

## **Supporting Information**

### **New substructure filters for removal of pan assay interference compounds [PAINS] from screening libraries and for their exclusion in bioassays**

Jonathan B. Baell\* and Georgina A. Holloway

#### **Table of Contents**

pp. S3-S7:	Table S1. Functional Group Filters (#) Used For the WEHI 93K HTS Library to Remove Inappropriate Compounds
pp. S8-S12:	Table S2. The six high throughput screening campaigns selected for analysis, and typical assay protocol
p. S13:	Table S3. Data for one particular validated hit, showing that the profile of an optimizable hit may not always be clean
pp. S14-S21:	Table S4. Structural classes with examples of the 362 compounds that registered as hits all six HTS campaigns selected for study
p. S22:	Table S5. Assay counts for benign versus problematic groups showing the greater number that hit 2-6 assays when expressed as a percentage of those that hit no assays ("Enrichment")
p. S23:	Table S6. Filter Family A ("Freq_Hit_5_morethan150.hits")
pp. S23-S25:	Table S7. Filter Family B ("Freq_Hit_5_less than150.hits")
pp. S25-S37:	Table S8. Filter Family C ("Freq_Hit_5_less than15.hits")
pp. S38-S41:	Table S9. Enrichment calculations for all problematic substructures with 4 or more member compounds. These are listed in order of population size.
pp. S42-S43	Table S10. Functional group prevalence in different screening libraries.
pp. S44-S45	Table S11. Rhodanines as problematic screening compounds, with discussion.
p. S46	Table S12. Phenolic mannich bases as problematic screening compounds, with discussion.
p. S47	Table S13. Hydroxyphenylhydrazones as problematic screening compounds, with discussion.
p. S48	Table S14. Alkylidene barbiturates as problematic screening compounds, with discussion.
p. S49	Table S15. Alkylidene-bearing heterocycles as problematic screening compounds, with discussion.

p. S50	Table S16. Fused tetrahydroquinolines as problematic screening compounds, with discussion.
p. S50	Table S17. Pyrroles as problematic screening compounds, with discussion.
p. S51	Table S18. Benzofurazans as problematic screening compounds, with discussion.
p. S52	Table S19. 2-Amino-3-carbonylthiophenes as problematic screening compounds, with discussion.
p. S53	Table S20. Catechols as problematic screening compounds, with discussion.
p. S53	Table S21. Quinones as problematic screening compounds, with discussion.
p. S54	Table S22. Azo as problematic screening compounds, with discussion.
p. S54	Table S23. Cyanopyridones as problematic screening compounds, with discussion.
p. S55	Table S24. Divinylketones as problematic screening compounds, with discussion.
p. S56	Table S25. Indoles as problematic screening compounds, with discussion.
p. S57	Table S26. Tertiary anilines as putative AlphaScreen®-specific problematic screening compounds.
p. S58	Table S27. Activity of certain anilines and olefins in the AlphaScreen® assay.
pp. S59-S60	Drugs containing PAINS substructures - extended discussion.
pp. S61-S84:	Figure S1. Assay interference substructures
pp. S85-S92:	Figure S2. Some drugs recognized by our filters for assay interference

TABLE S1. Functional Group Filters(#) Used For the WEHI 93K HTS Library to Remove Inappropriate Compounds. Subsequent Groups Used for the CTX 136K HTS Library are double hashed (##). These are currently blocked out but it is recommended that they be activated (unblocked) prior to the purchase of new libraries as a more expansive set of filters

```
#SYBYL/3DB HITLIST
#
#
#
#@CLASS REGLIST
#@DATABASE NONE
#@SOURCE in-house
#
#One hash means in original filter for WEHI 93K HTS Library
##Two hashes mean added for CTx 136K HTS Library (but blocked out here)
##The added ones may be new or simply broadened
##Many definitions are dependent on others (e.g. if allow sulfonates, others defs need modification)
#acid halide
C(=O)Hal
##...and related
##C(=S)Hal
#any het halide (includes sulfonyl halides)
HetHal
##acrylamide and related, including vinylpyridines etc:
##C(=O)C=CH2
##S(=O)C=CH2
##N#CC=CH2
##C(=CH2)C[5]:N:Hev:Hev:Hev:Hev:@5
##...and related
##C(=S)C=CH2
##acyl carbamate and related hot linear aryl carbonates/carbamates
##OK to go
##C(=O)NC(=O)Any[IS=O,S]
HetC(=O)O-[!R]C[8]:Hev(Any[NOT=O,S[TAC=2],C[TAC=4],N[TAC=3]]):Hev:Hev(Any[IS=Hal,C#N,C(F)(F)F,S(=O)=O,C(=O)&NOT=C(=O)OH]):Hev:Hev(Any[NOT=O,S[TAC=2],C[TAC=4],N[TAC=3]]):@8
HetC(=O)O-[!R]C[8]:Hev(Any[IS=Hal,C#N,C(F)(F)F,S(=O)=O,C(=O)&NOT=C(=O)OH]):Hev:Hev(Any[NOT=O,S[TAC=2],C[TAC=4],N[TAC=3]]):Hev:Hev(Any[NOT=O,S[TAC=2],C[TAC=4],N[TAC=3]]):@8
##acyl hydrazone (and oxime) not in ring:
##leave in (as well as carbazides, semicarbazides,
##acylsemicarbazides, hydrazones, acyl hydrazides)
##despite some reservation of Metal chelating for some of these
##O=CHetN=[!R]C(C)Any[IS=C,H]
##acyl hydrazide not in ring -leave in: see above
##O=CNH-[!R]NHC(=O)Het
##acyl imide - taken care of by imide filter
##O=CN(C(=O))C=O
##N-alkyl acyl sulfonamide
##O=C-[!R]N(C)S(=O)(=O)C
#acyl and ##sulfo cyanide
C(=O)C#N
##S(=O)C#N
##C(=S)C#N taken care of by thiocarbonyl filter
##acylimines
##Any[NOT=O,N]C(Any[NOT=O,N])=[!R]NC=O
#aldehyde
C(=O)H
##C(=S)H taken care of by thiocarbonyl filter
##aliphatic long chain
C[!r]H2CH2CH2C[!r]H2CH2CH3
#alkyl halide
BrC[TAC=4]H
ClC[TAC=4]H
##BrC[TAC=4]
##ClC[TAC=4]H2
##amidotetrazole
##N[1](C=NN=N@1)C=O
###aminals etc not in ring unless basic amine - harshest criteria - could revisit
##HetC[TAC=4]-[!R]Het
##HetC[TAC=4]N(Any[IS=H,C[TAC=4]])Any[IS=H,C[TAC=4]]
```

#anhydride/thioanhydrides  
C(=Any[IS=O,S])Any[IS=O,S]C=Any[IS=O,S]  
##anilines - not removed; medchem issue: specifics dealt with in frequent hitter filter  
##azepanes - not removed  
#azido/diazonium/diazo - specifics also in frequent hitters:  
N~N[f]  
##azo (and carbazones etc)  
##N[!r]=N  
#aziridines, epoxides etc  
N[1]CC@1  
##Hev[1]HetC@1  
##benzidine - not removed; medchem issue  
##betalactams/lactones, not thietanes, oxetanes, azetidines etc  
##Hev[1]C(=O)HetC~@1  
##beta-keto - in diketos  
##beta-amino ketones and related  
##CC[!r](=O)CH2CH2Het  
##CS[!r](=O)CH2CH2Het  
##biotin - not now  
##S[1]CH2C[5]HCH(CH@1CH2CH2CH2CH2C=O)NHC(NH@5)=O  
#boron  
B  
#carbamic acids  
NC(=O)OH  
##carbazide - leave in  
##NHC(=O)-[!R]NHNH  
##carbocation/anion and other wrongly charged atoms  
##checked to make sure sulfoxide retained  
##C[+1]  
##C[-1]  
S[+1]  
O[+1]  
O[TAC=3]  
Hal[+1]  
S(Any)(Any)(Any)(Any)Any  
##catechol  
##C[1]:C(:C:C:C(:C:@1)OH)OH  
##chromones - reactive; in Michael acceptors  
##crown ethers - main ones - not now  
##HetC~CHetC~CHetC~CHetC~CHetC~CHetC~C  
##coumarins - reactive; in Michael Acceptors; other specific ones in frequent hitters  
#cyanamides  
CH2NC#N  
#cyanate/thiocyanate  
OC#N  
SC#N  
#cyanohydrin  
N#CCOH  
##cycloalkanes - not removed  
#cyclohexadienes  
C[1]HC=CC=CCH@1  
C[1]HC=CCH2C=C@1  
##dialkynes  
##C#CC#C  
##diaminobenzenes - specific ones in frequent hitters  
##diketo (includes o-quinones) and beta-diketo and related hot keto  
##CC(=Het[!r])C(=Het[!r])C  
##CC(=O)C(=O)Any[IS=O,N]  
##CC(=Het[!r])S(=O)(=O)  
##CC(=O)C(=S)C - covered by thiocarbonyl  
##CC(=O)C[TAC=4]C(=O)C  
##CC(=O)C[TAC=4]S(=O)C  
##CC(=O)C[TAC=4]C#N  
##CC(=O)C[TAC=4]C(Hal)(Hal)Hal  
##disulfide and related  
S[TAC=2&!r]S[TAC=2&!r]  
SSS  
##enamines - in frequent hitters

##ester of hobt and su (and carbamates/carbonates) A

C(=O)O[!r]N:Hev

##C(=O)O[!r]NC(=O)

##esters - thio-type; some accounted for by thiocarbonyl removal

##CC(=S)O[!r]C

CC(=O)S[!r]C

##C(=S)S

##esters - aryl B

CC(=O)O[!r]C[TAC=3]

##ester - hot benzyl, non aryl C

##C[TAC=4]C(=O)O[!r]CH2C[8]:Hev(Any[NOT=O,S[TAC=2],C[TAC=4],N[TAC=3]]):Hev:Hev(Any[IS=Hal,C#N,C(F)(F)F,S(=O)=O,C(=O)Hev[NOT=OH]]):Hev:Hev(Any[NOT=O,S[TAC=2],C[TAC=4],N[TAC=3]]):@8

##C[TAC=4]C(=O)O[!r]CH2C[8]:Hev(Any[IS=Hal,C#N,C(F)(F)F,S(=O)=O,C(=O)Hev[NOT=OH]]):Hev:Hev(Any[NOT=O,S[TAC=2],C[TAC=4],N[TAC=3]]):Hev:Hev(Any[NOT=O,S[TAC=2],C[TAC=4],N[TAC=3]]):@8

##esters - other labile D

##CC(=O)OCHC(Hal)(Hal)Hal

##C[TAC=4]C(=O)O[!r]CHC(=O)Any[NOT=OH]

##C[TAC=4]C(=O)O[!r]CHC#N

##esters - other labile D

##COC[!r](=O)C(=O)

##COC(=O)S(=O)(=O)

##COC[!r](=O)C[TAC=4]S(=O)=O

##COC[!r](=O)C[TAC=4]C#N

##COC[!r](=O)C[TAC=4]C(Hal)(Hal)Hal

##COC[!r](=O)C[TAC=4]N(C(=O))C(=O)

##over halogenated rings

##1,2

##C[1]:C(:C(:Hev:Hev:Hev:@ 1)Br)Any[IS=Cl,Br]

##1,3

##C[1]:C(:Hev:C(:Hev:Hev:@ 1)Br)Any[IS=Cl,Br]

##1,4

##C[1]:C(:Hev:Hev:C(:Hev:@ 1)Br)Any[IS=Cl,Br]

##1,2,4

##Hev[1]:C(:C(:Hev:C(:Hev:@ 1)Hal)Hal)Hal

##1,2,3

##C[1](~C(:C(:Hev:Hev:Hev:@ 1)Hal)Hal)Hal

##1,3,5

##C[1](~C:C(:C:C(:C:@ 1)Cl)Cl)Cl

##hydrazine and other nucleophilic NH2

Any[IS=O,N]NH2

##hydroxylamine

##HON[TAC=3]C[TAC=4]

##hydroquinone - more related ones in frequent hitters

##C[1]:C:C:C(:C(:C:@ 1)OH)OH

##hydroxamic acid

##C(=O)NOH

##imide

##CC(=O)N-[!R]C(=O)C

###imidoyl chlorides

N=CHal

#imines -reactive ones

##Any[NOT=N,O]N=C[!r](Any[NOT=N])Any[NOT=N]

###imines - not completely clear whether ok or not; may need to reexamine in second pass

##Any[NOT=O,N]N=[!R]C(Any[IS=C,H])Any[IS=C,H]

##Any[NOT=O,N]N=[!R]C(C[TAC=4])C

##iodine

##I

#isocyanate, isothiocyanate - dealt with in ketenes

#ketene - includes allenes, carbodimides, isocyanates etc

Any=C=Any

##maleimides and surrogates

##C[1]C(NC(CH=@ 1)=O)=O

##C[1](CH2C(NC@ 1=O)=O)HetC[TAC=3]

##michael acceptors - more in frequent hitter file

CCH=C[!r]HC(=O)C

CCH=C[!r](C#N)C#N

CCH=C[!r](C(=O)Hev~Hev)C#N

##first, alpha-beta unsubstituted, unsaturated nitriles and ketones

##CCH=[!R]C(Any[IS=H,C])Any[IS=C#N,C(=O)&NOT=C(=O)Any[IS=N,O]]

##second, linear diactivated A:

##Any[IS=C,O]CH=[!R]C(Any[IS=C(=O),S(=O),C#N,Hal,C(Hal)(Hal)Hal&NOT=C(=O)OH])Any[IS=C(=O),S(=O),C#N&NOT=C(=O)OH]  
## linear diactivated B:  
##Hev[NOT=OH]C(=O)CH=[!R]CHC(=O)Hev[NOT=OH]  
## linear diactivated C:  
##Any[IS=C#N,C(=O)Hev[NOT=OH],S(=O)]CH=[!R]C(Any[IS=H,C])C[8]:Hev(Any[NOT=O,S[TAC=2],C[TAC=4],N[TAC=3]]):Hev:Hev(Any[IS=Hal,C#N,C(F)(F)F,S(=O)=O,C(=O)Hev[NOT=OH]]):Hev:Hev(Any[NOT=O,S[TAC=2],C[TAC=4],N[TAC=3]]):@8  
##Any[IS=C#N,C(=O)Hev[NOT=OH],S(=O)]CH=[!R]C(Any[IS=H,C])C[8]:Hev(Any[IS=Hal,C#N,C(F)(F)F,S(=O)=O,C(=O)Hev[NOT=OH]]):Hev:Hev(Any[NOT=O,S[TAC=2],C[TAC=4],N[TAC=3]]):Hev:Hev(Any[NOT=O,S[TAC=2],C[TAC=4],N[TAC=3]]):@8  
###thirdly, cyclic: activated coumarins; could revisit some of the less activated amides  
##C[1](C(=O)OC[5]:C:C:C:C:@5CH=@1)Any[IS=C(=O),C(=S),S(=O),C#N,Hal,C(Hal)(Hal)Hal&NOT=C(=O)OH]  
##and activated chromones;- could revisit some amides  
##C[1](=CHOC[5]:C:C:C:C:@5C@1=O)Any[IS=C(=O),C(=S),S(=O),C#N,Hal,C(Hal)(Hal)Hal&NOT=C(=O)OH]  
##mercaptothiazoles and related  
##S[TAC=2]C[2]:S:C:Any:N:@2  
##N-acyl and quaternary alkyl pyridines etc  
##CN[+1]:Hev -redundant by definition below  
##Broaden to include N-nitros, N oxides etc: can review  
##N[+1]:Hev  
##N-acyl pyrroles, imidazoles etc not in ring, but not ureas, carbamates  
##C[!r](=O)N(:Hev):Hev  
##naphthylamines -alpha and beta; leave in; medchem issue  
##C[1]:C:C[3]:C(:C:C:@1):C:C:C:C:@3-[!R]N  
##C[1]:C:C[3]:C(:C:C:@1):C:C:C(:C:@3)!R]N  
##C[1]H:CH:C[5]:C(:CH:CH:@1):CH:CH:CH:C:@5NH  
##C[1]H:CH:C[5]:C(:CH:CH:@1):CH:CH:CH:C:@5N(Any[IS=H,C(=O)C])Any[IS=H,C(=O)C]  
##C[1]H:CH:C[5]:C(:CH:CH:@1):CH:CH:C(:CH:@5)NH  
##nitro, including N-NO2 etc:  
##HevN[!r](~O[f])~O[f]  
CN[!r](~O[f])~O[f]  
#nitroso  
N[TAC=2]=O  
##N oxide and N-hydroxypyridine  
##Hev:N[+1]~[!R]O  
##N-oxide: broader  
##NO[f]  
##oxetanes, thietanes (see beta lactams)  
##oximes: not removed  
##O-N single bond: not removed  
#peroxide  
OO  
##phenyl carbonate and carbamate - hot ones removed above  
##phosphorous  
##P  
##phthalimides: not removed (except hot ones - above in triaryls)  
##polycyclic aromatic: not removed  
##polyene: not removed  
##quaternary nitrogen  
##CN[+1](C)(C)C  
##saponins - not removed  
#silicon  
Si  
##stilbenes - not removed  
##sulfur oxygen single bond (includes sulfinic/sulfonic acids/esters)  
##SO[!f]  
##sulfur-nitrogen single bond, exluding sulfonamides  
##NS[TAC=2]  
##sulfur-nitrogen single bond surrogate  
##C[1](C(=O)N):C(:C:C:C:C:@1)SCCC(=O)Any[NOT=OH]  
##thiocarbonyl  
##C=S  
#thiols  
SH  
##trichloromethyl  
##C(Cl)(Cl)Cl  
##tri (and di-) haloketones and amides (+esters) and related  
##C(=O)C[!r](Hal)Hal  
#trihaloimide  
OC(=NH)C(Hal)(Hal)Hal  
##isonitrile  
##N[TAC=2]~C[TAC=1]

##sulfone-activated rings and related below  
##2-halo and 2-sulfonylpyrimidines, but deactivated triazines OK  
##Any[IS=Hal,S(=O)(=O)]C[1]:N:Any(Any[NOT=N]):Any:Any(Any[NOT=N]):N:@ 1  
##4-halo and 4-sulfonylpyrimidines not too deactivated  
##C[1](Any[NOT=N]):N:Any(Any[NOT=N]):Any:Any(Any[IS=Hal,S(=O)(=O)]):N:@ 1  
##4-activated 2-halo and 2-sulfonyl pyridines  
##C[1]:N:Any(Any[IS=Hal,S(=O)(=O)C]):Any:Any(Any[IS=Hal,C(C)=NO,C#N,C(=O),C(F)(F)F,S(=O)=O&NOT=C(=O)OH]):Any:@ 1  
##2-activated 4-halo and 4-sulfonyl pyridines  
##C[1]:N:Any(Any[IS=Hal,C(C)=NO,C#N,C(=O),C(F)(F)F,S(=O)=O&NOT=C(=O)OH]):Any:Any(Any[IS=Hal,S(=O)(=O)C]):Any:@ 1  
##3-activated 2-halo and 2-sulfonyl pyridines  
##C[1]:N:Any(Any[IS=Hal,S(=O)(=O)C]):Any(Any[IS=Hal,C(C)=NO,C#N,C(=O),C(F)(F)F,S(=O)=O&NOT=C(=O)OH]):Any:Any:@ 1  
##5-activated 2-halo and 2-sulfonyl pyridines  
##C[1]:N:Any(Any[IS=Hal,S(=O)(=O)C]):Any:Any:Any(Any[IS=Hal,C(C)=NO,C#N,C(=O),C(F)(F)F,S(=O)=O&NOT=C(=O)OH]):@ 1  
##6-activated 2-halo and 2-sulfonyl pyridines and not too deactivated  
##C[1](Any[IS=Hal,C(C)=NO,C#N,C(=O),C(F)(F)F,S(=O)=O&NOT=C(=O)OH]):N:Any(Any[IS=Hal,S(=O)(=O)C]):Any(Any[NOT=N]):Any(Any[NOT=N]):Any:@ 1  
##3-activated 4-halo and 4-sulfonyl pyridines  
##C[1]:N:Any:Any(Any[IS=Hal,C(C)=NO,C#N,C(=O),C(F)(F)F,S(=O)=O&NOT=C(=O)OH]):Any(Any[IS=Hal,S(=O)(=O)C]):Any:@ 1  
##1,2,4-activated aryls (including activated phthalimides  
##C[1](Any[NOT=N,O]):C(Any[IS=Hal,C#N,C(=O),C(F)(F)F,S(=O)=O&NOT=C(=O)OH]):C(Any[IS=Hal,S(=O)(=O)C,C[r](=O)NC]):C(Any[NOT=N,O]):C(Any[NOT=N,O]):C(Any[IS=Hal,C#N,C(=O),C(F)(F)F,S(=O)=O&NOT=C(=O)OH]):@ 1  
##1,2,3-activated aryls  
##C[1](Any[IS=Hal,C#N,C(=O),C(F)(F)F,S(=O)=O&NOT=C(=O)OH]):C(Any[IS=Hal,S(=O)(=O)C,C[r](=O)NC]):C(Any[IS=Hal,C#N,C(=O),C(F)(F)F,S(=O)=O&NOT=C(=O)OH]):C(Any[NOT=N,O]):C(Any[NOT=N,O]):C(Any[NOT=N,O]):@ 1  
##activated pyrimidine phenol ethers:  
##C[1](Any[NOT=O][TAC=2],N[TAC=3]]):N:C(:Hev:Hev:N:@ 1)-[!R]Any[IS=O,S]C[8]:Hev(Any[NOT=O,S][TAC=2],C[TAC=4],N[TAC=3]):Hev:Hev(Any[IS=Hal,C#N,C(F)(F)F,S(=O)=O]):Hev:Hev(Any[NOT=O,S][TAC=2],C[TAC=4],N[TAC=3]):@ 8  
##C[1](Any[NOT=O][TAC=2],N[TAC=3]]):N:C(:Hev:Hev:N:@ 1)-[!R]Any[IS=O,S]C[8]:Hev(Any[IS=Hal,C#N,C(F)(F)F,S(=O)=O]):Hev:Hev(Any[NOT=O,S][TAC=2],C[TAC=4],N[TAC=3]):Hev:Hev(Any[NOT=O,S][TAC=2],C[TAC=4],N[TAC=3]):@ 8  
##C[1](Any[NOT=O][TAC=2],N[TAC=3]]):N:C(:N:Hev:Hev[NOT=N]:@ 1)-[!R]Any[IS=O,S]C[8]:Hev(Any[NOT=O,S][TAC=2],C[TAC=4],N[TAC=3]):Hev:Hev(Any[IS=Hal,C#N,C(F)(F)F,S(=O)=O]):Hev:Hev(Any[NOT=O,S][TAC=2],C[TAC=4],N[TAC=3]):@ 8  
##C[1](Any[NOT=O][TAC=2],N[TAC=3]]):N:C(:N:Hev:Hev[NOT=N]:@ 1)-[!R]Any[IS=O,S]C[8]:Hev(Any[IS=Hal,C#N,C(F)(F)F,S(=O)=O]):Hev:Hev(Any[NOT=O,S][TAC=2],C[TAC=4],N[TAC=3]):Hev:Hev(Any[NOT=O,S][TAC=2],C[TAC=4],N[TAC=3]):@ 8  
##activated 4-fluorosulphonamides  
##C[1](S(=O)(=O)NC:N):CH:CH:C(F):CH:CH:@ 1  
##thiazolinium halides and related  
##Het[1]:C(Hal):N:C:C:@ 1  
##reactive haloenamines  
##CC(C)=C(Hal)N(Any[IS=H,C[TAC=4]])Any[IS=H,C[TAC=4]]  
##terminal acetylenes  
##C#CH  
#####  
##Limiting number of certain slns#  
#####  
#  
##furans and thiophenes##  
##  
##S[1]:C:CH:CH:CH:@ 1<max=1>  
##O[1]:C:CH:CH:CH:@ 1<max=1>  
#  
##nitriles  
##C#N<max=2>  
#acids  
C(=O)OH<max=2>  
#  
##esters  
##C(=O)OC<max=2>  
#  
#thioethers  
C[TAC=4]S[TAC=2]C[TAC=4]<max=2>  
#  
#halogens  
Br<max=1>  
Cl<max=3>  
F<max=5>

**Supplementary Table S2.** The six high throughput screening campaigns using the WEHI 93K HTS library selected for analysis<sup>c</sup>

<b>HTS Campaign<sup>a</sup></b>	<b>Target Type</b>	<b>Detection method / anchor</b>	<b>Nature of interaction</b>	<b>Number of hits<sup>b</sup></b>
<b>A</b>	Protein-protein	AlphaScreen® / anti-GST Mab-conjugated acceptor beads, streptavidin-coated donor beads	Binding of GST-protein target to an amphipathic helical ligand	3,006
<b>B</b>	Protein-protein	AlphaScreen® / protein A-conjugated acceptor beads, streptavidin-coated donor beads.	Binding of receptor construct IL13R $\alpha$ 1-Fc to biotinylated cytokine IL13	4,086
<b>C</b>	Protein-protein	AlphaScreen® / anti-GST Mab-conjugated acceptor beads, streptavidin-coated donor beads	Binding of GST-protein target to an amphipathic helical ligand	3,145
<b>D<sup>c</sup></b>	Protein-protein	AlphaScreen® / anti-GST Mab-conjugated acceptor beads, streptavidin-coated donor beads	Binding of GST-protein target to an amphipathic helical ligand	746
<b>E</b>	Protein-protein	AlphaScreen® / nickel-coated acceptor beads, streptavidin-coated donor beads	SH2-binding domain: hexaHis-SOCS2 binding to biotinylated gp130	9,309
<b>F</b>	Protein-protein	AlphaScreen® / nickel-coated acceptor beads, streptavidin-coated donor beads	SH2-binding domain: hexaHis-SOCS3 binding to biotinylated gp130	14,745

a. We are not at liberty to reveal the targets for HTS Campaigns A, C and D.

b. Primary hits, classified as any compound that gave an inhibitory signal of 50% or more at the screening concentration tested (50 $\mu$ M except for HTS Campaign D {25 $\mu$ M}, single point determination).

c. A typical assay protocol is described in the standard operating procedure below. This particular protocol is used for follow-up from HTS Campaign D, where we delineate between the host protein as protein D and the competitive ligand as ligand D or peptide D. However, the protocol is essentially the same as that used as the primary screen against the WEHI 93K HTS library in HTS Campaign D, and is closely related to that used for HTS Campaigns A and C, whereas HTS Campaigns E and F used a different linking methodology. This involved the use of Nickel-HexaHis



and we have found that this gives rise to the highest hit rates most plausibly through the additional detection of mildly chelating compounds that interfere with the Nickel anchor. We have therefore abandoned this approach in our screening centre. HTS Campaign B also used a different linking technology as indicated.

## **Standard Operating Procedure: Routine AlphaScreen™ Assay using Proxiplates**

### **1. Introduction**

The aim of the routine Protein D AlphaScreen™ assay is to identify active small molecules against protein D. The compounds tested are generally synthesized in-house, however some outsourced compounds are also investigated. The screens are used to determine the potency of each compound. A GST-counter screen assay is also done to determine the verity of the results ie. protein D inhibition. To determine an accurate estimation of the IC<sub>50</sub>, the compounds are routinely tested at starting concentrations, 100µM and/or 1µM and serially titrated 3 fold over 11 dilutions.

The assay was developed using AlphaScreen™ technology which relies on hydrogel coated acceptor and donor beads which have functional groups for conjugation to a protein (GST-protein D or GST-Biotin) and a peptide (Biotin-peptide D) respectively. The beads come in close proximity when the protein and the peptides interact. Donor beads contain a photosensitiser that converts oxygen to an excited form of O<sub>2</sub> at an excitation of 680 nm. Energy is transformed from the singlet oxygen and reacts with chemiluminescers on the acceptor bead, resulting in light emission at 520 - 620 nm. Active library compounds when added to the reaction, reduce the intensity of the luminescence, dependent on the inhibition of proximity of the acceptor and donor beads. With this information, the IC<sub>50</sub> of each compound can be calculated

### **2. Assay Summary**

#### **2.1 Reagents and Materials**

GST-protein D and biotinylated GST proteins were prepared and provided by Peter Czabotar from Parkville which were stored as stock solutions at -80°C. The biotinylated peptide D was purchased from Aussep and stored as 500µM stock solutions in 100% DMSO at -80°C. The AlphaScreen™ GST (Glutathione-S-Transferase) Detection Kit was obtained from Perkin Elmer Lifesciences (Cat #6760603R). The Proxiplates, white 384 well flat-Bottom plates were purchased from Interpath Services, Melbourne (Cat #784075). The seals to cover the plates were purchased from Proscience, Melbourne (Cat#784075). DMSO was purchased from AnalaR. The 384 deep well plates and the Polypropylene 50 µL, V bottom polypropylene compound plates were purchased from Matrical.

#### **2.2 Preparation of compounds**

The chemists provided the compounds in the form of dry powder which were stored at -20°C. 10mM stocks were made with 100% DMSO the day before the assay was scheduled to be run. 12µL of 100% DMSO and 6µL of 10mM compound (ie. 3.333mM, final 100µM) was added to columns 1 and 12 in the Polypropylene 50 µL, V bottom compound plates. To achieve a final compound concentration of 1µM, in a separate matrical plate, 28µL of 100% DMSO and 2µL of 10mM compound was added to a well, mixed well, 2µL of this solution was taken and added to 38µL of 100% DMSO. 20µL of this solution was added to the test metrical plate. For the control wells 15µL 100% DMSO only was added to Lanes 23 and 24 of each plate.

#### **2.3 Control Compound Addition**

Several control compounds were included in the test plates. The control compounds used routinely are listed as follows.

[Positive Control]	[Stock]	[Start Conc]	[IC <sub>50</sub> , Protein D]
Control 1	0.5mM	15µM	0.06-0.2µM
Control 2	10mM	100µM	3-18µM
Control 3	5mM	150µM	0.7-1µM

#### **2.4 Titration of Compounds Automation**

The compound plates were serially diluted 2 fold using the MiniTrak *Titration\_11pt\_P30.wpt* protocol located in the C:\Packard\Plate Track\Bin folder.

- The plates were centrifuged prior to dilution.
- The Mini Trak program was opened and executed and the robot was set up with the following parameters:

Enter number of plates to be titrated:

**Parameters for Volumes**

Diluent to aspirate:	16µL
Sample to aspirate:	8µL
Volume to Dispense:	24µL
Volume to mix:	20µL

**Parameters for Stacker Setup**

Compound Plate/s:	Stacker 1 front
-------------------	-----------------

**Parameters for Deck Setup**

Tip Carrier:	MPD 4 (hooks facing left)
P30 Tips (only in column 1 and 12):	MPD 4
DMSO reservoir: 100% DMSO	MPD 6

\*50% DMSO is required to wash tips

- c. Once titrations are complete, the compound plate was immediately covered with a foil seal to prevent evaporation.

## 2.5 Buffer Preparation

The assay and bead buffers were prepared fresh on the day. Each titrated compound plate was assayed in duplicate. The following volumes were sufficient to run 12 Proxiplates (4 assay plates run in duplicate in each of protein D and counter assays)

**Assay Buffer**

[Stock]	[Final]	[Volume for 100 mL]
1M Hepes pH 7.4	50mM	5mL
1M DTT	10mM	1mL
4M NaCl	100mM	2.5mL
10% Tween-20	0.05%	0.5mL
10mg/mL Casein	0.1mg/mL	1mL
Milli-Q H <sub>2</sub> O		90mL

**Bead Buffer**

[Stock]	[Final]	[Volume for 100 mL]
1M Tris-HCL pH 7.5	50mM	5mL
10% Tween-20	0.01%	0.1mL
10mg/mL Casein	0.1mg/mL	1mL
Milli-Q H <sub>2</sub> O		93.9mL

## 2.6 Protein and Peptide Preparation

1. The assay and bead buffers were used to prepare the acceptor and donor solutions. Alphascreen™ beads are light sensitive and therefore prepared in a darkened room. 2.5 µL of beads were added per 1 mL of buffer.
2. The volume of protein or peptide added was calculated using the following formula:

$$\frac{C_1}{C_2} \times V_1 \times 2 = V_2$$

$C_1$  = Final Concentration of protein/peptide

$C_2$  = Stock Concentration of protein/peptide

$V_1$  = Total Volume of Acceptor/Donor Solution

$V_2$  = Volume of stock protein/Peptide to add to Acceptor/Donor solution

3. The assay components were prepared as separate Acceptor and Donor Solutions. The Acceptor Solution contains Acceptor beads and target protein, whilst the Donor Solution contains Donor beads and biotinylated peptide.

#### Protein D

[Acceptor Solution]	[mL]	[Donor Solution]	[mL]
Assay buffer	10mL	Assay buffer	10mL
Bead buffer	10mL	Bead buffer	10mL
Acceptor Beads	50µL	Donor Beads	50µL
114µM GST-protein D (or 3.51µL of a 1:10 dilution)	0.35µL	500µM Bt-Peptide D	0.32µL
<b>Final Protein</b>	<b>[1.0nM]</b>	<b>Final Peptide</b>	<b>[4nM]</b>

#### Counter-GST

[Acceptor Solution]	[mL]	[Donor Solution]	[mL]
Assay buffer	10mL	Assay buffer	8mL
Bead buffer	10mL	Bead buffer	8mL
Acceptor Beads	50µL	Donor Beads	50µL
77µM B-GST	1.04µL		
<b>Final Protein</b>	<b>[2nM]</b>		

4. When the solutions were prepared, they were left to incubate for 30 minutes at room temperature to allow the beads to bind to the protein and the peptide.
5. 5 µL of protein D solution or biotinylated-GST was added to columns 1-23 of the appropriate assay plates using the Multi-drop Combi (cassette #5). 5µL Assay/Bead buffer was added to column 24 (no protein).
6. Then selected *0.3 uL Tip transfer proxiplat\_stationary.wpt* program to Transfer 0.3µL of sample from the compound plate into each assay plate.

#### Parameters for Stacker Setup

Assay Plates:                      Stacker 3 front  
Compound Plates:                Stacker 2 rear

#### Parameters for Deck Setup

P10 Tips:                            MPD 4 (hooks facing left)

\*50% DMSO was needed to wash tips

7. Plates were then incubated for 30 mins at RT
8. 5µL of peptide D solution or donor solution (GST-counter) were added into assay plates (columns 1-24) using the Multi-drop Combi.
9. After final addition, tapped plates gently and sealed individually with adhesive film.
10. Plate were then incubated at RT for ~4hrs

#### 2.7 Assay Measurement

The plates were then loaded onto the right stacker of the Envision 2103 plate reader. Protocol *\*Alphascreen\_384well\_proxiplate\_Auto* was used to read the plates.

#### 2.8 Data Analysis

Data was displayed in an Excel format spreadsheet and imported into ActivityBase database. IC<sub>50</sub> for each compound calculated by using the *PROTEIN\_D\_IC50* and *COUNTER\_GST\_IC50* templates. The percent inhibition was calculated using the following equation:

$$\%Inhibition = 100 * \left(1 - \frac{(x - \mu^-)}{(\mu^+ - \mu^-)}\right)$$

$x$  = RFU obtained after compound treatment

$\mu^-$  = RFU obtained for the negative controls (no protein controls)

$\mu^+$  = RFU obtained for the positive controls (DMSO vehicle controls)

IC<sub>50</sub> values were then obtained by non-linear least squares fitting of the above data to XLfit3 equation 205:  
 $y = A + ((B - A) / (1 + ((C/x)^D)))$ .

The quality of the assay results were monitored by determination of the Z Prime factor for each assay plate, where Z Prime  $\Rightarrow$  0.5 for the results was considered as reliable (Zhang *et al*, J Biomol Screening, 4:67-73, 1999).

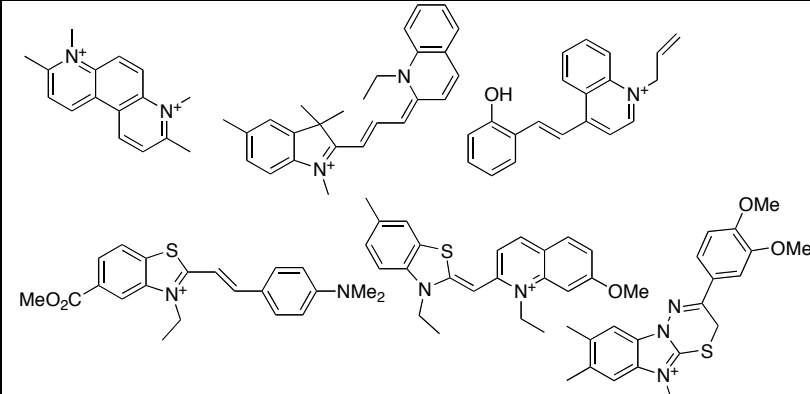
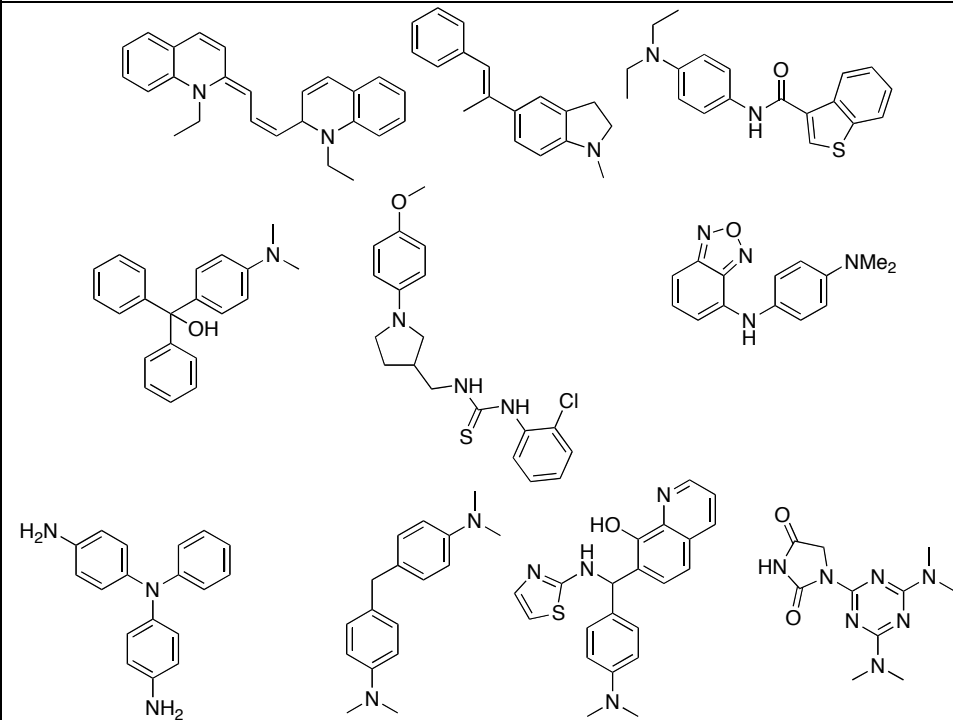
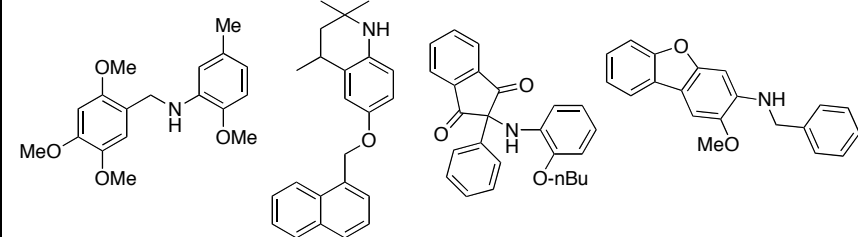
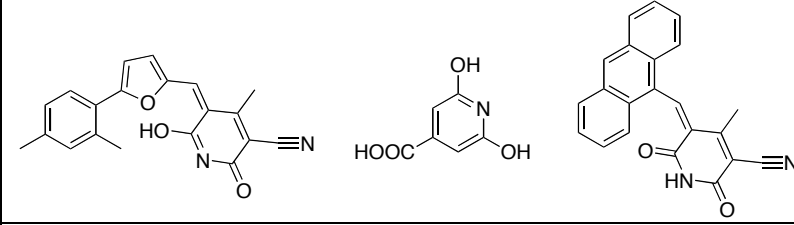
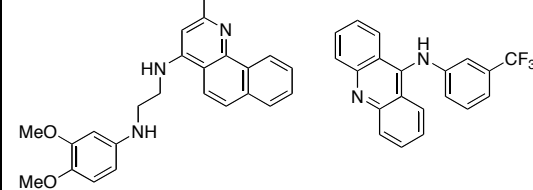
**Supplementary Table S3.** Data for one particular validated hit, showing that the profile of an optimizable hit may not always be clean

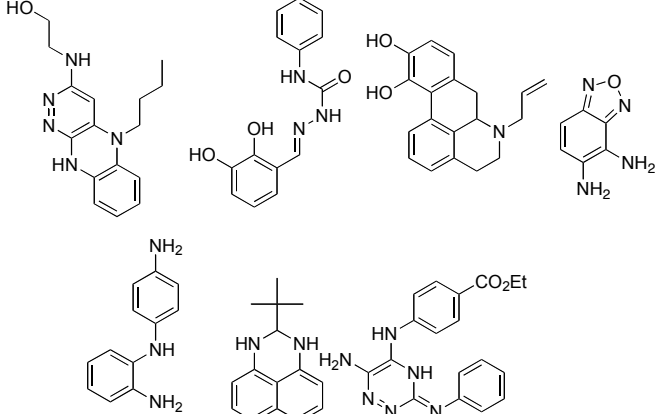
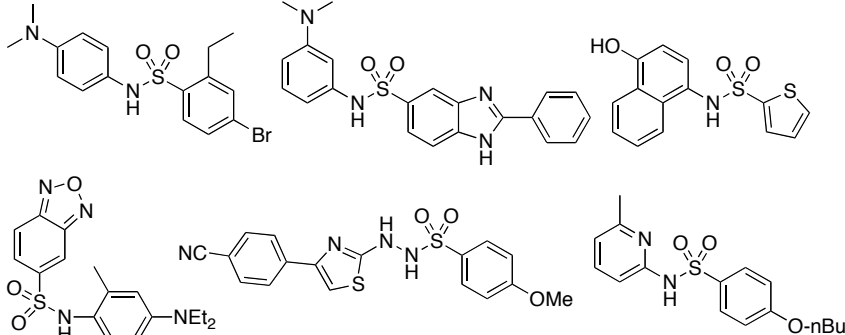
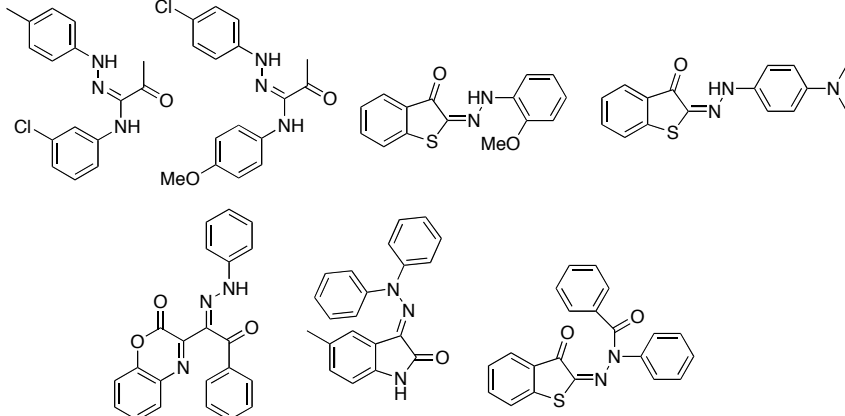
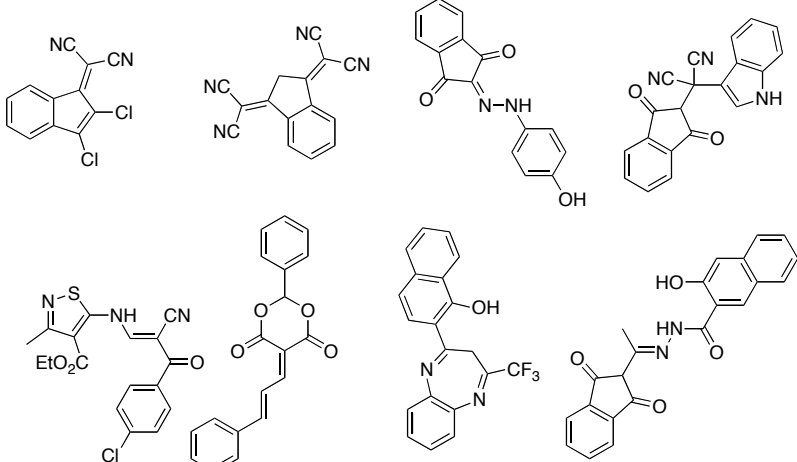
% inhibition at test concentration (50 $\mu$ M)						Count <sup>a</sup>
Screen A (50 $\mu$ M)	Screen B (50 $\mu$ M)	Screen C (50 $\mu$ M)	Screen D (25 $\mu$ M)	Screen E (50 $\mu$ M)	Screen F (50 $\mu$ M)	
<b>74</b>	<b>58</b>	<b>&lt;50</b>	<b>81</b>	<b>67</b>	<b>&lt;50</b>	<b>4</b>

a. In other words, this compound registered as a hit in four out of the six assays studied.

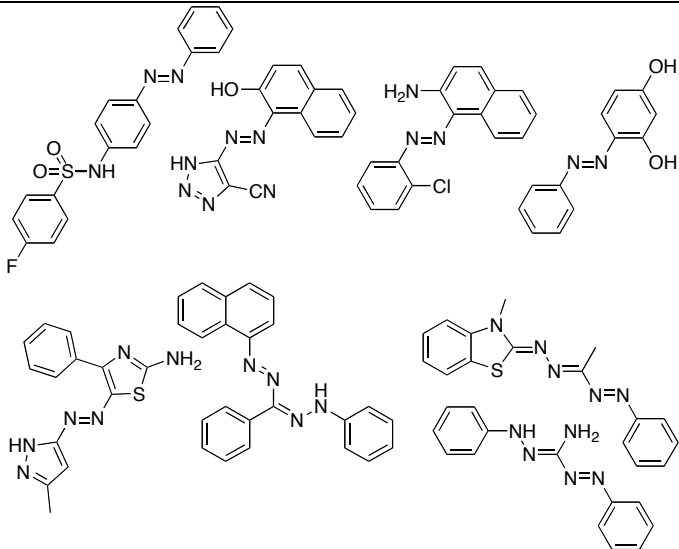
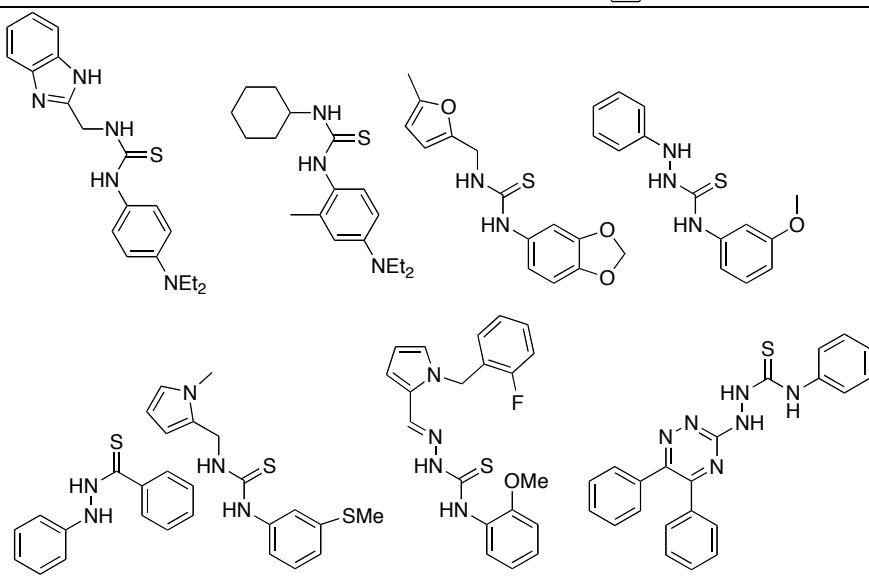
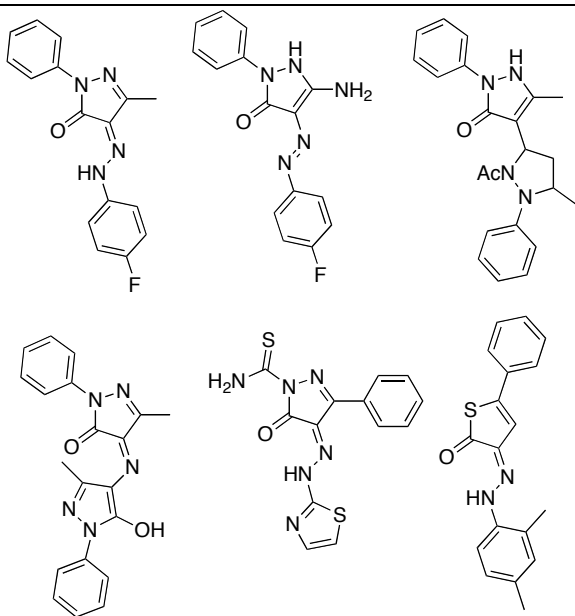
**Supplementary Table S4** - Structural classes with examples of the 362 compounds that qualified as hits all six HTS campaigns selected for study. Many of these compounds contain more than one identified assay interference moiety.

Class	Number	Examples
Quinone-like	51	
Rhodanine-like	50	
2-alkenyl furans	15	

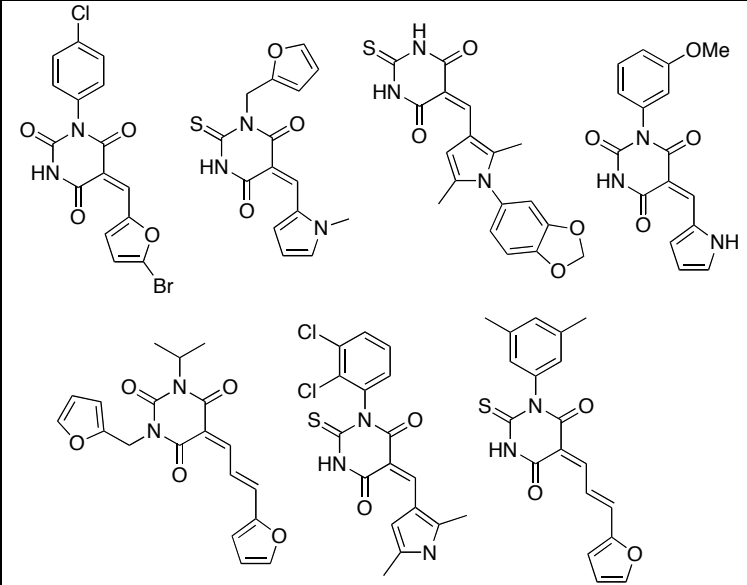
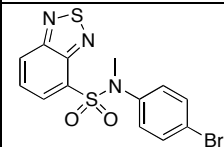
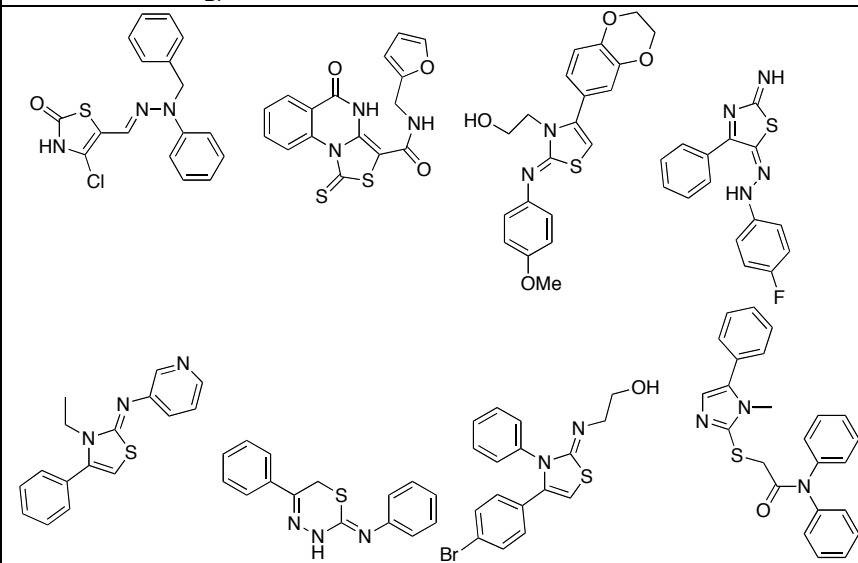
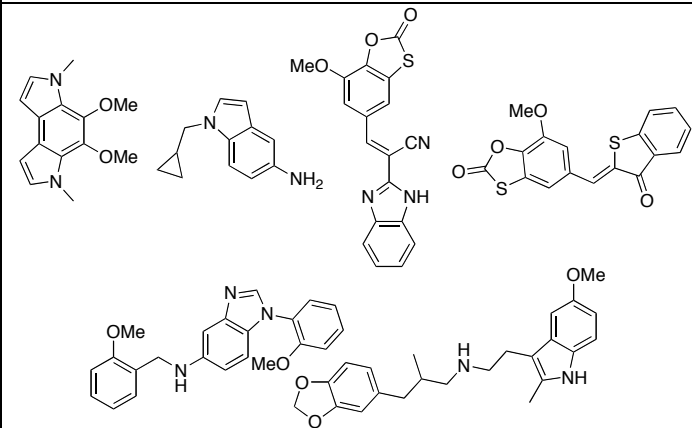
N-alkylated aromatic (charged) nitrogen {often dyes}	17	
Conjugated and aryl tertiary amines	65	
Anisidines	17	
2,6-dioxo--pyridine	3	
Fused amino-pyridines	5	

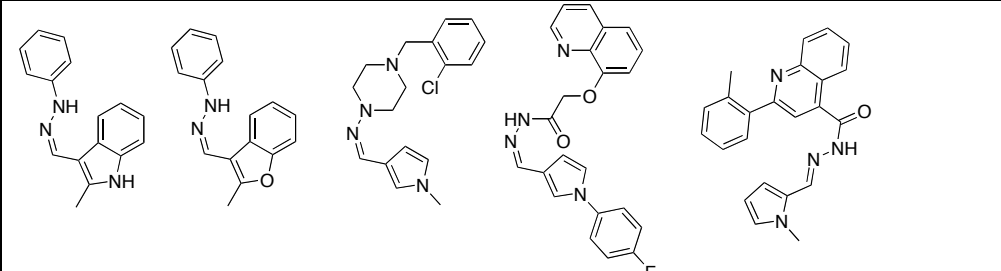
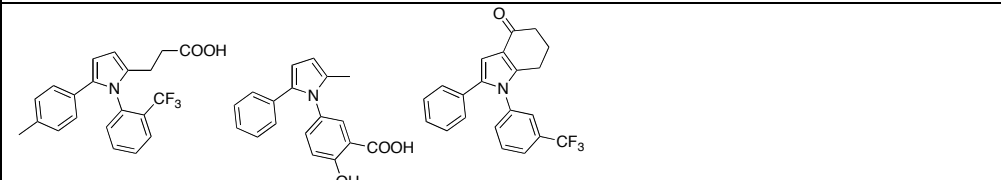
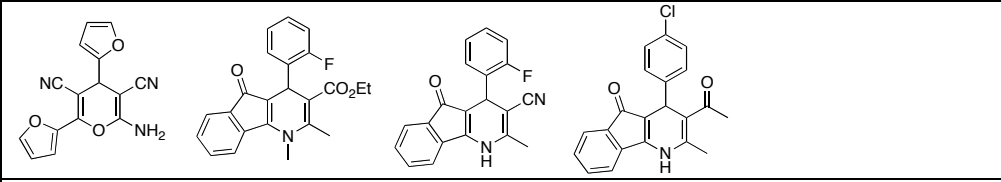
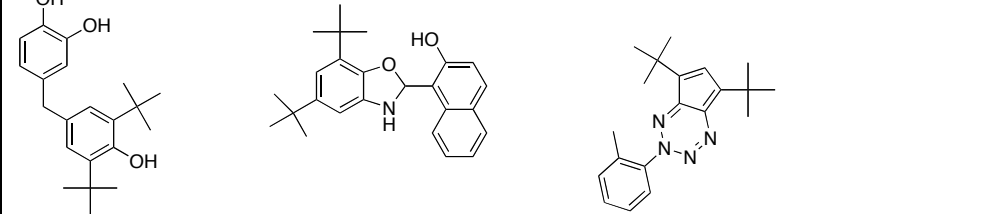

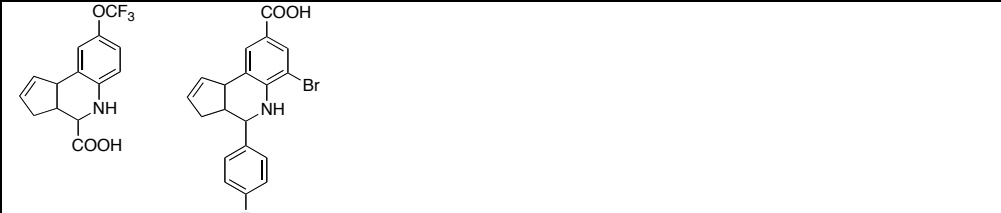
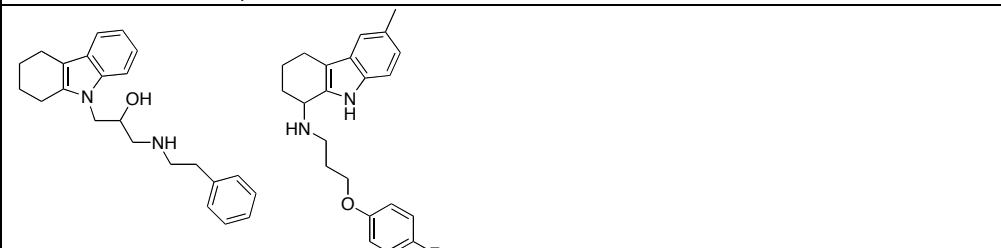
Phenylene-diamine, hydroquinone, catechol-like	21	
Certain aryl sulfonamides with donating group (either end)	16	
Alpha keto and isatin-like hydrazones and imines	13	
1,1-dinitrile / diketone / meldrum's acids	12	



Biarylazo and conjugated azo-imines	25	 <p>Chemical structures of biarylazo and conjugated azo-imines, including various azobenzene derivatives, azo-imines, and conjugated systems with aromatic and heterocyclic moieties.</p>
Thiohydrazide and electron rich phenyl-thioureas	14	 <p>Chemical structures of thiohydrazide and electron rich phenyl-thioureas, featuring various thiourea derivatives, thiohydrazides, and electron-rich phenyl groups.</p>
Unsaturated cyclic hydrazides and esters	12	 <p>Chemical structures of unsaturated cyclic hydrazides and esters, including various cyclic hydrazides, esters, and unsaturated systems with aromatic and heterocyclic moieties.</p>

Certain 2-amino-thiophenes	8	
Five-membered unsaturated hydrazides and lactone-like with exocyclic alkene	18	
Benzothiazine and miscellaneous sulfur-containing compounds	12	

Alkenyl barbiturates	8	
Electron deficient Benzoxa- diazole-like	5	
Cyclic thioureas, thiocarbamate s	10	
Electron rich indoles, benzoxa- thiolones, benzimidazole s	9	

Hydrazones of pyrrole aldehydes and similar	7	
1,2,5-carbopyrroles	7	
Certain dihydropyrans / pyridines	5	
Certain 1,3-di-t-Butyl aromatics	3	
Azulenes	2	
Fused cyclopentene-quinolines	2	
Certain 2,3-dialkyl indoles	4	

Miscellaneous	10	<p>The image displays 12 chemical structures:</p> <ul style="list-style-type: none"> <li>1. A benzimidazole derivative with a carboxylic acid group, an amino group, a cyano group, and an acetyl group.</li> <li>2. A fluorenone derivative with a hydroxymethyl group.</li> <li>3. A fluorenone derivative with a hydroxyl group.</li> <li>4. A naphthalene derivative with a chlorine atom, a carbonyl group, and a 3-hydroxypropylamino group.</li> <li>5. A pyrimidine derivative with two chlorine atoms, a cyano group, and an ethyl ester group.</li> <li>6. A benzimidazole derivative with a 2,4-diaminobenzoyl ester group.</li> <li>7. A piperazine derivative with a 3,5-dimethylphenyl group and a (E)-3-phenylprop-2-en-1-yl group.</li> <li>8. A pyrazolo[1,5-a]pyridine derivative with a 3-phenylprop-2-en-1-yl group.</li> <li>9. A quinazoline derivative with two chlorine atoms and a cyano group.</li> <li>10. A pyrazole derivative with a trifluoromethyl group, an ethyl ester group, and a cyano group.</li> <li>11. A 1,2,3,4,5-pentazapentalene derivative with a benzyl group.</li> <li>12. A pyrazole derivative with a bromine atom and a methyl group.</li> </ul>
---------------	----	--

**Supplementary Table S5.** Assay counts for benign versus problematic groups showing the greater number that hit 2-6 assays when expressed as a percentage of those that hit no assays (“Enrichment”)

Group	ASSAY COUNT							SLN	Enrichment
	6	5	4	3	2	1	0		2-6/0
BENIGN STRUCTURES									
amide	30	76	109	185	1142	3384	19625	CNHC(=O)C	7.9%
aminopyridine	0	3	4	5	18	34	167	C[1]:C(NH2):N:C:C:C:@1	18%
benzothiazole	5	8	10	14	66	163	746	C[1]:N:C:S:C(:@1):C:C:C:C:@1	14%
chlorophenyl	44	139	138	220	763	2030	12013	C[1]:C(Cl):C:C:C:C:@1	11%
Aryl N	33	64	76	159	1191	1757	9621	C[1]:C:N:C:C:C:@1	16%
nitrogenpyridine	3	11	13	28	127	278	1811	C[1]:C(N):N:C:C:C:@1	10%
									Av. 13%
SUSPECTED PROBLEMATIC STRUCTURES									
p_quinones	32	39	36	28	32	35	59	C[1]C(=O)C:=CC(=O)C:=@1	283%
rhodanine	16	41	21	26	32	39	60	C=C[1]SC(=S)NC(=O)@1	234%
rhodanines	17	50	30	41	85	109	305	C=C[1]SC(=Het)NC(=O)@1	41%
Imidazolin-2,5-dione	4	7	8	9	13	5	0	C=C[1]C(=O)NNC(=O)@1	infinity
123carbopyrrole	4	5	10	0	6	2	4	C[1]:C(CH2):N(C:C):C(C:C):CH:@1	625%
s	4	2	6	8	5	18	54	C[1]:C(NH2):S:C:C:@1	43%

TABLE S6

#  
#####  
## FILTER FAMILY A (Freq\_Hit\_5\_morethan150.hits ##  
#####  
#SYBYL/3DB HITLIST  
#  
#  
#  
#@CLASS REGLIST  
#@DATABASE NONE  
#@SOURCE in-house  
#  
#  
#the numbers refer to number of members in the 93K wehi library  
#  
#The first line in each case represents the historical name  
#  
#sixes\_A7.txt 483  
C[1](C(~Het~CHetC@1=Het)~Het)=C[!r]H<regId="ene\_six\_het\_A(483)">  
#BB\_OHzone\_2.txt 479  
C[1]:C:C:C(:C(:C:@1)OH)C=NN<regId="hzone\_phenol\_A(479)">  
#CC\_Me2\_p\_anils\_2.txt 478  
CH2N(CH2)C[1]:C:C(Any[IS=H,CH2,OCH2CH2]):C(N):CH:C:@1<regId="anil\_di\_alk\_A(478)">  
#indole\_3yl\_alk\_A.txt 461  
N[1](C(=C(C[4]:CH:C:C:C:@1:@4)C[TAC=4]H)Any[IS=CH2,C=:Het,CHN,CH(CH2)NHCH2,CH(CH2)CH2NHCH2])Any[IS=H,CH2]<regId="indol\_3yl\_alk(461)">  
#p\_quinones.txt 370  
Het=C[1]C=:CC(=Het)C=:C@1<regId="quinone\_A(370)">  
#HH\_azo.txt 324  
N[!r]=N<regId="azo\_A(324)">  
#iminone\_3.txt 321  
CC(=Het[!r])C(=Het[!r])Any[IS=C,S(=O)=O]<regId="imine\_one\_A(321)">  
#New\_Mannichs\_F.txt 296  
NC[TAC=4]C[1]:C(OH):C:C:C:@1<regId="mannich\_A(296)">  
#CC\_Me2\_N\_1.txt 251  
C[1]:C(:C:C:C(:C:@1)C=N)N(C[TAC=4])C[TAC=4]<regId="anil\_di\_alk\_B(251)">  
#CCpN\_OC\_1.txt 246  
C[1]:C(:C:C:C(:C:@1)N(C[TAC=4])Any[IS=H,C[TAC=4]])OC[TAC=4]<regId="anil\_di\_alk\_C(246)">  
#KK\_rhodaninesC\_2.txt 235  
N[1]C(=S)SC(=C)C(=O)@1<regId="ene\_rhod\_A(235)">  
#BB\_OHzone\_1.txt 215  
C[1](C:C:C:C(:C:@1)C=NN)OH<regId="hzone\_phenol\_B(215)">  
#fives\_K\_6NO.txt 201  
C[1](=C)C=NHetC(=O)@1<regId="ene\_five\_het\_A(201)">  
#CC\_Me2\_N\_5.txt 198  
C[1]:C(:C:C:C(:C:@1)C[TAC=4]Any[IS=OH,C=CH,NC[TAC=4]])N(C[TAC=4])C[TAC=4]<regId="anil\_di\_alk\_D(198)">  
#isatins\_1.txt 189  
O=C[2]C(=[!R]NN)C[5]:C(:C:C:C:C:@5)N@2<regId="imine\_one\_isatin(189)">  
#CC\_Me2\_C\_new.txt 186  
CHN(CH2)C[1]:CH:C(Any[IS=H,CH2]):C(CHAny[IS=H,CH]):CH:CH:@1<regId="anil\_di\_alk\_E(186)">  
#

# TABLE S7

#  
#####  
## FILTER FAMILY B (Freq\_Hit\_5\_lessthan150.hits ##  
#####  
#SYBYL/3DB HITLIST  
#  
#  
#  
#@CLASS REGLIST  
#@DATABASE NONE  
#@SOURCE in-house  
#With number of members remaining (in parentheses) in WEHI library after Freq\_Hit\_5\_morethan150.filter has already been applied  
#For historical reasons, the early name is include before the sln with the revised regid name.  
#

#158\_thiazene.txt 128  
C[1](Any[IS=CH2,C:C])=C(Any[IS=H,CH2,C=O])SC(N@1(Any[IS=H,CH,C:C]))=N[!r]<regId="thiaz\_ene\_A(128)">  
#pyrrole\_ArAlk.txt 118  
N[1](C[2]:Hev:C:C:C:@2)C(C[TAC=4])=CCH=C(C[TAC=4])@1<regId="pyrrole\_A(118)">  
#BBhydroquin\_2.txt 92  
C[1]:C:C:C(:C(:@1)OH)OH<regId="catechol\_A(92)">  
#fives\_B2ene.txt 90  
C[1](=C)C(N=CS@1)=O<regId="ene\_five\_het\_B(90)">  
#5quin.txt 89  
C[1]C(=Het[!r])C(=O)HetHev=@1<regId="imine\_one\_fives(89)">  
#furanone.txt 85  
C[1](C(C=CHet@1)=C)=Het<regId="ene\_five\_het\_C(85)">  
#misc9.txt 79  
CN[1]CH2CH2N(CH2CH2@1)N=CHC:Hev<regId="hzone\_pipzn(79)">  
#diketo\_1.txt 68  
C[1]:C:C:C:C(:@1)C(=O)C[TAC=4]C(=O)@1<regId="keto\_keto\_beta\_A(68)">  
#pyrrole\_zone1.txt 64  
N[1](C)CH=CHCH=C(C=NN)@1<regId="hzone\_pyrrol(64)">  
#enone\_A1mod1.txt 57  
C=[!R]C(Hev)-[R]C(=[!R]Het)-[R]C(=[!R]C)Hev<regId="ene\_one\_ene\_A(57)">  
#amin\_cyano\_1.txt 56  
C(C#N)(C#N)C(NH2)=CC#N<regId="cyano\_ene\_amine\_A(56)">  
#diketo\_ene\_1.txt 55  
C[1]:C:C:C:C(:@1)C(=O)C(=C)C(=O)@1<regId="ene\_five\_one\_A(55)">  
#DD\_sixNS\_4.txt 54  
C[1](C#N)C(=S)NHHev=HevHev=@1<regId="cyano\_pyridone\_A(54)">  
#EE\_tricycle.txt 51  
C[1]:C:C:C[4]:C(:C:@1)NCC[9]C@4C=CC@9<regId="anil\_alk\_ene(51)">  
#FF\_acridines.txt 46  
C[1]:C:C:C:C[5]:C:@1:C(:C[8]:C(:N:@5):C:C:C:C:@8)N<regId="amino\_acridine\_A(46)">  
#fives\_Nene.txt 46  
C[1](=C)C(=O)NNC(=O)@1<regId="ene\_five\_het\_D(46)">  
#KK\_EE\_amino\_5hets\_1.txt 45  
NH2C[1]:S:C(Hev):C(Hev):C(C=O):@1<regId="thiophene\_amino\_Aa(45)">  
#oxind\_ene\_1.txt 44  
NC=[!R]C[2]C(=O)C[5]:C(:C:C:C:C:@5)Het@2<regId="ene\_five\_het\_E(44)">  
#DD\_sulfonOH.txt 43  
C[1](:C(:C(:CH:C(:CH:@1)Hal)Hal)OH)S(=O)(=O)N<regId="sulfonamide\_A(43)">  
##S\_thiketone.txt 43  
CC(=S)C<regId="thio\_ketone(43)">  
#CCpsulfon\_OH.txt 41  
C[1]:C(:C:C:C(:C:@1)NHS(=O)=O)OH<regId="sulfonamide\_B(41)">  
#BB\_anils.txt 40  
C[1](NH2):CH:CH:C(Any[IS=O,N,CH2]):CH:CH:@1<regId="anil\_no\_alk(40)">  
#KK\_EE\_amino\_5hets\_3b.txt 40  
Any[IS=H,CH2,C:C]C[1]:S:C(NHC(=O)C):C(C(=O)O):C(Any[IS=C[2]:C:C:C:C:C:@2,C[2]:S:C:C:C:@2]):@1<regId="thiophene\_amino\_Ab(40)">  
#pyridiniums\_A.txt 39  
N[1:+1](Any[IS=CH3,O[TAC=1],CH2CH=CH2,CH2CH2OH,CH2C(=O)C,CH2C(=O)NHC:C,CH2CH3]):C:C:Hev:C[2]:C(:@1):CH:C(Any[IS=H,N]):C:C:@2<regId="het\_pyridiniums\_A(39)">  
#anthrone.txt 38  
C[1]:C(:C(:C:C:C:@1)C(C:C)=O)NHAny[NOT=C=O]<regId="anthranil\_one\_A(38)">  
#FF\_cyano\_3.txt 37  
NHN=C(C#N)C=Het[!r]<regId="cyano\_imine\_A(37)">  
#EE\_diazox\_2.txt 36  
N(C[1]:C:C:C:C:C:@1)S(=O)(=O)C[2]:C[3]:N:Any[IS=O,S]:N:C(:@3):C:C:C:@2<regId="diazox\_sulfon\_A(36)">  
#DiMeAnil\_zone.txt 35  
CH2N(CH2)C[1]:CH:CH:C(CH=NN(Any[IS=C(=O)CH2SC:N,C(=O)CH2Hev:Hev:N,C(=O)C:COH,C:N,CH2CHOH])):CH:CH:@1<regId="hzone\_anil\_di\_alk(35)">  
#KK\_rhodaninesC\_6.txt 33  
N[1]C(=S)SC[TAC=4]C(=O)@1<regId="rhod\_sat\_A(33)">  
#EE\_enamine\_2.txt 30  
NHN=CC(Any[IS=H,C])=C(C)-[!R]Any[IS=N,OH]<regId="hzone\_enamin(30)">  
#pyrrole\_ArAr.txt 29  
N[1](C[2]:Hev:C:C:C:@2)C(C[TAC=4])=CCH=C(C:C)@1<regId="pyrrole\_B(29)">  
#FF\_OH\_thiophene.txt 28  
S[1]C=CC(=C@1)OH<regId="thiophene\_hydroxy(28)">  
#FF\_cyano\_8.txt 27  
C[1](=C(C(=O)NC(=N@1)Het)C#N)C<regId="cyano\_pyridone\_B(27)">  
#barb\_imine\_1.txt 27  
C[1](C(=O)NC(=O)NC(=O)@1)=N<regId="imine\_one\_sixes(27)">



#dyes5A.txt 27  
CH2N(:-C:C)~C-:C=:CC~C:N<regId="dyes5A(27)">  
#FF\_Naphth\_4.txt 25  
C[1]:C:C:C[4]:C[5]:C(:C:C:C:C:@1:@5)N=CN@4<regId="naphth\_amino\_A(25)">  
#FF\_Naphth\_1.txt 25  
C[1]:C:C:C[4]:C[5]:C(:C:C:C:C:@1:@5)NHC[TAC=4]NH@4<regId="naphth\_amino\_B(25)">  
#enone\_2.txt 24  
CC(=O)CH=C(NHC)C(=O)OC<regId="ene\_one\_ester(24)">  
#misc3.txt 23  
S=C[1]C=:CHetC=:C@1<regId="thio\_dibenzo(23)">  
#cyanos\_1.txt 23  
C(C#N)(C#N)C(Any[IS=C#N,C=N])C#N<regId="cyano\_cyano\_A(23)">  
#naphthol\_A.txt 22  
C[1]:C(:CH:CH:CH:CH:@1):CH:C(:C(:CH:@1)OH)C(=O)NHN=C<regId="hzone\_acyl\_naphthol(22)">  
#misc11.txt 21  
O=CC[1]=C[2]N=C(CH2)C=C(OH)N(@2)N=C@1<regId="het\_65\_A(21)">  
#AAmisc2.txt 19  
N[1]:C(C[2]:C:C:C:C:@2):C(C[3]:C:C:C:C:@3):NH:C(C:Hev):@1<regId="imidazole\_A(19)">  
#ene\_cyano\_1.txt 19  
C(C#N)(C#N)=CC[1]:C:C:C:C:C:@1<regId="ene\_cyano\_A(19)">  
#GG\_acidzone.txt 19  
C[1](C(=O)OH):C(NHN=C):C:C:C:C:@1<regId="anthranil\_acid\_A(19)">  
#dyes3A.txt 19  
N[+1](:-C:C):=CCH=CN(C[TAC=4])C<regId="dyes3A(19)">  
#dhp\_bis\_amino\_CN.txt 19  
NH2C[1]=C(C#N)CH(C:C)C(C#N)=C(NH2)S@1<regId="dhp\_bis\_amino\_CN(19)">  
#BBparanils\_2\_mod.txt 18  
N~C[1]:Hev[IS=N,C(N)]:Hev[IS=N,CH,C(NH)]:C(Hev[IS=NH,OCH2]):N:N:@1<regId="het\_6\_tetrazine(18)">  
#EE\_Xenone.txt 17  
CC=C(Hal)C(=O)C<regId="ene\_one\_hal(17)">  
#imino\_cyano\_1.txt 17  
C(C#N)(C#N)=NNHC[1]:C:C:C:C:C:@1<regId="cyano\_imine\_B(17)">  
#158\_thiazene\_A5b.txt 17  
C[1](Any[IS=NC(=O)C:C,NH2])=C(-[!R]C(=O)NCH2)SC(N(@1)Any[IS=CH2CH=CH2,C:C])=S<regId="thiaz\_ene\_B(17)">  
#P\_rhodanines\_4E.txt 16  
S[1]C(=O)NC(=O)C(=CHAny[IS=CBr,C:CH:C(Hal):C:CHaL,C:CH:CH:CSCH2,C:C:C:C:C:C:C:COCH2,C[2]:C(CH2):N(CH2):C(CH2):C:@2])@1<regId="ene\_rhod\_B(16)">  
#DD\_65\_OOS\_1\_1.txt 15  
O[1]C(SC[4]:C:C(:C:C:C@1:@4)Any[IS=N,O])=Any[IS=O,S]<regId="thio\_carbonate\_A(15)">  
#KK\_EE\_amino\_5hets\_10a.txt 15  
N(CH2)(CH2)C[1]:O:C(C=NNHC(=Het)):CH:CH:@1<regId="anil\_di\_alk\_furan\_A(15)">  
#oxinde\_ene\_4.txt 15  
C[1](:C:C:C:C:C:@1)CH=[!R]C[2]C(=O)C[5]:C(:C:C:C:C:@5)S@2<regId="ene\_five\_het\_F(15)">  
#  
#

## # TABLE S8

#  
#####  
##                  FILTER FAMILY C (Freq\_Hit\_5\_lessthan15.hits                  ##  
#####  
#SYBYL/3DB HITLIST  
#  
#  
#  
#@CLASS REGLIST  
#@DATABASE NONE  
#@SOURCE in-house  
#  
#Numbers in parentheses refer to hits on database Lessthan15  
#For historical reasons, the early name is include before the sln with the revised regid name.  
#  
#Me2\_N\_5.txt 14  
C[1]:C(:C:C:C(N(Any[IS=H,C[TAC=4]])(Any[IS=H,C[TAC=4]])):C:@1)C[TAC=4]C[23]:C:C:C(:C:C:@23)N(Any[IS=H,C[TAC=4]])(Any[IS=H,C[TAC=4]])<regId="anil\_di\_alk\_F(14)">  
#CC\_anilzone.txt 14  
C[1](:CH:CH:C(NH2):CH:CH:@1)C=NNH<regId="hzone\_anil(14)">  
#pyraz\_OH.txt 14  
C[1](C[TAC=4])C(Any[IS=H,CH])=C(OH)N(C[2]:CH:CH:CH:CH:CH:@2)N=@1<regId="het\_5\_pyrazole\_OH(14)">  
#JJ\_DD\_phenothz.txt 13

C[1]H:CH:CH:CH:C[5]N(Any[IS=H,CH3,CH2CH])C[7]:C(~Any[IS=H,NHC[TAC=4],C:C]):C(~Any[IS=H,C:C]):C(Any[IS=H,NH2,OC[TAC=4]]):CH:C:@7SC:@1:@5<regId="het\_thio\_666\_A(13)">

#GG\_tricycle.txt 13

C[1]CC[3]:C:C:C:C:@3C(C[10]:C@1:C:C:C:C:@10)=CC<regId="styrene\_A(13)">

#P\_rhodaninesC\_4CB.txt 13

S[1]C(=NC:C)N(Any[IS=H,CH2CH2O,C:C])C(=O)C(=CHAny[IS=C:C:CCl,C:Het])@1<regId="ene\_rhod\_C(13)">

#dhp\_amino\_CNene.txt 13

NH2C[1]=C(C#N)CH(C:C)C(CH2)=C(C=C)O@1<regId="dhp\_amino\_CN\_A(13)">

#FF\_cyano\_12.txt 12

O=S(=O)C(C#N)=NNH<regId="cyano\_imine\_C(12)">

#KK\_thiourea\_3.txt 12

C[1]:C(NHC(=S)NHCH2CH2CH2N(CH2)C[2]:C:C:C:C:C:@2):C:C:C:C:@1<regId="thio\_urea\_A(12)">

#KK\_EE\_amino\_5hets\_2.txt 12

C[2]:C:C:C:C:C:(@2)NHC[1]:S:C(Any[IS=C=O,C#N,C(OH)=C]):C(N):C(Any[IS=C#N,C(:N):N]):@1<regId="thiophene\_amino\_B(12)">

#fives\_Ntac4.txt 12

C[1:TAC=4]C(=O)NNC(=O)@1<regId="keto\_keto\_beta\_B(12)">

#DD\_suber.txt 11

C[1]:C:C:C:C[5]:C:@1C(C[8]:C@5:N:Hev:C:C:@8)=O<regId="keto\_phenone\_A(11)">

#sixes\_A6.txt 11

C[1](C(CH2)=C(C#N)C(~O)~N~C(~O)@1)=CHC:C<regId="cyano\_pyridone\_C(11)">

#158\_thiazene\_A2.txt 11

C[1](Any[IS=Hal,N[+1](~C):C])=C(-[!R]C=N)SC(N@1)=O<regId="thiaz\_ene\_C(11)">

#KK\_DD\_thiophenezone\_1A.txt 11

C[1]:CH:CH:CH:CH:C(:@1):Het:C(Any[IS=H,OH,CH2]):C(:@1)CH=NNHAny[IS=C[2]:N:C:CH:S:@2,C:CH:CH,C:N:C:N:C:N,C:N:N:N:N]<regId="hzone\_thiophene\_A(11)">

#FF\_cyano\_18.txt 10

Hev:HevC(Any[IS=H,C#N])=C[1]C:=CC(=Any[IS=O,N[!r]])C:=C@1<regId="ene\_quin\_methide(10)">

#DD\_phenothz\_3.txt 10

C[1]:C:C:C[4]:C(:C:@1)SC[8]:CH:CH:C(:CH:C:@8C(C@4)N(Any[IS=H,C[TAC=4]])(Any[IS=H,C[TAC=4]]))Any[IS=H,O,S[TAC=2],C[TAC=4],N(Any[IS=H,C[TAC=4]]))Any[IS=H,C[TAC=4]]<regId="het\_thio\_676\_A(10)">

#KK\_110\_fives\_6A 10

C[1](=O)C(=CHAny[IS=C[2]:C:C:C:C:C:@2,C[2]:C:C:C:Het:@2])N=C(Hev:Hev:Hev)Any[IS=S,NHev:Hev]@1<regId="ene\_five\_het\_G(10)">

#JJ\_ureimiums.txt 9

N[+1](~Hev:Hev:Hev)Hev=O<regId="acyl\_het\_A(9)">

#KK\_CC\_ME2\_p\_im\_1D.txt 9

C[TAC=4]N(C[TAC=4])C[1]:CH:CH:C(C[2]:=N:-C:C:Hev:-@2):CH:CH:@1<regId="anil\_di\_alk\_G(9)">

#JJ\_FF\_dhp\_1.txt 9

N[1](Any[IS=C[TAC=4],H])C[2]:=C(C(=O)C:C:C)C(C)C(Any[IS=C=O,C#N])=C(CH3)@1<regId="dhp\_keto\_A(9)">

#KK\_thiourea\_4.txt 9

C[1]:C(NHC(=S)NHCH2CH2N(CH2)C[2]:C:C:C:C:C:@2):C:C:C:C:@1<regId="thio\_urea\_B(9)">

#JJ\_bim.txt 9

C[1]:CH:CH:C(NHCH2C[2]:C:C:C:C:C:@2):CH:C(:@1):N:CH:N(C):@1<regId="anil\_alk\_bim(9)">

#GG\_indolene.txt 9

C[1]:C:C:C[4]:C(:C:@1)C(C=N@4)=N[!r]<regId="imine\_imine\_A(9)">

#thiocarbz\_2.txt 9

C[1](~C:C:C:C:C:@1)NHC(=S)NNHC(=O)C[2]:Hev:Het:C:C@2<regId="thio\_urea\_C(9)">

#isatins\_2.txt 9

N[!r]=C[2]C(=O)C[5]:C(:C:C:C:C:@5)S@2<regId="imine\_one\_fives\_B(9)">

#dhp\_amino\_CNpyrz.txt 9

Any[IS=NH2,OH]C[1]=C(C#N)CH(C:C)C[2]:C(:N(C):N:C:@2)O@1<regId="dhp\_amino\_CN\_B(9)">

#CC\_ON\_3.txt 8

NH2C[4]:CH:N:C(OC[10]:C:C:C:C:C:@10):CH:CH:@4<regId="anil\_OC\_no\_alk\_A(8)">

#DD\_phenothz\_2.txt 8

C(=O)C[3]SC[5]:C:C:C:C:C:@5NC=@3<regId="het\_thio\_66\_one(8)">

#DD\_phenothz\_4.txt 8

C[1]:C:C:C[4]:C(:C:@1)Any[IS=S[TAC=2],C[TAC=4]]C[8]:C:C:C(:C:C:@8C@4=CC)Any[IS=H,Cl,C[TAC=4]]<regId="styrene\_B(8)">

#KK\_107\_SIM.txt 8

CH2S[TAC=2]C[1]:N(-[!R]CH2):C(C[2]:C:C:C:C:C:@2):CH:N:@1<regId="het\_thio\_5\_A(8)">

#NN\_CC\_Me2\_N\_2.txt 8

CH2N(CH2)C[1]=CHC[2]:C:C:C:C:C:@2)S[TAC=2]C[3]:C:C:C:C:C(:@3)@1<regId="anil\_di\_alk\_ene\_A(8)">

#P\_rhodaninesC\_4CA.txt 8

S[1]C(=[!R]NAny[IS=H,NHC:C])N(Any[IS=H,C:N:C:C:S])C(=O)C(=CHC:C(Any[IS=Cl,OCH]))@1<regId="ene\_rhod\_D(8)">

#P\_rhodaninesC\_4D.txt 8

S[1]C(=O)NC(=S)C(=CHC:C)@1<regId="ene\_rhod\_E(8)">

#p\_BnNH\_phenol.txt 8

C[1]:C:C:C:C:C(:@1)CH2NHC[2]:CH:CH:C(OH):CH:CH:@2<regId="anil\_OH\_alk\_A(8)">

#pyrrole\_AlkAralk.txt 8

N[1](C[TAC=4])C(CH2)=CHCH=C(C:C)@1<regId="pyrrole\_C(8)">

#thiocarbz\_1.txt 8

C[1](~C:C:C:C:C:@1)NHC(=S)NNHC[2]:C:C:C:C:C:@2<regId="thio\_urea\_D(8)">

#AAthiazole.txt 8

N[C(2]:C:C:C:C:C:@2)C[8]SC=C(N[charge=+1]=@8)C[13]:C:C:C:C:C:@13<regId="thiaz\_ene\_D(8)">  
#pyrrole\_ene2b.txt 8  
N[1]:C:C:C(CH=C[2]C(=O)NC(=Het)N@2):C(CH2):@1<regId="ene\_rhod\_F(8)">  
#158\_thiazene\_C1.txt 8  
C[1](SC[r])=C(CH(C)C)SC(N@1(Any[IS=H,CH2]))=O<regId="thiaz\_ene\_E(8)">  
#fives\_O.txt 7  
Hev[1]:Hev[3]:Hev(:Hev:Hev:Hev:@1)NHN(C@3=O)C<regId="het\_65\_B(7)">  
#FF\_misc\_3.txt 7  
C[1]:C:C:C[7]:C(:C:@1)C(C(=C@7OH)C)=O<regId="keto\_keto\_beta\_C(7)">  
#FF\_misc11.txt 7  
C[1]:C:C:C[7]:C(:C:@1):N:N:C(:N:@7)CH2C=O<regId="het\_66\_A(7)">  
#KK\_thiourea\_6.txt 7  
C[1]:C(NHC(=S)NHCH2C[2]:N:C:C:C:@2):C:C:C:C:@1<regId="thio\_urea\_E(7)">  
#KK\_EE\_amino\_5hets\_6a.txt 7  
CHCH2C[1]:S:C(NHC(=O)CCC=O):C(Any[IS=C(=O)O,C#N]):C(CH2):@1<regId="thiophene\_amino\_C(7)">  
#AA\_misc11.txt 7  
C(C[1]:CH:C:C(Any[IS=C[TAC=4],H]):CH:CH:@1)(C[2]:CH:CH:C(Any[IS=H,Cl]):CH:CH:@2)=Any[IS=NOCH2CH2CH2N(CH2)CH2,NOCH2CH2N(CH2)CH2,NNHC(=NH)NH2,CHN]<regId="hzone\_phenone(7)">  
#acid\_ar\_rhod.txt 7  
HOC(=O)C[1]:C:C:C:C(:C:@1)C:Hev:CCCH=C[2]C(=Het)NC(=Het)Het@2<regId="ene\_rhod\_G(7)">  
#FF\_cyano\_13.txt 7  
C[1](C[2]:C(:C:C:C(:N:@2)NH2)C#N)C(C=@1)=C)C#N<regId="ene\_cyano\_B(7)">  
#dhp\_amino\_CNCN.txt 7  
NH2C[1]=C(C#N)CH(C:C)C(C#N)=C(C:C)O@1<regId="dhp\_amino\_CN\_C(7)">  
#166\_het\_5\_q.txt 7  
N[1](C[2]:C:C:C:C:@2)N=C(C(=O)C[TAC=4]C(=O))@1<regId="het\_5\_A(7)">  
#fives\_K\_11.txt 6  
N[1]=CC(C(N@1)=S)=C<regId="ene\_five\_het\_H(6)">  
#EE\_5hets10.txt 6  
C[1]H=CO C(=CH@1)C(=S)N[10]CH2CH2HevCH2CH2@10<regId="thio\_amide\_A(6)">  
#FF\_cyano\_2.txt 6  
C=C(C#N)C(=NH)NN<regId="ene\_cyano\_C(6)">  
#KK\_111\_furzone\_1.txt 6  
C[1](O:C(Any[IS=H,CH2]):CH:CH:@1)C(Any[IS=H,CH2])=NNHC[2]:N:C:C:S:@2<regId="hzone\_furan\_A(6)">  
#KK\_sulfonam\_1.txt 6  
C[1]H:C(C(NHS(=O)(=O)C[2]:C:Hev:C:C:C:@2):CH:C(N(CH2)CH2):CH:CH:@1<regId="anil\_di\_alk\_H(6)">  
#FF\_Ncycle.txt 6  
N[1]C(=C(NC:C)N[6]C=CC=CC=@1@6)Any[IS=C(Hev)=CHC:C,C:O:C]<regId="het\_65\_C(6)">  
#thiocarbz\_5.txt 6  
C[1]NHNHC(=S)NNH@1<regId="thio\_urea\_F(6)">  
#oxind\_ene\_2.txt 6  
C[1](C:C:C:O:@1)CH=[!R]C[2]C(=O)C[5]:C(:C:C:C:C:@5)Het@2<regId="ene\_five\_het\_I(6)">  
#diketo\_B.txt 5  
O=C[2]C[TAC=4]CC(=O)C[7]:C@2:C:C:C:C:@7<regId="keto\_keto\_gamma(5)">  
#5quin2.txt 5  
C[1]:C:C:C[4]:C(:C:@1)C[7]=C[8]C(C@4=O)=CC=CC@8=NO@7<regId="quinone\_B(5)">  
#DD\_OHPyr\_1.txt 5  
HOC[2]:N:C(OH):C:C:C:@2<regId="het\_6\_pyridone\_OH(5)">  
#naphth\_D.txt 5  
C[1]:C(:CH:CH:CH:CH:@1):CH:C(:CH(:CH:@1))C=NNHAny[IS=C:C,C=S]<regId="hzone\_naphth\_A(5)">  
#DD\_thioest\_5.txt 5  
C[1]=C(SC(C=C@1)=S)N<regId="thio\_ester\_A(5)">  
#EE\_ketal.txt 5  
C[1]C(OC(C=@1)(O)C)O<regId="ene\_misc\_A(5)">  
#FF\_cyano\_17.txt 5  
O=C[2]C(=CC(=NN@2)C=O)C#N<regId="cyano\_pyridone\_D(5)">  
#FF\_misc\_7.txt 5  
C[1]C=CC[4]N(C=@1)C(NC)=C(N=@4)C[12]:C:C:C:C:N:@12<regId="het\_65\_Db(5)">  
#FF\_Ncycle\_2.txt 5  
N[1]C[2]:C(C(=N)C[6]:C:C:C:C:C@1:@6):C:C:C:C:@2<regId="het\_666\_A(5)">  
#KK\_108\_diazox\_3a.txt 5  
C[3]:C:C:C:C:C(:@3)N[1]CHCHN(CHCH@1)S(=O)(=O)C[4]:C[2]:N:S:N:C(:@2):C:C:C:@4<regId="diazox\_sulfon\_B(5)">  
#KK\_113\_dianiline\_2.txt 5  
C[1](NHCH2):CH:CH:C[2]:C(NHC(=O)NH@2):CH:@1<regId="anil\_NH\_alk\_A(5)">  
#KK\_sulfonam\_2.txt 5  
C[1]H:C(NHS(=O)(=O)C[2]:C:C:C(Het):C:C:@2):CH:C[3]:C(OCH2O@3):CH:@1<regId="sulfonamide\_C(5)">  
#KK\_107\_SIM\_4a.txt 5  
CHC[1]:N:N(C[3]:C:C:C:C:C:@3):S[2]:Het:Hev:C:C(:@2):@1<regId="het\_thio\_N\_55(5)">  
#DD\_enone\_1.txt 5

O=CC=CHOH<regId="keto\_keto\_beta\_D(5)">  
#P\_rhodaninesC\_4B.txt 5  
N[1]C(=NC(=O)C(CH2)=N@1)SC(=CHC:C)C(=O)@1<regId="ene\_rhod\_H(5)">  
#chalcimine.txt 5  
C:CCH=CHCH=NN(C[TAC=4])C[TAC=4]<regId="imine\_ene\_A(5)">  
#carbаз\_NSally1.txt 5  
C[1]:C:C:C:C:C(:@1):C[2]:N:N:C(SCH2C=O):N:C(:@2):N(CH2CH=CH2):@1<regId="het\_thio\_656a(5)">  
#pyrrole\_thiourea.txt 5  
N[1](C)CH=CHCH=C(CH2NHC(=S)NH)@1<regId="pyrrole\_D(5)">  
#pyrrole\_HetAr.txt 5  
N[1](C[2]:CH:Hev:Het:Hev:@2)C(C[TAC=4])=CHCH=C(C[TAC=4])@1<regId="pyrrole\_E(5)">  
#thiocarbз\_4.txt 5  
C[1](C:C:C:C:C:C:@1)NHC(=S)NNHC(=:N[r]):N[r]<regId="thio\_urea\_G(5)">  
#KK\_113\_dianisole\_2.txt 5  
C[1](OCH2):C(Any[IS=H,CH2]):CH:C(OCH2):C(CH2Any[IS=NHC(=O)CH2CH2CH2,CH(CH2)NHC(=S)NH]):CH:@1<regId="anisол\_A(5)">  
#pyrrole\_ArCN\_1.txt 5  
N[1](C[2]:C(C#N):C:C:Het:@2)CH=CHCH=C@1<regId="pyrrole\_F(5)">  
#dhp\_amino\_CNS.txt 5  
NH2C[1]=C(C#N)CH(C:C)C[2]:C(:C:C:S:@2)O@1<regId="dhp\_amino\_CN\_D(5)">  
#GG\_thiophene\_3.txt 4  
NHC[1]:N:C(C[2]:C:N:C(NH2):S:@2):C:S:@1<regId="thiazole\_amine\_A(4)">  
#BBparanils\_2\_1.txt 4  
N=C[1]NHC(NH)=C(NH)N=N@1<regId="het\_6\_imidate\_A(4)">  
#CCpCN\_O2.txt 4  
C[1]:C:C:C[4]:C(:C:@1OC[8]:CH:CH:C(:CH:CH:@8)NH):C:C:C:C:@4<regId="anil\_OC\_no\_alk\_B(4)">  
#DD\_phenothз\_5.txt 4  
C[1]:C:C:C[4]:C(:C:@1)C(C[8]=C(SC@4)SC=C@8)=C<regId="styrene\_C(4)">  
#DD\_apine\_1.txt 4  
C[1]:C:C:C[4]:C(:C:C:@1):C:C:C:@4<regId="azulene(4)">  
#EE\_5hets29.txt 4  
C[1](O:C(C[CH2]:C(CH2OC:C):CH:@1)C(=O)OH<regId="furan\_acid\_A(4)">  
#FF\_cyano\_19.txt 4  
Hev:CC[1]=CHC(C:O)=C(C#N)C(=O)NH@1<regId="cyano\_pyridone\_E(4)">  
#FF\_SS.txt 4  
C[1]=C(C(NC[5]:C@1:C:C:C:C:@5)(C)C)SSC@1=Hev<regId="anil\_alk\_thio(4)">  
#KK\_benzanilid\_1.txt 4  
C[1]H:C(NHC(=O)C[2]:C:C:C:C:C:@2):CH:C(N(CH2)CH2):CH:CH:@1<regId="anil\_di\_alk\_I(4)">  
#NN\_sixesB8\_2.txt 4  
CH2S[TAC=2]C[1]:N:C(C[2]:O:CH:CH:CH:@2):C(C[3]:O:CH:CH:CH:@3):N:N:@1<regId="het\_thio\_6\_furan(4)">  
#NN\_CC\_Me2\_N\_1.txt 4  
CH2N(CH2)C[1]=CC[2]:C:C:C:C:C(:@2)CH2@1<regId="anil\_di\_alk\_ene\_B(4)">  
#NN\_DD\_NNCON\_1\_1a.txt 4  
NH(C[1]:C:C:C:C:C:@1)N=C(C(=O)CH2)NHAny[IS=NH,C:C]<regId="imine\_one\_B(4)">  
#P\_fur\_anis.txt 4  
C[1]:CH:CH:CH:CH:C(:@1):O:C[2]:CH:C(NHCH2):C(OCH2):CH:C(:@2):@1<regId="anil\_OC\_alk\_A(4)">  
#KK\_116\_thipyr\_new.txt 4  
S=C[1]NHC=CC[2]=C(@1)C(=O)OC(@2)=CH<regId="ene\_five\_het\_J(4)">  
#pyrrole\_ArH.txt 4  
N[1](C[2]:C:CH:C(Any[IS=NH2,C:N]):C:CH:@2)CH=CHCH=CH@1<regId="pyrrole\_G(4)">  
#pyrrole\_ene.txt 4  
N[1](C)CH=CHCH=C(CH=C[2]C(=O)HetC=:Hev@2)@1<regId="ene\_five\_het\_K(4)">  
#amino\_cyano\_2a.txt 4  
C=CC(C#N)(C#N)C(C#N)=CNH2<regId="cyano\_ene\_amine\_B(4)">  
##Ar\_thiono\_S.txt 4  
C:CC(=S[f])S[TAC=2]CHAny[IS=CH2,C:C]<regId="thio\_ester\_B(4)">  
#isatin\_ene.txt 4  
O=C[1]C(=[!R]CHC[3]:C:C:N:C:@3)C[2]:C(:C:C:C:C:@2)N@1<regId="ene\_five\_het\_L(4)">  
#KK\_111\_thiophzone\_2a.txt 4  
C[1](S:CH:CH:C(Any[IS=H,CH2])):@1)CH=NNHC[2]:C:C:C:C:C:@2<regId="hzone\_thiophene\_B(4)">  
#dhp\_amino\_CNone.txt 4  
CH2S[f]C[1]=C(C#N)CH(C:C)C(C#N)C(=O)N@1<regId="dhp\_amino\_CN\_E(4)">  
#166\_het\_5\_u.txt 4  
N[1](C[2]:C:C:C:C:C:@2)N=C(NHC=O)CH2C(=O)@1<regId="het\_5\_B(4)">  
#JJ\_FF\_misc\_10\_B.txt 3  
C:CCH=CHCH=NN=C<regId="imine\_imine\_B(3)">  
#fives\_V.txt 3  
C[1](C:C:C(C[CH2]:C:C:@1)C[2]:N:C(NH2):S:C(CH3):@2<regId="thiazole\_amine\_B(3)">  
#sixesB3.txt 3

C[1](C=NC[4]:C(N@1):C:C:C:C:@4)=CHC=O<regId="imine\_ene\_one\_A(3)">  
#EE\_diazox\_1a.txt 3  
O(C[2]:C:C:C:C:C:@2)C[3]:C:C[1]:N:O:N:C(:@1):C:C:@3<regId="diazox\_A(3)">  
#chalc1.txt 3  
Hev[1]:Hev:Hev:Hev(:Hev:Hev:@1)CH=CHC(NC[13]:C:C[15]:C(:C:C:@13):N:C(:C:C:@15)N(C)C)=O<regId="ene\_one\_A(3)">  
#BB\_o\_anils\_3.txt 3  
NH2C[1]:N:C:C:C:C(OCH2C:C):@1<regId="anil\_OC\_no\_alk\_C(3)">  
#DD\_S\_thiaz2.txt 3  
CS[TAC=2]C[1]:N:CH:C:S:@1<regId="thiazol\_SC\_A(3)">  
#DD\_phenox.txt 3  
C[1]:C:C:C:C[9]:C:@1OC[12]:C(N@9CH2CH2):C:C:C:C:@12<regId="het\_666\_B(3)">  
#EE\_5hets21.txt 3  
C[1](O:C(CH2):CH:CH:@1)CH(OH)C#CC[TAC=4]<regId="furan\_A(3)">  
#FF\_azep.txt 3  
C[1](C(=CC=CC=C@1)NH)=NC<regId="colchicine\_A(3)">  
#JJ\_EE\_enamine\_1.txt 3  
CH2N(CH2)CH=CC(=O)C[1]:C(S[TAC=2]):S:C(Any[IS=C#N,C=O]):C:@1<regId="thiophene\_C(3)">  
#JJ\_enone\_Da.txt 3  
C[2]:C:C:C:C:C(:@2)N[1]C(=O)C(NHC[2]:C:C:C(OCH2):C:C:@2)=C(Hal)C(=O)@1<regId="anil\_OC\_alk\_B(3)">  
#KK\_107\_SIM\_3a.txt 3  
C[1]:C:C:C:C:C(:@1)CH2CH2N=C(@1)S[TAC=2]CH2C(=O)C[2]:C:C:C:C:C:@2<regId="het\_thio\_66\_A(3)">  
#KK\_110\_rhod\_6.txt 3  
N[1](C[2]:C(CH2):C:C:C:C:@2)C(=S)N(CH2Hev:Hev:Hev:Hev:Hev)CH2C(=O)@1<regId="rhod\_sat\_B(3)">  
#KK\_110\_rhod\_9.txt 3  
N[1](CH2)C(=S)NHC(=CHC[2]:C:C(Br):C:C:C:@2)C(=O)@1<regId="ene\_rhod\_I(3)">  
#KK\_111\_benzthioph\_1.txt 3  
C[1](S:C[2]:C:C:C:C:C:@2:C(CH2):@1)C(=O)CH2CH2<regId="keto\_thiophene(3)">  
#KK\_113\_polyimine\_3.txt 3  
N(CH2)(CH2)CH=NC(CH2)=NN(CH2)C:C<regId="imine\_imine\_C(3)">  
#KK\_114\_hetpyrdones\_1.txt 3  
C[1](CH2CH2):C(CH2):C[2]C(=Het[TAC=1])N(CHAny[IS=C(=O)O,C:C])C(Any[IS=H,SCH2])=NC(:@2):Het[TAC=2]:@1<regId="het\_65\_pyridone\_A(3)">  
#KK\_115\_Furthiaz\_1a.txt 3  
C[1](NHCH2C[2]:CH:CH:CH:O:@2):N:C(Hev:Hev:Hev(Any[IS=OCH2,CH2]):Hev:Hev):CH:S:@1<regId="thiazole\_amine\_C(3)">  
#KK\_116\_thipyr\_1a.txt 3  
N[1]:C(SCH):C(C#N):C(C[2]:C:C:C(OCH2):C:C:@2):CH:C(C:C):@1<regId="het\_thio\_pyr\_A(3)">  
#KK\_108\_melamime\_3.txt 3  
C[1]:C(NHCH2C[2]:O:CH:CH:CH:@2):N:C(NHC[3]:C(Any[IS=H,CH2,S[TAC=2],OCH,N[TAC=3]]):C:C(Any[IS=H,CH2,S[TAC=2],OCH,N[TAC=3]]):C(Any[IS=H,CH2,S[TAC=2],OCH,N[TAC=3]]):C:@3):N:C(:@1):C:C:C:C:@1<regId="melamine\_A(3)">  
#KK\_113\_dianiline.txt 3  
NH(C[1]:C:C:Hev:C:C:@1)C[2]:C:C:C(NHCH):C:C:@2<regId="anil\_NH\_alk\_B(3)">  
#KK\_110\_rhod\_3.txt 3  
N[1](C[2]:C:C:C:C:C:@2)C(=NC=O)SCH2C(=O)@1<regId="rhod\_sat\_C(3)">  
#KK\_EE\_amino\_5hets\_8a.txt 3  
C=CC(=O)NC[1]:S:C(C(=O)O):C(CH):C(C#N):@1<regId="thiophene\_amino\_D(3)">  
#KK\_CC\_o\_anisidine\_5C.txt 3  
Any[IS=H,CH2]OC[1]:CH:CH:CH:CH:C(:@1)NHCH2C[2]:N:C:C:N:@2<regId="anil\_OC\_alk\_C(3)">  
#NN\_sixesB8\_1.txt 3  
CH2S[TAC=2]C[1]=NC[2]C(CH2)=NN(C[3]:C:C:C:C:C:@3)C(=@2)N=N@1<regId="het\_thio\_65\_A(3)">  
#NN\_sixesB8\_3.txt 3  
CC(=O)CH2S[TAC=2]C[1]:N:C[2]:NH:C[3]:CH:CH:CH:CH:C(:@3):C(:@2):N:N:@1<regId="het\_thio\_656b(3)">  
#NN\_JJ\_thiziniums\_1.txt 3  
S[1]:C(NHC[2]:C(-:Any[IS=CH2,C:C]):C:C:C:C:@2):N[+1](CH2):C(C):CH:@1<regId="thiazole\_amine\_D(3)">  
#M1\_DD\_NNNS2a.txt 3  
C[1](=S)N(CH2C[2]:O:C:C:C:@2)C(C:C)=NNH@1<regId="thio\_urea\_H(3)">  
#KK\_116\_pyr\_2.txt 3  
N[1](C[2]:C:C:C:C:C:@2)C(=O)C(C#N)=CC(C#N)=N@1<regId="cyano\_pyridone\_F(3)">  
#KK\_110\_rhod\_8.txt 3  
N[1](C[2]:C:C:C:C:C:@2)C(=O)SCH(CH2C(=O)NHC:C)C(=O)@1<regId="rhod\_sat\_D(3)">  
#P\_rhodaninesC\_4A.txt 3  
CH2N[1]C(=Any[IS=S,N])HetC(=C[2]CH=CHC:CN(CH2)@2)C(=O)@1<regId="ene\_rhod\_J(3)">  
#iminophenol.txt 3  
C=N[!r]C[3]:C(OH):C:C:C:C:@3<regId="imine\_phenol\_A(3)">  
#JJ\_EE\_CNene\_1.txt 3  
O=C[1]SC[2]:C:C:C:C(OCH2):C(:@2)O@1<regId="thio\_carbonate\_B(3)">  
#QQ\_EE\_NS\_1.txt 3  
N=C[1]N=CNS@1<regId="het\_thio\_N\_5A(3)">  
#QQ\_EE\_NS\_2.txt 3  
N[1]SC[2]CNC:CC(=@2)C(=S)@1<regId="het\_thio\_N\_65A(3)">  
#QQ\_CC\_ME2\_pim\_1E.txt 3

CH2N(CH2)C[1]:CH:CH:C(CH=NN=C(C)C):CH:CH:@ 1<regId="anil\_di\_alk\_J(3)">  
#pyrrole\_iminium.txt 3  
N[1]C=CC=C(C=N(C)CC@ 1)@ 1<regId="pyrrole\_H(3)">  
#ene\_cyano.txt 3  
C(C#N)(C#N)=C(S)S<regId="ene\_cyano\_D(3)">  
#cyano\_cprop\_1a.txt 3  
C[1](C#N)(C#N)CH(C(=O)C)CH@ 1<regId="cyano\_cyano\_B(3)">  
#fives\_I.txt 3  
C[1]Any[IS=O,S]C(C(C=:@ 1)=O)=CC=O<regId="ene\_five\_het\_M(3)">  
#KK\_116\_conjugcyano.txt 3  
C:CC(=O)NHC(=O)C(C#N)=CHNHC:C<regId="cyano\_ene\_amine\_C(3)">  
#thiocarbz\_6.txt 3  
C[1](C:C:C:C:C:@ 1)NHC(=S)NHN=CC[2]:C:C:N:C:@ 2<regId="thio\_urea\_I(3)">  
#dhp\_amino\_CNCO.txt 3  
NH2C[1]=C(C#N)CH(C[2]:C:C:C:S:@ 2)C(C(=O)OC)=C(CH2)O@ 1<regId="dhp\_amino\_CN\_F(3)">  
#KK\_phthalimide\_3.txt 3  
C[3]:C:C(C(=O)NHC[4]:C:C:C:C(C(=O)OH):@ 4):C:C[1]:C(:@ 3)C(=O)N(CH2)C(=O)@ 1<regId="anthranil\_acid\_B(3)">  
#EE\_diazox\_3.txt 3  
ClC[2]:C:C[1]:N:O:N:C(:@ 1):C:C:@ 2<regId="diazox\_B(3)">  
#misc5.txt 3  
CC(=S)H<regId="thio\_aldehyd\_A(3)">  
#KK\_FF\_misc\_10\_1.txt 2  
C[TAC=4]NHC(C:C)=CHC(=S)NHC[1]:C:C:C:C:C:@ 1<regId="thio\_amide\_B(2)">  
#fives\_U.txt 2  
CH2CH2SCH2C[1]N=CNHC=@ 1<regId="imidazole\_B(2)">  
#fives\_W.txt 2  
O=CNHC[1]:C(C:C):N:C(CH2C#N):S:@ 1<regId="thiazole\_amine\_E(2)">  
#fivesY.txt 2  
CHNHC[1]:N:C(C[2]=CN=C[3]N(C=CS@ 3)@ 2):C:S:@ 1<regId="thiazole\_amine\_F(2)">  
#sixesB10.txt 2  
N[1]C(=O)CH=C(C)SC(=S)@ 1<regId="thio\_ester\_C(2)">  
#enone\_E.txt 2  
C(S)(N)=CHC=CHC=O<regId="ene\_one\_B(2)">  
#quinones2.txt 2  
O=C[1]C[2]:C:C:C:C:C(:@ 2)C[3]=C(OH)C(=O)NC[4]:C:C:C:C(:@ 1):C(:@ 3):@ 4<regId="quinone\_C(2)">  
#naphthol\_C.txt 2  
C[1]:C(:C(:C:C:C:@ 1)C(C)=O):C(:C:C:C:@ 1)Any[IS=OH,NH2]<regId="keto\_naphthol\_A(2)">  
#AAmisc5.txt 2  
CH(C[1]:C:C:C:C:C:@ 1)(C[2]:C:C:C:C:C:@ 2)C(=S)NH<regId="thio\_amide\_C(2)">  
#AA\_misc13.txt 2  
N[1](C(=O)C[2]:CH:C(C(=O)OH):CH:CH:C(:@ 2)C(=O)@ 1)(C[3]:C:CH:C(O):C:CH:@ 3)<regId="phthalimide\_misc(2)">  
#BBparanils\_5.txt 2  
C[1]:C(:C:C:C:C(:@ 1)NHS(=O)(=O))NHS(=O)=O<regId="sulfonamide\_D(2)">  
#BBorthoanils\_2.txt 2  
CHNHC[1]:CH:CH:CH:CH:C(NHCH):@ 1<regId="anil\_NH\_alk\_C(2)">  
#BB\_thiophenelactam.txt 2  
S[1]C(=C(C[4]=C@ 1NHC(C(=CH@ 4)C(=O)OH)=O)NH2)C(=O)NH<regId="het\_65\_E(2)">  
#CC\_naphth\_hydraz.txt 2  
C[1]H:CH:CH:CH:C[9]:CH:C(:CH:CH:C:@ 1:@ 9)NHNHC=O<regId="hziide\_naphth(2)">  
#CC\_ON\_6.txt 2  
CH2(C[4]:C(:CH:CH:C(:CH:@ 4)CH2NHC[TAC=4])OCH2)<regId="anisol\_B(2)">  
#DD\_thiocarbam\_5.txt 2  
C[1]C[TAC=4](SC(NC=@ 1)=S)<regId="thio\_carbam\_ene(2)">  
#DD\_thioamid\_2.txt 2  
C(N(CH)CH):CNHC(=S)CH<regId="thio\_amide\_D(2)">  
#DD\_phenaz.txt 2  
N[1]N=C(C[4]C(=C@ 1C)C=CC=@ 4)C<regId="het\_65\_Da(2)">  
#EE\_5hets20.txt 2  
S[1]:CH:CH:C(OCH2):C(C(=O)NHNH):@ 1<regId="thiophene\_D(2)">  
#EE\_sixNS.txt 2  
C[1]:CN=CC(=CNC)S@ 1<regId="het\_thio\_6\_ene(2)">  
#FF\_cyano\_A1.txt 2  
CH2CH(C#N)C(=O)C<regId="cyano\_keto\_A(2)">  
#FF\_cyano\_11.txt 2  
C[1](=C(NH2)N(C[7]:C(:C:C:C:C:@ 7)C(=O)OH)N=C@ 1C=O)Any[IS=C#N,C=S]<regId="anthranil\_acid\_C(2)">  
#FF\_Naphth\_2.txt 2  
C[1]:C:C:C[4]:C[5]:C(:C:C:C:C:@ 1:@ 5)C=NNH@ 4<regId="naphth\_amino\_C(2)">  
#FF\_Naphth\_3.txt 2

C[1]:C:C:C[4]:C[5]:C(:C:C:C:C:@1:@5)N=NN@4<regId="naphth\_amino\_D(2)">  
#FF\_misc\_4.txt 2  
C[1]N=C(SC=@1)NNS(=O)=O<regId="thiazole\_amine\_G(2)">  
#FF\_misc8.txt 2  
C[1]:C:C:C[4]:C(:C:@1):N:C(:N:C:@4)NHC[13]=NC(C=CNH@13)(CH2)CH2<regId="het\_66\_B(2)">  
#FF\_chromen\_2\_1.txt 2  
C[1]:C:C(OCH):CH:CH:C(:@1)C[2]:CH:CH:C(OCH2):CH:C(:@2)C(=O)O@1<regId="coumarin\_A(2)">  
#GG\_bicycle\_2.txt 2  
C[1]:N:C:C:C:C(:@1)C(NH2)=C(C(=O)~O[f])S@1<regId="anthranil\_acid\_D(2)">  
#GG\_bicycle\_5.txt 2  
C[1]:Hev:C:C:C:C(:@1):N:C(CH=C(OH)C):C(CH=C(OH)C):N:@1<regId="het\_66\_C(2)">  
#GG\_thiophene\_bis.txt 2  
C[1]=CSC(=C@1NH2)CH=CHC[13]=CC=CS@13<regId="thiophene\_amino\_E(2)">  
#GG\_4cycle.txt 2  
C[1]:C:C:C[4]:C(:C:@1):N:C[8]:C(:N:@4):C:C:C[13]:C:@8:C:C:C:C:@13<regId="het\_6666\_A(2)">  
#GG\_NSO2N.txt 2  
C:CNHS(=O)(=O)NHC:C<regId="sulfonamide\_E(2)">  
#CC\_Ar2\_p\_anils\_1.txt 2  
C[1]:C(:C:C:C:C:@1)N(C[TAC=3])C[TAC=3])NH2<regId="anil\_di\_alk\_K(2)">  
#JJ\_AAmisc1.txt 2  
N[1]=C(C[2]:C:C:C:C:C:@2)CH2C(OH)(C(F)(F)(F))N(Any[IS=C:C:C:C:C,C(=S)C':C:C:C:C:C])@1<regId="het\_5\_C(2)">  
#JJ\_enone\_D3.txt 2  
C[2]:C:C:C:C:C(:@2)C(=O)CH=C[1]C(=O)NHC(=O)C(=CHC[3]:C:C:C:C:C:@3)NH@1<regId="ene\_six\_het\_B(2)">  
#JJ\_AA\_steroid.txt 2  
O=C[1]CCC[2]C[3]C(=O)CC[4]CCCC(@4)C(@3)CCC(@2)=C@1<regId="steroid\_A(2)">  
#JJ\_ON\_indol\_4.txt 2  
C[1]:C[2]:C(:N:C:N:@2):C(OH):CH:C(:@1):C(C=O):C:N(C):@1<regId="het\_565\_A(2)">  
#JJ\_thiiniiums\_3.txt 2  
C[TAC=4]N[+1](C[TAC=4]OH)=CSCH<regId="thio\_imine\_ium(2)">  
#KK\_110\_fives\_7.txt 2  
C[1](=O)C(=CHNHC[2]:C(C(=O)OH):C:C:C:C:@2)N=C(C[3]:C:C:C:C:C:@3)O@1<regId="anthranil\_acid\_E(2)">  
#KK\_111\_furzone\_2.txt 2  
C[1](:O:C(Any[IS=H,CH2]):CH:CH:@1)C(Any[IS=H,CH2])=NNHC[2]:C:C:N:C:C:@2<regId="hzone\_furan\_B(2)">  
#KK\_111\_thiophbtz.txt 2  
C[1]((:S:C(Any[IS=H,CH2]):CH:CH:@1)C(Any[IS=H,CH2])C(=O)NHC[2]:N:C:C:S:@2<regId="thiophene\_E(2)">  
#KK\_114\_maleimide\_1.txt 2  
C:CCH2CH(C=O)N[1]C(=O)C[2]HCH2C=CCH2CH(@2)C(=O)@1<regId="ene\_misc\_B(2)">  
#KK\_107\_SIM\_2.txt 2  
C[1](C=O)(C:C)S[TAC=2]C=NNH@1<regId="het\_thio\_5\_B(2)">  
#KK\_EE\_amino\_5hets\_4a.txt 2  
NH2C[1]:S:C(NHC(=O)C[3]:C:C:C:C:C:@3):C(C#N):C(C[2]:Hev:Hev:Hev:Hev:Hev:@2):@1<regId="thiophene\_amino\_F(2)">  
#KK\_CC\_o\_anisidine\_5A.txt 2  
CH2OC[1]:CH:CH:C(CH2):CH:C(:@1)NHCH2C[2]:C(Any[IS=CH2,OCH2]):C:C:C:C:@2<regId="anil\_OC\_alk\_D(2)">  
#KK\_DD\_t\_Bu\_1C.txt 2  
CH3C(CH3)(CH3)C[1]:CH:C(C(CH3)(CH3)CH3):C(OCHN):C:CH:@1<regId="tert\_butyl\_A(2)">  
#KK\_DD\_thiophenezone\_1C.txt 2  
C[1]((:C:C:O:C(CH):@1)C=NNHC(=S)NH<regId="thio\_urea\_J(2)">  
#FF\_purine\_1.txt 2  
NHC[3]=NC(=NC[7]=NN=C(N@3@7)SC)NHC<regId="het\_thio\_65\_B(2)">  
#NN\_FF\_chromen\_7\_4.txt 2  
C[1]:C(CH2CH=CH2):C:C:C:C(:@1)CH=C(C(=O)NHC:C)C(=O)O@1<regId="coumarin\_B(2)">  
#M1\_DD\_NNNS.txt 2  
C[1](=S)N[2]C:CN=NC(@2)=NNH@1<regId="thio\_urea\_K(2)">  
#KK\_EE\_amino\_5hets\_5a.txt 2  
C:C:C:C:C:CC[1]:S:C(NHC(=O)C):C(C(=O)OH):C:@1<regId="thiophene\_amino\_G(2)">  
#QQ\_BB\_orthoanils\_1.txt 2  
NH2C[1]:C(NHCH(C)CHCH2):C:CH:CH:CH:@1<regId="anil\_NH\_alk\_D(2)">  
#QQ\_DD\_NNSO\_1.txt 2  
S=C[1]NHN=C(C[2]:CH:CH:C(OCH2):CH:CH:@2)O@1<regId="het\_thio\_5\_C(2)">  
#QQ\_DD\_thi\_azene.txt 2  
S=CC[1]:N[2]:C:C:C:C:C(:@2):C:C:@1<regId="thio\_keto\_het(2)">  
#KK\_115\_polyimine.txt 2  
C[1]~C(~N~N~C(~CH2)~CH2)~N~S~C~@1<regId="het\_thio\_N\_5B(2)">  
#o\_quinones.txt 2  
C[1](C:=CC:=CC@1=Het)=Het<regId="quinone\_D(2)">  
#amino\_furan\_ene.txt 2  
CH2N(CH2)C[1]:CH:CH:C(:O:@1)CH=CC#N<regId="anil\_di\_alk\_furan\_B(2)">  
#sixes\_last2.txt 2

O=C[1]C:CCH2NC(@1)=CH<regId="ene\_six\_het\_C(2)">  
#five\_pz\_sulf.txt 2  
C:CN[1]:N:C[2]CH2S[TAC=2]CH2C(:@2)C(:@1)NHC(=O)CH=CH<regId="het\_55\_A(2)">  
#bim\_SCH2.txt 2  
N[1]:C[2]:C:C:C:C:C(:@2):N(CH2):C(:@1)SCH2C(=O)NHN=CHCH=CH<regId="het\_thio\_65\_C(2)">  
#hydroquin\_enone.txt 2  
C[1](OH):C:C:C(OH):C(:C:@1)C(=[!R]CN)C=O<regId="hydroquin\_A(2)">  
#anthran\_OH.txt 2  
C[1](OH):C:C:C(NHC(=O)C:C):C(:C:@1)C(=O)OH<regId="anthranil\_acid\_F(2)">  
#pyrrole\_2one.txt 2  
N[1](CH2)C[2]=C(C:CC(=O)@2)C=C(CH2)@1<regId="pyrrole\_I(2)">  
#KK\_EE\_amino\_5hets\_7a.txt 2  
CHNHC[1]:S:C(CH):C(CH):C(C(=O)NHC:C):@1<regId="thiophene\_amino\_H(2)">  
#isatins\_3.txt 2  
C:CN[!r]=C[2]C(=Het)C[5]:C(:C:C:C:C:@5)N@2<regId="imine\_one\_fives\_C(2)">  
#DD\_acylzone\_3.txt 2  
C[1]:C(:C:C:C:@1)C(=O)NHN=C[2]C[3]:C(:C:C:C:C:@3)C[4]:C@2:C:C:C:C:@4<regId="keto\_phenone\_zone\_A(2)">  
#dyes7A.txt 2  
C[1]:C:C:C:C:C(:@1)N(CH2)CH=CHC=[!R]CHCH=CC=[R]NC[2]:C:C:C:C:C:@2<regId="dyes7A(2)">  
#pyridiniums\_B.txt 2  
C[1]:C:C:C:C:Hev(:@1):C(Any):Hev:N[+1](~C:C):Hev:@1<regId="het\_pyridiniums\_B(2)">  
#166\_het\_5\_t.txt 2  
N[1](C[2]:C:C:C:C:C:@2)N=C(CH2)CH(SC)C(=O)@1<regId="het\_5\_D(2)">  
#fives\_P\_1.txt 1  
C[1]:C(:C:C(C(=O)OH):C:C:@1)NHC[12]=NC(=CHS@12)C[18]:C:C:C(CH(CH)(CH)):C:C:@18<regId="thiazole\_amine\_H(1)">  
#fives\_P2.txt 1  
CH2NHC=NNHC[1]=NC(C:C)=CHS@1<regId="thiazole\_amine\_I(1)">  
#fives\_P3.txt 1  
C:CNHC(=O)C[1]N=NSC(NHC:C)=@1<regId="het\_thio\_N\_5C(1)">  
#fives\_Q.txt 1  
O=S(=O)(C:C)NHC[7]=NC(=C(S@7))C:C<regId="sulfonamide\_F(1)">  
#fives\_Q2.txt 1  
O=S(=O)(C:C)NHNHC[7]=NC(=C(S@7))C:C<regId="thiazole\_amine\_J(1)">  
#sixesB5.txt 1  
S[1]C[2]:C(C(CH2)=C@1CH2):C(:N:C:N:@2)NN=CC[19]=CC=CO@19<regId="het\_65\_F(1)">  
#enone\_H.txt 1  
C(=O)CH=C(OH)C(OH)=CHC(=O)C<regId="keto\_keto\_beta\_E(1)">  
#enone\_H2.txt 1  
C[1]H:CH:CH:CH:C[2]CH2C(=O)C(=C(CH2)CH2)C(:@2):@1<regId="ene\_five\_one\_B(1)">  
#meldrums2.txt 1  
C:CNHN=C(CH2)CH2C(CH2)=NNHC:C<regId="keto\_keto\_beta\_zone(1)">  
#thioimide\_2.txt 1  
C[TAC=4]S[TAC=2]C(=N-Hev:Hev:Hev:Hev)NHN=C<regId="thio\_urea\_L(1)">  
#fivesZ.txt 1  
C[1](SC(=CHC:C)N(C:C)N=@1)C=O<regId="het\_thio\_urea\_ene(1)">  
#multiCN.txt 1  
C[1](~C(:N:C[4]:C(:C:@1C#N):C(:C:C:C:@4)C#N)NH2)NH2)C#N<regId="cyano\_amino\_het\_A(1)">  
#chalc2.txt 1  
Hev[1]:Hev:Hev:Hev(:Hev:Hev:@1)CH=CHC(NHNHC[16]N(N=NN=@16)C)=O<regId="tetrazole\_hzide(1)">  
#naphthol\_B.txt 1  
C[1]:C(:CH:CH:CH:CH:@1):C(:C(:CH:CH:@1)C(=NC:C)CH2)OH<regId="imine\_naphthol\_A(1)">  
#misc1.txt 1  
C[1](~C(:CH:C[5]:C(:CH:@1):C(:CH:C(:CH:@5)C[15]:CH:CH:C(:C(:CH:@15)OCH2)OCH2)NHCH3)OCH2)OCH2<regId="misc\_anisole\_A(1)">  
#misc2.txt 1  
C[1]:C:C:C[7]:C(:C:@1)NC[13]=C(S@7)SC(=C@13)CH2<regId="het\_thio\_665(1)">  
#misc7.txt 1  
C[1]:C:C[3]:C(:C:C:@1)N(CH2CH2N[14]C(C[16]:C(C@14=O):C:C:C:C:@16)=O)CH2CH2@3<regId="anil\_di\_alk\_L(1)">  
#misc8.txt 1  
C[1](~C:C:C:C:@1)C[7]=CC(C[10]C(C(=C@7)OH)=COC=@10)=O)SCH2<regId="colchicine\_B(1)">  
#misc12.txt 1  
C[TAC=4]C[1]:CH:CH:C(C(=O)NHCH(CH2CH2SCH3)C(=O)OH):CH:CH:@1<regId="misc\_aminoacid\_A(1)">  
#AAmisc3.txt 1  
N[1]:C(C[2]:C:C:C:C:@2):C(C[3]:C:C:C:C:@3):N(N=[!R]C):C(NH2):@1<regId="imidazole\_amino\_A(1)">  
#AAmisc4.txt 1  
C(C[1]:C:C:C(OH):C:C:@1)(C[2]:C:C:C(OH):C:C:@2)OS(=O)=O<regId="phenol\_sulfite\_A(1)">  
#AAmisc7.txt 1  
C[1]:C:C:C[7]:C(:C:@1):N:C(:C:N:@7)CH2C(=O)C:C)CH2C(=O)C:C<regId="het\_66\_D(1)">  
#AAmisc8.txt 1



C[1](:CH:C(OCH2):C(OCH2):CH:CH:@1)C(=O)CH2N(CH2)C[2]:C:C:C(CH2)C:C:@2<regId="misc\_anisole\_B(1)">  
#AAmisc9.txt 1  
CH2C[1]=NN=NN(C[2]:CH:CH:C(OCH3):CH:CH:@2)@1<regId="tetrazole\_A(1)">  
#AAmisc9\_A.txt 1  
C[1](C)CH2C(=O)N[2]N=C(NH2)C(NH2)=C(@2)N=@1<regId="het\_65\_G(1)">  
#AA\_misc14.txt 1  
C(C:C)(C:C)(C:C)SC:CC(=O)OH<regId="misc\_trityl\_A(1)">  
#AA\_misc15.txt 1  
O=C(C[1]:CH:C(OCH3):N:C(OCH3):CH:@1)NHCH(CH2)CH2<regId="misc\_pyridine\_OC(1)">  
#AA\_misc17.txt 1  
N[1]CH(C:C)CH2C(=O)NHC(NH)=@1<regId="het\_6\_hydropyridone(1)">  
#AA\_misc18.txt 1  
C[1](C:C)CH2CH(C(=O)C)CH(C(=O)OH)CH2C(C:C)=@1<regId="misc\_stilbene(1)">  
#AA\_misc19.txt 1  
CH(C[1]:CH:CH:C(Cl):CH:CH:@1)C[2]:CH:CH:C(Cl):CH:CH:@2)OCH2CH2CH2C[3]N(CH3)CH=CHN=@3<regId="misc\_imidazole(1)">  
#BBparanils\_6.txt 1  
N[1]:CH:C(NH2):CH:CH:C(NHC:C):@1<regId="anil\_NH\_no\_alk\_A(1)">  
#BB\_o\_anils\_5.txt 1  
NH(C[1]:CH:CH:CH:CH:C(OH):@1)C[2]N=NC(=N)OC(NH2)=@2<regId="het\_6\_imidate\_B(1)">  
#BB\_anils\_misc.txt 1  
NH(C[1]:CH:CH:C(CH2):CH:CH:@1)CH2CH2C[2]:CH:CH:C(OCH2):CH:CH:@2<regId="anil\_alk\_B(1)">  
#BB\_anils2.txt 1  
C[1]:C:C:C[7]:C(:C:@1)C(C[12]:C@7:C:C:C(:C:@12)NH2)=CHC<regId="styrene\_anil\_A(1)">  
#BB\_aminall1.txt 1  
C[1]:C:C:C[4]:C(:C:@1)COC(N@4C(=O)CH2)(C(=O)OH)CH2<regId="misc\_aminall\_acid(1)">  
#BB\_meta\_anil2.txt 1  
N[1]:C(NH2):C(CH2):CH:C(CH2):C(NH2):@1<regId="anil\_no\_alk\_D(1)">  
#CC\_pCN\_O3.txt 1  
NH(C[1]:C:C:C:C:C:@1)C(C)(C)C[2]:CH:CH:C(OCH2):CH:CH:@2<regId="anil\_alk\_C(1)">  
#CCpNH\_OC\_1\_1.txt 1  
NH(C[1]:CH:CH:C(OCH3):C(OCH):CH:@1)C(=O)NHCH2CH2CH2N(CH3)C:C<regId="misc\_anisole\_C(1)">  
#CC\_CH2dioxy3.txt 1  
C[1]:C(:C:C[4]:C(:C:@1)CC@4)OCO@1<regId="het\_465\_misc(1)">  
#CC\_ON\_1\_3\_1.txt 1  
C[1](:CH:CH:C(OCH2):CH:C(C(=O)OH):@1)NHC:C<regId="anthranil\_acid\_G(1)">  
#CC\_N\_7.txt 1  
C[1]H:C(:CH:N:C[7]:C:@1:CH:CH:CH:CH:@7)CH2N[20]C[21]:C(CH2CH2@20):CH:CH:CH:CH:@21<regId="anil\_di\_alk\_M(1)">  
#CC\_7CO2H\_indol.txt 1  
C[1]:C:C:C[4]:C(:C:@1C(=O)OH)NHC(=C@4C:C)C:C<regId="anthranil\_acid\_H(1)">  
#DD\_thiourea.txt 1  
C:CN(CH2)CH2CH2CH2NHC(=S)NHC[1]:C(Hal):CH:C(CH2):CH:CH:@1<regId="thio\_urea\_M(1)">  
#DD\_thiobenz\_misc2.txt 1  
N[1]:C[2]:C(:C:C[5]:C:@1N=C(S@5)N)SC(=N@2)N<regId="thiazole\_amine\_K(1)">  
#DD\_NOS\_4.txt 1  
N=C[1]SC(=N)N=C@1<regId="het\_thio\_5\_imine\_A(1)">  
#DD\_thioamid.txt 1  
C[1]:C:C:C:N:C(:@1)C(=S)NHC[2]:C:C:C:C:C(:@2)OCH2<regId="thio\_amide\_E(1)">  
#DD\_phenothz\_6.txt 1  
C[1]CH2SC[2]:CH:CH:CH:C(:@2)C(OC:C)C(:@1):CH:CH:CH:CH:@1<regId="het\_thio\_676\_B(1)">  
#DD\_sulfon\_pyr.txt 1  
CH3C[5]:N:C(NHS(C[11]:CH:CH:C(:CH:CH:@11)OCH2CH2CH2)(=O)=O):CH:CH:CH:@5<regId="sulfonamide\_G(1)">  
#DD\_thionomorph.txt 1  
C(=O)(N[3]CCSCC@3)C[9]:C(:CH:CH:CH:CH:@9)SCH2<regId="thio\_thiomorph\_Z(1)">  
#DD\_enone\_3.txt 1  
C[1]:C:C[3]:C[4]:C(:C:@1):C:C:C:C:@4C(=CC@3=O)OCH2<regId="naphth\_ene\_one\_A(1)">  
#DD\_cf3.txt 1  
C[1]:C:C:C(:C[5]:C:@1:C(:C:C:C:@5)C(C=C@1C(F)(F)F)=O)N<regId="naphth\_ene\_one\_B(1)">  
#DD\_polycyc.txt 1  
C[1]:C:C:C:C[5]:C:@1:C:C[8]:C(:N:@5):N:C[12]:C(:C:@8N):C:C:C:C:@12<regId="amino\_acridine\_A(1)">  
#DD\_suber2.txt 1  
C[1]:C:C:C:C[5]:C:@1C(C[8]C@5=NHév=CC@8)=O<regId="keto\_phenone\_B(1)">  
#DD\_suber\_4.txt 1  
C[1]:CH:CH:C(OCH2):CH:C(:@1)C(=NNHC[2]:CH:CH:C(C(=O)OH):CH:CH:@2)C[3]:CH:C(OCH2):CH:CH:C(:@3)@1<regId="hzone\_acid\_A(1)">  
#DD\_sulfonAZ.txt 1  
C[1](:CH:CH:C(:CH:CH:@1)NH2)S(=O)(=O)NHC[19]:CH:CH:CH:N:N:@19<regId="sulfonamide\_H(1)">  
#EE\_pyrrole\_5.txt 1  
C[1]H=CHN(CH2)C[7]:C@1:C[9]CH=CHN(C:@9:C(:C:@7OCH2)OCH2)CH2<regId="het\_565\_indole(1)">  
#EE\_pyrrole\_6.txt 1

C[1](=C[2]C(=C(N@1C(O)=O)CH2)SCH2S@2)CH2<regId="pyrrole\_J(1)">  
#EE\_pz\_thz.txt 1  
S[1]C=CN=C@1C[6]CH=NNHC=@6NH2<regId="pyrazole\_amino\_B(1)">  
#EE\_5hets9.txt 1  
C[1](NHC(C[2]:CH:CH:CH:CH:CH:@2)=C(CH2)CH=@1)C(=O)OH<regId="pyrrole\_K(1)">  
#EE\_5hets30.txt 1  
C[1]((:O:C(C):CH:CH:@1)C(=O)NHC[1]:CH:CH:CH:CH:C(C(=O)OH):@1<regId="anthranil\_acid\_I(1)">  
#EE\_thiocarbz.txt 1  
Hev:CC(=S)NHNHC:Hev<regId="thio\_amide\_F(1)">  
#FF\_cyano\_10.txt 1  
C[1](=O)C(C(C#N)=CHN)C(N)C=C@1<regId="ene\_one\_C(1)">  
#FF\_cyano\_14a.txt 1  
C[1](=C[2]N(C(C=CN@2)=O)N=C@1C[11]NC=CC=@11)C#N<regId="het\_65\_H(1)">  
#FF\_cyano\_15.txt 1  
O=C[1]C(C#N)=NNCC@1<regId="cyano\_imine\_D(1)">  
#FF\_cyano\_16.txt 1  
C[1](:C[2]:C:C:C:C:@2:N:N:C:@1)C(C:C)C#N<regId="cyano\_misc\_A(1)">  
#FF\_misc8.txt 1  
C[1]:C:C:C[4]:C(:C:@1)N(C(C=C@4)(CH2)CH2)C(=O)NHC[23]:C:C:C(:C:@23)OCH2)OCH2<regId="ene\_misc\_C(1)">  
#FF\_misc12.txt 1  
C[1]:C:C:C[4]:C(:C:@1):N:C(C(:N:@4)C[11]:C:C:C:C:@11)C[17]:C(:C:C:C:C:@17)OH<regId="het\_66\_E(1)">  
#FF\_misc13.txt 1  
CH2C(OH)=C(C(=O)CH2)CHC#C<regId="keto\_keto\_beta\_F(1)">  
#FF\_zene.txt 1  
C[1]:C[2]:C(:C:C[5]:C:@1:C:C:C:C:@5)N(C(=N@2)COC(=O)C[18]:C:C(C:C(:C:@18)NH2)NH2)H<regId="misc\_naphthimidazole(1)">  
#FF\_naphth\_imine.txt 1  
C[1]((:C:C:C[4]:C[5]:C(:C:C:C:C:@1:@5)C(C=C@4)=N)N<regId="naphth\_ene\_one\_C(1)">  
#FF\_fluorenone.txt 1  
C[1]((:C:C[3]:C(:C[5]:C:C:C:C:C:@1:@5)C[11]:C(C@3=O):C:C:C:C:@11)OH<regId="keto\_phenone\_C(1)">  
#FF\_chromen\_6.txt 1  
C[1](C=NC[4]:C(O@1):C:C:C(Cl):C:@4)=O<regId="coumarin\_C(1)">  
#FF\_misc\_8\_1.txt 1  
C[1]=CN(C(C[5]:C:C:C:C:C@1:@5)(C#N)C(=S)S)C=O<regId="thio\_est\_cyano\_A(1)">  
#FF\_Bim\_urea.txt 1  
C[1](=NC[3]:CH:CH:CH:C:@3N@1C)NHC(NHC[20]:CH:C:C:C:CH:@20)=O<regId="het\_65\_imidazole(1)">  
#FF\_Ncycle\_3.txt 1  
NH(C:C)C[5]:C(C(=O)OH):C:C:C(:N:@5)C:C<regId="anthranil\_acid\_J(1)">  
#GG\_bicycle\_3.txt 1  
C[1]:C:C:C:C:C(:@1)SC(=NN=C[2]C=CC=CC=C@2)N(CH2)@1<regId="colchicine\_het(1)">  
#GG\_bicycle\_4.txt 1  
C[1]:CH:C(OCH2):C(OCH2):CH:C(:@1)C(C)=C(C)SCH2@1<regId="ene\_misc\_D(1)">  
#GG\_carbazole\_2.txt 1  
C[1]:CH:CH:CH:CH:C:@1)C(CH2)=C(C:C)N(-[!R]C:C)@1<regId="indole\_3yl\_alk\_B(1)">  
#GG\_anthril\_SO3\_2.txt 1  
NH2C[1]:C(OH):C:C(S(=O)(=O)OH):C:C:@1<regId="anil\_OH\_no\_alk\_A(1)">  
#GG\_thiophene\_2.txt 1  
S[1]:C:C:C(C[2]:C:S:C(NH2):N:@2):CH:@1<regId="thiazole\_amine\_L(1)">  
#GG\_thiophene\_4.txt 1  
C[1]=C(NH2)NN=C(@1)C[2]=C(CH2)OCH=CH@2<regId="pyrazole\_amino\_A(1)">  
#GG\_thiophene\_6.txt 1  
N[1]=NSC=C(@1)C[2]ON=C(C:C)N=@2<regId="het\_thio\_N\_5D(1)">  
#GG\_indoline.txt 1  
C[1]((:C:C:C[4]:C(C[6]CC[8]:C(C@6N@4):C:C:C:@8):C:@1)C[TAC=4]<regId="anil\_alk\_indane(1)">  
#GG\_tricycle3.txt 1  
C[1]:CH:CH:CH:CH:C(:@1)CH=CHC[2]C(C#N)CH2CHN(@2)@1<regId="anil\_di\_alk\_N(1)">  
#GG\_tricycle5.txt 1  
C[1]:C:C[2]:C:C:C:C:C(:@2):C:C(:@1)N(CH2)C(=O)C(C:CNHCH2)=N@1<regId="het\_666\_C(1)">  
#enone\_F.txt 1  
C(OH):CC(=O)CH=C(C)C<regId="ene\_one\_D(1)">  
#JJ\_CC\_ON\_indol\_2.txt 1  
C[1]:CH:CH:C(N(CH2)CH2):CH:C(:@1):C(S(=O)=O):CH:NH:@1<regId="anil\_di\_alk\_indol(1)">  
#JJ\_CC\_ON\_indol\_3.txt 1  
C[1]:CH:CH:C(NH2):CH:C(:@1):CH:CH:N(CH2):@1<regId="anil\_no\_alk\_indol\_A(1)">  
#JJ\_FF\_dhp\_2.txt 1  
S[1:TAC=2]C=C(C(C#N)C(C)(C=O)C(Any[IS=C=O.C#N])=C(NH2)@1<regId="dhp\_amino\_CN\_G(1)">  
#JJ\_FF\_dhp\_3.txt 1  
N[1]C=C(C=O)C(C[3]:C:C:C(N(CH2)CH2):C:C:@3)C[2]=C(~N~C(~S)~N~C(~N)~@2)@1<regId="anil\_di\_alk\_dhp(1)">  
#JJ\_FF\_anthrilNN\_1.txt 1

C[1]:C:C:C:C:C(:@1)C(=O)NHC[2]:C:C:C:C:C(:@2)C(=O)NHNHC[3]:N:C:C:S:@3<regId="anthranil\_amide\_A(1)">  
#JJ\_DD\_anthran.txt 1  
C[1]:C:C:C:C:C[5]:C:@1:C:C[8]:C(:C:@5C=NNHC[2]:C:C:C:C:C:@2):C:C:C:C:@8<regId="hzone\_anthran\_Z(1)">  
#JJ\_enone\_D2.txt 1  
C[1]:C:C:C:C:C(:@1)CHNC(=O)C(NHCH2)=CHC(=O)C[2]:C:C:C(COCH2):C:C:@2<regId="ene\_one\_amide\_A(1)">  
#JJ\_EE\_5hets17.txt 1  
S[1]:CH:CH:CH:C(:@1)C[2]NHN=C(C[4]:C:C:N:N:C:C:@4)C[3]:C(:C:C:C:C:@3)N=@2<regId="het\_76\_A(1)">  
#JJ\_EE\_5hets18.txt 1  
O[1]:CH:CH:CH:C(CH2NHC(=S)N(CH)CH2C[2]:C:C:C:C:C:@2):@1<regId="thio\_urea\_N(1)">  
#JJ\_FF\_chromen\_5.txt 1  
C[2]:C:C:C:C:C(:@2)N(CH)CHCHCHNHC(=O)C[1]C(CH2)=CHC(=O)OC(CH2)=@1<regId="anil\_di\_alk\_coum(1)">  
#JJ\_FF\_misc\_10.txt 1  
C[1]:C:C:C[2]:C:C:C:C:C(:@2):C(@1)CHC[TAC=4]NC(@1)=CHC(=O)N(CH2)CH2<regId="ene\_one\_amide\_B(1)">  
#JJ\_thiiniums\_2.txt 1  
C[1]:C:C:C:C:C(:@1)C[2]=NN[3]:C[4]:C:C:C:C:C(:@4):N[+1]:C(:@3)SC[TAC=4]@2<regId="het\_thio\_656c(1)">  
#KK\_110\_fives\_3.txt 1  
C[1](=O)C(=C(CH2)NHCH2CH2CH2)N=C(C[2]:C:C:C:C:C:@2)O@1<regId="het\_5\_ene(1)">  
#KK\_110\_imide 1  
C[2]:C:C:C:C:C(:@2)N[1]C(=O)C(SC[3]:C:C:C:C:C:@3)=CHC(=O)@1<regId="thio\_imide\_A(1)">  
#KK\_116\_thidhp\_1.txt 1  
N[1]HN=C(NH)SC(C:C)=C(C:C)@1<regId="dhp\_amidine\_A(1)">  
#KK\_thiourea\_2.txt 1  
C[1]H:C(NHC(=S)NHCHC[3]:O:C(CH):CH:CH:@3):CH:C[2]:C(OCH2O@2):CH:@1<regId="thio\_urea\_O(1)">  
#KK\_thiourea\_8.txt 1  
C[1]H:C(NHC(=S)NHC[2]:C:C:C:C:C:@2):CH:C(N(CH2)CH2):CH:CH:@1<regId="anil\_di\_alk\_O(1)">  
#KK\_107\_indole.txt 1  
O=C-[!R]N[1]:C:C:C[2]:C(NHC(=S)NH@2):@1<regId="thio\_urea\_P(1)">  
#KK\_107\_pyrazol\_1.txt 1  
C(F)(F)C(=O)NHC[1]:CH:N(CH2CH2OCH2C:C):N:CH:@1<regId="het\_pyraz\_misc(1)">  
#KK\_108\_diazox\_4.txt 1  
N[2]=NC[1]:N:Het:N:C(:@1)N=NC:C@2<regId="diazox\_C(1)">  
#KK\_108\_diazox\_5.txt  
C[1]H(OH)C[2]:N:Het:N:C(:@2)CH(OH)C=C@1<regId="diazox\_D(1)">  
#KK\_AA\_misc10.txt 1  
C[1](CH2CH2@1)(C(=O)NHC[2]:C:C:C(OCH2O):C:C:@2)S(=O)(=O)C:C<regId="misc\_cyclopropane(1)">  
#KK\_109\_enam.txt  
C[1]:CC(=O)C=C(@1)N=CHN(C[TAC=4])C[TAC=4]<regId="imine\_ene\_one\_B(1)">  
#KK\_109\_coumarin\_2.txt 1  
C[1]:C[2]C(C[3]:C:C:C:C:C:@3)=CHC(=O)OC(:@2):C:C(OCH2C:O:C):C:@1<regId="coumarin\_D(1)">  
#KK\_109\_furmorph\_1.txt 1  
C[3]:C(CH2):C(CH2):O:C(:@3)CH2NCH2CH(OCH2)CH2OC[1]:C:C:C[2]:C(OCH2O@2):C:@1<regId="misc\_furan\_A(1)">  
#KK\_110\_rhod\_2.txt 1  
N[1](C[2]:C:C:C:C:C:@2)C(=O)SCH(NHC[3]:C[4]:C:C:C:C:C(:@4):C:C:C:@3)C(=O)@1<regId="rhod\_sat\_E(1)">  
#KK\_110\_rhod\_5.txt 1  
N[1](C(=O)C[2]:C:C:C:C:C:@2)C(=NC[3]:C:C:C:C:C:@3)SCH2C(=O)@1<regId="rhod\_sat\_imine\_A(1)">  
#KK\_110\_rhod\_7.txt 1  
N[1](C[2]:C:C:C:C:C:@2)C(=O)SCH2C(=S)@1<regId="rhod\_sat\_F(1)">  
#KK\_110\_rhod\_13.txt 1  
N[1](CH2)C(=S)N(C:C)C(=NC:C)C(=NC:C)@1<regId="het\_thio\_5\_imine\_B(1)">  
#KK\_110\_rhod\_14.txt 1  
S[1]C(=NNH)SC(=NC:C)C(=NC:C)@1<regId="het\_thio\_5\_imine\_C(1)">  
#KK\_110\_fives\_5.txt 1  
C[1](=O)C(=CHC[2]:C:C(Hal):C:C:C(:@2)OCH2)N=C(SCH2)S@1<regId="ene\_five\_het\_N(1)">  
#KK\_112\_thiocarbam.txt 1  
CH2SC(=S)NHCH2C:C<regId="thio\_carbam\_A(1)">  
#KK\_113\_dianilide\_1.txt 1  
C[1](NHC(=O)CH2CH2C:C):CH:CH:C(CH2):C(NHC(=O)CH2CH2C:C):CH:@1<regId="misc\_anilide\_A(1)">  
#KK\_113\_lutidine\_1.txt 1  
C[1](NHC(=O)NHCH2CH2CH2):C(CH2):CH:C(Br):CH:C(CH2):@1<regId="misc\_anilide\_B(1)">  
#KK\_113\_mannich.txt 1  
C[1]:C(OCH2N(C:COCH2)CH2@1):CH:C:C:CH:@1<regId="mannich\_B(1)">  
#KK\_113\_mannich\_1.txt 1  
C[1]:C(OCH2N(CH2)CH2@1):CH:C(O):C(O):CH:@1<regId="mannich\_catechol\_A(1)">  
#KK\_113\_anilpip.txt 1  
NH(C[1]:CH:CH:C(CH(CH2)(CH2)):CH:CH:@1)CH2CH2N(CH2)(CH2)<regId="anil\_alk\_D(1)">  
#KK\_114\_acid.txt 1  
N[1]:C:C:C(C(~O)~O):C:C(:@1):C:C(C:C):C(C(=O)C:C):@1<regId="het\_65\_I(1)">  
#KK\_114\_alkylarylurea.txt 1

C[1](C(C[TAC=4])(C[TAC=4])NHC(=O)N(CH2CH2)CH2CH2CHCH2C:C):CH:CH:CH:C(C(=CH2)CH2):CH:@ 1<regId="misc\_urea\_A(1)">  
#KK\_114\_imipram\_1.txt 1  
C[1]H(N[2]:CH:N:CH:CH:@ 2)C[3]:CH:C(Br):CH:CH:C(:@ 3)CH2CH2C[4]:CH:CH:CH:CH:C(:@ 4)@ 1<regId="imidazole\_C(1)">  
#KK\_114\_arylstilbene.txt 1  
C(=CHCH2N[1]:CH:N:CH:CH:@ 1)(C:C)C:C<regId="styrene\_imidazole\_A(1)">  
#KK\_115\_Arthiaz\_2.txt 1  
C[1](NHC:CCH2):N:C(C[2]:C:C:N:C:C:@ 2):CH:S:@ 1<regId="thiazole\_amine\_M(1)">  
#KK\_115\_Arthiaz\_3.txt 1  
C[1](CH(CH2)CH2CH2NCH2C[2]:NH:C:C:C:@ 2):N:C(C[3]:C:C:C:C:@ 3):CH:S:@ 1<regId="misc\_pyrrole\_thiaz(1)">  
#KK\_115\_pyrrolester.txt 1  
N[1]H:C(C(=O)OCH2):C(CH2):C(CH2CH2):C(CH2CH2):@ 1<regId="pyrrole\_L(1)">  
#KK\_115\_thiobim.txt 1  
C[1](SCH2C(=O)NHC:C):N:C[2]:CH:C(Hal):C:CH:C(:@ 2):NH:@ 1<regId="het\_thio\_65\_D(1)">  
#KK\_116\_thianisole\_1.txt 1  
C[1](OCH2):C(OCH2):CH:C[2]C=CCHSC(:@ 2):CH:@ 1<regId="ene\_misc\_E(1)">  
#KK\_116\_thipyr\_2.txt 1  
N[1]HC(=S)CH(C#N)CH(C:C)CH=C(C:C)@ 1<regId="thio\_cyano\_A(1)">  
#KK\_116\_thipyr\_3.txt 1  
N[1]:C(S[TAC=2]C[3]:C(NH2):C:C:C:C:@ 3):C(C#N):C(C[2]:C:C:C:C:@ 2):C(C#N):C(NH2):@ 1<regId="cyano\_amino\_het\_B(1)">  
#KK\_116\_pyr\_1.txt 1  
N[1](C[2]:C:C:C(OCH2):C:C:@ 2)C(=O)C(C#N)=CC(N[3]:C:N:C:C:@ 3)=N@ 1<regId="cyano\_pyridone\_G(1)">  
#KK\_116\_benzofuran\_1.txt 1  
O[1]:C(C(~O)~O):C:C[2]:CH:C(OCH2):C(OCH2):CH:C(:@ 2):@ 1<regId="het\_65\_J(1)">  
#KK\_116\_diyneone.txt 1  
C#CC(=O)C#C<regId="ene\_one\_yne\_A(1)">  
#KK\_116\_aminonaphthol.txt 1  
C[1](NH2):C(OH):C(C=O):CH:C[2]:CH:CH:CH:CH:C(:@ 2):@ 1<regId="anil\_OH\_no\_alk\_B(1)">  
#KK\_111\_furzid\_3.txt 1  
C[1](~O:C(Any[IS=H,CH2]):CH:CH:@ 1)C(=O)NHN=C(Any[IS=H,CH2])C[2]:C:C(AnyAnyAnyC[3]:O:C:C:C:@ 3):C:C:C:@ 2<regId="hzone\_acyl\_misc\_A(1)">  
#KK\_EE\_amino\_5hets\_9a.txt 1  
S(=O)(=O)NHC[1]:S:C(CH):C(CH):C(C(=O)NH):@ 1<regId="thiophene\_F(1)">  
#KK\_CC\_o\_anisidine\_5B.txt 1  
CH2OC[1]:CH:CH:CH:CH:C(:@ 1)NHCH2CH(OH)CH2<regId="anil\_OC\_alk\_E(1)">  
#KK\_CC\_o\_anisidine\_5E.txt 1  
CH2OC[1]:CH:CH:CH:CH:C(:@ 1)NHCH(C=O)S<regId="anil\_OC\_alk\_F(1)">  
#KK\_DD\_t\_Bu\_1A.txt 1  
N[1]=NNN=C[2]C=CC=C(@ 2)@ 1<regId="het\_65\_K(1)">  
#KK\_DD\_thiophenezone\_1B.txt 1  
C[1]:CH:S:C(C=O):C(:@ 1)C(=O)NN=C(@ 1)NH2<regId="het\_65\_L(1)">  
#NN\_FF\_chromen\_7\_1.txt 1  
C[1]:C(Br):C:C[2]:C(:C:C:O:@ 2):C(:@ 1)CH=CC(=O)O@ 1<regId="coumarin\_E(1)">  
#NN\_FF\_chromen\_7\_2.txt 1  
C[1]:C:C:C:C:C(:@ 1)CH=C(C(=O)NHC[2]:C(Br):C:O:N:@ 2)C(=O)O@ 1<regId="coumarin\_F(1)">  
#NN\_FF\_chromen\_7\_3.txt 1  
C[1]:C(Hal):C:C(Hal):C:C(:@ 1)CH=C(C(=O)NH2)C(=NH)O@ 1<regId="coumarin\_G(1)">  
#NN\_FF\_chromen\_7\_5.txt 1  
C[1]:C:C:C:C:C(:@ 1)CH=C(C(=O)NHC[2]:S:C:C(C:S:CH):N:@ 2)C(=O)O@ 1<regId="coumarin\_H(1)">  
#NN\_sixesB8\_4.txt 1  
CH2S[TAC=2]C[1]:N:C[2]OC=NC:CC(:@ 2):N:N:@ 1<regId="het\_thio\_67\_A(1)">  
#NN\_CC\_ON\_8\_1.txt 1  
S(=O)(=O)(C[1]:C:N(CH2):C:N:@ 1)NHC[2]:C:N:N(CH2C:COCH2):C:@ 2<regId="sulfonamide\_I(1)">  
#NN\_CC\_ON\_8\_2.txt 1  
C[1]:C(OCH2O@ 1):C(CH2N[2]CH2CH2C:C@ 2):CH:CH:CH:@ 1<regId="het\_65\_mannich(1)">  
#NN\_CC\_ON\_8\_3.txt 1  
CH2OC:CCH2NHC[1]:CH:C[2]:N:CH:N(CH):C(:@ 2):CH:CH:@ 1<regId="anil\_alk\_A(1)">  
#NN\_JJ\_thiziniums\_2.txt 1  
N[1](C[2]:C:C:C:C:C:@ 2)CH=N[+1](C[3]:C:C:C:C:C:@ 3)C(=NC[4]:C:C:C:C:C:@ 4)N@ 1<regId="het\_5\_inium(1)">  
#NN\_CC\_Me2\_p\_anils\_5.txt 1  
CH2N(CH2)C[1]:C:C[2]:N:C(SCH2):S:C(:@ 2):C:C:@ 1<regId="anil\_di\_alk\_P(1)">  
#M15\_naphth\_F.txt 1  
C[1]:C(:CH:CH:CH:CH:@ 1):C(:CH:CH:CH:@ 1)C(CH2)=NNHC(=S)NHC:C:C<regId="thio\_urea\_Q(1)">  
#M1\_DD\_sixNS.txt 1  
C[1](~N:Hev:N:C(C[2]:O:CH:CH:CH:@ 2):N:@ 1)SC[TAC=4]<regId="thio\_pyridine\_A(1)">  
#KK\_108\_melamime\_1.txt 1  
N[1]:C(N(CH2)CH2):N:C(N(CH2)CH2):N:C(N(CH)C=O):@ 1<regId="melamine\_B(1)">  
#P\_isothidiaz\_1.txt 1  
C[1](C[2]:C:C:C:C:C:@ 2):N:C(N(CH2)CH2CH2NC(=O)C[3]:C(C(=O)OH):C:C:C:C:@ 3):S:N:@ 1<regId="misc\_phthal\_thio\_N(1)">  
#P\_pyracylzone.txt 1

N[1]:CH:CH:CH:CH:C(:@1)C(=O)NHN=CHC[2]:C(OCH2C(=O)OH):C:C:C:C:@2<regId="hzone\_acyl\_misc\_B(1)">  
#KK\_DD\_t\_Bu\_1B.txt 1  
CH3C(CH3)(CH3)C[1]:C(OH):C(C(CH3)(CH3)CH3):CH:C(CH2C[2]:CH:CH:C(OH):C:C:@2):CH:@1<regId="tert\_butyl\_B(1)">  
#QQ\_BB\_orthoanils.txt 1  
NH2C[1]:C(NH2):CH:CH:C[2]:N:O:N:C(:@1):@2<regId="diazox\_E(1)">  
#QQ\_BB\_orthoanils\_2.txt 1  
NH2C[1]:C(NHS(=O)=O):CH:C(NHCH2):C(Hal):CH:@1<regId="anil\_NH\_no\_alk\_B(1)">  
#QQ\_BB\_orthoanils\_3.txt 1  
NH2C[1]:C(N=C[2]CH=C~C~C=C@2):CH:CH:CH:CH:@1<regId="anil\_no\_alk\_A(1)">  
#QQ\_BB\_orthoanils\_3.txt 1  
NH2C[1]:C(N[2]:C:C:C:C:@2):CH:C(CH2):C(CH2):CH:@1<regId="anil\_no\_alk\_B(1)">  
#QQ\_DD\_thi\_eneaz.txt 1  
S=CC(CH2)=C(CH2)N(CH2)CH2<regId="thio\_ene\_amine\_A(1)">  
#QQ\_EE\_furane.txt 1  
C[1]:COC[2]CH2C(=O)OC(@2)@1<regId="het\_55\_B(1)">  
#thioimide\_4.txt 1  
OC(=O)CH2S[TAC=2]C(=NC#N)NHC[1]:C:C:C:C:C:@1<regId="cyanamide\_A(1)">  
#thioacetene.txt 1  
O=CC[1]=C(SC(=CHC)S@1)C=O<regId="ene\_one\_one\_A(1)">  
#sixes\_last1.txt 1  
O=C[1]NNC(Hev:Hev)=NC(@1)=CH<regId="ene\_six\_het\_D(1)">  
#acceptor.txt 1  
O=CCH=C(C#N)C<regId="ene\_cyano\_E(1)">  
#fur\_acid\_CN.txt 1  
HOC(=O)C[1]:C(OH):CH:C(:CH:CH:@1)C[2]:CH:CH:C(:O:@2)CH=C(C#N)C[3]:N:C:C:N:@3<regId="ene\_cyano\_F(1)">  
#acid\_ar\_zone.txt 1  
C[1]:C:C:C:C:C(:@1)N(C[2]:C:C:C:C:C:@2)N=CHC[3]:Hev:C(:CH:CH:@3)C[4]:C:C(C(=O)OH):C:C:C:@4<regId="hzone\_furan\_C(1)">  
#anilfuran.txt 1  
NH2C[1]:CH:CH:C(:CH:CH:@1)C[2]:CH:C(C=O):C(CH2):O:@2<regId="anil\_no\_alk\_C(1)">  
#acid\_zone\_fur.txt 1  
HOC(=O)C[1]:C:C:C:C:C:@1)NN=CHC[2]:Hev:C(:CH:CH:@2)C[3]:C:C:C:C:C:@3<regId="hzone\_acid\_D(1)">  
#acid\_furzonide.txt 1  
HOC(=O)C[1]:C:C:C:C:C:@1)C:Hev:CC=NNHC(=O)CH2O<regId="hzone\_furan\_E(1)">  
#DD\_OHPyr\_2.txt 1  
HOC[2]:N:C(NH2):Hev:C:C(CH2C(=O)O):@2<regId="het\_6\_pyridone\_NH2(1)">  
#fives\_B2het.txt 1  
C[1](=Het)C(N=CS@1)=O<regId="imine\_one\_fives\_D(1)">  
#pyrrole\_Aroxime.txt 1  
N[1](C[2]:C:C:C:C:C:@2)CH=CHCH=C(C=NOH)@1<regId="pyrrole\_M(1)">  
#pyrrole\_AlkAlkAlk.txt 1  
N[1](CHC[2]:CH:CH:C:CH:CH:@2)C(CH)=CHCH=C(CH)@1<regId="pyrrole\_N(1)">  
#pyrrole\_pentsub.txt 1  
N[1](CH2)C(CH2)=C(C(=O)C)C(C)=C(C:C)@1<regId="pyrrole\_O(1)">  
#pyrrole\_acrCN.txt 1  
N[1](C)CH=CHCH=C(CH=C(C#N)C[2]:N:C:C:S:@2)@1<regId="ene\_cyano\_G(1)">  
#pyrrole\_Arsulfon\_A.txt 1  
N[1](C[2]:C(NHS(=O)(=O)C[3]:S:C:C:C:@3):C:C:C:C:@2)CH=CHCH=CH@1<regId="sulfonamide\_J(1)">  
#pyrrole\_benzamide.txt 1  
N[1](C[2]:CH:CH:CH:CH:C(C(=O)NHCH(CH2)CH2OC:C):@2)CH=CHCH=CH@1<regId="misc\_pyrrole\_benz(1)">  
#thiocarbz\_3.txt 1  
C[1](;C:C:C:C:C:@1)NHC(=S)NNHCH=CHC=O<regId="thio\_urea\_R(1)">  
#meldrum\_1a.txt 1  
C[1](C(=O)CH2CCH2C(=O)@1)=C(NH)C=O<regId="ene\_one\_one\_B(1)">  
#dhp\_amino\_CNSS.txt 1  
NH2C[1]=C(C#N)CH(C:C)SC[TAC=4]S@1<regId="dhp\_amino\_CN\_H(1)">  
#anis\_122\_A.txt 1  
CH2OC[1]:CH:CH:CH:CH:C(:@1)NHC[2]:C[3]:C(OCH2):C:C:C(OCH2):C(:@3):N:C:C:@2<regId="het\_66\_anisole(1)">  
#anis\_112\_B.txt 1  
CH2OC[1]:CH:CH:C(OCH2):CH:C(:@1)NHC[2]:S:C:C(C[3]:C:C:C(OCH2):C:C:@3):N:@2<regId="thiazole\_amine\_N(1)">  
#pyridiniums\_C.txt 1  
C[1]~N(C:C)~C~C~C~C(~@1)~C[2]~N~C~C~C~N[+1](~@2)~N~@1<regId="het\_pyridiniums\_C(1)">  
#166\_het\_5\_p.txt 1  
N[1](C[2]:C:C:C:C[3]:C:C:C:C:C(:@3):@2)N=C(CH2)CH2C(=O)@1<regId="het\_5\_E(1)">

**Supplementary Table S9.** Enrichment calculations for all problematic substructures with 4 or more member compounds. These are listed in order of population size. This value can be used to match up with the structure as listed in Figure S1.

Search WEHI DB for structures in  
Freq\_Hit\_5\_norethan150.hits

								Total	Enrichment <sup>a</sup>
Filename	6	5	4	3	2	1	0		
<b>ene_six_het_A</b>	10	20	21	30	69	105	228	483	66%
<b>hzone_phenol_A</b>	5	4	7	17	208	82	156	479	154%
<b>anil_di_alk_A</b>	31	66	27	35	28	59	232	478	81%
<b>indol_3yl_alk</b>	6	15	23	18	51	80	268	461	42%
<b>quinone_A</b>	40	57	48	41	42	56	86	370	265%
<b>azo_A</b>	29	30	33	43	24	55	110	324	145%
<b>imine_one_A</b>	17	20	18	15	38	49	164	321	60%
<b>mannich_A</b>	2	4	13	15	59	57	146	296	64%
<b>anil_di_alk_B</b>	6	18	7	18	22	43	137	251	52%
<b>anil_di_alk_C</b>	15	23	13	10	26	29	130	246	67%
<b>ene_rhod_A</b>	16	41	21	26	32	39	60	235	227%
<b>hzone_phenol_B</b>	2	2	9	6	38	54	104	215	55%
<b>ene_five_het_A</b>	6	14	24	14	39	40	64	201	152%
<b>anil_di_alk_D</b>	12	10	18	11	24	28	95	198	79%
<b>imine_one_isatin</b>	1	6	12	8	35	46	81	189	67%
<b>anil_di_alk_E</b>	8	12	17	10	15	27	97	186	64%

Filter WEHI DB using  
Freq\_Hit\_5\_morethan150.hits  
Then Search Remaining Compounds using  
Freq\_Hit\_5\_less than150.hits

<b>thiaz_ene_A</b>	4	10	9	10	12	27	56	128	80%
<b>pyrrole_A</b>	1	16	13	14	11	21	42	118	131%
<b>catechol_A</b>	4	7	10	4	10	21	36	92	97%
<b>ene_five_het_B</b>	0	4	4	2	14	22	44	90	55%
<b>imine_one_fives</b>	7	11	9	5	9	18	30	89	137%
<b>ene_five_het_C</b>	3	9	7	7	7	13	39	85	85%
<b>hzone_pipzn</b>	1	7	8	13	8	9	33	79	112%
<b>keto_keto_beta_A</b>	3	2	0	3	10	15	35	68	51%
<b>hzone_pyrrol</b>	4	4	1	2	9	10	34	64	59%
<b>ene_one_ene_A</b>	1	2	3	5	6	16	24	57	71%
<b>cyano_ene_amine_A</b>	0	8	6	7	5	19	11	56	236%
<b>ene_five_one_A</b>	1	2	1	1	11	11	28	55	57%
<b>cyano_pyridone_A</b>	1	3	3	4	4	16	23	54	65%
<b>anil_alk_ene</b>	1	6	6	3	7	11	17	51	135%
<b>amino_acridine_A</b>	3	8	4	4	4	7	16	46	144%
<b>ene_five_het_D</b>	4	7	8	9	13	5	0	46	na
<b>thiophene_amine_Aa</b>	2	2	5	4	3	11	18	45	89%
<b>ene_five_het_E</b>	1	8	3	4	4	11	13	44	154%
<b>sulfonamide_A</b>	0	3	4	5	10	2	19	43	116%

thio_ketone	4	4	6	5	3	4	17	43	135%
sulfonamide_B	5	1	2	1	3	14	15	41	80%
anil_no_alk	2	1	1	3	5	8	20	40	60%
thiophene_amino_Ab	0	2	2	1	5	7	23	40	43%
het_pyridiniums_A	7	3	5	3	2	5	14	39	143%
anthranil_one_A	0	2	2	2	4	9	19	38	53%
cyano_imine_A	1	1	1	1	5	7	21	37	43%
diazox_sulfon_A	1	4	2	2	4	6	17	36	76%
hzone_anil_di_alk	1	4	5	0	5	7	13	35	115%
rhod_sat_A	0	6	6	6	6	7	2	33	1200%
hzone_enamin	0	0	3	2	9	1	15	30	93%
pyrrole_B	4	5	9	0	0	2	3	29	600%
thiophene_hydroxy	0	4	2	1	3	7	11	28	91%
cyano_pyridone_B	0	1	2	2	7	4	11	27	109%
imine_one_sixes	0	1	1	0	5	7	13	27	54%
dyes5A	8	3	4	1	0	2	9	27	178%
naphth_amino_A	0	1	2	2	2	3	15	25	47%
naphth_amino_B	1	9	3	5	5	1	1	25	2300%
ene_one_ester	0	1	0	1	4	11	7	24	86%
thio_dibenzo	2	2	5	3	1	2	8	23	163%
cyano_cyano_A	0	1	6	3	4	5	4	23	350%
hzone_acyl_naphthol	1	1	0	3	5	7	5	22	200%
het_65_A	0	0	2	2	2	7	8	21	75%
imidazole_A	2	2	2	2	2	2	7	19	143%
ene_cyano_A	3	2	1	0	3	4	6	19	150%
anthranil_acid_A	1	2	2	4	2	5	3	19	375%
dyes3A	5	0	1	2	3	1	7	19	157%
dhp_bis_amino_CN	0	1	5	7	1	4	1	19	1400%
het_6_tetrazine	2	2	3	2	0	4	5	18	180%
ene_one_hal	1	0	0	2	2	4	8	17	63%
cyano_imine_B	0	2	3	4	2	2	4	17	275%
thiaz_ene_B	1	2	1	0	1	1	11	17	36%
ene_rhod_B	0	1	5	1	4	4	4	16	275%
thio_carbonate_A	0	9	1	3	0	2	0	15	na
anil_di_alk_furan_A	0	0	2	1	6	4	2	15	450%
ene_five_het_F	1	1	0	1	2	6	4	15	125%

Filter WEHI DB using

Freq\_Hit\_5\_lessthan150.hits

Then Search Remaining Compounds using

Freq\_Hit\_5\_lessthan15.hits

anil_di_alk_F	2	1	3	2	0	1	5	14	160%
hzone_anil	0	1	1	0	1	1	10	14	30%
het_5_pyrazole_OH	1	0	1	3	4	3	2	14	450%
het_thio_666_A	1	3	3	0	0	3	3	13	233%
styrene_A	0	0	2	1	4	1	5	13	140%
ene_rhod_C	0	3	0	0	2	2	6	13	83%
dhp_amino_CN_A	0	0	2	2	6	0	3	13	333%
cyano_imine_C	0	3	1	2	2	2	2	12	400%
thio_urea_A	0	4	3	1	0	3	1	12	800%

thiophene_amino_B	2	2	1	0	1	3	3	12	200%
keto_keto_beta_B	0	0	2	1	3	0	6	12	100%
keto_phenone_A	0	2	0	0	1	1	7	11	43%
cyano_pyridone_C	2	2	0	3	3	1	0	11	na
thiaz_ene_C	1	1	1	0	4	1	3	11	233%
hzone_thiophene_A	2	3	2	1	0	0	3	11	267%
ene_quin_methide	0	2	2	1	3	2	0	10	na
het_thio_676_A	0	3	0	1	0	0	6	10	67%
ene_five_het_G	0	0	2	1	1	1	5	10	80%
acyl_het_A	0	2	2	1	0	2	2	9	250%
anil_di_alk_G	2	1	0	0	0	2	4	9	75%
dhp_keto_A	4	0	2	0	1	1	1	9	700%
thio_urea_B	0	0	2	2	2	0	3	9	200%
anil_alk_bim	1	1	2	1	1	0	3	9	200%
imine_imine_A	0	3	0	0	0	1	5	9	60%
thio_urea_C	0	1	1	1	3	2	1	9	600%
imine_one_fives_B	3	0	1	1	0	1	3	9	167%
dhp_amino_CN_B	0	2	2	1	1	1	2	9	300%
anil_OC_no_alk_A	0	1	0	1	2	1	3	8	133%
het_thio_66_one	1	0	1	1	2	2	1	8	500%
styrene_B	1	0	3	3	1	0	0	8	na
het_thio_5_A	1	0	1	2	0	1	3	8	133%
anil_di_alk_ene_A	0	2	1	0	2	1	2	8	250%
ene_rhod_D	0	1	1	1	0	1	4	8	75%
ene_rhod_E	0	3	1	1	1	0	2	8	300%
anil_OH_alk_A	0	4	3	0	1	0	0	8	na
pyrrole_C	0	1	2	2	1	1	1	8	600%
thio_urea_D	1	1	3	0	1	1	1	8	600%
thiaz_ene_D	1	1	1	1	0	1	3	8	133%
ene_rhod_F	0	2	1	1	1	0	3	8	167%
thiaz_ene_E	0	1	3	1	0	1	2	8	250%
het_65_B	0	1	2	0	0	3	1	7	300%
keto_keto_beta_C	0	1	1	1	1	3	0	7	na
het_66_A	0	2	1	1	0	1	2	7	200%
thio_urea_E	0	0	1	2	1	2	1	7	400%
thiophene_amino_C	0	1	1	0	1	0	4	7	75%
hzone_phenone	0	0	3	2	1	0	1	7	600
ene_rhod_G	0	2	1	2	0	2	0	7	na
ene_cyano_B	0	1	1	2	2	0	1	7	600%
dhp_amino_CN_C	1	0	0	2	1	2	1	7	400%
het_5_A	0	1	0	1	1	1	3	7	100%
ene_five_het_H	0	1	0	0	2	3	0	6	na
thio_amide_A	0	2	2	0	0	1	1	6	400%
ene_cyano_C	1	0	1	1	1	1	1	6	400%
hzone_furan_A	0	2	1	0	0	1	2	6	150%
anil_di_alk_H	1	1	0	2	1	0	1	6	500%
het_65_C	1	1	0	0	0	0	4	6	50%
thio_urea_F	0	1	0	1	4	0	0	6	na
ene_five_het_I	0	1	0	0	1	3	1	6	200%
keto_keto_gamma	1	1	1	0	1	1	0	5	na
quinone_B	0	1	2	2	0	0	0	5	na



het_6_pyridone_OH	1	0	0	0	3	0	1	5	400%
hzone_naphth_A	0	1	1	0	0	2	1	5	200%
thio_ester_A	0	1	3	0	0	0	1	5	400%
ene_misc_A	0	1	0	0	1	0	3	5	67%
cyano_pyridone_D	1	0	0	1	1	0	2	5	150%
het_65_Db	0	1	0	0	3	1	0	5	na
het_666_A	1	0	3	0	0	0	1	5	400%
diazox_sulfon_B	0	2	0	0	1	0	2	5	150%
anil_NH_alk_A	0	1	0	0	0	2	2	5	50%
sulfonamide_C	0	0	2	0	0	2	1	5	200%
het_thio_N_55	0	2	2	1	0	0	0	5	na
keto_keto_beta_D	0	2	2	1	0	0	0	5	na
ene_rhod_H	0	2	0	1	1	1	0	5	na
imine_ene_A	2	0	1	2	0	0	0	5	na
het_thio_656a	0	0	1	0	1	2	1	5	200%
pyrrole_D	1	0	0	0	1	3	0	5	na
pyrrole_E	0	2	0	0	0	0	3	5	67%
thio_urea_G	2	0	0	0	0	0	3	5	67%
anisol_A	0	1	1	0	0	0	3	5	67%
pyrrole_F	0	1	1	0	0	1	2	5	100%
dhp_amino_CN_D	0	1	1	0	0	1	2	5	100%
thiazole_amine_A	0	1	0	1	0	2	0	4	na
het_6_imidate_A	1	0	0	0	0	1	2	4	50%
anil_OC_no_alk_B	0	1	1	0	0	0	2	4	100%
styrene_C	0	1	0	0	0	1	2	4	50%
azulene	2	0	1	0	0	0	1	4	300%
furan_acid_A	0	0	1	0	1	0	2	4	100%
cyano_pyridone_E	0	1	0	0	1	1	1	4	200%
anil_alk_thio	1	3	0	0	0	0	0	4	na
anil_di_alk_I	1	2	0	0	0	0	1	4	300%
het_thio_6_furan	0	1	1	1	1	0	0	4	na
anil_di_alk_ene_B	0	2	0	0	1	0	1	4	300%
imine_ene_B	2	1	0	1	0	0	0	4	na
anil_OC_alk_A	1	0	0	0	2	0	1	4	300%
ene_five_het_J	0	0	1	0	1	0	2	4	50%
pyrrole_G	0	1	1	0	0	0	2	4	100%
ene_five_het_K	0	2	0	0	0	1	1	4	200%
cyano_ene_amine_B	0	0	2	1	0	0	1	4	300%
thio_ester_B	0	4	0	0	0	0	0	4	na
ene_five_het_L	0	0	1	1	0	1	1	4	200%
hzone_thiophene_B	0	1	1	1	0	1	0	4	na
dhp_amino_CN_E	0	0	3	1	0	0	0	4	na
het_5_B	0	1	1	0	0	1	1	4	200%

a. Total number of compounds that hit from 2-6 assays expressed as a percentage of those compounds that have hit none of the six assays. na = not applicable (enrichment = infinity because of absence of any compound that hit no assays). Note that a high count of 2 is usually due to interference with HTS assays E and F and may usefully point to chelators

**Supplementary Table S10.** A comparison of functional group prevalence in MDDR 2008.1, the inaugural WEHI 93K HTS Library, and the current new CTX 153K HTS Library (comprising a collaboration library, and compounds from Vendor A and Vendor B).<sup>a</sup>

Entry	Group	Prevalence in (%)				
		MDDR <sup>a</sup>	WEHI <sup>b</sup>	CTX 153K HTS Library		
				16K sub-Library <sup>c</sup>	87K sub-Library <sup>d</sup>	49K sub-Library <sup>e</sup>
1	Aromatic N	62	43	65	56	52
2	Amides	50	32	42	52	47
3	Alcohols	29	4.5	2.1	1.8	7.8
4	Carboxylic Acid	26	6.0	2.3	4.3	2.5
5	Lactams (not beta)	23	9.7	23	17	11
6	Phenols	12	3.8	1.7	1.8	2.1
7	Sulfonamides	10	6.8	19	19	8.7
8	Amines	10	7.1	6.4	10	36
9	Linear Ureas	6.3	4.0	4.6	4.6	2.8
10	Linear Carbamates	5.7	1.8	0	1.0	1.2
11	ketones	4.3	1.0	0	5.7	6.1
12	sulfones	3.8	1.6	3.8	2.8	1.2
13	Anilines (1°, 2°, 3° and cyclic or linear)	2.4	7.3	6.4	5.2	4.7
14	Cyclic acyl ureas	2.4	1.9	3.6	2.0	0.3
15	Cyclic carbamates	1.9	0.8	1.2	0.72	0.4
16	Cyclic imides (phthalimides)	1.6 (0.6)	2.7	0.06 (0)	1.5	1.2
17	Cyclic ureas	1.1	0.8	1.8	1.5	0.4
18	Linear Imides	0.18	0.6	0.7	0.08	0.13

a. MDDR 2008.1

b. Inaugural WEHI 93K HTS Library.

c. Collaborative HTS Library of 15,667 compounds established using separate criteria to those in Table 8, which we term the CTX 16K HTS Sub-Library.

d. 87,059 compounds from Vendor A, which we term the CTX 87K HTS Sub-Library

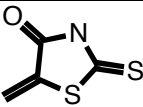
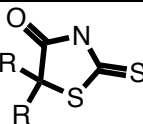
e. 49,144 compounds from Vendor B, which we term the CTX 49K HTS Sub-Library

It can be seen that there are similar overall trends in functional group prevalence for the WEHI 93K HTS Library compared with the new CTX 153K HTS Library, which we have segregated according to compound source (Collaborative Library, and Vendor A and Vendor B). That is, our filters for problematic compounds do not appear to have markedly altered functional group prevalence. Aromatic nitrogens are extremely common (entry 1) as are linear amides (entry 2) and then lactams (entry 5) and to some extent sulfonamides (entry 7), amines (entry 8) and then

anilines (entry 13) and alcohols (entry 3). Thereafter the remaining functional groups become significantly less prevalent. Even though trending the same way, for many functional groups there are significant fold differences depending on compound source. For example, the CTX 16K HTS Sub-Library has twice as many lactams as the CTX 16K HTS Sub-Library, whereas the latter has a very high percentage of amines (36%) and much higher than any other compound source. Similar variations are noted for alcohols, sulfonamides, ketones, cyclic carbamates, cyclic imides and linear imides.

We have also included analysis of MDDR 2008.1. Of note is the lower prevalence of alcohols, carboxylic acids, phenols, and linear carbamates in ordered screening compounds compared with MDDR, but the latter may be biased with highly populated classes (e.g. angiotensin converting enzyme inhibitors, Sweeteners, natural products etc) and this further analysis has not been undertaken.

**Table S11.** Rhodanines as problematic compounds.

Substructure <sup>a</sup>	Number of AlphaScreen® assays hit							Total Cpd	Enrichment <sup>b</sup>
	6	5	4	3	2	1	0		
 <b>ene_rhod_A</b>	16	41	21	26	32	39	60	235	227%
 <b>rhod_sat_A</b>	0	6	6	6	6	7	2	33	1200%

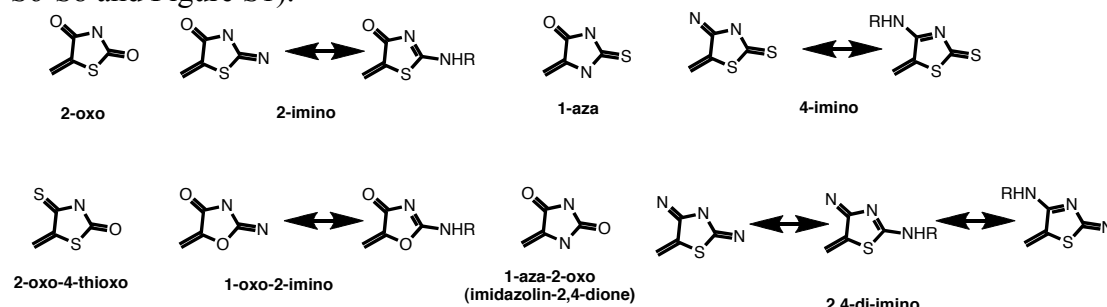
a. Name corresponds to name in Supplementary Table S6-S8 and Figure S1

b. Total number of compounds that hit from 2-6 assays expressed as a percentage of those compounds that have hit none of the six assays.

Rhodanines frequently occur as primary hits in our AlphaScreen® assays. There are alkylidene rhodanines with an enrichment of 227% for a total of 235 compounds, and saturated rhodanines, fewer in number (33) but with an astonishingly high enrichment of 1,200%. Focusing on the more common alkylidene rhodanines, these have been reported as inhibitors of anthrax lethal factor,<sup>21,22</sup> glycosyltransferase MurG,<sup>23</sup> SARS coronavirus,<sup>24</sup> PRL-3,<sup>25</sup> glycogen synthase kinase-3 $\beta$ ,<sup>26,27</sup> HIV-1 integrase,<sup>28,29</sup> extracellular signal-regulated kinase 2,<sup>30</sup> tau aggregation,<sup>31</sup> botulinum neurotoxin type A,<sup>32</sup> *Plasmodium falciparum* enoyl-acyl carrier protein reductase,<sup>33</sup> leucocyte migration (by stabilizing activated  $\alpha_M\beta_2$  integrin),<sup>34</sup> hepatitis C RNA polymerases,<sup>35-37</sup> tumour necrosis factor- $\alpha$ ,<sup>38,39</sup> UDP-galactopyranose mutase,<sup>40</sup> tyrosine kinase p56 Lck SH2 domain,<sup>41</sup> VHR phosphatase,<sup>42</sup> formylpeptide receptor,<sup>43</sup> RNA polymerase assembly,<sup>44</sup> cholesterol accumulation,<sup>45</sup> peptide deformylase,<sup>46</sup> human apurinic/aprimidinic endonuclease I,<sup>47</sup> aggrecanase-2,<sup>48</sup> *Trypanosoma brucei* dolicholphosphate mannose synthase<sup>49</sup>, *Helicobacter pylori* shikimate kinase,<sup>50</sup> protein tyrosine phosphatase-1B,<sup>51</sup> *Yersinia* tyrosine phosphatase,<sup>52</sup> retinoid x receptor- $\alpha$ ,<sup>53</sup> *Yersinia* protein kinase A,<sup>54</sup> and bacterial DNA adenine methyltransferases.<sup>55</sup>

Despite extensive downstream investigation reported by many of these studies and attempts at describing SAR, these rhodanines are often bereft of meaningful SAR and poor correlation is often observed with cell-based activity where investigated. Clues as to why can be gleaned from the work of Carlsson *et al.*<sup>40</sup> who report that rhodanines can undergo extremely facile reaction with nucleophiles and so reactivity-based SAR is often observed. Indeed, workers at Amgen<sup>36,37</sup> have elucidated the crystal structure of a rhodanine-hepatitis C NS5b RNA complex revealing the formation of a covalent bond. Furthermore, work by DuPont<sup>38</sup> and Bristol-Myers Squibb<sup>39</sup> shows that rhodanines can undergo light-induced irreversible binding to tumor necrosis factor receptor-1. Disturbingly, two different reaction sites on the rhodanine scaffold are involved. Moreover, rhodanines are known chelators of transition metals and this has also been shown in the context of a protein's active site.<sup>22</sup> Common substituents include phenolic OH, carboxylic acid, halogen and nitro. It is possible that the former two can contribute more to chelation and the latter to reactivity (including light induced) but this currently remains speculative. Abbott have recently provided supporting evidence that rhodanines could be problematic protein-reactive screening compounds.<sup>6,7</sup>

We have additionally found that variations of rhodanines such as those shown below can also be frequent hitters in our AlphaScreen® assays (See Supplementary Tables S6-S8 and Figure S1).

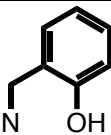


These also often appear as screening hits in the literature alongside the more common rhodanine parent structures in studies we have already cited above.<sup>21,23,25,26,38-42,51,52,55</sup>

We expand briefly for the “2-oxo” rhodanines and the “2-imino” rhodanines. The “2-oxo” rhodanines (thiazolidin-2,4-diones) have also been extensively studied for their peroxisome proliferator activated receptor gamma agonist activity as troglitazone-like compounds<sup>56-60</sup> and for their aldose reductase inhibition as epalrestat-like compounds.<sup>61,62</sup> They have also been reported as inhibitors of protein tyrosine phosphatase-1b and low molecular weight protein tyrosine phosphatase<sup>63</sup> and Pim-1 / Pim-2.<sup>64</sup> They are reported to also inhibit CFTR<sup>65</sup> and DYRK1A.<sup>66</sup> The SAR is generally confusing but especially so when the nitrogen is substituted<sup>61-63,65</sup> and hence the ring not given the possibility of being deactivated through deprotonation of the relatively acidic NH proton.

Similar comments apply to “2-imino” rhodanines, which have also been reported as inhibitors of cyclooxygenase and lipooxygenase,<sup>67</sup> HCV NS5B polymerase,<sup>68-71</sup> PTP1B,<sup>72</sup> mitogen-activated protein kinase phosphatase-1,<sup>73</sup> DNA synthesis,<sup>74</sup> and cysteine proteases.<sup>75</sup> They can also be highly cytotoxic.<sup>76,77</sup> We have not undertaken an in-depth analyses of this sub-class, but it is plausible that these compounds could also interfere in assays in many of the ways that the parent rhodanines do, and indeed light-induced protein reactivity has been reported.<sup>38,39</sup>

**Table S12.** Phenolic mannich bases as problematic compounds

Substructure <sup>a</sup>	Number of AlphaScreen® assays hit							Total Cpds	Enrichment <sup>b</sup>
	6	5	4	3	2	1	0		
 <b>mannich_A</b>	2	4	13	15	59	57	146	296	64%

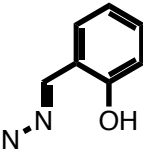
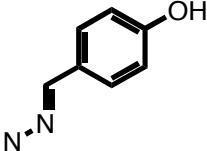
a. Name corresponds to name in Supplementary Table S6-S8 and Figure S1

b Total number of compounds that hit from 2-6 assays expressed as a percentage of those compounds that have hit none of the six assays.

Compounds with a 2-hydroxybenzylamine moiety frequently occur as primary hits in our AlphaScreen® assays with an enrichment of 64% for a total of 296 compounds. A brief survey of the literature revealed apparent hits for a number of different targets, including Hsp90,<sup>78</sup> L-type calcium channels,<sup>79</sup> bacterial ribosomal A-site,<sup>80</sup> human murine double mutant 2,<sup>81</sup> Mcl-1<sup>82</sup> and E2F.<sup>83</sup> As for the rhodanines, the molecules are often strikingly similar to each other. Disturbingly, this class of compounds can readily form highly reactive quinone methides<sup>84,85</sup> and they are also known to be metal chelators.<sup>86-90</sup> Perhaps not surprisingly, these compounds can be potently cytotoxic towards bacteria.<sup>91</sup> These compounds can also be highly cytotoxic to mammalian cells and susceptible to form intensely colored solutions (Theola Louie, personal communication).

While this paper was under review, the ability of this class to covalently modify proteins via the quinone methide was confirmed by Sanofi-Aventis and firms up these as problematic screening compounds.<sup>92</sup>

**Table S13.** Hydroxyphenylhydrazones as problematic compounds.

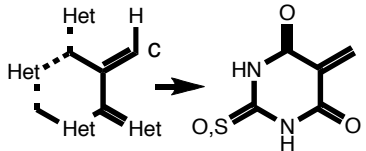
Substructure <sup>a</sup>	Number of AlphaScreen® assays hit							Total Cpds	Enrichment <sup>b</sup>
	6	5	4	3	2	1	0		
 <b>hzone_phenol_A</b>	5	4	7	17	208	82	156	479	154%
 <b>hzone_phenol_B</b>	2	2	9	6	38	54	104	215	55%

a. Name corresponds to name in Supplementary Table S6-S8 and Figure S1

b. Total number of compounds that hit from 2-6 assays expressed as a percentage of those compounds that have hit none of the six assays.

Hydrazones *per se* are close to being problematic in our assays (enrichment 28%) but scrutiny reveals certain sub-types of hydrazones are far more problematic than others and *cart blanche* removal of all hydrazones is unnecessary and likely to limit chemical diversity. Thus, 2- and 4-hydroxyphenylhydrazones frequently occur as primary hits in our AlphaScreen® assays with a respective enrichment of 154% and 55% for a total of 479 and 215 compounds respectively. These compounds often coexist in the literature with the other screening hits reported above<sup>21,41,74,75,78,91</sup> and have also been reported as inhibitors of type II topoisomerase,<sup>93</sup> anthrax lethal factor,<sup>94</sup> methionyl-tRNA synthetase,<sup>95</sup> 17 $\beta$ -hydroxysteroid dehydrogenase-1,<sup>96</sup> *Mycobacterium tuberculosis*<sup>97</sup> and Yersinia Type III secretion.<sup>98</sup> Closely related compounds such as the corresponding 2-hydroxybenzamides and the like are also usually co-reported screening hits.<sup>30,75,78-80,91,92,94,99,100</sup> These compounds can not only have unusual spectroscopic properties,<sup>101,102</sup> but they are also well known chelators.<sup>103,104</sup> Furthermore, the 2-hydroxyphenylhydrazones have been postulated to be able to form transient but highly reactive tautomeric quinone methide warheads<sup>105</sup> and unsurprisingly can be potentially cytotoxic.<sup>104</sup> Additionally, they are known to interfere in bioassays through the formation of robust macromolecular aggregates.<sup>2</sup> Intriguingly, the disproportionately high count of 2 for the 2-hydroxyphenylhydrazones is dominated by hits for HTS Campaigns E and F, which use a Nickel-based anchor and which would strongly support a component of interference through chelation. This is a good example where a compound may appear to be a hit in a primary assay and may give a readout in a cell-based secondary assay due to entirely separate mechanisms, the first in this example being chelation (or potentially aggregation under certain conditions) and the second mechanism being reactivity. Together, however, the two results in any given academic HTS laboratory may be convincing and publishable. This is also a good example why the selection of primary hit sets from six of our early screens that used relatively high screening concentrations, with the inclusion of two that would be susceptible to chelation interference in addition to other possible mechanisms of interference (apart from aggregation), from a library of typically commercially available sources, allows for maximum capture of compound classes likely to be discovered similarly and reported by HTS academics.

**Table S14.** Alkylidene barbiturates as problematic compounds

Substructure <sup>a</sup>	Number of AlphaScreen® assays hit							Total Cpds	Enrich- ment <sup>b</sup>
	6	5	4	3	2	1	0		
 <b>ene_six_het_A</b>	10	20	21	30	69	105	228	483	66%

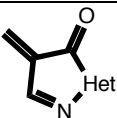
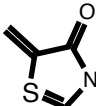
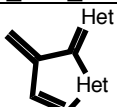
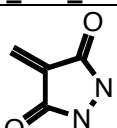
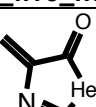
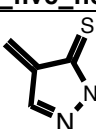
a. Name corresponds to name in Supplementary Table S6-S8 and Figure S1

b. Total number of compounds that hit from 2-6 assays expressed as a percentage of those compounds that have hit none of the six assays.

The substructure shown in Table S14 encodes for compounds that frequently occur as primary hits in our AlphaScreen® assays with an enrichment of 66% for a total of 483 compounds. These compounds are dominated by alkylidene barbiturates. These compounds are conspicuously prevalent as screening hits co-reported in studies already cited.<sup>23,26,29,41,42,51-55,74,100</sup> A brief survey of the literature reveals that they are also reported as inhibitors for Omi/HtrA2 protease,<sup>106</sup> potentiators of  $\Delta F508$ -CFTR chloride channel gating,<sup>107</sup> inhibitors of ERK,<sup>108</sup> dimethylarginine dimethylaminohydrolase,<sup>109</sup> fatty acid synthase,<sup>110</sup> PPM1D,<sup>111</sup>  $\alpha$ -glucosidase,<sup>112</sup>  $\alpha 4\beta 7$ -MAdCAM<sup>113</sup> and RNA Polymerase.<sup>114</sup> The structures involved are often very similar to each other and the SAR confusing. They are also Michael Acceptors and oxidants<sup>115-117</sup> and undergo light-induced irreversible covalent binding with proteins.<sup>39</sup> The thioxo counterparts in particular are prevalent in photoactive dyes that give rise to long-lived cytotoxic photoproducts.<sup>118</sup> We note that the exocyclic alkene renders these compounds distinct from marketed barbituric acid-based pharmaceuticals that have dialkyl substitution instead of the alkene and so would not be reactive.



**Table S15.** Alkylidenes of 5-membered heterocycles as problematic compounds

Substructure <sup>a</sup>	Number of AlphaScreen® assays hit							Total Cpds	Enrichment <sup>b</sup>
	6	5	4	3	2	1	0		
 <b>ene_five_het_A</b>	6	14	24	14	39	40	64	201	152%
 <b>ene_five_het_B</b>	0	4	4	2	14	22	44	90	55%
 <b>ene_five_het_C</b>	3	9	7	7	7	13	39	85	85%
 <b>ene_five_het_D</b>	4	7	8	9	13	5	0	46	na
 <b>ene_five_het_G</b>	0	0	2	1	1	1	5	10	80%
 <b>ene_five_het_H</b>	0	1	0	0	2	3	0	6	na

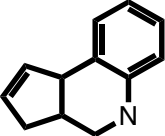
a. Name corresponds to name in Supplementary Table S6-S8 and Figure S1

b. Total number of compounds that hit from 2-6 assays expressed as a percentage of those compounds that have hit none of the six assays. na = not applicable (enrichment = infinity because of absence of any compound that hit no assays).

A number of 5-membered heterocycles bearing an exocyclic double bond interfere in our AlphaScreen® assays. In particular, the pyrazolidin-3,5-diones have a very high enrichment with no members that are clean. These sorts of compounds are prevalent in the screening literature already cited,<sup>21,23,28,40-42,45,50,51,66</sup> but are also reported in other screening literature as dual phosphodiesterase 1 and 5 inhibitors,<sup>119</sup> dual c-Src/Abl kinase inhibitors,<sup>120</sup> ubiquitin activating enzyme E1 inhibitors,<sup>121</sup> SARS-3CL<sup>pro</sup> inhibitors,<sup>122</sup> farnesoid X receptor inhibitors,<sup>123</sup> HIV-1 integrase inhibitors,<sup>124</sup> angiogenin<sup>125,126</sup> and 5-lipoxygenase.<sup>127</sup> Notably, these hits are dominated by pyrazolidin-3,5-diones. Similar to the other compound classes discussed above, these compounds tend to all look very similar and where reported, there is no conventional SAR apparent.

It is also somewhat disconcerting that the 5-alkenylimidazolin-4-one core in the **ene\_5\_het\_G** family of assay interference compounds corresponds precisely to that in the chromophore of green fluorescent protein.<sup>128</sup>

**Table S16.** Fused tetrahydroquinolines as problematic compounds

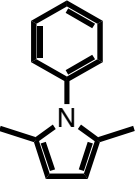
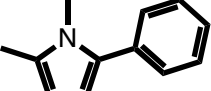
Substructure <sup>a</sup>	Number of AlphaScreen® assays hit							Total Cpd	Enrichment <sup>b</sup>
	6	5	4	3	2	1	0		
 <b>anil_alk_ene</b>	1	6	6	3	7	11	17	51	135%

a. Name corresponds to name in Supplementary Table S6-S8 and Figure S1

b. Total number of compounds that hit from 2-6 assays expressed as a percentage of those compounds that have hit none of the six assays.

Alkyl anilines feature strongly as problematic compounds in our assays. Amongst these, a distinctive class of tricyclic tetrahydroquinolines frequently appeared as primary hits in our AlphaScreen® assays and as shown, these totalled 51 in number and give rise to an enrichment value of 135%. These compounds also commonly occur in screening-based publications and are listed in works already cited.<sup>43,55,73</sup> They are also conspicuous in publications as inhibitors of anthrax edema factor,<sup>129</sup> low molecular weight protein tyrosine phosphatase,<sup>130</sup> MIF,<sup>131</sup> protein-tyrosine phosphatases,<sup>132</sup> Cdc25B dual specificity phosphatase<sup>133</sup> and polo-like kinase 1.<sup>134</sup> The structures involved are often very similar to each other and the SAR confusing. For example, weak but consistent inhibition against protein tyrosine phosphatase 1B is observed for the 6-F, 6-COOH and 6-OH analogs as well as the 8-H, 8-Br, 8-COOH and 8-Ac analogs.<sup>132</sup>

**Table S17.** Pyrroles as problematic compounds

Substructure <sup>a</sup>	Number of AlphaScreen® assays hit							Total Cpd	Enrichment <sup>b</sup>
	6	5	4	3	2	1	0		
 <b>pyrrole_A</b>	1	16	13	14	11	21	42	118	131%
 <b>pyrrole_B</b>	4	5	9	0	0	2	3	29	600%

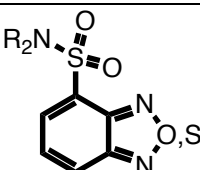
a. Name corresponds to name in Supplementary Table S6-S8 and Figure S1

b. Total number of compounds that hit from 2-6 assays expressed as a percentage of those compounds that have hit none of the six assays.

Certain pyrrole-containing compounds were prevalent as primary hits in our assays. A particularly troublesome group comprised N-aryl-2,5-dialkylpyrroles and N-alkyl-2-aryl-5-alkylpyrroles and as shown, these numbered some 118 and 29 respectively in our screening library with respective enrichment values of 131% and 600%. Very similar compounds are conspicuous in references already cited,<sup>21,47,55,73,114,126,130,132</sup> but have also been reported as inhibitors of HIV-1 fusion,<sup>135</sup> *M. tuberculosis* protein tyrosine phosphatase A inhibitors,<sup>136</sup> sphingosine 1-phosphate receptor agonists,<sup>137</sup> EphA4/EphA2 receptor inhibitors<sup>138</sup> and inhibitors of metabolic glutamate receptor 1.<sup>139</sup> An aryl carboxylate substituent frequently appears to be important for apparent

biological activity, for reasons that are unclear. As for the other PAINS described here, follow-up studies on close scrutiny reveal confusing SAR.<sup>140</sup>

**Table S18.** Benzofurazans as problematic compounds.

Substructure <sup>a</sup>	Number of AlphaScreen® assays hit							Total Cpds	Enrichment <sup>b</sup>
	6	5	4	3	2	1	0		
<div></div> <div><b>diazox sulfon A</b></div>	1	4	2	2	4	6	17	36	78%

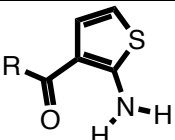
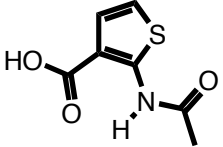
a. Name corresponds to name in Supplementary Table S6-S8 and Figure S1.

b. Total number of compounds that hit from 2-6 assays expressed as a percentage of those compounds that have hit none of the six assays.

Certain activated benzofurazans frequently occurred as hits in our primary screens. As shown, an exemplary class of benzofurazan sulfonamides, while not that common in our library and numbering 36, registered an enrichment of 78%. Of note, similar compounds are reported as hits in screening papers already cited<sup>33,93,98,130</sup> and are also reported to be inhibitors of HIV-1 reverse transcriptase inhibitors,<sup>141</sup> *M. tuberculosis* CYP51,<sup>142</sup> Yersinia type III secretion,<sup>143</sup> measles virus RNA-dependent RNA polymerase complex activity,<sup>144</sup> cholecystokinin-2 receptor antagonists<sup>145</sup> and thymidine monophosphate kinase.<sup>146</sup> SAR can be highly confusing.<sup>147</sup>

These compounds are highly unstable in vivo<sup>145</sup> and can be highly reactive sulfhydryl electrophiles,<sup>148</sup> used as fluoregenic and fluorescent labels,<sup>149</sup> and react rapidly with singlet oxygen.<sup>150</sup> Interestingly, though, this system can complex non-covalently with proteins<sup>142</sup> and has been the subject of a successful pharmacokinetic optimisation campaign, albeit one that required complete replacement of the benzofurazan with a phthalazine.<sup>145</sup> Abbott have recently provided supporting evidence that benzofurazans could be problematic protein-reactive screening compounds.<sup>6,7</sup>

**Table S19.** 2-Amino-3-carbonylthiophenes as problematic compounds

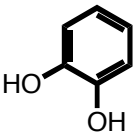
Substructure <sup>a</sup>	Number of AlphaScreen® assays hit							Total Cpds	Enrichment <sup>b</sup>
	6	5	4	3	2	1	0		
 <b>thiophene_amino_Aa</b>	2	2	5	4	3	11	18	45	94%
 <b>thiophene_amino_Ab</b>	0	2	2	1	5	7	23	40	43%

a. Name corresponds to name in Supplementary Table S6-S8 and Figure S1

b. Total number of compounds that hit from 2-6 assays expressed as a percentage of those compounds that have hit none of the six assays.

These compounds are less common in our screening library but as shown, register frequently as hits with an enrichment value of up to 94%. This class is conspicuous as hits in references already cited,<sup>27,74,78,96,107,128,134</sup> but has already been reported as inhibitors of tubulin,<sup>151</sup> reverse transcriptase-associated ribonuclease H,<sup>152</sup> HCV RNA polymerase<sup>153</sup> and FLT3 tyrosine kinase.<sup>154</sup> Notably, in two studies, this class appeared to comprise promiscuous, non-specific hits when investigated in more detail.<sup>155,156</sup> We note that these compounds can form intensely blue solutions (Luigi Aurelio, personal communication). Abbott have recently provided supporting evidence that 2-amino-3-carbonylthiophenes could be problematic protein-reactive screening compounds that cause protein thiol oxidation.<sup>157</sup>

**Table S20.** Catechols as problematic screening compounds.

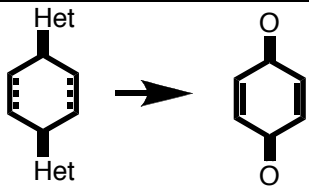
Substructure <sup>a</sup>	Number of AlphaScreen® assays hit							Total Cpd	Enrichment <sup>b</sup>
	6	5	4	3	2	1	0		
 <b>catechol A</b>	4	7	10	4	10	21	36	92	97%

a. Name corresponds to name in Supplementary Table S6-S8 and Figure S1

b. Total number of compounds that hit from 2-6 assays expressed as a percentage of those compounds that have hit none of the six assays.

Catechols are numerous (92) and problematic (enrichment 97%) in our screening campaigns. Catechols and hydroquinones are noticeable in screening literature already cited.<sup>2,25,75,78,94,125,131</sup> This activity may be due to oxidation to protein-reactive quinones.<sup>158</sup> Catechol-like templates are present in redox-reactive dyes such as pyocyanin.<sup>159</sup> Abbott have recently provided supporting evidence that catechols could be problematic protein-reactive screening compounds.<sup>6,7,157</sup>

**Table S21.** Quinones as problematic screening compounds

Substructure <sup>a</sup>	Number of AlphaScreen® assays hit							Total Cpd	Enrichment <sup>b</sup>
	6	5	4	3	2	1	0		
 <b>quinone_A</b>	40	57	48	41	42	56	86	370	265%

a. Name corresponds to name in Supplementary Table S6-S8 and Figure S1

b. Total number of compounds that hit from 2-6 assays expressed as a percentage of those compounds that have hit none of the six assays.

Quinones and quinone-like compounds are amongst the most problematic of the compounds we have in our library, being both numerous - numbering some 370, and giving rise to a very high enrichment factor of 265%. Quinones are conspicuous in screening literature already cited<sup>54,73,107,111,130,133</sup> and are known to be protein-reactive.<sup>160,161</sup> Abbott have recently provided supporting evidence that quinones could be problematic protein-reactive screening compounds.<sup>157</sup> We also find that many quinones in our library are substituted with electron donating groups, giving rise to strongly colored compounds.

**Table S22.** Azo compounds as problematic compounds.

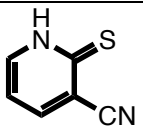
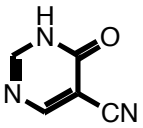
Substructure <sup>a</sup>	Number of AlphaScreen® assays hit							Total Cpds	Enrichment <sup>b</sup>
	6	5	4	3	2	1	0		
$R'-N=N-R''$ <b>azo_A</b>	29	30	33	43	24	55	110	324	145%

a. Name corresponds to name in Supplementary Table S6-S8 and Figure S1

b. Total number of compounds that hit from 2-6 assays expressed as a percentage of those compounds that have hit none of the six assays.

There are significant numbers of azo compounds in our screening library, numbering some 324 and these lead to a very high enrichment value of 145%. Of note, these compounds are conspicuous in references already cited,<sup>125,126,134</sup> but are also reported as screening hits that inhibit Alzheimer's-related fibril formation,<sup>162-164</sup> avian influenza neuraminidase<sup>165</sup> and Bcl-2.<sup>166</sup> These compounds are often strongly colored and in fact are often dyes in reported assay studies<sup>162-164</sup> and can also undergo photoisomerization and reaction with biological nucleophiles.<sup>167</sup> They have also been reported to give rise to false positive results mediated via aggregation phenomena<sup>2</sup> and can interfere with AlphaScreen® technology.<sup>168</sup>

**Table S23.** Cyanopyridones as problematic compounds.

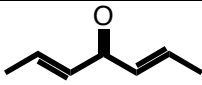
Substructure <sup>a</sup>	Number of AlphaScreen® assays hit							Total Cpds	Enrichment <sup>b</sup>
	6	5	4	3	2	1	0		
 <b>cyano_pyridone_A</b>	1	3	3	4	6	16	23	54	65%
 <b>cyano_pyridone_B</b>	0	1	2	2	7	4	11	27	109%

a. Name corresponds to name in Supplementary Table S6-S8 and Figure S1

b. Total number of compounds that hit from 2-6 assays expressed as a percentage of those compounds that have hit none of the six assays.

There are significant numbers of cyanopyridones, the most common numbering 54 and 27 with respective enrichment values of 65% and 109% respectively. Similar compounds are co-reported in references already cited.<sup>21,64,108,119</sup> We note that such compounds can be effective thiol-reactive electrophiles.<sup>169</sup>

**Table S24.** Divinylketones as problematic screening compounds

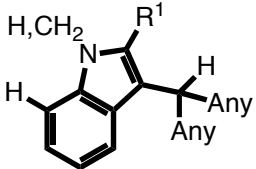
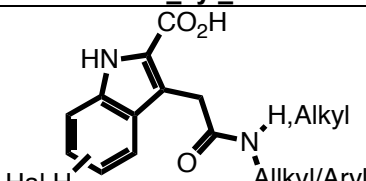
Substructure <sup>a</sup>	Number of AlphaScreen® assays hit							Total Cpds	Enrichment <sup>b</sup>
	6	5	4	3	2	1	0		
 <b>ene_one_ene_A</b>	1	2	3	5	6	16	24	57	71%

a. Name corresponds to name in Supplementary Table S6-S8 and Figure S1

b. Total number of compounds that hit from 2-6 assays expressed as a percentage of those compounds that have hit none of the six assays.

Divinylketones are problematic screening compounds, numbering 57 in our Library and registering an enrichment value of 71%. We note that these compounds are reported in screening literature already cited.<sup>74,111,124</sup> Abbott have recently provided supporting evidence that these could be problematic protein-reactive screening compounds.<sup>6</sup> Of relevance here is work from the Cravatt group that suggests vinyl ketones readily and selectively react with protein cysteines, whereas tosylate esters readily react with a variety of residue types, including Asp, Glu and Tyr (Weerapana *et al.*, *Nat. Chem. Biol.* **2008**, 4, 405-407).

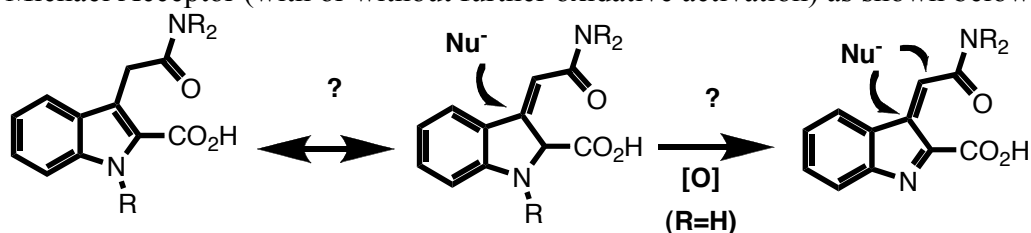
**Table S25.** Certain indoles as problematic screening compounds.

Substructure <sup>a</sup>	Number of AlphaScreen® assays hit							Total Cpds	Enrichment <sup>b</sup>
	6	5	4	3	2	1	0		
 <p><b>R<sup>1</sup> = CH<sub>2</sub>, C=Het, C:Het, CHN, CH(CH<sub>2</sub>)CH<sub>2</sub>NCH<sub>2</sub> indol_3yl_alk</b></p>	6	15	23	18	51	80	268	461	42%
	0	0	2	3	7	4	14	30	86%

a. Name corresponds to name in Supplementary Table S6-S8 and Figure S1

b. Total number of compounds that hit from 2-6 assays expressed as a percentage of those compounds that have hit none of the six assays.

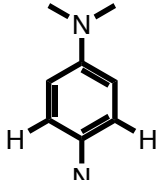

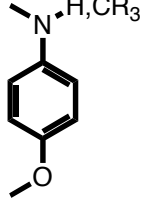
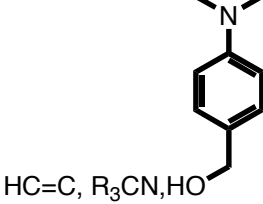
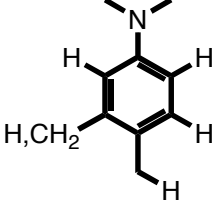
One of our largest classes of problematic compounds comprises 3-alkylindoles, numbering some 461 compounds although as shown the enrichment value of 42% is not so high. We have not refined this class further, but it seems likely that several subclasses are involved. Some of these may specifically interfere in the AlphaScreen® technology, but it is likely that other subclasses may be PAINS. For example, we have noted that a particular subclass, indole-3-acetamide-2-carboxylic acids, number thirty within this broad class and as shown, have a high assay interference enrichment of 86%. It can be envisaged that the indoline tautomer of this class, with the enecarbonyl projecting from the indoline 3-position, could act as a Michael Acceptor (with or without further oxidative activation) as shown below.



We have noticed these compounds in the screening literature reported as catalysts of peptide exchange,<sup>170</sup> mitogen-activated protein kinase phosphate-1 dual-specificity protein phosphatase inhibitors,<sup>171</sup> *Trypanosoma cruzi* trans-sialidase inhibitors,<sup>172</sup> and nitroglutathione reductase inhibitors.<sup>173</sup> As before, other PAIN structures already discussed are conspicuous in the latter three of these reports. We note that the broader class of indole-3-acetamides allowing any substituent off the 2 position (including H) yields 138 compounds with a relatively high enrichment 26%. This broader class is thus close to meeting our definition of being problematic in its own right, but like hydrazones, it is particular substituents that are likely to result in sub-classes that are particularly problematic. We believe there is the opportunity for many of our substructure filters to be thus refined further.



**Table S26.** Tertiary anilines as putative AlphaScreen®-specific problematic screening compounds.

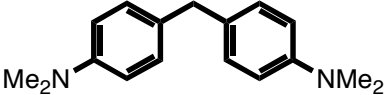
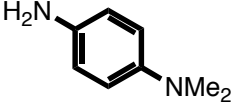
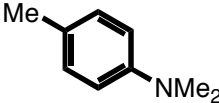
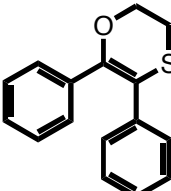
Substructure <sup>a</sup>	Number of AlphaScreen® assays hit							Total Cpd	Enrichment <sup>b</sup>
	6	5	4	3	2	1	0		
 <b>anil_di_alk_A</b>	31	66	27	35	28	59	232	478	81%
 <b>anil_di_alk_B</b>	6	18	7	18	22	43	137	251	52%
 <b>anil_di_alk_C</b>	15	23	13	10	26	29	130	246	67%
 <b>anil_di_alk_D</b>	12	10	18	11	24	28	95	198	79%
 <b>anil_di_alk_E</b>	8	12	17	10	15	27	97	186	64%

a. Name corresponds to name in Supplementary Table S6-S8 and Figure S1

b. Total number of compounds that hit from 2-6 assays expressed as a percentage of those compounds that have hit none of the six assays.

These compounds are less conspicuous in the screening literature that we have thus far cited. We propose that many of these compounds could interfere specifically with AlphaScreen®-technology, possibly through efficient quenching of singlet oxygen.<sup>175</sup> We note the involvement of such compounds in singlet oxygen reaction in the AlphaScreen® beads themselves (Figure 3). We felt it important to investigate this further and so deliberately sourced and tested individual compounds related to the AlphaScreen® bead chemistry. The results are shown in Table S27.

**Table S27.** Activity of certain anilines and olefins in the AlphaScreen® assay

Entry	Structure	Color of Solution	IC <sub>50</sub> (n=6)
1		Colorless	2.9 ± 0.2 µM
2		Rose	30 ± 3 µM
3		Colorless	>>100 µM
4		Colorless	>>100 µM

Here it is apparent that tertiary anilines do not interfere in AlphaScreen® assay technology *per se* as the 4-methyl analogue is inactive (entry 3). However, when suitably functionalised, they can do so and bis-(*N,N*-dimethylaminophenyl)methane, entry 1, consistently gives rise to an IC<sub>50</sub> of around 3µM. It is plausible that this is due to quenching singlet oxygen, which would make it around 30 times as effective in this regard as DABCO (Table 9) under the same conditions. However, in none of the classes of tertiary anilines in Table S26 is it exclusive that all six assays are hit, which would be expected if singlet oxygen quenching was the only mechanism involved. Thus it is plausible that other unknown mechanisms are also involved.

## Drugs containing PAIN substructures - extended discussion

Epalrestat, which is marketed in Japan as an aldose reductase inhibitor for treating diabetic neuropathy (and more recently launched in India), contains a rhodanine-based PAIN structure but does not appear to have poor pharmacokinetics. This compound contains an extended arylene group with a  $\gamma$ -methyl substituent. It may be that our filters have the potential to be refined further to identify in more detail specific components that give rise to reactivity and discriminate from those that do not within a given class. Similarly, troglitazone, rosiglitazone and pioglitazone are rhodanine-like thiazolidinedione-based orally active antidiabetic agents for the treatment of type 2 diabetes. However, in these cases, there is serious concern associated with metabolically-activated toxic by-products, some of which may derive from the thiazolidinedione core.<sup>190</sup> Similarly, we have found several marketed drugs to contain PAIN substructures (in the following discussion the compound numbers refer to those listed in Supplementary Figure S2). This is because they contain catechols or hydroquinones (levodopa (25), adrenaline (30), apomorphine (31), dopamine (32), isoprenaline (33), noradrenaline (34), dobutamine (36), carbidopa (37), methyldopa (38), rifampicin (6), benserazide (18)) or quinones (menadione (1), phytomenadione (2), epirubicin (5), mitoxantrone (8) (mitoxantrone), atavaquone (11), mitomycin (22), daunorubicin (23), idarubicin, (24)) or a masked quinoid (amsacrine (35)). These moieties are strongly associated with in vivo toxicity for reasons that also likely relate to their assay interference properties discussed herein.<sup>190</sup> In a significant study in 2002,<sup>191</sup> Hofmann La-Roche undertook a related exercise employing automated techniques as opposed to ours that took the more traditional route, relying more on the ability of an experience medicinal chemist to recognize and classify molecular structures. Despite the different approach, several catechol and quinone-based drugs were also identified as frequent hitters by the Roche group. These included dopamine, methyldopamine (methyldopa), benserazide, apomorphine and idebenone and so there is considerable overlap with those drugs similarly identified by us. The quinone-like clofazimine is also identified in both studies as a potentially problematic screening hit. Abbott has also identified the quinone-based cytotoxic doxorubicin as an ALARM-NMR-reactive drug.<sup>7</sup> In a number of cases, therefore, there appears to be a direct link between poorly tolerated marketed drugs and a PAIN-containing structure, where protein-reactivity, either directly or indirectly after metabolic activation, may be a common underlying mechanism. The work by Abbott supports this notion.<sup>7</sup>

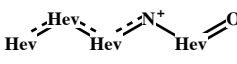
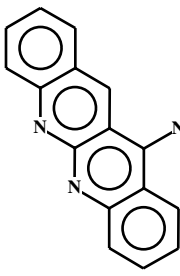
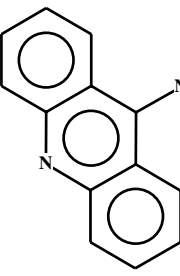
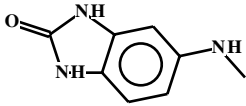
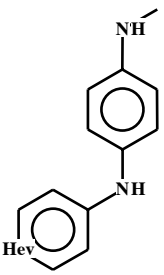
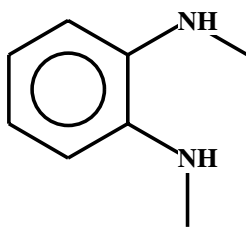
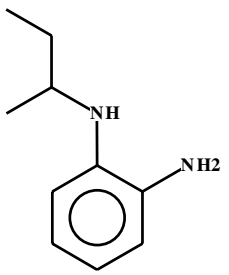
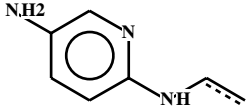
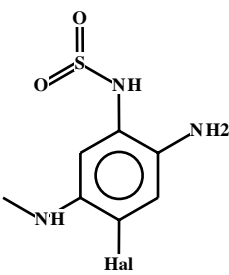
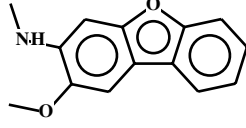
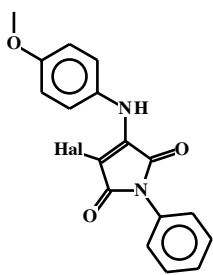
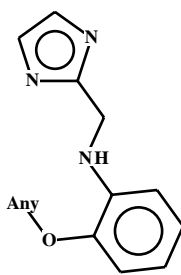
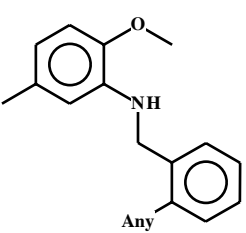
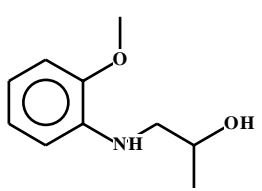
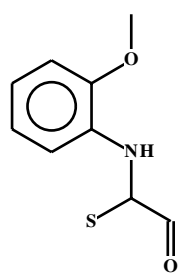
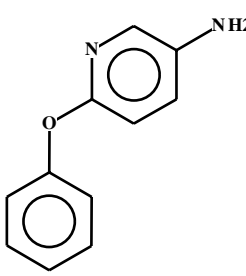
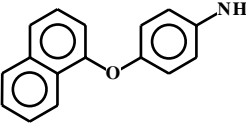
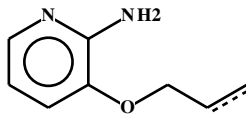
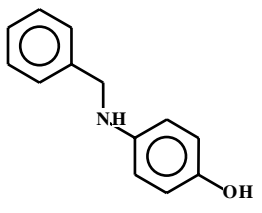
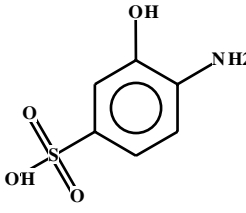
There are a number of other drugs in Supplementary Figure S2 that are also recognised by our assay interference filters but which we have not specifically categorized herein as PAINS because there is not the literature evidence to do so. Even so, the substructure involved in many of these also appears to be that which in the drug is also associated with its major metabolic liabilities. For example, the ethylthiomethyleneimidazole in cimetidine (41) is recognized as an assay interference moiety by our filters and it this group in cimetidine which is a major metabolic liability (oxidation of the sulfur atom and engagement and inhibitions of CYPs by the imidazole ring).<sup>192,193</sup> Similarly, olsalazine (13), balsalazide (14) and sulfasalazine (7) are recognized by our filters because they contain an azo group. This group is unstable in vivo and effectively acts as a prodrug by being metabolised by intestinal

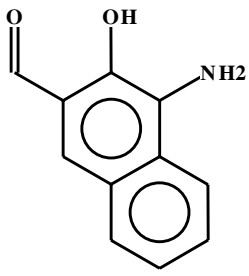
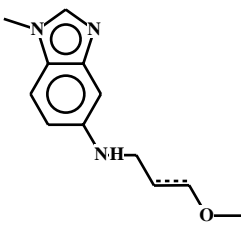
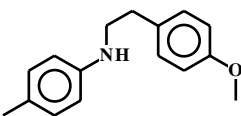
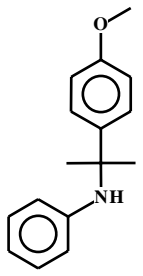
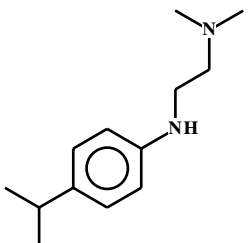
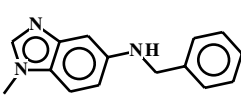
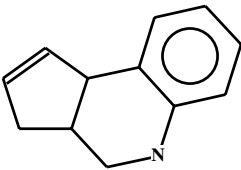
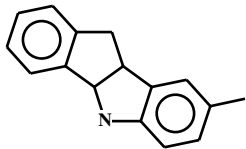
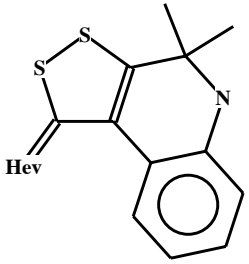
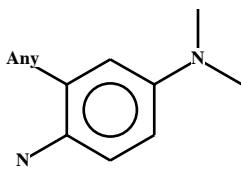
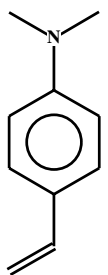
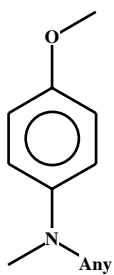
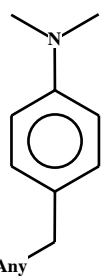
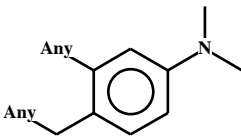
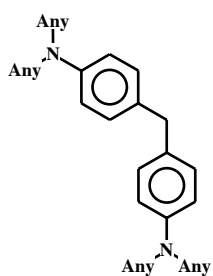
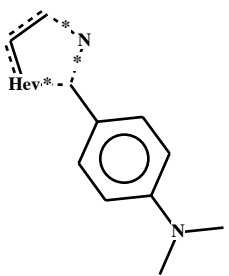
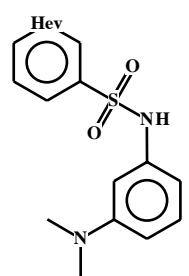
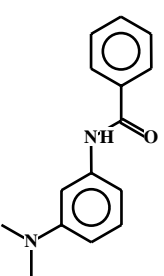
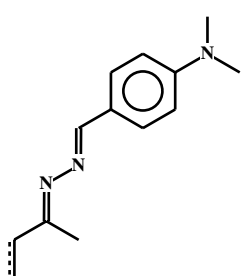
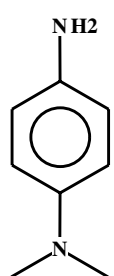
bacteria.<sup>194</sup> An azo group is recognised in the triazene of dacarbazine (20) and the azide in zidovudine (12, AZT) and in both cases is associated with instability or reactivity, though in the case of dacarbazine this reactivity also represents its principle mechanism of antitumor action (DNA methylation).<sup>195,196</sup> Tricyclics such as promethazine (43) and trimeprazine (44) are recognized by our assay interference filters because they contain a phenothiazine group, and it is the sulfur atom of this which is also one of the major metabolic liabilities.<sup>197</sup> The phenothiazine core has also been linked to phototoxicity.<sup>198</sup> Phenindione (26) is an old and dirty drug<sup>199</sup> and is recognized as a potential assay interference compound primarily by virtue of its 1,3-diketone group. Topotecan (9) is an antitumour drug that is recognized by our filters by virtue of the phenolic mannich base, which has been associated with aqueous instability with the formation of a reactive quinone methide.<sup>200</sup> Similarly, pyrroles related to those that we find are assay interference compounds can be DNA-reactive<sup>201</sup> and that in Atorvastatin has been linked to phototoxicity.<sup>202</sup>

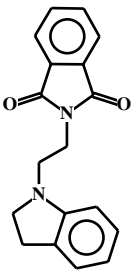
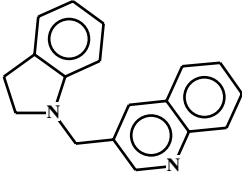
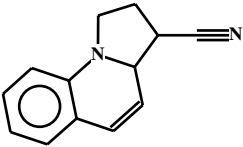
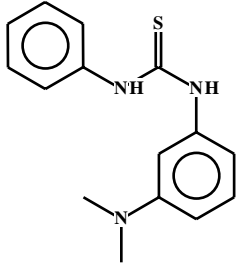
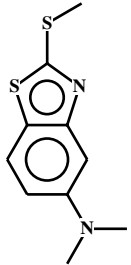
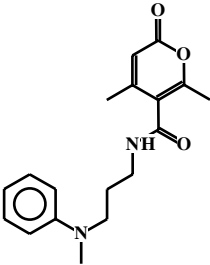
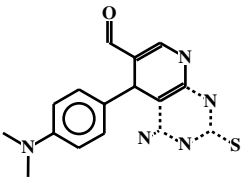
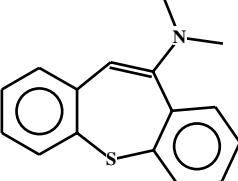
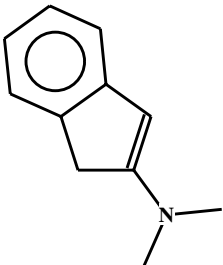
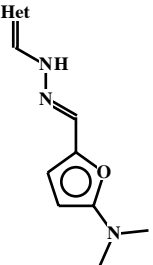
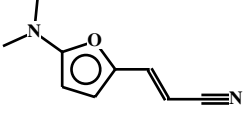
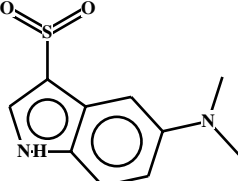
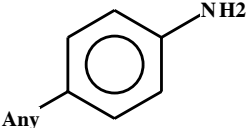
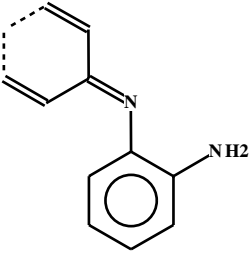
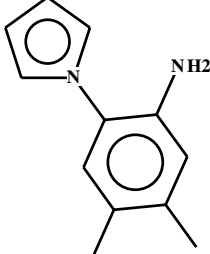
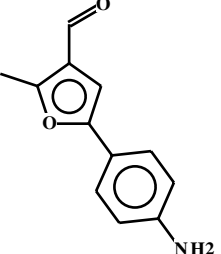
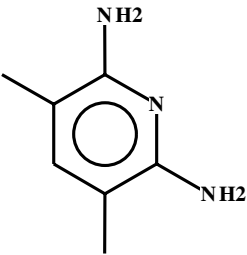
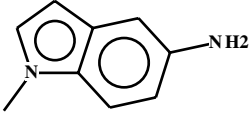
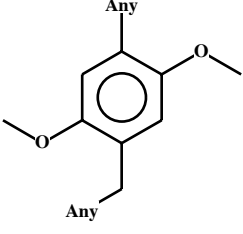
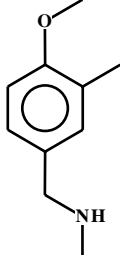
However, there are some drugs whose substructures recognised by our filters do not appear to impart any metabolic liability. These include the antifungals itraconazole (3), posaconazole (4), ketoconazole (17) and the antibiotic linezolid (16), all of which are recognized by our filters because of the presence of a tertiary aniline but this does not appear to introduce a metabolic liability in these particular cases, though we note that the electron-rich aromatic rings tend to be disfavoured by medicinal chemists as starting points because they are prone to metabolic oxidation. A similar argument applies for the tricyclics amitriptyline (40), nortriptyline (42), flupenthixol (45), zuclopenthixol (46) and thiothixene (47) which are also filtered out and so also appear in Supplementary Figure S2 because they contain a particular type of styrene functionality but which appears to be metabolically stable; though we note here too that the styrene core has been linked to the possibility of metabolically-induced toxicity.<sup>203</sup> Thiothixene, amitriptyline and nortriptyline also comprise three of the four CNS-active drugs that were filtered out in Table 7. The fourth CNS-active drug filtered out was frovatriptan, due to its indole-3-alkyl-containing substructure, but the pharmacokinetics of this compound appear to be acceptable and the indole nucleus not a metabolic liability.<sup>204</sup>

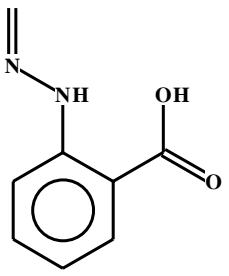
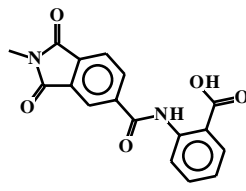
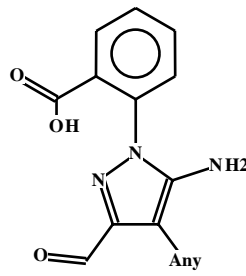
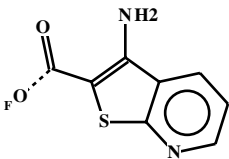
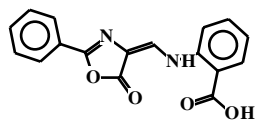
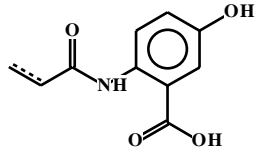
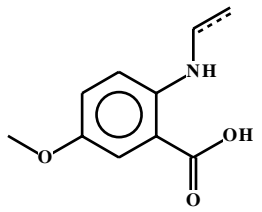
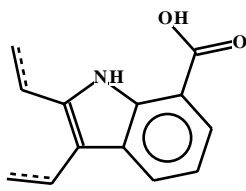
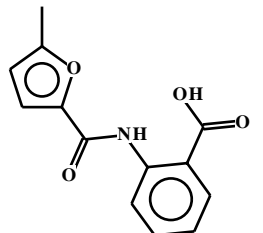
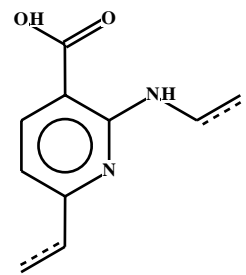
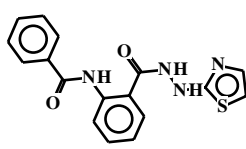
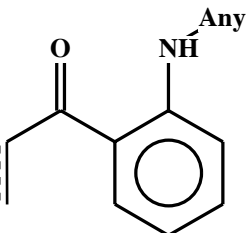
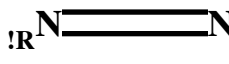
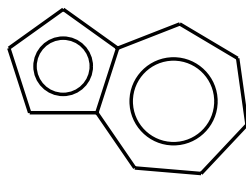
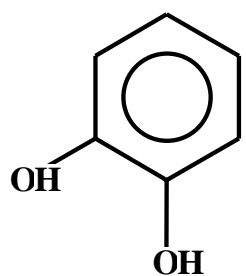
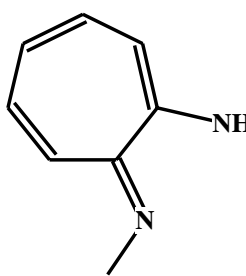
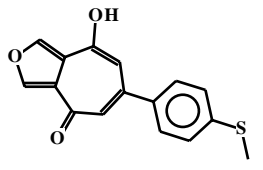
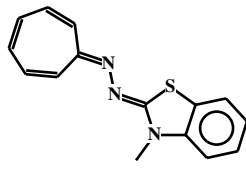
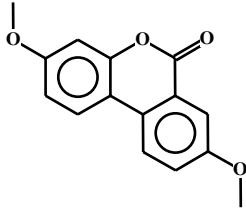
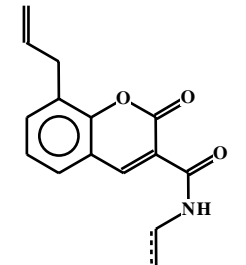
**FIGURE S1: Assay Interference structures broadly grouped according to type**

FREQ\_HIT\_5\_ALL\_SORTED

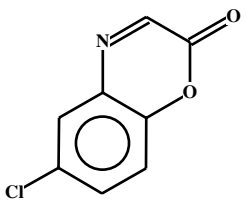
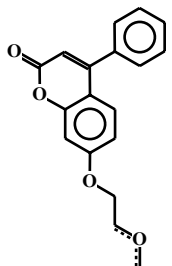
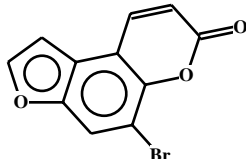
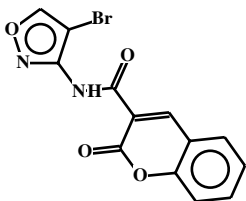
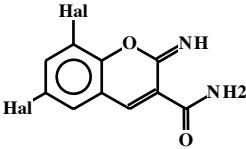
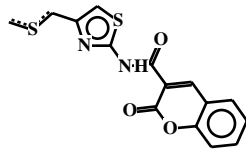
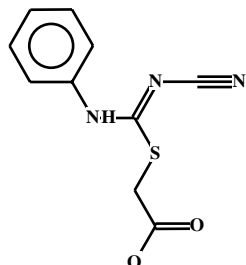
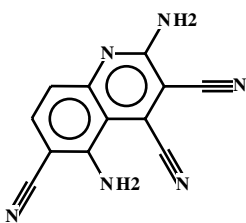
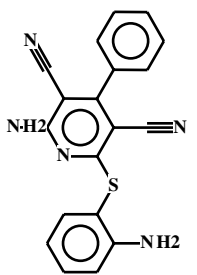
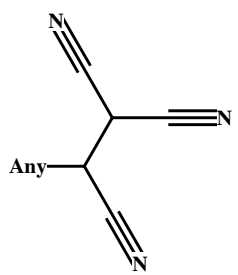
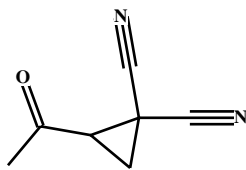
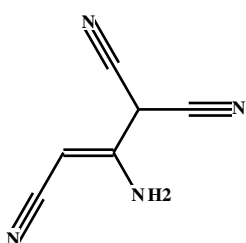
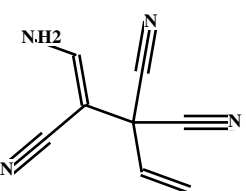
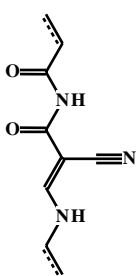
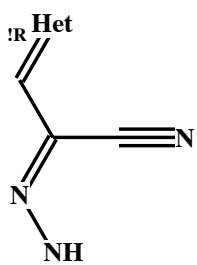
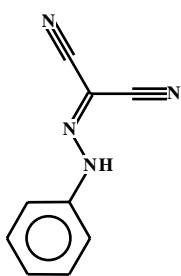
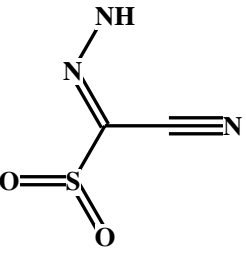
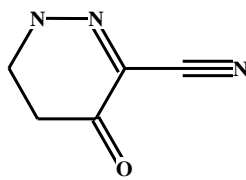
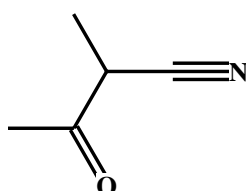
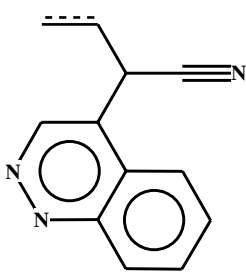
 <p><b>1:acyl_het_A(9)</b></p>	 <p><b>2:amino_acridine_A(1)</b></p>	 <p><b>3:amino_acridine_A(46)</b></p>	 <p><b>4:anil_NH_alk_A(5)</b></p>
 <p><b>5:anil_NH_alk_B(3)</b></p>	 <p><b>6:anil_NH_alk_C(2)</b></p>	 <p><b>7:anil_NH_alk_D(2)</b></p>	 <p><b>8:anil_NH_no_alk_A(1)</b></p>
 <p><b>9:anil_NH_no_alk_B(1)</b></p>	 <p><b>10:anil_OC_alk_A(4)</b></p>	 <p><b>11:anil_OC_alk_B(3)</b></p>	 <p><b>12:anil_OC_alk_C(3)</b></p>
 <p><b>13:anil_OC_alk_D(2)</b></p>	 <p><b>14:anil_OC_alk_E(1)</b></p>	 <p><b>15:anil_OC_alk_F(1)</b></p>	 <p><b>16:anil_OC_no_alk_A(8)</b></p>
 <p><b>17:anil_OC_no_alk_B(4)</b></p>	 <p><b>18:anil_OC_no_alk_C(3)</b></p>	 <p><b>19:anil_OH_alk_A(8)</b></p>	 <p><b>20:anil_OH_no_alk_A(1)</b></p>

			
<b>21:anil_OH_no_alk_B(1)</b>	<b>22:anil_alk_A(1)</b>	<b>23:anil_alk_B(1)</b>	<b>24:anil_alk_C(1)</b>
			
<b>25:anil_alk_D(1)</b>	<b>26:anil_alk_bim(9)</b>	<b>27:anil_alk_ene(51)</b>	<b>28:anil_alk_indane(1)</b>
			
<b>29:anil_alk_thio(4)</b>	<b>30:anil_di_alk_A(478)</b>	<b>31:anil_di_alk_B(251)</b>	<b>32:anil_di_alk_C(246)</b>
			
<b>33:anil_di_alk_D(198)</b>	<b>34:anil_di_alk_E(186)</b>	<b>35:anil_di_alk_F(14)</b>	<b>36:anil_di_alk_G(9)</b>
			
<b>37:anil_di_alk_H(6)</b>	<b>38:anil_di_alk_I(4)</b>	<b>39:anil_di_alk_J(3)</b>	<b>40:anil_di_alk_K(2)</b>

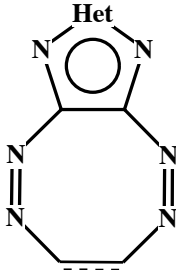
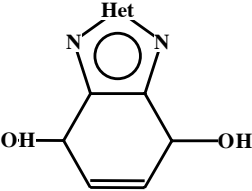
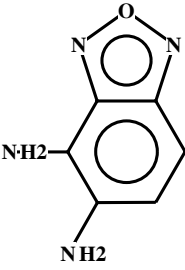
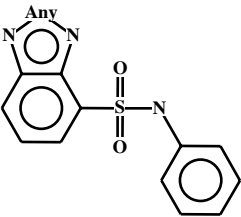
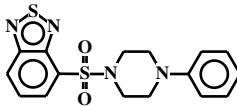
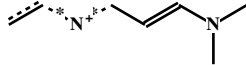
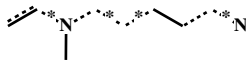
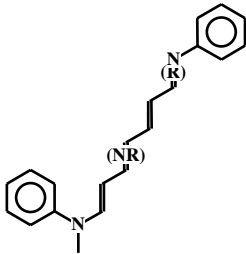
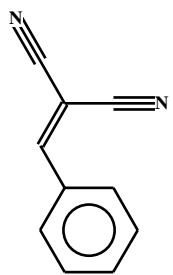
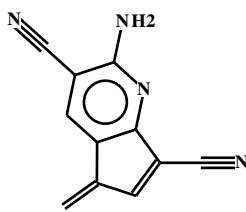
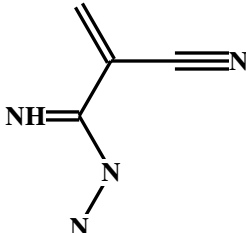
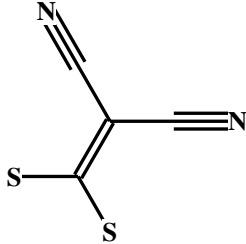
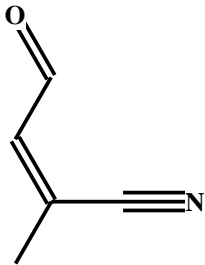
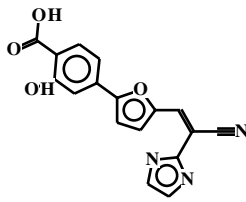
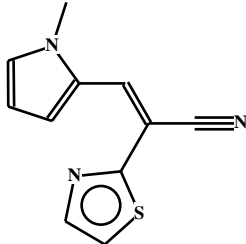
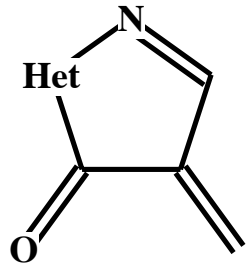
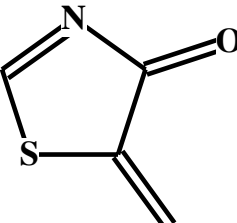
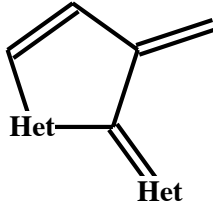
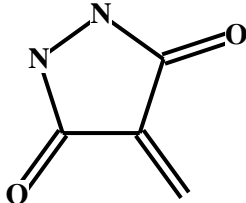
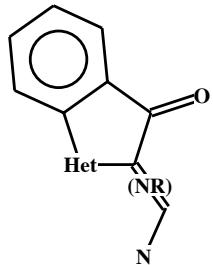
			
<b>41:anil_di_alk_L(1)</b>	<b>42:anil_di_alk_M(1)</b>	<b>43:anil_di_alk_N(1)</b>	<b>44:anil_di_alk_O(1)</b>
			
<b>45:anil_di_alk_P(1)</b>	<b>46:anil_di_alk_coum(1)</b>	<b>47:anil_di_alk_dhp(1)</b>	<b>48:anil_di_alk_ene_A(8)</b>
			
<b>49:anil_di_alk_ene_B(4)</b>	<b>50:anil_di_alk_furan_A(1~)</b>	<b>51:anil_di_alk_furan_B(2)</b>	<b>52:anil_di_alk_indol(1)</b>
			
<b>53:anil_no_alk(40)</b>	<b>54:anil_no_alk_A(1)</b>	<b>55:anil_no_alk_B(1)</b>	<b>56:anil_no_alk_C(1)</b>
			
<b>57:anil_no_alk_D(1)</b>	<b>58:anil_no_alk_indol_A(1)</b>	<b>59:anisol_A(5)</b>	<b>60:anisol_B(2)</b>

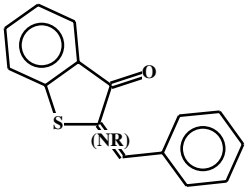
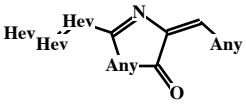
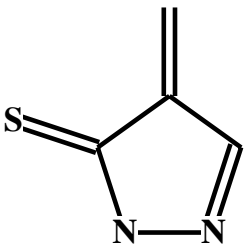
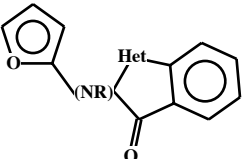
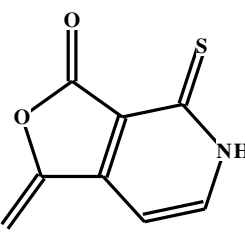
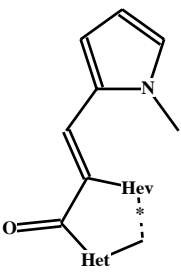
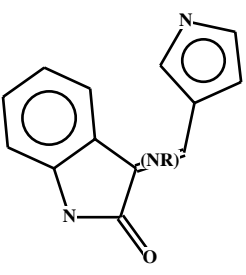
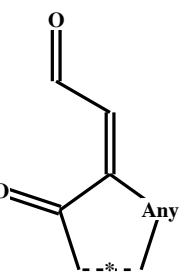
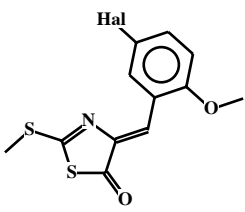
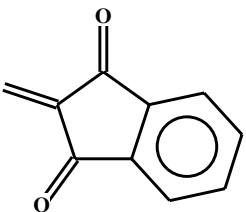
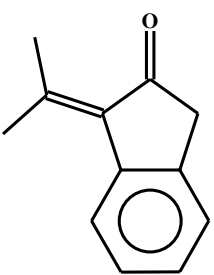
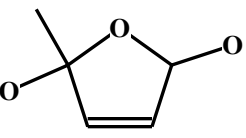
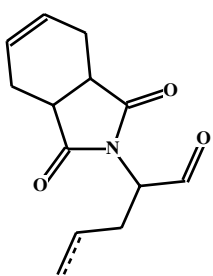
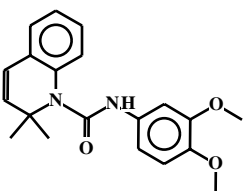
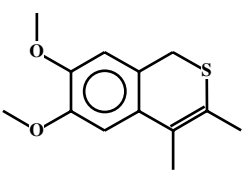
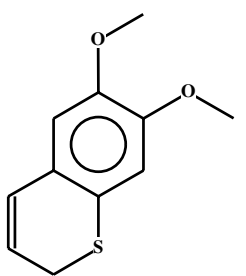
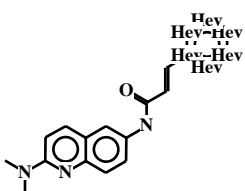
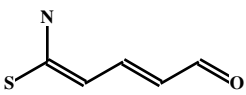
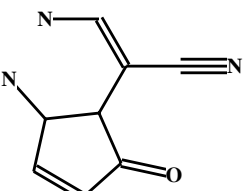
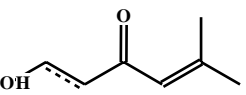
			
61:anthranil_acid_A(19)	62:anthranil_acid_B(3)	63:anthranil_acid_C(2)	64:anthranil_acid_D(2)
			
65:anthranil_acid_E(2)	66:anthranil_acid_F(2)	67:anthranil_acid_G(1)	68:anthranil_acid_H(1)
			
69:anthranil_acid_I(1)	70:anthranil_acid_J(1)	71:anthranil_acid_K(1)	72:anthranil_acid_L(38)
			
73:azo_A(324)	74:azulene(4)	75:catechol_A(92)	76:colchicine_A(3)
			
77:colchicine_B(1)	78:colchicine_het(1)	79:coumarin_A(2)	80:coumarin_B(2)



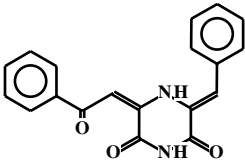
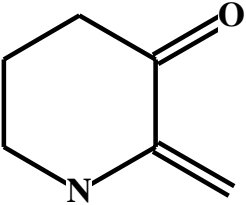
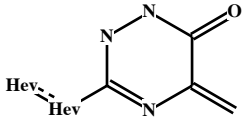
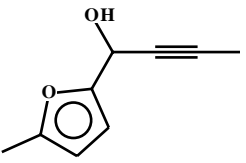
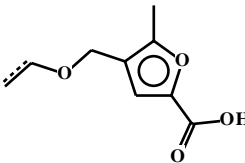
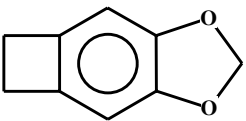
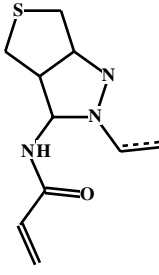
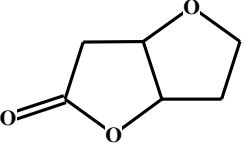
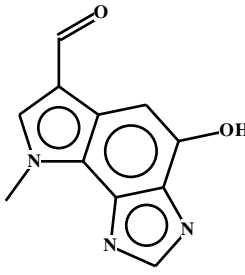
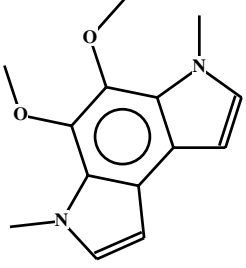
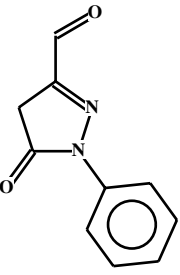
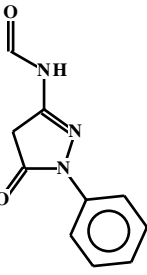
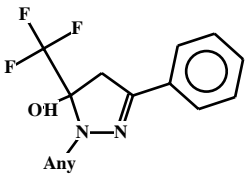
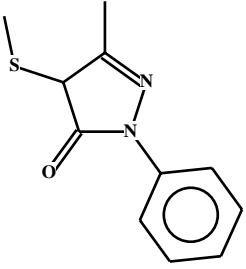
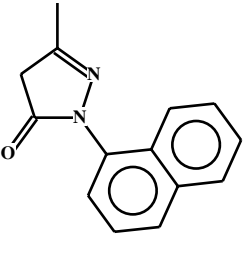
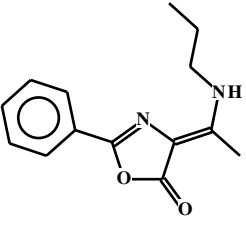
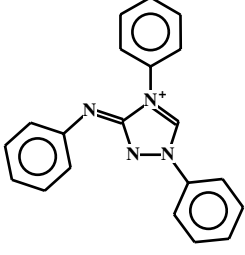
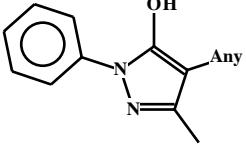
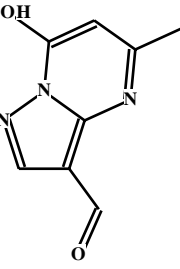
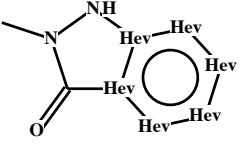
 <b>81:coumarin_C(1)</b>	 <b>82:coumarin_D(1)</b>	 <b>83:coumarin_E(1)</b>	 <b>84:coumarin_F(1)</b>
 <b>85:coumarin_G(1)</b>	 <b>86:coumarin_H(1)</b>	 <b>87:cyanamide_A(1)</b>	 <b>88:cyano_amino_het_A(1)</b>
 <b>89:cyano_amino_het_B(1)</b>	 <b>90:cyano_cyano_A(23)</b>	 <b>91:cyano_cyano_B(3)</b>	 <b>92:cyano_ene_amine_A(56)</b>
 <b>93:cyano_ene_amine_B(4)</b>	 <b>94:cyano_ene_amine_C(3)</b>	 <b>95:cyano_imine_A(37)</b>	 <b>96:cyano_imine_B(17)</b>
 <b>97:cyano_imine_C(12)</b>	 <b>98:cyano_imine_D(1)</b>	 <b>99:cyano_keto_A(2)</b>	 <b>100:cyano_misc_A(1)</b>

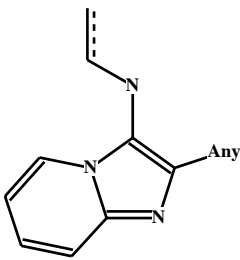
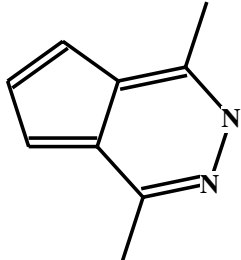
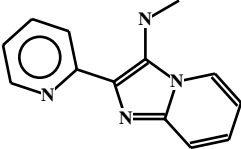
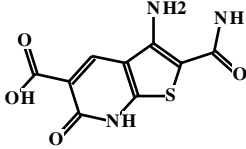
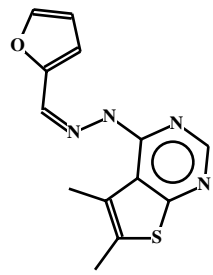
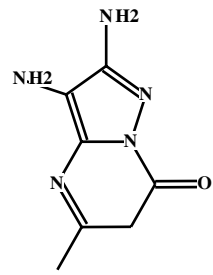
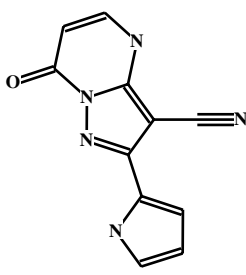
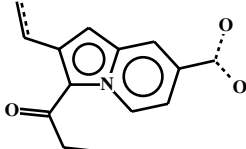
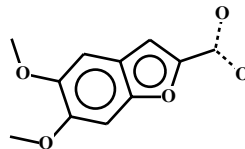
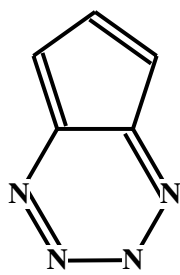
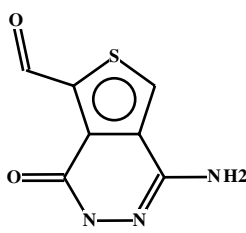
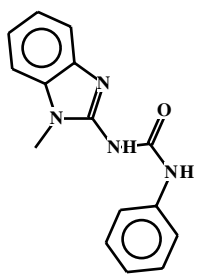
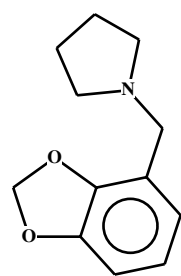
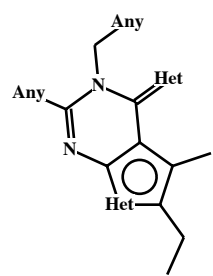
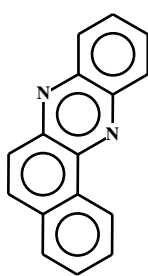
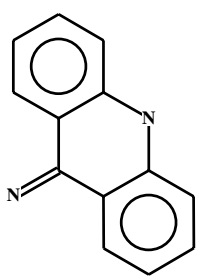
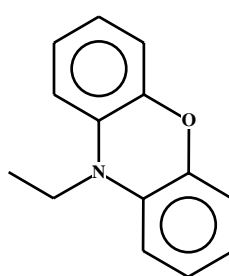
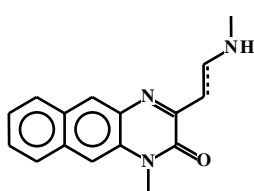
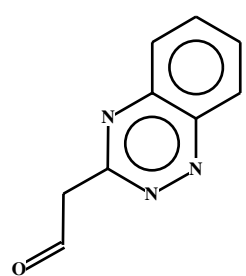
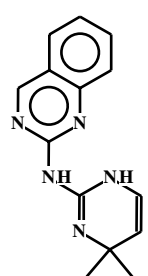
101:cyano_pyridone_A(54)	102:cyano_pyridone_B(27)	103:cyano_pyridone_C(11)	104:cyano_pyridone_D(5)
105:cyano_pyridone_E(4)	106:cyano_pyridone_F(3)	107:cyano_pyridone_G(1)	108:dhp_amidine_A(1)
109:dhp_amino_CN_A(13)	110:dhp_amino_CN_B(9)	111:dhp_amino_CN_C(7)	112:dhp_amino_CN_D(5)
113:dhp_amino_CN_E(4)	114:dhp_amino_CN_F(3)	115:dhp_amino_CN_G(1)	116:dhp_amino_CN_H(1)
117:dhp_bis_amino_CN(19)	118:dhp_keto_A(9)	119:diazox_A(3)	120:diazox_B(3)

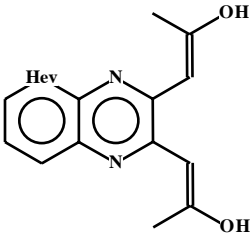
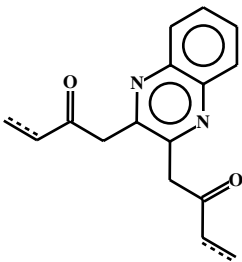
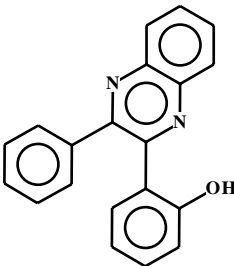
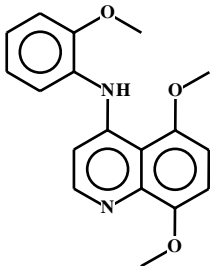
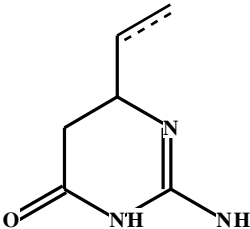
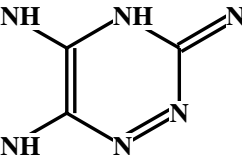
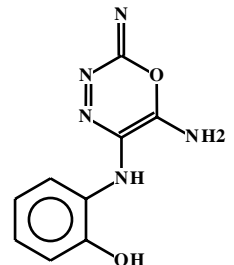
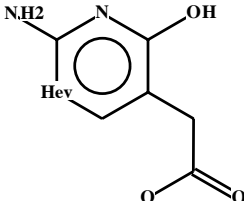
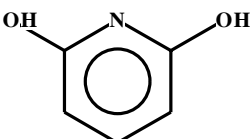
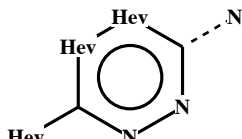
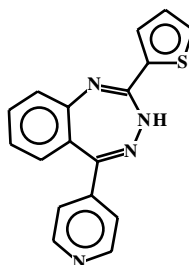
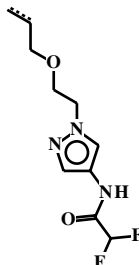
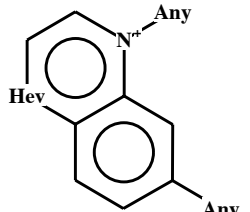
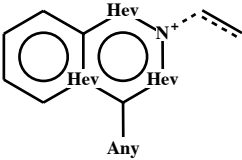
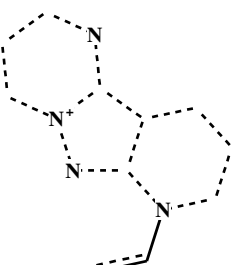
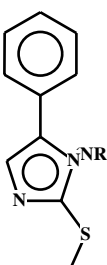
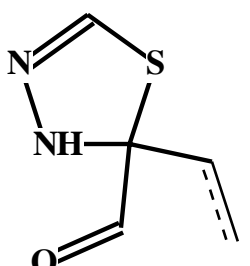
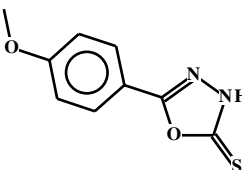
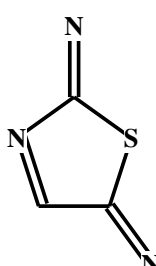
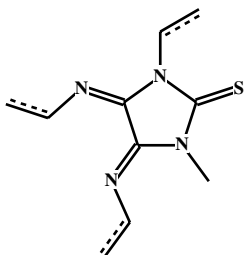
 <p><b>121:diazox_C(1)</b></p>	 <p><b>122:diazox_D(1)</b></p>	 <p><b>123:diazox_E(1)</b></p>	 <p><b>124:diazox_sulfon_A(36)</b></p>
 <p><b>125:diazox_sulfon_B(5)</b></p>	 <p><b>126:dyes3A(19)</b></p>	 <p><b>127:dyes5A(27)</b></p>	 <p><b>128:dyes7A(2)</b></p>
 <p><b>129:ene_cyano_A(19)</b></p>	 <p><b>130:ene_cyano_B(7)</b></p>	 <p><b>131:ene_cyano_C(6)</b></p>	 <p><b>132:ene_cyano_D(3)</b></p>
 <p><b>133:ene_cyano_E(1)</b></p>	 <p><b>134:ene_cyano_F(1)</b></p>	 <p><b>135:ene_cyano_G(1)</b></p>	 <p><b>136:ene_five_het_A(201)</b></p>
 <p><b>137:ene_five_het_B(90)</b></p>	 <p><b>138:ene_five_het_C(85)</b></p>	 <p><b>139:ene_five_het_D(46)</b></p>	 <p><b>140:ene_five_het_E(44)</b></p>

			
141:ene_five_het_F(15)	142:ene_five_het_G(10)	143:ene_five_het_H(6)	144:ene_five_het_I(6)
			
145:ene_five_het_J(4)	146:ene_five_het_K(4)	147:ene_five_het_L(4)	148:ene_five_het_M(3)
			
149:ene_five_het_N(1)	150:ene_five_one_A(55)	151:ene_five_one_B(1)	152:ene_misc_A(5)
			
153:ene_misc_B(2)	154:ene_misc_C(1)	155:ene_misc_D(1)	156:ene_misc_E(1)
			
157:ene_one_A(3)	158:ene_one_B(2)	159:ene_one_C(1)	160:ene_one_D(1)

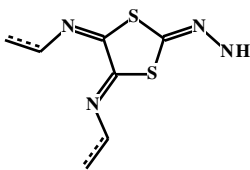
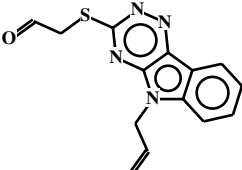
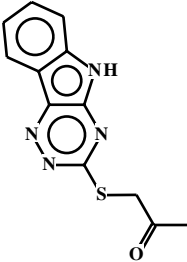
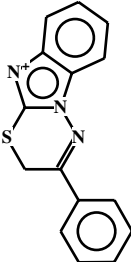
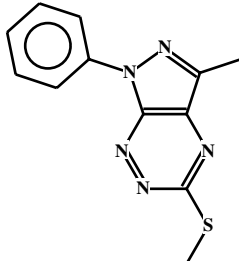
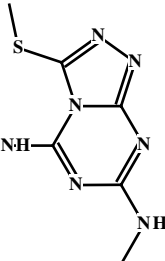
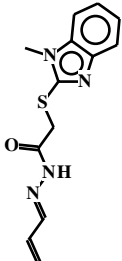
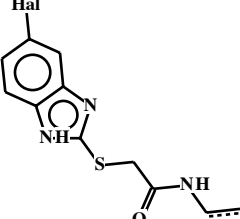
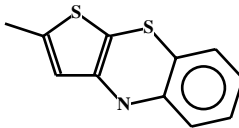
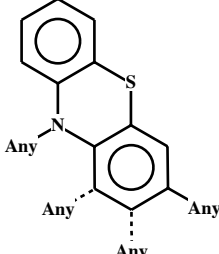
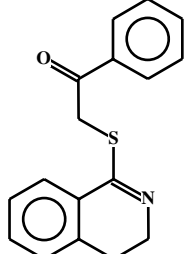
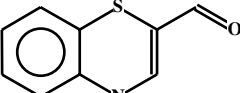
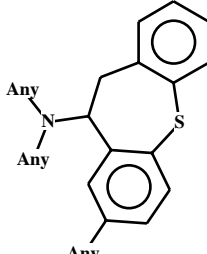
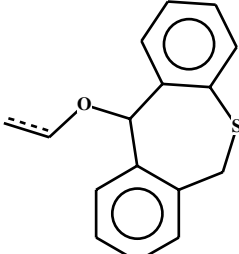
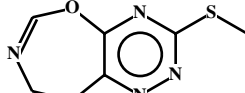
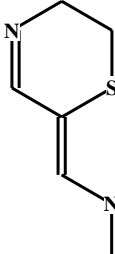
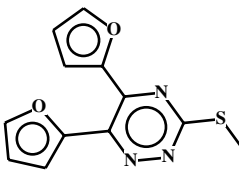
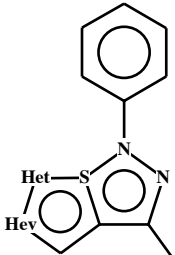
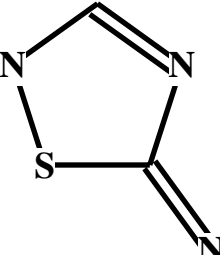
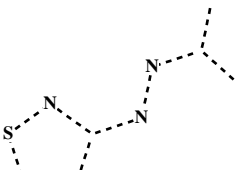
161:ene_one_amide_A(1)	162:ene_one_amide_B(1)	163:ene_one_ene_A(57)	164:ene_one_ester(24)
165:ene_one_hal(17)	166:ene_one_one_A(1)	167:ene_one_one_B(1)	168:ene_one_ynone_A(1)
169:ene_quin_methide(10)	170:ene_rhod_A(235)	171:ene_rhod_B(16)	172:ene_rhod_C(13)
173:ene_rhod_D(8)	174:ene_rhod_E(8)	175:ene_rhod_F(8)	176:ene_rhod_G(7)
177:ene_rhod_H(5)	178:ene_rhod_I(3)	179:ene_rhod_J(3)	180:ene_six_het_A(483)

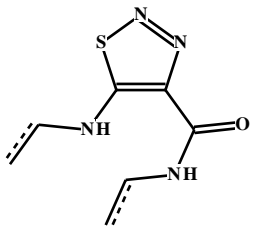
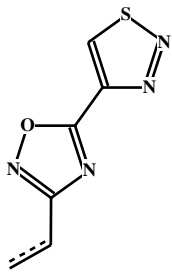
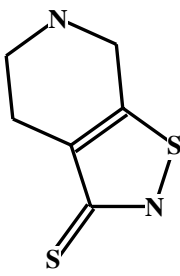
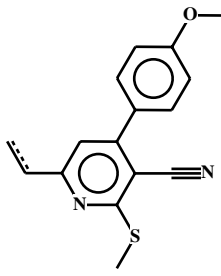
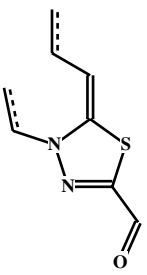
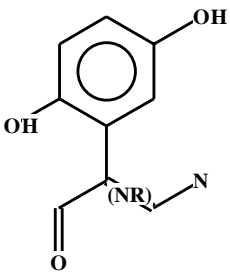
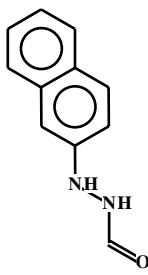
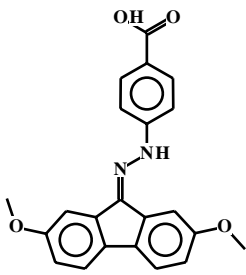
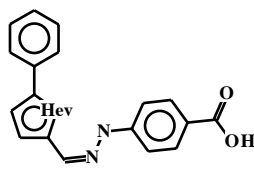
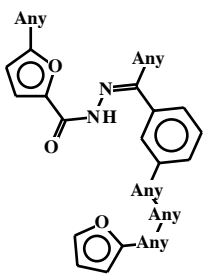
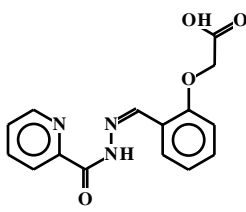
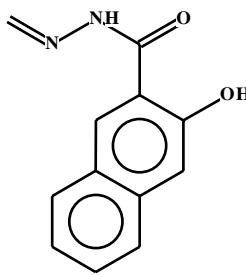
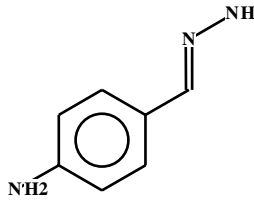
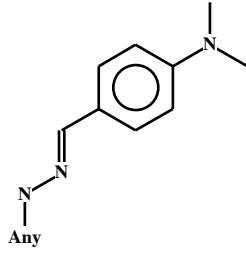
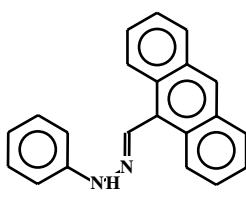
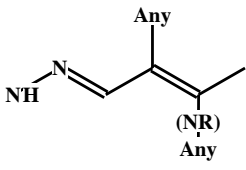
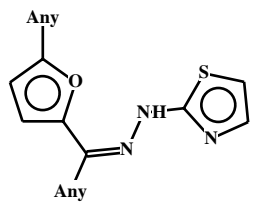
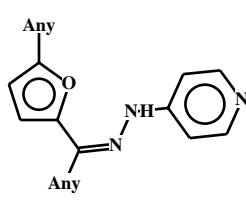
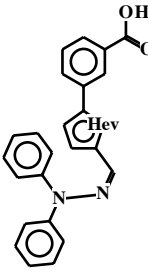
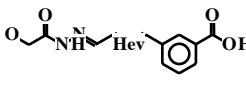
			
181:ene_six_het_B(2)	182:ene_six_het_C(2)	183:ene_six_het_D(1)	184:furan_A(3)
			
185:furan_acid_A(4)	186:het_465_misc(1)	187:het_55_A(2)	188:het_55_B(1)
			
189:het_565_A(2)	190:het_565_indole(1)	191:het_5_A(7)	192:het_5_B(4)
			
193:het_5_C(2)	194:het_5_D(2)	195:het_5_E(1)	196:het_5_ene(1)
			
197:het_5_inium(1)	198:het_5_pyrazole_OH(14)	199:het_65_A(21)	200:het_65_B(7)

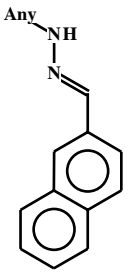
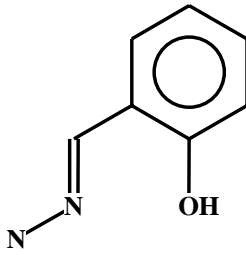
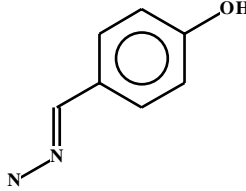
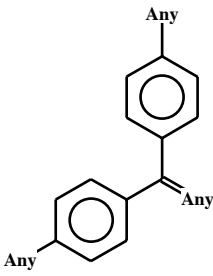
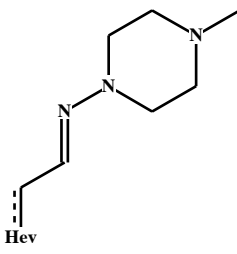
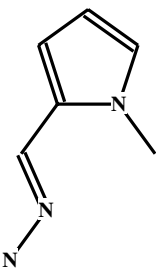
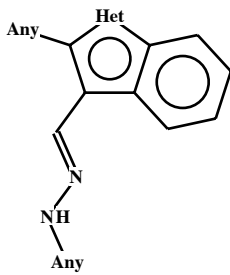
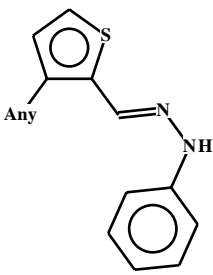
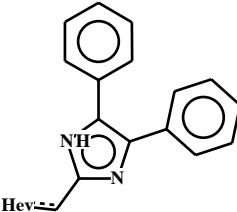
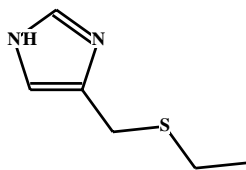
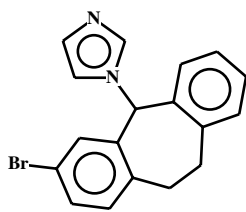
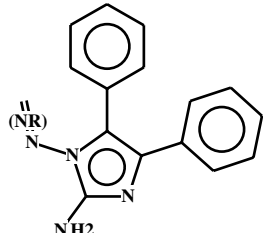
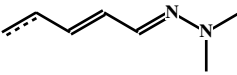
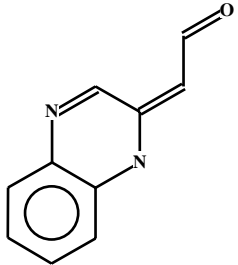
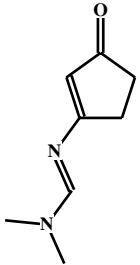
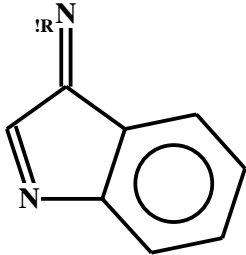
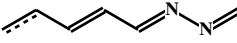
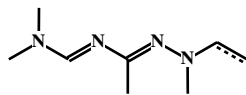
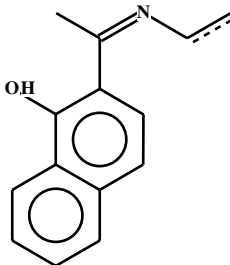
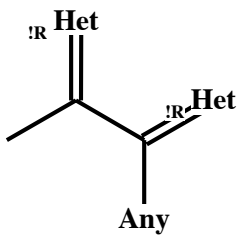
			
<b>201:het_65_C(6)</b>	<b>202:het_65_Da(2)</b>	<b>203:het_65_Db(5)</b>	<b>204:het_65_E(2)</b>
			
<b>205:het_65_F(1)</b>	<b>206:het_65_G(1)</b>	<b>207:het_65_H(1)</b>	<b>208:het_65_I(1)</b>
			
<b>209:het_65_J(1)</b>	<b>210:het_65_K(1)</b>	<b>211:het_65_L(1)</b>	<b>212:het_65_imidazole(1)</b>
			
<b>213:het_65_mannich(1)</b>	<b>214:het_65_pyridone_A(3)</b>	<b>215:het_6666_A(2)</b>	<b>216:het_666_A(5)</b>
			
<b>217:het_666_B(3)</b>	<b>218:het_666_C(1)</b>	<b>219:het_66_A(7)</b>	<b>220:het_66_B(2)</b>

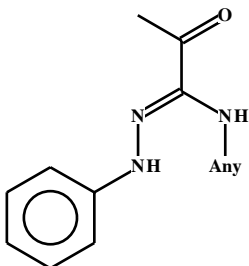
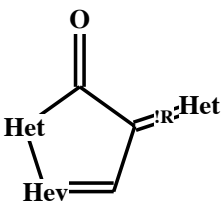
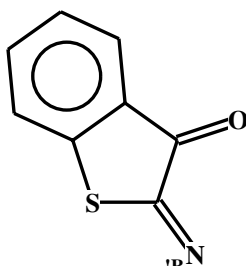
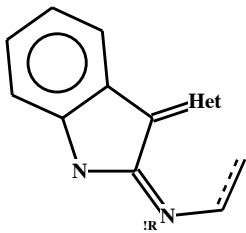
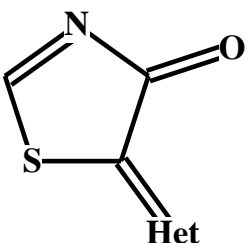
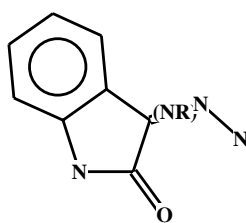
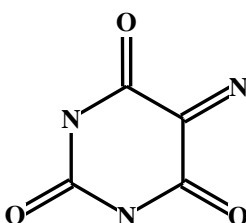
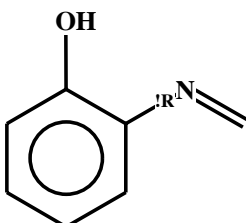
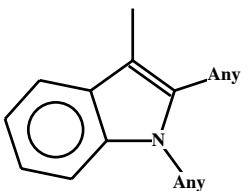
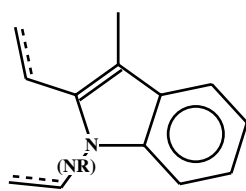
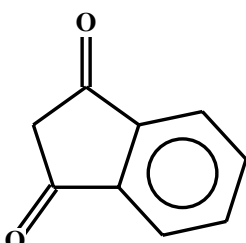
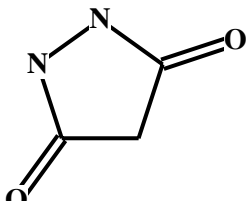
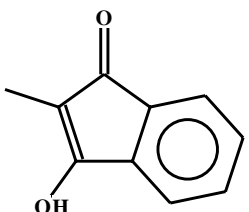
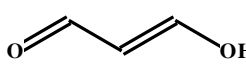
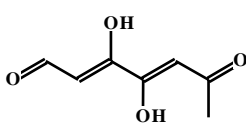
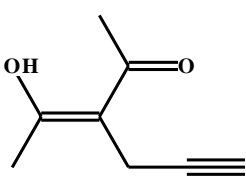
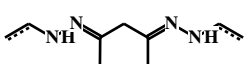
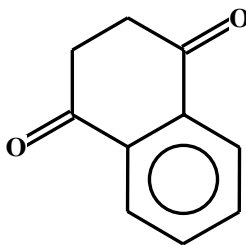
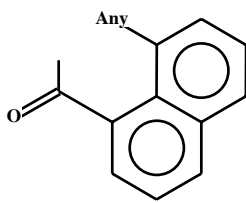
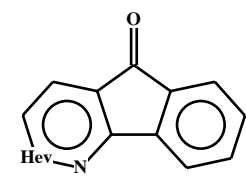
 <p>221:het_66_C(2)</p>	 <p>222:het_66_D(1)</p>	 <p>223:het_66_E(1)</p>	 <p>224:het_66_anisole(1)</p>
 <p>225:het_6_hydropyridone(~</p>	 <p>226:het_6_imidate_A(4)</p>	 <p>227:het_6_imidate_B(1)</p>	 <p>228:het_6_pyridone_NH2(1)</p>
 <p>229:het_6_pyridone_OH(5)</p>	 <p>230:het_6_tetrazine(18)</p>	 <p>231:het_76_A(1)</p>	 <p>232:het_pyraz_misc(1)</p>
 <p>233:het_pyridiniums_A(39)</p>	 <p>234:het_pyridiniums_B(2)</p>	 <p>235:het_pyridiniums_C(1)</p>	 <p>236:het_thio_5_A(8)</p>
 <p>237:het_thio_5_B(2)</p>	 <p>238:het_thio_5_C(2)</p>	 <p>239:het_thio_5_imine_A(1)</p>	 <p>240:het_thio_5_imine_B(1)</p>

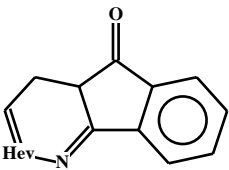
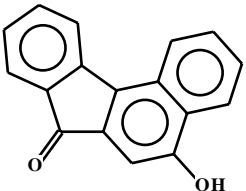
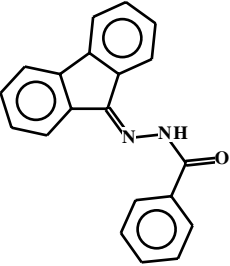
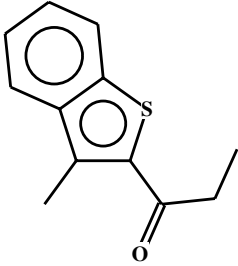
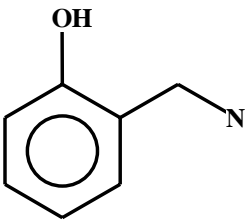
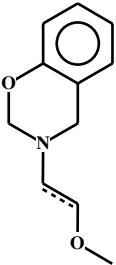
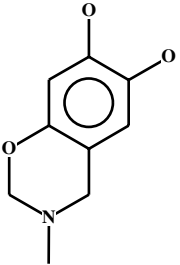
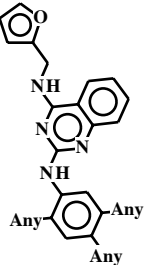
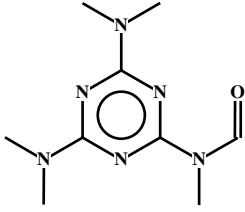
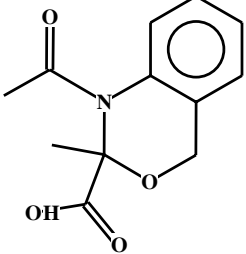
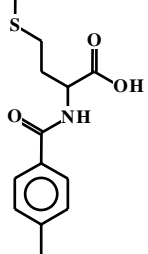
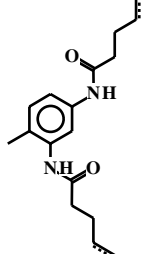
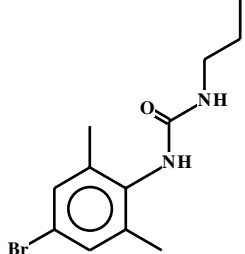
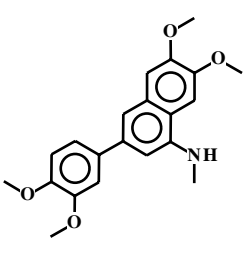
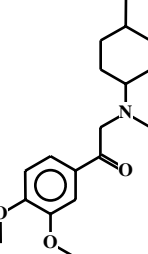
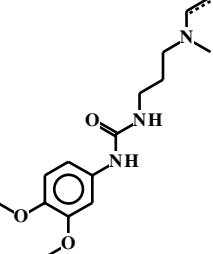
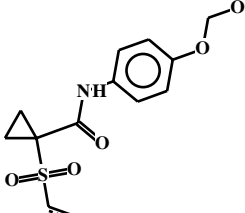
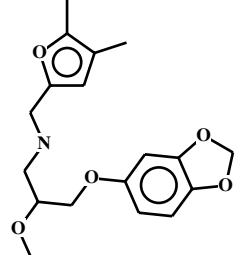
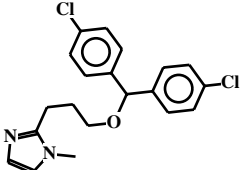
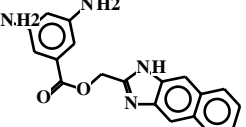


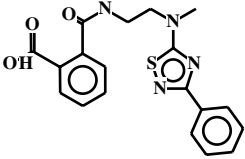
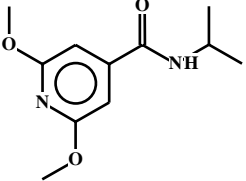
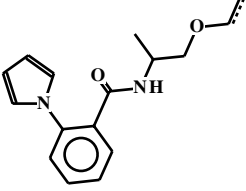
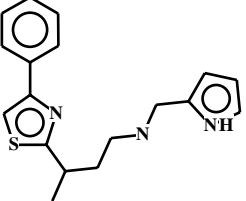
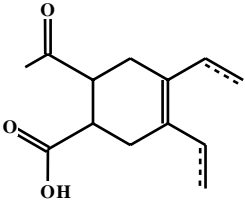
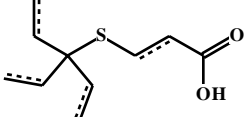
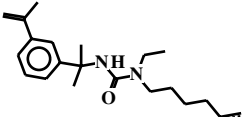
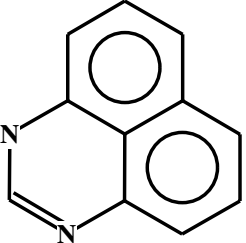
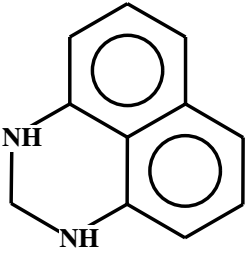
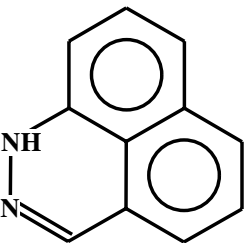
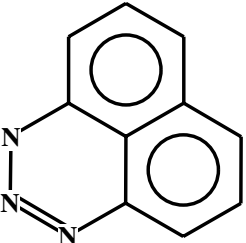
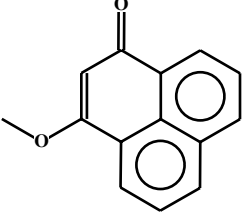
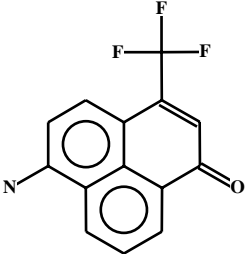
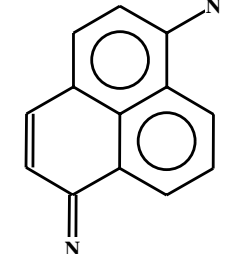
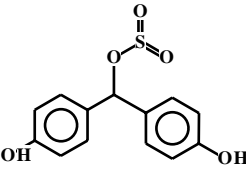
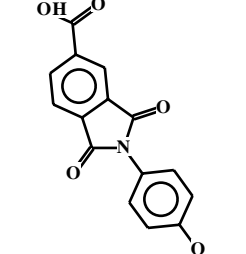
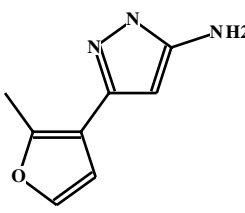
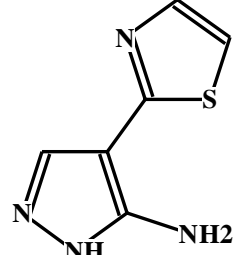
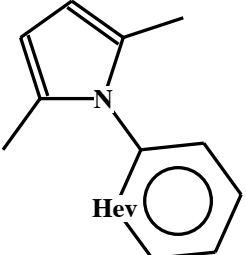
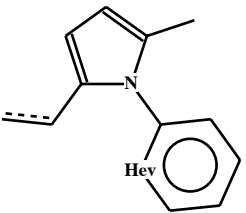
			
241:het_thio_5_imine_C(1)	242:het_thio_656a(5)	243:het_thio_656b(3)	244:het_thio_656c(1)
			
245:het_thio_65_A(3)	246:het_thio_65_B(2)	247:het_thio_65_C(2)	248:het_thio_65_D(1)
			
249:het_thio_665(1)	250:het_thio_666_A(13)	251:het_thio_66_A(3)	252:het_thio_66_one(8)
			
253:het_thio_676_A(10)	254:het_thio_676_B(1)	255:het_thio_67_A(1)	256:het_thio_6_ene(2)
			
257:het_thio_6_furan(4)	258:het_thio_N_55(5)	259:het_thio_N_5A(3)	260:het_thio_N_5B(2)

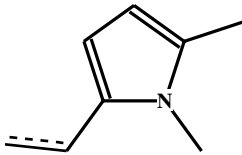
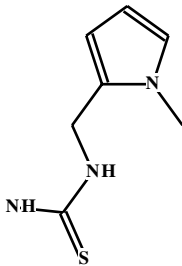
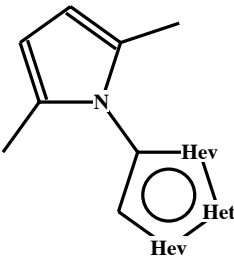
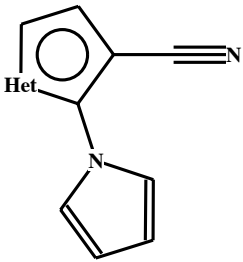
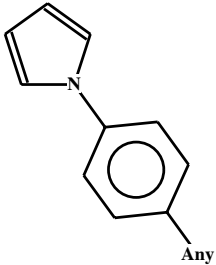
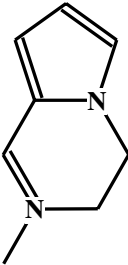
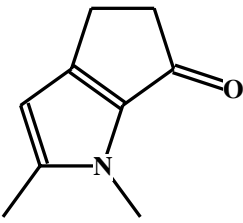
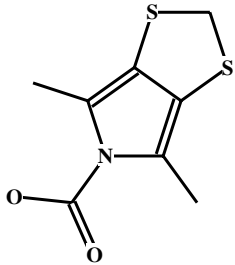
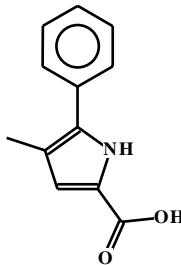
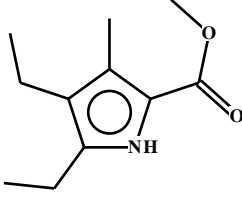
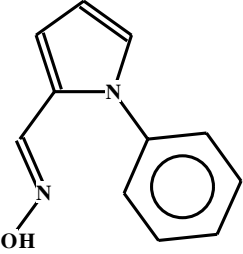
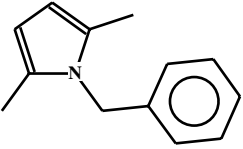
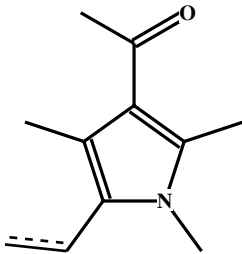
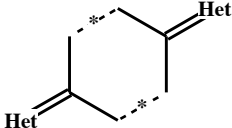
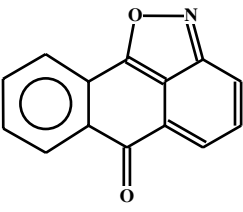
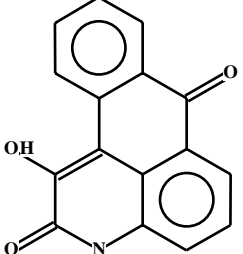
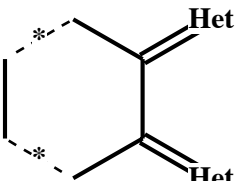
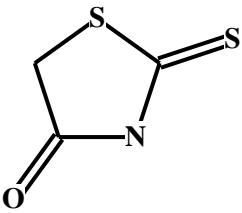
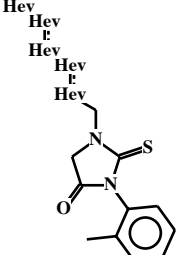
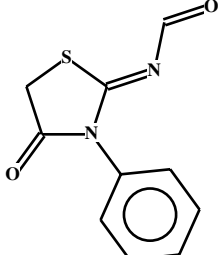
			
261:het_thio_N_5C(1)	262:het_thio_N_5D(1)	263:het_thio_N_65A(3)	264:het_thio_pyr_A(3)
			
265:het_thio_urea_ene(1)	266:hydroquin_A(2)	267:hziide_naphth(2)	268:hzone_acid_A(1)
			
269:hzone_acid_D(1)	270:hzone_acyl_misc_A(1)	271:hzone_acyl_misc_B(1)	272:hzone_acyl_naphthol(~)
			
273:hzone_anil(14)	274:hzone_anil_di_alk(35)	275:hzone_anthran_Z(1)	276:hzone_enamin(30)
			
277:hzone_furan_A(6)	278:hzone_furan_B(2)	279:hzone_furan_C(1)	280:hzone_furan_E(1)

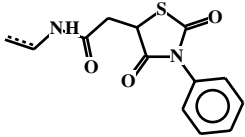
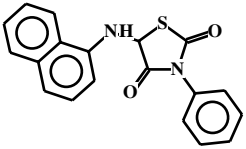
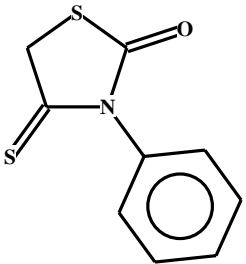
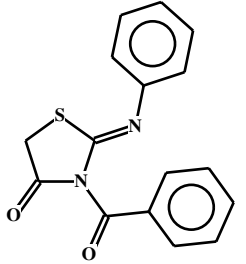
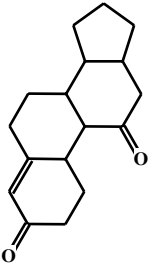
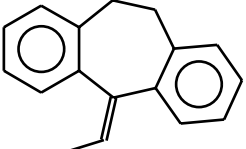
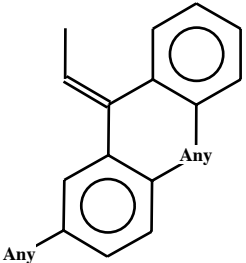
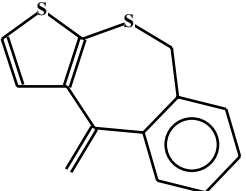
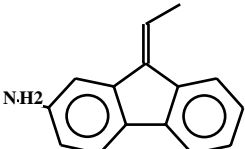
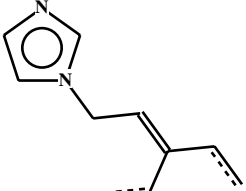
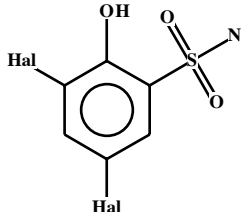
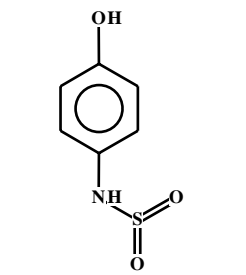
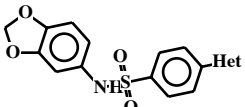
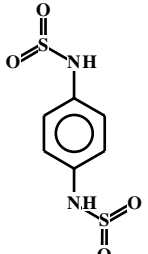
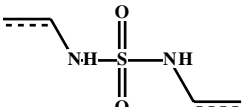
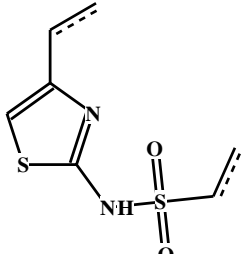
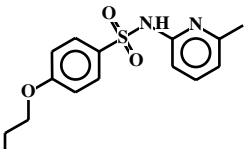
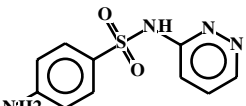
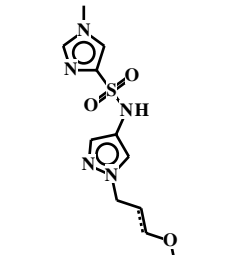
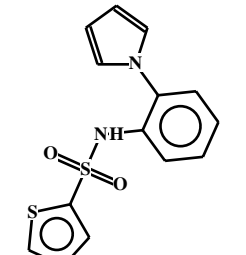
 <p>281:hzone_naphth_A(5)</p>	 <p>282:hzone_phenol_A(479)</p>	 <p>283:hzone_phenol_B(215)</p>	 <p>284:hzone_phenone(7)</p>
 <p>285:hzone_pipzn(79)</p>	 <p>286:hzone_pyrrol(64)</p>	 <p>287:hzone_thiophene_A(11)</p>	 <p>288:hzone_thiophene_B(4)</p>
 <p>289:imidazole_A(19)</p>	 <p>290:imidazole_B(2)</p>	 <p>291:imidazole_C(1)</p>	 <p>292:imidazole_amino_A(1)</p>
 <p>293:imine_ene_A(5)</p>	 <p>294:imine_ene_one_A(3)</p>	 <p>295:imine_ene_one_B(1)</p>	 <p>296:imine_imine_A(9)</p>
 <p>297:imine_imine_B(3)</p>	 <p>298:imine_imine_C(3)</p>	 <p>299:imine_naphthol_A(1)</p>	 <p>300:imine_one_A(321)</p>

 <p>301:imine_one_B(4)</p>	 <p>302:imine_one_fives(89)</p>	 <p>303:imine_one_fives_B(9)</p>	 <p>304:imine_one_fives_C(2)</p>
 <p>305:imine_one_fives_D(1)</p>	 <p>306:imine_one_isatin(189)</p>	 <p>307:imine_one_sixes(27)</p>	 <p>308:imine_phenol_A(3)</p>
 <p>309:indol_3yl_alk(461)</p>	 <p>310:indole_3yl_alk_B(1)</p>	 <p>311:keto_keto_beta_A(68)</p>	 <p>312:keto_keto_beta_B(12)</p>
 <p>313:keto_keto_beta_C(7)</p>	 <p>314:keto_keto_beta_D(5)</p>	 <p>315:keto_keto_beta_E(1)</p>	 <p>316:keto_keto_beta_F(1)</p>
 <p>317:keto_keto_beta_zone(~</p>	 <p>318:keto_keto_gamma(5)</p>	 <p>319:keto_naphthol_A(2)</p>	 <p>320:keto_phenone_A(11)</p>

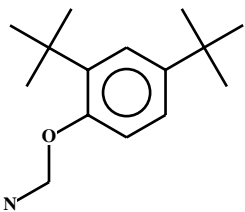
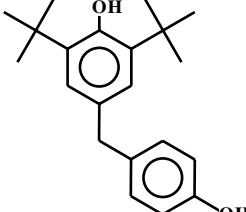
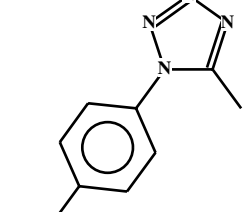
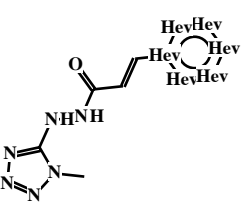
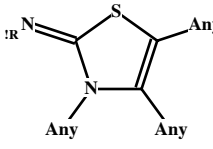
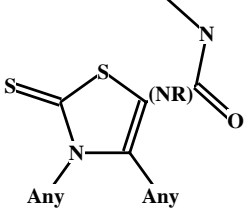
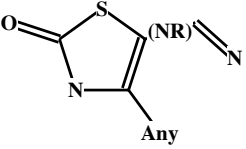
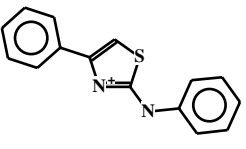
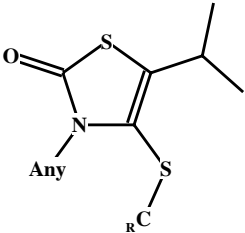
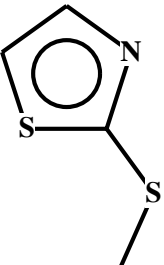
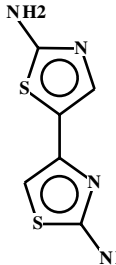
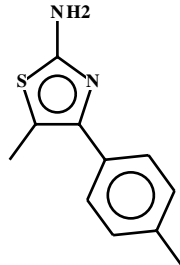
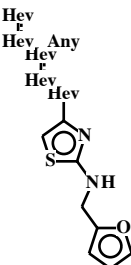
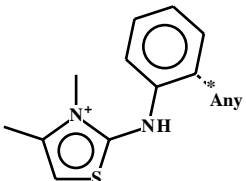
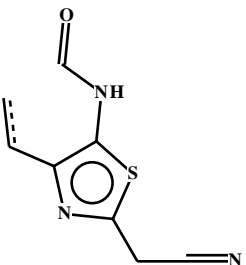
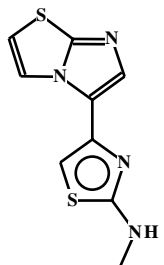
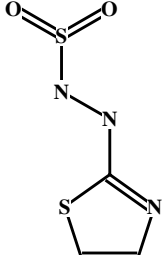
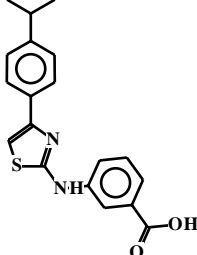
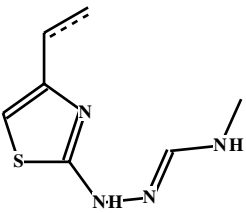
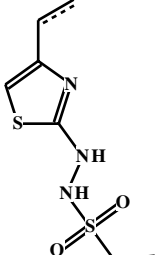
			
321:keto_phenone_B(1)	322:keto_phenone_C(1)	323:keto_phenone_zone_A(1)	324:keto_thiophene(3)
			
325:mannich_A(296)	326:mannich_B(1)	327:mannich_catechol_A(1)	328:melamine_A(3)
			
329:melamine_B(1)	330:misc_aminal_acid(1)	331:misc_aminoacid_A(1)	332:misc_anilide_A(1)
			
333:misc_anilide_B(1)	334:misc_anisole_A(1)	335:misc_anisole_B(1)	336:misc_anisole_C(1)
			
337:misc_cyclopropane(1)	338:misc_furan_A(1)	339:misc_imidazole(1)	340:misc_naphthimidazole(1)

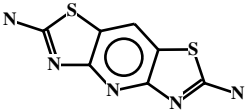
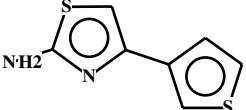
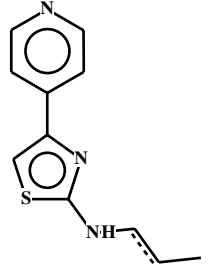
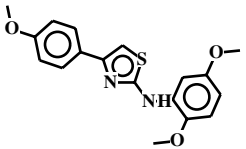
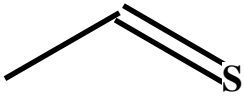
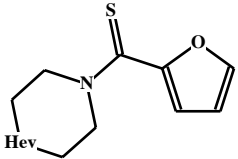
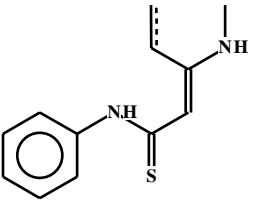
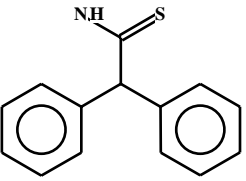
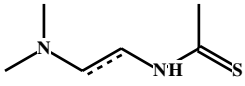
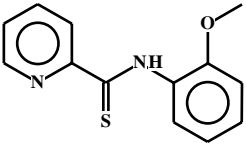
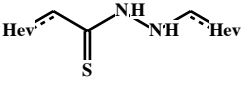
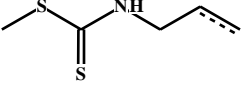
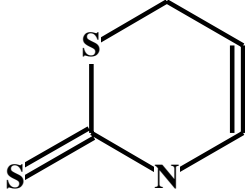
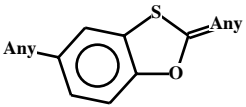
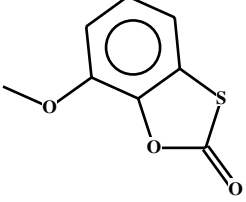
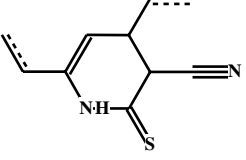
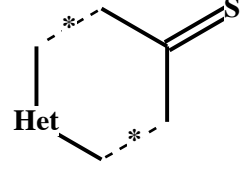
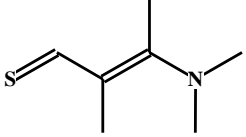
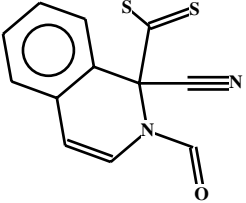
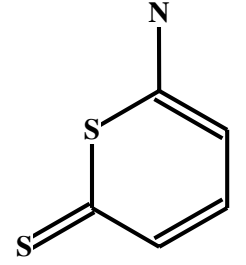
			
341:misc_phthal_thio_N(1)	342:misc_pyridine_OC(1)	343:misc_pyrrole_benz(1)	344:misc_pyrrole_thiaz(1)
			
345:misc_stilbene(1)	346:misc_trityl_A(1)	347:misc_urea_A(1)	348:naphth_amino_A(25)
			
349:naphth_amino_B(25)	350:naphth_amino_C(2)	351:naphth_amino_D(2)	352:naphth_ene_one_A(1)
			
353:naphth_ene_one_B(1)	354:naphth_ene_one_C(1)	355:phenol_sulfite_A(1)	356:phthalimide_misc(2)
			
357:pyrazole_amino_A(1)	358:pyrazole_amino_B(1)	359:pyrrole_A(118)	360:pyrrole_B(29)

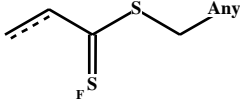
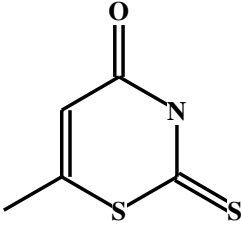
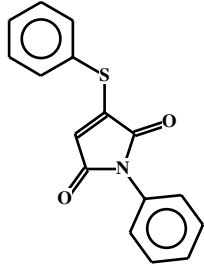
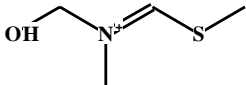
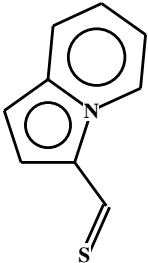
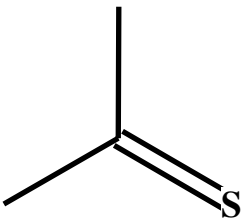
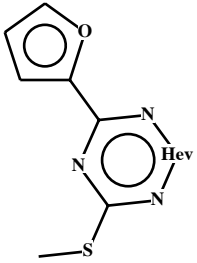
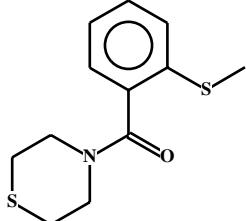
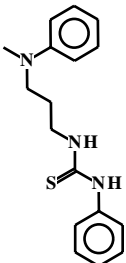
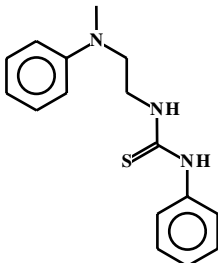
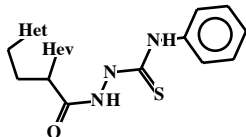
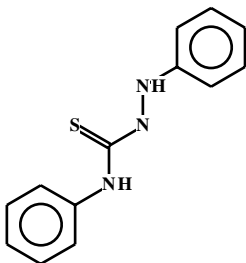
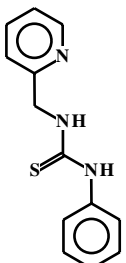
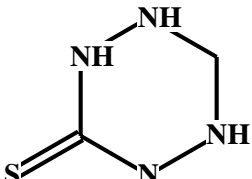
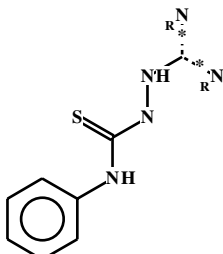
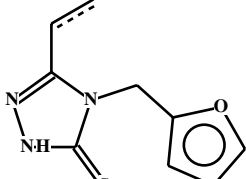
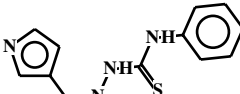
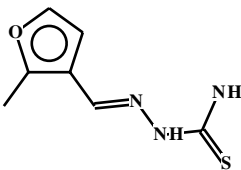
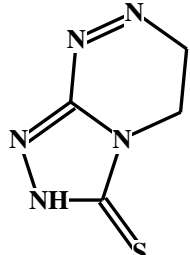
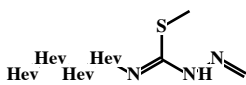
 <p>361:pyrrole_C(8)</p>	 <p>362:pyrrole_D(5)</p>	 <p>363:pyrrole_E(5)</p>	 <p>364:pyrrole_F(5)</p>
 <p>365:pyrrole_G(4)</p>	 <p>366:pyrrole_H(3)</p>	 <p>367:pyrrole_I(2)</p>	 <p>368:pyrrole_J(1)</p>
 <p>369:pyrrole_K(1)</p>	 <p>370:pyrrole_L(1)</p>	 <p>371:pyrrole_M(1)</p>	 <p>372:pyrrole_N(1)</p>
 <p>373:pyrrole_O(1)</p>	 <p>374:quinone_A(370)</p>	 <p>375:quinone_B(5)</p>	 <p>376:quinone_C(2)</p>
 <p>377:quinone_D(2)</p>	 <p>378:rhod_sat_A(33)</p>	 <p>379:rhod_sat_B(3)</p>	 <p>380:rhod_sat_C(3)</p>

			
381:rhod_sat_D(3)	382:rhod_sat_E(1)	383:rhod_sat_F(1)	384:rhod_sat_imine_A(1)
			
385:steroid_A(2)	386:styrene_A(13)	387:styrene_B(8)	388:styrene_C(4)
			
389:styrene_anil_A(1)	390:styrene_imidazole_A(~)	391:sulfonamide_A(43)	392:sulfonamide_B(41)
			
393:sulfonamide_C(5)	394:sulfonamide_D(2)	395:sulfonamide_E(2)	396:sulfonamide_F(1)
			
397:sulfonamide_G(1)	398:sulfonamide_H(1)	399:sulfonamide_I(1)	400:sulfonamide_J(1)

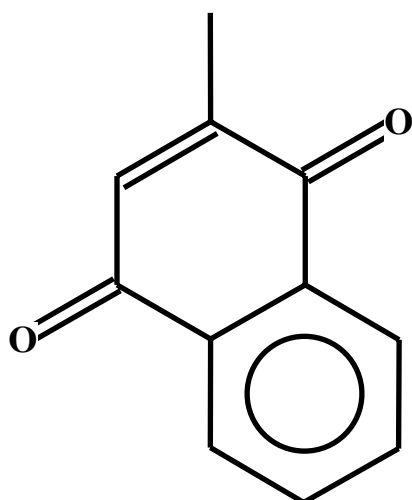


			
<b>401:tert_butyl_A(2)</b>	<b>402:tert_butyl_B(1)</b>	<b>403:tetrazole_A(1)</b>	<b>404:tetrazole_hzide(1)</b>
			
<b>405:thiaz_ene_A(128)</b>	<b>406:thiaz_ene_B(17)</b>	<b>407:thiaz_ene_C(11)</b>	<b>408:thiaz_ene_D(8)</b>
			
<b>409:thiaz_ene_E(8)</b>	<b>410:thiazol_SC_A(3)</b>	<b>411:thiazole_amine_A(4)</b>	<b>412:thiazole_amine_B(3)</b>
			
<b>413:thiazole_amine_C(3)</b>	<b>414:thiazole_amine_D(3)</b>	<b>415:thiazole_amine_E(2)</b>	<b>416:thiazole_amine_F(2)</b>
			
<b>417:thiazole_amine_G(2)</b>	<b>418:thiazole_amine_H(1)</b>	<b>419:thiazole_amine_I(1)</b>	<b>420:thiazole_amine_J(1)</b>

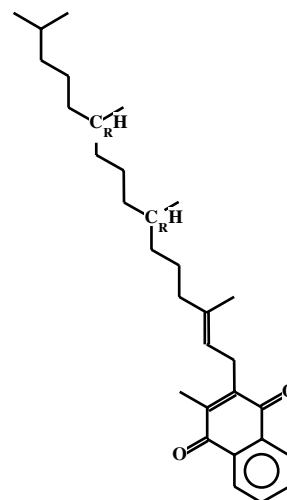
			
421:thiazole_amine_K(1)	422:thiazole_amine_L(1)	423:thiazole_amine_M(1)	424:thiazole_amine_N(1)
			
425:thio_aldehyd_A(3)	426:thio_amide_A(6)	427:thio_amide_B(2)	428:thio_amide_C(2)
			
429:thio_amide_D(2)	430:thio_amide_E(1)	431:thio_amide_F(1)	432:thio_carbam_A(1)
			
433:thio_carbam_ene(2)	434:thio_carbonate_A(15)	435:thio_carbonate_B(3)	436:thio_cyano_A(1)
			
437:thio_dibenzo(23)	438:thio_ene_amine_A(1)	439:thio_est_cyano_A(1)	440:thio_ester_A(5)

 <p>441:thio_ester_B(4)</p>	 <p>442:thio_ester_C(2)</p>	 <p>443:thio_imide_A(1)</p>	 <p>444:thio_imine_ium(2)</p>
 <p>445:thio_keto_het(2)</p>	 <p>446:thio_ketone(43)</p>	 <p>447:thio_pyridine_A(1)</p>	 <p>448:thio_thiomorph_Z(1)</p>
 <p>449:thio_urea_A(12)</p>	 <p>450:thio_urea_B(9)</p>	 <p>451:thio_urea_C(9)</p>	 <p>452:thio_urea_D(8)</p>
 <p>453:thio_urea_E(7)</p>	 <p>454:thio_urea_F(6)</p>	 <p>455:thio_urea_G(5)</p>	 <p>456:thio_urea_H(3)</p>
 <p>457:thio_urea_I(3)</p>	 <p>458:thio_urea_J(2)</p>	 <p>459:thio_urea_K(2)</p>	 <p>460:thio_urea_L(1)</p>

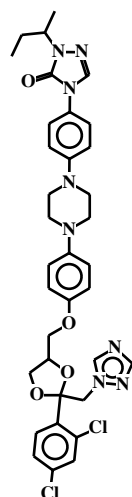
A benzene ring with a 'Hal' substituent at the 1-position and a thio-urea group (-NH-C(=S)-NH-) at the 2-position. The nitrogen of the thio-urea group is connected to a propyl chain ending in a dimethylamino group.	A thio-urea group (-NH-C(=S)-NH-) where the nitrogen is connected to a benzyl group and the other nitrogen is connected to a furan-2-ylmethyl group.	A thio-urea group (-NH-C(=S)-NH-) where the nitrogen is connected to a furan-2-ylmethyl group and the other nitrogen is connected to a benzofuran-2-yl group.	A thio-urea group (-NH-C(=S)-NH-) where the nitrogen is connected to a furan-2-yl group and the other nitrogen is connected to a furan-2-yl group with an (NR) substituent.
A thio-urea group (-NH-C(=S)-NH-) where the nitrogen is connected to a benzyl group and the other nitrogen is connected to a benzyl group with a naphthalen-1-yl substituent.	A thio-urea group (-NH-C(=S)-NH-) where the nitrogen is connected to a benzyl group and the other nitrogen is connected to a benzyl group with a propenal substituent.	A thiophene ring with an 'Any' substituent at the 2-position and a thio-urea group (-NH-C(=S)-NH-) at the 3-position, which is further connected to a benzyl group with a dimethylamino substituent.	A thiophene ring with a methoxy group at the 2-position and a thio-urea group (-NH-C(=S)-NH-) at the 3-position.
A thiophene ring with an 'Any' substituent at the 2-position and a thio-urea group (-NH-C(=S)-NH-) at the 3-position, which is further connected to a benzyl group with a naphthalen-1-yl substituent.	A thiophene ring with a methyl group at the 2-position and a thio-urea group (-NH-C(=S)-NH-) at the 3-position, which is further connected to a benzyl group with a propenal substituent.	A thiophene ring with a methyl group at the 2-position and a thio-urea group (-NH-C(=S)-NH-) at the 3-position, which is further connected to a benzyl group with a propenal substituent.	A thiophene ring with a methyl group at the 2-position and a thio-urea group (-NH-C(=S)-NH-) at the 3-position, which is further connected to a benzyl group with a propenal substituent.
A thiophene ring with a methyl group at the 2-position and a thio-urea group (-NH-C(=S)-NH-) at the 3-position, which is further connected to a benzyl group with a propenal substituent.	A thiophene ring with a methyl group at the 2-position and a thio-urea group (-NH-C(=S)-NH-) at the 3-position, which is further connected to a benzyl group with a propenal substituent.	A thiophene ring with a methyl group at the 2-position and a thio-urea group (-NH-C(=S)-NH-) at the 3-position, which is further connected to a benzyl group with a propenal substituent.	A thiophene ring with a methyl group at the 2-position and a thio-urea group (-NH-C(=S)-NH-) at the 3-position, which is further connected to a benzyl group with a propenal substituent.
A thiophene ring with a methyl group at the 2-position and a thio-urea group (-NH-C(=S)-NH-) at the 3-position, which is further connected to a benzyl group with a propenal substituent.	A thiophene ring with a methyl group at the 2-position and a thio-urea group (-NH-C(=S)-NH-) at the 3-position, which is further connected to a benzyl group with a propenal substituent.	A thiophene ring with a methyl group at the 2-position and a thio-urea group (-NH-C(=S)-NH-) at the 3-position, which is further connected to a benzyl group with a propenal substituent.	A thiophene ring with a hydroxyl group at the 2-position.



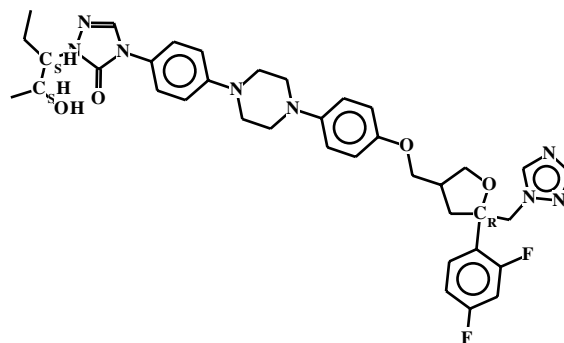
**1:Menadione (VitK)**



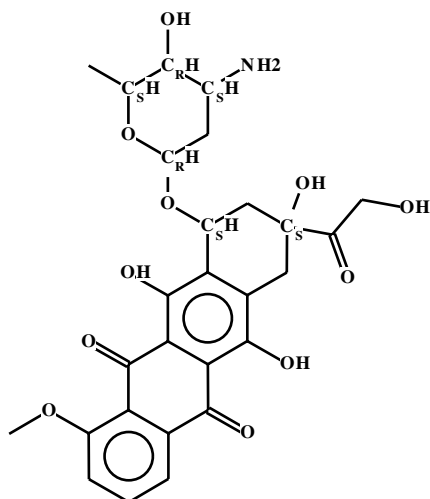
**2:Phytomenadione (VitK)**



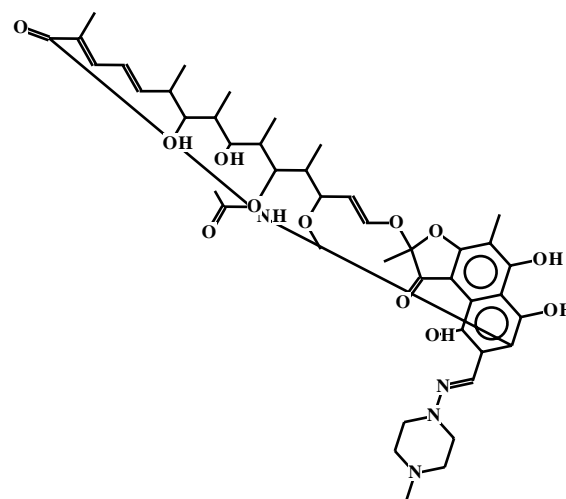
**3:itraconazole**



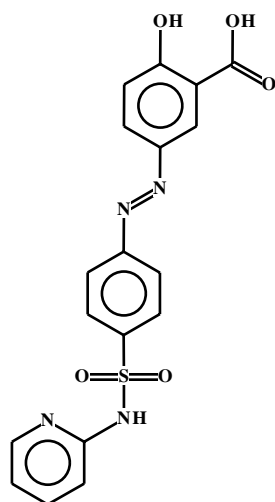
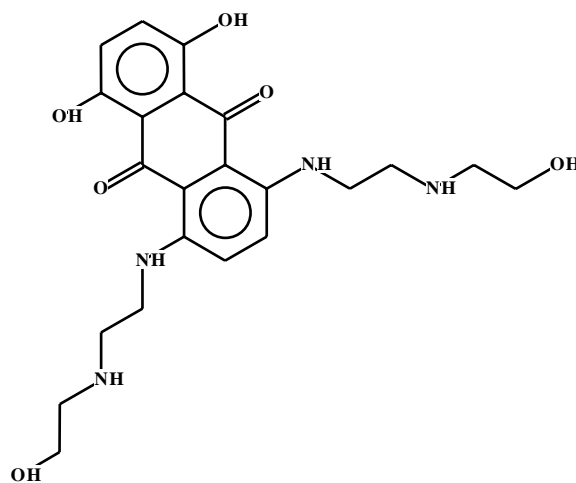
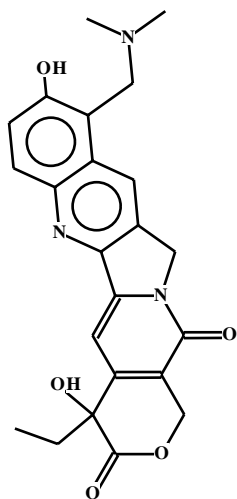
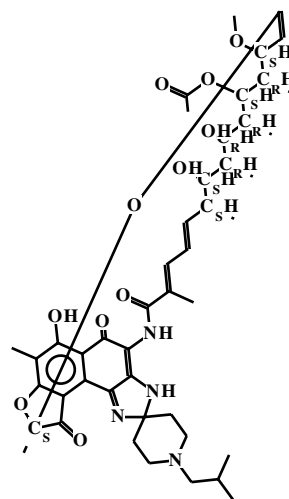
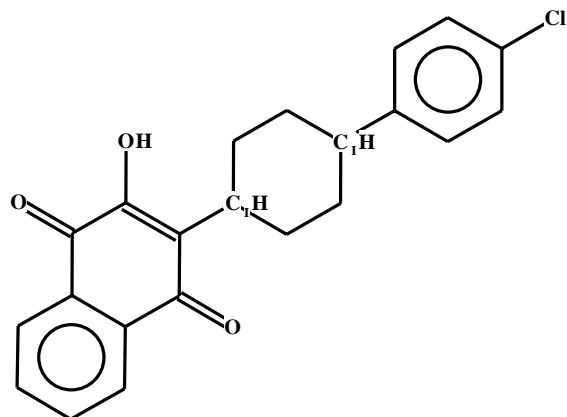
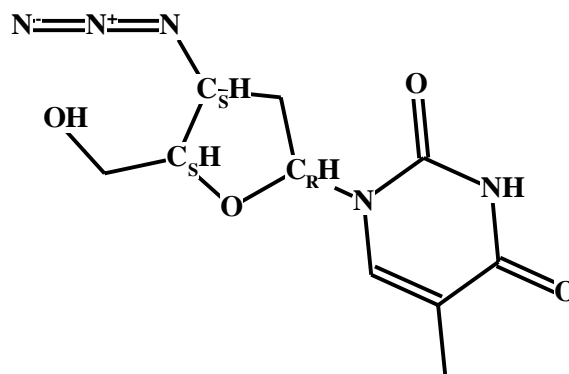
**4:posaconazole**

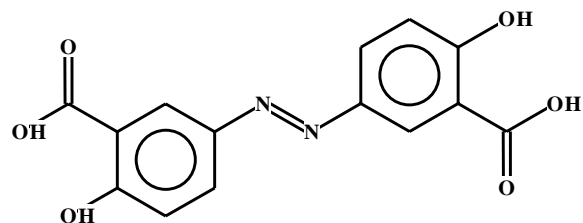


**5:epirubicin**

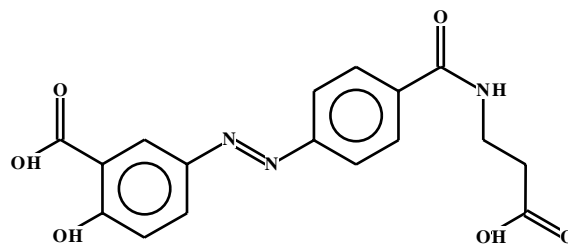


**6:rifampicin**

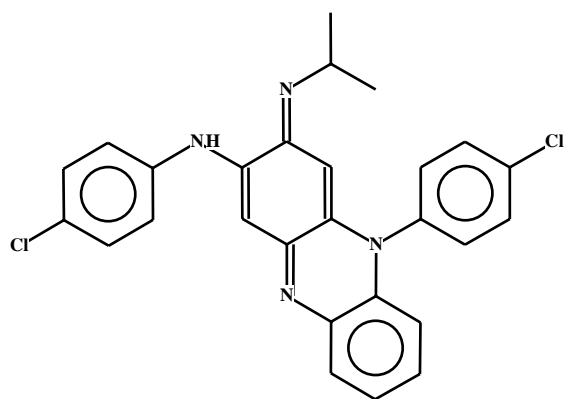
**7:sulfasalazine****8:mitozantrone****9:topotecan****10:rifabutin****11:atovaquone****12:zidovudine**



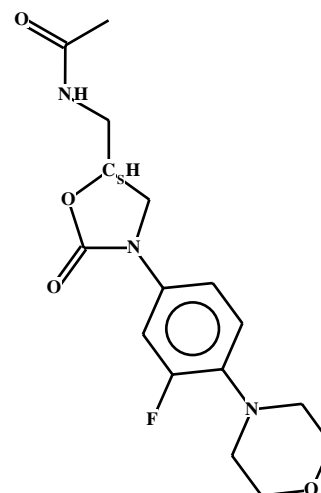
**13:olsalazine**



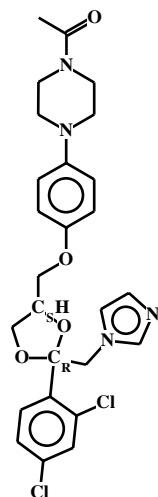
**14:balsalazide**



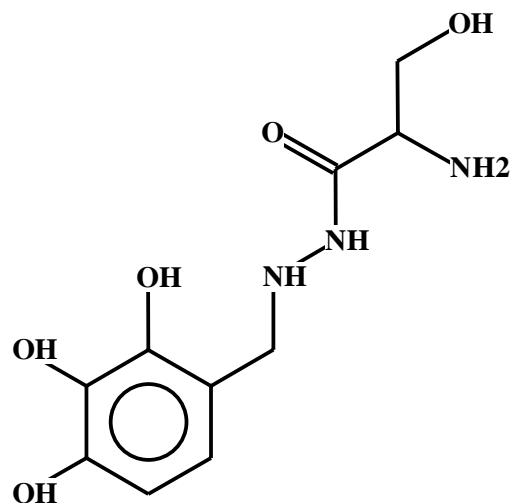
**15:clofazimine**



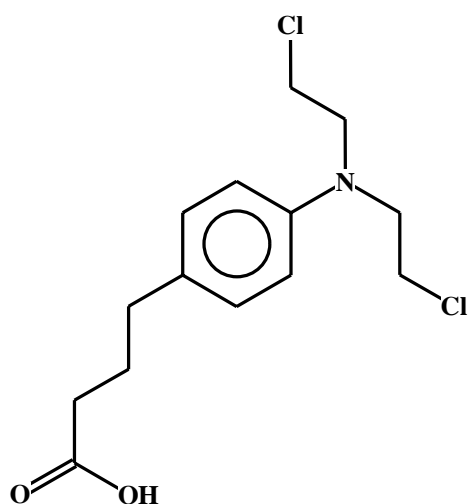
**16:linezolid**



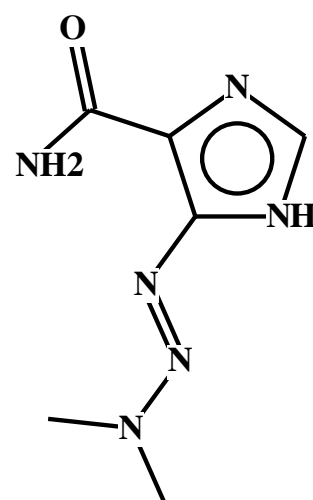
**17:ketoconazole**



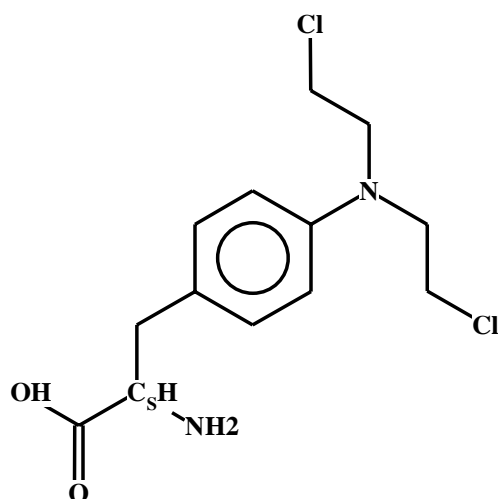
**18:benserazide**



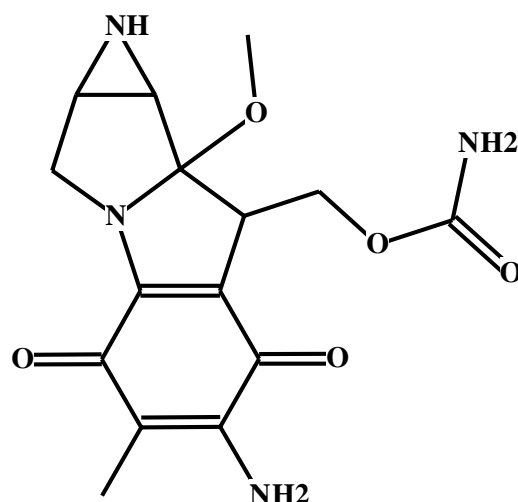
**19:chlorambucil**



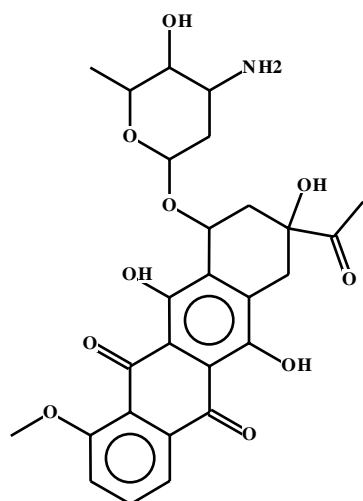
**20:dacarbazine**



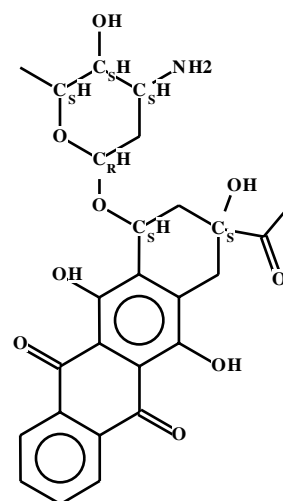
**21:melphalan**



**22:mitomycin**

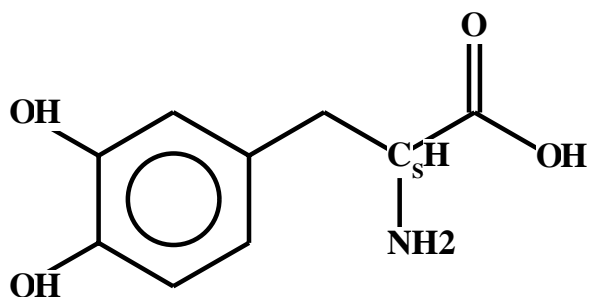
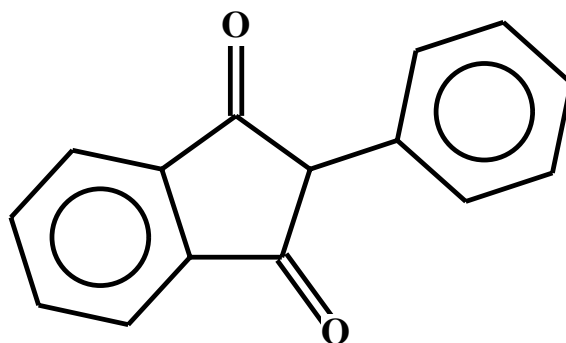
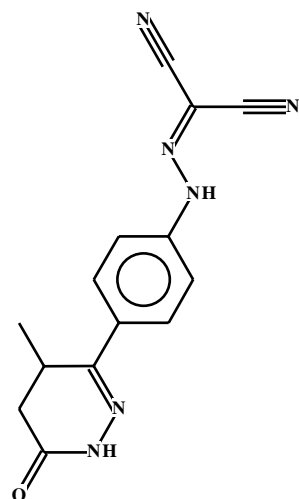
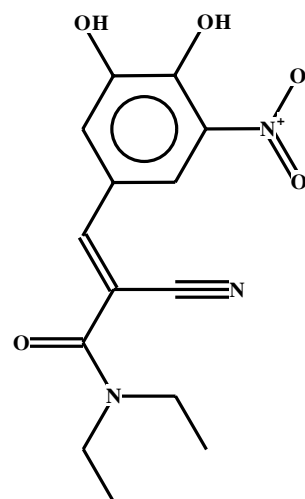
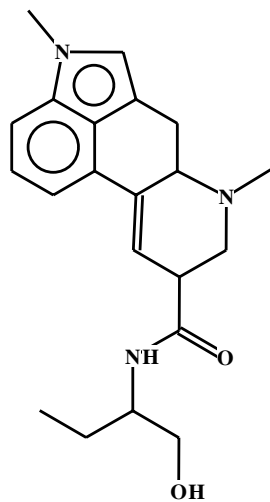
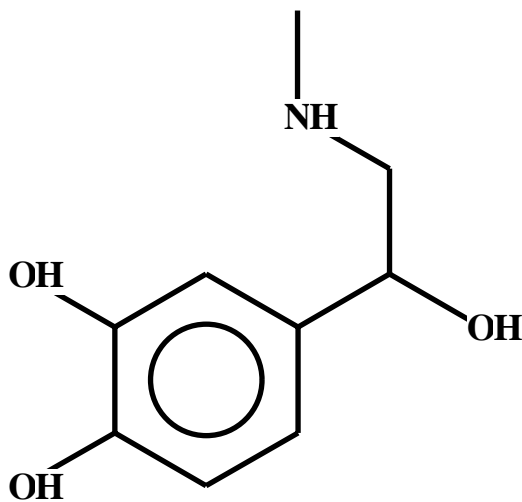


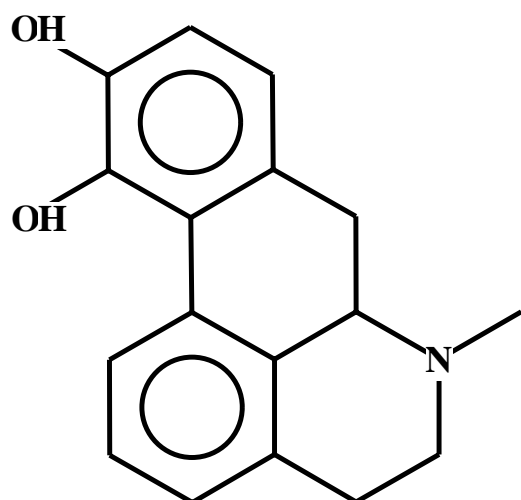
**23:daunorubicin**



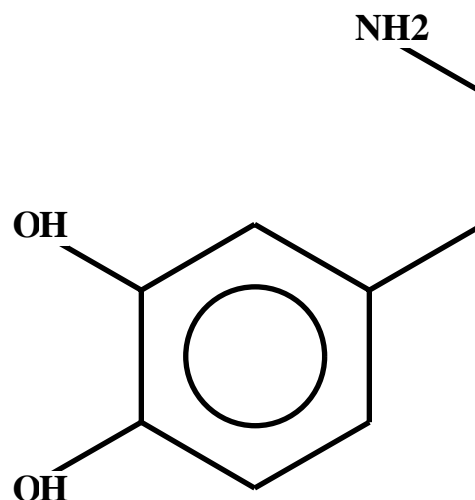
**24:idarubicin**



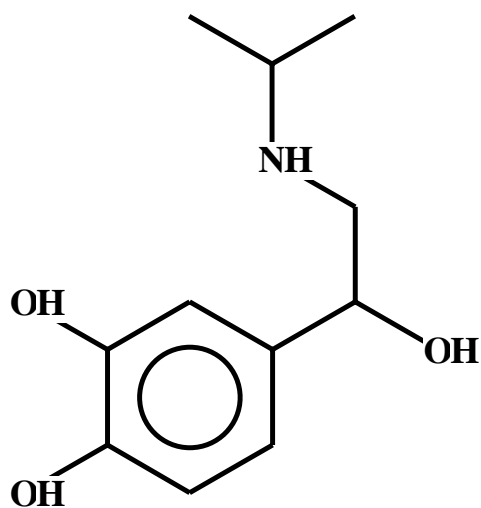
**25:levodopa****26:Phenindione****27:levosimendan****28:entacapone****29:methysergide****30:adrenaline**



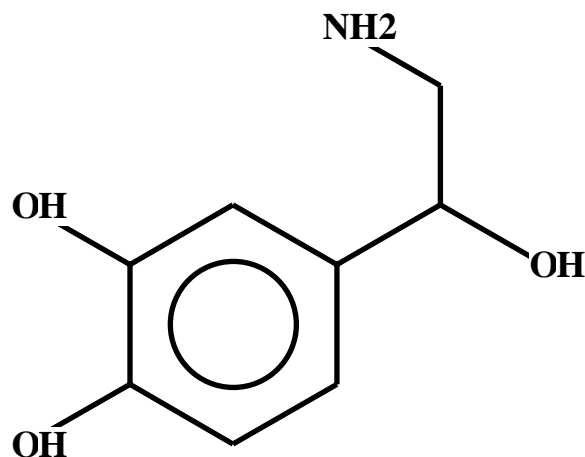
**31:apomorphine**



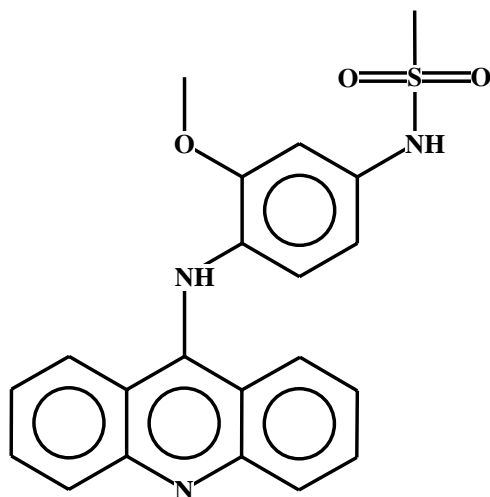
**32:dopamine**



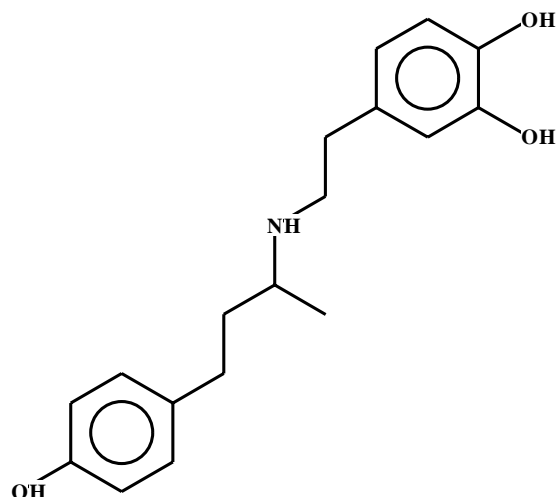
**33:isoprenaline**



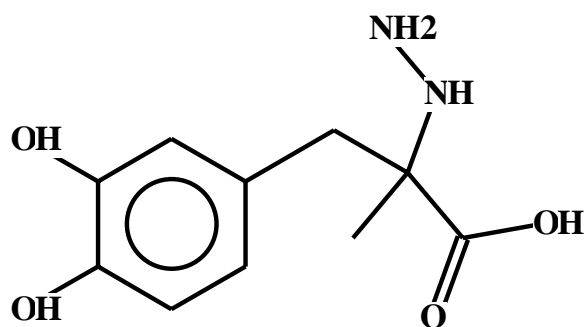
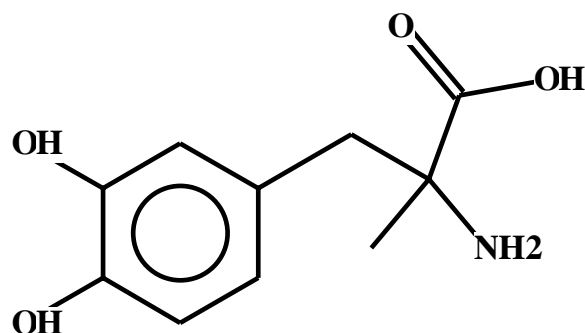
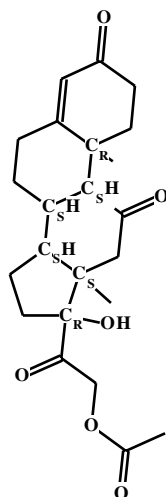
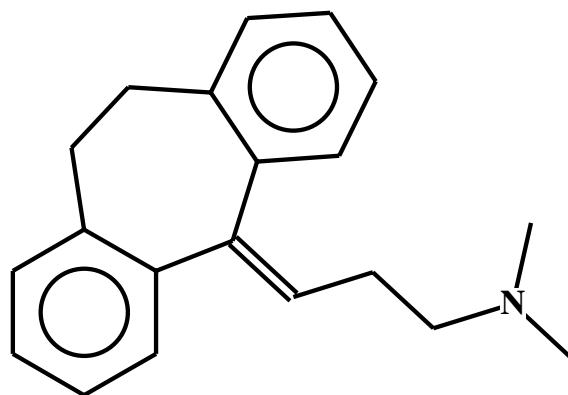
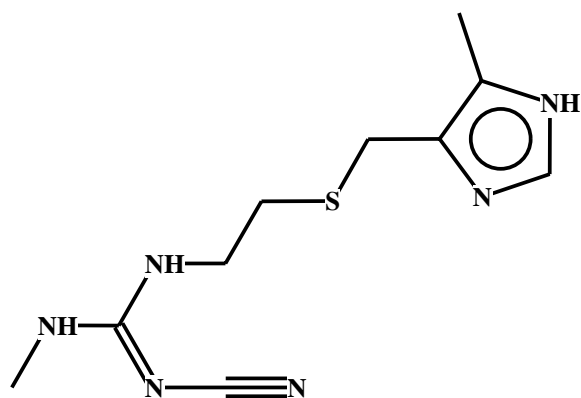
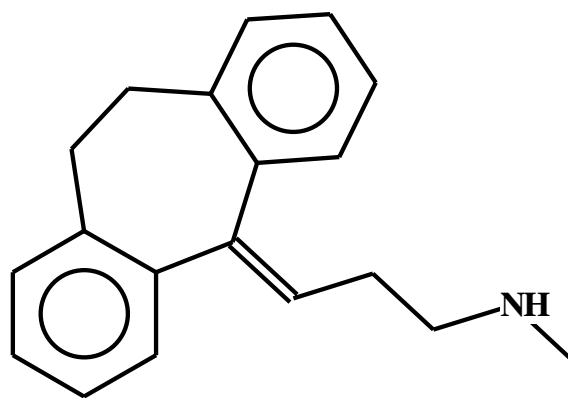
**34:noradrenaline**

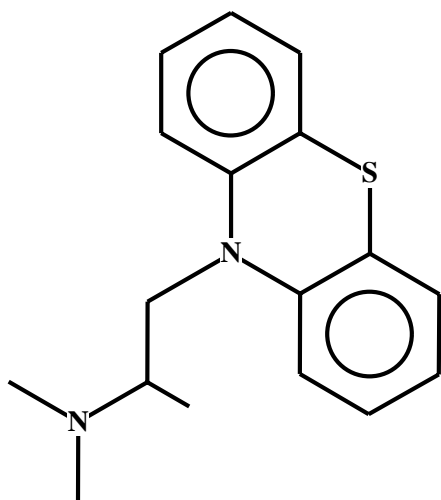
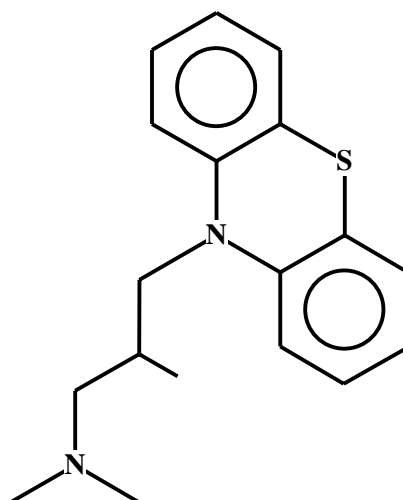
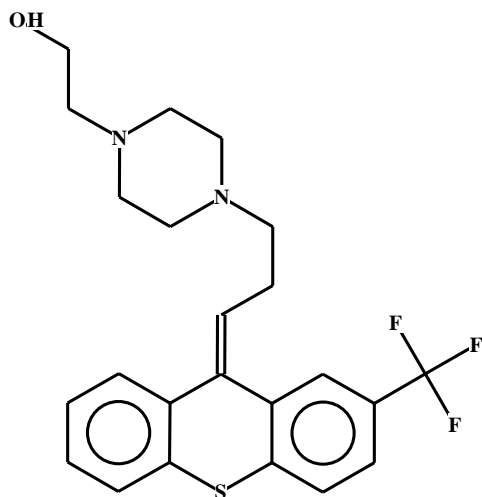
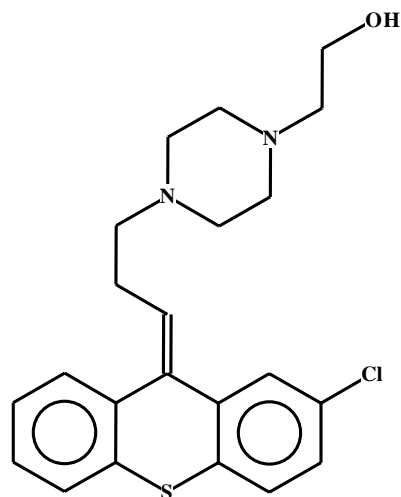
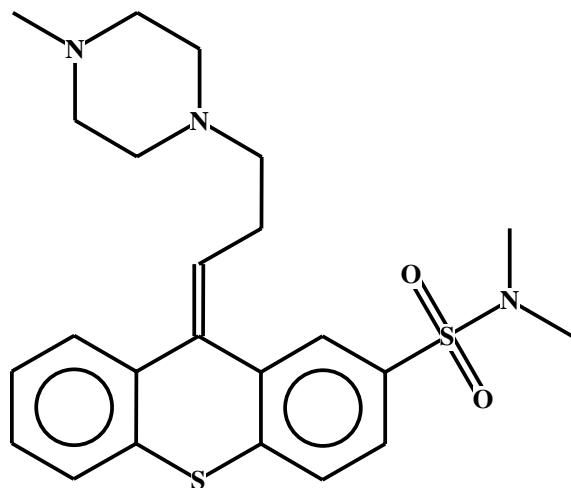


**35:amsacrine**



**36:dobutamine**

**37:carbidopa****38:methyldopa****39:cortisone acetate****40:amitriptyline****41:cimetidine****42:nortriptyline**

**43:promethazine\_hydrochloride****44:trimeprazine****45:flupenthixol****46:zuclopenthixol****47:thiothixene**