

Supplementary Materials

Table S1 p-values for the Chi-square test of differences in topology class distribution between the DrugAfter80 set and the other data sets.

	Clin. Cand	NP	DrugBefore80	HMDB	Bioact
DrugAfter80	0.2	<0.0001	<0.0001	<0.0001	<0.0001

Table S2 The p-values calculated with the non-parametric Wilcoxon rank-sum test for the statistical significance of the differences in RSC distributions.

		DrugAfter8	DrugBefore8	Clin.	HMDB	NP	Bioact
		0	0	Cand.			
DrugAfter8	1TR					<0.000	
0		-	0.5	0.25	0.27	1	0.25
	2TR					<0.000	
		-	0.12	0.49	0.39	1	0.02
	2TR+B_0						
		-	0.0004	0.9	0.05	0.0002	0.1
	2TR+B_1					<0.000	
		-	0.02	0.3	0.01	1	0.25
	3TR+B					<0.000	
		-	0.08	0.01	0.08	1	0.02
DrugBefore	1TR					<0.000	<0.000
80		-	-	<0.0001	0.495	1	1
	2TR					<0.000	<0.000
		-	-	<0.0001	0.06	1	1
	2TR+B_0					<0.000	<0.000
		-	-	<0.0001	0.11	1	1
	2TR+B_1					<0.000	<0.000
		-	-	<0.0001	0.82	1	1
	3TR+B					<0.000	<0.000
		-	-	<0.0001	0.37	1	1
Clin. Cand.	1TR					<0.000	
		-	-	-	<0.0001	1	0.51
	2TR					<0.000	<0.000
		-	-	-	0.001	1	1
	2TR+B_0					<0.000	<0.000
		-	-	-	0.002	<0.000	<0.000

						1	1
	2TR+B_1	-				<0.000	
			-	-	<0.0001	1	0.93
	3TR+B	-				<0.000	
			-	-	0.0007	1	0.28
HMDB	1TR					<0.000	<0.000
		-	-	-	-	1	1
	2TR					<0.000	<0.000
		-	-	-	-	1	1
	2TR+B_0					<0.000	<0.000
		-	-	-	-	1	1
	2TR+B_1					<0.000	<0.000
		-	-	-	-	1	1
	3TR+B					<0.000	
		-	-	-	-	1	0.0009
NP	1TR	-					<0.000
			-	-	-	-	1
	2TR						<0.000
		-	-	-	-	-	1
	2TR+B_0						<0.000
		-	-	-	-	-	1
	2TR+B_1						<0.000
		-	-	-	-	-	1
	3TR+B						<0.000
		-	-	-	-	-	1

Table S3 p-values calculated with the non-parametric Wilcoxon rank-sum test for the statistical significances of the differences in physiochemical properties between DrugAfter80 set and the other data sets. HAC is the Heavy Atom Count.

Compound set	Clin. Cand.	HMDB	Bioact.	DrugBefore80	NP
ClogP	<0.0001	<0.0001	<0.0001	0.34	0.34
HAC	0.05	<0.0001	<0.0001	<0.0001	0.61
MW	0.04	<0.0001	<0.0001	<0.0001	0.21

Table S4 p-values calculated with the non-parametric Wilcoxon rank-sum test for the statistical significances of the differences between DrugAfter80 and the other data sets for the topological descriptors.

Compound set	Clinical Candidates	HM	Bioactive compounds	DrugBefore80	NP
Fsp3	0.07	<0.0001	<0.0001	0.5	<0.0001
DFAAC	0.03	<0.0001	<0.0001	0.96	<0.0001
f_{MF}	0.002	<0.0001	<0.0001	0.33	<0.0001