SUPPLEMENTARY INFORMATION

TABLE S1 Top 4 BLAST hits comparing the putative and confirmed P450s encoded in the genome of *S. yanoikuyae* strain B2 to proteins in the uniprot database (http://www.uniprot.org/blast) (as per 26 Apr 2016). The four P450s which are not present in the *S. yanoikuyae* type strain (= ATCC 51230) are shaded grey. The putative enzyme names are based on the Prokka annotation prior to biochemical characterisation.

	ID number **GenBank accession no.** ***(Prokka annotation)***	**Uniprot accession no. (Source)**	**Sequence identity**
	Sya_B2_00569 KX496991 *(Cytochrome P450 monooxygenase PikC)*	A0A085K819 (*S. yanoikuyae* B1) K9CXW5 (*S. yanoikuyae* ATCC 51230) A0A0J9D1E6 (*S. yanoikuyae* B1) A0A084EQV7 (*S. yanoikuyae* B1)	99% 99% 99% 98%
	Sya_B2_01856 KX496992 *(Pentalenene oxygenase)*	A0A085K6G7 (*S. yanoikuyae* B1) A0A0J9D781 (*S. yanoikuyae* SHJ) K9CRH3 (*S. yanoikuyae* ATCC 51230) A0A084ELT3 (*S. yanoikuyae* B1)	98% 98% 98% 98%
	Sya_B2_02741 KX496993 *(Cytochrome P450-terp)*	A0A084ESR8 (*S. yanoikuyae* B1) A0A084E2B4 (*S. yanoikuyae* B1) W1S2M0 (*Sphingobium* sp. C100) A0A087NGZ3 (*Sphingobium* sp. ba1)	100% 98% 94% 93%
CYP101J4	Sya_B2_02767 KX496994 *(Camphor 5-monooxygenase)*	A0A084ESU3 (*S. yanoikuyae* B1) W1S6S1 (*Sphingobium sp*. C100) I9C229 (*Novosphingobium sp.* Rr 2-17) M5B4L7 and P00183 (*P. putida* ATCC 17453)	100% 80% 73% 45%
CYP101J3	Sya_B2_03538 KX496995 *(Camphor 5-monooxygenase)*	W1S6S1 (*Sphingobium sp*. C100) A0A084ESU3 (*S. yanoikuyae* B1) J1QPN8 (*Rhodococcus sp.* JVH1) M5B4L7 and P00183 (*P. putida ATCC* 17453)	53% 54% 47% 44%
CYP101J2	Sya_B2_03558 KX496996 *(Camphor 5-monooxygenase)*	W1S6S1 (*Sphingobium sp.* C100) A0A084ESU3 (*S. yanoikuyae* B1) I9C229 (*Novosphingobium sp.* Rr 2-17) A0A0R2WCS0 (SAR92 bacterium BACL16 MAG-120619-bin48)	79% 74% 83% 46%
	Sya_B2_04893 KX496997 *(Camphor 5-monooxygenase)*	A0A085K5M4 (*S. yanoikuyae* B1) A0A084EDI8 (*S. yanoikuyae* B1) A0A0J9D5Q1 (*S. yanoikuyae* SHJ) K9DAF9 (*S. yanoikuyae* ATCC 51230)	98% 98% 98% 97%
	Sya_B2_04918 KX496998 *(Camphor 5-monooxygenase)*	A0A085JZA6 (*S. yanoikuyae* B1) A0A084EP26 (*S. yanoikuyae* B1) A0A0J9CWL2 (*S. yanoikuyae* SHJ) K9D688 (*S. yanoikuyae* ATCC 51230)	100% 100% 100% 99%
	Sya_B2_05084 KX496999 *(Camphor 5-monooxygenase)*	A0A084E9S3 (*S. yanoikuyae* B1) K9CM96 (*S. yanoikuyae* ATCC 51230) J2D5J0 (*Sphingobium* sp. AP49) A0A0J9D5H5 (*S. yanoikuyae* SHJ)	100% 99% 99% 99%

TABLE S2 Pairwise sequence alignment of characterised members of the CYP101 family and CYP176A1 with the three new 1,8-cineole-hydroxylating P450s from *S. yanoikuyae* strain B2 using BLASTp (http://blast.ncbi.nlm.nih.gov/) showing alignment length/identities (%)/positives (%)/gaps (%).

P450	CYP101J2	CYP101J3	CYP101J4
CYP101J3	394/209 (53%)/269 (68%)/ 0 (0%)		
CYP101J4	410/302 (74%)/348 (85%)/ 0 (0%)	393/212 (54%)/273 (69%)/ 0 (0%)	
CYP101A1	393/173 (44%)/234 (59%)/ 1 (0%)	399/176 (44%)/241 (60%)/ 1 (0%)	409/179 (44%)/ 246 (60%)/ 1 (0%)
CYP101B1	399/186 (47%)/253 (63%)/ 8 (2%)	395/176 (45%)/235 (59%)/ 4 (1%)	398/180 (45%)/245 (61%)/ 6 (1%)
CYP101C1	397/157 (40%)/232 (58%)/ 9 (2%)	395/156 (39%)/218 (55%)/ 5 (1%)	395/149 (38%)/222 (56%)/ 5 (1%)
CYP101D1	401/171 (43%)/249 (62%)/ 11 (2%)	412/159 (39%)/227 (55%)/ 9 (2%)	405/160 (40%)/214 (59%)/ 7 (1%)
CYP101D2	418/175 (42%)/250 (59%)/ 11 (2%)	401/166 (41%)/231 (57%)/ 9 (2%)	410/172 (42%)/238 (58%)/ 7 (1%)
CYP176A1	349/92 (26%)/157 (44%)/ 1 (0%)	384/91 (24%)/168 (43%)/ 6 (1%)	369/105 (28%)/173 (46%)/ 3 (0%)

TABLE S3 Overview of the *S. yanoikuyae* strains for which a Whole Genome Shotgun (WGS) project is available at DDBJ/ENA/GenBank.

WGS accession prefix	Bioproject ID	Strain	Name used in this study	Size [Mb]
LVJD	PRJNA316001	B2	Sya_B2	5.9
JGVR	PRJNA241283	B1	Sya_B1_72	5.7
AFXE	PRJNA71691	XLDN2-5	Sya_XLDN2-5	5.4
JFFT	PRJNA239177	SHJ	Sya_SHJ	5.5
JPOU	PRJNA255061	B1	Sya_B1_75	5.2
AGZU	PRJNA52201	ATCC 51230	Sya_ATCC51230	5.5

TABLE S4 Relatedness of the six sequenced *S. yanoikuyae* strains and an outgroup taxon, *Sphingobium indicum* B90A (Sin_B90A = BioProject ID: PRJNA50313), calculated by counting the number of kMers (25-mers) shared between each pair of organisms.

No. of kMers	**Strain**	Sya_ B2	Sya_ B1_72	Sya_ B1_75	Sya_ SHJ	Sya_ XLDN2-5	Sya_ ATCC51230	Sin_ B90A
5904103	Sya_B2	100%	80%	80%	79%	78%	75%	46%
5681583	Sya_B1_72	76%	100%	77%	75%	74%	71%	46%
5154380	Sya_B1_75	68%	69%	100%	74%	76%	75%	45%
5496992	Sya_SHJ	73%	72%	80%	100%	76%	77%	45%
5354549	Sya_XLDN2-5	69%	69%	79%	73%	100%	72%	44%
5499826	Sya_ATCC51230	70%	69%	82%	77%	76%	100%	46%
4079365	Sin_B90A	31%	32%	35%	33%	33%	33%	100%

```
CYP101J2   -MEASVKGAAGQMTERPDNVPA-DRVFDFDIYR----DVP-EGLDFHQSWREIMRQAPHP 53
CYP101J3   ------MAEQAVLASPPSDVPA-DRIVDFDIYN----PFK-GQNDLHVAWMALRESTPHA 48
CYP101J4   -MEIGTIEPVGHDVLRPDNVPA-DRVFDFDIYR----DVP-EGSTLHHSWRALMDQVAYP 53
CYP101A1   -MTTETIQSNANLAPLPPHVPE-HLVFDFDMYN----PSN-LSAGVQEAWAVLQESNVPD 53
CYP101B1   -------------MEAPAHVPA-DRVVDIDIYM----PPGLAEHGFHKAWSDLSAGNP-A 41
CYP101C1   --------------MIPAHVPA-DRVVDFDIFN----PPG-VEQDYFAAWKTLL--DGPG 38
CYP101D1   MNAQTSTATQKHRVAPPPHVPG-HLIREIDAYD----LDG-LEQGFHEAWKRVQQPDTPP 54
CYP101D2   -MATNFDEAVRAKVERPANVPE-DRVYEIDMYA----LNG-IEDGYHEAWKKVQHPGIPD 53
CYP176A1   -MTA--------------TVASTSLFTTADHYHTPLGPDG-TPHAFFEALRDE--AETTP 42
                          *     .   * :               :

CYP101J2   LMWTPHNGGHWVALRSDLAETVMSDFERFSNHTVLVPKETAGEAYRLIPLSLDPPEHRPF 113
CYP101J3   VVWTPHNGGHWIALDPELIANVFGDSDRFSSFNVLVPKETAGEAYHFIPLSLDPPEHRPY 108
CYP101J4   VMWTPHNGGHWVVLRGDISDVVMSDSERFSNHTVLVPKETAGEAYRLIPLSLDPPAHQPF 113
CYP101A1   LVWTRCNGGHWIATRGQLIREAYEDYRHFSSECPFIPR-EAGEAYDFIPTSMDPPEQRQF 112
CYP101B1   VVWTPRNEGHWIALGGEALQEVQSDPERFSSRIIVLPK-SVGEMHGLIPTTIDPPEHRPY 100
CYP101C1   LVWSTANGGHWIAARGDVVRELWGDAERLSSQCLAVTP-GLGKVMQFIPLQQDGAEHKAF 97
CYP101D1   LVWTPFTGGHWIATRGTLIDEIYRSPERFSSRVIWVPR-EAGEAYDMVPTKLDPPEHTPY 113
CYP101D2   LIWTPFTGGHWIATNGDTVKEVYSDPTRFSSEVIFLPK-EAGEKYQMVPTKMDPPEHTPY 112
CYP176A1   IGWSEAYGGHWVVAGYKEIQAVIQNTKAFSNKGVTFPRYETGE-FELMMAGQDDPVHKKY 101
           : *:    ***:.           .   :*.    .     *:   ::    *   :  :

CYP101J2   RSLLNENLGPKPLRPIEQVVTDLAVSLIEGFRPKGRCNFTHEFAEQLPVRIFMRIVDLPV 173
CYP101J3   RKILNDNLYSSSVNPLEPKVRALTASLIDNFVANGRCDFVTEFAEQLPLRVFMQLVDLPT 168
CYP101J4   RKLLNDNLGPKPLKPVESKIVELTVSLIEGFRPNGQCDFMHEFAEKLPVQIFMQIVDLPI 173
CYP101A1   RALANQVVGMPVVDKLENRIQELACSLIESLRPQGQCNFTEDYAEPFPIRIFMLLAGLPE 172
CYP101B1   RQLLNAHLNPGAIRGLSESIRQTAVDLIEGFAAQGHCNFTAQYAEQFPIRVFMALVGIEA 160
CYP101C1   RTPVMKGLASRFVVALEPKVQAVARKLMESLRPRGSCDFVSDFAEILPLNIFLTLIDVPL 157
CYP101D1   RKAIDKGLNLAEIRKLEDQIRTIAVEIIEGFADRGHCEFGSEFSTVFPVRVFLALAGLPV 173
CYP101D2   RKALDKGLNLAKIRKVEDKVREVASSLIDSFAARGECDFAAEYAELFPVHVFMALADLPL 172
CYP176A1   RQLVAKPFSPEATDLFTEQLRQSTNDLIDARIELGEGDAATWLANEIPARLTAILLGLPP 161
           *      .       .   :   : .:::     *  :     :  :* .:   :  :

CYP101J2   EDLPKLKHLADQYTRPDGSIP------LDDVTKQFREYLRPVIEARRIKPGEDMISRMIN 227
CYP101J3   EHLPVLKQLADQFTRPDGSMT------PAEATTRFMEYVGPILNERRGSDRSDLLTAITR 222
CYP101J4   EDLPKLKHLADQFTRPDGSLT------YPEVAHLFRDYLMPVIAERRGGSGEDMISRMVN 227
CYP101A1   EDIPHLKYLTDQMTRPDGSMT------FAEAKEALYDYLIPIIEQRRQKPGTDAISIVAN 226
CYP101B1   SEAPRIRHWAECMTRPGMDMT------FDEAKAVFFDYVGPLVDARRETPGEDMISAMIN 214
CYP101C1   EDRPRLRQLGVQLTRPDGSMTV------EQLKQAADDYLWPFIEKRMAQPGDDLFSRILS 211
CYP101D1   EDATKLGLLANEMTRPSGNTPEEQGRSLEAANKGFFEYVAPIIAARRGGSGTDLITRILN 233
CYP101D2   EDIPVLSEYARQMTRPEGNTPEEMATDLEAGNNGFYAYVDPIIRARVGGDGDDLITLMVN 232
CYP176A1   EDGDTYRRWVWAITHVENPEE------GAEIFAELVAHARTLIAERRTNPGNDIMSRVIM 215
           ..           *:                      :   .:  *      * :: :

CYP101J2   GEVG-GRPLTDIEAENICIQVLVGGLDTVVNMLGFTFSHLAKDHALRRAIAADPSLIDDA 286
CYP101J3   GEVF-GRPLTDDEALRMAIQVMVGGLDTVVNFMSFTIQLLAQAPDIQDRLASDRSIYSAA 281
CYP101J4   GQVG-GRELTDVEAQNICMQVLVGGLDTVVNLLSFTLSWLAQDHDLRRALVADPALIDDA 286
CYP101A1   GQVN-GRPITSDEAKRMCGLLLVGGLDTVVNFLSFSMEFLAKSPEHRQELIERPERIPAA 285
CYP101B1   ADLGDGRRLTRDEALSVVTQVLIAGLDTVVNVLGFIMRELAGNPALRADLRQRGADILPV 274
CYP101C1   EPVG-GRPWTVDEARRMCRNLLFGGLDTVAAMIGMVALHLARHPEDQRLLRERPDLIPAA 270
CYP101D1   VEID-GKPMPDDRALGLVSLLLLGGLDTVVNFLGFMMIYLSRHPETVAEMRREPLKLQRG 292
CYP101D2   SEIN-GERIAHDKAQGLISLLLLGGLDTVVNFLSFFMIHLARHPELVAELRSDPLKLMRG 291
CYP176A1   SKID-GESLSEDDLIGFFTILLLGGIDNTARFLSSVFWRLAWDIELRRRLIAHPELIPNA 274
             :  *.         .   ::..*:*... .:.     *:        :

CYP101J2   LLEFFRRFPVVSSAREVLRDQEFEGVLLKAGDMVMAPTVVVAMDDARNEDPLEFRLGR-- 344
CYP101J3   INEALRRLPLVSSGRELRVDTEVDGVILRKGDMIIAPTELVALNPRMNEDPLRYDLNR-- 339
CYP101J4   LMEFLRRFPVVSSSREVRKDIEFEGVHLKSGDMVMAPTIVVTLDEDNNADPLDFRLGR-- 344
CYP101A1   CEELLRRFSLVADGRILTSDYEFHGVQLKKGDQILLPQMLSGLDERENACPMHVDFSR-- 343
CYP101B1   VHELFRRFGLVSIAREVRRDIEFHGVHLKAGDMIAIPTQVHGLDPRVNPDPLAIDPSR-- 332
CYP101C1   ADELMRRYPTVAVSRNAVADVDADGVTIRKGDLVYLPSVLHNLDPASFEAPEEVRFDRGL 330
CYP101D1   VEELFRRFAVVSDARYVVSDMEFHGTMLKEGDLILLPTALHGLDDRHHDDPMTVDLSR-- 350
CYP101D2   AEEMFRRFPVVSEARMVAKDQEYKGVFLKRGDMILLPTALHGLDDAANPEPWKLDFSR-- 349
CYP176A1   VDELLRFYGPAMVGRLVTQEVTVGDITMKPGQTAMLWFPIASRDRSAFDSPDNIVIER-- 332
             * :*    .  .*    :       :: *:       :   :      *      *

CYP101J2   KARQHSTFGKGSHTCPGAHLARMEMKVVLREWFARIPEFRIEDDAPLRYSNGIVGSVKPF 404
CYP101J3   KVRNHSVFGSGAHTCPGQFLARLEMKIFLEEWFDRIPSFELEPGQVLRHRGGIVGGCEPF 399
CYP101J4   KARKHSTFGRGSHTCPGAHLARMEMKIVLREWLARIPEFRLADGAELRFTNGIVGSVKPF 404
CYP101A1   QKVSHTTFGHGSHLCLGQHLARREIIVTLKEWLTRIPDFSIAPGAQIQHKSGIVSGVQAL 403
CYP101B1   KRARHSTFGSGPHMCPGQELARKEVAITLEEWLRRIPDFALGPNSDLSPVPGIVGALRRV 392
CYP101C1   APIRHTTMGVGAHRCVGAGLARMEVIVFLREWLGGMPEFALAPDKAVTMKGGNVGACTAL 390
CYP101D1   RDVTHSTFAQGPHRCAGMHLARLEVTVMLQEWLARIPEFRLKDRAVPIYHSGIVAAVENI 410
CYP101D2   RSISHSTFGGGPHRCAGMHLARMEVIVTLEEWLKRIPEFSFKEGETPIYHSGIVAAVENV 409
CYP176A1   TPNRHLSLGHGIHRCLGAHLIRVEARVAITEFLKRIPEFSLDPNKECEWLMGQVAGMLHV 392
               *  :. * * * *  * * *  : : *::  :*.* :          * *..   .

CYP101J2   VLEWPV------ 410
CYP101J3   VIRWPRQ----- 406
CYP101J4   MLEWDV------ 410
CYP101A1   PLVWDPATTKAV 415
CYP101B1   ELVWNT------ 398
CYP101C1   PLVWRA------ 396
CYP101D1   PLEWEPQRVSA- 421
CYP101D2   PLVWPIAR---- 417
CYP176A1   PIIFPKGKRLSE 404
            : :
```

FIGURE S1 Multiple sequence alignment (http://www.ebi.ac.uk/Tools/msa/clustalo/) of CYP101J2, CYP101J3 and CYP101J4 with previously characterised members of the CYP101 family and CYP176A.

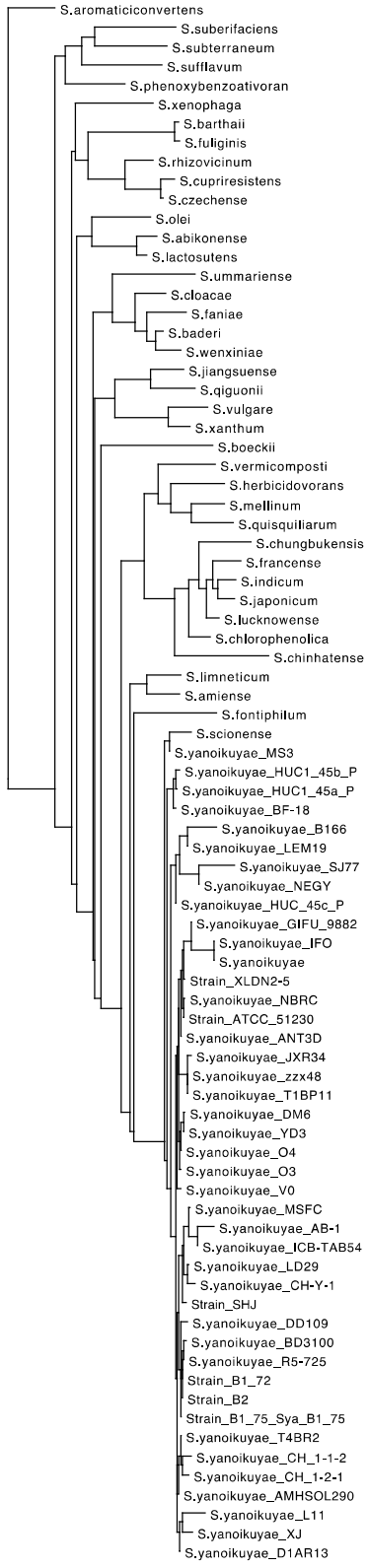

FIGURE S2 Phylogenetic relationship (based on the 16S rRNA gene inferred using the NJ method) between selected strains in the genus *Sphingobium*. Included in this analysis are 16s rRNA gene sequences representing each of the accepted type species in the genus obtained from LPSN (http://www.bacterio.net/) and 16S rRNA gene sequences of *S. yanoikuyae* strains available on GenBank (at 15 Dec 2015) and extracted from the draft genome sequences for each of the strains shown in Figure 1.

Additional experimental details:

Production of *S. yanoikuyae* B2 cells

A 2 L stirred tank bioreactor A containing 1.6 L of DM supplemented with 25 g L^{-1} glucose and 0.5 mL L^{-1} of 1,8-cineole was inoculated with a seed culture to attain an OD_{600} of 0.25. The temperature was maintained at 30 °C and the pH automatically controlled at pH 7.0 using 10% (w/v) H_3PO_4 or 10% (w/v) NH_4OH solution. The dissolved oxygen was controlled at 30% of saturation using a two-step cascade control which was used to maintain the dissolved oxygen above the specified set point. The agitator speed ranged from 200 rpm to 1200 rpm and airflow (supplemented with 30% pure O_2) ranged from 0.09 L min^{-1} to 1.5 L min^{-1}. Foaming was controlled via the automatic addition of 10% (v/v) polypropylene glycol 2000; 2.5 mL L^{-1} of the antifoam solution was added prior to inoculation. When the OD_{600} reached 28, glucose and 1,8-cineole feeds were started at rates of 3.5 g h^{-1} and 0.32 mL h^{-1}, respectively. At OD_{600} 34, 4 mL of 1 M $MgSO_4.7H_2O$ was added aseptically to the culture. At OD_{600} 63, the glucose feed was turned off and the 1,8-cineole feed increased to 3.2 mL h^{-1} for a further 5 h. The final OD_{600} was 66. A concentrated buffer stock (20x) was added to the culture to 1x (final concentration of 50 mM Tris pH 7.4, 50 mM KCl, 1 mM EDTA, 100 µM phenylmethylsulfonyl fluoride, 0.5 mM DL-dithiothreitol (DTT)) and aliquots were frozen at -80 °C without further processing.

Protein purification from *S. yanoikuyae* B2

Potential P450s were purified from *S. yanoikuyae* strain B2 cell extracts using a 3-step purification procedure. Cell aliquots were thawed and disrupted by 3 passages through a homogeniser (APV LAB2000). A clarified crude cell extract was obtained by centrifugation at 38400 *g* for 30 min at 4 °C and 0.2 µm filtration (Sartopore 2 300, Sartorius Stedim) and then fractionated using solid $(NH_4)_2SO_4$ to obtain 0-20%, 20-40% and 40-60% $(NH_4)_2SO_4$ cuts. The proteins were then purified using an ÄKTA Purifier UPC 100 (GE Healthcare Life

Sciences). Protein precipitates were dissolved in 50 mM Tris, 0.5 mM DTT, pH 7.4 (buffer A) and then loaded onto a 5 mL HiTrap Q HP ion exchange chromatography (IEX) column (GE Healthcare Life Sciences) equilibrated using buffer A. Proteins were eluted with a linear potassium chloride gradient over 20 column volumes up to 500 mM KCl. Eluted protein was detected at 280 nm and 417 nm (typical Soret absorbance of P450s). 1,8-cineole binding was tested by addition of 1,8-cineole to fractions absorbing at 417 nm. Two peak fractions exhibiting a typical type I absorbance shift, peak fraction 1 (PF 1) and peak fraction 2 (PF 2), from IEX of the 40-60% ammonium sulphate precipitation were further purified using a HiLoad Superdex 200 PG (GE Healthcare Life Sciences) gel filtration (GF) column equilibrated with 50 mM Tris, 150 mM KCl, 0.5 mM DTT, pH 7.4 at a flow rate of 1 mL min^{-1}. Protein purity after IEX and GF was determined using sodium dodecyl sulphate-polyacrylamide gel electrophoresis.

Partial sequencing of 1,8-cineole binding P450s

PF1 was concentrated using a centrifugal device with a 10 kDa molecular weight cut-off (Nanosep 10K Omega, Pall Life Sciences). It was then reduced with ca. 10 mM DTT in 7 M guanidine HCl, 0.1 M Tris, 1 mM EDTA, pH 9.5 for 2 h at 37 °C and alkylated with iodoacetic acid (ca. 15 mM final concentration) for 30 min at 37 °C in the dark before the alkylation was ended by the addition of further DTT (final concentration ca. 20 mM). Reduced, carboxymethylated (RCM) PF1 was recovered by reversed-phase high-performance liquid chromatography using a Brownlee C4 cartridge; the required fraction was identified by matrix-assisted laser desorption ionisation (MALDI)-MS. N-terminal sequencing of this fraction was performed on an Applied Biosystems Procise sequencer using a solid support (glass fibre disc) following traditional Edman degradation techniques. In addition, protein identity for PF 1 was confirmed by tryptic digest of the reduced and alkylated protein followed by MALDI-MS fingerprint mapping. RCM PF1 (ca. 20 µg) was dissolved in 100 µL of 1% NH_4HCO_3 and to this solution 4 µg of Sigma sequencing grade trypsin was added; the mixture was incubated overnight at 37 °C, dried in a SpeediVac and dissolved in 30 µL of

70% acetonitrile, 0.1% trifluoroacetic acid. The digest was then subjected to MALDI-MS using 2,5-dihydroxybenzoic acid and α-cyano-4-hydroxycinnamic acid as matrices. Protein identity for PF2 was determined by an in-gel tryptic digest (1) followed by MALDI-MS fingerprint mapping.

Reference:

1. **Jiménez CR, Huang L, Qiu Y, Burlingame AL.** 2001. In-gel digestion of proteins for MALDI-MS fingerprint mapping. Current protocols in protein science. **14:**16.4:16.4.1–16.4.5.