Combinatorial Chemistry

DNA Microarrays as Decoding Tools in Combinatorial Chemistry and Chemical Biology**

*Marina Lovrinovic and Christof M. Niemeyer**

Keywords:
combinatorial chemistry · DNA · immobilization · microarrays · proteins

Over the last few years, laterally microstructured arrays of DNA[1] and protein probes[2–4] have been developed as tools for high-throughput experimentation in biomedical research. These devices have the advantage of spatial addressability of the probes and require only small amounts of analyte. The development of protein microarrays is still obstructed by the intrinsic instability of many proteins, which leads to the loss of functionality during their automated deposition onto chemically activated surfaces. However, the application of DNA arrays is almost routine nowadays, as they are chemically stable and are often available off-the-shelf through academic and commercial suppliers. Hence, the application of DNA microarrays toward investigations of gene expression by quantitation of mRNA levels under variable environmental conditions offers a well-established approach in the fundamental and industrial research of biological systems.[5] In addition to these mainstream applications of DNA chips in genomics, alternative uses of these devices as tools for decoding combinatorial libraries in chemistry and chemical biology are currently emerging.

Figure 1. DNA-directed immobilization of DNA–protein conjugates with surface-bound capture oligomers of a DNA microarray.

As an example, DNA arrays can be used as programmed matrices for the site-selective immobilization of DNA oligomer-tagged proteins and small molecules (Figure 1). This approach, termed DNA-directed immobilization (DDI), offers a chemically mild procedure for the site-selective, highly parallel, and reversible attachment of protein libraries to solid supports.[6] As a particular advantage, the proteins retain their biological activity because they are attached to the surface through a short double-stranded DNA linker rather than being directly fixed at the surface by multiple covalent or noncovalent contacts, which may restrict their conformational freedom and could lead to (partial) denaturation of tertiary structure.[6b] The advantage of the DDI method for the fabrication of protein microarrays has been demonstrated for antibodies,[6c,f,g] receptors[6d] and enzymes.[6b,i]

Winssinger and co-workers have applied the principle of nucleic acid-directed immobilization to the screening of small-molecule libraries for compounds that bind to protein targets.[7–10] The technology is based on peptide nucleic acid (PNA)-encoded small-molecule probes, which are accessible through solid-phase combinatorial syntheses (Figure 2). The small-molecule portion of the probe is designed to bind to proteins in a mechanism-dependent manner, therefore discriminating between active proteins and proteins that are present in a latent or inactive form. The PNA portion of the probes functions as a code for the synthetic history of the small-molecule portion, and therefore permits deconvolution of the probe through hybridization at an oligonucleotide microarray. Array-based deconvolution allows the simultaneous analysis of multiple probes in a miniaturized format and has the potential of screening up to 400 000 probes in a solution volume of less than 300 μL. The initial validation of this approach was carried out with specific inhibitor probes whose design was based on the peptide substrate specificity of known proteases, such as cathepsins[7] and caspase.[8]

Recently, this methodology was applied to the discovery of new proteolytic activities in dust mite extracts.[10] House dust mites are a major source of allergens and contribute to the increased incidence of allergenic diseases such as bronchial asthma. Some of these allergens have protease activity, which has long been known to produce an allergic response. To elucidate the cellular mechanisms behind the development

[*] Dipl.-Chem. M. Lovrinovic, Prof. Dr. C. M. Niemeyer
Universität Dortmund
Fachbereich Chemie
Biologisch-Chemische Mikrostrukturtechnik
Otto-Hahn Strasse 6, 44227 Dortmund (Germany)
Fax: (+49) 231-755-7082
E-mail: christof.niemeyer@uni-dortmund.de

[**] We acknowledge financial support of our work by Deutsche Forschungsgemeinschaft (DFG) and by the research program "Molecular Basics of Biosciences" of the University of Dortmund.

DOI: 10.1002/anie.200500645

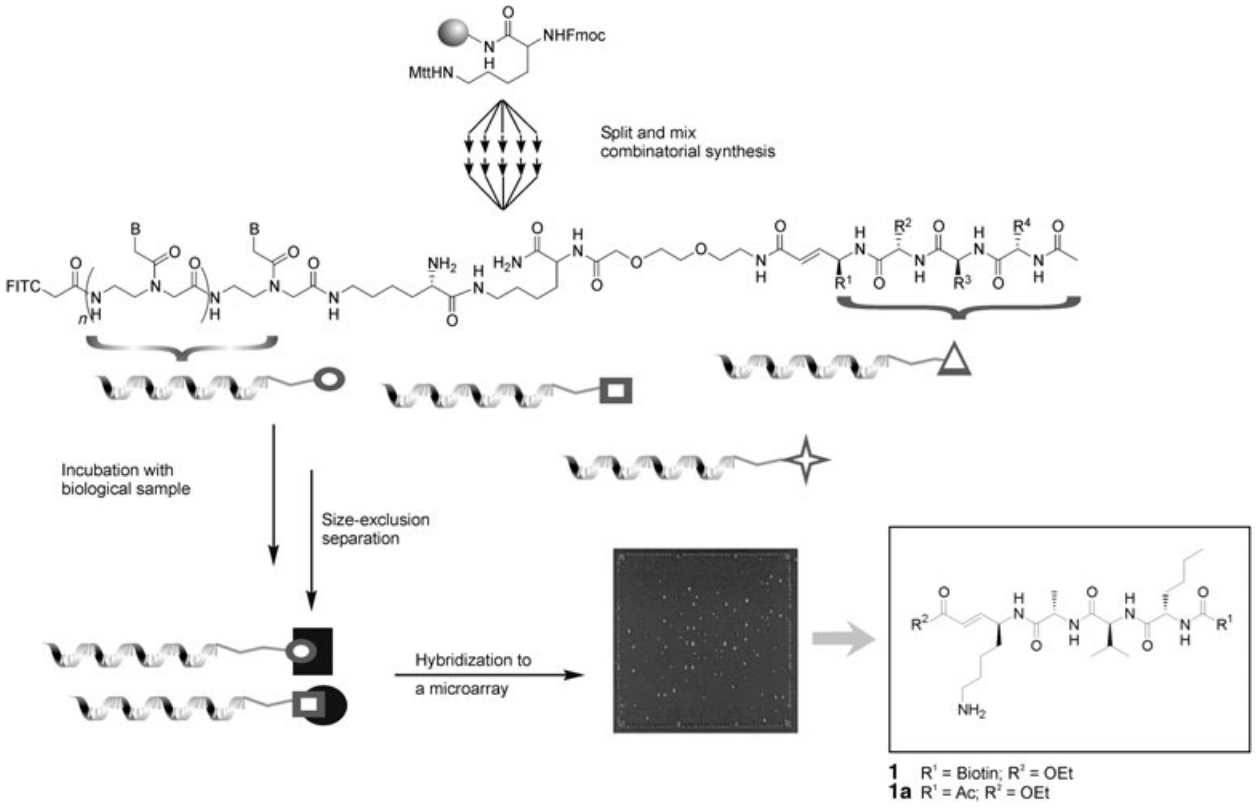

Figure 2. Synthesis and deconvolution of a PNA-encoded library (adapted from Ref. [10]). The fluorescein isothiocyanate (FITC) group enables fluorescence detection of the binding of all members of the library to the spatially encoded DNA array. Mtt = 4-methyltrityl; Fmoc = 9-fluorenylmethoxycarbonyl.

of allergies, it is essential to profile the activity of the proteases involved.

To this end, a 4000-member PNA-encoded tetrapeptide inhibitor library designed to target cysteine proteases was synthesized by combinatorial split-and-mix synthesis (Figure 2). The library was incubated with dust mite lysate, and the unbound probes were subsequently separated by spin filtration through a 30-kDa molecular weight cut-off filter. The retained samples that contained the protein-bound probes were hybridized to an oligonucleotide array containing capture oligomers for all members of the PNA-encoded inhibitor library. Fluorescence imaging allowed the identification of tetrapeptides that had bound to protease targets. Active inhibitors contained lysine and norleucine at position P1 and alanine at position P2 (Figure 2). Positions P3 and P4 of the inhibitor appeared to be less important for target binding; the most striking feature was that compounds with histidine, phenylalanine, or proline at position P3 were completely inactive as inhibitors. The probe with the highest intensity on the microarray had the inhibitor sequence Nle-Val-Ala-Lys (P1–P4). This probe (compound **1** in Figure 2) was re-synthesized with a biotinylated linker to isolate and identify the interacting proteins. Following incubation of **1** with the dust mite lysate, the probe and proteins attached to it were captured with streptavidin and the proteins were sequenced my mass spectrometry.

Two major proteins, Derp1 and Derp10 were identified. Derp1 is a 25-kDa protein homologous to the papain family of cysteine proteases, and Derp10 is a 33-kDa protein that is homologous to tropomyosin. Isolated Derp1 was then prepared from house mite fecal pellets by immunoaffinity chromatography by using an immobilized monoclonal antibody, and its substrate specificity was profiled with the tetrapeptide substrate library in a positional scanning format. The results confirmed that the major substrate specificity determinant for Derp1 is in the P2 position for the alanine residue. Moreover, a slight preference of Derp1 for basic amino acids in positions P1 and P3 was observed, as well as a preference for aliphatic amino acids such as isoleucine, proline, valine, leucine, and norleucine in position P4. Finally, the authors demonstrated the phenotypic relevance of Derp1 function in allergy progression by the inhibition of cleavage of CD25 (the α chain of the interleukin 2 cell-surface receptor) from T-cells with the tetrapeptide inhibitor **1a**.[9]

In a related approach, Melkko et al. used DNA-encoded self-assembling chemical (ESAC) libraries for the facile identification of small molecules that bind macromolecular targets.[11] In this case, libraries of organic molecules linked to individual DNA oligomers were assembled with DNA strands that contain complementary sequence stretches, thereby providing a code associated with each organic moiety. After incubation of the ESAC libraries with a protein of interest attached to a solid support and removal of unbound mate-

rials, the coding sequences of the selected compounds were decoded with a DNA microarray. Repetition of the selection cycle led to the affinity maturation of compounds that bind human serum albumin and bovine carbonic anhydrase with dissociation constants in the nanomolar range.[11]

The above examples impressively demonstrate how DNA microarrays can be used for the decoding of nucleic acid encoded small-molecule libraries. Hu et al. have recently applied DNA microarrays to the high-throughput screening of expressed enzyme libraries (Figure 3).[12] Their strategy, termed "Expression Display", is based on the ribosome-display technique, which allows the in-vitro expression of proteins in a cell-free translation reaction. The resulting polypeptides are thereby tagged with their own coding mRNAs.[13] The proteins were expressed from a cDNA library containing 384 different open-reading frames (ORFs) from yeast, four of which encoded known protein tyrosine phosphatases (PTPs), through in-vitro transcription and translation (steps 1 and 2 in Figure 3). The resulting library of ribosomal complexes was incubated with the activity-based probe **2** that had been immobilized on streptavidin magnetic beads, and which specifically and irreversibly binds PTPs (step 3). The mRNA was eluted and reverse-transcribed to generate the corresponding fluorescently labeled cDNA. The resulting cDNA library was then hybridized to the decoding DNA microarray (containing 384 cDNAs) for the parallel identification of functional PTPs. Positives were identified by the location of fluorescent spots on the array (Figure 3). Indeed, microarray analysis revealed the selective detection of the four PTPs from the library, which also contained ORFs for other enzyme classes (proteases, kinases, oxidoreductases) and nonenzyme proteins. Hence, this example of combining activity-based probes and array-based high-throughput identification suggests that it should be possible to screen thousands of proteins, all in a single reaction without the need for parallel cloning, expression, purification, and characterization of individual proteins.[12]

DNA microarrays have also been used recently in synthetic organic chemistry as a selection tool for the discovery of a new type of chemical reaction: the Pd-catalyzed carbon-carbon bond-forming reaction that generates an enone from an alkyne and alkene (Figure 4).[14] To this end, Liu and co-workers, whose work is focused on the development of DNA-templated organic syntheses,[15] prepared two pools of DNA-linked small-molecule substrates, each of which contains 12 different potentially reactive functional groups covalently linked to either the 5′- (pool A) or 3′-end (pool B) of an oligonucleotide. Pool A oligomers contain a "coding region" that uniquely identifies the substrate as well as one of 12 different "annealing regions". Pool B oligomers also contain a "coding region" which encodes the substrate and complements one of the 12 annealing regions in pool A (Figure 4a).

When pools A and B are combined in a single aqueous solution at nanomolar concentrations, specific Watson–Crick base-pairing assembles the compounds into 12×12 discrete pairs of substrates, which experience effective concentrations in the millimolar range. Substrates linked to noncomplementary oligomers experience nanomolar solution concentrations and hence do not react with each other at a significant rate. To allow the separation of reactive pairs, each substrate of pool B was covalently linked to its corresponding oligomer by a linker containing a biotin group and a cleavable disulfide bond (Figure 4b). After incubation under a set of chosen reaction conditions, the disulfide bonds were cleaved and the biotin group remained covalently linked only to pool A compounds after a bond-forming reaction between pool A and pool B substrates had occurred. Avidin affinity selection of the resulting solution separated biotinylated from non-biotinylated compounds. Reactive substrate pairs were then amplified by PCR. Because PCR amplification is extremely sensitive, femtomole quantities of substrate were sufficient for the entire reaction discovery process.

As both pools each contained 12 substrates, 144 different DNA-linked substrates of pool A were prepared to encode all heterocoupling combinations. Furthermore, DNA-linked substrates

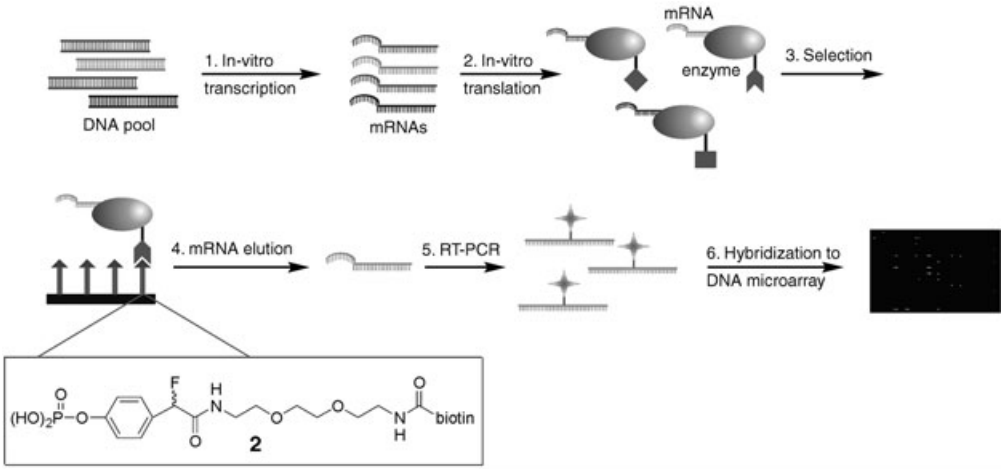

Figure 3. Expression display of an enzyme library (RT = reverse transcription). Also shown is the structure of the protein tyrosine phosphatase (PTP)-specific activity-dependent probe **2** (adapted from Ref. [12]).

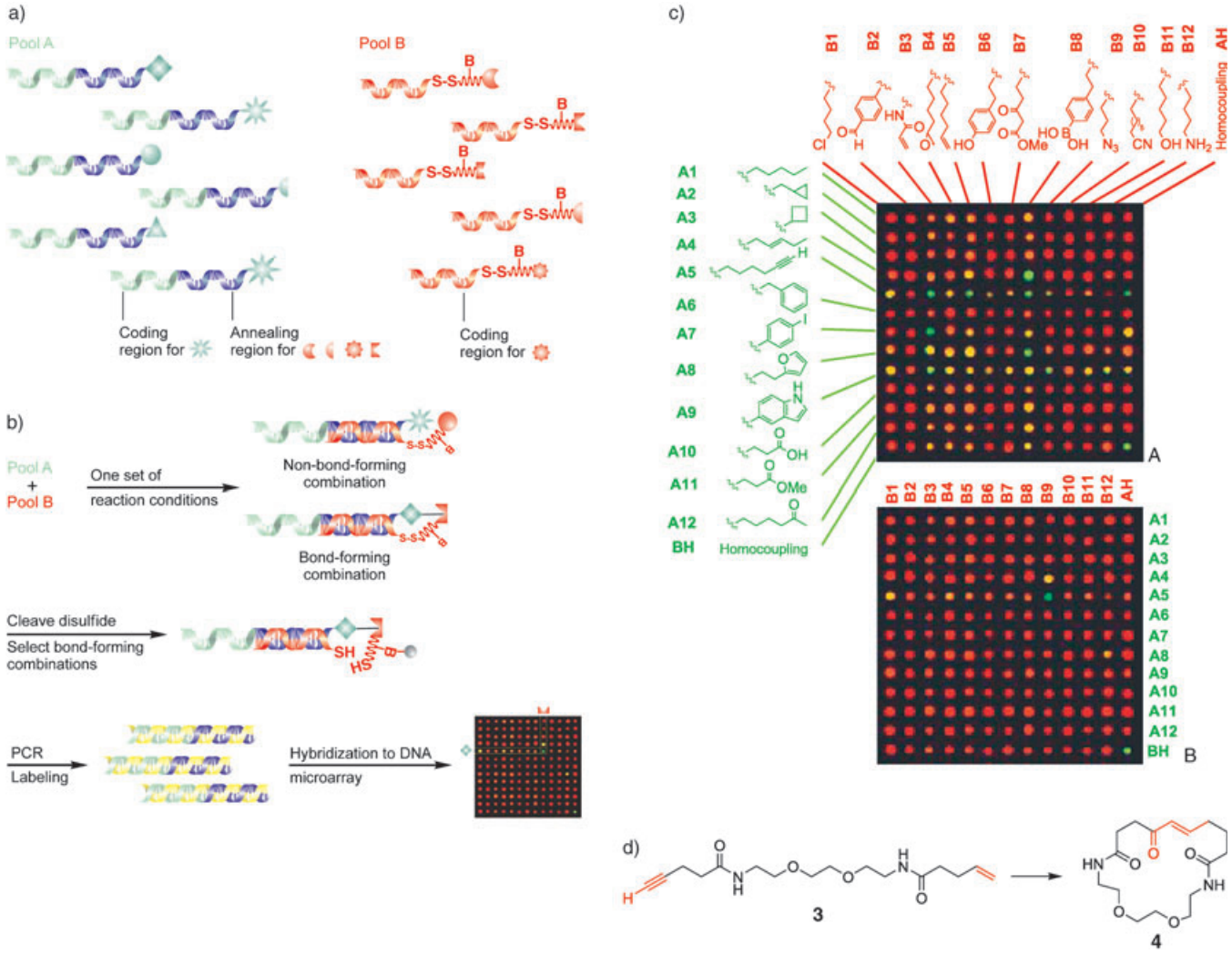

Figure 4. Discovery of a new type of chemical reaction that takes advantage of DNA-templated reactions (adapted from Ref. [14], array data reproduced with kind permission).

that encoded for the homocoupling of any of the 24 different substrates were prepared, bringing the total number of unique substrate combinations to 168. Although this approach required the preparation of a large number of DNA-linked substrates, these are typically synthesized at the nanomole-scale, and therefore, provide sufficient material for more than 1000 reaction discovery processes.

To enable a semiquantitative analysis of bond-forming efficiency, the mixture of the two pools A and B was amplified twice by PCR. Prior to bond formation and selection, PCR was carried out with Cy5-labeled primers while the post-selection mixture was amplified with Cy3-labelled primers. Equal amounts of the two PCR products were combined and hybridized to the DNA array containing all possible sequence combinations. The ratios of Cy3 (green) to Cy5 (red) fluorescence was determined for all array locations, and spots with green/red fluorescence ratios above 1.5 were considered positive. A pre-quantified internal standard (bottom right corner of the array in Figure 4c) was used as a positive control and as a reference to compare different arrays.

Initial validation tests were carried out with the known reaction between an alkyne (A5) and an azide (B9) in the presence of Cu^{I}. The observation of a single green spot (green/red ratio = 8.5) indicated the feasibility of the reaction discovery method (bottom array in Figure 4c). Similarly, the selection for bond formation after treatment of the pools with EDC/NHS (EDC = 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide, NHS = *N*-hydroxysuccinimide) led to the formation of a single positive spot, indicating a reaction between the amine (B12) and the carboxylic acid (A10) compound (green/red ratio = 15.6). After the successful "rediscovery" of known bond-forming reactions, the authors then examined the reactivity of the library in the presence of Pd^{II}. The first reaction, carried out for one hour at 37 °C, led to five strong positives (A7 + B3, A5 + B3, A4 + B8, A5 + B5, and A5 homocoupling), in addition to five weaker positives (A9 + B3, A8 + B3, A8 + B8, A5 + B8, and A5 + B9; top array in Figure 4c). The 10 putative bond-forming reactions were then examined in separate DNA-templated reactions, and gel electrophoretic analysis indicated that all five strong positives and three of the weak positives indeed

corresponded to authentic DNA-templated reactions, whereas two weak positives (A9 + B3 and A5 + B9) showed little or no product formation.

Repetition of the selection experiment under more stringent conditions (20 min at 25 °C) decreased the number of positive signals and suggested that Pd^{II}-mediated carbon-carbon bond formation between the terminal alkyne (A5) and terminal alkene (B5) proceeds efficiently to generate an enone product. For the detailed investigation of this reaction in a non-DNA-templated version, small-molecule substrate **3** was synthesized and subjected to intramolecular cyclization to yield enone **4** in the presence of Pd^{II}. It was observed that this macrocyclization occurs at the milligram scale with Na_2PdCl_4 (5 mol %) in the presence of 1 equivalent $CuCl_2$ in various solvents with yields larger than 90 %. Hence, the discovery of this alkyne–alkene coupling reaction indicates the value of searching a large number of substrate combinations for unexpected reactions. The authors anticipate that similar schemes will lead to the discovery of additional bond-forming reactions between simple and relatively unreactive functional groups.[14]

In conclusion, these examples demonstrate the versatility of DNA microarrays as tools for the decoding of complex libraries of DNA-tagged small-molecules[16] as well as biomolecular compounds. These developments have been made feasible through the extraordinary physicochemical stability of nucleic acids, their availability by solid-phase synthesis, and the resulting steady progress in DNA array technologies over the past 15 years. Although DNA chips have since become routine tools, many problems have yet to be solved, for instance, the accurate prediction and quantitation of hybridization efficiencies of individual probes. Far more challenging, however, is the extension of applications similar to those described herein to protein microarrays. This may allow even deeper insight into biological systems on a proteome-wide scale. Owing to the exquisite and delicate architecture of nature's universal tools, however, steps toward this goal will surely constitute a fascinating research area for creative chemical biologists in the near future.

Published online: ■■ ■■, 2005

[1] For reviews on DNA microarray technology, see: a) U. R. Mueller, D. V. Nicolau, *Microarray Technology and Its Applications*, Springer, Berlin, **2005**; b) R. L. Stears, T. Martinsky, M. Schena, *Nat. Med.* **2003**, *9*, 140 – 145; c) M. C. Pirrung, *Angew. Chem.* **2002**, *114*, 1326 – 1341; *Angew. Chem. Int. Ed.* **2002**, *41*, 1276; d) C. M. Niemeyer, D. Blohm, *Angew. Chem.* **1999**, *111*, 3039 – 3043; *Angew. Chem. Int. Ed.* **1999**, *38*, 2865 – 2869; e) E. M. Southern, K. Mir, M. Shchepinov, *Nat. Genet.* **1999**, *21*, 5 – 9.

[2] M. F. Templin, D. Stoll, M. Schrenk, P. C. Traub, C. F. Vohringer, T. O. Joos, *Trends Biotechnol.* **2002**, *20*, 160 – 166.

[3] D. S. Wilson, S. Nock, *Angew. Chem.* **2003**, *115*, 510 – 517; *Angew. Chem. Int. Ed.* **2003**, *42*, 494 – 500.

[4] D. S. Yeo, R. C. Panicker, L. P. Tan, S. Q. Yao, *Comb. Chem. High Throughput Screening* **2004**, *7*, 213 – 221.

[5] DNA microarray technology is a well-established tool in chemical genomics for the profiling of whole-genome transcripts for the identification and validation of targets of small molecules such as kinase inhibitors. For examples, see: a) C. Kung, K. M. Shokat, *ChemBioChem* **2005**, *6*, 523 – 526; b) D. W. Provance, Jr., C. R. Gourley, C. M. Silan, L. C. Cameron, K. M. Shokat, J. R. Goldenring, K. Shah, P. G. Gillespie, J. A. Mercer, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 1868 – 1873; c) K. M. Specht, K. M. Shokat, *Curr. Opin. Cell Biol.* **2002**, *14*, 155 – 159; d) A. S. Carroll, A. C. Bishop, J. L. DeRisi, K. M. Shokat, E. K. O'Shea, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 12578 – 12583.

[6] a) C. M. Niemeyer, T. Sano, C. L. Smith, C. R. Cantor, *Nucleic Acids Res.* **1994**, *22*, 5530 – 5539; b) C. M. Niemeyer, L. Boldt, B. Ceyhan, D. Blohm, *Anal. Biochem.* **1999**, *268*, 54 – 63; c) C. M. Niemeyer, B. Ceyhan, *Angew. Chem.* **2001**, *113*, 3798 – 3801; *Angew. Chem. Int. Ed.* **2001**, *40*, 3685 – 3688, ; d) M. Lovrinovic, R. Seidel, R. Wacker, H. Schroeder, O. Seitz, M. Engelhard, R. Goody, C. M. Niemeyer, *Chem. Commun.* **2003**, 822 – 823; e) U. Feldkamp, R. Wacker, W. Banzhaf, C. M. Niemeyer, *ChemPhysChem* **2004**, *5*, 367 – 372; f) R. Wacker, C. M. Niemeyer, *ChemBioChem* **2004**, *5*, 453 – 459; g) R. Wacker, H. Schroeder, C. M. Niemeyer, *Anal. Biochem.* **2004**, *330*, 281 – 287; h) F. Kukolka, C. M. Niemeyer, *Org. Biomol. Chem.* **2004**, *2*, 2203 – 2206; i) L. Fruk, C. M. Niemeyer, *Angew. Chem.* **2005**, *117*, DOI: 10.1002/anie.200462567; *Angew. Chem. Int. Ed.* **2005**, *44*, DOI: 10.1002/ange.200462567.

[7] N. Winssinger, J. L. Harris, B. J. Backes, P. G. Schultz, *Angew. Chem.* **2001**, *113*, 3254 – 3258; *Angew. Chem. Int. Ed.* **2001**, *40*, 3152 – 3155.

[8] N. Winssinger, S. Ficarro, P. G. Schultz, J. L. Harris, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11139 – 11144.

[9] N. Winssinger, R. Damoiseaux, D. C. Tully, B. H. Geierstanger, K. Burdick, J. L. Harris, *Chem. Biol.* **2004**, *11*, 1351 – 1360.

[10] J. Harris, D. E. Mason, J. Li, K. W. Burdick, B. J. Backes, T. Chen, A. Shipway, G. Van Heeke, L. Gough, A. Ghaemmaghami, F. Shakib, F. Debaene, N. Winssinger, *Chem. Biol.* **2004**, *11*, 1361 – 1372.

[11] S. Melkko, J. Scheuermann, C. E. Dumelin, D. Neri, *Nat. Biotechnol.* **2004**, *22*, 568 – 574.

[12] Y. Hu, G. Y. Chen, S. Q. Yao, *Angew. Chem.* **2005**, *117*, 1072 – 1077; *Angew. Chem. Int. Ed.* **2005**, *44*, 1048 – 1053.

[13] A related technology is based on the in-vitro translation of mRNA derivatized with a puromycin group at its 3′-end. The peptidyl-acceptor antibiotic puromycin covalently couples the mRNA with the encoded polypeptide chain formed at the ribosome, which results in the specific conjugation of the informative (mRNA) with the functional (polypeptide) moiety. Such covalent nucleic acid–protein conjugates have potential for the fabrication of protein microarrays: P. A. Lohse, M. C. Wright, *Curr. Opin. Drug Discovery Dev.* **2001**, *4*, 198 – 204; M. Kurz, K. Gu, A. Al-Gawari, P. A. Lohse, *ChemBioChem* **2001**, *2*, 666 – 672.

[14] M. W. Kanan, M. M. Rozenman, K. Sakurai, T. M. Snyder, D. R. Liu, *Nature* **2004**, *431*, 545 – 549.

[15] X. Li, D. R. Liu, *Angew. Chem.* **2004**, *116*, 4956 – 4979; *Angew. Chem. Int. Ed.* **2004**, *43*, 4848 – 4870.

[16] J. J. Diaz-Mochon, L. Bialy, L. Keinicke, M. Bradley, *Chem. Commun.* **2005**, 1384 – 1386.

