Supplemental Information for

Controlled oxidation of remote sp^3 C—H bonds in artemisinin via P450 catalysts with fine-tuned regio- and stereoselectivity

Kaidong Zhang, Brian M. Shafer, Matthew D. Demars II, Harry A. Stern, and Rudi Fasan*

Department of Chemistry, University of Rochester, Rochester, New York 14627

*fasan@chem.rochester.edu

Table of contents:

Figure S1. Chemical structures of the five probes used for high-throughput fingerprinting of the P450 libraries. The synthesis of these compounds was described earlier (Zhang et al., *J Am Chem Soc* **2011**, 133, 3242-45).

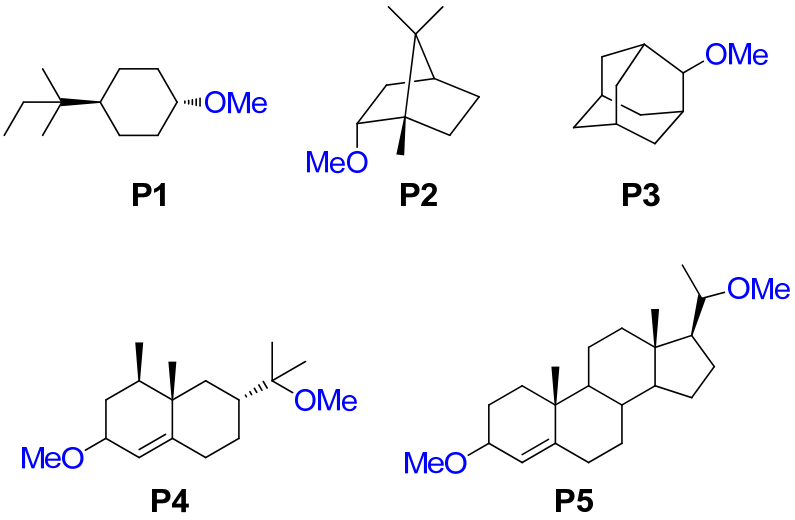


Figure S2. Plot of experimental *versus* calculated ART activity from multiple linear regression analysis of ART reactivity/fingerprint correlation across the P450 training set B (**Table S2**). Root mean standard deviation (RMSD): 0.179.

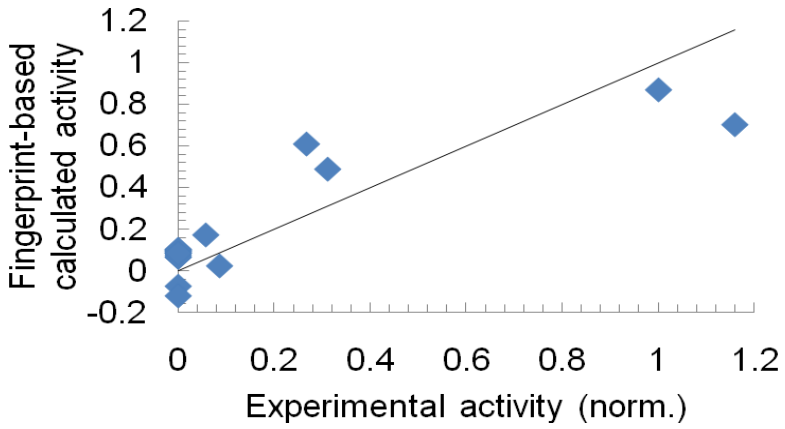


Figure S3. Ranking of the 50 II-E2-derived P450 variants according to their predicted artemisinin activity calculated based on the fingerprint-based model obtained using training set B (**Figure S2**).

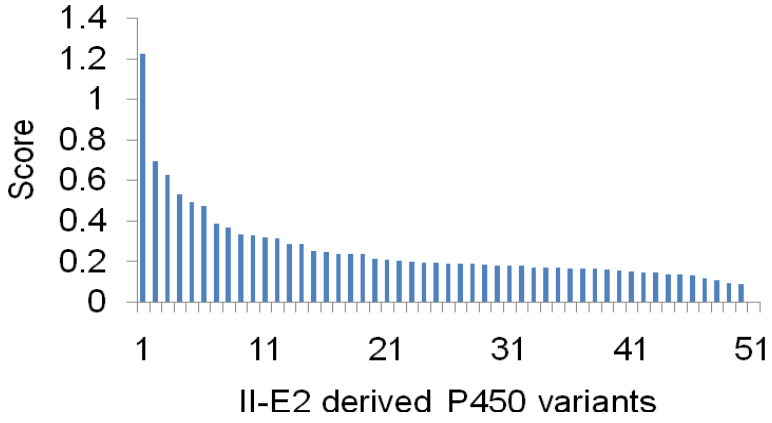


Figure S4. Visible spectra of the remaining P450 variants of Table 1 (in addition to those provided in **Figure 6**) recorded before (gray line) and after (red line) addition of artemisinin (1 mM), illustrating the artemisinin-induced shift of the heme spin state.

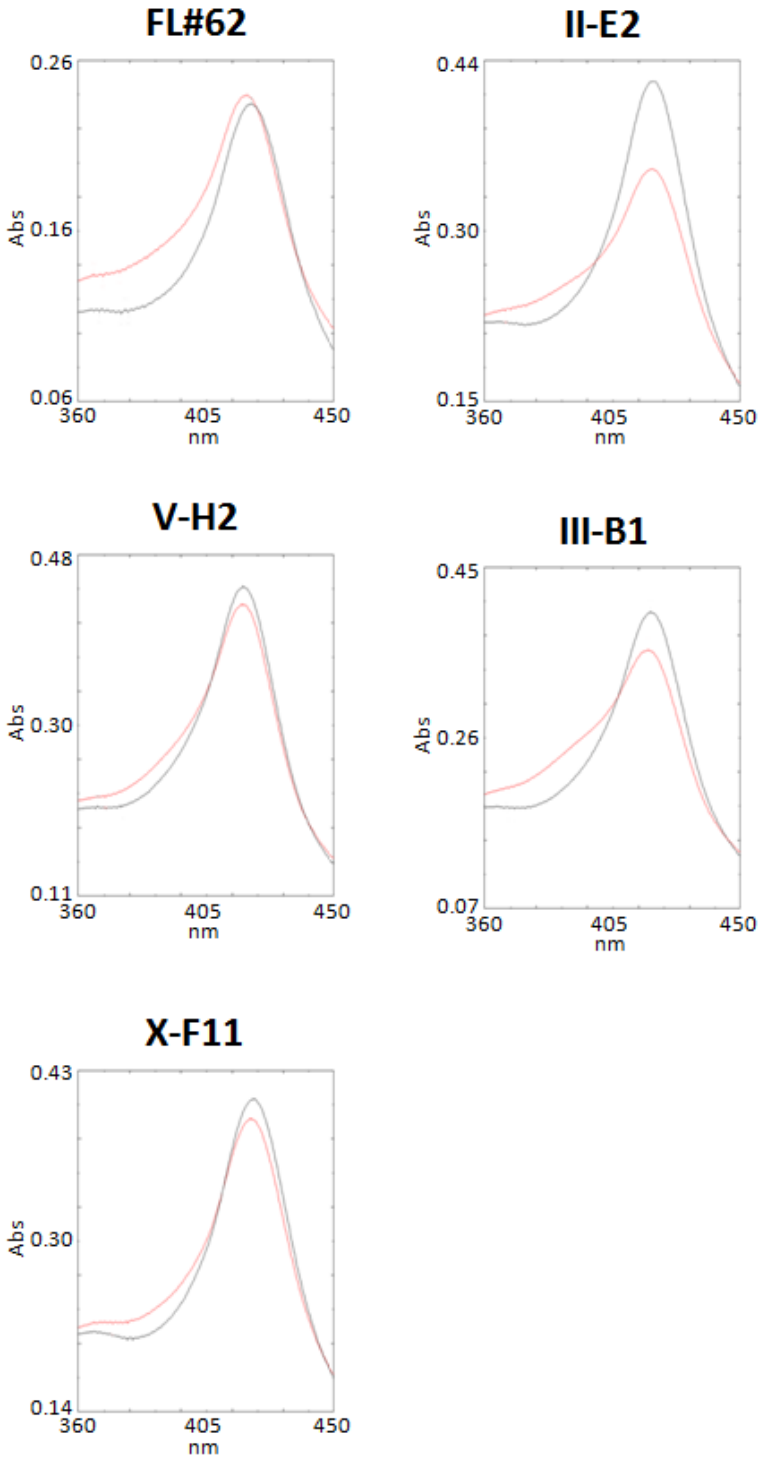


Figure S5. Representative plots of artemisinin-induced heme spin shift *versus* artemisinin concentration for the remaining P450 variants described in Table 1 (in addition to the data provided in **Figure 6**. The experimental data (dots) were fitted to a non-cooperative 1:1 binding model equation (solid line).

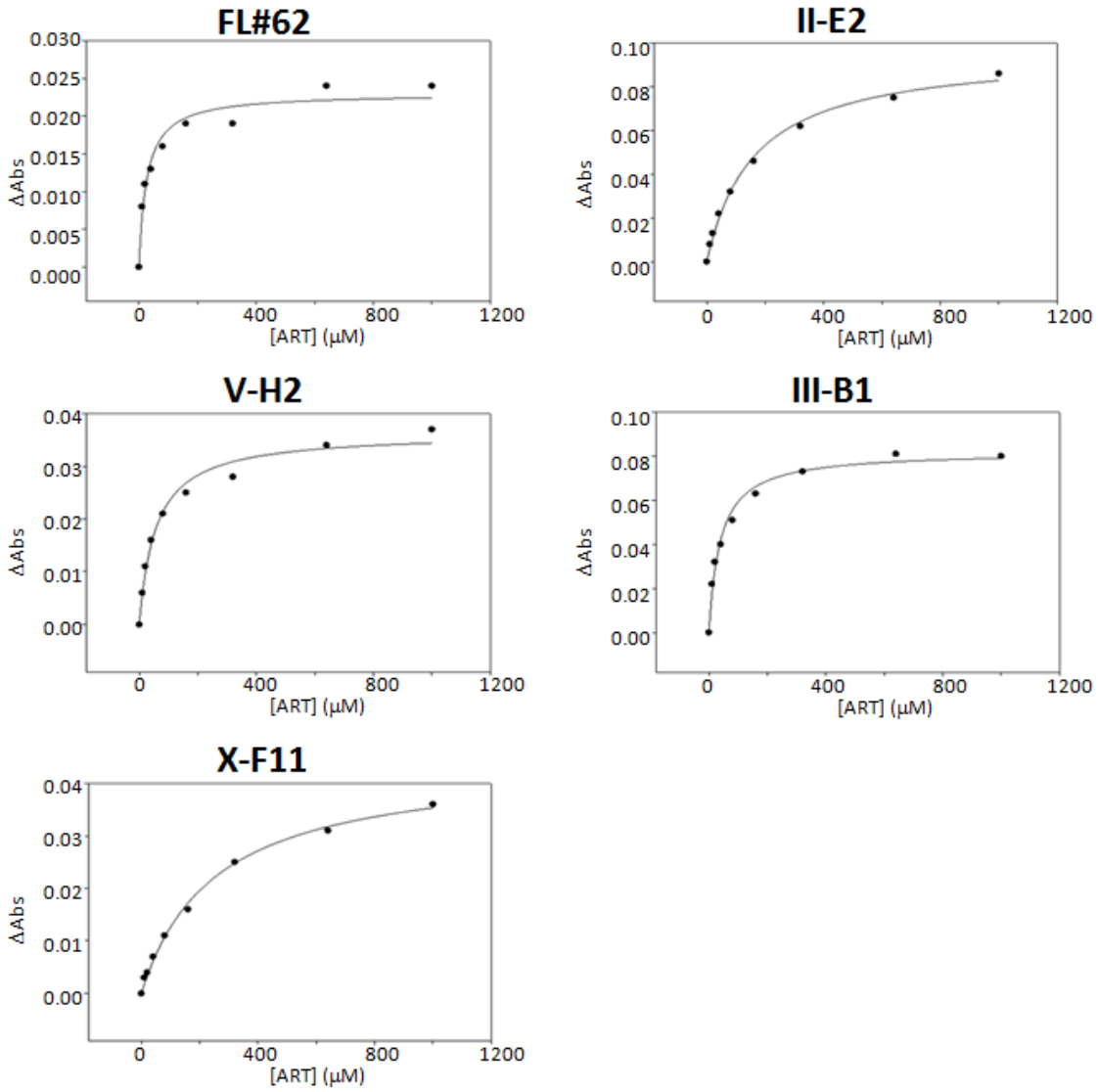
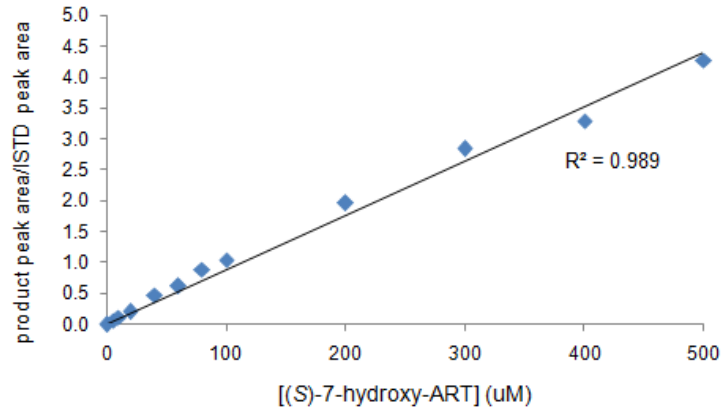
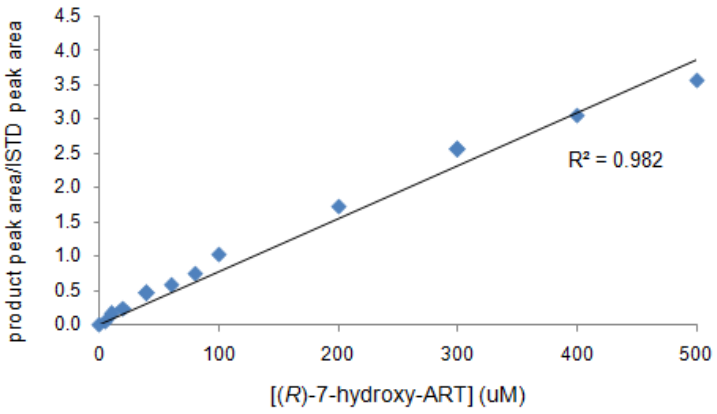


Figure S6. Calibration curves used for the quantification of the hydroxylated artemisinin derivatives **2** (top graph), **3** (middle graph), and **4** (bottom graph) by HPLC. The graphs report the ratio between the areas of the HPLC peaks corresponding to the hydroxy-artemisinin product and the internal standard (ISTD) plotted against the concentration of the hydroxy-artemisinin product.





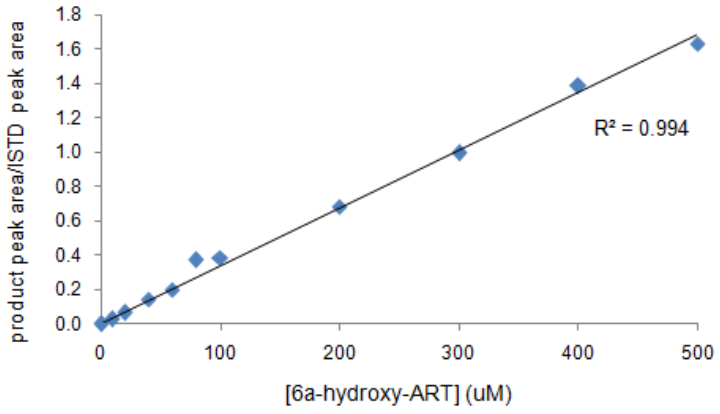


Table S1. Sequence of the oligonucleotides used for the preparation of the P450 libraries.

Primer	Sequence
74nnk_for	5'-GATAAAAACTTAAGTCAANNKCTTAAATTC-3'
74nnk_rev	5'-GAATTTAAGMNNTTGACTTAAGTTTTTATC-3'
78nnk_for	5'-AATTTNNKCGTGATAGTGTGGG-3'
78nnk_rev	5'-ACACTATCACGMNNAAATT-3'
81nnk_for	5'-CGTGATNNKGTGGGAGACGGGTTA-3'
81nnk_rev	5'-TAACCCGTCTCCCACMNNATCACG-3'
87nnk_for	5'-GACGGGTTANNKACAAGCTGGACGCATG-3'
87nnk_rev	5'-CATGCGTCCAGCTTGTMNNTAACCCGTC-3'
181nnk_for	5'-CGTGCANNKGATGAAGTAATGAACAAGC-3'
181nnk_rev	5'-GCTTGTTCATTACTTCATCMNNTGCACG-3'
184nnk_for	5'-GCACTGGATGAANNKATGAACAAG-3'
184nnk_rev	5'-CTTGTTCATMNNTTCATCCAGTGC-3'
78-87nnk_for	5'-AATTTNNKCGTGATAGTGTGGGAGACGGGTTANNKACAA-3'
78-87nnk_rev	5'-TTGTMNNTAACCCGTCTCCCACACTATCACGMNNAAATT-3'
78-81-82nnk_for	5'-AATTTNNKCGTGATNNKNNKGGAGACGGGTTA-3'
78-81-82nnk_rev	5'-TAACCCGTCTCCMNNMNNATCACGMNNAAATT-3'
81-82nnk_for	5'-CGCACGTGATNNKNNKGGAGACGGGTTA-3'
81-82nnk_rev	5'-TAACCCGTCTCCMNNMNNATCACGTGCG-3'
181-184nnk_for	5'-GTCCGTGCANNKGATGAANNKATGAACAAG-3'
181-184nnk_rev	5'-CTTGTTCATMNNTTCCTAMNNTGCACGGAC-3'

Table S2. Data corresponding to the P450 training sets used for calculating the fingerprint-based model for predicting ART reactivity: fingerprint components (= activity on probe **P1** to **P5**), ART turnovers (TTN), and relative ART activity (normalized to parent FL#62). The regio- and stereoselectivity of the variants are also indicated. The main table refers to training set A used for ranking the #FL62-derived variants, whereas training set B (ranking of II-E2-derived variants) included II-E2.

Variant	Fingerprint components					TTN	Rel. activity	Product distribution		
	P1	P2	P3	P4	P5			2	3	4
FL#62	11.875	10.800	5.500	5.000	9.111	339 ± 12	1.000	83%	10%	7%
0-#48	6.625	10.600	3.500	4.600	2.111	20 ± 1	0.059	27%	73%	0%
I-C1	6.941	0.727	1.232	0.988	0.983	0	0.000	-	-	-
I-F7	0.823	0.496	0.840	0.672	0.871	0	0.000	-	-	-
I-F9	0.892	0.495	0.872	0.622	0.872	0	0.000	-	-	-
I-G11	0.776	0.473	0.871	0.696	0.866	0	0.000	-	-	-
I-G3	2.312	7.194	5.930	4.174	1.250	0	0.000	-	-	-
II-C1	1.303	0.411	3.766	0.696	1.090	0	0.000	-	-	-
II-D10	0.572	0.236	0.737	0.592	0.885	0	0.000	-	-	-
II-G10	0.546	0.241	0.778	0.479	0.712	0	0.000	-	-	-
II-H4	0.519	0.209	0.620	0.433	0.941	0	0.000	-	-	-
III-A3	3.093	3.745	1.611	5.910	1.456	90 ± 13	0.267	70%	27%	3%
III-A4	0.670	0.322	0.791	0.549	0.930	0	0.000	-	-	-
III-B9	1.041	0.703	2.480	0.990	1.194	29 ± 2	0.086	0%	63%	37%
III-C7	2.791	3.071	1.915	4.736	1.630	106 ± 11	0.313	39%	59%	2%
III-C8	0.533	0.331	1.063	0.518	0.953	0	0.000	-	-	-
III-C9	0.509	0.277	0.870	0.465	0.967	0	0.000	-	-	-
III-E3	0.637	0.396	1.137	0.494	0.965	0	0.000	-	-	-
III-G3	0.775	0.300	0.870	0.563	0.870	0	0.000	-	-	-
III-H9	0.690	0.297	0.798	0.545	0.935	0	0.000	-	-	-
II-E2	3.626	3.144	1.172	6.469	1.267	393 ± 25	1.159	22	30	48

Table S3. Catalytic turnovers, regioselectivity, and stereoselectivity in artemisinin oxidation of the 50 top-scoring FL#62-derived P450 variants identified via fingerprint-based predictions. FL#62 is included for comparison. Stereoselectivity for C7-hydroxylation (expressed as diastereomeric excess, *de*) is indicated when higher than 60%. Variants are grouped according to their site-selectivity and highlighted as follows: red: variants with >85% 7(*S*)-selectivity; green: variants with >50% 7(*R*)-selectivity; blue: variants with >30% 6a-selectivity. Variants highlighted in darker colors correspond to those selected for further characterization (Table 1 in main text).

Variant	Product distribution			*% de (C7)*	TTN	Std Dev
	7(*S*)-OH-ART (2)	7(*R*)-OH-ART (3)	6a-OH-ART (4)			
FL#62	*83%*	*10%*	*7%*	*78% (S)*	*339*	*12*
IV-H4	100%	0%	0%	100% (*S*)	362	15
I-A1	100%	0%	0%	100% (*S*)	49	3
I-E2	100%	0%	0%	100% (*S*)	21	2
V-H2	96%	0%	4%	100% (*S*)	434	21
III-H2	95%	0%	5%	100% (*S*)	61	7
IV-C11	93%	0%	7%	100% (*S*)	454	17
I-G12	92%	0%	8%	100% (*S*)	413	6
I-H4	92%	0%	8%	100% (*S*)	498	23
I-D12	92%	0%	8%	100% (*S*)	287	8
VI-C2	91%	0%	9%	100% (*S*)	29	4
I-B12	91%	0%	9%	100% (*S*)	497	32
I-C12	89%	0%	11%	100% (*S*)	136	12
III-A11	89%	11%	0%	79% (*S*)	569	22
I-B4	89%	0%	11%	100% (*S*)	184	12
II-B12	81%	0%	19%	100% (*S*)	61	11
III-D4	80%	20%	0%	59% (*S*)	388	16
I-F5	79%	15%	6%	69% (*S*)	216	7
I-D2	79%	0%	21%	100% (*S*)	134	8
III-H10	78%	22%	0%		190	21
III-A5	78%	22%	0%		421	27
III-B2	77%	23%	0%		463	19

III-D9	77%	23%	0%		416	25
VI-G2	77%	20%	3%		475	16
II-C3	76%	24%	0%		602	16
II-H11	74%	18%	7%	60% (*S*)	103	10
VI-E10	72%	28%	0%		463	20
IV-C1	70%	28%	2%		462	21
II-D12	66%	22%	11%		22	6
VI-F8	66%	31%	3%		132	11
II-D1	65%	35%	0%		452	18
III-H4	64%	30%	6%		134	13
III-C10	63%	32%	4%		124	7
I-D1	62%	38%	0%		138	16
VI-F2	61%	33%	6%		105	11
IV-A11	61%	32%	7%		160	17
IV-F7	60%	34%	6%		119	12
I-C10	59%	32%	8%		153	9
II-H10	0%	100%	0%	100% (*R*)	270	8
III-B1	19%	81%	0%	62% (*R*)	403	17
I-B1	20%	75%	5%		388	15
II-E1	49%	51%	0%		426	23
III-B7	51%	49%	0%		75	4
V-D8	53%	47%	0%		428	19
IV-E10	53%	47%	0%		152	10
IV-C3	58%	42%	0%		422	13
II-H2	56%	40%	4%		67	3
II-E2	22%	30%	48%		393	25
II-A2	38%	29%	34%		21	6
III-F1	46%	21%	33%		45	6
II-D11	0%	0%	0%		0	0

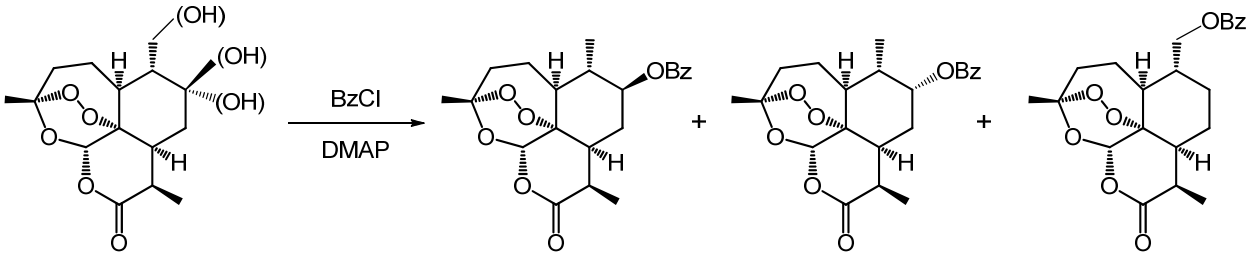
Table S4. Catalytic turnovers and regioselectivity in artemisinin oxidation of II-E2-derived P450 variants. II-E2 is included for comparison. The group highlighted in red corresponds to the ART-reactive variants identified among the 25 top-scoring P450s based on the fingerprint-driven predictions (ordered based on 6a-selectivity), with the underlining indicating those found among the 10 top-scoring ones. The groups highlighted in green and gray correspond to the ART-reactive variants found within the bottom 50% and bottom 10% fraction, respectively, of the 50-enzyme pool. Those highlighted in darker red were selected for further characterization (Table 1 in main text: [II-E2] II-G1 = X-E12; [II-E2] III-E3 = X-F11).

Variant	Product distribution			TTN	StdDev
	7(*S*)-OH-ART (2)	7(*R*)-OH-ART (3)	6a-OH-ART (4)		
II-E2	*22%*	*30%*	*48%*	*393*	*25*
[II-E2] II-G1 (="X-E12")	4%	2%	94%	113	12
[II-E2] III-E3 (="X-F11")	0%	8%	92%	376	19
[II-E2] III-D2	0%	8%	92%	46	2
[II-E2] XIV-B9	5%	5%	90%	87	*6*
[II-E2] VI-B10	0%	12%	88%	149	12
[II-E2] VI-H4	3%	12%	86%	70	4
[II-E2] II-E12	6%	9%	84%	115	17
[II-E2] XVII-E10	11%	8%	81%	171	9
[II-E2] II-D10	0%	23%	77%	105	11
[II-E2] VI-F4	11%	27%	63%	45	6
[II-E2] XII-F10	0%	43%	57%	88	*8*
[II-E2] II-A5	16%	33%	51%	58	7
[II-E2] XIII-B2	91%	0%	9%	474	23
[II-E2] VII-D3	0%	16%	84%	86	2
[II-E2] XV-E12	6%	13%	80%	59	9
[II-E2] VI-A2	0%	21%	79%	126	8

Materials, Methods, and Experimental Procedures

Reagents and Analytical Methods. Chemical reagents, substrates and solvents were purchased from Sigma-Aldrich, TCI, and Fluka. Silica gel chromatography purifications were carried out using AMD Silica Gel 60 230-400 mesh. UPLC/UV analyses were carried out on a Agilent 1200 spectrometer. 1D and 2D NMR experiments were carried out on a Bruker 500 MHz spectrometer. Data for ^{1}H NMR spectra are reported in the conventional form: chemical shift (δ ppm), multiplicity (s=singlet, d=doublet, t=triplet, q=quartet, m=multiplet, br=broad), coupling constant (Hz), integration). Data for ^{13}C NMR spectra are reported in terms of chemical shift (δ ppm). The oligonucleotides for the mutagenesis experiments were obtained from IDT DNA Technology. Restriction enzymes were purchased from New England Biolabs.

Derivatization and HPLC analysis of hydroxylated artemisinin products. To enable accurate measurement of the catalytic turnovers and regio/stereoselectivity of the P450 variants, the hydroxy-artemisinin products from the enzymatic reactions were derivatized with benzoyl chloride according to the scheme below followed by HPLC analysis of the resulting benzoyl ester derivatives.



Prior to analysis, the enzymatic reaction mixtures (0.5 mL) were added with 9-fluorenone (final conc.: 0.5 mM) as internal standard (ISTD) and extracted with 200 μL dicholoromethane. After evaporation of the organic solvent *in vacuo*, the crude product was re-dissolved in 100 μL dry

dichloromethane and reacted with excess of benzoyl chloride (1.25 μmol; 5 equiv.) in the presence of 4-dimethylaminopyridine (2.5 μmol; 10 equiv.) for 3 hours at room temperature to give the corresponding benzoylated derivatives. After evaporation of the solvent, the residue was dissolved in 100 μL acetonitrile and the mixture was separated by high-pressure reverse-phase chromatography using an Agilent 1200 UPLC instrument equipped with a photodiode UV-VIS array. Analytical conditions: Agilent XDB-C18 column (1.8 μm, 4.4 X 50 mm); UV detection: 230 nm; flow rate: 0.8 mL/min; solvent A: 1% trifluoroacetic acid in H_2O; solvent B: 1% trifluoroacetic acid in acetonitrile; gradient: 0-1 min: 20% B; 1-8 mins: 20% B to 90% B; 8-10 mins: 90% B. Retention times: 7.12 mins for 9-fluorenone (ISTD); 7.55 mins for 6a-OBz; 7.85 mins for 7(*R*)-OBz; 8.05 mins for 7(*S*)-OBz. Calibration curves were generated for each of the three hydroxylated artemisinin isomers using known concentrations of these compounds varying from 5 μM to 500 μM (Figure S6). The method showed a linear dependence of the signal (= analyte peak area/internal standard peak area) on the analyte concentration with a R-square value > 0.98 (see plots below) across the range of analyte concentration that were revelant for the enzyme characterization (2 μM – 1 mM). The detection limit of the method was determined to be about 2 μM.