

chemobabel

— Chemical Structures from MDL Molfiles, ChemDraw Files or SMILES Notations —
化学構造式を MOL ファイルや ChemDraw ファイル, SMILES 表記法から自動生成

Hironobu Yamashita (@aminophen)

September 19, 2022

この文書では、化学構造式を生成して L^AT_EX の文書中に挿入するためのパッケージである chemobabel.sty と同梱リソースの使い方を説明します。構造式は Open Babel の機能を利用することにより、MDL Molfile, ChemDraw ファイル, さらには SMILES 表記法を含むさまざまな化学データ形式から生成できます。以下の構造式はこの方法によって出力されたものです。

For English documentation, please refer to chemobabel-en.pdf.

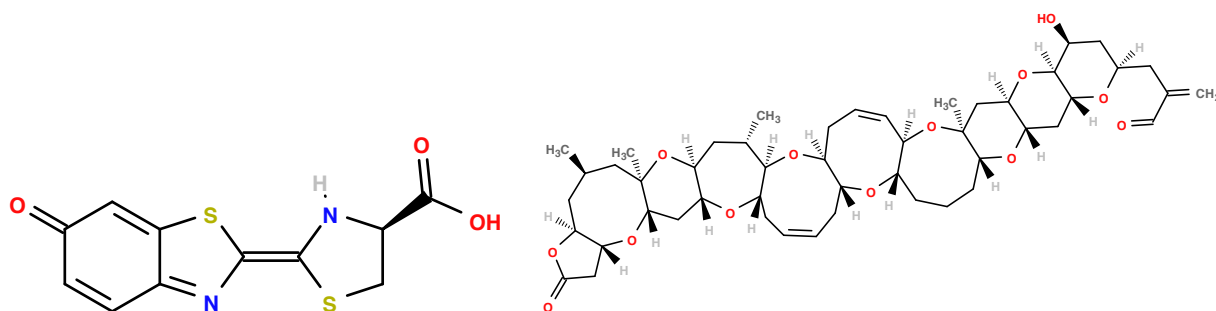


Fig. 1: Firefly luciferin & Brevetoxin A (from MDL Molfiles)

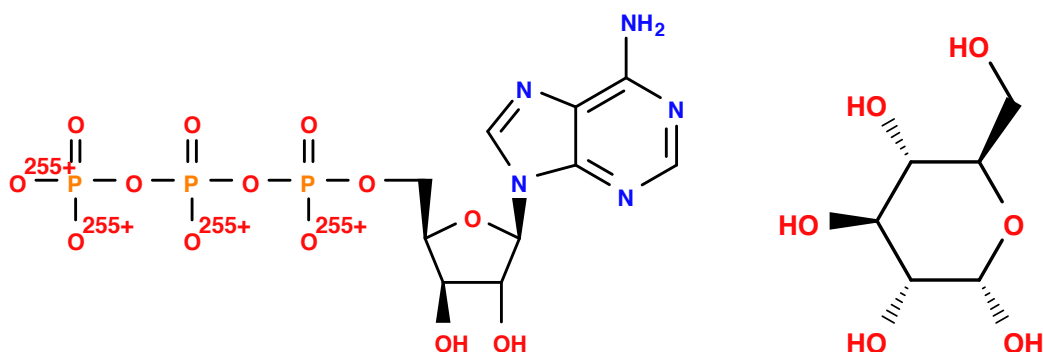


Fig. 2: ATP (Adenosine triphosphate) & Glucose (from ChemDraw files)

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>

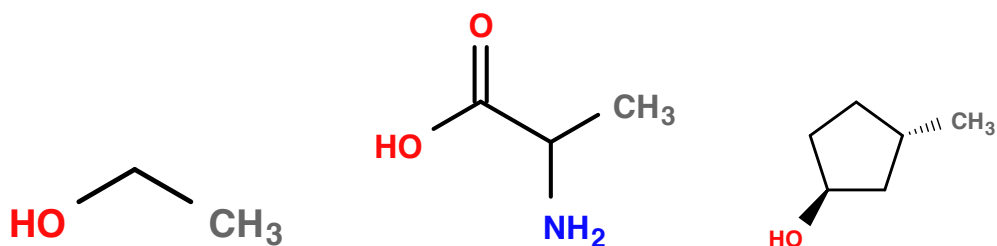


Fig. 3: Ethanol, Alanine & (1S,3S)-3-Methylcyclopentanol (from SMILES notations)

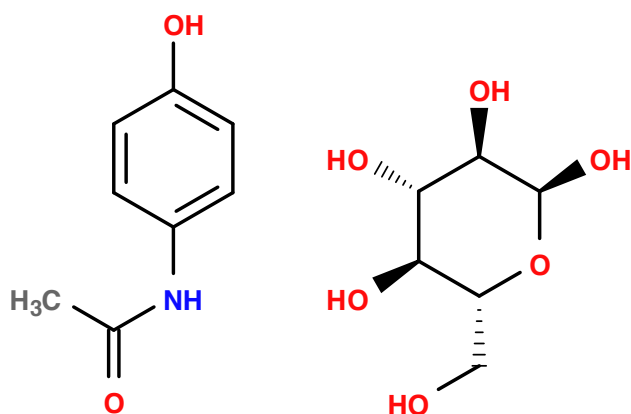


Fig. 4: Acetaminophen (Paracetamol) & α-D-Glucose (from SMILES notations)

このパッケージは GitHub で管理しています。

<https://github.com/aminophen/chemobabel>

問い合わせや改善案などは GitHub の Issues へお願いします。

Contents

1	はじめに	1
1.1	開発の動機	1
1.2	着想とアプローチ	2
2	パッケージを使う前に	3
2.1	依存ソフトウェアのインストール	3
2.2	パッケージの読み込み	3
3	基本的な使い方	4
3.1	化学構造式ファイルからの変換	4
3.2	SMILES 表記法からの変換	4
3.3	パッケージオプション	7
3.4	安全なタイプセットのために	8
4	高度な使い方	9
4.1	描画オプション	9
4.2	より複雑な構造式	11
5	互換性に関する注意	15
5.1	chemobabel パッケージの有利な点	15
6	この方法の限界と代替案	16
7	技術情報	17
8	更新履歴	18

1 はじめに

1.1 開発の動機

ご存じのとおり， \LaTeX は世界中で使われています。しかし，化学構造式を描くという点で考えると， \LaTeX 文書中に構造式を挿入する手段は非常に限られています。この目的を達成する確実かつ容易な方法が存在しないことは，化学分野における多くの研究者が \LaTeX システムを使わない要因になっていると考えられます。

すでに CTAN からいくつかのパッケージが利用可能です：

- \XLaTeX : 化学構造式を描くためのパッケージ集
- `chemfig`: TikZ を利用して構造式を描画するパッケージ

これらのパッケージは確実で，あらゆる構造式を一貫性をもって描画することができます。しかし，十分にその機能を活用できるようになるには相当の熟練を要します。

化学者向けの専用ツール (ChemDraw など) で PDF や EPS 形式で出力し，`\includegraphics` によって構造式を取り込むという手段もあります。この場合，化学用のフォーマット (`.cdx` や `.mol` など) と画像フォーマット (`.pdf` や `.eps` など) の両方で保存する必要があります。

新しいパッケージ `chemobabel.sty` は， \LaTeX 文書中に化学構造式を挿入する必要がある化学者に新たな選択肢を提供することでしょう。この方法では，Open Babel によって SVG 形式の構造式画像を生成し，さらに Inkscape によって SVG から PDF に変換します。どちらのプログラムもオープンソースかつクロスプラットフォームであり，盛んに開発が行われています。

■Open Babel について Open Babel はさまざまな化学データを扱うために特別に設計されたツールです。オープンな共同プロジェクトで，分子モデリング・化学・固体物性・生化学や関連分野のデータを誰もが検索・変換・分析・保存できることを目指しています。詳細は公式サイトを参照してください。

1.2 着想とアプローチ

最初に Open Babel を用いて SMILES 表記から化学構造式を生成するという Noel O'Blog の記事 [1], [2] とコメント [3] を見つけました。これらの投稿に示された手法は非常におもしろいと思いましたが、すべてのプログラムが `-shell-escape` オプションによって実行されてしまうことがあまり好きではありませんでした。それと同じころ、あるソース中の `\includegraphics` を別のソースに書き出すというマクロ [4] を偶然みつけ、このとき私は `-shell-escape` が必要な最小のコードだけを抽出するという着想を得ました。

この方法は ChemDraw ファイルを画像に変換する目的にも応用できると考えたのですが、この発想を実現したプロジェクトが見当たらず、ただ一つ見つけた Sourceforge の Chemdraw in L^AT_EX はすでに終了していました。そこで、化学構造式を L^AT_EX 文書中に挿入する手順を簡略化するパッケージを開発しようと考えました。

2 パッケージを使う前に

2.1 依存ソフトウェアのインストール

初めにコンピュータに Open Babel と Inkscape (または librsvg) をインストールし、コマンドライン実行ファイルである `obabel` と `inkscape` (または `rsvg-convert`) に `PATH` を通します。正しく `PATH` が通っているかどうかは以下のコマンドによって確認できます：UNIX 系の場合：

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

Windows の場合：

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

何らかのディレクトリ名が返ってきた場合は、正しくインストールできているはずです。

2.2 パッケージの読み込み

プリアンブルに

```
\usepackage{chemobabel}
```

と宣言して `chemobabel` パッケージを読み込みます。これは `graphicx` パッケージに依存しますので、もし (u)pL^AT_EX を使っていて DVI ドライバに `dvipdfmx` を使用する場合はドライバオプション `[dvipdfmx]` をクラスオプション (グローバル) に付けるのを忘れないようにしてください。

3 基本的な使い方

3.1 化学構造式ファイルからの変換

ATP.cdx は ATP (アデノシン三リン酸) の構造式を ChemDraw で描画したファイル、Brevetoxin A.mol は ChemSpider から入手できるファイルです (もとのファイル名は 9041092.mol となっています)。

`\chemobabel` コマンドは本文中で以下のように使います。ここでは元のファイルを `draw` というサブディレクトリに置いたので、ファイル名の前に `draw/` を付けました。もちろん \LaTeX ソースと同一ディレクトリに置くこともできます。

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

```

\LaTeX ファイル名が `test.tex` だとすると、以下のように `-shell-escape` オプションを与えてタイプセットします。このオプションを与えることで、 \LaTeX が外部プログラムである `obabel` や `inkscape` を呼んで実行できるようになります。

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

すると `chemobabelimgdir` 以下に「`chemobabelimg[NUM].pdf`」というファイルが生成します。得られた `test.dvi` に対して

```
$ dvipdfmx test.dvi
```

と実行すると、出力される「`test.pdf`」は図 5 のようになります。

構文は以下のとおりです。

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

引数として、中括弧の 1 つめにファイル名を与え、2 つめに `obabel` に送るオプション (例は第 4.1 節を参照) を指定してください。2 つめの引数は空にすることもできます。

また、オプションとして角括弧に `\includegraphics` に渡すオプションを指定することができます。このオプションは省略可能で、何も指定しない場合は `scale=1` が適用されます。

3.2 SMILES 表記法からの変換

\LaTeX ソースファイル中に直接 SMILES 表記を書き込んで、構造式に変換することもできます。例えば、`CCO` はエタノールの SMILES 表記、`CC(=O)Nc1ccc(cc1)O` はアセトアミノフェンの SMILES 表記です。`\smilesobabel` コマンドを以下のように使います：

```
\smilesobabel{CCO}{}  

```

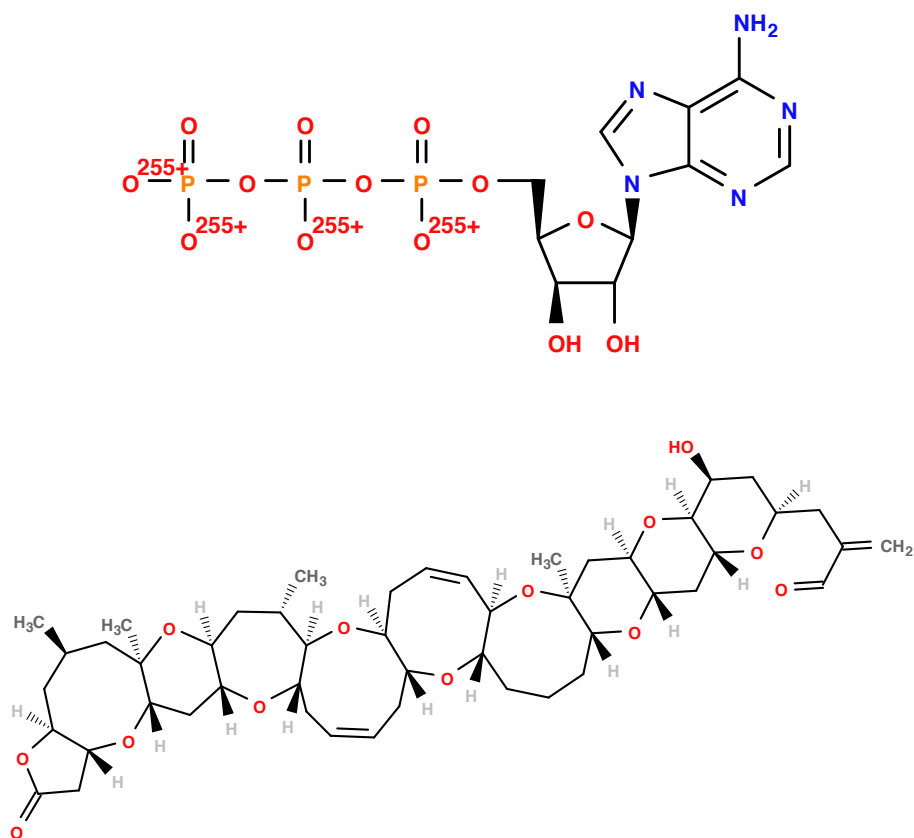


Fig. 5: Conversion from ATP.cdx and Brevetoxin A.mol

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

やはり `-shell-escape` オプションを付けてタイプセットします。

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

```
$ dvipdfmx test.dvi
```

すると今度は「smilesobabeling[NUM].pdf」のようなファイルが生成し、出力される「test.pdf」は図 6 のようになります。

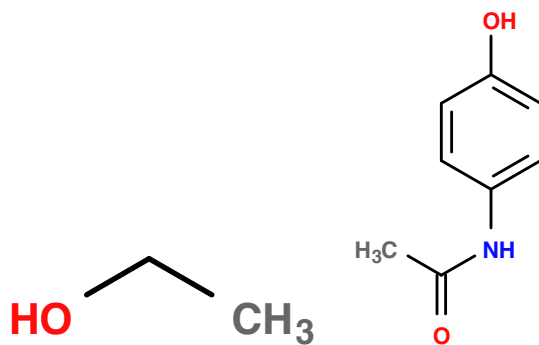


Fig. 6: Conversion from SMILES notation

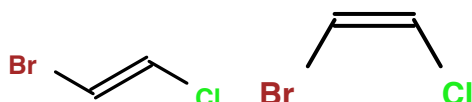
構文は以下のとおりです。

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

`\chemobabel` との唯一の違いは、中括弧の 1 つめに SMILES 表記法のテキストを与えることです。`\smilesobabel` の引数としては、任意の SMILES 表記を用いることができます（バックslash (\) やパーセント記号 (%) のような特別な文字が入っていてもかまいません）。なお、`\chemobabel` の引数についても同様に、特別な文字を含むことができる設計になっています。

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

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



旧バージョンのユーザに対する注意 (≤ chemobabel v0.6 [2015/06/29])

古いバージョンの `chemobabel` では、これらの \LaTeX で特別な意味を持つ文字に注意を払う必要がありました。従来私が紹介してきた回避策は“カジュアルな \LaTeX ユーザ”にカテゴリコードを手動で変更させていたため、好ましいものではありませんでした。

この問題を解決するため、パッケージ本体の側でこれらの文字をサポートするマクロを組み込みました (`chemobabel` v0.7 [2015/08/26] 以降)。これにより、`\` や `%` といった特別な文字を心配することなく、任意の SMILES 表記を直接引数に渡すことができるようになっています。従来の回避策 (`\begin{group} ... \end{group}` 内で `\catcode` を変更する方法) は今ではむしろ有害になっていますので、注意してください。

3.3 パッケージオプション

chemobabel パッケージは、デフォルトでは化学構造式を Inkscape によって PDF 形式の画像ファイルに変換して文書中に取り込みます。しかし、オプションで画像形式と変換プログラムを変更することができます。

dvips などの PDF 非対応のドライバを使用する場合は、EPS 形式の画像ファイルに変換したほうが便利でしょう。この場合

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

というように eps オプションを指定すれば、内部で Inkscape から出力される画像形式を PDF でなく EPS に変更します (chemobabel v0.9d [2016/02/28] 以降)。

また、画像変換プログラムを Inkscape から rsvg-convert に変更することもできます (chemobabel v0.9e [2016/03/07] 以降)。rsvg-convert は Inkscape より高速ですから、便利かもしれません。この場合は、以下のように librsvg オプションを指定します。

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

デフォルトは pdf オプションと inkscape オプションを指定した場合と等価です。

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

さらに、Open Babel v2.4.0 以降で生成される画像はデフォルトで余白が大きいため、pdfcrop (PDF の場合) あるいは ps2eps (EPS の場合) でクロップしています (chemobabel v0.9i [2022/09/12] 以降)。これをやめたい場合は nocrop オプションを指定してください。

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

3.4 安全なタイプセットのために

第3節で述べた基本的な使い方で、ほとんどの場合問題ないでしょう。しかし、`-shell-escape` オプションを付けてタイプセットするため、好むと好まざるとにかかわらず \LaTeX が**どんな外部コマンドでも**実行できてしまいます。つまり、知らないうちにプログラムが実行されてしまう可能性を示唆しており、これはとりわけ自分で書いた \LaTeX ソースでない場合には危険を伴います。この問題を回避するために、`\chemobabel` と `\smilesobabel` コマンドだけを別の \LaTeX ファイルに抽出するマクロを用意しています（参考：[TeX Forum \[4\]](#)）。

この場合は、パッケージを読み込む際に `extract` オプションを付与してください。

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

タイプセットの際には通常の

```
$ platex test.tex
```

というコマンドを使います。`extract` オプションを付けると、`chemobabel` パッケージは単に元の \LaTeX ソースから `\chemobabel` と `\smilesobabel` コマンドだけを抽出するはたらきをします。その結果、同じディレクトリに「ChemFigFile.tex」というファイルが生成します。これは元のソースに書かれていた `\chemobabel` と `\smilesobabel` コマンドをすべて含む最小の \LaTeX ソースで、これは直接 `pdflatex` でタイプセットできます。

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

欧文と同じ仕様に統一したため、抽出される「ChemFigFile.tex」では `graphicx` パッケージに `[dvipdfmx]` オプションを与えずに読み込んでいます。したがって、このタイプセットだけは `pdflatex` を用いてください。すると、第3節と同様に PDF 形式の図のファイルが生成しますので、あとは `extract` オプションを除去して再度

```
$ platex test.tex
```

```
$ dvipdfmx test.dvi
```

と実行すれば、適切に図を取り込んだ望みどおりの「test.pdf」が得られます。この方法を用いれば、タイプセットに要する時間を大幅に短縮することも期待できます。

4 高度な使い方

4.1 描画オプション

既に述べたとおり, `\chemobabel` と `\smilesobabel` の両方について, 2 つめの中括弧に *obabel options* を与えることができます。ここで, いくつかの有用な例を挙げておきます。

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

構造式中の全ての炭素原子を省略せずに描画するコマンドを実行します：

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

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

元素の着色を行わず, 末端の炭素原子を明示しない構造式を出力します：

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

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

Open Babel 2.x 系列で ChemDraw の図からファイル名を取り除くには, `-xd` オプションを付けます：

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

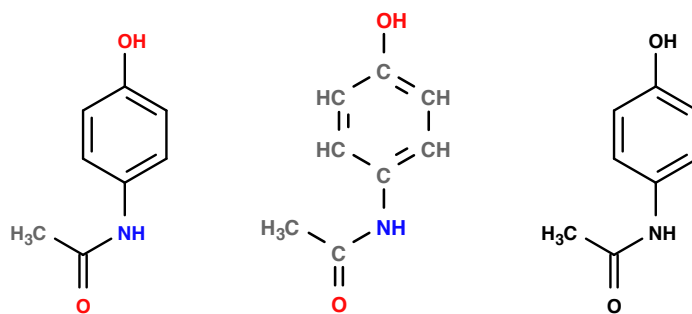


Fig. 7: Acetaminophen (Paracetamol): Nothing, `-xa`, `-xu`

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

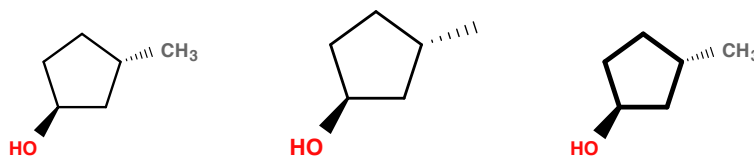


Fig. 8: (1S,3S)-3-Methylcyclopentanol: Nothing, `-xC`, `-xt`

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

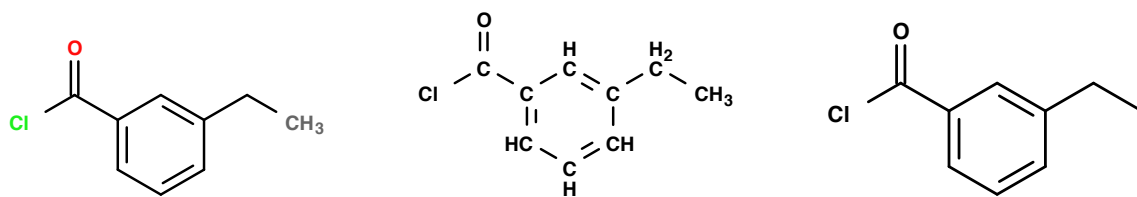


Fig. 9: 3-Ethylbenzoyl chloride: Nothing, -xa -xu, -xu -xC

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

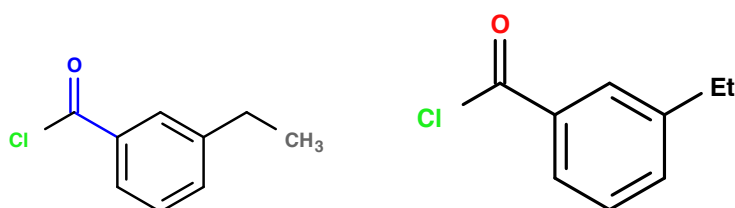


Fig. 10: 3-Ethylbenzoyl chloride: --highlight "cC=O blue", -xA --genalias

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

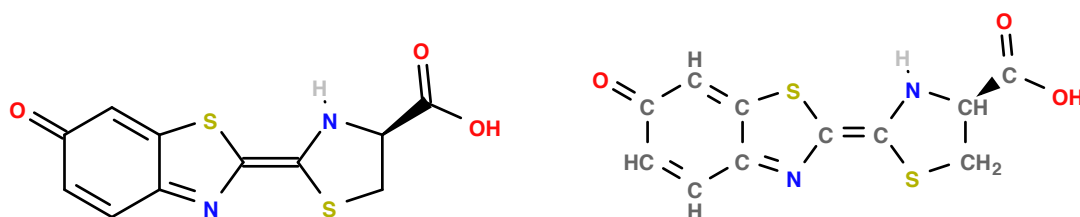


Fig. 11: Firefly luciferin (from MDL Molfiles): Nothing, -xa

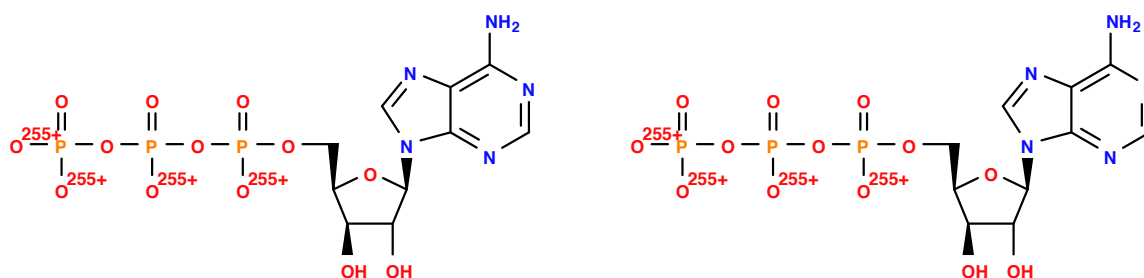


Fig. 12: ATP (from ChemDraw files): Nothing, -xd

より詳細には Open Babel documentation の obabel and babel または SVG depiction (svg) を参照してください。

4.2 より複雑な構造式

Open Babel は SMILES 表記から複雑な構造式を生成することができます。

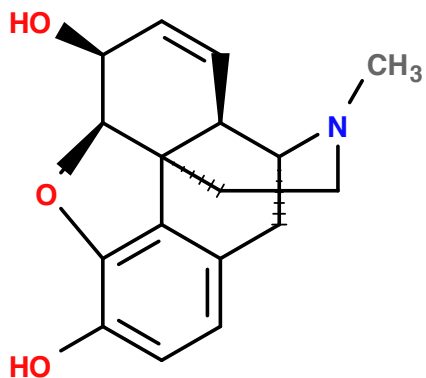


Fig. 13: (-)-Morphine

SMILES: CN1CC[C@]23c4c5ccc(c40[C@H]2[C@H](C=C[C@H]3[C@H]1C5)O)O

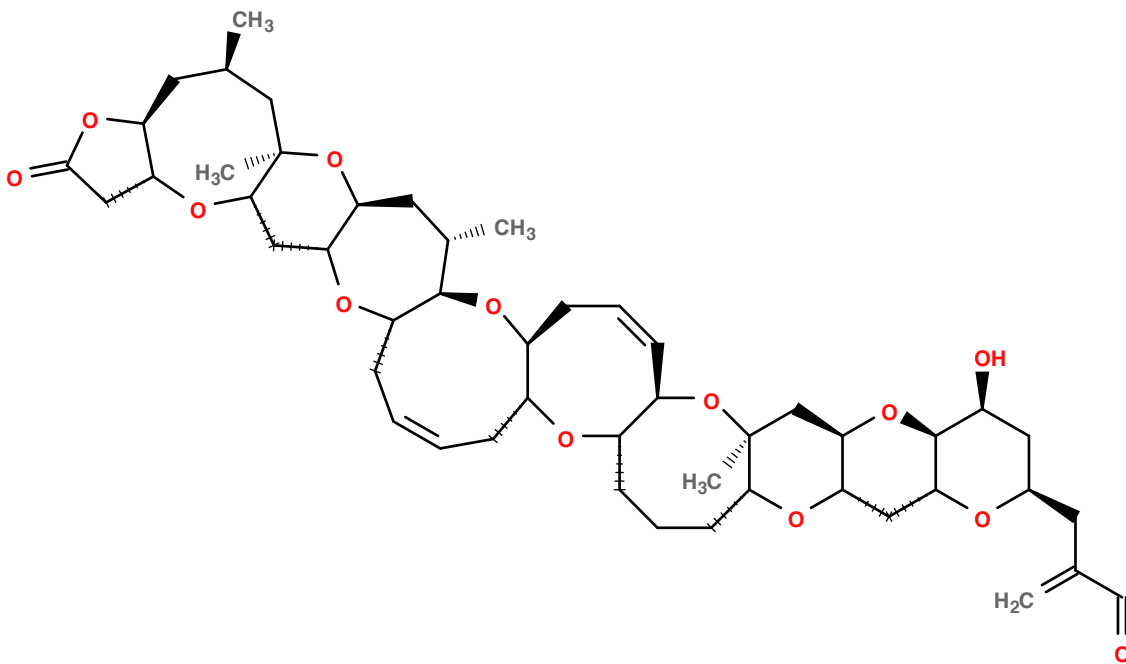


Fig. 14: Brevetoxin A

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]5C)C(=O)O)C[C@H](O)C[C@H]3O

しかし、時には結果が望ましくないことがあり、また使用するバージョン番号によって出力が異なることがあります。いくつかの例を示します：

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@]

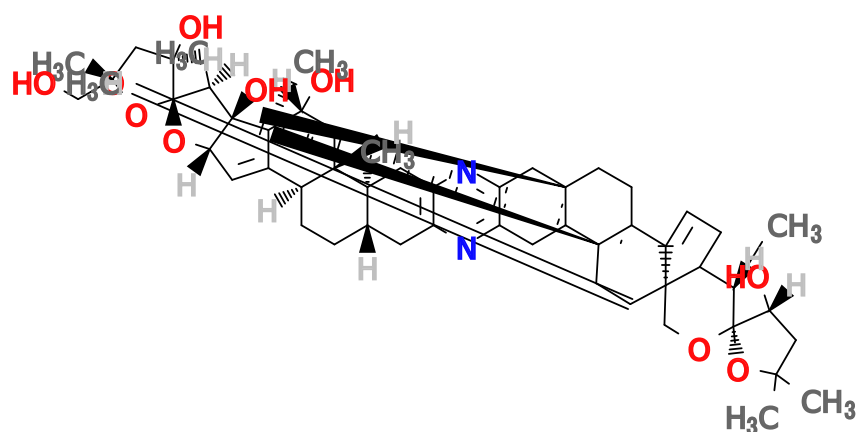


Fig. 15: Cephalostatin-1 (figure from Open Babel for Mac 2.3.1)

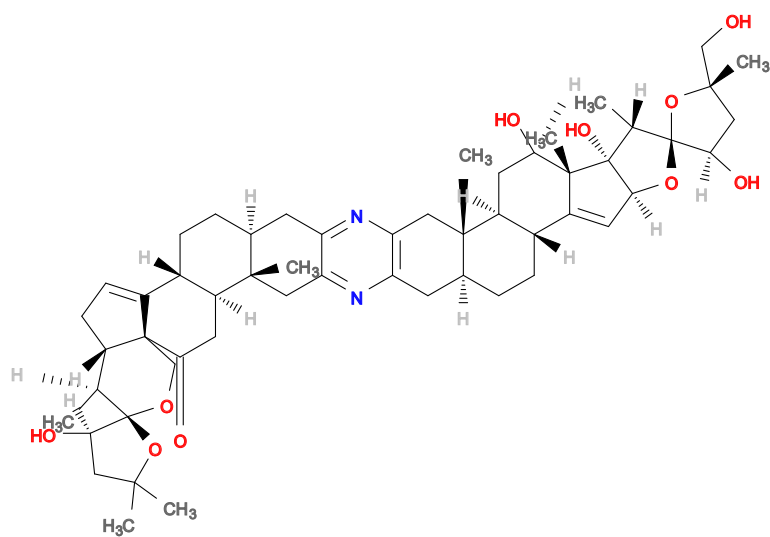


Fig. 16: Cephalostatin-1 (figure from Open Babel for Win 2.3.2)

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

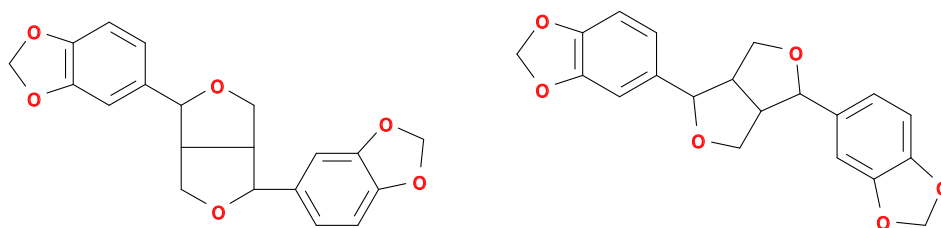


Fig. 17: Sesamin (figures from Open Babel for Mac 2.3.1; Normal and --gen2d)

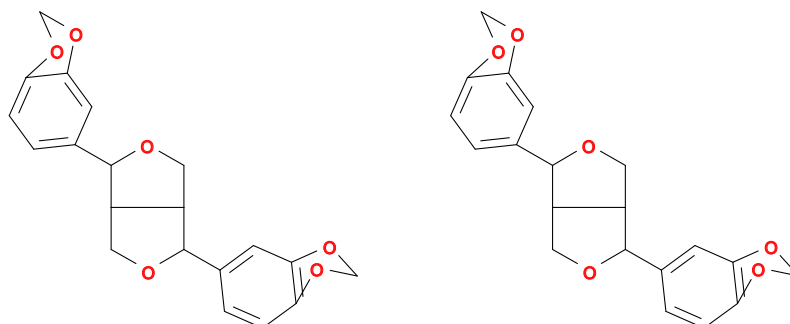


Fig. 18: Sesamin (figures from Open Babel for Win 2.3.2; Normal and --gen2d)

また、Open Babel は SMILES 表記から誤った構造式を生成することもあります。ホタルルシフェリンの図で、SMILES 表記から生成した構造式には二重結合が一つ欠けています。

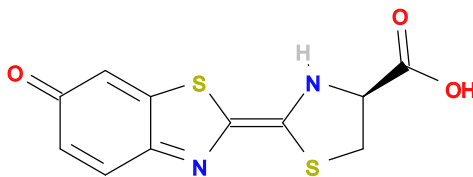


Fig. 19: Firefly luciferin: exact structure from ChemSpider ID 4588411

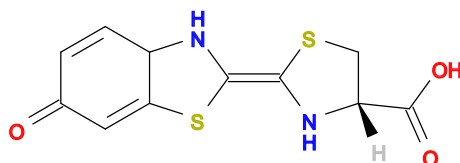


Fig. 20: Firefly luciferin?: output from Open Babel for Win 2.3.2

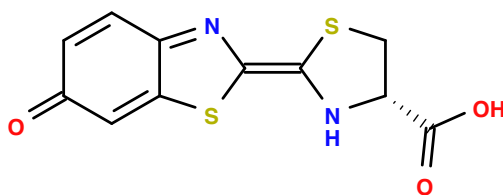


Fig. 21: Firefly luciferin?: output from current Open Babel

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

このような問題点は全て `.mol` または `.cdx` を用意して `\chemobabel` を利用することで回避できます。とはいえ、ただの文字列から複雑な構造式を生成しうるということ自体、おもしろいとは思いませんか？

5 互換性に関する注意

5.1 chemobabel パッケージの有利な点

Noel O'Boyle さんによる `\smiles` [1] あるいは Jakob Lykke Andersen さんによる `graphvizObabel.sty` が提供する `\obabel` [3] は以下のようにして使うことになっています:

```
\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}
```

しかし、`\smiles` はラスター画像 (.png) を生成してかつ `\includegraphics` に渡すオプションを扱えず、`\obabel` は拡張子 (.pdf) が欠落していたりオプション引数が空の場合にエラーが発生したりという理由で正しく処理できない可能性があります。これらの問題は `\smilesobabel` では解決されています。

6 この方法の限界と代替案

この方法は Open Babel に依存して構造式を生成しており，一切自分で構造式を描画することなく済ませることさえ可能です。もちろんある程度はそれらの構造を修正したりカスタマイズしたりできますが，この方法は出力の緻密な調節には不向きでしょう。また，コンピュータによって生成される構造式は時に見た目が不自然で，好ましくない場合もあるかもしれません。したがって，この方法は構造式の出力を完全に制御したいという意図がなく，構造式を簡便に挿入したい場合にのみ役に立つでしょう。出力に満足できない場合は， \LaTeX や chemfig のようなパッケージを代わりに使うことをご検討ください。

同様の目的を達成するために使えそうなパッケージとして，以下のようなものもあります：

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

7 技術情報

SMILES 表記法に使われる特殊文字の一覧は、SMILES Tutorial によると以下のようです。

- 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)

これらの文字が \LaTeX における特殊文字として解釈されることを防ぐ必要がありますが、バックスラッシュ \backslash はデフォルトのエスケープ文字であり、またパーセント記号 $\%$ もコメント文字としての特殊な意味を持っています。chemobabel v0.7 以上では、カテゴリーコードを一時的に変更することでこれらすべての文字を適切に扱うようになっています ($\backslash\text{verb}$ と同様の手法)。

To do (if possible):

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

8 更新履歷

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)