Table S1. SMARTS strings accounting for chemical diversity

descriptor name	SMARTS	description
	hydrogen-bond donors	
m001	[OH1][#6,#7][!O]	hydroxyl, but not acids
m002	[N;H]	N-H
m003	[S;H1]	S-H
m004	not used	
	hydrogen-bond acceptors	
m005	A=O	C=O, S=O, P=O
m006	[#8;X2]	–O–
m007	[#7;X3]	NR_3
m008	A=S	C=S, N=S, P=S
m009	[#6,#7][#16;X2][#1,#6]	–S–
	hydrophobic moieties	
m010	c1aaaaa1	aromatic 6-membered ring, at least one C
m011	c1aaaa1	aromatic 5-membered ring, at least one C
m012	C1CCCCA1	aliphatic six-membered ring including one hetero atom
m013	C1~A~C~C~C1	aliphatic 5-membered ring
m014	C1~A~C~C1	aliphatic 4-membered ring
m015	C1CC1	cyclopropyl
m016	C(C)(C)C	tert-butyl
m017	C(C)CC	sec-butyl
m018	C(C)C	iso-propyl
m019	[CD2;R0][CD2;R0][CD2;R0][CD2;R0]	n-butyl chain
	deprotonable groups	
m020	C(=O)[O;H1]	carboxylic acids
m021	S(=O)(=O)[OH1,NH2]	sulfonic acids and sulfon amides

Table S1 (continued)

m022	[NX4]	ternary nitrogens
m023	[#6;R]1~[#7;D2]~[#7;D2]~[#7;D2]~[#7;D2]	tetrazol
m024	P(=O)(~O)[OH1]	phosphoric and phosphonic acid
	protonable groups	
m025	[#6][NX3]([#6,#1])[#6,#1]	C–NR_2 not hydroxylamine
m026	CN~C(~[NH])~[NH2]	guanidinium
m027	not used	
m028	not used	
m029	not used	
	ring systems	
m030	c1aaaaa1	any aromatic 6-membered ring
m031	c1aaaa1	any aromatic 5-membered ring
m032	A1~C~A~A~A~A1	any aliphatic 6-membered ring
m033	A1~C~A~A~A1	any aliphatic 5-membered ring
m034	A1~C~A~A1	any aliphatic 4-membered ring
m035	C1[NH0]C1	3-membered ring (not epoxy)
m036	A1~C~A~A~A~A~A1	any aliphatic 7-membered ring
m037	C1OC([CH2]O)~C~C1	pentose sugar
m038	C1OC(O[C,H])C~C~C1	ketose sugar
m039	[#6]1@[#6]@[#6]2@C3@C@C@C4([C,H])@C([O,C,H])@CC@C4@C3@C@C@[#6]2@[#6]@[#6]1	steroids (including aromatic ring A)
	special drug-like functional groups	
m040	[#6]N(~O)(~O)	nitro
m041	[#6]C#N	cyano
m042	[#6]O[CH2][CH3,H]	methoxy, ethoxy
m043	[#6]S[CH2][CH3,H]	sulfoxymethyl, sulfoxyethyl
m044	[#6]S(=O)(=O)[!O]	sulfonyl
m045	[#6]S(=O)(=O)[NH]	amidosulfon
m046	[#6]C(=O)O[#6]	ester

Table S1 (continued)

m047	[CX4;R0]([C,H])O[CX4;R0]([C,H])	aliphatic ether, not ring
m048	[#6]C(=O)[NH][C,H]	amido
m049	[#6,H][NH]C(=O)[NH][#6,H]	urea
m050	[NH]C(=S)[NH]	thiourea
m051	CC#C	aliphatic alkyne
	alkanes	
m052	[CX4H3][#6]	primary carbon
m053	[CX4H2]([#6])[#6]	secondary carbon
m054	[CX4H1]([#6])([#6])[#6]	tertiary carbon
m055	[CX4]([#6])([#6])([#6])[#6]	quartary carbon
m056	C=[C;!CX2,C;!OX1,C;!NH2,C;!OH1]	alkenes without ketenes, allenes, enamines, enols
m057	[CX2]#[CX2]	alkynes
m058	[CX3]=[CX2]=[CX3]	allenes
m059	[CX2]#[CX2]	polyines
m060	[CX4][F,Cl,Br,I]	alkylhalogenides
	alcohols	
m061	[CH2]([OH1])[CX4]	primary alcohols
m062	[CH1]([OH1])([CX4])[CX4]	secondary alcohols
m063	[C;H0]([OH1])([CX4])([CX4])[CX4]	tertiary alcohols
	ethers excluding esters	
m064	[C;R0]O[C;R0]([!O])[H,#6]	aliphatic ether, not ring
m065	[C;R]OC([!O])[H,#6]	aliphatic ether, one or two rings
m066	cO[C;R0]([!O])[H,#6]	alkyl-aryl ether
m067	cOc	diaryl ether
m068	[C;R0][SX2][C;R0]([!O])[H,#6]	aliphatic thio ether, not ring
m069	[C;R][SX2]C([!O])[H,#6]	thio ether, rings allowed
m070	c[SX2][C;R0]([!O])[H,#6]	alkyl-aryl thio ether
m071	c[SX2]c	diaryl thio ether

Table S1 (continued)

	esters	
m072	COC(=O)C	aliphatic ester
m073	cC(=O)OC	aryl-alkyl ester
m074	CC(=O)Oc	alkyl-aryl ester
m075	cC(=O)Oc	diaryl ester
m076	COC(=S)C	aliphatic thio ester
m077	C[SX2]C(=O)C	aliphatic thio ester
m078	cC(=S)OC	aryl-alkyl thio ester
m079	cC(=O)[SX2]C	aryl-alkyl thio ester
m080	cC(=S)Oc	diaryl thio ester
m081	cC(=O)[SX2]c	diaryl thio ester
	other functional groups	
m082	COC(=O)OC	carbonate, carbonyloxy
m083	[#6][NH1]C(=O)[#6]	peptide
m084	[#6][NH0]([#6])C(=O)[#6]	amide
m085	[#6,#1]N([#6,#1])C(=O)[N]([#6,#1])[#6]	carbamide, urea derivatives
m086	[#6,#1]N([#6,#1])C(=S)[N]([#6,#1])[#6]	thiourea derivatives
m087	[#6]N[NH1]C(=O)[NH1]N[#6]	carbazide
m088	[#6]OO[#6,#1]	hydroperoxides
m089	[#6]OO[#6]	peroxides
m090	[#6][CH1]=O	aldehydes
	amines, ammonium, and enamines	
m091	C[NH2]	primary aliphatic
m092	C[NH1]C	secondary aliphatic
m093	CN(C)C	tertiary aliphatic
m094	CN(C)(C)C	quartary aliphatic
m095	C[NH3]	aliphatic ammonium
m096	c[NH2]	primary aromatic
m097	c[NH1]c	secondary aromatic
m098	cN(c)c	tertiary aromatic
m099	c[NH1]C	mixed secondary

Table S1 (continued)

m100	cN(c)C	mixed tertiary
m101	cN(C)C	mixed tertiary
m102	CN(c)([#6])[#6]	mixed quartary
	other thio compounds	
m103	[CH2]([SX2;H1])[CX4]	primary thiols
m104	[CH1]([SX2;H1])([CX4])[CX4]	secondary thiols
m105	[C;H0]([SX2;H1])([CX4])([CX4])[CX4]	tertiary thiols
m106	[SX2;D2][SX2;D2]	disulfides
	sulfoxides, sulfonamides	
m107	[#6][SX3](=O)[#6]	sulfoxide
m108	[#6][SX4](=O)(=O)[#6]	sulfoxide
m109	[#6][SX4](=O)(=O)[#7]	sulfonamide
m110	[#6][SX4](=O)(=O)~O	sulfonic acid
	other nitrogen containing compounds	
m111	[#6;R](=O)[#8;R][#6;R][#6;R]	lactones
m112	[#6;R](=O)[NH1;R][#6;R][#6;R]	lactimes
m113	[NH1][CX4;H1]C(=O)[O,N]	alpha-amino acids
m114	[NX1]~[NX2]~[NX2,NX1]	azido
m115	[CX3](=[OX1])[NX2]~[NX2]~[NX1]	acylazide
m116	[NX2](=[OX1])[!#7;!#8]	nitroso, no nitrites or nitrosamines
m117	[NX2](=[OX1])[#7]	nitrosamines
m118	[#6][NX3]([!#8])[NX3;H1]([!#8])	hydrazine
m119	C=[NX2][NX3;H1][!#7;!#8]	hydrazone
m120	[#6][NX3;H1][OX2;H1]	hydrylamines
m121	[#7v5](=[OX1])([!O])[!O]	N-oxide not nitro
	phosphorus containing compounds	
m122	[#6][PX3;H1,H2]	phosphines
m123	[#6][P;H1,H2](=O)	phosphine-oxides

Table S1 (continued)

m124	[#6][P+,PX5]	phosphonium
m125	[#6][P;H1,H2]=[CX3]	phosphorylene
m126	[#6][PX4;H1](=O)([OX2,#6,#7])	phosphinic acid H_3PO_2 derivatives
m127	[#6,#7][PX4;H0](=O)([OX2,#6,#7])[OX2]	phosphonic acid H_3PO_3 derivatives
m128	[#6,#8][PX4;H0](=O)([OX2,#6,#7])[OX2]	phosphoric acid H_3PO_4 derivatives
m129	[#6,#8][PX4;H0](=O)([OX2,#6,#7,SX2])[OX2,#6,#7,SX2]	H_3PO_4 derivatives
m130	[#6,#8][PX4;H0](=S)([OX2,#6,#7,SX2])[OX2,#6,#7,SX2]	H_3PSO_3 derivatives
	silicon containing compounds	
m131	[SiX4]([#6])([#6])([#6])[#6]	quartary silane
m132	[#6][SiX4;H1,H2,H3]	SiH_nC_{4-n}
m133	[#6][SiX4][F,Cl,Br,I]	halides
m134	[!#6][SiX4;H0]([#6])([#6])[#6]	hetero-trialkyl
m135	[#6][SiX4]O[SiX4][#6]	Si–O–Si
m136	[SiX4]([!#6])([!#6])([!#6])[!#6]	silicic acid derivatives
	boron containing compounds	
m137	[BX3]([#6])([#6])[#6]	trialkyl borane
m138	[BX3]([!#6])([!#6])[#6,#7,#8]	boronic acids and derivatives
m139	[BH1,BH2,BH3]	borohydrides
m140	[BX4]	quartary boron
m141	[BH1]([B,C,N])([B,C,N])([B,C])[B,C]	carboranes
	special rings frequently found in drugs	
m142	C1CCNCC1	piperidine
m143	N1CCNCC1	piperazine
m144	N1CCOCC1	morpholino
m145	C1OCCOC1	1,4-dioxan
m146	c12aaaac1c3aaaac3A2	fluorene and derivatives
m147	c12aaaac1Ac3aaaac3A2	9,10-dihydroanthracene and derivatives
m148	c12aaaac1A~Ac3aaaac3A2	10,11-dihydro-5H-dibenzo[a,d]-cycloheptene and derivatives

Table S1 (continued)

m149	c12aaaac1A(c3aaaaa3)~[A,a]~[A,a]~[A,a]~[A,a]2	benzodiazepines
m150	c1aaaaa1[C,N](c1aaaaa1)~A~A	diphenyl derivatives
m151	c12aaaaa1[#6]~aa2	indene, benzofuran, indole derivatives
m152	c12aaaaa1aaa2	benzoxazole, benzothiooxazole
m153	c12aaaaa1[!#6]~[!#6]~[!#6]2	benzotriazines, etc.
m154	[OX2][#6]1@[#6]@[#6]@[#6]2CC3C4C~CCCC4(CCN3)[#6]2@[#6]1	opioid ring system
m155	[C;R5,R6]([O;R5,R6][C;R5,R6])([C;R5,R6]~[C;R5,R6])O[C;R5,R6]([CH2]O)([C;R5,R6][O;R5,R6])[C;R5,R6]~[C;R5,R6]	1,4-diglycosides
m156	[C;R5,R6]([O;R5,R6][C;R5,R6][C;R5,R6][CH2]O)([C;R5,R6]~[C;R5,R6])O[CH2]([O;R5,R6][C;R5,R6])[C;R5,R6]~[C;R5,R6]	1,6-diglycosides
m157	[CX4,NH1]C1C2[SX2,CX4]~[C,CC]([CX4,SX2])~C(C(~O)(~O))N2C(=O)1	penicillins, carbapenems, etc.
m158	c12ccccc1[CX4]C3[CX4]C4C([NX3])C([OX2])=C(C(=O)[NX3])C(=O)C4([OH1])C([OH1])=C3C(=O)2	tetracyclines
m159	[A;r3]	any 3-membered ring
m160	[A;r4]	any 4-membered ring
m161	[A;r7]	any 7-membered ring
m162	[A;r8]	any 8-membered ring
m163	[A;r9]	any 9-membered ring
m164	[A;r10]	any 10-membered ring
m165	C1O[C,CC]OC1	cyclic acetales
m166	c1O[CX4]Oc1	aromatic dioxolo
	common amino derivatives	
m167	N(C[CH3])C[CH3]	diethylamino
m168	N([CH3])[CH3]	dimethylamino
	bridge head atoms	
m169	[C;R]([A;R])([A;R])([A;R])	tertiary bridge head carbon
m170	[N;R]([A;R])([A;R])([A;R])	tertiary bridge head nitrogen

Table S1 (continued)

	trihalogen groups	
m171	C(F)(F)F	trifluoromethyl
m172	C(Cl)(Cl)Cl	trichloromethyl

Table S2. SMARTS strings for reactive or unsuitable substructures compiled from ref.[1-5]

SMARTS string	description
[Br,Cl,I][CX4;CH,CH2]	alkyl halides
[S,C](=[O,S])[F,Br,Cl,I]	acid halides
C(=O)N=N=N	carbazides
COS(=O)O[C,c]	sulphate esters
COS(=O)(=O)[C,c]	sulphonates
C(=O)OC(=O)	acid anhydrides
OO	peroxides
C(=O)Oc1c(F)c(F)c(F)c(F)c1(F)	pentafluorophenyl esters
C(=O)Oc1ccc(N(=O)=O)cc1	paranitrophenyl esters
C(=O)Onnn	esters of HOBT
N=C=[S,O]	isocyanates & isothiocyanates
OS(=O)(=O)C(F)(F)F	triflates
P(=S)(S)S	Lawesson's reagent & derivatives
NP(=O)(N)N	phosphoramides
cN=N=N	aromatic azides
C(=O)C[N+,n+,NX4,nX4]	beta-carbonyl quaternary nitrogen
[N;R0][N;R0]C(=O)	acylhydrazine
[C+,Cl+,I+,P+,S+]	quaternary C, Cl, I, P, or S
C=P	phosphoranes
[Cl]C([C&R0])=N	chloramidines
[N&D2](=O)	nitroso
[P,S][Cl,Br,F,I]	P or S halides
N=C=N	carbodiimides
[N+]#[C-]	isonitriles
C(=O)N(C(=O))OC(=O)	triacyloximes
N#CC[OH]	cyanohydrins
N#CC(=O)	acyl cyanides
S(=O)(=O)C#N	sulfonyl cyanides
P(OCC)(OCC)(=O)C#N	cyanophosphonates
[N;R0]=[N;R0]C#N	azocyanamides

Table S2 (continued)

[N;R0]=[N;R0]CC=O	azoalkanals
[CD2;R0][CD2;R0][CD2;R0][CD2;R0][CD2;R0][CD2;R0][CD2;R0]	aliphatic unbranched chain > 6
[O;R1][C;R1][C;R1][O;R1][C;R1][C;R1][O;R1]	crown ethers
[SX2][SX2]	disulfides
[!a][SX2;H1]	thioles (not aromatic)
C1[O,S,N]C1	epoxides, thioepoxides, aziridines
c([OH])c([OH])c([OH])	2,3,4 trihydroxyphenyl
c([OH])c([OH])cc([OH])	2,3,5 trihydroxyphenyl
N=NC(S)N	hydrazothiourea
SC#N	thiocyanate
cC[N+,NX4]	benzylic quaternary nitrogen
C[O,S;R0][C;R0](=S)	thioesters
N[CH2]C#N	cyanamides
C1(=O)OCC1	four-membered lactones
P(=O)([OH])OP(=O)[OH]	di and triphosphates
N1CCC1=O	beta-lactams
[#6]P(=O)(~O)O[#6]	phosphonate esters
[#6][C;R0](=[N;R0])[#6]	imines (not ring)
[C;H1]=O	aldehydes
[#6]C(=O)C=C([H1])[#6]	Michael acceptors
C(=O)CC(C)[N;R0,O;R0]	beta-hetero substituted carbonyl
[#6]N-N[#6]	N-N
[#6]N-O[#6]	N-O
[#6][S;O0][N;H0]	N-S (not sulfonamides)
[S;R0][O;R0]	S-O (not ring)
P(c1aaaaa1)(c1aaaaa1)(c1aaaaa1)	triphenylphosphines
cN=Nc	diazonium
cC=CC=CC=Cc	polyene chain between aromatics
c1c2cccc3c2c4c(cc3)cccc4c1	pyrene fragments
C(=[O,S])[S,CF,CBr,CCl]	reactive carbonyls and sulfonyls
[S;R0]=[N;R0]	S=N (not ring)

Table S2 (continued)

NC(=S)N	thiourea
PCP	P–C–P
N#CCC#N	geminal dinitriles
[N;R0]~[N;R0]~[N;R0]	N~N~N (not ring)
[NH1]C[NH1]	geminal amines
C1=CC(=O)C=CO1	gamma-pyranone
c1ccccc1C2C=CN([H1])C=C2	1,4-dihydro-4-phenylpyridine
C1(C)(C)SC(C(=O)O)=C(C(=O)O)S1	2,2-dimethyl-4,5-dicarboxydithiole
c1c([OH1])c(C(C)(C)C)ccc1	o-tertbutylphenol
c1c([F,Cl])c([F,Cl])c([F,Cl])c([F,Cl])c1([F,Cl])	pentahalophenyl
N1=C[S,NH1]C(=[C,N,P][C,N,O,P])C1(=O)	1,3-thiazol/imidazol-4-one derivatives
CC#CC#CC	polyines
[Cl,Br,I]c1[c,n][c,n][c,n][c,n]n1	halodiazoles and triazoles
[NH2]cc[NH2]	o-diaminophenyl
[NH2]cac[NH2]	m-diaminophenyl
[NH2]caac[NH2]	p-diaminophenyl
O=C1[#6]~[#6]C(=O)[#6]~[#6]1	quinones
C=CC=CC=CC=C	polyenes
O1CCCCC1OC2CCC3CCCCC3C2	saponin derivatives
O=C1NCC2CCCCC21	cytochalasin derivatives
O=C1CCCC(N1)=O	cycloheximide derivatives
O1CCCCC1C2CCCO2	monensin derivatives
C12OCCC(O1)CC2	squalestatin derivatives

Table S3. Full list of computed descriptors

descriptor	description
MW	molecular weight
XlogP	calculated XlogP[a, 6]
SlogP	calculated SlogP[a, 7]
MR	molar refractivity[a, 7]
HBDON	number of hydrogen bond-donors
HBACC	number of hydrogen bond-acceptors
HALO	number of all halogen atoms
COOH	number of carboxylic acid groups
NOH	number of OH-groups
NH2	number of NH_2-groups
NO2	number of NO_2-groups
NNR3	number of protonatable NR_3-groups
SO1	number of S=O-groups
SO2	number of SO_2-groups
SO3	number of SO_3-groups
NACID	number of deprotonable acidic groups
CN	number of CN-groups
CF3	number of CF_3-groups
CCL3	number of CCl_3-groups
ESTER	number of ester-groups
FUNCGR	number of functional groups (NOH, ESTER, NO2, SO1, SO2, SO3, CN, CF3, CCL3, NNR3, NACID)
UNSUIT	number of unsuitable groups (according to ref.[5] also including occurences of SMARTS given in Table S2)
NRING	total number of rings
NR3	number of 3-membered rings
NR4	number of 4-membered rings
NR5	number of 5-membered rings
NR6	number of 6-membered rings
AR5	number of aromatic 5-membered rings

Table S3 (continued)

AR6	number of aromatic 6-membered rings
NROTB	number of rotatable bonds
LIPIA	number of violations of Lipinki's rule[b, 8]
GVW80	compound qualified within 80% range of drugs[b, 9]
GVW50	compound in prefered range of 50% of drugs[b, 9]
OPREA	Oprea's critria for 70% of drugs[b, 2]
HUTTER	total Hutter's drug-likeliness index[10]
MOLVOL	molecular van der Waals volume
GLOB	globularity[11]
DIPM	molecular dipole moment
MPOLAR	molecular polarizability[12]
IP	ionization potential
DIPDENS	dipolar density[13]
EHOMO	energy of highest occupied molecular orbital
ELUMO	energy of lowest unoccupied molecular orbital
DEHL	ELUMO – EHOMO
CHBAC	covalent hydrogen bond acidity[14]
CHBBA	covalent hydrogen bond basicity[14]
EHBAC	electrostatic hydrogen bond acidity[14]
EHBBA	electrostatic hydrogen bond basicity[14]
ESPMIN	minimum of molecular electrostatic potential (MEP)[15]
ESPMAX	maximum of MEP[15]
MESP	median of MEP[15]
M+ESP	median of positive MEP[15]
M–ESP	median of negative MEP[15]
V+	variance of positive MEP[15]
V–	variance of negative MEP[15]
VTOT	total variance of MEP[15]
BALESP	balance parameter of the MEP[15]
VXBAL	VTOT · BALESP[15]
QSUMH	sum of MEP derived atom centered point charges on hydrogen atoms[16]

Table S3 (continued)

QSUMN	sum of MEP derived atom centered point charges on nitrogen atoms[16]
QSUMO	sum of MEP derived atom centered point charges on oxygen atoms[16]
QSUMS	sum of MEP derived atom centered point charges on sulfur atoms[16]
QSUM–	sum of all MEP derived negative atom centered point charges[16]
QSUM+	sum of all MEP derived positive atom centered point charges[16]
PSA	polar surface area[17,18]
FPSA	fraction of polar surface from the total surface area
PSANEG	negative polar surface area (electrostatic potential < -25 kcal/mol)
PSAPOS	positive polar surface area (electrostatic potential > 10 kcal/mol)
DPSA1	| PSAPOS – PSANEG |
FPSA1	ratio of positively charged surface to the total surface[19]
FNSA1	ratio of negatively charged surface to the total surface[19]
HRSURF	ratio of surface on hydrogen atoms to the total surface area
CRSURF	ratio of surface on carbon atoms to the total surface area
HLSURF	ratio of surface on halogen atoms to the total surface area
HACSUR	ratio of the surface area belonging to atoms that are hydrogen bond acceptors to the total surface area
KHECA	Kier & Hall E-state on carbon atoms[20]
KHENI	Kier & Hall E-state on nitrogen atoms[20]
KHEOX	Kier & Hall E-state on oxygen atoms[20]
KHESU	Kier & Hall E-state on sulfur atoms[20]
KHEPH	Kier & Hall E-state on phosphorus atoms[20]
KHEHA	Kier & Hall E-state on halogen atoms[20]
SGECA	geometric E-state on carbon atoms[21]
SGENI	geometric E-state on nitrogen atoms[21]
SGEOX	geometric E-state on oxygen atoms[21]
SEGSU	geometric E-state on sulfur atoms[21]
SGEPH	geometric E-state on phosphorus atoms[21]
SGEHA	geometric E-state on halogen atoms[21]

Table S3 (continued)

G1TOP	gravitational G1 topological index[22]
G1GEO	gravitational G1 geometrical index[22]
T1E	topological electronic index 1[23]
T2E	topological electronic index 2[23]
MPHBD	minimal path length between two hydrogen bond donors
MPHBA	minimal path length between two hydrogen bond acceptors
MPHAC	minimal path length between a hydrogen bond donor and an acceptor
MGHBD	minimal distance between two hydrogen bond donors
MGHBA	minimal distance between two hydrogen bond acceptors
MGHAC	minimal distance between a hydrogen bond donor and an acceptor
RUGOS	rugosity (= molecular surface / MOLVOL)[24]
COHESI	cohesive index[25]

---

[a] Additional parameters required for boron and silicon containing compounds developped in our laboratory are given in Table S4.

[b] Using the computed XlogP values.

Table S4. Additional parameters used for the computation of XlogP, SlogP and molar refractivity (MR). According experimental data were taken from Hansch et al.[26] and the CRC Handbook of Chemistry and Physics.[27]

atom type	description	XlogP	SlogP	MR
C	A...C(H)...O	–0.310	n.a	n.a.
C	A…C(H)…S	0.136	n.a.	n.a.
C	A…C(R)…O (A≠H, R≠O,N,H)	–0.027	n.a.	n.a.
C	A…C(R)…S	0.013	n.a.	n.a.
C	A…C(X)…O (X = O,N)	–0.315	n.a.	n.a.
C	A…C(X)…S	0.027	n.a.	n.a.
C	$CA_{\geq 5}$ (hypervalent)	0.0	n.a.	n.a.
C	default	0.0	0.0	2.60
Ar	ArA	–0.987	0.0	0.0
N	C=N–H (guanidinium fragment)	–1.453	n.a.	n.a.
N	R...N...R	–0.275	n.a.	n.a.
N	R...N...X	0.251	n.a.	n.a.
N	$NR_{>3}$ (ammonium) (R≠O)	–5.442	n.a.	n.a.
N	A…N(=O)…A (N-oxide)	–1.44	n.a.	n.a.
N	AN(=O)A	0.1	n.a.	n.a.
N	default	n.a.	n.a.	3.00
O	R…O…R	0.83	n.a.	n.a.
O	R…O…X	0.0	n.a.	n.a.
O	X…O…X	0.0	n.a.	n.a.
O	$OA_{\geq 3}$ (hypervalent)	–4.0	n.a.	n.a.
O	default	0.0	0.0	1.50
S	C…S…C	0.73	n.a.	n.a.
S	SA_3 (hypervalent)	–4.0	n.a.	n.a.
S	default	0.0	0.0	7.80
P	PA_3	–0.98	n.a.	n.a.
P	default	0.0	0.0	5.50
Si	SiR_4 (R≠O, R≠H)	0 .787	0.0	7.24
Si	$SiOR_3$	–1.24	0.0	4.00

Table S4 (continued)

Si	default	0.0	0.0	4.00
B	BA_3	0.533	0.0	2.86
B	default	0.0	0.0	2.00
H	default	0.0	0.0	0.80
F	default	0.0	0.0	1.00
Cl	default	0.0	0.0	6.00
Br	default	0.0	0.0	8.60
I	default	0.0	0.0.	14.00

Designation of atoms corresponds to those used in the original literatur, i.e. X: any nitrogen or oxygen atom; R: any other atom; Ar: aromatic atom; A: any atom; –: single bond; =: double bond; ...: aromatic bond; n.a. = not applicable

References for Tables S1 – S4

(1) Hann, M.; Hudson, B.; Lewell, X.; Lifely, R.; Miller, L.; Ramsden, N. Strategic Pooling of Compounds for High-Throughput Screening. *J. Chem. Inf. Comput. Sci.* **1999**, *39,* 897-902.

(2) Oprea, T. I. Property distribution of drug-related chemical databases. *J. Comput.-Aided Mol. Des.* **2000**, *14,* 251-264.

(3) Rishton, G. M. Reactive Compounds and in vitro False Positives in HTS. *Drug Discov. Today* **1997**, *2,* 384-386.

(4) Anzali, S.; Barnickel, G.; Cezanne, B.; Krug, M.; Filimonov, D.; Poroikov, V. Discriminating between Drugs and Nondrugs by Prediction of Activity Spectra for Substances (PASS). *J. Med. Chem.* **2001**, *44,* 2432-2437.

(5) Flower, D. R. DISSIM: A program for the analysis of chemical diversity. *J. Mol. Graph. Model.* **1998**, *16,* 239-253.

(6) Wang, R.; Gao, Y.; Lai, L. Calculating Partition Coefficient by Atom-Additive Method. *Persp. Drug Discov. Des.* **2000**, *19,* 47-66.

(7) Viswanadhan, V. N.; Ghose, A. K.; Rebankar, G. R.; Robins, R. K. Atomic Physicochemical Parameters for Three Dimensional Structure Directed Quantitative Structure-Activity Relationships. 4. Additional Parameters for Hydrophobic and Dispersive Interactions and Their Application for an Automated Superposition of Certain Naturally Occuring Nucleoside Antibiotics. *J. Chem. Inf. Comput. Sci.* **1989**, *29,* 163-172.

(8) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **1997**, *23,* 3-25.

(9) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. A Knowledge-Based Approach in Designing Combinatorial or Medicinal Chemistry Libraries for Drug Discovery. 1. A Qualitative and Quantitative Characterization of Known Drug Databases. *J. Comb. Chem.* **1999**, *1,* 55-68.

(10) Hutter, M. C. Separating Drugs from Nondrugs: A Statistical Approach Using Atom Pair Distributions. *J. Chem. Inf. Model.* **2007**, *47,* 186-194.

(11) Meyer, A. Y. The Size of Molecules. *Chem. Soc. Rev.* **1986**, *15,* 449-475.

(12) Rinaldi, D.; Rivail, J. L. Calculation of molecular polarizabilities. Comparison of different methods. *Theor. Chim. Acta* **1974**, *32,* 243-251.

(13) Mu, L.; Drago, R. S.; Richardson, D. E. A Model Based QSPR Analysis of the Unified Non-Specific Solvent Polarity Scale. *J. Chem. Soc., Perkin Trans. 2* **1998**, 159-167.

(14) Cronce, D. T.; Famini, G. R.; De Soto, J. A.; Wilson, L. Y. Using Theoretical Descriptors in Quantitative Structure-Property Relationships: Some Distribution Equilibria. *J. Chem. Soc., Perkin Trans. 2* **1998**, 1293-1301.

(15) Breindl, A.; Beck, B.; Clark, T.; Glen, R. C. Prediction of the n-Octanol/Water Partition Coefficient, logP, Using a Combination of Semiempirical MO-Calculations and a Neuronal Network. *J. Mol. Model.* **1997**, *3,* 142-155.

(16) Beck, B.; Glen, R. C.; Clark, T. VESPA: A New, Fast Approach to Electrostatic Potential Derived Atomic Charges from Semiempirical Methods. *J. Comput. Chem.* **1997**, *18,* 744-756.

(17) Ertl, P.; Rohde, B.; Selzer, P. Fast Calculation of Molecular Surface Area as a Sum of Fragment Based Contributions and its Application to the Prediction of Drug Transport Properties. *J. Med. Chem.* **2000**, *43,* 3714-3717.

(18) Kelder, J.; Grootenhuis, P. D. J.; Bayada, D. M.; Delbressine, L. P. C.; Ploemen, J.-P. Polar Molecular Surface as a Dominating Determinant for Oral Absorption and Brain Penetration of Drugs. *Pharm. Res.* **1999***, 16,* 1514-5119.

(19) Stanton, D. T.; Egolf, L. M.; Jurs, P. C.; Hicks, M. G. *J. Chem. Inf. Comput. Sci.* **1992***, 32,* 306.

(20) Kier, L. B.; Hall, L. H. *Molecular Structure Description: The Electrotopological State*; Academic Press: San Diego, 1999.

(21) Brüstle, M.; Beck, B.; Schindler, T.; King, W.; Mitchell, T.; Clark, T. Descriptors, Physical Properties, and Drug-Likeness. *J. Med. Chem.* **2002***, 45,* 3345-3355.

(22) Katritzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M. Correlation of Boiling Points with Molecular Structure. 1. A Training Set of 298 Diverse Organics and a Test Set of 9 Simple Inorganics. *J. Phys. Chem.* **1996***, 100,* 10400-10407.

(23) Osmialowski, K.; Halkiewicz, J.; Kaliszan, R. Quantum Chemical Parameters in Correlation Analysis of Gas-Liquid Chromatographic Retention Indices of Amines. II. Topological Electronic Index. *J. Chromatogr. A* **1986***, 361,* 63-69.

(24) Meyer, Y. A. Molecular Mechanics and Molecular Shape. V. On the Computation of the Bare Surface Area of Molecules. *J. Comput. Chem.* **1988***, 9,* 18-24.

(25) Jorgensen, W. L.; Duffy, E. M. Predicting Drug Solubility from Monte Carlo Simulations. *Bioorg. Med. Chem. Lett.* **2000***, 10,* 1155-1158.

(26) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR 2. Hydrophobic, Electronic, and Steric Constants*; American Chemical Society: Washington, DC, 1995.

(27) *CRC Handbook of Chemistry and Physics*; 86 ed.; Taylor and Francis Group: Boca Raton, FL, 2005.