Ligation independent cloning (LIC) as a rapid route to families of recombinant biocatalysts from sequenced prokaryotic genomes

Daniel Bonsor,[a] Stephanie F. Butz,[a] Julianna Solomons,[a] Stephanie Grant,[b] Ian J. S. Fairlamb,[b] Mark J. Fogg*[a] and Gideon Grogan*[a]

Received 6th December 2005, Accepted 12th January 2006
First published as an Advance Article on the web 30th January 2006
DOI: 10.1039/b517338h

A technique is presented for the high throughput generation of families of recombinant biocatalysts sourced from prokaryotic genomes, providing rapid access to the naturally evolved diversity of enzyme specificity for biocatalyst discovery. The method exploits a novel ligation independent cloning strategy, based on the locally engineered vector pET-YSBLIC and has been used for the rapid generation of a suite of expression plasmids containing genes encoding a family of six Baeyer–Villiger monooxygenases (BVMOs) from *Mycobacterium tuberculosis* H37Rv (MTb). The six resultant recombinant strains of *E. coli* B834 (DE3) expressing the genes were assayed for oxygenating activity in respect of the target reaction; the resolution of bicyclo[3.2.0]hept-2-en-6-one. The analysis of biotransformations catalysed by growing cells of *E. coli* was complicated by the production of indole in the reaction mixtures, possibly resulting from the *in vivo* activity of *E. coli* tryptophanase. Four of the recombinant strains expressing different BVMOs catalysed the oxidation of one or more of four screening substrates, well above controls that had been transformed with the re-ligated parent vector. One of the recombinant strains, *E. coli* B834 (DE3) pDB5, expressing the Rv3049c gene from MTb, was found to effectively resolve the target substrate, yielding a 19% yield of (1*R*, 5*S*)-(+)-bicyclo[3.2.0]hept-2-en-6-one with >95% enantiomeric excess in a 4 L fermentation reaction.

Introduction

The steadily increasing number of available prokaryotic genome sequences promises to be a valuable source of genes encoding useful biocatalysts of as yet indeterminate chemical reactivity and selectivity. In the search for novel biocatalysts of desired function, the use of random mutagenesis methods to evolve biocatalysts *in vitro* to possess altered properties is becoming widespread through the increasing ease of use of error-prone PCR methods[1] and mutator strains of *Escherichia coli*.[2] However, a complete focus on such evolution methods would be in danger of ignoring the bounty of natural biocatalysts encoded by prokaryotic genomes that still remain uninvestigated and untapped. This is confirmed by recent reports both of biocatalysts of unprecedented activity resulting from amplification of genes from environmental DNA samples[3] and the dissection of the distinctive, and in some cases opposing, specificity of isofunctional enzymes that contribute to an overall whole-cell catalysed biocatalytic process.[4] Methods that allow rapid access to naturally evolved populations of isoenzymes with significant differences in sequence might be exploited for differences in reaction selectivity and perhaps for ease of crystallisation in structural enzymology studies.

In addition to the development of high-throughput selections and screens for the rapid evaluation of enzyme activity and selectivity, high throughput access to gene products for assay depends largely on traditional cut-and-paste methods of cloning, which can prove time consuming and unreliable. The development of 'ligation independent cloning' (LIC) in the 1990s obviated the use of a ligation reaction in the cloning protocol by using the exonuclease properties of a DNA polymerase to generate cohesive ends of sufficient length such that annealing was automatic.[5,6] This resulted in a rapid method for the generation of gene products that removed the ligation bottleneck that could be compromised by a number of factors such as temperature and ligase quality. In this technique (Fig. 1), the PCR purified product is incubated with T4 DNA polymerase, and uses the natural 3′-5′ exonuclease activity of the enzyme, in the presence of a single nucleotide, to produce highly specific single stranded overhangs at both ends of the insert. The single stranded sequences created in this way are complementary to overhangs in the linearised LIC engineered vector that have been derived in a similar manner, producing 5′ single stranded regions of 13 bases on the antisense strand and 14 bases in the sense strand of the vector that are not complementary to each other, thus ensuring directional cloning. A mixture of T4 DNA polymerase treated insert and vector are then simply annealed in a ten minute reaction prior to transformation (Fig. 1, **2**). The recombinant LIC vector mix can be directly transformed into an appropriate cloning or expression strain. The complementarity of the single stranded regions guarantees that the vector and insert stay annealed during the transformation process, and ligation occurs as a result of the action of the DNA repair enzymes in the host cell.

An important feature of the LIC method is that programs for the automated design of oligonucleotide primer sequences for use in the initial PCR reaction can be used, as the LIC process is independent of the target sequence, with no consideration of

[a]*York Structural Biology Laboratory, Department of Chemistry, University of York, Heslington, York, UK YO10 5YW. E-mail: grogan@ysbl.york.ac.uk; Fax: +44 (0)1904 328266; Tel: +44 (0)1904 328256*
[b]*Department of Chemistry, University of York, Heslington, York, UK YO10 5DD*

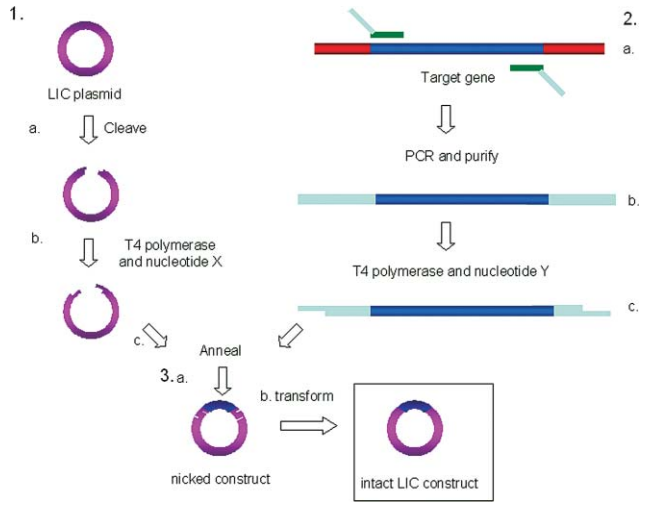


Fig. 1 Ligation independent cloning (LIC) of PCR products. 1. Preparation of LIC plasmid. (a) Vector is cleaved at the LIC site using a specific restriction endonuclease. Linearised vector is purified then (b) incubated with T4 DNA polymerase and a single deoxyribonucleotide (X = dATP, dTTP, dCTP or dGTP) to generate single stranded 5′ termini overhangs. (c) The vector is purified. 2. Preparation of PCR product for LIC. (a) Target gene is amplified by PCR using gene specific primers with LIC ends attached to their 5′ termini. (b) PCR product is purified then (c) incubated with T4 DNA polymerase and a single deoxyribonucletide (Y) complementary to that used for the LIC plasmid preparation, to generate a LIC insert with single stranded 5′ overhangs complementary to those generated in the vector. 3. (a) Purified plasmid and insert are mixed and annealed at room temperature (b) Nicks are repaired when mixture is transformed into cloning strain of *E. coli.*

engineered restriction sites for cloning. A variant of an LIC vector based on the Novagen plasmid pET-28a has been engineered in our research laboratories and used for the high-throughput generation of plasmid constructs for structural proteomics studies of pathogenic organisms such as *Bacillus anthracis.* The new vector, named pET-YSBLIC, features an N-terminal hexahistidine tag and restriction sites that allow for the transfer of the LIC region, if required, to other plasmids using established restriction/ligation cloning protocols. It occurred to us that the LIC strategy adopted in our laboratory for structural proteomics would also lend itself well to the rapid generation of families of recombinant 'designer' biocatalysts (described below), using the sequenced genomes of prokaryotes as a starting point for identifying potential open reading frames encoding useful enzymes.

The recent development of 'designer biocatalysts' has highlighted the usefulness of expressing genes encoding useful biocatalysts in recombinant systems in which amounts of enzyme many times larger than the wild-type may be obtained.[7] In addition, factors such as pathogenicity of the host wild-type strain, the peculiarities of wild-type organism biochemistry (esoteric carbon sources required for enzyme induction, overmetabolism or side reactions) or the competing activities of homologous enzymes in the wild-type organism can be eliminated in the recombinant biocatalyst. This would be particularly useful for oxidoreductase-dependent reactions, where problems of cofactor recycling render the cell-free biotransformation uneconomic. This strategy has recently proved particularly useful in delineating the contributions of various oxidoreductase enzymes from *Saccharomyces cerivisiae* to established whole-cell biotransformations using this familiar biocatalyst.[4]

The recombinant biocatalyst technique has also proved useful in the study of flavin containing monooxygenase enzymes catalysing the Baeyer–Villiger reaction [Baeyer–Villiger MonoOxygenases (BVMOs) or Baeyer–Villigerases (BVases) *e.g.* E.C. 1.14.13.22], with genes encoding enzymes from *Acinetobacter*,[8] *Comamonas* (formerly *Pseudomonas*),[9] *Brevibacterium*[10] and other environmental isolates[11] expressed in both recombinant *E. coli*[12] and *S. cerivisiae.*[7] A number of new enzymes have been described that display a wide range of selectivities that may not have been revealed by *in vitro* evolution techniques.

As part of a long-standing interest in Baeyer–Villigerases,[13,14] we chose to screen a readily available genomic DNA of an organism for genes encoding putative novel BVases that might be capable of the *resolution* of bicyclo[3.2.0]hept-2-en-6-one, one of a series of related substrates that are currently of interest in our laboratory for their potential use as precursors of ligands for transition metal catalysis.[14] The genome system chosen for study was *Mycobacterium tuberculosis* H37Rv,[15] whose genetics are comparatively well understood, and for which the relevant genomic and/or cosmid DNA was readily available from our collaborators. Using the primary sequence of the well-studied cyclohexanone monooxygenase (CHMO) from *Acinetobacter calcoaceticus* NCIMB 9871[16] as a model, we identified the six closest homologs in the MTb H37Rv genome[15] and chose these as our targets for LIC based development. Baeyer–Villigerases from *Mycobacterium tuberculosis* have recently aroused interest as the gene product encoded by Rv3854c, known as EtaA, is a monooxygenase that has been implicated in the sensitisation of the organism to the second-line antibiotic ethionamide.[17,18] Preliminary studies were reported on the substrate specificity of the enzyme, which, in conjunction with a now recognised sequence-motif for Baeyer–Villigerases,[19] identified it as an enzyme of that activity. The six gene targets, Rv0892 (predicted protein monomeric molecular weight 55.0 kDa), Rv0565c (54.5 kDa), Rv3854c (55.3 kDa), Rv1393c (55.3 kDa), Rv3049c (58.7 kDa) and Rv3083 (55.5 kDa) were subjected to a process of PCR, LIC, IPTG-induced expression and whole-cell biotransformation assay. Whilst the rapid throughput approach met with varying levels of success at each stage, a recombinant biocatalyst expressing gene Rv3049c has been created that catalysed the preparative-scale resolution of the target compound, bicyclo[3.2.0]hept-2-en-6-one.

Results and discussion

Target selection

The open-reading frames selected for the study were the six closest homologues to the gene encoding the cyclohexanone monooxygenase (AcCHMO) from *Acinetobacter calcoaceticus* NCIMB 9871, as identified by BLAST searching against the genome of MTb H37Rv on the PEDANT website (http://www.pedant.de/). The six closest homologues were, with amino acid sequence identities with AcCHMO in parentheses, Rv0892 (26%), Rv0565c (18%), Rv3854c (19%), Rv1393c (28%), Rv3049c (24%) and Rv3083 (19%). A section of a pairwise alignment of the relevant amino acid sequences, constructed using CLUSTALW[20] is shown in Fig. 2. In each case the six putative open reading frames

```
Rv1393c  IRLNTKVLAAEFDDEHSLWRVQ-----TDPGGEITARFLISACGILTVPK--LPDIDGVD 150
Rv3049c  IEFNSLVDRGYWDDDECRWHVF-----TADGREYVAQFLISGAGALHIPS--FPEIAGRD 167
Rv0892   IRFGATVVSARFDDG--RWVLR-----TDSGTESTVDFLISATGVLHHPR--IPPIAGLD 149
Rv3854c  IRFHHKVISADWSTAENRWTVH--IQSHGTLSALTCEFLFLCSGYYNYDEGYSPRFAGSE 155
Rv3083   IHYGLKVNTAEWSSRQCRWTVAGVHEATGETRTYTCDYLISCTGYYNYDAGYLPDFPGVH 157
Rv0565c  IEFNSYVRAADWDSSTDTWTVT--FEQNGVHKHYRSRFVFFGSGYYNYDEGYTPDFGGIE 160
         *.    *  . :.     * :               :::   *         * : * .

Rv1393c  SFEGVTMHTARWDHTQDLTGKRVGIIGTGASAVQVIPEMAPIVS----HLTVFQRTPIWC 206
Rv3049c  EFAGPAFHSAQWDHSIDLTGKRVAIVGTGASAIQIVPEIVGQVA----ELQLYQRTPPWV 223
Rv0892   DFRGTVFHSARWDHTVPLLGRRIAVIGTGSTGVQLVCGLAGVAG----KVTMFQRTAQWV 205
Rv3854c  DFVGPIIHPQHWPEDLDYDAKNIVVIGSGATAVTLVPALADSG---AKHVTMLQRSPTYI 212
Rv3083   RFGGRCVHPQHWPEDLDYSGKKVVVIGSGATAVTLVPAMAGSNPGSAAHVTMLQRSPSYI 217
Rv0565c  KFGGAVVHPQHWPEDLDYTGKKIVVIGSGATAVTLIPSLTDRA----EKVTMLQRSPTYL 216
          * *   *. :* .     .:.: ::*:*::.: ::  :.        .: : **:. :

         BVase motif
```

Fig. 2 Partial CLUSTALW pairwise alignment of six open-reading frames from the genome of *Mycobacterium tuberculosis* H37Rv, illustrating conservation of 'Baeyer–Villigerase' (BVase) motif [**FXGXXXHXXXW(P/D)**].

were observed to encode BVases as defined by the sequence motif [**FXGXXXHXXXW(P/D)**] described by Fraaije and co-workers.[19]

PCR of genes encoding six putative BVases and ligation independent cloning into pET-YSBLIC

The high-throughput amplification of genes from GC rich sources such as the DNA of actinomycetes like MTb does not, in our experience, allow a generality of approach, as conditions used for the amplification of one gene may be unsuitable for another. This proved to be the case for the six ORFs under consideration, each of which required subtly different conditions from either genomic or cosmid template. Table 1 lists the PCR conditions required for each individual gene. The amplification products of successful PCR reactions were cleaned and used in the LIC protocol to generate six recombinant plasmids, pDB1, pDB2, pDB3, pDB4, pDB5, pDB6, corresponding to products obtained from the insertion of Rv0892, Rv0565c, Rv3854c, Rv1393c, Rv3049c and Rv3083 into the YSBLIC plasmid. After transformation into Stratagene XL1-Blue competent cells, clean recombinant plasmids were prepared by miniprep for transformation into expression strains, and the identities of the inserts confirmed by sequencing.

Expression

Initial experiments on the transformation of plasmids DB1-DB6 into a selection of available expression strains of *Escherichia coli* [BL21 Gold, BL21 (DE3), B834 (DE3), Rosetta] indicated that transformation efficiencies were overall best in B834 (DE3), a methionine auxotroph routinely used in our laboratory for the preparation of selenomethionine derivatives of proteins for X-ray crystallographic studies. The six B834 (DE3) expression strains were thus numbered *E. coli* B834 (DE3) pDB1 through to *E. coli* B834 (DE3) pDB6. A series of temperature and inducer concentration experiments determined that in the interests of uniformity, consonant with the rapid throughput analysis of gene products, the highest levels of expression of most of the BVase genes was obtained using a growth protocol where cells were grown to an optical density A_{600} measurement of 0.5, followed by addition of 1 mM isopropyl-β-thiogalactopyranoside (IPTG), followed by growth at 20 °C overnight. An SDS-PAGE gel of samples of the total cell fraction from each recombinant strain compared to a control is shown in Fig. 3. The gel suggests that of the genes cloned in this study (Rv0892, Rv0565c, Rv3854c, Rv1393c, Rv3049c and Rv3083) genes Rv0892 (Lane 2) and Rv1393c

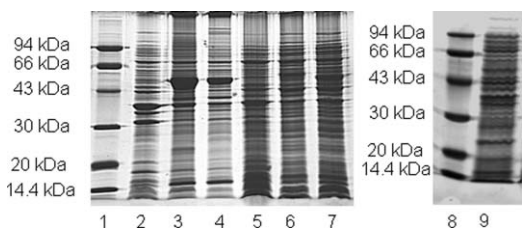


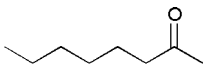
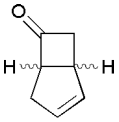
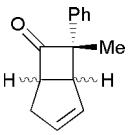
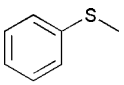
Fig. 3 Total cell extracts of *E. coli* strain B834 (DE3) expressing genes encoding putative Baeyer–Villigerases, analysed by SDS-PAGE. Lane 1, 8: Low molecular weight markers (BioRad); Lane 2, *E. coli* B834 (DE3) pDB1; Lane 3, pDB2; Lane 4, pDB3; Lane 5, pDB4; Lane 6, pDB5; Lane 7, pDB6. Lane 9, *E. coli* B834 (DE3) transformed with re-ligated YSBLIC with no insert.

Table 1 PCR conditions for amplification of LIC inserts for six genes encoding putative BVases

Gene	Template	Additions[a]	Melt	Anneal	Extend	Final extend
Rv0892	Cosmid MTCY31	—	94 °C, 150 s	66 °C, 30 s	72 °C, 120 s	72 °C, 120 s
Rv0565c	Genomic DNA	—	95 °C, 150 s	60 °C, 30 s	72 °C, 120 s	72 °C, 180 s
Rv3854c	Cosmid MTCY01A6	—	94 °C, 150 s	51 °C, 30 s	72 °C, 110 s	72 °C, 120 s
Rv1393c	Cosmid MTCY21B4	—	94 °C, 150 s	56 °C, 30 s	72 °C, 120 s	72 °C, 180 s
Rv3049	Genomic DNA	2 μl 25 mM $MgSO_4$	95 °C, 360 s	61 °C, 60 s	72 °C, 120 s	72 °C, 180 s
Rv3083	Genomic DNA	2 μl DMSO	95 °C, 360 s	60 °C, 60 s	72 °C, 150 s	72 °C, 180 s

[a] Additions to mix described in experimental section, replacing an equal volume of sterile deionised H_2O.

Table 2 Oxidations of four test substrates by recombinant strains of *E. coli* B834 (DE3) (pDB1 *etc.*) transformed with pET-YSBLIC-derived plasmids containing genes encoding six putative BVases

Substrate	% Conversion to oxidation products observed after 4 h						
YSBLIC	pDB1	pDB2	pDB3	pDB4	pDB5	pDB6	pET-YSBLIC
 **1**	—	—	39	—	50	5	—
 **2**	—	—	18	—	30	8	—
 **3**	—	—	—	—	Trace	—	—
 **4**	20	56	57	18	59	45	16

(Lane 5) were expressed poorly if at all, whereas genes Rv0565c, Rv3854c, Rv3049c and Rv3083 (Lanes 3, 4, 6 and 7) appeared to show significant expression as suggested by bands in the relevant molecular weight range of approximately 55–60 kDa. In strains expressing Rv0565c, Rv3854c, Rv3049c and Rv3083, significant amounts of protein were observed in the soluble fraction after cell disruption and centrifugation (data not shown). The data for strain *E. coli* B834 (DE3) pDB3, containing gene Rv3854c, whose expressed product EtaA is the monooxygenase responsible for sensitisation of MTb to ethionamide, are consistent with the accessibility of this enzyme reported in previous studies.[17,18]

Biotransformation of BVase substrates

Using protocols described by Stewart and coworkers,[11] growing cell biotransformations of a series of four substrates previously identified as substrates for Baeyer–Villigerases were performed. Typically, substrates were added to growing cells 45 min after induction with IPTG. The results in Table 2 show that recombinant strains *E. coli* B834 (DE3) pDB3, *E. coli* B834 (DE3) pDB5 and *E. coli* B834 (DE3) pDB6 exhibited the most substantial oxygenating activity, with significant conversions of octan-2-one, bicyclo[3.2.0]hept-2-en-6-one and phenyl-methyl sulfide, compared with control reactions that had been performed with *E. coli* B834 (DE3) transformed with the LIC plasmid re-ligated without insert. A significant amount of oxidation of thioanisole **4** was observed in the control reaction. Interestingly, *E. coli* pDB2 catalysed sulfide oxidation well above control levels, but did not catalyse Baeyer–Villiger type oxygen insertion reactions in this case. **1**, **2** and **4** had already been shown to be substrates for the enzyme EtaA, encoded by the gene Rv3854c expressed in *E. coli* pDB3.[18] However, the recombinant biocatalyst *E. coli* B834 (DE3) pDB5 expressing Rv3049c was found to be the most effective BVase biocatalyst created using the protocols described herein.

One notable phenomenon observed using growing cell reactions was the production of a metabolite (Fig. 4) that was independent of the biotransformation substrate added to the growing cells. This compound was observed in large amounts in GC chromatograms of ethyl acetate extracts of overnight test reactions and was also extracted from preparative scale reactions, complicating chromatography of biotransformation products. Isolation and ^{1}H NMR analysis of the metabolite revealed it to be indole. This is perhaps unsurprising as indole is known to be produced by strains of *Escherichia coli* by the action of the enzyme tryptophanase on tryptophan[21] and indeed, the production of indole by the bacterium forms part of the basis for commercial diagnostic

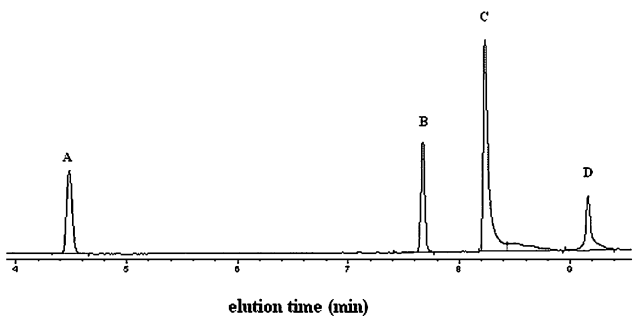


Fig. 4 GC chromatogram of biotransformation mixture of phenylmethylsulfide **4** using growing cells of *E. coli* B834 (DE3) pDB3 at 21 h biotransformation (A = phenylmethylsulfide; B = indole; C = phenylmethylsulfoxide; D = phenylmethylsulfone).

kits for the presence of *E. coli.* (*e.g.* Bactident® at http://service.merck.de/microbiology/tedisdata/prods). Indole has also been suggested to be a cell-signalling chemical with a role in quorum sensing in *E. coli*,[22] and to induce the expression of a number of xenobiotic transporter systems in this bacterium.[23] The production of indole in recombinant bacterial biotransformation systems used for whole-cell biotransformation has been noted before,[24] and it may have implications for downstream processing of large scale biotransformation product mixtures using such systems, or where indole is used an assay substrate for studies on oxidases where oxidation of added indole forms the basis for the detection or quantification of activity.[25]

Comparison of preparative biotransformations of bicyclo[3.2.0]hept-2-en-6-one using growing or rest cells induced with IPTG

Recombinants strains *E. coli* B834 (DE3) pDB3, *E. coli* B834 (DE3) pDB5 and *E. coli* B834 (DE3) pDB6 were assessed for their ability to transform bicyclo[3.2.0]hept-2-en-6-one using both growing and resting cell suspensions. The enantiomeric excesses of residual ketone were recorded (Table 3). With growing cells, *E. coli* B834 (DE3) pDB3 and *E. coli* B834 (DE3) pDB6 were observed to catalyse the required transformation somewhat poorly, although chiral GC analysis of the residual ketones revealed a preference for both gene products to transform the (1*S*, 5*R*)-(−)-enantiomer. The same was also true for *E. coli* B834 (DE3) pDB5, expressing Rv3049c, but in this case, rates of biotransformation were much higher, and the (1*R*, 5*S*)-(+)-ketone was resolved with an enantiomeric excess of >95%. The enantioselectivities of growing and resting cells for strains containing pDB3 and pDB5 were comparable, but resting cell transformations performed with pDB6 showed poor activity (Table 3). Resting cell reactions were easier to work up and extract, but required more catalyst per unit of substrate transformed, and of course required centrifugation and washing prior to use. It was therefore decided to attempt a preparative resolution of racemic **2** using a growing-cell fermentation, but with a view to keeping the reaction time as short as possible, to avoid the accumulation of indole in the medium.

A preparative resolution of **2** using *E. coli* B834 (DE3) pDB5, expressing Rv3049 in a fermenter was attempted. Using the same medium, growth, induction and substrate addition protocols as for shake-flask reactions, a fermenter-contained biotransformation of 3 g substrate at 0.75 g L^{-1} resulted in insufficient conversion for resolution (approximately 35%), yielding (1*R*, 5*S*)-(+)-**2** in 64% ee, and over a 22 h time period, during which a significant amount of indole was also produced. However, using a different protocol in which 1% glucose was added to the LB medium, the cells were grown to an optical density A_{600} of 1.8, induced and substrate added after 45 min growth. In this instance, the greater amount of biomass generated greatly increased the rate of reaction. Chiral GC showed that the (1*S*, 5*R*)-(−)-enantiomer of **2** was not detectable after 4 h. (Fig. 5). Work up and chromatography yielded a 19% yield of (1*R*, 5*S*)-(+)-**2** with an enantiomeric excess of >95%. The relatively poor yield may be attributed to losses during the extraction of a large aqueous volume and the consequent loss of volatile material during solvent evaporation. However, the rate and ease of the biotransformation process make us confident of increasing mass per unit volume yields for this biotransformation, leading to greater recovery of single enantiomer ketone.

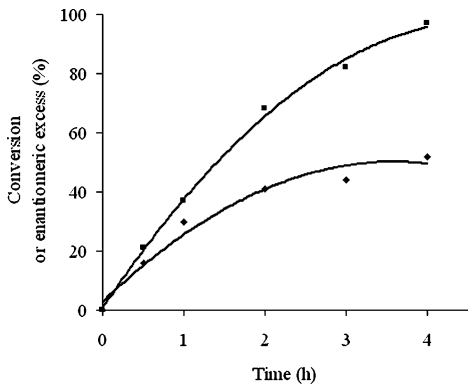


Fig. 5 Fermentation of bicyclo[3.2.0]hept-2-en-6-one by *E. coli* B834 (DE3) pDB5, expressing Rv3049. (= conversion measured by GC; ■ = enantiomeric excess of (1*R*, 5*S*)-(+)-**2**.

Whilst being far from an exhaustive assessment of the substrate specificity of these gene products, it is nevertheless noteworthy that these homologues, which share less than 30% sequence identity, should display the same enantiopreference in the same organism, where one might surmise that such an organism might be equipped with varied profile of activities to metabolise the range of carbon sources that may become available. The resolution of bicyclo[3.2.0]hept-2-en-6-one substrates remains a focus of research after a number of years, as homochiral **2** is an excellent starting material for the synthesis of prostanoids[26] and ligands for transition metal-catalysed organic synthesis.[27,28] Biocatalytic routes to enantio-enriched **2** and derivatives have involved asymmetric reduction using dehydrogenases[29] and lipase-catalysed resolution of the corresponding racemic alcohol.[30] The recombinant whole-cell approach to the resolution of **2**, described herein, is independent of exogenous nicotinamide cofactors. More

Table 3 Resolution of bicyclo[3.2.0]hept-2-en-6-one by both growing and resting cell preparations of *E. coli* B834 (DE3) transformed with pDB3, pDB5 and pDB6

Strain	% Conversion by GC	% Enantiomeric excess (1*R*, 5*S*)-(+) 2
Growing cells		
E. coli B834 (DE3) pDB3	31	39
E. coli B834 (DE3) pDB5	54	>99
E. coli B834 (DE3) pDB6	18	17
Resting cells		
E. coli B834 (DE3) pDB3	15	16
E. coli B834 (DE3) pDB5	49	93
E. coli B834 (DE3) pDB6	3	~0

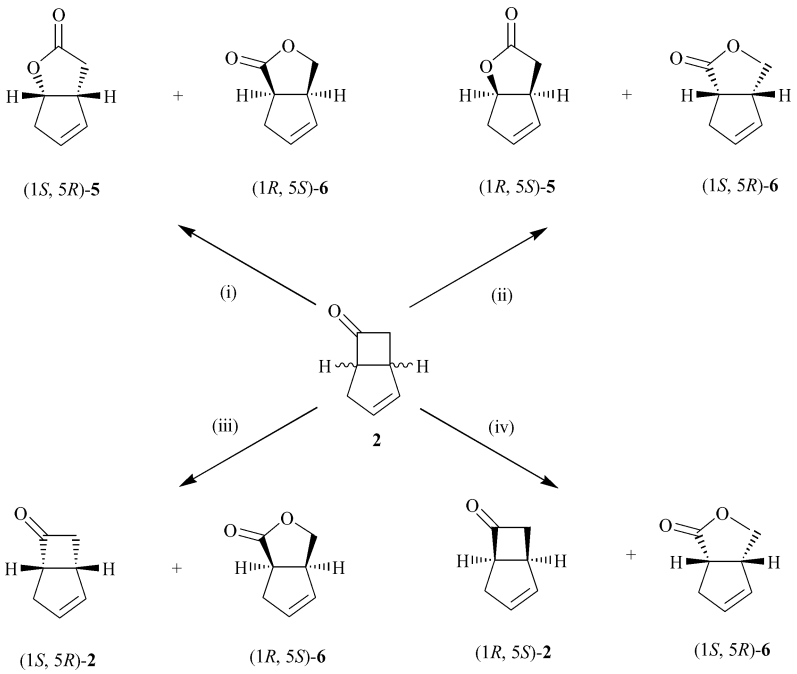


Fig. 6 Stereochemical outcomes of enantiodivergent and enantioselective biotransformations of bicyclo[3.2.0]hept-2-en-6-one by Baeyer–Villigerase catalysts showing major lactone products: (i) cyclohexanone monooxygenase from *Acinetobacter calcoaceticus* NCIMB 9871; (ii) 2,5,-diketocamphane 1,2-monooxygenase from *Pseudomonas putida* NCIMB 10007; (iii) whole cells of *Cunninghamella echinulata* NRRL 3655; HAPMO (hydroxyacetophenone monooxygenase) from *Pseudomonas fluorescens* (iv) Gene product of Rv3049c from *Mycobacterium tuberculosis* H37Rv.

noteworthy however, is the enantioselectivity of reaction in the current study. The cyclohexanone monooxygenase from *Acinetobacter calcoaceticus* NCIMB 9871, and engineered strains of *E. coli* expressing the CHMO gene, catalyse the 'enantiodivergent' biotransformation of **2** such that each enantiomer of the racemic starting material is transformed to a different regioisomeric lactone; each of near optical purity[31] (Fig. 6). A similar, if stereocomplementary reaction is catalysed by the 2,5-diketocamphane, 1,2-monooxygenase from camphor-grown *Pseudomonas putida* NCIMB 10007.[13] As such, these bacterial processes are not useful for the preparation of large amounts of the homochiral ketone. The filamentous fungus *Cunninghamella echinulata* reported to catalyse the resolution of **2**,[32] but our recent studies have shown that biotransformation of other ketones in the bicyclo[3.2.0]hept-2-en-6-on series by this system results in complex mixtures of products as a result of other oxidases and reductases in the fungus.[14] Hydroxyacetophenone monooxygenase (HAPMO) from *Pseudomonas fluorescens*, also exhibits an enantiopreference for the transformation of (1*R*, 5*S*)-(+)-**2**, leading to resolution of (1*S*, 5*R*)-(−)-**2**, but with only moderate enantioselectivity.[33]

The molecular determinants of the enantioselectivity of Baeyer–Villigerases remain unknown as the structure of the *Acinetobacter* enzyme or other enantioselective Baeyer–Villigerase in complex with a ketone or thioether substrate has not yet been determined, although a space-filling model based on numerous biotransformation results using AcCHMO has been published,[34] and random mutagenesis of the AcCHMO gene has created mutant enzymes of altered enantioselectivity.[1] The first structure of a Baeyer–Villigerase, reported recently,[35] from the thermophile *Thermobifida fusca* suggests that progress in analyzing the molecular determinants of selectivity in Baeyer–Villigerases may progress in the near future, thus facilitating rational redesign of these fascinating and useful enzymes.

Conclusion

In summary, six genes from *Mycobacterium tuberculosis* H37Rv thought to encode Baeyer–Villigerases were cloned and rapid expression and biotransformation experiments led to the isolation of a recombinant strain of *Escherichia coli* B834 (DE3), expressing the MTb gene Rv3049c that was competent for the target reaction—the resolution, by enantiomer-selective Baeyer–Villiger reaction, of the important synthetic intermediate (1*R*, 5*S*)-(+)-bicyclo[3.2.0[hept-2-en-6-one. We are currently carrying out detailed studies of the substrate specificity and solution biochemistry on the Rv3049c gene product and using the resolution described above to generate intermediates for further synthetic elaboration.

The study reported herein has demonstrated the application of a rapid throughput ligation independent cloning-expression strategy for creating and assessing a family of recombinant biocatalysts derived from the genomic DNA of a sequenced organism. We envisage that this system would be applicable to the high-throughput generation of large numbers of novel recombinant biocatalysts of other reaction types, allowing rapid access to the diversity of naturally evolved enzyme specificities for biocatalyst discovery programmes.

Experimental

General

Chemicals were purchased from Aldrich Chemical Company Poole, Dorset, UK Racemic **2** was purchased from Fluka.

Bicyclo[3.2.0]hept-2-en-6-one was purchased from Fluka and distilled prior to use in fermentation reactions. All restriction enzymes and corresponding buffers were purchased from New England Biolabs (NEB). Oligonucleotides were purchased from Invitrogen or from MWG Biotech AG (MWG). LIC qualified T4 DNA polymerase (T4 pol) and KOD Hot Start DNA polymerase (KOD) were from Novagen. Pfu Turbo C_x Hot Start was from Stratagene. *E. coli* strain BL21-Gold (DE3) were purchased from Stratagene and all others from Novagen.

Ligation independent cloning protocol

The engineering, construction and sequence of the pET-YSBLIC vector from pET28a3C (Novagen) will be published elsewhere. Pure pET-YSBLIC vector was linearised by incubation at 37 °C for 110 min in a 1 mL restriction enzyme digest reaction, that consisted of 50 μg pET-YSBLIC, 200U *Bse*RI, 100 μL NEB Buffer 2 and water to 1 mL. The linearised pET-YSBLIC was run out on an agarose gel and purified using a QIAquick gel extraction kit (Qiagen). In order to generate single stranded overhangs in the linearised vector, a T4 pol LIC preparation reaction was set up that contained 4.0 pmol of linearised pET-YSBLIC, 20U of T4 DNA polymerase, 40 μL of reaction buffer, 2.5 mM dTTP, 5 mM DTT and water to a final volume of 400 μL. The reaction was mixed, divided into four 100 μL aliquots and these were incubated in a thermal cycler for 30 min at 22 °C then for 20 min at 75 °C. The four reactions were recombined and the linearised vector was then purified using a QIAquick PCR purification kit (Qiagen). The DNA concentration was determined using an Eppendorf Biophotometer, thence adjusted to 50 ng μL^{-1} and stored at −20 °C until required.

Ligation-independent cloning: PCR of six genes encoding BVases from *Mycobacterium tuberculosis* H37Rv

PCR primers were designed using the resource at http://genome~ww2.stanford.edu/cgi~bin/SGD/web~primer. LIC-specific ends were added (Forward: CACCACCACCAC; Reverse: GAGGAGAAGGCGCGTTA) to each gene specific primer. The final primers used for each gene amplification are listed in Table 4. PCR reactions were found to be template/product specific and the successful conditions for each of the six genes are listed in Table 1. Usual PCR mixture consisted of template DNA (50 ng μL^{-1}) 0.5 μL; forward primer 0.4 μM; reverse primer 0.4 μM; 25 mM $MgSO_4$ 2 μL; 2 mM deoxynucleotide triphosphate mix 5 μL; KOD buffer 5 μL; KOD 1 μL and sterile deionised water to a final volume of 50 μL. Extra additions are detailed in Table 1.

PCR products were cleaned prior to LIC reaction using the PCR clean-up kit (Qiagen). The insert reaction contained: 0.2 pmol PCR product; 2 μL T4 pol buffer; 2 μL 25 mM stock dATP; 1 μL 100 mM stock dithiothreitol; 0.4 μL T4 pol; water to a final volume of 20 μL. The reaction was mixed and incubated at 22 °C for 30 min and stopped with further incubation at 75 °C for 20 min. 2 μL of the reaction mixture was added to 1 μL of pET-YSBLIC prepared vector (50 ng μL^{-1}) and the reaction incubated at room temperature for 10 min. 1 μL of 100 mM EDTA was added to give a final volume of 4 μL, the reaction mixed and left at room temperature for 10 min. A control reaction, in which the LIC reaction was substituted by 2 μL water, was subjected to identical conditions.

2 μL of LIC annealing reaction was added to 50 μL of NovaBlue singles competent cells. The cells were incubated on ice for 5 min then heat shocked at 42 °C for 30 s and again incubated on ice for 5 min. 100 μL of SOC medium was added and the cells incubated at 37 °C with shaking for 60 min. 150 μL of the reaction mixture was plated out onto LB agar containing 30 μg mL^{-1} kanamycin and the plates incubated overnight at 37 °C. Colonies were used to prepare stocks of plasmids for transformation into expression strains using the Wizard SVMiniprep kit (Promega).

Expression in *Escherichia coli* B834 (DE3)

A single colony of *E. coli* B834 (DE3) transformed with the relevant LIC plasmid construct was used to inoculate 2 mL of LB broth with 30 μg mL^{-1} kanamycin. This was grown at 37 °C overnight. 50 μL of this starter culture was used to inoculate 5 mL of LB broth with 30 μg mL^{-1} kanamycin and the culture grown until it had reached an optical density (A_{600}) of 0.5. At this stage, 1 mM isopropylthiogalactopyranoside (IPTG) was added, and growth continued at 20 °C overnight. The cells were harvested by centrifugation and resuspended in 0.5 mL of 50 mM Tris-HCl buffer pH 7.1, containing 20 μM phenylmethylsulfonylfluoride and 1 mM dithiothreitol (henceforth referred to as buffer). The total cell extracts were analysed by SDS-PAGE using established protocols.

Biotransformations of test substrates

A single colony of *E. coli* B834 (DE3) transformed with the relevant LIC plasmid construct was used to inoculate 2 mL of LB broth with 30 μg mL^{-1} kanamycin. This was grown at 37 °C overnight. 50 μL of this starter culture was used to inoculate 5 mL of LB broth with 30 μg mL^{-1} kanamycin and the culture grown until it had reached an optical density (A_{600}) of 0.5. At this stage, 1 mM IPTG was added, and growth continued at 30 °C for 45 min. Substrate (2.5 mg) in ethanol (50 μL) was then added and the progress of the biotransformation reaction monitored by sampling 0.5 mL of the reaction mixture, extracting with ethyl acetate

Table 4 PCR primers used for amplification of genes encoding putative BVases from genomic and cosmid DNA of *Mycobacterium tuberculosis* H37Rv

ORF	Forward primer	Reverse primer
Rv0892	CACCACCACCACATGACCGGGCGATGTCCGACGGTTG	GAGGAGAAGGCGCGTTATCAAGCGCTTTGAGGCCGACTAGTT
Rv0565c	CACCACCACCACATGAGCGTGACTCCAAACGCCGGCT	GAGGAGAAGGCGCGTTATCATGCCGCGCCGAACACCATTGCCT
Rv3854c	CACCACCACCACATGACCGAGCACCTCGACGTTGTCA	GAGGAGAAGGCGCGTTACTAAACCCCCACCGGGGCAGGCCTTT
Rv1393c	CACCACCACCACATGATGCCCGACTACCACGCACTGAT	GAGGAGAAGGCGCGTTACTAGCTGCTGATGCGGTAGTCGCCG
Rv3049c	CACCACCACCACATGAGCATTGCCGATACGGCTGCCAA	GAGGAGAAGGCGCGTTACTAGTTCGCCGCAGCGACCGTCGCG
Rv3083	CACCACCACCACATGAACCAGCATTTCGACGTCCTGA	GAGGAGAAGGCGCGTTATCAGGCCGCCGCGTGGTCTTGGGCC

(0.5 mL) and analysis of the organic extract by gas chromatography.

For resting cell biotransformations of **2**, a single colony of *E. coli* B834 (DE3) transformed with the relevant LIC plasmid construct was used to inoculate 5 mL of LB broth with 30 μg mL^{-1} kanamycin and the culture grown at 37 °C overnight. This was used to inoculate 500 mL LB broth with 30 μg mL^{-1} kanamycin and this large culture grown at 37 °C until it had reached an optical density (A_{600}) of 0.5. At this stage, 1 mM IPTG was added, and growth continued overnight at 20 °C. The cells were then harvested by centrifugation and washed once with 100 mL buffer. The cell pellet was then resuspended in 50 mL buffer and substrate (25 mg) in ethanol (0.5 mL) was added. The progress of the biotransformation reaction was monitored by sampling 0.5 mL of the reaction mixture, extracting with ethyl acetate (0.5 mL) and analysis of the organic extract by gas chromatography. When the reaction had reached approximately 50% conversion as determined by GC, or had reached an apparent plateau, the cells were removed by centrifugation and the supernatant extracted with 3 × 50 ml ethyl acetate. The combined organic fractions were dried over anhydrous magnesium sulfate, filtered and the solvent removed *in vacuo.* The residual ketone and product lactone were separated by flash chromatography using a gradient of 0–40% ethyl acetate in petroleum ether as the eluent.

Biotransformation of racemic bicyclo[3.2.0]hept-2-en-6-one in a fermenter with *E. coli* B834 (DE3) pDB5 as the catalyst

A single colony of *E. coli* B834 (DE3) pDB5 was used to inoculate a 10 mL starter culture of LB broth with 30 μg mL^{-1} kanamycin, which was then grown at 37 °C overnight. This was then used to inoculate 4 L of the same broth, to which had been added 1% glucose, in a 6 L operating volume glass autoclaveable fermentor from Applikon, fitted with a ADI1032 stirrer-controller and an AD1010 biocontroller. The culture was grown at 37 °C with a dissolved oxygen control loop determining the impeller speed. A dissolved oxygen level of 60% was maintained throughout the biotransformation, at an air flow rate of 2 L min^{-1}. Sodium hydroxide (2 M) was added throughout growth and biotransformation to keep the operating pH at 7.5. After the culture had reached an optical density A_{600} of 1.8, 1 mM IPTG was added, and growth continued for 45 min at 37 °C. Substrate (3 g, 0.028 mol, made up to 30 mL with absolute ethanol) was then added *via* syringe to the fermenter, and the biotransformation monitored using chiral GC. When the (1*S*, 5*R*)-(−)-2 ketone was no longer detectable, the culture was rapidly removed from the fermenter, the cells removed by centrifugation, and the supernatant extracted into ethyl acetate. Work up and chromatography were performed as described above, yielding 570 mg (0.00532 mol, 19%) of (1*R*, 5*S*)-(+)-**2**. The identity of the product lactone **6** was confirmed by ^{1}H NMR, which was in agreement with spectroscopic data in the literature.[36]

Gas chromatography

Non-chiral GC was performed on an Agilent 6890 gas chromatograph fitted with an HP5 column (30 m × 0.25 mm × 25 μm); inlet temperature 250 °C; detector temperature 320 °C; column temperature for analysis of biotransformations of substrates **1**, **2**, **3**, **4**: 100 °C for 3 min followed by a gradient of 100–200 °C at 5 °C min^{-1}. Chiral GC analysis was performed on the same instrument fitted with an Agilent Cyclosil B column; column temperature for analysis of ketone **2**, 100–200 °C at 5 °C min^{-1}. (1*S*, 5*R*)-(−)-**2**, 10.5 min; (1*R*, 5*S*)-(+)-**2**, 11.4 min.

Acknowledgements

The authors would like to thank the Nuffield foundation for a bursary award to D.B. and the Innovation, Research and Priming Fund at the University of York for an award to G.G. The work was part funded by the European Commission as SPINE (Structural Proteomics IN Europe), contract no. QLG2-CT-2002-00988 under the RTD programme "Quality of Life and Management of Living Resources". We also gratefully acknowledge Dr Stefano Benini and Dr Nadine Honore, of the Institut Pasteur, Paris, France, for the generous gift of relevant cosmid and genomic DNA from *Mycobacterium tuberculosis* H37Rv. We also thank Dr Jared Cartwright for assistance with fermentation techniques.

References

1 M. T. Reetz, B. Brunner, T. Schneider, F. Schulz, C. M. Clouthier and M. M. Kayser, *Angew. Chem., Int. Ed.*, 2004, **43**, 4075.
2 R. Carr, M. Alexeeva, A. Enright, T. S. C. Eve, M. J. Dawson and N. J. Turner, *Angew. Chem., Int. Ed.*, 2003, **42**, 4807.
3 L. S. Zhao, B. Han, Z. L. Huang, M. Miller, H. J. Huang, D. S. Malashock, Z. L. Zhu, A. Milan, D. E. Robertson, D. P. Weiner and M. J. Burk, *J. Am. Chem. Soc.*, 2004, **126**, 11156.
4 W. A. Kaluzna, T. Matsuda, A. K. Sewell and J. D. Stewart, *J. Am. Chem. Soc.*, 2004, **126**, 12827.
5 C. Aslandis and P. J. de Jong, *Nucl. Acids Res.*, 1990, **18**, 6069.
6 C. Li and R. M. Evans, *Nucl. Acids Res.*, 1997, **25**, 4165.
7 J. D. Stewart, K. W. Reed, J. Zhu, G. Chen and M. M. Kayser, *J. Org. Chem.*, 1996, **61**, 7652.
8 M. D. Mihovilovic, F. Rudroff, W. Kandioller, B. Grotzl, P. Stanetty and H. Spreitzer, *SYNLETT*, 2003, 1973.
9 H. Iwaki, Y. Hasegawa, S. Z. Wang, M. M. Kayser and P. C. K. Lau, *Appl. Environ. Microbiol.*, 2002, **68**, 5671.
10 M. D. Mihovilovic and P. Kapitan, *Tetrahedron Lett.*, 2004, **45**, 2751.
11 B. G. Kyte, P. Rouviere, Q. Cheng and J. D. Stewart, *J. Org. Chem.*, 2004, **67**, 12.
12 M. D. Mihovilovic, G. Chen, S. Z. Wang, B. Kyte, F. Rochon, M. M. Kayser and J. D. Stewart, *J. Org. Chem.*, 2001, **66**, 733.
13 R. Gagnon, G. Grogan, M. S. Levitt, S. M. Roberts, P. W. H. Wan and A. J. Willetts, *J. Chem. Soc., Perkin Trans. 1*, 1994, 2537.
14 I. J. S. Fairlamb, S. Grant, G. Grogan, D. A. Maddrell and J. C. Nichols, *Org. Biomol. Chem.*, 2004, **2**, 1831.
15 S. T. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M. A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead and B. G. Barrell, *Nature*, 1998, **393**, 537.
16 Y. C. J. Chen, O. P. Peoples and C. T. Walsh, *J. Bacteriol.*, 1988, **170**, 781.
17 T. A. Vannelli, A. Dykman and P.R. Ortiz de Montellano, *J. Biol. Chem.*, 2002, **277**, 12824.
18 M. W. Fraaije, N. M. Kamerbeek, A. J. Heidekamp, R. Fortin and D. B. Janssen, *J. Biol. Chem.*, 2004, **279**, 3354.
19 M. W. Fraaije, N. M. Kamerbeek, W. J. H. van Berkel and D. B. Janssen, *FEBS Lett.*, 2002, **518**, 43.
20 J. D. Thompson, D. G. Higgins and T. J. Gibson, *Nucleic Acids Res.*, 1994, **11**, 4673.
21 E. E. Snell, *Adv. Enzymol. Relat. Areas Mol. Biol.*, 1975, **42**, 287.
22 D. Wang, X. Ding and P. N. Rather, *J. Bacteriol.*, 2001, **183**, 4210.
23 H. Hirakawa, Y. Inazumi, T. Masaki, T. Hirata and A. Yamaguchi, *Mol. Microbiol.*, 2005, **55**, 1113.

24 E. M. Gillam, A. M. Aguinaldo, L. M. Notley, D. Kim, R. G. Mundkowski, A. A. Volkov, F. H. Arnold, P. Soucek, J. J. DeVoss and F. P. Guengerich, *Biochem. Biophys Res. Commun.*, 1999, **265**, 469.
25 Q. S. Li, J. Ogawa, R. D. Schmid and S. Shimizu, *Biosci., Biotechnol., Biochem.*, 2005, 293.
26 L. T. Boulton, D. Brick, M. E. Fox, M. Jackson, I. C. Lennon, R. McCague, N. Parkin, D. Rhodes and G. Ruecroft, *Org. Process Res. Dev.*, 1999, **6**, 138.
27 C. B. Dousson and S. M. Roberts, *ARKIVOC*, 2000, **1**, 882.
28 I. J. S. Fairlamb, S. Grant, A. C. Whitwood, J. Whitthall, A. S. Batsanov and J. C. Collings, 2005, *J. Organomet. Chem.*, **690**, 4462.
29 S. Butt, H. G. Davies, M. J. Dawson, G. C. Lawrence, J. Leaver, S. M. Roberts, M. K. Turner, B. J. Wakefield, W. F. Wall and J. A. Winders, *J. Chem. Soc., Perkin Trans. 1*, 1987, 903.
30 I. C. Cotterill, E. L. A. MacFarlane and S. M. Roberts, *J. Chem. Soc., Perkin Trans. 1*, 1988, 3387.
31 V. Alphand, A. Archelas and R. Furstoss, *Tetrahedron Lett.*, 1989, **30**, 3633.
32 J. Lebreton, V. Alphand and R. Furstoss, *Tetrahedron*, 1997, **53**, 145.
33 N. M. Kamerbeek, A. J. J. Olsthoorn, M. W. Fraaije and D. B. Janssen, *Appl. Environ. Microbiol.*, 2003, **69**, 419.
34 G. Ottolina, G. Carrea, S. Colonna and A. Rückemann, *Tetrahedron: Asymmetry*, 1996, **7**, 1123.
35 E. Malito, A. Alfieri, M. W. Fraaije and A. Mattevi, *Proc. Natl Acad. Sci. USA*, 2004, **101**, 13157.
36 T. Hudlicky, D. Bhaskar Reddy, S. V. Govindan, T. Kulp, B. Still and J. P. Sheth, *J. Org. Chem.*, 1983, **48**, 3422.

