Directed Evolution

DOI: 10.1002/ange.200600088

Evolving Proteins of Novel Composition**

*Jin Kim Montclare and David A. Tirrell**

The behavior of proteins can be altered significantly by the incorporation of noncanonical amino acids.[1–3] Changes in spectroscopic properties,[4] thermal stability,[5,6] and molecular recognition behavior[7,8] have been reported. In many cases,

[*] Prof. D. A. Tirrell
Division of Chemistry and Chemical Engineering
California Institute of Technology
1200 E. California Blvd.
Pasadena, CA 91125 (USA)
Fax: (+1) 626-793-8472
E-mail: tirrell@caltech.edu
J. K. Montclare
Department of Chemical and Biological Sciences
Polytechnic University
6 Metrotech Center
Brooklyn, NY 11201
and
Department of Biochemistry
SUNY Downstate Medical Center
450 Clarkson Street
Brooklyn, NY 11203 (USA)

[**] We thank Frances Arnold for use of several instruments. This work was supported by NIH GM 62523 and 5F32 GM67375-2 (D.A.T. and J.K.M.) and by the Beckman Institute of the California Institute of Technology (D.A.T.).

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

however, introduction of novel amino acids causes loss of protein function.[9,10] Herein we describe a new approach to protein engineering in which amino acid replacement is combined with directed evolution to create functional proteins of novel composition. Global replacement of the leucine residues of chloramphenicol acetyltransferase (CAT) by 5′,5′,5′-trifluoroleucine (TFL) results in a 20-fold reduction in the half-life ($t_{1/2}$) of thermal inactivation of the enzyme at 60 °C. Two rounds of random mutagenesis and screening yielded a variant of CAT containing three amino acid substitutions, which in fluorinated form demonstrates a 27-fold improvement in $t_{1/2}$, recovering the loss in thermostability caused by fluorination.

CAT confers chloramphenicol resistance in bacteria by catalyzing acetyl-group transfer from acetyl coenzyme A to the hydroxy groups of the antibiotic. Because CAT is expressed readily in bacterial cells and can be assayed easily for activity and thermostability,[11,12] it provides a convenient test system for evaluation of the capacity of directed evolution to restore the loss of function that can accompany global amino acid replacement in recombinant proteins. The enzyme functions as a homotrimer in which each polypeptide chain consists of 219 amino acids, including 13 leucine residues.[11] We have shown previously that TFL serves as an effective leucine surrogate with respect to bacterial protein synthesis.[6,13] In the experiments described herein, CAT is expressed in media depleted of leucine and supplemented with 500 μM TFL. The extent of leucine replacement is 76 ± 4 %, as shown by amino acid analysis and MALDI-TOF MS.

Two parameters are used commonly to describe the thermostability of an enzyme: the half-life ($t_{1/2}$), which shows loss of activity at elevated temperature (in these experiments, 60 °C), and the temperature (T_{50}) at which half of the initial activity is lost after 30 min.[14,15] Global replacement of leucine by TFL yields a fluorinated CAT (CAT T) for which $t_{1/2} = 5$ min and $T_{50} = 57$ °C (Figures 1 a and 2 and Table 1). The corresponding values for wild-type CAT (CAT L) are 101 min and 66 °C.

A library of CAT variants was created through error-prone PCR under conditions that yielded an average of 1–2 amino acid substitutions per gene, and transformed into the leucine auxotrophic LAM 1000 strain of *E. coli*. In the first round, lysates from 1848 colonies were screened for retention of activity after incubation at 60 °C for 1 h. Rescreening the 25 most active clones led to the identification of one variant, L1-C10 T, that exhibited a fourfold increase in $t_{1/2}$ and a value of T_{50} that was 3 °C higher than that of CAT T (Figures 1 a and 2 and Table 1). This variant was used as the parent for a second round of random mutagenesis and screening. For this round, lysates from 1760 colonies were screened for retention of activity after heating to 60 °C for 1 h. The 12 best clones were rescreened and yielded one variant, L2-A1 T, which was characterized by values of $t_{1/2}$ and T_{50} equal or superior to those of wild-type CAT L (Figures 1 a and 2, Table 1, and the Supporting Information). Two rounds of mutagenesis and screening thus yielded a

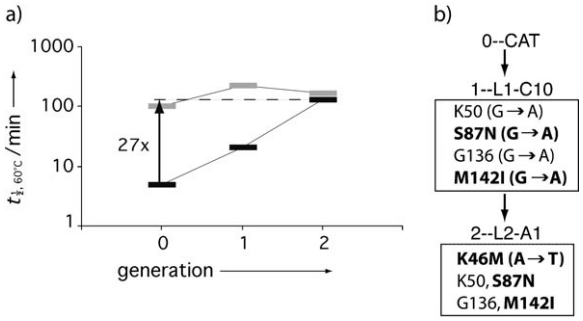

Figure 1. Thermostability landscape and identity of CAT and CAT variants. a) Evolutionary progression of thermostability of CAT and CAT variants expressed in media supplemented either with leucine (gray) or with TFL (black). Plot of the half-life for inactivation at 60 °C versus generation. The enzyme expressed in TFL-supplemented medium was screened for thermostability. b) Lineage and mutations in evolved CAT variants. Nucleotide mutations are shown in parentheses. Mutations that result in changes in protein sequence are shown in bold type. $t_{1/2,60°C}$ = half-life at 60 °C

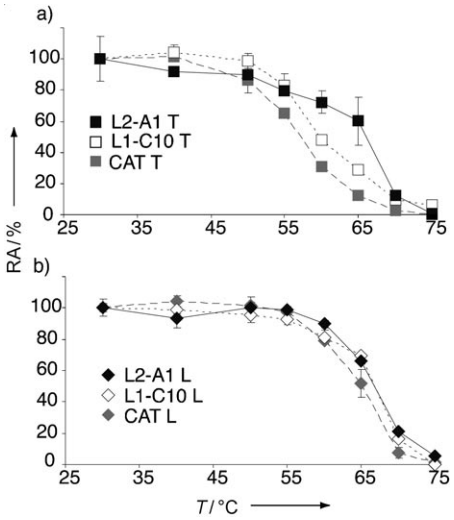

Figure 2. Residual activity versus temperature profiles. a) Residual activity versus temperature profiles of CAT T, L1-C10 T, and L2-A1 T. b) Residual activity versus temperature profiles of wild-type CAT L, L1-C10 L and L2-A1 L. All data represent an average of three trials in which error bars represent the standard deviation. Residual activity corresponds to the ratio of the activity after incubation for 30 min at the designated temperature divided by the initial activity at 30 °C. RA = residual activity.

Table 1: Kinetic parameters and thermostability of CAT variants.

Proteins	K_m [μM]		k_{cat}/K_m [10^4 μM^{-1} min^{-1}]		T_{50} [°C]		$t_{1/2}$ [min]			
	L	T	L	T	L	T	L[a]	T[a]	L[b]	T[b]
CAT	16.6 ± 3.0	19.9 ± 1.5	14.4 ± 2.7	8.4 ± 1.3	65.8	57	101	5	3	–
L1-C10	15.7 ± 2.4	11.3 ± 2.4	13.4 ± 2.2	7.7 ± 1.5	67	60	229	21	–	–
L2-A1	17.1 ± 2.4	10.0 ± 2.0	12.4 ± 1.9	12.0 ± 2.5	67	66.1	171	133	–	4

Kinetic parameters are the averages obtained in three experiments (± standard deviation). Values of $t_{1/2}$ and T_{50} are also averages from three experiments (standard deviation < 5 %). L represents enzyme expressed in leucine; T represents enzyme expressed in 5′,5′,5′-trifluoroleucine. [a] Half-life at 60 °C. [b] Half-life at 70 °C.

27-fold increase in $t_{1/2}$ and an increase of 9 °C in T_{50}. In contrast, little change in thermal stability was observed for CAT variants expressed in media supplemented with leucine.

The increase in thermostability observed in these experiments is accompanied by a modest increase in the specific activity of the fluorinated enzyme (Table 1). The first-generation variant, L1-C10 T, displayed no significant change in activity relative to CAT T; however, the specific activity of L2-A1 T was twice that of the fluorinated parent. For the enzymes expressed in leucine-supplemented media, the specific activity remained essentially unchanged (Table 1). This result was to be expected given that screening was conducted only on the fluorinated variants. The evolved variant L2-A1 T and wild-type CAT L were equally active under ambient conditions.

Mutations that improved the thermostability of CAT T were identified by DNA sequencing (Figure 1b). Among six nucleotide changes, three nonsynonymous mutations resulted in the amino acid substitutions K46M, S87N, and M142I (see the Supporting Information), and a non-sense mutation truncated the evolved polypeptide chain by one residue. None of the mutations were located in conserved positions of the enzyme as determined from sequence alignment. Of the three amino acid substitutions, only one (M142I) was reported in a homologous CAT protein (see the Supporting Information). Neither of the other two amino acid substitutions were identified in known CAT sequences, but both are considered conservative mutations.[16] Most significant, perhaps, was the fact that no leucine residues were eliminated from the protein.

A structural model of *E. coli* CAT I complexed with chloramphenicol revealed the sites of the mutations identified by directed evolution (Figure 3 and the Supporting Information).[17] Each of the mutations is located more than 15 Å from the chloramphenicol binding site, and none make direct contact with any of the TFL/leucine residues of the enzyme.

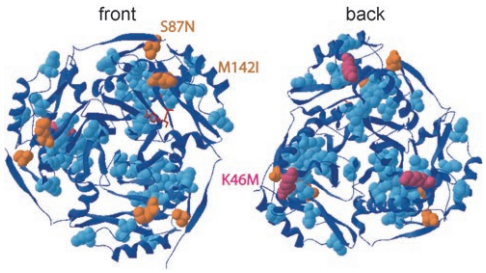

Figure 3. Structural model of the CAT trimer showing the three stabilizing mutations in L2-A1. Two of the mutations (S87N and M142I) from generation one are depicted in orange; the third mutation (K46M) from generation two is represented in pink. Residues highlighted in blue are the leucine/TFL residues of CAT; the chloramphenicol substrate is highlighted in red.

The work reported herein shows conclusively that it is possible to evolve functional proteins of novel composition. Although it is likely that incorporation of multiple copies of noncanonical amino acids into recombinant proteins will compromise function in many cases, directed evolution provides an effective means for adapting protein sequences to the inclusion of novel side chains and novel intramolecular interactions. We are exploring the limits of this approach to the engineering of proteins of novel composition, structure, and function. These results may also advance the more ambitious goal of evolving organisms with novel genetic codes.[18–20]

Experimental Section

Random mutagenesis: CAT libraries were created through error-prone PCR by using the GeneMorph Kit (Stratagene, La Jolla, CA) according to the manufacturer's protocols. The first-generation library was prepared by using template plasmid pCCCAT ($\approx$ 40 μg) and primers H1 (*Hind*III site in bold): 5′-cgt tct acc **aag ctt** cgc ccc gcc ctg cca ctc atc-3' and B1 (*Bam*HI site in bold): 5′-cgt tct acg c**gg atc c**at gga gaa aaa aat cac tgg ata tac cac c-3′. For the second-generation library, template plasmid (800 μg), and primers H2 (*Hind*III site in bold): 5′-gga gtc caa gct cag ctc tta **aag ctt** c-3′ and B2 (*Bam*HI site in bold): 5′-gca tca cca tca cca tca c**gg atc** cat g-3′ were employed. The library inserts and pCC vector (see the Supporting Information) were sequentially digested with *Hind*III and *Bam*HI, and ligated by using a insert/vector (3:1) ligation mixture for maximal ligation efficiency. The library was transformed into XL1 blue cells and grown in of 2XYT (100 mL) containing ampicillin (200 μg mL^{-1}). The amplified DNA was purified and transformed into leucine auxotrophic *E. coli* strain LAM1000 bearing the pREP4 plasmid, which carries the *lacI*^{q} gene for *lac* repressor.

Preparation of cell lysates for screening: Colonies of LAM1000/pCCCAT or pCCCAT mutants/pREP4 were transferred to a 96-deep-well plate containing M9 medium (200 μL) supplemented with glucose (0.2 %), thiamine (3.5 μg mL^{-1}), $MgSO_4$ (1 mM), $CaCl_2$ (0.1 mM), 20 amino acids (40 μg mL^{-1} of each amino acid), and antibiotics (ampicillin (200 μg mL^{-1}), kanamycin (35 μg mL^{-1})), and grown for 12 h at 30 °C at 80 % humidity at 250 rpm. The resulting starter culture plate was then used to inoculate a new 96-deep-well plate containing minimal M9 medium (400 μL, supplemented as above) and cells were grown for 12 h at 30 °C. The cells were washed twice with NaCl (0.9 %) solution prior to induction and resuspended in M9 medium supplemented with glucose (0.2 %), thiamine (3.5 μg mL^{-1}), $MgSO_4$ (1 mM), $CaCl_2$ (0.1 mM), 19 amino acids (40 μg mL^{-1} of each amino acid) plus TFL (500 μM), antibiotics (ampicillin (200 μg mL^{-1}), kanamycin (35 μg mL^{-1})), and isopropyl-β-D-galatopyranoside (IPTG; 1.0 mM), and induced for 6 h at 30 °C. Cells were harvested by centrifugation, stored at −80 °C, thawed at 37 °C in tris(hydroxymethyl)aminomethane (Tris)-Cl buffer solution (50 mM; pH 7.8) containing lysozyme (1.0 mg mL^{-1}; Sigma) and lysed by resuspension. The lysed cells were harvested by centrifugation and aliquots of the supernatant were used for analysis.

Thermostability screening: Lysed cells in 96-deep-well plates were heated in a water-bath set at 60 °C for 1 h for uniform heating. The lysates were clarified by centrifugation and aliquots (100 μL) were used for analysis. The activity of the lysates after heating was measured at room temperature as described in the Supporting Information.

Enzyme characterization: CAT and CAT variants grown in 20 amino acids and 19 amino acids plus TFL were expressed and characterized as described in the Supporting Information. The cell pellets were resuspended in Tris-Cl buffer solution (50 mM; pH 7.8) and lysozyme (0.5 mg mL^{-1}) on ice and lysed by sonication. Whole-cell lysates were clarified and subjected to purification by immobilized-metal affinity chromatography on Ni–nitrilotriacetic acid (NTA) columns according to the manufacturer's protocol (Qiagen). After loading, the column was washed with wash buffer 1A (Tris-Cl (50 mM; pH 7.8), imidazole (15 mM), glycerol (20 %), ethanol (5 %)), wash buffer 1B (Tris-Cl (50 mM; pH 7.8), imidazole (30 mM), glycerol

(20%), ethanol (5%)), and wash buffer 2 (Tris-Cl (50 mM; pH 7.8), imidazole (35 mM), glycerol (20%)). Purified protein was eluted with elution buffer 1 (Tris-Cl (50 mM; pH 7.8), imidazole (125 mM)) and elution buffer 2 (Tris-Cl (50 mM; pH 7.8), imidazole (250 mM)). The purity of each protein sample was monitored by PAGE analysis and verified by MALDI-TOF MS and amino acid analysis.

Determination of $t_{1/2}$ and T_{50}: Residual activity (RA) at various temperatures was calculated as the ratio of activity after heating divided by the initial activity measured at 30°C. Plots of RA versus temperature were used to calculate the temperature at which 50% of the enzyme activity was lost during an incubation period of 30 min (T_{50}).[14] The half lives of thermal inactivation were obtained by incubating purified proteins at specific temperatures. Aliquots were taken at various time intervals and activities were monitored. The data were fit to a first-order exponential-decay equation $V = V_o e^{-kt}$, where V_o represents the initial enzyme activity, k represents the inactivation rate constant, and t is the time in minutes.

Received: January 9, 2006
Published online: June 8, 2006

Keywords: amino acids · directed evolution · enzymes · protein engineering · thermostability

[1] A. J. Link, M. L. Mock, D. A. Tirrell, *Curr. Opin. Biotechnol.* **2003**, *14*, 603.
[2] N. Budisa, *Angew. Chem.* **2004**, *116*, 6586; *Angew. Chem. Int. Ed.* **2004**, *43*, 6426.
[3] L. Wang, P. G. Schultz, *Angew. Chem.* **2005**, *117*, 34; *Angew. Chem. Int. Ed.* **2005**, *44*, 34.
[4] J. H. Bae, et al., *J. Mol. Biol.* **2003**, *328*, 1071 (see the Supporting Information).
[5] Y. Tang, D. A. Tirrell, *J. Am. Chem. Soc.* **2001**, *123*, 11089.
[6] Y. Tang, G. Ghirlanda, W. A. Petka, T. Nakajima, W. F. DeGrado, D. A. Tirrell, *Angew. Chem.* **2001**, *113*, 1542; *Angew. Chem. Int. Ed.* **2001**, *40*, 1494.
[7] B. Bilgicer, X. Xing, K. Kumar, *J. Am. Chem. Soc.* **2001**, *123*, 11815.
[8] B. Bilgicer, K. Kumar, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 15324.
[9] N. Budisa, R. Huber, R. Golbik, C. Minks, E. Weyher, L. Moroder, *Eur. J. Biochem.* **1998**, *253*, 1.
[10] L. Wang, J. M. Xie, A. A. Deniz, P. G. Schultz, *J. Org. Chem.* **2003**, *68*, 174.
[11] W. V. Shaw, *Meth. Enzymol.* **1975**, *43*, 737.
[12] S. L. Turner, G. C. Ford, A. Mountain, A. Moir, *Protein Eng.* **1992**, *5*, 535.
[13] T. Panchenko, W. W. Zhu, J. K. Montclare, *Biotechnol. Bioeng.* **2006**, in press.
[14] H. M. Zhao, F. H. Arnold, *Protein Eng.* **1999**, *12*, 47.
[15] V. G. H. Eijsink, G. Vriend, B. Vandervinne, B. Hazes, B. Vandenburg, G. Venema, *Proteins Struct. Funct. Genet.* **1992**, *14*, 224.
[16] M. O. Dayhoff, R. M. Schwartz, B. C. Orcutt in *Atlas of protein sequence and structure*, *Vol. 5* (ed.: M. O. Dayhoff), National Biomedical Research Foundation, Washington, **1978**, p. 345.
[17] A. G. W. Leslie, J. M. Liddell, W. V. Shaw, *J. Mol. Biol.* **1986**, *188*, 283.
[18] J. M. Bacher, R. A. Hughes, J. T. F. Wong, A. D. Ellington, *Trends Ecol. Evol.* **2004**, *19*, 69.
[19] J. M. Bacher, J. J. Bull, A. D. Ellington, *BMC Evol. Biol.* **2003**, *3*, 24.
[20] R. A. Hughes, A. D. Ellington, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 1273.

