Module 5: Identifying Chemical Entities

By Sunghwan Kim

Questions

1. Go to the PubChem database (http://pubchem.ncbi.nlm.nih.gov) and search for omeprazole and esomeprazole. Fill in the table below with appropriate chemical representations for the two molecules and answer the following questions.

(1) What is the structural difference between omeprazole and esomeprazole?
(2) Do omeprazole and esomeprazole have the same InChI and InChIKeys as each other?
(3) Do omeprazole and esomeprazole have the same canonical SMILES? Explain why.

Omeprazol (from PubChem)	
CID	
IUPAC name	
Canonical SMILES	
Isomeric SMILES	
InChI	
InChIKey	
Esomeprazole (from PubChem)	
CID	
IUPAC name	
Canonical SMILES	
Isomeric SMILES	
InChI	
InChIKey	

2. Go to ChemSpider (http://www.chemspider.com) and search for omeprazole and esomeprazole. Fill in the table below with appropriate chemical representations for the two molecules and answer the following questions.

(1) Are the systematic names from ChemSpider the same as those from PubChem?

(2) Are the canonical SMILES from ChemSpider the same as those from PubChem?

(3) Are the InChI and InChIKeys from ChemSpider the same as those from PubChem?

Omeprazol (from ChemSpider)	
ChemSpider ID	
IUPAC name	
Canonical SMILES	
Isomeric SMILES	
InChI	
InChIKey	
Esomeprazole (from ChemSpider)	
ChemSpider ID	
IUPAC name	
Canonical SMILES	
Isomeric SMILES	
InChI	
InChIKey	

3. Compare the SMILES strings from PubChem with those from ChemSpider for the following compounds, in terms of how the two databases deal with perceived aromaticity of the molecules. Explain an advantage and a disadvantage of the SMILES strings used in each database.

	SMILES from PubChem	SMILES from ChemSpider
Benzene		
pyridine		
Pyrrole		
Furan		
Thiophene		
Selenophene		
Tellurophene		

4. Suppose that you are a project manager at Google, who are in charge of implementing a chemical search algorithm to the Google search. This algorithm accepts a chemical structure as an input through the search box on the Google homepage (http://www.google.com), but the input needs to be a text string that represents a chemical structure. Therefore, you need to choose a line notation that is most appropriate for this search system, among the canonical SMILES, InChI, and InChIKey. Choose only one and justify your choice over the others, based on what you have learned from this module and from Questions 1, 2 and 3).