Supplementary Information

Pattern-Based Sensing of Nucleotides in Aqueous Solution with a Multicomponent Indicator Displacement Assay

Andrey Buryak,[a] Alexei Pozdnoukhov[b] and Kay Severin*[a]

[a]Institut des Sciences et Ingénierie Chimiques, École Polytechnique Fédérale de Lausanne (EPFL), [b] Institute of Geomatics and Analysis of Risk, University of Lausanne (UNIL), 1015 Lausanne, Switzerland

General: Phosphate buffer (50 mM, pH 7.4) prepared with bidistilled water was used for all experiments. Stock solutions of the dyes (0.5 mM), $[Cp^*RhCl_2]_2$ (0.5 mM Rh), cAMP, AMP, ADP, ATP, GTP, PPi and PPi/cAMP (5.0 mM) were prepared with the buffer. The dyes Gallocyanine (Acros), Mordant Yellow 10 (Aldrich), Evans Blue (Aldrich) and the nucleotides cAMP (Fluka), AMP disodium salt (Fluka), ADP disodium salt (Fluka), ATP disodium salt (Aldrich), GTP disodium salt (Fluka) and sodium pyrophosphate decahydrate (Acros) were used as received. All spectra were measured after 1 h equilibration time with a *Lambda 40* spectrometer (Perkin Elmer).

UV/Vis Titrations. To reveal the spectroscopic changes generated upon coordination of the dyes **2**, **3**, and **4** to the Rh complex **1**, the following experiment was performed: the solutions of the respective dye (final concentrations: 40 μM for **2** and **3**, 20 μM for **4**) and **1** (final Rh concentration: 40 μM) were mixed in a cuvette and the UV-Vis spectrum was measured after an equilibration time of 1 hour. The same experiment was performed without complex **1**. The difference between the spectrum of the pure dye and the dye-Rh complex was calculated. To determine the association constant between the Rh complex and Mordant Yellow 10 (**2**), UV/Vis titration experiments were performed. The titration data were fitted to a 1:1 binding algorithm using the program DATAN 3.1, trial version (www.multid.se). The metal binding constant for Mordant Yellow 10 (**2**) was determind to be K = 3.7 (± 0.3) x 10^4 M^{-1}. Experimental conditions: [**2**] = 25 μM, [**1**] = 0 – 25 μM, 50 mM phosphate buffer, pH 7.4.

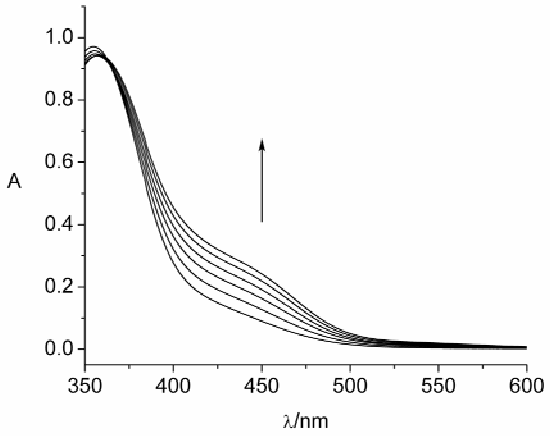

Figure S1. UV-Vis spectra of solutions of **2** (40 μM) upon addition of a variable amount of complex **1** (final Rh concentration: 0, 10, 20, 30, 40, and 50 μM). The spectra were recorded in H_2O (50 mM phosphate buffer, pH 7.4) after equilibration.

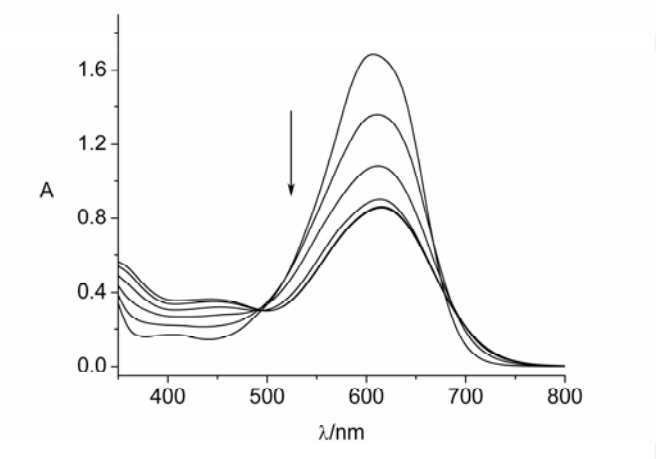

Figure S2. UV-Vis spectra of solutions of **4** (20 μM) upon addition of a variable amount of complex **1** (final Rh concentration: 0, 10, 20, 30, 40, and 50 μM). The spectra were recorded in H_2O (50 mM phosphate buffer, pH 7.4) after equilibration.

Multicomponent indicator displacement assay with ADP, GTP and ATP. The respective analyte (final concentration: 0.5 mM) was mixed with the solution of the sensor (final concentrations: [Mordant Yellow 10] = [Gallocyanine] = 40 μM, [Evans Blue] = 20 μM, [Rh] = 120 μM). A blank experiment was performed with no nucleotide being added. After the equilibration time of 1 hour the UV-Vis spectrum in the region 350-800 nm was recorded. The difference between the blank measurement and the respective analytes was calculated.

The discrimination of cAMP, AMP, ADP, ATP, GTP and PPi with a multicomponent indicator displacement assay. The respective analyte (final concentration: 1.0 mM) was mixed with the solution of the sensor (final concentrations: [Mordant Yellow 10] = [Gallocyanine] = 40 μM, [Evans Blue] = 20 μM, [Rh] = 120 μM). Six independent measurements were performed for each analyte. The intensities at five wavelengths (700, 625, 600, 540 and 440 nm) were chosen as the input variables for the linear discriminant analysis. These wavelengths correspond approximately to the maxima in the difference spectra for the dyes **2** – **4**. The analysis was performed with the commercial statistics program SYSTAT (version 11.0). A 100 % classification accuracy was achieved.

Atrificial neural network multi-layer perception analysis. Artificial neural networks (ANNs), especially the multi-layer perceptrons (MLP), are the main competitors to the state-of-the-art partial least squares (PLS) methods in terms of flexibility of the model and robustness of the results. In a number of applications it was observed that MLP outperforms PLS on non-linear problems.[1] However, it requires an input feature selection procedure. In the present study, the input features (wavelengths) were selected manually according to the maximum variation of the intensities.

MLP is one of the most popular models of ANNs. The main component of a MLP is the neuron, a unit which sums the inputs and performs a transformation via an activation function, which is responsible for non-linearity. In our case, the hyper-tangent transfer function was used. Neurons form a structure with some inputs representing exploratory variable, variable number of hidden layers including the activation function, and the outputs corresponding to the target variable(s). The particular MLP structure used for our analysis contained five input neurons (the number of selected input wavelengths), five hidden neurons arranged in one layer, and two output neuron representing the target variables values - concentrations. The structure of the MLP is presented in Figure S3.

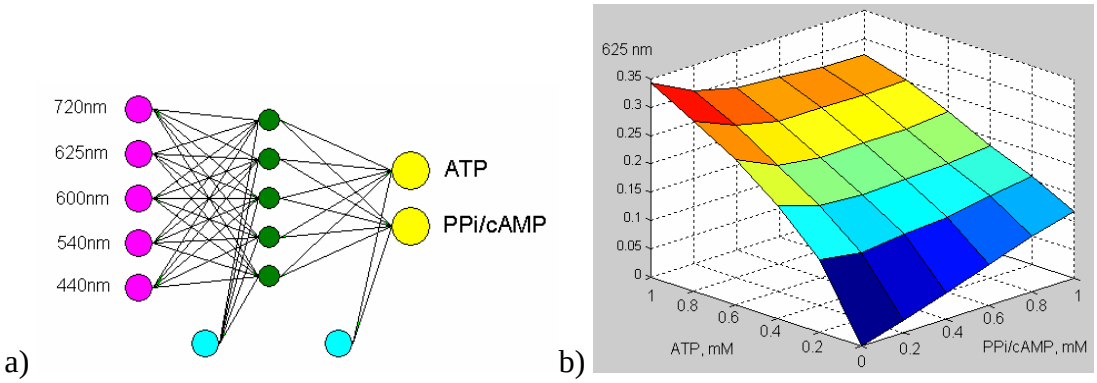

Figure S3. a) The structure of the MLP with five input neurons, one layer of five hidden neurons, and two outputs. b) An example of the response surface at 625nm.

The number of hidden neurons is subject to optimal configuration for a particular case study. In general, the complexity of the MLP must be consistent with the amount of information for training – there should be enough data to match the dependence. Choosing too many hidden neurons may lead to over-fitting. The MLP then loses its ability to generalize the information from the samples. In the case of over-fitting, the complexity of the model is too high for the task. The model fits the training data precisely; however it can not generalize when new samples are introduced. To avoid the risk of an over-fitting, a 6-fold cross-validation error was considered. This is a standard and commonly used approach in the model selection. In the cross-validation procedure, a fraction of the available training data is not used for training but for the estimation of the validation error of the model. In *K*-fold cross-validation, the data are split into *K* parts, each of them being consequently excluded from the training set and used to calculate the mean square error of the model. The results are then averaged over all splits, and, if necessary, after several repetitions of the described procedure.

The application of MLP proceeded first through the described structure selection phase. Then the model with the optimized structure was trained on the data, i.e. the weights of the neurons were adapted to minimize the mean squared error. This training employed minimization of the quadratic mean square error (MSE) cost function for optimizing the summation weights at each neuron. An error back-propagation algorithm was applied to calculate the gradient of the MSE and to adapt the weights. The Levenberg-Marquardt algorithm was used for this purpose. After the MLP was trained and the optimal weights were found, the predictions were computed by presenting the intensities at selected five wavelengths of the test samples to the MLP inputs. The results at the outputs were then computed using the trained MLP.

The software which implements the described MLP was developed in the research group of the Institute of Geomatics and Analysis of Risk, University of Lausanne. It is distributed freely for academic use with the book.[2] The authors are grateful to V. Timonin (Institute of Geomatics and Analysis of Risk, University of Lausanne) for providing the software. The results of the calculations are listed in the following table.

Table S1. Real and predicted concentrations of ATP and PPi/cAMP as determined.

	Concentrations added, mM		Concentrations found, mM		Absolute error, mM	
Entry	ATP	PPi/cAMP	ATP	PPi/cAMP	ATP	PPi/cAMP
1	0.10	0.30	0.06	0.33	0.04	0.03
2	0.10	0.70	0.06	0.73	0.04	0.03
3	0.30	0.70	0.25	0.74	0.05	0.04
4	0.30	0.90	0.25	0.90	0.05	0.00
5	0.50	0.10	0.46	0.10	0.04	0.00
6	0.50	0.30	0.46	0.32	0.04	0.02
7	0.70	0.10	0.65	0.10	0.05	0.00
8	0.70	0.50	0.65	0.46	0.05	0.04
9	0.90	0.50	0.85	0.45	0.05	0.05
10	0.90	0.70	0.86	0.80	0.04	0.10

References

1 H. Yang, P. R. Griffiths and J. D. Tate, *Anal. Chim. Acta*, 2003, **489**, 125.

2 M. Kanevski, A. Pozdnoukhov and V. Timonin, *Machine Learning Algorithms for Spatial Data Analysis and Modelling. Theory and Case Studies.* Press Polytechniques et Universitaires Romandes, Lausanne, 2007 (in press).