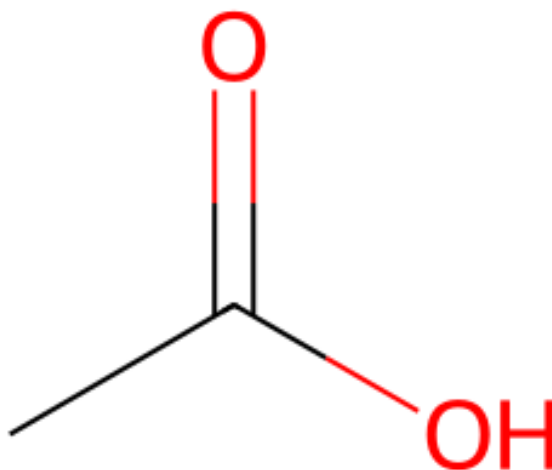


Appendice

Interpretazione della rappresentazione SELFIES

La molecola dell'acido acetico ha la seguente rappresentazione in notazione SMILES:

CC(=O)O



Acido Acetico

La conversione in SELFIES avviene applicando regole grammaticali che rispettano la valenza degli atomi e strutturano la molecola come sequenza lineare e ramificata. Il risultato è:

[C][C][=Branch1][C][=O][O]

La rappresentazione completa della derivazione avviene come da tabella:

Token SELFIES	Descrizione	Val. Iniz.	Val. Residua	Note
[C]	C metile (inizio catena)	4	3	Lineare
[C]	C centrale	4	3	Legato a C precedente
[=Branch1]	Apri ramo con legame doppio	–	–	Ramo da 1 token
[C]	C del gruppo carbonilico	4	2	Saturato da =O
[=O]	Ossigeno con doppio legame	2	0	Termina il ramo
[O]	Ossidrilico legato a C centrale	2	1	Prosegue catena lineare

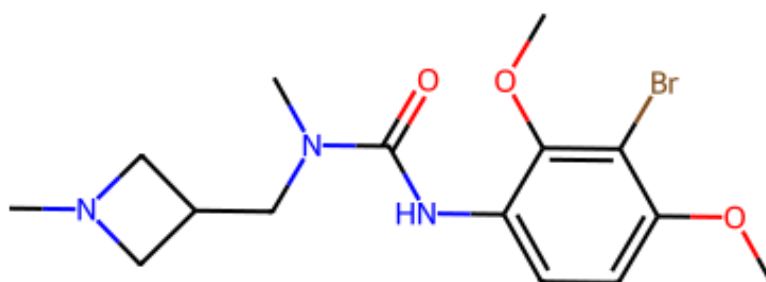
Table 1: Derivazione SELFIES per CC(=O)O: valenze e struttura molecolare.

La derivazione avviene tramite un sistema di produzione basato su stati X_n , che controllano la saturazione della valenza.

Il simbolo del carbonio che appare lungo la ramificazione principale è riprodotto nella ramificazione iniziata con il token [=Branch1]

Consideriamo la successiva molecola maggiormente strutturata facente parte del database di studio, rappresentata secondo SMILES con la sequenza.

COC1=CC=C(NC(=O)N(C)CC2CN(C)C2)C(OC)=C1Br



Questa è la rappresentazione in SELFIES

```
[C][O][C][=C][C][=C][Branch2][Ring1][Ring2][N][C]
[=Branch1][C][=O][N][Branch1][C][C][C][C][C][N]
[Branch1][C][C][C][Ring1][Branch1][C][Branch1]
{CH, {OCH
```

La parte iniziale [C][O][C][=C][C][=C] rappresenta, partendo dal gruppo metossilico [C][O][C] percorre l'anello aromatico fino a interpretare un successivo punto di diramazione e i due token che rappresentano 'l'apertura doppia di ciclo [Branch2][Ring1][Ring2]. Il ramo aperto con [Branch2] viene analizzato con [N][C] cui segue una sottodiramazione [=Branch1][C][=O] e l'Azoto che prosegue la catena.

[Branch1][C][C][C][C][N] percorre la catena fino all'atomo di azoto terminale

[Branch1][C][C][C] prepara la chiusura dell'anello piperidina-like

In successione il token [Ring1] chiude l'anello tra l'ultimo carbonio aggiunto nel passo in precedenza e uno dei precedenti carboni della catena laterale. [Branch1][C][Branch1] aprono piccole ramificazioni sostituenti (-CH, -OCH) La sequenza [Ring1][O][C][=C][Ring2][Ring1][Ring1][Br] ritorna all'anello iniziale, rappresenta l'ulteriore gruppo metossilico collegato e un ulteriore carbonio aromatico. I vari token [Ring2] etc sono relativi a come il parser di selfies ricostruisce la struttura a livello sintattico rispettando i valori di valenza ancora non soddisfatti per chiudere i cicli (sintatticamente e semanticamente). Chiude la rappresentazione il Bromo, sostituito presente nella catena aromatica