J|A|C|S
ARTICLES
Published on Web 02/01/2006

Pattern Recognition of Proteins Based on an Array of Functionalized Porphyrins

Huchen Zhou, Laura Baldini, Jason Hong, Andrew J. Wilson, and Andrew D. Hamilton*

Contribution from the Department of Chemistry, Yale University, P.O. Box 208107, 225 Prospect Street, New Haven, Connecticut 06520-8107

Received October 6, 2005; E-mail: andrew.hamilton@yale.edu

Abstract: A practical protein-detecting array is desirable for its potential application in proteomics, medical diagnostics, and pathogen detection. Here, we report a novel protein-detecting array based on porphyrins containing peripheral amino acids as protein surface receptors. The array of porphyrin receptors showed a unique pattern of fluorescence change upon interaction with certain protein samples. Both metal and nonmetal-containing proteins and mixtures of proteins gave distinct patterns, allowing their unambiguous identification. The composite pattern for each sample was subjected to principal component analysis (PCA) to generate a clustering map for more practical visualization. Increasing the number of porphyrin receptors from eight to sixteen gave improved resolution, suggesting that this array is expandable to give satisfactory resolution for any given sample system by carefully maximizing the chemical diversity of the receptors.

Introduction

The development of strategies for the detection of proteins remains a major problem whose solution is imperative for applications in proteomics, medical diagnostics, and pathogen detection.[1] All of these require the identification of protein species and their concentrations as biological markers of cellular responses to external stimuli, as markers for certain pathophysiological conditions or the presence of toxic pathogens. Protein-detecting arrays remain under-developed due to the lack of highly selective and specific binding agents that interact with protein surfaces through complementary interactions. Currently, antibodies are the only class of such ligands, but their application is limited by difficult preparation procedures and their instability. Although antibody fragments, nucleic acid aptamers, and peptoid protein ligands[2] have been explored, they still bear the problem of chemical instability and the need to label protein analytes with fluorescent tags.

Complementary to the highly specific recognition described above is the strategy of cross-responsive arrays that are inspired by the mammalian olfactory system.[3−5] Each nasal receptor binds to more than one odorant molecule, while each odorant molecule binds to more than one nasal receptor. Thus, a multicomponent odor generates a unique composite pattern of interactions with receptors that is distinctive and allows the recognition of a particular odor. An array of nonspecific or weakly interacting agents would give rise to distinctive fingerprints in response to an analyte. This strategy has been the basis of the so-called electronic nose that utilizes an array of conducting organic polymer films to generate a pattern of resistivities that provides a fingerprint for each odor after exposure. The cross-responsive strategy has also been illustrated in the cases of ATP/GTP differentiating guanadinium-tripeptide receptors[6] by Anslyn, chemically responsive dyes for detection of volatile organic compounds[4] (VOC) by Suslick, enzyme activity fingerprinting with substrate cocktails and HPLC techniques[7] by Reymond, oligonucleotide-based steroid-detecting sensors[8] by Stojanovic, an array of tin derivatives for sensing amine vapors[9] by Buriak, and the detection of 20 amino acids using IDA methods[10] by Severin. Most recently, this strategy has been demonstrated by Anslyn to distinguish proteins and glycoproteins using a series of compounds bearing amino acids and boronic esters as the binding functionalities.[11]

Our work on synthetic protein surface receptors has shown that certain tetraphenylporphyrins can bind to complementary protein surfaces with considerable selectivity. The hydrophobic core of the porphyrin primarily contributes to binding affinity, and the periphery can be modified with a mixture of hydrophobic and hydrophilic groups to enhance selectivity toward different proteins. Recently, we have reported a series of tetraphenylporphyrins functionalized with amino acid derivatives that bind to cytochrome *c* with affinities (K_d) ranging from

(1) Kodadek, T. *Chem. Biol.* **2001**, *8*, 105−115.
(2) Kodadek, T.; Reddy, M. M.; Olivos, H. J.; Bachhawat-Sikder, K.; Alluri, P. G. *Acc. Chem. Res.* **2004**, *37*, 711−718.
(3) Gouma, P.; Sberveglieri, G. *MRS Bull.* **2004**, *29*, 697−700.
(4) Suslick, K. S. *MRS Bull.* **2004**, *29*, 720−725.
(5) Freund, M. S.; Lewis, N. S. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2652−2656.
(6) McCleskey, S. C.; Griffin, M. J.; Schneider, S. E.; McDevitt, J. T.; Anslyn, E. *J. Am. Chem. Soc.* **2003**, *125*, 1114−1115.
(7) Goddard, J.-P.; Reymond, J.-L. *J. Am. Chem. Soc*. **2004**, *126*, 11116−11117.
(8) Stojanovic, M. N.; Green, E. G.; Semova, S.; Nikic, D. B.; Landry, D. W. *J. Am. Chem. Soc.* **2003**, *125*, 6085−6089.
(9) Baldauff, E. A.; Buriak, J. M. *Chem. Commun.* **2004**, 2028−2029.
(10) Buryak, A.; Severin, K. *J. Am. Chem. Soc.* **2005**, *127*, 3700−3701.
(11) Wright, A. T.; Griffin, M. J.; Zhong, Z.; McCleskey, S. C.; Anslyn, E. V.; McDevitt, J. T. *Angew. Chem., Int. Ed.* **2005**, *44*, 6375−6378.

10.1021/ja056833c CCC: $33.50

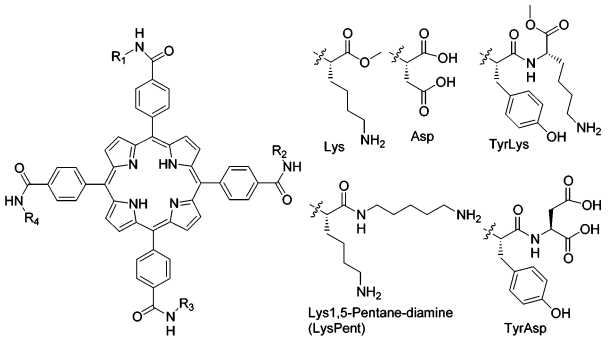
micromolar to nanomolar amounts.[12] These differential recognition properties prompted us to develop an 8-porphyrin array system that can detect different metal-containing proteins based on their distinct fluorescent quenching patterns upon binding to protein surfaces.[13] A synthetic tetra-*meso*-carboxylphenylporphyrin (TCPP) that has a hydrophobic surface area >300 Å^2 was chosen as the core of the receptors. The free-base TCPP derivatives have the advantage of being fluorescent; thus, protein tagging is not needed in detection. Furthermore, the carboxylate groups provide convenient handles for attaching a variety of functionalities for differential recognition of protein surfaces. A library of TCPPs conjugated with amino acids or amino acid derivatives (Chart 1) showed differential binding properties to a range of proteins due to their variable surface characteristics. Although none of the TCPP derivatives has antibody-like high affinity or specificity, together they form an array that gives a composite pattern for the identification of each individual protein.

In this paper, we describe the identification of both metal and nonmetal-containing proteins and protein mixtures using the porphyrin array combined with pattern recognition techniques. Upon surface binding to the porphyrin receptors, a protein such as lysozyme that does not contain a paramagnetic metal center gives a significantly weaker response than a metal-containing protein such as cytochrome *c*, making naked-eye recognition difficult. However, detection of nonmetalloproteins could be easily achieved using pattern recognition techniques. The series of composite patterns generated by protein samples was subjected to principal component analysis (PCA),[14] a linear transformation technique generally used for dimensionality reduction. PCA mapping converts the unique fingerprints of various samples into distinguishable clusters in a 3-D Euclidean space, allowing in turn for the unambiguous recognition of proteins of interest.

Experimental Procedures

General Methods. All proteins were purchased from Sigma Aldrich. Type I bovine milk α-lactalbumin and *Clostridium pasteurianum* ferredoxin were exchanged to 5 mM pH 7.4 phosphate buffer using a Micron centrifugal filter with a 3000 Da cutoff. The protein concentrations were standardized by the absorbance at 280 nm in 6 M GdmHCl, 20 mM phosphate, pH 6.5 ($\epsilon = 3.84 \times 10^4$ cm^{-1} M^{-1} chicken egg white lysozyme, 2.84×10^4 cm^{-1} M^{-1} α-lactalbumin), at 550 nm after reduction using sodium dithionite in 5 mM pH 7.4 buffer ($\epsilon = 2.94 \times 10^4$ cm^{-1} M^{-1} horse heart cytochrome *c*) or at 390 nm in 5 mM pH 7.4 phosphate buffer ($\epsilon = 3.00 \times 10^4$ cm^{-1} M^{-1} *C. pasteurianum* ferredoxin). All solvents were purchased from Mallinckrodt or Aldrich, and all reagents were purchased from Aldrich unless otherwise stated and used without further purification. Preparative TLC was conducted using Analtech 1000 mm silica gel precoated plates with fluorescent indicator active at UV245. ^{1}H NMR spectra were acquired on either Bruker DPX 400 or DPX 500 series spectrometers. Nominal mass spectrometry data were obtained by Dr. Walter McMurray at W. M. Keck Foundation Biotechnology Resource Laboratory of Yale University. Analytical HPLC was performed on a Rainin HP controller with a Rainin UV detector, both attached to a Dell Optiplex PC running Varian StarWorkstation software. UV absorbance was measured using a Molecular Devices Spectramax250 plate reader spectrophotometer. Fluorescence emission was measured using Analyst AD plate fluorescence reader.

Chart 1



	R_1	R_2	R_3	R_4
A	TyrAsp	TyrAsp	TyrAsp	TyrAsp
B	TyrAsp	TyrAsp	TyrAsp	LysPent
C	TyrAsp / Lys	Lys / TyrAsp	Lys	TyrAsp
D	Lys	Lys	Lys	LysPent
E	TyrLys / LysPent	LysPent / TyrLys	LysPent	TyrLys
F	Asp	Asp	Asp	Asp
G	TyrAsp / TyrLys	TyrLys / TyrAsp	TyrLys	TyrAsp
H	Lys / LysPent	LysPent / Lys	LysPent	Lys
I	TyrLys	TyrLys	TyrLys	TyrLys
J	Lys	Lys	Lys	Lys
K	Asp	Asp	Asp	TyrLys
L	Asp	Asp	Asp	Lys
M	Asp	Asp	Asp	LysPent
N	Asp	Asp	Asp	TyrAsp
O	Lys / Asp	Asp / Lys	Asp	Lys
P	Lys / TyrLys	TyrLys / Lys	TyrLys	Lys

Typical Experimental Conditions for the Fluorescent Array. The solutions of different porphyrins were prepared by dissolving each compound in a minimal volume of DMSO and then diluting to a final concentration of 5 μM in 5 mM pH 7.4 phosphate buffer containing 0.05% tween-20. Each porphyrin solution was loaded at 200 μL per

(12) Tsou, L. K.; Jain, R. K.; Hamilton, A. D. *J. Porphyrins Phthalocyanines* **2004**, *8*, 141−147.
(13) Baldini, L.; Wilson, A. J.; Hong, J.; Hamilton, A. D. *J. Am. Chem. Soc.* **2004**, *126*, 5656−5657.
(14) Chatfield, C.; Collins, A. J. *Introduction to Multivariate Analysis*; Chapman and Hall, Ltd.: London, 1983.

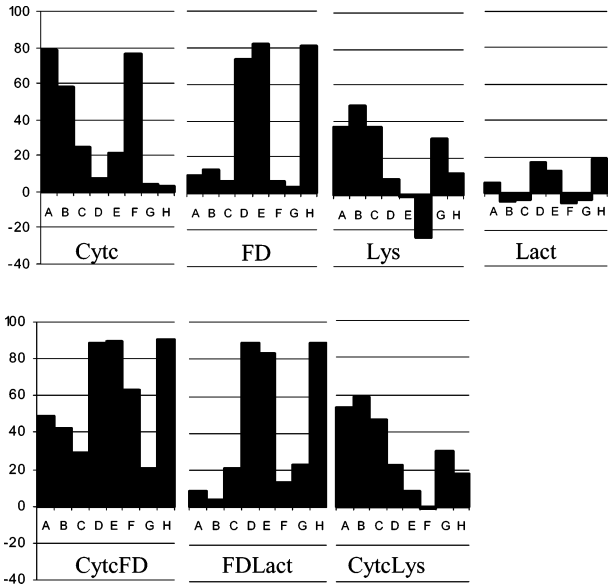


Figure 1. 8-Porphyrin array gave the fingerprints for protein samples.

well to a row in a multi-well plate, and the fluorescence emission was scanned. The excitation filter was GFP-wide blue excitation D455/70x, and the emission filter was Texas red D630/60m from Chroma Tech. To each porphyrin solution, 1.5 or 3 equiv of protein was added as a ~5 μL solution. The fluorescent emission was measured again after 5 min.

Discussion

The porphyrin-based receptors, TCPP derivatives **A**−**P**, have a variety of peripheral functionality composed of charged and hydrophobic amino acids and their derivatives, as shown in Chart 1. They were synthesized by coupling TCPP with a mixture of two peptidic derivatives followed by preparative TLC purification.[12] In forming the protein-detecting array, the porphyrins were first loaded onto a multi-well plate at 5 μM × 200 μL per well in 5 mM pH 7.4 phosphate buffer containing 0.05% tween-20. After measurement of their initial fluorescent intensity, each individual protein sample was added to eight different porphyrins, **A**−**H**, and the fluorescence intensity was scanned again. A plot of the change in fluorescence intensity as a percentage of the original $(((F_0 - F)/F_0)100\%)$ gave each protein sample an 8-D fingerprint, as shown in Figure 1.

Each protein surface has its own characteristic distribution of hydrophobic, neutral, and charged amino acid residues, which provide the basis for potential differential recognition by synthetic molecules (Figure 2). The varying shapes and sizes of the proteins will lead to different extents of contact with the near planar porphyrin scaffold. The affinity of a porphyrin toward a protein will depend on how well the peripheral functionalities and the macrocyclic core of the porphyrin complement the hydrophobic patches lined with polar or charged residues on the surface of the protein. Among the metal-containing proteins tested, basic cytochrome *c* (Cytc) and acidic ferredoxin (FD) gave distinctly different patterns since the charge and hydrophobicity distribution features on their respective surfaces are quite different (Figure 2). The paramagnetic iron in both proteins caused strong quenching of the fluorescence emission of those porphyrins to which they bind with high affinity. Despite lacking a paramagnetic metal center, lysozyme (Lys) and α-lactalbumin (Lact) both gave distinctive patterns,

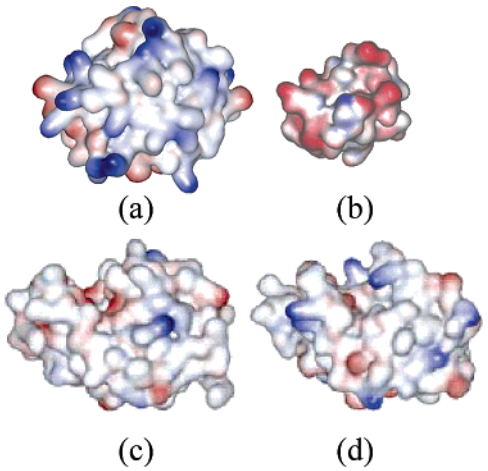


Figure 2. Electrostatic representation of (a) horse heart cytochrome *c*, (b) *C. pasteurianum* ferredoxin, (c) chicken egg white lysozyme, and (d) TypeI bovine milk α-lactalbumin. Hydrophobic patches are represented in white, acidic patches in red, and basic patches in blue.

as shown in Figure 1. In this case, the fluorescence changes are likely due to changes in the environment of the fluorophore caused by association with the protein surface. Binding to more polar or nonpolar surfaces will cause either a decrease or increase of fluorescence intensity, while possible energy transfer between the photoexcited porphyrin and protein residues such as tyrosine and tryptophan will further modify the responsive patterns. This observation opened the door to using carefully selected porphyrin arrays to detect all proteins regardless of whether they contain paramagnetic metal centers or not.

The previously reported recognition of different odors by an electronic nose is based on weak and nonspecific interactions of molecules in the vapor phase with an array of conducting polymer films.[5] The classification of different types of odors can be realized after processing the data matrix using mathematical transformations such as principal component analysis (PCA). This involves a mathematical procedure that transforms a number of correlated variables into a number of uncorrelated variables called principal components.[14] The first principal component (PC1) accounts for most of the variability in the data, and the rest of the *n* principal components have descending significance in presenting the variability. Thus, the first three principal components can represent the variability of the whole data set without losing much information. In our case shown below, the combination of the first three PCs contained 98% of the information.

In Figure 3, the PCA of the data obtained from 10 protein samples on an 8-porphyrin sensor array is shown in 3-D Euclidean space. Satisfactory sample resolution was achieved on these randomly selected samples either in their pure forms or as mixtures. Cytochrome *c* and ferredoxin are readily distinguishable, which is not surprising considering their opposite surface charges and their paramagnetic metal-containing prosthetic groups (Figure 3). Furthermore, the metal-free proteins lysozyme and α-lactalbumin are confidently distinguished from each other and from other protein samples. At the same time, this 8-porphyrin sensor array shows a clear concentration dependent response for some of the proteins. The solutions of cytochrome *c* at 7.5 μM (Cytc1.5e) and 15 μM (Cytc3e) can be distinguished in the PCA mapping as clusters **f** and **g**. This allows for an estimation of the concentration of a

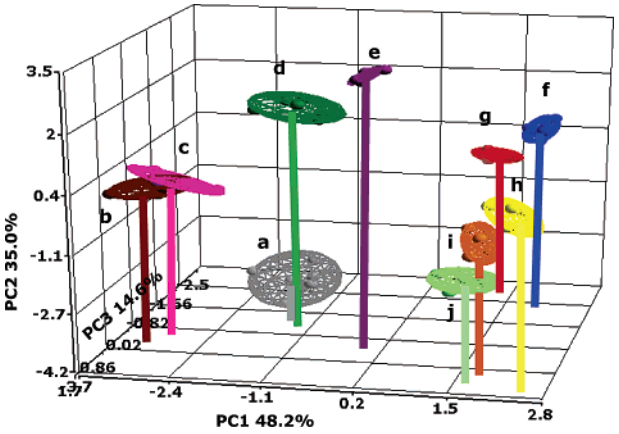


Figure 3. PCA mapping of the 10 samples identified by an 8-porphyrin array. Final concentrations: porphyrins 5 μM; proteins 7.5 μM (1.5e) or 15 μM (3e). Abbreviations: cytochrome *c* (Cytc), ferredoxin (FD), lysozyme (Lys), and α-lactalbumin (Lact). Samples: **a**-Lact1.5e, **b**-FD1.5e, **c**-FDLact1.5e, **d**-CytcFD1.5e, **e**-CytcFD3e, **f**-Cytc3e, **g**-Cytc1.5e, **h**-CytcLys3e, **i**-CytcLys1.5e, and **j**-Lys1.5e. Each sample was repeated 6 times over a period of 3 months to validate its reproducibility. Drop-bars are added to enhance the visualization of depth.

cytochrome *c* solution by the location of the sample point in the Euclidean space. The same concentration gradient is also observed for cytochrome *c* mixed with ferredoxin (clusters **d** and **e**) or with lysozyme (clusters **h** and **i**). Thus, the porphyrin array allows not only the detection of a protein but also an estimation of its concentration in pure or mixed forms.

Identification of a protein mixture is more demanding because interactions between the proteins can lead to the failure of even a highly selective receptor.[1] Our strategy of using receptors with modest affinity and selectivity has an advantage because it allows the sum of all protein−receptor and protein−protein interactions to be read. The mixture of cytochrome *c* and ferredoxin (**d** and **e**) is clearly distinguished from either cytochrome *c* (**f** and **g**) or ferredoxin (**b**) alone, as shown in Figure 3. The pattern of a mixture of two proteins is not a simple addition or average of the patterns of its two components; instead it is a composite pattern that reflects the equilibrium of the competitive binding of two proteins to the array and the interaction between the two proteins. The mixture of cytochrome *c* and lysozyme (**h** and **i**) is distinct from either of its components (**f** + **g** and **j**), suggesting that a strong-quenching protein like cytochrome *c* does not mask the pattern expression with its own signal when coexisting with a weak-quenching protein. As further evidence, the mixture of ferredoxin and α-lactalbumin (**c**) is distinguishable from either ferredoxin (**b**) or α-lactalbumin (**a**), although the pattern characteristic of α-lactalbumin is significantly weaker. In this case, **c** is located closer to **b** than **a**, indicating that although the pattern variation caused by the presence of α-lactalbumin in the mixture is a minor component of the resulting pattern, it is significant enough to distinguish ferredoxin (**b**) from a ferredoxin/α-lactalbumin mixture (**c**). Maximizing the resolving power of the porphyrin sensor array by increasing the diversity of the porphyrin receptors should give even better resolution for these neighboring clusters.

During the development of conducting polymer-based vapor detection systems, it has been found that increasing the number

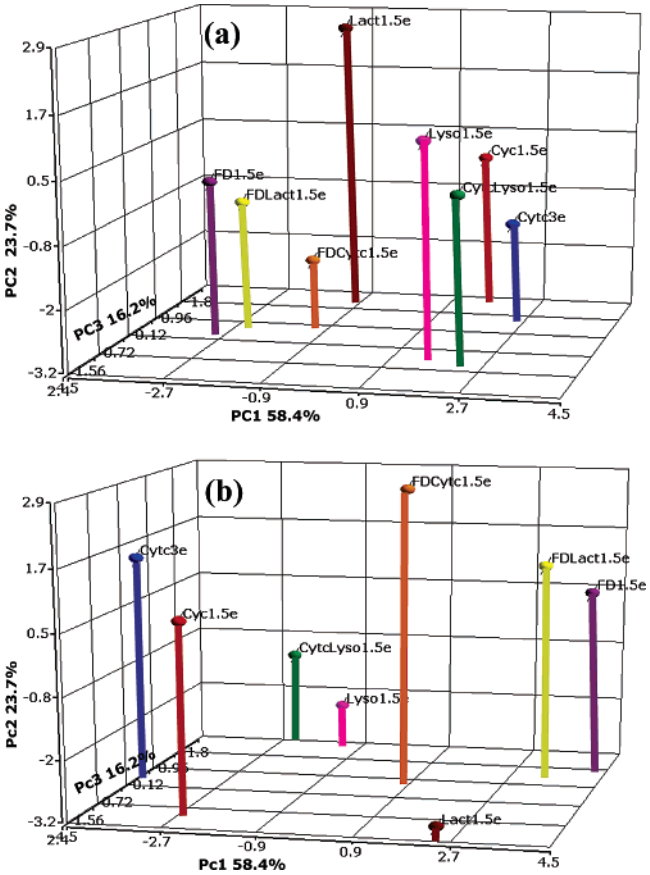


Figure 4. PCA mapping of protein samples against an 8-porphyrin array (a) or a 16-porphyrin array (b). To get the data matrix for b, the samples were tested on a 16-porphyrin array, while a is extracted from b by taking the data matrix of the first eight porphyrin receptors.

of detectors in an array increased the performance of the array.[15] This is because an increase in the diversity of the detectors leads to an increase in the information content of the responsive output. However, the effect was not linear. After a certain point, increasing the number of receptors lead to the resolving power reaching a plateau suggesting that the diversity is maximized within the available classes of sensors. The resolving power of our protein-detecting porphyrin sensor array, however, can at this point be increased by the addition of new porphyrin receptors to the array. As shown in Figure 4, the average Euclidean distances between samples are improved from 1.27 to 1.62 when the 8-porphyrin array is expanded to a 16-porphyirn array. The least well-resolved pairs are also improved: the distance from cytochrome *c* 7.5 μM to 15 μM (Cytc1.5e vs Cytc3e in Figure 4) increased from 0.37 to 0.51 and from ferredoxin 7.5 μM to the mixture of ferredoxin and α-lactalbumin 7.5 μM (FD1.5e vs FDLact1.5e in Figure 4) increased from 0.26 to 0.34. This observation suggests that by optimizing the number and diversity of the porphyrin receptors in an array, it should be possible to further expand the resolving power and detecting range of this protein-detecting system. With the assistance of statistical analysis such as the clustering of the variables,[16] the most diverse set of probes can be selected for a given detecting task. An optimized array for the detection of physiologically related biological samples is under development.

(15) Doleman, B. J.; Lonergan, M. C.; Severin, E. J.; Vaid, T. P.; Lewis, N. S. *Anal. Chem.* **1998**, *70*, 4177−4190.
(16) Grognux, J.; Reymond, J.-L. *ChemBioChem* **2004**, *5*, 826−831.

Conclusions

We report here a protein-detecting array composed of porphyrins modified with amino acids or amino acid derivatives at the periphery. Labeling of sample proteins is not required since the fluorescence of the porphyrin receptors serves as the monitoring signal. The unambiguous identification of a variety of proteins including both metal and nonmetal-containing proteins as well as mixtures was achieved with an 8-porphyrin array after processing the data matrix by principal component analysis (PCA) clustering. We have further shown that the introduction of new porphyrin receptors to this array can improve the resolving power. Considering the large number of porphyrins that can be rapidly synthesized using our simple functionalization procedure, these types of porphyrin arrays are promising for the detection of even more complex protein samples. An optimized resolution and detection range can be obtained for a given need by carefully selecting the members of the array. In our future work, we are modifying the macrocyclic template of the porphyrin to add shape-based information.

Acknowledgment. We thank the National Institutes of Health (GM 35208) for financial support of this work.

JA056833C