Article:
Alfaro-Chávez, AL, Liu, J-W, Porter, JL et al. (2 more authors) (Cover date: January 2019) Improving on nature's shortcomings: evolving a lipase for increased lipolytic activity, expression and thermostability. Protein engineering, design & selection : PEDS, 32 (1). pp. 13-24. ISSN 1741-0126

https://doi.org/10.1093/protein/gzz024

Improving on nature's shortcomings: evolving a lipase for increased lipolytic activity, expression and thermostability

Ana L. Alfaro-Chavez[1], Jian-Wei Liu[2], Joanne L. Porter[1], Adrian Goldman[3,4] David L. Ollis[1].

[1]Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia

[2]CSIRO Land and Water, Black Mountain, Canberra, ACT, 2601, Australia

[3]School of Biomedical Sciences, Astbury Centre for Structural Molecular Biology, University of Leeds, LS2 9JT

[4]Molecular and Integrative Biosciences Program, University of Helsinki, FIN-0018.

Corresponding Author:

David L, Ollis

Research School of Chemistry

Australian National University

Canberra, ACT, 2601, Australia

Tel: +61 254377

E-mail: david.ollis@anu.edu.au

Evolving a lipase for increased lipolytic activity, expression and thermostability

Abstract

An enzyme must be soluble, stable, active and easy to produce to be useful in industrial applications. Not all enzymes possess these attributes. We set out to determine how many changes are required to convert an enzyme with poor properties into one that has useful properties.

Lipase Lip3 from *Drosophila melanogaster* had been previously optimised for expression in *Escherichia coli*. The expression levels were good, but Lip3 was mainly insoluble with poor activity. Directed evolution was used to identify variants with enhanced activity along with improved solubility. Five variants and the wild-type (*wt*) enzyme were purified and characterised. The yield of the *wt* enzyme was just 2.2 mg per Litre of culture while a variant, produced under the same conditions, gave 351 mg. The specific activity of the best variant was 20 times higher than wild-type. All the purified variants showed increased thermal stability compared to the wild-type enzyme. The improved variants had between 5 and 9 changes compared with the *wt* enzyme. There were 4 changes that were found in all 30 final round variants for which sequences were obtained; three of these changes were found in the substrate binding domain.

Keywords: directed evolution / lip3 / lipase / solubility / thermo-stability.

Introduction

In this paper, we set out to evolve a protein with poor physical and catalytic properties to determine if a small number of changes could produce variants that might be useful to a chemist. To be suitable for industrial use, an enzyme must not only possess useful catalytic properties, it must also be soluble, easily produced and stable under the required process conditions (Jemli et al. 2016). One can screen naturally occurring enzymes for the desired attributes, but enzymes have evolved to satisfy the biological needs of the organism, not the industrial goals of humans: they may have one of the desired characteristics, but lack others. Rather than searching for a naturally enzyme with properties meeting all the relevant criteria, a chemist could modify a protein with some desirable properties to produce useful variants. This approach is only feasible if desired changes can be produced with a relatively small number of changes and if the technology to identify these changes is available. We set out to show that directed evolution was such a technology (Arnold 1998, 2018; Porter et al. 2016; M. T. Reetz and Jaeger 1999; Sheldon and Pereira 2017; Turner 2003, 2009). For this reason, we deliberately chose to start our experiments with a protein that was considered inappropriate for routine biochemical characterization.

The gene for Lip3, CG8823, from *Drosophila melanogaster* encodes a lipase that is expressed during the larval stage of fly development (Pistillo et al. 1998). The activity of the protein was assigned based on its sequence similarity with other lipases and without biochemical characterisation. Lipases are triacylglycerol ester hydrolases. As they can break down the triacylglycerides in fats, they are extremely useful in the food and cleaning industries as well as in other industrial areas. These applications have been described in comprehensive reviews (Gurung et al. 2013; Hasan et al. 2010; Houde et al. 2004; Kirk et al. 2002; Lobedanz et al.

2000) and will not be elaborated upon here. However, to be of any use in industry, an enzyme must possess appropriate physicochemical properties along with its ability to catalyse useful reactions.

The properties of Lip 3 were far from useful for practical applications. Prior to the start of this work, others had shown that Lip3 could be expressed at high level in *E. coli.*, but it was very insoluble and had only barely detectable activity. However, the protein exhibited 60.8% similarity to dog gastric lipase, for which the structure was known (PDB:1K8Q, (Roussel et al. 2002)). This observation suggested that the protein might be amenable to engineering studies and that the known structure might be used to understand the effects of changes brought about by these studies.

The Lip3 protein consists of 375 amino acids and has a molecular weight of 43 kDa. It exhibits sequence similarity to two other lipases (1K81 (Roussel et al. 2002) and 1HLG (Roussel et al. 1999). These lipases belong to the α/β hydrolase fold family of enzymes (Carr and Ollis 2009; Ollis et al. 1992). Proteins in the α/β hydrolase fold family typically have a core of eight β strands, with the second antiparallel to the others, and the strands connected by surface exposed α-helices. Like the serine proteases, these proteins have catalytic triads that consist of a nucleophile, an acidic residue and a histidine: in Lip3 these are S145, D319 and H350. Like many other α/β hydrolase fold enzymes, lipases have a substrate binding domain that emerges from Strand 7 and forms much of the active site. Part of this domain consists of a short flexible peptide, known as the lid, which can prevent access to the active site in one conformation and allow access to it in another; these are referred to as the open and closed conformations.

In this work, multiple rounds of directed evolution were used to improve the activity of Lip3 towards TGAs (Triacyl glycerides) with short aliphatic chains. In doing this, the solubility and the stability of Lip3 also improved. The object of these experiments was to produce variants that were easily purified and with a useful catalytic activity that could be detected in a high throughput assay. That is, the final variants had to be a suitable starting point for further evolution studies to produce variants that could be used for specific practical applications. The applications we had in mind required that the final variants from the current study should have detectable activity with TGAs with long aliphatic chains. Evolution towards more practical applications is described in subsequent papers.

Materials and methods.

Distilled water was purified with a MilliQ reagent system. All the consumables used for molecular biology including all the glassware, micropipette tips and micro-centrifuge tubes as well as growth media and buffers were sterilized using an ASB270BT autoclave (Astell Scientific). The autoclave was run at 121° C for 15 mins. Where possible, the chemicals used were of analytical grade. Ethanol, methanol, isopropanol, hydrochloric acid and glycerol were from Merck Pty. Ltd., VIC, Australia; sodium acetate was from ICN Biomedicals. Lysozyme was obtained from Novozyme and Tris from VWR. Imidazole, Triton X-100, ethyl acetate, Phenol red, tributyrin, glyceryl trioctanoate, 1-naphtyl palmitate, Alcalase®, agar and DTT (1,4 Dithiothreitol) were from Sigma. Glycine and sodium dodecyl sulfate (SDS) were from Amresco and bromophenol blue from LabChem. Yeast extract, Tryptone and nutrient broth were

obtained from Bacto. 1 Kb ladder as well as restriction enzymes *Nde* I and *Eco*RI were from New England Biolabs. FastAP, T4 ligase and Pfu polymerase were from Thermo Scientific. Taq Polymerase, dNTPs, agarose and Isolate II Plasmid mini kit were from Bioline while the Wizard®SV Gel and PCR Clean-Up kit were from Promega.

Heterologous expression of Drosophila melanogaster Lip3 lipase in Escherichia coli

The gene encoding Lip3 (CG8823) was codon optimised and synthesized by GeneArt (Thermo Fisher Scientific). After PCR amplification, the gene was inserted between the *Nde*I and *Eco*RI restriction sites of pETMCSIII. This vector (Stevenson et al. 2012) allows the leaky expression of recombinant proteins due in the presence of lactose in LB medium and the absence of the LacI gene in the plasmid and on the chromosome (Studier 2005). This vector adds an N-terminal hexa-histidine tag to the expressed protein so that it can be purified by nickel-affinity chromatography on an Ni-NTA column (Stevenson et al. 2012; Studier 2005). The *Nde1* site corresponds to the section of the gene that codes for the last histidine of the N-terminal His tag and the starting methionine of the Lip3 gene. The *Eco*RI site was 45 bases after the end of the Lip3 gene. The pETMCSIII vector with CG8823 inserted was used for all aspects of this work requiring the production of Lip3.

Library Creation

Error-prone Polymerase Chain Reaction (ep-PCR) was used to introduce mutations in the gene as described elsewhere (Wilson and Keefe 2001), so only a brief description will be given here. Mutations were introduced into the gene by including $MnCl_2$ and an increased amount of $MgCl_2$ and the non-proofreading Taq polymerase. The $MnCl_2$ and $MgCl_2$ were added to the

buffer supplied with the Tac polymerase. The PCR was carried out with a short extension time to ensure that the genes were shuffled (Zhao et al. 1998). The primers used were the pET 3 upstream 96 bp (base pair) primer (5'-CGA CTC ACT ATA GGG AGA CCA CAA C-3') and pET4 downstream 43 bp primer (5'-CCT TTC GGG CTT TGT TAG CAG-3'). Each library was a mix of the product of three PCR runs, each of which utilized different amounts of $MnCl_2$ to generate a large library that had components with different levels of mutagenesis. The three PCR reactions were performed in 50 µl reaction mixtures that contained 50-60 ng plasmid, $MnCl_2$ (0.1 or 0.2 or 0.3 mM), 1 µM of each primer (Integrated DNA Technologies), 5 mM $MgCl_2$ (diluted from a 50 mM solution provided by Bioline), 0.8 µM dNTPs and 5 U Biotaq (Taq polymerase, Bioline). PCR was performed using an Applied Biosystems Veriti Thermal Cycler with the following conditions: 94° C for 2 min before 35 cycles of (45° C for 10 secs, 72° C for 30 secs, 94° C for 10 secs) followed by 45° C for 2 min and 72° C for 5 min.

The amplified products from the different concentrations of $MnCl_2$ were combined and purified using the Promega® PCR purification kit. The purified DNA was digested with 20 U of each *Nde*I and *Eco*RI-HF and incubated at 37° C for 3 hours. The digestion reaction was carried out in the Cutsmart buffer provided with the restriction enzymes by New England Biolabs. The digested DNA was gel purified with the Promega® PCR purification kit. The pETMCSIII with the wild-type gene was cut with *Nde*I and *Eco*RI-HF as described above and the product treated with FastAP and incubated for 30 minutes at 37^{o} C. The reaction was stopped by incubating the mixture for 5 minutes at 65^{o} C, as specified by the manufacturer. The cut plasmid was gel purified with the Promega® PCR purification kit. T4 ligase was used to ligate the purified PCR product into the cut plasmid as specified by the manufacturer; the reaction was carried out overnight at 4° C. The recombinant plasmid was purified with the Promega® PCR purification

kit and eluted into 30 μL of water. The purified plasmid was transformed into *E. coli* BL21 (DE3) by electroporation and incubated at 37° C for 1 hour in YENB medium (7.5 g/L yeast extract and 8 g/L nutrient broth). The resulting library was plated on LB agar plates (5 g/L yeast extract, 10 g/L tryptone, 10g/L NaCl and 15 g/L agar) with ampicillin 50 μg/mL to determine the density of colonies.

Library screening

*Primary Screen***:** The library was plated with 300-400 mutants per LB agar plate. In a typical round of evolution, 100 agar plates were used. The plates, prepared with 2% glyceryl tri-octanoate sonicated in the medium after initial autoclaving, had a white color that disappeared in the presence of active lipases. Consequently, clear haloes formed around colonies expressing active lipases. The size of the haloes depended upon the activity of the lipases and the incubation temperature and time. As variants became more stable, the incubation temperature could be increased from 30° C to 37° C where the cells grew more quickly. In addition, as the variants became more active, the haloes grew faster so that the incubation times could be reduced. Table 1 gives the temperature and incubation time for each round of evolution. Although haloes were easily observed, their size could only be approximately estimated by eye so that a more accurate secondary screen was necessary. We picked about 10 colonies per plate for secondary screening (*i.e.* about 1000 colonies overall).

Secondary Screen: Colonies were picked for the secondary screen using sterile tips and cells from the colony grown in 96-well microtiter plates with 200 μl LB-amp and incubated overnight at 30° C or 37° C depending upon the temperature initially used to express the variants (Table 1). The plates were shaken at 1000 rpm in a Heidolph Microplate Shaker

(Titramax 1000). After 18 hours incubation, half of each well was transferred to the corresponding well of a fresh (replica) plate and lysed with 0.06 mg/mL lysozyme. The esterase activity was measured using pNP-C8 and monitoring the formation of the *p*-nitrophenol product at 405 nm. Along with the variants, the *wt* and the best variant from the previous round were grown in triplicate in each plate. For a variant to be chosen for further study, its activity had to be at least two times higher than the best variant from the previous round. The *wt* was grown to ensure that the growth protocol was reproducible from round to round.

The plasmids of the selected variants were isolated, purified using Qiagen Miniprep® kit and sequenced by the ACRF Biomolecular Resource Facility (BRF) at the Australian National University. Approximately equimolar amounts of DNA from each unique variant were mixed together and used as a template for the next round of directed evolution.

Enzyme purification

Transformants of the *wt* and the best variants were cultured in 10 mL LB-amp medium and incubated for five hours at 37° C. These starter cultures were used to inoculate one litre LB-amp medium, and the cultures incubated for 48 hours at 19° C for *wt* and 18 hours at 37° C for the mutants with shaking at 180 rpm. The cells were harvested and resuspended in 300 mM NaCl and 50 mM Tris-HCl pH 8 (Buffer A). The cells were lysed with a homogeniser and the insoluble materials were removed by centrifugation at 15000 g for 45 min at 4°C. The enzymes were purified using a 5 mL HisTrap column (Sigma, USA). The purification process was carried out by AKTA chromatography system (GE Healthcare) using 50 mM Tris-HCl, 300 mM NaCl and 500 mM imidazole as eluent. The enzyme was eluted with a gradient from 0 to

100% imidazole at a flow rate of 0.5 mL/min during 10 column volumes. Fractions of 5 mL were collected and analysed with SDS-PAGE. The fractions containing Lip3 were combined and dialysed overnight in Buffer A at 4° C. The enzyme was concentrated using Amicon filters (30 kDa MWCO).

Determination of esterase activity with pNP ester substrates

Esterase activity was determined spectrophotometrically with crude lysates and with purified enzymes using *p*-nitrophenyl esters as substrates. Preparation of the substrate mixture followed the method of Gupta and co-workers (Gupta et al. 2002). In this method, the substrate mixture was prepared by dissolving 30 mg of the ester in 10 ml of isopropanol and diluting it to 100 ml in buffer (50 mM Tris-HCl pH 8), with the addition of 2 % Triton X-100 for pNP-C16 and pNP-C18. To start the reaction, the substrate mixture was then added to the protein solution in Buffer A. Production of the *p*-nitrophenol product was measured at 405 nm on a microtiter plate reader (Spectramax M2e). A *p*-nitrophenol standard curve in the 10 – 600 µM range was used to quantitate product formation. One unit (U) of activity was defined as the amount of enzyme required to produce one µM p-nitrophenol/min/µg protein.

Thermal stability

The thermostability of the *wt* and variants was estimated by measuring their residual activity with pNP-C8 after heat treatment. The purified enzymes in Buffer A were incubated at different temperatures (30°C to 80°C) using a Bio-Rad C1000^{TM} Thermal Cycler. To get a reasonable level of activity the concentration of the *wt* enzyme was 1 mg/mL while that of the variants was 0.1 mg/mL. The thermal treatment was applied to 15 µL of enzyme for 45 min

followed for 30 min at 4° C. The activity assay was carried out with 10 µL of enzyme using *p*-nitrophenyl octanoate as substrate as described above. T_{50} values were calculated using a Hill type equation in Kaleidagraph as previously reported (Stevenson et al. 2008). T_{50} is defined as the temperature at which 50% of the original activity remains. The equation used to fit the curve was:

$$a = \frac{a_0 - (a_0 * T^n)}{T_{1/2}^n + T^n}$$

where: a is the proportion of activity of treated enzyme compared with untreated enzyme; a_0 the activity of the control enzyme (100%) and is measured as µM product/(min*ηM protein); T is temperature in °C; $T_{1/2}$ is the temperature at which half of the enzyme is inactivated; n is the Hill coefficient.

Activity Profile

pNP-C8 was used at different concentrations (0.005 to 0.1 mM) to determine the reaction rate under standard enzyme assay conditions, as described above for the esterase activity. The concentration of the *wt* enzyme was 60nM while the concentration of the variants was 20 nm, except for R7_59A variant for which 6 nM was used. The rate of hydrolysis did not plateau but decreased at higher substrate concentrations. Consequently, we fit the data to a substrate inhibition model:

$$turnover = \frac{k_{cat}[S]}{K_M + [S]\left(1 + \frac{[S]}{K_i}\right)}$$

where: turnover is $V_0/[E_0]$, V_0 is the initial velocity, $[E_0]$ is the concentration of total enzyme, k_{cat} is the turnover number (min^{-1}), [S] is the substrate concentration (mM), K_M is the equivalent of the Michaelis constant (mM) and K_i is the dissociation constant for the enzyme-substrate inhibitor complex (mM) (Kaiser 1980). The data were fit with GraphPad Prism v7.0, using default parameters except that k_{cat} was restrained to a maximum of 100,000 min^{-1}(about four times the maximal observed value), as otherwise the fit for the R7_59A variant was not well-behaved.

Lipase activity

To measure the hydrolysis of triglycerides, we developed a spectrophotometric lipase assay similar to that of Camacho-Ruiz *et al* (Camacho-Ruiz et al. 2015), who used a shift in pH generated by the release of free fatty acids to measure lipase activity. In this work, the lipases were placed in a 1 mM Tris-HCl buffer with phenol red to monitor the decrease of pH during the hydrolysis of various triglycerides. A standard curve, used to calculate the acid concentration at a given pH (Figure S2), was generated by titrating a phenol red solution with acetic acid. The ionisation of longer-chain organic acids differs somewhat from that of acetic acid, but we were interested in the differences between various lipases and not the precise lipase activity of individual variants. The substrates were tributyrin, trioctanoate, triolein (C18), and olive oil, which consists of saturated and unsaturated long chain (16-22C) triglycerides (Quintero-Flórez et al. 2015). The homo-triglyceride substrates were prepared in 1 mM Tris-HCl buffer, suspended by sonicating for 15 minutes and then 0.07 mM phenol red was added. Olive oil was prepared in the same way, except that only 0.04 % v/v of the oil was used. The assays were done in triplicate. The total volume of each assay was 200 μL, of which

10 μL was enzyme solution and 190 μL was suspended substrate. As the *wt* enzyme had worse activity than the variants, ten-fold more *wt* enzyme was used to obtain a detectable change in pH; the activity / per mole of enzyme was used for comparisons. The shift in pH was measured after 1 hour. The results could be used for comparisons and to indicate activity with various substrates.

Modelling studies

No experimental structure is currently available for Lip3. The CONSURF server (consurf.tau.ac.il) was used to determine the conserved residues and the most similar molecular structures (Ashkenazy et al. 2010). Analysis of the amino acid sequence revealed two structures with the highest similarity: 1K8Q dog gastric lipase (39.4% identity) and 1HLG human gastric lipase (38.3% identity). The structure of Lip3 was generated using ROSETTA as implemented in the ROBETTA server (http://robetta.bakerlab.org/) with default parameters. The starting model was dog gastric lipase (1K8Q). The structures were visualized with PYMOL (http://pymol.sourceforge.net/). The localization of the lid in the structure was done by protein sequence alignment of CG8823 lipase with 1K8Q sequence and the lid was defined based on previously published work (Khan et al. 2017).

Results

These experiments were carried out in three phases as shown schematically in Figure 1. In the first phase, the CG8823 gene was cloned into a plasmid for expression so that Lip3 could be characterized to provide data for designing the high-throughput screen used for the evolution in the next phase.

The evolution was carried out in the second phase, which cycled through three stages (Figure 1). In the first stage, large libraries of mutant genes were generated using well established protocols described in Methods (Arnold 1998, 2018; Jackson et al. 2013; Porter et al. 2016). In round 1, the *wt* gene was the starting gene for these experiments, while in later rounds, the combined pool of genes coding for the best variants from the previous round were used to generate the libraries. In the second and third stages, these libraries were screened for increased Lip3 activity. In the primary screen, large libraries (30 to 40,000) were tested for lipolytic activity with an agar plate assay. This was possible because *E. coli* does not have any lipases. This assay did not lend itself to a quantitative assessment of variant activity, but it could be used to identify a few hundred variants to be tested in a secondary screen. In addition to lipolytic activity, lipases have an esterase activity that lends itself to a more accurate plate assay, which was used in the final stage of the stage of evolution.

The purpose of the final phase was to test the best variants from the final round of evolution. These were purified and characterised to assess their suitability for further studies.

Phase 1: Expression and design of a high throughput assay

Codon-optimised CG8823 was cloned into the pETMCSIII plasmid (Neylon et al. 2000). This construct resulted in the fusion of an N-terminal His_6 tag so that the resulting Lip3 could be purified in a single step (Stevenson et al. 2012). This plasmid does not have a LacI gene so there is leaky expression when transformed cells are cultured in LB (Luria-Bertani) medium (Studier 2005). High levels of expression of *wt* CG8823 could be achieved in BL21 (DE3) when

the cells were grown at 37^{o} C, but the protein was mostly insoluble (Figure 2C). Low yields of soluble enzyme could be obtained by lowering the temperature to 30^{o} C (Figure 2C) or by increasing the aeration of the culture (Figure S1). The Lip3 protein was purified as described in Methods and its specific activity measured with pNP esters of C3, C8, C16 and C18 acids. The enzyme was most active with C8 esters. The activity with the C3 ester was four-fold lower than with the C8 ester, and there was little activity with the C16 and C18 substrates.

A high-throughput assay was developed to screen libraries. The method used agar plates in which lipids were added to the agar to make it cloudy. Activity was detected when the cloudiness disappeared around colonies that expressed active lipases. It was important to establish the conditions needed to detect the activity of the *wt* enzyme. Cells transformed with the wild-type gene were inoculated onto agar plates with ampicillin (amp) and lipids to determine the time of incubation required and the conditions for the formation of haloes. The variables in the design were: growth medium (M9 minimal media and LB), inducer (glucose and lactose), lipid (tributyrin and glyceryl tri-octanoate both in 0.06% v/v), calcium added as $CaCl_2$ (0, 0.2, 0.5, 1 and 2 mM) and temperature of incubation (30^{o} C and 37^{o} C). Haloes formed around colonies grown at 30° C on agar plates with LB, 1 mM lactose, 0.06% v/v glyceryl trioctanoate, and 0.2 mM Ca^{2+}. The haloes were easily seen after 5 days. More stringent conditions were needed to screen for variants with enhanced activity (see Methods).

Phase 2: Evolution

In stage 1, genes were randomly mutated (see Methods) by multiple ep-PCR reactions with short extension times to ensure that the altered genes were shuffled (Powell et al. 2001; Zhao

et al. 1998). The screening process was carried out in two stages. In the primary screen (Stage 2 in Figure 1), LB agar plates containing ampicillin and glyceryl trioctanoate were used to select colonies based on their ability to form haloes around them. In the initial rounds, few colonies formed haloes while in later rounds most did, so the incubation time was reduced and the incubation temperature increased so as to increase the stringency of selection (Table 1, Figure 3). From each round, we obtained between 384 and 1260 colonies that formed haloes (Table 1). Although this procedure allowed identification of variants with enhanced ability to degrade lipids, the discrimination between active variants was poor as halo size could not be measured accurately by eye. We therefore devised a more accurate secondary screen to identify the most active variants. In it, we transferred selected colonies to 96-well plates, cultured them, and assayed lysed cells for hydrolytic activity towards *p*-nitrophenyl octanoate. The activity of crude lysates increased steadily throughout the evolution, with a 200-fold improvement for variants in round seven (Figure 3, Table 2). In the final stage of each round, the genes of the variants with the highest activities were isolated, sequenced (to eliminate duplicates) and used as parents for the next generation. Table S1 shows the sequences of all the selected variants from each round. The average number of changes introduced in each round varied a great deal. Round 6 resulted in an average of 2.7 changes while in round 3, the average was *—1.6* changes: that is, there were on average more reversions to *w*t than changes to new sequences. At the end of round 7, there were an average of 6.4 changes in the sequence compared to the *wt* enzyme; overall, there was approximately one change in amino acid sequence per round of evolution.

Seven rounds of directed evolution were sufficient to identify variants of Lip3 with greatly improved activity as measured in crude lysates. Most of the changes that occurred in a single

round were neutral, as they were lost in subsequent rounds (Table S1). However, some of the changes that occurred in the early rounds were preserved and found in close to all the final round variants. For example, the P291L change occurred in round 1 and reached saturation point in round 6, while N268I first appeared in round 2 and was found in all but one variant selected in rounds 6 and 7. There are three changes, P291L, N268I and V201A, which occur in almost all the round 7 variants and occur in the five best variants that were purified and characterized. In addition, the M321L change occurs in 27 of the thirty round 7 variants and 4 of the 5 best variants of that round, indicating that these four changes are largely responsible for the observed improvements in stability and activity. There are some changes that occur in less than half of the sequences (see Table 3) but have persisted for a few rounds of evolution and are found in the five best variants. For example, the T250A change occurs in 13 of the 31 sequences and is found in one of the best five variants. This is also true for other changes (Table 2). From the sequence data, it was clear that the evolution had not converged under the current selection conditions. However, given that variants had about a 200-fold increase in activity as measured in crude lysates, we thought it appropriate to purify and characterise some of them.

Phase 3: Characterisation

After seven rounds of evolution, the *w*t and five variants were purified and tested for solubility, stability and catalytic activity. These five variants (R7_59A, R7_82E, R7_47D, R7_47E

and R7_12H) were chosen because they exhibited good activity and had a mix of changes[1]. They contained 5-9 amino acid changes, three of which were conserved: V201A, N268I and P291 (Table 2). The latter two changes occurred as single site changes in the first two rounds in variants R1_15B and R2_312H: hence these two variants were also selected for characterisation. Cells expressing the variants and *wt* Lip3 were grown under the same conditions (see Methods). The yields varied greatly: while only small quantities of the *wt* could be isolated due to poor solubility, the variants were obtained in significantly higher yield. Evolution had produced variants that enabled large quantities of purified protein to be easily obtained. However, it was not clear whether the increased production was due to increased solubility or expression.

To gauge the relative importance to expression and solubility, we compared the effects of temperature on the expression and solubility of variant R7_59A (the most active variant, see below) and *wt* by examining the total protein found in lysate, supernatant (soluble fraction) and pellet (insoluble fraction) by SDS-PAGE (Sodium dodecyl sulfate polyacrylamide gel electrophoresis) (Figure 2C). The different variants were grown under the same conditions, but grew at different rates and to different final optical densities (OD 680). The amounts of crude lysate loaded onto gels were normalised to account for differences in cell density. Temperature affected both the level of expression and the amount of soluble protein that was produced. The total protein in the lysate was similar for *wt* and R7_59A. There was a small amount of soluble *wt* protein at 19^{o} C, but this band is absent at higher temperatures (Figure

[1] The designation R7_59A denotes the round of evolution (7), the tray number (5) and the well coordinate within the tray (9A).

2C). In contrast, the amount of soluble variant increased with temperature, though there was also almost as much insoluble as soluble protein. It appeared that evolution had achieved increased activity in crude lysates by increasing the production of soluble protein. An attempt was made to quantitate the increase in solubility.

Solubility tests were performed with variant R7_59A and the *wt* protein. Purified samples of both proteins were concentrated in an Amicon concentrator. The *w*t protein started to precipitate at just 5 mg/mL while the variant did not precipitate at 100 mg/ml.

The thermostability of *wt* and selected variants were assessed by measuring the residual activity following incubation at elevated temperatures for 45 minutes, as described in Methods. The *wt* protein had a T_{50} value of 37° C, while the singly-mutated R1_15B (P291L) and R2_312H (N268I) had T_{50} values of 42.2 and 39° C, respectively. The selected variants from round 7 showed further increases in stability with T_{50} values that ranged from 45° C to 52 °C, corresponding to improvements of 8 to 16° C compared to *wt* (Table 2).

To determine if evolution affected the catalytic properties of the variants, their specific activities were measured. The specific activities of the *wt* and variants were measured for *p*-nitrophenyl esters: propionate pNP-C3, octanoate pNP-C8, palmitate pNP-C16 and stearate pNP-C18 (Table 2). The specific activities were measured with the same substrate concentrations as used for the C8 substrate in the secondary screen. All the variants from round 7 have similar activities toward pNP-C8, with R7_59A exhibiting the highest activity for pNP-C3 and pNP-C16 while R7_47D exhibited the highest for pNP-C18. Variants R7_47E and R7_82E showed very low activity on pNP esters with long chains (C16 and C18). The C8 substrate was used for selection in the evolution experiments and the activity with this

substrate is still higher than for other substrates, but the activities of all the other substrates exhibit greater enhancement. None of the enzymes had high activity with the long chain esters (C16 and C18). The variants identified in the early rounds show a decrease in specific activity but an increase in activity when measured with crude lysate, again suggesting that the solubility of the variants has been preferentially increased in these early rounds.

The esterase activities of the *wt* and variants were measured as a function of substrate concentration. When plotted, these data gave poor agreement with a two-parameter Michaelis-Menten curve. The shape of the curves could be reasonably described with a three-parameter substrate inhibition equation (Methods) (Figure 4). Given that the wt data were collected with a sample that was ten-fold more concentrated, the k_{cat} of R7_59A has increased, possibly by as much as forty-fold (≈27,000 min^{-1} for *wt* to over 100,000 for R7_59A). However, there are compensating changes in K_M and K_I so that observed maximum rate has increased by about a factor of 20. The variant and wt curves have very similar shapes with maxima that occur at the concentration of substrate used for selection in the secondary screen. Native gels (data not shown) obtained with the variants displayed multiple bands consistent with the variants existing in more than one oligomeric form. The native gels used for the variant enzymes failed to give any indication of the *wt* enzyme, due to its poor solubility and stability. Evolution appeared to have increased the activity of the variants, with increases in both the K_M and K_I such that maximal activity occurs at higher substrate concentrations. However, the enzymes are still substrate-inhibited: evolution has not relieved that aspect of the enzyme kinetics. The substrate inhibition may be due to non-productive binding of the alkyl chain in the active site or cap domain.

The lipase activity of the *wt* and round 7 variants was also measured (Figure 5). We used glyceryl tri-octanoate for the primary selection because the wild-type enzyme is more active on this than on tributyrin. The same result was obtained for the evolved variants suggesting, as expected, that directed evolution did not greatly alter the substrate specificity. The design of the primary screen ensured that both solubility and the lipolytic activity was selected for. As a result, the final round variants exhibited enhanced lipolytic activities with short chain tri-acid lipids (C4 and C8). These variants also exhibited low levels of activity with coconut oil (Figure 5).

Analysis of changes

Although a structure of Lip3 was not available for this work, there were structures of a dog lipase in the open and closed conformations. These were used to generate model structures of Lip3 with ROSETTA, as described in Methods. The observed sequence changes were mapped onto a schematic diagram of the Lip3 model structure (Figure 6) and an examination of the Rosetta structure provided some insight into the possible effects of the changes. Most of the changes occur in the substrate binding domain (V201A, N268I and P291L). These three changes were saturating; that is, they were present in all round 7 variants. Mutations P291L (R1_15B) and N268I (R2_312H) increase the stability of the protein. P291 is positioned within a helix composed of residues 288 to 294 (Figure 6A) and the substitution of a proline, which is a helix destabilising residue, with a leucine would likely stabilise the helix through more favourable hydrogen bond formation. The sidechain of N268 is found in a hydrophobic pocket: replacement of a polar asparagine residue with the hydrophobic isoleucine would strengthen the hydrophobic interactions, which often occurs in thermophilic proteins (Gromiha et al.

2013; Razvi and Scholtz 2006; Sengupta and Kundu 2012). Sequence comparisons using CONSURF (Table 3) indicate that residues at positions 201, 268 and 291 are highly variable. These changes appear to affect the stability of the protein and do not appear to be in positions to affect the way substrates might bind. In addition to saturating changes in the cap domain, the change N250I is found in about half of the round 7 variants. Again, this change appears to promote the stability of the domain as the Ile residue is a better fit in a hydrophobic cavity. This change scores 8 (of a possible 9) on the CONSURF scale; it is a conservative change, unlike the saturating changes found in the cap domain. Some of the changes in the cap domain are close to the hinge region of the lid (S210T and T225A). The exact effect of these changes on the lid is unclear without direct structural context. Neither residue is highly conserved in other sequences, but S210 scores 6 on the CONSURF conservation scale. In addition, these residues occur in less than a third of the round 7 variants.

There are a few changes that occur in the core domain of the protein and of these, two (M321L and K347E), are close to residues that form the catalytic triad: they probably affect catalysis. M321L occurs in all 30 of the round 7 variants and is highly conserved with a score of 8 on the CONSURF scale, while K347 is only present in about a third of the round 7 variants and is a variable residue with a score of 2 on the CONSURF scale. There is also a change in the oxy-anion hole (M59L), but this only occurs in 3 variants. The three changes in the first 50 residues, M20V, Y40F and E46G/K, occur in three, seven and two variants, respectively.

Discussion

Seven rounds of evolution were sufficient to meet the objectives set out at the start of this project: to take an enzyme with poor physical and catalytic properties and evolve it to identify variants with appropriate catalytic activities that could be easily purified to give an abundant supply of soluble and stable protein. Only a small number of mutations were necessary to meet these objectives. The pETMCSIII vector resulted in good expression of the *wt* and variant proteins. This expression did not appear to change during the evolution, but the solubility of the variants did. About half of the expressed R7_59A variant was soluble; this was sufficient to provide an abundant supply of purified protein. A single change in the first round was sufficient to give improved solubility while five to nine changes were sufficient to give improved solubility, stability and activity. There were four changes observed in most of the round 7 variants, a few of which were observed in earlier rounds. Evolution produced modest increases in specific activity with C8 substrates.

The *wt* enzyme and a selection of round 7 variants were purified and characterised. Although the characterisation confirmed that variants with acceptable properties had been identified, it did not appear that the evolution had converged. A few changes that appeared in earlier rounds became more common in the later rounds, suggesting that further improvements could have been obtained with more rounds of evolution. The final round variants exhibited low levels of activity with long chain aliphatic lipids and esters: they were therefore suitable starting points for evolution towards improved activity with these substrates, the subject of the next manuscript. Are the results we obtained generally applicable to other systems?

First, *wt* Lip3 expressed well and there was no discernible change in the expression of the variants during the seven rounds of evolution. Good expression is actually a prerequisite for this approach to work. Second, the improvements in solubility could be identified because the *wt* enzyme had a detectable amount of the required activity and a screen could be devised to test the activities of variants in a large library. Significant improvements were observed in the activity of variants as observed with crude lysates, but their specific activities showed only modest improvements. Third, we were fortunate that *E. coli* does not have lipase activity, as this underpins our screening strategy. As evolution progressed it was possible to alter conditions to make selection more stringent so we could identify the best variants.

Solubility and stability

Our results (Table 2) indicate that evolution and selection increased both the solubility and stability of Lip3. Why is this so? In its broadest sense, solubility refers to the amount of protein in solution, whether it be correctly folded and active or mis/un-folded and inactive. Our selection was for increased activity, which depends on correctly-folded protein. For this reason, we have assumed that increased solubility implies an increase in the amount of correctly folded and active protein. Our selection conditions demanded that variants fold correctly at 37° C, and gave no advantage to variants that might retain activity at higher temperatures; there was no selection pressure for an increase in stability. Strictly speaking, stability refers to the amount of free energy required to (reversibly) unfold a protein: practically, $T_{1/2}$ is measured using a non-equilibrium assay, and so its value depends upon the method used to measure unfolding. We used residual activity after heat treatment as a proxy for protein thermostability. In principle, therefore, increased protein solubility could have

been achieved without increased thermostability, as thermostability was not selected for. However, in all the round 7 variants there is an increase in $T_{1/2}$ over that of the *w*t enzyme. One explanation for this observation is that increased solubility is achieved by increasing the stability of the variants. This increase could be achieved via different changes in the variants.

Analysis of changes

Given the location and frequency of changes, it appears that the first-priority in producing more active enzyme in crude lysate was to stabilize the cap domain. Following this, the catalytic ability of variants was enhanced by changes near the triad residues along with changes near to the hinge points of the cap domain. Finally, sections at the start of the core structure appear to be changing, presumably to enhance the stability of the protein.

Prior use of directed evolution on lipases

Directed evolution has been applied to lipases previously (Schmidt et al. 2004; Svendsen 2000). In a few of these studies, the stereochemistry of the reaction catalysed by the lipase has been modified (Liebeton et al. 2000; Manfred T. Reetz et al. 1997). In other cases, the stability of the protein has been enhanced with respect to thermal stability (Akbulut et al. 2013; Khurana et al. 2011; Y. Liu et al. 2017; Madan and Mishra 2014; Yu et al. 2012) or organic solvents (Korman et al. 2013; Ogino and Ishikawa 2001; Manfred T. Reetz et al. 2010; Stepankova et al. 2013). In these studies, the investigators started with proteins that were soluble, stable and could be easily produced. It was thought that these starting enzymes were an ideal starting point for evolution. The present study is unusual in that the starting enzyme had none of these desired properties. However, a series of variants has been generated that

have the desired properties. In addition, we argue that our resulting series of variants constitute a better starting point for further evolution, as sequence differences will give a more diverse library. Some variants are more active while others are more stable so that a change in selection conditions may favour some, but not all the changes generated in the present study. In short, future evolution should be accelerated by using the genes of multiple variants as a starting point.

We propose that further evolution against a new, technologically-relevant target could lead to a useful product. For example, one might select for enhanced activity with substrates possessing longer aliphatic chains that occur in fats and oils. Some variants have activity with olive oil; one might evolve with the aim of enhancing the activity towards this and other natural oils for use in, for instance, a dishwashing detergent. Evolution towards these more ambitious targets would not have been achievable with the wt enzyme but are now possible by building on these newly evolved soluble enzymes. Such evolution is the subject of the next paper in this series.

Acknowledgments.

We acknowledge the staff at the BRF for performing the DNA sequencing. The authors wish to thank CSIRO for providing protein purification facility. A.A. is supported by a SENESCYT scholarship from the Ecuadorian government and A.G. acknowledges funding from the BBSRC (BB/M021610) and the Academy of Finland (1286429).

.

References

Akbulut, Nagihan, et al. (2013), 'Improved activity and thermostability of Bacillus pumilus lipase by directed evolution', *Journal of Biotechnology,* 164 (1), 123-29.

Arnold, Frances H. (1998), 'Design by Directed Evolution', *Accounts of Chemical Research,* 31 (3), 125-31.

--- (2018), 'Directed Evolution: Bringing New Chemistry to Life', *Angewandte Chemie (International ed. in English),* 57 (16), 4143-48.

Ashkenazy, Haim, et al. (2010), 'ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids', *Nucleic Acids Research,* 38 (Web Server issue), W529-W33.

Camacho-Ruiz, María de los Angeles, et al. (2015), 'A broad pH range indicator-based spectrophotometric assay for true lipases using tributyrin and tricaprylin', *Journal of Lipid Research,* 56 (5), 1057-67.

Carr, P. D. and Ollis, D. L. (2009), 'Alpha/beta hydrolase fold: an update', *Protein Pept Lett,* 16 (10), 1137-48.

Gromiha, M. Michael, et al. (2013), 'Hydrophobic environment is a key factor for the stability of thermophilic proteins', *Proteins: Structure, Function, and Bioinformatics,* 81 (4), 715-21.

Gupta, Namita, Rathi, Pooja, and Gupta, Rani (2002), 'Simplified para-nitrophenyl palmitate assay for lipases and esterases', *Analytical Biochemistry,* 311 (1), 98-99.

Gurung, Neelam, et al. (2013), 'A Broader View: Microbial Enzymes and Their Relevance in Industries, Medicine, and Beyond', *BioMed Research International,* 2013, 329121.

Hasan, Fariha, et al. (2010), 'Enzymes used in detergents: Lipases', *African Journal of Biotechnology,* 9 (31), 4836-44.

Houde, Alain, Kademi, Ali, and Leblanc, Danielle (2004), 'Lipases and their industrial applications', *Applied Biochemistry and Biotechnology,* 118 (1), 155-70.

Jackson, Colin J., et al. (2013), 'Structure and function of an insect α-carboxylesterase (αEsterase7) associated with insecticide resistance', *Proceedings of the National Academy of Sciences,* 110 (25), 10177-82.

Jemli, Sonia, et al. (2016), 'Biocatalysts: application and engineering for industrial purposes', *Critical Reviews in Biotechnology,* 36 (2), 246-58.

Kaiser, Peter M. (1980), 'Substrate inhibition as a problem of non-linear steady state kinetics with monomeric enzymes', *Journal of Molecular Catalysis,* 8 (4), 431-42.

Khan, Faez Iqbal, et al. (2017), 'The Lid Domain in Lipases: Structural and Functional Determinant of Enzymatic Properties', *Frontiers in Bioengineering and Biotechnology,* 5, 16.

Khurana, J., Singh, R., and Kaur, J. (2011), 'Engineering of Bacillus lipase by directed evolution for enhanced thermal stability: effect of isoleucine to threonine mutation at protein surface', *Mol Biol Rep,* 38 (5), 2919-26.

Kirk, Ole, Borchert, Torben Vedel, and Fuglsang, Claus Crone (2002), 'Industrial enzyme applications', *Current Opinion in Biotechnology,* 13 (4), 345-51.

Korman, Tyler P., et al. (2013), 'Dieselzymes: Development of a stable and methanol tolerant lipase for biodiesel production by directed evolution', *Biotechnology for Biofuels,* 6 (1), 70-70.

Liebeton, K., et al. (2000), 'Directed evolution of an enantioselective lipase', *Chem Biol,* 7 (9), 709-18.

Liu, Min, et al. (2001), 'Catalytic Acid−Base Groups in Yeast Pyruvate Decarboxylase. 1. Site-Directed Mutagenesis and Steady-State Kinetic Studies on the Enzyme with the D28A, H114F, H115F, and E477Q Substitutions', *Biochemistry,* 40 (25), 7355-68.

Liu, Yihan, et al. (2017), 'Improvement in thermostability of an alkaline lipase I from Penicillium cyclopium by directed evolution', *RSC Advances,* 7 (61), 38538-48.

Lobedanz, Sune, et al. (2000), 'Enzymes in Industrial Biotechnology', *Kirk-Othmer Encyclopedia of Chemical Technology* (John Wiley & Sons, Inc.).

Madan, Bhawna and Mishra, Prashant (2014), 'Directed evolution of Bacillus licheniformis lipase for improvement of thermostability', *Biochemical Engineering Journal,* 91, 276-82.

Neylon, C., et al. (2000), 'Interaction of the Escherichia coli replication terminator protein (Tus) with DNA: a model derived from DNA-binding studies of mutant proteins by surface plasmon resonance', *Biochemistry,* 39 (39), 11989-99.

Ogino, Hiroyasu and Ishikawa, Haruo (2001), 'Enzymes which are stable in the presence of organic solvents', *Journal of Bioscience and Bioengineering,* 91 (2), 109-16.

Ollis, David L., et al. (1992), 'The α/β hydrolase fold', *Protein Engineering,* 5 (3), 197-211.

Pistillo, D., et al. (1998), 'The Drosophila melanogaster lipase homologs: a gene family with tissue and developmental specific expression1', *Journal of Molecular Biology,* 276 (5), 877-85.

Porter, Joanne L., Rusli, Rukhairul A., and Ollis, David L. (2016), 'Directed Evolution of Enzymes for Industrial Biocatalysis', *ChemBioChem,* 17 (3), 197-203.

Powell, Keith A., et al. (2001), 'Directed Evolution and Biocatalysis', *Angewandte Chemie International Edition,* 40 (21), 3948-59.

Quintero-Flórez, Angélica, et al. (2015), 'The Fatty Acid Composition of Virgin Olive Oil from Different Cultivars Is Determinant for Foam Cell Formation by Macrophages', *Journal of Agricultural and Food Chemistry,* 63 (30), 6731-38.

Razvi, Abbas and Scholtz, J. Martin (2006), 'Lessons in stability from thermophilic proteins', *Protein Science : A Publication of the Protein Society,* 15 (7), 1569-78.

Reetz, M. T. and Jaeger, K. E. (1999), 'Superior biocatalysts by directed evolution', (200; BERLIN: SPRINGER-VERLAG BERLIN), 31-57.

Reetz, Manfred T., et al. (1997), 'Creation of Enantioselective Biocatalysts for Organic Chemistry by In Vitro Evolution', *Angewandte Chemie International Edition in English,* 36 (24), 2830-32.

Reetz, Manfred T., et al. (2010), 'Increasing the stability of an enzyme toward hostile organic solvents by directed evolution based on iterative saturation mutagenesis using the B-FIT method', *Chemical Communications,* 46 (45), 8657-58.

Roussel, Alain, et al. (1999), 'Crystal Structure of Human Gastric Lipase and Model of Lysosomal Acid Lipase, Two Lipolytic Enzymes of Medical Interest', *Journal of Biological Chemistry,* 274 (24), 16995-7002.

Roussel, Alain, et al. (2002), 'Crystal Structure of the Open Form of Dog Gastric Lipase in Complex with a Phosphonate Inhibitor', *Journal of Biological Chemistry,* 277 (3), 2266-74.

Schmidt, Marlen, et al. (2004), 'Directed Evolution of Lipases and Esterases', *Methods in Enzymology* (388: Academic Press), 199-207.

Sengupta, Dhriti and Kundu, Sudip (2012), 'Role of long- and short-range hydrophobic, hydrophilic and charged residues contact network in protein's structural organization', *BMC Bioinformatics,* 13 (1), 142.

Sheldon, Roger A. and Pereira, Pedro C. (2017), 'Biocatalysis engineering: the big picture', *Chemical Society Reviews,* 46 (10), 2678-91.

Stepankova, Veronika, et al. (2013), 'Strategies for Stabilization of Enzymes in Organic Solvents', *ACS Catalysis,* 3 (12), 2823-36.

Stevenson, Bradley J., Liu, Jian-Wei, and Ollis, David L. (2008), 'Directed Evolution of Yeast Pyruvate Decarboxylase 1 for Attenuated Regulation and Increased Stability', *Biochemistry,* 47 (9), 3013-25.

Stevenson, Bradley J., et al. (2012), 'Fermentative glycolysis with purified Escherichia coli enzymes for in vitro ATP production and evaluating an engineered enzyme', *Journal of Biotechnology,* 157 (1), 113-23.

Studier, F. William (2005), 'Protein production by auto-induction in high-density shaking cultures', *Protein Expression and Purification,* 41 (1), 207-34.

Svendsen, Allan (2000), 'Lipase protein engineering', *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology,* 1543 (2), 223-38.

Turner, Nicholas J. (2003), 'Directed evolution of enzymes for applied biocatalysis', *Trends in Biotechnology,* 21 (11), 474-78.
--- (2009), 'Directed evolution drives the next generation of biocatalysts', *Nature Chemical Biology,* 5, 567.
Wilson, D. S. and Keefe, A. D. (2001), 'Random mutagenesis by PCR', *Curr Protoc Mol Biol,* Chapter 8, Unit8.3.
Yu, Xiao-Wei, et al. (2012), 'Enhanced thermostability of a Rhizopus chinensis lipase by in vivo recombination in Pichia pastoris', *Microbial Cell Factories,* 11 (1), 102.
Zhao, H., et al. (1998), 'Molecular evolution by staggered extension process (StEP) in vitro recombination', *Nat Biotechnol,* 16 (3), 258-61.

Tables

Table 1. Experimental details of the continuous generation during directed evolution

Generation	Primary Screening			Secondary Screening		Improvement in activity[1]
	Library	Variant[3]	Conditions[2]	Variant[4]	Condition[5]	
1	40000	672	72 h, 30° C	10	30° C	10
2	40000	768	48 h, 30° C	12	30°C	12
3	40000	576	48 h, 30° C	10	30° C	22
4	40000	480	45 h, 30° C	7	30° C	15
5	32000	1248	24-48 h, 37° C	20	37°C	28
6	32000	384	24 h, 37° C	18	37° C	73
7	48000	960	24 h, 37° C	30	37° C	222

[1] Measured with pNP-C8 on the crude lysate for the best mutant of each round.

[2] Conditions became more stringent as evolution progressed; time required for halo formation decreased from 72 to 24 hours and the incubation temperature was increased from 30 to 37° C. For round 5, two sets of agar plates were used, one was incubated for 24 and the other was incubated for 48 hours. Ten variants were selected from each set.

[3] The number of variants selected in the primary screen.

[4] The number of variants selected in the secondary screen (see S1 for sequences).

[5] This is the temperature at which the selected colonies were grown in 96-well plates for a later assay for esterase activity.

Table II. Changes present in variants after seven rounds of directed evolution

	Changes																Improvement of activity in crude lysate[1]	Yield[2]	T_{50}[3] °C	Activity of Purified enzyme[4]			
Residue #	**2**	**20**	**40**	**59**	**101**	**188**	**201**	**210**	**225**	**250**	**268**	**269**	**281**	**291**	**305**	**321**	**pNP-C8**			**pNP-C3**	**pNP-C8**	**pNP-C16**	**pNP-C18**
WT	G	M	Y	M	K	F	V	S	T	E	N	L	R	P	K	M	1	2.2	37.3°C	15.9	96.6	0.4	1.6
R1_15B														L			10	ND	42.2°C	0.5	43.9	0.001	0.1
R2_312H											I						12	ND	39.0°C	1.8	55.3	0.01	0.03
R7_59A			F	L			A				I		C	L	R	L	218	340.8	45.0°C	99.9	147.6	3.1	4.2
R7_82E	R				R	L	A		A	G	I			L		L	198	235.8	49.6°C	74.1	148.1	0.8	0.9
R7_47D		V					A				I			L		L	222	172	52.9°C	71.4	148.8	2.6	5
R7_47E							A	T			T			L		L	193	24	51.5°C	56.9	150.1	0.2	0.3
R7_12H		V				L	A		A		I	Q		L			135	174	51.7°C	59	148.1	2.1	3.3

[1] ratio of activity measured with pNP-C8 on the crude lysate
[2] Protein purified in mg/L culture
[3] Temperature at which 50% of the original activity remains after the thermal attack
[4] Activity measured using purified enzyme on pNP-esters and expressed in µM nitrophenol/min/µg protein

Table 3. Position of each mutation in the model structure for CG8823

#	ID	20	40	46	59	174	185	188	201	210	211	225	230	250	268	291	305	321	347	351	# MUTs
	CG8823	**M**	**Y**	**E**	**M**	**A**	**T**	**F**	**V**	**S**	**N**	**T**	**D**	**E**	**N**	**P**	**K**	**M**	**K**	**L**	
1	R7_58D								A						T	L		L			4
2	R7_412F		F						A					G	I	L		L	E		8
3	R7_12H	V						L	A			A			I	L					7
4	R7_69H								A					G	I	L		L			6
5	R7_58F								A					G	I	L		L	E		6
6	R7_48A	V		G					A					G	I	L		L			9
7	R7_69G								A					G	I	L		L	E		6
8	R7_611E								A			A			I	L		L			7
9	R7_410H								A					G	I	L		L	E		6
10	R7_59A		F		L				A						I	L	R	L			8
11	R7_16D								A							L		L			3
12	R7_112D								A	T					T	L		L			5
13	R7_712E		F						A					G	I	L		L	E		7
14	R7_73H								A					G	I	L		L	E		6
15	R7_67B		F						A					G	I	L		L			8
16	R7_612G								A	T					T	L		L			5
17	R7_64C								A					G	I	L		L	E		6
18	R7_412D				L	V			A	T	S	A				L		L			8
19	R7_612B			K			A		A	T			G		I	L		L		Q	9
20	R7_48B		F						A					G	I	L		L	E		7
21	R7_107G								A					G	I	L		L	E		6
22	R7_47E								A	T					T	L		L			5
23	R7_82E							L	A			A		G	I	L		L			8
24	R7_106H					V			A	T		A			I	L		L			10
25	R7_410F								A	T		A			T	L		L			6
26	R7_106G								A	T					T	L		L			5
27	R7_411A		F		L				A			A			I	L		L			8
28	R7_47D	V							A						I	L		L			5
29	R7_67A		F						A					G	I	L					6
30	R7_95D					V			A	T			G		I	L		L			9
	Conservation Scale (consurf)	4	2	1	6	7	2	4	1	6	4	1	1	8	2	1	1	8	2	6	
	Position	e	e	e	b	b	b	b	b	e	3	e	e	e-f	b	e	e	b	e	b	

Figure legends

Figure 1. Schematic diagram to illustrate the stages in evolution. Primary screening involved selection with LB-amp agar plates enriched with 2% glyceryl trioctanoate. In the secondary screening the hydrolytic activity was measured with p-nitrophenyl octanoate ester.

Fig. 2 Comparison of evolution. **(A)** SDS-PAGE gels of fractions after purification of several variants of lip3 using 5 mL HisTrap column. Wild-type (I), variants R7_59A (II) and R7_47E (III). **(B)** Elution profiles of the lipases lip3 wild-type (I) and the variants R7_59A (II) and R7_47E (III). (**C**) Solubility test of mutant R7_59A and wt at different temperatures of incubation A=19°C, B=30°C, C=37°C.

Fig. 3 Screening activity of crude lysate during the directed evolution. Grey shadow shows the reduction of incubation time to see the haloes. Blue bars correspond to the increment of activity of the crude lysate in each round to hydrolyse pNP-C8. Spacing of the generations on the horizontal axis is indicative of the number of variants selected in the secondary screen. The orange circle shows the variants selected for final characterization.

Fig. 4 Activity profile using pNP-C8 showing the activity for *wt* (squares) and R7_59A (circles). *wt* lip3 was measured using 60 nM of enzyme, while R7_59A was measured with 6 nM of enzyme**.** Curve fitting was done in GraphPad Prism v7, yielding the following parameters: k_{cat}: $2.7*10^4$ and $10*10^4$ (min^{-1}); K_M = 0.0051 and 0.075 mM; K_i = 0.010 and 0.026 mM for *wt* and variant R7_59A respectively.

Figure 5. Lipolytic activity with phenol red using different substrates

Figure 6. Schematic diagram of lipase CG8823. Arrows represent β sheets, barrels α-helices. Catalytic triad S145, D318 and H350 are in green and residues of the oxyanion hole in orange. The most common mutations are in green circles while the other mutations are in blue circles.

Figures (artwork)

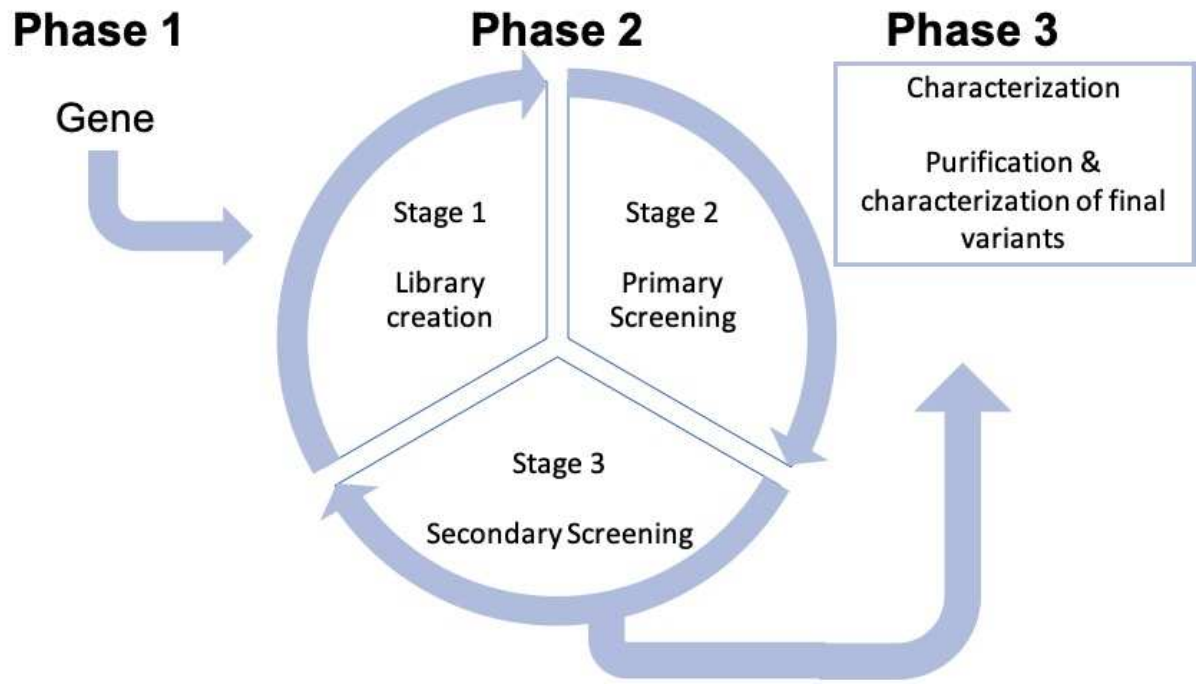

Fig. 1 Schematic diagram to illustrate the stages in evolution. Primary screening involved selection with LB-amp agar plates enriched with 2% glyceryl trioctanoate. In the secondary screening the hydrolytic activity was measured with p-nitrophenyl octanoate ester.

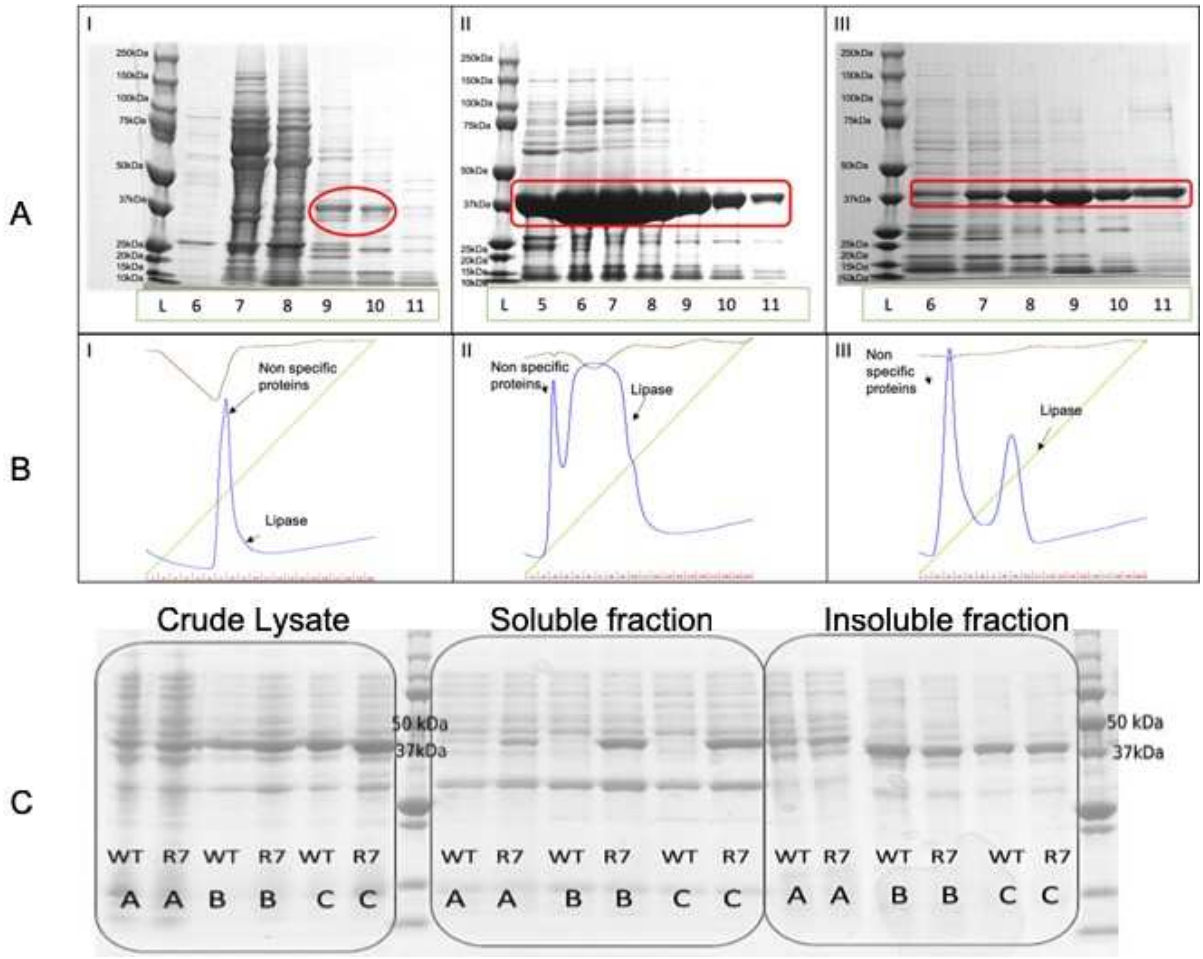

Fig. 2 Comparison of evolution. **(A)** SDS-PAGE gels of fractions after purification of several variants of lip3 using 5 mL HisTrap column. Wild-type (I), variants R7_59A (II) and R7_47E (III).**(B)** Elution profiles of the lipases lip3 wild-type (I) and the variants R7_59A (II) and R7_47E (III). (**C**) Solubility test of mutant R7_59A and wt at different temperatures of incubation A=19° C, B=30° C, C=37° C.

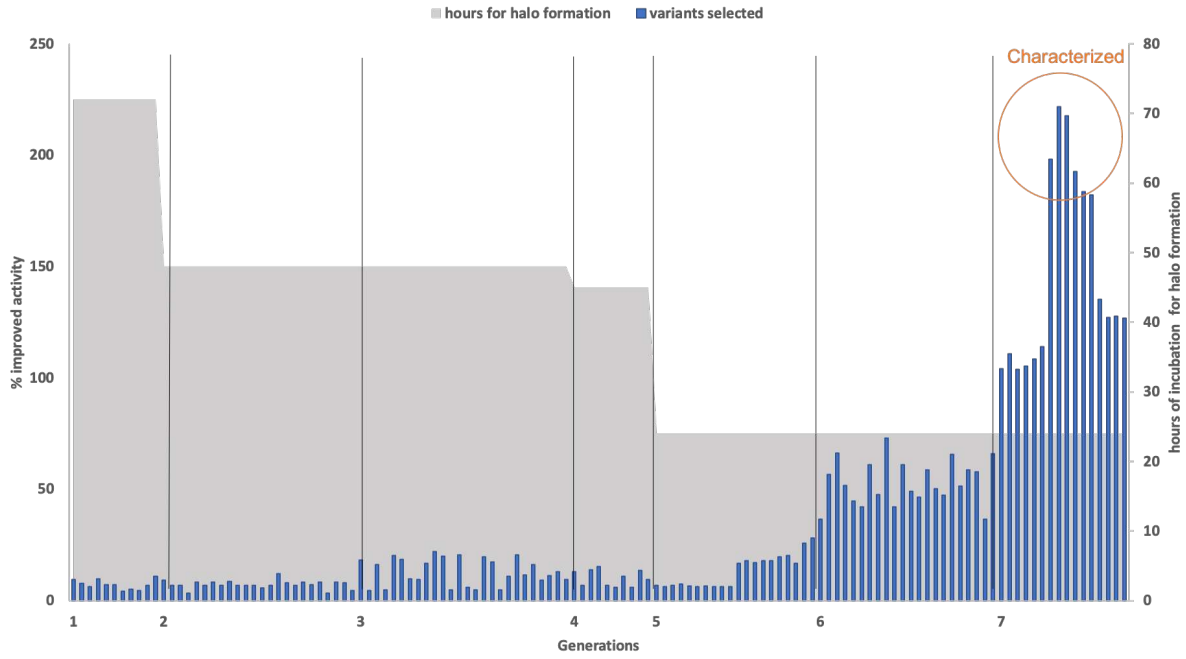

Fig. 3 Screening activity of crude lysate during the directed evolution. Grey shadow shows the reduction of incubation time to see the haloes. Blue bars correspond to the increment of activity of the crude lysate in each round to hydrolyse pNP-C8. Spacing of the generations on the horizontal axis is indicative of the number of variants selected in the secondary screen. The orange circle shows the variants selected for final characterization.

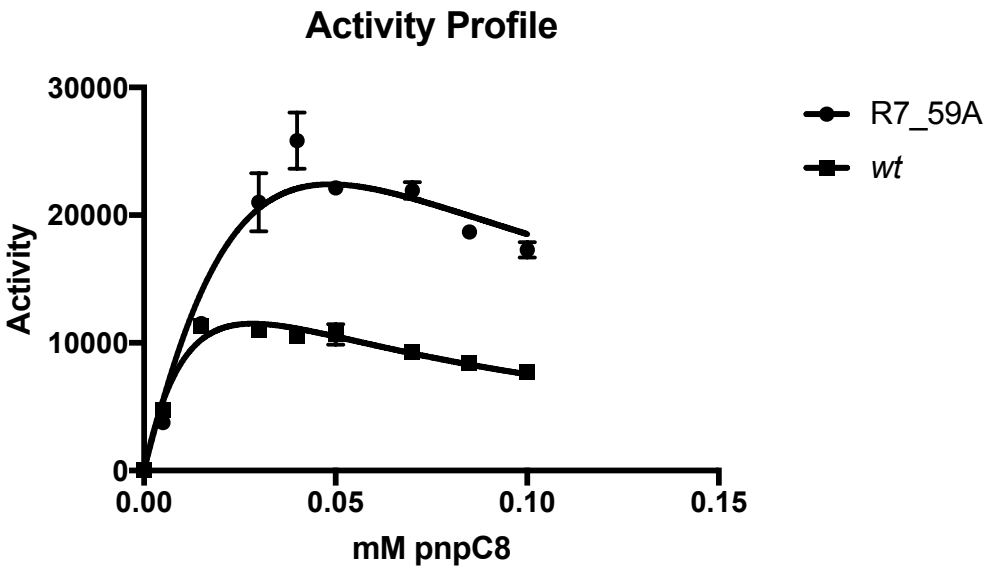

Fig. 4 Activity profile using pNP-C8 showing the activity for *wt* (squares) and R7_59A (circles). *wt* lip3 was measured using 60 nM of enzyme, while R7_59A was measured with 6 nM of enzyme**.** Curve fitting was done in GraphPad Prism v7, yielding the following parameters: k_{cat}: $2.7*10^4$ and $10*10^4$ (min^{-1}); K_M = 0.0051 and 0.075 (mM); K_i = 0.010 and 0.026 (mM)for *wt* and variant R7_59A respectively.

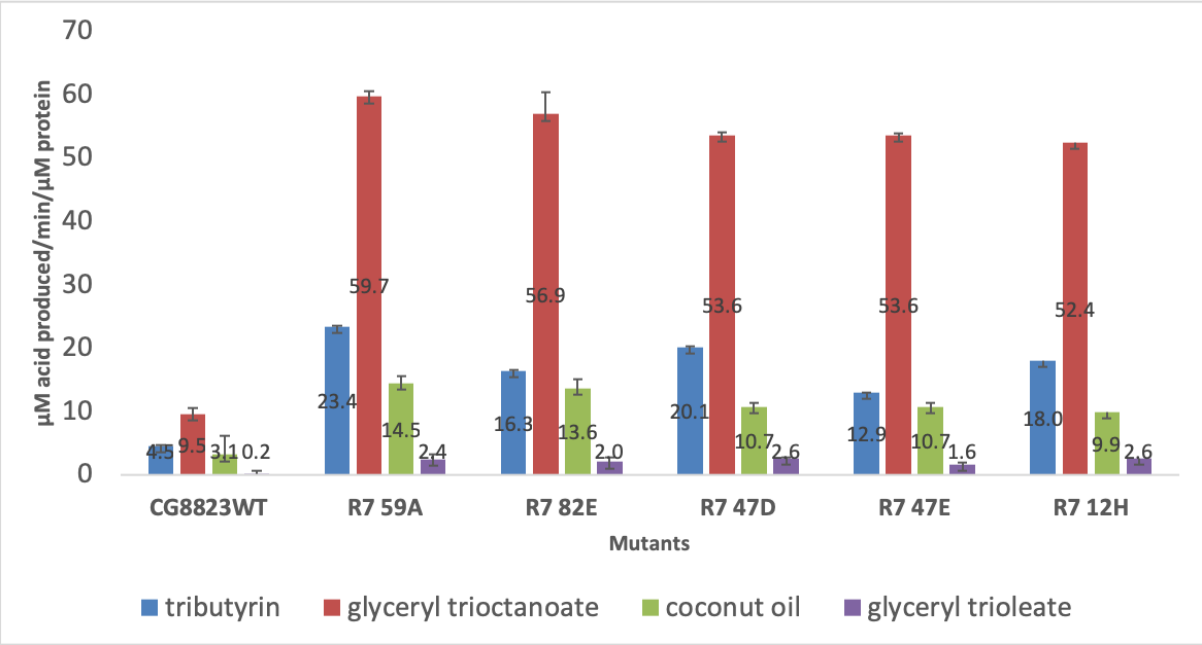

Fig. 5 Lipolytic activity with phenol red using different substrates

165x99 mm (300X300 DPI)

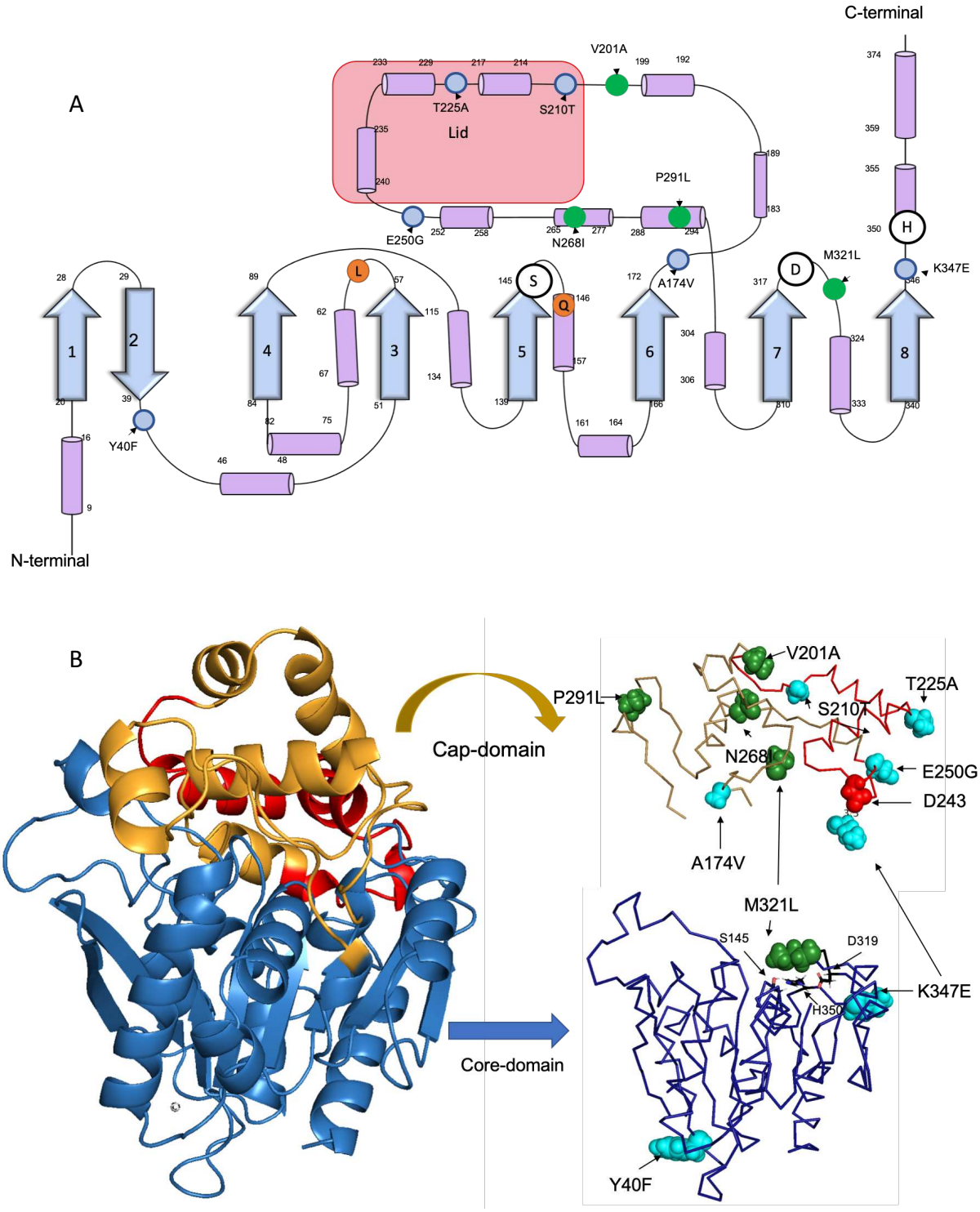

Fig. 6 (A) Schematic diagram of lipase lip3. Arrows represent β sheets, barrels α-helices. Catalytic triad S145, D318 and H350 are in black and residues of the oxyanion hole in orange. The most common mutations are in green circles while the other mutations are in blue circles. (B) Model of lip3. Left shows the complete molecule. The core-domain in blue, the cap domain in orange and the lid in red. Right shows the positions of the mutations. Common mutations in green spheres and the rest of mutation in cyan. Red spheres show the the residue interacting with K347

Supporting information

Figure S1. Comparison of the expression of the CG8823 lipase from the *wt* gene at different temperatures and increasing aeration using different volumes of culture in the same vessel a – 1L and b – 250ml.

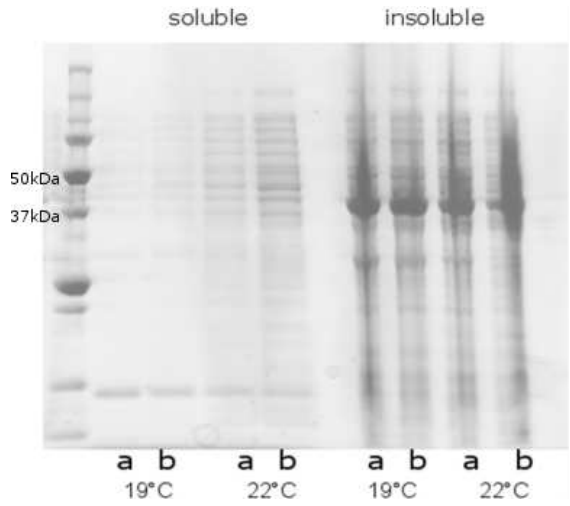

Figure S2. Acetic acid with phenol red standard curve

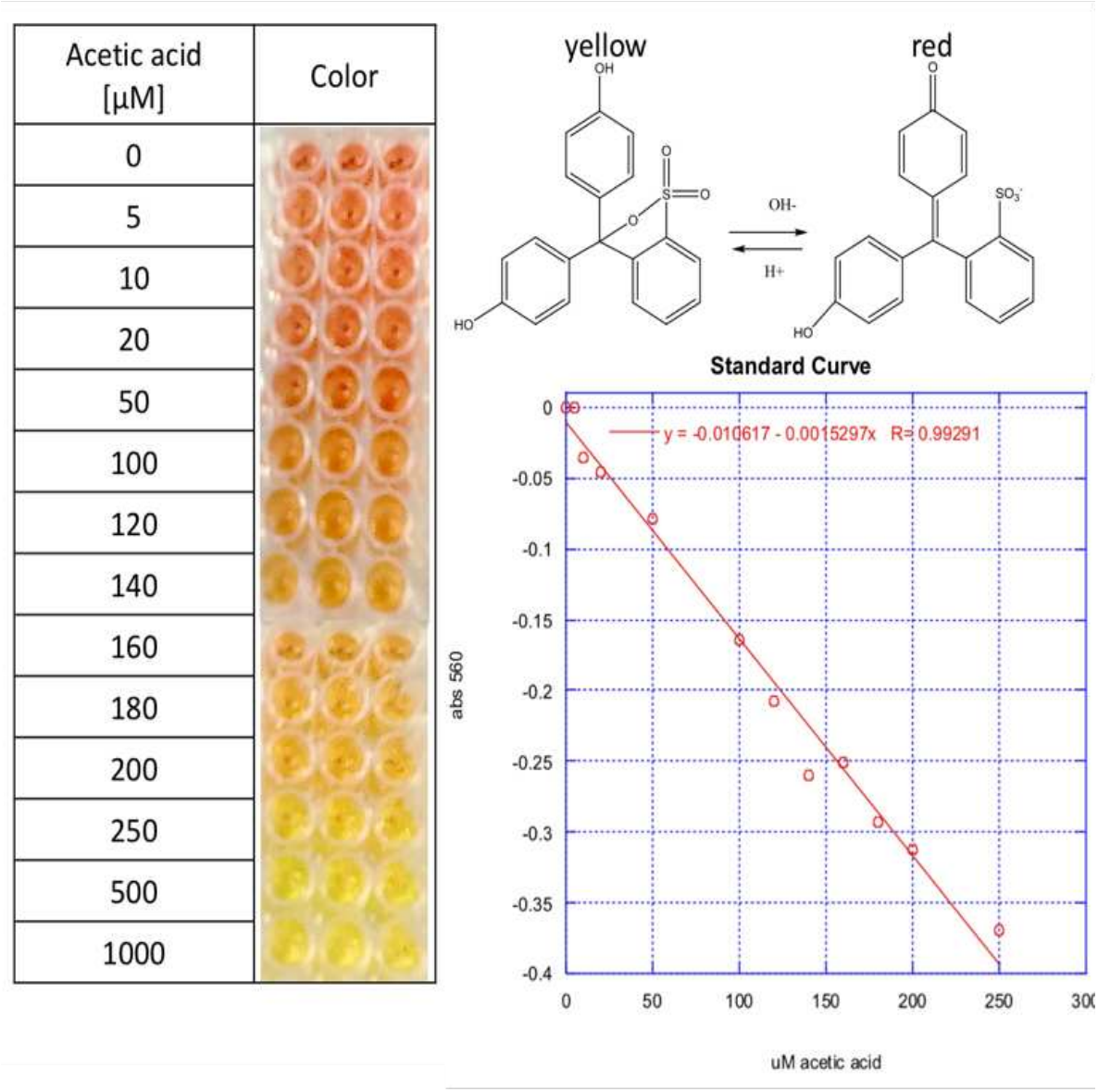

Table S1. Amino acid substitutions across seven rounds of directed evolution of lipase Lip3

	#	res →	28	41	49	56	61	78	174	201	245	248	291	# MUT
		wt	**S**	**S**	**N**	**M**	**S**	**M**	**A**	**V**	**E**	**D**	**P**	
ROUND 1	1	R1 12H								A				1
	2	R1 14H	R								V		L	3
	3	R1 15B											L	1
	4	R1 15C		G		V	G							3
	5	R1 17H						T						1
	6	R1 36C			K								Q	2
	8	R1 49A	R								V		L	3
	9	R1 66A										G		1
	10	R1 39E							V					1

	#	res →	3	28	35	40	46	78	87	101	174	185	201	211	223	242	268	276	291	361	362	363	# MUT
		wt	**S**	**S**	**M**	**Y**	**E**	**M**	**W**	**K**	**A**	**T**	**V**	**N**	**Q**	**Y**	**N**	**N**	**P**	**K**	**Y**	**V**	
ROUND 2	1	R2 28G		R								A											2
	2	R2 210E				F																	1
	3	R2 310C		T				T									I						3
	4	R2 312H															T						1
	5	R2 72F								R	V												2
	6	R2 712C									V	A											2
	7	R2 12A					K										I		L				3
	8	R2 311D	R					T									I						3
	9	R2 35B		R													Y	S	L	T		F	6
	10	R2 612F													L		I				F		3
	11	R2 510C			T				G							H	I		L				5
	12	R2 112D			I								A	D			I		Q			F	6

	#	res →	13	21	28	40	43	46	74	78	101	136	158	163	174	185	189	201	206	208	221	223	225	230	245	253
	#	wt	I	E	S	Y	K	E	S	M	K	Q	R	D	A	T	A	V	E	M	M	Q	T	D	E	E
ROUND 3	1	R3 28E								T												R				
	2	R3 212B								T																
	3	R3 310D									R								G							
	4	R3 311G									R				V			A					A			
	5	R3 411G				F																				
	6	R3 412D	V		R																					
	7	R3 511G																								
	8	R3 512C							I																	
	9	R3 612A	V		R								H	G												
	10	R3 310C	V																					G		
	11	R3 319D																						G		
	12	R3 19H					E										V				K					
	13	R3 111D			R			K												V						
	14	R3 612E	V		R																					
	15	R3 66C						K				R														
	16	R3 310B		G				K																		
	17	R3 59E														A										
	18	R3 111C																								G
	19	R3 314A						K																	V	
	20	R3 28D	V		R																					

	#	res →	268	274	291	293	303	321	333	337	363	364	373	# MUT
	#	wt	N	E	P	E	K	M	E	M	V	Y	S	
ROUND 3	1	R3 28E	T		L									4
	2	R3 212B	T		L									3
	3	R3 310D	I		L					T				5
	4	R3 311G	T		L									6
	5	R3 411G		G	L									3
	6	R3 412D	I			V								4
	7	R3 511G	I		L									2
	8	R3 512C			L									2
	9	R3 612A	I		L									6
	10	R3 310C	I			V								4
	11	R3 319D	I			V								3
	12	R3 19H	I		L		R							6
	13	R3 111D	I		L									5
	14	R3 612E	T		L									4
	15	R3 66C	I		L				V			F		6
	16	R3 310B			L									3
	17	R3 59E			L			L						3
	18	R3 111C	I		L						F			4
	19	R3 314A	I		L								S	4
	20	R3 28D	I			V								4

	#	res →	40	49	101	174	201	230	268	291	321	337	347	360	# MUT
	#	wt	Y	N	K	A	V	D	N	P	M	M	K	R	
ROUND 4	1	R4 312A			R	V						T			3
	2	R4 412E			R	V	A								3
	3	R4 19A	F								L				2
	4	R4 15G		K				G	I	L	L		E	C	7
	5	R4 28E								L					1

	#	res →	20	24	26	40	46	47	49	61	101	174	175	201	210	225	230	268	269	277	291	301	321	329	337	347
	#	wt	M	E	V	Y	E	S	N	S	K	A	A	V	S	T	D	N	L	S	P	D	M	K	M	K
ROUND 5	1	R5 13A	V				K																			
	2	R5 110A										V		A		A		I			L	N		R		
	3	R5 18F																			L		L			
	4	R5 12A												A	T			T			L		L			
	5	R5 16D				F						V		A	T			T			L		L			
	6	R5 19D																I			L		L			E
	7	R5 12D	V				K																			
	8	R5 19E							K									I			L		L			
	9	R5 19G	V				K			G								I		C	L					
	10	R5 510F												A	T			T			L		L			
	11	R5 37H																			L				T	
	12	R5 32C				F											G	I			L		L			
	13	R5 39A																			L					
	14	R5 312G						G			R	V	P					I	Q		L					
	15	R5 212C																			L		L			E
	16	R5 38B		V														I			L					
	17	R5 512G			D							V		A		A		I	Q		L		L			
	18	R5 34C												A	T			T			L		L			
	19	R5 312B																T			L		L			

	#	res →	360	369	# MUT
		wt	**R**	**K**	
ROUND 5	1	R5 13A			2
	2	R5 110A			7
	3	R5 18F			2
	4	R5 12A			5
	5	R5 16D			7
	6	R5 19D			4
	7	R5 12D			2
	8	R5 19E			4
	9	R5 19G			6
	10	R5 510F			5
	11	R5 37H			2
	12	R5 32C		E	6
	13	R5 39A			1
	14	R5 312G	C		8
	15	R5 212C	C		4
	16	R5 38B			3
	17	R5 512G			8
	18	R5 34C			5
	19	R5 312B			3

	#	res →	20	26	40	44	46	47	59	101	120	142	158	174	188	201	210	225	230	245	250	268	269	269	291	321
		wt	**M**	**V**	**Y**	**T**	**E**	**S**	**M**	**K**	**M**	**V**	**R**	**A**	**F**	**V**	**S**	**T**	**D**	**E**	**E**	**N**	**L**	**L**	**P**	**M**
ROUND 6	1	R6 15H											H	V		A	T					T			L	L
	2	R6 14H														A					G	I			L	L
	3	R6 12D	V												L	A		A				I	Q	Q	L	
	4	R6 19A			F									V		A	T					T			L	L
	5	R6 14F			F				L	R		I		V		A				G	G	I			L	L
	6	R6 15A	V				K									A					G	I			L	L
	7	R6 16B								R						A		A				I			L	
	8	R6 27D														A					G	I			L	L
	9	R6 29D		D						R				V		A		A				I	Q	Q	L	L
	10	R6 212G			F									V		A	T								L	L
	11	R6 311H						G			T			V		A	T					I			L	L
	12	R6 312A			F									V		A	T					I			L	L
	13	R6 15G														A					G	I			L	L
	14	R6 28A												V		A	T		G			I			L	L
	15	R6 111D				P								V		A	T	A				I			L	
	16	R6 112H			F											A					G	I			L	L

	#	res →	337	347	# MUT
		wt	M	K	
ROUND 6	1	R6 15H			7
	2	R6 14H		E	6
	3	R6 12D			7
	4	R6 19A			7
	5	R6 14F			11
	6	R6 15A			7
	7	R6 16B	T		6
	8	R6 27D		E	6
	9	R6 29D			9
	10	R6 212G			6
	11	R6 311H			8
	12	R6 312A			7
	13	R6 15G		E	6
	14	R6 28A			7
	15	R6 111D			7
	16	R6 112H		E	7

ROUND 7

	residue →	2	20	40	44	46	47	59	98	101	105	116	120	129	133	159	174	185	188	201	205	208	210
#	wt	G	M	Y	T	E	S	M	K	K	T	N	M	Y	K	P	A	T	F	V	M	M	S
1	R7 58D																			A			
2	R7 412F			F											E					A			
3	R7 12H		V																L	A			
4	R7 69H																			A		L	
5	R7 58F																			A			
6	R7 48A		V			G														A			
7	R7 69G																			A			
8	R7 611E								E							T				A			
9	R7 410H																			A			
10	R7 59A			F				L												A			
11	R7 16D																			A			
12	R7 112D																			A			T
13	R7 712E			F																A			
14	R7 73H																			A			
15	R7 67B			F										H						A	T		
16	R7 612G																			A			T
17	R7 64C																			A			
18	R7 412D							L									V			A			T
19	R7 612B					K												A		A			T
20	R7 48B			F																A			
21	R7 107G																			A			
22	R7 47E																			A			T
23	R7 82E	R								R									L	A			
24	R7 106H						G						T				V			A			T
25	R7 410F																			A			T
26	R7 106G																			A			T
27	R7 411A			F	A			L												A			
28	R7 47D		V																	A			
29	R7 67A			F								S								A			
30	R7 95D										A						V			A			T

		residue →	211	212	225	230	250	268	269	281	291	305	321	345	347	351	362	# MUT
	#	wt	N	K	T	D	E	N	L	R	P	K	M	F	K	L	Y	
ROUND 7	1	R7 58D						T			L		L					4
	2	R7 412F					G	I			L		L		E			8
	3	R7 12H			A			I	Q		L							7
	4	R7 69H					G	I			L		L					6
	5	R7 58F					G	I			L		L		E			6
	6	R7 48A					G	I		H	L		L	L				9
	7	R7 69G					G	I			L		L		E			6
	8	R7 611E			A			I			L		L					7
	9	R7 410H					G	I			L		L		E			6
	10	R7 59A						I		C	L	R	L					8
	11	R7 16D									L		L					3
	12	R7 112D						T			L		L					5
	13	R7 712E					G	I			L		L		E			7
	14	R7 73H					G	I			L		L		E			6
	15	R7 67B					G	I			L		L					8
	16	R7 612G						T			L		L					5
	17	R7 64C					G	I			L		L		E			6
	18	R7 412D	S		A						L		L					8
	19	R7 612B				G		I			L		L			Q		9
	20	R7 48B					G	I			L		L		E			7
	21	R7 107G					G	I			L		L		E			6
	22	R7 47E						T			L		L					5
	23	R7 82E			A		G	I			L		L					8
	24	R7 106H		R	A			I			L		L					10
	25	R7 410F			A			T			L		L					6
	26	R7 106G						T			L		L					5
	27	R7 411A			A			I			L		L					8
	28	R7 47D						I			L		L					5
	29	R7 67A					G	I			L							6
	30	R7 95D				G		I			L		L				F	9