TIN-a combinatorial compound collection of synthetically feasible multicomponent synthesis products.

AUTHOR(S)

Kristl V. Dorschner, David Toomey, Marian P. Brennan, Tim Heinemann, Fergal J. Duffy, Kevin B. Nolan, Dermot Cox, Mauro FA Adamo, Anthony J. Chubb

Supplementary Material - Dorschner, KV *et al.,* 2011

Table S1: MOE Descriptors calculated for each TIN molecule.

Descriptor Code	Type of Descriptor	Description
E	Internal energy	Value of the potential energy
weight	Physical properties	Molecular weight including implicit hydrogens
logP(o/w)	Physical properties	Log of the octanol/water partition coefficient including implicit hydrogens
lip_acc	Atom count and Bond count	The number of O and N atoms
lip_don	Atom count and Bond count	The number of OH and NH atoms
lip_violation	Atom count and Bond count	The number of violations of Lipinski's 'Rule of Five'.
lip_druglike	Atom count and Bond count	Displays 1 if violations of lip_violation < 2; otherwise 0
FCharge	Physical properties	Total sum of formal charges of molecule
a_acc	Pharmacophore atom type (descriptor considers only the heavy atoms of a molecule and may take into account implied protonation, deprotonation, keto/enol considerations and tautomerism at a biologically relevant pH)	Number of hydrogen bond acceptor atoms not including acidic atoms
a_don	Pharmacophore atom type	Number of hydrogen bond donor atoms not including basic atoms
a_acid	Pharmacophore atom type	Number of acidic atoms
a_base	Pharmacophore atom type	Number of basic atoms.
a_hyd	Pharmacophore atom type	Number of hydrophobic atoms.
chiral	Atom count and Bond count	Number of chiral centers
chiral_u	Atom count and Bond count	Number of unconstrained chiral centers
rings	Atom count and Bond count	Number of rings
logS	Physical properties	Log of the aqueous solubility
opr_brigid	Atom count and Bond count	The number of rigid bonds.
opr_leadlike	Atom count and Bond count	Displays 1 if violations of opr_violation < 2; otherwise 0
opr_nring	Atom count and Bond count	Number of rings bonds.
opr_violation	Atom count and Bond count	Number of violations of Oprea's lead-like criteria.

Descriptor Code	Type of Descriptor	Description
reactive	Physical properties	Indicator of the presence of reactive groups. Value > 0 indicates molecule contains a reactive group
a_aro	Atom count and Bond count	Number of aromatic atoms
a_count	Atom count and Bond count	Number of atoms including implicit hydrogens
a_heavy	Atom count and Bond count	Number of heavy atoms
a_ICM	Atom information content	Entropy of the element distribution in the molecule including implicit hydrogens (but not lone pair pseudo-atoms)
a_IC	Atom information content	Total atom information content = (a_ICM) x (n)
a_nH	Atom count and Bond count	Number of hydrogen atoms including implicit hydrogens
a_nB	Atom count and Bond count	Number of boron atoms
a_nC	Atom count and Bond count	Number of carbon atoms:
a_nN	Atom count and Bond count	Number of nitrogen atoms
a_nO	Atom count and Bond count	Number of oxygen atoms:
a_nF	Atom count and Bond count	Number of fluorine atoms:
a_nP	Atom count and Bond count	Number of phosphorus atoms
a_nS	Atom count and Bond count	Number of sulfur atoms
a_nCl	Atom count and Bond count	Number of chlorine atoms
a_nBr	Atom count and Bond count	Number of bromine atoms
a_nI	Atom count and Bond count	Number of iodine atoms
b_1rotR	Atom count and Bond count	Number of rotatable single bond not including conjugated single bonds
b_1rotN	Atom count and Bond count	Fraction of rotatable single bonds = b_1rotN divided by b_heavy.
b_ar	Atom count and Bond count	Number of aromatic bonds
b_count	Atom count and Bond count	Number of bonds including implicit hydrogens
b_double	Atom count and Bond count	Number of double bonds (not including aromatic bonds)

Table S2: Structural Fingerprints calculated for TIN and ZINC subsets.

Bit Position	Description	SMARTS pattern	ZINC clustered at 60% tanimoto similarity	Diversity subset of TIN	Fingerprint in ZINC	Fingerprint in TIN
0	Primary carbon	[CX4H3][#6]	3769	10911	1	1
1	Secondary carbon	[CX4H2]([#6])[#6]	3547	10845	1	1
2	Tertiary carbon	[CX4H1]([#6])([#6])[#6]	961	10450	1	1
3	Quaternary carbon	[CX4]([#6])([#6])([#6])[#6]	290	3075	1	1
4	Alkene	[CX3;$([H2]),$([H1][#6]),$(C([#6])[#6])]=[CX3;$([H2]),$([H1][#6]),$(C([#6])[#6])]	1680	3393	1	1
5	Alkyne	[CX2]#[CX2]	118	1139	1	1
6	Allene	[CX3]=[CX2]=[CX3]	13	0	1	
7	Alkylchloride	[ClX1][CX4]	126	746	1	1
8	Alkylfluoride	[FX1][CX4]	137	5100	1	1
9	Alkylbromide	[BrX1][CX4]	28	655	1	1
10	Alkyliodide	[IX1][CX4]	2	111	1	1
11	Alcohol	[OX2H][CX4;!$(C([OX2H])[O,S,#7,#15])]	384	1600	1	1
12	Primary alcohol	[OX2H][CX4H2;!$(C([OX2H])[O,S,#7,#15])]	206	814	1	1
13	Secondary alcohol	[OX2H][CX4H;!$(C([OX2H])[O,S,#7,#15])]	219	593	1	1
14	Tertiary alcohol	[OX2H][CX4D4;!$(C([OX2H])[O,S,#7,#15])]	44	353	1	1
15	Dialkylether	[OX2]([CX4;!$(C([OX2])[O,S,#7,#15,F,Cl,Br,I])])[CX4;!$(C([OX2])[O,S,#7,#15])]	343	636	1	1
16	Dialkylthioether	[SX2]([CX4;!$(C([OX2])[O,S,#7,#15,F,Cl,Br,I])])[CX4;!$(C([OX2])[O,S,#7,#15])]	288	114	1	1
17	Alkylarylether	[OX2](c)[CX4;!$(C([OX2])[O,S,#7,#15,F,Cl,Br,I])]	320	3817	1	1
18	Diarylether	[c][OX2][c]	47	608	1	1
19	Alkylarylthioether	[SX2](c)[CX4;!$(C([OX2])[O,S,#7,#15,F,Cl,Br,I])]	263	569	1	1
20	Diarylthioether	[c][SX2][c]	85	9	1	1
21	Oxonium	[O+;!$([O]~[!#6]);!$([S]*~[#7,#8,#15,#16])]	1	0	1	
22	Amine	[NX3+0,NX4+;!$([N]~[!#6]);!$([N]*~[#7,#8,#15,#16])]	184	1395	1	1
23	Primary aliph amine	[NX3H2+0,NX4H3+;!$([N][!C]);!$([N]*~[#7,#8,#15,#16])]	0	0		
24	Secondary aliph amine	[NX3H1+0,NX4H2+;!$([N][!C]);!$([N]*~[#7,#8,#15,#16])]	0	0		
25	Tertiary aliph amine	[NX3H0+0,NX4H1+;!$([N][!C]);!$([N]*~[#7,#8,#15,#16])]	123	169	1	1
26	Quaternary aliph ammonium	[NX4H0+;!$([N][!C]);!$([N]*~[#7,#8,#15,#16])]	14	12	1	1
27	Primary arom amine	[NX3H2+0,NX4H3+]c	496	465	1	1
28	Secondary arom amine	[NX3H1+0,NX4H2+;!$([N][!c]);!$([N]*~[#7,#8,#15,#16])]	0	0		
29	Tertiary arom amine	[NX3H0+0,NX4H1+;!$([N][!c]);!$([N]*~[#7,#8,#15,#16])]	0	0		
30	Quaternary arom ammonium	[NX4H0+;!$([N][!c]);!$([N]*~[#7,#8,#15,#16])]	0	0		

Bit Position	Description	SMARTS pattern	ZINC clustered at 60% tanimoto similarity	Diversity subset of TIN	Fingerprint in ZINC	Fingerprint in TIN
31	Secondary mixed amine	[NX3H1+0,NX4H2+;$([N]([c])[C]);!$([N]*~[#7,#8,#15,#16])]	137	78	1	1
32	Tertiary mixed amine	[NX3H0+0,NX4H1+;$([N]([c])([C])[#6]);!$([N]*~[#7,#8,#15,#16])]	46	1215	1	1
33	Quaternary mixed ammonium	[NX4H0+;$([N]([c])([C])[#6][#6]);!$([N]*~[#7,#8,#15,#16])]	1	0	1	
34	Ammonium	[N+;!$([N]~[!#6]);!$(N=*);!$([N]*~[#7,#8,#15,#16])]	16	12	1	1
35	Alkylthiol	[SX2H][CX4;!$(C([SX2H])~[O,S,#7,#15])]	9	138	1	1
36	Dialkylthioether	[SX2]([CX4;!$(C([SX2])[O,S,#7,#15,F,Cl,Br,I])])[CX4;!$(C([SX2])[O,S,#7,#15])]	180	114	1	1
37	Alkylarylthioether	[SX2](c)[CX4;!$(C([SX2])[O,S,#7,#15])]	238	569	1	1
38	Disulfide	[SX2D2][SX2D2]	95	57	1	1
39	1,2-Aminoalcohol	[OX2H][CX4;!$(C([OX2H])[O,S,#7,#15,F,Cl,Br,I])][CX4;!$(C([N])[O,S,#7,#15])][NX3;!$(NC=[O,S,N])]	31	0	1	
40	1,2-Diol	[OX2H][CX4;!$(C([OX2H])[O,S,#7,#15])][CX4;!$(C([OX2H])[O,S,#7,#15])][OX2H]	71	161	1	1
41	1,1-Diol	[OX2H][CX4;!$(C([OX2H])([OX2H])[O,S,#7,#15])][OX2H]	1	16	1	1
42	Hydroperoxide	[OX2H][OX2]	0	0		
43	Peroxo	[OX2D2][OX2D2]	24	1565	1	1
44	Organolithium compounds	[LiX1][#6,#14]	0	0		
45	Organomagnesium compounds	[MgX2][#6,#14]	0	0		
46	Organometallic compounds	[!#1;!#5;!#6;!#7;!#8;!#9;!#14;!#15;!#16;!#17;!#33;!#34;!#35;!#52;!#53;!#85]~[#6;!-]	0	0		
47	Aldehyde	[$([CX3H][#6]),$([CX3H2])]=[OX1]	77	187	1	1
48	Ketone	[#6][CX3](=[OX1])[#6]	1081	2836	1	1
49	Thioaldehyde	[$([CX3H][#6]),$([CX3H2])]=[SX1]	2	0	1	
50	Thioketone	[#6][CX3](=[SX1])[#6]	11	0	1	
51	Imine	[NX2;$([N][#6]),$([NH]);!$([N][CX3]=[#7,#8,#15,#16])]=[CX3;$([CH2]),$([CH][#6]),$([C]([#6])[#6])]	550	0	1	
52	Immonium	[NX3+;!$([N][!#6]);!$([N][CX3]=[#7,#8,#15,#16])]	127	0	1	
53	Oxime	[NX2](=[CX3;$([CH2]),$([CH][#6]),$([C]([#6])[#6])])[OX2H]	45	0	1	
54	Oximether	[NX2](=[CX3;$([CH2]),$([CH][#6]),$([C]([#6])[#6])])[OX2][#6;!$(C=[#7,#8])]	58	0	1	
55	Acetal	[OX2]([#6;!$(C=[O,S,N])])[CX4;!$(C(O)(O)[!#6])][OX2][#6;!$(C=[O,S,N])]	44	14	1	1
56	Hemiacetal	[OX2H][CX4;!$(C(O)(O)[!#6])][OX2][#6;!$(C=[O,S,N])]	8	0	1	
57	Aminal	[NX3v3;!$(NC=[#7,#8,#15,#16])]([#6])[CX4;!$(C(N)(N)[!#6])][NX3v3;!$(NC=[#7,#8,#15,#16])][#6]	3	0	1	

Bit Position	Description	SMARTS pattern	ZINC clustered at 60% tanimoto similarity	Diversity subset of TIN	Fingerprint in ZINC	Fingerprint in TIN
58	Hemiaminal	[NX3v3;!$(NC=[#7,#8,#15,#16])]([#6])[CX4;!$(C(N)(N)[!#6])][OX2H]	26	0	1	
59	Thioacetal	[SX2]([#6;!$(C=[O,S,N])])[CX4;!$(C(S)(S)[!#6])][SX2][#6;!$(C=[O,S,N])]	4	0	1	
60	Thiohemiacetal	[SX2]([#6;!$(C=[O,S,N])])[CX4;!$(C(S)(S)[!#6])][OX2H]	5	0	1	
61	Halogen acetal like	[NX3v3,SX2,OX2;!$(*C=[#7,#8,#15,#16])][CX4;!$(C([N,S,O])([N,S,O])[!#6])][FX1,ClX1,BrX1,IX1]	22	255	1	1
62	Acetal like	[NX3v3,SX2,OX2;!$(*C=[#7,#8,#15,#16])][CX4;!$(C([N,S,O])([N,S,O])[!#6])][FX1,ClX1,BrX1,IX1,NX3v3,SX2,OX2;!$(*C=[#7,#8,#15,#16])]	146	285	1	1
63	Halogenmethylen ester and similar	[NX3v3,SX2,OX2;$(**=[#7,#8,#15,#16])][CX4;!$(C([N,S,O])([N,S,O])[!#6])][FX1,ClX1,BrX1,IX1]	3	0	1	
64	NOS methylen ester and similar	[NX3v3,SX2,OX2;$(**=[#7,#8,#15,#16])][CX4;!$(C([N,S,O])([N,S,O])[!#6])][NX3v3,SX2,OX2;!$(*C=[#7,#8,#15,#16])]	91	0	1	
65	Hetero methylen ester and similar	[NX3v3,SX2,OX2;$(**=[#7,#8,#15,#16])][CX4;!$(C([N,S,O])([N,S,O])[!#6])][FX1,ClX1,BrX1,IX1,NX3v3,SX2,OX2;!$(*C=[#7,#8,#15,#16])]	94	0	1	
66	Cyanhydrine	[NX1]#[CX2][CX4;$([CH2]),$([CH]([CX2])[#6]),$(C([CX2])([#6])[#6])][OX2H]	0	0		
67	Chloroalkene	[ClX1][CX3]=[CX3]	96	351	1	1
68	Fluoroalkene	[FX1][CX3]=[CX3]	32	13	1	1
69	Bromoalkene	[BrX1][CX3]=[CX3]	14	278	1	1
70	Iodoalkene	[IX1][CX3]=[CX3]	1	0	1	
71	Enol	[OX2H][CX3;$([H1]),$(C[#6])]=[CX3]	110	26	1	1
72	Endiol	[OX2H][CX3;$([H1]),$(C[#6])]=[CX3;$([H1]),$(C[#6])][OX2H]	8	0	1	
73	Enolether	[OX2]([#6;!$(C=[N,O,S])])[CX3;$([H0][#6]),$([H1])]=[CX3]	290	73	1	1
74	Enolester	[OX2]([CX3]=[OX1])[#6X3;$([#6][#6]),$([H1])]=[#6X3;!$(C[OX2H])]	156	2	1	1
75	Enamine	[NX3;$([NH2][CX3]),$([NH1]([CX3])[#6]),$([N]([CX3])([#6])[#6]);!$([N]*=[#7,#8,#15,#16])][CX3;$([CH]),$([C][#6])]=[CX3]	389	0	1	
76	Thioenol	[SX2H][CX3;$([H1]),$(C[#6])]=[CX3]	1	0	1	
77	Thioenolether	[SX2]([#6;!$(C=[N,O,S])])[CX3;$(C[#6]),$([CH])]=[CX3]	191	26	1	1
78	Acylchloride	[CX3;$([R0][#6]),$([H1R0])](=[OX1])[ClX1]	1	0	1	
79	Acylfluoride	[CX3;$([R0][#6]),$([H1R0])](=[OX1])[FX1]	0	0		
80	Acylbromide	[CX3;$([R0][#6]),$([H1R0])](=[OX1])[BrX1]	0	0		
81	Acyliodide	[CX3;$([R0][#6]),$([H1R0])](=[OX1])[IX1]	0	0		
82	Acylhalide	[CX3;$([R0][#6]),$([H1R0])](=[OX1])[FX1,ClX1,BrX1,IX1]	1	0	1	

Bit Position	Description	SMARTS pattern	ZINC clustered at 60% tanimoto similarity	Diversity subset of TIN	Fingerprint in ZINC	Fingerprint in TIN
83	Carboxylic acid	[CX3;$([R0][#6]),$([H1R0])](=[OX1])[$([OX2H]),$([OX1-])]	401	396	1	1
84	Carboxylic ester	[CX3;$([R0][#6]),$([H1R0])](=[OX1])[OX2][#6;!$(C=[O,N,S])]	460	613	1	1
85	Lactone	[#6][#6X3R](=[OX1])[#8X2][#6;!$(C=[O,N,S])]	303	0	1	
86	Carboxylic anhydride	[CX3;$([H0][#6]),$([H1])](=[OX1])[#8X2][CX3;$([H0][#6]),$([H1])](=[OX1])	2	0	1	
87	Carboxylic acid derivative	[$([#6X3H0][#6]),$([#6X3H])](=[!#6])[!#6]	6344	8906	1	1
88	Carbothioic acid	[CX3;!R;$([C][#6]),$([CH]);$([C](=[OX1])[$([SX2H]),$([SX1-])]),$([C](=[SX1])[$([OX2H]),$([OX1-])])]	0	0		
89	Carbothioic S ester	[CX3;$([R0][#6]),$([H1R0])](=[OX1])[SX2][#6;!$(C=[O,N,S])]	14	0	1	
90	Carbothioic S lactone	[#6][#6X3R](=[OX1])[#16X2][#6;!$(C=[O,N,S])]	33	0	1	
91	Carbothioic O ester	[CX3;$([H0][#6]),$([H1])](=[SX1])[OX2][#6;!$(C=[O,N,S])]	0	0		
92	Carbothioic O lactone	[#6][#6X3R](=[SX1])[#8X2][#6;!$(C=[O,N,S])]	0	0		
93	Carbothioic halide	[CX3;$([H0][#6]),$([H1])](=[SX1])[FX1,ClX1,BrX1,IX1]	0	0		
94	Carbodithioic acid	[CX3;!R;$([C][#6]),$([CH]);$([C](=[SX1])[SX2H])]	1	0	1	
95	Carbodithioic ester	[CX3;!R;$([C][#6]),$([CH]);$([C](=[SX1])[SX2][#6;!$(C=[O,N,S])])]	1731	963	1	1
96	Carbodithiolactone	[#6][#6X3R](=[SX1])[#16X2][#6;!$(C=[O,N,S])]	8	0	1	
97	Amide	[CX3;$([R0][#6]),$([H1R0])](=[OX1])[#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
98	Primary amide	[CX3;$([R0][#6]),$([H1R0])](=[OX1])[NX3H2]	118	0	1	
99	Secondary amide	[CX3;$([R0][#6]),$([H1R0])](=[OX1])[#7X3H1][#6;!$(C=[O,N,S])]	167	487	1	1
100	Tertiary amide	[CX3;$([R0][#6]),$([H1R0])](=[OX1])[#7X3H0]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])]	63	171	1	1
101	Lactam	[#6R][#6X3R](=[OX1])[#7X3;$([H1][#6;!$(C=[O,N,S])]),$([H0]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
102	Alkyl imide	[#6X3;$([H0][#6]),$([H1])](=[OX1])[#7X3H0]([#6])[#6X3;$([H0][#6]),$([H1])](=[OX1])	24	0	1	
103	N hetero imide	[#6X3;$([H0][#6]),$([H1])](=[OX1])[#7X3H0]([!#6])[#6X3;$([H0][#6]),$([H1])](=[OX1])	10	0	1	
104	Imide acidic	[#6X3;$([H0][#6]),$([H1])](=[OX1])[#7X3H1][#6X3;$([H0][#6]),$([H1])](=[OX1])	22	0	1	
105	Thioamide	[$([CX3;!R][#6]),$([CX3H;!R])](=[SX1])[#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		

Bit Position	Description	SMARTS pattern	ZINC clustered at 60% tanimoto similarity	Diversity subset of TIN	Fingerprint in ZINC	Fingerprint in TIN
106	Thiolactam	[#6R][#6X3R](=[SX1])[#7X3;$([H1][#6;!$(C=[O,N,S])]),$([H0]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
107	Oximester	[#6X3;$([H0][#6]),$([H1])](=[OX1])[#8X2][#7X2]=,:[#6X3;$([H0]([#6])[#6]),$([H1][#6]),$([H2])]	15	0	1	
108	Amidine	[NX3;!$(NC=[O,S])][CX3;$([CH]),$([C][#6])]=[NX2;!$(NC=[O,S])]	377	5	1	1
109	Hydroxamic acid	[CX3;$([H0][#6]),$([H1])](=[OX1])[#7X3;$([H1]),$([H0][#6;!$(C=[O,N,S])])][$([OX2H]),$([OX1-])]	0	0		
110	Hydroxamic acid ester	[CX3;$([H0][#6]),$([H1])](=[OX1])[#7X3;$([H1]),$([H0][#6;!$(C=[O,N,S])])][OX2][#6;!$(C=[O,N,S])]	0	0		
111	Imidoacid	[CX3R0;$([H0][#6]),$([H1])](=[NX2;$([H1]),$([H0][#6;!$(C=[O,N,S])])])[$([OX2H]),$([OX1-])]	0	0		
112	Imidoacid cyclic	[#6R][#6X3R](=,:[#7X2;$([H1]),$([H0][#6;!$(C=[O,N,S])])])[$([OX2H]),$([OX1-])]	0	0		
113	Imidoester	[CX3R0;$([H0][#6]),$([H1])](=[NX2;$([H1]),$([H0][#6;!$(C=[O,N,S])])])[OX2][#6;!$(C=[O,N,S])]	0	0		
114	Imidolactone	[#6R][#6X3R](=,:[#7X2;$([H1]),$([H0][#6;!$(C=[O,N,S])])])[OX2][#6;!$(C=[O,N,S])]	0	0		
115	Imidothioacid	[CX3R0;$([H0][#6]),$([H1])](=[NX2;$([H1]),$([H0][#6;!$(C=[O,N,S])])])[$([SX2H]),$([SX1-])]	0	0		
116	Imidothioacid cyclic	[#6R][#6X3R](=,:[#7X2;$([H1]),$([H0][#6;!$(C=[O,N,S])])])[$([SX2H]),$([SX1-])]	0	0		
117	Imidothioester	[CX3R0;$([H0][#6]),$([H1])](=[NX2;$([H1]),$([H0][#6;!$(C=[O,N,S])])])[SX2][#6;!$(C=[O,N,S])]	0	0		
118	Imidothiolactone	[#6R][#6X3R](=,:[#7X2;$([H1]),$([H0][#6;!$(C=[O,N,S])])])[SX2][#6;!$(C=[O,N,S])]	0	0		
119	Amidine	[#7X3v3;!$(N([#6X3]=[#7X2])C=[O,S])][CX3R0;$([H1]),$([H0][#6])]=[NX2v3;!$(N(=[#6X3][#7X3])C=[O,S])]	24	0	1	
120	Imidolactam	[#6][#6X3R;$([H0](=[NX2;!$(N(=[#6X3][#7X3])C=[O,S])])[#7X3;!$(N([#6X3]=[#7X2])C=[O,S])]),$([H0](-[NX3;!$(N([#6X3]=[#7X2])C=[O,S])]=,:[#7X2;!$(N(=[#6X3][#7X3])C=[O,S])])]	0	0		
121	Imidoylhalide	[CX3R0;$([H0][#6]),$([H1])](=[NX2;$([H1]),$([H0][#6;!$(C=[O,N,S])])])[FX1,ClX1,BrX1,IX1]	0	0		
122	Imidoylhalide cyclic	[#6R][#6X3R](=,:[#7X2;$([H1]),$([H0][#6;!$(C=[O,N,S])])])[FX1,ClX1,BrX1,IX1]	0	0		
123	Amidrazone	[$([$([#6X3][#6]),$([#6X3H])](=[#7X2v3])[#7X3v3][#7X3v3]),$([$([#6X3][#6]),$([#6X3H])]([#7X3v3])=[#7X2v3][#7X3v3])]	0	0		
124	Alpha aminoacid	[NX3,NX4+;!$([N]~[!#6]);!$([N]*~[#7,#8,#15,#16])][C][CX3](=[OX1])[OX2H,OX1-]	2	0	1	

Bit Position	Description	SMARTS pattern	ZINC clustered at 60% tanimoto similarity	Diversity subset of TIN	Fingerprint in ZINC	Fingerprint in TIN
125	Alpha hydroxyacid	[OX2H][C][CX3](=[OX1])[OX2H,OX1-]	2	2	1	1
126	Peptide middle	[NX3;$([N][CX3](=[OX1])[C][NX3,NX4+])][C][CX3](=[OX1])[NX3;$([N][C][CX3](=[OX1])[NX3,OX2,OX1-])]	23	0	1	
127	Peptide C term	[NX3;$([N][CX3](=[OX1])[C][NX3,NX4+])][C][CX3](=[OX1])[OX2H,OX1-]	3	0	1	
128	Peptide N term	[NX3,NX4+;!$([N]~[!#6]);!$([N]*~[#7,#8,#15,#16])][C][CX3](=[OX1])[NX3;$([N][C][CX3](=[OX1])[NX3,OX2,OX1-])]	0	0		
129	Carboxylic orthoester	[#6][OX2][CX4;$(C[#6]),$([CH])]([OX2][#6])[OX2][#6]	4	0	1	
130	Ketene	[CX3]=[CX2]=[OX1]	1	0	1	
131	Ketenacetal	[#7X2,#8X3,#16X2;$(*[#6,#14])][#6X3]([#7X2,#8X3,#16X2;$(*[#6,#14])])=[#6X3]	186	0	1	
132	Nitrile	[NX1]#[CX2]	678	0	1	
133	Isonitrile	[CX1-]#[NX2+]	1	0	1	
134	Vinylogous carbonyl or carboxyl derivative	[#6X3](=[OX1])[#6X3]=,:[#6X3][#7,#8,#16,F,Cl,Br,I]	1472	1580	1	1
135	Vinylogous acid	[#6X3](=[OX1])[#6X3]=,:[#6X3][$([OX2H]),$([OX1-])]	187	150	1	1
136	Vinylogous ester	[#6X3](=[OX1])[#6X3]=,:[#6X3][#6;!$(C=[O,N,S])]	2005	3602	1	1
137	Vinylogous amide	[#6X3](=[OX1])[#6X3]=,:[#6X3][#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
138	Vinylogous halide	[#6X3](=[OX1])[#6X3]=,:[#6X3][FX1,ClX1,BrX1,IX1]	37	297	1	1
139	Carbonic acid dieester	[#6;!$(C=[O,N,S])][#8X2][#6X3](=[OX1])[#8X2][#6;!$(C=[O,N,S])]	14	0	1	
140	Carbonic acid esterhalide	[#6;!$(C=[O,N,S])][OX2;!R][CX3](=[OX1])[OX2][FX1,ClX1,BrX1,IX1]	0	0		
141	Carbonic acid monoester	[#6;!$(C=[O,N,S])][OX2;!R][CX3](=[OX1])[$([OX2H]),$([OX1-])]	0	0		
142	Carbonic acid derivatives	[!#6][#6X3](=[!#6])[!#6]	3590	2174	1	1
143	Thiocarbonic acid dieester	[#6;!$(C=[O,N,S])][#8X2][#6X3](=[SX1])[#8X2][#6;!$(C=[O,N,S])]	0	0		
144	Thiocarbonic acid esterhalide	[#6;!$(C=[O,N,S])][OX2;!R][CX3](=[SX1])[OX2][FX1,ClX1,BrX1,IX1]	0	0		
145	Thiocarbonic acid monoester	[#6;!$(C=[O,N,S])][OX2;!R][CX3](=[SX1])[$([OX2H]),$([OX1-])]	0	0		
146	Urea	[#7X3;!$([#7][!#6])][#6X3](=[OX1])[#7X3;!$([#7][!#6])]	41	0	1	
147	Thiourea	[#7X3;!$([#7][!#6])][#6X3](=[SX1])[#7X3;!$([#7][!#6])]	6	0	1	

Bit Position	Description	SMARTS pattern	ZINC clustered at 60% tanimoto similarity	Diversity subset of TIN	Fingerprint in ZINC	Fingerprint in TIN
148	Isourea	[#7X2;!$([#7][!#6])]=,:[#6X3]([#8X2&!$([#8][!#6]),OX1-])[#7X3;!$([#7][!#6])]	27	0	1	
149	Isothiourea	[#7X2;!$([#7][!#6])]=,:[#6X3]([#16X2&!$([#16][!#6]),SX1-])[#7X3;!$([#7][!#6])]	73	0	1	
150	Guanidine	[N;v3X3,v4X4+][CX3](=[N;v3X2,v4X3+])[N;v3X3,v4X4+]	262	4	1	1
151	Carbaminic acid	[NX3]C(=[OX1])[O;X2H,X1-]	1	0	1	
152	Urethan	[#7X3][#6](=[OX1])[#8X2][#6]	187	9	1	1
153	Biuret	[#7X3][#6](=[OX1])[#7X3][#6](=[OX1])[#7X3]	12	0	1	
154	Semicarbazide	[#7X3][#7X3][#6X3]([#7X3;!$([#7][#7])])=[OX1]	60	0	1	
155	Carbazide	[#7X3][#7X3][#6X3]([#7X3][#7X3])=[OX1]	3	0	1	
156	Semicarbazone	[#7X2](=[#6])[#7X3][#6X3]([#7X3;!$([#7][#7])])=[OX1]	46	0	1	
157	Carbazone	[#7X2](=[#6])[#7X3][#6X3]([#7X3][#7X3])=[OX1]	4	0	1	
158	Thiosemicarbazide	[#7X3][#7X3][#6X3]([#7X3;!$([#7][#7])])=[SX1]	21	0	1	
159	Thiocarbazide	[#7X3][#7X3][#6X3]([#7X3][#7X3])=[SX1]	4	0	1	
160	Thiosemicarbazone	[#7X2](=[#6])[#7X3][#6X3]([#7X3;!$([#7][#7])])=[SX1]	36	0	1	
161	Thiocarbazone	[#7X2](=[#6])[#7X3][#6X3]([#7X3][#7X3])=[SX1]	5	0	1	
162	Isocyanate	[NX2]=[CX2]=[OX1]	4	0	1	
163	Cyanate	[OX2][CX2]#[NX1]	0	0		
164	Isothiocyanate	[NX2]=[CX2]=[SX1]	13	0	1	
165	Thiocyanate	[SX2][CX2]#[NX1]	8	0	1	
166	Carbodiimide	[NX2]=[CX2]=[NX2]	3	0	1	
167	Orthocarbonic derivatives	[CX4H0]([O,S,#7])([O,S,#7])([O,S,#7])[O,S,#7,F,Cl,Br,I]	3	0	1	
168	Phenol	[OX2H][c]	156	3233	1	1
169	1,2-Diphenol	[OX2H][c][c][OX2H]	7	387	1	1
170	Arylchloride	[Cl][c]	203	1469	1	1
171	Arylfluoride	[F][c]	93	672	1	1
172	Arylbromide	[Br][c]	13	1021	1	1
173	Aryliodide	[I][c]	3	458	1	1
174	Arylthiol	[SX2H][c]	5	49	1	1
175	Iminoarene	[c]=[NX2;$([H1]),$([H0][#6;!$([C]=[N,S,O])])]	0	0		
176	Oxoarene	[c]=[OX1]	0	0		
177	Thioarene	[c]=[SX1]	0	0		
178	Hetero N basic H	[nX3H1+0]	732	2449	1	1
179	Hetero N basic no H	[nX3H0+0]	1510	5830	1	1
180	Hetero N nonbasic	[nX2,nX3+]	3644	10984	1	1
181	Hetero O	[o]	1145	5696	1	1
182	Hetero S	[sX2]	1058	0	1	
183	Heteroaromatic	[a;!c]	4507	10984	1	1

Bit Position	Description	SMARTS pattern	ZINC clustered at 60% tanimoto similarity	Diversity subset of TIN	Fingerprint in ZINC	Fingerprint in TIN
184	Nitrite	[NX2](=[OX1])[O;$([X2]),$([X1-])]	0	0		
185	Thionitrite	[SX2][NX2]=[OX1]	0	0		
186	Nitrate	[$([NX3](=[OX1])(=[OX1])[O;$([X2]),$([X1-])]),$([NX3+]([OX1-])(=[OX1])[O;$([X2]),$([X1-])])]	0	0		
187	Nitro	[$([NX3](=O)=O),$([NX3+](=O)[O-])][!#8]	353	4706	1	1
188	Nitroso	[NX2](=[OX1])[!#7;!#8]	69	0	1	
189	Azide	[NX1]~[NX2]~[NX2,NX1]	16	0	1	
190	Acylazide	[CX3](=[OX1])[NX2]~[NX2]~[NX1]	0	0		
191	Diazo	[$([#6]=[NX2+]=[NX1-]),$([#6-]-[NX2+]#[NX1])]	22	0	1	
192	Diazonium	[#6][NX2+]#[NX1]	3	0	1	
193	Nitrosamine	[#7;!$(N*=O)][NX2]=[OX1]	0	0		
194	Nitrosamide	[NX2](=[OX1])N-*=O	1	0	1	
195	N-Oxide	[$([#7+][OX1-]),$([#7v5]=[OX1]);!$([#7](~[O])~[O]);!$([#7]=[#7])]	71	0	1	
196	Hydrazine	[NX3;$([H2]),$([H1][#6]),$([H0]([#6])[#6]);!$(NC=[O,N,S])][NX3;$([H2]),$([H1][#6]),$([H0]([#6])[#6]);!$(NC=[O,N,S])]	100	77	1	1
197	Hydrazone	[NX3;$([H2]),$([H1][#6]),$([H0]([#6])[#6]);!$(NC=[O,N,S])][NX2]=[#6]	255	0	1	
198	Hydroxylamine	[NX3;$([H2]),$([H1][#6]),$([H0]([#6])[#6]);!$(NC=[O,N,S])][OX2;$([H1]),$(O[#6;!$(C=[N,O,S])])]	0	0		
199	Sulfon	[$([SX4](=[OX1])(=[OX1])([#6])[#6]),$([SX4+2]([OX1-])([OX1-])([#6])[#6])]	169	0	1	
200	Sulfoxide	[$([SX3](=[OX1])([#6])[#6]),$([SX3+]([OX1-])([#6])[#6])]	27	0	1	
201	Sulfonium	[S+;!$([S]~[!#6]);!$([S]*~[#7,#8,#15,#16])]	8	0	1	
202	Sulfuric acid	[SX4](=[OX1])(=[OX1])([$([OX2H]),$([OX1-])])[$([OX2H]),$([OX1-])]	0	5		1
203	Sulfuric monoester	[SX4](=[OX1])(=[OX1])([$([OX2H]),$([OX1-])])[OX2][#6;!$(C=[O,N,S])]	3	0	1	
204	Sulfuric diester	[SX4](=[OX1])(=[OX1])([OX2][#6;!$(C=[O,N,S])])[OX2][#6;!$(C=[O,N,S])]	0	0		
205	Sulfuric monoamide	[SX4](=[OX1])(=[OX1])([#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])])[$([OX2H]),$([OX1-])]	0	0		
206	Sulfuric diamide	[SX4](=[OX1])(=[OX1])([#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])])[#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
207	Sulfuric esteramide	[SX4](=[OX1])(=[OX1])([#7X3][#6;!$(C=[O,N,S])])[OX2][#6;!$(C=[O,N,S])]	2	0	1	
208	Sulfuric derivative	[SX4D4](=[!#6])(=[!#6])([!#6])[!#6]	77	5	1	1

Bit Position	Description	SMARTS pattern	ZINC clustered at 60% tanimoto similarity	Diversity subset of TIN	Fingerprint in ZINC	Fingerprint in TIN
209	Sulfonic acid	[SX4;$([H1]),$([H0][#6])](=[OX1])(=[OX1])[$([OX2H]),$([OX1-])]	31	0	1	
210	Sulfonamide	[SX4;$([H1]),$([H0][#6])](=[OX1])(=[OX1])[#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
211	Sulfonic ester	[SX4;$([H1]),$([H0][#6])](=[OX1])(=[OX1])[OX2][#6;!$(C=[O,N,S])]	34	0	1	
212	Sulfonic halide	[SX4;$([H1]),$([H0][#6])](=[OX1])(=[OX1])[FX1,ClX1,BrX1,IX1]	1	0	1	
213	Sulfonic derivative	[SX4;$([H1]),$([H0][#6])](=[!#6])(=[!#6])[!#6]	326	0	1	
214	Sulfinic acid	[SX3;$([H1]),$([H0][#6])](=[OX1])[$([OX2H]),$([OX1-])]	1	0	1	
215	Sulfinic amide	[SX3;$([H1]),$([H0][#6])](=[OX1])[#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
216	Sulfinic ester	[SX3;$([H1]),$([H0][#6])](=[OX1])[OX2][#6;!$(C=[O,N,S])]	4	0	1	
217	Sulfinic halide	[SX3;$([H1]),$([H0][#6])](=[OX1])[FX1,ClX1,BrX1,IX1]	0	0		
218	Sulfinic derivative	[SX3;$([H1]),$([H0][#6])](=[!#6])[!#6]	17	0	1	
219	Sulfenic acid	[SX2;$([H1]),$([H0][#6])][$([OX2H]),$([OX1-])]	0	0		
220	Sulfenic amide	[SX2;$([H1]),$([H0][#6])][#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
221	Sulfenic ester	[SX2;$([H1]),$([H0][#6])][OX2][#6;!$(C=[O,N,S])]	3	0	1	
222	Sulfenic halide	[SX2;$([H1]),$([H0][#6])][FX1,ClX1,BrX1,IX1]	0	0		
223	Sulfenic derivative	[SX2;$([H1]),$([H0][#6])][!#6]	312	246	1	1
224	Phosphine	[PX3;$([H3]),$([H2][#6]),$([H1]([#6])[#6]),$([H0]([#6])([#6])[#6])]	13	0	1	
225	Phosphine oxide	[PX4;$([H3]=[OX1]),$([H2](=[OX1])[#6]),$([H1](=[OX1])([#6])[#6]),$([H0](=[OX1])([#6])([#6])[#6])]	30	0	1	
226	Phosphonium	[P+;!$([P]~[!#6]);!$([P]*~[#7,#8,#15,#16])]	3	0	1	
227	Phosphorylen	[PX4;$([H3]=[CX3]),$([H2](=[CX3])[#6]),$([H1](=[CX3])([#6])[#6]),$([H0](=[CX3])([#6])([#6])[#6])]	3	0	1	
228	Phosphonic acid	[PX4;$([H1]),$([H0][#6])](=[OX1])([$([OX2H]),$([OX1-])])[$([OX2H]),$([OX1-])]	35	0	1	
229	Phosphonic monoester	[PX4;$([H1]),$([H0][#6])](=[OX1])([$([OX2H]),$([OX1-])])[OX2][#6;!$(C=[O,N,S])]	9	0	1	
230	Phosphonic diester	[PX4;$([H1]),$([H0][#6])](=[OX1])([OX2][#6;!$(C=[O,N,S])])[OX2][#6;!$(C=[O,N,S])]	41	0	1	
231	Phosphonic monoamide	[PX4;$([H1]),$([H0][#6])](=[OX1])([$([OX2H]),$([OX1-])])[#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		

Bit Position	Description	SMARTS pattern	ZINC clustered at 60% tanimoto similarity	Diversity subset of TIN	Fingerprint in ZINC	Fingerprint in TIN
232	Phosphonic diamide	[PX4;$([H1]),$([H0][#6])](=[OX1])([#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])])[#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
233	Phosphonic esteramide	[PX4;$([H1]),$([H0][#6])](=[OX1])([OX2][#6;!$(C=[O,N,S])])[#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
234	Phosphonic acid derivative	[PX4;$([H1]),$([H0][#6])](=[!#6])([!#6])[!#6]	146	0	1	
235	Phosphoric acid	[PX4D4](=[OX1])([$([OX2H]),$([OX1-])])([$([OX2H]),$([OX1-])])[$([OX2H]),$([OX1-])]	0	2		1
236	Phosphoric monoester	[PX4D4](=[OX1])([$([OX2H]),$([OX1-])])([$([OX2H]),$([OX1-])])[OX2][#6;!$(C=[O,N,S])]	8	0	1	
237	Phosphoric diester	[PX4D4](=[OX1])([$([OX2H]),$([OX1-])])([OX2][#6;!$(C=[O,N,S])])[OX2][#6;!$(C=[O,N,S])]	4	0	1	
238	Phosphoric triester	[PX4D4](=[OX1])([OX2][#6;!$(C=[O,N,S])])([OX2][#6;!$(C=[O,N,S])])[OX2][#6;!$(C=[O,N,S])]	7	0	1	
239	Phosphoric monoamide	[PX4D4](=[OX1])([$([OX2H]),$([OX1-])])([$([OX2H]),$([OX1-])])[#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
240	Phosphoric diamide	[PX4D4](=[OX1])([$([OX2H]),$([OX1-])])([#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])])[#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
241	Phosphoric triamide	[PX4D4](=[OX1])([#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])])([#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])])[#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
242	Phosphoric monoestermonoamide	[PX4D4](=[OX1])([$([OX2H]),$([OX1-])])([OX2][#6;!$(C=[O,N,S])])[#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
243	Phosphoric diestermonoamide	[PX4D4](=[OX1])([OX2][#6;!$(C=[O,N,S])])([OX2][#6;!$(C=[O,N,S])])[#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		

Bit Position	Description	SMARTS pattern	ZINC clustered at 60% tanimoto similarity	Diversity subset of TIN	Fingerprint in ZINC	Fingerprint in TIN
244	Phosphoric monoesterdiamide	[PX4D4](=[OX1])([OX2][#6;!$(C=[O,N,S])])([#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])])[#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
245	Phosphoric acid derivative	[PX4D4](=[!#6])([!#6])([!#6])[!#6]	89	2	1	1
246	Phosphinic acid	[PX4;$([H2]),$([H1][#6]),$([H0]([#6])[#6])](=[OX1])[$([OX2H]),$([OX1-])]	28	0	1	
247	Phosphinic ester	[PX4;$([H2]),$([H1][#6]),$([H0]([#6])[#6])](=[OX1])[OX2][#6;!$(C=[O,N,S])]	15	0	1	
248	Phosphinic amide	[PX4;$([H2]),$([H1][#6]),$([H0]([#6])[#6])](=[OX1])[#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
249	Phosphinic acid derivative	[PX4;$([H2]),$([H1][#6]),$([H0]([#6])[#6])](=[!#6])[!#6]	67	0	1	
250	Phosphonous acid	[PX3;$([H1]),$([H0][#6])]([$([OX2H]),$([OX1-])])[$([OX2H]),$([OX1-])]	0	0		
251	Phosphonous monoester	[PX3;$([H1]),$([H0][#6])]([$([OX2H]),$([OX1-])])[OX2][#6;!$(C=[O,N,S])]	0	0		
252	Phosphonous diester	[PX3;$([H1]),$([H0][#6])]([OX2][#6;!$(C=[O,N,S])])[OX2][#6;!$(C=[O,N,S])]	1	0	1	
253	Phosphonous monoamide	[PX3;$([H1]),$([H0][#6])]([$([OX2H]),$([OX1-])])[#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
254	Phosphonous diamide	[PX3;$([H1]),$([H0][#6])]([#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])])[#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
255	Phosphonous esteramide	[PX3;$([H1]),$([H0][#6])]([OX2][#6;!$(C=[O,N,S])])[#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
256	Phosphonous derivatives	[PX3;$([D2]),$([D3][#6])]([!#6])[!#6]	6	0	1	
257	Phosphinous acid	[PX3;$([H2]),$([H1][#6]),$([H0]([#6])[#6])][$([OX2H]),$([OX1-])]	0	0		
258	Phosphinous ester	[PX3;$([H2]),$([H1][#6]),$([H0]([#6])[#6])][OX2][#6;!$(C=[O,N,S])]	0	0		
259	Phosphinous amide	[PX3;$([H2]),$([H1][#6]),$([H0]([#6])[#6])][#7X3;$([H2]),$([H1][#6;!$(C=[O,N,S])]),$([#7]([#6;!$(C=[O,N,S])])[#6;!$(C=[O,N,S])])]	0	0		
260	Phosphinous derivatives	[PX3;$([H2]),$([H1][#6]),$([H0]([#6])[#6])][!#6]	5	0	1	
261	Quart silane	[SiX4]([#6])([#6])([#6])[#6]	0	0		
262	Non-quart silane	[SiX4;$([H1]([#6])([#6])[#6]),$([H2]([#6])[#6]),$([H3][#6]),$([H4])]	0	0		

Bit Position	Description	SMARTS pattern	ZINC clustered at 60% tanimoto similarity	Diversity subset of TIN	Fingerprint in ZINC	Fingerprint in TIN
263	Silylmonohalide	[SiX4]([FX1,ClX1,BrX1,IX1])([#6])[#6]	0	0		
264	Het trialkylsilane	[SiX4]([!#6])([#6])([#6])[#6]	0	0		
265	Dihet dialkylsilane	[SiX4]([!#6])([!#6])([#6])[#6]	0	0		
266	Trihet alkylsilane	[SiX4]([!#6])([!#6])([!#6])[#6]	0	0		
267	Silicic acid derivative	[SiX4]([!#6])([!#6])([!#6])[!#6]	0	0		
268	Trialkylborane	[BX3]([#6])([#6])[#6]	0	0		
269	Boric acid derivatives	[BX3]([!#6])([!#6])[!#6]	0	0		
270	Boronic acid derivative	[BX3]([!#6])([!#6])[!#6]	0	0		
271	Borohydride	[BH1,BH2,BH3,BH4]	0	0		
272	Quaternary boron	[BX4]	0	0		
273	Aromatic	a	6828	10995	1	1
274	Heterocyclic	[!#6;!R0]	9115	11000	1	1
275	Epoxide	[OX2r3]1[#6r3][#6r3]1	57	0	1	
276	NH aziridine	[NX3H1r3]1[#6r3][#6r3]1	3	0	1	
277	Spiro	[D4R;$(*(@*)(@*)(@*)@*)]	0	0		
278	Annelated rings	[R;$(*(@*)(@*)@*);!$([R2;$(*(@*)(@*)(@*)@*)])]@[R;$(*(@*)(@*)@*);!$([R2;$(*(@*)(@*)(@*)@*)])]	136	1993	1	1
279	Bridged rings	[R;$(*(@*)(@*)@*);!$([D4R;$(*(@*)(@*)(@*)@*)]);!$([R;$(*(@*)(@*)@*);!$([R2;$(*(@*)(@*)(@*)@*)])]@[R;$(*(@*)(@*)@*);!$([R2;$(*(@*)(@*)(@*)@*)])])]	136	1993	1	1
280	Sugar pattern 1	[OX2;$([r5]1@C@C@C(O)@C1),$([r6]1@C@C@C(O)@C(O)@C1)]	117	132	1	1
281	Sugar pattern 2	[OX2;$([r5]1@C(!@[OX2,NX3,SX2,FX1,ClX1,BrX1,IX1])@C@C@C1),$([r6]1@C(!@[OX2,NX3,SX2,FX1,ClX1,BrX1,IX1])@C@C@C@C1)]	106	123	1	1
282	Sugar pattern combi	[OX2;$([r5]1@C(!@[OX2,NX3,SX2,FX1,ClX1,BrX1,IX1])@C@C(O)@C1),$([r6]1@C(!@[OX2,NX3,SX2,FX1,ClX1,BrX1,IX1])@C@C(O)@C(O)@C1)]	18	123	1	1
283	Sugar pattern 2 reducing	[OX2;$([r5]1@C(!@[OX2H1])@C@C@C1),$([r6]1@C(!@[OX2H1])@C@C@C@C1)]	17	0	1	
284	Sugar pattern 2 alpha	[OX2;$([r5]1@[C@@](!@[OX2,NX3,SX2,FX1,ClX1,BrX1,IX1])@C@C@C1),$([r6]1@[C@@](!@[OX2,NX3,SX2,FX1,ClX1,BrX1,IX1])@C@C@C@C1)]	106	123	1	1
285	Sugar pattern 2 beta	[OX2;$([r5]1@[C@](!@[OX2,NX3,SX2,FX1,ClX1,BrX1,IX1])@C@C@C1),$([r6]1@[C@](!@[OX2,NX3,SX2,FX1,ClX1,BrX1,IX1])@C@C@C@C1)]	106	123	1	1
286	Conjugated double bond	*=*[*]=,#,:[*]	8746	11000	1	1
287	Conjugated tripple bond	*#*[*]=,#,:[*]	590	1137	1	1
288	Cis double bond	*/[D2]=[D2]/*	0	0		
289	Trans double bond	*/[D2]=[D2]/*	0	0		

Bit Position	Description	SMARTS pattern	ZINC clustered at 60% tanimoto similarity	Diversity subset of TIN	Fingerprint in ZINC	Fingerprint in TIN
290	Mixed anhydrides	[$(*=O),$([#16,#14,#5]),$([#7]([#6]=[OX1]))][#8X2][$(*=O),$([#16,#14,#5]),$([#7]([#6]=[OX1]))]	24	0	1	
291	Halogen on hetero	[FX1,ClX1,BrX1,IX1][!#6]	11	430	1	1
292	Halogen multi subst	[F,Cl,Br,I;!$([X1]);!$([X0-])]	1	12	1	1
293	Trifluoromethyl	[FX1][CX4;!$([H0][Cl,Br,I]);!$([F][C]([F])([F])[F])]([FX1])([FX1])	86	4897	1	1
294	C ONS bond	[#6]~[#7,#8,#16]	10000	11000	1	1
295	Charged	[!+0]	3340	9950	1	1
296	Anion	[-1,-2,-3,-4,-5,-6,-7]	1890	4729	1	1
297	Kation	[+1,+2,+3,+4,+5,+6,+7]	2299	9942	1	1
298	Salt	([-1,-2,-3,-4,-5,-6,-7]).([+1,+2,+3,+4,+5,+6,+7])	1890	4729	1	1
299	1,3-Tautomerizable	[$([#7X2,OX1,SX1]=*[!H0;!$([a;!n])]),$([#7X3,OX2,SX2;!H0]*=*),$([#7X3,OX2,SX2;!H0]*:n)]	9627	10999	1	1
300	1,5-Tautomerizable	[$([#7X2,OX1,SX1]=,:**=,:*[!H0;!$([a;!n])]),$([#7X3,OX2,SX2;!H0]*=**=*),$([#7X3,OX2,SX2;!H0]*=,:**:n)]	9627	10999	1	1
301	Rotatable bond	[!$(*#*)&!D1]-!@[!$(*#*)&!D1]	7396	11000	1	1
302	Michael acceptor	[CX3]=[CX3][$([CX3]=[O,N,S]),$(C#[N]),$([S,P]=[OX1]),$([NX3]=O),$([NX3+](=O)[O-])]	2489	642	1	1
303	Dicarbodiazene	[CX3](=[OX1])[NX2]=[NX2][CX3](=[OX1])	7	0	1	
304	CH-acidic	[$([CX4;!$([H0]);!$(C[!#6;!$([P,S]=O);!$(N(~O)~O)])][$([CX3]=[O,N,S]),$(C#[N]),$([S,P]=[OX1]),$([NX3]=O),$([NX3+](=O)[O-]);!$(*[S,O,N;H1,H2]);!$([*+0][S,O;X1-])]),$([CX4;!$([H0])]1[CX3]=[CX3][CX3]=[CX3]1)]	0	0		
305	CH-acidic strong	[CX4;!$([H0]);!$(C[!#6;!$([P,S]=O);!$(N(~O)~O)])]([$([CX3]=[O,N,S]),$(C#[N]),$([S,P]=[OX1]),$([NX3]=O),$([NX3+](=O)[O-]);!$(*[S,O,N;H1,H2]);!$([*+0][S,O;X1-])])[$([CX3]=[O,N,S]),$(C#[N]),$([S,P]=[OX1]),$([NX3]=O),$([NX3+](=O)[O-]);!$(*[S,O,N;H1,H2]);!$([*+0][S,O;X1-])]	0	0		
306	Chiral center specified	[$([*@](~*)(~*)(*)*),$([*@H](*)(*)*),$([*@](~*)(*)*),$([*@H](~*)~*)]	10081	11000	1	1

Table S3: Bioactive molecules using TIN scaffold substructures.

		Pubchem			CHEMbl		
Scaffold substructure	**Name**	**All compounds**	**Active compounds***	**No. Targets**	**All compounds**	**Positive Assay results***	**No. Targets**
o1nc(C)c(CCC(O)=O)c1C	MA01_00001	1337	11	1	35	114	17
o1nc(C)c(CCc2onc(C)c2[N+](=O)[O-])c1C	MA02_00001	0	0	0	0	0	0
o1nc(C)c(CCc2onc(C)c2[NH2+]C=O)c1C	MA03_00001	0	0	0	0	0	0
o1nc(C)c([N+](=O)[O-])c1CCC(C(=O)C)C(=O)C	MA04_00001	4	0	0	0	0	0
O1N=C(C)C([N+](=O)[O-])C12CC(=O)C(CC2)C	MA05_00001	10	0	0	0	0	0
o1nc(C)c([N+](=O)[O-])c1CCC(C(O)=O)C(O)=O	MA06_00001	5	0	0	0	0	0
o1nc(C)c([N+](=O)[O-])c1CCC=O	MA07_00001	6	0	0	0	0	0
o1nc(c2NC(=O)CCCc12)C	MA08_00001	6	0	0	0	0	0
[oH+]1nc(c2nc(O)ccc12)C	MA11_00001	0	0	0	0	0	0
O=Cc1c(cc(nc1C)C#CCCC)C=O	MA16_00001	0	0	0	0	0	0
FC(F)(F)CON1N=C(C=C(N=C1C)C=O)C#Cc1ccccc1	MA20_00001	0	0	0	0	0	0
FC(F)(F)CON1[NH+]=C(N=C(C=C1C#Cc1ccccc1)C=O)C	MA21_00001	0	0	0	0	0	0

* ChEMBL Positive Assay result: Activation/Inhibition either IC50, EC50, Ki < 10000 nM; Max Inhibition, Activity, Recovery > 50%. PubChem compounds returned as "Active in any Assay" from search.