JOC Article

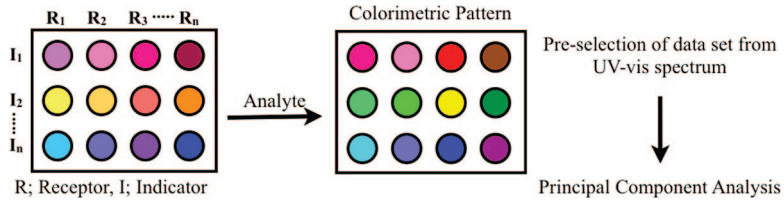
Guidelines for Pattern Recognition Using Differential Receptors and Indicator Displacement Assays

Masanori Kitamura, Shagufta H. Shabbir, and Eric V. Anslyn*

Department of Chemistry and Biochemistry, The University of Texas at Austin, Austin, Texas 78712

anslyn@ccwf.cc.utexas.edu

Received February 25, 2009



A colorimetric sensor array consisting of designed receptors and metal salts was utilized for discrimination of structurally similar carboxylic acid analytes (**25**−**31**). For the discrimination, the receptors do not need to have a 1:1 guest−receptor binding stoichiometry. More than 1:1 stoichiometry, such as 1:2 or 2:1, affords useful information for the discrimination. The importance of data preselection and data preprocessing methods prior to principal component analysis is discussed. These methods change a score plot of principal components in a manner that enhances discrimination. Using the higher order stoichiometry, and the preselection as well as preprocessing discussed herein, we found that an addition of only one designed receptor to an array of metal salts can discriminate the structurally similar analytes.

Introduction

Pattern recognition using "differential receptors" has become an active research area that mimics mammalian olfaction and gustation.[1−3] This method is a very powerful tool because it is possible to analyze diverse chemical structures and complex mixtures. In these methods, a synthetic receptor does not have to possess specific or very selective binding to a target analyte. Instead, it is only necessary that the receptors in the array bind differently to various analytes, thereby creating various patterns for the analytes. Nevertheless, some degree of rational design is still effective for differential receptors because the receptors must have some affinity to the analytes, as well as some selectivity.[1d] In this context, we are interested in exploring receptors that have affinity to the analytes of interest but are not designed and synthesized, alongside fully designed receptors.

(1) (a) *Electrnonic Noses & Sensors for the Detection of Explosives*; Gardner, J. W., Yinon, J., Eds.; NATO Science Series II: Mathematics, Physics and Chemistry; Kluwer Academic Publishers: Boston, 2004; Vol. 159. (b) Toko, K. *Sens. Actuators, B* **2000**, *64*, 205. (c) Albert, K. J.; Lewis, N. S.; Schauer, C. L.; Sotzing, G. A.; Stitzel, S. E.; Vaid, T. P.; Walt, D. R. *Chem. Rev.* **2000**, *100*, 2595. (d) Lavigne, J. J.; Anslyn, E. V. *Angew. Chem., Int. Ed.* **2001**, *40*, 3118.

(2) Selected recent papers for pattern recognition: (a) Baldini, L.; Wilson, A. J.; Hong, J.; Hamilton, A. D. *J. Am. Chem. Soc.* **2004**, *126*, 5656. (b) Zhou, H.; Baldini, L.; Hong, J.; Wilson, A. J.; Hamilton, A. D. *J. Am. Chem. Soc.* **2006**, *128*, 2421. (c) García-Acosta, B.; Albiach-Martí, X.; García, E.; Gil, L.; Martínez-Máñez, R.; Rurack, K.; Sancenón, F.; Soto, J. *Chem. Commun.* **2004**, 774. (d) Stojanović, M. N.; Green, E. G.; Semova, S.; Nikić, D. B.; Landry, D. W. *J. Am. Chem. Soc.* **2003**, *125*, 6085. (e) Paolesse, R.; Natale, C. D.; Nardis, S.; Macagnano, A.; D'Amico, A.; Pinalli, R.; Dalcanale, E. *Chem.—Eur. J.* **2003**, *9*, 5388. (f) Rakow, N. A.; Suslick, K. S. *Nature* **2000**, *406*, 710. (g) Suslick, K. S.; Rakow, N. A.; Sen, A. *Tetrahedron* **2004**, *60*, 11133. (h) Zhang, C.; Suslick, K. S. *J. Am. Chem. Soc.* **2005**, *127*, 11548. (i) Drew, S. M.; Janzen, D. E.; Buss, C. E.; MacEwan, D. I.; Dublin, K. M.; Mann, K. R. *J. Am. Chem. Soc.* **2001**, *123*, 8414. (j) Wright, A. T.; Anslyn, E. V.; McDevitt, J. T. *J. Am. Chem. Soc.* **2005**, *127*, 17405.

(3) Jurs, P. C.; Bakken, G. A.; McClelland, H. E. *Chem. Rev.* **2000**, *100*, 2649.

(4) Reviews: (a) Fabbrizzi, L.; Licchellli, M.; Taglietti, A. *Dalton Trans.* **2003**, 3471. (b) Suksai, C.; Tuntulani, T. *Chem. Soc. Rev.* **2003**, *32*, 192. (c) Wiskur, S. L.; Aït-Haddou, H.; Lavigne, J. J.; Anslyn, E. V. *Acc. Chem. Res.* **2001**, *34*, 963.

(5) Selected recent papers for IDAs: (a) Bonizzoni, M.; Fabbrizzi, L.; Piovani, G.; Taglietti, A. *Tetrahedron* **2004**, *60*, 11159. (b) Hortalá, M. A.; Fabbrizzi, L.; Marcotte, N.; Stomeo, F.; Taglietti, A. *J. Am. Chem. Soc.* **2003**, *125*, 20. (c) Fabbrizzi, L.; Leone, A.; Taglietti, A. *Angew. Chem., Int. Ed.* **2001**, *40*, 3066. (d) Fabbrizzi, L.; Marcotte, N.; Stomeo, F.; Taglietti, A. *Angew. Chem., Int. Ed.* **2002**, *41*, 3811. (e) Buryak, A.; Severin, K. *Angew. Chem., Int. Ed.* **2004**, *43*, 4771. (f) Klein, G.; Reymond, J.-L. *Angew. Chem., Int. Ed.* **2001**, *40*, 1771. (g) Dean, K. E. S.; Klein, G.; Renaudet, O.; Reymond, J.-L. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 1653. (h) Chow, C.-F.; Chiu, B. K. W.; Lam, M. H. W.; Wong, W.-Y. *J. Am. Chem. Soc.* **2003**, *125*, 7802. (i) Chow, C.-F.; Lam, M. H. W.; Sui, H.; Wong, W.-Y. *Dalton Trans.* **2005**, 475. (j) Han, M. S.; Kim, D. H. *Angew. Chem., Int. Ed.* **2002**, *41*, 3809. (k) Han, M. S.; Kim, D. H. *Tetrahedron* **2004**, *60*, 11251. (l) Takahashi, Y.; Tanaka, D. A. P.; Matsunaga, H.; Suzuki, T. M. *J. Chem. Soc., Perkin Trans. 2* **2002**, 759. (m) Hanshaw, R. G.; Hilkert, S. M.; Jiang, H.; Smith, B. D. *Tetrahedron Lett.* **2004**, *45*, 8721. (n) Aït-Haddou, H.; Wiskur, S. L.; Lynch, V. M.; Anslyn, E. V. *J. Am. Chem. Soc.* **2001**, *123*, 11296. (o) Wright, A. T.; Anslyn, E. V. *Org. Lett.* **2004**, *6*, 1341. (p) Folmer-Andersen, J. F.; Lynch, V. M.; Anslyn, E. V. *J. Am. Chem. Soc.* **2005**, *127*, 7986. (q) Zhang, T.; Anslyn, E. V. *Tetrahedron* **2004**, *60*, 11117. (r) Tobey, S. L.; Jones, B. D.; Anslyn, E. V. *J. Am. Chem. Soc.* **2003**, *125*, 4026.

10.1021/jo900433j CCC: $40.75
Published on Web 05/21/2009

For pattern recognition purposes, there are several advantages to the use of indicator displacement assays (IDAs).[4,5] In IDAs, a receptor is bound to an indicator through reversible covalent bonds or noncovalent interactions. When the receptor indicator complex interacts with the analyte the indicator is displaced, resulting in signal modulation. Hence, it is simple to set up an array with various combinations of receptors and commercially available indicators. Moreover, indicators and their concentrations can be chosen to optimize the response (e.g., affinity to receptors, optical properties of indicators, indicator−receptor ratio). In addition, complicated syntheses of indicator-linked receptors are not necessary. Yet, there are only a handful of investigations using IDAs for pattern recognition.[6,7]

In pattern recognition, the data of the sensor array require multivariate analysis such as principal component analysis (PCA) because the sensor array often yields innumerable and overlapping redundant data. Therefore, PCA is used to reduce the dimensionality of a data set to two or three principal components, which provide an adequate representation of the data. Prior to PCA, there are two important processes, data preselection and data preprocessing, which can provide more useful inputs to the PCA.

In this article, we report that (1) it is useful to set up an array of designed receptors and nondesigned receptors for discrimination of structurally similar analytes, (2) complex guest−receptor binding stoichiometries in IDAs for pattern recognition can create useful inputs although simple stoichiometry are preferred for the concentration detection of analytes, and (3) the data preselection and data preprocessing yield more useful input to the PCA.

Results and Discussion

An Array of Designed Receptors and Nondesigned Receptors. We decided to explore the use of an array of designed receptors and nondesigned receptors for discrimination of structurally similar analytes because highly selective binding is unnecessary for the pattern recognition. These two kinds of receptors compensate for various faults in the use of either alone. Nondesigned receptors overcome the need for syntheses inherent designed receptors, and designed receptors compensate for a lack of selective binding for nondesigned receptors.

As designed receptors, we exploited a potpourri of previously published receptors from our group that target carboxylic acid derivatives (Figure 1).[8−13] The amine and guanidinium groups in these structures are logical choices for analysis of carboxylic acid analytes. Boronic acid groups were utilized for diol and α-hydroxycarboxylic acid analytes. The addition of the analyte

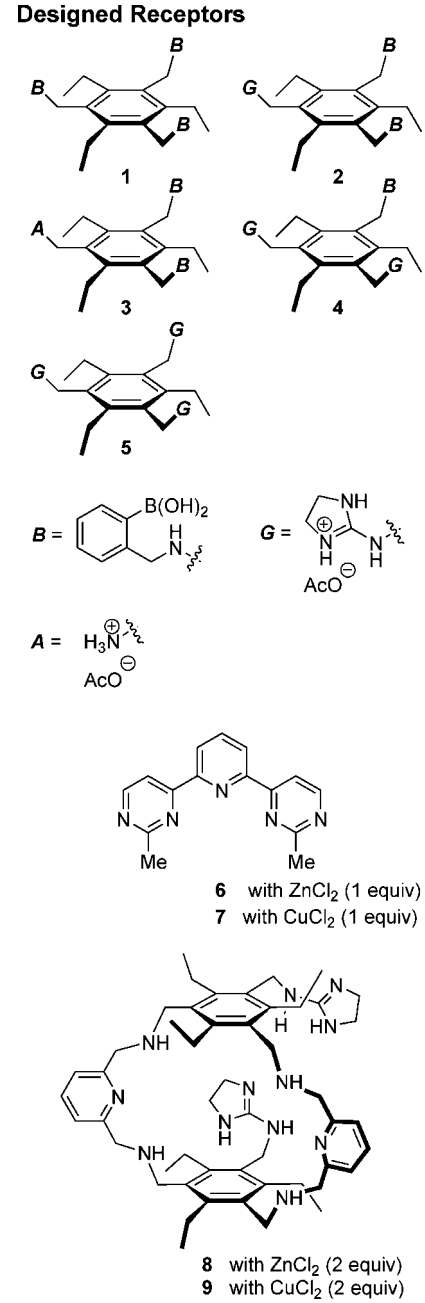
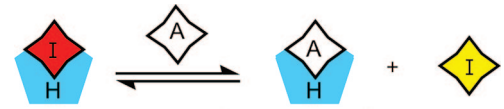


FIGURE 1. Structures of designed receptors.

SCHEME 1. Indicator Displacement Assay between Host−Indicator Complex and Analyte



H: Host, I: Indicator, A: Analyte

causes a change in the spectral response of the receptor indicator complexes solution as shown in Scheme 1. In the presence of the metal, 1:1 metal−ligand complexes for **6** and **7** are formed that have been discussed perviously.[14] These metal complexes

(6) (a) Greene, N. T.; Shimizu, K. D. *J. Am. Chem. Soc.* **2005**, *127*, 5695. (b) McCleskey, S. C.; Griffin, M. J.; Schneider, S. E.; McDevitt, J. T.; Anslyn, E. V. *J. Am. Chem. Soc.* **2003**, *125*, 1114. (c) Wiskur, S. L.; Floriano, P. N.; Anslyn, E. V.; McDevitt, J. T. *Angew. Chem., Int. Ed.* **2003**, *42*, 2070. (d) McCleskey, S. C.; Floriano, P. N.; Wiskur, S. L.; Anslyn, E. V.; McDevitt, J. T. *Tetrahedron* **2003**, *59*, 10089.

(7) (a) Buryak, A.; Severin, K. *J. Am. Chem. Soc.* **2005**, *127*, 3700. (b) Buryak, A.; Severin, K. *J. Angew. Chem. Int. Ed.* **2005**, *44*, 7935.

(8) Wiskur, S. L.; Lavigne, J. J.; Metzger, A.; Tobey, S. L.; Lynch, V.; Anslyn, E. V. *Chem.—Eur. J.* **2004**, *10*, 3792.

(9) Cabell, L. A.; Monahan, M.-K.; Anslyn, E. V. *Tetrahedron Lett.* **1999**, *40*, 7753.

(10) Wiskur, S. L.; Anslyn, E. V. *J. Am. Chem. Soc.* **2001**, *123*, 10109.

(11) Lavigne, J. J.; Anslyn, E. V. *Angew. Chem., Int. Ed.* **1999**, *38*, 3666.

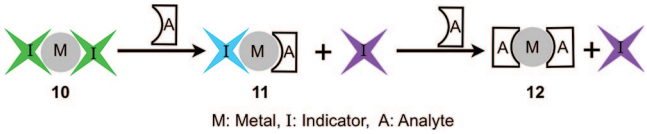
(12) (a) Metzger, A.; Anslyn, E. V. *Angew. Chem., Int. Ed.* **1998**, *37*, 649. (b) McCleskey, S. C.; Metzger, A.; Simmons, C. S.; Anslyn, E. V. *Tetrahedron* **2002**, *58*, 621.

(13) Manimala, J. C.; Wiskur, S. L.; Ellington, A. D.; Anslyn, E. V. *J. Am. Chem. Soc.* **2004**, *126*, 16515.

(14) Folmer-Andersen, J. F.; Aït-Haddou, H.; Lynch, V. M.; Anslyn, E. V. *Inorg. Chem.* **2003**, *42*, 8674.

(15) Aït-Haddou, H.; Wiskur, S. L.; Lynch, V. M.; Anslyn, E. V. *J. Am. Chem. Soc.* **2001**, *123*, 11296.

SCHEME 2. Coordination Modes of 1:2 Complexes for the Discrimination of Analytes



coordinate to amino acid analytes.[15] Compounds **8** and **9** have metal and guanidinium functional groups, both useful in coordination to carboxylate anions.

As nondesigned receptors, simple metal salts were considered because there are many coordination sites at a metal center. As a result, cross-reactivity results from the different coordination modes that can be obtained, as shown in Scheme 2. As a hypothetical example, if the binding constant between the analyte and the metal is small, the light green color of **10** is dominant. If the analyte displaces the indicator to make complex **11**, the light blue of **11** and the purple color of the free indicator appear. Lastly, pure purple of the free indicator dominates when the displacement by the analyte is complete (**12**). Hence, the color variation depends on the coordination character of each analyte to the metal. If there are fewer coordination sites, the resulting information is lower, as with only one coordination site that would give only two states. This complex guest−receptor binding stoichiometries in IDAs for pattern recognition can create useful information for the discrimination. Recent studies of the discrimination of amino acids or dipeptides by IDAs using nondesigned receptors such as a single Cp*Rh complex (Cp* = η^5- pentamethylcyclopentadienyl) or a mixture of $CuCl_2$ and $NiCl_2$ has been performed by Severin's group.[7]

In contrast to the use of a single receptor system, it is known that an array consisting of many types of receptors increases the effectiveness of discrimination.[2b,16] Therefore, we decided to use many types of receptors for the array. Metal salts **13**−**17** were selected on the basis of their hydrolysis constants and their large exchange rate constants for substitution of inner-sphere water ligands (Figure 2).[17]

Structures of Indicators and Analytes. The indicators (**18**−**24**) were chosen to represent a large diversity of structures but also because of their negative charges and ability to bind boronic acids and metal salts (Figure 3).

The analytes **25**−**31** in Figure 4 are components of wine.[18] They can be classified roughly into three groups, α-hydroxycarboxylic acid analytes, phenolic acid analytes, and α-amino acid analytes. They were chosen to represent very structurally similar analytes, hence challenging our methods. Compounds **25** and **26** differ only by one hydroxyl group, **27** and **28** by only a vinylene group, and **30** and **31** by only a methylene.

Initial Studies on 6−**9.** Receptors **6**−**9** were initially studied to determine their binding stoichiometry with indicators. If no isosbestic point in the UV−vis spectra is observed during a titration, it indicates a mixture of three or more species that include an indicator.[19] Higher order stoichiometries were found for all receptor−indicator combinations that were used for the

Non-Designed Receptors

13 $ZnCl_2$ **14** $CuCl_2$ **15** $CdCl_2$

16 $MnSO_4$ **17** $LaCl_3$

FIGURE 2. Structure of nondesigned receptors.

array analysis (see Supporting Information). Hence, the concept in Scheme 2 is applicable to designed receptors **6**−**9**. Designed receptors created by synthetic chemists are ordinarily utilized in 1:1 receptor−guest stoichiometries because this simplifies the mathematical equations that describe the binding isotherms. Therefore, a designed receptor that displays complex receptor−guest stoichiometries has attracted less attention. However, as shown herein, complex stoichiometries can be useful for pattern-recognition-based discrimination.

Coordination Modes of Metal Salt 13−**17.** As with the designed receptors, we first studied the binding properties of the metal salts. We measured UV−vis spectra of an indicator at constant concentration during a titration of a metal salt to delineate the coordination modes. In the titration experiments for each combination of metal salts **13**−**17** and indicators **18**−**24** (a total of 35 combinations), we observed four phenomena: (1) a precipitate appeared during the titration (**14** and **18**, **14** and **24**); (2) there was no response to addition of a metal salt (**13** and **22**, **16** and **20**, **16** and **22**); (3) during the titration, the absorbance increased gradually without change in the shape of the UV−vis spectrum (**14** and **22**, **15** and **22**); and (4) the UV−vis spectrum of the metal salt−indicator combinations behaved in a useful manner (all remaining combinations). One such UV−vis spectra using $CuCl_2$ and indicator **20** is shown in Figure 5, revealing the lack of an isosbestic point. Indeed, an isosbestic point was not commonly observed, with the exception of **15** and **18**, **15** and **20**. However, a binding study for **15** and **18**, **15** and **20** did not fit a 1:1 algorithm,[19,20] analogous to previous studies from our group (see Supporting Information).

Hence, the metal salts all display stoichiometries higher than 1:1 with indicators **18**−**24**. There are several possibilities for the structures of the complexes. Feasible association modes are depicted in Scheme 3. It is likely that an indicator has at least two coordination sites when an excess of a metal salt is used. Further, metal complexes are known to aggregate,[21] and hence different aggregates from different indicators and analytes results in the complex UV−vis spectra that affords information to be used in analyte discrimination.

Procedure for Discrimination. The sensor array was assembled in a 96-well plate by simply mixing each solution of receptors **1**−**9**, **13**−**17** and indicators **18**−**24** (total 98 combinations). After addition of each analyte of Figure 4 to the array, patterns are created for each analyte from the recorded UV−vis spectra of each well. Because the pattern is composed of multidimensional information, a mathematical treatment of the UV−vis spectra is needed to visualize and understand the data. Principal component analysis is one of chemometric tools used to reduce the number of dimensions in such a data set. However, before PCA, there was still a need to select a subset of the whole data set, because some of data likely will not be significant but

(16) (a) Ulmer, H.; Mitrovics, J.; Noetzel, G.; Weimar, U.; Göpel, W. *Sens. Actuators, B* **1997**, *43*, 24. (b) Mitrovics, J.; Ulmer, H.; Weimar, U.; Göpel, W. *Acc. Chem. Res.* **1998**, *31*, 307.

(17) Kobayashi, S.; Manabe, K. *Acc. Chem. Res.* **2002**, *35*, 209.

(18) Ribéreau-Gayon, P.; Glories, Y.; Maujean, A.; Dubourdieu, D. *Handbook of Enology: The Chemistry of Wine Stabilization and Treatments*; Wiley: New York, 2000; Vol. 2.

(19) Connors, K. A. *Binding Constants: The Measurement of Molecular Complex Stability*; Wiley: New York, 1987; Chapter 4.

(20) Piątek, A. M.; Bomble, Y. J.; Wiskur, S. L.; Anslyn, E. V. *J. Am. Chem. Soc.* **2004**, *126*, 6072.

(21) (a) Haiduc, I.; Edelmann, F. T. *Supramolecular Organometallic Chemistry*; Wiley-VCH: New York, 1999. (b) Ruben, M.; Rojo, J.; Romero-Salguero, F. J.; Uppadine, L. H.; Lehn, J.-M. *Angew. Chem., Int. Ed.* **2004**, *43*, 3644. (c) Fujita, M.; Tominaga, M.; Hori, A.; Therrien, B. *Acc. Chem. Res.* **2005**, *38*, 371.

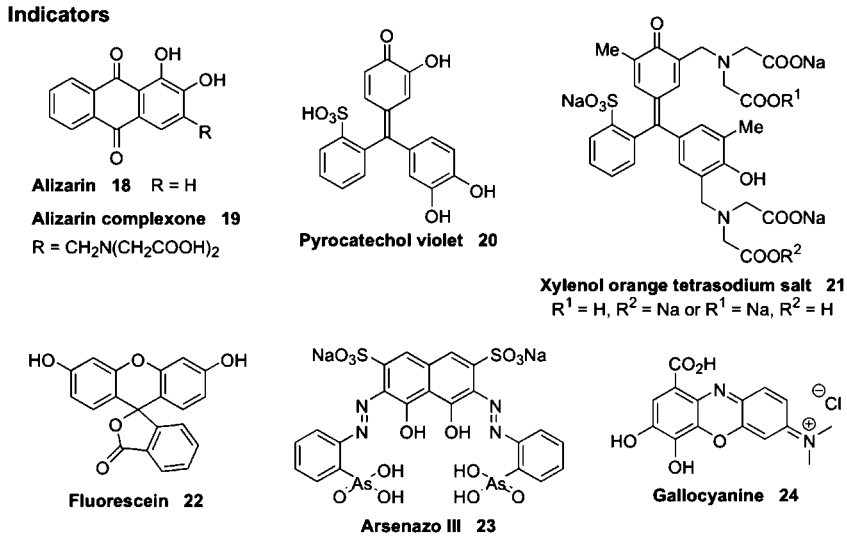


FIGURE 3. Structures of indicators.

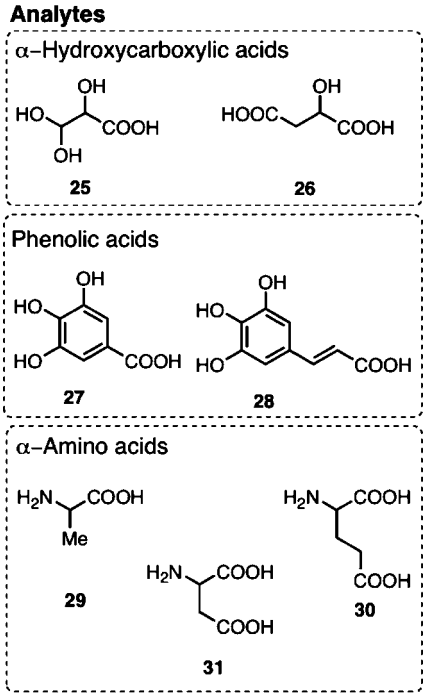


FIGURE 4. Structures of analytes.

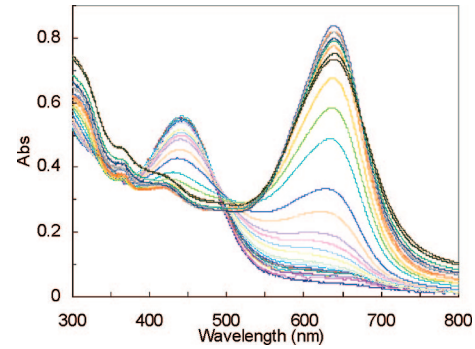
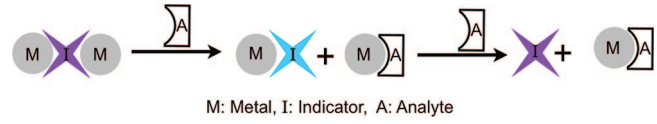


FIGURE 5. UV−vis spectra of the indicator **20** at constant concentration (50 μM) in the presence of $CuCl_2$ (0−200 μM). The spectra were recorded in 75% methanolic aqueous solution buffered by 10 mM HEPES at pH 7.5. An isosbestic point was not observed.

SCHEME 3. Feasible Different Scenario for Various Coordination Modes



rather will have a negative effect on the analysis because it represents noise.

Selection of Proper Indicator−Receptor Combinations in the Sensor Array. All of the 98 combinations of indicators and receptors are not useful for pattern recognition of the analytes (e.g., same selectivity to analytes). After measurement of the UV−vis spectra, the optimum combinations of indicators and receptors were derived on the basis of the following four criteria (see Table S-1 in Supporting Information): (1) there is no precipitation after addition of an analyte, (2) the data is reproducible in three repetitive UV−vis measurements, (3) there is at least one analyte among **25**−**31** that varies in its response in the UV−vis spectra, and (4) the response should arise from interaction to the receptor and not from direct interactions between the indicator and the analyte, such as a subtle difference of pH. To confirm point four, UV−vis spectra without receptors (only indicators and analytes) were compared to the data (see Supporting Information). Clearing these four criteria, 23 combinations of indicators and receptors were selected for further analysis.

Preselection. The data set consisting of UV−vis measurements for 23 combinations of receptors and indicators was still quite large for PCA. The data of UV−vis spectra collected every 5 nm through 400−700 nm (61 wavelengths) was used. Therefore, 61 wavelengths for 23 indicator−receptor combinations yielded 1403 variables. Hence, the size of the data matrix for seven analytes (**25**−**31**) with three repetitive experiments ($7 \times 3 = 21$) was $21 \times 1403 = 29{,}463$ (Figure 6).

All variables are not meaningful for the classification of the analytes. For instance, Feature 1 in Figure 7, which could come from several variables, is only indicating that the samples are carboxylic acids. Further, if variables that only indicate Feature 2 are used for the calculation, the result of classification will be simply α-hydroxycarboxylic acids, phenolic acids, and α-amino acids.

		1403 variables (columns)								
		23 combinations of the receptors and the indicators								
		1 and **18** 61 λ's					**16** and **23** 61 λ's			
		400 nm	405 nm	···	700 nm	···	400 nm	405 nm	···	700 nm
25	Experiment No. 1	0.285	0.308	···	0.008	···	0.452	0.435	···	0.063
	No. 2	0.29	0.312	···	0.002	···	0.453	0.436	···	0.065
	No. 3	0.298	0.32	···	0.001	···	0.448	0.432	···	0.064
26	No. 1	0.28	0.304	···	-0.006	···	0.462	0.445	···	0.064
	No. 2	0.279	0.303	···	0.005	···	0.465	0.446	···	0.056
	No. 3	0.291	0.316	···	0.002	···	0.468	0.45	···	0.058
27	No. 1	0.27	0.295	···	0.006	···	0.402	0.388	···	0.073
	No. 2	0.274	0.301	···	-0.002	···	0.408	0.394	···	0.072
	No. 3	0.277	0.304	···	0.002	···	0.41	0.396	···	0.073
28	No. 1	0.281	0.303	···	0.002	···	0.406	0.392	···	0.076
	No. 2	0.247	0.267	···	-0.009	···	0.41	0.396	···	0.074
	No. 3	0.259	0.281	···	-0.01	···	0.414	0.4	···	0.076
29	No. 1	0.259	0.29	···	0.003	···	0.419	0.406	···	0.083
	No. 2	0.273	0.305	···	0.001	···	0.432	0.419	···	0.082
	No. 3	0.273	0.304	···	0.004	···	0.416	0.404	···	0.08
30	No. 1	0.275	0.307	···	0.012	···	0.427	0.412	···	0.082
	No. 2	0.301	0.335	···	0.006	···	0.423	0.409	···	0.074
	No. 3	0.282	0.316	···	0.004	···	0.414	0.4	···	0.073
31	No. 1	0.3	0.333	···	0.018	···	0.416	0.402	···	0.076
	No. 2	0.294	0.327	···	0.005	···	0.43	0.416	···	0.078
	No. 3	0.285	0.319	···	0.008	···	0.42	0.406	···	0.077

21 samples (rows)

FIGURE 6. Data matrix consisting of 21 samples of seven analytes with three repetitive experiments and 1403 variables. Total data points in this matrix was 21 × 1403 = 29,463 data points.

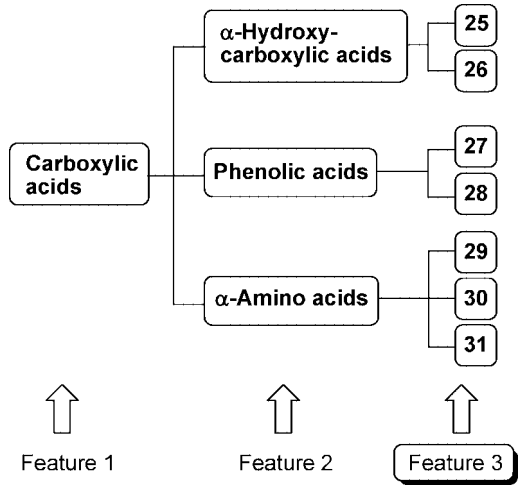


FIGURE 7. An appropriate feature for the classification of the analytes.

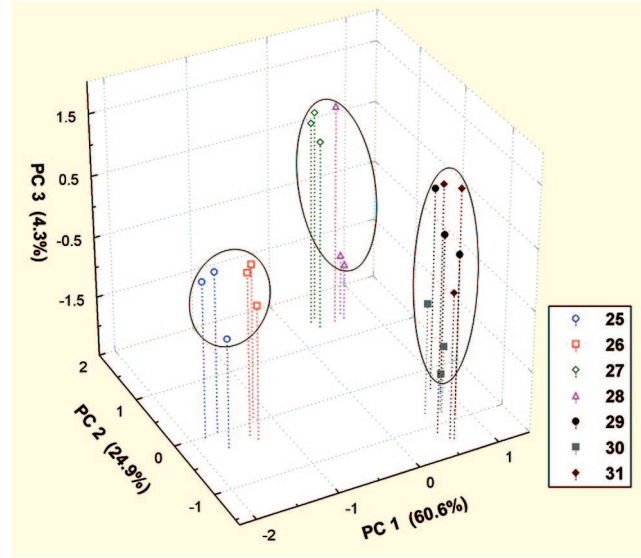


FIGURE 8. Three classes (α-hydroxycarboxylic acids, phenolic acids, and α-amino acids) in a score plot of the first three principal components (PCs) for **25**−**31**. The PCA was performed with a random selection of 230 variables and standardization preprocessing.

The later classification result was found when an arbitrary method of reducing the data was used. PCA for **25**−**31** showed only three classes when 230 variables of 1403 were randomly selected from the full data set (Figure 8). We show below that by selecting the appropriate 40 variables (Feature 3 in Figure 7), we can achieve full classification for **25**−**31**. This is an example that a reduced sensor array can sometimes have a better ability for classification of analytes using PCA.[22]

To accomplish the desired selection of Feature 3, we decided to use the two methods. One was based on an analysis of variance (ANOVA),[23] while the second was direct comparison of the data, because both methods have characteristic strengths. In general, an *F*-ratio in the ANOVA is used to test various

(22) (a) Holmberg, M.; Winquist, F.; Lundström, I.; Gardner, J. W.; Hines, E. L. *Sens. Actuators, B* **1995**, *26*−*27*, 246. (b) Natale, C. D.; Davide, F. A. M.; D'Amico, A.; Sberveglieri, G.; Nelli, P.; Faglia, G.; Perego, C. *Sens. Actuators, B* **1995**, *24*−*25*, 801. (c) Lidén, H.; Mandenius, C.-F.; Gorton, L.; Meinander, N. Q.; Lundström, I.; Winquist, F. *Anal. Chim. Acta* **1998**, *361*, 223. (d) Gibson, T. D.; Prosser, O.; Hulbert, J. N.; Marshall, R. W.; Corcoran, P.; Lowery, P.; Ruck-Keene, E. A.; Heron, S. *Sens. Actuators, B* **1997**, *44*, 413.

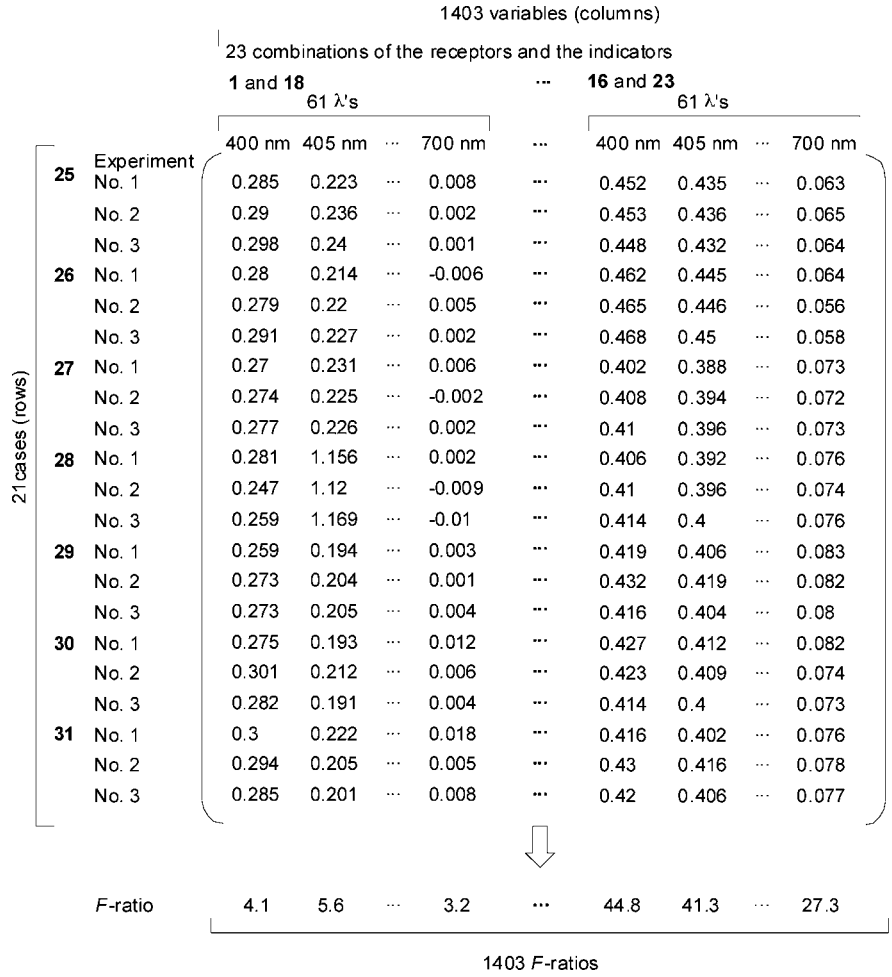


FIGURE 9. Matrix used for the calculation of the *F*-ratios.

statistical hypotheses about the means of the distributions. However, it could be used for the selection of variables (see Supporting Information for the details). The *F*-ratios shown in Figure 9 were calculated for each receptor−indicator combination and wavelength. We observed high *F*-ratios at particular indicator−receptor combinations and particular wavelengths when the analytes respond differently. Lastly, the variables used were selected based on cases that have $F > 300$ (the result of the PCA are shown later).

However, this *F*-ratio can be high if only one analyte reacts very differently, meaning that complete analyte discrimination would not be achieved. Also, the *F*-ratio can be low if all of the analytes react differently but these differences are relatively small, which may represent useful information for discrimination. Therefore, one should be cautious in using only ANOVA-based methods to carry out the preselection.

The second preselection we used involved direct comparison of the data. It is based on an idea that a large difference of the absorbances between analytes is the best information for the ultimate discrimination. Thus, subtractions of absorbances for all of the cases were done to create Figure 10 for comparisons.

First, the averages of the three experiments for each analyte were calculated. Next, subtraction of the averaged absorbances between each two analytes was calculated for all binary comparisons of analytes (21 cases). There are a total of 21 cases in the matrix because the each combination of two of the seven analytes are used ($_7C_2 = {}_7P_2/2! = 21$ cases). These subtracted values were transformed to the absolute values shown in Figure 10, because only differences are important. The largest differences of the absorbances given in Figure 10 indicate the particular indicator−receptor combinations and wavelengths that contain useful information to distinguish two analytes.

We created a matrix of a rank ordering, which corresponds to the order of the differences from first to 1403th in each row of Figure 10, so that we could know which combinations of indicators/receptors were the most effective for the discrimination of each two specific analytes (Figure 11). However, the indicator−receptor combinations and wavelengths that have a large difference for two specific analytes can be near the worst combinations for a different set of two analytes. For example, one indicator−receptor combination and wavelength, which was in first place of rank ordering in the row $|A_{27}-A_{25}|$ and therefore the best to distinguish analytes **25** and **27**, was the worst for the discrimination of **25** and **26** (1403th place in the row $|A_{26}-A_{25}|$). The poor selection between **25** and **26** represent a case of Feature 2 in Figure 7 (**25** and **26** belong to α-hydroxycarboxylic acids and only **27** is different).

Therefore, we needed to choose a subset of the indicator−receptor combinations and the wavelengths, which have differences in all rows of Figure 10. The subset was chosen in the

(23) (a) Ciosek, P.; Brzózka, Z.; Wróblewski, W. *Sens. Actuators, B* **2004**, *103*, 76. (b) Johnson, K. J.; Synovec, R. E. *Chemom. Intell. Lab. Syst.* **2002**, *60*, 225. (c) Johnson, K. J.; Rose-Pehrsson, S. L.; Morris, R. E. *Energy Fuels* **2004**, *18*, 844.

1403 variables (columns)

23 combinations of the receptors and the indicators

21 cases (rows)	**1** and **18** 61 λ's				···	**16** and **23** 61 λ's				Total Sum
	400 nm	405 nm	···	700 nm	···	400 nm	405 nm	···	700 nm	
A_{26}-A_{25}	0.008	0.005	···	0.003	···	0.014	0.013	···	0.005	24.5
A_{27}-A_{25}	0.017	0.013	···	0.001	···	0.044	0.042	···	0.009	79.3
A_{28}-A_{25}	0.029	0.030	···	0.009	···	0.041	0.038	···	0.011	108.5
A_{29}-A_{25}	0.023	0.013	···	0.001	···	0.029	0.025	···	0.018	108.0
A_{30}-A_{25}	0.005	0.006	···	0.004	···	0.030	0.027	···	0.012	101.0
A_{31}-A_{25}	0.002	0.013	···	0.007	···	0.029	0.026	···	0.013	101.7
A_{27}-A_{26}	0.010	0.008	···	0.002	···	0.058	0.054	···	0.013	65.6
A_{28}-A_{26}	0.021	0.024	···	0.006	···	0.055	0.051	···	0.016	92.7
A_{29}-A_{26}	0.015	0.008	···	0.002	···	0.043	0.037	···	0.022	91.9
A_{30}-A_{26}	0.003	0.012	···	0.007	···	0.044	0.040	···	0.017	84.9
A_{31}-A_{26}	0.010	0.019	···	0.01	···	0.043	0.039	···	0.018	85.5
A_{28}-A_{27}	0.011	0.016	···	0.008	···	0.003	0.003	···	0.003	38.7
A_{29}-A_{27}	0.005	0.000	···	0.001	···	0.016	0.017	···	0.009	51.7
A_{30}-A_{27}	0.012	0.019	···	0.005	···	0.015	0.014	···	0.004	51.3
A_{31}-A_{27}	0.019	0.026	···	0.008	···	0.015	0.015	···	0.004	57.3
A_{29}-A_{28}	0.006	0.016	···	0.008	···	0.012	0.014	···	0.006	58.1
A_{30}-A_{28}	0.024	0.036	···	0.013	···	0.011	0.011	···	0.001	61.7
A_{31}-A_{28}	0.031	0.042	···	0.016	···	0.012	0.012	···	0.002	66.7
A_{30}-A_{29}	0.018	0.020	···	0.005	···	0.001	0.003	···	0.005	24.3
A_{31}-A_{29}	0.025	0.027	···	0.008	···	0.000	0.002	···	0.005	27.5
A_{31}-A_{30}	0.007	0.007	···	0.003	···	0.001	0.001	···	0.001	23.5

FIGURE 10. Matrix for direct comparison-based preselection. The absolute values of the subtraction between absorbances of two analytes. A_X = the averaged absorbance of three experiments of analyte **X**.

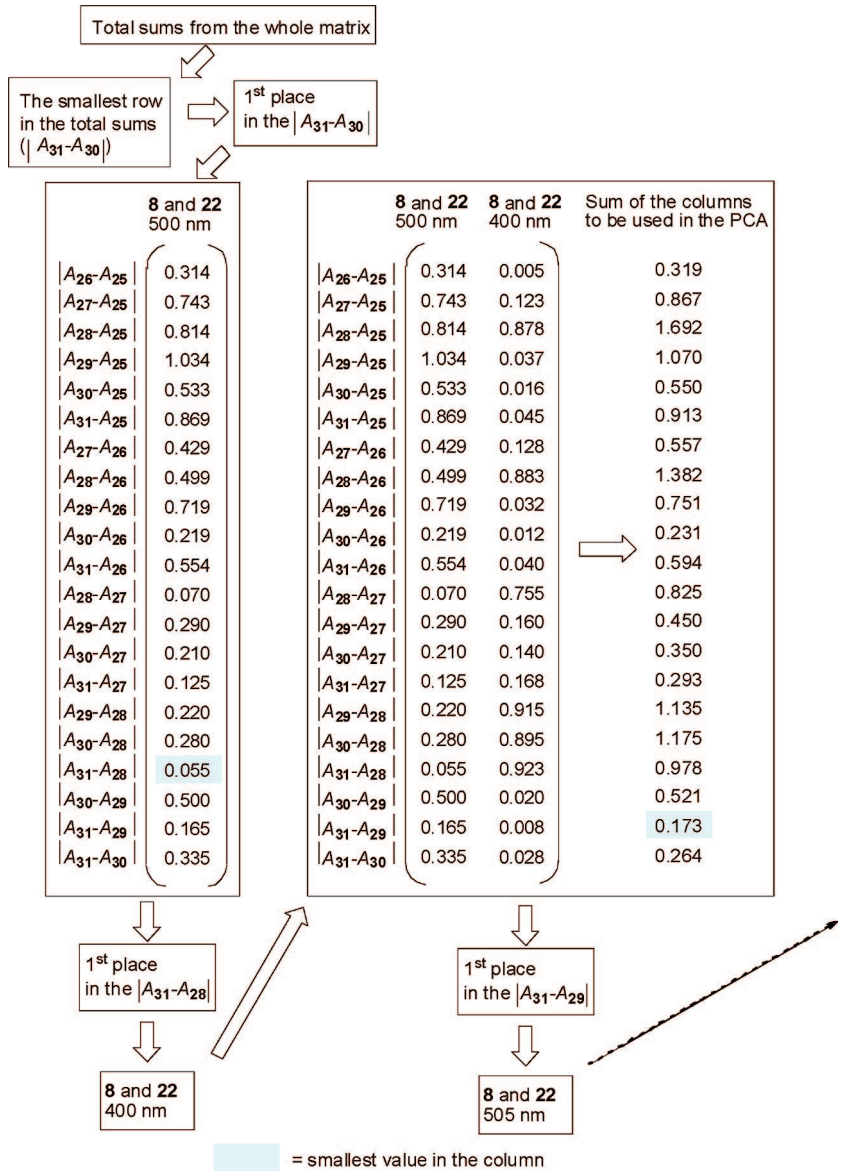
1403 variables (columns)

23 combinations of the receptors and the indicators

21 cases (rows)	**1** and **18** 61 λ's				···	**16** and **23** 61 λ's			
	400 nm	405 nm	···	700 nm	···	400 nm	405 nm	···	700 nm
A_{26}-A_{25}	688th	760th	···	948th	···	531th	550th	···	825th
A_{27}-A_{25}	923th	991th	···	1338th	···	542th	574th	···	1091th
A_{28}-A_{25}	861th	849th	···	1117th	···	711th	741th	···	1069th
A_{29}-A_{25}	868th	1015th	···	1346th	···	806th	851th	···	930th
A_{30}-A_{25}	1175th	1137th	···	1204th	···	749th	785th	···	1017th
A_{31}-A_{25}	1303th	974th	···	1137th	···	746th	782th	···	977th
A_{27}-A_{26}	995th	1076th	···	1335th	···	331th	350th	···	917th
A_{28}-A_{26}	872th	842th	···	1146th	···	506th	525th	···	948th
A_{29}-A_{26}	940th	1066th	···	1248th	···	623th	676th	···	843th
A_{30}-A_{26}	1263th	981th	···	1077th	···	537th	567th	···	887th
A_{31}-A_{26}	1037th	852th	···	1010th	···	582th	613th	···	868th
A_{28}-A_{27}	701th	560th	···	830th	···	1102th	1088th	···	1154th
A_{29}-A_{27}	1150th	1366th	···	1352th	···	797th	757th	···	1019th
A_{30}-A_{27}	818th	671th	···	1094th	···	771th	774th	···	1185th
A_{31}-A_{27}	772th	657th	···	1061th	···	866th	857th	···	1220th
A_{29}-A_{28}	1040th	721th	···	957th	···	816th	786th	···	1031th
A_{30}-A_{28}	630th	487th	···	828th	···	885th	909th	···	1348th
A_{31}-A_{28}	595th	455th	···	821th	···	918th	918th	···	1305th
A_{30}-A_{29}	427th	378th	···	935th	···	1293th	1086th	···	888th
A_{31}-A_{29}	365th	334th	···	867th	···	1389th	1261th	···	1033th
A_{31}-A_{30}	750th	735th	···	1025th	···	1331th	1259th	···	1302th

FIGURE 11. Matrix of rank ordering for direct comparison-based preselection. The rank ordering was done for each row.

SCHEME 4. Schematic Procedure of Direct Comparison-Based Preselection



following manner. The smallest sum of differences in one row means that the discrimination of these two analytes is the most difficult case. Therefore, selection of a column should be done to compensate this small sum. The highest place of rank order in the row of smallest total sum was used to choose a column of data that will be used in the PCA.

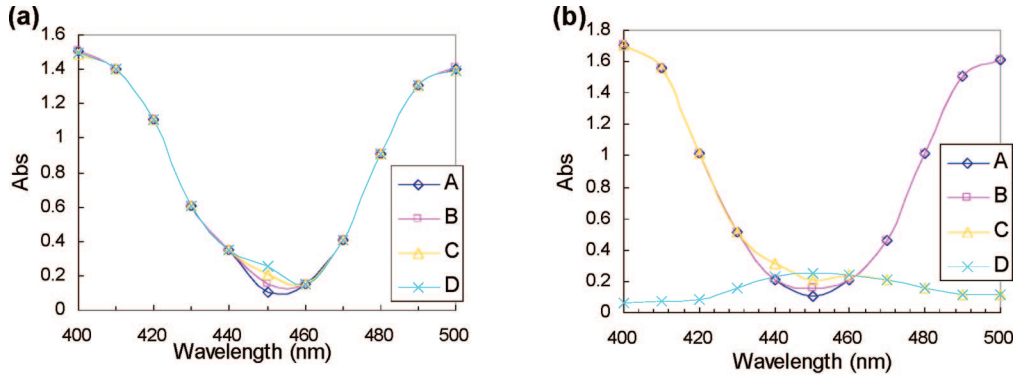
The procedure for the direct comparison-based preselection is illustrated in Scheme 4. As the first step in the preselection, the total sums of each column in the whole matrix (Figure 10) were used because there was no subset at the beginning. We first paid attention to the row $|A_{31}\text{-}A_{30}|$, because it had the lowest sum. The highest rank order in the $|A_{31}\text{-}A_{30}|$ now was found, corresponding to the column that involves receptor **8**, indicator **22**, and wavelength 500 nm, but row $A_{31}\text{-}A_{28}$ had the smallest value in the column of receptor **8**, indicator **22**, and wavelength 500 nm (shown in the light blue square in Scheme 4).

So we use this column in the final PCA, but we need to compensate for the lack of discrimination of **31** and **28**. Therefore, the first place of rank ordering in the row $A_{31}\text{-}A_{28}$ was found. It is receptor **8**, indicator **22**, and wavelength 400 nm. So, this column will be used in the final PCA also. Hence, in each case we found the lowest value of the sum and selected the one column that has the highest rank order to guide the next selection of a column. We again calculated the sums of each row for the expanded set of columns to be used. Finally, only 40 or 50 of the 1403 variables were selected for PCA (the results of PCA are shown later). This method was time-consuming because all of the cases needed to be compared.

Data Preprocessing. In chemometrics data preprocessing plays a very important role. Proper preprocessing can improve the results, but improper preprocessing leads to inaccurate interpretations. Unfortunately, no general guidelines exist to decide the suitable data preprocessing method.[3,24] In our study, it was instructive to examine several preprocessing strategies to determine which is best for PCA.

We decided to explore two preprocessing strategies, referred to as "mean centering" and "standardization" (see

(24) Brereton, R. G. *Chemometrics Data Analysis for the Laboratory and Chemical Plant*; Wiley: San Francisco, 2003; Chapter 4.



	$A_{400\ nm}$	$A_{450\ nm}$	$A_{500\ nm}$
A	1.5	0.1	1.4
	1.51	0.11	1.41
B	1.5	0.15	1.4
	1.52	0.16	1.42
C	1.5	0.2	1.4
	1.48	0.21	1.38
D	1.5	0.25	1.4
	1.49	0.26	1.39

	$A_{400\ nm}$	$A_{450\ nm}$	$A_{500\ nm}$
A	1.7	0.1	1.6
	1.72	0.11	1.62
B	1.7	0.15	1.6
	1.72	0.16	1.62
C	1.7	0.2	0.1
	1.72	0.21	0.12
D	0.05	0.25	0.1
	0.07	0.26	0.12

FIGURE 12. Imaginary UV−vis spectra of analytes **A**−**D** for demonstration of the two data preprocessing methods. Top: the averaged UV−vis spectra from two measurements. Bottom: the absorbances of two measurements at 400, 450, and 500 nm.

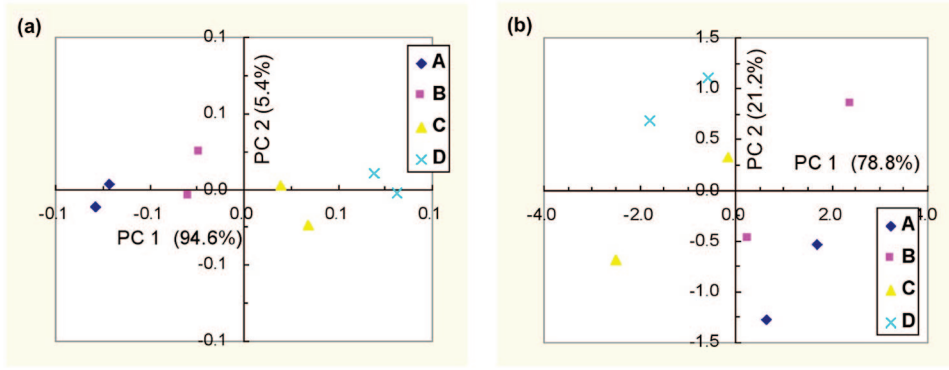


FIGURE 13. Score plots of first two PCs using the data set of Figure 12a. (a) Mean centering and (b) standardization.

Supporting Information for details). In mean-centering preprocessing the largest variances of each column dominate the PCA. Hence, small but meaningful variances become lost. On the other hand, standardization enhances small errors or meaningless data, because each variable has the same importance. To demonstrate these differences, the imaginary UV−vis spectra of analytes **A**−**D** with two repetitive experiments were created as shown in Figure 12. Using our direct comparison-based preselection method, 400, 450, and 500 nm were good choices for the PCA in both cases. In the case of Figure 12a, the mean centering was more practical than the standardization because standardization equally weights less relevant data at 400 and 500 nm (Figure 13). In the case of Figure 12b, standardization was more effective than the mean centering because mean centering deemphasized the data at 450 nm relative to that at 400 and 500 nm because the differences at 450 nm are small (Figure 14). So it is necessary to apply both preprocessing methods for the PCA using ANOVA-based preselection or our direct comparison-based preselection.

Principal Component Analysis. After ANOVA-based or our direct comparison-based preselection methods, PCA was applied with both mean centering and standardization preprocessing. The results are shown in Figures 15 and 16. The ANOVA-based preselection with mean centering worked well (Figure 15a), while analytes **30** and **31** overlapped with the standardization preprocessing (Figure 15b). The direct comparison-based preselection accomplished the discrimination of all analytes **25**−**31**, which means the data set includes the proper information for Feature 3 in Figure 7. The two data preprocessing methods showed acceptable results, although the shape of the score plots are different.

To compare the ability of a hybrid array of designed receptors and nondesigned receptors to a lack of design, only the data of nondesigned receptors (the variables chosen by the direct comparison-based preselection method) was employed for use in PCA (Figure 13). Analytes **27** and **31** in Figure 17a and **27** and **28** in Figure 17b were close to each other in the score plots. After addition of four variables from the designed receptor (**8**), good separation in the score plot was obtained (Figure 18). As a consequence, the array of nondesigned metal receptors and only one designed receptor (**8**) could discriminate the various structurally similar analytes.

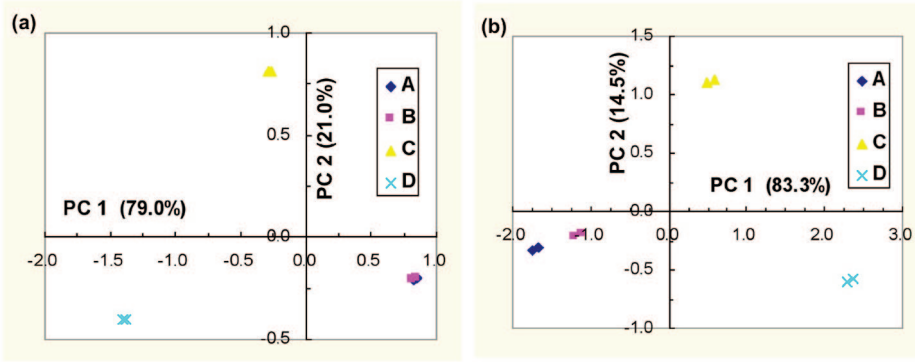


FIGURE 14. Score plots of first two PCs using the data set of Figure 8b. (a) Mean centering and (b) standardization. In score plot 1, the **A** and **B** overlap each other.

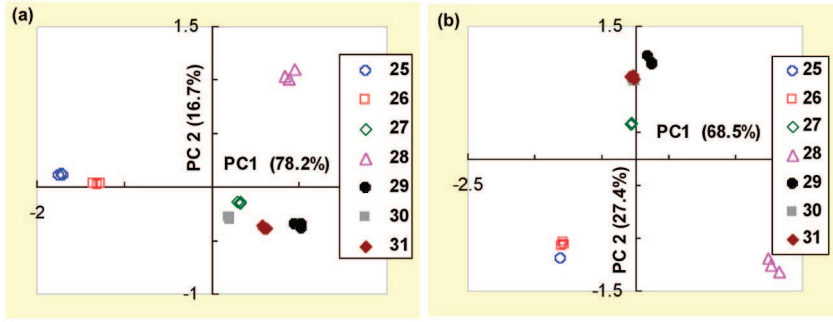


FIGURE 15. Score plots of 88 variables selected by $F > 300$ (ANOVA-based preselection). (a) Mean centering preprocessing and (b) standardization preprocessing.

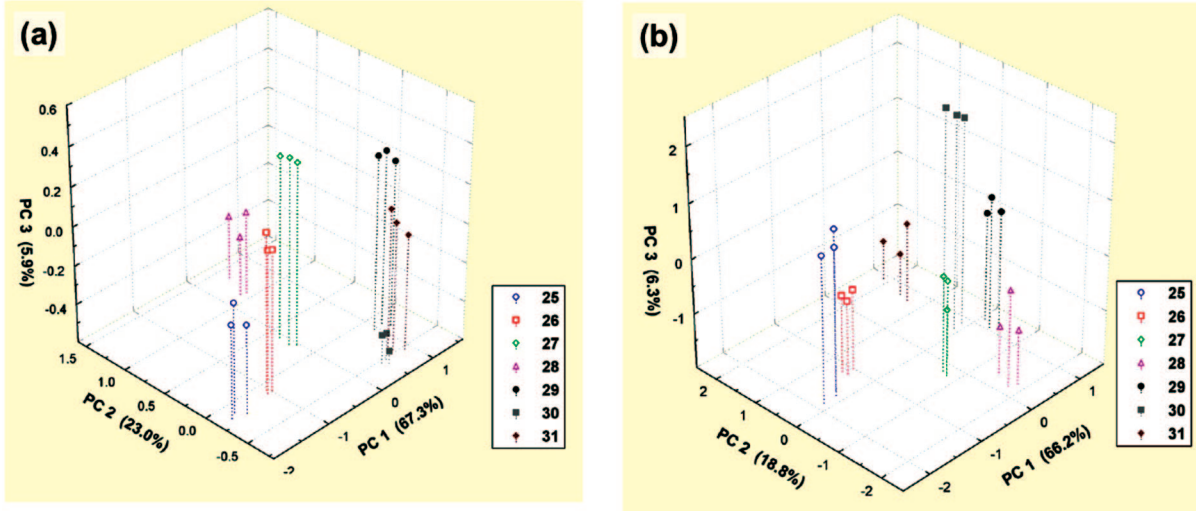


FIGURE 16. Score plots of 50 variables selected by direct comparison-based preselection. (a) Mean centering preprocessing and (b) standardization preprocessing.

Conclusion

In summary, we have studied the attribute need in the PCA discrimination of structurally similar analytes using an array of designed receptors and metal salts. In this system, the array of nondesigned metal salts and only one designed receptor could discriminate the seven analytes. The data from the metal salts have sufficient discrimination information because of the various coordination modes of the metal. On the basis of a direct-comparison data technique, we have also demonstrated that the choice of receptor−indicator−wavelength can be optimized to enhance the PCA. When a large amount of the data was used in the PCA, the analytes were split into only three structurally similar groups, such as α-hydroxycarboxylic acids, phenolic acids, and α-amino acids. However, it was possible to distinguish all of the analytes in the score plot when appropriate data was chosen by ANOVA-based or over direct comparison-based preselection. We are currently expanding this approach by using other commercially available receptors for analysis of mixtures of analytes.

Experimental Section

General. The UV−vis measurements were performed on a Bio-TEK Synergy HT Multi-Detection Microplate Reader using 96-well plates or a Beckman DU-640 spectrophotometer. pH mea-

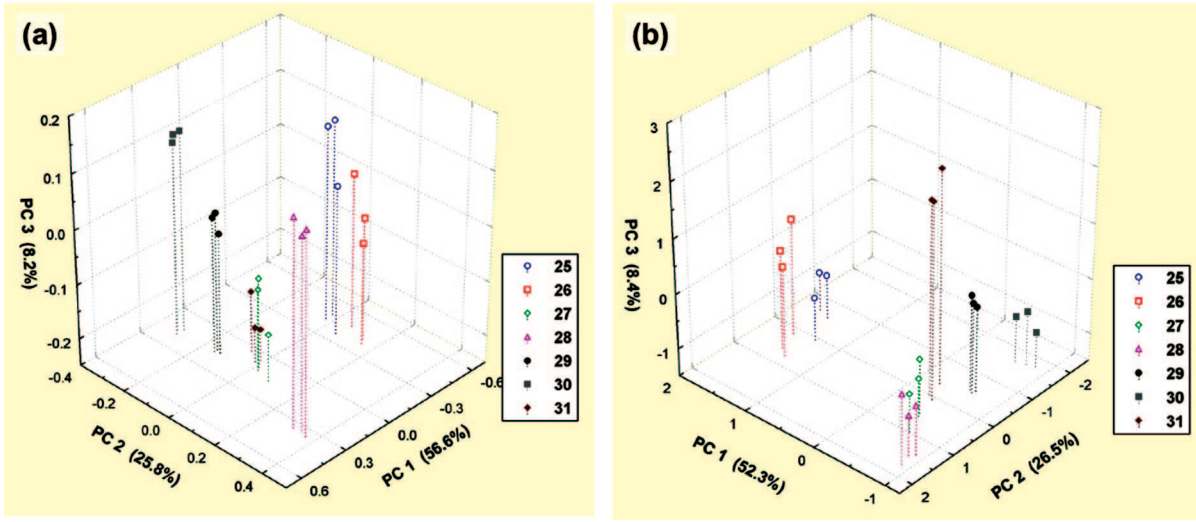


FIGURE 17. Score plots of 40 variables selected from UV−vis measurement of nondesigned metal receptors by direct comparison-based preselection. (a) Mean centering preprocessing and (b) standardization preprocessing.

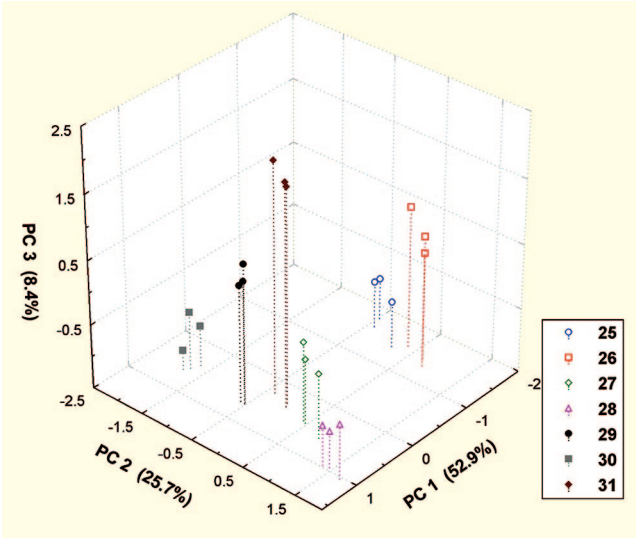


FIGURE 18. Score plot of the data set from nondesigned metal receptors (40 variables) and designed receptor (**11**) (4 variables). The data set was selected by direct comparison-based preselection. PCA was performed with standardization preprocessing.

surements were made using Orion 720A pH meter. PCA and ANOVA were performed using STATISTICA 5.5 or XLSTAT 7.5. Spectroscopic grade methanol was purchased from Fisher Scientific. Deionized water was used for preparation of stock solutions. Hydroscopic $ZnCl_2$ was dried under vacuum with heat before use. The other reagents were used as purchased from various commercial sources. Stock solutions of receptors, indicators, and analytes were prepared with 75% methanolic aqueous solution buffered with 10 mM HEPES (*N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid) at pH 7.5. Receptor solutions of **7** and **8** were prepared by mixing appropriate stoichiometry of the metal salt and the ligand (**7**, $ZnCl_2$:ligand = 2:1; **8**, $CuCl_2$:ligand = 2:1). Compounds that have poor solubility to methanolic water solution were dissolved using sonic waves or by standing overnight.

UV−vis Measurement for PCA. Each solution of the indicator, the receptor, and the analyte were mixed in a vessel of a 96-well plate. A 75% methanolic water solution buffered with 10 mM HEPES at pH 7.5 was added to the mixture to create a total volume of 300 μL. Final concentration of the indicators: [**18**] and [**19**] = 100 μM; [**20**], [**23**], and [**24**] = 50 μM; [**21**] and [**22**] = 30 μM. Final concentrations of the receptors and analytes were 150 μM and 1.0 mM, respectively. The UV−vis spectra (350−700 nm) were recorded after at least 5 min to allow the system to achieve equilibrium.

Acknowledgment. We gratefully acknowledge support from NSF (CHE-0716049) Welch Foundation (F-1151) for funding of this work. M.K. is grateful to the Japanese Society for the Promotion of Science for Young Scientists for a research fellow.

Supporting Information Available: Spectroscopic data of the ligand for **8** and **9**, the stoichiometry study of the ligand and metal ions for **8** and **9**, the binding stoichiometries of receptors **6**−**9** and metal salts **13**−**17** with the indicators, UV−vis spectra of indicators, indicator−analyte solutions and indicator-designed receptor solutions, UV−vis absorption spectra of indicator−receptor−analyte solutions for PCA. This material is available free of charge via the Internet at http://pubs.acs.org.

JO900433J