

Supporting Information:

High-resolution Cross Reactive Array for Alkaloids

Analysis of Array Classification Ability

The data collected with the array represent the fluorescence responses of 11 different sensors (**A-K**), to each of 15 analytes (**1-15**) at 5 or 6 different concentrations per analyte. There were six measurements taken for each analyte at each measured concentration. The background fluorescence signal, f_0 , for each batch of six measurements was taken as the average of the 0.0 concentration readings for each of the six measurements. Then, each fluorescence measurement was normalized by dividing by the corresponding f_0 for that analyte at that concentration.

Classification Schemes

Depending on the desired use of the sensor array there are several different classification schemes that can be used.

- I** *One class per analyte per concentration* - This classification scheme is useful if one wishes to determine concentration as well as analyte identity.
- II** *One class per analyte* - This method uses a single class to encompass all measurements for all concentrations of a single analyte. This allows the classifier to predict the class of an analyte, but it is unable to directly identify the concentration level. By grouping the measurements into fewer classes, however, we potentially gain more accuracy, as there are more measurements per class to train classifiers with.
- III** *One class for alkaloids, one class for steroids* - This allows the classifier to learn to distinguish alkaloids from steroids.

Table 1. Confusion matrix showing LOO classification errors for KNN ($k=6$). Here we use all 11 sensors and classification scheme **II**.

		Assigned Class														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
True Class	1	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	36	0	0	0	0	0	0	0	0	0	0	0	0	0
	3	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0
	4	0	0	0	30	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	3	27	0	0	0	0	0	0	0
	9	0	0	0	0	0	0	0	0	36	0	0	0	0	0	0
	10	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0
	11	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0
	13	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0
	14	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30
Total Errors:		3/468														
Accuracy Rate (95% conf.):		99.07% ± 0.72%														

Classifier Error Rates

Given a set of training data, the accuracy of a classifier trained with that data can be estimated, but never known exactly. We use the resampling technique called leave-one-out (LOO) to obtain an unbiased estimate of the error rate. We make the standard assumption that given a fixed number of trials, the number of missclassifications will be binomially distributed around the true rate, and use this to obtain 95% confidence intervals for the true error rate of the classifier [6].

We use nonparametric classification techniques. K-nearest neighbors is one of the simplest such techniques, yet it typically produces very good results. The method is computationally expensive, but not prohibitively so with our small number of training points (468). The resulting confusion matrix from the LOO error analysis of the KNN classifier used with classification scheme **II** is shown in Table 1. This shows the only classification errors occurring between **7** and **8**. We note all three of these errors are missclassifications at the lowest concentration level, and that **7** and **8** are well distinguished at higher concentrations. Furthermore, in Table 2, we have summarized the LOO error rates for several KNN classifiers with differing k values under each of the proposed classification schemes. The low rate of errors confirms that there is enough information in the feature vectors to accurately classify under any of the proposed schemes.

Additionally, we evaluated the LOO error estimates for several other classification techniques: support vector machines (SVMs) and classification trees.

SVM classifiers offer computationally efficient classification, which would become important if the number of training samples was dramatically increased. Specifically, we use C-SVM [8] for which a value of $C = 10$ provides good discriminatory power. We estimated error rates using a linear kernel as well as a radial basis function (RBF) kernel. The classifications were computed using the *libsvm* package [1], and results are reported in Table 2.

Classification trees are yet another family of non-parametric classification techniques. We use Ross Quinlan's C4.5 algorithm[7]. This algorithm recursively builds a binary decision tree, where each decision is a simple threshold measurement on a particular feature. The resulting decision tree classifier has the advantage of being extremely computationally cheap, as well as intuitively easy to understand. The method, however, is only able to use discriminate planes perpendicular to one of the axes, and thus performs poorly when optimal discriminate boundaries between classes are non-linear or linear but not perpendicular to any axis. In general, the LOO accuracy for the C4.5 decision trees was much poorer than the other techniques, however, under classification scheme **III** results were acceptable. The resulting decision tree is shown in Figure 1.

When comparing the performance of two classifier inference methods for a particular dataset, it is not always possible to confidently determine if one classifier is better than the other [3]. Therefore, we do not attempt to draw any conclusions about the best classifier method. It is evident from Table 1 that any of the classification methods will produce acceptably low error rates, with the possible exceptions of linear SVM under scheme **II** and C4.5 under either scheme **I** or **II**.

Table 2. Classifier method vs. classification scheme comparison matrix. All error rates are estimated to 95% confidence.

		Classification Scheme		
		I	II	III
Classifier	KNN K=3	99.49% ± 0.51%	99.49% ± 0.51%	99.61% ± 0.29%
	KNN K=4	99.61% ± 0.29%	99.61% ± 0.29%	99.61% ± 0.29%
	KNN K=5	98.64% ± 0.93%	98.64% ± 0.93%	99.61% ± 0.29%
	KNN K=6	99.07% ± 0.72%	99.07% ± 0.72%	99.61% ± 0.29%
	KNN K=7	98.64% ± 0.93%	98.43% ± 1.48%	99.61% ± 0.29%
	SVM Linear C = 10	99.61% ± 0.29%	96.94% ± 1.72%	99.61% ± 0.29%
	SVM RBF C = 10, $\gamma = 1$	99.61% ± 0.29%	99.28% ± 0.59%	99.61% ± 0.29%
	C4.5	85.79% ± 3.14%	88.87% ± 2.82%	98.86% ± 0.83%

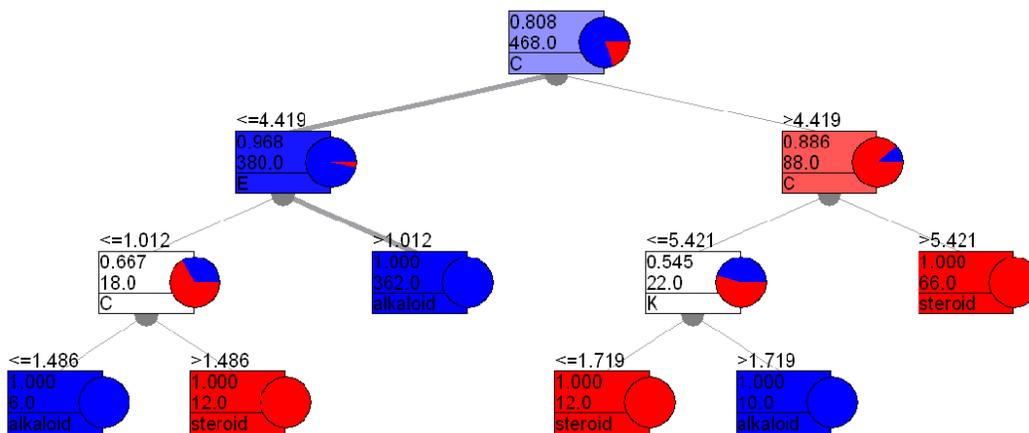


Figure 1: This graph shows the decision tree derived for classification scheme III using the C4.5 algorithm. Each leaf node is labeled at the bottom by the final classification decision, either "alkaloid" or "steroid". Non-leaf nodes are labeled by the sensor name on which the split will be made. The edges connecting a node to its children are labeled by the result of the decision. Also, each node is labeled with 2 other numbers. The upper number is the majority class probability and the lower number is the number of training samples at that node. The pie charts show the distribution of training samples at each node, and the node colors correspond to the majority class probability. This tree achieves a 98.86% ± 0.83% LOO accuracy rate.

Selection of Sensors

Given the very low error rates found with the full 11-sensor array, we want to know how the array can be simplified to reduce the number of sensors while still maintaining a low error rate. To do this, we used the LOO method to estimate the KNN (k=6) classification error rate for each possible subset of the original 11 sensors. This was done using the one class per analyte classification scheme (II). The results shown below are all arrays with a 95% confidence bound for the error rate below 2%. We have sorted the arrays by error-rate and array size.

One of the most accurate 6-sensor array is [BCDEGH] with error rate $98.9\% \pm 0.8\%$. The addition of sensor **K**, brings the error rate up to $99.1\% \pm 0.7\%$ accuracy which is the same as the full 11-sensor array. The 6-sensor array's LOO error-rate estimation corresponds to 4 of 468 missclassifications. All four missclassifications occur between **7** and **8** at the lowest concentration level. As it turns out, the error are exactly the same for this array under either classification scheme **I** or **II**.

Finally, we have also provided auxiliary tables that organize the same subset analysis in a different way. In Table 3 – Table 10 we have listed the top arrays with $N = 3, \dots, 10$ sensors. Also, since sensors **B** and **G** are included in almost every top ranked array, we were interested in how the sensor selection would work in their absence. Hence, we have included the top ranked arrays without sensors **B** and **G** as Table 11. We see that even with all of the other sensors, it is not possible to get better than a $95.5\% \pm 1.8\%$ accuracy. This shows the importance of sensors **B** and **G** to the classification of the chosen analytes.

Size	Sensors											Estimated Accuracy
7	A	B			E	F	G	H			K	$99.1\% \pm 0.7\%$
7	A	B			E		G	H		J	K	$99.1\% \pm 0.7\%$
7		B	C	D	E		G	H			K	$99.1\% \pm 0.7\%$
7		B	C	D	E	F	G	H				$99.1\% \pm 0.7\%$
7		B			E	F	G	H		J	K	$99.1\% \pm 0.7\%$
8		B	C	D	E	F	G	H			K	$99.1\% \pm 0.7\%$
8		B	C	D	E	F	G	H		J		$99.1\% \pm 0.7\%$
8		B	C	D	E		G	H	I		K	$99.1\% \pm 0.7\%$
8		B	C	D	E	F	G			J	K	$99.1\% \pm 0.7\%$
8	A	B	C	D	E	F	G				K	$99.1\% \pm 0.7\%$
8		B	C	D	E	F	G		I		K	$99.1\% \pm 0.7\%$
8	A	B	C	D	E		G	H			K	$99.1\% \pm 0.7\%$
8	A	B			E	F	G	H		J	K	$99.1\% \pm 0.7\%$
8		B	C	D	E		G	H		J	K	$99.1\% \pm 0.7\%$
8		B	C	D	E	F	G	H	I			$99.1\% \pm 0.7\%$
8	A	B			E	F	G	H	I		K	$99.1\% \pm 0.7\%$
9	A	B	C	D	E	F	G	H			K	$99.1\% \pm 0.7\%$
9		B	C	D	E	F	G		I	J	K	$99.1\% \pm 0.7\%$
9	A	B	C	D	E	F	G			J	K	$99.1\% \pm 0.7\%$
9		B	C	D	E		G	H	I	J	K	$99.1\% \pm 0.7\%$
9		B	C	D	E	F	G	H	I		K	$99.1\% \pm 0.7\%$
9	A	B			E	F	G	H	I	J	K	$99.1\% \pm 0.7\%$
9	A	B	C	D	E		G	H		J	K	$99.1\% \pm 0.7\%$
9		B	C	D	E	F	G	H		J	K	$99.1\% \pm 0.7\%$
9	A	B	C	D	E	F	G		I		K	$99.1\% \pm 0.7\%$

Continued on the next page

Size	Sensors											Estimated Accuracy
9		B	C	D	E	F	G	H	I	J		99.1% ± 0.7%
9	A	B	C	D	E		G	H	I		K	99.1% ± 0.7%
10	A	B	C	D	E	F	G	H		J	K	99.1% ± 0.7%
10		B	C	D	E	F	G	H	I	J	K	99.1% ± 0.7%
10	A	B	C	D	E	F	G	H	I		K	99.1% ± 0.7%
10	A	B	C	D	E		G	H	I	J	K	99.1% ± 0.7%
10	A	B	C	D	E	F	G		I	J	K	99.1% ± 0.7%
11	A	B	C	D	E	F	G	H	I	J	K	99.1% ± 0.7%
6		B	C	D	E		G	H				98.9% ± 0.8%
6		B			E	F	G	H			K	98.9% ± 0.8%
6		B			E	F	G			J	K	98.9% ± 0.8%
6		B	C	D			G	H			K	98.9% ± 0.8%
7		B	C	D		F	G	H			K	98.9% ± 0.8%
7		B			E	F	G	H	I		K	98.9% ± 0.8%
7		B	C	D	E		G	H	I			98.9% ± 0.8%
7	A	B			E	F	G			J	K	98.9% ± 0.8%
7	A	B		D	E		G		I		K	98.9% ± 0.8%
7		B	C	D	E	F	G				K	98.9% ± 0.8%
7	A	B		D	E		G	H		J		98.9% ± 0.8%
7	A	B		D	E	F	G				K	98.9% ± 0.8%
7	A	B	C	D			G	H			K	98.9% ± 0.8%
7	A	B		D	E		G	H			K	98.9% ± 0.8%
7		B	C	D			G	H	I		K	98.9% ± 0.8%
7		B			E		G	H	I	J	K	98.9% ± 0.8%
7		B	C	D	E		G	H		J		98.9% ± 0.8%
8	A	B		D	E	F	G		I		K	98.9% ± 0.8%
8		B	C	D		F	G	H	I		K	98.9% ± 0.8%
8	A	B	C		E		G		I	J	K	98.9% ± 0.8%
8	A	B	C	D		F	G	H			K	98.9% ± 0.8%
8	A	B		D	E		G		I	J	K	98.9% ± 0.8%
8	A	B	C	D	E		G			J	K	98.9% ± 0.8%
8	A	B	C	D	E	F	G	H				98.9% ± 0.8%
8	A	B		D	E	F	G	H			K	98.9% ± 0.8%
8	A	B		D	E		G	H	I		K	98.9% ± 0.8%
8		B			E	F	G	H	I	J	K	98.9% ± 0.8%
8	A	B		D	E		G	H	I	J		98.9% ± 0.8%
8	A	B	C	D			G	H	I		K	98.9% ± 0.8%
8	A	B		D	E		G	H		J	K	98.9% ± 0.8%
8		B	C	D		F	G	H		J	K	98.9% ± 0.8%
8	A	B		D	E	F	G			J	K	98.9% ± 0.8%
8	A	B			E	F	G		I	J	K	98.9% ± 0.8%
8		B	C	D	E		G	H	I	J		98.9% ± 0.8%
8	A	B			E		G	H	I	J	K	98.9% ± 0.8%
8	A	B		D	E	F	G	H		J		98.9% ± 0.8%

Continued on the next page

Size	Sensors											Estimated Accuracy
9	A	B		D	E	F	G	H	I	J		98.9% ± 0.8%
9	A	B	C	D	E		G		I	J	K	98.9% ± 0.8%
9	A	B		D	E	F	G	H		J	K	98.9% ± 0.8%
9	A	B		D	E	F	G		I	J	K	98.9% ± 0.8%
9	A	B	C	D	E	F	G	H		J		98.9% ± 0.8%
9		B	C	D		F	G	H	I	J	K	98.9% ± 0.8%
9	A	B		D	E	F	G	H	I		K	98.9% ± 0.8%
9	A	B	C	D		F	G	H	I		K	98.9% ± 0.8%
9	A	B	C		E		G	H	I	J	K	98.9% ± 0.8%
9	A	B	C	D		F	G	H		J	K	98.9% ± 0.8%
9	A	B	C	D	E	F	G	H	I			98.9% ± 0.8%
9	A	B		D	E		G	H	I	J	K	98.9% ± 0.8%
10	A	B	C	D		F	G	H	I	J	K	98.9% ± 0.8%
10	A	B		D	E	F	G	H	I	J	K	98.9% ± 0.8%
10	A	B	C	D	E	F	G	H	I	J		98.9% ± 0.8%
6		B	C	D			G		I	J		98.6% ± 0.9%
6	A	B		D	E		G	H				98.6% ± 0.9%
6		B			E		G	H		J	K	98.6% ± 0.9%
6	A	B			E		G			J	K	98.6% ± 0.9%
6		B	C	D			G			J	K	98.6% ± 0.9%
6		B	C	D	E		G			J		98.6% ± 0.9%
6		B	C	D	E	F	G					98.6% ± 0.9%
6	A	B		D	E		G				K	98.6% ± 0.9%
7	A	B			E	F	G	H		J		98.6% ± 0.9%
7		B		D	E	F	G		I		K	98.6% ± 0.9%
7	A	B		D	E	F	G	H				98.6% ± 0.9%
7	A	B		D	E		G			J	K	98.6% ± 0.9%
7		B	C	D	E	F	G			J		98.6% ± 0.9%
7		B	C	D			G		I	J	K	98.6% ± 0.9%
7	A	B		D	E		G	H	I			98.6% ± 0.9%
7	A	B	C		E	F	G	H				98.6% ± 0.9%
7		B	C	D	E		G		I	J		98.6% ± 0.9%
7	A	B	C		E		G			J	K	98.6% ± 0.9%
7		B	C	D		F	G	H		J		98.6% ± 0.9%
7		B	C	D	E	F	G		I			98.6% ± 0.9%
7		B			E	F	G		I	J	K	98.6% ± 0.9%
7		B	C	D			G	H		J	K	98.6% ± 0.9%
7		B	C	D		F	G		I	J		98.6% ± 0.9%
7		B	C	D	E		G			J	K	98.6% ± 0.9%
7	A	B	C	D			G			J	K	98.6% ± 0.9%
7		B		D	E	F	G			J	K	98.6% ± 0.9%
7		B	C	D		F	G			J	K	98.6% ± 0.9%
7	A	B	C	D	E		G	H				98.6% ± 0.9%
7		B	C	D	E		G		I		K	98.6% ± 0.9%

Continued on the next page

Size	Sensors											Estimated Accuracy
8	A	B	C	D		F	G			J	K	98.6% ± 0.9%
8		B	C	D	E		G		I	J	K	98.6% ± 0.9%
8		B	C	D		F	G		I	J	K	98.6% ± 0.9%
8		B		D	E	F	G		I	J	K	98.6% ± 0.9%
8		B	C	D		F	G	H	I	J		98.6% ± 0.9%
8	A	B		D	E	F	G	H	I			98.6% ± 0.9%
8		B	C	D	E	F	G		I	J		98.6% ± 0.9%
8	A	B	C		E		G	H		J	K	98.6% ± 0.9%
8	A	B	C		E		G	H	I		K	98.6% ± 0.9%
8	A	B	C	D	E	F	G			J		98.6% ± 0.9%
8		B		D	E	F	G	H		J	K	98.6% ± 0.9%
8	A	B	C	D	E		G		I		K	98.6% ± 0.9%
8		B	C	D			G	H	I	J	K	98.6% ± 0.9%
8	A	B	C	D	E		G	H		J		98.6% ± 0.9%
8	A	B	C	D			G	H		J	K	98.6% ± 0.9%
8	A	B			E	F	G	H	I	J		98.6% ± 0.9%
8	A	B	C	D			G		I	J	K	98.6% ± 0.9%
8	A	B	C		E	F	G	H		J		98.6% ± 0.9%
8	A	B	C	D	E		G	H	I			98.6% ± 0.9%
9	A	B	C	D			G	H	I	J	K	98.6% ± 0.9%
9		B		D	E	F	G	H	I	J	K	98.6% ± 0.9%
9	A	B	C	D	E		G	H	I	J		98.6% ± 0.9%
9	A	B	C	D	E	F	G		I	J		98.6% ± 0.9%
9	A	B	C	D		F	G		I	J	K	98.6% ± 0.9%
5		B	C	D			G			J		98.4% ± 1.0%
5		B			E	F	G				K	98.4% ± 1.0%
6		B	C	D			G	H		J		98.4% ± 1.0%
6	A	B			E	F	G	H				98.4% ± 1.0%
6	A	B	C	D			G			J		98.4% ± 1.0%
6		B	C	D		F	G			J		98.4% ± 1.0%
6	A	B			E	F	G				K	98.4% ± 1.0%
6	A	B		D	E		G		I			98.4% ± 1.0%
6	A	B			E		G	H			K	98.4% ± 1.0%
6		B	C		E	F	G	H				98.4% ± 1.0%
6		B		D	E	F	G			J		98.4% ± 1.0%
6		B	C	D			G		I		K	98.4% ± 1.0%
7		B	C		E	F	G	H	I			98.4% ± 1.0%
7		B		D	E	F	G	H		J		98.4% ± 1.0%
7	A	B			E		G	H	I		K	98.4% ± 1.0%
7		B	C	D			G	H	I	J		98.4% ± 1.0%
7	A	B	C	D			G		I	J		98.4% ± 1.0%
7		B	C		E	F	G		I	J		98.4% ± 1.0%
7	A	B		D	E	F	G			J		98.4% ± 1.0%
7		B		D	E		G	H	I	J		98.4% ± 1.0%

Continued on the next page

Size	Sensors										Estimated Accuracy
7	A	B		D		F	G	H		K	98.4% ± 1.0%
7		B	C		E	F	G	H		J	98.4% ± 1.0%
7	A	B			E	F	G	H	I		98.4% ± 1.0%
7		B	C		E		G	H		J	98.4% ± 1.0%
7		B		D	E		G		I	J	98.4% ± 1.0%
7	A	B	C	D	E		G			J	98.4% ± 1.0%
7	A	B			E		G		I	J	98.4% ± 1.0%
7		B	C	D		F	G		I		98.4% ± 1.0%
7	A	B	C	D		F	G			J	98.4% ± 1.0%
7	A	B	C	D	E	F	G				98.4% ± 1.0%
7	A	B	C	D	E		G				98.4% ± 1.0%
8		B		D	E	F	G	H	I	J	98.4% ± 1.0%
8		B		D	E	F	G	H	I		98.4% ± 1.0%
8	A	B	C			F	G	H		J	98.4% ± 1.0%
8		B	C		E		G	H	I	J	98.4% ± 1.0%
8	A	B	C	D		F	G		I	J	98.4% ± 1.0%
8	A	B	C	D		F	G	H		J	98.4% ± 1.0%
8	A	B		D	E	F	G		I	J	98.4% ± 1.0%
8	A	B	C	D		F	G	H	I		98.4% ± 1.0%
8	A	B	C	D		F	G		I		98.4% ± 1.0%
8	A	B	C	D	E		G		I	J	98.4% ± 1.0%
8		B	C		E	F	G	H	I	J	98.4% ± 1.0%
8	A	B	C		E	F	G	H	I		98.4% ± 1.0%
8	A	B	C		E	F	G			J	98.4% ± 1.0%
9	A	B	C	D		F	G	H	I	J	98.4% ± 1.0%
9	A	B	C		E	F	G	H	I	J	98.4% ± 1.0%
9	A	B	C		E	F	G	H		J	98.4% ± 1.0%
9	A	B	C		E	F	G	H	I		98.4% ± 1.0%
10	A	B	C		E	F	G	H	I	J	98.4% ± 1.0%
5	A	B		D	E		G				98.2% ± 1.1%
6	A	B		D	E		G			J	98.2% ± 1.1%
6		B	C	D		F	G				98.2% ± 1.1%
6		B	C	D		F	G	H			98.2% ± 1.1%
6	A	B	C	D			G	H			98.2% ± 1.1%
6		B	C		E		G			J	98.2% ± 1.1%
6		B	C	D	E		G				98.2% ± 1.1%
6	A	B			E	F	G			J	98.2% ± 1.1%
6		B		D	E	F	G				98.2% ± 1.1%
6		B			E	F	G	H	I		98.2% ± 1.1%
6	A	B	C	D	E		G				98.2% ± 1.1%
6		B	C	D	E		G		I		98.2% ± 1.1%
6		B		D	E		G		I		98.2% ± 1.1%
6		B		D	E		G			J	98.2% ± 1.1%
6		B			E		G	H	I		98.2% ± 1.1%

Continued on the next page

Size	Sensors										Estimated Accuracy	
7		B		D	E	F	G		I	J		98.2% ± 1.1%
7	A	B	C	D		F	G				K	98.2% ± 1.1%
7		B	C		E		G	H	I	J		98.2% ± 1.1%
7		B		D	E		G	H		J	K	98.2% ± 1.1%
7	A	B	C	D			G		I		K	98.2% ± 1.1%
7	A	B	C	D		F	G	H				98.2% ± 1.1%
7	A	B	C		E	F	G			J		98.2% ± 1.1%
7		B	C	D		F	G	H	I			98.2% ± 1.1%
7		B		D	E	F	G	H	I			98.2% ± 1.1%
7	A	B	C	D	E		G		I			98.2% ± 1.1%
7	A	B	C		E		G	H			K	98.2% ± 1.1%
7	A	B			E	F	G		I		K	98.2% ± 1.1%
7	A	B		D	E		G		I	J		98.2% ± 1.1%
7		B		D	E	F	G	H			K	98.2% ± 1.1%
7		B			E	F	G	H	I	J		98.2% ± 1.1%
7		B	C		E		G		I	J	K	98.2% ± 1.1%
7	A	B	C	D			G	H		J		98.2% ± 1.1%
7		B	C		E		G	H	I		K	98.2% ± 1.1%
8		B	C		E	F	G	H		J	K	98.2% ± 1.1%
8	A	B	C	D			G	H	I	J		98.2% ± 1.1%
8	A	B	C	D	E	F	G		I			98.2% ± 1.1%
8		B	C		E	F	G		I	J	K	98.2% ± 1.1%
8	A	B		D		F	G	H	I		K	98.2% ± 1.1%
8	A	B	C		E	F	G		I	J		98.2% ± 1.1%
8	A	B	C		E	F	G	H			K	98.2% ± 1.1%
8	A	B	C				G	H	I	J	K	98.2% ± 1.1%
8	A	B	C			F	G	H	I		K	98.2% ± 1.1%
8		B		D	E		G	H	I	J	K	98.2% ± 1.1%
8	A	B	C		E	F	G		I		K	98.2% ± 1.1%
9	A		C	D	E	F	G	H		J	K	98.2% ± 1.1%
9		B	C		E	F	G	H	I	J	K	98.2% ± 1.1%
9	A	B	C			F	G	H	I	J	K	98.2% ± 1.1%
9	A	B	C		E	F	G		I	J	K	98.2% ± 1.1%
5		B	C	D			G	H				98.0% ± 1.2%
5		B	C		E	F	G					98.0% ± 1.2%
5		B		D	E		G			J		98.0% ± 1.2%
6		B	C		E		G	H	I			98.0% ± 1.2%
6		B	C		E	F	G		I			98.0% ± 1.2%
6	A	B		D			G	H			K	98.0% ± 1.2%
6		B			E	F	G	H		J		98.0% ± 1.2%
6		B			E	F	G		I	J		98.0% ± 1.2%
6		B		D	E		G		I	J		98.0% ± 1.2%
6		B		D	E		G	H		J		98.0% ± 1.2%
6		B	C		E	F	G				K	98.0% ± 1.2%

Continued on the next page

Size	Sensors											Estimated Accuracy
6	A	B			E	F	G		I			98.0% ± 1.2%
6	A	B		D	E	F	G					98.0% ± 1.2%
6		B	C		E	F	G			J		98.0% ± 1.2%
7	A	B	C			F	G		I	J		98.0% ± 1.2%
7	A	B			E		G	H	I	J		98.0% ± 1.2%
7	A	B	C		E		G	H		J		98.0% ± 1.2%
7		B	C			F	G	H	I		K	98.0% ± 1.2%
7		B	C		E	F	G	H			K	98.0% ± 1.2%
7	A	B	C				G	H		J	K	98.0% ± 1.2%
7	A	B	C		E	F	G		I			98.0% ± 1.2%
7	A	B	C			F	G	H			K	98.0% ± 1.2%
7	A	B		D	E	F	G		I			98.0% ± 1.2%
7		B		D	E		G	H	I		K	98.0% ± 1.2%
7	A	B		D			G	H	I		K	98.0% ± 1.2%
7		B	C			F	G	H		J	K	98.0% ± 1.2%
7	A	B	C				G	H	I		K	98.0% ± 1.2%
7	A	B			E	F	G		I	J		98.0% ± 1.2%
7		B	C		E	F	G			J	K	98.0% ± 1.2%
7	A	B	C	D			G	H	I			98.0% ± 1.2%
7	A	B	C		E		G		I		K	98.0% ± 1.2%
8			C	D	E	F	G	H		J	K	98.0% ± 1.2%
8		B	C			F	G	H	I	J	K	98.0% ± 1.2%
8	A	B	C			F	G		I	J	K	98.0% ± 1.2%
8		B	C		E	F	G	H	I		K	98.0% ± 1.2%
8	A	B		D		F	G	H		J	K	98.0% ± 1.2%
8	A	B	C		E		G	H	I	J		98.0% ± 1.2%
10	A		C	D	E	F	G	H	I	J	K	98.0% ± 1.2%

Table 3. Size 3 arrays with highest KNN($k=6$) classification accuracy under classification scheme II.

Sensors										Estimated Accuracy
	B		D			G				94.4% \pm 2.0%
	B	C	D							94.0% \pm 2.1%
	B					G			K	93.3% \pm 2.2%
A	B					G				93.1% \pm 2.2%
	B		D	E						92.3% \pm 2.4%
	B			E		G				91.8% \pm 2.4%
	B				F	G				91.6% \pm 2.5%
	B	C				G				91.4% \pm 2.5%
	B		D		F					91.2% \pm 2.5%
A	B		D							91.2% \pm 2.5%
	B		D						J	89.9% \pm 2.7%
	B	C		E						89.9% \pm 2.7%
A	B			E						89.9% \pm 2.7%
	B			E	F					89.7% \pm 2.7%
A	B								J	89.7% \pm 2.7%
	B		D						K	89.3% \pm 2.8%
	B					G			J	88.9% \pm 2.8%
	B					G		I		88.7% \pm 2.8%
A	B				F					88.7% \pm 2.8%
	B		D					I		88.4% \pm 2.9%

Table 4. Size 4 arrays with highest KNN(k-6) classification accuracy under classification scheme II.

Sensors										Estimated Accuracy	
	B		D	E		G					97.2% ± 1.4%
A	B		D			G					97.2% ± 1.4%
	B			E	F	G					96.9% ± 1.5%
	B	C	D			G					96.7% ± 1.5%
	B		D		F	G					96.5% ± 1.6%
	B				F	G				K	96.3% ± 1.6%
	B			E		G	H				96.3% ± 1.6%
	B		D			G				K	96.3% ± 1.6%
	B			E		G				K	96.1% ± 1.7%
	B	C	D	E							96.1% ± 1.7%
	B		D			G				J	95.9% ± 1.7%
	B		D			G		I			95.9% ± 1.7%
	B	C		E		G					95.9% ± 1.7%
A	B				F	G					95.9% ± 1.7%
	B	C				G				J	95.7% ± 1.8%
	B	C	D		F						95.7% ± 1.8%
A	B		D	E							95.7% ± 1.8%
	B					G				J K	95.5% ± 1.8%
	B	C	D							J	95.5% ± 1.8%
A	B			E		G					95.5% ± 1.8%

Table 5. Size 5 arrays with highest KNN(k-6) classification accuracy under classification scheme II.

Sensors										Estimated Accuracy	
	B			E	F	G				K	98.4% ± 1.0%
	B	C	D			G				J	98.4% ± 1.0%
A	B		D	E		G					98.2% ± 1.1%
	B		D	E		G				J	98.0% ± 1.2%
	B	C		E	F	G					98.0% ± 1.2%
	B	C	D			G	H				98.0% ± 1.2%
	B			E		G	H			K	97.8% ± 1.2%
	B			E	F	G				J	97.8% ± 1.2%
	B		D	E		G				K	97.8% ± 1.2%
	B		D	E		G		I			97.8% ± 1.2%
	B		D	E	F	G					97.8% ± 1.2%
	B	C	D			G				K	97.8% ± 1.2%
	B			E		G				J K	97.6% ± 1.3%
	B			E	F	G	H				97.6% ± 1.3%
	B	C	D		F	G					97.6% ± 1.3%
	B	C	D	E		G					97.6% ± 1.3%
A	B				F	G				K	97.6% ± 1.3%
A	B			E	F	G					97.6% ± 1.3%
	B			E	F	G		I			97.4% ± 1.4%
	B		D		F	G		I			97.4% ± 1.4%

Table 6. Size 6 arrays with highest KNN(k-6) classification accuracy under classification scheme II.

Sensors										Estimated Accuracy	
	B			E	F	G			J	K	98.9% ± 0.8%
	B			E	F	G	H			K	98.9% ± 0.8%
	B	C	D			G	H			K	98.9% ± 0.8%
	B	C	D	E		G	H				98.9% ± 0.8%
	B			E		G	H		J	K	98.6% ± 0.9%
	B	C	D			G			J	K	98.6% ± 0.9%
	B	C	D			G		I	J		98.6% ± 0.9%
	B	C	D	E		G			J		98.6% ± 0.9%
	B	C	D	E	F	G					98.6% ± 0.9%
A	B			E		G			J	K	98.6% ± 0.9%
A	B		D	E		G				K	98.6% ± 0.9%
A	B		D	E		G	H				98.6% ± 0.9%
	B		D	E	F	G			J		98.4% ± 1.0%
	B	C		E	F	G	H				98.4% ± 1.0%
	B	C	D			G		I		K	98.4% ± 1.0%
	B	C	D			G	H		J		98.4% ± 1.0%
	B	C	D		F	G			J		98.4% ± 1.0%
A	B			E		G	H			K	98.4% ± 1.0%
A	B			E	F	G				K	98.4% ± 1.0%
A	B			E	F	G	H				98.4% ± 1.0%

Table 7. Size 7 arrays with highest KNN(k-6) classification accuracy under classification scheme II.

Sensors										Estimated Accuracy	
	B			E	F	G	H		J	K	99.1% ± 0.7%
	B	C	D	E		G	H			K	99.1% ± 0.7%
	B	C	D	E	F	G	H				99.1% ± 0.7%
A	B			E		G	H		J	K	99.1% ± 0.7%
A	B			E	F	G	H			K	99.1% ± 0.7%
	B			E		G	H	I	J	K	98.9% ± 0.8%
	B			E	F	G	H	I		K	98.9% ± 0.8%
	B	C	D			G	H	I		K	98.9% ± 0.8%
	B	C	D		F	G	H			K	98.9% ± 0.8%
	B	C	D	E		G	H		J		98.9% ± 0.8%
	B	C	D	E		G	H	I			98.9% ± 0.8%
	B	C	D	E	F	G				K	98.9% ± 0.8%
A	B			E	F	G			J	K	98.9% ± 0.8%
A	B		D	E		G		I		K	98.9% ± 0.8%
A	B		D	E		G	H			K	98.9% ± 0.8%
A	B		D	E		G	H		J		98.9% ± 0.8%
A	B		D	E	F	G				K	98.9% ± 0.8%
A	B	C	D			G	H			K	98.9% ± 0.8%
	B			E	F	G		I	J	K	98.6% ± 0.9%
	B		D	E	F	G			J	K	98.6% ± 0.9%

Table 8. Size 8 arrays with highest KNN(k-6) classification accuracy under classification scheme II.

Sensors											Estimated Accuracy
	B	C	D	E		G	H		J	K	99.1% ± 0.7%
	B	C	D	E		G	H	I		K	99.1% ± 0.7%
	B	C	D	E	F	G			J	K	99.1% ± 0.7%
	B	C	D	E	F	G		I		K	99.1% ± 0.7%
	B	C	D	E	F	G	H			K	99.1% ± 0.7%
	B	C	D	E	F	G	H		J		99.1% ± 0.7%
	B	C	D	E	F	G	H	I			99.1% ± 0.7%
A	B			E	F	G	H		J	K	99.1% ± 0.7%
A	B			E	F	G	H	I		K	99.1% ± 0.7%
A	B	C	D	E		G	H			K	99.1% ± 0.7%
A	B	C	D	E	F	G				K	99.1% ± 0.7%
	B			E	F	G	H	I	J	K	98.9% ± 0.8%
	B	C	D		F	G	H		J	K	98.9% ± 0.8%
	B	C	D		F	G	H	I		K	98.9% ± 0.8%
	B	C	D	E		G	H	I	J		98.9% ± 0.8%
A	B			E		G	H	I	J	K	98.9% ± 0.8%
A	B			E	F	G		I	J	K	98.9% ± 0.8%
A	B		D	E		G		I	J	K	98.9% ± 0.8%
A	B		D	E		G	H		J	K	98.9% ± 0.8%
A	B		D	E		G	H	I		K	98.9% ± 0.8%

Table 9. Size 9 arrays with highest KNN(k-6) classification accuracy under classification scheme II.

Sensors											Estimated Accuracy
	B	C	D	E		G	H	I	J	K	99.1% ± 0.7%
	B	C	D	E	F	G		I	J	K	99.1% ± 0.7%
	B	C	D	E	F	G	H		J	K	99.1% ± 0.7%
	B	C	D	E	F	G	H	I		K	99.1% ± 0.7%
	B	C	D	E	F	G	H	I	J		99.1% ± 0.7%
A	B			E	F	G	H	I	J	K	99.1% ± 0.7%
A	B	C	D	E		G	H		J	K	99.1% ± 0.7%
A	B	C	D	E		G	H	I		K	99.1% ± 0.7%
A	B	C	D	E	F	G			J	K	99.1% ± 0.7%
A	B	C	D	E	F	G		I		K	99.1% ± 0.7%
A	B	C	D	E	F	G	H			K	99.1% ± 0.7%
	B	C	D		F	G	H	I	J	K	98.9% ± 0.8%
A	B		D	E		G	H	I	J	K	98.9% ± 0.8%
A	B		D	E	F	G		I	J	K	98.9% ± 0.8%
A	B		D	E	F	G	H		J	K	98.9% ± 0.8%
A	B		D	E	F	G	H	I		K	98.9% ± 0.8%
A	B		D	E	F	G	H	I	J		98.9% ± 0.8%
A	B	C		E		G	H	I	J	K	98.9% ± 0.8%
A	B	C	D		F	G	H		J	K	98.9% ± 0.8%
A	B	C	D		F	G	H	I		K	98.9% ± 0.8%

Table 10. Size 10 arrays with highest KNN($k=6$) classification accuracy under classification scheme II.

Sensors											Estimated Accuracy
	B	C	D	E	F	G	H	I	J	K	$99.1\% \pm 0.7\%$
A	B	C	D	E		G	H	I	J	K	$99.1\% \pm 0.7\%$
A	B	C	D	E	F	G		I	J	K	$99.1\% \pm 0.7\%$
A	B	C	D	E	F	G	H		J	K	$99.1\% \pm 0.7\%$
A	B	C	D	E	F	G	H	I		K	$99.1\% \pm 0.7\%$
A	B		D	E	F	G	H	I	J	K	$98.9\% \pm 0.8\%$
A	B	C	D		F	G	H	I	J	K	$98.9\% \pm 0.8\%$
A	B	C	D	E	F	G	H	I	J		$98.9\% \pm 0.8\%$
A	B	C		E	F	G	H	I	J	K	$98.4\% \pm 1.0\%$
A		C	D	E	F	G	H	I	J	K	$98.0\% \pm 1.2\%$
A	B	C	D	E	F		H	I	J	K	$97.2\% \pm 1.4\%$

Table 11. Arrays with highest KNN($k=6$) classification accuracy under classification scheme II for arrays without sensors **B** and **G**.

Size	Sensors										Estimated Accuracy	
7			C	D	E	F		H		J	K	95.5% ± 1.8%
8	A		C	D	E	F		H		J	K	95.5% ± 1.8%
9	A		C	D	E	F		H	I	J	K	95.5% ± 1.8%
8	A		C	D	E			H	I	J	K	95.2% ± 1.9%
8			C	D	E	F		H	I	J	K	95.2% ± 1.9%
7	A			D	E	F		H		J	K	95.0% ± 1.9%
8	A			D	E	F		H	I	J	K	95.0% ± 1.9%
7	A		C	D	E			H		J	K	94.8% ± 2.0%
6			C	D	E	F		H		J		94.6% ± 2.0%
7	A		C	D	E	F		H		J		94.6% ± 2.0%
7			C	D	E	F		H	I	J		94.2% ± 2.1%
7	A		C	D	E			H	I		K	94.2% ± 2.1%
8	A		C	D	E	F		H	I	J		94.2% ± 2.1%
7	A		C	D	E	F		H			K	94.0% ± 2.1%
8	A		C	D	E	F		H	I		K	94.0% ± 2.1%
6			C	D	E	F		H			K	93.8% ± 2.1%
7	A		C	D		F		H		J	K	93.8% ± 2.1%
7			C	D		F		H	I	J	K	93.8% ± 2.1%
7	A			D	E			H	I	J	K	93.8% ± 2.1%
8	A		C	D	E	F			I	J	K	93.8% ± 2.1%
6	A		C	D	E			H			K	93.5% ± 2.2%
6			C	D	E			H		J	K	93.5% ± 2.2%
7			C	D	E	F		H	I		K	93.5% ± 2.2%
6	A			D	E	F		H			K	93.3% ± 2.2%
7	A			D	E	F		H	I		K	93.3% ± 2.2%
7	A		C	D	E	F				J	K	93.3% ± 2.2%
7			C	D	E	F			I	J	K	93.3% ± 2.2%
8	A		C	D		F		H	I	J	K	93.3% ± 2.2%
6				D	E	F		H		J	K	93.1% ± 2.2%
7			C	D	E			H	I	J	K	93.1% ± 2.2%
7	A		C	D	E				I	J	K	93.1% ± 2.2%
7				D	E	F		H	I	J	K	93.1% ± 2.2%
6			C	D	E	F				J	K	92.9% ± 2.3%
7	A			D	E	F		H	I	J		92.9% ± 2.3%
6	A			D	E			H		J	K	92.7% ± 2.3%
6	A		C	D				H		J	K	92.7% ± 2.3%
6			C	D		F		H		J	K	92.7% ± 2.3%
6	A			D	E	F		H		J		92.7% ± 2.3%
7	A		C	D	E			H	I	J		92.7% ± 2.3%
7	A		C	D				H	I	J	K	92.7% ± 2.3%

Visualization of Sensor Response

The chosen 6-sensor array [BCDEGH] achieves a LOO accuracy estimate of $98.9\% \pm 0.8\%$. In order to better understand how the individual sensor responses allow the analytes to be distinguished, we have visualized the data using 2 and 3 dimensional projections. Because the full sensor data for this optimal array is 6 dimensional, there is no way for a single 2D or 3D projection to encode all of the information that the KNN algorithm is able to make use of in its classification decisions. In fact, we know from the comprehensive subset analysis that the best 3-sensor array has KNN accuracy well below the [BCDEGH] array. Thus, we cannot expect to be able to separate all analytes at once with a single 3D projection. Instead, we have chosen to partition the analytes into several sets and visualize each set individually. The 15 sensors have been partitioned as follows: {1,2}, {3,4}, {5,6,7,8}, {9,10,11,12}, and {13,14,15}. In order to find the best sensors to visually separate each analyte subset, we used the Vizrank[5] module for the Orange machine learning suite[2]. Vizrank selects a subset of sensors that leads to the best KNN classifier performance in the reduced dimensionality space. This criterion leads to projections that tend to minimize the overlap of classes.

The projections in Figures 2-7 were created with the matplotlib[4] plotting package, using sensor subsets suggested by vizrank, selected from the set {B,C,D,E,G,H}. The plots show all the measured data points for each analyte as circles, and the mean value for each concentration as a square. The mean values are connected to form a concentration-parametrized curve. Some of the subsets (e.g {1,2}) separate easily in low dimensions, but others (e.g. {3,4}) are hard to visualize with a single snapshot, and only really start to separate in higher dimensional spaces. We needed a second image (Figure 4) for {3,4}, showing a different angle of the same plot, in order to show how the two analytes are separated in this 3D space.

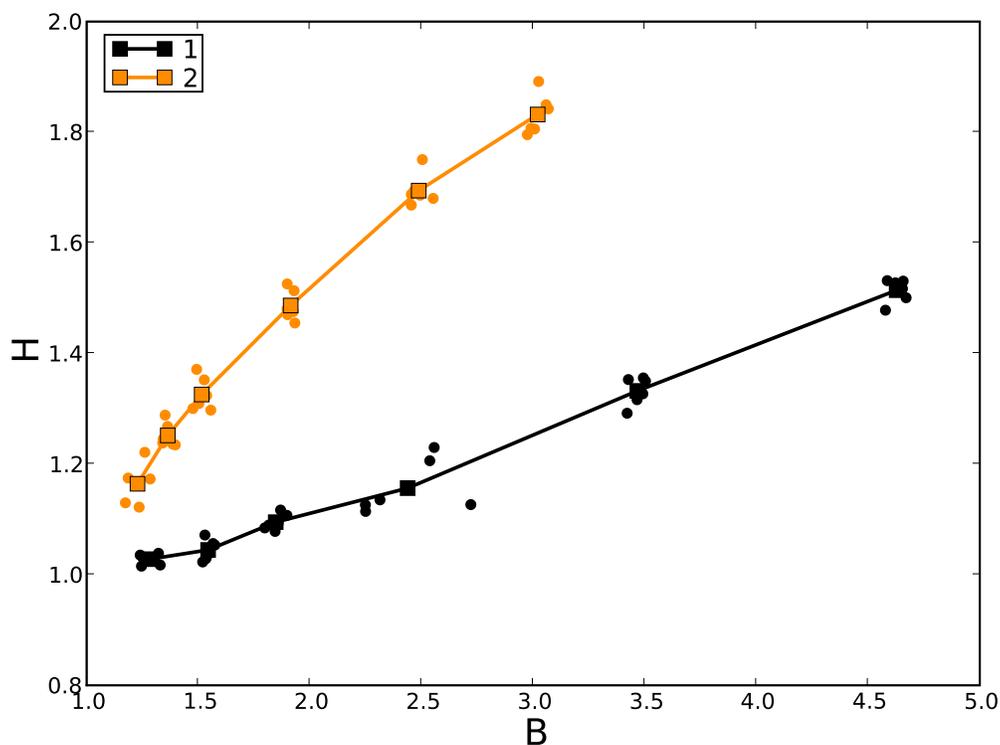


Figure 2: Visualization of analytes 1 and 2.

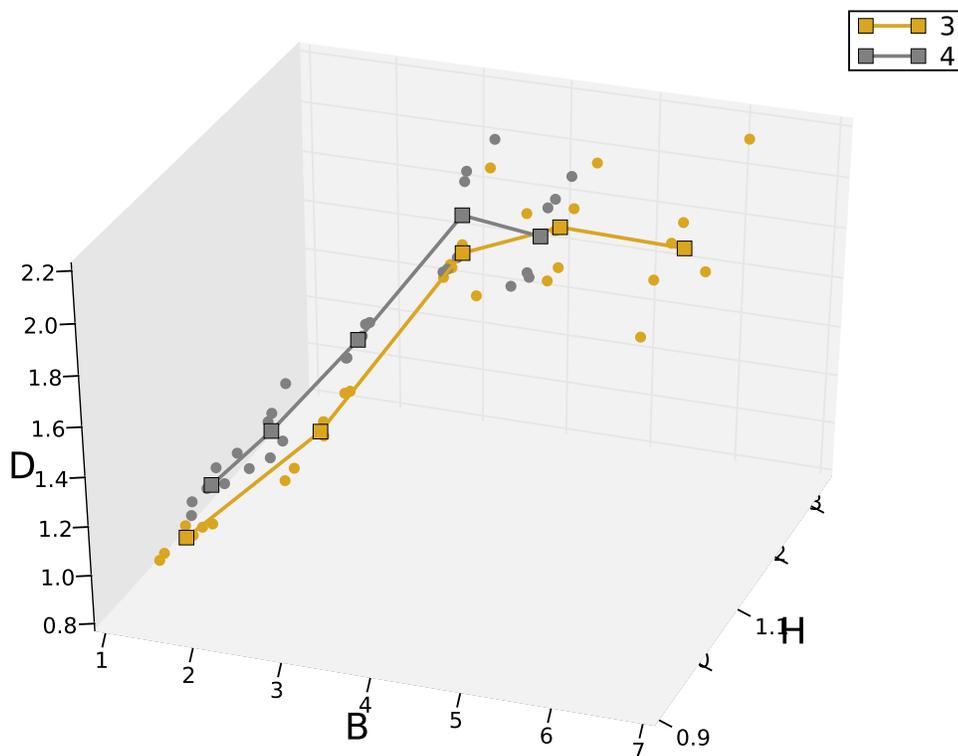


Figure 3: Visualization of analytes 3 and 4.

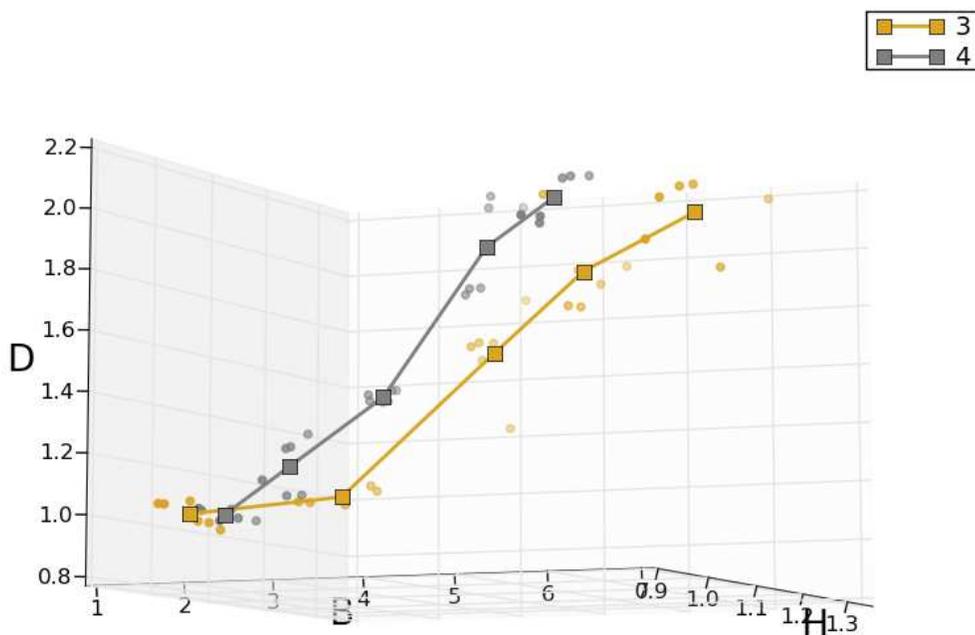


Figure 4: A second view of the same plot for 3 and 4. The curves are separated in this space, but it is difficult to see with a single 2D projection. This shows the higher concentration values separated.

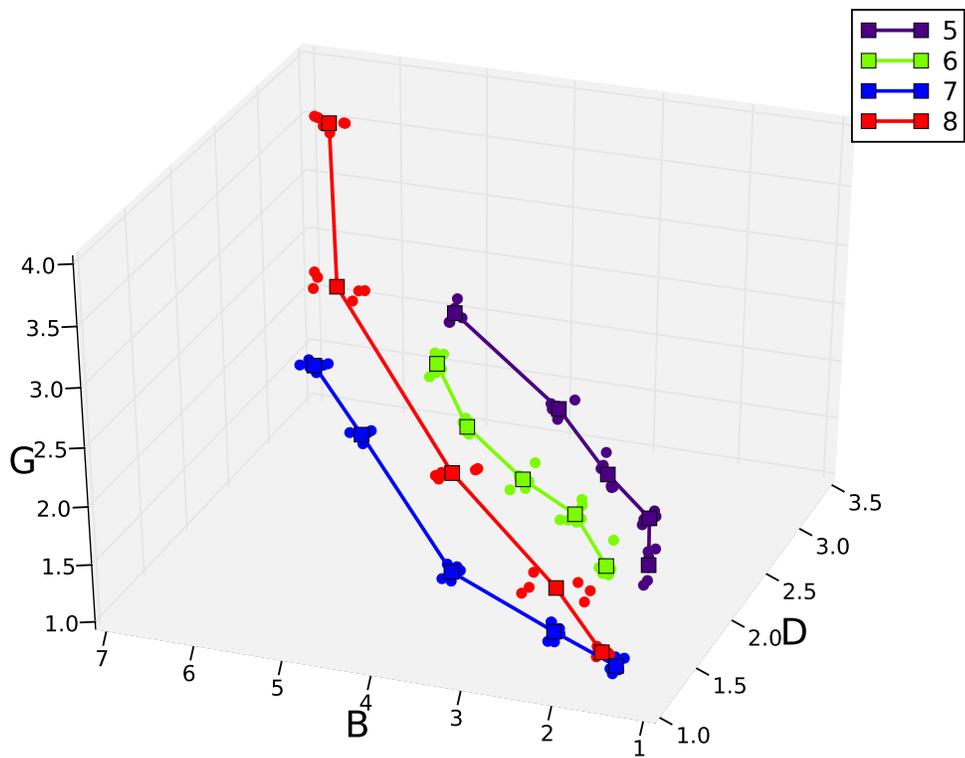


Figure 5: Visualization of analytes 5,6,7, and 8.

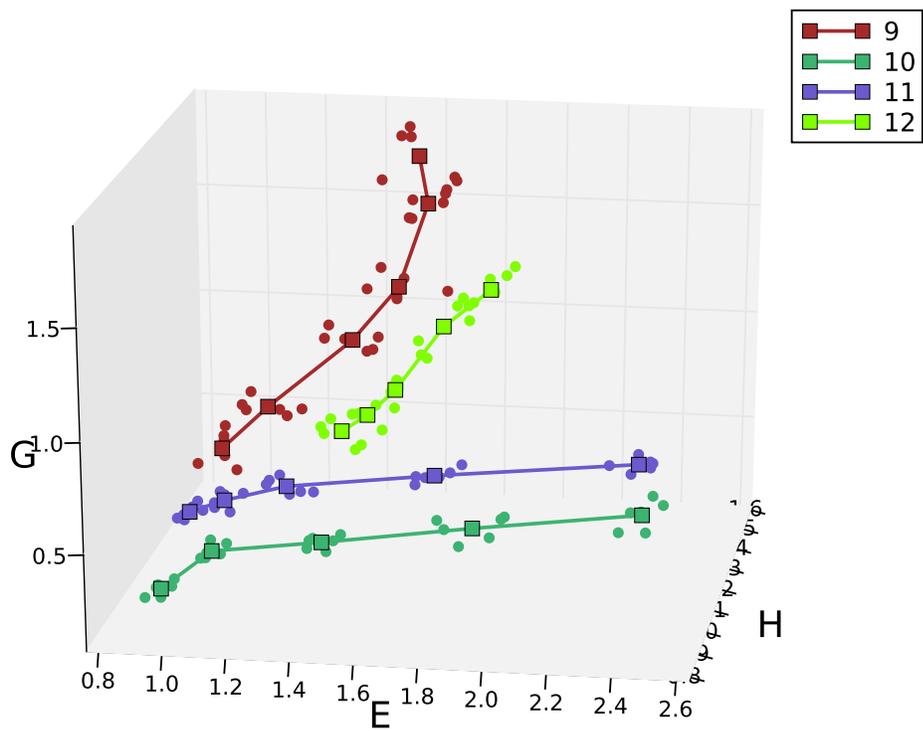


Figure 6: Visualization of analytes 9,10,11, and 12.

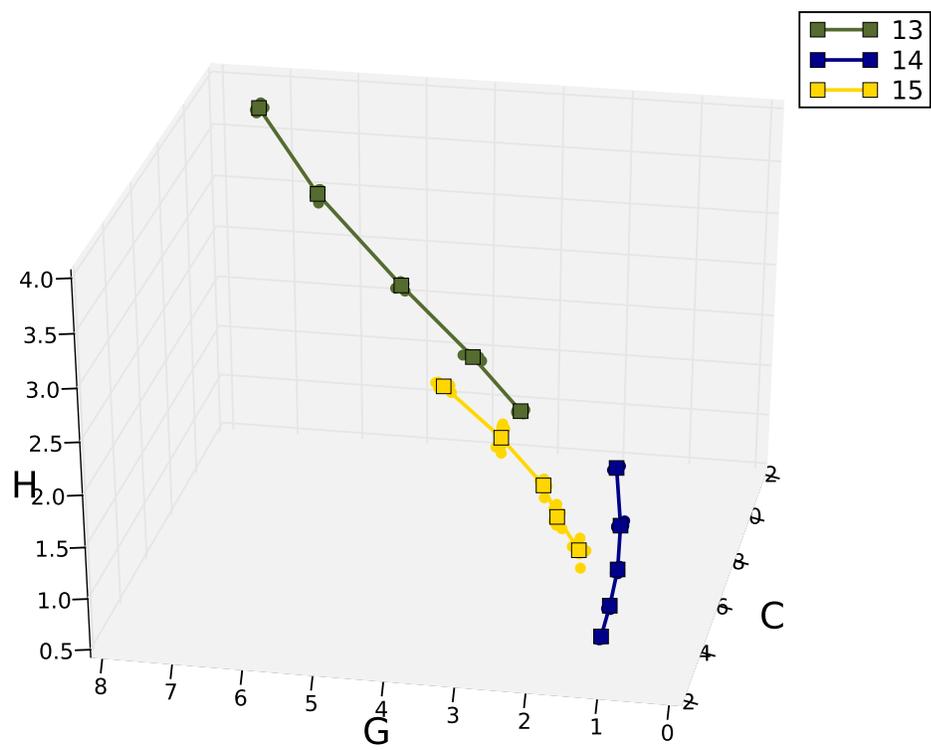


Figure 7: Visualization of analytes 13,14, and 15.

Analysis of Three-sensor Array [B,D,G]

As reported in Table 3, the three-sensor array [B,D,G], was found to have a $94.40\% \pm 2.03\%$ KNN($k=6$) classification accuracy as estimated by the LOO method using classification scheme II. We decided to analyze this array further to understand how such a small array was able to perform so well. In Table 12 we have summarized the LOO errors of the [B,D,G] array as a confusion matrix.

In order to visually understand how the classification errors coorespond to the actual overlap of sensor responses, we have provided several 3D plots of the [B,D,G] sensor response space. In Figures 8, 9, 10, and 11, we plot selected subsets of the sensors which cover the 15-sensor space. The concentration curves in these subsets have minimal overlap, and thus this 3-sensor array works well distinguishing these subsets. Also, looking at the confusion matrix in Table 12, several other larger subsets would separate well with these sensors.

The major errors in the [B,D,G] array are localized in two subsets of sensors. First, sensor subset [3, 4, 6, 7, 8, 9] is visualized in Figure 12. This set represents 12 of the 25 total errors for the [B,D,G] array. Analyte 4 was misclassified on 6 separate instances as one of the other analytes in this subset. Additionally, there were 6 errors between analytes 7, 8, and 9. Furthermore, another 11 of the 25 errors occur between analytes in the subset [2, 11, 12, 14], and are visualized in Figure 13.

Finally, all 15 analytes are plotted simultaneously in Figure 14. Because some of the analytes have a very high response, it is hard to see the detail of the separation of the less responsive analytes, but as discussed, almost all of the errors are accounted for in the subsets of Figure 12 and Figure 13.

Table 12. Confusion matrix showing LOO classification errors for the 3-sensor array [B,D,G], using classification method KNN ($k=6$) and classification scheme II.

	Assigned Class														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	34	0	0	0	2	0	0	0	0	0	0	0	0	0	0
2	0	36	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	30	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	2	24	0	2	1	0	1	0	0	0	0	0	0
5	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	1	26	3	0	0	0	0	0	0
9	0	0	0	0	0	0	0	2	34	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	28	1	0	1	0
12	0	1	0	0	0	0	0	0	0	0	0	25	0	4	0
13	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0
14	0	2	0	0	0	0	0	0	0	0	2	0	0	26	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	30

Total Errors: **25/468**
Accuracy Rate (95% conf.): **94.40% ± 2.03%**

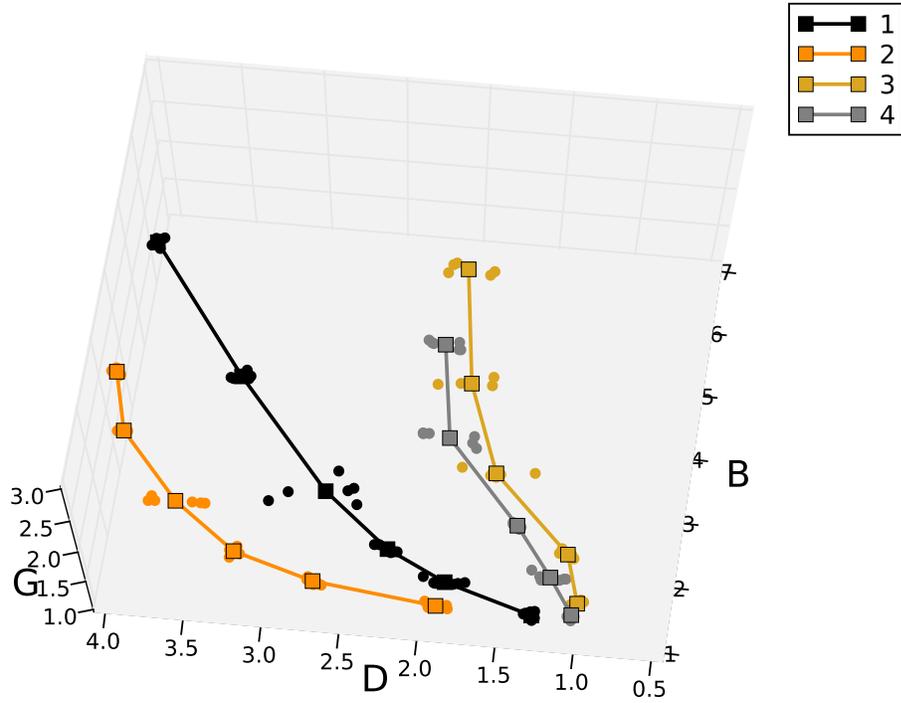


Figure 8: Visualization of analytes 1, 2, 3 and 4 with array [B,D,G]. There is some overlap between 3 and 4 at low concentrations, as is reported by the confusion matrix.

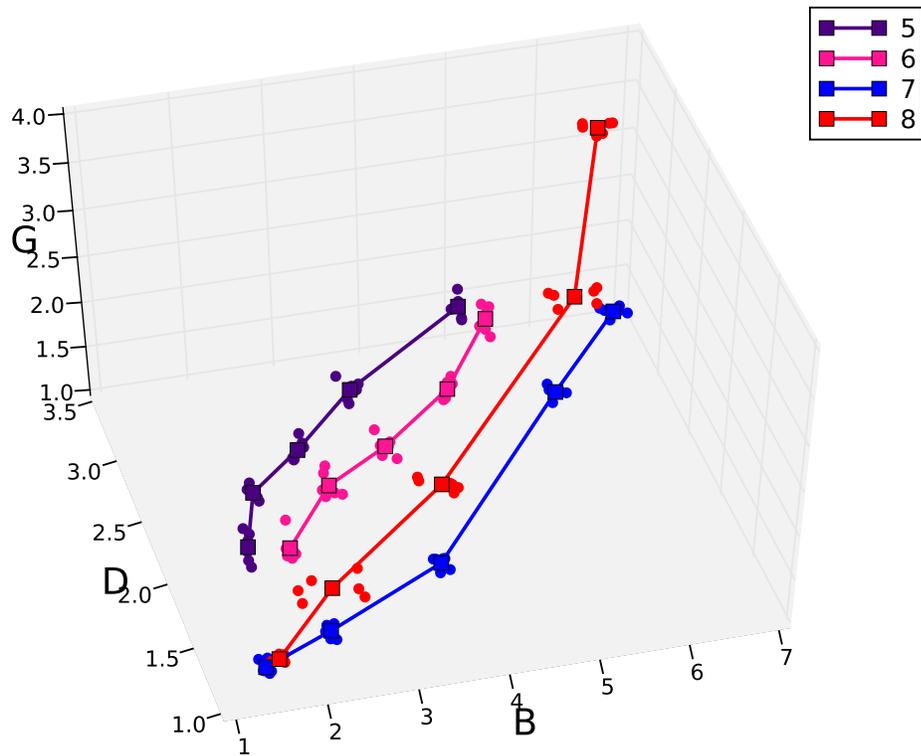


Figure 9: Visualization of analytes 5, 6, 7 and 8 with array [B,D,G]. There is some overlap between 7 and 8, as is reported by the confusion matrix.

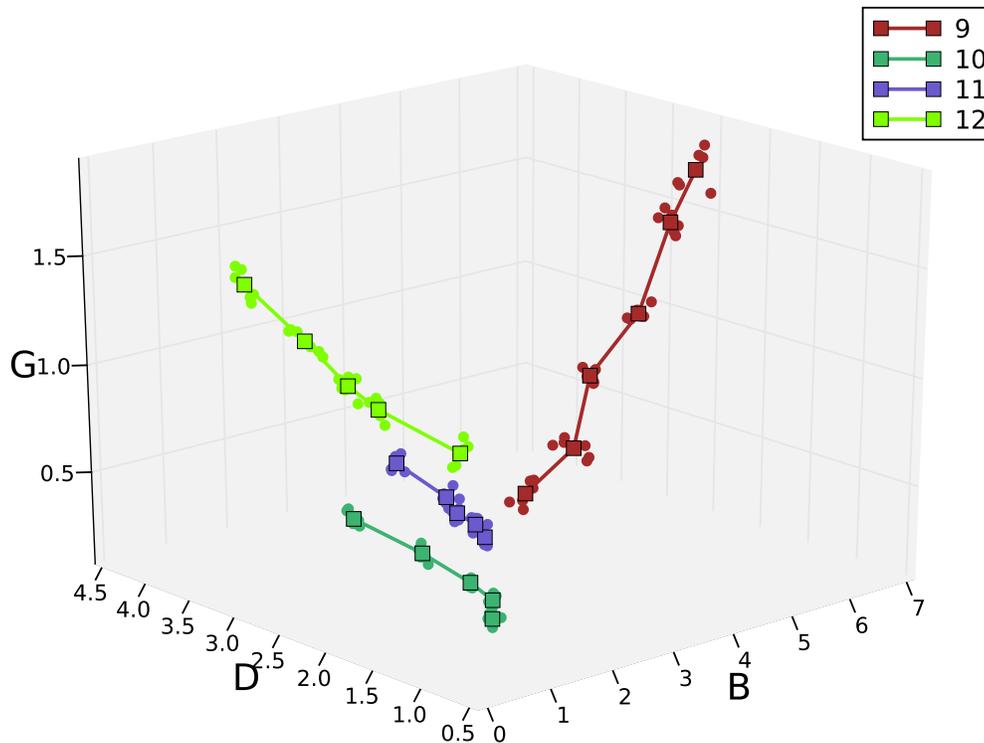


Figure 10: Visualization of analytes 9, 10, 11 and 12 with array [B,D,G]. There is some overlap between 11 and 12, as is reported by the confusion matrix.

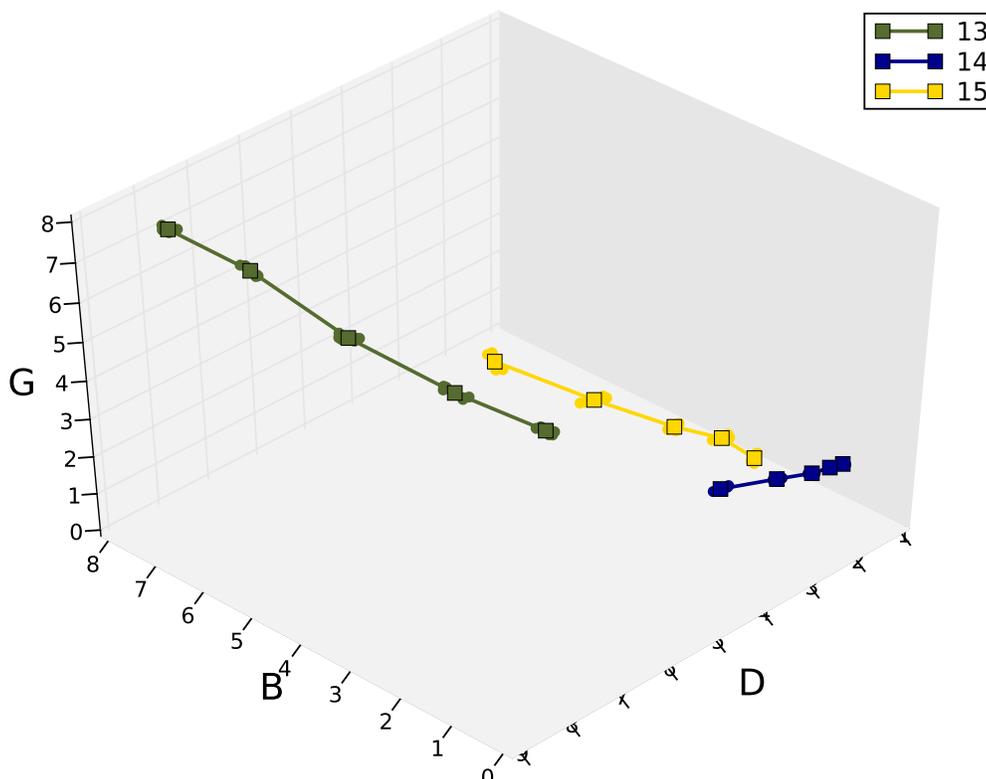


Figure 11: Visualization of analytes 13, 14 and 15 with array [B,D,G]. There not much overlap between these analytes, as is reported by the confusion matrix.

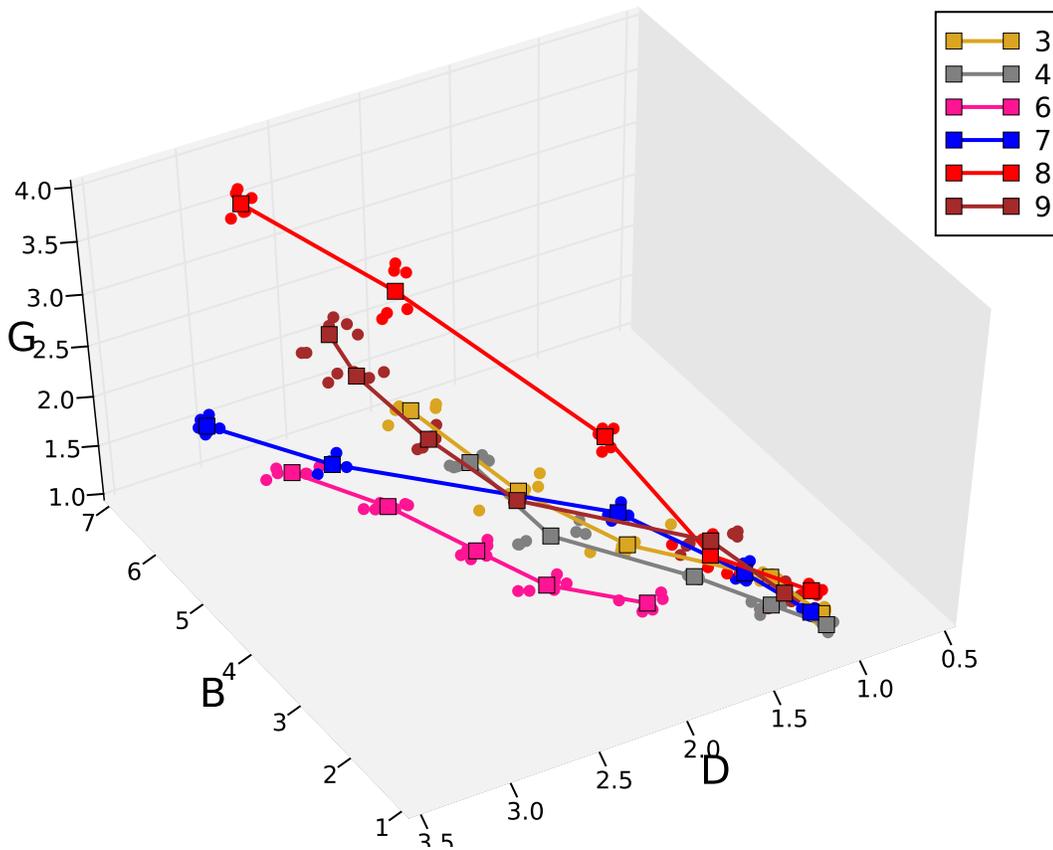


Figure 12: Visualization of analytes 3, 4, 6, 7, 8 and 9 with array [B,D,G]. According to the confusion matrix, this is one of the most overlapping sets of analytes. The KNN classifier classifies some 6 samples of analyte 4 as one of the other analytes. Also there are 6 other errors between 7, 8 and 9. This set represents 12 of of the 25 total errors of the [B,D,G] sensor array.

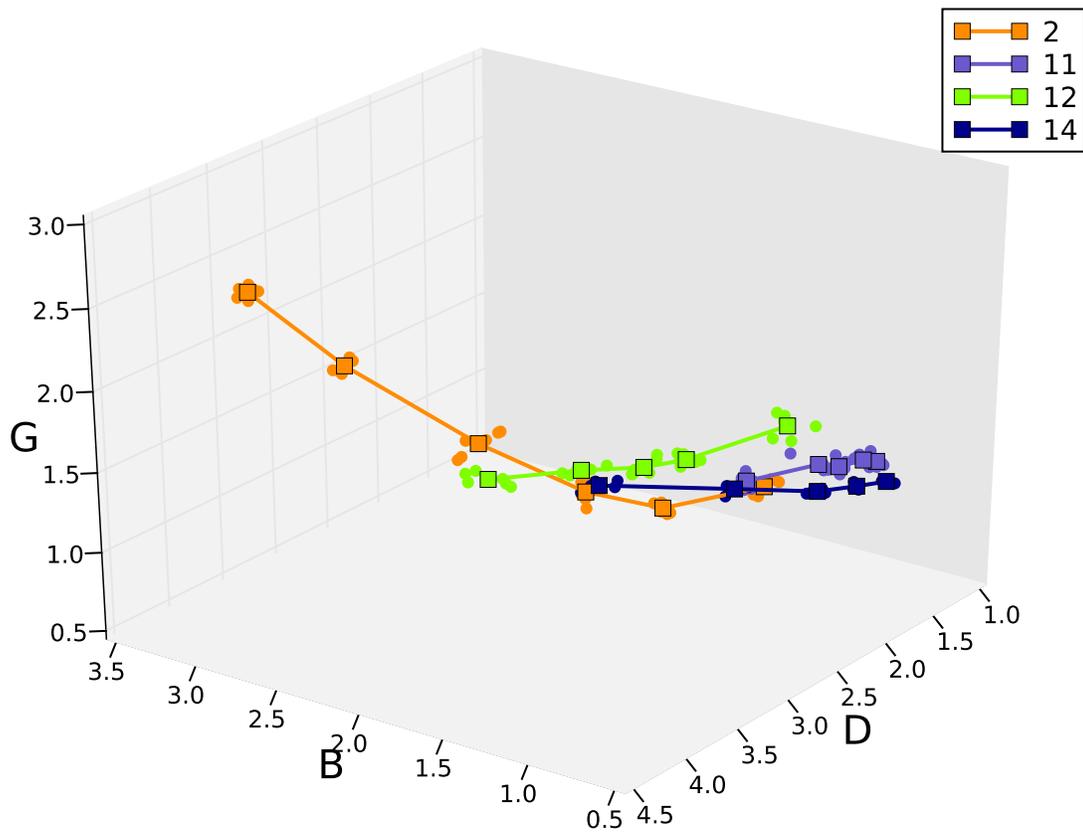


Figure 13: Visualization of analytes **2**, **11**, **12** and **14** with array $[B,D,G]$. According to the confusion matrix, this is also a very overlapping sets of analytes. This set represents 11 of the 25 total errors of the $[B,D,G]$ sensor array.

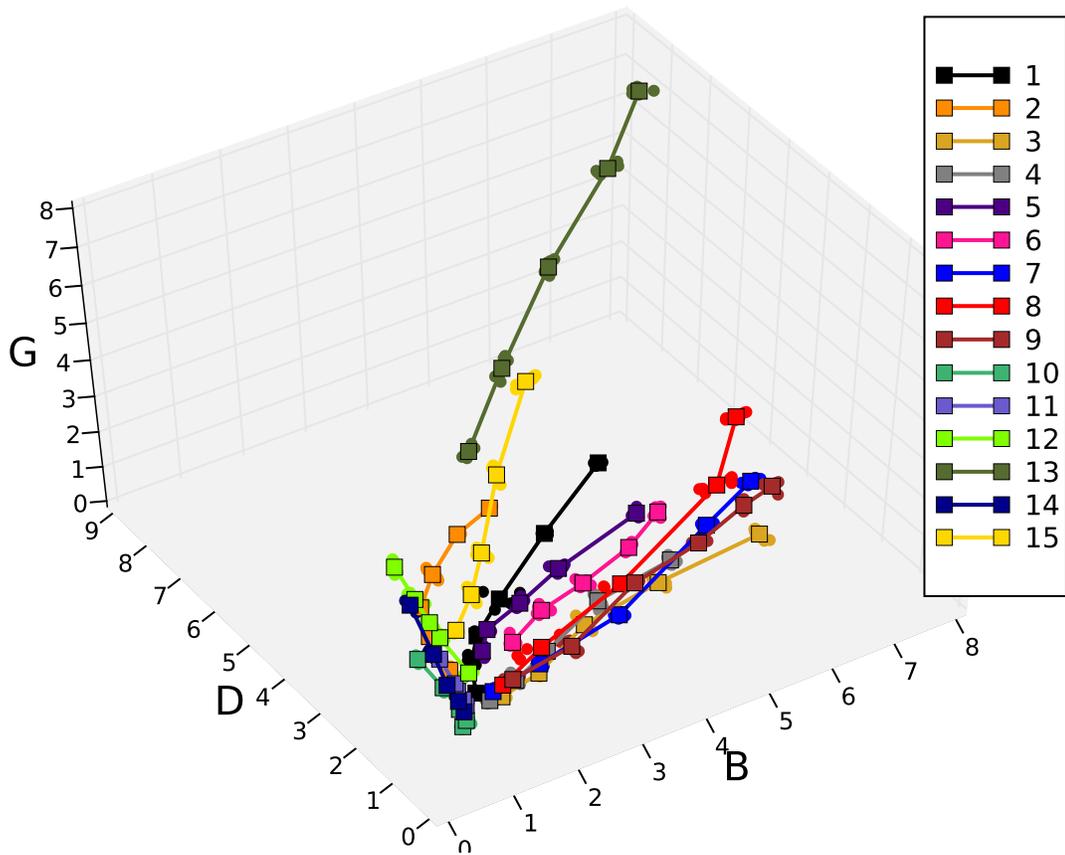


Figure 14: Visualization of all 15 analytes with array [B,D,G]. Because of the different scales of some of the analytes, the separation can be hard to see at this level of detail.

References

- [1] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [2] J. Demsar and B. Zupan. Orange: From experimental machine learning to interactive data mining. Technical report, Faculty of Computer and Information Science, University of Ljubljana, 2004.
- [3] Richard O. Duda, Peter E. Hart, and David G. Stork. *Pattern Classification*. John Wiley and Sons, 2001.
- [4] John D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science and Engineering*, 9(3):90–95, 2007.
- [5] Gregor Leban, Blaž Zupan, Gaj Vidmar, and Ivan Bratko. Vizrank: Data visualization guided by machine learning. *Data Min. Knowl. Discov.*, 13(2):119–136, 2006.
- [6] J. Kent Martin and D. S. Hirschberg. Small sample statistics for classification error rates II: Confidence intervals and significance tests. Technical report, University of California, Irvine, July 1996.
- [7] J. Ross Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, 1993.
- [8] Bernhard Schölkopf and Alexander J. Smola. *Learning with Kernels*. MIT Press, 2002.

Materials and Methods

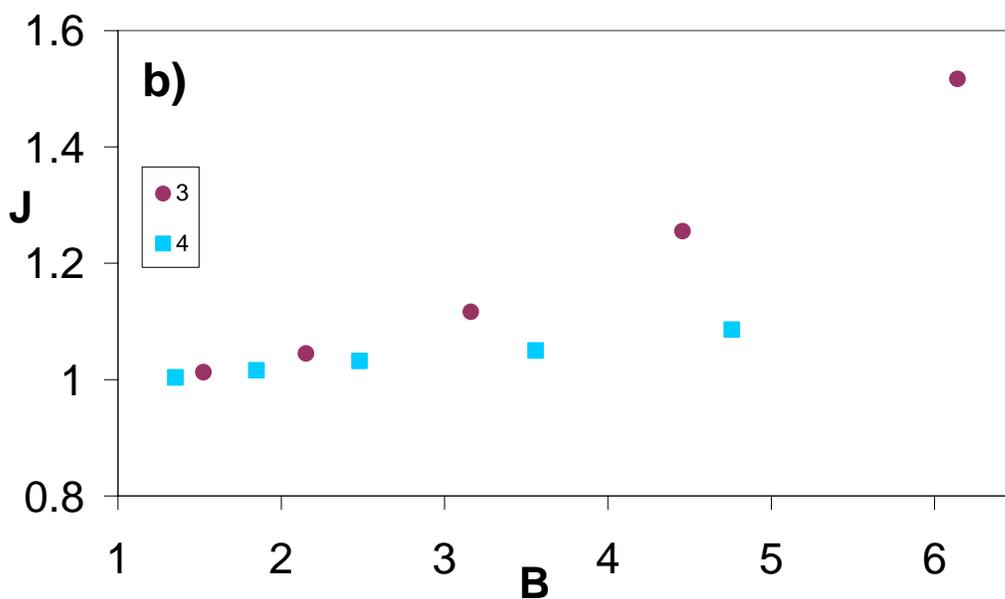
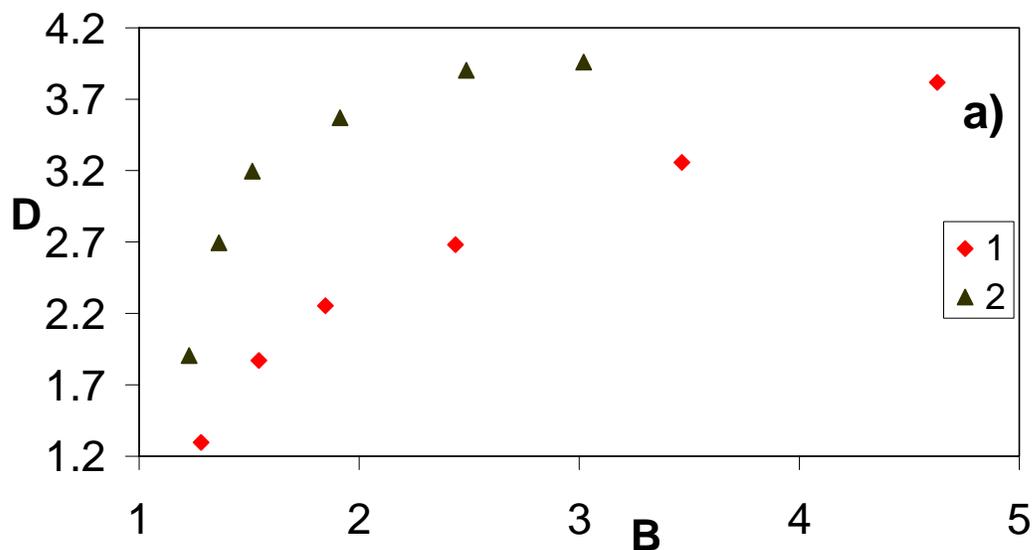
Materials. Three-way junction sensors **A, B, C, D, E, F, G, H, I, J, K** were made and HPLC purified by Integrated DNA Technologies, Inc (Coralville, IA, USA) and were used as received. DEPC-treated and nuclease-free water was purchased from Fisher Scientific (Fair Lawn, NJ, USA) and used for all buffers. The sequences are as following: **A**, 5'-fluorescein-ATC TCG GGA CGA CAG GAT TTT CCT CCA TGA AGT GGG TCG TCC C. **B**, 5'-fluorescein-ATC TCG GGA CGA CAG GAT TTT CCT CAA TGA AGT GGG TCG TCC C. **C**, 5'-fluorescein-ATC TCG GGA CGA CAG GAT TTT CCT CCA TGA AGT GG(NI)G TCG TCC C; NI-nitroindole base. **D**, 5'-fluorescein-ATC TCG GGA CGA C(NI)AG GAT TTT CCT CCA CGA AGT GG(NI)G TCG TCC C. **E**, 5'-fluorescein-ATC TCG GGA CGA C(NI)AG GAT TTT CCT (NI)CCA TGA AGT GG(NI)G TCG TCC C. **F**, 5'-fluorescein-ATC TCG GGA CGA CAG GAT TTT CCT CCA CGA AGT G(NI)G TCG TCC C. **G**, 5'-fluorescein-ATC TCG GGA CGA CAG GAT TTT CCT CAA TGA AGT GG(NI)G TCG TCC C. **H**, 5'-fluorescein-ATC TCG GGA CGA C(NI)AG GAT TTT CCT CAA TGA AGT GG(NI)G TCG TCC C. **I**, 5'-fluorescein-ATC TCG GGA CGA C(NI)AG GAT TTT CCT (NI)CAA TGA AGT GG(NI)G TCG TCC C. **J**, 5'-fluorescein-ATC TCG GGA CGA C(NI)AG GAT TTT CCT CAA TGA AGT GGG TCG TCC C. **K**, 5'-fluorescein-ATC TCG GGA CGA CAG GAT TTT CCT C(NI)A TGA AGT GGG TCG TCC C. Complementary quencher strand as: 5'-GTC GTC CCG AGA T-dabcyl. (-)-cocaine and (+)-cocaine were obtained through the National Institute of Drug Abuse. All other alkaloids and steroids were purchased from Sigma-Aldrich Co. (St. Louis, MO, USA).

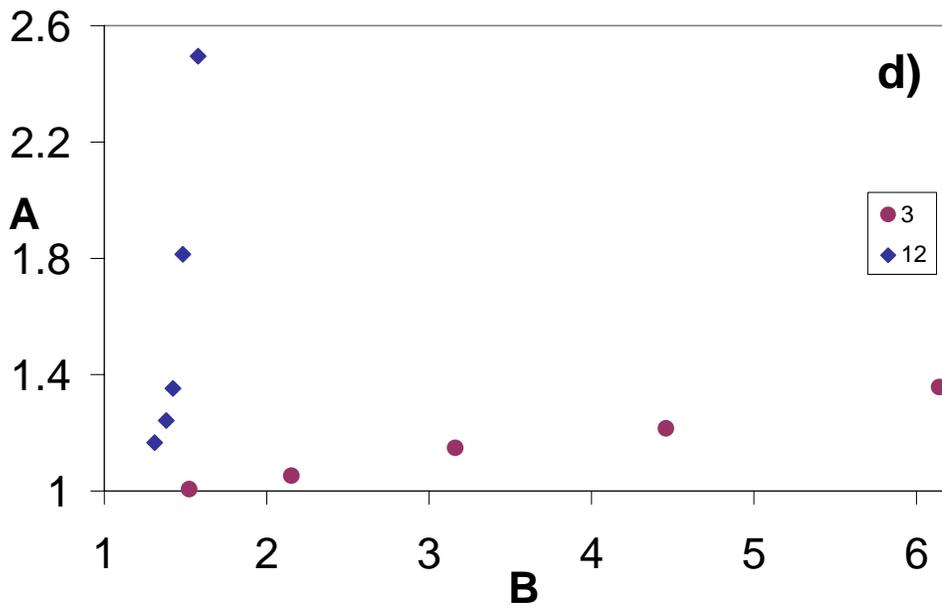
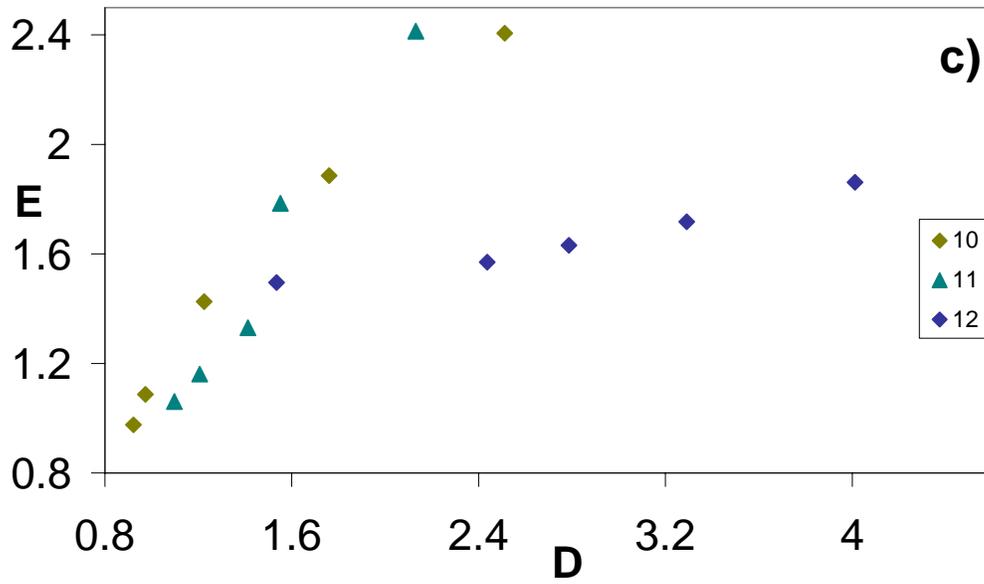
Instruments. The measurements were performed on a Perkin-Elmer Victor II microplate reader (Shelton, CT, USA) with a 485-nm excitation filter and a 535-nm emission filter. 384 well non-binding surface, flat bottom, black polystyrene assay plates (Corning, NY, USA) were used.

Measurements. All measurements were performed in the binding buffer containing 20 mM Tris-HCl pH7.4, 140 mM NaCl, 5 mM KCl, 2 mM MgCl₂. Mixture of sensors and quencher strand was incubated 5 min at room temperature, then a series of standard dilutions of all compounds (the stock solution of each compound was adjusted to pH7.4) were added to the mixture solutions to a final concentrations of 50 nM sensors and 150 nM (for sensors A-E, G, K), 250nM (F, H), 500nM (for sensors I, J) quencher strand. Measurements were performed after 30 min. The background fluorescence signal, F_0 , for each batch of measurements was taken as the average of 0.0 concentration fluorescence readings (without analytes) for each of those measurements. Then, each fluorescence measurement was normalized by dividing by the responding F_0 for that analyte at that concentration.

2D plots with data (-fold increase) from two sensors.

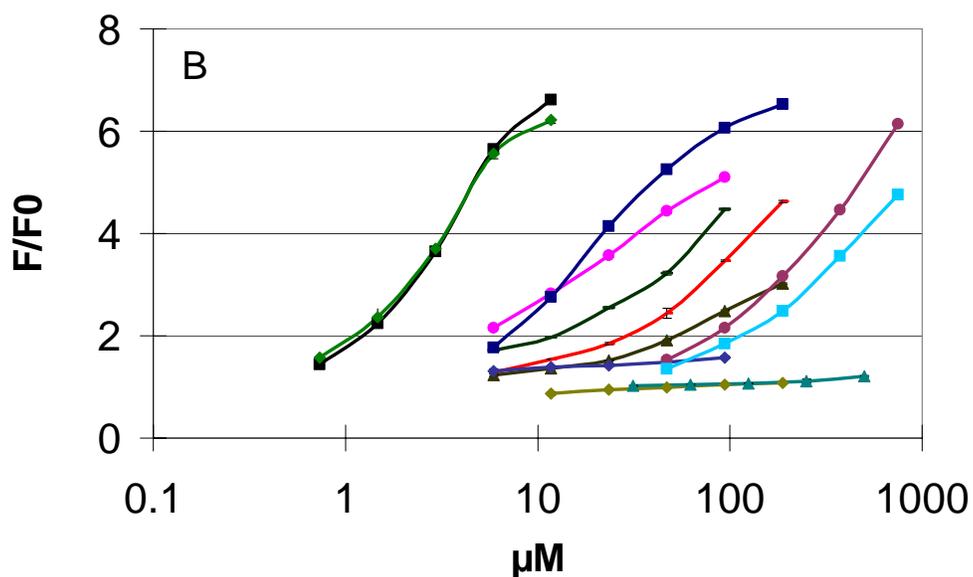
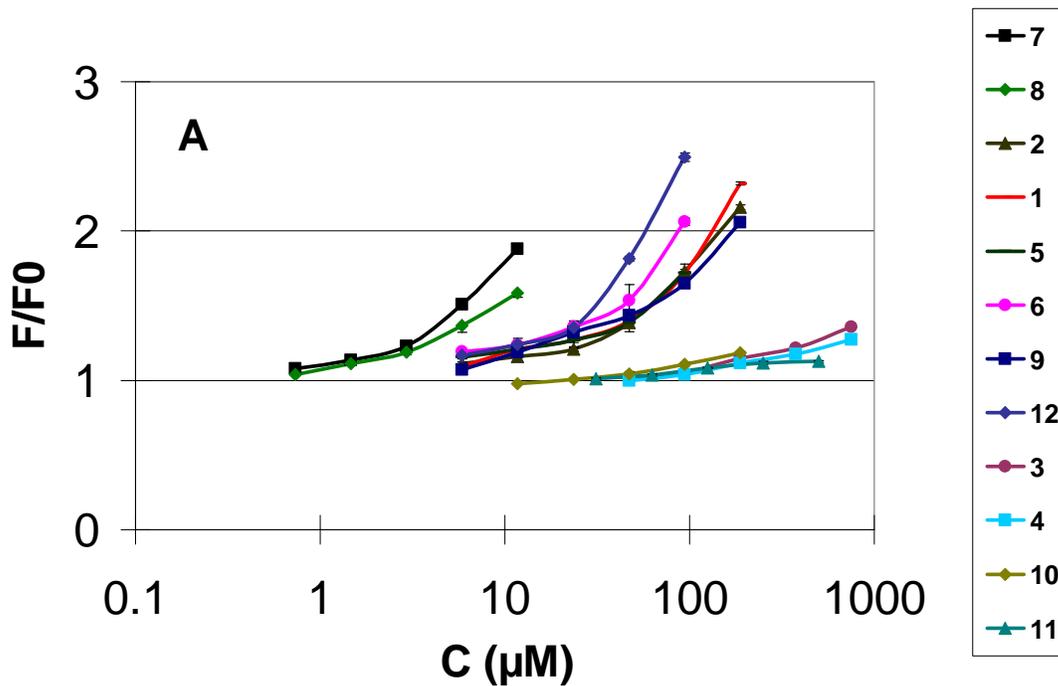
a) sensors B and D separate well strychnine (1) and brucine (2), D has two NI groups; b) sensors B and J distinguish (-)-cocaine (3) from (+)-cocaine (4); c) sensors D and E separate vindoline (12) from vinblastine (10) and vincristine (11); d) separation of (-)-cocaine (3) and vindoline (12) is based on weaker interactions with one particular subtype of junctions.

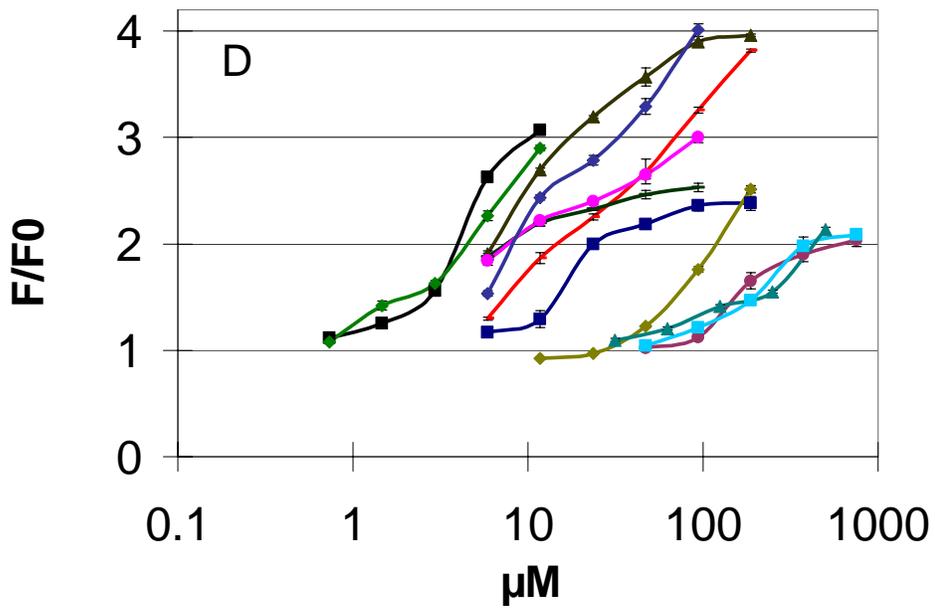
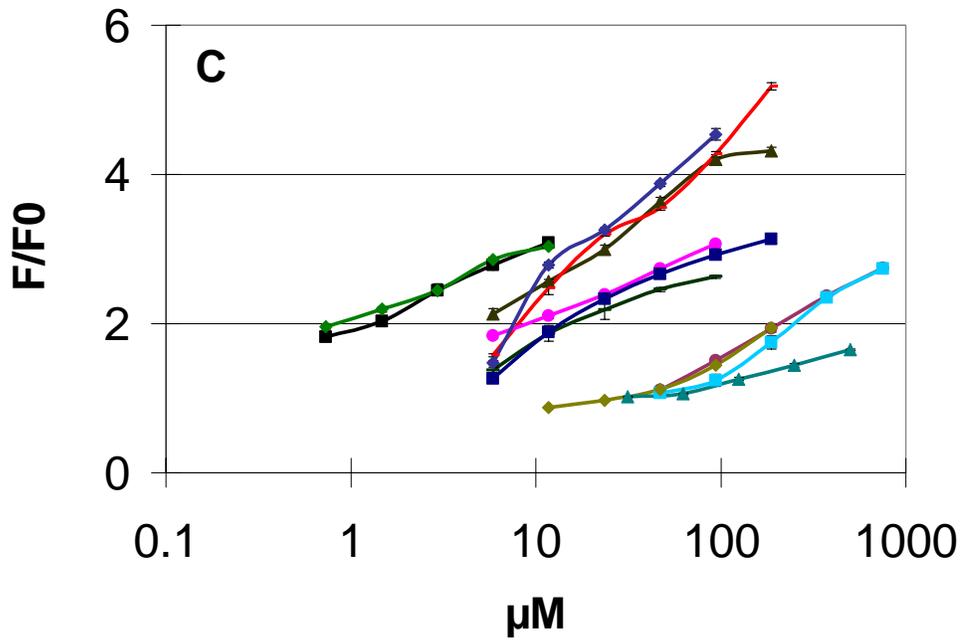


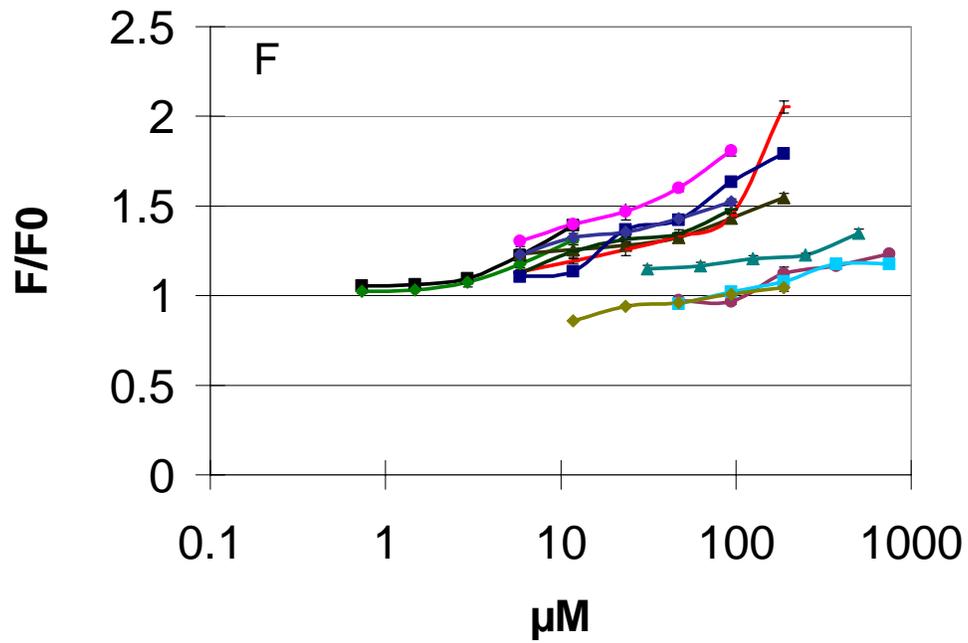
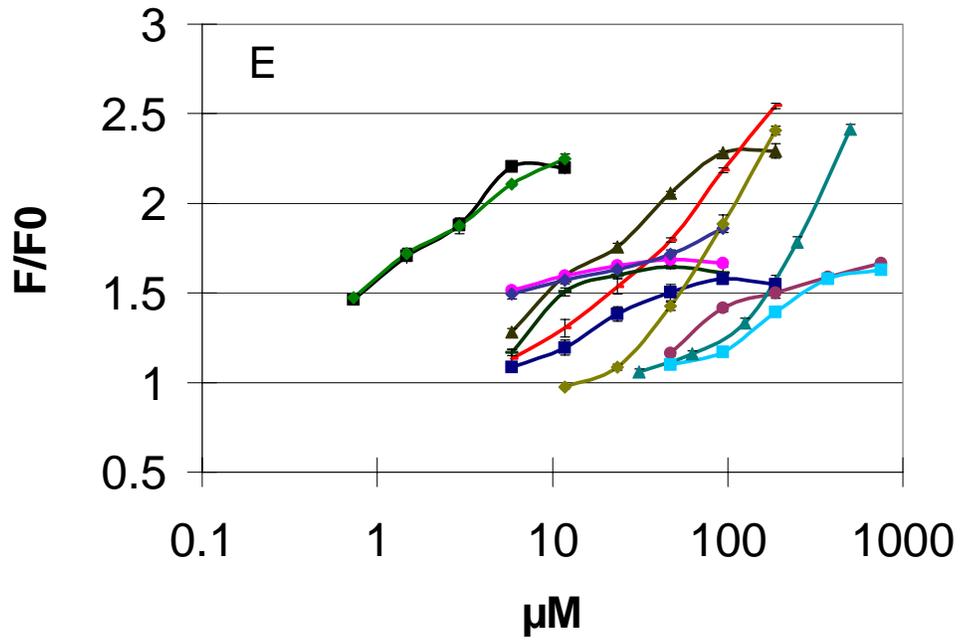


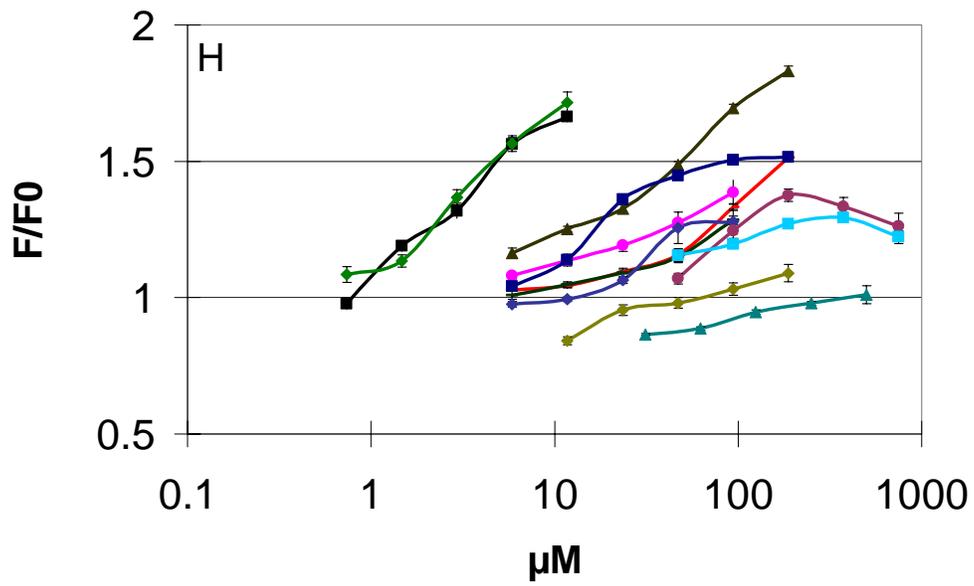
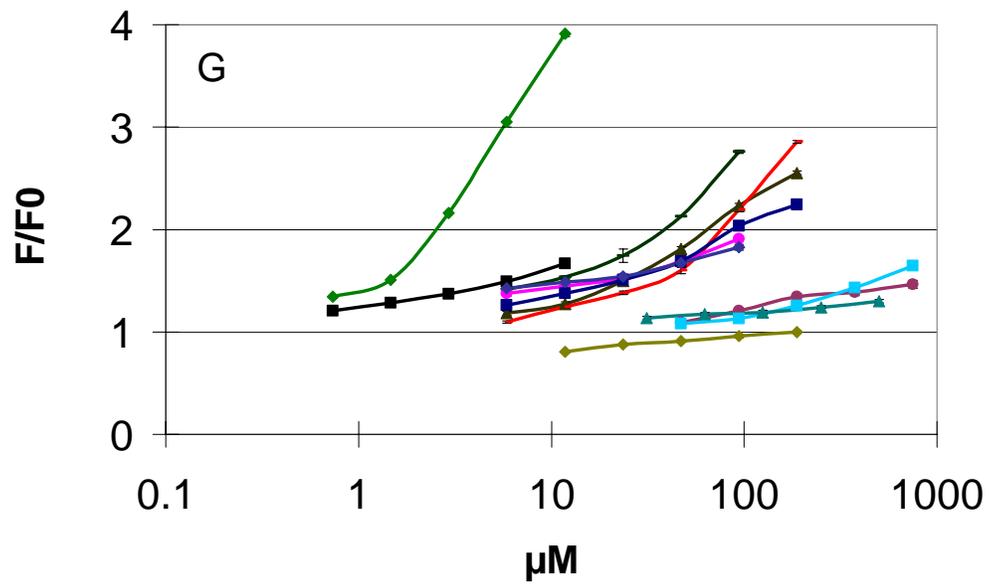
Response of 11 different sensors (A-K) to 12 alkaloids.

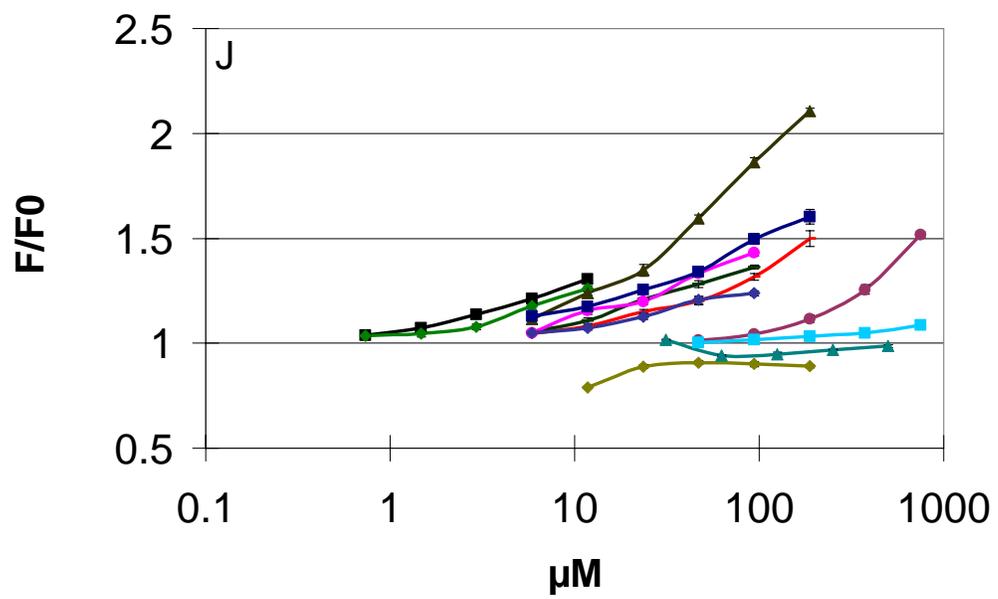
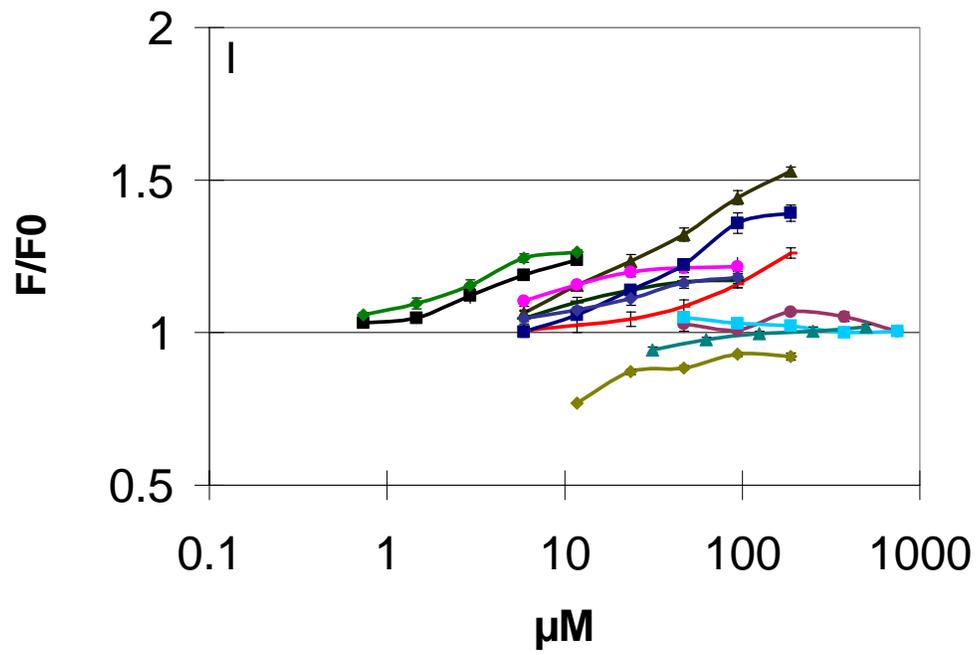
(1. strychnine: red bar; 2. brucine: olive green triangle; 3. (-) cocaine: plum circle; 4. (+) cocaine: sky blue square; 5. cinchonine: dark green bar; 6. cinchonidine: pink circle; 7. quinine: black square; 8. quinidine: green diamond; 9. methylergonovine: dark blue square; 10. vinblastine: dark yellow diamond; 11. vincristine: teal triangle; 12. vindoline: indigo diamond.)

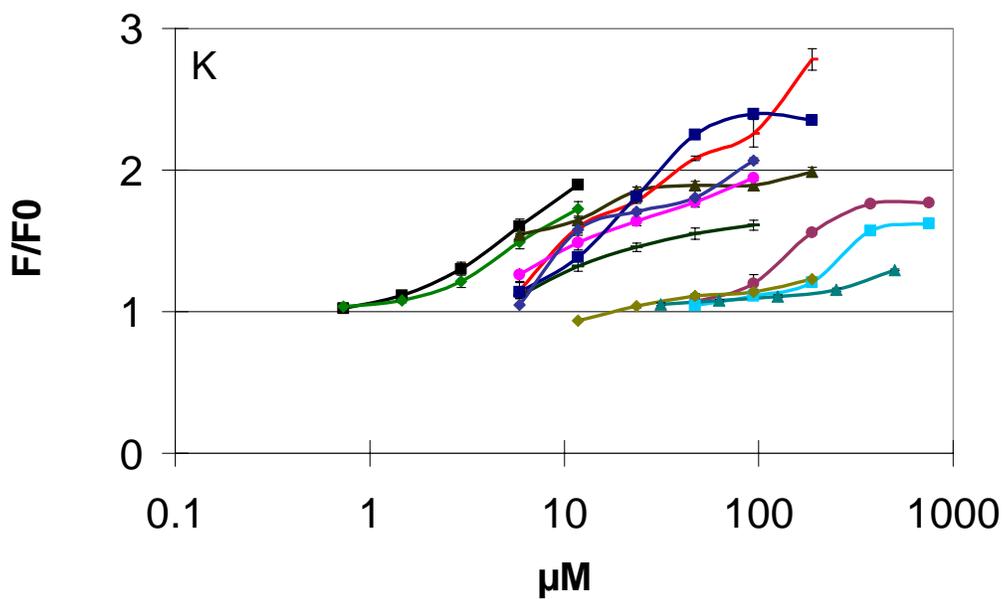












Response of 11 different sensors (A-K) to 3 steroids.

(13. deoxycholic acid: violet diamond; 14. deoxycorticosterone-21 glucoside: dark red triangle; 15. dehydroisoandrosterone-3-sulfate: bright green circle.)

