## **Supporting Information**

## Inverse QSPR/QSAR Analysis for Chemical Structure Generation (from y to x)

Tomoyuki Miyao, Hiromasa Kaneko and Kimito Funatsu\*.

Department of Chemical System Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyoku, Tokyo 113-8656, Japan

## **Corresponding Author**

\*E-mail: funatsu@chemsys.t.u-tokyo.ac.jp

**Table S1.** Statistics of cMLR of the alpha 1B adrenergic receptor QSAR model.

Statistics	C1	C2	C3	C4	MLR
Samples	92	160	141	207	600
Variables	13	14	11	14	14
Adjusted R <sup>2</sup>	0.421	0.619	0.508	0.454	0.273
<i>F</i> -statistics	6.52	20.83	15.48	14.19	18.30

Samples: the number of samples classified to the cluster, variables: the number of variables used for constructing a cMLR model. Adjusted  $R^2$  is an R squared adjusting the value based on the degree of freedom.

**Table S2.** Standard linear regression coefficients of cMLR for alpha1B adrenergic receptor QSAR model.

Descriptor	C1	C2	C3	C4	MLR
CIC	0.20	0.08	-0.40	0.20	0.11
R05	-0.06	-0.45	0.38	-0.31	-0.23
BR	0.11	-0.17	-0.34	0.02	-0.10
CH2R2	-0.05	0.50	0.17	0.14	0.36
CH3X	-0.05	0.23	-0.10	0.36	0.36
O2	0.09	0.42	-0.41	0.44	0.25
ArNR2	0.00	0.23		-0.13	0.05
X1	0.28	-0.78	-0.15	-0.71	-0.87
TPSA	-0.08	-0.12	0.44	-0.11	0.05
HBD	-0.05	0.13	-0.62	0.04	-0.07
RR	-0.47	-0.49		0.05	0.48
HBA	-0.31	0.37	0.02	0.32	0.23

aR -0.32 1.15 0.38 0.15

**Table S3.** Generation results for cluster 2 and 3 respectively. The constraints are ones in Table. 6 in main manuscript.

NF	Cluster2		Cluster3	Cluster3		
	Time (s)	NGS	Time (s)	NGS		
2	1.36	69	0.87	118		
3	0.8	138	0.8	1,028		
4	0.75	94	0.75	5,476		
5	0.75	51	0.86	20,633		
6	0.75	30	1.38	99,537		
7	0.77	0	4.78	372,242		
8	0.75	0	22.60	495,343		
9	0.75	0	75.37	230,446		
10	0.75	0	190.03	576,047		
11	0.75	0	346.68	850,288		
2-11	0.77	382	345.53	2,651,158		

NF: number of fragments combined. NGS: number of generated structures.

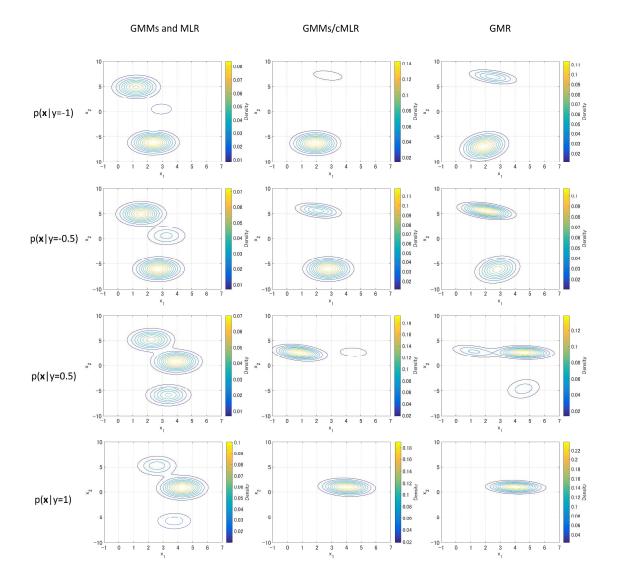
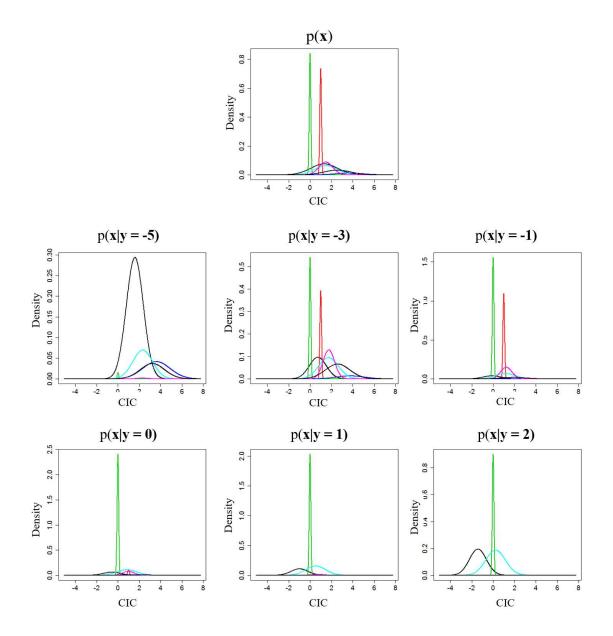
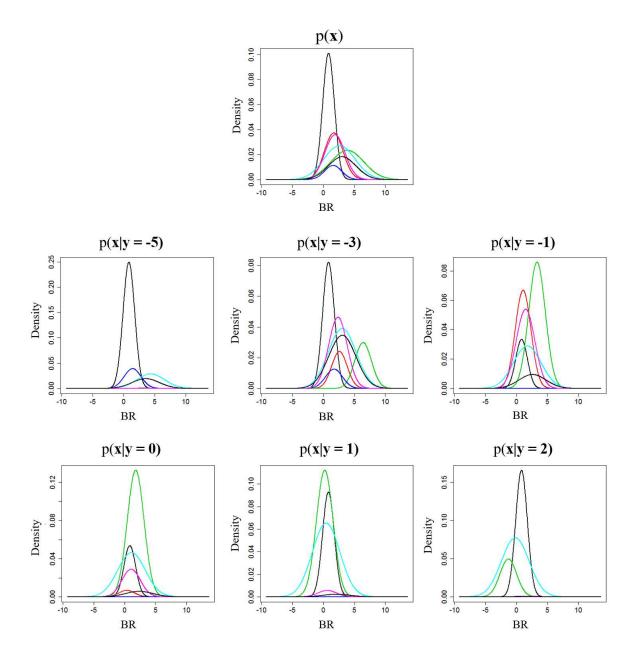


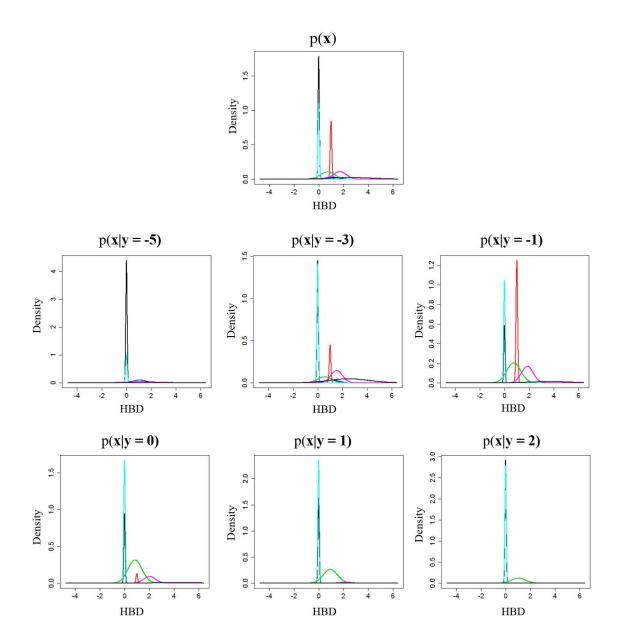
Figure S1. Contour plots of p(x|y) obtained by three methodologies for the simulation dataset.



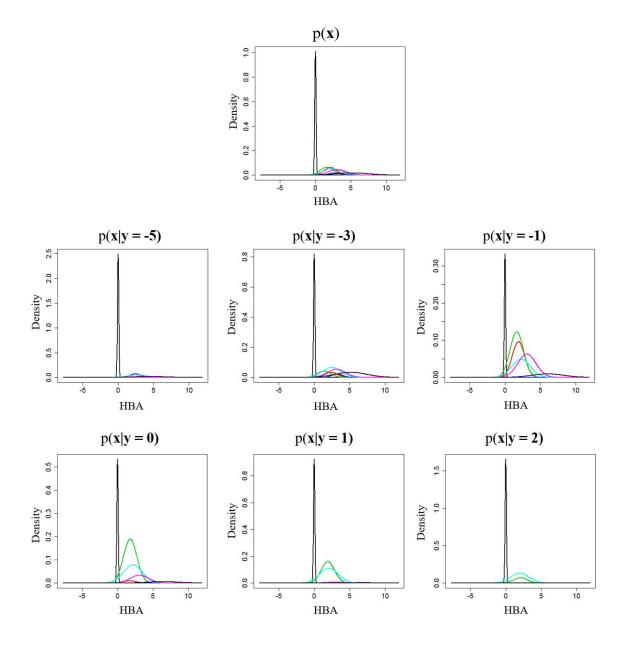
**Figure S2.** Marginal posterior distributions of CIC as well as corresponding prior distribution for y = -5, -3, -1, 0, 1, 2.



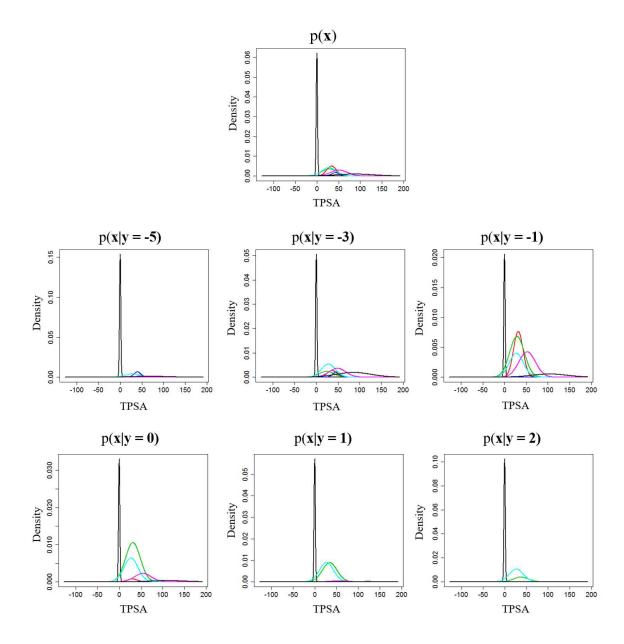
**Figure S3.** Marginal posterior distributions of BR as well as corresponding prior distribution for y = -5, -3, -1, 0, 1, 2.



**Figure S4.** Marginal posterior distributions of HBD as well as corresponding prior distribution for y = -5, -3, -1, 0, 1, 2.



**Figure S5.** Marginal posterior distributions of HBA as well as corresponding prior distribution for y = -5, -3, -1, 0, 1, 2.



**Figure S6.** Marginal posterior distributions of TPSA as well as corresponding prior distribution for y = -5, -3, -1, 0, 1, 2.