

Supporting Information Available:

SMARTS for reactive functional groups	p21
SMARTS for unsuitable leads	p22
SMARTS for unsuitable natural products	p23
SMARTS for Acids, Bases, Electrophiles and Nucleophiles	p24

SMARTS for reactive functional groups

R1	Reactive alkyl halides	<chem>[Br,Cl,I][CX4,CH,CH2]</chem>
R2	Acid halides	<chem>[S,C](=[O,S])[F,Br,Cl,I]</chem>
R3	Carbazides	<chem>O=CN=[N+]=[N-]</chem>
R4	Sulphate esters	<chem>COS(=O)O[C,c]</chem>
R5	Sulphonates	<chem>COS(=O)(=O)[C,c]</chem>
R6	Acid anhydrides	<chem>C(=O)OC(=O)</chem>
R7	Peroxides	<chem>OO</chem>
R8	Pentafluorophenyl esters	<chem>C(=O)Oc1c(F)c(F)c(F)c(F)c1(F)</chem>
R9	Paranitrophenyl esters	<chem>C(=O)Oc1ccc(N(=O)=O)cc1</chem>
R10	esters of HOBT	<chem>C(=O)Onnn</chem>
R11	Isocyanates & Isothiocyanates	<chem>N=C=[S,O]</chem>
R12	Triflates	<chem>OS(=O)(=O)C(F)(F)F</chem>
R13	lawesson's reagent and derivatives	<chem>P(=S)(S)S</chem>
R14	phosphoramides	<chem>NP(=O)(N)N</chem>
R15	Aromatic azides	<chem>cN=[N+]=[N-]</chem>

R16	beta carbonyl quaternary Nitrogen	<chem>C(=O)C[N+,n+]</chem>
R17	acylhydrazide	<chem>[N;R0][N;R0]C(=O)</chem>
R18	Quaternary C,Cl,I,P or S	<chem>[C+,Cl+,I+,P+,S+]</chem>
R19	Phosphoranes	<chem>C=P</chem>
R20	Chloramidines	<chem>[Cl]C([C&R0])=N</chem>
R21	Nitroso	<chem>[N&D2](=O)</chem>
R22	P/S Halides	<chem>[P,S][Cl,Br,F,I]</chem>
R23	Carbodiimide	<chem>N=C=N</chem>
R24	Isonitrile	<chem>[N+]#[C-]</chem>
R25	Triacyloximes	<chem>C(=O)N(C(=O))OC(=O)</chem>
R26	Cyanohydrins	<chem>N#CC[OH]</chem>
R27	Acyl cyanides	<chem>N#CC(=O)</chem>
R28	Sulfonyl cyanides	<chem>S(=O)(=O)C#N</chem>
R29	Cyanophosphonates	<chem>P(OCC)(OCC)(=O)C#N</chem>
R30	Azocyanamides	<chem>[N;R0]=[N;R0]C#N</chem>
R31	Azoalkanal	<chem>[N;R0]=[N;R0]CC=O</chem>

SMARTS for unsuitable leads

I1	Aliphatic methylene chains 7 or more long	<chem>[CD2;R0][CD2;R0][CD2;R0][CD2;R0][CD2;R0][CD2;R0][CD2;R0]</chem>
I2	Compounds with 4 or more acidic groups	<chem>[C,S,P](=O)[OH].[C,S,P](=O)[OH].[C,S,P](=O)[OH].[C,S,P](=O)[OH]</chem>
I3	Crown ethers	<chem>[O;R1][C;R1][C;R1][O;R1][C;R1][C;R1][O;R1]</chem>
I4	Disulphides	<chem>SS</chem>

I5	Thiols	[SH]
I6	Epoxides, Thioepoxides, Aziridines	C1[O,S,N]C1
I7	2,4,5 trihydroxyphenyl	c([OH])c([OH])c([OH])
I8	2,3,4 trihydroxyphenyl	c([OH])c([OH])cc([OH])
I9	Hydrazothiourea	N=NC(=S)N
I10	Thiocyanate	SC#N
I11	Benzylic quaternary Nitrogen	cC[N+]
I12	Thioesters	C[O,S;R0][C;R0](=S)
I13	Cyanamides	N[CH2]C#N
I14	Four membered lactones	C1(=O)OCC1
I15	Di and Triphosphates	P(=O)([OH])OP(=O)[OH]
I16	Betalactams	N1CCC1=O

SMARTS for unsuitable natural products

N1	Quinones	O=C1[#6]~[#6]C(=O)[#6]~[#6]1
N2	Polyenes	C=CC=CC=CC=C
N3	Saponin derivatives	O1CCCCC1OC2CCC3CCCCC3C2
N4	Cytochalasin derivatives	O=C1NCC2CCCCC21
N5	Cycloheximide derivatives	O=C1CCCC(N1)=O
N6	Monensin derivatives	O1CCCCC1C2CCCCO2
N7	Cyanidin derivatives	[OH]c1cc([OH])cc2=[O+]C(=C([OH])Cc21)c3cc([OH])c([OH])
N8	Squalestatin derivatives	C12OCCC(O1)CC2



SMARTS for Acids, Bases, Electrophiles and nucleophiles.

A1 e.g., carboxylic acid [OH1][P,C,S](=O)

A2 e.g., imide [NH1]([P,S]=O)([P,S]=O)

A3 [nH]1cnoc1=O

A4 [OH1]C1=NC=NO1

A5 [NH1]1C=NOS1=O

A6 [OH1]C1=NC(=O)CC1=O

A7 [OH1]C1NC(=O)C(=O)C1

A8 [nH1]1ncoc1=O

A9 [OH1]C1=NN=CO1

A10 [nH1]1[nH]cnc1=O

A11 [OH1]C1=N[NH1]C=N1

A12 [OH1]C1=NOC=C1

A13 [nH1]1occc1=O

A14 [OH1]c1oncc1

A15 [nH1]1ccc(=O)o1

A16 tetrazole [nH1]nnn

A17 [nH1](n)nn

A18 [OH1]C1=NC(=O)NO1

A19 [OH1]C1=NC(=O)ON1

A20 [nH1]1cnnc1C(F)(F)F

A21 [nH1]1cnc(n1)C(F)(F)F

A22 [nH1]1C(=O)CC(=O)O1

A23 [OH1]C1=CC(=O)NO1

A24 [OH1]C1=CC(=O)ON1

A25 benzosulphimide[NH1]1C(=O)c2ccccc2S1(=O)=O

A26 [OH1]C1=NS(=O)(=O)c2ccccc21

A27 [OH1]C1=NC(=O)c2ccccc21

A28 [OH1]C1=COC=CC1=O

A29 [OH1]C1=NSN=C1

A30 hyroxamic acid [OH1]NC(=O)

A31 trifluoromethyl sulphonamide [NH]S(=O)(=O)C(F)(F)F

A32 aryl sulphonamide [NH](c)S(=O)=O

A33 phenol [OH1]c1c[c,n]ccc1

ACID

[\$A1,\$A2,\$A3,\$A4,\$A5,\$A6,\$A7,\$A8,\$A9,\$A10,\$A11,\$A12,\$A13,\$A14,\$A15,\$A16,\$A17,\$A18,\$A19,\$A20,\$A21,\$A22,\$A23,\$A24,\$A25,\$A26,\$A27,\$A28,\$A29,\$A30,\$A31,\$A32,\$A33]

B1 primary amine [NH2][CX4]

B2 secondary amine [NH]([CX4])[CX4]

B3 tertiary amine [NX3]([CX4])([CX4])[CX4]

B4SUB [C,c](=N)N

B5EXC [C,c](=N)N[C,S](=O)

B6 [\$B4SUB;\$B5EXC]

B4 – B6 are amidines and guanidines but not their acylated derivatives

B7EXC n(:c)(:c):a

B7 [nH0;\$!(n-C);\$B7EXC]1ccccc1

B8EXC [N,n,+1]

B9 $[(\text{NH}_2)!:\text{c}], [(\text{NH}_1)(\text{CX}_4)!:\text{c}], [(\text{NH}_0)(\text{CX}_4)(\text{CX}_4)!:\text{c}]$

B7 – B9 are heterocyclic bases e.g., pyridines

BASE $[\text{B}1, \text{B}2, \text{B}3, \text{B}6, \text{B}7, \text{B}9; !\text{B}8\text{EXC}]$

E1 alkyl and aryl ketones and aldehydes $[\text{C}; \text{H}1](=\text{O}, \text{S})[\text{C}, \text{c}]$

E2 e.g., carboxylic esters $[\text{C}, \text{P}; \text{H}1](=\text{O}, \text{S})[\text{O}, \text{S}]$

E3 e.g., carbonates $[\text{C}](=\text{O})([\text{C}, \text{c}, \text{O}, \text{S}])[\text{C}, \text{c}, \text{O}, \text{S}]$

E4EXC $\text{C}(=\text{O})[\text{OH}1]$

E5EXC $\text{C}(=\text{O})[\text{SH}1]$

E6 $\text{C}(=\text{O}, \text{S})(\text{N})\text{Oc}$

E7 e.g., aryl carbamates $\text{C}1(=\text{O})\text{NS}(=\text{O})(=\text{O})[\text{C}, \text{c}] =, :[\text{C}, \text{c}]1$

E8SUB $\text{P}(=\text{O})[\text{O}, \text{S}]$

E9EXC $\text{P}[\text{OH}1]$

E10 $[\text{E}8\text{SUB}; !\text{E}9\text{EXC}]$

E11 $\text{c}(=\text{O})(\sim\text{c})\sim\text{c}$

E12EXC $[(\text{c}1(=\text{O})\text{ccn}([\text{C}, \text{c}])\text{cc}1), (\text{c}1(=\text{O})\text{n}([\text{C}, \text{c}])\text{cccc}1)]$

E12 $[\text{E}11; !\text{E}12\text{EXC}]$

E13 imides $\text{C}(=\text{O})-\text{N}-\text{C}=\text{O}$

ELEC $[\text{E}1, \text{E}2, \text{E}3, \text{E}6, \text{E}7, \text{E}10, \text{E}12, \text{E}13; !\text{E}4\text{EXC}; !\text{E}5\text{EXC}]$

N1 primary amines $[\text{NH}_2][\text{CX}_4]$

N2 secondary amines $[\text{NH}](\text{CX}_4)[\text{CX}_4]$

N4EXC $\text{N}=[\text{O}, \text{C}, \text{N}, \text{S}]$

N5EXC $\text{N}-[\text{C}, \text{c}, \text{N}]=[\text{C}, \text{c}, \text{N}, \text{n}, \text{O}, \text{S}]$

N6 [OH1][C,c,N;!\$(C=O)]

N7EXC [OH1]C=C

N8EXC [OH1]NC=[O,S]

N9 [\$([NH2]!:c),\$([NH1]([CX4])!:c),\$([NH0]([CX4])([CX4])!:c)]

N4 – N9 includes alcohols, hydroxylamines but excludes e.g., carboxylic acids

NUC [\$N1,\$N2,\$N6,\$N9;!\$N4EXC;!\$N5EXC;!\$N7EXC;!\$N8EXC]