

令和元年6月24日現在

機関番号：14301

研究種目：挑戦的研究（萌芽）

研究期間：2017～2018

課題番号：17K20043

研究課題名（和文）工業化学物質の受容体結合解析基盤の確立

研究課題名（英文）Construction of a platform for analysis of receptor binding by industrial chemicals

研究代表者

Brown John (Brown, John)

京都大学・医学研究科・講師

研究者番号：90583188

交付決定額（研究期間全体）：（直接経費） 5,000,000円

研究成果の概要（和文）：医薬品を含めて化学物質が複数の受容体を作用することが副作用と炎症反応を起こす原因にある。理想な安全性対策として全化学物質と受容体の相互作用を実験で評価する体制ができれば良いが、費用と材料の限界があるため、網羅的に実験で評価することが困難である。そのため、人工知能を用いて作用を予測し、検証回数を減らす手法が社会的にニーズがある。本研究では、化学物質と細胞の制御に關与する受容体の相互作用に対して、予測モデル開発に挑戦した。結果として核内受容体と代謝系CYP450酵素の化学物質との相互作用を高精度に予測できることを証明した。開発した手法は、規制機関で利用でき、人体への影響のリスク減少にも貢献できる。

研究成果の学術的意義や社会的意義

工場などで排出される空気に複数の化学物質が含まれている。人体に対する安全性を確認するため、化学物質の分子的な作用機序を明確にする必要がある。本研究は細胞制御と代謝に關連するタンパク質に対する化学物質の結合性を予測する手法を開発でき、次世代安全対策に貢献する技術を成立できた。また、開発した手法は、活用が期待される人工知能(AI)をどのように評価と解釈をすれば良いかという課題に大きく貢献し、学術的にも社会的にも意義がある。

研究成果の概要（英文）：Many side effects of drugs and inflammation reactions are the result of drugs and other chemicals interacting with multiple receptors. An ideal safety policy would be to test all chemicals against all receptors, but this is unrealistic due to expenses and limited resources. As a result, there is a societal need to apply machine learning (AI) techniques which can reduce the number of experimental validations required. In this research, we explored the challenge of developing machine learning models for protein families involved in cellular regulation, namely nuclear hormone receptors and metabolic CYP450 enzymes. Our results proved the ability to build highly predictive models for these protein families. The methods developed in the research can be used by regulatory agencies, and can contribute to reduction of risk on the human population.

研究分野：薬理情報学

キーワード：核内受容体 化学物質 人工知能 能動的学習 分子設計 安全評価 可視化

1. 研究開始当初の背景

性ホルモンや脂溶性ビタミンと似た作用を有する化学物質（化合物）があり、環境ホルモンと呼ばれる。この環境ホルモンの出元の一つは、工場の排出物にある。環境ホルモンとその他化学物質の安全性を解析する方法として、生体の受容体との作用機序を網羅的に調べることが望ましい。しかしながら、網羅的な評価は費用と労力に問題がある。そのため、コンピューター上で作用機序を予測することが期待されている。

研究代表者は化合物と受容体（タンパク質）の相互作用をコンピューター上で予測する手法を開発している。膨大なデータをブラックボックス的な人工知能で解析する手法が主流となってきたが、その代わりに、研究代表者は少量のデータと簡単なアルゴリズムでも同等な予測性能を得る手法と以前の研究で開発した。

研究代表者が開発した手法について大量の作用データの場合は予測能力を素早く獲得できることを本研究の開始当初に証明済であったが、性ホルモン等が含まれる受容体はデータがより少なく、この場合も同等な予測能力を獲得できるかが不明であった。また、ある受容体に対して作用機序が学習データに含まれない場合、その受容体を外部予測対象としてどの予測能力が得られるかも不明であった。また、例えば性ホルモン関連受容体に対する予測能力の評価統計が高くても、具体的に何を正確に予測できるのも予測できないのも明確にした方が、行政の関連安全性評価機構に役立つと考えられ、その具体性を満たす開発に取り組むことが望ましいと本研究の開始当初に発想した。

2. 研究の目的

こうした背景に、本研究の目的は次の通りである。

1. 相互作用ビッグデータのある受容体ファミリー（例：酵素）の代わりに、相互作用情報の少ない受容体ファミリー（例：核内受容体）に対しても、高予測性能を有する予測モデルを少量のデータから得られるかを評価する。
2. 相互作用モデルに含まれていない受容体に対しても、モデルが高予測性能を示すかを評価する。
3. 相互作用モデルが動的に予測能力を獲得しながら、具体的にどのように予測能力が向上しているかを明確にする手法を開発する。

3. 研究の方法

本研究は主に研究代表者が開発した「能動的ケモジェノミクス学習」(Future Medicinal Chemistry 2017 9:381)を基に実施した。各目的に対して、次のように研究を実施した。

1. ChEMBL という無償公開生命情報データベースから、相互作用の少ない核内受容体(NHR)とサイトクローム酵素(CYP450)の活性データを取得し整理した。実験に IC50 データを用いた。化学物質がタンパク質を阻害する活性値（濃度）閾値を 500nM、不活性値閾値を 5000nM(0.5uM)に設定した。この閾値を基にデータベースにある化合物とタンパク質の組（ペア）を「有活性」「無活性」に分別した。中間的な活性を示す(5000nM-500nM)のデータは排除した。こうした形で「有活性」と「無活性」の境界が明確となった。データ処理後、能動的ケモジェノミクス学習を繰り返し実施し、各受容体ファミリーが予測可能か否かを評価した。また、予測能力を獲得する速度も評価した。
2. 取得したデータにおいて、一つずつの受容体に対して、総合活性情報データから対象の受容体（と活性データ）を分離し、残っている総合活性データを用いて能動的学習で学習しながら分離した受容体の活性情報を予測できるかを評価した。人工知能を評価する際に用いられる統計の数種類を適用し予測能力を多面的に評価した。
3. 能動的学習の出力と評価統計の性質を解析する新手法（下記の発表論文を参照）を融合して、有活性と無活性の識別力を動的に可視化するプログラムを開発した。このプログラムから得られる画像を参考して能動的学習の進化過程を解析した。

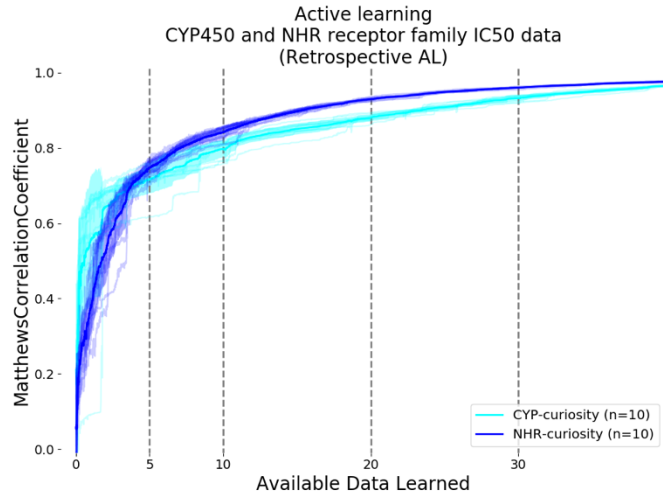
4. 研究成果

1. NHR と CYP450 の ChEMBL データ（2017年7月時点）を処理して、「有活性」と「無活性」に分類した（閾値：500nM 以下または 5uM 以上。中間活性データは排除。）。処理後のデータ密度（あらゆる組み合わせの中からデータが記録された点）は NHR 5%、CYP450 9%となった。能動的学習が両方のデータにおいて高予測性能を獲得できた(MCC 指標)。具体的に、CYP450 においては 9.6%のデータだけでモデルが MCC=0.8 を達成し、NHR においては 7.2%のデータだけで同予測性能を達成した（事頁右図参照）。一方、ランダムでデータを選択してモデルを作成した場合、MCC=0.8 の達成に必要なデータ量がそれぞれ 20.9%と 16.2%であった（発表論文参

照)。この結果から、能動的学習が少ない作用機序のデータでも高予測性能を獲得することができることを確認した。

- 一つずつの受容体の予測性について実験を実施し、結果を薬理的に解釈した。主に得られた結果として、作用機序が他の受容体と類似する受容体は、良い予測性能を示すことと明らかにした。例えば、女性ホルモン（エストロゲン）受容体は α 型と β 型があるが、 β 型とその他の核内受容体

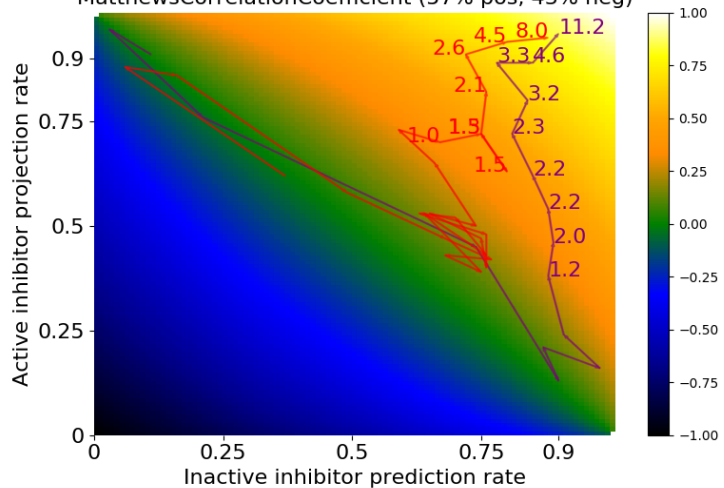
の薬理活性情報を効率よく選択すれば、エストロゲン α 受容体の作用機序を少量のデータで効率よく予測できる。また、この研究の実施により、「ビッグデータ」を学習させるとが予測性能を悪くする場合もあるとのことを明確にした。例えば男性ホルモン（アンドロゲン）受容体の作用機序を学習データから外して、アンドロゲン受容体以外の活性情報を能動的に学習しながらアンドロゲン受容体の作用機序を予測した場合、一定程度まで学習データを増やすと予測能力が向上されるが、それ以上学習させると、逆に外部検証対象となっている男性ホルモン受容体の予測性能がますます悪くなる。主流の考え方を覆すこの発見を報告した論文は査読評価により「VIP論文」と表彰された。



研究成果1 各種受容体ファミリーにおいてタンパク質を阻害する化合物に対して、能動的学習が少量の活性データだけでも阻害活性を示す化合物を識別できることを確認した。

- 予測性能の解釈について、まずは評価に用いる統計の性質を理論的に解釈する新手法を開発した。この手法は陽陰割合を基に、評価統計の定義(数式)からあらゆる結果(真陽率・偽陽率・真陰率・偽陰率)をまとめて、わかりやすい図を算出する。また、この基盤で各種の統計値を得る確率も出力し、応用として、ある予測問題に対して、「この結果を得る確率はどれくらいか」を答えるために開発した基盤を用いて解釈を補助する。この手法

NHR IC50 Active Projection, independent executions
(Values: Percent Data Learned)
MatthewsCorrelationCoefficient (57% pos, 43% neg)



研究成果3：能動的AIが具体的にどのように進化して予測能力を改善するかを分かりやすく描画する新手法を開発した。この新手法では、予測対象の活性・不活性の割合を基にあり得る統計値を描画し、どのデータ量がどの活性・不活性予測率を導いたかを明確にする。

は国内外で広く報道された。そして、代表研究者が単独でこの手法を用いて、能動的学習の進化過程（真陽率・真陰率）を分かりやすく描画する拡張（能動的写像・active projection）の開発にも成功した（前頁右図参照）。現在、「能動的写像」を応用する共同研究依頼を国内企業から受けて共同事業の計画を進めている。

その他に、本研究で開発した基盤を再利用し、動物における化学物質毒性を予測するコンクールにエントリーして良好な結果を発表した。

本課題の研究成果として、人工知能を用いた作用機序予測基盤を向上でき、解釈性と応用性を正確に評価できた。今後の産学連携研究開発に本基盤を適用することを期待できる。

5. 主な発表論文等

[1] (2018) Christin Rakers*, Rifat Ara Najnin, Ahsan Habib Polash, Shunichi Takeda, **J.B. Brown***

Chemogenomic active learning's domain of applicability on small, sparse qHTS matrices: a study using CYP450 and nuclear hormone receptor families

ChemMedChem, 13(6):511-521

[DOI: 10.1002/cmde.201700677](https://doi.org/10.1002/cmde.201700677) -- [PubMed ID: 29211346](https://pubmed.ncbi.nlm.nih.gov/29211346/)

[2] (2018) **J.B. Brown**

Classifiers and their metrics quantified

Molecular Informatics, 37(1-2):1700127

[DOI: 10.1002/minf.201700127](https://doi.org/10.1002/minf.201700127) -- [PubMed ID: 29360259](https://pubmed.ncbi.nlm.nih.gov/29360259/)

[3] (2018) **J.B. Brown**

Adaptive mining and model building of medicinal chemistry data with a multi-metric perspective

Future Medicinal Chemistry, 10(16):1885-1887

[DOI: https://doi.org/10.4155/fmc-2018-0188](https://doi.org/10.4155/fmc-2018-0188) -- [PubMed ID: 29966447](https://pubmed.ncbi.nlm.nih.gov/29966447/)

〔雑誌論文〕（計 3 件）

〔学会発表〕（計 12 件）

〔図書〕（計 1 件）

〔産業財産権〕

○出願状況（計 0 件）

名称：

発明者：

権利者：

種類：

番号：

出願年：

国内外の別：

○取得状況（計 0 件）

名称：

発明者：

権利者：

種類：

番号：

取得年：

国内外の別：

〔その他〕
ホームページ等

6. 研究組織

(1) 研究分担者 無し

研究分担者氏名：

ローマ字氏名：

所属研究機関名：

部局名：

職名：

研究者番号（8桁）：

(2) 研究協力者 無し

研究協力者氏名：

ローマ字氏名：

※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。