Chemistry & Biology, Vol. 12, 1291–1300, December, 2005, ©2005 Elsevier Ltd All rights reserved. DOI 10.1016/j.chembiol.2005.09.016

High-Throughput Screening of Enzyme Libraries: In Vitro Evolution of a β-Galactosidase by Fluorescence-Activated Sorting of Double Emulsions

Enrico Mastrobattista,[1,3,4] Valerie Taly,[1,2,3,5] Estelle Chanudet,[1] Patrick Treacy,[1,2,5] Bernard T. Kelly,[1] and Andrew D. Griffiths[1,2,5,*]
[1] Medical Research Council
Laboratory of Molecular Biology
Hills Road
Cambridge CB2 2QH
United Kingdom

Summary

We describe a completely in vitro high-throughput screening system for directed evolution of enzymes based on in vitro compartmentalization (IVC). Single genes are transcribed and translated inside the aqueous droplets of a water-in-oil emulsion. Enzyme activity generates a fluorescent product and, after conversion into a water-in-oil-in-water double emulsion, fluorescent droplets are sorted using a fluorescence-activated cell sorter (FACS). Earlier in vivo studies have demonstrated that Ebg, a protein of unknown function, can evolve to allow *Escherichia coli* lacking the *lacZ* β-galactosidase gene to grow on lactose. Here we demonstrate that we can evolve Ebg into an enzyme with significant β-galactosidase activity in vitro. Only two specific mutations were ever seen to provide this improvement in Ebg β-galactosidase activity in vivo. In contrast, nearly all the improved β-galactosidases selected in vitro resulted from different mutations.

Introduction

Directed evolution is a powerful technique for the engineering of enzymes. It is based on mimicking nature by performing iterative rounds of mutation/recombination and selection in the laboratory [1].

All evolutionary systems, whether in nature or in the laboratory, require a link between genotype (a nucleic acid that can be replicated) and phenotype (a functional trait such as a binding or catalytic activity). In addition, there has to be selective pressure for the desired phenotype.

In the laboratory, genotype-phenotype linkage is frequently achieved by physically linking genes to the proteins they encode by a variety of techniques, including display on phage, viruses, bacteria, and yeast, plasmid display, ribosome display, CIS display, and mRNA-peptide fusion. These "display technologies" have proven highly successful in the generation of binding proteins but have met with much less success selecting enzymes. Indirect selections—by binding to transition state analogs or enzyme inhibitors—have generally failed to produce potent catalysts [2]. Single-turnover, intramolecular selections of enzymes displayed on phage were demonstrated but these impose severe limitations, not least because selection is not directly for catalysis [2, 3]. To evolve proficient enzymes, the selection (or screen) should be simultaneous and direct for all enzymatic properties: substrate recognition, formation of a specific product, rate acceleration, and turnover.

In nature, linkage of genotype to phenotype is achieved by compartmentalization of genes in cells, and the selection of cells for a desired phenotype has also been used very successfully for directed evolution of proteins. Unfortunately, in vivo selections are usually, but not always (e.g., [4]), restricted to functions that affect the viability of the organism and are often complicated by the complex intracellular environment. In addition, there is often no selection system available for the desired activity.

Hence, directed evolution of enzymes frequently relies, not on selection, but on high-throughput screening (HTS) using chromogenic or fluorogenic substrates. However, when screening colonies on agar plates, or individual clones in microtiter plate wells, typically 10^3–10^4 clones, and rarely more than 10^5 clones, can be screened, even using sophisticated automated systems [3].

One way of greatly accelerating HTS is to use fluorescence-activated cell sorting (FACS), which can routinely sort >10^7 clones per hour, and has a series of other advantageous features [5]. FACS has already proven a highly successful technique to select proteins (notably antibodies) with high binding affinities [6–13]. In addition, FACS has significant potential to select for catalysis [3, 14]; however, so far, this approach has only been possible when the diffusion of product out of the cell can be restricted (e.g., [15]), or the product can be captured on the surface of the cell [16, 17], or onto microbeads [18].

We have developed a method for directed evolution, termed in vitro compartmentalization (IVC), which uses compartmentalization to link genotype and phenotype. However, instead of compartmentalizing genes in cells, as in nature, in IVC the genes are compartmentalized in aqueous microdroplets dispersed in a water-in-oil (w/o) emulsion [19]. IVC has been used to select proteins for catalysis [18, 20–24], binding [25–28], and regulation [29]. The common feature of previous IVC selections is that the product always ends up linked to the gene, either directly or via a microbead, and it is the genes themselves that are selected after breaking of the emulsion.

We have previously demonstrated that fluorescent aqueous droplets in w/o emulsions can be sorted using FACS and can be recovered intact without losing their contents if the primary w/o emulsion is first converted into a w/o/w double emulsion [30]. Here we demonstrate

*Correspondence: griffiths@isis.u-strasbg.fr
[2] Lab address: http://www-isis.u-strasbg.fr/lbc
[3] These authors contributed equally to this work.
[4] Present address: Utrecht Institute for Pharmaceutical Sciences (UIPS), Department of Pharmaceutics, Sorbonnelaan 16, 3584 CA Utrecht, The Netherlands.
[5] Present address: Institut de Science et d'Ingénierie Supramoléculaires (ISIS), Université Louis Pasteur, CNRS UMR 7006, Laboratoire de Biologie Chimique, 8 allée Gaspard Monge, BP 70028, F-67083 Strasbourg Cedex, France.

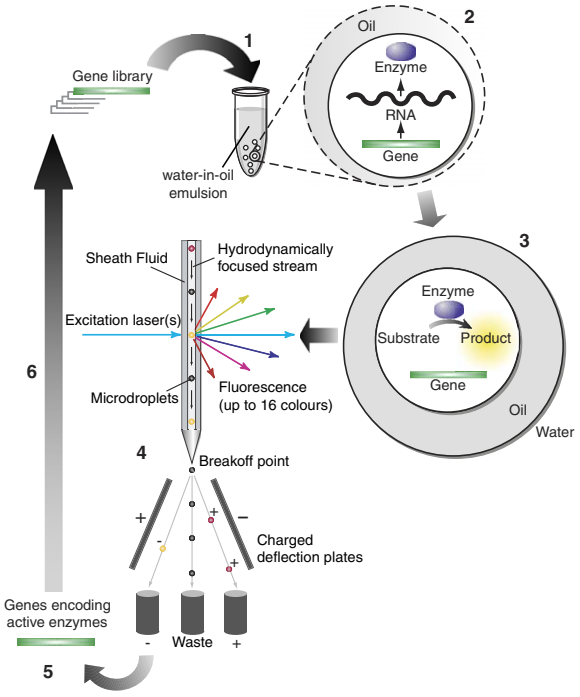

Figure 1. Selection of Double Emulsion Microdroplets Using a Fluorescence-Activated Cell Sorter (FACS)

An in vitro transcription/translation reaction mixture containing a library of genes encoding mutant enzymes is dispersed to form a water-in-oil (w/o) emulsion with typically one gene per aqueous microdroplet (1). The genes are transcribed and translated within their microdroplets (2). Proteins with enzymatic activity convert the nonfluorescent substrate into a fluorescent product and the w/o emulsion is converted into a water-in-oil-in-water (w/o/w) emulsion (3). Fluorescent microdroplets are separated from nonfluorescent microdroplets (or microdroplets containing differently colored fluorochromes) using a fluorescence-activated cell sorter (FACS) (4). Genes from fluorescent microdroplets, which encode active enzymes, are recovered and amplified using the polymerase chain reaction (5). These genes can be recompartmentalized for further rounds of selection (6).

that we can formulate double emulsions to create a completely in vitro double emulsion selection system based on IVC (Figure 1) in which the unbroken emulsion droplets are sorted, together with the genes they contain, using FACS. Single genes are transcribed and translated in vitro, compartmentalized in aqueous droplets in w/o emulsions together with a fluorogenic substrate. After conversion into a w/o/w double emulsion, fluorescent droplets are sorted using FACS.

The technique relies on using fluorogenic substrates, which are available for many different enzymes, and there is no need for the product of the enzymatic reaction to be coupled to genes, cells, or microbeads to allow selection.

We have used this double emulsion selection system for the in vitro evolution of the Ebg enzyme of *Escherichia coli*. The Ebg system has been used extensively over the past 3 decades as a model to study the evolution of novel enzyme functions in vivo [31–33]. These studies demonstrated that Ebg, an *E. coli* protein of unknown function and possessing negligible β-galactosidase activity, can evolve into an active β-galactosidase that can replace the function of the LacZ β-galactosidase. All mutants which could grow on lactose contained either a D92N substitution (class I mutants), a W976C substitution (class II mutants), or a combination of the two substitutions (class IV mutants).

In vitro evolution of Ebg using double emulsion IVC generated β-galactosidases which contained completely different mutations from those seen in all improved Ebg β-galactosidases evolved in vivo.

As described in the accompanying manuscript [34], FACS sorting of double emulsions also allows the selection of single bacterial cells which are compartmentalized in the internal aqueous droplets together with a fluorogenic substrate. This technique was used for the directed evolution of serum paraoxonase (PON1) with improved thiolactonase activity.

FACS of double emulsions, therefore, has the potential to greatly expand the range of enzymes that can be screened, using both in vitro and in vivo systems.

Results

Preparation and Characterization of Double Emulsions

In order to select as outlined in Figure 1, the emulsions should: (1) allow efficient transcription and translation; (2) be stable throughout the selection procedure (including flow sorting); (3) compartmentalize the gene-encoded proteins and the fluorescent reaction product; and (4) the majority of oil droplets should contain no more than a single aqueous droplet to prevent coselection of genes that do not encode for active β-galactosidases.

We have previously demonstrated that mineral oil-based w/o emulsions prepared with 10% aqueous volume fraction, 4.5% Span 80, and 0.48% Tween 80 can be re-emulsified with an equal volume of water phase (2% Tween 20 in PBS) to generate w/o/w double emulsions which can be sorted by FACS, while the contents of the aqueous droplets of the primary w/o emulsion remain intact [30].

However, although mineral oil-based w/o emulsions prepared with 4.5% Span 80 and 0.48% Tween 80 [30], or with 4.5% Span 80 and 0.5% Triton X-100 [18], allowed transcription and translation (data not shown), a fluorescence-based aqueous contents exchange assay [35] showed rapid exchange of small molecules between the aqueous droplets of w/o emulsions (see Figure S1 in Supplemental Data available with this article online).

We therefore investigated a variety of other formulations. The most suitable involved first making a w/o emulsion comprising the in vitro coupled transcription/translation mix (which forms the inner aqueous phase) and 1% (w/v) Span 60 and 1% (w/v) cholesterol in decane (which forms the oil phase). The w/o emulsion was then re-emulsified with 0.5% (w/v) Tween 80 in PBS to create a w/o/w emulsion with a continuous aqueous outer phase of PBS (Figure 2). The fluorescence-based aqueous contents exchange assay [35] showed little exchange of small molecules between the aqueous droplets (<4% after 2 hr incubation at 37ºC) (Figure S1).

To confirm that genes, the proteins they encode, and the fluorescent product are retained within the emulsion

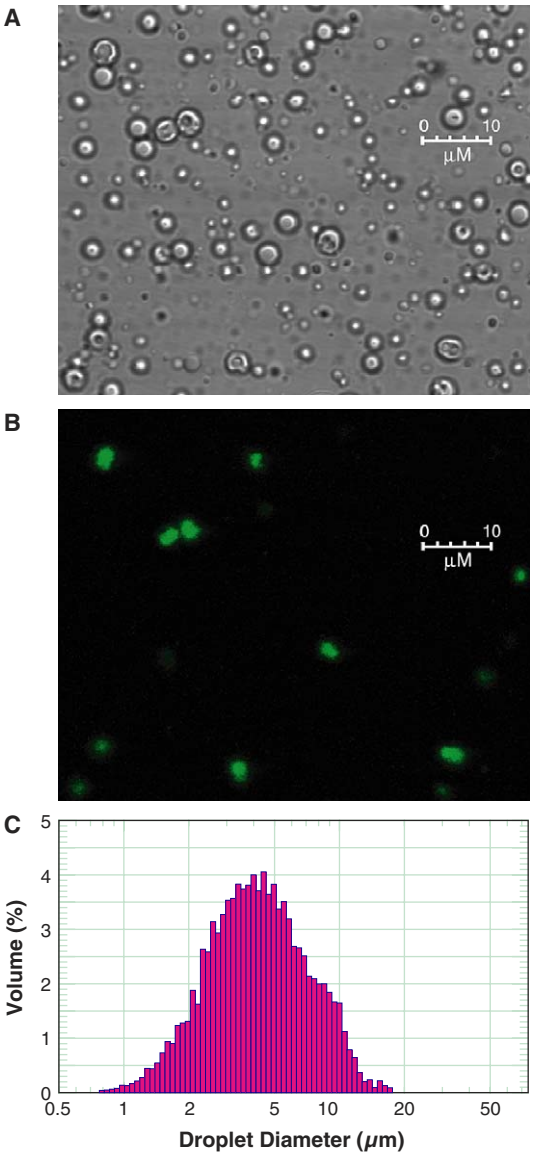

Figure 2. Water-in-Oil-in-Water Double Emulsions

Internal water droplets contain 4 μM fluorescein dextran (mw 500,000) and 30 μM BSA in PBS.

(A) A differential interference contrast micrograph of a double emulsion.

(B) A confocal epifluorescent micrograph of the same emulsion. Size bars represent 10 μm.

(C) The size distribution of aqueous droplets in the w/o/w double emulsion determined using single-particle optical sensing (by percentage of total volume).

The mean diameter of the oil droplets in the double emulsions is 4.7 μm.

droplets during the sorting process, a model selection for β-galactosidase activity was performed as described in Figure 1 using a fluorogenic substrate, fluorescein di-β-D-galactopyranoside (FDG). *lacZ* genes (encoding *E. coli* β-galactosidase) were enriched from a pool of frameshifted *lacZmut* genes (encoding an inactive protein). The *lacZ* genes were enriched 138-fold from a 1:1000 molar ratio of *lacZ*:*lacZmut* genes (Figure S2; Table S1).

Table 1. Kinetic Analysis of LacZ, Wild-Type Ebg, and Mutated Variants Using FDG

Enzyme	Relative k_{cat} (RFU s^{-1})		K_M (μM)		k_{cat}/K_M (RFU s^{-1} mM^{-1})
wt	N.D.		>4000		0.0315
EbgMut2a-C10	0.68	±0.06	24	±8	28.3
EbgMut2a-D1	0.42	±0.06	7.9	±5	53.2
EbgMut2a-D3	0.66	±0.04	34	±6	19.4
EbgMut2a-D5-2	0.29	±0.01	29	±5	10.0
EbgMut2a-D6	0.43	±0.04	17	±5	25.3
EbgMut2b-E4	0.44	±0.02	21	±2	21.0
EbgMut2b-E7-2	0.56	±0.03	28	±5	20.0
EbgMut2b-E8	0.43	±0.02	24	±4	17.9
Class I	1.8	±0.05	510	±26	3.53
Class II	1.5	±0.08	260	±30	5.77
Class IV	3.6	±0.44	29	±12	124
wt (Δ*ebgC*)	0.49	±0.06	15	±5	32.7
Class I (Δ*ebgC*)	0.74	±0.05	12	±3	61.7
Class II (Δ*ebgC*)	0.54	±0.04	110	±22	4.91
Class IV (Δ*ebgC*)	1.0	±0.05	53	±9	18.9
LacZ	187	±8.5	150	±16	1250

RFU, relative fluorescence units. N.D., not determined.

Distinguishing Differences in β-Galactosidase Activity by Flow Cytometry of Double Emulsions

Genes encoding several β-galactosidases with different catalytic efficiencies were used to assess the dynamic range of the system. Wild-type Ebg has negligible β-galactosidase activity with FDG as substrate. Ebg class I and class II mutants both have significant and similar levels of activity (with class II marginally the more efficient). Ebg class IV mutants hydrolyze FDG more efficiently again, but are still significantly less efficient than *E. coli* LacZ (Table 1).

Linear DNA constructs encoding wild-type Ebg, class I mutant, class II mutant, class IV mutant [36], or LacZ were entrapped inside aqueous compartments of decane w/o emulsions together with an in vitro translation system and FDG. After a 2 hr incubation at 30ºC, the w/o emulsions were converted into w/o/w double emulsions and analyzed by flow cytometry (Figure 3). Under these conditions, there are large fluorescence differences between wild-type Ebg, Ebg class I, Ebg class IV,

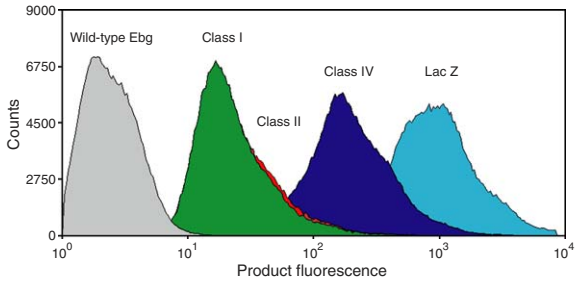

Figure 3. Discriminating Wild-Type Ebg, Class I, Class II, and Class IV Mutants, and LacZ in Double Emulsions Using Flow Cytometry

Genes encoding wild-type, class I, class II, or class IV Ebg and LacZ were translated for 2 hr at 30ºC cocompartmentalized in a w/o emulsion with the fluorogenic substrate FDG and 7-hydroxycoumarin-3-carboxylic acid. The sorter was triggered on coumarin fluorescence, thereby ignoring all o/w droplets (i.e., droplets lacking an internal aqueous phase). The levels of product fluorescence in oil droplets which contain internal aqueous droplets (gated on coumarin fluorescence) are plotted as histograms (250,000 events).

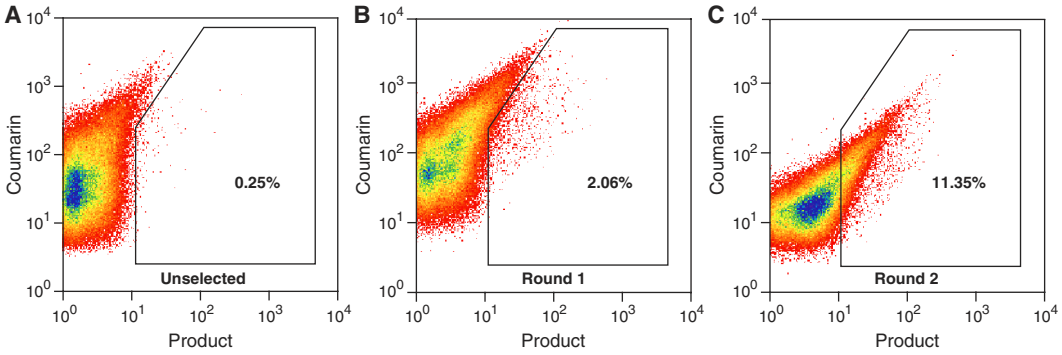

Figure 4. Selection of *ebg* Libraries

A randomly mutated library of *ebgAC* genes was selected as in Figure 1 using FDG as substrate. 7-hydroxycoumarin-3-carboxylic acid was cocompartmentalized with the genes and the levels of coumarin fluorescence and product fluorescence are displayed as dot-plots for each selection round. Only double emulsion compartments which contained 7-hydroxycoumarin-3-carboxylic acid were analyzed, allowing us to consider only w/o/w droplets (i.e., droplets having an internal aqueous phase). Sort gates were set in such a way that less than 0.01% of the double emulsion compartments of a negative control (double emulsions without DNA) was collected and nearly all of the double emulsion compartments of a positive control (Ebg class IV mutant). For each round of selection, 10^5 positive events were collected.

(A) Unselected library.

(B) Library after one round of selection.

(C) Library after two rounds of selection.

and LacZ, but the Ebg class I and class II mutants cannot be resolved.

Selection of *ebg* Gene Libraries for β-Galactosidase Activity

Constructs containing both the wild-type *ebgA* and *ebgC* genes, encoding the α domain and β domain of the Ebg enzyme, were randomly mutated by error-prone PCR [37], creating a library (EbgMut1) with an average of 4 ± 2 mutations per gene, of which 96.7% were transitions.

The library was selected as in Figure 1 using FDG as substrate. For each selection round, 10^5 positive events were collected and the DNA was amplified by PCR. Flow cytometry analysis of the double emulsion before sorting and after the first and second selection rounds are shown in Figure 4. Before selection, the percentage of positive droplets is very low (0.25%) but increases with each selection round up to 11%.

After two rounds of selection, the library was further mutated [37] to generate a new library (EbgMut2a) with an average of 9 ± 2 mutations per gene, and recombined by DNA shuffling [38], generating EbgMut2b, which contains 12 ± 3 mutations per gene. The libraries were selected for two rounds as above, but using more stringent sort gates.

Analysis of Selected Mutants from the First Round of Directed Evolution

Three hundred sixty-six clones from the selected EbgMut1 library were in vitro translated and assayed for FDG hydrolysis in 384-well plates using 0.1 mM FDG. Thirty percent had initial rates higher than wild-type Ebg enzyme, from which 50% were comparable to the class I mutant and 10% to the class II mutant. None of the clones had activity higher than the class II mutant. Sequencing of the 20 mutants with the highest activity revealed that 13 had mutations in *ebgC* leading to the premature termination of translation of the β subunit (Table 2, A). Moreover, two of these variants contained no mutations in the *ebgA* gene encoding the α subunit. A class I mutation (D92N) was observed in one variant but no class II mutations (W976C) were found. However, four variants containing an S975P mutation adjacent to the class II mutation were selected. The fact that the W976C amino acid substitution was not observed may be due to the bias for transition in the libraries because only a transversion event could generate this substitution.

Analysis of Selected Mutants from the Second Round of Directed Evolution

After selection, 1536 clones from the EbgMut2a (mutated) library, and 550 clones from the EbgMut2b (recombined) libraries, were in vitro translated and assayed for FDG hydrolysis as above. Sixty-eight percent of the EbgMut2a clones, and 41% of the EbgMut2b clones, had activities higher than wild-type Ebg. The majority of variants had activities higher than, or comparable to, the class I mutant (90% for EbgMut2a) and the average activity of the clones was significantly higher than those from the selected EbgMut1 library.

Individual clones (4% for each library) were selected based on their initial rates at 0.1 mM FDG and assayed to evaluate their V_{max} and K_M values. Forty-three percent of these clones had V_{max}/K_M values comparable to or higher than class II and 28 were used for further analysis. Fifty percent contained an early stop codon in the *ebgC* gene (Table 2, B). Of these, one third also contained the S975P mutation. Two other mutations were also observed frequently: S741C and N304D.

The presence of so many variants with frameshifts or stop codons in *ebgC* is partly explained by kinetic studies of wild-type Ebg and the class I, II, and IV mutants in which the *ebgC* gene is deleted (Table 1). Unlike wild-type Ebg, Ebg lacking the β domain encoded by *ebgC* shows significant FDG hydrolysis activity with a k_{cat}/K_M

Table 2. Sequence of Selected Ebg Mutants

	ebgA	*ebgC*
A1		
EbgMut1-A3	H73R/A458V/S691P	E50Stop
EbgMut1-B9	V243A/T272A/Y824C/C1003R	Q76Stop
EbgMut1-A8	none	I45T/L63F/FS77/88Stop[a]
EbgMut1-B6	none	I13T/Y100Stop
EbgMut1-E10	P17S/T1006A	Q9Stop
EbgMut1-H6	K301R	I13T/Y100Stop
EbgMut1-D11	V986I	Q76Stop
EbgMut1-C8	Q29R/I211V/F764L/F1019L	S16P/S44P/Y100Stop
EbgMut1-B3	E222G/I537P	T46Stop
EbgMut1-D6	V201A/Y502H/N573S	W20Stop
EbgMut1-A7	M79V/V280A/K436R/Y731C/**S741G**/N944S	Q21Stop
EbgMut1-H5	**Q89R**	Q73Stop
EbgMut1-G11	**Q89R**	E8Stop
EbgMut1-A12	H461R/K761R/**S975P**/F988L	A123V/FS135/182Stop[a]
A2		
EbgMut1-C1	V166A/**S975P**	V110A
EbgMut1-F11	W889C/**S975P**	NO
EbgMut1-G3	C226N/T623A/T624A/F764L/**S975P**	NO
EbgMut1-G9	A465T/E1027G	V132A
EbgMut1-C12	T443I	I41T/P66I
EbgMut1-G8	*D92N*/H783N/I811T/Q1030H	M1L/I3V
B1		
EbgMut2a-C7	D54A/N835D/**S975P**	FS77/88Stop[a]
EbgMut2a-D5-2	A34T/T228A/E281G/A916T/N1028S	Q76stop
EbgMut2a-F10	T703A	W20Stop
EbgMut2a-F11-2	I7F/H51R/A69V/I735V/I849V	Q12Stop
EbgMut2a-D6	T167A/N236S/**N304D**/I329V/**S741C**/**S975P**	FS77/88Stop[a]
EbgMut2a-D3	K422Q/**S741C**/**S975P**	FS77/88Stop[a]
EbgMut2a-D1	A158T/V201A/P698S/Q955R/**S975P**	Q21Stop
EbgMut2a-G11	**E498K**/I582V/H605R/A830V	R22H/R63Stop
EbgMut2a-B5	M335T/L719P	W20Stop
EbgMut2b-E3-2	W13S/F103S/N702D/N908S/**S975P**/S1026G	D5G/T46P/Y65P/Y100Stop
EbgMut2b-E7-2	**S741C**/**S975P**	C23R/R63Stop
EbgMut2b-D11	N237S/**N304D**/I329V/W826C/D864E	W20Stop
EbgMut2b-E4	*D92N*/D302G/**N304D**/A867V/G1018V	Q9Stop
EbgMut2b-A8	D97G/E639D/T745A/D793G/M857V	Q73Stop
B2		
EbgMut2a-C10	**E498K**/**Q881R**/**S975P**	**F10S**/**C23R**/**E90G**
EbgMut2a-D5	S39R/**Q282R**/E476G/**E498K**/**Q881R**/**S975P**	**F10S**/**C23R**
EbgMut2a-F9	**N2S**/**Q282R**/C452R/**E498K**/**Q881R**/S975P	**F10S**/**C23R**
EbgMut2a-C4	**N2S**/W123S/**Q282R**/**E498K**/**Q881R**/**S975P**	**F10S**/**C23R**
EbgMut2a-F7	**N2S**/**Q282R**/**E498K**/C725Y/**Q881R**/**S975P**	**F10S**/**C23R**
EbgMut2a-C11	**N2S**/**Q282R**/**E498K**/**Q881R**/**S975P**	**F10S**/**C23R**
EbgMut2a-C5	**N2S**/**Q282R**/**E498K**/**Q881R**/**S975P**	**F10S**/**C23R**
EbgMut2a-C8	**Q282R**/**E498K**/V727I/**Q881R**/**S975P**	**F10S**/**C23R**
EbgMut2a-G9	C225R/**S975P**	L45S/T96I/N130D
EbgMut2a-G7	K270R/V334A/M766V/N782S/**S975P**	NO MUTATION
EbgMut2a-G4	**N2S**/**Q282R**/**E498K**/**Q881R**/M911Y/**S975P**	**F10S**/**C23R**
EbgMut2a-B2	I116T/N331D/Q366R/P494L/E701D/K743R/T918A	D36G/Y46H/D56G/I127T/T40A
EbgMut2b-D8	**N2S**/V104A/**Q282R**/**E498K**/R692H/D722G/**Q881R**/**S975P**	**F10S**/**C23R**/K83R
EbgMut2b-E8	F383L/T398A/**Q881R**/**S975P**	**F10S**/K18R/**C23R**/C118Y

(A) First-generation mutants.
(B) Second-generation mutants.
(A1 and B1) Mutants with an *ebgC* gene containing a deletion or a point mutation leading to a stop codon.
(A2 and B2) Mutants containing a full-length *ebgC*. Amino acid positions found mutated more than once are shown in bold. The D92N mutation, observed in previous in vivo studies, is in italics.
[a] Location of frameshift (FS) and resulting stop codon (Stop).

higher than the class I and II mutants. This is mainly due to a large decrease in K_M, which is also observed in the class I mutant with *ebgC* deleted.

In the clones encoding a full-length β domain, some mutations appeared with high frequency. In the α subunit, 13 variants contained the S975P mutation, 11 of which contained an additional Q881R mutation, 10 an E498K mutation, 9 a Q282R mutation, and 7 an N2S mutation. Interestingly, three mutations were also conserved in the β subunit: 11 mutants contained the mutations F10S and C23R.

Eight mutants were selected for further kinetics analysis (Table 1). They all had k_{cat}/K_M values at least 300-fold higher than wild-type Ebg, which showed very little activity with FDG as substrate and had a very high K_M of >4 mM. The catalytic efficiencies (k_{cat}/K_M) of all eight

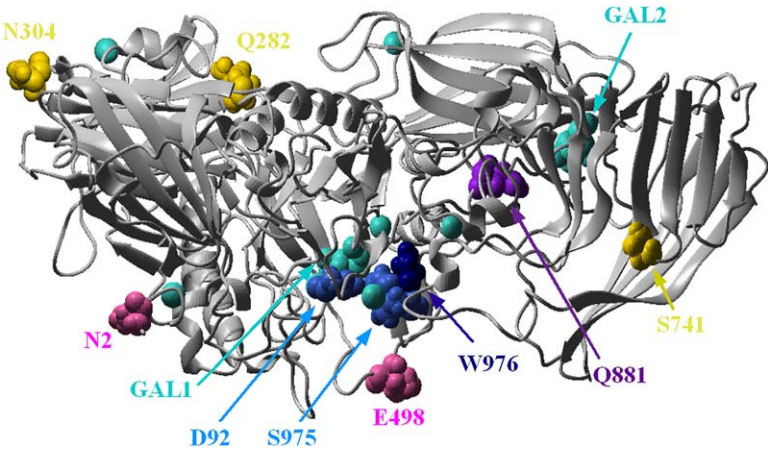

Figure 5. Model Structure of EbgA β-Galactosidase

The model was constructed by homology modeling using the structure of the LacZ β-galactosidase (Protein Data Bank code 1JZ7), with galactose in the binding site, as template [39]. The alignment between *ebgA* and the *lacZ* genes was recovered from the blast search that identified the template. The alignment of *lacZ* and the genes belonging to the same family as *ebgA* given by the ExPASy website (see http://us.expasy.org/cgi-bin/niceprot.pl?P06864) was used to define the highly conserved residues, which were all correctly aligned in the pairwise alignment. The residues of the model superimpose well with those of the template (rmsd = 0.67 Å), in particular those of the active site. One α subunit is shown in cartoon representation. The magnesium ions, sodium ions, and galactoses are shown in cyan. The model shows the location of residues which were frequently substituted in selected variants. Residues close to the galactose in the active site (less than 5 Å) are shown in blue and W976 in dark blue. The other residues are at the surface of the monomer: those at the interface between monomers are shown in pink (N2 and E498) and those facing the solvent in yellow (Q282, N304, S741). Residue Q881 is located less than 10 Å from the second galactose (which is not in the active site) and is shown in violet.

mutants were between those of class II and class IV mutants. All eight mutants had very similar relative k_{cat} values of around 0.5 RFU s^{-1}, slightly lower than the class I, II, and IV mutants, but with low K_M values of around 20 μM, comparable to class IV. This is consistent with the relatively low substrate concentrations used for selection and screening (0.5 mM and 0.1 mM FDG, respectively). A model of the EbgA subunit was constructed by homology modeling using the structure of the LacZ β-galactosidase [39] (Figure 5). All the frequently mutated amino acids in *ebgA*, except S975, seem to be on the surface of the monomer, suggesting a role in interaction between the subunits.

Discussion

We have developed a completely in vitro high-throughput screening (HTS) strategy for directed evolution of enzymes. This method is based on in vitro compartmentalization (IVC) of genes in aqueous microdroplets dispersed in a water-in-oil (w/o) emulsion [19]. IVC has previously been used to select proteins for catalysis [18, 20–24], binding [25–28], and regulation [29] using a format in which the product ends up linked to the gene, either directly or via a microbead, and it is the genes themselves that are selected after the breaking of the emulsion.

Here we describe an adaptation of the IVC system in which it is the unbroken emulsion droplets which are sorted by FACS, together with the genes they contain (Figure 1). The technique relies on using fluorogenic substrates, which are available for many different enzymes and which do not need to be coupled to the gene to allow selection. In order to allow the sorting of fluorescent droplets by FACS, the primary w/o emulsion is converted into a w/o/w double emulsion [30]. This technique greatly expands the range of enzymes that can be selected using IVC. We have demonstrated the use of this double emulsion selection system for the in vitro evolution of a β-galactosidase.

The evolution of β-galactosidase genes in *Escherichia coli* has proven to be an excellent model system to dynamically study the evolution of novel enzyme functions in vivo (for reviews, see [31–33]). Three decades ago, it was demonstrated that *E. coli* strains from which the *lacZ* β-galactosidase gene had been deleted (Δ*lacZ* strains) could evolve under selective pressure into variants that can grow with lactose as the only carbon source [40, 41]. Further studies revealed that this is due to the evolution of the *ebg* operon. The *ebg* operon consists of four genes: *ebgR* encodes a repressor that regulates the expression of *ebgABC*; *ebgB* encodes a protein whose function is unknown but which is structurally related to transport proteins; *ebgA* encodes the 118 kDa α subunit and *ebgC* the 20 kDa β subunit of the $\alpha_4\beta_4$ Ebg enzyme, the natural function of which is unknown. Wild-type Ebg enzyme has 31% protein homology with LacZ β-galactosidase but is a very feeble β-galactosidase and cannot support growth of Δ*lacZ* strains on lactose, even if the enzyme is constitutively expressed at levels up to 5% of total cellular protein [31].

Selection for spontaneous lactose-utilizing mutants from Δ*lacZ ebgR*$^-$ (constitutive) strains results in mutations in *ebgA* and produces only two classes of mutants: all class I mutants have a D92N substitution, while all class II mutants have a W976C substitution. The combination of the two substitutions (class IV mutants) further increases catalytic efficiency and expands the substrate range of the enzyme.

EbgA is part of an ancient clade of functional β-galactosidase that diverged from the paralogous LacZ β-galactosidase over 2 billion years ago [33]. In that time, EbgA has acquired a new, and unknown, function. EbgA differs from other members of its clade at only 2 of the 15 active site residues, and the two mutations required to bring it into line with the other clade members are the two substitutions found in the class IV mutant. This led to the conclusion that either these are the only acceptable amino acid substitutions to give β-galactosidase activity, or that all of the single point mutations that must arise as intermediates on pathways

to β-galactosidases with different sequences are too deleterious [33].

The results obtained in vitro using IVC were quite different from those from in vivo studies: of the evolved 48 mutants with significant β-galactosidase activity, only 2 contained the D92N substitution (class I) and none contained the W976C substitution (class II). Instead, a variety of alternative solutions are seen (Table 2). Particularly striking was the observation of many clones with mutations leading to frameshifts or stop codons in the *ebgC* gene, which encodes the small β subunit of the $\alpha_4\beta_4$ Ebg enzyme. Mutations in *ebgC* which remove the β subunit have never been observed during in vivo evolution experiments. The β subunit is an enigma; there is only one homolog in GenBank which encodes a hypothetical enzyme of unknown function, and no other β-galactosidases in the same family as Ebg and LacZ are known to be heteromultimers. The β subunit does not contain any active site residues, but loss of the β subunit reduces the rate-determining degalactosylation step in catalysis by 200-fold for the wild-type enzyme and 20-fold for the class I enzyme [42].

The fact that loss of the β subunit was never observed in the in vivo studies could be due to the fact that the β domain performs an important and unknown role in *E. coli*, creating a strong selective pressure to maintain the β domain. However, this seems unlikely, because loss of Ebg appears to have no ill effects on *E. coli* in the laboratory.

A clear difference between the in vivo studies and those presented here is that the substrates used for the in vivo selections were either lactose, lactulose, galactosyl-D-arabinose, or lactobionate, whereas in this study the substrate was FDG. Indeed, deleting the β subunit from wild-type Ebg results in a significant improvement of the FDG hydrolysis activity (Table 1).

Another difference between the selections by IVC and the in vivo studies is the mutation rate. In vivo, it would be extremely rare for the *ebg* gene to contain more than a single mutation, whereas in this study there were 4, 9, or 12 mutations per gene in the different libraries selected. This allows access to a region of sequence space that was not accessible in vivo.

We have demonstrated that, by using a fluorogenic enzyme substrate and flow sorting of double emulsions, it is possible to evolve an enzyme (Ebg) completely in vitro. In the case of Ebg, it would also have been possible to undertake a similar directed evolution experiment in vivo using FDG as a fluorogenic substrate and screening cells by FACS: libraries of *ebgA* and *ebgC* genes, mutated in vitro, could be transforming into *E. coli* in which the *lacZ*, *ebgA*, and *ebgC* genes were deleted. Indeed, flow cytometric assays of β-galactosidase in bacterial, yeast, and mammalian cells have been used quite extensively [43–48].

However, the in vitro double emulsion still has some advantageous features for directed evolution: multiple cycles of mutation/recombination and selection can be performed quickly, as there is no need to transform cells with DNA; the fluorogenic substrates do not have to penetrate cells; there is no requirement for the product to remain inside or be captured on the surface outside of cells; there are no constraints related to the need to retain the original function of the enzyme; enzymatic activities can be screened which would be toxic to a cellular host; and the use of highly purified cell-free translation systems [49] should greatly reduce the risk that endogenous enzyme activities will act on the substrate.

For some projects, the expression of enzymes in cells may still be advantageous. In such cases, when it is not possible to sort the cells directly by FACS, it is also possible to compartmentalize the cells in double emulsions, together with a fluorogenic assay system, and sort the double emulsions by FACS, as described in the accompanying manuscript [34].

We believe that HTS of double emulsions by FACS, as described in this and the accompanying manuscript [34], is a potentially powerful tool for directed evolution, both due to the speed of screening (>10^7 variants per hour) and the broad dynamic range. Indeed, flow cytometry of double emulsions can easily distinguish between wild-type Ebg, Ebg class I, Ebg class IV, and LacZ (Figure 3). Only the Ebg class I and class II mutants cannot be resolved. Furthermore, the ability to perform multiparameter flow sorting opens up the possibility of simultaneously selecting enzymes with two or more substrates, each of which is hydrolyzed to release a different colored fluorochrome, thereby allowing selection for substrate specificity [17].

Significance

We have developed a completely in vitro high-throughput screening (HTS) system (>10^7 hour^{-1}) for the directed evolution of enzymes based on FACS. We have previously described a system to select for catalysis based on in vitro compartmentalization (IVC) of genes in aqueous droplets in w/o emulsions [18–21, 24]. Here, we describe an adaptation of the IVC system based on HTS using fluorogenic substrates. The w/o emulsion is converted into a w/o/w double emulsion, and fluorescent droplets are sorted by FACS. Due to its speed, sensitivity, and dynamic range, HTS of cells by FACS has proven a powerful technique to screen proteins for binding and catalysis. However, as HTS of double emulsions is a completely in vitro system, multiple cycles of mutation/recombination and selection can be performed quickly, and enzymatic activities which are toxic to a cell can be screened. Furthermore, the range of enzymes that can be selected is greatly expanded, as many fluorogenic substrates and coupled enzyme assays are available and there is no requirement for the substrate to enter a cell, or for the fluorescent product to remain inside or trapped on the surface of a cell (as with in vivo FACS screens) or linked to the gene (as in earlier examples of IVC). FACS of cells compartmentalized in double emulsions can also be used for HTS of enzymes [34].

Here we demonstrate HTS of double emulsions to evolve Ebg into an enzyme with significant β-galactosidase activity. Earlier in vivo studies have shown that Ebg, a protein of unknown function, can evolve to allow *Escherichia coli* lacking the *lacZ* β-galactosidase gene to grow on lactose. Only two specific mutations were ever seen to provide this improvement in Ebg β-galactosidase activity. In contrast, nearly all the β-galactosidases selected in vitro resulted from different mutations.

Experimental Procedures

Synthesis of *lacZ* and *ebg* Genes

pIVEX2.2EM is a truncated version of pIVEX2.2bNde (Roche, Indianapolis, IN) lacking part of the LacZ α-peptide coding sequence. The wild-type *E. coli lacZ* gene and a variant *lacZ* gene with a frameshift mutation that removes an internal SacI site were cloned into pIVEX2.2EM, creating pIVEX2.2EM-LacZ and pIVEX2.2EM-lacZmut, respectively. The *E. coli ebgA* and *ebgC* genes were cloned as a single construct into pIVEX2.2EM, creating pIVEX2.2EM-ebgAC (Figure S3A). The *ebg* class I, class II, and class IV mutants were created by site-directed mutagenesis of pIVEX2.2EM-ebgAC. The deleted variant pIVEX2.2EM-ebgA with a deletion in the *ebgC* gene (Figure S3B) and the class I, class II, and class IV variants with the same deletion were constructed from the corresponding pIVEX2.2EM-ebgAC plasmids. Linear DNA constructs for in vitro translation and selection were generated from pIVEX2.2EM constructs by PCR. For further details, see Supplemental Data.

Creation of *ebg* Gene Libraries

A random mutagenesis library (EbgMut1) was constructed by error-prone PCR of *ebg* using base analogs essentially as described [37]. After two rounds of selection, the library was remutated following the same protocol. For further details, see Supplemental Data.

Shuffling of *ebg* Gene Libraries

EbgMut1 library DNA recovered after the second round of selection was shuffled essentially as described [38]. For further details, see Supplemental Data.

Preparation of In Vitro Transcription/Translation Mix

A 50 μl coupled in vitro transcription/translation (IVT) mixture (EcoProT7; Novagen/EMD Biosciences, Madison, WI), containing 0.1 nM library DNA and supplemented with 500 μM fluorescein di-β-D-galactopyranoside (FDG; Molecular Probes, Invitrogen, Carlsbad, CA) and 75 μM 7-hydroxycoumarin-3-carboxylic acid (Sigma Aldrich, St. Louis, MO), was prepared on ice according to the manufacturer's protocol. For model selections with LacZ β-galactosidase, the final concentration of FDG was 250 μM.

Preparation of Primary (w/o) Emulsions

A solution of 1% (w/v) Span 60 and 1% (w/v) cholesterol in decane (all from Sigma Aldrich) was prepared at 45ºC, then divided into 200 μl aliquots and placed in a heat block at 37ºC.

A hand-extruding device (mini extruder; Avanti Polar Lipids, Alabaster, AL) was fitted with a 19 mm Track-Etch polycarbonate filter with average pore size of 14 μm (GE Osmonics, Minnetonka, MN) inside the mini extruder. Two gas-tight 1 ml Hamilton syringes (Gas-tight 1001; Hamilton, Reno, NV) were used for extrusion. The extruder was prerinsed with 3 × 1 ml of decane.

For emulsification, 50 μl of IVT mixture was added to 200 μl of the preheated decane/surfactant mix and loaded into one of the syringes. The IVT mix was forced through the filter into the alternate syringe and directly forced back into the original syringe to complete one round of extrusion. In total, 7.5 rounds of extrusion were completed. The w/o emulsion formed was placed at 30ºC for 2 hr to allow in vitro transcription and translation.

Preparation of Secondary (w/o/w) Double Emulsions

For the second emulsification step, a 19 mm Track-Etch polycarbonate filter with an average pore size of 8 μm was fitted inside the extruder. The extruder was prerinsed with 3 × 1 ml PBS. Two hundred fifty microliters of the primary w/o emulsion was added to 750 μl of PBS containing 0.5% (w/v) Tween 80 (Sigma Aldrich) and loaded into one of the syringes. The IVT mix was forced through the filter into the alternate syringe and directly forced back into the original syringe to complete one round of extrusion. In total, 3.5 cycles of extrusion were performed. The double emulsions were collected on ice. The size distribution of the oil droplets was determined by single-particle optical sensing (SPOS) using an AccuSizer 780 optical particle sizer (Particle Sizing Systems, Santa Barbara, CA).

FACS Sorting of Double Emulsions

Double emulsions were diluted 25-fold in sterile-filtered PBS and run on a MoFlo flow cytometer (DakoCytomation, Fort Collins, CO) using PBS as sheath fluid. The cytometer was fitted with a 100 μm nozzle, an argon ion laser emitting at 488 nm, and an argon ion laser tuned at 350 nm. Two band-pass filters of 450 ± 30 nm and 530 ± 15 nm were used to detect the 7-hydroxycoumarin-3-carboxylic acid fluorescence and product fluorescence, respectively. The sorter was triggered on coumarin fluorescence, thereby ignoring all o/w droplets (i.e., droplets lacking an internal aqueous phase). The applied rate of sorting was 20,000 events per second. Sort gates were set so as to collect <0.01% of the double emulsion compartments of a negative control (double emulsions without DNA). For the second-generation libraries, in the last selection round, the gates were set to favor mutants with similar kinetics to class IV. One hundred thousand droplets were collected using purify mode in each selection round.

DNA Recovery from Sorted Double Emulsions

DNA from the sorted w/o/w compartments was precipitated by adding 0.1 volume (relative to the sorted volume) of 3 M sodium acetate (pH 5.2) and 0.7 volume of isopropanol in the presence of 20 μg glycogen. DNA was pelleted by centrifugation at 20,000 × g for 15 min at 4ºC, washed twice with 100 μl 70% ethanol, dried using a speedvac (Eppendorf, Cambridge, UK), and resuspended into 20 μl water. Ten microliters of this DNA was used in a 50 μl PCR reaction using Expand Long Template PCR mix with buffer 1 (Roche). Primers used after each selection round are described in Table S2. Reactions were performed using 300 μM primers and incubated for 2 min at 94ºC and subsequently subjected to ten cycles at 94ºC, 15 s; 55ºC, 30 s; 68ºC, 2 min, another 22 cycles with an increment in elongation time of 10 s/cycle, and a final incubation step for 7 min at 68ºC. PCR products were purified using a Qiaquick PCR Purification kit (Qiagen, Crawley, UK) between selection rounds and by electroelution using a GeBAflex dialysis kit (GEBA, Gene Bio-Application, Kfar Hanagid, Israel) after the last selection round.

Characterization of Selected Clones

Electroeluted DNA recovered after the last selection round of each cycle of directed evolution was cloned into pIVEX2.2EM (see Supplemental Data). Individual clones were amplified in 10 μl PCR reactions in a 384-well PCR plate (Abgene, Epsom, UK). The amplifications were performed using Expand Long Template PCR mix (Roche) with primers PIVB-4 and LMB2-10E, and cycled 32 times as above. One microliter of the PCR reaction was transferred to 7 μl of EcoProT7 IVT mix and incubated for 2.5 hr at 30ºC. Three microliters of this reaction mix was transferred to 20 μl of reaction buffer containing 0.1 mM FDG in a 384-well black microtiter plate (Nunc, Rochester, NY). Initial rates were measured by monitoring at 25ºC for 2 hr using a SpectraMAX Gemini XS fluorometer (Molecular Devices, Sunnyvale, CA) set at excitation 485 nm and emission 514 nm. Wild-type Ebg and the class I, II, and IV mutants were used as controls.

Kinetic Analysis of Selected Clones

After the second cycle of directed evolution, 60 clones with initial rates for 0.1 mM FDG hydrolysis at least two times higher than Ebg class II were selected from library EbgMut2a (mutated), and 23 were selected from library EbgMut2b (recombined). The initial rate of FDG hydrolysis was determined at 0.1 mM, 0.2 mM, 0.4 mM, and 0.8 mM FDG to estimate K_m and V_{max}. The 28 variants with the higher V_{max}/K_m were streaked and tested again. The DNA was extracted and used to retransform *E. coli*, and single colonies were assayed to ensure DNA segregation.

Linear DNA constructs encoding the selected mutants and controls (wild-type Ebg, the class I, II, and IV mutants, and the corresponding constructs lacking *ebgC*) were generated from pIVEX2.2EM constructs by PCR using primers LMB2-10E and PIVB-4 and the Expand Long Template PCR mix (Roche) (see Supplemental Data). DNA was purified using a Qiaquick PCR Purification kit (Qiagen). Five nanomolar linear constructs were used in 80 μl in vitro translation reactions using EcoProT7 IVT mix and incubated for 4 hr at 30ºC. To measure the K_M and V_{max}, 4 μl aliquots of the IVT reaction were assayed in 40 μl reaction buffer at different

concentrations of FDG ranging from 2 μM to 4 mM, as appropriate, in a 384-well black microtiter plate. The kinetic parameters were determined by fitting the maximum fluorescence rate at each FDG concentration to the Michaelis-Menten equation ($\upsilon = [E]_0[S]_0k_{cat}/[S]_0 + K_M$) using the Levenberg-Marquardt algorithm as implemented in KaleidaGraph (Synergy, Reading, PA) [50] (see Supplemental Data). Wild-type Ebg showed very little activity with FDG as substrate and it was not possible to accurately determine the k_{cat} and K_M because activity was only detectable at the highest FDG concentrations and at the highest FDG concentration achievable (4 mM) activity had not reached a plateau. Hence, it was only possible to say that the K_M is probably >4 mM and to estimate the k_{cat}/K_M from the maximum velocity at 4 mM FDG, assuming that $[S]_0 < K_M$ and $\upsilon \approx [E]_0[S]_0(k_{cat}/K_M)$.

The protein yield per translation reaction was determined for each of the selected mutants and controls by measuring incorporation of [^{35}S]methionine in a 20 μl IVT reaction. Insoluble material was removed by spinning 10 min at 25,000 × g, and 2.5 μl of the supernatant was run on a NuPAGE 10% Bis-Tris gel (Invitrogen) with MES buffer and full-length protein quantified using a phosphorimager (Storm 860, Molecular Dynamics). Protein yields varied by ±24%.

Supplemental Data

Supplemental Data include three figures and two tables and can be found with this article online at http://www.chembiol.com/cgi/content/full/12/12/1291/DC1/.

Acknowledgments

We acknowledge Richard Grenfell for assistance with FACS and Dr. Antoine Taly for making the Ebg structural model as well as for helpful discussion and critical reading of the manuscript. This research was supported by an EC Framework 5 Marie Curie Individual Fellowship to E.M., by the EC Framework 5 Marie Curie "ENDIRPRO" Training Network for V.T., and by the Medical Research Council.

Received: July 8, 2005
Revised: September 22, 2005
Accepted: September 22, 2005
Published: December 16, 2005

References

1. Tao, H., and Cornish, V.W. (2002). Milestones in directed enzyme evolution. Curr. Opin. Chem. Biol. *6*, 858–864.
2. Griffiths, A.D., and Tawfik, D.S. (2000). Man-made enzymes—from design to in vitro compartmentalisation. Curr. Opin. Biotechnol. *11*, 338–353.
3. Aharoni, A., Griffiths, A.D., and Tawfik, D.S. (2005). High-throughput screens and selections of enzyme-encoding genes. Curr. Opin. Chem. Biol. *9*, 210–216.
4. Firestine, S.M., Salinas, F., Nixon, A.E., Baker, S.J., and Benkovic, S.J. (2000). Using an AraC-based three-hybrid system to detect biocatalysts in vivo. Nat. Biotechnol. *18*, 544–547.
5. Georgiou, G. (2000). Analysis of large libraries of protein mutants using flow cytometry. Adv. Protein Chem. *55*, 293–315.
6. Feldhaus, M.J., Siegel, R.W., Opresko, L.K., Coleman, J.R., Feldhaus, J.M., Yeung, Y.A., Cochran, J.R., Heinzelman, P., Colby, D., Swers, J., et al. (2003). Flow-cytometric isolation of human antibodies from a nonimmune *Saccharomyces cerevisiae* surface display library. Nat. Biotechnol. *21*, 163–170.
7. Chen, G., Hayhurst, A., Thomas, J.G., Harvey, B.R., Iverson, B.L., and Georgiou, G. (2001). Isolation of high-affinity ligand-binding proteins by periplasmic expression with cytometric screening (PECS). Nat. Biotechnol. *19*, 537–542.
8. Hayhurst, A., and Georgiou, G. (2001). High-throughput antibody isolation. Curr. Opin. Chem. Biol. *5*, 683–689.
9. Wittrup, K.D. (2001). Protein engineering by cell-surface display. Curr. Opin. Biotechnol. *12*, 395–399.
10. Wentzel, A., Christmann, A., Adams, T., and Kolmar, H. (2001). Display of passenger proteins on the surface of *Escherichia coli* K-12 by the enterohemorrhagic *E. coli* intimin EaeA. J. Bacteriol. *183*, 7273–7284.
11. Harvey, B.R., Georgiou, G., Hayhurst, A., Jeong, K.J., Iverson, B.L., and Rogers, G.K. (2004). Anchored periplasmic expression, a versatile technology for the isolation of high-affinity antibodies from *Escherichia coli*-expressed libraries. Proc. Natl. Acad. Sci. USA *101*, 9193–9198.
12. Shusta, E.V., Holler, P.D., Kieke, M.C., Kranz, D.M., and Wittrup, K.D. (2000). Directed evolution of a stable scaffold for T-cell receptor engineering. Nat. Biotechnol. *18*, 754–759.
13. Nguyen, A.W., and Daugherty, P.S. (2005). Evolutionary optimization of fluorescent proteins for intracellular FRET. Nat. Biotechnol. *23*, 355–360.
14. Becker, S., Schmoldt, H.U., Adams, T.M., Wilhelm, S., and Kolmar, H. (2004). Ultra-high-throughput screening based on cell-surface display and fluorescence-activated cell sorting for the identification of novel biocatalysts. Curr. Opin. Biotechnol. *15*, 323–329.
15. Kawarasaki, Y., Griswold, K.E., Stevenson, J.D., Selzer, T., Benkovic, S.J., Iverson, B.L., and Georgiou, G. (2003). Enhanced crossover SCRATCHY: construction and high-throughput screening of a combinatorial library containing multiple non-homologous crossovers. Nucleic Acids Res. *31*, e126.
16. Olsen, M.J., Stephens, D., Griffiths, D., Daugherty, P., Georgiou, G., and Iverson, B.L. (2000). Function-based isolation of novel enzymes from a large library. Nat. Biotechnol. *18*, 1071–1074.
17. Varadarajan, N., Gam, J., Olsen, M.J., Georgiou, G., and Iverson, B.L. (2005). Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity. Proc. Natl. Acad. Sci. USA *102*, 6855–6860.
18. Griffiths, A.D., and Tawfik, D.S. (2003). Directed evolution of an extremely fast phosphotriesterase by in vitro compartmentalization. EMBO J. *22*, 24–35.
19. Tawfik, D.S., and Griffiths, A.D. (1998). Man-made cell-like compartments for molecular evolution. Nat. Biotechnol. *16*, 652–656.
20. Cohen, H.M., Tawfik, D.S., and Griffiths, A.D. (2004). Altering the sequence specificity of HaeIII methyltransferase by directed evolution using in vitro compartmentalization. Protein Eng. Des. Sel. *17*, 3–11.
21. Lee, Y.F., Tawfik, D.S., and Griffiths, A.D. (2002). Investigating the target recognition of DNA cytosine-5 methyltransferase HhaI by library selection using in vitro compartmentalisation. Nucleic Acids Res. *30*, 4937–4944.
22. Ghadessy, F.J., Ong, J.L., and Holliger, P. (2001). Directed evolution of polymerase function by compartmentalized self-replication. Proc. Natl. Acad. Sci. USA *98*, 4552–4557.
23. Ghadessy, F.J., Ramsay, N., Boudsocq, F., Loakes, D., Brown, A., Iwai, S., Vaisman, A., Woodgate, R., and Holliger, P. (2004). Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. Nat. Biotechnol. *22*, 755–759.
24. Doi, N., Kumadaki, S., Oishi, Y., Matsumura, N., and Yanagawa, H. (2004). In vitro selection of restriction endonucleases by in vitro compartmentalization. Nucleic Acids Res. *32*, e95.
25. Doi, N., and Yanagawa, H. (1999). STABLE: protein-DNA fusion system for screening of combinatorial protein libraries in vitro. FEBS Lett. *457*, 227–230.
26. Sepp, A., Tawfik, D.S., and Griffiths, A.D. (2002). Microbead display by in vitro compartmentalisation: selection for binding using flow cytometry. FEBS Lett. *532*, 455–458.
27. Yonezawa, M., Doi, N., Kawahashi, Y., Higashinakagawa, T., and Yanagawa, H. (2003). DNA display for in vitro selection of diverse peptide libraries. Nucleic Acids Res. *31*, e118.
28. Bertschinger, J., and Neri, D. (2004). Covalent DNA display as a novel tool for directed evolution of proteins in vitro. Protein Eng. Des. Sel. *17*, 699–707.
29. Bernath, K., Magdassi, S., and Tawfik, D.S. (2005). Directed evolution of protein inhibitors of DNA-nucleases by in vitro compartmentalization (IVC) and nano-droplet delivery. J. Mol. Biol. *345*, 1015–1026.
30. Bernath, K., Hai, M., Mastrobattista, E., Griffiths, A.D., Magdassi, S., and Tawfik, D.S. (2004). In vitro compartmentalization by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting. Anal. Biochem. *325*, 151–157.
31. Hall, B.G. (2003). The EBG system of *E. coli*: origin and evolution of a novel β-galactosidase for the metabolism of lactose. Genetica *118*, 143–156.

32. Hall, B.G. (1999). Experimental evolution of Ebg enzyme provides clues about the evolution of catalysis and to evolutionary potential. FEMS Microbiol. Lett. *174*, 1–8.
33. Hall, B.G., and Malik, H.S. (1998). Determining the evolutionary potential of a gene. Mol. Biol. Evol. *15*, 1055–1061.
34. Aharoni, A., Amitai, G., Bernath, K., Magdassi, S., and Tawfik, D.S. (2005). High-throughput screening of enzyme libraries: new thiobutyrolactonases by fluorescence activated sorting of single cells in emulsion compartments. Chem. Biol. *12*, this issue, 1281–1289.
35. Kendall, D.A., and MacDonald, R.C. (1982). A fluorescence assay to monitor vesicle fusion and lysis. J. Biol. Chem. *257*, 13892–13895.
36. Hall, B.G., and Zuzel, T. (1980). Evolution of a new enzymatic function by recombination within a gene. Proc. Natl. Acad. Sci. USA *77*, 3529–3533.
37. Zaccolo, M., Williams, D.M., Brown, D.M., and Gherardi, E. (1996). An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. J. Mol. Biol. *255*, 589–603.
38. Stemmer, W.P. (1994). Rapid evolution of a protein in vitro by DNA shuffling. Nature *370*, 389–391.
39. Juers, D.H., Heightman, T.D., Vasella, A., McCarter, J.D., Mackenzie, L., Withers, S.G., and Matthews, B.W. (2001). A structural view of the action of *Escherichia coli* (lacZ) β-galactosidase. Biochemistry *40*, 14781–14794.
40. Warren, R.A. (1972). Lactose-utilizing mutants of lac deletion strains of *Escherichia coli*. Can. J. Microbiol. *18*, 1439–1444.
41. Campbell, J.H., Lengyel, J.A., and Langridge, J. (1973). Evolution of a second gene for β-galactosidase in *Escherichia coli*. Proc. Natl. Acad. Sci. USA *70*, 1841–1845.
42. Calugaru, S.V., Hall, B.G., and Sinnott, M.L. (1995). Catalysis by the large subunit of the second β-galactosidase of *Escherichia coli* in the absence of the small subunit. Biochem. J. *312*, 281–286.
43. Nolan, G.P., Fiering, S., Nicolas, J.F., and Herzenberg, L.A. (1988). Fluorescence-activated cell analysis and sorting of viable mammalian cells based on β-D-galactosidase activity after transduction of *Escherichia coli* lacZ. Proc. Natl. Acad. Sci. USA *85*, 2603–2607.
44. Fiering, S.N., Roederer, M., Nolan, G.P., Micklem, D.R., Parks, D.R., and Herzenberg, L.A. (1991). Improved FACS-Gal: flow cytometric analysis and sorting of viable eukaryotic cells expressing reporter gene constructs. Cytometry *12*, 291–301.
45. Nir, R., Yisraeli, Y., Lamed, R., and Sahar, E. (1990). Flow cytometry sorting of viable bacteria and yeasts according to β-galactosidase activity. Appl. Environ. Microbiol. *56*, 3861–3866.
46. Wittrup, K.D., and Bailey, J.E. (1988). A single-cell assay of β-galactosidase activity in *Saccharomyces cerevisiae*. Cytometry *9*, 394–404.
47. Miao, F., and Kompala, D.S. (1993). A single-cell assay of β-galactosidase in recombinant *Escherichia coli* using flow cytometry. Biotechnol. Bioeng. *42*, 708–715.
48. Plovins, A., Alvarez, A.M., Ibanez, M., Molina, M., and Nombela, C. (1994). Use of fluorescein-di-β-D-galactopyranoside (FDG) and C12-FDG as substrates for β-galactosidase detection by flow cytometry in animal, bacterial, and yeast cells. Appl. Environ. Microbiol. *60*, 4638–4641.
49. Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., and Ueda, T. (2001). Cell-free translation reconstituted with purified components. Nat. Biotechnol. *19*, 751–755.
50. Huang, Z.J. (1991). Kinetic fluorescence measurement of fluorescein di-β-D-galactoside hydrolysis by β-galactosidase: intermediate channeling in stepwise catalysis by a free single enzyme. Biochemistry *30*, 8535–8540.