar M, Mani D. A Brief Overview of Diabetes. *Int J Pharm Pharm Sci,* 2011; 3(4): 22-27.
13. Liang YZ, Xie P, Chan K. Quality control of herbal medicines. J Chromatogr B Analyt Technol Biomed Life Sci 2004; 812: 53-70.
14. Zeng ZD, Liang YZ, Chau FT, Chen S, Daniel MK, Chan CO. Mass spectral profiling: an effective tool for quality control of herbal medicines. Anal Chim Acta 2007; 604: 89-98.
15. Zhou JL, Qi LW, Li P Herbal medicine analysis by liquid chromatography/time-of-flight mass spectrometry. J Chromatogr A 2009; 1216: 7582-94.
16. Duran AL, Yang J, Wang LJ, Sumner LW. Metabolomics spectral formatting, alignment and conversion tools (MSFACTs). Bioinformatics 2003; 19: 2283-93.
17. Stoyanovaa R, Nicholls AW, Nicholson JK, Lindon JC, Brown TR. Metabolomics by numbers - acquiring and understanding global metabolite data. J Magn Reson 2004; 1(70): 329.
18. Goodacre R, Vaidyanathan S, Dunn WB, Harrigan G, Kell DB. Practical metabolomics in drug discovery. Trends Biotechnol 2004; 22:245.
19. Keith MW, Uehara T, Khin TM, Yoshiaki S, Yoshiya O. Expert Opinion on Drug Discovery. Informa healthcare 2010; 5:249-263.
20. Cevallos JM, Reyes-De-Corcuera JI, Etxeberria E, Danyluk MD, Rodrick GE. Metabolomic analysis in food science: a review. Trends in Food Science & Technology 2009; 20: 557-556.
21. Vaclavik L, O. Lacina, J. Hajslova, Zweigenbaum J. Classification of Tempranillo wines according to geographic origin: combination of mass spectrometry based electronic nose and chemometrics. Anal Chim Acta 2010; 685(1):45-51.
22. Iriti M, Faoro F. Chemical Diversity and Defence Metabolism: How Plants Cope with Pathogens and Ozone Pollution. Int. J. Mol. Sci 2009; 10:3371-3399.