Supporting Data for

Machine learning-driven QSAR models for prediction of Metabolic mechanisms and Thyroid hormone disrupting effects of emerging pollutants in human body: a case study of Bisphenol analogues

Zijian Wang, Qingzhu Zhang[*], Wenxing Wang, Qiao Wang

Academician Workstation for Big Data Research in Ecology and Environment, Environmental Research Institute, Shandong University, Qingdao 266237, P. R. China

_____

[*]Corresponding authors. E-mail: zqz@sdu.edu.cn

Contents

Text S1 Standardization of molecular structural descriptors.

All values of molecular structural descriptors were standardized by the following eq. (S1) in QSAR modeling, so as to eliminate the effects of variable unit and data range of the different descriptors.

$$x^* = \frac{x - \min}{\max - \min} \qquad (S1)$$

Where x is the original value, x^* is standardized value, max is the maximum value of values corresponding to the descriptor, min is the minimum value of values corresponding to the descriptor.
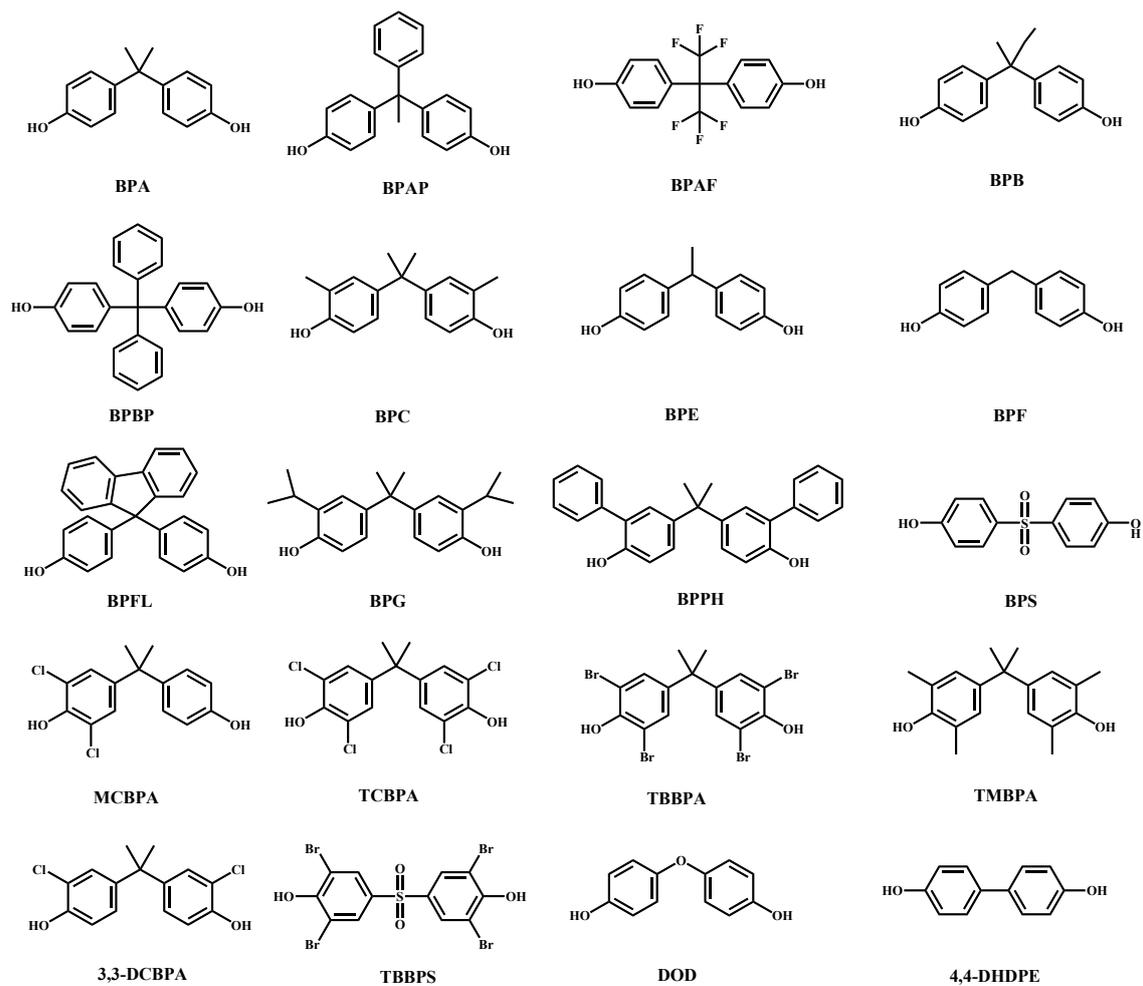
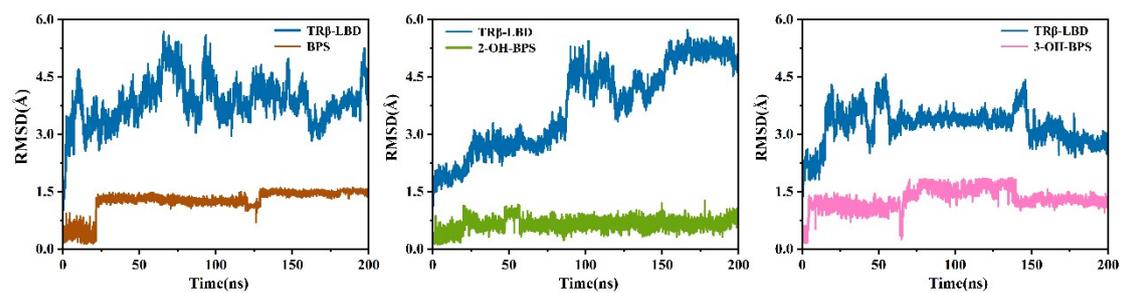**Figure S1** The chemical structures of 20 BPs.

Figure S2 The root-mean-square deviation (RMSD) values of MD simulations of (OH)-BPs within TRβ-LBD.
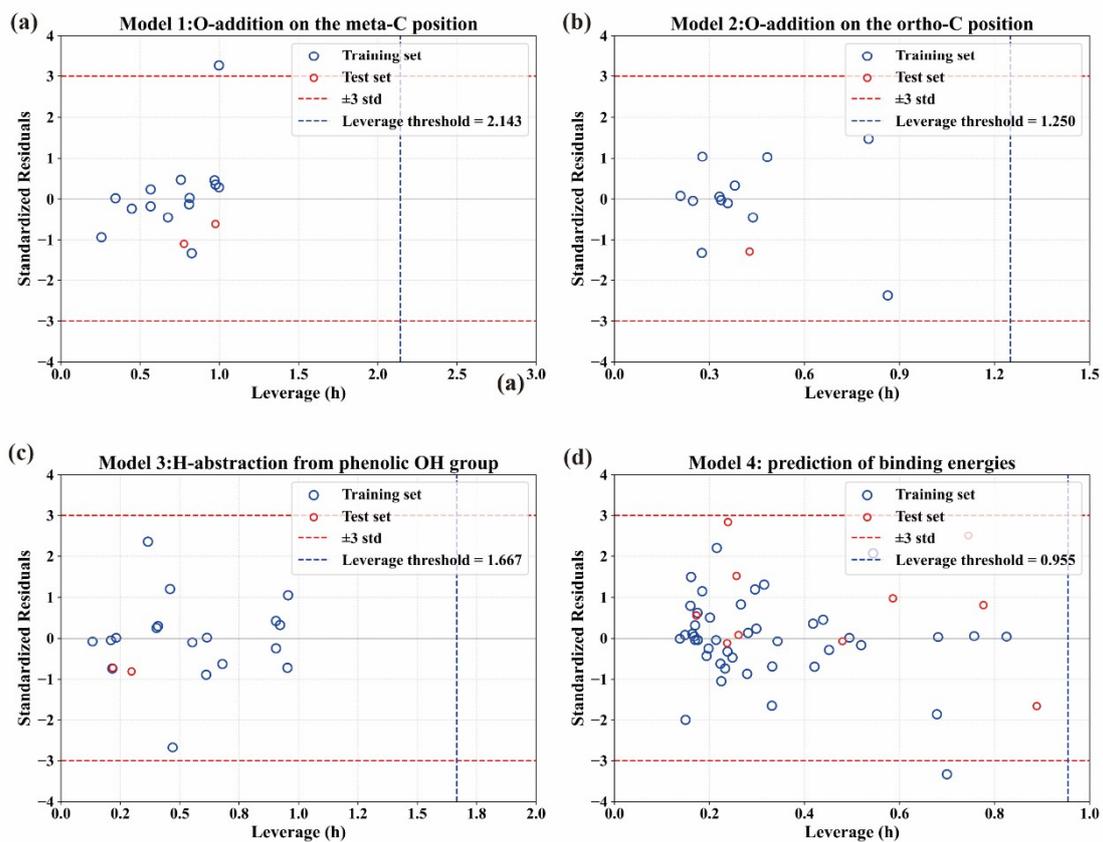
Figure S3 The Williams plots of QSAR models (1)-(4).
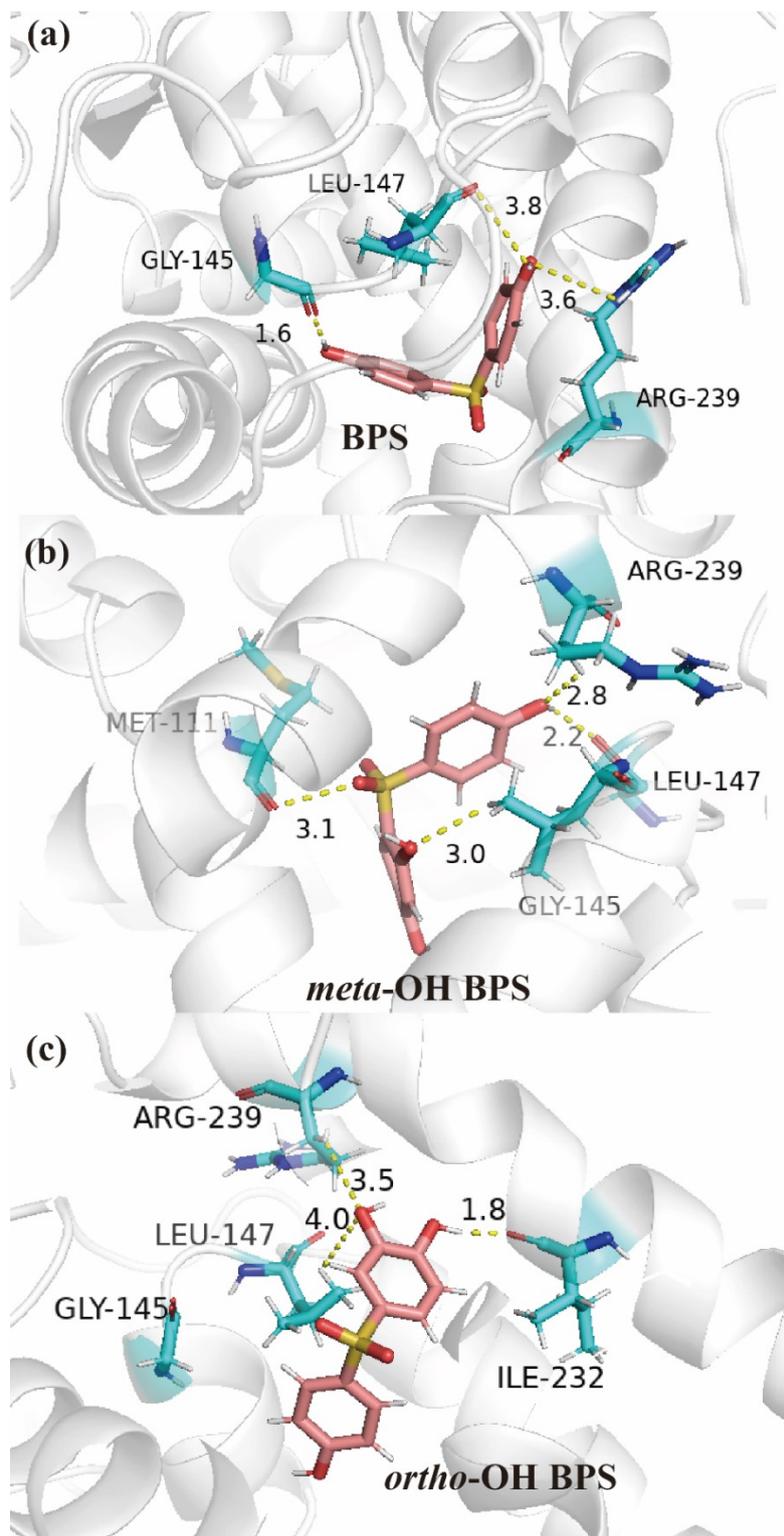
Figure S4 The binding models of (OH)-BPS in the ligand binding domain of TRβ-LBD.

Table S1 The energy data parameters of O-addition and H-abstraction.

No.	Name	Energy barriers		
		O-addition(C_m site)	O-addition(C_o site)	H-abstraction
1	BPA	13.62 [15.49]	11.83 [13.31]	8.13 [8.44]
2	BPAP	15.56 [17.30]	20.01 [22.77]	6.69 [6.67]
3	BPAF	18.95 [20.30]	10.46 [12.43]	8.83 [10.11]
4	BPB	13.95 [15.55]	12.13 [14.85]	9.19 [9.37]
5	BPBP	17.99 [19.87]	18.21 [20.74]	5.10 [4.80]
6	BPC	13.10 [15.14]	10.07 [16.19]	6.58 [6.92]
7	BPE	13.45 [15.33]	18.13 [19.49]	8.10 [8.40]
8	BPF	15.20 [16.83]	19.58 [22.29]	8.62 [8.43]
9	BPFL	16.02 [18.43]	11.02 [13.48]	7.57 [8.86]
10	BPG	14.43 [16.27]	12.47 [15.86]	6.52 [6.92]
11	BPPH	15.54 [20.26]	20.37 [26.95]	1.02 [0.57]
12	BPS	19.60 [22.99]	13.17 [15.10]	8.18 [8.75]
13	MCBPA	17.56 [19.80]	—	6.99 [6.85]
14	TCBPA	11.50 [17.78]	—	6.66 [6.58]
15	TBBPA	18.29 [21.14]	—	6.33 [6.27]
16	TMBPA	15.90 [19.08]	—	5.91 [6.46]
17	3,3-DCBPA	14.97 [16.64]	13.63 [15.55]	8.04 [7.59]
18	TBBPS	25.05 [25.92]	—	3.04[3.27]
19	DOD	13.88 [16.16]	20.61 [22.56]	6.41 [6.10]
20	4,4-DHBPE	15.60 [17.74]	20.38 [24.97]	6.81 [7.09]

Table S2 The definition of the relevant chemical descriptors.

Chemical descriptors	Definitions
D1	Mulliken change of $C_{O\text{-}p}$
D2	Mulliken change of $C_{O\text{-}m}$
D3	Mulliken change of $C_{O\text{-}o}$
D4	Mulliken change of $H_{H\text{-}OH}$
D8	The gap between E(HOMO) and E(LUMO)
D10	Electrochemical potential
AATS3s	averaged moreau-broto autocorrelation of lag 3 weighted by intrinsic state
AATS4Z	averaged moreau-broto autocorrelation of lag 4 weighted by atomic number
AATSC4c	centered moreau-broto autocorrelation of lag 4 weighted by gasteiger charge
AATSC0c	averaged and centered moreau-broto autocorrelation of lag 0 weighted by gasteiger charge
AATSC1c	averaged and centered moreau-broto autocorrelation of lag 1 weighted by gasteiger charge
AATS4s	averaged moreau-broto autocorrelation of lag 4 weighted by intrinsic state
AATS6d	averaged moreau-broto autocorrelation of lag 6 weighted by valence electrons
ABC	atom-bond connectivity index
AMID_X	averaged molecular ID on halogen atoms
ATS0s	moreau-broto autocorrelation of lag 0 weighted by intrinsic state
ATS3v	moreau-broto autocorrelation of lag 3 weighted by vdw volume
ATSC5d	centered moreau-broto autocorrelation of lag 5 weighted by sigma electrons
ATSC5dv	centered moreau-broto autocorrelation of lag 5 weighted by valence electrons
ATSC7p	centered moreau-broto autocorrelation of lag 7 weighted by polarizability
ATS8Z	moreau-broto autocorrelation of lag 8 weighted by atomic number
ATSC8se	centered moreau-broto autocorrelation of lag 8 weighted by sanderson EN
BCUTare-1l	first lowest eigenvalue of Burden matrix weighted by allred-rocow EN
BCUTse-1l	first lowest eigenvalue of Burden matrix weighted by sanderson EN
ETA_epsilon_4	ETA epsilon (type: 4)

GATS6p	geary coefficient of lag 6 weighted by polarizability
GATS6se	geary coefficient of lag 6 weighted by sanderson EN
GATS2pe	geary coefficient of lag 2 weighted by pauling EN
GATS2se	geary coefficient of lag 2 weighted by sanderson EN
GATS7d	geary coefficient of lag 7 weighted by sigma electrons
MATS3d	moran coefficient of lag 3 weighted by sigma electrons
MINsCH3	min of sCH3
MINsOH	min of sOH
SpAbs_A	SpAbs of adjacency matrix

Table S3 Selected molecular descriptors and corresponding values for model (1).

BPs	AATS4Z	D10	D4	ATS0s	ATSC8se	SpAbs_A	GATS6se	D3	D2
3,3-DCBPA	16.69892	0.118205	0.09642	139.8509	-0.10226	23.45396	1.260122	-0.12414	-0.12593
4,4-DHBPE	13.33333	0.10377	0.10679	125.1111	-0.49693	18.12129	0.73899	-0.13309	-0.06659
BPA	11.88172	0.104375	0.08866	140.6736	-0.15757	21.01931	0.623812	-0.12132	0.010582
BPAF	32.01075	0.11917	0.10372	513.7986	-5.04717	26.93094	1.68405	-0.13802	-0.10584
BPAP	15.29091	0.110075	0.088315	161.4514	-0.49228	28.6968	0.499616	-0.10971	-0.11733
BPB	12.68627	0.104	0.086562	144.9236	-0.89752	22.73429	0.532562	-0.13594	-0.11768
BPBP	18.28125	0.11124	0.094917	182.2292	-0.70238	36.14953	0.448408	-0.14463	-0.11329
BPC	11.71795	0.09848	0.087369	150.2292	-0.04047	23.45396	0.596415	-0.12837	-0.12987
BPE	12.27778	0.104435	0.089067	134.8889	-0.29223	20.33742	0.831555	-0.13506	-0.11777
BPF	11.2	0.10547	0.090187	129.3611	-0.45948	19.30404	1.230272	-0.13137	-0.11679
BPFL	19.24409	0.11859	0.093406	177.7847	-0.31421	36.56979	0.482897	-0.14131	-0.11794
BPG	12.28302	0.09917	0.084626	169.7847	0.570028	27.46491	0.527322	-0.13008	-0.12735
BPPH	15.09396	0.114965	0.078611	191.7847	0.073981	37.87286	1.046386	-0.11461	-0.14367
BPS	20.4	0.13535	0.150983	223.2161	-0.66856	21.01931	0.24495	-0.09258	-0.08181
DOD	14.72727	0.10441	0.102659	137.3611	-0.61552	19.30404	0.605214	-0.13413	-0.09729
MCBPA	19.45161	0.120095	0.090155	139.8509	-0.88195	23.1679	1.260122	-0.13469	-0.12342
TCBPA	27.02151	0.131855	0.114409	139.0281	1.134398	25.31556	1.659563	0	-0.1174
TBBPA	57.21505	0.13086	0.112779	136.32	0.273431	25.31556	1.362043	0	-0.14134
TBBPS	113.3333	0.15711	0.1429	218.8626	-1.08385	25.31556	0.39931	0	-0.10168
TMBPA	11.96454	0.09452	0.078142	159.7847	0.070185	25.31556	0.513488	0	-0.17015

Table S4 Selected molecular descriptors and corresponding values for model (2).

BPs	GATS2pe	GATS7d	ABC	GATS2se
3,3-DCBPA	0.516144	0.516144	14.693433	0.560595
4,4-DHBPE	0.605436	0.605436	10.784941	0.59866
BPA	0.562622	0.562622	13.14132	0.548142
BPAF	0.36922	0.36922	17.830166	0.396722
BPAP	0.609306	0.609306	17.163432	0.563349
BPB	0.578098	0.578098	13.689508	0.545388
BPBP	0.651111	0.651111	21.185545	0.577829
BPC	0.532215	0.532215	14.693433	0.525265
BPE	0.595497	0.595497	12.268104	0.567216
BPF	0.61057	0.61057	11.532488	0.583673
BPFL	0.604668	0.604668	21.771331	0.549089
BPG	0.566458	0.566458	17.659759	0.520833
BPPH	0.616716	0.616716	22.879054	0.557815
BPS	0.574597	0.574597	13.14132	0.628854
DOD	0.735432	0.735432	11.532488	0.814021
MCBPA	0.516144	0.516144	14.693433	0.560595
TCBPA	0.509791	0.509791	16.245546	0.59502
TBBPA	0.498761	0.498761	16.245546	0.573521
TBBPS	0.624739	0.624739	16.245546	0.738212
TMBPA	0.511097	0.511097	16.245546	0.509226

Table S5 Selected molecular descriptors and corresponding values for model (3).

BPs	BCUTare-1l	BCUTse-1l	AATS6d	AATSC4c	GATS6p	ATSC7p	ETA_epsilon_4	D8	D1
3,3-DCBPA	2.179866	2.426686	2.310185	-0.0022177	1.157158	-3.61442	0.533439	0.19867	-0.083256
4,4-DHBPE	2.176879	2.422806	3.6	-0.0024959	0.696953	-0.63954	0.498148	0.18646	0
BPA	2.182128	2.428055	2	-0.0011797	0.603738	-2.35692	0.485185	0.20251	-0.111085
BPAF	2.182868	2.428722	7.5	0.00681094	1.750393	14.18659	0.68709	0.20782	-0.310066
BPAP	2.176004	2.421961	2.704918	-0.0006162	0.601991	-0.08169	0.478161	0.19897	-0.176457
BPB	2.182325	2.428251	1.6875	-0.0010189	0.573895	-1.12007	0.481944	0.2028	-0.079247
BPBP	2.170577	2.416548	3.230769	0.00010769	0.63993	1.982227	0.473709	0.19118	-0.175749
BPC	2.177237	2.423158	1.760417	-0.0022178	0.689045	-0.96327	0.479085	0.20272	-0.080876
BPE	2.184828	2.430746	2.481481	-0.0011928	0.828974	-0.81456	0.488889	0.20329	-0.169044
BPF	2.186882	2.432794	3.444444	-0.0009549	1.265042	0.870145	0.493162	0.2038	-0.277066
BPFL	2.175134	2.42109	3.537975	-1.33E-05	0.668926	1.919387	0.478744	0.17756	-0.247449
BPG	2.176834	2.422756	1.506944	-0.0003278	0.627856	-2.81304	0.47037	0.2018	-0.079058
BPPH	2.169559	2.415496	2.925676	-0.0003387	1.012396	-3.36051	0.470563	0.17637	-0.074902
BPS	2.176627	2.437128	6.777778	-0.0021153	0.315007	-0.24969	0.559829	0.19658	0.891895
DOD	2.196881	2.442562	3.875	-0.0090017	0.694587	0.782659	0.524324	0.18544	-0.607018
MCBPA	2.17949	2.426546	2.296296	-0.0020502	1.157158	-3.61442	0.533439	0.19403	-0.081241
TCBPA	2.177671	2.425338	2.648148	-0.0028924	1.591985	-5.45416	0.581693	0.19721	-0.084706
TBBPA	2.176919	2.424426	2.475309	-0.00257	1.357653	-4.61685	0.526667	0.19442	-0.083657
TBBPS	2.172114	2.432585	7.728395	-0.0069636	0.359736	-2.79261	0.607692	0.1871	0.903527
TMBPA	2.172743	2.418659	1.42029	-0.002735	0.664457	0.055181	0.474269	0.20372	-0.080327

Table S6 Selected molecular descriptors and corresponding values for model (4).

BPs	ATSC5d	AATSC0c	AATSC1c	MINsOH	MATS3d	ATS3v	SpAbs_A	AATS4s	MINsCH3	AATS3s	ATSC5dv	ATS8Z	AMID_X	D4
						BPs								
3,3-DCBPA	32.30165	0.024025	-0.01323	9.474396	-0.27736	17873.25	23.45396	2.423138	2.02621	3.040746	32.30165	836	0.184943	0.09642
4,4-DHBPE	35.11111	0.031679	-0.01735	9.107689	-0.26974	10346.22	18.12129	2.837963	0	3.417778	35.11111	156	0	0.10679
BPA	40.38935	0.023425	-0.01295	9.296299	-0.29	15799.62	21.01931	2.4319	2.116827	2.989766	40.38935	324	0	0.08866
BPAF	-251.6	0.039854	-0.02019	9.139412	0.102348	18663.24	26.93094	11.94803	0	5.305556	-251.6	420	0.442177	0.10372
BPAP	20.335	0.020194	-0.0105	9.566868	-0.21807	20926.85	28.6968	2.580303	2.153864	2.88357	20.335	722	0	0.088315
BPB	27.39198	0.021601	-0.01192	9.369004	-0.21497	17256.67	22.73429	2.423203	2.138605	2.785948	27.39198	382	0	0.086562
BPBP	-61.5156	0.017962	-0.00883	9.837436	-0.03738	26054.08	36.14953	2.644531	0	2.805556	-61.5156	1212	0	0.094917
BPC	7.905325	0.019773	-0.01112	9.627452	-0.29215	19020.32	23.45396	2.531814	1.899787	3.034722	7.905325	513	0	0.087369
BPE	34	0.025693	-0.014	9.212352	-0.25252	13280.91	20.33742	2.524691	2.095994	3.003317	34	318	0	0.089067
BPF	21.14129	0.028436	-0.01532	9.12609	-0.24397	11264.12	19.30404	2.45	0	3.141762	21.14129	312	0	0.090187
BPFL	-82.1002	0.018389	-0.00904	9.847852	-0.03877	27650.55	36.56979	2.670604	0	2.86036	-82.1002	1091	0	0.093406
BPG	-98.2222	0.015437	-0.0087	10.0619	-0.08507	24617.24	27.46491	2.477289	2.086564	2.608871	-98.2222	1009	0	0.084626
BPPH	0.601638	0.015844	-0.00794	10.42081	-0.11067	28269.13	37.87286	2.675988	2.162012	2.830247	0.601638	1601	0	0.078611
BPS	-36.0341	0.033489	-0.01812	9.101477	-0.01336	13170.79	21.01931	4.706276	0	4.55364	-36.0341	312	0	0.150983
DOD	34.4384	0.039732	-0.02132	9.070534	-0.29329	9732.181	19.30404	3.128788	0	3.804444	34.4384	312	0	0.102659
MCBPA	32.30165	0.024065	-0.01312	9.342308	-0.27131	17873.25	23.1679	2.424423	2.02621	3.037714	32.30165	580	0.184899	0.090155
TCBPA	21.60777	0.024705	-0.01329	9.652493	-0.26019	19946.88	25.31556	2.416947	1.935593	3.085662	21.60777	1860	0.334581	0.114409
TBBPA	19.53109	0.023703	-0.01301	9.849854	-0.2763	20947.83	25.31556	2.351046	2.086827	2.932911	19.53109	6036	0.334581	0.112779
TBBPS	-23.8926	0.033955	-0.01776	9.655032	-0.09995	18319	25.31556	4.35473	0	4.500426	-23.8926	6024	0.334581	0.1429
TMBPA	-36.6854	0.017097	-0.00976	9.958604	-0.26996	22241.01	25.31556	2.58353	1.925828	3.055556	-36.6854	752	0	0.078142
						meta-oh-BPs								
3,3-DCBPA	-25.3691	0.033669	-0.01859	9.479604	-0.22194	18717.63	24.30505	3.080618	1.887043	3.327807	-25.3691	916	0.175695	0.315777
4,4-DHBPE	-3.728	0.044561	-0.02448	9.114492	-0.21028	11190.6	18.92605	3.940171	0	3.882479	-3.728	171	0	0.313428
BPA	-14.8858	0.033068	-0.0184	9.301508	-0.22581	16644	21.75585	3.107388	1.977661	3.276709	-14.8858	388	0	0.311605
BPAF	-182.346	0.049116	-0.02501	9.14462	0.077559	19507.62	27.73243	12.00945	0	5.53953	-182.346	484	0.423724	0.31409

BPAP	-14.8322	0.028287	-0.01495	9.572076	-0.17204	21771.23	29.4444	3.13231	2.014698	3.120081	-14.8322	801	0	0.31279
BPB	-25.8371	0.030513	-0.01696	9.374212	-0.15094	18101.05	23.48129	3.014937	2.033216	3.049808	-25.8371	446	0	0.312626
BPBP	-79.8958	0.024936	-0.01257	9.842644	-0.00745	26898.46	36.94404	3.104167	0	3.007066	-79.8958	1306	0	0.311741
BPC	-43.61	0.028064	-0.0158	9.63266	-0.22929	19864.7	24.30505	3.053489	1.802009	3.283642	-43.61	603	0	0.310234
BPE	2.058273	0.036185	-0.01995	9.21756	-0.21118	14125.29	21.08754	3.330409	1.956827	3.338164	2.058273	382	0	0.313676
BPF	14.40816	0.03994	-0.02185	9.131298	-0.23389	12108.5	20.11582	3.351563	0	3.541667	14.40816	376	0	0.313134
BPFL	-90.6975	0.025658	-0.01285	9.853061	-0.01965	28494.93	37.36068	3.134648	0	3.062684	-90.6975	1185	0	0.311705
BPG	-135.828	0.021898	-0.01237	10.06711	-0.03567	25461.62	28.21759	2.867246	2.01419	2.79343	-135.828	1175	0	0.310311
BPPH	-81.5926	0.022151	-0.01135	10.42602	-0.03889	29250.57	38.66661	3.219862	2.022846	3.020616	-81.5926	1755	0	0.316106
BPS	-80.5805	0.045367	-0.02378	9.106685	-0.03775	14050.34	21.75585	6.385947	0	4.741821	-80.5805	376	0	0.319103
DOD	10.69822	0.052511	-0.0282	9.075742	-0.21014	10522.95	20.11582	4.107639	0	4.445513	10.69822	376	0	0.315562
MCBPA	-26.3299	0.033689	-0.01856	9.363141	-0.22	18717.63	23.90522	3.10022	1.887043	3.323428	-26.3299	676	0.175651	0.31849
TCBPA	-42.5343	0.033637	-0.01827	9.855062	-0.23018	21983.56	26.69291	3.013132	1.947661	3.222618	-42.5343	6168	0.319074	0.325205
TBBPA	-70.3893	0.04618	-0.02321	9.66024	-0.13487	19389.89	26.69291	6.031006	0	4.692078	-70.3893	6156	0.319074	0.324291
TBBPS	-40.3783	0.034699	-0.01849	9.657701	-0.21012	20945.41	26.69291	3.078245	1.796426	3.389456	-40.3783	1956	0.319074	0.335393
TMBPA	-78.0265	0.024384	-0.01394	9.963812	-0.21655	23222.46	26.69291	3.127778	1.712865	3.31073	-78.0265	868	0	0.31337
ortho-oh-BPs														
3,3-DCBPA	26.36277	0.034166	-0.01898	9.478511	-0.19877	18664.02	24.38479	2.695272	1.928432	3.648267	26.36277	1018	0.175675	0.10509
4,4-DHBPE	40.2944	0.045044	-0.02513	9.112898	-0.20904	11136.99	19.12917	3.337607	0	4.346154	40.2944	220	0	0.090857
BPA	38.64014	0.033478	-0.0189	9.300415	-0.21997	16590.38	22.23695	2.68929	2.01905	3.621439	38.64014	394	0	0.09963
BPAF	-215.104	0.049432	-0.02587	9.143527	0.074292	19454.01	28.06496	11.81816	0	5.877849	-215.104	490	0.423762	0.094891
BPAP	17.678	0.027952	-0.01556	9.518637	-0.15158	20980.97	27.45943	2.875479	1.926533	3.44325	17.678	765	0	0.099016
BPB	33.17896	0.030889	-0.01742	9.373119	-0.14943	18047.44	23.94242	2.660639	2.066232	3.35696	33.17896	455	0	0.073988
BPBP	-38.6076	0.0252	-0.01298	9.841551	-0.01803	26844.84	37.28955	2.828914	0	3.240984	-38.6076	1422	0	0.110725
BPC	-20.405	0.028423	-0.01622	9.631567	-0.20972	19811.08	24.38479	2.765152	1.76062	3.558951	-20.405	618	0	0.07617
BPE	30.21124	0.036645	-0.02048	9.216467	-0.18358	14071.67	21.54427	2.858187	1.998216	3.716586	30.21124	388	0	0.093181
BPF	14.46939	0.040461	-0.02242	9.130205	-0.17826	12054.88	20.43091	2.868924	0	3.956481	14.46939	382	0	0.077888
BPFL	-57.7146	0.025933	-0.01326	9.851968	-0.02939	28441.32	37.71467	2.855598	0	3.298673	-57.7146	1299	0	0.105452

BPG	-129.379	0.022176	-0.01269	10.06602	-0.01625	25408.01	28.67284	2.627301	1.988786	2.991843	-129.379	1114	0	0.075846
BPPH	-10.7531	0.022342	-0.01163	10.42493	-0.06137	29059.89	39.07658	2.843682	2.064235	3.202908	-10.7531	1721	0	0.082486
BPS	12.161	0.045426	-0.02484	9.105592	-0.01389	13961.56	22.23695	4.829813	0	5.327829	12.161	382	0	0.122407
DOD	36.46746	0.052466	-0.02829	9.074649	-0.24783	10522.95	20.43091	3.853009	0	4.706197	36.46746	382	0	0.111291
MCBPA	27.8624	0.034099	-0.01906	9.379345	-0.21351	18664.02	24.38479	2.682122	1.928432	3.668157	27.8624	874	0.175655	0.092887
TCBPA	/	/	/	/	/	/	/	/	/	/	/	/	/	/
TBBPA	/	/	/	/	/	/	/	/	/	/	/	/	/	/
TBBPS	/	/	/	/	/	/	/	/	/	/	/	/	/	/
TMBPA	/	/	/	/	/	/	/	/	/	/	/	/	/	/