

Fast Differentiation of Bacteria causing Pharyngitis by Low Resolution Raman Spectroscopy and PLS-Discriminant Analysis

Cesar Mello,^{*a} Eza Sevéri,^a Emiliane Ricci,^a Antônio Marangoni,^a Lucinda Coelho,^a
Diórginis Ribeiro^b and Ronei J. Poppi^b

^aInstituto de Química, Universidade de Franca, Av. Dr. Armando Salles de Oliveira, 201,
CP 82, 14404-600 Franca - SP, Brazil

^bInstituto de Química, Universidade Estadual de Campinas, Cidade Universitária Zefferino Vaz s/n,
CP 6154, 13084-971 Campinas - SP, Brazil

A diferenciação de bactérias causadoras de faringites através de métodos microbiológicos clássicos é muito eficiente, na maior parte dos casos. Todavia, o elevado custo dos reagentes e o tempo necessário para tais determinações, cerca de 4 dias, podem causar sérias conseqüências, quando os pacientes são crianças, idosos ou adultos com baixa resistência imunológica. Assim, a pesquisa por métodos espectroscópicos de baixo custo, que permitam tais determinações com pouco uso de reagentes e em curtos intervalos de tempo é extremamente relevante. Neste trabalho os principais microrganismos causadores de faringites, *S. aureus*, *S. pyogenes* and *N. gonorrhoeae* foram avaliados. Foram preparadas sessenta dispersões para cada um dos microrganismos, usando solução fisiológica como solvente, e seus espectros adquiridos. Os espectros Raman foram obtidos usando um laser de diodo operando na região do infravermelho próximo. A análise dos espectros foi feita usando PLS-discriminante. Esta abordagem permitiu classificar corretamente 100% de todas as bactérias avaliadas e amostras reais provenientes do laboratório de análises clínicas, em reduzido intervalo de tempo (ca. 10 h), com uso de espectrômetro Raman portátil, que pode ser facilmente usado em Unidades de Tratamento Intensivo (UTI) e ambientes clínicos.

The diagnosis of the bacteria that cause pharyngitis through classical microbiological methods is a difficult, but usually very efficient task. However, the high cost of reagents and the time required for such identifications, about four days, could generate serious consequences, mainly when the patients concerned are children, the elderly or adults with low resistance. Thus, the search for new methods allowing a fast and reagentless differentiation of these microorganisms is extremely relevant. In this work, the main microorganisms responsible for pharyngitis, *S. aureus*, *S. pyogenes* and *N. gonorrhoeae* were studied. For each microorganism, sixty different dispersions were prepared using physiological solution as solvent. The Raman Spectra of these dispersions were recorded using a diode laser operating in the near infrared region. The PLS-discriminant analysis method was applied to bacteria differentiation through their respective Raman Spectra. This approach enabled the correct classification of 100% of all evaluated bacteria and unknown samples coming from clinical environment, in a reduced time interval (ca. 10 h), by using a low-cost, portable Raman spectrometer, which can be easily used in Intensive Care Units (ICU) and clinical environments.

Keywords: PLS-discriminant, Raman spectroscopy, pharyngitis

Introduction

A very common situation during the winter is the high

prevalence of bacterial pharyngitis. These infections are highly prevalent in children, the elderly and adult patients with some form of immunodeficiency or organic impairment, who may be severely compromised or may even, have lethal outcomes caused by the infection. The main difficulty found

*e-mail: camello@unifran.br

in the treatment of such cases is the long time necessary for the identification of the etiological agent responsible for the infection, so that the best antibiotic can be selected. Usually, the time necessary for the identification of the bacteria type by classical biological methods is approximately four days.¹ Any delay in identifying the etiological agent may produce severe consequences or even the patient's death. While the etiological agent is being identified, the clinical procedure is the simultaneous usage, by trial and error, of two or more antibiotics, based on theoretical clinical evidence. Although this procedure is necessary, it may not be effective and may even increase the resistance of bacteria to many types of antibiotics. The emergence of multi-resistant bacteria in our environment is a great risk since they can generate epidemic outbreaks that are difficult to control. Therefore, the development of new analytical methods that will allow the fast differentiation of bacteria in solid or liquid culture media is an intensively active research field.²

Nowadays, some molecular biology techniques capable of identifying pathogenic or non-pathogenic microorganisms have been developed.^{3,4} Most of these methods use basically DNA amplification methods^{5,6} and immune-electrophoresis^{7,8} to identify the type of microorganisms. However, contamination during the DNA amplification phase may produce false positive or false negative results.

Methods using infrared and Raman microspectrometry⁹⁻¹³ are also being developed, with excellent results for the discrimination of microorganisms in shorter time intervals (*ca.* 6 h). However, their application in clinical environments, intensive therapy centers and clinical analyses laboratories is limited by the cost of the microspectrometer. Methods using Surface Enhanced Raman Spectroscopy (SERS)¹⁴ are also being studied, leading to good results in the discrimination of microorganisms in short time intervals. Again, the high cost of the instrument and the chemical manipulations necessary for the obtention of a reproducible rugged surface to increase the Raman signal make their application difficult in clinical analyses laboratories and intensive care units. Thus, the development of simple spectroscopic methods involving minimum sample manipulation and no special reagents, as well as low-cost spectrometers is a research field in intense activity.

Low Resolution Raman Spectroscopy (LRRS) is a good alternative for the quantitative or qualitative analytical applications of Raman Spectroscopy, considering its low cost, portability, the use of a laser in the near infrared region, and the simplicity in coupling it with optical fibers. Such coupling makes *on-line* and *in situ* measurements possible in virtually any kind of environmental condition. Even though all the spectral features are not necessarily cleanly resolved with either near-IR or LRRS (as in near-IR, the bands observed in

LRRS are also broad-band due to the low resolution of the Raman spectrometer), the ability to use vibrational bands as fundamentals gives LRRS an inherent advantage over near-IR, which is a well-established spectroscopic technique widely used in Analytical Chemistry.¹⁵ Besides, contrary to near-IR spectroscopy, the strong interference of water absorption is not observed in LRRS, increasing its possibilities for analytical applications, mainly in the biomedical area. Finally, as in the case of near-IR spectroscopy, chemometric methods are required to obtain quantitative or qualitative information from LRRS spectra.

Chemometric methods for qualitative analysis are generally based on discriminant analysis.¹⁶ The Partial Least Square Discriminant Analysis (PLS-DA) objective, in our case, is to determine the differences between closely related Raman Spectra groups. A way of carrying out a discriminant analysis is to generate dummy variables representing the previously defined groups of bacteria. A dummy variable is one that can take only two or three values. They are used to indicate if the corresponding spectrum has the property (in our case, of belonging to the corresponding group) that the dummy variable represents. A model based on Partial Least Square (PLS) regression is then fit using the Low Resolution Raman Spectra as predictors, and the dummy variables as responses. This procedure is called PLS-DA.¹⁷ In this work, PLS-DA was applied to differentiate the evaluated bacteria through their Raman Spectra, using [1 0 0], [0 2 0] and [0 0 3] as dummy variables *Staphylococcus aureus*, *Streptococcus pyogenes* and *Neisseria gonorrhoeae*, respectively.

Thus, this work presents an alternative method to fast bacteria differentiation, using PLS-DA in conjunction with LRRS. This method fills all the requirements for a fast differentiation of evaluated microorganisms in clinical environments and Intensive Care Units (ICU): minimum sample preparation, portable Raman spectrometers, quickness, accuracy, automatic analyses and minimum use of reagents.

Experimental

Preparations of culture media and isotonic solution

The solid culture media Agar Mueller Hinton was prepared by dissolution of 17.0 g of Agar in 500 mL of deionized water in an Erlenmeyer, after sterilization in the autoclave at 121 °C, for 15 min. The distribution of culture media was done in sterilized Petri dishes under laminar flux, at 50 °C, by means of a sterilized 20 mL pipette. After complete solidification of the culture media, the dishes were incubated in a bacteriological stove at 36.5 °C for 24 h. The saline solution NaCl 0.85% was prepared by diluting 0.850 g of NaCl in 100 mL of deionized water. This solution

was distributed in capped tubes by using graduated pipettes sterilized in autoclave at 121 °C, for 15 min.

Bacterial sowing and growth conditions

In this work, the same experimental procedure was used for *Staphylococcus aureus*, *Streptococcus pyogenes* and *Neisseria gonorrhoeae*. Therefore, in this section we describe the sowing and growth methods for *S. aureus* only. The *Staphylococcus aureus* bacterium was sowed in a Mueller Hinton (MH) liquid culture medium equipped with a bacteriological disposable handle, and it was then incubated for growth in a bacteriological stove at 36.5 °C, for 18 h. After this period, the *Staphylococcus aureus* bacterium was transferred to a MH liquid culture medium again, and sowed in an Agar-MH solid culture medium by means of the draining technique.¹ Then, it was incubated again in a bacteriological stove for growth, at 36.5 °C, for 10 h, aerobically. After this period, the biomass was carefully collected by using sterile plastic inoculating loops, and dispersed in isotonic physiological solution. Knowing that when bacteria are dispersed on an isotonic media the solution becomes turbid, the number of bacteria dispersed in the isotonic solutions was estimated by using McFarland's Scale¹⁸ (which is basically a nefelometric scale consisting of barium sulfate particles dispersed in isotonic solution). The comparison between the turbidity of the bacteria dispersions with the McFarland's standards shown in Table 1 enabled the obtention of a stock dispersion containing approximately 5×10^8 bacteria/mL. This stock dispersion was prepared for each bacterial type separately.

Sixty successive 25 µL increments of stock dispersion were added to a quartz cuvette containing 1.1 mL of isotonic solution. The option for this stock dispersion was made after initial tests, so as to avoid the high gradients of Rayleigh scattering due to successive additions and the weak Raman signal due to diluted dispersions. By following the aforementioned procedure, different dispersions ranging from 0 to 1.4×10^8 bacteria/mL, with increments of 4.0×10^7 bacteria/mL for each addition were obtained.

Sample set: calibration, validation and unknown samples

The sample sets for each bacterial type consisted of sixty bacterium dispersions split in three independent sample set, 40 for calibration, 10 for validation and 10 for

prediction. For prediction, five unknown samples were used for each bacterial type. The unknown samples were provided by the Laboratory of Clinical Microbiology of the University of Franca. In the latter case, each bacterium type was isolated from the stool of five infected patients. Then, each isolated strain was sowed in Petri dishes containing Agar-Mueller Hinton solid culture medium, and incubated for growth in a bacteriological stove at 36.5 °C, for 10 h. After that, the biomass of each strain was carefully collected by using sterile plastic inoculating loops, and dispersed in isotonic physiological solution until the dispersion achieved approximately 8.0×10^7 bacteria/mL. This concentration was selected to avoid concentration extrapolation, once it is important to build a chemometric local model, the PLS-DA.

Raman measurements and sample set

Raman spectra were collected using an OceanOptics low resolution Raman spectrometer (Dunedin, FL, USA) mod. R-2001, with resolution of approximately 15 cm^{-1} , coupled to a near-infrared 785 nm multimode diode Laser adjusted to deliver 300 mW on the sample, and a thermoelectrically cooled 2048-element CCD array detector to measure spectra from 200 to 2800 cm^{-1} . The instrument was wavelength-calibrated with isopropyl alcohol and the dark current was subtracted from all the acquired spectra.

Three low resolution Raman spectra were acquired for each of the bacterial dispersions and for all the bacteria evaluated in this work, and the final spectrum was taken as the average of these three spectra, resulting in 60 spectra for each bacterial type, one summed spectra for each of 60 additions. As for the unknown samples, three spectra were acquired for each of them, and the final spectrum was taken as the average of these three spectra, resulting in 5 spectra for each sample.

The Rbase software version 3.0.1 (Raman Systems Inc., Watertown, MA, USA) running under Windows XP professional was used for spectrometer control and data capture.

Spectra pre-processing

The spectra were preprocessed in the following order: noise minimization by means of a wavelet¹⁹ filter with a

Table 1. McFarland's Scale

Turbidity number	0.5	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0	9.0	10.0
Approximate number of bacteria/mL $\times 10^8$	1.5	3.0	6.0	9.0	12.0	15.0	18.0	21.0	24.0	27.0	30.0

Daubachie (db4) base function, and subtraction of isotonic medium and cuvette spectrum using the Gram-Schmidt method.²⁰ Figure 1 presents the processed spectra of all bacteria studied in this work.

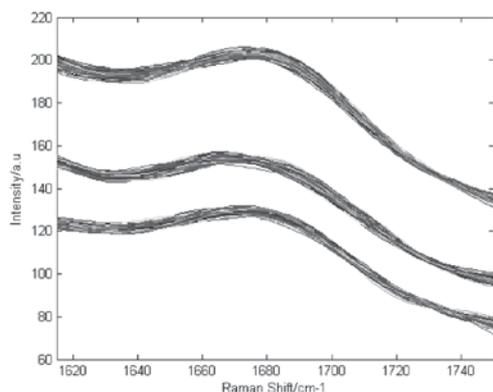


Figure 1. Typical low resolution Raman Spectra after pre-processing showing the three spectral sets used in this work plotted together. The spectra are vertically offset, for better visualization: (A) *Staphylococcus aureus*; (B) *Streptococcus pyogenes* and (C) *Neisseria gonorrhoeae*.

Computer programs

The programs for noise minimization, Gram-Schmidt method, and 2D calculations/visualizations were implemented by utilizing sub-routines from Matlab 4.0.

Results and Discussion

In Figure 1, the spectra are presented in the range 1615 to 1750 cm^{-1} , since there was high matrix fluorescence below 1615 cm^{-1} and no important signal was detected above 1750 cm^{-1} . Moreover, we wished to focus our analysis on the sulfide bands, which fall in this spectral region. By analyzing Figure 1, it is possible to observe that spectral changes after each bacterium addition are relatively small, since the increase in bacterium concentration in the solution after each addition is small, approximately 4×10^7 bacteria/mL.

Although the three types of bacterium display some particular spectral features, an attempt to differentiate between these bacteria through the meticulous analysis of the spectra by the naked eye is cumbersome due to band overlapping. Thus, the Principal Components Analysis (PCA) was applied so that the possible formation of Raman Spectra clusters related to the different types of evaluated bacteria could be observed, as shown in Figure 2.

It can be observed from Figure 2 that three very distinct groups were formed. Once the formation of the different grouping was observed, it was possible to build a PLS-DA model of classification for such bacteria. This was done by

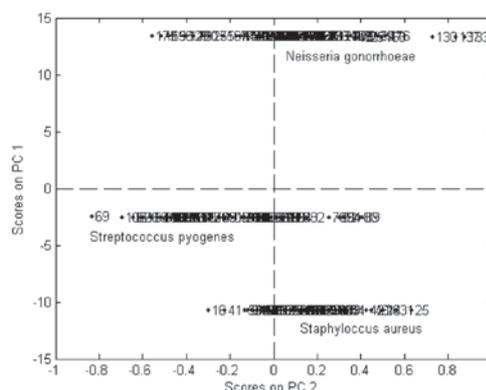


Figure 2. Plot of scores on PC 1 vs. scores on PC 2, for all the bacteria evaluated in this work.

attributing independent variables in the PLS-DA model. The following dummy variables were attributed: [1 0 0] to *Staphylococcus aureus*, [0 2 0] to *Streptococcus pyogenes*, and [0 0 3] to *Neisseria gonorrhoeae*. The PLS-DA model was built by using 4 Latent Variables (LV's) obtained by cross-validation, as shown in Figure 3. LV's were estimated as linear combinations of the wavenumbers that actually drive the spectral changes, and PRESS was the prediction residual error sum of squares. In this model, 96.1% and 97.5% of the variance were captured in the group of independent and dependent variables, respectively.

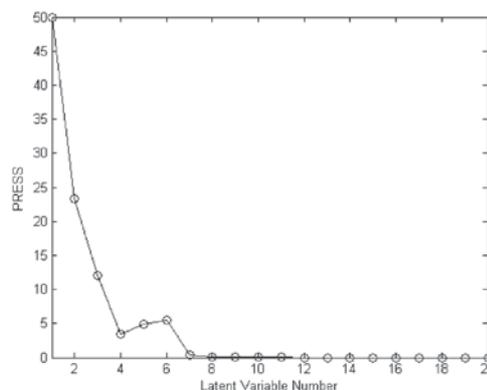


Figure 3. Plot of prediction residual sum of squares (PRESS) against the number of latent variables (LV's), where minimum PRESS and minimum number of LV's possible occur at 4 LV's.

Once the optimal number of LV's for the calibration set was selected, the minimum PRESS observed in Figure 4 and the minimum number of LV's were simultaneously employed to build a PLS-DA model for the classification of the evaluated bacteria, using 3 LV's. The results for calibration, prediction and unknown samples sets are shown in Figure 4, which presents a plot of the real against the predicted values of each class.

Analysis of Figure 4 reveals three distinct groups of spectra, indicating the possible use of PLS-DA with

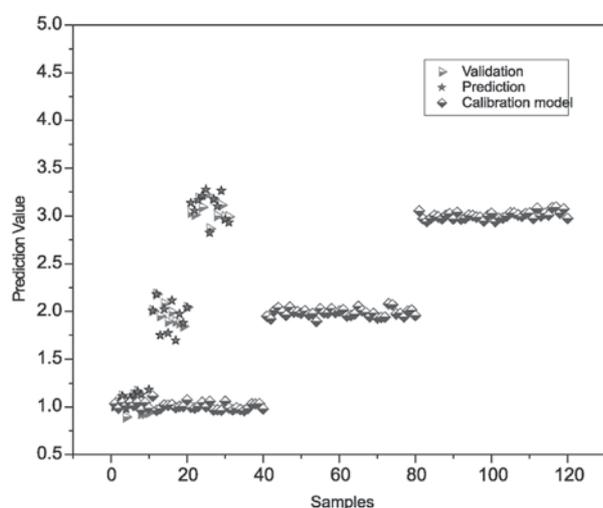


Figure 4. Plot of actual against predicted value using PLS-DA for calibration, prediction and predicted samples, plotted together for better visualization.

4 LV's to differentiate between the bacteria analyzed in this work. Moreover, the good agreement between the actual, validation and predict values, shown in Figure 4, enabled us to round the values obtained with the model toward the nearest integer value.

Analysis of Figure 4 reveals that the PLS-DA model with 4 latent variables was able to correctly classify 100% of all the types of bacteria in the calibration, prediction and unknown samples sets.

The differences captured by the PLS-DA model are basically due to the different conformations of the sulfide group in the cellular wall of the bacteria.^{22,23} This band reflects the fact that the cell wall of the bacteria used in this work consists of peptidoglycan crosslinked by polypeptide bridges, which varies among different species with changes in the conformation of the bacterial wall. Since localized sulfide mode frequencies in a given polypeptide are strongly affected by localized inter-peptide interaction and solvation, the electronic and molecular structures of peptide bond change due to interactions with surrounding peptides and solvent molecules.²⁴⁻²⁶

This fact indicates that the differences observed in the PLS-DA analysis are due to solvation of the sulfide groups, which changes the conformation of these polypeptides and the inter-peptide interactions, affecting the sulfide mode frequencies. This can be used as a powerful analytical tool for the fast differentiation of bacteria in biological samples. It is also possible to speculate on the conformation of peptides (e.g. α -helix, β -sheet, random coil, etc) present in the walls of the different types of bacteria, but this type of analysis is beyond the scope of this work. The most important point here is the possibility of obtaining a

powerful analytical tool for the fast differentiation of bacteria by means of simple systematic multivariate analysis and inexpensive Raman spectrometers.

Conclusions

In this work, we have shown that the use of a standard multivariate chemometric approach such as PLS-DA can be used in conjunction with LRRS to differentiate bacteria that cause pharyngitis. This method, which applies an inexpensive, portable Raman spectrometer in conjunction with multivariate chemometric method, can lead to the fast discrimination of pathogenic microorganisms in Intensive Therapy Centers and clinical environment, in a simple, inexpensive way.

Acknowledgments

The authors thank Coordenadoria de Aperfeiçoamento de Pessoal de Ensino Superior (CAPES) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) for fellowships and financial support of their research projects. Thanks are also given to Professor C. H. Collins, for her critical reading of the manuscript.

References

1. Koneman, E. W.; Allen, S. D.; Janda, W. M.; Schreckenberger, P. C.; Win Jr., W. C.; *Color Atlas and Textbook of Diagnostic Microbiology*, 5th ed, Lippincott, Raven Publishers: New York, 1997.
2. Mulvaney, S. P.; Keating, C. D.; *Anal. Chem.* **2000**, *72*, 153R.
3. Tang, Y. W.; Ellis, N. M.; Hopkins, M. K.; Smith, D. H.; Dodge, D. E.; Persing, D. H.; *J. Clin. Microbiol.* **1998**, *36*, 3674.
4. Tang, Y. W.; Procop, G. W.; Persing, D. H.; *Clin. Chem.* **1997**, *43*, 2021.
5. Fredricks, D. N.; Relman, D. A.; *J. Clin. Microbiol.* **1998**, *36*, 2810.
6. Mullis, K. B.; Faloona, F. A.; *Methods Enzymol.* **1987**, *155*, 49.
7. Grabar, P.; Burtin, P.; *Immuno-electrophoresis*, Elsevier: Amsterdam, 1964.
8. Bishop, M. L.; Duben-Engelkirk, J. L.; Fody, E. P.; *Clinical Chemistry: Principles, Procedures and Correlations*, Lippincott, Williams & Wilkins: London, 2000, 4th ed.
9. Maquelin, K.; Choo-Smith, L. P.; Van Vreeswijk, T.; Endtz, H. Ph.; Smith, B.; Bennet, R.; Bruining, H. A.; Puppels, G. J.; *Anal. Chem.* **2000**, *72*, 12.
10. Naumann, D.; Helm, D.; Labischinski, H.; *Nature* **1991**, *351*, 81.
11. Helm, D.; Labischinski, H.; Naumann, D.; *J. Microbiol. Methods* **1991**, *14*, 127.

12. Choo-Smith, L. P.; Marquelin, K.; van Vreeswijk, T.; Bruining, H. A.; Puppels, G. J.; Ngo Thi, N. A.; Kirschner, C.; Naumann, D.; Ami, D.; Villa, A. M.; Orsini, F.; Doglia, S. M.; Lamfarraj, H.; Sockalingum, G. D.; Manfait, M.; Allouch, P.; Endtz, H. P.; *Appl. Environ. Microbiol.* **2001**, *67*, 1461.
13. Kirschner, C.; Marquelin, K.; Pina, P.; Ngo Thi, N. A.; Choo-Smith, L. P.; Sockalingum, G. D.; Sandt, G. D.; Ami, D.; Orsini, F.; Doglia, S. M.; Allouch, P.; Manfait, M.; Puppels, G. J.; Naumann, D.; *J. Clin. Microbiol.* **2001**, *39*, 1763.
14. Goodacre, R.; Jarvis, R. M.; *Anal. Chem.* **2004**, *76*, 40.
15. Clarke, R. H.; Londhe, S.; Womble, M. E.; *Spectroscopy* **1998**, *13*, 28.
16. Vanderginste, B. G. M.; Massart, D. L.; Buydens, L. M. C.; de Jong, S.; Lewi, P. J.; Smeyers-Verbeke, J.; *Handbook of Chemometrics and Qualimetrics: Part B*, Elsevier: Amsterdam, 1998.
17. Louwerse, D. J.; Tates, A. A.; Smilde, A. K.; Koot, G. L. M.; Berndt, H.; *Chemometrics Intell. Lab. Syst.* **1999**, *46*, 197.
18. MacFaddin, J. F.; *Biochemical Tests for Identification of Medical Bacteria*, Lippincott, Williams & Wilkins: London, 2000.
19. Walczak, B.; Den Bogaert, B. V.; Massart, D. L.; *Anal. Chem.* **1996**, *68*, 1742.
20. Arfken, G. B.; Weber, J. H.; *Mathematical Methods for Physicists*, 4th ed, Academic Press: San Diego, California, 1995.
21. Wise, B. M.; Gallagher, N. B.; *PLS_Toolbox for use with Matlab, Version 1.5.1*, Eigenvector Technologies: Manson, 1995.
22. Harris, P.I.; Chapman, D. In *Infrared Spectroscopy of Biomolecules*; Mantsch, H.H.; Chapman, D., eds., John Wiley & Sons: New York, 1996.
23. Miyazawa, T.; *Poly- α -Amino Acids*, Marcel Dekker: New York, 1967.
24. Krimm, S.; Bandekar, J.; *Adv. Protein Chem.* **1986**, *38*, 181.
25. Ham, S.; Kim, J. -H.; Lee, H.; Cho, M.; *J. Chem. Phys.* **2003**, *118*, 3491.
26. Ham, S.; Cho, M.; *J. Chem. Phys.* **2003**, *118*, 6915.

Received: March 23, 2007

Published on the web: January 18, 2008

FAPESP helped in meeting the publication costs of this article.