

令和 3 年 6 月 4 日現在

機関番号：34315

研究種目：研究活動スタート支援

研究期間：2019～2020

課題番号：19K23813

研究課題名（和文）化合物ビッグデータを用いた炎症抑制物質の選択的開拓

研究課題名（英文）Selective exploration of anti-inflammatory substances using chemical big data

研究代表者

小川 慶子 (Ogawa, Keiko)

立命館大学・薬学部・助教

研究者番号：20844278

交付決定額（研究期間全体）：（直接経費） 2,100,000円

研究成果の概要（和文）：本研究は、化合物ビッグデータを利用して情報処理技術を用いた解析を行うことで、過去に蓄積されたデータから効率よく生物活性物質を開拓する手法の開発を目的とする。本研究成果として、機械学習の手法を用いることで、これまでに蓄積された情報（化合物データベース・論文）からトリテルペンの活性の有無（強弱）を判別する予測モデルを作成した。くわえてトリテルペン化合物について活性試験を実施し、数種の活性物質を見出した。活性試験の結果と予測結果を比較したところ、約8割の正答率で活性を予測できていることが検証された。

研究成果の学術的意義や社会的意義

本研究により、高い判別能を持つ抗HSV-1活性予測モデルの構築ができた。この予測モデルは種々のトリテルペン及びその配糖体についても適用が可能である。本予測モデルを用いることによって、抗HSV-1活性を持つトリテルペンを事前に想定して成分探索研究や構造誘導化を行うことができ、より効率的な活性物質開拓に繋がることが期待される。

研究成果の概要（英文）：The purpose of this study is to develop a new approach to identify the active compounds from the data in previous studies by using information processing technology. As a result of this research, we constructed a prediction model that discriminates the active or inactive of triterpenes by using machine learning methods. In addition, the activity assay of triterpenes were measured and several active compounds were found. The results of the activity tests were compared with the prediction results, and it was verified that the accuracy of the predictive model was approximately 80%.

研究分野：天然物化学、データサイエンス

キーワード：活性予測 トリテルペン 機械学習 天然物化学 ケモインフォマティクス 生物活性 ビッグデータ

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、研究者個人に帰属します。

1. 研究開始当初の背景

炎症性腸疾患は消化管に炎症・潰瘍を生じる自己免疫性炎症疾患である。クローン病と潰瘍性大腸炎に大別され、難病に指定されており、近年いずれも有病率が急激に増加していることから、早急な対策が必要とされる疾患である。近年、炎症性腸疾患の治療薬として生物学的製剤が積極的に使用されているが、費用が高いことから開発コストを抑えた低分子治療薬の開発が必要とされている。

低分子シード化合物の供給源として、申請者は天然資源化合物に着目した。天然資源化合物は多種多様な骨格の化合物の供給源であり、また伝承薬として古くから利用されているものもあることから、炎症抑制活性の指標にできると共に、その永い使用歴から安全性についても確立された有用な医薬品探索素材である。その一方で、多様な化合物の集合体として存在していることから、その中から目的に合致した単一の活性化合物を見つけ出すには多くの時間と労力が必要である。

昨今、データサイエンスに基づく研究が盛んに行われている。情報処理技術を基盤に化学領域の課題の解決を目指すケモインフォマティクス研究は比較的新しい研究分野であり、今後の実験化学領域においても発展が期待される分野である。そこで申請者は、情報処理技術を用いて化学領域の課題解決を目指すケモインフォマティクスの考え方を応用し、データサイエンスと天然物化学を組み合わせることで、炎症性腸疾患に有効性を示す新たな医薬品シード化合物の選択的開拓を目指す本研究を計画した。

2. 研究の目的

本研究テーマは、化合物ビッグデータを用いて、種々のデータサイエンスの手法を適用することにより、炎症性腸疾患に有効性を示す新たな医薬品シード化合物の選択的開拓を目指す。また、抗炎症物質に限らず、特定の活性をもつ化合物を選択的に開拓する手法の開発を広く目的とする。

天然物は多様な骨格の化合物が存在していることから、医薬品シーズの供給源として有用である。しかし、優秀な医薬品シーズの発見はセレンディピティに頼る部分が大きく、また、微量成分の入手や活性評価には相応の時間や手間を要する。本研究は、これまでに蓄積されたデータを解析することによりあらかじめ活性物質の構造特性を予測することで、天然物・低分子化合物からシード化合物の早期創出を可能にすることが大きな狙いである。

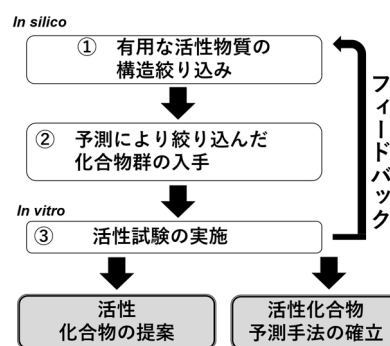


図 1. 研究の流れ

3. 研究の方法

研究目的達成のため、(1) 活性化合物群からのアプローチ (2) 骨格別化合物からのアプローチの二通りの方法で、活性化合物の予測・構造特性検討を行った。

(1) 活性化合物群からのアプローチ

まず「活性化合物群」収集のため、化合物-活性データベースとして ChEMBL を用いて抗炎症活性化合物の情報を集めた。次いで、「多様な骨格の化合物群」として、天然物データベース Dictionary of Natural Product (DNP) および化合物データベース PubChem を用い、天然物の生合成経路に則って化合物情報を収集した。「抗炎症活性化合物群」および「多様な骨格の化合物群」について、alvaDesc (Alvascience, software for molecular descriptors calculation) version 1.0.16, 2019) を用いてそれぞれ構造記述子を算出した。統計解析ソフトウェア JMP® 15 (SAS Institute Inc., Cary, NC, USA) を用いて、化合物を構造記述子によりクラスター解析(k-means 法)に適用した。クラスター分析の結果に基づいて、活性化合物の集約が見られるクラスターの構造特性を検討した。

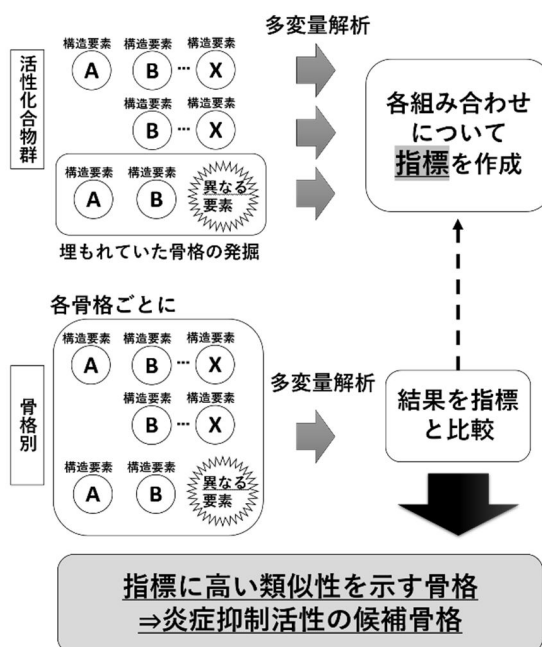


図 2. 研究方法の概略

(2) 化合物骨格からのアプローチ

「トリテルペンの抗 herpes simplex virus type 1 (HSV-1) 活性予測」

化合物骨格からのアプローチとして、トリテルペンに焦点を絞り、それらの抗 HSV-1 活性の有無(強弱)について、機械学習法を用いたクラス分類予測モデルを作成した。本テーマにおいて、抗 HSV-1 活性を対象としたのは、次の理由からである。抗 HSV-1 活性物質の既報は核酸塩基に関するものが大部分を占めており、(1)で示した活性データからのアプローチでは核酸塩基に類似した化合物しか評価ができないと考えられた。抗 HSV-1 活性評価を報告している論文は多いが、それらを包括的にまとめた報告はほとんどなく、ややフラグメンタリーな状況であったことが挙げられる。また、トリテルペンを対象とした理由として、トリテルペンは抗 HSV-1 活性物質を多く含むものの、わずかな構造の違い



• データベース(ChEMBL, DNP) および SciFinder からトリテルペン類の抗 HSV-1 活性のデータを収集

• 化合物⇒構造記述子への変換
• 学習用データセットの作成方法の検討
• 予測手法の選択・調整

• 各種評価指標より予測性能を確認

• ライブラリ由来トリテルペン類の予測結果と 活性試験の結果の比較により予測性能を検証

図 3. 予測モデル作成フローチャート

で大きく活性が変化することが知られており(参考文献¹⁾、人間の目で見ても構造活性相関を判断するのは困難であったことが挙げられる。以上の理由により、トリテルペンの抗 HSV-1 活性についてデータ分析・機械学手法でのアプローチができれば、その科学的意義は大きいと判断した。

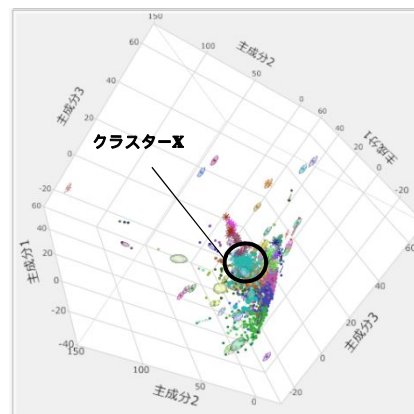
方法として、ChEMBL, DNP に記載された活性情報および SciFinder で検索した既報論文から、トリテルペンの抗 HSV-1 活性の情報を収集した。十分な情報の精査によりデータをクリーニングし、閾値 (IC_{50}) を $25 \mu M$ に設定した。すなわち、閾値よりも IC_{50} 値が小さい化合物を活性強群、 IC_{50} 値が大きい群を活性が弱(なし)群とした。得られたトリテルペンについて、alvaDesc を用いて構造記述子を計算した。得られた構造記述子のうち、欠損値を含むもの、値が一定であるもの、標準偏差が小さいもの、相関係数が大きいものは説明変数としては不適当であると考え、記述子削減を行った。

本データを用いて、Scikit-learn (ver. 0.20.3) の Random Forest 法および Logistic Regression 法により活性有無判別モデルを作成した。モデルはパラメータチューニングにより最適化し、5-fold クロスバリデーションにより評価した。続いて、作成した予測モデルの性能を確認する目的で、京都薬科大学 生薬学分野(中村誠宏 准教授) が所蔵する天然物ライブラリより 25 種類のトリテルペンをご提供いただき、Vero 細胞を用いた抗 HSV-1(HF 株) 活性試験を実施した。得られた活性試験の結果と、予測モデルによる活性有無予測結果を比較検討し、予測モデルの性能を確認した。

4. 研究成果

(1) 活性化化合物群からのアプローチ

はじめに、各種化合物データベースを用いて、「抗炎症活性化化合物の化学構造データセット」および化合物をランダムに収集した「対照用化学構造データセット」を作成した。作成したデータセットについて構造記述子を算出し、得られた構造特性に基づいてクラスタリングを行ったところ、良好なクラスタリング結果は得られなかった。対照用データセットに含まれる化合物に偏りが生じていることが原因と考えられたため、対照用化学構造データセットの作成方法を変更した。天然物の生合成経路による分類に基づいて 9447 種類の化合物を収集し、「対照用化学構造データセット」を作成した。上記データについて、構造記述子を計算し、記述子削減を行ったのち、k-means 法によりクラスタ分析を実施した(図 4)。結果として、クラスタ数を 100 に設定した場合に活性化化合物が比較的高い割合で集約しているクラスタ-X (35 化合物中 19 種類の活性化化合物, 54.3%) を見出した。これらのクラスタに含まれる化合物の構造的特徴として、purine 骨格もしくは 1H-imidazo[4,5-c]pyridine を構造中に持つことが見出された。しかし、全体的に見ると活性化化合物は複数のクラスタに散在しており、1 化合物で 1 クラスタを形成している例が 100 クラスタ中 39 クラスタであったことから、今回検討した条件では解釈が困難であると判断した。この原因として、対象化合物群の化合物数に対して、活性化化合物群の化合物数が少なかったことが考えられる。また、活性化化合物群のケミカルスペースに対して、対象化合物群のケミカルスペースが不釣り合いであったことも原因と思われる。今後、ケミカルスペースを考慮した更なる条件検討が必要と考えられる。



(2) 化合物骨格からのアプローチ

「トリテルペンの抗 herpes simplex virus type 1 (HSV-1) 活性予測」

データベース・既報論文からトリテルペンの抗 HSV-1 活性情報を収集した結果、合計で 472 化合物の活性情報が入手できた。データクリーニングの結果、最終的に 394 化合物がモデル構築に使用された。また、構造記述子を算出し、記述子削減を行った結果、約 700 種類の記述子がモデル構築に用いられた。Random Forest 法を用いた予測モデルについてクロスバリデーションにより性能評価を行ったところ、accuracy = 0.65, AUC = 0.72 程度の結果となった。なお、今回構築した抗 HSV-1 活性予測モデルは、様々な骨格のトリテルペン、サポニン、ステロールについて適用が可能である。

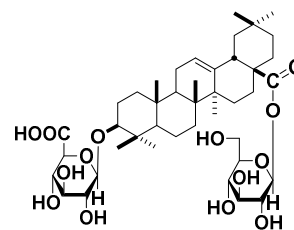


図 5. 抗 HSV-1 活性トリテルペン(1)

続いて、25 種類のトリテルペンについて Vero 細胞を用いた抗 HSV-1 (HF 株) 活性試験を実施した。その結果、*Pfaffia glomerata* 根部から得られた oleanane 型トリテルペン chikusetsusaponin IVa (1)をはじめ複数のトリテルペンが抗 HSV-1 活性を示すことを新たに見出した。(1)の抗 HSV-1 活性: $IC_{50} = 22 \mu M$) また、活性試験の結果と、今回作成した Random Forest 法による予測モデルの結果と比較したところ、83%の正答率で抗 HSV-1 活性の有無を予測することができた。特に、sterol 骨格をもつ化合物に対しては、予測結果が活性試験結果と全て一致しており、高い判別能を持つことが示された。

以上に示した研究成果は 2020 年度日本薬学会第 141 年会 (広島) にて発表しており、現在論文を執筆中である。また、2020 年度 Pacificchem (ハワイ) にて発表予定であったが、大会が延期となったため 2021 年度に発表予定である。

本研究テーマにおける今後の課題として、適用できる化合物のケミカルスペースの拡大化合物の極性、溶解性の問題が考えられる。 について、今回(2)で作成したモデルはトリテルペンに特化した予測モデルである。より汎用性を高めるため、多様な骨格の化合物に適用できる予測モデルの開発が望まれる。(1) で取り組んだ活性化合物からのアプローチではこの問題を解決できる可能性があるが、本手法は化合物の類似性に基づくため、既存の活性化合物報告が特定の骨格の化合物に偏っていた場合は適用が難しい。この点については、更なる方法上の工夫が必要であると考えられる。 極性、溶解性の問題について：今回(2)で作成したモデルでは、通常低極性物質であるトリテルペンと構造中に高極性の糖を持つサポニンを等しく構造記述子により扱っている。しかし、実際の活性試験では、測定溶媒への溶解性が大きく異なると考えられる。また、細胞レベルから動物レベルに引き上げる際、水への溶解性が試験結果に大きく影響することも想定される。したがって、今後の課題としては実際の活性試験を想定した物性の重みづけ等の調整を考慮する必要があると考える。

本研究により、高い判別能を持つ抗 HSV-1 活性予測モデルの構築ができた。この予測モデルは種々のトリテルペン及びその配糖体についても適用が可能である。本予測モデルを用いることによって、抗 HSV-1 活性を持つトリテルペンを事前に想定して成分探索研究や構造誘導化を行うことができ、より効率的な活性物質開拓に繋がることが期待される。

参考文献 1:

Ikeda, T.; Yokomizo, K.; Okawa, M.; Tsuchihashi, R.; Kinjo, J.; Nohara, T.; Uyeda, M. Anti-herpes virus type 1 activity of oleanane-type triterpenoids, *Biological & pharmaceutical bulletin*. **2005**, 28, pp. 1779–1781.

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件/うち国際共著 0件/うちオープンアクセス 0件）

1. 著者名 小川 慶子	4. 巻 57
2. 論文標題 ケモインフォマティクス手法を利用したマラリア予防薬剤の開発戦略	5. 発行年 2021年
3. 雑誌名 ファルマシア	6. 最初と最後の頁 62
掲載論文のDOI（デジタルオブジェクト識別子） 10.14894/faruawpsj.57.1_62	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計1件（うち招待講演 0件/うち国際学会 0件）

1. 発表者名 小川慶子、中村誠宏、笠香織、中村紗子、中嶋聡一、細木るみこ、松田久司
2. 発表標題 天然薬物由来トリテルペン類の抗単純ヘルペスウイルス活性と化学構造特性
3. 学会等名 日本薬学会第141年会（広島）
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
連携研究者	細木 るみこ (Hosoki Rumiko) (90270686)	立命館大学・薬学部・教授 (34315)	
連携研究者	中村 誠宏 (Nakamura Seikou) (20411035)	京都薬科大学・薬学部・准教授 (34306)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------