nature biotechnology

A simple selection strategy for evolving highly efficient enzymes

Martin Neuenschwander, Maren Butz, Caroline Heintz, Peter Kast & Donald Hilvert

Combining tunable transcription with an enzyme-degradation tag affords an effective means to reduce intracellular enzyme concentrations from high to very low levels. Such fine-tuned control allows selection pressure to be systematically increased in directed-evolution experiments. This facilitates identification of mutants with wild-type activity, as shown here for an engineered chorismate mutase. Numerous selection formats and cell-based screening methodologies may benefit from the large dynamic range afforded by this easily implemented strategy.

Genetic selection can greatly facilitate the search for rare catalysts in very large protein libraries[1]. Auxotrophic strains, which grow only if provided with a protein that functionally replaces a missing cellular enzyme, are frequently used for this purpose. However, growth represents an indirect (and imperfect) readout of catalyst activity. Even mediocre catalysts may provide sufficient activity for cells to grow at wild-type levels, making it difficult to distinguish the most active variants from their less effective counterparts. As a result, it can be difficult or impossible to optimize relatively inefficient enzymes through multiple rounds of mutagenesis and selection.

We faced this problem when we tried to improve an engineered chorismate mutase. The dimeric helical bundle chorismate mutase from *Escherichia coli* (EcCM) was successfully converted into a functional hexamer (hEcCM) by inserting a five-amino-acid hinge loop into the middle of the long H1 helix spanning the parent dimer[2], but the topological change was accompanied by a 2- to 3-order-of-magnitude decrease in activity (**Table 1**). Activity was partially recovered when hEcCM was subjected to two rounds of random mutagenesis and selection in a chorismate mutase–deficient *E. coli* strain (KA12/pKIMP-UAUC[3]; **Fig. 1a**)[4]. Nevertheless, the best variant, tEcCM, which contained three mutations and possessed a trimeric quaternary structure, still had a 14-fold lower k_{cat} value than the parent EcCM dimer (**Table 1**), and further improvements were not possible because tEcCM already conferred wild-type levels of growth to its host[4].

In theory, selection pressure in a complementation assay can be increased by decreasing the intracellular catalyst concentration. This might be accomplished by switching to a weakly active promoter, low gene dose or inefficient ribosomal binding sites for catalyst production. Such strategies have been profitably exploited for the directed evolution of aspartate transaminases[5], for example. As reclonning of a library is laborious and prone to loss of diversity, and the effect on selection pressure only qualitatively predictable, inducible systems that provide a tightly controlled and graded transcriptional response to an external inducer represent potentially attractive alternatives.

Regulable promoters that use arabinose[6] or tetracycline[7,8] as inducer compounds have been well characterized. Both allow homogeneous gene expression over a broad dynamic range. To construct selection plasmids that combine inducer dose-dependent gene expression with the convenience of high-copy plasmids, we opted for the tetracycline-inducible P_{tet} system[7], which does not require specifically engineered host strains. We replaced the weak, constitutive *bla* promoter on our selection plasmid pKECMB[2] (**Fig. 1a**) with a modified P_{tet} promoter cassette (**Fig. 1b**), which includes a downstream T7 promoter, to simplify protein overproduction of selected variants. As expected, auxotrophic KA12/pKIMP-UAUC cells harboring the gene for the very weakly active hEcCM under the control of this promoter system (on plasmid pKT) did not grow under selective conditions in the absence of inducer, but they regained prototrophy at high tetracycline concentrations (**Table 2**). In contrast, cells containing the more active tEcCM and EcCM variants grew even in the absence of inducer. Thus, background transcription affords sufficient amounts of these more active catalysts to fully satisfy the metabolic needs of the cell. Clearly, in this case, tight transcriptional control alone is not sufficient to reduce protein concentrations to a level low enough to allow discrimination between a moderately active catalyst (tEcCM) and the parent enzyme (EcCM).

To reduce protein concentration further, we fused an 11-amino-acid SsrA degradation signal[9] to the C terminus of the catalyst (**Fig. 1c**). The SsrA tag targets the catalyst for rapid degradation by the intracellular ClpXP protease. The efficacy of this strategy was established using green fluorescent protein as a reporter (**Supplementary Fig. 1** and **Supplementary Methods** online). The presence of the degradation tag in plasmid pKTS also increased the dynamic range of the chorismate mutase selection system, as demonstrated by

Table 1 Catalytic parameters of the evolved hinge-loop variants

Variant[a]	k_{cat} (s^{-1})	K_m (μM)	k_{cat}/K_m ($M^{-1}s^{-1}$)
EcCM[b]	14	350	41,000
hEcCM[b]	0.15	2,100	70
tEcCM[b]	1.0	34	30,000
EcCM-200/4[c]	12 ± 1	270 ± 20	45,000

[a]The sequences of the variants are provided in **Supplementary Figure 4** online. [b]Ref. 4, pH 6.5. The catalytic parameters of EcCM are similar at neutral and acidic pH. [c]This work. Protein production and characterization is described in **Supplementary Methods**. Kinetic measurements were performed at 20 °C in PBS (10 mM Na_2HPO_4, 2 mM KH_2PO_4, 137 mM NaCl, 2.7 mM KCl, pH 7.4) supplemented with 0.1 mg/ml bovine serum albumin.

Laboratory of Organic Chemistry, ETH Zurich, Hönggerberg HCI F 339, CH-8093 Zurich, Switzerland. Correspondence should be addressed to D.H. (hilvert@org.chem.ethz.ch).

Received 21 March; accepted 24 August; published online 16 September 2007; doi:10.1038/nbt1341

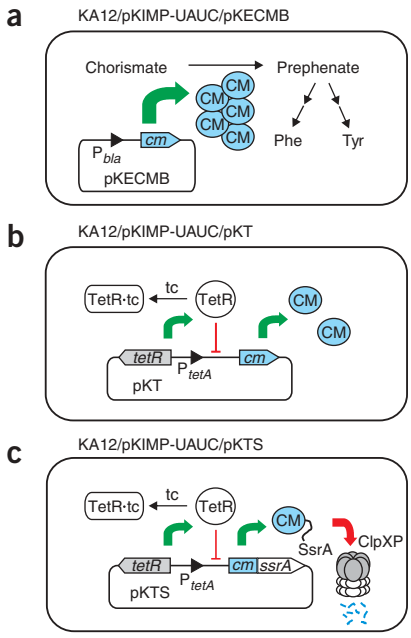

Figure 1 Selection plasmid design. (**a**) The catalyst gene (*cm*) on plasmid pKECMB[2] is constitutively expressed under control of the *bla* promoter. *E. coli* strain KA12 is deficient in chorismate mutase activity and also requires plasmid pKIMP-UAUC, which encodes two prephenate-processing enzymes, for growth on minimal medium[3]. (**b**) Selection plasmid pKT provides graded and homogeneous transcriptional control of catalyst production from promoter P_{tetA}. The *tetR* gene and its promoter region are located upstream of *cm*, so that the TetR repressor simultaneously regulates transcription of catalyst and TetR repressor genes[13]. The tetracycline-resistance determinant of the Tn*10* transposon (*tetA*) is integrated in the KA12 chromosome[3]. (**c**) Selection plasmid pKTS permits graded transcriptional control and limited enzyme half-life. The *ssrA* sequence is incorporated as a downstream genetic fusion to the catalyst gene. The resulting enzyme carries the degradation tag at its C terminus, and is directed to the ClpXP protease, where it is degraded. Plasmid construction details are provided in **Supplementary Methods** online.

complete suppression of cell growth in the absence of inducer for all tested chorismate mutase variants (**Table 2**). Restoration of growth upon addition of specific tetracycline concentrations on solid media roughly correlated with specific activity, distinguishing the most weakly active variant, hEcCM, from the more active tEcCM and EcCM variants. The correlation between growth rate, tetracycline concentration and specific activity was even more apparent in liquid culture (**Fig. 2a**). At the highest tetracycline concentrations tested, wild-type growth rates were achieved with the tEcCM and EcCM variants, but not with the weakly active hEcCM. Moreover, at intermediate tetracycline concentrations, cells harboring tEcCM grew more slowly than those with wild-type EcCM. The ability to control selection stringency simply by adjusting tetracycline concentration raises the possibility of evolving topologically novel catalysts that are more active than tEcCM.

To test the utility of this system in a directed evolution experiment, we inserted library fragments that encode the first 93 residues of hEcCM, diversified by error-prone PCR and DNA shuffling[4], into the pKTS acceptor vector in place of a stuffer fragment, in-frame with the last seven residues of hEcCM fused to the SsrA tag. After transformation of the KA12/pKIMP-UAUC selection strain (1.5×10^7 transformants), library clones were picked randomly and sequenced to check library quality. With the exception of the R44C substitution, which occurred in 40% of the sequences because of an apparent DNA shuffling artifact, mutations were evenly distributed over the entire hEcCM gene. Aliquots of the library were plated onto selective M9c plates containing varying amounts of tetracycline to identify active catalysts based on their ability to complement the chorismate mutase deficiency[3]. The number of complementing clones decreased with decreasing tetracycline concentration (**Fig. 2b**), consistent with the hypothesis that reducing intracellular enzyme concentration increases selection pressure.

Sequence analysis of 96 clones revealed a mutational bias in active variants that correlates roughly with selection stringency. The emergence and disappearance of specific mutations upon increasing selection pressure is illustrated in **Figure 2c** (see also **Supplementary Fig. 2** online). At a complementation frequency of 14 ± 5% (400 ng/ml tetracycline), an H66R mutation occurred frequently and was further enriched at a complementation frequency of 2 ± 1% (200 ng/ml tetracycline). An S42L mutation was also enriched under the latter conditions, whereas the R44C mutation, which was prevalent before selection, occurred less frequently, presumably because it provides no catalytic benefit. As selection pressure further increased (≤100 ng/ml tetracycline), the complementation frequency dropped to <0.1% and only false positives were observed. The latter had lost the degradation tag mainly through frameshift mutations, and were therefore presumably produced at elevated concentrations. The fact that the most stringent conditions only yielded false positives illustrates the importance of fine-tuning the selection pressure in these experiments to maximize the yield of highly active variants.

The genes of six clones selected at a complementation frequency of 2 ± 1% were retransformed, and the transformants grew faster than six out of seven variants selected at a complementation frequency of 14 ± 5% (**Supplementary Fig. 3** online). For six fast-growing variants, the degradation tag was replaced with a $(His)_6$ tag for protein purification, and the enzymatic activity of the four variants that could be produced in soluble form was determined *in vitro*. They showed uniformly high V_{max} values comparable to EcCM (**Supplementary Table 1** online). The highest catalytic efficiency was exhibited by variant EcCM-200/4, which contained four mutations relative to hEcCM (A9V, S42L, H66R, T87I; see also **Supplementary Fig. 4** online). It eluted from a gel filtration column as a trimer and catalyzed the rearrangement of chorismate to prephenate with a k_{cat} of 12 s^{-1} and a k_{cat}/K_m of 45,000 $M^{-1}s^{-1}$ (**Table 1**). The turnover number, which represents a 75-fold improvement over hEcCM and a tenfold improvement relative to the best previously characterized variant tEcCM, is similar to that of wild-type EcCM (**Table 1**). This result is notable as high k_{cat} values are

Table 2 Benchmark complementation assays with chorismate mutases having different activities

	pKT derivative encoding			pKTS derivative encoding		
[tc] (ng/ml)	EcCM	tEcCM	hEcCM	EcCM	tEcCM	hEcCM
0	+	+	0	0	0	0
10	+	+	0	0	0	0
50	+	+	+	+	+	0
300	+	+	+	+	+	+

+, cell growth; 0, no cell growth. Streak-outs of KA12/pKIMP-UAUC cells containing the indicated selection plasmids were evaluated after 2 d of incubation at 30 °C on M9c medium plates[14]. [tc] is the tetracycline concentration in M9c medium. pKT places the gene under control of the tetracycline-inducible P_{tet} system, pKTS additionally encodes a degradation tag fused to the catalyst.

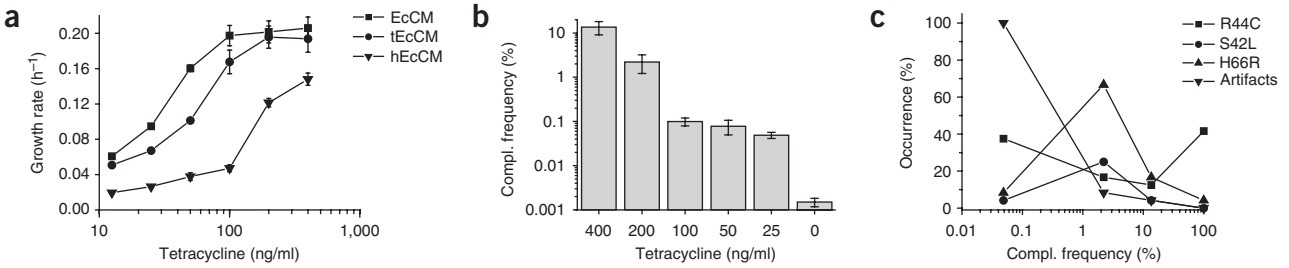

Figure 2 Tetracycline-dependent growth in selective M9c medium and influence of tetracycline concentration on the selection process. (**a**) KA12/pKIMP-UAUC cells were transformed with the pKTS selection plasmid encoding wild-type EcCM, tEcCM or hEcCM. Growth curves were determined for each transformant. Error bars indicate the s.d. of the curve fit in each growth experiment. (**b**) Complementation frequency among gene library members on selective M9c plates as a function of tetracycline concentration. (**c**) Mutation bias as a function of complementation frequency, determined using 24 sequences originating from three independent selection experiments for each selection regime. The occurrence of false positives lacking the degradation tag and the relatively frequent mutations H66R, S42L and R44C is plotted for different selection regimes. Clones grown on nonselective rich medium plates (100% complementation) were examined to assess library size, quality and sequence diversity before selection. See **Supplementary Figure 2** online for alignments of all 96 sequences used for this analysis and the **Supplementary Methods** for detailed experimental protocols for the liquid growth tests, library construction and selection experiments.

important for industrial biocatalysis, where high conversion of substrate to product is desired[10].

In conclusion, pairing a tunable promoter with a degradation tag can provide very low but adjustable catalyst concentrations within cells. By providing systematic control over selection stringency, this strategy facilitates the evolution of substantially more active enzymes than is possible with systems reliant on weak constitutive gene expression[4]. Extension of this approach for regulating stringency should benefit any cell-based selection[1,11] or screening[11,12] system in which the setting of a threshold for minimum activity allows the best variants to be distinguished from less interesting ones.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

This work was generously supported by the Schweizerischer Nationalfonds and the ETH Zurich.

AUTHOR CONTRIBUTIONS

M.N., P.K. and D.H. designed research; M.N., M.B. and C.H. performed the experiments; M.N., M.B., P.K. and D.H. analyzed data; M.N., P.K. and D.H. wrote the paper.

Published online at http://www.nature.com/naturebiotechnology
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Taylor, S.V., Kast, P. & Hilvert, D. *Angew. Chem. Int. Ed. Engl.* **40**, 3310–3335 (2001).
2. MacBeath, G., Kast, P. & Hilvert, D. *Protein Sci.* **7**, 325–335 (1998).
3. Kast, P., Asif-Ullah, M., Jiang, N. & Hilvert, D. *Proc. Natl. Acad. Sci. USA* **93**, 5043–5048 (1996).
4. Vamvaca, K., Butz, M., Walter, K.U., Taylor, S.V. & Hilvert, D. *Protein Sci.* **14**, 2103–2114 (2005).
5. Yano, T., Oue, S. & Kagamiyama, H. *Proc. Natl. Acad. Sci. USA* **95**, 5511–5515 (1998).
6. Khlebnikov, A., Risa, O., Skaug, T., Carrier, T.A. & Keasling, J.D. *J. Bacteriol.* **182**, 7029–7034 (2000).
7. Hansen, L.H., Ferrari, B., Sorensen, A.H., Veal, D. & Sorensen, S.J. *Appl. Environ. Microbiol.* **67**, 239–244 (2001).
8. Lutz, R. & Bujard, H. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
9. Karzai, A.W., Roche, E.D. & Sauer, R.T. *Nat. Struct. Biol.* **7**, 449–455 (2000).
10. Burton, S.G., Cowan, D.A. & Woodley, J.M. *Nat. Biotechnol.* **20**, 37–45 (2002).
11. Turner, N.J. in *Enzyme Assays* (ed. J.L. Reymond) 139–161, (Wiley-VCH, Weinheim, 2006).
12. Jestin, J.L. & Kaminski, P.A. *J. Biotechnol.* **113**, 85–103 (2004).
13. Hillen, W. & Berens, C. *Annu. Rev. Microbiol.* **48**, 345–369 (1994).
14. Gamper, M., Hilvert, D. & Kast, P. *Biochemistry* **39**, 14087–14094 (2000).