Characterisation of CYP101J2, CYP101J3 and CYP101J4, three 1,8-cineole-hydroxylating cytochrome P450 monooxygenases from *Sphingobium yanoikuyae* strain B2

Birgit Unterweger[a,b,c], Dieter M. Bulach[a,d], Judith Scoble[e], David J. Midgley[f], Paul Greenfield[g], Dena Lyras[a,b], Priscilla Johanesen[a,b], Geoffrey J. Dumsday[c]#

Department of Microbiology, Monash University, Clayton, VIC, Australia[a]; Infection and Immunity Program, Monash Biomedicine Discovery Institute, Monash University, Clayton, VIC, Australia[b]; CSIRO Manufacturing, Clayton, VIC, Australia[c]; Victorian Life Sciences Computation Initiative, The University of Melbourne, Carlton, VIC, Australia[d]; CSIRO Manufacturing, Parkville, VIC, Australia[e]; CSIRO Energy, North Ryde, NSW, Australia[f]; CSIRO Oceans & Atmosphere, North Ryde, NSW, Australia[g]

Running head: 1,8-Cineole-hydroxylating P450s from *S. yanoikuyae* B2

#Address correspondence to Geoffrey J. Dumsday, geoff.dumsday@csiro.au

ABSTRACT

We report the isolation and characterisation of three new cytochrome P450 monooxygenases: CYP101J2, CYP101J3 and CYP101J4. These P450s were derived from *Sphingobium yanoikuyae* B2, a strain that was isolated from activated sludge based on its ability to fully mineralise 1,8-cineole. Genome sequencing of this strain in combination with purification of native 1,8-cineole-binding proteins enabled identification of 1,8-cineole-binding P450s. The P450 enzymes were cloned, heterologously expressed (N-terminally-His_6-tagged) in *Escherichia coli* BL21 (DE3), purified and spectroscopically characterised. Recombinant whole-cell biotransformation in *E. coli* demonstrated that all three P450s hydroxylate 1,8-cineole using electron transport partners from *E. coli* to yield a product putatively identified as (1*S*)-2α-hydroxy-1,8-cineole or (1*R*)-6α-hydroxy-1,8-cineole. The new P450s belong to the CYP101 family and share 47% and 44% identity with other 1,8-cineole-hydroxylating members found in *Novospingobium aromaticivorans* and *Pseudomonas putida*. Compared to $P450_{cin}$ (CYP176A1), a 1,8-cineole-hydroxylating P450 from *Citrobacter braakii*, these enzymes share less than 30% amino acid sequence identity and hydroxylate 1,8-cineole in a different orientation. Expansion of the enzyme toolbox for modification of 1,8-cineole creates a starting point for use of hydroxylated derivatives in a range of industrial applications.

38 **IMPORTANCE**

39 CYP101J2, CYP101J3 and CYP101J4 are cytochrome P450 monooxygenases from
40 *S. yanoikuyae* B2 that hydroxylate the monoterpenoid 1,8-cineole. These enzymes
41 not only play an important role in microbial degradation of this plant-based chemical
42 but also provide an interesting route to synthesise oxygenated 1,8-cineole-
43 derivatives for applications as natural flavour and fragrance precursors or
44 incorporation into polymers. The cytochromes P450 also provide an interesting basis
45 from which to compare other enzymes with a similar function and expand the
46 CYP101 family. This could eventually provide enough bacterial parental enzymes
47 with similar amino acid sequences to enable *in vitro* evolution via DNA shuffling.

INTRODUCTION

The bicyclic monoterpenoid 1,8-cineole (1,3,3-trimethyl-2-oxabicyclo[2,2,2]octane) is a chemically stable saturated ditertiary ether (1) that is an abundant natural resource. It is a major component of many essential oils, in particular *Eucalyptus* oil (2). Whilst 1,8-cineole has its main applications in pharmaceutical, fragrance, food and cleaning industries, its oxidation provides a potential route to value-added chemicals. Hydroxylated 1,8-cineole derivatives are chiral compounds that can serve as building blocks and precursors with uses in organic chemistry (3). Hydroxylation of 1,8-cineole can be achieved chemically, however, these reactions often use a range of environmentally hazardous reagents and/or yield product mixtures (4-6). On the other hand, biooxygenation of 1,8-cineole using enzymes or whole cells can be performed under mild reaction conditions with the potential to yield products with high enantiopurity (6, 7).

A potential route to bio-derived 1,8-cineole derivatives is the utilisation of enzymes involved in bacterial 1,8-cineole metabolism. Bacterial 1,8-cineole metabolism was first examined in the late 1970s, when the isolation of a *Pseudomonas flava* strain capable of growing on 1,8-cineole as a sole source of carbon was reported (8). Subsequently, 1,8-cineole biohydroxylation has also been observed in several other bacterial species including *Rhodococcus* C1 (9), *Bacillus cereus* (10), *C. braakii* (11), *Novospingobium subterranea* (12) and a *Sphingomonas* sp. (13). The first bacterial 1,8-cineole-hydroxylating enzyme which was isolated and characterised was isolated from *C. braakii* (11). The initial oxidation of 1,8-cineole by this Gram-negative bacterium is catalysed by a P450 designated $P450_{cin}$ (CYP176A1) and with the aid of suitable electron transport partners yielded (1*R*)-6*β*-hydroxy-1,8-cineole (7, 11).

More recently, other P450s including CYP101A1 and CYP101B1 and the N252A $P450_{cin}$ mutant were shown to hydroxylate 1,8-cineole and yield a range of hydroxy-1,8-cineole derivatives (14, 15).

Cytochromes P450 are diverse oxidative hemoproteins that catalyse an enormous range of reactions including hydroxylations of unactivated carbon-hydrogen bonds (16). These enzymes are widespread in nature originating from all kingdoms of life, including more than 2150 bacterial P450s (http://drnelson.uthsc.edu/CytochromeP450.html) (17). Several bacterial P450s are involved in monoterpenoid oxidation, including $P450_{cam}$ from *P. putida* which was enriched by selection for growth on (1*R*)-(+)-camphor (18). $P450_{cam}$ or CYP101A1 was the first P450 crystallised (19) and has served as a model system for mechanistic studies and structure-function relationships. Since then many more monoterpene-oxidising P450s have been isolated, characterised, crystallised and used for laboratory-scale biooxidation reactions; the latest research in biooxidation of monoterpenes using bacterial monooxygenases has been recently reviewed (20).

Bacterial P450s (and their electron transport partners) are usually soluble proteins (21). For this reason, these proteins are easier to overexpress and purify compared to membrane-bound P450s of eukaryotic origin (20). They are particularly promising candidates for monoterpene biohydroxylations because of their relatively high activity and substrate specificity towards a number of monoterpenes (11, 20). With the recent availability of extensive metagenomic data, the number of gene sequences potentially encoding P450s has increased substantially (22), however, many of the

identified genes remain uncharacterised. The isolation and detailed characterisation of enzymes is a prerequisite for optimisation and use in biocatalytic reactions paving the way for applications in synthetic biology.

To identify and characterise new bacterial 1,8-cineole-hydroxylating P450s we sequenced the genome of *S. yanoikuyae* strain B2 isolated from activated sludge (23) that was capable of using 1,8-cineole as the sole source of carbon and energy. Purification of native 1,8-cineole-binding proteins from *S. yanoikuyae* strain B2 followed by recombinant expression and enzyme characterisation aided the identification of three new P450s capable of hydroxylating 1,8-cineole. This work has expanded the range of bacterial 1,8-cineole-oxidising enzymes, creating an opportunity for their exploitation in industrial biocatalysis and further understanding of enzyme mechanisms.

MATERIALS AND METHODS

Reagents

1,8-Cineole (99.5 % purity) was obtained from Felton Grimwade & Bosisto's Pty Ltd (Oakleigh South, Australia). Isopropyl-β-D-thiogalactopyranoside (IPTG), 2-adamantanone, β-ionone, α-terpineol, (1*R*)-(+)-camphor, (1*S*)-(–)-camphor and sodium hydrosulphite (85% purity) were from Sigma/Aldrich. (1*R*)-6β-Hydroxy-1,8-cineole was produced using *E. coli* strain DH5α harbouring the bicistronic plasmid pCW-P450$_{cin}$/Cdx (7) which was a gift from Professor James De Voss (University of Queensland). A mixture of chemically synthesised (1*S*)-2α-hydroxy-1,8-cineole and (1*S*)-2β-hydroxy-1,8-cineole was kindly provided by Stella Kyi (CSIRO). Restriction enzymes for molecular biology and Phusion High Fidelity DNA Polymerase for PCRs were from New England Biolabs. General reagents for media and buffer preparation were obtained from various commercial sources and were of reagent grade or better.

Strains and plasmid vector

Microorganisms capable of metabolising 1,8-cineole were isolated from activated sludge (obtained from Lilydale Sewage Treatment Plant, Victoria, Australia) using a modified continuous culture method (24) in combination with selection based on growth with 1,8-cineole as the sole source of carbon and energy (12). One isolate, *S. yanoikuyae* strain B2, was chosen for this study because of its ability to grow in relatively high concentrations of 1,8-cineole of up to 2 g L^{-1} (23). The *S. yanoikuyae* type strain DSM 7462 (= ATCC 51230) was obtained from the Leibniz Institute DSMZ - German Collection of Microorganisms and Cell Cultures GmbH. Unless

stated otherwise, *E. coli* strain BL21 (DE3) was used for protein expression and *E. coli* strain DH5α was used for cloning. The pET28a(+) vector was from Novagen.

Media

Liquid media used in this study were defined medium (DM; 10.64 g L^{-1} KH_2PO_4, 4 g L^{-1} $(NH_4)_2HPO_4$, 1.7 g L^{-1} citric acid monohydrate, 0.62 g L^{-1} $MgSO_4.7H_2O$ and 4.4 mg L^{-1} thiamine hydrochloride supplemented with 5 mL L^{-1} trace salts solution containing 2.0 g L^{-1} $Cu_2SO_4.5H_2O$, 0.08 g L^{-1} NaI, 3.0 g L^{-1} $Mn_2SO_4.H_2O$, 0.2 g L^{-1} $Na_2MoO_4.2H_2O$, 0.02 g L^{-1} boric acid, 0.5 g L^{-1} $CoCl_2.6H_2O$, 7.0 g L^{-1} $ZnCl_2$, 22.0 g L^{-1} $FeSO_4.7H_2O$, 0.5 g L^{-1} $CaSO_4.2H_2O$ and 1 mL L^{-1} H_2SO_4; DM was adjusted to pH 7.0 as required), DM 2 (461S) (12), terrific broth (TB; 12 g L^{-1} tryptone, 24 g L^{-1} yeast extract, 4 mL L^{-1} glycerol, 9.4 g L^{-1} K_2HPO_4 and 2.2 g L^{-1} KH_2PO_4) and Miller's lysogeny broth (LB) (25). Carbon sources, antibiotics and other additives were added as specified in the description of each experiment.

Metabolite characterisation

The metabolites produced by *S. yanoikuyae* strain B2 were purified and characterised using techniques described previously (12). Briefly, compounds produced during growth on 1,8-cineole were initially detected using gas chromatography (GC) of filtered culture supernatants. The metabolites were then identified in dichloromethane extracts from 500 mL cultures. Cells were grown in DM2 with 1,8-cineole as the sole carbon source in 2 L baffled flasks shaking at 200

rpm and 30 °C. After purification the compounds in the solvent extracts were identified using gas chromatography-mass spectrometry (GC-MS) and NMR.

Genome sequencing and strain identification

Genomic DNA was extracted from the pellet of a 10 mL culture grown in DM2 containing 0.5 mL L^{-1} 1,8-cineole using the DNeasy Blood and Tissue Kit (Qiagen) according to the manufacturer's instructions. Genomic DNA was prepared for Illumina MiSeq sequencing (University of New South Wales, Ramaciotti Centre) using the Nextera library preparation kit (Illumina). Quality filtering was performed using NESONI (version 0.128; https://github.com/Victorian-Bioinformatics-Consortium/nesoni) to remove low quality base calls and residual adaptor sequences from the read sets. A draft genome sequence was constructed using the VELVET *de novo* genome assembler version 1.2.07 (26); optimised assembly conditions were determined manually by evaluating assembly metrics from assemblies performed over a range of k-mer values. The draft genome sequence was annotated using Prokka version 1.7.1 (27); Prokka predicts both protein coding genes and RNA coding genes, with the function of protein coding regions being inferred by protein similarity to characterised proteins. Isolate identity was determined using the RDP classifier (28) using the partial 16S rRNA gene sequence extracted from the draft strain B2 genome.

Isolation of 1,8-cineole-binding P450s from *S. yanoikuyae* strain B2

177 *S. yanoikuyae* strain B2 cells were produced in a 2 L stirred tank bioreactor (Biostat
178 B, Sartorius Stedim) in fed-batch mode using glucose and 1,8-cineole as carbon
179 sources. Using a 3-step purification procedure, 1,8-cineole-binding P450s were
180 purified from crude cell extracts and identified using N-terminal sequencing and/or
181 matrix-assisted laser desorption ionisation (MALDI)-MS fingerprint mapping.
182 Additional experimental details can be found in the Supplementary Information.

183

184 **Molecular cloning**

185 Genes encoding proteins CYP101J2, CYP101J3 and CYP101J4 (systematic P450
186 names were assigned by Professor David R. Nelson), respectively, were amplified
187 by PCR using genomic DNA from *S. yanoikuyae* strain B2 as template. The following
188 primer pairs were used:
189 $fw_{CYP101J2}$ 5'-GGAATTCCATATGGAAGCAAGCGTGAAGGGC-3'
190 $rev_{CYP101J2}$ 5'-CCGCTCGAGTTAGACTGGCCATTCGAGCAC-3'
191 $fw_{CYP101J3}$ 5'-GGAATTCCATATGGCTGAACAGGCTGTTCTCG-3'
192 $rev_{CYP101J3}$ 5'-CCGCTCGAGTTACTGCCGCGGCCAGC-3'
193 $fw_{CYP101J4}$ 5'-GGAATTCCATATGGAAATCGGAACGATTGAACC-3'
194 $rev_{CYP101J4}$ 5'-CCGCTCGAGTTAGACATCCCATTCGAGCATG-3'
195 Restriction sites NdeI and XhoI introduced for cloning into the multiple cloning site of
196 pET28a(+) are underlined. The thermocycler was preheated to 98 °C. After initial
197 denaturation at 98 °C for 1 min, the genes were amplified by 30 cycles of
198 denaturation at 98 °C for 10 seconds followed by annealing and extension at 72 °C
199 for 40 seconds and a final extension step at 72 °C for 10 min. Original stop codons

(TGA) were swapped to TAA. After ligation into the vector, DNA sequencing was used to confirm that the gene was inserted correctly.

Expression and purification of recombinant P450s

Chemically competent *E. coli* BL21 (DE3) cells were freshly transformed with pET28a(+) harbouring genes encoding CYP101J2, CYP101J3 or CYP101J4, respectively and plated on LB agar containing 50 μg mL^{-1} kanamycin sulphate (Kan). A single transformant was used to inoculate 50 mL of TB (in a 250 mL non-baffled shake flask) containing 50 μg mL^{-1} Kan. Seed cultures were incubated with shaking at 180 rpm for 18 h at 30 °C and then this culture was used to inoculate 500 mL of TB supplemented with 50 μg mL^{-1} Kan in 2 L baffled shake flasks to an OD_{600} of approximately 0.1. When the OD_{600} reached 0.6-0.8, cultures were induced with 1 mM IPTG, grown for a further 18 h and then harvested by centrifugation at 6000 *g* for 20 min at 4 °C. Cell pellets were washed in 50 mM Tris, 150 mM NaCl buffer, pH 7.4 and then stored at -80 °C. Cells were resuspended in the same buffer containing 5 mM imidazole and then disrupted by two passages through an Avestin EmulsiFlex-C5 cell disruptor. After centrifugation, the clarified cell lysate was loaded onto a column with a bed volume of 2-3 mL of cobalt-charged resin (TALON® Superflow™ metal affinity resin; Clontech) equilibrated with 50 mM Tris, 150 mM NaCl and 5 mM imidazole, pH 7.4. After a washing step the His_6-tagged protein was eluted by a stepwise increase of imidazole. Fractions with similar purities as judged by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) were pooled and buffer exchanged into 50 mM Tris, pH 7.4 by repeated cycles of concentration and dilution using Amicon Ultra-15 Centrifugal Filter Units with a molecular weight cut-off

of 10 kDa. Concentrated protein solutions were then snap-frozen using a dry ice-ethanol (EtOH) bath and stored at -80 °C.

P450 functionality

Whole-cell biohydroxylation was achieved by expressing the cytochromes P450 in *E. coli* BL21 (DE3) as described above except for the following changes: Just after induction with 1 mM IPTG, 0.5 mL L^{-1} 1,8-cineole was added to the 500 mL cultures and the culture was incubated for a further 45 h. A 1 mL aliquot of the culture was then extracted with ethyl acetate, dried over Na_2SO_4 and analysed using GC-MS. GC-MS was performed using a Thermo Scientific TSQ8000 GC mass spectrometer using electron impact ionisation in the positive ion mode with an ionisation energy of 70 eV. The gas chromatography was performed with a Thermo Scientific TG-5MT column (15 m x 0.25 mm ID, 0.10 µm film thickness), with a temperature program of 50 °C for 2 min, then heating at 25 °C per min to 300 °C where the temperature was held for 3 min using an injection with a split of 10, an injector temperature of 300 °C and the transfer line was set to 300 °C. High-purity helium was used as carrier gas at a flow rate of 1.0 mL min^{-1}.

Spectroscopic characterisation of P450s

Spectra of oxidised protein solutions were prepared in 50 mM Tris, pH 7.4. 1,8-cineole induced absorbance shifts were demonstrated by adding 1 µL of 1,8-cineole to 1 mL of protein solution. The P450s were then reduced by the addition of a small amount of solid sodium dithionite and incubation for 1 min. Subsequently, carbon

monoxide (CO) was bubbled through the cuvettes for 30 seconds. Spectra were recorded from 350 - 700 nm. The concentration of functional P450 was estimated using CO difference spectroscopy using an extinction coefficient ε = 91 mM^{-1} cm^{-1} (29, 30).

Determination of 1,8-cineole dissociation constants (K_D)

The determination of dissociation constants was performed using a procedure similar to that described by Bell et al. (31). Enzymes were diluted to ca. 2 µM in 50 mM Tris, pH 7.4, or 50 mM Tris, 200 mM KCl, pH 7.4, respectively. Stock solutions of 1,8-cineole in EtOH with a concentration range of 1 - 600 mM were prepared by serial dilution. To 1 mL of protein solution, 1 µL aliquots of 1,8-cineole stock solutions of varying concentrations were added sequentially. After mixing, an absorbance spectrum between 350 and 450 nm was measured. Reference spectra of protein solutions containing EtOH instead of substrate solutions were subtracted from the sample absorbance spectra to obtain 1,8-cineole-induced difference spectra. Further 1 µL aliquots of 1,8-cineole stock solutions were added until the absorbance difference between peak and trough (ΔA) reached its maximum, the maximum total amount of added substrate solution was 1 % (*v/v*) of the sample. K_D values were calculated by fitting the ΔA against the 1,8-cineole concentration ([S]) to the hyberbolic function $\Delta A = (\Delta A_{max} \times [S])/(K_D + [S])$. Alternative substrates were screened for the induction of absorbance shifts by adding solutions of 2-adamantanone, β-ionone, (1*R*)-(+)-camphor, (1*S*)-(–)-camphor, α-terpineol and toluene in EtOH to the protein solutions to a final concentration of 1.2 mM (irrespective of substrate solubility) when the ΔA did not show any significant further

increase. The spin-state shifts corresponding to the ΔA were measured in the absence and presence of 0.2 M KCl for each substrate tested. The spin-state shift with 1,8-cineole and 0.2 M KCl was set at 100% for each P450 and all other type I shifts were compared with this value.

Applied and Environmental Microbiology

Sequence accession numbers

The assembly described in this manuscript is available as "Sphingobium_yanoikuyae_strain_B2_scaffolds.zip" available for download from the CSIRO Data Access Portal (https://data.csiro.au/dap/landingpage?pid=csiro:17150). Annotations for this assembly, along with contiguous sequences, sequences for putative and confirmed cytochromes P450 and raw reads are also available for download at CSIROs Data Access Portal. In addition, contiguous sequences over 1000 bp were also uploaded to GenBank and are available under BioProject PRJNA316001. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession LVJD00000000. The version described in this paper is version LVJD01000000. Sequences for putative and confirmed cytochrome P450 monooxygenases are also available at GenBank under the accessions KX496991 (Sya_B2_00569), KX496992 (Sya_B2_01856), KX496993 (Sya_B2_02741), KX496994 (Sya_B2_02767; CYP101J4), KX496995 (Sya_B2_03538; CYP101J3), KX496996 (Sya_B2_03558; CYP101J2), KX496997 (Sya_B2_04893), KX496998 (Sya_B2_04918) and KX496999 (Sya_B2_05084). The partial 16S ribosomal RNA gene sequence used to classify strain B2 is available at GenBank under the accession KX507119.

RESULTS

Strain identification and genomic analysis

Previously, a Gram-negative, rod-shaped bacterial isolate, designated strain B2, which was capable of using 1,8-cineole as a sole carbon and energy source was isolated from activated sludge (23). To characterise this strain further genome sequence analysis was performed on this isolate. Genome sequencing and assembly of the filtered reads using Velvet yielded a draft genome sequence of approximately 5.9 Mb containing 263 scaffolds. These scaffolds varied in length, and were distributed with mean, median and N50 lengths of 22468 bp, 7012 bp and 60180 bp, respectively. The longest scaffold was 178243 bp in length, while the average GC content of the genome was 62.3%. A 1484 bp segment of the 16S rRNA gene sequence was extracted from the draft strain B2 genome and used to classify the strain as *Sphingobium* using the RDP classifier (28). All characteristic 16S rRNA signatures that distinguish this genus (32) were present in the sequence of strain B2 and it matched closely (≥ 99%) with members of the species *S. yanoikuyae* including the type strain (ATCC 51230 = DSM 7462). It was hence designated as *S. yanoikuyae* strain B2. A comparison of 16S rRNA gene sequences from a selection of strains belonging to the genus *Sphingobium* is shown in the Supplementary Information (Figure S2). The *S. yanoikuyae* isolates appear as a single clade which contains no taxa from other species. The percentage identity of the *S. yanoikuyae* strains for which a draft genome sequence is available and an outgroup taxon (*Sphingobium indicum* B90A) was calculated by counting the number of kMers (25-mers) shared between each pair of organisms, effectively the D_2 metric (33, 34). The code used to calculate this relatedness metric differs from conventional D_2 tools in that it compensates for the effects of single mismatched bases, giving counts that

more closely approximate those obtained from multiple-alignment tools. The comparison shows that the *S. yanoikuyae* strains share 68 – 80% genome identity (Supplementary Information, Table S4). A comparison of the distribution of orthologous proteins in each of the six *S. yanoikuyae* strains (annotated using Prokka) was derived using Roary version 3.6.0 (35) and the data displayed using FriPan (http://drpowell.github.io/FriPan/) and approximately 1000 proteins are unique to strain B2 (Figure 1).

Metabolite characterisation

To determine the metabolic intermediates produced by *S. yanoikuyae* strain B2 from 1,8-cineole, GC analysis of aqueous samples taken from shake flask culture supernatants was performed. In exponential phase, residual 1,8-cineole was detected as well as three other compounds with different retention times. No 1,8-cineole or the compounds found in exponential cultures were detected in stationary phase cultures however, a further two additional compounds with longer retention times were observed (data not shown). Solvent extraction of 500 mL cultures provided a sufficient quantity of each compound to enable identification. Using previously described methods (12; NMR and GC-MS), the compounds found in the dichloromethane extracts from exponential phase cultures were identified as the α and β forms of hydroxylated 1,8-cineole and oxo-1,8-cineole (oxidised at position 2 or 6); only 5,5-dimethyl-4-(3'-oxobutyl)-4,5-dihydrofuran-2(3H)-one) was identified in solvent extracts from stationary phase cultures. These compounds have been described previously and are consistent with 1,8-cineole-derived products extracted from cultures of *N. subterranea* (12).

342

343 **Identification of P450s**

344 Genes encoding P450s in genome sequence

345 After annotation of the draft genome nine P450-like proteins were identified. The
346 nine proteins included six putative camphor-hydroxylating P450s, one putative α-
347 terpineol-oxidising P450, one putative cytochrome P450 PikC and one putative
348 pentalenene oxygenase. A BLAST search of the uniprot database
349 (http://www.uniprot.org/blast) showed that these putative P450s are similar to a
350 number of proteins including putative P450s from other *S. yanoikuyae* strains but
351 four of the P450s were absent from the type strain genome (Supplementary
352 Information, Table S1). Two of these genes, CYP101J2 and CYP101J3, appear to
353 be located on a single contiguous sequence, with complementary, overlapping ends
354 that appears to be a plasmid. This plasmid was 73051 bp in length, and had a GC
355 content of 61.2%. The sequence for the putative plasmid is available at the CSIRO
356 Data Access Portal (https://data.csiro.au/dap/landingpage?pid=csiro:17150).
357 BLASTn analyses of the putative plasmid indicate its closest matches are to
358 plasmids from *Sphingobium chlorophenolicum* (pSPHCH01), *Sphingobium* sp. YBL2
359 (2pYBL2-2) and *Sphingomonas wittichii* RW1 (pSWIT02), though it shares only low
360 identity (73%, over 10-20% query coverage) with these plasmids.

361

362 Growth of *S. yanoikuyae* strains B2 and DSM 7462 on selected monoterpenoids

363 To determine whether the ability to grow on 1,8-cineole is ubiquitous in *S.*
364 *yanoikuyae* strains, strain B2 and the type strain *S. yanoikuyae* strain DSM 7462 (=

365 ATCC 51230) were tested for growth in DM2 supplemented with either 0.25 to 1 mL
366 L^{-1} 1,8-cineole, 0.25 to 1 mL L^{-1} α-terpineol or 0.25 to 1 g L^{-1} (1*R*)-(+)-camphor or
367 (1*S*)-(–)-camphor as sole sources of carbon and energy. Growth in the
368 supplemented media at 30 °C and 200 rpm was monitored over a 7 day time period.
369 Following 7 days of incubation, visual examination showed that *S. yanoikuyae* DSM
370 7462 did not grow on any of the substrates tested whilst *S. yanoikuyae* strain B2
371 grew on both 1,8-cineole and α-terpineol, but showed no growth on either of the
372 camphor enantiomers.

373

374 Purification of 1,8-cineole-binding P450s from *S. yanoikuyae* strain B2

375 Confirmation of the genes encoding 1,8-cineole-hydroxylating P450s required
376 purification of 1,8-cineole-binding P450s from *S. yanoikuyae* strain B2. Proteins were
377 purified from extracts of *S. yanoikuyae* strain B2 cells grown in fed-batch mode using
378 glucose and/or 1,8-cineole as carbon sources. Binding of 1,8-cineole was tested in
379 partially purified protein fractions. Transition from the low-spin to the high-spin state
380 (36) upon 1,8-cineole-binding was detected in multiple peaks after ion exchange
381 chromatography (IEX) of ammonium sulphate (AS) precipitated fractions. Two
382 distinct peak fractions (PF), PF 1 and PF 2 from the 40-60% AS cut, were further
383 purified using gel filtration. After the 3-step purification procedure PF 1 was purified
384 to apparent homogeneity by SDS-PAGE analysis with an estimated mass of ca. 46
385 kDa, PF 2 appeared to still contain impurities (Figure 2). The purity was also
386 reflected by the Reinheitszahl (*RZ*), which is defined as the ratio between P450
387 absorbance at its typical Soret absorbance for the LS state (in this case 417 nm) and
388 the total protein absorbance at 280 nm ($A_{417\ nm}/A_{280\ nm}$). A higher *RZ* is indicative of

higher protein purity (11). PF 1 and PF 2 had an *RZ* of 1.4 and 0.6 respectively, which correlated with the purity observed by SDS-PAGE analysis (Figure 2).

After concentration, reduction and alkylation PF1 was further purified using reversed-phase high-performance liquid chromatography. N-terminal sequencing of the fraction with a molecular mass of approximately 46-47 kDa, the typical size for P450s, produced a readable sequence of 22 amino acids. Protein identity was confirmed by a tryptic digest of the reduced and alkylated protein followed by MALDI-MS fingerprint mapping. Together with the genome sequence, the MS data was used to identify a gene encoding a P450 (Sya_B2_03558; CYP101J2) with a length of 410 amino acids.

Protein identity for PF2 was determined by an in-gel tryptic digest of the protein with a mass of ca. 46 kDa followed by MALDI-MS fingerprint mapping. The peptide sequences gave more than 30% coverage of a P450 (Sya_B2_03538; CYP101J3) with a length of 406 amino acids. A third protein (Sya_B2_02767; CYP101J4) was identified as a potential 1,8-cineole-hydroxylating P450 based on high amino acid sequence identity (74%) to CYP101J2.

Enzyme characterisation

Recombinant expression and purification of P450s

410 To further characterise the identified P450s, the genes encoding the three identified
411 1,8-cineole-binding P450s were cloned into the pET28a(+) expression vector and
412 expressed in *E. coli* BL21 (DE3). This yielded N-terminally His_6-tagged proteins,
413 which were purified using immobilised metal affinity chromatography (IMAC). The
414 molecular weights estimated from the protein sequences including tags (CYP101J2
415 = 48.7 kDa, CYP101J3 = 47.8 kDa, CYP101J4 = 48.4 kDa) matched the main
416 expression products (Figure 3). Yield of active protein was estimated in the purified
417 pools by performing CO-difference spectroscopy in the presence of 1,8-cineole.
418 CYP101J2 and CYP101J3 were both recovered in good yields (>30 mg purified
419 protein per litre original culture). CYP101J4 was expressed in lower quantities as
420 judged by SDS-PAGE of IMAC fractions (data not shown) and this was reflected by
421 the amount of functional protein recovered (ca. 1.5 mg purified protein per litre
422 original culture).

423

424 Spectroscopic characterisation

425 All three isolated and purified P450s showed the characteristic spectroscopic
426 properties expected for P450s (Figure 4). Most importantly, the typical Soret
427 absorbance maximum (416-417 nm) largely shifted to lower wavelengths (392-396
428 nm) upon addition of 1,8-cineole. After reduction and CO binding the P450s
429 exhibited the typical absorbance maximum at ca. 450 nm (445-446 nm).

430

431 Substrate binding

The purified P450s were tested for their ability to bind 1,8-cineole. The dissociation constants (K_D) determined with 1,8-cineole as substrate are summarised in Table 1. K_D values are in the same order of magnitude for all three P450s, CYP101J4 shows slightly tighter substrate binding (by a factor of ca. 2). Binding of 1,8-cineole was increased by a factor of approximately 2 when 0.2 M KCl was present. Under the tested conditions, ΔA_{max} was largest using 1,8-cineole in the presence of KCl for CYP101J2, CYP101J3 or CYP101J4, respectively. Compared to the shift of each individual P450 with 1,8-cineole and 0.2 M KCl, addition of 1,8-cineole resulted in a 95% shift (CYP101J2), 85% (CYP101J3) and 90% (CYP101J4) shift in the absence of potassium chloride.

CYP101J2 and CYP101J3 were also screened for binding of alternative substrates (2-adamantanone, toluene, β-ionone, (1*S*)-(–)-camphor and (1*R*)-(+)-camphor) and percentage spin-state shifts are summarised in Table 2. The percentage type I spin-state shifts induced by these substrates were estimated to approximately ±5% by comparison to the shift of each individual P450 with 1,8-cineole in the presence of 0.2 M KCl. Whilst β-ionone triggered a 40% (CYP101J2) or 35% (CYP101J3) shift, binding of 2-adamantanone, (1*S*)-(–)-camphor and (1*R*)-(+)-camphor resulted in small shifts between 5 and 20%. Toluene addition resulted in no detectable spin-state shift. The enzymes did not show a type I shift with α-terpineol. However, there was a shift from ca. 410 nm to ca. 426 nm indicating a type II shift. Due to the low expression levels of CYP101J4, alternative substrates were not tested with this enzyme.

Functionality of recombinant P450s in *E. coli*

E. coli cultures expressing recombinant CYP101J2 in the presence of 1,8-cineole were solvent extracted and analysed using GC-MS. Forty five hours after addition of 1,8-cineole a peak with a retention time (RT) of 4.29 min was detected (Figure 5A). This peak had a mass increase of 16 units with a molecular peak at *m/z* = 170 and the typical fragmentation pattern of hydroxylated 1,8-cineole (6). It was identified with a 91-94% probability match to 1,3,3-trimethyl-2-oxabicyclo[2.2.2]octan-6-ol in the National Institute of Standards and Technology (NIST) 14 (MS Database and MS Search Program v.2.2) library containing spectra for 242466 chemical compounds. This peak was not detected in the negative control (*E. coli* BL21 (DE3) harbouring the pET28a(+) vector alone) (Figure 5B). Analysis of a mixture of chemically synthesised α and β forms of (1*S*)-2-hydroxy-1,8-cineole showed that these two stereoisomers can be separated using non-chiral GC-MS (RT of 4.19 and 4.29 min) (Figure 5C). Analysis of (1*R*)-6β-hydroxy-1,8-cineole (produced using $P450_{cin}$ isolated from *C. braakii*) using the same technique yielded a peak with a RT of 4.19 min (Figure 5D). When (1*R*)-6β-hydroxy-1,8-cineole was mixed with a solvent extract from a culture expressing CYP101J2 two distinct GC peaks were detected (RT of 4.19 min and 4.29 min) with highly similar MS data (Figure 5E). Since monohydroxylated 1,8-cineole from this study was separated from (1*R*)-6β-hydroxy-1,8-cineole) using non-chiral GC, and had the same RT as one of the α and β forms of (1*S*)-2-hydroxy-1,8-cineole we putatively identified the product of CYP101J2 as (1*S*)-2α-hydroxy-1,8-cineole or (1*R*)-6α-hydroxy-1,8-cineole. The same data was obtained when testing CYP101J3 and CYP101J4 under identical conditions (data not shown).

DISCUSSION

The industrial application of biocatalytic processes requires enzymes with specific catalytic ability and on many occasions these enzymes are discovered in microbial isolates with particular metabolic capabilities. To enable development of cost effective production of 1,8-cineole derivatives we previously isolated a number of microorganisms with the ability to utilise 1,8-cineole as a sole carbon and energy source (23). Unlike earlier studies during which microorganisms were isolated from environments that were likely to contain 1,8-cineole (e.g. soil proximate to *Eucalyptus* trees and *Eucalyptus* leaves) (8, 9, 11, 13), our approach was to interrogate a potentially more diverse microbial consortium – activated sludge. This approach is not without precedent. Terpene oxidation in a microorganism from sewage sludge was reported as early as in 1959 when an isolate that was identified as a pseudomonad was shown to utilise camphor as a carbon source (37).

In this study one isolate, strain B2, demonstrating a relatively high tolerance to 1,8-cineole (23), was chosen for further study. A draft genome sequence was determined and the strain B2 was identified as *S. yanoikuyae* based on its partial 16S rRNA gene sequence extracted from the genome sequence. Whilst there may be more than one copy of the 16S rRNA gene, this was not determined and it is unlikely that potential sequence variation would be significant enough to change the taxonomic classification. Interestingly, compared to other *S. yanoikuyae* draft genomes, the sequence of strain B2 is relatively large. A comparison between the sequenced strains of *S. yanoikuyae* is shown in Table S4 in the Supplementary Information. There appears to be more genetic difference between *Sphingobium yanoikuyae* strains than is normal for a typically defined bacterial species. This likely

505 accounts for the differences in genome size, and may indicate further taxonomic
506 scrutiny is warranted.

507

508 The isolate belongs to the group of sphingomonads, bacteria that are known to
509 degrade a diverse range of recalcitrant compounds in particular environmental
510 pollutants (38). This potential metabolic diversity may also be related to modification
511 and degradation of terpenes. Previously, P450s isolated from *N. aromaticivorans*
512 DSM12444 were shown to hydroxylate camphor (39) and a *Sphingomonas* isolate
513 was capable of metabolising 1,8-cineole (13). However, to our knowledge this is the
514 first report of 1,8-cineole metabolism by a sphingomonad from the genus
515 *Sphingobium*. A comparison of the distribution of orthologous proteins in each of six
516 sequenced *S. yanoikuyae* strains also indicates metabolic flexibility amongst
517 members of this species. A significant fraction of proteins was unique to each strain,
518 with approximately 1000 proteins unique to strain B2 (Figure 1).

519

520 Analysis of *S. yanoikuyae* strain B2 cultures grown with 1,8-cineole as a sole carbon
521 and energy source enabled identification of a series of compounds in solvent
522 extracts of culture supernatants. Three compounds were identified in exponential
523 phase cultures (the α and β forms of hydroxylated 1,8-cineole and oxo-1,8-cineole).
524 The presence of two forms of hydroxy-1,8-cineole may be due to the reduction of
525 oxo-1,8-cineole by the organisms rather than a direct product of 1,8-cineole
526 hydroxylation (1). These compounds are typical metabolites derived from 1,8-cineole
527 (8, 9, 12, 13). Based on the analytical techniques used the enantioselectivity of the
528 initial oxidation could not be established therefore the oxidation may occur at position

529 2 or 6. Although two different compounds were identified in aqueous samples from
530 stationary phase cultures only one compound was recovered from solvent extracts of
531 these cultures. The single compound identified as 5,5-dimethyl-4-(3'-oxobutyl)-4,5-
532 dihydrofuran-2(3H)-one). Williams et al. (1989) also detected this compound in
533 solvent extracts from crude extracts of *Rhodococcus* strain C1 incubated with 6-oxo-
534 1,8-cineole and concluded that this compound is not derived from metabolism of 1,8-
535 cineole but is in fact an artefact of the acidification and solvent extraction procedure.
536 A more likely product is 3-(1-hydroxy-1-methylethyl)-6-oxoheptanoic acid which
537 could be one of the compounds detected in aqueous supernatant samples however
538 production of this compound from 1,8-cineole by *S. yanoikuyae* strain B2 was not
539 confirmed.

540

541 The metabolites produced by *S. yanoikuyae* B2 included hydroxylated 1,8-cineole;
542 hydroxylation is typically the first step in the metabolism of 1,8-cineole. To gain
543 greater insight into this initial reaction, we sought to identify and characterise the
544 enzyme(s) involved in catalysis. Through sequencing of the selected strain's
545 genome, several genes potentially encoding P450s were identified based on
546 similarity to known P450 sequences. These included six genes putatively encoding
547 P450s that were similar to known camphor-hydroxylating P450s and one which was
548 similar to an α-terpineol-oxidising P450 (Supplementary Information, Table S1). This
549 was encouraging, as both, camphor and α-terpineol are monoterpenoids like 1,8-
550 cineole and camphor in particular shares structural similarity to 1,8-cineole.
551 However, sequences of P450s are diverse (40) and the hydroxylation of similar
552 substrates is often catalysed by a range of divergent P450s from different families

553 (41). In all, nine P450s (putatively annotated) were found and consequently further
554 characterisation of the native proteins was required to determine function.

555

556 As a first step towards elucidating the P450s involved in 1,8-cineole hydroxylation,
557 the metabolic ability of *S. yanoikuyae* strain B2 was compared to the type strain
558 (DSM 7462 = ATCC 51230). Unlike *S. yanoikuyae* B2, strain DSM 7462 was not able
559 to grow on 1,8-cineole or α-terpineol. Through comparison of the genomes from the
560 two strains, this observation correlated with the absence of orthologous proteins of
561 the four P450s in the type strain that were present in *S. yanoikuyae* B2
562 (Supplementary Information, Table S1). These genes were therefore implicated in
563 1,8-cineole (and/or α-terpineol) hydroxylation. However, it is also possible that
564 enzymes catalysing subsequent steps in the degradation of the tested
565 monoterpenoids are absent from the *S. yanoikuyae* strain DSM 7462 and hence
566 further investigation was required to confirm the function of these genes.

567

568 Other monoterpenoid oxidising P450s such as CYP176A1 ($P450_{cin}$) and CYP108A1
569 ($P450_{terp}$) were isolated and identified via purification from native cell extracts (11,
570 42). To facilitate a similar approach, a 3-step purification procedure was developed
571 to isolate 1,8-cineole-binding P450s from *S. yanoikuyae* strain B2 cell-free extracts.
572 Two proteins were purified and partial amino acid sequencing and subsequent
573 comparison with translated genes from the genome sequence resulted in the
574 identification of two 1,8-cineole-binding proteins, CYP101J2 and CYP101J3. A third
575 P450, CYP101J4, was identified based on sequence identity to CYP101J2. These
576 protein sequences represent three of the four P450s which were absent in the strain

577 ATCC 51230 draft genome. The three identified proteins belong to the cytochrome
578 P450 family CYP101. Pairwise sequence alignments of all characterised P450s from
579 the CYP101 family as well as $P450_{cin}$ (CYP176A1) showed that CYP101J2 shares
580 the highest sequence identity with CYP101J4 (74%). CYP101J3 shares sequence
581 identity with CYP101J2 (53%) and CYP101J4 (54%). When compared to previously
582 characterised members of the CYP101 family, all three newly discovered P450s are
583 most closely related to CYP101B1 and CYP101A1 sharing 45-47% and 44%
584 sequence identity, respectively. CYP101B1 has been shown to oxidise a broad
585 range of substrates including norisoprenoids (e.g. β-ionone), various aromatic
586 compounds and camphor (39, 43). CYP101A1 was characterised based on its
587 activity with camphor (18). More recently, both CYP101B1 and CYP101A1 have also
588 been shown to hydroxylate 1,8-cineole producing a range of hydroxy-1,8-cineole
589 isomers (14). The P450 from *C. braakii* is unique as no other bacterial enzymes have
590 been found that belong to the CYP176 family or use similar electron transfer
591 partners. However, despite the enzymes from *S. yanoikuyae* strain B2 and *C. braakii*
592 sharing less than 30% sequence identity, all oxidise the same substrate indicating a
593 high degree of flexibility within this large group of enzymes.

594

595 To characterise the cineole-binding proteins, CYP101J2, CYP101J3 and CYP101J4
596 were heterologously expressed in *E. coli* BL21 (DE3) as N-terminally His_6-tagged
597 fusions and purified using IMAC. Recovery of functional proteins varied significantly.
598 Whilst CYP101J2 and CYP101J3 were obtained in good yield, expression of
599 CYP101J4 under the selected conditions resulted in lower protein recovery. This
600 may be due to improper folding (and/or the formation of insoluble aggregates), which
601 could be resolved by modifying protein expression conditions. Spectroscopic

602 characterisation of the three purified recombinant proteins revealed the distinctive
603 P450 properties. All 3 recombinant enzymes showed the typical P450 absorbance
604 peaks and shifts and most importantly, 1,8-cineole-binding resulted in a large spin-
605 state shift indicating that 1,8-cineole may be the preferred substrate. K_D values using
606 1,8-cineole as substrate were in the same order of magnitude for all three P450s and
607 the presence of KCl appeared to result in tighter substrate binding (by a factor of ca.
608 2). Potassium ions have previously been shown to have a positive effect on
609 substrate binding by other P450s (39, 44). Within the structure of CYP101A1, Glu^{84},
610 Gly^{93}, Glu^{94} and Tyr^{96} have been shown to be involved in binding of potassium ions
611 (19, 45) with the latter three residues being conserved in the CYP101J enzymes
612 from *S. yanoikuyae* B2. The presence of potassium ions has as small effect on 1,8-
613 cineole-binding by CYP101J2, CYP101J3 and CYP101J4. It is not clear whether the
614 effect of potassium ions is specifically related to the interaction of potassium ions
615 with the enzymes or a non-specific effect due to altered ionic strength causing a
616 change in substrate solubility. The affinity of CYP101J2, CYP101J3 and CYP101J4
617 for 1,8-cineole is less than $P450_{cin}$ (11) and is also lower than other monoterpenoid-
618 hydroxylating P450s for their preferred substrates (18, 42).

619

620 Testing CYP101J2 and CYP101J3 with alternative substrates including compounds
621 which are converted by other P450s from the CYP101 family, did not lead to a type I
622 spin-state shift greater than that induced by 1,8-cineole. For example, β-ionone and
623 camphor triggered spin-state shifts of 35-40% and 10-15% with CYP101J2 and
624 CYP101J3, respectively. These shifts are relatively small compared to the large shift
625 observed when β-ionone binds to CYP101B1 (39) and camphor to CYP101A1 (46).
626 *S. yanoikuyae* B2 was able to use both 1,8-cineole and α-terpineol as carbon

627 sources for growth however, α-terpineol did not induce a type I spin-state shift in
628 either CYP101J2 and CYP101J3. This observation combined with the minimal shifts
629 triggered by the camphor enantiomers may indicate that 1,8-cineole is the
630 physiological substrate for the newly discovered P450s. Furthermore, despite being
631 shown to oxidise 1,8-cineole, CYP101A has a K_D value that is higher and a spin-
632 state shift that is smaller (14) than the P450s from *S. yanoikuyae*. These functional
633 variations due to sequence divergence and presumably structural differences
634 warrant further investigations and comparison.

635

636 Testing of CYP101J2, CYP101J3 and CYP101J4 in a recombinant whole-cell
637 biotransformation confirmed their proposed function. Presumably with the aid of
638 electron transport partners from *E. coli*, all three enzymes catalysed the oxidation of
639 1,8-cineole to a product tentatively identified as (1*S*)-2α-hydroxy-1,8-cineole or (1*R*)-
640 6α-hydroxy-1,8-cineole. The product from the *S. yanoikuyae* enzymes is
641 hydroxylated in the α-orientation compared to the product of the wild-type $P450_{cin}$
642 from *C. braakii* which hydroxylates in the ß-orientation (7). Due to the symmetric
643 nature of 1,8-cineole, NMR would also not enable differentiation between derivatives
644 that are hydroxylated in either the 2 or 6 position. Interestingly, even though the 1,8-
645 cineole-hydroxylating proteins from the two different microorganisms share limited
646 sequence similarity and hydroxylate in different orientation, both are able to use
647 electron transfer chains present in *E. coli* (11). Compared to the N242A $P450_{cin}$
648 mutant, CYP101A1 and CYP101B1, which each produce a series of hydroxylated
649 1,8-cineoles, the CYP101J enzymes from *S. yanoikuyae* B2 appear to hydroxylate
650 1,8-cineole more specifically (14).

651

652 This research has shown that *S. yanoikuyae* strain B2 harbours at least three genes
653 encoding enzymes with similar function; the presence of multiple genes that encode
654 functionally similar proteins suggests that the action of these proteins may be
655 essential for growth or survival of the bacterium within specific environments.
656 Functional redundancy is often the result of this essential need. Unlike other
657 monoterpene oxidising P450s including $P450_{cin}$ (11), $P450_{terp}$ (42) and $P450_{cam}$ (47),
658 the genes encoding the P450s were not adjacent to potential electron transport
659 partners in the draft genome sequence. This observation is probable as genes
660 encoding the enzymes involved in degradative pathways in sphingomonads are
661 often scattered throughout a number of gene clusters within the genome (including
662 plasmids) (31, 48, 49).

663

664 In this study three new bacterial P450s capable of 1,8-cineole hydroxylation have
665 been isolated from *S. yanoikuyae* strain B2. Whilst these P450s provide a basis of
666 comparison to other bacterial monoterpenoid-oxidising P450s for mechanistic
667 studies, these enzymes can also be used for process development and scaling up
668 biochemical production of hydroxylated 1,8-cineole derivatives. Furthermore, as the
669 number of enzymes in the CYP101 family grows enzyme modification using DNA
670 shuffling can be undertaken to engineer proteins with improved properties and
671 altered substrate specificities. We are currently identifying and testing potential
672 native electron transport partners from *S. yanoikuyae* strain B2 to increase product
673 yields in recombinant *E. coli*. Scale-up of the biotransformation will also enable the

isolation of the hydroxylated 1,8-cineole derivatives for a definitive structural characterisation and determination of the hydroxylation position.

ACKNOWLEDGMENTS

We would like to thank Carl Braybrook and Jo Cosgriff for GC-MS analyses and Lindsay Sparrow and Nick Bartone for amino acid sequencing. We would also like to acknowledge Stella Kyi for synthesis of a monohydroxylated 1,8-cineole standard and Professor James de Voss for providing the bicistronic plasmid containing the $P450_{cin}$ gene.

FUNDING INFORMATION

Assoc. Professor Dena Lyras was supported by Future Fellowship FT120100779 (Australian Research Council).

686 **REFERENCES**

687 1. **Carman RM, MacRae IC, Perkins MV.** 1986. The oxidation of 1,8-cineole by
688 *Pseudomonas flava*. Aust J Chem **39:**1739-1746.

689 2. **Simonsen JL, Owen LN.** 1953. The Terpenes, 2nd ed, vol 1. Cambridge
690 University Press, Cambridge, UK.

691 3. **Farlow AJ, Bernhardt PV, De Voss JJ.** 2013. Synthesis of oxygenated
692 cineole derivatives from cineole: utility of cytochrome $P450_{cin}$ as an
693 enantioselective catalyst. Tetrahedron Asymmetry **24:**324-333.

694 4. **de Boggiatto MV, de Heluani CS, de Fenik IJS, Catalán CAN.** 1987.
695 Regiospecific functionalization of the monoterpene ether 1,3,3-trimethyl-2-
696 oxabicyclo[2.2.2]octane (1,8-cineole). Synthesis of the useful bridged gamma
697 γ-lactone 1,3-dimethyl-2-oxabicyclo[2.2.2]octan-3→5-olide. J Org Chem
698 **52:**1505-1511.

699 5. **Asakawa Y, Matsuda R, Tori M.** 1986. Hydroxylation of menthols and
700 cineoles with m-chloroperbenzoic acid. Experientia **42:**201-203.

701 6. **Azerad R.** 2014. 1,8-Cineole: chemical and biological oxidation reactions and
702 products. Chempluschem **79:**634-655.

703 7. **Slessor KE, Hawkes DB, Farlow A, Pearson AG, Stok JE, De Voss JJ.**
704 2012. An *in vivo* cytochrome $P450_{cin}$ (CYP176A1) catalytic system for
705 metabolite production. J Mol Catal B Enzym **79:**15-20.

706 8. **MacRae IC, Alberts V, Carman RM, Shaw IM.** 1979. Products of 1,8-cineole
707 oxidation by a pseudomonad. Aust J Chem **32:**917-922.

708 9. **Williams RD, Trudgill PW, Taylor DG.** 1989. Metabolism of 1,8-cineole by a
709 *Rhodococcus* species: ring cleavage reactions. J Gen Microbiol **135:**1957–
710 1967.

10. **Liu WG, Rosazza JPN.** 1990. Stereospecific hydroxylation of 1,8-cineole using a microbial biocatalyst. Tetrahedron Lett **31:**2833-2836.

11. **Hawkes DB, Adams GW, Burlingame AL, Ortiz de Montellano PR, De Voss JJ.** 2002. Cytochrome $P450_{cin}$ (CYP176A), isolation, expression, and characterization. J Biol Chem **277:**27725-27732.

12. **Rasmussen J-AM, Henderson KA, Straffon MJ, Dumsday GJ, Coulton J, Zachariou M.** 2005. Two new biocatalysts for improved biological oxidation of 1,8-cineole. Aust J Chem **58:**912–916.

13. **Knight AR.** 2009. Ph.D. thesis. Murdoch University, Perth, Western Australia.

14. **Stok JE, Hall EA, Stone ISJ, Noble MC, Wong SH, Bell SG, De Voss JJ.** 2016. *In vivo* and *in vitro* hydroxylation of cineole and camphor by cytochromes P450CYP101A1, CYP101B1 and N242A CYP176A1. J Mol Catal B Enzym **128:**52-64.

15. **Meharenna YT, Slessor KE, Cavaignac SM, Poulos TL, De Voss JJ.** 2008. The critical role of substrate-protein hydrogen bonding in the control of regioselective hydroxylation in P450cin. J Biol Chem **283:**10804-10812.

16. **Urlacher VB, Lutz-Wahl S, Schmid RD.** 2004. Microbial P450 enzymes in biotechnology. Appl Microbiol Biotechnol **64:**317-325.

17. **Nelson DR.** 2009. The cytochrome P450 homepage. Hum Genomics **4:**59-65.

18. **Katagiri M, Ganguli BN, Gunsalus IC.** 1968. A soluble cytochrome P-450 functional in methylene hydroxylation. J Biol Chem **243:**3543-3546.

19. **Poulos TL, Finzel BC, Howard AJ.** 1987. High-resolution crystal structure of cytochrome P450cam. J Mol Bio **195:**687-700.

20. **Schewe H, Mirata MA, Holtmann D, Schrader J.** 2011. Biooxidation of monoterpenes with bacterial monooxygenases. Process Biochem **46:**1885-1899.

21. **Hannemann F, Bichet A, Ewen KM, Bernhardt R.** 2007. Cytochrome P450 systems—biological variations of electron transport chains. Biochim Biophys Acta **3:**330-344.

22. **Yang G, Ding Y.** 2014. Recent advances in biocatalyst discovery, development and applications. Bioorg Med Chem **22:**5604-5612.

23. **Dumsday G, Pilkington N, Straffon M, Zachariou M.** 2007. Using eucalypt oil as a carbon source for deriving other compounds – development of a microbial library. RIRDC Publication No 07/086.

24. **Dumsday GJ, Ocal G, Bridger JS, Zachariou M.** 2009. The use of oxygen uptake rate to monitor discovery of microbial and enzymatic biocatalysts. Biotechnol Bioeng **102:**673-683.

25. **Miller JH.** 1972. Experiments in molecular genetics. Cold Spring Harbour Laboratory Press: Cold Spring Harbor, NY.

26. **Zerbino DR, Birney E.** 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res **18:**821-829.

27. **Seemann T.** 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30:**2068-2069.

28. **Wang Q, Garrity GM, Tiedje JM, Cole JR.** 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol **73:**5261-5267.

29. **Omura T, Sato R.** 1964. The carbon monoxide-binding pigment of liver microsomes. I. Evidence for its hemoprotein nature. J Biol Chem **239:**2370-2378.

30. **Schenkman JB, Jansson I.** 2006. Spectral analyses of cytochromes P450, p 11-18. *In* Phillips IR, Shephard EA (ed), Cytochrome P450 protocols, 2nd ed. Humana Press, Totowa, NJ.

31. **Bell SG, Dale A, Rees NH, Wong L-L.** 2010. A cytochrome P450 class I electron transfer system from *Novosphingobium aromaticivorans*. Appl Microbiol Biotechnol **86:**163-175.

32. **Takeuchi M, Hamana K, Hiraishi A.** 2001. Proposal of the genus *Sphingomonas sensu stricto* and three new genera, *Sphingobium, Novosphingobium* and *Sphingopyxis*, on the basis of phylogenetic and chemotaxonomic analyses. Int J Syst Evol Microbiol **51:**1405-1417.

33. **Song K, Ren J, Reinert G, Deng M, Waterman MS, Sun F.** 2014. New developments of alignment-free sequence comparison: measures, statistics and next-generation sequencing. Brief Bioinform **15:**343-353.

34. **Chan CX, Bernard G, Poirion O, Hogan JM, Ragan MA.** 2014. Inferring phylogenies of evolving sequences without multiple sequence alignment. Sci Rep **4:**6504.

35. **Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J.** 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics **31:**3691-3693.

36. **Luthra A, Denisov IG, Sligar SG.** 2011. Spectroscopic features of cytochrome P450 reaction intermediates. Arch Biochem Biophys **507:**26-35.

37. **Bradshaw WH, Conrad HE, Corey EJ, Gunsalus IC, Lednicer D.** 1959. Microbiological degradation of (+)-camphor. J Am Chem Soc **81:**5507.

38. **Aylward FO, McDonald BR, Adams SM, Valenzuela A, Schmidt RA, Goodwin LA, Woyke T, Currie CR, Suen G, Poulsen M.** 2013. Comparison of 26 sphingomonad genomes reveals diverse environmental adaptations and biodegradative capabilities. Appl Environ Microbiol **79:**3724-3733.

39. **Bell SG, Wong L-L.** 2007. P450 enzymes from the bacterium *Novosphingobium aromaticivorans*. Biochem Biophys Res Commun **360:**666-672.

40. **Sezutsu H, Le Goff G, Feyereisen R.** 2013. Origins of P450 diversity. Philos Trans R Soc Lond B Biol Sci **368:**20120428.

41. **Schuler MA, Sligar SG.** 2007. Diversities and similarities in P450 systems: an introduction, p 1-26. *In* Sigel A, Sigel H, Sigel R (ed), The ubiquitous roles of cytochrome P450 proteins. John Wiley & Sons, Ltd, Chichester, England.

42. **Peterson JA, Lu J-Y, Geisselsoder J, Graham-Lorence S, Carmona C, Witney F, Lorence MC.** 1992. Cytochrome P-450$_{terp}$. Isolation and purification of the protein and cloning and sequencing of its operon. J Biol Chem **267:**14193-14203.

43. **Hall EA, Bell SG.** 2015. The efficient and selective biocatalytic oxidation of norisoprenoid and aromatic substrates by CYP101B1 from *Novosphingobium aromaticivorans* DSM12444. RSC Adv **5:**5762-5773.

44. **Deprez E, Gill E, Helms V, Wade RC, Hui Bon Hoa G.** 2002. Specific and non-specific effects of potassium cations on substrate–protein interactions in cytochromes P450cam and P450lin. J Inorg Biochem **91:**597-606.

806 45. **Westlake ACG, Harford-Cross CF, Donovan J, Wong LL.** 1999. Mutations
807 of glutamate-84 at the putative potassium-binding site affect camphor binding
808 and oxidation by cytochrome $P450_{cam}$. Eur J Biochem **265:**929-935.
809 46. **Sligar SG.** 1976. Coupling of spin, substrate, and redox equilibria in
810 cytochrome P450. Biochemistry **15:**5399-5406.
811 47. **Koga H, Yamaguchi E, Matsunaga K, Aramaki H, Horiuchi T.** 1989.
812 Cloning and nucleotide sequences of NADH-putidaredoxin reductase gene
813 (camA) and putidaredoxin gene (camB) involved in cytochrome P-450cam
814 hydroxylase of *Pseudomonas putida*. J Biochem **106:**831-836.
815 48. **Stolz A.** 2009. Molecular characteristics of xenobiotic-degrading
816 sphingomonads. Appl Microbiol Biotechnol **81:**793-811.
817 49. **Basta T, Keck A, Klein J, Stolz A.** 2004. Detection and characterization of
818 conjugative degradative plasmids in xenobiotic-degrading *Sphingomonas*
819 strains. J Bacteriol **186:**3862-3872.

820 **FIGURE LEGENDS**

821 **FIGURE 1** Overview of the distribution of orthologous proteins in each of the six
822 sequenced *S. yanoikuyae* strains (Sya_B2 = BioProject ID: PRJNA316001;
823 Sya_ATCC51230 = BioProject ID: PRJNA52201; Sya_XLDN2-5 = BioProject ID:
824 PRJNA71691; Sya_B1_75 = BioProject ID: PRJNA255061; Sya_SHJ = BioProject
825 ID: PRJNA239177; Sya_B1_72 = BioProject ID: PRJNA241283) displayed using
826 FriPan (http://drpowell.github.io/FriPan/). The region of continuous green across the
827 six lines (each representing a *S. yanoikuyae* strain) spanning from 0 to ca. 3000
828 indicates the proteins that are present in each of the strains. The region from ca.
829 3000 through to ca. 5500 shows the set of proteins that are present in two to five
830 strains. The region from ca. 5500 through to ca. 11400 is an overview of the proteins
831 that are unique to particular strains.

832 **FIGURE 2** Coomassie-stained SDS-PAGE (4-12% Bis-Tris, MES running buffer) of
833 PF 1 and 2 after IEX and GF. Lane 1: Molecular weight marker (BenchMark,
834 Invitrogen). Lane 2: PF 1 after IEX. Lane 3 and 4: PF 1 after GF. Lane 5: PF 2 after
835 IEX. Lane 3 and 4: PF 2 after GF.

836 **FIGURE 3** Coomassie-stained SDS-PAGE (4-12% Bis-Tris, MOPS running buffer) of
837 purified His_6-tagged P450s (ca. 2 µg) originating from *S. yanoikuyae* strain B2
838 recombinantly expressed in *E. coli* BL21 (DE3). A: Lane 1: Molecular weight marker
839 (SeeBlue Plus 2 Prestained, Invitrogen). Lane 2: Purified CYP101J2; Lane 3:
840 Purified CYP101J4. B: Lane 1: Molecular weight marker (SeeBlue Plus 2 Prestained,
841 Invitrogen). Lane 2: Purified CYP101J3.

842 **FIGURE 4** Absorbance spectra of (A) CYP101J2, (B) CYP101J3 and (C) CYP101J4
843 in 50 mM Tris, pH 7.4, of the oxidised (full black line), oxidised in the presence of

844 1,8-cineole (full grey line), reduced by the addition of sodium dithionite in the
845 presence of 1,8-cineole (broken black line) and reduced and in complex with CO in
846 the presence of 1,8-cineole (broken grey line) forms.

847 **FIGURE 5** GC-MS analysis and comparison with known chemical standards with
848 solvent extracts from *E .coli* cultures expressing CYP101J2. Only relevant peaks are
849 labelled with RT. Unlabelled peaks are impurities. (A) The solvent extract from a
850 culture expressing CYP101J2 in the presence of 1,8-cineole showed a peak at
851 4.29 min. The mass spectrum of this peak was identical to the mass spectra of the
852 hydroxy-1,8-cineole standards. (B) The negative control (solvent extract from *E. coli*
853 BL21 (DE3) culture harbouring the pET28a(+) vector alone) showed a peak at a RT
854 of 2.76 min corresponding to 1,8-cineole – no hydroxy-1,8-cineole was detected. (C)
855 The α and β forms of chemically synthesised (1*S*)-2-hydroxy-1,8-cineole separated
856 using non-chiral GC-MS (RT of 4.19 and 4.29 min) and showed the characteristic
857 mass spectra. (D) Purified (1*R*)-6β-hydroxy-1,8-cineole produced using $P450_{cin}$
858 showed a peak at a RT of 4.19 min and the characteristic mass spectrum. (E) A
859 mixture of purified (1*R*)-6β-hydroxy-1,8-cineole and CYP101J2 solvent extract
860 showed that the product of CYP101J2 separates from (1*R*)-6β-hydroxy-1,8-cineole.

861

FIGURE 1

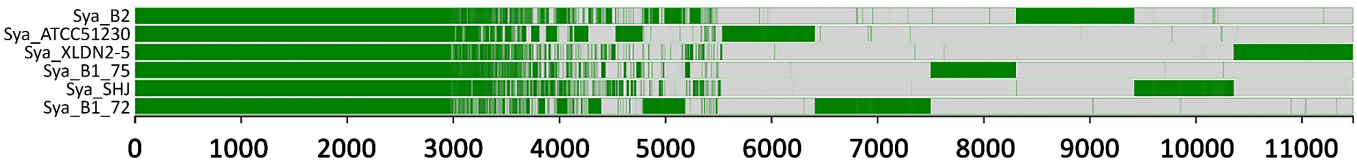

FIGURE 2

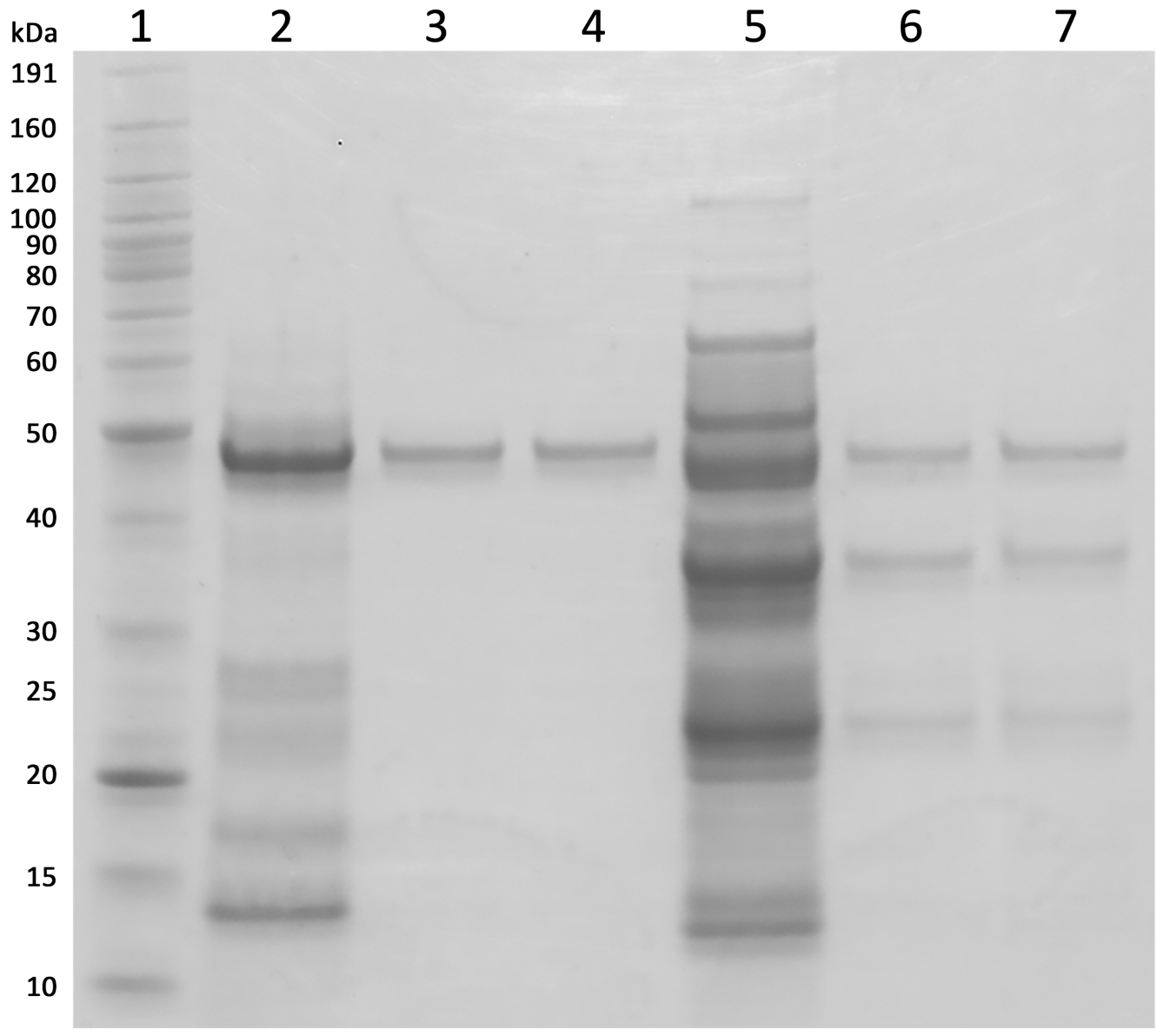

FIGURE 3

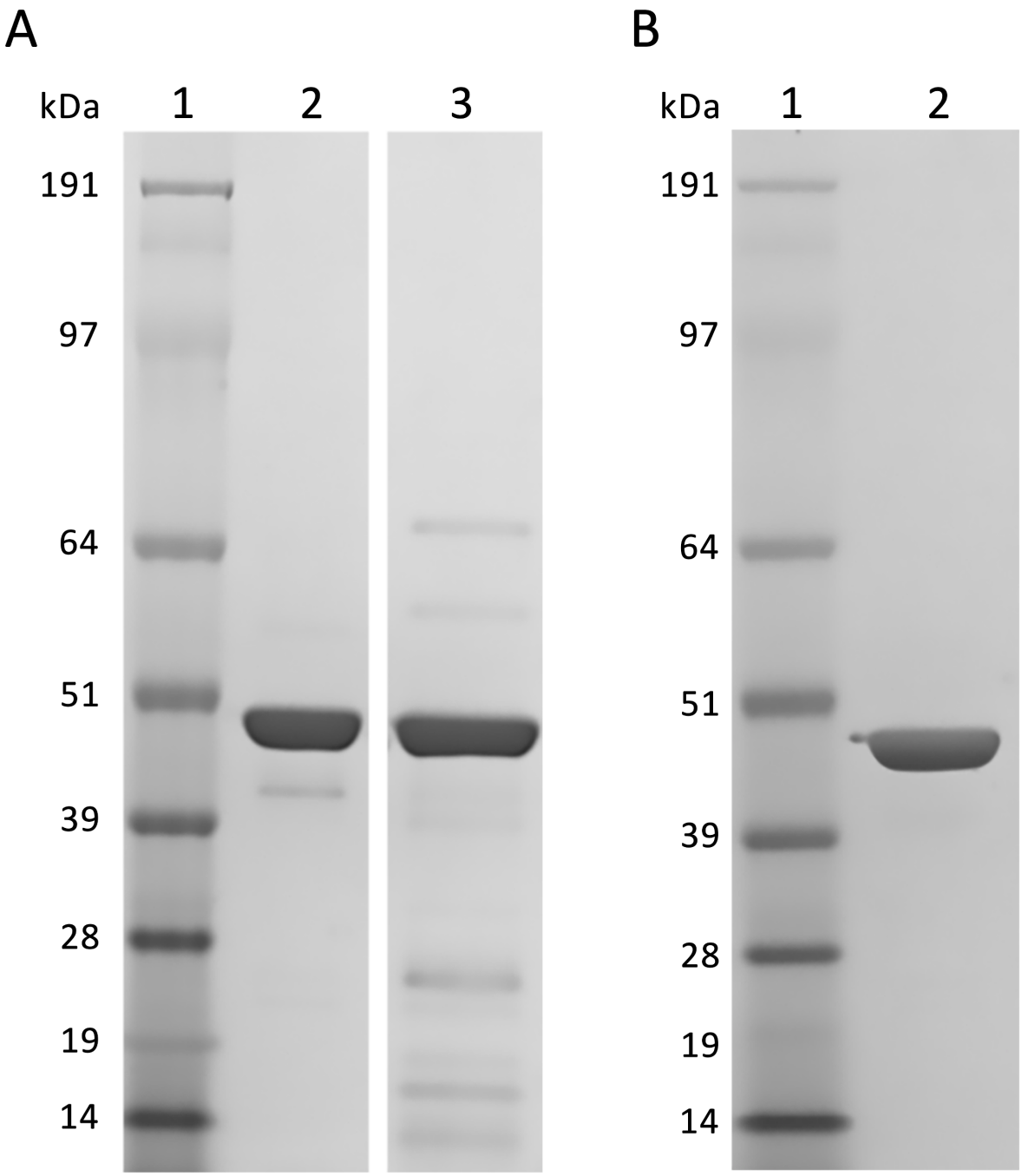

FIGURE 4

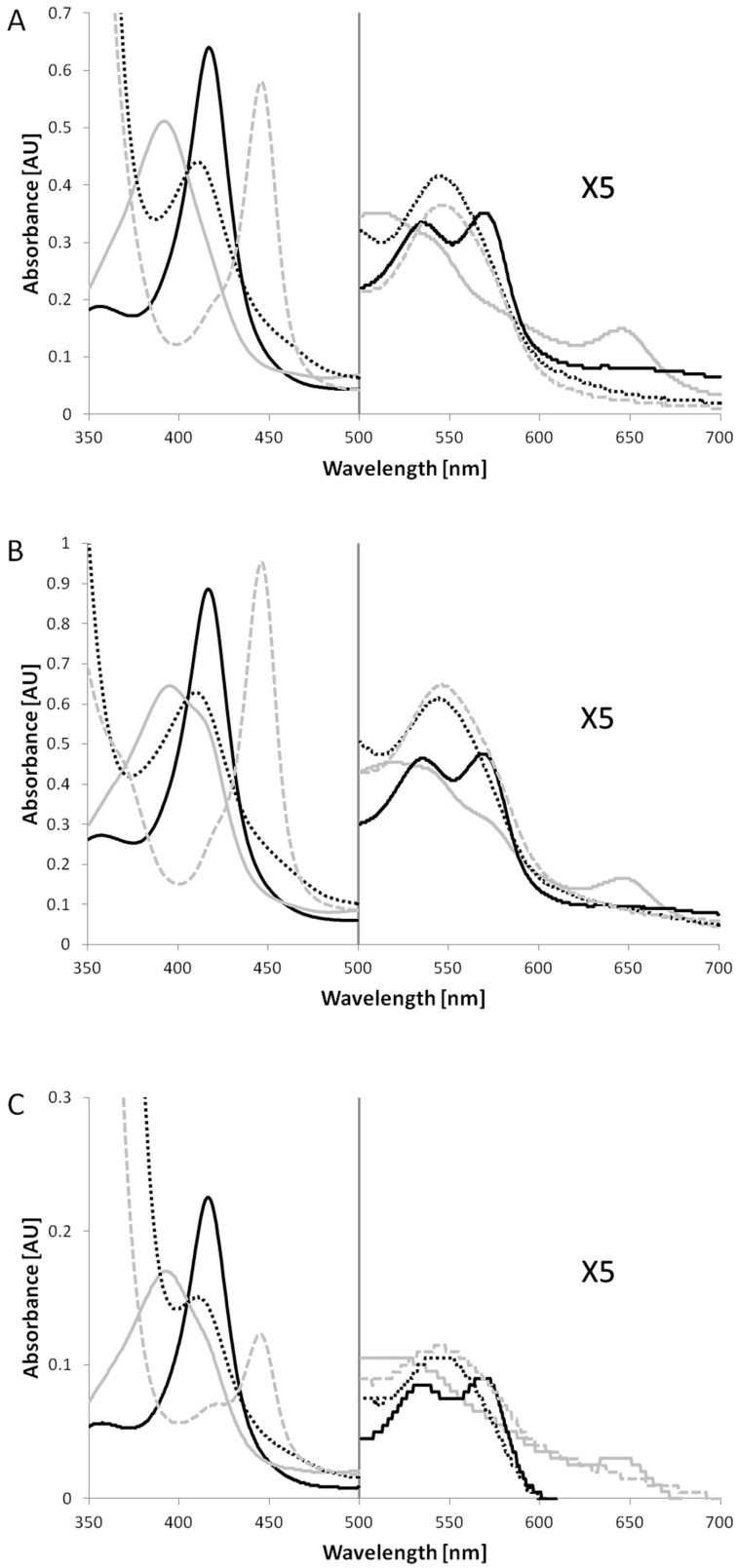

FIGURE 5

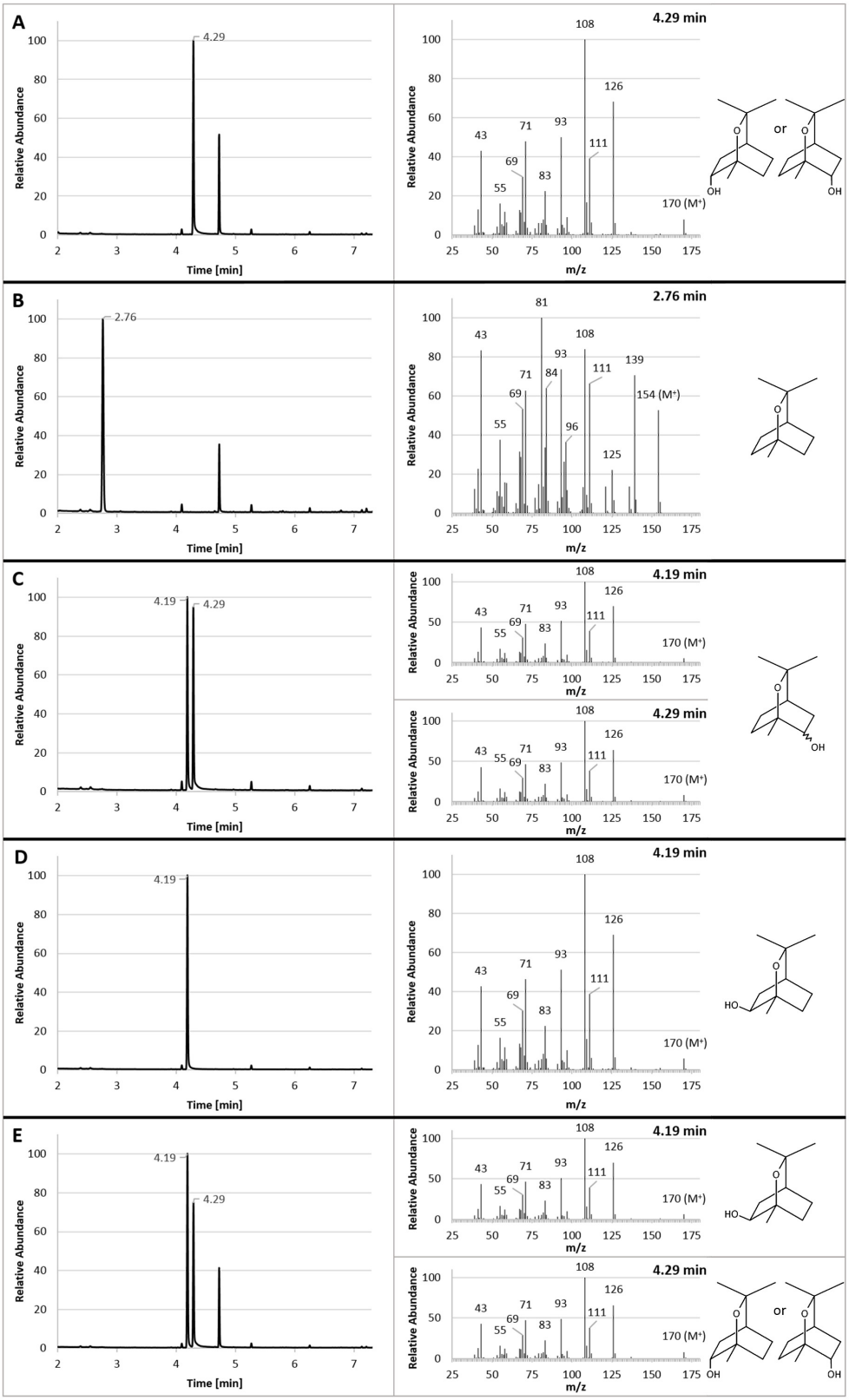

TABLE 1 K_D values for 1,8-cineole-binding in the absence and presence of 0.2 M potassium chloride in 50 mM Tris, pH 7.4. K_D values are reported as mean ± SD (n≥3).

	Dissociation constants K_D [µM]	
P450	**no KCl**	**presence of 0.2 M KCl**
CYP101J2	20.4 ± 2.6	10.1 ± 1.0
CYP101J3	22.1 ± 0.8	9.5 ± 0.5
CYP101J4	8.4 ± 0.6	5.1 ± 0.1

TABLE 2 Percentage absorbance difference induced by a range of alternative substrates in 50 mM Tris, pH 7.4, compared to ΔA_{max} induced by 1,8-cineole in the presence of 0.2 M KCl (estimated to within ca. ± 5%). Data obtained in the same buffer in the presence of 0.2 M KCl is in brackets.

Substrate	% Spin-state shift (type I)	
	CYP101J2	CYP101J3
β-ionone	40 (40)	35 (35)
(1*R*)-(+)-camphor	10 (15)	15 (15)
2-adamantanone	10 (10)	15 (15)
(1*S*)-(–)-camphor	5 (10)	5 (5)