SUPPLEMENTARY MATERIAL

General and Materials

General reagents and chemical substrates of analytical grade or higher quality were from Alfa-Aesar, Fisher Scientific and Sigma-Aldrich or their subsidiary companies. Solvents of HPLC quality were from Rathburn Chemicals (UK) and subsidiaries of Sigma-Aldrich and Merck. Buffer components were from Anachem, UK. NADPH (tetrasodium salt) was from Apollo Scientific and Melford Laboratories. Isopropyl-β-D-thiogalactopyranoside (IPTG) was from Melford Laboratories. Restriction enzymes, T4 DNA ligase and the associated buffers were from New England Biolabs. Taq and KOD Polymerases were from Merck Biosciences. Competent and supercompetent *E. coli* strains were from Stratagene. Oligonucleotides were from MWG Biotech, Germany. General molecular biology manipulations were carried out according to literature methods.[1] Amplified genes were fully sequenced by Geneservice, UK on an ABI 3730XL Prism DNA sequencer. UV/visible spectra and enzyme activity assays were run at 30 ºC on a Varian Cary 50 spectrophotometer. Gas chromatography (GC) was carried out on Thermo Finnigan Trace and 8000 Top instruments equipped with flame-ionisation detectors (FIDs) using DB-1 fused silica capillary columns and helium as the carrier gas. The injectors were maintained at 200 °C or 250 °C and the FIDs at 250 °C.

Directed evolution and screening procedures

A *Spe* I restriction site was introduced downstream of the haem domain-coding region of the CYP102A1 gene in pGLW11[2] using oligonucleotide 5′-GCTCATAATAC-GCCGCTACTAGTGCTATACGGTTCAAATATG-3′ and its reverse complement, resulting in silent mutations at residues 482 and 483. Error-prone PCR was carried out between this site and an *Eco*R I site upstream of the haem domain-coding region using the forward and reverse primers 5′-TCTCGAGAATTCATAATCATCGGAGAC-GCC-3′ and 5′-TGGATCCACTAGTAGCGGCGTATTATGAGC-3′. Libraries were constructed from wild-type CYP102A1 (WT) and mutant F87A templates under conditions designed to introduce 1-3 mutations per 1,000 bp according to the Stratagene GeneMorph protocol employed. Genes were amplified by 30 cycles of strand separation at 94 °C for 60 sec, annealing at 45 °C for 90 sec and extension at 68 °C for 110 sec + 2 sec per cycle. After digestion with *Eco*R I and *Spe* I, short fragments were reincorporated into pGLW11 (*Spe* I WT variant) using T4 DNA

Ligase, transformed into *E. coli* DH5α competent cells and grown for 36 hours on Luria-Bertani (LB) agar plates. Low levels of P450 expression occur from this vector without extraneous induction. Around 1500 colonies were screened. Those showing indigo formation[3,4] were isolated, transferred to fresh plates and grown for a further 36 hours to minimise false positives prior to DNA sequencing. Random-priming recombination was used to shuffle 11 variants representing 16 new mutations of possible interest.[5] The protocol given by Volkov and Arnold[6] was modified at stage 9, where Taq and KOD Polymerases were employed with 2 μL $MgSO_4$ rather than *Pfu* Polymerase. PCR was carried out on the assembly strands as described above, but using KOD Polymerase. Samples were digested, ligated and plated out as before. Of ~ 800 colonies, some 130 displayed indigo formation, and these were assayed *in vivo* against two substrates, naphthalene and propylbenzene. 3 mL LB growths containing 50 $\mu L \cdot mL^{-1}$ ampicillin were induced by adding IPTG to 0.5 mM, and after 1 hour made 1 mM in substrate. Following a further 36 hours of shaking, the whole cell turnovers were extracted into ethyl acetate and scrutinised for product formation levels and oxidation profile alterations using GC. A larger scale *in vivo* assay was then carried out on 12 selected variants in which the cells from induced 50-mL LB growths were harvested by centrifugation and resuspended in 50 mL *E. coli* Minimal Media prior to substrate addition. Four of these showed particular promise and were taken forward for purification and *in vitro* testing.

Protein expression and purification

Variants of interest were transferred into the pET28 vector using *Nco* I and *Bam*H I so that expression levels could be more tightly controlled. 30 $mL.L^{-1}$ of an overnight culture of *E. coli* JM109(DE3) harbouring the plasmid was inoculated into LB medium containing 0.4% (v/v) glycerol and 30 $mg.L^{-1}$ kanamycin and grown at 37 °C with shaking at 180 rpm to OD_{600} >1. Protein expression was induced by adding IPTG to 0.4 mM. After a further 12 h of growth at 30 °C cells were harvested by centrifugation. The red-brown pellet from each 1 L growth was re-suspended in 25 mL 40 mM potassium phosphate, pH 7.4, 1 mM in dithiothreitol (phosphate buffer). The cells were lysed by sonication, and cell debris was cleared by centrifugation at 37,500g for 30 min at 4 °C. The supernatant was loaded onto an Amersham-Pharmacia DEAE fast flow Sepharose column (200 x 50 mm) pre-equilibrated with phosphate buffer from which the protein was eluted using a linear gradient of 80-400 mM ammonium sulphate in phosphate buffer. The red P450 fractions were collected and concentrated by ultrafiltration, desalted using a Sephadex G-25 column pre-

equilibrated with phosphate buffer, and re-concentrated by ultrafiltration. The solution was centrifuged at 9,250g for 5 min at 4 °C and filter sterilised. FPLC anion-exchange purification was carried out on an Amersham-Pharmacia Source-Q column (120 x 26 mm) using a linear gradient of 0-30% 15x phosphate buffer. Fractions with $A_{404}/A_{280} > 0.25$ were collected, concentrated by ultrafiltration and filter sterilised before being stored at –20 °C in 50% (v/v) glycerol. Glycerol and salts were removed from proteins immediately prior to experiments using an Amersham Pharmacia 5 mL PD-10 column pre-equilibrated with 50 mM Tris buffer, pH 7.4.

DNA manipulations

Variants R47L/Y51F (RLYF) and F87A were prepared as described.[2] The RLYF couplet was introduced into variants KT2 and A330P using *Nco* I and *Afl* II in standard cloning procedures.[1] F87A was inserted into KT2 using the Stratagene Quik-Change kit and transformed into *E. coli* XL1 Blue Supercompetent cells, the mutagenic primers being 5′- GCAGGAGACGGGTTAGCGACAAGCTGGACGC-3′ and its reverse complement.

Table S1. Variant names, mutations and amino acid changes.

RLYF [a]	R47L	CGT → TTA
	Y51F	TAC → TTT
GVQ [a]	A74G	GCG → GGG
	F87V	TTT → GTT
	L188Q	CTG → CAG
KT2	A191T	GCA → ACA
	N239H	AAC → CAC
	I259V	ATT → GTT
	A276T	GCG → ACG
	L353I	CTA → ATA

[a] Site-directed mutagenesis variants

A330P	A330P	GCG → CCG
KSK19 [b]	F87A	TTT → GCG
	H171L	CAT → CTT
	Q307H	CAA → CAT
	N319Y	AAC → TAC
KT5 [b]	(P8P)	CCA → CCG
	F87A	TTT → GCG
	A330P	GCG → CCG
	E377A	GAG → GCG
	D425N	GAT → AAT

[b] Directed evolution variants from F87A template

NADPH turnover rate determinations

NADPH turnovers were run in 1250 µL of 50 mM oxygenated Tris, pH 7.4 at 30 ℃, containing 0.25 µM enzyme, 125 µg bovine liver catalase and 1 mM substrate added as a 100 mM stock in DMSO. Protein concentration was determined as described.[2] Assays were held at 30 ℃ for 1 min prior to NADPH addition as a 20 mg.mL^{-1} stock to a final concentration of ~320 µM (equivalent to 2AU). A period of ten seconds was allowed to elapse after NADPH addition before the absorbance decay at 340 nm was measured (Fig. S2c). The NADPH consumption rate was derived using ε_{340} = 6.22 mM^{-1} cm^{-1}, after correction for the leak rates recorded under the same conditions in the presence of substrate-free DMSO. Coupling in turnovers that failed to complete (specifically, those involving WT and F87A with pentane, octane and toluene) was

determined by running additional assays at enzyme concentrations of up to 2.5 μM. Because NADPH rates were lower under such conditions, the PFRs quoted for these turnovers were not achieved in practice. All data are means of at least three experiments with standard deviations less than 5% of the mean.

Product analysis

3 μL of internal standard (100 mM in DMSO) was added to 1000 μL of each completed turnover prior to extraction into 400 μL ethyl acetate or chloroform. Centrifugation was carried out at 19,000g for 3½ minutes in 1500 μL microcentrifuge tubes. Products were identified by matching the GC elution times observed to those of authentic samples. FID responses were calibrated using a representative equivalent for each product group on the assumption that isomeric mono-oxygenated products would give comparable responses (Table S2). Samples containing a range of concentrations of the chosen product and 1% v/v in DMSO in 50mM Tris, pH 7.4 were extracted as above. The integrated peak areas derived were expressed as ratios of the internal standard peak areas and plotted against product concentration.

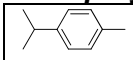
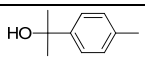
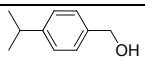
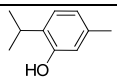
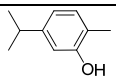
Table S2. Calibrants, internal standards and oven temperatures used in GC product analysis.

Substrate	Calibrant	Internal Standard	Oven temperature, [Column length]
Naphthalene	1-naphthol	4-benzylphenol	Held at 100 °C for 1 min then raised at 15 °C min^{-1} to 220 °C [7 m]
Propylbenzene	1-phenyl-1-propanol		Held at 60 °C for 1 min then raised at 15 °C min^{-1} to 150 °C [7 m]
p-Cymene	*p*-*α*,*α*-trimethyl benzylalcohol		
Toluene	*o*-cresol		
Octane	2-octanol		Held at 40 °C for 1 min then raised at 15 °C min^{-1} to 130 °C [7 m]
Pentane	3-pentanol	2-octanol	Raised from 70 °C to 90 °C at 1 °C min^{-1} then raised at 65 °C min^{-1} to 220 °C [60 m]

Spin shifts

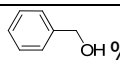
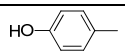
Substrate-induced spin state shifts were determined by pre-equilibrating 500 μL of 50 mM Tris (pH 7.4) containing the requisite variant for 2 min at 30 °C, adding 1 μL of substrate stock solution (100 mM in DMSO) and equilibrating for a further 1 min before recording the altered spectrum. The low-spin haem content was taken to be 100% initially and a notional 45% when A_{418} and A_{390} were at parity to compensate for the disparity in extinction coefficients. A linear relationship was then assumed over the intervening range and extrapolated where necessary. All data rounded to the nearest 5%.

Table S3. *In vitro* activity, selectivity and spin shifts of CYP102A1 and variants with *p*-cymene.

	Spin shift (%)	NADPH rate	PFR	C (%)	 % i-Pr-ol	 % Me-ol	 % thymol	 % carv	others
WT	10	467	168	36	82	2	3	7	6
A330P	< 5	1040	270	26	35	19	19	16	11
RLYF/A330P	10	1825	475	26	42	28	12	8	10
F87A	50	413	157	38	92	–	–	–	8
KSK19	65	2721	1442	53	76	2	–	–	22
GVQ	40	2799	784	28	71	1	–	–	28
KT5	40	1403	702	50	76	2	–	–	22

Rates in nmol·min^{-1}·(nmol P450)$^{-1}$. C = coupling. Products were *p*-α,α-trimethylbenzylalcohol (i-Pr-ol, 4.16 min), 4-isopropylbenzylalcohol (Me-ol, 5.07 min), thymol (5.10 min) and carvacrol (carv, 5.17 min).

Table S4. *In vitro* activity and selectivity of CYP102A1 and variants with toluene.

	NADPH rate	PFR	Coupling (%)	 % Benzylalcohol	 % *o*-cresol	 % *p*-cresol
WT	29	2.7	9.4	2	98	–
RLYF/KT2	368	136	37	3	95	2
A330P	189	85	45	1	98	1
RLYF/A330P (*)	363	189	52	1	97	1
F87A	11	0.1	0.8	22	78	–
KT5	217	6.3	2.9	95	5	–

Rates in nmol·min^{-1}·(nmol P450)$^{-1}$. Products were benzylalcohol (2.79 min), *o*-cresol (2.97 min) and *p*-cresol (3.19 min). (*) Percentages do not sum to 100 due to a minority product.

Table S5. *In vitro* activity, selectivity and spin shifts of CYP102A1 and variants with pentane.

Variant	Spin shift (%)	NADPH rate	PFR	Coupling (%)	% 3-one	% 2-ol	% 3-ol
WT	10	74	16	21	2	59	39
R47L/Y51F	5	361	162	45	–	58	42
RLYF/KT2	10	2010	1206	60	–	63	37
A330P	~ 0	959	623	65	–	62	38
RLYF/A330P	~ 0	1766	1183	67	–	63	37
GVQ	10	2107	590	28	1	47	52

Rates in nmol·min^{-1}·(nmol P450)$^{-1}$. Products were 3-pentanone (3-one, 12.1 min), 2-pentanol (2-ol, 12.4 min) and 3-pentanol (3-ol, 12.5 min).

Table S6. *In vitro* activity and selectivity of CYP102A1 and variants with octane.

Variant	NADPH rate	PFR	Coupling (%)	% 2-ol	% 3-ol	% 4-ol	% 3- & 4-octanones
WT	148	53	36	15	43	42	–
A330P	642	167	26	53	30	16	1
F87A	181	51	28	10	38	43	9
KT5	519	83	17	12	35	25	28

Rates in nmol·min^{-1}·(nmol P450)$^{-1}$. Products were 2-octanol (2-ol, 3.72 min), 3-octanol (3-ol, 3.66 min), 4-octanol (4-ol, 3.6 min) and 3- & 4-octanones (3.57 & 3.44 min).

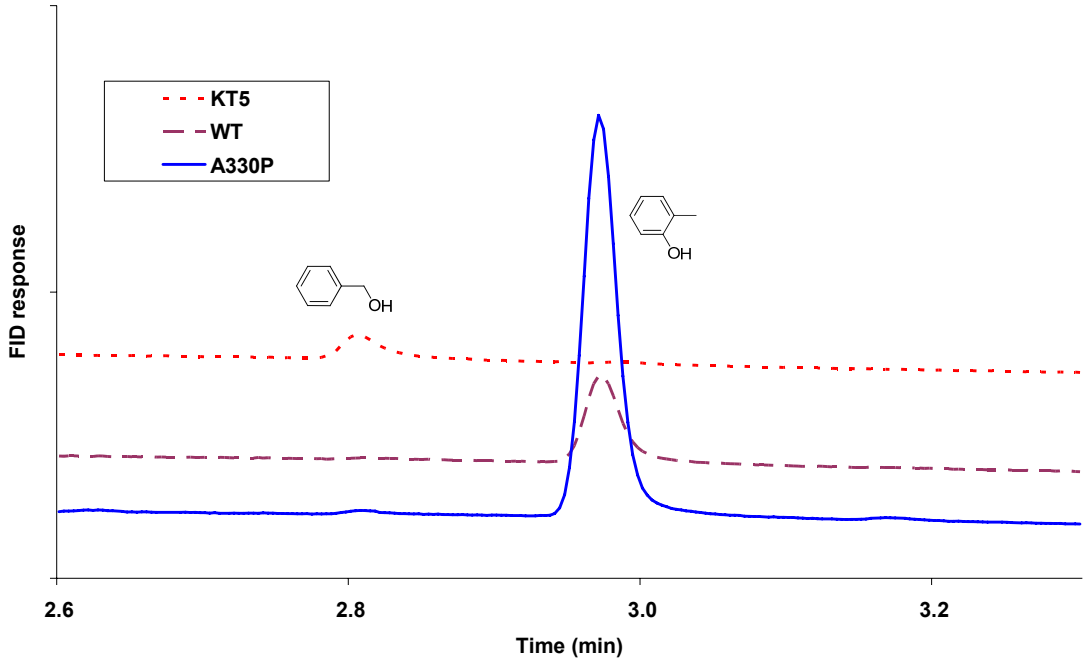


Fig. S1a Gas chromatograms of toluene oxidation by CYP102A1 and variants.

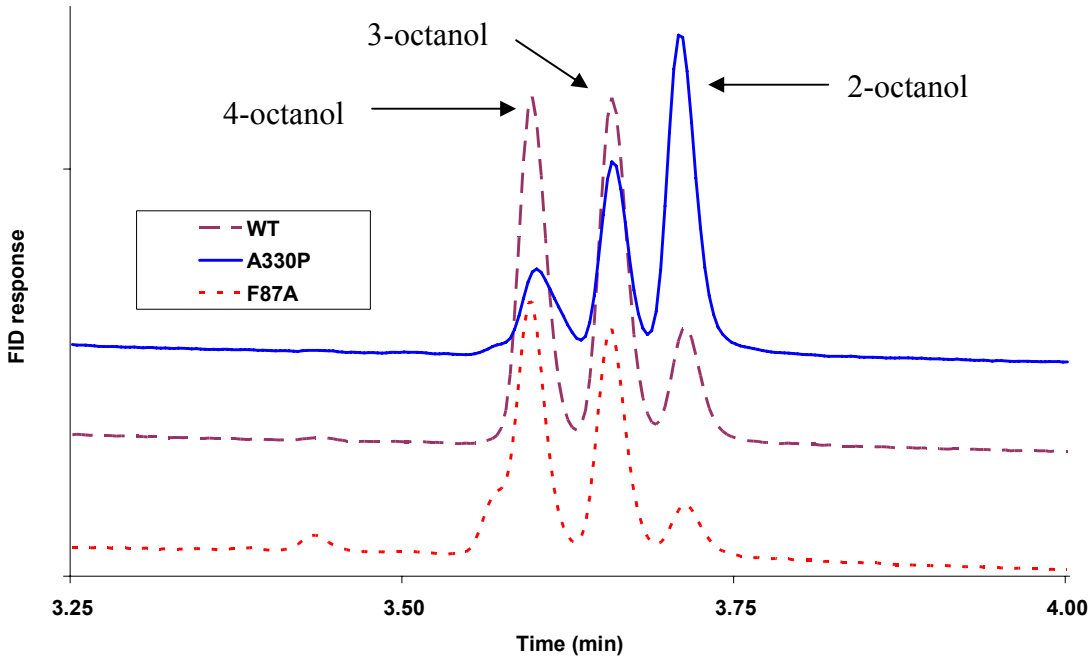


Fig. 1b Gas chromatograms of octane oxidation by CYP102A1 and variants.

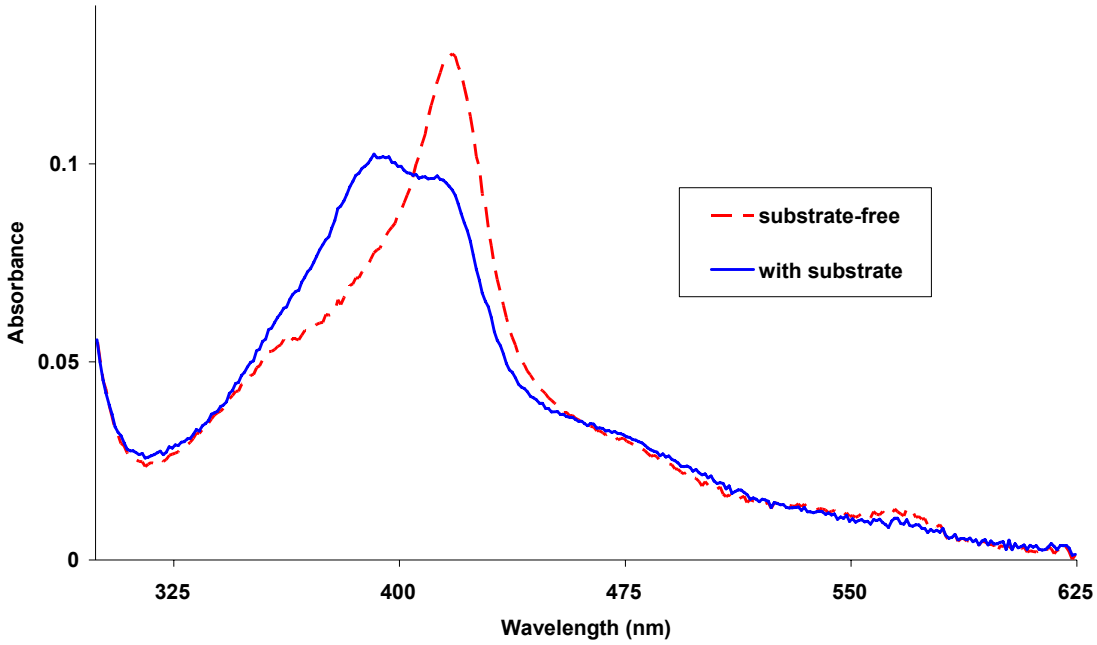


Fig. S2a Spin-shift response of KSK19 on addition of *p*-cymene.

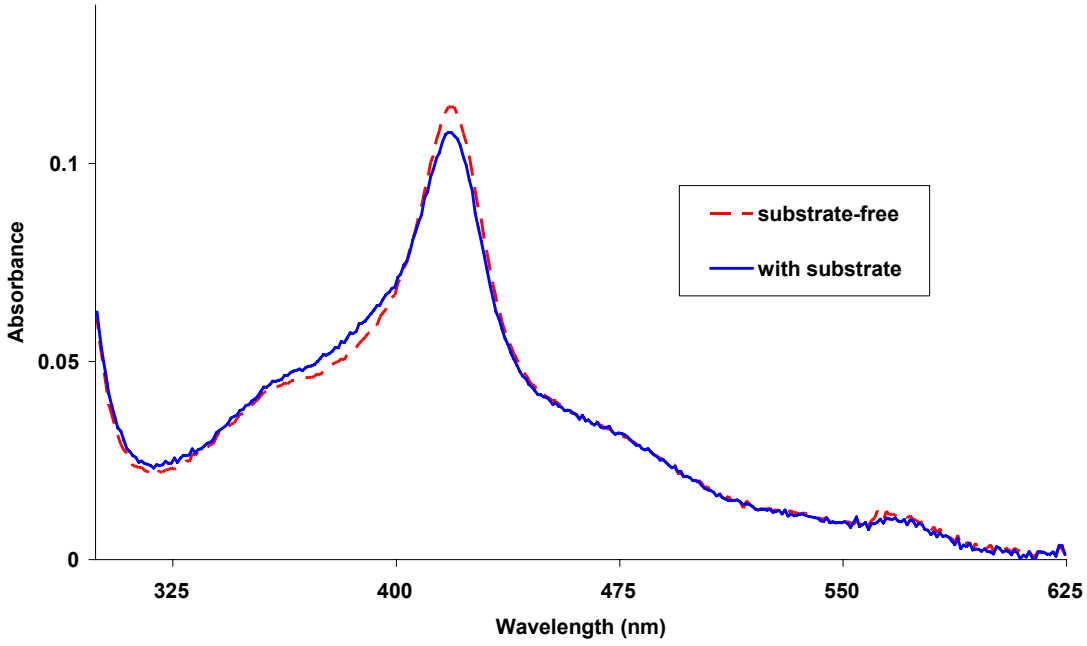


Fig. S2b Spin-shift response of RLYF/A330P on addition of *p*-cymene.

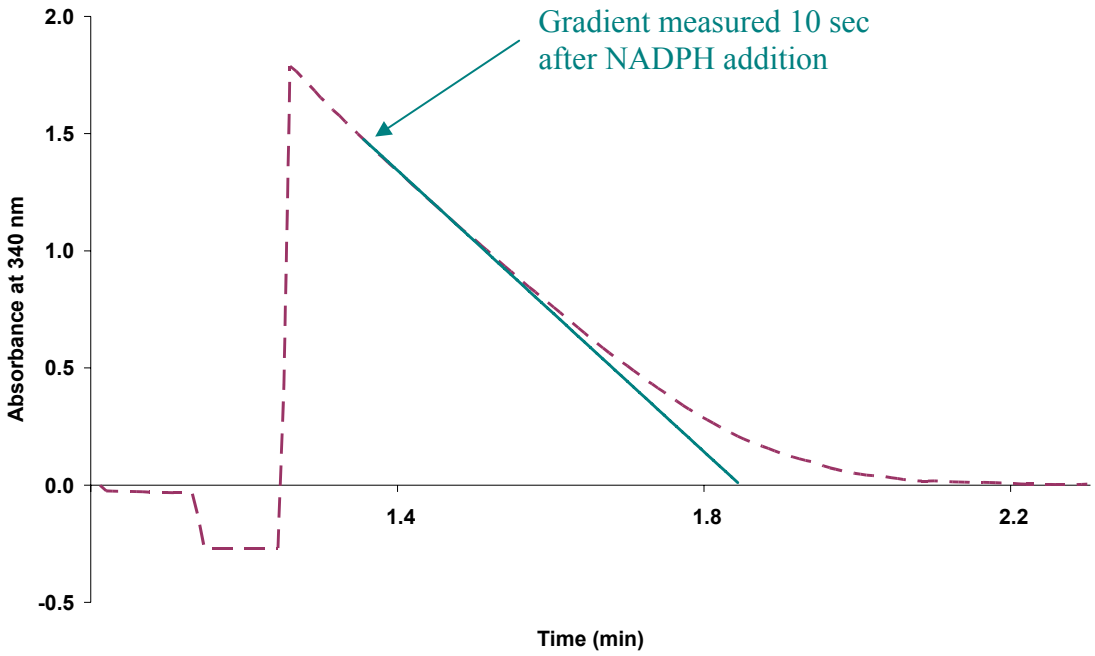


Fig. S2c NADPH turnover trace of RLYF/A330P with *p*-cymene.

References

1. J. Sambrook, E. F. Fritsch and T. Maniatis, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, New York, 1989.
2. A. B. Carmichael and L. L. Wong, *Eur. J. Biochem.*, 2001, **268**, 3117-3125.
3. 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**, 469-472.
4. Q. S. Li, U. Schwaneberg, P. Fischer and R. D. Schmid, *Chemistry*, 2000, **6**, 1531-1536.
5. Z. Shao, H. Zhao, L. Giver and F. H. Arnold, *Nucleic Acids Res.*, 1998, **26**, 681-683.
6. A. A. Volkov and F. H. Arnold, *Methods Enzymol.*, 2000, **328**, 447-456.