Protein Engineering, Design & Selection, 2016, vol. 29 no. 5, pp. 187–195
doi: 10.1093/protein/gzw006
Advance Access Publication Date: 15 March 2016
Original Article

OXFORD

Original Article

Investigating affinity-maturation strategies and reproducibility of fluorescence-activated cell sorting using a recombinant ADAPT library displayed on staphylococci

Mikael Åstrand[1], Johan Nilvebrant[1,2], Mattias Björnmalm[1,3], Sarah Lindbo[1], Sophia Hober[1], and John Löfblom[1,*]

[1]Division of Protein Technology, School of Biotechnology, KTH Royal Institute of Technology, Stockholm, Sweden, [2]Present address: Centre for Cellular and Biomolecular Research, The Donnelly Centre, University of Toronto, Toronto, Ontario, Canada, and [3]Present address: Department of Chemical and Biomolecular Engineering, The University of Melbourne, Parkville, Victoria, Australia

*To whom correspondence should be addressed. E-mail: lofblom@kth.se
Edited by Shohei Koide

Received 30 September 2015; Revised 22 January 2016; Accepted 12 February 2016

Abstract

During the past decades, advances in protein engineering have resulted in the development of various *in vitro* selection techniques (e.g. phage display) to facilitate discovery of new and improved proteins. The methods are based on linkage between genotype and phenotype and are often performed in successive rounds of selection. Since the resulting output depends on the selection pressures used and the applied strategy, parameters in each round must be carefully considered. In addition, studies have reported biases that can cause enrichment of unwanted clones and/or low correlation between abundance in output and affinity. We have recently developed a selection method based on display of protein libraries on *Staphylococcus carnosus* and isolation of affinity proteins by fluorescence-activated cell sorting. Here, we compared duplicate selections for affinity maturation using equilibrium binding at different target concentrations and kinetic off-rate selection. The results showed that kinetic selection is efficient for isolation of high-affinity binders and that equilibrium selection at subnanomolar concentrations should be avoided. Furthermore, the reproducibility of the selection was high and a clear correlation was observed between enrichment and affinity. This work reports on the reproducibility of bacterial display in combination with FACS and provides insights into selection design to help guide the development of new affinity proteins.

Key words: ADAPT, bacterial display, directed evolution, FACS, selection strategy

Introduction

Several directed evolution methods have been developed for combinatorial engineering of affinity proteins. They typically share the property of physical linkage of the protein-encoding gene and its polypeptide product, usually mediated by a host organism (e.g. in phage or bacterial display), to enable isolation of the target-binding proteins and subsequent propagation of the associated genes. Phage display has been the most widely used method for engineering of new affinity reagents, partly due to its straightforward workflow and efficient display of a wide range of different proteins (Chan *et al.*, 2014). Display on the surface of yeast or bacteria are attractive alternatives because fluorescence-activated cell sorting (FACS) can be

used for quantitative isolation of library members with desired properties and for monitoring the selection progression between rounds. By incubating the library-displaying cells with a fluorescently labeled target molecule, the fluorescence will be a direct measure of the amount of bound target to each individual cell and consequently also the respective affinities. The efficiency of cell-display methods is often further improved by using an expression-normalization tag to minimize potential biases due to differences in expression level (Löfblom, 2011; Gera *et al.*, 2013).

Some studies have demonstrated the occurrence of mutants within phage libraries that amplify at an abnormally high rate. Such biases can potentially lead to the occlusion of desirable library members (Menendez and Scott, 2005; Brammer *et al.*, 2008; Kuzmicheva *et al.*, 2009; Derda *et al.*, 2011; Matochko *et al.*, 2014; Nguyen *et al.*, 2014). Another reason for loss of diversity could be that certain gene products are toxic to the cells. If expression is allowed during growth, the affected cells do not propagate or propagate at a slower rate during amplification (Krebber *et al.*, 1996). Presumably, these effects vary greatly depending on what type of protein- or peptide library is displayed and together with which coat protein. It has been shown that a tightly regulated expression is key to decrease the toxicity of displayed heterologous proteins (Krebber *et al.*, 1996; Huang *et al.*, 2000). The same seems to be true also for yeast display where the GAL1 promoter generally is used, which enables inducible expression after the yeast growth phase (Boder and Wittrup, 1997; Gera *et al.*, 2013). This has been shown to allow for sequential library propagation with minimal introduction of diversity bias (Feldhaus *et al.*, 2003). *Escherichia coli* has also been adapted as a suitable cell-surface display platform, where the araBAD promoter is commonly used to regulate expression (Georgiou *et al.*, 1997; Daugherty *et al.*, 1999; Fleetwood *et al.*, 2014). In addition to various biases that might skew the enrichment, the applied selection pressure has a direct impact on the success of combinatorial protein engineering. The selection pressure is typically increased incrementally throughout rounds by, for example, decreasing the target concentration. For isolation of high-affinity binders (e.g. in affinity maturation efforts), so-called kinetic selections are often applied and theoretical considerations have suggested that it should be more efficient for enrichment of binders with high affinities in the low and subnanomolar range (Boder and Wittrup, 1998). Despite this, relatively poor correlation between enrichment and affinity is sometimes reported (Nilvebrant *et al.*, 2013; Reich *et al.*, 2014), making it necessary to individually screen clones in order to rank them.

Staphylococcal surface display is an alternative display technology using the gram-positive bacterium *Staphylococcus carnosus* that has been developed by our group (Wernérus and Ståhl, 2002; Kronqvist *et al.*, 2008). It has been used to successfully display fibronectin-binding domains (Liljeqvist *et al.*, 1999), peptide epitopes from *Salmonella enterica* (Nhan *et al.*, 2011) and libraries of single-domain camelid antibodies (nanobodies) (Fleetwood *et al.*, 2013), Affibody molecules (Kronqvist *et al.*, 2011; Lindberg *et al.*, 2013) and various libraries for epitope mapping (Rockberg *et al.*, 2008). ABD-derived affinity protein (ADAPT) is a scaffold based on one of the albumin-binding domains from streptococcal protein G that has also been displayed (Alm *et al.*, 2010; Nilvebrant *et al.*, 2011, 2013, 2014; Nilvebrant and Hober, 2013) and is the scaffold used in this study. Bispecific ADAPTs are small (46 amino acids) single-domain three-helical proteins, with retained affinity to albumin and an additional specificity (to for example a drug target) engineered into the opposite surface of the domain (Fig. 1a). The ADAPT molecules are displayed on the staphylococcal surface in fusion with two Z domains (Fig. 1b). These Z domains have a high-affinity for IgG, which allows for monitoring of the surface expression levels on individual cells in a flow cytometer. The cell-surface expression on staphylococcal cells is facilitated by a signal peptide at the N-terminus, which leads to secretion, and a C-terminal region including a LPETG motif that anchors the protein to the cell wall peptidoglycan by an endogenous sortase (Wernérus and Ståhl, 2002). The expression is under control of a constitutive lipase promoter from *S. hyicus*. Assuming that no biases exist, clones should be selected based on the ratio of amount of bound antigens per displayed binding proteins and varying the antigen concentration should enable discrimination between clones. Thus, there should be a correlation between clonal enrichment and binding affinity. VanAntwerp and Wittrup (2000) and Löfblom *et al.* (2005) have shown that it is practically possible to distinguish between two clones with very similar affinities.

Here, we sought to investigate the staphylococcal surface display platform for biases that could potentially hamper a selection. An affinity-maturation library based on previously engineered ADAPT binders for the human epidermal growth factor receptor 3 (ERBB3) was constructed and used to select candidates using FACS. ERBB3 is implicated in the progression of numerous cancers and is a potential therapeutic target (Zhang *et al.*, 2015). Two duplicate selections were separately performed and compared to evaluate the sequence agreement in the outputs of the different rounds of selection. The replicates from the successful sortings showed a large overlap between the output sequences, demonstrating that the method has a high reproducibility. Moreover, characterization of the output revealed that employing equilibrium selection at subnanomolar target concentrations was suboptimal for enrichment of high-affinity binders, but the kinetic selection resulted in a strong enrichment of in particular one dominating clone that indeed demonstrated the slowest dissociation rate. The results presented here reveal for the first time the reproducibility of bacterial display in combination with FACS for isolation of affinity proteins and include a quantitative assessment of the outcome from using kinetic screenings compared with equilibrium screenings and hence offers insights for optimally designing selection strategies in the future.

Materials and methods

Library assembly

Based on earlier candidates binding to ERBB3 isolated from a naïve ADAPT library (Nilvebrant *et al.*, 2013), an affinity-maturation library was designed where the same 11 positions were targeted for randomization using degenerate codons corresponding to amino acids with high prevalence in the first-generation candidates (Fig. 1a and Supplementary Fig. S1). The total size of the library was calculated to 4.4×10^6 protein variants and 1.8×10^7 gene variants. The library was encoded by a 189-bp ultramer oligonucleotide (Integrated DNA Technologies, IDT, San Diego, CA, USA) with the sequence 5′-CAGG ATCCTCTCGAGGATGAAGCCGTCGACGCGAATTCATTAGC TASKGCTAAAAVAKTAGCTCTGYACDTKCTTGACVNKARAGG AGTAAGTGACTATTACAAGGATCAATCGATAAAGCCAAAACT GTTGAAGGAGTACDKGCACTGDYAMKCGAAATTTTAVBAG CATTACCCGCTAGCTTTCCG-3′. The oligo was assembled and amplified by polymerase chain reaction (PCR) with Phusion polymerase (New England Biolabs (NEB), Ipswich, MA, USA) using external primers incorporating restriction sites XhoI and NheI starting with 100 pmol oligo. The construct was digested over night at 37°C using restriction enzymes XhoI and NheI (NEB) after gel extraction

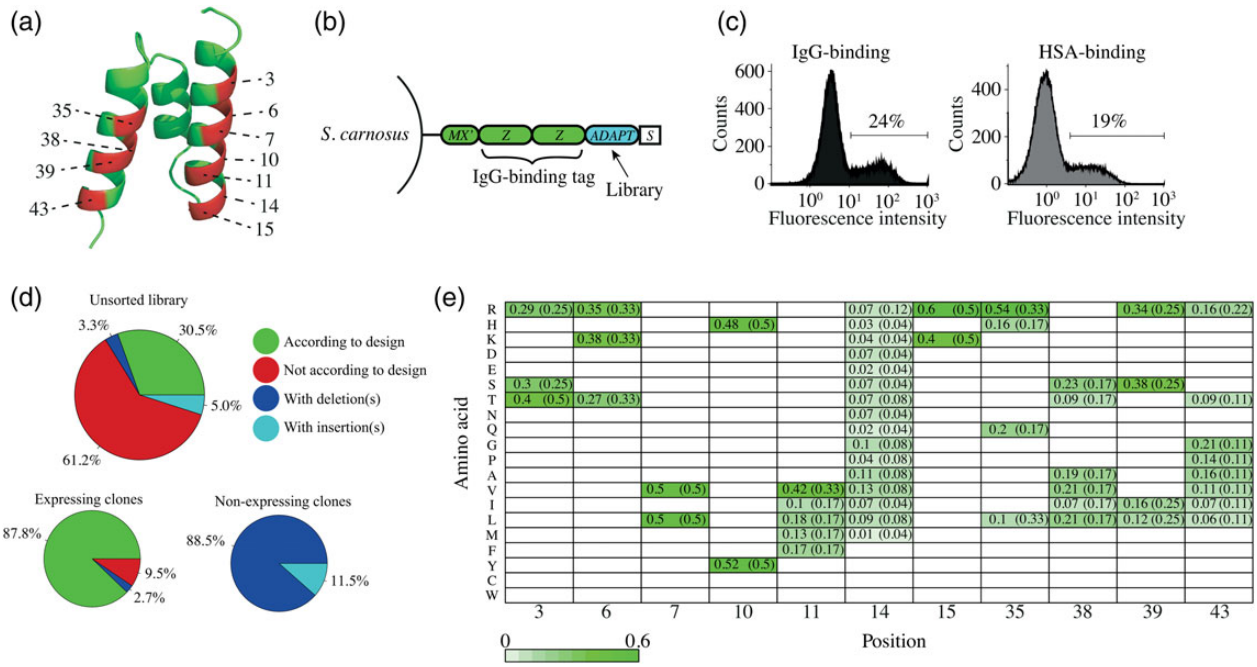

Fig. 1 Analysis of unsorted library. (**a**) Structure of the ABD scaffold on which the ADAPT library is based. Positions of the amino acids chosen for randomization are indicated (PDB: 1GJT). (**b**) Cartoon depicting the fusion protein displayed on the surface of *S. carnosus*. MX′ corresponds to the processed cell-wall anchoring region containing the LPETG motif. ADAPT and Z correspond to the library insert and the IgG-binding domains, respectively. S corresponds to the signal peptide from *S. hyicus*. (**c**) Flow-cytometry histograms showing fluorescence from bound IgG and HSA to *staphylococci* expressing the unsorted library. (**d**) Results from sequencing of random unsorted library members and members sorted based on their ability to bind fluorescent IgG, thus separating them in expressing and nonexpressing populations, respectively. Sequencing reads with incomplete data or poor quality were excluded. Sequences that agreed with the library design were labeled as 'allowed'. Pie charts show the proportion of clones that follow the library design compared with clones that do not in each population. (**e**) Amino acid distribution of the library members with a sequence in agreement with the design. The fraction for each observed amino acid out of all sequenced clones is indicated for every randomized position. Numbers in parenthesis are the theoretical distributions based on the codon usage in the design. The fractions are highlighted, and the intensity indicates the abundance.

purification from a 2% Seakem GTG agarose gel (Lonza, Basel, Switzerland) with a gel extraction kit (Qiagen, Venlo, Netherlands). The digested fragment was ligated into the staphylococcal surface display vector pSCABD1 (Nilvebrant *et al.*, 2011) digested with the same enzymes and purified by gel extraction. Approximately 2.2 pmol of vector was ligated with 11.2 pmol of library insert. Proper assembly of the library oligo was verified by agarose gel electrophoresis and ligation into the display vector was assessed by PCR and agarose gel electrophoresis.

Library transformation

The assembled library vector was purified by agarose gel extraction (Qiagen) using standard protocols and transformed by electroporation into electrocompetent *E. coli* SS320 (Lucigen). The transformed cells were grown over night in 2 × 500 ml tryptic soy broth (TSB) supplemented with 100 μg ml^{-1} ampicillin, and the plasmid DNA was extracted and purified using a Jetstar maxiprep kit (Genomed, Löhne, Germany). The purified plasmid was further purified by chloroform–phenol extraction and electroporated into electrocompetent *S. carnosus* according to previously described protocols (Löfblom *et al.*, 2007).

Sequencing

All sequencing reactions were performed using the method of Sanger with specific primers using a BigDye Terminator mix 3.1 (Invitrogen, Carlsbad, CA, USA) according to the instructions provided by the supplier. All sequencing reactions were run with PCR-amplified DNA fragments as template and analyzed on an ABI Prism 3700 DNA analyzer (Applied Biosystems, Foster City, CA, USA). Chromatograms were analyzed, and sequence logos were generated using the Geneious software v6.1 (Biomatters, Auckland, New Zealand). Pie charts were generated using the R language. Venn/Euler diagrams were generated using the VennDiagram package for the R language.

Protein labeling

Recombinant human ERBB3 (ERBB3; Sino Biological, Beijing, China) was biotinylated using Biotin-XX-Succinimidyl ester (Invitrogen) using the procedure recommended by the supplier. Briefly, a 100-fold molar excess of biotin was added to 50 μg of ERBB3 in 0.1 M $NaHCO_3$ buffer (pH 8.3) and incubated end-over-end for 1 h at room temperature. Labeled ERBB3 was dialyzed over night at 4°C against 1 × PBS using a Slide-a-lyzer dialysis casette (Pierce, Rockford, IL, USA) with a 3.5 kDa molecular weight cut-off. The concentration of the labeled protein was determined using absorbance analysis at 280 nm.

Fluorescence-activated cell sorting

Prior to each selection round, 100 ml cultures with TSB supplemented with yeast extract (TSB + Y; Merck, Darmstadt, Germany) and 20 μg ml^{-1} chloramphenicol were inoculated and incubated overnight. In Round 1, the culture was inoculated with cells corresponding to a 10-fold excess compared with the number of transformants reached during electroporations, which corresponds to a 100-fold

excess compared with the number of possible protein variants in the library. A 10-fold excess of cells compared with the number of sorted cells in the previous cycle was used as inoculant in all subsequent cycles. Before labeling, the cells were washed in PBS supplemented with 0.1% Pluronic (PBSP; BASF, Ludwigshafen, Germany) and subsequently incubated with biotinylated ERBB3 for 1 h at room temperature. The concentration and volume used in the incubation varied in each cycle. The ERBB3 incubation volume was adjusted in order to ensure an approximate ratio of 10^5 ERBB3 molecules per cell in order to minimize target depletion. As a second incubation step, cells were labeled with streptavidin R-phycoerythrin conjugate (SA-PE, Invitrogen) and human polyclonal IgG Alexa-647 conjugate (IgG-647) by incubating on ice for 30 min. The cells were washed as described above between each incubation step and prior to flow-cytometric sorting. A 5-fold excess of cells compared with the number of transformants were interrogated in Cycle 1 and in all following cycles a 10-fold excess compared with the number of sorted cells in the previous cycle. FACS was performed using an Astrios flow cytometer (Beckman Coulter, Brea, CA, USA). Sorted cells were collected in TSB + Y and used to inoculate 10 ml TSB + Y supplemented with 20 μg ml^{-1} chloramphenicol. Output cells were also spotted in a 10 × 10 grid on agar plates supplemented with chloramphenicol for viability testing and sequencing.

Flow-cytometric screening of selection output

Unique clones identified during sequencing were grown in 96-deep well plates in 1 ml TSB + Y. 10^6 cells, based on measurements of the cellular density by OD_{578}, were transferred from each culture, washed by centrifugation at 3000*g* at 4°C for 6 min and resuspended in PBSP. All clones were incubated with 5 nM biotinylated ERBB3 (Sino Biological) in PBSP for 1 h at room temperature. All subsequent washing and incubation steps were performed at 4°C or on ice. Detection and expression normalization were performed by incubating with SA-PE (Invitrogen) and IgG-647 for 30 min. ERBB3-binding was analyzed by interrogating 5×10^4 cells in a Gallios flow cytometer (Beckman Coulter, Brea, CA, USA) by measuring the median fluorescence intensity (MFI) of both SA-PE and IgG-647.

Growth-rate analysis

One colony each for a few chosen clones was used to inoculate 10-ml of TSB + Y supplemented with 20 μg ml^{-1} chloramphenicol. After an over-night incubation, the cultures were each used to inoculate 100 ml TSB + Y supplemented with 20 μg ml^{-1} chloramphenicol to a starting OD_{578} of 0.1. Cultures were grown at 37°C at 150 rpm, and the growth progression was monitored by withdrawing 100-μl samples at regular intervals and measuring the OD_{578} in a CLARIOstar plate reader (BMG Labtech GMBH, Ortenberg, Germany).

Sub-cloning

Individual *S. carnosus* clones were grown separately in 1 ml TSB + Y supplemented with 20 μg ml^{-1} chloramphenicol in 96-deep well plates and grown over night at 37°C. One microlitre from each culture was diluted in 50 μl sterile water of which 1 μl was subsequently used as a template in a PCR using general primers with EcoRI- and AscI-restriction sites, flanking the ADAPT gene. The PCR products were purified using a PCR purification kit (Qiagen) and digested using the appropriate restriction enzymes (NEB). Cleaved fragments were purified and ligated into a His_6-tag containing expression vector (Alm *et al.*, 2010; Nilvebrant *et al.*, 2011), which had been digested with the same enzymes, dephosphorylated using Antarctic phosphatase (NEB) and purified from a 2% agarose gel with a gel extraction kit (Qiagen). Ligated plasmids were used to heat-shock transform *E. coli* strain Top10 (Invitrogen) using standard protocols. Individual colonies carrying the correct inserts, as verified by PCR and sequencing as described above, were grown in TSB supplemented with 50 μg ml^{-1} kanamycin over night at 37°C, and plasmid DNA was purified using a plasmid miniprep kit (Qiagen). The plasmids were heat-shock transformed to *E. coli* strain Rosetta (DE5; Novagen, Madison, WI, USA) and plated on tryptone yeast extract agar plates supplemented with 50 μg ml^{-1} kanamycin and 20 μg ml^{-1} chloramphenicol.

Protein production and purification

Rosetta colonies were used to inoculate over-night cultures with TSB supplemented with 20 μg ml^{-1} chloramphenicol and 50 μg ml^{-1} kanamycin. One millilitre of culture was used the next day to inoculate 100-ml cultures with TSB + Y supplemented with 20 μg ml^{-1} chloramphenicol and 50 μg ml^{-1} kanamycin. At $OD_{600} \approx 0.6$–0.8, expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 1 mM and the cultures were incubated over night at 25°C. The next day, the cells were harvested by centrifugation at 3000*g*, 4°C for 20 min and resuspended in Tris-buffered saline (TST; 25 mM Tris–HCl, 200 mM NaCl, 1 mM EDTA, 0.05% (v/v) Tween20, pH 8.0). Cells were lyzed by sonication using a Vibra-cell sonicator (Sonics and Materials Inc., Newtown, CT, USA), and cell debris was removed by centrifugation at 10 000 g, 4°C for 20 min. Lysates were filtered through 0.45 μm filters and loaded onto HSA sepharose gravity-flow columns with 7.5 ml bed volume equilibrated with 10 column volumes (CV) TST. The columns were washed with 7 CV NH_4Ac, pH 5.5, and finally eluted in 1 ml fractions with HAc pH 2.8. Fractions containing protein, as determined by absorbance measurements at 280 nm, were dried overnight in a Savant AES2010 SpeedVac system (Thermo Scientific, Rockford, IL, USA). The fractions were resuspended in PBS, and the purity was confirmed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and mass spectrometry using a 6520 Accurate Q-TOF LC/MS (Agilent, Santa Clara, CA, USA). The concentrations were measured at least in duplicate with a bicinchoninic acid kit (Sigma-Aldrich) according to the supplier's recommendations using an amino acid analyzed ADAPT-variant as standard (Aminosyraanalyscentralen, Uppsala, Sweden). Secondary structure and stability were verified by circular dichroism (CD) spectroscopy using a Jasco J-810 spectropolarimeter (Jasco, Essex, UK). CD spectra were recorded at 250–195 nm at 25°C. Thermal stability was assessed by measuring CD at 221 nm with a temperature gradient from 25 to 95°C. An additional CD spectrum was recorded after cooling the sample to 25°C to determine the refolding capability.

Affinity measurements

The affinities to ERBB3 were measured by surface plasmon resonance (SPR) with a ProteOn XPR36 protein interaction array system (BioRad, Hercules, CA, USA). ERBB3 (Sino biological) was immobilized via amine coupling at 30 μl min^{-1} in 10 mM NaAc at pH 5.0 on a general layer medium (GLM) or a general layer compact (GLC) chip to ~2000–2500 RU. PBS supplemented with 0.005% (v/v) Tween20 (PBST) was used as running buffer, and all analyte injections were performed at 50 μl min^{-1}. All analyzed ADAPT variants were serially diluted in PBST to between 50 and 0.5 nM and injected with an association time of 300 s. The dissociation was monitored for 1000 or 2000 s. The chip surface was regenerated with 10 mM NaOH between injections. All SPR binding curves were fitted to a 1:1

Langmuir isotherm to calculate the association and dissociation kinetic constants k_a and k_d. The corresponding equilibrium dissociation constant, K_D, was calculated as k_d/k_a.

Results

Library construction and analysis

An affinity-maturation library was constructed, resulting in 5.5×10^7 variants that covered the theoretical library size on the DNA level more than 3-fold (Supplementary data, Fig. S1).

To assess the proportion of the unsorted library that expressed the recombinant proteins on the surface, bacteria were incubated with fluorescently labeled IgG and analyzed in the flow cytometer (Fig. 1c). Approximately 25% of the population was positive for IgG binding. An analogous experiment was also performed with fluorescently labeled HSA (Fig. 1c), which yielded similar results. Sequencing revealed that around 30% of the colonies contained full-length ADAPTs with sequences in concordance with the design. The rest of the sequences contained unintentional modifications resulting in a disrupted reading frame (Fig. 1d) and were not expected to yield properly expressed proteins. To investigate the cause of the nonexpressing population, cells from the expressing and nonexpressing populations were isolated with FACS. Around 90% of the cells sorted from the expressing population were viable, and sequence analysis

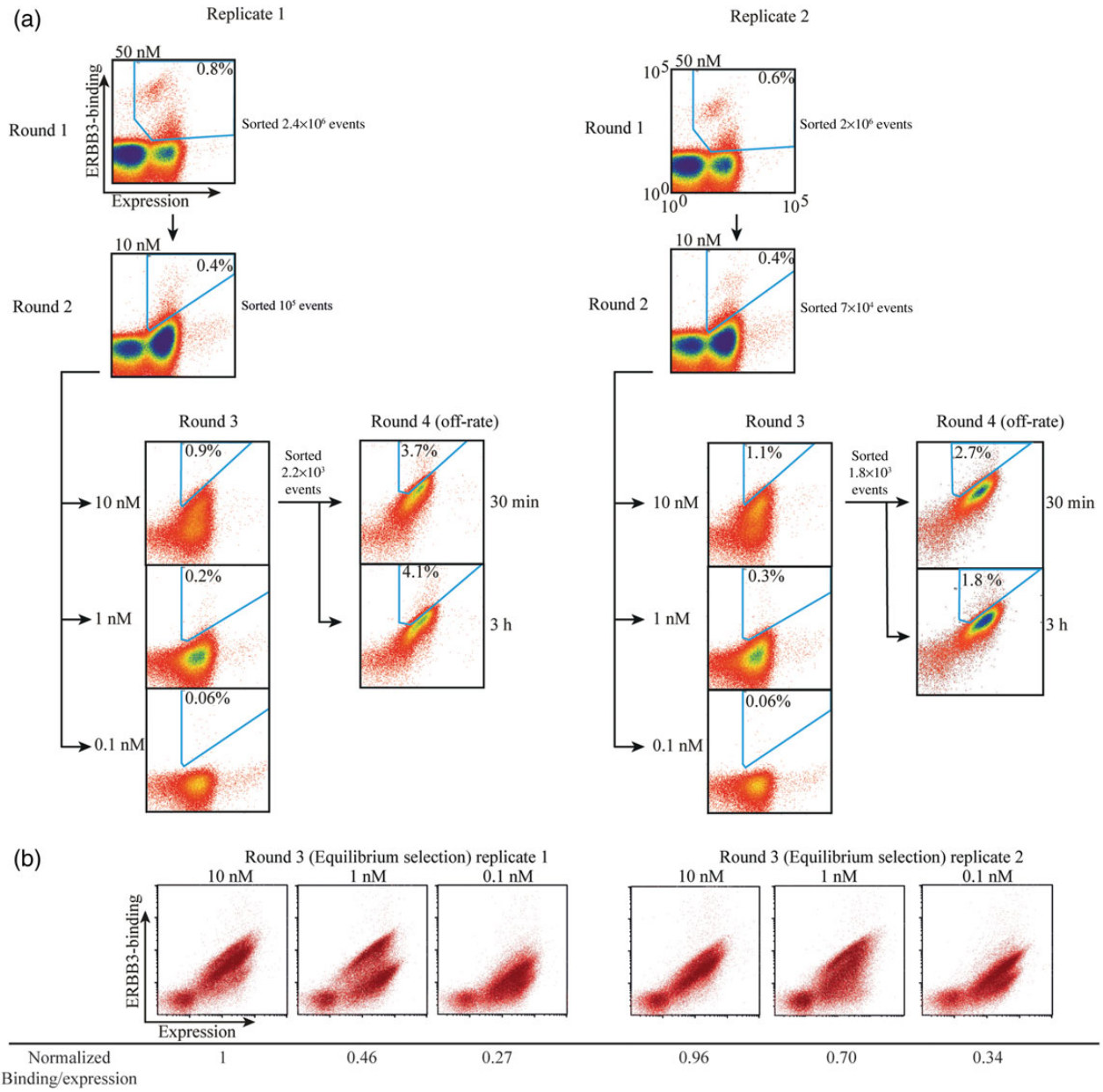

Fig. 2 Selection methodology and output analysis. (**a**) Density dot plots showing the selection order. Both separate selections are shown. The plots show R-phycoerythrin fluorescence on the *y*-axis (ERBB3-binding) and Alexa-647 fluorescence (IgG-binding, expression) on the *x*-axis. All graphs are plotted on a log-log scale. (**b**) Density dot plots from flow-cytometric analysis of the selection outputs of the indicated selection rounds (label on top of each dot plot). All samples were incubated with 1 nM ERBB3 and 50 000 events were recorded for each analysis. The ratio of the median binding fluorescence intensity and the median expression fluorescence intensity is indicated under each plot. The values from 10 nM were normalized to 1.

showed that around 88% of the ADAPTs were assembled according to the design. In contrast, only 70% of the cells from the nonexpressing population were viable, indicating that a part of the nonexpressing population was nontransformed bacteria that lacked antibiotic resistance. This is reflected in the percentage of expressing cells that bind IgG as discussed above, which is lower than the percentage of sequenced colonies carrying full-length ADAPTs. Furthermore, sequencing revealed that no full-length ADAPT sequences were present among the sorted clones from the nonexpressing population (Fig. 1d). The proportion of the expressing population also corresponded to the number of full-length sequences identified during sequencing of the unsorted library, which confirmed that a majority of the correct library members were displayed on the cell surface. The amino acid distribution among the correctly assembled library members also showed a good agreement with the design where each randomized position closely resembled the theoretically calculated distribution for the library design (Fig. 1e).

Staphylococcal surface display selection using FACS

To be able to draw conclusions from investigations of different selection strategies and comparisons of different selection methods, it is important to also explore the reproducibility of the method, i.e. will we achieve a similar binder repertoire when repeating the selection? Hence, all selections were performed in duplicate on different days (Fig. 2a). A clear enrichment of expressing cells could be seen after Round 1, while subsequent rounds also showed increasing enrichment of ERBB3-binding clones. In Round 3, the binding populations demonstrated ERBB3-concentration-dependent fluorescence signals, which was a further indication that clones binding to ERBB3 were isolated (Fig. 2a). Moreover, a fourth round of selection was performed on the clones from the 10-nM track in Round 3. Here, the aim was to select binders with the slowest off-rates, i.e. low dissociation rate constants (k_d) (Fig. 2a). The observed populations for each cycle from the two replicates were similar, indicating reproducibility of the selection method. For an initial comparison of the efficiency in isolating the variants with the highest affinity, the output populations from the three different equilibrium selections (0.1-, 1- and 10-nM tracks) were incubated with 1 nM ERBB3 and analyzed by flow cytometry. The analysis showed distinct patterns for all three different sorting strategies and most notably, that the selection at 0.1 nM had resulted in a lower enrichment of high-affinity binders (Fig. 2b).

DNA sequencing

The identities of the sorted clones from selection Rounds 3 and 4 were determined by DNA sequencing. The clones were numbered according to their overall abundance among sequenced colonies (Fig. 3a). All sorted clones in all tracks contained ADAPT sequences with intact reading frames. Only 16 out of the 1639 total sequenced clones contained mutations that did not agree with the library design. The sequencing results of the clones isolated from the kinetic selection (Round 4) revealed a high reproducibility showing that around 70% of the sequences were identical within the two replicate selections (Fig. 3a). Furthermore, the same clone dominated the outputs in both replicates (Fig. 3b). However, the similarity between clones from the replicates of the equilibrium selection (Round 3) was lower (Fig. 3a). Notably, the overlap decreased with a lower target concentration. The most striking difference was observed when comparing the 0.1-nM track, from which only one clone was identified in both replicates, indicating that enrichment occurred at random (Fig. 3a). Interestingly, the most highly enriched variants from the 0.1-nM track

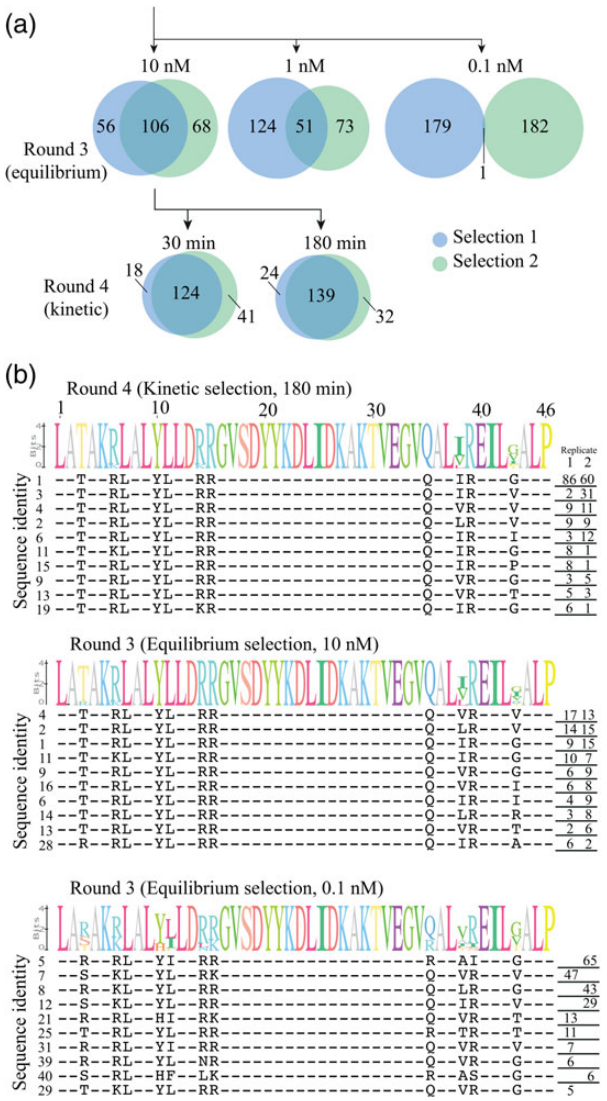

Fig. 3 Sequencing results from the selection outputs. (**a**) Venn/Euler diagrams depicting the overlap between the two selection attempts. All amino acid sequences identified during sequencing were matched to determine the sequence overlap. The section areas are proportional to the number of sequences. The arrows indicate the selection progression. (**b**) Sequence lists showing the 10 most abundant sequences from the 180-min track in Round 4, as well as the 10- and 0.1-nM tracks in Round 3. The number of times each sequence was observed in each separate selection during sequencing is shown to the right of each sequence.

were not discovered in the other tracks (Fig. 3b and Supplementary data, Fig. S2). The equilibrium sorting (Round 3) using the highest target concentration (10 nM) showed the highest reproducibility in this round with almost 50% identical clones between the two replicates. Of note is that the selection had not reached convergence to one or a few dominating clones at this stage (Fig. 3b). To verify that we had isolated the variants with highest fluorescence intensity, we performed an additional sorting from both a high-fluorescent population and a low-fluorescent population among the clones originating from the 1-nM track and compared it to the output (Fig. 4a). For this selection, 5 nM biotinylated ERBB3 was used for the sorting. The results showed that sequences from the population with lower binding signal were largely undiscovered during the selections. In contrast, the sequences from the

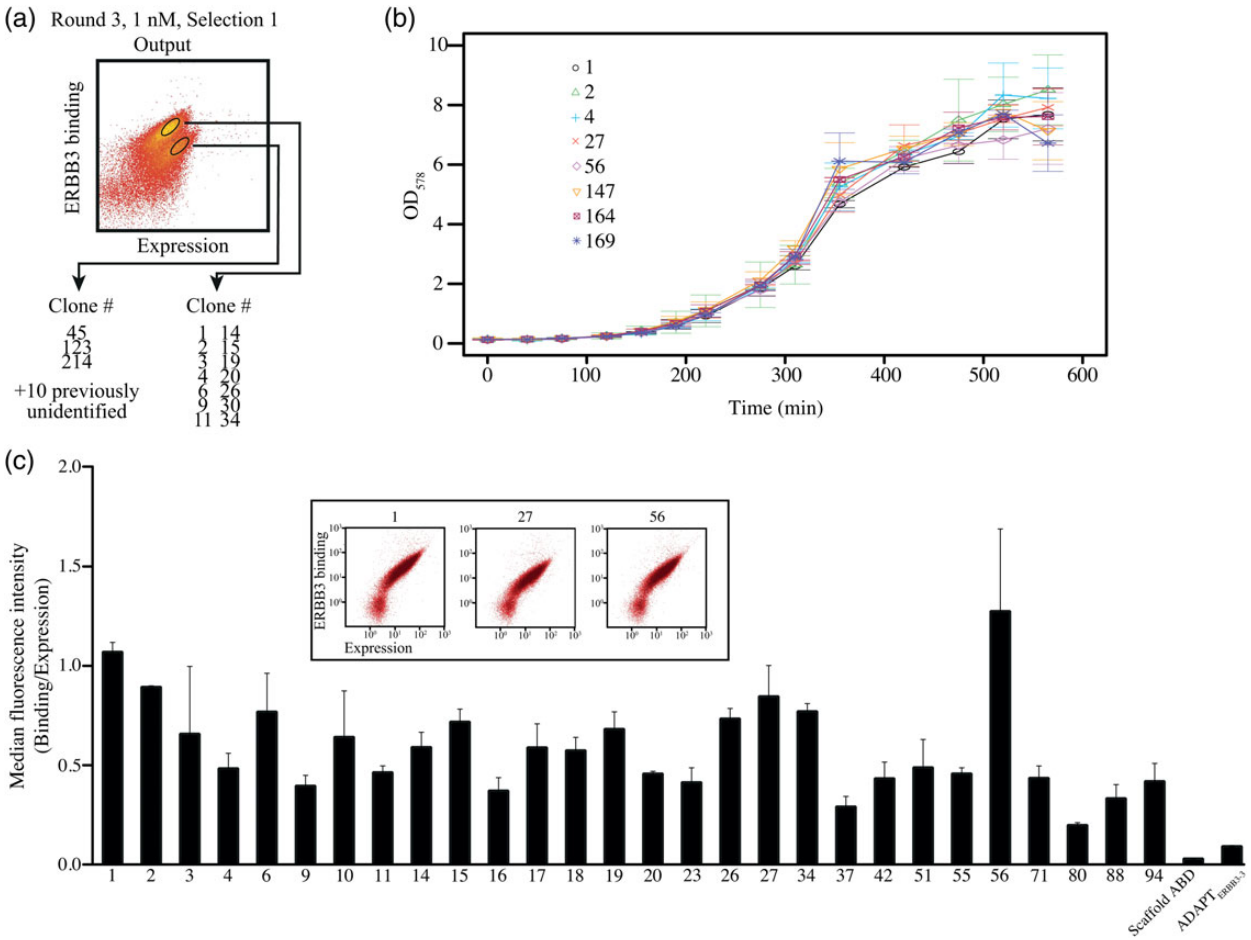

Fig. 4 Results from single-clone screening. (**a**) Sequence identities of clones sorted from each subpopulation in the indicated output clone pool. Identity numbers were determined based on the overall sequence abundance in Rounds 3 and 4. Thus, low-numbered clones were more frequent. Unidentified clones were only observed in this experiment and not during the selection. (**b**) Growth curves for eight staphylococcal clones grown individually. All measurements were done in triplicate on different days. Error bars indicate the standard deviation. (**c**) Single clones were grown in shake flasks and incubated in 1.5 ml microcentrifuge tubes. Bars represent median fluorescence intensities normalized to the expression level. Error bars indicate standard deviation for duplicate samples measured on different days. The nonbinding scaffold ABD and $ADAPT_{ERBB3\text{-}3}$ were included as negative control and reference, respectively.

population with a high binding signal were among the most frequent in the selection and all had been previously identified, thus verifying that the variants with high fluorescence signals had been isolated.

Growth-rate analysis

To investigate potential growth biases of the staphylococcal system, i.e. to rule out the possibility that clones were enriched due to significantly different generation times, a growth-rate assay was performed. Eight clones were chosen for this experiment based on the frequency at which they were observed during sequencing. The analysis demonstrated similar growth rates for all eight variants, suggesting that the most abundant clones in the output had not been enriched due to growth biases (Fig. 4b).

Single-clone affinity ranking

Individual clones, originating from Round 3 (10- and 1-nM tracks) and 4, were ranked using flow cytometry. The clones that were discovered only in the 0.1-nM track were not included as the binding signal of the outputs for this track was markedly lower than clones from the other tracks (Fig. 2b). Supplementary data, Figure S3 shows the results from a ranking of all clones by performing at least one measurement for each clone. Judging from the range of different binding signals that were observed for the screened clones, the chosen ERBB3 concentration can be considered appropriate. However, many clones have similar, albeit high, binding signals, which makes them difficult to distinguish with this concentration of target protein. Furthermore, many of the highest binding clones are among the most abundant, but the overall correlation between highly accumulated binders (low clone number) and high binding signal is relatively weak (Supplementary data, Fig. S3). A second screening attempt was made to verify the ranking among the clones giving the highest binding signal (Fig. 4c). Shake flasks were used the second time instead of deep-well plates, as well as incubation in 1.5 ml microcentrifuge tubes instead of microtiter plates, as this seemed to result in higher expression and better separation between the expressing and nonexpressing populations (see inlay plots in Fig. 4c and Supplementary data, Fig. S3). The ranking between the clones was not completely retained in the second screening attempt, although the strongest binding clones are the same in both experiments. A third screening attempt was made using 50 nM ERBB3 to more closely examine the clones that appeared only once during sequencing and had binding signals close to

background. The higher concentration of ERBB3 resulted in a higher binding signal for a majority of the clones compared with the background thus verifying that these clones were indeed selected based on affinity (Supplementary data, Fig. S4).

Affinity determination by SPR

The kinetic constants of the interactions with ERBB3 were determined using SPR. To get an estimation of the dynamic range of the screening assay, we included both high- and low-ranking clones. We also added clone #100, which had an unintended mutation (three-nucleotide deletion) in the last randomized position. For comparison, seven candidates that were only observed in the 0.1-nM track were included as well. The determined association rate constants (k_a) were in the range of 2.1×10^5–2.1×10^6 M^{-1} s^{-1} and the dissociation rate constants (k_d) between 8.1×10^{-4} and 2.7×10^{-2} s^{-1} (Table I and Supplementary data, Fig. S5). The affinities were relatively similar among most variants selected from the ranking assay, ranging from around 0.6 to ~2 nM, while one clone from the low-ranking variants had 9 nM.

As expected from the flow-cytometric analysis of the entire output pool, the clones from the 0.1-nM track had, on average, a lower affinity (ranging between 2 and 26 nM). Notably, Clone #1, which dominated the output from the kinetic selection, displayed the slowest dissociation, which suggests that the sorting indeed was quantitative and the applied strategy was appropriate for isolating variants with the lowest off-rate.

Eight of the binders were analyzed using CD spectroscopy, which confirmed alpha-helical secondary structure content and melting temperatures above 80°C (Supplementary data, Fig. S6). Due to the high stabilities, exact melting temperatures could not be determined in this experimental setting.

Table I: Affinity constants for selected clones originating from Rounds 3 and 4

Clone	k_a (M^{-1} s^{-1})	k_d (s^{-1})	K_D (nM)	Number of interactions
Clones observed in Rounds 3 (equilibrium at 1 and 10 nM) and 4				
1	$1.4(\pm0.1) \times 10^6$	$8.1(\pm1.0) \times 10^{-4}$	0.6	6
2	$2.1(\pm0.3) \times 10^6$	$1.2(\pm0.1) \times 10^{-3}$	0.6	3
4	$1.8(\pm0.2) \times 10^6$	$2.5(\pm0.3) \times 10^{-3}$	1.4	3
11	$1.7(\pm0.2) \times 0^6$	$1.3(\pm0.1) \times 10^{-3}$	0.8	6
27	$8.1(\pm0.5) \times 10^5$	$1.6(\pm0.1) \times 10^{-3}$	2.0	3
34	$8.6(\pm1.0) \times 10^5$	$1.9(\pm0.04) \times 10^{-3}$	2.2	5
37	$1.9(\pm0.2) \times 10^6$	$4.3(\pm0.5) \times 10^{-3}$	2.3	6
56	$1.4(\pm0.6) \times 10^6$	$1.2(\pm0.2) \times 10^{-3}$	0.9	6
80	$2.1(\pm0.1) \times 10^5$	$1.9(\pm0.1) \times 10^{-3}$	9.0	3
100	$1.1(\pm0.04) \times 10^6$	$2.0(\pm1.0) \times 10^{-3}$	1.8	3
Clones observed in Round 3 at 0.1 nM equilibrium				
8	$2.1(\pm0.1) \times 10^6$	$6.7(\pm1.5) \times 10^{-3}$	3	2
12	$9.1(\pm1.2) \times 10^5$	$5.3(\pm0.5) \times 10^{-3}$	6	2
21	$1.9(\pm0.2) \times 10^6$	$2.1(\pm0.03) \times 10^{-2}$	11	2
39	$1.0(\pm0.01) \times 10^6$	$2.7(\pm0.2) \times 10^{-2}$	26	2
43	$2.7(\pm0.3) \times 10^5$	$6.1(\pm0.3) \times 10^{-3}$	22	2
44	$3.8(\pm0.9) \times 10^5$	$7.3(\pm1.2) \times 10^{-3}$	19	2
45	$1.2(\pm0.1) \times 10^6$	$2.9(\pm0.3) \times 10^{-3}$	2	2

The association and dissociation kinetic constants (k_a and k_d) are shown as mean values from multiple measurements ± standard deviation. K_D was calculated as k_d/k_a. Affinity constants were determined from the number of total interaction sets indicated in the rightmost column. One or two dilution series were injected over two or three surfaces with immobilized ERBB3.

Discussion

In this study, we have investigated the reproducibility of bacterial display and impact of selection strategy on isolation of affinity proteins. As shown in Fig. 2a, the sort gates in the 10- and 1-nM tracks were placed in order to only include the cells with the highest binding signal in relation to the surface expression level.

In addition to equilibrium selections, we also included a kinetic selection with a theoretical selection pressure for binders with low dissociation rates. The complete sorting process, including all tracks, was repeated to strengthen findings regarding potential differences in output between the applied selection strategies. The aim was also to investigate how reproducible the output would be from bacterially displayed protein libraries when using FACS. Although the aim was to keep conditions and parameters as identical as possible between the duplicate selections, we expected that the unintended variability of an entire selection process over four cycles with amplification by cell growth between each round should be relatively high. However, the track with the kinetic selection in Round 4 resulted in a large overlap between the replicates, demonstrating that the staphylococcal display system combined with FACS yields robust and predictable results.

Although the reproducibility is clearly high, it is difficult to assess how the bacterial method combined with FACS compares with other methods, such as phage display, because reports on reproducibility in the literature are scarce. The results also showed that it is difficult to use very stringent conditions in terms of low target concentration in equilibrium selections for isolation of high-affinity binders, as has been previously suggested from theoretical simulations (Boder and Wittrup, 1998). It might be argued that a less stringent gate for the 0.1-nM track would have resulted in a more favorable outcome. However, the dot plot from the sorting reveals that the separation between the binding and nonbinding population is dramatically lower and a less stringent gate would include part of the nonbinders. Hence, incubating the library with this low concentration was not optimal, regardless of the gate parameters. Notably, the most enriched variants from the kinetic selection also had the slowest off-rates, demonstrating that the strategy was appropriate for isolation of high-affinity binders. No difference in growth rate was detected when comparing highly enriched clones with clones observed only once, which suggests that the enrichment was indeed based on the selection pressure.

Another advantageous aspect of cell-surface display techniques is the ability to individually screen the isolated clones for affinity. Ideally, this is performed using a range of target concentrations in order to determine the apparent K_D of the displayed binder. However, this is not practical when large numbers of clones need to be screened and we therefore performed the assay using one concentration. This successfully distinguished the highest affinity clones from the background. Comparing these results with the measured affinity constants showed that the affinities for the strongest binders were very similar and could not be accurately distinguished. Thus, this method is more appropriate for identifying positive clones and distinguishing them from clones with low affinity before determining the true affinities using appropriate biosensor assays.

In conclusion, the data presented here show that the staphylococcal display method is suitable for isolating proteins with desired affinities using FACS. It should be noted that the study was performed using a library with relatively restricted diversity. It would be interesting to conduct a similar investigation on a naïve library in the future. If speculating on the outcome of such investigation, the reproducibility might be even higher due to a lower frequency of binders compared with the library used here.

Finally, the most essential results of the study are from analysis on the reproducibility of selections from combinatorial protein libraries, detailed experimental comparison of the impact of different selection strategies on output and the correlation between output and affinity. Although directed evolution methods are well established and frequently used, these important aspects have been largely overlooked in the past. The results from our study should also translate to similar approaches for directed evolution (in particular to yeast and *E. coli* display) and offer guidelines for performing FACS-based selections as well as strengthen interpretations of results in the future.

Funding

This work was supported by the Swedish Research Council.

Supplementary data

Supplementary data are available at *PEDS* online.

References

Alm,T., Yderland,L., Nilvebrant,J., Halldin,A. and Hober,S. (2010) *Biotechnol. J.*, **5**, 605–617.

Boder,E.T. and Wittrup,K.D. (1997) *Nat. Biotechnol.*, **15**, 553–557.

Boder,E.T. and Wittrup,K.D. (1998) *Biotechnol. Prog.*, **14**, 55–62.

Brammer,L.A., Bolduc,B., Kass,J.L., Felice,K.M., Noren,C.J. and Hall,M.F. (2008) *Anal. Biochem.*, **373**, 88–98.

Chan,C.E.Z., Lim,A.P.C., MacAry,P.A. and Hanson,B.J. (2014) *Int. Immunol.*, **26**, 649–657.

Daugherty,P.S., Olsen,M.J., Iverson,B.L. and Georgiou,G. (1999) *Protein Eng. Des. Sel.*, **12**, 613–621.

Derda,R., Tang,S.K.Y., Li,S.C., Ng,S., Matochko,W. and Jafari,M.R. (2011) *Molecules*, **16**, 1776–1803.

Feldhaus,M.J., Siegel,R.W., Opresko,L.K., *et al.* (2003) *Nat. Biotechnol.*, **21**, 163–170.

Fleetwood,F., Devoogdt,N., Pellis,M., Wernery,U., Muyldermans,S., Ståhl,S. and Löfblom,J. (2013) *Cell. Mol. Life Sci.*, **70**, 1081–1093.

Fleetwood,F., Andersson,K.G., Ståhl,S. and Löfblom,J. (2014) *Microb. Cell Fact.*, **13**, 179.

Georgiou,G., Stathopoulos,C., Daugherty,P.S., Nayak,A.R., Iverson,B.L. and Curtiss,R. (1997) *Nat. Biotechnol.*, **15**, 29–34.

Gera,N., Hussain,M. and Rao,B.M. (2013) *Methods*, **60**, 15–26.

Huang,W., McKevitt,M. and Palzkill,T. (2000) *Gene*, **251**, 187–197.

Krebber,A., Burmester,J. and Plückthun,A. (1996) *Gene*, **178**, 71–74.

Kronqvist,N., Löfblom,J., Jonsson,A., Wernérus,H. and Ståhl,S. (2008) *Protein Eng. Des. Sel.*, **21**, 247–255.

Kronqvist,N., Malm,M., Göstring,L., *et al.* (2011) *Protein Eng. Des. Sel.*, **24**, 385–396.

Kuzmicheva,G.A., Jayanna,P.K., Sorokulova,I.B. and Petrenko,V.A. (2009) *Protein Eng. Des. Sel.*, **22**, 9–18.

Liljeqvist,S., Cano,F., Nguyen,T.N., Uhlén,M., Robert,A. and Ståhl,S. (1999) *FEBS Lett.*, **446**, 299–304.

Lindberg,H., Johansson,A., Härd,T., Ståhl,S. and Löfblom,J. (2013) *Biotechnol. J.*, **8**, 139–145.

Löfblom,J. (2011) *Biotechnol. J.*, **6**, 1115–1129.

Löfblom,J., Wernérus,H. and Ståhl,S. (2005) *FEMS Microbiol. Lett.*, **248**, 189–198.

Löfblom,J., Kronqvist,N., Uhlén,M., Ståhl,S. and Wernérus,H. (2007) *J. Appl. Microbiol.*, **102**, 736–747.

Matochko,W.L., Cory Li,S., Tang,S.K.Y. and Derda,R. (2014) *Nucleic Acids Res.*, **42**, 1784–1798.

Menendez,A. and Scott,J.K. (2005) *Anal. Biochem.*, **336**, 145–157.

Nguyen,K.T.H., Adamkiewicz,M.A., Hebert,L.E., *et al.* (2014) *Anal. Biochem.*, **462**, 35–43.

Nhan,N.T., Gonzalez de Valdivia,E., Gustavsson,M., Hai,T.N. and Larsson,G. (2011) *Microb. Cell Fact.*, **10**, 22.

Nilvebrant,J. and Hober,S. (2013) *Comput. Struct. Biotechnol. J.*, **70**, 3973–3985.

Nilvebrant,J., Alm,T., Hober,S. and Lofblom,J. (2011) *PLoS One*, **6**, e25791.

Nilvebrant,J., Åstrand,M., Löfblom,J. and Hober,S. (2013) *Cell. Mol. Life Sci.*, **70**, 3973–3985.

Nilvebrant,J., Åstrand,M., Georgieva-Kotseva,M., Björnmalm,M., Löfblom,J. and Hober,S. (2014) *PLoS One*, **9**, e103094.

Reich,L.L., Dutta,S. and Keating,A.E. (2014) *J. Mol. Biol.*, **427**, 2135–2150.

Rockberg,J., Löfblom,J., Hjelm,B., Uhlén,M. and Ståhl,S. (2008) *Nat. Methods*, **5**, 1039–1045.

VanAntwerp,J.J. and Wittrup,K.D. (2000) *Biotechnol. Prog.*, **16**, 31–37.

Wernérus,H. and Ståhl,S. (2002) *FEMS Microbiol. Lett.*, **212**, 47–54.

Zhang,N., Chang,Y., Rios,A. and An,Z. (2015) *Acta Biochim. Biophys. Sin. (Shanghai)*, **48**, 39–48.