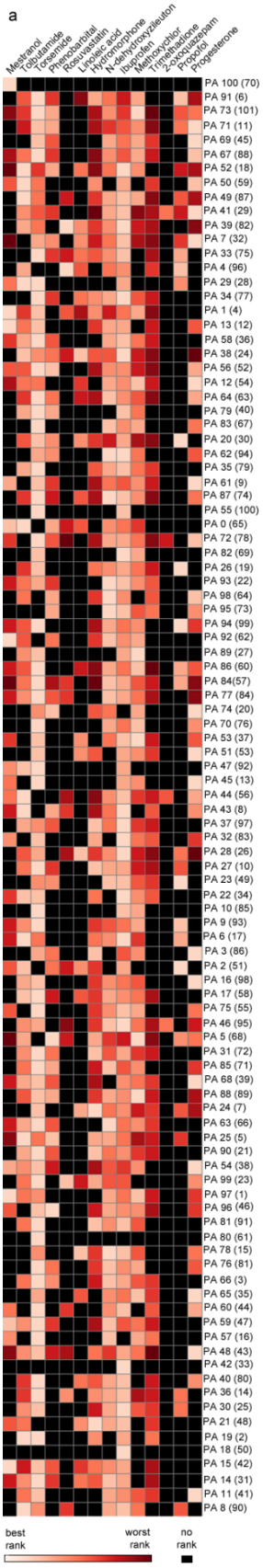
SUPPLEMENTARY INFORMATION

Calculation of Binding Site Volume

POVME was used to calculate the binding site volumes. Grid points were generated every 1.0Å and the center of the docking box was selected as the center of the InclusionBox with an edge length of 18Å (approximate length of the docking box used for docking). Discontiguous points were removed by setting the ContiguousPointsCriteria to 3. The ContiguousSeedBox was set with the same center as used in the initial points but the box was reduced to have an edge length of 5Å.



Supplementary Figure 1. SoM prediction using AutoDock Vina and the 14 ligand set in the pseudo-apo ensemble. The structure rank has been denoted.

Supplementary Table 1. Ranking of CYP2C9 SoMs by the methods tested in this study.

Substrate Name	Data Set[a]	Rank						
			Static Docking			Pseudo-Apo Ensemble		
		SMARTCyp alone	Vina Alone	Vina +SMARTCyp + QSAR	Vina+SMART Cyp + QSAR	Vina Alone	Vina +SMARTCyp + QSAR	Vina+SMART Cyp + QSAR
Flurbiprofen	T	2	4	1	1	1	1	1
Naproxen	N	1	n/a	n/a	n/a	1	1	1
Zolpidem	T	1	3	1	5	2	2	1
Tienilic acid	T	2	2	2	3	2	3	1
GV150526	T	2	1	1	1	4	1	1
Fluvistatin	N	7	2	3	1	1	1	2
Indomethacin	N	1	5	1	1	1	1	1
Tolbutamide	F,N	1	1	1	1	3	2	1
Desogestrel	N	1	n/a	n/a	n/a	1	1	1
Dibenzo[a h]anthracene	T	1	1	1	1	1	1	1
Losartan	N	1	n/a	n/a	n/a	1	1	1
Piroxicam	N	6	n/a	n/a	n/a	1	1	1
Mefenamic acid	N	1	4	1	1	1	1	1
Chlorpropamide	N	8	n/a	n/a	n/a	n/a	n/a	n/a
Amitriptyline	T	1	7	1	1	1	1	1
Celecoxib	T	1	n/a	n/a	n/a	5	1	1
Mestranol	F,N	1	n/a	n/a	n/a	6	1	1
Etodolac	N	4	n/a	n/a	n/a	10	8	1
17 alpha-ethinylestradiol	N	5	2	1	1	7	2	1
Galangin	T	3	1	2	1	1	1	1
Sildenafil	N	1	n/a	n/a	n/a	2	1	1
Aceclofenac	N	2	7	1	1	1	1	1
Tolterodine	T	15	3	6	2	1	3	1
Meloxicam	N	1	n/a	n/a	n/a	3	1	1
Acenocoumarol	T	4	5	3	1	7	2	1
Methoxychlor	F,N	1	n/a	n/a	n/a	1	1	1
Clozapine	N	1	5	1	1	4	1	1
Terbinafine	T	1	6	1	1	3	1	1
Tenoxicam	N	6	n/a	n/a	n/a	1	1	1
Methyl Eugenol	N	3	3	3	1	4	3	1
(s)-lansoprazole	N	8	n/a	n/a	n/a	1	2	2
N-dehydroxyzileuton	F,N	6	1	3	2	1	2	2
Ibuprofen	F,N	4	1	2	1	2	1	1
Diclofenac	T	1	3	1	1	1	1	1
Valproic acid	N	2	1	1	2	3	2	1
Arachidonic acid	N	9	n/a	n/a	n/a	n/a	n/a	n/a
Kaempferide	N	1	4	1	1	7	3	1
DMZ	T	1	n/a	n/a	n/a	1	1	1
Limonene	N	2	4	2	2	2	2	1

Safrole	N	2	2	2	1	1	2	1
Warfarin	T	4	n/a	n/a	n/a	1	2	1
Serotrodast	N	1	1	1	1	10	3	1
Lornoxicam	T	5	n/a	n/a	n/a	2	1	1
Fluoxetine	N	1	3	1	1	7	1	1
Phenytoin	T	1	5	1	1	1	1	1
Suprofen	N	2	7	1	1	2	2	1
Zafirlukast	N	3	2	2	1	2	2	1
Phenprocoumon	T	3	3	2	1	3	2	1
Ketamine	N	1	10	1	3	12	2	1
58C80	N	11	1	2	1	1	1	1
Delta-9-tetrahydrocannabinol	N	2	1	1	1	1	1	1
Torsemide	F,T	1	n/a	n/a	n/a	4	1	1
Linoleic acid	F,N	6	9	2	1	8	6	1
Hydromorphone	F,N	1	7	1	1	4	1	1
Carvedilol	N	4	n/a	n/a	n/a	13	11	1
TR-14035	N	1	n/a	n/a	n/a	1	1	1
Montelukast	N	22	8	8	3	6	3	1
Hexobarbital	T	1	6	1	1	5	1	1
Rosuvastatin	F,N	7	n/a	n/a	1	6	3	1
Tamoxifen	N	8	1	3	n/a	2	4	1
Zopliclone	N	1	8	1	3	1	1	1
Phenobarbital	F,N	2	3	2	1	3	1	1
s-MTPPA	N	3	9	2	1	1	1	1
2-oxoquazepam	F,N	1	1	1	2	n/a	n/a	n/a
rosiglitazone	N	2	n/a	n/a	n/a	1	1	1
sulfinpyrazone	N	3	1	1	1	1	1	1
trimethadione sulfate	F,N	1	3	1	1	1	1	1
alitretinoin (9-cis-retinoic acid)	N	8	n/a	n/a	n/a	1	2	2
flunitrazepam	N	3	1	1	1	5	3	1
muraglitazar	N	1	n/a	n/a	n/a	2	2	1
propofol	F,N	16	n/a	n/a	n/a	3	8	2
SCH351125	N	1	n/a	n/a	n/a	4	1	1
progesterone	F,N	25	1	8	2	1	8	2

[a] F= ligand used in optimizing the fitness function in the *SMARTCyp +Docking* approach and to rank protein structures in ensemble optimization, N= ligand used in the QSAR training set, and T= ligand used in the QSAR test set.