

PubChem のデータを用いた分子座標の作成

-データベースを用いて Gaussian16 の入力ファイルを作成する方法-

The generation of the molecular coordinates using PubChem data

-How to generate the input files of Gaussian16 from the data base -

寺前 裕之^{*1}

TERAMAE, Hiroyuki^{*1}

概要：NIH のデータベースである PubChem に収録されているデータを使用して、汎用分子軌道プログラムである Gaussian 16 の入力ファイルを作成する方法について述べている。座標データである.sdf ファイルを PubChem からダウンロードして、Openbabel を使用して入力ファイルに変換を行う。Openbabel を使用することで他のプログラムの入力を作成することも容易となる。

1. はじめに

近年、分子軌道計算は盛んに行われており様々な化学の学術論文において掲載されている。その中でおそらく最も使用されているのが Gaussian16 (略称 G16)およびそのインターフェースである Gaussview6 というプログラム¹である。他にも有名なものに GAMESS というプログラム²が存在するが、量子化学の初心者には使うのが難しい。本学情報科学研究センターにおいても G16/Gausview はサイトライセンスを使ってインストールされており本学の教職員・学生であれば使用することが可能となっている。

G16 プログラムを用いてある分子の分子軌道計算を行うためには、G16 プログラムへの入力データを作成しなければならない。ほとんどは内蔵されているデフォルト値を使用することで入力をする必要は無いが、分子の三次元構造については必ず入力しなければならない。分子式から Gausview を使用して三次元構造を作成していくことも可能だが、サイズが大きな分子についてはエキスパートにとっても大変な作業となってしまう。

そこでデータベースに含まれる実験値を使用して三次元構造を作成することを考えて、今回試行した結果を本研究ノートにまとめた。G16 を使ってみたいと思われる方の参考になれば幸いである。

2. 計算方法

^{*1} 城西大学理学部化学科・城西大学大学院理学研究科物質科学専攻

データベースはアメリカ国立衛生研究所(NIH)が公開しているオープンアクセスデータベースである PubChem³⁾を利用することにする。PubChem は、化学分子データベースの一つ。このシステムは、アメリカ国立衛生研究所(NIH)の下で国立医学図書館(NLM)の一部門である国立生物工学情報センター(NCBI)によって維持管理されている。ウェブユーザインタフェースを通して自由に接続する事ができ、数百万の化合物構造および記述のデータセットを FTP 経由でダウンロードすることが可能である。PubChem に集積されているのは 1000 原子および 1000 結合より少ない小さな分子である。

作業の流れとしてはデータベースにアクセスして、3 次元構造をあらわす sdf ファイルをダウンロードするが、そのままでは使えないので G16 の入力に合うようにデータを書き直す。入手とインストールが容易なオープンソースであることから Open Babel^{4,5)}を使用することとした。

3. 実際の計算手順

PubChem を利用するには次の URL にアクセスする。URL <https://pubchem.ncbi.nlm.nih.gov> この URL を入力することで、図 1 に示す検索窓のある画面になる。ここに化合物名などを入力することでデータベース中を検索することができる。

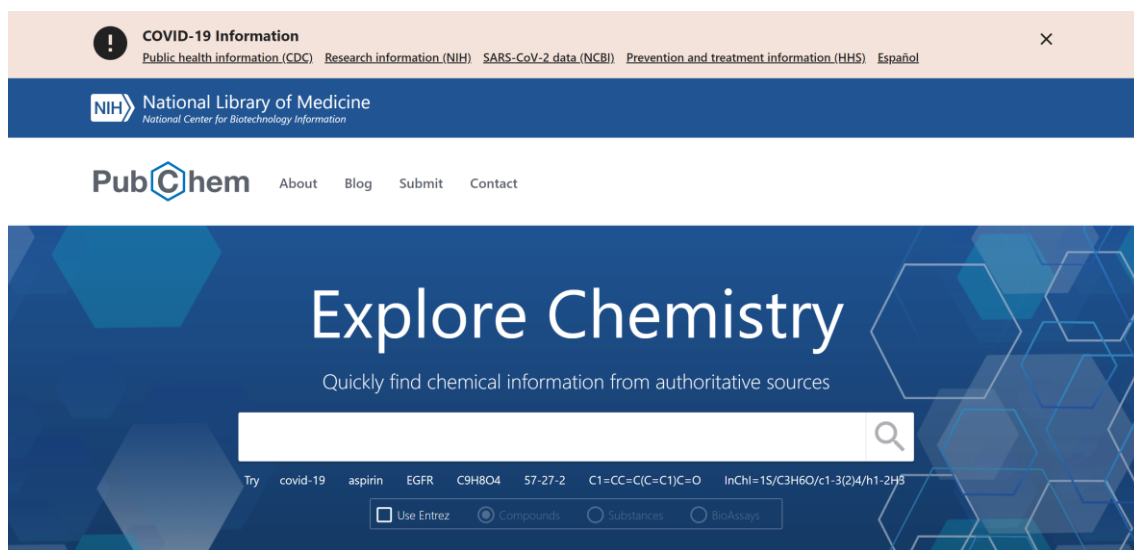


図 1 PubChem のホーム画面

例えばここで **testosterone** と入力することで **Testosterone** のデータを見ることが出来る。途中まで入力していくと図 2 に示すように検索窓の下に候補が表示されてとても便利である。

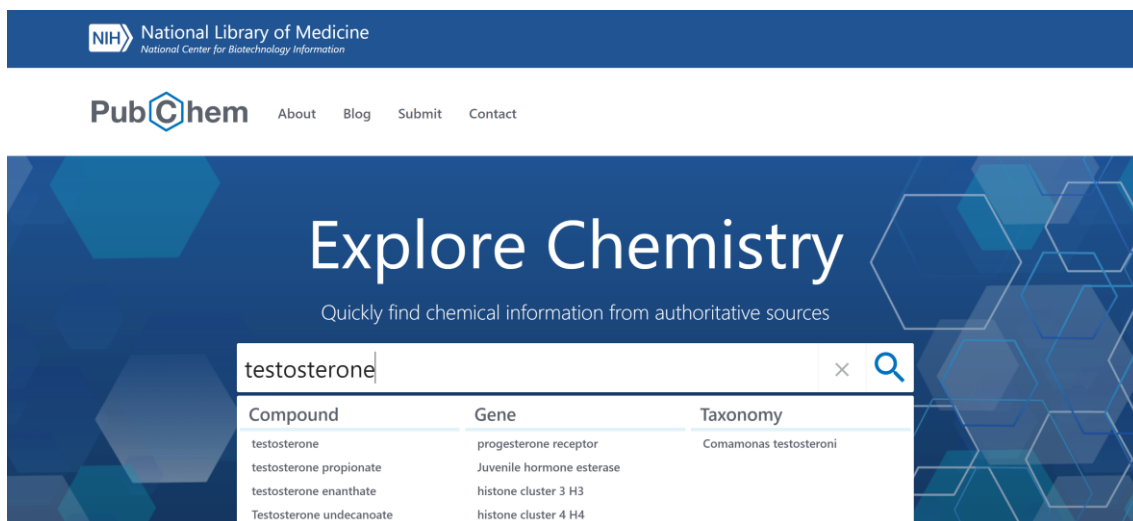


図 2 検索窓の下に現れる化合物名の候補

Enter キーを押すか検索の虫眼鏡アイコンを押せば検索が行われて図 3 に示したように結果が表示される。



図 3 検索結果の表示

実際にはもっと多くの候補が表示されるが、今は一番先頭にある Testosterone が求めるものである。この分子模型の画像をクリックすることで、図 4 に示した Testosterone のもっと詳しい画面が表示される。

Testosterone

PubChem CID 6013

Structure

2D 3D

Find Similar Structures

Chemical Safety

Irritant Health Hazard Environmental Hazard

Laboratory Chemical Safety Summary (LCSS) Datasheet

Molecular Formula $C_{19}H_{28}O_2$

testosterone
58-22-0
Testosteron

Cite Download

CONTENTS

Title and Summary

1 Structures

2 Names and Identifiers

3 Chemical and Physical Properties

4 Spectral Information

5 Related Records

6 Chemical Vendors

7 Drug and Medication Information

8 Food Additives and Ingredients

9 Pharmacology and Biochemistry

10 Use and Manufacturing

11 Identification

12 Safety and Hazards

13 Toxicity

図 4 テストステロンの詳細な結果画面

1.2 3D Conformer



Find Similar 3D Structures Get Image Download

Interactive Chemical Structure Model

☒ Ball and Stick
☐ Sticks
☐ Wire-Frame
☐ Space-Filling
☒ Show Hydrogens
☐ Animate

PubChem

Cite Download

CONTENTS

Title and Summary

1 Structures

2 Names and Identifiers

3 Chemical and Physical Properties

4 Spectral Information

5 Related Records

6 Chemical Vendors

7 Drug and Medication Information

8 Food Additives and Ingredients

9 Pharmacology and Biochemistry

10 Use and Manufacturing

11 Identification

12 Safety and Hazards

13 Toxicity

2 Names and Identifiers



2.1 Computed Descriptors



図 5 分子構造 3 次元表示の詳細画面

1.2 3D Conformer



Find Similar 3D Structure Download

Interactive Chemical Structure Model

☒ Ball and Stick
☐ Sticks
☐ Wire-Frame
☐ Space-Filling
☒ Show Hydrogens
☐ Animate

PubChem

Download

SDF Save Display
 JSON Save Display
 XML Save Display
 ASNT Save Display

Cite Download

CONTENTS

Title and Summary

1 Structures

2 Names and Identifiers

3 Chemical and Physical Properties

4 Spectral Information

5 Related Records

6 Chemical Vendors

7 Drug and Medication Information

8 Food Additives and Ingredients

図 6 sdf ファイルのダウンロード画面

ここで 3 次元構造のデータが必要なので **Structures** の場所にある分子模型のうち 3D の方をクリックする。すると図 5 に示した画面になる。ここで模型の右上にある **Download** をクリックすると

図 6 のような画面になる。

SDF の行の Save を選ぶことで名前を付けて保存のダイアログが表示されるのでファイル名入力すれば SDF 形式の構造データをダウンロードすることが出来る。なお Display を選ぶと画面上で SDF データを見ることができる。

ダウンロードした SDF データは Open Babel を利用することで、Gaussian16 の入力に作り変えることができる。

次に Open Babel をインストールしていく。Open Babel のサイトに行き、Windows の 64bit 版をダウンロードしてインストールする。URL は <http://openbabel.org/> である。

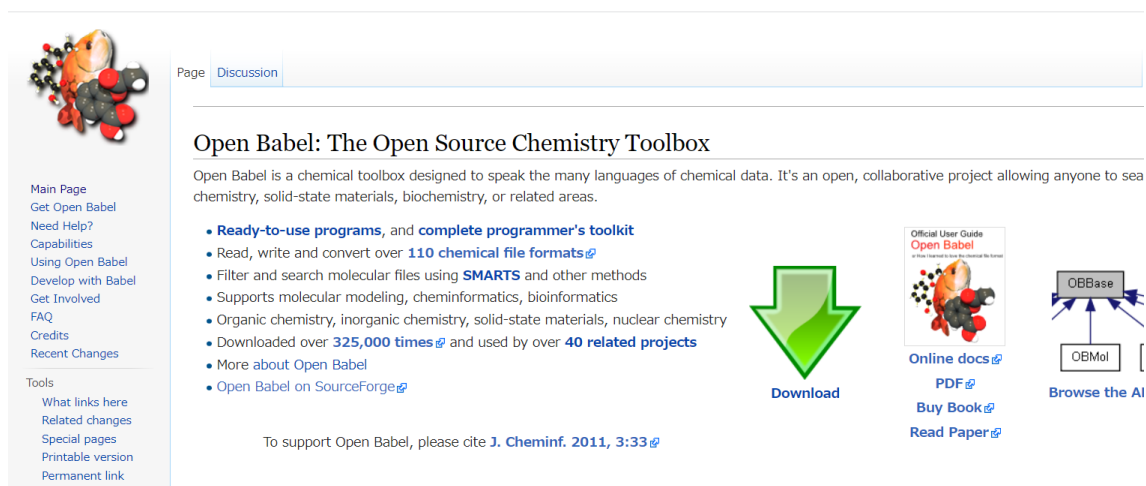


図 7 Open Babel のオープニング画面

この画面から Download を選ぶと図 8 のような画面になるので、Windows から Get the “latest installer” for 64-bit をクリックして選ぶ。

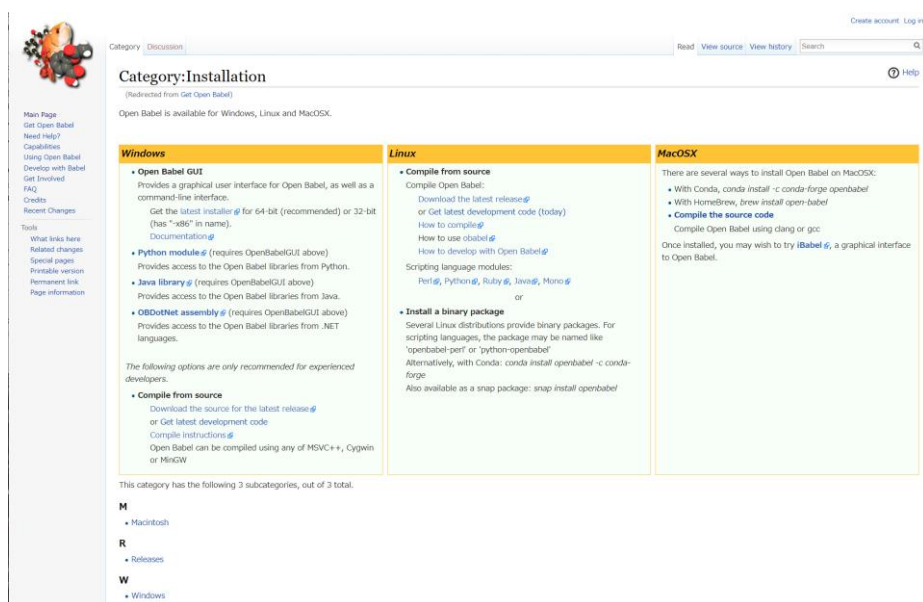


図 8 Open Babel のダウンロード画面

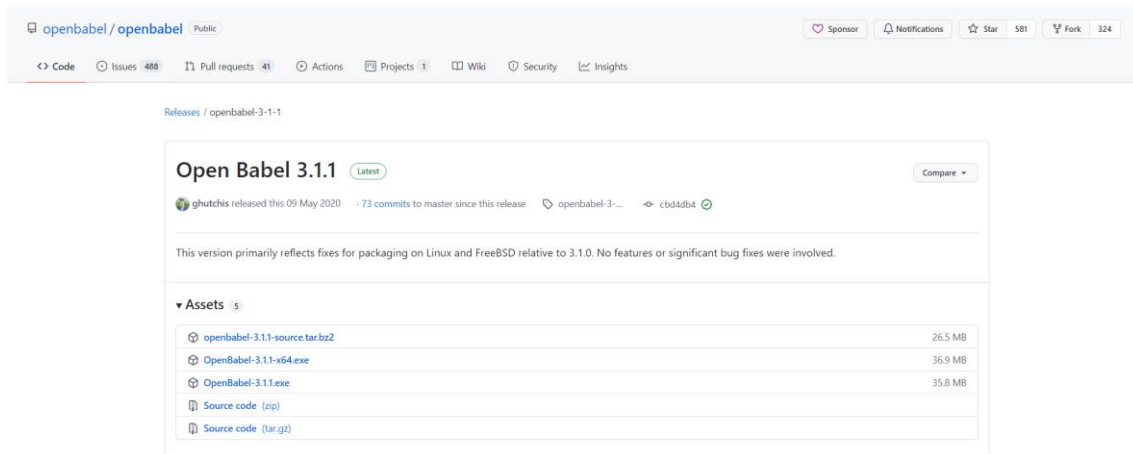


図 9 Open Babel のダウンロードの詳細画面

図 9 の画面の Assets の 2 番目にある、OpenBabel-3.1.1-x64.exe（あるいはその時点で最新のバージョン）をクリックしてダウンロードする。図 10 のようにダウンロードするファイル名の選択画面になるので、同じ名前のファイルがダウンロード先のフォルダーになればそのまま保存をクリック。ファイル名を変更したい場合はファイル名を変更して保存をクリックする。

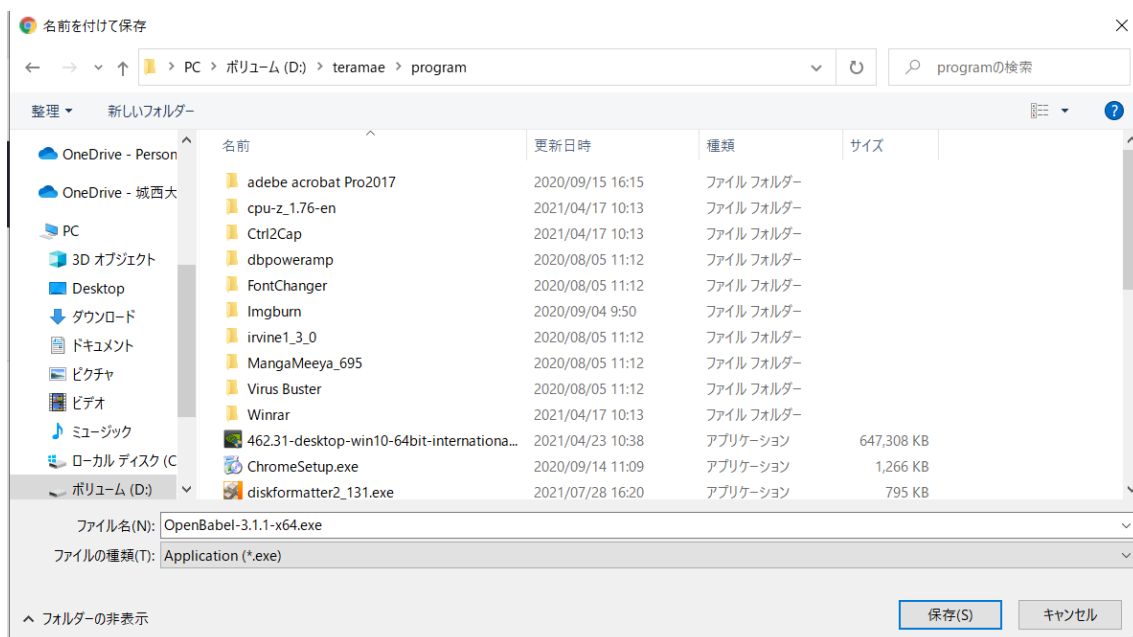


図 10 ダウンロード時のファイル名の選択画面

ダウンロードできたら、ダブルクリックすると図 11 の画面になるので Next をクリックすると、図 12 のライセンス承認の画面になるので I Agree（受け入れる）をクリックする。

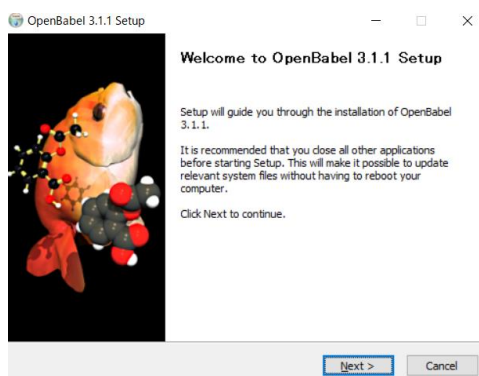


図 11 Open Babel インストール時の最初の画面

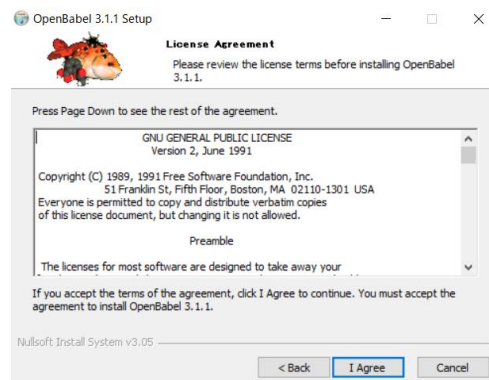


図 12 ライセンス承認画面

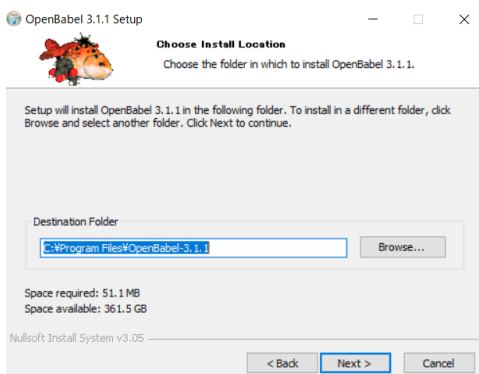


図 13 インストールフォルダーの指定

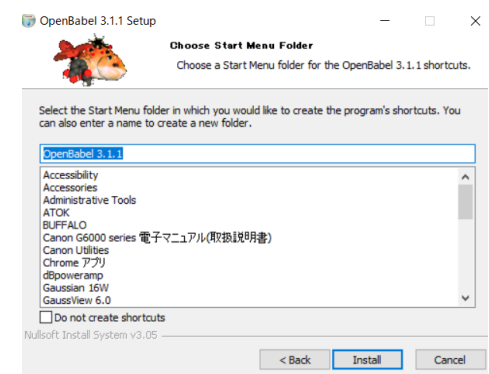


図 14 スタートメニューフォルダーの指定

図 13 のようにインストールするフォルダーを聞かれるので、そのままなら Next をクリックする。最後に図 14 のようにスタートメニューフォルダーの指定を聞かれるのでそのままであれば Install をクリックする。最後に図 15 の完了画面が出てインストールは終了する。

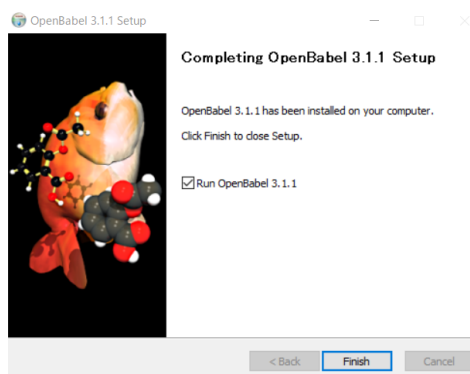


図 15 Open Babel インストールの終了

すぐに Open Babel を起動する場合は図 15 にあるように、Run OpenBabel 3.1.1 (3.1.1 はインストールした Open Babel のバージョン番号) にチェックをいれ、しない場合はチェックを外して Finish

をクリックする。

4K などの高解像度ディスプレイを使っていると Open Babel の画面が少しおかしくなることがあるのでその場合は以下のように修正しておく。問題が起こらなかった場合および通常の 2K ディスプレイを使用している場合はこれらの手続きは不要である。

図 16 に示したように、スタートメニューから Open Babel GUI を右クリックして、「その他」から「ファイルの場所を開く」を選ぶと、図 17 のようにフォルダー内のファイルが表示されるので、Open Babel GUI を右クリックして、プロパティ(R)を選ぶ。

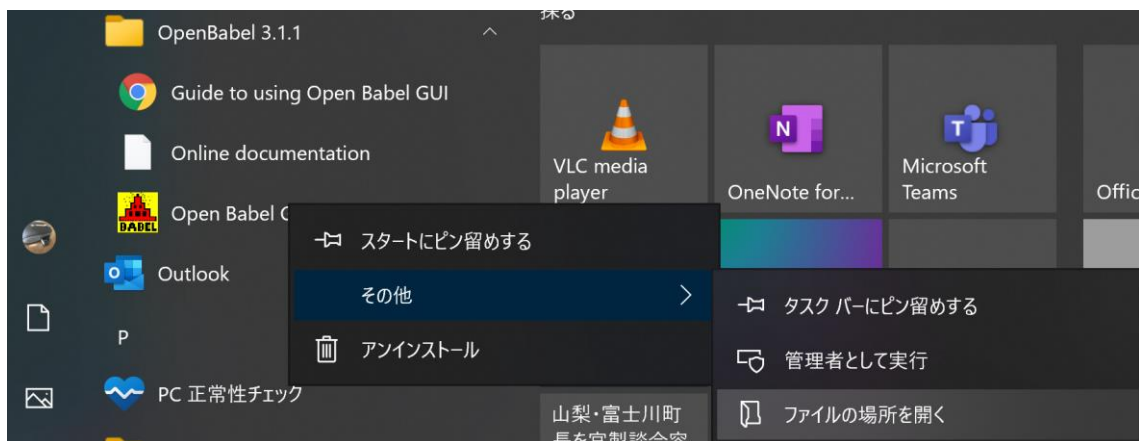


図 16 Open Babel GUI のある場所をさがす

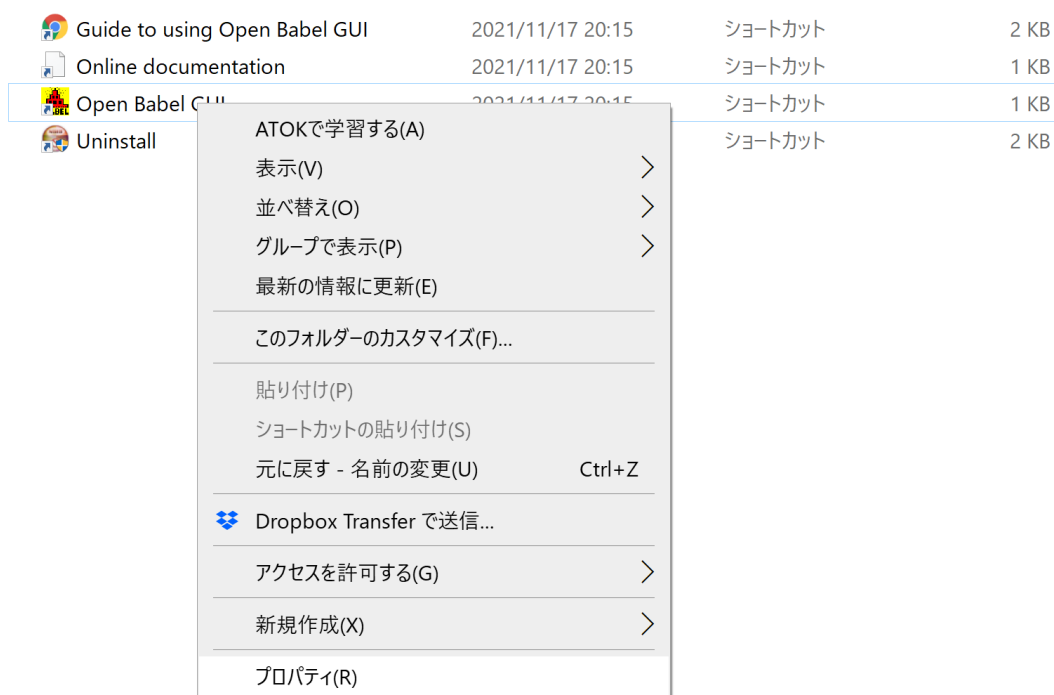


図 17 Open Babel GUI のプロパティを選ぶ



図 18 プロパティのダイアログ画面

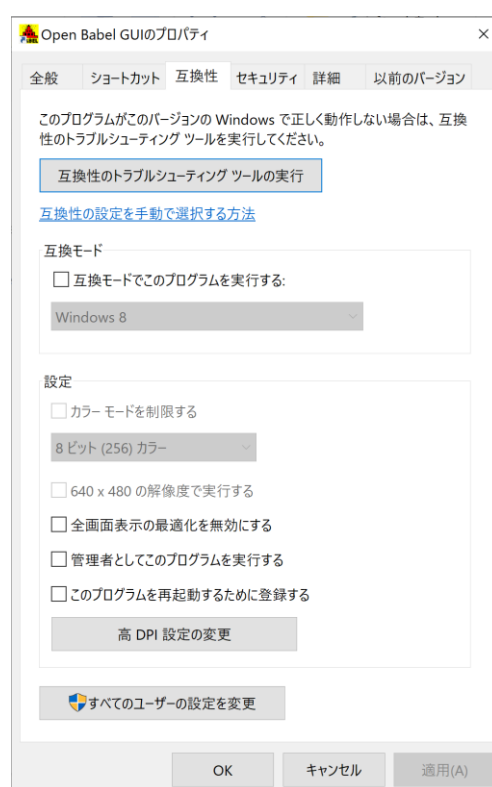


図 19 互換性のダイアログ画面

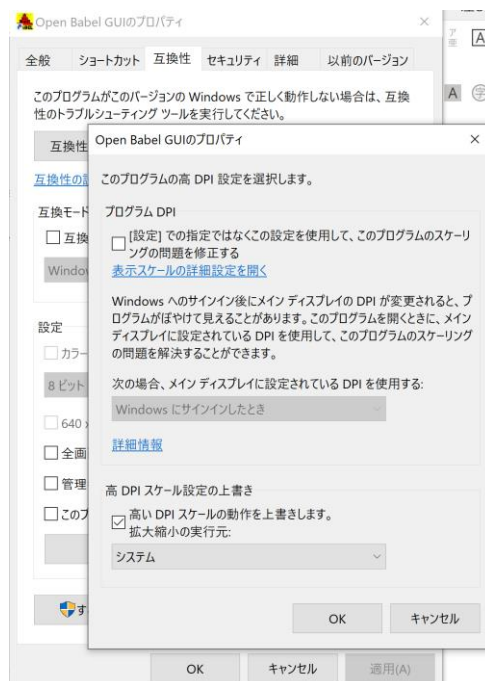


図 20 互換性タブから高 DPI 設定の変更を選ぶ

そうすると図 19 に示したダイアログが現れるので、互換性のタブをクリックすると図 19 の画面

になるので、高 DPI 設定の変更を選ぶ。すると図 20 で示したダイアログが現れるので、ここで「高 DPI 設定の上書き」で、「高い DPI スケールの動作を上書きします。」にチェックを入れて、拡大縮小の実行元はプルダウンメニューから「システム」を選ぶ。以上で正しく表示されるようになる。

以上で必要な準備ができた。Open Babel で Gaussian16 の入力を作るのは次のように行う。

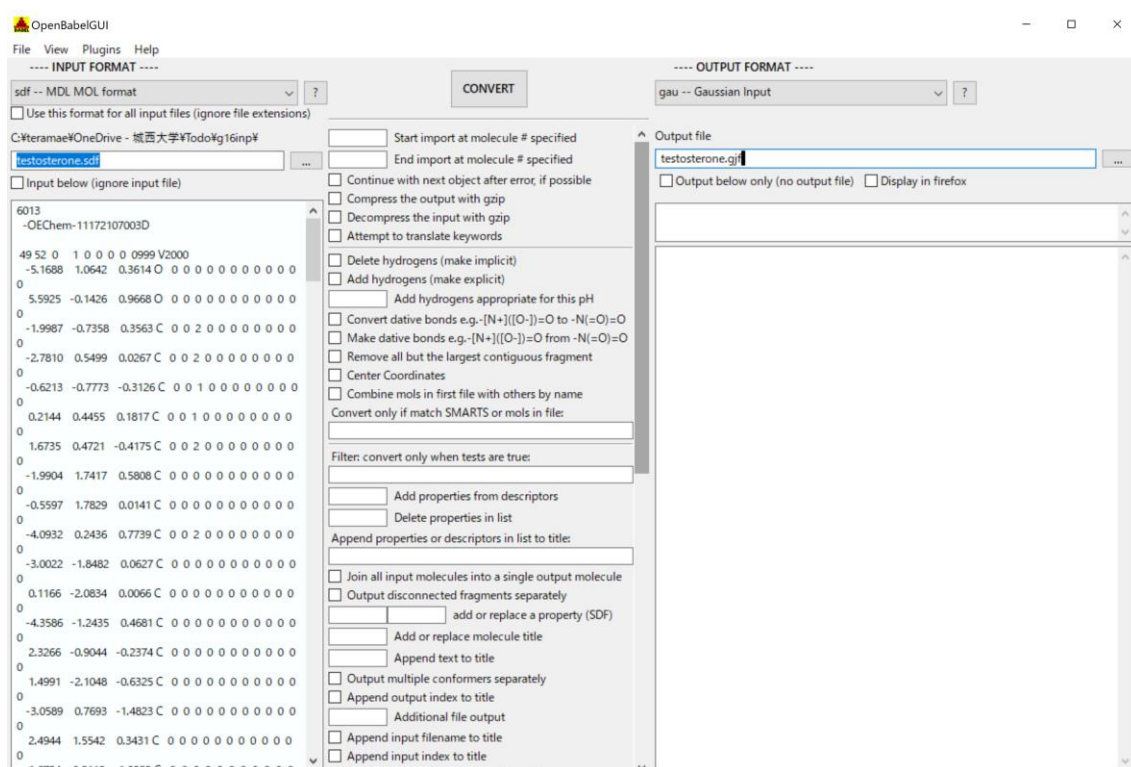


図 21 Open Babel の入力画面

プログラムを立ち上げると、図 21 の画面になるので、入力に sdf—MDL MOL format を、出力に gau—Gaussian Input を選んでおき、それぞれのファイル名を指定する。次に CONVERT をクリックすればファイル変換が実行される。Output file の名前にフルパスで日本語が入っていると Gausview がファイルを読めなくなるので注意する必要がある。

4. 実行例

表 1 に PubChem サイトからダウンロードされた sdf ファイルの最初の部分を示す。このように最初の部分に座標の値が書かれているため、変換プログラム（今の場合は Open Babel）により容易に変換することができる。

表 2 には Open Babel を用いて、testosterone.sdf ファイルから G16 の input ファイルである testosterone.gif に変換した結果の先頭部分を示す。ここで注意しないといけないのは、デフォール

トで変換されるのは 3 次元の座標だけであることで、その他の入力については図 21 の中央部に様々なオプションが選べるようになっているので各自選択する。デフォルトのままの場合は、当然追加されないため各自で変換された gjf ファイルに追記する必要がある。

表 1 sdf ファイルの先頭部分の内容例

6013															
-OEChem-11172101242D															
49 52 0	1	0	0	0	0	0	0999	V2000							
8.6500	2.1128	0.0000	O	0	0	0	0	0	0	0	0	0	0	0	0
2.0000	-2.1948	0.0000	O	0	0	0	0	0	0	0	0	0	0	0	0
7.3931	-0.1425	0.0000	C	0	0	2	0	0	0	0	0	0	0	0	0
7.3931	0.8575	0.0000	C	0	0	2	0	0	0	0	0	0	0	0	0
6.5271	-0.6425	0.0000	C	0	0	1	0	0	0	0	0	0	0	0	0
5.6610	-0.1425	0.0000	C	0	0	1	0	0	0	0	0	0	0	0	0
4.7510	-0.6493	0.0000	C	0	0	2	0	0	0	0	0	0	0	0	0
6.5271	1.3575	0.0000	C	0	0	0	0	0	0	0	0	0	0	0	0
5.6610	0.8575	0.0000	C	0	0	0	0	0	0	0	0	0	0	0	0
8.3393	1.1623	0.0000	C	0	0	2	0	0	0	0	0	0	0	0	0
8.3393	-0.4472	0.0000	C	0	0	0	0	0	0	0	0	0	0	0	0

表 2 openbabelGUI から出力される gjf ファイルの先頭部分

!Put Keywords Here, check Charge and Multiplicity.			
#			
6013			
O 1			
O	8.65000	2.11280	0.00000
O	2.00000	-2.19480	0.00000
C	7.39310	-0.14250	0.00000
C	7.39310	0.85750	0.00000
C	6.52710	-0.64250	0.00000
C	5.66100	-0.14250	0.00000
C	4.75100	-0.64930	0.00000
C	6.52710	1.35750	0.00000
C	5.66100	0.85750	0.00000
C	8.33930	1.16230	0.00000
C	8.33930	-0.44720	0.00000

G16 に対する入力 は 2 行目にある #行 に追記すれば良い。例えば #HF/6-31G(d,p) opt などとする。分子の電荷 0 とスピン多重度 1 は 6 行目にすでに記入されている。

5. おわりに

本研究ではフリーデータベースを利用して、G16 プログラムの入力データファイルを作成する方法について、データベースからの sdf ファイルのダウンロードのやり方から始まり、座標データの変換プログラムである OpenBabel のダウンロード・インストール・使い方について詳細に述べた。

我々の研究室においては、今まであまり研究対象としていなかった薬品の構造のモデルビルディングに相当苦勞をしていたが、本論文の方法に従って作成をしてもらうことにより、大幅に作成効率が改善した。同様の困難をされている方たちの参考になることを期待している。

参考文献

- 1) Gaussian 16, Revision C.01, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT (2016).
- 2) M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, J. A. Montgomery, Jr., *J. Comput. Chem.*, **14**, 1347-1363 (1993).
- 3) S. Kim, J. Chen, T. Cheng, et al. PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* 49(D1):D1388–D1395 (2021). [doi:10.1093/nar/gkaa971](https://doi.org/10.1093/nar/gkaa971)
- 4) N M O'Boyle, M Banck, C A James, C Morley, T Vandermeersch, and G R Hutchison. "Open Babel: An open chemical toolbox." *J. Cheminf.*, 3, 33 (2011). DOI:10.1186/1758-2946-3-33
- 5) The Open Babel Package, version 3.1.1 <http://openbabel.org> (accessed Dec 2021)