

In the format provided by the authors and unedited.

Engineering highly functional thermostable proteins using ancestral sequence reconstruction

Yosephin Gumulya^{1,5}, Jong-Min Baek¹, Shun-Jie Wun¹, Raine E. S. Thomson¹, Kurt L. Harris¹, Dominic J. B. Hunter ¹, James B. Y. H. Behrendorff ¹, Justyna Kulig², Shan Zheng¹, Xueming Wu³, Bin Wu⁴, Jeanette E. Stok¹, James J. De Voss ¹, Gerhard Schenk¹, Ulrik Jurva², Shalini Andersson ², Emre M. Isin^{2,6}, Mikael Bodén ¹, Luke Guddat¹ and Elizabeth M. J. Gillam ^{1*}

¹School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia. ²CVRM DMPK, Cardiovascular, Renal and Metabolism, IMED Biotech Unit, AstraZeneca, Gothenburg, Sweden. ³School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing, Jiangsu, China.

⁴College of Biotechnology and Pharmaceutical Engineering, Nanjing Tech University, Nanjing, Jiangsu, China. ⁵Present address: Commonwealth

Scientific and Industrial Research Organisation, Floreat, Western Australia, Australia. ⁶Present address: UCB Biopharma SPRL, Braine-l'Alleud, Belgium.

*e-mail: e.gillam@uq.edu.au

Supplementary information for:

Engineering highly functional thermostable proteins using ancestral sequence reconstruction

Yosephine Gumulya^{1†}, Jong-Min Baek¹, Shun-Jie Wun¹, Raine E. S. Thomson¹, Kurt L. Harris¹, Dominic J.B. Hunter¹, James B.Y.H. Behrendorff¹, Justyna Kulig², Shan Zheng¹, Xueming Wu³, Bin Wu⁴, Jeanette E. Stok¹, James J. De Voss¹, Gerhard Schenk¹, Ulrik Jurva², Shalini Andersson², Emre M. Isin^{2‡}, Mikael Bodén¹, Luke Guddat¹ and Elizabeth M. J. Gillam^{1*}

¹School of Chemistry and Molecular Biosciences, The University of Queensland, Australia.

² Cardiovascular and Metabolic Diseases, Innovative Medicines and Early Development, AstraZeneca, Gothenburg, Sweden.

³ School of Pharmacy, Nanjing University of Chinese Medicine, Nanjing, Jiangsu 210023, China.

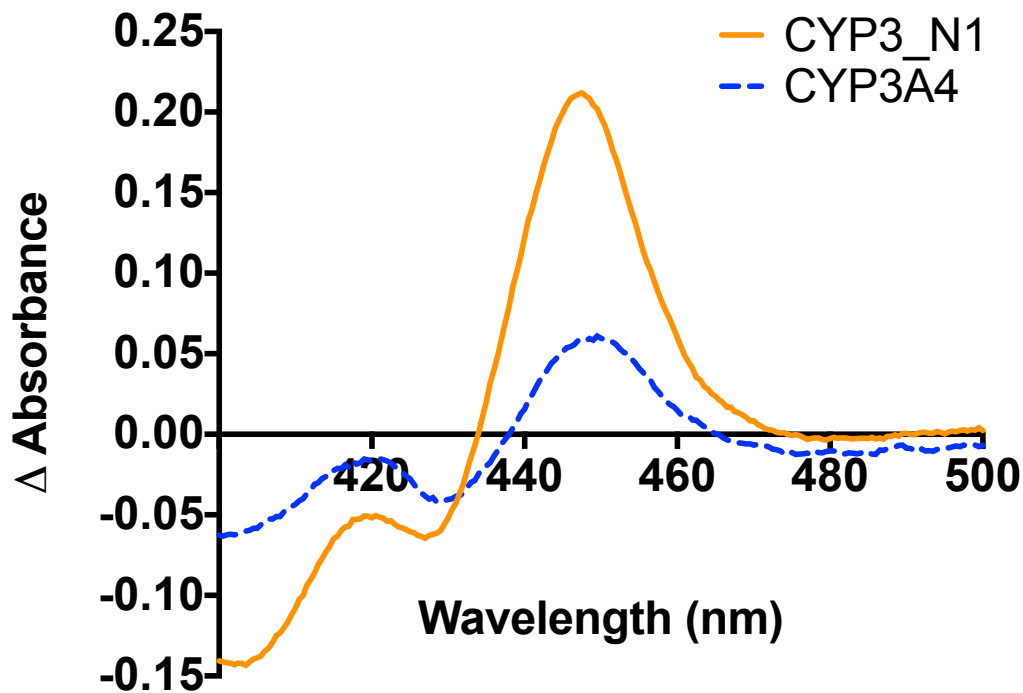
⁴ College of Biotechnology and Pharmaceutical Engineering, Nanjing Tech University, Nanjing, Jiangsu 211816, China.

[†]Current address: Commonwealth Scientific and Industrial Research Organisation (CSIRO), Floreat WA 6014, Australia.

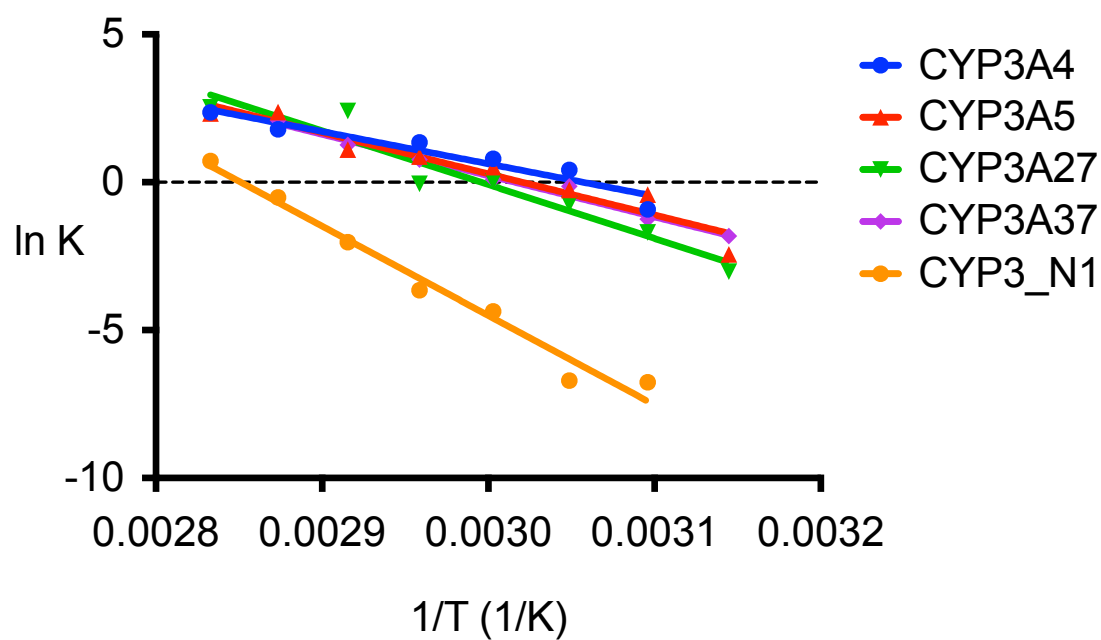
[‡] Current address: UCB BioPharma S.P.R.L. Chemin du Foriest, B-1420 Braine-l'Alleud, Belgium.

* Author to whom correspondence should be addressed

Supplementary Figures



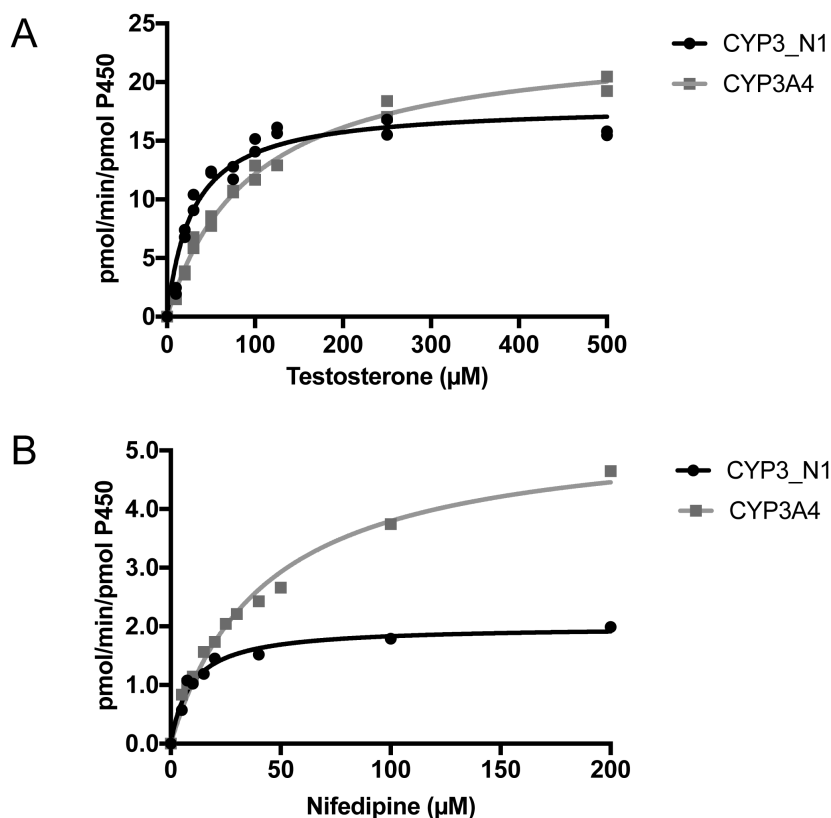
Supplementary Figure 1. The ancestral CYP3_N1 sequence expresses a properly folded P450 haemoprotein. The ancestral (CYP3_N1, solid line) and extant human (CYP3A4, dotted line) enzymes were expressed in *E. coli*. The membrane fraction was isolated from recombinant bacterial cultures and P450 content was quantified by Fe(II).CO vs. Fe(II) difference spectroscopy. The peak at ~450 nm in the CO-bound, reduced form indicates folded P450 containing an intact haem-thiolate linkage. The smaller peak at ~417 nm is characteristic of endogenous bacterial haemoproteins.



Supplementary Figure 2: Arrhenius plot for inactivation of CYP3_N1 and selected extant CYP3 enzymes.

Bacterial cells expressing each enzyme were heated to temperatures between 25 °C and 75 °C for 60 min, then cooled for 5 min at 4 °C and equilibrated to room temperature for 5 min prior to quantification of residual P450 content by Fe(II).CO *vs.* Fe(II) difference spectroscopy.

The data are means from three biological replicates. The data was fitted by linear regression using GraphPad Prism.

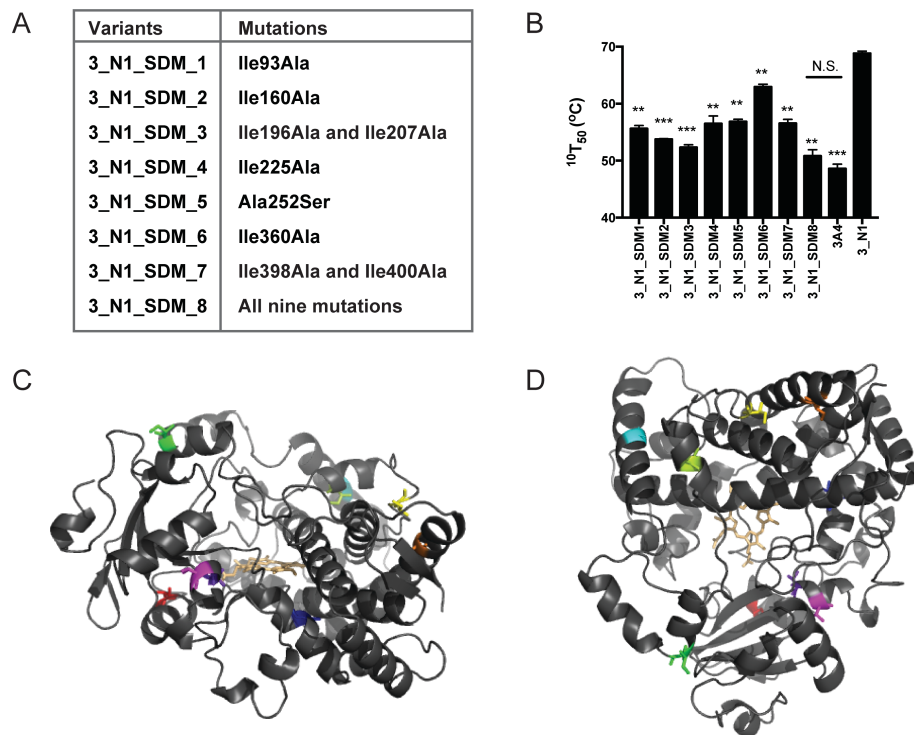


Supplementary Figure 3. Comparison of the catalytic efficiency of CYP3_N1 and CYP3A4 towards testosterone 6β-hydroxylation (A) and nifedipine oxidation (B). A.

Testosterone hydroxylation was carried out with 10 – 500 μM substrate and 50 nM P450 (co-expressed in bacterial membranes with hCPR) in 100 mM phosphate buffer (pH 7.4) at 37 °C for 10 min. B. Nifedipine oxidation was carried out similarly but with 5 – 200 μM substrate and 100 nM P450 in bacterial membranes, in 100 mM Tris buffer (pH 7.4) for 5 min. Product formation had been previously established to be proportional to time and enzyme concentration under these conditions. In each case metabolites were quantified by HPLC with reference to an internal standard (progesterone and halazepam for assays with testosterone and nifedipine respectively) and a standard curve prepared with authentic metabolite. Data represent individual measurements and lines of best fit were estimated by nonlinear regression in GraphPad Prism using the Michaelis-Menten model. Kinetic parameters are listed in Supplementary Table 5.

CYP3A27	MALLLAVFLTLIVVYGYWPGVFTKMGIPGPKPLPYFGTMLEYKGFNFDTCECFQKYGR	60
CYP3A4	MALLLAVFLVLLLYGTHSHGLFKKLGIPTPLPPLGNILSYHKGFCMFDMCHKKYGK	60
CYP3A5	MALLLAVFLVLLLYGTRTHGLFKRLGIPTPLPLGNVLSYRQGLWKFDTECYKKYGK	60
CYP3A37	-MALLLIIVALLIAGTWPFGLFKKLGIPTPLPFFGTCLFYRKGFLEFDTCECFKKYGK	59
CYP3_N1	MALLLAVFLTLILYGIWPGVGLFKKLGIPTPLPFFGTFLFYRKGFLEFDTCECFKKYGK	60
	:*.:.*: ** .*:.*:** ***:*. *.*:.*: ** ***:***:	
CYP3A27	IWGIYDGRQPVLCIMDKSMIKTVLIKECYNIFTNRRNFHLNGELFDALSVAEDDTWRRIR	120
CYP3A4	VWGFYDGGQPVLAITDPDMIKTVLVKECYSVFTNRRPFGPVGFMKSAISIAEDEEWKRLR	120
CYP3A5	MWGTYEGQLPVLAITDPDVIRTVLVKECYSVFTNRRSLGPVGFMSAISLAEDEEWKRLR	120
CYP3A37	IWGIYDGRQPAVVMDPQIIKTVLVKECYSTFTNRRRVLAGELRNAVSLAEDDQWKRLR	119
CYP3_N1	IWGFYDGRQPVLAIMDPAIIKTVLVKECYSTFTNRRNFGLNGPLESAISIAEDEQWKRLR	120
	:** *:*. *.:. * :*:.*:*****. ***** . * :*.*:.*:*.*:.*	
CYP3A27	SVLSPSFTSGRLKEMFGIMKQHSSTLLSGMKKQADKDQTEVKEFFGPYSMDVVTSTAFS	180
CYP3A4	SLLSPFTTSGKLEKMPVPIAQYGDVLRNLRREAEKGKPVTLKDVFGAYSMDVITSTSF	180
CYP3A5	SLLSPFTTSGKLEKMPVPIAQYGDVLRNLRREAEKGKPVTLKDVFGAYSMDVITSTSF	180
CYP3A37	TVLSPFTTSGKLEKMPFTMKHFGEMLVKNVQKRVKEDSSVPVKDFFGSYSMDVVTSTSF	179
CYP3_N1	TVLSPFTTSGKLEKMPFIMKHGQDILVKNIQKKAEDKDEPVMKDFGAYSMDVITSTSF	180
	:***:***:***. : :... *: :.:.:.:. : :*:.* *****:*.*:.*.	
CYP3A27	VDIDSLNNPSPDFVSNVKKMLKFDLFNPLFLLVALEPFTGPILEKMKFSFFPTAVTDFFY	240
CYP3A4	VNIDSLNNPQDPFVENTKKLLRDFDLDPFFLSITVFPFLIPILEVLNICVFPREVTFNLR	240
CYP3A5	VNIDSLNNPQDPFVESTKKFLKGFGLDPLFLSILFPFLTVPVEALNVSLFPKDTINFLS	240
CYP3A37	VNIDSMNPNKSPFVREMQLTKDFDPLFILAFVCPFLTPLMAKMNVSPFSDAVDFFL	239
CYP3_N1	VDIDSLNNPDPFVTNKKLLKFNFFNPLFILIVFPFLTVPVLEKMNISLFPKDVMDFFK	240
	*:***:***:***. :*: :*:.*:***: : ** *:*. :.:...* .:*	
CYP3A27	ASLAKIKSGRDTGNSTNRVDFLQLMIDSKGSDTKTGEEQTKGLTDHEILSQAMIFIFAG	300
CYP3A4	KSVKRMKESRLDQKHRVDFLQLMIDSQNSK---ETESHKALSDLELVAQSIIFIFAG	296
CYP3A5	KSVNRMKKSRLNDKQKRLDFLQLMIDSQNSK---ETESHKALSDLELVAQSIIFIFAG	296
CYP3A37	RSIDIKKDRKETHKGRVDFLQMMIESQKSDS--DGKNSHKALSDIEVLSQLAFIFIFAG	297
CYP3_N1	NAVKKIKEDRDKDNHGRVDFLQLMIDSQTSE---ENHSHKALTDVEILAQAIIIFIFAG	296
	: :*:.* . . *:***:***:***. . . *:.* *: :*:.*:***	
CYP3A27	YETSSSTMSFLAYNLATNHHVMTKLQEEIDTVFPNKAPIQYQALMQMDYLDQVLNESLRL	360
CYP3A4	YETSSSVLSFIMYELATHPDVQKQLQEEIDAVLPNKAPPTYDVLQMEYLDMMVNETLRL	356
CYP3A5	YETSSSVLSFTLYELATHPDVQKQLQEEIDAVLPNKAPPTYDAVQMEYLDMMVNETLRL	356
CYP3A37	YEPTSENTLCYLAYLLALHDPVQKQVNEIDTILPNKAPLTYEALMQLDYLDMAVNETLRL	357
CYP3_N1	YETSSSTLSFLAYNLATHPDVQKQLQEEIDATLPNKAPPTYDALMQMEYLDMMVNETLRL	356
	** :*.:.*: ** *:*. :* *:***: ***** *:***:***. :*:.*	
CYP3A27	YPIAPRLERVAKKTVEINGIVIPKDCIVLVPTWTLHRDPEIWSNPPEFKPERFSKENKES	420
CYP3A4	FPIAMRLERVCKKDVEINGMFIPKGWVVMIPSYALHRDPKYWTEPEKFLPERFSKKNKDN	416
CYP3A5	FPVAIRLERCTCKDVEINGVFIPKGWSMVIPTYALHDPKYWTEPEEFRPERFSKK-KDS	415
CYP3A37	YPLGGRIRERTCKKDVEINGVTIPKETIVVIPPYTLHRNSEYWPNPPEFRPERFSKENKDN	417
CYP3_N1	YPIGGRLERVCKKTVEINGVTIPKGTVVMIPVYVLRDPEYWPPEEFRPERFSKENKES	416
	:*. *.*:.* ***** ***: :*:.*:***: :* *:*** *****: *.:	
CYP3A27	IDPYTYMPFGAGPRNCIGMRFALIMIKLAMVEILQSFTFSVCDETEIPLMDNQGLLMPK	480
CYP3A4	IDPYTYTFFGSGPRNCIGMRFALNMKMLALIRVLQNFSPKPKETQIPLKLSLGGLLQPE	476
CYP3A5	IDPYTYTFFGTGPRNCIGMRFALNMKMLALIRVLQNFSPKPKETQIPLKLDQGLLQPE	475
CYP3A37	IDPYTYLFFGAGPRNCIGMRFALLTLKVAITSILQHFTFQVCKETQIPLKLLSMGLMTPE	477
CYP3_N1	IDPYTLFFGAGPRNCIGMRFALLTMKVAIVRLLQNFSPKPKETQIPLLELNTQGFLTPK	476
	*****: ***:*****: :*:.*: ***:.* *:***:***: :*:.*	
CYP3A27	RPIKLRLEARNTPSNTTATTLKSPTTSTHHHHHH	515
CYP3A4	KPVVLKVESRDGTVSGASTHHHHHH-----	501
CYP3A5	KPIVLKVDSDGTLSGESTHHHHHH-----	500
CYP3A37	KPIILKLVPRTNNAKAH--HHHHH-----	499
CYP3_N1	KPIVLKLVPSSTHHHHHH-----	493
	:*:.*:	

Supplementary Figure 4: Alignment of the expressed sequence of CYP3_N1 with those of the extant CYP3 forms studied here.

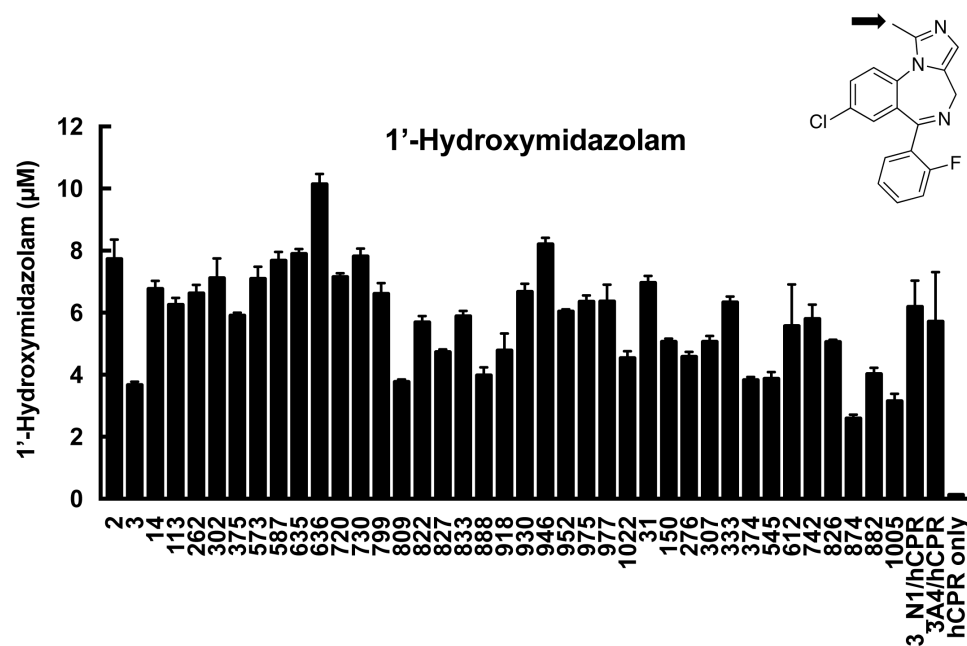


Supplementary Figure 5. Structural determinants of CYP3_N1 thermostability.

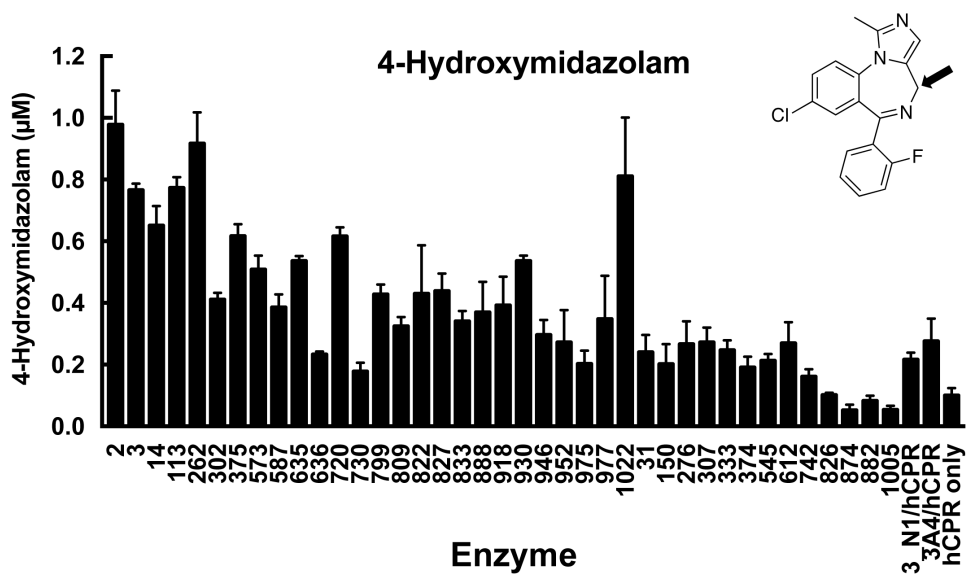
A. Mutations introduced into CYP3_N1 via site-directed mutagenesis (SDM) to decrease core hydrophobicity and thereby stability. Residues are numbered according to native extant CYP3A4 sequence rather than the expressed ancestral protein (to correspond with 3A4 X-ray structure, pdb:1TQN). Residue numbers are reduced by 10 in the expressed protein due to modification of the N-terminus for expression. B. Relative thermostability of SDM variants expressed in bicistronic format with hCPR. $10T_{50}$ is the temperature at which 50% of the protein remains intact after heating for 10 min. The data are means and range of two biological replicates of the $10T_{50}$ values interpolated from thermal denaturation curves fitted using a sigmoidal curve in GraphPad Prism. Asterisks indicate statistically significant differences to CYP3_N1 (*, $p < 0.01$; **, $p < 0.005$; ***, $p < 0.001$; two-tailed Students t -test). The $10T_{50}$ of 3_N1_SDM8 was not significantly different (N.S.) to that of CYP3A4 ($p=0.145$, two-tailed Students t -test). C, D. The location of the core-stabilising mutations

represented using homology model for CYP3_N1, constructed using SWISS-MODEL with the CYP3A4 X-ray structure (pdb:1TQN) as a template. C. Side view, with the I-helix shown running through the structure to the right of the haem. D. Top view, looking down on the active site from the distal side of the haem with the beta-sheet rich domain at the lower point of the lozenge-shaped fold. The haem is colored in gold and the mutated residues are shown in red (Ile93), orange (Ile160), yellow (Ile196), lime-green (Ile207), bright green (Ile225), cyan (Ala252), blue (Ile360), blue-purple (Ile 398) and magenta (Ile400).

A

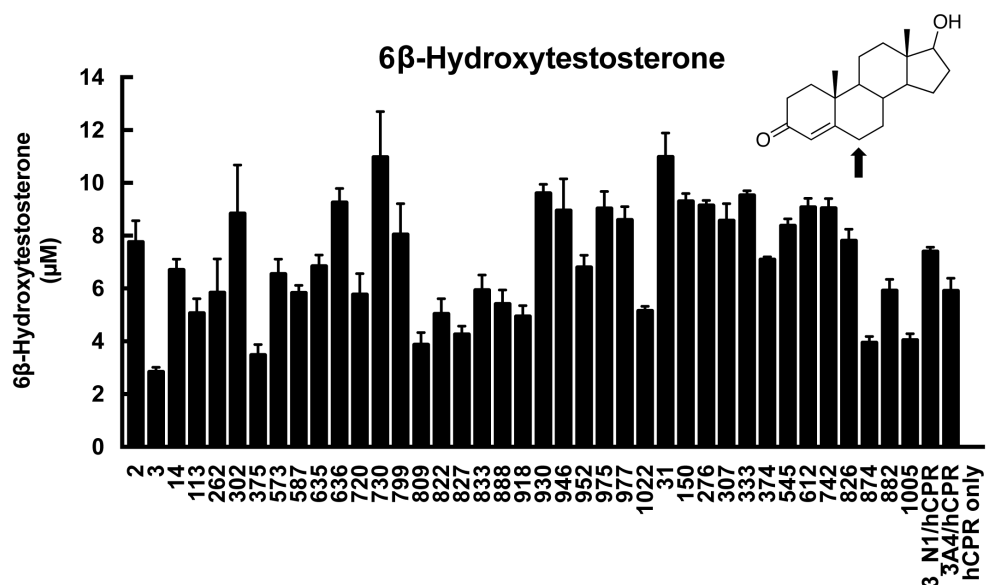


B

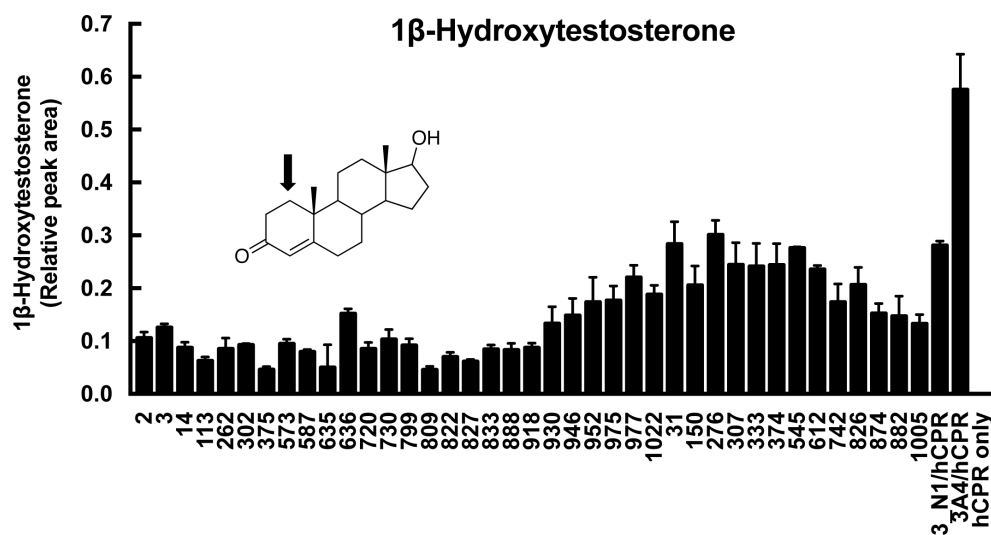


Supplementary Figure 6 (first part)

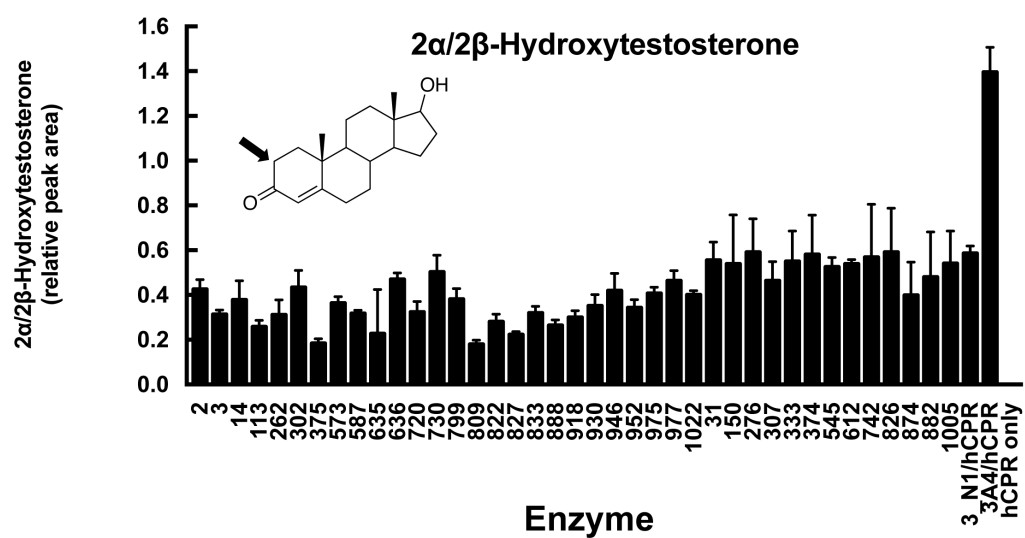
C



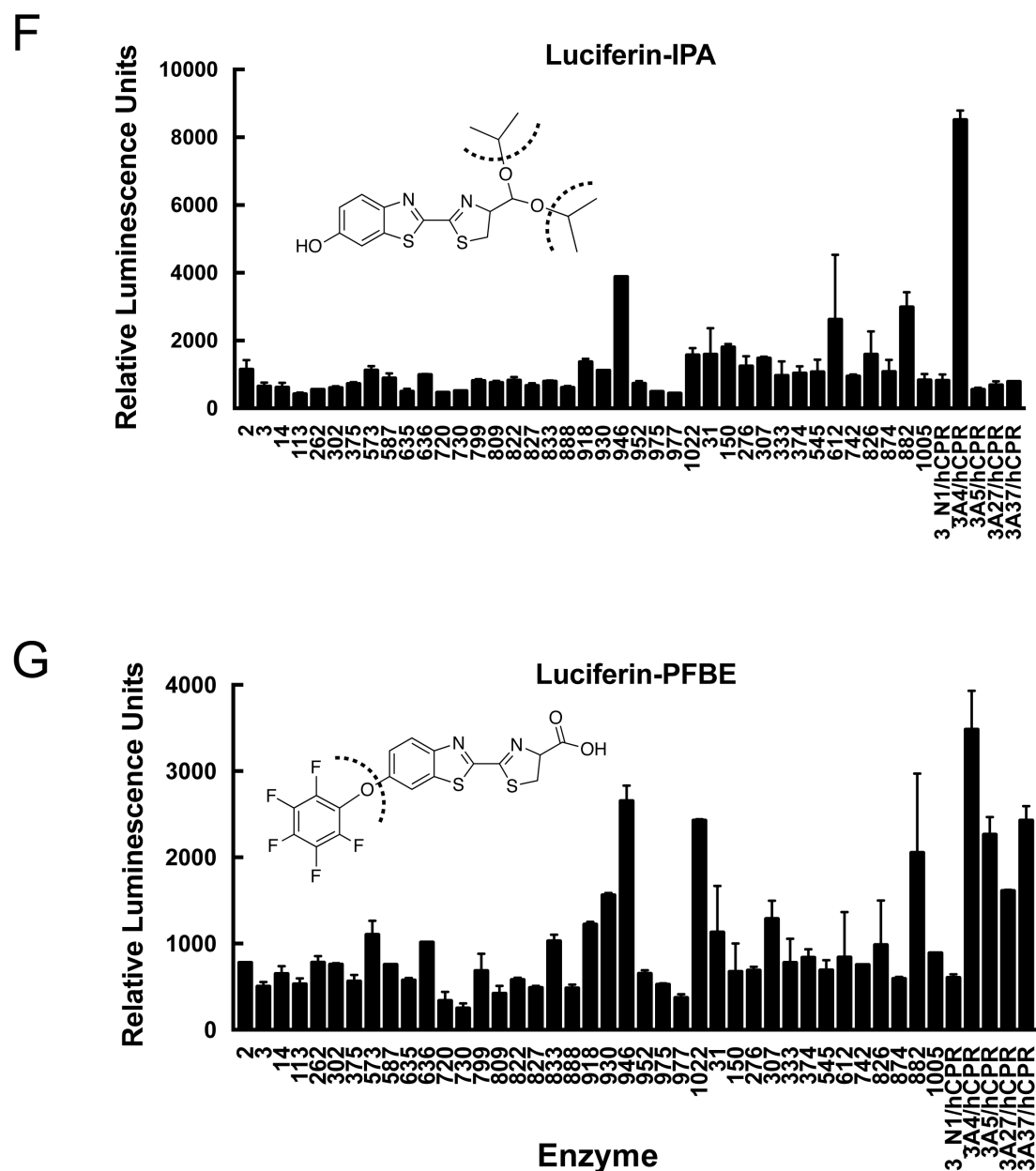
D



E



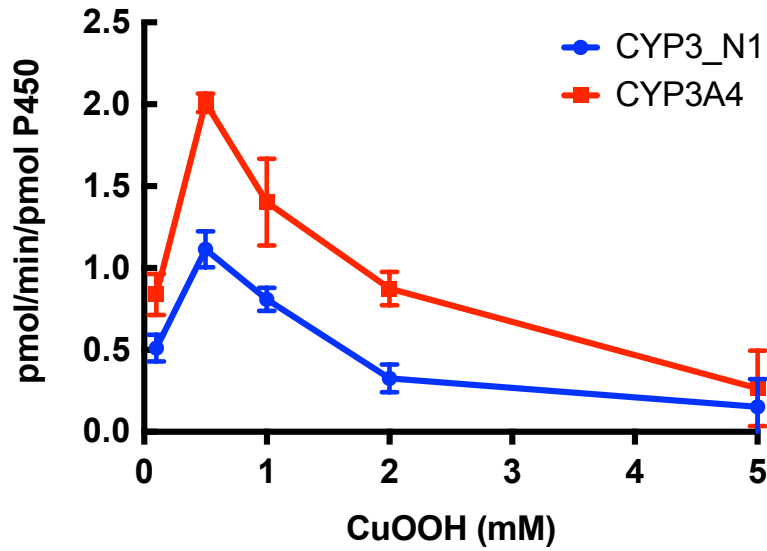
Supplementary Figure 6 cont.



Supplementary Figure 6: Comparison of activities of selected mutants from the ancestral cloud library towards CYP3 marker substrates

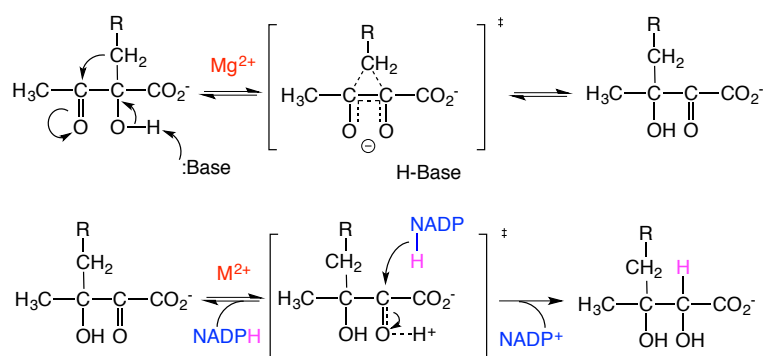
A,B. Midazolam 1' (A) and 4-hydroxylation (B). Bacterial membranes containing 0.2 μ M of the respective mutant or extant form co-expressed with hCPR were incubated for 120 min at 37 $^{\circ}$ C with 100 μ M midazolam and an NADPH-generating system. Incubations were

quenched by extraction with ethyl acetate and metabolites quantified by HPLC. Data are the mean $\pm$ standard deviation of $n = 3$ technical replicates. C, D, E. Testosterone 6β -(C), 1β -(D) and $2\alpha/2\beta$ -hydroxylation. Bacterial membranes containing $0.2\ \mu\text{M}$ of the respective mutant or extant form coexpressed with hCPR were incubated for 120 min at $37\ ^\circ\text{C}$ with $100\ \mu\text{M}$ testosterone and an NADPH-generating system. Incubations were quenched by extraction with ethyl acetate and testosterone metabolites were quantified by HPLC. For 1β - (D) and $2\alpha/2\beta$ -hydroxytestosterone (E) relative peak heights are shown, scaled to match those of the 6β -hydroxy metabolite assuming all metabolites have the same molar absorption coefficient as the authentic 6β -hydroxytestosterone. Data are the mean $\pm$ standard deviation of $n = 3$ technical replicates. F, G. Dealkylation of luciferin-IPA (luciferin isopropyl acetal; F) and PFBE (luciferin 6'-pentafluorobenzyl ether; G). Individual variants were expressed in duplicate. Resuspended cells containing $\sim 20\ \text{nM}$ of the respective P450 co-expressed with hCPR were incubated with $25\ \mu\text{M}$ luciferin PFBE or $1.5\ \mu\text{M}$ luciferin IPA in buffer for 1 h at $25\ ^\circ\text{C}$ with shaking at 400 rpm. Luciferin was detected in cell-free supernatants after addition of luciferin detection reagent containing luciferase. Data are the mean and range of $n = 2$ biological replicates. No activities were detected with other luciferin substrates tested (luciferins BE, ME, H, or CEE). Arrows represent positions of hydroxylations on testosterone and midazolam. Dotted lines indicate *O*-dealkylations of luciferin derivatives.



Supplementary Figure 7: Testosterone 6 β -hydroxylase activity of CYP3_N1 and CYP3A4 supported by cumene hydroperoxide

Testosterone (50 μ M) was incubated with 1 μ M CYP3_N1 or CYP3A4 in triplicate for 10 min. Reactions were initiated by addition of cumene hydroperoxide (CuOOH) to the concentrations indicated and 6 β -hydroxytestosterone was quantified by HPLC. The data are means $\pm$ standard deviations from three technical replicates.

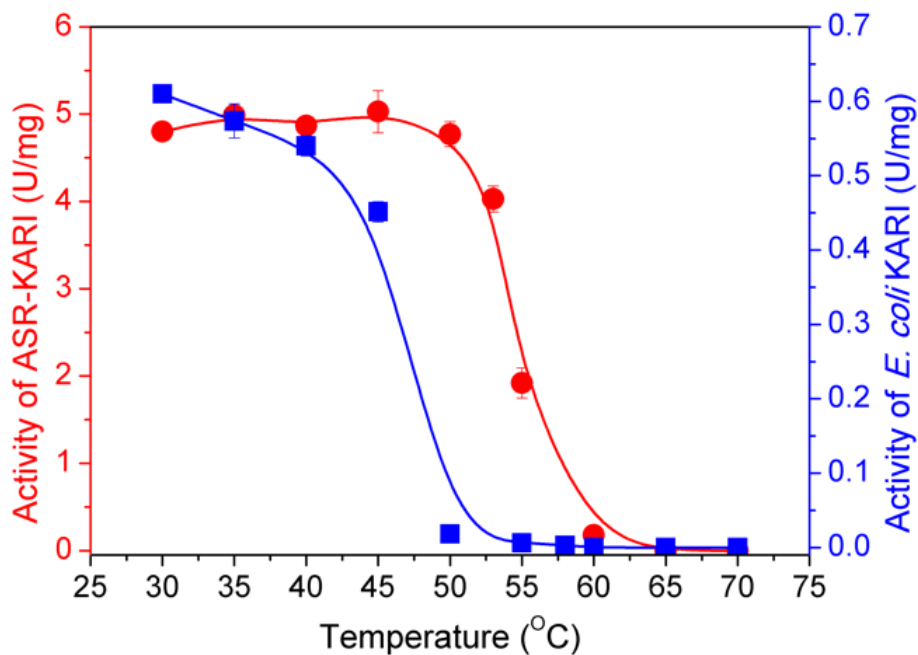


Supplementary Figure 8. The reactions catalysed by KARI. A. In the first step, the alkyl migration has a strict requirement for Mg²⁺, whereas the second step requires NADPH and a metal ion that can be Mn²⁺, Co²⁺, Ni²⁺ or Mg²⁺. The naturally occurring substrates for the reaction are 2-acetolactate (R=H) and 2-aceto-2-hydroxybutyrate (R=CH₃).

E_coli_KARI	MANYFNTLNLRQQLAQLGKCRFMGRDEFADGASYLQGKKVVIVGCCAQGLNQGLNMRDSG
Ancestral_KARI	MANYFNSIPWREALLQLGHCFMDSSEFADGVNALGKKIVIVGCCAQGLNQGLNLRDSG
	***** : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *
E_coli_KARI	LDISYALRKEAIAEKRASWRKATENGFKVGTYEELIPQADLVINLTPDKQHSDDVVRTVQP
Ancestral_KARI	LDVSYALRKEAIAEKQSWKNATENGFTVGTYDEMIPADLVGNLTPDKQHTSVITAVMP
	** : ***** * : * : * : * : * : * : * : * : * : * : * : * : * : *
E_coli_KARI	LMKDGAALGYSHGFNIVEVGEQIRKIDITVVMVAPKCPGTEVREEYKRGFGVPTLIAVHPE
Ancestral_KARI	LMKKGSAWYSHGFNIVEEGMQIRKIDITVMMAPKGPSEVRTEYVRGFGVPTLIAVHPE
	** : * : * : * : * : * : * : * : * : * : * : * : * : * : *
E_coli_KARI	NDPKGEGBMAIAKAWAAATGGHRAGVLESSFVAEVKSDLMGEQTILCGMLQAGSLLCFDKL
Ancestral_KARI	NDPEGKGAIAKALAVGTGGSRAVLESSFVAEVKSDLMGEQTILCGMLQAGSILCFDKM
	** : * : * : * : * : * : * : * : * : * : * : * : * : * : *
E_coli_KARI	VEEGTDPAYAEKLIQFGWETITEALKQGGITLMDRLSNPAKLRAVALSEQLKEIMAPLF
Ancestral_KARI	VEKIDPGYASKLIQYGWETITEALKWGGITNMDRLSNPAKIKAHLSKELKTIMAPLY
	** : * : * : * : * : * : * : * : * : * : * : * : * : * : *
E_coli_KARI	QKHMDDIISGEFSSGMMADWANDDKLLTWREETGKTAFETAPQY-EGKIGEQEYFDKGV
Ancestral_KARI	QKHMDDIISGEFSSTMMKDWANDDKNLLTWREATGKTAFEQTPAGAVIEISEQEYFDKGI
	***** * : * : * : * : * : * : * : * : * : * : * : * : * : *
E_coli_KARI	LMIAMVKAGVELAFETMVDSGIIIESAYYESLHELPLIANTIAKRLEYMNVVISDTAEY
Ancestral_KARI	LMVAMVKAGVELAFETMVDAGIKPESAYYESLHETPLIANLISRKLEYMNRVISDTAEY
	** : ***** : * : * : * : * : * : * : * : * : * : * : *
E_coli_KARI	GNYLFSYACVPLLK-PFMAELQPGDLGKAIP---EGAVDNGQLRDVNEAIRSHAIEQVGK
Ancestral_KARI	GCYLFDNACVPLLQKDFMAKVDTDVIGKGLNVSKDNSVSNRELVAVNDEIRNHPVEIVGK
	* : * : * : * : * : * : * : * : * : * : * : * : * : * : *
E_coli_KARI	KLRYMTDMKRIAVAG
Ancestral_KARI	KLRYMTAMKAVVKGGA
	** : * : * : * : * : * : * : * : * : * : * : *

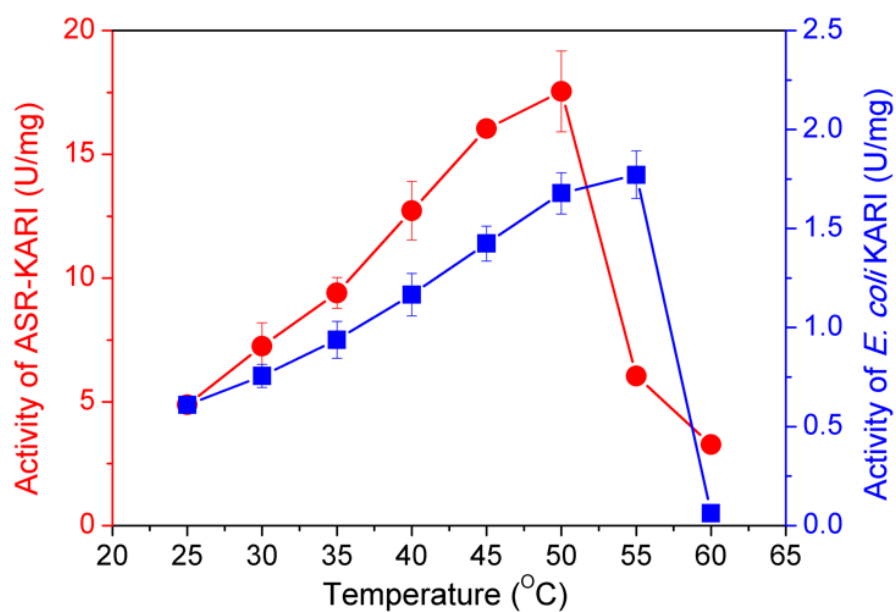
Supplementary Figure 9. Alignment of the ancestral KARI with *E. coli* KARI.

Sequences of the inferred KARI ancestor and extant *E. coli* form were aligned with Clustal Omega.



Supplementary Figure 10. Comparison of the thermal stability of wild type *E. coli* and ASR KARI.

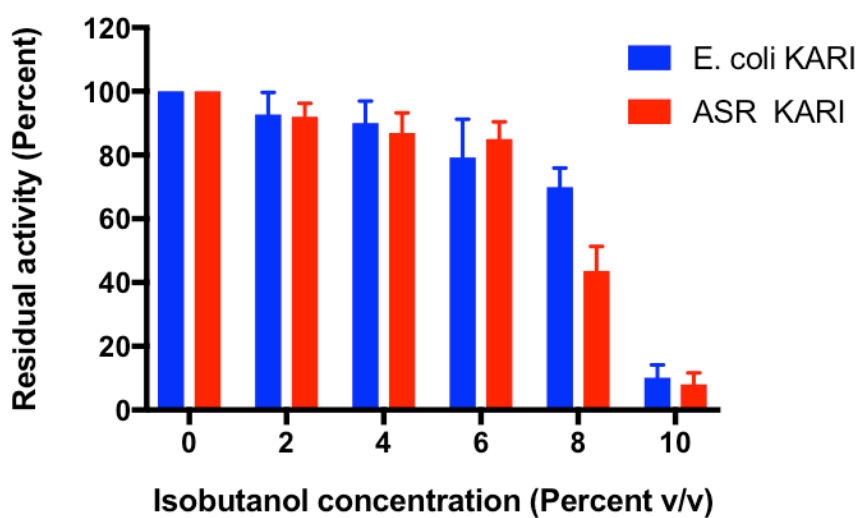
The thermal stability of wild type *E. coli* (blue squares) and ASR KARI (red circles) was determined by incubating the enzymes at various temperatures (30–70 °C) at pH 8.0 for 10 min, before recording the activity at 25 °C. The activity of the sample incubated at 30 °C was arbitrarily set as 100%. Data are the mean and range of duplicate measurements and the trend lines were drawn in Origin by interpolation between the means.



Supplementary Figure 11. Temperature dependence of the KARI catalytic activity.

The activities of wild type *E. coli* (blue squares, y axis at right) and ASR KARI (red circles, y axis at left) were measured in triplicate at various temperatures from 25 to 60 °C at pH 8.0.

Data are the mean and range of duplicate measurements and trend lines are interpolations between points.



Supplementary Figure 12. Effect of isobutanol concentration on enzyme activity.

The activity of wild type *E. coli* (blue columns) and ASR KARI (red columns) were measured in triplicate at 25 °C and pH 8.0 at isobutanol concentrations ranging from 0 to 10 % (v/v).

Data are the mean and range of duplicate measurements.

Supplementary Tables

Supplementary Table 1. Thermostability of extant vs ancestral CYP3 forms.

Variant	$^{60}T_{50}$ (°C) ^a	E_d (kJ/mol) ^b
CYP3A4	35 ± 3	91
CYP3A5	38 ± 2	116
CYP3A27	37 ± 3	151
CYP3A37	36 ± 4	116
CYP3_N1	66 ± 2	251

^a $^{60}T_{50}$ is the temperature at which 50% of the protein remains intact after heating for 60 min.

The thermostability data was plotted using Excel and $^{60}T_{50}$ values were extrapolated from the interpolated curve. The data shown are means and ranges derived from two biological replicates of the denaturation curve for each enzyme.

^b E_d is the energy of thermal deactivation, derived from Arrhenius plot shown in Supplementary Figure 2 fitted using linear regression with GraphPad Software. Three replicate cultures of bacterial cells expressing each enzyme were heated for 60 min at temperatures from 25-75 °C, then cooled for 5 min at 4 °C and equilibrated to room temperature for 5 min before P450 content was quantified by Fe(II)-CO versus Fe(II) difference spectroscopy.

Supplementary Table 2: Estimated half-lives of CYP3_N1 and selected extant forms.

T (°C)	Variants ^a	t _{1/2} (min) ^b
50	CYP3A4	2.1
	CYP3A5	1.1
	CYP3A27	4.5
	CYP3A37	2.4
	CYP3_N1	595.4
55	CYP3A4	0.5
	CYP3A5	0.9
	CYP3A27	1.6
	CYP3A37	0.8
	CYP3_N1	566.5
60	CYP3A4	0.4
	CYP3A5	0.5
	CYP3A27	0.7
	CYP3A37	0.6
	CYP3_N1	54.7
65	CYP3A4	0.2
	CYP3A5	0.3
	CYP3A27	0.8
	CYP3A37	0.3
	CYP3_N1	26.6
70	CYP3A4	n.a
	CYP3A5	0.2
	CYP3A27	0.1
	CYP3A37	0.2
	CYP3_N1	11.7
75	CYP3A4	0.1
	CYP3A5	0.1
	CYP3A27	n.a
	CYP3A37	0.1
	CYP3_N1	1.3

^a The ancestral (CYP3_N1) and extant (CYP3A4, CYP3A5, CYP3A27, CYP3A37) enzymes were expressed in *E. coli* as three biological replicate cultures.

^b Bacterial cells expressing each enzyme were heated to the temperatures indicated for 60 min, then cooled for 5 min at 4 °C and equilibrated to room temperature for 5 min prior to quantification of residual P450 content by Fe(II).CO vs Fe(II) difference spectroscopy. The P450 remaining intact in heated samples was calculated as a proportion of total P450 in untreated samples for each temperature. Estimated half-lives of extant and ancestor variants at various temperatures were derived from Arrhenius plot shown in Supplementary Figure 2.

Supplementary Table 3: Comparison of catalytic activity of CYP3_N1 and CYP3A4

Substrate	Footnote to Conditions	Activity	
		CYP3_N1	CYP3A4
Aminopyrine	a	83.5	58.1
Amlodipine	a	84.2	81.1
7-Benzyloxyresorufin	b	2.07	3.4
Clopidogrel	a	60.4	78.9
Clozapine	a	29.8	38.4
Coumarin	c	no activity	no activity
7-Ethoxycoumarin	c	no activity	no activity
Dextromethorphan	a	54.4	48.6
Diltiazem	a	17.7	37.4
Donepezil	a	8.7	18.6
7-Ethoxyresorufin	b	no activity	no activity

Erythromycin	a	22.3	45.3
Imipramine	a	55.9	46.8
Luc-CEE (luciferin 6' - chloroethyl ether)	d	no activity	no activity
Luc-H (6'-deoxyluciferin)	d	no activity	no activity
Luc-H-EGE (ethylene glycol ester of luciferin 6'-methyl ether)	d	no activity	no activity
Luc-IPA (luciferin isopropyl acetal)	e	0.16	0.6
Luc-ME (luciferin 6'-methyl ether)	d	no activity	no activity
Luc-ME-EGE (methyl ester of luciferin 6'-methyl ether)	d	no activity	no activity
Luc-PFBE (luciferin 6' - pentafluorobenzyl ether)	e	0.23	0.12

Midazolam	f	33	29
Methoxyresorufin	b	no activity	no activity
Nifedipine	g	1.79	3.74
Olanzapine	a	4.0	9.4
Omeprazole	a	9.5	42.4
7-Pentoxoresorufin	b	3.18	4.64
Rosiglitazone	a	3.3	4.1
Tamoxifen	a	57.0	82.5
Testosterone	h	16.3	12.3
Ticlodipine	a	95.0	97.8
Verapamil	a	25.6	60.7

^a Incubations were carried out using 10 μ M substrate for 20 min. Metabolite formation was quantified based on MS response using LC-MS^E and the activity is shown as the production of the major metabolite expressed as the percentage of total MS area of compound related material detected. Data are the means of duplicate measurements.

^b Incubations were carried out using 5 μ M substrate for 30 min. Metabolite formation was quantified by fluorescence with reference to a standard curve prepared with resorufin.

Activity is expressed as pmol resorufin produced per 30 min.

^c Incubations were carried out using 100 μ M substrate for 120 min.

^d Incubations with luciferins CEE (30 μ M), H (100 μ M), H-EGE (10 μ M), ME (100 μ M) and ME-EGE (30 μ M) were carried out at the substrate concentrations indicated and for 120 min.

^e Incubations with luciferins IPA (1.5 μ M) and PFBE (25 μ M) were carried out at the substrate concentrations indicated and for 60 min. Activity is expressed as pmol luciferin produced per 60 min.

^f Incubations were carried out using 1 mM midazolam for 120 min. 1'-Hydroxymidazolam formation was quantified by LC-MS and is expressed as μM 1'-hydroxymidazolam produced per 120 min.

^g Incubations were carried out using 100 μM nifedipine for 5 min. Formation of the pyridine metabolite was quantified by LC-MS and is expressed as metabolite produced per min per pmol P450.

^h Incubations were carried out using 100 μM testosterone for 10 min. Metabolite formation was quantified by LC-MS and is expressed as pmol 6 β -hydroxytestosterone produced per min per pmol P450.

Supplementary Table 4: K_d values for ligand binding to CYP3_N1 and CYP3A4

Ligand ^a	CYP3_N1		CYP3A4	
	$\Delta A_{\max} \times 10^3$	K _d (μM)	$\Delta A_{\max} \times 10^3$	K _d (μM)
Cholesterol	1.8 ± 0.2	2.3 ± 0.6	2.9 ± 0.2	4.7 ± 0.7
Progesterone	1.4 ± 0.1	13 ± 3	19 ± 6	126 ± 50
Androstenedione	3.1 ± 0.2	31 ± 7	12.2 ± 0.4	19 ± 3
Dehydroepiandrosterone	6.6 ± 0.4	30 ± 5	14.0 ± 0.5	6.7 ± 1.4
Testosterone	3.8 ± 0.8	24 ± 10	6.1 ± 0.4	6 ± 2
Dihydrotestosterone	1.3 ± 0.1	4.8 ± 1.5	10.2 ± 0.7	18 ± 3
Estradiol	3.3 ± 0.3	4.5 ± 1.0	0.3 ± 0.1	10 ± 7
Cortisol	6.5 ± 3.8	197 ± 150	14 ± 4	160 ± 60
Erythromycin	5.4 ± 0.1	8.2 ± 0.9	1.7 ± 0.1	4 ± 1
Cyclosporin	4.3 ± 0.2	3.9 ± 0.8	13 ± 1	5 ± 1
Troleandomycin	5.7 ± 0.1	9 ± 1	0.24 ± 0.01	1.7 ± 0.5
Quinidine	7.2 ± 0.4	120 ± 80	6 ± 1	48 ± 15
Ketoconazole	6.8 ± 0.4	0.04 ± 0.01	8.5 ± 0.3	0.002 ± 0.003

^a Recombinant bacterial membranes containing 0.1 μM of either CYP3_N1 or CYP3A4 were titrated with the ligands indicated. Ligands were added from concentrated stocks in methanol or water and an equivalent volume of the solvent was added to the reference cuvette during each titration. Spectra were acquired between 350 nm and 450 nm on an OLIS-modernised Aminco DW2a spectrophotometer. For each measurement, four spectra were averaged using the OLIS-DW2a instrument software.

^b Binding curves were prepared from measuring the difference between the absorbance at the peak and trough wavelengths for each set of spectra. K_d and $\Delta A_{\max}$ values and the respective

standard errors were determined using the quadratic form of the binding equation in Graphpad Prism.

Supplementary Table 5: Comparison of the catalytic efficiency of CYP3_N1 and CYP3A4 towards testosterone 6 β -hydroxylation and nifedipine oxidation

Substrate	Enzyme	$k_{\text{cat}}^{\text{a}}$ (min^{-1})	K_{m} (μM)	$k_{\text{cat}}/K_{\text{m}}$
Testosterone ^b	CYP3A4	23.9 ± 0.6	97 ± 6	0.25
	CYP3_N1	18.1 ± 0.7	30 ± 4	0.61
Nifedipine ^c	CYP3A4	5.4 ± 0.1	42 ± 3	0.13
	CYP3_N1	2.0 ± 0.1	9 ± 1	0.22

^a Kinetic parameters and the respective standard errors were estimated by nonlinear regression of substrate saturation curves in Graphpad Prism using the Michaelis-Menten Model and the data shown in Supplementary Figure 3.

^b Testosterone hydroxylation was carried out with 10 – 500 μM substrate and 50 nM P450 (coexpressed in bacterial membranes with hCPR) in 100 mM phosphate buffer (pH 7.4) at 37 °C for 10 min.

^c Nifedipine oxidation was carried out similarly but with 5 – 200 μM substrate and 100 nM P450 in bacterial membranes, in 100 mM Tris buffer (pH 7.4) for 5 min. Product formation had been previously established to be proportional to time and enzyme concentration under these conditions.

Supplementary Table 6: Potential hydrophobic interactions responsible for stabilization of CYP3_N1 relative to CYP3A4 and other extant forms analysed.

Site directed mutant	Mutation(s) ^a	Position in 3D structure	Potential interacting residues	Position in 3D structure	$\Delta^{10}T_{50}$ (°C) ^b
3_N1_SDM1	Ile93Ala	Helix B	Ile84	Sheet1	-11.9
3_N1_SDM2	Ile160Ala	Helix D	Ile175	Helix E	-16.1
3_N1_SDM3	Ile196Ala	Loop between helices E and F	Ala178	Helix E	-19.8
	Ile207Ala	Helix F	Ile303	Helix I	
3_N1_SDM4	Ile225Ala	Helix F'	Ile47	Loop before helix A	-13.1
3_N1_SDM5	Ala252Ser	Helix G	Phe203	Loop between helices F and F'	-13.3
3_N1_SDM6	Ile360Ala	Helix K	Leu449	Helix L	-14.7
3_N1_SDM7	Val398Ala	Helix K'	Leu433	Loop before helix L	-13.9
	Val400Ala		Phe65	Helix A	
3_N1_SDM8	All	See above	See above	See above	-19.3

^a Residues are numbered according to the CYP3A4 sequence rather than the expressed ancestor to correspond with the CYP3A4 structure, pdb:1TQN. Residue numbers are reduced by 10 in the expressed P450 due to N-terminal modifications.

^b Change in $^{10}T_{50}$ observed in the respective mutant relative to CYP3_N1.

Supplementary Table 7: Thermostability and sequence changes in selected mutants of the CYP3_N1 ancestral cloud library ranked by ¹⁰T₅₀.

Mutant number	P450 (nM) ^a	¹⁰ T ₅₀ ^b	Amino acids present at degenerate positions ^c									
			57	74	101	164	172	232	250	284	451	484
375	386	75.8	I	L	T	A	M	V	M	S	T	T
636	373	75.1	F	L	T	A	M	V	K	S	T	T
3	373	74.5	I	I	N	A	M	V	M	S	S	Q
952	704	74.5	I	L	N	A	M	I	M	S	T	T
587	431	74.4	I	L	T	V	M	I	M	S	S	T
635	672	74.4	I	L	T	A	V	I	M	S	T	T
975	585	74.0	I	I	T	A	M	V	K	N	T	T
888	661	73.7	F	L	T	A	V	I	M	S	S	Q
799	543	73.6	I	L	N	A	V	I	M	S	S	Q
946	402	73.6	I	L	T	A	M	V	M	N	T	Q
742	769	73.6	I	L	T	V	V	I	M	S	S	T
730	474	73.6	I	L	T	A	V	I	M	N	S	T
822	628	73.6	I	L	T	V	M	I	M	S	T	T
262	480	73.4	I	I	T	V	V	V	M	N	T	Q
918	474	73.4	I	L	T	A	M	I	M	S	S	Q
1005	632	73.3	I	L	T	A	M	I	M	S	T	T
809	771	73.1	I	L	T	A	V	V	M	S	T	Q
833	465	73.1	I	L	T	A	M	I	M	S	S	T
827	608	73.1	I	L	N	A	M	V	M	S	S	Q
573	420	73.0	I	L	T	A	M	I	M	N	S	T
720	547	73.0	I	L	T	A	M	V	M	S	S	T
302	574	72.9	I	I	N	A	V	V	M	N	S	Q
826	819	72.9	F	L	T	A	M	V	M	N	T	T
2	545	72.9	I	L	T	A	V	I	M	S	S	Q
977	661	72.8	I	L	T	A	M	V	K	S	S	Q
612	668	72.8	F	L	T	A	M	V	M	S	S	T
150	660	72.7	F	L	T	A	V	V	M	S	T	Q

14	605	72.6	I	L	T	A	V	I	M	S	S	T
113	605	72.6	I	L	T	A	V	V	M	S	T	T
874	729	72.4	F	L	N	A	M	V	M	S	T	Q
882	728	72.3	F	L	T	A	V	V	M	N	S	T
31	762	72.3	F	L	T	A	M	I	K	N	S	T
930	728	72.2	I	L	T	A	V	V	M	S	S	T
1022	237	72.2	I	L	T	A	M	V	M	N	S	T
374	643	71.8	F	I	T	V	M	V	M	S	S	Q
307	566	71.7	F	L	T	A	M	V	M	S	S	Q
3_N1	587	71.4	F	F	T	A	M	V	K	N	T	T
333	831	71.1	F	L	T	A	M	V	M	N	S	T
545	828	71.1	F	I	N	A	M	I	M	S	T	Q
276	838	69.9	F	I	T	A	M	V	M	N	S	Q

^a P450 expression was quantified by whole cell Fe(II).CO vs Fe(II) difference spectroscopy

^b $^{10}T_{50}$ is the temperature at which 50% of the protein remains intact after heating for 10 min.

Residual P450 content was quantified after heating to temperatures between 25 °C and 75 °C for 10 min, then cooling for 5 min at 4 °C and equilibrating to room temperature for 5 min.

^c Mutants were sequenced throughout the open reading frame. Shading indicates where residues in mutants were identical (dark grey) or different (unshaded or light grey) to the CYP3_N1 template (bolded). Residues are numbered according to native extant CYP3A4 sequence rather than the expressed ancestral protein (to correspond with 3A4 x-ray structure, pdb:1TQN). Residue numbers are reduced by 10 in the expressed protein due to modification of the N-terminus for expression.

Supplementary Table 8: Kinetic parameters for the ancestral and *E. coli* KARI enzymes

Kinetic parameter ^a	Ancestral KARI	<i>E. coli</i> KARI
K_m	$1240 \pm 20 \mu\text{M}$	$240 \pm 30 \mu\text{M}$
k_{cat}	$4.52 \pm 0.05 \text{ s}^{-1}$	$0.55 \pm 0.02 \text{ s}^{-1}$
Specific activity	4.88 U mg^{-1}	0.61 U mg^{-1}
k_{cat}/K_m	$3.65 \times 10^{-3} \mu\text{M}^{-1} \text{ s}^{-1}$	$2.29 \times 10^{-3} \mu\text{M}^{-1} \text{ s}^{-1}$

^a K_m and k_{cat} values and the standard errors in these parameters were determined by non-linear regression using the Michelis-Menten equation in Graphpad Prism and correspond to the lines of best fit drawn in Figure 5c.

Supplementary Table 9. Mutagenic oligonucleotides used in creating site directed mutagenesis variants of CYP3_N1

Variants	Mutation(s)	Primer	Sequence ^a
3_N1_SDM1	Ile93Ala	3_N1_93Ala_for	5'GGATCCGGCAATTATCAAAACCgcgCTGGTGAAAGAATGCTACAGCACCC3'
		3_N1_93Ala_rev	5'GGTGCTGTAGCATTCTTTCACCAGcgcGGTTTTGATAATTGCCGGATCC3'
3_N1_SDM2	Ile160Ala	3_N1_160Ala_for	5'CACTATGGCGACATCCTCGTGAAAAACgcgCAGAAAAAAGCCGAAAAAG ATGAGC3'
		3_N1_160Ala_rev	5'GCTCATCTTTTTTCGGCTTTTTTCTGcgcGTTTTTCACGAGGATGTCGCCATA GTG3'
3_N1_SDM3	Ile196Ala	3_N1_196Ala_207	5'GATATTGATAGCgcgAACAATCCGAACGATCCGTTTCGTGACCAATgcgAAA
	Ile207Ala	Ala_for	AAGCTTCTG3'
		3_N1_196Ala_207 Ala_rev	5'CAGAAGCTTTTTcgcATTGGTCACGAACGGATCGTTTCGGATTGTTcgcGCTA TCAATATC3'
3_N1_SDM4	Ile225Ala	3_N1_225Ala_for	5'GCTGTTTCATCCTGATTGTGgcgTTTCCGTTTCTGACACCGGTGCTG3'
		3_N1_225Ala_rev	5'CAGCACCGGTGTCAGAAACGGAAAcgcCACAATCAGGATGAACAGC3'
3_N1_SDM5	Ala252Ser	3_N1_252Ser_for	5'GGATGTGATGGACTTTTTCAAAACAgcGTGAAAAAATCAAAGAAGATC GC3'

		3_N1_252Ser_rev	5'GCGATCTTCTTTGATTTTTTTTCACgctGTTTTTGAAAAAGTCCATCACATCC 3'
3_N1_SDM6	Ile360Ala	3_N1_360Ala_for	5'GCAGATGGAGTACTTGGATATGGTTgcgAATGAAACCCTGCGCCTGTATCC 3'
		3_N1_360Ala_rev	5'GGATACAGGCGCAGGGTTTCATTcgcAACCATATCCAAGTACTCCATCTGC 3'
3_N1_SDM7	Val398Ala	3_N1_398Ala_400	5'GCACCGTTGTTATGATTCCGgcgTATgcgCTCCATCGTGATCCGGAATATTG Val400Ala Ala_for G3'
		3_N1_398Ala_400	5'CCAATATTCCGGATCACGATGGAGcgcATAcgcCGGAATCATAACAACGGT Ala_rev GC3'
1F1	-	-	5'GCTTAGGAGGTCATATGGCTCTGTTATTAGCAGTTTTTCTGACCCTGCT3'
1R10	-	-	5'TTGCTCACCATGCTCGAGGCGGCCGCGGTCTGACCGAGGAACCA3'

^a Altered codons are shown in lower case.

Supplementary Table 10. Mutagenic and silent oligonucleotides used in generating the CYP3 ancestor cloud library

Fragment	Primer	Position	Degeneracy	Sequence
A1	1F1	-	-	5'GCTTAGGAGGTCATATGGCTCTGTTATTAGCAGTTTTTCTGACCCTGCT3'
	1R1	57	IF	5'GCATTCGGTATCAAAATCCAGGAWGCCTTTACGATATTCCAGAAAGGTAC3'
A2	1F2	57	IF	5'GTACCTTTCTGGAATATCGTAAAGGCWTCCTGGATTTTGATACCGAATGC3'
	1R2	74	IL(F ^a)	5'CCATAATTGCCAGAACCGGCTGACGACCATCATAGAKGCCCCAGATTTTGCC G3'
B1	1F3	-	-	5'CGTCAGCCGGTTCTGGCAATTATGGATCCGGCAATTATC3'
	1R3	101	NT	5'CCAAAATTGCGACGATTGGTAAAGKTGCTGTAGCATTCTTTC3'
B2	1F4	101	NT	5'GAAAGAATGCTACAGCAMCTTTACCAATCGTCGCAATTTTGG3'
	1R4	164, 172	VA, VM	5'ATAGGCGCCAAAAATATCTTTTCAYATCCACCGGCTCATCTTTTTCCRCTTTTTT CTGAAT3'

C1	1F5	-	-	5'TGAAAGATATTTTTGGCGCCTATAGCATGGATGTTATTACCAGCACCAGCTTT AGCG3'
	1R5	232	IV	5'ACAGTGAAATATTCATTTTTTCCAGGAYCGGTGTCAGAAACGGAAAGATCAC 3'
C2	1F6	232	IV	5'GTGATCTTTCCGTTTCTGACACCGRTCCTGGAAAAAATGAATATTTCACTGT3'
	1R6	250	MK	5'GCGATCTTCTTTGATTTTTTTCACGGCGTTCWTGAAAAAGTCCATCAC3'
D1	1F7	-	-	5'GTGAAAAAAATCAAAGAAGATCGCAAAAAAGACAACCACAAAGATCGTGTT GATTTCT3'
	1R7	284	SN	5'CATCGGTCAGTGCTTTATGGCTATGGYTTTCTTCGCTGGTCTGGC3'
D2	1F8	284	SN	5'GCCAGACCAGCGAAGAAARCCATAGCCATAAAGCACTGACCGATG3'
	1R8	451	ST	5'CTGCAGCAGACGCACAATTGCAACTTTCATCGWCAGCAGTGCAAAGCGC

E1	1F9	-	-	5'ATGAAAGTTGCAATTGTGCGTCTGCTGCAGAATTTTAGCTTTCGTACCTGT3'
	1R9-a	484	QT	5'CAATCGGTTTTTTTCGGCGTCAGAAAACCCTGGGTATTCAGTTCAGCGGGATC
	1R9-b			TG3'
				5'CAATCGGTTTTTTTCGGCTGCAGAAAACCCTGGGTATTCAGTTCAGCGGGATC TG3'
E2	1F10-a	484	QT	5'CAGATCCCGCTGGAACTGAATACCCAGGGTTTTCTGACGCCGAAAAAACCGA
	1F10-b			TTG3'
				5'CAGATCCCGCTGGAACTGAATACCCAGGGTTTTCTGCAGCCGAAAAAACCGA
				TTG3'
	1R10	-	-	5'TTGCTCACCATGCTCGAGGCGGCCGCGGTCGACCGAGGAACCA3'
F	1F11	-	-	5'GCCTCGAGCATGGTGAGCAAGGGCGAGGAGCTGTTCAC3'
	1R12	-	-	5'CTTAAAGTTAAACAAAATTATTTCTAGAGCTCAGCTGCATCAATGATGATGAT GATGATG3'

^a Ile and Leu were introduced using the oligo noted whereas Phe was encoded by CYP3_N1 template at this position.