

## Differences in Functional Clustering of Endogenous and Exogenous Substrates Between Members of the CYP1A Subfamily

### Supplementary Material

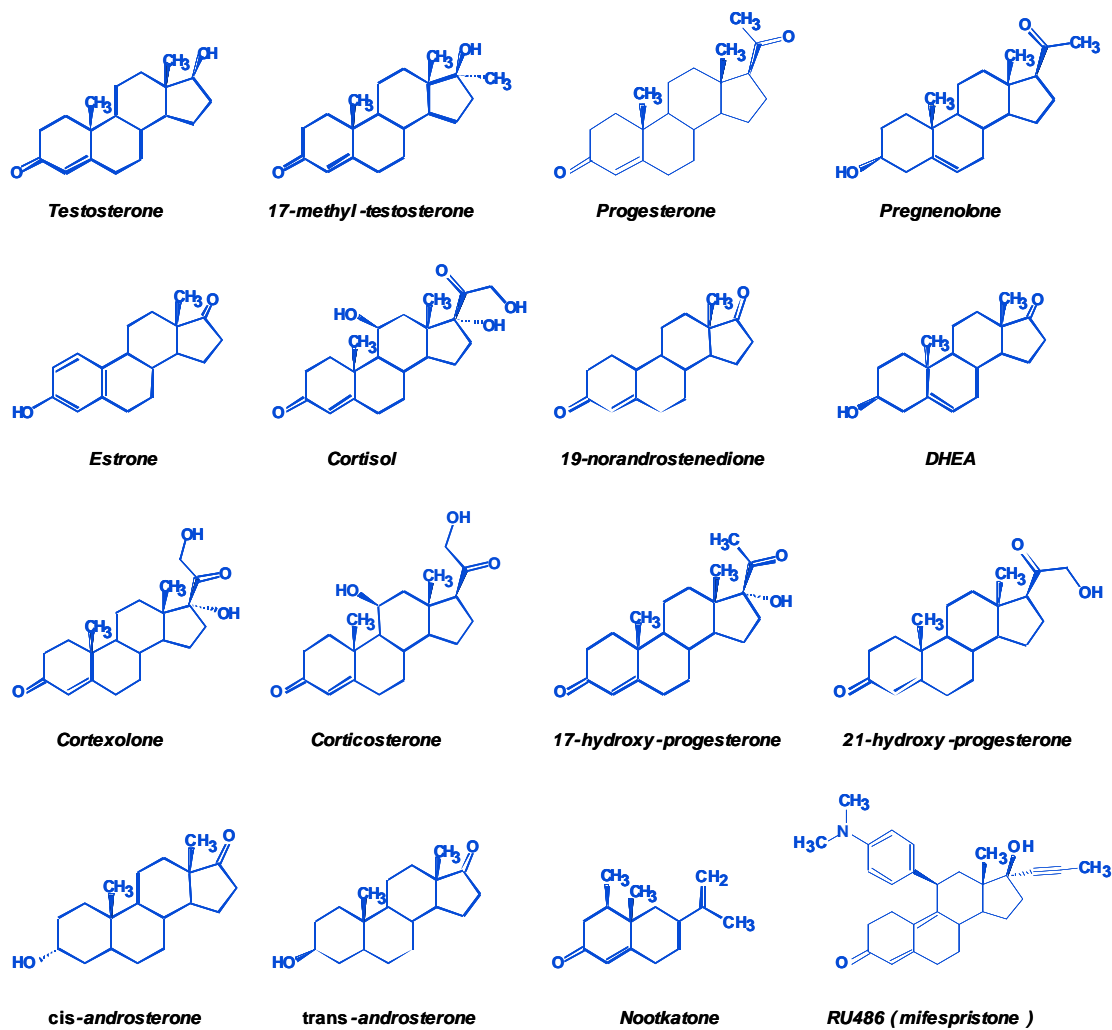


Fig. (S1). Structure of the steroid substrates used in this work.

**Table S1.** Observed rates of metabolite formation measured at a saturating concentration of steroid substrates. Metabolites were quantified by mass spectrometry after HPLC separation. Activities (units are  $\mu\text{V.s/min/mg}$  of microsomal protein) were calculated from initial velocities determined from metabolite concentration observed at different incubation times. The mammalian CYP1A enzymes were assayed in duplicate. An artificial mouse CYP1A1 - rabbit CYP1A2 chimera, the chimera 1ACh4, was also included. Enzymes were produced in yeast and incubations were performed with yeast microsomal fractions. This table is a non-processed data set

	Testosterone			RU486					17-methyl-testosterone			Cortisolone		
	T1	T2	T3	RU1	RU2	RU3	RU4	RU5	MT1	MT2	MT3	MT4	CTX1	CTX2
Hum 1A1-1	15968.6	142577.1	1693.1	207786.2	4229.5	7499.4	2682.8	5838.6	9044.8	113258.7	9658.6	8013.3	13116.4	50786.9
Hum 1A1-2	13320.3	118495.8	1407.1	336744.8	8750.9	5433.2	5061.3	10314.0	15129.7	118540.1	5619.6	12766.3	14679.8	68563.3
Hum 1A2-1	8.9	2082.5	8.9	11259.9	5.7	5.7	5.7	5.7	8.2	8.2	8.2	8.2	11.6	11895.9
Hum 1A2-2	8.9	3658.9	8.9	19089.3	5.7	5.7	5.7	5.7	8.2	8.2	8.2	8.2	11.6	9115.4
Mou 1A1-1	1588.8	67225.1	8.9	359506.6	5.7	5.7	5.7	1422.3	8.2	81412.3	8564.7	2429.0	28459.8	46343.1
Mou 1A1-2	980.2	45659.4	8.9	173040.0	5.7	5.7	5.7	967.0	8.2	53643.5	5378.3	4163.0	14154.4	14347.3
Rab 1A2-1	3596.5	82741.9	2103.9	1447504.5	5.7	5.7	5.7	5.7	8.2	63275.1	6716.7	1418.8	23720.8	61822.6
Rab 1A2-2	897.3	27970.9	2369.5	363536.2	5.7	5.7	5.7	5.7	8.2	36413.1	5792.7	1771.7	7307.9	19680.3
1ACh4	8.9	67956.8	4901.1	262670.3	5.7	5.7	5.7	665.5	8.2	79862.1	8382.6	5004.9	11.6	37463.7

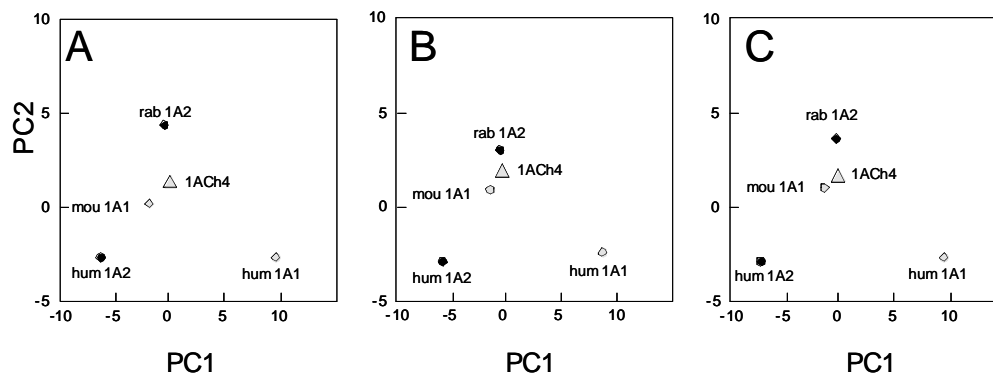
	Progesterone				Corticosterone				Pregnenolone				17-hydroxy-progesterone		DHEA
	PROG1	PROG2	PROG3	PROG4	PROG5	CTC1	CTC2	PREG1	PREG2	PREG3	PREG4	17HP1	17HP2	DHEA	
Hum 1A1-1	370480.7	281494.4	39430.5	22100.5	16171.6	20736.9	114565.3	259450.6	53248.9	58124.5	7286.5	35829.1	10599.6	1736.6	
Hum 1A1-2	285911.5	161149.2	22178.4	17003.9	13193.8	18948.7	71962.2	114860.5	22341.2	32804.7	88967.8	12512.7	11890.2	7690.4	
Hum 1A2-1	9.2	1315.8	2142.7	4292.6	19844.3	8.7	8.7	21.5	21.5	21.5	21.5	12.7	12.7	13.6	
Hum 1A2-2	9.2	1493.5	2431.1	2845.9	13156.3	8.7	8.7	21.5	21.5	21.5	21.5	12.7	12.7	13.6	
Mou 1A1-1	42550.1	158594.5	27784.1	12352.3	11663.8	17719.1	36029.6	16233.9	21.5	8397.0	21.5	11071.2	12.7	9972.8	
Mou 1A1-2	15884.1	84049.1	5656.1	1127.2	12436.6	8590.2	23908.0	14379.8	21.5	10115.9	21.5	14179.8	12.7	4011.2	
Rab 1A2-1	62340.4	226850.1	57741.0	49273.0	15679.9	24031.0	24289.1	9420.6	21.5	65718.9	17786.5	12.7	9394.7		
Rab 1A2-2	14950.1	105378.6	19904.8	18032.2	2000.7	4930.7	11366.5	8927.0	21.5	21.5	15982.8	1649.3	12.7	1550.6	
1ACh4	44029.9	171292.0	23072.1	5381.4	21657.4	36242.6	35247.7	27686.7	4594.4	21.5	21.5	16958.7	12.7	1335.1	

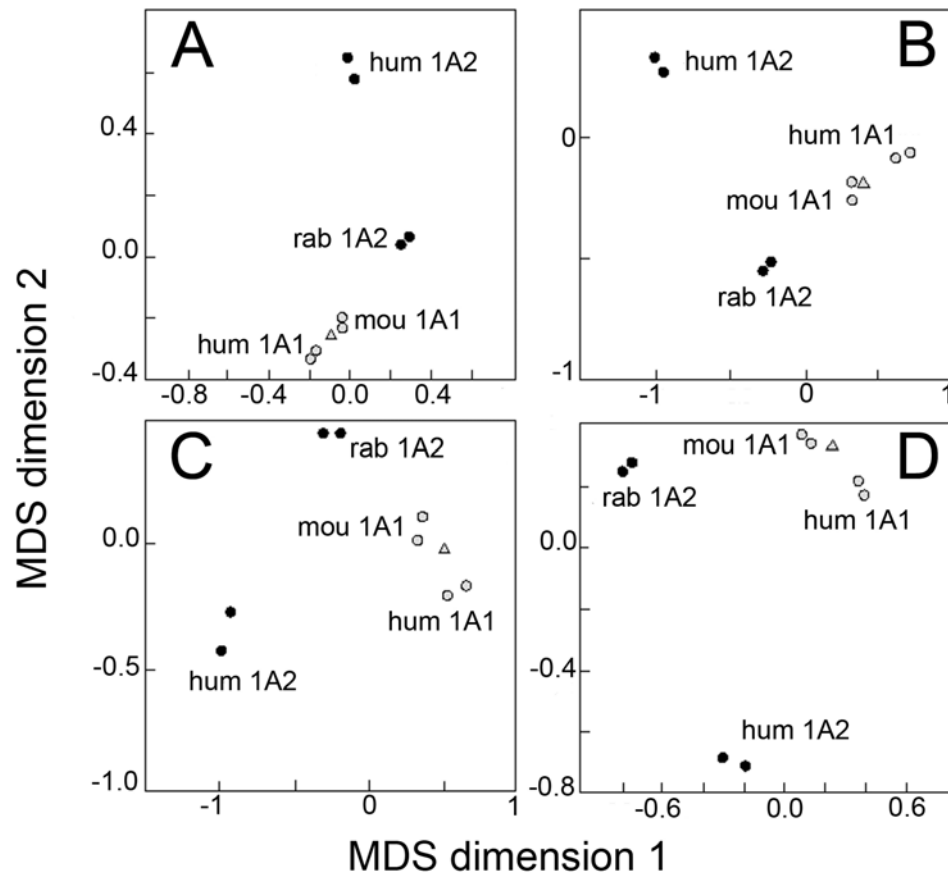
	DHEA		21-hydroxy-progesterone				Estrone				Nootkatone			
	DH2	DH3	21HP1	21HP2	21HP3	21HP4	EST1	EST2	NK1	NK2	NK3	NK4	NK5	NK6
Hum 1A1-1	14918.4	10735.9	41598.3	7968.3	167177.7	6.8	925421.8	56.2	18217.7	7378.2	33690.7	6110.2	37762.6	225429.4
Hum 1A1-2	14262.8	11507.1	16951.1	3140.8	168230.9	6.8	691001.1	56.2	19410.3	55321.3	21881.9	5247.4	33906.2	119705.2
Hum 1A2-1	13.6	29732.0	6.8	6.8	6491.1	6.8	56.2	56.2	7.8	7.8	7.8	7.8	7.8	7.8
Hum 1A2-2	13.6	8827.5	6.8	6.8	12780.8	6.8	56.2	56.2	7.8	7.8	7.8	7.8	7.8	7.8
Mou 1A1-1	7770.7	13.6	24070.9	6.8	24463.5	3025.1	422958.4	56153.0	7001.7	7.8	9526.2	7.8	10223.3	133203.7
Mou 1A1-2	3125.7	13.6	7816.6	6.8	16267.3	6413.6	202036.0	57058.5	4328.6	7.8	5890.0	7.8	10783.4	66713.9
Rab 1A2-1	6137.1	23499.7	60198.0	6.8	49081.4	10265.4	34488.2	56.2	7.8	7.8	26930.9	7.8	7.8	73688.3
Rab 1A2-2	1640.4	7808.8	26470.0	6.8	19427.8	1787.4	14083.2	56.2	7.8	7.8	8648.7	7.8	6309.3	23465.8
1ACh4	12540.8	13.6	22622.3	6.8	24688.0	6.8	478144.0	47845.9	7.8	7.8	7535.4	7.8	10630.2	122072.5

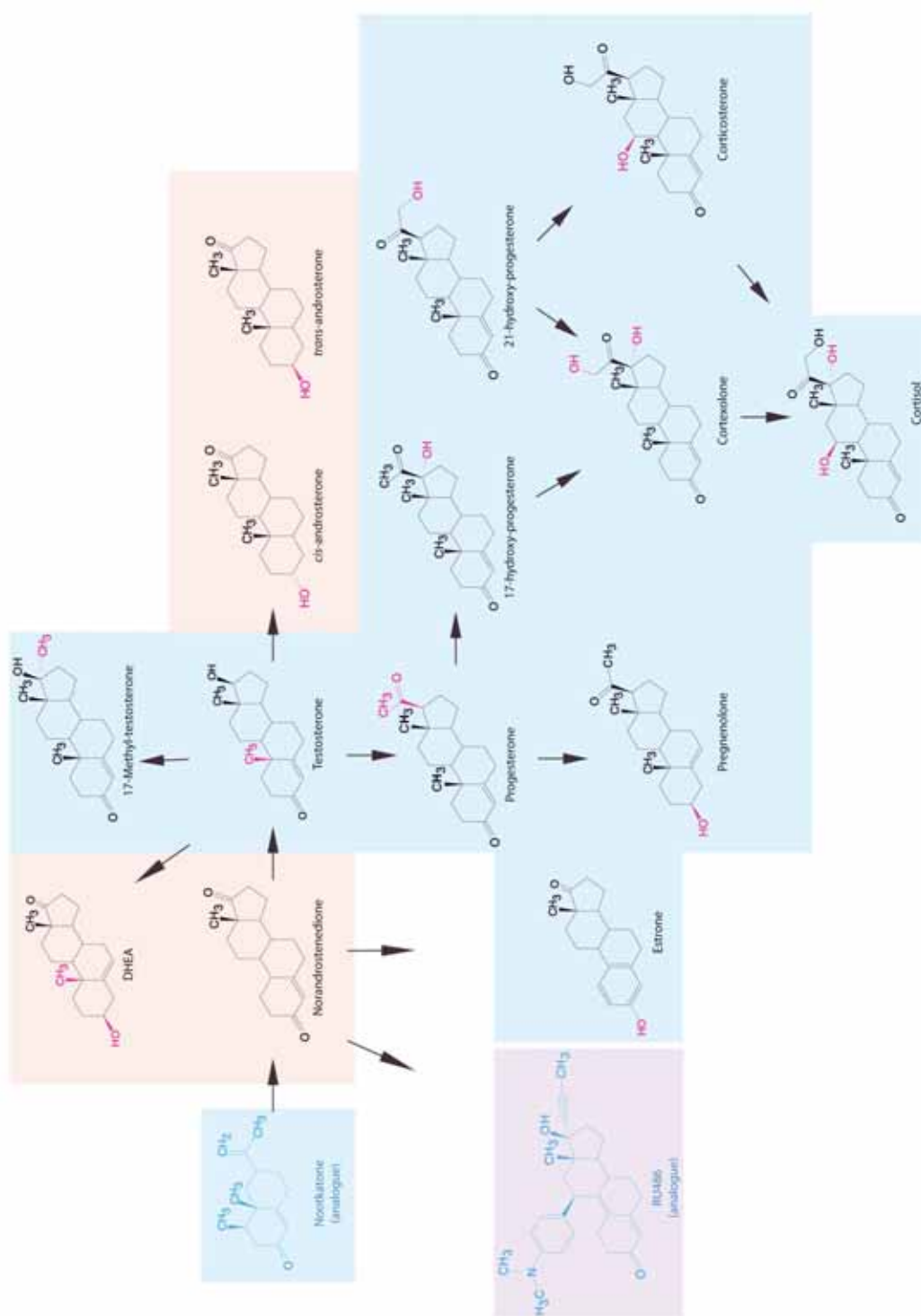
	Nootkatone		Cortisol				cis -Androsterone				Norandrosterone				trans -Androsterone	
	NK7	NK8	HC1	cAD1	cAD2	cAD3	cAD4	NAD1	NAD2	NAD3	NAD4	NAD5	tAD1	tAD2		
Hum 1A1-1	33744.4	22725.2	50649.0	11271.7	4723.5	86624.0	15.8	63350.3	64238.3	5.2	7661.0	6831.3	32122.7	31906.1		
Hum 1A1-2	15926.6	22725.2	34940.8	4627.2	8445.5	44510.3	15.8	38837.8	21644.7	5.2	4570.3	4075.4	11328.5	24377.3		
Hum 1A2-1	7.8	34303.7	13.2	15.8	14560.8	38146.9	15.8	9322.9	19108.4	27342.9	5.2	3925.5	18.1	20223.8		
Hum 1A2-2	7.8	7036.7	13.2	15.8	6597.7	6061.6	15.8	10944.7	22433.2	32100.7	5.2	4611.5	18.1	9882.7		
Mou 1A1-1	19040.8	9186.2	17901.5	8377.6	4998.4	39532.4	15.8	24894.1	14085.8	5.2	776.6	1449.0	2146.2	7065.0		
Mou 1A1-2	5044.3	2739.2	7909.4	3050.6	15.8	49447.1	15.8	13882.3	8942.9	5.2	1141.9	2130.8	3115.5	10386.3		
Rab 1A2-1	8950.5	56109.4	37051.1	14475.5	7181.7	87842.0	6748.8	64875.9	30281.6	5.2	9783.0	3243.1	6729.2	60889.9		
Rab 1A2-2	7.8	11541.2	7214.3	4970.0	1582.9	45154.8	3376.0	27400.2	14894.1	5.2	4629.2	2866.8	19025.3	32424.2		
1ACh4	17363.5	778.0	28039.8	8884.7	5128.0	76453.4	15.8	29745.5	59857.4	5.2	5.2	5.2	3794.2	12485.6		



**Fig. (S2).** Comparison of the principal component analyses on steroid dataset obtained with three different normalization procedures. (A). No normalization. (B). Normalization by the variance. (C). Normalization by the average substrate peak area. The first principal component accounts for 53.5 % of the variance in the initial dataset for panel A and for 60.2 % for both panels B and C. Grey circles, 1A1s; solid circles, 1A2s; grey triangle, chimera 1ACh4 (mostly mouse 1A1).



**Fig. (S3).** Comparison of MDS configuration plots obtained with four different models applied on steroid activity data. (A). Ratio MDS model, *stress* = 0.064. (B). Absolute MDS model, *stress* = 0.063. (C). Interval MDS model, *stress* = 0.064. (D). Polynomial MDS model, *stress* = 0.123. CYP1A1, grey; CYP1A2, black; chimera 1ACh4, triangle.



**Fig. (S4).** Step by step comparison of steroid and analogues from the simplest to the most complex. The background colour refers to CYP1A functional clusters and the substituent that differs from the previous molecule at each step is coloured in magenta.

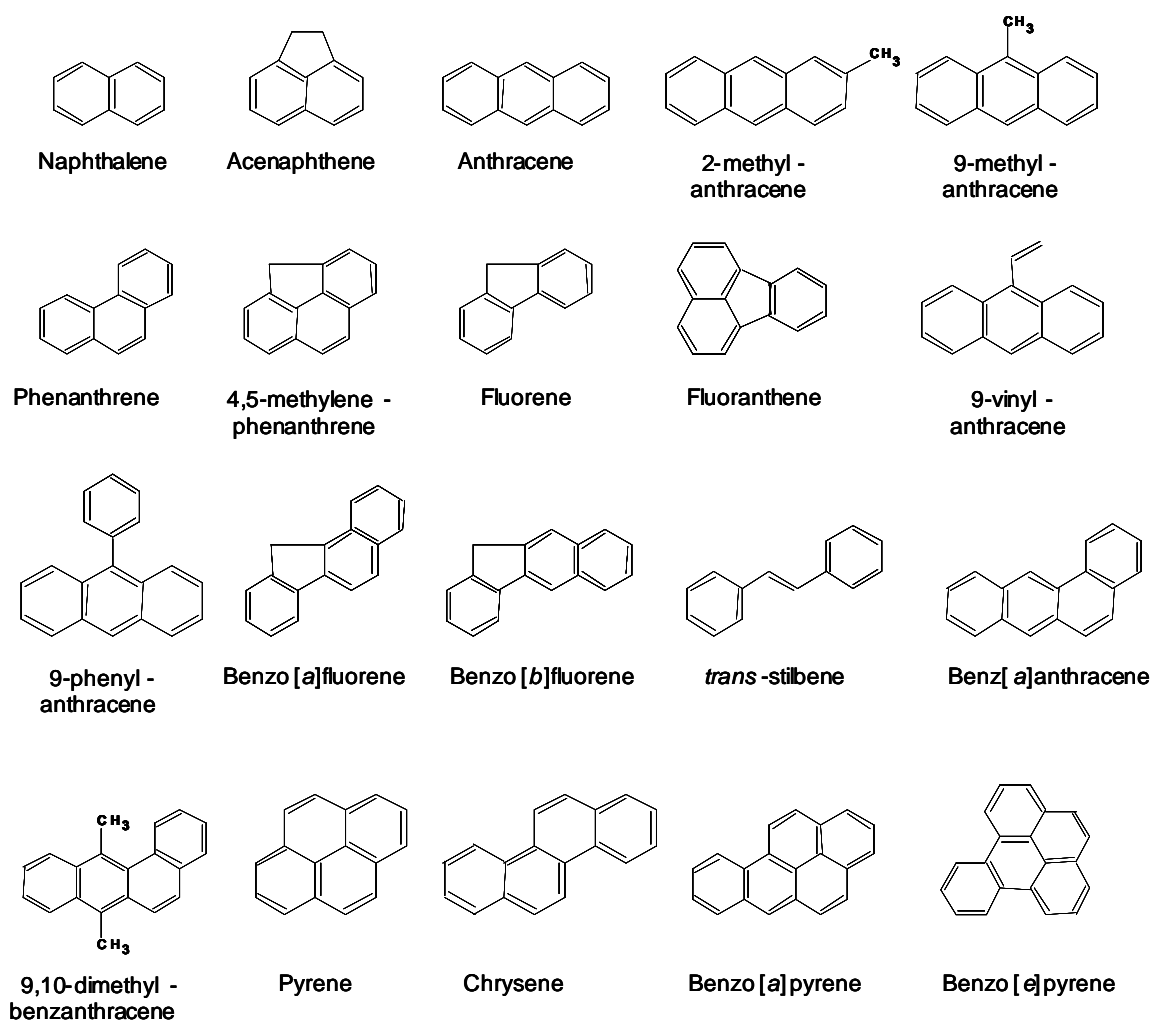


Fig. (S5). Structure of the PAH substrates used in this work.

**Table S2. Observed rates of metabolite formation measured at a saturating concentration of polycyclic aromatic hydrocarbon substrates. Metabolites were quantified by fluorescence after HPLC separation. Activities were calculated from initial velocities determined from metabolite concentration observed at different incubation times. Enzymes were produced in yeast and incubations were performed with yeast microsomal fractions at which recombinant human microsomal epoxide hydrolase was added. Metabolites were quantified by fluorescence after HPLC separation. Activities units are fluorescence arbitrary units/min/mg of microsomal protein. This table is a non-processed data set**

	EOR	MOR	EFEE	naphthalene		acenaphthene				trans-stilbene			
				naphthol-1 Rt= 13.4 min	naphthol-2 Rt= 13.8 min	Acen1 Rt= 10.6 min	Acen2 Rt= 13.3 min	Acen3 Rt= 13.8 min	Acen4 Rt= 16.4 min	Acen5 Rt= 33.4 min	Stib1 Rt= 3.8 min	Stib2 Rt= 6.0 min	Stib3 Rt= 6.2 min
Hum 1A1	414	123	453	837 111	946 290	10 352 881	5 543 982	5 571 793	1 005 072	1 603 802	102 225	103 336	43 282
Hum 1A2	68	108	36	242 715	278 470	2 397 986	3 147 100	7 587 007	264 116	607 818	3 410	25 336	30 789
Mou 1A1	821	262	59	1 024 309	1 294 024	2 434 277	7 754 512	6 462 494	289 345	331 111	40 413	87 611	59 619
Rab 1A2	17	207	323	471 789	576 444	2 759 935	5 340 558	9 330 774	291 482	350 360	8 880	15 867	23 523

	anthracene		2-methyl-anthracene			9-methyl-anthracene			phenanthrene				
	1,2-Diol Rt= 12.8 min	2man1 Rt= 10.5 min	2man2 Rt= 11.2 min	2man3 Rt= 11.5 min	2man4 Rt= 14.0 min	2man5 Rt= 14.4 min	9man1 Rt= 2.8 min	9man2 Rt= 4.2 min	Phe1 Rt= 10.2 min	Phe3 Rt= 12.6 min	Phe5 Rt= 16.3 min	Phe6 Rt= 16.7 min	Phe7 Rt= 17.2 min
Hum 1A1	10 851 304	5 034 649	857 804	5 810 073	19 752 853	24 148 599	37 542 671	20 917 129	1 160 198	575 780	1 725 960	771 272	910 460
Hum 1A2	994 461	57 979	52 781	10 501 105	3 642 542	1 027 601	3 099 816	4 738 593	307 717	42 389	321 945	111 941	175 464
Mou 1A1	15 130 582	1 074 805	294 962	4 684 619	19 811 595	22 934 834	12 527 464	22 687 667	1 475 968	489 406	3 311 146	1 000 236	1 882 976
Rab 1A2	5 324 524	64 842	97 608	858 056	9 186 853	2 825 621	5 125 815	8 706 386	428 531	58 779	588 051	138 288	399 105

	4,5-methylene-phenanthrene			fluorene			fluoranthene			9-phenyl-anthracene			
	Mphe1 Rt= 3.2 min	Mphe2 Rt= 3.6 min	Mphe3 Rt= 5.0 min	Mphe4 Rt= 5.9 min	Flu1 Rt= 3.5 min	Flu2 Rt= 5.3 min	Flu3 Rt= 5.7 min	Flua1 Rt= 1.7 min	Flua2 Rt= 3.2 min	Flua3 Rt= 3.6 min	Flua4 Rt= 5.6 min	Pant1 Rt= 4.7 min	Pant2 Rt= 5.9 min
Hum 1A1	1 706 036	316 189	1 390 711	5 044 405	247 940	785 375	474 180	282 591	157 925	73 570	1 491 810	425 777	16 519 379
Hum 1A2	453 873	2 378 449	64 051	5 121 140	158 830	4 079 965	247 608	1	4 995	1	40 350	28 209	2 308 200
Mou 1A1	2 334 887	103 153	4 031 200	8 375 819	1 088 120	1 273 435	908 761	315 816	131 240	84 433	1 832 330	196 575	3 972 726
Rab 1A2	611 144	1 286 706	403 704	6 152 602	800 500	1 395 591	1 540 200	112 716	12 020	4 207	582 840	31 370	32 176

	9-vinyl-anthracene				benzo[ a ]fluorene								
	Pant3 Rt= 8.3 min	Pant4 Rt= 9.1 min	Van1 Rt= 1.5 min	Van2 Rt= 2.3 min	Van3 Rt= 2.5 min	Van4 Rt= 3.2 min	Van5 Rt= 4.7 min	BaF1 Rt= 3.5 min	BaF2 Rt= 4.0 min	BaF3 Rt= 4.4 min	BaF4 Rt= 4.9 min	BaF5 Rt= 5.7 min	BaF6 Rt= 6.0 min
Hum 1A1	591 345	257 215	347 721	736 833	3 277 964	11 843 369	4 339 120	1 951 060	616 752	519 212	360 947	9 835 647	2 245 378
Hum 1A2	243 157	267 760	86 461	83 440	254 899	109 559	1 125 415	41 080	37 973	38 991	5 830	515 413	112 843
Mou 1A1	73 386	102 466	122 466	90 178	1 843 203	1 617 903	2 478 950	399 757	793 621	330 031	400 761	7 437 370	2 883 969
Rab 1A2	115 207	210 320	118 366	102 526	124 456	319 660	637 985	24 935	59 881	25 580	7 510	534 032	336 804

	benzo[ b ]fluorene			dimethylbenzanthracene			benzanthracene						
	BaF7 Rt= 6.2 min	BbF1 Rt= 3.5 min	BbF2 Rt= 5.7 min	BbF3 Rt= 5.9 min	Dmb1 Rt= 9.5 min	Dmb2 Rt= 9.8 min	Dmb3 Rt= 12.6 min	Dmb4 Rt= 13.5 min	Dmb5 Rt= 24.2 min	Bza1 Rt= 9.2 min	Bza3 Rt= 14.8 min	Bza4 Rt= 21.9 min	Bza5 Rt= 22.8 min
Hum 1A1	2 016 339	4 004 031	8 555 598	1 227 491	163 433	94 298	1 554 291	1 269 153	751 156	179 461	5 551 689	311 627	208 555
Hum 1A2	176 254	32 865	927 579	175 060	1 299	503	6 485	29 187	20 081	5 285	77 829	18 166	1 069
Mou 1A1	1 964 002	268 476	3 484 120	1 927 936	452 208	68 416	2 756 928	763 964	843 995	75 733	8 173 778	111 165	54 156
Rab 1A2	246 604	18 264	708 184	228 532	460	113	24 543	17 444	18 229	2 197	175 975	21 972	973

	chrysene				pyrene				benzo[ a ]pyrene				
	Bza6 Rt= 23.3 min	Chr1 Rt= 8.2 min	Chr2 Rt= 10.8 min	Chr3 Rt= 14.3 min	Chr4 Rt= 15.2 min	Chr5 Rt= 16.6 min	Chr6 Rt= 20.9 min	Pyr1 Rt= 13.5 min	Pyr2 Rt= 19.7 min	9,10diol-BaP Rt= 13.2 min	4,5diol-BaP Rt= 16.3 min	7,8diol-BaP Rt= 17.2 min	9OH-BaP Rt= 25.0 min
Hum 1A1	132 469	1 113 286	4 468 349	374 674	302 181	524 079	366 212	2 577 028	23 744 886	277 935	66 456	808 386	1 046 988
Hum 1A2	27 317	14 881	75 703	5 432	1 598	3 562	8 029	50 371	1 702 893	1	1	2 227	1
Mou 1A1	121 335	584 841	3 240 499	218 368	214 032	384 513	412 853	1 975 277	7 435 411	182 550	128 041	424 771	291 468
Rab 1A2	40 844	1 768	145 852	5 939	352	271	11 585	43 548	2 141 165	1	1	9 639	1

	benzo[ e ] pyrene						
	3OH-BaP Rt= 25.9 min	BeP1 Rt= 14.9 min	BeP2 Rt= 16.5 min	BeP3 Rt= 17.3 min	BeP4 Rt= 19.5 min	BeP5 Rt= 23.4 min	BeP6 Rt= 24.0 min
Hum 1A1	462 301	147 579	69 903	952 923	72 234	6 520 628	431 123
Hum 1A2	1	132	7 206	165	132	116 909	34 622
Mou 1A1	1 487 534	219 911	247 829	682 507	339 481	1 847 802	237 560
Rab 1A2	1	18 367	12 588	10 571	3 254	196 690	49 702

rabbit_1A2	..NAMSPPAALSVTELLLVSAVFLVFWAVRERPKVPPK	38
human_1A1	...MLFFISMSATEFLLASVIFCLVFWVIRASRPQVPPK	36
mouse_1A1	MPSMYGLPAFVVSATELLAVTVFCLGFWVIRTRTWVPPK	40
human_1A2	..MALSQSVPPFSATELLLASAIFCLVFWVIRLPRVPPK	38
rabbit_1A2	LKRLPGFNGMFLLGHLITLGNPHVALAFLERRYGQVFI	78
human_1A1	LKNEPGFNGMFLIGHMLTIGKNPHIALSRMSQQYGDVLIQI	76
mouse_1A1	LKTFPGFNGLFFIGHMLTVGKNPHLSLTRLQQYGDVLIQI	80
human_1A2	LKSPPEFNGMFLLGHVLTIGKNPHIALSRMSQRYGDVLIQI	78
rabbit_1A2	RIGSTFVWLSGLDTIQALVRQGDFFKGRPDLYSSSFIT	118
human_1A1	RIGSTFVWLSGLDTIRQALVRQGDFFKGRPDLYTFTLIS	116
mouse_1A1	RIGSTFVWLSGLNTIQALVRQGDFFKGRPDLYSFTLIT	120
human_1A2	RIGSTFVWLSGLDTIQALVRQGDFFKGRPDLYTSTLIT	118
rabbit_1A2	EQQSDFSPDSGPVWAARRRLAQDSIRSFASNPASSSS	158
human_1A1	NGQSDFSPDSGPVWAARRRLAQNGIKSFIASDPASSTS	156
mouse_1A1	NGKSDFFMDSGPVWAARRRLAQNALISFASDPASSS	160
human_1A2	DQSDFFSPDSGPVWAARRRLAQNALISFASDPASSSS	158
rabbit_1A2	CYLEEHVSQEAENLISRQELMAAVGRFDPYSQLVVSAAAR	198
human_1A1	CYLEEHVSKEAEVLISLQELMAGPGFNPYRQVVSVTN	196
mouse_1A1	CYLEEHVSKEANYLVSKLQVMAEVGHFDPYKLVVSVAN	200
human_1A2	CYLEEHVSKEAKALISRLQELMAGPGHFDPYQVVSVAN	198
rabbit_1A2	VIGAMCFGRHFFQOSEENLDVVRNSKFPVETASSSSPVD	238
human_1A1	VICAICFGRRYDRHNAELLSLVNINNNPGEVWVSGNPAF	236
mouse_1A1	VICAICFGQRYDRHDDQELLSVNLSEPEVTVSGVPAF	240
human_1A2	VIGAMCFQHFESSDENLSLVKNTHEFVETASSGNPLDF	238
rabbit_1A2	FPILRYLPNRPLORFKFNQRFNLFQKTVREHYEDFRN	278
human_1A1	IPILRYLPNPSLNAPFNLEKFFSFMQHMVKEHYKTFEKG	276
mouse_1A1	IPVLRVLPNSSLDAFQNLNDFYSPMKKLIKEHYRTFENG	280
human_1A2	FPILRYLPNAPLQRFKFNQRFNLFQKTVQEHYQDFDN	278
rabbit_1A2	SIQDITGALFKHSEKSKANGGLIPQ..EKIVNLVNDIFG	316
human_1A1	HIRDITDSLIEHCQEKQLDENANVQLSDEKINIVLDFG	316
mouse_1A1	HIRDITDSLIEHCQDKRKLDEANVQLSDDKVVITVLDLFG	320
human_1A2	SVRDITGALFKHSEKSKGPRASGNLIPQ..EKIVNLVNDIFG	316
rabbit_1A2	AGFDVTITALSWSLMYLVTNRSVQRKICSELDVAVGRPQ	356
human_1A1	AGFDVTITALSWSLMYLVTNRSVQRKICSELDTVIGRSRR	356
mouse_1A1	AGFDVTITALSWSLMYLVTNRSVQRKICSELDTVIGDRQ	360
human_1A2	AGFDVTITALSWSLMYLVTNRSVQRKICSELDTVIGRERR	356
rabbit_1A2	<u>PRLSDRPQLPYLEAFILEFRHSHPEFTIPHSTTRDTTL</u>	396
human_1A1	<u>PRLSDRSHLPYMEAFILEFRHSHPEFTIPHSTTRDTSL</u>	396
mouse_1A1	<u>PRLSDRPQLPYLEAFILEFRHSHPEFTIPHSTTRDTSL</u>	400
human_1A2	<u>PRLSDRPQLPYLEAFILEFRHSHPEFTIPHSTTRDTTL</u>	396
rabbit_1A2	NGFYIPKKECCIFINQWQINHDQKLVNPFSEFPERFLTAD	436
human_1A1	KGFYIPKGRCVFNQWQINHDQKLVNPFSEFPERFLTAD	436
mouse_1A1	NGFYIPKCCVFNQWQVNHDRLENGDPNEFRPERFLTAD	440
human_1A2	NGFYIPKCCVFNQWQVNHDFELMEDPSEFRPERFLTAD	436
rabbit_1A2	GAAINKPLSEKTLFGLGKRRKICGEIARWVFLFLAILL	476
human_1A1	GAIDKV.LSEKTLIFGMGKRRKICGEIARWVFLFLAILL	475
mouse_1A1	GTLDKR.LSEKTLFGLGKRRKICGEIARWVFLFLAILL	479
human_1A2	GTAINKPLSEKTLFGMGKRRKICGEIARWVFLFLAILL	476
rabbit_1A2	QRLEFSVPPGVVDLTFIYGLTMKHFRCEHVQARFRFSDD	516
human_1A1	QRVEFSVPLGVKVDMTPIYGLTMKHACCEHFQQLRS...	512
mouse_1A1	QQIEPKVSPGKVDMTPTTYGLTLKHARCEHFQVQRSSGF	519
human_1A2	QQLEFSVPPGVVDLTFIYGLTMKHARCEHVQARFRFSIN.	515

**Fig. (S6).** Multiple alignment of CYP1A amino acid sequences with positions characterizing differential steroid specificity coloured in blue. The J helix element is underlined.