



---

## 「目次」

---

### //// Perspective/Retrospective ////

大規模言語モデル（LLM）の創薬分野への応用とその展望

東京科学大学情報理工学院 関嶋政和

叢雲くすり (@souyakuchan) ... 1

### //// Cutting Edge ////

タンパク質言語モデルの創薬分野での利用

東京科学大学情報理工学院 大上雅史 ... 8

大規模言語モデルによる科学的知識の獲得と応用：化学分野における最新の研究事例

株式会社サイキンソー 山崎広之 ... 13

AI for Science と AI Scientist

fuku 株式会社 山田涼太 ... 22

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

SAR Presentation Award について、および受賞者コメント ... 29

### //// Activities ////

<報告>

第52回構造活性相関シンポジウム 開催報告

国立医薬品食品衛生研究所 古濱彩子 ... 32

<会告>

構造活性フォーラム 2025 会告

摂南大学薬学部 河合健太郎 ... 34

第53回構造活性相関シンポジウム 会告

近畿大学理工学部 川下理日人 ... 35

編集後記 ... 37

## //// Perspective/Retrospective ////

## 大規模言語モデル (LLM) の創薬分野への応用とその展望

東京科学大学情報理工学院 関嶋政和  
叢雲くすり (@souyakuchan)

### 1. はじめに

近年、人工知能 (AI) の急速な進歩に伴い、大規模言語モデル (Large Language Models; LLM) が広範な分野で実用化されつつある。LLM とは、大量のテキストデータを深層学習モデルで学習させることで、文脈を理解し、人間が生成したような自然な文章生成を可能にする AI モデルである。代表的な例としては、GPT シリーズ (OpenAI 社) や BERT (Google 社) があり、質問応答、文章要約、翻訳、情報抽出といった自然言語処理タスクで顕著な成果を挙げている [1, 2]。

LLM の実用例は医療や法律分野をはじめ多岐に渡る。医療分野では、電子カルテから患者情報を効率よく抽出し、診断支援や個別化医療への活用が進められている [3, 4]。また画像診断においては、読影レポート解析を通じた病変の早期発見や治療効果評価の支援が行われている [5, 6]。法律分野でも大量の判例や契約書の解析に LLM が応用され、判例検索や契約内容のリスク評価を迅速かつ正確に行うことで、弁護士業務の効率化が図られている [7]。

創薬・化学分野でも LLM は積極的に導入されており、化合物の設計や薬理活性予測に関する研究が活発化している。実際、一昨年開催された「LLM 創薬チャレンジ」では、多くの研究グループが LLM を利用して標的蛋白質に対する活性化化合物候補を設計し、専門家による評価でも興味深い成果を示している [8]。

LLM の基盤技術として Transformer モデルがある。Transformer は 2017 年に Vaswani らによって提唱され [9]、自己注意機構 (Self-Attention) を核としている。この自己注意機構は、系列データ内の要素間の関係性を動的かつ効率的に学習することが可能であり、従来のリカレントニューラルネットワーク (Recurrent neural network; RNN) や長短期記憶モデル (Long Short Term Memory; LSTM) と比べ、長距離の依存関係を効率よく処理することを実現した [9]。

このような Transformer 技術の成功により、自然言語処理のみならず、生物学や化学分野でも LLM の活用が急速に進展している。アミノ酸配列や分子構造をテキスト形式として扱い、これらの構造や性質を予測する研究が盛んに行われている [10, 11]。

### 2. 蛋白質言語モデルと化学言語モデル

LLM の技術的発展は、生物学や化学分野にも大きな影響を与えている。その応用例の一つが「蛋白質言語モデル」である。蛋白質言語モデルは、蛋白質を構成するアミノ酸配列を自然言語のように扱い、Transformer を中心とする深層学習モデルを通じて配列データを解析するものである [10]。従来の実験的な解析法では、蛋白質の機能や構造を明らかにするために多くの時間やコストが必要だったが、蛋白質言語モデルの登場により、この状況が大きく改善された。Transformer を用いたこれらのモデルは、大量の蛋白質配列データを自己教師あり学習し、蛋白質の進化的関連性や機能的特徴を迅速かつ高精度で抽出する能力を持つ [10]。

具体的な蛋白質言語モデルとしては Meta 社が開発した ESM (Evolutionary Scale Modeling) シリーズや ProtBERT が挙げられる。ESM シリーズは、数億の蛋白質配列を Transformer で学習し、未知の蛋白質の三次元構造を実験に匹敵する精度で予測する能力を示している。また ProtBERT は、蛋白質配列データをもとにその生物学的機能や相互作用を精度高く分類・予測するモデルであり、創薬や機能解析研究において重要な役割を果たしつつある [10]。

一方、化学分野でも LLM を活用した新たな取り組みとして「化学言語モデル」が注目されている。化学言語モデルとは、化学構造を SMILES や SELFIES などのテキスト表現で記述し、Transformer を用いて大量の分子情報を効率よく学習する手法である [11]。この手法によって、化合物の特性や薬理活性を高精度で予測したり、新薬候補となる分子構造を自動生成したりする

ことが可能となった。例えば、ChemBERTa や MolGPT は、その典型例である。ChemBERTa は特に分子の物性予測において高精度を示し、新薬開発の初期段階での化合物スクリーニングにおいて重要なツールとなっている[11]。また MolGPT は、薬理特性を指定したうえで新規分子を迅速かつ自動的に生成する能力を示しており、創薬研究における分子設計のプロセスを劇的に効率化する可能性を持つ[12]。

このように、蛋白質言語モデルと化学言語モデルはそれぞれの分野で画期的な進展をもたらしているだけでなく、相互に組み合わせることで創薬研究において新たな道を拓くことが期待されている。

### 3. 蛋白質言語モデルと化学言語モデルの潜在空間への応用

蛋白質言語モデルおよび化学言語モデルが創薬分野に与える影響は大きく、その可能性を最大限に引き出すための鍵となるのが「潜在空間 (Latent Space)」である。潜在空間とは、深層学習モデルが入力データの本質的な特徴を抽出して圧縮した、多次元的で連続的な空間表現である。Transformer モデルに代表される自己教師あり学習手法を通じて生成された潜在空間では、蛋白質や化学物質の構造的および機能的特性がコンパクトかつ効率的に表現されている[13]。

蛋白質言語モデルにおいて、潜在空間は蛋白質の進化的関係や立体構造、機能に関する情報を効果的に捕捉する役割を果たしている。Transformer を用いた蛋白質言語モデルは、大規模なアミノ酸配列データを自己教師ありで学習し、その過程で蛋白質配列の潜在的特徴を高次元ベクトルとして表現する。このエンベディングと呼ばれる表現を通じて、蛋白質間の類似性を測定したり、未知の蛋白質が持つ可能性のある機能を予測したりすることが可能になっている[1, 2, 13]。例えば、疾患関連の変異が蛋白質の機能や安定性にどのような影響を及ぼすかを、この潜在空間の位置関係から予測することも試みられている。

一方、化学言語モデルにおける潜在空間は、新規化合物の生成や分子の特性予測に役立てられている。Transformer ベースのモデルが膨大な分子データを学習することによって得られる潜在空間は、分子構造の化学的特徴を連続的かつ体系的に表現している。このような表現により、特定の化合物の周辺に類似性を持つ新規な化合物を生成したり、目的とする薬理特性や毒性の低減などを狙った分子設計を行うことが可能になっている[3, 4, 14]。

蛋白質言語モデルと化学言語モデルを組み合わせることで、潜在空間の統合的な利用も可能である。これにより、標的となる蛋白質に対して高い活性や特異性を示す新薬候補を、従来よりも効率のかつ高精度で設計できると期待される。この統合された潜在空間の活用は、創薬プロセスの大幅な効率化および新規性の向上につながり得る。

潜在空間の活用において、Variational Autoencoder (VAE) は特に重要な役割を果たしている。VAE は深層学習の一種であり、入力データを潜在空間内の確率分布として表現し、データの生成や特徴抽出を効率的に行うことができるモデルである[15]。VAE の特徴は、データの特徴を確率的に学習することで、単なる圧縮ではなく、意味的に連続的で滑らかな潜在表現を可能にする点である。

創薬分野において VAE が注目される理由は、化合物の生成や最適化プロセスにおいて多様性と新規性を両立できるためである。化学言語モデルの一種として、分子構造を SMILES 表記などで入力し、VAE を通じて潜在空間における分子の特徴を学習することにより、新規な化合物を生成できる。例えば、特定の薬理活性を持つ分子群から共通特性を抽出し、その特性を持つ新規化合物を潜在空間内で探索するといった活用が可能となる[15, 16]。

蛋白質言語モデルとの連携においても、VAE は有効な手法である。標的蛋白質に対するリガンド結合部位の相互作用情報を VAE で潜在空間に埋め込むことで、その相互作用を効果的に捉えた化合物設計が可能となる。これにより、蛋白質と化合物間の高精度な相互作用予測や、目的の相互作用を持つ新規リガンドの設計を高速に進めることが可能になる。

このように、VAE は蛋白質・化学言語モデルの潜在空間を創薬研究において効果的に活用するための強力な手法となり得る。今後、VAE を組み合わせた創薬研究がさらに進展することで、従来の方法では困難だった薬剤設計や新薬開発の新たな局面を切り開くことが期待される。

#### 4. IEV2Mol : 化学言語モデルと潜在空間を応用した分子設計手法

我々は、化学言語モデルと潜在空間を応用した創薬アプローチの一つとして、IEV2Mol (Interaction Energy Vector to Molecule) を提案した (図 1)。IEV2Mol は、蛋白質とリガンド間の相互作用エネルギーを考慮して分子を生成するモデルである [17]。従来、分子生成手法において蛋白質とリガンド間の結合親和性を考慮することが一般的であり、相互作用を考慮する場合も、相互作用情報としてはバイナリ形式の相互作用フィンガープリント (Interaction Fingerprint; IFP) が使用されてきた [18]。IFP は相互作用の有無を二値的に示すものであり、相互作用の強さや質を定量的に表現することは困難であった。また、ポケット構造を考慮していても、相互作用を定量的に考慮した例はあまり報告されていない。

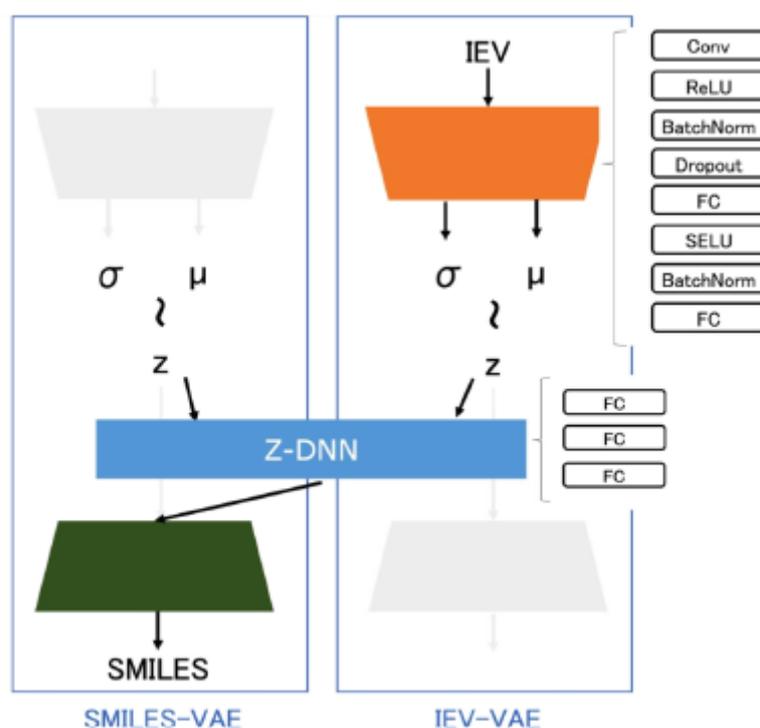


図 1. IEV2MOL のアーキテクチャ [17]

これに対し、IEV2Mol では蛋白質-リガンド相互作用を Interaction Energy Vector (IEV) という標的蛋白質とリガンド間の相互作用エネルギーを抽出したベクトルにより、相互作用の強度や特性をより詳細に捉えることを目指している。IEV は Yasuo らによって提案された指標であり、水素結合、静電相互作用、ファンデルワールス力などの相互作用タイプごとにドッキングシミュレーションで計算されたエネルギー強度を数値化し、ベクトル化したものである [19]。これを用いることで、蛋白質-リガンド間の相互作用を定量的に考慮できる。

IEV2Mol は、VAE を用いて、この IEV を潜在空間に埋め込んでいる。IEV を学習する VAE (IEV-VAE) と、化合物構造を SMILES 表記から学習する VAE (SMILES-VAE) の 2 種類のモデルを独立して訓練し、共有潜在空間において両者の情報を統合する。この統合は Z-DNN (3 層の畳み込みニューラルネットワーク) を介して行われ、これにより特定の標的蛋白質に望ましい相互作用特性を持つ化合物を生成できるよう設計されている。IEV2Mol の性能評価は、ドーパミン受容体 D2 (DRD2)、アデノシン受容体 A2 (AA2AR)、AKT セリン/スレオニンキナーゼ 1 (AKT1) の 3 つの標的蛋白質に対して行われた。既存の生成手法である JT-VAE や IFP-RNN と比較した結果、IEV2Mol が生成した化合物群は基準となる化合物の IEV に対して高い類似性を示し、構造的には異なる新規化合物も多数含まれていた。図 2 は、既知のヒット化合物の分布付近に化合物が生成されていることを示しており、IEV2Mol が標的蛋白質との相互作用情報を考慮することで、学習データとは異なる

る新規化合物の設計を可能にしていることを示している。また、生成化合物は化学的妥当性 (Validity)、一意性 (Uniqueness)、多様性 (Diversity) という評価指標において良好な結果を示した。また、生成した分子が標的蛋白質と良好なドッキング親和性を示すことも確認された [17]。

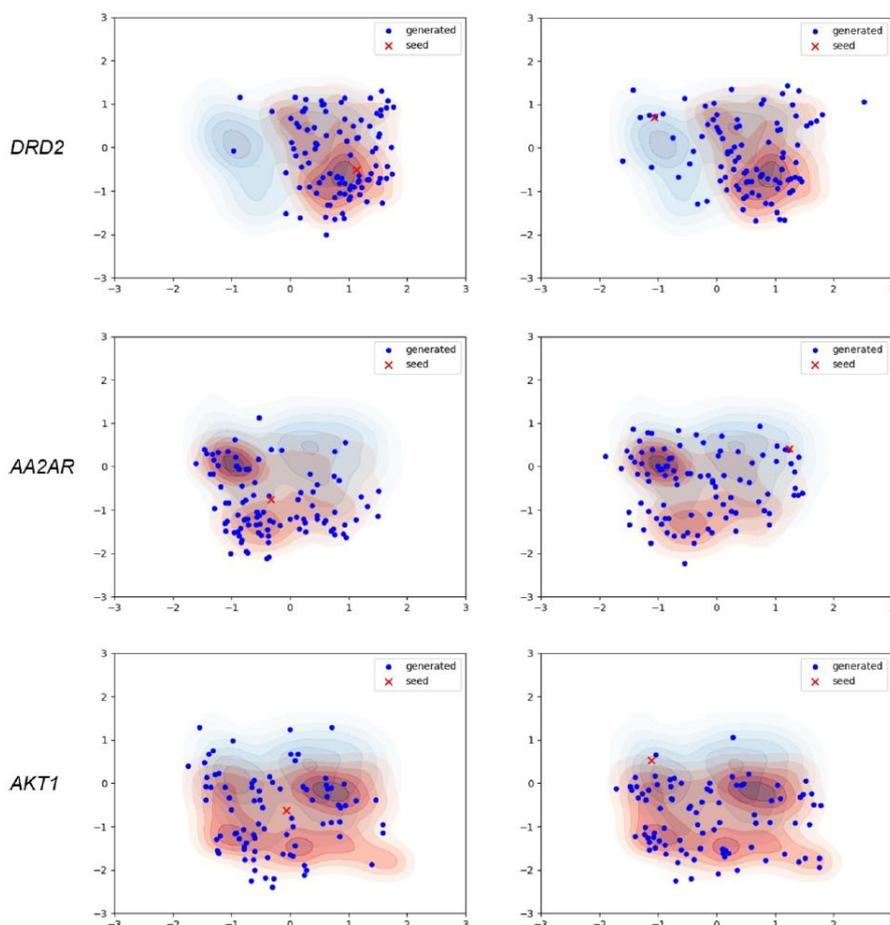


図 2. 標的蛋白質ごとの IEV2Mol 生成化合物の分布。ECFP4 フィンガープリント (2048 ビット) を主成分分析で二次元に削減して可視化。赤い十字はシード化合物、青の密度マップは DM-QP-1M データセットから無作為に抽出した 10,000 化合物、赤の密度マップは活性化化合物の分布。青点は各シード化合物から生成した化合物 [17]

## 5. LLM 創薬チャレンジの試み

LLM 創薬チャレンジは、LLM を創薬関連タスクに活用できるかどうかのフィージビリティスタディのために、SNS アカウント「創薬ちゃん (souyakuchan)」によって開催されたコンペティションである (図 3)。開催のきっかけは、2023 年 3 月に GPT-4 Technical Report [20] が OpenAI から公開された際に、ユースケースとして紹介されていた例の一つで「構造式の改変と開発業務受託機関 (Contract Research Organization) への注文」をさせたらできたという話があったのを受けて、「もはや LLM には創薬もできている！ シンギュラリティだ！」とネットの一部の層が盛り上がっていたという事案があり、「いや、実際のところどこまでできるのか？ (やれるとしても実務タスクのごく一部では?)」を検証するために本イベントを企画した、という経緯である。当時はまだチャット内でコード実行もできなかった段階であり、実際、化合物関連の化学計算などを ChatGPT にやらせてみるといい加減な数字を返してくることがほとんどであった。その頃 ChemCrow [21] などのプロジェクトも立ち上がったところで、当該グループの見解とし

ても、当時の LLM にできるのは簡単な置換基の付け替え程度で、直接 LLM に創薬化学計算をさせるというよりは外部ツール群を統合して処理を「手伝わせる」のが無難なアプローチであろう、といったことが言われていた [22]。

**叢雲くすり (創薬ちゃん) LLM創薬チャレン... @souyaku... · 3月30日**  
**#LLM創薬チャレンジ**  
**LLM が創薬している？ やれるものならやってほしい！**  
 .....というわけで LLM をどれだけ創薬に活かせるか大会を開催する。いつも通り賞金付き。今回は参加者全員に実費として ¥10k も支給する。  
 ルール: 画像参照、詳細→[github.com/souyakuchan/LLM](https://github.com/souyakuchan/LLM)...  
 参加登録: [forms.gle/1fAKmqHYp2BPQ2...](https://forms.gle/1fAKmqHYp2BPQ2...)

**LLM創薬チャレンジ**  
**総則**  
 ・ 所定の創薬標的 (本ページに詳細) に対し、大規模言語モデル (LLM) を利用して必ず LLM 自身が化合物の提案もしくは選別をするステップを含めたワークフローで10個の活性化化合物候補を提案すること。1参加者の応募は1セットまで。  
 ・ 提出デッドラインは 2023/04/30 24:00:00 (JST)  
 提出物:  
 ・ 実行可能なコード (Python/JavaScript/Javaなど)  
 ・ LLM との対話内容をマークダウン形式にしたもの  
 ・ GPT-4o/4o-mini などとアップロード  
 ・ 計算結果を LLM の外で実行した場合は、そのコード  
 ・ GitHub-Gist などとアップロード  
 ・ 提案の化合物 10個  
 ・ SMILES 形式で提出  
 ・ 評価方法: 参加者同士の相互評価と審査員による評価、定量指標を組み合わせる (後述)

**LLM創薬チャレンジ**  
**標的**  
 ・ CACHE #4 に乗っかる形で、CBL-Bタンパク質のTKBドメインを標的とする。  
 ・ 既知の結合化合物として約 900 個の特許化合物の情報が与えられている。  
 ・ それら既知化合物との構造類似性になるべく低く高活性な化合物候補を提案すること。\* 既知化合物は MorganSP での Tanimoto 値が 0.7 以下  
 ・ 既知のタンパク質-リガンド複合体構造あり。8GCY.PDB  
 → 同じ結合部位を狙うこと。

**LLM創薬チャレンジ**  
**総則**  
 ・ まずは参加登録 (参加フォーム) をすること。各参加者には参加報酬 (ホワイトマネー) として ¥10,000 ずつ支給する。\* 上限30名、先着順  
 ・ 参加報酬を受け取ったプレイヤーの義務:  
 ・ 提出デッドラインまでに提出物一式を提出すること  
 ・ 参加者全員分の提出化合物の評価を評価デッドラインまでに各自実施すること (他者からアップロード可)  
 ・ ルール:  
 ・ 本イベントは非営利活動。収益は全額、LLM 創薬研究会の活動に充てる。\* 収益は個人が MorganSP での Tanimoto 値が 0.7 以下  
 ・ 既知化合物の提案もしくは選別をさせる」ステップを入れること  
 ・ LLM に外部知識の入れ替え、embedding や finetune してよい  
 ・ LLM にできない任意の計算処理を LLM の外で実行してよいが、コードを LLM に生成させると、LLM を活用しているほど加算されるかもしれない (LLM 活用度) に加算  
 ・ 情報の変化が早いので、状況に応じて必要に応じてルール変更や追加を行う可能性があるが変更は行われること (LLM 活用度)  
 ・ 提案: 参加者間の相互評価と審査員評価の両方を基準点として、既知化合物との類似性指標 (LLM 活用度) を加算する  
 ・ 賞金はこれらが基準点で算出される。アップロードは必ずしも実施しない

**LLM創薬チャレンジ**  
**賞金**  
 得点上位 3 賞:  
 ¥80,000  
 ¥50,000  
 ¥30,000  
 ソーシャル創薬賞: 本イベント開催中に創薬に関する知見を Discussion で共有し、upvote 獲得数の多かった 3 名に授与。  
 ¥20,000 × 3 名  
 Discussion 投稿

図 3. LLM 創薬チャレンジ開催告知 [23]

チャレンジの基本ルールは、所定の創薬標的 (E3 ユビキチンリガーゼ CBL-B を採用した) に対し、LLM を利用して必ず LLM 自身が化合物の提案もしくは選別をするステップを含めたワークフローで、10 個の活性化化合物候補を提案せよ、というものである [24]。既知の結合化合物 (約 900 個の特許化合物) との構造類似性になるべく低く高活性なものの提案を目指してもらい、参加者同士の相互評価と審査員による評価、定量指標を組み合わせる順位を付けることとした。審査員は 5 名で、現役メドケムや計算化学者、構造生物学者を含むプロ達の布陣とした。評価項目としては、化学的な安定性・反応性、合成容易性、合成展開可能性、忌避構造の有無、活性がある期待度、の 5 項目と、手法の「LLM 活用度」を合わせた。採点作業用の Web アプリケーション (URL: [wisdom.chemical.space](https://wisdom.chemical.space)) を用意しており、今も公開中である。人智による構造式の評価ツールとしてご利用頂ける。

最終参加者数は 9 グループで、半分は院生で半分は研究職やその他の社会人であった。各参加者の専門分野・バックグラウンドは、ケモインフォマティクス、バイオインフォマティクス、薬学、有機化学合成、計算科学などから成り、「LLM 創薬チャレンジ」に比較的親和性が高いと思われる参加者層となった。各参加者から 10 構造式ずつ、合計 90 個の構造式が提出され、参加者同士の相互採点と審査員による採点が行われた。各参加者 (特に受賞者) の手法の詳細については本イベントに関する NeurIPS 2023 報告 [25] に譲るが、ここでは構造式の顔つきについて概観してみる (図 4)。今回、活性が既知の化合物が用意されていたわけだが、処理を LLM のみで完結させた参加者については、どうしても既知活性化化合物と似てしまうという傾向があった。一方、LLM と外部ツールの連携に有用性を見出した参加者たちの手法では、既知活性化化合物との類

似性が高くなく且つ採点上の評価も高いものが得られたという傾向があり、結果、受賞に至ったという構図となった。「LLM 創薬チャレンジ」の趣旨として、なるべく LLM を活用するという前提があったわけだが、実用的には外部ツールとの連携が必須であったと言えよう。

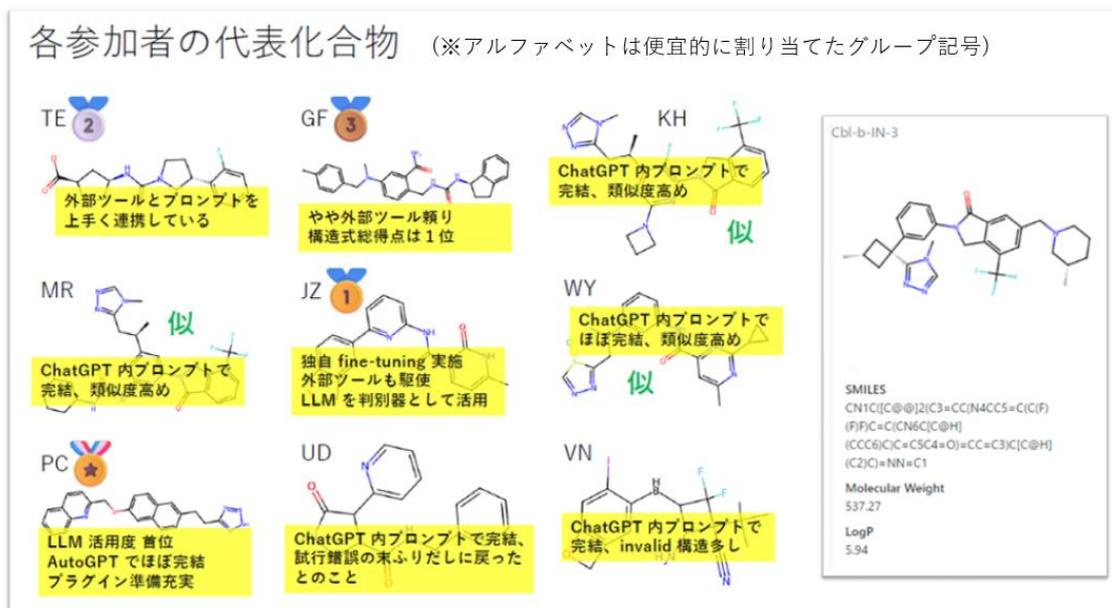


図 4. 参加者ごとの構造式の顔つきの傾向

本イベントの総括としては、少なくとも当時の LLM は「化学そのもの」を理解してはならず、計算も苦手なため、ツール群の統合のために使うのが無難な使い方であり、今後の進歩に期待するとすれば、構造式の表現や相互作用の表現を LLM-friendly な形にした上でそれに適したモデルアーキテクチャを開発し、化学ドメイン・創薬ドメインに特化させた LLM が必要であろう、といったところである。まずはデータ整備などにハードルはあるが、取り組んでいきたい。

また、参加者の皆さん、審査員の皆さん、そして本イベントに豪華副賞を提供してくれた株式会社アグロデザイン・スタジオ、iSiP 株式会社はこの場を借りて深く御礼申し上げます。

## 6. おわりに

本稿では、LLM が自然言語処理を超えて、蛋白質や化合物設計を含む創薬分野においても重要な役割を果たしつつあることを示した。化学言語モデルの VAE を用いた応用例として IEV2MOL を、LLM を創薬関連タスクに活用できるかどうかのフィージビリティスタディとして LLM 創薬チャレンジの事例を取り上げた。

LLM 創薬チャレンジは、専門分野やバックグラウンドが異なる参加者が、SNS を活用してリアルタイムに進行状況を共有しながら新たな分子設計を競うという、これまでの研究コンペとは異なるユニークな試みであった。また、参加者に報酬や副賞として焼き肉や焼き鳥が提供されるなど、イベント性や楽しさを重視した企画であり、多くの研究者に関心を持たれた。このような取り組みは、専門分野間の垣根を超えた新しい研究コミュニティ形成のきっかけとなり、創薬分野における LLM 活用の新たな可能性を広げることが示された。

しかし、創薬における LLM の応用はまだ初期段階であり、分子の物理化学的特性や生物学的活性の高精度な予測、実験データとの連携方法など、多くの課題も残っている。これらの課題を克服するためには、さらなる研究と技術開発が求められるだろう。今後、AI 技術の継続的な進展とともに、LLM を核とした手法が創薬プロセスにより一層深く組み込まれることにより、より迅速で効率的な新薬開発が可能となることを期待したい。

## 参考文献

- [1] Devlin, J., Chang, M. W., Lee, K., et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *arXiv preprint*, arXiv:1810.04805 (2018).
- [2] Brown, T. B., Mann, B., Ryder, N., et al. Language Models are Few-Shot Learners, *arXiv preprint*, arXiv:2005.14165 (2020).
- [3] Jiang, H., et al. (2023) Large Language Models in Healthcare: Applications, Challenges, and Future Prospects, *NPJ Digital Medicine*, **6**(1), 72.
- [4] Rasmy, L., Xiang, Y., Xie, Z., et al. Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction, *npj Digit. Med.*, **4**, 86 (2021).
- [5] Singhal, K., Azizi S., Tu, T., et al. Large language models encode clinical knowledge, *Nature*, **620**(7972), 172-180 (2023).
- [6] Moor, M., Banerjee, O., Abad, Z. S. H., et al. Foundation Models for Generalist Medical Artificial Intelligence, *Nature*, **616**(7956), 259-265 (2023).
- [7] Bommasani, R., Hudson, D. A., Adeli, E., et al. On the Opportunities and Risks of Foundation Models, *arXiv preprint*, arXiv:2108.07258 (2021).
- [8] LLM Drug Discovery Challenge 2023. Available at: <https://zenn.dev/retrosynthchan/articles/23b1a664f23046> (Accessed: 2025).
- [9] Vaswani, A., Shazeer, N., Parmar, N., et al. Attention is All You Need, *Advances in Neural Information Processing Systems*, **30**, 5998-6008 (2017).
- [10] Rives, A., Meier, J., Sercu, T., et al. Biological Structure and Function Emerge from Scaling Unsupervised Learning to 250 Million Protein Sequences, *Proc. Natl. Acad. Sci. USA*, **118**(15), e2016239118 (2021).
- [11] Chithrananda, S., Grand, G., Ramsundar, B. ChemBERTa: Large-Scale Self-Supervised Pretraining for Molecular Property Prediction, *arXiv preprint*, arXiv:2010.09885 (2020).
- [12] Bagal, V., Aggarwal, R., Priyakumar, U. D. MolGPT: Molecular Generation Using a Transformer-Decoder Model, *J. Chem. Inf. Model.*, **62**(9), 2064-2076 (2022).
- [13] Alley, E. C., Khimulya, G., Biswas, S., et al. Unified rational protein engineering with sequence-based deep representation learning, *Nat. Methods*, **16**(12), 1315-1322 (2019).
- [14] Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., et al. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules, *ACS Cent. Sci.*, **4**(2), 268-276 (2018).
- [15] Kingma, D. P., Welling, M. Auto-Encoding Variational Bayes, *arXiv preprint*, arXiv:1312.6114 (2013).
- [16] Jin, W., Barzilay, R., Jaakkola, T. Junction Tree Variational Autoencoder for Molecular Graph Generation, *arXiv preprint*, arXiv:1802.04364 (2018).
- [17] Ozawa, M., Nakamura, S., Yasuo, N., et al. IEV2Mol: Molecular Generative Model Considering Protein-Ligand Interaction Energy Vectors, *J. Chem. Inf. Model.*, **64**(18), 6969-6978 (2024).
- [18] Zhang, J., Chen, H. De Novo Molecule Design Using Molecular Generative Models Constrained by Ligand-Protein Interactions, *J. Chem. Inf. Model.*, **62**(14), 3291-3306 (2022).
- [19] Yasuo, N., Sekijima, M. Improved Method of Structure-Based Virtual Screening via Interaction-Energy-Based Learning, *J. Chem. Inf. Model.*, **59**(3), 1050-1061 (2019).
- [20] Achiam, J., Adler, S., Agarwal, S., et al. Gpt-4 technical report, *arXiv preprint*, arXiv:2303.08774 (2023).
- [21] Bran, A. M., Cox, S., Schilter, O., et al. Augmenting large language models with chemistry tools, *Nat. Mach. Intell.*, **6**(5), 525-535 (2024)
- [22] <https://x.com/andrewwhite01/status/1636215660542849025>
- [23] <https://x.com/souyakuchan/status/1641259176214298625>
- [24] [https://github.com/souyakuchan/LLM\\_DD\\_Challenge](https://github.com/souyakuchan/LLM_DD_Challenge)
- [25] Murakumo, K., Yoshikawa, N., Rikimaru, K., et al. LLM drug discovery challenge: A contest as a feasibility study on the utilization of large language models in medicinal chemistry, AI for Accelerated Materials Design-NeurIPS 2023 Workshop (2023).

//// Cutting Edge ////

## タンパク質言語モデルの創薬分野での利用

東京科学大学 大上雅史

### 1. はじめに

タンパク質言語モデル (Protein Language Model, PLM) は、自然言語処理分野で発展した言語モデル技術をタンパク質のアミノ酸配列解析に応用するものである。近年、公共データベースに登録されたタンパク質配列数は急増しており (現在、数億配列規模)、この膨大な進化的多様性を単一のモデルで学習する試みが行われている [1]。実際、配列データからのみ学習させた PLM がタンパク質の様々な情報 (アミノ酸の物理化学的性質、遠縁の相同性、2次構造や立体構造など) を捉えていることが示されている。PLM によって、変異の効果予測や残基間接触予測など様々なタスクで従来以上の精度が達成されている。PLM はこのように配列中に隠れた進化・構造・機能情報を抽出する強力な手法であり、バイオインフォマティクスや計算創薬における新たなパラダイムを形成しつつある [2]。本稿では、PLM の発展と意義について概観し、その基盤となる言語モデル技術、代表的モデル (ESM-2 や SaProt) の特徴、創薬研究や抗体医薬設計への応用例、そして今後の展望と課題について論じる。

### 2. 言語モデルの基本とタンパク質配列解析との関係

言語モデル (Language Model) とは、トークン (単語や文字など) からなる系列データに対し確率分布を定義し、その系列中の統計的規則性を学習するモデルである。すなわち、系列  $X = (x_1, x_2, \dots, x_L)$  に対してモデルが確率  $P(X)$  を与える (もしくは各位置での条件付き確率  $P(x_i | x_1, \dots, x_{i-1})$  を与える)。自然言語処理において言語モデルは機械翻訳や質問応答など幅広いタスクで利用されているが、一方の生物学分野では、プロファイル HMM (Hidden Markov Model) などが言語モデルに相当し、配列の位置ごとの残基出現確率を利用して相同性検索などに使われてきた歴史がある。深層学習に基づく言語モデルは、より長距離で複雑なトークン間の依存性を捉えられる点で有利であり、既知の全タンパク質配列をまとめて学習できるため、従来手法の限界を超える新たな可能性を示している [3]。

タンパク質配列は 20 種程度のアミノ酸文字の並びで表現される点で「言語」のようにも捉えることができ、実際「配列の統計には進化によって選択された構造・機能情報が刻まれている」という考え方は古くから提唱されてきた。これは言語学における分布仮説、「単語の意味はその出現文脈によって決まる」という考え方に通じるものであり、大量のタンパク質配列データから自己教師あり学習によってその「文法」を機械が学習できると期待される。実際、タンパク質配列も自然言語と同様に長距離の相互依存 (離れたアミノ酸同士の共変異や残基間相互作用) を含むため、Transformer [4] に代表されるような文脈全体を考慮できるアーキテクチャがタンパク質配列解析にも適していると考えられる。このようにタンパク質配列を「文章」とみなして言語モデルで学習させることで、配列に内在する進化規則を統計的に捉え、配列を高次元のベクトル表現にエンコードして構造や機能を予測したり、ある配列の出現確率 (尤度) からその配列変異の適応度を評価したりすることが可能となる [5]。

### 3. 自然言語の言語モデル

現代の言語モデルの多くは、2017 年に提案された Transformer アーキテクチャ [4] に基づいている。Transformer は自己注意機構により系列中の任意の位置同士の関係を動的に学習でき、並列計算によって長い系列でも効率よく学習可能にした点で画期的であった。Transformer の登場以降、自然言語処理分野では RNN (Recurrent Neural Network) や LSTM (Long Short-Term Memory) に代わって、Transformer ベースのモデルが主流となっている。中でも、大規模コーパスで事前

学習した汎用言語モデルが様々なタスクで飛躍的な性能向上をもたらすことが示され、GPT や BERT に代表される手法が確立された [6]。例えば OpenAI の GPT シリーズは系列の前方文脈から次のトークンを順次予測する方式（自己回帰型言語モデル）を採用しており、与えられたテキストに続く文章を自然に生成することに優れる。Google の BERT では系列中の一部トークンを隠して（マスクして）入力し、隠した単語を文脈から推定する事前学習を行うマスク言語モデルが採用されており、双方向の文脈情報を活用できるため優れた表現学習が可能になることが示されている。GPT 系は生成に、BERT 系はエンコーディングに強みがあり、タンパク質配列モデルでもこの違いを活かした応用がなされている。例えば ProGen [5] は GPT に類似した自己回帰型モデルでタンパク質配列を一文字ずつ生成することで新規配列創出に利用されており、実際に構造的・機能的に意味のあるタンパク質配列を生成できることが報告されている。ESM や ProtBERT など多くの PLM は BERT 流のマスク予測で事前学習されており、生成よりも配列表現の汎用性に重点を置いている。

モデルの大型化も言語モデル発展の重要な要素である。自然言語処理ではパラメータ数を増やしデータ量を拡充することで性能が向上し続ける傾向が報告されており、タンパク質の場合も基本的には同様である [2]。ただし計算資源とのトレードオフも大きく、どこまで大型化すべきかについては効率性の観点から議論がある。

#### 4. タンパク質言語モデル (PLM)

PLM では、タンパク質配列を「言語」と見立てて学習する。具体的には、大規模なタンパク質配列データベース (UniProt) やメタゲノム由来の配列データ (BFD など) をコーパスとして、自己教師あり学習によってモデルの事前学習を行う。この過程でモデルは配列中の進化的パターンを学習し、各配列を高次元ベクトルに埋め込み、進化・構造・機能に関する情報を内包した分散表現（埋め込み表現とも呼ばれる）を獲得する。得られたモデルは様々なタスクに転用可能であり、分類や回帰モデルの特徴量として用いたり、新規配列へのスコアリング（尤度評価、ゼロショット学習）に用いたりすることができる。ESM-1 モデルでは UniRef 等から得られた 2.5 億本のタンパク質配列を学習データとして、Transformer ベースの言語モデルを BERT 風に訓練し、配列間の遠縁の相同性検出や、二次構造・三次構造などの構造情報が自然に抽出されることを報告している [1]。このようにラベル無しの配列データのみで学習できることが PLM の利点であり、未知機能タンパク質やメタゲノムコンティグ配列（種が不明だがタンパク質だと思われる配列）も含め、利用可能な全配列情報を活用して汎用的な「知識」を引き出すことができる。得られたモデルはタスクに応じてファインチューニング（微調整）したり、埋め込み後のベクトルを入力とした一般的な教師あり学習を行うことで、特定の予測問題（天然変性領域予測、機能予測、局在予測、変異の影響予測など）に適用することが可能であり、多くの場合に精度の向上が見込める。

一方で、配列以外の情報を組み合わせる方法も検討されている。多重配列アラインメントから共進化情報を直接学習するモデルや、立体構造情報のアノテーションを組み込んで PLM 学習を行うモデルも登場している。後者の例の 1 つに SaProt があり、標準的な PLM では配列から構造を明示的に学習しないという限界に着目し、学習時に構造をエンコードしたトークンを追加することで性能向上を実現している [7]。

#### 5. 代表的なタンパク質言語モデル

##### 5.1 ESM-2 (Evolutionary Scale Modeling 2)

ESM-2 は Meta (旧 Facebook) 社の研究チームによって 2022 年に開発された PLM であり、2019 年に公開された ESM-1b の後継モデルである。Transformer を基盤とし、アミノ酸配列に対するマスク学習で構築された。ESM-2 はパラメータ数の異なる複数のモデルが用意されており、800 万 (8M)、3500 万 (35M)、1 億 5000 万 (150M)、6 億 5000 万 (650M)、30 億 (3B)、150 億 (15B) のうち好きなサイズのモデルを利用することができる（一般にパラメータ数の大きいモデルのほうが性能は良いが、利用する計算機環境の要求スペック（特に GPU VRAM）が高くなることに注意が必要である） [8]。

ESM-2 を用いた代表的な研究に ESMFold がある [9]。高速なタンパク質構造予測ツールとして知られる ESMFold は、AlphaFold2 [10]などが必要としていた多重配列アラインメントの検索を一切行わず、単一配列から直接タンパク質の立体構造を予測することに成功している。これにより数百残基程度のタンパク質であれば1分以内に構造予測が可能である。この速度を活かし、Meta社はESM Metagenome Atlasというメタゲノム由来タンパク質配列の予測構造集を公開している [9]。なお、構造予測の精度については、CASP等の性能検証の結果から ESMFold は AlphaFold2 にやや劣るとされている [11]。

## 5.2 SaProt (Structure-aware Protein Language Model)

SaProt は 2024 年に報告された新しい PLM であり、配列に加えて構造情報を取り入れたモデルである [7]。従来の PLM が配列情報のみを用いていたのに対し、SaProt では構造認識語彙 (structure-aware vocabulary) という手法を導入し、アミノ酸残基トークンに対応する立体構造上の情報を表す追加トークンを組み合わせて学習を行っている。具体的には、Foldseek [12] が内部で持っている VQ-VAE (Vector Quantised-Variational AutoEncoder) で射影される残基ごとの立体構造のアルファベット 20 種類を用いて、構造アルファベット  $20 \times$  アミノ酸アルファベット 20 の計 400 種類 (実際には<MASK>を入れて  $21 \times 21 = 441$  種類) のトークンを定義し、このトークン列を用いて言語モデル学習を行う。構造情報を組み込むことで、SaProt は各残基の立体的コンテキストを考慮した表現学習が可能となり、結果として多数のタスクで精度向上が報告されている。

## 6. インシリコ創薬への応用

PLM は薬剤標的のタンパク質に関する様々なプロセスで活用が期待されている。ここでは (1) 構造予測での利用、(2) バーチャルスクリーニングへの活用、(3) 変異効果予測、(4) 抗体医薬設計への活用を紹介し、応用可能性を述べる。

### 6.1 構造予測での利用

PLM が使われている ESMFold の構造予測は、前章で述べたとおり素早く予測構造を構築して創薬研究に供することができる。ただし、計算時間のアドバンテージは AlphaFold3 の登場によって多少薄れており、また ESMFold の予測精度は AlphaFold2 や AlphaFold3 に劣ることに注意が必要である。ESMFold に限らず、予測構造をタンパク質-リガンドドッキングや自由エネルギー摂動法計算による結合自由エネルギー推定などの化合物スクリーニングのために利用する取り組みが増えている [13, 14]。

### 6.2 バーチャルスクリーニングへの活用

前述のドッキング計算のような立体構造ベースのバーチャルスクリーニングのほか、立体構造を陽に用いずに配列情報から予測を行うリガンドベースのバーチャルスクリーニングにも PLM は活用されている。タンパク質配列から PLM で得た埋め込みベクトルと、化合物の情報 (例えば SMILES) の両方を用いて、化合物の結合活性を予測するモデルが開発されている [15]。Lamらが提案した BIND と呼ばれるモデルでは、スクリーニング性能が従来の構造ベース手法に匹敵する性能であると報告されている [15]。

### 6.3 変異効果予測

創薬標的となるタンパク質の変異が構造や機能に与える影響を予測することも重要な課題である。この問題に対し、PLM が「タンパク質として望ましい配列の並び」を学習できていると仮定すると、PLM によって与えられた配列に対する「(タンパク質としての) 文法の適切具合」を確率 (尤度) として評価できる。すなわち、ある変異が入った配列の尤度が大きく低下した場合、その変異はそのタンパク質にとって望ましくないと予測できる [16]。ESM-1b を改良した ESM-1v は、変異による機能変化予測に用いられており、変異株の影響の予測などに応用されている [17]。

## 6.4 抗体医薬設計への活用

抗体医薬品の台頭により、抗体の配列デザインや最適化もタンパク質工学の重要なテーマとなっている。抗体は抗原との親和性を決める領域として  $V_H$  および  $V_L$  領域内の CDR (Complementarity-Determining Region) ループを持つ一方で、フレームワークと呼ばれる領域では保存配列も多い。抗体独自の配列構造は一般のタンパク質とは異なる統計的特徴を持つため、抗体配列に特化した PLM (抗体言語モデル) が開発されている。ここでは、代表的な抗体言語モデルである AbLang、IgLM、REALM を紹介する。

- AbLang は Oxford 大学のグループが開発した抗体言語モデルであり、抗体のデータベースである Observed Antibody Space (OAS) の約 5 億件の配列を用いて訓練されている [18]。抗体言語モデルの有用性を示すタスクとして、シーケンスエラーなどで一部が欠失した抗体配列の復元という問題の正答率を評価しており、AbLang を用いることで OAS 中の約 40% の配列で欠失している N 末端 15 残基を高精度に推定できることを示した (ESM-1b による復元よりも優れていた)。
- IgLM は Johns Hopkins 大学のグループが開発した抗体言語モデルであり、 $V_H \cdot V_L$  のペア配列約 5.58 億件を学習して構築されている [19]。IgLM は学習時に「 $V_H$  または  $V_L$ 」「生物種 (ヒト、マウス、ラットなど)」のラベル情報も用いているため、例えば配列生成に用いる際にはこれらの条件を指定して生成することができる。
- REALM は筆者らが開発した抗体言語モデルであり、SaProt と同様に抗体の部位ごとのアノテーションを組み合わせてトークン化し、学習したものである [20]。製造上重要な抗体の変性温度の予測タスクなどで AbLang を上回る性能を示している。

抗原への親和性予測やエピトープ予測、適切な CDR 配列の生成、凝集や変性などの製造に関わる物性の予測、免疫原性の予測など、抗体医薬品に関連するタスクは多岐にわたる。もし抗体言語モデルが抗体の配列と機能をうまく結ぶことができているならば、近い将来これらのタスクが解決され、抗体医薬の理論的設計を実現することが期待される。

## 7. おわりに

PLM はタンパク質配列から膨大な情報を引き出す強力なツールとして台頭し、バイオインフォマティクス・計算創薬の分野に新風をもたらしている。しかし、使いこなすには相応の計算機スペックが要求されることと、埋め込み表現や予測モデルの解釈性に関する問題が存在する。また、単純に配列だけで学習するという方法だけでなく、SaProt のように事前知識を統合するという工夫の余地も大いに残されている。

なお、自然言語の大規模言語モデルが大量データとパラメータ数の増加によって高性能を得たのと同様に、PLM もまた大規模化の一途を辿っている。Meta 社が 2024 年に発表した ESM-3 [21] および ESM Cambrian [22] はより多くのデータ/パラメータで性能向上が図られており、今後様々なタスクへの利用が予想される。

ここで重要になるのは実験による検証との連携である。PLM が提示した予測や設計配列を実験的に検証・フィードバックすることで、新たな予測手法の開発・改善に繋がることが期待される。予測タスクが比較的簡便に解決できるようになった現在においては、創薬の現場で求められている、実際に役立つようなタスクを現実的に定義するという研究者の能力がより重要となるだろう。

なお、本稿の草稿に ChatGPT o3-mini Deep Research を活用した。

## 参考文献

- [1] Rives, A., Meier, J., Sercu, T., et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences, *Proc. Natl. Acad. Sci. USA*, **118**, e2016239118 (2021). doi: 10.1073/pnas.2016239118
- [2] Bepler, T., Berger, B. Learning the protein language: evolution, structure, and function, *Cell Syst.*, **12**, 654-669.e3 (2021). doi: 10.1016/j.cels.2021.05.017

- [3] Alley, E. C., Khimulya, G., Biswas, S., et al. Unified rational protein engineering with sequence-based deep representation learning, *Nat. Methods*, **16**, 1315-1322 (2019). doi: 10.1038/s41592-019-0598-1
- [4] Vaswani, A., Shazeer, N., Parmar, N., et al. Attention is all you need, *Adv. Neural Inf. Process. Syst.*, **30**, 5998-6008 (2017). doi: 10.5555/3295222.3295349
- [5] Madani, A., McCann, B., Naik, N., et al. ProGen: language modeling for protein generation, *arXiv*, 2004.03497 (2020).
- [6] Devlin, J., Chang, M.-W., Lee, K., et al. BERT: pre-training of deep bidirectional transformers for language understanding, *arXiv*, 1810.04805v2, (2018).
- [7] Su, J., Han, C., Zhou, Y., et al. SaProt: protein language modeling with structure-aware vocabulary, *Proc ICLR 2024*, (2024). <https://openreview.net/forum?id=6MRm3G4NiU>
- [8] Facebook Research. ESM available models, GitHub. <https://github.com/facebookresearch/esm?tab=readme-ov-file#available-models>
- [9] Lin, Z., Akin, H., Rao, R., et al. Evolutionary-scale prediction of atomic-level protein structure with a language model, *Science*, **379**, 1123-1130 (2023).
- [10] Jumper, J., Evans, R., Pritzel, A., et al. Highly accurate protein structure prediction with AlphaFold, *Nature*, **596**, 583-589 (2021). doi: 10.1038/s41586-021-03819-2
- [11] Elofsson, A. Progress at protein structure prediction, as seen in CASP15, *Curr. Opin. Struct. Biol.*, **80**, 102594 (2023). doi: 10.1016/j.sbi.2023.102594
- [12] van Kempen, M., Kim, S. S., Tumescheit, C., et al. Fast and accurate protein structure search with Foldseek, *Nat. Biotechnol.*, **42**, 243-246 (2024). doi: 10.1038/s41587-023-01773-0
- [13] Díaz-Rovira, A. M., Martín, H., Beuming, T., et al. Are deep learning structural models sufficiently accurate for virtual screening? application of docking algorithms to AlphaFold2 predicted structures. *J. Chem. Inf. Model.* **63**, 1668–1674 (2023) doi: 10.1021/acs.jcim.2c01270
- [14] Furui, K., Ohue, M. Benchmarking helixfold3-predicted holo structures for relative free energy perturbation calculations, *ACS Omega*. (2025). doi: 10.1021/acsomega.4c11413
- [15] Lam, H. Y. I., Guan, J. S., Ong, X. E., et al. Protein language models are performant in structure-free virtual screening, *Brief. Bioinform.*, **25**, bbae480 (2024). doi: 10.1093/bib/bbae480
- [16] Hie, B., Zhong, E. D., Berger, B., et al. Learning the language of viral evolution and escape, *Science*, **371**, 284-288 (2021). doi: 10.1126/science.abd7331
- [17] Meier, J., Rao, R., Verkuil, R., et al. Language models enable zero-shot prediction of the effects of mutations on protein function, *Proc NIPS '21*, 29287-29303 (2021). doi: 10.5555/3540261.3542504
- [18] Olsen, T. H., Moal, I. H., Deane, C. M. AbLang: an antibody language model for completing antibody sequences, *Bioinform. Adv.*, **2**, vbac046 (2022). doi: 10.1093/bioadv/vbac046
- [19] Shuai, R. W., Ruffolo, J. A., Gray, J. J. IgLM: infilling language modeling for antibody sequence design, *Cell Syst.*, **14**, 979-989.e4 (2023). doi: 10.1016/j.cels.2023.10.001
- [20] Nishino, T., Kato, N., Tsutaoka, T., et al. REALM: region-empowered antibody language model for antibody property prediction, *Proc IEEE BIBM 2024*, 7104-7106 (2024). doi: 10.1109/BIBM62325.2024.10822666
- [21] Hayes, T., Rao, R., Akin, H., et al. Simulating 500 million years of evolution with a language model, *Science*, **387**, 850-858 (2025). doi: 10.1126/science.ads0018
- [22] ESM Team. ESM Cambrian: revealing the mysteries of proteins with unsupervised learning, EvolutionaryScale Website, (2024). <https://evolutionaryscale.ai/blog/esm-cambrian>

//// Cutting Edge ////

## 大規模言語モデルによる科学的知識の獲得と応用： 化学分野における最新の研究事例

株式会社サイキンソー 山崎広之

### 1. はじめに

近年、深層学習をはじめとする人工知能 (AI: Artificial Intelligence) 技術の発展により、大規模言語モデル (LLM: Large Language Models) の性能が飛躍的に向上している。特に GPT (Generative Pre-trained Transformer) や BERT (Bidirectional Encoder Representations from Transformers) といったモデルは自然言語処理 (NLP: Natural Language Processing) タスクにおいて、従来の手法を大きく上回る成果を示している。LLM は、言語理解、文章生成、情報検索などの幅広い応用が期待される一方で、その学習プロセスや利用方法に関して多くの課題も抱えている。

本稿では LLM の基盤技術を概説した上でその学習プロセスや強化学習技術について解説する。さらに LLM の応用領域として、埋め込み表現、検索拡張生成、AI エージェントなどの技術についても解説し、科学および化学分野における適用事例を紹介する。

また、LLM の活用には情報管理やバイアスの問題、誤情報の生成 (ハルシネーション) のリスクなどが伴う。本稿では、これらの課題についても議論し、LLM の信頼性向上や安全な運用のための方向性についても言及する。

### 2. LLM の基盤技術と学習プロセス

#### 2.1 LLM の基盤技術

LLM の基本的なアーキテクチャとして、トランスフォーマー (Transformer) と呼ばれるニューラルネットワークが用いられる。従来のリカレントニューラルネットワーク (RNN) や畳み込みニューラルネットワーク (CNN) と比較し、トランスフォーマーは自己注意 (Self-Attention) 機構を採用しており、長文の文脈を効率的に捉えることができる。この技術の発展により、GPT や BERT などのモデルが開発され、文章の理解や生成において高い性能を発揮している [1,2]。

トランスフォーマーモデルの基本的な構造は、Vaswani らによって 2017 年に発表された論文 "Attention Is All You Need" で提案された [3]。この論文では、エンコーダとデコーダという 2 つの主要な構成要素からなる深層ニューラルネットワーク (DNN) としてトランスフォーマーが紹介されている。

BERT モデルは、この基本的なトランスフォーマーアーキテクチャを拡張し、双方向の文脈理解を可能にした革新的なアプローチを採用している。一方、GPT シリーズは、トランスフォーマーの一方方向型モデルを基盤として発展し、大規模な言語モデルの可能性を示した。これらのモデルの登場により、自然言語処理タスクの性能が飛躍的に向上し、後述の様々な応用分野での活用が進んでいる。

#### 2.2 LLM の学習プロセス

LLM の学習プロセスは大きく以下の 3 つに分かれる。

##### 2.2.1 事前学習 (Pre-training)

インターネット上の膨大なテキストデータを用いて、言語の統計的パターン、すなわち、単語の出現確率や文章の構造を学習し、文脈に基づいた推論能力を獲得する。例えば、2022 年に

発表された GPT-3.5 は事前学習データとして 約 570 GB のテキストデータを用いて、特定のタスクに依存しない汎用的な言語モデルが形成された。

## 2.2.2 ファインチューニング (Fine-tuning)

事前学習にて得られた汎用的な言語モデルに対して特定のタスクやドメインに特化したデータを用いて追加学習を行うことをファインチューニングと呼ぶ。このプロセスにより、例えば、医療、法務、金融そして化学などの事前学習にあまり含まれていない専門領域に関するデータセットを用いて追加学習を行うことでそれぞれの分野に関する正確な出力が可能となる。

また特定の人物やキャラクターがよく話すセリフや口調をデータセットとして学習させることによりその人物やキャラクターを模倣させた受け答えをさせることもできる。

一方で、ファインチューニングにも欠点があり、与えるデータセットの偏りが原因でモデルもその偏りを学習してしまう、特定のデータに過剰にフィットしすぎてしまい新しいデータや異なるタスクに対する汎化性が低下するといったことを引き起こす可能性があり、ファインチューニング後に性能評価を行う必要がある。

化学分野のデータを用いた LLM のファインチューニングの取り組みはこれまでも行われており、いくつかを例として紹介すると、Frey らは MOSES や ANI-1x データセットを用いて GPT-3 に対してファインチューニングを行い ChemGPT を構築した [4]。また Zhang らは PubChem, ChEMBL, ZINC などのインターネット上のデータを用いて Two-stage instruction tuning と呼ばれるファインチューニング法を適用し ChemLLM を構築することによって、ユーザーからの化学分野の質問への応答の性能が向上することを明らかにした [5]。

さらに Yu らは大規模な SMolInstruct と呼ばれるデータセットを構築し Galactica 6.7B, Llama 2, Code Llama, Mistral の複数の LLM をそれぞれファインチューニングし、「分子の物性を予測する」、「与えられた反応物と試薬に基づき生成物を提案する」といったタスクを実行させ性能の比較を行った。その結果 Mistral をファインチューニングした LLM モデル LlaSMolMistral が他の LLM をファインチューニングした LLM モデルや前述の ChemLLM を超え、最も高い性能を示すことを明らかにした [6]。

これらのことは化学分野においても LLM のファインチューニングが有効であることを示していると言える。

## 2.2.3 強化学習 (Reinforcement learning)

強化学習とは エージェント (AI) が試行錯誤しながら報酬を最大化する行動方針を学習する手法であり、2016 年に囲碁のプロ棋士を破った Google 社の AlphaGo [7] や化学の分野では分子設計ソフト REINVENT [8] に強化学習が用いられている。近年、NLP の分野でも活用が進んでおり、大規模言語モデル LLM の最適化にも応用されている。

特に LLM のチューニングにおいては、人間のフィードバックを模倣する報酬モデルを活用し、生成結果の質を向上させる目的で用いられる。例えば、OpenAI 社の ChatGPT では、人間の好みに基づく学習 (RLHF: Reinforcement Learning from Human Feedback) を用いて、より自然で有用な応答を生成するモデルが開発されている [9]。

さらに 2025 年 1 月に DeepSeek 社が発表した DeepSeek-R1 は、LLM のチューニングに特化した強化学習手法を導入したことで注目を集めた。DeepSeek-R1 では、PPO (Proximal Policy Optimization) の発展形である GRPO (Group Relative Policy Optimization) を適用し、より効率的な報酬最適化を実現している [10]。

興味深い点として、従来の膨大な学習データが必要な事前学習と異なり、DeepSeek-R1 は初期の学習データに限られているにもかかわらず、GRPO を用いることで、OpenAI 社の LLM モデルである o1 に匹敵する高い推論能力を示した [11]。これは、強化学習がファインチューニングとは異なるアプローチでモデル性能を向上させる可能性を示しており、今後の LLM 開発においても重要な役割を果たすと考えられる。

### 3. LLM の主な応用領域

本章では、LLM の主な応用技術である埋め込み表現、検索拡張生成 (RAG: Retrieval-Augmented Generation) および コンテキスト拡張生成 (CAG: Context-Augmented Generation)、AI エージェントの概要を説明した後、科学および化学における適用事例について解説する。

#### 3.1 埋め込み表現 (Embedding)

LLM における埋め込み表現 (以下、Embedding) では、単語、文、ドキュメントあるいはユーザーの入力情報を高次元のベクトル空間にマッピングし、数値情報へと変換する。数値情報へと変換してしまえば、数値情報に対して一般的に用いられている類似度計算、クラスタリングなどを適用することができる。これにより、例えば、完全一致、部分一致検索などと異なり、多少表現が異なってもおおよそ同じ意味の文章であると判断することができ、検索や推薦システムにおいて重要な役割を果たす。

Embedding の応用事例として以下のようなものがあげられる。

- 検索拡張生成
  - Embedding を使用して関連情報の検索精度を向上させ、より適切な回答生成を可能にする。詳しくは 3.2 にて説明する。
- テキスト分類
  - メールのスパムフィルタリングなど、テキストを数値情報に変換して分類アルゴリズムに適用する。
- 感情分析
  - 顧客レビューなどのテキストを Embedding にて変換し、感情分析モデルに入力することで、より正確な感情分析が可能になる。

Embedding の科学および化学における適用の試みは LLM 以前にもあり、タンパク質の構造情報をテキストとして扱った Pro2Vec、化学構造の情報をテキストとして扱った Mol2vec があった [12,13]。Mol2vec にてベクトル情報にした後、ベクトル空間上で距離が近い (つまり類似である) 化学構造は溶解度や薬効などの性質も近くスキップホールドホッピングができるのではないかと期待がされていた。さらに Kim らはファインチューニングした StructLLM 単独および StructLLM と Postive-Unlabeld 学習モデルを組み合わせることでバーチャルな結晶構造の合成容易性を予測できることを示した [14]。特に LLM を用いることで Bonding、Coordination など合成容易性の理由を列挙することが可能になり、提案した学習モデルの説明可能性を強調した。

#### 3.2 検索拡張生成 (RAG)

LLM はその性質から事前学習で得た情報のみでしか回答できず、例えば 2023 年に発表された ChatGPT-3.5 では「日本の 2025 年現在の日本の首相は誰であるか？」といったような直近の情報には当然のことではあるが正しい回答を行うことができない。回答できないだけならまだしもハルシネーションと呼ばれる、「さも正しい知識であるかのように振る舞い誤った回答をする」現象まで見られてしまうことがある [15]。

このような問題に対応するための一つに先述のファインチューニングがあるが、一方で検索拡張生成 (RAG) と呼ばれる手法もある [16]。RAG では以下のような手順を踏むことでユーザーへの LLM の回答を構築している (図 1)。

1. LLM に対する追加の情報を 3.1 で述べた Embedding を用いてベクトル情報へと変換しデータベースへ保存する。
2. ユーザーから質問が送信される。
3. LLM はユーザーからの質問を Embedding にてベクトル情報へ変換し、類似性検索などでデータベースに対して検索を行う。

4. 検索ヒットした質問に関連した情報を LLM へ返す。
5. LLM は抽出された情報の優先度付けなどを行い、上位の重要とみなした情報を要約する。なおこの際、これらの情報を持って回答できない場合は回答できないと返すように指示することでハルシネーションを起こす可能性を減らすことができる。
6. 要約をユーザーへの回答として出力する。

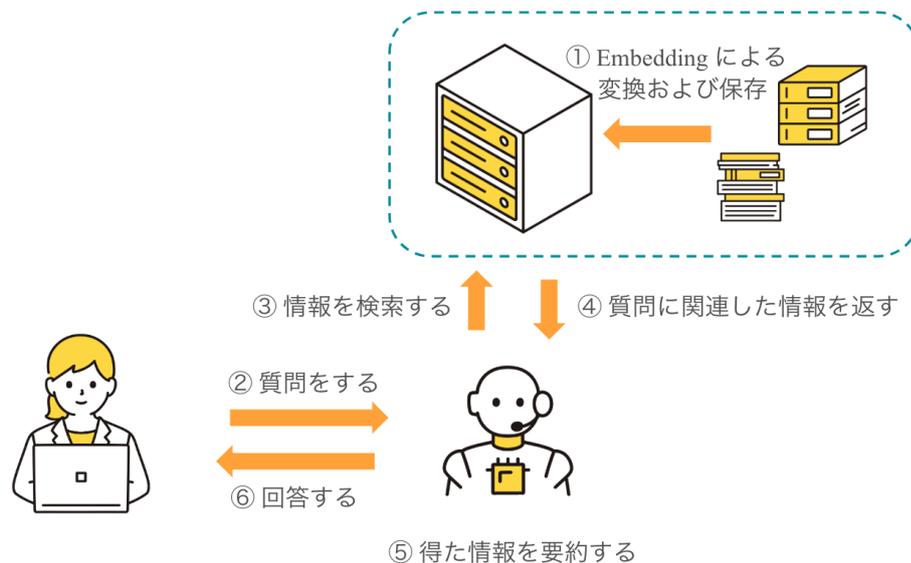


図 1. RAG を利用したワークフロー例

RAG により、ファインチューニングを行わずに LLM は事前学習以後の新規の情報や事前学習に含まれていない専門性のある情報を含み回答することができ、同時にハルシネーションが起こる可能性を低くすることができるようになった。また、RAG のように単純に LLM に質問を投げかけ回答を得るのではなく、別途 LLM に投げかけるシステムは後述の AI エージェントでも見られる。

RAG の応用事例として以下のようなものがあげられる。

- FAQ・ナレッジベース検索
  - 企業の内部文書を RAG によって検索し、LLM に要約や回答を生成させる。例えば LINE ヤフーは社内向け独自業務効率化ツールとして「SeekAI」を導入し業務効率化を図っている [17]。
- 対話型アシスタント
  - ユーザーからの問い合わせに対し RAG を活用して最新の情報を取得し、より適切な回答を提供する。例えば、近畿大学では従来の問い合わせ対応 AI チャットボットに RAG を導入することで、高度で柔軟な回答を可能とした [18]。

RAG は LLM の適用事例の中でも様々な検討がされており、例えば RAG を発展させたものとして CAG がある [19]。RAG の短所として、検索に時間がかかることがある、検索ヒット文章の優先度付けを誤ってしまうことなどが挙げられる。一方、CAG では Cache という言葉の文字通り、最初にすべての必要な情報を LLM にロードしておくことでこれらの短所をなくす工夫がされている。もちろん膨大なデータベースの場合はすべてを LLM にロードすることはできないので、ケースによっては RAG/CAG をハイブリッドさせて利用することも提案されている。

さらに、マルコフ決定過程と RAG を組み合わせることでデータベースでの検索が必要かどうかを判定する仕組みを持つ DeepRAG と呼ばれる手法も存在する [20]。DeepRAG では人間は

自身の情報が不足していると判断した時に Web 検索などで情報を追加で得ることを模倣し、LLM でも検索を行う前に Web やデータベースでの追加の検索が必要かどうかを判定する機能が追加されている。これにより応答の正確性を改善させ、また必要以上に検索に時間をかけるのを防止することができると考えられている。

また、RAG の考え方を応用しハルシネーションをより防ぐ取り組みとして Google 社によって提案された RIG (Retrieval Interleaved Generation) がある [21]。RAG では LLM がユーザーからの回答を生成する際にデータベースに必要な情報を検索するが、RIG ではユーザーへの回答を生成した後に回答の信頼性を担保するためにデータベース検索をする仕組みが作られており、これによりハルシネーションが起こる可能性を低くすることができた。

このように非常に発展を進めている RAG だが、化学の分野での応用として LLaMP があげられる [22]。LLaMP は Materials Project と呼ばれる結晶材料データベースを参照する RAG であり、化学分野でも RAG を導入することによりハルシネーションを抑えることを示している。

さらに Reed らは RAG と MIPRO (Multiprompt Instruction PProposal Optimizer) と呼ばれる手法を用いることでやはり化学分野でハルシネーションを抑えることができることを示している [23]。LLM を用いるのに不可欠なものとしてプロンプトがある。プロンプトは LLM に対して適切な応答を得るために与える指示文であり、例えば「ニトロベンゼンを部分構造に含む化合物を社内データベースから検索してください。」と言うプロンプトと、末尾にさらに「ただし、ない場合は『なし』と回答してください。」と追加したプロンプトでは後者のほうが正しく回答する事が多く、どのようなプロンプトにするべきかは非常に重要と言える。

MIPRO は LLM のプロンプトを最適化する手法の一つであり、人間がトライアンドエラーで行っていたプロンプトのチューニングをより効率的に行う手法であり、ハルシネーションを抑えることができるプロンプトの作成を試みている。

### 3.3. AI エージェント

AI エージェントとは、LLM を活用し、ユーザーの目的達成のために自律的に動作するシステムである。AI エージェントには Web 検索機能、データベース機能、コーディング機能や別途パソコンにインストールされているソフトウェアを実行する機能などが付与されており、ユーザーからタスクの依頼を受けた後、自問自答するようにシステムの内部にて「タスクを実行するためには自身に備わっている機能のうちどれをどのように使えばよいか？」を考え、ステップバイステップでタスクを実行していく。これまでの研究で単純にクエリを受け結果を返す場合よりも期待される結果を返すことが明らかになっており、業務の効率化や意思決定の支援を行う。

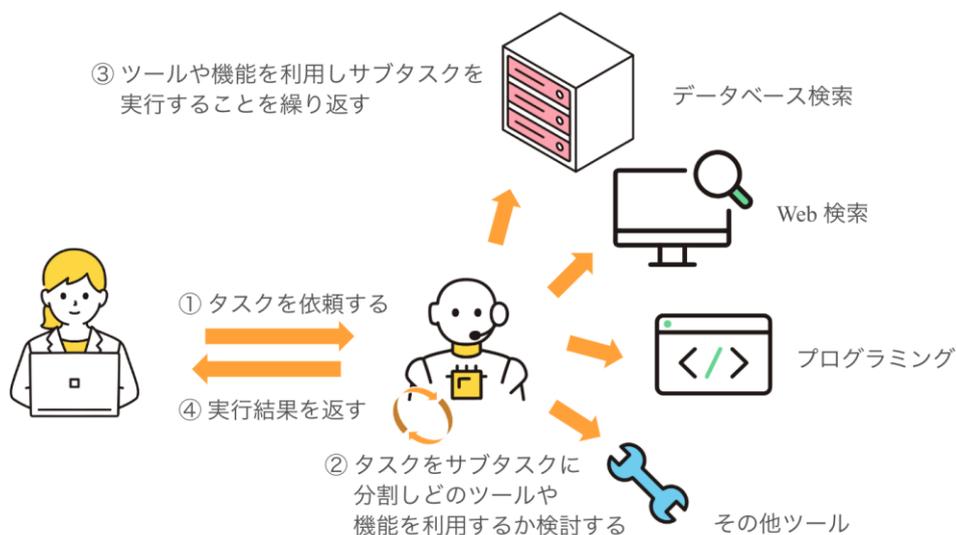


図 2. AI エージェントのワークフロー例

AI エージェントの応用事例として 2025 年 2 月に OpenAI から発表されたものとして Operator と DeepResearch がある [24,25]。

Operator は Web 検索機能などを有した AI エージェントであり、ホテルの予約やインターネット上での買い物を依頼することができる。

また DeepResearch では高い推論能力と Web 検索機能を有しており「ある分野における学術論文をピックアップしレポートにまとめてほしい」といった依頼に期待通り答えると話題になっている(筆者が書いているこの記事も瞬く間に書いてしまうのかもしれない)。

プログラミングコードのコードディングも AI エージェントによる恩恵を受け Cursor [26], Cline [27] と呼ばれるサービス/ソフトウェアが発表されることにより、それまでは Github Copilot [28] と呼ばれるソフトウェアなどの入力している途中のコードの続きを予測しあらかじめ表示する、いわゆるコードの補完が主な LLM の利活用であったのが、ユーザーから「自身のポートフォリオを表示するサイトを実装して」といった依頼をすると複数のファイルにまたがりコードを提案するといったことも行ってくれるようになった。

一方、AI エージェントの科学および化学の分野での利用例として最初に提案されたものの一つとして ChemCrow がある [29]。ChemCrow は LLM を科学の分野で利用できないかとトライアンドエラーがはじまった 2023 年 4 月に発表され AI エージェントが使うツールとしてパテントの情報や Web 検索機能を含み科学者であるユーザーからの質問に応答するような仕組みが作られていた。

ChemCrow が発表された当時は OpenAI 社の ChatGPT-4 と呼ばれる LLM モデルが発表された直後であり、筆者は創薬におけるタスクにどう利活用できるかを競う LLM 創薬チャレンジコンテストに参加し、既知活性化合物の構造に対して 3.1 で述べた Embedding を ChatGPT-3.5 や ChatGPT-4 を用いて計算し、類似性検索に用いたり説明変数として機械学習モデルに用いたりした。

しかし類似性検索、機械学習どちらも期待した結果を得ることができず、LLM の別の利活用として ChemCrow とは別の AutoGPT [30] と呼ばれる AI エージェントにドッキングや分子生成を行う機能を追加し LLM に指示をすることでそれらを行うことが可能であること示した [31,32]。

当時は LLM にクエリとして投げることでできるデータサイズの制限などもあり他の参加者も苦戦していたが現在 2025 年にまたコンテストがあるとどのような利活用が新たに提案できるのであろうかと考えると非常に面白い。

科学の分野での応用に話を戻すと、Ren らの CRES<sub>t</sub> には音声認識機能があり、研究室内で科学者が CRES<sub>t</sub> に話しかけるとデータベース検索を人間の代わりに行ったり実験の提案をしてもらったりする姿が見られる [33]。

また、AI エージェントのシステムの中であたかも複数人の人物が存在するかのように振る舞わせることでより期待される回答をユーザーに返す傾向があることがわかってきている。例えば Swanson らの Virtual Lab [34] では Principal Investigator や Scientific Critic ら 5 人にも及ぶ役割をアサインされた仮想の人物が存在し、例えばユーザーからの依頼に対して Machine Learning Specialist と Computational Biologist がディスカッションを繰り返すことでタスクをブラッシュアップし、最終的に提案されたものを Principal Investigator が評価し人間に返すといったことを行っている。

このような取り組みは人間の世界でも有識者同士が十分にディスカッションを繰り返すことでより洗練された回答を得ることができると考えると自然な流れなのかもしれない。Swanson らは最終的に Virtual Lab で提案されたナノボディを実験により合成することを達成している。

複数人の人物が存在するかのような振る舞いをする同様のシステムは Ma らや Stewart らにも提案されている (図 3) [35,36]。

チームリーダーと人間が  
ディスカッションを行う



図3. AI エージェントで採用されていたアーキテクチャ例。  
人間とエキスパートやエキスパート同士がディスカッションを行う  
仕組みになっている。

日本では高齢化の影響により労働人口が減っていていることを考えると人間の代わりをできる AI エージェントが存在することは非常にありがたく、将来、グループリーダーは人間で部下は全員 AI エージェントであるという未来が来るかもしれない。その場合、人間同様 AI エージェントにどう指示をすると的確な指示ができるのかが今後必要とされるかもしれない。例えば、科学の分野ではないが、ジュニアレベル相当のプログラマの能力を持つ Devin と呼ばれる AI エージェントが発表および販売されており、ユーザーが指示したコーディングやバグの修正などが可能であり、すでに業務で活用している企業も見られるようになっている [37]。

#### 4. LLM 利用時の問題点

これまで LLM がどのようなものであるかからどのように応用されているかを説明してきたが、製薬企業での利用時の問題点やまだ達成できていないことにも最後に触れておきたい。

##### 4.1. 情報漏洩のリスク

LLM の活用が広がる中で、特に企業や研究機関における機密情報の取り扱いが課題となっている。LLM は製薬企業においても、文献レビューの自動化、新薬候補の発見、臨床試験データの解析、規制対応のサポートなど、多くの分野での活用が期待されている一方で、未発表の研究成果や特許出願中の情報が外部に流出するリスクを抑えることが求められる。

多くの LLM サービスでは、ユーザーの入力データをモデルの改善に活用することがあるため、利用規約を確認し、必要に応じてデータの保存・学習に関する設定を適切なものにするのが重要である。

情報漏洩のリスクを軽減するための具体的な対策の一つとして、エンタープライズ向けの LLM サービスの活用があげられる。Microsoft 社の Azure OpenAI Service, AWS 社の Bedrock, Google 社の Vertex AI など、企業向けに提供される LLM サービスでは、データの保持・学習を制限できるオプションがあり、プライベート環境も提供していることが多い [38-40]。

また LLM を開発している企業や団体の中には、モデル自体をオープンソースとして公開している場合がある。このように公開されているモデルを自社のプライベート環境で運用することで外部へのデータ送信を防ぐことができる。ただし、ファインチューニングを行ったり 3 章で紹介した RAG や AI エージェントを運用するといったことまで行おうとするとある程度のコンピュータリソースやサービスを保守・運用するエンジニアといったリソースが必要になってくる。当然だがこれらを自社内で持つのは外部のサービスを利用することとのトレードオフになるので AI をどう利活用するかは企業で一体となって考えていく必要があることと言える。

## 4.2. 化学構造の理解

これまでの LLM の科学および化学への適用事例を通して、LLM はテキスト情報で与えられた化学構造の性質や特性をある程度理解しているのではないかと考えられるが、一方で Yan らは LLM は分子の表現への理解に対する整合性がない可能性があることを指摘している。Yan らは複数の化学構造を SMILES 文字列と IUPAC 命名法でそれぞれ表し、物性の予測などを LLM に指示した場合同じ結果が得られるかを調査しており一致率は 1 % 未満であることを明らかにした。LLM のファインチューニングによりある程度性能は改善したが、今後、記号的な分子表現と真の化学的理解のギャップを埋め、化学タスクにおける高い正確性と整合性の両立が可能になることが期待される [41]。

## 5. まとめ

本稿では、大規模言語モデル (LLM) の基本構造と学習プロセスを概説し、科学および化学分野における応用について詳述した。特に、埋め込み表現 (Embedding)、検索拡張生成 (RAG)、および AI エージェントの発展と実際の適用事例について紹介した。他の分野も含め広く浅い紹介であったため、より科学の分野を深く理解したい場合は Ramos らのレビューなどに目を通すとよいと思う [42]。

今後の展望として LLM と専門的なデータベース (Protein Data Bank, Materials Project など) をして製薬企業の自社データベースを網羅的に統合することで、より高度な予測・設計が可能になることや、RAG や AI エージェントを用いることで複雑なタスクを自律的に実行できることが期待される。

結論として LLM は科学および化学の分野において極めて有望な技術であり、その適用範囲は今後も拡大していくことが予想される。一方で、情報管理やモデルの信頼性向上といった課題に対応しながら、安全かつ効果的に活用していくことが求められる。

## 参考文献

- [1] Radford, A., Narasimhan, K., Salimans, T., et al. Improving language understanding by generative pre-training, *Proceedings of a Conference* (2018). [https://cdn.openai.com/research-covers/language-unsupervised/language\\_understanding\\_paper.pdf](https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf).
- [2] Devlin, J., Chang, M. W., Lee, K., et al. BERT: Pre-training of deep bidirectional transformers for language understanding, *arXiv* (2019). doi: 10.48550/arXiv.1810.04805.
- [3] Vaswani, A., Shazeer, N., Parmar, N., et al. Attention is all you need, *arXiv* (2017). doi: 10.48550/arXiv.1706.03762.
- [4] Frey, N. C., Soklaski, R., Axelrod, S. et al. Neural scaling of deep chemical models, *Nat. Mach. Intell.*, **5**, 1297–1305 (2023). doi: 10.1038/s42256-023-00740-3.
- [5] Zhang, D., Liu, W., Tan, Q., et al. ChemLLM: A chemical large language model, *arXiv* (2024). doi: 10.48550/arXiv.2402.06852.
- [6] Yu, B., Baker, F. N., Chen, Z. et al. LlaSMol: Advancing large language models for chemistry with a large-scale, comprehensive, high-quality instruction tuning dataset, *arXiv* (2024). doi: 10.48550/arXiv.2402.09391.
- [7] Silver, D., Huang, A., Maddison, C. et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **529**, 484–489 (2016). doi: 10.1038/nature16961.
- [8] Olivecrona, M., Blaschke, T., Engkvist, O. et al. Molecular de-novo design through deep reinforcement learning, *J. Cheminform.*, **9**, 48 (2017). doi: 10.1186/s13321-017-0235-x.
- [9] Ouyang, L., Wu, J., Jiang, X. et al. Training language models to follow instructions with human feedback, *arXiv* (2022). doi: 10.48550/arXiv.2203.02155.
- [10] DeepSeek-AI, Guo, D., Yang, D., et al. DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning, *arXiv* (2025). doi: 10.48550/arXiv.2501.12948.
- [11] [https://x.com/deepseek\\_ai/status/1881318130334814301](https://x.com/deepseek_ai/status/1881318130334814301).
- [12] Asgari, E., Mofrad, M. R. K. Continuous distributed representation of biological sequences, *PLoS ONE*, **10** (11), e0141287 (2015). doi: 10.1371/journal.pone.0141287.
- [13] Jaeger, S., Fulle, S., Turk, S. Mol2vec: Unsupervised machine learning approach with chemical intuition, *J. Chem. Inf. Model.*, **58** (1), 27–35 (2018). doi: 10.1021/acs.jcim.7b00616.

- [14] Kim, S., Schrier, J., Jung, Y. Explainable synthesizability prediction of inorganic crystal polymorphs using large language models, *chemRxiv* (2025). doi: 10.26434/chemrxiv-2024-ltncz-v3.
- [15] Huang, L., Yu, W., Ma, W., et al. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions, *ACM Trans. Inf. Syst.* (2024). doi: 10.1145/3703155.
- [16] Lewis, P., Perez, E., Piktus, A., et al. Retrieval-augmented generation for knowledge-intensive NLP tasks, *NeurIPS 2020*, *arXiv* (2020). doi: 10.48550/arXiv.2005.11401.
- [17] <https://www.lycorp.co.jp/ja/news/release/008806/>
- [18] <https://prtimes.jp/main/html/rd/p/000000164.000018339.html>
- [19] Chan, B. J., Chen, C. T., Cheng, J. H., et al. Don't do RAG: When cache-augmented generation is all you need for knowledge tasks, *arXiv* (2024). doi: 10.48550/arXiv.2412.15605.
- [20] Guan, X., Zeng, J., Meng, F., et al. DeepRAG: Thinking to retrieval step by step for large language models, *arXiv* (2025). doi: 10.48550/arXiv.2502.01142.
- [21] Radhakrishnan, P., Chen, J., Xu, B., et al. Knowing when to ask – Bridging large language models and data, *arXiv* (2024). doi: 10.48550/arXiv.2409.13741.
- [22] Chiang, Y., Hsieh, E., Chou, C. H., et al. LLaMP: Large language model made powerful for high-fidelity materials knowledge retrieval and distillation, *arXiv* (2024). doi: 10.48550/arXiv.2401.17244.
- [23] Reed, S. Augmented and programmatically optimized LLM prompts reduce chemical hallucinations, *chemRxiv* (2025). doi: 10.26434/chemrxiv-2025-rwgt8.
- [24] <https://openai.com/index/introducing-operator/>
- [25] <https://openai.com/index/introducing-deep-research/>
- [26] <https://www.cursor.com/ja>
- [27] <https://github.com/cline/cline>
- [28] <https://docs.github.com/ja/copilot>
- [29] Bran, A. M., Cox, S., Schilter, O., et al. Augmenting large language models with chemistry tools, *Nat. Mach. Intell.*, **6**, 525–535 (2024). doi: 10.1038/s42256-024-00832-8.
- [30] <https://github.com/Significant-Gravitas/AutoGPT>
- [31] Murakumo, K., Yoshikawa, N., Rikimaru, K., et al. LLM Drug Discovery Challenge: A contest as a feasibility study on the utilization of large language models in medicinal chemistry, *NeurIPS 2023* (2023), <https://openreview.net/pdf?id=kjUyIvko18>.
- [32] <https://x.com/souyakuchan/status/1641259176214298625>
- [33] Ren, Z., Zhang, Z., Tian, Y., et al. CRESSt – Copilot for real-world experimental scientist, *chemRxiv* (2023). doi: 10.26434/chemrxiv-2023-tnz1x-v4.
- [34] Swanson, K., Wu, W., Bulaong, N. L., et al. The Virtual Lab: AI agents design new SARS-CoV-2 nanobodies with experimental validation, *bioRxiv* (2024). doi: 10.1101/2024.11.11.623004.
- [35] Ma, K. AI agents in chemical research: An intelligent research assistant system, *chemRxiv* (2024). doi: 10.26434/chemrxiv-2024-6tv8c-v2.
- [36] Stewart, I., Buehler, M. J. Molecular analysis and design using generative artificial intelligence via multi-agent modeling, *Mol. Syst. Des. Eng.* (2024). doi: 10.1039/D4ME00174E.
- [37] <https://devin.ai/>
- [38] <https://azure.microsoft.com/ja-jp/products/ai-services/openai-service>
- [39] <https://aws.amazon.com/jp/bedrock/>
- [40] <https://cloud.google.com/vertex-ai?hl=ja>
- [41] Yan, B., Chen, A., Cho, K. Inconsistency of LLMs in molecular representations, *chemRxiv* (2024). doi: 10.26434/chemrxiv-2024-lnvbz.
- [42] Ramos, M. C., Collison, C. J., White, A. D. A review of large language models and autonomous agents in chemistry, *Chem. Sci.*, **16**(6), 2514-2572, (2025). doi: 10.1039/D4SC03921A.

//// Cutting Edge ////

---

## AI for Science と AI Scientist

fuku 株式会社 山田 涼太

---

### 1. はじめに

AI (Artificial Intelligence: 人工知能) の科学研究への活用は長い歴史を持つ。1956年にダートマス会議で初めて「Artificial Intelligence」[1]という言葉が登場して以来、様々な科学分野において様々な切り口で研究されてきた。

近年の基盤モデル (Foundation Model) の発展は、より広範な課題での AI 活用を後押ししている。従来の深層学習では、画像やテキストに意味づけをした大量のラベル付きデータが必要であり、その準備には多大なコストがかかっていた。一方、基盤モデルは膨大な生データ (ラベルのついていないデータ) を元に学習することで汎用的な能力を獲得している。例えば、従来の深層学習では、テキストの文書分類、情報抽出、要約、翻訳をそれぞれ行うために、個別のデータセットを用意し、専用モデルを学習させる必要があった。これに対し、LLM (Large Language Model: 大規模言語モデル) などの言語を扱う基盤モデルでは、プロンプト (モデルに与える指示や入力) に「以下のテキストを要約せよ」と命令すれば要約結果が得られ、「翻訳せよ」と命令すれば翻訳結果が得られる。このような基盤モデルの発展は、特に科学研究など専門家によるアノテーション (データへの注釈付け) が必要で、データセット構築に多大なコストがかかる領域での AI 活用を大きく加速させている。

現在、本テーマの盛り上がりが増している要因の1つは、AI スタートアップ Sakana AI が2024年8月に発表した The AI Scientist[2]である。これは仮説生成、実験、評価、論文執筆までを自律的に遂行するシステムであり、彼ら曰く「完全に自動化された科学的発見のための最初の包括的なフレームワーク」である。論文内では機械学習研究における、仮説生成から論文執筆までの結果を示し、生命科学・材料科学での応用も可能であると結んでいる (如何にして物理的な実験環境とシステムを接続するかという途方もなく大きな課題は残っている)。

これまで科学研究への AI の活用は AI for Science や AI for Scientific Discovery と銘打たれることが多かった。しかし、LLM の性能向上により擬人的にタスクを遂行する AI Agent (以下、Agent) が実現できるようになり、あらゆる領域で擬人化されたシステムに注目が集まるようになった。そのような背景のもと、科学研究においても例外に漏れず The AI Scientist が登場し、追従する形で Google からは2025年2月に AI co-scientist[3]という AI アシスタントが発表された。

著者は AI Scientist の概念に通じる領域で活動してきた一人である。東京大学農学部にて食物アレルギーの研究をしていたが、過去の文献を活用することでより効率的な研究ができるのではないかと思い立ち2017年同大工学部に転学部し、NLP (Natural Language Processing: 自然言語処理) を利用した論文解析に取り組み始めた。在学中の2018年に fuku 株式会社を創業し、AI による実験条件と実験結果の抽出およびデータベース化サービスを開発した。現在は広く科学研究の自動化に貢献するためのシステム開発を行っている。その立場から科学研究の自動化に関してこれまでの取り組みと、生成 AI の登場における変化を述べたい。

本稿ではこれまで科学研究にどのような形で AI が活用されてきたか AI for Science の取り組みをまとめた上で、Agent の台頭により生まれてきた AI Scientist という概念にも触れる。

### 2. AI for Science

科学研究への AI の活用は、有機化学における化合物の分子構造の推定[4]や数学における理論形成の自動化[5]に始まり、現在では多様な分野において仮説生成、実験計画、データ収集、分析など科学研究のあらゆるプロセスで研究されている (図1)。

AI for Science は AI の知識と対象分野の知識の双方が求められるため、学際的な取り組みになる。そのため異なる専門分野の研究者同士で協働するためにワークショップやシンポジウムが開

かれています。機械学習分野の国際会議である NeurIPS (Conference on Neural Information Processing Systems) と ICML (International Conference on Machine Learning) では 2021 年から AI for Science と題したワークショップを開いており、分野横断的に知見の共有を行なっている。国内においても JST 未来社会創造事業と JST ムーンショット型研究開発事業の代表 4 名が立ち上げた AI ロボット駆動科学イニシアティブが 2023 年 7 月からシンポジウムと研究会を開催している。

AI for Science に関する、化学、生物学、コンピューターサイエンス、材料科学、物理学などの分野における研究事例については AI for Science コミュニティがブログ記事にまとめているためそちらに譲る[7]。

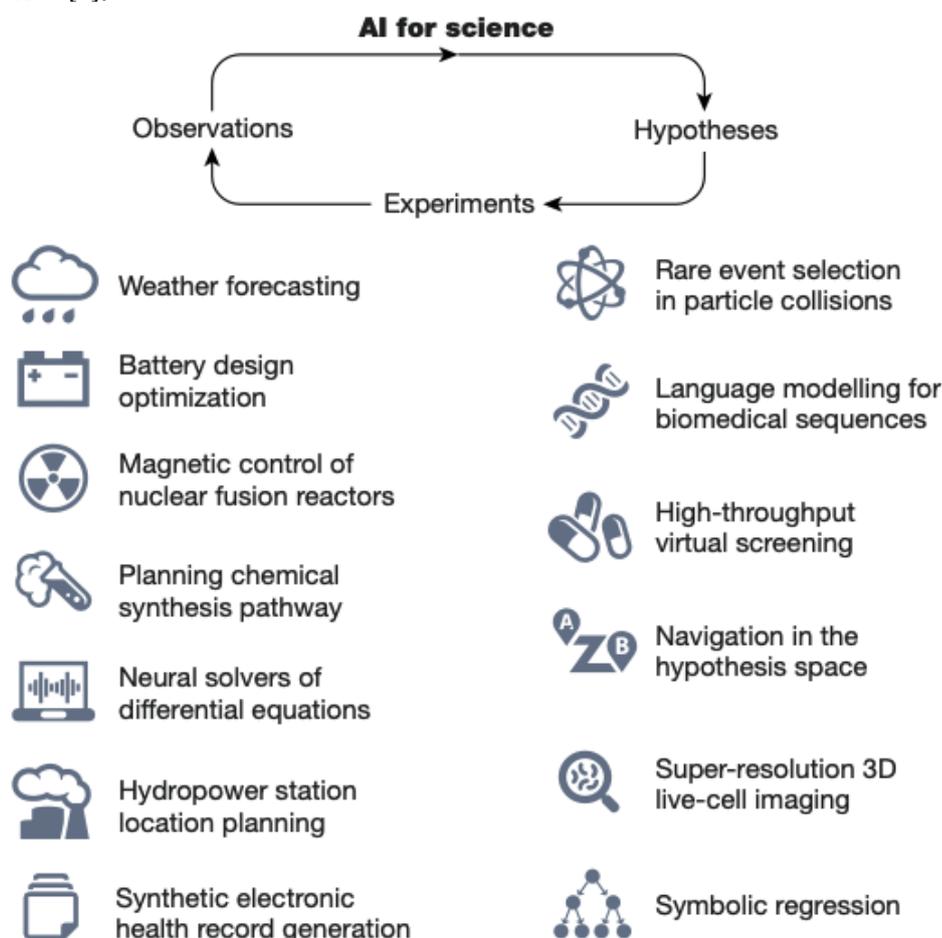


図 1. Science in the age of artificial intelligence.[6]

## 2.1 科学研究のための基盤モデル

基盤モデルの科学応用について、各領域で顕著な成果が挙げられている。DeepMind は数学分野において AlphaProof と AlphaGeometry2 を開発[8]し、国際数学オリンピックの問題を解決して銀メダル相当の成績を達成した。Meta は数学・制御理論分野でリアプノフ関数発見モデルを開発[9]し、力学系の安定性解析に貢献した。Wang らは進化的アルゴリズムに LLM を組み込んだ MOLLEO を開発[10]し、分子探索において既存手法を上回る性能を達成した。

基盤モデルを科学研究に適用する研究が盛んな一方で、科学研究のための基盤モデル (Scientific Foundation Models) も登場している。タンパク質の配列を言語とみなしたタンパク質言語モデル ESM3[11]は、自然界に存在しない機能性タンパク質の設計を可能にした。DNA 配列を言語とみなした DNABERT[12]は、転写因子結合部位の予測など遺伝子発現制御機構の解明に利用できる可能性を示した。気候科学では 100 万時間に及ぶ地球システムデータで事前学習した Aurora[13]は気象予測や海洋波のモデリングなどに利用できる。日本国内では理化学研究所が科

学研究基盤モデル開発プログラム（Advanced General Intelligence for Science Program: TRIP-AGIS）を立ち上げている[14]。

### 3. AI Scientist

前項の AI for Science では科学研究の一部のプロセスを AI によって自動化した取り組みを紹介した。これらの AI システムは、直接的にはプログラムから利用されるとしても、どのように利用するか意思決定は人間の科学者が行う。これに対して、エージェント技術の発展を基盤として実現された、あたかも人間のように自律的に計画を立てて研究を遂行する AI Scientist という概念が注目されている。

#### 3.1 Agent

無論 Agent の研究はそれ以前から続けられてきたが、ここでは昨今登場した基盤モデルを搭載した AI Agent のことを指して Agent と呼ぶ。Google は Agent とは「世界を観察し、自由に使えるツールを使って行動することで、目標を達成しようとするアプリケーション」と定義しており、図 2 に示すようなアーキテクチャで説明している[15]。

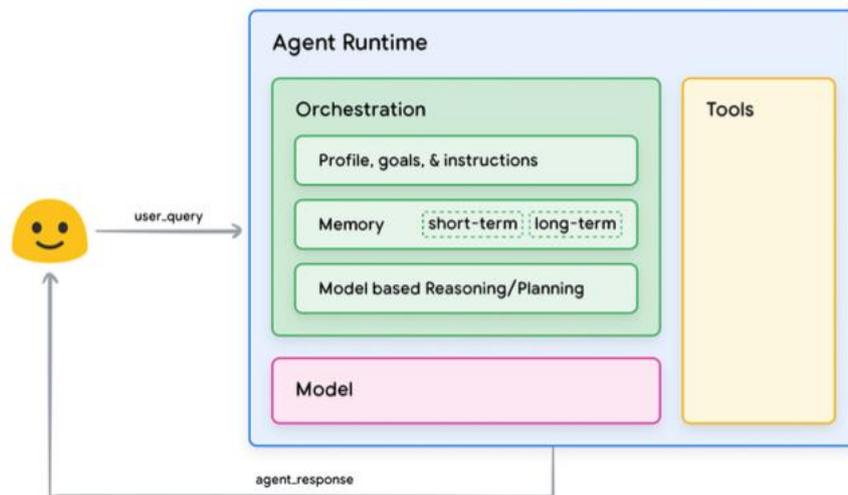


図 2. 一般的な Agent のアーキテクチャ[15]

何をもって Agent と呼ぶかは議論が尽きない問題[16]であるが、近年さまざまな領域で Agent の可能性に注目が集まっている。例えば、Minecraft 内で Agent に世界を探索させる Voyager[17] や、25 人の Agent に仮想的な街での生活をシミュレートさせる Generative Agents[18]、さらにはプログラマーやテスターなどの役割を担う Agent によるソフトウェア開発を実現した ChatDev[19]などが挙げられる。Agent 研究の盛り上がりを受けて、2025 年は「Agent 元年 (Agent era)」になるだろうと期待されている。

科学研究領域においても、研究計画や文献調査、論文執筆を遂行する Agent、AI Scientist が登場した。ここでは代表的な事例を 2 つ取り上げる。

#### 3.2 Sakana AI: The AI Scientist

Sakana AI が 2024 年 8 月に発表した The AI Scientist[2]は、科学的発見プロセスを自動化する包括的フレームワークである。このシステムは科学的プロセスの全段階を自律的に実行できる能力を有する。具体的には、新しい研究アイデアの生成、実験コードの作成・実行、結果の可視化、科学論文の執筆、さらにはシミュレーションされた査読プロセスまでを行う（図 3）。

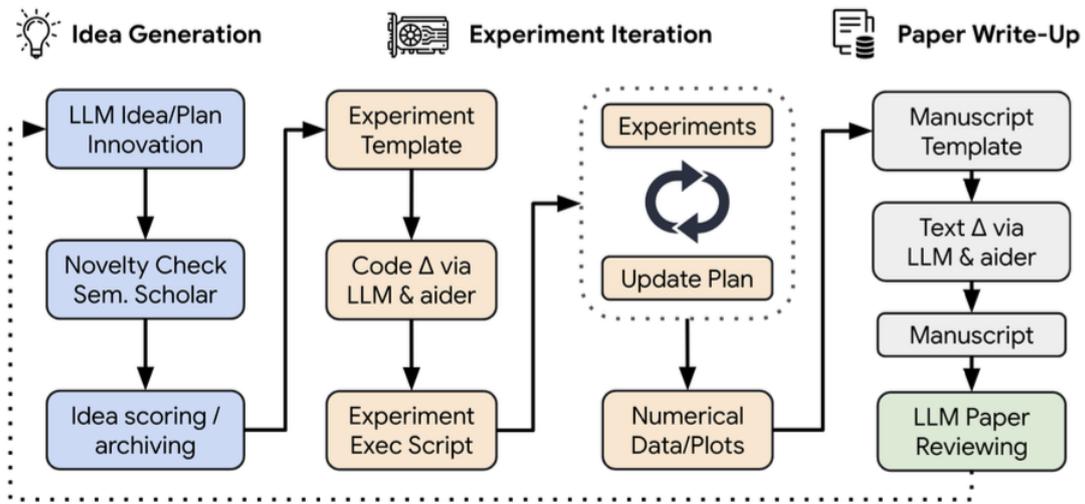


図 3. The AI Scientist 概略図.[2]

The AI Scientist は Chain-of-Thought[20]や Self-reflection[21]などの LLM フレームワークを活用して、研究アイデアを生成し実験計画を立案する。コード実装には Aider[22]というコーディネータアシスタントを使用して実験を実行し、結果に基づいて科学論文を作成する。さらに自動化された査読プロセスを用いて論文の質を評価する。

Sakana AI はこのシステムの有効性を示すため、機械学習における研究分野を対象に検証を行った。アイデアの論文化までにかかった費用はわずか 15 ドルであった。

### 3.3 Google: AI co-scientist

Google の AI co-scientist[3]は、複数の専門エージェントが協働するマルチエージェントシステムである。研究者が自然言語で研究目標を入力すると、エージェントたちは生成、議論、進化というプロセスを非同期かつ反復的に実行し、仮説の質を高めていく。AI co-scientist の概要を図 4 に示す。

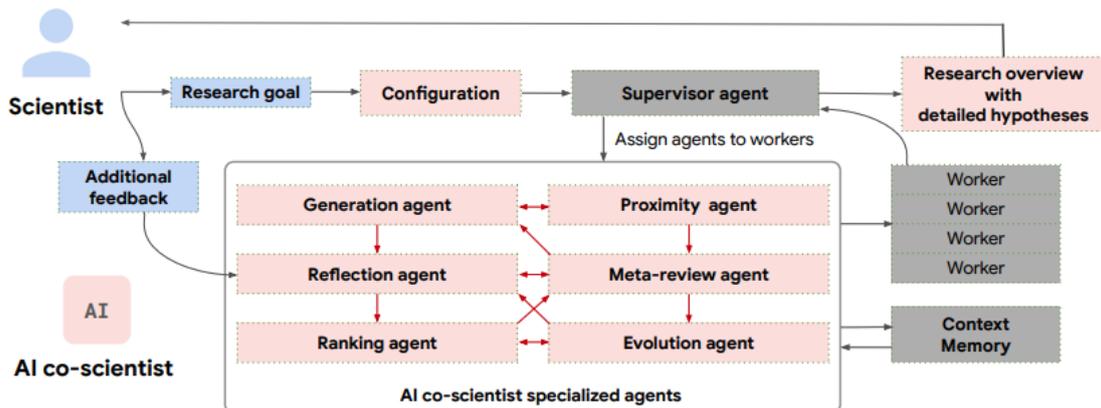


図 4. AI co-scientist 概略図 [3]

システムの中核は、Supervisor agent によって調整される専門エージェント群である。これらの Agent は科学的推論プロセスを模倣している。主要な専門エージェントは以下の 6 種類である。

1. **Generation agent** : 初期焦点領域と仮説を生成
2. **Reflection agent** : 仮説を正確さ、品質、新規性の観点で評価
3. **Ranking agent** : トーナメント形式で仮説を比較評価
4. **Proximity agent** : 仮説の近接グラフを計算し類似アイデアをクラスタリング

5. **Evolution agent** : トップランクの仮説を改良
6. **Meta-review agent** : 洞察を統合し、システム全体の改善を促進

Google は実験的検証として、AI co-scientist を三つの生物医学的応用（既存薬の転用、新治療ターゲットの提案、抗菌剤耐性のメカニズム解明）に適用した。特に注目すべきは、このシステムが急性骨髄性白血病に対する新しい薬物転用候補を提案し、後続の実験でその有効性が確認されたことである。また、肝線維症の治療ターゲット発見においても成功を収めている。現在 Google は世界中の研究機関を対象に Trusted Tester プログラムを通じてシステムへのアクセスを提供している。

#### 4. 展望: AI が科学する時代に何ができるか

著者はディープラーニングが研究領域を越え広く社会で注目されたころ(大学において機械学習サークルが登場し、工学部以外の学生もディープラーニングへの関心が高まっていた時期)に生命科学論文からの実験条件と実験結果の抽出に取り組み始めた。大規模な実験条件と実験結果のデータセットがあれば任意の実験結果を予測できるであろうと甘く考え、2018年3月に fuku 株式会社を創業した。それ以来、加速し続ける AI 研究に翻弄されている。当時、言語モデル研究の重要な転換点となった BERT[23]に大きな注目が集まったが、科学研究などの特定ドメインに対してはファインチューニングが必要であった。そのため、著者は長い時間をかけて科学研究からの情報抽出のためのデータセットを構築したが、2022年11月30日にリリースされた ChatGPT が非常に高い汎用性を示したことで度肝を抜かれ、独自のモデルを捨てモデルプロバイダーが提供する LLM を利用することとした。その後、有志と 2024年2月からレビュー論文自動生成システム GenSurv[24]の開発に取り組み始めた。しかし、2025年2月に OpenAI がリリースした deep research[25]機能に再び度肝を抜かれた。これはウェブ上の大量のテキスト、画像、PDF を探索し、分析し、洞察を統合するリサーチ機能であり、さまざまな領域における調査業務の効率化を実現した。そもそも我々の取り組みは文献調査の効率化を目的としており、その出力フォーマットとしてレビュー論文を採用していたに過ぎない。出力フォーマットにこだわらないのであれば deep research で十分だと考えシステム開発の優先度を下げ、むしろこのようなソリューションを評価するための科学研究に関するデータセットを作る方針に切り替えた。

著者は AI 技術全般の発展がこれからも加速度的に進むと確信している。特に計算機上のシミュレーションで完結する研究領域では、AI によって研究スピードがますます加速していくだろう。一方で、生命科学研究のように物理的な実験を伴う研究領域では、AI システムと現実の実験環境との接続が大きな課題となる。著者は、ラボラトリーオートメーションこそが、現実空間での実験を伴う研究領域が、シミュレーションで完結する研究領域のスピードに置いて行かれないようにするための鍵だと考えている。AI が実験計画を立案し、ロボットが実験を実行し、その結果を AI が解析して次の実験計画に活かす。このような循環的なシステムを構築することで、生命科学も飛躍的な進化を遂げることができるだろう。

科学と AI の共進化の道のりは始まったばかりだ。評価基盤の整備と並行して、AI と実験環境を結びつける技術開発にも注力し、真に役立つ AI 科学研究のエコシステム構築に貢献していきたい。

#### 謝辞

本稿の執筆にあたり、多くの方々からのご支援とご協力に深く感謝申し上げます。今の自分があるのは、主に二つのコミュニティのおかげです。2019年7月に国内版バイオハッカソンに飛び入り参加し、論文から実験条件と実験結果を抽出したいと宣言した際、向こうみずな若者を暖かく迎え入れてくださった参加者の皆様、とりわけその後も浮き沈みの激しい著者を受け入れてくださっているライフサイエンス統合データベースセンターの皆様には心より感謝申し上げます。

同じく 2019年7月に参加した Laboratory Automation 月例勉強会では、ロボット実験に取り組んでいる方々の中で自然言語処理に取り組む著者は異質でしたが、その後運営メンバーとして登

用していただきました。コアメンバーの皆様には常日頃からご支援いただき、深く感謝しております。

fuku 株式会社に携わってくださるすべての方々にも、この場を借りて御礼申し上げます。

また、生まれて半年の娘がいる生活の中、夜通しの執筆を支援してくれた妻には特別な感謝の気持ちを捧げます。

最後に、貴重な機会をくださり、執筆締め切りを大幅に超過したにも関わらず、校正および記事内容へのアドバイスをしてくださった日本薬学会構造活性相関部会の皆様には深謝いたします。大変ご迷惑をおかけしました。

## 参考文献

- [1] McCarthy, J., Minsky, M., Rochester, N., & Shannon, C. E. (1955). *A proposal for the Dartmouth summer research project on artificial intelligence*.
- [2] Lu, C., Lu, C., Lange, R. T., Foerster, J., Clune, J., & Ha, D. (2024). *The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery*. arXiv preprint arXiv:2408.06292.
- [3] Gottweis, J., Weng, W.-H., Daryin, A., Tu, T., Palepu, A., Sirkovic, P., Myaskovsky, A., Weissenberger, F., Rong, K., Tanno, R., Saab, K., Popovici, D., Blum, J., Zhang, F., Chou, K., Hassidim, A., Gokturk, B., Vahdat, A., Kohli, P., Matias, Y., Carroll, A., Kulkarni, K., Tomasev, N., Guan, Y., Dhillon, V., Vaishnav, E. D., Lee, B., Costa, T. R. D., Penadés, J. R., Peltz, G., Xu, Y., Pawlosky, A., Karthikesalingam, A., & Natarajan, V. (2025). Towards an AI co-scientist. *arXiv preprint arXiv:2502.18864*.
- [4] Buchanan, B. G., & Feigenbaum, E. A. (1978). Dendral and Meta-Dendral: Their applications dimension. In D. Michie (Ed.), *Expert Systems in the Micro-Electronic Age* (pp. 247-263). Edinburgh University Press.
- [5] Douglas B Lenat. Automated theory formation in mathematics. In *IJCAI*, volume 77, pages 833–842, 1977.
- [6] Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., ... & Zitnik, M. (2023). Scientific discovery in the age of artificial intelligence. *Nature*, 620, 47–60.
- [7] AI for Science Community. (n.d.). *AI for Science*. Retrieved March 22, 2025, from <https://ai4sciencecommunity.github.io/>
- [8] AlphaProof and AlphaGeometry teams. (2024, July 25). AI achieves silver-medal standard solving International Mathematical Olympiad problems. DeepMind Blog. Retrieved from <https://deepmind.google/discover/blog/ai-solves-imo-problems-at-silver-medal-level/>
- [9] Alfarano, A., Charton, F., & Hayat, A. (2024). Global Lyapunov functions: a long-standing open problem in mathematics, with symbolic transformers. arXiv preprint arXiv:2410.08304.
- [10] Wang, H., Skreta, M., Ser, C.-T., Gao, W., Kong, L., Streith-Kalthoff, F., Duan, C., Zhuang, Y., Yu, Y., Zhu, Y., Du, Y., Aspuru-Guzik, A., Neklyudov, K., & Zhang, C. (2024). Efficient evolutionary search over chemical space with large language models. *arXiv preprint arXiv:2406.16976*.
- [11] Hayes, T., Rao, R., Akin, H., Sofroniew, N. J., Oktay, D., Lin, Z., Verkuil, R., Tran, V. Q., Deaton, J., Wiggert, M., Badkundri, R., Shafkat, I., Gong, J., Derry, A., Molina, R. S., Thomas, N., Khan, Y., Mishra, C., Kim, C., Bartie, L. J., Nemeth, M., Hsu, P. D., Sercu, T., Candido, S., & Rives, A. (2024). Simulating 500 million years of evolution with a language model. *bioRxiv*.
- [12] Ji, Y., Zhou, Z., Liu, H., & Davuluri, R. V. (2021). DNABERT: pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome. *Bioinformatics*, 37(15), 2112–2120.
- [13] Bodnar, C., Bruinsma, W. P., Lucic, A., Stanley, M., Vaughan, A., Brandstetter, J., Garvan, P., Riechert, M., Weyn, J. A., Dong, H., Gupta, J. K., Thambiratnam, K., Archibald, A. T., Wu, C.-C., Heider, E., Welling, M., Turner, R. E., & Perdikaris, P. (2024). A foundation model for the Earth system. arXiv.
- [14] RIKEN. (n.d.). 最先端研究プラットフォーム連携(TRIP)事業本部科学研究基盤モデル開発プログラム Retrieved March 22, 2025, from <https://www.riken.jp/research/labs/trip/agis/>
- [15] Wiesinger, J., Marlow, P., & Vuskovic, V. (2024, September). *Agents*. Google.
- [16] Chase, H. (2024, June 28). *What is an AI agent?* LangChain. Retrieved March 22, 2025, from <https://blog.langchain.dev/what-is-an-agent/>

- [17] Wang, G., Xie, Y., Jiang, Y., Mandlekar, A., Xiao, C., Zhu, Y., Fan, L., & Anandkumar, A. (2023). *Voyager: An open-ended embodied agent with large language models*. arXiv.
- [18] Park, J. S., O'Brien, J. C., Cai, C. J., Morris, M. R., Liang, P., & Bernstein, M. S. (2023). *Generative Agents: Interactive Simulacra of Human Behavior*. arXiv
- [19] Qian, C., Liu, W., Liu, H., Chen, N., Dang, Y., Li, J., Yang, C., Chen, W., Su, Y., Cong, X., Xu, J., Li, D., Liu, Z., & Sun, M. (2024). *ChatDev: Communicative Agents for Software Development*. arXiv.
- [20] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35:24824–24837, 2022.
- [21] Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: Language agents with verbal reinforcement learning. *Advances in Neural Information Processing Systems*, 36, 2024.
- [22] Aider-AI. *Aider: AI pair programming in your terminal*. GitHub. <https://github.com/Aider-AI/aider>
- [23] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186.
- [24] GenSurv. (2025). *GenSurv: Generative Survey*. GitHub. <https://github.com/GenSurv/gensurv>
- [25] OpenAI. (2025, February 2). *Introducing deep research*. Retrieved March 22, 2025, from <https://openai.com/index/introducing-deep-research/>

## ///// 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。

### 2024 年度 SAR Presentation Award について

2024 年度は、第 52 回構造活性相関シンポジウムにおける 40 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（口頭発表・ポスター発表）を選考対象とすることとした。

### 2024 年度 SAR Presentation Award 受賞者（演題番号順）

口頭発表： 早川 大地（昭和大学大学院 薬学研究科）

ポスター発表： 寺山 慧（横浜市立大学大学院 生命医科学研究科）

ポスター発表： 村上 優貴（横浜市立大学大学院 生命医科学研究科）

### 受賞者の選考について

2024 年 12 月 12-13 日に各審査員からオンラインにて提出いただいた審査票を集計し、口頭発表 1 名、ポスター発表 2 名を受賞者として選出した。口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。審査員が共同演者となっている演題は選出不可としている。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。

### 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: スライドやポスターは、専門領域の異なる参加者にもわかりやすく作成されているか。
- 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 52 回構造活性相関シンポジウムに参加した 2024 年度常任幹事および幹事

## &lt;受賞者コメント&gt;

**KO04**

氏名 早川 大地 (はやかわ だいち)  
 所属 昭和大学大学院 薬学研究科  
 演題 量子化学計算とデータベース解析による CH/ $\pi$  および CH/N 相互作用の 3 次元解析

この度は、SAR Presentation Award を賜りましたこと、大変光栄に存じます。ご評価いただきました先生方、並びに日本薬学会構造活性相関部会の先生方に心より感謝申し上げます。

本研究では、含窒素複素環式化合物の周囲に形成し得る CH/ $\pi$  相互作用と CH/N 相互作用を、量子化学計算に基づいた分子相互作用場計算(MIF(QM))と Cambridge structural database (CSD)の解析により、三次元的に明らかにしました。CH/ $\pi$  や CH/N 相互作用といった弱い水素結合は、タンパク質とリガンド分子の間でも形成されることが知られています。これらの相互作用は創薬においても重要と考えられますが、基本的な特徴や特性について十分明らかになっていない点も多く残されています。本研究では、まずは CH/ $\pi$  や CH/N 相互作用の三次元的な描像を明らかにしたいという思いで検討を進めて参りました。

本研究においては、理論計算(MIF(QM))と実験データ(CSD)の解析を併用することで、理論的にも実験的にも確からしい結論に近づくことに最も注力いたしました。今後は、得られた CH/ $\pi$  や CH/N 相互作用の近似関数を用いた、創薬や SBDD のための分子モデリング法の検討を進めていきたいと考えています。

本研究は、JSPS 科研費 JP24K15167 の助成を受けたものです。同助成に感謝申し上げます。最後に、研究に関する多くのご助言をいただき、*in silico* 研究を実施するにあたり十分な研究環境を整えていただきました、昭和大学の合田浩明教授に心より感謝申し上げます。

**KP03**

氏名 寺山 慧 (てらやま けい)  
 所属 横浜市立大学大学院 生命医科学研究科  
 演題 分子生成 AI による V-ATPase 阻害剤の最適化

この度は、第 52 回薬学会構造活性相関シンポジウム SAR Presentation Award (ポスター賞) という名誉ある賞をいただき、大変光栄に存じます。ご評価頂いた審査員の先生方、日本薬学会構造活性相関部会の関係者の皆様、シンポジウム実行委員の皆様に、心より感謝申し上げます。

本研究では、V-ATPase 阻害剤のリード最適化を目的として、分子生成 AI ChemTSv2 を活用し、ヒット化合物 V-161 の構造最適化を試みました。具体的には、結合親和性を向上させつつ、膜透過性をコントロールすることを目標に、GNINA によるドッキング計算と logP の計算値を考慮した多目的最適化を実施しました。その結果、結合親和性のスコアが向上し、膜透過性が低いと予想される新規化合物の生成に成功しました。本研究はまだ発展途上にあり、現在、さまざまな分子構造の実験的検証を進めるとともに、膜透過性や阻害活性をより高精度に予測できるモデルの構築に取り組んでいます。

本研究は、共同研究者の皆様のご協力なしには成し得ませんでした。本プロジェクトは千葉大学の村田武士教授・鈴木花野助教による構造生物学的な研究成果と発見されたヒット化合物から始まったものです。阻害剤の設計・合成については、北海道大学の勝山彬助教、市川聡教授にご

尽力いただきました。また、インシリコでの分子設計全般にわたり、理化学研究所の大田雅照上級研究員、池口満徳教授、浴本享助教から貴重かつ具体的なアドバイスを頂戴しました。さらに、この研究成果は本研究室修士課程1年の戸板太陽さんが中心となって計算・解析を行い、得られたものです。この場を借りて深く感謝申し上げます。

#### KP15

氏名 村上 優貴 (むらかみ ゆうき)  
所属 横浜市立大学大学院 生命医科学研究科  
演題 細胞膜透過性を考慮した PROTAC リンカー生成 AI の開発

この度は、SAR Presentation Award (ポスター賞) という名誉ある賞をいただき大変光栄に思います。ご評価くださった先生方、日本薬学会構造活性相関部会の関係者の方々、構造活性相関シンポジウム実行委員の方々に厚く御礼申し上げます。

今回は、細胞膜透過性を考慮した PROTAC (proteolysis targeting chimera) リンカー生成 AI に関する内容を発表しました。タンパク質分解誘導剤である PROTAC は、次世代型医薬品として注目されています。しかし、2つのリガンドとそれらを繋ぐリンカーで構成される PROTAC の分子量は、一般的な低分子医薬品と比べて大きくなってしまい、細胞膜透過性が低い傾向にあります。細胞膜透過性を含む PROTAC の物理化学的特性は、リンカーの違いにより変化することが知られており、リガンドの組み合わせごとにリンカーを最適化する必要があります。そこで本研究では、当研究室で開発している分子構造生成 AI (ChemTSv2) を拡張し細胞膜透過性向上を目指した PROTAC のリンカー設計手法を開発しました。本手法を用いて PROTAC のリンカー設計を試みたところ、細胞膜透過性の予測値が高い PROTAC のリンカー設計に成功しました。本研究は、より優れた薬物動態を有する PROTAC 開発への応用が期待されます。

今回のシンポジウムでは、多くの方々から様々な角度でご意見をいただき、新たな視点や気づきを得ることができました。最後に、本研究の遂行や発表準備にあたりご指導いただきました、横浜市立大学 寺山准教授、石田祥一特任助教、国立医薬品食品衛生研究所 出水庸介先生、横浜市立大学生命情報科学研究室の皆様がこの場を借りて御礼申し上げます。今回の受賞を励みに、さらなる研究の発展に向けて精進してまいります。

## ///// Activities /////

## 第 52 回構造活性相関シンポジウム開催報告

日時：2024（令和 6）年 12 月 12 日（木）・13 日（金）

主催：日本薬学会構造活性相関部会

協賛：情報計算法学生物学会（CBI 学会）、日本農薬学会、日本薬学会医薬化学部会、日本薬学会レギュラトリーサイエンス部会、理論化学会、日本バイオインフォマティクス学会（JSBi）

会場：川崎市産業振興会館（神奈川県川崎市幸区堀川町 66 番地 20）1 階ホール・4 階展示場

第 52 回構造活性相関シンポジウムは、招待者を含めて 144 名のご参加登録を頂きました。2024 年度は 5 年ぶりに企業展示 7 社様のブースを設けることができ、ソフトウェアやシステムなどの最新動向についても参加者の皆様に情報を提供することができました。演題数は、特別講演 1 件、基調講演 1 件、招待講演 3 件、一般講演として口頭 7 件、ポスター 34 件となり、盛会のうちに終えることができました。これもひとえにご参加頂いた皆様と、実行委員の先生方やアルバイト並びに志鷹真由子部会長をはじめとする日本薬学会構造活性相関部会幹事の先生方のご助力、ご支援の賜物と存じます。主催の日本薬学会構造活性相関部会と、協賛頂いた諸学会に感謝するとともに、ご講演、ポスター発表いただいた皆様にお礼申し上げます。また、開催資金のご援助をいただいた日本薬学会並びに関係企業等の皆様に深くお礼申し上げます。

特別講演では、神戸大学の天能精一郎先生に、「特異な相互作用や機能に関わる複雑な電子状態と理論化学的アプローチ」についてご講演を頂きました。遷移金属の強い電子相関の解をもとめる理論展開にとどまらず、人工光合成の分野で高精度 *ab initio* 法で求めた理論予測を基に新たな実験を提案する研究成果でした。



ミニシンポジウム「レギュラトリーサイエンスと QSAR」を開催し、基調講演として国立医薬品食品衛生研究所の本間正充先生から「レギュラトリーサイエンスにおける AMES /QSAR の利用」、招待講演として、中外製薬株式会社の小山直己先生から「ICH M7 ガイドラインに準ずる QSAR を活用した医薬品不純物の変異原性予測評価の基礎と実際」、国立環境研究所の伊丹悠人先生から「生態リスク評価における QSAR の活用」、製品評価技術基盤機構の青柳智子先生から「OECD における QSAR の行政利用の推進と日本の貢献」のご講演を賜りました。日本薬学

会レギュラトリーサイエンス部会長を務めておられる本間先生のご講演は、レギュラトリーサイエンスの定義に始まり、化学物質管理や医薬品分野での構造活性相関の規制(レギュレーション)利用状況や関連する研究成果について幅広い教育的な内容でした。招待講演の先生方からも、ご専門分野に関する(定量的)構造活性相関研究とその応用状況等についてご紹介頂きました。

SAR Presentation Award は、以下の3名に決定いたしました。受賞おめでとうございます。

口頭発表(1名)

KO04: 早川大地(昭和大院薬)

「量子化学計算とデータベース解析による CH/π および CH/N 相互作用の3次元解析」

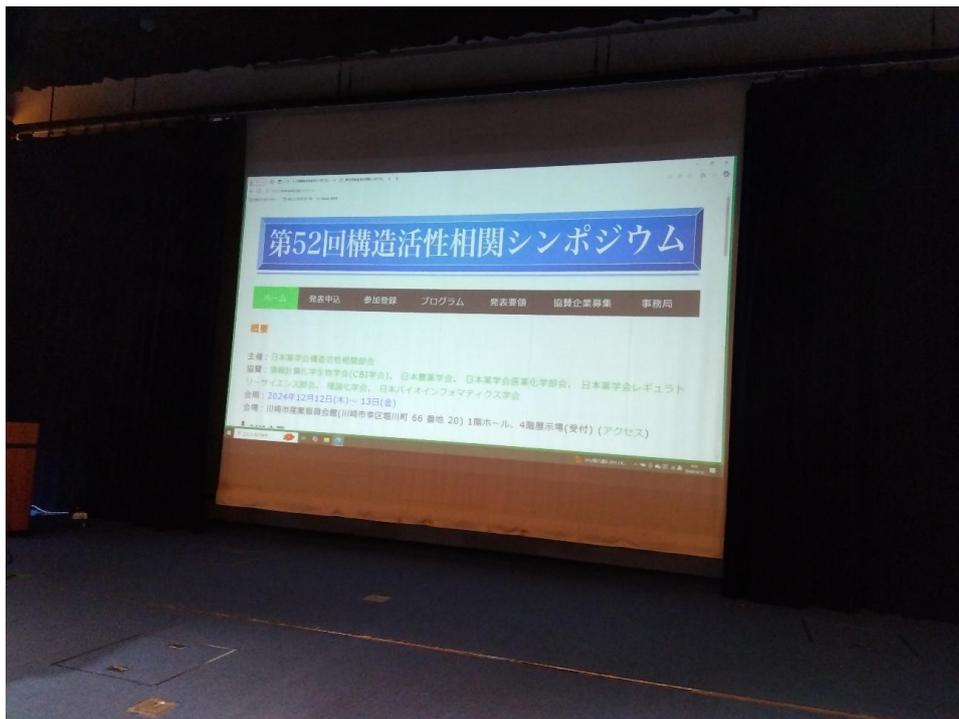
ポスター発表(2名)

KP03: 寺山慧(横浜市大院生命医)

「分子生成 AI による V-ATPase 阻害剤の最適化」

KP15: 村上優貴(横浜市大院生命医)

「細胞膜透過性を考慮した PROTAC リンカー生成 AI の開発」



第53回構造活性相関シンポジウムは、近畿大学の川下理日人実行委員長のもと大阪の地で2025年9月4日(木)～5日(金)に開催予定です。引続きよろしくお願ひ申し上げます。

第52回構造活性相関シンポジウム  
実行委員長 古濱 彩子

//// Activities ////

---

<会告>

構造活性フォーラム 2025

「AI・計算化学の産業応用と新展開」

主催： 日本薬学会構造活性相関部会

協賛： 情報計算化学生物学会（CBI学会）

会期： 2025年6月20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： <https://sites.google.com/view/sarforum2025>

**開催趣旨：**製薬関連企業では、計算化学や機械学習を含む先端的なインシリコ技術の開発や応用が進められており、多くのノウハウが蓄積されている。これら産業界における計算化学・機械学習の研究動向を知ることは、薬学研究者が新たな着想を得るうえで重要である。そこで、本フォーラムでは産業界を中心とした最先端の応用研究をご紹介いただき、AI・計算化学の新たな切り口を議論したい。

**プログラム：**

講演 1. 大川 和史（旭化成ファーマ）「AIと計算化学がもたらす新たな展望」

講演 2. 半田 耕一（Axcelead Tokyo West Partners）「創薬における実践的な機械学習モデルの追求」

講演 3. 準備中 キーワード：データ駆動型創薬、バーチャルスクリーニング、オープンサイエンス

講演 4. 森 健一（アステラス製薬）「様々な創薬モダリティ研究を効率化するためのAI活用事例と展望」

講演 5. 上原 彰太（塩野義製薬）「COVID-19 経口治療薬 S-217622(エンシトレルビル)創製における Virtual Screening の活用」

講演 6. 小野 聡（ゼウレカ）「環状ペプチドのコンフォメーションと膜透過性」

講演 7. 松尾 篤（中外製薬）中分子創薬プラットフォームの技術開発、LUNA18 の創薬研究における計算化学・SBDD

**参加登録および申込締切日：**6月6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。

**参加費：**一般 4000 円、日本薬学会・協賛学会会員 2000 円、学生 無料

**問合先：**

構造活性フォーラム 2025 実行委員会

摂南大学薬学部 河合健太郎（実行委員長）

〒573-0101 大阪府枚方市長尾峠町 45-1

Tel: 072-800-1235（直通）

E-mail: kentaro.kawai<@>pharm.setsunan.ac.jp（<@>を@に置換してください）

## ///// Activities /////

## &lt;会告&gt;

## 第 53 回構造活性相関シンポジウム

主催： 日本薬学会構造活性相関部会

会期： 2025 年 9 月 4 日（木）～5 日（金）

会場： 近畿大学東大阪キャンパス（大阪府東大阪市小若江 3 丁目 4 番 1 号）

## 討論主題

1. 生理活性物質の活性評価・構造展開・医農薬への応用
2. 基本パラメータ・基本手法・情報数理的アプローチ
3. 吸収・分布・代謝・毒性・環境毒性
4. in silico 技術（薬物-受容体相互作用計算、仮想スクリーニングなど）
5. バイオインフォマティクス
6. 分子情報処理（データベースを含む）・データ予測
7. その他

発表形式：口頭発表・ポスター発表

発表申込：6 月 1 日（日）～ 6 月 30 日（月）

要旨登録：7 月 1 日（火）～ 7 月 31 日（木）

## 招待講演（タイトルはいずれも仮題）：

木下 誉富（大阪公立大学）

「構造生物学を基盤としたキナーゼ創薬研究」

阿部 一啓（北海道大学）

「胃プロトンポンプの構造情報と AI によって駆動された新規阻害剤の de novo デザイン」

米倉 功治（理化学研究所）

「電子線三次元結晶構造解析とクライオ EM による化学特性の計測」

大上 雅史（東京科学大学）

「AlphaFold がもたらす計算創薬の変革」

## 参加登録費・懇親会参加費：

参加登録費	薬学会会員	一般*	8,000 円	学生	1,000 円	金額は不課税 (適用対象外)
	非会員	一般	10,000 円	学生	2,000 円	金額は税込額
懇親会参加費		一般	8,000 円	学生	8,000 円	金額は税込額

\*薬学会会員の一般はシニア会員、終身会員、永年会員を含む

参加登録申込：7月1日（火）～8月31日（日）

懇親会（定員50名\*）：THE LOUNGE（近畿大学東大阪キャンパス内）9月4日（木）開催

\*申込多数の場合は、申込順とさせていただきます。

その他、最新情報はホームページにてご確認ください。

HP：<https://www.phar.kindai.ac.jp/qsar2025/>

問い合わせ先：

第53回構造活性相関シンポジウム実行委員会

近畿大学理工学部 川下理日人（実行委員長）

E-mail: [sar2025@qsarj.org](mailto:sar2025@qsarj.org)

---

**部会役員人事**

2025 年度 常任世話人 2025/4/1 現在

部会長	竹田-志鷹 真由子 (北里大学 薬学部)
副部会長	田上 宇乃 (味の素 (株))
副部会長	前田 美紀 (農業・食品産業技術総合研究機構)
会計幹事	川下 理日人 (近畿大学 理工学部)
庶務幹事	河合 健太郎 (摂南大学 薬学部)
広報幹事	加藤 博明 (広島商船高等専門学校)
SAR News 編集長	合田 浩明 (昭和大学 薬学部)
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)

---

**構造活性関連部会の沿革と趣旨**

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たすこととなった。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**

日本薬学会構造活性相関部会誌 SAR News 第 48 号をお届けいたします。今号では、「大規模言語モデル (Large Language Models, LLM) の創薬への応用」をテーマに、第一線でご活躍の先生方にご寄稿をお願いいたしました。Perspective/Retrospective では、東京科学大学情報理工学院 関嶋政和先生、および叢雲くすり (@souyakuchan) 先生に創薬プロセスに有用な言語モデルの概説、および言語モデルに基づいた分子設計研究の将来展望をご紹介いただきました。Cutting Edge では、東京科学大学情報理工学院 大上雅史先生にタンパク質言語モデル (Protein Language Models, PLM) の発展と創薬分野での応用、株式会社サイキンソー山崎広之先生に LLM の基盤技術、学習法、応用領域、および問題点、fuku 株式会社 山田涼太先生に科学研究への人工知能 (Artificial Intelligence, AI) の活用の歴史、および擬人化され主体的に研究を行う AI Scientist について、ご紹介いただきました。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。

また、昨年 12 月に開催された第 52 回構造活性相関シンポジウムで SAR Presentation Award を受賞された先生方の受賞者コメント、第 52 回構造活性相関シンポジウムの開催報告、2025 年度に開催予定の構造活性フォーラム 2025 および第 53 回構造活性相関シンポジウムの会告も掲載いたしましたので、お目通しいただければ幸いです。(編集委員会)

SAR News No.48 2025 年 4 月 1 日

発行: 日本薬学会 構造活性相関部会長 竹田-志鷹 真由子

SAR News 編集委員会

(委員長) 合田 浩明、浴本 亨、遠藤 智史、仲西 功、原田 俊幸、幸 瞳

\*本誌の全ての記事、図表等の無断複写・転載を禁じます。