

Generating 3D Printable Models of Organic Compounds Using SMILES™

C Giménez Esteban

*Col·legi Sant Gabriel de Viladecans, Spain
carlos.gimenez@gmail.com*

Abstract. Although the three-dimensional structure is a key determining factor for most of organic compounds properties, nothing but flat representations of those compounds on a blackboard are available for high school students.

Today, technology enables an epic change in that situation. We can now easily generate 3D models of any organic compound to be seen, manipulated on a screen and even printed with a 3D-printer.

First, we must be able to properly communicate with the computer in order to “tell” the machine which is the molecule that we want to model. This communication is easily achieved using a de facto standard language for chemical structures named SMILES™ (Simplified Molecular Input Line Entry System).

SMILES™ is a simplified chemical language which provides easy communication between people and computers.

Once we have coded the molecule using SMILES™ language, we can use dozens of online services for extended functionality, such as getting chemical information, rendering 3D dynamic model on any screen and obtaining a file written in one of the 3D-printing standard formats.

In this talk, we are going to fully reproduce the sequence of actions required to get an on-screen 3D dynamic model and a 3D printable model of any given organic compound.

With this idea on mind, you can easily scale the sequence to your own needs; thus you can teach students to write SMILES™ code, or you can use easily available SMILES™ code of many organic compounds; you may 3D print your models, you can also create collections of compounds models, and so on.

This sequence has been successfully tested with high school students at Sant Gabriel

School in Viladecans, near Barcelona, Spain.

Keywords. 3D model, 3D printing, classroom experience, organic compounds, scalability, SMILES.

1. Why 3D modelling?

First of all, we must consider that modelling the reality forces us to do some simplifications, so the question is: how close to the original must be a model to become a good representation of this reality?

Of course, it's impossible to generate a “perfect” model. The only one model that fits exactly the reality is the reality itself.

Then, when we model an organic compound we must try to build a model as similar to the reality as possible. Molecules are 3D objects, so flat representations do not provide good models of them. However, there is a reason for not using 3D models of organic compounds in our classrooms: the technical difficulty related with building these models.

Although there are certain properties of organic compounds that depend on their 3D structure, the complexity required to build such models has prevented Chemistry and Biology teachers from using them in the classroom.

In this talk we are going to provide the technological tools that will allow everyone (students and teachers) to easily generate 3D models of any organic compound, to view and manipulate on the screen and also to build printable ones.

In the following section we'll review some examples supporting the benefits of working with 3D models in addition to classical flat representations of some interesting compounds.

1.1. Benzene vs. cyclohexane

By only using flat representations of these two compounds (Figure 1), probably our students would be tricked into thinking that these two compounds are really similar one to another.

In most cases, after viewing the following figures they are going to describe both molecules simply as “hexagons”.

Then, if we want to go deeper, we need to use 3D models (Figure 2) to properly highlight the structural difference between benzene (a real flat molecule) and cyclohexane (a non-flat molecule).

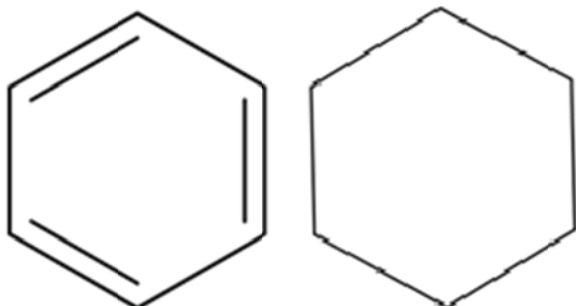


Figure 1. Flat models of benzene (left) and cyclohexane (right) generated from SMILES™ code at molview.org

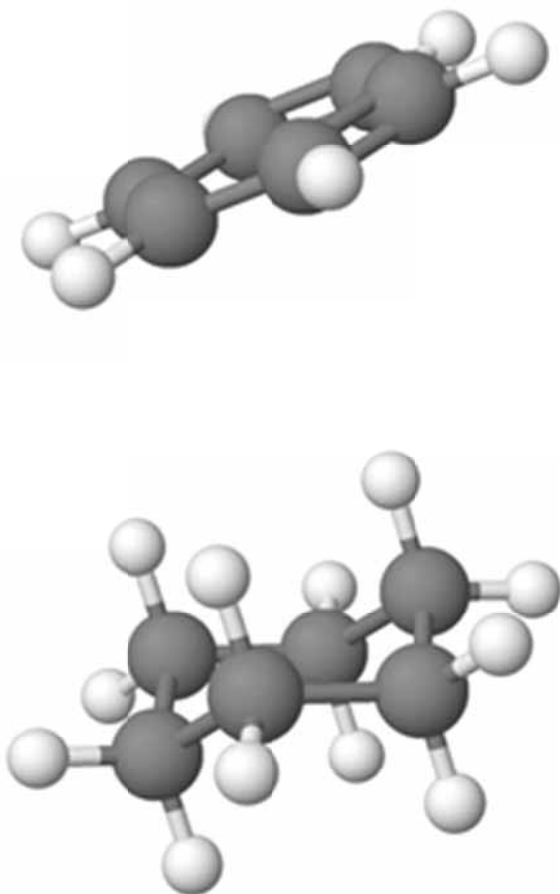


Figure 2. 3D models of benzene (top) and cyclohexane (bottom) generated from SMILES™ code at molview.org

1.2. *Trans* vs. *Cis* fatty acids

While monounsaturated fatty acids in their *cis*- configuration are widely considered as

healthy diet components (*cis*-oleic acid is the canonical example), many evidences prove the relationship between *trans*- fatty acids and a large set of cardiovascular diseases.

It seems much easier to understand such different behaviour by using 3D models of both kind of fatty acids (Figure 4) instead of simply using flat models (Figure 3).

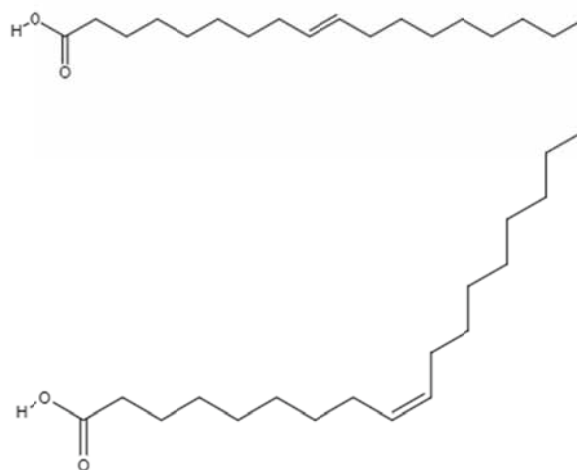


Figure 3. Flat models of *trans*-elaidic acid (top) and *cis*-oleic acid (bottom) generated from SMILES™ code at molview.org

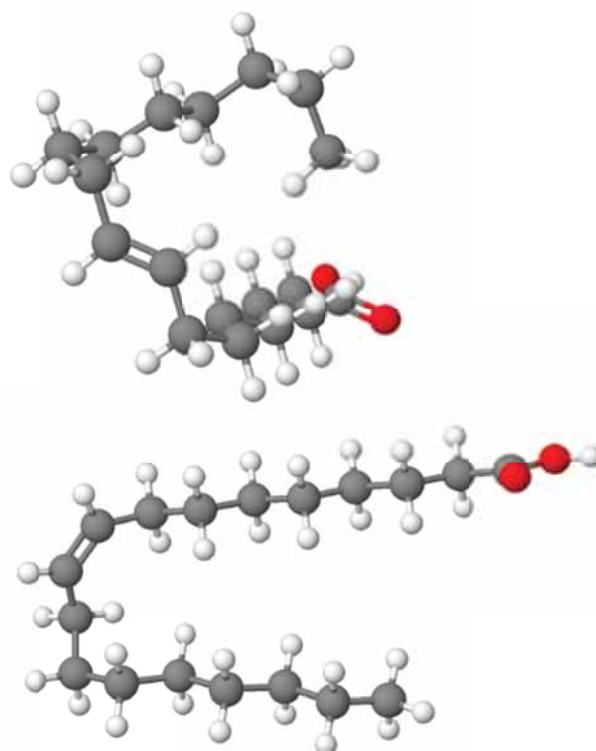


Figure 4. 3D models of *trans*-elaidic acid (top) and *cis*-oleic acid (bottom) generated from SMILES™ code at molview.org

2. How difficult is to code an organic molecule with SMILES™ language?

Easy, in fact coding an organic compound with SMILES™ is very easy. Let's see the main rules of SMILES™ coding [1]:

- 1) Atoms are represented by their atomic symbols.
- 2) We do not need to write hydrogens if the number of them attached to any atom conforms to the lowest normal valence consistent to explicit bonds.
- 3) Lower case letters specify atoms in aromatic rings.
- 4) Single, double, triple, and aromatic bonds are represented by the symbols -, =, #, and :, respectively. Adjacent atoms are assumed to be connected to each other by a single or aromatic bond (single and aromatic bonds may always be omitted).
- 5) Branches are specified by enclosing them in parentheses.
- 6) Cyclic structures are represented by breaking one bond in each ring. The bonds are numbered in no particular order, designating ring opening (or ring closure) bonds by a digit immediately following the atomic symbol at each ring closure.

With this simple set of rules, you can code most of the organic compounds; almost all the examples in this paper (Table 1). For more complex use cases, you can check the complete tutorial at the website of SMILES™ language developers.

Table 1. SMILES™ codes of the example compounds used in this paper

Compound	SMILES™ Code
benzene	c1ccccc1
cyclohexane	C1CCCCC1
transelaidic acid	CCCCCCC/C=C/CCCCCCCC(=O)O
cisoleic acid	CCCCCCC/C=C\CCCCCCCC(=O)O

3. Getting an 3D on screen model

Once you have coded your molecule using SMILES™, you need a computing engine capable of parsing this code and generate your

3D model. As of today, there are dozens of free online services that can do this work for you. All the 2D and 3D models that you can see in this paper have been generated using molview.org [2].

Then, you can create, for example, collections of models [3], and access to specific information of any compound using the SMILES™ code for querying databases, and so on.

4. Next step: getting a 3D printable model

Today we are witnessing the raising of a new technology: 3D printers are reducing dramatically their prize and overcoming their initial limitations, such as the long time needed to produce a single piece. It seems easy to predict that in a few years most of high schools will have a 3D printer. In this context, Chemistry and Biology teachers will welcome ideas taking advantage of this new resource. Physically putting the models into the hands of our students increases their understanding of the structures they can manipulate.

The sequence of actions you must do to get your model on your hands is very easy [4]:

- 1) Choose an organic compound
- 2) Code your compound using SMILES™
- 3) Transform your SMILES™ code into a specific chemistry language (.mol) file using an online service [5].
- 4) Get a 3D printable file (.stl) using another online service [6]
- 5) Print your model with your 3D printer.

5. Acknowledgements

I want to acknowledge first Dr. Ángel Herráez, at the Universidad de Alcalá de Henares, who first introduced me into the world of SMILES™ language.

Second, my colleague Angel Lucas, at Col·legi Sant Gabriel de Viladecans who shared with me our first steps using SMILES™.

Finally, Dr. Josep Maria Fernández Novell, at the Universitat de Barcelona, who encouraged me to develop these materials in the context of the "Crazy about biochemistry" [4], an excellence program addressed to high school students and funded by the Catalunya la Pedrera Foundation.

6. References

- [1] Daylight: SMILES - A Simplified Chemical Language,
<http://www.daylight.com/dayhtml/doc/theory/theory.smiles.html>
- [2] Bergwerf, Herman: Molview; 2015,
<http://molview.org/>
- [3] Giménez C. Banc de molecules 2014,
<http://carlosgimenez.info/molecules/>
- [4] Giménez C. Taller de Bioinformàtica 2018,
<http://carlosgimenez.info/bbq2018/3D.html>
- [5] US National Cancer Institute. File Generator 2017,
<https://cactus.nci.nih.gov/translate/>
- [6] US National Institutes of Health. 3D Print Exchange 2016,
<https://3dprint.nih.gov/>