Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry Letters
Bioorganic & Medicinal Chemistry Letters 15 (2005) 2447–2451

Site-specific immobilization of proteins in a microarray using intein-mediated protein splicing

Aparna Girish,[a] Hongyan Sun,[b] Dawn S. Y. Yeo,[a] Grace Y. J. Chen,[a,b] Teck-Khiang Chua[a] and Shao Q. Yao[a,b,*]

[a]*Department of Biological Sciences, National University of Singapore, 3 Science Drive 3, Singapore 117543, Singapore*
[b]*Department of Chemistry, National University of Singapore, 3 Science Drive 3, Singapore 117543, Singapore*

Received 17 February 2005; revised 17 March 2005; accepted 22 March 2005
Available online 12 April 2005

Abstract—One of the critical issues in the generation of a protein microarray lies in the choice of immobilization strategies, which ensure proteins are adhered to the glass surface while properly retaining their native biological activities. Herein, we report a bacterium-based, intein-mediated strategy to generate N-terminal cysteine-containing proteins which are then chemoselectively immobilized to a thioester-functionalized glass slide to generate the corresponding protein microarray. We also showed preliminary data of the strategy in a yeast host system.

In the post-genomic era, the primary aim for researchers is to fully characterize and understand all proteins encoded by the genome. In order to accomplish this, new proteomic technologies capable of high-throughput analysis and characterization of protein en masse have to be developed.[1] One of the emerging platforms—the protein microarray, has the potential to rapidly profile the entire proteome, thus is capable of revealing novel protein functions and mapping out comprehensive protein interaction networks of an organism.[2] An impending, and often overlooked, issue in protein microarray technologies lies in the successful development of robust strategies which allow efficient immobilization of proteins onto glass surfaces while maintaining their native biological functions.[3] This calls for the development of novel protein immobilization methods, which ensure that proteins are uniformly and stably attached to the solid surface.[4] We previously developed intein-mediated approaches to site-specifically biotinylate proteins both in vitro and in vivo, and subsequently immobilized them onto avidin-coated slides to generate the corresponding protein array.[5] We showed that proteins immobilized this way retained their native biological properties on the glass surface, making them suitable to be screened under a variety of assay conditions.[6] Recently, we showed that the intein-mediated protein biotinylation efficiency, both in vitro and in vivo, could be drastically improved.[5c] We now report, the use of the chemoselective chemical ligation reaction to site-specifically immobilize N-terminal cysteine-containing proteins to create the corresponding protein array. The key to our current strategy is the use of intein-mediated protein splicing for efficient expression of N-terminal cysteine-containing proteins.

Previous work in our group had established the specific immobilization of N-terminal cysteine-containing peptides onto a thioester-functionalized glass slide, which was subsequently used for high-throughput screening of kinase activities.[7] We now extend the strategy to the site-specific immobilization of proteins in a microarray format (Fig. 1). Briefly, a cysteine residue is introduced at the N-terminus of a target protein which is then fused to the C-terminus of an intein and subsequently expressed in a suitable host (bacterium or yeast). The resulting intein fusion could either be purified and cleaved in vitro or cleaved directly in vivo, to obtain the desired N-terminal cysteine-containing target protein. Subsequent immobilization of the protein occurs efficiently and chemoselectively under aqueous conditions, between the N-terminal cysteine on the protein and the thioester group on the slide, resulting in the

Keywords: Intein; N-terminal cysteine-containing protein; Protein array; Proteomics.
* Corresponding author. Tel.: +65 68741683; fax: +65 67791691; e-mail: chmyaosq@nus.edu.sg

0960-894X/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2005.03.079

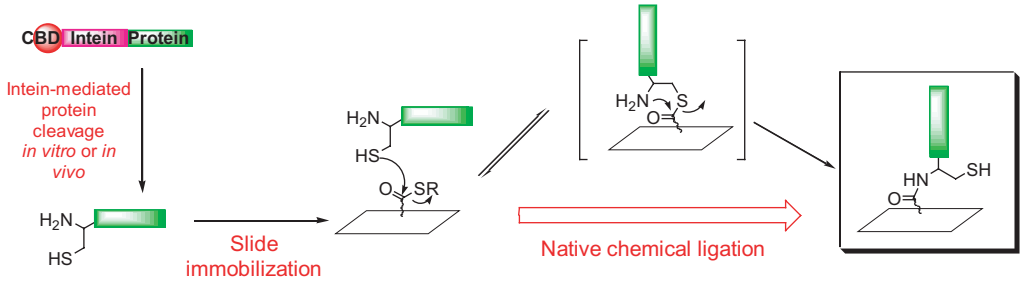

Figure 1. Protein immobilization using the native chemical ligation reaction with N-terminal cysteine-containing proteins expressed by intein-mediated protein splicing.

site-specific attachment of the protein. It is noted that recently, Camarero et al. independently introduced a similar strategy which immobilizes thioester-containing proteins (also generated by intein-mediated splicing) with cysteine-functionalized slides.[4d] Also from our previous work, it was shown that intein-mediated protein splicing serves as an efficient means to express N-terminal cysteine-containing proteins which could be further modified in vivo for bioimaging applications.[8] In the current study, we used it for the first time for protein microarray generation.

The Ssp DnaB intein-mediated strategy was originally designed to generate and purify N-terminal cysteine-containing proteins in a single-step in *E. coli*. The approach is direct and eliminates the need for extra steps (e.g., the use of additional proteases) or reagents (e.g., the addition of thiols during cleavage), all of which could complicate protein purification. In order to obtain a pure N-terminal cysteine-containing protein, it is necessary that a sufficient amount of the uncleaved CBD–intein fusion protein is obtained before binding to a chitin column. By performing the protein expression at room temperature for 12 h, we were able to obtain a substantial amount of the fusion (at least 50%; Lane 1 in Fig. 2a) before chitin column purification. In vitro cleavage of the fusion protein was further reduced by carrying out on-column loading and washings at 4 °C using lysis buffer (pH 8.5). On-column cleavage and purification conditions were also optimized by carrying out the on-column cleavage reaction at room temperature for 12 h using the cleavage buffer (pH 7.0). On an average, highly purified N-terminal cysteine-containing EGFP fractions could be routinely obtained with sufficient yield (~1.5 mg/mL; Lanes 2–4 in Fig. 2a) and directly used for spotting onto a slide. To confirm that the N-terminal cysteine-containing EGFP obtained this way is suitable for protein microarray generation, the concentration of the protein sample was adjusted to 1 mg/mL and other series dilutions, before spotting onto a thioester-functionalized glass slide.[9] Successful immobilization of EGFP was unambiguously confirmed by probing the slide with a Cy5-labeled anti-EGFP antibody (Fig. 2b). The spots arrayed on the glass slides were generally clear and defined and with relatively low background signals. As observed from the native fluorescence of the immobilized EGFP (Fig. 2c), relative spot intensity reached saturation at about 0.2–1 mg/mL, indicative of the upper limit of the amount that a protein

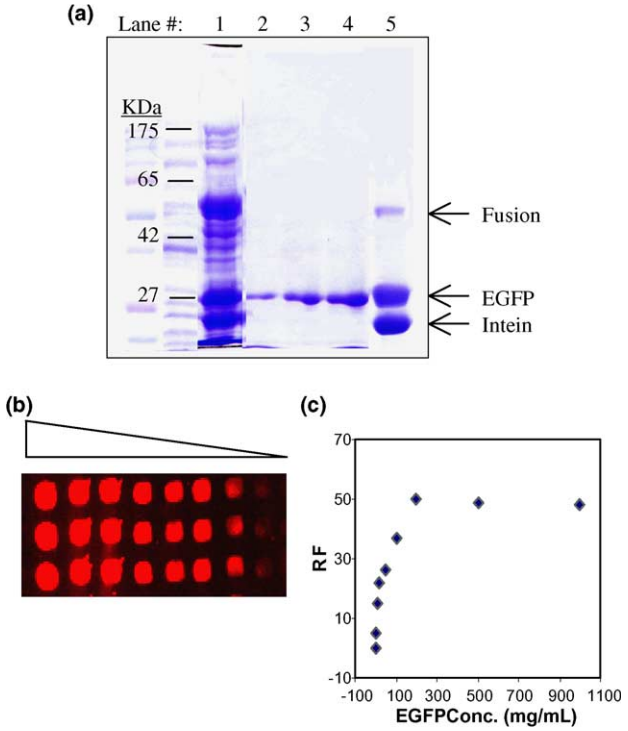

Figure 2. (a) Expression and purification of N-terminal cysteine-containing EGFP by in vitro intein-mediated cleavage on a chitin affinity column. Lane 1: Cell lysate containing overexpressed CBD–intein–EGFP fusion protein (~50 kDa). Both the fusion and in vivo-cleaved EGFP (i.e., the 27 kDa band) were observed. Lanes 2–4: The first three fractions eluted from chitin column. Lane 5: Proteins retained in the chitin column after elution. (b) Immobilization and detection of purified N-terminal cysteine-containing EGFP on a thioester slide. Varied concentrations of EGFP (left to right in mg/mL: 1, 0.5, 0.2, 0.1, 0.05, 0.02, 0.01, and 0.005) were spotted, in triplicate, and probed with Cy5-labeled anti-EGFP antibody. The native fluorescence of immobilized EGFP was also measured and plotted graphically in (c).

may be immobilized with this strategy. Significantly, as little as 0.005–0.01 mg/mL of EGFP was sufficient to give an easily detectable signal. This is in good agreement with the detection limit observed in other immobilization strategies.[4,5] The observation of native EGFP fluorescence also serves to confirm the proper folding of EGFP upon immobilization. The practicability and robustness of this immobilization strategy was further demonstrated through stringent washings and

prolonged storage of the EGFP slide, which showed a negligible loss in its native fluorescence (data not shown).

Having successfully validated the immobilization strategy with pure proteins, it was postulated that crude cell lysates may be directly spotted onto thioester glass slides, doing away with the tedious protein purification steps. This is possible because the native chemical ligation reaction between the N-terminal cysteine-containing proteins and thioester groups (on the slide) is highly chemoselective, even in the presence of other molecules in a crude cell lysate (e.g., other proteins and chemicals),[10] which, upon protein immobilization, may be easily removed by simple washing of the glass slide.[5b,8] In addition, yield of the substantial amount of N-terminal cysteine-containing proteins already present in the bacterial host (i.e., Lane 1 in Fig. 2a) may be further improved by modifications of cell growth/protein expression conditions.[8a] Crude cell lysates overexpressing N-terminal cysteine-containing EGFP and GST, respectively (rows 1 and 4 in Fig. 3), were spotted directly, without any purification, onto a thioester slide and the immobilization of these proteins was detected by native EGFP fluorescence and Cy5-labeled anti-GST antibody (A/B and C, respectively).[11] As a control, a commercial EGFP not having a N-terminal cysteine was also spotted (row 3). To confirm that endogenous proteins in a crude cell lysate (which do not normally contain an N-terminal cysteine) were not immobilized nonspecifically onto the glass slides, the crude lysate overexpressing EGFP was spiked with commercially available GST (not containing an N-terminal cysteine) before immobilization (row 2). As shown in Figure 3, only N-terminal cysteine-containing proteins, including EGFP and GST, were successfully immobilized on the thioester slide, whereas control proteins not having an N-terminal cysteine were readily washed off; the native fluorescence of EGFP disappeared upon slide washings (row 3 in Fig. 3A and B); while the commercial GST in the spiked sample was not detected upon Cy5-labeled anti-GST treatments (row 2 in Fig. 3C).

One obvious limitation with above system is that it uses a bacterial expression host, which in many cases cannot be used to express complex eukaryotic proteins due to the lack of post-translational processing and modification machineries in bacteria. The use of *Saccharomyces cerevisiae* as an expression host may solve this problem, and at the same time provide numerous advantages that other eukaryotic hosts lack, such as easy genetic manipulation and rapid growth under simple media. We therefore investigated the feasibility of using intein-mediated protein splicing to generate N-terminal cysteine-containing proteins in yeast and subsequent immobilization to thioester-derivatized slides. Again, EGFP and GST were used as models. As shown in Figure 4a, the pTWIN vectors, containing EGFP and GST fused to the C-terminus of an intein, respectively, were PCR-amplified, and cloned into a yeast expression vector,

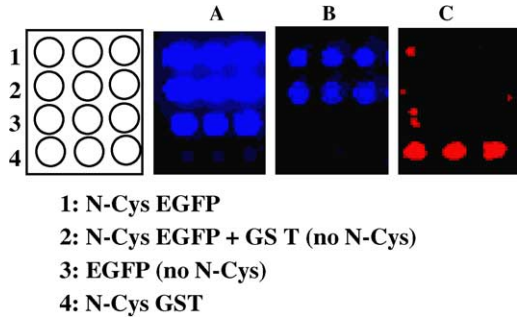

Figure 3. Direct spotting of whole cell lysates overexpressing N-terminal cysteine-containing proteins onto thioester-functionalized glass slides.[11] The samples were spotted in triplicates horizontally with spotting pattern shown on the left panel. Row 1: N-terminal cysteine-containing EGFP, 2: N-terminal cysteine-containing EGFP spiked with pure GST (1 mg/mL, no N-terminal cysteine), 3: commercial EGFP without N-terminal cysteine, 4: N-terminal cysteine-containing GST. Slides were treated differently followed by visualization: (A) visualized in FITC channel w/o washing; (B) visualized in FITC channel after slide washing; (C) visualized in Cy5 channel after slide washing and hybridization with Cy5-labeled anti-GST antibody.

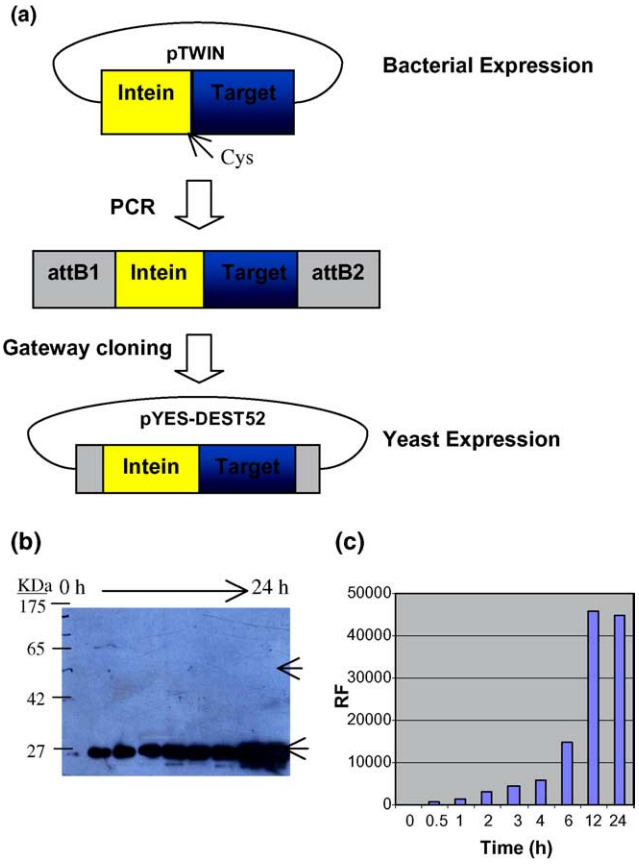

Figure 4. Intein-mediated expression of N-terminal cysteine-containing proteins in yeast.[12] (a) Gateway™ cloning of the two yeast expression constructs. (b) EGFP overexpression in yeast with different induction time (left to right: 0, 0.5, 1, 2, 3, 4, 6, 12, and 24 h). The gel was visualized by western blotting with anti-EGFP antibody and ECL™ kit. Therefore only EGFP-containing proteins may be visualized. Note that only the cleaved, N-terminal cysteine-containing EGFP (indicated by bottom arrow), and not intein–EGFP (top arrow), was detected throughout the process. (c) Overexpression of protein expression in live yeast cells was quantitated by periodic measurements of native EGFP fluorescence at different induction time (*x*-axis).

pYES-DEST52 (also known as destination vector; available from invitrogen), using the Gateway™ cloning strategy. The use of Gateway™ cloning makes the strategy adaptable for future high-throughput protein expression experiments. DNA sequencing was carried out to verify the presence of the insert in the destination vector. The final yeast expression constructs were chemically transformed into yeast, followed by protein expression and microarray immobilization.[12]

Figure 4b and c showed the time-dependent expression of EGFP in yeast. Given the intrinsically lower expression level of proteins in yeast as compared to that in a bacterial system, the overexpression of N-terminal cysteine-containing EGFP, in both the intein-fused and spliced forms, was visualized by Western blotting with an anti-EGFP antibody (Fig. 4b). Concurrently, the native EGFP fluorescence of the yeast cells was measured periodically, which serves as another means to determine the relative amount of properly folded EGFP expressed inside the cells (Fig. 4c). Both EGFP and GST (data not shown) were well-expressed in yeast, with the expression of the protein starting within the first 30 min of induction, and progressively increasing over a period of 12 h. Further increases in cell growth did not noticeably increase the yield of the protein. One important observation from the experiment is the complete absence of intein-fused protein (e.g., intein–EGFP fusion as indicated by top arrow in Fig. 4b) in the yeast cells throughout the duration of protein expression: the only overexpressed protein was the intein-spliced, N-terminal cysteine-containing EGFP (e.g., indicated by bottom arrow in Fig. 4b), indicating that a complete cleavage/splicing mediated by intein had occurred under yeast physiological conditions. The same phenomenon was also observed for the other protein (e.g., intein–GST). Subsequent modifications of yeast growth and protein expression processes under a variety of different conditions (i.e., temperature, time, concentration of chemical inducer, etc.) also failed to generate any detectable amount of the fusion protein (data not shown). To assess whether the in vivo cleaved EGFP (which contains an N-terminal cysteine) could be used directly for downstream protein microarray generation, the crude yeast cell lysate at the end of 12 h EGFP overexpression was taken directly, without further purifications, and spotted onto a thioester-functionalized slide, as described earlier. Unfortunately, we were not able to observe significant EGFP fluorescence on the slide after extensive washing (data not shown). This is likely due to a much lower expression of EGFP in yeast. Indeed, when GST was similarly expressed in yeast and enriched with a glutathione column, then spotted onto a thioester slide, we were able to observed successful GST immobilization with a Cy5-labeled anti-GST antibody (data not shown). We are currently investigating a number of strategies which may allow the enrichment of N-terminal cysteine-containing proteins expressed in yeast such that they may be used for protein microarray generation in high-throughput.

In conclusion, we have developed a facile and robust method to generate proteins for site-specific immobilization in a microarray format, while potentially preserving the biological activity of the protein in the process. This method has a number of key characteristics: (1) the site-specific immobilization of proteins via their N-terminal cysteine may allow the proteins to retain their homogeneous biological activities by virtue of their uniform orientation on the glass surface; (2) the generation of N-terminal cysteine-containing proteins using intein-mediated protein is direct and highly efficient (especially in yeast), allowing proteins to be used for protein microarray generation without tedious purification processes; and (3) the system is extremely robust as proteins are attached onto glass slides by means of a strong covalent bond. In our previous work,[5,8] we had shown that intein-mediated protein splicing events do not typically go to completion under both bacterial and mammalian host systems, consequently producing both the fusion and the cleaved product of a target protein. This again was confirmed in our current work when EGFP was expressed in *E. coli* (i.e., Fig. 2a). The fact that a complete protein splicing was observed in *S. cerevisiae* with the same protein (i.e., Fig. 4b) indicates that yeast might potentially provide an ideal host for highly efficient production of N-terminal cysteine-containing proteins, which are suitable for further downstream applications such as in vivo protein labeling for bioimaging and protein engineering experiments.[5,8]

Acknowledgements

Funding support was provided by the National University of Singapore (NUS) and the Agency for Science, Technology and Research (A*STAR) of Singapore.

References and notes

1. (a) Chen, G. Y. J.; Uttamchandani, M.; Lue, R. Y. P.; Lessaicherre, M. L.; Yao, S. Q. *Curr. Top. Med. Chem.* **2003**, *3*, 705–724; (b) Uttamchandani, M.; Walsh, D. P.; Yao, S. Q.; Chang, Y. T. *Curr. Opin. Chem. Biol.* **2005**, *9*, 4–13.
2. Zhu, H. et al. *Science* **2001**, *293*, 2101–2105.
3. For a recent review, see: Yeo, D. S. Y.; Panicker, R. C.; Tan, L. P.; Yao, S. Q. *Comb. Chem. High Throughput Screening* **2004**, *7*, 213–221.
4. (a) Hodneland, C. D.; Lee, Y. S.; Min, D. H.; Mrksich, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5048–5052; (b) Kindermann, N.; George, M.; Johnsson, N.; Johnsson, K. *J. Am. Chem. Soc.* **2003**, *125*, 7810–7811; (c) Yin, J.; Liu, F.; Li, X.; Walsh, C. T. *J. Am. Chem. Soc.* **2004**, *126*, 7754–7755; (d) Camarero, J. A.; Kwon, Y.; Coleman, M. A. *J. Am. Chem. Soc.* **2004**, *126*, 14730–14731.
5. (a) Lesaicherre, M. L.; Lue, R. Y. P.; Chen, G. Y. J.; Zhu, Q.; Yao, S. Q. *J. Am. Chem. Soc.* **2002**, *124*, 8768–8769; (b) Lue, R. Y. P.; Chen, G. Y. J.; Hu, Y.; Zhu, Q.; Yao, S. Q. *J. Am. Chem. Soc.* **2004**, *126*, 1055–1062; (c) Tan, L. P.; Lue, R. Y. P.; Chen, G. Y. J.; Yao, S. Q. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 6067–6070; (d) Tan, L. P.; Chen, G. Y. J.; Yao, S. Q. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5735–5738.
6. Chen, G. Y. J.; Uttamchandani, M.; Zhu, Q.; Wang, G.; Yao, S. Q. *ChemBioChem.* **2003**, *4*, 336–339.

7. (a) Uttamchandani, M.; Chan, E. W. S.; Chen, G. Y. J.; Yao, S. Q. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 2997–3000; (b) Lesaicherre, M. L.; Uttamchandani, M.; Chen, G. Y. J.; Yao, S. Q. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2079–2083; (c) Lesaicherre, M. L.; Uttamchandani, M.; Chen, G. Y. J.; Yao, S. Q. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 2085–2088.
8. (a) Yeo, S. Y. D.; Srinivasan, R.; Uttamchandani, M.; Chen, G. Y. J.; Zhu, Q.; Yao, S. Q. *Chem. Commun.* **2003**, 2870–2871; (b) Yeo, S. Y. D.; Srinivasan, R.; Chen, G. Y. J.; Yao, S. Q. *Chem. Eur. J.* **2004**, *10*, 4664–4672.
9. The cloning of the two pTWIN constructs was as previously described,[8a] and transformed into *E. coli* ER2566 host cells. The cells were grown in 1 L of LB medium containing 100 mg/L ampicillin at 37 °C in a 250 rpm air shaker. Protein expression was induced at $OD_{600} \sim 0.6$ with 0.5 mM IPTG (isopropyl-β-D-thiogalactoside) and left shaking overnight at room temperature. Cells were harvested by centrifugation (5000×*g*, 30 min, 4 °C), resuspended in lysis buffer (20 mM Tris–HCl, pH 8.5, 500 mM NaCl, and 1 mM EDTA) and lysed by sonication on ice. The cell debris was pelleted down by centrifugation (4000×*g*, 30 min, 4 °C) to give a clear lysate containing the intein-fusion proteins. A column packed with 10 mL of chitin beads was pre-equilibrated with column buffer (20 mM Tris–HCl, pH 8.5, 500 mM NaCl, and 1 mM EDTA). The clear lysate was loaded onto the column at a flow rate of 0.5 mL/min and washed with 10 vol of column buffer. The column was then flushed quickly with one column volume of cleavage buffer (20 mM Tris–HCl, pH 7.0, 500 mM NaCl, and 1 mM EDTA) before stopping the flow. The above procedures were carried out at 4 °C to prevent premature on-column cleavage of intein-tag. On-column incubation with the cleavage buffer took place for 20 h at room temperature with gentle agitation and the protein was eluted in 2 mL fractions. Each step of protein expression and purification was analyzed by 12% SDS-PAGE. Thioester slides were prepared as previously described.[7b] The N-terminal cysteine-containing proteins were adjusted to a stock concentration of 1 mg/mL (with PBS buffer, pH 7.4). Stocks with series of dilutions were also prepared. The protein was spotted onto the thioester-functionalized glass slides using an ESI SMA™ arrayer (Toronto, Canada). Prewash scanning of the slide was done where indicated. All slides were incubated for 3 h and subsequently washed with water for 20 min, followed by detection or storage at 4 °C. For detection of protein immobilization, slides were scanned with an ArrayWoRx™ either directly under FITC channel, or hybridized with Cy5-labeled anti-EGFP or anti-GST antibodies, respectively, for 1 h, and analyzed as previously described.[5]
10. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776–779.
11. For microarray spotting with crude cell lysates expressing a desired N-terminal cysteine-containing protein, bacterial cells were harvested after 18 h of protein expression, lysed in 1 X PBS (pH 7.4) and spotted immediately onto thioester slides.
12. Gateway™ cloning of the two yeast expression constructs was done based on protocols suggested by the vendor (Invitrogen, USA) with the pTWIN bacterial constructs as PCR templates. Protein expression in yeast was done by growing the cells in SD-Uracil+2% Glucose at 30 °C. 1 mL cultures were collected at specific time intervals and their EGFP fluorescence was measured. The cell lysates were also analyzed by SDS-PAGE and Western blotting with anti-EGFP (or anti-GST where applicable) antibody. Microarray experiments were performed as described.[9]