

# Protein “fingerprinting” in complex mixtures with peptoid microarrays

M. Muralidhar Reddy and Thomas Kodadek\*

Departments of Internal Medicine and Molecular Biology, Division of Translational Research, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX 75390-9185

Edited by Jack W. Szostak, Harvard University, Boston, MA, and approved July 25, 2005 (received for review February 15, 2005)

**We report here that microarrays comprised of several thousand peptoids (oligo-*N*-substituted glycines) are useful tools for the identification of proteins via a “fingerprinting” approach. By using maltose-binding protein, glutathione *S*-transferase, and ubiquitin, a specific and highly reproducible pattern of binding was observed when fluorescently labeled protein was hybridized to the array. A similar pattern was obtained when binding of an unlabeled protein to the array was visualized by secondary hybridization of a labeled antibody against that protein, showing that native proteins can be identified without the requirement for prior chemical labeling. This work suggests that small-molecule microarrays might be used for more complex fingerprinting assays of potential diagnostic value.**

profiling | proteomics | small molecule microarrays

Small-molecule microarrays (SMMs) (1–3) are becoming increasingly important tools in combinatorial chemistry. These arrays are generally produced by first synthesizing a combinatorial library on a suitable bead resin, separating the beads into the wells of microtiter plates, and then releasing the compounds from the beads (4, 5). The resulting solutions then are spotted robotically onto a chemically modified glass slide such that the library-derived molecule is attached covalently to the surface. Alternatively, methods exist for the synthesis of certain classes of compounds *in situ* on the array surface (6–10).

By far the most common application of SMMs has been as a versatile platform for library screening, usually with the goal of identifying small-molecule ligands for a given protein of interest (1, 11–13). However, little work has focused on the development of SMMs as analytical tools for biological research (14–16), with the notable exception of peptide arrays as tools for determining the substrate preferences for proteases and other protein-modifying enzymes (12, 17–21) and the binding preferences of antibodies (22–24). We hypothesized that hybridization of any particular protein to a SMM with thousands of features is likely to provide a unique pattern of binding to the array, allowing that factor to be identified by virtue of this “molecular fingerprint.” Indeed, screening experiments have shown qualitatively that different high-affinity ligands are usually, although not always, identified when different proteins are incubated with a combinatorial library (1). The fingerprinting idea simply extends this concept to the quantitative measurement of protein binding at most or all features on a SMM. The expectation is that whereas only a few molecules in a library will be high-affinity ligands for a given protein, a larger number would exhibit above-background binding with a broad spectrum of affinities. Quantitation therefore would provide a fingerprint unique to that protein, because it seems exceedingly unlikely that any two proteins would bind to thousands of different compounds with similar affinities.

Although this approach to protein identification seems reasonable, a number of important technical issues must be addressed to determine its feasibility. For example, libraries that are either large and/or rich in general protein-binding compounds must be made and arrayed, because the fingerprinting concept would not work if only a few compounds on an array

bound the target protein above background. Another critical issue is the reproducibility of this sort of experiment. Derivation of a fingerprint will be far more demanding in this regard than identifying a few of the brightest spots on the array, as is the case in ligand discovery screens.

We show here that arrays comprised of several thousand peptoids (oligo-*N*-substituted glycines) (25–28) can indeed be used to measure protein fingerprints in a highly reproducible fashion. We also demonstrate that this approach can be combined with an antibody-based “sandwich assay” to identify proteins in complex mixtures without chemical labeling of the sample. These studies suggest that peptoid-based SMMs could be developed as useful tools for protein identification in biological milieu.

## Materials and Methods

**General Remarks.** All chemicals and solvents were purchased from commercial suppliers and used without further purification. Ubiquitin (Ub) was obtained from Sigma–Aldrich, maltose-binding protein (MBP) was from New England Biolabs, goat anti-mouse IgG and Alexa Fluor 488 were obtained from Molecular Probes, and anti-glutathione *S*-transferase (GST) antibody was from Santa Cruz Biotechnology. All proteins were labeled by using standard protocols (27, 29). The slides were scanned by using a ScanArray ExpressHT Microarray Scanner (PerkinElmer) at 10- $\mu$ m resolution with 488- and 532-nm excitation lasers.

**Peptoid Library.** The library used in this study was synthesized by using the “submonomer” method (30) as described in ref. 27, except that microwave irradiation (31) was used to accelerate all of the synthetic steps.

**Preparation of Peptoid Microarrays.** Peptoid stock solutions were printed onto a chemically functionalized (maleimide) glass slides by using SpotArray 72 Microarray Printing System (PerkinElmer). The slides then were allowed to stand for 15 h on the printer platform and washed 1 h each with DMSO, dimethylformamide, tetrahydrofuran, and isopropanol. Slides were dried by centrifugation and stored under argon at room temperature. Full details of slide preparation are available from the authors on request.

**Microarray Hybridization and Image Analysis.** Peptoid microarrays were scanned and hybridized with various proteins as described in ref. 29. Slides were equilibrated with 1 $\times$  TBST (50 mM Tris/150 mM NaCl/0.1% Tween 20, pH 7.4) for 15 min and blocked with *Escherichia coli* lysate for 1 h at 4°C. After each step, microarray slides were rinsed in 1 $\times$  TBST before applying

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: MBP, maltose-binding protein; SMM, small-molecule microarray; Ub, ubiquitin.

\*To whom correspondence should be addressed. E-mail: thomas.kodadek@utsouthwestern.edu.

© 2005 by The National Academy of Sciences of the USA

the next protein. Each protein was diluted (see text for concentrations) with  $1\times$  TBST containing 100-fold excess of *E. coli* lysate and applied to the slides. Microarray slides then were incubated for 2 h at  $4^{\circ}\text{C}$  with gentle shaking. The slides were rinsed once and washed with  $1\times$  TBST ( $3\times 4$  min), then dried by centrifugation.

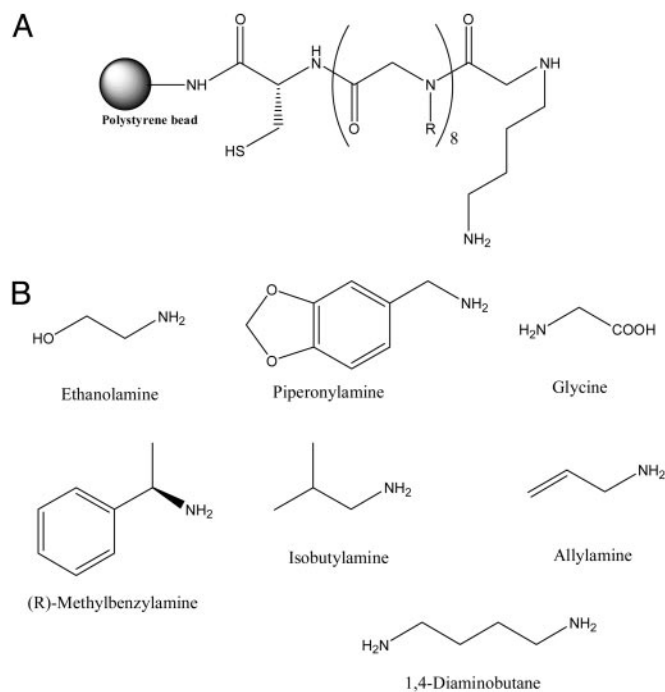
Microarrays were scanned with a ScanArray ExpressHT Microarray Scanner using blue 488-nm (fluorescein-labeled proteins) and green 543-nm (Cy3-labeled proteins) lasers at 100% power and 70% photomultiplier tube gain. All of the scanned images were analyzed by using the GENEPIX PRO 5.0 software (Axon Instruments, Union City, CA). Local background subtracted mean spot intensities were used for further analysis. All of the spot intensities of a slide before protein hybridization were subtracted from the same slide after protein hybridization in EXCEL to get the true signal intensity due to protein binding to each feature. These true signal intensities were used for downstream analysis with GENESPRING software (Silicon Genetics, Palo Alto, CA). Only features that gave positive signal intensity values on both experiments were used to determine reproducibility and standard correlation values. We used these intensities rather than a ratio in which the intensity at any given spot is referenced to that obtained at a control feature. This approach, which is used routinely in DNA microarray analysis, is problematic in this case. In a DNA microarray, or any array in which each feature is designed to capture a specific analyte, one can design controls that are not expected to bind anything specifically. In a protein-fingerprinting experiment, it is not possible to do so.

## Results

**Different Proteins Exhibit Unique and Reproducible Fingerprints When Hybridized to a Peptoid Microarray.** We constructed microarrays consisting of 7,680 different octameric peptoids spotted covalently on a maleimide-functionalized glass microscope slide by using a robotic pin spotter. The peptoid library was created by split and pool synthesis on  $500\text{-}\mu\text{m}$  polystyrene macrobeads (Rapp Polymere, Tübingen, Germany) by using the amines shown in Fig. 1. A C-terminal cysteine residue was included in each molecule to facilitate coupling to the array surface.

To these arrays was hybridized either fluorescein-labeled Ub, fluorescein-labeled GST, or Cy3-labeled MBP in the presence of a 100-fold excess of unlabeled proteins derived from a crude *E. coli* extract (to mimic a moderately abundant protein in a crude extract). The concentration of the labeled protein was 500 nM in each case. After washing, the pattern of binding of the labeled protein to the array was visualized by using a standard commercial microarray scanner. Each experiment was done twice in a completely independent fashion. The raw array images from the first set of hybridizations are shown in Fig. 2, along with an image of an array taken before protein hybridization (far left). The visible spots on this control array represent fluorescein-containing marker peptoids spotted as navigation aids. Hundreds of features on each array captured labeled protein at a level detectable above the background under these conditions, as can be better seen in the expanded regions shown in Fig. 2. These data confirm that a library of 7,680 peptoids is sufficiently rich in protein ligands to support a fingerprinting application.

Even with the naked eye, it can be seen that the binding patterns are distinct on each array (Fig. 2). To better illustrate and quantify this distinction, the fluorescent intensity at each feature (from 1 to 7,680) was quantified and, after subtraction of the background intensity (see *Materials and Methods*), was displayed as a color-coded bar graph with each line representing the background-subtracted intensity of a single feature using the GENESPRING software package (Fig. 3). Examination of these bar codes revealed that each protein provided a unique pattern of binding to the array, although, surprisingly, the Ub and GST data



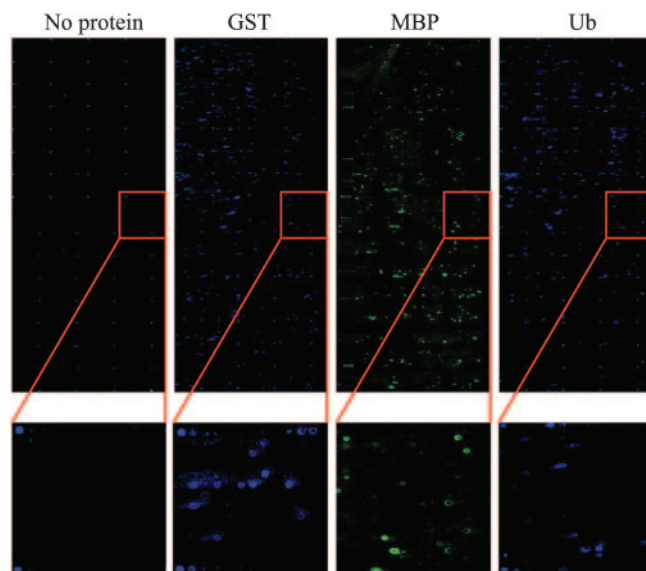
**Fig. 1.** The peptoid library used in this study. (A) General structure of the peptoid library synthesized on polystyrene macrobeads. (B) Amines used in the synthesis of the library.

sets superficially resembled one another, whereas that produced in the MBP hybridization was quite different. Focusing on almost any smaller set of features clearly illustrates the differences between the Ub and GST binding patterns (Fig. 3, bottom).

To address the critical issue of reproducibility and the uniqueness of the fingerprint in another way, all of the data were visualized as a series of scatter plots. When the two independent hybridizations for a given protein were compared, a high degree of correlation was obtained (Fig. 4A) ( $R = 0.97$  for GST,  $0.96$  for MBP, and  $0.97$  for Ub). An even better correlation was obtained if only the higher-intensity features were considered, as one would expect, because the data closest to the background tend to be the “noisiest” (see Fig. 6, which is published as supporting information on the PNAS web site) ( $R = 0.99$  for all proteins). To better display how many peptoids bind promiscuously to all of the proteins and how many are specific, all of the features that display an intensity value of  $\geq 10$ -fold above background were incorporated into a Venn diagram (see Fig. 7, which is published as supporting information on the PNAS web site); 191, 126, and 61 peptoids were found to be quite specific for GST, MBP, and Ub, respectively. We conclude that the technique is sufficiently reproducible and has a high enough degree of specificity to serve as a platform for protein fingerprinting.

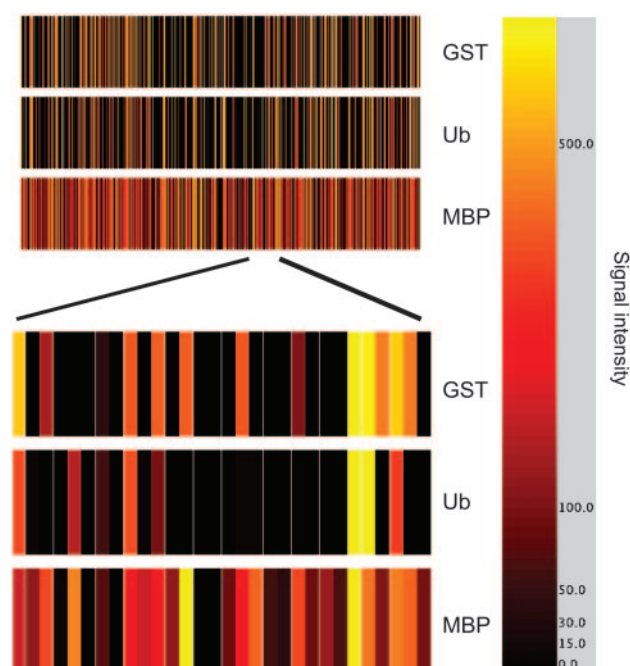
When the data sets obtained from two different protein hybridizations were compared, the correlation was far lower (see Fig. 4B) ( $R = 0.56$  for GST vs. Ub,  $0.28$  for GST vs. MBP, and  $0.25$  for Ub vs. MBP). Based on these data, we conclude that the three different proteins used in this study exhibit highly reproducible and distinctive patterns when hybridized to the peptoid microarray.

**An Antibody Sandwich Assay Allows Detection of the Binding Pattern of an Unlabeled Protein.** The proof of principle experiments presented above used purified proteins that had been labeled covalently and then mixed with the bacterial extract. Obviously, this type of protocol could not have been used to detect a native



**Fig. 2.** Protein profiling using peptoid microarrays. Images were obtained by incubating fluorescently labeled GST, MBP, or Ub to a peptoid microarray containing 7,680 different compounds. For comparison, a fourth array to which no protein has been hybridized is shown on the far left. In this image, the fluorescent spots are labeled peptoids that have been attached to the slide as navigation aids. These images were obtained by scanning the arrays with a standard commercial array scanner used for DNA microarray analysis after hybridization and washing. The *insets* provide a magnified view to illustrate the signal-to-noise ratio obtained in the hybridization of 500 nM labeled protein to the array.

protein. An alternative method that would achieve this goal would be to hybridize to the array a mixture of native proteins, then probe the chip with an antibody raised against a particular protein of interest, followed by a labeled secondary antibody (13). Therefore, we performed an experiment in which unlabeled GST was doped into a 100-fold excess of bacterial proteins and hybridized to the array. To visualize the GST binding pattern selectively, the array was then probed with an anti-GST antibody followed by a secondary antibody labeled with Alexa Fluor 488. As a control, a second array was treated in the same way, except that the GST was omitted. The fluorescent intensities at each feature then were measured, and the values obtained in the “–GST” experiment were subtracted from those in the “+GST” experiment to provide a corrected data set from which the binding patterns of the antibodies had been subtracted. Again, two completely independent experiments were conducted, and these experiments were highly reproducible (Fig. 5*A*) ( $R = 0.96$ ). A comparison of this data set with that obtained with the chemically labeled GST (Fig. 5*B*) revealed a similar, but not identical, protein fingerprint. The correlation coefficient was  $R = 0.84$ . Although errors introduced in the subtraction of the “antibody-only” control data could play some role in this reduced correlation, the similarity of this scatter plot to those generated by using data sets from the hybridization of different proteins (Fig. 4*B*) suggests an alternative explanation. The dramatically “off-diagonal” features clearly cluster into two groups, one of which registers a much higher signal intensity in the antibody-using experiment whereas the other provides a much higher signal in the experiment by using chemically labeled GST. This result suggests that the features in these two groups discriminate significantly between the labeled and native forms of the GST. In other words, the simplest explanation of the data are that native and fluorescein-labeled GST appear to the array as related, but different, proteins, although detailed binding studies of some of these peptoids to each form of the protein will

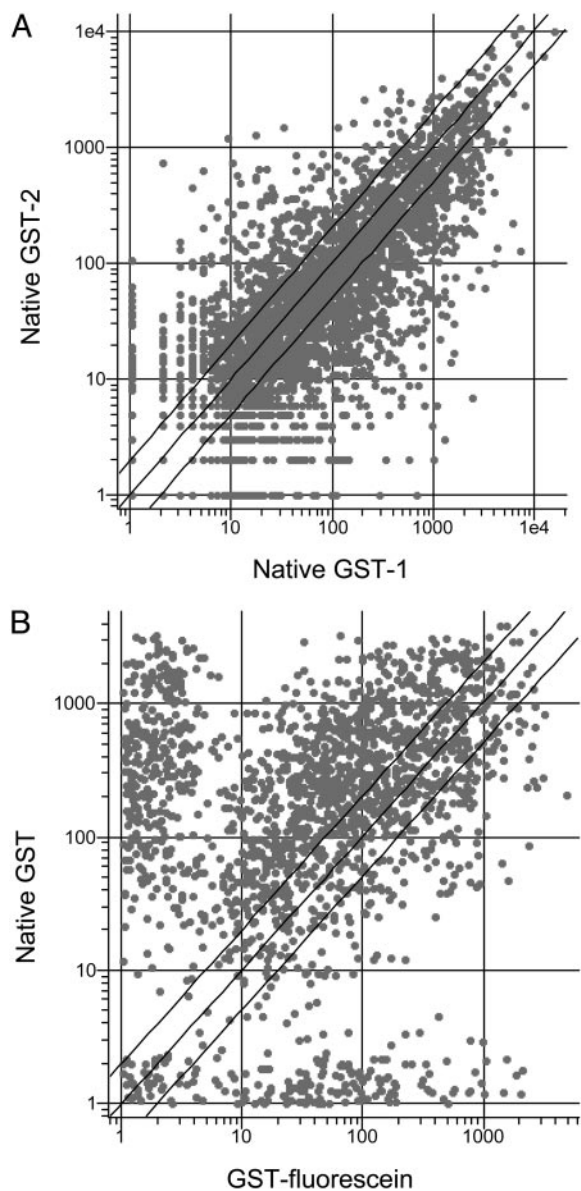


**Fig. 3.** A protein fingerprint. The intensities at each of the 7,680 features on the arrays shown in Fig. 2 (and the replicate experiments) were assigned a color code (shown on the right) and displayed as a bar code to allow visualization of the binding patterns. A portion of this bar code is expanded to show clearly the differences between the binding patterns.

be required to substantiate this hypothesis. In any case, this experiment shows clearly that a distinctive fingerprint of a native protein can be detected on the peptoid microarray by using an antibody-based sandwich assay.

**Sensitivity of the Peptoid Microarrays.** The experiments discussed above used the target protein at a concentration of 500 nM in the presence of a 100-fold excess of bacterial proteins. To determine whether more dilute proteins could be detected and fingerprinted by using the peptoid microarray, fluorescein-labeled GST solutions of 100 and 10 nM, again in a solution containing a 100-fold excess of bacterial proteins, were prepared and hybridized to the peptoid arrays. After washing, the arrays were scanned, and the intensity observed at each feature was quantified as described above. The data then were compared. This comparison revealed an excellent correlation (see Fig. 8, which is published as supporting information on the PNAS web site) between the data sets obtained at the different GST concentrations, with correlation coefficients of  $R = 0.95$ ,  $R = 0.97$ , and  $R = 0.95$  for comparison of the 500 vs. 100 nM, 100 vs. 10 nM, and 500 vs. 10 nM data sets, respectively. In other words, very similar protein fingerprints could be discerned at each of the protein concentrations examined. In another experiment, GST (500 nM) was serially diluted into an *E. coli* lysate to create solutions with a 500-fold (GST = 0.5% of total protein) or 1,000-fold (GST = 0.1% of total protein) excess of bacterial proteins. After hybridization to the arrays and detection as described above, the data sets were compared on a scatter plot (see Fig. 9, which is published as supporting information on the PNAS web site). The  $R$  values derived from these scatter plots were 0.96 for 100-fold vs. 500-fold, 0.97 for 100-fold vs. 1,000-fold, and 0.97 for 500 vs. 1,000-fold. In other words, these data show that a recognizable and consistent fingerprint is observed for a protein present in an extract at different levels and is not sensitive to the abundance of that protein. Another issue that can be addressed with these





**Fig. 5.** Fingerprinting native GST by using an antibody sandwich assay (see text for details). (A) Scatter plot comparing the two completely independent experiments using native GST. (B) Scatter plot comparing the data obtained with native GST and fluorescein-labeled GST (Fig. 2).

as one would expect based on the absolute protein concentration. These data suggest that we are approaching the sensitivity limit of the array somewhere between 100 and 10 nM fluorescein-labeled GST.

### Discussion

Both biological (32) and artificial (33, 34) sensors containing a large number of low to modest affinity receptors can differentiate between different molecules by distinguishing quantitative patterns of binding of an analyte to the different receptors. We have combined this concept with peptoid microarrays to create a powerful platform for the identification of proteins. Proof of principle experiments using chemically labeled proteins (Figs. 2–4) demonstrated that different proteins exhibit a different pattern of binding to the thousands of peptoids on the array. We also demonstrated that a protein fingerprint could be discerned when a native protein (GST) in a solution containing a large

excess of bacterial proteins was hybridized to the array (Fig. 5). Selective visualization of the GST binding pattern even in the presence of a large excess of other proteins was achieved by conducting a second hybridization with an anti-GST antibody. The pattern due to antibody binding was determined in a control experiment and subtracted from these data. This finding is critical in further development of these arrays as bioanalytical tools, because selective chemical labeling of a protein of interest in a complex mixture can rarely be achieved. A small amount of analytical work has been done to distinguish purified proteins with simpler arrays in other configurations (14–16). However, this work provides a previously undescribed demonstration that a glass slide-based SMM can “fingerprint” a native protein in a complex mixture. Because the proteins used in this study, GST, MBP, and Ub, are not related to one another and because the molecules in the library used to make the array (Fig. 1) were not biased in any way, we presume that the peptoid microarray is capable of fingerprinting almost any protein, although substantiation of this hypothesis obviously will require further study.

This array-based fingerprinting technique proved to be fairly sensitive, with a recognizable fingerprint produced by using a GST solution of only 10 nM in the presence of a 100-fold excess of bacterial proteins (Fig. 8). This issue was a major question to be addressed going into this study, because none of the peptoids were anticipated to be high-affinity ligands for GST or any other protein. Our experience with peptoid library screening suggests that even the best ligands in an unbiased library are likely to form complexes with  $K_D$  values in the low-micromolar range (27). However, the linearity of the signal intensities deviated from the expected values between 100 and 10 nM for labeled GST, indicating that the sensitivity limit for truly quantitative work was somewhere between 100 and 10 nM GST in this particular case. Nonetheless, this limit would be sufficient to fingerprint many highly and modestly abundant proteins in serum or cell extracts. Furthermore, this limit is not necessarily a general sensitivity limit of the array, because this limit will depend on the affinities of the protein analyzed for the peptoids on the array. For example, we have evidence that antibodies and certain other proteins bind with unusually high affinity to many peptoids (M.M.R. and T.K., unpublished results) on the microarray. It is also possible that much higher sensitivity could be achieved by using more intensely fluorescent dyes such as quantum dots or with detection schemes that employ some sort of signal amplification.

The experiments reported here are model studies and break no new biological ground. The presence and levels of a protein in a complex mixture could have been determined with SDS/PAGE and Western blotting. Nonetheless, these experiments demonstrate two critical points that were far from obvious before these studies. The first is that the patterns obtained are highly reproducible (Fig. 4) in completely independent experiments. The second is that even an array with a relatively modest number of features (<8,000) is capable of a high level of discrimination between two proteins. In other words, the collection of peptoids is sufficiently rich in specific ligands for any given protein that a unique pattern is produced. Indeed, it was quite interesting that a comparison between the binding patterns obtained with native GST and fluorescein-labeled GST were noticeably different. This finding suggests that the distinguishing power of the array may be sufficient to derive different patterns for the same protein with different posttranslational modifications. This application would be an attractive use of these arrays, because antibodies that recognize specific forms of a given protein are often difficult to obtain. To read the array pattern, only a standard antibody that recognized any form of the protein would be necessary.

Another potential application of this array technology is protein profiling for diagnostic purposes. By using other analytical platforms, particularly surface-enhanced laser desorption ionization mass spectrometry, several investigators have argued

