Talanta 87 (2011) 53–59

Contents lists available at SciVerse ScienceDirect

Talanta

journal homepage: www.elsevier.com/locate/talanta

# Laser-induced breakdown spectroscopy-based investigation and classification of pharmaceutical tablets using multivariate chemometric analysis

Ashwin Kumar Myakalwar[a], S. Sreedhar[a], Ishan Barman[b], Narahara Chari Dingari[b], S. Venugopal Rao[a], P. Prem Kiran[a], Surya P. Tewari[a], G. Manoj Kumar[a,*]

[a] Advanced Centre of Research in High Energy Materials (ACRHEM), South Campus, University of Hyderabad, Prof C R Rao Road, Central University Campus PO, Gachibowli, Hyderabad, Andhra Pradesh, 500046, India
[b] G.R. Harrison Spectroscopy Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

ARTICLE INFO

*Article history:*
Received 1 July 2011
Received in revised form 2 September 2011
Accepted 23 September 2011
Available online 1 October 2011

*Keywords:*
Laser induced breakdown spectroscopy
PCA
SIMCA
Classification
Pharmaceutical tablets

ABSTRACT

We report the effectiveness of laser-induced breakdown spectroscopy (LIBS) in probing the content of pharmaceutical tablets and also investigate its feasibility for routine classification. This method is particularly beneficial in applications where its exquisite chemical specificity and suitability for remote and on site characterization significantly improves the speed and accuracy of quality control and assurance process. Our experiments reveal that in addition to the presence of carbon, hydrogen, nitrogen and oxygen, which can be primarily attributed to the active pharmaceutical ingredients, specific inorganic atoms were also present in all the tablets. Initial attempts at classification by a ratiometric approach using oxygen (~777 nm) to nitrogen (742.36 nm, 744.23 nm and 746.83 nm) compositional values yielded an optimal value at 746.83 nm with the least relative standard deviation but nevertheless failed to provide an acceptable classification. To overcome this bottleneck in the detection process, two chemometric algorithms, i.e. principal component analysis (PCA) and soft independent modeling of class analogy (SIMCA), were implemented to exploit the multivariate nature of the LIBS data demonstrating that LIBS has the potential to differentiate and discriminate among pharmaceutical tablets. We report excellent prospective classification accuracy using supervised classification via the SIMCA algorithm, demonstrating its potential for future applications in process analytical technology, especially for fast on-line process control monitoring applications in the pharmaceutical industry.



## 1. Introduction

Laser induced breakdown spectroscopy (LIBS) is an atomic emission based technique where a pulsed laser light is focused on to the sample producing hot plasma [1–3]. The resultant plasma plume consists of electrons and ions of the sample constituents, which emit radiation as the plasma cools down. Typically, several peaks are observed in the LIBS spectra in the 200–1000 nm spectral region. The emission wavelength is characteristic of atoms/ions present in the plasma and area under the specific curve(s) (i.e. intensity of the specific spectral features) is proportional to concentration(s) of the analyte(s) contributing to it. In some cases molecular emissions are also observed [4,5]. The standard atomic and ionic emission wavelengths are tabulated in NIST database [6]. Although this technique was discovered a few decades back, it has received a great deal of attention from researchers in the past few years because of the many advantages it offers as a spectroscopic analytical tool. One of the most attractive features of LIBS is the lack of substantial sample preparation requirements, which considerably reduces the typical time needed for comparable detection processes. Samples in any form, be it solid, liquid or gas, can be used for the LIBS experiments. Solid samples can be used in their original form directly for the experiment. Furthermore, the total time needed to perform an experiment and acquire the results is very small. Moreover, with the availability of high power lasers and sensitive detectors it is now possible to perform LIBS experiments on samples located at few meters of distance (i.e. stand-off detection) [7]. These attributes make this technique suitable for on-site and remote characterization, which is of significant interest in a variety of industrial applications.

Particularly, on-site LIBS based detection, combined with chemometrics, can expedite quality assurance of the important pharmaceutical products. All pharmaceutical tablets typically comprise a mixture of organic active substances (also known as active pharmaceutical ingredient (API)) and excipients, which may have various inorganic elements present as additives or impurities. Importantly, LIBS can detect both organic and inorganic part of the

* Corresponding author. Tel.: +91 40 23138837; fax: +91 40 23012800.
*E-mail addresses:* manojsp@uohyd.ernet.in, my.hunn.manoj@gmail.com (G. Manoj Kumar).

0039-9140/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2011.09.040

**Table 1**
Details of pharmaceutical samples investigated in this study.

| Sl. No | Sample | Formula of the primary component | Ingredients | No. of spectra |
|---|---|---|---|---|
| 1 | Brufen | $C_{13}H_8O_2$ | Ibuprofen, Erythosine | 15 |
| 2 | Brufen-coated | $C_{13}H_8O_2$ | Ibuprofen, Erythosine, titanium dioxide | 15 |
| 3 | Glucosamine | $C_6H_{13}NO_5$ | Glucosamine sulphate, chondroitinsuplhate | 10 |
| 4 | Glucosamine-coated | $C_6H_{13}NO_5$ | Glucosamine sulphate, chondroitinsuplhate, titanium dioxide | 10 |
| 5 | Vitamin C | $C_6H_8O_6$ | Sodium ascorbate, ascorbic acid | 15 |
| 6 | Paracetamol | $C_8H_9NO_2$ | Paracetamol | 20 |

sample in a single step, which is not possible in other techniques such as ICP-OES or ICP-MS. The knowledge of organic part of the sample can be valuable in the drug discovery stage as well as in the production process of the drugs. Additionally, in cases of counterfeit pharmaceutical drugs, LIBS can be potentially used for extracting formulation information and, importantly, for screening them.

In this context, despite LIBS investigations of several materials have been reported in the literature, detection and compositional analysis of pharmaceutical samples has received much less attention [8–10]. In this paper, we investigate the feasibility of LIBS for routine pharmaceutical tablet investigation for compositional information and discrimination among tablets procured over the counter from local pharmacy. Herein, we first report our observations of the LIBS-based elemental analysis on common pharmaceutical tablets. Further, we attempted to classify these tablets into their corresponding functional groups based on a ratiometric approach. Finally, we incorporated multivariate chemometric analysis to exploit the multi-channel spectral dataset. Given the intrinsic advantages of these multivariate techniques in improving prediction accuracy, we anticipate that the combination of LIBS and chemometrics will be a powerful tool in future for rapid on-line process control monitoring (for example, see Mowery et al. [11] and Barrette and Turmel [12]) and counterfeit tablet detection [13] particularly in the pharmaceutical industry.

## 2. Brief description of multivariate classifiers

As is well-known in the field of analytical chemistry and process technology, simple ratiometric approaches, where the relative intensity of spectral lines is monitored, often do not provide acceptable classification results. To overcome this difficulty and to utilize the full extent of multi-channel data available, multivariate chemometric algorithms can be employed. For our application of discrimination of over-the-counter pharmaceutical samples using LIBS, we employ an unsupervised (principal component analysis, PCA) as well as a related supervised (soft independent modeling of class analogy, SIMCA) classification method [14–16]. Previous investigators have also employed these [17–21] as well as other chemometric methods, including partial least square discriminant analysis [22,23], neural networks [17,24] and discriminant function analysis [25,26], for processing LIBS data acquired from a diverse spectrum of samples ranging from high energy materials to rocks. The working concepts and the primary advantages of using PCA and SIMCA are briefly stated in the following paragraphs.

Principal component analysis (PCA) is one of the most extensively used multivariate statistical techniques in chemometrics and represents a powerful tool for exploratory data analysis and for making predictive models [27]. The linear multivariate PCA models are developed using orthogonal basis vectors (eigenvectors), which are called principal components, thereby reducing the high-dimensional LIBS data onto a lower dimensional space. In PCA one performs a linear mathematical transformation of the data into a new coordinate system such that the largest variance lies on the first axis and decreases thereafter for each successive axis. For the analysis presented in this article, PCA can be thought of as an unsupervised classification technique that separates the samples into clusters based on the variance of their corresponding LIBS spectra. Ideally, the separation obtained based on PCA analysis of LIBS spectra would correspond to the separate classes of over-the-counter pharmaceutical drugs tested in this study.

In contrast to PCA, SIMCA is a supervised classification technique. It is worth noting that the SIMCA method incorporates the application of PCA for dimensionality reduction [28,29]. Because of its supervised nature, it necessitates a training data set consisting of samples with a set of attributes (*e.g.* LIBS spectra) and, importantly, their class membership (*e.g.* type of pharmaceutical tablet). The primary idea of soft modeling refers to the fact that the classifier can identify samples as belonging to multiple (overlapping) classes and is not constrained to producing a classification of samples into strictly discrete (non-overlapping) classes. Importantly, SIMCA enables independent modeling of the classes as opposed to an overall variance modeling as performed in PCA. (The optimization of number of principal components retained in our models is further detailed in the ensuing Section 3.2) The class distance is estimated as the geometric distance (*e.g.* mean orthogonal distance) from the respective PC models. SIMCA-based predictive classification is subsequently performed by comparing the residual variance of the prospective sample with the mean residual variance of the training samples belonging to the specific class.

## 3. Materials and methods

### 3.1. Experimental

The over-the-counter drug samples were purchased from a local pharmacy. The details of the samples are provided in Table 1. For the coated samples, the spectra were first recorded directly with the coating. (Here, coated samples refer to the tablets that are available with colored coatings on them.) The protocol outlined in Missaghi and Fassihi [30] was used for removal of coating ensuring the relative flatness of sample. The potential variations introduced by this sample preparation technique are elaborated in Section 4.

To ensure that each laser pulse hits a fresh portion, the samples were translated using a motorized linear X–Y translation stage. Laser pulses with the energy of 25 mJ from a second harmonic of Nd:YAG laser at 532 nm (7 ns pulse width, 10 Hz repetition rate) were focused on to the sample using an 80 mm convex lens. The signal was collected using a lens system and was coupled to the spectrograph (Michelle ANDOR ME5000, coupled to an iSTAR DH734 ICCD). The resolving power of the spectrometer used was 5000. Spectra were recorded with an integration time of 1 μs and a delay of 0.5 μs. The delay refers to the time difference between the incidence of the laser pulse on the sample and opening of the ICCD gate. The Pockels cell output of the laser triggered a delay generator (SRS DG 535), which in turn provided a TTL (0–5 V) pulse to trigger the ICCD.

A set of ten tablets was used for each of the drug samples. Two spectra from each tablet were acquired after taking average over ten consecutive pulses. In this way, twenty spectra for each sample were recorded. However, a subset of these spectra for each sample (as listed in Table 1) was used for chemometric analysis after accounting for threshold signal-to-noise ratios and outlier rejection

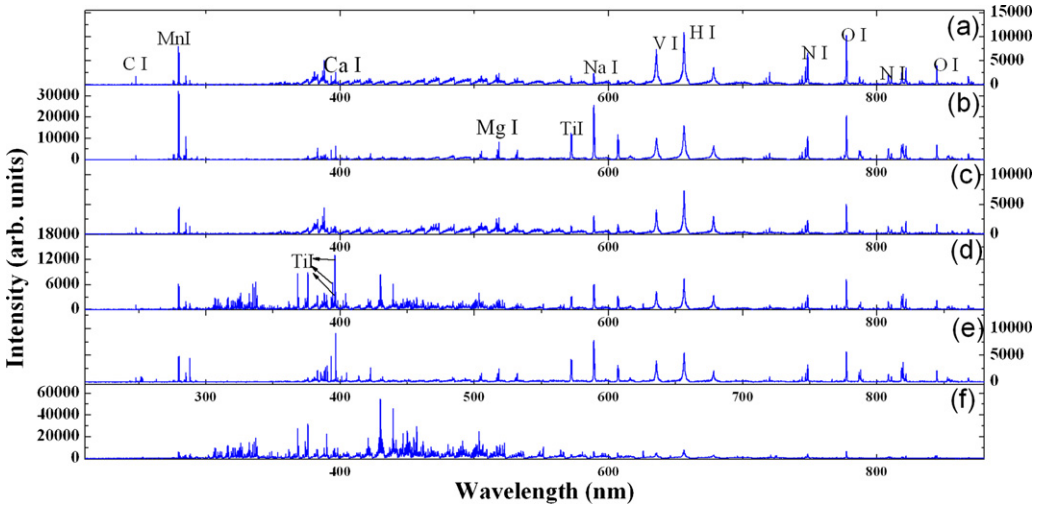


**Fig. 1.** LIBS spectra of the samples used in the study. (a) Paracetamol, (b) Vitamin C, (c) Brufen, (d) coated Brufen, (e) glucosamine, and (f) glucosamine-coated. OI refers to neutral oxygen and CaII refers to singly ionized calcium.

using Student's *t*-test employing the Mahalanobis distance function [31,32]. No other data pre-processing was performed prior to chemometric analysis to avoid the potential incorporation of spurious effects into our calibration models.

### 3.2. Data analysis

For both of the aforementioned methods, 85 spectral datasets acquired from pharmaceutical samples were used for analysis. Each spectrum contained 25,505 information pixels and further variable selection was not pursued in the analysis presented here. First, PCA models were created based on the entire spectral dataset using the Statistics Toolbox of MATLAB R2010b (Math Works, Natick, MA). Since the constructed PCA models were used only for visualization purposes (rather than for class prediction), no optimization was performed for determining the number of principal components to be retained (unlike for SIMCA, as described below).

SIMCA was performed on the spectral dataset in conjunction with the class membership information. In this investigation, 30 test samples (5 samples per each of the 6 classes of tablets) were randomly chosen and kept aside for prospective application. The construction of an independent test set is a standard chemometrics approach employed to diminish and/or examine for the presence of spurious correlations. Subsequently, the 55 training samples were used to develop the SIMCA models using a modified version of the LIBRA toolbox for MATLAB originally developed by Verboven and Hubert [33]. (It is worth emphasizing that separate PCA models were developed for each class of drug on the corresponding 55 training samples per iteration and that these PC models were not linked to the previous PCA implementation on 85 samples, which was only used for visualization purposes.) To ensure the reproducibility of the classification results, 100 iterations were performed to obtain an average value of the classification accuracy (where for every iteration re-splitting of the entire 85 sample dataset into training (55) and test (30) samples was performed). A key point in SIMCA model development is to decide how many of the principal components should be retained in the subsequent analysis. In an ideal situation, this number should equal the number of sample constituents. However, in real life applications (such as in our study), the number is rarely known *a priori*; moreover, correlations between sample constituents, system drift and noise also play a key role in the final number of principal components retained in real world situations. Here, we employed a standard leave-one-out cross-validation procedure to determine the number of PCs for each PCA model. The optimal number of PCs was observed to be in the range of 3–5 for each of the classes under investigation.

Finally, we employed equally weighted scaled orthogonal and score distances for assignment of class membership. In addition, an unclassification criterion was defined to prevent the misclassification of potential samples that were not close to the center of any of the PCA models. Similar to the method outlined in the study performed by Sirven et al. [34], we assumed that the distances of training sample to center of the corresponding class followed a normal distribution. This normal distribution was then employed to compute the probability of class membership of any test spectrum, given its distance to the center of the different classes. In the event that the membership probability for every class was observed to be less than 5%, we assigned it as an "unclassified" sample and removed it from further classification analysis. This protocol helps reduce the number of misclassifications and the decision threshold (i.e. statistical significance of 5%) can be tuned depending on the requirements for the specific monitoring technology.

## 4. Results and discussion

### 4.1. Spectral analysis

A representative LIBS spectrum for each of the six drug samples is plotted separately in Fig. 1(a)–(f), for the sake of clarity. All the spectra exhibited peaks corresponding to nitrogen, oxygen, hydrogen and carbon. These lines can be attributed to the primary components responsible for the action of the tablets (active pharmaceutical ingredients), which are organic molecules. Though the primary component of Brufen and Vitamin C do not contain any nitrogen in them, yet their LIBS spectra show the nitrogen peaks. The possible reason for observing nitrogen peaks in these samples could be the presence of other ingredients such as flavoring and coloring agents that are added to the pharmaceutical tablets. The spectra also exhibited peaks corresponding to iron, manganese, sodium, vanadium, magnesium, titanium and calcium. Except for vanadium, all the elements were detected when separate measurements were performed using ICP-OES (data not shown here). We suspect that the other metals could be components of excipients [35] (especially sodium, iron and titanium) or contaminants [36]. Differences were also observed with respect to coated and uncoated samples. While both the recordings showed peaks corresponding to carbon, hydrogen, oxygen and nitrogen, the coated spectra showed strong peaks corresponding to titanium. Table 2 shows the list of the

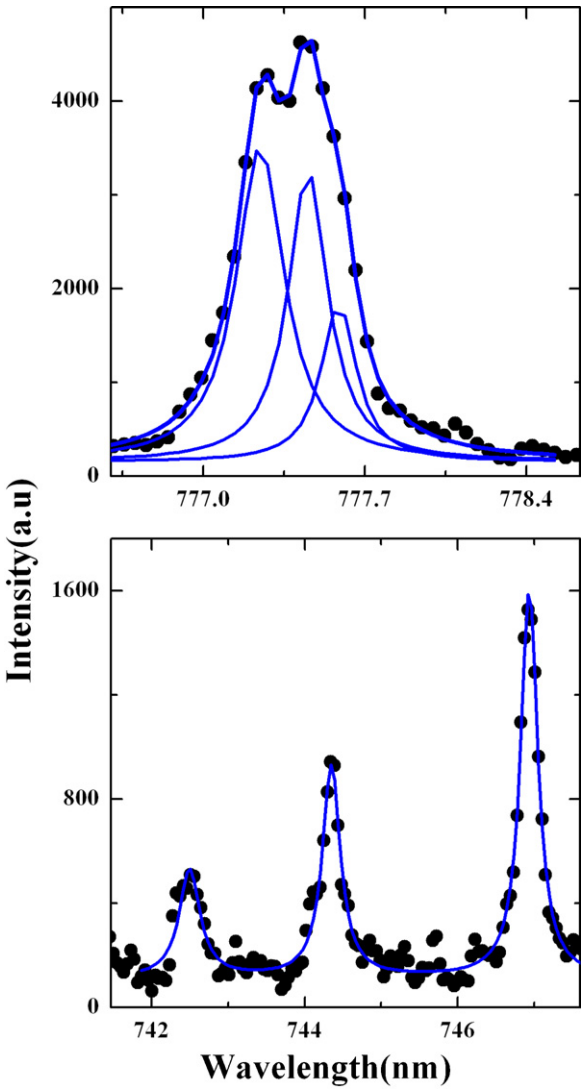


**Fig. 2.** (a) Oxygen and (b) nitrogen LIBS peaks from the Paracetamol spectra. The dots represent the experimental points and the solid lines are the Lorentzian fits.

**Table 2**
Different peaks observed in the LIBS spectra and the corresponding atomic elements.

| S. No. | Element | Wavelength (nm) |
|---|---|---|
| 1 | Carbon | 247.85 |
| 2 | Iron | 279.78, 283.59, 285.18 |
| 3 | Manganese | 279.10, 279.48, 380.96 |
| 4 | Sodium | 589.0, 589.60 |
| 5 | Vanadium | 251.16, 572.68, 635.70 |
| 6 | Oxygen | 777.19, 777.41, 777.53, 822.18, 822.76 |
| 7 | Nitrogen | 742.36, 744.23, 746.83, 818.48, 818.80, 821.63, 824.23 |
| 8 | Hydrogen | 656.27 |
| 9 | Magnesium | 518.36 |
| 10 | Titanium | 394.8, 395.6, 395.8, 399.8 |
| 11 | Calcium | 393.37, 396.86 |

elements observed in the spectra and their corresponding emission wavelengths.

### 4.2. Ratiometric analysis

We explore the potential of the ratiometric approach in identifying the tablets based on their respective oxygen to nitrogen intensity ratios. This approach was previously shown to yield reasonable identification of organic nitro-compounds, namely 4-nitroaniline and 4-nitrotoluene, by Rai et al. [37]. The oxygen peak at 777 nm (O) and nitrogen peaks at 742.36 nm ($N_1$), 744.23 nm ($N_2$), and 746.83 nm ($N_3$) were used for evaluating the O/N ratios. The oxygen peak at 777 nm is a triplet and was not fully resolved in our LIBS spectra. Three different O/N ratios were calculated corresponding to the peaks of nitrogen at $O/N_1$, $O/N_2$ and $O/N_3$. A direct evaluation of the O/N intensity ratios by considering the areas under the peaks resulted in very large values. This is understandable as density of the species is represented by $I/Ag$, where $I$ is the observed area under the peak, $A$ is the transition probability and $g$ is the degeneracy of the upper energy level involved in the transition. As the observed oxygen peak at 777 nm is a triplet, it was deconvolved using a triple Lorentzian fit. Fig. 2 shows a typical fit for the nitrogen using a Lorentzian fit and oxygen with a triple Lorentzian.

$I/Ag$ was calculated for three peaks corresponding to 777.19 nm $(I/Ag)_1$, 777. 41 nm $(I/Ag)_2$ and 777. 53 nm $(I/Ag)_3$ for oxygen. The values of $A$ and $g$ were taken from NIST database [6]. $I/Ag$ for oxygen was taken as the sum of these three ratios $[(I/Ag)_{oxygen} = (I/Ag)_1 + (I/Ag)_2 + (I/Ag)_3]$. The ratio of $(I/Ag)_{oxygen}/(I/Ag)_{nitrogen}$ was taken with $N_1$, $N_2$ and $N_3$. Table 3 shows the O/N ratios for all the samples. The ratios for all the spectra were first calculated. An elimination of ratios outside the mean ± standard deviation was performed. The ratios reported in the table are the averages over this set. We observed that except for Paracetamol and Vitamin C for which RSD were less than 10, the ratios showed a large variation. This could be a result of inhomogeneity in the sample consisting of various additives including the matrix. It was also observed that there is a general increasing trend of RSD from $O/N_3$ to $O/N_1$. In order to assess the effect of the ambient air, which also contains the oxygen and nitrogen, the spectra for Paracetamol were recorded under argon gas purging. This resulted in all the three ratios becoming close to each other and close to the actual value of 2. A CHNO analysis (Thermo Finnigan flash EA 112 analyzer) of Paracetamol also yielded a ratio of 1.99. Interestingly, this ratio matches very well with the O/N ratio of the primary component in Paracetamol as well as the value obtained by the LIBS spectra with argon purging.

### 4.3. Multivariate chemometric analysis

To enhance the accuracy for LIBS-based pharmaceutical tablet classification beyond that obtained by standard ratiometric approaches, PCA and SIMCA were employed. PCA was first performed on the entire 85 sample dataset for understanding the critical spectral features in the LIBS dataset as well as for probing the cluster behavior of the pharmaceutical samples.

Fig. 3 shows the first three principal components, which reveal the dimensions that explain most of the variance present in the dataset. These components, while abstract in form as they are obtained through a mathematical change of basis (*e.g.* via singular value decomposition), are useful in indicating the informative spectral features associated with the different samples. Here, we observed that the first PC appears strikingly similar to the glucosamine coated spectra (compared to Fig. 1) and the corresponding magnitudes of the scores for the glucosamine coated samples are larger than that for the other samples. On the other hand, PC2 shows a substantive influence of manganese alongside the contributions of the API elements such as carbon, hydrogen, nitrogen and oxygen. This is reflected in the higher PC scores for the Paracetamol and Vitamin C samples. The first two PCs explain 90.67% of the total variance in the dataset (as computed from the cumulative contribution of their eigenvalues). After incorporating the third PC, this metric rises to 96.29%. The relatively less significance of the third PC can also be visualized from Fig. 3, as some of the spectral features are repetitions from the previous PCs. The

**Table 3**
Oxygen to nitrogen ratios with the oxygen peak at 777 nm (O) and nitrogen peaks at 742.36 nm ($N_1$), 744.23 nm ($N_2$), and 746.83 nm ($N_3$).

| S No. | Sample | $O/N_1$ | $O/N_2$ | $O/N_3$ |
|---|---|---|---|---|
| 1 | Brufen | 3.08 ± 1.28 | 2.55 ± 0.37 | 1.95 ± 0.15 |
| 2 | Brufen-coated | 2.91 ± 0.50 | 2.51 ± 0.34 | 2.37 ± 0.41 |
| 3 | Glucosamine | 2.66 ± 0.75 | 3.15 ± 0.46 | 2.62 ± 0.31 |
| 4 | Glucosamine-coated | 3.65 ± 1.81 | 2.13 ± 0.28 | 2.57 ± 0.20 |
| 5 | Vitamin C | 2.83 ± 0.22 | 2.74 ± 0.27 | 2.23 ± 0.09 |
| 6 | Paracetamol | 2.32 ± 0.21 | 2.23 ± 0.15 | 1.80 ± 0.09 |
| | Paracetamol – argon purging | 1.86 ± 0.2 | 1.82 ± 0.12 | 1.93 ± 0.08 |

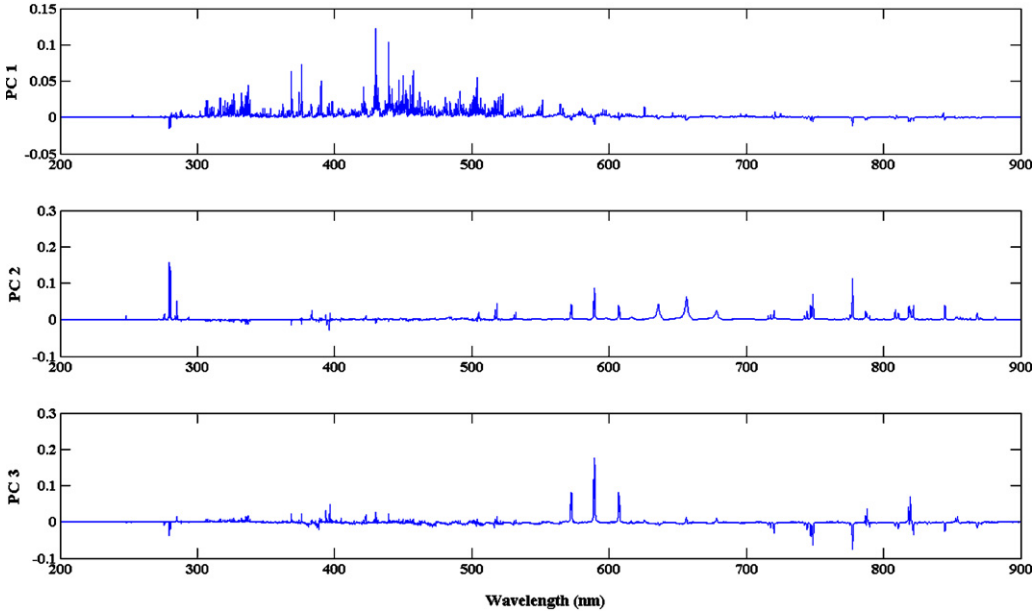


**Fig. 3.** The first three principal components corresponding to the entire spectral dataset acquired from the pharmaceutical samples. These three principal components, combined, explain 96.29% of the net variance in the dataset.

subsequent PCs (i.e. fourth, fifth, etc.) are fairly noisy and their incorporation deteriorates the quality of the model.

Fig. 4 shows the PC scores plot for LIBS analysis of the six different classes of pharmaceutical drugs (for the aforementioned first three PCs). Clearly, the samples of each class tend to cluster together and in almost all cases are fairly well separated from the other classes. Of these, Vitamin C and Paracetamol appear to be the easiest to distinguish based on their distance from the other classes. It is also interesting to probe the dispersion of the different classes along the PC directions. For example, Vitamin C and glucosamine tablets show a considerably larger dispersion as compared to both the coated and uncoated Brufen samples, which exhibit a more uniform pattern. This provides a novel insight into the tablet-to-tablet intra-class variations (*e.g.* arising from the heterogeneity of each tablet composition). Additionally, we can also observe outliers based on the spectral data, notably one coated glucosamine sample and two Paracetamol samples as seen in Fig. 4. Based on these results we can reasonably infer that: (a) the LIBS spectra provides vital information which can be used in routine sample monitoring for pharmaceutical tablets (or at least the ones used in this study) and (b) the PCA classification is able to identify the primary elements which help in distinguishing the various classes, which corresponds to the existing knowledge of the sample composition. The latter proves that there is a direct causality between PCA classification and chemical basis of the samples, rather than an arbitrary (potentially spurious) correlation, which cannot be successfully reproduced in prospective application. Nevertheless, while PCA is a valuable tool for recognizing similarities between sample types, it does not automatically provide class memberships due to its unsupervised nature. To assign class membership to the tested tablets, we have employed SIMCA.

As mentioned earlier, SIMCA computes a PCA model for each of the six classes of pharmaceutical samples and identifies the prospective samples based on their distance to the respective models. Based on this key concept and the previously stated unclassification criterion, class memberships were assigned to 30 prediction samples for each iteration. Fig. 5 shows a bar plot visualization of SIMCA classifications for a representative set of 30 test samples (for a specific iteration). In this representative case we observed that none of the samples were unclassified (unclassified samples are assigned a class membership of 0) and three of them are misclassified.

To obtain a more comprehensive sense of the results, we computed the average rate of unclassification, misclassification and

**Table 4**
SIMCA classification results obtained from 30 test samples over 100 iterations.

| Average rate of . . . | Correct classification | Wrong classification | Unclassification |
|---|---|---|---|
| Brufen | 0.908 | 0.076 | 0.016 |
| Brufen-coated | 0.934 | 0.062 | 0.004 |
| Glucosamine | 0.904 | 0.088 | 0.008 |
| Glucosamine-coated | 0.940 | 0.002 | 0.058 |
| Paracetamol | 0.968 | 0 | 0.032 |
| Vitamin C | 0.988 | 0 | 0.012 |
| Average | 0.9403 | 0.038 | 0.0217 |

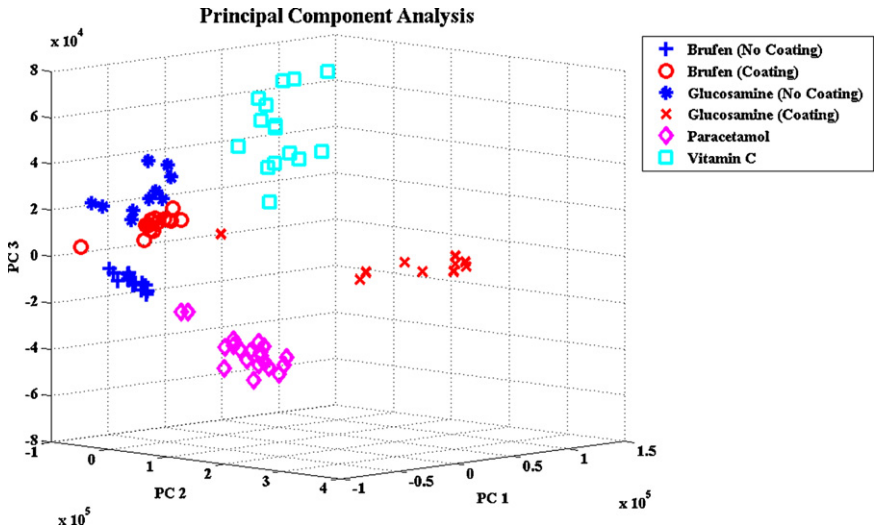


**Fig. 4.** PC scores plot of the first three principal components for the spectral dataset acquired from the six classes of pharmaceutical samples.

correct classification over 100 iterations (Table 4). We observe that on average we obtain acceptable rate of correct classification for all the classes of pharmaceutical samples (>90%). In addition, rate of unclassification is fairly low indicating that the acquired data consisted of few outliers and the training model was sensitive to all remaining samples (i.e. all but the outliers). We observed that the unclassification is relatively higher for the coated glucosamine (5.8%) and Paracetamol (3.2%) samples. This is not surprising based on the PCA scores plot (Fig. 4), where one could clearly observe one coated glucosamine and two Paracetamol spectral outliers. Furthermore, we find that the Paracetamol, Vitamin C and coated glucosamine samples are almost perfectly classified (i.e. there is no misclassification), which is again consistent with the clear separation of these samples observed in the PCA scores plot (Fig. 4). The others, namely the Brufen (both coated and uncoated) and uncoated glucosamine sample, are significantly more difficult to classify resulting in higher errors (7.6%, 6.2% and 8.8%, respectively).

An interesting point in this regard is the striking difference in classification accuracy for the coated and uncoated samples, both for Brufen and glucosamine. The significantly better classification accuracy for the coated samples (93.4% and 94% correct classification for Brufen and glucosamine, respectively, compared to 90.8% and 90.4% for their uncoated counterparts) can be primarily attributed to the distinctive coating composition of these tablets. Coating is mostly made of titanium oxide, which is evidently absent in the uncoated tablets as was shown in Fig. 1. Alternately, one may attribute the superior classification results obtained for coated samples to experimental errors/inaccuracies in removal of the coating from the respective tablets. Improper removal of the coating may lead to non-uniform surfaces (i.e.

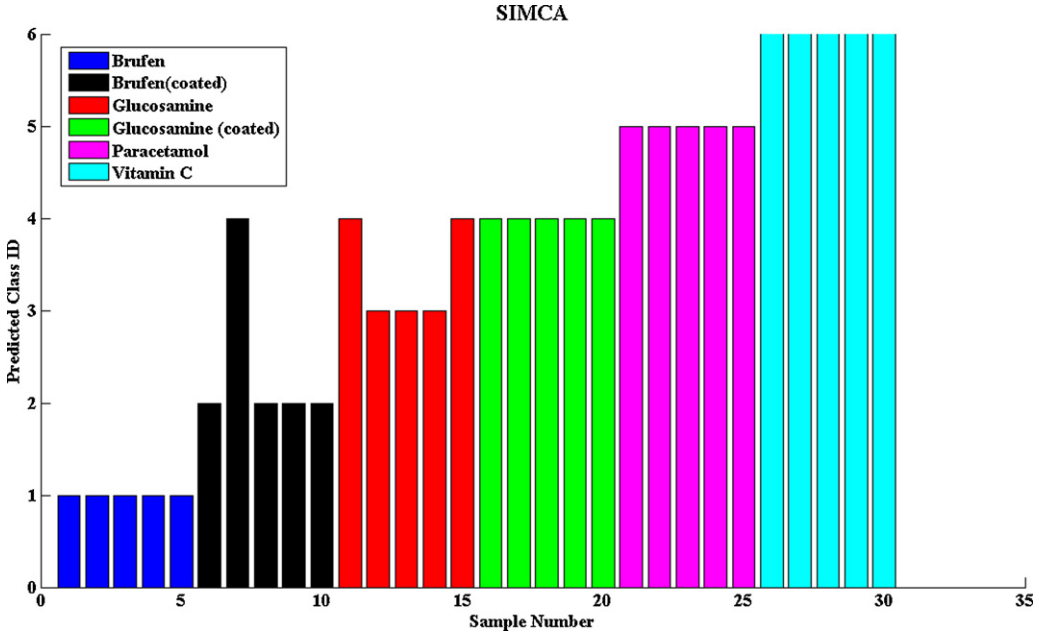


**Fig. 5.** Bar plot of SIMCA classifications for a representative set of 30 test samples. The test samples are ordered such that each successive subset of 5 samples belongs to a different class, as shown by the different bar coloring. Predicted Class ID are as follows: 0: unclassified; 1: Brufen; 2: Brufen (coated); 3: glucosamine; 4: glucosamine (coated); 5: Paracetamol; and 6: Vitamin C. Here, three misclassifications are observed, one corresponding to coated Brufen and two corresponding to glucosamine. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

roughness) at the microscopic scale, which may in turn result in noise in the LIBS measurements. In addition, the residual coating elements could contaminate the signal acquired at the LIBS spectrometer due to their (undesirable) presence in the plasma plume of the uncoated samples.

Finally, we note that for the Brufen (coated and uncoated) and the uncoated glucosamine tablets, the variance between the classes is of the scale of the variance within the respective classes, thereby impeding prospective SIMCA analysis. This problem can be potentially solved by employing an alternate classification method such as PLS-DA (partial least squares based discriminant analysis) in the future [38]. PLS-DA seeks to establish the maximum separation between classes as opposed to PCA which does not discriminate between class-to-class and intra-class variability in explaining the total variance in the dataset. Consequently, PLS-DA may present a better alternative when these two variability numbers are of the same order, as observed above for Brufen and uncoated glucosamine samples.

In summary, we have shown the potential of LIBS as an alternate method for in-line monitoring and analysis of different pharmaceutical drugs. The current study lays the feasibility foundation for further investigations of the robustness of the proposed approach as well as an evaluation of its true value with respect to conventional approaches such as ICP-OES. In addition, we envision that incorporation of feature selection approaches will lead to a significant reduction in computation time and a possible increase in accuracy and robustness of the classification models. This may also result in a concomitant simplification of the necessary hardware for LIBS data acquisition and will form the core of our future studies.

## 5. Conclusions

In this report, we have studied the effectiveness of routine monitoring of commercial pharmaceutical tablets using a combination of LIBS and chemometric methods. Oxygen to nitrogen ratios were calculated based on the spectra but nevertheless failed to provide an acceptable classification. Ratios for Paracetamol and Vitamin C showed a RSD of less than 10 with $O/N_3$ RSD for Vitamin C being 4.03. Large variations in the calculated ratios are a result of multiple components being present in the tablet. The observed O/N ratios can be further improved by performing the experiment in inert gas purging. In contrast to the ratiometric approach, PCA exhibited a clear visual diagnosis of the different classes of samples/spectra and also provides a physical interpretation of the classification results by identifying the key components that explain the variance in the dataset. Finally, SIMCA was employed for automatic prediction modeling with an average rate of approximately 94% correct classification. Based on the results obtained in this study, we expect that the combination of LIBS and chemometrics can be used successfully for quality control and routine monitoring of pharmaceutical tablets. Further, this combination can be used to screen and establish qualitative formulation differences between suspect counterfeit and authentic tablets, as long as appropriate training of the classification models is undertaken. In addition, the proposed approach is broad and general enough to be extended to similar (and otherwise intractable) applications in process control and on-site reaction monitoring. We envision that future work in the area of developing more robust classification methods, which can suitably treat even non-representative prospective samples, will enhance the feasibility of this approach in the industrial domain.

## Acknowledgements

GMK acknowledges the financial support DRDO, India. IB and NCD acknowledge the support of NIH National Center for Research Resources (Grant No. P41-RR02594). IB also acknowledges the support of Lester Wolfe Fellowship from the Laser Biomedical Research Center.

## References

[1] A.W. Miziolek, V. Palleschi, I. Schechter, Laser-induced Breakdown Spectroscopy (LIBS): Fundamentals and Applications, Cambridge University Press, New York, 2006, pp. 1–40.
[2] D.A. Cremers, L.J. Radziemski, Handbook of Laser-induced Breakdown Spectroscopy, Wiley & Sons Ltd., UK, 2006, pp. 23–50.
[3] J.P. Singh, S.N. Thakur, Laser-induced Breakdown Spectroscopy, Elsevier Science B.V., Amsterdam, Netherlands, 2007, pp. 3–19.
[4] M. Weidman, M. Baudelet, S. Palanco, M. Sigman, P.J. Dagdigian, M. Richardson, Opt. Express 18 (2010) 259–266.
[5] M. Baudelet, L. Guyon, J. Yu, J.-P. Wolf, Appl. Phys. Lett. 88 (2006) 063901–63903.
[6] NIST database of atomic spectral data, http://physics.nist.gov/PhysRefData/ASD.
[7] F.J. Fortes, J.J. Laserna, Spectrochim. Acta Part B 65 (2010) 975–990.
[8] M.D. Mowery, R. Sing, J. Kirsch, A. Razaghi, S. Bechard, R.A. Reed, J. Pharm. Biomed. Anal. 28 (2002) 935–943.
[9] J.-F. Archambault, A. Vintiloiu, E. Kwong, AAPS PharmSciTech 6 (2005) E253–E261.
[10] F.R. Doucet, P.J. Faustino, M. Sabsabi, R.C. Lyon, J. Anal. Atom. Spectrom. 23 (2008) 694–701.
[11] M.D. Mowery, R. Sing, J. Kirsch, A. Razaghi, S. Beĭchard, R.A. Reed, J. Pharm. Biomed. Anal. 28 (2002) 935–943.
[12] L. Barrette, S. Turmel, Spectrochim. Acta Part B 56 (2001) 715–723.
[13] A. Lanzarotta, K. Lakes, C.A. Marcott, M.R. Witkowski, A.J. Sommer, Anal. Chem. 83 (2011) 5972–5978.
[14] D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, L. Kaufman, Chemometrics, Elsevier Science Publishers BV, Amsterdam, 1988, pp. 1–488.
[15] H. Martens, T. Naes, Multivariate Calibration, John Wiley & Sons, Chichester, 1989, pp. 1–419.
[16] M.J. Adams, Chemometrics in Analytical Spectroscopy, RSC Analytical Spectroscopy Monographs, Cambridge, 2nd ed., 2004.
[17] J.B. Sirven, B. Bousquet, L. Canioni, L. Sarger, S. Tellier, M. Potin-Gautier, I. Le Hecho, Anal. Bioanal. Chem. 385 (2006) 256–262.
[18] J.D. Hybl, G.A. Lithgow, S.G. Buckley, Appl. Spectrosc. 57 (2003) 1207–1215.
[19] A. Erdem, A. Cilingiroglu, A. Giakoumaki, M. Castanys, E. Kartsonaki, C. Fotakis, D. Anglos, J. Archaeol. Sci. 35 (2008) 2486–2494.
[20] A.C. Samuels, F.C. DeLucia, K.L. McNesby Jr., A.W. Miziolek, Appl. Opt. 42 (2003) 6205–6209.
[21] F.M.V. Pereira, D.M.B.P. Milori, A.L. Venâncio, M. de S.T. Russo, P.K. Martins, J. Freitas-Astua, Talanta 83 (2010) 351–356.
[22] F.C. De Lucia, J.L. Gottfried, Spectrochim. Acta Part B 66 (2011) 122–128.
[23] J.L. Gottfried, F.C. De Lucia Jr., A.W. Miziolek, J. Anal. Atom. Spectrom. 24 (2009) 288–296.
[24] R. Sattmann, I. Monch, H. Krause, R. Noll, S. Couris, A. Hatziapostolou, A. Mavromanolakis, C. Fotakis, E. Larrauri, R. Miguel, Appl. Spectrosc. 52 (1998) 456–461.
[25] S.J. Rehse, Q.I. Mohaidat, Spectrochim. Acta Part B 64 (2009) 1020–1027.
[26] J. Diedrich, S.J. Rehse, J. Appl. Phys. 102 (2007) 014702–14708.
[27] R.G. Brereton, Chemometrics: Data Analysis for the Laboratory and Chemical Plant, Wiley Chichester, West Sussex, England/Hoboken, NJ, 2003.
[28] S. Wold, Pattern Recogn. 8 (1976) 127–139.
[29] G. Blomquist, E. Johansson, B. Söderström, S. Wold, J. Chromatogr. 173 (1979) 19–32.
[30] S. Missaghi, R. Fassihi, AAPS PharmSciTech 5 (2004) 1–8.
[31] D. Qi, A.J. Berger, Appl. Opt. 46 (2007) 1726–1734.
[32] N.C. Dingari, I. Barman, J.W. Kang, C. Kong, R.R. Dasari, M.S. Feld, J. Biomed. Opt. 16 (2011) 087009.
[33] S. Verboven, M. Hubert, Chemometr. Intell. Lab. Syst. 75 (2005) 127–136.
[34] J.-B. Sirven, B. Salle, P. Mauchien, J.-L. Lacour, S.M.G. Manhès, J. Anal. Atom. Spectrom. 22 (2007) 1471–1480.
[35] D.M. Brunette, P. Tengvall, M. Textor, P. Thomsen, Titanium in Medicine: Material Science, Surface Science, Engineering, Biological Responses and Medical Applications, 1st ed., Springer, 2001.
[36] N. Lewen, J. Pharm. Biomed. Anal. 55 (2011) 653–661.
[37] S.K. Rai, A.K. Rai, S.N. Thakur, Appl. Phys. B 91 (2008) 645–650.
[38] F.C. De Lucia Jr., J.L. Gottfried, C.A. Munson, A.W. Miziolek, Appl. Opt. 47 (2008) G112–G121.