BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

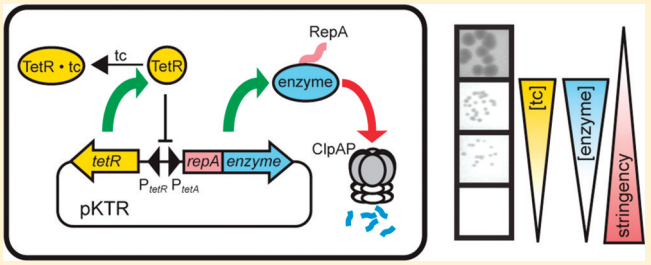
pubs.acs.org/biochemistry

An N-Terminal Protein Degradation Tag Enables Robust Selection of Highly Active Enzymes

Maren Butz, Martin Neuenschwander, Peter Kast, and Donald Hilvert*

Laboratory of Organic Chemistry, ETH Zurich, CH-8093 Zurich, Switzerland

ABSTRACT: Degradation tags are short peptide sequences that target proteins for destruction by housekeeping proteases. We previously utilized the C-terminal SsrA tag in directed evolution experiments to decrease the intracellular lifetime of a growth-limiting enzyme and thereby facilitate selection of highly active variants. In this study, we examine the N-terminal RepA tag as an alternative degradation signal for laboratory evolution. Although RepA proved to be less effective than SsrA at lowering protein concentrations in the cell, its N-terminal location dramatically reduced the occurrence of truncation and frameshift artifacts in selection experiments. We exploited this improvement to evolve a topologically redesigned chorismate mutase that is intrinsically disordered but already highly active for the conversion of chorismate to prephenate. After three rounds of mutagenesis and high-stringency selection, a robust and more nativelike variant was obtained that exhibited a catalytic efficiency (k_{cat}/K_M = 84000 M^{-1} s^{-1}) comparable to that of a natural dimeric chorismate mutase. Because of concomitant increases in catalyst yield, the level of intracellular prephenate production increased approximately 30-fold overall over the course of evolution.



Directed evolution is a powerful tool for exploring catalytic mechanisms and structure−function relationships in enzymes. It involves iterative rounds of mutagenesis, selection, and amplification in the laboratory. These three steps drive the improvement of enzymes, mimicking natural evolution. Survival selection systems are particularly beneficial for directed evolution because they couple the selection and amplification steps, allowing larger libraries of mutants to be analyzed than is possible with screening systems.[1,2]

Selection systems often take advantage of auxotrophic strains that are not able to grow unless they are supplied with a gene for an enzyme that can provide the missing metabolic activity. Enzymatic production of the essential metabolite then allows cellular growth, which is a multifactorial property that depends both on specific activity and on the concentration of the introduced enzyme. In an ideal case, increases in catalytic efficiency lead to proportional increases in growth rate (Figure 1). However, the metabolic needs of the cell could also be fully met by a poor enzyme if the catalyst is strongly overproduced. In the case of highly active enzymes, even tiny amounts of catalyst can supply a sufficiently high metabolite concentration to allow wild-type growth levels. In both cases, further evolution is difficult or impossible because cell growth in the selection system is already maximal, preventing the discrimination of moderately and highly active enzymes. Alternative strategies are needed here to increase the stringency of selection.

Increased stringency of an in vivo selection system can be achieved by various methods. One way to increase selection stringency is to reduce the amount of catalyst produced in the cell, because the turnover rate correlates with enzyme concentration.[3] Intracellular enzyme concentration may be controlled in several ways. These include (i) regulating promoter strength,[4−7] (ii) reducing the gene copy number,[8] (iii) using modified ribosome binding sites,[4,9] and (iv) using a degradation tag that directs the tagged protein to cellular proteases.[3]

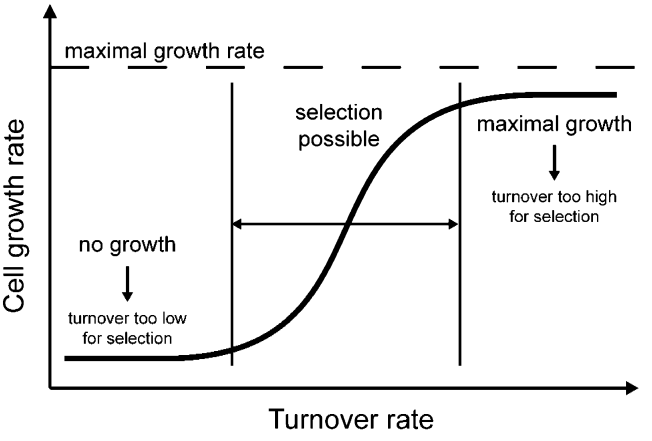


Figure 1. Dependence of growth rate on the turnover of the growth-limiting metabolic reaction catalyzed by the enzyme of interest. Discrimination between cells encoding enzymes with different activities is only possible in a selection window in which enzymatic turnover is not saturating for selection but sufficiently high to support weak cell growth.

Received: July 22, 2011
Revised: August 23, 2011
Published: September 14, 2011

Combining tunable transcription with an enzyme degradation tag has been shown to be an effective means of adjusting enzyme concentration over a very broad range. For example, this approach was successfully used to evolve highly active variants of a weakly active hexameric chorismate mutase (CM).[3] Protein production was regulated by a tetracycline promoter,[10] while a C-terminal SsrA tag, which directs the enzyme to the housekeeping ClpAP and ClpXP proteases,[11] was used to control the lifetime of the protein. This combination provided fine-tuned control over the intracellular CM concentration.[3] Selection stringency in this system could be steadily increased by decreasing inducer concentration, allowing variants with near wild-type (wt) activity to emerge.

The SsrA tag is an 11-amino acid sequence (AANDENYALAA) that can be appended to virtually any protein. Nevertheless, this strategy does have one major drawback. The C-terminal location of the tag allows a simple escape mechanism. Stop codon or frameshift mutations that lead to deletion of the tag protect the enzyme from degradation, allowing it to accumulate within the cell. The resulting higher catalyst concentrations facilitate complementation by mutants with lower intrinsic activities. In the evolution of the hexameric CM, such false positives severely outnumbered the truly improved variants under the most stringent selection conditions.[3] A potentially simple solution to this problem would be to move the degradation tag from the C-terminus to the N-terminus. One candidate for this purpose is a 16-amino acid segment from replication protein A (MNQSFISDILYADIES) that reportedly targets proteins to the ClpAP protease complex.[12,13] For example, when this sequence was attached to the N-terminus of the otherwise stable GFP, 50% of the protein was degraded within 30 min.

In this study, we compare the ability of the N-terminal RepA and the C-terminal SsrA tags to regulate the intracellular concentration of a growth-limiting enzyme and thus control selection stringency in a directed evolution experiment. We also assess the practical utility of the RepA tag by using it to evolve a highly active, monomeric CM (mMjCM), derived from the dimeric *Methanocaldococcus jannaschii* chorismate mutase (MjCM).[14] Although this topologically redesigned enzyme has near wild-type activity, it exhibits all the properties of a molten globule, including ANS binding, noncooperative unfolding, fast H−D exchange, and weak NMR peak dispersion.[15,16] Its further optimization demanded an extremely stringent selection system.

MATERIALS AND METHODS

Bacterial Strains. XL1-Blue (Stratagene) was used for general cloning and amplification purposes.[17] The CM-deficient *Escherichia coli* strain KA12/pKIMP-UAUC was the selection strain for the genetic complementation experiments.[18] The genotype of KA12 is Δ*(srlR-recA)306*::Tn*10*, Δ*(pheA-tyrA-aroF)*, *thi-1*, *endA-1*, *hsdR17*, Δ*(argF-lac)U169*, *supE44*. Plasmid pKIMP-UAUC provides genes encoding monofunctional forms of prephenate dehydrogenase and prephenate dehydratase from *Erwinia herbicola* and *Pseudomonas aeruginosa*, respectively, and also confers chloramphenicol resistance. CMs were overproduced in *E. coli* strain KA13, which is based on KA12 but additionally carries a chromosomally integrated, IPTG-inducible T7 RNA polymerase gene.[19,20]

DNA Manipulation. Molecular cloning was performed according to standard procedures.[21] Oligonucleotides were synthesized and purified by Microsynth. Chain termination chemistry[22] was applied for DNA sequencing using BigDye Terminator Cycle Sequencing Kit version 3.1 on an ABI PRISM 3100 Genetic Analyzer (PE Applied Biosystems).

Plasmid Construction. pKTR-R26 (3692 bp) was constructed by ligating a DNA fragment (49 bp) encoding the RepA degradation tag with the 3643 bp *Nde*I−*Nco*I fragment of pKTD-R26. Two primers were used to form the DNA fragment, RepA1#136 (5′-TATGA ACCAG AGCTT TATTA GCGAT ATTCT GTATG CGGAT ATTGA ATC) and Rep1#137 (5′-CATGG ATTCA ATATC CGCAT ACAGA ATATC GCTAA TAAAG CTCTG GTTCA), forming overhangs on each end compatible with either the *Nde*I or the *Nco*I restriction site. Plasmid pKTD-R26 originates from pKT-R26,[3] into which an additional restriction site, *Nco*I, was introduced as well as the Dps degradation tag (STAKLVKSKAT).[23] A 472 bp PCR fragment with primers DpsR26 (5′-GGAAT TCCAT ATGAG CACCG CGAAG CTGGT GAAAT CCATG GCGAC ATCGG AAAAC CCGT) and M13(−20) (5′-GTAAA ACGAC GGCCA GT) was generated with template pKT-R26. Following digestion with *Nde*I and *Xho*I, the 333 bp insert was ligated with the 3338 bp *Nde*I−*Xho*I fragment of pKT-R26, yielding the 3671 bp pKTD-R26. For pKTS-mMjCM (3428 bp), a *Pst*I restriction site was introduced via two silent mutations at the codons for Q88 and K89. mMjCM was amplified by PCR with primers T7 (5′-TAATA CGACT CACTA TAGGG) and mMjCMXhoPstv (5′-GGTGC TCGAG TGTTT CCTCA AGATA TTGCT TCTGC AGGGC CTTAT TATGC TCTAT) from pET-pATCH-mMjCM. After digestion, the 306 bp *Nde*I−*Xho*I fragment was subcloned into the 3122 bp *Nde*I−*Xho*I fragment of pKTS-EcCM.[3] pET-pATCH-mMjCM was obtained from pET-mMjCM[14] by inserting the pATCH sequence as previously described.[20] The same silent mutations were also introduced during construction of pKTR-mMjCM (3695 bp). Additionally, in pKTR-mMjCM, the *Nde*I restriction site was changed to *Nco*I. Primers mMjCMNdeNco (5′-AGATT TACAT ATGGC CATGG CCATC GAGAA ACTTG CTGAA ATTAG GA) and mMjCMXhoPstv were used for PCR on template pET-pATCH-mMjCM. After digestion with *Nco*I and *Xho*I, the 308 bp insert was ligated with the 3387 bp *Nco*I−*Xho*I fragment of pKTR-R26. pKTR-mMjCMneg was used as the acceptor vector for the library. It was obtained by digesting pKTR-mMjCM with *Hin*dIII within the CM gene. The overhangs were filled in with Klenow fragment to introduce a frameshift in the gene upon religation of the blunted ends, leading to the 3699 bp pKTR-mMjCMneg plasmid. For production of CM variants, the corresponding genes were amplified by PCR. The primers used for PCR were NcoI_NdeI (5′-ATGCG GATAT TGAAC ATATG GCCA) and M13(−20). The digested 306 bp *Nde*I−*Xho*I PCR fragments were subcloned into the 5364 bp *Nde*I−*Xho*I fragment of T7 promoter vector pET-pATCH.[20]

Library Construction. For the first library, the mMjCM gene was mutagenized by epPCR (GeneMorph II Random Mutagenesis Kit, Stratagene) using primer T7, which anneals upstream and downstream of the gene, and pKTR-mMjCM as a template at a final concentration of 0.2 ng/μL. The *Nco*I- and *Xho*I-digested 308 bp PCR fragment was inserted into the correspondingly digested 3387 bp pKTR-mMjCMneg vector replacing the frameshifted fragment. For the second and third libraries, the mMjCM genes were diversified by DNA shuffling.[24] The library gene pool (100−200 clones) was amplified with primer T7 by PCR and digested with DNase I. One microliter of 0.3 mg/mL DNase I was added to 50 μL of

PCR product and the mixture incubated for 20 s to 3 min, depending on the desired fragment length (20−100 bp). The digestion was quenched by addition of EDTA (ethylenediaminetetraacetic acid) and heat inactivation of the DNase I at 70 °C for 10 min. The DNA was purified by phenol/chloroform extraction. Additionally, the fragments were purified on a 3% agarose gel. The 20−100 bp fragments were extracted and purified using a NucleoSpin gel extraction kit (Macherey-Nagel). The DNA fragments were concentrated and reassembled by a PCR-like reaction, without the addition of exogenous primers [0.5 min at 95 °C; (0.5 min at 95 °C, 0.75 min at 45 °C, and 3 min at 72 °C) × 40; 5 min at 72 °C]. Without further purification, 8 μL reaction mixture was directly subjected to a 50 μL PCR using primer T7 to amplify the reassembled gene products [0.5 min at 95 °C; (0.5 min at 95 °C, 0.5 min at 50 °C, and 1 min at 72 °C) × 40; 5 min at 72 °C]. The purified PCR fragments were inserted into the pKTR-mMjCMneg vector using the *Nco*I and *Xho*I restriction sites as described for the first library. Afterward, the ligation product was used to transform KA12/pKIMP-UAUC cells by electroporation. A portion of the electroporated cells were plated on LB Amp^{150} Cam^{30} (LB agar plates containing 150 μg/mL sodium ampicillin and 30 μg/mL chloramphenicol) to measure library sizes, and the remaining cell culture was used to inoculate 50 mL of LB Amp^{150} Cam^{30} for preparation of frozen stocks, plasmid isolation, and selection experiments.

In Vivo Assays. The complementation and selection experiments were performed using M9c medium plates at pH 7.[18] M9c medium consists of Na_2HPO_4 (6 mg/mL), KH_2PO_4 (3 mg/mL), NH_4Cl (1 mg/mL), NaCl (0.5 mg/mL), 0.2% (w/v) D-(+) glucose, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, thiamin-HCl (5 μg/mL), 4-hydroxybenzoic acid (5 μg/mL), 4-aminobenzoic acid (5 μg/mL), 2,3-dihydroxybenzoic acid (1.6 μg/mL), L-Trp (20 μg/mL), sodium ampicillin (150 μg/mL), and chloramphenicol (30 μg/mL). Fifteen grams of agar were added per liter of medium. Where required, M9c was supplemented with L-Phe and L-Tyr to a final concentration of 20 μg/mL to yield M9c FY. For the complementation assay, the appropriate amount of tetracycline (tc) was added. Cell cultures were washed with ice-cold M9c salts [Na_2HPO_4 (6 mg/mL), KH_2PO_4 (3 mg/mL), NH_4Cl (1 mg/mL), and NaCl (0.5 mg/mL)] before being plated. Plates were incubated at 30 °C.

Dilution Plating. For dilution plating, KA12/pKIMP-UAUC cells also containing pKTR-mMjCM or pKTS-mMjCM were inoculated in LB Amp^{150} Cam^{30} at 30 °C and 230 rpm until an OD_{600} of 1 was reached. The cultures were diluted to OD_{600} values of 0.1, 0.001, and 0.00001. A 10 μL aliquot of each dilution was pipetted onto LB Amp^{150} Cam^{30} as well as onto the different M9c minimal medium plates containing the appropriate tc concentration. These aliquots contained approximately 10^5, 10^3, and 10 cells per 10 μL.

Library Selection. For the selection experiments, the overnight cultures of freshly transformed cells were used. The cultures were diluted to give an OD_{600} of 1 and 1 mL of the diluted sample was washed in ice-cold M9c salts before further dilution. On average, 1 mL of a KA12/pKIMP-UAUC cell culture with an OD_{600} of 1 corresponded to approximately 5×10^7 cells. To determine the correct number of cells in the culture, the sample was diluted 10^6- and 10^7-fold and plated on M9c FY. The number of cells that are able to grow under these nonselective conditions would correspond to a "complementation frequency" of 100%. For selection, the washed cell culture was diluted 10^4- and 10^5-fold and plated on M9c supplemented with the appropriate concentration of tc. Because a complementation frequency between 0.1 and 10% was targeted, 10^4- and 10^5-fold dilution should allow growth of 10−500 clones per plate, which are easily counted. After the cells had been plated, the plates were incubated for 3 days at 30 °C.

Protein Production and Purification. CM variants were produced with pET-pATCH plasmids[20] in host strain KA13; 500 mL of LB Amp^{150} were inoculated with an overnight preculture (5 mL of LB Amp^{150}). The culture was incubated at 37 °C until an OD_{600} of 1 had been reached. Protein production was induced by adding isopropyl 1-thio-β-D-galactopyranoside (IPTG) to a final concentration of 0.25 mM. After 20 h at 20 °C, the cells were harvested by centrifugation (5000 rpm for 10 min at 4 °C) and resuspended in 10 mL of PBS buffer (pH 8) (100 mM sodium phosphate and 160 mM NaCl) supplemented with 1 mg/mL lysozyme. The suspension was kept on ice for 30 min. The cells were sonicated, and the insoluble material was removed by centrifugation (16000 rpm for 30 min at 4 °C). The soluble fraction was loaded onto a column packed with 1 mL of Ni-NTA agarose (Qiagen) that had been pre-equilibrated with PBS (pH 6.5). The column was washed with 8 mL of PBS (pH 6.5) containing 10 mM imidazole and eluted with 8 mL of PBS (pH 6.5) containing 250 mM imidazole. The proteins were further purified by preparative size-exclusion chromatography using a Superdex 75 (26/60) FPLC column (PharmaciaBiotech) with PBS (pH 6.5) as the running buffer. The monomer fraction was pooled and concentrated to a final concentration of 15−20 μM. The proteins were stored at 4 °C. Protein concentrations were determined from the absorption at 280 nm. The extinction coefficient for each CM variant was calculated with the ProtParam tool of the ExPASy Bioinformatics Resource Portal (http://web.expasy.org/protparam/).

Enzyme Kinetics. All kinetic measurements were performed at 20 °C using final enzyme concentrations of 40 nM in PBS (pH 6.5) supplemented with 0.1 mg/mL bovine serum albumin (BSA). Initial rates were determined by monitoring the disappearance of chorismate spectrophotometrically at 310 nm (ε_{310} = 370 M^{-1} cm^{-1}). The rates were corrected for the corresponding uncatalyzed reaction. Kinetic parameters k_{cat} and K_M were calculated from the initial rates as described previously.[14]

Circular Dichroism Spectroscopy. Far-UV circular dichroism (CD) experiments were performed on a circular dichroism spectropolarimeter (model 202) from AVIV Instruments Inc. CD spectra were recorded at 20 °C in PBS (pH 6.5) with 15−20 μM protein (d = 0.2 cm). Spectra were obtained by averaging five wavelength scans taken in 0.5 nm steps, with a signal averaging time of 2 s and a bandwidth of 1 nm. Temperature denaturation curves were obtained by measuring at 222 nm as the samples were heated from 10 to 90 °C in 0.5 °C temperature steps and equilibration times of 60 s.

ANS Binding. All fluorescence experiments were performed at 20 °C in PBS (pH 6.5) with a fluorescence spectrometer from PTI Photon Technology International, using a 1 cm quartz cuvette. Spectra were obtained by averaging three wavelength scans with a signal averaging time of 1 s, in 2 nm steps, and excitation and emission slits of 10 nm. ANS (2 μM) was excited at 370 nm, and fluorescence was monitored at 400−600 nm in the presence and absence of 2 μM protein.

Determination of Enzyme Concentrations in the Selection Strain. Precultures of KA12/pKIMP-UAUC containing plasmid pKTR-mMjCM, pKTR-KY30, or pKTR-mMjCMneg were grown until an OD_{600} of 1 had been reached.

Main cultures of 50 mL of LB Amp^{150} Cam^{30} containing the appropriate tc concentration were inoculated with the pre-cultures to a final OD_{600} of 0.01 and incubated overnight at 30 °C and 250 rpm. The cells were harvested by centrifugation (6000 rpm for 5 min at 4 °C) and resuspended in 4 mL of PBS (pH 7.4). The cell suspensions were diluted to give an OD_{600} of 50. Lysozyme was added to a final concentration of 1 mg/mL, and 4 mL of the resulting suspension were incubated for 30 min at room temperature. The cells were sonicated, and the insoluble material was removed by centrifugation. The supernatant was loaded onto a column packed with 1 mL of Ni-NTA agarose (Qiagen) that had been pre-equilibrated with PBS (pH 7.4). The column was washed with 3 mL of PBS (pH 7.4) and eluted with 5 mL of PBS (pH 7.4) and 250 mM imidazole, although only the last 4 mL were collected. The collected samples were washed and concentrated with YM-3 filters (Microcon). The final volume for all samples was 250 μL. Catalytic activity was measured as described above, except that no BSA was added. The chorismate concentration was 200 μM, and 20 μL of the concentrated protein sample was added to a final volume of 800 μL. The disappearance of chorismate was measured spectrophotometrically at 274 nm (ε_{274} = 2630 M^{-1} cm^{-1}).

^{1}H NMR Spectroscopy. NMR experiments were performed at 26 °C on a 600 MHz Bruker Avance II 600 spectrometer in 20 mM sodium phosphate, 50 mM NaCl (pH 6.5), and 10% D_2O. The protein concentration was 280 μM.

Trypsin Digestion. A 100 μM CM sample (150 μL) was mixed with 150 μL of trypsin from bovine pancreas [10 μg/mL, 0.43 μM (Fluka)] to start the proteolysis reaction. The 300 μL mix was incubated at 25 °C. At different time points, 30 μL aliquots were removed and the reaction was quenched with 10 μL phenylmethanesulfonyl fluoride (50 mM in 2-propanol). Two microliters of 5× SDS (DTT) loading buffer was added to 6 μL of quenched reaction mix and heated for 5 min at 95 °C. One microliter was analyzed by high-density SDS−PAGE (GE Healthcare). The gel was run and developed with the PhastSystem from GE Healthcare. The developed gel was analyzed and quantified with ImageJ (National Institutes of Health, Bethesda, MD). The intensities were normalized to the band of the undigested sample.

RESULTS

Selection Strategies. The selection system for CM activity relies on the growth-limiting turnover of chorismate to prephenate, the biosynthetic precursor of the essential amino acids phenylalanine (Phe) and tyrosine (Tyr).[25] KA12/pKIMP-UAUC is a CM knockout strain that cannot grow in the absence of Tyr and Phe.[18] Transcription of the CM gene is typically controlled by the tunable tetracycline (tc) promoter where tc serves as a transcriptional inducer.[3,10] In this system, mMjCM fully complements the metabolic defect, even in the absence of tc. To increase stringency, the *ssrA* or *repA* sequence was introduced at either the 3′ or 5′ end of the chorismate mutase gene. The resulting plasmids pKTS and pKTR encode the SsrA-tagged or RepA-tagged proteins (Figure 2).

Regulation of Protein Concentration with RepA and SsrA. The ability of the RepA and SsrA tags to reduce the intracellular mMjCM concentration to growth-limiting levels was examined at different tc concentrations. Because the extent of protein production within the cell is directly proportional to the amount of tc added,[3] lowering the tc concentration is expected to increase selection stringency (Figure 3). Cells that

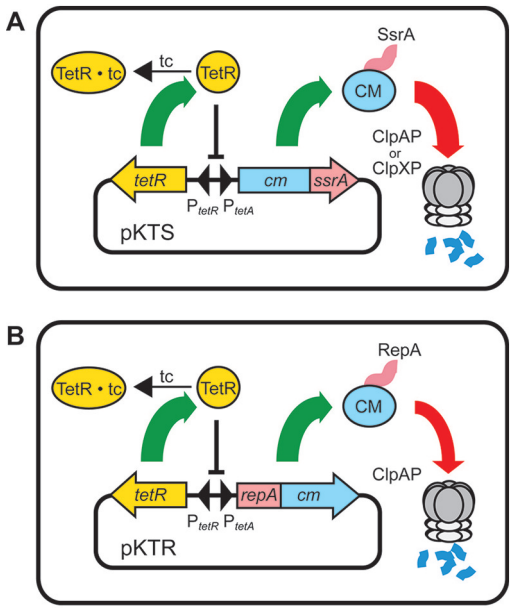


Figure 2. Strategies for increasing the stringency of a CM selection system. CM knockout selection strain KA12/pKIMP-UAUC[18] is not viable on M9c minimal medium unless complemented with a CM gene (*cm*), which is necessary for the production of Phe and Tyr. For both strategies shown here, the transcription of *cm* is controlled by the inducer tc, which binds to the tetracycline repressor (TetR). Upon tc binding, TetR detaches from the *tet* promoter region, leading to transcription of *tetR* and *cm*. *tetR* transcription is part of a feedback loop responsible for the tight transcriptional regulation by tc. KA12/pKIMP-UAUC is provided either with plasmid pKTS, encoding CM fused to a C-terminal SsrA tag (A), or with plasmid pKTR that encodes a CM fused to the N-terminal RepA tag (B). The tag directs the CM enzyme either to the ClpXP or ClpAP protease complexes for degradation.

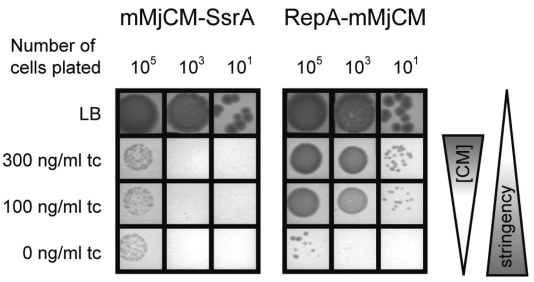


Figure 3. Complementation efficacy of wt mMjCM fused to the C-terminal SsrA tag or the N-terminal RepA tag at different inducer (tc) concentrations. Untagged mMjCM complements fully even at the lowest tc concentration (not shown).

produce SsrA-tagged mMjCM show significant reductions in growth levels at 300 ng/mL tc, whereas cells harboring RepA-mMjCM are still viable at 100 ng/mL tc. At tc concentrations lower than 100 ng/mL, the RepA tag is also successful in reducing the rate of cell growth. The SsrA tag is apparently a more efficient degradation signal than RepA because it affects growth already at higher inducer concentrations. Nevertheless, RepA is capable of lowering mMjCM levels to a growth-limiting concentration and should also be useful for regulating selection stringency in directed evolution experiments.

This control experiment provided additional insight into the frequency with which spontaneous background mutations inherent to selection systems that exploit the *tet* promoter arise.

dx.doi.org/10.1021/bi2011338 | *Biochemistry* 2011, 50, 8594−8602

As seen in Figure 3, a subfraction corresponding to 0.01−0.05% of the monoclonal RepA-tagged or SsrA-tagged cultures grew under the most stringent conditions (i.e., 10−50 colonies appeared on medium without tc). Analysis of these clones showed that such growth is enabled by spontaneous insertions and deletions within the tetracycline repressor gene (*tetR*), which abolish transcriptional repression of the *tet* promoter region (Figure 2). Unrepressed transcription leads to a high CM concentration, providing the cells with enough Phe and Tyr to permit growth. If the complementation frequency in individual selection experiments is greater than 1%, such artifacts are negligible, however.

Selection of Artifacts. The efficiency of selection strategies based on N- and C-terminal degradation tags was compared in the context of population diversification. mMjCM was mutagenized by error-prone PCR (epPCR), and the library was cloned in parallel into RepA and SsrA vectors. For the purpose of comparison, selection stringency in both cases was set so that 0.2−10% of the library members could complement the CM deficiency by varying the tc concentration. A 40-fold lower tc concentration was needed for the RepA selection experiments than for the SsrA system to afford a similar complementation frequency. At each complementation frequency, 11−17 clones were sequenced from each library, and the fraction of genuine mutants versus wt mMjCM and degradation tag artifacts was assessed. Under nonselective conditions ("100% complementation frequency"), some cells containing empty vectors were found. These were never observed under selection conditions (Figure 4). Because the entire mMjCM gene was randomized, stop codon mutations or frameshifts at the C-terminus that lead to truncation of the protein before the SsrA tag could arise, circumventing the degradation pathway. DNA sequences of clones obtained in the SsrA vector revealed that the number of artifacts derived from truncated mMjCM variants increased with increasing selection stringency. At a complementation frequency of 10%, one-third of all sequenced clones were false positives, and the fraction increased to 100% at higher stringencies. In contrast, such artifacts were never found in the RepA experiments. At a 10% complementation level, wt mMjCM was not found in either library, allowing separation of catalytically improved mutants from the parental protein. Together, these results illustrate the potential benefits of the N-terminal RepA degradation tag for controlling intracellular protein concentration in directed evolution experiments.

Directed Evolution Using the RepA Tag Strategy. mMjCM equipped with the RepA tag was subjected to three rounds of mutagenesis and selection. In the first selection round, cells were plated on minimal medium agar plates and protein production was induced with 10 ng/mL tc. The 200 largest colonies (10% of the complementers) were picked to inoculate a liquid culture. Then, the plasmid mixture of the pooled colonies was isolated and used as the template for DNA shuffling.[24] The next selection was performed at inducer concentrations of 2.5 and 5 ng/mL tc. Again, approximately 200 colonies from each plate were subjected to a second round of DNA shuffling. In the third and final selection, the tc concentration was further reduced to 0 and 2.5 ng/mL. The complementation frequency at each selection step ranged between 1 and 10% for each library. Finally, a set of four clones related by their mutation pattern was chosen for more detailed characterization (Table 1). Clones KW42 and KW46, selected in the second selection round, contain the double mutations A23fD/I76L and A23fD/N47D, respectively. Clones KY22 and KY30, identified in the third selection round, combined these three mutations (A23fD/N47D/I76L). Additionally, KY22 and KY30 contain the Y54H substitution and the Y54H and E66V

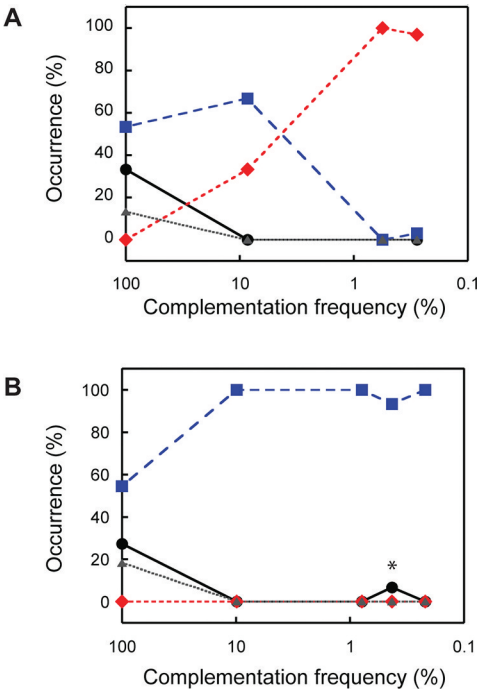


Figure 4. Comparison of the SsrA and RepA selection strategies. Complementing clones from the SsrA library (A) and RepA library (B) selected at varying complementation frequencies were sequenced and analyzed. The frequencies of occurrence of genuine mutants (filled blue squares, dashed line), parental mMjCM (filled black circles, solid line), mutants without the degradation tag (filled red diamonds, short dashed line), and the empty vector (filled gray triangle, dotted line) are plotted as a function of complementation frequency. The asterisk denotes a single parental mMjCM sequence found in the RepA-tagged library at a 0.4% complementation frequency; however, this is an artifact not caused by a loss of the degradation tag, but rather by a frameshift in the *tetR* gene.

Table 1. In Vivo Complementation Test and Protein Production Yields

variant	mutations[a]	selection round	in vivo test[b]	protein yield[c] (mg/L)
mMjCM	–	(parental)	–	10
KW42	A23fD, I76L	2	+	20
KW46	A23fD, N47D	2	+	10
KY22	A23fD, N47D, Y54H, I76L	3	++	40
KY30	A23fD, N47D, Y54H, E66V, I76L	3	++	80

[a]Nomenclature: mMjCM was generated by inserting an eight-amino acid hinge−loop sequence between residues 23 and 24 of the N-terminal helix of MjCM and replacing the six C-terminal residues QNKNKK with LEHHHHHH to facilitate purification; to preserve the numbering of the rest of the protein, the loop positions are designated as residues 23a−h.[29,30] [b]Complementation assay with selected clones after recloning and retransformation of the genes expressing the RepA-tagged CMs. Single-colony streak outs were performed on M9c plates without tc. After 3 days at 30 °C, the plates were evaluated for colony size (−, no growth; +, little growth; ++, good growth). [c]Protein yield was determined after Ni-NTA and size-exclusion chromatography of the untagged CMs.

dx.doi.org/10.1021/bi2011338 | *Biochemistry* 2011, 50, 8594−8602

substitutions, respectively. The frequency of occurrence of these mutations after the last selection round was 100% for I76L, 80% for A23fD, 40% for N47D, 20% for E66V, and 10% for Y54H. After the four variants had been recloned into fresh RepA vector (pKTR), they were retested for their complementation ability under the most stringent conditions (0 ng/mL tc). All variants grew quickly, forming single colonies that were 0.5−1 mm in diameter after 3 days at 30 °C. In contrast, wt mMjCM did not give rise to any colonies under these conditions (Table 1).

Biochemical Characterization of Evolved mMjCM Variants. The genes of the variants that conferred enhanced growth were cloned into the pET-pATCH vector[20] for protein production. The plasmid was transformed into the KA13 CM knockout strain,[19,20] which contains the T7 RNA polymerase for high-level gene expression. Wild-type mMjCM was used as a control. The His-tagged proteins were produced and purified by metal-affinity chromatography followed by size-exclusion chromatography. In contrast to the parent mMjCM protein, which forms dimers at high concentrations, all evolved variants remained monomeric below 100 μM protein. The protein yields for variants KW42 and KW46 from the second selection round are similar to that of mMjCM, whereas KY22 and KY30 from the third selection round have 4−8-fold higher yields. Thus, combination of the three mutations plus the additional Y54H mutation appears to be beneficial for protein production. The E66V mutation found in KY30 further increased the yield.

Steady-state kinetic analysis revealed that all four variants were 2−4-fold more active than mMjCM. KY30, which contains the most mutations (Figure 5), is also the most active catalyst.

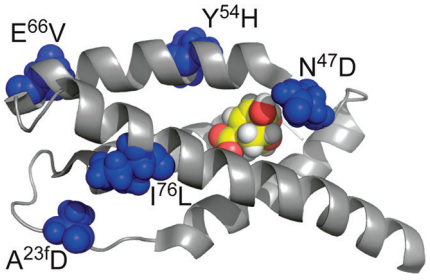


Figure 5. NMR structure of mMjCM with a bound transition-state analogue (yellow carbons) (Protein Data Bank entry 2gtv). The residues in mMjCM that were mutated in the evolved KY30 variant are highlighted in blue.

The k_{cat} for KY30 is 1.5-fold higher than that of mMjCM itself, while the K_M decreased 3-fold (Table 2). As a consequence of these changes, the catalytic efficiency of KY30 (k_{cat}/K_M = 84000 M^{-1} s^{-1}) is even 1.3 times greater than that of its natural counterpart, dimeric MjCM.

The higher activity of KY30 undoubtedly contributes to the observed growth advantage. However, the net increase in catalytic efficiency is modest. The accumulated mutations also significantly increase the intracellular stability of the enzyme. The intracellular concentrations of active enzyme were quantified for clone KY30 and the parental mMjCM, which were both produced in selection strain KA12/pKIMP-UAUC[18] at varying tc concentrations from the RepA selection plasmid. CM activity in cell extracts was measured after concentrating the CMs in the sample by affinity chromatography without a washing step to prevent potential loss of CM from the column. After the buffer had been exchanged and each sample had been concentrated to the same volume, 200 μM chorismate was added and the specific activity was determined. Using the steady-state parameters of the purified catalysts (Table 2), the concentration of enzyme present in the sample was calculated from the Michaelis−Menten equation. On average, the concentration of KY30 was 7-fold higher than that of mMjCM at all tc concentrations tested (Figure 6). These findings mirror

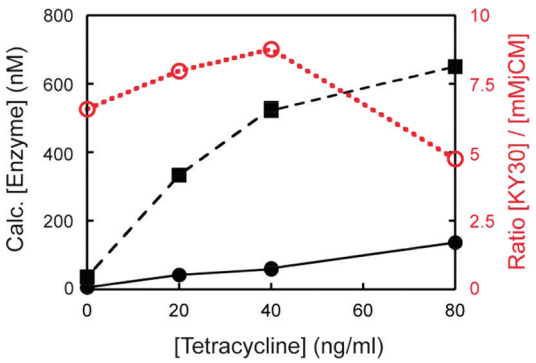


Figure 6. CM activity in extracts of cells expressing KY30 or mMjCM genes. The expression level is dependent on the concentration of the inducer tc. Cells were grown in LB supplemented with different tc concentrations as indicated. The crude cell extract was enriched for with CM by affinity chromatography. The eluate was tested for chorismate to prephenate conversion activity. The CM concentration was derived from the individual specific activities. The plot shows the calculated concentrations of KY30 (■, dashed line) and mMjCM (●, solid line) and the ratio between these two CMs (○, dotted line).

the production yields in KA13 obtained from constructs lacking the RepA tag (Table 1). Thus, the higher expression level for KY30 appears to be an intrinsic property of the evolved protein, independent of the RepA tag.

Higher intracellular enzyme concentrations could conceivably result from faster protein production (e.g., due to

Table 2. Kinetic Parameters of Untagged mMjCM and Its Variants[a]

variant	mutations[b]	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M (M^{-1} s^{-1})
mMjCM[c]	–	4.5 ± 0.4	220 ± 10	20000
KW42	A23fD, I76L	5.8 ± 0.3	160 ± 30	36000
KW46	A23fD, N47D	5.5 ± 1.2	130 ± 30	42000
KY22	A23fD, N47D, Y54H, I76L	7.1 ± 1.3	100 ± 20	70000
KY30	A23fD, N47D, Y54H, E66V, I76L	6.9 ± 0.8	80 ± 20	84000

[a]Assays for chorismate mutase activity were conducted in PBS buffer (pH 6.5) at 20 °C by monitoring the change in absorbance at 310 nm over time. [b]See Table 1 for nomenclature. [c]For comparison, literature values for mMjCM are as follows: k_{cat} = 3.2 s^{-1}; K_M = 170 μM; k_{cat}/K_M = 19000 M^{-1} s^{-1}. For MjCM, the values are as follows: k_{cat} = 3.2 s^{-1}; K_M = 50 μM; k_{cat}/K_M = 64000 M^{-1} s^{-1}.[14]

dx.doi.org/10.1021/bi2011338 | Biochemistry 2011, 50, 8594−8602

optimized codon usage) or from slower degradation by cellular proteases. A more stable tertiary structure, resulting from a decrease in molten globular character, for example, would be expected to confer greater resistance to proteases. A partial tryptic digest was performed on proteins KY30 and mMjCM to test their relative sensitivity toward proteases. As shown in Figure 7, the half-life of wt mMjCM is roughly

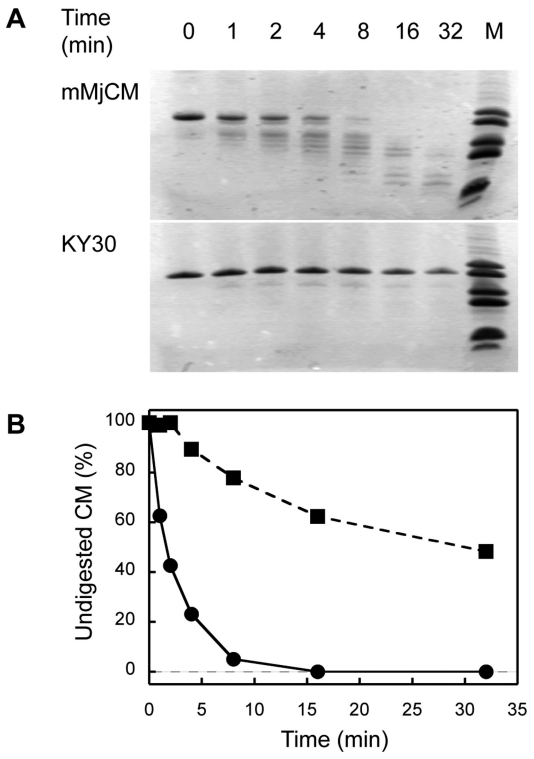


Figure 7. Partial tryptic digestion of wt mMjCM and KY30. Aliquots from the proteolysis reaction were quenched with phenylmethanesulfonyl fluoride after different incubation times. (A) SDS–PAGE gels of the quenched proteolysis reactions. Lane M contained peptide markers (16900, 14400, 10700, 8200, 6200, and 2500 Da). (B) Quantification of the partial digestion of KY30 (■, dashed line) and mMjCM (●, solid line). For each time point, the intensities of the highest protein band, corresponding to the undigested protein, were normalized to the intensity of the protein band at time zero. The starting concentrations of KY30 and mMjCM were the same (50 μM), and the intensity of the band corresponding to undigested protein at time zero is comparable for the two samples.

10 times shorter than that of KY30. To determine whether the increased proteolytic stability of KY30 derives from a more ordered fold, 8-anilinonaphthalene-1-sulfonate (ANS) binding was examined. When ANS binds to disordered proteins, its fluorescence increases significantly and a blueshift of the fluorescence maximum is observed.[15] We found that ANS shows a lower fluorescence and a less pronounced blueshift with KY30 than with wt mMjCM (Figure 8A), suggesting a less exposed hydrophobic core and a more compact fold. Another property of molten globules is their noncooperative thermal denaturation, which is observed for wt mMjCM. Thermal denaturation of KY30, however, shows a small but clear transition with a melting temperature (T_M) of 50 °C (Figure 8B). Furthermore, the ^{1}H NMR spectrum of KY30 shows greater peak dispersion than wt mMjCM, as expected for a more ordered protein (Figure 9).[15] Although KY30 is clearly

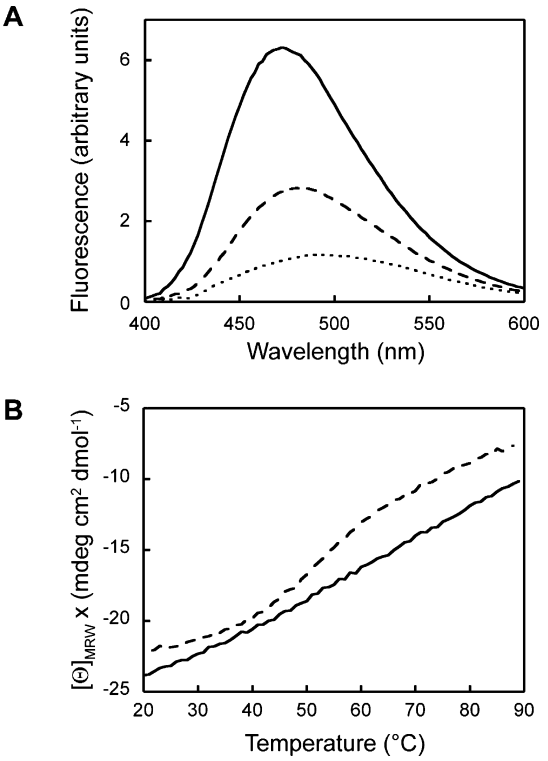


Figure 8. Biophysical characterization of KY30 and mMjCM. (A) Fluorescence measurement of ANS binding to mMjCM and KY30 at 20 °C. Spectra of free ANS (2 μM) without protein (···), ANS in the presence of 2 μM KY30 (---), and ANS in the presence of 2 μM wt mMjCM (—). (B) Thermal denaturation measured by CD at 222 nm of KY30 (---) and wt mMjCM (—). All measurements were made in PBS (pH 6.5).

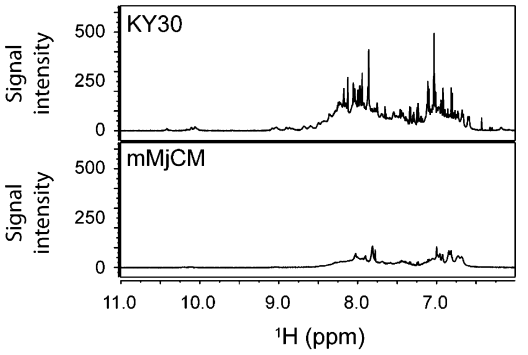


Figure 9. ^{1}H NMR spectrum of KY30 and mMjCM at 280 μM in 20 mM sodium phosphate and 50 mM NaCl (pH 6.5) in 10% D_2O.

less intrinsically disordered than wt mMjCM, its properties are not fully nativelike. In fact, the native dimeric MjCM shows much stronger cooperativity upon thermal denaturation and a T_M of almost 90 °C, a weaker fluorescence signal in the presence of ANS, and greater peak dispersion in the ^{1}H NMR spectrum.[15,19]

DISCUSSION

In vivo selection systems are powerful tools for the directed evolution of enzymes.[1,2] Most selection systems take advantage of auxotrophic strains, which are only able to grow if they are

dx.doi.org/10.1021/bi2011338 | Biochemistry 2011, 50, 8594–8602

provided with a functional enzyme that overcomes the auxotrophy. To monitor improvements in enzyme function over the course of a directed evolution experiment, the growth rate should ideally be proportional to the catalytic efficiency of the target enzyme. However, a strict correlation is seldom observed because growth also depends on the catalyst concentration. For example, selection systems often have problems distinguishing highly active enzymes from their moderately active variants. This represents an obstacle for further evolution. In such cases, strategies for increasing selection stringency are needed.

Because net catalytic turnover correlates with catalyst concentration, the easiest way to increase selection pressure is to decrease enzyme concentration. This can be accomplished by combining regulatable promotors to control the rate of protein synthesis with selective degradation tags to limit the lifetime of the protein in the cell. Our data show that the RepA degradation tag is an attractive alternative to the previously described SsrA tag[3] for directed evolution experiments. Although it is less effective than SsrA at directing appended proteins to housekeeping proteases, its N-terminal location dramatically reduces the frequency of occurrence of truncation and frameshift artifacts. As a consequence, it can be readily used to regulate selection stringency over a wide dynamic range.

The RepA tag selection strategy was successfully used here to further evolve mMjCM, a highly active, monomeric version of the CM from *M. jannaschii*.[14] After three rounds of mutagenesis and selection, the best variant, KY30, showed a >4-fold increase in catalytic efficiency (k_{cat}/K_M) compared to that of the starting enzyme. In addition, KY30 is produced at a 7-fold higher intracellular concentration than mMjCM (Figure 6). Thus, the new selection strategy afforded an approximately 30-fold increase in the level of production of prephenate from chorismate. The biophysical characterization of KY30 suggests that it is substantially better folded and more stable than its molten globular precursor (Figures 8 and 9). The more rigid fold also protects the enzyme against proteolytic degradation in the absence of the RepA tag (Figure 7).

Enhanced protease resistance and catalytic activity are important properties for biocatalysts in industrial applications.[26−28] Because high product production rates at high substrate concentrations are desirable, methods that afford catalysts with the highest possible V_{max} for a given V_{max}/K_M are important. As shown here, combining a regulatable promoter for protein production with a protein degradation tag is an effective strategy for accomplishing this goal. For this purpose, placement of the degradation signal at the N-terminus rather than at the C-terminus of the protein of interest offers significant tactical advantages. Nevertheless, the RepA tag is less effective than SsrA at reducing intracellular protein concentrations, so it may be possible to increase the dynamic range of this approach even further by identifying more efficient N-terminal degradation sequences.

AUTHOR INFORMATION

Corresponding Author

*E-mail: hilvert@org.chem.ethz.ch. Phone: +41-44-632-3176. Fax: +41-44-632-1486.

Funding

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

ACKNOWLEDGMENTS

We thank Caroline Heintz and Kathrin Székely for providing preliminary results and Dr. Marc-Olivier Ebert (nuclear magnetic resonance service of the ETH) for recording the NMR spectra.

ABBREVIATIONS

Amp, ampicillin; ANS, 8-anilinonaphthalene-1-sulfonate; BSA, bovine serum albumin; Cam, chloramphenicol; CD, circular dichroism; CM, chorismate mutase; epPCR, error-prone PCR; GFP, green fluorescent protein; IPTG, isopropyl 1-thio-β-D-galactopyranoside; k_{cat}, catalytic rate constant; K_M, Michaelis constant; LB, Luria-Bertani medium; M9c, minimal medium for chorismate mutase selection; MjCM, *M. jannaschii* chorismate mutase; mMjCM, monomeric chorismate mutase designed from MjCM; Ni-NTA, nickel nitrilotriacetic acid; NMR, nuclear magnetic resonance; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; SDS, sodium dodecyl sulfate; tc, tetracycline; TetR, tetracycline repressor; wt, wild type.

REFERENCES

(1) Taylor, S. V., Kast, P., and Hilvert, D. (2001) Investigating and engineering enzymes by genetic selection. *Angew. Chem., Int. Ed. 40*, 3310−3335.

(2) Jäckel, C., Kast, P., and Hilvert, D. (2008) Protein design by directed evolution. *Annu. Rev. Biophys. 37*, 153−173.

(3) Neuenschwander, M., Butz, M., Heintz, C., Kast, P., and Hilvert, D. (2007) A simple selection strategy for evolving highly efficient enzymes. *Nat. Biotechnol. 25*, 1145−1147.

(4) Huang, X., Zhang, X.-E., Zhou, Y.-F., Zhang, Z.-P., and Cass, A. E. G. (2006) Directed evolution of the 5′-untranslated region of the *phoA* gene in *Escherichia coli* simultaneously yields a stronger promoter and a stronger Shine-Dalgarno sequence. *Biotechnol. J. 1*, 1275−1282.

(5) Khlebnikov, A., Risa, O., Skaug, T., Carrier, T. A., and Keasling, J. D. (2000) Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *J. Bacteriol. 182*, 7029−7034.

(6) Khlebnikov, A., Skaug, T., and Keasling, J. D. (2002) Modulation of gene expression from the arabinose-inducible *araBAD* promoter. *J. Ind. Microbiol. Biotechnol. 29*, 34−37.

(7) Lutz, R., and Bujard, H. (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I_1-I_2 regulatory elements. *Nucleic Acids Res. 25*, 1203−1210.

(8) Bowers, L. M., LaPoint, K., Anthony, L., Pluciennik, A., and Filutowicz, M. (2004) Bacterial expression system with tightly regulated gene expression and plasmid copy number. *Gene 340*, 11−18.

(9) Chen, H., Bjerknes, M., Kumar, R., and Jay, E. (1994) Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of *Escherichia coli* mRNAs. *Nucleic Acids Res. 22*, 4953−4957.

(10) Saenger, W., Orth, P., Kisker, C., Hillen, W., and Hinrichs, W. (2000) The tetracycline repressor: A paradigm for a biological switch. *Angew. Chem., Int. Ed. 39*, 2042−2052.

(11) Karzai, A.W., Roche, E. D., and Sauer, R. T. (2000) The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nat. Struct. Biol. 7*, 449−455.

(12) Díaz-López, T., Lages-Gonzalo, M., Serrano-López, A., Alfonso, C., Rivas, G., Díaz-Orejas, R., and Giraldo, R. (2003) Structural changes in RepA, a plasmid replication initiator, upon binding to origin DNA. *J. Biol. Chem. 278*, 18606−18616.

(13) Hoskins, J. R., and Wickner, S. (2006) Two peptide sequences can function cooperatively to facilitate binding and unfolding by ClpA and degradation by ClpAP. *Proc. Natl. Acad. Sci. U.S.A. 103*, 909−914.

(14) MacBeath, G., Kast, P., and Hilvert, D. (1998) Redesigning enzyme topology by directed evolution. *Science 279*, 1958−1961.

(15) Vamvaca, K., Vögeli, B., Kast, P., Pervushin, K., and Hilvert, D. (2004) An enzymatic molten globule: Efficient coupling of folding and catalysis. *Proc. Natl. Acad. Sci. U.S.A. 101*, 12860−12864.
(16) Vamvaca, K., Jelesarov, I., and Hilvert, D. (2008) Kinetics and thermodynamics of ligand binding to a molten globular enzyme and its native counterpart. *J. Mol. Biol. 382*, 971−977.
(17) Bullock, W. O., Fernandez, J. M., and Short, J. M. (1987) XL1-Blue: A high efficiency plasmid transformig *recA Escherichia coli* strain with *β*-galactosidase selection. *BioTechniques 5*, 376−379.
(18) Kast, P., Asif-Ullah, M., Jiang, N., and Hilvert, D. (1996) Exploring the active site of chorismate mutase by combinatorial mutagenesis and selection: The importance of electrostatic catalysis. *Proc. Natl. Acad. Sci. U.S.A. 93*, 5043−5048.
(19) MacBeath, G., Kast, P., and Hilvert, D. (1998) A small, thermostable, and monofunctional chorismate mutase from the archaeon *Methanococcus jannaschii*. *Biochemistry 37*, 10062−10073.
(20) MacBeath, G., and Kast, P. (1998) UGA read-through artifacts: When popular gene expression systems need a pATCH. *BioTechniques 24*, 789−794.
(21) Sambrook, J., and Russell, D. W. (2001) *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press, Plainview, NY.
(22) Sanger, F., Nicklen, S., and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463−5467.
(23) Flynn, J. M., Neher, S. B., Kim, Y. I., Sauer, R. T., and Baker, T. A. (2003) Proteomic discovery of cellular substrates of the ClpXP protease reveals five classes of ClpX-recognition signals. *Mol. Cell 11*, 671−683.
(24) Stemmer, W. P. C. (1994) Rapid evolution of a protein in vitro by DNA shuffling. *Nature 370*, 389−391.
(25) Haslam, E. (1993) *Shikimic acid: Metabolism and metabolites*, Wiley, New York.
(26) Wu, X. C., Ye, R., Duan, Y., and Wong, S. L. (1998) Engineering of plasmin-resistant forms of streptokinase and their production in *Bacillus subtilis*: Streptokinase with longer functional half-life. *Appl. Environ. Microbiol. 64*, 824−829.
(27) Doran, P. M. (2006) Foreign protein degradation and instability in plants and plant tissue cultures. *Trends Biotechnol. 24*, 426−432.
(28) Fox, R. J., and Clay, M. D. (2009) Catalytic effectiveness, a measure of enzyme proficiency for industrial applications. *Trends Biotechnol. 27*, 137−140.
(29) Pervushin, K., Vamvaca, K., Vögeli, B., and Hilvert, D. (2007) Structure and dynamics of a molten globular enzyme. *Nat. Struct. Mol. Biol. 14*, 1202−1206.
(30) Woycechowsky, K., Choutko, A., Vamvaca, K., and Hilvert, D. (2008) Relative tolerance of an enzymatic molten globule and its thermostable counterpart to point mutation. *Biochemistry 47*, 13489−13496.