Evolved CYP102A1 ($P450_{BM3}$) variants oxidise a range of non-natural substrates and offer new selectivity options†

Christopher J. C. Whitehouse, Stephen G. Bell, Henry G. Tufton, Richard J. P. Kenny, Lydia C. I. Ogilvie and Luet-Lok Wong*

Received (in Cambridge, UK) 23rd November 2007, Accepted 19th December 2007
First published as an Advance Article on the web 11th January 2008
DOI: 10.1039/b718124h

The evolution of CYP102A1 variants with enhanced activity and altered specificity characteristics.

Cytochrome P450 (CYP) monooxygenases have a number of physiological roles including the biosynthesis of endogenous compounds and secondary metabolites, and the metabolism of drugs and xenobiotics.[1,2] They typically catalyse the insertion of an oxygen atom from atmospheric dioxygen into the C–H bonds of a variety of organic compound types, with potential applications that include the synthesis of fine chemicals and the preparation of drug metabolites. CYP102A1 ($P450_{BM3}$) from *Bacillus megaterium*, a fatty acid hydroxylase,[3] is one of the more extensively researched and engineered members of the CYP superfamily.[4] The wild-type enzyme (WT) turns over most substrates other than fatty acids at desultory rates. Site-specific mutagenesis and directed evolution have been employed, both separately and in tandem, to improve activity towards aromatics,[5,6] alkanes[7,8] and also pharmaceuticals,[9] a number of which are accepted as substrates by WT.[10] The challenge of selectivity control has also been addressed, and compounds not produced by WT have been successfully targeted.[11] Various screening procedures have been developed for directed evolution.[12] Indigo formation *via* indole oxidation,[13] one of the simplest, was first used in CYP102A1 to monitor saturation mutagenesis at specific sites.[14] By applying it as a preliminary activity screen and then sub-screening *in vivo* for selectivity as well as activity, we have been able to identify new variants with much enhanced activity (Fig. 1), two of which also have the ability to alter product profiles.

Error-prone PCR was carried out on the full haem domain-coding regions of the wild-type CYP102A1 gene and mutant F87A. Indigo formation provided a simple assay for identifying variants with improved performance potential. These were subjected to gene shuffling and a second round of indigo screening. Those that gave rise to blue colonies were assayed *in vivo* for naphthalene and propylbenzene oxidation activity. Product selectivity was also monitored. Four of the most promising were purified: A330P, a single mutation variant; A191T/N239H/I259V/A276T/L353I (KT2); F87A/H171L/Q307H/N319Y (KSK19); and F87A/A330P/E377A/D425N (KT5). All showed good stability characteristics, with little or no activity loss after fifteen months' storage at −20 °C. Naphthalene and propylbenzene were again used to assess the purified enzymes *in vitro*. WT, the F87A mutant and the highly active A74G/F87V/L188Q (GVQ) variant developed by Schmid and co-workers[14] were also tested as benchmarks.

All four new variants enhanced the product formation rate (PFR) of the wild-type enzyme with naphthalene[6] (3.1 nmol min^{-1} (nmol P450)$^{-1}$—henceforth abbreviated to min^{-1}) by at least an order of magnitude, A330P being the most effective at 155 min^{-1} (Table 1). In all cases 1-naphthol was the only GC-detectible product. WT was more active towards propylbenzene[15] at 615 min^{-1}, with 71% coupling of NADPH consumption to product formation, but the new variants again gave higher PFRs, KT2 being the fastest at 2205 min^{-1} (Table 1). WT and KT2 yielded ~99% 1-phenyl-1-propanol, but the other new variants directed oxidation away from the benzylic position. A330P gave 30% *o*-propylphenol, while variant KT5, in which mutations F87A and A330P occur in combination, gave 78% 1-phenyl-2-propanol, leveraging the selectivity bias of mutation F87A, which formed 53% 1-phenyl-2-propanol when acting alone (Fig. 2). Changes in haem spin state on substrate addition were determined for these and other substrates but generally correlated poorly with NADPH rates (Tables 1, S3 and S5†).

The R47L/Y51F (RLYF) couplet, which is known to increase CYP102A1 activity towards several substrates,[5,16] was then incorporated into variants A330P and KT2. In both cases NADPH consumption rates were further enhanced. With naphthalene, RLYF/A330P gave a PFR of 666 min^{-1}, the highest observed, while RLYF/KT2 gave the highest rate

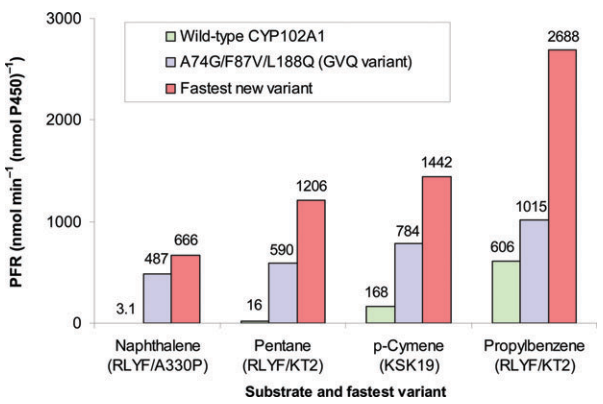

Fig. 1 *In vitro* activity of CYP102A1 variants with various substrates.

Department of Chemistry, University of Oxford, Inorganic Chemistry Laboratory, South Parks Road, Oxford, UK OX1 3QR. E-mail: luet.wong@chem.ox.ac.uk; Fax: +44 1865 272690
† Electronic supplementary information (ESI) available: Experimental details and additional data tables. See DOI: 10.1039/b718124h

Table 1 *In vitro* activity, selectivity and spin shifts of CYP102A1 and variants with naphthalene and propylbenzene[a]

	Naphthalene				Propylbenzene[b]						
	Spin shift (%)	NADPH rate	PFR	C (%)	Spin shift (%)	NADPH rate	PFR	C (%)	1-ol (%)	2-ol (%)	Phenol (%)
WT and generic accelerators											
WT	10	80	3.1	3.9	10	866	615	71	99	—	1
R47L/Y51F (RLYF)	5	166	32	19	5	2157	1682	78	99	1	—
KT2	20	490	113	23	25	2756	2205	80	98	1	1
RLYF/KT2	15	1306	496	38	15	3126	2688	86	99	—	—
Phe87 variants											
F87A	10	106	5.5	5.2	20	670	248	37	46	53	—
F87A/KT2	10	370	32	8.7	30	1664	566	34	54	45	—
KSK19	20	511	56	11	40	2079	1081	52	53	46	—
KT5	35	567	31	5.5	40	1062	701	66	20	78	—
GVQ	20	2116	487	23	30	3172	1015	32	77	14	8
A330P variants											
A330P	<5	483	155	32	<5	1810	706	39	68	2	30
RLYF/A330P	5	1306	666	51	5	2497	674	27	65	5	30[c]

[a] Rates in nmol min^{-1} (nmol P450)$^{-1}$. C = coupling (total product formation as % of NADPH consumption). Percentages do not always sum to 100 on account of minority products. 1-Naphthol was the only GC-detectible product formed in naphthalene turnovers. Means of at least three turnovers to within ±10%. [b] Products were 1-phenyl-1-propanol (1-ol), 1-phenyl-2-propanol (2-ol) and *o*-propylphenol (Phenol). [c] Includes 7% *p*-propylphenol.

with propylbenzene at 2688 min^{-1}. This approaches the rate reported for WT with fatty acid substrates (3200 min^{-1}).[17] Product profiles were little altered relative to the first generation variants, though RLYF/A330P gave some *p*-propylphenol. KT2 was also combined with mutation F87A to give a second generation variant capable of offering F87A-directed product profiles at higher rates (Table 1).

The range of regioselective outcomes obtained in propylbenzene turnovers led us to investigate other alkylbenzenes. The naturally occurring hydrocarbon, *p*-cymene is one of the more interesting as it is a precursor to four distinct flavouring compounds. WT (PFR 168 min^{-1}) gave predominantly *p*,α,α-trimethylbenzyl alcohol, along with small quantities of the two possible aromatic hydroxylation products, thymol (3%) and carvacrol (7%), and just 2% 4-isopropylbenzyl alcohol (Table S3†). Variant A330P, by contrast, yielded 19% thymol, 16% carvacrol and 19% 4-isopropylbenzyl alcohol. The new variants also enhanced activity, with KSK19 giving a PFR of 1442 min^{-1} (Fig. 1). A more dramatic selectivity shift was observed with toluene, KT5 yielding 95% benzyl alcohol and just 5% *o*-cresol where WT gave almost exclusively *o*-cresol

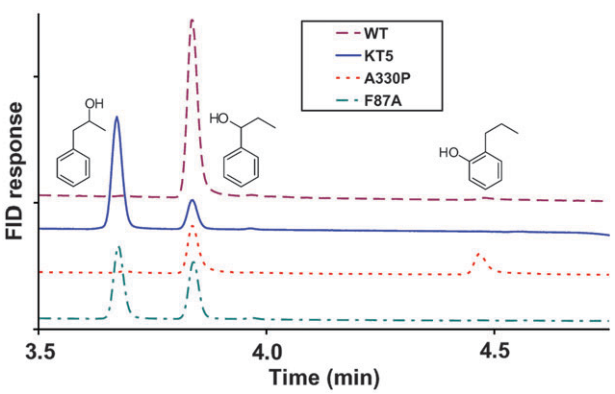

Fig. 2 Gas chromatography analysis of propylbenzene oxidation.

(98%) (Table S4, Fig. S1a†). Variant RLYF/A330P also gave 98% *o*-cresol, but at a PFR of 189 min^{-1} *versus* 2.7 min^{-1} for WT, coupling efficiency rising from 9.4% to 52%.

Although the new variants were selected on the basis of their affinity for aromatic substrates, KT2 and A330P also displayed high activity levels with pentane, particularly in combination with RLYF. RLYF/KT2 and RLYF/A330P gave NADPH rates of 2010 min^{-1} and 1766 min^{-1}, respectively, with 60% and 67% coupling *versus* 74 min^{-1} and 21% for WT (Table S5†). The GVQ variant also gave a rapid NADPH rate (2107 min^{-1}), though with coupling of just 28%. An NADPH turnover rate of 3250 min^{-1} with pentane has been reported by Arnold and co-workers for variant 139-3, though no PFR data are available.[18] A330P variants did not show altered pentane product profiles relative to WT, but selectivity shifts were observed with octane, A330P giving 53% 2-octanol *versus* 15% for WT, which yielded primarily 3- and 4-octanol (Table S6, Fig. S1b†). An identical level of 2-octanol formation (53%) was reported for variant 9-10A, a derivative of 139-3, rising to 89% when three additional mutations (V78F, A82S and A328F) were designed into the active site.[7,19]

KT2 thus increased NADPH consumption rates and coupling relative to WT across a range of non-natural substrates while leaving product profiles almost unaltered. F87A/KT2 and KSK19 displayed the same pattern of behaviour relative to base mutation F87A, increasing PFRs while giving F87A-like product distributions. The mutations comprising KT2 and F87A-free KSK19 may therefore be classified as generic rate accelerators for the substrates studied. KSK19 is generally more active than F87A/KT2, implying that its F87A-free derivative may prove more potent than KT2. The mutated residues in KT2 and KSK19 are all outside the active site (Fig. 3),[20] though Leu353 lies next to a substrate access channel residue, Met354, while Ala191 is close to the channel entrance.[21] KT2 mutations N239H and I259V have been reported previously.[22,23]

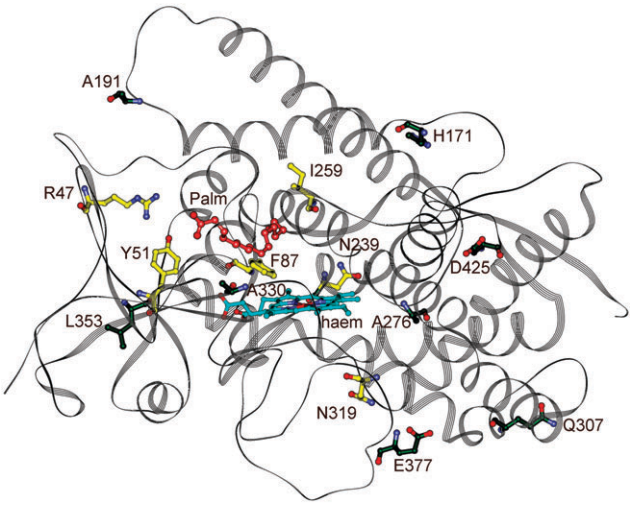

Fig. 3 The locations of the mutation sites.[20]

Variant A330P displayed more complex behaviour, always enhancing NADPH rates, but coupling significantly better than WT with certain substrates and less well with others. Though not part of the active site, this mutation frequently altered selectivity patterns, generally in a different sense to F87A. When in combination with F87A, however (as in KT5), rather than giving A330P-like product profiles it gave F87A-like profiles with selectivity shifts that exceeded those given by other F87A variants. The crystal structure of the A330P haem domain may help to elucidate the structural consequences of introducing a proline residue directly next to Pro329, a substrate access channel residue. It will be interesting to explore the effect of alternative substitutions at Ala330, and to establish how A330P combines with selectivity-directing mutations such as A264G, I263A, A82L and A328V.[5,7,24]

In summary, our screening protocol located CYP102A1 variants with significantly enhanced activity levels across an array of substrates. Typically, they improved coupling efficiencies and also combined well with site-specific mutations, attributes not shared by all activity-enhancing mutations.[4] The amino acid substitutions involved showed little overlap with those discovered using other methods, suggesting that directed evolution, as applied to CYP102A1, has yet to exhaust its potential. The new variants should make good starting-points for further protein evolution studies. A330P, with its striking selectivity characteristics and enhanced product formation rates, could be particularly useful in the pursuit of altered product profiles, *e.g.* in the development of phenol-forming variants[25] or terminal alkane oxidases.[26] The unusual selectivity patterns displayed by KT5 suggest that the F87A/A330P couplet could also make a promising evolutionary platform.

This work was supported by the EPSRC and BBSRC, UK (Grant ref. EP-D048559-1).

Notes and references

1 *Cytochrome P450: Structure, Mechanism and Biochemistry*, ed. P. R. Ortiz de Montellano, Kluwer Academic/Plenum Publishers, New York, 3rd edn, 2005.
2 *The Ubiquitous Roles of Cytochrome P450 Proteins*, ed. A. Sigel, H. Sigel and R. K. O. Sigel, John Wiley & Sons, Chichester, *Metal Ions in Life Sciences*, 2007, vol. 3.
3 M. J. Cryle, R. D. Espinoza, S. J. Smith, N. J. Matovic and J. J. De Voss, *Chem. Commun*., 2006, 2353–2355.
4 S. G. Bell, N. Hoskins, C. J. C. Whitehouse and L. L. Wong, in *The Ubiquitous Roles of Cytochrome P450 Proteins*, ed. A. Sigel, H. Sigel and R. K. O. Sigel, John Wiley & Sons, Chichester, *Metal Ions in Life Sciences*, 2007, vol. 3, ch. 14, pp. 437–476.
5 A. B. Carmichael and L. L. Wong, *Eur. J. Biochem*., 2001, **268**, 3117–3125.
6 Q. S. Li, J. Ogawa, R. D. Schmid and S. Shimizu, *Appl. Environ. Microbiol*., 2001, **67**, 5735–5739.
7 M. W. Peters, P. Meinhold, A. Glieder and F. H. Arnold, *J. Am. Chem. Soc*., 2003, **125**, 13442–13450.
8 R. Fasan, M. M. Chen, N. C. Crook and F. H. Arnold, *Angew. Chem., Int. Ed*., 2007, **46**, 8414–8418.
9 B. M. Vugt-Lussenburg, E. Stjernschantz, J. Lastdrager, C. Oostenbrink, N. P. Vermeulen and J. N. Commandeur, *J. Med. Chem*., 2007, **50**, 455–461.
10 G. Di Nardo, A. Fantuzzi, A. Sideri, P. Panicco, C. Sassone, C. Giunta and G. Gilardi, *J. Biol. Inorg. Chem*., 2007, **12**, 313–323.
11 V. B. Urlacher, S. G. Bell and L. L. Wong, in *Modern Biooxidation*, ed. R. D. Schmid and V. B. Urlacher, Wiley-VCH, Weinheim, 2007, ch. 4, pp. 99–122.
12 K. L. Tee and U. Schwaneberg, *Comb. Chem. High Throughput Screening*, 2007, **10**, 197–217.
13 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.
14 Q. S. Li, U. Schwaneberg, P. Fischer and R. D. Schmid, *Chemistry*, 2000, **6**, 1531–1536.
15 Q. S. Li, J. Ogawa, R. D. Schmid and S. Shimizu, *FEBS Lett*., 2001, **508**, 249–252.
16 V. B. Urlacher, A. Makhsumkhanov and R. D. Schmid, *Appl. Microbiol. Biotechnol*., 2006, **70**, 53–59.
17 J. H. Capdevila, S. Wei, C. Helvig, J. R. Falck, Y. Belosludtsev, G. Truan, S. E. Graham-Lorence and J. A. Peterson, *J. Biol. Chem*., 1996, **271**, 22663–22671.
18 A. Glieder, E. T. Farinas and F. H. Arnold, *Nat. Biotechnol*., 2002, **20**, 1135–1139.
19 P. Meinhold, M. W. Peters, M. M. Chen, K. Takahashi and F. H. Arnold, *ChemBioChem*, 2005, **6**, 1765–1768.
20 H. Li and T. L. Poulos, *Nat. Struct. Biol*., 1997, **4**, 140–146.
21 M. D. Paulsen and R. L. Ornstein, *Proteins: Struct. Funct. Genet*., 1995, **21**, 237–243.
22 P. C. Cirino and F. H. Arnold, *Angew. Chem., Int. Ed*., 2003, **42**, 3299–3301.
23 US Pat., 2003100744, 2003. World Pat., 2003008563, 2003.
24 R. J. Sowden, S. Yasmin, N. H. Rees, S. G. Bell and L. L. Wong, *Org. Biomol. Chem*., 2005, **3**, 57–64.
25 T. S. Wong, N. Wu, D. Roccatano, M. Zacharias and U. Schwaneberg, *J. Biomol. Screening*, 2005, **10**, 246–252.
26 O. Lentz, A. Feenstra, T. Habicher, B. Hauer, R. D. Schmid and V. B. Urlacher, *ChemBioChem*, 2005, **7**, 345–350.