Biocatalysis

DOI: 10.1002/ange.200503031

Substrate Ambiguity and Catalytic Promiscuity Within a Bacterial Proteome Probed by an Easy Phenotypic Screen for Aldehydes**

*Guillaume Krebs, Laure Hugonet, and John D. Sutherland**

Economic and environmental considerations are driving efforts to produce biocatalysts for organic synthesis. Fermentation and the use of whole cells and enzymes are becoming increasingly common in the industrial manufacture of chemicals, but many reactions cannot be catalyzed at present.[1] In those cases where a slight activity can be found by screening, directed evolution is proving a powerful way to improve the activity,[2] but in many cases the starting activity cannot be found. We are interested in finding, and evolving, background activities for a diverse range of reactions, some of which have precedence in nature, and some not. Our approach is based upon the expectation that amongst the proteome, and its mutational variants, there must exist the catalytic potential for many new reactions, either through substrate ambiguity, or catalytic promiscuity.[3] Developing screens for a range of different reactions by using a conventional approach is time consuming, however, so we have been attracted by the idea of using one screen for many reactions. One way in which this could be achieved is by building biosynthetic trees to a

[*] G. Krebs, L. Hugonet, Professor Dr. J. D. Sutherland
School of Chemistry
The University of Manchester
Oxford Road, Manchester M13 9PL (UK)
Fax: (+44) 161-275-4939
E-mail: john.sutherland@manchester.ac.uk

[**] We thank Nicola Glynn and Fabien Buchet for assistance, Dr. Matti T. Karp, University of Turku, Finland, for his kind gift of the *luxDCABE* vector pTetLux1, and The University of Manchester for financial support.

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

common intermediate, or class of intermediates, for which an easy assay exists (Figure 1 a). Thus, if a screen for the reaction of **Y** to **Z** exists, then the same screen can be used for the conversion of precursors **V**, **W**, and **X** into **Y**. If **Y** is carefully

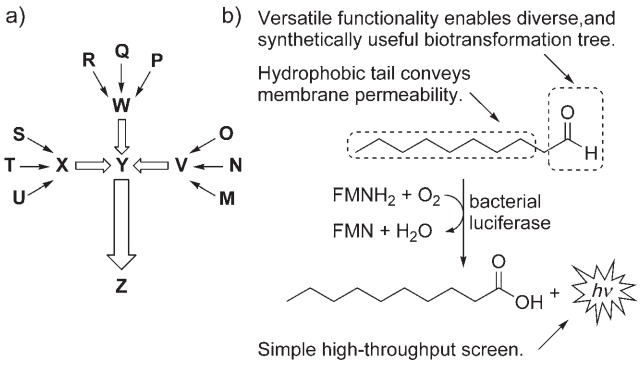


Figure 1. Building a biosynthetic product tree from a screenable reaction. a) Product tree based on the reaction **Y**→**Z**. b) The advantages of a screen using bacterial luciferase and light emission.

chosen, then it should be possible for different types of chemical reaction to be associated with these conversions. By optimizing the discovered activities and integrating them into the screening host, the process could be iterated to screen for additional chemical reaction types in the conversion of, for example, **M**, **N**, and **O** into **V** (Figure 1 a). In addition to the careful choice of **Y**, an easy screen as well as membrane-permeable precursors are crucial for this approach. Light emission is a very simple phenotype to screen, and bioluminescent bacteria can either be detected as colonies on plates or by fluorescence-assisted cell sorting (FACS).[4] The simplicity of a screen based on light emission made us consider a luciferase enzyme as our reporter, and the corresponding luciferin as the pivotal intermediate **Y**. In particular, the use of long-chain aldehydes as substrates for bacterial luciferase[5] was attractive given the central role of the carbonyl group in organic synthesis and the expected membrane permeability of such compounds (Figure 1 b).

An *Escherichia coli* luciferase expression system was developed to evaluate this approach (Figure 2). Our plan was then to introduce random fragments of the *E. coli* genome into pLH1 to generate a library of plasmids all encoding luciferase and over-expressing different *E. coli* genes. We hoped that this over-expression would mean that any trace activity of a particular protein would be amplified to the extent that it would be detectable. If the trace activity corresponded to the conversion of an aldehyde precursor into the aldehyde, then that aldehyde would be oxidized by luciferase, and light would be emitted. If this wild-type library did not yield the desired starting activities, then we planned to mutagenize it.

Before constructing and screening the library, we investigated whether different aldehydes would cause *E. coli*:pLH1 to emit light under inducing conditions. This was deemed important because the light emission would depend upon both membrane permeability and the substrate specif-

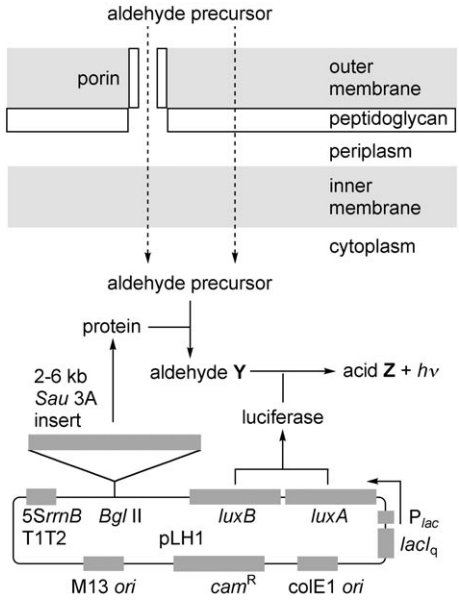
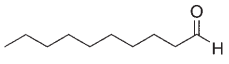
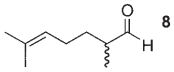
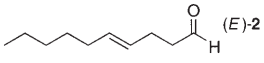
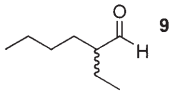
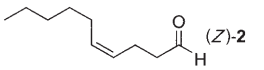
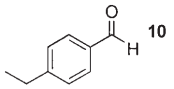
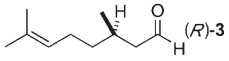
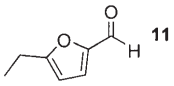
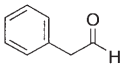
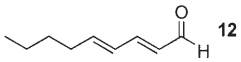
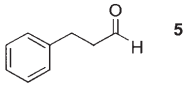
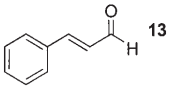
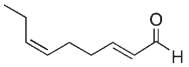
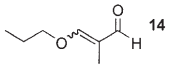
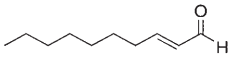
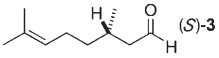


Figure 2. *E. coli* assay system. Based on the assumption that long-chain aldehydes and their precursors might freely diffuse into bacterial cells through the outer and inner membranes (and porins), a luciferase expression plasmid was used to transform *E. coli*. The plasmid pLH1 contains the *luxA* and *luxB* genes under the control of the strong inducible P_{lac} promoter. Upon induction with isopropylthio-β-D-galactoside, these genes direct the synthesis of the two subunits of bacterial luciferase. We used this basic system to investigate light emission caused by aldehydes; we used a library of *E. coli* genomic fragments cloned into the *Bgl* II site of pLH1 to screen the *E. coli* proteome for the oxidation of alcohols to aldehydes.

icity of luciferase. If many different aldehydes were to cause light emission in the assay then a greater range of chemical reactions could be screened for in subsequent evaluations of precursors. In the event, we found a wide range of aldehydes to cause light emission, as measured by luminometry (Table 1). The natural substrate for bacterial luciferase is a saturated long-chain aldehyde, most likely tetradecanal.[6] We have used shorter chain aldehydes to give slightly higher aqueous solubility, and we found excellent light emission from *E. coli*:pLH1 using decanal (**1**). Unsaturation is tolerated to varying degrees, depending on the position of the C=C bond, as shown by compounds **2**, **3**, **6**, **7**, and **12**. Branching is not accepted at the α-position (**8**, **9**, and **14**), but is at the β-position, although with a strong dependency on stereochemistry ((*R*)-**3** versus (*S*)-**3**). Aromatic substituents are tolerated at both the α- and β-positions (**4** and **5**), but conjugated aromatic aldehydes are not detectable substrates (**10**, **11**, and **13**), nor is the vinylogous ester **14**. Since all the compounds tested are nonpolar, and presumably membrane-permeable, we suspect that these trends in light emission reflect specificity for the luciferase substrate. Whatever the case, the wide range of substrates for the system suggests that a broad range of organic reactions could be screened through the use of appropriately chosen precursors. We are currently in the process of doing this, but herein report the results of some experiments on simple precursors because they validate the approach and provide evidence of substrate ambiguity, as

Table 1: Light emission from *E. coli*:pLH1 in the presence of various aldehydes.

Substrate	Light emission (RLU[a])	Not detectable as a substrate
 **1**	2.13	 **8**
 (*E*)-**2**	1.54	 **9**
 (*Z*)-**2**	1.15	 **10**
 (*R*)-**3**	0.20	 **11**
 **4**	0.18	 **12**
 **5**	0.07	 **13**
 **6**	0.04	 **14**
 **7**	0.03	
 (*S*)-**3**	0.01	

[a] Value of light emission at the maximum luminescence wavelength of 490 nm in luminescence units divided by the absorption of the sample at 600 nm (OD_{600}; see the Supporting Information for experimental procedures).

well as a remarkable example of catalytic promiscuity, in the proteome of *E. coli*.

The oxidation of an alcohol is one of the simplest ways in which an aldehyde can be synthesized, either chemically or biochemically, and we chose this reaction to evaluate screening of the *E. coli* proteome. Grossman and co-workers have also used bioluminescence to detect the oxidation of a long-chain alcohol to an aldehyde by *E. coli*.[7] In that work, however, the *Pseudomonas oleovorans* alcohol dehydrogenase, which is active on C_5–C_{12} linear alcohols, was heterologously expressed in *E. coli* because of the lack of such activity in the host. Our hope was that there would be trace activity of endogenous long-chain alcohol oxidation that would be amplified, and thus detectable, if the corresponding genes were over expressed. We selected the alcohol precursors to a good aldehyde substrate (**1**) and a poor one (**7**) to probe the sensitivity of our system. A genomic library was then constructed by cloning random 2–6 kb *E. coli Sau* 3A fragments into the *Bgl* II site of pLH1. 23000 colonies were grown on plates under inducing conditions, sprayed with *n*-decanol ($CH_3(CH_2)_9OH$, **15**), and examined for light emission in the dark with a CCD device attached to a sensitive camera. From this process 14 light-emitting colonies were selected, grown in a liquid medium, and the light emission in the presence of **15** checked by luminometry. The precise location of the selected DNA sequence in the *E. coli* genome, and, where possible, the size, were determined by sequencing both ends of the inserts (Table 2). In this way we found that over expression of three different genes caused light emission in the presence of **15**. The most commonly selected gene (*fucO*) encodes L-1,2-propanediol oxidoreductase, an enzyme involved in fucose metabolism.[8] The substrate specificity of this iron-containing alcohol dehydrogenase has not been fully mapped out, but it was previously thought to include 1,2-diols and short-chain alcohols only. The second class of insert contains genes from the ethanolamine utilization operon, and we think it likely that the product of the *eutG* gene is responsible for oxidation of **15** since it is homologous to the C terminus of *E. coli* alcohol dehydrogenase AdhE,[9] although, interestingly, none of our selected clones contained the *adhE* gene despite its likely representation in the library (see the Supporting Information). The last insert we isolated contains the *yphC* gene which encodes a putative member of the large class of zinc metalloenzyme medium-chain dehydrogenases;[10] however, light emission from these clones was lower. These results suggest that a number of alcohol dehydrogenases have a trace, low-level activity towards **15**, and it is likely that directed evolution could be used to increase this. We chose not to do this, however, and instead focused on screening our library for the precursor to the poor substrate **7**. Approximately 17000 colonies were screened for light emission in the presence of (*E*)-2-decenol (**16**), and one weakly bioluminescent clone was obtained. After purification of the colony and verification of the light-emitting phenotype in the presence of **16**, the sequence of the insert was determined. To our surprise, we found that the sequence (126809–129811) corresponded to the *lpd* gene, which encodes the E3 subunit of pyruvate dehydrogenase.[11] In the course of the action of pyruvate dehydrogenase, acetyl groups are transferred from the lipoyl domains (LDs) of the E2 subunit to coenzyme A, and the resultant LD dithiols are then oxidized to disulfides by the E3 subunit (Figure 3).[12] The lipoic acid is attached to a lysine residue of E2-LD, which itself is attached to the main body of E2 by a "swinging arm".[13] The dehydrogenase active site of E3 is at the end of a long, hydrophobic substrate-binding channel which accomodates the lipoyl lysine group of E2-LD,[14] and by implication,

Table 2: Inserts of plasmids obtained by screening the *E. coli*:pLH1-(*Sau* 3 A fragment) library for light emission in the presence of *n*-decanol (**15**).

Genome region[a]	Number of clones	Gene	Gene product
2926083–2937591	11[b]	*fucO*	L-1,2-propanediol oxidoreductase
2566061–2569982	1	*eutG*	putative alcohol dehydrogenase similar to the iron-containing alcohol dehydrogenase family
2671927–2675820	2[c]	*yphC*	hypothetical zinc-type alcohol dehydrogenase

[a] Region of the *E. coli* genome spanned by the insert or the range of inserts [http://genolist.pasteur.fr/Colibri/]. [b] Eight different clones plus three repeats. [c] One clone plus one repeat.

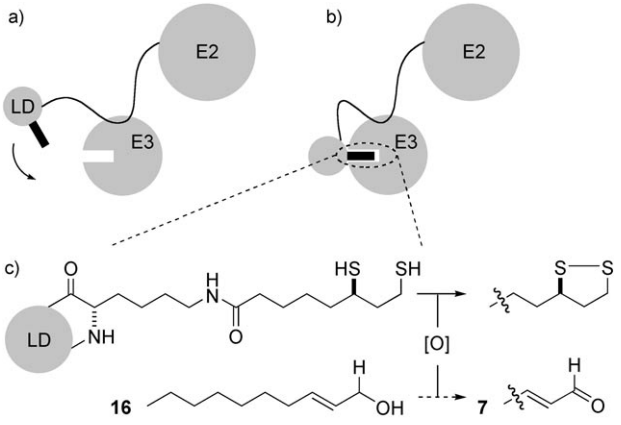


Figure 3. Catalytic promiscuity of lipoamide dehydrogenase. a) Schematic representation of part of the pyruvate dehydrogenase complex. The E3 subunit of the complex functions to oxidize the dihydrolipoamide group of a lipoyl domain (LD) of the E2 subunit.[13] b) A swinging arm[14] of E2 brings the LD into a position where the dihydrolipoamide group can dock into the active site of the E3 subunit. c) Molecular detail of the oxidative transformation of the LD dihydrolipoyllysine residue in comparison with the conversion of **16** into **7**. Although both transformations are oxidations, the conversion of **16** into **7** must be mechanistically distinct from the dithiol oxidation since it is known that the latter proceeds by way of a mixed disulfide with the enzyme.[15]

according to our results, the long chain of **16**, as a result of structural resemblance (Figure 3c)). However, the dehydrogenase uses a redox relay triad comprising an enzyme dithiol/disulfide, $FADH_2$/FAD, and NADH/NAD^+, and a mechanism not apparently adaptable to oxidation of an allylic alcohol such as **16**.[15] Thus, we wanted to prove that **16** was indeed oxidized directly to **7**, and not, for example, isomerized to **1**, which would also have resulted in light emission in the initial screen. Accordingly, we prepared a crude extract from the cell carrying the *lpd* clone and incubated it with **16** and NAD^+. ^{1}H NMR spectroscopic analysis after extraction into an organic phase clearly demonstrated the direct oxidation of **16** to **7** by the crude extract (Figure 4). We also found that the oxidation of **16** to **7** was inhibited by reduced lipoamide, a close analogue of the natural substrate for the *lpd* gene product (see the Supporting Information). These observations, coupled with the mechanism of the oxidation of the natural dithiol substrate,[15] suggest that the oxidation of the alcohol is an example of catalytic promiscuity. The actual mechanism by which **16** is oxidized to **7**, and the directed evolution of this activity will be addressed in future studies.

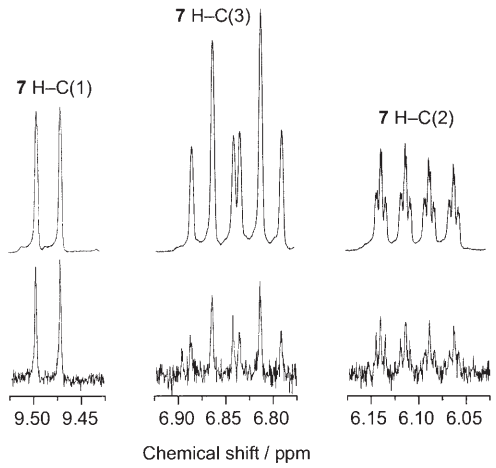


Figure 4. ^{1}H NMR analysis (300 MHz, $CDCl_3$) of the reaction of **16** with NAD^+ catalyzed by a cell-free extract of *E. coli* over-expressing the *lpd* gene. Comparison of the downfield region of the spectrum with that of a standard sample of **7** (upper) clearly shows that **16** is oxidized to **7**, and not isomerized to **1**.

The substrate ambiguity of the enzymes found to oxidize **15** to **1**, and the catalytic promiscuity observed in the oxidation of **16** to **7** by the *lpd* gene product were easily found by using the bacterial luciferase screen, and add to the growing number of reports concerning the metabolic plasticity of microbial proteomes. It is now hoped that other biocatalyst leads of use in mainstream organic chemistry can be obtained from this rich source of catalytic diversity.[16]

Received: August 25, 2005
Published online: November 29, 2005

Keywords: biocatalysis · biotechnology · enzymes · luminescence · oxidation

[1] A. Schmidt, J. S. Dordick, B. Hauer, A. Kiener, M. Wubbolts, B. Witholt, *Nature* **2001**, *409*, 258; M. Breuer, K. Ditrich, T. Habicher, B. Hauer, M. Kesseler, R. Stürmer, T. Zelinski, *Angew. Chem.* **2004**, *116*, 806; *Angew. Chem. Int. Ed.* **2004**, *43*, 788.

[2] W. P. C. Stemmer, *Nature* **1994**, *370*, 389; J. D. Sutherland, *Curr. Opin. Chem. Biol.* **2000**, *4*, 263; K. A. Powell, S. W. Ramer, S. B. del Cardayré, W. P. C. Stemmer, M. B. Tobin, P. F. Longchamp, G. W. Huisman, *Angew. Chem.* **2001**, *113*, 4068; *Angew. Chem. Int. Ed.* **2001**, *40*, 3948; M. T. Reetz, *Angew. Chem.* **2001**, *113*, 292; *Angew. Chem. Int. Ed.* **2001**, *40*, 284; S. Brakmann, K. Johnsson, *Directed Molecular Evolution of Proteins (or How to Improve Enzymes for Biocatalysis)*, Wiley-VCH, Weinheim, **2002**; M. T. Reetz, M. Bocola, J. D. Carballeira, D. Zha, A. Vogel, *Angew. Chem.* **2005**, *117*, 4264; *Angew. Chem. Int. Ed.* **2005**, *44*, 4192.

[3] P. J. O'Brien, D. Herschlag, *Chem. Biol.* **1999**, *6*, R91; S. Raillard, A. Krebber, Y. Chen, J. E. Ness, E. Bermudez, R. Trinidad, R. Fullem, C. Davis, M. Welch, J. Seffernick, L. P. Wackett, W. P. C. Stemmer, J. Minshull, *Chem. Biol.* **2001**, *8*, 891; U. T. Bornscheuer, R. Kazlauskas, *Angew. Chem.* **2004**, *116*, 6156; *Angew. Chem. Int. Ed.* **2004**, *43*, 6032; A. Aharoni, L. Gaidukov, O. Khersonsky, S. McQ. Gould, C. Roodveldt, D. S. Tawfik, *Nat. Genet.* **2005**, *37*, 73.

[4] M. J. Olsen, D. Stephens, D. Griffiths, P. Daugherty, G. Georgiou, B. L. Iverson, *Nat. Biotechnol.* **2000**, *18*, 1071; S. W. Santoro, L. Wang, B. Herberich, D. S. King, P. G. Schultz, *Nat. Biotechnol.* **2002**, *20*, 1044.

[5] a) T. O. Baldwin, T. Berends, T. A. Bunch, T. F. Holzman, S. K. Rausch, L. Shamansky, M. L. Treat, M. M. Ziegler, *Biochemistry* **1984**, *23*, 3663; b) M. T. Korpela, J. S. Kurittu, J. T. Karvinen, M. T. Karp, *Anal. Chem.* **1998**, *70*, 4457.

[6] S. Ulitzur, J. W. Hastings, *Proc. Natl. Acad. Sci. USA* **1979**, *76*, 265; W. A. Francisco, H. M. Abu-Soud, T. O. Baldwin, F. M. Raushel, *J. Biol. Chem.* **1993**, *268*, 24734.

[7] V. Minak-Bernero, R. E. Bare, C. E. Haith, M. J. Grossman, *Biotechnol. Bioeng.* **2004**, *87*, 170.

[8] A. Boronat, J. Aguilar, *J. Bacteriol.* **1979**, *140*, 320; C. Montella, L. Bellsolell, R. Pérez-Luque, J. Badía, L. Baldoma, M. Coll, J. Aguilar, *J. Bacteriol.* **2005**, *187*, 4957.
[9] I. Stojiljkovic, A. J. Bäumler, F. Heffron, *J. Bacteriol.* **1995**, *177*, 1357.
[10] J. Shafqat, J.-O. Höög, L. Hjelmqvist, U. C. T. Opperman, C. Ibàñez, H. Jörnvall, *Eur. J. Biochem.* **1999**, *263*, 305.
[11] D. J. Carothers, G. Pons, M. S. Patel, *Arch. Biochem. Biophys.* **1989**, *268*, 409.
[12] S. S. Mande, S. Sarfaty, M. D. Allen, R. N. Perham, W. G. J. Hol, *Structure* **1996**, *4*, 277.
[13] R. N. Perham, *Annu. Rev. Biochem.* **2000**, *69*, 961.
[14] A. Mattevi, A. J. Schierbeek, W. G. J. Hol, *J. Mol. Biol.* **1991**, *220*, 975; A. Mattevi, G. Oblomova, K. H. Kalk, W. J. H. van Berkel, W. G. J. Hol, *J. Mol. Biol.* **1993**, *230*, 1200.
[15] K. D. Wilkinson, C. H. Williams, Jr., *J. Biol. Chem.* **1979**, *254*, 852; A. Argyrou, J. S. Blanchard, B. A. Palfey, *Biochemistry* **2002**, *41*, 14580.
[16] M. Müller, *Curr. Opin. Biotechnol.* **2004**, *15*, 591.

