

METAFONT/METAPOSTを利用した化学構造式処理システムの開発

山路 章 (E-mail) akira.yamaji-molecoding@main.nifty.jp
 公開サイト (MoLeCoding研究室) <http://molecoding.cocolog-nifty.com/blog/>

はじめに

Molecular Coding Format(以降MCFと省略)は、線形表現で化学構造式を記述する新たな表記法である。現状、直接META言語で記述しておりコーディング(プログラミング)そのものである。ただし、ここで言う「コーディング」は、原子、結合番号の相対アドレス指定、変数のグルーピング化、マクロ置換等のプログラミング手法から来ている。MCF自体はMETA言語が表に出ない文法である。なお、分子の化学構造をMCFで表現することをMoLeCodingと略称している。

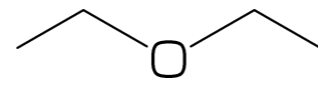
開発の経緯

構造式を高密度に収録する表の作成を目的として、「線形表現」を変換して「構造式フォント」を作るMETAFONTマクロを開発した。その後、同一ソースファイルをMETAPOSTでも処理できるよう改良し、MDL Molfileを出力する機能も追加した。開発したマクロ、MCF文法解説、TeX取込み例等は、上記サイトにおいて公開している。

MCF文法の概要

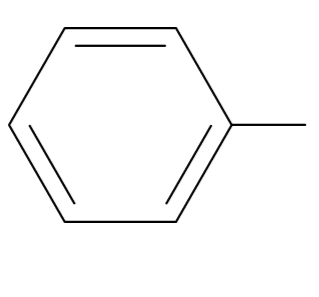
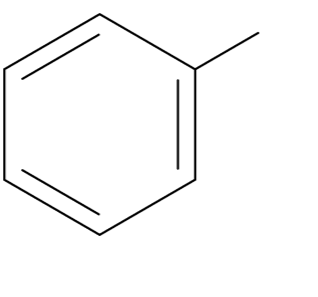
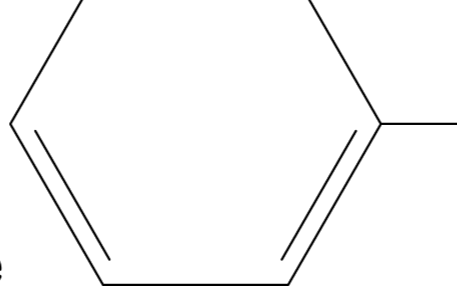
文法の詳細は配布ソフトウェアに含まれるMCF_MANUAL1_jp.pdfに記載している。代表的な化学構造式の線形表記法であるSMILES(Simplified Molecular Input Line Entry System)と比較するとその特徴を理解しやすい。以下にその主な特徴を挙げる。

(1)SMILESは原子(ノード)、MCFは結合(エッジ)優先

SMILES:CCOCC (指定のない結合は単結合) MCF: !2,0, !2 (指定のない原子はC) 

(2)MCFの結合(エッジ)は角度・長さ(ベクトル)※を含有

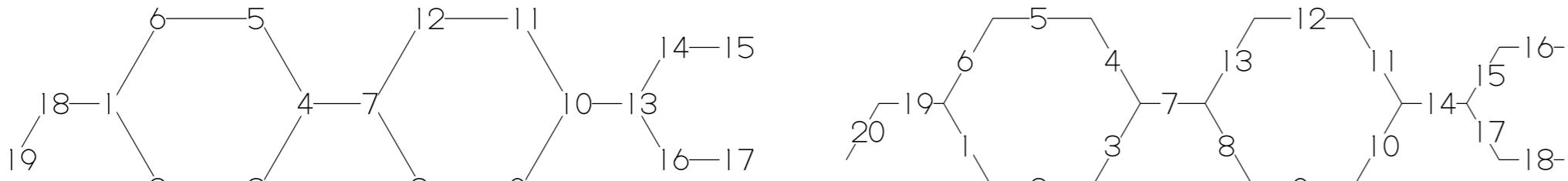
・SMILESは下の3つの構造式を区別できないが、MCFは区別できる。

SMILES: c1cccc(cc1)C  SMILES: 同左  SMILES: 同左 
 MCF: Ph,4>/Me MCF: ^30,Ph,4>/Me MCF: ^2,Ph,4>/Me

※紙の上に描画するための角度・長さであって、実際の分子構造を示すものではない。

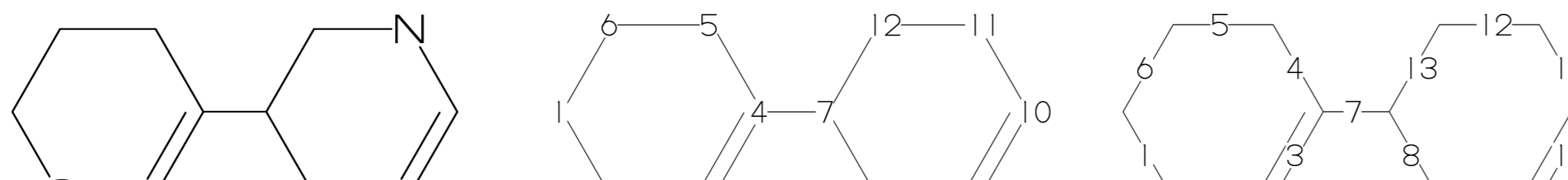
(3)MCFの原子、結合番号

・原子、結合は作成された順番に番号(アドレス)が割振られる。

?6,-3\,?6,-3\,!2,-2>/Et,1>/Et 

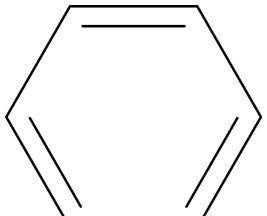
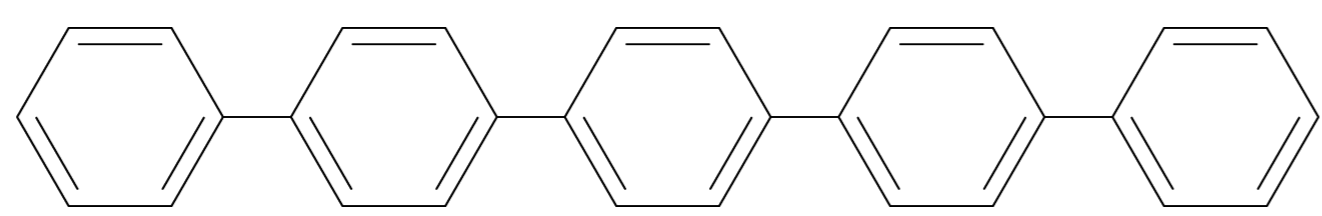
(4)原子、結合番号のグルーピング化、相対アドレス指定

・{ } 内の番号は1から振り直し、マイナス値はカレント番号から遡る番号数

?6,4\,?6,\$(3,-4)d1,-2>N,2>S 

(5)部分構造のマクロ置換

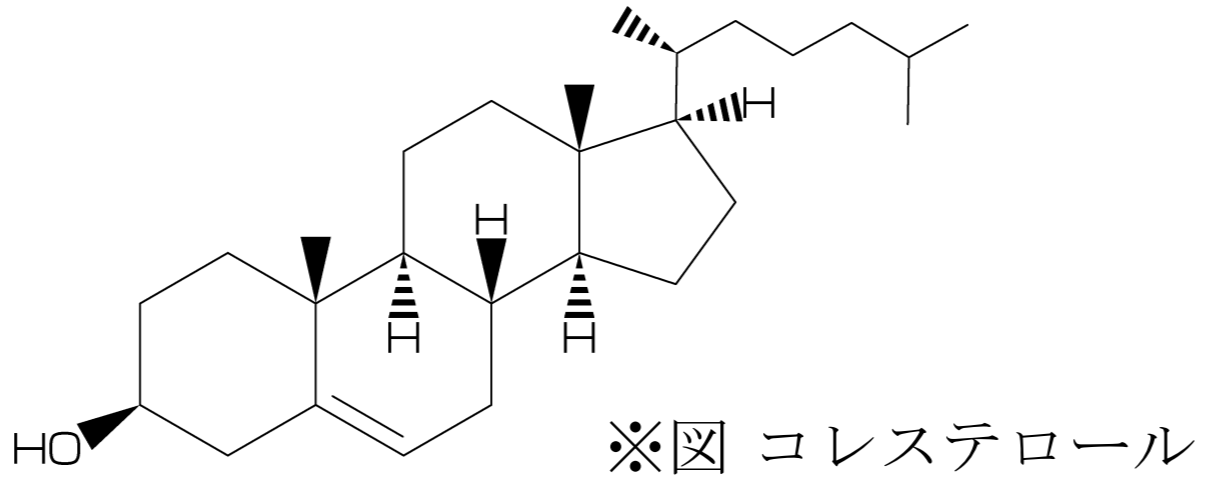
・ベンゼン環のように多用される部分構造を登録し、マクロ置換することができる。

Ph:= (?6,\$(-2,-4,-6)d1);  Ph,-3\,Ph,-3\,Ph,-3\,Ph,-3\,Ph 

(6)SMILESは文字列、MCFはCSV(Comma-Separated Values)

・MCFは改行、字下げ、コメントの挿入が可能

SMILES: [H]OC/4CC\3=C\CC1([H])(C[H])(CCO2(C)(C[H])(CCC12([H]))C(C)CCCC(C)C)C/3(C)CC4

MCF: ^30, % 回転角
 ?6,\$(-4,-2)?6,-4|?5,-14|d1, % 基本骨格
 @(10^180,^^-60,11,17)/*H, % 置換基
 9^60>*/H,1>*/OH, % //
 @(^60,4,12)*/Me, % //
 -1\17,/*Me,!4,iPr % 側鎖 

MoLeCodingシステムの利用法

(1)TeXを使用しない場合

- ・MCFを変換、生成したEPS/SVGファイルをワープロ、表計算ソフト等に貼り付ける。
- ・MCFを変換、生成したMDL Molfileを構造式エディター、分子モデルソフト等に取り込む。

(2)少数の構造式をTeXに取込む場合

- ・フォントは\char font番号、EPSファイルは\includegraphics{file名}を使って取り込む。
- ・MePoTeX(みなも氏作)、luamplib(Luatex)を利用して、TeX文書中に直接MCFを記入する。

(3)多数の構造式をTeXに取込む場合

- ・MCFをMETAFONT又はMETAPOSTで処理、同時に生成したLOGファイルをTeX文書中に挿入し、フォント又はEPSファイルを一斉に取り込む。

※このポスター中の構造式は、MCFをMETAFONTで処理して生成したpkフォントを\char (font番号)で取り込んでいる。

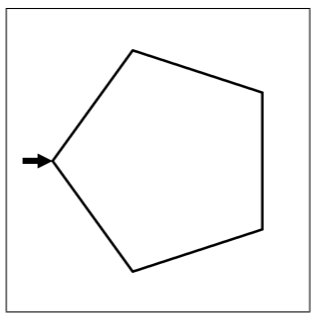
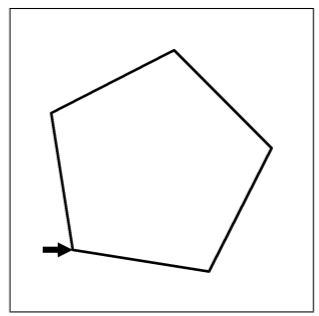
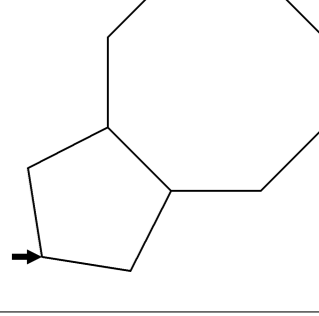
METAFONT、METAPOSTの使い分け

MoLeCodingは、同一ソースファイルをMETAFONT、METAPOSTどちらでも処理できる。その主な違いは以下の通り。

- | | |
|-------------------|---------------------------------|
| [METAFONT] | |
| (長所) | ・多数の構造式を扱う際、処理が早い(軽い) |
| | ・複数(最大256個)の構造式が1つのフォントファイルに収まる |
| (短所) | ・大きさに制限がある(最大4095ピクセル) |
| | ・カラーが使えない |
| | ・TeXを使う必要があり、単独では利用しにくい |
| [METAPOST] | |
| (長所) | ・大きさに制限がない |
| | ・カラーが使える |
| | ・TeXを使う必要がなく、単独で利用しやすい |
| (短所) | ・多数の構造式を扱う際、処理が遅い(重い) |
| | ・生成ファイル数が多くなりやすい |

MCFの入力例

マリントキシンの一種であるブレトキシシンAを例にしてMCFの作成工程を以下に示す。ブレトキシシンAは分子量867.07、分子式C₄₉H₇₀O₁₃と比較的規模の大きい分子である。MCFの文法上は特に制限はないが、可読性を考慮して構造を記述する順番は、基本骨格→基本骨格の結合種→基本骨格の原子→基本骨格の置換基(単原子)→基本骨格の置換基(複数原子)→別骨格(側鎖等)を基本とする。また、原点(左下)に近い位置から記述する。

【5員環】  MCF: ?5
 【角度調整】  MCF: ^^45,?5
 【縮合環】  MCF: ^^45,?5,-3|8

【基本骨格】
 MCF: ^^45,?5,-3|?8,-5|?6,-3|?7,-4|?9,-3|?8,-5|?8,\$(-3,-3,-3)?6

※\$(-3,-3,-3)?6: -3|?6,-3|?6,-3|?6

【結合番号表示】
 sw_numberingB:=1;
 MCF: ^^45,
 ?5,-3|?8,-5|?6,-3|?7,-4|?9,-3|?8,-5|?8,
 \$(-3,-3,-3)?6

【結合種変更】
 MCF: ^^45, 中略
 \$(26,36)d1

※\$(26,36)d1: 結合No.26,36を二重結合に変換

【原子番号表示】
 sw_numberingA:=1;
 MCF: ^^45,
 ?5,-3|?8,-5|?6,-3|?7,-4|?9,-3|?8,-5|?8,
 \$(-3,-3,-3)?6

【原子変更】
 MCF: ^^45, 中略
 @(5,6,15,16,27,28,39,40,47,48)0

※@(5,6...)0: 原子No.5,6..を0(酸素)に変換

【置換基導入】
 MCF: ^^45, 中略
 @(3,7,13,17,25,29,37,41,45)*H,
 @(4,14,18,26,30,42,46)*H,1>/O,
 @(8,19,38)*CH3,10>*/CH3,-1>/OH

※ / : 単結合, // : 二重結合, */ : 楔形,
 /* : 破線楔形

【置換基の角度、長さ調整】
 MCF: ^^45, 中略
 @(^-60,3,7,13,17,25,29,37,41,45)*H,
 @(^60,4,14,18,26,30,42,46)*H,1>/O,
 @('1.3,8^68,19,38^60)*CH3,10>*/CH3,
 -1>/OH

※^数値,^数値: 結合角度変更
 '数値,'数値: 結合長さ変更

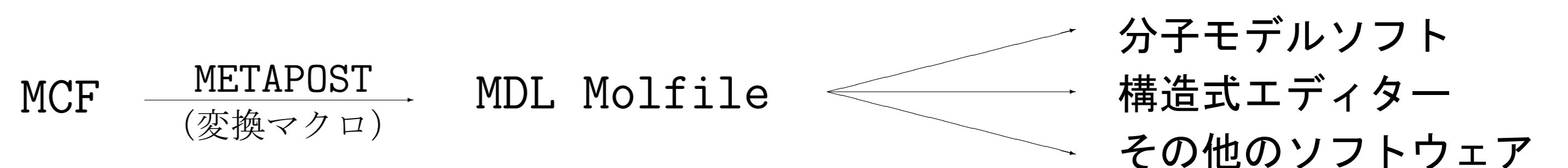
【側鎖導入】
 MCF: ^^45, 中略
 -3^wf,! ,/Me,! ,CHO

※-3\ : 原子番号 -3に移動
 ^wf : 結合を楔形に変換

【構造式完成】
 MCF: ^^45,
 ?5,-3|?8,-5|?6,-3|?7,-4|?9,-3|?8,-5|?8,
 \$(-3,-3,-3)?6,\$(26,36)d1,
 @(5,6,15,16,27,28,39,40,47,48)0,
 @(^-60,3,7,13,17,25,29,37,41,45)*H,
 @(^60,4,14,18,26,30,42,46)*H,1>/O,
 @('1.3,19,8^68,38^60)*CH3,10>*/CH3,
 -1>*/OH,-3^wf,! ,/Me,! ,CHO

MDL Molfileの出力機能

MCFを変換してMDL Molfile(代表的な構造式の結合表保存形式)を出力することができる。ほとんどの構造式エディター、分子モデルソフト等は、MDL Molfileを取り込み可能である。



まとめ、今後の課題

- (1)まとめ
- ・構造式(紙の上に描画するための情報を含む)を簡潔に記述できる
 - ・結合、原子番号表示等の機能を利用し、大規模な分子を容易に入力できる
 - ・MDL Molfileを介して構造式エディター、分子モデルソフト等を利用できる
- (2)今後の課題
- ・マニュアル類(英語版を含む)の拡充
 - ・多くのユーザーからの意見、要望を集約、それを基にした機能の追加