



Figure 1. Two almost identical molecules with markedly different canonical SMILES in RDKit. The edit distance between two strings is 22 (50.5% of the whole sequence).

Minimal DFS codes:

```
[ [ 0, 1, 6, 0, 6], [ 1, 2, 6, 0, 7], [ 2, 3, 7, 0, 6],
[ 3, 4, 6, 0, 6], [ 4, 5, 6, 2, 6], [ 5, 6, 6, 0, 6],
[ 5, 7, 6, 2, 6], [ 7, 8, 6, 2, 6], [ 8, 9, 6, 0, 9],
[ 8, 10, 6, 2, 6], [10, 11, 6, 2, 6], [11, 4, 6, 2, 6],
[ 3, 12, 6, 1, 8], [ 2, 13, 7, 0, 6], [ 0, 14, 6, 2, 6],
[14, 15, 6, 0, 6], [14, 16, 6, 2, 7], [16, 17, 7, 2, 6],
[17, 18, 6, 2, 7], [18, 0, 7, 2, 6], [18, 19, 7, 2, 6],
[19, 20, 6, 2, 6], [20, 21, 6, 0, 6], [20, 22, 6, 2, 16],
[22, 17, 16, 2, 6]]
```

and

```
[ [ 0, 1, 6, 0, 6], [ 1, 2, 6, 0, 7], [ 2, 3, 7, 0, 6],
[ 3, 4, 6, 0, 6], [ 4, 5, 6, 2, 6], [ 5, 6, 6, 0, 6],
[ 5, 7, 6, 2, 6], [ 7, 8, 6, 2, 6], [ 8, 9, 6, 0, 9],
[ 8, 10, 6, 2, 6], [10, 11, 6, 2, 6], [11, 4, 6, 2, 6],
[ 3, 12, 6, 1, 8], [ 2, 13, 7, 0, 6], [ 0, 14, 6, 2, 6],
[14, 15, 6, 0, 6], [14, 16, 6, 2, 7], [16, 17, 7, 2, 6],
[17, 18, 6, 2, 7], [18, 0, 7, 2, 6], [18, 19, 7, 2, 6],
[19, 20, 6, 0, 6], [19, 21, 6, 2, 6], [21, 22, 6, 2, 16],
[22, 17, 16, 2, 6]]
```

Example from <https://arxiv.org/pdf/1802.04364.pdf>

Junction Tree Variational Autoencoder for Molecular Graph Generation