

構造活性相関部会・ニュースレター <1 April, 2022>

SAR News No.42

「目次」

///// Perspective/Retrospective ///// ソフトマテリアル分野の計算科学 産業技術総合研究所 青柳 岳司	•••	1
///// Cutting Edge ///// データサイエンスに基づく分子設計・材料設計・プロセス設計の研究紹介 明治大学 金子 弘昌		10
///// SAR Presentation Award ///// 受賞コメントおよび講演要旨		17
///// Activities ///// <報告>		
第 49 回構造活性相関シンポジウム 開催報告 近畿大学 仲西 功		33
<会告> 構造活性フォーラム 2022 会告 明治薬科大学 植沢 芳広		36
編集後記		37

ソフトマテリアル分野の計算科学

(国研) 産業技術総合研究所 青柳岳司

1. はじめに

医薬開発に対する計算科学の利用は長い歴史があり、定量的構造活性相関(QSAR)、ホモロジ ーモデリング、ドッキングスタディ、ケムインフォマティクス、バイオインフォマティクスとい うような様々な手法や、量子化学計算、分子力学(MM)/分子動力学(MD)計算などの理論計算な どが広く用いられてきた。一方、材料設計の分野においても医薬の QSAR に対応して、活性で はなく物性と構造との相関を解析するアプローチは定量的構造物性相関(QSPR)と呼ばれて、原 子団寄与法などをはじめとして古くから用いられている。理論計算に関しても、化学反応解析、 触媒設計などへの量子化学計算の適用は、医薬の分野における酵素反応解析などと同様に、様々 な材料や反応系を対象に取り組まれてきた。本解説では、医薬分野と比較的近いソフトマテリア ルを対象とした計算科学に関する取り組みを紹介する。筆者の限られた知識ではあるが、医薬分 野におけるアプローチとの類似性や違いなどに関して焦点を当てられればと思う。

2. ソフトマテリアルの階層構造

ソフトマテリアルとは合成高分子、液晶、コロイド、生体膜、生体高分子などの柔らかい物質 の総称である。プラスチックなど通常の感覚ではソフトとはいい難いものも、金属やセラミック スなどのハードマテリアルと対比してソフトマテリアルの範疇に含まれる。ソフトマテリアルは、 柔らかいという特徴に加えて分子の集合体であることがハードマテリアルとの違いとして挙げ られる。そして通常ソフトマテリアルにおいては、生体膜、ミセルなどに観察されるように多数 の分子が自己組織化して、より高次の構造が形成される。また合成高分子などにおいては分子構 造自体に多様な分布が存在するのが常である。タンパク質の場合、個々のタンパク質においてア ミノ酸配列はおおよそ固定されており、分子量も均一である。それに応じて3次元構造も、変性 したような場合でない限りおおよそ一義的に決まる。しかし、合成高分子の場合、例えば同じポ リエチレンと呼ばれる高分子材料においても(-CH₂-CH₂-)というモノマー構造が同じであるだけ で、分子量が異なっていても同じポリエチレンと呼ばれる。さらに通常、分子量は均一ではなく、 一つのポリエチレン材料の中にも様々な長さの分子が混ざった状態で分子量分布を持つ。さらに 紛らわしいことに、モノマーが直線的に結合している直鎖分子だけではなく、分岐や架橋した構 造を持つ場合もある。

単一のモノマーユニットからなる合成高分子でもこのような多様な構造が存在するのに加え て、コポリマーといわれる複数種のモノマーが共重合してできる高分子材料はさらに複雑である。 合成高分子的な見方をすると、タンパク質は基本 20 種のアミノ酸、DNA は4 種類の核酸からな るコポリマーということができ、例えばタンパク質であれば、その配列によって個々に区別され ている。しかしながら、合成高分子の場合は同じモノマー種からなるコポリマーは基本ひとまと めにして扱われる。かろうじてモノマーユニットがランダムにつながっているランダムコポリマ ーか、同一のモノマーがある程度の長さのブロックを構成し、そのブロックがつながっているブ ロックコポリマーかという違いはあるが、シーケンスによって基本的に区別されない。そういう 意味では生体高分子に比較して極めてアバウトな取り扱いである。

タンパク質にも一次構造から四次構造までの階層的な構造が存在するが、ソフトマテリアルに おいても分子構造の多様性に加えて、高次構造の多様性が存在する。"マテリアル"は通常分子 集合体であり、非晶、液晶、結晶状態などが存在しうる。また合成高分子の結晶は単結晶である ことはなく、通常、図1に模式図を示すように、複数の分子が集合した状態で結晶領域と非晶領 域が混在している。 さらに図のように、同じ分子鎖でも一部は 折りたたまって結晶を構成し、残りは非晶を 構成するというような場合もあり、タンパク 質でいうところの三次構造と四次構造の境が 明確でない。

このようにソフトマテリアルには複雑な、 あるいは逆の見方をすればアバウトに定義さ れた階層構造が存在しているため、数値とし て明確に定義された構造情報の入力が求めら れる計算科学によるアプローチを複雑にして いる。



3. ソフトマテリアルの計算科学

3.1 構造物性相関

医薬の分野、特に低分子医薬設計における QSAR は 1960 年代の Hansch-Fujita 法に遡るようで ある[1]。Hansch-Fujita 法では分配係数 LogP やハメット定数 σ など分子、あるいは分子に存在す る官能基の持つ特性値と活性との相関を解析する。その後分子構造式そのもの、あるいは3次元 構造や量子化学計算で得られる様々な物理量などが QSAR パラメータとして用いられているの は改めて紹介するまでもないと思う。

一方、医薬以外においても、主に化学工学において QSPR として原子団寄与法が提案されたの は 1955 年に遡るが[2]、普及し出したのは、気液平衡に関する臨界温度、臨界圧力、臨界体積な どを推算した Joback 法[3]の頃からのようである。その後さまざまな改良や新たな手法の提案が 行われ、プロセスシミュレーション等に必要な諸物性値の推算に用いられてきた。

高分子材料の分野においても、モノマーユニットの化学構造から原子団寄与法を用いて材料物 性を予測する試みが行われてきた。高分子材料のQSPRとして古くから広く用いられている方法 として van Krevelen 法[4]がある。この van Krevelen 法を解説した書籍は 1972 年に初版が出版さ れているので、低分子医薬のQSAR が用いられだした時期と大きくは変わらない。表1に van Krevelen 法で推算されている物性値の例をあげる。このように多様な物性値を実験データに基づ いた推算式により推算することができる。

表 1.	van	Krevelen	法により	ノ推算て	できる物性の例	ij
------	-----	----------	------	------	---------	----

密度、熱膨張係数、ガラス転移温度	
結晶融点	
溶解度パラメータ(SP 値)、凝集エネルギー密度	
表面張力	
熱容量、融解エンタルピー	
体積弾性率、ずり弾性率、伸張弾性率、ポアソン比	
引張強度、破断強度	
屈折率、誘電率	等

さらに 1993 年には Bicerano[5]により、あらかじめ規定された原子団を用いることなく、モノ マーユニットの化学構造を構成する元素種と結合情報から、グラフ理論を用いて記述子となるパ ラメータを抽出し、物性推算を行う手法が報告された。このアプローチは後に述べるマテリアル ズインフォマティクスに通じるものであると考えられる。

このような QSPR の手法は、低分子、あるいはモノマーユニットの構造に由来するものであり、 前節で述べたような高分子鎖としての構造や、高次構造の効果は基本的に考慮していない。コポ リマーを扱うような場合でも、基本はモノマーの組成比に応じた加成則により物性値を推算する ので推算精度にはおのずと限界があり、適応できる対象は基本的にシーケンスがランダムな高分 子鎖で、相分離などの高次構造を持たない均一な材料に限定される。とはいえ、タンパク質や DNA を構成するアミノ酸や核酸の組成だけを与えて、その生理的な機能を予測することはほぼ 不可能であるのに対して、合成高分子の場合は限られた範囲ではあるが物性予測が可能であるこ とが特徴である。

3.2 分子シミュレーション

ソフトマテリアルに対する理論計算を 用いたアプローチも、医薬と同様、様々 な目的で利用されてきた。例えば、合成 反応設計に対する量子化学計算の適用な どは、材料においても医薬と同様に広く 行われている。また屈折率、誘電率のよ うな高次構造や分子間相互作用の寄与が 低いような光学的、電気的物性などは、 単分子、あるいは高分子の場合、一部を 切り取ったオリゴマー構造の量子化学計 算によりおおよその物性値を計算するこ とが可能である。図 2 に各種高分子の固 精細は省くが、3 量体程度の部分構 造を用いて密度汎関数法計算により得ら





れる分極率や、コンフォメーション、および Bicerano 法により推算した密度の値などを組み合わせて推算を行っている。

このように、高分子の一部を切り出して計算することにより予測できる物性もあるが、強度、 弾性率のような機械的物性、結晶融点、非晶のガラス転移温度のような熱的物性など、分子集合 体の構造が大きく寄与するような物性を予測するには、できるだけ現実に近い集合体の構造を対 象に計算を行うことが必要になる。このような分子集合体を扱うために古典的力場を用いた MM/MD シミュレーションが行われる。ソフトマテリアルにおいても、用いられる MM/MD の 基本原理は医薬分野と全く同じで、アプリケーションプログラムも GROMACS[7]などの汎用 MD パッケージが用いられる。シミュレーションできる時空間スケールも通常の計算機環境では、空 間スケールが nm オーダー、時間スケールが μ sec オーダー程度であることも変わらない。生体 高分子系でも周期境界条件を考慮して水中の挙動をシミュレーションするようなことが一般的 に行われると思うが、材料の場合は、通常、固体状態あるいは溶液におけるミセルのような分子 集合体の挙動を調べないといけないので、ユニットセルに分子を詰めて、周期境界条件を考慮し た計算を行うのがデフォルトである。高分子鎖の場合、凝集状態において分子がグロビュール状 に各々独立した状態で凝集しているのではなく、コイル状に大きく広がった状態をとり、多数の



図 3. 高分子集合体の周期境界条件イメージ(左図)とポリスチレン非晶の分子モデル(右図)

分子鎖が入り組んだ状態で存在する。通常工業的に用いられる分子量の高分子鎖の広がりは数十 nmのオーダーであるため、現実的にシミュレーションできるユニットセルのサイズより大きい。 よって一本の分子鎖の頭の部分が周期境界条件により自分の尻尾のイメージと相互作用するよ うな場合が起こりうる。図3にそのような高分子集合体の計算モデルを示す。このように自分自 身のイメージと相互作用することによる弊害はもちろんあるが、MDの時間スケールは限られて おり、高分子鎖全体の運動性に対応する時間スケールまでシミュレーションが実行されるわけで はないので、結果的に周期境界条件の弊害が顕著に表れないことが多い。

このような MM/MD シミュレーションを用いて予測される特性の例を表2に挙げる。

表 2. 分子力学/分子動力学により推算できる物性の例

密度、熱膨張係数、ガラス転移温度 結晶融点 溶解度パラメータ(SP値)、凝集エネルギー密度 表面張力 熱容量、熱伝導度 体積弾性率、ずり弾性率、伸張弾性率、ポアソン比 ガス拡散係数 溶解自由エネルギー、吸着自由エネルギー 等

ただし、当然のことながら、現実の材料構造を正しくモデリングできるかどうかにより、得ら れた推算値と実験値との整合性が大きく左右される。図3右図に示すような非晶構造のシミュレ ーションで得られた物性値は、結晶部分を持つような実際の材料の実験値と一致しないのは当然 のことである。またシミュレーションで扱える時間スケールも実際の観察とは大きく異なるので、 注意が必要である。例えば材料を変形した場合の弾性挙動などを予測する際でも、シミュレーシ ョンで可能な変形速度は現実には起こりえないような高速である場合が多いので、それで得られ た弾性挙動をダイレクトに実験と比較してよいかどうかは慎重に判断しなければならない。

3.3 粗視化分子シミュレーション

前節で紹介したような分子シミュレーションは非常に計算コストがかかり、多くの材料が持つ 多相構造を扱うことは現実的には困難である。そこで分子構造を粗視化したモデルを用いた MD シミュレーションがソフトマテリアルの分野でもよく行われる。生体系を取り扱う場合も様々な 粗視化分子モデルが用いられてきているが、粗視化 MD シミュレーションにおいては MARTINI force field(あるいは MARTINI model)[8]がよく知られている。MARTINI ではアミノ酸、核酸、 糖、脂質など様々なパラメータが整備されているが、生体高分子に加えて合成高分子の力場パラ メータも報告されており、材料科学における MARTINI の利用に関しても最近総説が出ている[9]。 総説によると、すでに 50 種の異なる合成高分子の構造ライブラリが提供されているようである。

MARTINIでは、図4に示すように水素を除いた原子4個を基本粗視化ユニットとしている。 一般的に化学構造を反映したボトムアップな粗視化を行おうとした場合、この程度の粗視化が限 界だといわれている。さらに MARTINIを用いた研究は主に分子鎖の3次元構造を再現すること を目的としている例が多く、材料分野のシミュレーションの目的として重要な、物性予測を行っ た事例は少ない。

一方、高分子物理の分野においては、より粗視化の進んだモデルを用いて、からみあいとそれ に起因する粘弾性のような高分子の普遍的な構造と特性に関する研究が進められてきた。その一 例がビーズースプリングモデルである。このビーズースプリングモデルはその名の通り、粗視化 ユニットであるビーズがバネで結合されて高分子鎖を表している。ビーズ間の相互作用やバネの 関数形によっていくつかのモデルがあるが、現在でも最もよく使われているのは Kremer と Grest が提案した[10]、ビーズ間の相互作用に Lennard-Jones ポテンシャルを用い、バネとして有限伸 び切り効果を持つ FENE ポテンシャルを用いたもので、Kremer-Grest モデルとして高分子シミュ レーションの分野では広く知られたモデルである。



Figure 1. Martini mapping examples of selected molecules. A) Standard water particle representing four water molecules; B) the organic solvent toluene; C) dimyristoylphosphatidylcholine (DMPC) lipid; D) poly(ethylene oxide) (PEO); E) the 1-octyl-3-methylimidazolium tetrafluoroborate $[C_8mim]^+[BF_4]^-$ ionic liquid; F) poly(3-hexylthiophene) (P3HT); G) the 1,3,5-benzenetricarboxamide (BTA) self-assembling molecule; H) C_{60} fullerene; I) the surface of graphene. Martini CG beads are shown as cyan transparent beads overlaying the atomistic structure.

図 4:MARTINI の粗視化マッピングの例。参考文献[9]、Figure1 より引用 (Licensed under CC BY 4.0)

図 5 に分子モデルとビーズースプリングモデルの対応のイメージを示すが、このように高度 に粗視化されたモデルの場合、化学構造を正しく反映するような粗視化は困難で、からみ点間分 子量のような実験とシミュレーション両者で観察される物理量に基づいてスケーリングを行う というトップダウン的なアプローチにより実材料とのマッピングを行う。例えば粘弾性測定によ るポリスチレンのからみ点間分子量 Me が 18,000、ビーズースプリングモデルの粗視化 MD シミ ュレーションにより得られたからみ点間ビーズ数 Ne が 35 であった場合、1 ビーズはポリスチレ ン分子量 514、約5 モノマーユニットに相当すると考える(注:1990年のオリジナルの論文では Ne≒35 と見積もっており、ここではその当時のスケールを紹介したが、その後の検討[11]によ り Ne の値は 70-80 程度の大きな値であると報告されている)。



図 5:分子モデル(ポリスチレン)とビーズースプリングモデルの対応例

このビーズ-スプリングモデルは元々溶融高分子の構造とダイナミクスを対象としたモデル であるが、相互作用パラメータ等を調整することにより、ガラス状態を取り扱うことも可能であ る。筆者ら[12][13]は、熱可塑性エラストマーモデルとして、ミクロ相分離構造を形成するガラ ス状ブロックとゴム状ブロックからなるブロックコポリマーのモデリングを行い、エラストマー の弾性挙動予測を行った。図 6 にシミュレーションの概要を示すが、自己無撞着場理論(SCFT) から得られたブロックコポリマーのミクロ相分離構造に基づいてモデリングした粗視化分子モ デルを、実験と同様に粗視化 MD シミュレーションにおいて伸長しながら、応力-ひずみ挙動 を解析する。

化学構造を正確に反映することが困難なトップダウン型の粗視化モデルを用いる以上、得られ る物性を実験値と定量的に合わせることは難しいが、ソフトマテリアルのデザイン、特にコスト 面が重要になる工業的には、同じ化学構造を用いて組成やシーケンスを設計することにより、物 性を制御することがよく行われるので、このような定性的なシミュレーションで傾向を示すこと でも、材料開発を進めていく上で有用な知見となる。



図 6 熱可塑性エラストマーモデルのシミュレーションの流れ(上図)と得られた応カーひずみ曲 線(下図)の例(トリブロックコポリマーとジブロックコポリマーの比較)

3.4 マテリアルズインフォマティクス

直接医薬開発に関わっていない筆者から見ると、医薬の分野においては、Hansch-Fujita 法の時 代からケムインフォマティクス、バイオインフォマティクスまで、情報科学的手法は継続的に利 用され、大きく発展して来たように見える。一方マテリアルの分野では、先に紹介した QSPR は、材料構造の多階層性や求められる特性の多様性のため、限られた用途、範囲でしか用いられ ず、永らくの間それほど目立った進歩がなかった。しかしながら 2011 年、当時のオバマ米大統 領が"Materials Genome Initiative"[14]という、情報技術や計算科学を駆使し、優れた新材料の発見 から実用化までのスピードを2倍に早める計画を立ち上げたことを発端に、材料設計の分野にお けるインフォマティクスの活用、マテリアルズインフォマティクス(MI)が脚光を浴びるようにな ってきた。

MI は当初、無機材料の分野で発展してきたが、最近ではソフトマテリアル、特に高分子材料 への MI の適用が期待されている[15][16]。高分子 MI の基本的なアプローチは、ケムインフォマ ティクスと同様、分子構造(高分子の場合モノマーの化学構造)に基づいて、様々な記述子を抽 出し、機械学習を用いて実験あるいは理論計算で得られる物性値との回帰を行うことにより、物 性推算システムを構築する。しかしソフトマテリアルの場合、先に述べているように分子構造の みで決定される特性は多くなく、たとえ高度な記述子や機械学習の手法を適用しても、基本的に 従来の QSPR の域を出ない。古典的な QSPR を超えた精度、実用性を得るためには、分子構造に 加えて、材料の加工時のプロセス条件、それにより生成される高次構造まで考慮した物性予測、 すなわちプロセスー構造-物性相関が重要である。ただ、高次構造の観察には電子顕微鏡等限ら れた手法によるしかなく、機械学習が必要とする大量のデータを用意することは困難である。そ こで先に紹介したような粗視化シミュレーションにより得られる高次構造を用いて MIを進めて いくというアプローチが期待されている。

高次構造を考慮した MI の事例として、筆者らは前節で紹介した熱可塑性エラストマーモデルの粗視化 MD シミュレーションのデータに基づいて、深層学習を用いて 3 次元相分離構造と応 カーひずみ曲線の回帰を行った[17]。実際の材料においては、規則正しい最安定な相分離構造が 得られるとは限らず、非平衡な状態で構造が凍結する場合が多い。図 7 に検討の流れを、図 8 にシミュレーションで求めた応力-ひずみ曲線と相分離構造から予測した結果を示す。このよう に高次構造を加味した深層学習により、シミュレーションより高速で、かつこれまでの QSPR を超えた物性予測が可能になる。



図7 深層学習(3D-CNN)による相分離構造からの応カーひずみ曲線の予測の流れ



図 8 検証サンプルにおける応カーひずみ曲線、(青)粗視化 MD、(赤)深層学習による予測。 横軸にひずみ、縦軸に応力を示す。

4. おわりに

ソフトマテリアル分野の計算科学に関して、限られた範囲であるが筆者の知見に基づいて紹介 させていただいた。材料開発においても基本的なアプローチは医薬開発と同様で、QSP(A)R、理 論計算、インフォマティクスなどの手法が場合に応じて利用されてきている。ここでは触れなか ったが、実験においてもコンビナトリアルケミストリー、ハイスループットスクリーニング、自 動実験などのアプローチも一部の材料開発で利用されている。材料側から見ると医薬開発におけ る計算科学の利用は、"薬理活性を有する化合物の探索"というようにターゲットが明快で、取 り組みやすいように見える。しかしながら、薬理活性の高い化合物が見つかったからと言って、 医薬品として世に出すためには、毒性、副作用、安定性、合成コストなど様々な壁があり、簡単 なものではないというのは医薬業界に籍を置かないものにとっても自明である。さらに分子設計 が出来ても、製剤に進む際には様々な高次構造の制御が必要になってくるというのも、ソフトマ テリアルと共通する部分である。

それでも、医薬開発において計算科学が一定の地位を占めているのは、薬理活性のように「そ れがないと薬にならない」という、ここ一番の要所におけるスクリーニングに使うという位置づ けが確立されているからではないかと考える。一方、材料の分野では様々な素材と用途があり、 開発のスキームや抑えるべき要所も多岐にわたるため、計算科学の利用法が確立されておらず、 「わかる人にしか使えない」、「どこから手を付けていいのかよくわからない」状態が多くみられ る。材料の多様性を考えると一朝一夕にブレークスルーが起きるとは考えにくいが、材料分野の 計算科学の利用に関しても、要所を抑えた効率的な利用法が普及していくことが望まれる。一方、 医薬開発においても、製剤などソフトマテリアルの計算科学が貢献できる分野が多々あると思わ れるので、医薬-マテリアル間の交流が今後ますます進むことを期待する。

謝辞

本稿で紹介した事例の一部は、JSPS 科研費 新学術領域研究 「次世代物質探索のための離散 幾何学」17H06464の助成を受けたものです。

参考文献

- Hansch, C., Fujita T. ρ-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure, J. Am. Chem. Soc., 86, 1616-1626 (1964). doi: 10.1021/ja01062a035
- [2] Lydersen, A. L. Estimation of Critical Properties of Organic Compounds, Univ. Wisconsin Coll. Eng., Eng. Exp. Stn. rept. 3, (1955).
- [3] Joback, K. G., Reid, R. C. Estimation of Pure-Component Properties from Group-Contributions, *Chem. Eng. Commun.*, 57, 233-243 (1987). doi: 10.1080/00986448708960487
- [4] van Krevelen, D. W. te Nijenhuis K. Properties of Polymers, 4th completely revised ed., Elsevier (2009).
- [5] Bicerano, J. Prediction of Polymer Properties, 3rd ed., CRC press (2002).
- [6] Tanaka, T, Shirai, H., Aoyagi, T. Proceedings of PPS-22: 22nd annual meeting of the Polymer Processing Society, Yamagata, Japan (2006).
- [7] <u>https://www.gromacs.org/</u>
- [8] Marrink, S. J., Tieleman D. P. Perspective on the Martini model, *Chem. Soc. Rev.*, 42, 6801-6822 (2013). doi: 10.1039/c3cs60093a
- [9] Alessandri. R., Grünewald, F., Marrink, S. J. The Martini Model in Materials Science, Adv. Mater., 33, 2008635 (2021). doi: 10.1002/adma.202008635
- [10] Kremer K., Grest, G. S. Dynamics of entangled linear polymer melts: A molecular-dynamics simulation, J. Chem. Phys., 92, 5057-5086 (1990). doi: 10.1063/1.458541
- [11] Likhtman, A. E., Sukumaran, S. K., Ramirez, J. Linear Viscoelasticity from Molecular Dynamics Simulation of Entangled Polymers, *Macromolecules*, 40, 6748–6757 (2007). doi: 10.1021/ma070843b
- [12] Aoyagi, T., Honda, T., Doi, M. Microstructural study of mechanical properties of the ABA triblock copolymer using self-consistent field and molecular dynamics, J. Chem. Phys., 117, 8153-8161 (2002). doi: 10.1063/1.1510728

- [13] Aoyagi T. Coarse-grained molecular dynamics study of elasticity of block copolymers with cubic symmetrical morphology, *Polymer*, **243**, 124624 (2022). doi: 10.1016/j.polymer.2022.124624
- [14] https://www.mgi.gov/
- [15] Audus D. J., de Pablo, J. J. Polymer Informatics: Opportunities and Challenges, *ACS Macro Lett.*, **6**, 1078-1082 (2017). doi: 10.1021/acsmacrolett.7b00228
- [16] Chen, L. et al., Polymer informatics: Current status and critical next steps, Mater. Sci. Eng. R: Reports, 144, 100595 (2021). doi: 10.1016/j.mser.2020.100595
- [17] Aoyagi, T. High-throughput prediction of stress-strain curves of thermoplastic elastomer model block copolymers by combining hierarchical simulation and deep learning, *MRS Advances*, 6, 32-36 (2021). doi: 10.1557/s43580-021-00008-1

///// Cutting Edge /////

データサイエンスに基づく分子設計・材料設計・プロセス設計の 研究紹介

明治大学 金子弘昌

1. はじめに

医薬品をはじめとする高機能性材料を研究・開発・製造する際、化学データや化学工学データ を活用して機械学習を行い、分子設計・材料設計・プロセス設計・プロセス管理を効率化するこ とが一般的となっている。分子設計では、化合物の化学構造・分子構造を数値化した分子記述子 x と活性・物性・特性 y との間で数理モデル y = f(x) を構築する。構築されたモデルに基づい て、新たな化学構造に対する活性・物性・特性の値を予測したり、活性・物性・特性が所望の値 となる化学構造を設計したりする。

材料設計では、材料の物性・活性・特性 y と材料の合成条件や製造条件 x との間で数理モデ ル y = f(x) を構築する。モデルに基づいて、実験前もしくは製造前に材料の合成条件や製造条 件から材料の物性・活性・特性を予測したり、目標の材料を達成するための合成条件や製造条件 を設計したりする。

プロセス設計では、目標の性能をもつ装置やプラントを設計するため、プロセスシミュレーションもしくは実験を行う際のプロセス条件 x と装置・プラントの性能 y との間で数理モデル y = f(x) を構築することで、装置やプラントの目標の性能を達成するためのプロセス条件を効率的に設計できる。

プロセス管理では、温度や圧力などの簡単に測定可能なプロセス変数 x と製品品質を代表す る濃度・密度などの測定が困難なプロセス変数 y との間で数理モデル y = f(x) を構築する。モ デルを用いることで、簡単に測定可能なプロセス変数の値から、測定が困難なプロセス変数の値 をリアルタイムかつ連続的に推定できる。推定値を実測値の代わりに使用することで効率的なプ ロセス制御を達成できる。

分子設計・材料設計における研究・開発はケモインフォマティクスやマテリアルズインフォマ ティクス、プロセス設計やプロセス管理における研究・開発はプロセスインフォマティクスと呼 ばれ、それぞれ活発な議論がなされている。それぞれの分野において、説明変数 x と目的変数 y との間で予測精度の高い数理モデル y = f(x) を構築することが重要である。モデル構築手法と して、線形手法には partial least squares regression (PLS) [1], リッジ回帰, least absolute shrinkage and selection operator, elastic net [2] があり、非線形手法には、サポートベクター回帰 (support vector regression, SVR) [3], 決定木 [4], ランダムフォレスト [5], gradient boosting [6], deep neural network (DNN) [7] がある。サンプル数が非常に大きい場合は DNN のようなフィッティングパ ラメータの多い手法が有効である一方で、サンプル数が小さい場合は、サンプルに含まれるノイ ズにもモデルが適合してしまう過学習に注意してモデルを構築しなければならない。データセッ トごとにモデルの予測性能を適切に評価して、過学習のない予測精度の高いモデル構築手法を選 択する必要がある。

モデルを運用して分子設計・材料設計・プロセス設計・プロセス管理をする際は、モデルの適 用範囲 (Applicability Domain, AD) [8] を考慮する必要がある。AD とはモデルが本来の予測性能 を発揮できる x のデータ領域のことである。新しいサンプルの x の値が AD 内であれば y の 予測値を信頼できるが、AD 外であれば信頼できない。与えられたデータセットに基づいて AD を設定し、新しいサンプルの y の値を予測する際は AD 内か AD 外か検討する。本誌では、 この AD に関連して、モデルが AD 内の新しいサンプルを予測する際の誤差を推定する研究例 を紹介する。

AD を考慮することで的確な予測を達成できる一方で、分子設計・材料設計・プロセス設計に 必要なことは、物性・活性・特性 y の目標値からそれを実現するための特徴量(合成条件・プ ロセス条件など)x を導くことである。これをモデルの逆解析と呼ぶ。ただ一般的な逆解析で行われていることは、x の仮想サンプルをコンピュータで大量に生成し、それらをモデルに入力して y の値を予測し、予測値が良好な仮想サンプルを選択する、すなわち順解析を網羅的に繰り 返す擬似的な逆解析にすぎない。これでは人が事前に想定した x の探索範囲における y を予測 することにすぎず、当初想定しない条件でこそ発現する新機能の探索には全く対応できない。

著者の研究室では、y の値から x の値を直接的に予測する、すなわち数理モデルを直接的に 逆解析する手法「直接的逆解析法」を開発している [9,10]。数理モデルを解析することで、y の 目標値から直接 x の値を自由自在に予測できる。さらに直接的逆解析法では y が複数、すなわ ち物性・活性・特性が複数存在する場合でも、すべての物性の目標値を満たす x の値を提案で きる。直接的逆解析法を実行するための無料の Python ライブラリもある [11]。本誌では、直接 的逆解析法を活用した分子設計および材料設計に関する研究例を紹介する。

モデルの逆解析により合成条件 x が提案された後、実際に合成して物性・活性・特性 y の値 を測定する。測定値が目標を満たしていれば開発は終了となるが、目標を満たしていなければ、 そのデータをデータセットに追加してモデルを再構築し、次の x を提案する。このように実験・ モデル構築・モデルの逆解析を繰り返すことを適応的実験計画法と呼ぶ。特に y の目標値が高 い場合にはベイズ最適化 (Bayesian Optimization, BO) [12] により効率的に目標までの合成条件 を探索できる。BO は、ガウス過程回帰モデルにより得られる y の予測値とその分散により計 算される獲得関数を指標にして、内挿だけでなく外挿も探索することで効率的な適応的実験計画 法を達成しようとする手法である。さらに近年、上述した直接的逆解析法により BO を凌ぐ合 成条件探索の効率化が達成されたため、本誌の最後に紹介する。

2. 新しいサンプルにおける予測誤差の推定

数理モデルを用いて AD 内の新しいサンプルの y の値を予測する際に、予測誤差を推定する 手法 [13] を紹介する。AD 内のサンプルであるため、モデル構築用データにおける推定精度と 同程度の精度で予測できるはずであるが、偶然の相関によりモデル構築用データにおけるモデル の推定性能は高く見積もられてしまう。偶然の相関とは、本来は特徴量の間の相関はない、もし くは相関が弱くても、実際のサンプル群において、ノイズ等の影響により偶然に相関が出たり、 強くなったりする現象である。偶然の相関の影響がモデルに反映されることが、モデルの過学習 を意味する。サンプル数が小さいほど、そして x の数が大きいほど偶然の相関が起こりやすく、 モデルが過学習しやすい。

モデルの予測精度を見積もるためには偶然の相関の影響を考慮する必要がある。偶然の相関の 程度を推定するため y-randomization (もしくは y-scrambling) に着目した。y-randomization とは、 モデル構築用データにおいて、x の値はそのままで y の値のみサンプル間でシャッフルし、モ デル構築およびモデル構築用データの推定を行う手法である。y のみシャッフルしていることか ら x と y の間の相関関係はないと考えられ、y-randomization では y の推定誤差が大きくなる はずであるが、特にサンプル数が小さく x の数が大きい場合には、偶然の相関により y の推定 誤差が小さくなってしまう。実際のモデル構築時の推定誤差が小さくても、y-randomization で の推定誤差も小さければ、そのデータセットは偶然の相関が起こりやすいといえ、モデル構築用 データを用いて見積もられる推定性能が本来の性能より高く評価されてしまう。

偶然の相関の影響を除外してモデルの予測性能を評価する指標が chance correlation-excluded MAE (MAE_{CCE}) [13] である。MAE (Mean Absolute Error) は y の誤差の絶対値の平均であり、MAE_{CCE} は以下の式で表される。

$$MAE_{CCE} = MAE_{TRAIN} + (MAE_{MEAN} - MAE_{yRAND})$$
(1)

ここで、MAE_{TRAIN} はモデル構築用データにおける MAE、MAE_{MEAN} は y の推定値をモデル構 築用データの y の平均値としたときの MAE、MAE_{yRAND} は y-randomization 時の MAE である。 (MAE_{MEAN} - MAE_{yRAND}) により偶然の相関による MAE の大きさが分かり、それを MAE_{TRAIN} に足し合わせることで、新しいサンプルに対する MAE を推定する。偶然の相関の影響がない ときは、MAE_{yRAND} は MAE_{MEAN} と等しくなり、MAE_{CCE} は MAE_{TRAIN} と一致する。 y-randomization は複数回実施されるため、MAE_{CCE} も複数の値が得られる。これらの値の分布が 新しいサンプルを予測したときの MAE の分布を表すと考えられる。

水への溶解度が測定された 1290 個の化合物のデータセット logS [14] と、環境毒性として pIGC₅₀ が測定された 1213 個の化合物のデータセット Tox [15] を用いて、新しいサンプルにお ける予測誤差の推定に関する検証を行った。分子記述子は RDKit [16] により計算された 2 次元 記述子である。モデル構築用データのサンプル数を 30 として、残りをモデル検証用データとし た。モデル構築手法は PLS である。

MAE_{CCE}のヒストグラムと、テストデータにおける実際の MAE (MAE_{TEST})を図 1 に示す。 MAE_{TEST} はモデル検証用データにおける MAE である。MAE_{CCE} が分布として推定され、実際 の MAE_{TEST} が分布の中心付近であることがわかる。MAE_{CCE} により新しいサンプルの予測誤差 を適切に推定できることを確認した。



MAE_{CCE} (histogram), MAE_{TEST} (red point) MAE_{CCE} (histogram), MAE_{TEST} (red point) (a) logS (b) Tox

図 1. MAEcce のヒストグラム [13]。MAEtest はモデル検証用データにおける MAE を意味 する

3. 直接的逆解析法による分子設計

物性・活性・特性 y の目標値から、分子記述子や合成条件・製造条件 x を直接的に予測でき る直接的逆解析法が可能な回帰分析手法は、Gaussian mixture regression (GMR) [10] および Generative topographic mapping regression (GTMR) [17] である。GMR や GTMR により構築され たモデルでは、x と y の間のすべての関係、すなわち同時確率分布が複数の正規分布の重ね合 わせ (混合正規分布) で表現される。この同時確率分布に基づいて、確率の乗法定理とベイズの 定理により、y の値が与えられたときの事後確率分布である x の確率分布を獲得でき、この分 布も混合正規分布で表現される。x の確率分布が高くなる値が y の値をモデルに入力したとき の x の予測値となり、直接的逆解析が達成される。GMR や GTMR は y が複数の場合も対応 でき、GTMR ではデータの可視化もモデルの順解析もしくは逆解析と同時に達成可能である。 なお GMR におけるパラメータ最適化に、expectation-maximization アルゴリズムだけでなく変 分ベイズ法を用いることで、予測精度の向上が達成された [10]。x に制約条件が存在したとし てもそれを考慮した x の予測が可能である [18]。

今回は GTMR を実データに応用した解析結果を紹介する。使用したデータセットは、水への 溶解度を S [mol/L] としたときの log(S) が測定された 1290 個の化合物のデータセット [14] である。重複している 16 化合物は削除した。化合物の分子記述子 200 個を RDKit [16] を用い て計算した。865 化合物をモデル構築用データ、409 化合物をモデル検証用データとした。同じ 値を持つサンプルの割合が 0.8 を超える記述子を除去した後、遺伝的アルゴリズムに基づく PLS を用いて 10 個の記述子を選択した。

GTMR によりデータの可視化をした結果を図 2 に示す。化合物が二次元平面全体に分散していることがわかる。また二次元平面上において近くに位置している化合物は logS の値も近い化 合物であり、logS の値が大きい領域や小さい領域に分かれている。GTMR により化合物群を良好に可視化できた。

GTMR と比較のため用いたガウシアンカーネルによる SVR により、モデル構築用データを 用いてモデルを構築し、モデル検証用データの logS を推定した結果を図 3 に示す。GTMR と比 較して SVR の方が、化合物が対角線付近に固まっており、GTMR と比較して SVR の推定性能 が高かった。図 3(b)では対角線から外れるサンプルも散見されたが、GTMR により全体的に logS の値を推定できた。

続いて logS の目標値を比較的水に溶けやすい化合物がもつ値である 1.6 として、全化合物 を用いて構築した GTMR モデルを逆解析した。y = 1.6 としたときの x (分子記述子)の値が得 られ、その値を用いて化合物探索を行った。探索に用いた化合物セットは、Enamine 社 [18] か ら提供され、Initiative for Parallel Bioinformatics (IPAB) [19] によって編集された約 250 万化合物 である。モデルの逆解析によって推定された記述子セットとユークリッド距離の近い化合物を選 択した。選択された化合物の一例を図4に示す。ヒドロキシ基やフッ素基などの官能基をもつ、 極性化合物が選択されていることがわかる。GTMR によって適切にモデルの逆解析に基づく化 合物探索を行えることを確認した。



図 2. GTMR によるデータの可視化結果 [17]。カラーバーは logS の値に対応する





図 4. logS の目標値を 1.6 とした際の GTMR による直接的逆解析の結果 [17]

4. 分子設計・材料設計・プロセス設計のための直接的逆解析法による適応的実験計画

直接的逆解析法によりベイズ最適化を凌ぐ実験条件探索の効率化が達成された例 [20] を紹介する。合成条件 x と物性・活性・特性 y の間の複雑な非線形関数を、x で合成して y の結果が得られる系と仮定して、適応的実験計画法を実施する。非線形関数は Rastrigin 関数 [21] に 基づく以下の式を使用した。

$$y = -10d + \sum_{i=1}^{d} \left(10\cos(2\pi x_i) - x_i^2 \right)$$
(2)

ここで d は x の数、 x_i は i 番目の x を表す。d=5,10,15 として、それぞれにおいて設定 した y の目標値を越える x の候補を探索することを目的とする。

30 サンプルを最初のデータセットとして、ランダムにサンプルを選択 (Random)、BO により サンプルを選択、GMR に基づく直接的逆解析法のいずれかにより、次のサンプルを一つ選択す る。GMR では、数理モデルとして x と y の間のすべての関係を複数の正規分布の重ね合わせ で表現することで、ベイズの定理と確率の乗法定理から、y が与えられた際の x の事後確率分 布を計算できる。これにより y の目標値から直接 x の値を瞬時に予測できる。

それぞれの方法で選択されたサンプルの x の値から、式(2)で y の値を計算する。この値が y の目標値を超えた場合は終了するが、超えていない場合は再度それぞれの探索手法で次のサンプルを一つ選択する。これを y の目標値を超えるまで繰り返す。ただし、繰り返し回数が 100 回 に到達したら (合計の実験回数が 130 になったら)、強制的に終了とする。

yの目標値を超える x の探索にかかった手法ごとの実験回数の平均値を表1 に示す。Random では目標を達成できない状況において、BO と比較して大幅に少ない実験回数で、GMR により y の目標値を達成した。特に BO では x が多くなると実験回数も大きくなってしまったが、 GMR では少ない実験回数で効率的に目標まで到達できた。直接的逆解析法により効果的な適応 的実験計画が可能となることを確認した。

各手法における、実験ごとの y の値の推移の例を図 5 に示す。BO では y の値が全く向上 しない状況において、GMR では実験ごとに y の値が向上し、さらに最後の実験において既存 の実験データにおける y の値を超越する y の値を達成した。GMR により、適切に外挿領域の サンプルを探索できること、および既存の y の値を超越する実験結果を獲得することが可能で あることを確認した。



表1. 手法ごとの目標到達までの実験回数 [20]

- 14 -

5. おわりに

今回はケモインフォマティクス・マテリアルズインフォマティクス・プロセスインフォマティ クスの研究例として、新しいサンプルにおける予測誤差の推定、直接的逆解析法による分子設計、 分子・材料・プロセスの設計のための直接的逆解析法による適応的実験計画について紹介した。 同じ研究分野において、モデルの予測精度向上のための新たな記述子の研究 [22]、ポリマー材 料 [23] や熱電変換材料 [24] や金属有機構造体 [25] をはじめとする様々な高機能性材料の設 計、所望の性能を満たす化学プラントの設計 [26]、プロセスの推定制御のためのソフトセンサ - [27,28]、スペクトル解析 [29] など活発な議論がなされている。

著者は協力者と一緒に、本誌でも説明した化学分野で有用なデータ解析・機械学習に基づいた 分子設計・材料設計・プロセス設計をプログラミングなしに容易な操作で実施できるウェブサー ビスを開発中である (2022 年春リリース予定)。データアクセス・化学構造生成・実験計画法・ 分子記述子計算・データの可視化・データの前処理・回帰分析・適応的実験計画法・ベイズ最適 化・時系列データ解析・モデル最適化・直接的逆解析の機能がすべて搭載されている。興味のあ る方は著者にご連絡いただきたい。化学・化学工学データと機械学習を組み合わせることで、分 子設計・材料設計・プロセス設計・プロセス管理が効率化することを期待する。

参考文献

- Wold, S., Sjöström, M., Eriksson, L. PLS-regression: a basic tool of chemometrics, *Chemom. Intell. Lab. Syst.*, 58, 109–130 (2001).
- [2] Li, Z. T., Sillanpaa, M. J. Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection, *Theor: Appl. Genet.*, **125**, 419–435 (2012).
- [3] Bishop, C. M. Pattern recognition and machine learning, Springer, New York (2006).
- [4] Bruce, C. L., Melville, J. L., Pickett, S. D., et al. Contemporary QSAR classifiers compared, J. Chem. Inf. Model., 47, 219–227 (2007).
- [5] Palmer, D. S., O'Boyle, N. M., Glen, R. C., et al. Random forest models to predict aqueous solubility, J. Chem. Inf. Model., 47, 150–158 (2007).
- [6] Voyant, C., Notton, G., Kalogirou, S., et al. Machine learning methods for solar radiation forecasting: A review, *Renew. Energ.*, 105, 569–582 (2017).
- [7] Goh, G. B., Hodas, N. O., Vishnu, A. Deep learning for computational chemistry, J. Comput. Chem., 38, 1291–1307 (2017).
- [8] 金子弘昌, 化学のための Python によるデータ解析・機械学習入門, オーム社 (2019).
- [9] Shimizu, N., Kaneko, H. Direct inverse analysis based on Gaussian mixture regression for multiple objective variables in material design, *Mat. Des.*, **196**, 109168 (2020).
- [10] Kaneko, H. Extended Gaussian mixture regression for forward and inverse analysis, *Chemom. Intell. Lab. Syst.*, 213, 104325 (2021).
- [11] https://datachemeng.com/dcekit/
- [12] 金子弘昌, Python で学ぶ実験計画法入門 ベイズ最適化によるデータ解析, 講談社 (2021).
- [13]Kaneko, H. Estimation of predictive performance for test data in applicability domains Using y-randomization, J. Chemom., 33, e3171 (2019).
- [14] Hou, T. J., Xia, K., Zhang, W. et al. ADME evaluation in drug discovery. 4. prediction of aqueous solubility based on atom contribution approach, *J. Chem. Inf. Comput. Sci.*, 44, 266–275 (2004).
- [15] http://www.cadaster.eu/node/65.html
- [16] https://www.rdkit.org/
- [17] Kaneko, H. Data Visualization, regression, applicability domains and inverse analysis based on generative topographic mapping, Mol. Inf., **38**, 1800088 (2019).
- [18] http://www.enamine.net/
- [19] http://www.ipab.org/initiative-for-parallel-bioformatics-ipab?set_language=en
- [20] Kaneko, H. Adaptive design of experiments based on Gaussian mixture regression, *Chemom. Intell. Lab. Syst.*, **208**, 104226 (2021).
- [21] https://www.sfu.ca/~ssurjano/rastr.html
- [22] Nitta F., Kaneko, H. Two- and three-dimensional quantitative structure-activity relationship models based on conformer structures, *Mol. Inf.*, **40**, 2000123 (2021).

- [23] Takano, S., Kaneko, H. Monomer design of polymer materials with high refractive index and high glass transition temperature, *J. Comput. Chem. Jpn.*, **18**, 115-121 (2019).
- [24] Yoshihama, H., Kaneko, H. Design of thermoelectric materials with high electrical conductivity, high Seebeck coefficient, and low thermal conductivity, *Anal. Sci. Adv.*, **2**, 289-294 (2021).
- [25] Yuyama, S., Kaneko, H. Correlation between the metal and organic components, structure property, and gas-adsorption capacity of metal-organic frameworks, J. Chem. Inf. Model., 61, 5785-5792 (2021).
- [26] Iwama, R., Kaneko, H. Design of ethylene oxide production process based on adaptive design of experiments and Bayesian optimization, J. Adv. Man. Proc., **3**, e10085 (2021).
- [27] Kaneko, H. Estimating the Reliability of Predictions in Locally Weighted Partial Least-Squares Modeling, J. Chemom., 35, e3364 (2021).
- [28]Kaneko, H. Beware of r^2 even for test datasets: using the latest measured y-values (r^2_{LM}) in time series data analysis, *J. Chemom.*, **33**, e3093 (2019).
- [29] Kaneko, H., Kono, S., Nojima, A., et al. Transfer learning and wavelength selection method in NIR spectroscopy to predict glucose and lactate concentrations in culture media using VIP-Boruta, *Anal. Sci. Adv.*, 2, 470-479, (2021).

///// SAR Presentation Award /////

SAR Presentation Award について

「SAR Presentation Award」は、構造活性相関シンポジウムにおける若手研究者の発表を奨励 し、構造活性相関研究の発展を促進するため、2010年度に創設された。当初は応募制とし て審査対象講演の募集を行った。2012年度からは、正式名称を「構造活性相関シンポジウ ム優秀発表賞」(英語表記 SAR Presentation Award)と定めた。

2021 年度 SAR Presentation Award について

2021 年度は、第49回構造活性相関シンポジウム(オンライン開催)における45歳以下の発表者(日本薬学会会員または受賞後に日本薬学会に入会いただける方)による一般講演(ロ頭発表・ポスター発表)を選考対象とすることとした。

2021 年度 SAR Presentation Award 受賞者(五十音順)

口頭発表 :	郡	除実	(横浜市立大学大学院)
ポスター発表:	伊藤	朱里	(横浜市立大学大学院)
ポスター発表:	武富	啓	(横浜市立大学大学院)
ポスター発表:	柴橋	佑希	(横浜市立大学大学院)
ポスター発表:	佐藤	朋広	(横浜市立大学大学院)

受賞者の選考について

2021年11月18日にオンラインにて各審査員から提出頂いた審査票を集計し、ロ頭発表1 名、ポスター発表4名を受賞者として選出した。ロ頭発表の審査は点数方式、ポスター発 表の審査は3演題選出するという方式で行った。授賞式は、第49回構造活性相関シンポジ ウムの閉会式において行った。後日、受賞者には、賞状と副賞を贈呈した。なお、審査に あたっての審査項目は下記の通りである。

口頭発表審査項目

a) 講演要旨: 講演要旨は発表内容を反映して適切に作成されているか。

- b) 講演資料: スライドは専門領域の異なる参加者にもわかりやすく、見やすく、かつ発 表時間に見合って適切に作成されているか。
- c) プレゼンテーション:発表時に参加者にわかりやすく説明しているか。
- d) 研究の目的:研究の背景と目的、先行研究との関係、研究の新規性あるいは有用性が 明確になっているか。
- e) 研究成果:価値のある成果が得られているか。
- f) 質疑応答: 質問等に対し、的確な応答・議論がなされたか。活発な討論がなされたか。
- g) 将来性:研究内容について、将来の発展が期待できるか。

審査員

第49回構造活性相関シンポジウムに参加した2021年度常任幹事および幹事

<受賞者コメント>

KO-01

- 氏名 郡 聡実(こおり さとみ)
- 所属 横浜市立大学 生命医科学研究科

演題 DNA メチル化制御薬の開発:維持メチル化因子 UHRF1 の機能阻害剤のリード化合物の同定

この度は、第49回構造活性相関シンポジウム優秀発表賞(口頭)という名誉ある賞を頂戴し、 大変光栄に存じます。ご評価いただきました先生方、ならびに日本薬学会構造活性相関部会の先 生方に厚く御礼申し上げます。

本発表では、DNA 維持メチル化因子である UHRF1 の阻害剤探索について報告しました。DNA メチル化は細胞の形質を決定する重要なエピジェネティック修飾ですが、その異常は細胞のが ん化と関連することが知られています。UHRF1 は様々ながん細胞で高発現し、異常な DNA メ チル化によるがん細胞の異常増殖と関連することから、創薬の標的分子として注目されていま す。我々は、複製と連携した DNA 維持メチル化の初期に形成される UHRF1 と DNA Ligase 1 の 複合体に着目し、この結合を阻害する化合物の探索を行いました。化合物ライブラリから計算科 学で候補化合物を選別し、生化学実験で結合評価を行いました。その結果、5-amino-2,4dimethylpyridine (5A-DMP) を同定し、この化合物の UHRF1 への結合様式を X 線結晶構造解析で 評価しました。さらに、ツメガエル卵抽出液を用いた無細胞実験で、5A-DMP は UHRF1 と DNA Ligase 1 の結合を阻害することを明らかにしました。今後、5A-DMP をリード化合物とした構造 展開でより活性の高い UHRF1 の機能阻害剤の開発が可能であると考えています。

最後に、本研究の実施の機会を与えてくださり、その遂行にあたって終始熱心な学術的指導を 賜りました有田恭平教授をはじめ、共同研究でお世話になりました先生方に深謝申し上げます。

DNAメチル化制御薬の開発:維持メチル化因子 UHRF1の機能阻害剤のリード化合物の同定

○郡聡実¹、柴橋佑希¹、浴本亨¹、西山敦哉²、吉見早恵 ¹、長門石暁²、大田雅照³、津本浩平²、中西真²、池口 満徳^{1,3}、有田恭平¹

(¹横浜市大·生命医、²東京大·医科研、³理研) E-mail: aritak@yokohama-cu.ac.jp

1. 背景と目的

DNA メチル化異常はがん化の原因の一つであ る。哺乳類の DNA メチル化は主に CG 配列中の シトシン塩基の 5 位の炭素に起こる。DNA メチル 化は遺伝子サイレンシングに関与し、細胞固有の 遺伝子発現パターンを定義する重要なエピジェ ネティック修飾である。一方で、がん細胞中では DNA メチル化パターンの異常がみられ、ゲノム不 安定化やがん抑制遺伝子のサイレンシングを引 き起こす。したがって、DNA メチル化制御因子を 標的とした薬剤はがん治療薬になると考えられ ている。

UHRF1 は DNA 維持メチル化に必須の因子で あるが、多くのがん細胞中で高発現していること が報告されている [1]。また、がん細胞中の UHRF1 のノックダウンはがん抑制遺伝子の発現 を回復させるので、UHRF1 はがん治療薬の標的分 子として潜在性を持つ。

我々は、UHRF1 の TTD ドメインと Lys126 が トリメチル化された DNA リガーゼ 1 (LIG1) が強 い親和性で結合し、これが UHRF1 を複製部位へ リクルートすることを明らかにした [2]。さらに、 この複合体の X 線結晶構造解析から、LIG1 との 高親和性での結合には UHRF1 TTD の Arg-binding cavity と LIG1 の Arg121 の相互作用が重要である ことを報告した [3]。UHRF1 と LIG1 の相互作用 は複製と連携した DNA 維持メチル化の初期で形 成されるので、UHRF1 と LIG1 の結合阻害はがん 細胞中で高発現している UHRF1 の機能阻害に繋 がることが期待される。本研究では、UHRF1 TTD の Arg-binding cavity を標的とし、LIG1 との結合 を阻害する UHRF1 機能阻害剤の開発を目的とし た。

2. 方法

1) In silico スクリーニング

ドッキングシミュレーション、構造揺らぎと水 和水を考慮した全原子 MD シミュレーション、結 合自由エネルギー計算、動的ファーマコフォアス クリーニングにより、化合物ライブラリーから UHRF1 TTD の Arg-binding cavity に結合する候補 化合物を選抜した。

2) UHRF1 TTD に結合する化合物の同定

熱安定性実験により、UHRF1 TTD に結合する 化合物を同定した。化合物とUHRF1 TTD の解離 定数は等温滴定カロリーメトリー (ITC) で定量 的に評価した。

3) UHRF1 TTD と化合物の X 線結晶構造解析

同定した化合物とUHRF1 TTD の結合様式を解 明するために共結晶化を行った。得られた結晶の 回折強度データは Photon Factory (PF) BL-5A で得 た。位相決定は単体の UHRF1 TTD の結晶構造 (PDB: 5YYA)をサーチモデルとして分子置換法 で行い、複合体構造を決定した。

4) <u>化合物によるUHRF1:LIG1 相互作用の阻害効</u> <u>果の検証</u>

化合物によるUHRF1とLIG1の結合の阻害効果 を検証するため、アフリカツメガエルの卵抽出液 を用いた阻害実験を行った。卵抽出液に化合物を 添加後、抗LIG1 抗体で免疫沈降し、共沈する UHRF1 を検出することで化合物の阻害効果を評 価した。

3. 結果とまとめ

In silico スクリーニングで化合物ライブラリー 中の約 20 万の化合物から 130 個の候補化合物を 選抜した。この 130 個の化合物を用いた熱安定性 実験の結果、UHRF1 TTD に結合する化合物を 2 個同定した。次に、最も安定して結合した化合物 である 5-amino-2,4-dimethylpyridine (5A-DMP) と UHRF1 TTD の解離定数を見積もるために、ITC 測定を行った。その結果、UHRF1 TTD と 5A-DMP は $K_{\rm D}$ = 19.3 μ M で結合することが分かった。



図-1 UHRF1 TTD:5A-DMPのX線結晶構造

複合体の全体構造(上図)と、Arg-binding cavity(下図)を示す。5A-DMPの $|F_0|$ - $|F_c|$ オミットマップを 4 σ で示す。5A-DMP は UHRF1 TTDのArg-binding cavity に結合し ていた。

5A-DMP と UHRF1 TTD の結合様式を明らかにす るため、次に複合体のX線結晶構造解析を行った。 1.45 Å 分解能での立体構造の決定に成功し、 5A-DMPは UHRF1 TTDの Arg-binding cavity に結 合することを明らかにした (図-1)。立体構造情報 に基づき、5A-DMPの認識に関与する UHRF1 TTD の5つのアミノ酸残基をそれぞれアラニン残基に 置換した変異体を作製した。これらの変異体を用 いた ITC 測定の結果、すべてのアラニン変異体で 5A-DMP の認識に影響を及ぼしたことから、 5A-DMP は UHRF1 TTD の Arg-binding cavity に特 異的に結合することが分かった。さらに、5A-DMP の結合による UHRF1 と LIG1 の結合阻害を検証す るために、アフリカツメガエルの卵抽出液を用い た阻害実験を行った。その結果、5A-DMPの濃度 依存的に UHRF1 と LIG1 の結合が阻害されること を明らかにした (図-2)。

本研究では、構造生物学と計算科学を組み合わ せたスクリーニングにより、大規模化合物ライブ ラリーから効率的に候補化合物を選別し、UHRF1 TTD の Arg-binding cavity に結合する 5A-DMP を 同定した。さらに、5A-DMP は UHRF1 と LIG1 の 結合を阻害することを明らかにした。5A-DMP を リード化合物としたさらなる構造最適化により、 がん細胞における UHRF1 の機能阻害剤となるこ とが期待される。



図-2 5A-DMP による UHRF1:LIG の結合阻害 アフリカツメガエルの卵抽出液を用いた阻害 実験の結果。5A-DMP の濃度依存的に UHRF1:LIG の結合が阻害された。

4. 謝辞

本研究を遂行するにあたり、化合物ライブラ リーを提供してくださいました東京大学 創薬機 構、回折実験でお世話になりました PF ビームラ インスタッフの皆様にこの場をお借りして感謝 申し上げます。

5. 参考文献

- 1. Sidhu H. and Capalash, N. Tumor Biol. 39, 1010428317692205 (2017),
- 2. Ferry L. et al., Mol Cell. 67, 550-565.e5 (2017),
- 3. Kori S. et al., Structure. 27, 485-496.e7 (2019),

<受賞者コメント>

KP-05

氏名 伊藤 朱里(いとう あかり)

所属 横浜市立大学大学院 生命医科学研究科

演題 中分子シクロスポリンAとシクロスポリンEの分子ダイナミクスの比較

この度は栄誉ある賞を頂戴し、大変嬉しく存じます。会期中は様々な研究分野の方々とのディ スカッションを通じて多くの気づきや学ぶことがあり、非常に意義深いものとなりました。 本研究は、経口投与可能な中分子医薬品シクロスポリン A (CsA) と、その代謝物であるシクロ スポリン E (CsE)の膜透過性の差が生まれる要因に、全原子モデルでの分子動力学シミュレー ションから迫るものであります。CsA と CsE で水中・膜界面・膜中の3箇所の環境での構造分 布に違いがあることがわかっただけでなく、CsA と CsE で1残基異なる部分における主鎖の N-メチル基の有無が、環状構造の形状に影響を及ぼしていることもわかりました。しかし、水中・ 膜界面・膜中の3箇所の環境以外での計算を行っていないため、水中から膜中における、一連の 膜透過過程で起こる構造変化は分かっておりません。一連の計算は通常のスーパーコンピュー タの計算資源では行うことができないため、現在はスーパーコンピュータ「富岳」を利用した計 算を実施しております。

これらの研究を進めるにあたり、温かなご指導をいただきました池口満徳教授をはじめ、多く のアドバイスをくださった浴本亨助教、理化学研究所の山根努上級研究員、生命情報科学研究室 の皆様に心より感謝申し上げます。

中分子シクロスポリンAとシクロスポリンEの 分子ダイナミクスの比較

○伊藤朱里¹、浴本亨¹、山根努²、池口満徳^{1,2}
 (¹横浜市大・生命医、²理研・R-CCS)
 E-mail: w205406f@yokohama-cu.ac.jp

1. 背景と目的

免疫抑制剤シクロスポリンA(CsA)は、非天 然アミノ酸を含む 11 残基の環状ペプチド分子で あり、大きな分子量(1202.6)にあるにも関わら ず、膜透過性を持つ経口投与可能な中分子医薬品 である。CsAは PPI 阻害と膜透過性を両立した代 表的な中分子医薬品として知られているが、その 膜透過メカニズムの詳細は明らかにされておら ず、膜透過性予測が困難である。結晶構造解析で 得られた構造から、CsA では親水溶媒中では環が 拡がった open 構造、疎水溶媒中では環がタイト に閉じた closed 構造をとり、open/closed 構造をダ イナミックに変化させながら膜透過すると想定 されている。しかし、水中、膜中、及び、透過中 の膜水界面中の CsA の構造(変化)情報は報告さ れていないため、その膜透過メカニズムは明らか ではない。また、CsA の代謝物シクロスポリン E

(CsE)では、ほんの1残基の主鎖のNメチル基 の有無だけの違いにも関わらず、膜透過性が約10 倍異なることが知られている。[1] CsA と CsE の 膜透過過程中の構造が比較できれば、膜透過性を 決定する主要因が明らかにできると期待される。

これまでに我々は、CsA と CsE の膜水系におけ る水中および膜中に両者を配置した MD シミュレ ーションを各 1-3 µ s 実施し、CsA および CsE の特 徴的な構造変化傾向の違いを PCA 解析から明ら かにした。本発表では両者の特徴的な違いをより 詳細に調べるために、拡張アンサンブル手法 gREST 法[2]を用いて、膜透過過程中で特徴的な、 水中・膜水界面中・膜中における、CsA と CsE の 構造サンプリングを実施し、各々の環境で出現頻 度の高い構造を明らかにすることを目的とする。 また、CsA と CsE の比較解析を実施し、1 残基の 違いが構造分布にどのように影響を及ぼすのか 明らかにする。

2. 方法

CsA は 11 残基中 8 残基が非天然アミノ酸(7 残基が N メチルアミノ酸、残り 1 残基が d 体アラ ニン)であり、N メチルアミノ酸については新規 に charmm 力場を開発し利用した。gREST 法は、 GENESIS を利用して実施した。CsA 重心と膜分子 重心間の Z 座標距離(膜面に垂直な座標)を固定 し、水中・膜水界面・膜中における構造サンプリ ングを実施した。

3. 結果とまとめ

CsA と CsE について、水中・膜水界面・膜中 の3つの場所で収束した構造分布を得ることがで きた。結晶構造解析で示唆されている、open/closed 構造で trans/cis が変化する残基の二面角の時系列 を解析したところ、150 ns 程度のレプリカ計算で 収束することがわかった。最終的に 300 ns まで実 行し、構造分布の解析を実施した。

主鎖構造の特徴を明らかにするために、構造群 を PCA により分類した。その結果、CsA では、 水中では open 構造、膜中では closed 構造近傍の 構造の固規頻度が高く、膜水界面では open/closed 構造の両方が出現することがわかった。構造群は、 結晶構造で解かれている open/closed 構造だけで なく、中間的な構造も出現することもわかった。 一方で、CsE の構造は CsA よりも広い構造分布を 持つことがわかった。特に、膜中の構造分布に顕 著な違いがあることがわかった。

CsA と CsE の主鎖構造の違いを生む要因を明 らかにするために、分子の尖り具合を解析した。 CsA と CsE 共に、closed 構造で主鎖の曲率が大き くなる部分は共通しているが、その曲率の分布が CsA の方がシャープであることがわかった。特に、 CsA と CsE で1残基異なる場所付近の違いが顕著 であることもわかった。これらは、主鎖 N メチル 基の有無で、その部位の環状構造のタイトさが変 化し、その結果、closed 構造の安定性が変化する ことを示唆している。

4. 参考文献

1. Ahlbach L C, Lexa W K, Bockus T A, Chen V, Crews P, Jacobson P M, Lokey S R, *Future Med. Chem.* 7, 2121-30 (2015).

2. Motoshi Kamiya, Yuji Sugita, J ChemPhys, 149, 072304 (2018).

<受賞者コメント>

KP-06

- 氏名 武富 啓(たけどみ けい)
- 所属 横浜市立大学 生命医科学研究科

演題 Effect of Water Molecules on the Activating S810L Mutation of the Mineralocorticoid Receptor

この度は、第 49 回構造活性相関シンポジウム SAR Presentation Award を頂き、誠にありがと うございます。審査員の先生方、並びに日本薬学会構造活性相関部会の先生方に心より御礼申 し上げます。オンラインではありましたが多くのアカデミアの先生方や製薬企業の方々と議論 させて頂くことができ、貴重なご意見を頂く事が出来ました。

本研究では、血中に存在するイオン濃度の維持に関わっている硬質コルチコイド受容体を題 材とし、疾患を引きおこす活性型 Ser810Leu 変異の活性化メカニズムを分子動力学シミュレーシ ョンにより解明することを試みました。その結果、変異したロイシン残基は"栓"のような役割 を果たし、周囲に存在する水分子から硬質コルチコイド受容体に結合しているリガンドを遮蔽 し、リガンドの結合を安定化させる働きがあることが分かりました。今後は、他のリガンドや他 の蛋白質にも得られた知見を活用していきたいと考えております。

最後に本研究を進めるにあたりご指導を賜りました、横浜市立大学の池口満徳教授、浴本亨助 教、理化学研究所の大田雅照上級研究員に感謝申し上げます。

Effect of Water Molecules on the Activating S810L Mutation of the Mineralocorticoid Receptor

o武富啓^{1,2}、大田雅照³、浴本亨¹、池口満徳^{1,3}
 (¹横浜市大・生命医、²田辺三菱製薬(株)、³理研)
 E-mail: takedomi.kei@mp.mt-pharma.co.jp

1. 背景と目的

核内受容体の一つ鉱質コルチコイド受容体 (MR)を活性化する S810L 変異は、若年性高血圧 や妊娠性高血圧を生じることが知られている。内 因性のステロイドホルモンであるプロゲステロ ンは迅速な解離を伴う受動的拮抗薬として野生 型(WT) MR に結合する。一方、活性型 S810L 変 異体には解離が遅いフルアゴニストとして結合 し、プロゲステロンの生物活性の変化が疾患の原 因の一つと考えられている。

S810L変異に伴い、プロゲステロンの活性がア ンタゴニストからアゴニストへ変化してしまう 構造的な要因を、結晶構造のみから考察するには 困難がある。(図1)核内受容体リガンド結合ド メインでは、ヘリックス12(H12)の構造変化が アゴニスト/アンタゴニスト活性を調節する機構 として知られている。しかし、MRの結晶構造で は、WTとS810L変異体でH12の構造に違いがな く、両者とも活性型構造をとっている。さらに、 S810Lは、H12とリガンドを挟んだ反対側に位置 しており、なぜ、S810LがH12の構造変化で調整 されるアゴニスト/アンタゴニスト活性の変化を 引き起こすのか理解が困難であった。



図1:野生体と変異体の結晶構造比較

実験的に、S810L変異体ではプロゲステロンの

解離速度が WT よりも速くなることが知られてい る。[1] WT と S810L 変異体では、プロゲステロン の結合親和性 (K_d) は同程度であるが、半減期 (t_{1/2}) が WT (約 10 分) から約5倍長くなる。これは、 解離速度定数 (k_{off}) で WT: 1.0×10^{-3} s⁻¹、S810L: 0.2 × 10^{-3} s⁻¹ となる。K_d と k_{off} から、会合速度定数 (k_{on}) を見積もると、WT: 1.0×10^{5} M⁻¹s⁻¹、S810L: 0.8× 10^{5} M⁻¹s⁻¹ となり、k_{on} については同程度の値を示 す。これらの実験結果によると、S810L 変異によ って koff が大きく変化することから、変異に伴う 活性変化の要因がプロゲステロンの解離過程に あることを示唆している。

本研究では、S810L 変異が解離過程に及ぼす影響を明らかにするために、分子動力学シミュレーション (MD)を用いた解離過程解析を実施した。 Steered Molecular Dynamics (SMD)を用いて、プロ ゲステロンを結合状態から解離させた。S810L は 結合ポケットの奥側に位置しており、リガンドが 解離すると、その隙間に水分子が流入する。リガ ンドと水分子の置き換わりは、結合/解離過程の重 要な要素であるため、アポ状態構造で MD を実施 し、結合ポケット内の水分子の分布を WT と S810L で比較解析した。これらの結果を統合し、 S810L 変異に伴う解離過程の変化を説明できる、 Water inflow model を提案した。[2]

2. 方法

WT および S810L 変異体におけるプロゲステ ロンの解離過程の解析には SMD を用いて、リガ ンドの重心と 810 番目の残基の重心間距離が離れ るように力を加えた。WT および S810L 変異体に おけるリガンドフリー体での水分子の解析は、複 合体構造からプロゲステロンを除いた構造を用 いた MD 計算により実施した。MD は、Gromacs を用いて実施した。

3. 結果とまとめ

SMD でプロゲステロンの解離過程をシミュレ ーションした結果、WT に結合したプロゲステロ ンでは短時間で解離が始まることがわかった。 (図 2 a) プロゲステロン重心の位置(赤線)は、 7 ns で大きく変化(図 2a, ①)し、15 ns 付近で2 段階目の変化が起こる(図 2a, ②)。対照的に、 S810L 変異体では、15 ns 付近まで重心の変化がな く、大きな変化は1度だけであった(図 2b, ③)。 SMD 中でかかる力は線形であるため、WT ではプ ロゲステロンの結合が弱いが、S810L では結合が 強固であることを意味する。これらの結果は、 S810L 変異によってプロゲステロンの解離が起こ りやすいことを示唆しており、実験結果と一致し ていた。



解離過程の違いを引き起こす物理的な要因を 明らかにするために、リガンド解離に伴い、置き 換わると想定される水分子の分布を解析した。解 離過程シミュレーション中において、結合ポケッ トの大きさは WT と S810L で一致していたため、 解離過程の違いは、流入する水分子の違いに起因 する。アポ状態の構造で WT と S810L の MD を実 施し、結合ポケット内の水分子の分布を得た。(図 3) WT では水の流入部分が、プロゲステロン横と S810部分の2箇所存在していた(図 3a)が、S810L 変異体では、プロゲステロン横の箇所のみに減少 していた。また、プロゲステロン横の流入場所も、 狭くなっており、S810L では水分子の流入が起こ りにくくなっていることがわかった。これらの違 いは、S810L 変異によって、疎水的で嵩高い側鎖 が存在することで、水分子の通り道が制限されて しまったこととして解釈できる。



図3:アポ状態での水分子のパスウェイ

以上の結果より、S810L 変異に伴い、プロゲス テロンの解離が遅くなり、その原因は水分子の流 入経路の制限にあることがわかった。これらを統合し、S810L 変異による解離過程の違いを、水分子の流入から理解する、Water inflow model を提案した。(図 4)

Water inflow model



WTでは、水分子の流入口が十分に確保されて おり、水分子の流入がしやすい状態にある。その ため、十分な水分子の流入は、プロゲステロンの 解離をアシストできる。対照的に、S810L変異体 では、疎水的な嵩高い側鎖が流入口に存在するこ とになり、水分子の流入を制限する。そのため、 プロゲステロンの解離をアシストできず、結合状 態を維持しやすくなる。Water inflow model による と、変異体で解離速度が遅くなる実験結果を、H12 から遠いS810L付近で起こる水分子流入の制限に よって説明できる。

4. 参考文献

[1] Geller D S. Science. 289, 119–123 (2000).

[2] Takedomi K, Ohta M, Ekimoto T, Ikeguchi M. J. Chem. Inf. Model. 61, 3583–3592 (2021).

<受賞者コメント>

KP-15

- 氏名 柴橋 佑希(しばはし ゆうき)
- 所属 横浜市立大学 生命医科学研究科

演題 維持メチル化因子 UHRF1 の阻害剤探索:分子動力学シミュレーションを組み合わせ たインシリコスクリーニング

この度は、第49回構造活性相関シンポジウム SAR Presentation Award を賜り、大変光栄に存じ ます。ご評価いただきました審査員の先生方、ならびに日本薬学会構造活性相関部会の先生方に 深く御礼申し上げます。本シンポジウムでは、様々な専門家の方々と議論を交わし、研究につい て多くのご意見を頂くことができました。

本研究は、複数の計算科学的な手法を用いて、がん治療薬として期待される、UHRF1の機能 阻害剤を開発するというものになります。静的な構造を使ったドッキングシミュレーションと 分子動力学シミュレーション (MD)を使った動的情報を組み合わせたスクリーニングを実施し、 機能阻害剤となるヒット化合物を得ることができました。この MD を組み合わせた探索は、膨 大な数の化合物ライブラリの中から効率的に候補化合物の選別が行えることを実証し、今後の 阻害剤探索の置ける新たな方法論を提示できたものと考えております。

最後に、本研究を進めるにあたりご指導賜りました池口満徳教授、浴本亨助教、理化学研究所 の山根努上級研究員に感謝申し上げます。また、共同研究者として多大なお力添えをいただきま した有田恭平教授、郡聡実さんに感謝申し上げます。

維持メチル化因子 UHRF1 の阻害剤探索: MD を組み合わせたインシリコスクリーニング

○柴橋佑希¹、浴本亨¹、郡聡実¹、山根努²、有田恭平¹、 池口満徳^{1,2}

(¹横浜市大·生命医、²理研·R-CCS) E-mail: w195418g@yokohama-cu.ac.jp

1. 背景と目的

エピジェネティックな変化の蓄積は、腫瘍形成 の主な原因の一つである。DNAメチル化パターン の異常はゲノムの不安定化、そして様々ながん細 胞でがん抑制遺伝子のサイレンシングを引き起 こす。そのため、DNAメチル化の制御因子を標的 とした薬剤開発はがん治療薬に発展する可能性 がある。

Ubiquitin-like containing PHD and RING finger domain 1(UHRF1)は DNA 維持メチル化に必要不 可欠な因子である[1,2]。さらに、UHRF1 は種々の がん細胞において過剰発現しており、ノックダウ ンすることでがん抑制遺伝子の発現が回復する ことが報告されている[3]。このことから UHRF1 の阻害は有望ながん治療薬の創薬標的である。

UHRF1 は 5 つの機能ドメインから構成されて いる。特に、tandem Tudor domain(TTD)のペプチ ド結合溝(TTD-groove)は分子内、分子外相互作用 の結合プラットフォームとして機能している。 TTD-groove と DNA リガーゼ 1 の Lys126 ジ/トリ メチル化体(LIG1K126me3) との相互作用は、 UHRF1を DNA 複製の場に誘導するトリガーとな っている。そのため、本研究では TTD-groove と LIG1K126me3 の結合を阻害する分子の発見を目 的とした。



図-1 TTD-LIG1K126me3 相互作用

2. 方法

20 万超の化合物ライブラリから、阻害剤候補と なる化合物の探索を実施した。化合物の探索は、 ドッキングシミュレーションや MM-GBSA、分子 動力学シミュレーション(MD)、ファーマコフォア によるインシリコスクリーニングを利用した。ド ッキングシミュレーションには Glide を、MD の プログラムパッケージには GROMACS を、ファ ーマコフォアスクリーニングには LigandScout を それぞれ使用した。

3. 結果とまとめ



図-2 In silico スクリーニングサイクル

静的構造を使ったドッキングシミュレーショ ンと、分子動力学シミュレーション(MD)を使 った動的情報の両方を組み合わせたインシリコ スクリーニングサイクルを構築し、ヒット化合物 探索を実施した(図-2)。ドッキングシミュレーシ ョン結果から、ドッキングスコアと MM-GBSA ス コアにより結合親和性を評価し、その結合ポーズ の安定性を全原子 MD の RMSD によって評価す る。さらに、MD のトラジェクトリから動的ファ ーマコフォアを作成し、ファーマコフォアスクリ ーニングによって候補化合物を取得した。一般的 に、in silico の、どの計算スコア、および、指標 が実験的に評価された親和性と相関するか分か らないため、本スクリーニングでは、複数のスコ アや指標からトップの化合物を選出し、最終的に、 130 個のヒット化合物候補を取得した。

全原子 MD では、ドッキングシミュレーション から得た複合体構造を使い、10 ns MD を 3 回独立 に実行した。リガンドの相対配置の RMSD の平均 値 と標準偏差によって、"Stable"と"Partially Stable"、"Unstable"の 3 パターンに分類することで、 ポーズの安定性を評価した。Stable に分類された、 図-3 の(a),(b)の 2 個の化合物は Thermal stability assay による結合実験でヒットした。特に、(a)の 化合物はスクリーニングサイクルの4 つの指標の 内、ドッキングスコアを除いた 3 つで優れた評価 を示していた。



図-3 RMSD によるポーズ安定性の評価

130 個の化合物に対して、実験的に結合実験を 実施した結果、2つのヒット化合物を得ることが できた。そのうち1つについて、UHRF1とヒット 化合物複合体の結晶構造を解いた結果、結晶構造 中の結合ポーズはMD中で確認されたポーズとほ とんど同一であった。また、ヒット化合物の結晶 構造から1µs MDを実施したところ、化合物は結 合ポーズで安定であり、タンパク質-リガンド間の 相互作用も予測通り安定に形成されていた。

2個のヒット化合物は MD に基づいた指標で高 く評価されていた。MD の指標を用いたとき、候 補化合物におけるヒット化合物の割合は 10.5 % で、一般的なハイスループットスクリーニングの ヒット率1%よりも良い結果となった。これらの 結果は、標的タンパク質に結合する化合物のスク リーニングにおいて、構造揺らぎや水分子を考慮 する MD が、有用な手法であることを示している。

本研究で得られたヒット化合物は、UHRF1 と LIG1K126me3 の結合を阻害することが確認され たが、結合能は不十分であった。ヒット化合物は 分子量の小さいフラグメント化合物だったため、 構造展開を行うことで活性の向上を図っている。 そこで、機械学習をベースに分子生成を行う ChemTSにより、リード化合物の候補を発生させ る。そして、自由エネルギー摂動法により結合自 由エネルギーを計算することで、リードを最適化 する手法を検討している。



図-4 ChemTS と相対 FEP 計算 を活用した構造展開

4. 謝辞

本研究を行うにあたり、化合物ライブラリを 提供していただきました東京大学創薬機構に深 く感謝申し上げます。

本研究、および、東京大学創薬機構の化合物 サンプル提供は国立研究開発法人日本医療研究 開発機構(AMED)創薬等ライフサイエンス研究支 援基盤事業創薬等先端技術支援基盤プラットフ オーム(BINDS)の課題番号 JP20am0101109 の支援 を受けました(支援番号 1746(インシリコ 1614))。

5. 参考文献

- 1. Sharif J, et al. Nature 450, 908-912 (2007),
- 2. Bostick M, et al. Science 317, 1760-1764 (2007),
- 3. Ashraf W, et al. Oncotarget 8, 51946-51962 (2017),

<受賞者コメント>

KP-19

氏名 佐藤 朋広(さとう ともひろ)

所属 横浜市立大学 生命医科学研究科

演題 転移学習を用いた活性予測モデルの新規化合物シリーズに対するモデル適用性の改善

この度、第49回薬学会構造活性相関シンポジウム優秀発表賞(ポスター)を賜りまして、大 変光栄に存じます。昨年に続き厳しい社会情勢の中、リモート開催によって研究発表の機会をい ただきましたこと、運営委員会の先生方に厚く御礼申し上げます。

機械学習に基づく予測モデルは予測適用範囲が学習データに強く依存しており、新規性の高 い化合物に対する活性等の予測が外挿となり予測精度が低下してしまうことが課題となってお ります。本研究では転移学習と半教師あり学習を組み合わせることで、新規化合物シリーズの実 験データが少ない場合でも過学習を抑制して効率的な予測モデルのチューニングを可能とする 学習フレームワークを開発いたしました。予測モデルの性能評価における内挿・外挿の違いは、 しばしば既知データを用いたベンチマークと実際の創薬研究に適用した場合とで予測性能が乖 離してしまう原因となっており、今後も機械学習を用いた手法がより創薬の現場に対して貢献 できるよう研究を進めていきたいと考えております。

最後に、本研究の共同発表者の寺山慧先生、本間光貴先生に感謝を申し上げます。

転移学習を用いた活性予測モデルの新規化合物 シリーズに対するモデル適用性の改善

○佐藤朋広¹、寺山慧¹、本間光貴²
 (¹横浜市大・生命医科学研究科、²理化学研究所)
 E-mail: sato.tom.wa@yokohama-cu.ac.jp

1. 背景と目的

近年、創薬分野において深層学習に代表される 機械学習を用いた予測は、標的タンパク質に対す る阻害活性や体内動態の予測など、様々な形で利 用されている。我々の研究グループでも、2015~ 2019 年度に実施した AMED 創薬インフォマティ クス事業において心毒性リスクの予測を目的と して ChEMBL. PubChem 等の公共データベースお よび独自に測定したアッセイ結果を統合するこ とで約 30 万化合物の hERG 阻害情報を収載した データベースを構築し、それに基づく hERG 阻害 活性予測モデルを構築し発表している[1,2]。本予 測モデルは遺伝的アルゴリズムを用いた記述子 選択と Support Vector Machine による判別分析を 組み合わせることで、テスト化合物セットに対し て正答率 98.4%、Kappa 統計量 0.733 を記録し、 市販されている hERG 予測モデルを大きく上回る 予測精度を達成した。

一方で、機械学習を用いた予測モデルの弱点と して、予測適用範囲が学習データに強く依存する ことが知られている[3]。前述のhERG予測モデル においても新規性の高い化合物に対しては予測 精度が低下してしまうという問題が確認されて いる。本研究では、予測モデルの新規化合物シリ ーズに対する予測性能を事後的に改善する事を 目的として、EGFR 阻害剤、COX-2 阻害剤、A2a 阻害剤の公知情報を利用して、あらかじめ特定の 化合物シリーズを除外して深層学習モデルを構 築し、除外した化合物シリーズのデータを二次的 に追加学習させることで予測性能の変化を検証 した。

2. 方法

追加学習による予測モデルの適用範囲の拡大 について、図1に概略を示す手順で検証を行った。 検証に用いるデータセットには、ChEMBL データ ベース(v28)より EGFR, COX-2, A2a 受容体の阻害 剤情報を収集し、IC50, pKi などを標準化した pchembl_value が設定された化合物から1. 変異体 に対する実験情報、2. アゴニスト活性を有する化



図-1 深層学習モデル比較の概要

合物、3. 分子量150 未満および800以上の化合物、 4. 無機化合物、を除去することで、それぞれ5,266、 3,314、4,479 化合物の活性情報を取得した。各標 的タンパク質の化合物セットについて、仮想的に 予測対象の新規化合物シリーズとみなす化合物 群を抽出するため、ECFP4, FCFP4, MACCSKeys フィンガープリントを用いた構造クラスタリン グと構造類似性計算を組み合わせて類似の骨格 構造を持つそれぞれ1,077、694、657 化合物を予 測対象として選択し、ランダムに75%を学習用デ ータ、25%をテスト用データとして分割した。

予測モデルとしては、ECFP4 と Tensorflow を用 いて pchembl_value の回帰を行う深層学習モデル を、①予測対象以外の阻害剤データを全て学習し たグローバルモデル、②予測対象の化合物シリー ズのみを学習したローカルモデル、③グローバル モデルに予測対象の化合物シリーズを追加学習 してファインチューニングしたモデル、の3パタ ーンで構築し、テスト用に学習から除外しておい た予測対象化合物セットに対する予測性能を検 証した。深層学習の設定は中間層の数、ノード数、 ドロップアウト率等のハイパーパラメータにつ いて、学習データの 5 フォールドの交差検定で RMSE 値が最小となるよう Optuna で最適化を実 施して決定した。追加学習の際には元の予測モデ ルについて 1. 全情報をそのまま使用する、2. 重 みガウスノイズを加える、3. 重みに一定割合で乱数を加える、の3種から Optuna により交差検定での RMSE 値が最小となる設定を選択した。

3. 結果とまとめ

EGFR、COX-2、A2a 受容体それぞれについて、 各種予測モデルについてテストセットに対する 予測性能 (\mathbf{R}^2 値) を図2に示す。いずれの標的タ ンパク質に関してもグローバルモデルの予測性 能は学習データから除外した化合物シリーズに 対しては顕著に低下し(EGFR:0.707→0.277、 COX-2:0.565→0.306、A2a 受容体:0.681→0.081)、 グローバルモデル<ローカルモデル<追加学習 モデルの順に予測性能が改善していくという結 果が得られた。追加学習モデルの R² 値は EGFR:0.492、COX-2:0.59、A2a 受容体:0.522 とな り、追加学習により元のグローバルモデルでは予 測適応範囲外であった化合物シリーズに対して も高い予測精度を達成することに成功した。また、 追加学習モデルとローカルモデルを比較した場 合、ハイパーパラメータのチューニングに関して は、追加学習モデルはネットワーク構造などのハ イパーパラメータが元のグローバルモデルと同 じ設定に固定される関係から、ローカルモデルの 方がより広範な設定から予測対象の化合物シリ ーズに適した条件を選択することができる。本検 証において追加学習モデルはハイパーパラメー タのチューニングに関しては有利なはずのロー カルモデルを上回る予測性能を示しており、グロ ーバルモデル構築の際に学習した多様な骨格の 阻害剤情報から、対象の化合物シリーズの予測に 有用な情報を転移学習によって獲得することで 高い予測性能を達成していることが示唆された。



図-2 各標的に対する予測モデルの性能

上記結果に加え、追加学習に用いるデータをラ ンダム選択により減らすことで、より実測データ が少ない状況を想定した予測性能についても比較検討を行った。COX-2 に関して化合物数 5%~50%でランダム選択を各 10 回実施した結果について、選択化合物数と R²値の関係を図 3 に箱ひげ図として示す。

追加学習モデルはいずれのデータ量の場合も ローカルモデルの予測性能を上回ったものの、追



図-3 COX-2について追加学習に用いるデ ータ数を変えた場合の予測性能

加学習データを元の 5%まで減らした場合には予 測性能が元のグローバルモデル以下へと低下し ており、データ量が不足する場合には追加学習時 に少数のデータに対してオーバーフィッティン グすることで予測性能が低下する可能性が示唆 された。

本研究により深層学習を用いて多様な構造を 学習して構築したグローバルモデルを元にして 特定の化合物シリーズを追加学習させることで、 対象の化合物シリーズに特化した高性能の予測 モデルを構築可能であることが示唆された。今後、 より多くの標的タンパク質や条件設定での検証 を実施することで、多様な条件で安定して予測性 能の向上が期待できる頑健な予測モデル構築手 順を確立することを目指している。

4. 謝辞

本研究は AMED 産学連携による次世代創薬 AI 開発事業の一環として実施された。

5. 参考文献

- 1. Sato T, Yuki H, Ogura K, Honma T, *PLOS ONE* **13**, 1–18 (2018),
- 2. Ogura K, Sato T, Yuki H, Honma T. Sci Rep., **9** (1):12220 (2019).
- 3. Mathea M, Klingspohn W, Baumann K, *Mol Inform*. **35**(5), 160-180 (2016),

///// Activities /////

第49回構造活性相関シンポジウム開催報告 日時:令和3年11月18日(木)・19日(金) 会場:オンライン開催 主催:日本薬学会構造活性相関部会 協賛:日本農薬学会・情報計算化学生物学会・日本化学会 後援:日本農芸化学会

11 月 18~19 日の 2 日間にわたって、第 49 回構造 活性相関シンポジウムを開催した(図 1)。新型コロナ ウイルスの感染拡大を防ぐために、前回と同様にオン ラインによる開催とした。オンラインツールには、 前回ほとんどトラブルがなかったことから、特別・ 招待講演と口頭発表のセッションに Zoom を、ポスタ 一発表、ランチョンセッション及び懇親会に Remo を 用いた。また、前回、参加費を無料とすることにより 多くの参加者を集め、構造活性相関研究の裾野を広げ ることができたことから、今回も無料とした。その 結果、参加登録 279 名(うち学生 56 名)、最多同時 接続数は両日ともに150名程度と多くの参加者があり、 前回と同規模の開催となった。

今回のシンポジウムでは、医農薬設計研究の新たな ブレイクスルーを創出するため、新規創薬手法や様々 な手法の適用事例についての発表を募ったところ、 一般ロ頭発表8件、ポスター発表32件の演題登録が あった。また、特別講演と二つの招待講演の他、両日 ともに2演題のランチョンセミナーを企画した。



図1 ファルマシアに掲載した開催案内

特別講演には、部会長を含め長く構造活性相関部会の運営に多大な貢献をしてこられた大阪 大学大学院薬学研究科 高木達也教授に「計量薬学の歴史と未来」と題してご講演をいただいた (図 2)。また、初日の招待講演では、株式会社 Epigeneron 研究開発部長の今枝泰宏先生から、 多様な医薬品モダリティが注目される中での創薬化学と計算化学の連携について、研究事例を 交えて紹介いただいた。2日目の招待講演では、クライオ電顕によるタンパク質構造解析が注目 されている昨今であるが、約10年前から Spring-8 で運用が開始されている SACLA (X線自由 電子レーザー施設)におけるタンパク質の結晶構造解析の現状と最近の成果について、和歌山県 立医科大学薬学部の中津亨教授にご講演いただいた。



図2 特別講演の様子(終了時、部会幹事にはビデオオンにして拍手を依頼した。講演番号隣が高木先生)

一般ロ頭発表では、インシリコ技術を活用したリード化合物探索をはじめ、タンパク質構造 モデル評価、薬効・薬剤耐性予測、相互作用解析、手法開発等、多様な研究テーマの成果が紹介 され、活発な質疑応答があった。Zoom Webinar によるロ頭講演・発表は、大きなトラブルなく 実施することができたが、質疑応答に際し座長が Q&A に寄せられた質問を代読し演者が回答 する形式は、多くの学会でも採用されているものの、議論が深まりにくく、改善が必要ではない かと感じている。

ポスター発表は、Remo Conference を用いて実施した。セッション直前に、各発表の1分間の ショートプレゼンテーション動画を視聴したのちに、発表ブースでポスターを掲示し討論する ようにした。オンラインであるが、図 3 に示すように各ブースの混み具合がわかるため、効率 よく発表を聴講できるように感じた(人気のある発表は常に混んでいたが)。各ブースでは、 適度に質問者が入れ替わってディスカッションが行われており、リアルな学会と変わらない印象 を受けた。しかし、複数の参加者が個々にポスター画面をスクロールすると意図しない方向に 画面が動き、非常に見づらい状況になることがあり、この点が改善されればと思われた。今回、 発表者にはショートプレゼンテーション用の動画の他に、シンポジウム開催 1 週間前から当日 まで視聴できる 5 分以内の説明動画を作成してもらい、参加者のみが視聴できるようにした。 これらの動画により、参加者が議論を深めたり、時間の都合で聴講できなかった発表の概要を 知ることができたりしたのではないかと思う。

以上の研究発表の他に、両日ともランチョンセッション(12 時~13 時)を企画し、企業からのセミナーを4件(各日2件)実施した。Remo Conferenceのセミナーモードを用いたが、トラブルなく実施することができた。昼食の提供がないため、あらかじめシンポジウム参加者にはお弁当を持参で参加していただくようお願いしておいたところ、両日ともに80名前後の出席があった。

初日のセッション終了後には、Remo Conference で懇親会(情報交換会)を開催した。40名 程度のご参加があり、発表の質疑応答の続きや情報交換、コロナ禍でなかなか会うことができな い方とのご挨拶など、様々な目的で活用いただけたのではないかと思う。18時の中締め後も、 数名の方には20時近くまで歓談していただけた。

SAR Award は、審査希望のあった発表を対象に、部会幹事のオンライン投票の結果を集計し、 受賞者を決定した。以下に受賞者を紹介する。今年も、学生による優秀な発表が多く、受賞者 5名のうち、3名を学生が占めた(図4)。受賞者の皆様にはお祝い申し上げるとともに、受賞を 励みに、今後もご活躍されることを期待します。また、審査していただいた先生方にも御礼申し 上げます。



図 3. ポスターセッション時の Remo の画面

KO-01 郡聡実(こおりさとみ) 横浜市立大学大学院生命医科学研究科 D3 DNA メチル化制御薬の開発:維持メチル化因子 UHRF1の機能阻害剤のリード化合物の同定

KP-05 伊藤朱里(いとうあかり) 横浜市立大学大学院生命医科学研究科 M2 中分子シクロスポリンA とシクロスポリンE の分子ダイナミクスの比較

KP-06 武富啓(たけどみけい) 横浜市立大学大学院生命医科学研究科 D3 Effect of Water Molecules on the Activating S810L Mutation of the Mineralocorticoid Receptor

KP-15 柴橋佑希(しばはしゆうき)横浜市立大学大学院生命医科学研究科 特任助手 維持メチル化因子 UHRF1 の阻害剤探索:

分子動力学シミュレーションを組み合わせたインシリコスクリーニング

KP-19 佐藤朋広(さとうともひろ)横浜市立大学大学院生命医科学研究科 特任准教授 転移学習を用いた活性予測モデルの新規化合物シリーズに対するモデル適用性の改善



図 4. SAR Award 受賞式における受賞者の様子

今回のシンポジウムは、ほぼ前回のオンラインシステムの運用を踏襲したため、大きなトラブ ルなく実施することができた。種々のご助言、情報、資料等をご提供いただいた第48回シンポ ジウムの実行委員の先生方に感謝いたします。また、これらをもとに本シンポジウムを企画、 調整、運営していただいた実行委員の先生方、アルバイトの学生諸氏にも謝意を表します。 最後に、協賛・展示いただいた企業の皆様、特別講演・招待講演をお引き受け下さいました 先生方、発表者、一般参加の皆様に厚く御礼申し上げます。

第49回構造活性相関シンポジウム

実行委員長 仲西 功

///// Activities /////

<会告>

構造活性フォーラム 2022

「In silico アプローチによる毒性予測研究および周辺分野の現状と展望」

主催: 日本薬学会構造活性相関部会

協賛・後援: 日本化学会、計算情報科学生物学会(CBI学会)、日本毒性学会

会期: 2022 年 6 月 3 日 (金)

会場: Zoom によるオンライン開催

フォーラムホームページ: http://www.qsarj.org/forum2022 (4月公開予定)

開催趣旨: 化学物質の安全性を確保するために、毒性の評価は必須の課題である。毒性物質を効率的に識別する方法として、*in silico* 技術を用いた予測法の確立には大きな期待が寄せられている。しかし、毒性は一般に複雑な発現メカニズムを伴うことから、予測が困難な対象として知られている。毒性発現過程には、組織・臓器への暴露に関連する体内動態の要因、生体内高分子であるタンパク質や核酸との相互作用、そしてそれに続く多様な生化学経路が関与する。そのため、単一のタンパク質を解析対象とすることは困難であることが多い。肝毒性の様に同じ表現型であっても、原因物質ごとに異なる機序で誘発される例もある。一方、予測モデルの説明性に対する要望は極めて強い。本フォーラムでは、このような毒性関連の生理的なイベントを *in silico* 技術でひもとき、解析し、予測する研究について議論を深めたい。様々な形で毒性関連の研究に携わっている方々にとって有意義な意見交換の場となれば幸いである。

プログラム:

基調講演「毒性発現メカニズムに基づく一般化学品の毒性予測—AI-SHIPS プロジェクト—」 船津 公人(奈良先端科学技術大学院大学)

講演 1. 「Applicable Artificial Intelligence Method to Drug Metabolism and Pharmacokinetics -Comparison of Various Methods for Metabolic Active Sites-」

笹原 克則 (Otsuka Pharmaceutical Development &

Commercialization, Inc.)

講演2.「大規模変異原性データを用いた第二回 Ames/QSAR 国際チャレンジプロジェクト」 古濱 彩子(国立医薬品食品衛生研究所)

講演 3. 「AI 創薬の基盤とデータ統合」

水口 賢司(大阪大蛋白質研究所・医薬基盤健康栄養研究所)

講演 4. 「拡散方程式の ADMET 予測モデルへの適用」

日高 中(武田薬品工業)

講演 5. 「副作用研究における AI の可能性」

奥野 恭史(京都大学大学院医学研究科)

参加登録および申込締切日: 5月20日(金)までに、フォーラムホームページから事前参加登録をお願いいたします。参加人数が上限に達しましたら、参加登録を打ち切らせていただきますのでご了承ください。

参加費: 一般会員・関連学会員 1000 円 非会員 2000 円 学生 無料

問合先: 構造活性フォーラム 2022 実行委員会 植沢 芳広 (実行委員長)
 〒204-8588 東京都清瀬市野塩 2-522-1 明治薬科大学
 Tel: 042-495-8983 E-mail: uesawa@my-pharm.ac.jp

部会役員人事	
2022 年度 常任世話人	
部会長	本間光貴(理化学研究所)
副部会長	服部一成(塩野義製薬(株))
副部会長	竹田–志鷹 真由子(北里大学 薬学部)
会計幹事	川下理日人(近畿大学 理工学部)
庶務幹事	杉本 学(熊本大学大学院 先端科学研究部)
広報幹事	加藤博明(広島商船高等専門学校)
SAR News 編集長	幸 瞳(理化学研究所)
ホームページ委員長	高木達也(大阪大学大学院 薬学研究科)

構造活性相関部会の沿革と趣旨

1970年代の前半、医農薬を含む生理活性物質の活性発現の分子機構、立体構造・電子構造の計算や活性 データ処理に対するコンピュータの活用など、関連分野のめざましい発展にともなって、構造活性相関と 分子設計に対する新しい方法論が世界的に台頭してきた。このような情勢に呼応するとともに、研究者の 交流と情報交換、研究発表と方法論の普及の場を提供することを目的に設立されたのが本部会の前身の構 造活性相関懇話会である。1975年5月京都において第1回の「懇話会」(シンポジウム)が旗揚げされ、1980 年からは年1回の「構造活性相関シンポジウム」が関係諸学会の共催の下で定期的に開催されるようになっ た。

1993年より同シンポジウムは日本薬学会医薬化学部会の主催の下、関係学会の共催を得て行なわれることとなった。構造活性相関懇話会は1995年にその名称を同研究会に改め、シンポジウム開催の実務担当グループとしての役割を果すこととなった。2002年4月からは、日本薬学会の傘下組織の構造活性相関部会として再出発し、関連諸学会と密接な連携を保ちつつ、生理活性物質の構造活性相関に関する学術・研究の振興と推進に向けて活動している。現在それぞれ年1回のシンポジウムとフォーラムを開催するとともに、部会誌のSAR Newsを年2回発行し、関係領域の最新の情勢に関する啓蒙と広報活動を行っている。本部会の沿革と趣旨および最新の動向などの詳細に関してはホームページを参照頂きたい。

(https://sar.pharm.or.jp/)

編集後記

今号では「マテリアル分野の計算化学」をテーマとして、2人の先生にご執筆をお願いしました。産業 技術総合研究所の青柳先生からは、ソフトマテリアル分野での計算科学の活用についてさまざまな事例を 交えてご紹介いただきました。明治大学の金子先生からは、予測誤差の推定、直接的逆解析法を活用した 分子設計例、適応的実験計画法といった多岐にわたる手法についてそれぞれご紹介いただきました。

本部会では医農薬分野および生理活性物質関連の研究が主ですが、マテリアル分野における化合物最適 化と共通する観点が多いと感じております。みなさまのご研究の参考になれば幸いです。ご多忙の中、快 くご執筆していただいた先生方に深く感謝申し上げます。

昨年11月に開催した構造活性相関シンポジウムのご報告および6月の構造活性フォーラムの会告も掲載 しておりますので、お目通しいただければ幸いです。

(編集委員会)

SAR News No.42 2022 年 4 月 1 日

発行:日本薬学会 構造活性相関部会長 本間光貴

SAR News 編集委員会

(委員長)幸 瞳、河合健太郎、清田泰臣、合田浩明、田上宇乃、仲西 功

*本誌の全ての記事、図表等の無断複写・転載を禁じます。