VIP Ligand Design

DOI: 10.1002/anie.201207005

Single-Round, Multiplexed Antibody Mimetic Design through mRNA Display**

*C. Anders Olson, Jeff Nie, Jonathan Diep, Ibrahim Al-Shyoukh, Terry T. Takahashi, Laith Q. Al-Mawsawi, Jennifer M. Bolin, Angela L. Elwell, Scott Swanson, Ron Stewart, James A. Thomson, H. Tom Soh, Richard W. Roberts, and Ren Sun**

There is a pressing need for new technologies to generate robust, renewable recognition tools against the entire human proteome.[1] Hybridoma-based monoclonal antibodies,[2] the standard for protein reagents, are undesirable for this task because of the number of animals, amount of target, time, and effort required to generate each reagent. Here, we developed a unified approach to solve this problem by integrating four distinct technologies: 1) a combinatorial protein library based on the 10th fibronectin type III domain of human fibronectin (10Fn3),[3] 2) protein library display by mRNA display,[4] 3) selection by continuous-flow magnetic separation (CFMS),[5] and 4) sequence analysis by high throughput sequencing (HTS).[6] Next generation sequencing has revolutionized many fields of biology, and is increasingly being used to improve ligand design efforts.[7] The result of our integrated approach is the ability to perform selection-based protein design in a single round—an entirely in vitro, highly scalable, and multiplexed process. Statistical analysis of input and selected pools reveals the key factors in the success of this first trial were the high uniformity of the input pools and excellent fold enrichment. Our results also demonstrate that highly functional binders ($K_D \approx 100$ nM or better) are present with a frequency of >1 in 10^9 in our library.

To begin, we needed to devise an appropriate selection protocol and library creation format. Typically, in vitro selections require multiple rounds of modest sequential enrichment, followed by small-scale sequencing of functional clones (Figure 1 A). Indeed, the need to generate a target-specific library at each round provides a significant limitation towards parallelizing and accelerating selections. In contrast, for identification of ligands after a single round of CFMS mRNA display (Figure 1 A,B), only a single naive library pool must be synthesized for any number of targets, drastically decreasing the effort needed for ligand discovery.

To integrate CFMS mRNA display with HTS, we adapted a protein scaffold with variable regions that can be easily read by Illumina HTS. Previously, we had designed, created, and optimized a high-complexity library termed e10Fn3[5,8]—based on the 10Fn3 scaffold developed by Koide and co-workers.[3] This scaffold contains only two random sequence regions, the BC loop (7 residues) and the FG loop (10 residues, Figure 1 C), which can be read by paired end sequencing using customized primers (Figure 1 D and Figure S1 in the Supporting Information). Additionally, the simplicity of the scaffold enables rapid, accurate binder reconstruction using oligonucleotides for validation without the need for cloning into bacteria.

Generally, mRNA display selections use large libraries (10^{12}–10^{14} sequences) with low copy number (3–10 copies) for protein design.[5,8–10] To achieve a single-round selection, we needed to balance the input diversity with three factors: 1) the number of clones we could sequence, 2) the fold enrichment in a single round of CFMS, and 3) the inherent frequency of functional clones in our library. One lane of an Illumina GAIIx yields 20–30 million sequences, thus we reasoned that clones enriched to greater than 1 in 1 million would be identified with 20–30 copies and therefore would be identifiable above the statistical background. Our previous work indicated that CFMS enrichment was >1000-fold per round,[5] thus enabling us to identify functional clones occurring at a frequency of about 1 in every 1 billion sequences in the naive pool (Figure 1 A). Prior yeast display work used much smaller libraries to isolate functional clones[11] supporting a complexity of about 10^9.

We also needed to increase the copy number in our library to detect functional sequences versus background clones by statistics alone. In conventional mRNA display selections

[*] Dr. C. A. Olson, J. Diep, Dr. I. Al-Shyoukh, Dr. L. Q. Al-Mawsawi, Prof. R. Sun
Department of Molecular and Medical Pharmacology
University of California, Los Angeles, CA 90095 (USA)
E-mail: rsun@mednet.ucla.edu
J. Nie, J. M. Bolin, A. L. Elwell, S. Swanson, Dr. R. Stewart, Prof. J. A. Thomson
Morgridge Institute for Research, Madison, WI 53707 (USA)
Prof. T. T. Takahashi, Prof. R. W. Roberts
Department of Chemistry, Department of Chemical Engineering
University of Southern California, Los Angeles, CA 90089 (USA)
Prof. J. A. Thomson
Department of Cellular & Regenerative Biology
University of Wisconsin, Madison, WI 53706 (USA)
and
Department of Molecular, Cellular & Developmental Biology
University of California, Santa Barbara, CA 93106 (USA)
Prof. H. T. Soh
Departments of Materials and Mechanical Engineering
University of California, Santa Barbara, CA 93106 (USA)

[**] Funding for this work was provided by the NIH, 5R01AI085583 (R.W.R., H.T.S., and R.S.). C.A.O. was supported by the NCI Cancer Education Grant, R25 CA 098010. J.A.T was supported by the Charlotte Geyer Foundation and the Morgridge Institute for Research.

Supporting information for this article, including a detailed Methods Section, is available on the WWW under http://dx.doi.org/10.1002/anie.201207005.

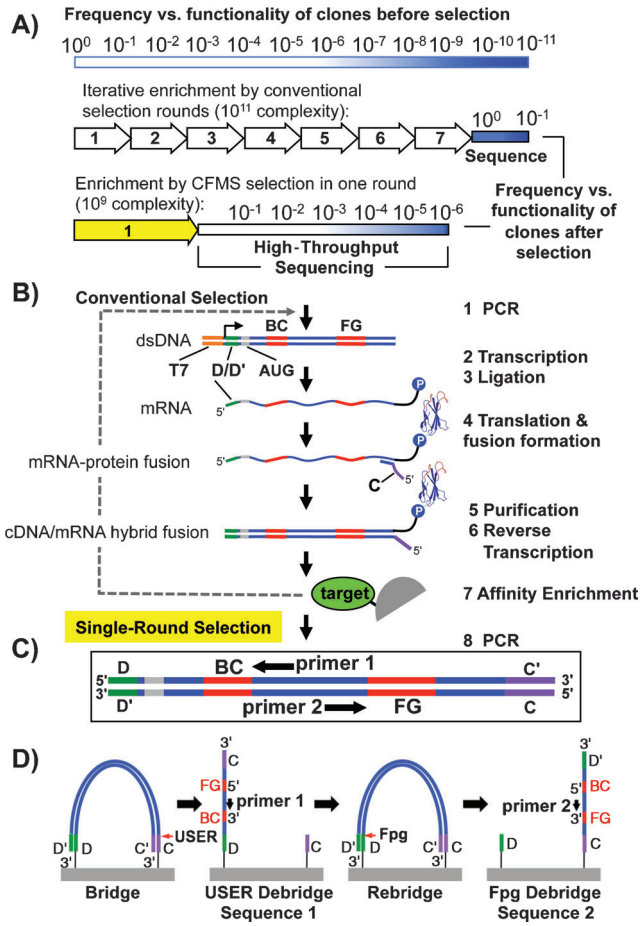


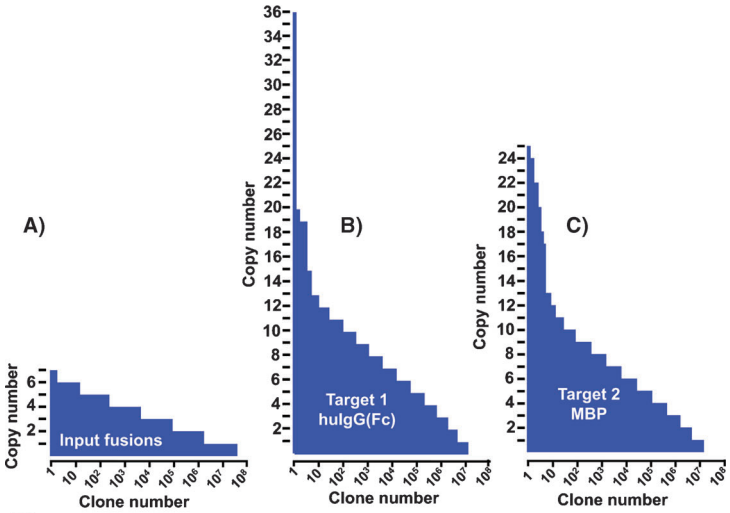
Figure 1. Selection scheme. A) Graphical representation of functional sequence enrichment by conventional or single-round CFMS mRNA display. Functionality, a combination of specificity and affinity, is depicted by a gradient from white (nonfunctional, more common) to dark blue (high functionality, least common). In conventional selection, many rounds of enrichment are performed until most clones are functional. In our single round selection described here, a lower complexity library (about 10^9) combined with improved enrichment efficiency by CFMS (about 1000-fold)[5] enables identification of functional clones >1 in 10^6 by Illumina sequencing. B) Naive mRNA display library synthesis steps are illustrated (steps 1–6). The e10Fn3 library was adapted for Illumina sequencing by inserting one of the annealing regions necessary for bridge amplification ("D") in the 5′ untranslated region. The second chip-annealing/bridge-amplification region ("C") is added by the reverse transcription primer (step 4). C) For high-throughput selection we identify ligands after one round of selection by sampling the semienriched pools through Illummina HTS (D) using the incorporated annealing regions for bridge amplification and e10Fn3-specific sequencing primers.

with low copy numbers, 1–2 copies of each functional clone are retained by the first enrichment (Figure S2 A). However, even if only about 0.01 % of the nonfunctional clones are retained because of our CFMS protocol,[5] the individual copy number of these clones (1 copy for each random, sequence-independent carryover) would be indistinguishable from the functional clones. Therefore, here we aimed to achieve a copy number of 1000 for each sequence, allowing adequate separation of functional from nonfunctional clones (Figure S2B).

In any random in vitro selection, it is essential that synthesis of the library be as unbiased as possible. This is especially true for our goal of identifying the best binders after one round of selection where the initial frequency cannot be determined for each of the ≈ 1 billion unique library members. After the affinity enrichment step, the observed frequencies of selected clones are dependent on both initial frequency and binding efficiency. Therefore, large skews in clone representation in the initial pool will hinder identification of the clones with the best binding efficiency. Thus, we sequenced a fraction of both the DNA library (step 1, Figure 1 B) and purified fusions (step 6, Figure 1 B). Using statistical models, we determined the copy number distribution of each library and determined, for the first time, that mRNA display achieves desired target complexities (here about 0.9×10^9 unique purified fusions) with highly uniform clone frequencies (see Figures S3 and S4 in the Supporting Information).

To validate our method, we chose two ubiquitous, highly used targets: maltose binding protein (MBP) and human IgG(Fc). MBP is commonly used for purification and solubilization of fusion proteins.[12] Human IgG ligands may be useful for detecting antigen-reactive antibodies in vitro or for purification of IgGs from complex mixtures. Using our CFMS method described previously,[5] we performed affinity enrichment with both targets in parallel (see the Supporting Information). Next, samples were amplified by PCR below saturation prior to performing sequencing on an Illumina GAIIx instrument. Following sequencing, we ranked each unique clone by copy number. Figure 2 A demonstrates the clone frequency landscape for the input library (purified fusions prior to selection) and the two semienriched pools that had undergone one round of affinity selection versus IgG(Fc) (Figure 2 B) or MBP (Figure 2 C). The two selected pools contained significantly over-represented clones relative to the frequency distribution of the random carryover as evident by the skewing of the copy number landscape at the highest end. An important feature of our strategy is clones that have high copy numbers but appear in pools selected against different targets are likely to be matrix-binding sequences and can therefore be eliminated from future study (Figure 2 D,E).

The clones with the highest copy number that were unique to each target were rapidly validated to bind a target. To do this, we developed a strategy for reconstruction, expression, and testing entirely in vitro, without the need for cloning into bacteria. We used synthetic oligonucleotides covering the loop regions to amplify and extend each clone from the selected pools. We then used the crude PCR reactions as template for coupled in vitro transcription and translation. Next, the pull-down efficiencies of the expressed binders on beads with or without target were assayed by western blot (Figure S5 and Figure 2 D,E). The top two unique, putative IgG(Fc) binders (eFn-anti-IgG1 and 2) and the top five MBP binders (eFn-anti-MBP1-5) were pulled-down specifically by their respective targets with no detectable binding to the beads alone. The frequencies of the best binders validate the predicted outcome that the best IgG(Fc) binder, eFn-anti-IgG1, was estimated to be enriched approximately 1500-fold.



D)

eFn-anti-IgG#	Copies	BC loop sequence	FG loop sequence	Target specific sequence	Target binding	K_D (M)
1	36	**PATTYTL**	**PCNQCQLRPN**	Yes	++	2.8×10^{-8}
-	20	PSFAFTV	PLKLINLTWP	No		
2	19	**GSAKTTV**	**LGTFFLKKSM**	Yes	+	
-	19	PTPMLAV	WWTFPWSPTL	No		
3	15	**LTLEIAV**	**WWEYIERLIR**	Yes	–	
-	15	KIMRWAI	PLVVDIWSYE	No		

E)

eFn-anti-MBP#	Copies	BC loop sequence	FG loop sequence	Target specific sequence	Target binding	K_D (M)
-	25	PTPMLAV	WWTFPWSPTL	No		
-	24	LSGCGMV	SLVRQHFYRD	No		
1	22	**AGLWLGL**	**YRASWTPNYS**	Yes	++	2.8×10^{-7}
2	20	**QAPHLFV**	**QFMYLLLPGR**	Yes	++	1.3×10^{-7}
-	18	PSFAFTV	PLKLINLTWP	No		
3	17	**NMRSWFL**	**WPRTPPYTLW**	Yes	++	
-	13	LGRPVWV	DYWHYLLFLN	No		
4	13	**QLYTRML**	**IPSFFLYHIS**	Yes	++	
-	13	QMEMPFV	FNRTYPMLRP	No		
5	13	**LRWTSRV**	**MAWPWLLAHW**	Yes	+	

Figure 2. Clone frequency distribution landscape. A–C) Sequences are ranked by copy number, illustrating the landscape of variants in the input and selected pools. (D,E) The top sequences are listed for each pool. Target-specific sequences (denoted in bold) indicate clones that were not observed in the other selection pool. Reconstructed clones were analyzed for target binding via western blot analysis. + + indicates a strong binding, + indicates faint binding and – indicates no binding within the limit of detection. Affinities are from Figures S6 and S7.

Furthermore, even lower frequency binders (e.g., eFn-anti-MBP4-5) displayed greater enrichment than previously achieved using our batch-based affinity enrichment method (see the Supporting Information).[5]

We first attempted to determine the eFn-anti-IgG1 binding constant using surface plasmon resonance (SPR), but were not able to regenerate the chip surface in a timely manner without denaturing the immobilized IgG. As an alternative, we used an equilibrium-binding assay to measure pull-down of biotinylated, monomeric eFn-anti-IgG1 onto polystyrene multiwell plates coated with limiting amounts of IgG at room temperature. The amount of eFn-anti-IgG1 bound to IgG coated plates was measured by subsequently adding excess SA-HRP, followed by washing and detection of luminescence (Figure 2D and Figure S6A). Using nonlinear regression, the binding isotherm was fit to a K_D of 28 ± 6 nM. The Hill coefficient was 1.005, indicating there was no cooperativity arising from the surface or avidity effects. Also, we demonstrated eFn-anti-IgG1 recognizes IgG at the consensus site demonstrated to bind numerous proteins and peptides[13] by competition with protein G (Figure S6B).

We used SPR to determine the binding kinetics and affinity of the top two MBP binders at 25 °C (Figure 2E and Figure S7). MBP binding to eFn-anti-MBP1 immobilized on a SA sensor chip demonstrated a $k_{on} = 1.69\times10^{5}\,M^{-1}s^{-1}$ and $k_{off} = 4.77\times10^{-2}\,s^{-1}$ resulting in a K_D of 282 nM. MBP binding to eFn-anti-MBP2 demonstrated a $k_{on} = 1.01\times10^{4}\,M^{-1}s^{-1}$ and $k_{off} = 1.30\times10^{-3}\,s^{-1}$ resulting in a K_D of 129 nM. While eFn-anti-MBP1 displays lower affinity than eFn-anti-MBP2, both display excellent functionality in avidity-enhanced detection assays.

To maximize the use of an affinity reagent, it is valuable to be able to introduce modifications that enable simple detection of targets in vitro or in vivo. Previously, we have demonstrated use of mRNA display-selected ligands similar to and beyond that of antibodies in inhibition or detection assays both intracellularly and in vitro.[8,9,14] Here, we demonstrate two methods for detection in vitro by linking two validated IgG ligands (eFn-anti-IgG1 and 2) and five validated MBP ligands (eFn-anti-MBP1-5) to alkaline phosphatase (AP) or horseradish peroxidase (HRP). To link our fibronectins with AP, we generated direct, genetic fusions as demonstrated previously,[15,16] using a new vector designed to simplify cloning (Figure 3A and Figure S8). Alternatively, for detection through HRP, we fused the C terminus of our e10Fn3s with a short peptide that can be specifically biotinylated by the *E. coli* biotin ligase, BirA (Figure 3B and Figures S9 and S10). This construct can then be detected using streptavidin-conjugated HRP (SA-HRP). Since AP forms a stable dimer (Figure 3C) and SA a tetramer, the multivalent display of fibronectins may enhance target detection through avidity.

We first tested our enzyme-linked e10Fn3s for target detection using ELISA-like assays (Figure S11). eFn-anti-IgG1 displayed saturated detection with no background while eFn-anti-IgG2 was less efficient with significant background in a *p*-nitrophenyl phosphate (PNPP) reaction assay (Figure S11A). We subsequently used eFn-anti-IgG1-AP in a mock sandwich assay, similar to antigen-reactive IgG detection assays, using SA-coated polystyrene multiwell plates to pull down various concentrations of biotinylated IgG(Fc). eFn-anti-IgG1-AP was able to detect IgG(Fc) at a wide range of concentrations spanning three orders of magnitude using a single experimental condition (Figure 3D).

We then tested target detection by eFn-anti-IgG1 through SA-HRP. First, we compared the efficiency and specificity of eFn-anti-IgG1-SA-HRP relative to conventional polyclonal anti-human IgG(Fc) in luminescence-based ELISAs (Figure 3E). Detection was comparable to the commercially

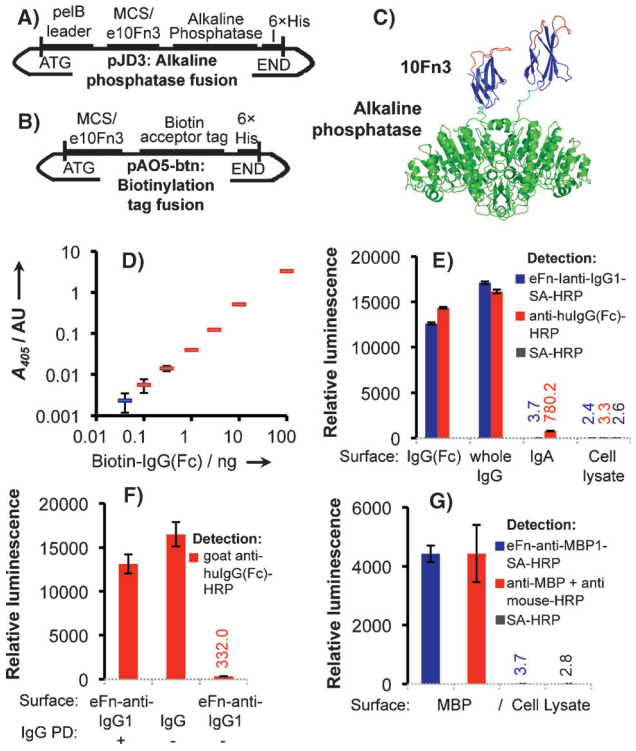


Figure 3. Validation of binders for use in enzyme-linked detection assays. a) pJD3 vector for creating e10Fn3-alkaline phosphatase fusions and b) pAO5-btn vector for generating fusions to a biotin acceptor peptide tag. c) Structural model of 10Fn3-AP fusions. The 10Fn3 scaffold is blue, randomized loops are red, and alkaline phosphatase is green. 10Fn3 domains (PDBID: 1FNF and 1FNA) are joined to AP (PDBID: 1ALK) by a GSGGSG linker (cyan). d) Mock sandwich antibody-free ELISA assay for IgG detection. (Blue is the no-IgG control). e) Comparison of eFn-anti-IgG-SA-HRP to anti human IgG-HRP. f) eFn-anti-IgG1 efficiently captures IgG (Pull-down, "PD") for detection by polyclonal anti-huIgG-HRP g) Comparison of eFn-anti-MBP1-SA-HRP to a commercially available monoclonal. Binding assays were performed in triplicate.

available reagent with both Fc and full antibody. The background from mammalian cell lysate was equivalent to either polyclonal anti-IgG(Fc) or SA-HRP alone, within error. The eFn-anti-IgG1-SA-HRP background binding to IgA was negligible, similar to the signal from cell lysate, and significantly lower than the commercially available antibody. Figure 3F demonstrates that eFn-anti-IgG1 captures IgG efficiently and is also highly specific, binding well to human IgG but not appreciably to goat IgG. We also demonstrate the use of eFn-anti-IgG1-SA-HRP in western blotting, as IgG in the low nanogram range is detectable after SDS PAGE and transfer to nitrocellulose (Figure S12A).

We similarly conjugated the MBP binders to SA-HRP for luminescence-based detection assays. eFn-anti-MBP1 displayed the highest detection efficiency (Figure S11B) and exhibited efficiency of detection comparable to a commercially available anti-MBP monoclonal antibody, with background equal to SA-HRP alone, within error (Figure 3G). eFn-anti-MBP1-SA-HRP was similarly effective in antibody-free western blotting as demonstrated by detecting expression of an MBP-fusion protein (Figure S12B).

This study demonstrates approximately 10^9 sequences were sufficient for selecting an IgG binder with an affinity of 28 nM, roughly equivalent to long established existing reagents (protein A and protein G) targeting this domain.[17] Also, this affinity is typical of selections using high complexity 10Fn3-based libraries.[5,8,9] This indicates that biases in the multistep selection process may preclude selecting higher affinity binders that theoretically exist in higher complexity libraries. Our CFMS platform requires only minimal target material (100 pMoles or 1–10 μg) and enables performing affinity enrichment in parallel. Without further improvements, about 12 single-round selections could be analyzed per lane on an Illumina HiSeq instrument (through barcoding, see the Supporting Information). By scaling up the initial naive pool synthesis and automating the affinity enrichment step, high throughput ligand generation may be attainable. A proteome-wide affinity reagent resource based on the highly validated 10Fn3 antibody mimetic scaffold[18] would greatly impact molecular biology research as well as human health by providing a wealth of diagnostics and potential therapeutics.

Received: August 29, 2012
Published online: November 5, 2012

Keywords: antibody mimetics · directed evolution · ligand design · mRNA display · selection methods

[1] O. Stoevesandt, M. J. Taussig, *Proteomics* **2007**, *7*, 2738.
[2] G. Köhler, C. Milstein, *Nature* **1975**, *256*, 495.
[3] A. Koide, C. W. Bailey, X. Huang, S. Koide, *J. Mol. Biol.* **1998**, *284*, 1141.
[4] R. W. Roberts, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 12297.
[5] C. A. Olson, J. D. Adams, T. T. Takahashi, H. Qi, S. M. Howell, T. T. Wu, R. W. Roberts, R. Sun, H. T. Soh, *Angew. Chem.* **2011**, *123*, 8445; *Angew. Chem. Int. Ed.* **2011**, *50*, 8295.
[6] D. R. Bentley, et al., *Nature* **2008**, *456*, 53.
[7] U. Ravn, F. Gueneau, L. Baerlocher, M. Osteras, M. Desmurs, P. Malinge, G. Magistrelli, L. Farinelli, M. H. Kosco-Vilbois, N. Fischer, *Nucleic Acids Res.* **2010**, *38*, e193; S. T. Reddy, X. Ge, A. E. Miklos, R. A. Hughes, S. H. Kang, K. H. Hoi, C. Chrysostomou, S. P. Hunicke-Smith, B. L. Iverson, P. W. Tucker, A. D. Ellington, G. Georgiou, *Nat. Biotechnol.* **2010**, *28*, 965; M. Cho, Y. Xiao, J. Nie, R. Stewart, A. T. Csordas, S. S. Oh, J. A. Thomson, H. T. Soh, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 15373; G. V. Kupakuwana, J. E. Crill 2nd, M. P. McPike, P. N. Borer, *PLoS One* **2011**, *6*, e19395; A. T. Bayrac, K. Sefah, P. Parekh, C. Bayrac, B. Gulbakan, H. A. Oktem, W. Tan, *ACS Chem. Neurosci.* **2011**, *2*, 175; S. Hoon, B. Zhou, K. D. Janda, S. Brenner, J. Scolnick, *Biotechniques* **2011**, *51*, 413.
[8] C. A. Olson, H.-I. Liao, R. Sun, R. W. Roberts, *ACS Chem. Biol.* **2008**, *3*, 480.
[9] H. I. Liao, C. A. Olson, S. Hwang, H. Deng, E. Wong, R. S. Baric, R. W. Roberts, R. Sun, *J. Biol. Chem.* **2009**, *284*, 17512.
[10] T. T. Takahashi, R. W. Roberts, *Methods Mol. Biol.* **2009**, *535*, 293.
[11] D. Lipovšek, S. M. Lippow, B. J. Hackel, M. W. Gregson, P. Cheng, A. Kapila, K. D. Wittrup, *J. Mol. Biol.* **2007**, *368*, 1024.
[12] R. B. Kapust, D. S. Waugh, *Protein Sci.* **1999**, *8*, 1668.
[13] W. L. DeLano, M. H. Ultsch, A. M. de Vos, J. A. Wells, *Science* **2000**, *287*, 1279.

[14] F. N. Ishikawa, H. K. Chang, M. Curreli, H. I. Liao, C. A. Olson, P. C. Chen, R. Zhang, R. W. Roberts, R. Sun, R. J. Cote, M. E. Thompson, C. Zhou, *ACS Nano* **2009**, *3*, 1219.
[15] C. Suzuki, H. Ueda, E. Suzuki, T. Nagamune, *J. Biochem.* **1997**, *122*, 322.
[16] Z. Han, E. Karatan, M. D. Scholle, J. McCafferty, B. K. Kay, *Comb. Chem. High Throughput Screening* **2004**, *7*, 55.
[17] K. Saha, F. Bender, E. Gizeli, *Anal. Chem.* **2003**, *75*, 835.
[18] L. Bloom, V. Calabro, *Drug Discovery Today* **2009**, *14*, 949.

