Biointerphases (2012) 7:53
DOI 10.1007/s13758-012-0053-7

Information Visualization to Enhance Sensitivity and Selectivity in Biosensing

Osvaldo N. Oliveira Jr. · Felippe J. Pavinatto · Carlos J. L. Constantino · Fernando V. Paulovich · Maria Cristina F. de Oliveira

Received: 20 June 2012 / Accepted: 6 August 2012 / Published online: 22 August 2012

Abstract An overview is provided of the various methods for analyzing biosensing data, with emphasis on information visualization approaches such as multidimensional projection techniques. Emphasis is placed on the importance of data analysis methods, with a description of traditional techniques, including the advantages and limitations of linear and non-linear methods to generate layouts that emphasize similarity/dissimilarity relationships among data instances. Particularly important are recent methods that allow processing high-dimensional data, thus taking full advantage of the capabilities of modern equipment. In this area, now referred to as e-science, the choice of appropriate data analysis methods is crucial to enhance the sensitivity and selectivity of sensors and biosensors. Two types of systems deserving attention in this context are electronic noses and electronic tongues, which are made of sensor arrays whose electrical or electrochemical responses are combined to provide "finger print" information for aromas and tastes. Examples will also be given of unprecedented detection of tropical diseases, made possible with the use of multidimensional projection techniques. Furthermore, ways of using these techniques along with other information visualization methods to optimize biosensors will be discussed.

This article is part of the Topical Collection "In Focus: Future of Biosensors".

O. N. Oliveira Jr. (✉) · F. J. Pavinatto
Instituto de Física de São Carlos, Universidade de São Paulo, CP 369, São Carlos, SP 13560-970, Brazil
e-mail: chu@ifsc.usp.br

C. J. L. Constantino
Faculdade de Ciências e Tecnologia, UNESP Univ Estadual Paulista, Presidente Prudente, SP 19060-900, Brazil

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

1 Introduction

The generation and availability of so much information in electronic media, including scientific data, has sparked research and development of computational and statistical tools to handle such information, within a new scientific paradigm of data-driven scientific discovery—in some situations referred to as e-science [1]. E-science refers to a computationally intensive science, typically making use of highly distributed computer networks or a science dealing with large amounts of data for which grid computing is used. Working within the e-science paradigm normally involves cloud computing and parallel processing, required to handle the massive amounts of data. In a broader setting, it may refer to the application of modern computational methods of data mining, data visualization, information retrieval and other technologies for knowledge generation from data.

For sensing and biosensing, which have become ubiquitous in modern systems in our society, state-of-the-art technologies lead to massive amounts of data of various natures. Sensors and biosensors may be based on principles of detection exploring electrical, electrochemical, optical, spectroscopic properties, to name just a few [2]. In biosensing, in particular, dealing with biological systems and even in vivo experiments poses additional challenges owing to the variability of biological samples. In investigating sensor configurations, hundreds if not thousands of measurements may be performed to characterize a set of

biological samples, and a single measurement may actually consist of a spectrum of values. Therefore, in studying a single biosensing system they face many instances of the problem of pattern recognition from data, in which the goal is to identify the capability of discriminating samples, given their characterizing output by one or multiple sensor configurations.

In recent years a number of data analysis methods have been employed, with many issues of sensors design and discrimination of similar samples being dealt with methods and approaches from chemometrics [3], which is the science dedicated to data-driven discovery approaches applied to chemical systems. In this context, the methods possibly most relevant for sensing stem from multivariate data analysis (see a review on biosensors in Lindholm-Sethson et al. [4]). Typical data analysis covers both exploratory techniques such as principal component analysis (PCA) and cluster analysis for discrimination; as well as supervised techniques such as linear discriminant analysis (LDA), soft independent modeling of class analogy (SIMCA) or Partial Least Squares Discriminant Analysis (PLSDA) for classification [5]. In particular, the need of handling data from many sensor configurations simultaneously drives an interest on exploratory approaches that can help users to interactively identify the solutions that deserve further investigation. These will possibly require additional analysis with traditional supervised pattern recognition techniques, which is the central topic in this review.

In contrast to most reviews on biosensors (see [6, 7]), here we shall not dwell upon the materials for the sensing units or on the principles of detection. We shall rather concentrate on data analysis methods, particularly the exploratory data visualization techniques recently introduced in biosensing [8]. This review is organized as follows. Section 2 brings a brief introduction to the main concepts and methods of information visualization. The increasing trend toward the usage of a wider variety of data analysis methods is highlighted in Sect. 3, while the specific use of information visualization for sensing and biosensing appears in Sect. 4. Section 5 closes the paper with conclusions and outlook.

2 Exploratory Data Visualization: Concepts and Methods

The amount of data generated in different fields over the last decades has grown so substantially that data analysis represents now a major challenge. Technologies to store and retrieve data are well established and increasingly affordable, but our interpretation capacity is limited. In order to reduce the gap between data collection and data exploration, display and interpretation, use can be made of data mining and data visualization. Visual analytic techniques are attractive for complex data analysis because they generate interactive visual representations that potentially benefit from the human visual channel to speed-up interpretation of complex (large and/or high-dimensional) data [9].

Visualization methods and techniques are usually categorized into two fields: scientific visualization (*SciVis*) and information visualization (*InfoVis*) [10] (sometimes closely related with multivariate, or multidimensional data visualization, known from statistics). *SciVis* visual representations are built upon data representing objects and concepts associated with real or simulated physical phenomena, such as weather simulations or computer tomography scans. Resulting data are spatial and embedded in 1D, 2D or 3D spaces (as the objects they represent), and usually the visualization model is a straightforward representation of the geometry of the underlying objects. *InfoVis* representations are built from abstract entities that do not necessarily have a physical or geometric representation, such as census data or web pages returned from a user query. Typically, the data instances are multidimensional, describing entities that consist of multiple measurements or attributes, not necessarily of a spatial nature. While spatial objects may be associated with abstract attributes, abstract objects can also be associated with spatial attributes (e.g., demographics data are commonly associated with a 2D spatial location, or cartographic maps may display abstract entities). Therefore, the distinction between these fields is blurred. From an end user perspective, a major difference is that abstract visualizations can be more difficult to interpret, as they do not rely on familiar object representations.

Techniques applied in biosensing are typically from *InfoVis,* since the output of sensor measurements is data in a high-dimensional space, e.g., spectrum of values. Figure 1 shows a representation of the pipeline for mapping data into abstract visual representations, or the *visual mapping pipeline*, as described by Card et al. [9]. Raw data are transformed and organized into data tables, from which graphical representations are derived by means of visual mappings. Such graphical representations are then displayed to users who can interact with them as a means of exploring the underlying data. In this process, new data transformations or new visual mappings may be required. There is a wide range of visualization techniques—or visual mappings—targeted at multidimensional data, most of which adopt the overall approach of mapping each data instance to a graphical marker, which may be a single pixel, or a line or an icon. Detailed reviews of *InfoVis* techniques may be found elsewhere [10, 11]. In the following we focus initially on a specific class of techniques, known as multidimensional projections, which are proving

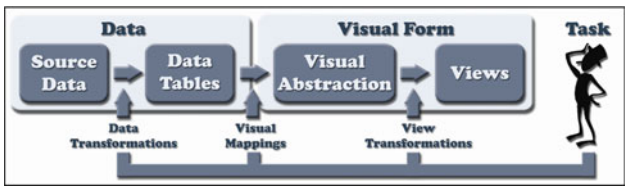

Fig. 1 Visual mapping pipeline: from data to visual abstractions and user interaction. The raw data represented in the *leftmost block* (source data) are transformed into tables that can be visualized with distinct paradigms, referred to here as visual abstraction. The process of view transformation is responsible for deciding the final format of the graphical display to the end user. The *bar* at the *bottom* was included to illustrate that the whole process can be interactive, with the user choosing the methods to transform data, visualize the mapping and even modify the views. *Adapted from* Ref. [9]

promising to create visual representations of biosensing data that afford exploratory analysis.

The goal of multidimensional projection techniques is to convey global similarity relationships amongst high-dimensional data instances by generating a two-dimensional embedding of the data. A projection technique maps each data element to a visual marker placed on a plane, so that markers depicting similar instances are placed close, whereas those depicting dissimilar instances are placed apart from each other. It requires a measure of similarity/dissimilarity to be defined, usually approximated by some distance function defined in the high-dimensional data space. Such techniques are closely related with dimensionality reduction and multidimensional scaling (MDS) [12] approaches, which are normally classified into *linear* or *non-linear techniques* [13]. Examples of linear techniques are *Principal Component Analysis* (PCA) [14] and *Classical Scaling* [12]. Linear techniques may fail to recover non-linear structures such as clusters of arbitrary shapes or curved manifolds that may be present in the data. If this is the case, non-linear dimension reduction tends to provide superior performance in projecting the data on lower-dimensional spaces.

A mathematical formulation of the projection problem follows: let $X = \{x_1, x_2, \ldots, x_n\}$ be the data set, and $\delta(x_i, x_j)$ a dissimilarity (distance) function defined between two different instances. Let $Y = \{y_1, y_2, \ldots, y_n\}$ be the set of visual markers corresponding to X, and $d(y_i, y_j)$ a distance function amongst them. A projection technique is an injective function $f: X \rightarrow Y$ which seeks to make $|\delta(x_i, x_j) - d(f(x_i), f(y_j))| \approx 0$, $\forall x_i, x_j \in X$ [15]. Different formulations of the error function and different approaches to its minimization result in several possible choices for the mapping function f. The error function is as a measure of the information lost in the projection procedure. If the mapping is effective, perceived clusters of visual markers indicate groups of highly correlated data instances (similar content), and markers placed apart and in different clusters can be related to dissimilar instances. In this review, we shall comment upon visualizations created mainly with two non-linear techniques, viz. Sammon's Mapping [16] and interactive document map (IDMAP) [17].

The error function minimized in Sammon's Mapping is given by

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

where δ is a measure of the dissimilarity between samples x_i and x_j, and d is the distance among their projections y_i and y_j onto a 2D plot.

For IDMAP, the error function is defined as

$$S_{IDMAP} = \frac{\delta(x_i, x_j) - \delta_{\min}}{\delta_{\max} - \delta_{\min}} - d(y_i, y_j)$$

where δ_{min} and δ_{max} are the minimum and maximum distances between the samples. It is based on a fast dimension reduction strategy referred to as *Fastmap* [18], which is employed to generate an initial placement of the data points that is improved with the *Force Scheme* [15], a strategy that mimics a placement approach based on simulating mass-spring models typically employed for drawing graph models [19].

The quality of the low-dimensional embedding achieved with a projection may depend on various factors, including properties of the data and behavior of the distance function, as well as user goals. Apart from very general guidelines or recommendations, it is difficult to anticipate which technique will output the best projection of a given data set, or which dissimilarity function better captures the relevant data behavior. In fact, defining which is best from a set of alternative layouts is itself a difficult research question. Another issue is computational cost, as one wants to generate two-dimensional embeddings at interactive rates. A recently published solution was shown capable to process millions of instances within minutes [20], implying feasibility to process very large datasets.

In order to illustrate how projections can be used, we show in Fig. 2 an IDMAP projection of the Iris flower dataset, available and described at the UCI Machine Learning repository.[1] This dataset is well-known to the pattern recognition, machine learning and visualization research communities, and widely employed to illustrate usage and performance of classification and visualization algorithms. It describes 150 Iris flower samples of three different species: iris virginica, iris versicolour and iris setosa, providing 50 samples from each class. Each sample flower is described by four different measures, namely *sepal length*, *sepal width*, *petal length* and *petal width*,

[1] Available at the UCI Machine Learning Repository (http://archive.ics.uci.edu/ml/).

measured in centimeters. It is known that, based on these four descriptive attributes, one class is linearly separable from the other two, which are not themselves linearly separable from each other (footnote 1).

Let us now comment on the projection in Fig. 2: in an effective projection mapping, dissimilar data samples, according to the values of their describing attributes, are positioned farther apart than samples that are more similar, which are positioned closer. Notice that, although the projection visually resembles the *scatterplots* typically employed to display the relationship between two data attributes, it has a distinct interpretation, as no attribute is being mapped to either the horizontal or vertical axes. The placement of the data samples in the two-dimensional space is relative and only indicates global proximity, or similarity. In the projection view in Fig. 2, each circle depicts a flower sample, with the color mapping flower type. It has been computed considering the four descriptive attributes simultaneously, using the Euclidean distance as an approximation of dissimilarity. Observing the color coded projection one notes that the setosa flowers are very different from the virginica and versicolour, whereas these latter two are not fully distinguishable, as some green and red samples are actually very close, i.e., similar. One infers that taking these four attributes to describe the samples may cause some flowers to be mistakenly classified as verginica or versicolour. So, other additional measures would be needed in order to correctly identify all the flowers. On the other hand, we also know the projection is effective, in that it reveals information about the data set that is known to be correct.

The previous example illustrates how a particular data sample x_i is described by multiple attributes, i.e., $x_i = \{x_{i1}, x_{i2}, \ldots, x_{im}\}$, that actually determine the global relationships amongst data instances. A visualization such as the previous one, obtained by projecting the data, does not convey the contribution of the different attributes to an observed behavior. One may resort to alternative high-dimensional data visualizations to investigate the role of attributes on data behavior. A particularly expressive technique for this goal is *parallel coordinates* [21], which again departs from the conventional approach of mapping attributes to orthogonal coordinate axes of a Cartesian plane, as in *scatterplots*. In parallel coordinates an axis is associated with each data attribute and used to map its range, but the axes are arranged in parallel on the plane. A data instance is represented as a polyline that will cross the attribute axes at the point determined by the value of the corresponding attribute. This solution enables visualizing a relatively large number of attributes on a single planar representation, since—unlike scatterplots—it can display more than two or three attributes simultaneously. It has been shown useful to highlight patterns on the data and functional dependencies amongst multiple data attributes, particularly when data sets are not too large—otherwise strong overlapping of lines can severely hamper user interpretation [10]. Later on we shall discuss how this technique has been applied, in connection with projection-based visualizations, to optimize the performance of biosensors.

Figure 3 shows a *parallel coordinates* visualization of the Iris dataset. In this view, each polyline depicts a flower sample, i.e., they correspond to the same circles shown in Fig. 2. Again, line color identifies the flower's type. The four vertical axes map the range of values of the four measurements. It is noted that sepal length and width are not suitable attributes to differentiate the flowers, since they show considerable overlap of the polylines representing flowers of different types. Therefore, it is not possible to characterize the flower only with these measurements. On the other hand, when inspecting the petal attributes one observes that different flower types have quite different measures, as indicated by the good separation of the crossing lines of different colors at the corresponding axes. This plot allows one to infer that the setosa flowers have petal length and width considerably smaller than those of verginica and versicolour, on this particular dataset, and thus it is possible to differentiate the setosa

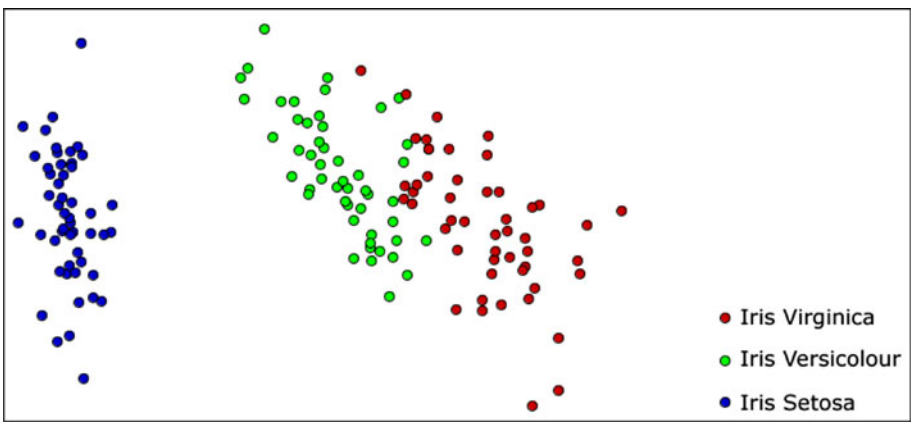

Fig. 2 Projection of the Iris flower dataset, in which each circle represents an Iris flower sample, and its color indicates the type of the flower. One observes that one type of flower (Iris setosa) is easily distinguished from the other two, which in turn are not clearly separable

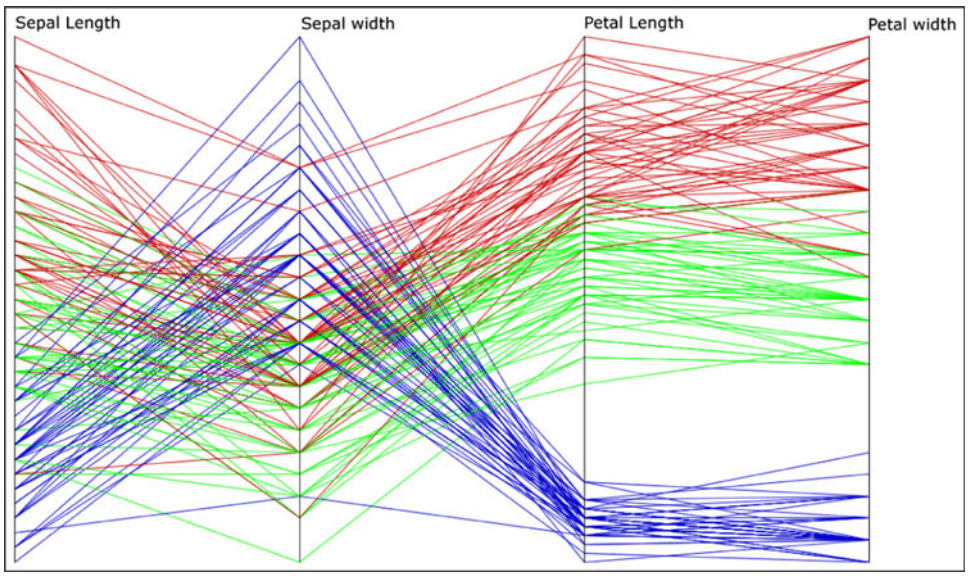

Fig. 3 Parallel coordinates visualization of the same Iris flower data set depicted in Fig. 2. Each polyline maps a flower sample, and its *color* indicates the flower type. This visualization shows that petal measurements are more effective to distinguish the flower types than the sepal measurements

flowers from the other two. Not all samples of virginica and versicolor can be distinguished, however, as there is some degree of overlap, again confirming what we know about the data.

3 Trends in the Use of Data Analysis Methods

The complexity inherent in biological, imaging and other types of sensing data has motivated application of a variety of statistical and computational methods, ranging from artificial neural networks [22] to visualization techniques [23, 24]. In a number of cases, the data are generated by a wide range of sensing devices, obtained by an equally large variety of sensor types. These may include electrical, electrochemical or optical sensors, satellite images, traffic (see for instance Medeiros et al. [25]) and spectroscopic techniques. In problems that generate large amounts of correlated data, as in the measurements in multiple brain areas obtained over time with electrode arrays, it is essential to employ sophisticated data-analysis methods. This was discussed by Reed and Kaas [26], including the challenges to analyze large-scale neuronal recording data. The final goal in this type of exam is to relate stimulus properties to the response of individual neurons and neuronal networks. The authors mentioned as one of the challenges the need to take into account the data dependencies arising from the multi-electrode recordings and consider the non-linear nature of dependency among the variables of interest.

In addition to processing huge amounts of data, sensing and biosensing systems also face the problems arising from the so-called dimensionality curse [27]. These problems may be addressed with feature selection methods [28] coupled with data cleaning and fusion. For traffic events in a major French city, Medeiros et al. [25] combined analytical methods with data management strategies to handle spatio-temporal data. Feature selection is essential in many data analysis problems, including biosensor optimization. The work by Paulovich et al. [29], for instance, deals with feature selection in the context of seeking to optimize sensor performance (this is further discussed in Sect. 4).

Sensing is also crucial for real-time monitoring of fabrication processes in the high tech industry, as in the production of semiconductor wafers. A major difficulty is to develop control systems that can both handle a lot of data in a short time period while simultaneously providing adequate feedback. This issue was discussed by Yang and Chen [30], who described optical emission spectroscopy as a suitable, noninvasive monitoring method. The major difficulty in using this spectroscopy method, however, is the huge amount of information obtained. Real-time detection of faults could be achieved by implementing a model allowing direct matching of patterns characteristic of good samples. Another example of control of fabricated structures is directly related to biosensing, in that 3-D microdomains were formed with photolithography combined with laser excimer technology [31] to serve as template for investigating cell growth. For microfluidic lab-on-a-chip, which promises to revolutionize sensing and biosensing, Yoon et al. [32] stated that full realization of the advantages of these new systems depends on implementing effective data-analysis methods. They exemplified the importance of novel approaches by introducing a pattern-mining method in the analysis of large-scale biological data obtained from high-throughput biochip experiments.

In the remainder of this section, we shall focus on two topics associated with the processing of large amounts of data, namely usage of multivariate analysis and data processing in applications related to electronic noses and tongues.

3.1 Multivariate Data Analysis

The use of computational methods has been advocated [33] for drug discovery using libraries of drug candidates integrated with data from biosensors based on surface plasmon resonance. For sensing based on impedance spectroscopy, Lindholm-Sethson et al. [34] showed the suitability of PCA to analyze data collected over a range of frequencies, for the PCA score plots could depict an objective overview of the various interactions in a complex system. They provide an indication of the presence of specific interactions that cause grouping(s) in the data and also reveal the time dependence of an interaction process and the relative size. The same applied to the combination of multivariate analysis and electrochemical impedance to study interactions with a phospholipid monolayer [35]. Furthermore, multivariate data analysis may be applied to complex number matrix representations of the impedance spectroscopy data [4], in the so-called complex number chemometrics [3]. As confirmed later in our discussion in the context of electronic tongues, Lindholm-Sethson et al. [4] argued that "multifrequency impedance data are best studied by taking all frequencies into account at once and not by studying the frequency response at each frequency separately".

In a review paper, Saurina [36] addresses recent achievements in wine characterization using chemometric analysis of physicochemical data, as identified from representative papers published in the last decade. They emphasize that data handled in wine characterization is typically multivariate in nature, comprising a list or array of values. Data thus obtained from suitable analytical methods may be combined into a data matrix in which each line refers to a wine sample, and each column describes a measured variable. This data may be treated with chemometric methods [37]. The authors listed PCA and cluster analysis as complementary techniques often adopted in exploratory studies; whereas LDA and SIMCA as techniques for classifying wines into pre-established categories or groups. Artificial Neural Networks and Partial Least Squares Regression are sometimes employed for purposes of identifying correlation, e.g., uncovering potential relationships of physicochemical variables with sensorial attributes. They survey many contributions on wine characterization, providing an extensive table that includes information on the data analyzed and the chemometric methods employed.

3.2 Electronic Tongues and Noses

Among the many systems employing multivariate data analysis, particularly relevant for biosensing are those related to electronic tongues and noses [38–56]. The latter comprise arrays of chemical sensors, whose response constitutes a taste or odor pattern, respectively. They rely on the concept of global selectivity, according to which the measurements yield a "finger print" of the liquid or vapor under study. Several kinds of sensing elements and detection methods have been studied for e-noses and mainly e-tongues [45, 51, 57–62], which allow applicability in fields as food [57, 62–66], wines [67], water [68] and pharmaceutical analysis [66]. The importance of the e-tongues and e-noses to biosensing stems from the possible extension through the incorporation of sensing units capable of molecular recognition [69–72].

The principles behind the combination of measurements to establish patterns have been discussed in [47, 73]. The latter authors mentioned the relevance of "soft" measuring techniques, i.e., ones that collect multiple information variables with low, partially overlapping, specificity. Since the latest developments in the application of multivariate data analysis to e-tongues have been reviewed in [38], and the use of information visualization for systems based on the e-tongue concept is described in the next section, we shall turn to electronic noses. Wedge et al. [74] investigated e-noses made with arrays of organic field-effect transistors to detect airbone analytes in real time, with a time-lag of only 4 s. Data processing made use of genetic programming, which was proven adequate to deal with the multiple parameters involved in the sensor arrays. Zhang et al. [75] combined Fisher Discriminant Analysis (FDA) [76] with Sammon's mapping [16] to distinguish among seven samples including fuels and drinks. Figure 4a shows that Sammon's mapping itself does not yield a reasonable clustering of the data. This was attributed to fluctuations of temperature, humidity and sample concentration, which caused the data to be dispersed. However, when Sammon's mapping was used in conjunction with FDA, much better distinction was attained, as shown in Fig. 4b.

Volatile compounds produced by bacteria from processed poultry were identified upon treating the data from an electronic nose with Sammon's mapping and artificial neural networks [77]. In a similar work, Byun et al. [78] also employed Sammon's mapping to assess the malodour in pig slurry. For complex samples, such as those associated with distinct aromas, electronic noses and chemometric analysis have been used in conjunction [79]. Neural networks have also been combined with discrete wavelet transform (DWT) to obtain calibration curves for the simultaneous quantification of Cd^{2+} and Pb^{2+} in solution, where the principle of detection was potentiometry [80].

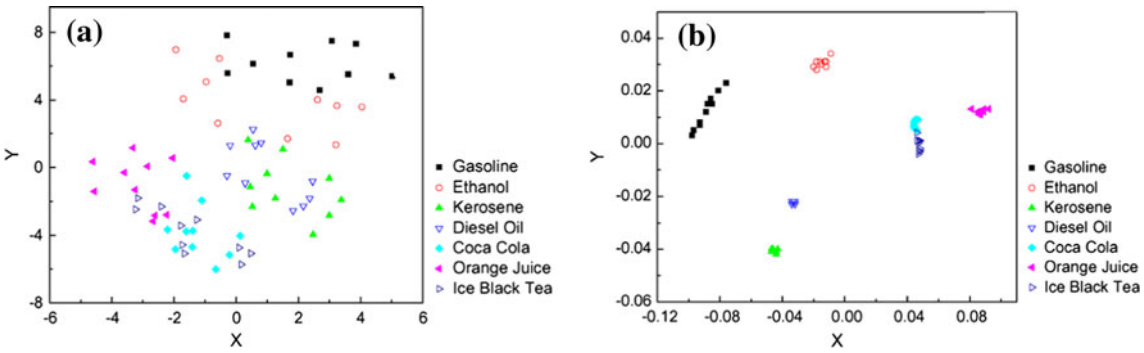

Fig. 4 **a** Plot of the data for multiple transistors using Sammon's mapping, for which it is clear the different liquids analyzed cannot be distinguished. **b** For the FDA-MSM result, distinct clusters could be identified. The liquids analyzed are listed in the insets. *Reproduced with permission from 75*

The variety of statistical and computational methods to analyze data from e-noses is evident from inspecting recent papers in the field, as is the case of e-noses used to characterize several odors [81] and for discriminating volatile organic compounds (VOCs) [82].

To summarize, the performance of e-tongues and e-noses obviously depends on an adequate choice of materials and film architectures for the sensing units, and of suitable principles of detection. But a successful application ultimately depends on the data analysis, which may require a suite of tools for a single case. As emphasized by Zhang et al. [75], the pattern recognition method has become an important part of the e-nose technique.

4 Information Visualization Applied to Sensing and Biosensing

The term "information visualization" has only recently been associated with sensing and biosensing [8], though many works discussed in Sect. 3 already employed some form of visual representation. In this section we shall demonstrate that employing sophisticated data treatment techniques are also crucial for optimizing sensing and biosensing performance. This is true for several aspects akin to analytical tasks, from the choice of suitable sensing units to the identification of features with higher distinguishing ability. For instance, applications that require several sensors incur in a dramatic increase in the number of possible parameter configurations [83]. Optimization can be performed by comparing distinct detection methods. Freitas et al. [84] showed aroma patterns could be better distinguished by using gas sensor arrays (similar to an e-nose) than with chromatography techniques. Figure 5 shows good separation of coffee samples according to the geographic origin upon using Sammon's mapping (a) and PCA (b).

Computational methods are essential to correlate data from sensors and human taste perception. For example, Della Lucia et al. [85] found evidence that extrinsic or non-sensory characteristics of food, such as brand names, affect consumers' choice. In another example, Ferreira et al. [86] applied machine learning methods to correlate data from electronic tongues to the human taste for coffee samples. The concept of electronic tongue has been discussed also in connection with chemometrical data analysis, considering data from a multimicrobial biosensor chip [87]. In the analysis of wines, for instance, in addition to electronic tongues, research has been conducted to characterize wines on the basis of compositional profiles. Saurina [36] reviewed the potential descriptors of wine and its quality, where information on the contents of low molecular organic acids, volatile species, polyphenols, amino acids, biogenic amines and inorganic species is processed with several methods, including cluster analysis and PCA.

Artificial intelligence methods allowed the production of noninvasive glucose monitors for diabetic human subjects [88]. Sensing was performed by measuring the electric current generated in the transport of glucose that interacted with glucose oxidase in a hydrogel placed on the skin surface. The glucose concentration in the blood could be estimated with a combination of methods, involving the theory of mixtures of experts (MOE) using a superposition of multiple linear regressions and switching algorithm. In the MOE method, the unknown coefficients were determined with the Expectation Maximization algorithm.

Visualization techniques are useful not only to assist the biosensing tasks per se, but also in integrated systems where sensing is coupled to other types of information. For example, a platform of biosensing to detect tropical diseases could be developed by integrating biosensors with spatial technology, as in Saxena et al. [89] who applied remote sensing and global positioning system (GPS) to identify areas affected by malaria epidemics.

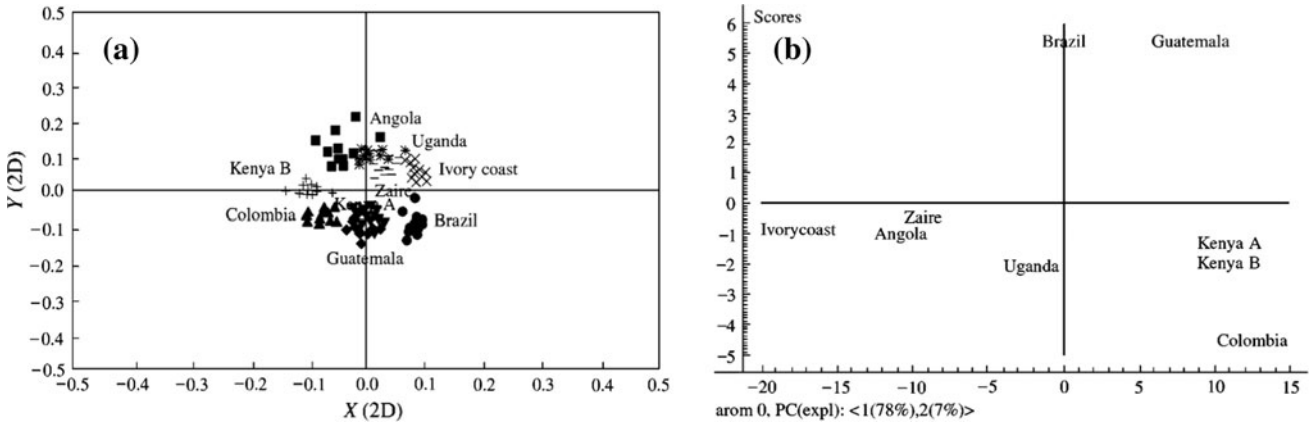

Fig. 5 **a** The patterns of nine coffees (*Arabica*—Brazil, Colombia, Guatemala and Kenya (A and B); *Robusta*—Angola, Ivory Coast, Uganda and Zaire) analyzed with an electronic sensor array appear almost superimposed in a Sammon's mapping plot. **b** Good distinction was achieved when the data were plotted in a PCA diagram, with the first two principal components characterizing the *Arabica* and *Robusta* varieties. *Reproduced with permission from* Ref. [84]

In the sensing field, where the identification of samples is basically a classification task, the performance of the sensing devices has improved with the aid of machine learning and information visualization methods for treating data. This is the case of e-tongues, discussed earlier, which are being used in the analysis of liquids such as wines, fruit juices, coffee, milk and beverages. Electrochemical measurements and impedance spectroscopy are among the most prominent principles of detection. Riul et al. [38, 90] reported a very sensitive e-tongue based on impedance spectroscopy and ultrathin films (nanometers in thickness) deposited onto interdigitated electrodes, whose experimental setup is given in Fig. 6a. Because a large number of samples and measurements are needed to distinguish between very similar samples, applying chemometric or pattern recognition methods is inevitable. PCA is the most popular tool to analyze e-tongue data. However, sophisticated tools combining machine-learning and data mining approaches and information visualization techniques have been applied recently.

Information visualization introduces three main advantages. The first and most obvious is the possibility of treating the whole dataset rather than specific parts of the data. For example, instead of applying PCA just to the impedance value at particular frequencies, the whole impedance vs. frequency curves can be processed automatically. The second advantage is related to the ample choice of projection techniques to map the data. In addition to the linear techniques, such as PCA, non-linear methods can be employed, as we shall comment upon below. The third advantage is the possible optimization of sensing performance that goes beyond exploiting the whole data, for instance employing feature selection strategies to maximize inter-cluster distances while minimizing intra-cluster distances [29].

Moraes et al. [8] compared Sammon's mapping and IDMAP as strategies to plot impedance data from sensors made with layer-by-layer (LbL) [91] films in order to detect phytic acid in solution. The real and imaginary components of the impedance were analyzed concomitantly. Significantly, better distinction ability was achieved with different projection techniques for the distinct sensing units. While for the sensor made with LbL films of poly(allylamine chloride) (PAH) alternated with polyvinyl sulfonic acid (PVS) IDMAP proved more efficient, for the unit with phytase layers alternated with PAH better results were obtained with Sammon's mapping. Figure 6b shows the plot obtained with Sammon's mapping after a data standardization procedure. With the specific interaction between phytic acid and phytase, one should expect a much superior performance for the sensing unit containing LbL films of phytase. That PAH/PVS LbL film efficiency to detect phytic acid could be explained by a detailed analysis of the whole curves, which was only possible with the visualization methods. It should be stressed that the distinction performance achieved using linear PCA was much worse.

The power of visualization methods has been combined with an extended e-tongue technology [72, 92] to solve a major problem in biosensing for clinical diagnosis of two tropical diseases, namely Leishmaniasis and Chagas' Disease caused by Trypanosoma cruzi. It so happens that even in sophisticated immunoassays, many false positives occur [93, 94]. Perinotto et al. [72] addressed this problem with impedance spectroscopy measurements with a sensor array containing four sensing units, two of which had immobilized antigens with molecular recognition capability toward anti-Leishmania and anti-T. Cruzi antibodies in LbL films. A cartoon with the biosensing device (one sensing unit) is given in Fig. 7, which also shows the capacitance versus

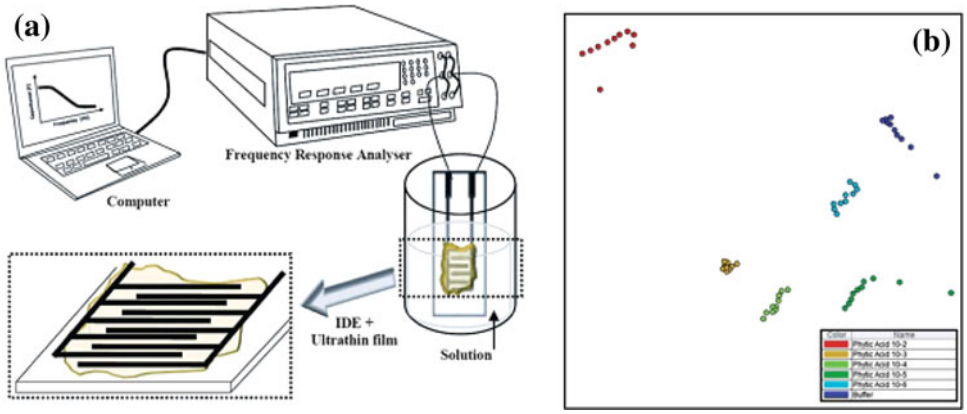

Fig. 6 **a** Illustrative diagram of the experimental setup used in impedance spectroscopy measurements for e-tongues. **b** Sammon's mapping plot of data with standardization for the electrical impedance data obtained with the PAH/phytase sensing unit. The *color* represents different samples of phytic acid, in addition to the buffer. The axes are not labeled, as the relative distances give the degree of dis(similarity) among the samples. *Reproduced with permission from* Refs. [8, 38]*, respectively*

frequency measurements for antibody solutions at 10^{-5} mg/mL for three of the sensing units. The latter were, respectively, a bare electrode, an electrode containing 5 bilayers of PAMAM/PVS (poly(amidoamine) generation 4 dendrimer/poly(vinyl sulfonic acid)), which is a non-specific sensor, and an electrode containing 5 bilayers of PAMAM/proteoliposome (biosensor). The biosensor clearly presents a distinct response for solutions containing antibodies. Even for the mixture of antibodies, the capacitance curve was practically the same as that for the positive anti-*L. amazonensis* IgGs. The latter reveals specific interactions occur upon immersion of the electrode in the mixture solution, with only the positive anti-*L. amazonensis* antibodies binding to the electrode.

By applying PCA to data such as those in Fig. 7, it was possible to distinguish between the samples made with a buffer to which various concentrations of antibodies were added [72]. However, when all the "real" samples made with blood serum of infected animals were included full distinction could not be reached, encouraging investigation of other projection techniques. By way of illustration we show in Figs. 8 and 9 visualizations of the impedance spectroscopy data obtained with one sensor (the bare electrode) for all the samples. Not surprisingly, with the lack of specificity in interaction with the analytes (antibodies), the distinction is rather poor. But a visual inspection of Fig. 9 already shows that a non-linear technique, namely Sammon's mapping, offers a better response than the PCA plot shown in Fig. 8.

The full distinction with Sammon's mapping was achieved upon employing the impedance data of the four sensing units mentioned above. This is shown in Fig. 10.

Another evidence of the superiority of non-linear methods for biosensing was obtained by plotting the data from the four sensors with PCA, shown in Fig. 11. It is observed the distinction is good, but not perfect, in contrast to the Sammon's Mapping plots. Other non-linear techniques, IDMAP included, were also considered, but results were inferior to those obtained with Sammon's mapping. At present, it is not clear why non-linear techniques have performed better in biosensing data. We hypothesize that the specific interactions between the materials in the sensing units and the analytes, owing to molecular recognition processes, may cause the electrical responses to depend on the various parameters in a highly non-linear fashion.

The IDMAP technique was also employed with light-addressable potentiometric sensors (LAPS) as an efficient tool to eliminate cross-talk between sensor units with micrometric size produced by semiconductor technology [95]. In the LAPS described, the detection of penicillin G was attained by monitoring the variation of ions in solution, at a fixed photocurrent, for 16 points illuminated by infrared light emitting diodes (IR-LEDs). Eight points were modified with a 6-bilayer LbL film of single-walled carbon nanotubes (SWCN) and poly(amidoamine) dendrimer (PAMAM). This film was deposited on the gate insulator of the chip, and the enzyme penicillinase was adsorbed on the top. The reaction of the penicillinase with penicillin G in solution generates free H^+ ions on the electrode surface, and the porous structure of the LbL facilitates its diffusion to the chip surface. Due to the close proximity of the modified and non-modified points of detection (especially those adjacent each other) there was some influence of neighboring points, i.e. cross-talk. Thus, a direct analysis of the voltage versus time curves of the sensors (with constant-current) reveals that both modified and unmodified points have the same trend of responses. In the plot obtained with the IDMAP projection, the modified and unmodified sensors were clearly separated in two clusters.

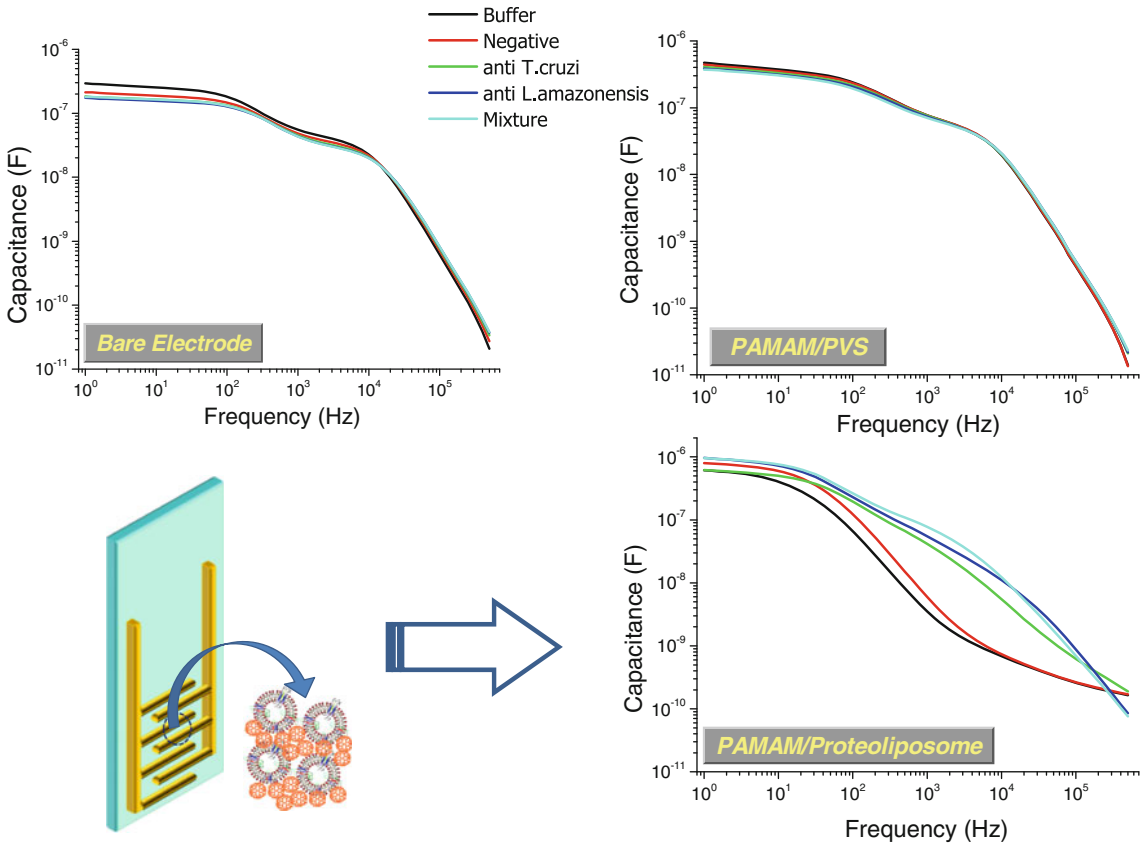

Fig. 7 On the *bottom left* a schematic diagram is shown for the sensing device, where a LbL film containing antigens in proteoliposomes is deposited onto an interdigitated electrode. The other panels bring capacitance versus frequency curves for three electrodes immersed into 10^{-5} mg/mL antibody solutions, as indicated. Note that distinction between the samples is much superior with the electrode containing a 5-bilayer LbL film of PAMAM/proteoliposome (*lower right*). *Reproduced with permission from* Ref. [72]

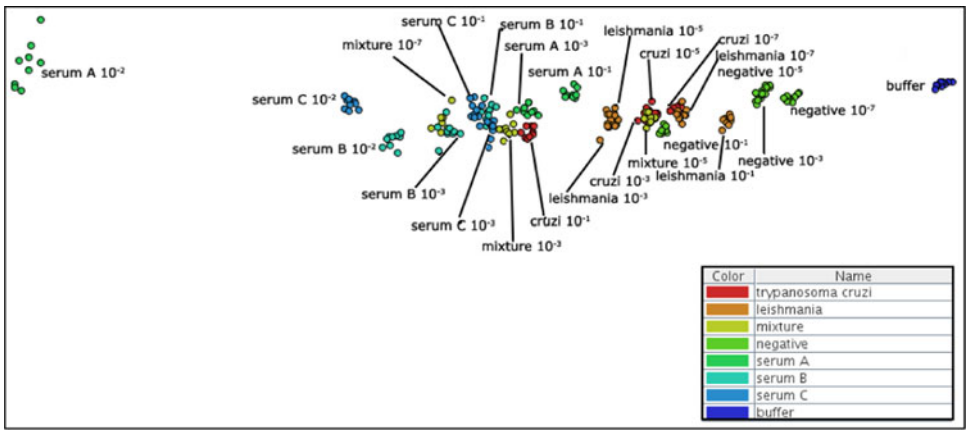

Fig. 8 Projection using PCA of the electrical impedance data obtained with the bare electrode for *L. amazonensis* and *T. Cruzi* samples with different concentrations, as follows. Serum A contained negative antibodies, serum B contained anti-*Leishmania* antibodies, serum C contained anti-*T. Cruzi* antibodies. The other samples were the buffer, and the so-called synthetic samples made with the buffer to which anti-*Leishmania*, anti-*T. Cruzi* and negative antibodies were added. The mixtures were synthetic samples with anti-*Leishmania*, anti-*T. Cruzi* antibodies together. *Reproduced with permission from* Ref. [92]

Moreover, the technique allowed the recognition and grouping of different samples containing glucose, pure buffer and penicillin G with three different concentrations. Once again, the authors tried several projection methods available in a free platform called *PEx-Sensors* (see below) [29], and IDMAP provided the best classification results.

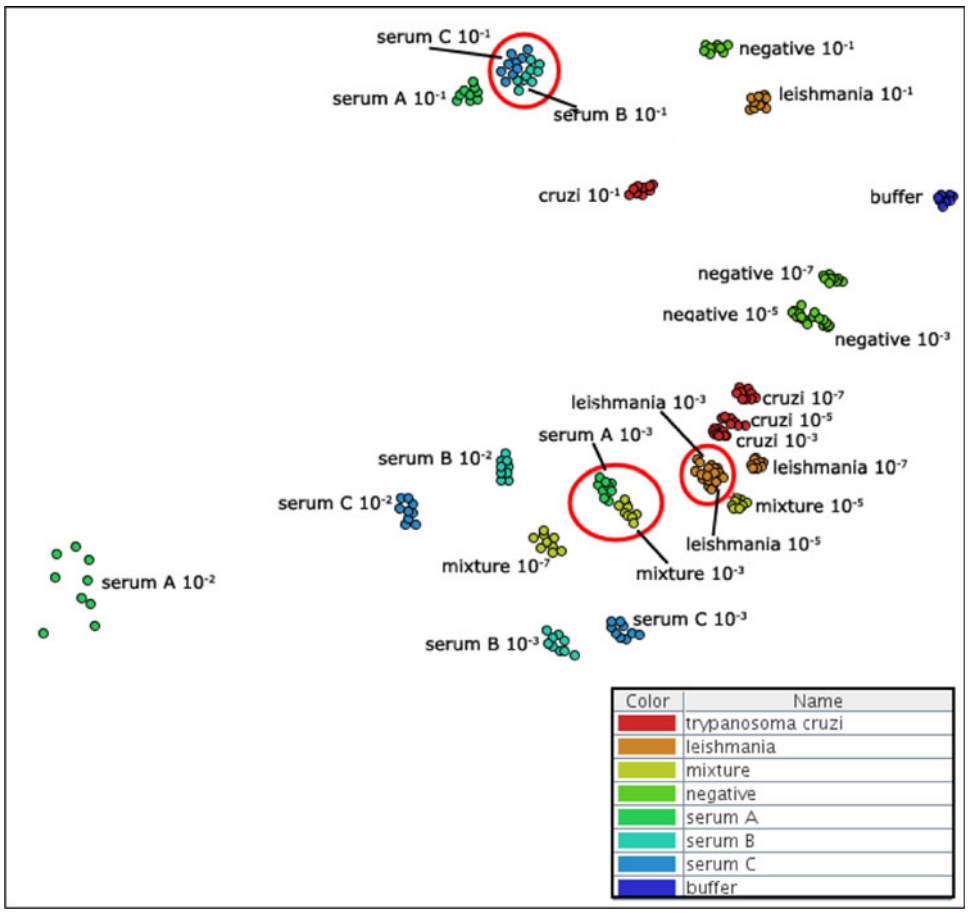

Fig. 9 Projection using Sammon's mapping of the same samples in Fig. 8. Though data points from different samples are still mixed (*circled in red*), the distinction is better than in Fig. 8 where PCA was used. *Reproduced with permission from* Ref. [92]

With regard to the third advantage of information visualization methods, one may mention the optimization of biosensor performance using feature selection coupled with visualizations obtained with projection techniques. Paulovich et al. [29] used *Parallel Coordinates* (PC) visualizations [21] of capacitance data of a PAH/PVS sensing unit, obtained much in the same way as the aforementioned measurements, for aqueous solutions containing the analyte phytic acid to be detected. Owing to the lack of specific interaction, the distinguishing ability of this sensing unit was expected to be poor. Indeed, this seems to be the case judging by the Parallel Coordinates plot in Fig. 12.

With such visualization and computation of the silhouette coefficient [96] for each measured value at a particular frequency, one may conceive ways to select frequencies and enhance the distinguishing ability. The silhouette is a metric for evaluating the quality of a data cluster that varies between -1 and 1, where higher values indicate better cluster quality. The silhouette coefficient is given by:

$$S = \frac{1}{n}\sum_{i=1}^{n}\frac{(b_i - a_i)}{\max(a_i, b_i)}$$

where a_i is the average of the distances between the ith data point and all other points of the same cluster, and b_i is the minimum distance between the ith data point and all other points from the other clusters.

Choosing the most suitable frequencies for distinguishing the sample amounts to feature selection, which can be done quantitatively using the silhouette coefficients. Paulovich et al. [29] employed a genetic algorithm to scan the whole data space of cluster silhouettes and automatically identify the best frequencies for distinction. Figure 13 depicts a parallel coordinates visualization for the 10 best frequencies selected, where a better distinction capability is readily observed in comparison with Fig. 12. The improvement was confirmed with multidimensional projections of the data obtained using IDMAP [17]. The importance of a systematic search for the features leading to optimization is highlighted by the analysis of the silhouette coefficients in Fig. 13. While most of the frequencies selected had high coefficients (represented by blue color), one particular frequency was denoted by a red box. This means this frequency, when considered in isolation, does not lead to good distinction for the different

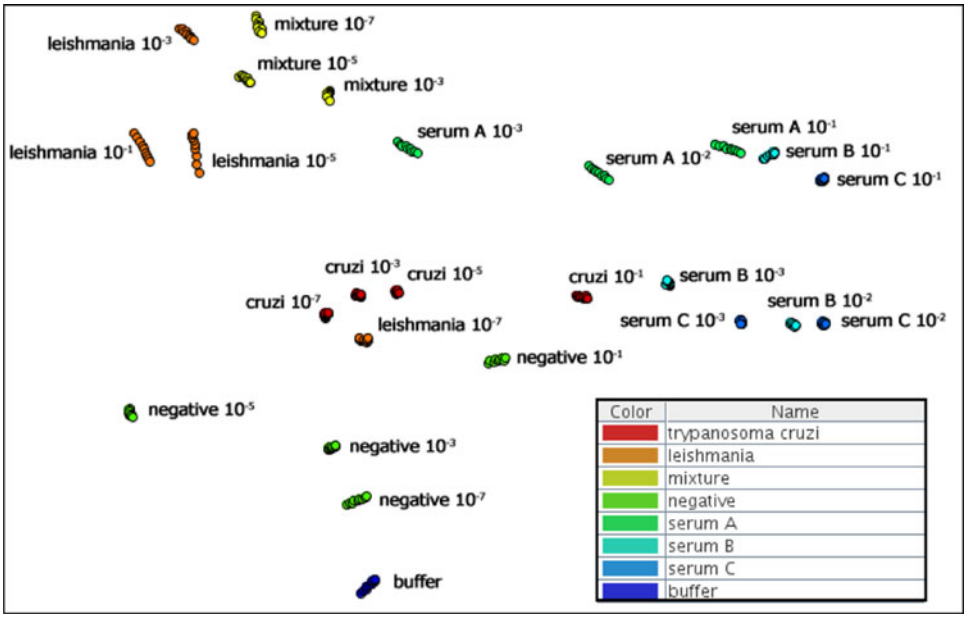

Fig. 10 Projection using Sammon's mapping of the capacitance data obtained with four sensors for all the samples shown in Fig. 8. All samples can now be clearly separated. *Reproduced with permission from* Ref. [92]

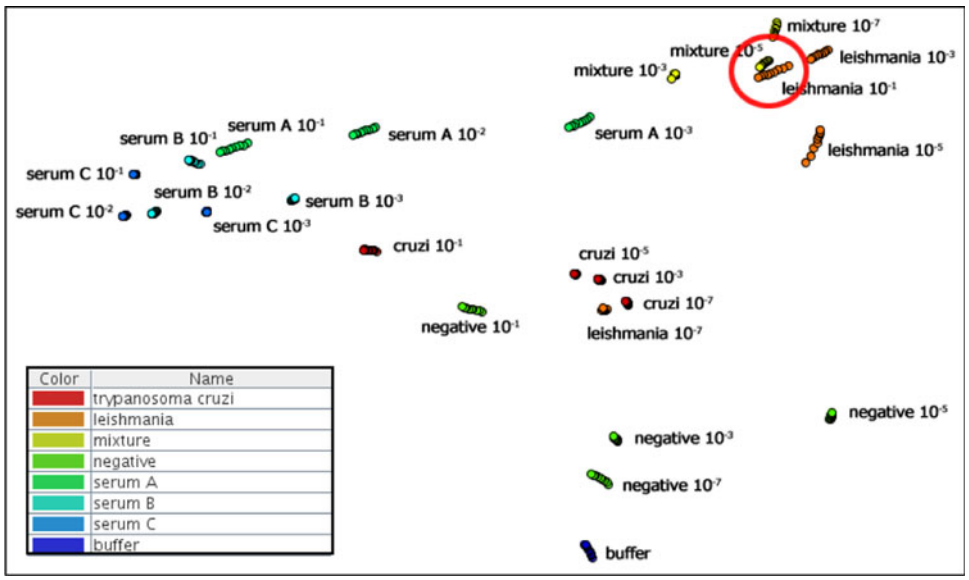

Fig. 11 The same data in Fig. 10, now projected with PCA. In contrast to the Sammon's mapping plots, now some samples could not be distinguished (*circled in red*). *Reproduced with permission from* Ref. [92]

samples. However, used in conjunction with other frequencies it improves the overall distinguishing ability of the system.

4.1 Systems Available

Several visualization systems for data analysis are available, and a brief review of pros and cons of commercial and freely available systems is given in [97]. For specific applications, *Nature Methods* published a special issue on methods to visualize biological data [98], including genome sequences, macromolecular structures, phylogenetic trees, cells, and organisms. Specifically for data from sensors and biosensors, to our best knowledge the only system is the *Projection Explorer Sensors (PEx-Sensors)* [29]. The *PEx-Sensors* platform was designed to handle large datasets, such as those reported by Siqueira Jr. et al. [95] who analyze multiple impedance versus frequency curves from many sensors simultaneously. *PEx-Sensors* implements several projection techniques that may be tested in search for the most appropriate for a given application. It also allows for obtaining parallel coordinate

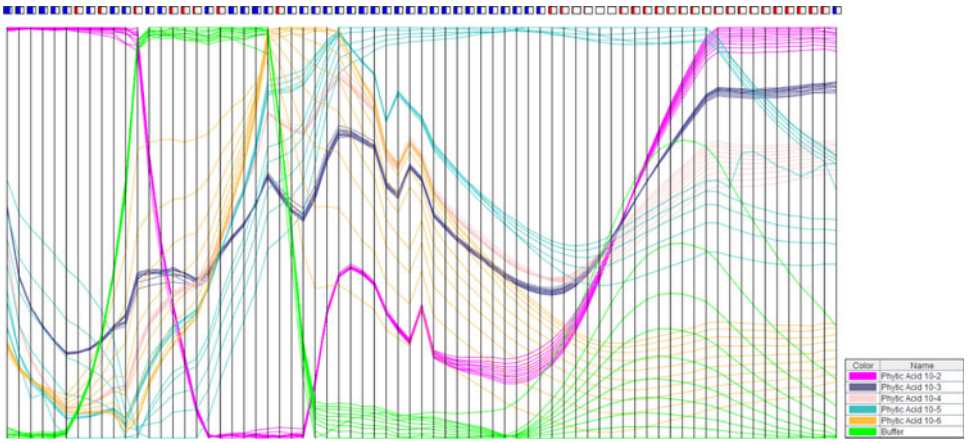

Fig. 12 Visualization of capacitance data with a sensing unit made of a PAH/PVS LbL film deposited onto an interdigitated gold electrode, using the Parallel Coordinates technique. The x-axis is the frequency and the y-axis gives normalized values for the capacitance. Note that for some small concentrations of phytic acid (denoted by *different colors*), there is overlap of the graphs. The *little boxes* on the *top* of the figure represent the silhouette coefficient for each data attribute. *Blue boxes* indicate frequencies that are useful for distinguishing the samples whereas the opposite applies for the *red boxes*. *Reproduced with permission from* Ref. [29]

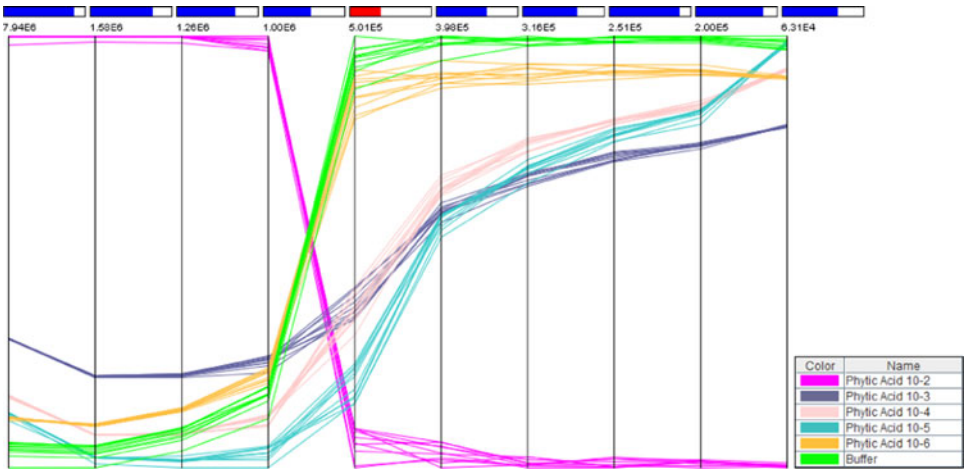

Fig. 13 Visualization with parallel coordinates of the same data in Fig. 12, but now only with 10 selected frequencies to improve the distinguishing ability. The *boxes* representing the silhouette coefficients are almost all *blue*, for an optimization procedure was performed. *Reproduced with permission from* Ref. [29]

plots of the data frequencies to help specialists understand the responses of impedance spectroscopy data. It provides modules to compare the similarity of different sensing units, thus supporting analysis of reproducibility of nominally equal units, and a visual optimization module to support the selection of frequency ranges that render more discriminant sensors. The results reported in Ref. [29], discussed above, were all obtained with *PEx-Sensors*. Furthermore, the techniques implemented in the platform are potentially applicable to other detection principles (i.e. optical absorption and electrochemistry), and *PEx-Sensors* is currently being adapted to work with practically any kind of output data from sensors and biosensors. *PEx-Sensors* is freely available for non-commercial use and may be accessed at http://www.icmc.usp.br/∼paulovic/pexsensors/.

5 Conclusions and Perspectives

In this review paper we have advocated the use of computational methods, especially from the information visualization field, to treat the large amounts of data normally generated in sensing and biosensing. We emphasized the three main advantages of using information visualization, namely: (i) possibility of treating whole datasets in a fast way; (ii) choice of suitable projection techniques; (iii) possibility of optimizing sensing performance upon combining with other computational methods. One of our goals was then to try and disseminate the importance of these tools, not only out of necessity because treating a lot of data manually is no longer feasible but also because many new opportunities arise with data-intensive discovery. In this context, the outlook for this area is extremely

promising. Since the information visualization methods, such as those implemented in PEx-Sensors, are completely generic, they may be applied to images, videos and text as well. Associated with biosensing, in particular, one can now envisage clinical diagnosis intelligent systems that consider not only the data obtained with the biosensors and imaging methods but also prior information about specific patients and diseases. Much in the same way as expert systems for diagnosis in general, the time has come to integrate the knowledge acquired in biosensing into a platform that takes advantage of the tremendous amount of electronic information about any given topic relevant for our society.

Acknowledgments This work was supported by FAPESP, CNPq, CAPES and nBioNet (Brazil).

References

1. Hey T, Tansley S, Tolle K (2009) The fourth paradigm—data intensive scientific discovery. Microsoft Research, Redmond
2. Eggins B (1996) Biosensors: an Introduction. Wiley and B.G. Teubner, Stuttgart
3. Geladi P, Nelson A, Lindholm-Sethson B (2007) Anal Chim Acta 595:152–159
4. Lindholm-Sethson B, Nystrom J, Malmsten M, Ringstad L, Nelson A, Geladi P (2010) Anal Bioanal Chem 398:2341–2349
5. Gorban AN, Kégl B, Wunsch DC, Zinovyev A (2007) Principal manifolds *for* data visualization *and* dimension reduction. Springer, Berlin
6. Luong JHT, Male KB, Glennon JD (2008) Biotech Adv 26:492–500
7. Chambers JP, Arulanandam BP, Matta LL, Weis A, Valdes JJ (2008) Curr Issues Mol Biol 10:1–12
8. Moraes ML, Maki RM, Paulovich FV, 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
9. Card SK, Mackinlay JD, Shneiderman B (1999) Readings in information visualization: using vision to think. Morgan Kaufmann Publishers Inc, San Francisco
10. Oliveira M, Levkowitz H (2003) IEEE Trans Vis Comput Gr 9:378–394
11. Grinstein G, Trutschl M, Cvek U (2001) Proceedings of the 7th data mining conference KDD workshop, pp 7–19
12. Torgeson WS (1965) Psychometrika 30:379–393
13. Paulovich FV, Nonato LG, Minghim R, Levkowitz H (2008) IEEE Trans Vis Comput Gr 14:564–575
14. Jolliffe IT (2002) Principal component analysis. Springer, New York
15. Tejada E, Minghim R, Nonato LG (2003) Inf Vis 2:218–231
16. Sammon JW (1969) IEEE Trans Comput 18:401–409
17. Minghim R, Paulovich FV, Lopes AA (2006) IS&T/SPIE symposium on electronic imaging—visualization and data analysis, vol 6060. pp S1–S12
18. Faloutsos C, Lin K (1995) ACM SIGMOD, pp 163–174
19. Di Battista G, Eades P, Tamassia R, Tollis IG (1999) Graph drawing—algorithms for the visualization of graphs. Prentice-Hall, Upper Saddle River
20. Paulovich FV, Silva CT, Nonato LG (2010) IEEE Trans Vis Comput Gr 16:1281–1290
21. Inselberg A, Dimsdale B (1990) Proceedings of the IEEE visualization (Vis'90), pp 361–375
22. Bishop CM (2005) Neural networks for patterning recognition. Clarendon Press, Oxford
23. Gehlenborg N, O'Donoghue SI, Baliga NS, Goesmann A, Hibbs MA, Kitano H, Kohlbacher O, Neuweger H, Schneider R, Tenenbaum D, Gavin A-CC (2010) Nat Methods 7:S56–S68
24. Walter T, Shattuck DW, Baldock R, Bastin ME, Carpenter AE, Duce S, Ellenberg J, Fraser A, Hamilton N, Pieper S, Ragan MA, Schneider JE, Tomancak P, Hériché J-KK (2010) Nat Methods 7:479
25. Medeiros CB, Joliveau M, Jomier G, De Vuyst F (2010) Geoinformatica 14:279–305
26. Reed JL, Kaas JH (2010) Neural Netw 23:673–684
27. Hinneburg A, Aggarwal CC, Keim DA (2000) Proceedings of the 26th international conference on very large data bases (VLDB'00), pp 506–515
28. Guyon I, Elisseeff A (2003) J Mach Learn Res 3:1157–1182
29. Paulovich FV, Moraes ML, Maki RM, Ferreira M, Oliveira ON Jr, Oliveira MCF (2011) Analyst 136:1344–1350
30. Yang R, Chen RS (2010) Sensors 10:5703–5723
31. Duncan AC, Weisbuch F, Rouais F, Lazare S, Baquey C (2002) Biosens Bioelectron 17:413–426
32. Yoon S, Benini L, De Micheli G (2006) IEEE Trans Comput Aided Design Integr Circuits Syst 25:353–372
33. Danielson UH (2009) Fut Med Chem 1:1399–1414
34. Lindholm-Sethson B, Geladi P, Koeppe RE, Jonsson O, Nilsson D, Nelson A (2007) Langmuir 23:5029–5032
35. Lindholm-Sethson B, Geladi P, Nelson A (2001) Anal Chim Acta 446:121–131
36. Saurina J (2010) Trac Trend Anal Chem 29:234–245
37. Brown SD, Tauler R, Walczak B (2009) Comprehensive chemometrics, chemical and biochemical data analysis, vol 3. Elsevier, Amsterdam
38. Riul A Jr, Dantas CAR, Miyazaki CM, Oliveira ON Jr (2010) Analyst 135:2481–2495
39. Aoki PHB, Caetano W, Volpati D, Riul A Jr, Constantino CJL (2008) J Nanosci Nanotechnol 8:4341–4348
40. Cabral FPA, Bergamo BB, Dantas CAR, Riul A Jr, Giacometti JA (2009) Rev Scientific Instrum 80:026107
41. Gay M, Apetrei C, Nevares I, Del Alamo M, Zurro J, Prieto N, De Saja JA, Rodríguez-Méndez ML (2010) Electrochim Acta 55:6782–6788
42. Apetrei C, Apetrei IM, Villanueva S, De Saja JA, Gutierrez-Rosales F, Rodriguez-Mendez ML (2010) Anal Chim Acta 663:91–97
43. Rodríguez-Méndez ML, Parra V, Apetrei C, Villanueva S, Gay M, Prieto N, Martinez J, De Saja JA (2008) Microchim Acta 163:23–31
44. Rodriguez-Mendez ML, Gay M, De Saja JA (2009) J Porphyr Phtalocyanines 13:1159–1167
45. Kobayashi Y, Habara M, Ikezazki H, Chen R, Naito Y, Toko K (2010) Sensors 10:3411–3443
46. Shen HF, Habara M, Toko K (2008) Sensor Mater 20:171–178
47. Ivarsson P, Kikkawa Y, Winquist F, Krantz-Rülcker C, Höjer N-E, Hayashi K, Toko K, Lundström I (2001) Anal Chim Acta 449:59–68
48. Vlasov YG, Legin AV, Rudnitskaya AM, Damico A, Di Natale C (1997) J Anal Chem 52:1087–1092

49. Vlasov YG, Legin AV, Rudnitskaya AM, Di Natale C, Damico A (1996) Russ J Appl Chem 69:848–853
50. Mimendia A, Gutierrez JM, Opalski LJ, Ciosek P, Wroblewski W, Del Valle M (2010) Talanta 82:931–938
51. Chudy M, Grabowska I, Ciosek P, Filipowicz-Szymanska A, Stadnik D, Wyzkiewicz I, Jedrych E, Juchniewicz M, Skolimowski M, Ziolkowska K, Kwapiszewski R (2009) Anal Bioanal Chem 395:647–668
52. Ciosek P, Grabowska I, Brzozka Z, Wroblewski W (2008) Microchim Acta 163:139–145
53. Ciosek P, Wroblewski W (2007) Analyst 132:963–978
54. Ciosek P, Wroblewski W (2006) Sens Actuator B Chem 114:85–93
55. Yu HC, Wang J, Xiao H, Liu MA (2009) Sens Actuator B Chem 140:378–382
56. Yu HC, Wang YW, Wang J (2009) Sensors 9:8073–8082
57. Ciosek P, Wroblewski W (2011) Sensors 11:4688–4701
58. Vlasov YG, Ermolenko YE, Legin AV, Rudnitskaya AM, Kolodnikov VV (2010) J Anal Chem 65:880–898
59. Del Valle M (2010) Electroanalysis 22:1539–1555
60. Bratov A, Abramova N, Ipatov A (2010) Anal Chim Acta 678:149–159
61. Winquist F (2008) Microchim Acta 163:3–10
62. Scampicchio M, Ballabio D, Arecchi A, Cosio SM, Mannino S (2008) Microchim Acta 163:11–21
63. Rehman A, Iqbal N, Lieberzeit PA, Dickert FL (2009) Monatshefte Fur Chemie 140:931–939
64. Escuder-Gilabert L, Peris M (2010) Anal Chim Acta 665:15–25
65. Ghasemi-Varnamkhasti M, Mohtasebi SS, Siadat M (2010) J Food Eng 100:377–387
66. Baldwin EA, Bai JH, Plotto A, Dea S (2011) Sensors 11:4744–4766
67. Zeravik J, Hlavacek A, Lacina K, Skladal P (2009) Electroanalysis 21:2509–2520
68. Vlasov YG, Legin AV, Rudnitskaya AM (2008) Russ J Gen Chem 78:2532–2544
69. Pavinatto FJ, Fernandes EGR, Alessio P, Constantino CJL, De Saja JA, Zucolotto V, Apetrei C, Oliveira ON, Rodriguez-Mendez ML (2011) J Mater Chem 21:4995–5003
70. Siqueira JR, Abouzar MH, Poghossian A, Zucolotto V, Oliveira ON, Schoning MJ (2009) Biosens Bioelectron 25:497–501
71. Caseli L, Moraes ML, Zucolotto V, Ferreira M, Nobre TM, Zaniquelli MED, Rodrigues UP, Oliveira ON Jr (2006) Langmuir 22:8501–8508
72. Perinotto 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) Anal Chem 82:9763–9768
73. Ivarsson P, Holmin S, Hojer NE, Krantz-Rulcker C, Winquist F (2001) Sens Actuator B Chem 76:449–454
74. Wedge DC, Das A, Dost R, Kettle J, Madec MB, Morrison JJ, Grell M, Kell DB, Richardson TH, Yeates S, Turner ML (2009) Sens Actuator B Chem 143:365–372
75. Zhang SP, Xie CS, Fan CQ, Zhang QY, Zhan Q (2007) Sens Actuator B Chem 127:399–405
76. Mika S, Rätsch G, Weston J, Schölkopf B, Müller K-R (1999) Proceedings of the IX IEEE conference on neural networks for signal processing, pp 41–48
77. Arnold JW, Senter SD (1998) J Sci Food Agric 78:343–348
78. Byun HG, Persaud KC, Khaffaf SM, Hobbs PJ, Misselbrook TH (1997) Comput Electron Agric 17:233–247
79. Rodriguez SD, Monge ME, Olivieri AC, Negri RM, Bernik DL (2010) Food Res Int 43:797–804
80. Cartas R, Mimendia A, Legin A, Del Valle M (2010) Talanta 80:1428–1435
81. Distante C, Leo M, Siciliano P, Persaud KC (2002) Sens Actuator B Chem 87:274–288
82. Setkus A, Olekas A, Senuliene D, Falasconi M, Pardo M, Sberveglieri G (2010) Sens Actuator B Chem 146:539–544
83. Petersson H, Klingvall R, Holmberg M (2009) Sens Actuator B Chem 142:435–445
84. Freitas AMC, Parreira C, Vilas-Boas L (2001) J Food Compos Anal 14:513–522
85. Della Lucia SM, Minim VPR, Silva CHO, Minim LA, Ceresino EB (2010) Boletim do Centro de Pesquisa de Processamento de Alimentos 28:11–24
86. Ferreira EJ, Pereira RCT, Delbem ACB, Oliveira ON, Mattoso LHC (2007) Electron Lett 43:1138–1139
87. Reul T, Harmeling C, Spener F, Knoll M, Zaborosch C (2000) Anal Chem 72:2022–2028
88. Kurnik RT, Oliver JJ, Waterhouse SR, Dunn T, Jayalakshmi Y, Lesho M, Lopatin M, Tamada J, Wei C, Potts RO (1999) Sens Actuator B Chem 60:19–26
89. Saxena R, Nagpal BN, Srivastava A, Gupta SK, Dash AP (2009) Indian J Med Res 130:125–132
90. Riul A Jr, Dos Santos DS Jr, Wohnrath K, Di Tommazo R, Carvalho ACPLF, Fonseca FJ, Oliveira ON Jr, Taylor DM, Mattoso LHC (2002) Langmuir 18:239–245
91. Decher G, Hong JD, Schmitt J (1992) Thin Solid Films 210:831–835
92. Paulovich FV, Maki RM, Oliveira MCF, Colhone MC, Santos FR, Migliaccio V, Ciancaglini P, Perez KR, Stabeli RG, Perinoto AC, Oliveira ON Jr, Zucolotto V (2011) Anal Bioanal Chem 400:1153–1159
93. Nouir NB, Gianinazzi C, Gorcii M, Müller N, Nouri A, Babba H, Gottstein B (2009) Trans R Soc Trop Med Hyg 103:355–364
94. Singh S, Sivakumar R (2003) J Postgrad Med 49:55–60
95. Siqueira JR Jr, Maki RM, Paulovich FV, Werner CF, Poghossian A, Oliveira MCF, Zucolotto V, Oliveira ON Jr, Schöning MJ (2010) Anal Chem 82:61–65
96. Tan P-N, Steinbach M, Kumar V (2005) Introduction to data mining. Addison-Wesley Longman Publishing Co, Boston
97. Telea AC (2007) Data visualization: principles and practice. A. K. Peters Ltd, Wellesley
98. O'Donoghue SI, Gavin A-C, Gehlenborg N, Goodsell DS, Hériché J-K, Nielsen CB, North C, Olson AJ, Procter JB, Shattuck DW, Walter T, Wong B (2010) Nat Methods 7:S2–S4