

chemobabel

— Chemical Structures from MDL Molfiles, ChemDraw Files or SMILES Notations —

Hironobu Yamashita (@aminophen)

September 19, 2022

This document (chemobabel-en.pdf) describes the usage of chemobabel.sty, a L^AT_EX package for **generating chemical structural formulas** to be inserted in your documents. The formulas can be generated **from many kinds of chemical data formats**, including MDL Molfiles, ChemDraw files and even from SMILES notations, with the help of Open Babel and Inkscape. The formulas below are generated using this method.

日本語版ドキュメントは chemobabel-ja.pdf を参照してください。

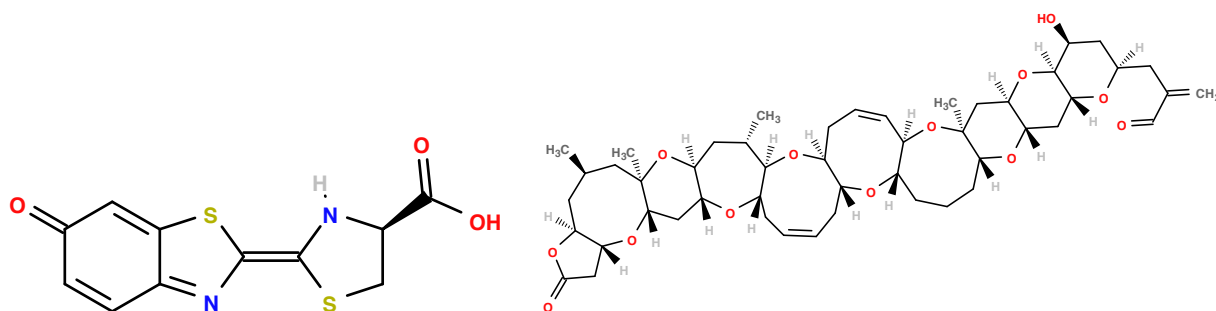


Fig. 1: Firefly luciferin & Brevetoxin A (from MDL Molfiles)

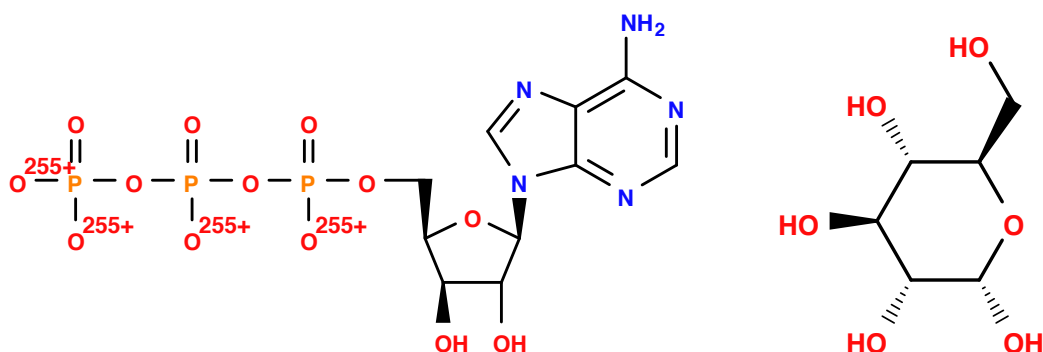


Fig. 2: ATP (Adenosine triphosphate) & Glucose (from ChemDraw files)

Note: with Open Babel 3.1.1 the chemical structure of ATP from the .cdx file becomes wrong.

See <https://github.com/openbabel/openbabel/issues/338>

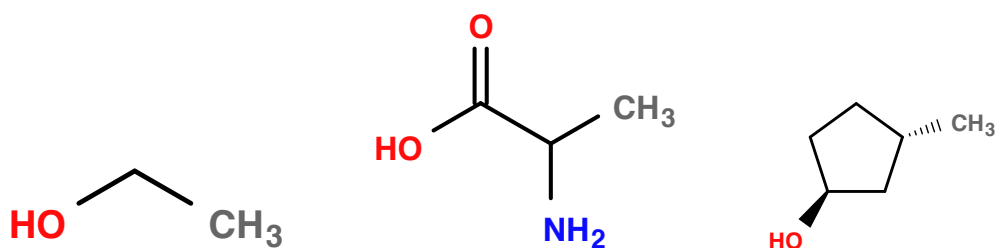


Fig. 3: Ethanol, Alanine & (1S,3S)-3-Methylcyclopentanol (from SMILES notations)

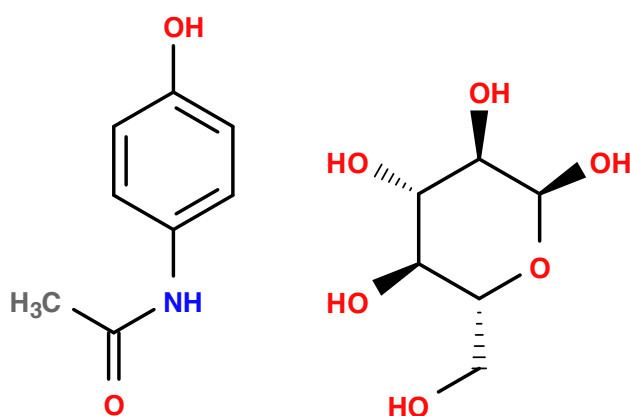


Fig. 4: Acetaminophen (Paracetamol) & α -D-Glucose (from SMILES notations)

This package is maintained on GitHub:
<https://github.com/aminophen/chemobabel>
 If you have questions, please let me know.

Contents

1	Introduction	1
1.1	Motivation for Development	1
1.2	Idea and Approach	2
2	Before you begin	3
2.1	Installing dependent softwares	3
2.2	Loading the package	3
3	Basic Usage	4
3.1	Conversion from Chemical Structure Data Files	4
3.2	Conversion from SMILES Notations	4
3.3	Package options	7
3.4	For Safer Typesetting...	8
4	Advanced Usage	9
4.1	Options for Depiction	9
4.2	More complex structural formulas	11
5	Note for Compatibility	15
5.1	Advantages of chemobabel	15
6	Limitations and Alternatives	16
7	Technical information	17
8	Version History	18

1 Introduction

1.1 Motivation for Development

As you already know, \LaTeX is being used all over the world. However, when it comes to drawing chemical structural formulas, ways of inserting formulas into \LaTeX documents are very limited. I think it is mainly due to lack of both reliable and simple methods of doing this that many of the researchers in chemistry are reluctant to use \LaTeX system.

Some \TeX / \LaTeX macros and packages are already available from CTAN:

- $\text{\X}\text{\LaTeX}$: a set of packages for drawing chemical structural formulas
- chemfig: a package which draws molecules using TikZ

These packages are reliable and almost all kinds of structural formulas can be drawn in a consistent way. However, a lot of practice will be required before one can come to make full use of them.

You can also insert structural formulas with `\includegraphics` of PDF or EPS files exported from special softwares for chemists, such as ChemDraw and other similar programs. In this case, you have to save files in both chemical formats (such as `.cdx` or `.mol`) and graphical formats (`.pdf` or `.eps`) manually.

This new package, `chemobabel.sty`, will offer a new choice for chemists who need chemical structures inserted in their \LaTeX documents. In this method, we use Open Babel for generating chemical structural formulas in SVG format, and Inkscape for converting SVG to PDF. Both programs are open source and cross-platform, and being actively maintained.

■About Open Babel Open Babel is a toolbox specially designed to handle many kinds of chemical data. It is an open, collaborative project allowing anyone to search, convert, analyze, or store data from molecular modeling, chemistry, solid-state materials, biochemistry, or related areas. See Official Website for detail.

1.2 Idea and Approach

First I found a post on Noel O’Blog [1], [2] and the comment [3] about generating chemical structural formulas from SMILES notations using Open Babel. I thought the method described in these posts were very interesting, but I didn’t like all the programs could be executed by `-shell-escape` option. At the same time, I encountered a macro for copying all `\includegraphics` from one source file to another [4], which made me conceive of an idea of extracting minimal source code which needs `-shell-escape` option.

I thought this method would be applicable to converting ChemDraw files into graphics, but I could not find any project attempting to realize this idea. I found a project named Chemdraw in \LaTeX on Sourceforge, but now it was closed. This drove me to develop a new package to simplify the process for inserting chemical structural formulas into \LaTeX documents.

2 Before you begin

2.1 Installing dependent softwares

First, you have to install Open Babel and Inkscape (or librsvg) on your computer, and export PATH to the command-line binary of both programs, obabel and inkscape (or rsvg-convert). You can confirm by executing following commands: In UNIX operating system:

```
$ which obabel
$ which inkscape      (if you choose inkscape)
$ which rsvg-convert  (if you choose librsvg)
```

In Windows operating system:

```
> where obabel
> where inkscape      (if you choose inkscape)
> where rsvg-convert  (if you choose librsvg)
```

And when you get some command echo, installation should be successful.

2.2 Loading the package

In the preamble of your document, declare

```
\usepackage{chemobabel}
```

to load chemobabel package. The package depends on graphicx.

3 Basic Usage

3.1 Conversion from Chemical Structure Data Files

Suppose you have `ATP.cdx`, a file with chemical structural formula of ATP (Adenosine triphosphate) drawn in ChemDraw, and `Brevetoxin A.mol`, a MDL Molfile which can be obtained from ChemSpider (originally `9041092.mol`).

Use `\chemobabel` command as follows. I put original files in a subdirectory `draw`, so added `draw/` before the actual filename. Of course you can put in the same directory with your \LaTeX file.

```
\chemobabel[width=90mm]{draw/ATP.cdx}{}  
\chemobabel[width=120mm]{draw/Brevetoxin A.mol}{}  

```

Then, typeset with `-shell-escape` option. Remember to use this option, otherwise it won't allow \LaTeX to call any external command-line programs, such as `obabel` and `inkscape`!

```
$ pdflatex -shell-escape test.tex
```

You will see in `chemobabelimgdir` subdirectory "`chemobabelimg[NUM].pdf`", and output file `test.pdf` will be Figure 5:

The syntax is:

```
\chemobabel[options]{filename}{obabel options}
```

As parameters, *filename* should be placed in the first braces, and *obabel options* (see Section 4.1 for some examples) can be added in the second braces. You can leave the second parameter empty.

You can give *options* to be passed to `\includegraphics` command as optional parameters in the box brackets. This option is not necessary, so can be omitted. In this case the default option "`scale=1`" is passed to `\includegraphics` command.

3.2 Conversion from SMILES Notations

You can also write SMILES notations directly in your \LaTeX file to be converted into structural formulas. For example, `CCO` is the SMILES notation of ethanol, and `CC(=O)Nc1ccc(cc1)O` is that of acetaminophen (paracetamol). Use `\smilesobabel` command as follows:

```
\smilesobabel{CCO}{}  
\smilesobabel[width=35mm]{CC(=O)Nc1ccc(cc1)O}{}  

```

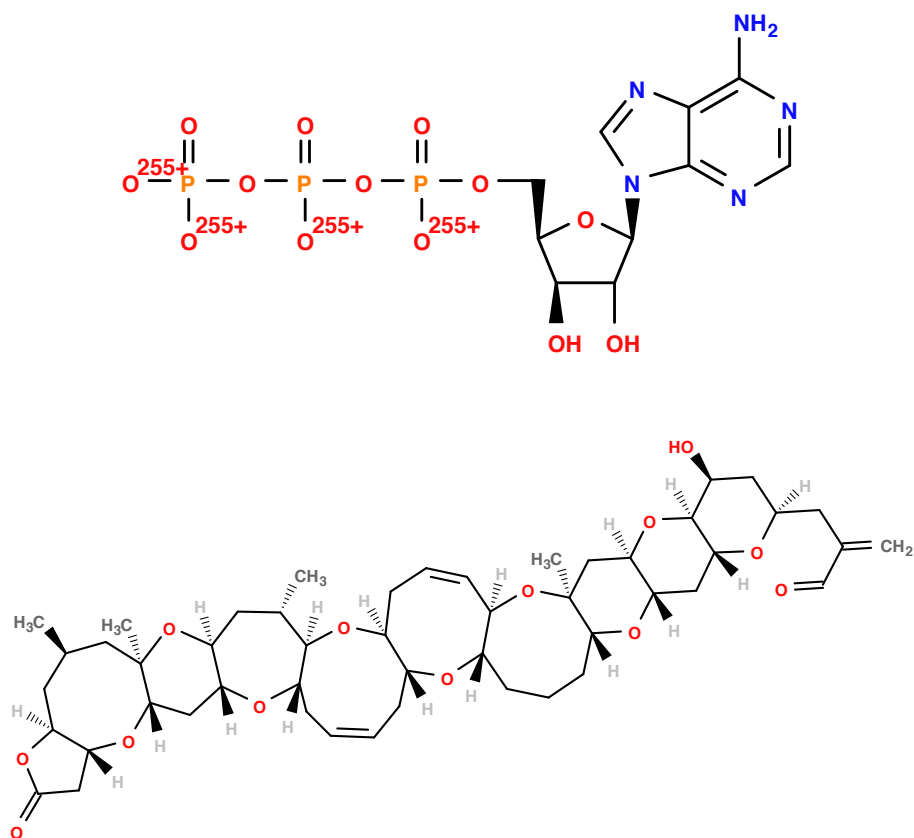


Fig. 5: Conversion from ATP.cdx and Brevetoxin A.mol

Again typeset with `-shell-escape` option:

```
$ pdflatex -shell-escape test.tex
```

This time you will get “smilesobabeling[*NUM*].pdf” and “test.pdf” as Figure 6.

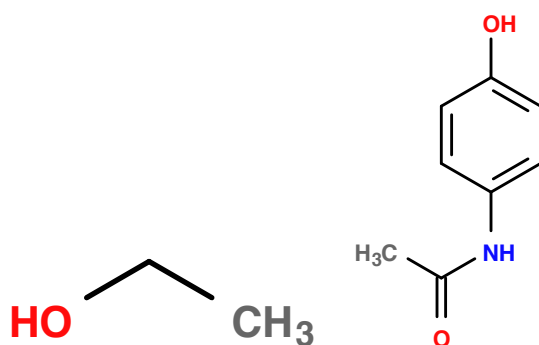


Fig. 6: Conversion from SMILES notation

The syntax is:

```
\smilesobabel[options]{SMILES notation}{obabel options}
```

The only difference from `\chemobabel` is that *SMILES notation* should be placed in the first braces. You can put any kinds of SMILES notations in parameters of `\smilesobabel`, even

when they contain some special characters such as a backslash (\) and a percent symbol (%). Similarly, parameters of `\chemobabel` can also contain these special characters.

```
\smilesobabel[width=28mm]{Cl/C=C/Br}{}
```

```
\smilesobabel[width=28mm]{Cl/C=C\Br}{}
```



Notes for users of earlier versions (\leq chemobabel v0.6 [2015/06/29])

In earlier versions of chemobabel, it was necessary for users to pay another attention when using these \LaTeX special characters. However, the workaround I had suggested before was far from desirable because “casual \LaTeX users” were forced to change category codes of these characters manually.

To solve this problem, I prepared supporting macros inside the package itself (\geq chemobabel v0.7 [2015/08/26]). Now you can put any kinds of SMILES notations directly without worrying about special characters (\backslash , %, etc). Note that the undesirable workaround (like changing `\catcode` inside `\begingroup ... \endgroup`) has now become rather harmful.

3.3 Package options

By default, chemobabel converts chemical structures into PDF images using Inkscape. However, you can switch the image format and the conversion program if you like.

When you are using drivers which do not support PDF figures (such as dvips), it will be helpful if chemobabel converts them into EPS images. For this purpose, you can use:

```
\usepackage[eps]{chemobabel}
```

The option `eps` switches the image format (that is, output format from Inkscape) into EPS (\geq chemobabel v0.9d [2016/02/28]).

You can also change the image conversion program from “Inkscape” to “rsvg-convert” (\geq chemobabel v0.9e [2016/03/07]). The program `rsvg-convert` is faster than `inkscape`, so it will be useful. For this purpose, use `librsvg` option:

```
\usepackage[librsvg]{chemobabel}
```

The default behavior is equivalent to the `pdf` and `inkscape` options:

```
\usepackage[pdf,inkscape]{chemobabel}
```

Moreover, Open Babel v2.4.0 or later versions generate images with a rather large margin. To avoid this, the images are cropped using `pdftocrop` for PDF or `ps2eps` for EPS (\geq chemobabel v0.9i [2022/09/12]). If you want to disable this cropping, please add `nocrop` option:

```
\usepackage[nocrop]{chemobabel}
```

3.4 For Safer Typesetting...

With the basic method described in Section 3, you will get desired output in almost any situation. However, it is sometimes NOT desirable to use `-shell-escape` option, because with that option \LaTeX can call *any* external command-line programs whether you like it or not. This means that there is a chance for unknown programs to be executed, and this can be dangerous especially in case that you don't write your own \LaTeX code. To avoid this problem, I provide a workaround for extracting all `\chemobabel` and `\smilesobabel` commands to another \LaTeX file (Reference: \TeX Forum [4]).

All you have to do is to add `extract` option when loading the package:

```
\usepackage[extract]{chemobabel}
```

When typesetting, usual command

```
$ pdflatex test.tex
```

can be used. With `extract` option, all `chemobabel` does is extracting all `\chemobabel` and `\smilesobabel` commands from the original \LaTeX file. You will get "ChemFigFile.tex" in the same directory, a minimal \LaTeX source file which includes all `\chemobabel` and `\smilesobabel` commands. You can simply typeset it by

```
$ pdflatex -shell-escape ChemFigFile.tex
```

and you will get PDF figures in the same way as described in Section 3. After this, typeset your own document again (this time you should remove `extract` option):

```
$ pdflatex test.tex
```

This time you will get desired output "test.pdf" which includes figures properly. This method will also save you a lot of time for typesetting.

4 Advanced Usage

4.1 Options for Depiction

As I have already mentioned, in both `\chemobabel` and `\smilesobabel`, you can give some *obabel options* in the second braces. Here I will introduce some examples.

- `\smilesobabel{CC(=O)Nc1ccc(cc1)O}{-xa}`

executes a command which gives a formula with all carbon atoms shown:

```
$ obabel -:"CC(=O)Nc1ccc(cc1)O" -O smilesobabelimg[NUM].svg -xa
```

- `\smilesobabel{CC(=O)Nc1ccc(cc1)O}{-xu -xC}`

gives one without element-specific atom coloring or terminal carbon atoms:

```
$ obabel -:"CC(=O)Nc1ccc(cc1)O" -O smilesobabelimg[NUM].svg -xu -xC
```

- `\chemobabel{ATP.cdx}{-xd}`

You can add `-xd` option to remove filenames from ChemDraw figures with Open Babel 2.x.

```
$ obabel -:"ATP.cdx" -O chemobabelimg[NUM].svg -xd
```

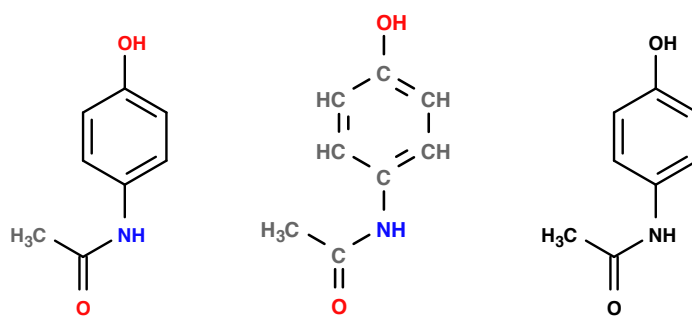


Fig. 7: Acetaminophen (Paracetamol): Nothing, `-xa`, `-xu`

SMILES: CC(=O)Nc1ccc(cc1)O

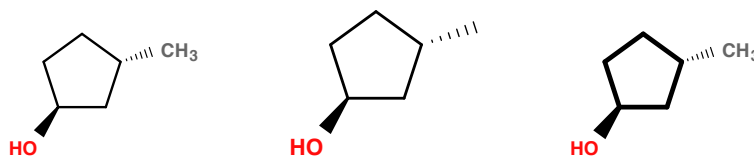


Fig. 8: (1S,3S)-3-Methylcyclopentanol: Nothing, `-xC`, `-xt`

SMILES: C1[C@H](C)C[C@@H](O)C1

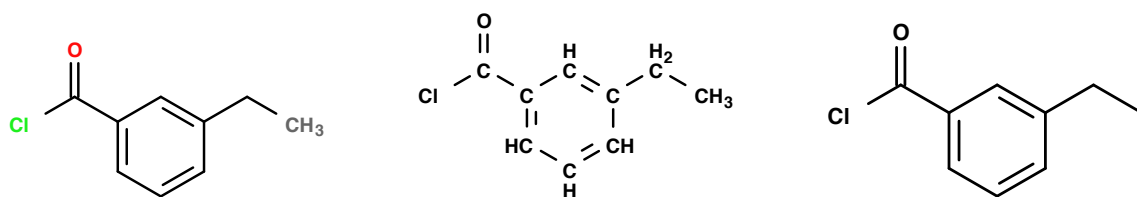


Fig. 9: 3-Ethylbenzoyl chloride: Nothing, -xa -xu, -xu -xC

SMILES: CCc1cccc(c1)C(=O)Cl

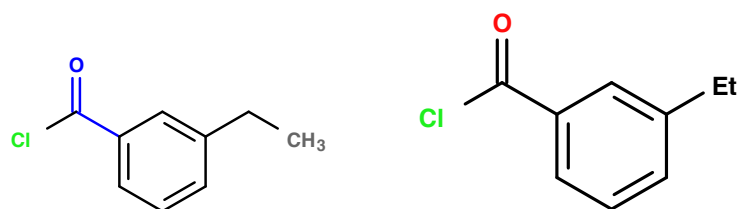


Fig. 10: 3-Ethylbenzoyl chloride: --highlight "cC=O blue", -xA --genalias

SMILES: CCc1cccc(c1)C(=O)Cl

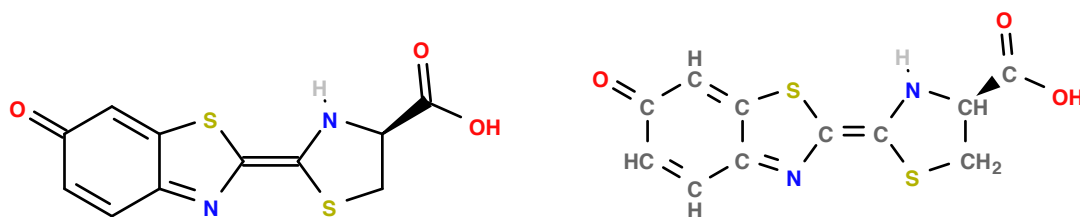


Fig. 11: Firefly luciferin (from MDL Molfiles): Nothing, -xa

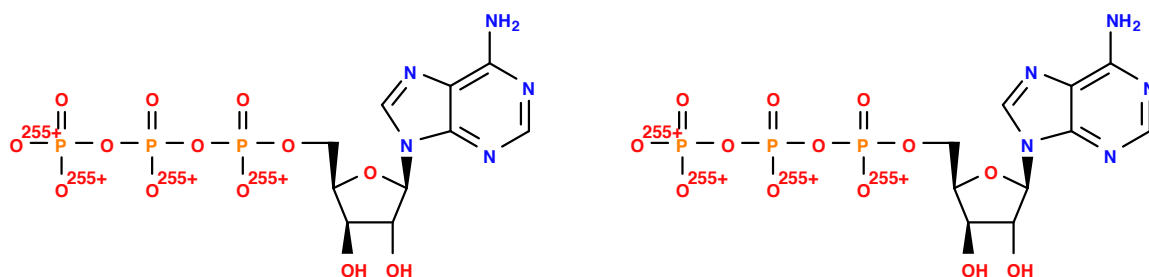


Fig. 12: ATP (from ChemDraw files): Nothing, -xd

For more details, see obabel and babel and SVG depiction (svg) in Open Babel documentation.

4.2 More complex structural formulas

Open Babel can generate more complex structural formulas from SMILES notations.

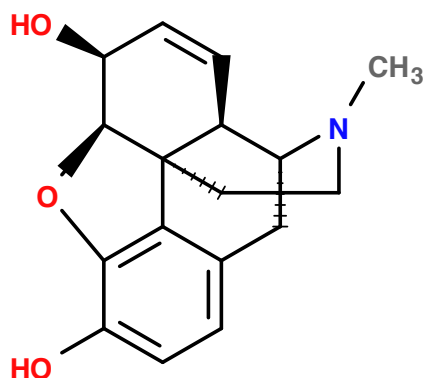


Fig. 13: (-)-Morphine

SMILES: CN1CC[C@]23c4c5ccc(c4O[C@H]2[C@H](C=C[C@H]3[C@H]1C5)O)O

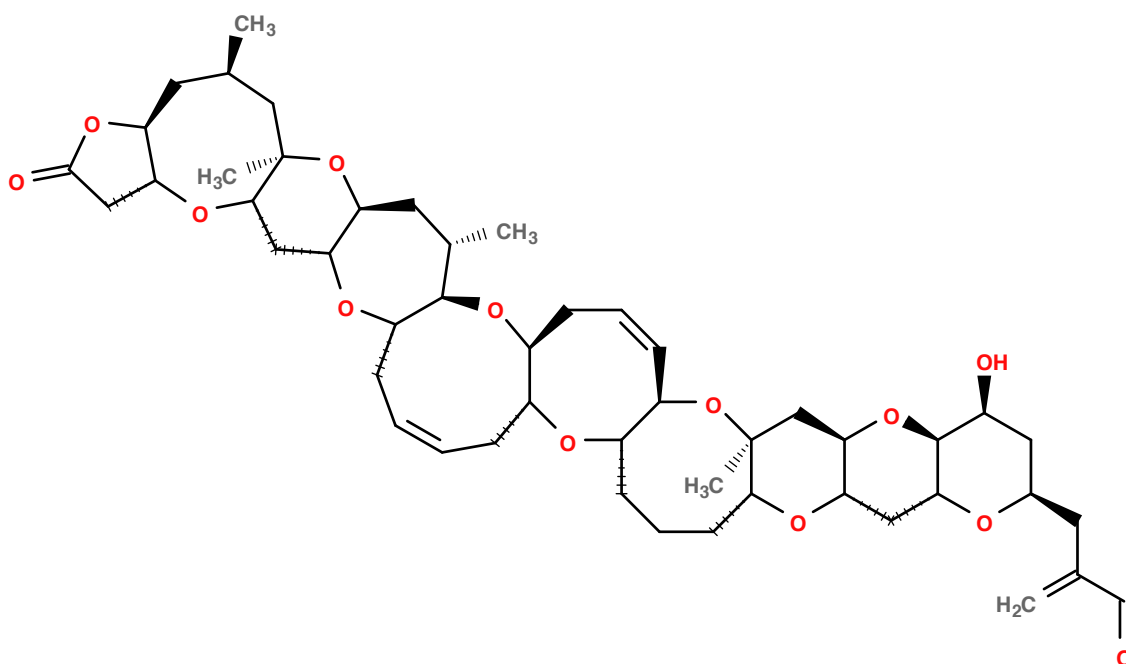


Fig. 14: Brevetoxin A

SMILES (Sorry but too long):

C[C@@H]1C[C@H]2[C@@H](CC(=O)O2)O[C@H]3C[C@@H]4[C@H](C[C@@H]([C@@H]5[C@@H](O4)C/C=C\C[C@H]6O[C@H]7C[C@H]8O[C@H]9C[C@H]10O[C@H]11C[C@H]12O[C@H]13C[C@H]14O[C@H]15C[C@H]16O[C@H]17C[C@H]18O[C@H]19C[C@H]20O[C@H]21C[C@H]22O[C@H]23C[C@H]24O[C@H]25C[C@H]26O[C@H]27C[C@H]28O[C@H]29C[C@H]30O[C@H]31C[C@H]32O[C@H]33C[C@H]34O[C@H]35C[C@H]36O[C@H]37C[C@H]38O[C@H]39C[C@H]40O[C@H]41C[C@H]42O[C@H]43C[C@H]44O[C@H]45C[C@H]46O[C@H]47C[C@H]48O[C@H]49C[C@H]50O[C@H]51C[C@H]52O[C@H]53C[C@H]54O[C@H]55C[C@H]56O[C@H]57C[C@H]58O[C@H]59C[C@H]60O[C@H]61C[C@H]62O[C@H]63C[C@H]64O[C@H]65C[C@H]66O[C@H]67C[C@H]68O[C@H]69C[C@H]70O[C@H]71C[C@H]72O[C@H]73C[C@H]74O[C@H]75C[C@H]76O[C@H]77C[C@H]78O[C@H]79C[C@H]80O[C@H]81C[C@H]82O[C@H]83C[C@H]84O[C@H]85C[C@H]86O[C@H]87C[C@H]88O[C@H]89C[C@H]90O[C@H]91C[C@H]92O[C@H]93C[C@H]94O[C@H]95C[C@H]96O[C@H]97C[C@H]98O[C@H]99C[C@H]100O[C@H]101C[C@H]102O[C@H]103C[C@H]104O[C@H]105C[C@H]106O[C@H]107C[C@H]108O[C@H]109C[C@H]110O[C@H]111C[C@H]112O[C@H]113C[C@H]114O[C@H]115C[C@H]116O[C@H]117C[C@H]118O[C@H]119C[C@H]120O[C@H]121C[C@H]122O[C@H]123C[C@H]124O[C@H]125C[C@H]126O[C@H]127C[C@H]128O[C@H]129C[C@H]130O[C@H]131C[C@H]132O[C@H]133C[C@H]134O[C@H]135C[C@H]136O[C@H]137C[C@H]138O[C@H]139C[C@H]140O[C@H]141C[C@H]142O[C@H]143C[C@H]144O[C@H]145C[C@H]146O[C@H]147C[C@H]148O[C@H]149C[C@H]150O[C@H]151C[C@H]152O[C@H]153C[C@H]154O[C@H]155C[C@H]156O[C@H]157C[C@H]158O[C@H]159C[C@H]160O[C@H]161C[C@H]162O[C@H]163C[C@H]164O[C@H]165C[C@H]166O[C@H]167C[C@H]168O[C@H]169C[C@H]170O[C@H]171C[C@H]172O[C@H]173C[C@H]174O[C@H]175C[C@H]176O[C@H]177C[C@H]178O[C@H]179C[C@H]180O[C@H]181C[C@H]182O[C@H]183C[C@H]184O[C@H]185C[C@H]186O[C@H]187C[C@H]188O[C@H]189C[C@H]190O[C@H]191C[C@H]192O[C@H]193C[C@H]194O[C@H]195C[C@H]196O[C@H]197C[C@H]198O[C@H]199C[C@H]200O[C@H]201C[C@H]202O[C@H]203C[C@H]204O[C@H]205C[C@H]206O[C@H]207C[C@H]208O[C@H]209C[C@H]210O[C@H]211C[C@H]212O[C@H]213C[C@H]214O[C@H]215C[C@H]216O[C@H]217C[C@H]218O[C@H]219C[C@H]220O[C@H]221C[C@H]222O[C@H]223C[C@H]224O[C@H]225C[C@H]226O[C@H]227C[C@H]228O[C@H]229C[C@H]230O[C@H]231C[C@H]232O[C@H]233C[C@H]234O[C@H]235C[C@H]236O[C@H]237C[C@H]238O[C@H]239C[C@H]240O[C@H]241C[C@H]242O[C@H]243C[C@H]244O[C@H]245C[C@H]246O[C@H]247C[C@H]248O[C@H]249C[C@H]250O[C@H]251C[C@H]252O[C@H]253C[C@H]254O[C@H]255C[C@H]256O[C@H]257C[C@H]258O[C@H]259C[C@H]260O[C@H]261C[C@H]262O[C@H]263C[C@H]264O[C@H]265C[C@H]266O[C@H]267C[C@H]268O[C@H]269C[C@H]270O[C@H]271C[C@H]272O[C@H]273C[C@H]274O[C@H]275C[C@H]276O[C@H]277C[C@H]278O[C@H]279C[C@H]280O[C@H]281C[C@H]282O[C@H]283C[C@H]284O[C@H]285C[C@H]286O[C@H]287C[C@H]288O[C@H]289C[C@H]290O[C@H]291C[C@H]292O[C@H]293C[C@H]294O[C@H]295C[C@H]296O[C@H]297C[C@H]298O[C@H]299C[C@H]300O[C@H]301C[C@H]302O[C@H]303C[C@H]304O[C@H]305C[C@H]306O[C@H]307C[C@H]308O[C@H]309C[C@H]310O[C@H]311C[C@H]312O[C@H]313C[C@H]314O[C@H]315C[C@H]316O[C@H]317C[C@H]318O[C@H]319C[C@H]320O[C@H]321C[C@H]322O[C@H]323C[C@H]324O[C@H]325C[C@H]326O[C@H]327C[C@H]328O[C@H]329C[C@H]330O[C@H]331C[C@H]332O[C@H]333C[C@H]334O[C@H]335C[C@H]336O[C@H]337C[C@H]338O[C@H]339C[C@H]340O[C@H]341C[C@H]342O[C@H]343C[C@H]344O[C@H]345C[C@H]346O[C@H]347C[C@H]348O[C@H]349C[C@H]350O[C@H]351C[C@H]352O[C@H]353C[C@H]354O[C@H]355C[C@H]356O[C@H]357C[C@H]358O[C@H]359C[C@H]360O[C@H]361C[C@H]362O[C@H]363C[C@H]364O[C@H]365C[C@H]366O[C@H]367C[C@H]368O[C@H]369C[C@H]370O[C@H]371C[C@H]372O[C@H]373C[C@H]374O[C@H]375C[C@H]376O[C@H]377C[C@H]378O[C@H]379C[C@H]380O[C@H]381C[C@H]382O[C@H]383C[C@H]384O[C@H]385C[C@H]386O[C@H]387C[C@H]388O[C@H]389C[C@H]390O[C@H]391C[C@H]392O[C@H]393C[C@H]394O[C@H]395C[C@H]396O[C@H]397C[C@H]398O[C@H]399C[C@H]400O[C@H]401C[C@H]402O[C@H]403C[C@H]404O[C@H]405C[C@H]406O[C@H]407C[C@H]408O[C@H]409C[C@H]410O[C@H]411C[C@H]412O[C@H]413C[C@H]414O[C@H]415C[C@H]416O[C@H]417C[C@H]418O[C@H]419C[C@H]420O[C@H]421C[C@H]422O[C@H]423C[C@H]424O[C@H]425C[C@H]426O[C@H]427C[C@H]428O[C@H]429C[C@H]430O[C@H]431C[C@H]432O[C@H]433C[C@H]434O[C@H]435C[C@H]436O[C@H]437C[C@H]438O[C@H]439C[C@H]440O[C@H]441C[C@H]442O[C@H]443C[C@H]444O[C@H]445C[C@H]446O[C@H]447C[C@H]448O[C@H]449C[C@H]450O[C@H]451C[C@H]452O[C@H]453C[C@H]454O[C@H]455C[C@H]456O[C@H]457C[C@H]458O[C@H]459C[C@H]460O[C@H]461C[C@H]462O[C@H]463C[C@H]464O[C@H]465C[C@H]466O[C@H]467C[C@H]468O[C@H]469C[C@H]470O[C@H]471C[C@H]472O[C@H]473C[C@H]474O[C@H]475C[C@H]476O[C@H]477C[C@H]478O[C@H]479C[C@H]480O[C@H]481C[C@H]482O[C@H]483C[C@H]484O[C@H]485C[C@H]486O[C@H]487C[C@H]488O[C@H]489C[C@H]490O[C@H]491C[C@H]492O[C@H]493C[C@H]494O[C@H]495C[C@H]496O[C@H]497C[C@H]498O[C@H]499C[C@H]500O[C@H]501C[C@H]502O[C@H]503C[C@H]504O[C@H]505C[C@H]506O[C@H]507C[C@H]508O[C@H]509C[C@H]510O[C@H]511C[C@H]512O[C@H]513C[C@H]514O[C@H]515C[C@H]516O[C@H]517C[C@H]518O[C@H]519C[C@H]520O[C@H]521C[C@H]522O[C@H]523C[C@H]524O[C@H]525C[C@H]526O[C@H]527C[C@H]528O[C@H]529C[C@H]530O[C@H]531C[C@H]532O[C@H]533C[C@H]534O[C@H]535C[C@H]536O[C@H]537C[C@H]538O[C@H]539C[C@H]540O[C@H]541C[C@H]542O[C@H]543C[C@H]544O[C@H]545C[C@H]546O[C@H]547C[C@H]548O[C@H]549C[C@H]550O[C@H]551C[C@H]552O[C@H]553C[C@H]554O[C@H]555C[C@H]556O[C@H]557C[C@H]558O[C@H]559C[C@H]560O[C@H]561C[C@H]562O[C@H]563C[C@H]564O[C@H]565C[C@H]566O[C@H]567C[C@H]568O[C@H]569C[C@H]570O[C@H]571C[C@H]572O[C@H]573C[C@H]574O[C@H]575C[C@H]576O[C@H]577C[C@H]578O[C@H]579C[C@H]580O[C@H]581C[C@H]582O[C@H]583C[C@H]584O[C@H]585C[C@H]586O[C@H]587C[C@H]588O[C@H]589C[C@H]590O[C@H]591C[C@H]592O[C@H]593C[C@H]594O[C@H]595C[C@H]596O[C@H]597C[C@H]598O[C@H]599C[C@H]600O[C@H]601C[C@H]602O[C@H]603C[C@H]604O[C@H]605C[C@H]606O[C@H]607C[C@H]608O[C@H]609C[C@H]610O[C@H]611C[C@H]612O[C@H]613C[C@H]614O[C@H]615C[C@H]616O[C@H]617C[C@H]618O[C@H]619C[C@H]620O[C@H]621C[C@H]622O[C@H]623C[C@H]624O[C@H]625C[C@H]626O[C@H]627C[C@H]628O[C@H]629C[C@H]630O[C@H]631C[C@H]632O[C@H]633C[C@H]634O[C@H]635C[C@H]636O[C@H]637C[C@H]638O[C@H]639C[C@H]640O[C@H]641C[C@H]642O[C@H]643C[C@H]644O[C@H]645C[C@H]646O[C@H]647C[C@H]648O[C@H]649C[C@H]650O[C@H]651C[C@H]652O[C@H]653C[C@H]654O[C@H]655C[C@H]656O[C@H]657C[C@H]658O[C@H]659C[C@H]660O[C@H]661C[C@H]662O[C@H]663C[C@H]664O[C@H]665C[C@H]666O[C@H]667C[C@H]668O[C@H]669C[C@H]670O[C@H]671C[C@H]672O[C@H]673C[C@H]674O[C@H]675C[C@H]676O[C@H]677C[C@H]678O[C@H]679C[C@H]680O[C@H]681C[C@H]682O[C@H]683C[C@H]684O[C@H]685C[C@H]686O[C@H]687C[C@H]688O[C@H]689C[C@H]690O[C@H]691C[C@H]692O[C@H]693C[C@H]694O[C@H]695C[C@H]696O[C@H]697C[C@H]698O[C@H]699C[C@H]700O[C@H]701C[C@H]702O[C@H]703C[C@H]704O[C@H]705C[C@H]706O[C@H]707C[C@H]708O[C@H]709C[C@H]710O[C@H]711C[C@H]712O[C@H]713C[C@H]714O[C@H]715C[C@H]716O[C@H]717C[C@H]718O[C@H]719C[C@H]720O[C@H]721C[C@H]722O[C@H]723C[C@H]724O[C@H]725C[C@H]726O[C@H]727C[C@H]728O[C@H]729C[C@H]730O[C@H]731C[C@H]732O[C@H]733C[C@H]734O[C@H]735C[C@H]736O[C@H]737C[C@H]738O[C@H]739C[C@H]740O[C@H]741C[C@H]742O[C@H]743C[C@H]744O[C@H]745C[C@H]746O[C@H]747C[C@H]748O[C@H]749C[C@H]750O[C@H]751C[C@H]752O[C@H]753C[C@H]754O[C@H]755C[C@H]756O[C@H]757C[C@H]758O[C@H]759C[C@H]760O[C@H]761C[C@H]762O[C@H]763C[C@H]764O[C@H]765C[C@H]766O[C@H]767C[C@H]768O[C@H]769C[C@H]770O[C@H]771C[C@H]772O[C@H]773C[C@H]774O[C@H]775C[C@H]776O[C@H]777C[C@H]778O[C@H]779C[C@H]780O[C@H]781C[C@H]782O[C@H]783C[C@H]784O[C@H]785C[C@H]786O[C@H]787C[C@H]788O[C@H]789C[C@H]790O[C@H]791C[C@H]792O[C@H]793C[C@H]794O[C@H]795C[C@H]796O[C@H]797C[C@H]798O[C@H]799C[C@H]800O[C@H]801C[C@H]802O[C@H]803C[C@H]804O[C@H]805C[C@H]806O[C@H]807C[C@H]808O[C@H]809C[C@H]810O[C@H]811C[C@H]812O[C@H]813C[C@H]814O[C@H]815C[C@H]816O[C@H]817C[C@H]818O[C@H]819C[C@H]820O[C@H]821C[C@H]822O[C@H]823C[C@H]824O[C@H]825C[C@H]826O[C@H]827C[C@H]828O[C@H]829C[C@H]830O[C@H]831C[C@H]832O[C@H]833C[C@H]834O[C@H]835C[C@H]836O[C@H]837C[C@H]838O[C@H]839C[C@H]840O[C@H]841C[C@H]842O[C@H]843C[C@H]844O[C@H]845C[C@H]846O[C@H]847C[C@H]848O[C@H]849C[C@H]850O[C@H]851C[C@H]852O[C@H]853C[C@H]854O[C@H]855C[C@H]856O[C@H]857C[C@H]858O[C@H]859C[C@H]860O[C@H]861C[C@H]862O[C@H]863C[C@H]864O[C@H]865C[C@H]866O[C@H]867C[C@H]868O[C@H]869C[C@H]870O[C@H]871C[C@H]872O[C@H]873C[C@H]874O[C@H]875C[C@H]876O[C@H]877C[C@H]878O[C@H]879C[C@H]880O[C@H]881C[C@H]882O[C@H]883C[C@H]884O[C@H]885C[C@H]886O[C@H]887C[C@H]888O[C@H]889C[C@H]890O[C@H]891C[C@H]892O[C@H]893C[C@H]894O[C@H]895C[C@H]896O[C@H]897C[C@H]898O[C@H]899C[C@H]900O[C@H]901C[C@H]902O[C@H]903C[C@H]904O[C@H]905C[C@H]906O[C@H]907C[C@H]908O[C@H]909C[C@H]910O[C@H]911C[C@H]912O[C@H]913C[C@H]914O[C@H]915C[C@H]916O[C@H]917C[C@H]918O[C@H]919C[C@H]920O[C@H]921C[C@H]922O[C@H]923C[C@H]924O[C@H]925C[C@H]926O[C@H]927C[C@H]928O[C@H]929C[C@H]930O[C@H]931C[C@H]932O[C@H]933C[C@H]934O[C@H]935C[C@H]936O[C@H]937C[C@H]938O[C@H]939C[C@H]940O[C@H]941C[C@H]942O[C@H]943C[C@H]944O[C@H]945C[C@H]946O[C@H]947C[C@H]948O[C@H]949C[C@H]950O[C@H]951C[C@H]952O[C@H]953C[C@H]954O[C@H]955C[C@H]956O[C@H]957C[C@H]958O[C@H]959C[C@H]960O[C@H]961C[C@H]962O[C@H]963C[C@H]964O[C@H]965C[C@H]966O[C@H]967C[C@H]968O[C@H]969C[C@H]970O[C@H]971C[C@H]972O[C@H]973C[C@H]974O[C@H]975C[C@H]976O[C@H]977C[C@H]978O[C@H]979C[C@H]980O[C@H]981C[C@H]982O[C@H]983C[C@H]984O[C@H]985C[C@H]986O[C@H]987C[C@H]988O[C@H]989C[C@H]990O[C@H]991C[C@H]992O[C@H]993C[C@H]994O[C@H]995C[C@H]996O[C@H]997C[C@H]998O[C@H]999C[C@H]1000O[C@H]1001C[C@H]1002O[C@H]1003C[C@H]1004O[C@H]1005C[C@H]1006O[C@H]1007C[C@H]1008O[C@H]1009C[C@H]1010O[C@H]1011C[C@H]1012O[C@H]1013C[C@H]1014O[C@H]1015C[C@H]1016O[C@H]1017C[C@H]1018O[C@H]1019C[C@H]1020O[C@H]1021C[C@H]1022O[C@H]1023C[C@H]1024O[C@H]1025C[C@H]1026O[C@H]1027C[C@H]1028O[C@H]1029C[C@H]1030O[C@H]1031C[C@H]1032O[C@H]1033C[C@H]1034O[C@H]1035C[C@H]1036O[C@H]1037C[C@H]1038O[C@H]1039C[C@H]1040O[C@H]1041C[C@H]1042O[C@H]1043C[C@H]1044O[C@H]1045C[C@H]1046O[C@H]1047C[C@H]1048O[C@H]1049C[C@H]1050O[C@H]1051C[C@H]1052O[C@H]1053C[C@H]1054O[C@H]1055C[C@H]1056O[C@H]1057C[C@H]1058O[C@H]1059C[C@H]1060O[C@H]1061C[C@H]1062O[C@H]1063C[C@H]1064O[C@H]1065C[C@H]1066O[C@H]1067C[C@H]1068O[C@H]1069C[C@H]1070O[C@H]1071C[C@H]1072O[C@H]1073C[C@H]1074O[C@H]1075C[C@H]1076O[C@H]1077C[C@H]1078O[C@H]1079C[C@H]1080O[C@H]1081C[C@H]1082O[C@H]1083C[C@H]1084O[C@H]1085C[C@H]1086O[C@H]1087C[C@H]1088O[C@H]1089C[C@H]1090O[C@H]1091C[C@H]1092O[C@H]1093C[C@H]1094O[C@H]1095C[C@H]1096O[C@H]1097C[C@H]1098O[C@H]1099C[C@H]1100O[C@H]1101C[C@H]1102O[C@H]1103C[C@H]1104O[C@H]1105C[C@H]1106O[C@H]1107C[C@H]1108O[C@H]1109C[C@H]1110O[C@H]1111C[C@H]1112O[C@H]1113C[C@H]1114O[C@H]1115C[C@H]1116O[C@H]1117C[C@H]1118O[C@H]1119C[C@H]1120O[C@H]1121C[C@H]1122O[C@H]1123C[C@H]1124O[C@H]1125C[C@H]1126O[C@H]1127C[C@H]1128O[C@H]1129C[C@H]1130O[C@H]1131C[C@H]1132O[C@H]1133C[C@H]1134O[C@H]1135C[C@H]1136O[C@H]1137C[C@H]1138O[C@H]1139C[C@H]1140O[C@H]1141C[C@H]1142O[C@H]1143C[C@H]1144O[C@H]1145C[C@H]1146O[C@H]1147C[C@H]1148O[C@H]1149C[C@H]1150O[C@H]1151C[C@H]1152O[C@H]1153C[C@H]1154O[C@H]1155C[C@H]1156O[C@H]1157C[C@H]1158O[C@H]1159C[C@H]1160O[C@H]1161C[C@H]1162O[C@H]1163C[C@H]1164O[C@H]1165C[C@H]1166O[C@H]1167C[C@H]1168O[C@H]1169C[C@H]1170O[C@H]1171C[C@H]1172O[C@H]1173C[C@H]1174O[C@H]1175C[C@H]1176O[C@H]1177C[C@H]1178O[C@H]1179C[C@H]1180O[C@H]1181C[C@H]1182O[C@H]1183C[C@H]1184O[C@H]1185C[C@H]1186O[C@H]1187C[C@H]1188O[C@H]1189C[C@H]1190O[C@H]1191C[C@H]1192O[C@H]1193C[C@H]1194O[C@H]1195C[C@H]1196O[C@H]1197C[C@H]1198O[C@H]1199C[C@H]1200O[C@H]1201C[C@H]1202O[C@H]1203C[C@H]1204O[C@H]1205C[C@H]1206O[C@H]1207C[C@H]1208O[C@H]1209C[C@H]1210O[C@H]1211C[C@H]1212O[C@H]1213C[C@H]1214O[C@H]1215C[C@H]1216O[C@H]1217C[C@H]1218O[C@H]1219C[C@H]1220O[C@H]1221C[C@H]1222O[C@H]1223C[C@H]1224O[C@H]1225C[C@H]1226O[C@H]1227C[C@H]1228O[C@H]1229C[C@H]1230O[C@H]1231C[C@H]1232O[C@H]1233C[C@H]1234O[C@H]1235C[C@H]1236O[C@H]1237C[C@H]1238O[C@H]1239C[C@H]1240O[C@H]1241C[C@H]1242O[C@H]1243C[C@H]1244O[C@H]1245C[C@H]1246O[C@H]1247C[C@H]1248O[C@H]1249C[C@H]1250O[C@H]1251C[C@H]1252O[C@H]1253C[C@H]1254O[C@H]1255C[C@H]1256O[C@H]1257C[C@H]1258O[C@H]1259C[C@H]1260O[C@H]1261C[C@H]1262O[C@H]1263C[C@H]1264O[C@H]1265C[C@H]1266O[C@H]1267C[C@H]1268O[C@H]1269C[C@H]1270O[C@H]1271C[C@H]1272O[C@H]1273C[C@H]1274O[C@H]1275C[C@H]1276O[C@H]1277C[C@H]1278O[C@H]1279C[C@H]1280O[C@H]1281C[C@H]1282O[C@H]1283C[C@H]1284O[C@H]1285C[C@H]1286O[C@H]1287C[C@H]1288O[C@H]1289C[C@H]1290O[C@H]1291C[C@H]1292O[C@H]1293C[C@H]1294O[C@H]1295C[C@H]1296O[C@H]1297C[C@H]1298O[C@H]1299C[C@H]1300O[C@H]1301C[C@H]1302O[C@H]1303C[C@H]1304O[C@H]1305C[C@H]1306O[C@H]1307C[C@H]1308O[C@H]1309C[C@H]1310O[C@H]1311C[C@H]1312O[C@H]1313C[C@H]1314O[C@H]1315C[C@H]1316O[C@H]1317C[C@H]1318O[C@H]1319C[C@H]1320O[C@H]1321C[C@H]1322O[C@H]1323C[C@H]1324O[C@H]1325C[C@H]1326O[C@H]1327C[C@H]1328O[C@H]1329C[C@H]1330O[C@H]1331C[C@H]1332O[C@H]1333C[C@H]1334O[C@H]1335C[C@H]1336O[C@H]1337C[C@H]1338O[C@H]1339C[C@H]1340O[C@H]1341C[C@H]1342O[C@H]1343C[C@H]1344O[C@H]1345C[C@H]1346O[C@H]1347C[C@H]1348O[C@H]1349C[C@H]1350O[C@H]1351C[C@H]1352O[C@H]1353C[C@H]1354O[C@H]1355C[C@H]1356O[C@H]1357C[C@H]1358O[C@H]1359C[C@H]1360O[C@H]1361C[C@H]1362O[C@H]1363C[C@H]1364O[C@H]1365C[C@H]1366O[C@H]1367C[C@H]1368O[C@H]1369C[C@H]1370O[C@H]1371C[C@H]1372O[C@H]1373C[C@H]1374O[C@H]1375C[C@H]1376O[C@H]1377C[C@H]1378O[C@H]1379C[C@H]1380O[C@H]1381C[C@H]1382O[C@H]1383C[C@H]1384O[C@H]1385C[C@H]1386O[C@H]1387C[C@H]1388O[C@H]1389C[C@H]1390O[C@H]1391C[C@H]1392O[C@H]1393C[C@H]1394O[C@H]1395C[C@H]1396O[C@H]1397C[C@H]1398O[C@H]1399C[C@H]1400O[C@H]1401C[C@H]1402O[C@H]1403C[C@H]1404O[C@H]1405C[C@H]1406O[C@H]1407C[C@H]1408O[C@H]1409C[C@H]1410O[C@H]1411C[C@H]1412O[C@H]1413C[C@H]1414O[C@H]1415C[C@H]1416O[C@H]1417C[C@H]1418O[C@H]1419C[C@H]1420O[C@H]1421C[C@

Cephalostatin-1 (Example from Wikipedia):

C[C@@](C)(O1)C[C@@H](O)[C@@]1(O2)[C@H](C)[C@H]3CC=C4[C@]3(C2)C(=O)C[C@H]5[C@H]4CC[C@H]5

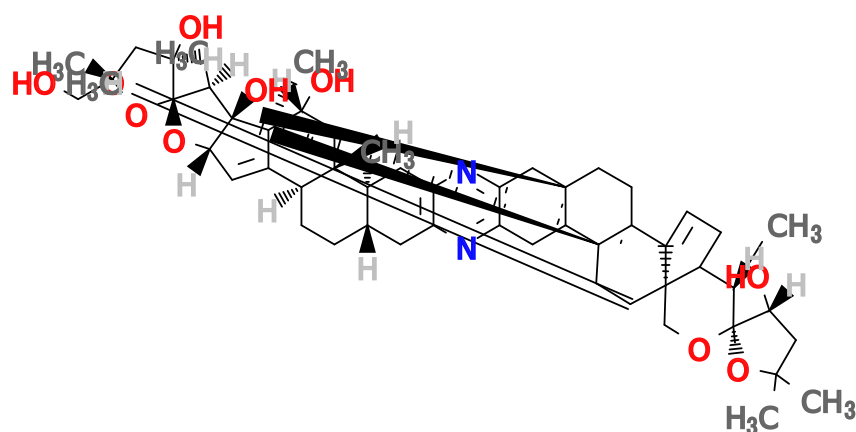


Fig. 15: Cephalostatin-1 (figure from Open Babel for Mac 2.3.1)

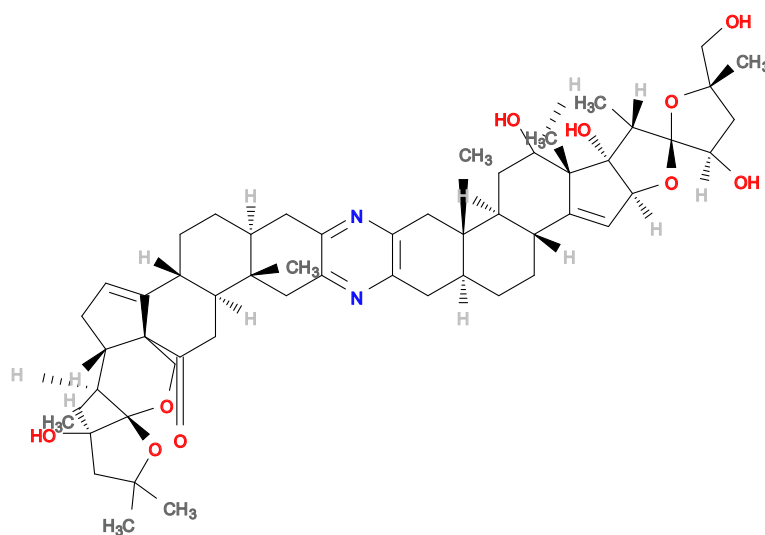


Fig. 16: Cephalostatin-1 (figure from Open Babel for Win 2.3.2)

Sesamin: c1cc2c(cc1C3C4COC(C4C03)c5ccc6c(c5)OC06)OC02

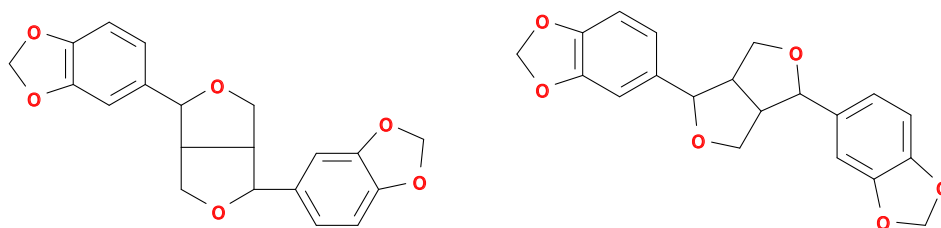


Fig. 17: Sesamin (figures from Open Babel for Mac 2.3.1; Normal and --gen2d)

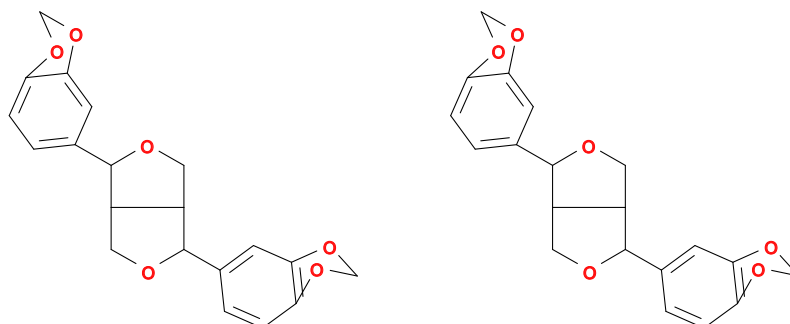


Fig. 18: Sesamin (figures from Open Babel for Win 2.3.2; Normal and --gen2d)

Also, Open Babel may generate wrong structures from SMILES notations. Here we can see the formula generated from a SMILES notation lacks one double bond.

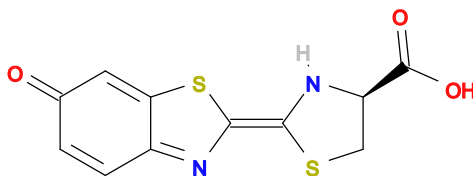


Fig. 19: Firefly luciferin: exact structure from ChemSpider ID 4588411

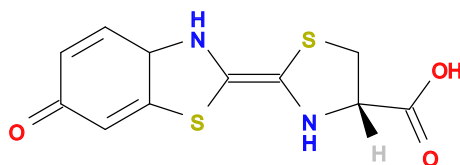


Fig. 20: Firefly luciferin?: output from Open Babel for Win 2.3.2

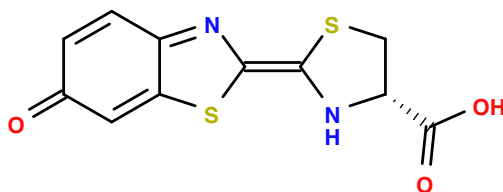


Fig. 21: Firefly luciferin?: output from current Open Babel

SMILES: C1[C@@H](N/C(=C\2/nc3c(=CC(=O)C=C3)s2)/S1)C(=O)O

We can avoid all these problems by using `\chemobabel` with `.mol` or `.cdx` files, instead of `\smilesobabel`. However, the fact that many complex structures can be generated from only one character string is very interesting, isn't it?

5 Note for Compatibility

5.1 Advantages of chemobabel

The macros `\smiles` (written by Noel O’Boyle [1]) and `\obabel` (depends on `graphvizObabel.sty` by Jakob Lykke Andersen’s [3]) can be used as follows:

```
\newcounter{smilescounter}
\setcounter{smilescounter}{1}
\newcommand{\smiles}[1]{
  \immediate\write18{obabel -:"#1" -O smilesimg\arabic{smilescounter}.png}
  \includegraphics{smilesimg\arabic{smilescounter}.png}
  \addtocounter{smilescounter}{1}
}
\usepackage{graphvizObabel}
\smiles{CCO}
\obabel[scale=0.6]{CCO}
```

However, `\smiles` command creates raster graphics (.png) and has no options to pass to `\includegraphics`, and `\obabel` command sometimes does not work because of lack of file extension (.pdf) and possibility of giving errors with empty optional parameters. These problems are solved in `\smilesobabel`.

6 Limitations and Alternatives

This method relies on Open Babel for generating structural formulas, even without drawing anything by yourself. Of course you can modify and customize these structures to some extent, but this method will not be suitable for fine control of the output. Also, computer-generated formulas may sometimes be unnatural and undesirable. Therefore this method will be useful only when you need a simple method for inserting structural formulas with minimal modification. If you are not satisfied with the output, consider using \LaTeX or chemfig packages instead.

I also found following packages which can be used for similar purpose:

- mol2chemfig: convert chemical structures from MDL molfile format to chemfig source code

7 Technical information

According to SMILES Tutorial, special characters are used in SMILES syntax are the followings:

- Roman alphabets: A-Z, a-z
- Numbers: 1-10
- Brackets: [] ()
- Others:
 - * (unspecified atomic number)
 - . (disconnection)
 - + (charge sign)
 - - (single bond or charge sign)
 - = (double bond)
 - # (triple bond)
 - \$ (quadruple bond)
 - : (aromatic bond)
 - % (used when more than 10 ring closures)
 - /, \ (configuration around double bonds)
 - @ (tetrahedral chirality)
 - > (reaction)

It is important to prevent these characters from being translated as \LaTeX special characters. However, a backslash (\backslash) is a default escape character in \LaTeX , and a percent symbol (%) also has a special meaning as a comment character. In chemobabel (higher than v0.7), all these characters are handled properly by changing category codes temporarily (like $\backslash\text{verb}$).

To do (if possible):

- Check whether `obabel` and `inkscape` are successfully installed or not, before running programs.

8 Version History

2014/12/01 v0.1	Made public as <code>smilesobabel.sty</code>
2014/12/02 v0.2	Add options which can be passed to <code>obabel</code> .
2014/12/07 v0.3	Change name of package: <code>chemobabel.sty</code> Images are stored in <code>chemobabelimgdir</code> . Add <code>\chemobabel</code> command.
2014/12/09 v0.4	Fix a bug: (Thanks: Yusuke Terada) Extra spaces at the end of lines are removed.
2014/12/20 v0.5	Add <code>extract</code> option.
2015/06/29 v0.6	Improve warning messages.
2015/08/26 v0.7	Solve <code>\catcode</code> -related problems in <code>\smilesobabel</code> .
2015/08/27 v0.8	Improve <code>\smilesobabel</code> a little.
2015/08/28 v0.9	Improve <code>\smilesobabel</code> : (Thanks: ZR) Exclude ϵ -TeX dependency.
2015/08/29 v0.9a	Improve <code>\chemobabel</code> a little.
2016/01/06 v0.9b	Support LuaTeX-0.85.0 and later versions.
2016/02/09 v0.9c	Fix a bug; forgotten in v0.9b.
2016/02/28 v0.9d	Add <code>eps</code> and <code>pdf</code> options.
2016/03/07 v0.9e	Add <code>librsvg</code> and <code>inkscape</code> options.
2016/10/23 v0.9f	Incorporate <code>extract</code> option into main package.
2018/01/23 v0.9g	Minor refactor of code.
2022/09/11 v0.9h	Support Inkscape version 1.0 command line syntax.
2022/09/12 v0.9i	Crop images before inclusion, add <code>nocrop</code> option.
2022/09/19 v0.9j	Check shell-escape status to show better error message.

Important!

- In Version 0.2, the number of parameters set in `\smilesobabel` is changed!
- From Version 0.3, the package name is changed to `chemobabel.sty`.
- From Version 0.7, the `\catcode`-related workaround is rather harmful.

References

- [1] Cheer up your L^AT_EX with SMILES support – Noel O’Blog
- [2] Cheer up your L^AT_EX with SMILES support II – Noel O’Blog
- [3] L^AT_EX: Graphviz and OpenBabel – Jakob Lykke Andersen
- [4] 文章内の画像のみを表示する方法 – T_EX Forum
(How to extract only figures in L^AT_EX source)
- [5] 化学構造式を T_EX で (1) : 自動化による簡単生成 – Acetaminophen’s diary
(Chemical Structural Formula in L^AT_EX (1): An Easy Method by Auto-generation)
- [6] 化学構造式を T_EX で (2) : 自動化の注意点と解消法 – Acetaminophen’s diary
(Chemical Structural Formula in L^AT_EX (2): Important Notes for Auto-generation and Solution)
- [7] 化学構造式を T_EX で (3) : 補足事項 – Acetaminophen’s diary
(Chemical Structural Formula in L^AT_EX (3): Supplement)

Note: Acetaminophen’s diary is my own blog! (Sorry, but only in Japanese)

I added a post about this package on the 8th day of T_EX & L^AT_EX Advent Calendar 2014 (in Japanese):

- 誰でも簡単！ 化学構造式を L^AT_EX に取り込むパッケージ – Acetaminophen’s diary
(An easy way to insert chemical structural formulas into L^AT_EX documents)