Excluded Hetcodes

CO ACP ADP AMP ANP ATP
H4B
CDP CMP COA CAA CAO CTP
FAD FMN
GDP GMP GNP GTP
NAD NAP NAI NDO NDP
MCA
SCA
UMP
1CM
2BM 2MO 6WO 6HE 749 7HE
A71 A72 A48 AF3 AIO ALF AMS APW ARS AST AUC
B69 B70 BEF BEF BF2 BH1 BPT BRO
C1O C2C CAC CAD CAS CFM CFN CFO CFQ CLF CLP CM6 CN1 CNB CNF CUB CUM CUN CUO
DP4 DRN DW2 DWC
EMC EMT F3S FCO FDC FEO FES FNE FS1 FS2 FS3 FS4 FSO
HC0 HC1 HF3 HF5 HGB HGC HGI HO3
IAP IBZ IDO IOD ISU IUM
JM1
KYS KEG
LCP LPT
MAC MAP MBO MGF MGO MM4 MMC MNH MNR MN5 MN6 MO1 MO2 MO3 MO4 MO5 MO6 MO7 MOO MOS MTD MW1 MW2 MW3
NA2 NA5 NA6 NAO NAW NBF NCO ND4 NFS NI1 NI2 NI3 NPB
O4M OC1 OC2 OC3 OC4 OC5 OC6 OC7 OC8 OCL OCM OCN OCO OF1 OF2 OF3
PBA PBC PBM PCL PEU PHF PHG PMB PPB
REO RHX RTC R4A R6A
SBD SBL SBO SE4 SES SF3 SF4 SIF SM2 SM3 SM4 SRB SRD SUF
T42 TAS
UNK UNX
V35 V36 V7O VA1 VO4
WCC WO4 WO5
XCC
YBT
ZO3

SMARTS Patterns

Including SMARTS

OC([C@@H]1[C@H]([C@@H]([C@@H](C(O1)O)O)O)O)

Excluding SMARTS

# mannose
OC([C@@H]1[C@H]([C@@H]([C@@H](C(O1)O)O)O)O)
C([C@@H]([C@@H]([C@H]([C@H](C=O)O)O)O)O)O

# galactose
OC([C@@H]1[C@@H]([C@@H]([C@H](C(O1)O)O)O)O)
C([C@@H]([C@H]([C@H]([C@@H](C=O)O)O)O)O)O

# glucose
OC([C@@H]1[C@H]([C@@H]([C@H](C(O1)O)O)O)O)
C([C@@H]([C@@H]([C@H]([C@@H](C=O)O)O)O)O)O

xylose
OC1[C@H]([C@@H]([C@H](C(O1)O)O)O)
C([C@@H]([C@H]([C@@H](C=O)O)O)O)O

N-acetylglucosamin
CC(=O)N[C@@H]1[C@H]([C@@H]([C@H](O[C@H]1O)CO)O)O

N-actetylgalactosamine
CC(=O)N[C@H]1[C@@H]([C@H](C(O[C@@H]1O)CO)O)O

N-acteylneruaminic acid
CC(=O)N[C@@H]1[C@H](CC(O[C@H]1[C@@H]([C@@H](CO)O)O)(C(=O)O)O)O

N-glycolylneruaminic acid
C1[C@@H]([C@H]([C@@H](OC1(C(=O)O)O)[C@@H]([C@@H](CO)O)O)NC(=O)CO)O

long linkers
[CH2;!R][CH2;!R][CH2;!R][CH2;!R]

non-organic elements
[!C;!c;!S;!s;!#1;!N;!n;!O;!o;!P;!p;!Cl;!F;!Br]