Electronic Supplementary Material (ESI) for Reaction Chemistry & Engineering.

Generating Molecules with Optimized Solubility using Iterative Graph Translation

Camille Bilodeau[1], Wengong Jin[2], Hongyun Xu[3], Jillian Emerson[3], Sukrit Mukhopadhyay[3], Tom Kalantar[3], Tommi Jaakkola[2], Regina Barzilay[2], Klavs F. Jensen (kfjensen@mit.edu)[1]*

[1]Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, United States

[2]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139, United States

[3]Dow Chemical Company, Midland, MI 48674, United States

*Corresponding Author

Supplementary Information

1. Iterative Graph-to-Graph Translation

All framework, architecture, and training parameters for the iterative graph-to-graph translation procedure can be found in Table 1. The code used for training can be found here: github.com/cbilodeau2/g2g_optimization.

Table S1. Framework, architecture, and training parameters for the iterative graph-to-graph translation.

Framework Parameters	
Cutoff Improvement, α	0.8 LogS
Number of Iterations	3
Cutoff SA Score	3.5
Training Parameters	
Batch Size	32
Number of CPU	9
Dropout	0.0
Learning Rate	1e-3
Clip Norm	2.0
Beta	0.3
Number of Epochs	10 (per iteration)

Anneal Rate	0.9
Architecture Parameters	
Activation	ReLU
Aggregation	Sum
RNN Type	LSTM
Hidden Size	270
Embedding Size	270
Latent Size	4
Tree Depth	20
Graph Depth	20
Tree Iterations	1
Graph Iterations	3

2. Directed Message Passing Neural Network (DMPNN)

All architecture and training parameters for the DMPNN can be found in Table 2. The resulting model for predicting solubility had R = 0.790, MAE =0.789, and RMSE = 1.038 (Figure 1).

Table S2. Architecture and training parameters for the DMPNN.

Training Parameters	
Batch Size	50
Dropout	0.0
Learning Rate	1e-3
Max Learning Rate	1e-2
Number of Epochs	100
Architecture Parameters	
Activation	ReLU
Aggregation	Mean
Ensemble Size	1
Depth	3
FFN Hidden Size	300
FFN Depth	2
Split Sizes	0.8, 0.1, 0.1
Split Type	Random

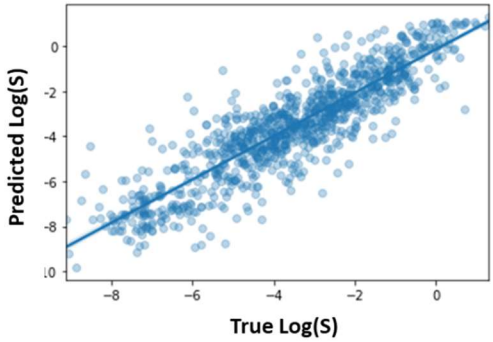

Figure S1. Comparison of the true and predicted Log solubilities of the test set molecules from AqSolDB.

Table S3. Source, catalog number, and physical properties for molecules used in solubility experiments.

Selected Molecules	Supplier	Catalog number	Molar Mass (g/mol)	Density (g/mL)
Dodecane	Fisher	117590250	170.33	0.75
Methyl butyrate	Sigma Aldrich	246093	102.13	0.90
4-Chlorobutyric acid	Sigma Aldrich	C29835	122.55	1.24
2-Heptanone	Sigma Aldrich	537683	114.19	0.82
N-Butoxymethyl acrylamide	Fisher	B106025ML	157.21	0.98
Ethyl Formate	Sigma Aldrich	112682	74.08	0.92
Octamethyltrisiloxane	Sigma Aldrich	O025725ML	236.53	0.82
Pinacolone	Sigma Aldrich	P45605	100.16	0.80
Dimethyl Glutarate	Fisher	G018525G	160.17	1.09
N-Methylformamide	Fisher	F005925G	59.07	1.01
Toluene	Fisher	T290-1	92.14	0.87
4-Fluoroaniline	Fisher	F3800	111.12	1.17

3. Aqueous solubility measurement by experiments

For each selected molecules, aqueous solutions of varies concentrations are prepared to determine the solubility of the molecule. Table S3 summarize the solubility behavior and Figure 2 & 3 display the image of each solution in Table 3.

Table S4. Solubility behavior of the aqueous solution with the selected molecules at each concentration.

Selected Molecules	**Target & Actual Concentration**															
	0.20%		**0.50%**		**1%**		**3%**		**7%**		**15%**		**34%**		**95%**	
Dodecane*	0.18%	X	0.44%	X	0.92%	X	3.11%	X	6.89%	X	15.77%	X	33.60%	X	94.93%	X
Methyl butyrate	0.17%	O	0.54%	O	1.14%	O	3.60%	X	7.14%	X	15.18%	X	33.52%	X	94.95%	X
4-Chlorobutyric acid	0.20%	O	0.70%	O	1.37%	O	4.07%	O	6.15%	O	13.78%	O	32.20%	X	94.89%	O
2-Heptanone*	0.17%	O	0.47%	O	0.92%	X	2.91%	X	7.27%	X	14.85%	X	33.25%	X	94.98%	X
N-Butoxymethyl acrylamide	0.16%	O	0.48%	O	0.96%	O	3.00%	X	7.27%	X	13.68%	X	34.68%	X	95.17%	X
Ethyl Formate	0.18%	O	0.30%	O	0.84%	O	2.73%	O	7.08%	O	14.94%	X	34.82%	X	95.18%	O
Octamethyl trisiloxane*	0.19%	X	0.41%	X	0.97%	X	2.93%	X	6.91%	X	14.82%	X	33.04%	X	94.96%	X
Pinacolone	0.14%	O	0.43%	O	0.90%	X	2.89%	X	6.80%	X	15.05%	X	34.24%	X	95.04%	X
Dimethyl Glutarate	0.21%	O	0.46%	O	0.98%	O	2.92%	O	6.49%	X	14.34%	X	34.05%	X	94.90%	X
N-Methyl formamide	0.20%	O	0.53%	O	0.96%	O	2.89%	O	6.79%	O	14.64%	O	33.93%	O	94.93%	O
Toluene*	0.13%	O	0.43%	X	0.88%	X	2.92%	X	6.99%	X	13.07%	X	34.12%	X	94.99%	X
4-Fluoroaniline	0.18%	O	0.49%	O	0.98%	O	2.84%	X	6.68%	X	14.47%	X	32.70%	X	94.92%	X

Note: “X” represents insoluble, where the solution is cloudy or phase separation. “O” represents soluble, where the solution is clear.

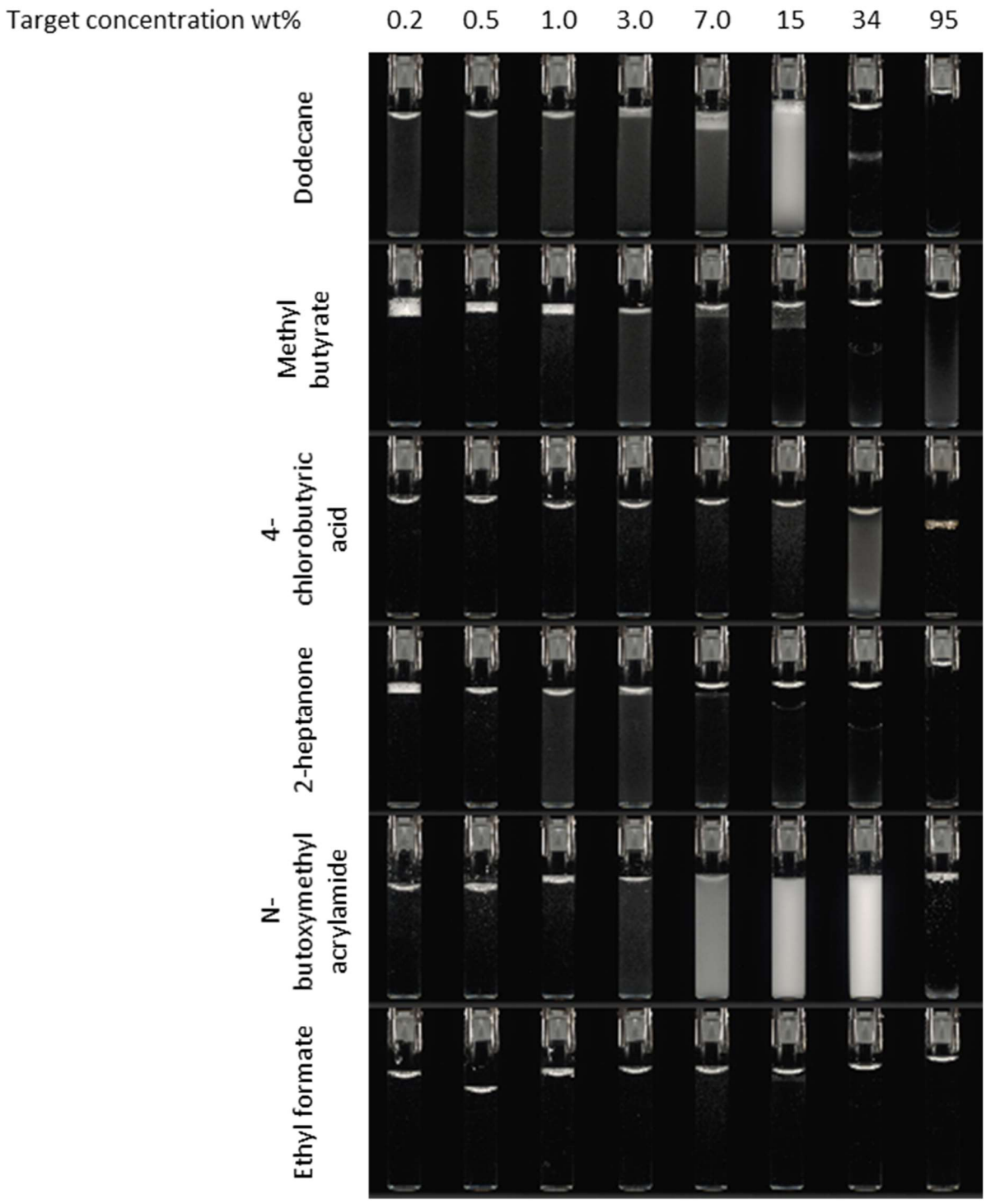

Figure S2. Images of first set of solvents in aqueous solutions.

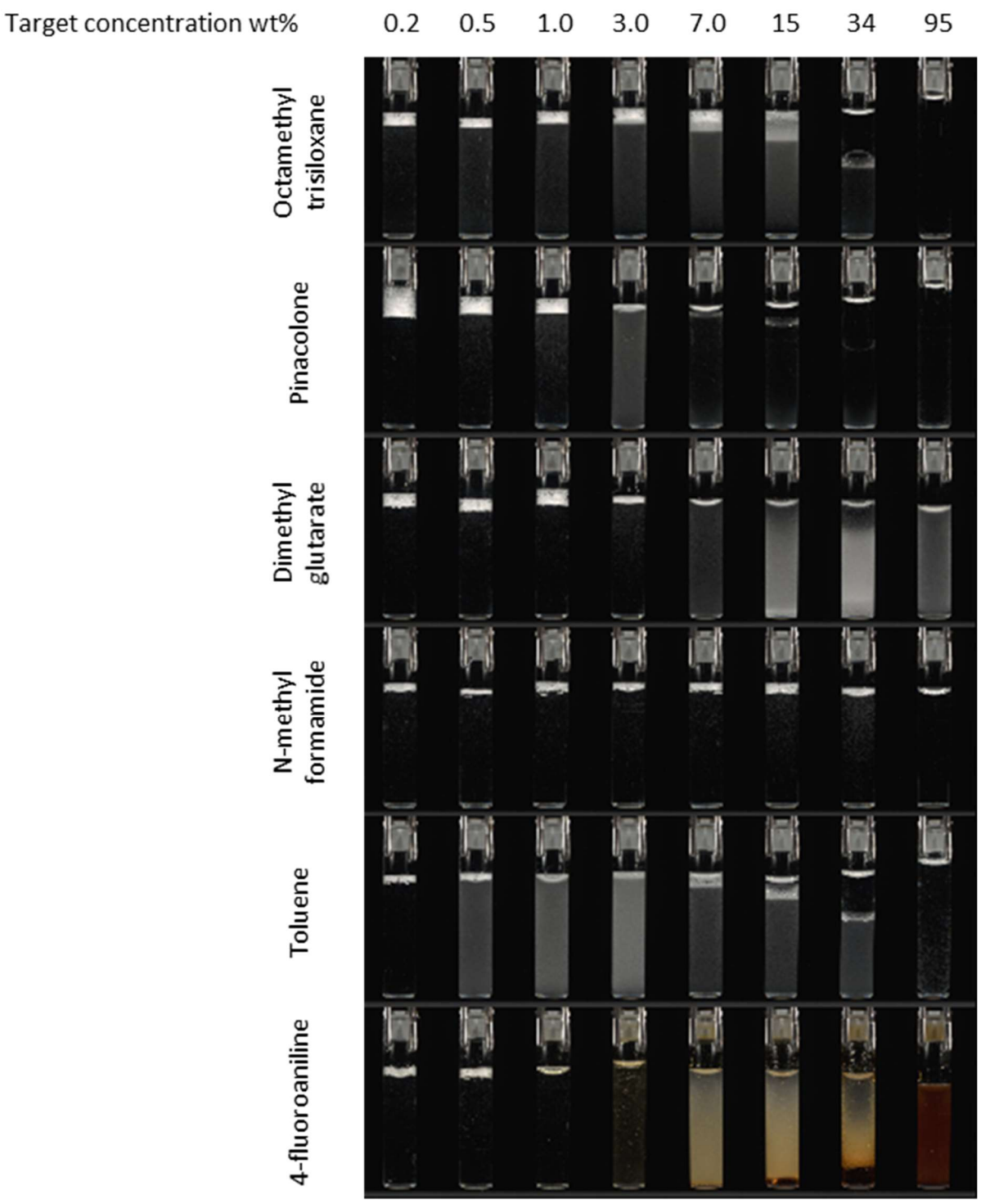

Figure S3. Images of second set of solvents in aqueous solutions.

Table S5. 100 randomly selected from the molecules generated with good SA score and good RA score.

SMILES	Tree Found
O=C(O)C1=Cc2ccc(C(=O)O)c(C(=O)O)c2C(=O)C1	TRUE
O=C(NCCCS(=O)(=O)O)c1cc(N=Nc2ccc(O)cc2)ccc1O	TRUE
CN(CC(O)COc1cccnc1)C(C)(C)N	FALSE
Cc1c(N)ccc(Nc2ccc(S(=O)(=O)O)cc2C(=O)O)c1S(=O)(=O)O	TRUE
O=CCc1ccc(S(=O)(=O)Nc2cccc3cc(S(=O)(=O)O)ccc23)cc1	TRUE
CCN(C)C(=O)CCN(C)C(=O)c1cccc(S(=O)(=O)Nc2ccc(S(=O)(=O)O)cc2)c1	TRUE
Cn1c(=O)n(CO)c(=O)c2ncn(CO)c21	FALSE
O=C1C=C2C=CC=CC2c2ccc(S(=O)(=O)O)cc21	FALSE
CCn1nc2c(=O)n(C)c(CC=O)nc2c1N	FALSE
NC(C(=O)O)C(=O)NC(C(=O)O)c1cccc(O)c1	FALSE
NCCCCCOCCOCCOCCO	TRUE
CCc1ccc(S(=O)(=O)O)c(O)c1	TRUE
CNCCCn1cnc2nc3cncnc3nc21	FALSE
CC(=O)NC1CCC(N)CC1	TRUE
NCNC(=O)NCNC(=O)Nc1ccccc1	TRUE
CC(C(=O)NCC(=O)O)c1ccc(Oc2ccccc2C(=O)O)cc1	TRUE
OCCNCc1nccs1	TRUE
O=C(O)c1cc(S(=O)(=O)O)c2c(c1NO)C(=O)CCC2=O	FALSE
CN(C)Cc1c(Cl)c(Cl)cc2c1C(=O)c1cc(O)cc(C(=O)O)c1C2=O	FALSE
O=CNCCOCCOCCOCCOCCOCn1cnc2c1c(=O)n(CCCO)c(=O)n2CCCO	FALSE
CCOC(=O)NC1CCC(O)CC1	TRUE
NCCCCC(O)C1(O)C(=O)NC(=O)NC1=O	FALSE
C=CCNC(=O)CCc1cccnc1	TRUE
CSc1cc(S(=O)(=O)O)cc2ccccc12	TRUE
Cc1cc(O)cc(C(=O)N(CCO)c2c(O)cc(O)cc2O)c1O	FALSE
O=CNCOC(=O)c1ccccc1	TRUE
Cc1cc(S(=O)(=O)O)c(O)c2nc(S(=O)(=O)O)sc12	TRUE
CN(C)CC(Cl)NCC(=O)O	FALSE
NNc1cc(S(=O)(=O)O)c2c(c1S(=O)(=O)O)C(=O)CCC2=O	FALSE
O=C(Oc1ccccc1N(CCO)CCO)c1ccccc1O	TRUE
Cn1c(=O)n(CCO)c(=O)c2c1ncn2CCC(=O)O	TRUE
NCc1nc2cncnc2n1CCCO	TRUE
N#Cc1ccc(O)c(N(O)C(=O)Nc2ccc(S(=O)(=O)O)cc2)c1	FALSE
Cc1ncc2nc(N)c(C)nc2n1	FALSE
CN(C)CCNC(=O)Nc1c(S(=O)(=O)O)cc(O)c2c1OCO2	FALSE
CC(=O)Nc1ccc(N)cc1CNC(=O)CCC(=O)NCOC(C)C(=O)OCC(=O)N(C)CCO	FALSE
Cc1cc(P(=O)(O)O)ccc1N(O)c1ccc(S(=O)(=O)O)cc1	FALSE
COCC1=CCCC1	TRUE
CN(C)C(=O)COC(=O)C(O)CO	TRUE
CC(CN)c1c(S(=O)(=O)O)c(O)cc2ccccc12	FALSE

COc1ccc(NCP(=O)(O)O)cc1N(C)c1ccccc1	TRUE
O=C(NCCO)Nc1ccc(O)c2c1-c1ccccc1C2CO	FALSE
CC(=O)N(C)C1CC=C(NC(=O)CNC(=O)CN)CC1	FALSE
O=C(O)c1cc(O)nc(O)n1	TRUE
N=Cc1cc2cc(S(=O)(=O)O)ccc2c2ccccc12	FALSE
CNC(=O)Oc1ccccc1OC(C=O)CO	FALSE
NCC(O)COC(=O)c1cccnc1	TRUE
CNC(=O)C(C)(C(=O)NCC(=O)O)C1CCC(O)CC1	FALSE
CNCCC(O)(COC)c1ccc(O)c(O)c1	FALSE
CNCCNCCNCCNC(=O)Cn1c(=O)ccn(C)c1=O	TRUE
O=C(O)c1c(C(=O)N(CCO)c2ccc(O)c(O)c2)ccc(O)c1O	FALSE
Nc1ccc(S(=O)(=O)NC(N)N)cc1	FALSE
CNCC(=O)OCNC(=O)c1ccco1	TRUE
Cc1cnn(-c2cccc(S(=O)(=O)Nc3ccc(S(=O)(=O)O)nc3)c2)c1N	FALSE
CCNCC1CO1	TRUE
Cc1cc(O)c(O)c(Cc2cc(O)c(O)c(O)c2)c1	TRUE
CNCC(O)COC(=O)c1cnc2ccccc2n1	TRUE
CNCC(=O)Nc1ccc2c(c1)C(=O)NC2=O	TRUE
O=P(O)(O)CSCCO	FALSE
CCNC(=O)OCC(=O)NCC(=O)OCCOC(=O)CN	TRUE
O=S(=O)(O)NO	FALSE
COCNC(=O)c1ccc2cccnc2c1	TRUE
O=Cc1cc(O)n(CCO)n1	FALSE
O=C(Nc1ccc(S(=O)(=O)NC(CO)NCCO)cc1)Nc1cccc(S(=O)(=O)O)c1	FALSE
CNC(=O)NCC(=O)OC	TRUE
CNC(=O)N(C)c1ccc(OC)cc1	TRUE
OCC(O)(COc1ccc(O)c(O)c1O)c1ccc(O)c(O)c1	FALSE
CCNc1cc(NCCO)n(CCO)n1	FALSE
NCCC(N)NC(=O)CCC(=O)O	FALSE
NC(=O)Sc1c(O)c(S(=O)(=O)O)cc2ncn(CO)c12	FALSE
O=C(NCCCCO)C(O)c1ccccc1	TRUE
CNCC(N)(C(=O)O)c1cccc2cccnc12	FALSE
CCOCCN(CCO)CCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCO	TRUE
Cc1c(S(=O)(=O)O)cc2ccc(S(=O)(=O)O)cc2c1N	FALSE
NC(NCC(=O)O)c1cccc(S(=O)(=O)Nc2ccc(S(N)(=O)=O)nc2)c1	FALSE
CC(=O)OCN(CCO)C(=O)C1CCCCC(O)C1	FALSE
O=C(O)COC1CCCCC1OCCO	FALSE
CC(=O)CCc1ccn[nH]1	FALSE
O=S(=O)(O)c1nc2ccccc2s1	TRUE
CC(C)NCC(O)COCNC(=O)c1ccco1	FALSE
COc1cc([N+](=O)[O-])ccc1S(=O)(=O)O	TRUE
NNc1cc(O)c(NC(=O)c2ccc(O)cc2)cc1S(=O)(=O)O	FALSE
N=C(N)NCc1ccc(O)c2c1-c1ccc(S(=O)(=O)O)cc1CC2	FALSE

N=Cc1cc(O)c(S(=O)(=O)O)cc1OCCOc1ccc(OCCO)cc1	FALSE
NCc1oc(N)cc1S(=O)(=O)O	FALSE
CN(C)C(=O)C(C)(O)c1cc(C(=O)O)ncn1	FALSE
CC(=O)CCC(=O)NC(CO)CO	TRUE
NCC(=O)NCC(=O)OCCOC(=O)CN	TRUE
NCCNc1ccc(S(=O)(=O)O)c(-c2ccc(S(=O)(=O)O)cc2)c1	TRUE
Nc1ccc(-c2ccc(Nc3ccc(S(=O)(=O)O)cc3)cc2)cc1	TRUE
CCOCCOC(C)(C)OCCO	TRUE
CC(C)(O)CC(N)c1ccc(Nc2ccc(S(=O)(=O)O)cc2)cc1	FALSE
O=P(O)(O)CCN(CCO)c1cc(CO)ncn1	TRUE
NCCCNCc1cc2c(cc1S(=O)(=O)O)CCC=C2S(=O)(=O)O	FALSE
O=C(NC(CO)C(=O)O)C(=O)S(=O)(=O)O	FALSE
NC(=O)Oc1ccco1	TRUE
CN(C)C(=N)Nc1ccc2cnn(-c3cccc(S(=O)(=O)O)c3)c2c1	FALSE
COC(=O)COc1ccc(N(C)C(=O)CO)cc1	TRUE
COC(CN(C)C(=O)NS(=O)(=O)c1ccc(N(C)C(=O)N(C)c2ccc(O)cc2)cc1)C(=O)O	FALSE
O=C(O)C(CO)NCc1c2c(cc(O)c1S(=O)(=O)O)-c1ccccc1C2	FALSE

Table S6. 100 randomly selected from the molecules generated with good SA score and bad RA score.

SMILES	Tree Found
NC(O)C(O)(c1ccccc1)c1cc(O)cc(O)c1O	FALSE
O=C(CO)Nc1cc(C(O)C(O)CO)cc(O)c1O	FALSE
NN(CCc1ccc(S(=O)(=O)O)cc1)CC1C(=O)C(=O)C(=O)C1=O	FALSE
CC(NC=Nc1ccc(OCCOC(=O)CNC(=O)CN)cc1OCCOC(=O)CN)C(=O)O	FALSE
CC(NC=Nc1ccc(Oc2ccc(Oc3ccccc3S(=O)(=O)O)cc2)cc1)C(=O)O	FALSE
O=CNCCCCOCCOCCOCCOCCOCCOCNC(=O)c1cc(O)cc(O)c1Nc1ccc(O)c(O)c1	FALSE
O=C(O)CC(O)(CC(=O)O)Cc1cc(C(=O)O)c2c(c1O)C(=O)OC2	FALSE
O=CNCCOCCOCCOCCOCCOCCOCCOCNC(=O)c1cc(S(=O)(=O)O)ncc1O	FALSE
CNC(N)c1cccc(S(=O)(=O)O)c1S(=O)(=O)c1ccc(O)cc1	FALSE
CNC(=O)OC(C(=O)N(C)C)c1cc2c(=O)c3ccccc3oc2cc1O	FALSE
O=S(=O)(O)C=Nc1cccc(S(=O)(=O)Nc2ccc(S(=O)(=O)O)nc2)c1O	FALSE
Nc1ccc(C(O)NC(N)N)c(-c2ccccc2)c1	FALSE
O=C1C(F)C(=O)N(CO)C(=O)N1CO	FALSE
NC(=O)NCCOCCOCCOC(=O)CCC(=O)NCOCCO	FALSE
CC(O)COc1ccc(O)c2cc(O)ccc12	FALSE
OCc1ccccc1-n1nccc1O	TRUE
Nc1c(I)c(O)c(C(=O)Nc2ccc(S(=O)(=O)O)cc2)c(O)c1S(=O)(=O)O	FALSE
Nc1cc(S(=O)(=O)O)c2c(c1C(=O)O)CC(O)CC2=O	FALSE
O=C1OCc2c(O)c(O)cc(OC(O)O)c21	FALSE
OC=C(C(O)O)C(O)(c1ccccc1)C(O)c1ccccc1	FALSE
CC(=O)Nc1ccc(O)c(NC(=O)c2cc(O)c(S(=O)(=O)O)cc2O)c1	TRUE
CN(C)C(=O)C(O)C(O)c1cc2c(cc1O)OCO2	TRUE

CNCOc1cc(OC(N)=O)cc(OCc2ccc(S(=O)(=O)O)cc2)c1O	FALSE
O=C(CCO)OCCOCCOCCOCn1cnc2ncn(CO)c21	FALSE
O=C(O)C(=O)CCCn1cnc2c1c(=O)n(CO)c(=O)n2CNC(CO)CO	FALSE
CCN(CC)C(N)NCc1ccc(O)c(C(=O)O)c1C(=O)O	FALSE
CN(C)CC(O)c1cc(O)cc(O)c1OC(=O)c1cc(O)c(O)c(O)c1	FALSE
CN(C)C(=O)C(O)c1cc(O)c2c(c1)OCO2	FALSE
NC(N)C1(O)OC(=O)c2ccccc21	FALSE
OCCC(O)C(O)C(O)c1cccnc1	FALSE
COc1c(S(=O)(=O)O)cc2c(c1O)CCC(O)=C2	FALSE
CCOCCOCCOCCOC(=O)C(C)(O)c1nccn1C	FALSE
Cc1c(N)c2c(c(O)c1S(=O)(=O)O)-c1ccccc1C(O)C2=O	FALSE
Cc1cc(CO)c(C(=O)O)c(P(=O)(O)O)c1C(=O)O	FALSE
Cc1c(N(C(=O)N(CCO)c2ccc(O)c(O)c2)c2ccc(O)c(P(=O)(O)O)c2)ccc(O)c1O	FALSE
CN(C)CC(O)c1cc(O)c(Oc2ccc(Oc3ccc(O)c(O)c3)cc2O)c(O)c1	FALSE
CCC(=O)NCCNC(=O)CCC(=O)Nc1ccc(O)cc1C(=O)OCCOCCOCCOCCOC(=O)CNC(=O)CN	FALSE
CNC(CO)COC(=O)C1Cc2ccccc2C1O	FALSE
COC1CC(S(=O)(=O)O)c2ccccc21	FALSE
O=CNCCCCOCCOCCOCCOCCOCCOCNC(=O)c1cc(S(=O)(=O)O)c(O)cc1O	FALSE
N=Cc1ccc(S(=O)(=O)O)c(NC(CC(=O)O)Cc2ccc(S(=O)(=O)O)cc2)c1	FALSE
CCN(CC)CCOCCOC1=CC(=O)C(=O)c2cc(O)ccc21	FALSE
COc1c2c(cc(O)c1P(=O)(O)O)COC2=O	FALSE
Cc1cc(N(C(=O)c2cc(O)cc(O)c2O)C(C)C(=O)O)cc(S(=O)(=O)O)c1	FALSE
O=C(O)C(c1ccc(S(=O)(=O)O)cc1)(c1cccc(O)c1)S(=O)(=O)O	FALSE
CC(N)(N)c1ccc(N)c(S(=O)(=O)c2ccc(N)cc2)c1	FALSE
O=S(=O)(O)c1nc(O)nc(Nc2c(S(=O)(=O)O)ccc(S(=O)(=O)O)c2O)n1	FALSE
O=S(=O)(O)c1cc(O)c(C(O)C(O)c2ccccc2)cc1O	FALSE
CCN(CC)CCOC(=O)COc1ccc(OCCOCCOCCOC(=O)CNC(=O)CN)c2c1C(=O)Cc1ccccc1-2	FALSE
COc1c(S(=O)(=O)O)c(O)cc2c1ncn2CN	FALSE
CN(C)C(=O)C(=O)Sc1cc(S(=O)(=O)O)cc(S(=O)(=O)O)c1O	FALSE
CNC(=O)NCc1ccc(OC)c2c1C(=O)Cc1ccccc1-2	FALSE
O=S(=O)(O)c1ccc(CO)c(C(O)C(O)CO)c1	FALSE
O=C(O)CCCNC(=O)C(O)Cc1ccc(O)c2ccccc12	FALSE
CNC(=O)OC(C(=O)N(C)C)c1ccc(C(=O)O)cc1O	FALSE
O=C1CC(C(=O)O)CC=C1S(=O)(=O)O	FALSE
NCCCn1c(O)nc2cc(S(=O)(=O)O)cc(S(=O)(=O)O)c21	FALSE
O=c1c2c(nc(F)n2C(CO)CO)n(CCO)c(=O)n1CCO	FALSE
COc1ccc(S(=O)(=O)O)c(O)c1NC(=O)N(CCO)c1ccc(O)c(P(=O)(O)O)c1	FALSE
CC(C)(O)C(=O)NCCCCCNC(=O)C1CC(C(=O)O)c2ccccc21	TRUE
O=C1NC(=O)C(NCCCNC(=O)C(=O)NCCO)C(=O)N1	TRUE
CNC(N)=NCc1ccc(S(=O)(=O)O)c2ccccc12	FALSE
O=CNCCCCOCCOCCOCCOCCOCCOCNC(=O)c1cc(O)cc(O)c1S(=O)(=O)Nc1ccc(O)cc1	FALSE
O=C(NO)NC(CO)(CO)COC(=O)OCCOCN1C(=O)CCC1=O	TRUE
Oc1cc(O)c2c(c1O)-c1ccccc1C(O)C2O	FALSE

CS(=O)(=O)c1ccc(C(O)C(CO)c2ccc(O)cc2)cc1O	FALSE
O=C(O)C1c2ccccc2N(CO)CC1O	FALSE
O=S(=O)(O)c1nc(COCCOCCOCCOCCOCCOCCOCCOCCO)nc(OCCOCCOCCOCCOCCO)n1	FALSE
CC(C)(O)C(=O)NCCCCCNC(=O)C(C)(O)CCCn1cnc2cnc(N)nc21	FALSE
O=C(O)c1c(O)ccc(C(O)C(=O)NCC(O)C(=O)O)c1C(=O)O	FALSE
Cc1cc(S(=O)(=O)O)c(C(=O)O)c(C(=O)c2ccsc2)c1O	FALSE
CN(C)C(=O)Cc1c(C(=O)O)cc(O)c(O)c1O	FALSE
O=C(O)CC1CNCCC1O	TRUE
O=P(O)(O)CNc1cc(O)c(S(=O)(=O)O)cc1O	TRUE
COc1cc(CS(=O)(=O)O)c2c(c1O)C(=O)OC2	FALSE
CN(C)C(=O)C(O)c1cc(O)c(C(=O)c2ccccc2)c(O)c1O	FALSE
O=S(=O)(O)c1ccc(S(=O)(=O)C(c2ccccc2)S(=O)(=O)O)cc1	FALSE
O=C(O)CNC(=O)NCNC(=O)NCn1ccnc1NCCO	FALSE
O=C(O)Oc1c(O)c(O)c(S(=O)(=O)O)c(O)c1O	FALSE
O=C1Cc2ccccc2-c2c1ccc(O)c2N(CCO)C(CO)C(=O)O	FALSE
O=C1OCc2c1cc(CO)c(P(=O)(O)O)c2S(=O)(=O)O	FALSE
CCN(CC)C(=O)c1ccc(O)cc1C(=O)NCOCCOCCOCCOC(=O)CNC(=O)CN	FALSE
O=S(=O)(O)c1ccc(C(O)C(O)CO)c(C(O)c2ccccc2)c1	FALSE
OC(O)(O)C(O)(O)c1ccco1	FALSE
O=NC(=O)c1cc(O)c(S(=O)(=O)O)c2c1CNC2=O	FALSE
O=C(NCO)C(O)c1ccc2c(ncn2CO)c1O	FALSE
Oc1ccc(COc2ccc(O)c(C(O)O)c2)cc1O	FALSE
O=C(O)COCCOCCOCCOCCOCCOCCOCCOCCOc1nc(OCCO)nc(OCCO)n1	TRUE
O=C(O)CN(CC(=O)O)c1cc(N(CCO)CCO)cc2c1C(=O)CCC2=O	FALSE
NN(C(=O)C(CO)CO)C1=CC(=O)CCC1	FALSE
CSc1nc(NCS(=O)(=O)O)nc(S(=O)(=O)O)n1	FALSE
NC(N)=NCc1cc(O)cc(Oc2ccc(S(=O)(=O)O)cc2)c1	TRUE
NC(N)C(O)(c1cccc(OCC(=O)O)c1)c1cc(O)cc(O)c1C(=O)O	FALSE
NC(N)Nc1c(O)c(S(=O)(=O)O)cc2c1CCC=C2	FALSE
O=P(O)(O)CN(CCCCCCN(CCO)CCO)CCCCCCn1cnc2ccc(O)cc21	TRUE
CN(c1ccccc1)c1ccc(S(=O)(=O)O)cc1N=NC(N)=O	FALSE
O=C(O)Cc1ccc2c(c1S(=O)(=O)O)CC(O)O2	FALSE
O=C(O)CCNCC1CCC(NCCO)CC1O	FALSE
O=C(O)CCCNC(=O)C(O)c1ccn[nH]1	TRUE
O=S(=O)(O)c1c(C(O)O)cc(O)c2c1Cc1ccccc1O2	FALSE

Table S7. 100 randomly selected from the molecules generated with bad SA score and good RA score.

SMILES	Tree Found
NCNCC1CC=C(N)CC1	FALSE
CCP(=O)(O)C1CCCCC1	FALSE
CNP(N)(=O)OCCOC(=O)CN	FALSE
OCCNC(CO)n1ccnc1	FALSE

C=C(CO)c1nc(S(=O)(=O)O)nc(S(=O)(=O)O)n1	FALSE
NNS(=O)(=O)C(O)C1CCCO1	FALSE
NNc1cc(S(=O)(=O)O)cc2c1OC(S(=O)(=O)O)C2	FALSE
CN(C)CNCOP(=O)(O)Oc1cccc2occc12	FALSE
CC(O)n1cnc2ncncc21	FALSE
CN(CN)c1cnn(Cn2nccc2C#N)c1	FALSE
NC1CCC(O)(C(O)NCCO)CC1	FALSE
O=[PH](O)Oc1ccc(S(=O)(=O)O)cc1O	FALSE
CN(N)c1nc2nc[nH]c2n1C	FALSE
CC(O)CNCC(O)COC1=CCC(C(=O)O)CC1	TRUE
O=S(=O)(O)CC(O)Br	FALSE
NCCCNc1c(P(=O)(O)O)cc2c(c1O)CC(C(=O)O)C2	FALSE
CSC1C=Cc2cc(S(=O)(=O)O)ccc2C1	FALSE
O=CNCCOCCOCCOCCOCCOCCOCCOCNC(=O)C1CO1	FALSE
NC1CCc2c(S(=O)(=O)O)cc(NC=O)c(S(=O)(=O)O)c2C1=O	FALSE
NC(N)Nc1ncc(S(=O)(=O)O)[nH]1	FALSE
O=P(O)(Nc1ccc(S(=O)(=O)O)nc1)c1cccnc1	FALSE
O=C(COC(=O)CNC(=O)COC(=O)NCOCn1cc(CO)nc1S(=O)(=O)O)NCCCS(=O)(=O)[O-]	FALSE
O=CNCCCCOCCOCCOCCOCCOCCOCNC(=O)COCCOC(=O)c1cc(O)c2c(c1O)CC(C(=O)O)C2	FALSE
CCC(C)NCC(O)COCNC(=O)C1CCCC1O	FALSE
COCC1OC1C1CO1	FALSE
O=c1c2c(ncn2COCC(O)CO)n(COCCO)c(=O)n1CO	FALSE
NC(CO)(CO)c1nc(O)nc(O)n1	FALSE
COC(N)C1CCCO1	FALSE
COC(C)(CNC(=O)C(C)(C)O)NC(=O)c1cnc[nH]1	FALSE
CNC(N)(N)C(C)C	FALSE
CC(=NO)C1CCCC(N)C1	FALSE
CNC(=O)CC(C)(O)c1nc(O)nc(N(C)C)n1	FALSE
CN(C)CNc1ccc(N2CC2S(=O)(=O)O)cc1	FALSE
OCCOCNC(NCO)n1ccnc1	FALSE
NCNc1oc(S(=O)(=O)O)cc1S(=O)(=O)O	FALSE
NCNC(N)NC1CCC(N)CC1	FALSE
O=CNCCNC(=O)NCOCCOCNC1CCCCC1NCCO	FALSE
CNC(CC(=O)O)C1CCC(O)C1	FALSE
CCC(=O)NCCNC(=O)CCC(=O)N(C)CC(=O)OCCOC1CCC(COCCOC(=O)CNC(=O)CN)C1	FALSE
CN(C)C(=O)C(O)c1cc(O)c(O)c(C(=O)CO)c1O	FALSE
C=CC(C)(O)C1=CC=C(O)CC1	FALSE
CN(C)C1CC=C(NC=O)CC1	FALSE
O=P(O)(O)CNCC1CCCC(O)C1	TRUE
CNC(=O)C(O)c1c[nH]ccc1=O	TRUE
CC(c1ccn[nH]1)C(C)(C)N	FALSE
CNC(N)NC(=O)Nc1nccc(S(=O)(=O)O)n1	FALSE
Nc1ccc(P(=O)(O)O)o1	FALSE

CNC(=O)C(O)Cn1cncc1OC	TRUE
CCC(C)(O)CC(N)C(N)=O	FALSE
NC(=O)C(N)(NCCCC(=O)O)c1ccnn1C(=O)[O-]	FALSE
NC(N)OC1=c2ccccc2=C(S(=O)(=O)O)CC1	FALSE
CCOP(=O)(OCC)OCOCN(C)CCOCCOCCOCCOCCOc1cnn(CO)c1	FALSE
O=[SH]CC(CO)(CC(=O)O)c1c(O)cc(C(=O)O)c2ccccc12	FALSE
OCNc1ncc2ncn(C(Cl)(CO)CO)c2n1	FALSE
O=S(=O)(O)C1=NC(S(=O)(=O)[O-])CN1CCO	FALSE
CN(CC(=O)O)C(N)Nc1ccc(O)c(NC(=O)CNC(=O)CN)c1N=Nc1ccc(O)cc1	FALSE
O=C(NCO)Nc1cc(N2CC2S(=O)(=O)O)cc(S(=O)(=O)O)c1	FALSE
O=[N+]([O-])C(O)CO	TRUE
CCNC1(N)CCC(S(=O)(=O)O)c2ccccc21	FALSE
NC(O)NCCO	FALSE
O=C(NO)NC1=CCC(S(=O)(=O)O)C1	FALSE
OCCNC(CO)NC1=CCC(CO)CC1	FALSE
N#CCNCN	FALSE
COC(=O)CN(C)C(=O)CNC(=O)N(C)CC(=O)OCCN(C)C(=O)CNC(=O)N(C)CC(=O)OCCNC(=O)C	
NC(=O)COC(=O)c1cc(O)cc2oc(=O)ccc12	FALSE
O=C(O)CC(O)(C(=O)O)C1C=CC(=O)c2ccccc21	FALSE
COC(=O)C(O)Cc1nc(O)nc(O)n1	FALSE
CNCCC(O)C1CCC(C(=O)OC)C1	TRUE
OCC(O)OCOC(O)(c1ccccc1)c1ccc(O)cc1	FALSE
CCC(O)C(N)C(C)(C)O	TRUE
NCCNC(N)NCc1ccc(S(=O)(=O)O)c2c1C(=O)OC2	FALSE
CC(C)NCC(O)COc1ccc(NC(=O)NCNC(=O)CN)c2nc(S(=O)(=O)O)sc12	FALSE
CP(=O)(O)c1cc2c(cc1O)C=C(S(=O)(=O)O)CC2	FALSE
O=[SH]Cc1nc2c(ncn2CO)n1CCO	FALSE
Cc1nc(O)nc(C(C)C(=O)N(C)C)n1	FALSE
CC(O)(CNC(=O)CNC(=O)CN)C(O)(c1cc(O)cc(O)c1)c1cc(O)cc(O)c1O	FALSE
O=C(NCO)N(CO)c1ncc[nH]1	FALSE
O=C(O)C(O)(C(=O)NCO)c1cc2scnc2cc1O	FALSE
COC(=O)C(NC(=O)C1=CCCC(O)C1)C(=O)O	FALSE
NC(=O)NC(N)(CO)NC(=O)C1CCC(O)CC1	FALSE
NCC(O)Cc1c[nH]ccc1=O	FALSE
O=C1CC(=O)N(O)CN1	FALSE
O=C(Br)N(O)C(CO)CO	FALSE
OCC(O)SC1CCC(O)CC1	FALSE
CNC(N)NN	FALSE
CNC(N)n1nccc1N	FALSE
CNCC(O)c1nc2cncnc2[nH]1	TRUE
CC(=N)C1CCCC(N)C1	FALSE
Cc1nc2ncn(CCO)c2n1CC(C)(O)CN	FALSE
NCn1nc(S(=O)(=O)O)cc1S(=O)(=O)O	FALSE

OCCN(O)c1nc(O)nc(O)n1	FALSE
O=C(O)CC1(C(=O)NCCO)CCCC1C(=O)O	FALSE
O=P(O)(O)c1c(NCO)cc2ncn(CO)c2c1O	FALSE
CO[PH](O)(O)OCCN(C)C(C)=O	FALSE
N#CCc1nc(O)nc(O)n1	FALSE
CC(NC(=O)C([O-])C(=O)O)C(=O)O	FALSE
CN(C)CC(O)c1c[nH]c(S(=O)(=O)O)cc1=O	FALSE
NC(N)Nc1nc2[nH]cnc2[nH]1	FALSE
CN(C)CNC1C=Cc2cc(S(=O)(=O)O)ccc2C1	FALSE
Cn1c(=O)n(CO)c(=O)c2c(S(=O)(=O)O)ccc(NC(N)NCN)c21	FALSE
NC(C(=O)[O-])C(=O)O	TRUE

Table S8. 100 randomly selected from the molecules generated with bad SA score and bad RA score.

SMILES	Tree Found
COC(=O)CN(C(=O)CO)C(=O)COC(=O)c1ccc(OC(=O)OCC(=O)N(C)C)c(OC(=O)C(O)CO)c1	FALSE
O=S(=O)(O)C1=c2ccccc2=CCC1(CO)NCCO	FALSE
CN(C)CC(O)C1=CCC(NC(=O)CO)C1	FALSE
N=C(N)NC1CC(C(=O)O)C(C(=O)O)C1	TRUE
CS(=O)(=O)C1=CCC(N)C(C(=O)O)C1	FALSE
C=C(O)Oc1nc(O)nc(O)n1	FALSE
O=P(O)(O)c1cc(P(=O)(O)O)c2c(c1O)OCO2	FALSE
O=C1C(O)OC(O)(O)C1O	FALSE
O=CNC1(CO)CCCCC1O	FALSE
OCCCC(O)c1ncc(O)n1CO	FALSE
CC(O)C(N)C(O)c1cc(O)cc2c1C(O)Cc1ccccc1-2	FALSE
O=C(N(CO)CCO)N(CCO)C1CCC(O)C1	FALSE
CN(CC(O)COC(=O)CN)CC(O)COC(=O)CN(C)C(=O)COC(=O)CN	FALSE
OCCOCCOCC(O)(CO)COCCOCCOC1=C(O)CCC1	FALSE
CC(=O)NC1=CCC(O)CC1C(=O)O	FALSE
O=S(=O)(O)CCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOc1nc(OCCO)nc(N(CCO)CCO)n1	FALSE
CNCN1C(=O)C(N)N=C1N	FALSE
NC(N)C(O)C(O)C(O)C(O)C(=O)O	FALSE
OC(O)NC(O)n1ccnc1	FALSE
Nc1ccc(S(=O)(=O)O)c2c1CC(O)C2C(=O)O	FALSE
O=CNCCOCCOCCOCCOCNC(=O)COCCOCNC(=O)COCOc1c(O)cc(O)c2ncsc12	FALSE
COCN1N=C(C)C(CN(C)C(C)=O)C1O	FALSE
CCCC(N)C(N)C(O)CC	FALSE
O=CNCCCCOCCOCCOCCOCCOCCOCCOCNC(=O)COCCOC(=O)C1=C(O)C(O)C(O)C(O)C1O	FALSE
OCCOCCOCCOCCOCCOC(O)COCCOC(O)COc1cc(O)cs1	FALSE
O=C1NC(=O)N(CO)C(=O)C1[O-]	FALSE
NNC1=CC(OP(=O)(O)O)=C(OP(=O)(O)O)C2C=CC=CC12	FALSE

O=P(O)(O)OCCOCCOCCOCOCOCOCOCCOCCOCCOCCOCCOC1=C(O)CCC1	FALSE
CC1c2cc(S(=O)(=O)O)ccc2N(C)CC1N	FALSE
O=C(NCCO)N1C(=O)N(CO)C(CO)=C2C1N=CN2CO	FALSE
O=C(NCC(O)C(O)C(O)C(O)CO)c1cc(O)c(-c2ccccc2)c(CO)c1O	FALSE
CC1=NC(N)=C(C(=O)NC(C)(C)C(=O)O)C(C(=O)O)C1O	FALSE
O=S(=O)(O)c1ccc(C([O-])O)c(Nc2nc(NCCO)nc(OCCO)n2)c1O	FALSE
CNCN1C(=O)CC(S(=O)(=O)O)C1=O	FALSE
O=P(O)(O)c1cc(O)c(N2C(O)=NC(O)C2O)c2scnc12	FALSE
O=S(=O)(O)c1cc(O)c2c(c1C(O)(O)C(O)CO)-c1ccccc1CC2	FALSE
CNC1CC(C(=O)O)C(C(=O)O)CC1C(=O)O	FALSE
NC(=O)c1cc(O)c(S(=O)(=O)O)c2c1C(=O)C(O)CC2=O	FALSE
CN(C)C(=O)C(O)c1cc(O)c(O)c(C(=O)O)c1C(N)(C(=O)O)c1ccccc1	FALSE
NC(CC1CNCC(N)C1N)C(=O)O	FALSE
O=P(O)(O)CCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOCCOC1 =CC=C(O)CC1	FALSE
COC(=O)C(NC(=O)C(O)CC(=O)O)c1nc(O)nc(O)n1	FALSE
O=CNc1nc(Nc2nc(NCCO)nc(S(=O)(=O)O)n2)nc(S(=O)(=O)O)n1	FALSE
CC(C(=O)n1c(=O)cc[nH]c1=O)C(C)(C(=O)N(C)C)C(=O)N(C)C	FALSE
CCN(CC)C(C)CNC(=O)c1cc(O)cc(N2CC2O)c1C(=O)O	FALSE
CC1=CC(=O)C(O)C(C)(C(=O)NCC(=O)O)C1	FALSE
O=C(O)C1=C(C(=O)O)C(CO)C(O)(C(=O)O)C(CO)=C1	FALSE
CNCC(COc1ccc(N2CC2S(=O)(=O)O)cc1)NC(=O)CO	FALSE
O=CNCCOCCOCCOCCOCCOCCOCCOCNC(=O)COCOc1nc(OCCO)nc(Oc2nc(O)nc(O)n2)n1	FALSE
O=C(NC(CO)CO)C1=C(S(=O)(=O)O)NC(O)=CC1	FALSE
O=CCCN(CCO)C(CO)(CO)COCCOCCOCCOCOCCOCCOCCOCCOCCOCCOCCOCCOCCO CCOCNCCO	FALSE
CNC(N)=Nc1cc(S(=O)(=O)O)c(N)c2c1CCC(N)C2	FALSE
CNC(N)NC(N)C1=CC(=O)Cc2ccccc21	FALSE
COC(O)CC1=CNC(OCCO)=C(OCCO)C1	FALSE
CC(N)(CNC(N)=O)c1cc(O)cc2c1OCC(S(=O)(=O)O)=C2	FALSE
O=S(=O)(O)C=Nc1cnc(S(=O)(=O)O)cc1S(=O)(=O)O	FALSE
O=S(=O)([O-])c1c(CCCO)cc(O)c2ncn(C(O)CO)c12	FALSE
O=C1C(=O)C2(O)C=C(O)C=CC2N1CCCO	FALSE
CNC(C)C1C(O)CC(N)CN1C	FALSE
COC(O)c1cc(Oc2ccc(Oc3ccc(O)c(C(=O)NCNC(=O)CN)c3)cc2)ccc1Oc1ccc(O)c(C(=O)NCC(=O)O)c1	FALSE
O=C(O)C(=O)c1cc(O)c(N2CC2S(=O)(=O)O)c(O)c1	FALSE
CC(C)(O)NC(N)C(O)c1cc(O)cc(O)c1	FALSE
O=S(=O)(O)CCSC(CO)C(O)CO	FALSE
O=C(NC(C(O)CO)C(O)C(O)C(O)CO)c1ccccc1	FALSE
O=C(O)C(CO)n1nc(O)c(O)c1S(=O)(=O)O	FALSE
OCc1cc(O)c2c(c1O)CC(O)C(O)C2	TRUE
CCCNC(=O)C(C)(O)c1nc(O)nc(O)n1	FALSE

O=C1C(F)C(=O)N(COCCO)C(=O)N1CO	FALSE
CC(C)(CNCC(=O)O)NCC(O)COC(=O)Cn1ncc2c(=O)[nH]cnc21	FALSE
CCCCCCOc1ccc(OCCOCCOCCOCCOCCOc2cc3c(cc2S(=O)(=O)O)-c2ccccc2C(O)C3)cc1S(=O)(=O)O	FALSE
[NH3+]C(C(=O)NC(CO)C(=O)O)c1c(O)cnc(O)c1O	FALSE
COCN1CC(=O)N=C1C(C)O	FALSE
OC(O)C(O)C1=CCCC1	FALSE
CN(C)CCc1cc(O)c(O)c2c1C(CC=O)CC(O)C2	FALSE
O=C(O)CC(c1ccc(S(=O)(=O)O)cc1)C(c1ccccc1)C(Cc1ccc(S(=O)(=O)O)cc1)c1ccccc1P(=O)(O)O	FALSE
O=C(NCCO)N1C(=O)N(CO)C(NCCO)=C2C1N=CN2CCO	FALSE
CCN(C)CC1CC(=O)N(C)C1N	FALSE
O=C(O)Oc1cc2c(cc1O)C(=O)C1C=CC=CC1C2=O	FALSE
OCn1nc(O)c(O)c1O	FALSE
CNc1ccc2c(c1S(=O)(=O)O)C(C)C(O)C2N	FALSE
O=P(O)(O)c1ccc(S(=O)(=O)O)cc1NC(Cc1ccc(S(=O)(=O)O)cc1)C(CO)c1ccc(O)cc1	FALSE
O=CNCCOCCOCCOCCOCCOCCOCCOCNC1c2ccccc2CC(O)C1O	FALSE
N=C(N)Sc1cc(S(=O)(=O)O)cc2c1C(S(=O)(=O)O)CC(S(=O)(=O)O)C2	FALSE
OCOc1cc2c(c(OCO)c1OCO)OC(O)C2	FALSE
O=C(O)CC(NC(=O)C(O)(CC(=O)O)Cc1ccn[nH]1)C(=O)O	FALSE
O=C1Cc2c(c(S(=O)(=O)O)cc(O)c2P(=O)(O)O)C=C1O	FALSE
O=C1C2C=CC=CC2C(=O)N1CCCO	FALSE
OCCN(Cc1ccc(O)c(O)c1)C(O)C(O)c1c(O)cc(O)cc1O	FALSE
O=CNCCOCCOCCOCCOCNC(=O)COCCOCNC(=O)COCOc1cc(O)c(O)c2scnc12	FALSE
O=COc1cc(OC=O)c(OC(O)C(O)(O)O)c(O)c1O	FALSE
NC1=C2C=CC(S(=O)(=O)O)=CC2CC(S(=O)(=O)O)=C1O	FALSE
COC(N)C(N)C(O)c1cc(O)cc(S(=O)(=O)Nc2ccc(O)cc2)c1	FALSE
COC(O)OCn1cnc2cnc(O)nc21	FALSE
O=C(NC(O)C(=O)O)SC1CCC(S(=O)(=O)O)CC1	FALSE
O=C1c2c(ccc(P(=O)(O)O)c2CO)CCC1S(=O)(=O)O	FALSE
O=C1C(O)C(O)(C(=O)O)OC1S(=O)(=O)O	FALSE
CC1NCC(CC(N)=O)C1O	FALSE
CC(=O)n1nc(S(=O)(=O)O)c2c1C(=O)N(C)CN2O	FALSE
O=c1cc[nH]c(=O)n1C(O)CO	FALSE
CC(N)CC1CC(C(=O)O)C(C(=O)O)C1	FALSE