Supporting Information

Comprehensive and Automated Linear Interaction Energy Based Binding-Affinity Prediction for Multifarious Cytochrome P450 Aromatase Inhibitors

*Marc van Dijk,[1] Antonius M. ter Laak,[2] Jörg D. Wichard,[2] Luigi Capoferri,[1] Nico P. E. Vermeulen,[1] Daan P. Geerke[1]**

[1]AIMMS Division of Molecular Toxicology, Department of Chemistry and Pharmaceutical Sciences, Vrije Universiteit Amsterdam, De Boelelaan 1108, 1081 HZ Amsterdam, The Netherlands, and [2]Bayer AG, Pharmaceuticals Division, Müllerstrasse 178, D-13353 Berlin, Germany

* d.p.geerke@vu.nl

Rule-based protocols for resolving eight types of protein-ligand interactions using in-house python software

Hydrogen-bonded contacts: Pre-select all heavy atom contacts within 0.41 nm, not involving water molecules or ions. Identify donor-acceptor pairs according to their SYBYL atom type (acceptors: N.3, N.2, N.1, N.acid, N.ar, O.3, O.co2, O.2, S.m, S.a; donors: N.3, N.2, N.acid, N.am, N.ar, N.4, N.pl3, N.plc, O.3) as assigned by OpenBabel version 2.3.2.[1] Label the pairs as hydrogen bonded interaction if *(i)* the

donor atoms has at least one covalently bonded hydrogen atom, *(ii)* the angle between any identified donor – hydrogen – acceptor pair does not deviate more than 50˚ from its ideal in-plane (180˚) orientation,[2] and *(iii)* the angle between the acceptor heavy atom neighbor – acceptor – hydrogen pair does not deviate more than 90˚ from its ideal in-plane (180˚) orientation. The positions of hydrogen atoms not part of trigonal planar donors are locally optimized. Interactions identified as hydrogen bonds were labeled as acceptor-donor (hb-ad), donor-acceptor (hb-da) or both (hb).

Water mediated hydrogen-bonded contacts: Pre-select all water oxygen atoms within a range[3] of 0.25 to 0.4 nm of any ligand or protein atom. For each water molecule, identify possible ligand donor – water – protein acceptor and ligand acceptor – water – protein donor pairs according to their SYBYL atom types (see *Hydrogen-bonded contacts*). Label a pair as water-mediated hydrogen bond if *(i)* the donor atom has at least one covalently bonded hydrogen atom, *(ii)* the Θ angle (water oxygen – donor hydrogen – donor) is larger than 100˚,[3] and *(iii)* the ω angle (acceptor – water oxygen – donor hydrogen) is within the range 75˚-140˚.[3] A water molecule is only allowed to participate as donor with a maximum of two hydrogen bonds. In case of more than two possible contacts, the ones with a Θ closest to 110˚ and/or smallest hydrogen bond distances are kept. Interactions identified as water mediated hydrogen bonds were labeled 'wb-ad' (acceptor-donor), 'wb-da' (donor-acceptor), or both (wb).

Halogen-bonded contacts (xb): Pre-select all heavy atom contacts within 0.41 nm between a proximal nitrogen, oxygen or sulfur atom and a halogen atom (I,Br,Cl,F) not involving water molecules. Identify acceptor and donor groups and label as halogen bond if *(i)* the halogen has only one covalent carbon as neighbor (donor), *(ii)* the angle between the donor-halogen-acceptor pair does not deviate more than 30˚ around the optimal halogen donor angle of 165˚, and *(iii)* the angle between the

acceptor-halogen-donor pair does not deviate more than 30° around the optimal halogen acceptor angle of 120°.[4]

Salt bridging contacts: Pre-select all heavy atom contacts within 0.55 nm (value from reference 5 + 0.15 nm) between amino acid and ligand positive and negative charge centers defined by their SYBYL atom types as positive charges on the side-chain nitrogen atoms of arginine, lysine and histidine (SYBYL: N.pl3, N.4, N.ar), negative charges on the carboxyl groups in aspartic acid and glutamic acid (SYBYL: O.co2), positively charged ligand centers (SYBYL: N.4, N.am, S.3, C.cat), and negatively charged ligand centers (SYBYL: O.co2, S.3, S.O2). A physiologically relevant pH is assumed (pH ~7.4). Interactions identified as salt bridging were labeled as negative ligand to positive amino-acid (sb-np) or positive ligand to negative amino-acid (sb-pn). Salt bridging contacts have precedence over hydrogen-bonded contacts if they involve the same pair of heavy atoms.

Cation-π interaction (pc): Pre-select all heavy atom contacts within 0.7 nm between ligand positive charge centers (see *salt bridging contacts*) and aromatic amino-acids. A contact is labeled as cation-π interaction if the distance between the cation and the amino-acid ring center is smaller than 0.6 nm,[6] and the offset between the ring center and the cation projected onto the ring plane is less than 0.2 nm. If the cation involves a tertiary or quaternary amine, the angle between the ring center, the covalent heavy atom neighbor of the amine closest to the ring center, and the amide nitrogen should be larger than 90° as filter for cation-π interactions going "through" the ligand.

Aromatic stacking: contacts between aromatic rings are labeled as π- or T-stacking interactions (ps, ts respectively) if *(i)* the center-of-mass between the two rings is within 0.75 nm,[7] *(ii)* the offset distance between the center-of-mass of one ring projected onto the other is no more than 0.2 nm, and *(iii)* the angle between the

normals of both rings does not deviate more than 30˚ from the ideal 180˚ for π-stacking and 90˚ for T-stacking. π-stacking was evaluated between the aromatic amino-acid (F,H,W,Y) rings and ligand aromatic rings resolved using a DFS-based cycle detection algorithm to find the set of smallest unique rings in the structure, with additional check on planarity and aromaticity.

Hydrophobic interactions (hf): contacts are labeled hydrophobic if both atoms involved are carbons within a distance of 0.4 nm and they only have carbon or hydrogen atoms as covalent neighbors. If the hydrophobic interactions are part of a ring system involved in π- or T-stacking interactions they are removed. If one ligand atom has a hydrophobic interaction with multiple atoms of the same amino-acid, the one with the smallest distance is retained.

Heme-coordination (hc): ligand aromatic nitrogen atoms were labeled as possible heme iron coordinating if *(i)* the distance between the nitrogen atom and the heme iron was below 0.35 nm, *(ii)* the distance between the nitrogen projected onto the plane of the heme is less than 0.1 nm, and *(iii)* the normal angle between the aromatic ring structure and the heme plane is between 90˚ +/- 45˚ (cone fit).[8]

Figure S1. Time series of electrostatic ($V_{lig-surr}^{el}$, in violet) and van der Waals interaction energies ($V_{lig-surr}^{vdW}$, in blue) between a ligand and its surrounding during MD for a random and representative selection of six simulations, as indicated per simulation by the ID of the ligand and (starting) pose. Black dashed and solid lines in the lower panel are fitted splines to the time series of $V_{lig-surr}^{el}$ and $V_{lig-surr}^{vdW}$, respectively, after FFT filtering. In the upper panels, gradients (grad) of the fitted splines are presented (in kJ mol^{-1} ps^{-1}).

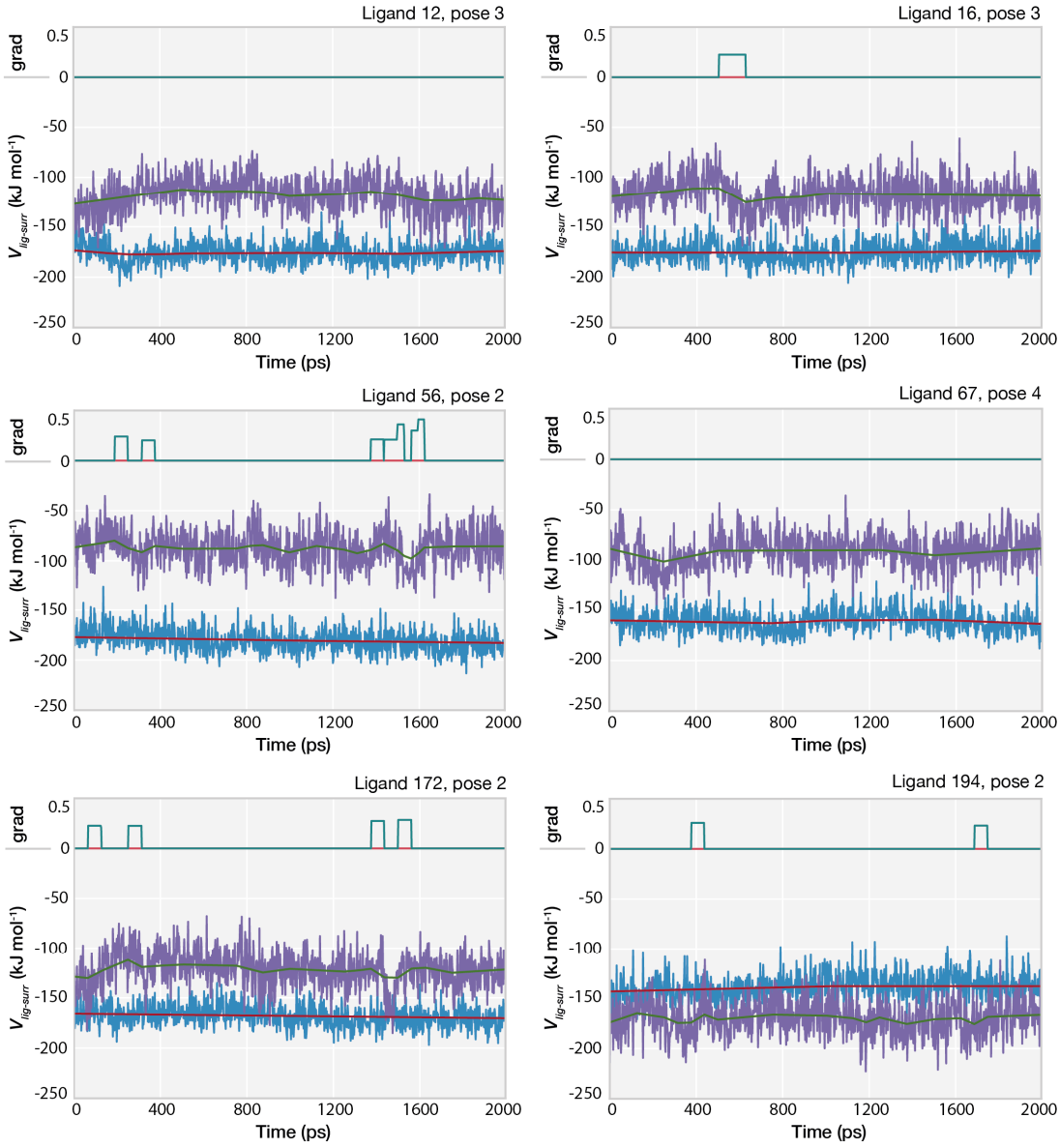


Figure S2. Multivariate normal distribution analysis of ΔV^{vdW} and ΔV^{el} energy values. The red ellipsoid marks the 97.5% confidence limit of the fitted distribution. Ligand simulations located outside the confidence limit are colored red and labeled with their ligand ID followed by the simulation ID. Corresponding fitted distribution for both energy values are displayed along the axis.

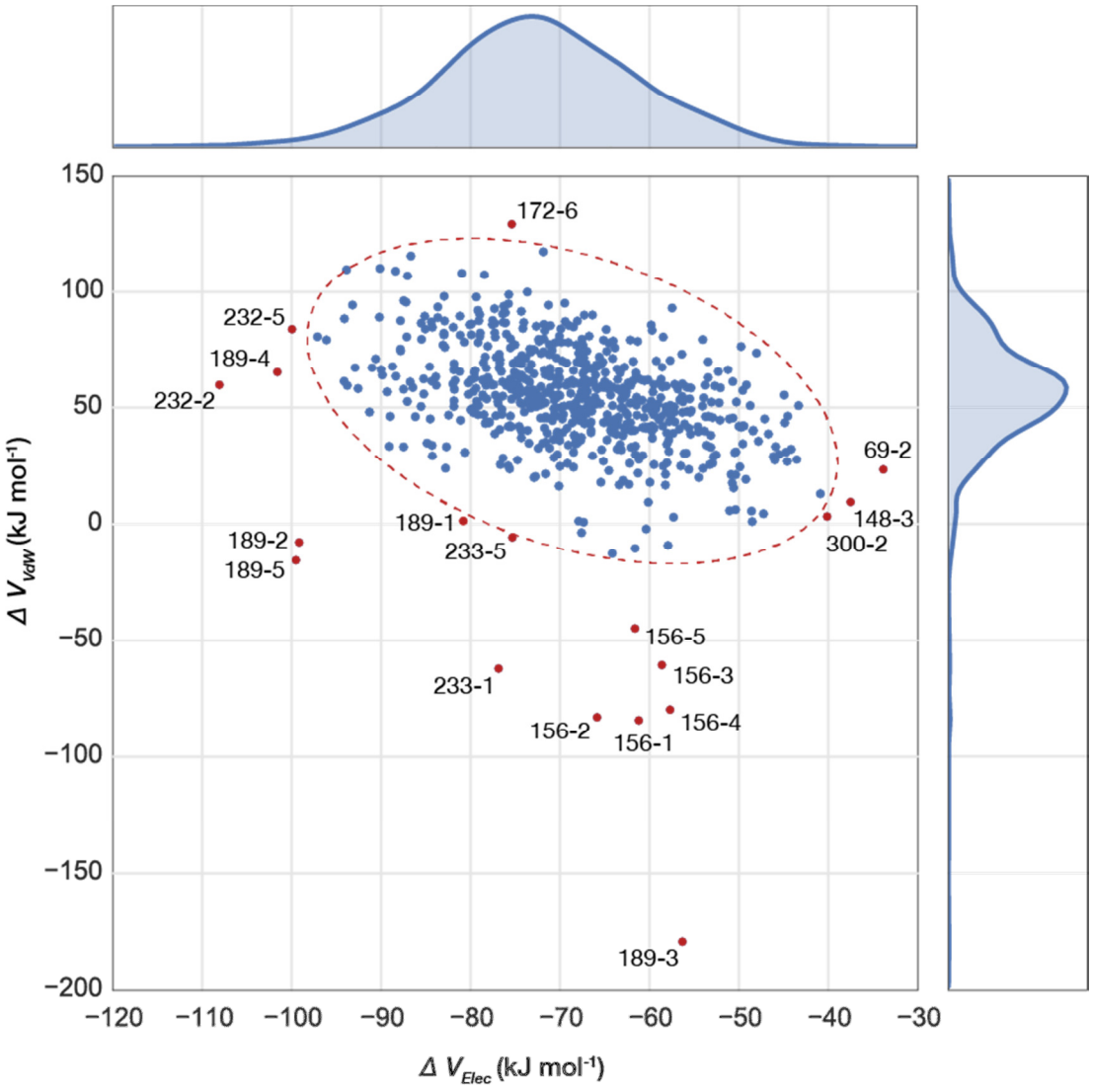


Figure S3. Bootstrap cross-validation results for the combined model 2 and 3 illustrating a unified model with RMSE >= 3.4 kJ.mol^{-1}, and a separation into two individual models with RMSE values lower than 3.4 kJ.mol^{-1} (idealized distributions are shown as gray lines). Bootstrap cross-validation was performed 24000 times with random samples ranging in size from 14 to 84 cases, with steps of 2 cases. Shades of blue color in the matrix plot represent the number of times the particular ligand (x-axis) occurred in all of the bootstrap models within an RMSE range (y-axis) that were accepted according to the statistical acceptance criteria (see Methods section). Additional mixing of the ligands in the accepted models with an RMSE lower than 3.4 kJ.mol^{-1} was performed to validate that the segregation between models 2 and 3 was not due to an absence of the ligands in the overlap region of the two models.

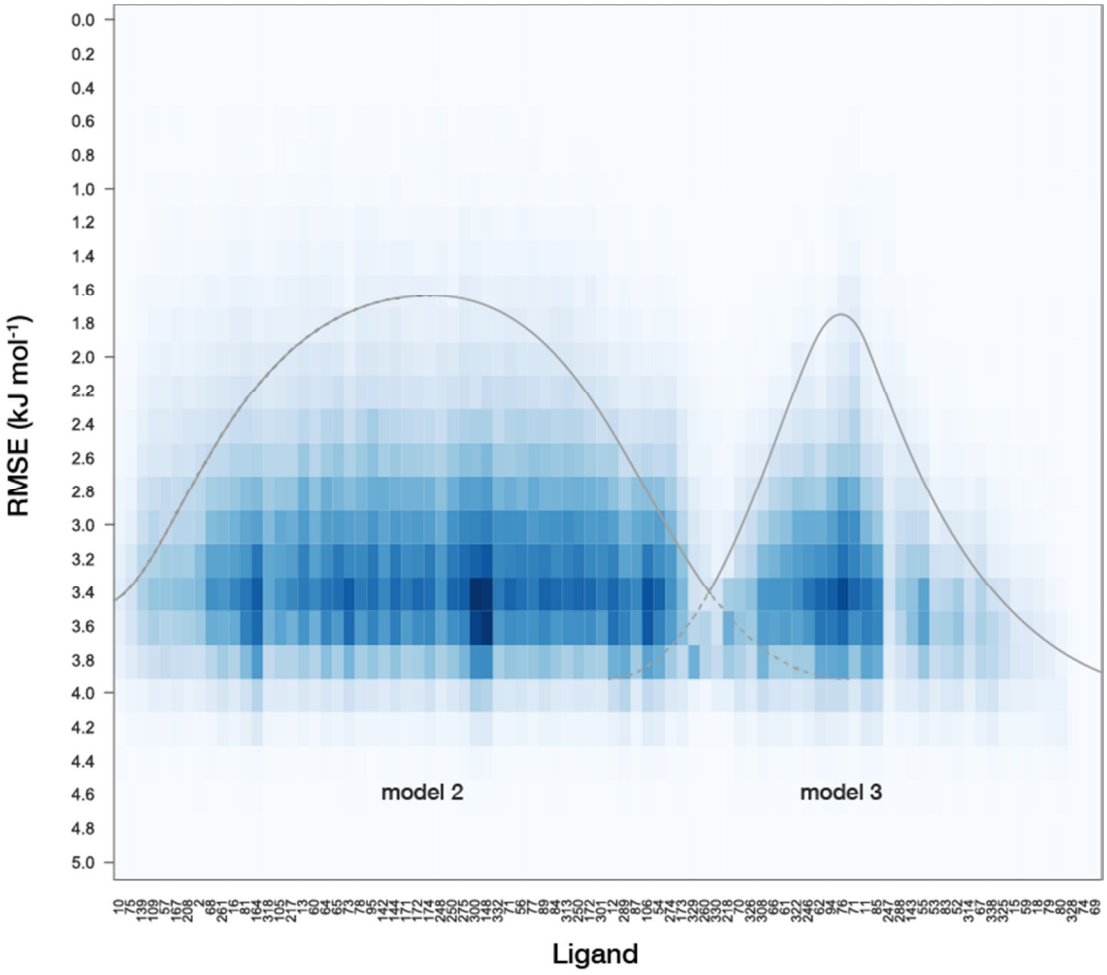


Figure S4. Absolute ΔG_{pred} error values (residuals) for each compound, when compared to the experimentally obtained ΔG_{obs} when calculated using model 1 (blue), model 2 (green) or model 3 (purple) derived from the stochastic approximate inference. Values for the residuals are represented as vertically stacked bars.

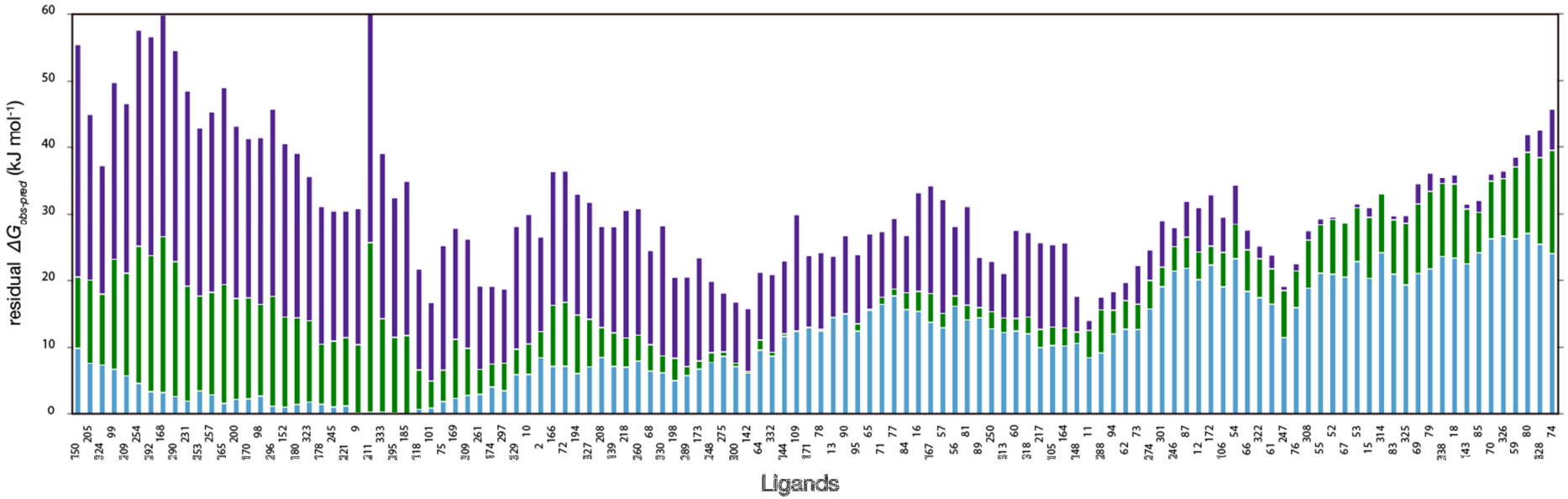


Figure S5. Stick representation (gray) of the dominant binding pose (pose with the highest weight W_i in the LIE binding affinity model, eq. 2) of 10 steroidal aromatase inhibitors explained by LIE binding affinity prediction model 1. The displayed poses for these 10 compounds were derived from representative snapshots of the molecular dynamics simulation and illustrate the similarity in binding pose when compared to the natural substrate of Cytochrome P450 19A1 (PDB code 3EQM, cartoon representation), 4-androstene-3-17-dione (ASD, cyan stick representation). Protein residues (including the heme group, HEME) involved most frequently in polar protein-ligand interactions in more than 50% of the simulation time are displayed in stick representation.

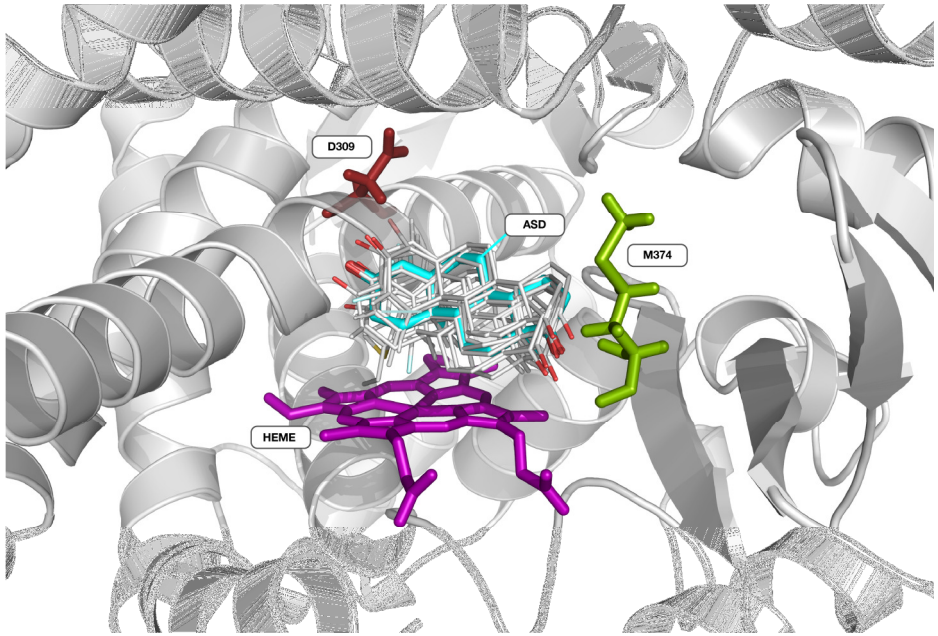


Figure S6. Molecular structures of a selection of compounds (with ligand ID) from final LIE models 1, 2 and 3.

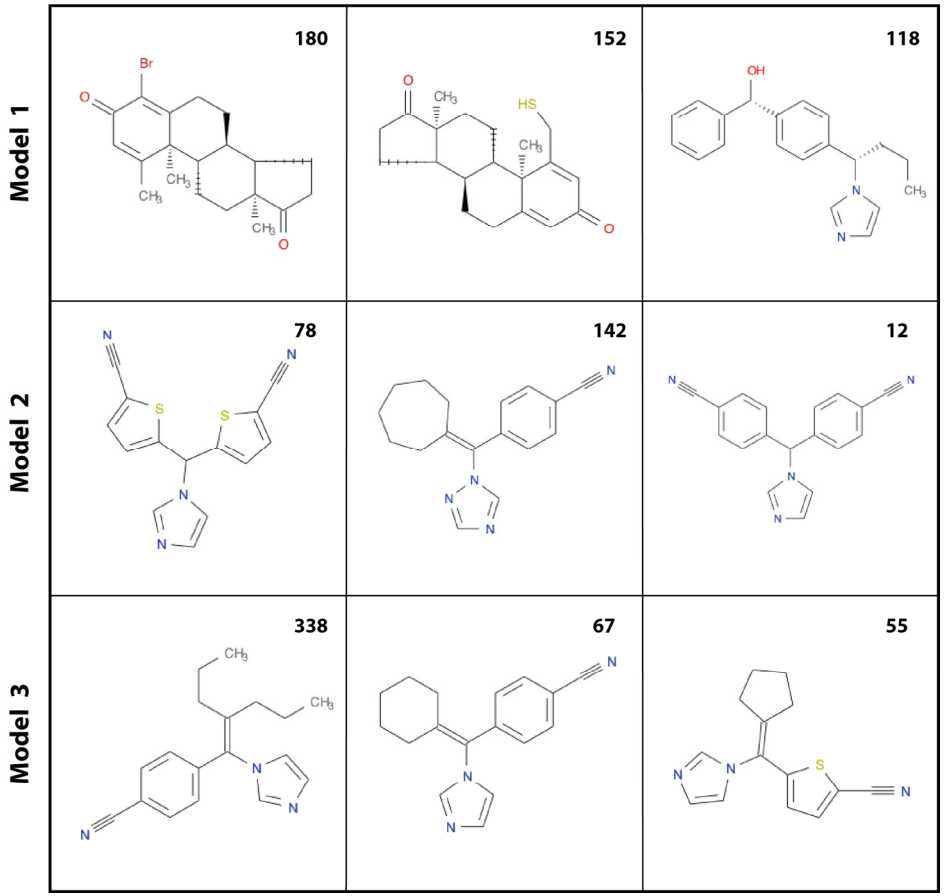
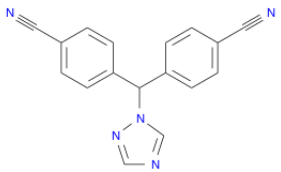
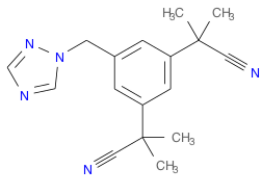
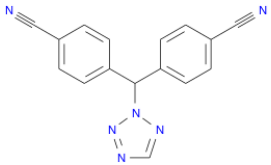
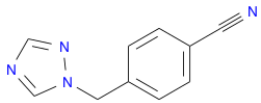
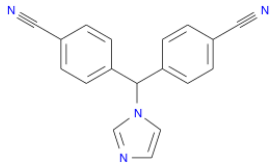
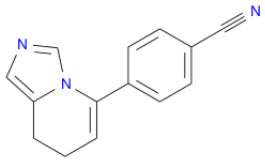
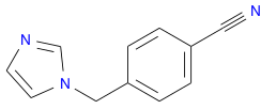
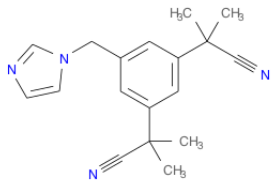
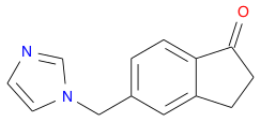
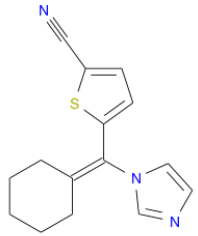


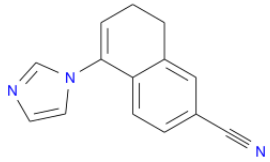
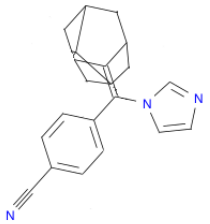
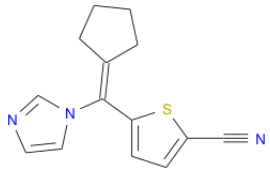
Table S1: Dataset of 132 putative aromatase inhibitors. Listed are the ligand ID used throughout the study; 2D structure depiction (generated using Open Babel version 2.3.2); molecular mass (MW); experimentally determined inhibition constants (pK_i) provided by *Bayer Pharma AG*, and corresponding experimental Gibbs free energy values calculated as $\Delta G_{\mathrm{obs}} = -RT \ln(10^{pKi})$ at a temperature T = 305 K; the predicted binding affinity (ΔG_{pred}); $W\Delta V_{\mathrm{VdW}}$ and $W\Delta V_{\mathrm{Elec}}$ are the weighted and MD averaged differences in the Van der Waals and Coulomb components of nonbonded interaction energies after applying Boltzmann weighting to the selected poses; model ID used for the prediction; and the number of simulations (Nr. poses) used for the prediction. The 6 duplicate compounds are labeled as such in the structure column. The numerical data in this table, together with the canonical SMILES string, are also made available as spreadsheet (vandijk_CYP19_supp_S4.xlsx) as part of the supporting information.

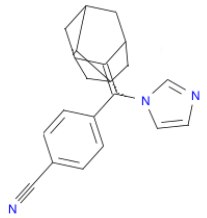
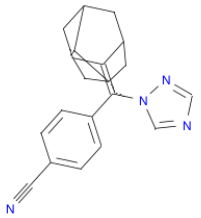
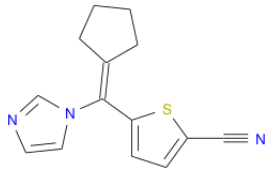
ID	Structure	MW (u)	pK_i	ΔG_{obs} (kJ.mol^{-1})	ΔG_{pred} (kJ.mol^{-1})	$W\Delta V_{\mathrm{VdW}}$ (kJ.mol^{-1})	$W\Delta V_{\mathrm{Elec}}$ (kJ.mol^{-1})	Model	Nr. poses
2	 Letrozole	285.30	8.19	-47.82	-51.54	46.37	-75.54	2	1

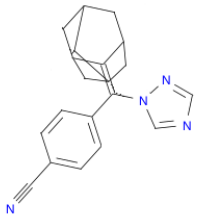
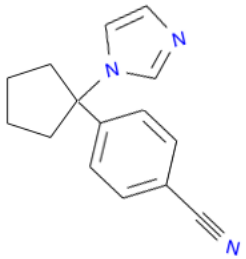
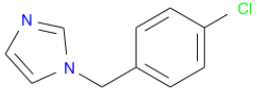
9	 Anastrozole	293.37	7.68	-44.84	-44.95	28.33	-76.37	1	1
10		286.29	7.55	-44.09	-48.22	48.48	-71.53	2	2
11		184.20	6.59	-38.48	-40.18	63.14	-52.95	3	2

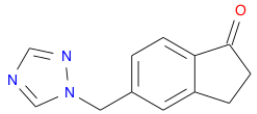
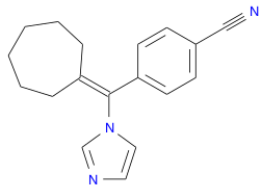
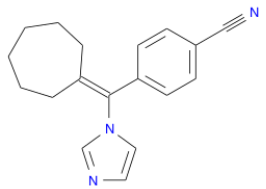
12		284.31	9.92	-57.92	-54.15	66.12	-81.86	2	1
13	 Fadrozole	221.26	7.96	-46.48	-46.27	59.97	-70.44	2	1
15		183.21	8.15	-47.59	-46.31	44.75	-58.92	3	1

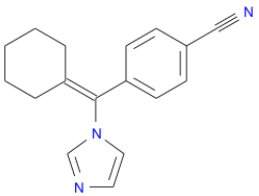
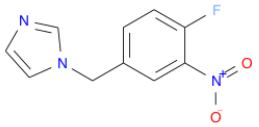
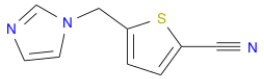
16		292.38	8.66	-50.57	-53.81	87.02	-84.39	2	1
18		212.25	8.82	-51.50	-50.20	77.10	-67.43	3	1
52		269.37	9.62	-56.17	-57.82	59.06	-69.25	3	2

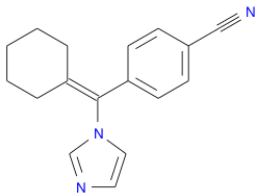
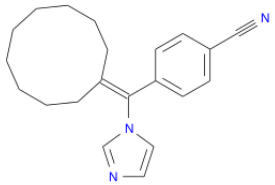
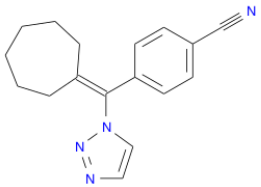
53		221.36	8.72	-50.92	-51.25	51.68	-65.48	3	1
54		315.41	9.92	-57.92	-53.30	81.62	-82.94	2	1
55		255.34	9.60	-56.06	-58.27	53.41	-73.20	3	2

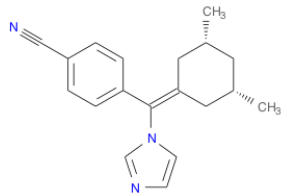
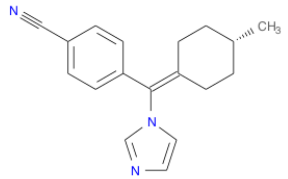
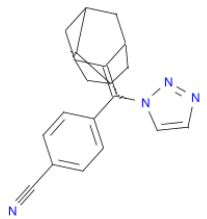
56		315.41	9.27	-54.13	-53.75	69.14	-81.05	2	2
	Duplicate of 54								
57		316.40	8.47	-49.46	-52.46	79.57	-80.10	2	1
59		255.34	9.43	-55.06	-54.34	69.74	-70.67	3	2
	Duplicate of 55								

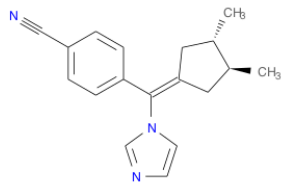
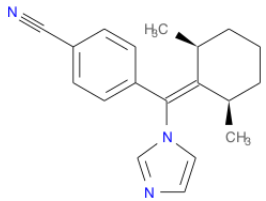
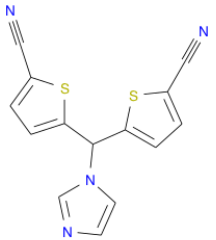
60		316.40	8.37	-48.87	-51.51	58.80	-76.48	2	1
	Duplicate of 57								
61		237.30	7.85	-45.84	-48.27	39.88	-59.67	3	2
62		192.65	6.85	-40.00	-42.42	29.67	-52.56	3	1

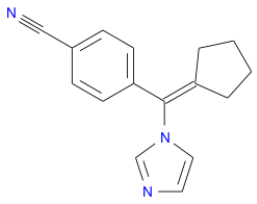
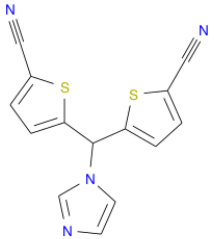
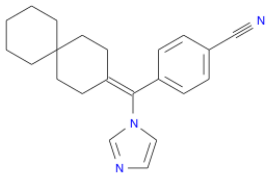
64		213.24	7.28	-42.51	-43.85	42.03	-64.64	2	2
65		277.36	8.68	-50.68	-51.14	89.49	-81.17	2	1
66	 Duplicate of 65	277.36	9.28	-54.19	-57.17	58.24	-73.11	3	2

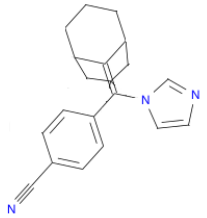
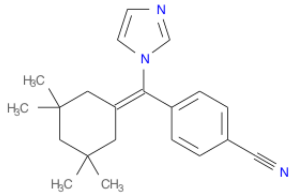
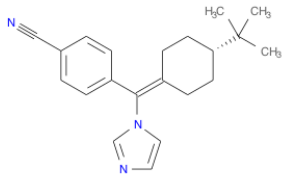
67		263.34	9.32	-54.42	-54.93	62.10	-70.32	3	2
68		221.19	7.57	-44.20	-48.17	44.60	-70.51	2	2
69		189.24	7.38	-43.09	-40.76	61.76	-54.67	3	2

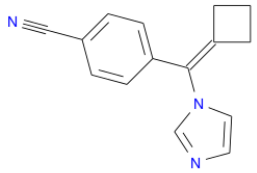
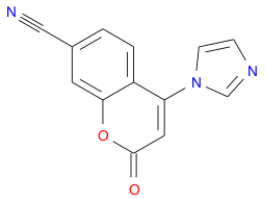
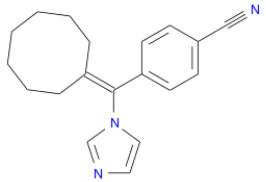
70		263.34	9.72	-56.76	-58.20	80.30	-76.66	3	2
	Duplicate of 67								
71		319.44	9.23	-53.90	-52.91	86.62	-83.24	2	2
72		278.35	6.82	-39.82	-33.63	75.53	-75.54	1	2

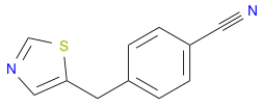
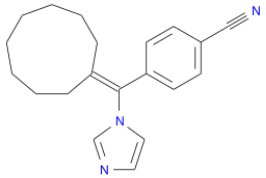
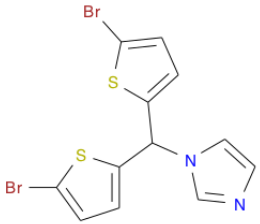
73		291.39	9.54	-55.71	-54.00	27.18	-72.68	2	2
74		277.36	9.92	-57.92	-52.37	56.64	-67.39	3	1
75		316.40	7.40	-43.21	-49.16	71.51	-73.51	2	2

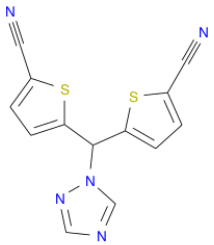
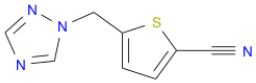
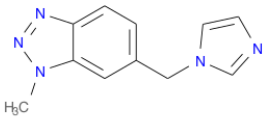
76		277.36	9.92	-57.92	-60.01	61.98	-75.57	3	1
77		291.39	8.57	-50.04	-50.23	82.23	-78.92	2	1
78		296.37	9.20	-53.72	-54.57	49.06	-79.97	2	1

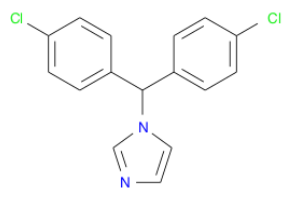
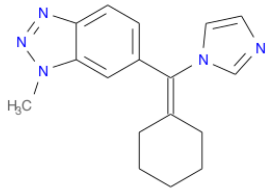
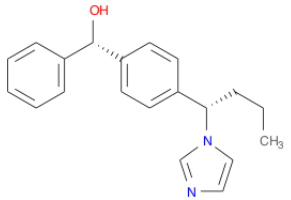
79		249.31	9.17	-53.55	-52.17	43.94	-64.38	3	3
80	 Duplicate of 78	296.37	10.73	-62.65	-60.96	59.70	-77.66	3	1
81		331.45	10.01	-58.45	-60.89	67.10	-91.00	2	1

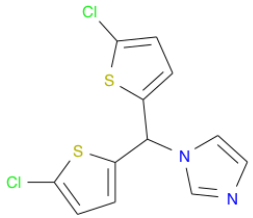
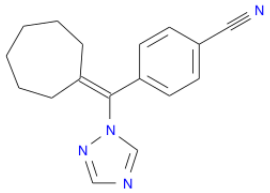
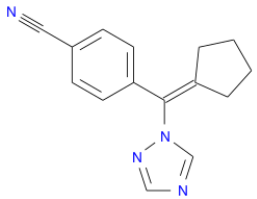
83		303.40	9.42	-55.00	-56.01	67.42	-72.24	3	2
84		319.44	9.14	-53.37	-51.45	64.71	-77.72	2	3
85		319.44	10.07	-58.80	-56.94	59.73	-73.36	3	2

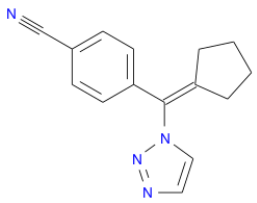
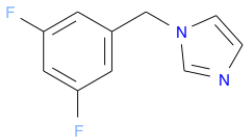
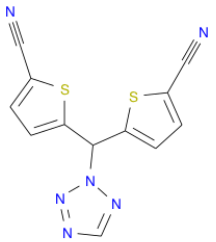
87		235.28	8.92	-52.09	-56.95	79.61	-74.86	3	1
89		237.21	7.36	-42.98	-42.05	57.71	-64.48	2	1
90		291.39	9.57	-55.88	-56.28	62.04	-84.13	2	2

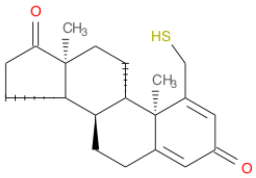
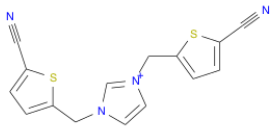
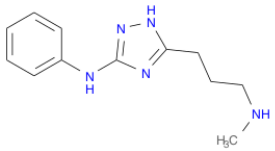
94		200.26	6.92	-40.41	-43.49	35.02	-51.57	3	2
95		305.42	9.16	-53.49	-53.56	49.67	-78.03	2	2
98		404.14	8.15	-47.59	-50.07	29.23	-84.27	1	1

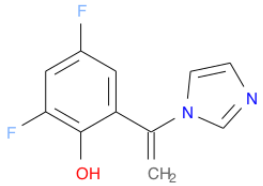
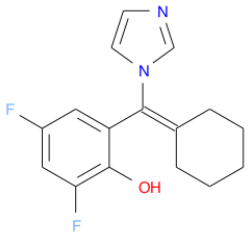
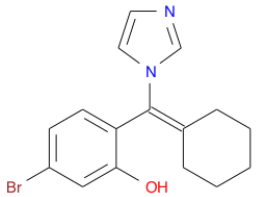
99		297.36	6.52	-38.07	-44.05	35.96	-68.23	1	1
101		190.23	5.82	-33.98	-33.12	5.73	-51.04	1	1
105		213.24	7.92	-46.25	-48.70	50.88	-72.39	2	1

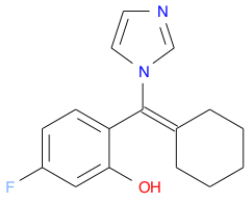
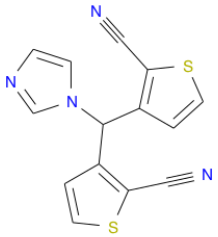
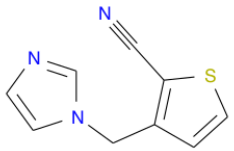
106		303.19	10.00	-58.39	-53.70	52.67	-79.32	2	1
109		293.37	8.77	-51.21	-51.11	61.60	-78.24	2	1
118		306.40	7.85	-45.84	-46.48	-3.68	-67.59	1	1

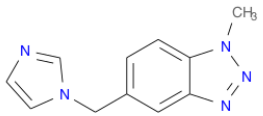
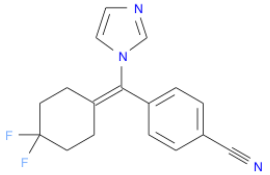
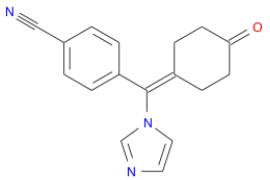
139		315.24	8.39	-48.99	-53.86	40.17	-66.52	2	1
142		278.35	8.70	-50.80	-50.50	56.25	-75.68	2	2
143		250.30	8.49	-49.57	-50.30	73.36	-66.96	3	2

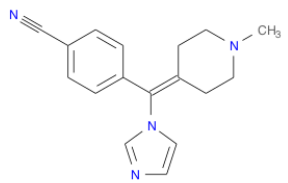
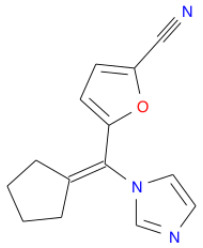
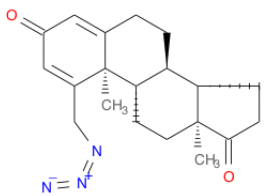
144		250.30	7.46	-43.56	-44.99	77.27	-71.20	2	1
148		194.18	6.46	-37.72	-36.22	32.98	-53.13	2	1
150		298.35	6.60	-38.54	Outlier	Outlier	Outlier	Outlier	Outlier

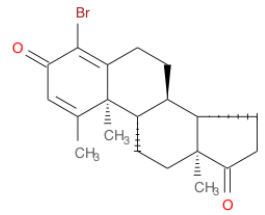
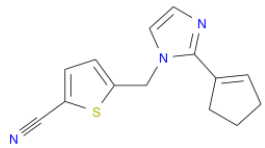
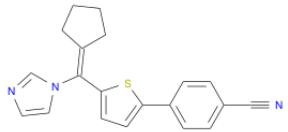
152		330.48	8.85	-51.68	-52.65	59.39	-93.78	1	1
156		311.41	6.14	-35.85	Outlier	Outlier	Outlier	Outlier	Outlier
159		231.30	4.36	-25.46	Outlier	Outlier	Outlier	Outlier	Outlier

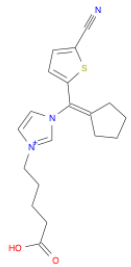
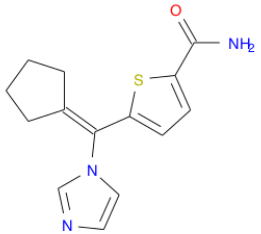
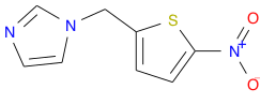
164		222.19	6.54	-38.19	-41.55	70.87	-64.10	2	1
165		290.31	7.43	-43.38	-44.88	63.34	-88.35	1	1
166		333.22	7.80	-45.55	Outlier	Outlier	Outlier	Outlier	Outlier

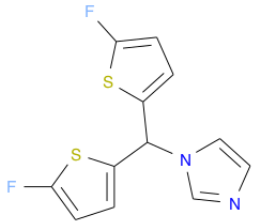
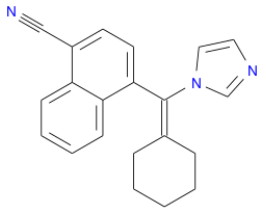
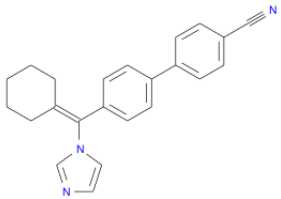
167		272.32	8.82	-51.50	-55.52	92.71	-87.47	2	2
168		296.37	5.77	-33.69	-36.61	92.49	-86.17	1	2
169		189.24	5.32	-31.06	-28.91	47.56	-58.37	1	1

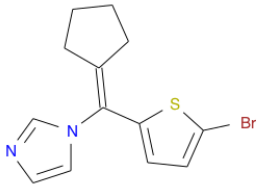
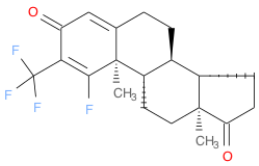
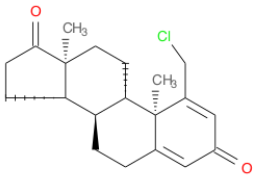
170		213.34	5.16	-30.13	-32.23	52.72	-65.94	1	1
171		299.32	8.92	-52.09	-52.49	53.65	-78.18	2	2
172		277.32	8.85	-51.68	-49.17	94.49	-79.26	2	2

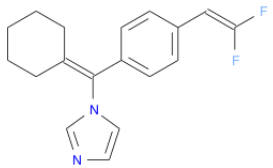
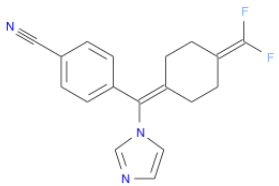
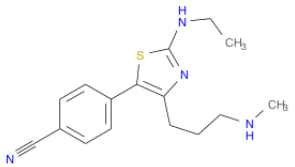
173		278.35	7.02	-40.99	-41.51	53.09	-60.81	2	1
174		239.27	7.15	-41.75	-44.78	23.31	-62.74	2	2
178		339.43	8.85	-51.68	-50.32	23.85	-82.79	1	1

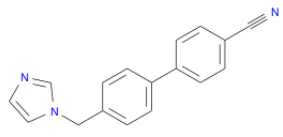
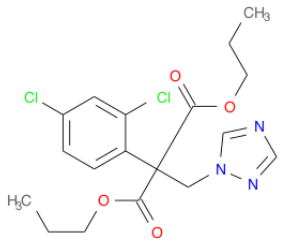
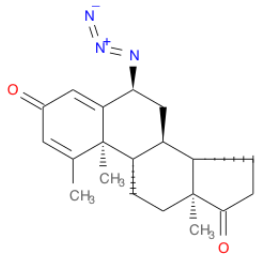
180		377.32	7.11	-41.52	-39.95	44.36	-74.84	1	3
182		255.34	3.85	-22.48	-24.93	54.71	-56.33	1	1
185		331.43	8.42	-49.17	-49.17	37.48	-84.77	1	1

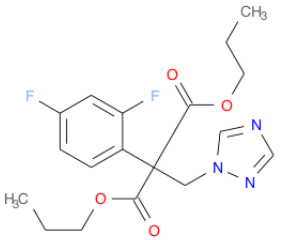
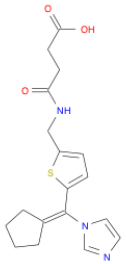
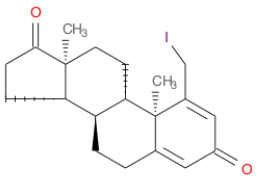
189		356.46	5.68	-33.17	Outlier	Outlier	Outlier	Outlier	Outlier
194		273.35	6.02	-35.15	-29.74	66.43	-66.99	1	1
198		209.23	7.32	-42.74	-38.23	22.83	-64.52	1	1

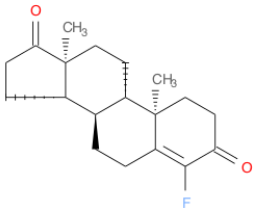
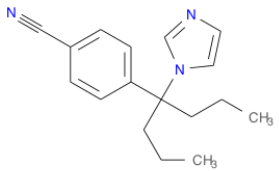
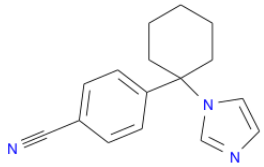
200		282.30	7.25	-42.33	-44.36	44.56	-81.10	1	1
205		313.40	8.52	-49.75	-42.64	88.38	-94.10	1	2
208		339.43	8.60	-50.22	-54.34	47.44	-79.44	2	2

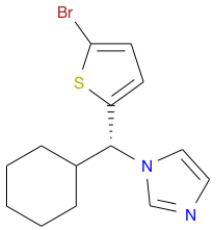
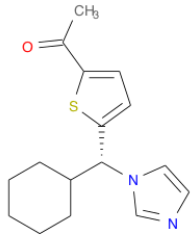
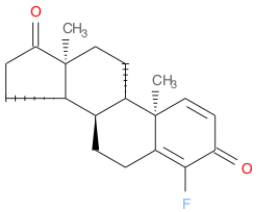
209		309.23	6.89	-40.23	-45.37	51.32	-72.04	1	1
210		370.38	5.38	-31.41	Outlier	Outlier	Outlier	Outlier	Outlier
211		332.86	6.64	-38.77	-39.16	59.39	-78.27	1	1

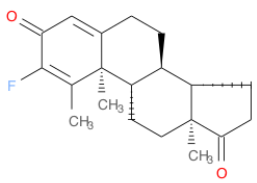
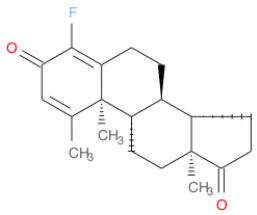
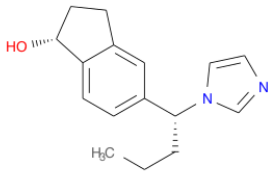
217		300.35	7.85	-45.84	-48.84	54.35	-66.08	2	1
218		311.33	8.82	-51.50	-55.40	56.25	-71.59	2	1
220		300.42	4.24	-24.76	Outlier	Outlier	Outlier	Outlier	Outlier

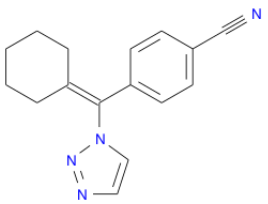
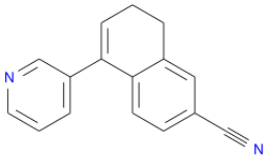
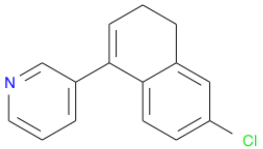
221		259.31	6.18	-36.09	-34.93	43.15	-66.64	1	1
223		414.28	5.44	-31.76	Outlier	Outlier	Outlier	Outlier	Outlier
231		339.43	7.31	-42.68	-44.45	64.26	-87.24	1	2

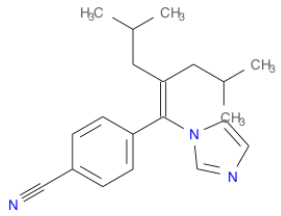
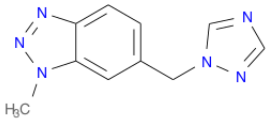
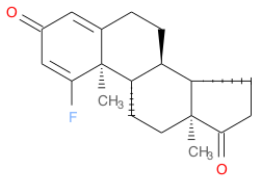
232		381.37	5.00	-29.20	Outlier	Outlier	Outlier	Outlier	Outlier
233		359.44	4.35	-25.40	Outlier	Outlier	Outlier	Outlier	Outlier
235		424.32	6.60	-37.99	-39.22	23.45	-74.23	1	1

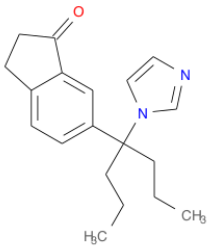
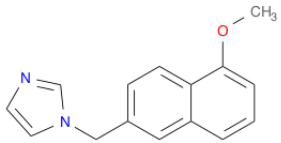
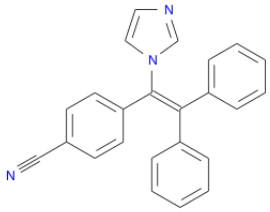
245		304.40	7.46	-43.56	-44.57	19.87	-72.90	1	1
246		267.34	9.20	-53.72	-57.81	90.97	-76.10	3	1
247		251.33	7.85	-45.84	-46.21	48.03	-58.18	3	1

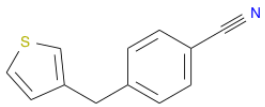
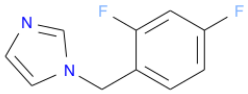
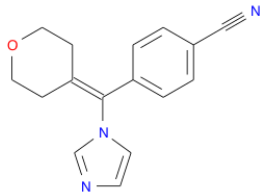
248		325.27	8.30	-48.46	-49.34	51.10	-73.92	2	2
250		288.41	8.89	-51.91	-50.08	34.74	-71.92	2	1
253		302.38	7.04	-41.11	-37.66	57.48	-76.12	1	2

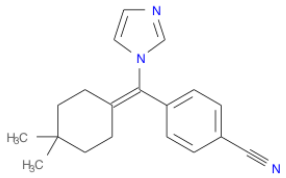
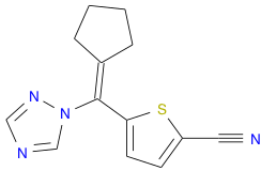
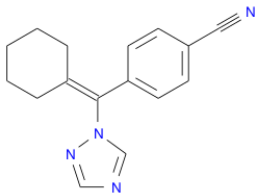
254		316.41	7.24	-42.28	-47.08	57.10	-88.50	1	1
257		316.41	7.51	-43.85	-41.76	62.87	-82.81	1	1
260		256.34	8.60	-50.22	-53.10	36.00	-67.68	2	1

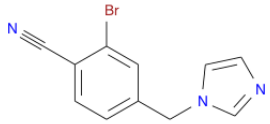
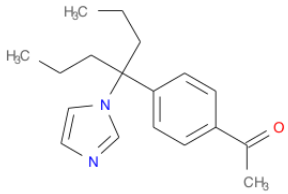
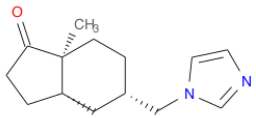
261		264.33	7.52	-43.91	-46.28	107.45	-78.46	2	2
274		232.28	8.37	-48.87	-45.07	43.82	-66.52	2	1
275		241.72	7.23	-42.22	-42.81	27.70	-61.19	2	1

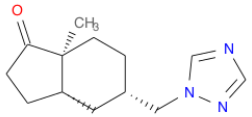
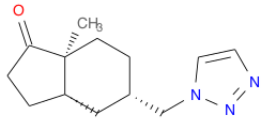
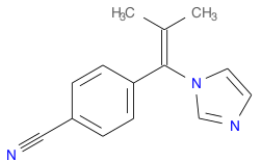
288		307.43	9.80	-57.22	-56.23	75.32	-73.36	3	1
289		214.23	6.51	-38.01	-40.58	70.12	-59.23	2	1
290		302.38	6.92	-40.41	-43.98	34.40	-68.23	1	1

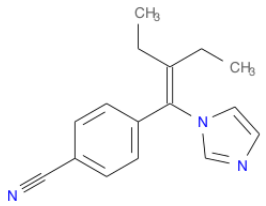
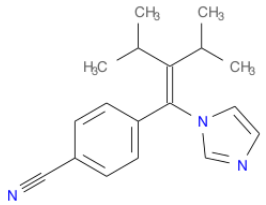
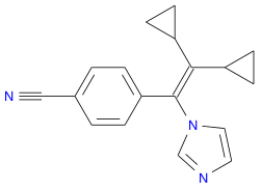
292		296.41	7.80	-45.55	-49.28	61.20	-94.13	1	1
295		238.28	6.68	-39.01	-39.11	39.01	-71.41	1	1
296		347.41	7.62	-44.49	-45.89	57.79	-87.48	1	1

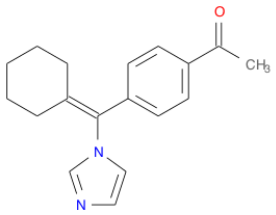
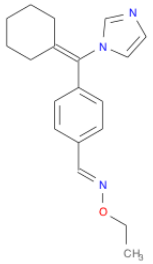
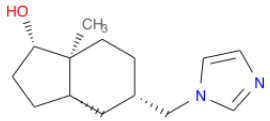
297		199.27	5.60	-32.70	-29.55	23.73	-51.96	1	2
300		194.18	6.36	-37.14	-38.72	30.35	-56.10	2	1
301		265.31	8.66	-50.57	-47.88	77.36	-75.11	2	2

308		291.39	9.89	-57.75	-59.54	58.02	-76.17	3	2
309		256.33	7.37	-43.03	-40.51	29.81	-70.30	1	1
313		264.33	8.21	-47.94	-45.59	37.22	-67.59	2	4

314		262.11	8.44	-49.28	-49.17	71.91	-65.60	3	1
318		284.40	8.13	-47.47	-49.56	68.71	-76.45	2	4
322		232.32	8.39	-48.99	-52.52	38.67	-63.20	3	2

323		233.31	6.42	-37.49	-35.83	45.38	-68.75	1	2
324		233.31	6.80	-39.71	-46.26	-64.14	-12.72	1	1
325		233.37	8.22	-48.00	-47.62	53.14	-60.80	3	2

326		251.33	9.15	-53.43	-55.06	91.24	-74.20	3	2
327		279.38	7.96	-46.48	-52.89	71.33	-73.83	2	1
328		275.35	9.64	-56.29	-52.18	50.10	-66.79	3	1

329		280.36	8.57	-50.04	-53.66	55.45	-70.18	2	1
330		309.41	8.10	-47.30	-49.89	29.42	-60.75	2	1
332		234.34	8.00	-46.71	-47.40	55.04	-70.01	2	1

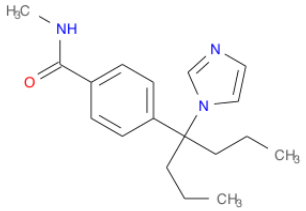
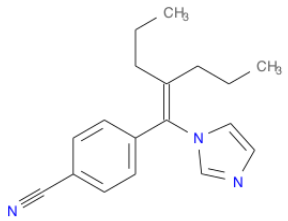
333		299.41	7.41	-43.27	-43.49	50.48	-81.86	1	1
338		279.38	10.00	-58.39	-57.34	58.87	-73.64	3	2

Table S2: Averaged difference in the Coulomb component of the nonbonded protein-ligand interaction energy (in kJ.mol^{-1}, after FFT trajectory filtering) for up to 8 selected docking poses for the dataset of 132 putative aromatase inhibitors represented by the ligand ID (Table S1). The docking poses selected by the Boltzmann weighting scheme during training of the respective model the ligand belongs to (Table S1) are shown as gray shaded cells.

ID	1	2	3	4	5	6	7	8
2	-66.244	-67.544	-59.492	-75.539				
9	-75.989	-78.494	-73.475	-75.537	-76.368			
10	-71.896	-70.730	-75.100	-78.579	-70.148			
11	-52.207	-47.238	-54.027	-49.752	-43.326	-48.410		
12	-82.942	-74.142	-81.858	-67.144				
13	-61.514	-57.458	-51.064	-49.564	-70.439			
15	-48.021	-53.764	-52.703	-51.222	-58.923			
16	-67.001	-71.694	-84.388	-79.900	-80.963			
18	-59.701	-60.972	-58.727	-57.146	-67.425			
52	-71.152	-59.268	-74.585	-60.353	-71.082	-66.530		
53	-44.541	-61.814	-48.386	-54.466	-65.478			
54	-68.319	-64.338	-65.711	-69.410	-82.935			
55	-68.407	-73.584	-55.445	-72.582	-70.115			
56	-66.413	-81.208	-56.599	-80.679	-75.435			
57	-73.390	-85.065	-69.221	-81.148	-71.405			
59	-60.356	-61.323	-68.246	-71.838	-63.153	-59.601		
60	-73.557	-65.200	-72.911	-74.612	-77.401			
61	-55.814	-44.940	-50.688	-60.770	-57.823			
62	-50.000	-52.559	-44.600	-45.520	-46.690			
64	-64.944	-63.543	-64.335	-62.985	-56.300			

65	-59.501	-74.585	-70.063	-81.174	-67.731		
66	-74.496	-71.461	-73.092	-64.493	-57.818		
67	-67.894	-69.610	-61.488	-70.317	-63.534	-58.055	
68	-65.932	-71.101	-70.354	-66.994	-66.616	-67.261	
69	-49.146	-33.851	-56.363	-49.565	-52.107	-47.804	
70	-67.309	-77.293	-62.050	-58.161	-64.826	-64.177	
71	-82.973	-69.724	-63.684	-83.648	-79.525		
72	-74.745	-59.691	-58.032	-62.618	-77.288		
73	-68.176	-72.729	-62.492	-72.341			
74	-52.073	-63.602	-61.292	-67.386	-50.464	-65.803	
75	-62.831	-62.953	-73.453	-65.415	-74.260		
76	-88.392	-68.238	-70.119	-77.823	-73.985	-75.569	
77	-72.856	-72.374	-64.039	-76.345	-78.918		
78	-72.326	-79.970	-60.166	-74.739	-65.422		
79	-59.493	-58.197	-57.535	-65.345	-66.606		
80	-76.697	-75.026	-74.271	-77.661			
81	-73.006	-78.347	-91.001	-71.164	-87.814	-81.845	
83	-70.691	-63.274	-67.365	-74.318	-66.362	-59.861	
84	-79.160	-73.315	-77.137	-77.224			
85	-82.984	-69.545	-74.458	-71.792			
87	-62.140	-69.072	-56.311	-58.543	-75.534	-57.208	
89	-57.167	-58.577	-64.762	-64.475	-58.454		
90	-79.580	-57.635	-84.322	-80.253	-71.636	-74.560	-83.877
94	-43.535	-50.353	-53.067	-47.099	-53.272		
95	-72.728	-78.551	-74.482	-77.880	-69.111		
98	-84.267	-79.685	-70.636	-72.413	-67.908		
99	-56.490	-68.787	-64.104	-68.230	-74.743		
101	-58.000	-52.311	-49.419	-52.844	-51.041		
105	-49.887	-67.574	-72.386	-61.441	-58.862		
106	-59.420	-70.842	-79.324	-65.907			

109	-87.133	-77.525	-78.244	-50.587	-70.531	-82.857	
118	-67.429	-62.643	-71.602	-60.128	-67.587		
139	-57.920	-66.520	-78.270	-59.665	-61.567		
142	-75.990	-66.448	-67.112	-67.358	-75.410		
143	-67.843	-64.999	-66.228	-64.586	-64.096		
144	-64.925	-71.196	-63.624	-62.992	-65.125		
148	-46.131	-44.750	-37.505	-53.135			
150	-61.610	-73.179	-66.463	-70.461	-86.169		
152	-76.790	-79.635	-81.960	-77.908	-93.779		
156	-61.193	-65.848	-58.603	-57.687	-61.617		
159	-70.525		-75.222	-67.801	-50.363		
164	-49.219	-68.049	-53.453	-60.410	-58.694		
165	-84.263	-88.355	-79.945	-85.519	-72.645		
166	-61.568	-81.837		-69.073	-71.738		
167	-63.138	-85.476	-90.115	-68.640	-69.274	-53.621	
168	-87.051	-87.206	-67.082	-73.639	-85.118		
169	-58.704	-57.658	-50.277	-40.883			
170	-68.998	-69.469	-55.781	-60.028	-65.944		
171	-78.182	-73.211	-57.779	-66.348	-76.488	-76.408	
172	-79.369	-79.181	-75.758	-74.041	-72.639	-75.385	
173	-60.810	-62.454	-67.717	-63.528	-71.082	-67.122	
174	-62.170	-66.736	-58.120	-62.296	-57.235	-54.578	-55.568
178	-87.075	-89.990	-77.173	-82.785	-80.626		
180	-68.486	-81.734	-76.494	-78.639	-78.365		
182	-62.117	-70.459	-68.484	-57.298	-56.329		
185	-66.924	-67.050	-75.448	-84.861	-76.719	-71.164	-84.503
189	-77.565	-99.151	-56.304	-101.593	-99.511		
194	-71.281	-66.986	-64.541	-71.852			
198	-61.399	-57.348	-59.796	-59.947	-64.522		
200	-76.391	-81.102	-69.171	-71.719	-71.593		

205	-90.179	-63.884	-65.333	-74.851	-78.249	-94.100	-87.836	-79.758
208	-68.174	-70.303	-79.684	-72.061	-67.432	-79.162		
209	-75.787	-73.786	-63.457	-67.924	-64.703	-72.045		
210	-90.592	-63.134	-79.491	-91.284	-72.962			
211	-88.988	-78.270	-76.915	-79.908	-84.299			
217	-63.268	-67.031	-58.211	-72.838	-66.081			
218	-65.973	-75.628	-86.452	-67.883	-66.930	-71.587		
220	-62.936	-81.708	-83.657	-62.891	-76.427			
221	-56.650	-53.921	-66.638	-48.505				
223	-97.112	-92.904	-87.488	-93.191	-80.234			
231	-93.859	-73.673	-86.675	-86.753	-85.223			
232	-96.106	-108.064	-87.471	-79.090	-99.950			
233	-76.861	-82.685	-89.850	-70.678	-75.320	-89.093		
245	-75.595	-54.523	-76.962	-72.895	-72.892			
246	-76.105	-76.297	-66.022	-67.288	-71.110	-78.954		
247	-59.802	-72.123	-81.919	-58.179	-59.463			
248	-70.005	-69.686	-68.480	-73.255	-74.451			
250	-67.528	-73.004	-70.451	-71.917	-71.651			
253	-76.560	-58.500	-68.070	-69.256	-83.073			
254	-62.219	-79.388	-68.926	-89.316	-83.823			
257	-87.846	-53.399	-76.823	-64.111	-73.616			
260	-74.295	-85.242	-67.682	-65.897	-74.315	-64.208		
261	-68.154	-67.123	-78.455	-65.102	-60.365	-72.104		
274	-60.223	-66.516	-55.317	-64.603	-48.580			
275	-61.191	-44.164	-58.458	-55.576	-53.845			
288	-74.580	-62.089	-92.566	-73.365	-88.723			
289	-62.641	-62.761	-63.850	-59.229	-62.556			
290	-63.997	-68.231	-55.386	-85.533	-71.209			
292	-82.889	-94.133	-76.668	-71.577	-80.083			
295	-71.414	-50.748	-59.729	-58.058	-59.758	-54.007		

296	-68.457	-70.897	-84.567	-77.284	-76.726	-87.484	-78.573
297	-50.871	-52.939	-45.875	-53.720	-45.975		
300	-49.558	-40.130	-56.102	-50.335	-50.660	-49.746	
301	-71.797	-71.145	-69.906	-67.788	-74.431	-76.086	
308	-74.933	-70.021	-78.085	-74.330	-72.160		
309	-68.514	-71.125	-70.299	-74.276	-69.826		
313	-70.038	-70.426	-67.592	-68.807	-67.627		
314	-55.484	-48.730	-56.243	-65.595	-53.458	-54.056	
318	-75.767	-78.968	-77.236	-75.343	-74.753	-69.859	
322	-62.045	-64.031	-69.750	-63.169	-65.127		
323	-67.348	-72.539	-68.656	-60.592	-63.591		
324	-64.917	-61.639	-70.656	-64.138	-64.741		
325	-56.824	-55.356	-62.386	-55.878	-44.875	-53.146	
326	-61.438	-62.788	-75.720	-59.702	-67.316	-70.873	
327	-69.833	-61.800	-78.632	-73.826			
328	-59.031	-68.110	-66.793	-56.068	-52.935	-57.677	
329	-70.176	-66.387	-68.800	-82.439	-67.915		
330	-78.063	-60.751	-81.802	-62.964	-67.344	-68.399	
332	-61.992	-68.490	-70.007	-63.367	-77.319		
333	-81.857	-66.597	-67.822	-75.783	-79.074	-65.135	
338	-60.549	-73.602	-71.570	-75.205	-73.693		

Table S3: Averaged difference in the Van der Waals component of the nonbonded protein-ligand interaction energy (in $kJ.mol^{-1}$, after FFT trajectory filtering) for up to 8 selected docking poses for the dataset of 132 putative aromatase inhibitors represented by the ligand ID (Table S1). The docking poses selected by the Boltzmann weighting scheme during training of the respective model the ligand belongs to (Table S1) are shown as gray shaded cells.

ID	1	2	3	4	5	6	7	8
2	60.995	79.554	59.724	46.370				
9	81.778	46.958	66.355	75.820	28.326			
10	45.536	36.321	65.194	61.565	59.612			
11	51.995	4.421	79.381	76.385	50.802	39.988		
12	98.039	84.206	66.124	66.649				
13	64.808	92.983	45.151	57.286	59.974			
15	73.572	59.530	59.223	66.217	44.755			
16	53.681	83.247	87.017	86.150	108.174			
18	42.177	61.965	54.201	43.911	77.098			
52	72.890	49.334	64.050	41.075	85.417	39.322		
53	46.307	69.916	40.309	19.919	51.678			
54	46.501	55.271	67.723	87.178	81.619			
55	38.130	51.825	47.930	55.998	65.464			
56	67.009	65.375	53.227	78.267	57.800			
57	60.828	83.965	41.162	81.298	65.154			
59	41.720	78.590	60.538	74.168	56.739	56.597		
60	53.104	69.692	64.403	56.993	59.691			
61	58.646	55.311	61.065	41.485	36.245			
62	42.025	29.673	26.821	29.899	38.704			
64	43.715	57.874	40.352	49.864	47.913			

65	46.885	58.847	80.919	89.487	49.270		
66	68.958	45.425	75.154	43.491	31.270		
67	73.403	47.127	63.111	62.102	61.970	34.296	
68	68.423	66.889	38.790	74.412	49.101	43.065	
69	59.604	23.323	73.047	42.198	44.616	35.614	
70	56.677	82.303	51.915	50.251	83.744	40.962	
71	82.155	45.148	38.595	93.441	62.603		
72	70.555	48.236	26.214	47.805	86.319		
73	77.069	20.304	43.184	54.082			
74	58.416	66.651	21.772	56.640	30.745	57.965	
75	41.609	57.139	34.744	68.339	72.529		
76	109.104	52.009	16.215	60.649	48.125	61.978	
77	74.960	79.503	78.774	86.856	82.231		
78	78.586	49.058	61.815	55.287	17.850		
79	25.421	21.267	44.523	59.105	52.224		
80	70.686	77.144	51.573	59.704			
81	57.567	43.217	67.104	63.504	74.506	52.845	
83	49.462	64.794	61.451	91.543	89.972	50.010	
84	78.130	73.854	71.740	55.225			
85	81.433	50.583	62.346	85.931			
87	79.192	78.146	60.101	58.711	79.785	50.264	
89	59.870	67.320	74.322	57.716	80.485		
90	64.973	46.229	61.215	60.480	52.049	41.966	63.099
94	27.662	21.047	52.299	45.134	25.420		
95	50.863	69.418	55.960	43.869	60.614		
98	29.228	56.772	47.137	32.604	1.286		
99	24.821	50.976	58.683	35.961	27.454		
101	48.718	65.342	42.113	28.893	5.735		
105	24.456	46.347	50.884	50.997	60.442		
106	50.653	44.411	52.673	51.754			

109	80.546	87.646	61.602	15.467	49.980	76.056	
118	68.990	40.301	21.422	9.326	-3.682		
139	-9.145	40.170	43.825	36.902	29.850		
142	59.350	48.424	57.630	0.817	53.573		
143	74.028	61.089	66.066	58.129	71.195		
144	64.971	77.274	42.343	58.242	79.182		
148	29.032	30.879	9.427	32.984			
150	-10.473	56.061	49.893	54.010	46.095		
152	73.081	4.736	60.019	43.566	59.390		
156	-84.218	-82.892	-60.649	-79.657	-44.751		
159	78.052		69.103	70.609	6.184		
164	19.126	87.377	62.775	49.612	62.000		
165	84.679	63.343	61.778	86.479	50.339		
166	33.536	68.933		33.961	66.154		
167	53.481	79.502	110.248	43.466	59.443	33.874	
168	106.967	95.495	69.045	99.903	89.471		
169	45.969	50.913	52.349	12.950			
170	85.478	95.072	48.334	50.421	52.717		
171	53.655	68.642	44.057	29.275	47.100	90.430	
172	97.757	91.985	94.818	86.757	85.130	129.125	
173	53.091	46.869	44.355	44.344	58.416	61.727	
174	51.816	80.867	59.088	16.967	48.007	41.837	41.152
178	52.565	66.331	46.502	23.849	30.799		
180	24.301	66.610	49.191	61.822	71.891		
182	43.570	62.609	84.843	2.956	54.714		
185	56.825	58.945	41.827	34.820	69.190	32.672	45.129
189	85.878	-7.806	-179.490	65.165	-15.645		
194	94.241	66.432	77.929	117.389			
198	46.737	38.988	52.701	17.849	22.834		
200	54.872	44.565	25.920	41.265	35.198		

205	89.032	44.456	31.976	56.053	66.945	88.376	87.528	49.612
208	44.777	39.623	47.842	53.550	64.547	46.993		
209	24.840	62.006	45.961	37.090	46.958	51.319		
210	71.026	31.901	65.889	48.045	51.080			
211	46.203	59.387	51.701	81.467	33.855			
217	34.149	47.324	16.661	52.754	54.347			
218	58.544	23.582	58.528	69.601	50.656	56.247		
220	18.675	61.586	88.891	54.203	60.944			
221	40.824	54.105	43.152	1.084				
223	80.608	67.126	33.208	94.273	55.582			
231	109.711	57.299	115.576	60.912	55.743			
232	79.226	59.641	96.159	50.870	83.856			
233	-62.170	33.233	63.871	28.484	-5.686	33.413		
245	42.253	39.660	52.563	19.870	56.207			
246	90.974	93.749	63.860	89.915	65.876	80.744		
247	85.675	58.918	84.697	48.028	41.295			
248	24.680	39.303	30.869	49.803	52.113			
250	53.743	56.817	46.476	34.739	64.088			
253	65.371	42.079	34.843	69.285	77.060			
254	25.981	58.183	50.990	56.681	59.495			
257	75.279	46.280	67.023	50.801	40.223			
260	68.687	62.223	35.996	51.520	37.099	39.387		
261	50.071	56.478	107.451	58.521	-2.093	73.333		
274	41.705	43.823	48.647	59.842	5.573			
275	27.702	32.179	24.757	23.316	39.931			
288	58.140	59.373	58.093	75.317	67.633			
289	55.269	40.769	75.791	70.121	56.687			
290	44.810	34.396	38.616	50.439	63.655			
292	87.395	61.202	25.452	64.859	85.777			
295	39.015	17.606	39.970	39.140	48.996	23.760		

296	49.001	55.621	67.528	38.753	40.624	57.792	54.490
297	21.172	26.011	33.283	46.138	27.415		
300	34.467	3.248	30.354	31.044	45.528	41.623	
301	79.665	72.903	84.010	54.710	70.947	86.599	
308	63.663	35.947	76.505	40.244	45.529		
309	60.169	63.308	29.813	63.784	78.193		
313	62.489	71.477	37.224	60.396	56.587		
314	53.670	50.812	52.446	71.910	69.807	50.611	
318	59.329	90.099	80.462	60.640	59.711	50.096	
322	32.991	42.790	70.162	49.858	43.646		
323	54.486	55.037	61.565	45.819	32.236		
324	59.758	40.274	53.207	-12.722	36.134		
325	57.885	58.317	60.435	30.498	43.503	38.125	
326	54.369	67.366	98.776	83.383	71.581	74.774	
327	49.526	61.532	55.408	71.328			
328	51.578	69.587	50.096	43.013	45.721	70.306	
329	55.543	60.986	56.023	46.766	54.185		
330	62.798	29.417	57.708	37.959	44.099	38.227	
332	40.022	65.854	55.044	16.740	92.694		
333	50.479	60.930	60.131	41.914	87.491	26.959	
338	61.411	56.319	46.086	77.171	61.914		

References:

(1) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminform.* **2011**, *3*, 33.

(2) Baker, E. N.; Hubbard, R. E. Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97–179.

(3) Jiang, L.; Kuhlman, B.; Kortemme, T.; Baker, D. A "solvated rotamer" approach to modeling water-mediated hydrogen bonds at protein-protein interfaces. *Proteins: Struct., Funct., Bioinf.* **2005**, *58*, 893–904.

(4) Auffinger, P.; Hays, F. A.; Westhof, E.; Ho, P. S. Halogen bonds in biological molecules. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16789–16794.

(5) Barlow, D. J.; Thornton, J. M. Ion-pairs in proteins. *J. Mol. Biol.* **1983**, *168*, 867–885.

(6) Gallivan, J. P.; Dougherty, D. A. Cation-pi interactions in structural biology. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 9459–9464.

(7) McGaughey, G. B.; Gagné, M.; Rappé, A. K. pi-Stacking interactions. Alive and well in proteins. *J. Biol. Chem.* **1998**, *273*, 15458–15463.

(8) Lonsdale, R.; Houghton, K. T.; Żurek, J.; Bathelt, C. M.; Foloppe, N.; de Groot, M. J.; Harvey, J. N.; Mulholland, A. J. Quantum Mechanics/Molecular Mechanics Modeling of Regioselectivity of Drug Metabolism in Cytochrome P450 2C9. *J. Am. Chem. Soc.* **2013**, *135*, 8001–8015.