Colorimetric Sensor Array for Qualitative Water Analysis

Shu Ann Chan, A Jun-Seok Lee, B and Young-Tae Chang A, C, D, E

A Department of Chemistry, National University of Singapore, 117543, Singapore.
B Department of Chemistry, New York University, New York, NY 10003, USA.
C NUS MedChem Program of the Office of Life Sciences Institute, National University of Singapore, 117543, Singapore.
D Laboratory of Bioimaging Probe Development, Singapore Bioimaging Consortium, Agency for Science, Technology and Research (A*STAR), Biopolis, Singapore 138667, Singapore.
E Corresponding author. Email: chmcyt@nus.edu.sg

A chemosensor array comprising 45 off-the-shelf colorimetric dyes, dubbed the Singapore Tongue (SGT), that is capable of discriminating different brands of bottled water and waters of different geographical attribute is described. Twelve kinds of bottled waters were tested by the SGT, and changes of absorbance spectra were analyzed by unsupervised classification methods to validate the SGT system for water analysis. All 12 bottled waters were discriminated at 1 × concentration, and SGT could distinguish the identity of samples of the waters diluted up to 100 times, except distilled waters. Following the study of 63 tap waters in different mass rapid transit stations in Singapore, two distinct clusters were observed from a principal component analysis plot, which correspond to the origin of the tap water. The successful discrimination and identification of in this study demonstrates the practical application of the SGT as a simple tool for water analysis.

Manuscript received: 14 May 2009.
Manuscript accepted: 23 July 2009.

Introduction

Colorimetric analysis has received much attention because it is a simple and fast detection method. [1] For over a hundred years, organic compounds that change colour in response to analytes have remained popular for use in methods to detect specific chemicals. The attention had caused rapid development in the synthesis of a large numbers of organic dyes that can be used as chromogenic sensors.

Perhaps the most familiar application of colorimetric analysis is still based on the use of a single indicator molecule of a target molecule, however, some scientists have begun to develop analytical methods based on a combination of dyes arranged in array formats. [2-3] The concept of this combinatorial sensing approach relies on the cross-reactivity of individual probes. Responses of each dye provide a unique fingerprint pattern, and the distinct patterns enable the system to differentiate and identify different analytes when multi-component analysis is employed. In past decades, a variety of array-based sensors have been developed and have proved the general applicability of combinatorial sensing strategies. For example, Suslick et al. demonstrated the sensing capability of a metallolporphyrin dye array for the discrimination of four sugars, [4-12] amines, [5] and 12 softdrinks. [6] Wolfbeis et al. used fluorescence indicators to discriminate five metal cations. [7] In a related work, our group has developed a colorimetric dye array that was able to distinguish 43 metal nitrates [8] and 23 carbohydrates in aqueous solution. [9] We also successfully designed a counterion-free sensor array to distinguish four metal cations independent of 15 counter anions contained in the mixture. [10] The development of the New York Tongue-1 (NYT-1) sensor array was particularly noteworthy as it was developed for solution-phase analysis. Not only it does not require a probe immobilization process, but the homogeneous interaction between probe and analyte also provides a fast and stable response signal. This array platform provides a clue to overcoming the challenge in developing specific receptors for the solution-based analysis of complex mixtures. Because the absorption spectra of dyes are affected by various interactions (including pH and Lewis acid–base), it does not rely on specific binding or interaction modes. Coupled with its ability to transcend the relative lack of specificity for a particular analyte, differential arrays can emerge as a powerful technique for single and multi-analyte sample identification. As an extension of the NYT-1 identification of cations, we investigated its applicability in water analysis.

Drinking water is an important factor for life and it contains minerals that play a crucial role in human nutrition. [11-13] Waters consist of different contents of cations and anions, and the amount of different ion contents define the water’s quality such as its taste, odour, hardness, and alkalinity. In particular, high levels of heavy metal cations in water can have severe consequences for humans. For example, mercury contamination of Minamata lake water in Japan resulted in blindness of its consumers. [14] Hence, it is imperative that an effective but cost-efficient water analysis method be made accessible to everyone.

Research into water purification and treatment (and subsequently into water analysis) has been growing rapidly for the past decade. [15] The ability to ensure efficient provision of a safe and sustainable water supply is a tangible concern internationally.
A colorimetric dye array was developed initially for the discrimination of bottled waters and tap waters from different parts of Singapore. To our surprise, only the Boon Lay sample was clustered as qualitatively different from the rest of the five samples. Samples represented locations in the West (Jourong East, Boon Lay), East of Singapore. As a primary test, we chose a few stops that represented locations in the West (Jourong East, Boon Lay), East of Singapore.

Results and Discussion

A colorimetric dye array was developed initially for the discrimination of bottled waters. The dye set consists of different scaffold of dyes, i.e., triphenyl, azo, naphthyl, acridine, and hydroxyquinones, which are known to be indicators for various organic and inorganic compounds. Hence, this dye array ultimately senses samples based on their different anion and cation contents, and their response to pH variation. From the previous 47 off-the-shelf-dye array, a further selection of 45 from 47 dyes was made based on their responsiveness to cations and anions, and the dye stability (see Accessory Publication Table S1).

**Discrimination of 12 Bottled Waters**

The selected 45 off-the-shelf dyes, dubbed the Singapore Tongue (SGT), were chosen for the analysis of waters, which comprised 12 bottled waters; spring, carbonated, and mineral. Bottled waters were acquired from different shops and places to improve the variability of the samples.

We used a principal component analysis (PCA) to discriminate water samples using logarithm values of fold-changes (for a detailed description see the Experimental section). PCA is a data mining procedure that is often used for rudimentary analysis of a large amount of data, and it offers the advantage of dimension reduction and component decomposition, which enables data rendering in a small number of variables. PCA aims to define linear combinations of the variables that represent the largest amount of variability in the total data matrix. In this study, colour change patterns of 45 colorimetric dyes were compressed into two axes, which represent the major variance of the total response data. Because this non-supervised analysis method relies on the discrimination power of the dye array rather than the mathematical prediction model, we used PCA as a method to evaluate the performance of the dye array.

From the PCA plot, we were able to identify different clusters for all the waters (Fig. 1). The water samples within one cluster contained subtle different amounts of ions than the waters in the rest of the clusters based on their nutritional labels (the ion contents of all bottled waters are summarized in Table S2). Notably, distillation waters (N1 and N6) and medicated water (N5) were clustered with the control point that represents the response of the dye array to de-ionized (DI) water. While the SGT consists of various indicators, the SGT cannot distinguish the medicated water that consists of only fibrogen.

Encouraged by the successful discrimination of the 12 bottled water samples using SGT, we extended this investigation to study the effects of dilution on the discrimination of N1–N12 (Fig. 2). To maximize the discrimination performance, we first selected the most responsive 21 dyes out of the 45 dyes, and these were used for discrimination of the diluted N1–N12 samples (Table S1). The water samples were diluted 10, 100, and 1000 times. At 10 and 100 times dilution, discrimination of the water samples was still possible, although some degree of the discrimination had been lost for certain dyes. Finally, at a 1000 times dilution, the discrimination of water samples was not possible and all the test solutions fell within the same class as the control point (DI water). Hence, we concluded that the SGT successfully discriminated the samples of 12 bottled waters up to a dilution of 100 times based on their different nutritional value as determined by the ion contents in the bottle.

**Analysis of Real-World Samples: 63 Singapore Tap Waters**

Next, we decided to use SGT to distinguish tap waters from different parts of Singapore. For a systematic approach towards potable water analysis, we decided to initiate the study with the mass rapid transit (MRT) stations, which are tangible landmarks of Singapore. As a primary test, we chose a few stops that represented locations in the West (Jourong East, Boon Lay), East (Bedok), South (City Hall), and Central Singapore (Orchard). To our surprise, only the Boon Lay sample was clustered as qualitatively different from the rest of the five samples. Samples...
from the Boon Lay MRT station were re-collected on a different day and at another tap at the same station to ensure data reproducibility (Fig. 3). The validation test result showed high reproducibility and it was confirmed that the difference was not because of sampling bias.

With the results from these five MRT water samples setting the standard for this part of the experiment, we expanded samples to all 61 MRT stations (before the opening of the circle line and the EW28 and EW29 stations), collected on the same day. From the PCA of the total MRT station data, two distinct clusters were observed (C1 and C2). EW26 (Lakeside) and EW27 (Boon Lay) were observed to deposit in cluster C2 (Fig. 4a). This result was consistent with the experiment performed on five representative MRT stations. Two classes were obviously separated with 59 out of 61 test samples (96.7%) falling into one class. It suggests broad similarities in the qualities in Singapore, except for the
having a slight difference in pH and ion contents. (Table 2). The results show that waters from C1 and C2 clusters attributed to pH variation between the different water samples to investigate if the differential clustering observed could be indicators, pH measurements of the water samples were also taken by chromatography (Table 1). Because the SGT is a composite of inductively coupled plasma optical emission spectroscopy (ICP-OES) and ion chromatography (Table 1). Because the SGT is a composite of different commercial dyes and include dyes used as pH indicators, pH measurements of the water samples were also taken to investigate if the differential clustering observed could be attributed to pH variation between the different water samples (Table 2). The results show that waters from C1 and C2 clusters have a slight difference in pH and ion contents. Finally, we attempted to find correlations between the qualities of MRT tap waters and waters from the surrounding area to evaluate the practical applicability of the SGT. For this investigation, we selected two home (domestic) waters near EW26 and EW27, and five other home waters in the C1 area (home waters were denoted as HW). As expected, HW1, HW7, HW21, and HW19 were clustered within cluster C1 and HW27 was clustered in cluster C2 (Fig. 5a). In contrast, we found that HW26 was clustered in C1 rather than C2. Despite the proximity of EW26 and HW26, the pipeline supplying water to the two places were different based on consultation with the PUB. While the network of water pipelines are all linked and connected, there are pumps and valves installed at various points on the pipelines to control the water flow and its direction. We found that along pipe AC there was a valve that separated the supply of water to region X and region Y as indicated in the map in Fig. 5.

In a preliminary study to compare the clustering pattern for Singapore tap waters against waters collected from different parts of Malaysia, Japan, and India, Singapore's tap water appeared to converge into one cluster, where for the other three countries, even waters from the same country can show substantial differences (Fig. S1). Although a comparison of potable water within Singapore may show a degree of variability, however, this variability in the quality of water samples from around Singapore is less significant compared with the degree of variability in tap water qualities' of Malaysia, Japan, and India. Hence, the ability of the SGT to separate water samples from different countries into distinct clusters demonstrates the clustering sensitivity that the artificial tongue can potentially achieve.

Next, we selected 10 dyes that showed the highest differential response to study what is the major mechanism of water discrimination (the selected dyes are listed in Table S3, Accessory Publication). Changes in the electronic environment of chromophores can induce a colour change. This happens by (i) hydrogen bonding, (ii) protonation and deprotonation of anions, (iii) Lewis acid–base interactions, and (iv) metal chelations. Although some of the well-known metal chelators, methylthymol blue, alizarin, alizarin red S, and xylenol orange, caused the strongest response as a result of metal cation recognition, the differential responses of many of the other dyes seemed to be a combination of the effects of the cations and the counter anions in the solutions.

SGT demonstrated a successful discrimination of various water samples in a training set by using fingerprints of the dye response rather than quantifying each ingredient in the samples. Complex mixtures, such as real-life water samples, cannot therefore be analyzed by this method. However, the analysis of complex samples requires a reference training set to establish a proper analysis method to characterize the identity of the samples. The collected response data obtained from SGT provides a suitable platform for further application to more sophisticated samples.

Conclusions
Using a 45 colorimetric dye array, we developed a SGT that was able to discern 12 different bottled waters. Different levels of ion content in them caused the observed discrimination. To further demonstrate the applicability of the SGT to real world samples, we observed the classification of 61 MRT station tap waters into two clusters. Elemental analysis strongly suggests that waters from cluster 2 (C2: EW26 and EW27) came from the Choa Chu Kang waterworks (W1), whereas 60 other MRT stations’ taps (classified as C1) were supplied by other waterworks (W2). An extension of this study to incorporate newly opened EW28 and EW29 classified the two stations under C2 by SGT. This showed that the source of water came from the same source to the left of EW26, which is different from the rest of the MRT stations.
Fig. 4. (a) PCA plot of tap waters from 61 MRT stations. Two distinct clusters are observed and denoted as cluster 1 (C1) and cluster 2 (C2). (b) Representative Singapore map with MRT network. Tap waters from the white background area were clustered as C1, and tap waters from the dark background were clustered as C2.

Table 1. Elemental analysis result of C1 and C2 clustered samples, and published elemental analysis data from the Public Utility Board (PUB)\(^\text{[14]}\)

<table>
<thead>
<tr>
<th>Ions</th>
<th>Elemental analysis [ppm]</th>
<th>PUB information [ppm]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cl(^-)</td>
<td>34.885 (± 6.26)</td>
<td>38.48</td>
</tr>
<tr>
<td>NO(^3)</td>
<td>0.42 (± 0.028)</td>
<td>0.42</td>
</tr>
<tr>
<td>SO(^2)(^-)</td>
<td>45.99 (± 5.50)</td>
<td>50.53</td>
</tr>
<tr>
<td>F(^-)</td>
<td>0.43 (± 0.057)</td>
<td>0.47</td>
</tr>
<tr>
<td>Si</td>
<td>1.36</td>
<td>1.46</td>
</tr>
<tr>
<td>Na(^+)</td>
<td>20.545 (± 6.064)</td>
<td>19.75</td>
</tr>
<tr>
<td>K(^+)</td>
<td>3.41</td>
<td>3.24</td>
</tr>
<tr>
<td>Mg(^2+)</td>
<td>2.61 (± 0.085)</td>
<td>2.42</td>
</tr>
<tr>
<td>Ca(^2+)</td>
<td>27.325 (± 2.04)</td>
<td>28.8</td>
</tr>
</tbody>
</table>

\(^A\) Measured in total alkalinity.
\(^B\) Measured in total hardness.

Table 2. \(pH\) measurements for different MRT tap water samples

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Sample</th>
<th>Ave. (pH)</th>
<th>Std dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>DI water</td>
<td>6.91</td>
<td>0.10</td>
</tr>
<tr>
<td>C1</td>
<td>EW13</td>
<td>7.28</td>
<td>0.0058</td>
</tr>
<tr>
<td>C1</td>
<td>EW18</td>
<td>7.56</td>
<td>0.068</td>
</tr>
<tr>
<td>C2</td>
<td>EW26</td>
<td>7.81</td>
<td>0.0055</td>
</tr>
<tr>
<td>C2</td>
<td>EW27</td>
<td>8.01</td>
<td>0.023</td>
</tr>
</tbody>
</table>

This investigation successfully demonstrated the potential for the SGT to be developed into a rapid portable water analytical tool. A further systematic study that includes different types of water samples from all around the world can be studied for the development of an in-situ environmental diagnostic tool for water analysis.

**Experimental**

**Materials and Softwares**

Spectroscopy grade dimethyl sulfoxide (DMSO, 99.9% purity) was purchased from Acros. DI water (18 MΩ) was prepared using a Picosystem filtering system from Hydro service and supply company. \(pH\) indicators were purchased from Chem Service, Sigma, Fluka, Janseen, and Aldrich. All stock solutions of dyes were prepared in 50.0 mL Eppendorf tubes and wrapped with aluminum foil. Polystyrene dark, clear flat-bottom 384-well plates were purchased from Greiner. Clear, flat-bottom 384-well plates were purchased from Biotek. Bottled water samples were bought from National Trade Union Congress and Cold Storage. The same samples were bought on two different days. All home and MRT tap waters were collected on the same day, from two different taps in 15.0 mL Eppendorf tubes. Water samples were stored in a refrigerator at 4°C and before analysis the samples were allowed to equilibrate to room temperature. All UV-vis spectroscopy data were recorded from 350 to 750 nm using a plate reader (Molecular Device, Spectra Max Plus 384). Principal component analyses were performed using Matlab v7 2007b. Graphical data were recorded using a HP3010 scanner, with a resolution of 1200 dpi. Image analyses were performed using Adobe Photoshop v7.2.

**Dye Preparation**

Appropriate amounts (in grams) of the solid forms of the off-the-shelf indicators were dissolved in DMSO to make up 10–100 mM solutions. The DMSO solution was then diluted with DI water to make up the dye solutions with concentrations listed in the Accessory Publication.
The dyes (30 µL) were placed into the 384-well plates and the water samples (30 µL) were then added. For optimization of dye volume needed, an appropriate volume of the dyes (20 µL or 30 µL) were added to the 384-well plates followed by the addition of water samples in a 1:1 ratio. For most experiments, the following plate format was used. Each dye filled one column and every test sample filled two rows as shown in Fig. S3 of the Accessory Publication. Identical duplicate rows provided duplicate experimental data. It was also a check for intra-plate deviation. The plate was then shaken three times with an interval of 30 s between each agitation. It was ensured that there were no bubbles in the test solutions. The plate was left for 5 min to ensure full colour development. To confirm the reproducibility of data, similar experiments were performed using another plate on a different day to check for inter-plate deviation. The spectroscopic data was obtained using the plate reader by recording the absorbance spectra between 350 and 750 nm at 5 nm intervals.

Array Preparation and Data Acquisition

The dyes were chosen to achieve maximum sensitivity and simplify the pattern of response to the analytes. For dyes that change their absorbance intensity without a significant change in λmax, the fold change was calculated using the formula in Eqn 1:

\[
F = \frac{I_{\text{max}}'}{I_{\text{max}}}.
\]  

where \(I_{\text{max}}'\) is the absorbance value of the dye in the presence of samples and \(I_{\text{max}}\) is the absorbance value of the dye in the presence of the control (DI water).

Absorbance Spectra Data Processing and Multivariate Analysis

Fold changes of absorbance from the λmax, between the analytes and the control from the same column were calculated. Theoretically, fold changes can be calculated by comparing the absorbance values at any arbitrarily chosen wavelength. However, this gives poor sensitivity and a lower signal to noise ratio, which is undesirable. Hence, absorbance values from preset λmax were chosen to achieve maximum sensitivity and simplify the data treatment process. Fold change was calculated based on its pattern of response to the analytes. For dyes that change their absorbance intensity without a significant change in λmax, the fold change was calculated using the formula in Eqn 1:

\[
F = \frac{I_{\text{max}}'}{I_{\text{max}}},
\]  

For other samples with two different λmax, a normalization factor of \(I_1/I_2\) is multiplied to Eqn 1 to yield the following formula:

\[
F = \left(\frac{I_1}{I_2}\right) \left(\frac{I_2'}{I_1'}\right),
\]  

where \(I_1\) and \(I_1'\) are the absorbance values of the dye at \(\lambda_1\) while \(I_2\) and \(I_2'\) are the absorbance values at \(\lambda_2\). The logarithmic values of \(F\), \(\log F\), were used as inputs for PCA. Log F values were also used for all other statistical treatment. Visual images were also taken using a high-resolution scanner for qualitative visual inspection.

Accessory Publication

The Accessory Publication contains full list of dyes and colour images of tap water discrimination. It is available from the Journal’s website.

Acknowledgement

We gratefully acknowledge the National University of Singapore (NUS) for financial support (Young Investigator Award: R-143-000-353-123). J.-S.L. was supported by a Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund: KRF-2005-C00088).

References