

SMILES 표기법

규칙 6가지

1. 각 원자는 표준 원소기호로 나타낸다.
2. 수소 원자는 표기를 생략한다.
3. 이웃에 위치한 원자는 서로 결합되어 있다.
4. 이중결합은 "="으로, 삼중결합은 "#"으로 표기한다.
5. 결합 가지는 괄호로 표기한다.
6. 고리 구조는 서로 연결되어 있는 원자에 숫자로 표시한다.

1. 각 원자는 표준 원소기호로 나타낸다.
2. 수소 원자는 표기를 생략한다.

SMILES 표기법에서 각 원자는 해당하는 원소 기호로 표시하고, 수소 원자 (H)는 생략한다. 예를 들어, 탄소는 C, 질소는 N, 산소는 O, 염소는 Cl 등으로 표기하는 것이다.

3. 이웃에 위치한 원자는 서로 결합되어 있다.
4. 이중결합은 "="으로, 삼중결합은 "#"으로 표기한다.

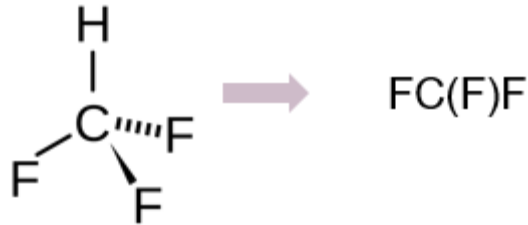
각 결합은 . - = # : / ≡ 의 8가지 기호로 표현한다. 이 때, 단일 결합은 -로 표현되고 주로 생략된다. 결합의 예시로는, 이중 결합은 =, 삼중 결합은 #, ...으로 표현한다. 1.5중은 :.이고, 결합하지 않음은 .이다. (공유 결합이 없고 이온 결합 같은 경우.)

$\text{CH}_3\text{-N=C=O}$ 수소 (H)를 모두 제거하고, 단일 결합을 생략하면? $\Rightarrow \text{CN=C=O}$

methyl isocyanate 분자

5. 결합 가지는 괄호로 표기한다.

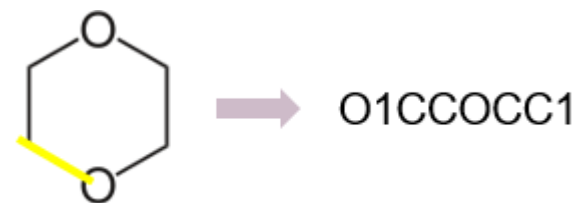
분자의 가지는 괄호()로 표현한다. 이 때, 괄호 안에 포함된 첫 번째 원자와 괄호가 끝나고 나오는 첫 번째 원자가 같은 원자에 연결되어 있다.



6. 고리 구조는 서로 연결되어 있는 원자에 숫자로 표시한다.

SMILES 표기법에서 고리는 어떤 랜덤한 한 지점의 결합을 끊고, 해당 끊긴 부분의 원자 두 개에 번호를 표시하는 방식으로 표기한다.

예로 dioxane 분자의 한 부분의 결합을 아래와 같이 끊으면, O 분자로부터 반시계방향으로 돌아가면서 SMILES 표기법으로 분자를 표현할 수 있다. 이 때, 해당 고리의 번호는 1이다.



6. 고리 구조는 서로 연결되어 있는 원자에 숫자로 표시한다.

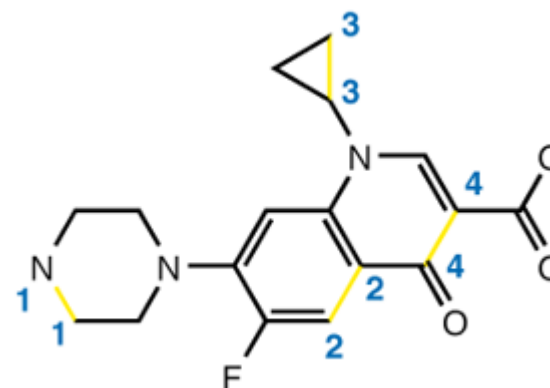
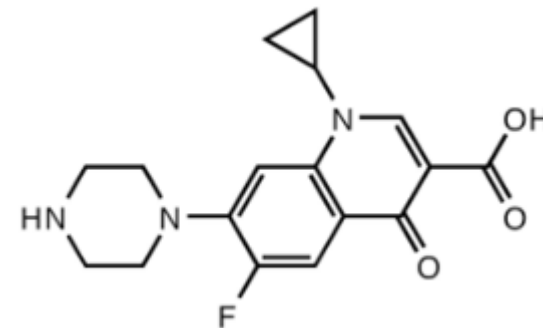
방향족은 탄소 화합물이 평면의 고리 형태로 결합하여 안정적인 구조를 가지는 aromatic ring을 포함하고 있는 것을 말한다. 이 aromatic ring은 앞 슬라이드에 설명한 고리 (ring)을 표현하는 것과 동일하게 표시하지만, 여기에 포함된 B, C, N, O, P, S 원자를 소문자로 표시한다.



pyridine 분자의 SMILES 표기법

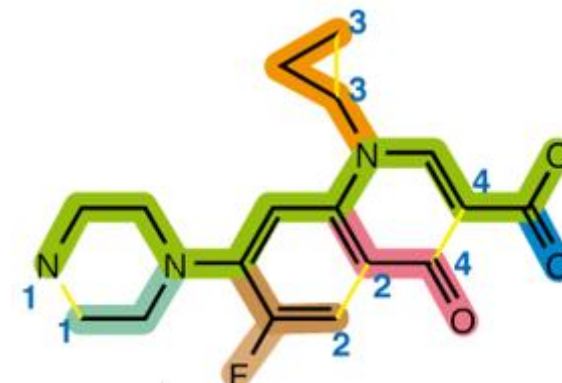
Example

분자들로부터 수소 원자들을 제거함.
분자의 고리들에 있는 결합을 하나씩 제거 및 고리들마다 번호를 매김.



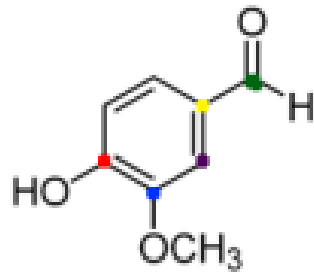
특정 시작 원자를 정하고, 해당 시작 원자로부터 depth-first search (DFS)를 통해 SMILES string을 만들어낸다. (고리의 결합을 2에서 끊었기 때문에 tree 형태로 표현할 수 있고, DFS가 가능하다.) 이 때, branch를 만나면 SMILES string에 역시 추가하고, 더 이상 진행할 수 없는 dead-end를 만나면 끝.

N1CCN(CC1)C(C(F)=C2)=CC(=C2C4=O)N(C3CC3)C=C4C(=O)O



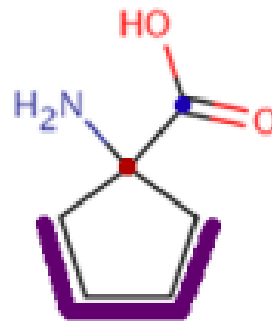
Example

바닐린 (Vanillin)



O=Cc1ccc(O)c(OC)c1

시클로루신



NC1(CCCC1)C(=O)O

어려울 거 같은데...

실제로 사용할 때는 MarvinSketch 같은 Cheminformatics 툴킷에서 분자를 그린 다음에 SMILES로 저장해서 사용할 수 있다고 함.

MarvinSketch에선 분자를 그린 후에 마우스를 드래그해서 분자를 선택한 후, Edit -> Copy as Smiles를 선택하거나, 우클릭한 후 Copy As를 누르고 여기서 ChemAxon SMILES나 Daylight SMILES를 선택할 수 있다.

어려울 거 같은데...

웹에서 사용할 수 있는 프로그램들도 있다고 함.

<https://chemdrawdirect.perkinelmer.cloud/js/sample/index.html>

분자를 그린 후에 Structure에서 Get SMILES 를 선택하면 됨.

rdkit?

```
from rdkit import Chem
```

```
from rdkit.Chem import Draw
```

```
smiles = 'CCc1ccc2[nH]ccc2c1' #example
```

```
mol = Chem.MolFromSmiles(smiles)
```

유의점

SMILES화하는 알고리즘이 죄다 달라서 똑같은 분자여도 SMILES가 다를 수 있다.