

Supplementary Information

Probing the origin of estrogen receptor alpha inhibition via large-scale QSAR study

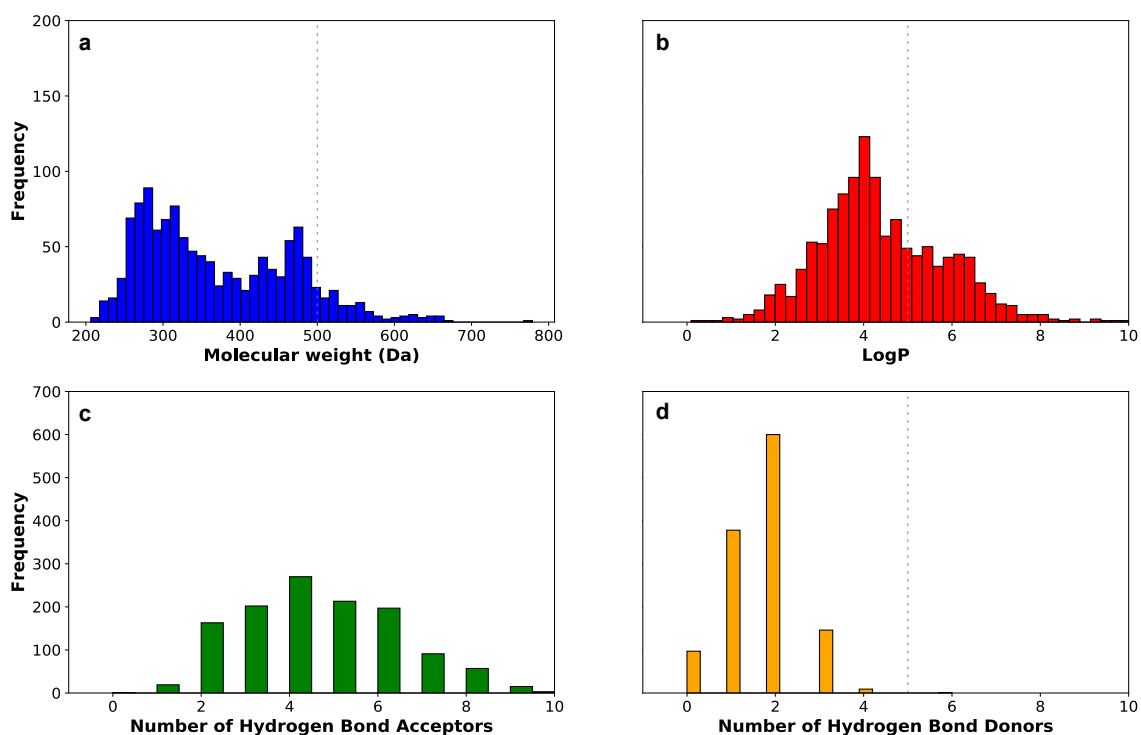
Naravut Suvannang^a, Likit Preeyanon^b, Aijaz Ahmad Malik^a, Nalini Schaduangrat^a,
Watshara Shoombuatong^a, Apilak Worachartcheewan^b, Tanawut Tantimongcolwat^c, and
Chanin Nantasenamat^{*1}

^a*Center of Data Mining and Biomedical Informatics, Faculty of Medical Technology, Mahidol University, Bangkok
10700, Thailand*

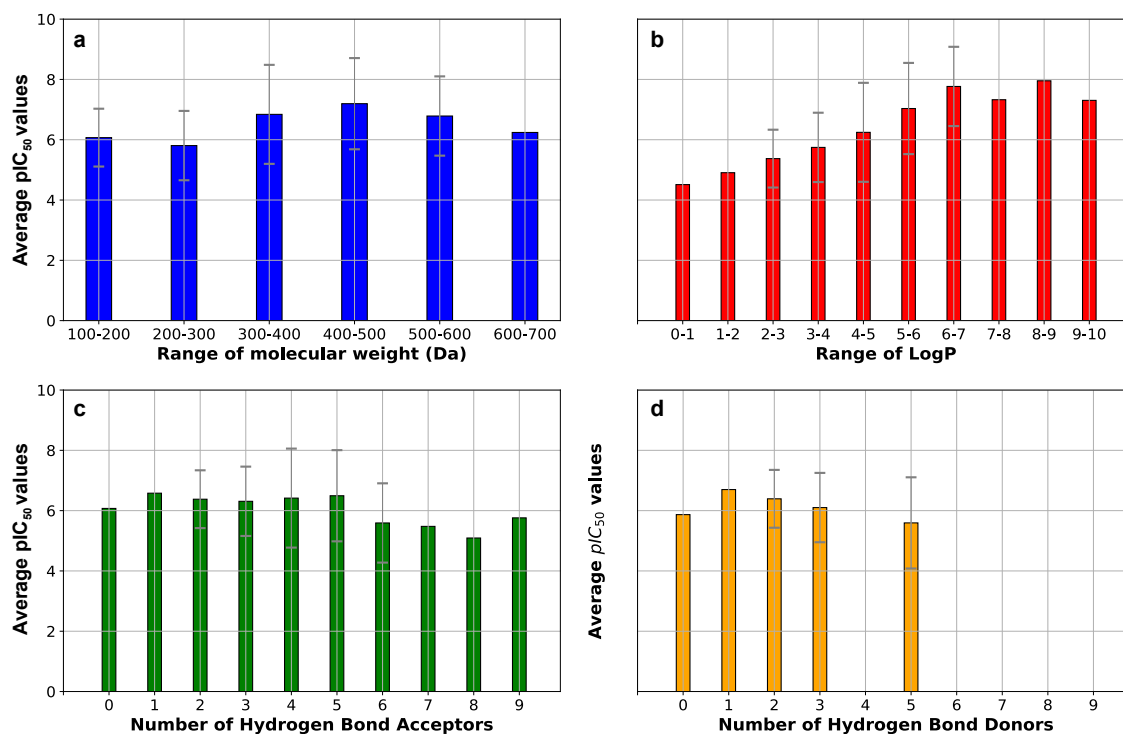
^b*Department of Community Medical Technology, Faculty of Medical Technology, Mahidol University, Bangkok 10700,
Thailand*

^c*Center for Research and Innovation, Faculty of Medical Technology, Mahidol University, Bangkok 10700, Thailand*

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



Supplementary Figure 1. Histogram plots of frequency count of Ro5 descriptors for ER α inhibitors. This consisted of molecular weight (a), LogP (b), number of hydrogen bond acceptors (c) and number of hydrogen bond donors (d).



Supplementary Figure 2. Histogram plots of mean pIC₅₀ values for Ro5 descriptors for ER α inhibitors. This consisted of molecular weight (a), LogP (b), number of hydrogen bond acceptors (c) and number of hydrogen bond donors (d).

Supplementary Table 1Summary of predictive performance for QSAR model of ER α inhibitory activity.

Fingerprint class	Training set		10-fold CV set		External set		$R^2 - Q_{cv}^2$	$Q^2 - Q_{ext}^2$
	R_{Tr}^2	RMSE $_{Tr}$	Q_{CV}^2	RMSE $_{CV}$	Q_{Ext}^2	RMSE $_{Ext}$		
AtomPairs 2D count	0.93	0.38	0.73	0.72	0.73	0.53	0.20	0.003
AtomPairs 2D	0.85	0.54	0.67	0.80	0.68	0.62	0.18	-0.01
CDK fingerprinter	0.87	0.51	0.68	0.78	0.71	0.56	0.19	-0.03
CDK extended	0.84	0.55	0.69	0.78	0.67	0.65	0.16	0.02
CDK graph only	0.81	0.60	0.66	0.81	0.70	0.58	0.15	-0.04
E-State	0.80	0.63	0.66	0.82	0.64	0.71	0.14	0.02
Klekota-Roth count	0.91	0.41	0.71	0.75	0.72	0.54	0.20	-0.01
Klekota-Roth	0.82	0.60	0.65	0.82	0.70	0.59	0.16	-0.04
MACCS	0.86	0.52	0.73	0.72	0.71	0.58	0.13	0.03
PubChem	0.84	0.57	0.70	0.76	0.71	0.56	0.13	-0.01
Substructure count	0.94	0.34	0.73	0.72	0.73	0.52	0.21	-0.004
Substructure	0.87	0.51	0.69	0.77	0.68	0.63	0.18	0.02