

# プログラミング言語 Python を活用と文化財材料分析の研究 —量子化学計算の一例—

高木 秀明<sup>a,b</sup>

プログラミング言語 Python と組み合わせることできる量子化学計算パッケージ PSI4 と既報の量子化学計算パッケージ Gaussian09W との構造最適化計算結果との比較を行った。比較に用いたのは染料の発色成分ベルベリンである。計算結果は、全エネルギーがそれぞれ、PSI4 で -1128.3897886、Gaussian で -1128.3897939 Hartree であった。PSI4 の特徴は、量子化学計算の手法を Python のコード内に記述し、量子化学計算をし、その結果を引き続き統計処理や機械学習、深層学習などの処理も行えること、また、初期構造も処理したのち量子化学計算をはじめることができるなどである。今後、多彩な活用方法があると思われる。

## 1. はじめに

著者はこれまで、文化財の材料分析に量子化学計算を取り入れ、顔料や染料といった色材の量子化学計算結果と光吸収特性との考察を試みてきた [1-4]。文化財に使用された材料分析する場合、多くは、固体状態を対象とするため、結晶状態の構造から計算をスタートする量子化学計算結果との比較が容易であると考えたためである。

化学におけるデータサイエンスや科学技術計算では、実験や計測結果のデータの統計処理だけでなく、応用化学や化学工学における合成反応における生成物の予測にも利用されている。マテリアルズインフォマティクス分野では、計算科学の自動化や過去の研究成果のデータを機械学習や深層学習を通して、新たな新規材料の開発に利用されている [5,6]。

量子化学計算パッケージは、商用の Windows 版 Gaussian09W パッケージを用いて研究を行ってきた。これまでの著者の文化財関連以外の研究分野で 1 分子を対象とした量子化学計算を研究していたため、文化財材料にも同様に 1 分子を対象に計算を行ってきた。本論文では、多種ある 1 分子を対象とした量子化学計算パッケージのうち、オープンソースの量子化学計算パッケージの PSI4 を用いて、既報の量子化学計算との比較を報告する。

## 2. プログラミング言語 Python と科学計算

本報告では、量子化学計算パッケージ PSI4 を取り上げている。PSI4 は、GNU Lesser General Public License version 3 (L-GPL3.0) のライセンスで配布されているオープンソースのパッケージである [7]。実行型の Psithon、プログラム言語 Python で用いるモジュール型の PsiAPI が公開されている。ソースコード公開の他にバイナリパッケージが、Linux、macOS、Windows 版が提供されている。Python と組み合わせて使う場合は、Python のパッケージ管理システム conda の環境下で行うための Anaconda デイストリビューション用のインストーラーが用意されている。

プログラム言語 Python は、科学技術の分野では、統計処理やベクトル計算などのアプリケーションと比較されながら、普及していった。他のプログラム言語と処理速度では劣るもののモジュールやパッケージ、ライブラリ [注 1] と呼ばれる関数群が多種類公開されている。関数群をユーザーの目的に応じてインストールする場合、Python のインストールコマンドを実行すると、現在の Python 環境に応じたバージョンのファイル群を自動的にダウンロード、インストールされる。科学技術系でよく使われるものとして、数値解析用ライブラリ SciPy[8]、データ解析支援ライブラリ pandas[9]、ケモインフォマティクスと機械学習のためのモジュール群 RDKit[10] などがある。Python にはコマンドラインを利用した対話型で行う方法とプログラムコード（ソース）を記述し、保存し、Python のコマンドでこの保存したコードを実行する方法がある。

データサイエンスでは、数値データを計算することが主流であるように思われるが、テキストなど文字列、図像を対象に行われていることもある [11]。複雑な化学構造を文字列で表現する SMILES(Simplified Molecular-Input Line-Entry Sysrem) 記法が 1988 年に提唱された [12]。構造式を描画する際に便利な記法ではあるが、著者の経験から、化学分野では、学術論文や書籍の執筆・出版などには利用されていなかったと記憶している。しかしながら、化学構造を描画し、それらをデータベースや Web を利用したシステムなどで運用する際は、文字列を扱うほう便利であると考えられる。RDKit モジュールを使用すると SMILES 記法をプログラミングに導入することができるので、化学構造を自動生成し、あるいは化学的特徴のある構造を記法で描画し、プログラムに取り込むことができる。

### 3. PSI4 を用いた量子化学計算の一例

既報で文化財の色材に利用される顔料や染料のモデル化合物を対象とし、量子化学計算パッケージ Gaussian を用いた化合物の最適化構造や電子スペクトルも計算結果を報告している。計算をはじめ際の初期構造は、いずれも報告にある X 線構造解析の結果から得られた結晶構造のデータファイル Crystallographic Information File (cif) を採用している。計算が収束するよう初期構造の一部を水素などの簡単な原子に置き換えモデル化合物としたり、振動により収束しにくい結合を結合長や結合角、二面角などを固定したりして収束に導いている。

本報告では、Gaussian と PSI4 の初期構造入力から計算終了までの過程を比較するため、既報の中で、初期構造の調整がほとんど必要なかったヨウ化ベルベリンを計算対象とした [4]。この化合物の構造最適化計算にて比較することにした。

構造最適化計算と振動数計算は、密度汎関数を用い、基底関数は 6-31G(d) を用いた。入力ファイルは表 1 のとおりである。PSI4 の Anaconda プロンプトから実行し、Psithon のバージョンは、1.4rc2.dev1 である。

既報では Gaussian09W パッケージで構造最適化（キーワード：opt）と振動数計算（同：freq）を 1 回の実行でおこなった。このとき、計算時間は 2 時間 36 分 27 秒で計算終了時の全エネルギーは -1128.3897939 Hartree であった。PSI4 での計算時間は 14 時間 26 分 4 秒と計算に時間がかかり、全エネルギーは -1128.3897886 Hartree であった。構造最適化のみ（入力ファイルは表 2）では 1 時間 29 分 49 秒であるので、振動数計算の部分に時間がかかっている。計算過程におけるアルゴリズムなどが異

なるので時間の比較は困難であるが、全エネルギーについては、両者の間には差がないと判断している。

表1 PSI4 の構造最適化と振動数計算の入力ファイル (抜粋)

---

```
# (コメント行)

Molecule{
1,1
O
(原子座標 略)
H
}

set basis 6-31g(d)

optimize('b3lyp')
b3lyp_e, b3lyp_wfn = frequencies('scf', return_wfn=True, dertype=1)

[EOF]
```

---

表2 PSI4 の構造最適化の入力ファイル (抜粋)

---

```
# (コメント行)

(表1 と同一のため 略)
}

set basis 6-31g(d)

optimize('b3lyp')
b3lyp_e, b3lyp_wfn = frequencies('scf', return_wfn=True, dertype=1)

[EOF]
```

---

PSI4 を用いた量子化学計算では、構造最適化計算から既報の Gaussian09W とほぼ同じ計算結果が得られた。オープンソースであるため、紙媒体の詳しい参考書籍が出版されておらず、PSIcode 内の Tutorial や Docs、Forum のサイトなどを利用しながら行う必要があった [7a]。

PSI4 は、2022 年 2 月現在、Windows インストーラー版においては、Python3.8 のバージョン向けが最新となっている。Linux や MacOS 向けや conda でインストール場合は、Python3.9 に対応している。

#### 4. おわりに

Python を使用する際の注意点は Python のバージョンに対応するライブラリやパッケージ、モジュールが必要となる。Python のもつインストーラーで最適なバージョンを検索してインストールできるが、バージョンによって機能が制限される場合があるので注意が必要である。複数のライブラリをインストールする場合、ライブラリが対応している古いバージョンの Python を使用しなければならないことになる。

Python のコード内に PsiAPI を用いて量子化学計算を実行することで、分子構造の生成を自動化したり、得られた計算結果を機械学習や深層学習の方法を用いて処理することも容易になると予想される。

#### 注

[注 1]Python では、関数やクラスをまとめたものモジュールといい、モジュールを複数集めたものをパッケージ、パッケージを複数集めたものをライブラリと呼ぶ。

#### 参考文献

- [1] 高木秀明、文化財の材料分析に供する理論化学的展開について－無機金属錯体の計算化学の点から－、文化財情報学研究第 15 号、pp.37-46、2018 年。
- [2] 高木秀明、文化財の材料分析における反射スペクトル測定への理論化学的展開の試み、文化財情報学研究第 16 号、pp. pp.19-26、2019 年。
- [3] 高木秀明、天然染料ウコンのモデル化合物の計算化学の一例、文化財情報学研究第 17 号、pp.15-22、2020 年。
- [4] 高木秀明、黄色天然染料の光吸収特性への理論化学的試み、文化財情報学研究第 18 号、pp.13-22、2021 年。
- [5] 知京豊裕、「マテリアルズインフォマティクスの現状と課題」- 海外の動向と日本の挑戦 -、情報知識学会誌 27 巻、4 号、pp.297-304、2017 年。
- [6] 伊藤 聡、日本のマテリアルズインフォマティクス研究、人工知能、34 巻、3 号、pp.325-329、2019 年。
- [7](a)PSI4 のオフィシャルページ、<https://psicode.org/>、2022 年 2 月 16 日閲覧。  
(b) Daniel G. A. Smith, Lori A. Burns, Andrew C. Simmonett, Robert M. Parrish, Matthew C. Schieber, Raimondas Galvelis, Peter Kraus, Holger Kruse, Roberto Di Remigio, Asem Alenaizan, Andrew M. James, Susi Lehtola, Jonathon P. Misiewicz, Maximilian Scheurer, Robert A. Shaw, Jeffrey B. Schriber, Yi Xie, Zachary L. Glick, Dominic A. Sirianni, Joseph Senan O' Brien, Jonathan M. Waldrop, Ashutosh Kumar, Edward G. Hohenstein, Benjamin P. Pritchard, Bernard R. Brooks, Henry F. Schaefer III, Alexander Yu. Sokolov, Konrad Patkowski, A. Eugene DePrince III, Uğur Bozkaya, Rollin A. King, Francesco A. Evangelista, Justin M. Turney, T. Daniel Crawford, and C. David Sherrill, "PSI4 1.4: Open-source software for high-throughput quantum chemistry", J. Chem. Phys. 152(18) 184108 (2020).
- [8]SciPy のオフィシャルページ、<https://scipy.org/>、2022 年 2 月 16 日閲覧。
- [9]pandas のオフィシャルページ、<https://pandas.pydata.org/>、2022 年 2 月 16 日閲覧。
- [10]RDKit のオフィシャルページ、<https://www.rdkit.org/>、2022 年 2 月 16 日閲覧。
- [11] 村上征勝、『文化を計る－文化計量学序説』、朝倉書店、2015 年。
- [12]David Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28(1), pp.31-36, 1988.

所属：

<sup>a</sup> 吉備国際大学外国語学部外国学科

<sup>b</sup> 吉備国際大学文化財総合研究センター