Anal Bioanal Chem (2011) 400:1153–1159
DOI 10.1007/s00216-011-4853-2

ORIGINAL PAPER

Using multidimensional projection techniques for reaching a high distinguishing ability in biosensing

Fernando V. Paulovich · Rafael M. Maki · Maria C. F. de Oliveira · Marcelle C. Colhone · Fabiana R. Santos · Vanessa Migliaccio · Pietro Ciancaglini · Katia R. Perez · Rodrigo G. Stabeli · Ângelo C. Perinoto · Osvaldo N. Oliveira Jr · Valtencir Zucolotto

Received: 15 December 2010 / Revised: 9 February 2011 / Accepted: 15 February 2011 / Published online: 25 March 2011

Abstract Recent advances in the control of molecular engineering architectures have allowed unprecedented ability of molecular recognition in biosensing, with a promising impact for clinical diagnosis and environment control. The availability of large amounts of data from electrical, optical, or electrochemical measurements requires, however, sophisticated data treatment in order to optimize sensing performance. In this study, we show how an information visualization system based on projections, referred to as Projection Explorer (PEx), can be used to achieve high performance for biosensors made with nanostructured films containing immobilized antigens. As a proof of concept, various visualizations were obtained with impedance spectroscopy data from an array of sensors whose electrical response could be specific toward a given antibody (analyte) owing to molecular recognition processes. In addition to discussing the distinct methods for projection and normalization of the data, we demonstrate that an excellent distinction can be made between real samples tested positive for Chagas disease and Leishmaniasis, which could not be achieved with conventional statistical methods. Such high performance probably arose from the possibility of treating the data in the whole frequency range. Through a systematic analysis, it was inferred that Sammon's mapping with standardization to normalize the data gives the best results, where distinction could be made of blood serum samples containing 10^{-7} mg/mL of the antibody. The method inherent in PEx and the procedures for analyzing the impedance data are entirely generic and can be extended to optimize any type of sensor or biosensor.

Keywords Biosensors · Nanostructured films · Leishmaniasis · Chagas disease · Projection techniques · Information visualization

Electronic supplementary material The online version of this article (doi:10.1007/s00216-011-4853-2) contains supplementary material, which is available to authorized users.

F. V. Paulovich · R. M. Maki · M. C. F. de Oliveira
Instituto de Ciências Matemáticas e de Computação,
USP, CP 668,
São Carlos 13560-970, São Paulo, Brazil

M. C. Colhone · F. R. Santos · V. Migliaccio · P. Ciancaglini
Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, USP,
Ribeirão Preto 14040-901, São Paulo, Brazil

K. R. Perez
Departamento de Biofísica, Escola Paulista de Medicina,
Universidade Federal de São Paulo, UNIFESP,
São Paulo CEP04044-020, São Paulo, Brazil

R. G. Stabeli
Universidade Federal de Rondônia (UNIR) and Fundação
Oswaldo Cruz—Fiocruz Noroeste,
Rondônia 76801-974, Brazil

Â. C. Perinoto · O. N. Oliveira Jr (✉) · V. Zucolotto
Instituto de Física de São Carlos, USP, CP 369,
13560-970, São Carlos, São Paulo, Brazil
e-mail: chu@ifsc.usp.br

Introduction

The various tools akin to nanotechnology have allowed the fabrication of sensing units with high sensitivity, in which molecular recognition processes can be exploited, and synergy may be achieved with the distinct materials comprising the units [1–5]. Biosensors made with layer-by-layer (LbL) films [6] where enzymes or antigens are immobilized are good examples of such systems, yielding the high sensitivity and selectivity required for clinical diagnosis [7–9]. The methods of detection in these

biosensors can involve optical [10], electrochemical [11], or impedance spectroscopy [1] measurements, all of which may generate large amounts of data. Traditional statistical methods, e.g., principal component analysis (PCA) [12], are therefore normally employed to treat the data, which is essential for good distinction to be made among very similar samples [13]. A biosensor based on the extension of the electronic tongue concept [14] was so obtained where a sensor array consisting of three electrodes led to good distinction between positive and negative antibodies of the zoonose Pasteurellosis [1]. Essential for the latter result was the use of one sensing unit containing the immobilized antigen (capable of molecular recognition) for Pasteurellosis, in addition to PCA for processing the impedance spectroscopy data. This approach is obviously generic and may be applied to various types of samples.

There are however limitations in distinguishing a large number of samples because the PCA plots tend to be overcrowded. Moreover, a linear technique such as PCA is not always suitable to deal with multidimensional data (see below). This has proven to be the case in the attempts to produce biosensors that could distinguish between Leishmaniasis and Chagas disease here. In order to circumvent such difficulties, in this paper we use information visualization methods [15] that allow one to consider the whole impedance vs. frequency curves for all sensing units and all samples. Since these visualization methods have only recently been applied to analytical chemistry in general [16, 17] and biosensing in particular [18, 19], in the "Methodology" section, we provide details of the importance of projection techniques and a brief description of the Projection Explorer (PEx) system [20], before presenting our results. We shall concentrate on the use of the projection techniques to analyze the data, as many physicochemical and biological aspects of the work have been discussed in a previous paper [21].

Methodology

Sensor array and impedance measurements

Following the concept of an extended electronic tongue [1], here we employed four interdigitated electrodes as sensing units, two of which contained LbL films [6] with polyamidoamine dendrimer generation 4 (G4 PAMAM, purchased from Aldrich) layers alternated with immobilized antigens from *Leishmania amazonensis* or *Trypanosoma cruzi* protozoans (for Chagas), which are capable of recognizing specifically anti-*L. amazonensis* and anti-*T. cruzi* IgG molecules, respectively. The choice of dendrimers was based on their suitability as matrix for immobilization of proteins in biosensors, mainly because of their branched, porous structure [21]. The other two sensing units comprised a bare interdigitated electrode and an electrode coated with an LbL film made with PAMAM and poly(vinyl sulfonic acid) (PVS). The experimental procedures to produce the sensing units and measure the electrical impedance were essentially the same as in references [1, 21], and therefore only a few details are given here. The sensing units were obtained by coating an interdigitated electrode with 50 digit pairs, each having 10 μm width, 0.1 μm height, and 10 μm apart from each other, with LbL films. The latter were 6-bilayer films prepared in three architectures: (a) layers of polyamidoamine generation 4 dendrimer (PAMAM) alternated with layers of PVS, which were adsorbed during 5 min from solutions at 2.0 mg/mL for PAMAM and PVS, and (b) layers of PAMAM alternated with proteoliposome from *L. amazonensis* or *T. cruzi* protozoans, where the latter were prepared in a concentration of 0.3 mg/mL in a 5-mM TRIS buffer at pH 7.5. The adsorption time for the proteoliposomes was 15 min. The film fabrication and the impedance measurements were carried out at 20 °C.

Impedance spectroscopy data were acquired for 25 different substances (solutions at various concentrations), with nine samples each, totaling 225 instances for the data set. These substances included a buffer, used as reference, synthetic samples with purified antibodies related to *Leishmania*, *T. cruzi* and a negative control (mouse serum in the absence of antibodies against parasites), and real samples made with the total serum of infected BALBc mice with Leishmaniasis and Chagas disease. For each sample, the real and imaginary parts of the electrical impedance were measured at 58 distinct frequencies in the range from 1 Hz to 500 kHz. Though the synthetic samples could be distinguished with traditional statistical methods employing linear techniques, such as PCA, the same did not apply when the real samples were included. We have therefore resorted to the multidimensional projection techniques implemented in a suite of tools referred to as PEx [20, 22].

Use of multidimensional projections

Considering each frequency as a data attribute, or data dimension, if m frequencies are considered, a sample can be interpreted as an m-dimensional vector embedded on an $\Re^m$ definition space. In traditional statistical analysis, data instances with more than three dimensions are known as multivariate or hypervariate data. On the information visualization field such data are called multidimensional [15]. There exist several information visualization techniques to aid users on interpreting multidimensional data. Here we use projection techniques to support the analysis of similarity relationships between the samples. Typically, a projection technique maps the m-dimensional data instances

into graphical elements (spheres, circles, etc.) on a visual space (2D or 3D), retaining as much as possible the similarity relationships between data instances calculated on the $\Re^m$ space. As long as the distance measure successfully captures the dissimilarity between data samples, the resulting visual representation is capable of conveying information on data behavior. If two elements are closely placed on the visual space, one infers that the data instances they represent are very similar; otherwise the instances are less related.

Formally, let $\mathbf{X}=\{x_1,x_2,\ldots,x_n\}$ be a set of n data instances on $\Re^m$ with $\delta=\Re^m\times\Re^m \rightarrow \Re$ a distance (or dissimilarity) function between instances on the m-dimensional space. Let also $\mathbf{Y}=\{y_1,y_2,\ldots,y_n\}$ be a set of points in $\Re^p$, with $p=\{1,2,3\}$, and $d=\Re^p\times\Re^p \rightarrow \Re$ a distance (normally Euclidean) between points on the p-dimensional space. A multidimensional projection technique is an injective function $f=\mathbf{X} \rightarrow \mathbf{Y}$ which attempts to minimize the differences between δ and d, i.e., to make $|\delta(x_i, x_j)-d(f(x_i), f(x_j))|$ approach zero $\forall x_i, x_j \in \mathbf{X}$.

In general, projection techniques can be split into two major groups according to the nature of the projection function used: linear techniques and non-linear techniques [20]. Linear techniques create linear combinations of the data attributes, defining them in a new orthogonal basis of lower dimension. Well-known high-precision techniques, in terms of distance preservation, are *Classical Scaling* [23] and PCA [12]. Although linear techniques are precise to reveal linear structures embedded on a data set, if an m-dimensional data space presents non-linear structures, such as clusters of arbitrary shapes or curved manifolds—which is very common on data sets with more than a dozen attributes (dimensions)—these techniques fail to properly capture such structures (for further discussion on the limitations of linear techniques see [24]). Hence, non-linear techniques are better solutions for projecting the data on a low-dimensional space. In applying non-linear techniques, one normally starts by defining a cost function based on the differences among the distances between the elements in the visual projected space and the desired distances computed between the data instances in the original m-dimensional space. Then one attempts to minimize this function. An example of non-linear technique is Sammon's mapping [25], which is used here and has a cost function given by

$$S=\frac{1}{\sum_{i<j}\delta(x_i,x_j)}\sum\frac{\left(d(y_i,y_j)-\delta(x_i,x_j)\right)^2}{\delta(x_i,x_j)}$$

where δ is a measure of the distance between two data samples x_i and x_j, and d is the distance among their projections y_i and y_j onto the visual space. S can be interpreted as a measure indicating the amount of information lost in the projection process. It is minimized using an iterative non-linear steepest descent optimization approach based on its gradient to find a (local) minimum.

The Projection Explorer system

PEx implements many projection strategies, including traditional ones from the literature, such as PCA and Sammon's mapping, while others have been recently proposed targeted at specific data domains and at handling very large data sources, such as Interactive Document Map (IDMAP) [26], Projection by Clustering [27], and the Least Square Projection [24]. With PEx, it is possible to create and interactively explore visual representations of multidimensional data sets, enabling data analysts to employ their visual ability to recognize structures or patterns based on data similarity. PEx provides different visual representations that allow observing and interacting with a set of multidimensional data with the goal of identifying groups of highly similar elements. It is available free of charge upon request.

Different projection techniques have specific properties which make them more or less effective, depending on the data set characteristics. In this study, we investigated PCA, Sammon's mapping, and IDMAP, all known to be precise techniques regarding distance preservation in the projected space. They are computationally expensive, which in this particular case was not an issue, as data consist of a reduced number of samples. As discussed in the following section, PCA did not produce good results on this data. Though only results with Sammon's mapping are shown, in several cases results from both IDMAP and Sammon's were equivalent, and in a few situations results with Sammon's were slightly better.

Results and discussion

Figure 1 illustrates how poor the distinction among the data samples is when only the bare electrode was used, which should be expected as there is no molecular recognition ability toward the analytes. As we shall discuss later on, distinction can be improved if instead of PCA we employ Sammon's mapping, as shown in Electronic Supplementary Material Figure S1. In the PCA plot of Fig. 1, each data sample (visually represented as circles) consists of a set of measures taken from a particular sensing unit. Circle colors depict the substances at a given concentration, for which measurements were taken. Different substances overlap in the PCA placement, an indication that the sensor measurements are not effective to distinguish them. In PCA, the first step is to compose an attribute covariance matrix, and

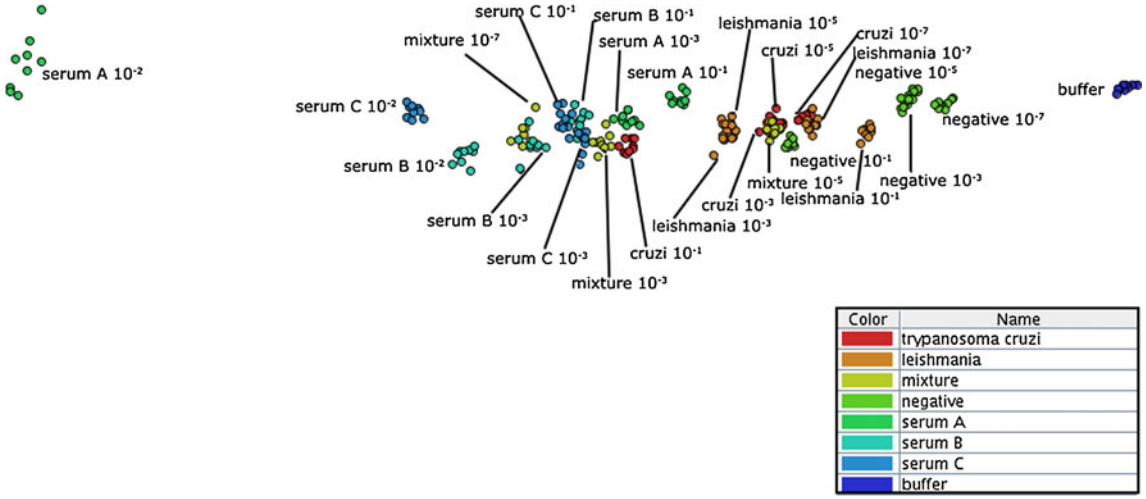

Fig. 1 PCA projection of electrical impedance data for *L. amazonensis* and *T. cruzi* samples with different concentrations of serum A (containing negative antibodies), serum B (containing anti-*Leishmania* antibodies), and serum C (containing anti-*T. cruzi* antibodies). The serum samples were obtained by dissolving the serum in the TRIS buffer. Also shown are measurements for the TRIS buffer, and the synthetic samples made with the buffer to which anti-*Leishmania*, anti-*T. cruzi* and negative antibodies were added. The samples marked with mixture are synthetic samples with anti-*Leishmania* and anti-*T. cruzi* antibodies mixed together. The measurements were obtained using only one sensor, namely the bare electrode. There is no clear separation between the different types of samples, indicating that the measured frequencies and the employed projection technique cannot distinguish these samples. Note that the axes were not labeled, as the relative distances give the degree of dis(similarity) among the samples

then apply a spectral decomposition, with the p eigenvectors with highest eigenvalues being selected to recover the Cartesian coordinates of the projected elements. An interesting property of PCA is that the resulting p-dimensional space exhibits the smallest quadratic deviation from X (the matrix describing the nm-dimensional data samples) between all possible p-dimensional spaces ($p<m$). In other words, among all matrices $A_{n \times m}$ with rank$\leq p$, the matrix resulting from the PCA process is the one that minimizes $\|(X - A)^2\|_2$ [28]. Therefore, PCA preserves as much as possible the (Euclidean) distances between the data instances when projecting them on a lower-dimensional space [23].

As mentioned above, if Sammon's mapping with prior data standardization is employed to obtain the placement for the same samples and sensing unit of Fig. 1, some improvement in substance differentiation is reached, as observed in Electronic Supplementary Material Figure S1. If now a sensing unit containing the antigen for *Leishmania* is used, a much better distinction is achieved—even without combining the response of the other sensing units—as shown in Fig. 2. The importance of introducing a data normalization procedure is as follows. On computing the distances among the m-dimensional vectors describing the data samples, if measures in one or more vector dimensions are on a different (larger) scale compared to the others, the resulting visual representations are distorted. A bias is therefore introduced into the projection process, leading to results that reflect mostly the difference between these dominant data attributes, while the discrimination power of the remaining ones is lost. In order to surpass this problem a data transformation called *standardization* can be applied to the m-dimensional vectors. In this transformation, if $\overline{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$ is the average value computed for the jth attribute and $\sigma_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{ij} - \overline{x}_j\right)^2}$ is its standard deviation, this transformation is obtained making $x'_{ij} = \left(x_{ij} - \overline{x}_j\right)/\sigma_j$ for $1 \leq i \leq n$ and $1 \leq j \leq m$, thus mapping the values of this particular attribute so that their average vanishes and their standard deviation is one [29].

A complete distinction was obtained on the data from the four sensors, converting the electrical impedance measures into capacitance (i.e., using only the imaginary part of the impedance), with the visualization that employs Sammon's mapping and data standardization. Figure 3 shows that even the real samples, corresponding to the serum of negative, *Leishmania* and *T. cruzi* antibodies, are completely separated from each other in the visual mapping of the data. Moreover, samples containing mixtures of anti-*Leishmania* and anti-*T. cruzi* antibodies could also be distinguished, and the data points are placed close to the samples containing *Leishmania*. In subsidiary visualizations, we verified that PCA—even with standardization and using all the electrodes and capacitance measures—does not yield the same level of distinguishability as Sammon's mapping, as shown in Figure S2 (see Electronic Supplementary Material). Other non-linear techniques implemented in PEx were also

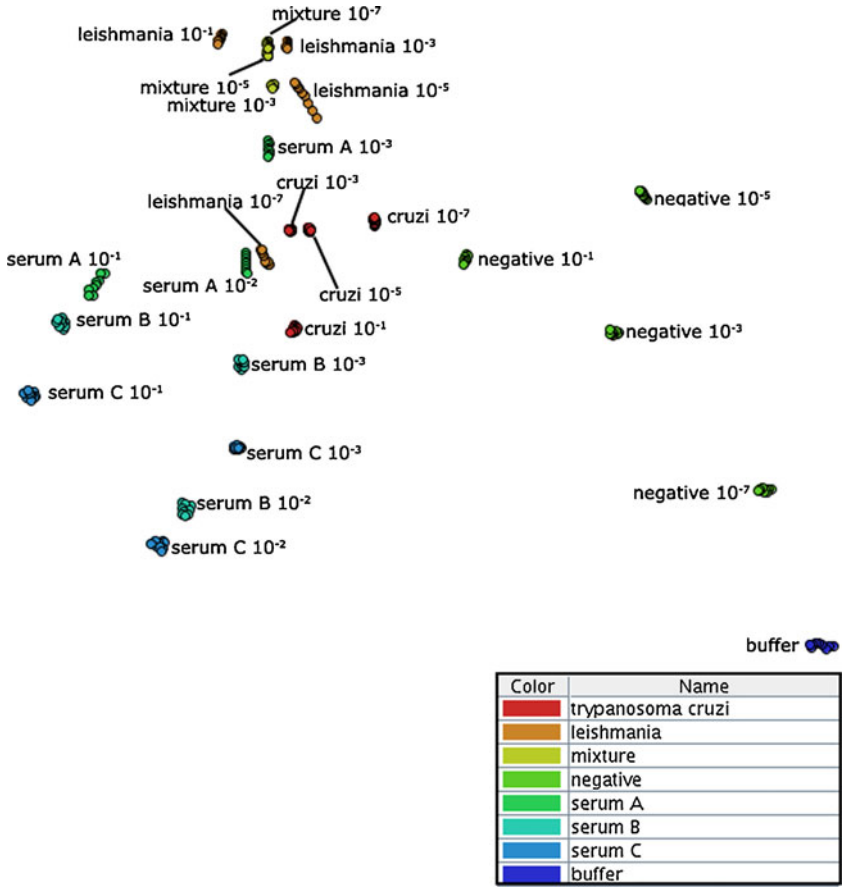

Fig. 2 Sammon's mapping with standardization for the data collected (same as in Fig. 1) with only one sensor containing the Proto *Leishmania* antigen immobilized in an LbL film with PAMAM

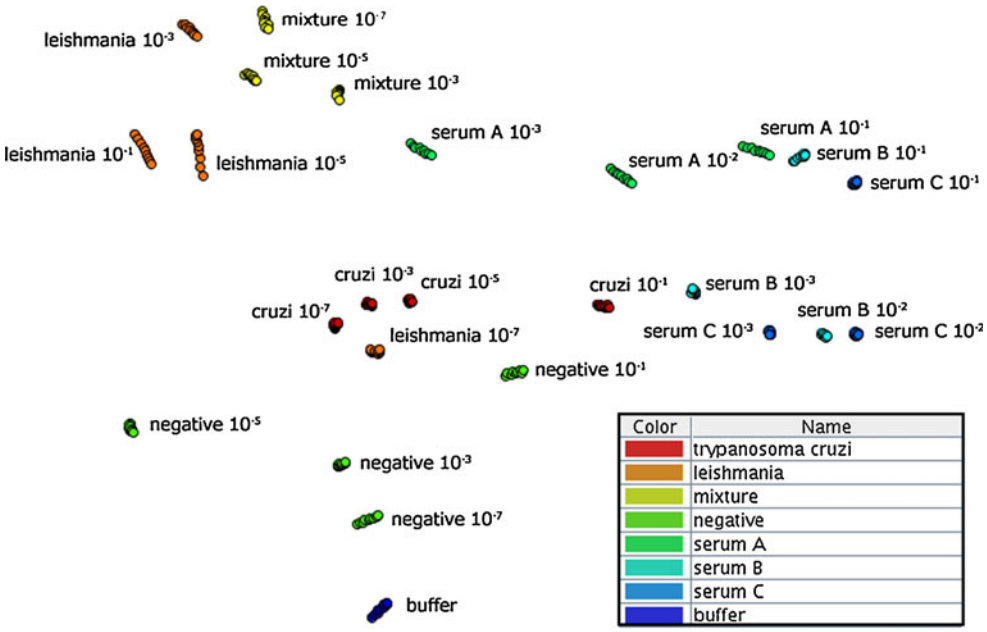

Fig. 3 Sammon's mapping projection for all the samples shown in Fig. 1, using the data from the four sensors converted into capacitance, with the standardization procedure applied to the *m*-dimensional data samples. Clear separation is observed for all samples

tried, but the results were not as good as with Sammon's mapping.

The electrical response in impedance measurements depends on several factors, including the geometric capacitance of the interdigitated electrode, the conductivity and charge storage ability of the film coating the electrode, the conductivity of the solution in contact with the film, and the double-layer formed at the film/solution interface [30]. The contribution from each of these factors is frequency dependent. For instance, at high frequencies (above 10 kHz) the response is normally dominated by the geometric capacitance of the whole system. At low frequencies (below 200 Hz), the double-layer contribution is the most important one. The latter is highly sensitive to any changes in the solution because it depends strongly on the interaction between the analytes and the film-forming materials [31]. Also worth noting was the close placement of all samples with high antibodies concentrations (0.1 M), even though still distinguishable, far away from the other data points. This means that the contribution to the electrical response from the solution predominated. In the results presented here, an efficient distinction for the serum samples was only achieved with sensing units containing antigens capable of molecular recognition toward the antibodies in the solution. With such specific interactions, one should expect the electrical response to incorporate cooperative, magnifying effects whereby trace amounts of the analyte may induce large changes in the electrical properties of the interface [32]. This may be the reason why a non-linear projection technique was successful while traditional statistical methods based on linear techniques, such as PCA, failed.

It should be said, however, that we cannot theoretically prove the superiority of non-linear methods—especially because the molecular-level mechanisms behind the so-called molecular recognition are still not well established—we can only suggest the occurrence of non-linear relationships between some data dimensions. Otherwise, the same layout quality, in terms of separating the samples, would be verifiable when employing a linear technique such as PCA. In this sense, the visual representations and the exploratory framework provided by PEx, a platform that implements several alternative data projection approaches and interactive data exploration functionalities, are valuable tools to understand the underlying data characteristics which would be difficult to discover employing only standard statistical approaches.

Finally, we emphasize that the visualization information methods are especially suitable for classification, as proven here with the good distinguishing ability between samples of Leishmaniasis and Chagas' diseases that normally lead to false positives. However, parameters such as limit of detection and resolution are not obtained in a straightforward manner, and regression techniques need to be combined with the projections. The same applies to calibration curves. For the data from the synthetic samples, i.e., those with antibodies added to a buffer, a calibration curve could be obtained using the electrical measurements without treating the data with the visualization methods [21]. One of our goals was to show that visualizations can actually tell when an automatic classification is possible based on the outcome of the sensors. For instance, we demonstrated that sample distinction was not possible with the bare electrode, but a complete distinction could be achieved by combining the data from the four sensors. In fact, the core benefit of the visualizations is allowing researchers to investigate different scenarios before resorting to analytical classification methods which require training and testing datasets and may not work well because the data are not fit for classification.

Conclusions and perspectives

The combination of sensing arrays containing immobilized antigens—capable of molecular recognition—with projection techniques allowed unprecedented distinction between serum samples infected with *Leishmania* and *T. cruzi*, which is important for public health programs in tropical countries. Because the electrical measurements take only a few minutes, this procedure represents a fast, low-cost approach for clinical diagnosis. In the proof-of-concept experiments reported here, the samples could be distinguished but a calibration was not obtained for a straightforward determination of the concentration of a new, unknown sample. Obtaining such calibration will also require data treatment methods beyond the conventional statistical methods, which again reinforces the importance of resorting to information visualization. Furthermore, one may now seek different visual representations to help researchers understand the non-linear relationships amongst data dimensions, providing additional insight on which data attributes better describe the phenomenon under study.

The methods used here are entirely generic, both with regard to the fabrication of sensing units and the data processing, therefore amenable to be exploited in diversified applications. For example, it may be used with data for sensing based on anion receptor chemistry [33], and in clinical diagnosis. In fact, a revolution may take place in clinical diagnosis when information visualization methods become commonplace, for the whole data from exams and information from the patients will be explored, rather than only employing a minute part of the data from the clinical measurements. In this context, in a recent paper, we combined multidimensional projections techniques with other information visualization methods for optimizing the

performance of a biosensor to detect phytic acid [34]. An important feature of the latter work was that the multivariate nature of impedance spectroscopy data could be exploited in the optimization, by choosing whether the real or imaginary part (or both) of the data should be used in the visualization, and selecting the frequencies that led to the best discrimination ability. That information visualization methods are particularly suitable for data of different natures was demonstrated with the analysis of image collections using the tool Projection Explorer for Images [35]. This tool incorporates several similarity-based multidimensional projections and other point placement techniques to create interactive visualizations that aid user-driven feature detection from images. Analogously to the plots shown here for data points with sensing measurements, one needs to evaluate similarity amongst all pairs of images, which may be done by computing distances on the feature space. In addition to its application to image databases, which is useful for clinical diagnosis, the paper also illustrates how the tool can be used in simultaneous analysis of different data types, such as text and images, offering a common visual representation for data expressed in different modalities.

Acknowledgments This work was supported by FAPESP, GENOPROT, CNPq, and Capes (Brazil).

References

1. Zucolotto V, Daghastanli KRP, Hayasaka CO, Riul A Jr, Ciancaglini P, Oliveira ON Jr (2007) Anal Chem 79:2163–2167
2. Wei D, Bailey MJA, Andrew P, Ryhanen T (2009) Lab Chip 9:2123–2131
3. Lvov Y, Ariga K, Ichinose I, Kunitake TJ (1995) Am Chem Soc 117:6117–6123
4. Mrksich M, Whitesides GM (1995) Trends Biotechnol 13:228–235
5. Siqueira JR Jr, Caseli L, Crespilho FN, Zucolotto V, Oliveira ON Jr (2010) Biosens Bioelectron 25:1254–1263
6. Decher G (1997) Science 277:1232–1237
7. Lu M, Lee DJ, Xue W, Cui TH (2009) Sensors & Actuators A-Phys 150:280–285
8. Ariga K, Hill JP, Ji QM (2007) Phys Chem Chem Phys 9:2319–2340
9. Ariga K, Hill JP, Lee MV, Vinu A, Charvet R, Acharya S (2008) Sci Technol Adv Mater 9:014109
10. Yuan PF, Ma Q, Meng RZ, Wang C, Dou WC, Wang GN, Su XG (2009) J Nanosci Nanotech 9:3092–3098
11. Crespilho FN, Ghica ME, Florescu M, Nart FC, Oliveira ON Jr, Brett CMA (2006) Electrochem Commun 8:1665–1670
12. Jolliffe IT (2002) Principal component analysis. Springer, New York
13. Briggman KL, Abarbanel HDI, Kristan WB (2005) Science 307:896–901
14. Riul A Jr, Dos Santos DS Jr, Wohnrath K, Di Tommazo R, Carvalho AACPLF, Fonseca FJ, Oliveira ON Jr, Taylor DM, Mattoso LHC (2002) Langmuir 18:239–245
15. Oliveira MCF, Levkowitz H (2003) IEEE Trans Vis Comput Graph 9:378–394
16. Ivosev G, Burton L, Bonner R (2008) Anal Chem 80:4933–4944
17. Wiklund S, Johansson E, Sjostrom L, Mellerowicz EJ, Edlund U, Shockcor JP, Gottfries J, Moritz T, Tryg J (2008) Anal Chem 80:115–122
18. Siqueira JR Jr, Maki RM, Paulovich FV, Werner CF, Poghossian A, De Oliveira MCF, Zucolotto V, Oliveira ON Jr, Schöning MJ (2010) Anal Chem 82:61–65
19. Moraes ML, Maki RM, Paulovich PV, Rodrigues Filho UP, De Oliveira MCF, Riul A Jr, De Souza NC, Ferreira M, Gomes HL, Oliveira ON Jr (2010) Anal Chem 82:3239–3246
20. Paulovich FV, Oliveira MCF, Minghim R (2007) The projection explorer: a flexible tool for projection-based multidimensional visualization. In: Proceedings of the XX Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2007), Washington, C, USA: IEEE Computer Society, pp. 27–36
21. Perinoto AC, Maki RM, Colhone MC, Santos FR, Migliaccio V, Daghastanli KR, Stabeli RG, Ciancaglini P, Paulovich FV, de Oliveira MCF, Oliveira ON Jr, Zucolotto V (2010) Novel biosensors for rapid and efficient diagnosis of leishmaniasis: innovations in bioanalytics for a neglected disease. Anal Chem 82:9763–9768
22. Tejada E, Minghim R, Nonato LG (2003) Information Visualization 2:218–231
23. Cox TF, Cox MAA (2001) Multidimensional scaling. Chapman & Hall/CRC, Boca Raton
24. Paulovich FV, Nonato LG, Minghim R, Levkowitz H (2008) IEEE Trans Vis Comput Graph 14:564–575
25. Sammon JW (1969) IEEE Trans Comput 18:401–409
26. Minghim R, Paulovich FV, Andrade Lopes A (2006) Content-based text mapping using multi-dimensional projections for exploration of document collections. In: Erbacher RF, Roberts JC, Gröhn MT, Börner K (eds) Proceedings of IS&T/SPIE Conference—Visualization and Data Analysis, SPIE, 2006, p. 60600S
27. Paulovich FV, Minghim R (2006) Text map explorer: a tool to create and explore document maps. In: Proceedings of the 10th International Conference on Information Visualization (IV'06), Washington, DC, USA: IEEE Computer Society, p. 245–251
28. Mardia KV, Kent JT, Bibby JM (1995) Multivariate analysis (probability and mathematical statistics). Academic, New York
29. Tan P-N, Steinbach M, Kumar V (2005) Introduction to data mining. Addison-Wesley Longman, Boston
30. Taylor DM, Macdonald AG (1987) J Phys D Appl Phys 20:1277–1283
31. Oliveira ON Jr, Riul A Jr, Leite VBP (2004) Braz J Phys 34:73–83
32. Taylor DM, Oliveira ON Jr, Morgan H (1989) Thin Solid Films 173:L141–L147
33. Caltagirone C, Gale PA (2009) Chem Soc Rev 38:520–563
34. Paulovich FV, Moraes ML, Maki RM, Ferreira M, Oliveira ON Jr, De Oliveira MCF (2011) Analyst 136(7):1344–1350
35. Eler DM, Nakazaki MY, Paulovich FV, Santos DP, Andery GF, Oliveira MCF, Batista Neto JES, Minghim R (2009) Vis Comput 25:923–937