Supporting Information

© Wiley-VCH 2006

69451 Weinheim, Germany

Evolving proteins of novel composition

Jin Kim Montclare and David A. Tirrell[*]

[*] Prof. David A. Tirrell
Division of Chemistry and Chemical Engineering
California Institute of Technology
1200 E. California Boulevard
Pasadena, CA 91125-4100, USA

Jin Kim Montclare
Department of Chemical and Biological Sciences
Polytechnic University
6 Metrotech Center
Brooklyn, NY 11201, USA
and Department of Biochemistry
SUNY Downstate Medical Center
450 Clarkson Avenue
Brooklyn, NY 11203, USA

Complete Reference [4]: J. H. Bae, M. Rubini, G. Jung, G. Wiegand, M. H. J. Seifert, M. K. Azim, J. S. Kim, A. Zumbusch, T. A. Holak, L. Moroder, R. Huber, N. Budisa, *Journal of Molecular Biology* **2003**, *328*, 1071-1081.

Methods

Thermostability screening: Thermostability as measured by the activity after heating to 60°C for 1 hour was determined in a buffer containing 50 mM Tris, pH 7.8, 1 mM 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB, Pierce), 0.1 mM chloramphenicol, and 0.4 mM acetyl coenzyme A. As a control, a row of parent protein was arrayed along with the library in the 96-well plate. Those colonies displaying highest activity after heat treatment were selected and re-screened against the parent. In this second screen, the initial activity was assayed at room temperature prior to heating. The clone that exhibited the best thermostability and initial activity comparable to that of the parent was then expressed, purified and characterized.

Large-scale expression: Minimal M9 medium supplemented with 0.2% glucose, 3.5 μg/mL thiamine, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 20 amino acids (at 40 μg/mL of each amino acid), and antibiotics (ampicillin 200 μg/mL, kanamycin 35 μg/mL) was inoculated with an overnight culture of LAM1000/pCCCAT/pREP4 or LAM1000/pL1-C10/pREP4 or LAM1000/pL2-A1/pREP4. The cultures were grown to an OD_{600} of 0.8 to 1.0. Cells were sedimented at 6000 rpm at 4°C , washed twice with ice-cold 0.9% NaCl, and resuspended in M9 medium supplemented with 0.2% glucose, 3.5 μg/mL thiamine, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, ampicillin 100 μg/mL, 19 amino acids containing either 353 μM leucine or 500 μM TFL. Protein expression was induced 10 minutes after the medium shift by adding isopropyl-β-D-thiogalactoside (IPTG) to a final

concentration of 1 mM and incubated at 37°C. Cells were harvested after 3 hours by centrifugation at 6000 rpm at 4°C and stored at –80°C.

Activity assay on purified proteins: CAT and CAT mutants were assayed in 50 mM Tris pH 7.8, 1 mM 5,5'-dithiobis-(2-nitrobenzoic acid) (DTNB, Pierce), 0.1 mM chloramphenicol, 0.4 mM acetyl coenzyme A, 0.25 mg/mL BSA and 2 mM imidazole according to previously reported conditions.[1] Rates of CAT activity were calculated from the rate of 2-nitro-5-thiobenzoic acid formation (as the increase in absorbance at 412 nm over time) using a 96-well spectrophotometer (SPECTRAmax, Molecular Devices). Rates were determined in triplicate.

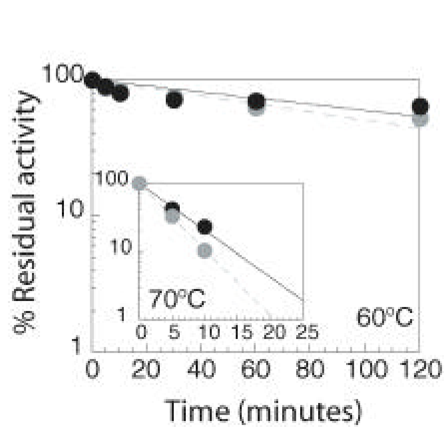

Figure S1. Kinetics of thermal inactivation of L2-A1 T (black circles) and wild-type CAT L (grey circles) at 60°C. Inset shows the kinetics of thermal inactivation of L2-A1 T and wild-type CAT L at 70°C. Data represent an average of 3 trials (standard deviation <5%).

Table S1 Library mutations

DNA pos	Generation 1 L1-C10	Generation 2 L2-A1	Mutations
137		A→T°	M_{46}: AAG →ATG
156	G→A	G→A	K_{50}: AAG →AAA
260	G→A	G→A	N_{87}: AGT → AAT
411	G→A	G→A	G_{136}: GGG→GGA
426	G→A	G→A	I_{142}: ATG → ATA
664	A→T°	A→T°	N_{222}→stop
total	5	6	DNA
mut.	2	3	protein

°Transversion mutations.
Highlighted in blue are the nonsynomyous mutations.

Table S2 Amino acid substitutions in evolved mutants

aa substitution	variant	dist. from substrate (Å)[a]	secondary structure	residue occurring at this position[b]
K46M	L2-A1	21	α helix 2	$K_5/T_3/S_2/A_4/D_3$/R/N/H_2/Q_2/P/E/G/L/V
S87N	L1-C10, L2-A1	17	surface loop	S_6/Q_2/G/K_4/E_2/C/V/Y/I/F_2/T
M142I	L1-C10, L2-A1	far	β-sheet G	$M_3/H_2/A_2/T_7/V_5$/N/L_6/F/S/I

[a]Shortest distance from this residue to the substrate; far signifies that the residue is facing away from the substrate. [b]Indicates the amino acid at the particular position from a set of homologous proteins[2-23] from the references below.

References

1. Panchenko, T. P., Zhu, W. W. & Montclare, J. K. Influence of global fluorination on chloramphenicol acetyltransferase activity and stability. *Biotechnol. Bioeng.* (2006) *in press*.
2. Bruggemann, H. et al. The genome sequence of *Clostridium tetani*, the causative agent of tetanus disease. *Proc. Natl. Acad. Sci. USA* **100**, 1316-1321 (2003).
3. Cerdeno-Tarraga, A. M. et al. Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science* **307**, 1463-1465 (2005).
4. Deppenmeier, U. et al. The genome of *Methanosarcina mazei*: Evidence for lateral gene transfer between bacteria and archaea. *J. Mol. Microb. Biotech.* **4**, 453-461 (2002).
5. Duchaud, E. et al. The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. *Nat. Biotech.* **21**, 1307-1313 (2003).
6. Galagan, J. E. et al. The genome of *M-acetivorans* reveals extensive metabolic and physiological diversity. *Gen. Res.* **12**, 532-542 (2002).
7. Gilmour, M. W., Thomson, N. R., Sanders, M., Parkhill, J. & Taylor, D. E. The complete nucleotide sequence of the resistance plasmid R478: defining the backbone components of incompatibility group H conjugative plasmids through comparative genomics. *Plasmid* **52**, 182-202 (2004).
8. Grady, R. & Hayes, F. Axe-Txe, a broad-spectrum proteic toxin-antitoxin system specified by a multidrug-resistant, clinical isolate of *Enterococcus faecium*. *Mol. Microb.* **47**, 1419-1432 (2003).
9. Heidelberg, J. F. et al. The genome sequence of the anaerobic, sulfate-reducing bacterium *Desulfovibrio vulgaris* Hildenborough. *Nat. Biotech.* **22**, 554-559 (2004).
10. Horinouchi, S. & Weisblum, B. Nucleotide-sequence and functional map of Pc194, a plasmid that specifies inducible chloramphenicol resistance. *J. Bact.* **150**, 815-825 (1982).
11. Kehrenberg, C. & Schwarz, S. Plasmid-borne florfenicol resistance in *Pasteurella multocida*. *J. Antimicro. Chemo.* **55**, 773-775 (2005).
12. Kleerebezem, M. et al. Complete genome sequence of *Lactobacillus plantarum* WCFS1. *Proc. Natl. Acad. Sci. USA* **100**, 1990-1995 (2003).
13. Ko, H. W., Jiang, J. & Edery, I. Role for Slimb in the degradation of *Drosophila* period protein phosphorylated by Doubletime. *Nature* **420**, 673-678 (2002).
14. Kuwahara, T. et al. Genomic analysis of *Bacteroides fragilis* reveals extensive DNA inversions regulating cell surface adaptation. *Proc. Natl. Acad. Sci. USA* **101**, 14919-14924 (2004).
15. Nolling, J. et al. Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bact.* **183**, 4823-4838 (2001).
16. Rabus, R. et al. The genome of *Desulfotalea psychrophila*, a sulfate-reducing bacterium from permanently cold Arctic sediments. *Env. Microb.* **6**, 887-902 (2004).

17. Rasko, D. A. et al. The genome sequence of *Bacillus cereus* ATCC 10987 reveals metabolic adaptations and a large plasmid related to *Bacillus anthracis* pXO1. *Nuc. Acids Res.* **32**, 977-988 (2004).
18. Saldarelli, P., Minafra, A. & Martelli, G. P. The nucleotide sequence and genomic organization of grapevine virus B. *J. Gen. Virol.* **77**, 2645-2652 (1996).
19. Seo, J. S. et al. The genome sequence of the ethanologenic bacterium *Zymomonas mobilis* ZM4. *Nat. Biotech.* **23**, 63-68 (2005).
20. Shirai, T. et al. High-resolution crystal structure of *M-protease*: phylogeny aided analysis of the high-alkaline adaptation mechanism. *Prot. Eng.* **10**, 627-634 (1997).
21. Wang, P. Z., Projan, S. J., Henriquez, V. & Novick, R. P. Origin recognition specificity in Pt181 plasmids is determined by a functionally asymmetric palindromic DNA element. *Embo J.* **12**, 45-52 (1993).
22. White, O. et al. Genome sequence of the radioresistant *bacterium Deinococcus radiodurans* R1. *Science* **286**, 1571-1577 (1999).
23. Xu, J. et al. A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science* 299, 2074-2076 (2003).