Detailed validation dataset results

This document provides details on the framework’s performance in predicting for the substructure probabilities and the ranking of candidate molecular structures for each molecule in our validation set, as well as the experimental $^1$H and $^{13}$C NMR spectra used as input. The data is ordered by the number of possible constitutional isomers generated by OMG using the molecular formula (SI section 3.5), starting with the largest. For each molecule, the SMILES string and molecular formula corresponding to the true molecular structure are listed first.

“Index of correct structure” denotes the rank of the correct structure as predicted for by our framework, with an index of “0” meaning that the correct structure was ranked as the most likely structure by our framework, “1” meaning the correct structure was the second ranked structure, and so on. An index of “-1” indicates that the correct structure was not generated as part of the ranked list of structures predicted for by our molecular graph generation algorithm. This index is reported alongside the total number of possible constitutional isomers for this molecular formula as generated by OMG.

“True structure loss” is the binary cross-entropy (BCE) loss between the predicted substructure probabilities and the correct structure substructure labels (SI section 3.5). A lower loss indicates better agreement between the ML-predicted substructure probabilities and the true substructure label, with 0 being perfect agreement.

The true structure is shown next along with the experimental $^{13}$C and $^1$H NMR spectra. The solvent in which the NMR spectra was collected in is noted in parentheses when available.

The 10 structures with the lowest BCE loss generated by our framework, sorted from lowest (most likely) to highest, are shown next. For each of the top 10 predicted molecular structures, the BCE loss between its ML-predicted substructure probabilities and the true structure’s substructure label is displayed below an image of the predicted molecule’s 2D structure.

For each molecule, additional details for the substructure predictions are shown on the following page. First, the top 10 highest probability substructures predicted for by our ML model are shown as SMARTS strings alongside their respective predicted probabilities. Next, the 10 highest probability true positive substructures (“best positives”) are shown. These are the substructures present in the true structure that the model predicted to be present with a high probability. Substructures listed under “best negatives”, “worst negatives”, and “worst positives” correspond to the lowest probability true negatives, highest probability false positives, and lowest probability false negatives, respectively.
Example 0 true smiles: N=C(N)NCCCC(=O)O formula: C5H11N3O2
Index of correct structure: 8 of 677501
True structure loss: 0.030339
True structure:

Experimental $^{13}$C NMR (solvent: D2O)

Experimental $^1$H NMR (solvent: D2O)

Top predicted structures (loss):

- 0.021364
- 0.024313
- 0.02474
- 0.025733
- 0.028428
- 0.029822
- 0.030128
- 0.030274
- 0.030339
- 0.030498
| Substructures                        | prob  | Substructures                        | prob  |
|--------------------------------------|-------|--------------------------------------|-------|
| [CX4H2][{#6}][#6]                   | 0.9972| [#7X3][#6H2]                         | 0.9146|
| [#7X3H2]                            | 0.9838| [OX2H1]                              | 0.8938|
| [CX3](=OX1)C                        | 0.9587| [CX4H2][CX4H2]                       | 0.8604|
| [#7][#6H2][#6H2]                    | 0.9538| [CX4H2]C=C=O                         | 0.841 |
| [#7][#6H2]                          | 0.9306| [CX4H2][{CX4H2}][CX3H0]              | 0.7779|

**Best positives**

| Substructures                        | prob  | Substructures                        | prob  |
|--------------------------------------|-------|--------------------------------------|-------|
| [CX4H2][{#6}][#6]                   | 0.9972| C=C=CC=C=C                           | 0.0   |
| [#7X3H2]                            | 0.9838| [CX2H1][CX2H0][CX3H1]=[CX3H0]       | 0.0   |
| [CX3](=OX1)C                        | 0.9587| [CX2H0][=OX2H0][CX2H0]              | 0.0   |
| [#7][#6H2][#6H2]                    | 0.9538| [CX3H0]([CX4H2])[CX2H0]              | 0.0   |
| [#7][#6H2]                          | 0.9306| [CX2H0][=OX2H0][CX3H0]              | 0.0   |
| [OX2H1]                             | 0.8938| CC#CCC=C                             | 0.0   |
| [CX4H2][CX4H2]                      | 0.8604| C=C=CCCC=C                           | 0.0   |
| [CX4H2]C=C=O                        | 0.841 | [CX2H0][=OX2H0][CX4H0]              | 0.0   |
| [CX4H2][{CX4H2}][CX3H0]             | 0.7779| C=C=CCCC=C                           | 0.0   |

**Worst negatives**

| Substructures                        | prob  | Substructures                        | prob  |
|--------------------------------------|-------|--------------------------------------|-------|
| [CX4H2][{NX3H2}][CX4H2]             | 0.5724| [OX1H0]=[OX1H0][{#8}][CX4H2]        | 0.0418|
| [#7H2][#6H2]                        | 0.5509| [#7][#6][#7]=[#7]                   | 0.0897|
| [#7][#6][#6][#6X3]                  | 0.3595| [#8][=#6][#6H2]                     | 0.1089|
| [#7][#6][#6][#6][#7]                | 0.2792| [CX3H0][=OX1H0][{OX2H1}][CX4H2]    | 0.134 |
| [#7][#6][=OX2H0][#6][#6]            | 0.272 | [#8][=#6][#8]                        | 0.2531|
| [#7][#6][#6][#6][#7]                | 0.2362| [CX3][=OX1]D                         | 0.2953|
| [#8][=#6H0][#6H1]                   | 0.1883| OCC[CH2]                             | 0.3515|
| #6H1                                | 0.1817| [#7][#6H0]                           | 0.5193|
| #7X3H0                              | 0.1753| [#7X3H1]                             | 0.5285|
| [#7][#6][#6X3]                      | 0.1204| [CX4H2][{NX3H1}][CX4H2]             | 0.5364|
Example 1 true smiles: Cn1ccc(=O)n(C)c1=O formula: C6H8N2O2
Index of correct structure: 0 of 623393
True structure loss: 0.031526
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

- 0.031526
- 0.03894
- 0.041776
- 0.042338
- 0.044092
- 0.0446
- 0.045601
- 0.047015
- 0.047597
- 0.047685
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [#7X3][#6H3]               | 0.9986 | [#6H3][#7][#6X3] | 0.8865 | [#7X3][#6H3] | 0.766 |
| [#6H1]                      | 0.988  | [#6X3][#7][#6X3] | 0.8583 | [#6H1] | 0.58 |
| [#6H3][#7]                  | 0.969  | [#7][#6][#6X3]  | 0.8576 | [#6H3][#7] | 0.55 |
| [CX4H3]                     | 0.9559 | [#7][#6][#6][#6X3] | 0.8361 | [CX4H3] | 0.5 |
| [#6X3][#6X3]                | 0.9119 | [#7][#6][#6X3]  | 0.8361 | [#6X3][#6X3] | 0.5 |

| worst negatives | prob | worst positives | prob |
|-----------------|------|-----------------|------|
| [CX4H3][NX3H0]        | 0.5633 | [cX3H1][#nX3H0][cX3H1] | 0.1523 |
| [#X][#6][#6X3] | 0.4202 | [#7][#6][#7] | 0.2353 |
| [#8][#6][#6X3] | 0.4273 | [#7][#6][#6][#7] | 0.2377 |
| [OX2H2][#6][#6] | 0.3802 | [OX2H0][#cX3H0][cX3H1] | 0.2437 |
| [#6X3][#6H2][#6X3] | 0.367 | [#8][#6][#6H1][#6H1] | 0.259 |
| [#8][#6][#6H2] | 0.3665 | [#8][#6][#6H1][#6H1] | 0.3397 |
| [OX2H1] | 0.3365 | O=[eX3] | 0.4465 |
| [#6H1][#6H2] | 0.3315 | [#7][#6X3H0][#6X3H1] | 0.4526 |
Example 2 true smiles: CNC(=O)c1cccnc1 formula: C7H8N2O
Index of correct structure: -1 of 376372
True structure loss: 0.02822
True structure:
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [#6X3] [#6X3]               | 0.9993 | OX2H0 ] | CX4H2 ] | CX4H3 ] | CX4H1 ] | 0.1382 |
| [#6H1]                      | 0.9993 | | | | | |
| [#6X3H0] [#6X3H0]           | 0.9883 | OX2H0 ] | CX4H2 ] | CX4H3 ] | CX4H1 ] | 0.0 |
| [#7] [#6] [#6X3]            | 0.9696 | OX2H1 ] | CX4H1 ] | CX4H1 ] | CX4H1 ] | 0.0 |
| [#7] [#6] [#6] [#6X3]       | 0.8944 | OX2H1 ] | CX4H1 ] | CX4H1 ] | CX4H1 ] | 0.0 |
| [#6X3] [#6X3] [#6X3]        | 0.8868 | OX2H0 ] | CX4H2 ] | CX4H3 ] | CX4H1 ] | 0.0 |
| [cH]                        | 0.8516 | CX4H0 ] | CX4H2 ] | CX4H1 ] | CX4H1 ] | 0.0 |
| [#6X3] [#6X3H0] [#6X3H1]    | 0.8217 | CX4H1 ] | CX4H2 ] | CX4H0 ] | CX4H0 ] | 0.0 |
| [cH] [cH]                   | 0.6439 | | | | | |
| [cX3H1] (cX3H1) [cX3H0]    | 0.6405 | | | | | |
| worst negatives             | prob | worst positives | prob | worst negatives | prob |
| [#7] [#6X3H0] [#6X3H1]      | 0.6519 | CX4H3 ] | CX4H1 ] | 0.1382 |
| [#7] [#6H0] [#6H1]          | 0.5931 | | | 0.1653 |
| [#8] [#6] [#6X3]            | 0.5206 | | | 0.208 |
| [OX2H1]                     | 0.4645 | | | 0.2701 |
| [#8] [#6H0] [#6H1]          | 0.3266 | | | 0.2975 |
| [#7X3H2]                    | 0.3261 | cX3H1 ] | cX3H1 ] | 0.316 |
| [#6X3] [#6] [#6H3]          | 0.2444 | cX3H1 ] | cX3H1 ] | 0.3174 |
| [#7] [#6] [#6H1]             | 0.2241 | | | 0.3245 |
| [#7] [#6] [#7]              | 0.2103 | | | 0.3677 |
| [cH] cO                     | 0.2070 | | | 0.4011 |
Example 3 true smiles: NC(=O)CCC(N)=O formula: C5H10N2O3
Index of correct structure: 3 of 371534
True structure loss: 0.024882
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

0.020751  0.022857  0.023199  0.024882  0.025053

0.025267  0.02767  0.028047  0.028804  0.029394
| Top predicted substructures | prob | Best positives | prob | Best negatives | prob |
|----------------------------|------|----------------|------|----------------|------|
| [CX4H2][{#6}][#6]          | 1.0  | O=[CX3H0][CX4H2][CX4H2] | 0.9637 | [CX2H1][#][CX2H0][CX3H1][#][CX3H0] | 0.0  |
| [CX3]([=OX1])C             | 0.9966 | [CX3][#][OX1])O | 0.9542 | [CX2H0][#][CX2H0][CX4H0] | 0.0  |
| [#7X3H2]                   | 0.9756 | [CX4H1][#][NX3H2][#][CX4H2][#][CX3H0] | 0.8848 | [CX2H0][#][CX2H0][CX3H0] | 0.0  |
| [CX4H2][#][CX4H2][CX4H2]   | 0.9706 | O=[CX3H0][CX4H2][CX4H2] | 0.882 | [CX2H0][#][CX2H0][CX3H0] | 0.0  |
| [CX4H2][CX4H1][CX4H2]      | 0.9706 | [CX3H0][=OX1])O | 0.9542 | [CX2H0][#][CX2H0][CX3H0] | 0.0  |
| [OX1H0]=#][CX3H0][#][CX4H2] | 0.7335 | O=[CX3H0][CX4H2][CX4H2] | 0.2988 | [CX2H0][#][CX2H0][CX3H0] | 0.0  |
| [#6][#6][#6]               | 0.5612 | #][=OX1])O | 0.2028 | [CX2H0][#][CX2H0][CX3H0] | 0.0  |
| [CX4H2][[CX4H2]][CX4H2]    | 0.5612 | O=[OX1H0][#][NX3H1][#][CX4H2] | 0.2545 | [OX1H0][#][OX2H1][#][CX4H2] | 0.0  |
| [OX1H0][=OX1H0][#][CX4H2]  | 0.2988 | O=[CX3H0][CX4H2][CX4H2] | 0.2269 | [OX1H0][#][OX2H1][#][CX4H2] | 0.0  |
| [#7X3H1]                   | 0.3015 | [OX1H0][#][NX3H1][#][CX4H2] | 0.2545 | [OX1H0][#][OX2H1][#][CX4H2] | 0.0  |
| [OX1H0][=OX1H0][#][OX2H1]  | 0.2988 | O=[CX3H0][CX4H2][CX4H2] | 0.2269 | [OX1H0][#][OX2H1][#][CX4H2] | 0.0  |
| [#7][#6][#6][#6]           | 0.1517 | [OX1H0][#][OX2H1][#][CX4H2] | 0.3908 | [OX1H0][#][OX2H1][#][CX4H2] | 0.0  |
| [OX1H0][=OX1H0][#][OX2H1]  | 0.118  | [OX1H0][#][OX2H1][#][CX4H2] | 0.3908 | [OX1H0][#][OX2H1][#][CX4H2] | 0.0  |
Example 4 true smiles: NC(\text{-O})CCC(N)C(\text{-O})O formula: C5H10N2O3
Index of correct structure: -1 of 371534
True structure loss: 0.020849
True structure:

Experimental 13C NMR (solvent: N/A)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

1. \begin{align*}
\text{H}_2\text{N} &\text{O} \\
\text{H} &\text{O}
\end{align*}
0.040077

2. \begin{align*}
\text{H}_2\text{N} &\text{OH} \\
\text{H}_2\text{N} &\text{OH}
\end{align*}
0.044054

3. \begin{align*}
\text{H}_2\text{N} &\text{OH} \\
\text{OH} &\text{OH}
\end{align*}
0.045911

4. \begin{align*}
\text{H}_2\text{N} &\text{N} \text{H} \\
\text{H}_2\text{N} &\text{OH}
\end{align*}
0.046316

5. \begin{align*}
\text{H}_2\text{N} &\text{N} \text{H} \\
\text{O} &\text{OH}
\end{align*}
0.046643

6. \begin{align*}
\text{H}_2\text{N} &\text{O} \\
\text{H}_2\text{N} &\text{OH}
\end{align*}
0.047568

7. \begin{align*}
\text{H}_2\text{N} &\text{N} \text{H} \\
\text{H}_2\text{N} &\text{OH}
\end{align*}
0.047568

8. \begin{align*}
\text{H}_2\text{N} &\text{N} \text{H} \\
\text{H}_2\text{N} &\text{OH}
\end{align*}
0.048043

9. \begin{align*}
\text{H}_2\text{N} &\text{N} \text{H} \\
\text{O} &\text{OH}
\end{align*}
0.048543

10. \begin{align*}
\text{H}_2\text{N} &\text{N} \text{H} \\
\text{H}_2\text{N} &\text{OH}
\end{align*}
0.048938
| Top predicted substructures                              | prob | best positives                              | prob | best negatives                              | prob |
|----------------------------------------------------------|------|---------------------------------------------|------|---------------------------------------------|------|
| [CX4H2][{#6}][{#6}]                                     | 0.9992 | [CX2H1][{CX2H0}][{CX3H1}]=[CX3H0]           | 0.0  | [CX2H1][{CX2H0}][{CX3H1}]=[CX3H0]           | 0.0  |
| [CX3][={OX1}][{#6}]                                     | 0.9823 | CC=CCC[C]                                  | 0.0  | CC=CCC[C]                                  | 0.0  |
| [OX2H1]                                                 | 0.9796 | CC=CCC[C]                                  | 0.0  | CC=CCC[C]                                  | 0.0  |
| [CX4H2][{CX4H2}]                                        | 0.8941 | C=CCCC[C]                                  | 0.0  | C=CCCC[C]                                  | 0.0  |
| [#7H2][{#6}]                                            | 0.8829 | C=CCCC[C]                                  | 0.0  | C=CCCC[C]                                  | 0.0  |
| [#7X3H2]                                                | 0.8733 | CC#CCC=C                                  | 0.0  | CC#CCC=C                                  | 0.0  |
| [#8]=[{#6}][{#6}]                                      | 0.8292 | [CX3H1] ={[CX3H1]}= [CX2H0]                | 0.0  | [CX3H1] ={[CX3H1]}= [CX2H0]                | 0.0  |
| [#7][{#6}][{#6X3}]                                      | 0.8207 | C=CCCC[C]                                  | 0.0  | C=CCCC[C]                                  | 0.0  |
| [CX4H2][{CC}[O]                                         | 0.8164 | CCC=CC=C                                  | 0.0  | CCC=CC=C                                  | 0.0  |
| OCC{CH2}                                                | 0.7604 | [CX3H0] ={[CX3H1]}= [{OX2H0}][{CX2H0}]     | 0.0  | [CX3H0] ={[CX3H1]}= [{OX2H0}][{CX2H0}]     | 0.0  |
| worst negatives                                          | prob | worst positives                              | prob | worst positives                              | prob |
| [#7X3H1]                                                | 0.4857 | [#7H2][{#6}][{#6}]                         | 0.1693 | [#7H2][{#6}][{#6}]                         | 0.1693 |
| [#7][{#6}][{#6}][{#7}]                                 | 0.3365 | [#8]=[{#6}][{#6}][{#6}]                   | 0.2032 | [#8]=[{#6}][{#6}][{#6}]                   | 0.2032 |
| [#7X3H0]                                                | 0.3296 | [CX3H0]=[{OX1H0}][{OX2H1}][{CX4H1}]        | 0.2892 | [CX3H0]=[{OX1H0}][{OX2H1}][{CX4H1}]        | 0.2892 |
| [#8]=[{#6}][{#6X2}]                                     | 0.2799 | [#7][{#6}][{#6}][{#6}][{#6}][{#6}][{#7}]  | 0.4205 | [#7][{#6}][{#6}][{#6}][{#6}][{#6}][{#7}]  | 0.4205 |
| [#6H1][{#6}]                                            | 0.2698 | [CX3]=[{O}][{OX2H1}]                      | 0.451 | [CX3]=[{O}][{OX2H1}]                      | 0.451 |
| [#7][{#6}][{#6}][{#6}][{#7}]                           | 0.258 | [#8]=[{#6}][{#6}][{#6}]                   | 0.5107 | [#8]=[{#6}][{#6}][{#6}]                   | 0.5107 |
| [OH][{CX4H}][{O}]                                       | 0.2145 | O={XC3}[{CX4H}]                            | 0.5163 | O={XC3}[{CX4H}]                            | 0.5163 |
| [CX4H1][{OX2H1}][{OX4H1}][{CX4H1}]                     | 0.1968 | [CX4H2][{CX3}=O]                           | 0.5288 | [CX4H2][{CX3}=O]                           | 0.5288 |
| [CX4H2][{CX4H2}][{CX4H2}]                               | 0.1663 | [#6H1]                                    | 0.6315 | [#6H1]                                    | 0.6315 |
| [#7][{#6}][{#6}][{#6}][{#6}][{#7}]                     | 0.1604 | [CX4H2][{CX4H2}][{CX4H1}]                 | 0.6329 | [CX4H2][{CX4H2}][{CX4H1}]                 | 0.6329 |
Example 5 true smiles: O=Cl[\text{nH}]cc(CO)c(=O)[\text{nH}]1
Formula: C5H6N2O3
Index of correct structure: -1 of 354279
True structure loss: 0.035503
True structure:

Experimental 13C NMR (solvent: DMSO-d6)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):
### Top predicted substructures

| Structure | prob |
|-----------|------|
| [#6H1]    | 0.9707 |
| [#6X3] [#6X3] | 0.9662 |
| [#7] [#6] [#6X3] | 0.8608 |
| [OX2H1] [CX4H2] [#6X3H0] | 0.8345 |
| [#6X3H1] [#6X3H0] | 0.8231 |

### Best positives

| Structure | prob |
|-----------|------|
| [#6H1]    | 0.9707 |
| [#6X3] [#6X3] | 0.9662 |
| [#7] [#6] [#6X3] | 0.8608 |
| [OX2H1] [CX4H2] [#6X3H0] | 0.8345 |

### Best negatives

| Structure | prob |
|-----------|------|
| [#6X3H1] [#6X3H0] | 0.8231 |
| [#8] [#6] [#6] [#6X3] | 0.6931 |

### Worst negatives

| Structure | prob |
|-----------|------|
| [#7X3H0] [#6X3H1] | 0.2766 |
| [#6X3H1] [#6X3H0] | 0.2739 |
Example 6 true smiles: NCCc1ccc(G)c1 formula: C8H11NO
Index of correct structure: 0 of 193269
True structure loss: 0.024764
True structure:

Experimental 13C NMR (solvent: CD3OD)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

- H2N-CH2-Ph-OH: 0.024764
- H2N-CH2-C6H5: 0.026646
- H2N-CH2-C6H5: 0.029193
- H2N-CH2-C6H5: 0.030488
- C8H11NH: 0.035378
- H2N-CH2-C6H5: 0.039365
- C8H11NH: 0.041407
- C8H11NH: 0.042634
- C8H11NH: 0.043999
- C8H11NH: 0.046839
| Top predicted substructures | prob | best positives | prob | worst negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [#6H1]                      | 0.9993 | [CX4H2]|{CX4H2}|{CX3H0} | 0.9351 |
| [CX4H2]|{#6}|{#6} | 0.9958 | [#7]|{#6}|{#6X3} | 0.8705 |
| [cH]|{cH} | 0.9816 | [#6H1]|{#6H1} | 0.8443 |
| [#6X3]|{#6X3}|{#6X3}|{#6X3} | 0.9614 | [CX4H2]|{#6X3H2}|{CX4H2} | 0.8428 |
| [cX3]|{#6X3} | 0.9439 | [cH] | 0.8401 |
| [CX4H2]|{#6}|{#6} | 0.9958 | [CX4H0]|{OX2H1}|{CX3H0}|{CX4H1} | 0.0 |
| [cH]|{cH} | 0.9816 | [CX4H0]|{OX2H1}|{CX4H3}|{CX4H0} | 0.0 |
| [#6X3]|{#6X3}|{#6X3}|{#6X3} | 0.9614 | [OX2H1]|{CX4H0}|{CX4H1}|{1} | 0.0 |
| [cX4H2]|{#6X3H2}|{cX4H2} | 0.9351 | [CX3H0]|{OX2H1}|{CX4H3}|{CX4H0} | 0.0 |
| [#6H1]|{#6H1} | 0.8443 | [CX4H0]|{OX2H1}|{CX4H3}|{CX4H2}|{CX4H1} | 0.0 |
| [cX4H2]|{#6X3H2}|{cX4H2} | 0.8428 | [CX4H1]|{OX2H1}|{CX4H3}|{CX4H0} | 0.0 |
| [cH] | 0.8401 | [OX2H0]|{CX4H2}|{CX4H0}|{CX4H3} | 0.0 |
| [cX3]|{#6X3H1}|{cX3H1} | 0.8379 | [CX3H2]|{#6H1}|{CX4H0}|{OX2H1} | 0.0 |

| Best positives | prob | Best negatives | prob |
|----------------|------|----------------|------|
| [#6H1] | 0.9993 | [CX4H0]|{CX4H3}|{CX4H2}|{CX4H2}|{CX3H0} | 0.0 |
| [CX4H2]|{#6}|{#6} | 0.9816 | [OX2H1]|{CX4H0}|{CX4H3}|{CX4H2}|{CX4H1} | 0.0 |
| [#6X3]|{#6X3}|{#6X3}|{#6X3} | 0.9614 | [OX2H0]|{CX4H0}|{CX4H1}|{1} | 0.0 |
| [cX4H2]|{#6X3H2}|{cX4H2} | 0.9351 | [CX3H0]|{OX2H1}|{CX4H3}|{CX4H0} | 0.0 |
| [#6H1]|{#6X3}|{#6X3} | 0.8443 | [OX2H0]|{CX4H2}|{CX4H3}|{CX4H0} | 0.0 |
| [cX4H2]|{#6X3H2}|{cX4H2} | 0.8428 | [OX2H0]|{CX4H2}|{CX4H0}|{CX4H3} | 0.0 |
| [cH] | 0.8401 | [OX2H0]|{CX4H2}|{CX4H0}|{CX4H3} | 0.0 |
| [cX3]|{#6X3H1}|{cX3H1} | 0.8379 | [CX3H2]|{#6H1}|{CX4H0}|{OX2H1} | 0.0 |

| Worst positives | prob | Worst negatives | prob |
|------------------|------|-----------------|------|
| [#7]|{#6}|{#6H3} | 0.8705 | [cX3H0]|{cX3H1}|{cX3H1}|{cX3H0} | 0.1148 |
| [#6H1]|{#6H2} | 0.6122 | [#6X3H1]|{#6X3H1}|{#6X3H1}|{#6X3H1}|{#6X3H1} | 0.141 |
| [cX3H1]|{cX3H1}|{cX3H1} | 0.5428 | [cH]|{cO} | 0.2988 |
| [CX4H2]|{cX3H0} | 0.4241 | [OX2H1] | 0.3204 |
| [#6X3]|{#7}|{#6X3} | 0.3728 | [CX4H2]|{CX4H2} | 0.3796 |
| [#61]|{#6}|{#6}|{#6}|{#6}|{#7} | 0.2982 | [OX2H1]|{cX3} |{c} | 0.3968 |
| [#7]|{#6X3H1} | 0.2617 | [#6]|{#6}|{#6}|{#6}|{#6} | 0.427 |
| [#7]|{#6H0}|{#6H1} | 0.2514 | [#6]|{#6H0}|{#6H1} | 0.4371 |
| [#6H2]|{#7}|{#6X3} | 0.2111 | [#6]|{#6}|{#6}|{#6X3} | 0.4734 |
| [cX3H0]|{cX3H1}|{cX3H0}|{cX3H0}|{cX4H2} | 0.1989 | [#7]|{#6H2}|{#6H2} | 0.5587 |

----------------------------------------------------------------------------------------------------------------------------------------
Example 7 true smiles: NCCCCC(N)=O formula: C6H14N2O2
Index of correct structure: 0 of 143634
True structure loss: 0.021445

True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

1. 0.021445
2. 0.032197
3. 0.033694
4. 0.034373
5. 0.034638

1. 0.034681
2. 0.036561
3. 0.03884
4. 0.040982
5. 0.041044
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [CX4H2][{#6}]/{#6}         | 1.0  | [CX4H2][{CX4H2}][CX4H2] | 0.9238 | | |
| {#7X3H2}                    | 0.9909 | OCC[{CH2}] | 0.9076 | | |
| [O2H1]                      | 0.9893 | [CX4H2]CC=O | 0.8824 | | |
| [CX4H2][CX4H2]              | 0.968 | [CX4H2][{CX4H2}][CX4H1] | 0.8577 | | |
| [O2H1]                      | 0.9354 | [{#7}][{6H2}][{6H2}] | 0.8489 | | |

| worst negatives | prob | worst positives | prob |
|------------------|------|-----------------|------|
| [{#8}][{6}][{6H2}] | 0.569 | CCCCCC | 0.069 |
| [{#7}][{6}][{6H0}] | 0.5649 | [{#8}][{6H0}][{6H1}] | 0.1173 |
| [CX4H2]CX3=O     | 0.5159 | [CX3H0][{OX1H0}][{OX2H1}][{CX4H1}] | 0.2924 |
| O=[CX3H0][{CX4H2}][CX4H2] | 0.451 | [{#6}][{6H2}] | 0.4028 |
| [CX4H2][{OX2H1}][{CX4H2}] | 0.3368 | [{#6}][{6H1}] | 0.4525 |
| [{#7X3H1}]       | 0.275 | [CX3][{OX1H0}] | 0.479 |
| [{#7}][{6}][{6}][{6}][{#7}] | 0.2494 | [{#7}][{#6}][{6K3}] | 0.4943 |
| [CX4H2][{N3H1}][{CX4H2}] | 0.2047 | [{#8}][{#6}][{#8}] | 0.57 |
| [{#6H1}][{6H2}][{6}][{6}][{#7}] | 0.1833 | [CX4H1][{NX3H2}][{CX4H2}][{CX3H0}] | 0.6514 |
| [{#7}][{6}][{6}][{#7}] | 0.143 | [{#8}][{#6H0}][{6H1}] | 0.6631 |
Example 8 true smiles: NCCCCC(N)=O
formula: C6H14N2O2
Index of correct structure: 0 of 143634
True structure loss: 0.021405
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):
### Top predicted substructures

| Substructure                          | prob |
|--------------------------------------|------|
| [CX4H2][CX4H2][CX4H2]               | 0.9238 |
| [#7X3H2]                            | 0.9903 |
| [OX2H1]                             | 0.9892 |
| [CX4H2][CX4H2]                      | 0.9679 |
| [CX3](=[OX1])C                      | 0.9363 |
| OCC[CH2]                            | 0.9076 |
| [OX2H1]                             | 0.9892 |
| [CX4H2]                             | 0.9679 |
| [OX2H1][CX4H2]                      | 0.9363 |
| OCC[CH2][CX4H2]                     | 0.9076 |
| [OX4H2][OX4H2]                      | 0.9139 |
| [OX2H1][OX2H1][OX4H1]               | 0.9363 |
| [#7][#6H2][#6H2]                    | 0.8567 |
| worst negatives                     | prob |
| [OX4H2][OX4H2][OX4H2]               | 0.5139 |
| [CX4H2][OX2H1][CX4H2]               | 0.4554 |
| [CX4H2][OX2H1][CX4H2]               | 0.3334 |
| [#7X3H1]                            | 0.2746 |
| [#7X3H1][OX4H2][OX4H2]              | 0.1832 |
| [#7][#6][#6][#6][#7]                | 0.2487 |
| [#6H1][#6H2][#6][#6][#7]            | 0.1446 |

### best positives

| Substructure                          | prob |
|--------------------------------------|------|
| [CX4H2][CX4H2][CX4H2]               | 1.0 |
| [#7X3H2]                            | 0.9903 |
| [OX2H1]                             | 0.9892 |
| [CX4H2][CX4H2]                      | 0.9679 |
| [CX3](=[OX1])C                      | 0.9363 |
| OCC[CH2]                            | 0.9076 |
| [OX2H1]                             | 0.9892 |
| [OX4H2][OX4H2]                      | 0.9139 |
| [OX2H1][OX2H1][OX4H1]               | 0.9363 |
| [#7][#6H2][#6H2]                    | 0.8567 |

### worst negatives

| Substructure                          | prob |
|--------------------------------------|------|
| [OX4H2][OX4H2][OX4H2]               | 0.5139 |
| [CX4H2][OX2H1][CX4H2]               | 0.4554 |
| [OX4H2][OX2H1][CX4H2]               | 0.3334 |
| [#7X3H1]                            | 0.2746 |
| [#7X3H1][OX4H2][OX4H2]              | 0.1832 |
| [#7][#6][#6][#6][#7]                | 0.2487 |
| [#6H1][#6H2][#6][#6][#7]            | 0.1446 |

### Additional constraints

- [CX4H2][CX4H2][CX4H2] = [CX4H2][CX4H2][CX4H2]
- [#7X3H2] = [CX4H2][CX4H2][CX4H2]
- [OX2H1] = [CX4H2][CX4H2][CX4H2]
- [CX4H2][CX4H2] = [CX4H2][CX4H2][CX4H2]
- [CX3](=[OX1])C = [CX4H2][CX4H2][CX4H2]
Example 9 true smiles: Nc1cccc(C(=O)O)c1 formula: C7H7NO2
Index of correct structure: 4 of 141060
True structure loss: 0.027383
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

- 0.020355
- 0.023262
- 0.023546
- 0.027078
- 0.027383
- 0.028523
- 0.029103
- 0.029256
- 0.029341
- 0.02951
| Top predicted substructures | prob | best positives | prob | best negatives | prob | worst negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|----------------|------|
| [#6X3] | 0.9953 | [O=] | 0.0047 | [OX2H0] | 0.0099 | [OX2H0] | 0.0099 |
| [O=] | 0.9934 | [O=] | 0.0066 | [OX2H0] | 0.0099 | [OX2H0] | 0.0099 |
| [cH] | 0.9704 | [cH] | 0.0296 | [OX2H0] | 0.0099 | [OX2H0] | 0.0099 |
| [cH] | 0.9217 | [cH] | 0.0783 | [OX2H0] | 0.0099 | [OX2H0] | 0.0099 |
| [#6X3] | 0.9036 | [#6X3] | 0.0964 | [OX2H0] | 0.0099 | [OX2H0] | 0.0099 |
| [#6X3] | 0.8839 | [#6X3] | 0.1161 | [OX2H0] | 0.0099 | [OX2H0] | 0.0099 |
| [O=] | 0.8684 | [O=] | 0.1316 | [OX2H0] | 0.0099 | [OX2H0] | 0.0099 |
| [cH] | 0.8621 | [cH] | 0.1379 | [OX2H0] | 0.0099 | [OX2H0] | 0.0099 |
| [cH] | 0.8292 | [cH] | 0.1708 | [OX2H0] | 0.0099 | [OX2H0] | 0.0099 |
| [#6X3] | 0.784  | [#6X3] | 0.216  | [OX2H0] | 0.0099 | [OX2H0] | 0.0099 |
Example 10 true smiles: O=C(O)C\text{C}cc1ccc\text{n}c1 formula: C7H7NO2
Index of correct structure: 1 of 141060
True structure loss: 0.050539
True structure:

Experimental 13C NMR (solvent: DMSO-d6)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

0.035522
0.037206
0.038745
0.039656
0.040429

0.040517
0.040547
0.042879
0.043882
0.046072
### Top predicted substructures

| Substructure | Prob |
|--------------|------|
| [#6X3][#6X3] | 0.9987 |
| [cH]         | 0.9935 |
| [#6H1]       | 0.9921 |
| [#6X3H1][#6X3H0] | 0.9772 |
| [cH][cH]     | 0.9702 |
| [#6X3][#6X3][#6X3][#6X3] | 0.9642 |
| [cX3H1][cX3H1][cX3H0] | 0.8783 |
| [cX3H1][cX3H1] | 0.8771 |

### Best positives

| Substructure | Prob |
|--------------|------|
| [#6X3][#6X3] | 0.9987 |
| [cH]         | 0.9935 |
| [#6H1]       | 0.9921 |
| [#6X3H1][#6X3H0] | 0.9772 |
| [cH][cH]     | 0.9702 |
| [#6X3][#6X3][#6X3][#6X3] | 0.9642 |
| [cX3H1][cX3H1][cX3H0] | 0.8783 |
| [cX3H1][cX3H1] | 0.8771 |

### Best negatives

| Substructure | Prob |
|--------------|------|
| [#6X3][#6X3] | 0.9987 |
| [cH]         | 0.9935 |
| [#6H1]       | 0.9921 |
| [#6X3H1][#6X3H0] | 0.9772 |
| [cH][cH]     | 0.9702 |
| [#6X3][#6X3][#6X3][#6X3] | 0.9642 |
| [cX3H1][cX3H1][cX3H0] | 0.8783 |
| [cX3H1][cX3H1] | 0.8771 |

### Worst negatives

| Substructure | Prob |
|--------------|------|
| [#6X3][#6X3] | 0.9987 |
| [cH]         | 0.9935 |
| [#6H1]       | 0.9921 |
| [#6X3H1][#6X3H0] | 0.9772 |
| [cH][cH]     | 0.9702 |
| [#6X3][#6X3][#6X3][#6X3] | 0.9642 |
| [cX3H1][cX3H1][cX3H0] | 0.8783 |
| [cX3H1][cX3H1] | 0.8771 |

### Worst positives

| Substructure | Prob |
|--------------|------|
| [#6X3][#6X3] | 0.9987 |
| [cH]         | 0.9935 |
| [#6H1]       | 0.9921 |
| [#6X3H1][#6X3H0] | 0.9772 |
| [cH][cH]     | 0.9702 |
| [#6X3][#6X3][#6X3][#6X3] | 0.9642 |
| [cX3H1][cX3H1][cX3H0] | 0.8783 |
| [cX3H1][cX3H1] | 0.8771 |

### Worst negatives

| Substructure | Prob |
|--------------|------|
| [#6X3][#6X3] | 0.9987 |
| [cH]         | 0.9935 |
| [#6H1]       | 0.9921 |
| [#6X3H1][#6X3H0] | 0.9772 |
| [cH][cH]     | 0.9702 |
| [#6X3][#6X3][#6X3][#6X3] | 0.9642 |
| [cX3H1][cX3H1][cX3H0] | 0.8783 |
| [cX3H1][cX3H1] | 0.8771 |
Example 11 true smiles: NC(CCC(=O)O)C(=O)O
formula: C5H9NO4
Index of correct structure: 0 of 92537
True structure loss: 0.028545
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

0.028545  0.037194  0.039214  0.042218  0.042586
0.042857  0.043528  0.043671  0.044099  0.044258
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [CX4H2][#6][#6]            | 0.9996 | [CX3]=[OX1][O] | 0.8933 | | |
| [OX2H1]                    | 0.9984 | [CX3]=O[OX2H1] | 0.8039 | | |
| [CX3]=[OX1]C               | 0.9976 | O=[CX3H0][CX4H2][CX4H2] | 0.7848 | | |
| [#8]=[#6][#8]              | 0.9208 | [#6H1]          | 0.7701 | | |
| OCC[C]H2                   | 0.905  | [#8][#6][#6H2] | 0.7503 | | |
| [CX4H2][#6][#6]            | 0.9996 | [CX2H1][CX2H0][CX3H1]=[CX3H0] | 0.0 | | |
| [OX2H1]                    | 0.9984 | CC=CC=CC        | 0.0 | | |
| [CX3]=[OX1]C               | 0.9976 | [CX3H0]=[CX3H2][{CX4H3}][CX4H2] | 0.0 | | |
| [#8]=[#6][#8]              | 0.9208 | [#6H3][#6X3][#6X3]=[#6X3H2] | 0.0 | | |
| OCC[C]H2                   | 0.905  | CCCC=CC         | 0.0 | | |
| [CX3]=[OX1]O               | 0.8933 | CC=CC           | 0.0 | | |
| [OX2H1]                    | 0.8039 | [CX3H1]=[CX3H1][CX2H0] | 0.0 | | |
| O=[CX3H0][CX4H2][CX4H2]    | 0.7848 | CC=CC=CC        | 0.0 | | |
| [#6H1]                     | 0.7701 | C=CC=CC         | 0.0 | | |
| [#8][#6][#6H2]             | 0.7503 | [CX3H1]=[#6X3H1][#6X3H0]=[#6X3H1] | 0.0 | | |

| worst negatives | prob | worst positives | prob |
|------------------|------|-----------------|------|
| [#8][#6][#6][#6][#6][#6]=[#8] | 0.5963 | [#7H2][#6X4H1][#6X3] | 0.1517 |
| [#8][#6][#6][#6][#6][#6]=[#8] | 0.3958 | [CX4H1][{NX3H2}][{CX4H2}][CX3H0] | 0.0617 |
| [#7X3H1]         | 0.3574 | [#7X3H2]        | 0.2904 |
| [CX4H2][OX2H1][OX2H1] | 0.3479 | [#6X1H0]=[#6X1H0][#6H1] | 0.3363 |
| [#8][#6][#6][#6X3] | 0.3249 | [OX1H0]=[CX3H0][{#8}][CX4H2] | 0.3559 |
| [#7][#6H1][#6H2][#6] | 0.2829 | [CX3H0]=[OX1H0][{OX2H1}][CX4H1] | 0.4671 |
| [#7][#6H1][#6H2][#6] | 0.2256 | [OX1H0]=[#OX1H0][#OX2H1][CX4H1] | 0.5111 |
Example 12: true smiles: Nc1cccc(O)c1N
Formula: C₆H₈N₂O
Index of correct structure: 3 of 75211
True structure loss: 0.022729
True structure:

Experimental ¹³C NMR (solvent: DMSO)

Experimental ¹H NMR (solvent: CDCl₃)

Top predicted structures (loss):

- 0.021308
- 0.02231
- 0.022402
- 0.022729
- 0.022749
- 0.023122
- 0.024044
- 0.02418
- 0.025169
- 0.025648
| Top predicted substructures | prob | [cH] | 0.9648 |
|----------------------------|------|------|--------|
| [#6H1]                     | 0.9989 |      |        |
| [#6X3]                     | 0.9954 |      |        |
| [#6X3H0]                   | 0.9932 |      |        |
| [#6X3][#6X3][#6X3]        | 0.9837 |      |        |
| [#6X3][#6X3][#6X3][#6X3]  | 0.9791 |      |        |

| best positives            | prob | best negatives        | prob |
|----------------------------|------|-----------------------|------|
| [#6H1]                     | 0.9989 | [OX1H0][CX3H0][CX3H1][CX4H1][CX4H2] | 0.0 |
| [#6X3]                     | 0.9954 | [OX2H0][CX4H2][CX4H2][CX4H1] | 0.0 |
| [#6X3H0]                   | 0.9932 | [OX2H0][CX4H2][CX4H2][CX4H1] | 0.0 |
| [#6X3][#6X3]              | 0.9837 | [OX2H1][CX4H1][CX4H1][CX4H1] | 0.0 |
| [#7][#6][#6X3]            | 0.9791 | [OX2H1][CX4H1][CX4H1][CX4H1] | 0.0 |

| worst negatives           | prob | worst positives       | prob |
|----------------------------|------|-----------------------|------|
| [#6X3][#7][#6X3]          | 0.7913 | [#6][#6][#6][#6][#6] | 0.1071 |
| [#6X3][#7][#6X3]          | 0.7362 | [#7][#6][#6][#7]     | 0.1733 |
| [#6][#6][#6][#6][#7]      | 0.6031 | [cX3H0][cX3H1][cX3H0][CX2H1] | 0.3688 |
| [#6X3H1][#6X3H0][#6X3H1] | 0.5985 | [#8][#6][#6][#6X3]   | 0.4102 |
| [#7][#6X3H1]              | 0.5688 | [#8][#6][#6][#6X3]   | 0.4976 |
| [#7][#6X3H1]              | 0.5175 | [OX2H0][cX3][c]      | 0.6236 |
| [#6H1r5][#7]              | 0.4436 | [cH][cO]             | 0.6685 |
| [cX3H1][cX3H1][cX3H1]     | 0.3278 | [#72H][#6H0]        | 0.7382 |
| [#6X3H2]                  | 0.3046 | [OX2H1]             | 0.7828 |
Example 13 true smiles: O=C(O)c1ccc(O)nc1 formula: C6H5NO3
Index of correct structure: -1 of 67881
True structure loss: 0.023869
True structure:

![Molecular structure](image)

Experimental 13C NMR (solvent: DMSO-d6)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

1. ![Structure 1](image) - 0.022012
2. ![Structure 2](image) - 0.028095
3. ![Structure 3](image) - 0.033325
4. ![Structure 4](image) - 0.034622
5. ![Structure 5](image) - 0.034832
6. ![Structure 6](image) - 0.035773
7. ![Structure 7](image) - 0.035795
8. ![Structure 8](image) - 0.03586
9. ![Structure 9](image) - 0.036143
10. ![Structure 10](image) - 0.036339
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [#6H1]                       | 0.9991 | [#6X3H1][#6X3H0] | 0.9599 | [cX3H1][cX3H1][cX3H0] | 0.9263 |
| [#6X3][#6X3] | 0.9979 | [cX3H1][cX3H1][cX3H0] | 0.9263 | [cX3H1][cX3H1][cX3H0] | 0.9599 |
| [cH][cH]                     | 0.9821 | [cX3H1][cX3H1][cX3H0] | 0.9599 | [cX3H1][cX3H1][cX3H0] | 0.9979 |
| [#6X3H1][#6X3H0]             | 0.9604 | [cX3H1][cX3H1][cX3H0] | 0.9263 | [cX3H1][cX3H1][cX3H0] | 0.9979 |
| [cX3H1][cX3H1][cX3H0]        | 0.9664 | [cX3H1][cX3H1][cX3H0] | 0.9263 | [cX3H1][cX3H1][cX3H0] | 0.9604 |
| [cX3H1][cX3H1][cX3H0]        | 0.8554 | [cX3H1][cX3H1][cX3H0] | 0.9263 | [cX3H1][cX3H1][cX3H0] | 0.9604 |
| [cX3H1][cX3H1][cX3H0]        | 0.7644 | [cX3H1][cX3H1][cX3H0] | 0.9263 | [cX3H1][cX3H1][cX3H0] | 0.8554 |
| [cX3H1][cX3H1][cX3H0]        | 0.5307 | [cX3H1][cX3H1][cX3H0] | 0.9263 | [cX3H1][cX3H1][cX3H0] | 0.5307 |
| [cX3H1][cX3H1][cX3H0]        | 0.4282 | [cX3H1][cX3H1][cX3H0] | 0.9263 | [cX3H1][cX3H1][cX3H0] | 0.4282 |
| [cX3H1][cX3H1][cX3H0]        | 0.4053 | [cX3H1][cX3H1][cX3H0] | 0.9263 | [cX3H1][cX3H1][cX3H0] | 0.4053 |
| [cX3H1][cX3H1][cX3H0]        | 0.3276 | [cX3H1][cX3H1][cX3H0] | 0.9263 | [cX3H1][cX3H1][cX3H0] | 0.3276 |
| [cX3H1][cX3H1][cX3H0]        | 0.3043 | [cX3H1][cX3H1][cX3H0] | 0.9263 | [cX3H1][cX3H1][cX3H0] | 0.3043 |
| [cX3H1][cX3H1][cX3H0]        | 0.2957 | [cX3H1][cX3H1][cX3H0] | 0.9263 | [cX3H1][cX3H1][cX3H0] | 0.2957 |
| [cX3H1][cX3H1][cX3H0]        | 0.2504 | [cX3H1][cX3H1][cX3H0] | 0.9263 | [cX3H1][cX3H1][cX3H0] | 0.2504 |
| [cX3H1][cX3H1][cX3H0]        | 0.1533 | [cX3H1][cX3H1][cX3H0] | 0.9263 | [cX3H1][cX3H1][cX3H0] | 0.1533 |
| [cX3H1][cX3H1][cX3H0]        | 0.1361 | [cX3H1][cX3H1][cX3H0] | 0.9263 | [cX3H1][cX3H1][cX3H0] | 0.1361 |

| worst negatives | prob | worst positives | prob | worst negatives | prob |
|------------------|------|-----------------|------|-----------------|------|
| [cX3H1][cX3H1][cX3H0] | 0.0381 | [cX3H1][cX3H1][cX3H0] | 0.0381 | [cX3H1][cX3H1][cX3H0] | 0.5307 |
| [cX3H1][cX3H1][cX3H0] | 0.1533 | [cX3H1][cX3H1][cX3H0] | 0.1533 | [cX3H1][cX3H1][cX3H0] | 0.4282 |
| [cX3H1][cX3H1][cX3H0] | 0.2571 | [cX3H1][cX3H1][cX3H0] | 0.2571 | [cX3H1][cX3H1][cX3H0] | 0.4053 |
| [cX3H1][cX3H1][cX3H0] | 0.3034 | [cX3H1][cX3H1][cX3H0] | 0.3034 | [cX3H1][cX3H1][cX3H0] | 0.3276 |
| [cX3H1][cX3H1][cX3H0] | 0.361  | [cX3H1][cX3H1][cX3H0] | 0.361  | [cX3H1][cX3H1][cX3H0] | 0.3043 |
| [cX3H1][cX3H1][cX3H0] | 0.3825 | [cX3H1][cX3H1][cX3H0] | 0.3825 | [cX3H1][cX3H1][cX3H0] | 0.2957 |
| [cX3H1][cX3H1][cX3H0] | 0.4119 | [cX3H1][cX3H1][cX3H0] | 0.4119 | [cX3H1][cX3H1][cX3H0] | 0.2504 |
| [cX3H1][cX3H1][cX3H0] | 0.4151 | [cX3H1][cX3H1][cX3H0] | 0.4151 | [cX3H1][cX3H1][cX3H0] | 0.2409 |
| [cX3H1][cX3H1][cX3H0] | 0.4361 | [cX3H1][cX3H1][cX3H0] | 0.4361 | [cX3H1][cX3H1][cX3H0] | 0.2264 |
Example 14 true smiles: Oc{eq}[nH{eq}]c2ccc(ccc12 formula: C8H7NO
Index of correct structure: 3 of 59121
True structure loss: 0.016676
True structure:

![Chemical structure](image)

Experimental 13C NMR (solvent: DMSO-d6)

![NMR spectrum](image)

Experimental 1H NMR (solvent: D2O)

![NMR spectrum](image)

Top predicted structures (loss):

1. ![Structure 1](image) - 0.014749
2. ![Structure 2](image) - 0.015962
3. ![Structure 3](image) - 0.016278
4. ![Structure 4](image) - 0.016676
5. ![Structure 5](image) - 0.018786
6. ![Structure 6](image) - 0.01948
7. ![Structure 7](image) - 0.020223
8. ![Structure 8](image) - 0.020537
9. ![Structure 9](image) - 0.022308
10. ![Structure 10](image) - 0.023469
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|---------------|------|
| [#6H1]                      | 0.9997 | [#6X3H1][#6X3H0] | 0.9876 | [#6X3H1] | 0.9814 |
| [#6X3][#6X3]                | 0.9992 | [cX3H1][cX3H1][cX3H0] | 0.9876 | [cX3H1][cX3H1][cX3H0] | 0.9876 |
| [cH][cH]                    | 0.9981 | [#6H1][#6H1] | 0.9794 | [#6H1][#6H1] | 0.9794 |
| [#6X3H1][#6X3H0]            | 0.9978 | [cX3H1][cX3H1][cX3H0] | 0.9814 | [cX3H1][cX3H1][cX3H0] | 0.9814 |
| [cX3H1](#cX3H1)[cX3H1]     | 0.9976 | [cX3H1][cX3H1][cX3H0] | 0.9814 | [cX3H1][cX3H1][cX3H0] | 0.9814 |
| [#6X3H1][#6X3H0]            | 0.9974 | [cX3H1][cX3H1][cX3H0] | 0.9814 | [cX3H1][cX3H1][cX3H0] | 0.9814 |
| [cX3H1][cX3H1][cX3H1]      | 0.9962 | [cX3H1][cX3H1][cX3H0] | 0.9814 | [cX3H1][cX3H1][cX3H0] | 0.9814 |
| [cX3H1][cX3H1]              | 0.9962 | [cX3H1][cX3H1][cX3H0] | 0.9814 | [cX3H1][cX3H1][cX3H0] | 0.9814 |

| worst negatives | prob | worst positives | prob |
|-----------------|------|----------------|------|
| [#6X3H1][#6X3H0][#6X3H1] | 0.6089 | [cX3H0][cX3H0][cX3H0][cX3H0] | 0.1816 |
| [cX3H1][cX3H1][cX3H1] | 0.425 | [cX3H1][cX3H1][cX3H0] | 0.2116 |
| [#7][#6][#6][#6X3] | 0.222 | [cX3H1][cX3H1][cX3H0] | 0.2718 |
| [#7][#6][#6][#6] | 0.1837 | [cH]CO | 0.307 |
| [#6][#7] | 0.1277 | [#7X3H1] | 0.3084 |
| [cH][cH] | 0.1139 | [OX2H1] | 0.4599 |
| [#7X3H0] | 0.1139 | [#7][#6X3H1] | 0.526 |
| [#7][#6][#6X3] | 0.1137 | [#7][#6X3H1][#6H1] | 0.5433 |
| [#6X3H1][#7X3H0] | 0.1058 | [#6H1][#7] | 0.567 |
| [cX3H1][cX3H0][cX3H1] | 0.1032 | [#7][#6X3H0][#6X3H1] | 0.5792 |
Example 15 true smiles: NC(=O)CC(N)C(=O)O
formula: C4H8N2O3
Index of correct structure: 0 of 58024
True structure loss: 0.033268
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

- 0.033268
- 0.035903
- 0.037272
- 0.03757
- 0.037613
- 0.03893
- 0.041735
- 0.042075
- 0.042357
- 0.042735
Top predicted substructures

| Structure                  | Prob | Structure                  | Prob |
|----------------------------|------|----------------------------|------|
| [CX3]=[OX1]C              | 0.9811 | [#8]=[#6H0][#6H1]         | 0.7567 |
| [OXYH1]                   | 0.9512 | O=[CX3][CX4H]             | 0.7458 |
| [#7X3H2]                  | 0.8666 | [CX3]=[OX1]D             | 0.726  |
| [#8]=[#6][#8]             | 0.848  | [#6H1][#6H2]             | 0.7157 |
| [#7H1]                    | 0.794  | [#7H2][#6H1]             | 0.6911 |

Best positives

| Structure                  | Prob | Structure                  | Prob |
|----------------------------|------|----------------------------|------|
| [CX3]=[OX1]C              | 0.9811 | [CX2H1][CX2H0][CX3H1]=[CX3H0] | 0.0 |
| [OXYH1]                   | 0.9512 | [CX3H0]=[CX3H1]([CX4H2])[CX2H0] | 0.0 |
| [#7X3H2]                  | 0.8666 | [CX3H0]=[CX3H1]([OXY2H0])[CX2H0] | 0.0 |
| [#8]=[#6][#8]             | 0.848  | CC=CCC=C                | 0.0 |
| [#6H1]                    | 0.794  | CC=CCC=C                | 0.0 |
| [#8]=[#6H0][#6H1][#6H1]   | 0.7567 | [CX3H0]=[CX3H2]([CX4H3])[CX4H0] | 0.0 |
| O=[CX3][CX4H]             | 0.7458 | C=CC=CCC               | 0.0 |
| [CX3]=[OX1]D             | 0.726  | [CX3H0]=[CX3H1]([CX4H3])[CX3H1] | 0.0 |
| [#6H1][#6H2]              | 0.7157 | C=CCC=C                | 0.0 |
| [#7H2][#6H1]              | 0.6811 | [#6H3][#6X3][#6X3]=[#6X3H2] | 0.0 |

Worst negatives

| Structure                  | Prob | Structure                  | Prob |
|----------------------------|------|----------------------------|------|
| [#7][#6H2]                | 0.58  | [#8]=[#6][#6][#6][#6][#8]  | 0.0702 |
| [#6H1][#6H1]              | 0.5766 | [#7H2][#6H0]              | 0.1389 |
| [#7X3][#6H2]              | 0.5574 | [#8]=[#6H0][#6H1]       | 0.2347 |
| [#7X3H1]                  | 0.5368 | [CX4H1][#NX3H2][CX4H2][CX3H0] | 0.305 |
| [#8][#6][#6][#6X3]        | 0.4624 | [CX4H2][#CX4H1][CX3H0]    | 0.3183 |
| [#8][#6H1][#6H1]          | 0.4136 | [#8]=[#6][#6H2][#6H1]    | 0.3464 |
| [CX4H0]                   | 0.4038 | O=[CX3H0][CX4H2][CX4H1]   | 0.3882 |
| [#8][#6][#6H1][#6H1][#6H1] | 0.2866 | [CX3H0]=[OX1H0][OX2H1][CX4H1] | 0.5026 |
| [CX4H2][#NX3H1][CX4H1]    | 0.2677 | [CX4H2][#6][#6]          | 0.5386 |
Example 16 true smiles: NC(=O)c1cccnc1 formula: C6H6N2O
Index of correct structure: -1 of 31495
True structure loss: 0.014517
True structure:

Experimental 13C NMR (solvent: DMSO)

Experimental 1H NMR (solvent: d2o)

Top predicted structures (loss):

0.019746 0.023498 0.025849 0.027113 0.027799
0.028293 0.028402 0.030001 0.030352 0.030973
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [#6H1]                      | 1.0  | [#6H1]        | 1.0  | [OX2H0]       | 0.0  |
| [#6X3][#6X3]                | 0.9989 | [OX2H1] | 0.9989 | [OX2H1][#6H1] | 0.0  |
| [#7][#6][#6X3]              | 0.9754 | [OX2H1][#7][#6H1] | 0.9754 | [OX2H1][#6H1][#7] | 0.0  |
| [cH]                        | 0.9649 | [OX2H1][#7][#6H1] | 0.9649 | [OX2H1][#6H1][#7] | 0.0  |
| [#6X3H1][#6X3H0]            | 0.9639 | [OX2H1][#7][#6H1] | 0.9639 | [OX2H1][#6H1][#7] | 0.0  |

| worst negatives | prob | worst positives | prob |
|----------------|------|----------------|------|
| [#7][#6H0][#6H1] | 0.5251 | [OX2H1][#7][#6H1] | 0.5251 |
| [#7][#6X3H0][#6X3H1] | 0.4934 | [OX2H1][#7][#6H1] | 0.4934 |
| [cH][cO]        | 0.2763 | [OX2H1][#7][#6H1] | 0.2763 |
| [#6][#6][#6X3]  | 0.2686 | [OX2H1][#7][#6H1] | 0.2686 |
| [#7][#6X3H1]    | 0.1733 | [OX2H1][#7][#6H1] | 0.1733 |
| [#7][#6X3H1]    | 0.1682 | [OX2H1][#7][#6H1] | 0.1682 |
| [O2XH1]         | 0.1197 | [OX2H1][#7][#6H1] | 0.1197 |
| [#7][#6][#6X3]  | 0.1191 | [OX2H1][#7][#6H1] | 0.1191 |
| [#7][#7]        | 0.1179 | [OX2H1][#7][#6H1] | 0.1179 |
| [#7][#6][#7]    | 0.109  | [OX2H1][#7][#6H1] | 0.109  |
Example 17 true smiles: CN(C(=N)NC(=N)N formula: C4H11N5
Index of correct structure: 3 of 30817
True structure loss: 0.01683
True structure:

\[
\begin{array}{c}
\text{Experimental }^{13}\text{C NMR (solvent: D2O)} \\
200 \quad 150 \quad 100 \quad 50 \quad 0 \\
\text{Experimental }^1\text{H NMR (solvent: d2o)} \\
12 \quad 10 \quad 8 \quad 6 \quad 4 \quad 2 \quad 0 \quad -2 \\
\end{array}
\]

Top predicted structures (loss):

1. 0.015959
2. 0.016287
3. 0.016689
4. 0.01683
5. 0.016899
6. 0.017047
7. 0.017159
8. 0.017536
9. 0.018149
10. 0.018632
Top predicted substructures

| Structure | prob |
|-----------|------|
| [CX4H3]   | 0.9746 |
| [#7][#6](#7)=#7 | 0.9737 |
| [#6H3][#7] | 0.9719 |
| [#7][#6H0]=#7 | 0.946 |
| [#7][#6H0][#7] | 0.944 |

| Structure | prob |
|-----------|------|
| [CX4H3]   | 0.9746 |
| [#7][#6](#7)=#7 | 0.9737 |
| [#6H3][#7] | 0.9719 |
| [#7][#6H0]=#7 | 0.946 |
| [#7][#6H0][#7] | 0.944 |

best positives

| Structure | prob |
|-----------|------|
| [CX4H3]   | 0.9746 |
| [#7][#6](#7)=#7 | 0.9737 |
| [#6H3][#7] | 0.9719 |
| [#7][#6H0]=#7 | 0.946 |
| [#7][#6H0][#7] | 0.944 |

| Structure | prob |
|-----------|------|
| [CX4H3]   | 0.9746 |
| [#7][#6](#7)=#7 | 0.9737 |
| [#6H3][#7] | 0.9719 |
| [#7][#6H0]=#7 | 0.946 |
| [#7][#6H0][#7] | 0.944 |

worst negatives

| Structure | prob |
|-----------|------|
| [CX4H3][NX3H1] | 0.4833 |
| [#7][#7] | 0.471 |
| [#7][#6][#6X3] | 0.3816 |
| [#6][#6][#7][#6][#7]1 | 0.20 |
| [#7][#6H0][#6H1] | 0.2613 |
| [#7][#6][#6][#7] | 0.2198 |
| [#7][#7H1] | 0.2163 |
| [#7][#6H3][#6X3] | 0.18 |
| [#6H1] | 0.1377 |

| Structure | prob |
|-----------|------|
| [CX4H3][NX3H1] | 0.4833 |
| [#7][#7] | 0.471 |
| [#7][#6][#6X3] | 0.3816 |
| [#6][#6][#7][#6][#7]1 | 0.20 |
| [#7][#6H0][#6H1] | 0.2613 |
| [#7][#6][#6][#7] | 0.2198 |
| [#7][#7H1] | 0.2163 |
| [#7][#6H3][#6X3] | 0.18 |
| [#6H1] | 0.1377 |
Example 18 true smiles: Cc1=nHc2ccc=cc1 formula: C9H9N
Index of correct structure: 0 of 29511
True structure loss: 0.013431
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: d2O)

Top predicted structures (loss):

| Structure | Loss  |
|-----------|-------|
| ![Structure 1](image1.png) | 0.013431 |
| ![Structure 2](image2.png) | 0.014915 |
| ![Structure 3](image3.png) | 0.015945 |
| ![Structure 4](image4.png) | 0.016185 |
| ![Structure 5](image5.png) | 0.018507 |
| ![Structure 6](image6.png) | 0.019129 |
| ![Structure 7](image7.png) | 0.020373 |
| ![Structure 8](image8.png) | 0.020373 |
| ![Structure 9](image9.png) | 0.021081 |
| ![Structure 10](image10.png) | 0.022922 |
### Top predicted substructures

| Substructure | Prob  |
|--------------|-------|
| [#6H1]       | 0.9998 |
| [#6X3]       | 0.9997 |
| [#6H3]       | 0.9996 |
| [CX4H3]      | 0.9995 |

### Best positives

| Substructure | Prob  |
|--------------|-------|
| [#6H1]       | 0.9998 |
| [#6X3]       | 0.9998 |
| [#6H3]       | 0.9996 |
| [CX4H3]      | 0.9995 |

### Best negatives

| Substructure | Prob  |
|--------------|-------|
| [#6X3H1]     | 0.9958 |
| [#6H0]       | 0.9949 |
| [cH]         | 0.993  |
| [#6X3H1]     | 0.9926 |

### Worst positives

| Substructure | Prob  |
|--------------|-------|
| [#6X3H1]     | 0.6176 |
| [#6H1]       | 0.4272 |
| [cX3H1]      | 0.2115 |
| [cX3H0]      | 0.1673 |
| [#7]         | 0.1381 |
| [#6X3]       | 0.1244 |
| [cX3H0]      | 0.0927 |
| [cH]         | 0.0587 |

### Worst negatives

| Substructure | Prob  |
|--------------|-------|
| [#6X3H1]     | 0.6176 |
| [#6H1]       | 0.4272 |
| [cX3H1]      | 0.2115 |
| [cX3H0]      | 0.1673 |
| [#7]         | 0.1381 |
| [#6X3]       | 0.1244 |
| [cX3H0]      | 0.0927 |
| [cH]         | 0.0587 |
Example 19 true smiles: COc1cccc(N)c1 formula: C7H9NO
Index of correct structure: 1 of 29421
True structure loss: 0.022269
True structure:

Experimental $^{13}$C NMR (solvent: CDCl$_3$)

Experimental $^1$H NMR (solvent: CDCl$_3$)

Top predicted structures (loss):

- 0.020336
- 0.022269
- 0.026384
- 0.029129
- 0.032354
- 0.032459
- 0.036699
- 0.0424
- 0.042488
- 0.04278
| Top predicted substructures | prob     | best positives | prob     | best negatives | prob     | worst negatives | prob     | worst positives | prob     |
|-----------------------------|---------|---------------|---------|---------------|---------|----------------|---------|----------------|---------|
| #6H1 | 0.9992 | [cH][cH] | 0.9391 | #6H1[#6H1] | 0.9992 | [OX2H0][OX2H0][OX2H0][OX2H0][OX2H0] | 0.0 |
| #6X3]#6X3 | 0.9868 | [CX4H3][CX4H3][CX4H3][CX4H3][CX4H3] | 0.9129 | #7H2[#6H0] | 0.9129 | [cX3H1][cX3H1][cX3H1][cX3H1][cX3H1] | 0.0 |
| CX4H3 | 0.9649 | [cX3H1][cX3H1][cX3H1][cX3H1][cX3H1] | 0.9118 | #6X3#6X3 | 0.9118 | [cX3H1][cX3H1][cX3H1][cX3H1][cX3H1] | 0.0 |
| #6X3H1][#6X3H0] | 0.9553 | [CX4H3][CX4H3][CX4H3][CX4H3][CX4H3] | 0.9129 | #7H2[#6H0] | 0.9129 | [cX3H1][cX3H1][cX3H1][cX3H1][cX3H1] | 0.0 |
| #7]#6]#6X3 | 0.95 | [CX4H3][CX4H3][CX4H3][CX4H3][CX4H3] | 0.9129 | #6X3H2 | 0.9129 | [cX3H1][cX3H1][cX3H1][cX3H1][cX3H1] | 0.0 |
| #6X3H2 | 0.2192 | [cX3H1][cX3H1][cX3H1][cX3H1][cX3H1] | 0.7187 | #6X3H2 | 0.2192 | [cX3H1][cX3H1][cX3H1][cX3H1][cX3H1] | 0.0 |
Example 20 true smiles: COc1ccccc1N
Formula: C7H9NO
Index of correct structure: 0 of 29421
True structure loss: 0.017036
True structure:

Experimental $^{13}$C NMR (solvent: CDCl3)

Experimental $^1$H NMR (solvent: CDCl3)

Top predicted structures (loss):

![Chemical structures with loss values]

| Structure | Loss Value |
|-----------|------------|
|           | 0.017036   |
|           | 0.02047    |
|           | 0.022668   |
|           | 0.022859   |
|           | 0.024341   |
|           | 0.027768   |
|           | 0.027878   |
|           | 0.032987   |
|           | 0.034698   |
|           | 0.035239   |
| Top predicted substructures | prob | best positives | prob | best negatives | prob | worst negatives | prob | worst positives | prob |
|-----------------------------|------|----------------|------|----------------|------|-----------------|------|----------------|------|
| [#6H1]                      | 0.9979 | [#6X3H1][#6X3H0] | 0.9036 | [#7X3H1] [#6H2] | 0.3931 | [#7H2][#6H0] | 0.183 |
| [#6X3][#6X3]                | 0.9911 | [cH][cH]        | 0.8995 | [OX2H1][#6H2] | 0.3794 | [#7H2][#6H0] | 0.213 |
| [cH]                        | 0.9658 | [#6H1][#6H1]   | 0.8945 | [OX2H1][#6H2] | 0.3564 | [#7][#6H0][#6H1] | 0.3398 |
| [#7][#6][#6X3]              | 0.9588 | [cX3H1][#6X3H1] | 0.8733 | [OX2H0][#6X3H1] | 0.341 | [#7][#6X3H0][#6X3H1] | 0.4599 |
| [#6X3][#6X3][#6X3][#6X3]   | 0.9265 | [#7][#6][#6][#6X3] | 0.8635 | [#6X3H0][#6X3H1] | 0.3371 | [#7][#6][#6][#6][#6] | 0.5398 |
| [#7][#6][#6X3][#6X3][#6X3] | 0.9036 | [#6H1][#6H1]   | 0.8945 | [#6X3H2][#7H2] | 0.3174 | [#7][#6][#6][#6] | 0.6248 |
| [OX2H1][#6H2]              | 0.9979 | [OX2H0][#6X3H1] | 0.8733 | [OX2H0][#6X3H1] | 0.2989 | [cH][cO]       | 0.6338 |
| [#7][#6][#6][#6][#7]       | 0.9265 | [cX3H1][#6X3H1] | 0.8606 | [#8][#6H0][#6H1] | 0.2977 | [#7][#6][#6][#6][#6][#7] | 0.7336 |
| [#6H1][#6X3][#6X3][#6X3]   | 0.9036 | [cX3H1][#6X3H1] | 0.8733 | [#6X3H2][#7H2] | 0.2977 | [#7][#6][#6][#6][#6][#7] | 0.7413 |
| [cX3H1][#6X3]               | 0.9658 | [cX3H1][#6X3H1] | 0.8733 | [#6X3H2][#7H2] | 0.2977 | [#7][#6][#6][#6][#6][#7] | 0.7598 |
Example 21 true smiles: CCCCCC=CC(=O)O formula: C8H14O2
Index of correct structure: 0 of 28834
True structure loss: 0.008703
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

0.008703 0.039766 0.048218 0.051362 0.055961

0.056054 0.058186 0.06541 0.070585 0.074826
| Top predicted substructures | prob | Best positives | prob | Best negatives | prob |
|-----------------------------|------|----------------|------|---------------|------|
| [CX4H2][{#6}]|{#6} | 1.0 | [CX4H3][{#6}] | 0.9991 | [CX4H3][{#6}] | 0.0 |
| [CX4H3] | 0.9998 | O={#6}[{#6}]={#6X3} | 0.9989 | [CX4H3][{#6}] | 0.0 |
| [CX4H3][CX4H2] | 0.9997 | [CX3]{={OX1}}O | 0.996 | [CX4H3][{#6}] | 0.0 |
| [{#6H3}[{#6}][{#6}] | 0.9995 | [#8]=[#6][{#8}] | 0.9975 | [{#6H1} | 0.9994 |
| [#6H1] | 0.9994 | [#6]#C | 0.0 |
| [CX4H3][{#6}] | 0.9991 | CCCCCC | 0.5444 | [{#6X3}[{#6}][{#6X3}] | 0.5747 |
| [{#6X3}[{#6}][{#6X3}] | 0.3697 | [CX4H2][{CX4H2}][{CX4H2}][{CX4H2}] | 0.5653 | [{#6H3}[{#6}][{#6X3}] | 0.0 |
| [#6H1][{#6}][{#6X3}] | 0.0341 | [CX4H2][{CX4H2}][{CX4H2}] | 0.6269 | [#8][{#6}][{#6X3}] | 0.1584 |
| [CX4H2][{#6}][{#6X3}] | 0.033 | [#8][{#6}][{#6X3}] | 0.6786 | [{#6X3}[{#6}][{#6X3}] | 0.1566 |
| [OX1H0][{OX1H0}][{OX2H1}][{OX2H1}] | 0.0917 | [{#8}]=[{#6H0}][{#6H1}] | 0.8167 | [CX4H3][{#6H1}][{#6H2}] | 0.0723 |
| [OX1H0][{OX1H0}][{OX2H1}][{CX4H1}] | 0.0723 | [{#6H1}][{#6H2}] | 0.823 |
Example 22 true smiles: CCCCCC=CC(=O)O formula: C8H14O2
Index of correct structure: 0 of 28834
True structure loss: 0.008863
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: d2o)

Top predicted structures (loss):

0.008863 0.035556 0.048876 0.053914 0.055939
0.056741 0.063285 0.066607 0.067133 0.078098
| Substructure | Probability | Substructure | Probability |
|--------------|-------------|--------------|-------------|
| [CX4H2][{#6}][#6] | 1.0 | O=[{#6}][#6]=[{#6}][#6] | 0.9993 |
| [CX4H3][CX4H2] | 0.9999 | [CX2H0][{CX2H0}][{CX2H0}] | 0.9991 |
| [CX4H3] | 0.9999 | [CX2H0][{CX2H1}][{CX2H1}] | 0.9983 |
| [CX4H3][#6] | 0.9995 | [{#8}][{#6}][{#6}][{#6}][{#6}] | 0.9977 |
| [#6H3][#6][#6] | 0.9995 | [{#6X3}][{#6X3}] | 0.9963 |
| best positives | prob | best negatives | prob |
| [CX4H2][{#6}][#6] | 1.0 | [{#6X2}][{#6H1}][{#6X2}] | 0.0 |
| [CX4H3][CX4H2] | 0.9999 | CCCCCC | 0.0 |
| [CX4H3] | 0.9999 | [{#7}][{#6H1}][{#6X2}] | 0.0 |
| [CX4H3][#6] | 0.9995 | [CX2H0][{CX2H1}][{CX2H1}][{CX2H1}] | 0.0 |
| [#6H1] | 0.9993 | [CX2H0][{CX2H1}][{CX2H1}] | 0.0 |
| O=[{#6}][#6]=[{#6X3}] | 0.9991 | [CX2H0][{CX2H1}][{CX4H0}] | 0.0 |
| [CX4H3][=C] | 0.9999 | [CX2H0][{CX2H0}][{CX4H0}] | 0.0 |
| [#8][{#6}][{#6H1}]={#6H1} | 0.9977 | [CX2H0][{CX2H1}][{CX4H1}] | 0.0 |
| [#6X3][{#6X3}] | 0.9963 | [{#6H2}][{#6}][{#6X2}] | 0.0 |
| worst negatives | prob | worst positives | prob |
| [#6X3][{#6}][{#6}][#6] | 0.5245 | CCCCCC | 0.2862 |
| [#8][{#6}][#6H2] | 0.2748 | [CX4H2][{CX4H2}][{CX4H2}][{CX4H2}] | 0.5175 |
| [CX4H2][=C=O] | 0.2187 | [CX4H2][{CX3H}] | 0.544 |
| OCC[CH2] | 0.2143 | CCCCCC | 0.5959 |
| [#6H3][{#6}][{#6X3}] | 0.2035 | [CX4H2][{CX4H2}][{CX4H1}] | 0.6958 |
| [CX4H2][{CX4H2}][{CX3H0}] | 0.1177 | [{#6H1}][{#6H2}] | 0.746 |
| [CX3H0][{#O1H0}][{#O2H1}][{CX2H0}][{CX3H0}] | 0.1001 | [CX4H2][{CX4H2}][{CX3H1}] | 0.7582 |
| [CX3H0][{#O1H0}][{#O2H0}][{#O2H0}][{CX2H0}][{CX3H1}] | 0.0912 | [{#8}][{#6}][{#6}][{#6X3}] | 0.8162 |
| [#6H1][{#6H1}] | 0.0748 | [{#8}][{#6}][{#6}][{#6X3}] | 0.8426 |
Example 23 true smiles: NCC(=O)CCC(=O)O formula: C5H9NO3
Index of correct structure: 0 of 27953
True structure loss: 0.019908
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

0.019908  0.034361  0.040346  0.041791  0.043131
0.044076  0.045778  0.048403  0.048685  0.049151
| Top predicted substructures | prob | prob |
|----------------------------|------|------|
| [CX3](=OX1)C               | 1.0  | [CX4H2][CX3]=O | 0.988 |
| [CX4H2][OX1][#6]          | 0.9997 | [OX2H1] | 0.9756 |
| [OX1H0]=C[OX3H0][#6][CX4H2]| 0.9985 | [#7X3][#6H2] | 0.9679 |
| [OX1H0]=C[OX3H0][#6][CX4H2]| 0.9963 | [CX4H2]CC=O | 0.9206 |
| [OX4H2][CX3]=O             | 0.988 | [#7][#6H2] | 0.9169 |

| best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| [CX3](=OX1)C   | 1.0  | CC=CCC#C      | 0.0  |
| [OX1H0]=C[OX3H0][#6][CX4H2]| 0.9997 | CC=CC#CC | 0.0  |
| [OX4H2][OX1][#6] | 0.9985 | C=CC=CC#C | 0.0  |
| [CX4H2][CX3]=O | 0.988 | C=CCC#C      | 0.0  |
| [OX2H1]        | 0.9756 | [CX4H2][CX3]=O | 0.0  |
| [#7X3][#6H2]   | 0.9679 | [CX2H0][#7X3][#6H2] | 0.0  |
| [CX4H2]CC=O    | 0.9206 | [CX2H0][#7X3][#6H2] | 0.0  |
| [#7][#6H2]     | 0.9169 | CCCC#C#C     | 0.0  |

| worst negatives | prob | worst positives | prob |
|-----------------|------|-----------------|------|
| [#7][#6H2][#6H2]| 0.6196 | [CX4H2][NX3H2][CX3H0] | 0.0503 |
| [#7X3H1]        | 0.4542 | [#7H2][#6H2] | 0.1256 |
| [OX4H2][OX1H0][#6] | 0.3455 | [#8][#6][#6H2] | 0.359 |
| [#7][#6][#6X3]  | 0.3301 | [#6X3][#6H2][#7] | 0.3748 |
| [OX1H0]=C[OX3H0][#6][CX4H2]| 0.2928 | [#8][#6][#6][#6][#6]=[#8] | 0.4503 |
| [OX2H1][CX4H2][#6X3H0]| 0.2697 | [#8]=[#6][#6][#6][#6][#6]=[#8] | 0.6313 |
| [#7H2][#6H0]    | 0.2321 | [#7][#6][#6X3] | 0.6397 |
| [#8][#6][#6H2][#8] | 0.2189 | [#7X3H2] | 0.7007 |
| [OX1H0]=C[OX3H0][#8][#6] | 0.211 | [OX1H0]=C[OX3H0][#8][#6] | 0.7035 |
| [#6H2][#7][#6X3] | 0.1922 | [#6X3H0][CX4H2][CX4H2][#6X3H0] | 0.7186 |
Example 24 true smiles: O=C(O)C1CC(O)CN1 formula: C5H9NO3
Index of correct structure: 0 of 27953
True structure loss: 0.032393
True structure:
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [#6H1]                      | 0.9989 | [#6H1][#6H2]   | 0.9335 |                  |      |
| [CX4H2][#6]                | 0.9987 | [O2H1]         | 0.922 |                  |      |
| [OX1][#6H2]                | 0.9942 | [#8]=[#6H0][#6H1] | 0.8957 |                  |      |
| [#7X3][#6H2]               | 0.9414 | [C3]==[OX1]    | 0.8849 |                  |      |
| best positives             | prob | best negatives | prob | worst positives | prob |
| [#6H1]                      | 0.9989 | [C2H1][C2H0][C3H1]=[C3H0] | 0.0 |                  |      |
| [CX4H2][#6]                | 0.9987 | [C4H0]==[C4H2][C4H3][C4H2] | 0.0 |                  |      |
| [CX1][=O][#6]              | 0.9942 | [C4H0]=[C4H1][C4H2][C4H0] | 0.0 |                  |      |
| [#8]=[#6][#8]              | 0.9446 | [#6H3][#6X3][#6X3]=[#6X3H2] | 0.0 |                  |      |
| [#7X3][#6H2]               | 0.9414 | C=CC=CCMC      | 0.0 |                  |      |
| [#6H1][#6H2]               | 0.9335 | [C4H0][C4H2][C4H1] | 0.0 |                  |      |
| [OX2H1]                     | 0.922  | [#6H3][#6H1][#6H1]=[#7]  | 0.0 |                  |      |
| [#8]=[#6H0][#6H1]          | 0.8957 | [C3H0]=[C3H0][C4H1][C4H2] | 0.0 |                  |      |
| OCC[C2H2]                  | 0.8849 | [C2H0][C2H0][C3H0] | 0.0 |                  |      |
| [CX1][=O][#X1]             | 0.8714 | [#6X3][#6][#6][#6H3] | 0.0 |                  |      |
| worst negatives            | prob | worst positives | prob | worst negatives | prob |
| [#8][#6][#6][#6X3]         | 0.5012 | [C4H2][NX3H1][C4H1] | 0.164 |                  |      |
| [CX4H2][C4H2][C4H1]        | 0.4633 | [#6H1][#6H2][#6H2] | 0.2168 |                  |      |
| [C4H2][#6][O]              | 0.4506 | [#7][#6H2][#6H1] | 0.2253 |                  |      |
| [C4H2][C4H2]               | 0.4158 | [C3]=[O][O2H1] | 0.2954 |                  |      |
| [#7X3H2]                   | 0.4148 | C4H2)=[CH]([CH] | 0.3104 |                  |      |
| [CX4H2][C3]=O              | 0.3846 | [C4H1][O2H1][C4H2][C4H2] | 0.3143 |                  |      |
| [#8][#6H1][#6H1]           | 0.3578 | [#6H1][#6H2][#6][#6][#7] | 0.3234 |                  |      |
| [#7][#6][#6][#6X3]         | 0.2444 | [#6][#6][#6][#6][#7] | 0.3717 |                  |      |
| [C4H1][NX3H2][C4H2][C3H0]  | 0.2364 | [#7X3H1]      | 0.3754 |                  |      |
Example 25 true smiles: O=C(O)C1CC(O)CH1 formula: C5H9NO3
Index of correct structure: 0 of 27953
True structure loss: 0.032454
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

0.032454  0.039223  0.042535  0.047224  0.052116
0.053022  0.053465  0.053893  0.05523  0.056236
| Substructure | prob | best positives | prob | best negatives | prob |
|--------------|------|----------------|------|----------------|------|
| #6H1        | 0.9954 | [CX4H]O | 0.9004 | [CX3](=O)[OX2H1] | 0.8361 |
| [CX3](=O)[OX1]C | 0.9942 | O[CX4H][CX4H2] | 0.8266 | [CX2H1]O[CX3H0][CX4H1] | 0.8266 |
| [CX4H2][#6][#6] | 0.9744 | [CX2H1]O[CX4H1] | 0.8266 | [CX2H1][CX3H1][CX4H2] | 0.8266 |
| #6H1][#6H2] | 0.96 | OCC[CX3H] | 0.7816 | [CX4H2][#6][#6H2] | 0.8266 |
| #6H1]#6H2 | 0.96 | [CX3H0][CX3H0]#CX4H | 0.7816 | [OX2H1]#[OX2H1][OX2H1] | 0.7816 |
| [CX4H]O | 0.9004 | [CX3H0][CX3H0][CX4H1] | 0.7816 | [CX3H0][CX3H0][CX4H1] | 0.7816 |
| O[CX4H][CX4H2] | 0.8361 | [CX3H0][CX3H0][CX4H1] | 0.7816 | [CX3H0][CX3H0][CX4H1] | 0.7816 |
| #6H1][#6H2] | 0.8266 | [CX3H0][CX3H0][CX4H1] | 0.7816 | [CX3H0][CX3H0][CX4H1] | 0.7816 |
| [CX4H2][#6H1] | 0.7606 | [CX3H0][CX3H0][CX4H1] | 0.7816 | [CX3H0][CX3H0][CX4H1] | 0.7816 |

worst negatives
| Substructure | prob | worst positives | prob |
|--------------|------|----------------|------|
| #6H1][#6H1] #6H1 | 0.4266 | [CX3][=O][OX2H1] | 0.14 |
| #6][#6]=#6 | 0.4079 | [#7H1][#6X4H1][#6X3] | 0.2058 |
| #6][#6][#6] #6X3 | 0.3946 | [OX2H1][([NX3H1])[OX2H1]] | 0.2465 |
| [#6H1][#6H1] | 0.3982 | [OX2H1][([OX2H1])[OX2H1]] | 0.2565 |
| #7H3H2 | 0.3711 | [#8][#6][#6] #6 | 0.2721 |
| [CX4H2][#6H1] | 0.3122 | [OX2H1][([OX2H1])[OX2H1]] | 0.3053 |
| #6H1][#6H1][#6] | 0.3048 | [OX2H1][([NX3H1])[OX2H1]] | 0.3349 |
| [#7][#6][#6X3] | 0.2688 | [#6][#6H1][#6H1] | 0.3536 |
| O[CX4H][CX4H2][CX4H1] | 0.2429 | [CX3](=O)[OX1] | 0.4072 |
Example 26 true smiles: O=C(O)C(=O)c1ccccc1 formula: C₈H₈O₂

Experimental ¹³C NMR (solvent: CDCl₃)

Experimental ¹H NMR (solvent: d₂O)

Top predicted structures (loss):

- 0.023907
- 0.042161
- 0.046921
- 0.054306
- 0.054605
- 0.055526
- 0.058364
- 0.058816
- 0.059834
- 0.060167
| Substructures | prob | Substructures | prob |
|---------------|------|---------------|------|
| [#6H1]        | 0.9973 | [CX3]==[OX1]C | 0.9338 |
| [#6X3][#6H2][#6X3] | 0.9812 | [#6X3H1][#6X3H0] | 0.8959 |
| [#6X3]         | 0.977  | [#8X1]=[#6X3][#6H2][#6H0] | 0.8538 |
| O=[#6][#6][#6X3] | 0.9614 | [#6X3][#6X3][#6X3][#6X3] | 0.7715 |
| [CX4H2][#6][#6] | 0.9439 | [#8]=[#6][#8] | 0.757 |

**Best Positives**

| Substructures | prob | Substructures | prob |
|---------------|------|---------------|------|
| [#6H1]        | 0.9973 | [CX4H1][NX3H2][X4H3][X4H1] | 0.0 |
| [#6X3][#6H2][#6X3] | 0.9812 | [CX4H0][NX3H1][X4H3][X4H2][X4H1] | 0.0 |
| [#6X3]         | 0.977  | [#6H3][#6H1][#7][#7] | 0.0 |
| O=[#6][#6][#6X3] | 0.9614 | [CX4H1][OX2H1][CX4H3][CX4H0] | 0.0 |
| [CX4H2][#6][#6] | 0.9439 | [#8][#6H2][#6H2][#6X2] | 0.0 |
| [CX3][=OX1]C | 0.9338 | [CX4H1][NX3H2][X4H2][X4H0] | 0.0 |
| [#6X3H1][#6X3H0] | 0.8959 | [CX4H2][NX2H0][X4H1] | 0.0 |
| [#8X1]=[#6X3][#6H2][#6H0] | 0.8538 | [CX4H1][NX3H0][X4H3][CX4H1] | 0.0 |
| [#6X3][#6X3][#6X3][#6X3] | 0.7715 | [CX4H2][NX3H1][X4H3] | 0.0 |
| [#8]=[#6][#8] | 0.757 | [CX4H1][NX3H1][X4H3][X4H2] | 0.0 |

**Best Negatives**

| Substructures | prob | Substructures | prob |
|---------------|------|---------------|------|
| [CHX3][=C]C  | 0.3903 | [#6][#6][#6][#6][#6][#6][#6][#6] | 0.1447 |
| [CHX3][CHX3] | 0.3433 | [#6X3H1][#6X3H1][#6X3H0][#6X3H1] | 0.1576 |
| [#6X3][#6X3]=[#6X3][#6X3] | 0.3093 | [CX3H0][=OX1H0][OX2H1][X4H2] | 0.1624 |
| [#8]=[#6H0][#6H1] | 0.2914 | [cX3H1][cX3H1][cX3H1] | 0.1718 |
| O=[#6][#6]=[#6X3] | 0.2842 | [OX2H1] | 0.315 |
| [OX1H0]=CX3H0][X4H2][CX3H0] | 0.2613 | [#8]=[#6][#6H2] | 0.3455 |
| [CX4H2][=CX3H0][CX3H0] | 0.2419 | [CX3][=O][OX2H1] | 0.4395 |
| [CX3H1][=CX3H1][=CX3H0] | 0.2409 | [cH][cH] | 0.4548 |
| [#6H1][#6H2] | 0.2295 | [cX3H1][cX3H1][cX3H0] | 0.5049 |
| [OX1H0]=CX3H0][CX3H1][=CX3H1] | 0.2126 | [#6H1][#6H1] | 0.6281 |
Example 27 true smiles: O=C(O)CCCCC(=O)O formula: C6H10O4
Index of correct structure: 0 of 19323
True structure loss: 0.011026
True structure:
Example 28 true smiles: CC(CC(=O)O)CC(=O)O formula: C6H10O4
Index of correct structure: 0 of 19323
True structure loss: 0.027036
True structure:

![Chemical structure](image)

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

1. ![Structure 1](image) 0.027036
2. ![Structure 2](image) 0.05023
3. ![Structure 3](image) 0.050384
4. ![Structure 4](image) 0.053722
5. ![Structure 5](image) 0.058615
6. ![Structure 6](image) 0.060752
7. ![Structure 7](image) 0.062412
8. ![Structure 8](image) 0.064204
9. ![Structure 9](image) 0.066972
10. ![Structure 10](image) 0.06846
| Top predicted substructures                  | prob  | best positives                  | prob  | best negatives                  | prob  |
|---------------------------------------------|-------|---------------------------------|-------|---------------------------------|-------|
| [OX2H1]                                     | 0.9997| [CX2H1] # [CX2H0] # [CX3H1] = [CX3H0] | 0.0   | [OX2H1] # [CX2H1] # [CX2H0] | 0.0   |
| [#8]=[#6] # [#6]                             | 0.9996| [OX1] # [OX2H0] # [OX2H1] | 0.0   | [OX1] # [OX2H1] | 0.0   |
| [CX3] (= [OX1]) C                            | 0.9993| [CX2H1] # [CX2H0] # [CX3H1] = [CX3H0] | 0.0   | [CX2H1] # [CX2H0] | 0.0   |
| [CX3] (=O) [OX2H1]                           | 0.9991| [OX1] # [OX2H1] # [OX2H0] | 0.0   | [OX1] # [OX2H1] | 0.0   |
| [OX2H1]                                     | 0.9997| [CX2H1] # [CX2H0] # [CX3H1] = [CX3H0] | 0.0   | [OX2H1] # [CX2H1] | 0.0   |
| [#8]=[#6] # [#6]                             | 0.9996| [OX1] # [OX2H0] # [OX2H1] | 0.0   | [OX1] # [OX2H1] | 0.0   |
| [CX3] (= [OX1]) C                            | 0.9993| [OX1] # [OX2H0] # [OX2H1] | 0.0   | [OX1] # [OX2H1] | 0.0   |
| [CX3] (=O) [OX2H1]                           | 0.9991| [OX1] # [OX2H0] # [OX2H1] | 0.0   | [OX1] # [OX2H1] | 0.0   |
| [#8]=[#6] # [#6]                             | 0.9996| [OX1] # [OX2H0] # [OX2H1] | 0.0   | [OX1] # [OX2H1] | 0.0   |
| [CX4H2] ([#6]) # [#6]                       | 0.9692| [OX1] # [OX2H0] # [OX2H1] | 0.0   | [CX2H0] # [OX1] | 0.0   |
| [CX3H1] (=# [CX3H1]) # [#6]                  | 0.9785| [OX2H0] # [CX2H0] # [CX3H1] = [CX3H0] | 0.0   | [CX2H0] # [OX1] | 0.0   |
| [8]=# [6] # # [8]                           | 0.9785| [OX2H0] # [CX2H0] # [CX3H1] = [CX3H0] | 0.0   | [CX2H0] # [OX1] | 0.0   |
| [CX3H1] (=# [CX3H1]) # [#6]                  | 0.9785| [OX2H0] # [CX2H0] # [CX3H1] = [CX3H0] | 0.0   | [CX2H0] # [OX1] | 0.0   |
| worst negatives                              | prob  | worst positives                | prob  | worst negatives                | prob  |
| [#8]=[#6] # [#6]                             | 0.8976| [CX2H1] # [CX2H0] # [CX3H1] = [CX3H0] | 0.0   | [OX1] # [OX2H1] # [OX2H0] | 0.0   |
| [OX2H1]                                     | 0.8976| [CX2H1] # [CX2H0] # [CX3H1] = [CX3H0] | 0.0   | [OX1] # [OX2H1] # [OX2H0] | 0.0   |
| [CX3H0] (=# [OX1H1]) ([OX2H1]) ([OX4H1])     | 0.8575| [CX2H1] # [CX2H0] # [CX3H1] = [CX3H0] | 0.0   | [OX1] # [OX2H1] # [OX2H0] | 0.0   |
| [OX2H1]                                     | 0.8575| [CX2H1] # [CX2H0] # [CX3H1] = [CX3H0] | 0.0   | [OX1] # [OX2H1] # [OX2H0] | 0.0   |
| [OX1] # [OX2H0] # [OX2H1]                   | 0.7204| [CX2H1] # [CX2H0] # [CX3H1] = [CX3H0] | 0.0   | [OX1] # [OX2H1] # [OX2H0] | 0.0   |
| [OX2H1]                                     | 0.7204| [CX2H1] # [CX2H0] # [CX3H1] = [CX3H0] | 0.0   | [OX1] # [OX2H1] # [OX2H0] | 0.0   |
| [OX2H1]                                     | 0.5838| [CX2H1] # [CX2H0] # [CX3H1] = [CX3H0] | 0.0   | [OX1] # [OX2H1] # [OX2H0] | 0.0   |
| [OX2H1]                                     | 0.5838| [CX2H1] # [CX2H0] # [CX3H1] = [CX3H0] | 0.0   | [OX1] # [OX2H1] # [OX2H0] | 0.0   |
| [OX2H1]                                     | 0.5367| [CX2H1] # [CX2H0] # [CX3H1] = [CX3H0] | 0.0   | [OX1] # [OX2H1] # [OX2H0] | 0.0   |
| [OX2H1]                                     | 0.5367| [CX2H1] # [CX2H0] # [CX3H1] = [CX3H0] | 0.0   | [OX1] # [OX2H1] # [OX2H0] | 0.0   |
| [OX2H1]                                     | 0.5217| [OX2H0] # [OX1H0] # [OX2H1] = [OX2H0] | 0.0   | [OX1] # [OX2H1] # [OX2H0] | 0.0   |
| [OX2H1]                                     | 0.5217| [OX2H0] # [OX1H0] # [OX2H1] = [OX2H0] | 0.0   | [OX1] # [OX2H1] # [OX2H0] | 0.0   |
| [OX2H1]                                     | 0.4633| [OX2H0] # [OX1H0] # [OX2H1] = [OX2H0] | 0.0   | [OX1] # [OX2H1] # [OX2H0] | 0.0   |
| [OX2H1]                                     | 0.4633| [OX2H0] # [OX1H0] # [OX2H1] = [OX2H0] | 0.0   | [OX1] # [OX2H1] # [OX2H0] | 0.0   |
| [OX2H1]                                     | 0.4464| [OX1H0] # [OX2H1] # [OX2H0] | 0.0   | [OX1H0] # [OX2H1] # [OX2H0] | 0.0   |
| [OX2H1]                                     | 0.4464| [OX1H0] # [OX2H1] # [OX2H0] | 0.0   | [OX1H0] # [OX2H1] # [OX2H0] | 0.0   |
| [OX2H1]                                     | 0.2848| [OX1H0] # [OX2H1] # [OX2H0] | 0.0   | [OX1H0] # [OX2H1] # [OX2H0] | 0.0   |
| [OX2H1]                                     | 0.2848| [OX1H0] # [OX2H1] # [OX2H0] | 0.0   | [OX1H0] # [OX2H1] # [OX2H0] | 0.0   |
| [OX2H1]                                     | 0.2233| [OX1H0] # [OX2H1] # [OX2H0] | 0.0   | [OX1H0] # [OX2H1] # [OX2H0] | 0.0   |
| [OX2H1]                                     | 0.2233| [OX1H0] # [OX2H1] # [OX2H0] | 0.0   | [OX1H0] # [OX2H1] # [OX2H0] | 0.0   |
Example 29 true smiles: COC\(\text{\(\equiv\)}\text{O})\text{CCC(\(\equiv\)}\text{O})\text{O}

Formula: C_6H_{10}O_4

Index of correct structure: 0 of 19323

True structure loss: 0.011848

True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: d2o)

Top predicted structures (loss):

- 0.011848
- 0.029618
- 0.031528
- 0.034056
- 0.048556
- 0.049282
- 0.050412
- 0.050809
- 0.055695
- 0.055804
Example 30 true smiles: CN1CC(=O)NCl=N formula: C4H7N3O
Index of correct structure: 0 of 16898
True structure loss: 0.02136
True structure:

Experimental 13C NMR (solvent: DMSO)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

- 0.02136
- 0.028479
- 0.030813
- 0.031246
- 0.032402

- 0.032854
- 0.035052
- 0.035694
- 0.036352
- 0.036589
| Top predicted substructures | prob | best positives | prob | worst negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [#7][#6H0]=[#7] | 0.9591 | [#7][#6[#6][#7]] = [#7] | 0.9161 | [#7][#6][#6H0] | 0.9137 |
| [#7][#6][#6H3] | 0.9448 | [#6][#6][#7] | 0.9137 | [#6][#6][#6H3] | 0.9137 |
| [#6][#6H3][#7] | 0.9423 | [#6][#6H3][#7] | 0.9137 | [#6][#6H3][#6][#7] | 0.9137 |
| [#7][#6][#6H3] | 0.8906 | [#6][#6H3][#7] | 0.9137 | [#6][#6H3][#6][#7] | 0.9137 |
| [#6][#6H3][#7] | 0.8653 | [#6][#6H3][#7] | 0.9137 | [#6][#6H3][#6][#7] | 0.9137 |

| best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| [#7][#6H0]=[#7] | 0.9591 | [#7][#6][#6][#7] = [#6][#6][#7] | 0.0 |
| [#7][#6][#6H0] | 0.9448 | [#6][#6][#7] = [#6][#6][#7] | 0.0 |
| [#6][#6H3][#7] | 0.9423 | [#6][#6H3][#7] | 0.0 |
| [#6H3][#7] | 0.935 | [#6][#6][#7] | 0.0 |
| [#6][#6H3][#7] | 0.8906 | [#6][#6H3][#7] | 0.0 |
| [#7][#6H3][#6][#7] | 0.8653 | [#6][#6H3][#7] | 0.0 |

| worst negatives | prob | worst positives | prob |
|-----------------|------|-----------------|------|
| [#7][#6][#6][#6X3] | 0.3453 | [#6H][#6][#6][#7] | 0.1109 |
| [#6H1] | 0.2854 | [#6H][#6][#6][#7] | 0.116 |
| [#6X3] | 0.2757 | [#6H][#6][#6][#7] | 0.2475 |
| [#7][#6H0][#6][#6X3] | 0.2753 | [#6H][#6][#6][#7] | 0.2662 |
| [#7][#6H0][#6X3] | 0.2702 | [#6H][#6][#6][#7] | 0.473 |
| [#6X3][#6][#6X3] | 0.2385 | [#6H][#6][#6][#7] | 0.5192 |
| [#7][#6H0][#6H1] | 0.2225 | [#6H][#6][#6][#7] | 0.5537 |
| [#7][#6H0][#6][#6X3] | 0.1786 | [#6H][#6][#6][#7] | 0.5837 |
| O=[#6][#6][#6X3] | 0.1695 | [#6H][#6][#6][#7] | 0.6003 |
| [C4X4H2][#6][#6] | 0.1655 | [#6H][#6][#6][#7] | 0.656 |
Example 31 true smiles: N=C(N)NCC(=O)O formula: C3H7N3O2
Index of correct structure: 6 of 15726
True structure loss: 0.026552
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

- 0.020384
- 0.021752
- 0.026122
- 0.026445
- 0.026499
- 0.026551
- 0.026552
- 0.027227
- 0.027468
- 0.027487
| Top predicted substructures | prob  | best positives | prob  | best negatives | prob  |
|-----------------------------|-------|----------------|-------|----------------|-------|
| [#7][#6H2]                  | 0.9313| [CX2H0][#6H0][#7H2] | 0.742 | [CX2H0](=O)[CX2H0] | 0.0   |
| [#7X3H2]                    | 0.9283| [OX1H0][#7X3H0][#7] | 0.7324| [OX1H0][#7X3H0][#7] | 0.0   |
| [#7X3][#6H2]                | 0.856 | [CX2H0][#6H0][#7H2] | 0.7294| [CX2H0][#6H0][#7H2] | 0.0   |
| [CX3](=O)[OX1]C             | 0.8397| [CX2H0][#6H0][#7H2] | 0.6917| [CX2H0][#6H0][#7H2] | 0.0   |
| [#6X3][#6H2][#7]            | 0.82  | [CX2H0][#6H0][#7H2] | 0.6779| [CX2H0][#6H0][#7H2] | 0.0   |
| best positives              | prob  | best negatives  | prob  | worst positives | prob  |
| [#7][#6H2]                  | 0.9313| [CX2H0][#6H0][#7H2] | 0.0   | [CX2H0][#6H0][#7H2] | 0.0   |
| [#7X3H2]                    | 0.9283| [OX1H0][#7X3H0][#7] | 0.0   | [OX1H0][#7X3H0][#7] | 0.0   |
| [#7X3][#6H2]                | 0.856 | [CX2H0][#6H0][#7H2] | 0.0   | [CX2H0][#6H0][#7H2] | 0.0   |
| [CX3](=O)[OX1]C             | 0.8397| [CX2H0][#6H0][#7H2] | 0.0   | [CX2H0][#6H0][#7H2] | 0.0   |
| [#6X3][#6H2][#7]            | 0.82  | [CX2H0][#6H0][#7H2] | 0.0   | [CX2H0][#6H0][#7H2] | 0.0   |
| worst negatives             | prob  | worst positives | prob  | worst positives | prob  |
| [#7][#6][#6][#7]            | 0.4249| [CX3H0][=O][OX1H0][#7][#7] | 0.1549| [CX3H0][=O][OX1H0][#7][#7] | 0.0   |
| [#7H2][#6H2]                | 0.4205| [CX3][=O][OX2H0] | 0.1707| [CX3][=O][OX2H0] | 0.0   |
| [CX4H3]                     | 0.375 | [OX2H0] | 0.2216| [OX2H0] | 0.0   |
| [#6X3][#7X3][#6X3]          | 0.3605| [CX2H0][#6H0][#7H2] | 0.2506| [CX2H0][#6H0][#7H2] | 0.0   |
| [#6X3][#7][#6X3]            | 0.2295| [OX1H0][#6H0][#7H2] | 0.3006| [OX1H0][#6H0][#7H2] | 0.0   |
| [CX4H3][OX2H0]              | 0.2196| [OX1H0][#6H0][#7H2] | 0.3498| [OX1H0][#6H0][#7H2] | 0.0   |
| [#7X3H0]                    | 0.2149| [CX3][=O][OX1H0][#7][#7] | 0.3662| [CX3][=O][OX1H0][#7][#7] | 0.0   |
| O=#6][#6][#6X3]             | 0.2102| [NH1][#6][#7] | 0.37 | [NH1][#6][#7] | 0.0   |
| [CX4H3][#X3H0]              | 0.1828| [#7][#6][#7] | 0.4565| [#7][#6][#7] | 0.0   |
| [CX3H0][=O][OX1H0][#6X3H2]  | 0.1801| [#8][#6][#8] | 0.4577| [#8][#6][#8] | 0.0   |
Example 32 true smiles: O=C(O)c1ccc(O)cc1 formula: C7H6O3
Index of correct structure: 0 of 15458
True structure loss: 0.015458
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

0.015458  0.016028  0.016751  0.023018  0.027588
0.029742  0.029754  0.030721  0.031412  0.031722
| Top predicted substructures | prob   | best positives | prob   | best negatives | prob   |
|-----------------------------|--------|----------------|--------|----------------|--------|
| [#6X3][#6X3]               | 0.9996 | [#6X3][#6X3][#6X3][#6X3] | 0.9746 |
| [#6H1]                     | 0.9944 | [#6X3][#6X3][#6X3][#6X3] | 0.9746 |
| [#8]=[#6][#8]             | 0.995  | O=#6][#6][#6X3] | 0.9355 |
| [cH][cH]                   | 0.9895 | [cX3H1][cX3H1][cX3H0] | 0.9325 |
| [#6X3H1][#6X3H0]          | 0.9843 | [cH]            | 0.8623 |

| worst negatives            | prob   | worst positives | prob   |
|-----------------------------|--------|----------------|--------|
| [cX3H1][cX3H1][cX3H1]      | 0.7692 | [#6][#6][#6][#6][#6] | 0.4336 |
| [cX3]=[#O]                 | 0.5981 | OX2H][cX3][c]    | 0.4412 |
| [#8][#6][#6H1]            | 0.5847 | [cX3H0][cX3H1][cX3H1][cX3H0] | 0.5008 |
| [OX1H0]=[#cX3H0][cX3H1]   | 0.5116 | [#6X3H1][#6X3][#6X3][#6X3] | 0.5032 |
| O=ox3]                     | 0.4958 | [cH][O]          | 0.5104 |
| [#8][#6][#6][#6X3]        | 0.3654 | [cX3][=O][OX2H1] | 0.7037 |
| [#8][#6][#6H1][#6H1]      | 0.295  | OX2H1          | 0.7737 |
| [#8][#6][#6H1][#6H1]      | 0.2343 | [#6][#6][#6H1]   | 0.7992 |
| [#8][#6][#6H1][#6H1]      | 0.1717 | [cX3H1][=O][#6X3] | 0.8104 |
| [#8][#6][#6][#6][#8]      | 0.164  | [#8][#6H0][#6H1] | 0.8359 |
Example 33 true smiles: CCC(C)C(H)C(=O)O formula: C6H13NO2
Index of correct structure: 0 of 14628
True structure loss: 0.020782
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: #N/A)

Top predicted structures (loss):

0.020782
0.032973
0.035784
0.038178
0.040487
0.04087
0.041271
0.041349
0.041494
0.042415
### Top predicted substructures

| Structure                  | prob  |
|----------------------------|-------|
| [CX4H3]                    | 1.0   |
| [CX4H3][#6][#6]            | 0.9993|
| [OX2H1]                    | 0.951 |
| [CX4H3][#6]                | 0.9978|
| [CX4H2][#6][#6]            | 0.9951|
| [OX2H1][#6H1]              | 0.9104|
| [CX4H3][#6][#6]            | 0.9978|
| [CX4H2]([#6][#6])          | 0.9951|
| [CX4H3][#6][#6]            | 0.9978|
| [CX4H3][#6][#6]            | 0.9951|
| [CX4H3][CX4H2]             | 0.9696|
| [CX4H3][CX4H1]             | 0.8132|
| [CX4H3][#6][#6]            | 0.9993|
| [OX2H1][#6H1]              | 0.9104|
| [CX4H3][#6][#6]            | 0.9978|
| [OX2H1][#6H1]              | 0.9104|
| [CX4H3][#6][#6]            | 0.9978|
| [OX2H1][#6H1]              | 0.9104|
| [CX4H3][CX3H1]             | 0.1391|
| [CX4H3][#6][#6]            | 0.9993|
| [CX4H3][#6][#6]            | 0.9978|
| [OX2H1][#6H1]              | 0.9104|
| [CX4H3][#6][#6]            | 0.9978|
| [OX2H1][#6H1]              | 0.9104|
| [CX4H3][#6][#6]            | 0.9978|
| [OX2H1][#6H1]              | 0.9104|
| [CX4H3][#6][#6]            | 0.9978|
| [OX2H1][#6H1]              | 0.9104|
| [CX4H3][#6][#6]            | 0.9978|
| [OX2H1][#6H1]              | 0.9104|
| [CX4H3][#6][#6]            | 0.9978|
| [OX2H1][#6H1]              | 0.9104|

### best positives

| Structure                  | prob  |
|----------------------------|-------|
| [OX2H1]                    | 0.951 |
| [#6H1][#6][#6]             | 0.9993|
| [CX2H0]                    | 0.9978|
| [OX2H1][#6H1][#6H2]        | 0.9951|
| [CX2H0][#6H1][#6H2]        | 0.9701|
| [CX2H0][#6H1][#6H2]        | 0.9696|
| [OX2H1][#6H1][#6H2]        | 0.951 |
| [CX2H0][#6H1][#6H2]        | 0.9104|
| [CX2H0][#6H1][#6H2]        | 0.8587|
| [OX2H1][#6H1][#6H2]        | 0.8132|

### worst negatives

| Structure                  | prob  |
|----------------------------|-------|
| [CX4H2][CX4H2]             | 0.6286|
| [CX4H2][CX4H2]             | 0.5128|
| [CX4H2][CX4H2]             | 0.3307|
| [CX4H2][CX4H2]             | 0.3005|
| [CX4H2][CX4H2]             | 0.2648|
| [CX4H2][CX4H2]             | 0.2337|
| [CX4H2][CX4H2]             | 0.2278|
| [CX4H2][CX4H2]             | 0.2163|
| [CX4H2][CX4H2]             | 0.2044|
| [CX4H2][CX4H2]             | 0.1429|
Example 34 true smiles: CCC(C)(N)C(=O)O formula: C6H13NO2
Index of correct structure: 0 of 14628
True structure loss: 0.020715
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: d2o)

Top predicted structures (loss):

- \( \text{HO-CONH} \)
- \( \text{COOH-NH} \)
- \( \text{NH}_2\text{CHCH}_2\text{COOH} \)
- \( \text{HO-CONH} \)
- \( \text{NH}_2\text{CHCH}_2\text{COOH} \)

0.020715 0.03305 0.035674 0.038283 0.040448

- \( \text{HO-CONH} \)
- \( \text{HO-CONH} \)
- \( \text{HO-CONH} \)
- \( \text{HO-CONH} \)
- \( \text{HO-CONH} \)

0.040841 0.041151 0.04141 0.041454 0.042373
| Top predicted substructures | prob | best positives | prob | worst negatives | prob |
|----------------------------|------|----------------|------|-----------------|------|
| [CX4H3]                    | 1.0  | [CX4H3][CX4H2] | 0.9698 | OCC[CH2]        | 0.6303 |
| [#6H3][#6][#6]             | 0.9993 | [OX2H1]        | 0.951 | [CX4H2][CC=O]  | 0.5085 |
| [CX4H3][#6]                | 0.9978 | [#6H1]         | 0.912 | [#7H2][#6H0]   | 0.3281 |
| [CX4H2][#6][#6]            | 0.9952 | [#6H1][#6H2]  | 0.8584 | [CX4H2][#7][#6H0][#6H1] | 0.2309 |
| [CX3][#6]                  | 0.9698 | #OCC[CH2]      | 0.1428 | [CX4H2][#6H3][#6H1] | 0.2986 |
| [CX4H3][#6H1]              | 0.9978 | C=CC=CC=CC     | 0.0  | [CX4H2][#7][#6H0][#6H1] | 0.2681 |
| [CX3][O]                   | 0.9993 | [CX2H0][#7]    | 0.0  | [#6H3][#6][#6][#6H3] | 0.2298 |
| [#6H1][#6H2]               | 0.912  | C=CC=CC=CC     | 0.0  | [CX4H2][#7][#6H0][#6H1] | 0.2201 |
| [CX4H3][CX4H1]             | 0.8102 | #C=CC=CC=CC    | 0.0  | [CX4H2][#6H3][#6H1] | 0.206 |
| [CX2H1]                    | 0.9510 | C=CC=CC=CC     | 0.0  | [CX4H2][#6H3][#6H1] | 0.1459 |
| [#6H1][#6H2]               | 0.912  | C=CC=CC=CC     | 0.0  | [CX4H2][#6H3][#6H1] | 0.0  |
| [OX2H1]                    | 0.951  | C=CC=CC=CC     | 0.0  | [CX4H2][#6H3][#6H1] | 0.0  |
| [CX4H3][CX4H1]             | 0.8102 | C=CC=CC=CC     | 0.0  | [CX4H2][#6H3][#6H1] | 0.0  |

**Top predicted substructures probabilities:**

| Substructure | Probability |
|--------------|-------------|
| [CX4H3]       | 1.0         |
| [CX4H3][#6]   | 0.9993      |
| [CX4H3][#6][#6] | 0.9978    |
| [CX4H2][#6][#6] | 0.9952   |
| [CX3][#6]     | 0.9698      |
| [CX4H3][#6H1] | 0.9978      |
| [CX3][O]      | 0.9993      |

**Best predicted positives probabilities:**

| Substructure | Probability |
|--------------|-------------|
| [CX4H3][CX4H2] | 0.9698 |
| [OX2H1]        | 0.951  |
| [CX2H0][#7]    | 0.0    |
| [CX4H2][#6H3][#6H1] | 0.8584 |
| [CX4H2][#6H3]  | 0.8102  |

**Best predicted negatives probabilities:**

| Substructure | Probability |
|--------------|-------------|
| OCC[CH2]     | 0.6303      |
| [CX4H2][CC=O]| 0.5085     |
| [CX4H2][#6H0]| 0.3281     |
| [CX4H2][#6][#6] | 0.2986  |
| [CX4H2][#6H1] | 0.2681     |
| [CX4H2][#6H0][#6H1]| 0.2309  |
| [CX4H2][#6][#6][#6H3]| 0.2298  |
| [CX4H2][#6][#6][#6H3]| 0.2201  |

**Worst predicted positives probabilities:**

| Substructure | Probability |
|--------------|-------------|
| OCC[CH2]     | 0.6303      |
| [CX4H2][CC=O]| 0.5085     |
| [CX4H2][#6H0]| 0.3281     |
| [CX4H2][#6][#6] | 0.2986  |
| [CX4H2][#6H1] | 0.2681     |
| [CX4H2][#6H0][#6H1]| 0.2309  |
| [CX4H2][#6][#6][#6H3]| 0.2298  |
| [CX4H2][#6][#6][#6H3]| 0.2201  |
Example 35 true smiles: CCC(C)(N)C(=O)O formula: C6H13NO2
Index of correct structure: 0 of 14628
True structure loss: 0.021957
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

1. \(\text{HO}\text{C}\text{NH}_2\text{C}\text{HN}\text{C}\text{O}\text{HO}\) 0.021957
2. \(\text{HO}\text{C}\text{NH}_2\text{C}\text{HN}\text{C}\text{O}\text{HO}\) 0.032026
3. \(\text{NH}_2\text{C}\text{HN}\text{C}\text{O}\text{HO}\) 0.033232
4. \(\text{NH}_2\text{C}\text{HN}\text{C}\text{O}\text{HO}\) 0.037734
5. \(\text{NH}_2\text{C}\text{HN}\text{C}\text{O}\text{HO}\) 0.039877
6. \(\text{HO}\text{C}\text{NH}_2\text{C}\text{HN}\text{C}\text{O}\text{HO}\) 0.043138
7. \(\text{HO}\text{C}\text{NH}_2\text{C}\text{HN}\text{C}\text{O}\text{HO}\) 0.044492
8. \(\text{NH}_2\text{C}\text{HN}\text{C}\text{O}\text{HO}\) 0.044893
9. \(\text{HO}\text{C}\text{NH}_2\text{C}\text{HN}\text{C}\text{O}\text{HO}\) 0.047251
10. \(\text{HO}\text{C}\text{NH}_2\text{C}\text{HN}\text{C}\text{O}\text{HO}\) 0.047937
| Top predicted substructures | prob     | [CX3H]                | 1.0  | [CX3H][#6][#6] | 0.9996 | #O=[CX3][CXH4] | 0.9175 | [CX4H3][#6] | 0.9987 | [#6H3]=[#6][#6] | 0.9859 |
|-----------------------------|----------|-----------------------|------|----------------|--------|----------------|--------|----------------|--------|----------------|--------|
| best positives              | prob     | [CX4H3][#6][#6]       | 0.9996 | [CX4H2][#6][6][6] | 0.9974 | [#O2H1]       | 0.8583 | [CX3][#6H1]  | 0.9872 | [#8]=[#6][#6] | 0.8435 |
| worst negatives             | prob     | [CX4H3][#6][#6]       | 0.9996 | [CX3H2][#6][#6] | 0.9974 | C=[O][#6][#6]  | 0.9872 | [CX4H2][#6][#6] | 0.9859 | C=CC=CC#C          | 0.0   |
| best positives              | prob     | [CX4H3][#6][#6]       | 0.9996 | [CX3H2][#6][6][6] | 0.9974 | [#O2H1]       | 0.8583 | [CX3][#6H1]  | 0.9872 | [#8]=[#6][#6] | 0.8435 |
| worst negatives             | prob     | [CX4H3][#6][#6]       | 0.9996 | [CX3H2][#6][6][6] | 0.9974 | [#O2H1]       | 0.8583 | [CX3][#6H1]  | 0.9872 | [#8]=[#6][#6] | 0.8435 |
Example 36 true smiles: CN1CCW(C)C1=O formula: C5H10N2O
Index of correct structure: 0 of 14072
True structure loss: 0.02034
True structure:
| Top predicted substructures                          | prob | best positives                          | prob | worst negatives                          | prob |
|-----------------------------------------------------|------|-----------------------------------------|------|-----------------------------------------|------|
| [#7X3][#6H2]                                       | 0.968| [CX4H2][CX4H2]                          | 0.818| [#7X3H1]                                | 0.6854|
| [#7][#6H2]                                         | 0.9601| [#6H2][#7][#6X3]                        | 0.7848| [CX4H2][CX4H2][CX3H0]                   | 0.5299|
| [#7][#6H2][#6H2]                                   | 0.9204| [#6H3][#7][#6X3]                        | 0.7758| [CX4H2][#6X3][#6X3]                     | 0.5005|
| [#7X3][#6H3]                                       | 0.8688| [CX4H2][#NX3H0][CX4H2]                  | 0.7597| [#7][#6][#6][#6X3]                      | 0.4992|
| [#6H3][#7]                                         | 0.8488| [#7X3H0]                                | 0.7487| [#7][#6][#6][#6X3]                      | 0.467|
| best positives                                     |      | best negatives                          | prob |
| [#7X3][#6H2]                                       | 0.968| C=CC=CC#C                               | 0.0  | worst positives                         | prob |
| [#7][#6H2]                                         | 0.9601| [CX2H0][#CX2H1][CX3H0]                  | 0.0  | [#7][#6H2][#6H2]                        | 0.6854|
| [#7][#6H2][#6H2]                                   | 0.9204| [CX2H1][CX2H0][CX4H1][OX2H1]            | 0.0  | [CX4H2][#6X3][#6X3]                     | 0.5299|
| [#6H3][#7]                                         | 0.8688| C=CCC#C                                 | 0.0  | [CX4H2][#6X3][#6X3]                     | 0.5005|
| [#6H3][#7][#6X3]                                   | 0.8488| [CX3H0][#CX3H1]{(#OX2H0)[CX2H0]}        | 0.0  | [CX4H2][#6X3][#6X3]                     | 0.5005|
| [#6H3][#7][#6X3]                                   | 0.7848| CCC=CC#C                                | 0.0  | [CX4H2][#6X3][#6X3]                     | 0.5005|
| [#6H3][#7][#6X3]                                   | 0.7758| [CX3H1][#CX3H2]{#CX2H0}                 | 0.0  | [CX4H2][#6X3][#6X3]                     | 0.5005|
| [CX4H2][#6X3][#6X3][#6X3]                          | 0.7597| [CX3H1][#CX3H2][#CX2H0]                 | 0.0  | [CX4H2][#6X3][#6X3]                     | 0.5005|
| [CX4H2][#6X3][#6X3][#6X3]                          | 0.7487| [CX4H2][#CX4H3][#CX2H0]                 | 0.0  | worst negatives                         | prob |
| worst negatives                                     |      | worst positives                         | prob |
| [#7X3H1]                                           | 0.6854| [#6][#6][#7][#6][#7]                    | 0.2071| [#7][#6][#6][#7]                        | 0.4594|
| [CX4H2][#OX3H0][#CX4H2]                            | 0.5299| [#7][#6H0][#7]                          | 0.4379| [CX4H2][#6X3][#6X3]                     | 0.4784|
| [CX4H2][#OX3H0][#CX4H2]                            | 0.5005| [#7][#6][#7]                            | 0.4784| [CX4H2][#6X3][#6X3]                     | 0.4784|
| [CX4H2][#OX3H0][#CX4H2]                            | 0.4992| [#7][#6][#6][#7]                        | 0.4594| [CX4H2][#6X3][#6X3]                     | 0.4594|
| [CX4H2][#OX3H0][#CX4H2]                            | 0.467| [#6H3][#7][#6H2]                        | 0.6348| [CX4H2][#6X3][#6X3]                     | 0.6473|
| [CX3][#6X][#7][#7]                                 | 0.4186| [CX4H3]                                 | 0.6473| [#7][#6][#6][#6X3][#7]                  | 0.6488|
| [CX3][#6X][#7][#7][#7]                             | 0.4131| [#7][#6H2][#6H2][#7]                    | 0.6488| [CX4H2][#6X3][#6X3]                     | 0.6674|
| [#7][#6][#6][#6][#7]                               | 0.3332| [#6H3][#7X3H0][#6X4H2][#6X4H2]          | 0.741| [NH1][#6][#7]                           | 0.3333|
| [NH1][#6][#7]                                       | 0.3333| [#7X3H0]                                | 0.7487| | |
Example 37 true smiles: O=C\text{Cl}C\text{NC}(=\text{O})\text{N}1\text{ formula: C}_4\text{H}_6\text{N}_2\text{O}_2

Index of correct structure: 0 of 12102

True structure loss: 0.026577

True structure:

Experimental $^{13}$C NMR (solvent: DMSO)

Experimental $^{1}$H NMR (solvent: D$_2$O)

Top predicted structures (loss):

\[
\begin{array}{cccccc}
\text{structures} & \text{loss} \\
\hline
\text{O=C\text{Cl}C\text{NC}(=\text{O})\text{N}1} & 0.026577 \\
\text{O=C\text{Cl}C\text{NC}(=\text{O})\text{N}1} & 0.036055 \\
\text{O=C\text{Cl}C\text{NC}(=\text{O})\text{N}1} & 0.036253 \\
\text{O=C\text{Cl}C\text{NC}(=\text{O})\text{N}1} & 0.036614 \\
\text{O=C\text{Cl}C\text{NC}(=\text{O})\text{N}1} & 0.037828 \\
\text{O=C\text{Cl}C\text{NC}(=\text{O})\text{N}1} & 0.03805 \\
\text{O=C\text{Cl}C\text{NC}(=\text{O})\text{N}1} & 0.039202 \\
\text{O=C\text{Cl}C\text{NC}(=\text{O})\text{N}1} & 0.039352 \\
\text{O=C\text{Cl}C\text{NC}(=\text{O})\text{N}1} & 0.040693 \\
\text{O=C\text{Cl}C\text{NC}(=\text{O})\text{N}1} & 0.043407 \\
\end{array}
\]
| Top predicted substructures | prob | best positives | prob |
|-----------------------------|------|----------------|------|
| [CX4H2](#{6})#{6}           | 0.9954 | [#7]#{6H2}    | 0.8652 |
| [CX4H2](CX4H2)[CX3H0]      | 0.922  | [CX4H2]CX3=O  | 0.8137 |
| [CX3](=OX1)cC              | 0.8961 | [CX4H2]CC=O   | 0.7189 |
| [CX4H2][CX4H2]             | 0.8837 | [#6X3][#7X3][#6X3] | 0.652 |
| O=[CX3H0][CX4H2][CX4H2]    | 0.8827 | [#6]=[#6][#6] | 0.6518 |
| [CX4H2](#{6})#{6}           | 0.9954 | [CX3H0]({=[CX3H2]}{=[CX4H3]}{=[CX4H0]} | 0.0 |
| [CX4H2](CX4H2)[CX3H0]      | 0.922  | [CX4H0]={{CX4H3}}{=[CX4H2]}{=[CX4H1]}{=[CX3H1]} | 0.0 |
| [CX4H2][CX4H2]             | 0.8837 | CC=CC#CC       | 0.0 |
| O=[CX3H0][CX4H2][CX4H2]    | 0.8827 | [CX4H0]{{CX4H3}}{=[CX4H2]}{=[CX4H1]}{=[CX4H1]} | 0.0 |
| [#7][#6H2]                 | 0.8652 | C=C=CCC#C      | 0.0 |
| [CX4H2][CX3]=O             | 0.8137 | CC=CCC#C       | 0.0 |
| [CX4H2][CC=O]              | 0.7189 | [CX2H0][#CX2H0][#CX2H0] | 0.0 |
| [#6X3][#7X3][#6X3]         | 0.652  | [CX2H1][#CX2H0][#CX3H1][#CX3H0] | 0.0 |
| [#6H2][#6H2]               | 0.6169 | [CX3H1][#CX3H2][#CX4H0] | 0.0 |

| worst negatives | prob | worst positives | prob |
|-----------------|------|----------------|------|
| [#6]=[#6][#6]   | 0.6518 | [#7][#6][#6][#6][#7] | 0.1724 |
| [CX3](=OX1)O    | 0.6105 | [NX3H1][#NX3H1][#CX4H2] | 0.1782 |
| [#8][#6][#6H2] | 0.6076 | [CX3H0][=OX1H0][{=NX3H1}]{=[CX4H2]} | 0.2742 |
| [#7][#6][#6X3] | 0.5004 | [#7][#6][#6][#6X3] | 0.394 |
| [OCC(CH2)]     | 0.4911 | [#7X3H1] | 0.4338 |
| [OX1H0]=CX3H0(#{8})[CX4H2] | 0.3048 | [#7][#6][#7] | 0.4994 |
| [OX2H1]        | 0.3046 | [#6X3][#7][#6X3] | 0.5718 |
| [#6X3H0][CX4H2][CX4H2][#6X3H0] | 0.2809 | [#7][#6H0][#7] | 0.5769 |
| [#6H3][#7]     | 0.2805 | [#7X3][#6H2] | 0.5941 |
Example 38 true smiles: CN1CC(=O)NC1=O formula: C4H6N2O2
Index of correct structure: 0 of 12102
True structure loss: 0.027381
True structure:

Experimental 13C NMR (solvent: DMSO-d6)

![Experimental 13C NMR](image)

Experimental 1H NMR (solvent: D2O)

![Experimental 1H NMR](image)

Top predicted structures (loss):

1. ![Structure 1](image) - 0.027381
2. ![Structure 2](image) - 0.031975
3. ![Structure 3](image) - 0.035806
4. ![Structure 4](image) - 0.036366
5. ![Structure 5](image) - 0.037764
6. ![Structure 6](image) - 0.038849
7. ![Structure 7](image) - 0.039835
8. ![Structure 8](image) - 0.041758
9. ![Structure 9](image) - 0.04187
10. ![Structure 10](image) - 0.041875
| Top predicted substructures | prob | best positives | prob | best negatives | prob | worst positives | prob |
|-----------------------------|------|----------------|------|----------------|------|----------------|------|
| [#7X3][#6H3]               | 0.985| [CX3]=[OX1]C   | 0.781| [CX3][=OX1][C  | 0.781| [#8][#6][#8]   | 0.7127|
| [CX4H3]                    | 0.9838| [CX3][#6][#8] | 0.781| [CX4H2][NX3H0][CX3H0] | 0.6611| [CX4H3][NX3H0] | 0.6615|
| [#6H3][#7]                 | 0.9497| [CX4H2][CX3H0] | 0.6152| [CX4H2]#OX1#C  | 0.6151| [CX4H2][NX3H0] | 0.651|
| [#6H3][#7][#6X3]           | 0.8954| [#7][#6H0][#7] | 0.6152| [OX2H0][CX4H2][OX2H0] | 0.6151| [OX2H0][CX4H2][OX2H0] | 0.6151|
| [#7][#6H0][#7]             | 0.8333| [CX4H2][CX4H3][CX2H0] | 0.6152| [CX4H1][CX4H1][CX4H1] | 0.6152| [CX4H1][CX4H1][CX4H1] | 0.6152|
| [#7][#6][#7]               | 0.781| [OX2H0][CX4H2][CX4H1] | 0.6152| [OX2H0][CX4H2][CX4H1] | 0.6152| [OX2H0][CX4H2][CX4H1] | 0.6152|
| [CX4H3]=[OX1][C            | 0.6611| [OX2H0][CX4H2][CX4H1] | 0.6152| [OX2H0][CX4H2][CX4H1] | 0.6152| [OX2H0][CX4H2][CX4H1] | 0.6152|
| [CX4H2][NX3H0][CX3H0]      | 0.6611| [OX2H0][CX4H2][CX4H1] | 0.6152| [OX2H0][CX4H2][CX4H1] | 0.6152| [OX2H0][CX4H2][CX4H1] | 0.6152|
| [OX2H0][CX4H2][CX4H1]      | 0.6611| [OX2H0][CX4H2][CX4H1] | 0.6152| [OX2H0][CX4H2][CX4H1] | 0.6152| [OX2H0][CX4H2][CX4H1] | 0.6152|
| [OX2H0][CX4H2][CX4H1]      | 0.6611| [OX2H0][CX4H2][CX4H1] | 0.6152| [OX2H0][CX4H2][CX4H1] | 0.6152| [OX2H0][CX4H2][CX4H1] | 0.6152|
| [OX2H0][CX4H2][CX4H1]      | 0.6611| [OX2H0][CX4H2][CX4H1] | 0.6152| [OX2H0][CX4H2][CX4H1] | 0.6152| [OX2H0][CX4H2][CX4H1] | 0.6152|
| [OX2H0][CX4H2][CX4H1]      | 0.6611| [OX2H0][CX4H2][CX4H1] | 0.6152| [OX2H0][CX4H2][CX4H1] | 0.6152| [OX2H0][CX4H2][CX4H1] | 0.6152|

---
Example 39 true smiles: O=CC=Cc1ccccc1 formula: C9H8O
Index of correct structure: 1 of 10441
True structure loss: 0.026198
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: MeOD)

Top predicted structures (loss):
| Top predicted substructures       | prob   | best positives       | prob   | best negatives       | prob   |
|-----------------------------------|--------|----------------------|--------|----------------------|--------|
| [#6H1]                            | 0.9998 | [#8]                | 0.1025 | O=[#6]               | 0.125  |
| [#6X3][#6X3]                      | 0.9996 | [CHX3]              | 0.1177 | [#6H0]=[#6H1]        | 0.1122 |
| [cX3H1][cX3H1][cX3H0]             | 0.9882 | [OX2H1][#6X3]=[#6X3]| 0.1189 | [#6H0][#6H1]        | 0.1177 |
| [#6H1][#6X3H0]                   | 0.9793 | [CX4H2][OX2H1][#6X3]| 0.1225 | [OX1H0][#6X3]=[#6X3]| 0.1225 |
|                                   |        | [#6X3][#6X3][#6X3]  | 0.1225 | [CX3H1]=([#6X3][#6X3]| 0.1159 | [#8][#6X3][#6X3]=[#6X3]| 0.1225 |
|                                   |        | [#6X3][#6X3]        | 0.1159 | [CX3H1]=([#6X3][#6X3]| 0.1181 | [#8][#6X3][#6X3]==[#6X3]| 0.1159 |
|                                   |        | [cX3H1][cX3H1][cX3H0]| 0.0927 | [CX3H1]==[#6X3][#6X3]| 0.1207 | [OX1H0][#6X3][#6X3]==[#6X3]| 0.1207 |
|                                   |        | [cX3H1][cX3H1][cX3H0]| 0.0819 | [OX1H0][#6X3][#6X3]==[#6X3]| 0.1225 | [OX1H0][#6X3][#6X3]==[#6X3]| 0.1225 |
|                                   |        | [cX3H1][cX3H1][cX3H0]| 0.0819 | [OX1H0][#6X3][#6X3]==[#6X3]| 0.1225 | [OX1H0][#6X3][#6X3]==[#6X3]| 0.1225 |
|                                   |        | [cX3H1][cX3H1][cX3H0]| 0.0819 | [OX1H0][#6X3][#6X3]==[#6X3]| 0.1225 | [OX1H0][#6X3][#6X3]==[#6X3]| 0.1225 |
|                                   |        | [cX3H1][cX3H1][cX3H0]| 0.0819 | [OX1H0][#6X3][#6X3]==[#6X3]| 0.1225 | [OX1H0][#6X3][#6X3]==[#6X3]| 0.1225 |

worst negatives prob worst positives prob

O=[#6]=[#6][#6X3] 0.4865 [#8][#6][#6H1]==[#6H1] 0.1025
[#8]=[#6H][#6X3][#6X3] 0.2813 [CHX3]==[cC] 0.1177
[#8]=[#6X0][#6H1] 0.1435 O=[#6][#6]==[#6X3] 0.1122
[cX3H0][cX3H1][cX3H1][cX3H0] 0.1225 [CX3H1][=[CX3H1][cX3H1] 0.1159
[#6X2][#6X3][#6X3][#6X3] 0.1225 [CX3H1]==[cX3H1] 0.1526
[#8][#6X0][#6H1] 0.0999 [#6X3][#6X3][#6X3][#6X3] 0.1569
[cX3H0][cX3H1][cX3H1][cX3H0] 0.0927 [CX3H1]==[#6X3][#6X3]| 0.1891
[OX1H0][cX3H0][cX3H1][cX3H1] 0.0819 [OX1H0][#6X3][#6X3]| 0.2076
[#8][#6]=[#6][#6X3] 0.08 [OX1H0][#6X3][#6X3] 0.2111
[cX3H0][cX3H1][cX3H1][cX3H0] 0.0574 [CX3H1]==[OX1H0][cX3H1] 0.2295
------------------------------------------------------------------------------------------------------------------------
Example 40 true smiles: O=C(O)c1ccccn1 formula: C6H5NO2
Index of correct structure: 0 of 10337
True structure loss: 0.012249
True structure:

Experimental 13C NMR (solvent: DMSO)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

0.012249 0.021685 0.022058 0.023854 0.023959
0.025189 0.02572 0.026606 0.027178 0.027365
| Top predicted substructures | prob | best positives | prob | worst negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [6H1]                       | 0.9999 | [6H1] | 0.9854 |
| [6X3]                       | 0.9999 | [6X3] | 0.9823 |
| [6X3] [6X3] [6X3] [6X3]    | 0.9979 | [6X3] [6X3] | 0.9737 |
| [cH] [cH]                   | 0.9959 | [cH] [cH] | 0.936 |
| [6X3H1] [6X3H0]            | 0.9854 | [6X3H1] [6X3H0] | 0.6739 |
| [cX3H1] [cX3H1] [cX3H1] [cX3H1] | 0.9823 | [cX3H1] [cX3H1] [cX3H1] | 0.6009 |
| [cX3H1] [cX3H1]            | 0.9737 | [cX3H1] [cX3H1] | 0.4033 |
| [OX2H1] [OX2H2] [OX2H1]   | 0.9459 | [OX2H1] [OX2H2] | 0.2602 |
| [OX2H1] [OX2H1]            | 0.936 | [OX2H1] [OX2H1] | 0.2469 |
| [cX3H1] [cX3H1] [cX3H1]    | 0.9214 | [cX3H1] [cX3H1] [cX3H1] | 0.2194 |
| [cX3H1] [cX3H1] [cX3H1]    | 0.816 | [cX3H1] [cX3H1] [cX3H1] | 0.1811 |
Example true smiles: O=C(O)c1cccnc1 formula: C6H5NO2
Index of correct structure: -1 of 10337
True structure loss: 0.013704
True structure:

Experimental 13C NMR (solvent: DMSO-d6)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

- 0.019113
- 0.022086
- 0.026728
- 0.026819
- 0.028933
- 0.02929
- 0.029411
- 0.030296
- 0.030497
- 0.031821
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [#6H1]                      | 1.0  | [#6X3H1][#6X3H0] | 0.9764 | [#6X3][#6X3][#6X3] | 0.9601 |
| [#6X3][#6X3]                | 0.997 | [#6X3][#6X3][#6X3] | 0.9601 | [#6X3][#6X3][#6X3] | 0.9601 |
| [#7][#6][#6][#6X3]         | 0.9848| [#6H1][#7][#6H1] | 0.8798 | [#6X3][#6X3] | 0.9997 |
| [#8][#6][#6][#6]           | 0.9844| [#6H1][#7][#6H1] | 0.8798 | [#6X3][#6X3][#6X3] | 0.9997 |
| [cH]                        | 0.9829| [cX3H1][#mX2H0][cX3H0] | 0.8575 | [#6X3][#6X3][#6X3] | 0.9997 |
| [cX3H1][#mX2H0]             | 0.9764| [#6X3][#6X3][#6X3] | 0.9997 | [#6X3][#6X3][#6X3] | 0.9997 |
| [cX3H1][#mX2H0]             | 0.9601| [cX3H1][#mX2H0][cX3H0] | 0.8575 | [#6X3][#6X3][#6X3] | 0.9997 |
| [cX3H1][#mX2H0][cX3H0]     | 0.9597| [cX3H1][#mX2H0][cX3H0] | 0.8575 | [#6X3][#6X3][#6X3] | 0.9997 |
| [cX3H1][#mX2H0][cX3H0]     | 0.8798| [cX3H1][#mX2H0][cX3H0] | 0.8575 | [#6X3][#6X3][#6X3] | 0.9997 |

| worst negatives | prob | worst positives | prob |
|-----------------|------|----------------|------|
| [cH][cO]        | 0.4969| [cX3][=O][OX2H1] | 0.0398 |
| [#8][#6H0][#6H1] | 0.429 | [OX2H1] | 0.4263 |
| [#8][#6H0][#6H1] | 0.4127| [#6X3H1][#6X3H1][#6X3H0] | 0.6266 |
| [#8][#6X3H0][#6X3H1] | 0.357 | [#6X3H1][#6X3H1][#6X3H0] | 0.6266 |
| [O][AxH][cX3][c] | 0.3183| [cX3H1][#mX2H0][cX3H1] | 0.6865 |
| [O][AxH][cX3][c] | 0.2553| [cX3H1][#mX2H0][cX3H1] | 0.6865 |
| [O][AxH][cX3][c] | 0.168 | [cX3][=O][OX1] | 0.7289 |
| [cX3H0][{cX3H1}][{cX3H0}][OX2H1] | 0.1244| [cX3][=O][OX1] | 0.7289 |
| [cX3H0][{cX3H1}][{cX3H0}][OX2H1] | 0.1183| [cX3H1][{cX3H1}][cX3H1] | 0.7856 |
Example 42 true smiles: CCCCCCC(=O)O formula: C8H16O2
Index of correct structure: 0 of 9984
True structure loss: 0.006326
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

0.006326  0.024303  0.027016  0.028162  0.030015
0.036825  0.044071  0.044708  0.045575  0.049083
### Top Predicted Substructures

| Substructure | prob  |
|--------------|-------|
| [CX4H2][#6] | 0.9999 |
| [#6H3][#6][#6] | 0.9999 |
| [CX4H3][#6] | 0.9996 |
| [CX4H3][CX4H2] | 0.999 |
| [CX3][#6H3] | 0.9999 |
| [CX4H3][#6] | 0.9996 |
| [CX4H3][CX4H2] | 0.999 |
| [CX3][#6H3][#6] | 0.9999 |
| [CX4H3][#6][#6] | 0.9999 |

### Best Positives

| Substructure | prob  |
|--------------|-------|
| [CX4H2][#6] | 0.9999 |
| [#6H3][#6][#6] | 0.9999 |
| [CX4H3][#6] | 0.9996 |
| [CX4H3][CX4H2] | 0.999 |
| [CX3][#6H3] | 0.9999 |

### Best Negatives

| Substructure | prob   |
|--------------|--------|
| [CX4H2][#6] | 0.9999 |
| [#6H3][#6][#6] | 0.9999 |
| [CX4H3][#6] | 0.9996 |
| [CX4H3][CX4H2] | 0.999 |
| [CX3][#6H3] | 0.9999 |

### Worst Negatives

| Substructure | prob   |
|--------------|--------|
| [#6H1][#6H2] | 0.4448 |
| [#6H1] | 0.3232 |
| [CX3H0][#6H3][#6H2] | 0.2239 |
| [CX4H3][CX4H1] | 0.1484 |
| [CX4H2][#6H1] | 0.1437 |
| [#6H3] | 0.1234 |
| [CX4H3][#6H2] | 0.1131 |
| [CX4H1][#6H1][#6H2] | 0.1007 |
| [#6H3][#6][#6H3] | 0.096 |

---

### Worst Positives

| Substructure | prob   |
|--------------|--------|
| [OX1H0][#8][#8][#8][CX4H2] | 0.6574 |
| [#6H1] | 0.6788 |
| O[X3H0][CX4H2][CX4H2] | 0.7064 |
| [CX4H2][CX4H1] | 0.7251 |
| [OX1H0][#8][#8][#8][CX4H2] | 0.7729 |
| [OX1H0][#8][#8][#8][CX4H2] | 0.8335 |
| [OX1H0][#8][#8][#8][CX4H2] | 0.838 |
| [OX1H0][#8][#8][#8][CX4H2] | 0.8454 |
| [OX1H0][#8][#8][#8][CX4H2] | 0.8628 |
| [OX1H0][#8][#8][#8][CX4H2] | 0.8756 |
Example 43 true smiles: NCCc1c[nH]c1 formula: C5H9N3
Index of correct structure: 1 of 8824
True structure loss: 0.022963
True structure:

\[
\begin{align*}
&\text{Experimental }^1\text{C NMR (solvent: D}_2\text{O)} \\
&\text{Experimental }^1\text{H NMR (solvent: D}_2\text{O)} \\
&\text{Top predicted structures (loss):}
\end{align*}
\]

0.021891, 0.022963, 0.024396, 0.032793, 0.033342,
0.035301, 0.052169, 0.052178, 0.052482, 0.05325
Top predicted substructures  prob  
[\#6H1] 0.9983  [#6X3][#6X3] 0.9092  
[CX4H2][{#6}][#6] 0.9917  [#7X3][#6H2] 0.878  
[#7X3H2] 0.9176  [#7][#6][#6X3] 0.829  
[#7][#6H2][#6H2] 0.9176  [#7][#6H2] 0.8177  

best positives  prob  best negatives  prob  
[\#6H1] 0.9983  [OX2H0][CX4H2][CX4H1][CX4H1]1 0.0  
[CX4H2][{#6}][#6] 0.9917  [OX2H0][CX4H2][CX4H1][CX4H1]1 0.0  
[cH] 0.9369  [OX2H0][CX4H2][CX4H1][OX2H1][CX4H3] 0.0  
[#7X3H2] 0.9196  [OX2H0][#CX2H0][CX2H0] 0.0  
[#7][#6H2][#6H2] 0.9176  [OX2H0][CX4H2][CX4H2][OX2H1][CX4H3] 0.0  
[#6X3][#6X3] 0.9092  [OX2H1][CX4H1][CX4H2][CX4H1]1 0.0  
[#7X3][#6H2] 0.878  [OX2H1][CX4H1][OX2H0][CX4H2][CX4H1] 0.0  
[#7][#6][#6][#6X3] 0.8601  [OX2H1][{OX2H0}][{CX4H3}][CX4H0] 0.0  
[#7][#6][#6X3] 0.829  [OX2H1][{OX2H0}][{CX4H1}][CX4H0] 0.0  
[#7][#6H2] 0.8177  [OX2H1][{OX2H0}][{CX4H1}][#6H1][#6H1]#7 0.0  

worst negatives  prob  worst positives  prob  
[cX3H1][{nX2H0}][cX3H0] 0.5003  [#7][#6][#6][#6][#6][#7] 0.2249  
[#7][#7] 0.4353  [cX3H1][{nX3H1}][cX3H0] 0.2641  
[#7X3H0] 0.3889  [#7][#6X3H0][#6X3H1] 0.2764  
[#6X3H1][#7X3H0] 0.2955  [#7][#6H0][#6H1] 0.3113  
[cH][cH] 0.2886  [#6X3][#7][#6X3] 0.3797  
[CX4H2][CX3H] 0.2386  [#6X3][#7][#6X3] 0.4547  
[#6H2][#7][#6X3] 0.2327  [#7][#6X3H1] 0.4658  
[CX4H3] 0.2311  [#6][#6][#7][#6][#7] 0.4963  
[CX4H2][{NX3H1}][CX4H2] 0.2053  [#6H1][#7][#6H1] 0.5013  
[#7][#7H1] 0.1818  [#7][#6][#6][#7] 0.5562  

------------------------------------------------------------------------------------------------------------------------
Example 44 true smiles: CC(O)(CC(O)C(=O)O formula: C₆H₁₂O₄
Index of correct structure: 6 of 8605
True structure loss: 0.047542
True structure:

Experimental ¹³C NMR (solvent: CDCl₃)

Experimental ¹H NMR (solvent: D₂O)

Top predicted structures (loss):

1. 0.038973
2. 0.044147
3. 0.044396
4. 0.046009
5. 0.046128
6. 0.046747
7. 0.047542
8. 0.048719
9. 0.049301
10. 0.049788
Example 45 true smiles: O=C(O)CCC(O)C(=O)O formula: C5H8O5
Index of correct structure: 1 of 8115
True structure loss: 0.028053
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

1. ![Structure 1](image1) 0.024143
2. ![Structure 2](image2) 0.028053
3. ![Structure 3](image3) 0.032335
4. ![Structure 4](image4) 0.034373
5. ![Structure 5](image5) 0.037587
6. ![Structure 6](image6) 0.037689
7. ![Structure 7](image7) 0.039094
8. ![Structure 8](image8) 0.040344
9. ![Structure 9](image9) 0.043423
10. ![Structure 10](image10) 0.046571
| Top predicted substructures | prob | best positives | prob | worst negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [CX4H2][[6]][6]           | 0.9993 | [CX3]=[OX1][0] | 0.9786 | [6H1][6H1] | 0.4845 |
| [CX3]=[OX1]C              | 0.9979 | [6H3][6X3][6X3]=[6X3H2] | 0.0 | [6][6][6H2][6][6] | 0.2572 |
| [OX2H1]                   | 0.9968 | [OX1H0]=[OX1H0][{8}][{8}] | 0.0271 | [6H1][6H1][6H1] | 0.2031 |
| OCC[CH2]                  | 0.9962 | [6H3]=[6H3][6H3][6H3] | 0.0 | [8][6][6][6][6][6][6] | 0.3892 |
| [8]=={6}=[6]             | 0.9805 | [CX3]=[OX1][0] | 0.9786 | [8][6]=={6}=[6] | 0.1938 |
| [8][6][6]=={6}           | 0.9885 | [CX3]=[OX1][0] | 0.9786 | [6H1][6H1][6H1] | 0.2176 |
| best positives            | prob | best negatives | prob | worst positives | prob |
| [CX4H2][[6]][6]           | 0.9993 | [CX2H1][CX2H0][CX3H1]=[CX3H0] | 0.0 | [6H1][6H1] | 0.4845 |
| [CX3]=[OX1]C              | 0.9979 | [6H3][6X3][6X3]=[6X3H2] | 0.0 | [6][6][6H2][6][6] | 0.2572 |
| [OX2H1]                   | 0.9968 | [OX1H0]=[OX1H0][{8}][{8}] | 0.0271 | [6H1][6H1][6H1] | 0.2031 |
| OCC[CH2]                  | 0.9962 | [6H3]=[6H3][6H3][6H3] | 0.0 | [8][6][6][6][6][6][6] | 0.3892 |
| []8=[6]==[6]             | 0.9805 | [CX3]=[OX1][0] | 0.9786 | [8][6]=={6}=[6] | 0.1938 |
| []8][6][6]==[6]          | 0.9885 | [CX3]=[OX1][0] | 0.9786 | [6H1][6H1][6H1] | 0.2176 |
Example 46 true smiles: \( O=C(\text{O})\text{CCC}(=\text{O})\text{C}(=\text{O})\text{O} \) formula: C5H6O5

Index of correct structure: 0 of 7597
True structure loss: 0.008162
True structure:

Experimental 13C NMR (solvent: DMSO)

Experimental 1H NMR (solvent: d2o)

Top predicted structures (loss):

```
0.008162  0.042883  0.049861  0.049914  0.06086

0.062934  0.064672  0.066055  0.068078  0.069021
```
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [CX3]=([OX1])C             | 1.0  | O=[CX3H0][CX4H2][CX4H2] | 0.9903 |                  |      |
| [CX4H2][#6][#6]           | 0.9999 | [CX4H2][CX3]=O | 0.9811 |                  |      |
| [OX1]O                     | 0.9997 | OCC[CH2] | 0.9543 |                  |      |
| [#8]=[#6][#8]             | 0.9986 | [CX4H2][CX4H2][CX3H0] | 0.9938 | [#8]=[#6][#6]=[#8] | 0.9427 |

| worst negatives | prob | worst positives | prob |
|-----------------|------|-----------------|------|
| [CX4H3][CX3H0]  | 0.3481 | [#8][#6][#6][#6][#6]=[#8] | 0.514   |
| [OX2H0][CX3H0][CX4H2] | 0.3122 | [#8][#6][#6][#6]=[#8] | 0.5427   |
| [OX1H0][OX2H0][CX4H2] | 0.2447 | [#8]=[#6][#6][#6][#6][#6][#8] | 0.7041   |
| [OX1H0][OX1H0][OX2H0][CX4H2] | 0.1992 | [#8][#6][#6][#6][#6][#8] | 0.7116   |
| [OX1H0]=([OX1H0][OX2H0])| 0.1964 | [#8][#6][#6][#6][#6][#6][#8] | 0.7351   |
| [OX1H0][OX1H0][OX2H0][OX2H0][CX4H2] | 0.1642 | [#8][#6][#6][#6][#6][#6][#6][#8] | 0.805   |
| [OX1H0][OX1H0][OX2H0][OX2H0][OX2H0][CX4H2] | 0.1515 | [#8][#6][#6][#6][#6][#6][#6][#6][#8] | 0.8141   |
| [OX1H0][OX1H0][OX2H0][OX2H0][OX2H0][OX2H0][CX4H2] | 0.1462 | [#8][#6][#6][#6][#6][#6][#6][#6][#6][#8] | 0.8203   |
| [OX1H0][OX1H0][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0][CX4H2] | 0.1116 | [#8][#6][#6][#6][#6][#6][#6][#6][#6][#6][#8] | 0.8269   |
| [OX1H0][OX1H0][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0][CX4H2] | 0.1064 | [#8][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#8] | 0.8935   |

---
Example 47 true smiles: O=C(O)C1CCCN1 formula: C5H9NO2
Index of correct structure: 0 of 6935
True structure loss: 0.024354
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: d2o)

Top predicted structures (loss):

0.024354 0.040611 0.045086 0.049682 0.051446
0.052438 0.053394 0.054517 0.05542 0.055889
| Substructure                                                                 | prob | Substructure                                                                 | prob |
|-----------------------------------------------------------------------------|------|-----------------------------------------------------------------------------|------|
| `[CX4H2]([#6])[#6]`                                                          | 0.9998 | `[#7X3][#6H2]`                                                              | 0.89  |
| `[CX4H2][CX4H2]`                                                            | 0.9861 | `[#6H1][#6H2]`                                                              | 0.8996 |
| `[CX3]=([OX1])C`                                                             | 0.9834 | `[#6H1][#6H2]`                                                              | 0.8844 |
| `[CX4H2][CX4H2][CX4H1]`                                                      | 0.972  | `O=[CX3][CX4H1]`                                                           | 0.9593 |
| `OCC[CH2]`                                                                   | 0.9593 | `O=[CX3][CX4H1]`                                                           | 0.9264 |

**Best Positives**

| Substructure                                                                 | prob | Substructure                                                                 | prob |
|-----------------------------------------------------------------------------|------|-----------------------------------------------------------------------------|------|
| `[CX4H2]([#6])[#6]`                                                          | 0.9998 | `[#7X3][#6H2]`                                                              | 0.89  |
| `[CX4H2][CX4H2]`                                                            | 0.9861 | `[#6H1][#6H2]`                                                              | 0.8996 |
| `[CX3]=([OX1])C`                                                             | 0.9834 | `[#6H1][#6H2]`                                                              | 0.8844 |
| `[CX4H2][CX4H2][CX4H1]`                                                      | 0.972  | `OCC[CH2]`                                                                  | 0.9264 |
| `OCC[CH2]`                                                                   | 0.9593 | `OCC[CH2]`                                                                  | 0.9593 |

**Worst Negatives**

| Substructure                                                                 | prob | Substructure                                                                 | prob |
|-----------------------------------------------------------------------------|------|-----------------------------------------------------------------------------|------|
| `[CX4H2][#6][O]`                                                             | 0.5716 | `[#7][#6H1][#6H2]`                                                          | 0.2458 |
| `[CH2X4]([0])[CX4H2]`                                                        | 0.4952 | `[#6H1][#6H2]`                                                              | 0.3091 |
| `[#8][#6][#6H2]`                                                             | 0.4921 | `[#6H1][#6][#6][#6H1][#7]`                                                 | 0.3217 |
| `[#7X3H2]`                                                                   | 0.4455 | `[#7][#6H1][#6H2]`                                                          | 0.393  |
| `O=(CX3H0)[CX4H2][CX4H2]`                                                    | 0.3793 | `[#7][#6H1][#6H2]`                                                          | 0.4148 |
| `[CX4H2][CX3]=O`                                                             | 0.3654 | `[#7][#6H1][#6H2]`                                                          | 0.4438 |
| `[CX4H2][OX2H0][CX4H2]`                                                      | 0.3401 | `[#7][#6H1][#6H2]`                                                          | 0.4651 |
| `[#7][#6H2][#6H1]`                                                           | 0.2564 | `[#7][#6H1][#6H2][#7]`                                                      | 0.5972 |
| `[#7][#6X4H1][#6X3]`                                                         | 0.2398 | `[#6][#6H1][#6H2][#6][#7]`                                                 | 0.6125 |
Example 48 true smiles: CCOC(=O)CC#N formula: C5H7NO2
Index of correct structure: 0 of 6308
True structure loss: 0.028354
True structure:

- Experimental 13C NMR (solvent: CDCl3)
- Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

1. 0.028354
2. 0.042572
3. 0.046532
4. 0.048316
5. 0.050451
6. 0.053284
7. 0.055955
8. 0.055974
9. 0.056476
10. 0.058853
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [CX4H3]                     | 0.9878 | [CX4H2][{CX3H0}][CX2H0] | 0.7995 | [CX3H1](=[CX3H2])[CX3H1] | 0.0 |
| [CX4H3][#6]                | 0.8479 | [#8][#6][#6H2] | 0.7796 | [CX3H1](=[CX3H2])[CX3H2] | 0.0 |
| [#6][#7]                   | 0.8474 | [#8][#6][#6H2][#6X2] | 0.7432 | [CX3H1](=[CX3H2])[CX3H0] | 0.0 |
| [CX2H0][{CX4H2}][#6X3H0]  | 0.8455 | [#6H1] | 0.6899 | [OX1H0](=[OX1H1])[OX1H0] | 0.0 |
| [CX4H3][{CX4H2}]           | 0.8052 | [CX4H2][{#6}][#6] | 0.6827 | [OX1H0][{OX1H1}][OX1H0][OX1H2] | 0.0 |
| best positives             | prob | best negatives | prob |
| [CX4H3]                     | 0.9878 | [CX3H1](=[CX3H2])[CX3H1] | 0.0 |
| [CX4H3][#6]                | 0.8479 | [CX3H1](=[CX3H2])[CX4H2] | 0.0 |
| [#6][#7]                   | 0.8474 | [CX4H3][CX3H0][CX4H2][CX3H1] | 0.0 |
| [CX4H2][{CX3H0}][CX2H0]    | 0.7995 | C=CC=CC=CC | 0.0 |
| [#8][#6][#6H2]             | 0.7796 | [CX4H0][{CX4H2}][{CX4H2}][{CX4H1}][CX4H1] | 0.0 |
| [OX1H0][{OX1H1}][OX1H2]    | 0.6696 | [CX3H1](=[CX3H2])[CX4H2] | 0.0 |
| worst negatives             | prob | worst positives | prob |
| [#6H1]                     | 0.6899 | [CX4H2][{CX3}]=O | 0.0745 |
| [#8][#6][#6][#6X3]        | 0.4729 | [OX1H0][{OX1H0}][{#8}][CX4H2] | 0.1265 |
| [#6E1][#6H1]              | 0.373 | [CX3H0][{OX1H0}][{OX2H0}][{CX4H2}] | 0.1695 |
| [OX1H0]                     | 0.2989 | [CX4H2][{#6}][0] | 0.3515 |
| [#8][#6][#8]              | 0.2475 | [OX2H0][{CX3H0}][{CX4H2}] | 0.3763 |
| [OX1H0][{OX2H0}][{CX4H2}] | 0.2435 | [CX3][{OX1}]=O | 0.3783 |
| O=[{CX3}][{CX4H}]         | 0.2361 | [#8][{#6X3}][#6H2][#6H0] | 0.3915 |
| C1CC1                      | 0.2336 | [OX1H0][{#6}][#6H2][#6X2] | 0.4134 |
| [#8][#6H0][#6H1]          | 0.214 | [#6H2][#6X2] | 0.4931 |
| OCC[CH2]                   | 0.2077 | [#8]=#8 | 0.4994 |
Example 49 true smiles: CC\(\text{O}\)CC\(\text{(O}\)C\(\text{(O}\)O formula: C\(6\text{H}_{10}\text{O}_3\)
Index of correct structure: 0 of 6069
True structure loss: 0.016838
True structure:

![Chemical structure](image)

**Experimental 13C NMR** (solvent: DMSO)

**Experimental 1H NMR** (solvent: D\(_2\)O)

Top predicted structures (loss):

1. ![Structure 1](image) 0.016838
2. ![Structure 2](image) 0.054685
3. ![Structure 3](image) 0.055422
4. ![Structure 4](image) 0.064634
5. ![Structure 5](image) 0.073687
6. ![Structure 6](image) 0.077413
7. ![Structure 7](image) 0.078502
8. ![Structure 8](image) 0.082401
9. ![Structure 9](image) 0.082986
10. ![Structure 10](image) 0.083879
| Substructure                                      | prob  | Substructure                                      | prob  |
|--------------------------------------------------|-------|--------------------------------------------------|-------|
| [{CX3}](=[{OX1}])C                              | 0.9999 | O=[{CX3H0}][{CX4H2}][{CX4H1}]                   | 0.9608 |
| [{CX4H3}]                                        | 0.9999 | [{CX4H2}]{{#6}}{{#6}}                            | 0.9539 |
| [#6H3]{{#6}}{{#6}}                               | 0.9953 | [{CX3}]=[{OX1}]O                               | 0.9525 |
| [{CX4H3}]{{#6}}                                 | 0.9894 | [{CX4H2}]{[CX3]}=O                              | 0.95   |
| [{OX1H0}]=[{CX3H0}]{{#6}}{{#6}}{{CX4H2}}         | 0.9584 | [{#6H1}]                                     | 0.9455 |

**best positives**

| Substructure                                      | prob  | Substructure                                      | prob  |
|--------------------------------------------------|-------|--------------------------------------------------|-------|
| [{CX3}](=[{OX1}])C                              | 0.9999 | O=CC=CC=C#C                                     | 0.0   |
| [{CX4H3}]                                        | 0.9999 | [{CX2H0}]{{#6}}{[CX2H0]}                       | 0.0   |
| [#6H3]{{#6}}{{#6}}                               | 0.9953 | CCC=C#C#C                                      | 0.0   |
| [{CX3H3}]{{#6}}                                 | 0.9894 | [{CX2H0}]{{#6}}{[CX4H0]}                       | 0.0   |
| [{OX1H0}]=[{CX3H0}]{{#6}}{{#6}}{{CX4H2}}         | 0.9584 | CC=CC=C#C                                      | 0.0   |
| O=[{CX3H0}][{CX4H2}][{CX4H1}]                   | 0.9584 | [{CX4H3}][{CX2H0}]                             | 0.0   |
| [{CX4H2}]{{#6}}{{#6}}                            | 0.9539 | [{CX3}]=[{OX1}]O                              | 0.0   |
| [{CX4H2}]{{#6}}                                 | 0.95   | [{CX2H0}][{CX2H1}]{{#6}}                       | 0.0   |
| [#6H1]                                           | 0.9455 | [{CX2H0}][{CX2H1}]{{#6}}                       | 0.0   |

**worst negatives**

| Substructure                                      | prob  | Substructure                                      | prob  |
|--------------------------------------------------|-------|--------------------------------------------------|-------|
| [#8]{{#6}}{{#6H2}}                               | 0.7481 | OCC[CH2]                                       | 0.1562 |
| O[{{CX4H}].[CX4H2}}                              | 0.4988 | [#8]{{#6}}{{#6]]=[#8]                           | 0.3936 |
| [#8]{{#6}}{{#6}}[6X3}                              | 0.3854 | [{CX4H2}]CC=O                                  | 0.397  |
| [#6H2]{{#6}}{{#6H1}}                              | 0.3465 | [{CX3H}][{OX1H0}]{{#6}}                       | 0.5962 |
| [{OX1H0}]=[{CX3H0}]{{#8}}{{#8}}{{CX4H2}}         | 0.3268 | [#6X3]{{#6X3}}                                  | 0.6026 |
| [{CX4H2}][{OX4}][{CHX4}]                         | 0.3212 | [#8]=[#6]{{#6]]=[#8]                           | 0.7001 |
| [{CX3H0}]=[{OX1H0}]{{#6X4}}{[CX4H2}}             | 0.2723 | [{CH4}][{CH3X4}][{CH2X4}]                      | 0.7511 |
| [#8]{{#6}}{{#6}{{#6}}{[#6]}{[#6]}{[#6]}{[#6]}   | 0.2611 | [#8]=[#6]{{#6H2}}{{#6H1}                       | 0.7651 |
| [{CX4H2}][{#6}]}[O]                              | 0.2494 | [{CX3}]=[{O}][{OX2H1}]                       | 0.7694 |
| [{CX4H2}]{{#6}}]O                                | 0.2027 | [{CX4H3}][{CX4H1}]                             | 0.7729 |
Example 50 true smiles: COc1ccccc1O formula: C7H8O2
Index of correct structure: 1 of 5977
True structure loss: 0.008976

True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

- 0.006759
- 0.008976
- 0.011707
- 0.011993
- 0.013973
- 0.032486
- 0.032625
- 0.035157
- 0.040894
- 0.049135
Top predicted substructures                  prob
[#6X3][#6X3][#6X3][#6X3]                  0.9979  [#6X3H1][#6X3H0]                  0.9792
[#6H1]                                    0.9976  [cH][cH]                        0.9617
[#6X3][#6X3]                              0.9973  [cX3H1][cX3H1][cX3H0]           0.9536
[cH]                                      0.9864  [#6H1][#6H1]                   0.9455
[#8][#6][#6][#6X3]                       0.9828  [#8][#6H0][#6H1]                0.8961

best positives                            prob           best negatives                  prob
[#6X3][#6X3][#6X3][#6X3]                  0.9979  [CX4H1][NX3H1][CX4H3][CX4H2][CX4H1] 0.0  
[#6H1]                                    0.9976  [CX4H1][NX3H2][CX4H2][CX4H1]     0.0  
[#6X3][#6X3]                              0.9973  [CX4H1][NX3H2][CX4H1]            0.0  
[cH]                                      0.9864  [CX4H1][NX3H0][CX4H3]           0.0  
[#8][#6][#6][#6X3]                       0.9828  [CX4H1][NX3H2][CX4H2][CX4H0]     0.0  
[cH][cH]                                   0.9973  [CX4H1][NX3H0][CX4H2][CX4H0]     0.0  
[cX3H1][cX3H1][cX3H0]                   0.9536  [CX4H1][NX3H1][CX4H3][CX4H0]     0.0  
[#6H1][#6H1]                              0.9455  [CX4H1][NX3H1][CX4H2][CX4H2]     0.0  
[#8][#6H0][#6H1]                         0.8961  [CX4H1][NX3H1][CX4H2][CX4H0]     0.0  

worst negatives                           prob           worst positives                 prob
[#6X3H1][#6X3H1][#6X3H0][#6X3H1]          0.4438  [OX2H1]                        0.3102
[#8][#6H1][#6H1]                          0.223   [cX3H0][cX3H1][cX3H0]             0.3169
[#8][#6H][#6X3][#6X3H]                   0.1817  [OX2H][cX3][cX3]                   0.3804
[cX3H1][cX3H0][cX3H0]                    0.1601  [#6][#6][#6][#6][#6]             0.4274
[#8][#6][#6H2]                            0.1257  [#6][#6][#6][#6][#6][#6]         0.5898
[cX3H0][cX3H1][cX3H0][OX2H1]              0.1126  [cX3H1][cX3H1][cX3H1]             0.7576
o[cH]                                     0.1048  [CX4H3]                         0.8693
[#8][#6][#6][#6][#6][#6][#6]             0.0975  [cX4H3][OX2H0]                    0.885
[cX3H1][cX3H0][cX3H1][cX3H0]              0.0953  [cH][cO]                         0.8922
[cX3H0][cX3H1][cX3H1][cX3H0]              0.0953  [#8][#6H0][#6H1]                  0.8961

------------------------------------------------------------------------------------------------------------------------
Example 51 true smiles: Nccccc(H)nH formula: C5H7N3
Index of correct structure: 0 of 5951
True structure loss: 0.015978
True structure:

![Chemical structure diagram]

Experimental 13C NMR (solvent: Benzene-d6)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

1. C5H7N3 0.015978
2. C5H7N3 0.021855
3. C5H7N3 0.023035
4. C5H7N3 0.023718
5. C5H7N3 0.025275
6. C5H7N3 0.028579
7. C5H7N3 0.029698
8. C5H7N3 0.031166
9. C5H7N3 0.031357
10. C5H7N3 0.031773
Example 52 true smiles: C1CN2CCN1CC2 formula: C6H12N2
Index of correct structure: 0 of 5002
True structure loss: 0.010484
True structure:

Experimental $^{13}$C NMR (solvent: CDCl$_3$)

Experimental $^1$H NMR (solvent: CDCl$_3$)

Top predicted structures (loss):

```
| Structure | Loss   |
|-----------|--------|
| ![Structure 1](image1.png) | 0.010484 |
| ![Structure 2](image2.png) | 0.030006  |
| ![Structure 3](image3.png) | 0.035419  |
| ![Structure 4](image4.png) | 0.036905  |
| ![Structure 5](image5.png) | 0.037147  |
| ![Structure 6](image6.png) | 0.038104  |
| ![Structure 7](image7.png) | 0.038266  |
| ![Structure 8](image8.png) | 0.038602  |
| ![Structure 9](image9.png) | 0.040673  |
| ![Structure 10](image10.png) | 0.041273  |
```
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [#7X3][#6H2]               | 0.9364 | [#7][#6H2][#6H2] | 0.6546 | [#7X3][#6H2][#7] | 0.6298 |
| [#7][#6H2][#6H2][#7]      | 0.9076 | [#7][#6][#6][#7] | 0.6298 | [#7][#6][#6][#7] | 0.6207 |
| [#6][#7][#6][#6][#7][1]   | 0.8213 | [#7X3H0]        | 0.55  | [#7][#6H2]     | 0.6834 |
| [#6H2][#7][#6H2]          | 0.7831 | [#7X3H1]        | 0.55  | [#6H2][#6H2]   | 0.6834 |

| worst negatives | prob | worst positives | prob |
|-----------------|------|----------------|------|
| [#7X3H0]        | 0.55  | [#6H3][#6H2][#6H2][#7] | 0.3561 |
| [#6H3][#7]      | 0.3664 | [#7X3H0]        | 0.6298 |
| [#7][#6][#6][#6][#7] | 0.318 | [#7X3H0]        | 0.6207 |
| [#7X3H2]        | 0.2662 | [#7][#6][#6][#7] | 0.6207 |
| [#6H2][#6H2][#6H2][#7] | 0.2369 | [#7][#6][#6][#7] | 0.6207 |
| [#6H1][#6H1]    | 0.2203 | [#7][#6H2]     | 0.6834 |
| [#6][#6][#6][#6][#7][1] | 0.2197 | [#6H2][#7][#6H2][#7] | 0.7831 |
| [#6H1][#6H2][#6H2][#7] | 0.2188 | [#6H2][#6H2][#6H2][#7] | 0.8213 |
| [#6H1][#6H2]    | 0.2164 | [#7][#6H2][#6H2][#7] | 0.9076 |
| [#6H1]          | 0.1993 | [#7X3][#6H2]   | 0.9364 |
Example 53 true smiles: O=c1cc[nH]c(=O)[nH]1 formula: C4H4N2O2
Index of correct structure: 5 of 4792
True structure loss: 0.030097

True structure:

[Chemical structure image]

Experimental 13C NMR (solvent: DMSO)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

1. Structure 1: 0.027947
2. Structure 2: 0.02828
3. Structure 3: 0.028299
4. Structure 4: 0.028616
5. Structure 5: 0.028707
6. Structure 6: 0.030097
7. Structure 7: 0.030224
8. Structure 8: 0.030357
9. Structure 9: 0.03056
10. Structure 10: 0.030795
| Top predicted substructures                     | prob   | best positives                           | prob   | best negatives                          | prob   | worst negatives                           | prob   |
|-------------------------------------------------|--------|------------------------------------------|--------|-----------------------------------------|--------|------------------------------------------|--------|
| [#6H1]                                          | 0.9941 | [cX3H1][{cX3H1}][cX3H0]                  | 0.7864 |                                        |        |                                          |        |
| [#6X3][#6X3]                                    | 0.9803 | [#7][#6][#6][#6X3]                       | 0.738  |                                        |        |                                          |        |
| [#6X3H1][#6X3H0]                                | 0.9093 | [cH][cH]                                 | 0.7282 |                                        |        |                                          |        |
| [cH]                                            | 0.8626 | [#7][#6H0][#6H1]                         | 0.6851 |                                        |        |                                          |        |
| [#7][#6][#6X3]                                  | 0.8431 | [#7][#6X3H0][#6X3H1]                     | 0.6356 |                                        |        |                                          |        |
| best positives                                  | prob   | best negatives                           | prob   | worst positives                         | prob   | worst negatives                           | prob   |
| [#6H1]                                          | 0.9941 | [OX2H0][{CX4H2}][{CX4H2}][{CX4H1}][{CX4H1}] | 0.0    |                                        |        |                                          |        |
| [#6X3][#6X3]                                    | 0.9803 | [OX2H0][{CX4H2}][{CX4H1}][{CX4H1}][{CX4H1}] | 0.0    |                                        |        |                                          |        |
| [cX3H1][{cX3H1}][cX3H0]                         | 0.9093 | [CX4H0][{NX3H1}][{CX4H3}][{CX4H3}][{CX4H3}][{CX4H3}] | 0.0    |                                        |        |                                          |        |
| [cH][cH][cH]                                    | 0.8626 | [CX4H1][{CX4H3}][{CX4H3}][{CX4H3}][{CX4H3}] | 0.0    |                                        |        |                                          |        |
| [#6X3H1][#6X3H0]                                | 0.8431 | [CX4H0][{CX4H3}][{CX4H3}][{CX4H3}][{CX4H3}] | 0.0    |                                        |        |                                          |        |
| [#7][#6][#6X3]                                  | 0.7864 | [CX4H0][{CX4H1}][{CX4H1}][{CX4H1}][{CX4H1}] | 0.0    |                                        |        |                                          |        |
| [cX3H1][{cX3H1}][cX3H0]                         | 0.738  | [OX2H0][{OX2H1}][{OX2H1}][{OX2H1}][{OX2H1}] | 0.0    |                                        |        |                                          |        |
| [cH][cH][cH]                                    | 0.7282 | [OX2H0][{CX4H1}][{CX4H1}][{CX4H1}][{CX4H1}][{CX4H1}] | 0.0    |                                        |        |                                          |        |
| [#7][#6H0][#6H1]                                | 0.6851 | [CX4H2][{CX4H0}][{CX4H2}][{CX4H2}][{CX4H2}] | 0.0    |                                        |        |                                          |        |
| [#7][#6X3H0][#6X3H1]                            | 0.6356 | [OX2H0][{CX4H2}][{CX4H1}][{CX4H1}][{CX4H1}][{CX4H1}] | 0.0    |                                        |        |                                          |        |
| worst negatives                                 | prob   | worst positives                         | prob   | worst negatives                           | prob   | worst negatives                           | prob   |
| [#8][#6H0][#6H1]                                | 0.4699 | [cX3H1][{cX3H1}][{X3H1}]                | 0.0977 |                                        |        |                                          |        |
| [OX2H1]                                         | 0.4697 | [#7][#6X3H1]                            | 0.1612 |                                        |        |                                          |        |
| [#6X3][#6X3][#6X3][#6X3]                        | 0.4444 | [#8][#6][#6H1][#6H1]                     | 0.2121 |                                        |        |                                          |        |
| [#8][#6][#6][#6]                                | 0.4013 | [#7][#6][#6H1][#6][#7]                   | 0.2122 |                                        |        |                                          |        |
| [#8][#6][#6][#6X3]                              | 0.3816 | [#7][#6H0][#7]                          | 0.278  |                                        |        |                                          |        |
| [#7X3H2]                                        | 0.3228 | [#7X3H1]                                | 0.4124 |                                        |        |                                          |        |
| [OX2H1][{cX3}][c]                               | 0.3076 | O=#6][#6][#6X3]                          | 0.4135 |                                        |        |                                          |        |
| [cH][cH][cH]                                    | 0.2842 | [OX1H0]=[{cX3H0}][{cX3H1}]              | 0.4206 |                                        |        |                                          |        |
| [#7][#7]                                        | 0.2616 | O=[{cX3}]                               | 0.4341 |                                        |        |                                          |        |
| [cX3]=[{OX1}]O                                  | 0.2715 | [#7][#6][#7]                            | 0.463  |                                        |        |                                          |        |
Example 54 true smiles: NCc1ccccn1 formula: C6H8N2
Index of correct structure: 0 of 4358
True structure loss: 0.012161
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

0.012161
0.020249
0.021371
0.025413
0.030034

0.031805
0.035658
0.035862
0.037262
0.037527
Top predicted substructures

| Substructure                          | prob | prob |
|---------------------------------------|------|------|
| [#6H1]                                | 0.9999 | 0.9518 |
| [#7][#6][#6X3]                        | 0.9927 | 0.9483 |
| [cH][cH]                              | 0.9887 | 0.9449 |
| [#6X3][#6X3]                          | 0.9842 | 0.935 |
| [cX3H1][cX3H1][cX3H1]                | 0.9703 | 0.9303 |

Best positives

| Substructure                          | prob | prob | prob |
|---------------------------------------|------|------|------|
| [#6H1]                                | 0.9999 | [OX2H0][CX4H2][CX4H2][CX4H1][CX4H1] | 0.0 |
| [#7][#6][#6X3]                        | 0.9927 | [CX4H0][OX2H1][OX2H1][OX2H1][OX2H1][OX2H1] | 0.0 |
| [cH][cH]                              | 0.9887 | [#8][#6H1][#6H2][#6H1][#6H1] | 0.0 |
| [#6X3][#6X3]                          | 0.9842 | [OX2H0][CX4H2][CX4H2][CX4H1][OX2H0] | 0.0 |
| [cX3H1][cX3H1][cX3H1][cX3H1][cX3H1]  | 0.9703 | [OX2H0][CX4H2][CX4H1][CX4H1][OX2H1] | 0.0 |
| [cH]                                  | 0.9518 | [OX2H0][CX4H2][CX4H1][CX4H1][CX4H1][CX4H1] | 0.0 |
| [#6H1][#6H1]                          | 0.9483 | [CX4H1][OX2H1][OX2H1][OX2H1][OX2H1][OX2H1] | 0.0 |
| [cX3H1][nX2H0][cX3H1]                 | 0.9449 | [OX2H1][CX4H1][CX4H1][CX4H1][CX4H1][CX4H1] | 0.0 |
| [#6X3][#6X3][#6X3][#6X3]              | 0.935 | [OX2H0][CX4H2][CX4H1][CX4H1][CX4H1][CX4H1] | 0.0 |
| [cX3H1][cX3H1][cX3H1][cX3H1]          | 0.9303 | [OX2H1][CX4H1][CX4H1][CX4H1][CX4H1][CX4H1] | 0.0 |

Best negatives

| Substructure                          | prob | prob | prob |
|---------------------------------------|------|------|------|
| [#6X3H1][#6X3H1][#6X3H0][#6X3H1]     | 0.4853 | [#7][#6][#6][#7] | 0.2992 |
| [CX4H3]                               | 0.4706 | [#7X3H2] | 0.5031 |
| [#7X3H1]                              | 0.2605 | [#6X3][#7][#6X3] | 0.5142 |
| [cX3H1][cX3H3]                        | 0.2578 | [#7H2][#6H2] | 0.528 |
| [#7][#6][#7]                          | 0.244 | [#7X3][#6H2] | 0.6095 |
| [#6H1][#7][#6H1]                      | 0.2378 | [#7][#6H2] | 0.6933 |
| [#6H3][#7]                            | 0.1691 | [#7][#6H0][#6H1] | 0.7268 |
| [cX4H2][CX3H]                         | 0.1506 | [#7][#6X3H0][#6X3H1] | 0.7429 |
| [#7][#6H0][#7]                        | 0.1391 | [#6X3][#6H2][#7] | 0.8122 |
| [#7H2][#6H0]                          | 0.1267 | [#6X3H1][#6X3H0] | 0.8232 |
Example 55 true smiles: Ccoccc(W)al formula: C6H8N2

Index of correct structure: 0 of 4358

True structure loss: 0.017386

True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

![Structures with corresponding losses]

- 0.017386
- 0.023935
- 0.02415
- 0.024627
- 0.02566
- 0.025368
- 0.027708
- 0.03032
- 0.032023
- 0.035372
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [H]                         | 0.9981 | [cX3H1][{cX3H1}][{cX3H0}] | 0.977 | [OX1H0]=[{cX3H0}][{cX3H0}] | 0.0 |
| [H]                         | 0.9943 | [OX2H0][{cX3H0}][{cX3H0}][{cX3H0}] | 0.0 | [OX2H0][{cX3H0}][{cX3H0}][{cX3H0}] | 0.0 |
| [CX4H3]                     | 0.9924 | [OX2H0][{cX3H0}][{cX3H0}][{cX3H0}] | 0.0 | [OX2H0][{cX3H0}][{cX3H0}][{cX3H0}] | 0.0 |
| [H]                         | 0.9852 | [OX2H0][{cX3H0}][{cX3H0}][{cX3H0}] | 0.0 | [OX2H0][{cX3H0}][{cX3H0}][{cX3H0}] | 0.0 |
| [H]                         | 0.9783 | [OX2H0][{cX3H0}][{cX3H0}][{cX3H0}] | 0.0 | [OX2H0][{cX3H0}][{cX3H0}][{cX3H0}] | 0.0 |

| worst negatives             | prob | worst positives | prob |
|-----------------------------|------|-----------------|------|
| [H][{H}]                    | 0.6413 | [{H}][{H}]     | 0.3514 |
| [H][{H}]                    | 0.554  | [{H}][{H}][{H}][{H}][{H}] | 0.422 |
| [H][{H}][{H}]               | 0.5275 | [{H}][{H}][{H}] | 0.4561 |
| [H][{H}][{H}]               | 0.277  | [{H}][{H}][{H}] | 0.4617 |
| [H][{H}][{H}]               | 0.2544 | [{H}][{H}][{H}] | 0.5142 |
| [H][{H}][{H}]               | 0.2539 | [{H}][{H}][{H}] | 0.5202 |
| [H][{H}][{H}]               | 0.2366 | [{H}][{H}][{H}] | 0.5282 |
| [H][{H}][{H}]               | 0.1952 | [{H}][{H}][{H}] | 0.5860 |
| [H][{H}][{H}]               | 0.1843 | [{H}][{H}][{H}] | 0.6043 |

| worst negatives             | prob | worst positives | prob |
|-----------------------------|------|-----------------|------|
| [H][{H}][{H}]               | 0.1842 | [{H}][{H}][{H}] | 0.6043 |

| worst negatives             | prob | worst positives | prob |
|-----------------------------|------|-----------------|------|
| [H][{H}][{H}]               | 0.1842 | [{H}][{H}][{H}] | 0.734 |
Example 56 true smiles: NNc1ccccc1 formula: C6H8N2
Index of correct structure: 1 of 4358
True structure loss: 0.019757
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

0.018309  0.019757  0.021042  0.024399  0.032753

0.035244  0.043157  0.04647  0.050038  0.051222
| Substructure | Top Predicted Substructures | Best Positives | Best Negatives | Worst Negatives | Worst Positives | Prob |
|--------------|-----------------------------|----------------|---------------|----------------|----------------|------|
| [cH][cH]     | 0.9995                      | [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cH][cH]     | 0.9994                      | [cH]             | 0.9875        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
| [cX3H1][cX3H1][cX3H0] | 0.9884        | [cX3H1][cX3H1][cX3H1] | 0.9844        | [cX3H1][cX3H1][cX3H1] | 0.9844        |      |
Example 57 true smiles: Cc1cccc(N)n1 formula: C₆H₈N₂
Index of correct structure: 0 of 4358
True structure loss: 0.016728
True structure:

Experimental ¹³C NMR (solvent: CDCl₃)

Experimental ¹H NMR (solvent: CDCl₃)

Top predicted structures (loss):

- 0.016728
- 0.022537
- 0.023097
- 0.023302
- 0.024167
- 0.025751
- 0.029256
- 0.030268
- 0.030902
- 0.037129
| Top predicted substructures | prob | best positives | best negatives | prob | worst negatives | worst positives | prob |
|-----------------------------|------|----------------|----------------|------|----------------|----------------|------|
| [#6H1]                      | 0.9989 | [#7] [#6] [#6X3] | [OX2H0] [CX4H2] [CX4H1] [CX4H1] | 0.0 | [#6X3] [#6X3] | [#7] [#6X3] | 0.2438 |
| [cH][cH]                    | 0.9967 | [#6] [#6H0] | [OX2H0] | 0.2453 |
| #6X3 [#6X3]                 | 0.9881 | [#6] [#6H0] | [OX2H0] | 0.2453 |
| [cX3H1][cX3H1][cX3H0]      | 0.9852 | [cH]          | [OX2H1] | 0.4221 |
| [cH]                        | 0.9807 | [#6] [#6H0] | [OX2H0] | 0.6721 |
| #6X3 [#6X3]                 | 0.9724 | [#6X3] [#6X3] | [OX2H0] | 0.7687 |
| [cX3H1][cX3H1][cX3H0]      | 0.968 | [#6H3] [#6] | [cX3H1] | 0.8309 |
| [#6X3] [#6H0]               | 0.9596 | [#6] [#6X3] | [cX3H1] | 0.8473 |

---

| #6X3 [#7X3] [#6X3]          | 0.7699 | [#7] [#6H0] | [OX2H0] | 0.2438 |
| #7X3H1                      | 0.629  | [#7X3H0]    | [OX2H0] | 0.2453 |
| #6X3H1 [#6X3H1] [#6X3H0]   | 0.5472 | [#7] [#6H0] | [OX2H0] | 0.2453 |
| #7X3H1 [#6X3H1]             | 0.2591 | [#7] [#6H0] | [OX2H0] | 0.2453 |
| [cX3H1][cX3H0][cX3H1]      | 0.2153 | [#6] [#6H0] | [OX2H0] | 0.2453 |
| #6H3 [#7]                   | 0.2029 | [#6H3] [#6H0] | [OX2H0] | 0.2453 |
| #7X3H0                      | 0.2019 | [#6H3] [#6X3] | [OX2H0] | 0.2453 |
| #7X3H1 [#6] [#6] [#7]      | 0.1778 | [#cX3H1] [cX3H1] | [OX2H0] | 0.2453 |
| [cX3H1] [#6X3H0] [cX3H1]   | 0.1711 | [#7] [#6X3H0] | [OX2H0] | 0.2453 |

---

| Worst predicted substructures | prob | Worst predicted substructures | prob |
|--------------------------------|------|--------------------------------|------|
| [cX3H1][cX3H1][cX3H0]        | 0.9881 | [cX3H1][cX3H1][cX3H0]        | 0.9881 |
| [cX3H1][cX3H1][cX3H0]        | 0.9852 | [cX3H1][cX3H1][cX3H0]        | 0.9852 |
| [cX3H1][cX3H1][cX3H0]        | 0.968 | [cX3H1][cX3H1][cX3H0]        | 0.968 |
| [cX3H1][cX3H1][cX3H0]        | 0.9596 | [cX3H1][cX3H1][cX3H0]        | 0.9596 |
| [cX3H1][cX3H1][cX3H0]        | 0.9539 | [cX3H1][cX3H1][cX3H0]        | 0.9539 |
Example 58 true smiles: NCCCNCCCNN formula: C7H19N3
Index of correct structure: 0 of 4058
True structure loss: 0.010241
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

0.010241 0.016137 0.020638 0.022335 0.023653

0.023821 0.025178 0.028091 0.031008 0.03336
Top predicted substructures

| Structure                                      | prob          |
|------------------------------------------------|---------------|
| [#7X3H2]                                      | 1.0           |
| [CX4H2][(#6)][#6]                             | 0.9995        |
| [#7X3][#6H2]                                  | 0.9936        |
| [#7][#6H2]                                    | 0.9891        |
| [#7H2][#6H2]                                  | 0.9885        |
| best positives                                | prob           |
| [#7X3H2]                                      | 1.0           |
| [CX4H2](#6)[#6]                               | 0.9995        |
| [#7X3][#6H2]                                  | 0.9936        |
| [#7][#6H2]                                    | 0.9891        |
| [#7H2][#6H2]                                  | 0.9885        |
| worst negatives                               | prob           |
| [#7H2][#6H1]                                  | 0.5241        |
| [CX4H2][(#6)][#6]                             | 0.4136        |
| [#6H1]                                        | 0.3811        |
| [CX4H2][(NX3H2)][CX4H1]                       | 0.2582        |
| [#7][#6][#7]                                  | 0.2526        |
| worst positives                               | prob           |
| [#7H2][#6H2][#7]                              | 0.1199        |
| [CX4H2][#6H2][#6H2]                           | 0.1152        |
| [CX4H3][#6]                                   | 0.1152        |

---
Example 59 true smiles: NCCCC(=O)O formula: C5H11NO2
Index of correct structure: 0 of 3703
True structure loss: 0.006702
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

- 0.006702
- 0.042572
- 0.05026
- 0.060383
- 0.062468

- 0.064524
- 0.065611
- 0.066211
- 0.068058
- 0.069829
| Substructure                          | Prob | Substructure                          | Prob |
|--------------------------------------|------|--------------------------------------|------|
| [CX4H2][{#6}][#6]                   | 0.9999 | [CX4H2][{CX4H2}][CX4H2]              | 0.9688 |
| [OX2H1]                             | 0.9971 | [7][#6H2][#6H2]                      | 0.9334 |
| [CX3](=O)[OX2H1]                    | 0.9947 | [#7][#6H2]                           | 0.9257 |
| [CX4H2][CX4H2]                      | 0.9898 | [OX2H1]                               | 0.9436 |
| [OX2H1]                             | 0.9999 | [CX2H1][#6H2][CX3H]                  | 0.9371 |
| [CX3](=O)[OX2H1]                    | 0.9947 | [#7][#6H2]                           | 0.9334 |
| [CX4H2][{#6H2}][#6H2]               | 0.9688 | [OX2H1]                               | 0.9436 |
| [OX2H1]                             | 0.9999 | [CX2H1][#6H2][#6H2]                  | 0.9579 |
| [CX3](=O)[OX2H1]                    | 0.9947 | [CX2H1][#6H2][#6H2]                  | 0.9579 |
| [CX4H2][{#6H2}][#6H2]               | 0.9688 | [OX2H1]                               | 0.9436 |
| [CX3](=O)[OX2H1]                    | 0.9947 | [OX2H1]                               | 0.9436 |
| [CX4H2][{#6H2}][#6H2]               | 0.9688 | [OX2H1]                               | 0.9436 |
| [CX3](=O)[OX2H1]                    | 0.9947 | [OX2H1]                               | 0.9436 |
| [CX4H2][{#6H2}][#6H2]               | 0.9688 | [OX2H1]                               | 0.9436 |
| [OX2H1]                             | 0.9999 | [CX2H1][#6H2][#6H2]                  | 0.9371 |
| [CX3](=O)[OX2H1]                    | 0.9947 | [#7][#6H2]                           | 0.9257 |
| [CX4H2][{#6H2}][#6H2]               | 0.9688 | [OX2H1]                               | 0.9436 |
| [OX2H1]                             | 0.9999 | [CX2H1][#6H2][#6H2]                  | 0.9371 |
| [CX3](=O)[OX2H1]                    | 0.9947 | [#7][#6H2]                           | 0.9257 |
| [CX4H2][{#6H2}][#6H2]               | 0.9688 | [OX2H1]                               | 0.9436 |
| [OX2H1]                             | 0.9999 | [CX2H1][#6H2][#6H2]                  | 0.9371 |
| [CX3](=O)[OX2H1]                    | 0.9947 | [#7][#6H2]                           | 0.9257 |
| [CX4H2][{#6H2}][#6H2]               | 0.9688 | [OX2H1]                               | 0.9436 |
| [OX2H1]                             | 0.9999 | [CX2H1][#6H2][#6H2]                  | 0.9371 |
| [CX3](=O)[OX2H1]                    | 0.9947 | [#7][#6H2]                           | 0.9257 |
| [CX4H2][{#6H2}][#6H2]               | 0.9688 | [OX2H1]                               | 0.9436 |
| [OX2H1]                             | 0.9999 | [CX2H1][#6H2][#6H2]                  | 0.9371 |
| [CX3](=O)[OX2H1]                    | 0.9947 | [#7][#6H2]                           | 0.9257 |
| [CX4H2][{#6H2}][#6H2]               | 0.9688 | [OX2H1]                               | 0.9436 |
| [OX2H1]                             | 0.9999 | [CX2H1][#6H2][#6H2]                  | 0.9371 |
| [CX3](=O)[OX2H1]                    | 0.9947 | [#7][#6H2]                           | 0.9257 |
| [CX4H2][{#6H2}][#6H2]               | 0.9688 | [OX2H1]                               | 0.9436 |
| [OX2H1]                             | 0.9999 | [CX2H1][#6H2][#6H2]                  | 0.9371 |
| [CX3](=O)[OX2H1]                    | 0.9947 | [#7][#6H2]                           | 0.9257 |
| [CX4H2][{#6H2}][#6H2]               | 0.9688 | [OX2H1]                               | 0.9436 |
| [OX2H1]                             | 0.9999 | [CX2H1][#6H2][#6H2]                  | 0.9371 |
| [CX3](=O)[OX2H1]                    | 0.9947 | [#7][#6H2]                           | 0.9257 |
| [CX4H2][{#6H2}][#6H2]               | 0.9688 | [OX2H1]                               | 0.9436 |
| [OX2H1]                             | 0.9999 | [CX2H1][#6H2][#6H2]                  | 0.9371 |
| [CX3](=O)[OX2H1]                    | 0.9947 | [#7][#6H2]                           | 0.9257 |
| [CX4H2][{#6H2}][#6H2]               | 0.9688 | [OX2H1]                               | 0.9436 |
| [OX2H1]                             | 0.9999 | [CX2H1][#6H2][#6H2]                  | 0.9371 |
Example 60 true smiles: NCCCC(=O)O formula: C5H11NO2
Index of correct structure: 0 of 3703
True structure loss: 0.013067
True structure:

![Chemical structure diagram]

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

- ![Structure 1] 0.013067
- ![Structure 2] 0.044319
- ![Structure 3] 0.044563
- ![Structure 4] 0.053856
- ![Structure 5] 0.055854

- ![Structure 6] 0.05827
- ![Structure 7] 0.060447
- ![Structure 8] 0.060671
- ![Structure 9] 0.062129
- ![Structure 10] 0.063969
| Top predicted substructures | prob | best positives | prob | worst negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [CX4H2][[\#6]][\#6]       | 0.9997 | [CX2H1][\#6][\#6][CX4H1][\#6] | 0.9267 | [CX4H1][\#6][\#6] | 0.2864 |
| [OX2H1]                    | 0.9955 | [CX2H1][\#6][\#6][CX4H1][\#6] | 0.9198 | [CX4H1][\#6][\#6] | 0.4032 |
| [CX1][\#6][CX1]C           | 0.9939 | [CX2H1][\#6][\#6][CX4H1][\#6] | 0.9149 | [CX4H1][\#6][\#6] | 0.5129 |
| [OX2H1][\#6][CX4H1]        | 0.9903 | [CX2H1][\#6][\#6][CX4H1][\#6] | 0.9055 | [CX4H1][\#6][\#6] | 0.5139 |
| [CX3][\#6][CX3]H2          | 0.9354 | [CX2H1][\#6][\#6][CX4H1][\#6] | 0.8731 | [OX2H1][\#6][\#6] | 0.8079 |
| [OX2H1][\#6][CX4H1]        | 0.9267 | [CX2H1][\#6][\#6][CX4H1][\#6] | 0.8731 | [OX2H1][\#6][\#6] | 0.8115 |
Example 61 true smiles: CCCCCC(N)C(=O)O formula: C5H11NO2
Index of correct structure: 0 of 3703
True structure loss: 0.01618
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

- 0.01618
- 0.040724
- 0.043792
- 0.047883
- 0.052978
- 0.054491
- 0.055547
- 0.056181
- 0.057161
- 0.058635
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [CX4H3]                     | 1.0  | [CX4H3][CX4H2] | 0.9839 | [CX2H1][CX2H0][CX3H1] | 0.0 |
| [#6H3][#6][#6]              | 0.9994 | [O2H1]              | 0.9725 | [CX2H0][#6H1]              | 0.0 |
| [CX4H2][#6][#6]             | 0.9958 | [#7X3H2]              | 0.9698 | [CX2H1][#6H1]              | 0.0 |
| [CX4H3][#6]                 | 0.9953 | O=#[CX3][CX4H]          | 0.9613 | [O2H1][#6H1]              | 0.0 |
| [CX3][#6]                   | 0.9974 | O=#[CX3][CX4H]          | 0.953  | [CX2H1][#6H1]              | 0.0 |
| [OX2H1]                     | 0.9725 | O=#[CX3][CX4H]          | 0.953  | [CX2H1][#6H1]              | 0.0 |
| [#6H1]                      | 0.9698 | O=#[CX3][CX4H]          | 0.953  | [CX2H1][#6H1]              | 0.0 |
| [#7X3H2]                    | 0.9613 | O=#[CX3][CX4H]          | 0.953  | [CX2H1][#6H1]              | 0.0 |

| worst negatives | prob | worst positives | prob |
|-----------------|------|-----------------|------|
| [#6H3][#6][#6]  | 0.7254 | [CX4H2][CX4H2]  | 0.3276 |
| [#8] [#6][#6H1][#6H1] | 0.6122 | [CX4H2][CX4H2]  | 0.361  |
| [#7H2][#6H0]    | 0.55  | [#8][#6H0][#6H1] | 0.4191 |
| [#7][#6H0][#6H1] | 0.4041 | [CX4H2][CX4H2]  | 0.4832 |
| [CX4H2][#6][#6H3] | 0.3063 | [#7][#6][#6X3]  | 0.555  |
| #6H3][#6H1][#6H1] | 0.2858 | [OX2H1][OX1H0]  | 0.7282 |
| [#6H1][#6H1]    | 0.2632 | [#6H1][#6H2]    | 0.7758 |
| [CX4H3][CX4H1]  | 0.2398 | [OX2H1][OX1H0]  | 0.8049 |
Example 62 true smiles: Nc1ccc(O)cc1

Index of correct structure: 0 of 3639
True structure loss: 0.014456
True structure:

Experimental 13C NMR ( solvent: CDCl3)

Experimental 1H NMR ( solvent: CDCl3)

Top predicted structures (loss):

0.014456
0.01564
0.017236
0.018273
0.019053
0.033679
0.038067
0.038271
0.038935
0.039351
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [#6H1]                      | 0.9951 | [cH][cH] |      | [OX1H0]=[CX3H0][CX4H1][CX4H2][CX4H3] | 0.0 |
| [#6X3][#6X3]                | 0.9937 | [OX2H0][CX4H2][CX4H1][CX4H1][CX4H1] | 0.0 |
| [#6X3][#6X3][#6X3][#6X3]    | 0.9583 | [OX2H0][CX4H2][CX4H1][CX4H1][CX4H1] | 0.0 |
| [#6X3H1][#6X3H0]           | 0.9366 | [#8][#6H1][#6H2][#6H1]=[#8] | 0.0 |
| [cH]                        | 0.9182 | [OX2H1][CX4H2][CX4H1][CX4H1][CX4H1] | 0.0 |
|                          |      | [OX2H1][CX4H2][CX4H1][CX4H1][CX4H1] | 0.0 |
| [*#7][#6][#6X3]            | 0.8843 | [OX2H0][CX4H2][CX4H1][CX4H1][CX4H1] | 0.0 |
| [#6X3][#6X3][#6X3]          | 0.8590 | [OX2H1][CX4H2][CX4H1][CX4H1][CX4H1] | 0.0 |
| worst negatives             | prob | worst positives | prob |-----------------|------|
| [cX3H1][cX3H1][cX3H1]      | 0.7017 | [cX3H0][cX3H1][cX3H1][cX3H0] | 0.4082 |
| [#6X3][#7][#6X3]           | 0.4324 | [#6X3H1][#6X3H1][#6X3H0] | 0.5256 |
| [#6][#6][#6][#6][#7]       | 0.3094 | [OX2H1] | 0.55 |
| [#7][#6][#6][#6][#7]       | 0.2147 | [cH][cO] | 0.5598 |
| [#6][#6][#6][#6][#7]       | 0.1769 | [#7][#6X3H0][#6X3H1] | 0.5652 |
| [#7][#6X3H1]               | 0.1769 | [#8][#6][#6][#6X3] | 0.5818 |
| [#6X3][#6X3][#6X3][#6X3]   | 0.1658 | [#7][#6H0][#6H1] | 0.5827 |
| [cX3H1][cX3H1][cX3H1]      | 0.1526 | [#8][#6H0][#6H1] | 0.595 |
| [#6X3][#7][#6X3]           | 0.1443 | [OX2H1][cX3][c] | 0.6068 |
| [cX3H1][cX3H1][cX3H1]      | 0.1398 | [#7H2][#6H0] | 0.6205 |
Example 63 true smiles: Cc1ccc[nH]c1=O  formula: C6H7NO
Index of correct structure: 2 of 3639
True structure loss: 0.026615
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

0.021851
0.025291
0.026615
0.028358
0.03163
0.03252
0.033951
0.034559
0.038276
0.040865
Top predicted substructures | prob | 
|--------------------------|------| 
| [#6X3] [#6X3]           | 0.9994 | 
| [#6H1]                   | 0.9987 | 
| [CX4H3] [#6]             | 0.9981 | 
| [CX4H3]                  | 0.9971 | 
| [#6H3] [#6H0]            | 0.9876 | 

Best positives | prob | Worst negatives | prob | 
|--------------------------|------|-----------------|------| 
| [#6X3] [#6X3]           | 0.9994 | [#6H3] [#6] [#6H0] | 0.969 | 
| [#6H1]                   | 0.9987 | [#7] [#6] [#6X3] | 0.9507 | 
| [CX4H3] [#6]             | 0.9981 | [#7] [#6H0] | 0.9422 | 
| [CX4H3]                  | 0.9971 | [#6H0] [#6] [#6] | 0.9314 | 
| [#6H3] [#6H0]            | 0.9876 | [#6X3] [#6X3] [#6X3] [#6X3] | 0.9178 |
Example 64 true smiles: CC(=O)c1ccccc[nH]1 formula: C6H7NO
Index of correct structure: 0 of 3639
True structure loss: 0.022279
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

![Chemical structures with loss values](image)
| Top predicted substructures | prob  | best positives | prob  | best negatives | prob |
|-----------------------------|-------|----------------|-------|----------------|-------|
| [#6H1]                      | 0.9964| [OX1H0]=[CX3H0][CX4H3] | 0.9259|                 |       |
| [#6X3][#6X3]                | 0.9911| [OX2H1][CX4H2][CX4H2][CX4H1][CX4H1] | 0.0   |                 |       |
| [CX4H3]                     | 0.9884| [OX2H0][CX4H2][CX4H0][OX2H0] | 0.0   |                 |       |
| [#6H3][#6][#6]              | 0.9632| [OX2H0][CX4H2][CX4H1][CX4H1][CX4H1] | 0.0   |                 |       |
| [OX1H0]=[CX3H0][CX4H3]      | 0.9604| [OX2H1][CX4H1][CX4H1]([CX4H2])[CX4H2] | 0.0   |                 |       |
| [OX2H0][CX4H2][CX4H0][OX2H0] | 0.9225| [OX2H1][CX4H2][CX4H1][CX4H1][CX4H1] | 0.0   |                 |       |
| [OX2H1][CX4H1][CX4H1][CX4H2][CX4H2] | 0.9045| [OX2H0][CX4H1][CX4H1][CX4H1][CX4H1] | 0.0   |                 |       |
| [OX2H1][CX4H2][CX4H2][CX4H2] | 0.8653| [OX2H0][CX4H1][CX4H1][CX4H1][CX4H1] | 0.0   |                 |       |
| [OX2H0][CX4H2][CX4H3][CX4H3] | 0.8462| [OX2H0][OX2H0][OX2H0][CX4H2][CX4H2][CX4H2][CX4H2][CX4H2] | 0.0   |                 |       |

| worst negatives | prob  | worst positives | prob |
|-----------------|-------|-----------------|-------|
| O=[#6][#6]=[#6X3] | 0.5372| [#7][#6H0][#6H1] | 0.2855|
| [CX3H3]+C       | 0.3737| [#7][#6X3H0][#6X3H1] | 0.2927|
| [CX4H2][#6][#6] | 0.2578| [#6H1r5][#7] | 0.3125|
| [#7X3H2]        | 0.2476| [cX3H1][cX3H1][cX3H1] | 0.329 |
| [#7H2][#6H0]    | 0.2456| [#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6][#6] |
Example 65 true smiles: Cc1cccc(=O)[nH]1 formula: C6H7NO
Index of correct structure: 0 of 3639
True structure loss: 0.018889
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):
Top predicted substructures

| Substructure | prob  |
|--------------|-------|
| #6H1         | 0.9989 |
| #6X3#6X3     | 0.9986 |
| cX3H1        | 0.9943 |
| #6X3H3#6X3   | 0.9881 |
| #6X3H1#6X3H0 | 0.9857 |

Best positives

| Substructure | prob  |
|--------------|-------|
| #6H1         | 0.9989 |
| #6X3#6X3     | 0.9986 |
| cX3H1        | 0.9943 |
| #6X3H3#6X3   | 0.9881 |
| #6X3H1#6X3H0 | 0.9857 |

Worst negatives

| Substructure | prob  |
|--------------|-------|
| #8#6H0#6H1  | 0.6411 |
| #6X3H1#6X3H1#6X3H0#6X3H1 | 0.5062 |
| #8#6#6#6X3  | 0.3913 |
| cX3H1#6X3H0#6X3H0  | 0.2753 |
| #7X3H1      | 0.2596 |
| #6X3H1#6X3H1#6X3H0#6X3H0  | 0.2345 |
| #8#6H1#6H1  | 0.2197 |
| #6H3#6H0#6H0#6H0  | 0.1723 |
| #6H3H0#6H0#6H1  | 0.1685 |
| #7X2H0#6H0  | 0.1585 |

---

Example 66 true smiles: Nc1ccc(O)cc1 formula: C6H7NO
Index of correct structure: 0 of 3639
True structure loss: 0.015917
True structure:

![Chemical structure diagram](image)

Experimental 13C NMR (solvent: DMSO)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

![Predicted structures](image)
| Top predicted substructures | prob | best positives | prob |
|-----------------------------|------|----------------|------|
| [#6H1]                      | 0.994| [cH]           | 0.8872 |
| [#6X3] [#6X3]               | 0.9916| [cX3H1]       | 0.8649 |
| [#6X3] [#6X3] [#6X3]        | 0.9268| [#6H1] [#6H1] | 0.8164 |
| [cH]                        | 0.8966| [#7] [#6]      | 0.7911 |

| worst negatives | prob | best negatives | prob |
|-----------------|------|----------------|------|
| [#6X3] [#6X3]   | 0.7149| [cX3H1]       | 0.2257 |
| [#6X3] [#6X3]   | 0.3167| [#8] [#6H0]   | 0.4292 |
| [#6X3] [#6X3]   | 0.1862| [#7] [#6H0]   | 0.4944 |
| [#6X3] [#6X3]   | 0.1815| [OX2H0]       | 0.5059 |
| [#6X3] [#6X3]   | 0.1616| [#6H0]        | 0.5097 |
| [#6X3] [#6X3]   | 0.1569| [OX2H1]       | 0.5117 |
| [#6X3] [#6X3]   | 0.1421| [OX2H0]       | 0.538 |
| [#6X3] [#6X3]   | 0.1382| [OX2H0]       | 0.5859 |
| [CHX3] #C       | 0.1368| [cH]cO        | 0.6107 |
Example 67 true smiles: Cc1ccc(=O)[nH]c1 formula: C6H7NO
Index of correct structure: 7 of 3639
True structure loss: 0.025019
True structure:
### Top predicted substructures

| Substructure                          | Prob   |
|---------------------------------------|--------|
| [#6X3][#6X3]                          | 0.9991 |
| [cX3H1][cX3H1][cX3H0]                | 0.9874 |
| [#6H1]                                | 0.9987 |
| [cXH][cXH]                            | 0.9977 |
| [CX4H3][#6]                           | 0.9958 |
| [cX3H1][#6X3][#6X3][#6X3]             | 0.9926 |

### Best positives

| Substructure                          | Prob   |
|---------------------------------------|--------|
| [#6X3][#6X3][#6]                      | 0.9991 |
| [cX3H1][cX3H1][cX3H0][cX3H1]         | 0.9874 |
| [#6H1]                                | 0.9987 |
| [cXH][cXH]                            | 0.9977 |
| [CX4H3][#6]                           | 0.9958 |
| [cX3H1][#6X3][#6X3][#6X3]             | 0.9926 |

### Worst negatives

| Substructure                          | Prob   |
|---------------------------------------|--------|
| [cX3H1][cX3H1][cX3H0][cX3H1]         | 0.6229 |
| [cX3H1][cX3H1][cX3H0][cX3H0]         | 0.5198 |
| [cXH][cXH]                            | 0.4919 |
| [cX3H1][cX3H1][cX3H1][cX3H0][cX3H0]  | 0.4295 |
| [cX3H1][cX3H1][cX3H0][cX3H0][cX3H1]  | 0.3516 |
| [O2XH][cX3][cX3H1][cX3H0][cX3H0]     | 0.3225 |
| [O2XH][cX3][cX3H1][cX3H0][cX3H0]     | 0.2778 |
| [O2XH][cX3][cX3H1][cX3H0][cX3H0]     | 0.2355 |
| [O2XH][cX3][cX3H1][cX3H0][cX3H0]     | 0.2172 |

---

### Best negatives

| Substructure                          | Prob   |
|---------------------------------------|--------|
| [cX3H1][cX3H1][cX3H0][cX3H1]         | 0.9874 |
| [cX3H1][cX3H1][cX3H0][cX3H0]         | 0.9847 |
| [cXH][cXH]                            | 0.9847 |
| [cX3H1][cX3H1][cX3H0][cX3H1][cX3H0]  | 0.9825 |
| [cX3H1][cX3H1][cX3H0][cX3H1][cX3H0]  | 0.981  |

### Worst positives

| Substructure                          | Prob   |
|---------------------------------------|--------|
| [cX3H1][cX3H1][cX3H0][cX3H1]         | 0.9874 |
| [cX3H1][cX3H1][cX3H0][cX3H0]         | 0.9847 |
| [cXH][cXH]                            | 0.9847 |
| [cX3H1][cX3H1][cX3H0][cX3H1][cX3H0]  | 0.9825 |
| [cX3H1][cX3H1][cX3H0][cX3H1][cX3H0]  | 0.981  |

---

### Worst negatives

| Substructure                          | Prob   |
|---------------------------------------|--------|
| [cX3H1][cX3H1][cX3H0][cX3H1]         | 0.9874 |
| [cX3H1][cX3H1][cX3H0][cX3H0]         | 0.9847 |
| [cXH][cXH]                            | 0.9847 |
| [cX3H1][cX3H1][cX3H0][cX3H1][cX3H0]  | 0.9825 |
| [cX3H1][cX3H1][cX3H0][cX3H1][cX3H0]  | 0.981  |
Example 68 true smiles: Cc1ccc(O)cn1 formula: C₆H₇NO
Index of correct structure: 0 of 3639
True structure loss: 0.014215
True structure:

Experimental ¹³C NMR (solvent: DMSO)

Experimental ¹H NMR (solvent: CDCl₃)

Top predicted structures (loss):

- 0.014215
- 0.01648
- 0.016908
- 0.01717
- 0.018078

- 0.020607
- 0.020746
- 0.021498
- 0.02281
- 0.038462
### Top predicted substructures

| Structure | prob |
|-----------|------|
| [#6H1]    | 0.9995 |
| [#6X3]    | 0.9992 |
| [#6X3]    | 0.9968 |
| [#6H1]    | 0.9949 |
| [#7]      | 0.9946 |

### Best positives

| Structure | prob |
|-----------|------|
| [#6H1]    | 0.9995 |
| [#6X3]    | 0.9992 |
| [#6X3]    | 0.9968 |
| [#6H1]    | 0.9949 |
| [#7]      | 0.9946 |

### Worst negatives

| Structure | prob |
|-----------|------|
| [#6H1]    | 0.8243 |
| [#6X3]    | 0.6093 |
| [#6H1]    | 0.4328 |
| [#6H1]    | 0.2702 |
| [#6H0]    | 0.1784 |
| [#6X3]    | 0.1745 |
| [#6X0]    | 0.1509 |
| [#6H0]    | 0.1391 |
| [#6H0]    | 0.1218 |

### Best negatives

| Structure | prob |
|-----------|------|
| [#6H1]    | 0.0 |
| [#6X3]    | 0.0 |
| [#6X3]    | 0.0 |
| [#6H1]    | 0.0 |
| [#7]      | 0.0 |

### Worst positives

| Structure | prob |
|-----------|------|
| [#6H1]    | 0.0 |
| [#6X3]    | 0.0 |
| [#6X3]    | 0.0 |
| [#6H1]    | 0.0 |
| [#7]      | 0.0 |

---
Example 69 true smiles: O=C1CCC(=O)N1O formula: C4H5NO3
Index of correct structure: 0 of 3337
True structure loss: 0.036294
True structure:

Experimental $^{13}$C NMR (solvent: DMSO-d$_6$)

Experimental $^1$H NMR (solvent: D$_2$O)

Top predicted structures (loss):

- 0.036294
- 0.042469
- 0.042616
- 0.04262
- 0.043709
- 0.044092
- 0.045782
- 0.047579
- 0.048092
- 0.048961
| Substructure | Top Predicted Probability | Top Predicted Substructures |
|-------------|--------------------------|-----------------------------|
| [CX3](=OX1)C | 0.9693 | [CX4H2]CC=O |
| [7]= [6] | 0.9048 | [7X3H1] |
| OX2H1 | 0.7909 | O=[CX3][CX4H] |
| [CX3](=OX1)O | 0.765 | [CX3]O=[O][OX2H1] |
| [7]=[6][6X3] | 0.7026 | [CX4H2][CX3]=O |

| Substructure | Best Positives Probability | Best Positives |
|-------------|--------------------------|----------------|
| [CX3](=OX1)C | 0.9693 | [CX3H0]=[CX3H2][CX4H3][CX4H1] |
| [OX2H1] | 0.7909 | [6X3H2]=[6][6H2][8H] |
| CX4H2=C=O | 0.5649 | [6X3H1]=[6X3H1][6X3H0]=[6X3H1] |
| [7][6] | 0.4891 | CC=CC=CC |
| [CX4H2][CX3]=O | 0.4783 | [CX3H0]=[OX1H0][NX3H0][CX4H2] |
| [7]=[6][6][6]= [6][6][6] | 0.4352 | [CX2H1][CX2H0][CX3H1]=[CX3H0] |

| Substructure | Top Predicted Probability | Top Predicted Substructures |
|-------------|--------------------------|-----------------------------|
| [CX3](=OX1)C | 0.9693 | [CX3H0]=[CX3H2][CX4H3][CX4H1] |
| [OX2H1] | 0.7909 | [6X3H2]=[6][6H2][8H] |
| CX4H2=C=O | 0.5649 | [6X3H1]=[6X3H1][6X3H0]=[6X3H1] |
| [7][6] | 0.4891 | CC=CC=CC |
| [CX4H2][CX3]=O | 0.4783 | [CX3H0]=[OX1H0][NX3H0][CX4H2] |
| [7]=[6][6][6]= [6][6][6] | 0.4352 | [CX2H1][CX2H0][CX3H1]=[CX3H0] |
Example 70 true smiles: O=C(O)CCCC(=O)O, formula: C5H8O4
Index of correct structure: 0 of 3240
True structure loss: 0.006856

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

0.006856 0.049658 0.051419 0.053381 0.067022
0.069207 0.072057 0.073864 0.078546 0.078714
Top predicted substructures

| Substructure | prob |
|--------------|------|
| [#8]=[#6] | 0.9995 |
| [CX4H2]=([#6])=[#6] | 0.9994 |
| [CX3]=([Ox1])C | 0.9993 |
| [OX2H1] | 0.999 |
| [CX3]=[#6] | 0.9987 |

Best positives

| Substructure | prob |
|--------------|------|
| [#8]=[#6] | 0.9995 |
| [CX4H2]=([#6])=[#6] | 0.9994 |
| [CX3]=([Ox1])C | 0.9993 |
| [OX2H1] | 0.999 |
| [CX3]=[#6] | 0.9987 |

Worst negatives

| Substructure | prob |
|--------------|------|
| [#8][#6][#6] | 0.4232 |
| [#8][#6][#6] | 0.2917 |
| [#8][#6][#6] | 0.2645 |
| [#6][#6][#6] | 0.2518 |
| [#6][#6][#6] | 0.1751 |
| [#6][#6][#6] | 0.1518 |
| [#6][#6][#6] | 0.1502 |
| [CX4H2][CX4H2][CX4H2] | 0.1401 |
| [#8][#6][#6][#6] | 0.131 |

Best negatives

| Substructure | prob |
|--------------|------|
| [#8]=[#6] | 0.9995 |
| [CX4H2]=([#6])=[#6] | 0.9994 |
| [CX3]=([Ox1])C | 0.9993 |
| [OX2H1] | 0.999 |
| [CX3]=[#6] | 0.9987 |

Worst positives

| Substructure | prob |
|--------------|------|
| [#8][#6][#6][#6][#6] | 0.4232 |
| [#8][#6][#6][#6][#6] | 0.2917 |
| [#8][#6][#6][#6][#6] | 0.2645 |
| [#6][#6][#6][#6][#6] | 0.2518 |
| [#6][#6][#6][#6][#6] | 0.1751 |
| [#6][#6][#6][#6][#6] | 0.1518 |
| [#6][#6][#6][#6][#6] | 0.1502 |
| [CX4H2][CX4H2][CX4H2] | 0.1401 |
| [#8][#6][#6][#6][#6] | 0.131 |
Example 71 true smiles: CCC(=O)O)C(=O)O formula: C5H8O4
Index of correct structure: 0 of 3240
True structure loss: 0.022372
True structure:

Experimental 13C NMR (solvent: CDCl3, DMSO-d6)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

0.022372
0.036184
0.046682
0.052213
0.056183
0.057518
0.060492
0.061483
0.061659
0.063964
| Substructure                        | Probability | Top predicted substructures | Probability |
|------------------------------------|-------------|-----------------------------|-------------|
| [\#8]=[\#6][\#8]                 | 0.9992      | [CX4H2][\#6][\#6]          | 0.9796      |
| [\#8]=[\#6]                      | 0.9999      | [OXY1]                      | 0.968       |
| [\#8]=[\#6]                      | 0.9974      | [CX4H3][CX4H2]              | 0.9487      |
| [\#6H3][\#6][\#6]               | 0.9916      | [CX4H3][\#6]               | 0.8938      |

**Best positives**

| Substructure                        | Probability | Best negatives | Probability |
|------------------------------------|-------------|----------------|-------------|
| [\#8]=[\#6][\#8]                 | 0.9992      | [CX2H1][CX2H0][CX3H1]=[CX3H0] | 0.0          |
| [\#8]=[\#6]                      | 0.9999      | CCC=CC#CC       | 0.0          |
| [\#6H3][\#6][\#6]               | 0.9916      | [\#7][\#6][\#6]=[\#6][\#7] | 0.0          |
| [CX4H2][\#6][\#6]               | 0.9796      | [\#7][\#6][\#6]=[\#6][\#7] | 0.0          |
| [CX4H3][\#6]                     | 0.968       | CCC=CC#CC       | 0.0          |
| [OX2H1]                           | 0.9487      | [OXY1][OXY2]     | 0.0          |
| [OCC][CH2]                        | 0.9179      | CCC=CC#CC       | 0.0          |
| [CX4H3][\#6]                     | 0.8938      | [\#6X3H1]=[\#6X3H1][\#6X3H0]=\#6X3H1 | 0.0          |

**Worst negatives**

| Substructure                        | Probability | Worst positives | Probability |
|------------------------------------|-------------|-----------------|-------------|
| [\#8]=[\#6][\#6H2]               | 0.6206      | O=[\#6][\#6][\#6X3] | 0.0707      |
| [\#8]=[\#6][\#6]=[\#8]          | 0.5807      | [CX4H1][\#6X3][\#6X0][\#6X0] | 0.3057      |
| [\#8]=[\#6][\#8]                | 0.5149      | [\#6X3][\#6][\#6X3] | 0.4135      |
| [\#6][\#6H3]                     | 0.4777      | [\#8][\#8][\#6][\#6X3] | 0.4636      |
| [CX4H0]                           | 0.4307      | [\#6X0][\#8X0][\#8X0][\#8X0] | 0.5555      |
| [\#8][\#6][\#6H1][\#6H2]       | 0.217       | [\#6H1][\#6H2] | 0.6022      |
| [\#8][\#6H1]                     | 0.2032      | O=[\#6][\#6X0][\#6X0] | 0.6614      |
| O=[CX4H1][\#6X0][\#6X0][\#6X0][\#6X2] | 0.1817  | [CX4H2][\#6H1] | 0.6936      |
| O=[CX4H1][\#6X0][\#6X0][\#6X0][\#6X2] | 0.1784  | [\#8][\#6H0][\#6H1] | 0.7015      |
Example 72 true smiles: COC(=O)CC(=O)OC  formula: C5H8O4
Index of correct structure: 0 of 3240
True structure loss: 0.014072
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

- 0.014072
- 0.037693
- 0.040552
- 0.056913
- 0.065685
- 0.067063
- 0.068575
- 0.069116
- 0.070151
- 0.070197
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [#8]=[#6][#8]               | 0.9995 | [OX1H0]=[CX3H0][OX2H0][CX4H3] | 0.9199 |
| [CX3]=([=OX1])O            | 0.9985 | [CX4H3]         | 0.9055 |
| [CX3]=([=OX1])C            | 0.9961 | [CX4H2][CX3]=O  | 0.8437 |
| [OX1H0]=[CX3H0][#8][CX4H2] | 0.9675 | [CX4H2][CX3H0][CX3H0] | 0.7876 |
| [#8][#6][#6H2]             | 0.9433 | [CX4H3][OX2H0]  | 0.785  |
| [OX1H0][CX3H0]             | 0.9395 | [CX4H3][OX2H0]  | 0.785  |
| [CX4H2][OX2H0][CX4H3]      | 0.9199 | [CX2H0][CX2H0]  | 0.785  |
| [OX4H3]                     | 0.9055 | [#7][#6][#6][#6]=[#7] | 0.0   |
| [CX4H2][#8][#3]            | 0.8437 | [CX3H1][OX2H0][CX2H0] | 0.0   |
| [OX4H2][CX4H0][OX3H0]      | 0.7876 | [CX3H1][CX4H1][CX4H2] | 0.0   |
| [OX4H2][OX2H0]             | 0.785  | [CX3H1][CX4H2][OX2H0] | 0.0   |
| worst negatives            | prob | worst positives | prob | worst positives | prob |
| [OX1H0][=OX3H0][OX2H1]     | 0.5191 | [OX1H0][=OX3H0][OX2H1][OX4H2] | 0.6372 |
| [OX2H1]                     | 0.5179 | [#6X3][#6H2][#6X3] | 0.6722 |
| [CX4H2][OX2H0][OX4H3]      | 0.5075 | [#6][#6][#6]=[#6] | 0.2881 |
| [#8][=OX1H0][=OX2H1][=O]  | 0.5191 | [OX1H0][=OX3H0][OX2H1][OX4H2] | 0.6372 |
| OCC[CH2]                    | 0.4732 | [#8][#6][#6][#6X3] | 0.6722 |
| [OX4H2][#8]=#6=O           | 0.3523 | [#8X1][#6X3][#6H2][#6X3] | 0.6963 |
| O=[OX3][CX4H1]             | 0.2589 | [OX3H1][OX2H0][OX4H2] | 0.7071 |
| [#8][=OX1H0][=OX2H1][=O]  | 0.2368 | [OX1H0][=OX3H0][OX2H1][OX4H2] | 0.7075 |
| [#6H1][#6][#8]             | 0.208  | [CX4H3][OX2H0]  | 0.785  |
| [#6H1]                      | 0.199  | [CX4H2][OX3H0][OX2H0] | 0.7876 |
Example 73 true smiles: COClCCC(OC)O1 formula: C6H12O3
Index of correct structure: 0 of 3020
True structure loss: 0.009313
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

```
0.009313  0.021802  0.041438  0.04503  0.052808
0.056076  0.057516  0.058041  0.05979  0.061794
```
| Top predicted substructures | prob |
|-----------------------------|------|
| [OX2H0][CX4H1][OX2H0]      | 1.0  |
| [CX4H3]                     | 0.9973 |
| [CX4H2][#6][#6]             | 0.9969 |
| [CX4H0]                     | 0.9959 |
| [#6H1]                      | 0.9938 |

| best positives | prob |
|----------------|------|
| [OX2H0][CX4H1][OX2H0]      | 1.0  |
| [CX4H3]                     | 0.9973 |
| [CX4H2][#6][#6]             | 0.9969 |
| [CX4H0]                     | 0.9959 |
| [#6H1]                      | 0.9938 |

| worst negatives | prob |
|----------------|------|
| [#8][#6][#6H2]                          | 0.2965 |
| [OX2H0]                                    | 0.1441 |
| [OH][CX4H]                                 | 0.1313 |
Example 74 true smiles: COC(CC(C)=O)OC formula: C6H12O3
Index of correct structure: 0 of 3020
True structure loss: 0.009806
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

![Predicted structures with their respective losses]
| Top predicted substructures | prob |
|----------------------------|------|
| [CX4H3]                    | 1.0  |
| [CX4H3][=O(1)]C            | 0.9997 |
| [CX4H3][O(1)]=O(CX3H0)     | 0.9993 |
| [CX4H3][O(1)]=O(CX3H0)     | 0.9978 |
| [CX4H3][O(1)]=O(CX3H0)     | 0.9976 |

| best positives          | prob |
|-------------------------|------|
| [CX4H3]                 | 1.0  |
| [CX4H3][=O(1)]C         | 0.9997 |
| [CX4H3][O(1)]=O(CX3H0)  | 0.9993 |
| [CX4H3][O(1)]=O(CX3H0)  | 0.9978 |
| [CX4H3][O(1)]=O(CX3H0)  | 0.9976 |

| worst negatives         | prob |
|-------------------------|------|
| OCC[CH2]                | 0.5148 |
| [O(1)][#6][#6][#6]=O(8) | 0.3456 |
| [CX4H3][O(1)]=O(CX3H0)  | 0.2748 |
| [O(1)][#6][#6H2][#6H1]  | 0.254 |
| [CX4H3][O(1)]=O(0)      | 0.2386 |
| [O(1)][#6][#6]=O(8)     | 0.2345 |
| [O(1)][#6][#6H2][#6H3]  | 0.2013 |
| [O(1)][#6][#6H2][#6H3]  | 0.2009 |
| [O(1)][#6][#6H2][#6H3]  | 0.1827 |
| [O(1)][#6][#6H2][#6H3]  | 0.1803 |
Example 75 true smiles: CCOC(=O)CC(C)O  formula: C6H12O3
Index of correct structure: 0 of 3020
True structure loss: 0.031697
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

0.031697  0.039367  0.043515  0.046002  0.048921

0.049044  0.04967  0.050377  0.051692  0.052115
Top predicted substructures prob

[CX4H3] 1.0 [#6H3][#6][#6] 0.9953
[OX1] 0.9986 [CX4H3][CX4]O 0.9908
[OX1] 0.9984 [OX2H1] 0.9903
[CX4H3][#6] 0.9979 [CX4H2][#6][#6] 0.9799
[CX1][#6H1] 0.9971 [CX4H2][#6][#6] 0.9706

best positives prob best negatives prob

[OX1] 1.0 C=CC=C=O 0.0
[OX1] 0.9986 [CX2H0][#6H2][#6H1] 0.0
[CX4H3][#6] 0.9979 [CX2H0][#6H2][#6H1] 0.0
[CX4H3][#6] 0.9971 [#6H2][#6H1] 0.0
[CX4H3][#6] 0.9953 C=CC=C=O 0.0
[CX4H3][#6] 0.9908 [CX2H0][#6H2][#6H1] 0.0
[CX4H3][#6] 0.9799 [#6H2][#6H1] 0.0
[CX4H3][#6] 0.9706 [CX2H0][#6H2][#6H1] 0.0

worst negatives prob worst positives prob

[OX1][#6H2][#6H1] 0.9255 [OX1][#6H2][#6H1] 0.1878
[OX1][#6H2][#6H1] 0.7836 [OX1][#6H2][#6H1] 0.2528
[OX1][#6H2][#6H1] 0.6518 [OX1][#6H2][#6H1] 0.2538
[OX1][#6H2][#6H1] 0.634 [OX1][#6H2][#6H1] 0.2555
[OX1][#6H2][#6H1] 0.5426 [OX1][#6H2][#6H1] 0.2812
[OX1][#6H2][#6H1] 0.38 [OX1][#6H2][#6H1] 0.3086
[OX1][#6H2][#6H1] 0.3667 [OX1][#6H2][#6H1] 0.3252
[OX1][#6H2][#6H1] 0.3665 [OX1][#6H2][#6H1] 0.3313
[OX1][#6H2][#6H1] 0.2805 [OX1][#6H2][#6H1] 0.3742
[OX1][#6H2][#6H1] 0.2401 [OX1][#6H2][#6H1] 0.4506
Example 76 true smiles: COC(=O)C(C)(C)CO formula: C6H12O3
Index of correct structure: 0 of 3020
True structure loss: 0.026019
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

1. 0.026019
2. 0.03807
3. 0.040402
4. 0.041319
5. 0.041943
6. 0.043469
7. 0.049183
8. 0.049239
9. 0.050928
10. 0.052917
| Top predicted substructures | prob |
|-----------------------------|------|
| [CX4H3]                     | 1.0  |
| [#6H3][#6][#6]              | 1.0  |
| [CX4H3][#6]                 | 0.9996|
| [#6][#6]                    | 0.9969|
| [CX3][#OX1][#6]             | 0.9969|

| best positives             | prob |
|-----------------------------|------|
| [CX4H3]                     | 1.0  |
| [#6H3][#6][#6]              | 1.0  |
| [CX4H3]                     | 0.9996|
| [#6][#6]                    | 0.9969|
| [CX3][#OX1][#6]             | 0.9969|

| worst negatives            | prob |
|-----------------------------|------|
| [#6H1]                      | 0.9187|
| [CX3][#OX2H1]               | 0.6856|
| [#6H1][#6]                  | 0.6163|
| [CX3][#6]                   | 0.5652|
| [#6][#6]                    | 0.5512|
| [#6H0][#6H1]                | 0.5503|
| [OX][#CX4H]                 | 0.5489|
| [CX4H3][#6][#6H1]           | 0.52  |
| [CX4H3][#6][#6]             | 0.4838|
| [CX3][#OX1][#6H1]           | 0.4567|
Example 77 true smiles: CCCCC(O)C(=O)O
formula: C6H12O3
Index of correct structure: 0 of 3020
True structure loss: 0.014868
True structure:

![Structure Image]

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

1.  
   ![Structure Image]
   0.014868

2.  
   ![Structure Image]
   0.043137

3.  
   ![Structure Image]
   0.049736

4.  
   ![Structure Image]
   0.051526

5.  
   ![Structure Image]
   0.051833

6.  
   ![Structure Image]
   0.053758

7.  
   ![Structure Image]
   0.054139

8.  
   ![Structure Image]
   0.054479

9.  
   ![Structure Image]
   0.054607

10.  
    ![Structure Image]
    0.057478
| Top predicted substructures                   | prob   | best positives                  | prob   | best negatives                  | prob   |
|-----------------------------------------------|--------|---------------------------------|--------|---------------------------------|--------|
| [CX4H2][{#6}][#6]                           | 0.9999 | [CX4H2][{#6}][#6][#6]           | 0.9999 | [CX4H2][{#6}][#6][#6]           | 0.9969 |
| [#6H3][#6][#6]                               | 0.9999 | [CX4H2][{#6}][#6][#6]           | 0.9999 | [CX4H2][{#6}][#6][#6]           | 0.9969 |
| [CX4H3]                                      | 0.9987 | [CX4H2][{#6}][#6][#6]           | 0.9999 | [CX4H2][{#6}][#6][#6]           | 0.9969 |
| [CX3][{#6X1}][#6]                            | 0.9984 | [OCC][CH2]                      | 0.9989 | [CX3][{#6X1}][#6][#6]           | 0.9969 |
| [OX2H1]                                      | 0.9982 | [CX3][{#6X1}][#6]               | 0.9986 | [CX3][{#6X1}][#6][#6]           | 0.9969 |
| best positives                                | prob   | best negatives                  | prob   | worst negatives                 | prob   |
| [CX4H2][{#6}][#6][#6]                        | 0.9999 | C=CC=CC#C                       | 0.0    | [CX4H2][{#6}][#6][#6]           | 0.4451 |
| [#6H3][#6][#6][#6]                           | 0.9999 | CCC=CC#C                        | 0.0    | [CX4H2][{#6}][#6][#6]           | 0.4013 |
| [CX4H3]                                      | 0.9987 | [CX3H1][#6][#6][#6][CX3H1]      | 0.0    | [#8][#6][#6][#6][#6][#8]        | 0.2414 |
| [CX3][{#6X1}][#6][#6]                        | 0.9994 | [#6H1][#6][#6][#6][#6][#6][#8]  | 0.0    | [#8][#6][#6][#6][#6][#8]        | 0.1468 |
| [OX2H1]                                      | 0.9982 | [CX3H1][#6][#6][#6][#6][#6][#8] | 0.0    | [#8][#6][#6][#6][#6][#6][#8]    | 0.1437 |
| worst negatives                               | prob   | worst positives                 | prob   | worst positives                 | prob   |
| [CX4H2][{#6}][#6]                            | 0.4451 | [#8][#6][#6][#6][#6][#6][#8]    | 0.2349 | [CX4H2][{#6}][#6]               | 0.4013 |
| [CH3][CH2]                                   | 0.4001 | [OX2H1][#6][#6]                 | 0.2531 | [OX2H1][#6][#6]                 | 0.3404 |
| [OX2H1][{#6}[#6][#6]                         | 0.2414 | [CX4H2][{#6}][#6]               | 0.3469 | [CX4H2][{#6}][#6]               | 0.2414 |
| [#8][#6][#6][#6][#6][#6]                     | 0.2159 | [#8][#6][#6][#6][#6][#6][#8]    | 0.5352 | [#8][#6][#6][#6][#6][#6][#8]    | 0.1468 |
| [OX2H1][{#6}[#6][#6]                         | 0.1437 | [OX2H1][{#6}][#6][#6][#6][#6][#8]| 0.5994 | [OX2H1][{#6}][#6][#6][#6][#6][#8]| 0.1468 |
| [OX2H1][{#6}[#6][#6]                         | 0.1303 | [OX2H1][{#6}][#6][#6][#6][#6][#8]| 0.6093 | [OX2H1][{#6}][#6][#6][#6][#6][#8]| 0.1468 |
| [CX4H2][{#6}][#6][#6]                        | 0.1208 | [OX2H1][{#6}][#6][#6][#6][#6][#8]| 0.6379 | [OX2H1][{#6}][#6][#6][#6][#6][#8]| 0.1468 |
| [OX2H1][{#6}][#6][#6][#6][#6][#6][#8]       | 0.1111 | [OX2H1][{#6}][#6][#6][#6][#6][#8]| 0.6845 | [OX2H1][{#6}][#6][#6][#6][#6][#8]| 0.1468 |
Example 78 true smiles: CC(C)CC(O)(=O)O formula: C6H12O3
Index of correct structure: 0 of 3020
True structure loss: 0.01545
True structure:

![Chemical structure image]

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

- ![Structure 1] 0.01545
- ![Structure 2] 0.04494
- ![Structure 3] 0.051578
- ![Structure 4] 0.059144
- ![Structure 5] 0.06263
- ![Structure 6] 0.064006
- ![Structure 7] 0.064863
- ![Structure 8] 0.066002
- ![Structure 9] 0.066504
- ![Structure 10] 0.06662
### Top predicted substructures

| Structure | prob | Structure | prob |
|-----------|------|-----------|------|
| [#6H3]#6#6 | 1.0  | [CX4H3]#6 | 0.9942 |
| [CX4H3] | 0.9995 | [CX3]=O@[OX2H1] | 0.9895 |
| [#6H1] | 0.9985 | [#8]=[#6][#8] | 0.989 |
| [OX2H1] | 0.998 | [CX4H2]([#6])[#6] | 0.9887 |
| [CX3]=[OX1])C | 0.998 | [CX3]=@([OX1])O | 0.994 |

### best positives

| Structure | prob | Structure | prob |
|-----------|------|-----------|------|
| [#6H3]#6#6 | 1.0  | [CX3H0]=[CX3H1]@[OX2H0][CX2H0] | 0.0 |
| [CX4H3] | 0.9995 | [CX2H1][CX2H0][CX3H1]@[CX3H0] | 0.0 |
| [#6H1] | 0.9985 | CC@C=C@C | 0.0 |
| [OX2H1] | 0.998 | [CX2H0][#CX2H1][CX4H0] | 0.0 |
| [CX3]=[OX1])C | 0.998 | [CX2H0][#CX2H1][CX3H0] | 0.0 |
| [CX4H3]#6 | 0.9942 | CC@CCC=C | 0.0 |
| [CX2H1][OX2H1] | 0.9895 | [#6X2]@[#6H1][#6X2] | 0.0 |
| [#8]=[#6][#8] | 0.989 | [CX3H0]=@([CX3H1])[CX4H2][CX2H0] | 0.0 |
| [CX4H2][#6][#6] | 0.9887 | C@CCC=C | 0.0 |
| [CX3]=[OX1])O | 0.9694 | [#7][#6]=@[#6][#6][#7] | 0.0 |

### worst negatives

| Structure | prob | Structure | prob |
|-----------|------|-----------|------|
| [#6X3]#6#6#6H3 | 0.4718 | [#8][#6][#6][#8] | 0.144 |
| [CH3]CC[OH] | 0.4167 | [CX4H][O]CO | 0.2109 |
| [#8]=[#6][#6][#6X3] | 0.3104 | [CX4H2][#CH][#CH] | 0.5079 |
| [CX4H2][CX4H2] | 0.2503 | [CX3H0][@OX1H0][OX2H1][CX4H1] | 0.529 |
| [CX4H1][OX2H1][CX4H2][CX4H2] | 0.1786 | [#8][#6H0][#6H1] | 0.5358 |
| [#8][#6H1][#6H1] | 0.1505 | [CX4H1][OX2H1][CX4H2][CX3H0] | 0.5521 |
| [#8][#6H1][#6H1] | 0.146 | [CX4H2][@CX4H1][CX4H1] | 0.6033 |
| [#8]=#6[#6H2][#6H2] | 0.1447 | O@[CX4H][CX4H2] | 0.7054 |
| [#8]=#6[#6H1][#6H1] | 0.1338 | [#8]=#6[#6X4H1][#6X3H0] | 0.7262 |
| CCCC | 0.1186 | [CX4H][O] | 0.729 |
Example 79 true smiles: CCC(O)(CC)(=O)O  
Index of correct structure: 0 of 3020  
True structure loss: 0.008672  
True structure:
| Top predicted substructures | prob | 
|----------------------------|------|  
| [#6H3][#6][#6]             | 1.0  | [CX4H2][#6][#6]                  | 0.9969 |
| [CX4H3]                    | 1.0  | [OX2H1]                         | 0.9963 |
| [CX4H3][#6]                | 0.9998 | [#8][#6][#6]=[#8]               | 0.9937 |
| [CX4H3][CX4H2]             | 0.9986 | [OX2H1][CX4H0][CX4H2][CX4H3]    | 0.984  |
| [CX3](=O[OX1])C            | 0.9998 | [OX2H1][CX4H0][CX4H2][CX4H3]    | 0.984  |

| best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| [CX4H3][#6][#6] | 1.0  | CC=#CCC#C         | 0.0  |
| [CX4H3]        | 0.9998 | [OX2H1][O[OX1]]   | 0.0  |
| [CX4H3][#6]    | 0.9986 | [CX2H0][OX2H1][CX2H0][CX2H0][CX2H0] | 0.0  |
| [CX4H2][(#6)][#6] | 0.9969 | [OX2H1][OX2H1][CX2H0][CX2H0][CX2H0] | 0.0  |
| [OX2H1]        | 0.9963 | [OX2H1][OX2H1][CX2H0][CX2H0][CX2H0] | 0.0  |
| [#8][#6][#6]=[#8] | 0.9937 | [CX3H1][=O[OX1]]   | 0.0  |
| [OX2H1][OX2H0][CX4H2][CX4H3] | 0.9984 | [#7][#6H1][#6X2]      | 0.0  |
| [#8][#6][#6H2]  | 0.9778 | [CX2H1][OX2H0][CX3H1][=O[OX1]] | 0.0  |

| worst negatives | prob | worst positives | prob |
|-----------------|------|-----------------|------|
| [#8][#6H0][#6H1] | 0.581 | OCC[CH2]        | 0.4724 |
| [CX4H1]O        | 0.425 | [CX4H2][CC=O]   | 0.6979 |
| [CX4H2][OX4H3][CX4H1] | 0.4098 | [#8][#6][#6][#8] | 0.7739 |
| [#8][#6H0][#6H1] | 0.3691 | [CX3][=O][OX2H1] | 0.8499 |
| [OX4H1][OX4H2]  | 0.3389 | [CX4H2][OX4H3][CX4H0] | 0.892 |
| [#6H1]         | 0.3331 | [CX3][=O[OX1]]  | 0.9214 |
| [CX3H0][=O[OX1H0]][OX2H1][OX4H1] | 0.324 | [#6X3][#6][#6][#6H3] | 0.9237 |
| [CX4H][O][O]   | 0.3008 | [#8]=[#6][#8]   | 0.9762 |
| [#6H1][#6H2]   | 0.2877 | [#8][#6][#6H2]  | 0.9778 |
Example 80 true smiles: CC\(\text{C})\text{C}(\text{O})\text{C}(\equiv\text{O})\) formula: C\(6\)H\(12\)O\(3\)

Index of correct structure: 0 of 3020
True structure loss: 0.020035
True structure:

Experimental 1\(3\)C NMR (solvent: CD\(3\)C\(3\))

Experimental 1\(H\) NMR (solvent: D\(2\)O)

Top predicted structures (loss):

- \(0.020035\)
- \(0.051365\)
- \(0.057001\)
- \(0.057718\)
- \(0.06121\)
- \(0.061549\)
- \(0.061987\)
- \(0.063275\)
- \(0.063397\)
- \(0.064033\)
| Top predicted substructures | prob | best positives | prob |
|-----------------------------|------|----------------|------|
| [#6H3][#6][#6]              | 1.0  | [#6H1]         | 0.9954 |
| [CX4H3]                     | 1.0  | [#8]=[#6][#8]  | 0.9947 |
| [OX2H1]                     | 0.9991 | [CX4H3][#6]    | 0.9943 |
| [CX3]([=OX1])C              | 0.9984 | [CX3][=OX1])O | 0.9802 |
| [CX1][=O][OX2H1]            | 0.9977 | [CHX4][=CH3X4)]CHX3X4 | 0.9068 |
| best positives              | prob | best negatives | prob |
| [#6H3][#6][#6]              | 1.0  | [CX2H0][=CX3H1)[=OX2H0][=CX2H0] | 0.0 |
| [CX4H3]                     | 1.0  | [CX2H1]=[CX2H0][=CX3H1][=CX3H0] | 0.0 |
| [OX2H1]                     | 0.9991 | C=C=C=C=C=C     | 0.0 |
| [CX3]([=OX1])C              | 0.9984 | CC#CCC=CC      | 0.0 |
| [CX3][=O][OX2H1]            | 0.9977 | [CX2H0][=CX2H1][=CX4H0] | 0.0 |
| [#6H1]                      | 0.9954 | CC#CCC=CC      | 0.0 |
| [#8]=[#6][#8]              | 0.9947 | C=C=C=C=C=C     | 0.0 |
| [CX4H3][#6]                | 0.9943 | [CX2H0][=CX2H1][=CX3H0] | 0.0 |
| [OX2H1]                     | 0.9982 | [CX2H0][=CX2H1][=CX3H1] | 0.0 |
| [CX4H3][=CH3X4)CH3X4]       | 0.9068 | C=C=C=C=C       | 0.0 |
| worst negatives             | prob | worst positives | prob |
| [CH3CC[O=H]                 | 0.6814 | [#8][#6][#6][#8] | 0.1115 |
| [#6X3][#6][#6][#6H3]        | 0.6175 | [CX4H][O=CO]   | 0.2115 |
| [#8][#6][#6H1][#6H1]        | 0.3635 | [#6H][=6X4H1][=6X3H0] | 0.3629 |
| [#8][#6][#6][#6X3]          | 0.2167 | [CX4H][=OX2H1][=CX4H2][=CX3H0] | 0.3949 |
| [#8][#6H1][#6H1]            | 0.2121 | O[=CX4H][=CX4H2] | 0.415 |
| CCCCCCC                     | 0.1994 | [CX4H2][=CH1][=CH1] | 0.507 |
| [#6H1][#6H1]               | 0.1652 | [CX4H2][=OX4H1][=CX4H1] | 0.5643 |
| [OX2H1][=OX3H1][=OX2H1][=CX4H2] | 0.1641 | O=C[C=CH2]     | 0.5726 |
| [CX4H2][=CX3]=O             | 0.1207 | [#8][=6H0][=6H1] | 0.5812 |
| [CX4H2][=CX4H2]             | 0.0943 | [CX4H][O=O]    | 0.5939 |
Example 81 true smiles: CC(O)C(N)=O formula: C4H9NO3
Index of correct structure: 0 of 2840
True structure loss: 0.019032

True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

0.019032  0.032936  0.03333  0.036132  0.040698
0.04141  0.047184  0.048481  0.050476  0.050609
| Top predicted substructures | prob | best positives | prob | best negatives | prob | worst positives | prob | worst negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|-----------------|------|----------------|------|
| [CX4H3]                     | 0.9999 | [CX3H](=O)(=O)C | 0.9875 | [OX2H1]       | 0.9848 | [CH3]CC[OH]     | 0.2184 | [CH3]CC[OH]     | 0.2293 |
| [OX2H1]                     | 0.9918 | [CH3]CC[OH]     | 0.9859 | [#8]=[#6H0][#61] | 0.9232 | [CH3]CC[OH]     | 0.2184 | [CH3]CC[OH]     | 0.2293 |
| [#6H1]                      | 0.9875 | [CH3]CC[OH]     | 0.9789 | [CX4H3][#6]   | 0.9918 | [CH3]CC[OH]     | 0.2184 | [CH3]CC[OH]     | 0.2293 |
| [CX4H3][#6]                 | 0.9918 | [CH3]CC[OH]     | 0.9246 | [OX2H1]       | 0.9848 | [CH3]CC[OH]     | 0.2184 | [CH3]CC[OH]     | 0.2293 |
| [OX2H1]                     | 0.9918 | [CH3]CC[OH]     | 0.9246 | [OX2H1]       | 0.9848 | [CH3]CC[OH]     | 0.2184 | [CH3]CC[OH]     | 0.2293 |
| [OX2H1]                     | 0.9918 | [CH3]CC[OH]     | 0.9246 | [OX2H1]       | 0.9848 | [CH3]CC[OH]     | 0.2184 | [CH3]CC[OH]     | 0.2293 |
Example 82 true smiles: CC(O)C(N)=O formula: C4H9NO3
Index of correct structure: 1 of 2840
True structure loss: 0.031016
True structure:
Example 83 true smiles: CC(O)C(N)(=O)O formula: C4H9NO3
Index of correct structure: 0 of 2840
True structure loss: 0.019013
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

| Structure | Loss   |
|-----------|--------|
| ![Structure 1](image1.png) | 0.019013 |
| ![Structure 2](image2.png) | 0.0327   |
| ![Structure 3](image3.png) | 0.035489 |
| ![Structure 4](image4.png) | 0.037584 |
| ![Structure 5](image5.png) | 0.038551 |
| ![Structure 6](image6.png) | 0.038865 |
| ![Structure 7](image7.png) | 0.042675 |
| ![Structure 8](image8.png) | 0.043992 |
| ![Structure 9](image9.png) | 0.044356 |
| ![Structure 10](image10.png) | 0.045511 |
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [CX4H3]                     | 1.0  | [CX2H1][CX2H0][CX3H1]=[CX3H0] | 0.0  | [CX2H1][CX2H0][CX3H1]=0.0 |
| [OX2H1]                     | 0.9821 | [CX2H0][CX2H1][CX4H2] | 0.0  | [CX2H0][CX2H1][CX4H2] | 0.0  |
| [#6H3][#6][#6]              | 0.9819 | [#6X2][#6H1][#6X2] | 0.0  | [#6X2][#6H1][#6X2] | 0.0  |
| [OX2H1][CX4H3][CX4]O        | 0.9773 | CC=CCC#C        | 0.0  | CC=CCC#C        | 0.0  |
| [#6H1]                      | 0.9466 | [CX3H0]=[CX3H1][(CX4H2)][CX2H0] | 0.0  | [CX3H0]=[CX3H1][(CX4H2)][CX2H0] | 0.0  |
| O=CX3(CX4H)                 | 0.9127 | [CX3H1]=[CX3H2][CX2H0] | 0.0  | [CX3H1]=[CX3H2][CX2H0] | 0.0  |
| [#8]=[#6H0][#6H1]           | 0.9064 | CCC=CC#A       | 0.0  | CCC=CC#A       | 0.0  |
| worst negatives             | prob | worst positives | prob | worst positives | prob |
| [#8][#6]=[#8]               | 0.5384 | [#6X3][#6][#6][#6H3] | 0.2192 | [#6X3][#6][#6][#6H3] | 0.2192 |
| [#6H3][#6][#6X3]            | 0.5148 | [#8][#6][#6][#6X3] | 0.2821 | [#8][#6][#6][#6X3] | 0.2821 |
| [CH3][C(OH)]                | 0.2652 | [#8][#6H1][#6H1] | 0.328 | [#8][#6H1][#6H1] | 0.328 |
| [#8][#6][#8]                | 0.2423 | [#6H1][#6H1][#6H1][#7] | 0.3899 | [#6H1][#6H1][#6H1][#7] | 0.3899 |
| [CX4H1][#6][#6]             | 0.2382 | [#6H1][#6H1][#6H1][#6H1] | 0.4927 | [#6H1][#6H1][#6H1][#6H1] | 0.4927 |
| [#7][#6H0][#6H1]            | 0.2256 | [#7H2][#6X4H1][#6X3] | 0.6013 | [#7H2][#6X4H1][#6X3] | 0.6013 |
| [#8][#6H4H1][#6X3H0]        | 0.2222 | [CX4H2](#6X3H0)[#6H1][#7] | 0.6032 | [CX4H2](#6X3H0)[#6H1][#7] | 0.6032 |
| [#8][#6H5][#7]              | 0.2187 | [#7H2][#6H1] | 0.6123 | [#7H2][#6H1] | 0.6123 |
| [CX4H1](O)CO                | 0.211  | [CX3H0]=[CX3H1][#6H0][#6X2H1][#6X3H0] | 0.6604 | [CX3H0]=[CX3H1][#6H0][#6X2H1][#6X3H0] | 0.6604 |
Example 84 true smiles: O=C(=O)C=CC(=O)O formula: C₅H₆O₄
Index of correct structure: 0 of 2762
True structure loss: 0.016883
True structure:

Experimental 13C NMR (solvent: CDCl₃)

Experimental 1H NMR (solvent: D₂O)

Top predicted structures (loss):

![Predicted structures with their losses]

0.016883  0.05633  0.059739  0.071016  0.071765
0.074151  0.07869  0.081863  0.082281  0.083658
| Top predicted substructures   | prob   | best positives   | prob   | best negatives   | prob   |
|--------------------------------|--------|------------------|--------|------------------|--------|
| [#8]=[#6][#8]                 | 0.9998 | [CX3]=O[OX2H1]  | 0.9559 |                   |        |
| [CX3]=[OX1]O                  | 0.9992 | [CHX3]=C        | 0.9552 |                   |        |
| [#6H1]                        | 0.9781 | [OX2H1]         | 0.9543 |                   |        |
| O=C[CH2][CX3H1]               | 0.9628 | [OX1H0]=([CX3H0][CX4H2][CX3H1] | 0.9305 |                   |        |
| [CX3]=[OX1]C                  | 0.9561 | [CX4H2][#6][#6] | 0.928  |                   |        |

| worst negatives               | prob   | worst positives  | prob   | worst negatives   | prob   |
|--------------------------------|--------|------------------|--------|------------------|--------|
| [CX3H0][#6][OX1H0][OX2H1][CX3H0] | 0.6393 | [CX3H1]=([CX3H1][CX3H0] | 0.2616 |                   |        |
| [CX4H2][CC]=O                  | 0.6297 | [#8]=[#6][#6H1] | 0.3787 |                   |        |
| [OX2H0][CX3H0][CX4H2]         | 0.3754 | [#8]=[#6][#6H2][#6H1] | 0.4918 |                   |        |
| [#8][#6][#6][#8]              | 0.1611 | [OX1H0]=([CX3H0][CX3H1]=([CX3H1] | 0.5227 |                   |        |
| [OX4H2][CX4H2]                | 0.1489 | O=C[CH3]        | 0.6198 |                   |        |
| [CX3H0][#6][OX1H0][OX2H0][CX4H2] | 0.1372 | O=[#6][#6][#6X3] | 0.6245 |                   |        |
| [#8][#6][#6][#6][#6][#8]      | 0.1308 | [#6X3][#6X3][#6X3] | 0.6268 |                   |        |
| [#6X3][#6]                    | 0.1296 | [CX4H2][CX3H]  | 0.6295 |                   |        |
| [CX3H1][O]                    | 0.1002 | [#8][#6][#6H1] | 0.663  |                   |        |
Example 85 true smiles: C=C(C(=O)O)C(=O)O formula: C5H6O4
Index of correct structure: 0 of 2762
True structure loss: 0.01308
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

- 0.01308
- 0.078413
- 0.083502
- 0.086334
- 0.095294
- 0.098375
- 0.099115
- 0.100337
- 0.107447
- 0.109385
| Top predicted substructures | prob  |
|-----------------------------|-------|
| [#8]=[#6][#8]               | 0.9999|
| [CX3]((=[OX1])O            | 0.9999|
| [#6X3H2]                   | 0.9996|
| [CX3]=([OX1])C              | 0.9991|
| [CX3]=([OX1])O([OX2H1])    | 0.9978|
| [CX3]=([OX1])O             | 0.9999|
| [CX3]=([OX1])O             | 0.9996|
| [CX3]=([OX1])C              | 0.9991|
| [CX3]=([OX1])O([OX2H1])    | 0.9978|
| [CX3]=([OX1])O             | 0.9999|
| [CX3]=([OX1])O             | 0.9996|
| [CX3]=([OX1])C              | 0.9991|
| [CX3]=([OX1])O([OX2H1])    | 0.9978|

| best positives | prob  | best negatives | prob  |
|----------------|-------|----------------|-------|
| [#8]=[#6][#8] | 0.9999| [#6X2][#6H1][#6] | 0.0   |
| [CX3]=([OX1])O| 0.9999| [CX2H0][#CX2H1][CX4H0] | 0.0   |
| [#6X3H2]      | 0.9996| [CX4H1][CX4H1][CX4H1][CX2H0] | 0.0   |
| [CX3]=([OX1])C| 0.9991| [CX2H0][#CX2H0][CX2H0] | 0.0   |
| [CX3]=([OX1])O| 0.9996| [OX2H0][CX4H1][CX4H1][CX4H1][CX4H1] | 0.0   |
| [CX3]=([OX1])O| 0.9996| [CX4H1][CX4H1][CX4H1][CX2H0] | 0.0   |
| [CX3]=([OX1])C| 0.9991| [CX4H1][CX4H1][CX4H1][CX2H0] | 0.0   |
| [CX3]=([OX1])O| 0.9996| [OX2H0][CX4H1][CX4H1][CX2H0] | 0.0   |

| worst negatives | prob  | worst positives | prob  |
|-----------------|-------|-----------------|-------|
| [OX2H0][CX3H0][CX4H2] | 0.4063| [#8][#6][#6][#6][#6][#6] | 0.0349|
| [#6H1]          | 0.2111| [#8][#6][#6][#6][#6][#6] | 0.1646|
| [CX3H0]=([OX1H0])([OX2H0])| 0.1825| [#8][#6][#6][#6][#6][#6] | 0.4123|
| [#8][#6][#6][#6][#6]=[#8] | 0.1558| [CX4H2]CCC0 | 0.5706|
| [#6][#6][#6]=[#6][#6][#6] | 0.1497| [CX3H0]=([CX3H2])([CX4H2])[CX3H0] | 0.7751|
| [#6X3H1][#6X3H0] | 0.1425| [#8][#6][#6]=[#6X3] | 0.777|
| [CX4H2][CX3H]   | 0.1416| [#8][#6][#6][#6][#6X3] | 0.7802|
| [#8][#6]=[#6]=[#6] | 0.1256| [#6]=[#6][#6][#6H2] | 0.8025|
| [#8][#6]=[#8]   | 0.1026| [CX3H2]=([CX3H0] | 0.8101|
| [CX4H3]         | 0.0873| OCC[CH2] | 0.8142|
Example 86 true smiles: COC(=O)C=CC(=O)O formula: C₅H₆O₄
Index of correct structure: 0 of 2762
True structure loss: 0.040183
True structure:

![Chemical structure image]

**Experimental 13C NMR (solvent: CDCl₃)**

**Experimental 1H NMR (solvent: D₂O)**

Top predicted structures (loss):

- ![Structure 1](image1) with loss 0.040183
- ![Structure 2](image2) with loss 0.047029
- ![Structure 3](image3) with loss 0.0508
- ![Structure 4](image4) with loss 0.051253
- ![Structure 5](image5) with loss 0.051843
- ![Structure 6](image6) with loss 0.052214
- ![Structure 7](image7) with loss 0.052267
- ![Structure 8](image8) with loss 0.052804
- ![Structure 9](image9) with loss 0.053883
- ![Structure 10](image10) with loss 0.056626
| Top predicted substructures | prob  | best positives | prob  | best negatives | prob  | worst positives | prob  | worst negatives | prob  |
|-----------------------------|-------|----------------|-------|----------------|-------|----------------|-------|----------------|-------|
| [#6H1]                      | 0.9951  | [CX3]#[OX1]C  | 0.7348  | [OX2H1]        | 0.715 |
| [#8]=[#6]=[#8]             | 0.9951  | [OX2H1]       | 0.7214  | [OX1H0]=[CX3H0]=[OX2H0]=[CX4H3] | 0.6375  |
| [#6X3]=[#6X3]              | 0.8261  | [OX1H0]       | 0.737   | [OX1H0]=[CX3H0]=[OX2H0]=[CX4H3] | 0.6375  |
| [#8]=[#6]=[#6X3]           | 0.737   | [OX1H0]       | 0.737   | [OX1H0]=[CX3H0]=[OX2H0]=[CX4H3] | 0.6375  |
| [#6H1]                      | 0.9951  | [CX4H3]=[NX3H1]=[CX4H3]=[NX3H1]=| 0.6018  | [OX2H0]([NX3H0])=[OX2H0]       | 0.0    |
| [#6H1]                      | 0.9951  | [CX4H2]=[NX3H1]=[NX3H1]=| 0.6018  | [OX2H0]([NX3H0])=[OX2H0]       | 0.0    |
| [#6X3]=[#6X3]              | 0.8261  | [CX4H2]=[NX3H0]=| 0.6018  | [OX2H0]([NX3H0])=[NX3H0]       | 0.0    |
| [#6X3]=[#6X3]              | 0.8261  | [CX4H2]=[NX3H0]=| 0.6018  | [OX2H0]([NX3H0])=[NX3H0]       | 0.0    |
| [#6X3]=[#6X3]              | 0.8261  | [OX2H0]([NX3H0])=[OX2H0]       | 0.0    | [OX2H0]([NX3H0])=[NX3H0]       | 0.0    |
| [#6X3]=[#6X3]              | 0.8261  | [OX2H0]([NX3H0])=[NX3H0]       | 0.0    | [OX2H0]([NX3H0])=[NX3H0]       | 0.0    |
| [#6X3]=[#6X3]              | 0.8261  | [OX2H0]([NX3H0])=[NX3H0]       | 0.0    | [OX2H0]([NX3H0])=[NX3H0]       | 0.0    |
| [#6X3]=[#6X3]              | 0.8261  | [OX2H0]([NX3H0])=[NX3H0]       | 0.0    | [OX2H0]([NX3H0])=[NX3H0]       | 0.0    |
| [#6X3]=[#6X3]              | 0.8261  | [OX2H0]([NX3H0])=[NX3H0]       | 0.0    | [OX2H0]([NX3H0])=[NX3H0]       | 0.0    |
Example 87 true smiles: O=Cc1cccc(O)c1 formula: C7H6O2
Index of correct structure: 2 of 2390
True structure loss: 0.014748
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

0.012713 0.013394 0.014748 0.030149 0.044794

0.045856 0.048388 0.049905 0.051083 0.051486
Top predicted substructures

| Substructure | prob | O=[#6][#6][#6X3] | 0.9974 |
|--------------|------|-------------------|--------|
| #6H1         | 0.9999 | [cH][cH] | 0.991 |
| [C6H1]1=O[=6] | 0.9996 | [cH] | 0.9868 |
| #6X3[#6X3][#6X3][#6X3] | 0.9994 | [#6H1][#6H1] | 0.9561 |
| #6X3H1[#6X3H0] | 0.9985 | [cX3H1][{cX3H1}][cX3H0] | 0.9441 |

Best positives

| Substructure | best negatives | prob |
|--------------|----------------|------|
| {#6X3}[^#6X3] | 1.0 | {OX2H0}({CX4H2}){CX4H2}{{CX4H1}O}X2H0 | 0.0 |
| #6H1          | 0.9999 | {CX4H1}{{NX3H2}C}{CX4H2}C{CX3H1} | 0.0 |
| C6H1=O[=6]   | 0.9996 | {CX4H0}{{NX3H1}{{CX4H3}}}{CX4H2}{{CX4H1}}X4H1 | 0.0 |
| #6X3[#6X3][#6X3][#6X3] | 0.9994 | {X2H0O}{{CX4H2}C}{CX4H2}{{CX4H1}}{CX4H1}1 | 0.0 |
| #6X3H1[#6X3H0] | 0.9985 | {OX2H0}1{CX4H2}{{OX2H1}C}{CX2H0} | 0.0 |
| O=[#6][#6][#6X3] | 0.9974 | {CX4H1}{{NX3H0}C}{CX4H2}{{CX4H1}}X4H1 | 0.0 |
| [cH][cH]     | 0.9991 | {CX4H1}{{NX3H1}{{CX4H3}}}{CX4H2}C | 0.0 |
| #6H1         | 0.9868 | {CX4H1}{{OX2H1}C}{CX4H2}{{CX4H0}} | 0.0 |
| {#6H1}[#6H1] | 0.9561 | {CX4H1}{{NX3H1}{{CX4H2}C}{CX4H0} | 0.0 |
| [eX3H1]{{cX3H1}C}X3H0] | 0.9441 | {CX4H0}{{NX3H1}{{CX4H2}C}{CX4H2}{{CX4H1}}X4H1 | 0.0 |

Worst negatives

| Substructure | worst positives | prob |
|--------------|----------------|------|
| [CX3]=([OX1]C | 0.5104 | {cX3H1}{{cX3H0}C}{cX3H0} | 0.1209 |
| [#8]=[#6][#6H1][#6H1] | 0.3707 | {cX3H0}1{cX3H1}C{cX3H0}1{OX2H1} | 0.1317 |
| [cX3H0]{{cX3H1}C}X3H0] | 0.2449 | {cH}O | 0.2764 |
| O=[cX3H1] | 0.1634 | {OX2H}{{cX3}C}c | 0.2979 |
| [#8]=[#6H0][#6H1] | 0.1448 | [#8]1[#6][#6][#6][#6] | 0.3558 |
| [#8]=[#6H][#6X3]=[#6X3H1] | 0.1373 | [OX2H1] | 0.5037 |
| [#8][#6H][#6X3][#6X3H] | 0.1056 | [#8][#6X3][#6H1] | 0.6102 |
| [#8][#6H][#6X3][#6X3H] | 0.0906 | [#8][#6H][#6X3][#6X3H] | 0.6396 |
| [OX1H0][{cX3H0}C]X3H1] | 0.0877 | [#6X3H1]{{#6X3H1}C}{#6X3H0}X3H1 | 0.7175 |
| [C6H1]1=[{C6H1}C]X3H1] | 0.0872 | [#6][#6][#6][#6][#6][#6] | 0.7588 |
Example 88 true smiles: clccc2[nH]ccc2c1 formula: C8H7N
Index of correct structure: 0 of 2370
True structure loss: 0.008669
True structure:
Example 89 true smiles: \texttt{N\#CC(\textcolor{red}{c}ccc\textcolor{red}{c}\textcolor{red}{c})}\ formula: C8H7N
Index of correct structure: 0 of 2370
True structure loss: 0.010376
True structure:

![Chemical structure](image)

Experimental $^{13}$C NMR (solvent: CDCl$_3$)

---

Experimental $^1$H NMR (solvent: CDCl$_3$)

---

Top predicted structures (loss):

- ![Structure 1](image) 0.010376
- ![Structure 2](image) 0.037804
- ![Structure 3](image) 0.038478
- ![Structure 4](image) 0.038617
- ![Structure 5](image) 0.039793
- ![Structure 6](image) 0.042286
- ![Structure 7](image) 0.042591
- ![Structure 8](image) 0.04324
- ![Structure 9](image) 0.043593
- ![Structure 10](image) 0.044861

---


| Substructures                                                                 | Probability | Substructures                                                                 | Probability |
|-------------------------------------------------------------------------------|-------------|-------------------------------------------------------------------------------|-------------|
| [#6H1]                                                                        | 0.9911      | [#7][#6][#6][#6X3]                                                            | 0.8348      |
| [#6X3][#6X3]                                                                  | 0.974       | [#6][#7]                                                                     | 0.8272      |
| [cH]                                                                           | 0.9498      | [#6X3][#6X3][#6X3][#6X3]                                                     | 0.8072      |
| [#6H1][#6H1]                                                                  | 0.8974      | [cX3H1][{cX3H1}][cX3H0]                                                      | 0.7411      |
| [cH][cH]                                                                       | 0.8905      | [#6X3H1][#6X3H0]                                                             | 0.7025      |

| Best Positives                                                                | Probability | Best Negatives                                                                | Probability |
|-------------------------------------------------------------------------------|-------------|-------------------------------------------------------------------------------|-------------|
| [#6H1]                                                                        | 0.9911      | [OX2H0][CX4H2][CX4H2][CX4H1][CX4H1]                                          | 0.0         |
| [#6X3][#6X3]                                                                  | 0.974       | [OX2H1][CX4H0][CX4H2][CX4H0]                                                | 0.0         |
| [cH]                                                                           | 0.9498      | [OX2H1][CX4H2][CX4H1]                                                       | 0.0         |
| [#6H1][#6H1]                                                                  | 0.8974      | [OX2H0][CX4H2][CX4H2][CX4H1][OX2H0]                                        | 0.0         |
| [cH][cH]                                                                       | 0.8905      | [#8][#6][#6][#6][#6][#6][#6][#6][#6]                                        | 0.0         |

| Worst Negatives                                                               | Probability | Worst Positives                                                               | Probability |
|-------------------------------------------------------------------------------|-------------|-------------------------------------------------------------------------------|-------------|
| [#6X3][#7][#6X3]                                                              | 0.3284      | [#6X3H1][#6X3H1][#6X3H0][#6X3H1]                                             | 0.4732      |
| [#7][#6][#6][#6X3]                                                            | 0.2673      | [CX2H0][CX4H2][#6X3H0]                                                      | 0.5449      |
| [cX3H1]=cX3H3                                                                 | 0.2171      | [CX2H0][#6X1][CX4H2]                                                        | 0.551       |
| [#7][#6][#6X3]                                                                 | 0.2057      | [CX4H2][#6][#6]                                                             | 0.5849      |

| Worst Positives                                                               | Probability | Worst Positives                                                               | Probability |
|-------------------------------------------------------------------------------|-------------|-------------------------------------------------------------------------------|-------------|
| [#7][#7][#6][#6X3]                                                            | 0.1484      | [cX3H1][{cX3H1}][cX3H1]                                                      | 0.5978      |
| [#6X3][#7][#6X3]                                                              | 0.1366      | [cX3H1][{cX3H1}][cX3H1]                                                      | 0.6208      |
| [#6H3][#6][#6X3]                                                               | 0.1208      | [#6H2][#6X2]                                                                | 0.6864      |
| [cX3H0][{cX3H1}][{cX3H0}]                                                     | 0.1184      | [cX3H1][{cX3H1}][cX3H0]                                                      | 0.7411      |
| [CX4H2][{CX4H2}][{cX3H0}]                                                     | 0.1108      | [#6X3][#6X3][#6X3][#6X3]                                                    | 0.8072      |
Example 90 true smiles: clcc2[nH]ccc2cl1 formula: C8H7N
Index of correct structure: 0 of 2370
True structure loss: 0.015342
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

0.015342
0.016765
0.024085
0.02584
0.02696

0.026964
0.027941
0.03207
0.035016
0.042892
| Top predicted substructures | prob  | best positives | prob  | best negatives | prob |
|-----------------------------|-------|----------------|-------|----------------|------|
| [cH]                        | 1.0   | [cX3H1][cX3H1][cX3H1] | 0.9553 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.0   |
| [cH][cH]                    | 0.9982| [cX3H1][cX3H1][cX3H1] | 0.9553 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.0   |
| [cX3H1][cX3H1][cX3H1]      | 0.9977| [cX3H1][cX3H1][cX3H1] | 0.9553 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.0   |
| [cX3H1][cX3H1][cX3H1]      | 0.9976| [cX3H1][cX3H1][cX3H1] | 0.9553 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.0   |
| [cX3H1][cX3H1][cX3H1]      | 0.9952| [cX3H1][cX3H1][cX3H1] | 0.9553 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.0   |
| [cX3H1][cX3H1][cX3H1]      | 0.9744| [cX3H1][cX3H1][cX3H1] | 0.9553 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.0   |
| [cX3H1][cX3H1][cX3H1]      | 0.9553| [cX3H1][cX3H1][cX3H1] | 0.9553 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.0   |
| [cX3H1][cX3H1][cX3H1]      | 0.9436| [cX3H1][cX3H1][cX3H1] | 0.9553 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.0   |
| [cX3H1][cX3H1][cX3H1]      | 0.935 | [cX3H1][cX3H1][cX3H1] | 0.9553 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.0   |
| [cX3H1][cX3H1][cX3H1]      | 0.9074| [cX3H1][cX3H1][cX3H1] | 0.9553 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.0   |

| worst negatives | prob  | worst positives | prob |
|-----------------|-------|-----------------|------|
| [#6][#6][#6][#6][#6][#7][1] | 0.519 | [cX3H1][cX3H1] | 0.1423 |
| [#6][#6][#6][#6][#6][#6][1] | 0.3843 | [cX3H1][cX3H1] | 0.265 |
| [cX3H1][cX3H1][cX3H1] | 0.3199 | [cX3H1][cX3H1][cX3H1] | 0.2913 |
| [cX3H1][cX3H1][cX3H1] | 0.1613 | [cX3H1][cX3H1][cX3H1] | 0.2929 |
| [cX3H1][cX3H1][cX3H1] | 0.1012 | [cX3H1][cX3H1][cX3H1] | 0.2982 |
| [cX3H1][cX3H1][cX3H1] | 0.0932 | [cX3H1][cX3H1][cX3H1] | 0.3431 |
| [cX3H1][cX3H1][cX3H1] | 0.073 | [cX3H1][cX3H1][cX3H1] | 0.3842 |
| [cX3H1][cX3H1][cX3H1] | 0.0708 | [cX3H1][cX3H1][cX3H1] | 0.4345 |
| [cX3H1][cX3H1][cX3H1] | 0.0694 | [cX3H1][cX3H1][cX3H1] | 0.4396 |
| [cX3H1][cX3H1] | 0.0646 | [cX3H1][cX3H1][cX3H1] | 0.7965 |
Example 91 true smiles: COCCOCC(=O)O formula: C5H10O4
Index of correct structure: 0 of 1865
True structure loss: 0.026528
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

0.026528
0.03734
0.044413
0.046428
0.047324
0.047717
0.049079
0.050718
0.050792
0.05223
| Top predicted substructures | prob   | best positives | prob   | best negatives  | prob  |
|----------------------------|--------|----------------|--------|-----------------|-------|
| [CX4H2][{6}][O]           | 0.9939 | [OX2H1]        | 0.9492 | [OX2H1][{6}][O] | 0.0   |
| [CX3][=OX1][C]            | 0.9898 | [#8]=[#6][#8]  | 0.9316 | [#8]=[#6][#8]  | 0.0   |
| [#8] [#6][#6H2]           | 0.9656 | [#8]=[#6][#6][#8] | 0.896 | [#8]=[#6][#6][#8] | 0.0 |
| [#8] [#6] [#6H2]          | 0.9514 | [#8]=[#6][#6]=[#8] | 0.825 | [#8]=[#6][#6]=[#8] | 0.0 |
| [CX3][=OX1][O]            | 0.9505 | [OX2H0][CX4H2][CX4H2][OX2H0] | 0.8082 | [OX2H0][CX4H2][OX2H0] | 0.0 |
| [OX2H1]                   | 0.9492 | [CX3][=OX1][C] | 0.9898 | [OX2H1][{6}][O] | 0.0 |
| [#8] [#6][#6H2]           | 0.9656 | [#7][#6][#6][#7] | 0.0 | [#7][#6][#6][#7] | 0.0 |
| [CX3][=OX1][O]            | 0.9505 | CC=CCC=CCC=CCC | 0.0 | CC=CCC=CCC=CCC | 0.0 |
| [OX2H1]                   | 0.9492 | [CX2H0][{6}][CX2H0][OX2H0] | 0.0 | [CX2H0][{6}][CX2H0][OX2H0] | 0.0 |
| [#8]=[#6][#8]             | 0.9316 | [OX2H0][{6}][#8] | 0.0 | [OX2H0][{6}][#8] | 0.0 |
| [#8] [#6] [#6]=[#8]       | 0.8986 | [OX2H0][{6}][OX2H0][CX4H2] | 0.0 | [OX2H0][{6}][OX2H0][CX4H2] | 0.0 |
| [O] [CX3][=OX1][C]        | 0.7644 | [OX2H0][{6}][OX2H0][OX2H0] | 0.2785 | [OX2H0][{6}][OX2H0][OX2H0] | 0.2785 |
| [OX2H0][OX1][{6}][OX2H0]  | 0.7529 | [OX2H0][OX2H0][{6}][OX2H0] | 0.3665 | [OX2H0][OX2H0][{6}][OX2H0] | 0.3665 |
| O[CX4H2][{6}][OX2H0]      | 0.6905 | [OX2H0][OX2H0][OX2H0][OX2H0] | 0.3847 | [OX2H0][OX2H0][OX2H0][OX2H0] | 0.3847 |
| [OX2H0][{6}][OX1][OX2H0]  | 0.6822 | [OX2H0][OX2H0][OX2H0][CX4H2] | 0.4226 | [OX2H0][OX2H0][OX2H0][CX4H2] | 0.4226 |
| [#8]=[#6H0][#6H1]         | 0.5431 | [OX2H0][OX2H0][OX2H0][OCX2H2] | 0.465 | [OX2H0][OX2H0][OX2H0][OCX2H2] | 0.465 |
| [OX2H2][OCX2H2][{6}][O]   | 0.5052 | [OX2H0][OX2H0][OX2H0][OX2H0] | 0.4731 | [OX2H0][OX2H0][OX2H0][OX2H0] | 0.4731 |
| [OX2H0][OCX2H2][{6}][O]   | 0.4774 | [OX2H0][OX2H0][OX2H0][OX2H0] | 0.5149 | [OX2H0][OX2H0][OX2H0][OX2H0] | 0.5149 |
| [#8]=[#6][#8]             | 0.4417 | [OX2H0][OX2H0][OX2H0][OX2H0] | 0.5459 | [OX2H0][OX2H0][OX2H0][OX2H0] | 0.5459 |
| [OX2H2][OCX2H2][{6}][#8]  | 0.4139 | [#8]=[#6][#6][#8] | 0.5688 | [#8]=[#6][#6][#8] | 0.5688 |
| O=CX3[CX4H2]              | 0.3913 | [CX4H2][OX2H0][CX4H2][OX2H0] | 0.5734 | [CX4H2][OX2H0][CX4H2][OX2H0] | 0.5734 |
Example 92 true smiles: Cc1cccc(N)c1 formula: C7H9N
Index of correct structure: 0 of 1755
True structure loss: 0.01404
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

- 0.01404
- 0.014554
- 0.016638
- 0.01955
- 0.020187
- 0.026088
- 0.026829
- 0.038456
- 0.039593
- 0.041978
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|---------------|------|
| [#6H1]                      | 0.9981 | [#6H3] | 0.9376 | [#5H1] [#6X3] | 0.9016 |
| [#6X3] [#6X3]               | 0.9977 | [#6X3H1] [#6X3H0] | 0.9803 | [#6X3H1] [#6X3H0] | 0.9448 |
| [cH] [cH]                   | 0.9839 | [cH] [cH] | 0.9155 | [cX3H1] [cX3H0] | 0.8958 |
| [#6X3] [#6X3] [#6X3] [#6X3] | 0.9803 | [#6H3] [#6H0] | 0.9376 | [#6X3H1] [#6X3H0] | 0.9448 |
| [cX3H1] [cX3H0] [cX3H1] [cX3H0] | 0.9977 | [cX3H1] [cX3H0] | 0.9155 | [cX3H1] [cX3H0] | 0.8958 |

| worst negatives | prob | worst positives | prob |
|-----------------|------|----------------|------|
| [#6] [#6] [#6] [#6] [#6] [#6] [#6]   | 0.403  | [cX3H1] [cX3H1] [cX3H0] | 0.113  |
| [cX3H1] [cC3] | 0.392  | [cX3H1] [cX3H0] | 0.2834 |
| [#6X3] [#7] [#6X3] | 0.3811 | [#6X3] [#6X3] | 0.4359 |
| [cX4H2] [#6] [#6] | 0.1712 | [#6X3] [#6X3] | 0.491 |
| [#7] [#6] [#6H3] | 0.1604 | [#6X3] [#6] [#6] | 0.5626 |
| [cX3H1] [cX3H1] [cX3H0] | 0.1428 | [cX3H1] [cX3H1] [cX3H0] | 0.6209 |
| [cX3H3] [cX3H3] | 0.1351 | [cX3H1] [cX3H1] [cX3H1] | 0.6398 |
| [cX4H2] [cX3H1] | 0.1084 | [cX4H3] [#6] | 0.8394 |
Example 93 true smiles: Cc1ccccc1N formula: C7H9N
Index of correct structure: 0 of 1755
True structure loss: 0.015657
True structure:

![Molecule Structure]

Experimental 13C NMR (solvent: CDCl3)

![13C NMR Spectrum]

Experimental 1H NMR (solvent: CDCl3)

![1H NMR Spectrum]

Top predicted structures (loss):

1. ![Structure 1] - 0.015657
2. ![Structure 2] - 0.018607
3. ![Structure 3] - 0.019227
4. ![Structure 4] - 0.027272
5. ![Structure 5] - 0.029253
6. ![Structure 6] - 0.029578
7. ![Structure 7] - 0.034941
8. ![Structure 8] - 0.039918
9. ![Structure 9] - 0.04112
10. ![Structure 10] - 0.048189
| Top predicted substructures | prob | best positives | prob | best negatives | prob | worst negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|----------------|------|
| [#6X3][#6X3] | 0.9979 | [#6X3][#6X3][#6X3][#6X3] | 0.9424 | [#6X3][#6X3][#6X3][#6X3] | 0.9326 | [#6X3][#6X3][#6X3][#6X3] | 0.9289 |
| [#6H1] | 0.9962 | [#6H1][#6H1][#6H1][#6H1] | 0.9450 | [#6H1][#6H1][#6H1][#6H1] | 0.9450 | [#6H1][#6H1][#6H1][#6H1] | 0.9450 |
| [CX4H3][#6] | 0.9958 | [OX2H1][CX4H2][CX4H2][CX4H2][CX4H2] | 0.9436 | [OX2H1][CX4H2][CX4H2][CX4H2][CX4H2] | 0.9436 | [OX2H1][CX4H2][CX4H2][CX4H2][CX4H2] | 0.9436 |
| [CX4H3] | 0.9887 | [OX2H0][CX4H2][CX4H2][CX4H2] | 0.9436 | [OX2H0][CX4H2][CX4H2][CX4H2] | 0.9436 | [OX2H0][CX4H2][CX4H2][CX4H2] | 0.9436 |
| [CH1][CH1] | 0.9576 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.9436 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.9436 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.9436 |
| [cX3H1][#6] | 0.8925 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.9436 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.9436 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.9436 |
| [cX3H1][#6] | 0.8825 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.9436 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.9436 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.9436 |
| [cX3H1][#6] | 0.8556 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.9436 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.9436 | [OX2H0][CX4H2][CX4H2][CX4H2][CX4H2] | 0.9436 |

| worst negatives | prob | worst positives | prob | worst negatives | prob |
|-----------------|------|-----------------|------|-----------------|------|
| [#6X3][#6X3][#6X3][#6X3] | 0.3797 | [#7H2][#6H0] | 0.1473 | [#6X3][#6X3][#6X3][#6X3] | 0.324 | [#7H2][#6H0] | 0.0758 |
| [#CHX3]=[C]C | 0.322 | [#7X3H2] | 0.4268 | [#6X3][#6X3][#6X3][#6X3] | 0.2535 | [#cX3H1][#cX3H1][#cX3H1][#cX3H1] | 0.421 |
| [#6X3][#6]=[#6X3][#6X3] | 0.2174 | [#7][#6X3][#6X3][#6X3] | 0.4638 | [#6X3][#6X3][#6X3][#6X3] | 0.2059 | [#6X3][#6X3][#6X3][#6X3] | 0.5048 |
| [#6X3][#6X3][#6X3][#6X3] | 0.1997 | [#7][#6X3][#6X3][#6X3] | 0.5873 | [#6X3][#6X3][#6X3][#6X3] | 0.1941 | [#cX3H1][#cX3H1][#cX3H1][#cX3H1] | 0.5915 |
| [CX4H2][CX3]=C | 0.1888 | [CX4H3][CX3H0] | 0.6959 | [#6X3][#6X3][#6X3][#6X3] | 0.1866 | [#6X3][#6X3][#6X3][#6X3] | 0.744 |
Example 94 true smiles: CNc1ccc(c1) formula: C7H9N
Index of correct structure: 0 of 1755
True structure loss: 0.009893
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

0.009893 0.041262 0.042546 0.043419 0.044812

0.046776 0.048976 0.054249 0.054935 0.056296
| Top predicted substructures | prob   | [cH] | 0.9739 |
|-----------------------------|--------|------|--------|
| [#6H1]                      | 0.9997 | [cX3H1] | {cX3H1} | {cX3H0} | 0.9728 |
| [#6X3] [#6X3]               | 0.9949 | [cX7H0] | {cX6H0} | {cX6H1} | 0.9542 |
| [cH] [cH]                   | 0.9899 | [cX3H1] | {cX3H1} | {cX3H0} | 0.9153 |
| [cX3H1] [#6X3H0]            | 0.9763 | [cX6H1] | #7     | #6     | 0.8853 |

| best positives             | prob   | [cX3H1] | {cX3H1} | {cX3H0} | 0.0 |
|-----------------------------|--------|---------|--------|--------|-----|
| [#6H1]                      | 0.9997 | [cX2H0] | #6X3H0 | #6X3H0 | 0.0 |
| [#6X3] [#6X3] [#6X3] [#6X3] | 0.9949 | [cX4H0] | {cX4H1} | {cX4H3} | 0.0 |
| [cX7H1]                     | 0.9899 | [cX2H1] | #6X3H0 | #6X3H0 | 0.0 |
| [cH] [cH]                   | 0.9763 | [cX2H0] | #6X3H0 | #6X3H0 | 0.0 |
| [cX3H1] [#cX3H1] [cX3H0]   | 0.9739 | [cX3H1] | {cX3H1} | {cX3H0} | 0.0 |
| [cX6H0] [#cX3H1] [cX3H0]   | 0.5759 | [cX4H3] | #6X3H0 | #6X3H0 | 0.0 |
| [cX7H0] [#6X3H1]            | 0.2969 | [cX2H1] | #6X3H0 | #6X3H0 | 0.0 |
| [cX7H1] [#6X3H1]            | 0.2331 | [cX2H0] | #6X3H0 | #6X3H0 | 0.0 |
| [cX4H2] [#cX3H1] [#cX3H0]  | 0.1981 | [cX6H0] | #6X3H0 | #6X3H0 | 0.0 |

| worst negatives            | prob   | [cX3H1] | {cX3H1} | {cX3H0} | 0.0 |
|-----------------------------|--------|---------|--------|--------|-----|
| [#6H3] [#6X3]               | 0.5759 | [cX4H3] | #6X3H0 | #6X3H0 | 0.0 |
| [#6X3] [#6X3] [#6X3] [#6]   | 0.474  | [cX7H0] | {cX6H0} | {cX6H1} | 0.0 |
| [#6X3] [#6X3] [#6X3]        | 0.2969 | [cX7H0] | {cX6H0} | {cX6H1} | 0.0 |
| [#6X3] [#6X3] [#6X3]        | 0.2498 | [cX7H0] | {cX6H0} | {cX6H1} | 0.0 |
| [cX6H1] [#cX3H1]           | 0.715  | [cX6H0] | #6X3H0 | #6X3H0 | 0.0 |
| [cX6H1] [#6X3H1]            | 0.7185 | [cX6H0] | #6X3H0 | #6X3H0 | 0.0 |
| [cX4H2] [#cX3H1] [#6X3]    | 0.7668 | [cX6H0] | #6X3H0 | #6X3H0 | 0.0 |
| [cX6H3] [#cX3H1] [#6X3]    | 0.8312 | [cX6H0] | #6X3H0 | #6X3H0 | 0.0 |
| [cX6H3] [#cX3H1] [#6X3]    | 0.8725 | [cX6H0] | #6X3H0 | #6X3H0 | 0.0 |

| worst positives            | prob   | [cX3H1] | {cX3H1} | {cX3H0} | 0.0 |
|-----------------------------|--------|---------|--------|--------|-----|
| [#6H3] [#6X3]               | 0.5759 | [cX4H3] | #6X3H0 | #6X3H0 | 0.0 |
| [#6X3] [#6X3] [#6X3] [#6]   | 0.474  | [cX7H0] | {cX6H0} | {cX6H1} | 0.0 |
| [#6X3] [#6X3] [#6X3]        | 0.2969 | [cX7H0] | {cX6H0} | {cX6H1} | 0.0 |
| [#6X3] [#6X3] [#6X3]        | 0.2498 | [cX7H0] | {cX6H0} | {cX6H1} | 0.0 |
| [cX6H1] [#cX3H1]           | 0.715  | [cX6H0] | #6X3H0 | #6X3H0 | 0.0 |
| [cX6H1] [#6X3H1]            | 0.7185 | [cX6H0] | #6X3H0 | #6X3H0 | 0.0 |
| [cX4H2] [#cX3H1] [#6X3]    | 0.7668 | [cX6H0] | #6X3H0 | #6X3H0 | 0.0 |
| [cX6H3] [#cX3H1] [#6X3]    | 0.8312 | [cX6H0] | #6X3H0 | #6X3H0 | 0.0 |
| [cX6H3] [#cX3H1] [#6X3]    | 0.8725 | [cX6H0] | #6X3H0 | #6X3H0 | 0.0 |

| worst positives            | prob   | [cX3H1] | {cX3H1} | {cX3H0} | 0.0 |
|-----------------------------|--------|---------|--------|--------|-----|
| [#6H3] [#6X3]               | 0.5759 | [cX4H3] | #6X3H0 | #6X3H0 | 0.0 |
| [#6X3] [#6X3] [#6X3] [#6]   | 0.474  | [cX7H0] | {cX6H0} | {cX6H1} | 0.0 |
| [#6X3] [#6X3] [#6X3]        | 0.2969 | [cX7H0] | {cX6H0} | {cX6H1} | 0.0 |
| [#6X3] [#6X3] [#6X3]        | 0.2498 | [cX7H0] | {cX6H0} | {cX6H1} | 0.0 |
| [cX6H1] [#cX3H1]           | 0.715  | [cX6H0] | #6X3H0 | #6X3H0 | 0.0 |
| [cX6H1] [#6X3H1]            | 0.7185 | [cX6H0] | #6X3H0 | #6X3H0 | 0.0 |
| [cX4H2] [#cX3H1] [#6X3]    | 0.7668 | [cX6H0] | #6X3H0 | #6X3H0 | 0.0 |
| [cX6H3] [#cX3H1] [#6X3]    | 0.8312 | [cX6H0] | #6X3H0 | #6X3H0 | 0.0 |
| [cX6H3] [#cX3H1] [#6X3]    | 0.8725 | [cX6H0] | #6X3H0 | #6X3H0 | 0.0 |
Example 95 true smiles: Cc1ccnc(C)c1 formula: C7H9N
Index of correct structure: 0 of 1755
True structure loss: 0.013298
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

![Top predicted structures](image)
Top predicted substructures

| Substructure | prob  |
|--------------|-------|
| [#6X3][#6X3] | 0.9933|
| [#6X3][#6][#6] | 0.9932|
| [#6H1] | 0.9931|
| [CX4H3] | 0.9926|
| [#6X3][#6X3][#6X3][#6X3] | 0.9925|

Best positives

| Substructure | prob  | Best negatives | prob  |
|--------------|-------|----------------|-------|
| [#6X3][#6X3] | 0.9933| [#8][#6H1][#6H2][#6H1]=#[8] | 0.0   |
| [#6X3][#6][#6] | 0.9932| [OX2H0][CX4H2][CX4H2][CX4H1][CX4H1]1 | 0.0   |
| [#6H1] | 0.9931| [OX1H0][#6H1][CX4H1][#6H1][CX4H1]1 | 0.0   |
| [CX4H3] | 0.9926| [OX2H0][CX4H2][CX4H2][CX4H1][CX4H1]1 | 0.0   |
| [#6X3][#6X3][#6X3][#6X3] | 0.9925| [OX2H0][CX4H2][CX4H1][#6H1] | 0.0   |

Worst negatives

| Substructure | prob  | Worst positives | prob  |
|--------------|-------|-----------------|-------|
| [CX4H2][(#6)][#6] | 0.6798| [cX3H1][#6X3H0][cX3H0] | 0.237  |
| [CX4H2][CX4H2] | 0.6223| [#7][#6H0][#6H1] | 0.3474 |
| [#6][#6][#6][#6][#6]1 | 0.4063| [cX3H1][#6X2H0][cX3H1] | 0.5246 |
| [cX3H1][#6X3H1][cX3H1] | 0.3669| [#7][#6X3H0][#6X3H1] | 0.5552 |
| [cX3H1][#6X2H0][cX3H0] | 0.2575| [#6X3H1][#6X3H0][#6X3H1] | 0.5771 |
| [cX3H1][#6X3H1][#6X3H0][CX4H3] | 0.2169| [#7][#6][#6H3] | 0.5941 |
| [#6H3][#6H0][#6H1][#7] | 0.2112| [#6X1][#7][#6X3] | 0.6808 |
| [#6H1][#6H2] | 0.1912| [#6][#6][#6][#6][#6][#7]1 | 0.6813 |
| [cX3H0][#6X3H1][cX3H1][cX3H0] | 0.1486| [#7][#6][#6X3] | 0.887  |
| [CX4H2][CX4H2][CX3H1] | 0.1246| [#6H1][#6H1] | 0.8948 |
Example 96 true smiles: Cc1ccc(N)cc1 formula: C7H9N
Index of correct structure: 0 of 1755
True structure loss: 0.012272
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

- 0.012272
- 0.012897
- 0.013551
- 0.021013
- 0.021516

- 0.022322
- 0.024314
- 0.031965
- 0.035487
- 0.048218
Top predicted substructures

| Substructure | prob  |
|--------------|-------|
| [#6X3][#6X3] | 0.9963|
| [#6H1]       | 0.9929|
| [CX4H3][#6]  | 0.9899|
| [#6X3][#6X3][#6X3][#6X3] | 0.9894|
| [CX4H3]      | 0.9874|
| [#6X3][#6X3][#6X3][#6X3][#6X3] | 0.9894|

best positives

| Substructure | prob  |
|--------------|-------|
| [#6X3][#6X3] | 0.9963|
| [#6H1]       | 0.9929|
| [CX4H3][#6]  | 0.9899|
| [#6X3][#6X3][#6X3][#6X3] | 0.9894|
| [CX4H3]      | 0.9874|

worst negatives

| Substructure | prob  |
|--------------|-------|
| [cX3H1][cX3H1][cX3H0] | 0.9414|
| [cX3H1][cX3H1][cX3H0] | 0.7615|
| [#6X3][#6X3][#7] | 0.4605|
| [cX3H1][cX3H1][cX3H0][CX4H3] | 0.2914|
| [#7X3H1] | 0.2149|
| [cX3H1][cX3H0][cX3H0] | 0.2082|
| [#6X3][#6X3][#7] | 0.2082|
| [6X3][#7][#6X3] | 0.1316|
| [6H3][#6] | 0.125|

---

best negatives

| Substructure | prob  |
|--------------|-------|
| [#6X3][#6X3] | 0.9963|
| [#6H1]       | 0.9929|
| [CX4H3][#6]  | 0.9899|
| [#6X3][#6X3][#6X3][#6X3] | 0.9894|
| [CX4H3]      | 0.9874|

worst positives

| Substructure | prob  |
|--------------|-------|
| [cX3H1][cX3H1][cX3H0] | 0.7615|
| [#6X3][#6X3][#7] | 0.4605|
| [cX3H1][cX3H1][cX3H0][CX4H3] | 0.2914|
| [#7X3H1] | 0.2149|
| [cX3H1][cX3H0][cX3H0] | 0.2082|
| [#6X3][#6X3][#7] | 0.2082|
| [6X3][#7][#6X3] | 0.1316|
| [6H3][#6] | 0.125|
Example 97 true smiles: O=C1CCC(=O)C1 formula: C₆H₈O₂
Index of correct structure: 0 of 1578
True structure loss: 0.007099
True structure:

Experimental ¹³C NMR (solvent: CDCl₃)

Experimental ¹H NMR (solvent: CDCl₃)

Top predicted structures (loss):

0.007099  0.032071  0.032482  0.069982  0.0849
0.091144  0.111186  0.119501  0.120469  0.121847
Example 98 true smiles: CCl CCC(=O) O1 formula: C6H10O2
Index of correct structure: 0 of 1567
True structure loss: 0.013298
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

- 0.013298
- 0.062732
- 0.063131
- 0.076545
- 0.078268
- 0.080148
- 0.082432
- 0.085663
- 0.08738
- 0.088436
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|----------------------------|------|----------------|------|----------------|------|
| [CX4H2][{#6}]#{6}          | 1.0  | [#8][#6][#6]   | 0.9888 | [#8][#6][#8]   | 0.9856 |
| [CX4H3]                    | 0.999| [#8][#6][#8]   | 0.9856 | [CX3]{=OX1}O   | 0.9756 |
| [#6H3][#6][#6]             | 0.9979|[CX4H3][#6]    | 0.9845 | [CX4H3][#6]   | 0.9710 |
| [OX1][OX1][C]              | 0.9955| OCC[CH2]       | 0.9411 | [CX4H3][#6]   | 0.9684 |

| worst negatives | prob | worst positives | prob |
|-----------------|------|-----------------|------|
| [#8][#6H0][#6H1]| 0.3823| [#8][#6][#6][#6][#6]=[#8]| 0.2797 |
| [OX1][=O][OX2H1]| 0.3644| [#6H1][#6H2][#6H2]| 0.3485 |
| [OX2H1]        | 0.3437| [CX4H2][CX3]=O | 0.5141 |
| [OX2H2][{CX4H2}][{CX4H2}]| 0.2266| [OX2H0][OX3H0][CX4H2]| 0.5787 |
| [OX3H1]        | 0.2063| [OX3H0]=([OX1H0])([OX2H0])=[CX4H2]| 0.5794 |
| [OX4H2][{CX4H3}][{CX4H2}]| 0.1943| [OX4H2][{CX4H2}][{OX3H0}]=| 0.6016 |
| [#8][#6][#6][#6][#6H3] | 0.1765| [OX2H0][{OX4H1}][{OX4H2}][{CX4H2}][{CX4H2}]=| 0.6804 |
| [#6H3][#6][#6][#6H3] | 0.1456| O=[OX3H0][{OX4H2}][{CX4H2}][{CX4H2}]=| 0.7081 |
| [OX2H1][{OX2H1}][{OX4H2}][{OX3H0}]= | 0.1372| C1OCCC1 | 0.737 |
Example 99 true smiles: CC1CCCC(=O)O1 formula: C6H10O2
Index of correct structure: 0 of 1567
True structure loss: 0.012951
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

- 0.012951
- 0.053548
- 0.057716
- 0.0581
- 0.068806
- 0.078512
- 0.078936
- 0.079061
- 0.079458
- 0.082227
### Top predicted substructures

| Structure                  | Prob | Structure                  | Prob |
|----------------------------|------|----------------------------|------|
| [CX4H2][#6][#6]             | 0.9999 | [#8][#6][#6]               | 0.9887 |
| [CX4H3]                     | 0.9994 | OCC[CH2]                   | 0.978 |
| [CX3]([OX1]C               | 0.9969 | [CX4H3][#6]               | 0.9714 |
| [CX3]([OX1]O               | 0.9945 | [#6H1]                    | 0.9636 |
| [#6H1][#6][#6]             | 0.9938 | [#8][#6][#6H2]            | 0.9499 |

### Best positives

| Structure                  | Prob | Structure                  | Prob |
|----------------------------|------|----------------------------|------|
| [CX4H2][#6][#6]             | 0.9999 | [#7][#6][#6][#6][#6]       | 0.0  |
| [CX4H3]                     | 0.9994 | [CX3H0]([OX2H0])[CX2H0]    | 0.0  |
| [CX3]([OX1]C               | 0.9969 | C=CC=CC=CC               | 0.0  |
| [CX3]([OX1]O               | 0.9945 | OC=CC=CC=CC              | 0.0  |
| [#6H1][#6][#6]             | 0.9938 | [CX3H0]([OX2H1])[CX4H1]    | 0.0  |
| [#8][#6][#6]               | 0.9887 | [CX3H1]([OX2H2])[NX3H0]   | 0.0  |

### Worst negatives

| Structure                  | Prob | Structure                  | Prob |
|----------------------------|------|----------------------------|------|
| [OX3]([Ox2H1]             | 0.5415 | [#8][#6][#6][#6][#6][#6]   | 0.3076 |
| [#6H1][#6H1]              | 0.4763 | O[CX4H][CX4H2]             | 0.4531 |
| O=[CX3][CX4H]             | 0.1904 | [OX2H0][CX4H1][CX4H2][CX4H2] | 0.4649 |
| [OX2H1]                   | 0.1717 | [OX2H0][CX3H0][CX4H2]      | 0.5592 |
| [#8][#6H0][#6H1]          | 0.15  | [CX4H1]O                   | 0.5862 |
| [CX4H2][CX4H2][CX4H2]     | 0.1423 | [CX3H0]([OX2H0])[CX4H2]    | 0.588 |
| [OX3H0]([OX2H0])[CX4H1]   | 0.1377 | [CX4H2]([OX2H2])[CX4H2]    | 0.6125 |
| [#6X3][#6][#6][#6H3]      | 0.1335 | CCC=CCC                    | 0.7465 |
| [OX3H0]([OX2H0])[CX4H1][CX4H2] | 0.1298 | [OX2H0][CX4H1][CX4H3]     | 0.7621 |
| [#6H1][#6H2][#6H2]        | 0.1261 | [CX4H2][CX3]O             | 0.7856 |
Example 100 true smiles: CC=CC(=O)OCC formula: C6H10O2
Index of correct structure: 0 of 1567
True structure loss: 0.013429
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

0.013429
0.058213
0.064394
0.070942
0.073778
0.075099
0.075477
0.080694
0.083619
0.084162
| Substructure                                                                 | Prob  |
|------------------------------------------------------------------------------|-------|
| [CX4H3]                                                                      | 1.0   |
| [CX4H3][#6]                                                                  | 0.9998|
| [#6H1]                                                                       | 0.9996|
| [CHX3][=C]C                                                                  | 0.9961|
| O=[#6][#6]=[#6X3]                                                             | 0.9909|
| best positives                                                               | prob  |
| [CX4H3]                                                                      | 1.0   |
| [CX4H3][#6]                                                                  | 0.9998|
| [#6H1]                                                                       | 0.9996|
| [CHX3][=C]C                                                                  | 0.9961|
| O=[#6][#6]=[#6X3]                                                             | 0.9909|
| worst negatives                                                               | prob  |
| [#6H3][#6][#6]                                                                | 0.7024|
| [O2XH1]                                                                      | 0.5216|
| [#6H3][#6][#6X3]                                                              | 0.501 |
| [CX3][=O][OX2H1]                                                              | 0.2613|
| [#8H][#6H2][#6H1]                                                             | 0.2237|
| [O6XH1]=[#6X3H0]                                                               | 0.2202|
| [CX4H2][CX3][=C]                                                              | 0.2159|
| [#6H1][#6H1]                                                                  | 0.1752|
| [#6H1][#6H2]                                                                   | 0.1688|
Example 101 true smiles: \text{CC=C(C(=O)OC} formula: \text{C}_6\text{H}_{10}\text{O}_2

Index of correct structure: 0 of 1567

True structure loss: 0.016688

True structure:

![Chemical Structure]

Experimental \text{13C NMR} (solvent: CDCl₃)

Experimental \text{1H NMR} (solvent: CDCl₃)

Top predicted structures (loss):

- 0.016688
- 0.054286
- 0.063237
- 0.066651
- 0.067998
- 0.069045
- 0.069707
- 0.072092
- 0.073025
- 0.07791
| Top predicted substructures | prob  |
|-----------------------------|-------|
| [CX4H3][CX3]               | 0.9987|
| [CHX3][=C]                 | 0.9959|
| [CX4H3][#6]                | 0.9983|
| [#6H1]                     | 0.9888|

| best positives             | prob  | best negatives            | prob  |
|-----------------------------|-------|---------------------------|-------|
| [CX4H3]                     | 1.0   | [CX2H0][#]#[CX2H1][]# [CX4H0] | 0.0   |
| [CX4H3][CX3]                | 0.9987| CCC=CC=CC                 | 0.0   |
| [CHX3][=C]                  | 0.9959| [CX2H0][#]#[CX2H0][CX2H0] | 0.0   |
| [CX4H3][#6]                 | 0.9953| [#6X2][#6H1][=C]         | 0.0   |
| [#6H1]                      | 0.9883| [CX2H0][#]# [CX2H2][]# [CX4H2] | 0.0   |
| [#6X3][#6X3]                | 0.9818| [#7][#6H1][#6X2]         | 0.0   |
| [CX4H3][OX2H0]              | 0.9761| [CX4H1][#]# [CX4H1][CX2H0] | 0.0   |
| [#8]=[#6][#6]               | 0.9703| CC#CCC=CC                 | 0.0   |
| [CX3]==[OX1]O               | 0.9578| [CX2H1][#]# [CX2H0]      | 0.0   |
| [OX1H0]==[CX3H0][OX2H0][CX4H3] | 0.9462| [#7][#6][#7][#7]       | 0.0   |

| worst negatives            | prob  | worst positives            | prob  |
|-----------------------------|-------|---------------------------|-------|
| [#6X3H1][#6X3H0]            | 0.5715| [#6H3][#6][#6X3]          | 0.2463|
| [#8][#6H0][#6H1]            | 0.5222| [#6H3][#6][#6H0]         | 0.2868|
| [CX3H1]==[CX3H1][][CX3H0]   | 0.3951| [#6H3][#6][#6]           | 0.363 |
| [#6H1][#6H2]                | 0.3063| [CX4H3][CX4H0]          | 0.4756|
| [CHX3]==[CHX3]              | 0.3   | [#6X1H0]==[CX3H0][CX3H0][CX4H3] | 0.4879|
| OCC[CH2]                    | 0.2782| [CX4H3][CX3H0][CX3]=0    | 0.5088|
| [OX2H1]                     | 0.2497| [#6H3][#6]==[#6][#6H3]  | 0.5642|
| [CX4H1][=][#6][#6]          | 0.2193| [CX3H0][#]# [OX2H0][]# [CX3H0] | 0.6444|
| [#6X3]=[#6X3][#6X3]==[#6X3] | 0.1917| [#6X3H1][#]#[CX3H0][#6X3] | 0.684 |
| [#8]=[#6H0][#6H1]           | 0.169 | [CX3H0][#]# [CX3H1][#]# [CX4H3][#]# [CX3H0] | 0.7601|
Example 102 true smiles: NCC(N)C(=O)O formula: C₃H₈N₂O₂
Index of correct structure: 3 of 1492
True structure loss: 0.033238
True structure:

Experimental ¹³C NMR (solvent: D₂O)

Experimental ¹H NMR (solvent: D₂O)

Top predicted structures (loss):

- 0.027831
- 0.029552
- 0.032872
- 0.033238
- 0.033761
- 0.035094
- 0.035784
- 0.035784
- 0.035784
- 0.035919
| Top predicted substructures | prob  | best positives | prob  | best negatives | prob  |
|-----------------------------|-------|----------------|-------|----------------|-------|
| [#7X3H2]#6H2               | 0.9757 | [CX2H1]#6H2[#6H3]#X3H1=[CX3H0] | 0.00  | [CX2H0]#6H0[#6H1] | 0.0814 |
| [#7]#6H2                   | 0.9703 | [CX4H2][CX4H2] | 0.0905 | [CX4H2][NX3H2]#X3H1 | 0.0933 |
| [CX3]=[OX1]C              | 0.9503 | [CX3][=OX1]O #X2H1=[CX3H0] | 0.00  | [CX3H0]#6H0[#6H1] | 0.0268 |
| [#7X3H2]                   | 0.8974 | [OX2H1] | 0.6901 | [CX3H0]#6H0[#6H1] | 0.2894 |
| [#8]#6H2#6X3              | 0.7911 | [CX4H2]#6H2[#6H3]#X3H1 | 0.00  | [OX2H1] | 0.3873 |
| [#8]#6H2#6X3#7            | 0.747  | [CX2H0]#6H0[#6H1]#7 | 0.0533 | [CX3H0]#6H0[#6H1] | 0.3505 |
| [CX3][=OX1]O              | 0.704  | [CX3][=OX1]O [CX3H0]#6H0[#6H1] | 0.00  | [CX3H0]#6H0[#6H1] | 0.373  |
| [#8]#6H2#6X3#7            | 0.6932 | [CX2H0]#6H0[#6H1]#7 | 0.0533 | [CX3H0]#6H0[#6H1] | 0.3959 |

| worst negatives | prob  | worst positives | prob  | worst negatives | prob  |
|------------------|-------|-----------------|-------|-----------------|-------|
| [#6H2]#7]#6X3    | 0.747  | [#8]#6H0[#6H1] #6X3 | 0.0533 | [CX4H2][NX3H2]#X3H1 | 0.247  |
| [#7]#6H2#6H2    | 0.6776 | [CX4H2][NX3H2]#X3H1 | 0.0533 | [CX4H2][NX3H2]#X3H1 | 0.0268 |
| [CX4H2][CX4H2]  | 0.5541 | [#7]#6H0[#6X3]#6X3 | 0.0633 | [OX2H1]#6H0[#6H1] | 0.3226 |
| [#7]#6H2#6H2#7 | 0.3901 | [CX3H0]=[OX1H0][=OX2H1]#X3H1 | 0.00  | [CX3H0]#6H0[#6H1] | 0.3505 |
| [CX4H2][NX3H2]#X3H1 | 0.3779 | [#7]#6H2#6H1 | 0.0533 | [CX3H0]#6H0[#6H1] | 0.373  |
| [CX4H2][CX3]=O  | 0.3691 | [#6H1]#6H2 | 0.0533 | [CX3]=O[OX2H1] | 0.3505 |
| [#7]#6H0#6H1    | 0.2894 | O=[CX3][CX4H1] | 0.0533 | [CX3]=O[OX2H1] | 0.373  |
| [#7]#6H1#7      | 0.2414 | [#6H1]#6H1 | 0.0533 | [OX2H1]#6H0[#6H1] | 0.3959 |
Example 103 true smiles: O=C1CCCCC1 formula: C5H9NO
Index of correct structure: 0 of 1318
True structure loss: 0.021931
True structure:

![Chemical structure](image)

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

- ![Structure 1](image) 0.021931
- ![Structure 2](image) 0.039314
- ![Structure 3](image) 0.053053
- ![Structure 4](image) 0.056398
- ![Structure 5](image) 0.057907
- ![Structure 6](image) 0.061226
- ![Structure 7](image) 0.066954
- ![Structure 8](image) 0.067872
- ![Structure 9](image) 0.068434
- ![Structure 10](image) 0.069597
Example 104 true smiles: COC(=O)CC(C)=O formula: C5H8O3
Index of correct structure: 0 of 1131
True structure loss: 0.014216
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: d2o)

Top predicted structures (loss):

![Top predicted structures](image-url)
| Substructure | Prob   | Substructure | Prob   |
|--------------|--------|--------------|--------|
| [CX3](= [OX1])C | 0.9999 | [CX4H2][CX3] = O | 0.979  |
| [CX4H3] | 0.996  | [#6H3][#6H0] | 0.9559 |
| [#8]=[#6][#6] | 0.9906 | [CX4H3][CX3] | 0.9502 |
| [OX1H0]=[CX3H0][OX2H0][CX4H3] | 0.988  | [OX1H0]=[CX3H0][#8][CX4H2] | 0.9338 |
| [CX3](= [OX1])O | 0.9944 | [CX4H3][CX3H0] | 0.9332 |

**Best positives**

| Substructure | Prob   | Substructure | Prob   |
|--------------|--------|--------------|--------|
| [CX3](= [OX1])C | 0.9999 | [CX3H0](= [OX1])C | 0.9996 |
| [CX4H3] | 0.996  | [CX4H3] | 0.996  |
| [OX1H0]=[CX3H0][OX2H0][CX4H3] | 0.988  | [#8]=[#6][#6] | 0.9906 |
| [CX3](= [OX1])O | 0.9844 | [#6][#8] | 0.9906 |
| [CX4H2][CX3]=O | 0.979  | [OX1H0]=[CX3H0][OX2H0][CX4H3] | 0.988  |
| [#6H3][#6H0] | 0.9559 | [OX1H0]=[CX3H0][OX2H0][CX4H3] | 0.988  |
| [CX4H3][CX3] | 0.9502 | [CX3](= [OX1])O | 0.9844 |

**Worst negatives**

| Substructure | Prob   | Substructure | Prob   |
|--------------|--------|--------------|--------|
| OCC[CH2] | 0.3999 | [#6X3][#6][#6] | 0.1515 |
| [CX4H2][#C#C] | 0.3695 | [CX4H2][CX3H0][CX3H0] | 0.5711 |
| [CX4H2][#6][O] | 0.3341 | [#8][#6][#6] | 0.572  |
| [#6X3][#6X3] | 0.2794 | O=[#6][#6][#6][#6] | 0.619  |
| [CX3H0]=[OX1H0][OX2H1][CX4H2] | 0.2406 | [#6H3][#6X3H0][#6H2] | 0.6379 |
| [#8]=[#6][#6] | 0.1996 | [#6][#6][#6] | 0.6951 |
| [#6X3][#6H2][#8] | 0.199 | [OX1H0]=[CX3H0][CX4H2][CX3H0] | 0.6999 |
| [OX3H1][CX4H2][#6X3H0] | 0.1949 | [OX1H0][OX2H0][CX3H0][CX4H2] | 0.7147 |
| [#8]=[#6][#6][#6][#6] | 0.1804 | [#8][#6][#6H2] | 0.7324 |
Example 105 true smiles: CC(C)(=O)C(=O)O formula: C5H8O3
Index of correct structure: 0 of 1131
True structure loss: 0.01553
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

- ![Structure 1](image1.png) 0.01553
- ![Structure 2](image2.png) 0.039704
- ![Structure 3](image3.png) 0.070354
- ![Structure 4](image4.png) 0.074231
- ![Structure 5](image5.png) 0.076095
- ![Structure 6](image6.png) 0.083627
- ![Structure 7](image7.png) 0.085353
- ![Structure 8](image8.png) 0.08562
- ![Structure 9](image9.png) 0.092913
- ![Structure 10](image10.png) 0.096611
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [CX3]=[OX1]C               | 1.0  | [CX3]=([OX1])O | 0.986| [CX3]=[OX1]O   | 0.9764|
| [CX4H3]                    | 0.994| [CX3]=[OX1]O   | 0.986| [CX4H3]        | 0.9764|
| [#6H3]=[#6]=[#6]           | 0.9971| [CX3]=([OX1])O | 0.986| [#6H3]=[#6]=[#6]| 0.9725|
| [CX4H3]=[#6]              | 0.9946| [CX3]=([OX1])O | 0.986| [CX4H3]=[#6]   | 0.9725|
| [OX2H1]                    | 0.9936| [CX3]=([OX1])O | 0.986| [OX2H1]        | 0.9725|
| #8]=[#6]=[#6]             | 0.9764| [CX3]=([OX1])O | 0.986| #8]=[#6]=[#6]  | 0.9725|
| [CX3]=[OX1]O              | 0.986 | [CX3]=([OX1])O | 0.986| [CX3]=[OX1]O   | 0.9725|
| [CX4H3]                    | 0.9994| [CX3]=([OX1])O | 0.986| [CX4H3]        | 0.9725|
| [CX3]=([OX1])O             | 0.986| [CX3]=([OX1])O | 0.986| [CX3]=([OX1])O | 0.9725|
| [CX3]=([OX1])O             | 0.986| [CX3]=([OX1])O | 0.986| [CX3]=([OX1])O | 0.9725|
| [CX3]=([OX1])O             | 0.986| [CX3]=([OX1])O | 0.986| [CX3]=([OX1])O | 0.9725|

| worst negatives | prob | worst positives | prob |
|-----------------|------|-----------------|------|
| [CX4H2]=[CX3]   | 0.6438| [CX3]=([OX1])O | 0.986|
| [CX3H0]=([OX1H0])([OX2H1])([OX4H1]) | 0.6319| [CX3]=([OX1])O | 0.986|
| [#6H0]=[#6]=[#6] | 0.5364| [CX3]=([OX1])O | 0.986|
| [CX4H2]=CC=O    | 0.4105| [CX3]=([OX1])O | 0.986|
| [CX3]=([OX1])O | 0.3886| [CX3]=([OX1])O | 0.986|
| O=#6[#6]=#6H3  | 0.3837| [CX3]=([OX1])O | 0.986|
| [#8]=[#6][#6]=#8 | 0.3575| [CX3]=([OX1])O | 0.986|
| [CX4H2]=[#6][#6] | 0.2431| [CX3]=([OX1])O | 0.986|
| CCCCCCCCC      | 0.1383| [CX3]=([OX1])O | 0.986|
Example 106 true smiles: CC(=O)CCC(=O)O formula: C₅H₈O₃
Index of correct structure: 0 of 1131
True structure loss: 0.004847
True structure:

Experimental 13C NMR (solvent: CDCl₃)

Experimental 1H NMR (solvent: D₂O)

Top predicted structures (loss):

- 0.004847
- 0.080461
- 0.083891
- 0.083916
- 0.08424
- 0.091077
- 0.098337
- 0.105358
- 0.106729
- 0.111187
### Top predicted substructures

| Substructure | prob | Substructure | prob |
|--------------|------|--------------|------|
| [CX3](=[OX1])C | 1.0 | [CX4H3](=[CX3H0]) | 0.9941 |
| [CX4H2][[6]][#6] | 0.9995 | O=[CX3H0][CX4H2][CX4H2] | 0.993 |
| [CX4H3][CX3] | 0.9983 | [CX4H2]([CX4H2])[CX3H0] | 0.9922 |
| [OX1H0][CX3H0][CX4H3] | 0.997 | [#8]=[#6][#8] | 0.9914 |
| [#6H3][#6H0] | 0.995 | [CX4H2][CX3]=O | 0.9872 |

### best positives

| Substructure | prob | Substructure | prob |
|--------------|------|--------------|------|
| [CX3](=[OX1])C | 1.0 | C=CC=CC#C | 0.0 |
| [CX4H2][[6]][#6] | 0.9995 | CCC#CC#C | 0.0 |
| [OX1H0][CX3H0][CX4H3] | 0.997 | [#6X2][#6H1][#6X2] | 0.0 |
| [#6][#6] | 0.9983 | [CX2H0][#CX2H1][CX3H0] | 0.0 |
| [CX4H2][CX3H0][CX4H3] | 0.9941 | [CX2H0][#CX2H1][CX3H0] | 0.0 |
| [OX1H0][CX3H0][CX4H3] | 0.9941 | [CX2H0][#CX2H1][CX3H0] | 0.0 |
| [#8]=[#6][#8] | 0.9914 | [CX2H0][#CX2H1][CX3H0] | 0.0 |
| [CX4H2][CX3]=O | 0.9872 | CCC=CC#C | 0.0 |

### worst negatives

| Substructure | prob | Substructure | prob |
|--------------|------|--------------|------|
| CCC=CCC | 0.2577 | [#8][#6H0][#6H1] | 0.0775 |
| [CX4H2][CX3H0][CX4H3] | 0.2181 | [OX1H0][OX2H0][CX4H2] | 0.1606 |
| [CX3H0][[OX1H0]][[OX2H0]][CX4H2] | 0.1606 | OCC[CH2] | 0.8379 |
| [#8][#6H0][#6H1] | 0.1477 | [#6X3][#6X3H0][#6H2] | 0.8718 |
| O=[CX3H0][CX4H2][CX3H0][CX4H2] | 0.1172 | [CX4H3][#6] | 0.8552 |
| [CX4H2][CX3H0][CX4H3] | 0.0851 | [#6X3][#6X3H0][#6H2][#6X3H0] | 0.9004 |
| [6X3][#6X3] | 0.0825 | [OX1H0][[CX4H3]][CX4H2] | 0.9064 |
| [#8][#6][#6][#8] | 0.0791 | [OX1H0][[CX3H0]][#8][#8] | 0.909 |
| [OX2H0][#6][#6] | 0.0775 | [CX4H2][#CX3]=O | 0.921 |
| [OX1H0][[OX2H0]][CX3H0] | 0.0613 | [OX2H1] | 0.9392 |

---
Example 107 true smiles: O=C(=O)CC(O)=O formula: C4H6O5
Index of correct structure: 0 of 1119
True structure loss: 0.029603
True structure:
| Substructure                                             | Prob | Substructure                                             | Prob |
|----------------------------------------------------------|------|----------------------------------------------------------|------|
| [CX3]([=][OX1])C                                        | 0.9995 | [#8][#6][#6H2]                                          | 0.9694 |
| [#8][=][#6][#8]                                         | 0.9988 | OCC[CH2]                                                | 0.9374 |
| [OX2H1]                                                 | 0.9982 | [#6H1]                                                 | 0.8988 |
| [CX3]([=][OX1])O                                        | 0.9935 | [#6H1][#6H2]                                            | 0.7982 |
| [OX2H1][=][#6][#6]                                      | 0.9911 | [CX4H2][CX3]=O                                           | 0.7758 |

**Best Positives**

| Substructure                                             | Prob | Substructure                                             | Prob |
|----------------------------------------------------------|------|----------------------------------------------------------|------|
| [CX3]([=][OX1])C                                        | 0.9995 | [#6H3][#6X3][#6X3]=[#6X3H2]                               | 0.0  |
| [#8][=][#6][#8]                                         | 0.9988 | [CX2H1][=][CX2H0][CX3H1]=[#CX3H0]                         | 0.0  |
| [OX2H1]                                                 | 0.9982 | CC#CCC=C                                                | 0.0  |
| [CX4H2][=][#6][#6]                                      | 0.9911 | [CX3H0]=([CX3H2])([CX4H3])[CX4H2]                        | 0.0  |
| [#8][#6][#6H2]                                          | 0.9694 | [CX3H0]=([CX3H1])([CX4H3])[CX3H1]                        | 0.0  |
| OCC[CH2]                                                | 0.9374 | CCC=CC#C                                                | 0.0  |
| [#6H1]                                                 | 0.8988 | [#6H3][#6H1][#6H1]=[#7]                                  | 0.0  |
| [#6H1][#6H2]                                            | 0.7982 | [CX3H1]=([CX3H1][CX2H0]                                  | 0.0  |
| [CX4H2][=][CX3]=O                                       | 0.7758 | [CX3H0]=([CX3H0])([CX4H3])[CX4H2]                        | 0.0  |

**Best Negatives**

| Substructure                                             | Prob | Substructure                                             | Prob |
|----------------------------------------------------------|------|----------------------------------------------------------|------|
| [CX4H2][=][#6][#6]                                      | 0.3773 | [#8]=[#6][#6][#6]=[#6]                                   | 0.0659 |
| [#6H1][#6H1]                                            | 0.3479 | [CX4H1][=][OX2H1][CX4H2][CX3H0]                          | 0.0977 |
| [#6H1][#6H2]=[#6H2]                                      | 0.2463 | [OX1H0]=([CX3H0][CX4H1][=][OX2H1])[CX4H2]               | 0.1663 |
| [CX4H2][=][OX1]                                         | 0.1937 | [CX3][=][CX4H]                                           | 0.1757 |
| [#8][#6H1][#6H1]                                       | 0.1826 | [#6H1]=[#6X4H][#6X3H0]                                  | 0.2003 |
| [CX4H1][=][OX2H1][=][CX4H2][=][CX4H1]                  | 0.1557 | [#8]=[#6H0][#6H1]                                        | 0.2463 |
| [#7][#6][#6X3]                                          | 0.1468 | [#8][#6][#6]=[#8]                                        | 0.2987 |
| [CX4H2][=][CX4H1][=][CX4H0]                            | 0.1286 | [CX4H2][=][CX4H][=]                                      | 0.4384 |
| [OX1H0][=][CX3H0][=][CX4H2]                            | 0.1186 | [OX1H0][=][OX1H0][CX4H][=]                               | 0.4442 |

**Worst Positives**

| Substructure                                             | Prob | Substructure                                             | Prob |
|----------------------------------------------------------|------|----------------------------------------------------------|------|
| [CX4H2][=][#6][#6]                                      | 0.1107 | [#8]=[#6][#6H2][#6H1]                                    | 0.4543 |
Example 108 true smiles: NCc1ccco1 formula: C5H7NO
Index of correct structure: 0 of 1024
True structure loss: 0.028721
True structure:

![Chemical structure]

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

- ![Structure 1](image1.png) 0.028721
- ![Structure 2](image2.png) 0.043498
- ![Structure 3](image3.png) 0.052437
- ![Structure 4](image4.png) 0.059929
- ![Structure 5](image5.png) 0.060608
- ![Structure 6](image6.png) 0.062206
- ![Structure 7](image7.png) 0.065483
- ![Structure 8](image8.png) 0.066442
- ![Structure 9](image9.png) 0.067302
- ![Structure 10](image10.png) 0.068634
Top predicted substructures

Best positives

Worst negatives
**Example 109** true smiles: OCC1C(CO1)O

**formula:** C₆H₁₂O₂

**Index of correct structure:** 0 of 903

**True structure loss:** 0.016167

**True structure:**

![Chemical structure diagram]

**Experimental ¹³C NMR (solvent: CDCl₃)**

![Experimental ¹³C NMR spectrum]

**Experimental ¹H NMR (solvent: CDCl₃)**

![Experimental ¹H NMR spectrum]

**Top predicted structures (loss):**

1. ![Structure 1](image1) - 0.016167
2. ![Structure 2](image2) - 0.019617
3. ![Structure 3](image3) - 0.02226
4. ![Structure 4](image4) - 0.032337
5. ![Structure 5](image5) - 0.034418
6. ![Structure 6](image6) - 0.040589
7. ![Structure 7](image7) - 0.041438
8. ![Structure 8](image8) - 0.041947
9. ![Structure 9](image9) - 0.042096
10. ![Structure 10](image10) - 0.046171
Top predicted substructures

| Substructure                                                                 | Prob  |
|------------------------------------------------------------------------------|-------|
| [CX4H2][(#6)][#6]                                                           | 1.0   |
| OCC[CH2]                                                                    | 0.9997|
| [#8][#6][#6H2]                                                               | 0.999 |
| [CX4H2][CX4H2][CX4H1]                                                       | 0.997 |
| [CX4H2][#6][O]                                                               | 0.9963|

best positives

| Substructure                                                                 | Prob  |
|------------------------------------------------------------------------------|-------|
| [CX4H2][(#6)][#6]                                                           | 1.0   |
| OCC[CH2]                                                                    | 0.9997|
| [#8][#6][#6H2]                                                               | 0.999 |
| [CX4H2][CX4H2][CX4H1]                                                       | 0.997 |
| [CX4H2][#6][O]                                                               | 0.9963|

worst negatives

| Substructure                                                                 | Prob  |
|------------------------------------------------------------------------------|-------|
| [#6X4H2][#6H1][#8]                                                           | 0.5738|
| [#8][#6][#6][#6][#6][#6]                                                     | 0.5118|
| [#8][#6H1][#6H1]                                                             | 0.42  |
| O[CX4H1][CX4H2][CX4H1]                                                       | 0.4182|
| [OH][CX4H]                                                                  | 0.361 |
| [CX4H1][OX2H1][CX4H2][CX4H1]                                                 | 0.2831|
| [#6H1][#6H1]                                                                 | 0.2603|
| [CX4H1][CX4H2][CX4H2][CX4H1]                                                 | 0.1723|
| [CX4H2][OX2H0][CX4H2]                                                        | 0.1527|
| [#6][#6][#6][#6][#6][#6][#6][#6]                                             | 0.1491|
Example 110 true smiles: CCCCCC(=O)O formula: C6H12O2
Index of correct structure: 0 of 903
True structure loss: 0.004445
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

- 0.004445
- 0.042294
- 0.056974
- 0.05836
- 0.061383
- 0.062021
- 0.0807
- 0.085363
- 0.092392
- 0.095194
| Top predicted substructures | prob | | prob |
|----------------------------|------||----------------------------|------|
| [CX4H2][{#6}]|{#6} | 1.0 | CX4H3|[#6] | 0.9981 |
| [CX4H2][{#6}]|{#6} | 0.9998 | CX4H3 | | 0.9971 |
| [CX4H3][CX4H2] | 0.9993 | | | | 0.9947 |
| [CX3]=[O][OX2H1] | 0.9991 | | | | 0.9835 |
| [CX3]=[O][OX1]C | 0.9999 | | | | 0.9755 |
| [CX4H2][{#6}]|{#6} | 0.9998 | | | 0.90 |
| [CX4H3][CX4H2] | 0.9993 | | | | 0.8214 |
| [CX3]=[O][OX2H1] | 0.9991 | | | | 0.8511 |
| [CX3]=[O][OX1]C | 0.9999 | | | | 0.8614 |
| [CX4H3][{#6}][#8] | 0.9991 | | | | 0.8714 |
| [CX4H3] | 0.9971 | | | | 0.8788 |
| [CX4H3] | 0.9971 | | | | 0.8848 |
| [CX4H2][{#6}][#6H2] | 0.9755 | | | | 0.9015 |
| worst negatives | prob | | prob |
|----------------------------|------||----------------------------|------|
| [#6H1][{#6}][#6H2] | 0.3048 | | | | 0.6494 |
| [CX3H0]=[O][OX1H0][{OX2H1}][CX4H1] | 0.2195 | | | | 0.688 |
| [#6H1] | 0.1738 | | | | 0.8214 |
| [#6H1][{#6}][#6H2] | 0.1248 | | | | 0.8511 |
| [#8]=[#6H0][{#6}][#6H1] | 0.0998 | | | | 0.8614 |
| [CH4][{CH3X4}][CH2X4] | 0.0806 | | | | 0.8741 |
| [CH3]=O[OK] | 0.0799 | | | | 0.8788 |
| [#6H1][{#6H2}][{#6H2}] | 0.071 | | | | 0.8848 |
| [#6H3][{#6}][#6X3] | 0.0699 | | | | 0.8945 |
| | | | | | 0.9015 |
Example 111 true smiles: CCO(=O)C(C)C formula: C6H12O2

Index of correct structure: 0 of 903
True structure loss: 0.022851
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

0.022851
0.044646
0.047757
0.050781
0.053552
0.05806
0.058515
0.058806
0.058854
0.059992
Top predicted substructures          prob
[CX4H3]                                1.0     [CX3]=[OX1]O                                  0.9329
[#6H3][#6][#6]                         0.9999  [CX4H2][{#6}][O]                              0.9303
[CX4H3][#6]                            0.999   [CX4H3][CX4H1]                               0.8783
[CX3]=[OX1]C                           0.9873  [#6H3][#6][#6X3]                             0.8551
[#8]=[#6][#8]                          0.9925  [CX4H3][CX4H2]                               0.8471

best positives                       prob     best negatives                      prob
[CX4H3]                                1.0     C=CC=CC#C                                    0.0
[^6H3][^6][^6]                         0.9999  [CX2H0][^6][CX2H1][^6H0]                   0.0
[CX4H3][#6]                            0.999   [CX3H0][^6][CX3H1][{OX2H0}][CX2H0]     0.0
[CX1]=[^6][^6]                         0.9873  CC#CCC=C                                    0.0
[#8]=[^6][^6]                          0.9825  [CX2H0][^6][CX2H1][^6H0]                   0.0
[CX1]=[^6][^6]                         0.9329  [CX3H0][^6][CX3H1][{CX4H2}][CX2H0]     0.0
[CX4H2][{#6}][O]                      0.9303  CCC=CCC=C                                   0.0
[CX4H3][CX4H1]                        0.8783  [#6X2][#6H1][#6X2]                         0.0
[^6H3][^6][^6X3]                      0.8551  C=C=C=C                                       0.0
[CX4H3][CX4H2]                        0.8471  CCC=CCC=C                                   0.0

worst negatives                      prob     worst positives                      prob
[OX2H1]                                0.7563  [CX3H0][{OX1H0}][{OX2H0}][CX4H1]     0.1421
[^6H3][^6][^6][^6H3]                   0.6401  [^8]=[^6H0][^6H1]                         0.1627
[CX1]=[^6][OX2H1]                      0.614   [CHX4][(CH3X4)](CH3X4)                  0.2393
[^6H3][^6H0]                           0.5914  [CX4H1][(CX4H3)][{CX4H3}][CX3H0]    0.3492
[CX4H3][CX4H0]                        0.5165  O=[CX3][CX4H]                              0.3638
[OX1H0]=[CX3H0][{^8}][CX4H2]           0.4751  [CH3][^6][^8]                             0.5874
[CX4H2][CX3]=^0                        0.3545  [#3][^6H0][^6H1]                        0.6496
[^6H3][^6H0][^6H3]                     0.3298  [CX4H3][CX4]O                             0.6655
[OX2H1][CX4H2][^6X3H0]                 0.319   [#6H1]                                     0.7254
[CX3H0][=^6X1H0][{OX2H1}][CX4H1]       0.2715  [CX4H2][{OX2H0}][CX4H3]                 0.7587

------------------------------------------------------------------------------------------------------------------------
Example 112 true smiles: CC(C)CCC(=O)O formula: C6H12O2
Index of correct structure: 0 of 903
True structure loss: 0.006743
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

0.006743  0.033622  0.0486  0.066608  0.069353
0.072996  0.076831  0.087612  0.094259  0.103345
| Top predicted substructures | prob | best positives | prob | best negatives | prob |
|-----------------------------|------|----------------|------|----------------|------|
| [#6H3][#6][#6]              | 1.0  | [#8]=[#6][#8]  | 0.9914 |                 |      |
| [CX4H2][#6][#6]             | 0.9991 | [CX4H3][#6]   | 0.9866 |                 |      |
| [CX3](=O)[O2H1]             | 0.999 | [CX3]=[O2H1]O | 0.985 |                 |      |
| [CX4H3]                     | 0.9986 | [O2H1]         | 0.9754 |                 |      |
| [CX3](=[O2H1])C             | 0.9974 | [CH4][([CH3]4)[CH3]4] | 0.9549 |      |

| worst negatives | prob | worst positives | prob |
|----------------|------|----------------|------|
| [CX4H2][([CX4H2])][CX4H2] | 0.404 | [CX4H2][([CX4H2])][CX4H2] | 0.5845 |
| [CX3H0]=[O1H0][O2H1][CX4H1] | 0.3546 | [#8][#6][#6H2] | 0.7081 |
| [CX4H2][CX4H2][CX4H2][CX4H2] | 0.2974 | OCC[CH2] | 0.7314 |
| [O2H1] | 0.277 | [CX4H2][CX4H2] | 0.7585 |
| [#6H3][#6][#6H3] | 0.1705 | [#6H1][#6H2] | 0.7799 |
| [#6][#6][#6H3] | 0.1614 | [CX4H2][CC] | 0.8074 |
| [#6][#6][#6H3] | 0.1584 | [CX4H2][CX3]=O | 0.8337 |
| [CX4H2][([CX4H1])][CX3H0] | 0.1363 | [CH4][([CH3]4)[CH2]4] | 0.8760 |
| [#8][#6H0][#6H1] | 0.0971 | [#6H1] | 0.8814 |
| [#8][#6H0][#6H1] | 0.0934 | [CX4H2][([CX4H2])][CX3H0] | 0.8892 |
Example 113 true smiles: C=COCCCCO formula: C6H12O2
Index of correct structure: 0 of 903
True structure loss: 0.006823
True structure:

![Chemical structure diagram]

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

1. ![Predicted structure 1] 0.006823
2. ![Predicted structure 2] 0.029637
3. ![Predicted structure 3] 0.046928
4. ![Predicted structure 4] 0.063585
5. ![Predicted structure 5] 0.066065

6. ![Predicted structure 6] 0.068251
7. ![Predicted structure 7] 0.072528
8. ![Predicted structure 8] 0.076116
9. ![Predicted structure 9] 0.078152
10. ![Predicted structure 10] 0.080675
| Substructure                                      | Probability | Substructure                                      | Probability |
|--------------------------------------------------|-------------|--------------------------------------------------|-------------|
| CX3H1[^CX3H2][OX2H0]                            | 0.9999      | [CX4H2]{{#6}}{O}                                 | 0.9969      |
| [#8][#6]=[#6H2]                                 | 0.9998      | OCC{CH2}                                         | 0.9961      |
| [CH2X3]=cC                                      | 0.9988      | [CH2X4]{{O}|CX4H2}|CX4H2               | 0.9938      |
| CX4H2[^#6][^6]                                 | 0.9983      | [CX3H2]=^CX3H1                                   | 0.9918      |
| best positives                                   |             | best negatives                                   |             |
| CX3H1[^CX3H2][OX2H0]                            | 0.9999      | CX2H0[^#CX2H1]^CX3H0                             | 0.0         |
| [#8][#6]=[#6H2]                                 | 0.9998      | [#6X2][#6H1][#6X2]                               | 0.0         |
| CH2X3/=C                                        | 0.9988      | [#7][#6][#6][#6][#7]                             | 0.0         |
| CX4H2[^#6][^6]                                 | 0.9983      | CX2H0[^#CX2H0][CX2H0]                            | 0.0         |
| CX4H2[^#6][O]                                  | 0.9969      | C=C=C=cC=C=C                                     | 0.0         |
| OCC{CH2}                                        | 0.9961      | CX2H1[^CX2H0]|CX3H1]=^CX3H0                             | 0.0         |
| [6X3H2]                                         | 0.9949      | cC#CCC#C                                         | 0.0         |
| CH2X4[^O]|CX4H2|^CX4H2                               | 0.9938      | CX2H0[^#CX2H1]^CX4H2                             | 0.0         |
| CX3H2]=^CX3H1                                   | 0.9918      | CX2H0[^#CX2H0]^CX4H0                             | 0.0         |
| worst negatives                                  |             | worst positives                                   |             |
| CX4H2[^CX4H2]|CX4H1                   | 0.4017      | [#8][#6][#6][#6][#6][#8]                         | 0.4203      |
| [#8][#6][#6H2]                                 | 0.3617      | CX4B2[^OX2H1]|CX4B2                               | 0.5112      |
| [#8][#6][#6H1]                                 | 0.3321      | [#6H2][#8][#6H1]                                 | 0.6835      |
| [#8][#6][#6][#6]                               | 0.2209      | CX4H2[^OX2H0]|CX4H2                             | 0.7659      |
| CX4H2[^CX4H2]|CX3H1                   | 0.2048      | CX4H2[^CX4H2]|CX4H2|[CX4H2][CX4H2]                  | 0.9231      |
| O|CX4H1|^CX4H2                               | 0.2016      | CX3H0|^CX4H2                                  | 0.9254      |
| CH3][=cC][C                                      | 0.1509      | [OX2H1]|                                     | 0.9329      |
| CX4H2[^CX3H]                                   | 0.1259      | [#6H1]|                                     | 0.9364      |
| CX4H1[^O]                                      | 0.1192      | [CX3H][O]                                     | 0.9484      |
| [#8][#6][#6H2][#8]                             | 0.1084      | [CH2X4][^O]|CX4H2]                           | 0.9484      |
Example true smiles: CC(=O)OC(C)(C)C
formula: C₆H₁₂O₂
Index of correct structure: 0 of 903
True structure loss: 0.007652
True structure:

Experimental 13C NMR (solvent: CDCl₃)

Experimental 1H NMR (solvent: CDCl₃)

Top predicted structures (loss):

0.007652 0.036201 0.041301 0.061614 0.064295
0.064928 0.07172 0.075579 0.076905 0.080629
| Top predicted substructures | prob | prob |
|-----------------------------|------|------|
| [#6H3][#6H0]               | 0.9996 | [CX4H3][#6]  | 0.9958 |
| [CX4H3]                    | 0.9993 | [CX4H3][CX4H0][CX4H3] | 0.9946 |
| [CX4H3][CX4H0]             | 0.9987 | [CX4H3][CX3]  | 0.9811 |
| [#6H3][#6][#6]             | 0.9972 | [CX4H3][CX4H0][{CX4H3}][OX2H0] | 0.9765 |
| [CX4H3][CX4H0][OX2H0]      | 0.9967 | [#8]=[#6][#8] | 0.9605 |

| best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| [#6H3][#6H0]   | 0.9996 | [#7][#6][#6][#6][#7] | 0.0 |
| [CX4H3]        | 0.9993 | CCC#CC#C   | 0.0 |
| [CX4H3][CX4H0] | 0.9987 | [#7][#6][#6][#6][#7] | 0.0 |
| [CX4H3][CX4H0] | 0.9972 | [CX2H0][#C][CX2H0][CX2H0] | 0.0 |
| [CX4H3][CX4H0][#6] | 0.9967 | [CX2H0][#C][CX2H1][CX2H0] | 0.0 |
| [CX4H3][#6]    | 0.9958 | [CX4H1][{OX2H0}][{CX4H1}][OX2H0] | 0.0 |
| [CX4H3][CX4H0][{CX4H3}][OX2H0] | 0.9946 | [CX3H1][{OX2H0}][{OX2H0}] | 0.0 |
| [CX4H3][CX3]   | 0.9911 | [#7][#6H1][#6X2] | 0.0 |
| [CX4H3][{CX4H3}][OX2H0] | 0.9765 | [CX2H0][#C][CX2H1][{CX3H0}] | 0.0 |
| [#8]=[#6][#8]  | 0.9606 | [#6X2][#6H1][#6X2] | 0.0 |

| worst negatives | prob | worst positives | prob |
|------------------|------|----------------|------|
| [#6H0][#6H2]    | 0.5603 | [OX1H0]=[OX1H0][OX1H0][{CX4H3}] | 0.7527 |
| OCC[CH2]         | 0.5255 | [OX1H0][{OX1H0}][{OX2H0}][{CX4H3}] | 0.8171 |
| [OX2H1]          | 0.4812 | [OX1H0][{OX1H0}][{OX2H0}][{CX4H3}] | 0.877 |
| [#6H3][#6][#6H] | 0.2951 | [OX2H0][{OX1H0}][{OX2H0}][{CX4H3}] | 0.9063 |
| [#8][#6]=[#8]   | 0.2695 | [CH3][#6][#8] | 0.9194 |
| [CX4H2][{#6}][#6] | 0.2636 | [CX3]=[OX1]C | 0.9364 |
| [CX4H2][OX2H2]  | 0.2233 | [CX4H3][CX3H0] | 0.9375 |
| [CX3][#O][OX2H1] | 0.2005 | [CX3]=[OX1]O | 0.9389 |
| [#8][#6]=[#6X3] | 0.1911 | [#6H0][{#6H3}][{#6H3}][#8] | 0.9579 |
| [#8]=[#6][#8]   | 0.176 | [#8]=[#6][#8] | 0.9606 |
Example 115 true smiles: CC(C)CCC(=O)O formula: C6H12O2
Index of correct structure: 0 of 903
True structure loss: 0.007335
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

1. [Structure Image] 0.007335
2. [Structure Image] 0.034777
3. [Structure Image] 0.044111
4. [Structure Image] 0.063887
5. [Structure Image] 0.066391
6. [Structure Image] 0.075492
7. [Structure Image] 0.078784
8. [Structure Image] 0.081609
9. [Structure Image] 0.089869
10. [Structure Image] 0.099078
### Top predicted substructures

| Substructure | Probability |
|--------------|-------------|
| #6H3][6][6] | 1.0         |
| [CX3][=O][OX2H1] | 0.9996     |
| [CX4H3] | 0.9992         |
| [CX4H2][][6][6] | 0.9981     |
| [CX3][=O][O][CX1] | 0.9976     |

### Best positives

| Substructure | Probability |
|--------------|-------------|
| #6H3][6][6] | 1.0         |
| [CX3][=O][OX2H1] | 0.9996     |
| [CX4H3] | 0.9992         |
| [CX4H2][][6][6] | 0.9981     |
| [CX3][=O][O][CX1] | 0.9976     |

### Best negatives

| Substructure | Probability |
|--------------|-------------|
| #6H3][6][6] | 0.0         |
| [CX2H0][=C][CX2H1][CX3H0] | 0.0     |
| [CX4H2][][6][6] | 0.0     |
| [CX3][=O][O][CX1] | 0.0     |
| [CX3][=O][O][CX2H0] | 0.0     |
| [CX3][=O][O][CX4H2] | 0.0     |

### Worst positives

| Substructure | Probability |
|--------------|-------------|
| [CX3H0][=O][CX2H1][CX4H1] | 0.4638     |
| [CX4H2][=O][CX4H2][CX3H0][CX4H1] | 0.7047     |
| [CX4H2][=C][CX4H2][CX4H2] | 0.2871     |
| [CX4H2][=C][CX4H2][CX4H2] | 0.2199     |
| [CX4H2][=C][CX4H2][CX4H2] | 0.2009     |
| [CX4H2][=C][CX4H2][CX4H2] | 0.1997     |

### Worst negatives

| Substructure | Probability |
|--------------|-------------|
| [CX3H0][=O][CX2H1][CX4H1] | 0.7015     |
| [CX4H2][=C][CX4H2][CX4H2] | 0.7051     |
| [CX4H2][=C][CX4H2][CX4H2] | 0.7062     |
| [CX4H2][=C][CX4H2][CX4H2] | 0.7485     |
| [CX4H2][=C][CX4H2][CX4H2] | 0.7533     |
| [CX4H2][=C][CX4H2][CX4H2] | 0.7896     |
| [CX4H2][=C][CX4H2][CX4H2] | 0.8125     |
| [CX4H2][=C][CX4H2][CX4H2] | 0.8385     |
| [CX4H2][=C][CX4H2][CX4H2] | 0.9088     |
| [CX4H2][=C][CX4H2][CX4H2] | 0.9167     |
Example 116 true smiles: OC1CCC(O)CC1 formula: C6H12O2
Index of correct structure: 0 of 903
True structure loss: 0.012793
True structure:

Experimental 13C NMR (solvent: DMSO)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

![Structures with corresponding losses](image-url)
| Top predicted substructures                  | prob   | best positives                  | prob   | worst negatives                  | prob   |
|---------------------------------------------|--------|---------------------------------|--------|----------------------------------|--------|
| [CX4H2][([#6])[#6]]                       | 1.0    | [CX4H2][([CX4H2])][CX4H1]     | 0.9566 | [CX4H2][([CX4H2])][CX4H2]     | 0.6622 |
| OCC[C1H2]                                  | 0.9986 | [#6H1][#6H2]                   | 0.8993 | [#6H0][#6H1][#6H2]               | 0.411  |
| [#8][#6][#6H2]                             | 0.988  | [#6H1]                          | 0.853  | [#8][#6H0][#6H1][#6H2]           | 0.478  |
| [CX4H2][CX4H2]                             | 0.9741 | [#6X4H2][#6H1][#8H]            | 0.852  | [CX4H1][CX4H2][CX4H2]            | 0.7806 |
| best positives                              |        | best negatives                  |        | worst positives                  |        |
| [CX4H2][([#6])[#6]]                       | 1.0    | [CX3H0][=([CX3H1])][CX2H0]    | 0.0    | [CX4H0][=([CX4H1])][CX4H1]     | 0.2112 |
| OCC[C1H2]                                  | 0.9986 | CC#C=C                         | 0.0    | [CX4H0][#6H2][#6H2][#6H1]       | 0.2112 |
| [#8][#6][#6H2]                             | 0.988  | C=C=C#C                         | 0.0    | [CX4H0][#6H0][#6H1][#6H2]       | 0.478  |
| [CX4H2][CX4H2]                             | 0.9741 | [CX4H0][=([CX3H1])][=([CX3H0])| 0.0    | [CX4H0][#6H0][#6H1][#6H2]       | 0.478  |
| OCC[C1H2]                                  | 0.9986 | [CX2H1][#6H2][#6H1][=([CX3H0])| 0.0    | [CX4H0][#6H1][#6H2][#6H1]       | 0.2112 |
| [#6H1][#6H2]                               | 0.8993 | [CX4H1][=([CX3H1])][=([CX3H0])| 0.0    | [CX4H0][#6H1][#6H2][#6H1]       | 0.2112 |
| worst negatives                             |        | worst positives                 |        | worst positives                  |        |
| [CX4H2][([CX4H2])][CX4H2]                 | 0.6622 | [#6H1][#6H2][#6H2][#6H1]       | 0.2112 | [CX4H0][#6H1][#6H2][#6H1]       | 0.2112 |
| [#8][#6H0][#6H1]                           | 0.411  | [CX4H1][=([CX4H1])][=([CX4H0])| 0.478  | [CX4H0][#6H1][#6H2][#6H1]       | 0.2112 |
| [CX4H2][([CX4H2])][CX4H0]                 | 0.3947 | [OH][CX4H]                      | 0.5191 | [CX4H0][#6H1][#6H2][#6H1]       | 0.2112 |
| [CX4H2][([CX4H2])][CX4H2][CX4H2]          | 0.3146 | [#8][#6][#6][#6][#6][#6][#6]   | 0.52   | [CX4H0][#6H1][#6H2][#6H1]       | 0.2112 |
| [CX4H1][([CX4H2])][([CX4H2])][CX4H2]     | 0.2233 | [CX4H0][=([CX3H1])][=([CX3H0])| 0.5359 | [CX4H0][#6H1][#6H2][#6H1]       | 0.2112 |
| [CX4H1][([CX4H2])][([CX4H2])][CX4H2]     | 0.2213 | [#6H1][([#6H2)][#6H2]          | 0.6636 | [CX4H0][#6H1][#6H2][#6H1]       | 0.2112 |
| [#8][#6][#6][#6]                           | 0.2099 | [#8][#6][#6][#6][#6][#6][#6][#6] | 0.7286 | [CX4H0][#6H1][#6H2][#6H1]       | 0.2112 |
| [CX4H1][([CX4H2])][([CX4H2])][CX4H1]     | 0.2015 | O=CX4H1][=([CX3H1])][=([CX3H0])| 0.7783 | [CX4H0][#6H1][#6H2][#6H1]       | 0.2112 |
| [CX4H1][([CX4H2])][CX4H0]                 | 0.1937 | [CX4H1][([CX2H1])][([CX4H2])][| 0.7806 | [CX4H0][#6H1][#6H2][#6H1]       | 0.2112 |
| [CX4H2][([CX4H1])][CX4H1]                 | 0.1927 | [CX4H2][=([CX4H2])][CX4H2]     | 0.852  | [CX4H0][#6H1][#6H2][#6H1]       | 0.2112 |
Example 117 true smiles: COC(OC)N(C)C formula: C5H13NO2

Index of correct structure: 0 of 900
True structure loss: 0.022358
True structure:

Experimental 13C NMR (solvent: CCl4)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

0.022358 0.022391 0.029448 0.034865 0.034961
0.035292 0.036968 0.037916 0.040704 0.042838
Top predicted substructures

| Structure                        | prob  |
|----------------------------------|-------|
| [CX4H3]                          | 0.9868|
| [#7X3][#6H2]                    | 0.8904|
| [#7][#6H2]                      | 0.8376|
| [#8][#6][#6H2]                  | 0.8376|
| [OX2H0][CX4H1][OX2H0]           | 0.8311|
| [#7X3][#6H3]                    | 0.7403|
| [#6H3][#7]                      | 0.6847|
| [#6H2][#7]                      | 0.6589|
| [CX4H2][#6][#6H0]               | 0.5379|
| [#6H2][#6H3]                    | 0.5316|
| [OX2H0][OX2H0][OX2H0]           | 0.4523|
| [CX4H3]                          | 0.3649|
| [OX2H0][OX2H0][OX2H0]           | 0.3619|

Best positives

| Structure                        | prob  |
|----------------------------------|-------|
| [CX4H3]                          | 0.9868|
| [CX4H3][CX4H1][OX2H0]            | 0.8311|
| [#7X3][#6H3]                    | 0.7403|
| [CX4H3][OX2H0]                   | 0.6847|
| [#6H3][#7]                      | 0.6589|
| [CX4H2][OX2H0][OX2H0]           | 0.6589|
| [#6H2][#7]                      | 0.6589|
| [CX4H3][NX3H0]                   | 0.342 |
| [OX2H0][OX2H0][OX2H0]           | 0.3619|
| [CX4H3][#6H1]                   | 0.3619|
| [CX4H3][NX3H0]                   | 0.342 |
| [OX2H0][OX2H0][OX2H0]           | 0.3619|

Best negatives

| Structure                        | prob  |
|----------------------------------|-------|
| [CX4H3]                          | 0.9868|
| [CX4H3][CX4H1][OX2H0]            | 0.8311|
| [#7X3][#6H3]                    | 0.7403|
| [CX4H3][OX2H0]                   | 0.6847|
| [#6H3][#7]                      | 0.6589|
| [CX4H2][OX2H0][OX2H0]           | 0.6589|
| [#6H2][#7]                      | 0.6589|
| [CX4H3][NX3H0]                   | 0.342 |
| [OX2H0][OX2H0][OX2H0]           | 0.3619|
| [CX4H3][#6H1]                   | 0.3619|
| [CX4H3][NX3H0]                   | 0.342 |
| [OX2H0][OX2H0][OX2H0]           | 0.3619|

Worst positives

| Structure                        | prob  |
|----------------------------------|-------|
| [CX4H3]                          | 0.9868|
| [CX4H3][CX4H1][OX2H0]            | 0.8311|
| [#7X3][#6H3]                    | 0.7403|
| [CX4H3][OX2H0]                   | 0.6847|
| [#6H3][#7]                      | 0.6589|
| [CX4H2][OX2H0][OX2H0]           | 0.6589|
| [#6H2][#7]                      | 0.6589|
| [CX4H3][NX3H0]                   | 0.342 |
| [OX2H0][OX2H0][OX2H0]           | 0.3619|
| [CX4H3][#6H1]                   | 0.3619|
| [CX4H3][NX3H0]                   | 0.342 |
| [OX2H0][OX2H0][OX2H0]           | 0.3619|

Worst negatives

| Structure                        | prob  |
|----------------------------------|-------|
| [CX4H3]                          | 0.9868|
| [CX4H3][CX4H1][OX2H0]            | 0.8311|
| [#7X3][#6H3]                    | 0.7403|
| [CX4H3][OX2H0]                   | 0.6847|
| [#6H3][#7]                      | 0.6589|
| [CX4H2][OX2H0][OX2H0]           | 0.6589|
| [#6H2][#7]                      | 0.6589|
| [CX4H3][NX3H0]                   | 0.342 |
| [OX2H0][OX2H0][OX2H0]           | 0.3619|
| [CX4H3][#6H1]                   | 0.3619|
| [CX4H3][NX3H0]                   | 0.342 |
| [OX2H0][OX2H0][OX2H0]           | 0.3619|
Example 118 true smiles: CN(C)CC(=O)O formula: C4H9NO2
Index of correct structure: 0 of 896
True structure loss: 0.018376
True structure:

![Chemical structure image](image)

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

- ![Substructure A](image) with prob 0.018376
- ![Substructure B](image) with prob 0.024773
- ![Substructure C](image) with prob 0.030801
- ![Substructure D](image) with prob 0.040296
- ![Substructure E](image) with prob 0.043706

Top predicted substructures

- [CX4H3] with prob 0.9839
- [CX4H3][NX3H0] with prob 0.8289
| best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| [CX3](=[OX1])C | 0.9517 | [CX2H0](#[OX1])O | 0.7896 |
| [#6H3][#7]     | 0.9203 | [OX2H1][#8] | 0.7793 |
| [#8]=[#6][#8]  | 0.8577 | [#7][#6H2] | 0.7557 |
| [CX4H3]        | 0.9839 | C=CCCC#C | 0.0 |
| [#7X3][#6H3]   | 0.9544 | CCC#CC#C | 0.0 |
| [CX3](=[OX1])C | 0.9517 | [CX2H0](#[OX1])O | 0.0 |
| [#6H3][#7]     | 0.9203 | [CX2H0](#[CX2H1])O | 0.0 |
| [#8]=[#6][#8]  | 0.8577 | C=CCCC#C | 0.0 |
| [CX4H3][NX3H0] | 0.8289 | [CX2H0](#[CX2H1])O | 0.0 |
| [#6H3][#7][#6H2] | 0.8185 | [CX4H1](#[CX4H2])O | 0.0 |
| [CX3](=[OX1])O | 0.7896 | [CX2H0](#[OX1])O | 0.0 |
| [#7X3H0]       | 0.7793 | [OX1H0][#8] | 0.0 |
| [#7][#6H2]     | 0.7557 | CCC#CC#C | 0.0 |

| worst negatives | prob | worst positives | prob |
|-----------------|------|-----------------|------|
| [CX4H3][OX2H0] | 0.5766 | [OX2H2](#[OX1H0])O | 0.1351 |
| [#6H3][#7][#6X3] | 0.3678 | [OX2H1] | 0.2571 |
| [#6H2][#7][#6X3] | 0.3098 | [OX2H1] | 0.3188 |
| [#8][#6][#6]=[#8] | 0.2945 | [#7][#6][#6X3] | 0.4147 |
| [#7X3H1]       | 0.2937 | [OX1H0]=([OX3H0])O | 0.4513 |
| [CX4H2][#6]O]  | 0.2519 | [#6X3][#6H2][#7] | 0.4727 |
| [OX1H0]=([OX3H0])O | 0.2169 | [CX4H2][OX3H0][CX4H3] | 0.6611 |
| [#8]=[#6H0][#6H1] | 0.2107 | [#8][#6][#6H2] | 0.6755 |
| [#8][#6][#6][#6X3] | 0.2044 | [#7X3][#6H2] | 0.7197 |
| [CH2X4][O]O[CX4H2] | 0.1983 | [CX4H2][CX3]O | 0.7255 |
Example 119 true smiles: CC(C)(N)(C(=O)O) formula: C4H9NO2
Index of correct structure: 0 of 896
True structure loss: 0.010581
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

0.010581
0.023607
0.035324
0.036309
0.036552

0.037491
0.03961
0.041186
0.045326
0.046072

Top predicted substructures
[#6H3][#6][#6]                    prob  0.998
{CX4H3}[#6]                               0.9769
| best positives | prob   | best negatives | prob   |
|----------------|--------|----------------|--------|
| #6H3][#6][#6]  | 0.998  | C=CCCC#C       | 0.0    |
| [CX4H3]        | 0.9966 | [CX2H1][CX2H0][CX3H1]=[CX3H0] | 0.0    |
| [OX1H1]        | 0.9966 | [OX1H1]=CCMH0][CX4H0] | 0.0    |
| [CX4H3]        | 0.9952 | CC=CCCC#C      | 0.0    |
| [CX4H3][CX4H0] | 0.9883 | CCC=CCCC#C     | 0.0    |
| [CX4H3][CX4H0] | 0.9867 | CC=CCCC#C      | 0.0    |
| [OX1H0]=[CX3H0][CX4H0][CX4H3] | 0.9952 | #6H3][#6H0] | 0.9646 |
| [CX4H3][#6H3][#6H0] | 0.9646 | #7][#6][#6H3] | 0.8187 |
| [OX2H1]        | 0.9693 | [OX2H1]=CCMH0][CX4H0] | 0.0    |
| [OX2H1]        | 0.9867 | [OX2H1]=CCMH0][CX4H0] | 0.0    |

| worst positives | prob   | worst negatives | prob   |
|-----------------|--------|-----------------|--------|
| [CX4H3][CX4H0]O| 0.6106 | [CH3]CC[OH]     | 0.1509 |
| [#8][#6][#6]=[#8] | 0.4852 | [#7H2][#6H0]   | 0.569  |
| [CH3][#6][#8]  | 0.2555 | [CX3]=O[OX2H1] | 0.5965 |
| [7X3H1]        | 0.1904 | [#7][#6][#6X3] | 0.6135 |
| [CX3H0]=([OX1H0][([OX2H1])][CX4H1] | 0.154 | [#6H3][#6][#6X3] | 0.6418 |
| [7X3H1]=([OX1H0][([OX2H1])][CX4H1] | 0.1539 | [#7][#6][#6X3] | 0.801  |
| [#8][#6][#6] | 0.137  | [#7X3H2]       | 0.8151 |
| [#8][#6H0][#6H1] | 0.1316 | [OX2H1]        | 0.8171 |
| [CX4H2][CC]=O  | 0.1212 | [#7][#6][#6H3] | 0.8187 |
Example 120 true smiles: OCC1CCCH1 formula: C5H11NO
Index of correct structure: 0 of 864
True structure loss: 0.017582
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

Top predicted substructures:

| Substructure               | prob |
|----------------------------|------|
| [CX4H2][(#6)][#6]            | 0.9998|
| [#7][#6H2][#6H2]             | 0.9635|
| Structure | prob | Structure | prob |
|-----------|------|-----------|------|
| [CX4H2][CX4H2] | 0.9937 | [CX4H2] | 0.9998 |
| [CX4H2][{CX4H2}][CX4H1] | 0.9937 | [#7][#6H2] | 0.9882 |
| OCC[CH2] | 0.9659 | [OX2H1] | 0.9499 |

**best positives**

| Structure | prob | Structure | prob |
|-----------|------|-----------|------|
| [CX4H2][{#6}][#6] | 0.9998 | [CX3H0][=#CX3H1]({OX2H0})[CX2H0] | 0.0 |
| [#7X3][#6H2] | 0.9973 | C=CC=CC=C | 0.0 |
| [CX4H2][{CX4H2}][CX4H1] | 0.9937 | C=CC=C=C | 0.0 |
| [#7][#6H2] | 0.9882 | [CX3H0][=#CX3H1][{CX4H1}][CX2H0] | 0.0 |
| OCC[CH2] | 0.9659 | [TX2H1]#[CX4H1]#[OX2H0]#[CX2H0] | 0.0 |
| [#6H1][#6H2] | 0.9635 | CC=CC=CC | 0.0 |
| [CX4H2][{CX4H2}] | 0.9607 | CC=CC=C=C | 0.0 |
| [#6H1] | 0.9599 | [CX3H1][=#CX3H1][CX2H0] | 0.0 |
| [#6H1][#6H2] | 0.9563 | [#6X3H1][=#6X3H1][#6X3H0][=#6X3H1] | 0.0 |
| [OX2H1] | 0.9499 | [CX3H1][=#CX3H2][CX2H0] | 0.0 |

**worst negatives**

| Structure | prob | Structure | prob |
|-----------|------|-----------|------|
| [#8][#6][#6H2] | 0.7279 | [CX4H2][O][CHX4] | 0.3232 |
| CH2X4[O][CX4H2] | 0.4326 | [#6H1r5][#7] | 0.3377 |
| [#6H1][#6][#6H2] | 0.3638 | [#7X3H1] | 0.425 |
| [#6H1][#6][{#6}][#6][#7][1] | 0.351 | [CX4H1][{NXH3}][{CX4H2}][CX4H2] | 0.4256 |
| CCCCCC | 0.3119 | [#6][#6][#6][#6][#7][1] | 0.4221 |
| [#6H2][#7][#6H2] | 0.2933 | [#8][#6H2][#6H1] | 0.4455 |
| O1{CX4H1}[CX4H2] | 0.2463 | [#6H1][#6H2][#6][#6][#7] | 0.4624 |
| [CX4H2][CX4H2][CX4H2][CX4H2] | 0.2156 | [#6H1][#6H2][#6][#6][#7] | 0.4735 |
| CH2X4[O][CX4H2][CX4H2] | 0.2076 | [CX4H2][{OX2H1}][CX4H1] | 0.5672 |
| [CX4H2][{OX2H1}][CX4H2] | 0.2025 | [#7][#6H1][#6H2r5] | 0.6475 |
Example 121 true smiles: Cc1ccc(O)ccc1 formula: C7H8O

Index of correct structure: 0 of 746
True structure loss: 0.009113
True structure:

Experimental 1H NMR (solvent: D2O)

Top predicted substructures (loss):

1. HO-phenyl
2. Phenol
3. Alcohol
4. Ether
5. Furane
6. Allyl

Top predicted substructures

| #6X3| #6X3 | prob   | #6H3| #6| #6 | prob   |
|-----|-----|-------|-----|---|---|-------|
|     |     | 0.9987|     |   |   | 0.9822|

Probabilities for top predicted substructures.
| Structure                  | prob | Structure                  | prob |
|----------------------------|------|----------------------------|------|
| [6X3][6X3][6X3][6X3]      | 0.9973 | [cH][cH]                  | 0.9801 |
| [6H1]                     | 0.9971 | [6X3H1][6X3H0]            | 0.9753 |
| [CX4H3][6]                | 0.9935 | [6H3][6H0]                | 0.9711 |
| [CX4H3]                    | 0.9914 | [61][6][6][6][6][6]       | 0.9661 |
| **best positives**         |      | **best negatives**        |      |
| [6X3][6X3]                | 0.9987 | [OX2H0][CX4H2][CX4H2][CX4H1] | 0.0  |
| [6X3][6X3][6X3][6X3]      | 0.9973 | [6H][6H1][6H2][6H1]==[6H] | 0.0  |
| [6H1]                     | 0.9971 | [CX4H1][NX3H2][CX4H2][CX4H1] | 0.0  |
| [CX4H3][6]                | 0.9935 | [OX1H0][NX3H1][CX4H1]     | 0.0  |
| [CX4H3]                    | 0.9914 | [CX4H1][NX3H1][CX4H3][CX4H2] | 0.0  |
| [6H3][6]                  | 0.9822 | [CX4H1][NX3H1][CX4H2][CX4H1] | 0.0  |
| [6H3]                      | 0.9801 | [CX4H1][NX3H1][CX4H3][CX4H2] | 0.0  |
| [6X3H1][6X3H0]            | 0.9753 | [CX4H1][NX3H0][CX4H2][CX4H0] | 0.0  |
| [6H3][6H0]                | 0.9711 | [6H3][6H1][6][6][6][6]    | 0.0  |
| [61][6][6][6][6][6]       | 0.9661 | [CX4H1][OX2H1][CX4H2][CX2H0] | 0.0  |
| **worst negatives**       |      | **worst positives**       |      |
| [cX3H1][cX3H1][cX3H1]     | 0.7674 | [OX2H1]                   | 0.534 |
| [cX3H0][cX3H1][cX3H0][CX4H3] | 0.5336 | [6X3H1][6X3H1][6X3H0][6X3H1] | 0.5378 |
| [cX3H1][cX3H0][cX3H0]     | 0.2326 | [6H][6][6][6][6]         | 0.6170 |
| [cX3H0][cX3H1][cX3H0][OX2H1] | 0.2013 | [OX2H][cX3][c]           | 0.6537 |
| [cX3H][6][6]              | 0.1295 | [6H][6H][6H][6H][6H]     | 0.6667 |
| [6X3][6][6][6][6][6]      | 0.1127 | [cX3H][cX3H][cX3H][cX3H][cX3H] | 0.6875 |
| [6H3][6]==[6X3][6X3][6]   | 0.1111 | [cH][c]                   | 0.7683 |
| [cX3H0][cX3H1][cX3H0][OX2H1] | 0.0968 | [6H1][6H1][6H1][6H1][6H1] | 0.8509 |
| [cX3H][cX3][cX3]          | 0.0938 | [6H3][6][6][6][6]        | 0.8858 |
| [6][6X3][6X3][6X3][6X3]   | 0.0863 | [CX4H3][cX3H0]           | 0.8944 |
Example 122 true smiles: Cc1ccc1c1O
Index of correct structure: 0 of 746
True structure loss: 0.007141

True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

| Structure | Loss   |
|-----------|--------|
| ![Structure 1] | 0.007141 |
| ![Structure 2] | 0.010702 |
| ![Structure 3] | 0.013357 |
| ![Structure 4] | 0.041378 |
| ![Structure 5] | 0.064152 |
| ![Structure 6] | 0.073154 |
| ![Structure 7] | 0.075647 |
| ![Structure 8] | 0.07753  |
| ![Structure 9] | 0.077702 |
| ![Structure 10] | 0.078329 |
| ![Structure 11] | 0.9994   |
| ![Structure 12] | 0.9938   |
|                       | prob       |                       | prob       |
|-----------------------|------------|-----------------------|------------|
| \[\#{H1}\]           | 0.9985     | \[\#{6H3}\][\#{6}][\#{6}] | 0.98       |
| \[\#{6X3}\][\#{6X3}][\#{6X3}][\#{6X3}] | 0.9983     | \[cH\][cH]           | 0.9766     |
| \[cX4H3\]            | 0.9977     | \[\#{6X3H1}\][\#{6X3H0}] | 0.9761     |
| \[cX4H3\][\#{6}]    | 0.9964     | \[\#{6X3}\][\#{6}][\#{6}][\#{6H3}] | 0.97       |
| **best positives**    |            |                       |            |
| \[\#{6X3}\][\#{6X3}] | 0.9994     | \[OX1H0\][\{cX3H0\}[\{cX4H1\}]|\{cX4H1\}] | 0.0        |
| \[\#{6H1}\]         | 0.9985     | \[OX2H0\][\{cX4H2\}]|\{cX4H1\}] | \{cX4H1\}] | 0.0        |
| \[\#{6X3}\][\#{6X3}][\#{6X3}][\#{6X3}] | 0.9983     | \[cX4H1\][\{cX3H2\}][\{cX4H2\}]|\{cX3H1\}] | 0.0        |
| \[cX4H3\]            | 0.9977     | \[cX4H0\][\{cX3H1\}][\{cX4H3\}][\{cX4H2\}] | 0.0        |
| \[cX4H3\][\#{6}]    | 0.9964     | \[\#{8}\][\#{6H1}][\#{6H2}][\#{6H1}] | \{c8\}    | 0.0        |
| \[\#{6H3}\][\#{6H0}] | 0.9938     | \[cX4H0\][\{cX3H1\}][\{cX4H2\}][\{cX4H1\}] | 0.0        |
| \[\#{6H3}\][\#{6}][\#{6}] | 0.98      | \[cX4H1\][\{cX3H0\}][\{cX4H2\}]|\{cX4H1\}] | 0.0        |
| \[cH1][cH]           | 0.9766     | \[cX4H1\][\{cX3H0\}][\{cX4H2\}] | \{cX4H0\}] | 0.0        |
| \[\#{6X3H1}\][\#{6X3H0}] | 0.9761     | \[cX4H1\][\{cX3H1\}][\{cX4H3\}][\{cX4H2\}] | 0.0        |
| \[\#{6X3}\][\#{6}][\#{6}][\#{6H3}] | 0.97      | \[OX2H0\][\{cX4H2\}][\{cX4H2\}]|\{cX4H1\}] | 0.0        |
| **worst negatives**   |            |                       |            |
| \[\#{6X3H1}\][\#{6X3H1}][\#{6X3H0}][\#{6X3H1}] | 0.3628     | \[cX3H0\][\{cX3H1\}][\{cX3H0\}][\{cX3H0\}][\{cX3H0\}] | 0.3608     |
| \[cX3H0\][\{cX3H1\}][\{cX3H1\}][\{cX3H0\}] | 0.3205     | \[\#{8}\][\#{6H0}][\#{6H0}] | \{c8\}    | 0.643      |
| \[cX3H3\][\{cX3H3\}] | 0.1941     | \[\#{8}\][\#{6}][\#{6}][\#{6X3}] | 0.7155     |
| \[cX3H0\][\{cX3H1\}][\{cX3H0\}][\{OX2H1\}] | 0.19      | \[OX2H1\][\{cX3\}][\{c8\}] | \{c8\}    | 0.7402     |
| \[cX3H1\][\{cX3H0\}][\{cX3H0\}] | 0.1843     | \[OX2H1\] | \{cX3H1\} | 0.7458     |
| \[\#{8}\][\#{6X3}][\#{6X3}][\#{6X3}][\#{6H3}] | 0.1509     | \[cX3H0\][\{cX3H1\}][\{cX3H0\}][\{cX4H3\}] | 0.7845     |
| \[\#{6H3}][\#{6}][\#{6X3}] | 0.1331     | \[cX3H1\][\{cX3H1\}][\{cX3H1\}] | \{c8\}    | 0.7979     |
| \[\#{6X3}\][\#{6}][\#{6H3}] | 0.1099     | \[cH1][cO] | \{c8\}    | 0.8557     |
| o[cH]                | 0.092      | \[\#{6H3}\][\#{6}][\#{6X3}] | \{c8\}    | 0.8741     |
| \[cH3\][\#{6}][\#{6}] | 0.0723     | \[\#{6H1}\][\#{6H1}] | \{c8\}    | 0.8827     |

------------------------------------------------------------------------------------------------------------------------
Example 123 true smiles: Cc1cccc(O)c1 formula: C7H8O
Index of correct structure: 0 of 746
True structure loss: 0.00733
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

| Structure | Loss |
|-----------|------|
| ![Structure 1](image1) | 0.00733 |
| ![Structure 2](image2) | 0.009841 |
| ![Structure 3](image3) | 0.01076 |
| ![Structure 4](image4) | 0.041811 |
| ![Structure 5](image5) | 0.060376 |
| ![Structure 6](image6) | 0.071828 |
| ![Structure 7](image7) | 0.073556 |
| ![Structure 8](image8) | 0.076027 |
| ![Structure 9](image9) | 0.076688 |
| ![Structure 10](image10) | 0.076963 |

Top predicted substructures
[#6X3][#6X3][#6X3][#6X3] prob 0.9998
[#6X3H1][#6X3H0] prob 0.9959
| best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| [6X3]|[6X3]|0.9997 | [6X3]|[6X3]|0.9953 |
| [CX4H3] |0.9993 | [cH]|cH |0.9884 |
| [#6H1] |0.9991 | [CX4H3]| [6]|0.9831 |
| [#6H3]| [6H0] |0.9972 | [6X3]| [6]| [6]| [6H3] |0.9772 |

| worst negatives | prob | worst positives | prob |
|----------------|------|----------------|------|
| [cX3H0]|([cX3H1]|([cX3H0]|([cX4H3] |0.3974 | [cX3H1]|([cX3H0]|([cX3H0] |0.4304 |
| [cX3H0]|([cX3H1]|([cX3H0]|([OX2H1]|0.2473 | [cX3H0]|([cX3H1]|([cX3H0]|([OX2H1] |0.5777 |
| [cX3H0]|([cX3H1]|([cX3H0]|([cX3H0] |0.2423 | [6]| [6X3]| [6]| [6X3]| [6]| [6] |0.59 |
| [cX3H1]|([cX2H0]|([cX3H1]|0.1624 | [OX2H1] |0.6697 |
| [#8]| [#6H1]| [#6H1]|0.1325 | [OX2H1]| [cX3]| [c] |0.7578 |
| [#8]| [#6H1]| [#6X3]|0.0895 | [cX3H1]|([cX3H1]|([cX3H1] |0.8513 |
| [cH]| [6]| [6]|0.0789 | [ch]|cO |0.8614 |
| [CX3]| [O]|0.0753 | [6]| [6]| [6]| [6]| [6]|1 |0.8717 |
| [CH3]| [CH3]|0.0711 | [CX4H3]|([cX3H0] |0.8890 |
Example 124 true smiles: COCCC(=O)OC formula: C5H10O3
Index of correct structure: 0 of 739
True structure loss: 0.015744
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

| Substructure | Prob  |
|--------------|-------|
| [CX4H3]      | 0.9947|
| [CX4H2]{[#6]}[#6] | 0.9747|
Example 125 true smiles: CCC(C)(O)(=O)O formula: C5H10O3
Index of correct structure: 0 of 739
True structure loss: 0.009601

True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

| Structure | Loss |
|-----------|------|
| ![Structure 1](image) | 0.009601 |
| ![Structure 2](image) | 0.046423 |
| ![Structure 3](image) | 0.061202 |
| ![Structure 4](image) | 0.06516 |
| ![Structure 5](image) | 0.065722 |
| ![Structure 6](image) | 0.0741 |
| ![Structure 7](image) | 0.07679 |
| ![Structure 8](image) | 0.07681 |
| ![Structure 9](image) | 0.078138 |
| ![Structure 10](image) | 0.078969 |

Top predicted substructures

| Substructure | prob | loss |
|--------------|------|------|
| [#8][#6][#6] | 1.0 | 0.9927 |
| [#8]=[#6][#8] | | |
|      | prob |      | prob |
|------|------|------|------|
| [CX4H3]| 0.9999 | [#6H3]#/6H0] | 0.9914 |
| [CX4H3]#/6 | 0.9998 | [CX4H2]#/6]#/6 | 0.9913 |
| [OX2H1] | 0.9993 | [CH3]CC[OH] | 0.9907 |
| [CX3]#/OX1]C | 0.9988 | [CX4H3]#/CX4H2] | 0.9868 |
|      |      |      |      |
| best positives |      |      |      |
| [#6H3]#/6]#/6 | 1.0 | CC=CC#CC | 0.0 |
| [CX4H3] | 0.9999 | C=C=C#C | 0.0 |
| [CX4H3]#/6 | 0.9998 | CC=CC#C | 0.0 |
| [OX2H1] | 0.9993 | CC#CCC#C | 0.0 |
| [CX3]#/OX1]C | 0.9988 | [CX2H1]#/CX2H0]#/CX3H1]#/CX3H0] | 0.0 |
| [#8]#/6]#/8 | 0.9927 | CCC#CCC#C | 0.0 |
| [CX4H3]#/6H0] | 0.9914 | [OX3H0]#/OX3H1]#/OX2H0]#/OX2H1] | 0.0 |
| [CX4H3]#/6]#/6 | 0.9913 | [#6X2]#/6H1]#/6X2] | 0.0 |
| [OH]#/CX4H] | 0.9907 | [#7]#/6]#/6]#/6]#/7] | 0.0 |
| [CX4H3]#/CX4H2] | 0.9868 | [CX2H0]#/CX2H1]#/CX3H1] | 0.0 |
| worst negatives |      |      |      |
| [#8]#/6H0]#/6H1] | 0.6323 | [#8]#/6]#/6]#/8] | 0.5714 |
| [CX3H0]#/OX1H0]#/OX2H1]#/CX4H1] | 0.6148 | [CX4H2]#CC=O | 0.6551 |
| [#6H1] | 0.3919 | [#6H3]#/6]#/6H3] | 0.711 |
| [OH]#/CX4H] | 0.3251 | [CX4H2]#/OX4H3]#/CX4H0] | 0.722 |
| [#8]#/6]#/6]#/6X3] | 0.3041 | [#6H3]#/6]#/6X3] | 0.76 |
| #8]#/6H0]#/6H1] | 0.2186 | [#6X3]#/6]#/6]#/6H3] | 0.8131 |
| O#6H]#/6H]#/6H0] | 0.1529 | [OX2H1]#/CX4H0]#/CX4H2]#/CX4H3] | 0.8562 |
| [#6H]#/6X4H]#/6X3H0] | 0.1568 | [CH3]#/6]#/8 | 0.8638 |
| O#CX4H]#/CX4H0] | 0.1345 | [#8]#/6]#/6H2] | 0.8913 |
Example 126 true smiles: \( \text{O=C(O)CCCCO} \) formula: \( \text{C}_5\text{H}_{10}\text{O}_3 \)

Index of correct structure: 0 of 739
True structure loss: 0.007485

True structure:

![Chemical structure](image)

Experimental \( ^{13}\text{C} \) NMR (solvent: CDC13)

![NMR spectrum](image)

Experimental \( ^1\text{H} \) NMR (solvent: D2O)

![NMR spectrum](image)

Top predicted substructures (loss):

| Structure                  | prob  |
|----------------------------|-------|
| \([\text{CX}_4\text{H}_2]\{[\#6]\}\{[\#6]\} \) | 1.0   |
| \([\text{CX}_3]=\{[\text{OX}_1]\}\text{O} \) | 0.9973|
| \([\text{CX}_4\text{H}_2]\{[\#6]\}\{[\#6]\} \) | 0.007485 |
| \([\text{CX}_3]=\{[\text{OX}_1]\}\text{O} \) | 0.046104 |
| \([\text{CX}_4\text{H}_2]\{[\#6]\}\{[\#6]\} \) | 0.048947 |
| \([\text{CX}_3]=\{[\text{OX}_1]\}\text{O} \) | 0.07109 |
| \([\text{CX}_4\text{H}_2]\{[\#6]\}\{[\#6]\} \) | 0.076818 |
| \([\text{CX}_4\text{H}_2]\{[\#6]\}\{[\#6]\} \) | 0.08025 |
| \([\text{CX}_4\text{H}_2]\{[\#6]\}\{[\#6]\} \) | 0.080297 |
| \([\text{CX}_4\text{H}_2]\{[\#6]\}\{[\#6]\} \) | 0.082838 |
| \([\text{CX}_4\text{H}_2]\{[\#6]\}\{[\#6]\} \) | 0.083664 |
| \([\text{CX}_4\text{H}_2]\{[\#6]\}\{[\#6]\} \) | 0.083805 |
[\#8]=[#6][\#8] 0.9993  [CX4H2][CX4H2] 0.9973
OCC[CH2] 0.9989  [#8][#6][#6H2] 0.997
[CX2H1] 0.9989  [CX4H2][CX4H2][CX4H2] 0.9894
[CX1](=OX1)C 0.9976  O=[CX3H0][CX4H2][CX4H2] 0.9851

best positives
prob  best negatives
prob

[CX4H2][(#6)[#6] 1.0  [CX3H0]==[CX3H1][{OX2H0}][CX2H0] 0.0
[\#8]=[#6][#8] 0.9993  [CX2H0][#CX2H1][cX3H0] 0.0
OCC[CH2] 0.9989  CCC=CC==C 0.0
[CX2H1] 0.9989  C=CC=CC=C 0.0
[CX1](=OX1)C 0.9976  [CX2H0][#CX2H1][CX4H0] 0.0
[CX4H2][CX4H2] 0.9973  C=CC=C=CC 0.0
[CX3H0](=OX1)C 0.9976  [CX2H1][#CX2H0][CX3H1]=[CX3H0] 0.0
[CX4H2][CX4H2] 0.9973  [CX2H0][#CX2H1][CX4H0] 0.0

worst negatives
prob  worst positives
prob

[CX4H2][CX4H2][CX4H1] 0.8182  [CX4H2][{OX2H1}][CX4H2] 0.7688
[\#8H] 0.5007  [CH2X4](O)[CX4H2] 0.8153
[#8H][#6H2] 0.4171  [CX3H0]==[OX1H0][{OX2H1}][CX4H1] 0.8454
[CX3H0][=OX1H0][{OX2H1}][CX4H1] 0.3302  [CX4H2][CC=O] 0.8848
[\#8H][#6][#6][#6/#8] 0.3113  [CX4H2][CX3]=O 0.8888
[\#8H][#6H2][#6H1] 0.3111  [CX4H2][OX1H0][{OX81}][{CX4H2}][CX3H0] 0.9355
[CX2H0][CX3H0][CX4H2] 0.2699  [CX3H2][{OX4H2}][CX4H0] 0.9543
[CX2H0][CX4H2][CX3H0] 0.1407  [CX4H2][CX4H2][CX4H2][CX4H2] 0.9638
[\#8][#6][#6][#6][#6][#8] 0.1407  [CX4H2][OX1H0][{OX81}][{OX81}][{CX4H2}][CX4H2] 0.9759
Example 127 true smiles: CC(C(O)C(=O)O formula: C5H10O3
Index of correct structure: 0 of 739
True structure loss: 0.011302
True structure:

Experimental 13C NMR (solvent: DMSO)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

| Substructure | Loss  |
|--------------|-------|
| 
| 0.011302 |
| 0.050509 |
| 0.061464 |
| 0.062112 |
| 0.06217 |
| 0.063761 |
| 0.070679 |
| 0.072699 |
| 0.077151 |
| 0.079581 |

Top predicted substructures

| Prob | Substructure |
|------|--------------|
| 0.999 | CX3(=O1)C |
| 0.996 |          |
|                | prob |                | prob |
|----------------|------|----------------|------|
| [CX4H3]        | 0.9998 | [CX3]=([OX1])O | 0.9949 |
| [CX4H3][#6]   | 0.9997 | [OX2H1]        | 0.9937 |
| [#8]=[#6][#8] | 0.9985 | [CHX4][([CH3X4])][CH3X4] | 0.9933 |
| [#6H1]        | 0.9981 | [CX3]=O[OX2H1] | 0.9722 |

**best positives**

|                | prob |                | prob |
|----------------|------|----------------|------|
| [#6H3][#6][#6] | 0.9999 | [CX2H1][#CX2H0][CX3H1]= [CX3H0] | 0.0  |
| [CX4H3]        | 0.9998 | CC#CCC#C       | 0.0  |
| [CX4H3][#6]   | 0.9997 | CCC#CCC#C      | 0.0  |
| [#8]=[#6][#8] | 0.9985 | [CX2H0]=[#CX2H0][CX2H0] | 0.0  |
| [#6H1]        | 0.9981 | [CX3H0]=([CX3H1]) ([CX4H2])[CX2H0] | 0.0  |
| [CX3]=([OX1])O| 0.996  | CCC=CC=C       | 0.0  |
| [CX3]=([OX1])C| 0.9949 | CC=CC=CC       | 0.0  |
| [OX2H1]       | 0.9937 | [#6X2]=#6H1] [#6X2] | 0.0 |
| [CHX4][([CH3X4])[CH3X4]] | 0.9933 | CCC=CC=C       | 0.0  |
| [CX3]=O[OX2H1]| 0.9722 | [CX3H0]=([CX3H1]) ([OX2H0])[CX2H0] | 0.0  |

**worst negatives**

|                | prob |                | prob |
|----------------|------|----------------|------|
| [#6H1][#6H2]  | 0.5933 | [#8]=[#6H1][#6H1] | 0.3302 |
| [CHX4][([CH3X4])[CH2X4]] | 0.2751 | [#6H1][#6H1] | 0.4853 |
| [#8]=[#6][#6H2]| 0.2661 | [OH][CX4H] | 0.5715 |
| [CX4H1][([CX4H3])[CX4H3])[CX4H2] | 0.254 | [#8]=[#6][#6H1][#6H1] | 0.6868 |
| OCC[CH2]      | 0.2407 | [CX4H0] | 0.6907 |
| [CX4H1][([OX2H1])[([CX4H2])[CX3H0]] | 0.2355 | [#8]=[#6H0][#6H1] | 0.7058 |
| [CX4H2]=CC=O  | 0.2329 | [#8]=[#6H0][#6H1] | 0.7568 |
| [#8]=[#6][#6][#6X3]| 0.197 | [CX3H0]=([OX1H0])[([OX2H1])[CX4H1] | 0.7572 |
| [CX4H2][CX3]=O | 0.1727 | [CH3][CC=OH] | 0.7739 |
| [CH3][#6][#6]| 0.1348 | O=[CX3][CX4H] | 0.776 |

------------------------------------------------------------------------------------------------------------------------
Example 128 true smiles: OCC(O)C(O)C(O)CO formula: C5H12O5
Index of correct structure: 0 of 734
True structure loss: 0.010997
True structure:

![Structural formula]

Experimental $^{13}$C NMR (solvent: D$_2$O)

![$^{13}$C NMR spectrum]

Experimental $^1$H NMR (solvent: D$_2$O)

![$^1$H NMR spectrum]

Top predicted structures (loss):

| Structure | Loss  |
|-----------|-------|
| ![Structure 1] | 0.010997 |
| ![Structure 2] | 0.013642 |
| ![Structure 3] | 0.019935 |
| ![Structure 4] | 0.022565 |
| ![Structure 5] | 0.022975 |
| ![Structure 6] | 0.022978 |
| ![Structure 7] | 0.023036 |
| ![Structure 8] | 0.024061 |
| ![Structure 9] | 0.024443 |
| ![Structure 10] | 0.025249 |

Top predicted substructures

| Substructure | prob |
|--------------|------|
| ![Substructure 1] | 0.9997 |
| ![Substructure 2] | 0.9862 |
| Compound | Score | Compound | Score |
|----------|-------|----------|-------|
| #8][#6][#6][#8] | 0.9966 | [CX4H]O | 0.9561 |
| #8][#6][#6H2][#8] | 0.9958 | OCC[CH2] | 0.9414 |
| [CX4H](O)CO | 0.9954 | [#6H1] | 0.9312 |
| #8][#6][#6H2] | 0.9884 | O[CX4H][CX4H2] | 0.9255 |

### Best Positives

| Compound | Score | Compound | Score |
|----------|-------|----------|-------|
| [OX2H1] | 0.9997 | [CX2H1]=[CX2H0][CX3H1]=[CX3H0] | 0.0 |
| [#8][#6][#6][#8] | 0.9966 | CCCCC=C | 0.0 |
| [#8][#6][#6H2][#8] | 0.9958 | [#7][#6]=[#6][#6]=[#7] | 0.0 |
| [CX4H](O)CO | 0.9954 | CC=CC=C | 0.0 |
| [#8][#6][#6H2] | 0.9884 | [CX3H0][CX4H2][CX3H1]=[CX3H0] | 0.0 |
| [CX4H2][#6]()[O] | 0.9862 | CC=CCCC=C | 0.0 |
| [CX4H]O | 0.9861 | [#7][#6][#6]=[#7] | 0.0 |
| OCC[CH2] | 0.9414 | [OX1H0]=[CX3H0][CX2H0][CX2H1] | 0.0 |
| [#6H1] | 0.9312 | O[CX4H][CX4H2] | 0.0 |

### Worst Negatives

| Compound | Score | Compound | Score |
|----------|-------|----------|-------|
| [CX4H2][CX4H2] | 0.4875 | [#6X4H1][#6X4H1][#6X4H1] | 0.2404 |
| [#6H2][#6H1][#6H1][#6H2] | 0.3715 | [CX4H](O)[(CH)](CH) | 0.2744 |
| [CH2X4]([O])[CX4H2] | 0.3071 | [CX4H1][#OX2H1][[CX4H2]][CX4H2] | 0.4353 |
| [CX4H1][#OX2H0][CX4H2] | 0.2426 | [#8][#6][#6][#6][#6][#8] | 0.4735 |
| [OX2H0][CX4H2][CX4H1][CX4H2] | 0.2124 | [#8][#6H1][#6H1] | 0.4889 |
| [#6H1][#6H2][#6H2] | 0.159 | [CX4H1][#OX2H1][[CX4H2]][CX4H2] | 0.6617 |
| [OX2H1][CX4H2][#OX2H0][CX4H2] | 0.1416 | [#6H1][#6H1] | 0.7223 |
| [OX2H0][CX4H2][#OX2H1][CX4H2] | 0.0988 | [CX4H1][#OX2H1][[CX4H2]][CX4H2] | 0.7763 |
| O[CX4H2][#CX4H2] | 0.0701 | [OX2H1][CX4H1][CX4H1][O2XH1] | 0.7805 |
| [CX4H2][OX2H0][CX4H2] | 0.07 | O[CX4H][#CX4H][#CX4H] | 0.7868 |
Example 129 true smiles: OCC(O)C(O)C(O)CO formula: C5H12O5
Index of correct structure: 0 of 734
True structure loss: 0.012278
True structure:

Experimental $^{13}$C NMR (solvent: D2O)

Experimental $^1$H NMR (solvent: D2O)

Top predicted structures (loss):

- Top predicted substructures                   prob
- [OX2H1]                                       0.9997          [#8][#6][#6H2]                                0.9788

- 0.012278
- 0.015361
- 0.022517
- 0.022808
- 0.023509

- 0.024467
- 0.024473
- 0.025082
- 0.025642
- 0.026691

Top predicted substructures

[OX2H1]  prob  0.9997  [#8][#6][#6H2]  0.9788
Example 130 true smiles: O=C(O)CC(=O)C(=O)O

Index of correct structure: 0 of 729

True structure loss: 0.025576

True structure:

![Chemical structure](image)

Experimental $^{13}$C NMR (solvent: D$_2$O)

Experimental $^1$H NMR (solvent: D$_2$O)

Top predicted structures (loss):

- $[^{13}C]X(=^{13}O)C$: 0.9999
- $[^{13}O]X1$: 0.8852
- $[^{13}O]X2H1$: 0.862491
- $[^{13}O]X3$: 0.865158
- $[^{13}O]X4$: 0.865795
- $[^{13}O]X5$: 0.869105
- $[^{13}O]X6$: 0.870087
- $[^{13}O]X7$: 0.870771
- $[^{13}O]X8$: 0.8852
Example 131 true smiles: C1CNCCN1 formula: C5H12N2
Index of correct structure: 0 of 619
True structure loss: 0.012311

Top predicted structures:

Top predicted substructures

Top predicted substructures

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)
| best positives | prob   | best negatives | prob   |
|----------------|--------|----------------|--------|
| [CX4H2]([#6])[#6] | 0.9663 | [CX4H2][#6]   | 0.7754 |
| [#7] [#6H2]     | 0.9604 | [CX4H2][#6X3H1][CX4H2] | 0.7481 |
| [#7] [#6H2] [#6H2] | 0.9363 | [#7X3H1] | 0.7274 |
| [#7] [#6H2] [#6H2] [#7] | 0.8547 | [#7X3H2] | 0.6358 |

| worst negatives | prob   | worst positives | prob   |
|-----------------|--------|-----------------|--------|
| [#7X3H2]       | 0.6358 | [CX4H2][#6H2][CX4H2] | 0.3753 |
| [#6H1] [#6H2] [#6H2] | 0.4555 | [#7] [#6] [#6] [#6] [#7] | 0.5167 |
| [#6H1] [#6H2]   | 0.3999 | [#7] [#6] [#6] [#6] [#7] | 0.6099 |
| [CX4H2][#6X3H1][CX4H2] | 0.39 | [#7X3H1] | 0.7274 |
| [#7] [#6] [#6] [#6] [#6] [#7] | 0.3829 | [CX4H2][#6X3H1][CX4H2] | 0.7481 |
| [#6H1] | 0.3811 | [#6H2] [#7] [#6H2] | 0.7754 |
| [#7H2] [#6H2]   | 0.3281 | [CX4H2][CX4H2] | 0.8451 |
| [CX4H3]         | 0.2858 | [#7] [#6H2] [#6H2] [#7] | 0.8547 |
| [CX4H2][CX4H2][CX4H1] | 0.2241 | [#7] [#6H2] [#6H2] | 0.9363 |
| [#7H2] [#6H1]   | 0.2211 | [#7] [#6H2] | 0.9604 |
Example 132 true smiles: NC(CO)C(=O)O formula: C₃H₇NO₃
Index of correct structure: 0 of 576
True structure loss: 0.021597
True structure:

Experimental ¹³C NMR (solvent: D₂O)

Experimental ¹H NMR (solvent: D₂O)

Top predicted structures (loss):

| Structure | Loss |
|-----------|------|
| [CX₃](-[OX₁])C | 0.9709 |
| [CX₃](-[OX₁])O | 0.824 |
| 0.021597 | 0.036092 | 0.041058 | 0.04681 | 0.049029 |
| 0.049246 | 0.05125 | 0.051951 | 0.053941 | 0.054471 |

Top predicted substructures

[CX₃](-[OX₁])C 0.9709  [CX₃](-[OX₁])O 0.824
| Structure | Probability | Structure | Probability |
|-----------|-------------|-----------|-------------|
| [O2H1]    | 0.9666      | [#8]=[#6H0][#6H1] | 0.0081      |
| [#8]=[#6H0][#8] | 0.8923        | O=[CX3][CX4H] | 0.7977      |
| [CX4H2][{#6}][O] | 0.864 | [#7X3H2] | 0.7683      |
| [CX3]={O}[O2H1] | 0.8283 | [CX4H2]CC=O | 0.7541      |

**Best Positives**

| Structure | Probability | Structure | Probability |
|-----------|-------------|-----------|-------------|
| [CX3]=[=O1]C | 0.9709 | [CX2H1][=CX2H0][CX3H1]=[CX3H0] | 0.0 |
| [O2H1] | 0.9666 | CC=CCC#C | 0.0 |
| [#8]=[#6][#8] | 0.8923 | CC=CC#CC | 0.0 |
| [CX4H2][{#6}][O] | 0.864 | [CX2H0][=CX2H1][CX4H2] | 0.0 |
| [CX3]={O}[O2H1] | 0.8283 | [CX4H2][{CX4H3}][CX2H0] | 0.0 |
| [CX3]=[=O1]O | 0.824 | [#6H2][#6][#6X2] | 0.0 |
| [#8]=[#6H0][#6H1] | 0.8081 | [#6X3][#6][#6][#6H3] | 0.0 |
| O=[CX3][CX4H] | 0.7797 | [#6X2][#6H1][#6X2] | 0.0 |
| [#7X3H2] | 0.7683 | [CX2H0][=CX2H0][CX4H0] | 0.0 |
| [CX4H2]CC=O | 0.7541 | CCC#CC=C | 0.0 |

**Worst Negatives**

| Structure | Probability | Structure | Probability |
|-----------|-------------|-----------|-------------|
| [#8][#6][#6H2] | 0.5467 | [#8][#6H0][#6H1] | 0.1484      |
| [#7][#6][#6H0] | 0.3305 | [CX4H1][=CX2H0][CX4H2][CX3H0] | 0.2368      |
| [CX4H2][{OX2H1}][O2H1][CX4H2] | 0.3033 | OCC[CX2H2] | 0.2631      |
| [#8][#6][#8] | 0.2452 | [#8][#6H1][#6H2] | 0.3436      |
| [CH2X4][O][CX4H2] | 0.2288 | [CX4H2][O][CX4H4] | 0.3481      |
| [#7][#6H0][#6H1] | 0.2254 | [#8][=CX2H0][=CX2H1][#6X3] | 0.4181      |
| [#7X3H1] | 0.2071 | [#8][=CX2H0][=CX2H1][#6X3] | 0.419      |
| [CX4H2][O2H1] | 0.2015 | [CX4H2][=OX2H1][CX4H1] | 0.4192      |
| [CX4H2][CX4H2] | 0.1807 | [#7][#6][#6K3] | 0.5001      |
| [#8][#6][#6H1][#6H1] | 0.1802 | [#7H2][#6H1] | 0.5054      |
Example 133 true smiles: NC(O)C(=O)O formula: C3H7NO3
Index of correct structure: 0 of 576
True structure loss: 0.023511
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: d2o)

Top predicted structures (loss):

- \([\text{CX4H2}]\{\#6}\)\{0\} with prob 0.9831
- \([\#7X3H2]\) with prob 0.7948
[O]X\[H]1\] 0.9727  [\#8]=[\#6\#0]\[\#6\#1\] 0.7974
[CX3]=([O]X\[1]\)\)C  0.9698  O=[CX3][CX4\#0] 0.7700
[\#8]=\[\#6\#0\] 0.8216  [\#8]=[\#6\#2]\[\#6\#1\] 0.7487
[CX4\#2]CC=O  0.7985  [\#8\#0]\[\#6\#2]\[\#6\#1\] 0.7033

best positives

[CX4\#2]\[\#6\] 0.9831  CX2\#1\)\#(CX2\#0] 0.9831  CX3H1]=\[CX3\#0\] 0.0
[O]X\[1\] 0.9727  CC=CCC\#C 0.0
[CX3]=([O]X\[1\])\)C  0.9698  CC=C\#C\#C 0.0
[\#8]=\[\#6\#0\] 0.8216  \[CX4\#2\]([CX4\#3]\)[CX2\#0]\[\#6\#2\] 0.0
[CX4\#2]CC=O  0.7985  \[\#6\#2\]\#\[\#6\#2\] 0.0
[\#7X3\#2\] 0.7948  \[\#6\#3\]\#\[\#6\#3\] 0.0
[\#8]=\[\#6\#0\] 0.7974  \[\#6\#2\] 0.0
O=[CX3][CX4\#0] 0.7708  \[CX4\#2\]([CX4\#0]\)[CX2\#0]\[\#6\#2\] 0.0
[\#8]=\[\#6\#2]\[\#6\#1\] 0.7487  CCC\#C\#C 0.0
[\#8\#0]\[\#6\#2]\[\#6\#1\] 0.7033  \[CX2\#0]\)\#\[\#CX2\#1\] 0.0

worst negatives

[\#8]=\[\#6\] 0.6951  \[CX4\#1\]([NX3\#2]\)[CX4\#2]\)[CX3\#0\] 0.1569
[\#7\#2]=[\#6\#0\] 0.4281  \[\#8]=\[\#6\#0\][\#6\#1\] 0.2227
[\#7\#2]=\[\#6\] 0.4152  \[\#6\#1\] 0.407
[CX4\#2]([CX4\#2]\)[CX4\#2]\] 0.4071  \[\#7\#2\]\#\[\#6\#1\] 0.4928
O([CX4\#2]\)[CX4\#2]\)\] 0.4031  \[CX3\#0\]=([O]X\#1\#)[\#(CX2\#1]\)[CX4\#1]\] 0.436
[CX2\#4\]\)[OX]\[CX4\#2]\] 0.3962  \[\#8]=\[\#6\]\#\[\#6X3\] 0.4564
[\#8]=\[\#6\#2\] 0.3961  OCC\#C\#2\] 0.4856
[CX4\#2]\[\#(O\#X2\#0\])\)[CX4\#2]\] 0.3322  \[CX4\#2\]\)[O\)\)[CX4\#4\]\] 0.4961
[\#8]=\[\#6\#0\] 0.2676  \[\#7]\)[\#6\]\#\[\#6\#3\] 0.5293
[O]X\#0\]([CX4\#2]\)[CX4\#2]\)[OX2\#0\] 0.2431  \[\#7\#2\]\#\[\#4\#1\] 0.5381

------------------------------------------------------------------------------------------------------------------------
Example 134 true smiles: CCCCC(C)=O formula: C7H14O
Index of correct structure: 0 of 556
True structure loss: 0.006778
True structure:

![Structural diagram]

Experimental $^{13}$C NMR (solvent: CDCl$_3$)

Experimental $^1$H NMR (solvent: D$_2$O)

Top predicted structures (loss):

| Structure | Loss |
|-----------|------|
| ![Substructure 1] | 0.006778 |
| ![Substructure 2] | 0.034776 |
| ![Substructure 3] | 0.035885 |
| ![Substructure 4] | 0.049086 |
| ![Substructure 5] | 0.054234 |
| ![Substructure 6] | 0.060119 |
| ![Substructure 7] | 0.061917 |
| ![Substructure 8] | 0.067532 |
| ![Substructure 9] | 0.072748 |
| ![Substructure 10] | 0.079484 |

Top predicted substructures

| [CX4H3] | prob | [CX4H3][CX4H2] | prob |
|---------|------|----------------|------|
|         | 1.0  |                 | 0.9995 |
| best positives | prob | best negatives | prob |
|----------------|------|---------------|------|
| [CX4H3]        | 1.0  | [CX2H0][#[CX2H1]][#CX3H0] | 0.0  |
| [CX4H2][#6][#6] | 0.9999 | [CX2H0][#CX2H0][CX2H0] | 0.0  |
| [CX3](#OX1)C   | 0.9997 | [CX3H0][#OX2H0][CX2H0] | 0.0  |
| [CX4H3][#6]    | 0.9996 | [OX1H0][#CX3H0][#CX4H2] | 0.9769 |
| [CX4H3][CX4H2] | 0.9995 | [CX2H1][#CX2H0][CX3H1] | 0.0  |
| [OX1H0][#CX3H0] | 0.9987 | [CX4H1][#OX2H0][#CX4H2] | 0.0  |

| worst negatives | prob | worst positives | prob |
|-----------------|------|----------------|------|
| [#6H1]          | 0.4646 | [CX4H2][#CX4H1][#CX4H2][#CX4H2] | 0.4765 |
| [#6H1][#6H2]    | 0.4448 | [CX4H2][#CX4H2][CX4H2] | 0.6509 |
| [CX4H2][#CX4H1][#CX3H0] | 0.3074 | [CX4H2][#CX4H2] | 0.7293 |
| [CX4H2][#CX4H3][#CX4H1] | 0.2626 | [CX4H2][#CX4H2][#CX4H2] | 0.7635 |
| [#6X3][#6][#6] | 0.2416 | O=[CX3H0][#CX4H2][#CX4H2] | 0.807 |
| [#6H1][#6H2][#6H2] | 0.2262 | [CX4H2][#OX2H0]| 0.8609 |
| O=[CX3H0][#CX4H2][#CX4H1] | 0.1597 | [CX4H2][#CX4H3][#CX4H2] | 0.8841 |
| [#6X3][#6][#6][#6] | 0.1377 | CCCCCC | 0.9076 |
| [#8]=[#6][#6H2][#6H1] | 0.1072 | [#6H3][#6X3H0][#6H2] | 0.9153 |
| [#6H3][#6][#6X3] | 0.1032 | [CX4H2][#CX3]=O | 0.9182 |
Example 135 true smiles: CCCC(=O)CCC formula: C7H14O
Index of correct structure: 0 of 556
True structure loss: 0.008486
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

Top predicted substructures

| Substructure | prob   | Substructure | prob   |
|--------------|--------|--------------|--------|
| [CX4H3][#6] | 0.9999 | [CX4H3][CX4H2] | 0.9976 |
| Compound | Prob | Compound | Prob |
|----------|------|----------|------|
| [CX4H3]  | 0.9998 | [OX1H0]=[CX3H0]([#6])|CX4H2]  | 0.9847 |
| [OX1]C   | 0.9997 | [CX4H2]([CX4H3])|CX4H2]  | 0.9812 |
| [#6H3][#6] | 0.9994 | [CX3H0]=|OX1H0|)([CX4H2] ]|CX4H2]  | 0.8176 |
| [CX4H2]([#6])|#6] | 0.9999 | [CX2H0],[#2X1]X0]|CX3H0] | 0.0 |
| [CX4H3]  | 0.9998 | CCC#CC#C | 0.0 |
| [CX3](=#[OX1]C) | 0.9997 | C=CC#CC | 0.0 |
| [CX4H2]([#6])|#6] | 0.9994 | CCC#CC | 0.0 |
| [OX1H0]=|OX1H0|[CX4H2]  | 0.9976 | [CX2H1],[#2X1]X0]|CX3H1]=|CX3H0] | 0.0 |
| [OX1H0]=|CX3H0]([#6][CX4H2]  | 0.9847 | [CX2B1],[#2X2H0]X0]|CX3H1]=|CX3H0] | 0.0 |
| [CX4H2]([CX4H3])|C4H2] | 0.9812 | [CX2H1],[#2X2H0]X0]|CX4H1]=[OX2H1] | 0.0 |
| [CX4H2]=|C3]|0 | 0.8176 | [CX2H0],[#2X2H0]X0]|CX2H0] | 0.0 |
| [CX3H0]=|OX1H0|[CX4H2]  | 0.816 | CCC#CC|C | 0.0 |

| Compound | Prob | Compound | Prob |
|----------|------|----------|------|
| [CX4H3]  | 0.9999 | [OX1H0]=[CX3H0]([#6][CX4H2]  | 0.9847 |
| [CX4H3]  | 0.9998 | CCC#CC#C | 0.0 |
| [CX3](=#[OX1]C) | 0.9997 | C=CC#CC | 0.0 |
| [CX4H2]([#6])|#6] | 0.9994 | CCC#CC | 0.0 |
| [OX1H0]=|OX1H0|[CX4H2]  | 0.9976 | [CX2H1],[#2X1]X0]|CX3H1]=|CX3H0] | 0.0 |
| [OX1H0]=|CX3H0]([#6][CX4H2]  | 0.9847 | [CX2B1],[#2X2H0]X0]|CX3H1]=|CX3H0] | 0.0 |
| [CX4H2]([CX4H3])|C4H2] | 0.9812 | [CX2H1],[#2X2H0]X0]|CX4H1]=[OX2H1] | 0.0 |
| [CX4H2]=|C3]|0 | 0.8176 | [CX2H0],[#2X2H0]X0]|CX2H0] | 0.0 |
| [CX3H0]=|OX1H0|[CX4H2]  | 0.816 | CCC#CC|C | 0.0 |
Example 136 true smiles: CC(=O)O)C(=O)O formula: C₄H₆O₄
Index of correct structure: 0 of 502
True structure loss: 0.015123
True structure:

Experimental ¹³C NMR (solvent: DMSO)

Experimental ¹H NMR (solvent: D₂O)

Top predicted structures (loss):

| Structure | Loss |
|-----------|------|
| ![Structure 1](image1) | 0.015123 |
| ![Structure 2](image2) | 0.036391 |
| ![Structure 3](image3) | 0.059477 |
| ![Structure 4](image4) | 0.062217 |
| ![Structure 5](image5) | 0.064527 |

| Structure | prob |
|-----------|------|
| ![Structure 1](image1) | 0.9995 |
| ![Structure 2](image2) | 0.9596 |
| ![Structure 3](image3) | 0.96825 |
| ![Structure 4](image4) | 0.969131 |
| ![Structure 5](image5) | 0.970913 |
| Structure                  | Prob   | Structure                  | Prob   |
|---------------------------|--------|---------------------------|--------|
| \( \#8=\#6=\#8 \)        | 0.9982 | \([\text{OX2H1]}\)       | 0.0491 |
| \([\text{CX3}]\{=\text{OX1}\}\)\(O\) | 0.9972 | \([\text{CX3}]\{=\text{O}\}\)[\text{OX2H1}] | 0.8703 |
| \([\text{CX3}]\{=\text{OX1}\}\)\(C\)   | 0.9945 | \([\text{CX4H3}]\)\(\#6\) | 0.0499 |
| \([\#6H3]\)\(\#6\)\(\#6\)                  | 0.9752 | \([\#8]\)\(\#6H0\)\(\#6H1\) | 0.8234 |

**Best Positives**

| Structure                  | Prob   | Structure                  | Prob   |
|---------------------------|--------|---------------------------|--------|
| \([\text{CX4H3]}\)                       | 0.9995 | \([\text{CX2H1}]\{=\text{CX2H0}\}\)[\text{CX3H1}]={\#6H0} | 0.0    |
| \([\#8]=\#6]=\#8 \)                  | 0.9982 | \(\text{C}=\text{C}=\text{C}\#\text{C} \)          | 0.0    |
| \([\text{CX3}]\{=\text{OX1}\}\)\(O\)   | 0.9972 | \(\text{C}=\text{C}=\text{C}\#\text{C} \)          | 0.0    |
| \([\text{CX3}]\{=\text{OX1}\}\)\(C\)   | 0.9945 | \(\text{C}=\text{C}=\text{C}\#\text{C} \)          | 0.0    |
| \([\#6H3]\)\(\#6\)\(\#6\)                  | 0.9752 | \([\text{CX3H0}]\{=\text{CX3H2}\}\{=\text{CX4H2}\}\{=\text{CX4H2}\} \)  | 0.0 |
| \([\#6H1]\)                          | 0.9596 | \(\text{C}=\text{C}=\text{C} \)           | 0.0    |
| \([\text{OX2H1}]=\#6\)                   | 0.9491 | \(\text{C}=\text{C}=\text{C} \)           | 0.0    |
| \([\text{CX3}]\{=\text{O}\}\)[\text{OX2H1}] | 0.8703 | \([\#7]\)\(\#6]\{=\#6]\)\(\#6]\={\#7} \)       | 0.0    |
| \([\text{CX4H3}]\)\(\#6\)                 | 0.8499 | \([\#6X3H1]\)\{=\#6X3H1]\)\{=\#6X3H0]\={\#6X3H1} \) | 0.0 |
| \([\#8]\)\(\#6H0\)\(\#6H1\)                  | 0.8234 | \([\text{CX3H0}]\{=\text{CX3H2}\}\{=\text{CX4H3}\}\{=\text{CX4H2}\} \)  | 0.0 |

**Worst Negatives**

| Structure                  | Prob   | Structure                  | Prob   |
|---------------------------|--------|---------------------------|--------|
| \([\#6H1]\)\(\#6H1\)                       | 0.5942 | \(\text{O}=\#6]=\#6]\)\(\#6X3\) \)     | 0.1034 |
| \([\#6X3]\)\(\#6\)\(\#6\)\(\#6X3\)             | 0.5008 | \(\text{CH3}=\text{C}=\text{C}\) \)     | 0.0    |
| \([\#8]\)\(\#6H1]\)\(\#6H1\)                 | 0.3213 | \(\text{O}=\#6]=\#6]\)\(\#6H0\) \)     | 0.3483 |
| \([\text{CX4H4}]\)O                               | 0.2991 | \([\#8]\)\(\#6]=\#6]\)\(\#6X3\) \)     | 0.3989 |
| \([\text{CX4H4}]\)O\(\#6\)\(\#6\)\(\#6\)\(\#6X3\) | 0.2863 | \([\text{CX3H0}]\{=\text{OX1H0}\}\{=\text{OX2H1}\}\)\(\text{CX4H1} \) | 0.5252 |
| \([\#8]\)\(\#6]\)\(\#6\)\(\#6\)\(\#6X3\) | 0.2159 | \(\text{\#6X3]\)\(\#6]\)\(\#6X3\) \)     | 0.5336 |
| \([\#8]\)\(\#6]=\#6H1]\)\(\#6H1\)               | 0.2106 | \(\text{CX4H3}\)\(\text{CX4H1} \)       | 0.7748 |
| \([\text{CH3}]\)\(\#6\)\(\#6\)\(\#6\)\(\#6\)\(\#6X3\) | 0.1993 | \([\#8]=\#6]=\#6]\)\(\#6H1\) \)     | 0.7992 |
| \([\#8]\)\(\#6]=\#6H2\)\(\#6\)\(\#6\)\(\#6\)\(\#6X3\) | 0.1807 | \(\text{O}=\text{\#6X3}\)\(\text{\#6H1} \)     | 0.8114 |
| \([\#8]\)\(\#6]=\#6H1\)\(\#6H1\)\(\#6H3\)\(\#6X3\) | 0.1615 | \([\#8]=\#6]=\#6]\)\(\#6H1\) \)     | 0.8234 |
Example 137 true smiles: COC(=O)C(=O)OC formula: C$_4$H$_6$O$_4$

Index of correct structure: 0 of 502
True structure loss: 0.02006
True structure:

![Chemical structure image]

Experimental $^{13}$C NMR (solvent: CDCl$_3$)

Experimental $^1$H NMR (solvent: CDCl$_3$)

Top predicted structures (loss):

```
Top predicted substructures                   prob
[#8]=[#6][#8]                                 0.966           [CX4H3]                                       0.617
```

```
0.02066                                   0.027254                   0.027367                   0.027367                   0.03017
0.053531                                   0.059798                   0.064389                   0.067154                   0.067883
```
| best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| [#8]=[#6][#8]  | 0.966| [#8]=[#6][#6]=[#8] | 0.0289|
| [OX1H0]=([CX3H0][OX2H0][CX4H3]) | 0.9409| [OX2H1] | 0.4729|
| [CX4H3][OX2H0] | 0.9008| [#6X3][#6X3] | 0.3049|
| [CX3]=([OX1])0 | 0.8793| [#8][#6][#6H2] | 0.2585|
| [#8]=[#6][#6][#8] | 0.6565| [#8][#6][#6][#6X3] | 0.2551|
| [OX1H0]=([CX3H0][OX2H0][CX4H3]) | 0.9409| [OX2H1] | 0.4729|
| [CX4H3][OX2H0] | 0.9008| [#6X3][#6X3] | 0.3049|
| [CX3]=([OX1])0 | 0.8793| [#8][#6][#6H2] | 0.2585|
| [#8]=[#6][#6][#8] | 0.6565| [#8][#6][#6][#6X3] | 0.2551|
| [OX1H0]=([CX3H0][OX2H0][CX4H3]) | 0.9409| [OX2H1] | 0.4729|
| [CX4H3][OX2H0] | 0.9008| [#6X3][#6X3] | 0.3049|
| [CX3]=([OX1])0 | 0.8793| [#8][#6][#6H2] | 0.2585|
| [#8]=[#6][#6][#8] | 0.6565| [#8][#6][#6][#6X3] | 0.2551|
| [OX1H0]=([CX3H0][OX2H0][CX4H3]) | 0.9409| [OX2H1] | 0.4729|
| [CX4H3][OX2H0] | 0.9008| [#6X3][#6X3] | 0.3049|
| [CX3]=([OX1])0 | 0.8793| [#8][#6][#6H2] | 0.2585|
| [#8]=[#6][#6][#8] | 0.6565| [#8][#6][#6][#6X3] | 0.2551|
| [OX1H0]=([CX3H0][OX2H0][CX4H3]) | 0.9409| [OX2H1] | 0.4729|
| [CX4H3][OX2H0] | 0.9008| [#6X3][#6X3] | 0.3049|
| [CX3]=([OX1])0 | 0.8793| [#8][#6][#6H2] | 0.2585|
| [#8]=[#6][#6][#8] | 0.6565| [#8][#6][#6][#6X3] | 0.2551|
| [OX1H0]=([CX3H0][OX2H0][CX4H3]) | 0.9409| [OX2H1] | 0.4729|
| [CX4H3][OX2H0] | 0.9008| [#6X3][#6X3] | 0.3049|
| [CX3]=([OX1])0 | 0.8793| [#8][#6][#6H2] | 0.2585|
| [#8]=[#6][#6][#8] | 0.6565| [#8][#6][#6][#6X3] | 0.2551|
| [OX1H0]=([CX3H0][OX2H0][CX4H3]) | 0.9409| [OX2H1] | 0.4729|
| [CX4H3][OX2H0] | 0.9008| [#6X3][#6X3] | 0.3049|
| [CX3]=([OX1])0 | 0.8793| [#8][#6][#6H2] | 0.2585|
| [#8]=[#6][#6][#8] | 0.6565| [#8][#6][#6][#6X3] | 0.2551|

worst negatives

| worst negatives | prob | worst positives | prob |
|-----------------|------|-----------------|------|
| [OX2H1] | 0.4729 | [#8]=[#6][#6]=[#8] | 0.0289|
| [#8][#6][#6H2] | 0.2505 | [#8][#6][#6]=[#8] | 0.0408|
| [#8][#6][#6]=[#6X3] | 0.2551 | O=CC=O | 0.0622|
| [#6H1] | 0.2511 | [#8][#6]=[#8] | 0.0289|
| [OX1H0]=([CX3H0][OX2H0][CX4H3]) | 0.1924 | [CX3][=([OX1])0] | 0.1869|
| [OX1H0]=([OX1H0][OX2H0][CX3H0]) | 0.1158 | [CX3][=([OX1])0][OX2H0][CX3H0] | 0.2151|
| [OX1H0]=([OX1H0][OX2H0][CX3H0]) | 0.1112 | [OX2H1] | 0.4729|
| [OX1H0]=([OX1H0][OX2H0][CX3H0]) | 0.1097 | [#8][#6][#6][#8] | 0.6565|
| [#8][#6H1][#6H1] | 0.0815 | [CX3][=([OX1])0[OX1H0][OX2H0][CX3H0]] | 0.8793|
| [CX4H3][O][CO] | 0.0758 | [CX4H3][OX2H0] | 0.9008|

------------------------------------------------------------------------------------------------------------------------
Example 138 true smiles: Oclocccncl formula: C5H5NO
Index of correct structure: 0 of 371
True structure loss: 0.01149
True structure:

$$\text{Experimental } ^{13} \text{C NMR (solvent: CDCl}_3\text{)}$$

$$\text{Experimental } ^1\text{H NMR (solvent: CDCl}_3\text{)}$$

Top predicted structures (loss):

| Structure | Loss   |
|-----------|--------|
| ![Structure 1](example.png) | 0.01149 |
| ![Structure 2](example.png) | 0.012264 |
| ![Structure 3](example.png) | 0.016269 |
| ![Structure 4](example.png) | 0.038359 |
| ![Structure 5](example.png) | 0.042365 |
| ![Structure 6](example.png) | 0.051   |
| ![Structure 7](example.png) | 0.055451 |
| ![Structure 8](example.png) | 0.05981 |
| ![Structure 9](example.png) | 0.060799 |
| ![Structure 10](example.png) | 0.07308 |

Top predicted substructures

| Substructure | prob | loss   |
|--------------|------|--------|
| ![Substructure 1](example.png) | 0.9999 | ![Substructure 2](example.png) | 0.9911

[Image of the molecule with OH group and NH group connected to a benzene ring]
| best positives          | prob | best negatives          | prob |
|-------------------------|------|-------------------------|------|
| [#6H1]                  | 0.9999 | [OX2H0][cX3H1][cX3H1]  | 0.0  |
| [#6X3][#6X3]           | 0.9996 | [#7][#6][#6][#6][#6X3] | 0.9807|
| [cH][cH]                | 0.9967 | [#6][#6][#6][#6][#6][#7] | 0.9789|
| [cH]                    | 0.9946 | [#6X3H1][#6X3H0]       | 0.9628|
| worst negatives         | prob | worst positives         | prob |
| [#7][#6H0][#6H1]       | 0.7332 | [OX2H][cX3H1][c]       | 0.347|
| [#7][#6X3H0][#6X3H1]   | 0.7118 | [cH]cO                 | 0.3983|
| [#6X3][#6X3][#6X3]     | 0.3251 | [#8][#6H0][#6H1]       | 0.517|
| [#7X3H2]               | 0.3039 | [cX3H1][#6H0][#6H1]    | 0.6028|
| O=[#6][#6][#6X3]       | 0.2547 | [#6H1][#7][#6H1]       | 0.743|
| [#7H2][#6H0]           | 0.2275 | [#6X3H1][#6X3H0][#6X3H1] | 0.7644|
| [#6X3][#7X3][#6X3]     | 0.1575 | [OX2H1]                | 0.7846|
| [#7H][#6X3H1]          | 0.1489 | [#8][#6][#6][#6X3]     | 0.8555|
| o[cH]                  | 0.1355 | [#6X3][#7][#6X3]       | 0.9162|
| [#7X3H1]               | 0.1295 | [#6H1][#6H1]           | 0.9212|
Example 139 true smiles: O\text{clcc}[nH]ccl formula: C5H5NO
Index of correct structure: 1 of 371
True structure loss: 0.04109
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

| Structure     | Loss       |
|---------------|------------|
| ![Structure 1](image1) | 0.03829   |
| ![Structure 2](image2) | 0.04109   |
| ![Structure 3](image3) | 0.046299  |
| ![Structure 4](image4) | 0.05235   |
| ![Structure 5](image5) | 0.056054  |
| ![Structure 6](image6) | 0.056478  |
| ![Structure 7](image7) | 0.057749  |
| ![Structure 8](image8) | 0.058017  |
| ![Structure 9](image9) | 0.060548  |
| ![Structure 10](image10) | 0.060651  |

Top predicted substructures

| Substructure | prob | Loss |
|--------------|------|------|
| \(#6H1\)     | 0.9992 | 0.7819 |\(#8\)=\(#6H0\)[#6H1] |
| Structure | prob  | Structure | prob  |
|-----------|-------|-----------|-------|
| [6X3]l[6X3] | 0.9049 | [CX3]=[OX1]C | 0.7285 |
| [6X3H1][6X3H0] | 0.8087 | O=[#6][#6][6X3] | 0.6833 |
| [CHX3]=[C]C | 0.8007 | O=[#6][#6]=[6X3] | 0.6584 |
| [6H1][6H1] | 0.7824 | [#7][#6][#6][6X3] | 0.6362 |

**Best positives**

| Structure | prob  | Structure | prob  |
|-----------|-------|-----------|-------|
| [6H1] | 0.9992 | [OX2H0][CX4H2][CX4H2][CX4H1][CX4H1]1 | 0.0 |
| [6X3][6X3] | 0.9849 | [OX2H0][CX4H2][CX4H2][CX4H1][OX2H0] | 0.0 |
| [6X3H1][6X3H0] | 0.8087 | [OX2H0][CX4H2][CX4H1][CX4H1][OX2H0] | 0.0 |
| [6H1][6H1] | 0.7824 | [CX4H0][CX4H2][CX4H2][CX4H2][CX4H0] | 0.0 |
| [#8]=[#6H0][#6H1] | 0.7819 | [OX2H1][CX4H1][CX4H1][CX4H2][CX4H1]1 | 0.0 |
| O[#6][#6]=[6X3] | 0.6833 | [CX4H0][CX4H1][CX4H1][CX4H1][CX4H2] | 0.0 |
| [#7]=#6[6X3] | 0.6362 | [CX4H0][CX4H1][CX4H1][CX4H1][CX4H2] | 0.0 |
| [#6X3][6X3][6X3][6X3] | 0.606 | [OX2H0][CX4H2][CX4H1][CX4H1][CX4H1]1 | 0.0 |
| [#7]=#6[6X3] | 0.549 | [OX2H0][CX4H2][CX4H1][CX4H1]1 | 0.0 |
| [cH][cH] | 0.4907 | [CX4H0][CX4H1][CX4H1][CX4H1][CX4H2] | 0.0 |

**Best negatives**

| Structure | prob  | Structure | prob  |
|-----------|-------|-----------|-------|
| [6X3][6X3][6X3][6X3] | 0.9049 | [CX3]=[OX1]C | 0.7285 |
| [6X3H1][6X3H0] | 0.8087 | O=[#6][#6][6X3] | 0.6833 |
| [CHX3]=[C]C | 0.8007 | O=[#6][#6]=[6X3] | 0.6584 |
| [6H1][6H1] | 0.7824 | [#7][#6][#6][6X3] | 0.6362 |

| Structure | prob  | Structure | prob  |
|-----------|-------|-----------|-------|
| [OX2H0][CX4H2][CX4H2][CX4H1][CX4H1]1 | 0.0 |
| [OX2H0][CX4H2][CX4H2][CX4H1][OX2H0] | 0.0 |
| [OX2H0][CX4H2][CX4H1][CX4H1][OX2H0] | 0.0 |
| [OX2H0][CX4H2][CX4H2][CX4H2][CX4H0] | 0.0 |
| [OX2H1][CX4H1][CX4H1][CX4H2][CX4H1]1 | 0.0 |
| [CX4H0][CX4H1][CX4H1][CX4H1][CX4H2] | 0.0 |
| [CX4H0][CX4H1][CX4H1][CX4H1][CX4H2] | 0.0 |
| [OX2H0][CX4H2][CX4H1][CX4H1][CX4H1]1 | 0.0 |
| [OX2H0][CX4H2][CX4H1][CX4H1][CX4H1]1 | 0.0 |
| [OX2H0][CX4H2][CX4H1][CX4H1][CX4H1]1 | 0.0 |
| [OX2H0][CX4H2][CX4H1][CX4H1][CX4H1]1 | 0.0 |
| [OX2H0][CX4H2][CX4H1][CX4H1][CX4H1]1 | 0.0 |

**Worst negatives**

| Structure | prob  | Structure | prob  |
|-----------|-------|-----------|-------|
| [CHX3]=[C]C | 0.8007 | O=[cX3] | 0.0846 |
| [CX3]==[OX1]C | 0.7285 | [#6][#6][#6][#6][#6][#7]1 | 0.104 |
| O=[#6][#6]=[6X3] | 0.6584 | [cX3H1]([cX3H1]cX3H1) | 0.1165 |
| [#8]=[#6H0][#6H1][#6H1] | 0.5599 | [#6H1][#7][#6H1] | 0.1263 |
| [#7X3H2] | 0.5528 | [#7H][#6X3H1] | 0.1638 |
| [CHX3]=[CHX3] | 0.5333 | [#6X3H1][#6X3H1][#6X3H0][#6X3H1] | 0.1797 |
| O=C[CX3H] | 0.5135 | [cX3H1]([cX3H1]cX3H0) | 0.2238 |
| [CX3H1]==[CX3H1][CX3H0] | 0.4813 | [#7X3H1] | 0.2569 |
| [#X3H1][#X3H1][#X3H1] | 0.4549 | [cH] | 0.2571 |
| [#7]=#6H0[#6H1] | 0.4496 | [#6X3][#7X3][#6X3] | 0.2596 |
Example 140 true smiles: CC(C)=CC(=O)O formula: C5H8O2

Index of correct structure: 0 of 329
True structure loss: 0.015726

True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

| Structure                  | prob   |
|----------------------------|--------|
| [CX4H3]                    | 0.9998 |
| [#6H3][#6][#6]             | 0.9876 |
| #6H3]#6H0 | 0.9993 | [#8]=[#6][#8] | 0.9839 |
| #6H3]#6H0 | 0.9958 | [CX4H3]#6 | 0.9644 |
| #6H3]#6H0 | 0.9956 | [#6X3]#6X3 | 0.9566 |
| #6H3]#6H0 | 0.994 | [CX3]=#OX1]0 | 0.9282 |

**best positives**

| prob | best negatives |
| prob |
| prob |
| prob |
| prob |

| prob | worst negatives | prob |
| prob |
| prob |

| worst negatives | prob | worst positives | prob |
| prob |

| prob | worst negatives | prob |
| prob |
| prob |

| prob | worst negatives | prob |
| prob |
| prob |
Example 141 true smiles: C≡C(CC)(=O)O formula: C5H8O2
Index of correct structure: 0 of 329
True structure loss: 0.012184
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

Top predicted substructures

| Substructure | prob   | Loss                        |
|--------------|--------|-----------------------------|
| [CX4H3][#6] | 0.9998 | [#8]=[#6][#8]               |
|              |        |                             | 0.9953                       |
\[ \text{[CX4H3]} \quad 0.9996 \quad \text{[CX3]}=\text{[OX1]} \quad 0.9943 \]
\[ \text{[CX4H2][\#6][\#6]} \quad 0.9992 \quad \text{[CX3]}=\text{[OX1]} \quad 0.9912 \]
\[ \text{[CX4H3][CX4H2]} \quad 0.9988 \quad \text{[CX4H2][CX3]}=\text{C} \quad 0.9701 \]
\[ \text{[\#6H3][\#6][\#6]} \quad 0.9954 \quad \text{[CX3]}(=\text{O})[\text{OX2H1}] \quad 0.967 \]

**best positives**

| Structure | prob | Structure | prob |
|-----------|------|-----------|------|
| \[ \text{[CX4H3][\#6]} \] | 0.9998 | \[ \text{[CX2H0][\#X][CX2H0][CX2H0]} \] | 0.0 |
| \[ \text{[CX4H3]} \] | 0.9996 | \[ \text{[CX2H0]}[\#X][CX2H1][CX4H2] \] | 0.0 |
| \[ \text{[CX4H2][\#6][\#6]} \] | 0.9992 | \[ \text{[CX2H0]}[\#X][CX2H1][CX4H0] \] | 0.0 |
| \[ \text{[CX4H3][CX4H2]} \] | 0.9988 | \[ \text{CCC}[\text{cc}[\text{C}]] \] | 0.0 |
| \[ \text{[\#6H3][\#6][\#6]} \] | 0.9954 | \[ \text{[\#6H2][\#6][\#6X2]} \] | 0.0 |
| \[ \text{[\#8]=\[\#6][\#8]} \] | 0.9953 | \[ \text{[CX2H0]}[\#X][CX2H1][CX4H1] \] | 0.0 |
| \[ \text{[CX3]=\[OX1]} \] | 0.9943 | \[ \text{[\#6X2][\#6H1][\#6X2]} \] | 0.0 |
| \[ \text{[\#6]=\[OX1][\#6]} \] | 0.9912 | \[ \text{[CX2H0][\#X][CX2H0][CX4H1] \] | 0.0 |
| \[ \text{[CX4H2][\#6X3]} \] | 0.9701 | \[ \text{[CX2H0][\#X][CX2H1][CX2H0] \} | 0.0 |
| \[ \text{[CX3]=\[O][\#X][\#X4H1]} \] | 0.967 | \[ \text{[CX4H1][\#X][CX2H1][CX4H1][CX2H0] \} | 0.0 |

**worst negatives**

| Structure | prob | Structure | prob |
|-----------|------|-----------|------|
| \[ \text{[\#6H1]} \] | 0.6156 | \[ \text{[\#6X3][\#6][\#6][\#6H3]} \] | 0.2924 |
| \[ \text{[CX4H3]=\[C][C]} \] | 0.405 | \[ \text{[CX4H2][CC][C]} \] | 0.4542 |
| \[ \text{[OX1][\#6][\#6X3]} \] | 0.257 | \[ \text{[CX3H0]}[\#X][CX3H2][CX3H2][CX4H2][CX3H0] \} | 0.6438 |
| \[ \text{[\#8]=\[\#6X0][\#6H1]} \] | 0.2322 | \[ \text{[OCC][\#H2]} \} | 0.6192 |
| \[ \text{[\#6X3H1][\#6X3H0]} \] | 0.2003 | \[ \text{[\#6H2][\#6X3H0][\#6H2]} \} | 0.628 |
| \[ \text{[\#6H1][\#6H2]} \] | 0.1959 | \[ \text{[\#6X3H2]} \} | 0.724 |
| \[ \text{[\#6H1][\#6H2]} \] | 0.1929 | \[ \text{[CX3H2][\#X][CX3H0]} \} | 0.7368 |
| \[ \text{[\#8]=\[\#6H1]} \] | 0.1691 | \[ \text{[CX3H2][\#X][CX3H0][\#6][\#6]} \} | 0.7982 |
| \[ \text{[CX4H2][CX3H]} \] | 0.1475 | \[ \text{[CX3H2][\#X][CX3H0][\#X][\#X] \} | 0.802 |
| \[ \text{[\#X][\#X][\#][\#X][\#X][\#X]} \] | 0.1475 | \[ \text{[\#X][\#X][\#X][\#X][\#X][\#X]} \} | 0.8201 |
Example 142 true smiles: CC=C(C(=O)O formula: C5H8O2
Index of correct structure: 0 of 329
True structure loss: 0.010051
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

Top predicted substructures
| Formula                     | Prob | Formula                     | Prob |
|-----------------------------|------|-----------------------------|------|
| [CX4H3][CX3]               | 0.9994 | [#8]=[#6][#8]              | 0.9811 |
| [CX4H3][#6]                | 0.9993 | [#6H3][#6]=[#6][#6H3]      | 0.9632 |
| [CX3](=O)[OX1]C            | 0.9878 | [#6H3][#6]=[#6X3]          | 0.9621 |
| [#6H3][#6H0]               | 0.9868 | [CX3](=O)[OX2H1]           | 0.9621 |
| **best positives**         |      | **best negatives**         |      |
| [CX4H3]                    | 1.0  | [CX2H0][#6X2][#6H1][#6X2]  | 0.0  |
| [CX4H3][CX3]               | 0.9994 | [CX2H0][#6X2][#CX2H0][#CX2H0] | 0.0  |
| [CX4H3][#6]                | 0.9993 | [OX2H0][#CX2H0][CX4H2][CX4H2][CX4H1][CX4H1] | 0.0  |
| [CX3](=O)[OX1]C            | 0.9878 | [CX2H0][#CX2H1][CX4H2]     | 0.0  |
| [#6H3][#6H0]               | 0.9868 | [CX2H0][#CX2H1][CX4H1]     | 0.0  |
| [#6X3][#6X3]               | 0.9855 | [CX2H0][#CX2H1][#CX2H0]    | 0.0  |
| [#8]=[#6][#8]              | 0.9811 | [CX2H0][#CX2H1][CX4H0]     | 0.0  |
| [#6H3][#6]=[#6][#6H3]      | 0.9632 | [CX4H1][#CX4H1][#CX4H1][#CX4H1][CX4H0] | 0.0  |
| [#6H3][#6]=[#6X3]          | 0.9621 | [OX2H0][#CX4H2][CX4H1][#CX4H1][#CX4H1][CX4H1] | 0.0  |
| [CX3](=O)[OX2H1]           | 0.9621 | [CX4H2][#CX4H0][#CX2H0]    | 0.0  |
| **worst negatives**        |      | **worst positives**        |      |
| O=[#6][#6][#6X3]           | 0.4539 | CHX3](=C)C                 | 0.3154 |
| [6X3H1][#6X3H0]            | 0.2332 | CH3CC(OH)                  | 0.4226 |
| [#8][#6]=[#6][#6X3]        | 0.211  | [#6H1]                      | 0.459  |
| [CX3H0][#6X1H0][#OX2H1][CX4H2] | 0.1473 | [6X3H1]==[6X3H0]           | 0.5751 |
| [#8][#6]=[#6][#6H1]        | 0.1394 | [#6X3H1]==[6X3H0]          | 0.6785 |
| [#6X3][#6]=[#6][#6H3]      | 0.1384 | [#6X2][#6][#6H3]           | 0.6814 |
| [#8][#6]=[#6][#6]          | 0.0977 | [CX3H0]==[6X1H0][#OX2H1][CX3H0] | 0.7244 |
| [#8][#6]=[#6][#6X3][#8]    | 0.0905 | [#8][#6X3][#6X3][#6X3][#6H3] | 0.7712 |
| [CX4H2][CC==O]             | 0.0869 | [CX4H1][CX3H1]             | 0.7722 |
| [#8]=[#6][#6H0][#6H1]      | 0.0788 | [CX3H0][#CX3H1][#CX4H1][#CX4H3][#CX3H0] | 0.8203 |
Example 143 true smiles: CCCCC(=O)O formula: C5H10O2
Index of correct structure: 0 of 303
True structure loss: 0.005996
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

Top predicted substructures

| Structure | prob  |
|-----------|-------|
| [CX4H2]{(#6)}{#6} | 0.9999 |
| [CX4H3]{#6} | 0.9967 |
[CH3][O][#6][#6] 0.9994  [O][#6][#6] 0.9963
[CH3][=O][O][#6] 0.999  [O][#6] 0.9944
[CH3][=O][O][#6] 0.9988  [O][#6] 0.9863
[CH3][=O][O][#6] 0.9983  [O][#6] 0.9793

best positives

[CH3][=O][#6][#6] 0.9999  C=CC=C#C  0.0
[CH3][O][O][#6] 0.9999  CCC=CC#C  0.0
[CH3][=O][O][#6] 0.9988  CCC#CC#C  0.0
[CH3][=O][O][#6] 0.9983  CCC#CC#C  0.0
[CH3][O][O][#6] 0.9967  C=CCC#C  0.0
[CH3][=O][O][#6] 0.9963  C=CCC#C  0.0
[CH3][=O][O][#6] 0.9944  [O][#6] 0.0
[CH3][=O][O][#6] 0.9863  [O][#6] 0.0
[O][#6] 0.9793  [O][#6] 0.0

worst negatives

[CH3][=O][O][#6][#6] 0.6204  [O][#6][#6] 0.6338
[CH3][=O][O][#6][#6] 0.4429  [O][#6] 0.7568
[CH3][=O][O][#6][#6] 0.2647  [O][#6] 0.8257
[CH3][=O][O][#6][#6] 0.2375  [O][#6] 0.8315
[CH3][=O][O][#6][#6] 0.236  [O][#6] 0.8508
[CH3][=O][O][#6][#6] 0.1805  [O][#6] 0.9073
[CH3][=O][O][#6][#6] 0.1211  [O][#6] 0.9144
[CH3][=O][O][#6][#6] 0.0927  [O][#6] 0.9252
[CH3][=O][O][#6][#6] 0.0594  [O][#6] 0.9437

------------------------------------------------------------------------------------------------------------------------
Example 144 true smiles: O=C(O)C=CC(=O)O formula: C4H4O4
Index of correct structure: 0 of 301
True structure loss: 0.035775
True structure:

\[
\begin{align*}
&\text{Experimental } 13\text{C NMR (solvent: D2O)} \\
&\text{Experimental } 1\text{H NMR (solvent: D2O)} \\
&\text{Top predicted structures (loss):}
\end{align*}
\]

Top predicted substructures

| Structure | prob  |
|-----------|-------|
| [6x3][6x3] | 0.8107 |

0.035775 0.036655 0.039464 0.041437 0.041872
0.044606 0.0447 0.047843 0.047968 0.049387
Example 145 true smiles: O=C(O)C=CC(=O)O formula: C₄H₄O₄
Index of correct structure: 4 of 301
True structure loss: 0.039178

True structure:

Experimental 13C NMR (solvent: DMSO)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

Top predicted substructures

| Structure | Loss |
|-----------|------|
| #8=[#6][#8] | 0.032792 |
| [CX3](=O)[OX2H1] | 0.033734 |
| [CX3](=O)[OX2H1] | 0.037724 |
| [CX3](=O)[OX2H1] | 0.039167 |
| [CX3](=O)[OX2H1] | 0.039178 |
| [CX3](=O)[OX2H1] | 0.040963 |
| [CX3](=O)[OX2H1] | 0.041943 |
| [CX3](=O)[OX2H1] | 0.042753 |
| [CX3](=O)[OX2H1] | 0.044155 |
| [CX3](=O)[OX2H1] | 0.044979 |
| [CX3](=O)[OX2H1] | 0.07435 |

Top predicted substructures

| Structure | prob |
|-----------|------|
| [#8]=[#6][#8] | 0.9843 |
| (CX3)(=O)(OX2H1) | 0.7435 |
| best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| [#6H1]=[#6][#8] | 0.9843 | CCC#CC#C | 0.0 |
| [#6H1] | 0.9612 | [OX2H0][CX4H2][CX4H1](CX4H1) | 0.0 |
| [OX2H0][CX4H3] | 0.9554 | [CX4H1](NX3H1)(CX4H1) | 0.0 |
| [OX2H0][CX4H2] | 0.9554 | [CX4H1] | 0.0 |
| [OX2H0][CX4H2] | 0.9177 | [CX2H0][CX2H0] | 0.0 |
| [OX2H0][CX4H2] | 0.7641 | [CX4H1] | 0.0 |
| [OX2H0][CX4H2] | 0.7435 | [OX2H0]1[CX4H2][CX4H1] | 0.0 |
| [OX2H0][CX4H2] | 0.707 | [OX2H0][CX4H2] | 0.0 |
| [OX2H0][CX4H2] | 0.707 | [OX2H0][CX4H2] | 0.0 |
| [OX2H0][CX4H2] | 0.707 | [OX2H0][CX4H2] | 0.0 |
| [OX2H0][CX4H2] | 0.707 | [OX2H0][CX4H2] | 0.0 |
| [OX2H0][CX4H2] | 0.707 | [OX2H0][CX4H2] | 0.0 |
| [OX2H0][CX4H2] | 0.707 | [OX2H0][CX4H2] | 0.0 |

| worst negatives | prob | worst positives | prob |
|-----------------|------|-----------------|------|
| O=[#6][#6][#6X3] | 0.5864 | [6X3][#6X3]=[6X3][#6X3] | 0.0226 |
| [OX2H0][CX3H1]=# [O] | 0.4692 | [CX3H0][CX3H1]=[CX3H0][CX3H1] | 0.0487 |
| [OX2H0][CX3H1]=[#] | 0.4523 | [CX3H0]=# | 0.076 |
| [OX2H0][CX3H1]=[#] | 0.3834 | [OX2H0]=# | 0.0949 |
| [OX2H0][CX3H1]=[#] | 0.3786 | [OX2H0]=# | 0.097 |
| [OX2H0][CX3H1]=[#] | 0.3112 | [OX2H0]=# | 0.139 |
| [OX2H0][CX3H1]=[#] | 0.3085 | [OX2H0]=# | 0.148 |
| [OX2H0][CX3H1]=[#] | 0.2756 | [OX2H0]=# | 0.181 |
| [OX2H0][CX3H1]=[#] | 0.2747 | [OX2H0]=# | 0.187 |
| [OX2H0][CX3H1]=[#] | 0.2615 | [OX2H0]=# | 0.222 |
Example 146 true smiles: O=C(O)C≡CC(=O)O
formula: C₄H₄O₄

Index of correct structure: 0 of 301
True structure loss: 0.026882

True structure:

Experimental ¹³C NMR (solvent: CD₃OD)

Experimental ¹H NMR (solvent: D₂O)

Top predicted structures (loss):

Top predicted substructures

- [{#8}=#6][#8] with prob 0.9941
- [CX₃]({'=O'}[OX₂H₁]) with prob 0.8909
|                  | prob  |                  | prob  |
|------------------|-------|------------------|-------|
| [CX3]=[OX1])O   | 0.9839| [CX3]=[OX1])C   | 0.8059|
| [#6X3]=[#6X3]   | 0.9369| O=#6]==#6]==#6X3) | 0.7132|
| [#6H1]          | 0.9104| [#6X3]=[#6X3H0] | 0.7068|
| [OX2H1]         | 0.8964| O=#6]==#6]==#6X3) | 0.6379|
| **best positives** |        | **best negatives** |        |
| [#8]=[#6]=[#8]  | 0.9941| CCC#CC#C         | 0.0   |
| [CX3]=[OX1])O   | 0.9839| [CX4H0]=([NX3H1])([CX4H3])([CX4H2])=[CX4H1] | 0.0   |
| [#6X3]=[#6X3]   | 0.9369| [#7]==[#6]=[#6X2] | 0.0   |
| [#6H1]          | 0.9104| [CX4H2]=[CX4H3][CX2H0] | 0.0   |
| [OX2H1]         | 0.8964| [CX4H1]=([NX3H1])([CX4H3][CX4H2] | 0.0   |
| [CX3]=[OX1])C   | 0.8059| O=#6]==#6]==#6X3) | 0.0   |
| [OX2H1]         | 0.8909| [CX4H2]=([NX3H0][CX4H3] | 0.0   |
| [CX3]=[OX1])C   | 0.7132| [CX4H1][CX4H2][CX4H3] | 0.0   |
| [OX2H1]         | 0.7068| [CX2H0]=[#CX2H0][CX2H0] | 0.0   |
| [#8]==#6]==#6X3] | 0.5765| [CX2H0][#CX2H0][CX2H0] | 0.0   |
| **worst negatives** |        | **worst positives** |        |
| O=#6]==#6]==#6X3) | 0.6379| [#8]==#6]==#6]==#6]==#8] | 0.1644|
| [CX3H0]=([OX1H0])([OX2H1])([CX3H0] | 0.6203| [CX3H1]=([CX3H1])([CX3H0] | 0.179 |
| [#8]==#6]==#6]==#6X3) | 0.3569| [CX3H0][CX3H1][CX3H1][CX3H0] | 0.184 |
| [eH]            | 0.304 | [OX1H0]=([CX3H0][CX3H1][CX3H1] | 0.2162|
| [#8]==#6]==#6]==#6] | 0.2699| [CHX3]=[CHX3] | 0.2722|
| [#8]==#6]==#6] | 0.26 | [#8]==#6]==#6]=[#6H1] | 0.2795|
| [#6X3H1]==#6X3H0] | 0.2421| [#8]=[#6]=[#6H1] | 0.281 |
| [eX3H1]=[eX3H1][eX3H0] | 0.1815| [#6X3]=[#6X3][#6X3] | 0.2926|
| [eH]=eO        | 0.1709| [CHX3]=([eC]C] | 0.3094|
| [#6X3]==[#6X3][#6X3][#6X3] | 0.1728| O=[CX3H0] | 0.3492|

------------------------------------------------------------------------------------------------------------------------
Example 147 true smiles: O=Cc1ccccc1 formula: C7H6O
Index of correct structure: 0 of 261
True structure loss: 0.010196
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

- \(#6H1\) 0.9998
- \(#6X3\)#6X3#6X3#6X3#6X3 0.9415

Top predicted substructures

- \(#6H1\) 0.010196
- \(#6X3\) 0.04115
- \(#6X3\) 0.076293
- \(#6X3\) 0.693436
- \(#6X3\) 0.096076

- \(#6X3\) 0.110224
- \(#6X3\) 0.11064
- \(#6X3\) 0.110902
- \(#6X3\) 0.111885
- \(#6X3\) 0.112736

Top predicted substructures

- \(#6H1\) 0.9998
- \(#6X3\) 0.9415
| Structure | Best Positives | Best Negatives |
|-----------|----------------|----------------|
| `[6x3]` | 0.9992 | [cH][cH] 0.0501 |
| `[CX3]H1` | 0.9967 | O=[6] `[#6][#6][#6x3]` 0.8302 |
| `[6x3H]1` | 0.9905 | `[cX3H1]`(`[cX3H1]`)[`cX3H1]` 0.827 |
| `[#6H1]` | 0.9548 | [CX3][`=[OX1]`]C 0.8128 |

| Structure | Worst Positives | Worst Negatives |
|-----------|----------------|----------------|
| [cX3](`=[OX1]`)C | 0.8128 | [cX3](`=[OX1]`)C 0.5652 |
| O=`[CX3]`H | 0.4753 | `=[#6][#6][#6][#6][#6][#6]` 0.6042 |
| `[#6x3]` | 0.3759 | `[cX3H1]`(`[cX3H1]`)[`cX3H1]` 0.6778 |
| `[cXH3]` | 0.3655 | `[#6x3H]1`(`[#6x3H]1`)[`#6x3H]1` 0.6039 |
| `[#6x3H]1` | 0.2878 | `[cH]` 0.8049 |
| `[#6H]` | 0.2744 | O=`[#6][#6][#6][#6][#6][#6]` 0.8302 |
| `[#6x3]` | 0.2684 | [cH][cH] 0.8501 |
| `[#8]` | 0.2366 | `[#6x3]`(`[#6x3]`)[#6x3]` 0.9415 |
| `[#6H0]` | 0.2219 | `[#6H1]`(`[#6H1]`)[`#6H1]` 0.9548 |
Example 148 true smiles: Nc1ccccc1 formula: C6H7N
Index of correct structure: 0 of 245
True structure loss: 0.008435
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

Top predicted substructures

[1]  

prob  0.9989  [cH]  0.9652
| worst negatives | prob  | best positives | prob  |
|-----------------|-------|----------------|-------|
| [#6X3][#7][#6X3] | 0.4874 | [#7][#6H0] | 0.4249 |
| [#6][#6][#6][#6][#6][#7] | 0.4765 | [#7][#6H0][#6H1] | 0.5986 |
| [#7][#6X3H1] | 0.3176 | [#7][#6X3H0][#6X3H1] | 0.66 |
| [#6][#6X3H1] | 0.2547 | [#7][#6X3H2] | 0.6981 |
| [#6X3H0][#6X3H1][#6X3H1][#6X3H0] | 0.2382 | [#6X3H1][#6X3H1][#6X3H0][#6X3H1] | 0.753 |
| [#6H1][#7][#6H1] | 0.2292 | [#7][#6][#6][#6X3] | 0.9116 |
| [cH][cO] | 0.2147 | [#6X3H1][#6X3H0] | 0.9163 |
| [cX3H1][#6X3H1][#6X3H1] | 0.2034 | [#6H1][#6H1] | 0.9209 |
| [#6X3][#7][#6X3] | 0.18 | [#6][#6][#6][#6][#6] | 0.9426 |
| [cX3H0][#6X3H1][#6X3H1][#6X3H0] | 0.1547 | [#6][#6][#6][#6] | 0.9426 |

| best positives | prob  | best negatives | prob  |
|-----------------|-------|----------------|-------|
| [#6X3][#6X3] | 0.9979 | [#7][#6][#6X3] | 0.9612 |
| [cH][cH] | 0.9964 | [cX3H1][#cX3H1][#cX3H0] | 0.955 |
| [#6X3][#6X3][#6X3][#6X3] | 0.9879 | [#6][#6][#6][#6][#6][#6][#6] | 0.9426 |
| [cX3H1][#cX3H1][#cX3H1][#cX3H0] | 0.9798 | [#6][#6H1] | 0.9209 |
| [cH][cH][cH][cH][cH] | 0.9964 | [#6][#6][#6][#6][#6][#6][#6] | 0.9426 |
| [#6X3][#6X3][#6X3][#6X3] | 0.9879 | [#O][#6X2H1][#C][#4XH2][#C][#4XH1] | 0.0 |
| [cX3H1][#cX3H1][#cX3H1][#cX3H0] | 0.9798 | [#O][#6X2H1][#C][#4XH2][#C][#4XH1] | 0.0 |
| [cH][cH][cH][cH][cH] | 0.9652 | [#C][#4XH0][#N][#3XH1][#C][#4XH3][#C][#4XH2][#C][#4XH1] | 0.0 |
| [#7][#6][#6X3][cX3H1][#cX3H1][#cX3H0] | 0.9612 | [#O][#6X2H1][#C][#4XH2][#C][#4XH1] | 0.0 |
| [#6][#6][#6][#6][#6][#6][#6] | 0.9426 | [#C][#4XH0][#O][#2XH1][#C][#4XH3][#C][#4XH2][#C][#4XH1] | 0.0 |
| [#6H1][#6H1] | 0.9209 | [#O][#6X2H1][#C][#4XH1][#C][#4XH1][#C][#4XH2][#C][#4XH1] | 0.0 |

----------
Example 149: true smiles: CC(CO)(CO)CO
Formula: C5H12O3
Index of correct structure: 0 of 238
True structure loss: 0.024948

True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

Top predicted substructures                   prob
[OX2H1]                                       0.9999          [OH][CX4H]                                    0.8238
| Structure                  | Score  | Structure                  | Score  |
|---------------------------|--------|---------------------------|--------|
| [CH3]CC[OH]               | 0.7833 | [CH3]CC[OH]               | 0.7833 |
| OCC[CH2]                  | 0.7512 | [CH3]CC[OH]               | 0.7833 |
| [CH3]CC[OH]               | 0.7512 | OCC[CH2]                  | 0.7512 |
| [CH3]CC[OH]               | 0.7512 | [CH3]CC[OH]               | 0.7833 |
| [CH3]CC[OH]               | 0.7512 | [CH3]CC[OH]               | 0.7833 |
| [CH3]CC[OH]               | 0.7512 | [CH3]CC[OH]               | 0.7833 |
| [CH3]CC[OH]               | 0.7512 | [CH3]CC[OH]               | 0.7833 |
| [CH3]CC[OH]               | 0.7512 | [CH3]CC[OH]               | 0.7833 |
| [CH3]CC[OH]               | 0.7512 | [CH3]CC[OH]               | 0.7833 |

---

The table above lists the best and worst structures based on certain criteria, with their respective scores.
Example 150 true smiles: OCc1ccccc1, formula: C5H6O2

Index of correct structure: 0 of 226
True structure loss: 0.023588

True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

| Structure  | prob  |
|------------|-------|
| #6H1       | 0.9999|
| #6X3H1[#6X3H0] | 0.8704|
| Structure | Probability | Structure | Probability |
|-----------|-------------|-----------|-------------|
| [#6X3][#6X3] | 0.9911 | [CX4H2][[#6]][#O] | 0.0323 |
| [#6H1][#6H1] | 0.9161 | [#6X3][#6H2][#8] | 0.8345 |
| [#6][#6][#6X3] | 0.9149 | [cH][cH] | 0.8209 |
| [OX2H1] | 0.9142 | [cX3H1][([cX3H1])[cX3H1]] | 0.7764 |

**Best Positives**

| Structure | Probability | Structure | Probability |
|-----------|-------------|-----------|-------------|
| [#6H1] | 0.9999 | [CX2H0][#CX2H1][CX4H0] | 0.0 |
| [#6X3][#6X3] | 0.9911 | [CX4H0][[CX4H2]][[CX4H2]][[CX4H1]][CX4H1] | 0.0 |
| [#6H1][#6H1] | 0.9161 | [CX4H0][[NX3H1]][[CX4H3]][[CX4H2]][CX4H1] | 0.0 |
| [#6][#6][#6X3] | 0.9149 | [CX4H2][[NX3H1]][CX4H3] | 0.0 |
| [OX2H1] | 0.9142 | [CX4H0][[NX3H0]][[CX4H3]][CX4H1] | 0.0 |
| [#6X3][#6H2][#8] | 0.8704 | [CX4H0][[NX3H0]][[CX4H3]][CX4H1] | 0.0 |
| [OX2H1] | 0.8523 | [cX3H0][[NX3H0]][[CX4H3]] | 0.0 |
| [cH][cH] | 0.8209 | [CX4H0][[CX4H3]][CX2H0] | 0.0 |
| [cX3H1][[cX3H1]][cX3H1] | 0.7764 | [CX4H0][[CX4H3]][CX4H0] | 0.0 |

**Worst Negatives**

| Structure | Probability | Structure | Probability |
|-----------|-------------|-----------|-------------|
| [CHX3][=C]C | 0.7159 | [CX4H2][[OX2H1]][cX3H0] | 0.0949 |
| [CHX3][=CHX3] | 0.4380 | [#][#6][#6H2][#8] | 0.1477 |
| [#6H1][#6H2] | 0.4271 | [CX2H1][CX4H2][#6X3H0] | 0.1822 |
| [#6][#6][#6X3] | 0.3879 | [cH] | 0.3261 |
| [CX4H2][CX3]=C | 0.3824 | [#][#6][#6][#8] | 0.3543 |
| [CX3H1][=CX3H1][OX2H0] | 0.3152 | [#][#6][#6H2][#8][6X3] | 0.4084 |
| [OX2H1] | 0.292 | [CX4H0][[OX2H1]][#6X3H0][#8X2H0] | 0.4557 |
| [OX2H1][=CX3H1] | 0.2425 | [#][#6H0][#6H1] | 0.4871 |
| [CX4H0] | 0.2304 | [#][#6][#6H2] | 0.5545 |
| [CX4H2][CX3H] | 0.2241 | [#6X3][#6X3][#6X3][#6X3] | 0.5709 |
Example 151 true smiles: OCc1ccccc1 formula: C5H6O2
Index of correct structure: 0 of 226
True structure loss: 0.017968
True structure:

\[ \text{Experimental } ^{13}\text{C NMR (solvent: CDCl}_3 \text{)} \]

\[ \text{Experimental } ^1\text{H NMR (solvent: CD}_3\text{OD) } \]

Top predicted structures (loss):

| Structure  | prob  |
|------------|-------|
| #6H1       | 0.9995|
| #6X3#6H2#8 | 0.9504|
|     | prob | best positives | prob | best negatives | prob |
|-----|------|----------------|------|----------------|------|
| [#6H1] | 0.9995 | [CX2H0]([CX2H1]) | 0.0 | #6H1 | 0.9995 |
| [CX4H2][#6][O] | 0.9965 | [#6X3H1][#6X3H0] | 0.0473 | #6X3][#6X3] | 0.9931 |
| [OX2H1][CX4H2][#6X3H0] | 0.9632 | [cH][cH] | 0.9207 | [OX2H1] | 0.9605 |
| best positives | prob | best negatives | prob | worst negatives | prob |
| [#8][#6]=[#6X3] | 0.6777 | [#8][#6H][#6X3][#6X3H] | 0.2424 | [#8][#6][#6X3] | 0.524 |
| [CHX3][=C] | 0.4629 | [#8][#6H][#6H1] | 0.2536 | [CHX3][=C] | 0.3372 |
| O[CX4H2][CX3H1] | 0.3193 | [#8][#6][#6H] | 0.3999 | [CX4H2][CX3H]=C | 0.3193 |
| [CX4H2][O][C] | 0.2514 | [#8][#6][#6H1][#6H1] | 0.4312 | [O][CX4H2][CX3H1] | 0.2514 |
| [CHX3][=C][H][C] | 0.2479 | [#8][#6][#6H2][#8] | 0.4949 | [CX4H2][=C] | 0.2479 |
| [OX2H1][#6X3H] | 0.2123 | [cX3H1][cX2H0][cX3H1] | 0.6041 | [OX2H1][=C][X3H1][OX2H0] | 0.2123 |
| worst negatives | prob | worst positives | prob | worst positives | prob |
| [#8][#6]=[#6X3] | 0.6777 | [#8][#6H][#6X3][#6X3H] | 0.2424 | [#8][#6][#6X3] | 0.524 |
| [CHX3][=C] | 0.4629 | [#8][#6H][#6H1] | 0.2536 | [CHX3][=C] | 0.3372 |
| O[CX4H2][CX3H1] | 0.3193 | [#8][#6][#6H] | 0.3999 | [OX2H1][=C][X3H1][OX2H0] | 0.2123 |
| [OX2H1][#6X3H] | 0.2123 | [cX3H1][cX2H0][cX3H1] | 0.6041 | [OX2H1][=C][X3H1][OX2H0] | 0.2123 |
Example 152 true smiles: CC(N(=O)O formula: C3H7NO2
Index of correct structure: 0 of 207
True structure loss: 0.011071
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

Top predicted substructures

[CX4H3] [CX4H1]

0.996 0.9063
| Structure | Best Positives | Prob | Best Negatives | Prob |
|-----------|----------------|------|----------------|------|
| [CX4H3][#6] | 0.9967 | O=[CX3][CX4H] | 0.0844 |
| [#6H3][#6][#6] | 0.9886 | [#7][#6][#6H3] | 0.8753 |
| [CX3][=OX1]C | 0.9771 | [CX3][=OX1]0 | 0.8642 |
| [#8]=[#6][#8] | 0.9273 | [#7][#6][#6X3] | 0.8588 |

| Structure | Worst Positives | Prob | Worst Negatives | Prob |
|-----------|-----------------|------|-----------------|------|
| [CX4H3] | 0.9996 | [CX3H0]=([CX3H1])([CX4H2])=[CX2H0] | 0.0 |
| [CX4H3][#6] | 0.9967 | C=CCCCC#C | 0.0 |
| [#6H3][#6][#6] | 0.9886 | CC=CC#CC | 0.0 |
| [CX3][=OX1]C | 0.9771 | C=CC=CC#C | 0.0 |
| [#8]=[#6][#8] | 0.9063 | [CX2H0]=[CX2H1][OX3H0] | 0.0 |
| O=[CX3][CX4H] | 0.8944 | CCC#CC#C | 0.0 |
| [#7][#6][#6H3] | 0.8753 | [CX2H1][CX2H0][CX3H1]=[CX3H0] | 0.0 |
| [CX3][=OX1]0 | 0.8642 | CC#CCC=C | 0.0 |
| [#7][#6][#6X3] | 0.8558 | CCC#CC#C | 0.0 |
Example 153 true smiles: CC(N)C(=O)O

Index of correct structure: 0 of 207
True structure loss: 0.011341

True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

- Top predicted substructures
  - prob
  - [CX4H3] 0.9993
  - [#8]=[#6H0][#6H1] 0.9175
| Structure  | prob  | Structure  | prob  |
|------------|-------|------------|-------|
| [CX4H3][#6] | 0.9934 | [#7X3H2]  | 0.9148 |
| [#6H3][#6][#6] | 0.9923 | [#8]=[#6][#6] | 0.9043 |
| [CX3](=[OX1])C | 0.985  | [CX4H3][CX4H1] | 0.89  |
| O+[CX3][CX4H] | 0.9442 | [OX2H1] | 0.8655 |

**best positives**

| Structure  | prob  |
|------------|-------|
| [CX4H3] | 0.9993  |
| [CX4H3][#6] | 0.9934  |
| [#6H3][#6][#6] | 0.9923  |
| [OX3](=[OX1])C | 0.985  |
| O+[CX3][CX4H] | 0.9442  |
| [#8]=[#6H0][#6H1] | 0.9175  |
| [CX3H0](=[CX3H1])=[CX3H0][CX3H1][CX3H1][CX3H0] | 0.0  |
| [CX4H3][CX4H1] | 0.89  |

**worst negatives**

| Structure  | prob  |
|------------|-------|
| [#7][#6H0][#6H1] | 0.4468  |
| [#8]=[#6][#6H1][#6H1] | 0.3211  |
| [CX4H1] | 0.2573  |
| [#6X4H3][#6][#8X3H] | 0.1881  |
| [OH][CX4H] | 0.1558  |
| [#6X4H3][#6][#8X3H] | 0.137   |

**prob**

| Structure  | prob  |
|------------|-------|
| [CX4H3] | 0.0148  |
| [OX3][#6] | 0.0   |
| CC=CC#C | 0.0  |
| CCC#CC | 0.0  |
| CCC#CC=C | 0.0  |
| CCC#CC#C | 0.0  |
| CCC=C | 0.0  |
| [CX4H3][CX4H1] | 0.0  |
| [OX2H1] | 0.0  |

**prob**

| Structure  | prob  |
|------------|-------|
| [CX4H3] | 0.4315  |
| [CX4H3][#6] | 0.303  |
| [#6H3][#6][#6] | 0.2511  |
| [OX3](=[OX1])C | 0.2011  |
| O+[CX3][CX4H] | 0.1832  |
| [#8]=[#6H0][#6H1] | 0.1558  |
| [CX3H0](=[CX3H1])=[CX3H0][CX3H1][CX3H1][CX3H0] | 0.0  |
| [CX4H3][CX4H1] | 0.0  |

**prob**

| Structure  | prob  |
|------------|-------|
| [CX4H3] | 0.0148  |
| [OX3][#6] | 0.0   |
| CC=CC#C | 0.0  |
| CCC#CC | 0.0  |
| CCC#CC=C | 0.0  |
| CCC#CC#C | 0.0  |
| CCC=C | 0.0  |
| [CX4H3][CX4H1] | 0.0  |
| [OX2H1] | 0.0  |
Example 154 true smiles: O=CCCC(=O)O  formula: C4H6O3
Index of correct structure: 0 of 195
True structure loss: 0.029319

True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

Top predicted substructures prob [#6H1]  prob [#6H1]  prob [#6H1]  prob [#6H1]  prob [#6H1]  prob [#6H1]
best positives
[CX3H1](=O)[#6] 0.9911 [OX2H1] 0.0726
[CX3H2][{#6}][{#6}] 0.9787 [CX4H2][CC=O] 0.8403
[#{6}][#{6}] 0.927 [CX4H2][CX4H2] 0.8351
[CX3](=OX1)O 0.8876 OCC[CH2] 0.8214

worst negatives
[#{6}X3H1][#{6}X3H0] 0.3981 [CX4H2][CX3H] 0.0117
[OX2H0]CC=O 0.873 [OX2H2][#6H2] 0.0
[CX3H1](=OX1)[#6] 0.9911 [CX2H1]#[CX2H0][CX3H0]=[CX3H0] 0.0
[CX4H2][{#6}][{#6}] 0.9787 CC=CC#C 0.0
[#{6}][#{6}] 0.927 [CX3H1]==[CX3H1][CX2H0] 0.0
[CX3](=OX1)O 0.8876 CCC#CC#C 0.0
[#{6}H1] 0.873 CCC=CC#C 0.0
[OX2H1] 0.8726 [CX2H2][#6H2]=[#6][#{6}][#6X2] 0.0
[CX4H2][CX4H2] 0.8351 [CX2H0][#6H4]==[CX2H1][cX3H0] 0.0
OCC[CH2] 0.8214 CC=CCC#C 0.0
Example 155 true smiles: CCC(=O)C(=O)O formula: C4H6O3
Index of correct structure: 0 of 195
True structure loss: 0.008142
True structure:

Experimental 13C NMR (solvent: DMSO)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

Top predicted substructures

| Substructure | prob |
|--------------|------|
| [CX3](=\[OX1\])C | 0.9999 |
| [OX1H0]=\[CX3H0\]({#6})[CX4H2] | 0.9965 |
best positives  prob  best negatives  prob

[OX1H0]=[CX3H0][CX4H2][CX4H3]  0.9994  [CX4H3][#6]  0.9962
[CX4H2][{#6}][#6]  0.999  [#6H3][#6][#6]  0.995
[CX4H2][{CX4H3}][CX3H0]  0.997  [CX4H3][CX4H2]  0.9906
[CX4H3]  0.9966  [#8]=[#6][#8]  0.9831

worst negatives  prob  worst positives  prob

[OX1H0]=[CX3H0][{#8}][CX4H2]  0.469  OCC[CH2]  0.5116
[#8][#6][#6H2]  0.3429  [CX4H2][CC=O]  0.5538
[#8][#6][#6][#6]=[#8]  0.2988  [OX1H0]=[{OX2H1}][CX3H0]  0.5577
[OX2H0][CX3H0][CX4H2]  0.2184  [#63][#6][#6][#6H3]  0.6751
[OX1H0]=[{#6X3}][#6H2][#6H0]  0.1557  [OX3]==[O]([OX2H1]  0.6898
[#8][#6][#6][#6X3]  0.1495  [#8][#6][#6]=[#8]  0.7467
[CX3H0]=[{OX1H0}][{OX2H1}][CX4H2]  0.1464  O=CC=O  0.8112
[#63][#6H2][#6X3]  0.0971  [#8]=[#6][#6]=[#8]  0.8589
CCCCCC  0.0713  [OX2H1]  0.8999
Example 156 true smiles: CC(=O)CC(=O)O formula: C4H6O3
Index of correct structure: 0 of 195
True structure loss: 0.015394
True structure:

![Molecule Structure]

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

1. \[\text{CX}_3\](\text{=OX}_1)\text{C}\] prob 1.0 [\text{OX}_1\text{H}_0]=[\text{CX}_3\text{H}_0][\text{CX}_4\text{H}_3] prob 0.9074
2. \[\text{CX}_3\]OC\] prob 0.015394
3. \[\text{OX}_1\text{H}_0\] prob 0.050129
4. \[\text{OX}_1\text{H}_0\] prob 0.063121
5. \[\text{OX}_1\text{H}_0\] prob 0.079611
6. \[\text{OX}_1\text{H}_0\] prob 0.08312
7. \[\text{OX}_1\text{H}_0\] prob 0.087835
8. \[\text{OX}_1\text{H}_0\] prob 0.088136
9. \[\text{OX}_1\text{H}_0\] prob 0.094407
10. \[\text{OX}_1\text{H}_0\] prob 0.097968
11. \[\text{OX}_1\text{H}_0\] prob 0.106133
| Structure                        | prob  |
|---------------------------------|-------|
| [CX4H3]                         | 0.997 |
| [CX4H3][CX3]                   | 0.9862|
| [CX4H3][CX3H0]                 | 0.978 |
| [CX4H2][{CX3H0}][CX3H0]       | 0.9753|
| [CX3H0][OX2H0]                 | 0.4627|
| [CXH2][{CX3H0}][CX3H0]        | 0.3838|
| [OX1H0][{OX2H1}][CX4H2]       | 0.2683|
| [OX2H1]                        | 0.1744|

**Best positives**

| Structure                        | prob  |
|---------------------------------|-------|
| [CX3H0][{OX1H0}][{CX4H3}][CX4H2] | 0.6666|
| [CX3H0][{OX1H0}]                | 0.5715|
| [OX1H0][{OX2H1}]                | 0.5936|
| [OX2H1]                        | 0.687 |

**Best negatives**

| Structure                        | prob  |
|---------------------------------|-------|
| [CX3H0][{OX1H0}][{CX4H3}][CX4H2] | 0.3529|
| [CX3H0][{OX1H0}]                | 0.3211|
| [OX1H0][{OX2H1}]                | 0.5449|
| [OX2H1]                        | 0.5715|

| Structure                        | prob  |
|---------------------------------|-------|
| [CX3H0][{OX1H0}][{CX4H3}][CX4H2] | 0.5888|
| [OX1H0][{OX2H1}]                | 0.5671|
| [OX2H1]                        | 0.5936|

**Worst positives**

| Structure                        | prob  |
|---------------------------------|-------|
| [CX3H0][{OX1H0}][{CX4H3}][CX4H2] | 0.6666|
| [OX1H0][{OX2H1}]                | 0.5715|
| [OX2H1]                        | 0.5936|

| Structure                        | prob  |
|---------------------------------|-------|
| [CX3H0][{OX1H0}][{CX4H3}][CX4H2] | 0.687 |
| [OX1H0][{OX2H1}]                | 0.687 |

**Worst negatives**
Example 157 true smiles: CCCC(=O)CC formula: C6H12O
Index of correct structure: 0 of 193
True structure loss: 0.004278
True structure:

Experiment 13C NMR (solvent: CDCl3)

Experiment 1H NMR (solvent: D2O)

Top predicted structures (loss):

| Structure | Loss  |
|-----------|-------|
| [CX4H3][CX4H2] | 0.004278 |
| [CX4H3][#6] | 0.043035 |
| [CX4H3] | 0.063048 |
| [CX4H3][#6] | 0.068929 |
| [CX4H3][#6] | 0.08883 |

| Structure | Loss  |
|-----------|-------|
| [CX4H3][CX4H2] | 0.097767 |
| [CX4H3][CX4H2] | 0.103281 |
| [CX4H3][CX4H2] | 0.110448 |
| [CX4H3][#6] | 0.113741 |
| [CX4H3][#6] | 0.117429 |

Top predicted substructures

| Substructure | prob | Loss  |
|--------------|------|-------|
| [CX4H3][#6] | 1.0  | 0.9997 |
Example 158 true smiles: CC(=O)C(C)(C)C formula: C6H12O
Index of correct structure: 0 of 193
True structure loss: 0.007709
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

0.007709  0.056767  0.065416  0.07255  0.074018

0.074218  0.074584  0.089033  0.099302  0.103097

Top predicted substructures

[6H3][6][6] prob 0.9999 [CX4H3][CX4H0][CX4H3] 0.9969
| Structure | prob | Structure | prob |
|-----------|------|-----------|------|
| [CX3](=[OX1])C | 0.9995 | [CX4H3][CX4H0] | 0.0915 |
| [CX4H3] | 0.9993 | [CX4H3][CX3] | 0.9662 |
| [CX4H3][#6] | 0.9987 | [CX3H0](=[OX1H0])([CX4H3])[CX4H0] | 0.9079 |
|[#6H3][#6H0] | 0.9984 | [CX4H3][CX3H0] | 0.8839 |
| **best positives** | | **best negatives** | |
| [#6H3][#6][#6] | 0.9999 | CCC#CC#C | 0.0 |
| [CX1](=[OX1])C | 0.9995 | C=CC=CC#C | 0.0 |
| [CX4H3] | 0.9993 | [CX3H0](=[CX3H1])([OX2H0])[CX2H0] | 0.0 |
| [CX4H3][#6] | 0.9987 | [CX2H0][#CX2H1][CX3H0] | 0.0 |
| [#6H3][#6H0] | 0.9984 | CC=CCC#C | 0.0 |
| [CX4H3][CX4H0][CX4H3] | 0.9969 | [#7][#6]=[#6][#6][#7] | 0.0 |
| [CX4H3][CX3] | 0.9915 | C=CC=CC=CC | 0.0 |
| [CX4H3][CX3H0] | 0.9907 | [CX2H1][#CX2H0][CX3H1][#CX3H0] | 0.0 |
| [CX4H3][CX4H0] | 0.8839 | [CX2H0][#CX2H1][CX2H0] | 0.0 |
| **worst negatives** | | **worst positives** | |
| [#6H1] | 0.5581 | [#6H3][#6][#6][#6H3] | 0.3797 |
| [CX4H2][CX3]=O | 0.5449 | [#6H3][#6][#6X3] | 0.7535 |
| [#6X3][#6][#6][#6H3] | 0.4247 | [OX1H0]=[CX3H0][CX4H3] | 0.8426 |
| [CX4H3][CX4H1] | 0.3629 | [OX1H0]=[CX3H0][CX4H0][CX4H3] | 0.8432 |
| [#8]=[#6H0][#6H1] | 0.2848 | [CX4][([CX4H3])([CX4H3])][CX4H3] | 0.8489 |
| [CX4H2][CC]=O | 0.2573 | [CX4H3][CX3H0] | 0.8839 |
| [CX4H1][([CX4H3])([CX4H3])][CX3H0] | 0.2202 | [CX3H0][#OX1H0][([CX4H3])[CX4H0] | 0.9079 |
| On=[CX3][CX4H] | 0.1856 | [CX4H3][CX3] | 0.9662 |
| On=[#6][#6][#6H0] | 0.1538 | [CX4H3][CX4H0] | 0.9915 |
| [CHX4][([CH3X4])][CH3X4] | 0.1528 | [CX4H3][CX4H0][CX4H3] | 0.9969 |
Example 159 true smiles: CCCCC=O formula: C6H12O
Index of correct structure: 0 of 193
True structure loss: 0.006741
True structure:

Experimental $^{13}$C NMR (solvent: CDCl$_3$)

Experimental $^1$H NMR (solvent: CDCl$_3$)

Top predicted structures (loss):

Top predicted substructures

| Substructure | prob  |
|--------------|-------|
| [CX3H1](=O)[#6] | 1.0   |
| [CX4H3]      | 0.9973|
Example 160 true smiles: CC(=O)CC(C)C formula: C6H12O
Index of correct structure: 0 of 193
True structure loss: 0.009541
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

```
| Structure | prob   |
|-----------|--------|
| [OX1H0]=[CX3H0][CX4H3] | 0.009541 |
| [OX1H0][#6H3][#6][#6] | 0.043412 |
| [OX1H0][#6H3][#6][#6] | 0.047147 |
| [OX1H0][#6H3][#6][#6] | 0.070068 |
| [OX1H0][#6H3][#6][#6] | 0.076053 |
| [OX1H0][#6H3][#6][#6] | 0.083305 |
| [OX1H0][#6H3][#6][#6] | 0.09189  |
| [OX1H0][#6H3][#6][#6] | 0.096687 |
| [OX1H0][#6H3][#6][#6] | 0.099413 |
```
| Formula                          | Best Positives | Best Negatives | Worst Negatives | Worst Positives |
|---------------------------------|----------------|----------------|-----------------|-----------------|
| [CX4H3]                         | 1.0            | [CX4H3][CX3]   | 0.0981          | [CX4H2][CX4H2]  |
| [CX3]=([OX1])C                  | 0.9996         | [#6H3][#6H0]   | 0.9626          | [CX4H2][CX4H2]  |
| [CX4H3][#6]                     | 0.9993         | [CX3H0]=([OX1H0])([CX4H3])|[CX4H2]| 0.9611          | [CX4H2][#6H0]  |
| [CX4H3][CX3H0]                  | 0.9983         | [CX4H1]([CX4H3])([CX4H3])|[CX4H2]| 0.9469          | [CX4H2][#6H3]  |
|                                  |                |                |                 |                 |
| best positives                  | prob           | best negatives | prob            | worst positives |
| [#6H3][#6][#6]                  | 1.0            | [CX2H0]=([CX2H1])|[CX3H0] 0.0          | [#6H1][#6H2] 0.4289 |
| [CX4H3]                         | 1.0            | C=CC=CC#C       0.0          | [CX4H2][CX4H2] 0.4298 |
| [CX3]=([OX1])C                  | 0.9996         | CCC=CC#C        0.0          | [#6X3][#6H6][#6H0] 0.4424 |
| [CX4H3][#6]                     | 0.9993         | CCC=CC#C        0.0          | [#6X3][#6X3][#6H0] 0.4424 |
| [CX4H3][CX3H0]                  | 0.9983         | [CX2H0]=([CX2H1])|[CX3H0] 0.0          | [#6X3][#6X3][#6H0] 0.4424 |
| [OX1H0]=([CX4H3])|[CX4H3])|[CX4H2]| 0.1103 | [CX4H1]([CX4H3])|[CX4H3] 0.5663 |
| [CX4H2][CX4H2]                  | 0.11           | O=[CX3H0][CX4H2][CX4H1] 0.4758 |
| [OX1H0]=([CX3H0])|[CX4H2])|[CX4H0] 0.0946 | [#6H1] 0.588 |
| [#6X3][#6][#6][#6H3]            | 0.1103         | [CH4][([CH3X4][CH2X4]) 0.5663 |
| [OX1H0]=([CX3H0])|[CX4H2][CX4H0] 0.0846 | [CH4][([CH3X4][CH3X4]) 0.7787 |
| [CX4H3]                         | 0.0749         | [#6X3][#6][#6][#6H3] 0.7862 |
| O=[CX3H0][CX4H2][CX4H2]         | 0.0741         | [CX4H2][CX3]=O 0.8085 |

------------------------------------------------------------------------------------------------------------------------
Example 161 true smiles: CCCCC(C)=O formula: C6H12O
Index of correct structure: 0 of 193
True structure loss: 0.003486
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

Top predicted substructures

[CX4H3] 0.9998
[CX4H3][CX4H2] 0.9985
| best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| [CX4H3]        | 0.9998 | [CX2H0]#([CX2H1])[CX3H0] | 0.0 |
| [CX3](=[OX1])C | 0.9998 | CC=CC#C          | 0.0 |
| [#6H3][#6][#6] | 0.9996 | CCC=CC#C         | 0.0 |
| [CX4H2]([#6])[#6] | 0.9995 | [CX2H0]#([CX2H0])[CX2H0] | 0.0 |
| [OX1H0]=[CX3H0][CX4H3] | 0.9953 | [CX2H1]#([CX2H0])[CX3H1]=[CX3H0] | 0.0 |
| [CX3H0](=[OX1H0])([CX4H3])[CX4H2] | 0.9871 | [CX3H0]([OX2H0])[CX2H0] | 0.0 |
| worst negatives | prob | worst positives | prob |
| [#6X3][#6][#6][#6H3] | 0.2448 | [CX4H2]([CX4H2])[CX4H2] | 0.752 |
| [CX4H2][CX4H2][CX4H2][CX4H2] | 0.2329 | CCC#CC          | 0.7559 |
| [#6H1][#6H2] | 0.1905 | [CX4H2][CX4H2] | 0.7812 |
| [#6H3][#6][#6X3] | 0.1187 | [CX4H2][CX4H2] | 0.7812 |
| [#6H1] | 0.1172 | [CX4H2][CX3]=0 | 0.9109 |
| [#6H3][#6][#6][#6H3] | 0.1162 | [#6H3][#6XH0][#6H2] | 0.9166 |
| [#6H2][#6XH0][#6H2] | 0.0517 | [CX4H2][CX4H2][CX3H0] | 0.9218 |
| [OX1H0]=[CX3H1][CX4H2][CX4H2] | 0.0493 | [#6H3][#6H0] | 0.9492 |
| [#8]=[#6H0][#6H1] | 0.0469 | [CX4H2]([CX4H3])[CX4H2] | 0.9667 |
| [CX4H3][CX4H1] | 0.0448 | O=[CX3H0][CX4H2][CX4H2] | 0.9765 |
Example 162 true smiles: CCCCC1C01 formula: C6H12O
Index of correct structure: 0 of 193
True structure loss: 0.010964
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

Top predicted substructures

| Substructure          | Prob |
|-----------------------|------|
| [CX4H3]               | 0.9999 |
| [CX4H3][#6]           | 0.9924 |
| best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| [#6H3][#6][#6] | 0.9994 | [#6H1] | 0.9904 |
| [CX4H2][#6][#6] | 0.9993 | [#6H2][#6H1r3] | 0.9852 |
| [CX4H3][CX4H2] | 0.9983 | [CX4H]O | 0.985 |
| [OX2H0][CX4H2][CX4H1] | 0.9978 | C10C1 | 0.9811 |
| best positives | prob | best negatives | prob |
| [CX4H3] | 0.9999 | [CX2H1[#CX2H0][CX3H1]=[CX3H0]] | 0.0 |
| [#6H3][#6][#6] | 0.9994 | [CX2H0][#CX2H1][CX3H0] | 0.0 |
| [CX4H2][#6][#6] | 0.9993 | C=C=[CX4H4] | 0.0 |
| [CX4H3][CX4H2] | 0.9983 | C=C=C=C=C | 0.0 |
| [OX2H0][CX4H2][CX4H1] | 0.9978 | C=C=C=C=C | 0.0 |
| [CX4H3][#6] | 0.9924 | [CX3H0][#CX3H1][#OX2H0][CX2H0] | 0.0 |
| [#6H1] | 0.9904 | [CX2H0][#CX2H0][CX2H0] | 0.0 |
| [#6H2][#6H1r3] | 0.9852 | [CX4H3][#X3H0] | 0.0 |
| [CX4H]O | 0.985 | [#7][#6][#6][#7] | 0.0 |
| C10C1 | 0.9811 | [CX3H0][#CX3H1][#OX2H2][CX2H0] | 0.0 |
| worst negatives | prob | worst positives | prob |
| [CX4H2][CX4H3][CX4H1] | 0.7099 | [OX2H0][CX4H1][CX4H2][CX4H2] | 0.4635 |
| [#8][#6H1][#6H1] | 0.4617 | [CX4H4][CX4H2][CX4H2] | 0.5669 |
| [#6H1][#6H1] | 0.4114 | [#6H1][#6H2][#6H2] | 0.5649 |
| [CX4H2][CX4H2][CX4H2][OX2H1] | 0.3018 | [CX4H2][CX4H2][CX4H2] | 0.607 |
| O[CX4H][CX4H2][CX4H1] | 0.3017 | [CX4H2][CX4H2][CX4H1] | 0.6125 |
| [#6H2][#6H1][#6H1][#6H2] | 0.2298 | [CX4H2][CX4H3][CX4H2] | 0.639 |
| [OX2H1] | 0.2169 | [OX2H0][OX2H0][CX4H1][CX4H2] | 0.7261 |
| [#6X4H1][#6X4H1][#6X4H1] | 0.1329 | [CX4H2][#OX2H0][CX4H1] | 0.7728 |
| [CX4H1][OX2H0][CX4H2][CX4H1] | 0.1243 | [CX4H2][O][CX4] | 0.7802 |
| [CX4H2][CH][CH] | 0.1047 | [CX4H2][CX4H2] | 0.7905 |
Example 163 true smiles: CCC(=O)C(C)C formula: C₆H₁₂O
Index of correct structure: 0 of 193
True structure loss: 0.011109
True structure:

Experimental ¹³C NMR (solvent: CDCl₃)

Experimental ¹H NMR (solvent: D₂O)

Top predicted structures (loss):

Top predicted substructures                   prob
[OX1H0]=[CX3H0][CX4H2][CX4H3]                 0.9893
Example 164 true smiles: COC(CN)OC formula: C4H11NO2
Index of correct structure: 0 of 181
True structure loss: 0.005701
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

Top predicted substructures prob
[OX2H0][CX4H1][OX2H0] 0.9996
[CX4H3] 0.9748
| Structure                      | prob  | Structure                      | prob  |
|--------------------------------|-------|--------------------------------|-------|
| [7H2][6H2]                    | 0.9969| [7X3H2]                        | 0.0738|
| [7][6H2]                      | 0.9935| [CX4H3][OX2H0]                 | 0.9611|
| [7X3][6H2]                    | 0.9909| [7][6H2][6H1]                  | 0.8323|
| [CX4H1][OX2H0][OX2H0][CX4H2] | 0.9851| [8][6][6H2]                    | 0.8168|
| best positives                |       | best negatives                 |       |
| [OX2H0][CX4H1][OX2H0]         | 0.9996| [6X2][6H1][6X2]                | 0.0   |
| [7H2][6H2]                    | 0.9969| CC=CCC#C                       | 0.0   |
| [7][6H2]                      | 0.9935| [CX2H1][CX2H0][CX3H1][CX3H0]  | 0.0   |
| [7X3][6H2]                    | 0.9909| [OX1H0][CX3H0][CX2H0][CX2H1]  | 0.0   |
| [CX4H1][OX2H0][OX2H0][CX4H2] | 0.9851| CCC#GCC#C                      | 0.0   |
| [CX4H3]                       | 0.9748| [CX2H0][CX2H1][CX4H2]          | 0.0   |
| [7X3H2]                       | 0.9738| [OX2H0][OX2H0][CX4H2]          | 0.0   |
| [CX4H3][OX2H0]                | 0.9611| [CX2H0][CX2H1][CX2H0]          | 0.0   |
| [7][6H2][6H1]                 | 0.8323| [CX2H0][CX2H1][CX2H0]          | 0.0   |
| [8][6][6H2]                   | 0.8168| [CX2H0][CX2H1][CX2H0]          | 0.0   |
| worst negatives               |       | worst positives                |       |
| OCC[CH2]                      | 0.3533| [#6H1][6H2]                    | 0.581 |
| [CX4H2][NX3H2][CX4H2]         | 0.3351| [#6H1]                         | 0.6654|
| [CX4H2][CX4H2]                | 0.1797| [CX4H2][NX3H2][CX4H2]          | 0.7087|
| [#6H2][7][6H2]                | 0.1736| [OX4H][OX4H2]                  | 0.7146|
| [CX4H2][6][6H2]               | 0.1499| [CX4H][O]                      | 0.775 |
| [7X3H1]                       | 0.1424| [#6][6][6H2]                   | 0.8168|
| [6H2][8][6H1]                 | 0.1335| [7][6H2][6H1]                  | 0.8323|
| [7][6H2][6H2]                 | 0.1166| [CX4H3][OX2H0]                 | 0.9611|
| [CX4H3][OX2H0][CX4H2]         | 0.0628| [7X3H2]                        | 0.9738|
| [CX4H2][O][CX4]               | 0.0599| [CX4H3]                        | 0.9748|
|------------------------------------------------------------------------------------------------------------------------|
Example 165 true smiles: OCCNCCO formula: C₄H₁₁NO₂

Index of correct structure: 0 of 181
True structure loss: 0.008394

True structure:

![Chemical Structure Image]

Experimental ¹³C NMR (solvent: CDCl₃)

Experimental ¹H NMR (solvent: D₂O)

Top predicted structures (loss):

1. ![Predicted Structure 1](image) 0.008394
2. ![Predicted Structure 2](image) 0.019748
3. ![Predicted Structure 3](image) 0.020764
4. ![Predicted Structure 4](image) 0.025876
5. ![Predicted Structure 5](image) 0.027609

Top predicted substructures

| Substructure | prob | prob |
|--------------|------|------|
| [#7]#6H2     | 0.9767 | CX₄H₂{{OX2H1}}CX₄H₂ | 0.8669 |
[OX2H1] 0.9684 [O][#6H2] 0.0593
[#8][#6][#6H2] 0.9306 [CH2X4][O][CX4H2] 0.7909
[CX4H2][(#6)][O] 0.9222 [CX4H2][CX4H2] 0.7214
[#7X3][#6H2] 0.9003 [#6H2][#7][#6H2] 0.6729

best positives prob best negatives prob
[#7][#6H2] 0.9767 [CX2H0][#CX2H1][CX3H0] 0.0
[OX2H1] 0.9684 CCC=CC#C 0.0
[#8][#6][#6H2] 0.9306 CC=CC#CC 0.0
[CX4H2][(#6)][O] 0.9222 C=CC=CC#C 0.0
[#7X3][#6H2] 0.9003 [CX3H0]=([CX3H1])[CX4H2][CX2H0] 0.0
[CX4H2][OX2H1][CX4H2] 0.8669 CCC#CC=C 0.0
[#7][#6H2][#6H2] 0.8593 [CX2H1][#CX2H0][CX3H1]=[CX3H0] 0.0
[CH2X4][O][CX4H2] 0.7909 C=CCC#C 0.0
[CX4H2][CX4H2] 0.7214 [CX3H0]=[CX3H1][CX4H1][CX2H0] 0.0
[#6H2][#7][#6H2] 0.6729 [#6X2][#6H1][#6X2] 0.0

worst negatives prob worst positives prob
[#6H1] 0.3835 [#7X3H1] 0.247
OCC[CH2] 0.3122 [CX4H2][NX3H1][CX4H2] 0.4356
[#7X3H2] 0.2735 [#6H2][#7][#6H2] 0.6729
[CX4H2][NX3H0][CX4H2] 0.2625 [CX4H2][CX4H2] 0.7214
[#7X3H0] 0.2549 [CH2X4][O][CX4H2] 0.7909
[#7][#6H2][#6H1] 0.1802 [#7][#6H2][#6H2] 0.8593
[#6H1][#6H2] 0.1761 [CX4H2][OX2H1][CX4H2] 0.8669
[CX4H2][OX2H1][CX4H1] 0.1491 [#7X3][#6H2] 0.9003
[#8][#6H1][#6H1] 0.131 [CX4H2][(#6)][O] 0.9222
[#7H2][#6H1] 0.107 [#8][#6][#6H2] 0.9306

=================================================================

Example 166 true smiles: CC(N)(CO)CO formula: C4H_{11}NO_{2}

Index of correct structure: 0 of 181
True structure loss: 0.016605
True structure:
best positives

| [OX2H1] | 0.9986 | [O7X3H2] | 0.9377 |
| [CX4H3] | 0.998 | [CX4H2]{[OX2H1]}{CX4H0} | 0.9033 |
| [6H3][6H0] | 0.9711 | [CH3]{CC}{OH} | 0.8724 |
| [CX4H3]{CX4H0} | 0.9568 | [6H1] | 0.7673 |

worst negatives

| [6H1] | 0.7673 | [6H1][6H2][6H0][6H2][8] | 0.2105 |
| [OH]{CX4H} | 0.7577 | [6X3H3][6H0][6H2][6H2] | 0.3025 |
| [7X2H2] | 0.554 | [7X2H2][6H0] | 0.3644 |
| [CX4H2]{[OX2H1]}{CX4H1} | 0.4928 | OCC{C}R2 | 0.5283 |
| [CX4H0] | 0.4697 | [CX4H2]{{6}}{0} | 0.5446 |
| [6H1][6H2][6H2] | 0.2533 | [CH3]{CC}{OH} | 0.8724 |
| [7X3H1] | 0.252 | [CX4H2]{[OX2H1]}{CX4H0} | 0.9033 |
| [7X3][6H2] | 0.2324 | [7X3H2] | 0.9377 |
| [6][6][6][6] | 0.2322 | [CX4H3]{6} | 0.9544 |
Example 167 true smiles: \(	ext{CC(=O)CC(O)O}\) formula: C_4H_8O_3

Index of correct structure: 0 of 172
True structure loss: 0.019179
True structure:

Experimental 13C NMR (solvent: CDCl₃)

Experimental 1H NMR (solvent: D₂O)

Top predicted structures [loss]:

- \([\text{CX4H3}]\) 1.0
- \([\text{CX4H3}]\)#6 0.995

Top predicted substructures

\[
\begin{array}{c}
\text{prob} \\
0.019179 \\
0.060651 \\
0.061101 \\
0.065334 \\
0.068097 \\
0.075476 \\
0.084514 \\
0.086433 \\
0.087185 \\
0.093735
\end{array}
\]
| Structure | Probabilities | Structure | Probabilities |
|-----------|---------------|-----------|---------------|
| \([#6H3]|[#6]|[#6]|\) | 0.9996 | \([#6]|[#6]|[#6]|\) | 0.9888 |
| \([CX3]|=#|[OX1]|\)C | 0.9988 | \([#8]|=#|[#6]|[#8]|\) | 0.9872 |
| \([OX2]|H|\) | 0.9985 | \([CX4]|H|O|\) | 0.9811 |
| \([CX4]|H|2|[#6]|[#6]|\) | 0.9971 | \([CX3]|=#|[OX1]|\)O | 0.9754 |

**Best Positives**

| Structure | Probabilities |
|-----------|---------------|
| [CX4]|H|3 | 1.0 |
| \([#6]|H|3]|[#6]|[#6]|\) | 0.9996 | \([#8]|=#|[#6]|[#8]|\) | 0.9872 |
| \([CX3]|=#|[OX1]|\)C | 0.9988 | \([CX4]|H|2|([#6]|[#6]|[#6]|\) | 0.9971 |
| \([#6]|H|1|\) | 0.9888 | \([CX4]|H|2|([#6]|[#6]|\) | 0.995 |

**Worst Negatives**

| Structure | Probabilities |
|-----------|---------------|
| \([#8]|=#|[#6]|[#6]|\) | 0.6222 | \([CX3]|H|0|([#6]|H|0|)#|[#6]|H|1|\] | 0.5622 |
| \([CX4]|H|0|([#6]|H|0|)#|[#6]|H|1|\] | 0.5987 | \([OX2]|H|1|([#6]|H|0|)#|[#6]|H|1|\] | 0.5622 |

**Worst Positives**

| Structure | Probabilities |
|-----------|---------------|
| \([#8]|=#|[#6]|H|1|\] | 0.4505 | \([CX4]|H|1|([#6]|H|1|)#|[#6]|H|1|\] | 0.4505 |
| \([#8]|=#|[#6]|H|1|\] | 0.4505 | \([#6]|H|1|([#6]|H|1|)#|[#6]|H|1|\] | 0.4505 |

| Structure | Probabilities |
|-----------|---------------|
Example 168 true smiles: O=C1C=CC(=O)C=C1 formula: C6H4O2
Index of correct structure: 0 of 160
True structure loss: 0.018842
True structure:

Experimental 13C NMR (solvent: CDC13)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

Top predicted substructures
[#6X3][#6X3] prob 0.9888 [#6X3][#6X3][#6X3]=[#6X3] 0.8337
| Structure | prob | Structure | prob |
|-----------|------|-----------|------|
| O=[#6][#6]=[#6X3] | 0.9266 | #6X3H1][#6X3H0] | 0.0331 |
| [CX3]=[#OX1]C | 0.9264 | #OX3H0][CX3H0][CX3H1]=[CX3H1] | 0.0819 |
| [#8]=[#6][#6]=[#6][#6]=[#8] | 0.8614 | [#6H1] | 0.8038 |
| O=[#CX3H] | 0.8553 | #CX3H0][CX3H1]=[CX3H1][CX3H0] | 0.774 |

**best positives**

| Structure | prob | Structure | prob |
|-----------|------|-----------|------|
| O=[#6][#6]=[#6X3] | 0.9888 | #6H3][#6X4][#6H3] | 0.0 |
| [#6X3][#6X3] | 0.9264 | CX3H2][N3H1][CX4H3] | 0.0 |
| [CX3]=[#OX1]C | 0.9264 | CX4H1][N3H1][CX4H3][CX4H2] | 0.0 |
| [#8]=[#6][#6]=[#6][#6]=[#8] | 0.8614 | CX4H1][N3H0][CX4H3][CX4H1] | 0.0 |
| O=[#CX3H] | 0.8553 | CX4H1][N3H2][CX4H3][CX4H1] | 0.0 |
| [#6X3][#6X3][#6X3]=[#6X3] | 0.8337 | #6H1][#6H1][#6H1] | 0.0 |
| [OXH0][#CX3H0][CX3H1]=[CX3H1] | 0.8195 | #6H1][#6H1][#6H1] | 0.0 |
| [CX3H0][CX3H1][#6H1][CX3H1][CX3H0] | 0.774 | #CX4H1][N3H2][CX4H2][CX3H1] | 0.0 |

**worst negatives**

| Structure | prob | Structure | prob |
|-----------|------|-----------|------|
| O=[#6][#6]=[#6X3] | 0.7434 | #6X3H1]=[#6X3H1][#6X3H0][#6X3H1] | 0.0869 |
| [#6X3][#6X3][#6X3][#6X3] | 0.6991 | CCC=CCC | 0.5204 |
| O=CC=O | 0.6991 | [CX3]=#CX3 | 0.5204 |
| [#8]=[#6][#6]=[#8] | 0.3372 | CC=CCC=C | 0.5605 |
| [cH] | 0.3148 | [CHX3]=#C | 0.5863 |
| [#6][#6][#6][#6][#6][#6][#6] | 0.2821 | #6H1][#6H1][#6][#6][#6][#6] | 0.6332 |
| [#8]=[#8][#6][#6][#6][#6][#6][#8] | 0.2751 | #CX3H1]==[CX3H1][CX3H0] | 0.6345 |
| [CX3H0][#CX3H0][CX3H0] | 0.2598 | #8]=[#6H0][#6H1] | 0.6419 |
| [#8][#6][#6][#6][#6][#6] | 0.2581 | #8]=[#6][#6][#6][#6] | 0.7204 |
| [#6H1][#6H1] | 0.2391 | #6X3][#6X3]==[#6X3][#6X3] | 0.7419 |
Example 169 true smiles: CC(C)(C)C#N formula: C5H9N
Index of correct structure: 0 of 125
True structure loss: 0.006086
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

Top predicted substructures

Top predicted substructures
| best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| [CX4H3][CX4H0] | 0.9822 | [OX2H0][CX4H2][CX4H1][CX4H1][CX4H1][OX2H0] | 0.0  |
| [CX4H0][#(NX1H0)][CX4H0] | 0.9328 | [OX1H0][CX3H0][CX4H1][CX4H1][CX4H2] | 0.0  |
| [CX4H3][#6] | 0.8335 | [OX2H0][CX4H1][CX4H1][CX4H1] | 0.0  |
| [CX4H3][CX4H0] | 0.8198 | [OX2H1][CX4H2][CX4H2][CX4H1][CX4H1] | 0.0  |
| [#6][#7] | 0.8108 | [OX2H1][CX4H2][CX4H2][CX4H1] | 0.0  |
| [CX4H3][CX4H1][CX4H3][CX4H3] | 0.788 | [OX1H0][CX3H1][CX4H0][CX4H0] | 0.0  |
| [#6][#7] | 0.7521 | [OX1H0][CX3H1][CX4H0][CX4H0] | 0.0  |

| worst negatives | prob | worst positives | prob |
|-----------------|------|-----------------|------|
| [#6] | 0.3413 | [#6][#6][#6][#6] | 0.788 |
| [#7] | 0.2642 | [#6][#7] | 0.7521 |
| [#7][#7] | 0.2325 | [CX4H3] | 0.788 |
| [CHX3][CHX3][CX4H2] | 0.2317 | [CX4H1][CX4H1][CX4H1][CX4H1] | 0.0  |
| [CHX3][CHX3] | 0.1879 | [#6][#6][#6][#6] | 0.8108 |
| [#7][#7][#6][#6] | 0.1472 | [CX4H3][CX4H0] | 0.8198 |
| [#7][#6][#6][#6] | 0.1214 | [CX4H3][#6] | 0.8335 |
| [#6][#7][#6][#6] | 0.1136 | [CX2H0][#(NX1H0)][CX4H0] | 0.9328 |
| [#7][#6][#6] | 0.1089 | [CX4H3][CX4H0][CX4H3] | 0.9822 |
Example 170 true smiles: N\#Cc1ccccc1 formula: C7H5N
Index of correct structure: 0 of 121
True structure loss: 0.012092
True structure:

![Chemical Structure Image](image)

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CD3OD)

Top predicted structures (loss):

| Structure | Probability |
|-----------|-------------|
| #6H1      | 0.012092    |
| #6X3      | 0.018296    |
| #6X3#6X3  | 0.051488    |
| #6H1#6H1  | 0.060653    |
| #6H1      | 0.061253    |
| #6X3#6X3  | 0.061892    |
| #6X3#6X3#6X3 | 0.064748 |
| #6X3#6X3#6X3#6X3 | 0.066741 |
| #6X3#6X3#6X3#6X3 | 0.068065 |

Top predicted substructures

| Substructure | Probability |
|--------------|-------------|
| #6H1 #6X3   | 0.9992      |
| #6X3 #6X3   | 0.9988      |
| #6X3 #6X3 #6X3 #6X3 | 0.9985 |
| #6X3 #6X3 #6X3 #6X3 #6X3 #6X3 #6X3 | 0.9834 |
| #6X3 #6X3 #6X3 #6X3 #6X3 #6X3 #6X3 #6X3 #6X3 #6X3 | 0.9803 |
| #6X3 #6X3 #6X3 #6X3 #6X3 #6X3 #6X3 #6X3 #6X3 #6X3 #6X3 #6X3 | 0.9606 |
best positives  prob  best negatives  prob
[#6H1]  0.9992  [#8][#6H1][#6H2][#6H1]=[#8]  0.0
[#6X3][#6X3]  0.9988  [CX4H0][OX2H0][CX4H3][CX4H2][CX4H1]  0.0
[#6X3][#6X3][#6X3][#6X3][#6X3]  0.9985  [OX2H0][CX4H2][OX2H0][CX4H0]  0.0
[cH][cH]  0.9992  [OX2H0][CX4H2][CX4H2][CX4H1][OX2H0]  0.0
[#6X3H1][#6X3H0]  0.9834  [OX4H1][OX2H0][CX4H3][CX4H0]  0.0
cX3H1][cX3H1][cX3H0][cX3H1]  0.9803  [OX2H1][CX4H1][CX4H1][CX4H1]  0.0
[#6H1][#6H1]  0.9606  [OX1H0][CX3H0][CX4H1][CX4H1][CX4H2]  0.0
cX3H1][cX3H1][cX3H1]  0.9276  [OX2H0][CX4H2][CX4H1][CX4H1][CX4H1]  0.0
worst negatives  prob  worst positives  prob
[#7][#6][#6][#6X3]  0.7941  [#7][#6][#6][#6X3]  0.5001
[#7][#6X3H0][#6X3H1]  0.5595  [#7][#6][#6][#6X3]  0.556
[#7][#6H0][#6H1]  0.5407  [#6X3H1][#6X3H1][#6X3][#6X3]  0.7164
[#6X3][#7][#6X3]  0.4874  [#6][#6][#6][#6][#6H1]  0.9276
[#61][#6][#6][#6][#6]  0.3077  [#6H1][#6H1]  0.9606
[cX3H1][cX3H1][cX3H1]  0.1785  [#6X3H0][#6X3H0]  0.9834
[6X3][#7X3][#6X3]  0.1699  [cH]  0.9912
-----------------------------------------------------------------------------------------------
| best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| [CX4H2][CX4H2] | 0.8359 | [#6H3][#7X3H0][#6X4H2][#6X4H2] | 0.7393 |
| [#7X3][#6H3] | 0.8244 | [CX4H2][#CX4H2][CX4H2] | 0.727 |

| worst negatives | prob | worst positives | prob |
|-----------------|------|-----------------|------|
| [#6H1][#6H2] | 0.5697 | [#6][#6][#6][#7][#7] | 0.1489 |
| [#7X3H2] | 0.3576 | [CX4H3] | 0.2449 |
| C1CCC1 | 0.2572 | [#6H2][#7][#6H2] | 0.2926 |
| [#6H1] | 0.2456 | [#6H3][#7][#6H2] | 0.4922 |
| [#7X3H1] | 0.2283 | [#7X3][#6H2] | 0.5065 |
| [#6H1][#6H2][#6H2] | 0.173 | [CX4H2][CX4H2][CX4H2][CX4H2] | 0.6648 |
| CCCCCC | 0.1713 | [#7][#6H2] | 0.7116 |
| [#7][#6H1][#6H2r5] | 0.1611 | [CX4H2][#CX4H2][CX4H2] | 0.727 |
| [CX4H2][#NX3H2][CX4H2] | 0.1259 | [#6H3][#7X3H0][#6X4H2][#6X4H2] | 0.7393 |
| [CX4H2][#CX4H2][CX4H2] | 0.1252 | [#7][#6H2][#6H2] | 0.752 |
Example 172 true smiles: Oc1ccccc1 formula: C6H6O
Index of correct structure: 0 of 98
True structure loss: 0.005278

True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

| Structure  | prob  |
|------------|-------|
| ![Structure 1](image)  | 0.005278 |
| ![Structure 2](image)  | 0.046617  |
| ![Structure 3](image)  | 0.057299  |
| ![Structure 4](image)  | 0.06161   |
| ![Structure 5](image)  | 0.08174   |
| ![Structure 6](image)  | 0.083112  |
| ![Structure 7](image)  | 0.104978  |
| ![Structure 8](image)  | 0.13359   |
| ![Structure 9](image)  | 0.148344  |
| ![Structure 10](image) | 0.166134  |

Top predicted substructures:

- [#6X3]#[#6X3] prob: 0.9998, [cX3H1][{cX3H1}]{cX3H0}
- [#6X3]#[#6X3]#[#6X3]#[#6X3] prob: 0.9997, [#6X3H1]#6X3H0]
- [cH][cH] prob: 0.9991, [#6]1[#6]1[#6]1[#6]1
| best positives | prob | worst negatives | prob |
|----------------|------|----------------|------|
| [#6X3][#6X3]  | 0.9988 | [cX4H0][{cX4H1}][{cX4H2}][{cX4H1}] | 0.0 |
| [#6X3][#6X3][#6X3][#6X3] | 0.9997 | [Ox2H0][{cX4H2}][{cX4H1}][{cX4H0}] | 0.0 |
| [cH][cH] | 0.9991 | [cX4H0][{cX4H1}][{cX4H2}][{cX4H1}] | 0.0 |
| [#6H1] | 0.9981 | [cX4H0][{cX4H0}][{cX4H1}] | 0.0 |
| [cH] | 0.9853 | [Ox2H0][{cX4H2}][{cX4H1}][{cX4H0}] | 0.0 |
| [cX3H1][{cX3H1}]][{cX3H0}] | 0.9851 | [Ox1H0][{cX3H0}][{cX4H1}][{cX4H2}][{cX4H0}] | 0.0 |
| [#6X3H1][#6X3H0] | 0.9816 | [cX4H1][{cX4H1}][{cX4H2}][{cX4H0}] | 0.0 |
| [#6][#6][#6][#6][#6] | 0.9808 | [Ox2H0][{cX4H2}][{cX4H2}][{cX4H1}][{cX4H1}] | 0.0 |
| [cX3H1][{cX3H1}][{cX3H1}] | 0.9659 | [cX4H1][{cX3H1}][{cX4H3}][{cX4H2}] | 0.0 |
| [#6H1][#6H1] | 0.9543 | [cX2H1][{cX2H0}][{cX3H1}][{cX3H0}] | 0.0 |

| worst positives | prob | worst positives | prob |
|----------------|------|----------------|------|
| [cX3H0][{cX3H1}][{cX3H0}][{Ox2H1}] | 0.1679 | [Ox2H1] | 0.6808 |
| [#7][#6][#6][#6X3] | 0.2303 | [#6X3H1][#6X3H1][#6X3H0][#6X3H1] | 0.7855 |
| [cX3H0][{cX3H1}][{cX3H1}][{cX3H0}] | 0.2131 | [#8][#6][#6][#6X3] | 0.8662 |
| [cX3H1][{cX3H0}][{cX3H1}] | 0.2107 | [{cX3H1}:{c}] | 0.8725 |
| [#8][#6H1][#6H1] | 0.1858 | [cH][cO] | 0.9132 |
| [cX3H0][{cX3H1}][{cX3H0}][{Ox2H1}] | 0.1813 | [#8][{#6H0}][#6H1] | 0.9141 |
| [#8][#6H][#6X3][#6X3H] | 0.1665 | [#6H1][#6H1] | 0.9543 |
| [#6X3][#7][#6X3] | 0.1618 | [cX3H1][{cX3H1}][{cX3H1}] | 0.9659 |
| [#6][#6][#6][#6][#7] | 0.1551 | [#6][#6][#6][#6][#6][#6] | 0.9808 |
| [#7][#6][#6X3] | 0.1461 | [#6X3H1][#6X3H0] | 0.9816 |
Example 173 true smiles: NCCCCCN, formula: C5H14N2

Index of correct structure: 0 of 97
True structure loss: 0.005111

True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

0.005111 0.023611 0.036452 0.042226 0.042658

0.045164 0.04917 0.05025 0.052515 0.059539

Top predicted substructures

[#7X3H2] 0.9977 0.9976 0.997 0.9659
[#7][#6H2][#6H2] 0.9278
[#7X3][#6H2] 0.9225
Example 174 true smiles: OCC(O)C(O)CO formula: C4H10O4
Index of correct structure: 0 of 92
True structure loss: 0.016919
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

0.016919  0.020553  0.02169  0.022483  0.023601

0.024666  0.0254  0.026699  0.028374  0.029403

Top predicted substructures

| Substructure | prob    | Substructure | prob    | Substructure | prob    |
|--------------|---------|--------------|---------|--------------|---------|
| [CX4H2][[#6]| 0.9975  | [CX4H1)(O)CO | 0.8638  | [CX4H2][OX2H1]| 0.8315  |
| [OX2H1]     | 0.9968  | [CX4H1)(OX2H1)| 0.8025  | [CX4H2][OX2H2][#8] | 0.8025  |
best positives

| Structure                        | prob | Structure                        | prob |
|---------------------------------|------|---------------------------------|------|
| [CX4H2][#6][O]                  | 0.9975 | [CX2H1][#6][O]                  | 0.9955 |
| [OX2H1]                         | 0.9865 | [CX2H1][#6][O]                  | 0.9960 |
| [ox2h1]                         | 0.9858 | [CX2H1][#6][O]                  | 0.9871 |
| [CX4H1][O]CO                    | 0.8638 | [CX2H1][#6][O]                  | 0.9960 |
| [CX4H2][OX2H1][CX4H1]           | 0.8315 | [CX2H1][#6][O]                  | 0.9865 |
| OCC[CH2]                        | 0.8025 | [CX2H1][#6][O]                  | 0.9955 |
| [CX4H2][O][CHX4]                | 0.7424 | [CX2H1][#6][O]                  | 0.9858 |
| OCC[CH2]                        | 0.7424 | [CX2H1][#6][O]                  | 0.9871 |

worst negatives

| Structure                        | prob | Structure                        | prob |
|---------------------------------|------|---------------------------------|------|
| [CX4H2][CX4H2]                  | 0.4858 | [CX4H2][OX2H1][CX4H2]           | 0.4395 |
| [CH2X4][O][CX4H2]               | 0.4395 | [CX2H1][#6][O]                  | 0.9955 |
| [CX4H2][OX2H1][CX4H2]           | 0.371  | [CX2H1][#6][O]                  | 0.9960 |
| O[CX4H2][CX4H2][O]              | 0.2279 | [CX2H1][#6][O]                  | 0.9865 |
| [CX4H1][OX2H0][CX4H2]           | 0.2056 | [CX2H1][#6][O]                  | 0.9955 |
| [CX4H1][OX2H1][CX4H2][CX4H2]    | 0.1785 | [CX2H1][#6][O]                  | 0.9960 |
| [O][CX4H1][CX4H2][O]            | 0.1477 | [CX2H1][#6][O]                  | 0.9865 |
| [OX2H0][O][CX4H2][CX4H2][OX2H0] | 0.1118 | [CX2H1][#6][O]                  | 0.9871 |
| [CX4H1][OX2H1][CX4H1][CX4H1]    | 0.1081 | [CX2H1][#6][O]                  | 0.9960 |
| [CX4H2][O][OX2H1][CX4H2]        | 0.0987 | [CX2H1][#6][O]                  | 0.9960 |

worst positives

| Structure                        | prob | Structure                        | prob |
|---------------------------------|------|---------------------------------|------|
| [CX4H2][CX4H2]                  | 0.1591 | [CX4H2][OX2H1][CX4H2]           | 0.1591 |
| [CH2X4][O][CX4H2]               | 0.1591 | [CX2H1][#6][O]                  | 0.2583 |
| [CX4H2][OX2H1][CX4H2]           | 0.2611 | [CX2H1][#6][O]                  | 0.2583 |
| O[CX4H2][CX4H2][O]              | 0.2855 | [CX2H1][#6][O]                  | 0.2611 |
| [OX2H0][OX2H0][CX4H2]           | 0.3125 | [CX2H1][#6][O]                  | 0.2855 |
| [CX4H1][OX2H0][CX4H2]           | 0.3125 | [CX2H1][#6][O]                  | 0.2855 |
| [OX2H0][O][CX4H2][CX4H2][OX2H0] | 0.3703 | [CX2H1][#6][O]                  | 0.3125 |
| [O][CX4H1][CX4H2][O]            | 0.5707 | [CX2H1][#6][O]                  | 0.3703 |
| [CX4H1][OX2H1][CX4H1][CX4H1]    | 0.5736 | [CX2H1][#6][O]                  | 0.5707 |
| [OX2H2][O][OX2H1][CX4H2]        | 0.5941 | [CX2H1][#6][O]                  | 0.5736 |
| [OX2H1][OX2H1][CX4H2]           | 0.5981 | [CX2H1][#6][O]                  | 0.5941 |
Example 175 true smiles: OCC(O)C(O)CO formula: C4H10O4
Index of correct structure: 0 of 92
True structure loss: 0.016768
True structure:

![Chemical structure](image)

Experimental 13C NMR (solvent: CD3OD)

![13C NMR spectrum]

Experimental 1H NMR (solvent: D2O)

![1H NMR spectrum]

Top predicted structures (loss):

| Structure | Prob |
|-----------|------|
| [CX4H2]([#6])O | 0.9976 |
| [OX2H1] | 0.997 |
| [#8][#6][#6H2][#8] | 0.9868 |
| [CX4H2]([OX2H1])O[C4H1] | 0.9726 |
| [OX2H1]COCX4H2 | 0.8343 |
| OCC[CH2] | 0.8025 |

Top predicted substructures

| Substructure | prob |
|--------------|------|
| [CX4H2]([#6])O | 0.9976 |
| [OX2H1]COCX4H2 | 0.997 |
| [#8][#6][#6H2][#8] | 0.9868 |
| [OX2H1]COCX4H2 | 0.9726 |
| [OX2H1]COCX4H2 | 0.8343 |
| OCC[CH2] | 0.8025 |
best positives

\[
\begin{align*}
[CX4H2][\{\#6\}][O] & \quad \text{prob} \quad 0.996 \quad \text{CX4H2}[O][CHX4] \quad \text{prob} \quad 0.743 \\
[CX4H2][\{\#6\}][O] & \quad \text{prob} \quad 0.9768 \quad \text{CX4H2}[\{\#6\}][\#6H1] \quad \text{prob} \quad 0.6936
\end{align*}
\]

worst negatives

\[
\begin{align*}
[CX4H2][CX4H2] & \quad \text{prob} \quad 0.4851 \quad [\#6H2][\#6H1][\#6H1][\#6H2] \quad \text{prob} \quad 0.1637 \\
[CX4H2][OX2H0][CX4H2] & \quad \text{prob} \quad 0.4406 \quad [\#8][\#6][\#6][\#6][\#6] \quad \text{prob} \quad 0.2587 \\
[CX4H2][O][CX4H2] & \quad \text{prob} \quad 0.3707 \quad [\#8][\#6][\#6H1][\#6H1] \quad \text{prob} \quad 0.2632 \\
O[CX4H2][CX4H2][O] & \quad \text{prob} \quad 0.2245 \quad \text{prob} \quad 0.2898 \\
[CX4H2][OX2H0][OX2H0][CX4H2] & \quad \text{prob} \quad 0.1964 \quad \text{prob} \quad 0.3079 \\
[CX4H2][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0] & \quad \text{prob} \quad 0.1786 \quad \text{prob} \quad 0.379 \\
[CX4H2][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0] & \quad \text{prob} \quad 0.1469 \quad \text{prob} \quad 0.5761 \\
[CX4H2][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0] & \quad \text{prob} \quad 0.1147 \quad \text{prob} \quad 0.5827 \\
[CX4H2][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0] & \quad \text{prob} \quad 0.1096 \quad \text{prob} \quad 0.5944 \\
[CX4H2][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0][OX2H0] & \quad \text{prob} \quad 0.1004 \quad \text{prob} \quad 0.6005
\end{align*}
\]
Example 176 true smiles: OCC=CCO formula: C4H8O2
Index of correct structure: 0 of 72
True structure loss: 0.00449
True structure:

![Structure Image]

Experimental 13C NMR (solvent: DMSO-d6)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

0.00449
0.065389
0.104814
0.10675
0.125735
0.131649
0.143793
0.146275
0.148439
0.148442

Top predicted substructures

| Substructure | prob |
|--------------|------|
| [CHX3]=[C]C  | 0.9998|
| [#6H1]       | 0.9996|
| [#8][#6H2][#6H]=[#6X3] | 0.9992|
| [CX3H1]>({CX3H1})[CX4H2] | 0.9912|
| [CX4H2]>({#6})[O] | 0.9907|
| 0[CX4H2][CX3H1] | 0.9892|
| Structure | Best Positives | Prob  | Structure | Best Negatives | Prob  |
|-----------|----------------|-------|-----------|----------------|-------|
| [CX4H2](#CX3)=C | 0.998 | | [#8](#6)=[#6X3] | 0.9767 |
| [#8H](#6H2)=#6H1 | 0.9952 | | [#6X3](#6H2) [#8] | 0.9551 |
| best positives | | | worst negatives | | |
| [CHX3](=C)C | 0.9998 | | [CX2H0](#CX2H1) [#6X3] | 0.0 |
| [#6H1] | 0.9996 | | [CX2H0] [#CX2H1] [CX4H0] | 0.0 |
| [#8](#6H2)=[#6X3] | 0.9992 | | [CX2H1] [#CX2H0] [CX3H1] [#6X3] | 0.0 |
| [CX4H2](#CX3)=C | 0.998 | | C=CC=C#C | 0.0 |
| [#6X3](#6H2)=#6H1 | 0.9952 | | [#6X2](#6H1) [#6X2] | 0.0 |
| [CX3H1] [#CX3H1] [CX4H2] | 0.9912 | | CC#CCC#C | 0.0 |
| worst negatives | | | worst positives | | |
| [#6H1] [#6H1] | 0.3787 | | [CX2H0] [#CX2H1] [CX4H1] | 0.0 |
| [OH] [CX4H] | 0.2697 | | [CX4H2] [CX3H] | 0.0 |
| [CX3H] [CX4H] | 0.1495 | | [#6H1] [#6H2] | 0.0 |
| [#8] [#6H1] [#6H1] | 0.1355 | | [CHX3] [#CHX3] | 0.0 |
| [#8] [#6] [#6X3] | 0.1217 | | [CX4H2] [#OX2H1] [CX3H1] | 0.0 |
| [CX4H3] [OX2H0] | 0.1053 | | [OX2H1] | 0.0 |
| [CX3H1] [#CX3H1] [CX4H1] | 0.1026 | | [#8] [#6H2] [#6X3H] [#6X3H] | 0.0 |
| worst negatives | | | worst positives | | |
| [#6H1] [#6H1] | 0.1026 | | [CX2H0] [#CX2H1] [CX4H1] | 0.0 |
| [OH] [CX4H] | 0.1026 | | [CX4H2] [CX3H] | 0.0 |
| [CX3H] [CX4H] | 0.1026 | | [#6H1] [#6H2] | 0.0 |
| [#8] [#6H1] [#6X3] | 0.1026 | | [CX4H2] [OX2H1] [CX3H1] | 0.0 |
| [CX4H3] [OX2H0] | 0.1026 | | [OX2H1] | 0.0 |
| [CX3H1] [#CX3H1] [CX4H1] | 0.1026 | | [#8] [#6H2] [#6X3H] [#6X3H] | 0.0 |
Example 177 true smiles: COCC(C)=O
formula: C₄H₈O₂
Index of correct structure: 0 of 72
True structure loss: 0.006544
True structure:

**Experimental 13C NMR (solvent: CDCl₃)**

**Experimental 1H NMR (solvent: CDCl₃)**

Top predicted structures (loss):

1. [CX₄H₃][CX₃] 1.0  [CX₄H₃][#6][#6][#6] 0.9977
2. [CX₄H₃][CX₃H₀] 0.9999  [CX₄H₃][#6H₀][#6] 0.9962
3. [CX₄H₃] 0.9998  [CX₄H₃][#6] 0.9908
4. [CX₄H₃][CX₃] 0.115004
5. [CX₄H₃][CX₃H₀] 0.130328
6. [CX₄H₃] 0.131932
7. [CX₄H₃][CX₃] 0.134524
8. [CX₄H₃][CX₃H₀] 0.138948
9. [CX₄H₃] 0.140668
10. [CX₄H₃][CX₃] 0.141956
| best positives                      | prob | best negatives                      | prob |
|------------------------------------|------|-------------------------------------|------|
| [CX4H3][CX3]                       | 1.0  | CCC#CC#C                            | 0.0  |
| [CX4H3][CX3H0]                     | 0.9999 | C=CC=CC=CC                          | 0.0  |
| [CX4H3]                            | 0.9998 | [CX3H0]=([CX3H1])                  | 0.0  |
| [OX1][OX1]=(OX1)C                  | 0.9995 | C=CC=[CX4H3]                       | 0.0  |
| [OX1][OX1]=(OX1)C                  | 0.9999 | C=CC=[CX3H0]                       | 0.0  |
| [OX1][OX1]=(OX1)C                  | 0.9994 | [CX2H0]=([CX2H1])                  | 0.0  |
| [OX1][OX1]=(OX1)C                  | 0.9997 | [CX2H0]=([CX2H1])                  | 0.0  |
| [OX1][OX1]=(OX1)C                  | 0.9992 | [CX2H0]=([CX2H1])                  | 0.0  |
| [OX1][OX1]=(OX1)C                  | 0.9990 | [CX2H0]=([CX2H1])                  | 0.0  |
| [OX1][OX1]=(OX1)C                  | 0.9996 | [CX2H0]=([CX2H1])                  | 0.0  |
| worst negatives                    | prob | worst positives                     | prob |
| OCC[CX2]                           | 0.3472 | [#8][#6][#6]=[#8]                   | 0.7237|
| [OX1][#6][#6H2]                    | 0.318  | [#8]=([#6][#6H2])                   | 0.7293|
| [OX1][OX1][#6][#6H1]               | 0.2955 | [CX4H2][CX3]=O                      | 0.7488|
| [OX1][OX1][#6]                     | 0.293  | [CX3H0]=([OX1H0])                  | 0.7952|
| [OX1][OX1][#6][#6H1]               | 0.2557 | [CX3H0]=([OX1H0])                  | 0.833 |
| [OX1][OX1][#6][#6][#6X3]           | 0.2477 | [CX3][#6][#6H1][#6H2]               | 0.9026|
| [OX1][OX1][#6][#6][#6H2]           | 0.2464 | [CX3][#6][#6H3][#6H2]               | 0.906 |
| [OX1][OX1][#6][#6][#6H2]           | 0.1551 | [CX4H2][OX4H2]                     | 0.9204|
| [OX1][OX1][#6][#6][#6H2]           | 0.1506 | [CX4H2][OX4H2]                     | 0.928 |
| [OX1][OX1][#6][#6][#6H2]           | 0.1477 | [CX4H2][OX4H2][OX2H0]              | 0.9368|
Example 178 true smiles: CCCC(=O)O

Formula: C4H8O2

Index of correct structure: 0 of 72
True structure loss: 0.00878

True structure:

![Chemical Structure](image)

Experimental 13C NMR (solvent: CDCl3)

![13C NMR Spectrum](image)

Experimental 1H NMR (solvent: D2O)

![1H NMR Spectrum](image)

Top predicted structures (loss):

- [CX3](=O)[OX1]C: 0.999
- [CX4H3][CX4H2]C: 0.9942
- [CX4H2]([#6])[#6]: 0.9989
- [#8]=[#6][#8]: 0.9941
- [#6H3][#6][#6]: 0.9989
- [CX3](=O)[OX2H1]: 0.9921

Top predicted substructures

- [CX3](=O)[OX1]C: 0.999
- [CX4H3][CX4H2]: 0.9942
- [#8]=[#6][#8]: 0.9941
- [CX3](=O)[OX2H1]: 0.9921
best positives prob  best negatives prob

[OX2H1] 0.9977  [CX3]{=[OX1]}O 0.9877
[CX4H3] 0.9972  [OX2H1] 0.9872

worst negatives prob  worst positives prob

[OX2H1][OX1H0][{OX2H1}][CX4H1] 0.5366  [#8][#6][#6H2] 0.4587
[#6H0][#6H1] 0.3821  [OX1H0]=[CX3H0][{#8}][CX4H2] 0.5066
[#8][#6H0][#6H1] 0.3477  O=[CX3H0][CX4H2][CX4H2] 0.5954
O=[CX3][CX4H] 0.2241  [CX4H2][CX3]=O 0.6263
[CX4H3][CX4H1] 0.1645  OCC[CH2] 0.6402
[#6H1] 0.1644  [#6H3][#6][#6H3] 0.7231
[#6H1][#6H2] 0.1625  [CX4H2][CC]=O 0.7422
[#8][#6][#6][#8] 0.1511  [CX4H2][{CX4H2}][CX3H0] 0.7681
[#8][#6][#6]=[#8] 0.1134  [CX3H0][{OX1H0}][{OX2H1}][CX4H2] 0.8081
[CX3][{OX1}]O 0.1129  [CX4H2][CX4H2] 0.8794

------------------------------------------------------------------------------------------------------------------------
Example 179 true smiles: CC(C)=O\ O \ formula: C4H8O2
Index of correct structure: 0 of 72
True structure loss: 0.005506
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

Top predicted substructures

prob
[CX4H3] 1.0
[#6H3][#6][#6] 0.9999 [CX4H3][CX4H1] 0.9659
[CX4H3][#6] 0.9998 [CX3]=[OX1]O 0.9634

0.005506
0.077914
0.084234
0.101048
0.107408
0.113703
0.124453
0.127398
0.129836
0.132785
| best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| [CX3](=[OX1])C | 0.9977 | O=[CX3][CX4H] | 0.9352 |
| [CX3](=[OX2H1]) | 0.9969 | [OX2H1] | 0.9339 |
| [CX4H3] | 1.0 | [CX2H1][CX2H0][CX3H1] | [CX3H0] | 0.0 |
| [CX4H3][#6][#6] | 0.9999 | C=CC=CC#C | 0.0 |
| [CX4H3][#6] | 0.9998 | CCC=CC#C | 0.0 |
| [CX3](=[OX1])C | 0.9969 | [CX2H0][#][CX2H1][CX3H0] | 0.0 |
| [CX3](=[OX1])O | 0.9968 | [CX2H0](=#[CX2H1])[cX3H0] | 0.0 |
| [CX2H0](=##CX2H1)[cX3H0] | 0.9634 | [CX2H0][CX2H0] | 0.0 |
| [CX2H0][#][CX2H0] | 0.9352 | [CX2H0][#][CX2H0] | 0.0 |
|worst negatives | prob | worst positives | prob |
| [#6]=|#6[#6H1][#6H1] | 0.4888 | [CH3][CC][OH] | 0.5814 |
| [#6X3][#6][#6H3] | 0.2938 | [CH4][CH3]CH3] | 0.7996 |
| [CX3H0][=[OX1H0]][OX2H1][CX4H2] | 0.286 | [#6][#6H0][#6H1] | 0.8055 |
| [CX4H2][#6][#6] | 0.2048 | [#6X3][#6][#6X3] | 0.8144 |
| [CX4H2][#6][#6H3] | 0.1777 | [CX4H1][[OX1H0]][[OX2H1]][CX4H1] | 0.864 |
| [CX4H2][#6] | 0.168 | [#6][[#6][#6H0][#6H1] | 0.8774 |
| [CX4H2][#6][#6] | 0.0983 | [#6][#6H1] | 0.9054 |
| [CX4H2][#6H1][CH][CH] | 0.0814 | [OX2H1] | 0.9079 |
| [#6H1][#6H2] | 0.0726 | O=[CX3][CX4H] | 0.9352 |

------------------------------------------------------------------------------------------------------------------------
Example 180 true smiles: CO(C)CCO formula: C5H12O2
Index of correct structure: 0 of 69
True structure loss: 0.013422
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

Top predicted substructures
| best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| [CX4H3] | 0.9924 | [CX4H1] | 0.9263 |
| [#6H1] | 0.9916 | [CX4H2] | 0.8977 |
| [CX4H3] | 1.0 | [CX2H0] | 0.0 |
| [CX4H3] | 0.9974 | CCC=C | 0.0 |
| [CX4H3] | 0.9944 | CCC=CC | 0.0 |
| [CX4H3] | 0.9924 | [#7]=[#6]=[#6] | 0.0 |
| [CX4H3] | 0.9916 | C=C=CC | 0.0 |
| [CX4H3] | 0.9701 | CCC=C | 0.0 |
| [CX4H3] | 0.9668 | [C3H0]==[C3H0]==[C3H0] | 0.0 |
| [#6H3] | 0.9402 | [CX2H0] | 0.0 |
| [CX4H3] | 0.9263 | [CX2H1]=[CX2H0]=[CX3H1]=[CX3H0] | 0.0 |
| [CX4H2] | 0.8977 | CCC=C | 0.0 |
| worst negatives | prob | worst positives | prob |
| [#8] | 0.54 | [OX2H0] | 0.2309 |
| [#8] | 0.429 | [O2H] | 0.4421 |
| [CX4H3] | 0.3905 | O=C | 0.5162 |
| [OH] | 0.292 | [CX4H2] | 0.5304 |
| [CX4H2] | 0.2727 | [CX4H2] | 0.5396 |
| [#8]==[#6]=[#6] | 0.2527 | [CHX4]=[CHX4]=[CH2X4] | 0.5901 |
| [CX4H2] | 0.2359 | [CX4H2] | 0.6575 |
| O=[CX4H2] | 0.1783 | [CX4H2] | 0.6716 |
| [CX4H2] | 0.1760 | [CX4H1] | 0.761 |
| [CX4H1] | 0.1731 | [OX2H0] | 0.8019 |
Example 181 true smiles: COO(C)(C)OC formula: C₅H₁₂O₂

Index of correct structure: 0 of 69

True structure loss: 0.004759

True structure:

[Experimental 13C NMR (solvent: CDCl₃)]

[Experimental 1H NMR (solvent: CDCl₃)]

Top predicted structures (loss):

0.004759  0.014062  0.022775  0.045376  0.045577

0.04945  0.053543  0.062637  0.064374  0.066537

Top predicted substructures

[CX₄H₃] [CX₄H₃][CX₄]O [CX₄][CX₄][CX₄]O
| Structure | prob | Structure | prob |
|-----------|------|-----------|------|
| [CX4H3][CX4H0] | 0.9935 | [#6H3][#6H0] | 0.9627 |
| [CX4H3][OX2H0] | 0.9935 | [#6H0][({#6H3})[{#6H3}][#8)] | 0.9389 |
| **best positives** | | **best negatives** | |
| [CX4H3] | 0.9999 | [CX2H0][({#CX2H1})[cX3H0] | 0.0 |
| [CX4H3][#6] | 0.9997 | CCC=C=C=C | 0.0 |
| [CX4H3][CX4]O | 0.9995 | C=CC=C=C | 0.0 |
| [CX4H3][CX4H0] | 0.9935 | [CX2H0][({#CX2H0})[CX2H0] | 0.0 |
| [CX4H3][OX2H0] | 0.9935 | [#7][#6][={#6][#6][#7] | 0.0 |
| [#6H3][#6][#6] | 0.9887 | C=C=C=C=C | 0.0 |
| [#6H3][#8][#6H0][#8] | 0.9717 | [#6X2][#6H1][#6X2] | 0.0 |
| [CH3][#6][#8] | 0.9704 | [CX2H0][({#CX2H1})[CX2H0] | 0.0 |
| [#6H3][#6H0] | 0.9627 | CC=CCC=C | 0.0 |
| [#6H0][({#6H3})[{#6H3}][#8] | 0.9389 | [CX3H0][({OX2H0})[CX2H0] | 0.0 |
| **worst negatives** | | **worst positives** | |
| [#6H1] | 0.5702 | [CX4H3][CX4H0][{CX4H3}]{OX2H0] | 0.9098 |
| [OX2H0][CX4H1][OX2H0] | 0.5465 | [CX4H3][CX4H0][CX4H3] | 0.9222 |
| [#8][#6][#6H2] | 0.4975 | [#6H0][({#6H3})[{#6H3}][#8] | 0.9389 |
| [#8][#6H0][#6H1] | 0.2987 | [#6H3][#{6H0} | 0.9627 |
| [OX2H0][CX4H1][OX2H0] | 0.2463 | [CH3][#{6}[#6] | 0.9704 |
| OCC[CH2] | 0.1506 | [#6H3][#8][#6H0][#8] | 0.9717 |
| [#8][#6H1][#6H1] | 0.1153 | [#6H3][#6][#6] | 0.9887 |
| [#8][#6H1][#6H1] | 0.0941 | [CX4H3][OX2H0] | 0.9935 |
| [#8][#6][#6][#8] | 0.0532 | [CX4H3][CX4]O | 0.9995 |
Example 182 true smiles: C=O(C)C(=O)O
Formula: C₄H₆O₂
True structure:

Experimental ¹³C NMR (solvent: CDCl₃)

Experimental ¹H NMR (solvent: D₂O)

Top predicted structures (loss):

| Structure                          | prob  |
|------------------------------------|-------|
| [CX₄H₃]                            | 0.9997|
| [CX₄H₃][#6]                        | 0.9997|
| [CX₄H₃][CX₃]                       | 0.9997|

Top predicted substructures

| Structure                          | prob  |
|------------------------------------|-------|
| [#6H₃][#6]=[#6X₃]                  | 0.9926|
| [#6H₃]                            | 0.9866|
| [OX₂H₁]                            | 0.9865|
| Structure | prob | Structure | prob |
|-----------|------|-----------|------|
| \([\text{CX3}]\{(=O)\}[\text{OX2H1}]\) | 0.9976 | \([\#8]=\{\#6\}\{\#8\}\) | 0.9827 |
| \([\text{CX3}]\{\{\text{OX1}\}\}\text{C}\) | 0.9935 | \([\text{CX3}]\{\{\text{OX1}\}\}\text{O}\) | 0.9798 |
| best positives | prob | best negatives | prob |
| \([\text{CX4H3}]\) | 0.9997 | \([\text{CX2H0}]\{\{\text{CX2H1}\}\}[\text{CX4H0}]\) | 0.0 |
| \([\text{CX4H3}\{\#6}\) | 0.9997 | \([\text{CX2H0}]\{\{\text{CX2H1}\}\}[\text{CX4H2}]\) | 0.0 |
| \([\text{OX2H1}]\{\{\text{CX3}\}\}\) | 0.9935 | \([\#6H2]\{\#6\}\{\#6X2\}\) | 0.0 |
| \([\#8]=\{\#6\}\{\#8\}\) | 0.9926 | CCC\#CC\#C | 0.0 |
| worst negatives | prob | worst positives | prob |
| \([\#6H1]\) | 0.9866 | \([\text{CX3H2}]=\{\text{CX3H0}\}\) | 0.2111 |
| \([\text{CX4H3}]\{\{\text{CX3H2}\}\}\{\text{C}\}\) | 0.9658 | \([\text{CX2X3}]\{\{\text{CX3H2}\}\}\{\text{C}\}\) | 0.3434 |
| \([\text{CX3H1}]\{\{\text{CX3H1}\}\}[\text{CX4H3}]\} | 0.7783 | \([\text{CH3}\{\text{CC}\}\{\text{OH}\}\}) | 0.3681 |
| \([\text{CX3H1}]\{\{\text{CX3H1}\}\}[\text{CX4H3}]\{\{\text{CX4H1}\}\}) | 0.563 | \([\text{CX3H2}]=\{\text{CX3H0}\}\{\text{CX3}]=\{\text{O}\}\) | 0.3915 |
| \([\#6X3]\{\#6\}]=\{\#6\}\{\#6X3\}\) | 0.5421 | \([\text{CX3H2}]=\{\text{CX3H0}\}[\text{CX3}]=\{\text{O}\}\) | 0.4052 |
| \([\#8]=\{\#6\}\{\#8\}\) | 0.5275 | \([\text{CX4H3}]\{\text{CX3H0}\}\) | 0.4053 |
| \([\#8]=\{\#6\}\{\#8\}\) | 0.4829 | \([\text{CX3H2}]=\{\text{CX3H0}\}[\{\#6\}]=\{\#6\}\) | 0.4073 |
| \([\#6X3]=\{\#6\}\{\#6X3\}\) | 0.4016 | \([\#6H3]=\{\#6H0\}\{\#6H0\}\) | 0.4215 |
| \([\#6X3]=\{\#6\}\{\#6X3\}\) | 0.4309 | \([\text{CX3H2}]=\{\text{CX3H2}\}[\{\text{CX4H3}\}\{\text{CX3H0}\}\) | 0.4381 |
| \([\text{CH3H3}]\{\text{CH3H3}\}\) | 0.3962 | \([\text{CX4H3}]\{\text{CX3H0}\}[\{\text{CX3H2}\}\) | 0.4399 |
Example 183 true smiles: Cl-COCCCl formula: C5H8O
Index of correct structure: 0 of 66
True structure loss: 0.00998
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

| Structure | Loss |
|-----------|------|
| ![Structure 1](image1) | 0.06968 |
| ![Structure 2](image2) | 0.077758 |
| ![Structure 3](image3) | 0.08154 |
| ![Structure 4](image4) | 0.103122 |
| ![Structure 5](image5) | 0.1057 |
| ![Structure 6](image6) | 0.116916 |
| ![Structure 7](image7) | 0.118393 |
| ![Structure 8](image8) | 0.120263 |
| ![Structure 9](image9) | 0.123239 |
| ![Structure 10](image10) | 0.127778 |

Top predicted substructures

| Substructure | prob |
|--------------|------|
| ![Substructure 1](image11) | 0.9973 |
| ![Substructure 2](image12) | 0.9936 |
| ![Substructure 3](image13) | 0.9934 |
| ![Substructure 4](image14) | 0.9545 |
| ![Substructure 5](image15) | 0.9385 |
| ![Substructure 6](image16) | 0.9379 |
| best positives | prob | worst negatives | prob |
|----------------|------|----------------|------|
| [C\text{X4H2}][\text{O}][\text{C}] | 0.9545 | [C\text{X4H3}][\text{CX3}] | 0.6603 |
| [C\text{X4H3}][\text{CX3}] | 0.9629 | [O][\text{C}] | 0.6832 |
| [C\text{X4H2}][\text{C}] | 0.2898 | [C\text{X4H2}] | 0.745 |
| [C\text{X4H2}] | 0.2509 | [C\text{X4H2}] | 0.7567 |
| [C\text{X4H2}][\text{C}] | 0.1956 | [C\text{X4H2}] | 0.795 |
| [C\text{X4H2}] | 0.1729 | [C\text{X4H2}] | 0.7987 |
| [C\text{X4H2}] | 0.1534 | [C\text{X4H2}] | 0.8047 |
| [C\text{X4H2}] | 0.1256 | [C\text{X4H2}] | 0.8114 |

| worst negatives | prob | best positives | prob |
|-----------------|------|----------------|------|
| [C\text{X4H3}] | 0.5227 | [C\text{X4H2}][\text{C}] | 0.6603 |
| [C\text{X4H3}][\text{CX3}] | 0.3269 | [C\text{X4H2}][\text{C}] | 0.6832 |
| [C\text{X4H2}][\text{C}] | 0.2898 | [C\text{X4H2}][\text{C}] | 0.745 |
| [C\text{X4H2}] | 0.2509 | [C\text{X4H2}][\text{C}] | 0.7567 |
| [C\text{X4H2}][\text{C}] | 0.1956 | [C\text{X4H2}][\text{C}] | 0.795 |
| [C\text{X4H2}] | 0.1729 | [C\text{X4H2}][\text{C}] | 0.7987 |
| [C\text{X4H2}] | 0.1534 | [C\text{X4H2}][\text{C}] | 0.8047 |
| [C\text{X4H2}] | 0.1256 | [C\text{X4H2}][\text{C}] | 0.8114 |

------------------------------------------------------------------------------------------------------------------------
| Best Positives | Prob | Best Negatives | Prob |
|---------------|------|----------------|------|
| [cH][cH]      | 0.9129 | [cH]           | 0.773 |
| [#6H1][#6H1] | 0.9086 | [cX3H1][{cX3H1}][cX3H0] | 0.7308 |
| [#6X3][#6X3] | 0.9998 | [cX3H1]=(O)[#6] | 0.9664 |
| [cX3H1]=(O)[#6] | 0.9129 | [cH][#6H1] | 0.9086 |
| [cX3H1]=[cX3H1][cX3H1] | 0.8774 | [cX3H1]=[cX3H1][cX3H1] | 0.848 |
| [#6X3H1][#6X3H0] | 0.773 | [cX3H1]=[cX3H1][cX3H0] | 0.7308 |

| Worst Negatives | Prob | Worst Positives | Prob |
|-----------------|------|-----------------|------|
| [#8]=[#6]=[#6X3] | 0.7761 | [#8]=[#6]=[#6] | 0.7761 |
| [cX3]=([OX1])C | 0.5206 | [#8]=[#6]=[#6]=[#6X3] | 0.5884 |
| [cX3]=([OX1])O | 0.4888 | [cX3]=([OX1])O | 0.4678 |
| [cX3]=([OX1])O | 0.4635 | [cX3]=([OX1])O | 0.4678 |
| [#8]=[#6H]=[#6X3] | 0.3871 | [#8]=[#6H]=[#6X3] | 0.3871 |
| [cX3]=([cX3])2 | 0.3216 | [cX3]=([cX3])2 | 0.3216 |
| [cX3]=([cX3])2 | 0.3098 | [cX3]=([cX3])2 | 0.3098 |
| O=[cX3H1][#6H1] | 0.285 | [cX3]=([cX3])2 | 0.285 |
Example 185 true smiles: CC=C=CO formula: C5H10O
Index of correct structure: 0 of 65
True structure loss: 0.007174
True structure:

![Structural formula]

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

| Structure | Prob  |
|-----------|------|
| [CX4H3]   | 1.0  |
| [CX4H3][CX3] | 1.0  |
| [CX4H3][#6H0] | 0.998|

Top predicted substructures

| Substructure         | prob |
|----------------------|------|
| [#6H3][#6][#6]       | 0.9927|
| [#8][#6H2][#6H1][#6H0] | 0.9879|
| [CHX3](=C)C          | 0.9836|
| Structure | Prob | Structure | Prob |
|----------|------|----------|------|
| [CX4H3][CX3H0] | 0.9978 | [OX2H1][CX4H2][#6X3H0] | 0.2203 |
| [CX4H3][#6] | 0.9962 | [CX4H2][CX3H1] | 0.2134 |
| [CX4H3][#6H3][#6H0] | 0.998 | [CX4H3][#6H3][#6][#6H3] | 0.1816 |
| [OH][CX4H1] | 0.1804 | [OH][CX4H1] | 0.1365 |
| [#8][#6][#6X3] | 0.1291 | [#6][#6H1][#6H1] | 0.1207 |
| [#6H3][#6X3H0][#6H2] | 0.1066 | [CX3H1][#6H2][#6H2] | 0.1033 |
| [CX3H1][#6][#6H3] | 0.1033 | [#8][#6H3][#6H3][#6X3] | 0.1066 |
| [OH][CX4H3] | 0.0977 | [#8][#6H3][#6H3][#6X3] | 0.0977 |

**Best Positives**

| Structure | Prob | Structure | Prob |
|----------|------|----------|------|
| [CX4H3] | 1.0 | [CX2H0][#6][CX3H0] | 0.0 |
| [CX4H3][CX3] | 1.0 | [CX2H0][#6][CX4H2] | 0.0 |
| [#6H3][#6H0] | 0.998 | CCC#CC#C | 0.0 |
| [CX4H3][CX3H0] | 0.9978 | [CX2H0][#6][CX3H1][#6H0] | 0.0 |
| [CX4H3][#6] | 0.9962 | C#C=CC=C#C | 0.0 |
| [#6H3][#6][#6] | 0.9927 | [CX2H0][#6][CX2H0] | 0.0 |
| [8][#6H2][#6H1][#6H0] | 0.9879 | [#6X2][#6H1][#6X2] | 0.0 |
| [CX3][#6][C] | 0.9836 | [CX2H0][#6][CXH0] | 0.0 |
| [CX4H2][#6][#6][O] | 0.9772 | [OX1H][#6][CX3H0][CX2H0][#CX2H1] | 0.0 |
| [#6H1] | 0.9699 | [OX1H][#6][CX2H0] | 0.0 |

**Best Negatives**

| Structure | Prob | Structure | Prob |
|----------|------|----------|------|
| [OX2H1][CX4H2][#6X3H0] | 0.2203 | [CX4H2][#6H0][#6H2] | 0.3513 |
| [CX4H2][CX4H2] | 0.2134 | [CX4H3][#6H1][#6H3] | 0.5218 |
| [#6H3][#6][#6][#6H3] | 0.1816 | [CX4H2][OX2H1][CX3H1] | 0.6781 |
| [OH][CX4H1] | 0.1804 | [CX4H3][CX3H0][CX4H3] | 0.7829 |
| [#8][#6][#6][#6X3] | 0.1365 | O[CX4H2][CX3H1] | 0.7858 |
| [#6][#6H1][#6H1] | 0.1291 | [#6X3][#6H2][#8] | 0.8034 |
| [CX4H3][CX3H1] | 0.1207 | [CX3H0][#6][CX3H1][#6H3][#6H3] | 0.8119 |
| [#6H3][#6X3H0][#6H2] | 0.1066 | [CX4H2][CX3][#6][#6H3][#6H3][#6X3] | 0.8354 |
| [CX3H1][#6][#6H3][#6H2] | 0.1033 | [#8][#6H2][#6][#6H3][#6X3] | 0.8539 |
| [OH][CX4H3] | 0.0977 | [#8][#6H2][#6H1] | 0.8616 |
Example 186 true smiles: CC(C)CC=O formula: C5H10O
Index of correct structure: 0 of 65
True structure loss: 0.01415
True structure:

![Structure Image]

Experimental 13C NMR (solvent: CDCl3)

![13C NMR Graph]

Experimental 1H NMR (solvent: D2O)

![1H NMR Graph]

Top predicted structures (loss):

- Top predicted substructures

  - [CX4H3] [CX3H1]([=O][#6]) [CX3]([=OX1])C
  - prob: 0.9998 0.9997 0.9995
  - Index: 0 0 0
  - Loss: 0.01415 0.030954 0.048717

- [CX4H3] [CX4H1]([CX4H3])([CX4H3])[CX4H2]
  - prob: 0.9998
  - Index: 0
  - Loss: 0.067829

- [CX4H3] [CX4H1]([CX4H3])([CX4H3])[CX4H2]
  - prob: 0.9998
  - Index: 0
  - Loss: 0.071756

- [CX4H3] [CX4H1]([CX4H3])([CX4H3])[CX4H2]
  - prob: 0.9998
  - Index: 0
  - Loss: 0.081982

- [CX4H3] [CX4H1]([CX4H3])([CX4H3])[CX4H2]
  - prob: 0.9998
  - Index: 0
  - Loss: 0.100525

- [CX4H3] [CX4H1]([CX4H3])([CX4H3])[CX4H2]
  - prob: 0.9998
  - Index: 0
  - Loss: 0.156875

- [CX4H3] [CX4H1]([CX4H3])([CX4H3])[CX4H2]
  - prob: 0.9998
  - Index: 0
  - Loss: 0.172101

- [CX4H3] [CX4H1]([CX4H3])([CX4H3])[CX4H2]
  - prob: 0.9998
  - Index: 0
  - Loss: 0.176257
best positives

| Compound               | Prob   |
|------------------------|--------|
| [CX4H3]               | 0.9998 |
| [CX3H1] (=[OX1](C=C=C)C) | 0.9997 |
| [CX4H3][#6]            | 0.9994 |
| [CX4H1][#6][#6]        | 0.9985 |
| [CX4H2][#6H0][#6H1]    | 0.9165 |
| [CX3H0][#6H1][#6H2]    | 0.8287 |

worst negatives

| Compound               | Prob   |
|------------------------|--------|
| [CX4H2][#6H0][#6H1]    | 0.2868 |
| [O=CX3]                 | 0.2263 |
| [CX4H1][#6H0][#6H1]    | 0.1848 |
| [CX4H2][#6H0][#6H1]    | 0.0991 |
| [CX4H2][#6H1][#6H2]    | 0.0932 |
| [CX4H2][#6H0][#6H1]    | 0.0659 |
| [CX4H1][#6H1]          | 0.0595 |

---

[Note: The table entries represent chemical structures and their probabilities as computed by a certain algorithm or method.]
Example true smiles: OCC1CC1 formula: C5H10O
Index of correct structure: 0 of 65
True structure loss: 0.010513
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

Top predicted substructures

| Substructure                     | prob   |
|----------------------------------|--------|
| [CX4H2][#6][#6]                  | 1.0    |
| [CX4H2][#6][0]                   | 0.9711 |
| [OX2H1]                          | 0.9696 |
| [CX4H2][CX4H2]                   | 0.8752 |
| [CX4H1][#6][#6]                  | 0.8618 |
| [CX4H2][CX4H2]                   | 0.8589 |
| Structure | best positives | prob | worst negatives | prob |
|-----------|----------------|------|-----------------|------|
| OCC(CH2)  | OCC(CH2)(CX4H2)(CX4H1) | 0.9691 | OCC(CH2)(CX4H2)(CX4H1) | 0.9691 |
| C1CCC1    | C1CCC1         | 0.8499 | C1CCC1         | 0.8499 |

**Best Positives**

| Structure | prob | Best Negatives | prob |
|-----------|------|----------------|------|
| (CX4H2)(#6)[O] | 1.0 | C=C=CCOC =C=CCOC | 0.0   |
| [OX2H1]   | 0.9696 | [OX2H1]([CX4H2])([CX4H2]) | 0.0   |
| OCC(CH2)  | 0.9691 | (CX4H2)(CX4H2)(OX2H1)(CX4H2) | 0.0   |
| [CX4H1]   | 0.8369 | (CX4H2)(#6)(CX4H2)(CX4H2) | 0.0   |
| [CX4H2]   | 0.8499 | [CX4H2](CX4H2)(CX4H2) | 0.0   |
| C1CCC1    | 0.8499 | [CX4H2](CX4H2)(OX2H1) | 0.0   |

**Worst Negatives**

| Structure | prob | Worst Positives | prob |
|-----------|------|-----------------|------|
| [CX4H2]   | 0.6896 | [CX4H2](O)(CX4H2) | 0.4939 |
| [#8](#6)[O] | 0.5911 | [CX4H2](#6)(CX4H2)(CX4H2)(CX4H2) | 0.6199 |
| [O](#6)   | 0.3961 | [CX4H2](#6)(CX4H2)(CX4H2) | 0.63   |
| [OX2H1]   | 0.2758 | [CX4H2](#6)(CX4H2)(CX4H2) | 0.6596 |
| [CH2X4]   | 0.1841 | [CX4H2](#6)(CX4H2)(CX4H2) | 0.6865 |
| [CX4H3]   | 0.1803 | [CX4H2](#6)(CX4H2)(CX4H2) | 0.7794 |
| [OX4H2]   | 0.1787 | [CX4H2](#6)(CX4H2)(CX4H2) | 0.8243 |
| [OX2H1]   | 0.1749 | [CX4H2](#6)(CX4H2)(CX4H2) | 0.8369 |
| [C1CCC1]  | 0.1528 | [OX2H1](#6)(CX4H2)(CX4H2) | 0.8499 |
| [CX4H2]   | 0.1224 | [OX2H1](#6)(CX4H2)(CX4H2) | 0.8589 |

---
Example 188 true smiles: C=CC(C)CO formula: C5H10O
Index of correct structure: 0 of 65
True structure loss: 0.008143
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

Top predicted substructures

prob

[OX2H1]
1.0

[6H1]
0.9999

[#6X3]=[#6][#6][#6H3]
0.9972

[OX2H1]
0.9912

[#8H][#6H2][#6H1]
0.981
| Structure         | Prob. | Structure                  | Prob. |
|-------------------|-------|----------------------------|-------|
| [CH3]CC[OH]       | 0.9591| [CX3H][CX4H]               | 0.9580|
| [CHX3](=C)C       | 0.9994| [CHX3][CX4H]               | 1.0   |
| [CX4H3][#6]      | 1.0   | [CHX3][CX4H]               | 0.9994|
| [CX4H3]=#6H2]    | 0.9994| [CX4H3][#6H1][#6H2]        | 0.9998|
| [CX4H3][#6H1]    | 0.9999| [CX4H3][#6H1][#6H2]        | 0.9994|
| [CX4H3][#6H2]    | 0.9999| [CX4H3][#6H1][#6H2]        | 0.9994|

**Best Positives**

| Structure         | Prob. | Structure                  | Prob. |
|-------------------|-------|----------------------------|-------|
| [CX4H3][#6H1]    | 0.9999| [CX4H3][#6H1][#6H2]        | 0.9998|
| [CX4H3][#6H2]    | 0.9997| [CX4H3][#6H1][#6H2]        | 0.9997|
| [CX4H3][#6H3]    | 0.9997| [CX4H3][#6H1][#6H2]        | 0.9997|

**Worst Negatives**

| Structure         | Prob. | Structure                  | Prob. |
|-------------------|-------|----------------------------|-------|
| [CHX3]=#6H2]     | 0.3084| [CHX3]=#6H2]               | 0.2143|
| [CHX3][#6H3]     | 0.1368| [CHX3][#6H3]               | 0.1465|
| [CHX3][#6H1]     | 0.1328| [CHX3][#6H1]               | 0.1456|
| [CHX3][#6H2]     | 0.1287| [CHX3][#6H2]               | 0.1391|
| [CHX3][#6H3]     | 0.1243| [CHX3][#6H3]               | 0.1284|

| Structure         | Prob. | Structure                  | Prob. |
|-------------------|-------|----------------------------|-------|
| [CHX3][#6H1][#6H0]|0.1183 | [CHX3][#6H1][#6H0]        | 0.1183|

---

Note: The structures are represented in a simplified form for clarity.
Example 189 true smiles: CN(C)=O formula: C3H7NO
Index of correct structure: 0 of 59
True structure loss: 0.008732
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

Top predicted substructures                   prob
[#7X3][#6H3]                                  0.9998          [#7X3H0]                                      0.7694
[#6H3][#7]                                    0.9995          [#8]=[#6H1][#7]                               0.7625
[CX4H3]                                       0.9971          [#6H1]                                        0.7588
| Best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| #6X3][#7][#6X3] | 0.9385 | #6X3][#7X3H0] | 0.9385 |
| #6X3][#7X3H0] | 0.9675 | #6X3][#7X3H0] | 0.9675 |

| Worst negatives | prob | worst positives | prob |
|-----------------|------|-----------------|------|
| #7X3][#6X3] | 0.9998 | #7X3][#6X3] | 0.9998 |
| #6X3][#7] | 0.9995 | #7X3H0] | 0.7071 |
| #6X3][#7] | 0.9385 | #6X3][#7] | 0.9385 |
| [CX4H3][#6X3H0] | 0.8575 | [CX4H3][#6X3H0] | 0.8575 |
| #7X3H0] | 0.7694 | #7X3H0] | 0.7694 |
| #6H1] | 0.7588 | #6H1] | 0.7588 |
| #6X3H1][#O1H0][#NX3H0] | 0.707 | #6X3H1][#O1H0][#NX3H0] | 0.707 |
| #6X3H1][#7X3H0] | 0.6201 | #6X3H1][#7X3H0] | 0.6201 |

| Contextual | prob | Contextual | prob |
|------------|------|------------|------|
| [CX3H][#O1H0] | 0.4001 | #6X3H1][#7X3H0] | 0.6201 |
| #OX2H1] | 0.3915 | #6X3H1][#7X3H0] | 0.6201 |
| [CX4H3][#O1H0] | 0.3771 | #6X3H1][#7] | 0.7588 |
| #6X1][#7] | 0.3371 | #6X1][#7] | 0.7588 |
| [OX1H0] | 0.3302 | [OX1H0] | 0.3302 |
| [O#8][#6X3][#7] | 0.2687 | [O#8][#6X3][#7] | 0.2687 |
| [O#7][#6X3][#7] | 0.2288 | [O#7][#6X3][#7] | 0.2288 |
| [O#7][#6X3][#7] | 0.192 | [O#7][#6X3][#7] | 0.192 |
| [O#7][#6X3][#7] | 0.1903 | [O#7][#6X3][#7] | 0.1903 |
Example 190 true smiles: NCC(=O)O formula: C2H5NO2
Index of correct structure: 0 of 47
True structure loss: 0.010644
True structure:

![Chemical Structure](image)

Experimental 13C NMR (solvent: D2O)

---

Experimental 1H NMR (solvent: D2O)

---

Top predicted structures (loss):

1. 0.010644
2. 0.036594
3. 0.045586
4. 0.046442
5. 0.060313

---

Top predicted substructures

| Substructure | prob |
|--------------|------|
| [CX3]=([OX1])C | 0.9844 |
| [OX1]=([OX1])C | 0.9755 |
| [#7X3]=[#6H2] | 0.8644 |
| [OX1]=([OX1])C | 0.8606 |
| [#8]=[#6][#8] | 0.8547 |
| best positives | prob     | worst negatives | prob     |
|----------------|----------|-----------------|----------|
| [CX3]=([OX1])C | 0.9844   | [OX1]=([OX1])H2 | 0.9588   |
| [#7]=[#6][#8]  | 0.9195   | [O#3]=([O#3])H2 | 0.6126   |
| [#7]=[#6][#8]  | 0.9154   | [O#3]=([O#3])H2 | 0.6126   |
| [CX4H2][CX3]=O | 0.8698   | [OX4][#3][#6X3] | 0.7404   |
| [CX3][=([OX1])O] | 0.8606  | [OX4][#3][#6X3] | 0.7404   |
| [CX3][=([OX1])O] | 0.8547  | [OX4][#3][#6X3] | 0.7404   |
| [CX3][=([OX1])O] | 0.8547  | [OX4][#3][#6X3] | 0.7404   |
| [CX3][=([OX1])O] | 0.8547  | [OX4][#3][#6X3] | 0.7404   |
| [CX3][=([OX1])O] | 0.8547  | [OX4][#3][#6X3] | 0.7404   |
| [CX3][=([OX1])O] | 0.8547  | [OX4][#3][#6X3] | 0.7404   |

| worst negatives | prob     | worst positives | prob     |
|-----------------|----------|-----------------|----------|
| [OX4][#7][#6X3] | 0.2507   | [CX3][=([OX1])O] | 0.8606   |
| [OX4][#7][#6X3] | 0.2507   | [CX3][=([OX1])O] | 0.8606   |
| [OX4][#7][#6X3] | 0.2507   | [CX3][=([OX1])O] | 0.8606   |
| [OX4][#7][#6X3] | 0.2507   | [CX3][=([OX1])O] | 0.8606   |
| [OX4][#7][#6X3] | 0.2507   | [CX3][=([OX1])O] | 0.8606   |
| [OX4][#7][#6X3] | 0.2507   | [CX3][=([OX1])O] | 0.8606   |
| [OX4][#7][#6X3] | 0.2507   | [CX3][=([OX1])O] | 0.8606   |

*--------------------------------------------------------------------------------------------------------*
Example 191 true smiles: O=C=CC(N)=O formula: C3H5NO
Index of correct structure: 0 of 46
True structure loss: 0.014223
True structure:

\[
\begin{align*}
&\text{Experimental } ^{13}\text{C NMR (solvent: D}_2\text{O)} \\
&\text{Experimental } ^{1}\text{H NMR (solvent: D}_2\text{O)} \\
\end{align*}
\]

Top predicted structures (loss):

\[
\begin{align*}
&\text{Top predicted substructures} & \text{prob} \\
&\{\#6X3\} & 0.014223 \\
&\{\#6X3\} & 0.046104 \\
&\{\#6X3\} & 0.05967 \\
&\{\#6X3\} & 0.060387 \\
&\{\#6X3\} & 0.081697 \\
\end{align*}
\]

Top predicted substructures prob

\[
\begin{align*}
&\{\#6X3\} & 0.9808 \\
&\{\#6X3\} & 0.9226 \\
&\{\#6X3\} & 0.896 \\
&\{\#6X3\} & 0.8817
\end{align*}
\]
| best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| [#7X3H2]       | 0.9299 | O=#6][#6][#6X3] | 0.807 |
| [#6H1]         | 0.9268 | [#7H2][#6H0]   | 0.7796 |
| [#6X3][#6X3]   | 0.9808 | CCCOCOC=C       | 0.0  |
| [OX2H0][OX4H1][CX4H1] | 0.9431 | [OX1H0][#CX2H0][#CX2H0] | 0.0  |
| [#7X3H2]       | 0.9299 | [OX2H0][CX3H0][CX4H1][CX4H1][CX4H1][CX4H1] | 0.0  |
| [#6H1]         | 0.9268 | [OX2H0][CX4H2][CX4H1][CX4H1][CX4H1][CX4H1] | 0.0  |
| [#6X3H2]       | 0.9226 | [OX2H0][#CX2H1][CX4H2] | 0.0  |
| [#6X3H1][#6X3H0] | 0.896  | [OX2H0][#CX2H1][CX4H0] | 0.0  |
| [CHX3][=C]     | 0.8817 | [#6H2][#6][#6X2] | 0.0  |
| O=#6][#6][#6X3] | 0.807  | [OX2H0][CX4H2][CX4H1][CX4H1][CX4H1] | 0.0  |
| [#7H2][#6H0]   | 0.7796 | [OX4H1][OX2H0][CX4H2][CX2H0] | 0.0  |
| worst negatives | prob | worst positives | prob |
| [OX2H1]        | 0.5556 | [#7][#6X3H0][#6X3H1] | 0.5274 |
| [CHX3]=CHX3    | 0.4558 | [CX3H1]==[CX3H2][CX3H0] | 0.6151 |
| [#8][#6H0][#6H1] | 0.4225 | [#8][#6H0][#6H1] | 0.7066 |
| [#8][#6][#6X3] | 0.4083 | [#8][#6X0][#6H1] | 0.7112 |
| [OX2H0][#OX1]O | 0.3966 | [#7][#6][#6][#6X3] | 0.7204 |
| [#6][#6][#8]  | 0.3516 | [#7][#6][#6X3] | 0.7261 |
| [OX2H1][=CX3H1][CX3H0] | 0.3497 | [CX3][=OX1]C | 0.7582 |
| O=#6][#6][#6X3] | 0.3404 | O=C[CX3H] | 0.7693 |
| [#7][#6][#6X3] | 0.3309 | [#7H2][#6H0] | 0.7796 |
Example 192 true smiles: clonconel formula: C4H4N2
Index of correct structure: 0 of 46
True structure loss: 0.017585
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

0.017585  0.018729  0.023258  0.059451  0.061355
0.061403  0.061855  0.067281  0.072427  0.073715

Top predicted substructures

| Substructure | prob     | Substructure                  | prob     |
|--------------|----------|-------------------------------|----------|
| [#6H1]       | 0.9993   | [#7][#6][#6][#6X3]           | 0.0105   |
| [#6X3][#6X3]| 0.977    | [#6H1][#6H1]                 | 0.7943   |
| [cH]         | 0.9504   | [#6H1][#7][#6H1]             | 0.7546   |
| best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| [#6H1]         | 0.993| [OX2H0][CX4H2][CX4H2][CX4H1][CX4H1] | 0.0  |
| [#6X3]         | 0.977| [OX2H0][CX4H2][CX4H0][OX2H0]         | 0.0  |
| [cH]           | 0.954| [OX2H0][CX4H2][CX4H2][CX4H1][OX2H0] | 0.0  |
| [#7][#6][#6X3] | 0.8217| [OX2H1][CX4H1][CX4H1][CX4H1][CX4H1] | 0.0  |
| [#7][#6][#6X3] | 0.8105| [#8][#6H1][#6H2][#6H1]=[#8]          | 0.0  |
| [#6H1][#6H1]  | 0.7943| [CX4H0][(OX2H1)][(CX4H3)][(CX4H2)][CX4H1] | 0.0  |
| [#6H1][#7][#6H1]| 0.7546| [CX4H1][(OX2H0)][(CX4H2)][CX4H0]    | 0.0  |
| [cX3H1][cX2H0][cX3H1]| 0.6421| [OX2H0][CX4H2][CX4H1][CX4H1][CX4H1] | 0.0  |
| [cX3H1][cX2H1][cX3H1]| 0.6359| [OX2H0][CX4H1][(CX4H1)][CX4H1][CX4H3] | 0.0  |
| worst negatives | prob | worst positives | prob |
| [#6X3][#6X3][#6X3][#6X3] | 0.7448| [#7][#6][#6][#6][#7] | 0.2938 |
| [#6X3][#6H0]  | 0.6703| [#7][#6H1][#7]       | 0.339 |
| [cX3H1][cX3H1][cX3H0]| 0.5854| [#7][#6][#7]         | 0.4802|
| [#7][#6X3][#6X3H1]| 0.5144| [#6X3][#7][#6X3]    | 0.6103|
| [cX3H1][cX2H0][cX3H0]| 0.418| [cX3H1][{cX3H1}][cX3H1] | 0.6359|
| [#7][#6H0][#6H1]| 0.3842| [cX3H1][{cX2H0}][cX3H1] | 0.6421|
| [#7][#7]       | 0.3635| [#6H1][#7][#6H1]    | 0.7546|
| [#6X3][#7X2][#6X3]| 0.3396| [#6H1][#6H1]         | 0.7943|
| [#6H1][#6][#6][#6][#7] | 0.3349| [#7][#6][#6][#6X3]  | 0.8105|
| [#6X3][#6X3][#6X3][#6X3][#6X3][#6X3] | 0.3137| [#7][#6][#6X3]    | 0.8217|
Example 193 true smiles: Cc1ccco1 formula: C5H6O
Index of correct structure: 0 of 45
True structure loss: 0.009801
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

Top predicted substructures
[#6H1] 0.9995 [cX4H3][#6] 0.9925
[cH][cH] 0.999  [cH] 0.991
[#6X3][#6X3][#6X3][#6X3] 0.9983 [cX3H1]{[cX3H1]}[cX3H0] 0.986

0.009801 0.036472 0.141793 0.158804 0.162743

0.16676 0.167792 0.176873 0.178773 0.183347
| Best positives | prob | Best negatives | prob |
|----------------|------|---------------|------|
| [#6X1] | 0.9995 | [#OX1H0]= [#CX3H0][#CX4H1][#CX4H2] | 0.0 |
| [#8] | 0.999 | [#CX3H0][#CX4H2][#CX4H1] | 0.0 |
| [#6X3][#6X3] | 0.9983 | CCC#CC#C | 0.0 |
| [#6X3][#6X3] | 0.998 | [#OX2H1][#CX4H2][#CX4H1] | 0.0 |
| [#6X3][#6X3] | 0.996 | [#CX4H0][#NX3H1][#CX4H0] | 0.0 |
| [#6X3] | 0.9925 | [#CX4H1][#NX3H0][#CX4H3] | 0.0 |
| [#6X3][#6X3] | 0.986 | [#OX2H1][#CX4H1][#CX4H2][#CX4H1] | 0.0 |
| [#6X3][#6X3] | 0.986 | [#OX2H1][#CX4H1][#CX4H2][#CX4H1] | 0.0 |
| Worst negatives | prob | Worst positives | prob |
| [#6X3][#6X3][#6X3] | 0.3007 | [#8][#6H][#6X3][#6X3] | 0.2426 |
| [#CX4H2][#6X3] | 0.2846 | o[#8] | 0.3342 |
| [#6X3][#6X3][#6X3][#6X3] | 0.2358 | [#8][#6H][#6X3][#6H] | 0.4695 |
| [#6X3][#6X3] | 0.2092 | [CX3][#6][#8] | 0.5045 |
| [#OX2H][#6X3] | 0.1938 | [#6X3][#6X3][#6X3][#6X3] | 0.7585 |
| [#6X3][#6X3][#6X3][#6X3] | 0.1833 | [#8][#6H][#6H] | 0.7773 |
| [#8][#6X3][#6X3][#6X3] | 0.1577 | [CX3][#6][#6H] | 0.844 |
| [#8][#6X3][#6X3][#6X3] | 0.1193 | [CX4H3][#CX4H0][#OX2H] | 0.8865 |
| [7][#6][#6X3][#6X3] | 0.0804 | [#6H][#6H] | 0.9041 |
Example 194 true smiles: CNC(=N)N formula: C2H7N3
Index of correct structure: 0 of 39
True structure loss: 0.012255
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

Top predicted substructures

| Substructure | prob |
|--------------|------|
| [#7X3][#6H3] | 0.9971 |
| [#6H3][#7][#6X3] | 0.9427 |
| [#6H3][#7] | 0.9757 |
| [#7][#6][#7] | 0.8712 |
| [CX4H3] | 0.9644 |
| [#7X3H1] | 0.7627 |
| best positives | prob | worst negatives | prob |
|----------------|------|-----------------|------|
| [\#7X3] [\#6H3] | 0.9971 | [\#7H1] [\#6H0] [\#7X3] [\#6H3] | 0.5542 |
| [\#6H3] [\#7] | 0.9757 | [\#7] [\#6H0] [\#6H1] | 0.2709 |
| [\#7] [\#6H0] [\#6H3] | 0.9543 | [\#7] [\#6H0] [\#6H2] [\#7] | 0.27 |
| [\#7X3H] [\#7X3] [\#6X3] | 0.1657 | [\#7] [\#7] [\#6H0] | 0.1581 |
| [\#7X3H] [\#7X3] [\#6X3] | 0.0491 | [\#7] [\#6H0] [\#7] | 0.1494 |
| [\#7] [\#6H0] | 0.1383 | [\#7] [\#6H0] [\#7] | 0.1283 |
| [\#7X3H] [\#6X3] | 0.0279 | [\#7] [\#6H0] [\#7] | 0.0279 |
Example 195 true smiles: \(\text{CC(O)C(=O)O}\) formula: \(\text{C}_3\text{H}_6\text{O}_3\)

Index of correct structure: 0 of 38

True structure loss: 0.008204

True structure:

![Chemical structure diagram]

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

- \([\text{CX}_4\text{H}_3]\) with prob 0.9999
- \([\text{CX}_4\text{H}]\text{O}\) with prob 0.9812
- \([\#6\text{H}_3][\#6][\#6]\) with prob 0.9767
- \([\text{OX}_2\text{H}1]\) with prob 0.9767
- \([\text{CH}_3][\#6][\#8]\) with prob 0.9437

Top predicted substructures

- \([\text{CX}_4\text{H}_3]\) with prob 0.9999
- \([\#6\text{H}_3][\#6][\#6]\) with prob 0.9967
- \([\text{OX}_2\text{H}1]\) with prob 0.9973
- \([\text{CH}_3][\#6][\#8]\) with prob 0.9912
Example 196 true smiles: O=C(O)CCO formula: C3H6O3
Index of correct structure: 0 of 38
True structure loss: 0.007232
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

| Structure | Prob  |
|-----------|------|
| [CX4H2]{[#6],[#6]} | 0.007232 |
| [#8]=[#6][#8] | 0.079487 |
| [CX3]{[=OX1]}C | 0.085344 |
| [CX4H2]{[#6],[O]} | 0.089833 |
| [CX4H2]{[#6],[#6H2]} | 0.092632 |
| [CX4H2]{[#6],[#6]} | 0.09286 |
| [#8]=[#6][#8] | 0.101687 |
| [OX1]{[O]} | 0.111257 |
| [OX1]{[O]} | 0.112307 |
| [OX1]{[O]} | 0.112885 |

Top predicted substructures

| Substructure | Prob  |
|--------------|------|
| [CX4H2]{[#6],[#6]} | 0.9983 |
| [#8]=[#6][#8] | 0.9953 |
| [CX3]{[=OX1]}C | 0.9952 |
| [CX4H2]{[#6],[#6H2]} | 0.99727 |
| [CX4H2]{[#6],[O]} | 0.9746 |
| [CX4H2]{[#6],[#6]} | 0.9727 |
| [OX1]{[O]} | 0.9628 |
Example 197 true smiles: C(O)(=O)O formula: C3H6O3
Index of correct structure: 0 of 38
True structure loss: 0.010427

True structure:
| Structure                  | Best Positives prob | Best Negatives prob |
|---------------------------|---------------------|---------------------|
| `CX4H3`                   | 0.9954              | `CH3`[#6]#[8]        | 0.9265              |
| `CX4H3`[#6]               | 0.993               | `CX4H3`[CX4H1][OX2H1] | 0.9068              |
| `CX4H3`[#6]               | 0.9999              | `CC=CX4H3`           | 0.0                 |
| `CX4H3`[#6]               | 0.9968              | `C=CX4H3`            | 0.0                 |
| `CX3`{=O}                 | 0.9964              | `CX4H3`{=CX3H2}      | 0.0                 |
| `CX4H3`{=O}               | 0.9954              | `CCC=CX4H3`          | 0.0                 |
| `CX4H3`{=O}               | 0.993               | `#6X2`{=H1}#6X2     | 0.0                 |
| `CX3`{=O}                 | 0.9902              | `CX3H0`{=CX3H2}      | 0.0                 |
| `#6H3`{=O}                | 0.9837              | `CX3H1`{=CX3H1}      | 0.0                 |
| `O#6X4H1`                 | 0.9517              | `CX3H1`{=CX3H1}      | 0.0                 |
| `#8=#6X3`                 | 0.9265              | `CX3H0`{=OX2H0}      | 0.0                 |
| `CX4H3`{=O}               | 0.9068              | `CX3H1`{=OX2H0}      | 0.0                 |

| Structure                  | Worst Positives prob | Worst Negatives prob |
|---------------------------|---------------------|---------------------|
| `CX4H2`{=O}               | 0.2077              | `CH3`{=O}           | 0.3688              |
| `#6H1`{=O}                | 0.1839              | `#8H1`{=OX2H0}      | 0.3703              |
| `CX4H1`{=O}               | 0.1673              | `#8H1`{=OX2H0}      | 0.3781              |
| `#8`{=O}                  | 0.1497              | `[#8][#6]#[8]`       | 0.506               |
| `CX4H1`{=O}               | 0.1369              | `[#8][#6]#[6]`       | 0.6438              |
| `O#6X4H1`                 | 0.1142              | `#8`{=O}#6H0#6H1    | 0.6957              |
| `CX3H0`{=O}               | 0.1111              | `#8`{=O}#6H0#6H1    | 0.7147              |
| `CX4H2`{=O}               | 0.111               | `#8`{=O}#6H0#6H1    | 0.7241              |
| `#8`{=O}                  | 0.1024              | `#6H3`{=O}#6X3      | 0.7607              |
| `CX4H2`{=O}               | 0.1024              | `#6H3`{=O}#6X3      | 0.7709              |
Example 198 true smiles: O=C(CO)CO  formula: C3H6O3
Index of correct structure: 0 of 38
True structure loss: 0.013689
True structure:

Experimental 13C NMR (solvent: N/A)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

0.013689  0.060303  0.063898  0.065795  0.070775
0.072064  0.072961  0.073991  0.081297  0.091921

Top predicted substructures

[OX1H0]=([OX1H0][#6][#6][#6][#6][#6][#6][#6])
 prob 0.9969  0.7643
[OX2H1]  0.9969  0.7369
[OX3H0]  0.9969  0.6381

OCC[CH2]
| Compound | Prob  | Compound | Prob  |
|----------|-------|----------|-------|
| [CX4H2][CX3]=O | 0.8879 | [CX4H2][#6][O] | 0.5832 |
| [CX3H0]//[OX1H0][CX4H2][CX4H2] | 0.8812 | [CX4H2][OX2H1][CX3H0] | 0.567 |

**Best Positives**

| Compound | Prob  | Compound | Prob  |
|----------|-------|----------|-------|
| [CX3](=O)[CX3]C | 0.9996 | CC=CC#CC | 0.0 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.9969 | C=CC=CC| 0.0 |
| [OX2H1] | 0.9841 | C=CCC=C | 0.0 |
| [CX4H2][CX3]=O | 0.9969 | CC=CC#C | 0.0 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 | [#6X2][#6H1][#6X2] | 0.0 |

**Worst Negatives**

| Compound | Prob  |
|----------|-------|
| [#8][#6][#6H2] | 0.5254 |
| [OX2H1][CX4H2][CX3H0][CX4H2] | 0.43 |
| [CX4H2][CC=O] | 0.4289 |
| [#8][#6][#6X3] | 0.2862 |
| [OX2H1][CX4H2][#6X3H0] | 0.4421 |

| Compound | Prob  |
|----------|-------|
| [CX4H2][#6][#6H2] | 0.2629 |
| [#8][#6][#8] | 0.1799 |
| [OX4H2][#6X3H0][#6H2][#8] | 0.5346 |
| [OX2H1][#6X3H0][#6H2][#8] | 0.5346 |
| [OX4H2][#6X3H0][#6H2][#8] | 0.5346 |
| [OX2H1][#6X3H0][#6H2][#8] | 0.5346 |

**Best Negatives**

| Compound | Prob  | Compound | Prob  |
|----------|-------|----------|-------|
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 | [CX4H2][OX2H1][CX3H0] | 0.567 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 | [CX4H2][OX2H1][CX3H0] | 0.567 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 | [CX4H2][OX2H1][CX3H0] | 0.567 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 | [CX4H2][OX2H1][CX3H0] | 0.567 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 | [CX4H2][OX2H1][CX3H0] | 0.567 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 | [CX4H2][OX2H1][CX3H0] | 0.567 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 | [CX4H2][OX2H1][CX3H0] | 0.567 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 | [CX4H2][OX2H1][CX3H0] | 0.567 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 | [CX4H2][OX2H1][CX3H0] | 0.567 |

**Worst Positives**

| Compound | Prob  |
|----------|-------|
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 |

**Worst Negatives**

| Compound | Prob  |
|----------|-------|
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 |
| [OX1H0][CX3H0][#6][CX4H2] | 0.8812 |
Example 199 true smiles: \text{CC(CC)(=O)O} formula: C_4H_4O_2

Index of correct structure: 0 of 32
True structure loss: 0.004418

True structure:

![Chemical Structure](image)

Experimental $^{13}$C NMR (solvent: CDCl$_3$)

![Experimental $^{13}$C NMR](image)

Experimental $^1$H NMR (solvent: D$_2$O)

![Experimental $^1$H NMR](image)

Top predicted structures (loss):  

| Structure                               | prob   | loss      |
|-----------------------------------------|--------|-----------|
| [CX$_4$H$_3$][CX$_2$H$_0$]              | 0.9993 | 0.004418  |
| $[\text{CX}_2\text{C}]$                 | 0.9983 | 0.025247  |
| [CX$_2$H$_0$][#[CX$_2$H$_0$]][CX$_3$H$_0$] | 0.9928 | 0.043696  |
| $[\text{CX}_2\text{H}_0]\text{(#}{\text{CX}_2\text{H}_0}\text{)}$ | 0.9910 | 0.064242  |
| $[\text{CX}_2\text{H}_0]\text{(#}{\text{CX}_2\text{H}_0}\text{)}$ | 0.9928 | 0.072622  |

| Structure | prob   | loss      |
|-----------|--------|-----------|
| $[\text{CX}_4\text{H}_3]$              | 0.9983 | 0.00809   |
| $[\text{CX}_2\text{H}_0]\text{(#}{\text{CX}_2\text{H}_0}\text{)}$ | 0.9983 | 0.099047  |
| $[\text{CX}_2\text{H}_0]\text{(#}{\text{CX}_2\text{H}_0}\text{)}$ | 0.9928 | 0.102659  |
| $[\text{CX}_2\text{H}_0]\text{(#}{\text{CX}_2\text{H}_0}\text{)}$ | 0.9910 | 0.107462  |
| $[\text{CX}_2\text{H}_0]\text{(#}{\text{CX}_2\text{H}_0}\text{)}$ | 0.9928 | 0.115173  |

Top predicted substructures

- $[\text{CX}_4\text{H}_3][\text{CX}_2\text{H}_0]$ 0.9993 $[\#\text{C}]\text{(#}{\#\text{C}}\text{)}$ 0.9138
- $[\text{CX}_2\text{H}_0]\text{(#}{\text{CX}_2\text{H}_0}\text{)}][\text{CX}_3\text{H}_0]$ 0.9983 $[\text{CX}_3]\text{(#}{\text{OX}_1}\text{)}\text{O}$ 0.9026
- $[\text{CX}_2\text{H}_0]\text{(#}{\text{CX}_2\text{H}_0}\text{)}][\text{CX}_3\text{H}_0]$ 0.9928 $[\text{CX}_4\text{H}_3]$ 0.8999

Top predicted substructures

- $[\text{CX}_4\text{H}_3][\text{CX}_2\text{H}_0]$ 0.9993 $[\#\text{C}]\text{(#}{\#\text{C}}\text{)}$ 0.9138
- $[\text{CX}_2\text{H}_0]\text{(#}{\text{CX}_2\text{H}_0}\text{)}][\text{CX}_3\text{H}_0]$ 0.9983 $[\text{CX}_3]\text{(#}{\text{OX}_1}\text{)}\text{O}$ 0.9026
- $[\text{CX}_2\text{H}_0]\text{(#}{\text{CX}_2\text{H}_0}\text{)}][\text{CX}_3\text{H}_0]$ 0.9928 $[\text{CX}_4\text{H}_3]$ 0.8999
| best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| [CX4H3][CX2H0] | 0.9993 | [OX1H0]=[CX3H0][CX4H1][CX4H1][CX4H2] | 0.0 |
| [S][CX2][#C] | 0.9983 | [CX3H1][=][CX3H2][CX4H2] | 0.0 |
| [CX2H0][#][CX2H0][CX3H0] | 0.9928 | [CX3H0][=][CX3H2][CX4H2][CX4H0] | 0.0 |
| [6X3][#][#][6H3] | 0.9964 | [CX3H1][=][CX3H2][NX3H0] | 0.0 |
| [CX4H3][#6] | 0.9561 | [CX3H2][=][CX3H1][CX4H0][OX2H1] | 0.0 |
| [#][=][#][#] | 0.9138 | [CX3H1][=][CX3H2][CX4H0] | 0.0 |
| [CX1][=][OX1][O] | 0.9026 | [CX4H1][NX3H2][CX4H2][CX3H1] | 0.0 |
| [CX4H3] | 0.8999 | [CX3H1][=][CX3H2][CX3H0] | 0.0 |
| [#6H3][#6H0] | 0.7969 | [CX3H0][=][CX3H2][CX4H2][CX4H2] | 0.0 |
| [CX1][=][OX1][C] | 0.7356 | [CX3H0][=][CX3H2][CX4H3][CX4H2] | 0.0 |

| worst negatives | prob | worst positives | prob |
|-----------------|------|-----------------|------|
| [#8][#6][#6][#8] | 0.2293 | [OX2H1] | 0.5969 |
| [#6H1] | 0.1463 | [CX3][=][O][OX2H1] | 0.7048 |
| [CX3H0][=][OX1H0][OX2H1][CX3H0] | 0.1024 | [CX3][=][OX1][C] | 0.7356 |
| [#8][#6][#6H2] | 0.0799 | [#6H3][#6H0] | 0.7969 |
| [#8][#6][=][#8] | 0.0743 | [CX4H3] | 0.8999 |
| [CX4H2][CX3]=O | 0.07 | [CX3][=][OX1][O] | 0.9026 |
| [CX3H][O] | 0.0694 | [#8][=][#6][#8] | 0.9138 |
| [#6H1][#6H2] | 0.0576 | [CX4H3][#6] | 0.9561 |
| [CX3H0][=][OX1H0][OX2H1][CX4H2] | 0.0632 | [#6X3][#][#][#][#6H3] | 0.9864 |
| [#8][#6][#6X3] | 0.0562 | [CX2H0][=][CX2H0][CX3H0] | 0.9928 |
Example 200 true smiles: cloconel formula: C5H5N
Index of correct structure: 0 of 27
True structure loss: 0.016342
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

Top predicted substructures                  prob
[#6H1]                                        0.9998          [cX3H1]{{cX3H1}}{cX3H1}                       0.9991
[cH][cH]                                      0.999           [#7]{#6}{#6}{#6X3}                            0.9837
[#6X3]{#6X3}                                  0.9984          [cH]                                          0.9816
best positives | prob | best negatives | prob
--- | --- | --- | ---
[#6X3]|#6X3]|#6X3]|#6X3] | 0.9969 | [#6]|#6]|#6]|#6]|#7] | 0.9665
[#7]|#6]|#6X3] | 0.9914 | [cX3H1]|(cX3H1)|cX3H0] | 0.9503
worst negatives | prob | worst positives | prob
--- | --- | --- | ---
[cX3H1]|(cX3H1)|cX3H0] | 0.9503 | [#6H1]|#7]|#6H1] | 0.5928
[#6X3H1]|#6X3H0] | 0.9077 | [#6X3]|#7]|#6X3] | 0.7718
[#6X3H1]|#6X3H1]|#6X3H0]|#6X3H1] | 0.8683 | [cX3H1]|(nX2H0)|cX3H1] | 0.9044
[#6]|#6]|#6]|#6]|#6]|#6] | 0.7428 | [#6H1]|#6H1] | 0.9494
[#7]|#6X3H0]|#6X3H1] | 0.6768 | [#6]|#6]|#6]|#6]|#6]|#7] | 0.9665
[#7]|#6H0]|#6H1] | 0.6699 | [cH] | 0.9816
[#7X3H2] | 0.4489 | [#7]|#6]|#6]|#6X3] | 0.9837
[#7H2]|#6H0] | 0.2631 | [cX3H1]|(cX3H1)|cX3H1] | 0.9891
[cX3H1]|(nX2H0)|cX3H0] | 0.2243 | [#7]|#6]|#6X3] | 0.9914
[cH]cO | 0.1757 | [#6X3]|#6X3]|#6X3]|#6X3] | 0.9969
------------------------------------------------------------------------------------------------------------------------
Example 201 true smiles: c1ccncc1 formula: C5H5N
Index of correct structure: 0 of 27
True structure loss: 0.015397
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

Top predicted substructures

| Substructure          | prob  | Substructure          | prob  |
|-----------------------|-------|-----------------------|-------|
| [6H1]                 | 0.9999| [#7] [#6] [#6X3]      | 0.9972|
| [cH][cH]              | 0.9996| [cH]                  | 0.9843|
| [#6X3] [#6X3]         | 0.9984| [#6] [#6] [#6] [#6]   | 0.9775|
| best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| 0.9968 | 0.9967 | 0.9968 | 0.9967 |
| 0.9999 | 0.9996 | 0.9994 | 0.9996 |
| 0.9968 | 0.9967 | 0.9968 | 0.9967 |
| 0.9999 | 0.9996 | 0.9994 | 0.9996 |
| 0.9999 | 0.9996 | 0.9994 | 0.9996 |
| 0.9999 | 0.9996 | 0.9994 | 0.9996 |
| 0.9999 | 0.9996 | 0.9994 | 0.9996 |
| 0.9999 | 0.9996 | 0.9994 | 0.9996 |
| 0.9999 | 0.9996 | 0.9994 | 0.9996 |
| 0.9999 | 0.9996 | 0.9994 | 0.9996 |
| 0.9999 | 0.9996 | 0.9994 | 0.9996 |

| worst negatives | prob | worst positives | prob |
|-----------------|------|-----------------|------|
| 0.949 | 0.9078 | 0.8516 | 0.7913 |
| 0.8868 | 0.8868 | 0.8868 | 0.8868 |
| 0.9512 | 0.9512 | 0.9512 | 0.9512 |
| 0.9775 | 0.9775 | 0.9775 | 0.9775 |
| 0.9843 | 0.9843 | 0.9843 | 0.9843 |
| 0.9967 | 0.9967 | 0.9967 | 0.9967 |
| 0.9968 | 0.9968 | 0.9968 | 0.9968 |
Example 202 true smiles: C1CCOC1 formula: C4H8O
Index of correct structure: 0 of 22
True structure loss: 0.006421
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

Top predicted substructures

[0.006421] [0.068256] [0.093125] [0.121414] [0.135288]
[0.137885] [0.150093] [0.150942] [0.153123] [0.155557]
| OCC[CH2]            | 0.9472 | [CH2X4](O)[CX4H2]          | 0.741 |
|---------------------|--------|-----------------------------|-------|
| [OX2H0][CX4H2][CX4H2][CX4H2] | 0.9278 | [CX4H2][CX4H2][CX4H2][CX4H2] | 0.7930 |
| best positives      | prob   | best negatives              | prob  |
| [CX4H2][(#6)][#6]  | 0.9896 | [OX1H0]=[CX3H0][CX4H1][CX4H2][CX4H2] | 0.0   |
| [#8][#6][#6H2]     | 0.9911 | [#6H3][#6H1][#6H1]=[#7]     | 0.0   |
| [CX4H2][(#6)][O]   | 0.9606 | [CX2H0][#CX2H1][CX3H0]     | 0.0   |
| [CX4H2][CX4H2]     | 0.9278 | [CX3H0][=CX3H1][CX4H1][CX2H0] | 0.0   |
| [CX4H2][O][CX4H2]  | 0.8752 | C=CC=CC=C                         | 0.0   |
| [OX2H0][CX4H2][CX4H2][CX4H2] | 0.7862 | [#7][#6]=[#6][#6][#6]=[#7] | 0.0   |
| [OX2H1]            | 0.7903 | [CX3H0][=CX3H2][CX4H2][CX4H1] | 0.0   |
| worst negatives    | prob   | worst positives              | prob  |
| [#8][#6][#6][#6][#6][#6] | 0.1766 | C1OCCC1                     | 0.295 |
| [#8][#6][#6][#8]   | 0.1541 | [CX4H2][OX2H0][CX4H2]       | 0.6094|
| [CX4H3][#6]        | 0.154  | [CX4H2][OX2H0][CX4H2]       | 0.7178|
| [OX2H1]            | 0.1356 | [CH2X4](O)[CX4H2]           | 0.741 |
| [CH3][#6][#8]      | 0.1124 | [CX4H2][CX4H2][CX4H2][CX4H2] | 0.7903|
| [#6H1][#6H2]       | 0.111  | [CX4H2][CX4H2][CX4H2][CX4H2] | 0.796 |
| [#8][#6][#6H2][#8] | 0.1057 | [CH2X4](O)[OX2H0][CX4H2]    | 0.8612|
| [#6H1]             | 0.1035 | [CX4H2][CX4H2]             | 0.8752|
| CCCCCC              | 0.1033 | [OX2H0][CX4H2][CX4H2][CX4H2] | 0.9278|
| [OX2H0][CX4H1][CX4H3] | 0.0868 | OCC[CH2]                    | 0.9472|

------------------------------------------------------------------------------------------------------------------------
Example 203 true smiles: CCCC=O formula: C4H8O
Index of correct structure: 1 of 22
True structure loss: 0.030431
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

Top predicted substructures

0.020199
0.030431
0.107416
0.143562
0.167901

0.171477
0.174714
0.183314
0.188553
0.211463

Top predicted substructures

[CX3H1](=O)[#6] prob 0.9997 [#6H3][#6][#6] 0.9862
[OX1]=C[#6] 0.9994 [#6X3][#6][#6][#6H3] 0.7682
[CX4H3][#6] 0.9981 [CX4H3][CX4H1] 0.7426
| best positives                          | prob | best negatives                          | prob |
|----------------------------------------|------|-----------------------------------------|------|
| [CX4H3]                                | 0.9968 | [CX4H2]{[#6]}{[#6]}                     | 0.7244 |
| [#6H1]                                 | 0.9951 | [CX4H3]{CX4H2}                          | 0.6874 |
| [CX3H1]{=#O}{[#6]}                     | 0.9997 | C=CCCC#C                                | 0.0000 |
| [CX3]{=#[OX1]}C                        | 0.9994 | CCC=CCC#C                               | 0.0000 |
| [CX4H3]{[#6]}                         | 0.9981 | [CX2H1]{=[CX2H0]C}{=[CX3H1]}{=[CX3H0]C} | 0.0000 |
| [CX4H3]                                | 0.9968 | [CX3H0]{=[CX3H1]}{=[OX2H0]C}{=[CX2H0]C} | 0.0000 |
| [#6H1]                                 | 0.9951 | C=CCC#C                                 | 0.0000 |
| [#6H3]{[#6]}{[#6]}                     | 0.9862 | [CX2H0]{=[CX2H1]}{=[CX3H0]C}            | 0.0000 |
| [#6X3]{[#6]}{[#6]}{[#6H3]}             | 0.7682 | [CX3H2]{=[CX3H0]C}{=[OX2H0]C}{=[CX4H2]C} | 0.0000 |
| [CX4H2]{[#6]}{[#6]}                    | 0.7244 | [CX2H0]{=[CX2H1]}{=[CX3H0]C}            | 0.0000 |
| [CX4H3]{CX4H2}                         | 0.6874 | C=CCC#C                                 | 0.0000 |
| [CX4H2]{CX4H2}                         | 0.5008 | [CX2H0]{=[CX2H1]}{=[CX3H1]}{=[CX3H0]C} | 0.0000 |

| worst negatives                        | prob | worst positives                         | prob |
|----------------------------------------|------|-----------------------------------------|------|
| [CX4H3]{CX4H1}                         | 0.7426 | [CX4H2]{=[CX4H2]C}{=[CX3H1]}            | 0.0019 |
| O=[CX3]{CX4H}                          | 0.6661 | [OX1H0]{=[CX3H1]}{=[CX4H2]C}{=[CX4H2]C} | 0.0219 |
| [#6H1]{[#6H]}                          | 0.5494 | [CX4H2]{=[CX3H1]}                      | 0.0652 |
| [#8]=[#6H1]{[#6H]}                     | 0.423  | [CX4H2]{=[CX3H1]}{=[OX1H0]C}{=[CX4H2]C} | 0.1516 |
| [#6H3]{[#6]}{[#6]}{[#6H3]}             | 0.3802 | [#6H1]{[#6H]}{[#6H]}                   | 0.174 |
| [CX3H1]{CX4H}                          | 0.3545 | [CX3H1]{=[OX1H0]C}{=[CX4H2]C}           | 0.2456 |
| [CX4]{=[CH3X4]}{=[CH3X4]C}            | 0.3178 | [CX4H2]{=[CX3H1]}{=[OX1H0]C}{=[CX4H2]C} | 0.2582 |
| [#6H3]{[#6]}{[#6X3]}                   | 0.3024 | [CX4H2]{=[CX4H3]C}{=[CX4H2]C}           | 0.4085 |
| [CX3H1]{=[OX1H0]C}{=[CX4H1]}           | 0.1718 | [CX4H2]{=[CX4H2]C}                      | 0.5008 |
| [#8]=[#6]{=[6H1]}{[#6H1]}             | 0.1664 | [CX4H3]{=[CX4H2]C}                      | 0.6874 |

------------------------------------------------------------------------------------------------------------------------
Example true smiles: COCCOC formula: C4H10O2
Index of correct structure: 0 of 20
True structure loss: 0.006334
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

| Structure | prob   |
|-----------|--------|
| [CX4H2]{[#6]}{[O]} | 0.006334 |
| [#8]{[#6]}{[#6H2]}{[#8]} | 0.034641 |
| [CX4H2]{[#6H2]}{[O]}{[CX4H2]} | 0.045662 |
| [OX2H0]{[CX4H2]}{[CX4H2]}{[OX2H0]} | 0.046683 |
| [OX2H0] | 0.060173 |

| Structure | prob   |
|-----------|--------|
| [CX4H2]   | 0.060918 |
| [OX2H0]   | 0.06407  |
| [OX2H0]   | 0.083101 |
| [OX2H0]   | 0.095612 |
| [OX2H0]   | 0.097936 |

Top predicted substructures

[CX4H2]{[#6]}{[O]} | 0.9976 | [#8]{[#6]}{[#6H2]}{[#8]} | 0.8997
[#8]{[#6]}{[#6H2]} | 0.9624 | [OX2H0]{[CX4H2]}{[CX4H2]}{[OX2H0]} | 0.8085
[CX4H2] | 0.9415 | [CX4H3] | 0.8737
|       | best positives |       | best negatives |       |
|-------|----------------|-------|----------------|-------|
|       | prob           |       | prob           |       |
| O[CH4H2] [CX4H2]O | 0.9395 | [CH2X4] (O) [CX4H2] | 0.9041 |
| [CH2X4] (O) [CX4H2] | 0.9041 | [CX4H2] [OX2H0] | 0.8446 |
|       |               |       | [CX4H3] [OX2H0] | 0.8446 |
|       |               |       | [CH2X4] (O) [CX4H2] | 0.9041 |
|       |               |       | [CX4H2] [OX2H0] | 0.8446 |
|       |               |       | [CX4H3] [OX2H0] | 0.8446 |
| worst negatives | prob | worst positives | prob |
| [CX4H2] [OX2H0] [CX4H2] | 0.5821 | O[CH2] [CX4H3] [OX2H0] [CX4H2] | 0.6928 |
| OCC [CH2] | 0.3817 | [#8] [#6] [#6] [#8] | 0.6971 |
| [CX4H2] [O] [CXH4] | 0.3618 | [CX4H2] [OX2H0] | 0.8434 |
| [CX4H2] | 0.2719 | [CX4H3] [OX2H0] | 0.8446 |
| [#7X3] [#6H2] | 0.2673 | [CH2X4] [O] [CX4H2] | 0.9041 |
| [CX4H2] | 0.2498 | [O] [CX4H2] [OX2H0] [CX4H2] | 0.9395 |
| [#6H1] [#6H2] | 0.2334 | [#8] [#6] [#6H2] [#8] | 0.8997 |
| [OXH1] | 0.1687 | [CH2X4] (O) [CX4H2] | 0.9041 |
| [#6H1] | 0.1593 | O[CH4H2] [OX2H0] | 0.9395 |
| O[CH4H2] | 0.0793 | [CX4H2] [OX2H0] [CX4H2] | 0.9415 |
Example 205 true smiles: C(\(\text{O}\))C(\(\text{C}\))O

Index of correct structure: 0 of 20

True structure loss: 0.005398

True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

| Structure | Loss |
|-----------|------|
| \([\text{CX}_4\text{H}_3]\) | 0.005398 |
| \([\text{CX}_4\text{H}_3]\)[\#6] | 0.036003 |
| \([\text{CX}_4\text{H}_3]\)[\text{CX}_4\text{O}] | 0.045253 |
| \([\text{CX}_4\text{H}_3]\)[\#6][\#8] | 0.047717 |
| \([\text{CX}_4\text{H}_3]\)[\#6] | 0.055175 |

Top predicted substructures

| Substructure | prob | Substructure | prob |
|--------------|------|--------------|------|
| \([\text{CX}_4\text{H}_3]\) | 0.9998 | \([\text{OX}_2\text{H}_1]\) | 0.9901 |
| \([\text{CX}_4\text{H}_3]\)[\#6] | 0.9996 | \([\text{CX}_4\text{H}]_0\) | 0.9898 |
| \([\text{CX}_4\text{H}_3]\)[\text{CX}_4\text{O}] | 0.9977 | \([\text{CH}_3][\#6][\#8]\) | 0.9815 |
best positives  prob    best negatives  prob

[CX4H3]  0.9998  [CX2H1]#[CX2H0][CX3H1]=[CX3H0]  0.0
[CX4H3][#6]  0.9996  C=CC=CC#C  0.0
[CX4H3][CX4]O  0.9977  C=CCC=C  0.0
[OX2H1]  0.9962  C=CCCC=C  0.0
[OX2H1]  0.9961  C=CCCC=C  0.0
[OX2H1]  0.9901  C=CCCC=C  0.0
[CX4H3]  0.9898  [CX3H0]==[CX3H1][CX4H2][OX2H0]  0.0
[CX4H3][CX4H1][OX2H1]  0.9723  [CX3H0]==[CX3H1][OX2H0][OX2H0]  0.0
[OX2H1]  0.9666  [OX1H0]==[CX3H0][CX2H0]#[CX2H1]  0.0

worst negatives  prob    worst positives  prob

[CX4H1][OX2H1][CX4H1][CX4H1]  0.5259  [#6H3]#[6]#6[#6H3]  0.5228
[CX4H1](#6H3)[#6H3][#6H3][#6H3]  0.4094  [OX2H1][CX4H1][CX4H1][OX2H1]  0.6797
[CX4H1][CX4H1][OX2H1][CX4H1]  0.247  [#6H1][#6H1]  0.8192
[CH3][CH4][CH3][CH4][CH3][CH4]  0.2416  [CX4H1](#6)[#6H3]  0.8403
[CX4H2](#6)[CX4H1]  0.1218  [#8][#6H1][#6H1][#6H3]  0.8717
[OX2H1][#6H1][#6H2]  0.0845  [#8][#6][#6][#8]  0.8951
[CX4H1][OX2H1][OX2H1][CX4H1]  0.0837  [CH3][CC[OH]]  0.9001
[CH3][#6H2][#6H1][#8H]  0.0573  [OX2H0][OX2H0][CX4H1]  0.9094
[OX2H0][OX2H0][CX4H1][CX4H3]  0.0567  [OH][CX4H4]  0.9459
Example 206 true smiles: CCC(=O)O

**Formula:** C₃H₆O₂

**Index of correct structure:** 0 of 18

**True structure loss:** 0.00581

**True structure:**

![Chemical structure diagram]

**Experimental 13C NMR (solvent: CDCl₃)**

**Experimental 1H NMR (solvent: D₂O)**

**Top predicted structures (loss):**

| Structure | Loss  |
|-----------|-------|
| [CX₄H₃] | 0.06581 |
| [OX₁H₂][CX₃H₃][CX₄H₂][CX₄H₃] | 0.09395 |
| [CX₃][=OX₁][CX₄H₂][CX₄H₃][#6] | 0.10935 |
| [CX₄H₃][#6]=[#6][#8] | 0.110616 |
| [OX₁H₂][#6][#8] | 0.130769 |
| [OX₁H₂][CX₃H₃][CX₄H₂][CX₄H₃] | 0.155919 |
| [OX₁H₂][CX₃H₃][CX₄H₂][CX₄H₃] | 0.162787 |
| [OX₁H₂][CX₃H₃][CX₄H₂][CX₄H₃] | 0.179026 |
| [OX₁H₂][CX₃H₃][CX₄H₂][CX₄H₃] | 0.201812 |
| [OX₁H₂][CX₃H₃][CX₄H₂][CX₄H₃] | 0.203705 |

**Top predicted substructures**

| Substructure | Prob  |
|--------------|-------|
| [CX₄H₃] | 0.9988 |
| [OX₁H₂][CX₃H₃][CX₄H₂][CX₄H₃] | 0.9986 |
| [CX₄H₃][#6] | 0.9985 |
| [OX₁H₂][CX₃H₃][CX₄H₂][CX₄H₃] | 0.9943 |
| [OX₁H₂][CX₃H₃][CX₄H₂][CX₄H₃] | 0.9904 |
| [OX₁H₂][CX₃H₃][CX₄H₂][CX₄H₃] | 0.972 |
best positives                                prob            best negatives                                prob

[CX4H3]                                       0.9988          [CX3H0](=O)=[CX3H1][CX3H2][CX3H3]             0.0
[CX3](=[OX1])C                                0.9986          [CX3H0]==[CX3H1][CX3H2][CX3H3]             0.0
[OX1H0]([OX1H0])O                             0.9904          [CX3H0]==[OX1H0][CX3H1]                  0.0
[CX4H3]                                       0.9985          [CX3H0]==[OX1H0][CX3H1][CX3H2][CX3H3]     0.0
[CX4H3][CX4H2]                                0.9958          [OX1H0]([OX1H0])O                          0.0
[CX3](=[OX1])C                                0.9904          [OX1H0]([OX1H0])O                          0.0
[CX4H3][OX2H1]                                0.9352          [CX3](=[OX1])C                            0.0
[OX2H1]                                       0.9304          [OX1H0]([OX1H0])O                          0.0

worst negatives                               prob            worst positives                               prob

[CX4H2][CX3][OX1]                             0.0249          [OX1H0]==[CX3H0][OX1H0][CX4H2]             0.0
[CX4H2][CX3]                                 0.1907          [CX3H0]==[CX4H2][CX3H1][CX4H2]             0.0
[CX4H2][CX3]                                 0.1886          [CX3H0]==[OX1H0][OX2H1][CX4H2]             0.0
[CX4H2][CX3]                                 0.1804          [OX1H0]==[CX3H0][OX2H1][CX4H2]             0.0
[OX1H0]==[CX3H0][OX1H0][CX4H2]                0.0981          [OX1H0]==[CX3H0][OX2H1][CX4H2]             0.0
[OX1H0]==[CX3H0][OX2H1][CX4H2]                0.0943          [OX1H0]==[CX3H0][OX2H1][CX4H2]             0.0
[OX1H0]==[CX3H0][OX2H1][CX4H2]                0.0935          [OX1H0]==[CX3H0][OX2H1][CX4H2]             0.0
[OX1H0]==[CX3H0][OX2H1][CX4H2]                0.0656          [OX1H0]==[CX3H0][OX2H1][CX4H2]             0.0
[OX1H0]==[CX3H0][OX2H1][CX4H2]                0.0571          [OX1H0]==[CX3H0][OX2H1][CX4H2]             0.0

-----------------------------
Example 207 true smiles: CCCCCCC(C)CN
 formula: C5H13N
 Index of correct structure: 0 of 17
 True structure loss: 0.013327
 True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

Top predicted substructures

[CX4H3] 1.0
[#6H3][#6][#6] 0.9999
[CNH3][#6] 0.9994
best positives        prob                best negatives              prob
[CX4H3][CX4H2]      0.9907            [#7X3][#6H2]                  0.9193
[CX4H2][[#6]][#6]    0.9789            [#6H1][#6H2]                  0.8595

worst negatives       prob                worst positives              prob
[CX4H2][CX4H2][CX4H2] 0.5598          [#6H1][#6H2][#6H2]              0.3591
[CX4H2][[CX4H3]][CX4H2] 0.436           [CX4H1][[CX4H3]][[CX4H2]][CX4H2] 0.389
[#6H1][#6H1]          0.4247          [#7][#6H2][#6H1]                0.4531
[#6H3][#6H1][#6H1][#7] 0.3729          [CX4H2][[CX4H3]][CX4H1]            0.4596
[CX4H2][[CX4H2]][CX4H1] 0.355           [#6H3][#6][#6][#6H3]              0.4993
[#7X3H1]              0.3259          [CX4H2][[NX3H2]][CX4H1]            0.5139
[#7H2][#6H1]          0.2999          [CHX4][[CH3X4]][CH2X4]             0.598
[CX4H2][[NX3H1]][CX4H2] 0.2291          [CX4H3][CX4H1]                  0.79
[#7][#6H2][#6H2]      0.2033          [#7H2][#6H2]                     0.841
[CX4H2][[CX4H2]][CX4H2] 0.2029          [#6H1][#6H2]                     0.8595

------------------------------------------------------------------------------------------------------------------------
Example 208 true smiles: C(C)C(C)N formula: C5H13N
Index of correct structure: 0 of 17
True structure loss: 0.009488
True structure:

Experimental 13C NMR (solvent: CDCl3)

Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

0.009488
0.031909
0.032797
0.038671
0.043152

0.048335
0.055476
0.060535
0.064514
0.066122

Top predicted substructures

| Substructure | prob   | |
|--------------|--------|------------|
| [CX4H3]     | 1.0    | [CX4H3]    |
| [#6H3][#6]  | 0.9999 | [CX4H1]    |
| [CX4H3][#6] | 0.9994 | [CH4]      |
| best positives | prob | best negatives | prob |
|----------------|------|----------------|------|
| [CX4H3]        | 1.0  | [CX4H3]        | 1.0  |
| [#6H1][#6][#6] | 0.9999 | [CX4H3][#6][#6] | 0.9999 |
| [CX4H3][#6]    | 0.9994 | [CX4H3][#6]    | 0.9994 |
| [CX4H3][#6]    | 0.9998 | [CX4H3][#6]    | 0.9998 |
| [CX4H3][CX4H1] | 0.9914 | [CX4H3][CX4H1] | 0.9914 |
| [#7X3H2]       | 0.9732 | [CX4H3][CX4H1] | 0.9732 |
| [CX4H3][#6][#6] | 0.9999 | [CX4H3][#6][#6] | 0.9999 |
| [CX4H3][#6][#6] | 0.9999 | [CX4H3][#6][#6] | 0.9999 |
| [CX4H3][#6][#6] | 0.9999 | [CX4H3][#6][#6] | 0.9999 |

| worst negatives | prob | worst positives | prob |
|-----------------|------|-----------------|------|
| [CHX4][#6][#6][#6] | 0.5107 | [CX4H3][#6][#6][#6] | 0.5107 |
| [#6H1][#6][#6] | 0.5006 | [CX4H3][#6][#6][#6] | 0.5006 |
| [CX4H3][#6][#6][#6] | 0.4811 | [CX4H3][#6][#6][#6] | 0.4811 |
| [CX4H3][#6][#6][#6] | 0.3441 | [CX4H3][#6][#6][#6] | 0.3441 |
| [CX4H3][#6][#6][#6] | 0.3111 | [CX4H3][#6][#6][#6] | 0.3111 |
| [CX4H3][#6][#6][#6] | 0.294 | [CX4H3][#6][#6][#6] | 0.294 |
| [CX4H3][#6][#6][#6] | 0.2593 | [CX4H3][#6][#6][#6] | 0.2593 |
| [CX4H3][#6][#6][#6] | 0.208 | [CX4H3][#6][#6][#6] | 0.208 |
| [CX4H3][#6][#6][#6] | 0.1816 | [CX4H3][#6][#6][#6] | 0.1816 |
| [CX4H3][#6][#6][#6] | 0.1577 | [CX4H3][#6][#6][#6] | 0.1577 |
Example 209 true smiles: C\[N\]^+\}(C)[O-] formula: C3H9NO
Index of correct structure: -1 of 17
True structure loss: 0.013455
True structure:

Experimental 13C NMR (solvent: D2O)

Experimental 1H NMR (solvent: D2O)

Top predicted structures (loss):

| Structure | Loss |
|-----------|------|
| #7X3 #6H2 | 0.03226 |
| #6H2 #6H1 | 0.033502 |
| CX4H3 | 0.033637 |
| #7 #6H2 | 0.034276 |
| #7 #6H2 #6H1 | 0.051929 |

Top predicted substructures

| Substructure | Prob | Substructure | Prob |
|--------------|------|--------------|------|
| #7X3 #6H2 | 0.9249 | #7X3H1 | 0.3659 |
| CX4H3 | 0.8394 | #7X3H2 | 0.3623 |
| #7 #6H2 | 0.6938 | #7 #6H2 #6H1 | 0.2895 |
Example 210 true smiles: CNCC(C)C formula: C5H13N
Index of correct structure: 0 of 17
True structure loss: 0.01223
True structure:
Experimental $^{13}$C NMR (solvent: CDCl$_3$)

Experimental $^1$H NMR (solvent: CDCl$_3$)

Top predicted structures (loss):

| Structure | Probability |
|-----------|-------------|
| [CX4H3]  | 1.0         |
| [#6H3][#6] | 0.9997     |
| [#7X3][#6H3] | 0.9988     |
| [CX4H3][#6] | 0.9955     |
| [#6H3][#7] | 0.9836     |

Best positives

| Structure | Probability |
|-----------|-------------|
| [CH4][(CH3)4] | 0.8869     |
| [CX4H3] | 0.867       |
| [CX4H3][NX3H1] | 0.8502    |
| [#6H1] | 0.7825      |
| [#7X3][#6H2] | 0.6849     |

Best negatives

| Structure | Probability |
|-----------|-------------|
| [CH4][(CH3)4] | 0.8869     |
| [CX4H3] | 0.867       |
| [CX4H3][NX3H1] | 0.8502    |
| [#6H1] | 0.7825      |
| [#7X3][#6H2] | 0.6849     |
Example 211 true smiles: CC(C)(C)CN formula: C5H13N
Index of correct structure: 0 of 17
True structure loss: 0.011019
True structure:
Experimental 13C NMR (solvent: CDCl3)

 Experimental 1H NMR (solvent: CDCl3)

Top predicted structures (loss):

| Structure | Probability |
|-----------|-------------|
| NH₂       | 0.011019    |
| NH₂       | 0.02531     |
| NH₂       | 0.041095    |
| NH₂       | 0.05675     |
| NH₂       | 0.057282    |
| NH₂       | 0.057635    |
| NH₂       | 0.06509     |
| NH₂       | 0.065881    |
| NH₂       | 0.068679    |
| NH₂       | 0.071354    |

Top predicted substructures

| Substructure | Probability |
|--------------|-------------|
| [CX₄H₃]     | 0.9997      |
| [#₆H₃][#₆]  | 0.9991      |
| [CX₄][{CX₄H₃}][{CX₄H₃}][CX₄H₃] | 0.9984 |
| [CX₄H₃][#₆] | 0.9960      |
| [#₇X₃H₂]    | 0.9903      |

Best positives

| Probability |
|-------------|
| 0.9741      |
| 0.9204      |
| 0.6844      |
| 0.5967      |
| 0.589       |
| worst negatives | prob | worst positives | prob |
|-----------------|------|-----------------|------|
| [#6H1]          | 0.5967 | [#7X3H0]        | 0.1188 |
| [#7X3H1]        | 0.4838 | [#7X3H2]        | 0.1987 |
| [#7X3H2]        | 0.3169 | [#7][#6H2]      | 0.5401 |
| [#6H3][#7]      | 0.2847 | [#7X3][#6H2]    | 0.589 |
| [#7][#6H2]      | 0.2358 | [#6H3][#6H0]    | 0.9204 |
| [#6H1][#6H2]    | 0.1789 | [CX4H3][CX4H0]  | 0.9741 |
| [#7X3][#6H3]    | 0.1607 | [CX4H3][#6]     | 0.9903 |
| [#6H2][#6H1]    | 0.1085 | [CX4H3][CX4H0][CX4H3] | 0.996 |
| [#6H2][#7][#6H2] | 0.0996 | [CX4][{CX4H3}][{CX4H3}] | 0.9984 |

Example 212 true smiles: CC(C)CCH formula: C5H13N
Index of correct structure: 0 of 17
True structure loss: 0.014666
True structure:
Experimental 1H NMR (solvent: CDCl3)

Experimental 13C NMR (solvent: CDCl3)

Top predicted structures (loss):

- [C6H3][#6][#6] 0.9996
- [C6H3] 0.9978
- [C7H3][#6H2] 0.9903
- [C4H3][#6] 0.9893
- [C7H2][#6H2] 0.9732

Top predicted substructures

| [C6H3][#6][#6] | 0.9699 |
| [C6H3] | 0.9186 |
| [C7H3][#6H2] | 0.8439 |
| [C4H3][#6] | 0.8272 |
| [C7H2][#6H2] | 0.7821 |

best positives prob

| [C6H3][#6][#6] | 0.9996 |
| [C6H3] | 0.9978 |
| [C7H3][#6H2] | 0.9903 |
| [C4H3][#6] | 0.9893 |
| [C7H2][#6H2] | 0.9732 |

best negatives prob
```
| best positives                               | prob  |
|------------------------------------------------|-------|
| CCCCC#C                                      | 0.9996|
| [CX4H3]                                      | 0.9978|
| [CX4H3]                                     | 0.9903|
| [CX4H3]                                     | 0.9893|
| [CX4H3]                                     | 0.9732|
| [CX4H3]                                     | 0.9699|
| [CX4H3]                                     | 0.9186|
| [CX4H2]                                     | 0.8439|
| [CX4H2]                                     | 0.8272|

| worst negatives                               | prob  |
|------------------------------------------------|-------|
| CCCCC#C                                      | 0.0004|
| [CX4H3]                                      | 0.0003|
| [CX4H3]                                     | 0.0004|
| [CX4H3]                                     | 0.0003|
| [CX4H3]                                     | 0.0003|
| [CX4H3]                                     | 0.0003|
| [CX4H3]                                     | 0.0003|
| [CX4H3]                                     | 0.0003|
| [CX4H3]                                     | 0.0003|
| [CX4H3]                                     | 0.0003|
| [CX4H3]                                     | 0.0003|

Example 213 true smiles: CCCCNC formula: C5H13N
```

**Index of correct structure:** 0 of 17

**True structure loss:** 0.016642

**True structure:**
Experimental $^{13}$C NMR (solvent: CDCl$_3$)
worst negatives                          prob            worst positives                          prob
[#6H1]                                   0.8169          [#7][#6H2][#6H2]                               0.1223
[#6H1][#6H2]                             0.6088          [CX4H2]([NX3H1])[CX4H2]                         0.1841
[CX4H3][CX4H1]                           0.5615          [CX4H2][CX4H2][CX4H2]                           0.353
[CX4H4][CX3X4][CX2X4]                   0.5281          [CX4H2][CX4H2]                                 0.5532
[CX4H2][[NX3H1]][CX4H1]                 0.4053          [#6H3][#7][#6H2]                               0.6932
[CX4H2][[CX4H3]][CX4H1]                 0.36            [#7X3][#6H2]                                  0.7106
[#6H3][#6H1][#6H1][#7]                  0.3215          [CX4H2][CX4H3][CX4H2]                           0.7297
[#7][#6H2][#6H1]                        0.2725          [#7X3H1]                                      0.7637
CCCCCC                                   0.1812          [#7][#6H2]                                    0.796
[#7X3H2]                                 0.1655          [CX4H3][NX3H1]                                 0.8863

--------------------------------------------------------------------------------------------------