Supplementary Information

Unraveling the origin of aromatase inhibitory activity via proteochemometric modeling

Saw Simeon[1], Ola Spjuth[3], Maris Lapins[3], Sunanta Nabu[1], Virapong Prachayasittikul[2], Jarl E. S. Wikberg[3], and Chanin Nantasenamat*[1]

[1] *Center of Data Mining and Biomedical Informatics, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand*

[2] *Department of Clinical Microbiology and Applied Technology, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand*

[3] *Department of Pharmaceutical Biosciences, Uppsala University, Uppsala SE751 24, Sweden*

*Corresponding author. E-mail: chanin.nan@mahidol.ac.th

Supplementary Table 1

Summary of mutational information and pIC_{50} of aromatase variants

No.	Compound	Position on aromatase sequence													pIC_{50}
		119	**124**	**130**	**133**	**235**	**308**	**309**	**310**	**320**	**395**	**474**	**475**	**476**	
1	4_OHA	K	C	K	I	F	P	A	T	F	I	I	H	D	0.8729
2	4_OHA	K	C	K	I	L	P	D	T	F	I	I	H	D	1.2441
3	4_OHA	K	C	K	W	F	P	D	T	F	I	I	H	D	1.2757
4	4_OHA	K	C	K	Y	F	P	D	T	F	I	I	H	D	1.1739
5	4_OHA	K	C	K	I	F	P	D	T	F	F	I	H	D	1.5686
6	4_OHA	K	C	K	I	F	P	D	T	F	I	M	H	D	0.9830
7	4_OHA	K	C	K	I	F	P	D	T	F	I	N	H	D	0.9547
8	4_OHA	K	C	K	I	F	P	D	T	F	I	W	H	D	1.1079
9	4_OHA	K	C	K	I	F	P	D	T	F	I	Y	H	D	0.8013
10	4_OHA	K	C	K	I	F	F	D	T	F	I	I	H	D	2.0000
11	4_OHA	K	C	K	I	F	P	D	S	F	I	I	H	D	0.6990
12	4_OHA	K	C	K	I	F	P	D	T	F	I	I	H	D	1.2218
13	7α-APTADD	K	C	K	I	F	P	A	T	F	I	I	H	D	0.5575
14	7α-APTADD	K	C	K	I	L	P	D	T	F	I	I	H	D	0.6402
15	7α-APTADD	K	C	K	W	F	P	D	T	F	I	I	H	D	0.5686
16	7α-APTADD	K	C	K	Y	F	P	D	T	F	I	I	H	D	0.6968
17	7α-APTADD	K	C	K	I	F	P	D	T	F	F	I	H	D	0.5901
18	7α-APTADD	K	C	K	I	F	P	D	T	F	I	M	H	D	0.8386
19	7α-APTADD	K	C	K	I	F	P	D	T	F	I	N	H	D	0.8041
20	7α-APTADD	K	C	K	I	F	P	D	T	F	I	W	H	D	0.9431
21	7α-APTADD	K	C	K	I	F	P	D	T	F	I	Y	H	D	0.9914
22	7α-APTADD	K	C	K	I	F	F	D	T	F	I	I	H	D	1.2218
23	7α-APTADD	K	C	K	I	F	P	D	S	F	I	I	H	D	0.6757
24	7α-APTADD	K	C	K	I	F	P	D	T	F	I	I	H	D	0.7721
25	Aminoglutethimide	K	C	K	I	F	P	A	T	F	I	I	H	D	0.1367
26	Aminoglutethimide	K	C	K	I	L	P	D	T	F	I	I	H	D	–0.2304
27	Aminoglutethimide	K	C	K	W	F	P	D	T	F	I	I	H	D	–0.6812
28	Aminoglutethimide	K	C	K	Y	F	P	D	T	F	I	I	H	D	–0.8692
29	Aminoglutethimide	K	C	K	I	F	P	D	T	F	F	I	H	D	–0.5051
30	Aminoglutethimide	K	C	K	I	F	P	D	T	F	I	M	H	D	–0.2304
31	Aminoglutethimide	K	C	K	I	F	P	D	T	F	I	N	H	D	–0.3979
32	Aminoglutethimide	K	C	K	I	F	P	D	T	F	I	W	H	D	0.1367
33	Aminoglutethimide	K	C	K	I	F	P	D	T	F	I	Y	H	D	0.0655
34	Aminoglutethimide	K	C	K	I	F	F	D	T	F	I	I	H	D	–0.7243
35	Aminoglutethimide	K	C	K	I	F	P	D	S	F	I	I	H	D	0.1739
36	Aminoglutethimide	K	C	K	I	F	P	D	T	F	I	I	H	D	–0.7404

Supplementary Table 1

Continued ...

No.	Compound	Position on aromatase sequence													pIC_{50}
		119	**124**	**130**	**133**	**235**	**308**	**309**	**310**	**320**	**395**	**474**	**475**	**476**	
37	CGS20267	K	C	K	I	F	P	A	T	F	I	I	H	D	2.0969
38	CGS20267	K	C	K	I	L	P	D	T	F	I	I	H	D	2.8861
39	CGS20267	K	C	K	W	F	P	D	T	F	I	I	H	D	2.7447
40	CGS20267	K	C	K	Y	F	P	D	T	F	I	I	H	D	2.7212
41	CGS20267	K	C	K	I	F	P	D	T	F	F	I	H	D	2.8239
42	CGS20267	K	C	K	I	F	P	D	T	F	I	M	H	D	3.3979
43	CGS20267	K	C	K	I	F	P	D	T	F	I	N	H	D	3.0458
44	CGS20267	K	C	K	I	F	P	D	T	F	I	W	H	D	3.3979
45	CGS20267	K	C	K	I	F	P	D	T	F	I	Y	H	D	3.3979
46	CGS20267	K	C	K	I	F	F	D	T	F	I	I	H	D	2.7959
47	CGS20267	K	C	K	I	F	P	D	S	F	I	I	H	D	3.0458
48	CGS20267	K	C	K	I	F	P	D	T	F	I	I	H	D	2.8539
49	ICID1033	K	C	K	I	F	P	A	T	F	I	I	H	D	0.6990
50	ICID1033	K	C	K	I	L	P	D	T	F	I	I	H	D	1.7447
51	ICID1033	K	C	K	W	F	P	D	T	F	I	I	H	D	1.8239
52	ICID1033	K	C	K	Y	F	P	D	T	F	I	I	H	D	1.2757
53	ICID1033	K	C	K	I	F	P	D	T	F	F	I	H	D	1.4318
54	ICID1033	K	C	K	I	F	P	D	T	F	I	M	H	D	2.0969
55	ICID1033	K	C	K	I	F	P	D	T	F	I	N	H	D	1.6990
56	ICID1033	K	C	K	I	F	P	D	T	F	I	W	H	D	2.2218
57	ICID1033	K	C	K	I	F	P	D	T	F	I	Y	H	D	2.0969
58	ICID1033	K	C	K	I	F	F	D	T	F	I	I	H	D	1.4318
59	ICID1033	K	C	K	I	F	P	D	S	F	I	I	H	D	2.0969
60	ICID1033	K	C	K	I	F	P	D	T	F	I	I	H	D	1.5686
61	MDL101003	K	C	K	I	F	P	A	T	F	I	I	H	D	1.4559
62	MDL101003	K	C	K	I	L	P	D	T	F	I	I	H	D	1.7447
63	MDL101003	K	C	K	W	F	P	D	T	F	I	I	H	D	1.7959
64	MDL101003	K	C	K	Y	F	P	D	T	F	I	I	H	D	1.6576
65	MDL101003	K	C	K	I	F	P	D	T	F	F	I	H	D	1.8861
66	MDL101003	K	C	K	I	F	P	D	T	F	I	M	H	D	1.7959
67	MDL101003	K	C	K	I	F	P	D	T	F	I	N	H	D	1.7959
68	MDL101003	K	C	K	I	F	P	D	T	F	I	W	H	D	1.9586
69	MDL101003	K	C	K	I	F	P	D	T	F	I	Y	H	D	1.7959
70	MDL101003	K	C	K	I	F	F	D	T	F	I	I	H	D	1.3372
71	MDL101003	K	C	K	I	F	P	D	S	F	I	I	H	D	1.6990
72	MDL101003	K	C	K	I	F	P	D	T	F	I	I	H	D	1.9208

Supplementary Table 1

Continued ...

No.	Compound	Position on aromatase sequence													pIC_{50}
		119	**124**	**130**	**133**	**235**	**308**	**309**	**310**	**320**	**395**	**474**	**475**	**476**	
73	MR20492	K	Y	K	I	F	P	D	T	F	I	I	H	D	0.4815
74	MR20492	K	C	K	I	F	P	D	T	F	I	I	H	E	–0.7482
75	MR20492	K	C	K	I	F	P	D	T	F	I	I	H	N	0.0362
76	MR20492	K	C	K	I	F	P	D	T	C	I	I	H	D	–0.7284
77	MR20492	K	C	K	I	F	P	D	T	F	I	I	A	D	0.3468
78	MR20492	K	C	K	I	F	P	D	T	F	I	T	H	D	–1.0000
79	MR20492	E	C	K	I	F	P	D	T	F	I	I	H	D	0.1308
80	MR20492	V	C	K	I	F	P	D	T	F	I	I	H	D	–0.7931
81	MR20492	Y	C	K	I	F	P	D	T	F	I	I	H	D	–0.7574
82	MR20492	K	C	N	I	F	P	D	T	F	I	I	H	D	–0.6304
83	MR20492	K	C	K	I	F	P	D	T	F	I	I	H	D	–0.5944
84	MR20494	K	Y	K	I	F	P	D	T	F	I	I	H	D	0.4815
85	MR20494	K	C	K	I	F	P	D	T	F	I	I	H	E	0.8239
86	MR20494	K	C	K	I	F	P	D	T	F	I	I	H	N	0.8239
87	MR20494	K	C	K	I	F	P	D	T	C	I	I	H	D	0.0269
88	MR20494	K	C	K	I	F	P	D	T	F	I	I	A	D	1.0000
89	MR20494	K	C	K	I	F	P	D	T	F	I	T	H	D	–0.4472
90	MR20494	E	C	K	I	F	P	D	T	F	I	I	H	D	0.2757
91	MR20494	V	C	K	I	F	P	D	T	F	I	I	H	D	–0.1903
92	MR20494	Y	C	K	I	F	P	D	T	F	I	I	H	D	0.7696
93	MR20494	K	C	N	I	F	P	D	T	F	I	I	H	D	0.6990
94	MR20494	K	C	K	I	F	P	D	T	F	I	I	H	D	0.6383
95	MR20814	K	Y	K	I	F	P	D	T	F	I	I	H	D	–0.6435
96	MR20814	K	C	K	I	F	P	D	T	F	I	I	H	E	–0.6702
97	MR20814	K	C	K	I	F	P	D	T	F	I	I	H	N	–0.3054
98	MR20814	K	C	K	I	F	P	D	T	C	I	I	H	D	–1.0000
99	MR20814	K	C	K	I	F	P	D	T	F	I	I	A	D	0.0362
100	MR20814	K	C	K	I	F	P	D	T	F	I	T	H	D	–1.0000
101	MR20814	E	C	K	I	F	P	D	T	F	I	I	H	D	–0.9912
102	MR20814	V	C	K	I	F	P	D	T	F	I	I	H	D	–1.0000
103	MR20814	Y	C	K	I	F	P	D	T	F	I	I	H	D	–0.5752
104	MR20814	K	C	N	I	F	P	D	T	F	I	I	H	D	0.6383
105	MR20814	K	C	K	I	F	P	D	T	F	I	I	H	D	–1.0346
106	Vorozole	K	C	K	I	F	P	A	T	F	I	I	H	D	2.3188
107	Vorozole	K	C	K	Y	F	P	D	T	F	I	I	H	D	3.0969
108	Vorozole	K	C	K	I	F	P	D	T	F	F	I	H	D	3.3010

Supplementary Table 1

Continued ...

No.	Compound	Position on aromatase sequence													pIC_{50}
		119	**124**	**130**	**133**	**235**	**308**	**309**	**310**	**320**	**395**	**474**	**475**	**476**	
109	Vorozole	K	C	K	I	F	P	D	T	F	I	Y	H	D	3.6990
110	Vorozole	K	C	K	I	F	F	D	T	F	I	I	H	D	3.0000
111	Vorozole	K	C	K	I	F	P	D	S	F	I	I	H	D	3.3010
112	Vorozole	K	C	K	I	F	P	D	T	F	I	I	H	D	3.0458

Supplementary Table 2

List of SMILES notation for aromatase inhibitors used herein

	Compound	SMILES
1	4_OHA	C[C@]12CCC3C(CCC4=C(O)C(=O)CC[C@]34C)C1CCC2=O
2	7alpha-APTADD	C[C@]12CCC3C(C1CCC2=O)C(CC1=CC(=O)C=C[C@]31C)SC1=CC=C(N)C=C1
3	AG	CCC1(CCC(=O)NC1=O)C1CCC(N)CC1
4	CGS20267	[C-]#[N+]C1=CC=C(C=C1)C(N1C=NC=N1)C1=CC=C(C=C1)[N+]#[C-]
5	ICID1033	CC(C)([N+]#[C-])C1=CC(=CC(CN2C=NC=N2)=C1)C(C)(C)[N+]#[C-]
6	MDL101003	C[C@]12CCC3C(CCC4=CC(=O)[C@H]5COC[C@@]34C5)C1CCC2=O
7	MR20492	ClC1=CC=C(C=C1)C1CN2C=CC=C2C(=O)\C1=C/C1=CC=NC=C1
8	MR20494	ClC1=CC=C(C=C1)C1CN2C=CC=C2C(=O)\C1=C/C1=CN=CC=C1
9	MR20814	[H]C([H])(O)C1=CC2=C(C=C1)C(=O)C(CC1=CC=NC=C1)=C2N
10	Vorozole	CN1N=NC2=C1C=C(C=C2)C(N1C=NC=N1)C1=CC=C(Cl)C=C1