Available online at www.sciencedirect.com

ScienceDirect

Procedia Engineering 120 (2015) 501 – 506

Procedia Engineering

www.elsevier.com/locate/procedia

EUROSENSORS 2015

Optimization and evaluation of a hapten microarray using chemometric methods

Peter Carl[a,b], Rudolf J. Schneider[a]*

[a] *BAM Federal Institute for Materials Research and Testing, D-12205 Berlin, Germany*
[b] *Institute of Chemistry, Humboldt-Universität zu Berlin, D-12489 Berlin, Germany*

Abstract

The demands for the assessment of water quality are increasing steadily, making it necessary to routinely monitor multiple contaminants in water samples. For this application a hapten microarray was developed. In order to reach the required low detection limits a design of experiments (DoE) approach was used to optimize the assay performance. Here we show that a Box-Behnken design plan is an adequate choice for the straightforward exploration of hapten microarray assay parameters. For both read-out systems studied (fluorophore-labelled detection antibodies or enzymatic signal development followed by reflectometric scan, respectively), it was possible to significantly extend the measurement ranges. Furthermore, it could be shown that multivariate data analysis, here partial least squares regression (PLS), can improve the prediction accuracy of "unknown" samples when used as calibration model, compared to classical, univariate data evaluation methods.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of the organizing committee of EUROSENSORS 2015

Keywords: Proteine microarray; design of experiments; multivariate data analysis; partial least squares regression; small molecule detection.

1. Introduction

The measurement of anthropogenic markers in wastewater samples is becoming a strategy to monitor fluctuating influent concentrations and removal efficiencies. The surveillance of water quality demands for measurements of multiple marker substances. Traditional immunoassays are adequate high-throughput tools to monitor the distribution of pollutants in waterways. However, only a single analyte at a time can be detected. In contrast, protein microarrays proved their potential in the simultaneous analysis of multiple analytes, and also in environmental analysis[1].

* Corresponding author. Tel.: +49-30-8104-1151; fax: +49-30-8104-71151.
E-mail address: rudolf.schneider@bam.de

1877-7058 © 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license
(http://creativecommons.org/licenses/by-nc-nd/4.0/).
Peer-review under responsibility of the organizing committee of EUROSENSORS 2015
doi:10.1016/j.proeng.2015.08.684

The idea was first described by Ekins [2] and there exist multiple examples of analytical microarrays detecting e.g. antibiotics in milk[3], steroids in urine[4], pesticides[5], mycotoxins[6], e.g. in cereals[7].

In principle hapten microarrays are analogous to indirect, competitive immunoassays, with the difference that several primary antibodies with different selectivities are used for the competitive reaction. Fig. 1 shows the measurement principle of a hapten microarray: In the first step (A), the so-called printing process, conjugates of the protein (ovalbumin) and the hapten are immobilized on the *N*-hydroxysuccinimide activated surface of a glass slide. The proteins are arranged in the shape of spots. The encoding of the corresponding analyte is defined by using a two dimensional grid called subarray. The analytes covalently bound to the proteins and the analytes in the sample compete for the antigen binding sites of the analyte-specific antibodies (B). Bound antibodies are visualized by antibody-fluorophore and enzyme conjugates (C), respectively. The signal intensities are then measured by a microarray fluorescence scanner or by a reflectometric scan, using a flatbed scanner.

The univariate evaluation of the assay is done via determination of the median signal intensity of each spot. The reduced mean intensity of each spot repetition are then plotted against the concentration of the analyte on logarithmic scale. A four-parametric equation is used to fit the calibration curve.

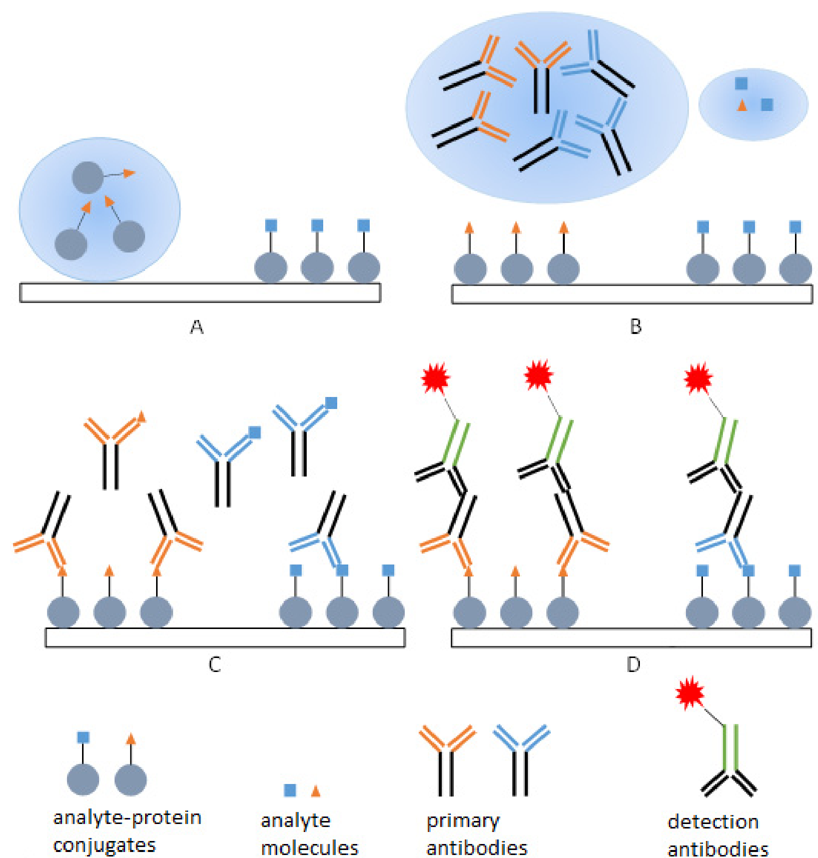

Figure 1: Flow scheme of a hapten microarray. A) Immobilization of hapten protein conjugates. B and C) Competitive reaction of analyte molecules and immobilised haptens for the antigen binding sites of the primary antibodies. D) Detection with labelled antibodies. Non-bound reagents are washed away after the steps A, C and D

In order to achieve the required sensitivity and robustness of the method, many assay parameters have to be optimized. Furthermore, spot inhomogeneities, fluctuating background and a large amount of data cause massive evaluation problems. To overcome these problems, two chemometric methods were studied to improve the assay performance: first, a so-called design of experiments (DoE) for rapid optimization of the assay parameters; second, the use of multivariate data analysis for the calibration of the assay. The commercial software "The Unscrambler® X." (CAMO) was used in this study for DoE as well as PLS.

PLS is a principle component regression (PCR) based method. Therefore a data matrix **X**, containing n · m measure variables is correlated to a target matrix **Y**, containing 1 · m target components (PLS-1). The model is calculated using the system of equations given in equation 1:

$$X = T P^T + E$$
$$Y = U Q^T + F \qquad (1)$$

Where **T** and **U** are projections on **X** and **Y** and were called **X**- and **Y**-Scores. **P** and **Q** are called Loadings of **X** and **Y** and are generated by a principle component analysis. The regression is done in such a way, that **T** and **U** show the highest possible covariance. In order to optimize the root means square error of prediction (RMSEP, equation 2), the PLS models were validated by "leave-one-out" cross validation.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \qquad (2)$$

Multivariate data analysis, especially PLS has shown great performance for the classification of genes in the evaluation of genetic microarrays [8,9]. However, to the best of our knowledge, it has not yet been tested for the application in hapten microarrays. To investigate the efficiency of those two chemometric methods for this purpose, we have selected two microarray formats for the determination of caffeine (CAF) and carbamazepine (CBZ). CBZ, an anticonvulsant e.g. for the treatment of epilepsy, is being released into the environment in significant amounts. It is almost not degraded in wastewater treatment plants and is found in the influent of wastewater treatment plants at concentrations of up to 3.8 µg/L and in mean concentration in surface waters around 0.5 µg/L [10]. CAF is well-known for its psychoactive/stimulating properties and is today an accepted anthropogenic marker [11].

2. Results

2.1. Optimization of the assay performance using DoE

A Box-Behnken design plan was used for optimization of the hapten microarray. The efficiency of the method was demonstrated by a microarray system for the determination of caffeine in water samples. In this study we investigated the influence of a) the concentration of the (primary and secondary) antibodies, b) the concentration of the spotted hapten protein conjugate, c) the influence of the pH of three different buffers d) the incubation times of the antibodies and in the case of reflectometric scan detection e) the signal development time.

The Box-Behnken design plans were carried out by combining the middle-level range of each parameter with the high and low-level range of the other investigated parameters, respectively. The maximal signal intensity (SMax) and the signal-to-noise ratio (S/B) have been selected as response variables. Each experimental design was evaluated by Analysis of Variance (ANOVA) in order to obtain information about the influence of the parameters on the response variables, followed by generating a quadratic model (response-surface-model, Fig. 2) to ascertain the optimal levels of each parameter.

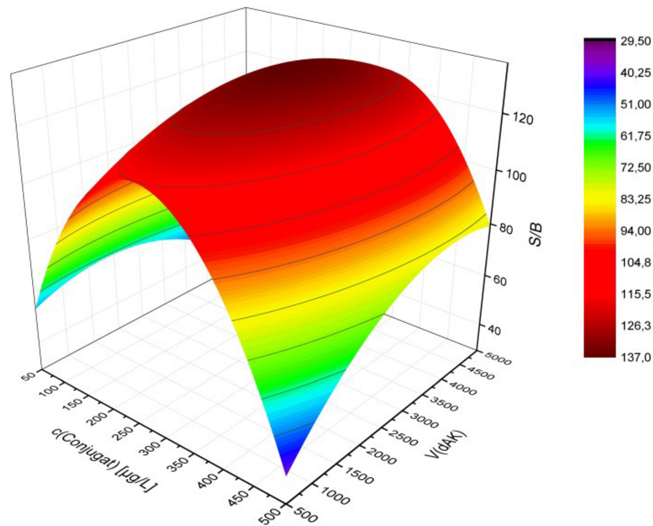

Figure 2. Three-dimensional response-surface model, demonstrating the influence of two assay parameters on the signal to noise (S/B) ratio

The dilution of anti-caffeine-antibodies and the pH values of buffers revealed to exhibit the most significant influences on the S/B of the microarray. Other parameters had a statistically significant influence on the S_{Max} but did not influence the S/B. Hence, improving them might only increase the measurement range, but not the limit of detection, which is restricted by the S/B [12].

To validate the optimization two calibration curves for each detection method were recorded: under initial microarray conditions and under optimized conditions (Fig. 3)

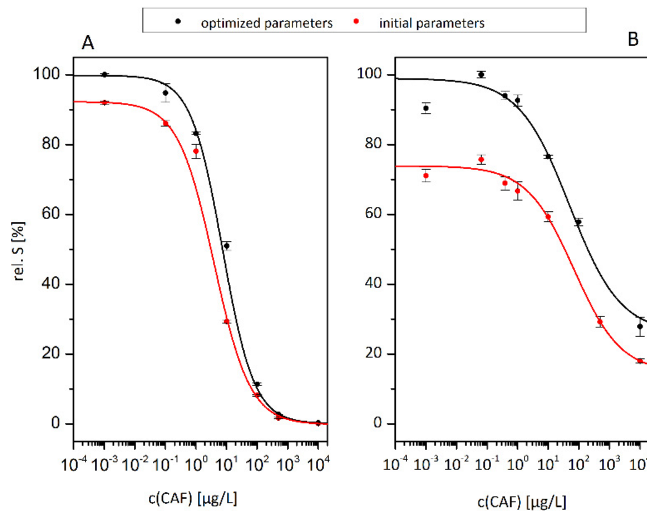

Figure 3 Calibration curves for CAF microarrays using initial and optimized parameters. A: Fluorimetric detection. B: Refelctometric detection

After optimization, a wider measurement range (here determined as the difference between the upper and lower asymptote) could be achieved for both detection methods, with a maximal 5.5 times higher range. In case of the reflectometric scan detection a sensitivity increase by 27 % could be obtained by comparing the inflection points of the standard curves, which are used in classical immunoassays as a measure of sensitivity.

2.2. Multivariate data analysis of microarray calibrations

For evaluation of the microarray, a PLS was shown to give the best calibration models. Therefore, a data processing method including transformation of the microarray data into greyscale data, dimension reduction and logarithmic transformation of the corresponding Y matrix was established. Adequate calibration models with a minimal RMSEP could be achieved by convolution and logarithmic transformation of the microarray data.

In two different examples (hapten microarrays for either CAF or CBZ) it could be shown that PLS calibration models lead to a more correct prediction of "unknown" samples than univariate calibration, because information e.g. of the signal background, is included in the calibration model. For comparing both evaluation methods (uni- and multivariate) internal (CAF microarray) and external (CBZ) validations were performed and the prediction errors were compared. As it can be seen in Fig. 4 for the internal validation of the CAF microarray, even lower concentrations of the analyte can be detected: in this case the measurement range is set to a limit of 30 % relative error.

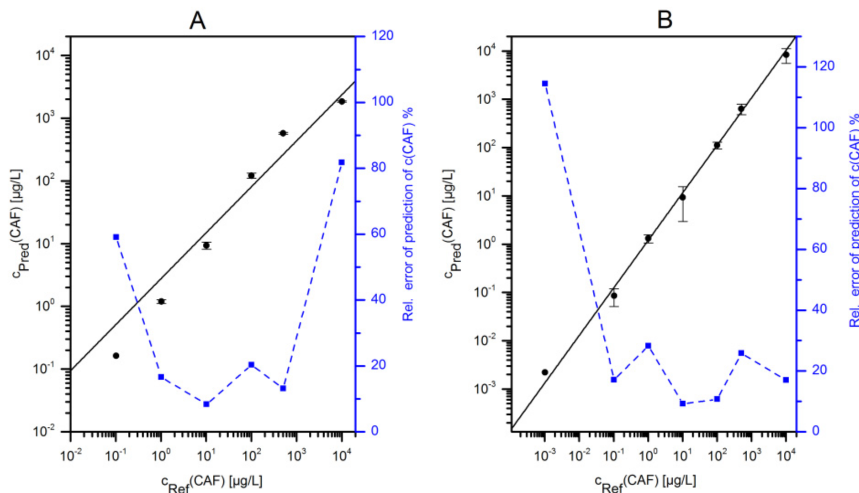

Figure 4. In-house validation of the CAF microarray. A: Univariate calibration. B: PLS calibration.

The univariate calibration model is limited by an upper and lower detection limit, whereas the PLS calibration model only provides a lower measurement range, which is even lower than for the univariate calibration. (0.1 µg/L and 1 µg/L). Similar results were obtained performing an external validation of a CBZ microarray: The achieved prediction errors of the PLS calibration were smaller, compared to the univariate calibration.

Furthermore, another multivariate data analysis method, the so-called support vector machine regression, was examined for application in the calibration of the microarrays. However, compared to the PLS and univariate data analysis, higher prediction errors were obtained due to "over-fitting" of the model.

3. Conclusion

Two different chemometric methods were investigated for the optimization of hapten micorarrays: design of experiments and multivariate data analysis. Both methods revealed a very good performance for fast optimization of the studied microarrays.

With the use of DoE, it was possible to gain an insight in the influence of parameters on the performance of the microarray. After final optimization, a hapten microarray with two different detection methods exhibited an extension of the measurement range. In addition, in case of the reflectometric detection method, even an increase in sensitivity could be achieved.

Furthermore, it could be demonstrated that PLS calibration of hapten microarrays leads to more precise calibration models than univariate calibration. Lower prediction errors are achieved by including important information, e.g. the background signal, in the calibration. Thus, lower limits of detection compared to univariate calibration can be reached. In addition, the method is faster than univariate calibration because the amount of data is significantly reduced.

4. References

[1] Z. Fan, Y. S. Keum, Q. X. Li, W. L. Shelver, L. H. Guo, Sensitive immunoassay detection of multiple environmental chemicals on protein microarrays using DNA/dye conjugate as a fluorescent label. *J. Environ. Monitor*. 14 (2012) 1345-1352.

[2] R. Ekins, F. Chu, E. Biggart, Development of microspot multi-analyte ratiometric immunoassay using dualfluorescent-labelled antibodies, *Anal. chim. acta* 227 1989 73-96.

[3] B.G. Knecht, A. Strasser, R. Dietrich, E. Martlbauer, R. Niessner, M.G. Weller, Automated microarray system for the simultaneous detection of antibiotics in milk, *Anal. Chem.* 76 (2004) 646-654.

[4] H. Du, Y. Lu, W. Yang, M. Wu, J. Wang, S. Zhao, M. Pan, J. Cheng, Preparation of Steroid Antibodies and Parallel Detection of Multianabolic Steroid Abuse with Conjugated Hapten Microarray, *Anal. Chem.* 76 (2004) 6166-6171

[5] E. Belleville, M. Dufva, J. Aamand, L. Bruun, L. Clausen, C.B. Christensen, Quantitative microarray pesticide analysis, *J. Immunol. Methods* 286 (2004) 219-229

[6] J.C. Sauceda-Friebe, X.Y. Karsunke, S. Vazac, S. Biselli, R. Niessner, D. Knopp, Regenerable immuno-biochip for screening ochratoxin A in green coffee extract using an automated microarray chip reader with chemiluminescence detection, *Anal. chim. acta* 689 (2011) 234-242.

[7] S. Oswald, X.Y.Z. Karsunke, R. Dietrich, E. Märtlbauer, R. Niessner, D. Knopp, Automated regenerable microarray-based immunoassay for rapid parallel quantification of mycotoxins in cereals, *Anal. Bioanal. Chem.* (2013) 1-11.

[8] A.L. Boulesteix, PLS dimension reduction for classification with microarray data, Statistical applications in genetics and molecular biology 3 (2004) Article33.

[9] D.V. Nguyen, D.M. Rocke, Tumor classsification by partial least squares using microarray gene expression data, Bioinformatics, 18 (2002) 39-50.

[10] A. Bahlmann, M. Weller, U. Panne, R. Schneider, Monitoring carbamazepine in surface and wastewaters by an immunoassay based on a monoclonal antibody, *Anal. Bioanal. Chem.* 395 (2009) 1809-1820.

[11] I.J. Buerge, T. Poiger, M.D. Müller, H.-R. Buser, Caffeine, an anthropogenic marker for wastewater contamination of surface waters, *Environ. Sci. Technol.* (2003) 691–700.

[12] W. Luo, M. Pla-Roca, D. Juncker, Taguchi design-based optimization of sandwich immunoassay microarrays for detecting breast cancer biomarkers, Anal Chem, 83 (2011) 5767-5774.