Skin Research and Technology 2010; ■■: 1–6
Printed in Singapore · All rights reserved
doi: 10.1111/j.1600-0846.2010.00461.x

Use of Raman spectroscopy for the early detection of filaggrin-related atopic dermatitis

F. J. González[1], J. Alda[2], B. Moreno-Cruz[3], M. Martínez-Escanamé[3], M. G. Ramírez-Elías[1], B. Torres-Álvarez[3] and B. Moncada[3]

[1]Coordinación para la Innovación y la Aplicación de la Ciencia y la Tecnología, Universidad Autónoma de San Luis Potosí, SLP, México, [2]Applied Optics Complutense Group, University Complutense of Madrid, Madrid, Spain and [3]Dermatology Department, Hospital Central 'Dr Ignacio Morones Prieto', Universidad Autónoma de San Luis Potosí, SLP, México

Background: Filaggrin (FLG) gene mutations, which result in complete or incomplete loss of proFLG/FLG peptides, have been reported as an important predisposing factor for atopic dermatitis (AD) and secondary atopic phenotypes such as atopic asthma.
Method: The presence of the protein FLG in the skin was evaluated at birth on 12 infants using Raman spectroscopy; these 12 infants were monitored for 1 year to see whether they developed AD. Three different statistical analysis procedures, two of which involved principal component analysis (PCA), were performed on the Raman spectra in order to determine the FLG content.
Results: The infants who had a lower FLG content, determined using any of the three statistical analysis procedures proposed, were also the ones that clinically developed AD.
Conclusion: The results suggest that Raman spectroscopy and statistical analysis such as PCA could be used as an early detection procedure for FLG -related AD and as a possible quantitative marker for FLG gene mutations.

Key words: atopic dermatitis – filaggrin – Raman spectroscopy – principal component analysis

Abbreviations: FLG: filaggrin; AD: atopic dermatitis; PCA: principal component analysis; SC: stratum corneum; NMF: natural moisturizing factor.

Accepted for publication 4 June 2010

Filaggrin (FLG) is a key protein required for the formation of the stratum corneum (SC) barrier (1); it plays an important role in the maintenance of the skin barrier function (2) and is essential for SC hydration (1).

FLG gene mutations have been reported as an important predisposing factor for atopic dermatitis (AD) and secondary atopic phenotypes such as atopic asthma (3), which suggests that a skin barrier defect is a primary key event leading to allergic sensitization and development of AD and related allergic phenotypes (3).

Kezic et al. (1) used confocal Raman microspectroscopy to measure natural moisturizing factor (NMF) as a function of SC depth, showing that individuals who are carriers of FLG-null mutations have lower levels of the NMF in the SC and a higher transepidermal water loss as compared with non-carriers; these findings suggest that the measurement of these parameters could be used as a marker of FLG gene status (1).

Nemoto-Hasebe et al. (2), using immunohistochemistry, showed that FLG gene mutations result in complete or incomplete loss of proFLG/FLG peptides.

Principal component analysis (PCA) is a multivariate technique that can be applied to a data set without any prior knowledge about its nature, making it an unsupervised data analysis technique (4, 5). However, the correct analysis of the results of the PCA requires an understanding of the type of data being evaluated and how the given results are relevant to the problem.

In this work, the presence of the protein FLG in the skin is detected using Raman spectroscopy and PCA as an early detection procedure for FLG-related AD and as a possible marker for FLG gene mutations. The technique proposed produces quantitative parameters that positively identify those patients who developed AD by using a non-invasive optical procedure at birth, when no sign or symptom of the disease was evident.

SRT 461 | Journal Name | Manuscript No. | B | Dispatch: 29.6.10 | Journal: SRT | CE: Guna | Author Received: | No. of pages: 6 | PE: Chris/Aswani

Materials and Methods

Twelve healthy infants born without complications from healthy mothers participated in this study. Informed consent was obtained from the mothers of all the participants and the study was approved by the ethics committee of the Hospital 'Dr Ignacio Morones Prieto'.

Raman scattering measurements were performed on the right thigh for each infant at birth; the measurements were performed at room temperature using a Raman Systems R3000 spectrometer (Ocean Optics, Dunedin, FL, USA) with a 785 nm laser diode, a spectral resolution of 8 cm^{-1} and a laser power of 90 mW. The irradiance of this laser diode is below the ANSI standard for skin and none of the participants showed any kind of discomfort when the measurement was performed. The measurements were performed in the 200–1800 cm^{-1} spectral range and the instrument was calibrated using a Teflon standard every day before each round of measurements.

The Raman measurements included one *in vitro* Raman spectrum of pure FLG human recombinant protein (GenWay Biotech Inc., San Diego, CA, USA) in order to have a fixed spectrum that would help us assess the amount of FLG in the measured spectra. For this measurement, a quartz cuvette and a 300 mW laser power were used.

The participants of this study were monitored for 1 year to see whether any of them developed AD. The dermatologist's diagnosis of AD was made according to standard criteria in the presence of a chronic or chronically relapsing pruritic dermatitis with the typical morphology and distribution (6, 7).

Three different types of statistical analyses were performed on the measured spectra to obtain quantitative parameters related to the problem. The first analysis consisted in calculating the cross-correlation between the spectra in order to determine the similarities between spectra and relate them to clinical findings. Also, the PCA method was applied to the set of data. The second analysis consisted in analyzing the FLG content of each measured spectrum by reconstructing the measured spectra using only the principal components relevant to the FLG and the third statistical approach consisted in constructing a two-dimensional plot of the first and second principal components and comparing the location of the spectrum of FLG with the location of the measured spectra.

Using the cross-correlation matrix, the relation between a given set of spectra can be obtained (8); the element (i, j) of this matrix quantifies the similarity of spectrum i to spectrum j. In this work, it is of interest to determine how the FLG spectrum is related to the patients' spectra; as a first approach, the cross-correlation matrix was used to find these similarities. However, there are other statistical techniques, such as the PCA, that provide a more in-depth analysis of the structure of the variance–covariance matrix between spectra. It is worth noting that the cross-correlation matrix is a normalized version of the variance–covariance matrix of the data (8, 9).

The N measured spectra were sampled at M different wavenumbers and arranged as an $M \times N$ matrix that can be seen as M, N-dimensional vectors. Applying the PCA technique to these data, the PCA yields three types of results: N principal components, an $N \times N$ matrix containing the coefficients for the transformation between the original data and the principal components, and N eigenvalues describing the importance of the corresponding principal components. The original N experimental spectra are transformed into a new set of N 'synthetic' spectra called principal components. These principal components are independent of each other; statistically, we could say that they are uncorrelated, as opposed to the original ones, which show a non-negligible cross-correlation between them. Besides, PCA yields an $N \times N$ matrix with the coefficients to transform the original spectra into the new uncorrelated ones, and vice versa. Also, these principal components are given in decreasing order of importance, which means that the first principal component can explain the largest quantity of the variance of the original data, the second one explains more variance than the third, and so on. In the terminology used by the PCA, we call these variance contributions as eigenvalues. Mathematically, they are the diagonal of the covariance matrix among principal components. One of the advantages of the PCA is that, by evaluating the relative importance of the consecutive principal components, it is possible to reduce the dimension of the data set by finding a smaller collection of spectra that explain a given and acceptable amount of variance. On the other hand, when advancing towards the last principal component, the consecutive PC can be grouped together and form processes that are typically explained as noise contributions (10, 11).

This grouping is based on a robust statistical criterion that involves the number of spectral points measured and the desired level of confidence for the grouping of the principal components.

Taking into account the properties of PCA, a method was developed to analyze Raman spectra; the method consisted of including the FLG spectrum with the original data set containing the Raman spectra of the patients, and analyzing the results obtained by applying the PCA method. The eigenvalues were grouped together to reduce the dimension of the data and only the first independent principal components were taken and the rest were considered to be noise contributions. Afterwards, the principal components that are FLG related are identified. A new set of reconstructed spectra is generated using only the FLG-related principal components and compared with the original ones. The square mean deviation between the original and the reconstructed spectra is calculated for each spectrum in order to determine the contribution of the FLG-related principal components to the original spectra. A larger square mean deviation will mean that those spectra will have a lower FLG contribution, which might indicate a degraded or the absence of FLG. At the same time, we identify those principal components that make a lower contribution to the FLG spectrum. The coefficients of the patients' spectra for these FLG-non-related components were used to locate these spectra with respect to the FLG. Then we can identify those spectra located farther apart from the FLG position.

All the mathematical analysis was performed using the computational package Matlab (The Mathworks Inc., Natick, MA, USA) and its Statistics toolbox.

Results

From the 12 infants measured and monitored, three of them developed AD in the course of 1 year.

A statistical analysis was performed on the 12 Raman spectra obtained from the infants at birth and the Raman spectrum of the protein FLG. These 13 spectra were measured over a spectral range from 300.17 to 1783.99 cm^{-1}, with 1484 data points. Therefore, the data matrix used in this analysis has a dimension of 1484×13. The measured spectra were processed in order to remove fluorescence.

Figure 1 shows three spectra corresponding to FLG (black solid line), a patient who developed AD (gray dashed line), and another one who remained healthy (gray solid line).

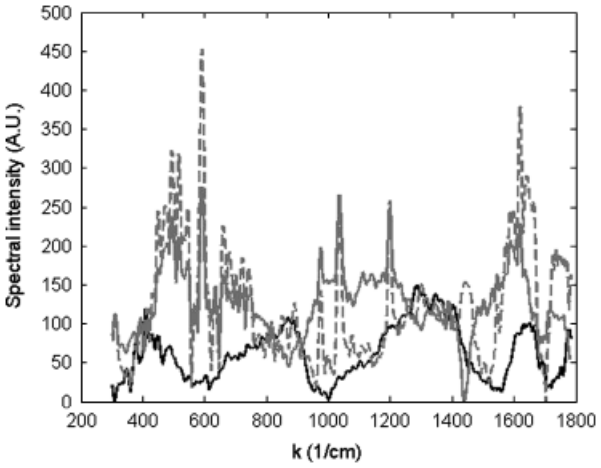

Fig. 1. Plot of three spectra used in this analysis. The solid black spectrum corresponds to the filaggrin, the dashed gray one is for a patient who developed atopic dermatitis (patient #1), and the solid gray spectrum corresponds to a healthy patient (patient #4). These spectra have been pre-processed to eliminate fluorescence.

Figure 2(a) shows the absolute value of the cross-correlation matrix between spectra. The absolute value was used because some of the elements were negative. The FLG data location is labeled as FLG. Figure 2(b) shows the correlation of the patient's spectra with the FLG spectrum; the lower values of the cross-correlation correspond to the patients who later developed the disease (patients #1, #9, and #11 are plotted as gray circles). It is worth noting that other patients (such as #4 and #7) also have low cross-correlation values. From these plots, it can be seen that the cross-correlation can identify those spectra that have a lower FLG component, which would indicate a larger probability of developing AD.

In order to obtain a more in-depth analysis of the influence of the FLG spectrum over the measured spectra, a PCA was applied to the dataset, producing 13 eigenvalues, which are plotted in Fig. 3. Using a statistical criterion to group together the principal components (10), we find that, with a level of confidence of 99%, the first eight components are statistically independent from each other, and after the eighth component, there is a statistical dependence between them that may be associated with noise (10). The first eight principal components explain 99.84% of the total variance of the data; this limit has been represented in Fig. 3 as a solid vertical line.

The criterion to select these principal components is based on the inner connections that could

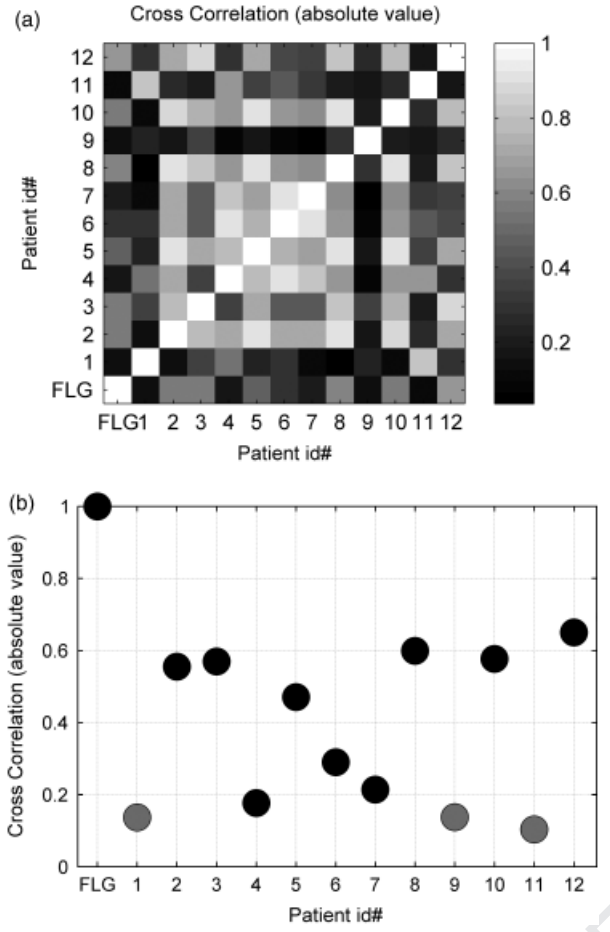

Fig. 2. (a) Cross-correlation matrix of the spectra as a gray-level map and (b) shows the cross-correlation of the filaggrin (FLG) spectrum with the spectra of the patients. FLG labels the FLG data location [the patients developing atopic dermatitis (AD) are plotted with gray circles].

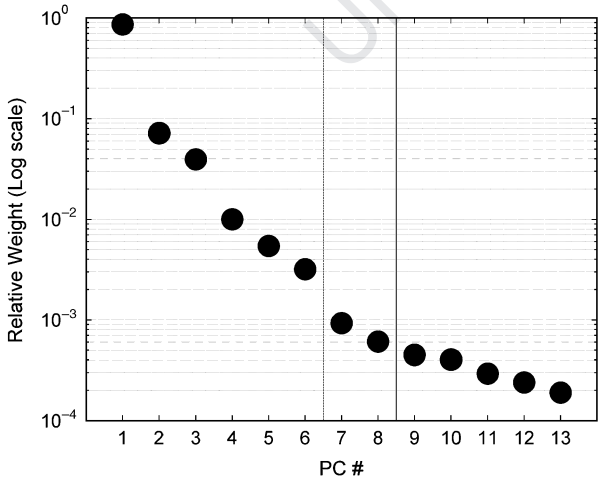

Fig. 3. Semilog plot of the relative weight of the eigenvalues of the principal components. After the eighth principal component (PC), the successive principal components can be combined into processes that can be interpreted as noise. The dashed line represents a limit when considering the filaggrin contribution.

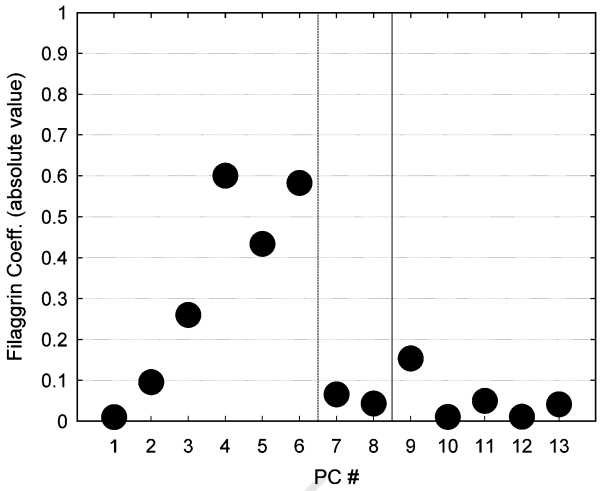

Fig. 4. Coefficient of the filaggrin spectrum in terms of the principal components. The solid line represents the limit in the number of significant principal components when a statistical discrimination criterion is applied. The dashed line is placed to indicate that we neglect the contribution of the principal components #7 and #8.

be established because of statistical uncertainties related to the normality of the data. Then, some principal components cannot be taken individually because of these inner relations. Together, these linked principal components describe a so-called process. Typically, these processes are associated with noise. In our analysis, we have preferred to limit our study to those principal components that can certainly be considered as independent.

Because we are mainly interested in the FLG contribution, special attention is focused on the expansion of the FLG spectrum in terms of the principal components. Figure 4 represents the coefficients of this expansion. It can be seen that the first two principal components make an almost negligible contribution to the FLG spectrum, and that the main contributions emerge from principal components #3, #4, #5, and #6. Therefore, in order to reduce the dimension of the original data while maintaining the FLG-related information, we will only consider up to principal component #6. The criterion has been to neglect those contributions smaller than those from principal components above #8. This limit has been represented as a dashed line both in Figs 3 and 4.

The contribution of the FLG spectrum in the original spectra can be calculated by first reconstructing the original data using only those principal components relevant to the FLG. This means that the reconstruction is carried out by selecting principal components #3, #4, #5, and #6.

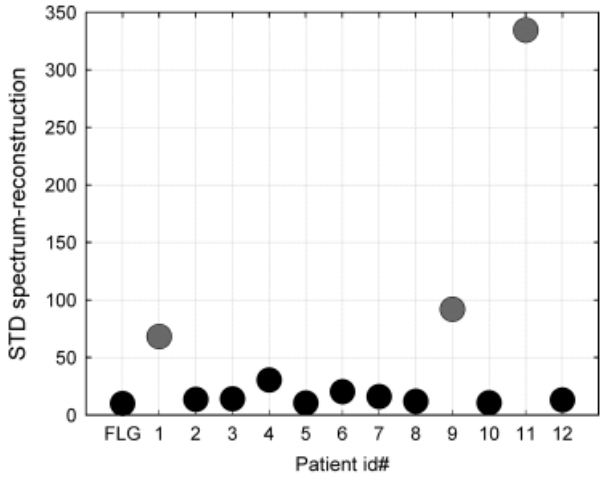

Fig. 5. Quadratic difference between the original spectra and the ones reconstructed using only those principal components with relevance in the filaggrin spectrum. The gray circles represent those patients developing atopic dermatitis (AD).

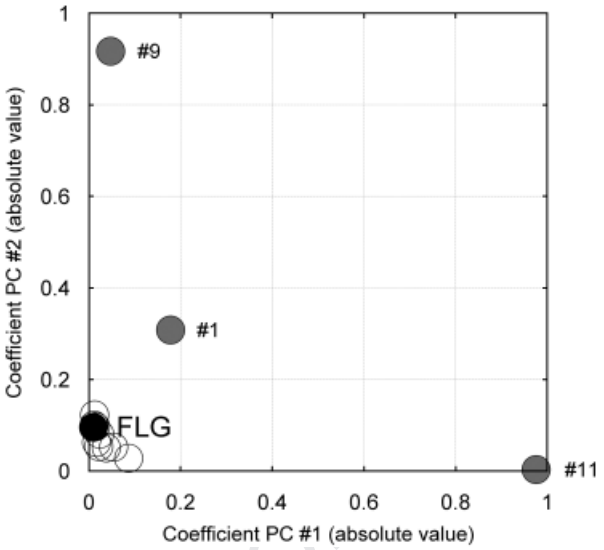

Fig. 6. Location of the spectra in the principal component (PC) space using only the first and second PCs. The gray dots represent those patients developing the disease. The black dot represents the filaggrin spectrum.

It is worth noting that these four principal components explain only 5.83% of the total variance of the data; however, their contribution is quite relevant in order to analyze the amount of FLG in the patients' spectra. The reconstructed spectra are compared with the original ones and the standard deviation of the difference is evaluated. Figure 5 shows this evaluation for the 13 spectra. We may see that the FLG spectrum shows the minimum standard deviation, as it should be because we are using the most significant principal components to the FLG. At the same time, those spectra showing a larger standard deviation are the ones who have a lower FLG component and who were also the ones who eventually developed AD (patients #1, #9, and #11). This identification is made with better specificity than when using only the cross-correlation between spectra.

As a complementary check of the usefulness of the PCA method, the behavior of the first two principal components was evaluated; these two components represent 93.86% of the total variance of the original data. As mentioned before, the coefficients of the FLG spectrum for these two principal components are almost negligible (see Fig. 4). Therefore, we can say that these two principal components are not FLG-related and they represent the lack of FLG in the Raman spectra. The original data can be represented as a bi-dimensional plot having as coordinates the absolute value of the coefficients of the spectra for the first and the second principal components (Fig. 6). The gray dots correspond to those infants who developed AD; the rest of the patients are grouped together around the location of the FLG spectrum (represented as a black dot). This result also indicates that this approach identifies patients who developed the disease.

Discussion

Loss-of-function mutations in the FLG gene contribute to the most prevalent skin disorders of cornification, ichthyosis vulgaris, and also show a strong association with AD; however, the detection of these mutations by genotyping consists of an invasive procedure that requires expensive equipment that is generally not easily available.

Raman spectroscopy is a non-invasive procedure that can provide information on the molecular structure of the skin. The analysis of the cross-correlation matrix between spectra and the PCA were used in order to extract relevant and quantitative information from the Raman spectra.

Before applying the PCA method, the cross-correlation matrix was calculated; this matrix showed that those spectra showing a cross-correlation closer to zero with the FLG spectrum were the ones corresponding to the infants who developed AD; however, spectra of subjects who remained healthy also resulted in low cross-correlation values. The advantage of using the cross-correlation method is that it does not need a database of Raman spectra. The correlation between spectra can be evaluated using only

the measured spectrum and the Raman spectrum of FLG.

A new approach to improve the specificity that involves PCA was proposed to identify those principal components related to the FLG spectrum. These selected principal components were used to reconstruct the measured spectra and later compared with the original ones. The difference between the reconstructed and the original data represents the lack of FLG in the spectra. Comparison of these results with the clinical evaluation 1 year after the measurements showed that those subjects with less FLG developed AD. The advantage of using this method is that it does not depend strongly on the set of data and it was useful in identifying the subjects who later developed the disease.

A third method used to analyze the content of FLG in the Raman spectra was performed by constructing a two-dimensional graph of the first two principal components of the dataset where the FLG made a lower contribution. Using this graph, the subjects who later developed the disease were the ones who were located farther apart from the FLG spectrum.

From the findings reported above, it can be concluded that it is possible to determine which infants are more susceptible to develop AD by taking a Raman spectra of their skin at birth, when no sign or symptom of the disease had developed, and analyzing the content of the protein FLG by any of the three different statistical analysis techniques presented. This quantitative approach could be used as an early detection procedure for FLG-related AD and as a possible marker for FLG gene mutations.

Acknowledgements

This work was completed during a stay of Javier Alda at the Universidad Autonoma de San Luis Potosi (Mexico). This stay has been funded by the Consejo Nacional de Ciencia y Tecnología (CONACyT) under Grant CB-2006-60349, by Fondos Mixtos-San Luis Potosi (FOMIX-SLP) under Grant FMSLP-C01-87127, and by the Spanish Ministry of Education and Science under the project TEC2006-01882.

References

1. Kezic S, Kemperman PMJH, Koster ES et al. Loss-of-function mutations in the filaggrin gene lead to reduced level of natural moisturizing factor in the stratum corneum. J Invest Dermatol 2008; 128: 2117–2119.
2. Nemoto-Hasebe I, Akiyama M, Nomura T, Sandilands A, McLean WHI, Shimizu H. Clinical severity correlates with impaired barrier in filaggrin-related eczema. J Invest Dermatol 2009; 129: 682–689.
3. Weidinger S, Illig T, Baurecht H et al. Loss-of-function variations within the filaggrin gene predispose for atopic dermatitis with allergic sensitizations. J Allergy Clin Immunol 2006; 118: 214–219.
4. Jollifee IT. Principal component analysis, 2nd edn. New York: Springer-Verlag, 2002.
5. Krafft C, Steiner G, Beleites C, Salzer R. Disease recognition by infrared and Raman spectroscopy. J Biophoton 2009; 2: 13–28.
6. Hannifin JM, Rajka G. Diagnostic features of atopic dermatitis. Acta Derm Venereol 1980; 92: 44–47.
7. Williams HC et al. The U.K. Working Party's diagnostic criteria for atopic dermatitis. I. Derivation of a minimum set of discriminators for atopic dermatitis. Br J Dermatol 1994; 131: 383–396.
8. Morrison DF. Multivariate statistical methods, 3rd edn. Singapore: McGraw-Hill, 1990.
9. Rothe H, Duparr A, Jacobs S. Generic detrending of surface profiles. Opt Eng 1994; 33: 3023–3030.
10. López-Alonso JM, Alda J, Bernabeu E. Principal component characterization of noise for infrared images. Appl Opt 2002; 41: 320–331.
11. López-Alonso JM, Alda J. Operational parametrization of the $1/f$ noise of a sequence of frames by means of the principal component analysis in focal plane arrays. Opt Eng 2004; 42: 1915–1922.

Address:
F. J. González
Coordinación para la Innovación y la Aplicación de la Ciencia y la Tecnología, Universidad Autónoma de San Luis Potosí, Sierra Leona 550, Lomas 2da. Sección, 78210, San Luís Potosí, SLP, México
Tel: +52 444 825 0183 ext 232
Fax: +52 444 825 0198
e-mail: javier.gonzalez@uaslp.mx

Author Query Form

Journal	**SRT**
Article	**461**

Dear Author,

During the copy-editing of your paper, the following queries arose. Please respond to these by marking up your proofs with the necessary changes/additions. Please write your answers clearly on the query sheet if there is insufficient space on the page proofs. If returning the proof by fax do not write too close to the paper's edge. Please remember that illegible mark-ups may delay publication.

Query No.	Description	Author Response
Q1	Author: If there are 10 or more authors, please list the first three authors, then et al. If there are 9 authors or fewer, please list all authors for the reference [7].	