This is an Accepted Manuscript, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this Accepted Manuscript with the edited and formatted Advance Article as soon as it is available.

You can find more information about Accepted Manuscripts in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal’s standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this Accepted Manuscript or any consequences arising from the use of any information it contains.
A high throughput whole cell flow cytometer screening toolbox was developed and validated by identifying improved variants (1.3-7-fold) for three hydrolases (esterase, lipase, cellulase). The screening principle is based on coupled enzymatic reaction using glucose derivatives which yield upon hydrolysis a fluorescent-hydrogel-layer on the surface of *E. coli* cells.

Flow cytometer-based screening systems are, despite their extremely high throughput (up to $1.8\times10^7$ events per hour) and a competitively high sensitivity, rarely used in directed evolution campaigns. This can be attributed to challenges in the compartmentalization technology such as double emulsion formation, dispersity of the emulsion, substrate and product compatibility, and stability of compartments [1-3]. Flow cytometry whole cell based screening systems for directed evolution require fluorogenic substrates which can diffuse inside the cell and remain entrapped upon conversion into fluorescent products [4-7]. The latter has been reported in directed evolution campaigns for three specific enzymes (P450 monooxygenase, N-acetylgalactosaminidase, protease) [5-7]. Therefore, novel screening principles which are generally applicable are of high interest for protein engineering and biocatalysis. Recently, we reported a proof of concept for a novel screening principle named Fur-Shell which is based on fluorescent hydrogel formation around *E. coli* cells with a phytase as an example. In the Fur-Shell screening platform, a whole cell acts as a compartment in which $H_2O_2$ is formed through a coupled reaction of phytase and glucose oxidase; subsequently a fluorescent monomer is co-polymerized in a fluorescent polyester hydrogel shell formed around *E. coli* cells expressing active phytase variants [8].

Significantly, the proof of concept of the Fur-Shell technology overcomes technical limitations in flow cytometry-based screening systems in terms of compartmentalization and leakage of a fluorescent substrate and/or fluorescent product. In the current report we advanced the screening principles of the Fur-Shell technology into a general high throughput screening toolbox for directed evolution of hydrolases by establishing and validating screening protocols for three hydrolases: a p-nitrobenzyl esterase from *Bacillus licheniformis* (pNBEBL) [9], a *Bacillus subtilis* lipase A (BSLA) [10], and a cellulase (CelA2) isolated from a metagenome library by Streit et al. [11]. Validation was performed for each of the three hydrolases through a single round of directed evolution by screening an epPCR random mutagenesis library. All three hydrolases address enzymes which are of significant synthetic and industrial importance (e.g. esterases and lipases are used in laundry detergents, in the synthesis of pharmaceuticals, and in food processing [12]); cellulases are applied in e.g. depolymerisation of cellulose and in food industry [13-15]). Fig. 1 shows the principle of the Fur-Shell technology in four steps. In Step 1 the gene diversity library is generated by epPCR. Subsequently, insert and vector fragments are cloned by PLCing and the mutant library is transformed and expressed in *E. coli* cells [16]. Step 2 comprises the Fur-Shell technology, which relies on the conversion of a substrate (β-D-(+)-glucose pentaacetate for esterase and lipase; cellobiose for cellulase) into β-D-glucose. Subsequently, in a glucose oxidase-coupled reaction $H_2O_2$ and glucono-δ-lactone are produced. Through a Fenton reaction, radical species from $H_2O_2$ are generated, initiating a PEG based co-polymerization of the fluorescent Polyfluorol 70 monomer [8]. *E. coli* cells expressing active enzyme variants are surrounded by a fluorescent hydrogel shell and can in Step 3 be analysed and sorted by flow cytometer at rates of around 5000 events per second. The sorted *E. coli* cells expressing active hydrolase variants are plated on agar plates and afterwards transferred for screening in microtiter plates (MTPs) (Step 4a). Alternatively, as reported here in Step 4b, performance parameters such as low cell survival of *E. coli* cells (<8 %) were...
Fig. 1: Flow cytometer-based sorting principle of the Fur-Shell toolbox for hydrolases in four steps. **Step 1:** Library generation by epPCR and subsequent cloning using PLICing. **Step 2:** *E. coli* BL21-Gold (DE3) cells producing enzyme variants are incubated with substrate, glucose oxidase and fluorescent labelled monomers (Polyfluor 570). Fluorescent hydrogel is formed around cells expressing active enzyme variants (+) which allows in **Step 3** analysis and sorting (enrichment) by flow cytometer at a rate of about 5000 events per second. Finally the *E. coli* BL21-Gold (DE3) cells are either plated on LB agar plates (Step 4a) or used for plasmid isolation and transformation (Step 4b), in both cases followed by transferring clones into MTPs.

optionally addressed by introducing a plasmid isolation step of sorted cells to rescue mutated genes that encoded improved hydrolase variants in non-viable *E. coli* cells [8]. Isolated plasmids are subsequently transformed into competent *E. coli* BL21-Gold (DE3) cells and most beneficial variants were identified after screening of hydrolase clones (at least 360 clones per hydrolase) in 96-well MTP format. Confocal microscopy images showing an overlay of transmission and fluorescence were recorded in order to show the difference in the fluorescent signal of *E. coli* BL21-Gold (DE3) cells producing either active hydrolases (pET22b(+)-pNBEBL, BSLA, or CelA2) or cells harbouring pET22b(+) (Fig. 2) (see ESI†). For *E. coli* BL21-Gold (DE3) cells expressing active hydrolase variants (Fig. 2D: esterase pNBEBL; 2E: lipase BSLA; 2F: cellulase CelA2) a strong fluorescent signal was detected which confirms hydrogel formation through incorporation of Polyfluor 570. *E. coli* BL21-Gold (DE3) cells harbouring pET22b(+) (Fig. 2A-C) showed little to no fluorescence. In Fig. 3 flow cytometry analysis of Fur-Shell labelled *E. coli* BL21-Gold (DE3) cells are shown and confirm the visual impression of recorded confocal microscopy images. Comparison of Fig. 3 (A/D; B/E; C/F) shows a significant...
difference in fluorescence intensity among E. coli cells expressing active hydrolases and those harbouring an empty vector. The fluorescent intensity of E. coli cells harbouring an empty vector was around 20-fold lower compared to E. coli cells expressing wildtype hydrolases (Fig. 3D: esterase pNBEBL; 3E: lipase BSLA; 3F: cellulase CelA2). One epPCR random mutagenesis library per hydrolase gene (esterase pnbegl, lipase bsla and cellulase cela2) was generated using 0.1 or 0.2 mM MnCl₂ with an average mutation frequency of 2.2 (pnbegl, 0.1 mM MnCl₂), 13.2 (bsla, 0.2 mM MnCl₂), and 4.2 (cela2, 0.2 mM MnCl₂) mutations per kb (see ESI†). The populations of all three mutant libraries were analysed by flow cytometry and showed a reduced fluorescent signal due to the presence of inactive hydrolase variants (Fig. 3G: esterase pNBEBL; 3H: lipase BSLA; 3I: cellulase CelA2). The sorting gate P1 was set to minimize background fluorescent signal (0.9%, see Fig. 3A/B/C) from E. coli BL21-Gold (DE3) cells harbouring an empty vector (Y-axis). This means >99% of the negative, non-fluorescent population is excluded in the sorting gate and in return solely positive, active cells are sorted. Additionally, P1 was adjusted using calibration beads with the forward scatter laser (X-axis) to contain only single cells being smaller than 3 µm in size in order to exclude associated cells [17]. E. coli BL21-Gold (DE3) cells expressing active hydrolase variants were sorted and collected (Fig. 1, Step 3). DNA of 5×10⁴ sorted E. coli cells was recovered by using option 4b (Fig. 1, Step 4b): plasmid isolation to maintain diversity and to compensate for the survival rate (7 %) [8]. Subsequently, isolated plasmids were transformed into E. coli BL21-Gold (DE3) cells and plated on LB-agar (see ESI†). Variants of the three hydrolases (in total 2160 clones) were transferred into 96-well MTPs for subsequent screening. In case of pNBEBL the identical substrate (β-D-(+)-glucose pentaacetate) was employed for flow cytometry and MTP screening (see principle and procedure in ESI†); in case of CelA2 the similar substrate 4-methylumbelliferyl-β-D-celllobioside (4-MUC) was employed, and in case of the BSLA p-nitrophenyl acetate (pNPA) was used due to better sensitivity and established screening procedures in 96-well MTP format [18, 19]. Enrichment factors were determined by activity measurements and calculated by dividing the percentage of active populations after and before flow cytometry sorting (number of clones analysed before sorting: 180 variants per hydrolase; number of clones analysed after flow cytometry sorting: 360 variants (pNBEBL) or 900 variants per hydrolase (BSLA; CelA2)). In detail, enrichment was efficient for the CelA2 population (1.7-fold enrichment, 38% active population in sorted libraries). In case of pNBEBL (1.3-fold enrichment, 41% active population in sorted libraries) and BSLA (1.5-fold enrichment, 44% active population in sorted libraries) low enrichment was obtained which was unexpected due to the adjustment of the P1 sorting gate. A decrease reaction time -hydrogel formation from 10 min to 2 min, as well as decrease in monomer concentration did not yield a further improvement in enrichment factors. Nevertheless, in all cases a sufficiently high number of active hydrolase variants was obtained for 96-well MTP screening. Two improved variants out of rescreening of 8 variants for pNBEBL, 14 variants for BSLA and 8 variants for CelA2 in MTP format were kinetically characterized in detail and compared to the corresponding wildtypes (Fig. S1; Table 1). The esterase pNBEBL variant E1 (E256G, G401V) showed an impressive 7-fold higher kₐ and 2-fold lower Kᵢₐ value compared to wildtype pNBEBL, indicating a high specific activity. In E1 (E256G, G401V), the E256G substitution has not been reported yet and is located on surface of E. coli BL21-Gold (DE3) cells were incubated with β-D-(+)-glucose pentaacetate (A/B/D/E/G/H) or cellobiose (C/F/I) as substrates. A/B/C: E. coli BL21-Gold (DE3) pT22b(+) cells were used as a negative control to determine the threshold fluorescence. D/E/F: E. coli BL21-Gold (DE3) pT22b(+) pNBEBL, BSLA, CelA2 wildtype cells were used as a positive control. Random mutagenesis libraries were generated with epPCR: 0.1 mM MnCl₂; pNBEBL (G) 0.2 mM MnCl₂ for BSLA (H), and 0.2 mM MnCl₂ for CelA2 (I).
Table 1: Kinetic characterization of pNBEBL, BSLA and CelA2 was performed regarding $K_m$, $k_{cat}$ and U/mg. pNBEBL was kinetically characterized at 22°C with the fluorometric Amplite screening system (see Fig. S2A ESI†). BSLA was kinetically characterized with the colourimetric substrate pNPA (at 22°C, pH 6.5) (see Fig. S2B ESI†). CelA2 was kinetically characterized with the fluorogenic substrate 4-MUC (at 30°C, pH 7.2) (see Fig. S2C ESI†). Units are defined as: 1 U of pNBEBL catalyzes the conversion of 1 mmol β-D-(+)-glucose pentaacetate per minute, 1 U of BSLA catalyzes the conversion to 1 µmol 4-MU per minute, 1 U of CelA2 catalyzes the conversion to 1 µmol 4-MU per minute.

<table>
<thead>
<tr>
<th>Enzyme (substitution)</th>
<th>$K_m$ [µM]</th>
<th>$k_{cat}$ [1/s]</th>
<th>U/g [µmol/min*mg]</th>
</tr>
</thead>
<tbody>
<tr>
<td>pNBEBL wildtype</td>
<td>540±170</td>
<td>26.1±2</td>
<td>481±24</td>
</tr>
<tr>
<td>E1 (E256G, G401V)</td>
<td>270±50</td>
<td>186±4</td>
<td>3415±68</td>
</tr>
<tr>
<td>E2 (F313S)</td>
<td>300±100</td>
<td>169±7</td>
<td>3106±107</td>
</tr>
<tr>
<td>BSLA wildtype</td>
<td>1259±92</td>
<td>3.4±0.1</td>
<td>10.7±0.8</td>
</tr>
<tr>
<td>L1 (Y130D)</td>
<td>1893±241</td>
<td>4.3±0.2</td>
<td>13.4±0.7</td>
</tr>
<tr>
<td>L2 (R60Q, S166G)</td>
<td>1807±243</td>
<td>4.2±0.2</td>
<td>13.1±0.7</td>
</tr>
<tr>
<td>CelA2 wildtype</td>
<td>170±8</td>
<td>0.15±0.2</td>
<td>0.14±0.02</td>
</tr>
<tr>
<td>C1 (V37A, E275G, E398V)</td>
<td>189±5</td>
<td>0.26±0.3</td>
<td>0.22±0.03</td>
</tr>
<tr>
<td>C2 (N135S)</td>
<td>172±9</td>
<td>0.26±0.2</td>
<td>0.23±0.03</td>
</tr>
</tbody>
</table>

CelA2. Position E275 is located on the surface between two β-sheets, position E398 is located in the middle of a α-helix, and position V37 is not covered by the homology model which was generated on the basis of a glycosyl hydroxylase family 9 (GH9) cellobiosidase from *Clostridium cellulovorans* (PDB ID: 3RX7). None of the three substituted positions in CelA2 variant C1 have been reported yet. Compared to standard directed enzyme evolution campaigns (low mutagenesis frequency, 1200-15000 variants screened per round; 1.5-2.5-fold improved activity [21]) the obtained campaigns (low mutagenesis frequency, 1200-15000 variants) are of outstanding success (Table 1). Units are defined as: 1 U of pNBEBL catalyzes the conversion of 1 mmol β-D-(+)-glucose pentaacetate per minute, 1 U of BSLA catalyzes the conversion to 1 µmol 4-MU per minute, 1 U of CelA2 catalyzes the conversion to 1 µmol 4-MU per minute.

Notes and references