

# SMILES 記法を利用した薬物設計ソフトウェアの開発

— MolSmile —

## 1. 背景

2003 年の 4 月 14 日にヒトゲノム解読完了の宣言がなされ、現在ポストゲノムと呼ばれる時代を迎えている。医薬品開発においては、ゲノム情報から病気の原因を特定し、それを治療する薬物を理論的にデザインする「ゲノム創薬」というプロセスを組むようになった。従来の創薬方法では、数万種類の化合物の活性を調べる必要があり、コストと時間が膨大に必要であった。また、研究者の経験や勘に頼るところも多く、偶然に活性薬物を発見することも少なくなかった。一方ゲノム創薬では、理論的に標的 DNA や標的タンパク質を特定することにより、効果が高く副作用の少ない薬物を低コストで効率的に開発することができる。

ゲノム創薬におけるキーポイントは、Structure-Based Drug Design (SBDD) である。SBDD とは、薬物標的である生体高分子の 3 次元立体構造をもとにして、その形に合った薬物分子をコンピュータ上で設計する方法である。SBDD に必要なものは、生体高分子と化合物の 3 次元立体構造情報であるが、構造既知のタンパク質や核酸については Protein Data Bank (<http://www.rcsb.org/pdb/>) より 3 次元立体構造情報を得ることができる。では、化合物の 3 次元立体構造はというと、Protein Data Bank に 3 次元立体構造が登録されているものもあるが、その数には限りがある。また、無償で提供されている化合物データベースは存在するのだが、それらのほとんどは化合物の 3 次元構造情報ではなく、2 次元構造情報が登録されている。それら化合物の 2 次元構造情報から 3 次元構造に変換するソフトウェアが有償で提供されているが、その多くは数百万円もするソフトウェアであり、容易に使えるものではない。

そこで、私は、化合物の 2 次元構造情報から 3 次元構造に変換し、化合物の構造最適化を行うことのできるソフトウェアの開発を思い立つに至った。さらには、SBDD を容易に行えるようにするため、薬物標的タンパク質と設計化合物の親和性を高速かつ簡単に評価できるソフトウェアの開発も行う。

## 2. 目的

分子の化学構造を表記する方法に、SMILES 記法 (Simplified Molecular Input Line Entry Specification syntax) という方法がある。SMILES 記法とは、グラフ理論に基づいて考案された表記方法で、化合物中の原子をノード、原子間の結合をエッジと見立て、化学構造を 1 行の文字列で表す方法である。SMILES による化合物表記例を図 1 に示す。SMILES は文字列であるために、化合物のデータベースを作成するのに大変扱いやすく、広く用いられている。しかしながら、SMILES から得られる情報は、化合物の原子種とそれらの結合情報だけで、原子の座標情報を得ることが出来ない。また、水素原子の結合情報が欠失していることが、SMILES の欠点である。そこで、本プロジェクトでは SMILES から化合物の 3 次元立体構造を構築するプログラムの作成を第 1 の目的とする。

さらに、構築した化合物の 3 次元立体構造を用いて、薬物標的タンパク質との結合様式

や結合親和性を評価するプログラムの作成を第2の目的とする。

|        |       |         |              |                       |
|--------|-------|---------|--------------|-----------------------|
| 構造式    |       | $O=C=O$ | $HC\equiv N$ |                       |
| 分子名    | エタノール | 二酸化炭素   | シアン化水素       | アセチルサリチル酸             |
| SMILES | CCO   | O=C=O   | C#N          | CC(=O)Oc1ccccc1C(=O)O |

図 1. SMILES による化合物表記例

### 3. 開発の内容

#### SMILES からの原子の種類と結合情報の抽出

OpenBabel (<http://openbabel.sourceforge.net/>) のライブラリを用いて、SMILES 文字列から化合物を構成する原子の情報と、それらの原子の結合情報を抽出した。さらに、これらの情報から角度を構成する3つの原子のペア、二面角を構成する4つの原子のペア、平面性を維持するのに関係する4つの原子のペア、キラル中心に結合している4つの原子のペアを抽出した。D-alanine での例を図 2 に示す。

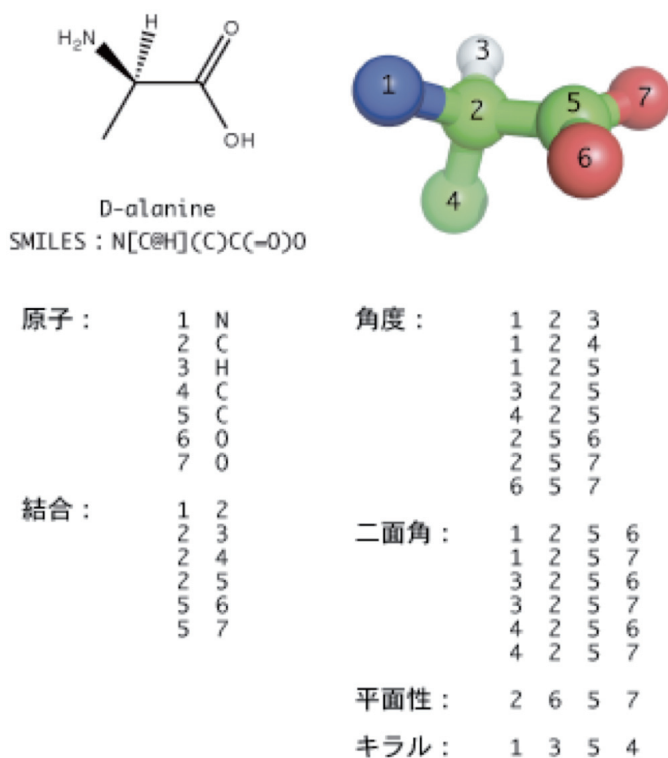


図 2. SMILES からの原子の種類と結合情報の抽出.

## 原子の種類と結合情報からの3次元構造構築

SMILES 文字列から得られた原子種と結合情報から分子の3次元立体構造を構築するために、距離幾何学法を用いた。距離幾何学法により簡単に3次元立体構造を作成した後に、分子力学法によって3次元構造の最適化を行った。分子力学法とは、原子を球に、結合をバネに見立て、古典力学に基づいて化合物のポテンシャルエネルギーを計算する方法である。SMILES 文字列から3次元構造の構築を行った例を図 3 に示す。

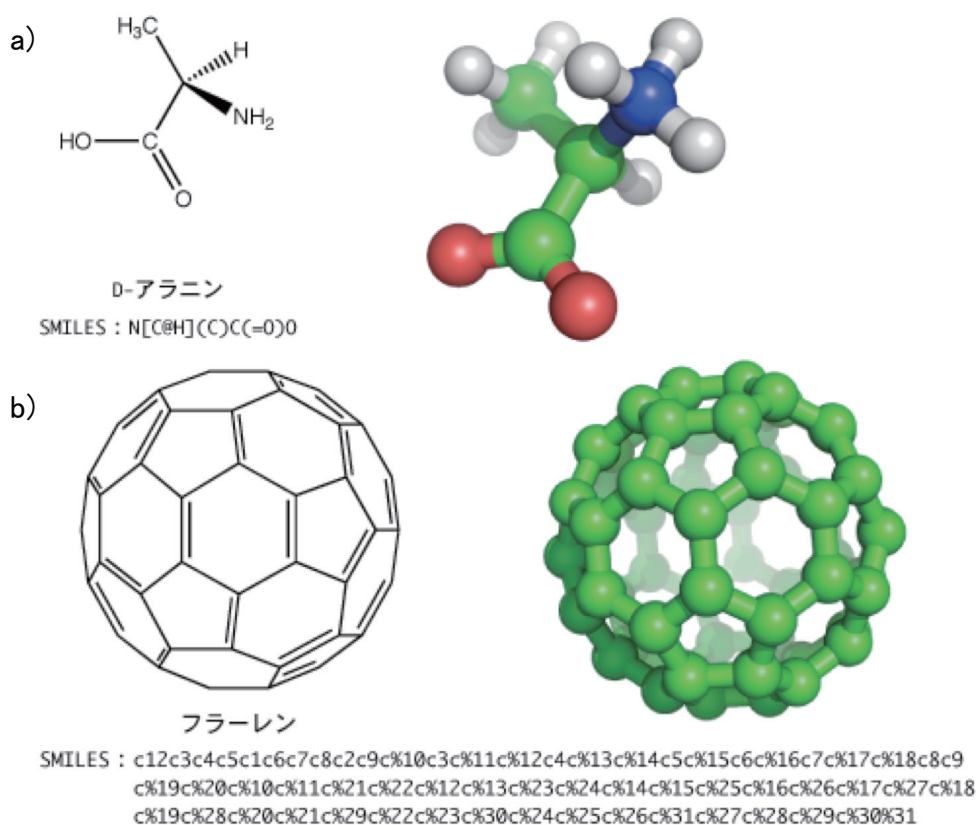


図 3. SMILES 文字列からの化合物3次元構造構築. a) D-アラニン, b) フラーレン.

## 薬物標的タンパク質との親和性評価

Protein Data Bank (<http://www.pdb.org/pdb/>) に登録されたタンパク質結晶構造をもとに、薬物標的タンパク質と薬物の結合様式、親和性を評価した。薬物を標的タンパク質ヘドッキングさせるために、タンパク質内部の疎水性領域を計算し、その疎水性領域と薬物の慣性主軸を重ね合わせた。薬物ドッキングの結果を次頁図 4 に示す。

### 4. 従来の技術(または機能)との相違

現在は、数百万円する高価なプログラムの一部の機能として、SMILES 文字列から化合物の3次元構造を構築できるものが存在する。しかし、本プロジェクトで開発したプログラムは、既存のプログラムと同等の機能を有しており、本プログラムを無償で提供することによって誰でも容易に化合物の立体構造を作成することが可能となる。

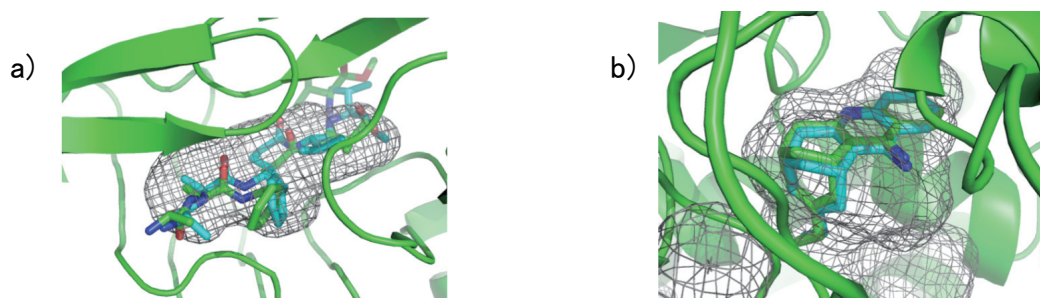


図 4. 薬物標的タンパク質への薬物ドッキングの結果. 疎水性領域を灰色のメッシュ表示, X線結晶構造を緑色, 予測した薬物の結合様式を水色で表示した. a) HIV-1 protease-hydroxyethylene 複合体 (PDB ID: 1AAQ), b) Acetylcholinesterase-huprine X 複合体 (PDB ID: 1E66).

## 5. 期待される効果

本プロジェクトで開発したプログラムを用いることで、既存の化合物データベースに登録されたSMILES文字列から化合物の3次元立体構造を容易に作成することが可能となった。また、データベースに登録されていない化合物であっても、図5のようにSMILESの一部を変えることで、新たな化合物の作成も可能である。

さらに、標的タンパク質と化合物の3次元構造を利用した薬物ドッキングと親和性評価の精度も良く、新規薬剤開発用のツールとして非常に有用であると考えられる。

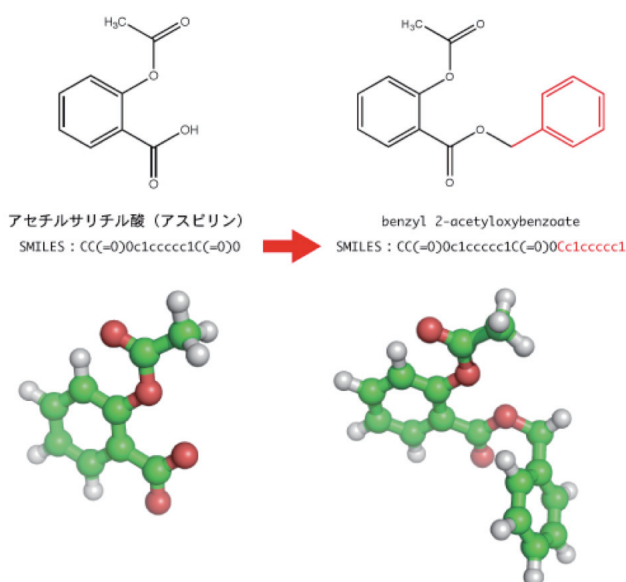


図 5. SMILES を利用した薬物設計.

## 6. 普及(または活用)の見通し

本ソフトウェア「MolSmile」は、創薬研究を行っている研究者を対象に作成した。コンピュータの操作に不慣れな研究者でも容易に化合物の3次元構造構築、化合物の医薬品としての有効性の評価を行えるように開発を行った。今後、さらなる機能の充実と改良を行いながら、Web上にて無償で公開することを目標としている。

(公開予定先: <http://bukka.p.chiba-u.jp/molsmile/>)

## 7. 開発者名(所属)

開発代表者: 藤 秀義 (千葉大学大学院医学薬学府)

共同開発者: 辰巳 絢子 (千葉大学大学院医学薬学府)