Simultaneous Determination of Ca, P, Na, K, Mg, Cl, Urea, Glucose, and Creatinine in Urine by Energy Dispersive X-Ray Spectrometry and Partial Least Squares Regression

Gisele G. Bortoleto^{a*}, Karen Goraieb^a, Rafael A. Sousa^b, and Maria Izabel M. S. Bueno^a

^a University of Campinas, Institute of Chemistry, Department of Analytical Chemistry, 13083-971 Campinas, SP Brazil

^b Department of Chemistry, Federal University of Juiz de Fora, 36036-900 Juiz de Fora - MG, Brazil

INTRODUCTION

Urine is considered a particularly important biological fluid since its elemental composition reflects kidney performance in regulating electrolyte and water metabolism, and also because it is the vehicle that carries off many superfluous and toxic substances from the body (1).

Some papers dealing with urine analysis have pointed out the advantages of the multielemental techniques such as inductively coupled plasma optical emission spectrometry (ICP-OES) and inductively coupled plasma mass spectrometry (ICP-MS) (2-4). Nevertheless, electroanalytical and spectrophotometric techniques are the mostly applied for routine urine analysis (5), since they are very efficient and low in operational cost. The drawbacks of these latter techniques are high reagent cost and, ultimately, residue treatment. In addition, these techniques cannot simultaneously determine both inorganic and organic species.

X-ray fluorescence (XRF) is an emission qualitative and quantitative analytical method where the atoms are excited by interaction with high energy electrons, X-ray photons, and charged particles (6). The production of X-ray photons is a result of the ejection of an electron from the inner shell (K or L) of

ABSTRACT

Quantitative detection of some substances (inorganic or organic) in urine is important for the detection of an anomaly in a person's health. The aim of this study was to simultaneously determine the species most frequently monitored in clinical urine tests using X-ray spectrometry (XRS) and partial least square regression.

Urine samples were obtained from the Clinical Hospital of the University of Campinas for the evaluation by this proposed method. Reference values were obtained using inductively coupled plasma atomic emission spectrometry and the standard methods used at the Clinical Hospital. Prediction tests were also performed.

Partial least square regression is useful to establish a quantitative correlation between the instrumental responses and the content of each analyte. The obtained calibration and the cross-validation models presented good correlation, generally higher than 0.90. Prediction tests showed good results when compared with reference values. It is possible to use a simple spectrum to simultaneously determine several inorganic and organic species by XRS. The overall performance of this method was good, quantification of the analytes varied from 4.9 mg L⁻¹ (Ca) to 5549 mg L⁻¹ (Cl), presenting an average for the RSDs of 11%. The method requires minimal sample preparation, offers easy data treatment, and speed of analysis.

an atom in contact with the beam (photoelectric effect) and subsequent electronic transition between the atomic orbitals. The emission of X-ray photons is a characteristic specific for each element of the Periodic Table, occurring in well-defined energies (7).

XRF is generally used to determine the multielemental chemical composition of elements with atomic number (Z) between 11 and 92. Its detectability is higher when Z is medium or high. When Z is low, XRF presents some difficulties because the absorption coefficients and the fluorescence yields are low. Depending on the spectral resolution of the detector in use, spectral overlapping can occur. In addition, light elements not only scatter the incident X-ray beam but also present little absorption of radiation. Thus, the XRF determination of organic samples, rich in carbon, is difficult because of the great amount of scattered radiation (6,7).

Scattering occurs when X-ray photons interact with the outer electrons of the target element and can be of three types: Rayleigh (coherent scattering), Compton (incoherent scattering), and Raman (8-10). Rayleigh scattering is produced by elastic interaction between the X-ray photon (from the incident beam) and the atoms of the sample without loss of energy in the collision process. But, when the scattered photons give up a small part of their energy during the collision, the scattering is referred to as Compton (7). Raman scattering depends on the incident

*Corresponding author.
E-mail: gbortoleto@gmail.com
Tel.: +55 19 3413-1086
Fax: +55 19 3413-1702

beam energies and on the atomic number, according to the observations of Suzuki (8).

Traditionally, in XRF the scattering portion is avoided, deleted, or used to correct background radiation. Thus, characteristic XRF peaks are used for elemental characterization and quantification, while the scattering region is considered a drawback in the spectrum (7).

On the other hand, by using chemometric tools, X-ray spectrometry (XRS) considers not only the characteristic elemental lines but also the scattering portion (in the case of this paper, from 17.00 to 24.00 keV). This approach makes it possible to study organic compounds by XRS (11-13).

Although the scattering region is very informative when combined with a multivariate mathematical tool, XRS has only been used since 1990, first by Wang et al. (14) and then by other research groups (11, 15-20). All observed interesting associations between the Compton/Rayleigh scatterings and sample information. According to Henrich et al. (16), these phenomena include information on atomic number, particle size, and specific mass of the irradiated compounds. Other papers have also reported some of the advantages of quantitative analysis (17-20).

The chemometric tool employed in this work is Partial Least Square Regression (PLSR), which is a bilinear and non-iterative method based on the properties of the Nonlinear Iterative Partial Least Squares (NIPALS) algorithm. It compresses the data matrix X to its most relevant factors, developing a mathematical model. These compressed variables are used as regressors for y (variable to be studied, e.g., "analyte concentration" or "density" of samples). In its basic form, denoted as PLSR 1, it is applied for one single Y variable. It can also be modi-

fied to accommodate two or more variables (PLSR 2) (21).

The aim of this paper was to simultaneously determine important inorganic elements, whose fluorescence lines do not appear in the spectra, as well as organic compounds that produce scattering. A method for the direct and simultaneous determination of Ca, P, Na, K, Mg, Cl, urea, glucose, and creatinine in urine samples, the species most frequently monitored in clinical urine tests, was developed using XRS allied to PLSR.

EXPERIMENTAL

Samples

The urine samples were obtained daily from the Clinical Hospital of the University of Campinas, Brazil (HC-UNICAMP). Aliquots of 100 mL were randomly taken from the available samples and stored up to 24 hours in volumetric vessels containing 2 mL of 0.1 mol L⁻¹ HNO₃ (Merck & Co., Philadelphia, PA, USA) at 5 °C.

X-Ray Analysis

The irradiations were performed with an EDXRF spectrometer (Shimadzu, Model EDX 700, Kyoto, Japan), equipped with a Si(Li) detector. The excitation source was a Rh X-ray tube, operated at 15 kV and 100 mA.

About 2.0 mL of each acidified urine sample was placed into measuring cups (X-cells, Chemplex Industries, Inc., Palm City, FL, USA) mounted with 2.5 µm thick Mylar™ films (DuPont Tejin Films, USA). Afterwards, the samples were irradiated for 100 seconds using a beam collimation of 10 mm.

Statistical Analysis

For each constituent of interest, several PLSR 1 models were built after applying some pre-processing such as smoothing and normalization. To proceed with the multivari-

ate calibrations, the spectral data were "mean-centered" and treated by the PLSR 1 tool, using The Unscrambler® software (version 9.2, CAMO, USA). The spectral variables were put together in a matrix X ($n \times p$), where "n" is the number of rows (corresponding to the number of samples) and "p" is the number of columns, which is equal to the number of channels or energy intervals in the spectrum. In this work, the p value in the X matrix was 2047, while the number of samples n was 16 for each optimized model. The ($n \times 1$) y vector contains the sample constituent concentrations obtained by the reference methods.

The reference concentrations of the inorganic elements (Ca, P, Na, K, Mg, and Cl) were obtained by inductively coupled plasma optical emission spectrometry (ICP-OES) as described elsewhere (22), whereas urea, glucose, and creatinine concentrations were provided by HC-UNICAMP.

Then PLSR 1 was applied to the whole spectrum (from 0 to 40.92 keV) for the determination of the inorganic species or by using only the scattering portion (from 17.00 to 24.00 keV) for the determination of the organic species. Then the "root mean square errors of validation" (RMSEV) values for each multivariate model were evaluated through the "leave-one-out" validation procedure, considering the number of latent variables used in the building of each model.

A prediction test was also applied for some external urine samples to evaluate the predictive power of the method proposed in this paper in relation to the reference values obtained by ICP-OES (for the inorganic elements) or by standard methods used at the Clinical Hospital (for glucose, urea, and creatinine).

RESULTS AND DISCUSSION

Since urine is a biological fluid, it is rich in carbon, oxygen, hydrogen, and nitrogen, having atoms with a low atomic number. Thus, they need a low amount of energy to emit X-ray peaks (less than 1 keV) to fluoresce and scattering of excess incident energy (from the X-ray beam) by the sample is eliminate. So, the smaller the atomic number, the greater the tendency to scatter incident radiation. By irradiating organic samples, scattering is the main process and the peaks related to X-ray scattering are more pronounced. Besides, in order to detect these elements, it would be necessary to use a vacuum, high irradiation times (more than 500 seconds for the sample), and a source of radiation with high intensity, such as synchrotron light. This makes the analysis extremely long, expensive, and not always available, which certainly are drawbacks for routine analysis.

From Figure 1, it can easily be seen that in conventional XRF it is impossible to obtain qualitative or quantitative information about element concentrations due to the pronounced scattering of the incident beam. The fluorescence peaks are hidden by the scattering of the incident beam. Therefore, the use of chemometrics is important to extract information from the samples.

For the inorganic compounds, PLSR 1 established a quantitative relation between the instrumental responses (the whole X-ray spectrum) and each analyte. This quantitative relation allowed the construction of PLSR 1 curves for both calibration and cross-validation models for the inorganic elements (see Figure 2, next page).

The concept of correlation refers to the linear interrelation between two variables, x and y , and can be represented by values between

Fig. 1. Superimposed X-ray spectrum for urine samples showing strong scattering signals from the Rh X-ray source.

0 and 1. A value equal to 1 corresponds to a perfect fit of the regression equation and a value equal to 0 reflects no linear relation at all. Thus, the correlation parameters were evaluated in this work (21). In Figure 2 it can be seen that a good correlation between the elemental concentrations and the XRS spectra was achieved for all inorganic elements since the calibration and cross-validation models have very similar profiles. All multivariate calibration models showed excellent profiles, and good values were achieved for the correlation coefficients (around 0.98 – 0.99). Also, excellent coefficients of validation values for most of the elements were achieved (generally higher than 0.90), except for Ca, which had the poorest validation coefficient (0.84). The data obtained for both the calibration and validation models are listed in Table I.

Bilinear modeling methods, such as PLSR, are designed for situations where co-linearity exists among the original variables. Common information in the original variables were used to build new variables that reflect the underlying (“latent”) structure. These variables are therefore called latent variables. The latent variables are estimated as linear functions of both the original variables and the observations, thereby the name bilinear (23).

Generally, the multivariate calibration models use different methodologies to determine prediction error of a model. For a model to produce satisfactory results, it must explain most of the accumulated variance of the data with a minimum number of latent variables, minimizing the noise influence. In this condition, a model can be obtained that is easy to interpret (24).

TABLE I
Correlation Coefficient Values for Both Calibration and Validation Models Obtained via PLSR 1 for All Inorganic Elements

Element	Correlation Coefficient		Element	Correlation Coefficient	
	Calibration	Validation		Calibration	Validation
Ca	0.99992	0.84705	K	0.99751	0.99558
P	0.98802	0.96488	Mg	0.96973	0.90379
Na	0.99367	0.98957	Cl	0.99982	0.99275

Fig. 2 (a - f). Multivariate calibration and validation models for the determination of (a) Ca (with 3 latent variables), (b) P (with 3 latent variables), (c) Na (with 2 latent variables), (d) K (with 2 latent variables), (e) Mg (with 5 latent variables), and (f) Cl (with 2 latent variables).

TABLE II
Correlation Coefficient Values for Calibration and Validation Models Obtained via PLSR 1 for the Organic Compounds

Compounds	Correlation Coefficient	
	Calibration	Validation
Glucose	0.99165	0.94337
Urea	0.99925	0.95103
Creatinine	0.99093	0.96018

Fig. 3. Multivariate calibration and validation models with 4 latent variables for the determination of (a) glucose, (b) urea, and (c) creatinine.

Thus, a model containing a number of latent variables lower than the ideal number will result in under fitting because it does not consider all the information contained in the data set. On the other hand, if the number of latent variables is higher than ideal, an increase in the noise influence will occur and the result is an over fitting of data. Thus, the selection of a correct number of latent variables must to be done carefully (24).

The necessary number of latent variables to construct calibration and validation models for the inorganic elements is 3 for Ca and P, 5 for Mg, and 2 is enough for Na, K, and Cl.

For quantification of the organic compounds (glucose, urea, and creatinine), as cited before, only the scattering region is used, totaling only 350 energy channels or 350 p columns in the X-matrix. The content of the organic compounds and the scattering portion also are well correlated, and construction of the multivariate models (calibration and validation models) using PLSR 1 for their determination is possible. Also, good correlation coefficients were obtained for the calibration and validation models of the three organic compounds studied (Table II). The number of latent variables needed to construct the models is 4 for all of the organic compounds.

The multivariate models of calibration and validation, obtained by PLSR 1, are shown in Figure 3.

Considering prediction tests, two different sets were used: a training set and an independent validation set. The training set is used to construct the multivariate model (this step is called “modeling”), while the validation set corresponds to unknown samples used only for prediction. Thus, the prediction error (for other samples) can be estimated from the differences observed for this validation set (25).

Table III shows the predicted results for the simultaneous determination of Ca, P, Na, K, Mg, Cl, urea, glucose, and creatinine in different urine samples. In most cases, the predicted concentrations by the XRS method were in good agreement with the reference values.

On the other hand, some determinations can present higher errors when low concentrations are measured, as observed for glucose. In this case, this result could be improved by increasing the number of replicates and using longer irradiation times. These alternatives could make the obtained models more robust.

Technically, the most important feature of this new method is the fact that no photoelectric peaks are observed for any element (Figure 1). The quantification is possible due to the fact that, besides having photon absorption/emission, the relatively low-Z elements substantially scatter the incident photon beam. The scatter profile is deeply affected by the presence and concentration of each analyte and small variations in scattering profiles correspond to the analytical signals. By processing the whole spectra (for the inorganic elements) or only the scattering portion (for the organic compounds) via PLSR 1, important sample information can be extracted, resulting in a “fingerprint” of the sample, when compared to standards. PLSR 1 demonstrated the existence of a correlation between the spectral data (from the whole XRS spectra or only from the scattering portion) and the concentration of the studied analytes.

TABLE III
Comparison Between Predicted Values Using the Calibration Models
for the Different Analytes and Reference Values
(Obtained by ICP-OES for the Elements and
by Standard Methods for the Organic Compounds)

Analyte	Sample	Predicted Values (mg L ⁻¹)	Reference Values (mg L ⁻¹)
Ca	1	4.9 ± 0.3	5.5 ± 0.3
	2	23 ± 1	24 ± 1
	3	176 ± 2	176 ± 9
P	1	103 ± 4	125 ± 6
	2	328 ± 6	305 ± 15
	3	504 ± 9	564 ± 28
Na	1	238 ± 75	310 ± 15
	2	1232 ± 58	1250 ± 62
	3	3761 ± 56	3800 ± 90
K	1	133 ± 38	144 ± 7
	2	278 ± 18	308 ± 15
	3	1703 ± 27	1746 ± 85
Mg	1	15 ± 2	31 ± 1
	2	18 ± 2	27 ± 1
	3	11 ± 5	9.6 ± 0.5
Cl	1	264 ± 41	212 ± 11
	2	1408 ± 92	1464 ± 70
	3	5549 ± 94	5878 ± 98
Urea	1	ND*	0.5 ± 0.1
	2	20 ± 3	15.5 ± 0.7
	3	50 ± 9	50 ± 3
Glucose	1	0.6 ± 0.5	0.5 ± 0.1
	2	14 ± 3	10 ± 1
	3	101 ± 14	100 ± 5
Creatinine	1	5.6 ± 4.8	8.8 ± 0.5
	2	305 ± 39	280 ± 14
	3	1116 ± 18	1120 ± 53

*ND = not detected.

CONCLUSION

The chemometric tool PLSR 1 was used to construct multivariate models that allowed the simultaneous determination of mineral elements (Ca, P, Na, K, Mg, Cl) and organic compounds (urea, glucose, and creatinine) in human urine. Good correlation coefficients for the obtained calibration and validation models were achieved, indicating good accuracy. The main advantages of the proposed method (high analytical throughput, no chemical residue, minimal sample preparation, and easy data treatment) suggest that it can be very useful for routine urine analysis.

ACKNOWLEDGMENTS

The authors acknowledge Dr. Nelci F. Höher from the Department of Clinical Pathology – HC UNICAMP for the urine samples, Dr. Solange Cadore for the ICP OES determinations, and Dr. Carol H. Collins for the English editorial review of the manuscript. GGB, RAS, and KG thank CNPq for fellowships and financial support.

Received February 23, 2011.

REFERENCES

1. J. Bennett, F. Plum, N.G. Gordon, P.K. Juha, L. M. Gerald, K. C. Robert, Cecil, Tratado de Medicina Interna, (20th ed), Koogan S.A. Guanabara, Rio de Janeiro, in Portuguese (1997).
2. L. Bizaio, F.J. Krug, Revista Analytica. 11, 34, in Portuguese (2004).
3. L. Xiaoqing, Y. Xiaogang, Y.J. Chonghuai, Inherit. Metab. Dis. 29, 86, supplement (2006).
4. A. Taylor, S. Branch, M.P. Day, M. Patriarca, M.J. White, Anal. At. Spectrom. 22, 415 (2007).
5. Fleury Laboratories. Exams Handbook of Fleury Laboratories, Laboratório Fleury S/C Ltda, São Paulo (1999).
6. D.A. Skoog, D.M. West, F.J. Holler, S.R. Crouch, Fundamentals of Analytical Chemistry (6th ed), Saunders College Publishing, Philadelphia (2004).
7. R. Jenkins, X-ray Fluorescence Spectrometry (2nd ed), Wiley-Interscience, New York, (1999).
8. T. Suzuki, J. Phys. Soc. Jpn. 22, 1139 (1967).
9. T. Suzuki, T. Kishimoto, T. Kaji, T. Suzuki, J. Phys. Soc. Jpn. 29, 730 (1970).
10. Y. Mizuno, Y. Ohmura, J. Phys. Soc. Jpn. 22, 445 (1967).
11. M.I.M.S. Bueno, T.P.O. Castro, A.M.S. Souza, E.B.S. Oliveira, A.P. Teixeira, Chemom. Intell. Lab. Syst. 78, 96 (2005).
12. F.M. Verbi, E.R. Pereira Filho, M.I.M.S. Bueno, Microchim. Acta. 150, 131 (2005).
13. G.G. Bortoleto, L.C.M. Pataca, M.I.M.S. Bueno, Anal. Chim. Acta. 539, 283 (2005).
14. Y.D. Wang, X.N. Zhao, B.R. Kowalski, Appl. Spectrosc. 44, 998 (1990).
15. T. Kessler, P. Hoffmann, T. Greve, H.M. Ortner, X-ray Spectrom. 31, 383 (2002).
16. A. Henrich, P. Hoffmann, H.M. Ortner, T. Greve, H. Itzel, Fresenius J. Anal. Chem. 368, 130 (2000).
17. K. Goraieb, A.S. Lopes, C.A. Sato, M.G. Segatelli, V.P. Silva, J.C. Verzoto, M.I.M.S. Bueno, J. Chemom. 20, 455 (2006).
18. N. Nagata, P.G. Peralta-Zamora, R.J. Poppi, C.A. Perez, M.I.M.S. Bueno, X-Ray Spectrom. 35, 79 (2006).
19. F.M.V. Pereira, E.R. Pereira-Filho, M.I.M.S. Bueno, J. Agric. Food Chem. 54, 5723 (2006).
20. M. Blanco, M. Castillo, A. Peinado, R. Beneyto, Anal. Chim. Acta. 581, 318 (2007).
21. H. Martens, T. Naes, Multivariate Calibration, Wiley, London, 1989.
22. G.G. Bortoleto, R.A. Sousa, S. Cadore, At. Spectrosc. 31, 89 (2010).
23. Tutorial of the Unscrambler 9.2 software.
24. P.J. Gemperline, Practical guide to Chemometrics (2nd ed), CRC Press Taylor & Francis, New York (2006).
25. K. Faber, B.R. Kowalski, Appl. Spectrosc. 51, 660 (1997).