

令和元年6月5日現在

機関番号：82626

研究種目：基盤研究(C) (一般)

研究期間：2016～2018

課題番号：16K05711

研究課題名(和文)有機分子立体配座コードプログラムの開発

研究課題名(英文)Development of Conformational Code for Organic Molecules (CCOM) Program

研究代表者

和泉 博 (Izumi, Hiroshi)

国立研究開発法人産業技術総合研究所・エネルギー・環境領域・主任研究員

研究者番号：20356455

交付決定額(研究期間全体)：(直接経費) 2,300,000円

研究成果の概要(和文)：IUPAC命名法P-94.2順位則に代わる、有機分子立体構造(立体配座)ビッグデータで共通部分構造を識別するのに必要となる修正案を国際科学会議IUPACに提案した。オリジナルの三次元化学構造検索機能を活用し、理論計算の一つである、密度汎関数法(DFT)計算のための初期構造を自動生成するConfFragGenerationプログラム、RingFragGenerationプログラムや立体配座コード文字列を出力するConfCodeView3プログラムを開発した。プログラムを活用した免疫抑制剤ラパマイシン分子の解析から、オキサリル基の新規水素結合様式を見出した。

研究成果の学術的意義や社会的意義

2017年ノーベル化学賞を受賞したクライオ電子顕微鏡法の単粒子解析のように、より詳細な有機分子立体構造(立体配座)の分類の要請がある。そこで、研究成果を基に立体配座ビッグデータで共通部分構造を識別するのに必要となる順位則を国際科学会議に提案し、国際基準として認められるための取り組みを行っている。本規則及び開発したプログラムは、溶液状態にある有機分子の正確な未知の立体構造の検証に利用できる。さらに、ディープラーニング解析と組み合わせ、タンパク質立体構造解析等に応用することにより、新規機能性有機分子の設計への展開が期待できる。

研究成果の概要(英文)：The revised sequence rule of P-94.2 of the IUPAC Rules for Nomenclature of Organic Chemistry for an identification of the 3D maximal common substructures has been proposed. The auto-generation program of initial structures for density functional theory (DFT) calculations (ConfFragGeneration), the ring structure-enabled auto-generation program of initial structures for DFT calculations (RingFragGeneration), and the codification program of organic structures, conformations, (ConfCodeView3) using the original method of 3D chemical structure search were constructed. The new hydrogen-bonding of oxalyl group in an immunosuppressant rapamycin was found by using the programs.

研究分野：構造有機化学

キーワード：立体配座 機能性有機分子 IUPAC命名法 三次元化学構造検索 プログラム ケモインフォマティクス
赤外円二色性分光法 創薬

様式 C - 19、F - 19 - 1、Z - 19、CK - 19 (共通)

1. 研究開始当初の背景

薬剤、繊維、液晶などに使用される機能性有機分子は日本を支えるものづくりに欠かせない素材であり、分子レベルでの構造特性を突き詰めることが国際競争力確保につながっていく。抗がん剤タキソールなどの機能性有機分子の多くは原子間結合の回転の自由度によりその形(立体配座)を変化させながら機能を発揮している。しかしながら、その機能発揮に深く関わる、ピコ秒オーダーの精度の高い構造(立体配座)変化の情報は容易に取り出すことができなかった。特に創薬に役立つ精度の高い立体配座情報を簡便に取り出す技術が求められていた。

また、有機分子の立体構造を表示する多くのソフトウェアではボタン一つで光学活性分子のキラリティーを表示することが可能である。一方で、ゴーシュプラス、トランスなどの有機分子の立体配座を簡便に表示するソフトウェアは存在していなかった。

SMILES (Simplified molecular input line entry specification syntax) 記法は、化学構造式を ASCII 符号の英数字で文字列化した表記法であり (D. Weininger, A. Weininger, J. L. Weininger, *J. Chem. Inf. Comput. Sci.*, **29**, 97-101 (1989))、その拡張である SMARTS (Smiles ARbitrary Target Specification) 記法を用いることにより、化学データベースからの構造検索が可能になっていた。2次元の化学構造式はこの SMARTS 記法により容易に部分構造検索が行えるようになっていたが、3次元の部分構造の検索は容易ではなかった。研究開始当初の時点で、理論計算から得られた最適化構造と X 線結晶構造解析から得られた分子構造との比較では構造の揺れの問題があり、分子モデルの重ね合わせくらいしか構造比較が出来なかった。

2. 研究の目的

申請者は世界で初めて有機分子の立体配座を簡便に記述する Conformational Code for Organic Molecules (CCOM) プログラム (図 1) を開発しており、本プログラムを用いることで薬剤の機能に深く関わる構造変化をデータマイニング可能であることを明らかにした。そこで、CCOM プログラムをさらに発展させ、さまざまな手法の構造データを連携させて精度の高い立体配座変化情報を蓄積する解析技術を開発する。これにより、立体配座変化を利用した機能性分子の設計に活用し、立体配座に関する構造有機化学の基盤を確立する。

CCOM プログラムにより、有機分子中に多数存在する立体配座をわかりやすく記述し、かつ密度汎関数法(DFT)を用いた構造最適化でギブズ自由エネルギー値の異なる構造をきちんと区別することができる。さらに、特許取得した相同性判定手法(特許 5672596)をプログラムに組み込み SMARTS 記法と組み合わせることで、異なる分子でも 3D フラグメント構造を簡単に検索することが可能である。

ところが、開発した CCOM プログラムは、立体配座に関する IUPAC (International Union of Pure and Applied Chemistry) 命名法に準拠した方法を採用しており、その中に、絶対配置 *R*, *S* の決定に使われる優先則 (Cahn-Ingold-Prelog 則) が含まれていた。通常同じ分子の立体配座を比較するには問題なかったが、別の絶対配置の記述である *D*, *L* とは違い、異なる分子の比較には適さなかった。すなわち、キラル中心に結合した 4 個の結合の相対的な位置関係で *R*, *S* が決定されるため、例えば、水素原子が塩素原子に置き換わると *R*, *S* が変わるケースが起こりえる。このことから、二面角を決定するための 4 個の原子の選択において、データベースに保存された分子と検索を行うフラグメントとの間で選択される原子が違ってくるケースが起こることが大きな課題になっていた。そこで、IUPAC 命名法とは異なる規則を提案し、共通する構造であればどんな 3D フラグメントでも対応可能な CCOM プログラムを開発する。

本目的を達成するため、(1)構造解析データ連携 CCOM プログラムの開発、(2)CCOM プログラムを用いた 3D フラグメントデータベースの構築の研究テーマに取り組む。

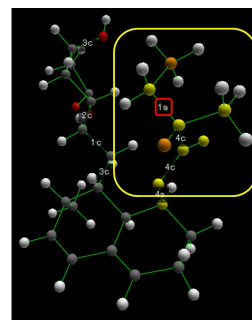
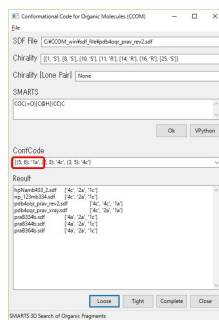


図 1. CCOM プログラム

3. 研究の方法

(1)構造解析データ連携 CCOM プログラムの開発

ConfFragGeneration, RingFragGeneration プログラムは、オープンソースである Pybel 及び OpenBabel のライブラリーの他に RDKit, wxPython, VPython, NumPy, pywinauto, GaussView のライブラリー等を利用し、python スクリプト言語により作成した。

アミノ酸などでは L-システインのように *R/S* 表記の代わりに *D/L* 表記が一般に使われている。これは、L-システインのみ *R*-配置になり、それ以外のアミノ酸は *S*-配置となるためである。このように、*D/L* 表記とは違い、*R/S* 表記は異なる分子の比較には使用できない。立体配座に関する命名法の規則である IUPAC Rule P-94.2 にも *R/S* 表記の規則が一部含まれており、異なる分子の比較に不具合が生じた。そこで、Rule P-94.2 に代えて、決定できるなら、たとえ原子番号が大きくなくとも中心の結合から近い順に二面角決定のための原子を選択するという規則をプログラムに導入し解析を行った。

さらに、DFT 計算で Gibbs 自由エネルギー値が小さく、存在比の大きいフラグメント分子の立体配座データのみデータベースに格納した。その際、環構造については index の小さい bond を切断し、ダミー原子としてアスタチンをそれぞれの原子位置に置き、再結合させたデータを格納した。また、フラグメント分子データベースに格納される構造データには、二面角分類コードだけではなく、選択された4個の原子及び詳細な二面角の情報を登録した。クエリとなる、環構造についてはアスタチンを付加した共通する構造ファイルを検索し、それぞれの構造ファイルについて、二面角位置コードの一致を確認して OpenBabel の SetTorsion 機能を使ってデータベースに格納された構造データの二面角情報に基づき順に個々の bond の立体配座を発生させた。発生させた構造からアスタチン原子を取り除いた分子構造をそのまま DFT 計算の入力ファイルとして保存した。

(2)CCOM プログラムを用いた 3D フラグメントデータベースの構築

ConfFragGeneration, RingFragGeneration プログラムを活用し、1万を超えるキラル有機分子立体配座構造の赤外円二色性(VCD)スペクトルコード化解析を行った。実測のスペクトルを再現できた立体配座構造について、Gibbs 自由エネルギー値がおおよそ 1 kcal/mol の範囲内にある異性体について、877 個の構造ファイルを選択し、フラグメント分子の立体配座データベースに格納した。フラグメントデータベースの有効性を検証するため、分子量 914 のマクロライド環を有するラパマイシンの VCD スペクトル解析及びエリスロマイシン、エベロリムス、テムシロリムス、プリオスタチン 1 のマクロライド分子の立体配座探索を行った。

免疫抑制剤であるラパマイシン分子のオキサリル基の新規水素結合様式について、これまでの核磁気共鳴 (NMR) を用いた溶液の解析では X 線結晶構造とよく似た構造をとると帰属されており議論になったことから、得られた立体配座構造について NMR の DFT 計算もあわせて行った。この解析において、立体配座コード文字列を出力する ConfCodeView3 プログラムを改良し、X 線結晶構造データと比較するため cif ファイルにも対応できるようにした。さらに、IUPAC Rule P-94.2 の不具合(図 2)に対応した CCOM プログラムを作成し、3D 部分構造検索を行った。

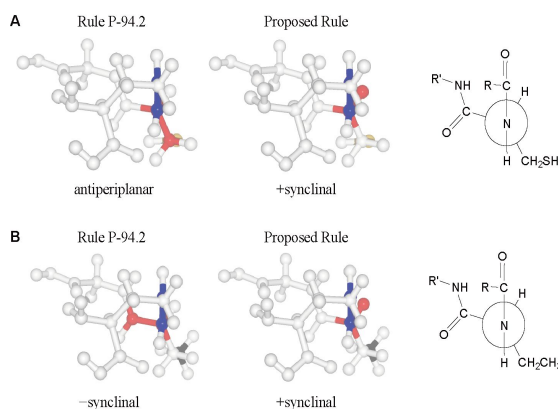


図 2 . P-94.2 と提案した規則による立体配座の帰属 (A: グルタチオン、B: オフタルミン酸)

4 . 研究成果

(1)構造解析データ連携 CCOM プログラムの開発

創薬に向けたリード有機分子の構造最適化において、最大共通部分構造 (MCS) の比較が必要である。特に近年、MCS 三次元構造比較を行うための重ね合わせ以外の手法のニーズが高まっている。異なる有機分子において MCS の立体配座を比較するには、二面角を決めるための 4 原子が一致しなくてはならない。IUPAC 有機化学命名法 (Blue Book) P-94.2 には、不斉炭素の R, S の定義に使用する順位則 (P-92) を利用することが定められている。しかしながら、P-94.2 はこの順位則の関係で同一の分子にしか適用できない制約が生じていた。

P-94.2 には、4 原子の選択において、

- すべての原子あるいは置換基の組が異なっている場合、順位則 (P-92) で優先する組を選択する。
- 一つの原子あるいは置換基の組のみ異なっている場合には、その組を選択する。
- すべての原子あるいは置換基の組が同一である場合には、最も二面角が小さくなる組を選択する。

という基準が定められている。

構造解析データ連携 CCOM プログラムの開発を進める中で、P-94.2 はほとんどのケースで適用できるものの、原子あるいは置換基の一つの組のみ異なっている場合にはその組を選択するという規則と順位則との間で齟齬を生じるケースで数少ない例外がみられることが判明した。その点を考慮して、二面角を決定するための 4 個の原子の選択において、データベースに保存された分子と検索を行うフラグメントとの間で選択される原子が違ってくる問題を解決した。

さらに、P-94.2 の基準 a), b) を融合した、a) すべての原子あるいは置換基の組が異なっている場合、あるいは一つの原子あるいは置換基の組のみ異なっている場合には、n 層と n+1 層の各々の原子の順位付け (P-92.1.5 階層的有向グラフの探索を参照のこと) の間に、一つの原子あるいは置換基が唯一のものならその組を選択するという規則を除いて、順位則 (P-92) で優先する組を選択するという規則(図 2) を IUPAC に提案した (H. Izumi, *Chem. Int.* **40**, 36-37 (2018))。提案した新たな規則は例外なく有機分子の MCS の比較に適用可能である。

ConfFragGeneration の作成において、SMARTS 記法で表されるフラグメントには水素原子の情

せることが出来ることから所期の目的のデータベースを構築できた。データベースを検証したところ、分子量 300 程度の有機分子であれば MCS を選択して初期構造を自動生成できることがわかった。また、フラグメントデータをあらかじめ準備すればエリスロマイシン、エベロリムス、テムシロリムス、プリオスタチン 1 のマクロライド分子についても初期構造を自動発生させることが可能だった。さらに、ラパマイシン分子について得られた存在比の大きい立体配座構造データを用いて VCD スペクトル解析を行い、これまで報告されていなかったオキサリル基の水素結合様式を見出した。

C=O 伸縮振動に帰属される VCD バンドに明らかな溶媒効果が観測され、 CDCl_3 中の 1620 cm^{-1} 付近の負のバンドは $\text{DMSO}-d_6$ 中で消失した(図 5)。分子モデル計算では、 1620 cm^{-1} 付近の最も低波数側の水素結合したアミド C=O 伸縮振動の VCD バンドはピペコリル ケトアミド及びオキサリル基を含む *rapZ* フラグメント(図 4)の立体配座を強く反映しており、*rapZa* ($2c4c5c4a$) 構造パターンのみが負のバンドを再現した(図 6)。さらに、*rapZa* 構造パターンを伴う 4 個の立体配座構造が他の構造より計算上約 2 kcal/mol ほど安定であり、その平均化予測 VCD スペクトルは実測の CDCl_3 のスペクトルとよい一致を示した(図 5)。

これまで、 $\text{DMSO}-d_6$ 中のラパマイシン分子の NMR 解析と分子動力学計算結果から、溶液状態の立体配座構造は X 線結晶構造と類似の構造をとると考えられてきた。そこで、前述の 4 個の立体配座構造について、NMR の DFT 計算も行い、 CDCl_3 中の major な異性体のケミカルシフトとよい一致を示すことを確認した。ここで、NMR と VCD の主要な違いはデータ取り込みのタイムスケールであった(NMR: ミリ秒から秒のオーダー、VCD: ピコ秒のオーダー)。NMR タイムスケールの間に 3 個以上の局所安定化立体配座構造変化が起きる場合には、その平均化構造は 1 個の分子モデルで表現することはできず、VCD 解析の分子モデル表現のように、複数の分子モデルをその存在比とともに記述する必要がある。

これまでの通説を覆し、重クロロホルム中では、結晶構造とは異なる水素結合をもつ *rapZa* 構造パターンを伴う立体配座をとっていることが明らかとなった。

作成したプログラムについては、知的財産の観点から、立体配座コード変換機能および 3D フラグメント構造検索機能を有する CCOM プログラムと、RingFragGeneration プログラムから立体配座コード文字列を出力する機能のみ取り出した ConfCodeView3 プログラムを、科学技術振興機構(JST) researchmap 上で公開した。ダウンロード数も順調の伸び(<http://researchmap.jp/read0004613/%E8%B3%87%E6%96%99%E5%85%AC%E9%96%8B/>)、延べ 147 件に至っている。

ラパマイシン分子の新規水素結合様式を見出したように、溶液状態にある有機分子の正確な立体配座が検証されている例は少ない。これは有機分子全体の立体配座の記述の仕方が規定されていないことも一因である。2017 年ノーベル化学賞を受賞したクライオ電子顕微鏡法の単粒子解析のように、より詳細な立体配座の分類の要請もある。現在、P-94.2 を融合した規則の IUPAC 有機化学命名法(Blue Book)への採用を目指した活動を行っている。これを足がかりに、動的な立体配座変化の比較を目的とした、分子全体を記述するための立体配座の国際規格制定に向けた動きにつなげ、立体配座と機能との相関解析の深化に努めていく。

謝辞：ラパマイシン分子の VCD 測定を行ってくださった、北海道大学谷口透助教授及び門出健次教授に深謝する。

5. 主な発表論文等

[雑誌論文](計 2 件)

- 1) H. Izumi, Consideration of the sequence rule in rule P-94.2, *Chemistry International*, 査読有, **40**, 36-37 (2018). DOI: 10.1515/ci-2018-0323.
- 2) H. Izumi, L. A. Nafie, R. K. Dukor, Three-Dimensional Chemical Structure Search Using the Conformational Code for Organic Molecules (CCOM) Program, *Chirality*, 査読有, **28**, 370-375 (2016). DOI:10.1002/chir.22596

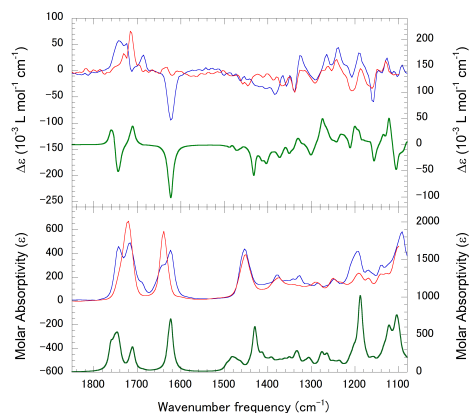


図 5. ラパマイシンの VCD 及び IR スペクトルの実測と予測の比較 (赤: $\text{DMSO}-d_6$ 、青: CDCl_3 、緑: B3LYP/6-31G*)

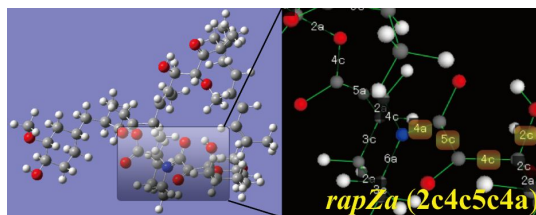


図 6. *rapZa* ($2c4c5c4a$) 構造パターン

〔学会発表〕(計12件)

- 1) 和泉 博、インフォマティクスにむけた IUPAC 有機化学命名法 P-94.2 順位則修正の提案、日本化学会 第99 春季年会 (2019)、2019 年
- 2) 和泉 博、IUPAC 有機化学命名法 (Blue Book) P-94.2 中の順位則に関する考察、第29 回基礎有機化学討論会、2018 年
- 3) 和泉 博、Consideration of the Sequence Rule in Rule P-94.2 of IUPAC "Blue Book" Nomenclature of Organic Chemistry, Chirality 2018、2018 年
- 4) 和泉 博、Supersecondary Structure Code (SSSC) of Interferon , , and 、JCUP IX、2018 年
- 5) 和泉 博、環構造対応 3D フラグメント構造検索初期構造自動作成プログラムの構築、第40 回ケモインフォマティクス討論会、2017 年
- 6) 和泉 博、ConfCodeView3 プログラムを用いた溶液状態のラパマイシンの立体配座解析、第28 回基礎有機化学討論会、2017 年
- 7) 和泉 博、CCOM(Conformational Code for Organic Molecules)プログラムを用いたラパマイシンの立体配座解析、VCD ROA ユーザー会&勉強会 2017、2017 年
- 8) 和泉 博、Conformational Analysis of Rapamycin in Solution Using the Conformational Code for Organic Molecules Program、Chirality2017(ISCD-29)、2017 年
- 9) 和泉 博、Development of Conformational Code for Organic Molecules (CCOM) Program for the Input of Artificial Intelligence、JCUP VIII、2017 年
- 10) 和泉 博、Development of Conformational Code for Organic Molecules (CCOM) Program for the Input of Artificial Intelligence、日本化学会第97 春季年会(2017)、2017 年
- 11) 和泉 博、3D フラグメント構造検索を用いた初期構造自動作成プログラムの構築、第39 回ケモインフォマティクス討論会、2016 年
- 12) 和泉 博、三次元化学構造検索にむけた立体配座コード規則の代替案、第27 回基礎有機化学討論会、2016 年

〔図書〕(計1件)

- 1) H. Izumi, in *Protein Supersecondary Structures, Methods and Protocols*, 2nd ed., ed. by A. E. Kister, Springer Nature, New York, **2019**, Vol. 1958, Chap. 17, pp. 329-340.

〔その他〕

ホームページ等

- 1) プログラム公開ホームページ

<http://researchmap.jp/read0004613/%E8%B3%87%E6%96%99%E5%85%AC%E9%96%8B/>

- 2) 研究紹介 URL

<http://staff.aist.go.jp/izumi.h/>

6 . 研究組織

(2)研究協力者

研究協力者氏名：ローレンス ネイフィー

ローマ字氏名：Laurence A. Nafie

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。