人工知能が拓く
生命科学の新時代

金景珍（キム・ギョンジン）

この本の作られ方

本書は人工知能とともに編纂された研究書です。人間が資料を選定して構成を組み立て、AIモデルが文章を起草し事実関係をクロスチェックしました。

その基盤となっているのは、Google NotebookLM上でまとめられた1冊のノートブックです。ここには2026年8月までに発表された332編のarXivプレプリントと71件の国際学術セミナー発表資料が集約されています。英語、中国語、日本語の資料が含まれています。ノートブックによって生成された中核的な要約が全体の骨格を形成し、各章の内容は2026年2月以降に公開された原著論文や関連資料を再調査することで肉付けされました。

このノートブックは一般に公開されています。以下のリンクから、本書の根拠となった資料を直接ご確認いただけます。

https://notebooklm.google.com/notebook/df55fd37-1b59-426c-93f5-670b807eac5f

このリソースは、研究者や学生の皆様が研究や学習に役立てられるよう公開されています。本書の記述に不確かな点がある場合は、ノートブック内の原典に戻ってご確認ください。

各セクションの末尾には、学術引用形式で参考文献が記載されています。引用された論文は、著者名とタイトルをarXiv APIおよびCrossrefと照合し、1件ずつ検証しました。実在を確認できなかった資料は本文から除外してあります。

人工知能によって合成された文章には本質的な限界があります。誤りが含まれる可能性もあります。検証の責任は読者の皆様にあります。

金景珍（キム・ギョンジン）

目次

	本書はどのように制作されたか 情報源となったNotebookLMノート、公開URL、および相互検証手法
1	**AI主導の生物学R&Dにおけるパラダイムシフト** 深夜の研究室で行き来するピペットとキーボード、そして20億ドル・10年の壁
	1.1 ドライラボとウェットラボの相補的融合、そして学際的なキャリアモデル
	1.2 生物学的知識の伝播と生成：学界と産業界の変化
2	**生体データの幾何学的表現学習とマルチモーダルアライメント** 共有座標多様体における配列・構造・テキストの整列
	2.1 タンパク質・遺伝子・テキストのモダリティを横断するマルチモーダルアライメントと対照学習
	2.2 高次元マルチオミクスデータを低次元潜在表現へと圧縮する埋め込みの原理
3	**タンパク質立体構造予測の革新** 10^47という天文学的な立体配座の可能性から単一の折りたたみ状態を解く機械
	3.1 アミノ酸配列からの高精度立体構造予測の原理と数学的基礎
	3.2 AlphaFold、AlphaFold2、AlphaFold3のアーキテクチャの進化と高分子複合体相互作用モデリング
	3.3 実験的構造プロテオミクスデータと機械学習（ML）モデルの相互誘導
4	**生成型生物学モデルを用いたタンパク質と抗体のde novo設計** ガウス拡散ノイズから自然界に存在しない新規タンパク質を造形する
	4.1 拡散モデルとフローマッチングによる物理的タンパク質構造の生成
	4.2 開発可能性を最大化する抗体CDR設計と選好性最適化
	4.3 Proteo-R1：タンパク質構造の理解にマルチタスク推論と言語的思考の連鎖を注入する
5	**ゲノムスケールのDNA言語モデルと生物学的コードの解読** 4文字のアルファベットで記された30億塩基の文章を読み解く言語モデル
	5.1 全ゲノムDNA配列におけるヌクレオチド言語モデルの事前学習
	5.2 1塩基解像度でメガベース規模のコンテキストを読み解くEvoとEvo2のアーキテクチャ
6	**微生物ゲノム解析と機能推論** 未培養微生物のゲノムから生息地と二次代謝産物を推論する
	6.1 GGBoundによる微生物の生理的限界予測とゲノムからの代謝産物探索
	6.2 MicroFuse：オペロン構造と微生物制御経路推論のためのエキスパート融合フレームワーク

7	**AIによるバーチャルスクリーニング（VS）と化合物設計** 合成前にイン・シリコで1000億個の仮想低分子をスクリーニング・選別する
	7.1 標的ポケットと化合物の3次元分子ドッキング、および薬剤-標的親和性（DTA）バーチャルスクリーニング（VS）
	7.2 アクティブラーニングと極度なデータ不足を克服するSPADEフレームワーク
8	**トランスクリプトーム主導の創薬候補化合物探索** がんの代謝およびトランスクリプトーム回路図に沿って創薬設計を逆転させる
	8.1 患者の代謝経路と制御構造を考慮した標的抗がん剤のバーチャルスクリーニング（VS）
	8.2 大規模遺伝子発現（トランスクリプトーム）データを用いた低分子化合物の逆設計
	8.3 副作用を軽減する薬剤再設計を導く先例誘導型コ・サイエンティスト・アーキテクチャ
9	**1細胞データの前処理と埋め込み技術** 単一細胞トランスクリプトームプロファイルから技術的ドロップアウトとノイズを除去する技術
	9.1 1細胞トランスクリプトミクス（scRNA-seq）とマルチオミクス多様体アラインメント
	9.2 1細胞トランスクリプトミクスにおける固有ノイズ低減に向けた補完と基盤モデルのベンチマーク
10	**1細胞状態遷移と動的摂動応答解析** 離散的なスナップショットから連続的な細胞軌道とエピジェネティック・ランドスケープを再構成する
	10.1 Chreodeによる時間的細胞発生軌跡とトランスクリプトミクス摂動モデリング
	10.2 Sigmoid Attentionを用いた数万個の遺伝子にわたる動的競合相互作用の計算
11	**AIエージェント駆動の合成生物学と遺伝子回路設計** 階層的フィードバックを通じて遺伝子回路を設計・検証する自律型エージェント
	11.1 ゲノムおよび制御配列の自動設計に向けた大規模バイオロジカルエージェントの展開
	11.2 自律的遺伝子回路設計のための階層的検証に基づく強化学習（GenCircuit-RL）
12	**未来の自律型・自律駆動型ラボ（SDL）** 無人の夜間に生体アッセイを実行する自律型ロボットラボ
	12.1 リキッドハンドリングロボットのためのスクリプト自動導出と最適化された自律的PCR・実験実行（AutoPCR）
	12.2 実験的検証と連携したマルチエージェント自律型クローズドループ創薬（Latent-Y）
13	**バイオセキュリティ、デュアルユースガバナンス、および生成バイオAIの安全ガイドライン** 生成設計技術がデュアルユースリスクをもたらす際のガバナンスとスクリーニングフィルターの確立
	13.1 AI設計病原体のリスクとウイルス合成シナリオにおけるスクリーニングの極めて重要な役割
	13.2 生物学基盤モデルのための初期安全性専門家統合フレームワーク

14	**バイオAIのグローバルガバナンス、レギュラトリーサイエンス、人間の説明責任** 規制の枠組み、再現性ベンチマーク、そして人間の知恵が持つ不変の必要性
	14.1 EU AI法におけるバイオAIの診断安全性とシステミックリスク管理
	14.2 臨床的信頼性、再現性の危機の克服、そして人間の知恵と責任の不可欠な役割

第1章：AI主導の生物学R&Dにおけるパラダイムシフト

1.1 ドライラボとウェットラボの相補的融合、そして学際的なキャリアモデル

午前2時、実験室は静まり返っています。蛍光灯は半分だけ点いており、遠心分離機が低く唸りを上げています。大学院生が96ウェルプレートに試薬を1つずつ移しています。手首が痛み始めると、振り返って隣の机のノートパソコンを確認します。画面上では、学習損失曲線が一歩ずつ下降しています。ピペットとキーボードを行き来するこの光景は、今や生命科学の研究室における日常の夜の姿となっています。

生物学は長い間、実際に手を動かす学問でした。グレゴール・メンデルはエンドウ豆を1株ずつ掛け合わせることで遺伝の法則を明らかにしました。1953年には、ロザリンド・フランクリンが撮影した1枚のX線回折写真がDNA二重らせんの構造を解き明かしました。2003年には、ヒトゲノム計画が30億すべての塩基対（bp）を解読しました。これらの成果の舞台となったのは、常に試験管や培養器を備えた物理的な実験室、すなわちウェットラボでした。ウェットラボは、生きた細胞が実際にどのように反応するかを研究者が直接目撃する場所です。その現場に代わる場所は、今も存在しません。

真の障壁となるのは費用です。1つの創薬候補化合物を実験台から薬局の棚へと送り届けるまでには、平均して10年から15年かかります。研究開発費の総額は2兆ウォン（15億ドル以上）を超えます。第1相臨床試験に進んだ創薬候補化合物のうち、10個中9個は薬事承認を得ることなく姿を消します。何万もの化合物を1つずつ合成して細胞に投与するのは、堅実な手法です。しかし、時間がかかりすぎます。失敗が確認されるのは大半が実験台の上であり、その時点ですでに時間と資金は消え去ってしまっています。

ドライラボは、物理的な試薬の代わりにデータを扱います。冷蔵庫を満たすのは生体試料ではなくサーバー室の熱気であり、その原材料となるのはペタバイト規模のオミクス測定データと数億個に及ぶタンパク質の原子座標です。化学的に合成可能な低分子化合物の数は、およそ10の60乗と推定されています。地球上のすべての砂粒を数え上げたとしても、この数字には到底及びません。ドライラボは物理法則と幾何学的ディープラーニングを活用して、この広大な倉庫の中を探索します。この学習手法は、分子を単なるテキスト文字列としてではなく、3次元の幾何学的オブジェクトとして扱います。かつては何ヶ月もかかっていた候補化合物のスクリーニングが、今では数時間で完了します。

ドライラボが研究の周辺から中心へと躍り出た決定的な瞬間を1つ挙げるならば、それは2021年です。その年、DeepMindはアミノ酸配列のみからタンパク質の3次元折りたたみ構造を予測する人工知能システム、AlphaFoldを公開しました[7]。それまで、1つのタンパク質構造を解明するには結晶を成長させ、X線を照射し、何年もの歳月を費やす必要がありました。AlphaFoldは

その作業をわずか数分にまで短縮し、予測された構造の公開データベースは2億件を超えました。世界中の何百万人もの研究者がこのデータベースに依存しています。構造を知ることと生物学的機能を理解することは別であるという異論は、依然としてもっともです。それでもなお、研究のスタートラインは完全に塗り替えられました。かつては構造を解明すること自体が学術論文として成立していましたが、今日では構造は実験を行うための単なる前提条件にすぎなくなっています。

本当の変革は、一方のラボが他方を打ち負かすことにあるのではありません。2つのラボが互いに返信をやり取りし始めることにあるのです。ドライラボが数億個の仮想分子の中から十数個の有望な候補を選び出して送ると、ウェットラボがそれらを実際に合成し、結合定数や細胞毒性を測定します。それらの実験結果は再び数値となってモデルへと戻ります。モデルは自らの予測がどこで誤っていたかを特定し、パラメータを更新します。この往復運動は「クローズドループ（Closed Loop）」と呼ばれます。この手紙のやり取りの周期が短くなればなるほど、研究の進む速度は上がります。

1つの完全なループがどのように回るのかを見れば、このプロセスが具体的に分かります。月曜日の朝、モデルは標的タンパク質の結合ポケットに適合するように設計された12個の候補分子を絞り込みます。火曜日に化学合成が始まり、木曜日には結合親和性を測定するアッセイが完了します。12個のうち9個は、予測よりもはるかに弱くしか結合しませんでした。過去であれば、このような結果は失望を呼び、そこで終わっていたはずです。しかし現在では、その9個の失敗が貴重な学習データとなります。モデルはどの原子配置が過信を招いたのかを検証し、翌週の候補リストを再構成します。失敗がデータになる瞬間、ラボの時間は直線的に流れるのをやめ、螺旋を描いて前進するようになります。

2026年までに、研究室は人の手ではなく機械がこうした手紙をやり取りする段階に入りました。ProtoPilotは、実験の意図を受け取ってプロトコル文書を作成し、それを機械が実行可能なコードへと変換して、実行結果を検証したうえでプロトコル自体を書き直すマルチエージェントシステムです [1]。Latent-Yは、単一のプロンプトから文献マイニング、標的プロファイリング、候補設計、実験的検証に至る抗体探索パイプライン全体を自律的に駆動させました [2]。また、ウェットラボでの操作においてピペットを扱うロボットアームを訓練するための身体性シミュレーションを実証した研究も現れ [3]、複数ベンダーのハードウェアを単一のソフトウェア環境に統合する自律型ラボのオペレーティングシステムもリリースされました [4]。Ginkgo Bioworksは、2026年末までに自社の全R&Dパイプラインを自律型ラボのインフラへと移行する計画を発表しました [10]。バークレーのある無機合成ラボでは、人の介入なしに17日間連続で稼働しました。

ここで1つの誤解が生じます。ロボットがワークフローを処理するため、人間が脇に追いやられるという誤解です。実際には、その逆が成り立ちます。自律型ラボが増加するにつれて、どのような問いを立てるべきかを決定する人間の役割はむしろ大きくなります。機械は実験の反復を加速させますが、自ら価値ある問いを立てることはできません。コンピュータ上で細胞の応答をシミュレートする仮想細胞モデルの忠実度を測定するベンチマークが登場しているものの、物理的なウェットラボの測定結果との間には依然として大きな乖離が残っています [6]。部門の垣根をなくしたAIネイティブなバイオテック組織が効果的に機能するかどうかを検証した研究では、従来の企業の階層構造をそのまま自律型エージェントに当てはめるだけではうまく機能しないと指摘されています [5]。

このループが途切れる地点は明確です。コンピュータシミュレーション内で強固に結合する分子であっても、物理的な細胞膜を通過できないことは頻繁にあります。マウスに対して安全な化合物が、ヒトの肝細胞では毒性を示すことも少なくありません。さらに、予測モデルの学習に用いられるデータは、成功した実験に大きく偏っています。失敗した試みは論文として発表されることが滅多になく、記録から消え去ってしまいます。この世界に存在しないデータからモデルが学習することはできません。この事実は、自律型ラボのもう一つの重要な価値を際立たせています。人間の研究者なら苛立ちのあまり廃棄してしまったかもしれない失敗実験のすべてを、ロボットは一つ残らず記録するのです。

その結果、研究室における人材の構成が変わりました。かつての製薬会社は、分子生物学者、有機化学者、計算生物学者が別々のフロアで順番に作業を進める、組み立てラインのように運営されていました。今日では、双方の言語を話せる人材が最前線でプロジェクトを率いています。ウェットラボの実験に根ざした研究者は、モデルの出力を盲目的に受け入れることはなく、その数値が生物学的に妥当であるかを問い直します。計算科学に長けた研究者は、培養皿ごとのロット間差異や試薬の純度のばらつきを、モデル内の明示的な不確実性の項へと変換します。これら2つのスキルセットが1人の人物に備わったとき、会議に費やす時間は消滅します。

労働市場のデータはこの変化を明確に反映しています。2026年第2四半期において、米国内のバイオインフォマティクス関連の求人は631件あり、提示された給与は14万8,000ドルから21万5,000ドルの範囲でした [8]。タンパク質構造モデリングと実運用レベルの機械学習（ML）の双方を扱えるシニア研究者は、24万5,000ドルから32万5,000ドルの基本給を得ています [9]。同年、米国のバイオ医薬品セクター全体では約4万2,700人分の職が削減され、2024年の2万9,000人から増加しました。それにもかかわらず、調査対象となったバイオ医薬品企業の64%が積極的な採用を継続しており、41%は2026年にさらなる人員拡大を見込んでいます [9]。一方では人員削減が行われる一方で、他方では空きポストが埋まらないままになっています。製薬企業の43%が、必要なデジタル能力を備えた応募者を見つけられないと回答しました。

大学のカリキュラムも徐々に適応しつつあります。生物学専攻の学生が微分方程式とPythonを学び、計算機科学専攻の学生が細胞生物学の実習実験を履修するダブルメジャープログラムが拡大しています。障壁となるのは成績証明書ではなく、物理的な手触りです。コーディングはオンライン講義を通じて学ぶことができますが、培地に生えたカビの匂いや、バランスが崩れた遠心分離機特有のガタつきを認識するには、実際に研究室に立たなければなりません。逆に、ベンチでの実験に10年を費やした人物は、実験ノートの生データをなぜそのままモデルの学習パイプラインに投入できないのかを理解できないことがよくあります。これら2つの感性を1人の研究者の中で融合させるには時間が必要です。このボトルネックこそが、現在の採用難の真の原因なのです。

この人材ギャップがすぐに埋まることはないと私は見ています。細胞培養で10年の経験を持つ研究者であれば、数か月でPythonを習得できるでしょう。しかし、モデルがなぜ特定の分子を選択したのかに疑問を抱く直感を養い、その懐疑心を翌週の実験アッセイ設計へと昇華させる力は、数本の論文を読んだだけで身につくものではありません。深夜2時にピペットとキーボードを交互に行き来している大学院生こそが、これからの10年を牽引していく存在となるはずです。

参考文献

[1] Jiang, Y., et al. (2026). A Self-Evolving Agentic System for Automated Generation and Execution of Biological Protocols. arXiv preprint arXiv:2606.31763. https://arxiv.org/abs/2606.31763

[2] Latent Labs Team, et al. (2026). Latent-Y: A Lab-Validated Autonomous Agent for De Novo Drug Design. arXiv preprint arXiv:2603.29727. https://arxiv.org/abs/2603.29727

[3] Liu, Z., et al. (2026). An Embodied Simulation Platform, Benchmark, and Data-Efficient Augmentation Framework for Wet-Lab Robotics. arXiv preprint arXiv:2606.12936. https://arxiv.org/abs/2606.12936

[4] Gao, J., et al. (2025). UniLabOS: An AI-Native Operating System for Autonomous Laboratories. arXiv preprint arXiv:2512.21766. https://arxiv.org/abs/2512.21766

[5] Wang, Y. (2026). Do AI-Native Biotechs Need Departments? Benchmarking Company World Models for AI-Driven Drug Development. arXiv preprint arXiv:2607.18696. https://arxiv.org/abs/2607.18696

[6] Brouwer, E. D., et al. (2026). AssayBench: An Assay-Level Virtual Cell Benchmark for LLMs and Agents. arXiv preprint arXiv:2605.10876. https://arxiv.org/abs/2605.10876

[7] Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583-589. https://doi.org/10.1038/s41586-021-03819-2

[8] CompBioJobs. (July 2026). Bioinformatics Job Market Q2 2026: 631 Jobs, $148K-$215K Avg, Salary Transparency Surges. https://www.compbiojobs.com/blog/bioinformatics-job-market-q2-2026

[9] KORE1. (2026). Biotech AI Hiring 2026: Where Bio Meets ML. https://www.kore1.com/biotech-ai-hiring-2026/

[10] Ginkgo Bioworks. (2026). Autonomous Lab. https://www.ginkgo.bio/autonomous-lab

1.2 生物学的知識の伝播と生成：学界と産業界の変化

朝8時、ある研究者がコーヒーを片手に文献検索ウィンドウを開きます。過去24時間の間に、関連する論文が40本投稿されていました。それらの要旨を読むだけでも2時間かかります。そのうち30本は読まなくてもよかったものであり、読まなかった10本の中に、彼が6か月間頭を悩ませていた問いへの答えが隠されています。どちらがそうであるかは、実際に読んでみて初めて明らかになります。

出版された知識の量は、人間が読む速度をすでに追い越してしまいました。2026年には、世界全体の学術論文の発表数が600万本を超えると予測されています。これは2025年の550万本から、わずか1年で50万本増加したことを意味します[10]。事前の査読を経ずに原稿を共有するプレプリントサーバーも、同様のペースで拡大しました。2025年単年だけでも、bioRxivとmedRxivに6万4,229本の新規原稿がアップロードされ、2024年の約5万6,000本から増加しました。これらの約80%は、平均250日をかけて後に査読付き学術誌に掲載されます。正式な出版の前に、半年以上もの間オープンに流通していることになります。

読む速度よりも先に、書く速度が変化しました。2025年に米国癌学会に投稿された7,177本の原稿のうち、36%に人工知能によって生成された文章が含まれていました。bioRxivに論文を投稿する生命科学分野の研究者の生産量は52.9%増加しました[10]。それと同時に、学術誌の編集者は査読者の確保に苦慮しています。投稿数が急増する一方で、人間の査読者の総数は横ばいのままだからです。査読を人工知能に任せられるかどうかを検証した研究は、モデルが原稿の実際の欠陥を特定する能力と、欠陥を見つけたと主張する傾向とは別物であると警告しています[6]。

従来の知識獲得方法は、手作業による蓄積でした。ひとつの標的タンパク質を理解するためには、過去の数百本に及ぶ論文を手作業で読み込み、まとめなければなりませんでした。その過程で、他のチームが記録した毒性による失敗報告を見落とし、再び同じ袋小路に行き当たることもしばしばありました。物理学や材料科学から生まれた測定技術がキャンパスの中庭を渡るまでに、たとえ隣の研究棟にあったとしても、何年もかかることがよくありました。こうした知識の分断は「知識のサイロ（Knowledge Silo）」と呼ばれます。同じキャンパスで働きながら、完全に別々の階で生活しているような状態です。

大規模言語モデル（LLM）と生物医学知識グラフは、こうしたサイロを打破しました。知識グラフとは、遺伝子、タンパク質、薬剤、疾患をノードとし、それらを関係性を示すエッジで結んだ地図のようなものです。地下鉄の路線図を思い浮かべるとわかりやすいでしょう。駅が遺伝子を表し、路線が相互作用を表しています。ある駅を出発した薬剤が、数駅先にある疾患にどのように到達するかを視覚的にたどることができます。このような地図を手作業で論文を読んで作成し

ようとすれば、人間の生涯ほどの時間がかかってしまいますが、モデルは何百万本もの論文テキスト、特許、臨床試験の記録を一括して読み込み、そのつながりを描き出します。課題は、この地図が毎月描き直される点にあります。2026年には、生物医学知識グラフが拡張され構造が変化しても、モデルが過去に獲得した関連性を忘れないようにする継続学習手法が導入されました[3]。言語モデルが生物医学ツールを直接呼び出すよう訓練するためのデータセットも公開されました[1]。メタボロミクスにおいては、統合された情報源で学習した領域特化型モデルが代謝物の相互作用グラフを予測し始めました[2]。研究論文から微生物の代謝物生合成データを抽出・構造化する自動文献マイニングシステムも登場しました[4]。

科学的な問いかけの構造も今や変化しました。研究者が「KRAS G12D変異膵臓がんにおいて免疫チェックポイント阻害薬への耐性を誘導する代謝バイパス経路を特定し、その経路を遮断する合成致死標的を発見せよ」と入力します。数秒のうちに、モデルは酵素反応ネットワーク、動物モデルのトランスクリプトームデータ、失敗した臨床試験の記録などを精査し、優先順位付けされた候補リストと実験計画の草案を返します。数年前であれば、この作業は博士研究員の丸2か月もの時間を費やしていたはずです。

AI研究エージェントは、このワークフローをさらに一歩先へと進めます。FutureHouseが開発したRobinは、自律的に仮説を生成し、実験を設計し、結果を解釈することで、緑内障治療薬であるリパスジルを乾燥型加齢黄斑変性の治療候補として特定しました。最初の構想から論文の完成までにかかった期間は2か月半でした[8]。同機関が開発したKosmosは、個人の研究者なら6か月かかる分析をわずか1日で完了できると謳っており、その推定精度は約80%に達します。課題となるのは、残りの20%です。大腸がんの脆弱性を調査するマルチスケールAIスウォーム研究では、モデルの尤もらしい推論と、物理的な哺乳類細胞の生物学との間にあるギャップを埋めることが、依然として主要なボトルネックであることが浮き彫りにされています[7]。同年には、エージェント時代における科学的発見を再定義する提案も登場しました[5]。

知識が伝達される経路もまた進化しました。かつては、学会のロビーが情報の発信拠点となっていました。発表後のコーヒーブレイクで交わされる10分間の雑談が、その後の研究の方向性を左右していたのです。今日では、プレプリントがオンライン上に公開された瞬間に、世界中の研究室がまさにその日のうちにその方法論を再現します。論文とともにソースコードと学習済みモデルの重みを公開することが標準的な慣行となったため、再現にかかる期間は数か月からわずか1日へと短縮されました。その代償として、検証されていない主張も同じ速度で拡散することになります。査読前の原稿が6か月間出回っている間に、その知見を引用する複数の後続論文がすでに積み上がってしまいます。これは、安全装置を犠牲にして速度を手に入れたトレードオフといえます。

産業資本はこの変化にいち早く着目しました。Alphabet傘下のIsomorphic Labsは、21億ドルのシリーズB投資ラウンドを調達し、2026年上半期におけるこの分野のベンチャー投資全体の約87%を占めました。同社はイーライリリーおよびノバルティスと提携を結びました。リリーとの契約には4500万ドルの契約一時金と最大17億ドルの成果連動型マイルストーンが含まれ、ノバルティスとの契約には3750万ドルの契約一時金と最大12億ドルに達するマイルストーンが含まれていました。これら2つの提携を合わせると30億ドル規模に迫ります[9]。2026年1月には、イーライリリーとNVIDIAが創薬に特化したAI研究施設を設立するため、5年間で10億ドルを共同投資することに合意しました。サノフィは自己免疫疾患に関する単一の契約において、1億6000万ドルの契約一時金と最大25億6000万ドルのマイルストーン支払いを約束しました。Recursionは競合するExscientiaを6億8800万ドルで買収しました。

見出しに躍る数字は華々しい祝祭のように見えますが、臨床での実績は依然として冷静な現実を物語っています。AIが見出した173のプログラムが臨床開発段階に入っているものの、最終的な薬事承認を獲得したものはまだ一つもありません。第1相の通過率は80%を超えるものの、第2相では約40%にまで低下します。それでも、私はこの変化をバブルだとは考えていません。Insilico Medicineは、標的の発見から第2相臨床試験まで創薬候補化合物をわずか30か月で進展させました。従来の手作業による創薬で同じ道のりを辿るには、6年から8年を要します。その薬剤に最終的な有効性があるかどうかは、まだ見極める必要があります。しかし、開発の時計はすでにまったく異なる速度で動いています。

学界と産業界の関係は逆転しました。かつては、大学が基礎的な原理を解明して論文を発表し、製薬企業が数年後にその知見を取り入れて開発を進めていました。今日では、膨大な計算インフラを持つ企業が生物学の基盤モデルを構築し、大学がそのライセンスを受けて利用しています。ハイスループットスクリーニング（HTS）機器を持たない小規模な研究室であっても、クラウド上で数億個もの仮想分子をフィルタリングできます。その見返りとして、大手製薬企業は大学病院が収集した患者コホートのデータセットを学習用入力として取り込んでいます。双方が互いにとってのモジュール型コンポーネントとなっているのです。

この逆転には弊害も伴います。生物学の基盤モデルを学習させるために必要な計算資源は、大半の大学の年間研究予算を超えています。モデルを構築する側とライセンスを受けて利用する側が分かれると、研究課題を決定する主導権は構築する側へと傾きます。企業が非公開を選択した学習データセットは論文に現れず、非公開データは第三者が独立して検証することができません。そのため、オープンなデータセットと公開ベンチマークを維持することが、かつてないほど重要になっています。全員が同じ試験問題に基づいて採点されて初めて、科学的な価値とマーケティング上の主張を区別できるのです。

研究者の日常的なルーティンも変化しました。毎朝文献を読むのに費やしていた2時間は30分に短縮され、その代わりに、モデルが生成した要約が正確かどうかを原著論文に戻って確認する時間が増えました。要約が誤りを犯す箇所は予測可能です。実験条件の例外、統計的検定の前提条件、そして著者が慎重に付け加えた限定的な記述です。これら3つの点をチェックする規律を保てるかどうかが、研究者の力量を決定づけます。ツールの処理が加速するほど、懐疑的な視点を保つという規律の価値は高まります。

研究者にとって核心となる問いは、一点に絞られます。何を記憶しているかではなく、何を問うのかということです。今や機械のほうが素早く答えを見つけ出します。まだ誰も投げかけていない問いを見極める眼識は、厳然として人間の側に残されています。

参考文献

[1] Gao, X., et al. (2026). BioTool: A Comprehensive Tool-Calling Dataset for Enhancing Biomedical Capabilities of Large Language Models. arXiv preprint arXiv:2605.05758. https://arxiv.org/abs/2605.05758

[2] Ku, D., et al. (2026). MetaboLLM: a metabolomics-specialized large language model for biochemical knowledge integration and predictive metabolite graph construction. arXiv preprint arXiv:2608.06253. https://arxiv.org/abs/2608.06253

[3] Radwan, Y. A., et al. (2026). CMKL: Modality-Aware Continual Learning for Evolving Biomedical Knowledge Graphs. arXiv preprint arXiv:2605.10510. https://arxiv.org/abs/2605.10510

[4] Sun, X., et al. (2025). Literature Mining System for Nutraceutical Biosynthesis: From AI Framework to Biological Insight. arXiv preprint arXiv:2512.22225. https://arxiv.org/abs/2512.22225

[5] Zheng, Y., et al. (2026). Rethinking Scientific Discovery in the Agentic Era. arXiv preprint arXiv:2607.03863. https://arxiv.org/abs/2607.03863

[6] Wang, J., et al. (2026). When AI reviews science: Can we trust the referee?. arXiv preprint arXiv:2604.23593. https://arxiv.org/abs/2604.23593

[7] Baker, C., et al. (2026). Autonomous mechanistic discovery of colorectal cancer vulnerabilities via multi-scale AI swarms. arXiv preprint arXiv:2607.16262. https://arxiv.org/abs/2607.16262

[8] Chemical & Engineering News. (May 2026). AI companies introduce new agent-based tools for scientific discovery. https://cen.acs.org/pharmaceuticals/drug-discovery/ai-companies-introduce-agent-based-research-tools/104/web/2026/05

[9] BioSpace. (2026). Lilly, Novartis Sign AI Partnership with Alphabet's Isomorphic. https://www.biospace.com/lilly-novartis-sign-ai-partnership-with-alphabet-s-isomorphic

[10] Futurity Publishing. (2026). AI Is Reshaping Scientific Publishing in 2026: Biology and Medicine Are Growing Fastest. https://futurity-publishing.com/ai-scientific-publishing-2026/

第2章：生体データの幾何学的表現学習とマルチモーダルアライメント

2.1 タンパク質・遺伝子・テキストのモダリティを横断するマルチモーダルアライメントと対照学習

研究室のホワイトボードの前に3人の研究者が立っています。まったく同じタンパク質を見つめながら、それぞれがまったく異なるノートを開きます。配列を扱う研究者は、20種類のアミノ酸を表す1文字表記の長い連なりを書き留めます。構造を扱う研究者は、個々の原子ごとにx、y、z座標という3つの数値を書き留めます。文献を読む研究者は、「このキナーゼは、G1/S期移行時に標的タンパク質のセリン残基にリン酸基を転移する」と書きます。3冊のノートには共通する文字が1文字もありません。それでも3冊すべてが、同じ物理的分子を指し示しています。

生命は、自らの内部情報を常に異なる表現へと変換しています。遺伝情報は、4種類の化学塩基からなる1次元の配列として保存されます。転写と翻訳によって、その暗号はアミノ酸の鎖へと変換されます。原子間の引力と斥力に導かれ、その鎖は3次元空間の中で折りたたまれます。折りたたまれたタンパク質は他の分子と結合し、細胞内のシグナルを伝達します。人間は、その結果として得られる生物学的観察結果を学術論文の文章として記録します。物理的な形式は4回変化しますが、その根底にある実体は1つのままです。

過去10年間のモデルは、こうした表現のうち一度に1つしか扱うことができませんでした。配列専用の言語モデルは線形な文字列を読み、構造専用のモデルは空間座標を解析し、生物医学の言語モデルは研究文献を個別に処理していました。その結果、一次配列は判明しているものの三次構造が未解明であるタンパク質に直面した際、ワークフローは停滞しました。また、新たに構造が解明されたものの機能記述のアノテーションが付与されていないタンパク質に遭遇した際にも、同様に停滞しました。3冊のノートを並べて同時に読み解くことができるシステムは存在しなかったのです。

マルチモーダルアライメントとは、異種多様なデータ形式を共有された潜在空間へと埋め込む手法です。これは、司書が主題に厳格に従って本を書架に並べる作業に似ています。表紙の色や製本様式が異なっていても、同じ主題を共有する本は隣同士に並べられます。この体系化を推進するアルゴリズムのエンジンが対照学習です。同じタンパク質に由来する配列ベクトル、構造座標ベクトル、テキスト記述ベクトルは正例ペア（Positive Pairs）として結びつけられます。無関係なタンパク質同士は負例ペア（Negative Pairs）を形成します。最適化はInfoNCE損失関数を用いて行われます[1]。正例ペアは互いに引き寄せ合い、負例ペアは互いを押し離します。CLIPは2021年に画像と自然言語の文章を結びつけることでこのパラダイムを実証し[2]、計算生物学はその数学的構文をそのまま取り入れました[3]。

アライメントという概念は抽象的に聞こえるかもしれませんが、数値に置き換えるとその仕組みは明快になります。配列エンコーダーはアミノ酸鎖を取り込んで1,024次元のベクトルを出力します。構造エンコーダーは原子座標を受け取り、同一の長さのベクトルを出力します。テキストエンコーダーは説明文を処理し、同じ次元のベクトルを生成します。当初、これら3つのベクトルの間には数学的な関係性はありません。しかし学習が進むにつれて、同一のタンパク質に由来する3つのベクトル間のコサイン類似度は0.1から0.9近くまで上昇します。損失関数の内部には、温度パラメータと呼ばれる小さなスカラー値が存在します。この値を下げると、モデルは紛らわしい負のペアに対してより厳しくペナルティを科すようになり、上げるとペナルティは緩和されます。研究プロジェクト全体の成否が、しばしばこの単一の数値ハイパーパラメータにかかっています。

最初期のブレイクスルーは、生物学的データベースの検索において現れました。配列、構造、機能テキストを同時に対照学習によってアライメントさせたProTrekは、Swiss-Protから精選された1,400万ペアと、UniRef50からフィルタリングされた2,500万ペアからなる計4,000万ペアで学習されました [4]。配列から機能への検索において、性能は従来手法と比べて30倍に向上し、機能から配列への検索は60倍に向上しました。タンパク質同士の構造比較は、FoldseekやMMseqs2よりも100倍高速に動作しました。ProTrekは、11の下流ベンチマークタスクのうち9つでESM-2を上回る性能を示しました。このモデルをホストする公開サーバーでは、50億個のタンパク質の埋め込みがあらかじめ計算されています。

学習データの品質と構成は、下流の成果を左右します。Swiss-Protのアノテーションは、手作業で精選され人間によって検証された文章で構成されています。一方、UniRef50から収集された2,500万件のレコードには、かなりのノイズが含まれています。ノイズを含むペアは精度を低下させると考えられがちですが、実証的な結果はまさにその正反対でした。幅広くノイズの多いデータセットが、狭く純粋なコレクションによって生じた構造的な空白を埋めたのです。正式な標準語のみで辞書を編纂すると現場で使い物にならなくなるのと同様です。その後、ProTrekの公開サーバーにはNCBIからさらに7億件のタンパク質配列が統合されました。

2026年6月、ある基盤モデルがこれら3つのモダリティすべてを単一のアーキテクチャへと統合しました。BioMatrixは、低分子化合物の配列と3Dコンホメーション、タンパク質の配列と構造、そして自然言語を単一の離散トークン空間へとマッピングします [5]。外部の専用エンコーダーやモダリティ固有の出力ヘッドは一切使用していません。言語モデルQwen3の1.7Bおよび4Bパラメータ版をベースに構築され、追加で3,044億個のマルチモーダルトークンを学習しました。6つの生物学的カテゴリーにまたがる80の評価タスクにおいて、77のベンチマークで最高水準あるいはそれに近い結果を達成しました。以前のアーキテクチャに対する際立った優位性は、読み取ることができるあらゆるモダリティを直接生成することもできる点にあります。

空間構造のトークン化もまた軽量化が進んでいます。2026年5月に公開されたYetiは、Look-up Free Quantizationを用いて3D原子座標を離散トークンへと変換します [6]。ESM3の10分の1のパラメータ数で動作しながら、より高いコードブック利用率とトークンの多様性を達成しました。大規模な事前学習を行うことなく、配列と3D座標を同時に生成するようゼロから学習され、自身の10倍の規模を持つモデルに匹敵する生成忠実度を示しました。容量の大きい空間座標ファイルをコンパクトなトークン列へと変換する処理が、今や研究室のワークステーション1台で実行できるようになっています。

アライメントが成功すると、検索インターフェースは一変します。研究者が「キナーゼの触媒ポケットに結合し、少なくとも3本の水素結合を形成するタンパク質ドメインを特定せよ」というプロンプトを入力すると、モデルはそのテキストプロンプトの埋め込みベクトルに最も近い構造コンホメーションやアミノ酸配列を検索して提示します。ウェットラボでの実験を経たことのない未知のタンパク質が、候補として浮上します。タンパク質のペアがどのように相互作用するかを自由な自然言語で記述できるシステムが2026年5月に登場し[7]、同様のアライメント原理によってゼロショット条件付きRNA設計も可能になりました[8]。

しかし、アライメントは融合と同義ではありません。対照学習によって結びつけられた2つのモダリティは、潜在空間において物理的に重なり合うわけではないのです。画像ベクトルとテキストベクトルは、それぞれ別々の狭い円錐状の多様体内にクラスターを形成し、両者の間には空白の回廊が残されます。この現象はモダリティギャップ（Modality Gap）と呼ばれています[9]。ニューラルネットワークは層を経るごとに角度のばらつきを狭める傾向があるため、2つのクラスターは初期化の段階で乖離し、そのまま別個のクラスターとして固定されます。この状態でもコサイン類似度による順位付けには十分ですが、2つの空間が単一の連続体として統合されたと主張するのは誤りです。

このギャップがなぜ残り続けるのかについては、さまざまな説明がなされています。重みの初期化に伴う幾何学的なアーティファクトを指摘する見解もあれば、対照損失に内在する数学的性質に起因するとする見解もあります。最適化ステップの不足を反映しているという主張や、データ表現のバイアスを指摘する声もあります。温度パラメータの値や埋め込みの次元数、モダリティ間におけるデータセット規模の不均衡は、いずれもギャップの幅を変化させます。実用面において、このギャップは検索時ではなく生成時に障害となります。テキストベクトルの座標から直接3次元構造を生成しようとしても失敗するのは、その領域に構造ベクトルが存在しないためです。それは、空白の回廊の上に家を建てようとするようなものです。

連続座標を離散トークンへとマッピングする際に生じるコストも測定されています。2026年4月に発表された論文は、これを幾何学的アライメント税（Geometric Alignment Tax）と名付け

ました[10]。滑らかで連続的な幾何学的曲面を離散的なビンに無理やり押し込めると、本来の空間距離の尺度が歪んでしまいます。同一のエンコーダーアーキテクチャにおいて、離散的な交差エントロピー損失を連続的な回帰ヘッドへと置き換えたところ、幾何学的な歪みは8.5分の1に低減しました。さらに際立った発見は、3つの構造コンホメーションを比較した際の結果です。連続的な幾何学的目的関数の下ではそれらの距離の差が1.3倍にとどまっていたのに対し、離散的なトークン化の下では3,000倍にも乖離しました。トークン化には代償が伴うのです。

アライメントされた潜在空間は、実験の手順そのものを再構築します。従来、ヒット化合物を特定するには、標的タンパク質を選定し、数十万種類もの候補化合物を合成して順番にスクリーニングする必要がありました。現在では、研究者が標的の特性をテキストで記述し、潜在空間内でそのテキストベクトルの近傍に位置する数百種類の候補だけを合成すればよくなっています。2026年4月の研究では、触媒残基をあらかじめ指定することなく、新規酵素の設計に成功しました[11]。この成果は、基盤となるモデルが配列、構造、機能の記述を統合的に処理できたからこそ可能になったものです。合成化学が探索してきた反応空間は、進化が生み出したもののほんの一部にすぎません。DNAを通じてプログラム可能な化学空間は、はるかに広大です。

私は、この技術を「翻訳」と表現するほうがより正確であると考えています。私たちは、3つの生物学的言語を並べた多言語辞書を構築しました。この辞書によって、探索は加速し、実験候補のリストは絞り込まれます。しかし、優れた辞書があるからといって、別個の2つの言語が同一になるわけではありません。次世代の課題は、パラメータ数を増やすことではなく、幾何学的代償をより小さく抑える数学的表現を発見することにあります。

参考文献

[1] van den Oord, A., Li, Y., & Vinyals, O. (2018). Representation Learning with Contrastive Predictive Coding. arXiv preprint arXiv:1807.03748. https://arxiv.org/abs/1807.03748

[2] Radford, A., Kim, J. W., Hallacy, C., et al. (2021). Learning Transferable Visual Models From Natural Language Supervision. arXiv preprint arXiv:2103.00020. https://arxiv.org/abs/2103.00020

[3] Xu, M., Yuan, X., Miret, S., & Tang, J. (2023). ProtST: Multi-Modality Learning of Protein Sequences and Biomedical Texts. arXiv preprint arXiv:2301.12040. https://arxiv.org/abs/2301.12040

[4] Su, J., He, Y., You, S., et al. (2025). A trimodal protein language model enables advanced protein searches. Nature Biotechnology. https://doi.org/10.1038/s41587-025-02836-0

[5] Pei, Q., et al. (2026). BioMatrix: Towards a Comprehensive Biological Foundation Model Spanning the Modality Matrix of Sequences, Structures, and Language. arXiv preprint arXiv:2606.22138. https://arxiv.org/abs/2606.22138

[6] Giri, N., Farrell, S., & Bouchard, K. E. (2026). Yeti: A compact protein structure tokenizer for reconstruction and multi-modal generation. arXiv preprint arXiv:2605.09981. https://arxiv.org/abs/2605.09981

[7] Fei, X., et al. (2026). PPI2Text: Captioning Protein-Protein Interactions with Coordinate-Aligned Pair-Map Decoding. arXiv preprint arXiv:2605.08924. https://arxiv.org/abs/2605.08924

[8] Klypa, R., et al. (2026). Multimodal Alignment and Preference Optimization for Zero-Shot Conditional RNA Generation. arXiv preprint arXiv:2605.23961. https://arxiv.org/abs/2605.23961

[9] Liang, V. W., Zhang, Y., Kwon, Y., Yeung, S., & Zou, J. (2022). Mind the Gap: Understanding the Modality Gap in Multi-modal Contrastive Representation Learning. arXiv preprint arXiv:2203.02053. https://arxiv.org/abs/2203.02053

[10] Raju, P. C. (2026). The Geometric Alignment Tax: Tokenization vs. Continuous Geometry in Scientific Foundation Models. arXiv preprint arXiv:2604.04155. https://arxiv.org/abs/2604.04155

[11] Rector-Brooks, J., et al. (2026). General Multimodal Protein Design Enables DNA-Encoding of Chemistry. arXiv preprint arXiv:2604.05181. https://arxiv.org/abs/2604.05181

2.2 高次元マルチオミクスデータを低次元潜在表現へと圧縮する埋め込みの原理

1個の細胞をシーケンサーに投入すると、2万個の数値からなる配列が得られます。それぞれの値は、特定の遺伝子の転写産物が何回読み取られたかをカウントしたものです。画面上では、スプレッドシートが右へと果てしなく広がっています。しかし、その2万個の項目の大半はゼロです。遺伝子が真に発現していない「真のゼロ」と、遺伝子は活性化していたものの機器の検出をすり抜けてしまった「偽のゼロ」が混在しています。後者はドロップアウト（Dropout）と呼ばれます。こうしたデータセットを初めて開いた大学院生は、これら2種類のゼロを区別する単純な方法が存在しないという現実に直面することになります。

数値の数が膨大であるからといって、情報量が多いとは限りません。2万次元の空間では、すべてのデータポイント間の距離がほぼ等しくなってしまいます。距離を計算しても、異なるクラスターを効果的に分離することはできません。この現象は「次元の呪い」として知られています。ここに測定ノイズが加わると、2つの細胞が同一の生物学的サブタイプに属するのか、それとも異なる細胞系譜に属するのかを判断することが不安定になります。

幸いなことに、遺伝子は独立して機能しているわけではありません。細胞周期を制御する遺伝子モジュールは一斉に活性化します。エネルギー代謝や免疫カスケードを調節するモジュールも同様です。これは、2万個の照明器具がわずか数十個のスイッチに配線されているオフィスビルに似ています。名目上の見かけの次元は2万ですが、独立した制御レバーの数ははるかに少ないのです。この薄く連続的な幾何学的部分空間は、「低次元多様体」と呼ばれます。

変分オートエンコーダー（VAE）は、こうした根底にあるレバーを発見するために設計されたアーキテクチャです。エンコーダーとなるニューラルネットワークが、2万次元の入力ベクトルを32次元または64次元の確率分布へと圧縮します。そしてデコーダーネットワークが、このコンパクトな潜在ベクトルだけを用いて元の2万個の遺伝子カウントを再構成します。学習では、出力が入力とどれだけ正確に一致するかを測る再構成誤差と、潜在分布が標準的な事前分布からどれだけ乖離しているかを測るカルバック・ライブラー情報量という2つの目的関数項を同時に最小化します。2018年に発表されたscVIは、このフレームワークを1細胞データへと厳密に適用した最初のモデルでした[1]。転写産物のカウントを負の二項分布でモデル化することにより、技術的なドロップアウトと本物の生物学的ゼロを分離することに成功しました。

デコーダーのタスクを詳しく見ると、なぜこのアーキテクチャがノイズに対して頑健であるのかがよく分かります。デコーダーは、わずか32次元のベクトルから2万個の遺伝子発現値を再構成しなければなりません。情報容量が制限されているため、モデルは遺伝子ネットワークの共発現ルールを内面化せざるを得なくなります。たとえば、ある特定の細胞周期マーカーが活性化して

いれば、関連する経路の遺伝子群も活性化している可能性が極めて高いということを学習します。これらの生物学的な依存関係がネットワークの重みに埋め込まれると、測定機器のドロップアウトによって生じた欠測のゼロは、共調節されている遺伝子を参照することで補完できるようになります。データ圧縮は、実質的に自動的な欠損値補正エンジンとして機能するのです。

次元削減は、生物学的意味を保ちながら技術的ノイズを排除します。異なるシーケンシングの実行（ラン）から生じるばらつきは、真の生物学的プロファイルを再構成する上での有用性が極めて低く、潜在ベクトルの中に持続的な表現として定着することはありません。これに対して、細胞の分化状態、組織の同一性、腫瘍の進行段階などは正確な再構成に不可欠であり、保持されます。このように圧縮された潜在空間の中では、距離尺度が生物学的な意味を取り戻します。もし希少がん患者のトランスクリプトームベクトルが免疫療法応答性クラスタの隣にマッピングされれば、その治療法の適用を検討する合理的な根拠となります。正常細胞から悪性化に至る軌跡を逆にたどれば、分岐点における極めて重要な転写因子を特定することもできます。

潜在空間は視覚的な誤解を招くことがあります。2万次元を2次元の2軸へと投影するUMAPプロットは、色鮮やかな群島のように見えます[10]。そして、島と島の間の距離は生物学的な関係性として安易に解釈されがちです。しかし、UMAPは局所的な近傍のトポロジーを保持する一方で、大域的な距離を歪めてしまいます。2次元平面上で遠く離れて描かれている二つの細胞クラスタが、高次元の生物学的空間においても必ずしも遠いとは限りません。プレゼンテーションで見栄えのするプロットであっても、慎重に解釈する必要があります。厳密な結論は、2次元の可視化からではなく、32次元の潜在ベクトルそのものから導き出さなければなりません。

ここまでは教科書的な理解の概要です。2023年以降、この領域にははるかに大規模な基盤モデルが参入してきました。代表的な構造としては、3300万個の細胞で学習されたscGPTのような1細胞基盤モデルが挙げられます[2]。これらは遺伝子をトークンとして扱い、Transformerによって処理します。しかし、2026年の検証研究によって、こうしたアプローチに対して重大な疑義が投げかけられることになりました。

2026年2月18日に発表された論文は、冷静な評価を下しました[3]。ボストン大学の2名の研究者は、標準的な正規化と線形手法を組み合わせた手堅いパイプラインが、複数のベンチマークにおいて大規模な基盤モデルと同等かそれ以上の性能を発揮することを実証しました。未見の細胞型や未知の生物種を扱う分布外（out-of-distribution）タスクでは、単純な線形ベースラインのほうが大規模モデルを凌駕したのです。これは、細胞状態に関する生物学の主要な要素が、線形表現によって効果的に捉えられることを示しています。2026年の他のベンチマークでも同様の結果が報告されており、さまざまな生物学的システムにおいて、ゼロショット埋め込みの性能が高変動遺伝子を選択する基本的なベースラインを下回りました。

なぜこのようなことが起きたのでしょうか。自然言語とは異なり、遺伝子発現表には本質的な順序が存在しません。文章では単語の順序が意味を決定しますが、表における遺伝子の列の順序は任意です。Transformerは連続する文脈の解析を得意としますが、この問題の構造ではその能力を発揮する余地が限られていました。さらに、評価ベンチマークが寛容すぎた面もあります。細胞型の境界が明確で離散的なデータセットでは、基本的な主成分分析（PCA）でも十分に明瞭な分離が可能です。真の難所は、境界が重なり合い連続的なグラデーションを描く複雑な組織にあり、そこではあらゆる計算手法が一様に苦戦を強いられています。

生物学の大規模モデルにおける内部表現についても詳細な検証が行われてきました。2026年2月25日のある研究では、52の評価を通じて141件の幾何学的および位相幾何学的（トポロジー的）仮説を立てて検証しました[4]。そのうち27件が肯定的に確認され、21件があいまいで、63件が決定的に否定されました。最も厳格な基準を適用すると残ったのは15件未満であり、妥当な肯定的知見の割合は約10%にまで低下しました。それでもなお、有意義な構造は残っていました。遺伝子の埋め込み近傍は真正なトポロジー的秩序を示し、多様体に沿った測地線距離はユークリッド距離よりも調節相互作用を良好に予測しました。正準相関分析によってscGPTとGeneformerの位置合わせを行ったところ、相関は0.80に達しました。個々の遺伝子がどこにマッピングされるかについては見解が分かれたものの、両モデルは遺伝子空間全体の幾何学的構造について一致していました。

これらのモデルを単に切り捨てるべきではありません。2026年3月、研究者たちはscGPT内部から造血系譜の分化を記述する多様体を抽出し、その潜在幾何から直接、機能的アルゴリズムを導き出しました[5]。また、構造のわずかな改良によっても顕著な性能向上が得られます。例えば、ソフトマックスアテンションをシグモイドアテンションに置き換えたことで、6つのデータセット全体において細胞型のクラスタリング分離度が25%向上しました[6]。スパースオートエンコーダーを用いて表現を精査したところ、明示的な調節ロジックが著しく乏しい一方で、整然とした生物学的オントロジーが存在することが明らかになりました[7]。

ここで、私は明確な立場をとります。過去3年間、この分野はパラメータ規模を主要な評価基準として扱ってきました。その傾向は論文の出版数をいたずらに増やしたものの、臨床現場での実用性は限られたものでした。医療機関が必要としているのは、新たに入院した患者の組織に適用した際にも揺るぎなく機能する妥当な表現です。学習データに含まれていない細胞型において線形ベースラインモデルが大規模モデルを上回ったという事実は、これら大規模なアーキテクチャが生物学の基本法則ではなく、データセット固有の癖を記憶していたにすぎないことを示唆しています。

資本と研究の焦点は、データ生成層へと移行しつつあります。50種のがんモデルと1,100種類以上の創薬化合物にわたる1億個の細胞および6万件の摂動実験を網羅したTahoe-100Mアトラスが公開されました[8]。Arc InstituteのVirtual Cell Atlasは6億個以上の細胞を集約しました。細胞イメージングの分野では、単一のJUMP Cell Paintingデータセットだけで115テラバイトに達します[9]。114カ国から5,000人以上が参加した2026年のVirtual Cell Challengeにおいて、上位入賞者の提出物は単一の巨大な基盤モデルに依存するのではなく、疑似バルク（pseudobulk）要約統計量と順位に基づく検定に依拠していました。

計測技術は拡張を続けています。現在では、同一の単一細胞内において遺伝子発現、クロマチンアクセシビリティ、タンパク質のリン酸化状態を同時に読み取ることができます。分子カウントに組織の2次元座標を加える空間トランスクリプトミクスも標準的な手法となりました。特徴量の軸が増えるにつれて、圧縮はより高度な処理を要求するようになります。異質な測定モダリティを単一の潜在ベクトルへと統合するためには、まず各測定軸にどれだけの重みを割り当てるかを決めなければなりません。前節で観察されたモダリティギャップはここでも再び現れます。単一の細胞から得られた2つの異なる測定ベクトルが、同一の近傍へとマッピングされないのです。

圧縮という全体的な方向性そのものは正しいものでした。2万個の数値次元を数十次元にまで削減することは、計算生物学において依然として中核を占めています。しかし、圧縮エンジンの物理的な規模を拡大することと、真に効果的な圧縮を達成することとは全くの別物です。2026年の実証結果は、この両者の間に明確な一線を画しました。徹底した正規化、厳密なベースライン、そして分布外（out-of-distribution）の検証は、依然として不可欠です。これら3つの柱を迂回したモデルは、既存のベンチマークで高いスコアを獲得できたとしても、未知の組織を前にしたときには立ち行かなくなります。

参考文献

[1] Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., & Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. Nature Methods, 15(12), 1053-1058. https://doi.org/10.1038/s41592-018-0229-2

[2] Cui, H., Wang, C., Maan, H., et al. (2024). scGPT: toward building a foundation model for single-cell multi-omics using generative AI. Nature Methods, 21(8), 1470-1480. https://doi.org/10.1038/s41592-024-02201-0

[3] Souza, H., & Mehta, P. (2026). Parameter-free representations outperform single-cell foundation models on downstream benchmarks. arXiv preprint arXiv:2602.16696. https://arxiv.org/abs/2602.16696

[4] Kendiukhov, I. (2026). What Topological and Geometric Structure Do Biological Foundation Models Learn? Evidence from 141 Hypotheses. arXiv preprint arXiv:2602.22289. https://arxiv.org/abs/2602.22289

[5] Kendiukhov, I. (2026). Discovery of a Hematopoietic Manifold in scGPT Yields a Method for Extracting Performant Algorithms from Biological Foundation Model Internals. arXiv preprint arXiv:2603.10261. https://arxiv.org/abs/2603.10261

[6] Sadashivaiah, V., et al. (2026). Better Models, Faster Training: Sigmoid Attention for single-cell Foundation Models. arXiv preprint arXiv:2604.27124. https://arxiv.org/abs/2604.27124

[7] Kendiukhov, I. (2026). Sparse autoencoders reveal organized biological knowledge but minimal regulatory logic in single-cell foundation models: a comparative atlas of Geneformer and scGPT. arXiv preprint arXiv:2603.02952. https://arxiv.org/abs/2603.02952

[8] Zhang, J., et al. (2025). Tahoe-100M: A Giga-Scale Single-Cell Perturbation Atlas for Context-Dependent Gene Function and Cellular Modeling. bioRxiv. https://doi.org/10.1101/2025.02.20.639398

[9] Muñoz, A. F., et al. (2026). JUMP-lite: Compact, reproducible benchmarking of cell representations. arXiv preprint arXiv:2608.07632. https://arxiv.org/abs/2608.07632

[10] Becht, E., McInnes, L., Healy, J., et al. (2019). Dimensionality reduction for visualizing single-cell data using UMAP. Nature Biotechnology, 37(1), 38-44. https://doi.org/10.1038/nbt.4314

第3章：タンパク質立体構造予測の革新

3.1 アミノ酸配列からの高精度立体構造予測の原理と数学的基礎

物語は一本の試験管から始まります。1950年代後半、米国国立衛生研究所のクリスチャン・アンフィンセンは、リボヌクレアーゼと呼ばれる酵素を高濃度の尿素溶液に浸しました。リボヌクレアーゼはRNAを切断する微小なハサミのように機能します。尿素は折りたたまれたポリペプチド鎖を変性させます。実験台のガラス瓶の中で、酵素はもつれた糸のようになりました。そして、RNAを切断する能力は失われました。

アンフィンセンはそこで止まりませんでした。彼は尿素をゆっくりと透析して除去しました。すると、ほどけていた鎖は自発的に再び折りたたまれました。分子のハサミは再びRNAを切断するようになったのです。人の手を一切介することなく、その鎖は本来の天然構造を見つけ出しました。この実験から導き出された結論は「アンフィンセンのドグマ」として知られています。すなわち、タンパク質の天然の立体構造は、そのアミノ酸配列の中にすべてコードされているというものです。天然構造は自由エネルギーの大域的最小値に対応します [1]。アンフィンセンは1972年にノーベル化学賞を受賞しました。

タンパク質においては、形が機能を決定します。酵素が化学反応を触媒するには、精密な形状の触媒部位を持つ必要があります。抗体が病原体を中和するためには、病原体の分子表面に適合する結合腕が必要です。コラーゲン繊維が皮膚や腱にかかる機械的張力を支えるには、三重らせんのロープ状に巻きつく必要があります。たとえ一次配列が同一であっても、三次元の幾何学的構造が崩れれば、タンパク質はその生化学的有用性を失います。その結果、一次元の配列から直接その空間構造を解読することは、数十年にわたり生命科学における中心的な課題であり続けました。

その障壁となったのは計算の実行可能性でした。1969年、サイラス・レビンサールは紙の上で組み合わせの可能性を計算しました [2]。100個のアミノ酸からなる小さなタンパク質を考えてみます。主鎖の二面角であるφ（ファイ）角とψ（プサイ）角、そして側鎖のχ（カイ）角は、それぞれ複数の離散的な状態を取り得ます。控えめに見積もっても、取り得る異なるコンホメーションの数は10の47乗を超えます。もし1つのコンホメーションを探索するのに1ピコ秒（1兆分の1秒）かかるとすれば、その空間全体を探索するには宇宙の年齢よりも長い時間が必要になります。それにもかかわらず、生きた細胞の内部では、タンパク質は1秒未満で折りたたまれます。この逆説は「レビンサールのパラドックス」として知られています。

その解決の鍵はエネルギーランドスケープにありました。折りたたまれるポリペプチドは、すべてのコンホメーションを無作為に探索するわけではありません。自由エネルギーランドスケープは平らな平面ではなく、漏斗（ファネル）のような形をしています。ポリペプチド鎖はこのファ

ネルの壁を滑り落ちていきます。どの方向に転がっても最小点に向かって下降し、自由エネルギーが低下するにつれてアクセス可能なコンホメーション空間は狭まっていきます。階段の段差を一段ずつ降りるのではなく、鎖は直接底部へと滑り落ちます。配列が構造をコードしているという命題は、配列が折りたたみファネルの形状を作り出していると言い換えることができます。

計算科学におけるブレークスルーは、物理シミュレーションではなく、進化の歴史からもたらされました。今日利用可能なタンパク質配列は、数百万年にわたる進化の選択の結果を表しています。ヒトのヘモグロビンと魚類のヘモグロビンは、アミノ酸の位置の約半分が異なっていますが、折りたたまれた3次元構造はほぼ同一です。機能的な形状が遥かな時間を超えて維持された一方で、個々の文字（アミノ酸）は変異しました。この事実から極めて重要な手がかりが浮かび上がります。それは、3次元空間で物理的に接触している残基対は、連動して変異するという点です。

地下鉄の座席ベンチが分かりやすい例えになります。2人の乗客が隣り合って座っているとき、片方が膝の上に大きなリュックサックを置くと、隣の乗客はスペースを空けるために肩をすぼめます。このような連動した調整が起こるのは、まさに彼らが隣同士に座っているからです。タンパク質も同じ論理に従います。残基15がかさ高いアミノ酸に変異すると、3次元空間でそれに隣接する残基80は、立体衝突を避けるためにより小さな残基へと変異します。逆に、3次元空間で互いに離れている残基同士は、進化的な圧力を及ぼし合いません。このようなペアでの変異パターンは「共進化（Co-evolution）」と呼ばれます。

深層学習（DL）アーキテクチャは、こうした進化のシグナルを大規模に読み取ります。UniRefやBFDといった膨大なデータベースから、標的タンパク質に関連する数万件の相同配列を取得します。これらの配列を垂直に積み重ねることで、多重配列アラインメント（MSA）が得られます。これは、同じ文献の数万点に及ぶ歴史的版を、大きな閲覧室の机いっぱいに並べる作業に似ています。ある列はすべての版で不変のまま保たれます。また、ある列のペアは版をまたいで同期して変化します。

複雑なのは、間接的なシグナルの伝播です。残基AがBに接触し、BがCに接触している場合、AとCは直接触れ合っていなくても統計的な相関を示してしまいます。2011年、MorcosとMarksが率いる研究チームは、こうした推移的な相関を除去するために直接結合解析（DCA）を導入しました[3][4]。統計物理学の最大エントロピー定式化を取り入れたこのアプローチは、直接的な空間結合と間接的な統計的ノイズを分離します。Marksらは、実験による結晶構造を参照することなく、配列データのみを用いて15のタンパク質ファミリーの天然の折りたたみ構造を正確に予測しました。二乗平均平方根偏差はおおむね3〜5オングストロームの範囲に収まりました。原

子の直径がおよそ1〜2オングストロームであることを考慮すると、これは主要なトポロジー構造を捉えたといえます。

ニューラルネットワークは、これらの予測を確率的な距離分布へと変換します。2つの残基間の単一の決定論的な距離を出力するのではなく、2〜22オングストロームの空間的区間を64個の離散ビンに分割し、各ビンに確率を割り当てます。この表現形式がディストグラム（Distogram）です。モデルは「距離は8オングストロームである」と断定するのではなく、「距離が7〜9オングストロームの間にある確率が高い」と出力します。2020年にDeepMindが発表した第1世代のAlphaFoldは、このディストグラムを物理的なエネルギーポテンシャルに組み込み、勾配降下法によって3次元座標を復元しました[5]。

このアプローチには本質的なトレードオフが存在します。共進化シグナルを抽出するには、十分な深さを持つ相同配列アライメントが必要です。読書机の上にテキストの版が3冊しか置かれていなければ、相関する列を特定することは不可能です。MSAに含まれる相同配列の数が数百本を下回ると、予測精度は急激に低下します。ウイルスのオーファン遺伝子や抗体の超可変相補性決定領域は、このようなデータが乏しいカテゴリーに該当します。自然界に進化的類縁体を持たないde novo設計の人工タンパク質も、同様の壁に直面します。

この限界が、代替となる新たな道を生み出しました。タンパク質言語モデルは、MSAの構築を完全に省略します。アライメントされていない数億もの配列を自然言語のテキストのように読み込むことで、アミノ酸の並びの文法を直接内部化するのです。これは、形式的な文法規則を暗記することなく、多読によって言語的直感を身につける過程に似ています。MSAの検索ステップを排除することで、推論は数十倍も高速化されます。高度に保存されたタンパク質においてアライメント依存型システムよりもわずかに精度が劣るものの、進化的類縁体を持たないオーファン配列を評価する際には言語モデルが優位に立ちます[8]。

予測が外れる可能性もあるため、現代のアーキテクチャは自己評価による信頼度指標を出力します。Predicted Local Distance Difference Test（pLDDT）は、残基ごとに0から100までの信頼度スコアを割り当てます。90を超えるスコアは側鎖のロータマーに至るまで信頼度が高いことを示し、70から90のスコアは信頼できる主鎖トポロジーを示します。50未満のスコアは、多くの場合、天然変性領域を表しています。研究者たちはこれらの信頼度スコアを手がかりに、構造モデルを色分けによって読み解きます。すなわち、濃い青色のドメインを信頼する一方で、オレンジ色のループについては物理的な実験を通じて検証を行うのです。

この段階で数理物理学が登場します。折りたたまれたタンパク質を空間内で90度回転させても、机の上のコーヒーマグを回転させてもマグカップのままであるのと同様に、その物理的な実体は

変わりません。すべての3次元回転と並進の群は、特殊ユークリッド群SE(3)と呼ばれます。妥当な予測モデルは、この対称性を尊重しなければなりません。原子間距離や結合エネルギーなどのスカラー量は、空間的な回転に対しても変化しない必要があります。この性質を不変性（Invariance）と呼びます。逆に、3次元座標の出力を生成する際、入力系を回転させたなら出力もまったく同じように回転して生成されなければなりません。この性質が同変性（Equivariance）です。

学習中に各タンパク質をランダムに回転させた数千ものパターンを与えることで、純粋な力任せのデータ拡張を通じてこれらの対称性を学習させようと試みることも可能です。しかし、これは計算資源の浪費になります。原理的な解決策は、幾何学的不変量をニューラルネットワークのアーキテクチャに直接埋め込むことです。Invariant Point Attention（IPA）は、各残基の位置に局所的な基準座標系を設定し、厳密にその局所座標系内で空間的距離を測定します。Equivariant Graph Neural Networks（EGNN）は、相互作用する原子ペアを結ぶ方向ベクトルに沿ってのみ座標の更新を制限します[6]。経験的に回転を学習する必要を省くことで、学習データを大幅に節約できると同時に、予測された構造における非物理的な幾何学的歪みを防ぐことができます。

下流の生物学的タスクにおいて、3次元構造表現が常に1次元配列表現より優れていると考えるのは直感的に思えるかもしれません。しかし、2026年6月の評価研究によって、それが普遍的な真実ではないことが明らかになりました [7]。この研究では、3つの生物学的ベンチマークにわたり、AlphaFold2の内部表現と配列のみを扱う言語モデルESM-2の内部表現を比較評価しました。結合親和性の予測では、AlphaFold2の0.307に対してESM-2が0.449の相関を達成して上回りました。コンホメーションの柔軟性予測においてもESM-2がリードし、0.764に対して0.824のAUROCを記録しました。AlphaFold2がESM-2を上回った唯一のベンチマークはアロステリック部位の特定であり、AUROCで0.548対0.485とわずかな差にとどまりました。また同研究は、単純な残基レベルのデータ分割を行うと、ベンチマーク性能が27%から39%も過大評価される点も浮き彫りにしました。

アンフィンセンは正しかったのです。すなわち、配列が構造を決定します。しかし、静的な3次元構造を計算するだけでは、すべての生物学的な問いが解決するわけではありません。タンパク質がどれほど強く結合するのか、そしてどれほど動的に揺らぐのかは、多くの場合、1次配列そのものの中により明確に記録されています。構造予測は最終目的地ではなく、極めて重要な分岐点なのです。

参考文献

[1] Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. Science, 181(4096), 223-230. https://doi.org/10.1126/science.181.4096.223

[2] Levinthal, C. (1969). How to fold graciously. Mossbauer Spectroscopy in Biological Systems: Proceedings of a Meeting held at Allerton House, 22-24.

[3] Marks, D. S., Colwell, L. J., Sheridan, R., Hopf, T. A., Pagnani, A., Zecchina, R., & Sander, C. (2011). Protein 3D structure computed from evolutionary sequence variation. PLoS ONE, 6(12), e28766. https://doi.org/10.1371/journal.pone.0028766

[4] Morcos, F., Pagnani, A., Lunt, B., Bertolino, A., Marks, D. S., Sander, C., Zecchina, R., Onuchic, J. N., Hwa, T., & Weigt, M. (2011). Direct-coupling analysis of residue coevolution captures native contacts across many protein families. PNAS, 108(49), E1293-E1301. https://doi.org/10.1073/pnas.1111471108

[5] Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., et al. (2020). Improved protein structure prediction using potentials from deep learning. Nature, 577(7792), 706-710. https://doi.org/10.1038/s41586-019-1923-7

[6] Satorras, V. G., Hoogeboom, E., & Welling, M. (2021). E(n) Equivariant Graph Neural Networks. arXiv preprint arXiv:2102.09844. https://arxiv.org/abs/2102.09844

[7] Chauhan, K. (2026). When Does Structure Help? The Information Bonus of AlphaFold2 Representations over Protein Language Models. arXiv preprint arXiv:2606.04228. https://arxiv.org/abs/2606.04228

[8] Zhang, Y., Deng, N., Song, X., Bi, Z., Wang, T., Yao, Z., et al. (2025). Advanced Deep Learning Methods for Protein Structure Prediction and Design. arXiv preprint arXiv:2503.13522. https://arxiv.org/abs/2503.13522

3.2 AlphaFold、AlphaFold2、AlphaFold3のアーキテクチャの進化と高分子複合体相互作用モデリング

2020年11月30日、オンラインで開催されたCASP14コンファレンスにおいて、数百人の構造生物学者がコンピュータのモニターの前に座っていました。2年ごとに開催されるこのコンペティションは、実験によって構造が解明されているものの未公開であるブラインドターゲットを用いて、構造予測アルゴリズムを評価するものです。採点ダッシュボードにAlphaFold2の指標が表示された瞬間、会議の場は静まり返りました。30年近くにわたりCASPを共同創設し率いてきたジョン・モルト（John Moult）は、根本的な意味において、この問題は解決されたと述べました。

定量的なスコアも彼の言葉を裏付けていました。AlphaFold2は、すべての評価ターゲットにおいて、Global Distance Test（GDT_TS）スコアの中央値92.4を達成しました。100点満点の尺度において、90を超えるスコアは、実験による結晶構造と事実上区別がつかないものと見なされます。主鎖原子の平均二乗平均平方根誤差は0.96オングストローム［1］であり、これは炭素原子1個の半径よりも小さい誤差でした。2位のチームが達成したスコアの中央値は70点台前半にとどまりました。

その2年前、そのアーキテクチャはまったく異なるものでした。2018年のCASP13に参加した第1世代のAlphaFoldは、2段階のシステムとして動作していました。上流の畳み込みニューラルネットワークが残基間の距離分布図（ディストグラム）を予測し、そのマップに基づいてRosetta由来の下流の物理エネルギー最小化モジュールが3次元座標を構築していました［2］。その性能は強力なものでしたが、2つのモジュールの間には構造的な境界が存在していました。下流の物理エンジンで生じた誤差を逆伝播させて、上流の畳み込み層を調整することができなかったのです。

AlphaFold2はその障壁を取り払い、配列入力を直接3次元座標へと変換する、統合されたエンドツーエンドで微分可能なニューラルネットワークへとパイプラインを再構築しました。その中核に位置するのがEvoformerブロックです。Evoformerは、生物種間の進化的変異を追跡するMSA表現と、残基iとjの間のペアごとの空間的関係をマッピングするペア表現という、2つの独立した情報台帳を同時に保持します。連続する48個のブロックを通じて、これら2つの台帳は互いに絶え間なく更新し合います。新たに抽出された配列の相関関係がペアの空間的特徴を洗練させ、構造化されたペアの幾何構造が配列の解釈を導きます。

ペア表現に適用される数学的制約の1つに「三角不等式」があります。地図上の3つの都市を考えてみてください。ソウルから大田（テジョン）までの距離が140キロメートル、大田から大邱

（テグ）までが150キロメートルである場合、ソウルから大邱までの距離が500キロメートルになることはあり得ません。残基AがBに近接し、BがCに近接しているなら、AとCの間の距離も幾何学的な制約に従う必要があります。Evoformerは、すべての残基の三つ組（トリプレット）に対して三角乗算更新（triangular multiplicative update）と三角自己注意（triangular self-attention）を適用し、矛盾する空間配置を排除します。これにより、明示的な物理エネルギー方程式を解くことなく、物理的に整合のとれた空間マップが得られます。

精緻化されたペア表現はStructure Moduleへと渡されます。このモジュールでは、各残基が専用の位置ベクトルと向きベクトルを持ち、3次元空間に浮かぶ三角形の剛体として概念化されます。Invariant Point Attentionがこれら剛体の相対的な位置と向きを調整し、主鎖および側鎖のすべてのロータマー角を同時に決定します。こうして得られた座標はアーキテクチャ内で3回再循環（リサイクル）されますが、これは著者が原稿を執筆し、見直し、推敲を重ねる過程に似ています。

このモデルが社会に与えた広範な影響は、最初の論文そのものよりも、その結果として公開されたデータベースによるものが大きいです。DeepMindとEMBL-EBIはAlphaFold Protein Structure Databaseを設立し、予測された構造を誰もが自由に利用できるようにしました。2024年までに、このリポジトリには2億1400万件以上のタンパク質構造が登録され、2021年の最初の公開から500倍の規模へと拡大しました。すべての国連加盟国から240万人を超える研究者がこのプラットフォームにアクセスしました [3]。かつては専門の研究室で1つのタンパク質構造を解明するのに何年も費やしていましたが、今日では大学院生が昼食の合間に遺伝子識別子を検索し、完成した座標ファイルをダウンロードしています。

続いて過渡期となる開発ステップがありました。AlphaFold2は孤立した単一の鎖を効果的に折りたたむことができたものの、2つの異なる鎖がどのように集合するかをモデル化することには苦戦していました。研究者たちは、個別のタンパク質配列を人工的なポリグリシン-セリンリンカーで連結し、モデルに単一の鎖として認識させる工夫を凝らしました。2022年に導入されたAlphaFold-Multimerは、複合体の集合に真正面から取り組みました [12]。明示的な鎖の境界定義を与え、生物種の起源ごとに相同なMSAをペアリングすることにより、「生体内で2つのタンパク質が共進化した場合、相互作用する界面の残基には相関した変異が見られる」という原理を活かしました。その時点で、研究の焦点は単一分子から分子複合体へと移行しました。

2024年、AlphaFold3はモデリングの領域を拡張し、生体分子複合体全体を全原子分解能でカバーできるようにしました [4]。内部アーキテクチャも刷新されました。計算負荷の高かったEvoformerが簡素化されてPairformerとなり、MSA処理ブロックは4つに削減されました。剛体フレームに基づくStructure Moduleに代わり、生成的なDiffusion Moduleが導入されまし

た。拡散モジュールは、ランダムなガウス座標ノイズの塊を受け取り、連続するデノイジングステップを通じてノイズを反復的に除去することで、その下にある鮮明な原子座標を浮かび上がらせます。これは、朝霧が晴れて都市のビルの鮮明な輪郭が現れる様子に似ています。

このアーキテクチャにより、AlphaFold3はタンパク質-タンパク質複合体に加え、DNA二重らせん、RNA鎖、リン酸化などの翻訳後修飾、低分子化合物の創薬候補化合物、亜鉛や鉄などの構造金属イオンを、単一の統一された座標空間内でモデル化できるようになりました。タンパク質とリガンドの結合ポーズの精度を評価するPoseBustersベンチマークにおいて、AlphaFold3は76.4%の成功率を達成し、50%台にとどまっていた従来のドッキングプログラムを凌駕しました[4]。

性能の向上は、すべてのカテゴリーで一様だったわけではありませんでした。2025年に発表された独立したベンチマークは、詳細な見解を提示しました[5]。単量体タンパク質において、AlphaFold3はTM-scoreで0.817を達成し、AlphaFold2の0.812と比較しても統計的にはわずかな差にとどまりました。LDDT指標においても、AlphaFold3の優位はわずかなものでした（0.787に対して0.792）。一方で多量体については、DockQスコアが0.568に対して0.589となり、相互作用界面の78.2%を正確に同定しました。また、核酸複合体や抗原抗体複合体において明確な改善を達成しました。対象領域の拡大によって単量体の精度が劇的に向上したわけではありませんが、対応可能な生体分子の実体の多様性を拡張したこと自体が、重要な節目となりました。

ブラインド方式によるCASPの評価でも、これらの結果がそのまま反映されました。CASP16では、125のエキスパートチームと36の自動サーバーによって、3つのステージにわたり85の標的が評価されました[6]。評価者らは、単一ドメインのタンパク質標的においてフォールディングの誤りが一つもなかったことを確認しました。上位の成績を収めたグループは、主に独自のMSAキュレーションパイプラインを組み合わせたAlphaFold2およびAlphaFold3を使用していました。対照的に、多量体複合体の標的では30%以上が依然として課題として残り、なかでも抗体・抗原のエピトープ予測が最も高い難度を示しました。

2026年の主要な潮流は、オープンアクセス化と推論の高速化です。AlphaFold3をめぐる商用アクセスの制限が、オープンソースの代替モデルの急速な開発を促しました。MIT出身の研究者らによって開発されたBoltzファミリーは、モデルの重み、アーキテクチャ、学習コードを、寛容なMITライセンスの下で公開しました。Boltz-2は3次元座標とともに結合親和性を予測し、自由エネルギー摂動（FEP）シミュレーションと同等の精度を1,000倍以上の速度で実現しています。OpenFoldコンソーシアムは2026年3月13日にOpenFold3のメジャーリリースを発表し、完

全な学習データセットを公開するとともに、抗体・抗原のモデリングを重視しつつAlphaFold3に匹敵する性能を確立しました[7]。

速度とスケールに関する研究も急速に進展しました。2026年5月に発表されたDCFoldは、反復的な拡散ノイズ除去を単一のフォワードパスに置き換えることで、AlphaFold3に匹敵する構造精度を維持しながら推論速度を15倍に加速させました[8]。2026年3月にはFold-CPがコンテキスト並列処理を用いてメモリの限界を突破しました。64基のNVIDIA B300 GPUを活用することで、30,000残基を超える巨大な複合体をモデル化し、CORUMタンパク質複合体データベースの90%以上をカバーしました[9]。これにより、予測可能な上限を数千残基にとどめていた従来の制約は取り払われました。

下流の設計応用は、現代の構造予測の成熟ぶりを示しています。Chai Discoveryは、結合することが知られている抗体が存在しない高難度の52の標的を評価し、標的ごとに20個以下の設計候補化合物を合成したところ、物理的なアッセイにおいて16%が結合に成功しました。また、ミニタンパク質バインダーにおいては、5つの標的全体で68%の成功率を達成しました[10]。数万個の候補分子を必要としていた従来のスクリーニングパイプラインと比べると、現在では標的の検証作業がわずか1枚のマルチウェルプレート内に収まります。

次のベンチマークはすでに形作られています。CASP17は2026年の競技会に向けた標的の選定を進めており、9月1日まで実験による座標データの提出を受け付けています。その注目の焦点はコンホメーションアンサンブルです。単一の静的なスナップショットではなく、タンパク質が溶液中でとる動的な構造状態の完全な分布を出力することがモデルに求められています[11]。静的な構造予測の時代は、今まさに終わりを告げようとしています。

参考文献

[1] Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583-589. https://doi.org/10.1038/s41586-021-03819-2

[2] Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., et al. (2020). Improved protein structure prediction using potentials from deep learning. Nature, 577(7792), 706-710. https://doi.org/10.1038/s41586-019-1923-7

[3] Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., et al. (2024). AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. Nucleic Acids Research, 52(D1), D368-D375. https://doi.org/10.1093/nar/gkad1011

[4] Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature, 630(8016), 493-500. https://doi.org/10.1038/s41586-024-07487-w

[5] Peng, C., Ni, W., Liu, Q., Hu, G., & Zheng, W. (2025). A comprehensive benchmarking of AlphaFold3 for predicting biomacromolecules and their interactions. Briefings in Bioinformatics, 26(6), bbaf616. https://doi.org/10.1093/bib/bbaf616

[6] Yuan, R., Zhang, J., Kryshtafovych, A., Schaeffer, R. D., Zhou, J., Cong, Q., & Grishin, N. V. (2025). CASP16 protein monomer structure prediction assessment. bioRxiv. https://doi.org/10.1101/2025.05.29.656942

[7] OpenFold Consortium. (March 13, 2026). OpenFold Consortium Announces Major OpenFold3 Update and Public Release of Training Data for Reproducible Biomolecular AI. https://www.businesswire.com/news/home/20260313170622/en/

[8] Zhang, Z., Feng, Y., Song, Y., Qiu, K., Zhou, H., & Ma, W.-Y. (2026). DCFold: Efficient Protein Structure Generation with Single Forward Pass. arXiv preprint arXiv:2605.17899. https://arxiv.org/abs/2605.17899

[9] Lin, D., Chu, S., Iyer, V., Lee, Y., St John, J., Boyd, K., et al. (2026). Fold-CP: A Context Parallelism Framework for Biomolecular Modeling. arXiv preprint arXiv:2603.14806. https://arxiv.org/abs/2603.14806

[10] Chai Discovery Team. (2025). Zero-shot antibody design in a 24-well plate. bioRxiv. https://doi.org/10.1101/2025.07.05.663018

[11] Protein Structure Prediction Center. (2026). Call for Targets: CASP17. https://predictioncenter.org/casp17/callForTargets.cgi

[12] Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., et al. (2022). Protein complex prediction with AlphaFold-Multimer. bioRxiv. https://doi.org/10.1101/2021.10.04.463034

3.3 実験的構造プロテオミクスデータと機械学習（ML）モデルの相互誘導

タンパク質構造データバンク（PDB）の座標ファイルを開くと、原子座標の長いリストが表示されます。すべての位置は、結晶格子構造として捉えられたり、電子顕微鏡下で急速凍結されたりして、ある一瞬の状態で静止しています。それはまるでアルバムの中の証明写真のようです。しかし、生体細胞内のタンパク質は静止した写真のように機能するわけではありません。グルコースがヘキソキナーゼに近づくと、その2つの構造ローブは二枚貝のようにパチンと閉じます。トランスポーターは、内向き状態と外向き状態の間でゲートを交互に切り替えます。シグナル伝達タンパク質は、結合相手と結合する際に主鎖全体を再構成します。

機械学習（ML）モデルは、こうした静止スナップショットのアルバムから学習したため、その構造的バイアスを受け継いでいます。α-シヌクレインやタウ、p53の転写活性化ドメインのようなタンパク質は、溶液中で絶えずコンホメーションを変化させます。それらの配列をAlphaFoldに入力すると、断片的で信頼度の低い予測が出力されます。モデルが正常に動作していないわけではありません。本質的に多様な形態をとる存在に対して、単一の決定論的な構造を求めてしまう点に誤りがあるのです。

2026年の『Nature Methods』誌に掲載されたコメンタリーは、この転換をこの分野における次の主要な目標として強調しました[1]。著者らは、構造生物学が単一のコンホメーションの予測から、動的なコンホメーションアンサンブルの平衡分布と存在比率の予測へと移行しなければならないと主張しています。そのためには正解データとなるベンチマークデータセットが必要ですが、現在それらには標準化された評価指標や分析ツールが不足しています。

初期の試みでは、モデルのサンプリングダイナミクスの内部に解決策が求められました。完全なアライメントをAlphaFold2に入力するのではなく、MSAをランダムにサブサンプリングすることで、ネットワークは同一のタンパク質に対して多様なコンホメーション状態を出力するよう誘導されます。系統樹の分岐ごとにアライメントをグループ化すると、特定の進化系統に特有のコンホメーションが生成されます。2024年に『Nature』誌に掲載された研究では、このサブサンプリングアプローチを用いてキナーゼやトランスポーターの活性型および不活性型の両方のコンホメーションを抽出しました[11]。単一の写真ではなく、小さなアルバムを生成したといえます。しかし、このような計算によるサンプリングでは、生体細胞内における各状態の生理学的な存在比率（パーセンテージ）を決定することはできません。平衡比率を確定するには、物理的な実験が必要です。

実験的構造プロテオミクスは、欠けていた制約条件を提供します。架橋質量分析（XL-MS）はその代表的な例です。研究者は、一定の長さのスペーサーアームを持つ化学架橋剤をタンパク質

溶液に添加します。広く使われているDSS架橋剤は、11.4オングストロームのスペーサーアームを持っています。このアームが2つのリシン残基の側鎖を共有結合で結びつけます。リシンの側鎖には柔軟性があるため、架橋された残基同士のα炭素間距離はおよそ30オングストローム以下になります。架橋されたタンパク質をプロテアーゼで消化し、その断片を質量分析計に導入することで、溶液中でどの残基のペアが物理的に近接していたかが明らかになります[2][3]。これは、タンパク質を微小な糸で縛り上げ、残った結び目同士の距離を測定するようなものです。

水素・重水素交換質量分析（HDX-MS）は、相補的な動態情報を提供します。タンパク質を重水にさらすと、主鎖に沿ったアミド水素が重水素と交換されます。柔軟で溶媒がアクセスしやすい表面のループ領域は数秒以内に交換されますが、密にパッキングされた疎水性コア領域は何時間も交換されないまま残ります。時間経過に伴う質量の増加を追跡することで、どの領域が開いており、どの領域が溶媒から遮蔽されているかをマッピングできます [4]。HDX-MSは原子座標を提供するものではありませんが、局所的な柔軟性と溶媒アクセシビリティを直接マッピングします。

どちらの実験手法も、単独では完全ではありません。XL-MSは数十箇所の距離制約をもたらすものの、残りの原子位置は未特定のまま残されます。HDX-MSが提供するのも座標セットではなく、領域ごとの保護値にとどまります。それとは対照的に、深層学習（DL）モデルは完全な原子座標を出力しますが、複数の選択肢の中から生理学的に正しいコンホメーションを選択することに苦戦します。こうした背景から、実験による制約を生成モデルの事前分布に直接融合させるハイブリッド手法の開発が進められました。

2026年5月に公開されたAIMS-Foldは、この統合的フレームワークを象徴する例です [5]。同手法は、XL-MSの距離境界とHDX-MSの溶媒アクセシビリティプロファイルを、拡散モデルに基づく共折りたたみ（co-folding）モデルへと直接埋め込みます。実験データを事後的なフィルタリングに用いるのではなく、実験測定値そのものが座標生成の軌道を誘導します。研究者らは、双方のデータセットを統合することで相乗的な精度向上が得られたと報告しています。その性能は、本来は自然に結合しないタンパク質間に非天然の相互作用を強制するPROTACsや分子グルーのような、誘導近接（induced proximity）標的において特に顕著でした。人工的に設計されたこれらの複合体は学習データにおける進化的な情報が極めて乏しいため、AIMS-FoldはBoltz-2のような誘導のないモデルよりも、誘導近接複合体において高い精度を達成しました。

同様の統合は、クライオ電子顕微鏡（cryo-EM）の領域も変革しています。2026年8月12日時点で、Electron Microscopy Data Bank（EMDB）には60,895件の3次元密度マップが登録されていました [6]。未処理のcryo-EM密度マップは霧がかかった写真に似ており、分子全体の輪郭は判別できるものの、個々の原子位置は曖昧なままです。構造生物学者は手作業や最適化エンジン

を用いてこの密度データに原子モデルを当てはめますが、この工程は歴史的に、PhenixやRosettaなどのソフトウェアを通じた多大な手動リファインメントを必要としていました。

2026年2月に発表されたCryoNet.Refineは、このリファインメントパイプラインを1ステップの拡散モデルへと変換しました [7]。立体化学的な幾何学的制約とともに、密度マップとの相関を損失関数に組み込むことで、マップとモデルの相関性と立体化学的品質の双方を向上させました。ICLR 2026に採択されたこのツールは、公開ウェブサーバーとして利用可能です。同じ研究室から発表されたCryoLVMは、自己教師あり学習によって未処理のcryo-EM密度マップ上で大規模ビジョンモデルを事前学習し、マップの明瞭度の向上、解像度の引き上げ、配向バイアスに起因する欠損領域の補完を実現しています [8]。CryoDynaはコンホメーションの不均一性に直接取り組み、単一の実験データセットから複数の分離した機能状態を抽出します [9]。

電子顕微鏡は、単離・精製されたタンパク質にとどまらない領域へと進化を遂げました。クライオ電子トモグラフィー（cryo-ET）は、精製を行わずに無傷の細胞切片を急速凍結し、傾斜連続撮影（チルトシリーズ）の投影像を取得して3次元ボリュームを再構成します。こうして得られたトモグラムは、本来の細胞環境におけるリボソームやプロテアソームを捉えます。課題は、トモグラムのノイズが多く、コントラストが低い点です。深層学習（DL）モデルは、予測された3次元構造をテンプレートとして使用し、トモグラムを走査して合致する密度シグネチャを特定することで粒子ピッキングを支援します。これにより、無傷の細胞ボリュームから生体高分子マシンの空間ディレクトリが抽出されます。

情報の流れは逆方向にも進みます。X線結晶構造解析は、長年にわたり位相問題に悩まされてきました。回折パターンは反射強度を記録するものの、位相角の情報が失われてしまうためです。従来、位相を決定するには相同な結晶構造を用いた分子置換法が必要であり、関連する構造が存在しない場合、プロジェクトは何年も停滞しました。今日では、信頼度の高いAI予測モデルが初期探索モデルとなり、座標誤差が1オングストローム未満であれば位相を解決できます。クライオ電子顕微鏡（cryo-EM）においても、予測モデルを低分解能マップへドッキングさせることで、何日もかかっていた手作業による主鎖のトレース作業が数時間に短縮されます [10]。

異なる実験手法を統合する統合的構造モデリングが主流となっています。X線小角散乱（SAXS）は分子全体のエンベロープ（外形）と回転半径の測定値を提供します。核磁気共鳴（NMR）は、柔軟なドメインに対する原子間距離や二面角の拘束条件をもたらします。XL-MSは界面の接触部位をマッピングし、クライオ電子顕微鏡（cryo-EM）は複合体全体の構造の輪郭を描き出します。単一の手法だけで完全な構造を解明できるものはありません。ニューラルネットワークは、統一された座標系の中でこれら異種の拘束条件を重み付けすることに優れており、すべての

実験的拘束条件を同時に満たすコンホメーションのアンサンブルを特定するという最適化問題として、構造生物学を再定義しています。

この関係性は、計算が物理的な実験に取って代わるものと捉えるべきではありません。実験はモデルに経験的な拠り所（アンカー）を提供し、モデルは実験ではまばらにしか得られない原子座標を補完し、人間の研究者がその両者の間で生物学的な妥当性を評価します。最新のデータが示すように、実験データによって導かれたモデルは誘導のないアーキテクチャを一貫して凌駕し、AIが生成した草案を土台にすることで、実験による精密化は急速に進展します。

明確にしておくべき原則が一つあります。それは、予測モデルが実験構造を代替するものではないということです。ニューラルネットワークは、自らの誤りを自律的に診断することはできません。とりわけ誘導近接（induced proximity）標的のようにデータが乏しい領域では、誤ったコンホメーションに対してモデルが高い信頼度スコアを付与する事例が現在も報告されています。質量分析による30箇所の架橋距離拘束条件があれば、そうしたリスクを排除できます。一見妥当なモデルの配置が物理的な実験室での測定結果と矛盾するとき、研究者は再計算を行うための具体的な理由を得ます。ウェットラボのアッセイとニューラルネットワークは対立関係にあるのではなく、同じ荷車を前後から押す2人のパートナーなのです。前に立つ者が進むべき前方の道を見通し、後ろに立つ者がどれだけの抵抗が残されているかを把握します。

参考文献

[1] Wankowicz, S. A., & Bonomi, M. (2026). From possibility to precision in macromolecular ensemble prediction. Nature Methods, 23, 1100-1108. https://doi.org/10.1038/s41592-026-03084-z

[2] Leitner, A., Faini, M., Stengel, F., & Aebersold, R. (2016). Crosslinking and Mass Spectrometry: An Integrated Approach to Understand the Structure and Function of Molecular Machines. Trends in Biochemical Sciences, 41(1), 20-32. https://doi.org/10.1016/j.tibs.2015.10.008

[3] Rappsilber, J. (2011). The beginning of a beautiful friendship: Cross-linking/mass spectrometry and modelling of proteins and multi-protein complexes. Journal of Structural Biology, 173(3), 530-540. https://doi.org/10.1016/j.jsb.2010.10.014

[4] Engen, J. R. (2009). Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS. Analytical Chemistry, 81(19), 7870-7875. https://doi.org/10.1021/ac901154s

[5] Shtrikman, A., Simchi, N., Ran Shchory, M., Brodsky, S., Seger, E., & Pevzner, K. (2026). Co-folding model guided by structural proteomics. arXiv preprint arXiv:2605.26192. https://arxiv.org/abs/2605.26192

[6] EMBL-EBI. (August 12, 2026). Electron Microscopy Data Bank (EMDB) Statistics. https://www.ebi.ac.uk/emdb/

[7] Huang, F., Yu, X., Xu, K., & Zhang, Q. C. (2026). CryoNet.Refine: A One-step Diffusion Model for Rapid Refinement of Structural Models with Cryo-EM Density Map Restraints. arXiv preprint arXiv:2602.22263. https://arxiv.org/abs/2602.22263

[8] Fu, W., Shu, K., Xu, K., & Zhang, Q. C. (2026). CryoLVM: Self-supervised Learning from Cryo-EM Density Maps with Large Vision Models. arXiv preprint arXiv:2602.02620. https://arxiv.org/abs/2602.02620

[9] Zhang, C., et al. (2025). CryoDyna: Multiscale end-to-end modeling of cryo-EM macromolecule dynamics with physics-aware neural network. arXiv preprint arXiv:2510.16510. https://arxiv.org/abs/2510.16510

[10] Terashi, G., Wang, X., Prasad, P. R., & Kihara, D. (2023). Deep learning methods for protein structure modeling in cryo-EM density maps. Current Opinion in Structural Biology, 79, 102534. https://doi.org/10.1016/j.sbi.2023.102534

[11] Wayment-Steele, H. K., Ojoawo, A., Otten, R., Apitz, J. M., Pitsawong, W., Hömberger, M., Ovchinnikov, S., Colwell, L., & Kern, D. (2024). Predicting multiple conformations via sequence clustering and AlphaFold2. Nature, 625(7996), 832-839. https://doi.org/10.1038/s41586-023-06832-9

第4章 生成型生物学モデルを用いたタンパク質と抗体のde novo設計

4.1 拡散モデルとフローマッチングによる物理的タンパク質構造の生成

モニタースクリーン上に点が散らばっています。3次元座標空間の中に、目に見える規則性もなく数百個の点が分布しています。研究者が実行ボタンをクリックします。点は少しずつ位置を変えていきます。およそ50ステップの計算を経て、点はヘリックス（らせん構造）へと整列し、その両脇をプリーツシート（ひだ状構造）が挟み込み、界面にはノギスのような形をした間隙が残されます。これまで地球上に存在しなかったタンパク質が誕生します。プロセス全体にかかる時間はわずか数分です。

天然に存在するタンパク質は、数十億年にわたる生物学的な試行錯誤の賜物です。自然選択が満たすべき試験は、生存と繁殖という2点だけでした。特定のがんバイオマーカーを分離して結合させたり、市販のプラスチックを分解したりするといった要求は、進化という試験の出題範囲から完全に外れています。その結果、20世紀のタンパク質工学は、主に指向性進化（Directed Evolution）を通じて天然タンパク質を改変することに注力していました。これは、何十万ものランダムなアミノ酸変異体を作製し、機能を持つ変異体をスクリーニングするという手法です。しかしこれは、複数階建ての巨大な図書館全体を探索するのに、最初の本棚だけをひっくり返して探しているようなものです。わずか100残基のタンパク質であっても、その組み合わせ空間は20の100乗通りもの異なる配列に及び、これは観測可能な宇宙に存在する全原子の数すら上回ります。

生成型人工知能は、拡散モデルを皮切りに、この探索空間に対して異なるアプローチを取りました。この手法は、定型のないノイズの中から徐々に構造を浮かび上がらせていきます。学習時には、天然の結晶構造にガウスノイズを段階的に加えていき、完全なランダム状態に至らせます。生成時には、ニューラルネットワークがこの軌跡を逆再生します。秩序のない原子の雲から出発し、段階的にノイズを除去（デノイズ）することで、整然とした3次元の折り畳み構造を現出させるのです。標的タンパク質の座標は「条件付け（Conditioning）」として与えられます。これは生成前にあらかじめ提示される空間的な青写真です。標的となる結合ポケットの原子レベルの輪郭が与えられると、モデルはそのポケットに正確に収まるように設計された、相補的な主鎖骨格を生成します。主鎖とは、タンパク質の構造的背骨を形成する、共有結合した連続的なN-Cα-C-O鎖のことを指します。

この生成パラダイムが科学界で広く注目を集めたのは、2023年にワシントン大学のデヴィッド・ベイカー研究室が『Nature』誌にRFdiffusionを発表したときでした。彼らは、構造予測ネットワークであるRoseTTAFoldの内部表現を生成拡散フレームワークへと適応させることで、すでに構造の折り畳み原理を学習していたネットワークを活用しました。それらのルールを逆転

させて3次元座標を生成することにより、ウェットラボの実験において綺麗に折り畳まれる人工タンパク質を生み出すことに成功したのです[1]。その後3年のうちに、数多くの派生アーキテクチャが登場しました。

これらの計算は、厳密な物理的対称性に従わなければなりません。3次元空間でタンパク質を回転させたり平行移動させたりしても、その生化学的な同一性が変わることはありません。モデルがこの対称性を考慮できなければ、同一の分子を別個の実体として扱い、計算資源を浪費してしまいます。そのため、構造生成モデルには「同変性（Equivariance）」が組み込まれています。これは、入力座標を回転させると、出力も全く同様に回転した結果が得られるという性質です。路線図を上下逆さまにしても、駅の接続関係や順序が変わらないのと似ています。また、基本的な化学的制約も満たす必要があります。アミノ酸を連結する平面的なペプチド結合は1.33オングストローム付近に厳密に制約されており、結合角も定められた物理的許容範囲から外れることはできません。こうした幾何学的制約に違反する構造は、画面上ではもっともらしく見えても、溶液中では正しく折りたたまれません。標準的なアルファヘリックスやベータシートが自発的に現れるのは、モデルがこれら立体化学的不変量を厳格に維持しているからです。

代表的な設計タスクの一つが「モチーフスキャフォールディング（Motif Scaffolding）」です。これは、短く機能的な構造セグメントを空間上に固定し、その正確な幾何学的構造を維持できるように、その周囲に全く新しいタンパク質スキャフォールド（足場）を生成する手法です。中和抗体に認識される特定のエピトープ残基のみを維持してワクチン免疫原を設計することは、このタスクの代表例です。機能モチーフが「釘」だとすれば、生成されたスキャフォールドはその釘を所定の位置にしっかりと固定する「フレーム」の役割を果たします。

フローマッチング（Flow Matching）は、その後に登場したアルゴリズムの進化形です。拡散モデルが目標とする分布に向かって確率的な経路をたどるのに対し、フローマッチングは確率微分方程式ではなく常微分方程式を用い、初期ノイズと最終的なコンホメーションを結ぶ直線的な確率軌跡を学習します。経路がより直線的であれば、必要な数値積分のステップ数が少なくて済み、推論時の計算量を削減できます。2026年3月に発表されたPI-Mambaは、フローマッチングと状態空間モデルを統合し、単一の24 GB GPU上で2,000残基を超える巨大タンパク質を生成しました。局所的な結合幾何構造の違反率0%を達成するとともに、自己整合性TM-score（scTM）で0.91を記録しています [3]。2026年6月には、ProHiFloが階層的フローマッチング戦略を採用しました。これは、微細な原子レベルの詳細を配置する前に大まかなトポロジーのスキャフォールドを構築する手法であり、酵素の触媒ポケット主鎖の設計成功率を、RFdiffusionの41.2%に対して58.9%まで引き上げました [4]。

固定された配列長を必要とする制約も解決されました。2026年7月に登場したGPFlowは、生成サンプリングの軌跡の中でタンパク質の長さを動的に決定し、16のモチーフスキャフォールディング・ベンチマーク中10項目で1位を獲得しました [5]。設計者はもはや、事前に任意の鎖長を指定しておく必要がなくなったのです。

3次元の主鎖を生成することは課題の半分にすぎません。その幾何構造へと物理的に折りたたまれるアミノ酸配列を特定する必要があります。この課題は「逆フォールディング（Inverse Folding）」と呼ばれ、目的の3次元立体構造を読み取り、それを生み出す一次配列コードを決定する作業です。2022年に導入されたProteinMPNNは、長らくこの工程の中核を担ってきました。既知の主鎖から天然配列を復元するタスクにおいて、物理ベースのRosettaが32.9%であったのに対し、ProteinMPNNは天然配列との一致率52.4%を達成し [2]、しかもわずか数秒で処理を実行しました。2026年3月に発表されたRefoldは、局所的な構造フラグメントのマッチングとニューラル配列生成を組み合わせることでこれをさらに前進させ、局所的な精度と鎖全体のグローバルな一貫性を同時に捉えることに成功しました [6]。

2025年12月3日、ベイカー研究室がRFdiffusion3をオープンソースライセンスで公開したことで、重要な進展がもたらされました。RFdiffusion2が残基と原子の混合表現で動作していたのに対し、バージョン3は個々の原子を基本的な設計単位として扱い、設計者が指定した原子対の間に明示的な水素結合を設定できるようにしました。全原子分解能で動作するにもかかわらず、推論はバージョン2よりも10倍高速でした。計算設計された96種類の金属加水分解酵素を合成し、in vitroで評価したところ、5種類が触媒活性を示し、そのうちの1つの設計であるZETA_1は、従来の計算設計酵素と比べて桁違いに高い触媒効率を発揮しました。システインプロテアーゼにおいては、評価された190種類の設計のうち35種類が測定可能な酵素活性を示しました[7]。

これらの数値には客観的な評価が求められます。96件中5件のヒットは5%の成功率を表し、190件中35件は18%です。2025年8月に発表されたメタアナリシスでは、15種類の生体標的に対して実験的に特徴づけられた3,766種類のde novoバインダーを集約し、物理的な結合の全体的な成功率が11.6%であることを明らかにしました[8]。コンピュータが妥当な分子候補を大量に生成する一方で、物理的なウェットラボのアッセイで実際に結合するのはおよそ10個に1個です。研究助成金のプレゼンテーションでは80%という計算上の成功率が頻繁に引用されますが、それらの数値は通常、インシリコでニューラルネットワークが別のニューラルネットワークを採点した結果を反映しているにすぎません。ウェットラボのアッセイプレートが示す現実はそれとは異なります。

その結果、研究の焦点は生成量の最大化から実験での脱落の最小化へと移行しつつあります。2026年3月に発表されたProteina-Complexaは、生成設計と構造予測器に導かれた配列最適化

を統合しました。予測されたモノマーライブラリからドメイン間インターフェースを抽出することで、複合体構造に関する基盤モデルを事前学習するための合成データセットTeddymerを構築し、続いて推論時計算量を割り当てて候補群を絞り込みました[9]。2026年7月に公開されたChamaileonは、従来の単一標的・単一状態というパラダイムを超え、複数の標的およびコンホメーション状態に同時に作用するようバインダー生成を再構成するとともに、CROSS評価ベンチマークを確立しました[10]。表面のコンホメーションを変化させる動的なウイルス性タンパク質に結合させるには、こうした多状態対応能力が不可欠です。

評価プロトコルも並行して進化しています。以前は、予測された配列を同じタンパク質立体構造予測ネットワークに再入力し、出力が設計図と一致するかどうかを確認することで、計算機上で設計を検証していました。その限界は学習データの共有にありました。すなわち、生徒モデルが同じデータで学習した試験官によって採点されていたのです。2026年の最新ワークフローでは、独立したモデルファミリー間での交差検証を採用し、AlphaFold3、Boltz-1、Rosettaに対して候補を同時に評価しています。3,766種類のバインダーを評価したメタアナリシスでは、界面に焦点を当てた接触指標が、全体的な構造信頼度スコアよりもはるかに高い信頼性でウェットラボでの結合成功を予測できることが示されました[8]。

ランダムなノイズから分子の形状を生成する能力は、今や確立されています。残された課題は、それらの形状が物理的な溶液中で安定性と活性を維持できるかどうかを検証することです。その検証には、今なお実験台でピペットを手にする研究者の存在が欠かせません。

参考文献

[1] Watson, J. L., Juergens, D., Bennett, N. R., et al. (2023). De novo design of protein structure and function with RFdiffusion. Nature, 620(7976), 1089-1100. https://doi.org/10.1038/s41586-023-06415-8

[2] Dauparas, J., Anishchenko, I., Bennett, N., et al. (2022). Robust deep learning-based protein sequence design using ProteinMPNN. Science, 378(6615), 49-56. https://doi.org/10.1126/science.add2187

[3] Wu, T., & Zhu, L. (2026). PI-Mamba: Linear-Time Protein Backbone Generation via Spectrally Initialized Flow Matching. arXiv preprint arXiv:2603.26705. https://arxiv.org/abs/2603.26705

[4] Wang, C., Cleti, M., & Jano, P. (2026). ProHiFlo: Hierarchical Flow Matching with Functional Guidance for De Novo Protein Generation. arXiv preprint arXiv:2606.11243. https://arxiv.org/abs/2606.11243

[5] Cheng, C., Ni, Z., Qu, Y., et al. (2026). Variable-Length Generative Protein Design via Generalized Poisson Flow. arXiv preprint arXiv:2607.09039. https://arxiv.org/abs/2607.09039

[6] Zhu, Y., et al. (2026). Refold: Refining Protein Inverse Folding with Efficient Structural Matching and Fusion. arXiv preprint arXiv:2603.14350. https://arxiv.org/abs/2603.14350

[7] Genetic Engineering & Biotechnology News. (December 3, 2025). RFdiffusion3 Now Open Source, Designs DNA Binders and Advanced Enzymes. https://www.genengnews.com/topics/artificial-intelligence/rfdiffusion3-now-open-source-designs-dna-binders-and-advanced-enzymes/

[8] Overath, M. D., Rygaard, A., Jacobsen, C. P., et al. (2025). Predicting Experimental Success in De Novo Binder Design: A Meta-Analysis of 3,766 Experimentally Characterised Binders. bioRxiv. https://doi.org/10.1101/2025.08.14.670059

[9] Didi, K., Zhang, Z., Zhou, G., et al. (2026). Scaling Atomistic Protein Binder Design with Generative Pretraining and Test-Time Compute. arXiv preprint arXiv:2603.27950. https://arxiv.org/abs/2603.27950

[10] Cao, H., Cheng, S., Liu, M., et al. (2026). Chamaileon: Cross-Context Binder Design with Contextualized Modeling and Mixed Sampling. arXiv preprint arXiv:2607.23518. https://arxiv.org/abs/2607.23518

4.2 開発可能性を最大化する抗体CDR設計と選好性最適化

実験台の上に1本の試験管が置かれています。昨日の朝には透明だった液体が、今日は白く濁っています。底には白い沈殿物がたまっています。設計された抗体分子が凝集してしまったのです。ほんの数日前、ナノモル濃度の結合親和性を示し、研究室中を沸き立たせた候補でした。その候補は、直ちに廃棄されます。

抗体はY字型の糖タンパク質です。その2本の結合アームの先端には、相補性決定領域（CDR）と呼ばれる6つの超可変ループが存在します。これらのループが抗原の表面に接触して結合します。6つのうち、重鎖の3番目のループであるCDR-H3は、モデリングが極めて困難であることで知られています。その配列の長さは多種多様で、主鎖のコンホメーションは極めて柔軟です。抗原結合の特異性に最も大きく寄与する一方で、その構造的な柔軟性ゆえに計算による予測は極めて困難になります。

商業的な指標からも、このループの重要性が浮き彫りになります。世界で最も売れている医薬品上位10品目のうち、半分以上を治療用抗体が占めています。単一のモノクローナル抗体を市場に投入するには、10年以上の歳月と数億ドルの費用が必要です。その費用の大部分は、最初の候補を作り出すことではなく、その後の段階で失敗した候補を廃棄することから生じています。研究室では数万種類もの変異体をスクリーニングして一握りのリードを単離しますが、前臨床動物試験や初期の臨床開発の段階で10個中9個が失敗に終わります。初期の計算設計の段階で失敗の原因を特定できれば、開発資金と時間を大幅に節約できます。

抗体工学への人工知能の適用は、複数の方法論的な分岐を経て進展しました。2026年2月に発表されたAbFlowは、フローマッチングを全原子抗体設計に適応させ、抗原の構造的特徴を抽出する等変エンコーダーと、CDR-H3ループを洗練する速度場ネットワークを結合させました[1]。2026年7月に公開されたABOPDは、中核となる学習のダイナミクスを刷新しました。従来の拡散モデルは、ノイズが加えられた正解構造のみから学習します。しかし、自律的な生成を行う際、モデルがたどる中間状態は自己生成されたアーティファクトとなります。この分布の乖離により、予測されたループ座標は生成の軌跡に沿ってドリフトしてしまいます。ABOPDはOn-Policy Distillationを取り入れ、天然構造を正解データとして教師あり学習を行いながら、モデル自身が生成した軌跡を学習状態として用います。RAbDベンチマークにおいて、これによりCDR-H3の座標誤差は2.37オングストロームから1.95オングストロームへと0.42オングストローム減少しました[2]。1オングストロームは1億分の1センチメートルです。原子の直径の半分に相当する差が、抗体が標的に結合するかどうかを決定づけるのです。

高い結合親和性を持つ結合分子を設計することは、開発における課題の半分を解決したにすぎません。その候補分子は、製造可能であり、かつ臨床的にも実用可能でなければならないのです。これが「開発可能性（Developability）」です。第一に、抗体は溶液中で凝集してはなりません。第二に、高濃度であっても低い粘度を維持する必要があります。150 mgの皮下注射を行うには、製剤が注射針をスムーズに通過できなければなりません。第三に、チャイニーズハムスター卵巣（CHO）細胞のバイオリアクターにおいて高い発現力価を達成する必要があります。第四に、ヒトの免疫系がその分子を外来の免疫原として認識しないことが求められます。これら4つの要件は、常に互いに拮抗し合っています。疎水性残基を導入すると結合親和性が向上することが多い一方で、凝集傾向も高まってしまいます。お弁当箱のひとつの区画を広げれば、別の区画が必然的に狭くなってしまうのと同じです。

構造モデルは系統的な失敗パターンに直面することがあります。2026年5月に発表されたEvoStructは、抗体設計に適用された等変グラフニューラルネットワークが「語彙の崩壊（vocabulary collapse）」に陥る問題を指摘しました。高い配列再現率の指標を達成しているものの、生成される出力は狭い範囲のアミノ酸、とりわけチロシンやグリシンに著しく偏っていたのです[9]。これらの残基は抗体のループ領域に天然状態で豊富に存在するため、無差別に挿入するだけで評価スコアが不当につり上がってしまっていました。これは、すべての試験問題に対して暗記した同じ解答を書き込む生徒に似ています。EvoStructは、タンパク質言語モデル由来の進化学的事前分布と構造的特徴を結びつけることで、このバイアスを解決しました。2026年7月には、抗原表面を解析するために特別に設計されたマルチモーダル基盤モデルも登場しています[10]。

選好最適化（Preference Optimization）は、こうした多目的のトレードオフに対処するための枠組みを提供します。その数学的論理は、大規模言語モデル（LLM）を人間の選好に整合させるために用いられる手法と同様のものです。すべての物理的特性に対して明示的なスカラー目標値を定義するのではなく、どちらの変異体が優れているかを示すペアごとの候補比較をモデルに提示します。Direct Preference Optimization（DPO）は、個別の報酬モデルを必要とせずに、これらペアごとの選好データから直接パラメータを最適化し、望ましくない特徴を抑えつつ望ましい特性を生成する確率を高めます。凝集や粘度のような特性は、異なるアッセイ緩衝液や測定機器をまたぐと絶対的な定量化が困難であるため、この手法は抗体工学に非常によく適しています。一方で実験者は、同一条件下で2つの候補を評価し、どちらの凝集が少なかったかを容易に判定できます。これにより、最適化問題がラボのアッセイで直接答えられる形の問いへと再定義されるのです。

抗体の実験データセットは、この枠組みに自然に組み込むことができます。精密な定量的発現データは限られているものの、高発現クローンと非発現変異体を区別する二値的な記録は豊富に存

在します。2026年6月の研究では、この特性が利用されました。定量的に特性評価された1,254個の配列からなる小規模なデータセットとともに、ラクダ科動物への免疫によって得られた400万個のラベルなし抗体配列を弱教師ありシグナルとして活用したのです。IMGTナンバリングに基づいて可変長の配列を標準化し、マスクされた対数尤度を集約することで、この手法はほぼすべてのベンチマークにおいて既存の発現予測モデルを凌駕する性能を示しました[3]。

積極的な選好アラインメントにはリスクが伴います。特定の目的に対して過度に最適化すると、モデルが基礎的な生物物理学的事前知識（バイオフィジカル・プライア）を忘却し、予測上の結合親和性は高くても実際にはフォールディングしない配列を生成してしまうおそれがあります。2026年5月に発表されたProteinOPDは、個別の特性に特化した教師モデルからトークンレベルの知識を統合生徒モデルへと転移させることでこの破滅的忘却を緩和し、従来の強化学習ループと比較して8倍の学習速度向上を達成しました[4]。

標準化された評価フレームワークへの支持が広がりつつあります。過去3年間で抗体設計に関する論文が数十本発表されましたが、それぞれが非標準的な指標を用い、SAbDabデータベースのばらばらなバージョンで評価を行っていました。2026年3月に公開されたCHIMERA-Benchは、エピトープごとに層別化したベンチマークデータセットを確立しました[5]。多目的最適化に向けて、BOATはベイズ最適化によって複数のインシリコ予測モデルを統合し[6]、CrossAbSenseは特性特異的エンコーダーを活用して高額な生物物理学的アッセイを削減しました[7]。こうした進歩にもかかわらず、開発可能性（developability）の予測は結合親和性の予測ほど堅牢ではなく、同一の分子に対して独立したアルゴリズムが矛盾するスコアを提示することも頻繁にあります[6]。強い結合力があれば製造可能性も保証されると過信することは、実験台の上で白濁したバイアル瓶を再び生み出す結果に直結します。

免疫原性には特有の難しさがあります。凝集や粘性が化学や物理の現象に起因するのに対し、免疫原性は個人差の大きいヒトの免疫レパートリーに依存します。人工的に改変された抗体が異物と認識されると、患者の体内で抗薬物抗体（ADA）が産生され、薬効が中和されるばかりか、重篤な過敏症反応を引き起こす危険が生じます。そのため設計者は、CDRループをヒト生殖細胞系列のフレームワークに移植することで抗体候補をヒト化します。しかし、ヒトのフレームワークに厳格に適合させようとするとコンホメーションの自由度が制約を受け、結合親和性の低下を招くことが頻繁にあります。こうして、繊細なバランス調整が繰り返されます。

臨床での検証も着実に進展しています。Absci社がプロラクチン受容体（PRLR）を標的として開発したABS-201は、完全な生成AI設計による抗体創薬候補として初めてヒト臨床試験へと進み、2025年12月に投与を開始しました。2026年6月24日に発表された第1相試験の中間データでは、4段階の静脈内用量漸増コホート（150、450、900、1,800 mg）で評価された健康成人32名

において重篤な有害事象は報告されず、数か月ごとの投与間隔を可能にする薬物動態プロファイルが示されました[8]。対象となる適応症には、これまで非侵襲的な注射によるバイオ医薬品の選択肢が存在しなかった男性型脱毛症や子宮内膜症が含まれています。

こうした臨床上のマイルストーンを解釈する際には、冷静で厳格な姿勢が求められます。第1相臨床試験は安全性と薬物動態を評価するものであり、治療効果そのものはまだ証明されていません。加えて、「AI設計」と謳われる候補物質の多くは、ゼロから完全にde novo生成されたものではなく、既存のモノクローナル抗体を計算によって親和性成熟させたものです。親和性成熟は、現代のアルゴリズムにおいて最も成熟した応用領域です。計算機上のまったくの白紙からde novo生成され、臨床試験に到達した抗体は依然としてごく稀です。この分野の現在地を客観的に評価するためには、こうした違いを見極めることが不可欠です。

実験台の上で試験管が白濁する光景は、今でも毎週のように見られます。以前との違いは、生成モデルが試薬を混ぜ合わせる前の段階で、開発不能な候補にあらかじめ警告を出せるようになったことです。予測精度はまだ人間による物理的な測定の域には達していないものの、研究室で100個の候補を合成して99個を廃棄していた時代からは、すでに大きく前進しています。

参考文献

[1] Wang, W., Zhang, Y., Wei, Z., & Huang, W. (2026). AbFlow: End-to-end Paratope-Centric Antibody Design by Interaction Enhanced Flow Matching. arXiv preprint arXiv:2602.07084. https://arxiv.org/abs/2602.07084

[2] Yang, Z., He, J., Xie, J., et al. (2026). ABOPD: Antibody CDR Design via On-Policy Distillation. arXiv preprint arXiv:2607.18835. https://arxiv.org/abs/2607.18835

[3] Sun, J. Q., Babaie, M., Hou, W., Crowley, M., & Young, D. (2026). Preference-based Antibody Expression Ranking: Scaling with Large-scale Weak Supervision. arXiv preprint arXiv:2607.16263. https://arxiv.org/abs/2607.16263

[4] Zhang, Y., Cao, H., Jiang, Z., et al. (2026). ProteinOPD: Towards Effective and Efficient Preference Alignment for Protein Design. arXiv preprint arXiv:2605.10189. https://arxiv.org/abs/2605.10189

[5] Ahmed, M., et al. (2026). CHIMERA-Bench: A Benchmark Dataset for Epitope-Specific Antibody Design. arXiv preprint arXiv:2603.13431. https://arxiv.org/abs/2603.13431

[6] Rao, J., et al. (2026). BOAT: Navigating the Sea of In Silico Predictors for Antibody Design via Multi-Objective Bayesian Optimization. arXiv preprint arXiv:2604.13980. https://arxiv.org/abs/2604.13980

[7] Crouzet, S. J. (2026). Biologically-Grounded Multi-Encoder Architectures as Developability Oracles for Antibody Design. arXiv preprint arXiv:2604.09369. https://arxiv.org/abs/2604.09369

[8] Absci Corp. (June 24, 2026). Absci Announces Positive Interim Phase 1 Data from the HEADLINE Trial of ABS-201, a Novel Antibody Targeting the Prolactin Receptor (PRLR). https://investors.absci.com/news-releases/news-release-details/absci-announces-positive-interim-phase-1-data-headlinetm-trial/

[9] Ahmed, M., et al. (2026). EvoStruct: Bridging Evolutionary and Structural Priors for Antibody CDR Design via Protein Language Model Adaptation. arXiv preprint arXiv:2605.21485. https://arxiv.org/abs/2605.21485

[10] Shi, X., et al. (2026). Antigen-specific Antibody Multi-modal Foundation Model for Functional Antibody Design. arXiv preprint arXiv:2607.20057. https://arxiv.org/abs/2607.20057

4.3 Proteo-R1：タンパク質構造の理解にマルチタスク推論と言語的思考の連鎖を注入する

大学院生が画面上で解析ファイルを開きます。そこには「-1.8」という単一の数値が表示されています。これは、この特定の点変異を導入することで標的との結合がその数値分だけ弱まることを示しています。学生は指導教官に、なぜその変異が結合を阻害するのかを尋ねます。しかし、モデルが単一のスカラー値しか出力しないため、指導教官もそれを説明できません。なぜ210番目の残基が重要なのか、具体的にどの水素結合や塩橋が破壊されたのかは、ニューラルネットワークの重みの中に閉じ込められたままです。学生は、ただ盲目的にその数値を信じるべきかどうかの判断を迫られます。

近年、生成モデルの精度が向上するにつれて、この解釈可能性の壁はますます高くなりました。モデルは全原子精度で構造を構築する一方で、どの残基が機能特異性をもたらしているのかについての明示的な説明を一切提供しませんでした。設計上の決定が拡散モデルの連続的なデノイズ軌道の中に埋め込まれていたため、人間の研究者にはそのプロセスを評価したり誘導したりするための明確な手がかりがありませんでした。

このような不透明さは、実験が失敗した際に大きなコストとなります。ある研究室が計算機上で設計された結合分子を100個合成し、そのうち90個が結合しなかったと仮定してみましょう。研究者は次の実験計画を立てなければなりません。しかし、失敗の構造的な原因が分からないままでは、どのパラメータを変更すべきかを判断することは不可能です。唯一残された手段は、ハイパーパラメータを微調整し、再び見通しの立たないまま100個の候補を生成することだけです。計算による生成が加速した一方で、実験における原因究明とトラブルシューティングは停滞してしまいました。

スタンフォード大学および提携機関の25名の研究者からなる共同チームが2026年5月1日に発表したProteo-R1は、この課題に正面から取り組みました。ICML 2026で発表されたこのモデルは、構造の理解と構造の生成を分離することにより、明示的な推論（R1で表記）を取り入れています。

従来のアーキテクチャでは、標的構造を入力として受け取り、一度の処理で座標を出力するという形で、理解と生成が単一のエンドツーエンドのネットワーク内に統合されていました。この手法は高速であるものの、途中の解釈可能性に欠けています。それはまるで、材料を鍋に放り込んで蓋を閉め、途中で味見もせずに完成したシチューを食卓に出すようなものです。塩を入れすぎたかどうかは、調理がすべて終わってからでなければ分かりません。

Proteo-R1は、ワークフローを2つの専門モジュールに分離しています。上流のコンポーネントは「理解エキスパート（Understanding Expert）」です。これは、タンパク質配列、3次元座標、機能に関する文献のアノテーションを同時に処理し、標的への結合や結合特異性を左右する重要な残基を導き出すマルチモーダル大規模言語モデル（MLLM）です。下流のモジュールは「生成エキスパート（Generation Expert）」です。これはAlphaFold3スタイルの拡散モデルであり、特定された接触残基を固定された幾何学的制約として受け取り、3次元主鎖とアミノ酸配列を同時に生成します。これは、周囲の立体化学をモデル化する前に、どの生化学的接触が不可欠であるかを確定させるという、人間の構造生物学における推論プロセスを反映しています。

思考の連鎖（Chain-of-Thought: CoT）推論の実装は、生物物理学的な幾何構造に特化されています。標準的な言語モデリングにおいて、CoTは最終的な答えを提示する前に、中間的な自然言語のステップを生成することを意味します[2]。しかし、冗長な自然言語トークンを拡散モデルのデノイジング軌道に直接注入すると、空間座標の更新が不安定になり、文章としてはもっともらしくても幾何構造が歪んでしまいます。Proteo-R1は、中間推論をハードな制約として渡される離散的な残基インデックスへと蒸留することで、この課題を解決しています。推論は示唆的な論評であることをやめ、明示的な幾何学的設計図となるのです[1]。

定量的なベンチマークは、この設計の成果を反映しています。SAbDabの抗体複合体で学習された多量体CDR再設計タスクにおいて、CDR-H1の座標誤差は1.33オングストロームに低下し、従来のベンチマークであるMFDesign（1.61オングストローム）を上回りました。CDR-H2では、誤差が1.44オングストロームから1.13オングストロームへと減少しました。また、RAbDデータセットにおけるCDR-H3の再設計では、dyMEANの0.407に対して0.801というDockQスコアを達成しました。lDDTで測定された局所構造の一致度は0.9693に達し、評価されたすべてのベースラインを上回っています[1]。

同じ性能表からは、ある低い指標も明らかになっています。天然配列復元率は10.75%にとどまり、天然抗体に見られるアミノ酸の約10個に1個しか一致していません。これが失敗を意味するのか、それとも成功を意味するのかは、設計の意図によって異なります。もし目的が天然抗体の複製にあるなら、10.75%は低い数値です。しかし、同一の結合ポケット内で同じ生化学的機能を果たす代替的な配列解を発見することが目的であるならば、結論は変わってきます。異なる配列組成によって正確な3次構造のフォールディングと結合親和性を達成することは、特許の制約を回避し、宿主の免疫原性を逃れるうえで大きな利点をもたらします。私はこの結果を前向きに解釈しています。天然配列復元率を主要なベンチマークとして扱うことは、現代の生成能力から取り残された旧来の評価基準を反映しているにすぎません。

このようなハイブリッド推論モデルの学習には、特殊なペアデータセットが必要です。タンパク質構造データベースには原子座標が記録されている一方で、学術論文誌には人間が書いた文章が収められており、この2つを結びつける学習コーパスは歴史的に不足していました。2026年2月に公開されたAFD-Instructionは、抗体の機能アノテーションを自然言語のインストラクションチューニング用データセットへと構造化することでこの欠落を埋め、言語モデルが抗体のメカニズムについて推論できるようにしました[8]。

Proteo-R1は、推論能力を強化した生体分子AIへのより広範な潮流の一部です。2026年7月に発表されたS1-Omniは、生物学的理解、予測モデリング、生成設計にまたがるマルチモーダルな科学的推論を統合しました[3]。2026年3月に発表されたAgent Rosettaは、大規模言語モデル（LLM）にRosetta生体高分子モデリングスイートを組み込み、複雑な設計タスクを自律的に解決させました[4]。1月には、AutoBinder AgentがModel Context Protocol（MCP）を介して個別の設計ツールを接続し、バインダー探索ワークフローのエンドツーエンドの自動化を実現しました[5]。Latent-Yはこれを発展させて自律型ラボの検証ループを構築し[6]、共同生成フレームワークは自然言語の機能仕様を一次配列に直接結びつけました[7]。商業バイオテクノロジーの分野では、設計からスクリーニングまでを自動化したハイスループットパイプラインが現在では標準となっており[10]、計算機上のプロトタイプと製造バイオリアクターとの距離を縮めています。

この分野の学術文献は、依然として断片的な状態にとどまっています。研究によって配列復元率、座標RMSD、結合自由エネルギーなどの異なる指標が混在して性能報告に用いられており、同じ表の中で残基レベルのモデルと原子レベルのモデルが並べて評価されています。2026年3月の総説論文ではこうした現状を概観し、表現形式、条件付けフレームワーク、ベンチマーク基準の統合に向けた整理が行われました[9]。

こうした進化を決定づけているのが、ツールを使用するエージェント型ワークフローです。初期のモデルはすべての予測を内部だけで完結させていました。一方、近年のアーキテクチャは外部の計算エンジンを呼び出し、構造予測ツールの実行や結合自由エネルギーの計算を行い、初期の測定値が基準に届かない場合には設計パラメータを修正します。この計算ループは人間による反復的な設計プロセスを再現したものであり、一晩がかりの計画作業を、午前2時に自動生成される自律的な実行ログへと変貌させました。

もっとも、依然として注意は必要です。生成された推論テキストは説得力を持って読めるため、無批判な信頼を生んでしまう恐れがあります。言語モデルが「この残基は2.8オングストロームで水素結合を形成する」と主張した際、研究者が手動での検証を行わずにその主張を受け入れてしまうかもしれません。その空間的距離は、実際の計算に基づくものかもしれない一方で、単に

学習テキストでよく見られる表現を繰り返しただけという可能性もあります。Proteo-R1は、推論の出力を明示的な残基インデックスに結びつけることでこの問題に対処し、下流の構造出力によって推論の主張を直接検証できるようにしました。これにより、解釈可能性は単なる修辞ではなく、検証可能なものとなります。

依然として解決すべき限界も存在します。上流の理解モジュールが誤った残基を特定した場合、拡散プロセス中に厳密な幾何学的制約を覆すことができないため、下流の生成エンジンはその誤った指示を忠実に実行してしまいます。論文内のアブレーション研究では、正解となる接触残基を与えることで構造誤差が減少し、原子間の立体衝突が最小限に抑えられることが示されています[1]。下流の生成性能は、上流の推論精度の制約を受け続けます。推論モジュールが成功すれば構造の忠実度が向上し、標的を誤認すれば生成エンジンは無効な構造を作り出すことになります。

机に向かう大学院生の話に戻りましょう。現在の解析出力では、スカラー値のスコアと並んで重要な残基の明示的なリストが提示されます。大学院生はそのリストを精査し、特定の残基の選択には同意しつつ、他の選択肢は却下することができます。更新した制約条件を再入力することで、モデルは別の構造配置を生成します。ウェットラボでの検証には依然として数週間を要しますが、研究者はアッセイでどのような仮説が検証されているのかを正確に把握できるようになりました。

参考文献

[1] Wu, F., Xuan, W., Qi, H., et al. (2026). Proteo-R1: Reasoning Foundation Models for De Novo Protein Design. arXiv preprint arXiv:2605.02937. https://arxiv.org/abs/2605.02937

[2] Wei, J., Wang, X., Schuurmans, D., et al. (2022). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. arXiv preprint arXiv:2201.11903. https://arxiv.org/abs/2201.11903

[3] Zhao, J., et al. (2026). S1-Omni: A Unified Multimodal Reasoning Model for Scientific Understanding, Prediction, and Generation. arXiv preprint arXiv:2607.15686. https://arxiv.org/abs/2607.15686

[4] Teneggi, J., et al. (2026). Protein Design with Agent Rosetta: A Case Study for Specialized Scientific Agents. arXiv preprint arXiv:2603.15952. https://arxiv.org/abs/2603.15952

[5] Ge, F., et al. (2026). AutoBinder Agent: An MCP-Based Agent for End-to-End Protein Binder Design. arXiv preprint arXiv:2602.00019. https://arxiv.org/abs/2602.00019

[6] Latent Labs Team, et al. (2026). Latent-Y: A Lab-Validated Autonomous Agent for De Novo Drug Design. arXiv preprint arXiv:2603.29727. https://arxiv.org/abs/2603.29727

[7] Chen, X., et al. (2026). Co-Generative De Novo Functional Protein Design. arXiv preprint arXiv:2605.00948. https://arxiv.org/abs/2605.00948

[8] Luo, L., et al. (2026). AFD-INSTRUCTION: A Comprehensive Antibody Instruction Dataset with Functional Annotations for LLM-Based Understanding and Design. arXiv preprint arXiv:2602.04916. https://arxiv.org/abs/2602.04916

[9] Wanasekara, S. H., et al. (2026). Generative Modeling in Protein Design: Neural Representations, Conditional Generation, and Evaluation Standards. arXiv preprint arXiv:2603.26378. https://arxiv.org/abs/2603.26378

[10] Gao, J., et al. (2025). HelixDesign-Binder: A Scalable Production-Grade Platform for Binder Design Built on HelixFold3. arXiv preprint arXiv:2505.21873. https://arxiv.org/abs/2505.21873

第5章：ゲノムスケールのDNA言語モデルと生物学的コードの解読

5.1 全ゲノムDNA配列におけるヌクレオチド言語モデルの事前学習

深夜3時、研究室のモニターには4つの文字だけが流れています。A、C、G、Tです。どれほどスクロールしても、5番目の文字が現れることはありません。アデニン（Adenine）、シトシン（Cytosine）、グアニン（Guanine）、チミン（Thymine）。この4つの文字が、人間を構築するための完全な取扱説明書を構成しています。この説明書は30億塩基対（bp）に及びます。標準的な200ページの書籍として印刷すれば、5,000冊以上を埋め尽くす分量になります。

その5,000冊の中で、タンパク質の翻訳を指示する文章は全体の1.5%未満にすぎません。残りの98.5%は、歴史的に「ジャンクDNA」として軽視されていました。しかし今日、その言葉を使う研究者はいません。その98.5%は、どの遺伝子が、どの組織で、発生のどの段階で、どの程度の転写強度で活性化するかを決定する、遺伝的なオペレーティングシステムとして機能しています。これが「非コードゲノム（Non-coding Genome）」です。それはレシピというよりも、厨房の作業スケジュールに似ています。スケジュールが乱れれば、レシピの正確さにかかわらず厨房の作業は停止してしまいます。

ゲノム基盤モデルは、このスケジュールを解読するために構築された計算エンジンです。何千もの細菌、植物、哺乳類、そしてヒトの全ゲノムを新聞の束のように読み込むことで、生の遺伝子配列を一括して処理します。学習を開始する前に、連続した配列を離散化する必要があります。この分割処理が「トークン化（Tokenization）」です。配列を重複する6-merにまとめたり、頻出するモチーフからバイト対辞書を学習したり、あるいは個々のヌクレオチドを単一のトークンとして扱ったりします。これらの選択には明確なトレードオフが存在します。より粗いトークンは計算を高速化しますが、1塩基の解像度は精度を保証します。単一の塩基変異がどのように疾患を引き起こすかを追跡するには、単一ヌクレオチドの解像度が求められます。

学習は、マスク言語モデリング（Masked Language Modeling）を通じて、穴埋め問題のように進められます。配列トークンの15%をランダムにマスクし、隠された塩基を予測するようにネットワークを訓練します。人間のアノテーターがラベルを付与することはなく、本来の生体配列そのものが正解となります。人手によるラベル付けを必要としないため、この枠組みは自己教師あり学習（Self-Supervised Learning）と呼ばれます。マスクされた1つの塩基を予測するために、モデルは周辺にある数万から数十万の塩基をスキャンし、この最適化を何兆もの学習トークンにわたって繰り返します。

このような反復学習を通じて、ネットワークは生物学的構文を自律的に内面化していきます。かつては特徴付けに何十年もの生化学アッセイを要した制御モチーフである、遺伝子活性化スイッチ（プロモーター）、長距離エンハンサー（エンハンサー）、異常なクロストークを防ぐ境界（イ

ンスレーター）、そしてエキソンとイントロンの境界（スプライス部位）などを区別できるようになります。2021年にDNABERTがこの方向性の先駆けとなり[1]、2024年後半に『Nature Methods』誌で発表されたNucleotide Transformerは、850のゲノムで学習を行い、ヒトゲノムに関する18のタスクにおいて性能ベースラインを刷新しました[2]。

多様な生物種を横断してモデルを学習させる背景には、進化論的な論理が存在します。ヒトとマウスが進化学的に分岐してから8000万年が経過しても、あるゲノム領域が不変のままであるなら、その領域の変異は生物の生存にとって致命的だったということです。これらの配列には、40億年におよぶ進化の実験が凝縮されています。ヒトゲノムだけで学習したモデルでは、このような比較記録にアクセスできません。それに対して850の生物種を横断して学習したモデルは、どの遺伝子座を改変してはならないのかを認識できます。機械は、比較ゲノミクスという古典的な道具を実質的に再発明したのです。

モデリングの規模はさらに進展しました。2025年12月に公開されたNucleotide Transformerバージョン3（NTv3）は、パラメータ数が800万から6億5000万であり[3]、一般的な言語モデルの基準からすれば控えめです。しかし、そのコンテキストウィンドウは単一塩基解像度で100万塩基対（bp）におよびます。8兆個以上のヌクレオチドを用いた事前学習の後、動植物24種にわたる約1万6000の機能トラックとアノテーションラベルによるファインチューニングが施されました。7種にわたる106のベンチマークタスクでの評価において、遺伝子構造解析ではAUGUSTUS、SpliceAI、SegmentNTなどの長年使われてきたツールを凌駕し、発現量やタンパク質存在量の予測でも10億パラメータのAgroNTモデルを上回る精度を達成しました。1メガベースのウィンドウの処理速度は、HyenaDNAやCaduceusと比べて約10倍高速でした[4]。

これと並行して、Google DeepMindは統合的なアーキテクチャによるアプローチを推し進めました。2026年1月28日付の『Nature』第649巻で発表されたAlphaGenomeは、1メガベースのDNA入力を処理し、遺伝子発現、転写開始、クロマチンアクセシビリティ、ヒストン修飾、転写因子結合、クロマチンコンタクトマップ、スプライス部位の利用といった数千ものゲノム機能トラックを同時に予測します[5]。調節領域の変異影響に関する26のベンチマーク課題において、25の課題で従来の最高水準モデルと同等以上の性能を達成し、配列のみからあらゆる機能モダリティを同時に予測できる初のモデルとなりました。

この能力が臨床現場でどのように活きるかを考えてみましょう。原因不明の発達遅滞を抱える小児がいるとします。全ゲノムシーケンスを行っても、タンパク質をコードするエキソン領域には変異が見つかりません。しかし、発生において極めて重要な遺伝子から8万塩基対（bp）離れた位置で、シトシン1つがチミンに変異している一塩基多型（SNP）が確認されました。従来、こうしたたった1つの変化の性質を解明するには、数ヶ月におよぶウェットラボでの細胞アッセイ

が必要でした。一方、ゲノム言語モデルは野生型と変異型の配列を取り込み、内部の潜在表現がどれほど大きく変化するかを測定します。大きなズレが見られれば、その変異が転写因子の結合モチーフを破壊したことを意味します。この推論にかかる時間はわずか1秒です。

この1秒の推論が、臨床遺伝学における長年のボトルネックを解消します。患者が遺伝子検査の結果を受け取る際、変異は「病原性（Pathogenic）」「良性（Benign）」「意義不明の変異（VUS：Variant of Uncertain Significance）」の3つに分類されます。VUSと判定された患者やその家族は、予防的手術を受けるべきか、家族の遺伝子検査を行うべきかといった決断を下せないまま、何年間も先行きが見えない状態に置かれます。非コード領域の変異は実験による検証が難しいため、高い割合でVUSに分類されてしまいます。ゲノム言語モデルの本質的な価値は、新たな生物学的知見を発見することだけにとどまらず、このように未分類のまま滞留している臨床変異の山を解消することにあるのです。

ベンチマークのスコアカードで高い評価を得ているからといって、ゲノムモデリングの課題が解決されたわけではありません。2025年後半に公開されたNABenchは、標準化されたデータ分割とメタデータを備えた162種類のハイスループットアッセイにわたる260万件の変異配列を収集し、ゼロショット、少数ショット、転移学習、教師あり学習の各条件下で29種類の基盤モデルを評価しました[6]。その性能にはばらつきがあり、DNAベンチマークで優れていたモデルがRNAタスクでは性能低下を起こしました。2026年5月に発表されたViroBenchは、極めて重大な弱点を浮き彫りにしました。18種類のウイルスゲノムベンチマークにわたって66種類の塩基配列モデルを評価したところ、系統学的に離れた系統や、採取時期が異なるデータを評価する際に性能が急激に低下することが明らかになったのです。生成タスクにおいては、モデルが高い尤度を割り当てた配列が、実際の物理的な機能を果たさない事例が頻発しました。統計的な妥当性と生物学的な実効性は乖離しています。分類学的に多様なデータで事前学習されたコンパクトなモデルが、元の大型モデルを67.5%上回る性能を示し、学習データの構成が単なるパラメータ規模よりも重要であることが実証されました[7]。

2026年4月には、650配列からなるベンチマークであるMechanistic Invariance Testを通じて根本的な批判が提起されました[8]。研究者らは、調節モチーフの空間的配置をずらしたり、間隔を広げたり、鎖の向きを反転させたりすることで、自己回帰型、マスク型、双方向状態空間モデルという5種類の異なるアーキテクチャを評価しました。その結果、モデルは空間的位置を追跡できていないことが判明しました。モデルが見かけ上示していた感度は、配列における生のA/T含有率と強く相関しており、その相関係数は0.78から0.96に達していました。Evo2-1BとCaduceusは、調節モチーフが機能しない位置へと移動させられた配列に対して、より高い適応度スコアを付与し、生物学的な真実を逆転させてしまいました。塩基組成は、空間的位置よりも46倍も強い影響を及ぼしていたのです。対照的に、わずか100個のパラメータを持つ単純な位置

特異的重み行列は、ほぼ完璧な精度でこのテストに合格しました。モデル規模の拡大は、こうしたアーキテクチャ上のバイアスを軽減するどころか、むしろ増幅させました。

私はこれらの知見を失敗の物語とは捉えていません。これらは、ゲノムの文法書がまだ半分程度しか書かれていないことを示しています。機械は語彙を記憶したものの、空間的な構文を理解するのに苦戦しているのです。当面の最優先課題はパラメータを10倍に増やすことではなく、アーキテクチャの設計目標を再構築することです。応用研究はすでにこの方向へと進んでいます。2026年3月の研究では、埋め込みマッチングを通じて大規模ゲノムモデルからmRNAの表現を蒸留し、同クラスでトップの性能を維持しながらモデルサイズを200分の1に縮小しました[9]。2026年8月には、ゲノム深層学習（DL）における不確実性推定の信頼性を評価する実証研究が発表されました[10]。確信度のキャリブレーション（信頼度較正）を明示することは、生の予測スコアよりも重要です。患者の診療にあたる臨床医にとって、0.97という予測スコアが信頼に足るものかどうかを知ることこそが、第一の要件だからです。

参考文献

[1] 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. https://doi.org/10.1093/bioinformatics/btab083

[2] Dalla-Torre, H., Gonzalez, L., Mendoza-Revilla, J., et al. (2025). Nucleotide Transformer: building and evaluating robust foundation models for human genomics. Nature Methods, 22(2), 287-297. https://doi.org/10.1038/s41592-024-02523-z

[3] Boshar, S., Evans, B., Tang, Z., et al. (2025). A foundational model for joint sequence-function multi-species modeling at scale for long-range genomic prediction. bioRxiv. https://doi.org/10.64898/2025.12.22.695963

[4] InstaDeep. (February 18, 2026). Modelling the Genome with NTv3. https://instadeep.com/2026/02/modelling-the-genome-with-ntv3/

[5] Avsec, Z.̌, Latysheva, N., Cheng, J., Novati, G., Taylor, K. R., et al. (2026). Advancing regulatory variant effect prediction with AlphaGenome. Nature, 649(8099), 1206-1218. https://doi.org/10.1038/s41586-025-10014-0

[6] Li, Z., Ma, R., Tan, J., Tan, C., & Zheng, S. (2025). NABench: Large-Scale Benchmarks of Nucleotide Foundation Models for Fitness Prediction. arXiv preprint arXiv:2511.02888. https://arxiv.org/abs/2511.02888

[7] Ye, D., Hu, F., Hu, H., Hu, S., Tan, Y., Ouyang, W., Li, S. Z., Cui, J., & Dong, N. (2026). ViroBench: Benchmarking Nucleotide Foundation Models on Viral Genomics Tasks. arXiv preprint arXiv:2605.25388. https://arxiv.org/abs/2605.25388

[8] Cheng, B., & Zhang, J. (2026). The Mechanistic Invariance Test: Genomic Language Models Fail to Learn Positional Regulatory Logic. arXiv preprint arXiv:2604.06549. https://arxiv.org/abs/2604.06549

[9] Haidari, R., Martin, S., & Allard, M. (2026). Distilling Genomic Models for Efficient mRNA Representation Learning via Embedding Matching. arXiv preprint arXiv:2604.08574. https://arxiv.org/abs/2604.08574

[10] Saran, S., Ghanbari, M., & Ohler, U. (2026). Uncertainty-Aware Deep Learning for Genomics Applications: Insights from an Empirical Study. arXiv preprint arXiv:2608.11054. https://arxiv.org/abs/2608.11054

5.2 1塩基解像度でメガベース規模のコンテキストを読み解くEvoとEvo2のアーキテクチャ

長編小説の読解力テストを、ある一つの条件のもとで受ける場面を想像してみてください。それは、「ランダムに選ばれた2ページしか読んではいけない」という条件です。登場人物の名前を把握したり、文章の文体を評価したりすることはできますが、第3部でなぜその登場人物が命を落とすのかを読み解くことはできません。2023年までのゲノム深層学習（DL）モデルは、まさにこのような制約のもとで動作していました。

このボトルネックは、すべてのトークン間の相互作用を総当たりで計算するTransformerの自己注意機構（セルフアテンション）に起因していました。配列の長さが2倍になると計算量は4倍になり、配列の長さに対して2乗で増大します。その結果、実用的なモデルは数千塩基対（bp）程度のコンテキストに制限されていました。しかし、真核生物のゲノムはそのような短い区間だけで機能しているわけではありません。数十万から数百万塩基対（bp）も離れた場所にあるエンハンサーが、遠く離れたプロモーターを活性化します。この物理的な距離は、クロマチンループ（Chromatin Looping）によってDNAが3次元空間で折りたたまれることで橋渡しされます。これは、2次元の路線図上では遠く離れて見える2つの地下鉄の駅が、直接の乗り換え通路で結ばれている様子に似ています。コンテキストウィンドウが狭いモデルでは、その乗り換え通路を認識することができません。

突破口となったのは、標準的なアテンション機構の置き換えでした。2023年に発表されたHyenaDNAは、アテンションの代わりに長畳み込み（long convolution）を採用しました[1]。同年登場したMambaは、制御工学の状態空間モデル（State Space Model: SSM）を深層学習（DL）に適応させました[2]。状態空間モデルは、過去の配列履歴を連続的な隠れ状態に圧縮し、それを線形に前方へと進めます。計算コストは配列の長さに応じて線形にスケールします。コンテキストを100倍に拡張しても、計算量は1万倍ではなく100倍に増えるだけで済みます。

このアーキテクチャをゲノミクスに応用したことで、Evoが誕生しました。2024年11月に『Science』誌に掲載された初期のEvoモデルは、2兆7,000億個の原核生物のヌクレオチドを取り込み、分子スケールとゲノムスケールのモデリングを統合しました[3]。続いてEvo 2が登場しました。2025年2月にプレプリントとして公開され、2026年3月4日付の『Nature』誌第652巻に正式に掲載されました[4]。学習は12万8,000種以上にわたる9兆3,000億個のヌクレオチドに及びました。細菌や古細菌に焦点を当てていた前身モデルとは異なり、Evo 2はヒト、植物、菌類を含む真核生物のゲノムを取り入れました。公開されたモデルサイズは、2兆4,000億トークンで学習された7Bパラメータモデルと、9兆3,000億トークンで学習された40Bパラメータモデルの2

種類です。いずれも100万トークンのコンテキストウィンドウを1塩基解像度で処理します。StripedHyena 2アーキテクチャに基づいて構築されており、最適化されたTransformerと比べて約3倍高速に学習が行われました。

StripedHyenaという名前は、シマウマの縞模様のように層を交互に積み重ねるそのアーキテクチャを反映しています。大半の層にはアテンションを用いない長畳み込みと状態空間演算子が採用され、その間に疎なアテンション（スパースアテンション）層が散りばめられています。低コストな畳み込み層が長距離にわたって情報を伝達する一方で、アテンション層は重要な局面で精密な計算を実行します。これは、時折高速な客車を連結して走る貨物列車に似ています。列車全体を高速客車だけで編成すれば莫大なコストがかかり、すべてを貨物車だけで編成すれば精度が低下してしまいます。このバランスを見出したことが、Evo 2の学習効率の高さをもたらしました。

1メガベースのコンテキストウィンドウの有用性は、CRISPR治療薬の設計において極めて明白です。標的切断部位を選択する際、ゲノム編集はその局所的な塩基だけでなく、それ以上の影響を及ぼします。例えば、30万塩基対（bp）離れたエンハンサーが標的を失ったり、隣接する遺伝子の発現が2倍になったりすることがあります。狭いウィンドウしか扱えないモデルでは、こうしたシステム全体にわたる影響を評価できません。Evo 2は編集前後の1メガベース領域全体を取り込み、下流の転写への波及効果をシミュレーションします。

モデル公開1周年にまとめられた実証検証結果は、その能力の高さを実証しました[5]。バクテリオファージの設計では、生成された285個のファージのうち16個が複製に成功し、非標的菌株を損なうことなく標的の宿主細菌に感染しました。エピゲノム設計では、予測されたクロマチンアクセシビリティと実験測定値との相関が0.92から0.95のAUROCに達し、細胞種特異的な24種類の設計のうち4種類が標的細胞種において2倍以上のアクセシビリティ差を達成しました。家畜ゲノムの変異分類においては、マスクされた変異条件下で0.921のAUROCを記録しました。ヒト遺伝学では、乳がんと卵巣がんのリスクを左右する遺伝子であるBRCA1のディープミューテーションスキャンデータがブラインドテストとして用いられました。Evo 2は尤度スコアのみで機能喪失変異と良性多型を分離し、その内部の埋め込みに基づいて構築された分類器は、事前のファインチューニングなしで0.95のAUROCと0.88のAUPRCを達成しました。

生物学的コードを読むことと書くことは異なりますが、Evo 2はその両方をこなします。初期配列のプロンプトが与えられると、染色体スケールの連続したDNAを生成します。生成と推論時探索（Inference-Time Search）を組み合わせることで、標的とする成果を得ることができます。具体的には、候補集団を生成し、指定された制御特性を持たない変異体を刈り込んでいきま

す。研究者らはこの手法を用いて、標的細胞種でのみ選択的にクロマチンを開く合成制御エレメントを設計し、ゲノムの解釈からde novoゲノム合成への移行を示しました。

この能力は倫理的な問題を提起します。機能的な生物学的コードを書くことができる機械は、病原性を持つ配列を合成することも可能だからです。バイオセキュリティ上のリスクに対処するため、Evo 2の研究チームは真核生物の宿主を標的とするウイルス配列を学習コーパスから除外し[4]、ネットワークがヒトや動物の病原体を学習するのを防ぎました。これはViroBenchの知見とも重なります。統計的な尤度が物理的な生存可能性を保証するわけではありませんが、そのギャップを恒久的な安全障壁として頼りにするのは危険です。学習データセットのキュレーションは依然として人間の政策的決定事項であり、将来の関係者によって変更される可能性があります。

導入実績の指標は、インフラとしての規模の大きさを反映しています。GitHubでのダウンロード数は8万8000件、フォーク数は380件、Hugging Faceでのダウンロード数は10万件以上、APIクエリ数は7Bモデルで200万件、40Bモデルで600万件、OpenGenome2学習コーパスのダウンロード数は4万8000件、プレプリントでの引用は200件以上に達しています。このモデルは今や研究における標準となっています。

しかし、アーキテクチャ上の課題は依然として残されています。2026年4月、Evo2-7Bが3次元クロマチントポロジーを認識できているかを検証する評価が行われました[6]。トポロジカル関連ドメイン（TAD）境界と収束型CTCFループアンカーを含む1メガベースのウィンドウを評価対象とし、機能的モチーフを破壊した際に尤度スコアが低下するかを調べ、生成された配列が収束型CTCFループを再構成できるかをテストしました。Evo 2は双方の評価で不合格となりました。機能的なCTCF破壊とランダムな対照群を区別できず、TAD境界も部分的にしか復元できませんでした。局所的な配列モチーフは習得したものの、高次の3次元空間構造の学習には至らなかったのです。その理由の一端は、3次元クロマチンの区画化を持たない原核生物が学習データに大量に含まれていたことにあります。論文の著者は、コンテキストウィンドウをこれ以上拡大するだけでは不十分であり、細胞型のコンテキストや3次元接触行列をモデルのアーキテクチャに直接組み込む必要があると結論づけています。

予期せぬ創発的能力も現れました。2025年11月の研究では、Evo 2にいくつかのコンテキスト例を与えることでインコンテキスト学習（In-Context Learning）を実行できることが示され[7]、その性能は事例数に応じて対数線形にスケールしました。かつては人間の言語モデルに特有のものと考えられていたインコンテキスト学習が、A、C、G、Tのみで厳格に学習されたネットワークにおいて発現したのです。これは、インコンテキスト学習が人間の言語だけに特有の性質では

なく、大規模な予測的配列モデリング全般に共通する一般的な創発特性であることを示しています。

応用工学の進展も急速です。CRISPR-Cas12のガイドRNA効率を予測する研究では、基盤モデルの埋め込みと実験的なクロマチンアクセシビリティデータを融合させることで、データが乏しい条件下でも精度を維持することに成功しました[8]。細胞型特異的な制御エレメントを工学的に設計するフレームワークであるCtrl-DNAは、生成言語モデルと制約付き強化学習を組み合わせ、標的とする細胞系譜でのみ特異的に活性化する配列を生成しました[9]。私はこの方向性に同意します。単に生の配列データを増大させてスケーリングするだけでは、収穫逓減に直面することになります。実験による機能測定値を目的関数に統合する準メカニスティックなフレームワークこそが、次の段階を主導するでしょう[10]。コンテキストウィンドウを100万塩基対（bp）にまで拡大したことは、重要なマイルストーンでした。しかし、広い窓から外を眺めることができても、そこに広がる文法を理解できるとは限りません。窓はすでに開かれました。その先にあるものを読み解くための取り組みは、現在進行形で進められています。

参考文献

[1] Nguyen, E., Poli, M., Faizi, M., Thomas, A., Birch-Sykes, C., Wornow, M., et al. (2023). HyenaDNA: Long-Range Genomic Sequence Modeling at Single Nucleotide Resolution. arXiv preprint arXiv:2306.15794. https://arxiv.org/abs/2306.15794

[2] Gu, A., & Dao, T. (2023). Mamba: Linear-Time Sequence Modeling with Selective State Spaces. arXiv preprint arXiv:2312.00752. https://arxiv.org/abs/2312.00752

[3] Nguyen, E., Poli, M., Durrant, M. G., Kang, B., Katrekar, D., et al. (2024). Sequence modeling and design from molecular to genome scale with Evo. Science, 386(6723), eado9336. https://doi.org/10.1126/science.ado9336

[4] Brixi, G., Durrant, M. G., Ku, J., Naghipourfar, M., Poli, M., et al. (2026). Genome modelling and design across all domains of life with Evo 2. Nature, 652, 1349-1361. https://doi.org/10.1038/s41586-026-10176-5

[5] Arc Institute. (March 4, 2026). Evo 2: One Year Later. https://arcinstitute.org/news/evo-2-one-year-later

[6] Lee, U. (2026). Probing 3D Chromatin Structure Awareness in Evo2 DNA Language Model. arXiv preprint arXiv:2604.07196. https://arxiv.org/abs/2604.07196

[7] Breslow, N., Mishra, A., Revsine, M., Schatz, M. C., Liu, A., & Khashabi, D. (2025). Genomic Next-Token Predictors are In-Context Learners. arXiv preprint arXiv:2511.12797. https://arxiv.org/abs/2511.12797

[8] Dehghani Amirabad, A., Zhang, Y., Moskalev, A., Rajesh, S., Mansi, T., Li, S., Prakash, M., & Liao, R. (2025). Multimodal Modeling of CRISPR-Cas12 Activity Using Foundation Models and Chromatin Accessibility Data. arXiv preprint arXiv:2506.11182. https://arxiv.org/abs/2506.11182

[9] Chen, X., Ma, S., Lin, R., Lin, J., & Wang, B. (2025). Ctrl-DNA: Controllable Cell-Type-Specific Regulatory DNA Design via Constrained RL. arXiv preprint arXiv:2505.20578. https://arxiv.org/abs/2505.20578

[10] Kovaceˇvic´, L., Gaudelet, T., Opzoomer, J., Triendl, H., Whittaker, J., Uhler, C., Edwards, L., & Taylor-King, J. P. (2025). No Foundations without Foundations: Why semi-mechanistic models are essential for regulatory biology. arXiv preprint arXiv:2501.19178. https://arxiv.org/abs/2501.19178

第6章 微生物ゲノム解析と機能推論

6.1 GGBoundによる微生物の生理的限界予測とゲノムからの代謝産物探索

森の地面からスプーン1杯の土をすくい取ってみてください。そこには数十億もの個々の細菌が含まれています。それを研究室に持ち帰り、寒天培地に薄く塗り広げて3日間待ちます。プレートの上にはいくつのコロニーが育つでしょうか。運が良くても、せいぜい100個程度です。残りの数十億の細菌は、寒天の上で何の反応も示しません。死んでいるからではなく、私たちが用意した食事が彼らの好みに合わなかったからです。

標準的な実験室の培地で育たない微生物は、「難培養微生物（Unculturable Microorganisms）」と呼ばれます。偏性嫌気性菌の中には、大気中の微量な酸素に触れただけで死滅するものもあります。また、隣接する他の細菌種が特定のアミノ酸代謝産物を供給してくれない限り分裂できない菌も存在します。微生物学者たちは何十年も前からこの限界を理解していましたが、なす術がありませんでした。培養できなければ観察できず、観察できなければ研究することもできないからです。

メタゲノムシーケンシングはこの障壁を回避しました。研究者たちは単離した菌株を培養するのではなく、土壌から直接環境全DNAを抽出し、配列を解読して、断片化されたリードを計算によって生物の再構成ゲノム、すなわちメタゲノムアセンブルゲノム（MAGs）へと組み立てます。2026年1月に公開されたgcMetaデータベースには、104,266件の環境サンプルからアセンブルされた270万個のMAGsが登録されています[4]。種レベルの閾値でクラスタリングすると109,586個の明確なグループが形成され、そのうち69,248個（全体の63%）は正式な学名を持たない新規の分類群です。ロックフェラー大学の研究者たちは、単一の森林土壌サンプルから2.5テラベースの配列を解読し、数百個の完全な細菌ゲノムを復元しましたが、その99%以上がこれまで特徴付けられていなかった未知の種でした[5][10]。

しかし、その後に続くジレンマは即座に生じます。配列を手に入れたとしても、その生物の物理的な生息地まではわからないということです。4℃の深海海溝で繁栄しているのか、それとも80℃の熱水噴出孔で生きているのでしょうか。pH 3の酸性温泉を好むのか、それとも海洋と同等の塩分濃度を必要とするのでしょうか。どのような栄養素を消費するのでしょうか。従来、こうした特徴を明らかにするには経験的な培養が必要でした。温度、塩分濃度、炭素源を体系的に変化させながら何百回もの培養実験を行い、表のたった1行を埋めていたのです。

2026年5月14日に発表されたGGBoundは、このワークフローを逆転させることを提案しました[1]。上海交通大学などの研究チームによって開発されたこの自律型エージェントは、生のゲノム配列を入力として受け取り、微生物株の生理的生存限界（生存可能な温度範囲、最適pH、塩分耐性、基質利用プロファイル、細胞形態）を統合された推論フレームワーク内で予測します。著

者らは、International Journal of Systematic and Evolutionary Microbiology（IJSEM）、NCBI、BacDiveデータベースから収集した1,525株を対象に、6,448問からなる評価スイートを構築しました。

GGBoundの内部アーキテクチャを調べると、3つのコンポーネントから構成されていることがわかります。ゲノム言語モデルであるLucaOneが塩基配列を埋め込みベクトルへと変換します。これらの表現は、Qwen言語モデルの入力トークン空間へと渡されます。この言語モデルは、相同株を求めてデータベースを検索する検索モジュールと、ゲノムスケール代謝モデル（GEM）を実行する計算ツールを連携させます。GEMは、生物の代謝ネットワークを数学的なフラックスバランス方程式として表現します。エージェントはインシリコで特定の栄養素を除去し、その生物がバイオマス生産を維持できるかどうかを計算します。かつて研究者が実験台で行っていた培地の摂動実験を、コード上で実行するのです。

学習には、特殊な強化学習メカニズムが組み込まれています。最適化の過程で、研究者たちはランダムに並び替えられた非生物学的なゲノムベクトルを断続的に注入します。もしモデルが、真正なベクトルを与えられたときと破損したベクトルを与えられたときで同じ回答を出力した場合、その応答はゲノムの解釈から導かれたものではなく、言語モデルが元々持っていた事実的知識によって生成されたことになります。GGBoundはそのような応答にペナルティを与え、真正なゲノム配列が正確な予測を直接導き出したときにのみ報酬を与えます。これは、問題文を読まずに生物学の試験を解く生徒を見つけ出す、アルゴリズムによる試験監督のように機能します。

この設計により、40億パラメータのモデルが、はるかに大規模な汎用システムと競合できるようになります。GGBoundは、微生物の生理学的境界を予測するタスクにおいて、DeepSeek-V3.2、DeepSeek-R1、GLM-4.7、Kimi-K2と同等またはそれを上回る性能を示しました[1]。ゲノム表現を正確に解析することは、モデルパラメータを100倍に増やすことよりも効果的であることが証明されたのです。

評価された特性の中でも、基質利用能は実用的な価値を持っています。未培養株がどの糖を代謝し、どのアミノ酸を合成できないかを把握できれば、培地にどの栄養素を添加すべきかが特定できます。未培養の生物を培養するために専用の培地を設計する学問分野であるカルチャロミクス（Culturomics）では、従来、試行錯誤を通じて何百もの培地組成を用意する必要がありました。代謝モデルが組み合わせのレシピを計算によって絞り込むことで、実験用のアッセイプレートの数は数百枚から数十枚へと激減します。

ゲノムに含まれているのは、基礎的な生理学的マーカーだけではありません。細菌は、抗生物質や色素、毒素といった二次代謝産物を生成するための遺伝子群を、生合成遺伝子クラスター

（BGC）として知られる連続したゲノムカセットにまとめています。これらは、単一の工場フロアに整然と配置されたモジュール式の製造組み立てラインによく似ています。ポリケチド合成酵素（PKS）や非リボソームペプチド合成酵素（NRPS）は、基質ユニットをつなぎ合わせる逐次的な酵素組み立て機械として機能します。これらの酵素モジュールの順序と種類を読み取ることで、研究者は生成される分子の核となる化学骨格を紙の上に描き出すことができるのです。

antiSMASHは、この計算によるマイニングを自動化します。2025年にリリースされたバージョン8.0では、認識可能なクラスターの分類が81から101へと拡大されました[2]。antiSMASH Databaseのバージョン5には、5万4,800の細菌ゲノム、833の古細菌ゲノム、421の真菌ゲノムが収録されており、49万7,429箇所のBGC領域が特定されています。これは前回のリリースにおける23万1,534クラスターの2倍以上にあたります[3]。これらのクラスターの大部分については、生成される最終的な化合物は依然として不明のままです。地図は存在するものの、地名が空白になっているような状態なのです。

こうした空白を埋める方法論が登場しつつあります。ロックフェラー大学の研究者らは、未培養の土壌細菌からBGC配列を抽出し、元の宿主株を培養する必要性を回避して、コードされたペプチドを化学合成によって直接合成しました。合成された分子のうち、エルタシジン（erutacidin）は細菌膜のカルジオリピン脂質に結合して膜の完全性を損なわせ、薬剤耐性を持つ臨床病原体に対する殺菌効果を示しました[5]。

並行して、de novo分子生成も加速しています。2026年にマックマスター大学とスタンフォード大学の研究者らによって開発されたSyntheMol-RLは、強化学習を適用して抗生物質の創薬候補化合物を設計します[6]。約15万種類の市販ビルディングブロックと検証済みの50種類の化学反応を組み合わせることで、460億個の分子からなるアクセス可能なケミカルスペースが生み出されます。モデルはこの空間を探索し、合成が容易な候補を選択します。著者らは設計された58種類の分子を合成して評価し、そのうち6種類がAcinetobacter baumanniiを含むESKAPE病原体に対して強力な抗菌活性を示しました。

非構造化された文献の統合は、ゲノムマイニングを補完します。2025年12月の研究では、言語モデルを活用して学術論文から微生物株と代謝物のペアを抽出し[7]、Corynebacterium glutamicumとEscherichia coliが栄養補助食品（ニュートラシューティカル）の生合成において繰り返し用いられる主力菌株であることを特定しました。発見された菌株をバイオ製造へと移行させるには、標準化されたデータオントロジーが必要です。2026年2月18日に公開されたPREFERは、精密発酵コミュニティ向けの標準化オントロジーを確立し[8]、異なるバイオファウンドリにまたがるハイスループットバイオリアクターが統一された命名規則の下で発酵データを

記録できるようにしました。共通の用語体系がなければ、機関横断的なデータセットを統合することはできず、基盤モデルを学習させることもできません。

配列は物理的な生理機能と同義ではありません。GGBoundの出力は依然として予測にとどまります。ゲノム深層学習（DL）における不確実性を評価した2026年8月の実証研究では、モデルの信頼度スコアと物理的な実験精度が乖離した事例が報告されています[9]。それでもなお、私はこの計算科学的アプローチの方向性を支持します。経験的な培地組成の試行錯誤に3年間も迷走する代わりに、研究者は優先順位が付けられた10種類の候補組成リストを受け取ることができます。この技術は実験室での実験をなくすのではなく、実験の順序を整理するものです。その順序付けこそが、3年間の労力と3ヶ月の差を生み出すのです。

参考文献

[1] Huang, H., Gong, X., Wang, J., Bai, L., Xiao, X., Zhao, W., & Liang, S. (2026). GGBound: A Genome-Grounded Agent for Microbial Life-Boundary Prediction. arXiv preprint arXiv:2605.14442. https://arxiv.org/abs/2605.14442

[2] Blin, K., Shaw, S., Vader, L., et al. (2025). antiSMASH 8.0: extended gene cluster detection capabilities and analyses of chemistry, enzymology, and regulation. Nucleic Acids Research, 53(W1), W32-W38. https://doi.org/10.1093/nar/gkaf334

[3] The antiSMASH database version 5. (2026). Nucleic Acids Research, 54(D1), D522-D526. https://doi.org/10.1093/nar/gkaf1210

[4] Sun, Y., et al. (2025). gcMeta 2025: a global repository of metagenome-assembled genomes enabling cross-ecosystem microbial discovery and function research. Nucleic Acids Research, 54(D1), D724. https://academic.oup.com/nar/article/54/D1/D724/8307362

[5] Burian, J., et al. (2025). Bioactive molecules unearthed by terabase-scale long-read sequencing of a soil metagenome. Nature Biotechnology. https://doi.org/10.1038/s41587-025-02810-w

[6] Swanson, K., et al. (2026). SyntheMol-RL: a flexible reinforcement learning framework for designing easily synthesizable antibiotics. Molecular Systems Biology. https://doi.org/10.1038/s44320-026-00206-9

[7] Sun, X., Sarmah, N., & Guo, M. (2025). Literature Mining System for Nutraceutical Biosynthesis: From AI Framework to Biological Insight. arXiv preprint arXiv:2512.22225. https://arxiv.org/abs/2512.22225

[8] Amigó, T., Tan, S. Z. K., Luu Phanthanourak, A., et al. (2026). PREFER: An Ontology for the PREcision FERmentation Community. arXiv preprint arXiv:2602.16755. https://arxiv.org/abs/2602.16755

[9] Saran, S., Ghanbari, M., & Ohler, U. (2026). Uncertainty-Aware Deep Learning for Genomics Applications: Insights from an Empirical Study. arXiv preprint arXiv:2608.11054. https://arxiv.org/abs/2608.11054

[10] Rockefeller University. (September 2025). Hundreds of new bacteria, and two potential antibiotics, found in soil. https://www.rockefeller.edu/news/38239-hundreds-of-new-bacteria-and-two-potential-antibiotics-found-in-soil/

6.2 MicroFuse：オペロン構造と微生物制御経路推論のためのエキスパート融合フレームワーク

地下鉄の車両は一列に連結され、一人の運転士によって統括されます。細菌の遺伝子も同様に配置されています。乳糖（ラクトース）の代謝に必要な3つの構造遺伝子は染色体上で隣接して並んでおり、上流のプロモーターというスイッチが作動すると、3つすべてが一度に転写されます。この機能単位がオペロン（Operon）です。単一のオペレーターの制御下で連続したポリシストロニックmRNAへと転写される遺伝子クラスターを指します。

1961年、フランソワ・ジャコブ（François Jacob）とジャック・モノー（Jacques Monod）は大腸菌（Escherichia coli）における乳糖オペロン（Lacオペロン）の仕組みを解明しました[1]。培地に乳糖を加えると遺伝子のスイッチが入り、ブドウ糖（グルコース）が共存すると抑制されます。細菌は必要なときにだけ分解酵素を産生し、細胞内のエネルギーと基質を節約します。この発見以降、オペロンは原核生物が変動する環境に合わせて代謝フラックスを適応させる基本的な制御機構として確立されました。

モノーの二段階増殖（ダイオキシー）曲線は、2つの異なる増殖相を示します。グルコースとラクトースの双方が含まれる培地において、大腸菌はまずグルコースを消費し、指数関数的に増殖します。グルコースが枯渇すると、増殖は一時的に停止します。この二段階増殖の遅滞期（ダイオキシーラグ）の間に、細菌はlacオペロンを活性化させ、その後にラクトースを利用した指数関数的増殖を再開します。その短い停滞期は、遺伝子のスイッチを切り替えるのに必要な物理的時間を捉えています。これは、増殖の動力学を通じて目に見えない制御回路を追跡した実験でした。

ゲノム配列のみからオペロンの境界を特定することは困難です。隣接する遺伝子間の距離（遺伝子間距離）は大きく異なり、スペーサーが短い場合は同一オペロンであることを示す場合が多いものの、例外も数多く存在します。内部の副プロモーターが下流の遺伝子の転写を独立して駆動することもありますし、内在性のrho非依存性転写ターミネーターがあいまいなヘアピン構造を示すことも少なくありません。単純なヒューリスティックルールをそのまま適用すると、必然的に誤りが生じます。

2001年から2009年にかけて開発された手法は、遺伝子間距離、機能アノテーションの一致度、共発現の相関を確率的分類器に統合したものでした[2][3][4]。これらは、大腸菌や枯草菌（Bacillus subtilis）のように詳細に研究されたモデル生物では良好に機能しました。しかし、未培養のメタゲノムアセンブリにおいて壁に突き当たりました。特徴づけられていない門に属する遺伝子には、機能アノテーションも実験的な発現データセットも存在しないためです。

このことが、ゲノムを生物学的言語として解析する基盤モデルの開発を促しました。2025年7月に公開されたBacformerは、細菌ゲノムをタンパク質トークンの配列として扱います[5]。これは単語によって構成される文章に似ています。30億個のタンパク質配列からなる130万個のゲノムで学習され、各タンパク質はESM-2を通じて埋め込みが行われた後、Transformer層を通じて順次処理されます。著者らは、ロングリードRNAシーケンシングによって予測されたオペロンの境界を検証し、AlphaFold3によってタンパク質間相互作用を裏付けました。

2026年5月9日に発表されたMicroFuseは、このアーキテクチャを拡張しました[6]。タンパク質自体の同一性（構造を考慮した言語モデルであるProstT5によって解析）と、ゲノムのシンテニーコンテキスト（Bacformerによってモデル化）という、生物学における2つの異なる系統のエビデンスを調和させています。2つの埋め込みベクトルを単純に連結するだけでは、タンパク質の機能とゲノム構成が乖離している事例を解決する仕組みが備わっていないため、うまくいきません。

MicroFuseは、4つの特化モジュールで構成されるMixture-of-Expertsアーキテクチャを採用しています。具体的には、タンパク質エキスパート、ゲノムコンテキストエキスパート、コンセンサスエキスパート（双方のモダリティが一致する事例を処理）、そしてコンフリクトエキスパート（矛盾するシグナルの解決に特化）の4つです。ルーティングネットワークが結合埋め込みを処理し、4人のエキスパートにルーティング重みを分配します。これは4人の専門家からなる委員会に似ており、意見が対立する事例は特にコンフリクト専門家へとルーティングされます。最適化にはInfoNCE対照アライメントを使用し、相補的なモダリティ同士を引き合わせながら、判断が難しい矛盾した遺伝子ペアの重み付けを高めています。

微生物生物学において、矛盾するシグナルは頻繁に発生します。細菌は、薬剤耐性を拡散させる主要なメカニズムである水平遺伝子移動（HGT）を介して、系統の壁を越えて連続した遺伝子カセットを獲得します。また、バクテリオファージは宿主の染色体にゲノムDNAを挿入します。こうした遺伝子座では、隣接する遺伝子がまったく無関係な機能を実行することが珍しくありません。逆に、機能的に連動している代謝経路の酵素群が、環状染色体の正反対の端に離れて位置している場合もあります。空間的なシンテニーのみに頼ると誤りが生じますし、個々のタンパク質の相同性のみに依存しても同様です。コンフリクトエキスパートは、まさにこうしたエッジケースを仲裁するために設計されています。

著者らは、OMGメタゲノムコーパスから精選した10万組の遺伝子ペアを含むベンチマークデータセット「OG-Operon100K」を導入しました。同一鎖上で近接している5万組のペアが正例を構成し、距離が離れているか反対向きの5万組の陰性対照によって均衡が保たれています。このコーパスは、データセット分割間での配列リークを防ぐため、メタゲノムスキャフォールド単位

で厳格に区分され、学習用68,692組、検証用14,169組、テスト用17,139組のペアに分割されています。

データセットをスキャフォールド単位で厳格に分割することは不可欠です。同一のメタゲノム断片に由来する遺伝子ペアが学習セットとテストセットの双方に含まれてしまうと、モデルは一般化可能なシンテニールールを学習するのではなく、特定の配列のアーティファクトを記憶してしまいます。その結果、評価スコアは見かけ上高くなりますが、汎化性能は頭打ちになります。著者らは、名目上のスコアが低くなったとしても誠実な評価指標を確保するため、スキャフォールドレベルでの分割を徹底しました。

得られた数値は示唆に富んでいます。MicroFuseは0.6616のAUROCを達成したのに対し、単純なベクトル連結では0.6587、ProstT5単体では0.5884、Bacformer単体では0.5841でした[6]。マルチモーダルな融合は単一モダリティのベースラインを0.07以上のAUROCで上回った一方で、単純な連結に対するエキスパートルーティングによる限界ゲインは0.003でした。異なる生物学的視点を融合することには明らかな価値があるものの、ルーティング機構の改良による効果は漸進的なものです。論文では、配列類似性がベースラインモデルを誤った方向へ導いてしまうような難度の高い事例において、性能向上が集中していたと指摘されています。

0.66というAUROCは、一見すると控えめに見えるかもしれません。2026年5月21日に発表されたTransformerベースのオペロン予測モデルは、大腸菌（E. coli）において87.5%の正解率を達成し、複数種混合のベンチマークにおいて84.96%のF1スコアを記録しています[7]。しかし、これら2つの研究は異なる問題に取り組んでいます。後者が詳細にアノテーションされたモデル生物を評価対象としたのに対し、MicroFuseはアノテーションが付与されていない断片的なメタゲノムアセンブリを扱ったのです。教科書を持ち込める試験を受けることと、白紙の状態から取り組むことには明確な違いがあります。

ゲノム言語モデルに関する警告的な知見も明らかになりました。2026年4月8日の研究では位置摂動テストが導入され、5つのゲノム基盤モデルが、制御モチーフを非機能的な位置へ移動させた配列に対して、より高い適合度スコアを割り当ててしまうことが示されました[8]。100個のパラメータを持つ位置重み行列が数十億パラメータのモデルを上回ったことは、生物学的な構文が空間的な位置に依存していることを浮き彫りにしています。空間的な位置は単なる背景の文脈ではなく、それ自体が機能を持つコンテンツなのです。

エンジニアはこうした制御マップに大きく依存しています。合成生物学者がシャーシ株に人工代謝経路を導入する際、アノテーションされていない内因性オペロンにDNAを挿入すると、宿主の制御回路が破壊されてしまいます。2026年5月に公開されたような、自動化された遺伝子回路

設計のための強化学習ツール[9]が生体内で安全に機能するためには、包括的な内因性回路マップが必要です。生物学的廃水フィルターの性能を予測する物理誘導型AIモデルも、微生物群集の機能プロファイルを入力として使用しています[10]。微生物群集工学は、どの遺伝子が協調して制御されているかを特定することから始まります。

前節で取り上げた生合成遺伝子クラスターは、主にオペロンとして機能します。複雑な抗生物質を合成するには、共通の制御オペレーターの下で統制され、順次作用する10種類の酵素のカスケードが必要です。クラスターの境界を誤認すると不完全な合成経路が生じ、代替宿主で異種発現させた際に標的分子を産生できなくなります。難培養メタゲノムから新規抗生物質を発見することは、オペロンの境界を正確に特定することに直結しています。

オペロン予測は世間の大きな注目を集める分野ではありません。AUROCが0.003向上したとしても、業界ニュースになることはめったにありません。しかし、その0.003という数値は、相反する生物学的証拠を処理するために個別のエキスパートを定義するという、規律ある緻密なエンジニアリングを示しています。難培養微生物ゲノムの制御構造を解読する取り組みは、まだ始まったばかりなのです。

参考文献

[1] Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. Journal of Molecular Biology, 3(3), 318-356. https://doi.org/10.1016/S0022-2836(61)80072-7

[2] Ermolaeva, M. D., White, O., & Salzberg, S. L. (2001). Prediction of operons in microbial genomes. Nucleic Acids Research, 29(5), 1216-1221. https://doi.org/10.1093/nar/29.5.1216

[3] Price, M. N., Huang, K. H., Alm, E. J., & Arkin, A. P. (2005). A novel method for predicting operons in genome sequences. Nucleic Acids Research, 33(3), 880-892. https://doi.org/10.1093/nar/gki232

[4] Mao, F., Dam, P., Chou, J., Olman, V., & Xu, Y. (2009). DOOR: a database for prokaryotic operons. Nucleic Acids Research, 37(Database issue), D459-D463. https://doi.org/10.1093/nar/gkn757

[5] Wiatrak, M., et al. (2025). A contextualised protein language model reveals the functional syntax of bacterial evolution. bioRxiv. https://doi.org/10.1101/2025.07.20.665723

[6] Cho, S. (2026). MicroFuse: Protein-to-Genome Expert Fusion for Microbial Operon Reasoning. arXiv preprint arXiv:2605.08815. https://arxiv.org/abs/2605.08815

[7] Assaf, R., & Fakhri, B. (2026). Transformer-based operon prediction using textual representations of gene pairs. Bioinformatics Advances, 6(1), vbag140. https://doi.org/10.1093/bioadv/vbag140

[8] Cheng, B., & Zhang, J. (2026). The Mechanistic Invariance Test: Genomic Language Models Fail to Learn Positional Regulatory Logic. arXiv preprint arXiv:2604.06549. https://arxiv.org/abs/2604.06549

[9] Flynn, N. (2026). GenCircuit-RL: Reinforcement Learning from Hierarchical Verification for Genetic Circuit Design. arXiv preprint arXiv:2605.14215. https://arxiv.org/abs/2605.14215

[10] Uzma, Cholet, F., Quinn, D., Smith, C., You, S., & Sloan, W. (2025). EnviroPiNet: A Physics-Guided AI Model for Predicting Biofilter Performance. arXiv preprint arXiv:2504.18595. https://arxiv.org/abs/2504.18595

第7章：AIによるバーチャルスクリーニング（VS）と化合物設計

7.1 標的ポケットと化合物の3次元分子ドッキング、および薬剤-標的親和性（DTA）バーチャルスクリーニング（VS）

夜11時の実験室でのことです。ロボットアームが透明な384ウェルプレートの上にかざされています。それぞれのウェルへ、個別の化合物を含んだ一滴の液滴が落ちていきます。翌朝、蛍光プレートリーダーがどのウェルの光が消えたかを記録します。暗くなったウェルはヒット化合物の候補を示しています。これがハイスループットスクリーニング（HTS）です。物理的な試験管の中に実際の化合物を投入し、生物学的応答を経験的に測定する手法です。

低分子化合物の創薬における最初のステップは、鍵が鍵穴にはまるように、標的タンパク質の触媒ポケットや結合ポケットに適合する低分子有機化合物を特定することです。鍵穴の輪郭は不規則であり、特定の鍵しか適合しません。HTSは何十万もの鍵を1本ずつ試すことで機能します。試薬には数十万ドルの費用がかかり、アッセイの読み出しには数ヶ月を要します。そのような網羅的な探索を行った後でさえ、スクリーニングの対象は製薬企業の化合物保管庫に収蔵されている数百万種類の化合物に留まります。

その障壁となっているのは、ケミカルスペース（化学空間）の規模です。創薬に適した低分子化合物の数は、およそ10の60乗個と推定されています。それに比べれば、数百万種類の化合物ライブラリなど一滴の水にすぎません。ここ数年、オンデマンドで合成可能なバーチャルライブラリがこのギャップを埋め始めています。2026年8月現在、Enamine REAL Spaceは検索および合成が可能な945億個以上の分子を提供しており[6]、2025年9月時点（830億個）から100億個以上拡大しました。これらの分子は、在庫のある約20万個のビルディングブロックに対して検証済みの169の並列合成プロトコルを適用することによって生成されます。注文された化合物は、80%を超える合成成功率で3〜4週間で到着します。それらは紙の上の理論的な構造にとどまるのではなく、宅配便で手元に届けられるのです。また、無料のZINC-22データベースには370億個以上の購入可能な分子が収録されており、その中には分子ドッキングに即座に利用できる事前計算済みの3次元コンホメーションが45億個含まれています[5]。

ロボット液体分注機を用いて945億個の物理的な分子を評価することは不可能です。鍵をインシリコでドッキングさせなければなりません。AutoDock Vinaのような従来の分子ドッキングソフトウェアは、モンテカルロ探索を用いて物理的な力場全体でリガンドのポーズをサンプリングし[4]、立体障害を排除しながら低エネルギーのポーズを保持します。精度は高いものの速度が遅く、1つのリガンドをドッキングさせるのに数十秒を要します。10億個の化合物を評価しようとすれば、数千台のCPUを何週間も稼働させ続けなければなりません。

分子ドッキングは、相互に関連する2つの課題を結びつけています。ポーズ予測（結合ポケット内におけるリガンドの3次元の平行移動、回転、および二面角コンホメーションの決定）と、スコアリング（結合自由エネルギーをスカラー量として推定すること）です。アルゴリズムが空間的に妥当なポーズを生成しながら、不正確な結合スコアを出してしまうことは珍しくありません。計算においては、ファンデルワールス反発力、水素結合、疎水性脱溶媒和ペナルティ、芳香族のπ-πスタッキングを考慮する必要があります。これらの物理的な力の総和を一握りの経験的ポテンシャル項へと近似することが、計算誤差を生み出す原因となります。

DiffDockはこの計算課題を再定義し[1]、ドッキングをエネルギー最適化探索ではなく生成的な拡散プロセスとして扱いました。DiffDockはリガンドの自由度を、3次元空間の並進、剛体としての3次元回転、および内部のねじれ二面角という3つの要素にパラメータ化します。この多様体上で拡散プロセスを逆行させることで、結合ポーズを直接生成します。PDBBindベンチマークにおいて、top-1予測が実験による結晶構造ポーズの2オングストローム以内に収まった割合は38%に達し、従来の物理ベースのドッキングの23%や、従来の深層学習（DL）モデルの20%を上回りました。結晶構造ではなく、計算によって予測されたAlphaFoldのタンパク質構造で評価した場合、その性能差はさらに広がりました。従来の手法が10.4%に低下したのに対し、DiffDockは21.7%を維持しました。

これらのアーキテクチャの基盤には、同変グラフニューラルネットワーク（Equivariant Graph Neural Networks）が存在します。同変性により、3次元の入力座標を回転させると、出力される立体配置もまったく同一に回転することが保証されます。プラスチック製の分子模型を90度回転させても、その内部の結合距離が変わらないのと同じです。この幾何学的な不変性をネットワークの重みに組み込むことで、モデルは任意の角度にわたる回転変換を再学習する学習計算リソースを浪費せずに済みます。DiffDockは、与えられたリガンドに対して数十個の候補ポーズをサンプリングし、専用の信頼度モデルを用いて順位付けを行い、確信が持てない場合には低いスコアを付与します。

こうした進歩には客観的な評価が必要です。PoseBustersの研究では、深層学習（DL）によるドッキングモデルが生成したポーズを標準的な物理的・化学的基準に照らして再評価しました[7]。その結果、正解座標に近いスコアでありながら、深刻な原子同士の立体衝突、歪んだ結合角、あるいは物理的に不合理な立体化学を含むポーズが多数見つかりました。また、新規な配列同一性を持つ標的タンパク質に対しては性能が低下しました。深層学習（DL）によるドッキングは、あらゆる基準において従来の物理エンジンを一様に凌駕しているわけではありません。そのため、実際のパイプラインでは、ニューラルネットワークを用いて初期ポーズを迅速に生成した後、物理ベースの力場によるエネルギー最小化を行う手法が採られています。このハイブリッ

ド戦略は、ニューラル表現から高速性を引き出しつつ、物理法則によって立体化学の妥当性を担保する実用的なアプローチです。

ポーズが特定されると、解離定数（Kd）、阻害定数（Ki）、または半数阻害濃度（IC50）として表される薬剤-標的結合親和性（DTA）予測を通じて、結合の強さを定量化する必要があります[2]。数値が小さいほど強力な結合を示します。マイクロモル濃度の親和性は弱く、ナノモル濃度はリード最適化に適しており、ピコモル濃度は極めて強力な結合を意味します。初期のモデルは1次元の配列文字列から親和性を予測していましたが、現代のシステムは3次元の原子座標を直接読み取ります。粗視化構造表現を用いたTerraBindは、従来のベンチマークと比較して推論を26倍に高速化し、親和性予測の精度を約20%向上させました[8]。この高速化により、評価可能な候補の数が大幅に拡大します。

計算速度の実用的なインパクトは、Boltz-2によって明確に示されています[3]。自由エネルギー摂動法（FEP）は結合自由エネルギー計算のゴールドスタンダードとされていますが、1つの化合物を評価するだけでも数時間のGPU計算を要します。Boltz-2は、FEP+ベンチマークにおいてピアソン相関係数0.62を達成し、既存のオープンソースFEPパイプラインに匹敵する精度を維持しながら、1,000倍以上高速に動作しました。また、140個のブラインド複合体を対象としたCASP16の親和性推定チャレンジにおいても、競合するすべての提出モデルを上回る成績を収めました。計算時間を数時間から数秒へと短縮することは、この作業の本質を一変させます。すなわち、10個の候補を確認する検証ツールから、100万個の化合物をスクリーニングする探索エンジンへと転換させるのです。

ウェットラボでの検証が、この能力を如実に実証しています。2026年8月5日、Deep Originは、がん細胞が抗腫瘍免疫を抑制するために利用する酵素であるCD73を標的として、800億個の化合物を含む仮想ライブラリに対するスクリーニング結果を発表しました[9]。予測された上位183個の候補化合物を実際に合成し、in vitroで試験したところ、56個の活性ヒットが得られ、ヒット率は30.6%に達しました。同じ標的に対する以前の機械学習（ML）を用いたスクリーニングでは、335個の化合物を試験してわずか1個の弱い阻害剤しか見つからなかったため（ヒット率0.3%）、実に100倍もの向上を示したことになります。主要なリード化合物は、天然基質の従来のヌクレオチド骨格を模倣することなく、細胞レベルで570 nMのIC50を達成し、新規なケミカル・スターティング・ポイントを確立しました。183個の化合物を合成する作業は、単一のアカデミアの研究室でも1か月で処理できる実験規模ですが、従来の探索で同等の成果を得るには数千回もの物理的アッセイが必要でした。

すべての標的がバーチャルスクリーニング（VS）で容易に攻略できるわけではありません。KRASやMYCのように、結合表面が浅く特徴に乏しいタンパク質に対しては、生成モデルも苦戦

を強いられます。深い触媒クレフトであれば低分子化合物の鍵を容易に受け入れることができますが、平坦なタンパク質表面には足がかりとなるポケットがほとんど存在しません。ShallowBenchは、浅いポケットを持つ標的においてモデルの性能がどれほど急激に低下するかを定量化するための標準化されたベンチマークを確立しました[10]。困難な問題の境界条件を明確に定義することは、技術が進歩するうえで不可欠です。

新たに確立されつつあるワークフローは極めて整然としています。朝のうちに標的の座標データが読み込まれ、日中には計算クラスタが数十億個の候補化合物をスクリーニングし、夕方には優先順位の付けられた50〜100個の分子リストがまとめられます。かつては384ウェルプレートへ無差別に試薬を注入していた自動分注ロボット（リキッドハンドラー）は、今やモデルによって指定されたウェルに対してのみ試薬をピペッティングしています。

参考文献

[1] Corso, G., Stärk, H., Jing, B., Barzilay, R., & Jaakkola, T. (2022). DiffDock: Diffusion Steps, Twists, and Turns for Molecular Docking. arXiv preprint arXiv:2210.01776. https://arxiv.org/abs/2210.01776

[2] Khan, J., & Shuvo, M. H. (2026). Recent Advances in Deep Learning-Based Drug-Target Binding Affinity Prediction. arXiv preprint arXiv:2608.13797. https://arxiv.org/abs/2608.13797

[3] Passaro, S., Corso, G., Wohlwend, J., et al. (2025). Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. bioRxiv. https://doi.org/10.1101/2025.06.14.659707

[4] Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry, 31(2), 455-461. https://doi.org/10.1002/jcc.21334

[5] Tingle, B. I., Tang, K. G., Castanon, M., et al. (2023). ZINC-22: A Free Multi-Billion-Scale Database of Tangible Compounds for Ligand Discovery. Journal of Chemical Information and Modeling, 63(4), 1166-1176. https://doi.org/10.1021/acs.jcim.2c01253

[6] Enamine. (August 18, 2026). REAL Space. https://enamine.net/compound-collections/real-compounds/real-space-navigator

[7] Buttenschoen, M., Morris, G. M., & Deane, C. M. (2023). PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. arXiv preprint arXiv:2308.05777. https://arxiv.org/abs/2308.05777

[8] Rossi, M., Pederson, R., Wang-Henderson, M., et al. (2026). TerraBind: Fast and Accurate Binding Affinity Prediction through Coarse Structural Representations. arXiv preprint arXiv:2602.07735. https://arxiv.org/abs/2602.07735

[9] Deep Origin. (August 5, 2026). Deep Origin Introduces Novel Virtual Screening Architecture; Achieves 100-fold Hit Rate Jump on Challenging Target. https://deeporigin.com/news/virtual-screening-architecture-100-fold-hit-rate/

[10] Reddy, S., & Liu, S. (2026). ShallowBench: Benchmarking Generative Drug Design Models on Shallow-Pocket Targets. arXiv preprint arXiv:2606.06717. https://arxiv.org/abs/2606.06717

7.2 アクティブラーニングと極度なデータ不足を克服するSPADEフレームワーク

ホワイトボードに、ある標的タンパク質の名前が書かれています。その下には、「12」という1つの数字だけが記されています。この12という数字は、学術論文などの公開文献において、このタンパク質への結合が確認されている化合物の総数を表しています。わずか12個。両手で数え切れてしまうほどの数です。教師あり深層学習（DL）モデルの学習には、通常、数万件規模のデータポイントが必要となります。希少遺伝性疾患や新興感染症の病原体、あるいは構造が解明されたばかりの未創薬（undrugged）標的タンパク質に取り組む際、研究現場は日常的にこのようなホワイトボードの現実に直面します。

わずか12件のデータポイントで標準的なニューラルネットワークを学習させると、深刻な過学習（オーバーフィッティング）が生じます。モデルはこれら12個の分子に共通する特定の化学官能基を丸暗記してしまい、その官能基が現れるたびに高い予測結合スコアを付与してしまいます。これは、特定の練習問題を丸暗記した生徒が見慣れない試験問題で失敗するのと同様であり、新規なケミカルスキャフォールド（化学骨格）に対する予測は信頼性のない根拠を欠いた数値になってしまいます。しかし創薬において、最も価値の高いリード候補化合物は新規スキャフォールドから得られます。よく知られたケモタイプ（化学構造群）は、すでに競合他社によって特許化されているからです。

アクティブラーニング（能動学習）は、問い合わせループ（クエリーループ）の仕組みを変えることで、このデータ不足を克服します。モデルに静的な学習データを与えるのではなく、モデル自身が自らのデータ収集を主導するのです。サロゲートモデル（代理モデル）は、個々の候補分子に対して予測親和性スコアと明確な不確実性の指標を共に出力します。そして、自身の予測の不確実性が最も高い領域、すなわち物理的なアッセイデータが最大の情報利得をもたらすケミカルスペース（化合物空間）を特定し、それらの候補を研究者に提示します。ウェットラボはその少量のバッチを合成してアッセイを行い、その結果をモデルにフィードバックしてネットワークの重みを更新します。この能動的なループを繰り返す姿は、見慣れた書架を眺めるのではなく、図書館の中で最も理解の進んでいない本棚を自ら探して手に取る研究者に似ています。

候補選択のバランスを取るには戦略が必要です。スコア上位の分子ばかりを選択すると、短期的には実験でのヒット数を最大化できますが、探索範囲はすでに知られているケミカルスペースに限定されてしまいます（活用：Exploitation）。一方で、不確実性の高い候補ばかりを選択すると、未踏の領域を素早くマッピングできますが、強力なリード分子の特定は遅れてしまいます（探索：Exploration）。実際の創薬現場では、各実験バッチの半分を高信頼度のヒット候補に、残りの半分を新規かつ不確実なケモタイプに割り当てることで、両者のバランスを取ってい

ます。このバランスをどう設計するかという点にこそ、パイプラインを導く人間の判断が生かされます。

実証実験の結果も、このアプローチの効率性の高さを示しています。HASTENフレームワークは、Enamine REALライブラリに含まれる15億6,000万個のリード様化合物（lead-like compounds）を評価しました[1]。ライブラリ全体のわずか1%を反復的にドッキングするだけで、真の上位1,000件のヒット化合物の90%を回収し、全体のドッキング計算の99%を削減することに成功しました。2025年3月に『Nature Computational Science』誌に掲載された研究では、このアプローチを数十億個規模の化合物に拡大適用し、実験的に検証された新規の結合分子を単離することに成功しています[2]。

こうしたパイプラインにも限界が存在します。アクティブラーニングにおけるサロゲートモデルは通常、2次元分子グラフ上で動作するため、3次元の受容体コンホメーションや結合ポーズを考慮できません。6つの標的を対象にアクティブラーニングのサロゲートモデルを評価した研究では、高スコアを獲得した学習化合物に共通する構造断片をモデルが頻繁に記憶してしまう傾向が浮き彫りになりました [3]。このような帰納的バイアスが存在するものの、パイプラインはDUD-Eベンチマークから真の活性分子を単離し、Enamine REALから高スコアの化合物を回収することに成功しており、利用者がモデルの前提条件を理解していれば実用的な有用性を持つことが実証されています。

2026年5月に発表されたSPADEは、この課題をサンプル効率の挑戦として捉えて取り組みました[4]。同論文は、初期のヒット・トゥ・リード探索を生き残る候補リガンドが5%未満にすぎないという厳しい現実の指摘から始まります。SPADEは、事前の生物活性データが全く存在しない新規の標的タンパク質に対してもスクリーニングを開始できるように設計されました。ベンチマーク評価全体を通じて、10個の高品質な活性リガンドを確保するのに必要とされた物理的アッセイの反復回数は平均40回でした。他の深層学習（DL）やベイズ最適化のベースラインと比較したベンチマークにおいて、SPADEはサンプル効率の中央値を7%から32%向上させ、最も近い競合手法と比べて候補プールを10倍高速にスコアリングしました。著者らはソースコードとデータセットをオープンに公開しています。反復的なアクティブラーニングにおいては速度が極めて重要です。ウェットラボのサイクルごとにサロゲートモデルの再学習と数億個の仮想分子の再スコアリングが必要となるためです。もし再スコアリングに24時間かかればウェットラボの作業は中断してしまいますが、スコアリングが数時間で完了すれば、計算上の反復と実験室でのアッセイが同じ週のうちに同期して循環します。

40回の実験的反復を実験室の実務指標に換算すると、その影響の大きさがよくわかります。化合物を合成してその結合定数を測定するには、数日間の時間と数百ドルの費用がかかります。40回

の物理的アッセイであれば、1人の化学者が2か月で完了できます。これと同等の発見率をランダムスクリーニングで達成しようとすれば、数千回もの経験的アッセイが必要になります。これによりプロジェクトの予算構造が変わり、大学院生の研究期間が短縮され、化学試薬の無駄や有害な溶媒の廃棄が削減されます。

アクティブラーニングにおける核心的な要件は、誠実な不確実性推定です。モデルが誤った予測に対して自信を示してしまうと、スクリーニングループ全体が偽陽性の領域へと迷い込んでしまいます。研究チームは、多様なモデル初期化における分散を定量化するためにディープアンサンブルを配備したり、推論時にモンテカルロドロップアウトを利用したりしています。2026年6月には、分子拡散モデルに特化した事後的不確実性推定法が発表されました [5]。また、多目的のバランスを取ることも不可欠です。実用的な薬剤は、標的に結合するだけでなく、水溶解性、肝細胞における代謝安定性、そして低い細胞毒性を維持しなければなりません。BOATは、多目的ベイズ最適化を通じてこれらのトレードオフを統括・調整します [6]。

ループの反対側には、分子設計生成モデルが位置しています。現代の大規模言語モデル（LLM）は、自然言語の文と同じように化学SMILES文字列を生成します。ここでよく生じる失敗パターンは、原子価の規則に違反したり環構造を閉じ損ねたりする、化学的に無効な文字列を生成してしまうことです。検証可能な化学的報酬を用いた強化学習によって、モデルが化学的妥当性を保つように訓練されます[8]。検証は人間による手動のスコアリングを介さず、自動化されたルールによって進められるため、学習は長期にわたる反復にも容易に拡張できます。生成モデルが新規のスキャフォールドを提案し、アクティブラーニングが合成すべき最も情報量の多い候補化合物を絞り込み、ウェットラボのアッセイ結果が生成事前分布を洗練させます。これら3つのエンジンが同期して機能します。

このループ内に潜むリスクも管理が必要です。スクリーニングのサイクルが進むにつれて、推奨される候補化合物は似通った構造モチーフへと収束していく傾向があります。初期の反復で高いスコアを獲得した単一のスキャフォールドがその後の学習バッチを占有し、モデルのバイアスを強化してしまうのです。このような雪だるま式の偏りを防ぐため、創薬チームはすべてのバッチにおいてスキャフォールドの多様性制約を義務付け、予測スコアが低くても構造的に異なる分子を意図的に含めます。そのわずかなトレードオフを受け入れることで、モデルの大域的な探索能力が損なわれずに維持されます。

最終的な検証は、常に実験台の上で行われます。分子生成モデルを評価したある研究では、インシリコのスコアがin vitroの細胞アッセイと頻繁に乖離することが確認されています[7]。このギャップを埋める商業的なマイルストーンが2026年に達成されました。Insilico Medicineは、TNIKに対する7万8,000個の仮想阻害剤を生成し、多特性フィルタリングを適用して上位60個の

分子を合成して、16.7%のヒット率を達成しました。こうして得られたリード候補化合物であるrentosertibは、2026年7月に特発性肺線維症を対象とした第3相臨床試験に進み[9]、AIが設計した低分子化合物として初めてヒトを対象とした第3相試験に到達しました。

まったく同じアクティブラーニングの原理が、低分子化合物の化学以外の領域にも適用されます。2つまたは3つの遺伝子を同時にノックアウトする組み合わせCRISPR摂動スクリーニングは、指数関数的な組み合わせ爆発に直面します。アクティブラーニングが次に実施すべき多重遺伝子ノックアウトを選択することで、実験の反復回数を大幅に削減します[10]。低分子化合物を評価する場合でも遺伝子ノックアウトを評価する場合でも、基本原則は同じです。盲目的にすべてをテストするのではなく、モデルが何を把握していないのかを計算し、そこへ優先的にラボ実験を向けることです。

ホワイトボードの例に戻りましょう。プロジェクトは依然として12種類の既知の結合分子から始まります。違いは、翌週の実験がどのように選定されるかにあります。以前は完全に人間の直感に頼っていましたが、現在ではモデルが最も不確実性の高い領域を提示し、研究者が優先順位付けされた候補化合物を明確な根拠を持って評価できるようになりました。100回未満の集中的な実験によって、ナノモル濃度のリード候補化合物が手の届くところまで引き寄せられます。

参考文献

[1] Sivula, T., Yetukuri, L., Kalliokoski, T., Käsnänen, H., Poso, A., & Pöhner, I. (2023). Machine Learning-Boosted Docking Enables the Efficient Structure-Based Virtual Screening of Giga-Scale Enumerated Chemical Libraries. Journal of Chemical Information and Modeling, 63(18), 5773-5783. https://doi.org/10.1021/acs.jcim.3c01239

[2] Luttens, A., Cabeza de Vaca, I., Sparring, L., et al. (2025). Rapid traversal of vast chemical space using machine learning-guided docking screens. Nature Computational Science, 5(4), 301-312. https://doi.org/10.1038/s43588-025-00777-x

[3] Kim, J., Nam, J., & Ryu, S. (2024). Understanding active learning of molecular docking and its applications. arXiv preprint arXiv:2406.12919. https://arxiv.org/abs/2406.12919

[4] Nandakumar, R., Fauber, B., & Chakrabarti, D. (2026). SPADE: Faster Drug Discovery by Learning from Sparse Data. arXiv preprint arXiv:2605.05370. https://arxiv.org/abs/2605.05370

[5] Seij, P., Naesseth, C. A., Mandt, S., et al. (2026). Uncertainty Estimation for Molecular Diffusion Models. arXiv preprint arXiv:2606.13451. https://arxiv.org/abs/2606.13451

[6] Rao, J., Gonzalez Hernandez, F., Gerard, L., et al. (2026). BOAT: Navigating the Sea of In Silico Predictors for Antibody Design via Multi-Objective Bayesian Optimization. arXiv preprint arXiv:2604.13980. https://arxiv.org/abs/2604.13980

[7] Osman, N., Lembo, V., Bottegoni, G., et al. (2025). From In Silico to In Vitro: Evaluating Molecule Generative Models for Hit Generation. arXiv preprint arXiv:2512.22031. https://arxiv.org/abs/2512.22031

[8] Ouyang, M., Lan, H., & Lin, W. (2026). Adopting Reinforcement Learning with Verifiable Rewards for Molecular Generation. arXiv preprint arXiv:2607.19044. https://arxiv.org/abs/2607.19044

[9] Insilico Medicine. (July 2026). Insilico Initiates Phase III Clinical Trial for Rentosertib, Its AI-Empowered TNIK Inhibitor for Idiopathic Pulmonary Fibrosis. https://insilico.com/news/xmjsn4l091-insilico-initiates-phase-iii-clinical-tr

[10] Qin, J., Wessels, H.-H., Fernandez-Granda, C., et al. (2024). Active learning for efficient discovery of optimal gene combinations in the combinatorial perturbation space. arXiv preprint arXiv:2411.12010. https://arxiv.org/abs/2411.12010

第8章：トランスクリプトーム主導の創薬候補化合物探索

8.1 患者の代謝経路と制御構造を考慮した標的抗がん剤のバーチャルスクリーニング（VS）

寒天プレート上の液体培地は、最初は澄んだ赤色をしています。がん細胞を加えてから24時間以内に、その色は鮮やかな黄色へと変わります。これは乳酸が蓄積したことを示す化学的指標です。1920年代のドイツで、オットー・ワールブルクはこの色の変化を不思議に思いながら観察していました。培養器内には酸素が豊富に存在していました。酸素が存在するとき、正常な細胞は1分子のグルコースから30単位以上のエネルギー通貨であるATPを生成します。しかし、がん細胞はその効率的な酸化的リン酸化経路を迂回し、グルコース1分子あたりわずか2個のATPしか生成しない乳酸発酵へとシフトしていました。がん細胞は代謝効率を増殖速度と引き換えにしたのです。この現象はワールブルク効果として知られています。がん細胞は生体エネルギー効率を犠牲にして、バイオマス生成と細胞分裂を加速させているのです。

がん細胞が再構築するのは、グルコースの異化代謝だけにとどまりません。アミノ酸輸送、脂肪酸合成ライン、脂質膜の組み立て経路を再編しますが、この変化は代謝リプログラミングと呼ばれます。急速に分裂する細胞が必要とするのは、単なるATPエネルギーだけではありません。新たな娘細胞を合成するには、DNAのためのヌクレオチド、膜のためのリン脂質、タンパク質のためのアミノ酸が必要となります。発酵経路は産生するATP分子の数が少ない一方で、中間前駆体プールを豊富に満たします。がん細胞は発電所を縮小して原材料の倉庫を拡張しているのです。このことは、多くの腫瘍が窒素と炭素骨格の双方を供給するグルタミンに大きく依存している理由を説明しています。

再構築された代謝回路は、本質的な脆弱性をもたらします。特定の代謝酵素は正常な細胞では不要ですが、がん細胞の生存には不可欠となります。それらの特定の酵素を阻害することで、腫瘍を選択的に兵糧攻めにすることができます。その概念的論理はきわめて明快です。

臨床試験の実績は厳しい現実を突きつけています。IACS-010759などのミトコンドリア複合体I阻害剤やデビミスタットのような代謝調節薬は、2026年までに臨床試験で相次いで失敗を経験しました[1]。失敗の主な要因は代謝の不均一性です。大腸がん患者10人を調べれば、その代謝回路図は患者ごとに異なります。KRAS変異を持つ患者は野生型の患者とは異なる代謝依存性を示し、低酸素状態にある腫瘍中心部の細胞は血管が豊富な外周部に隣接する細胞とは異なる挙動をとります。周囲の免疫細胞、血管内皮細胞、がん関連線維芽細胞を含む腫瘍微小環境（TME）は、相互に連結した代謝生態系として機能しています。単一の代謝阻害剤がすべての患者に一律に効果を発揮することはありません。

ゲノムスケール代謝モデル（GEMs）は、この複雑さに対処します。都市の地下鉄路線図を思い浮かべてみてください。駅は代謝物を表し、線路は酵素が触媒する反応を表します。ヒトの代謝再構築モデルには、1万を超える個別の反応線路が含まれています。このネットワークにフラックスバランス解析（FBA）が組み合わされます。フラックスは、各線路を走る電車の運行量を表します。すべての代謝駅において総流入量と総流出量が一致することを求める定常状態の仮定のもとで、FBAは細胞のバイオマス産生を最大化するために必要な代謝フラックスの最適分布を計算し[2]、物理的なアッセイを行うことなく線形計画法の数式を解くだけで細胞の代謝状態を推定します。

基本的なネットワークマップはヒトの間で共通して保存されているものの、代謝の動態は個人によって異なります。患者の腫瘍生検サンプルのトランスクリプトームをシーケンシングすることで、特定の酵素をコードする遺伝子の相対的な発現レベルが明らかになります。インシリコでは、低発現の酵素を介した反応流束（フラックス）は制限され、発現が亢進（アップレギュレート）した酵素を介したフラックスは拡大されます。これを手作業で実行することは不可能です。1万件もの反応に対して同時に境界制約を設定する必要があり、極めて重要な反応をたった一つでも誤って制約すると、ネットワークモデル全体が停止してしまう可能性があるためです。単一の生検プロファイルから患者特異的な活性代謝マップを再構築するために、トランスクリプトームの発現量をフラックス制約へと変換するアルゴリズムが開発されました。

こうした個別化された代謝モデルにおいて単一遺伝子ノックアウトのシミュレーションを反復して実行することで、特定の患者の腫瘍にとってどの酵素ノードが致死的な脆弱性となるかを特定できます。2014年の研究では、さまざまながん細胞株を対象にこのアプローチが実証され、in vitroで腫瘍の増殖を阻害する代謝上の弱点を予測することに成功しました [3]。腫瘍微小環境も同じフレームワーク内でシミュレーションされます。血管新生が乏しく酸素が不足している腫瘍の中心部では、計算上の酸素取り込み速度を下げることで、がん細胞が代替となる嫌気的経路へとフラックスを迂回させる様子をモデル化します。同一の物理的腫瘍の内部であっても、低酸素状態の中心部にある細胞と、酸素が十分に行き届いた周辺部にある細胞とでは、異なる治療的介入が必要となります。

純粋に計算によるフラックスモデリングに内在する限界も明らかになりました。FBAは、細胞内の代謝産物濃度が静的な定常状態の平衡にあると仮定します。しかし、生きた細胞は動的です。酵素はアロステリック調節を受け、転写産物の量が活性を持つタンパク質レベルへと線形に変換されるわけではありません。こうしたギャップを埋めるためにハイブリッド手法が登場しました。2026年、トリプルネガティブ乳がんを研究する研究者らは、代謝調整の最小化（MOMA: Minimization of Metabolic Adjustment）から得られたフラックス値をランダムフォレスト分類器の数値特徴量へと変換しました。制約ベースのFBAのみに依存した場合、必須遺伝子の同定

における感度は0.37にとどまりましたが、機械学習（ML）を組み込むことで感度は0.55へと向上しました [4]。物理的なフラックス方程式が基礎となる事前情報を形成し、教師あり学習モデルが経験的なギャップを埋めたのです。

がんの薬理学における主要な障壁の一つが、代謝の可塑性です。経路Aを阻害すると、腫瘍は迂回路となる経路Bを活性化させてしまいます。治療効果を持続させるには、2つの標的を同時に阻害する必要があります。合成致死（Synthetic Lethality）とは、遺伝子Aまたは遺伝子Bのいずれか一方のみに摂動を与えても細胞は生存しますが、両方に同時に摂動を与えると死滅する遺伝子相互作用を指します。2本の橋がかかった川を思い浮かべてください。片方を破壊しても交通は迂回できますが、両方を破壊すれば通行は完全に遮断されます。これが、PARP阻害剤がBRCA欠損の卵巣がんや乳がんを選択的に死滅させる理由です。2026年には、代謝ネットワークと転写制御回路を併せて評価する計算フレームワークによって、複数遺伝子の合成致死および合成用量致死の組み合わせが特定されました [5]。

2026年6月、クイーンズ大学ベルファストの研究者らは、マルチスケール自律型科学発見エンジンであるOctopusを発表しました[6]。特化型言語モデルエージェントの群れが生物学的仮説を生成し、それらは物理法則に基づく代謝流束モデルおよび実験に基づくCRISPR依存性データセットと照合されて即座に検証されます。テキスト上では説得力があるように見えても生物物理学的制約に反する仮説は、計算の過程で刈り込まれます。大腸がんのトランスクリプトームに対して先入観を持たずに適用された際、同システムはインスリン様成長因子2（IGF2）が5-フルオロウラシル（5-FU）化学療法への耐性を媒介する主要なドライバーであることを特定しました。

検証指標は厳密でした。偽発見率は0.0292であり、IGF2高発現群と低発現群の患者コホート間で全生存期間を分けるログランク検定のp値は0.0007でした。治療を受けたマウスコホートにおける生体内（in vivo）の腫瘍体積減少は、0.0373のp値を達成しました[6]。AIエージェントによって自律的に生成された計算上の仮説が、生きた動物モデルにおいて実証されたのです。

その3か月前、同研究室は大腸がんを補完的な角度から検討していました。ニューロシンボリック・ワールドモデルを構築し、遺伝的復帰をインシリコでシミュレーションしたのです[7]。83名の患者プロファイル全体において、トランスクリプトーム予測の相関は0.268に達しました。得られた生物学的知見は驚くべきものでした。変異型KRASが5-FUの毒性に対する予期せぬ緩衝材として機能し、計算上でPIK3CA変異を野生型へと戻すと5-FU耐性が高まりました。代償的なシグナル伝達が細胞生存経路を回復させたため、遺伝子変異を修復したことでかえって薬物応答が悪化したのです。

私はこの研究の方向性を支持します。相関係数0.268はベンチマークスコアとしては控えめであり、83名の患者プロファイルも小さなサンプルです。それでもこの研究が価値を持つのは、反証可能でメカニズムに基づいた生物学的仮説を自然言語で明確に提示しているからです。生物学的メカニズムを説明することなく高い精度を達成するブラックボックス型の予測モデルは、腫瘍内科医の臨床処方を導くことができません。根底にある因果経路を説明するモデルこそが、的を絞った実験室での検証を動機づけます。2026年8月に発表されたMetaboLLMは、代謝産物の相互作用グラフを再構築するために、メタボロミクスの知識を用いて言語モデルをファインチューニングしました[8]。記号計算と言語モデリングの融合が進んでいます。

正常組織に不可欠な経路を保護しながら、がん細胞が依存する代謝経路を選択的に遮断すること——これがAI主導のがんメタボロミクスが目指す目的の要約です。この選択性は全身性の臨床毒性を最小限に抑えます。精密な標的化によって化合物の投与量を低減できるため、肝機能や心機能を保護することができます。

しかし、シミュレーションされた代謝ネットワークは、生きたヒトの生理機能と完全に同一ではありません。予測された脆弱性が臨床的有効性につながるかどうかは、細胞培養や動物モデルを通じた検証を必要とします。Octopusのチームがマウスを用いたin vivo試験を実施したのはそのためです。この「計算から実験へ」の橋渡しを加速させることが、今後数年間の中心的な課題となります。続く節では、代謝流束方程式を計算するのではなく、病的な遺伝子発現シグネチャーを直接逆転させる分子を設計するという補完的なアプローチを検討します。

参考文献

[1] Frontiers in Immunology. (2026). Cancer metabolism: from the Warburg effect to precision therapy. https://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2026.1793553/full

[2] Orth, J. D., Thiele, I., & Palsson, B. Ø. (2010). What is flux balance analysis? Nature Biotechnology, 28(3), 245-248. https://doi.org/10.1038/nbt.1614

[3] Yizhak, K., Gaude, E., Le Dévédec, S., et al. (2014). Phenotype-based cell-specific metabolic modeling reveals metabolic liabilities of cancer. Nature Communications, 5, 4349. https://doi.org/10.1038/ncomms5349

[4] Kim, B. K., et al. (2026). Integrating Genome-Scale Metabolic Modeling with Machine Learning Improves Gene Essentiality Prediction in Triple-Negative Breast Cancer. International Journal of Molecular Sciences. https://doi.org/10.3390/ijms27115059

[5] Barrena, N., et al. (2025). Beyond synthetic lethality in large-scale metabolic and regulatory network models via genetic minimal intervention set. Bioinformatics Advances, 6(1), vbaf319. https://academic.oup.com/bioinformaticsadvances/article/6/1/vbaf319/8384237

[6] Baker, C., Ren, T., Rafferty, K., Wang, H., & McDade, S. (2026). Autonomous mechanistic discovery of colorectal cancer vulnerabilities via multi-scale AI swarms. arXiv preprint arXiv:2607.16262. https://arxiv.org/abs/2607.16262

[7] Baker, C., Ren, T., Rafferty, K., & Wang, H. (2026). Contextual Invertible World Models: A Neuro-Symbolic Agentic Framework for Colorectal Cancer Drug Response. arXiv preprint arXiv:2603.02274. https://arxiv.org/abs/2603.02274

[8] Ku, D., Kwak, M. G., Pasquel, F. J., & Li, J. (2026). MetaboLLM: a metabolomics-specialized large language model for biochemical knowledge integration and predictive metabolite graph construction. arXiv preprint

arXiv:2608.06253. https://arxiv.org/abs/2608.06253

[9] Hanahan, D. (2022). Hallmarks of cancer: new dimensions. Cancer Discovery, 12(1), 31-46. https://doi.org/10.1158/2159-8290.CD-21-1059

8.2 大規模遺伝子発現（トランスクリプトーム）データを用いた低分子化合物の逆設計

図書館で1冊の本が誤った場所に戻されたとしても、その本が見失われるだけで、蔵書全体の機能は保たれます。従来の創薬は歴史的に、その見失われた1冊の本を探し出すように進められてきました。すなわち、特定の病因タンパク質を同定し、その結合ポケットに適合する分子を作り出すアプローチです。この標的中心型のアプローチは数々のブレイクスルーをもたらしました。しかし、アルツハイマー病、関節リウマチ、あるいは臓器線維症などに適用した場合、標的ベースの創薬はしばしば行き詰まりを見せました。これらの複雑な病態は、単に1冊の本の場所がずれているのではなく、本棚が20架まるごと倒れて乱雑になっているような状態を表しているからです。

こうした現実を背景として、表現型創薬（Phenotypic Drug Discovery）が登場しました。これは、どの単一タンパク質を阻害すべきかを問うのではなく、細胞のどのような全体的状態を回復させるべきかを問いかけるアプローチです。トランスクリプトームは細胞状態の読み取り値を提供します。すなわち、細胞内のすべての遺伝子の活性を追跡する発現ベクトルです。疾患組織と健常組織のトランスクリプトームベクトルを比較することで、表現型の乖離が明らかになります。その目標は、その疾患シグネチャーを反転させ、健常なベースラインへと戻す低分子化合物を発見することです。

2006年に発表されたブロード研究所（Broad Institute）のConnectivity Map（CMap）は、164種類の薬剤摂動に対する遺伝子発現応答を体系化することで、この概念を実証しました[1]。その基本ロジックは明快でした。薬剤によって誘導される発現シグネチャーが、疾患シグネチャーの数学的な反転となる分子を選択することです。2017年には、L1000プラットフォームによってこのリポジトリが100万件以上のプロファイルへと拡張されました[2]。ヒトの全2万遺伝子を測定する代わりに、L1000は978個の代表的なランドマーク遺伝子をアッセイし、計算による補完を通じて残りのトランスクリプトームを推論することで、プロファイルあたりのコストを100分の1に削減しました。2025年には、L2S2検索エンジンが167万8,000件の化合物およびCRISPRノックアウトシグネチャーを統合検索インターフェースへと統合しました[3]。

測定された978個のマーカーから2万個の遺伝子を補完・推論できるのは、遺伝子制御のモジュール性に基づいています。遺伝子は協調して活性化および不活性化する共制御ネットワークの中で機能しているためです。図書館でどの閲覧席が利用されているかを観察すれば、どの学術分野が研究されているかが分かります。少数の代表的な貸出記録から、図書館全体の利用動態を再構

成できるのと同じです。補完された値は統計的な近似値ですが、この妥協策によって、網羅的に測定するにはコストが高すぎた何百万件もの摂動に対するスクリーニングが可能になりました。

このリソースはドラッグリポジショニングを可能にしました。ブロード研究所の研究者らは、がん細胞株全体で4,518種類の非抗がん薬をスクリーニングし、がん細胞を選択的に死滅させる複数の糖尿病治療薬、抗生物質、抗炎症薬を同定しました[10]。既存の分子を転用することは、すでに確立された臨床上の安全性データを活用できるため、初期段階の毒性評価を回避することができます。

リポジショニングは既存の化合物ライブラリからの選択にとどまります。必要な分子が存在しない場合には、de novoで設計しなければなりません。トランスクリプトームに基づく逆設計はこれを実現します。目標とする遺伝子発現ベクトルを明示的な入力条件として与え、相補的な化学構造をゼロから組み立てるのです。既知の化学リガンドを模倣するようにモデルを学習させるのではなく、指定された生物学的効果をもたらす化学的要因を生成するタスクをモデルに課すのです。

条件付き拡散モデルが、この逆設計を実行します。拡散モデルは、初期のガウス分布から段階的にノイズを除去していくことで構造を生成します。分子生成において、ノイズはランダム化された原子座標や結合グラフに相当します。このノイズ除去の軌道を望ましいトランスクリプトーム変化で条件付けることにより、標的とする生物学的表現型を誘導するケミカルスペースへと生成を導きます。2026年5月に発表されたCUREは、処理前後の対となるトランスクリプトームから目標指向型の摂動の埋め込みを抽出し、拡散モデルを誘導します[4]。これにより、解明された3D標的タンパク質の座標を必要とせずに生理活性分子を生成する、生成的逆問題として創薬設計を位置づけています。

2026年7月に発表されたJoPMolは、生物学的有効性と化学的実現可能性のバランスを取るために、結合多目的条件付け[5]を導入しました。生物学的なトランスクリプトームのベクトルのみで条件付けると、化学的に不安定で合成不可能な構造が生み出されます。一方で、化学的ルールのみで条件付けると、合成は可能でも生物学的に不活性な分子が生成されます。これら双方の制約を同時に両立させることが、表現型創薬における実用上の課題です。

合成可能性は常に立ちふさがる障壁です。ソフトウェア上ではいかなる原子配置も描くことができますが、実際の化学合成には、安定した反応中間体を伴う実現可能な多段階の反応経路が必要となります。最新のアーキテクチャは、合成容易性（SA）スコアや医薬品らしさの定量的推定（QED）指標を生成損失関数に直接組み込み、生成サンプリングを合成可能なケミカルスペース内に制約しています。

逆設計には高精度な順方向応答予測モデル、すなわち特定の化学構造に曝露された際に細胞のトランスクリプトームがどのように変化するかを予測するモデルが必要です。2026年5月に発表され、240万件の1細胞マウス胚プロファイルで事前学習された細胞ワールドモデルであるChreodeは、Norman perturb-seqデータセットにおいて摂動予測誤差を0.2121から0.1858へと低減させ、12.4%の性能向上を達成しました[6]。また、2026年3月に発表された検索拡張型アーキテクチャは、予測を生成する前に過去の類似した摂動アッセイをメモリから検索する仕組みを備えており、適合する細胞型のコンテキストで条件付けることが予測精度に不可欠であることを実証しました[7]。

従来の単一標的を対象とする薬理学パラダイムは転換期を迎えています。2026年7月に発表された総説論文では、ポリファーマコロジーと多標的創薬設計について論じられており、がんや神経変性疾患においては、経路の冗長性や代償性耐性によって単一標的阻害薬が失敗することが頻繁にあると指摘されています[8]。トランスクリプトームで条件付けられた生成設計は、目的の表現型が1つのタンパク質の調節によって達成されるか、あるいは5つのタンパク質の調節によって達成されるかにかかわらず、包括的な表現型のエンドポイントに向けて最適化を行うため、多標的ポリファーマコロジーに極めて自然に対応します。

疾患の発現シグネチャーを反転させるアプローチには注意すべき点があります。疾患細胞において発現が上昇または低下したすべての遺伝子を元に戻したとしても、それが治療的な回復を保証するわけではありません。トランスクリプトームの変化には病態を直接駆動するものがある一方で、代償的な生体防御反応を表しているものもあるためです。機能している生体防御機構を抑制してしまうと、感染症の治療において無差別に熱を下げることでかえって病状を悪化させるのと同様に、疾患の転帰を悪化させてしまいます。主要な原因因子と二次的な生体応答とを明確に区別することが、依然として不可欠です。

データの限界も存在します。L1000プロファイルはプラスチック製ディッシュ上の不死化単層培養細胞株に由来しており、生体内の完全なヒト肝臓組織や脳組織とは異なります。ランドマークアッセイは978個の遺伝子を測定して残りの遺伝子発現量を補完するものであり、バルクシーケンシングは何百万もの細胞集団全体の発現量を平均化してしまいます。その結果、ハイコンテンツ細胞イメージングを統合する研究が進められています。6種類の蛍光染色にわたる115テラバイトのJUMP Cell Paintingデータセット[9]は、トランスクリプトミクスデータを補完する形態学的表現型解析を提供しています。

バルクシーケンシングによる平均化は、細胞の不均一性を覆い隠してしまいます。腫瘍生検組織には悪性上皮細胞、浸潤免疫リンパ球、間質線維芽細胞が含まれていますが、これらを破砕してバルクライセートとして均質化すると、個々の細胞型特有の応答がかき消されてしまいます。1

細胞トランスクリプトミクスは、個々の細胞を別々にプロファイリングし、治療的摂動に対する細胞型特異的な応答を追跡することで、この課題を解決します。

トランスクリプトミクス逆設計の利点は、その生物学的な柔軟性にあります。疾患標的の3次元結晶構造が未知であっても、細胞の表現型の変化に基づいて設計を進めることができます。これにより、根本的な疾患の病因が部分的にしか解明されていない段階であっても、創薬を開始することが可能になります。アスピリンをはじめとする歴史的な治療薬も同様に機能していました。シクロオキシゲナーゼ阻害というメカニズムが解明される前の何世紀もの間、ヤナギの樹皮の煎じ液は痛みを和らげてきました。病的な細胞状態を逆転させる分子の設計は、そうした経験的な伝統を現代の計算ツールを用いて応用するものです。

参考文献

[1] Lamb, J., Crawford, E. D., Peck, D., et al. (2006). The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science, 313(5795), 1929-1935. https://doi.org/10.1126/science.1132939

[2] Subramanian, A., Narayan, R., Corsello, S. M., et al. (2017). A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell, 171(6), 1437-1452. https://doi.org/10.1016/j.cell.2017.10.049

[3] Marino, G. B., et al. (2025). L2S2: chemical perturbation and CRISPR KO LINCS L1000 signature search engine. Nucleic Acids Research, 53(W1), W338. https://academic.oup.com/nar/article/53/W1/W338/8123452

[4] Xu, Z., Zhang, Z., Wang, L., Liu, Z., Liu, Q., Wu, S., & Wang, L. (2026). Reading the Cell, Designing the Cure: Perturbation-Conditioned Molecular Diffusion for Function-Oriented Drug Design. arXiv preprint arXiv:2605.15243. https://arxiv.org/abs/2605.15243

[5] Yuan, H., Li, C., Ma, W., Murata, T., & Jiang, Y. (2026). Gene Expression-Informed Jointly Controlled Generative Modeling for Precision Molecular Design. arXiv preprint arXiv:2607.11978. https://arxiv.org/abs/2607.11978

[6] Qiu, M., Zheng, G., Xu, Y., Zhang, R., Ding, Y., Long, Q., & Chen, T. (2026). Chreode: A Cell World Model for One-Step Temporal Dynamics and Perturbation Prediction. arXiv preprint arXiv:2605.28111. https://arxiv.org/abs/2605.28111

[7] Di Francesco, A. G., Rubbi, A., & Liò, P. (2026). Retrieval-Augmented Generation for Predicting Cellular Responses to Gene Perturbation. arXiv preprint arXiv:2603.07233. https://arxiv.org/abs/2603.07233

[8] Han, T., Pan, Z., Ge, W., & Zhao, Q. (2026). Beyond SBDD: Geometric Deep Learning in Polypharmacology and Multi-target Drug Design. arXiv preprint arXiv:2607.20550. https://arxiv.org/abs/2607.20550

[9] Muñoz, A. F., et al. (2026). JUMP-lite: Compact, reproducible benchmarking of cell representations. arXiv preprint arXiv:2608.07632. https://arxiv.org/abs/2608.07632

[10] Corsello, S. M., Nagari, R. T., Spangler, R. D., et al. (2020). Discovering the anticancer potential of non-oncology drugs by systematic viability profiling. Nature Cancer, 1(2), 235-248. https://doi.org/10.1038/s43018-019-0018-6

8.3 副作用を軽減する薬剤再設計を導く先例誘導型コ・サイエンティスト・アーキテクチャ

第1相臨床試験病棟のテレメトリーモニター上を、心電図の波形が流れていきます。ある瞬間、特定の区間が伸びます。QT間隔です。治験担当医の手が止まります。そのまま放置すれば、QT延長は致死的な心室性不整脈であるトルサード・ド・ポアンツ（Torsades de Pointes）を引き起こしかねません。創薬において最も大きな代償を伴う結末は、分子が標的に結合しないことではなく、効果的に結合しながらも深刻なオフターゲット毒性を引き起こすことなのです。

臨床試験に進む創薬候補化合物の約10個中9個は薬事承認を得るに至らず、開発中止の大きな割合を予期せぬ毒性が占めています。低分子化合物が心筋細胞のhERGカリウムチャネルを阻害すると、心臓の再分極が遅延してQT間隔が延長し、不整脈が誘発されます。hERGチャネルのポア（細孔）腔は疎水性で広々としており、医薬品化学ライブラリによく見られるモチーフである塩基性アミンを含む親油性骨格を捉えやすいため、頻繁にオフターゲットのリスクをもたらします。化学者が好むケモタイプは、往々にしてhERGチャネルが結合するものと一致してしまうのです。

肝臓もまた、それに匹敵する毒性学的リスクを抱えています。肝細胞内のシトクロムP450（CYP）酵素は生体異物を代謝します。ある薬剤がCYPアイソフォームを阻害すると、併用薬が体外へ排出されなくなり、毒性を示す血漿中濃度にまで蓄積してしまいます。あるいは、肝代謝によって不活性な親化合物が反応性のキノン中間体やエポキシド中間体に変換され、細胞内タンパク質を共有結合で修飾してしまうこともあります。薬物性肝障害（DILI）は、製薬企業に毎年数億ドルもの損失をもたらしています。重篤な特異体質性DILIは1万人に1人という低い頻度でしか発生しないため、初期の臨床試験では見逃されやすく、また前臨床の動物毒性モデルはヒトにDILIを引き起こす化合物の約半分を見落としてしまいます[1]。

こうした安全上のリスクを計算によって予測する技術は進歩してきました。hERG阻害予測において、分子フィンガープリントを用いたアンサンブルモデルは交差検証で84.9%の精度と0.887のAUROCを達成しましたが、分布外（OOD）の化合物セットで評価するとAUROCは0.786へと低下しました[1]。肝毒性に関しては、300種類の薬剤に対するヒト細胞の応答をプロファイリングして2025年に公開されたトキシコゲノミクスリソースが、特異度を100%に保った状態で88%の感度でDILIを検出しました[2]。これらの分類器は危険な分子を特定できるものの、受動的なフィルターとして機能するにすぎません。つまり、有効性を維持しながら毒性を軽減するための具体的な化学的修飾を提示することなく、分子にリスクのフラグを立てるだけなのです。

2026年7月3日にEugene Kim氏とChamgil Hong氏によって発表され、ICML 2026 AI for Science Workshopに採択されたPRECEDEは、この限界を克服するものです[3]。「先例誘導型コ・サイエンティスト（precedent-guided co-scientist）」というその名の通り、これは確立された親薬剤を受け取り、標的への結合親和性を維持しつつ、指定された副作用を排除する構造修飾を提案します。ゼロからのde novo合成ではなく、標的を絞った化学的再設計を実行するのです。

その中核となる基盤は、確立されたエビデンスに根ざしている点にあります。PRECEDEは、薬物と有害事象の関連性、バイオメディカル知識グラフ、そして構造修飾によって安全性の懸念が解決された過去の創薬化学の事例研究を横断して推論を行います。大規模言語モデル（LLM）は、明確な生物物理学的ルールとヒューマンインザループの検証チェックポイントのもとで、この知識検索を統括します。著者らは、生成された仮説が監査可能であり、反証可能であり、かつ確立された薬理学的原理に基づいたものでなければならないと強調し[3]、指針のない無秩序な生成探索を制限しています。

バイオメディカル知識グラフは、医薬品、標的タンパク質、有害な表現型、代謝経路を、相互に接続されたノードと関係性を示すエッジとして結びつけます。研究者はこのネットワークにクエリを実行することで、薬物による酵素阻害から下流経路の摂動、さらには臨床症状の発現に至るまでの多段階のメカニズム経路を追跡できます。単一のリスク確率を出力するだけのブラックボックス型の分類器とは異なり、知識グラフは自然言語によるメカニズムの説明を提示します。

構造再設計のワークフローは、創薬化学の工房のように機能します。生成コンポーネントが修飾された化学構造を提案し、オーケストレーターが知識グラフの経路と照合して危険なモチーフにフラグを立てます。オフターゲットの懸念が特定されると、創薬化学者がかつて同様の懸念を解決した先例を検索し、バイオアイソスター（生物学的等価体）による置換を適用します[4]。たとえば、電子豊富な芳香環が酸化されて肝細胞タンパク質と結合する反応性キノンイミンへと変化する場合、フッ素原子を導入して代謝酸化をブロックします。また、脆弱な環における代謝の懸念には、オキセタン環に置換することで対処します。バイオアイソスターは、立体的な体積や電子的性質を保ちつつ、体内での代謝運命を変化させます。こうして、重要なファーマコフォア領域は保存され、トキシコフォアのみが修飾されます。

関連するアーキテクチャも並行して進歩しました。51名の研究者によって執筆されたGoogleのCo-Scientistシステムは、2026年に『Nature』誌で発表されました[5]。このシステムは専門化されたマルチエージェント集団（スワーム）を展開しており、生成エージェントが科学的仮説を提案し、批評エージェントが査読シミュレーションを通じてそれに疑問を投げかけ、監督エージェントが候補の選定を仲介します。医薬品再配置（ドラッグリパーパシング）や薬剤耐性メカニ

ズムの解明において検証が行われ、急性骨髄性白血病に関する予測はin vitroで実証されました。また、Latent Labsが開発したLatent-Yは抗体設計プロジェクトを自律的に遂行し、9つの標的のうち6つにおいて1桁のナノモーラー水準の親和性で実験的な結合を達成しました。これは人間の専門家による作業日程の56倍の速さです[6]。

生成速度が向上するにつれて、安全性のベンチマークが不可欠になります。2026年7月に発表されたMolSafeEvalは、分子生成モデルが危険な、あるいは反応性や毒性を持つケモタイプ（化学型）を生成していないかを4つのタスク区分にわたって評価します[7]。これは、従来のベンチマークが新規性や結合親和性を評価する一方で、分子の安全性リスクを見過ごしていたという限界に対処するものです。

臨床パイプラインの統計は、この状況を反映しています。ASCO 2026で発表された分析によると、63社においてAIが設計した117の分子がヒト臨床試験に進んでおり、そのうち60件（51.3%）が第1相試験を完了し、8件（6.8%）が第2相試験を完了しました[8]。第1相試験の通過率は80%から90%の間に達し、従来の基準値であった50%を上回りました。しかし、第2相試験の成功率は従来の40%というベンチマーク付近にとどまっています[8]。AIが設計した低分子化合物でFDAの最終承認を取得したものはまだ存在せず、現在はrentosertibが第3相試験で評価を進めている段階です。

これらの指標は、第1相試験で評価される低分子化合物の薬物動態や安全性プロファイルの最適化において、人工知能が優れていることを示しています。第2相試験は治療効果、すなわち選択した標的を調節することが患者の生存率を改善するかどうかを評価します。根底にある生物学的な標的仮説に欠陥があれば、どれほど化学的な改良を施しても臨床的な有効性を生み出すことはできません。

第2相試験での開発中止は、化学的な問題というよりも根本的には生物学的な障壁によるものです。特定のキナーゼを阻害することで疾患の病態が反転するという仮説に誤りがあれば、どれほどそのキナーゼに結合するよう分子を最適化しても治療効果は得られません。第8.1節で述べた代謝フラックスモデルや、第8.2節で詳述したトランスクリプトームシグネチャーが価値を持つのは、それらが標的の選定を導くからです。標的探索パイプラインを化学的最適化エンジンへと直接接続することが、今後の第2相試験の成功率を左右することになります。

PRECEDEで設計されているように、人間のレビューによるチェックポイントを設けることは、依然として適切な設計上の選択です。科学的なワークフローには、バイオアイソスター置換の根拠となった過去の事例を特定し、それを裏付ける文献や実験データの対照表を参照できるよう

な、監査可能な記録が不可欠です。第2相試験での臨床的な脱落を減らすためには、特定の化学修飾がなぜ行われたのかを説明できるシステムが求められます。

第1相試験のテレメトリーモニターに話を戻すと、QT間隔が延長した際に臨床医が必要としているのは、解釈不能な確率スコアではありません。どの化学的部分構造がhERGポアと結合したのか、過去のどの薬剤がそのリスクに遭遇したのか、そして主標的への有効性を保ちながら、どのバイオアイソスター置換によってオフターゲット相互作用を排除したのかを特定する因果関係の記録です。創薬におけるAIの永続的な貢献は、そうした監査可能な記録を確立することにあります。

参考文献

[1] Anand Phani Kumar, K. (2025). Recent advances in AI-based toxicity prediction for drug discovery. Frontiers in Chemistry. https://doi.org/10.3389/fchem.2025.1632046

[2] Bergen, V., et al. (2025). A large-scale human toxicogenomics resource for drug-induced liver injury prediction. Nature Communications. https://www.nature.com/articles/s41467-025-65690-3

[3] Kim, Y., & Hong, C. (2026). A Precedent-Guided Co-Scientist for Side-Effect-Aware Drug Redesign. arXiv preprint arXiv:2607.02944. https://arxiv.org/abs/2607.02944

[4] Meanwell, N. A. (2011). Synopsis of some recent tactical application of bioisosteres in drug design. Journal of Medicinal Chemistry, 54(8), 2529-2591. https://doi.org/10.1021/jm1013693

[5] Gottweis, J., Weng, W.-H., Daryin, A., Tu, T., Sirkovic, P., et al. (2026). Accelerating scientific discovery with Co-Scientist. Nature. https://www.nature.com/articles/s41586-026-10644-y

[6] Latent Labs Team (Schmon, S. M., Pretorius, D., Mathis, S., et al.). (2026). Latent-Y: A Lab-Validated Autonomous Agent for De Novo Drug Design. arXiv preprint arXiv:2603.29727. https://arxiv.org/abs/2603.29727

[7] Xu, T., Cao, X., Zhu, Z., Ding, K., & Chen, H. (2026). MolSafeEval: A Benchmark for Uncovering Safety Risks in AI-Generated Molecules. arXiv preprint arXiv:2607.00464. https://arxiv.org/abs/2607.00464

[8] Khairil, L., et al. (2026). AI in Drug Discovery: Clinical Failures, Regulatory Reality, and the Validation Crisis Behind the Hype. Pharmaceuticals, 19(6), 916. https://www.mdpi.com/1424-8247/19/6/916

[9] Redfern, W. S., Carlsson, L., Davis, A. S., et al. (2003). Relationships between preclinical cardiac electrophysiology, clinical QT interval prolongation and torsade de pointes for a broad range of drugs. Cardiovascular Research, 58(1), 32-45. https://doi.org/10.1016/S0008-6363(02)00846-5

第9章：1細胞データの前処理と埋め込み技術

9.1 1細胞トランスクリプトミクス（scRNA-seq）とマルチオミクス多様体アラインメント

ジューススタンドでフルーツミキサーが回転しています。イチゴ、バナナ、ホウレンソウが一緒に粉砕され、均一な緑色の飲み物になります。このブレンドされたミックスジュースをどれほど綿密に分析したとしても、個々のイチゴ本来の純粋な風味を取り戻すことはできません。2010年代以前のトランスクリプトームプロファイリングも、これとまったく同じように行われていました。バルクRNAシーケンシング（bulk RNA-seq）として知られるこの手法では、研究者が数百万個の不均一な細胞を含む組織サンプル全体をすり潰し、総mRNAを抽出して、検体全体における平均化された総合的な発現プロファイルを測定していました。その欠点は、数学的な平均値が希少な細胞亜集団を覆い隠してしまうことです。固形腫瘍の1%未満を占めるがん幹細胞や、稀な薬剤耐性残存細胞、あるいは免疫抑制を司る制御性T細胞などは、残り99%の細胞の発現シグナルにかき消されてしまいました。たとえその1%こそが腫瘍の再発を引き起こす亜集団であったとしても、バルク解析による平均化はその存在を覆い隠してしまったのです。疾患の生物学的な駆動要因は、統計的平均の外側に存在します。

1細胞RNAシーケンシング（scRNA-seq）はミキサーを引退させ、細胞を1個ずつ単離して分析することを可能にしました。地下鉄の改札機が通勤客を1人ずつ通すように、マイクロ流体デバイスは個々の細胞を一列に流します。人の髪の毛よりも細いマイクロ流体流路の内部で、個々の細胞は油で覆われたナノリットル単位の液滴（ドロップレット）の中に封入されます。各液滴は、単一の細胞のための独立した反応容器として機能します。液滴の中に封入されるのは細胞だけではありません。各液滴には、バーコード化された数百万本もの短いDNAオリゴヌクレオチドで修飾された微小なハイドロゲルビーズも取り込まれます。液滴の内部で細胞が溶解すると、これらのオリゴヌクレオチドが細胞内のポリアデニル化されたmRNA転写産物を捕捉し、2つの分子タグを付加します。1つは細胞バーコード（そのビーズ上のすべてのオリゴヌクレオチドで共有され、一括シーケンシングを行った後に転写産物を計算によって元の細胞へと割り当てることを保証します）、もう1つは分子バーコード（UMI：Unique Molecular Identifier）です。UMIは、図書館の個々の蔵書に割り当てられる受入番号のように機能します。mRNA分子はシーケンシングの前にPCRによって増幅されるため、UMIがあることで、PCRによる複製コピーが別個の生物学的転写産物として誤って重複カウントされるのを防ぐことができます。このワークフローには品質管理のためのフィルタリングが必要です。2個の細胞を取り込んだ液滴は人工的なダブレットを生み出し、溶解した細胞から漏れ出た環境RNAは周囲の液滴を汚染し、死滅しかけている細胞ではミトコンドリア由来の転写産物の割合が高くなるからです。実験台の上は静かです。マイクロ流体チップが毎秒数千個もの液滴を生成している間、技術者はただ試薬チューブを移送するだけで済みます。

シーケンシングのスループットは急速に拡大しました。2015年、Drop-seqは4万4,808個の網膜細胞を同時にプロファイリングし、39種類の異なる細胞型を同定しました[1]。同じ学術誌の号において、inDropも同様の液滴マイクロ流体技術を用いて胚性幹細胞のプロファイリングを行いました[2]。9年後の2024年10月、10x GenomicsはGEM-Xアーキテクチャを発表し、1回のランで最大256万個の単一細胞を処理可能にするとともに、ライブラリ調製コストを1細胞あたり1セント未満へと引き下げました[3]。個々の細胞における2万個の遺伝子プロファイリングがチューインガム1パックよりも安価になり、実験のキャパシティは数万個の細胞規模から数百万個の規模へと拡大しました。

生成されるデータセットは、1細胞あたり2万遺伝子次元の行列となります。人間の視覚は2万次元の空間を認識できないため、低次元への埋め込みが1細胞インフォマティクスの中核を担っています。埋め込みは、2万個の遺伝子発現カウントを、細胞のアイデンティティを捉える数十の潜在座標へと圧縮します。これは、3次元の地球儀を2次元の世界地図に投影することに似ています。投影による歪みを完全に排除することはできませんが、機能的に関連する細胞同士は近くに集まる必要があります。埋め込み多様体において、類似した細胞状態は島のように集まり、幹細胞から成熟細胞への分化の軌跡は海岸線のように現れます。解析者のモニター上では、散布図の個々の点がそれぞれ1つの細胞を表しています。前処理手法によって、この空間的投影が生物学的な真理（グラウンドトゥルース）を確実に反映するよう担保されます。

データセットの規模が拡大するにつれて、技術的なバッチ効果が主要な課題として浮上しました。バッチ効果とは、実験の実施回（ラン）の間に生じる、生物学的要因ではない系統的かつ技術的なばらつきを指します。まったく同じ血液サンプルを処理した場合でも、試薬ロット、シーケンサー（シーケンシング機器）、臨床検体の採取施設、あるいは処理日が異なれば、名目上の発現カウントは変化してしまいます。これは、同じ音楽の録音を異なるスピーカーで再生するとそれぞれ異なる音の特性が生じるのと同じです。楽曲自体は変わりませんが、音響出力が変化するのです。バッチ効果を補正しないまま放置すると、アルゴリズムはソウルで採取されたT細胞と釜山で採取されたT細胞を別の細胞型として分類してしまい、細胞オントロジーではなく病院の施設ごとに埋め込み多様体を分断してしまいます。Mutual Nearest Neighbors（MNN）は、異なるバッチ間でお互いを最近傍と認識し合う細胞ペアを特定し、その間の並進ベクトルを計算することでこの問題を補正します[4]。Harmonyは細胞を低次元空間に埋め込み、クラスタリングと混合モデルによる線形補正を交互に繰り返すことで、一般的なノートパソコン上でも数十万個の細胞を数分で整列（アライメント）させます[5]。scVIは深層生成モデルを採用し、バッチ識別子を条件付けした変分自己符号化器（VAE）を用いて、生物学的な変動と技術的ノイズを直接分離します[6]。バッチ補正には慎重さが求められます。技術的なアーティファクトを除去

するあまり、疾患患者と健常対照群を区別する真の生物学的表現型まで消し去ってしまってはならないからです。

マルチモーダルアッセイの拡大は、マルチオミクスのアライメントという新たな課題をもたらしました。アッセイごとに測定する生物学的モダリティは異なります。scRNA-seqは遺伝子発現をプロファイリングし、scATAC-seqはオープンクロマチンのアクセス可能性を測定し、CITE-seqは抗体由来タグを介して細胞表面タンパク質のエピトープをカウントします[7]。そして、それぞれが異なるスケール分布とゼロスパース構造を示します。これは、異なる言語で印刷された3枚の地図帳を重ね合わせるようなものです。LIGERなどの行列分解ツールは、マルチモーダルデータを共有因子とデータセット固有の因子に分解し、共通の座標を抽出します[8]。最適輸送（OT）フレームワークはこれを数学的にモデル化し、ある1細胞の経験的確率分布を別の分布へと移動させる輸送コストを最小化するマッピング計画を計算します。動的最適輸送モデルは、静的な時間スナップショットをつなぎ合わせ、連続的な発生・分化の軌跡を再構築します[9]。

こうした前処理フレームワークによって、ヒト細胞アトラス（Human Cell Atlas：HCA）イニシアチブの実現が可能になりました。2016年に設立されたこの国際コンソーシアムには、約100か国から3,600人以上の研究者が参加しています。2026年までに、同プロジェクトは450件の発表済み研究を通じて、約1万人の組織ドナーから1億個以上の単一細胞のプロファイリングを実施し、肺、神経、心臓、免疫系など18の生物学的ネットワークにまたがる初期のリファレンスドラフトを構築しました [10]。その後のマイルストーンでは、包括的な臓器アトラスを構築するために、人口動態的に多様な人類集団にわたって数十億個の単一細胞をプロファイリングすることを目指しています。10年間にわたり数百もの研究室で生成されたデータセットを統一された参照座標系へと統合することは、バッチ補正およびマルチオミクスのアライメントアルゴリズムに直接依存していました。かつてバルクのホモジネート（組織破砕液）の中に埋もれて見えなくなっていた個々の細胞が、今ではヒトの参照マップ上の個別の座標として保持されています。

参考文献

[1] Macosko, E. Z., Basu, A., Satija, R., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell, 161(5), 1202-1214. https://doi.org/10.1016/j.cell.2015.05.002

[2] Klein, A. M., Mazutis, L., Akartuna, I., et al. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell, 161(5), 1187-1201. https://doi.org/10.1016/j.cell.2015.04.044

[3] 10x Genomics. (October 15, 2024). 10x Genomics Delivers 'Single Cell for a Single Cent' with New Chromium Launches. https://www.prnewswire.com/news-releases/10x-genomics-delivers-single-cell-for-a-single-cent-with-new-chromium-launches-302276878.html

[4] Haghverdi, L., Lun, A. T. L., Morgan, M. D., & Marioni, J. C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbours. Nature Biotechnology, 36(5), 421-427. https://doi.org/10.1038/nbt.4091

[5] Korsunsky, I., Millard, N., Fan, J., et al. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nature Methods, 16(12), 1289-1296. https://doi.org/10.1038/s41592-019-0619-0

[6] Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., & Yosef, N. (2018). Deep generative modeling for single-cell transcriptomics. Nature Methods, 15(12), 1053-1058. https://doi.org/10.1038/s41592-018-0229-2

[7] Stoeckius, M., Hafemeister, C., Stephenson, W., et al. (2017). Simultaneous epitope and transcriptome measurement in single cells. Nature Methods, 14(9), 865-868. https://doi.org/10.1038/nmeth.4380

[8] Welch, J. D., Kozareva, V., Ferreira, A., et al. (2019). Single-cell multi-omic integration compares and contrasts features of brain cell identity. Cell, 177(7), 1873-1887. https://doi.org/10.1016/j.cell.2019.05.006

[9] Ling, Y., et al. (2025). CellStream: Dynamical Optimal Transport Informed Embeddings for Reconstructing Cellular Trajectories from Snapshots Data. arXiv preprint arXiv:2511.13786. https://arxiv.org/abs/2511.13786

[10] Human Cell Atlas. (Accessed August 18, 2026). About the Human Cell Atlas. https://www.humancellatlas.org/learn-more/about-the-human-cell-atlas/

9.2 1細胞トランスクリプトミクスにおける固有ノイズ低減に向けた補完と基盤モデルのベンチマーク

教室の出席簿を思い浮かべてみてください。生徒は実際に教室の中にいますが、先生が名前を呼んだときにうつむいていたため、出席記録には欠席と記されてしまいます。この誤りは、生徒が物理的にいなかったからではなく、抽出（サンプリング）方法の問題を反映しています。これと同様のサンプリングの失敗が、1細胞ゲノミクスでも常に発生しています。それがドロップアウト（Dropout）です。細胞内で活発に転写されている遺伝子がシーケンシング中に検出されず、発現値「0」として記録されてしまう現象です。哺乳類の単一細胞に含まれる総mRNA量は、およそ1ピコグラム（1兆分の1グラム）にすぎません。逆転写酵素がこのごく微量のmRNAをcDNAへと変換する過程で多くの分子が失われ、液滴マイクロ流体技術による捕捉効率は平均して10%から20%にとどまります。シーケンシングのリード数は何百万もの個々の細胞に分散して割り当てられるため、細胞あたりのシーケンシング深度はバルクアッセイに比べて浅くなります。その結果、発現レベルの低い転写産物はサンプリングから漏れてしまうことが頻繁に生じます。これにより、1細胞発現行列（細胞×遺伝子）の要素の80%から95%がゼロで占められることになります。その影響は甚大です。ヘルパーT細胞の決定的なマーカーであるCD4遺伝子が、本物のヘルパーTリンパ球の多くにおいてゼロとして記録されてしまうのです。これらのゼロカウントを誤って分類すると、細胞のクラスタリングが歪み、希少な細胞集団が断片化してしまいます。

ここで本質的な区別が必要となります。観測されたゼロというカウントは、生物学的なものなのでしょうか、それとも技術的なものなのでしょうか。生物学的ゼロは、実際に転写活性がないことを反映しています。例えば肝細胞では、神経細胞のニューロフィラメント遺伝子の転写産物が正真正銘ゼロとなります。一方、技術的ゼロは、活発に発現している遺伝子が測定中にドロップアウトすることで発生します。行列上では、どちらも全く同じ数値のゼロとして現れます。両者を識別することは極めて重要です。技術的ドロップアウトを生物学的サイレンシング（発現停止）として扱えば重要な表現型マーカーを見落とすことになり、逆に真の生物学的サイレンシングに誤って値を補完してしまうと、実際には存在しない細胞状態を作り出してしまうからです。補完アルゴリズムは、技術的ノイズを生物学的なサイレンシングから統計的に分離し、欠損した値を復元することを目指しています。

インピュテーション（欠損値補完）戦略は、方法論的に3つの系統に沿って発展してきました。第1の系統は、グラフベースの近傍拡散に依拠しています。MAGICは転写の類似性に基づいてk近傍細胞親和性グラフを構築し、水中のインクのように発現値をグラフ全体に拡散させて、近傍の状態に基づいて欠損したエントリを平滑化します [1]。拡散タイムステップパラメータtの選択

は極めて重要です。tを小さく設定しすぎると技術的ドロップアウトが未補正のまま残り、tを大きく設定しすぎるとデータセットが過剰に平滑化され、明確に異なる細胞同士が均一な平均値へと同質化してしまいます。第2の系統は、確率的パラメトリックカウントモデルを使用します。1細胞のカウントデータは、過分散を捉える負の二項分布に従います。Deep Count Autoencoders（DCA）およびscVIは、ニューラルネットワークを学習させてゼロ過剰負の二項（ZINB）分布のパラメータを推定し、遺伝子ごとの潜在的な平均発現量とドロップアウト確率の双方を明示的に推定します [2]。これらのオートエンコーダーは、近傍の生の値を単にコピーするのではなく、大域的な分布法則を学習します。第3の系統は、1細胞基盤モデルを使用します。scGPT（3,300万個の1細胞データで事前学習）およびGeneformer（3,000万個の1細胞データで事前学習）は、Transformerアーキテクチャを細胞生物学に適応させ [3][4]、細胞を文、遺伝子を個々の単語トークンとして扱います。scGPTは遺伝子の識別子とビン分割された発現値をトークン化しますが、Geneformerは細胞ごとの発現量の大きさで遺伝子をランク付けします。マスクされた言語モデルが文脈から欠落した単語を予測するのと同様に、1細胞Transformerは共発現のコンテキストからマスクされた遺伝子発現を予測するように学習され、ドロップアウトした値を補完できる潜在的な細胞の埋め込みを生成します。

インピュテーションには重大な失敗モードが存在します。それは過剰平滑化（Over-smoothing）です。近傍の細胞から値を補完することは、真の生物学的異質性を消し去ってしまうリスクを伴います。2018年の研究では、インピュテーションアルゴリズムが生データには存在しない人工的な遺伝子間相関や偽のマーカー遺伝子を頻繁に生成し [5]、ノイズから人工的なシグナルを作り出してしまうことが示されました。2020年の包括的なベンチマークでは、細胞クラスタリング、軌跡推論、および差分発現解析にわたって18種類のインピュテーション手法が評価されました [6]。このベンチマークにより、いくつかのインピュテーションツールは未補完の生データを分析する場合よりも劣る下流結果を生み出し、すべての評価指標において優位に立つ単一のツールは存在しないことが明らかになりました。クラスタリング精度で上位にランクされた手法が、差分発現解析では低いスコアを記録しました。2026年の大規模な比較分析でもこれらの知見が補強され、各手法の相対的な性能はデータセットの背景やタスクの定式化に大きく依存することが示されました [7]。その結果、現代の1細胞解析パイプラインでは、確定的な統計検定に補完済みカウント行列を使用することを避けています。研究者たちはscVIのような確率モデルによって生成された潜在空間内でクラスタリングを実行するものの、マーカー発現の検証は未補完の生のカウントデータに対して行っています。ゼロの除去は慎重に進めなければなりません。補完前後の妥当性確認を行わずに検証不十分なインピュテーションアルゴリズムを実行することは、教室の様子を確認することなく、欠席の記録を出席の印へと書き換えてしまうことと同じです。

1細胞基盤モデルをめぐる議論も同様に先鋭化しています。数千万個の細胞を用いた事前学習によって、あらゆる下流タスクで汎用的な性能が得られるという期待は、厳格なベンチマークによる検証を受けることになりました。2025年の『Nature Methods』誌の研究では、遺伝的摂動に対する細胞応答の予測において、深層学習（DL）アーキテクチャが単純な線形回帰のベースラインを上回ることができなかったと報告されています [8]。2026年2月には、標準的なベンチマーク全体において、パラメータを持たない表現が基盤モデルと同等か、あるいはそれを上回る結果を頻繁に示したことが実証されました [9]。さらに2026年6月、VCBenchは、摂動予測、異種間転移、遺伝子制御ネットワークの推論、発生擬時間の順序付けにおいて、Geneformer、scGPT、UCE、TranscriptFormer、Arc Stateを、事前に登録されたベースラインモデルと比較してベンチマーク評価を行いました [10]。その結果、5つの評価軸のうち4つで、単純な線形ベースラインや最近傍モデルが、ディープな基盤モデルと同等以上の性能を示しました。摂動予測タスクでは、加法ベースラインが0.890を記録したのに対し、Geneformerは0.627、scGPTは0.545という結果でした [10]。未処理の対照細胞に平均的な摂動変化を加算するだけの単純な手法が、数百万個のパラメータを持つ深層モデルを上回ったのです。

これは、基盤モデルに有用性がないことを意味するものではありません。機構的解釈可能性の研究によれば、内部のアテンション層が体系化された遺伝子制御プログラムを捉えており、これらのモデルは、ラベルのない1細胞データセットに細胞オントロジーの注釈を転移する際に優れた効果を発揮します。言語モデルも歴史的に、アーキテクチャとデータスケーリングが成熟する前に同様の厳しい検証を経験しました。重要なのはベンチマークの規律です。複雑なニューラルネットワーク構造を妥当性確認するには、まず単純な線形ベースラインを基準にベンチマークを行う必要があります。評価パイプラインにベースライン対照を組み込むことこそが、2026年の手法面における最も重要な教訓です。1細胞マトリックスの欠損値を補正するにあたっては、複雑さよりも正確さが求められ、未解明の細胞状態が持つ微細なシグナルを確実に維持することが不可欠です。

参考文献

[1] van Dijk, D., Sharma, R., Nainys, J., et al. (2018). Recovering gene interactions from single-cell data using data diffusion. Cell, 174(3), 716-729. https://doi.org/10.1016/j.cell.2018.05.061

[2] Eraslan, G., Simon, L. M., Mircea, M., Mueller, N. S., & Theis, F. J. (2019). Single-cell RNA-seq denoising using a deep count autoencoder. Nature Communications, 10, 390. https://doi.org/10.1038/s41467-018-07931-2

[3] Cui, H., Wang, C., Maan, H., et al. (2024). scGPT: toward building a foundation model for single-cell multi-omics using generative AI. Nature Methods, 21(8), 1470-1480. https://doi.org/10.1038/s41592-024-02201-0

[4] Theodoris, C. V., Xiao, L., Chopra, A., et al. (2023). Transfer learning enables predictions in network biology. Nature, 618(7965), 616-624. https://doi.org/10.1038/s41586-023-06139-1

[5] Andrews, T. S., & Hemberg, M. (2018). False signals induced by single-cell imputation. F1000Research, 7, 1740. https://doi.org/10.12688/f1000research.16613.2

[6] Hou, W., Ji, Z., Ji, H., & Hicks, S. C. (2020). A systematic evaluation of single-cell RNA-sequencing imputation methods. Genome Biology, 21, 218. https://doi.org/10.1186/s13059-020-02132-x

[7] Iwashita, Y., et al. (2026). A Large-Scale Comparative Analysis of Imputation Methods for Single-Cell RNA Sequencing Data. arXiv preprint arXiv:2603.24626. https://arxiv.org/abs/2603.24626

[8] Ahlmann-Eltze, C., Huber, W., & Anders, S. (2025). Deep-learning-based gene perturbation effect prediction does not yet outperform simple linear baselines. Nature Methods, 22. https://www.nature.com/articles/s41592-025-02772-6

[9] Souza, H., & Mehta, P. (2026). Parameter-free representations outperform single-cell foundation models on downstream benchmarks. arXiv preprint arXiv:2602.16696. https://arxiv.org/abs/2602.16696

[10] Weidener, L., et al. (2026). VCBench: A Multi-Dimensional Benchmark for Single-Cell Foundation Models. bioRxiv. https://www.biorxiv.org/content/10.64898/2026.06.18.733146v1

第10章：1細胞状態遷移と動的摂動応答解析

10.1 Chreodeによる時間的細胞発生軌跡とトランスクリプトミクス摂動モデリング

実験台の上にマルチウェル培養プレートが置かれています。各ウェルには、同じ日に同一の培地で播種された細胞が入っています。6時間が経過した時点で、研究者は1列目を取り出します。細胞内mRNAを単離するために細胞膜が溶解され、その過程でそれらの細胞は破壊されます。12時間後には2列目が処理されて溶解されます。24時間、48時間、72時間。1列ずつ、プレートは空になっていきます。実験が終わったとき、残るのは5つのデータファイルだけであり、アッセイを生き延びた細胞はゼロでした。

1細胞RNAシーケンシングは破壊的に機能します。細胞のトランスクリプトーム状態を読み取るには、細胞を破壊する必要があるからです。その結果、1つの生きた細胞が時間をかけて発生していく様子を連続した動画として記録することはできません。私たちが手にするのは、離散的な時点において何万もの異なる個々の細胞を捉えた、5枚の静的なスナップショット写真です。それぞれの写真は、異なる個別の細胞群を捉えています。しかし、生物学的な探究が焦点を当てるのは時間的ダイナミクスです。多能性幹細胞が心筋細胞に分化するのか神経細胞に分化するのかを特定することや、炎症刺激の数日後に健康な肝細胞がどこへ状態遷移するのかをマッピングすることなどがこれに該当します。

発生生物学者のコンラッド・ワディントンは1957年、起伏のある丘を転がり落ちるボールという1枚の図によってこの概念を定式化しました。その丘には、分岐する谷が刻まれています。ボールは無作為に転がるのではなく、刻まれた谷の輪郭に従って進みます。彼はこの地形をエピジェネティックランドスケープと名付け、分岐した個々の谷を、ギリシャ語の「必要な道」に由来するChreodeと名付けました。細胞は無作為な表現型状態へと漂流するのではなく、あらかじめ定められた発生の谷に沿って運河化（canalize）されます[1]。70年後の現在、1細胞ダイナミクスにおける基盤モデルは、その理論的系譜をこの概念に辿ることができます。

静的なスナップショットから時間的方向性を再構成するための最初の計算フレームワークがRNA Velocityでした。新生の未スプライシングpre-mRNAにはイントロンが含まれていますが、成熟したスプライシング済みmRNAではイントロンが除去されています。各遺伝子における未スプライシング転写産物とスプライシング済み転写産物の比率を測定することで、転写発現が上昇しているのか下降しているのかが示され、単一のスナップショットプロファイルから方向微分ベクトルを推定することができます[2]。2020年、scVeloは動的キネティクスモデルによって定常状態の仮定を緩和し、このフレームワークを拡張しました[3]。

しかし、推定されたベロシティベクトルは相反する方向を指すことが少なくありません。2026年6月に『PLOS Computational Biology』誌で発表された比較評価では、3つのデータセットにわたって5つのRNAベロシティアルゴリズムのベンチマークを実施し、局所的なベクトルの一貫性、手法間の一致度、およびシーケンシングリード深度への感度を評価しました[4]。同一のデータセットに異なるアルゴリズムを適用したところ、互いに矛盾する軌跡が生成されました。2026年3月に『Cell Reports Methods』誌に掲載されたより大規模なベンチマークでは、17の独立した研究にわたって15の手法を評価し、精度、安定性、使いやすさのすべてにおいて同時に優れている手法は存在しないことが明らかになりました[5]。ヒト胎児前脳のデータセットでは、競合する2つの手法が同一の細胞クラスターに対して正反対の発生の流れをマッピングしました。ベロシティベクトルは未スプライシングのイントロン数に依存していますが、これは遺伝子特異的なキネティクスや技術的ノイズの影響を受けやすい脆弱な生物学的シグナルなのです。

その結果、近年の手法は局所的な速度の矢印から、大域的な分布輸送へと移行しました。「6時間時点で観察された細胞集団の分布を、24時間時点で観察された分布へと輸送するための最小コストのマッピング計画は何か」という問いが核心となったのです。これが最適輸送（Optimal Transport: OT）です。移動距離の合計を最小限に抑えて最寄りの保管倉庫を満たすようにトラックの運行計画を立てるのと同様に、最適輸送は個々の細胞を1対1で追跡することなく、確率分布全体を後続の時点へとマッピングします。CellStreamは動的最適輸送を潜在表現に組み込むことで、離散的なスナップショットから連続的な発生軌道を再構築しました[6]。

これはベクトル場（Vector Fields）へと直接つながります。高次元のトランスクリプトーム空間において、個々の細胞は1つの座標点です。すべての点に方向性を持った速度ベクトルを割り当てるとベクトル場が形成されます。多様体上のどこに置かれても、ベクトルの流れに従うことで細胞の将来の軌道をたどることができます。ニューラル常微分方程式（Neural ODEs）は、ニューラルネットワークを用いてこのベクトル場をパラメータ化し、連続的な導関数を微分方程式ソルバーに渡します。Neural ODEsは離散的な層を積み重ねるのではなく時間を連続的にモデル化するため、段階的な階層を跳躍するのではなく滑らかに進行する生物学的遷移と合致します。

常微分方程式（ODE）の数値積分は計算上のボトルネックをもたらします。軌道を解くには始点から終点まで細かなステップで数値積分を行う必要があり、データセットが変わるたびに最初からやり直さなければならないためです。2026年5月27日に公開されたChreodeは、このアーキテクチャを逆転させました[7]。ウォディントンのエピジェネティックランドスケープを直接反映した細胞世界モデル（Cell World Model）として設計されたChreodeは、計算負荷をオフラインの事前学習へと移し、下流の軌道推論を単一ステップの遷移作用素によって実行します。この遷移作用素は、谷底へと下る勾配降下流（運河化）、接平面に沿った回転運動、そして確率的拡散

項という3つの異なる動的要素に分解され[7]、ウォディントンの傾斜、等高線、揺らぎをそのまま数学的な項へと変換しています。

評価指標はこの効率性を反映しています。7つのデータセットにわたるマウス胚アトラスの240万個の細胞で事前学習されたChreodeは、Weinrebの造血系ベンチマークおよびVeresの膵島分化ベンチマークにおいて、Sinkhorn距離でde novoベースライン、PI-SDE、PRESCIENTを上回りました。転移性能も注目に値する結果を示しました。Chreodeの発生軌道表現を遺伝子状態の埋め込みとしてGEARS摂動モデルに注入したところ、Norman Perturb-seqにおいて発現変動上位20遺伝子の予測誤差が0.2121から0.1858へと減少し、12.4%のエラー削減を達成しました[7]。発生分化の学習によってCRISPRノックアウト応答の予測精度が向上したことは、細胞発生と遺伝子摂動が共通のエピジェネティックな多様体を巡るという仮説を支持しています。事前学習されたバックボーンは、クローン系統タグを見ることなく、Weinrebデータにおけるゼロショットのクローン運命予測で動的最適輸送のベースラインに匹敵する性能を示しました[7]。

データセットの規模も並行して拡大しました。Tahoe-100Mは、1,100種類の低分子化合物で処理された50種類のがん細胞株にわたり約6万件の摂動条件をアッセイし、1億個の1細胞をプロファイリングしました [8]。2026年1月には、Tahoe Therapeutics、Arc Institute、Chan Zuckerberg Biohubが共同イニシアチブを発表し、これを4倍となる1億2,000万個の細胞および22万5,000件の摂動プロファイルへと拡張することを明らかにしました [8]。こうしたデータを処理するためのスケーラブルなアーキテクチャも登場しました。SCALEは摂動予測を条件付き輸送問題として再定式化し、Tahoe-100Mにおける従来のベンチマークと比較して、事前学習を12.51倍高速化しつつ、ピアソン相関（PDCorr）を12.02%、発現変動遺伝子の重複率を10.66%向上させました [9]。

ワールドモデル（World Model）の概念は、ロボット工学やゲームの強化学習に端を発しています。ロボットのワールドモデルは、物理的な行動を起こす前に、ロボットアームを動かすことが環境をどのように変化させるかを内部でシミュレーションします。行動を入力すると、モデルは次の環境状態を予測します。これを細胞生物学に応用すると、行動は薬剤処理やCRISPRノックアウトに相当し、状態はトランスクリプトームプロファイルに相当するため、実験台でのアッセイを行う前にインシリコでのスクリーニングが可能になります。

これらのモデルの有用性は、薬剤の組み合わせを評価する際に顕著に表れます。がん幹細胞をインシリコで候補分子に曝露させることで、細胞集団がアポトーシスへと向かうのか、それとも薬剤耐性を持つ残存状態を表す隣の谷へと迂回するのかを予測できます。100種類の候補分子の組み合わせを実験的にスクリーニングするには、4,950通りのペアワイズアッセイを試験する必要がありますが、計算によって優先順位を付けることで試験スケジュールを再編できます。

依然として慎重な姿勢が求められます。Arc Instituteが主催した第1回Virtual Cell Challengeの結果は、現在の限界を物語っています。114カ国から5,000人の登録者と1,200の提出チームが集まりましたが、上位にランクインしたモデルですら、すべての評価指標において単純な統計的ベースラインを上回ることはできませんでした。入賞した提出モデルは深層学習（DL）と古典的な統計的特徴量を組み合わせたものであり [10]、エンドツーエンドの深層アーキテクチャ単体では摂動予測を解決できていないことを示しています。賞金総額10万ドルで2026年8月20日に開始される第2回コンペティションでは、評価対象が複数遺伝子の組み合わせノックアウトや未知の細胞型へと拡張されています [10]。

死滅した細胞の5つの破壊的なスナップショットから連続的な時間ダイナミクスを再構築することは、極めて困難な課題です。それでも研究者らは、この問題をウォディントンのエピジェネティックランドスケープの中に位置づけ、240万個の1細胞にわたってモデルを学習させることで谷の地形図を作成しました。ベクトル場には不確実性が含まれるものの、トランスクリプトームのスナップショットに時間という次元を加えることは、細胞の状態遷移をモデル化するための動的な基盤を提供します。

参考文献

[1] Waddington, C. H. (1957). The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology. George Allen & Unwin.

[2] La Manno, G., Soldatov, R., Zeisel, A., et al. (2018). RNA velocity of single cells. Nature, 560(7719), 494-498. https://doi.org/10.1038/s41586-018-0414-6

[3] Bergen, V., Lange, M., Peidli, S., Wolf, F. A., & Theis, F. J. (2020). Generalizing RNA velocity to transient cell states through dynamical modeling. Nature Biotechnology, 38(12), 1408-1414. https://doi.org/10.1038/s41587-020-0591-3

[4] Ancheta, S., et al. (2026). Challenges and progress in RNA velocity: Comparative analysis across multiple biological contexts. PLOS Computational Biology, 22(6), e1014303. https://doi.org/10.1371/journal.pcbi.1014303

[5] Luo, Y., et al. (2025). Benchmarking RNA velocity methods across 17 independent studies. Cell Reports Methods. https://www.cell.com/cell-reports-methods/fulltext/S2667-2375(26)00067-6

[6] Ling, Y., Zhang, P., Zhang, Z., & Zhou, P. (2025). CellStream: Dynamical Optimal Transport Informed Embeddings for Reconstructing Cellular Trajectories from Snapshots Data. arXiv preprint arXiv:2511.13786. https://arxiv.org/abs/2511.13786

[7] Qiu, M., Zheng, G., Xu, Y., Zhang, R., Ding, Y., Long, Q., & Chen, T. (2026). Chreode: A Cell World Model for One-Step Temporal Dynamics and Perturbation Prediction. arXiv preprint arXiv:2605.28111. https://arxiv.org/abs/2605.28111

[8] Arc Institute. (January 20, 2026). Tahoe Therapeutics, Arc Institute, and Biohub Partner to Generate the Largest Perturbation Dataset for Virtual Cell Models. https://arcinstitute.org/news/tahoe-arc-biohub

[9] Chen, S., Yu, L., Jin, K., et al. (2026). SCALE: Scalable Conditional Atlas-Level Endpoint transport for virtual cell perturbation prediction. arXiv preprint arXiv:2603.17380. https://arxiv.org/abs/2603.17380

[10] Arc Institute. (December 9, 2025). Virtual Cell Challenge 2025 Wrap-Up: Winners and Reflections. https://arcinstitute.org/news/virtual-cell-challenge-2025-wrap-up

10.2 Sigmoid Attentionを用いた数万個の遺伝子にわたる動的競合相互作用の計算

初めてアテンションヒートマップを表示すると、整然とした格子が現れます。水平軸と垂直軸の双方に文が並び、マトリックス全体のセルの強度が変化します。代名詞「it」の行を選択すると、先行詞である名詞に対応する単一のセルが明るく光り、周囲のセルは暗いままになります。その行の値の合計は正確に1.0になります。

Softmaxはこの正規化を強制します。生のロジットの指数関数を計算し、その総和で割ることで、値を厳密な確率分布へと変換します。2017年のTransformerアーキテクチャが自己アテンションにSoftmaxを組み込んだのには理由があります[1]。自然言語において、単数の代名詞は1つの指示対象に特定されます。「it」が「cat」を指しているならば、同時に「table」を指すことはありません。各投票者が1票のみを投じる学級選挙のように、1つのトークンにより高いアテンション重みを割り当てると、他のすべてのトークンに使える割り当てが減るというゼロサムの計算が行われます。

細胞生物学においては、この前提が破綻します。マスター転写因子であるp53が活性化すると、数百個の下流標的遺伝子が同時に転写されます。MYCについても同様です。遺伝子Aが活性化するからといって、遺伝子Bを不活性化させる必要はありません。生きた細胞は1票ずつ割り当てる選挙のようには機能せず、数十個のスイッチが同時にオンになる電気配電盤のように動作します。Softmaxアテンションを適用すると、モデルはこの配電盤をゼロサムの選挙として扱うことを強いられます。すなわち、1つの標的遺伝子に高いアテンション重みを割り当てると、共調節される他の標的遺伝子の重みが強制的にゼロへと押し下げられてしまいます。

ゲノム規模で見ると、この数値的な歪みは明白になります。ヒトゲノムは約20,000個のタンパク質コード遺伝子をコードしています。アテンションの各行の合計が1.0でなければならないとすると、遺伝子1個あたりの平均一様重みは1/20,000（0.00005）になります。p53のようなマスター調節因子が300個の下流標的を制御する場合、アテンションを均等に配分すると各標的に割り当てられる重みはおよそ0.003となり、シグナルは背景ノイズの中に埋もれてしまいます。逆に、モデルが3個の標的にアテンションを集中させると、残りの297個の真の制御関係はゼロへと抑制されてしまいます。正規化の仕組みによって、本物の生物学的な共調節が見えなくなってしまうのです。

Sigmoid Attentionは、この行ごとの除算を排除します。QueryベクトルとKeyベクトルの内積（次元数の平方根でスケーリングされたもの）は、遺伝子ペアごとに独立した要素ごとのSigmoid関数へと渡されます。各セルは行の合計が1.0になることを要求されることなく、0から

1の範囲内の値を出力します。調節因子は300個の標的に同時に強く結合することも、0個に結合することも可能です。ある相互作用が隣接する相互作用を希釈することがないため、相互作用の絶対的な強度が保たれます。

Appleの研究者たちは2024年に、Sigmoid Attentionが大規模言語モデル（LLM）において効果的に機能することを実証しました。Sigmoid Attentionを使用するTransformerは普遍的関数近似器であり続け、初期学習時のアテンションノルムを調整すればSoftmaxと同等の性能を発揮します。彼らが開発したFlashSigmoidカーネルは、Nvidia H100 GPU上でFlashAttention-2よりも17%高速に推論を実行しました[2]。

1細胞基盤モデルにおいてSoftmaxをSigmoid Attentionに置き換える生物学的な検証結果は、2026年4月29日に発表されました[3]。6つの1細胞データセットにおいて、Softmaxのベースラインと比較して細胞型のクラスタリング分離度が25%向上し、学習スループットは最大10%向上しました[3]。この利点は勾配のダイナミクスに由来します。Sigmoid関数の導関数は0.25を超えることがなく、そのヤコビ行列は完全に対角行列のままです。対照的にSoftmaxは密なヤコビ行列を持ち、1つの要素を変更するとその行全体のすべての要素が変動してしまいます。1億6000万パラメータのモデルを用いたストレステストでは、Softmaxは1万倍の勾配爆発を起こして学習が崩壊したのに対し、Sigmoid Attentionは数値的に安定したままでした。研究チームのTritonSigmoidカーネルは、H100ハードウェア上で515 TFLOPSに達しました[3]。

演算の規模を考えてみてください。個別の遺伝子2万個をトークン化すると、1細胞あたり20,000×20,000、すなわち4億要素のアテンション行列が生成されます。この非圧縮の行列をGPUのVRAMに読み込もうとすると、メモリ不足（OOM）エラーが発生します。FlashAttentionは、中間アテンションマップを実体化させることなく、高速なSRAMに行列をタイリング（分割配置）することで、このメモリの壁を回避しました[4]。Sigmoid Attentionが実際の生物学的ワークフローに導入されたのは、現代のGPUメモリカーネルがそれを支えられるまでに成熟したからです。数学的な修正とハードウェア上での実行は密接に結びついています。

その実用的な影響は、遺伝子制御ネットワーク（GRN）の推論において明確に現れます。制御ネットワークを推論すると、左側に制御因子、右側に標的、そして割り当てられた信頼度スコアからなる相互作用テーブルが生成されます。Softmaxで学習されたモデルでは、上位の相互作用が高いスコアを示すものの、その後にスコアが急落し、実験的に検証された相互作用が200位以降に落ちて偽陰性（False Negative）エラーを引き起こします。一方、Sigmoid Attentionモデルではスコアの分布が緩やかに減少するため、実際に存在する複数の正当な制御関係が同時に高い信頼度スコアを維持できます。

明確に区別しておくべき点があります。生のアテンション重みは、直接的な因果関係を持つ生物学的制御と等価ではありません。2026年3月の機械論的解釈可能性（mechanistic interpretability）に関する研究では、Geneformerの第5層にある4,065個のアクティブなスパースオートエンコーダー特徴量を分析し、140万個の下流機能結合をマッピングしました[5]。その結合性には大きな偏りがあり、1.8%の特徴量が密なハブとして機能し、上位20個のハブのうち8個には識別可能な生物学的アノテーションが認められませんでした。特徴量のアブレーション（除去）テストでは、冗長率が0.59（3つの特徴量）および0.74（2つの特徴量）となり、非線形なシナジーは皆無でした[5]。GeneformerとscGPTを比較した関連研究でも、スパースオートエンコーダーは整理されたオントロジー的な細胞知識を抽出するものの、捉えられる因果的な制御ロジックは限定的であることが示されました[6]。アテンション重みは、生化学的な因果的活性化というよりも、統計的な共発現の相関関係を反映しているのです。

さらなるベンチマークも慎重な見方を後押ししています。2026年2月、パラメータフリーの表現が複数の下流ベンチマークにおいて1細胞基盤モデルを上回ったことが研究者によって報告されました[7]。2025年の『Nature Methods』のベンチマークでは、GEARS、scGPT、scFoundationを単純な線形ベースラインと比較評価し、単一遺伝子および二重遺伝子の摂動タスクにおいて深層モデルが線形対照群を超えられないことが判明しました[8]。ICML 2026の論文ではこの力学が説明されています。1細胞のばらつきの大部分は摂動のないベースラインの恒常性シグナルを反映しており、摂動によって誘発される変化はスパース（希薄）にとどまります。そうした微細なシグナルを分離しない限り、モデルのパラメータ数を増やしても収穫逓減に陥るのです[9]。

アーキテクチャの改変は、Sigmoid Attentionだけにとどまりません。2025年12月に発表されたSpikGPTは、細胞タイプのアノテーションのために、スパイキングニューラルネットワークのダイナミクスをアテンション層へと統合しました[10]。また他のアーキテクチャでは、アテンションを完全に迂回し、連続的なグラフダイナミクスを直接モデル化しています。共通するテーマは、自然言語処理の前提を生体システムへと無批判に適用することをやめる点にあります。

SoftmaxをSigmoid Attentionに置き換えること自体は、わずか数行のコードを変更するだけで済みます。しかし、その数行を実装するためには、遺伝子制御の物理的原理を再評価し、専用のGPU実行カーネルを記述し、1億6000万パラメータのネットワーク全体にわたる勾配の安定性を分析する必要がありました。生物学的事実に基づいて設計された数学的な適応は、細胞ネットワークをモデル化するための確固たる基盤を確立しています。

参考文献

[1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. NeurIPS 2017. arXiv preprint arXiv:1706.03762. https://arxiv.org/abs/1706.03762

[2] Ramapuram, J., Danieli, F., Dhekane, E., et al. (2024). Theory, Analysis, and Best Practices for Sigmoid Self-Attention. arXiv preprint arXiv:2409.04431. https://arxiv.org/abs/2409.04431

[3] Sadashivaiah, V., Dasoulas, G., Mueller, J., & Ghosh, S. (2026). Better Models, Faster Training: Sigmoid Attention for single-cell Foundation Models. arXiv preprint arXiv:2604.27124. https://arxiv.org/abs/2604.27124

[4] Dao, T., Fu, D. Y., Ermon, S., Rudra, A., & Ré, C. (2022). FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness. arXiv preprint arXiv:2205.14135. https://arxiv.org/abs/2205.14135

[5] Kendiukhov, I. (2026). Exhaustive Circuit Mapping of a Single-Cell Foundation Model Reveals Massive Redundancy, Heavy-Tailed Hub Architecture, and Layer-Dependent Differentiation Control. arXiv preprint arXiv:2603.11940. https://arxiv.org/abs/2603.11940

[6] Kendiukhov, I. (2026). Sparse autoencoders reveal organized biological knowledge but minimal regulatory logic in single-cell foundation models: a comparative atlas of Geneformer and scGPT. arXiv preprint arXiv:2603.02952. https://arxiv.org/abs/2603.02952

[7] Souza, H., & Mehta, P. (2026). Parameter-free representations outperform single-cell foundation models on downstream benchmarks. arXiv preprint arXiv:2602.16696. https://arxiv.org/abs/2602.16696

[8] Ahlmann-Eltze, C., Huber, W., & Anders, S. (2025). Deep-learning-based gene perturbation effect prediction does not yet outperform simple linear baselines. Nature Methods. https://doi.org/10.1038/s41592-025-02772-6

[9] Jiang, W., Liu, Y., Cai, Y., Gao, E., Dong, J., Abbasnejad, E., Yao, L., & Shi, J. Q. (2026). What Makes a Representation Good for Single-Cell Perturbation Prediction? ICML 2026. arXiv preprint arXiv:2605.19343. https://arxiv.org/abs/2605.19343

[10] Huang, M., & Kamaleswaran, R. (2025). SpikGPT: A High-Accuracy and Interpretable Spiking Attention Framework for Single-Cell Annotation. arXiv preprint arXiv:2512.03286. https://arxiv.org/abs/2512.03286

第11章：AIエージェント駆動の合成生物学と遺伝子回路設計

11.1 ゲノムおよび制御配列の自動設計に向けた大規模バイオロジカルエージェントの展開

深夜2時、研究室には蛍光灯のうなる音だけが響いています。実験台の上にはノートが開かれており、大学院生がプライマーの長さを数えています。アニーリング温度を手計算し、Gibson Assembly反応の化学量論比を計算しては再計算し、消しては書き直します。翌朝にアッセイが失敗したとき、その根本原因はたいていそのノートのどこかに隠れています。計算がわずか2℃狂っただけで、3日間の実験台での作業が水泡に帰してしまうのです。

合成生物学は、生体を標準化された部品へと分解し、それらを再構築する工学分野です。研究者たちは、大腸菌（Escherichia coli）や出芽酵母（Saccharomyces cerevisiae）のゲノムをモジュール化されたパーツに断片化し、それらをつなぎ合わせることで望ましい生物学的機能を組み込みます。従来は石油に由来していたバイオプラスチックの前駆体を合成するように微生物をプログラミングしたり、がん特異的な表面抗原と結合したときにのみサイトカインを分泌するように免疫細胞を改変したりする例が挙げられます。部品の標準化が最初に定式化されたのは、BioBrick標準でした[1]。あらゆる遺伝子パーツの5'末端および3'末端の隣接領域に同一の制限酵素切断部位を配置することで、同一の酵素反応を用いて任意の部品を切断・連結できるようにしました。これは、あらゆる家電製品を問題なく差し込めるように壁のコンセントの規格を統一することに似ています。

学生たちがこれらの標準を実践する国際的な舞台が、iGEM大会（iGEM Competition）です。2026年の大会では世界中から約300チームが登録し、「Software & AI」をはじめとする専用のプロジェクトトラックが設けられました[2]。参加チームは新たに作製したパーツを「Registry of Standard Biological Parts（標準生物パーツ登録所）」に登録し、毎年11月にパリで開催されるグランドジャンボリー（Grand Jamboree）でその成果を発表します。この登録所は公共図書館のように機能します。あるチームが返却した1冊の本が、次のチームにとっての出発点となる基盤として役立つのです。リポジトリが拡張するにつれて、新しく加わる学部生たちは一から（de novoで）始めるのではなく、すでに確立された部品を基盤にして開発を進めることができます。

生成人工知能は、このリポジトリに新たな展開をもたらしました。『Cell Systems』誌の総説では、現在の時期を転換点と位置づけています[3]。物理的なウェットラボに足を踏み入れる前に、モジュール化されたパーツ、遺伝子回路、そして完全なゲノムを計算機上で設計する時代に入ったのです。配列設計では、複数の制約を同時に調和させる必要があります。コドン最適化（Codon Optimization）は、宿主生物が持つ特定のtRNA存在量プールに合わせてコーディン

グ配列を調整します。縮重コドンは同一のアミノ酸へと翻訳されるため、リボソームの進行を妨げるmRNAの二次構造（ヘアピン構造）を避けつつ、宿主の翻訳バイアスに適合するように選択を行い、転写開始点、翻訳開始モチーフ、終結配列、ポリアデニル化シグナルを正確に配置します。

研究者がAIエージェントに提示する仕様は、今や具体的な形をとっています。チャイニーズハムスター卵巣（CHO）細胞において、抗体の重鎖および軽鎖の合成を正確に1対2の化学量論比で駆動するデュアルプロモーター発現プラスミドを、実行可能なロボットプロトコルとともに設計するというものです。この化学量論比が崩れると、抗体鎖のミスフォールディングが起き、細胞内で凝集してしまいます。以前は、この比率のバランスを取るために数か月に及ぶ経験的なプロモータースクリーニングが必要でした。現在では、エージェントが最適化された塩基配列、遺伝子パーツの配置、そして実行可能なリキッドハンドリングスクリプトを単一の統合パッケージとして提示します。研究者の役割は、生成されたパッケージを評価し、ストレステストを行うことへと移行しています。

実務上のボトルネックは実行にあります。回路の設計自体は1日で完了しますが、それを生きた細胞へとクローニングするために必要な物理的プロトコルには、数百もの数値パラメータが含まれています。ピペッティング容量、サーマルサイクラーの昇降温速度と保持時間、遠心分離速度、抗生物質による選抜濃度などを明示的に指定しなければなりません。かつて研究者たちは、紙の付録資料や試薬マニュアルを調べ上げてこれらのパラメータを組み立てていました。現在では、大規模言語モデル（LLM）エージェントがこのギャップを埋めています。自然言語で記述された生物学的目標を、構造化されたプロトコルや自動リキッドハンドラー用の機械可読な実行コードへと翻訳しているのです。

2026年6月に発表されたProtoPilotは、この能力を定量化しました[4]。著者らは、検証済みの98の標準作業手順書から導出された294件の合成生物学および分子生物学タスクからなるベンチマークを構築しました。ProtoPilotは、単一のモノリシックなシステムを適用するのではなく、意図の解析、ハードウェアおよび数値制約の検証、実行コードの生成を分離し、成功したプロトコルを後から検索できるモジュール型のスキルライブラリに保存します。評価対象となったプロトコルは、専門家による評価の90.2%において上位3位以内にランク付けされました。実行コードにコンパイルされたプロトコルは、89.5%の事例で仮想ドライラン検証を通過し、実機のOpentronsリキッドハンドラーにおいて88.24%の実行合格率を達成しました。同一タスクにおける従来のベースラインAIツールの合格率は32.35%でした。この性能差は、言語的な流暢さによるものではなく、ハードウェアを意識した検証ゲートによってもたらされたものです。研究者らは、エージェントが作成したプロトコルによって生成されたGibson Assembly産物をサンガーシーケンシングで検証し、実験現場での物理的な実用性を実証しました。

身体性を持つラボラトリーロボティクスもこれに歩調を合わせて進歩しました。2026年6月に公開されたPipetteベンチマークは、液体の分注、培養容器の搬送、ハードウェア操作、精密ツールの配置にわたる12種類の異なるウェットラボタスクを網羅しています[5]。タスクあたりわずか30回の人間の専門家による遠隔操作デモンストレーションを与えた場合、ベースラインポリシーの平均成功率は60.3%でしたが、シミュレーション拡張データを統合することで成功率は40.4%から71.8%へと向上しました。ロボットアームに物理的な触覚の繊細さを教え込むには30回の手本では不十分ですが、物理シミュレータが学習データ分布の不足を補完します。市販のハードウェアメーカーもこれに応じて適応を進めました。Opentronsは2026年2月にボストンで自然言語によるqPCRワークフロー生成を実演し、2026年4月にリリースされたApp 9.0には自動ワークフローシミュレーションを直接統合しました[6]。これにより、物理ハードウェアでの実稼働を行う前に、シミュレーション上でプロトコルの力学的・物理的動作を検証できるようになりました。

機械可読性を実現するには、プロトコルの標準化が必要です。生物学の文献では、実験手法が本文、表、図、補足資料のPDFなどに分散して記載されています。人間はこうした断片を目視で統合して理解できますが、機械には構造化されたデータが必要です。現在進められている標準化の取り組みでは、試薬の識別子、濃度、インキュベーション時間、温度、ハードウェアの型番といった個別のメタデータフィールドへと実験プロトコルを構造化し、文章による記述を機械が実行可能なパラメータへと変換しています。

重要な知見の1つは、プロンプトエンジニアリングだけでは不十分であるという点です。言語モデルを高分子モデリングスイートRosettaと統合したAgent Rosettaプロジェクトでは、構造化された環境がなければ、いくら流暢な文章を作成できてもドメイン固有の正確なコマンドを生成できないことが実証されました[7]。フォーマットが正しくないコマンドは、ドメインツールをエラー通知なしで密かに停止させたり、破損した出力を返したりする原因になります。抗体工学においては、Latent-Yエージェントが9つの標的に対する創薬キャンペーンを完遂し、その67%の標的において1桁ナノモーラー台の親和性を持つ実験的結合を達成しました。これにより、専門家による設計ワークフローを56倍高速化し[8]、数週間に及ぶ手作業をわずか数時間に圧縮しました。

エージェントの有用性は、その検証チェックポイント、すなわち自らの出力を棄却し、デバッグし、改良する能力によって決まります。現代の言語モデルであれば、どれも説得力のあるプロトコル文書を生成できます。真に必要なのは、そのプロトコルを物理的なハードウェアの制約に通し、構文エラーや実行エラーを検出してスクリプトを修正し、ウェットラボからのフィードバックをその後の反復へと組み込むことです。Design（設計）、Build（構築）、Test（試験）、

Learn（学習）を自動化されたループへと繋ぎ合わせることで、ラボは手作業のワークステーションから統合されたシステムへと変貌します。

能力の向上に伴い、安全性への配慮もより重要になります。2026年6月に発表されたバイオセキュリティのベンチマークであるABC-Benchは、自動リキッドハンドラーのコーディング、DNAアセンブリ設計、DNA合成スクリーニング回避の3項目にわたり言語モデルエージェントを評価しました[9]。テストされたエージェントは、3つのタスクすべてにおいて人間の専門家の中央値を上回る性能を示しました。3つ目のタスクはバイオセキュリティ上のリスクを浮き彫りにしています。生物学的設計の自動化能力が高まるにつれて、危険な配列のスクリーニングを回避する技術的ハードルが低下するためです。商業バイオファウンドリ市場はこの成長を反映しており、2025年の16億5,000万ドルから2026年には20億1,000万ドルへと拡大し、年率21.3%で成長しています[10]。実験記録は今や、消えてしまう鉛筆のメモではなくサーバーログに保存されるようになり、実行されたすべてのプロトコルに対して監査可能な追跡記録が提供されています。

参考文献

[1] Shetty, R. P., Endy, D., & Knight, T. F. (2008). Engineering BioBrick vectors from BioBrick parts. Journal of Biological Engineering, 2, 5. https://doi.org/10.1186/1754-1611-2-5

[2] iGEM Foundation. (2026). About the iGEM Competition. https://competition.igem.org/about

[3] Kim, N., Carluccio, G. D., Zhang, K., & Collins, J. J. (2026). Generative AI for synthetic biology: Designing biological parts, circuits, and genomes. Cell Systems, 17, 101533. https://doi.org/10.1016/j.cels.2026.101533

[4] Jiang, Y., et al. (2026). A Self-Evolving Agentic System for Automated Generation and Execution of Biological Protocols. arXiv preprint arXiv:2606.31763. https://arxiv.org/abs/2606.31763

[5] Liu, Z., Jin, H., Du, Z., et al. (2026). Pipette: An Embodied Simulation Platform, Benchmark, and Data-Efficient Augmentation Framework for Wet-Lab Robotics. arXiv preprint arXiv:2606.12936. https://arxiv.org/abs/2606.12936

[6] The Robot Report. (April 2026). Opentrons introduces dynamic simulation, visualization for AI-generated lab workflows. https://www.therobotreport.com/opentrons-introduces-dynamic-simulation-visualization-ai-generated-lab-workflows/

[7] Teneggi, J., Turzo, S. M. B. A., Marwah, T., Bietti, A., Renfrew, P. D., Mulligan, V. K., & Golkar, S. (2026). Protein Design with Agent Rosetta: A Case Study for Specialized Scientific Agents. arXiv preprint arXiv:2603.15952. https://arxiv.org/abs/2603.15952

[8] Schmon, S. M., et al. (2026). Latent-Y: A Lab-Validated Autonomous Agent for De Novo Drug Design. arXiv preprint arXiv:2603.29727. https://arxiv.org/abs/2603.29727

[9] Liu, A. B., Nedungadi, S., Cai, B., Kleinman, A., Bhasin, H., & Donoughe, S. (2026). ABC-Bench: An Agentic Bio-Capabilities Benchmark for Biosecurity. arXiv preprint arXiv:2606.11150. https://arxiv.org/abs/2606.11150

[10] The Business Research Company. (2026). Biofoundry-as-a-Service Global Market Report 2026. https://www.researchandmarkets.com/reports/6241453/biofoundry-as-a-service-global-market-report

11.2 自律的遺伝子回路設計のための階層的検証に基づく強化学習（GenCircuit-RL）

コンピューターのモニター上では、電子信号はシャープな矩形波として振動し、オンとオフの状態間を明確に切り替わります。しかし、そのまったく同じ論理回路を大腸菌（Escherichia coli）にクローニングし、蛍光顕微鏡下で観察すると、読み取り結果にはノイズが現れます。一部の細胞が明るく蛍光を発する一方で、隣接する同一クローンは暗いままです。同じ培養容器内にあり、まったく同じDNAプラスミドを持つ細胞同士であっても、ばらつきのある異なる応答を示すのです。

遺伝子回路は、電子工学の論理を細胞内の生化学へと変換するものです。この分野は、2000年に『Nature』誌の同一号に掲載された二つの画期的な研究によって幕を開けました。一つは、互いに抑制し合う二つの転写リプレッサーを用いて双安定メモリを保持する「遺伝子トグルスイッチ（Genetic Toggle Switch）」[1]、もう一つは、三つのリプレッサー遺伝子を環状フィードバックループに配置して自律的な周期的振動を生み出す「リプレシレーター（Repressilator）」[2]であり、それぞれ生物学的なメモリレジスタと生物学的時計を実現しました。2016年には「Cello」によって設計の自動化が導入されました。ユーザーがハードウェア記述言語Verilogで回路論理を指定すると、ソフトウェアがあらかじめ特性定義されたNOTゲートやNORゲートを用いて、具体的なDNA配列へとコンパイルします[3]。半導体設計自動化のワークフローが、そのまま生きた細胞内へと移植されたのです。

しかし、生きた細胞は絶縁されたシリコン基板ではありません。細胞内分子は絶えずブラウン運動を行っており、転写や翻訳は本質的に確率論的な現象となります。転写リプレッサーのコピー数が一桁台前半（わずか数個）にまで低下すると、たった1分子のタンパク質の確率的な結合や解離によって出力状態が反転してしまいます。このノイズをモデル化するには、化学マスター方程式（Chemical Master Equation）とギレスピー（Gillespie）の確率的シミュレーション[4]が必要となり、モンテカルロサンプリングを通じて個々の反応イベントをシミュレーションします。また、細胞内のリソース競合も制約をもたらします。活性を持つリボソームやRNAポリメラーゼの総量は限られており、それらに合成回路の発現という負荷をかけると、宿主の増殖速度が低下します。さらに、下流の遺伝子モジュールは上流の駆動要素に対して「レトロアクティビティ（Retroactivity：逆行性負荷）」を及ぼします[5]。エレクトロニクスにおいては、絶縁された部品同士は接続されても独立して動作しますが、細胞内では下流の負荷が加わることで上流モジュールの反応速度論的特性が変化してしまいます。これは、地下鉄の乗換通路で発生した混雑がボトルネックとなり、遠く離れたホームの乗降ペースまで低下させてしまう現象に似ています。

2026年5月14日に発表されたGenCircuit-RLは、この課題を自動コード生成の問題として再構成しました[6]。このモデルは、生のDNA文字列を直接生成するのではなく、pysbol3ライブラリを対象としたPythonスクリプトを記述し、それを実行することで標準的なSynthetic Biology Open Language（SBOL）形式の回路設計を生成します。機械可読な設計ファイルには、構造的なパーツの接続関係や制御トポロジーが記録されます。この中間表現によってモジュール単位での検証が可能になります。長いDNA配列の文字列を人間が読み解くのは困難ですが、SBOLファイルはどのプロモーターがどのリプレッサーを駆動しているかを明示し、解釈しやすい回路図を生成してくれます。

その中核となるアーキテクチャは、階層型報酬の定式化を中心に据えています。回路の成否に対して0または1を与える情報量の乏しい二値報酬ではなく、GenCircuit-RLは検証を5つの階層的な段階（Pythonの構文実行、SBOL仕様への準拠、有効な遺伝子パーツの割り当て、標的論理関数との構造的な回路トポロジーの一致）に分解します。学習は、単純な単一ゲート設計から複雑な複数入力論理ネットワークへと進む4段階のカリキュラムに従います。著者らは、9つのタスク定式化にわたる6つの標準的な回路クラスの4,753個の回路を含むベンチマークスイート「SynBio-Reason」を導入しました。階層型報酬で学習されたモデルは、機能推論タスクにおいて二値報酬のベースラインを14〜16パーセントポイント上回りましたが、カリキュラムを取り除くと最適化は破綻しました。

代謝負荷は、実験室において物理的な形をとって現れます。重い人工回路を保持する大腸菌（E. coli）は、液体培養での増殖が遅くなります。回路の発現が強まるにつれて細胞の増殖は遅くなり、標準的な2日間の培養期間が4日間に延び、選択圧によって細胞は変異を起こして合成遺伝子を切り出すようになります。その結果、最適化の報酬は、論理的な忠実性と、細胞負荷およびSN比（不均一な単一細胞間におけるオン状態とオフ状態の間の動的な誘導倍率の範囲を測定したもの）とのバランスを取る必要があります。ソフトウェア上で論理真理値表をクリアすることは最初のチェックポイントにすぎず、最終到達点ではありません。

階層型報酬の教育的な有用性は明らかです。合格・不合格という二値の成績を受け取る学生は具体的な誤りを特定できませんが、項目別の評価シートがあれば、論理のどの部分で失敗したかを特定できます。強化学習エージェントもまったく同様に動作します。コードが実行され、パーツが正しく割り当てられた一方で、トポロジー的な接続性に問題があったことを知ることで、パラメータ最適化が適切に導かれます。未知の遺伝子パーツで評価されたGenCircuit-RLは、トポロジー的に有効な回路を生成し、人間の介入なしに教科書に載っているような標準的な回路トポロジーを自律的に再発見しました。

相補的なアーキテクチャも同様の解決策に到達しました。2026年1月に発表されたGenAI-Netは、生成的な反応提案エージェントと数値シミュレーション環境を組み合わせることで、生体分子反応ネットワークを設計します[7]。自然言語による挙動の仕様が与えられると、生成エージェントが反応ネットワークを提案し、常微分方程式（ODE）シミュレーターと確率的シミュレーターが動力学的出力をスコアリングします。これにより、決定論的領域と確率論的領域の双方において、用量反応曲線、論理ゲート、発振器、堅牢な完全適応ネットワークを設計します。同一の機能仕様を満たす多様な候補トポロジーを生成することは、実験における冗長性をもたらします。もし実験台（ベンチ）で1つのトポロジーが機能しなかった場合でも、代替の候補が残されているためです。

制御DNA工学において、Ctrl-DNAは制約付き強化学習を適用して細胞タイプ特異的な合成プロモーターやエンハンサーを設計し[8]、非標的系統における漏出性発現を抑えつつ、標的細胞タイプにおける転写活性化を同時に最大化しました。生成された配列には、標的細胞の発現プロファイルに合致する真正な転写因子結合モチーフが含まれていました。これと並行して外部からの回路制御も進歩しました。光パルスを用いて遺伝子転写を制御するオプトジェネティクス（光遺伝学）は、従来、急峻な非線形応答曲線への対処に苦慮していましたが、2026年の制御理論研究ではパルス幅変調（PWM）によってこの課題を解決しました[9]。光強度を調整することなく、さまざまなデューティ比で照射を高速にパルス化することで、平均発現レベルを滑らかに調整できるようにしたのです。

シミュレーターは依然として数学的な近似にとどまり、報酬関数は研究者が指定する実験的パラメータに依存します。2026年1月に『Nature』誌に掲載された研究は、こうしたモデルをハイスループットな実証データに基づかせました[10]。ライス大学の研究チームはCLASSICを開発し、数十万種類に及ぶ多様な遺伝子回路の設計をヒト細胞に同時に導入しました。ロングリードシーケンシングによってどの細胞がどの回路変異体を受け取ったかを追跡し、ショートリードRNAシーケンシングによって設計ごとの転写出力を定量化することで、回路レイアウトと生物学的性能を結びつける包括的なデータセットをマッピングしました。このデータセットで学習させたモデルは、未知の回路設計の性能を正確に予測しました。研究チームが予測された40種類の回路を実際に合成して実験室で検証したところ、40種類すべてが計算による予測と一致しました。

このようなハイスループットマッピングは、設計サイクルを一変させます。以前は、研究者が30個の候補回路を選び、分子クローニングに何週間も費やして性能を評価し、次の候補群を選ぶという手順を踏んでいました。現在では、モデルが数万個の候補回路の予測性能をインシリコで評価し、物理的なクローニングとアッセイによる検証を行う少数の候補を優先順位付けします。物理的な実験がなくなるわけではありません。その位置づけが、初期の探索的スクリーニングから最終的な標的検証へと移行するのです。

遺伝子回路工学における人工知能の実質的な貢献は、構造化された階層的検証とハイスループットな実証的ベンチマークにあります。ソフトウェア上の理想化された矩形波と、ノイズを伴う細胞の増殖曲線との間の隔たりは依然として存在します。しかし現在では、その隔たりを体系的に埋めるための標準的な測定指標と自動設計ツールが存在しています。

参考文献

[1] Gardner, T. S., Cantor, C. R., & Collins, J. J. (2000). Construction of a genetic toggle switch in Escherichia coli. Nature, 403(6767), 339-342. https://doi.org/10.1038/35002131

[2] Elowitz, M. B., & Leibler, S. (2000). A synthetic oscillatory network of transcriptional regulators. Nature, 403(6767), 335-338. https://doi.org/10.1038/35002125

[3] Nielsen, A. A. K., Der, B. S., Shin, J., et al. (2016). Genetic circuit design automation. Science, 352(6281), aac7341. https://doi.org/10.1126/science.aac7341

[4] Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry, 81(25), 2340-2361. https://doi.org/10.1021/j100540a008

[5] Del Vecchio, D., Ninfa, A. J., & Sontag, E. D. (2008). Modular cell biology: retroactivity and insulation. Molecular Systems Biology, 4, 161. https://doi.org/10.1038/msb4100204

[6] Flynn, N. (2026). GenCircuit-RL: Reinforcement Learning from Hierarchical Verification for Genetic Circuit Design. arXiv preprint arXiv:2605.14215. https://arxiv.org/abs/2605.14215

[7] Filo, M., Rossi, N., Fang, Z., & Khammash, M. (2026). GenAI-Net: A Generative AI Framework for Automated Biomolecular Network Design. arXiv preprint arXiv:2601.17582. https://arxiv.org/abs/2601.17582

[8] Chen, X., Ma, S., Lin, R., Lin, J., & Wang, B. (2025). Ctrl-DNA: Controllable Cell-Type-Specific Regulatory DNA Design via Constrained RL. arXiv preprint arXiv:2505.20578. https://arxiv.org/abs/2505.20578

[9] Espinel-Ríos, S. (2026). Switching-time bioprocess control with pulse-width-modulated optogenetics. arXiv preprint arXiv:2511.22893. https://arxiv.org/abs/2511.22893

[10] Rai, K., O'Connell, R., et al. (2026). Ultra-high-throughput mapping of genetic design space. Nature, 650, 1035-1044. https://doi.org/10.1038/s41586-025-09933-9

第12章：未来の自律型・自律駆動型ラボ（SDL）

12.1 リキッドハンドリングロボットのためのスクリプト自動導出と最適化された自律的PCR・実験実行（AutoPCR）

深夜2時の研究室を思い浮かべてみてください。大学院生が384ウェルプレートの前に座っています。繰り返されるピペッティング操作により、右手の親指は感覚を失っています。プランジャーを押し込み、戻し、再び押し込みます。この動作は本日だけで4,000回以上繰り返されました。各ウェルに吐出される液量は5マイクロリットルであり、涙1滴のおよそ10分の1の量にすぎません。

根本的な問題は親指の疲労ではなく、データのばらつきです。午前のうちに手際よくピペッティングされたプレートと、疲れ切った研究者が深夜に準備したプレートとでは、得られる測定値が異なります。ピペッティング時の傾き角度、チップの液体への浸漬深度、プランジャーを戻す速度のわずかな違いが、吐出液量を変化させます。火曜日と金曜日にまったく同じプロトコルを実行しても、微妙な食い違いが生じます。生命科学における再現性の危機は、理論的な概念の誤りから生じたのではなく、人間の指先の震えから始まったのです。

この危機の規模は2016年に数値化されました。『Nature』誌が1,576人の研究者を対象に、他の科学者が発表した実験を再現できなかったことがあるかを調査したところ、70%以上が「ある」と回答しました。さらに半数以上が、自分自身の過去の実験結果さえ再現できなかった経験を報告しています[6]。典型的な実験手法（Methods）の記述には、「試薬をよく混合し、室温で30分間インキュベートした」といったものがあります。何回の垂直反転を行えば「よく混合した」ことになるのでしょうか。「室温」とは22℃でしょうか、それとも26℃でしょうか。明文化されていない暗黙の慣習が、そうした手法の記述の曖昧さを埋めていたのです。

液体はそれぞれ異なる物理的・レオロジー的特性を示します。純水はスムーズに吸引されます。粘性の高いグリセロールはゆっくりと吸い上がり、プラスチック製ピペットチップの内壁に付着します。界面活性剤の溶液は吸引時に泡立ちを生じます。人間の実験者は触覚を伴う経験を通じてこれらの違いを学びますが、ロボットは直感的に学習することができません。その代わりに、ロボット工学ではリキッドクラス（Liquid Classes）と呼ばれる仕組みを用います。これは特定の試薬クラスに対して、吸引速度、待機時間、ブローアウト圧力、液滴のタッチオフパラメータなどを指定するキャリブレーション参照テーブルです。正確に調整されていればロボットの精度は人間の能力を凌駕しますが、調整が不十分であれば、ロボットは1つの誤りを10,000回そのまま繰り返すことになります。

ロボットは退屈を感じません。これこそがリキッドハンドリング自動化の核心的な利点です。この概念には歴史的な経緯があります。2009年、アベリストウィス大学のRoss King率いる研究チ

ームは、ロボット「Adam」を導入しました。Adamは酵母のオーファン遺伝子の機能に関する仮説を自律的に立て、アッセイを設計し、実験の測定結果をもとに仮説を更新しました。これは人間の介入なしに新たな科学的発見を生み出した最初の機械となりました[7]。その後の数十年間にわたり、絶え間ない障壁となっていたのはハードウェアの断片化でした。Hamilton社製ロボット用に書かれたスクリプトコードはTecan社のプラットフォームでは動作せず、研究室のハードウェアを変更するたびにプロトコルを書き直す必要がありました。PyLabRobotは、オープンソースでハードウェアに依存しないPythonインターフェースを提供することでこの断片化を解消し[1]、統一されたコマンド体系を用いてHamilton STAR、Tecan EVO、Opentrons OT-2システムを制御できるようにしました。デッキレイアウトや実験器具（ラボウェア）を共通のコードとして標準化することで、各研究室は検証済みのプロトコルを直接ダウンロードして実行できるようになり、実験手法の記述は曖昧な文章から実行可能なスクリプトへと変貌を遂げたのです。

ロボットによる実験の実行は、2025年後半にさらなる進展を見せました。研究者らは、ポリメラーゼ連鎖反応（PCR）増幅を、静的なサーマルサイクラーの加熱ブロック操作から、動的なモーションコントロールプロセスへと再構築しました [2]。ペルチェ素子による加熱ブロックの温度を上下させるのではなく、ロボットアームが、垂直方向に一定の温度勾配が設定された静止オイルバスの中に、密封されたピペットチップを繰り返し浸漬させます。チップを深く下げることで変性温度に達し、少し引き上げることでアニーリングおよび伸長反応の温度帯に到達します。ロボットアームの垂直方向の移動によってサーマルサイクルが実行される仕組みです。このシステムは、密封されたチップバレルによってサンプルの蒸発や交差汚染を排除しながら、増幅効率とシーケンシングの正確さにおいて市販のハイエンドサーマルサイクラーと同等の性能を達成しました。PCRのスループット向上は、サーマルサイクラーブロックを追加購入することから、ロボットの動作スケジュールを最適化することへと移行しつつあります。

公開されたプロトコルを実行可能なコードへと変換する作業は、マルチエージェントアーキテクチャを通じて自動化されつつあります。2026年6月に発表されたProtoPilotは、文献の手順のパース、物理的なハードウェア制約の確認、実行スクリプトの生成、そして実行失敗時の反復的なエラーデバッグという作業を、専門化されたAIエージェント間で分担します [3]。98件の標準作業手順書から抽出された294のタスクで評価したところ、上位3つの候補プロトコルのうち90.2%がドメイン専門家によって実行可能と判定されました。また、コード検証ゲートの通過率は89.5%、Opentronsの実機における実行成功率は88.24%に達し、従来のベースラインAIツールの性能（32.35%）を大きく上回りました。組み立てられたDNA産物は、サンガーシーケンシングによって検証されました。

ハードウェアとソフトウェアの通信には、標準化されたプロトコルが必要です。2026年6月に導入されたLaboratory Agent Protocol（LAP）は、AIエージェントと物理機器との間のオープンな通信仕様を確立しました [4]。Model Context Protocol（MCP）がエージェントをソフトウェアツールと接続し、A2Aがエージェント群同士を接続する一方で、エージェントとウェットラボのハードウェアとの間の通信インターフェースには標準仕様が欠けていました。物理的な実験室での行動は不可逆的であり、一度こぼれた試薬を元に戻すことはできません。LAPは、署名付きの機器機能マニフェスト、シングルテナントによる機器の占有、危険な操作に対する人間の承認トークン、ならびに校正メタデータと不確実性の範囲を含む標準化された単位アノテーションを義務付けています。

マシンビジョンシステムは、リアルタイムの運用モニタリングを提供します。2026年に市販されているリキッドハンドラーには統合カメラが組み込まれており、メニスカスの液面の高さを測定し、気泡やチップの詰まりを検知すると直ちに吸引を停止します。デッキ上におけるわずか数ミリメートルのプレートの位置ずれも、リアルタイムのコンピュータビジョンによってフラグが立てられ、無人で行われる夜間バッチ処理の失敗を未然に防ぎます。

ロボットシステムも故障やエラーと無縁ではありません。チップが詰まることもあれば、グリッパーがマイクロプレートを落とすこともあり、試薬リザーバーが予定より早く枯渇することもあります。自動化の真の価値は、監査可能な網羅的ログを生成できる点にあります。人間の技術者がピペッティングのわずかなずれを記録することは稀ですが、自動化ロボットはすべての工程においてタイムスタンプ付きのセンサーテレメトリを記録します。

実験メタデータの標準化も進んでいます。2026年7月に公開されたSciSchema.orgは、科学的プロセスの記述を構造化するためのスキーマを確立しました [8]。実験変数を標準化されたフィールドに構造化することで、遠隔地のロボットが複数の分散拠点でプロトコルをパースして実行できるようになり、発表された論文を直接実行可能なワークフローへと変換することが可能になります。

運用規模の拡大は、実験室の力学を変化させます。Ginkgo Bioworksは2026年3月にクラウドラボを開設し、ボストンにある105台の自動ロボットワークセルを接続しました[5]。実証ベンチマークにおいて、同プラットフォームは3万6,000件の生体反応を実行し、15万点のアッセイデータポイントを生成しました。この運用の転換の中心にあるのは稼働率です。人間が作業する実験室が週40時間稼働するのに対し、自律型クラウドラボは168時間連続で稼働し、必要な物理的設置面積を3分の1に削減します。

アクセスの敷居は低くなりました。かつては何十万ドルもの費用がかかり、専任のオートメーションエンジニアが必要でしたが、Opentrons Flexのような最新のベンチトップ機器により、大学院生でも短いPythonスクリプトを用いてカスタムプロトコルを実行できるようになりました。これにより、自動化は企業の施設を超えて学術機関のラボへと広がっています。

自動化は人間の責任をなくすのではなく、変化させます。自動化プラットフォームの運用には、リキッドクラスのキャリブレーション、デッキレイアウトの設定、コードの記述、エラーログのデバッグが必要です。手作業によるピペッティングは、計算論的なプロトコル設計へと置き換えられます。

ロボットは生物学的な直感を持っていません。その真価は決定論的な実行力、すなわち1万回目の分注サイクルを1回目とまったく同じように実行できる点にあります。実験台での暗黙知がオープンなソフトウェアスクリプトとしてコード化されれば、地球の裏側でもまったく同じ忠実度で実験を実行できます。

参考文献

[1] Wierenga, R. P., Golas, S. M., Ho, W., Coley, C. W., & Esvelt, K. M. (2023). PyLabRobot: An open-source, hardware-agnostic interface for liquid-handling robots and accessories. Device, 1(4), 100111. https://doi.org/10.1016/j.device.2023.100111

[2] Beguin, V., Grétillat, J., Kaminskaite, K., Juzenas, S., Burgi, P.-Y., & Charmet, J., et al. (2025). High-fidelity robotic PCR amplification. arXiv preprint arXiv:2512.23877. https://arxiv.org/abs/2512.23877

[3] Jiang, Y., et al. (2026). A Self-Evolving Agentic System for Automated Generation and Execution of Biological Protocols (ProtoPilot). arXiv preprint arXiv:2606.31763. https://arxiv.org/abs/2606.31763

[4] Zhu, L., et al. (2026). LAP: An Agent-to-Instrument Protocol for Autonomous Science. arXiv preprint arXiv:2606.03755. https://arxiv.org/abs/2606.03755

[5] Ginkgo Bioworks. (March 2, 2026). Ginkgo Bioworks launches Ginkgo Cloud Lab, powered by autonomous lab infrastructure. https://www.prnewswire.com/news-releases/ginkgo-bioworks-launches-ginkgo-cloud-lab-powered-by-autonomous-lab-infrastructure-302700458.html

[6] Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. Nature, 533(7604), 452-454. https://doi.org/10.1038/533452a

[7] King, R. D., Rowland, J., Oliver, S. G., et al. (2009). The automation of science. Science, 324(5923), 85-89. https://doi.org/10.1126/science.1165620

[8] D'Souza, J., et al. (2026). SciSchema.org: A Multidisciplinary Collection of Schemas for Structured Scientific Process Descriptions. arXiv preprint arXiv:2607.27955. https://arxiv.org/abs/2607.27955

[9] Holland, I., & Davies, J. A. (2020). Automation in the life science research laboratory. Frontiers in Bioengineering and Biotechnology, 8, 571777. https://doi.org/10.3389/fbioe.2020.571777

12.2 実験的検証と連携したマルチエージェント自律型クローズドループ創薬（Latent-Y）

2026年3月下旬、ロンドンの研究チームはAIシステムに、「指定された疾患タンパク質を標的とする高親和性抗体を設計せよ」という自然言語プロンプトを与えました。そのプロンプトの後、研究者は一切介入しませんでした。システムは公開された文献を精査し、標的タンパク質の構造を解析し、結合エピトープをマッピングし、候補配列を生成し、計算によって設計をフィルタリングし、実験室での合成に向けた候補を選定しました。

そのシステムがLatent-Yです[1]。実験的検証によってその性能が確認されました。9種類の異なる標的タンパク質にわたり、システムは6種類について検証済みのナノボディバインダーを生成し、標的あたりの成功率は67%に達しました。結合親和性は1桁ナノモル（nM）台のKd値に達しました。システムは人間の指導なしに候補を選定し、人間の専門家の作業期間と比べて56倍の速さで創薬キャンペーンを完了させ[2]、数週間を数時間へと圧縮しました。

中核となる生成エンジンは、原子分解能の分子拡散モデルであるLatent-X2です。これは環状ペプチドや小型タンパク質バインダーの設計にも適応可能であり、単一のエージェントが治療モダリティを横断して設計を行うことを可能にしています。

67%の成功率は、9つの標的のうち3つではバインダーが得られなかったことも意味しています。人間のチームも同様に失敗に直面しますが、実現不可能なプロジェクトを中止するまでに数か月を費やすことが少なくありません。自律型ループは数日で失敗を特定し、別の標的へとリソースを再配分します。自動化されたスピードの価値は、早期に失敗を見極め、実験リソースを再配分できる点にあります。

自律駆動型ラボ（SDL）は、設計（Design）、構築（Build）、試験（Test）、学習（Learn）という「DBTL」の反復ループで動作します。初期のクローズドループの実証は化学合成の分野で現れました。2020年、リバプール大学は自律走行型ロボット化学者を導入し、8日間で688回の光触媒実験を実行して、基準物質の6倍の活性を持つ触媒配合を発見しました[3]。2023年には、ローレンス・バークレー国立研究所のA-Labが17日間の連続稼働で58回の試みのうち41種類の新規標的化合物を合成しました[4]。同年にカーネギーメロン大学のCoscientistは、大規模言語モデル（LLM）を自動合成ハードウェアと連携させました[5]。

DBTLループは、図書館の階段を上るように機能します。1つの階に上がり、本を手に取り、内容を確認し、必要に応じて階段を降りて別の書架を探します。人間が1つの階に何日も費やすのに対し、自動化システムはわずか数時間でこのサイクルを回します。

試験（Test）フェーズは、運用の要となる段階です。設計された抗体配列はDNAプラスミドへと合成され、細胞内で発現し、精製された後、標的タンパク質に対して評価されます。結合キネティクスは表面プラズモン共鳴（SPR）によって測定されます。金センサーチップ上に標的タンパク質を固定し、その表面に抗体溶液を流すことで、光の反射角度の変化から結合速度（kon）と解離速度（koff）をリアルタイムで追跡します。センサーグラムの曲線が次の設計（Design）フェーズへのフィードバックとなり、予測誤差をもとにモデルの更新が進められます。

活性を示さなかった設計も、破棄されることなく保持されます。結合しなかった配列は、配列空間における境界領域を定義します。人間が主導するプロジェクトでは、陰性のアッセイ結果が公開されることはほとんどなく、プロジェクトの終了とともに消滅してしまいます。自律型ループは、陽性結果と陰性結果の双方を統一されたスキーマで保存し、陰性結果が持つ統計的価値を捉えます。

ベイズ最適化は探索ループを統括し、次にテストすべき候補を選択します。霧に包まれた山を進むように、このアルゴリズムは過去にサンプリングした高度に基づいて未観測の地形を推定します。『Digital Discovery』誌に掲載された2026年のベンチマーク研究では、ベイズ最適化によって目的の材料を特定するために必要な実験の反復回数が18分の1に削減されたことが確認されました[6]。抗体工学において、多目的ベイズ最適化は、水溶性、熱安定性、低い免疫原性とともに結合親和性のバランスをとります[7]。

自律型ループは生物学的な発見を推進してきました。2026年6月のマルチスケールAIスウォーム研究では、教師なしで大腸がんのトランスクリプトームをスクリーニングし、インスリン様増殖因子2（IGF2）を5-フルオロウラシル耐性に関連する主要な脆弱性として特定しました[8]。複数のコホートにわたり統計的に有意であったこの知見は、独立したマウス試験におけるin vivoでの腫瘍縮小によって裏付けられました。研究者らは機構的な解釈可能性を重視し、ニューラルモデルと生物物理学的シミュレーターを結合させて、特定の標的が選定された理由を検証しました。

進歩には客観的な見極めが必要です。自律駆動型ラボ（SDL）は、レベル0からレベル5までの自律性に分類されます。大半の商用プラットフォームは、プロトコルがデジタル化されデータが機械可読であるレベル2で稼働しています。異常を検知した際にクローズドループを自律的に調整するレベル3のプラットフォームは依然として稀であり、レベル4は狭いタスク領域に限定され、レベル5の完全な自律性は存在しません[6]。ベイズ最適化は、実験のばらつきが蓄積する多段階のプロトコルへと規模を拡大する際に困難に直面します。Andrew Cooper氏が指摘したように、完全自律型の自律駆動型ラボ（SDL）は依然として開発途上にあります。

能力の向上は安全性への配慮をもたらします。治療薬を継続的に設計できるエンジンは、危険な化合物を同じ速度で設計することもできるからです。2026年7月に導入されたMolSafeEvalは、生成化学における安全リスクを評価するためのベンチマーク指標を確立しました[9]。従来のベンチマークは構造の新規性と親和性を優先する一方で、毒性や反応性の化学的有害性を除外していました。合成が開始されてからでは生成後のスクリーニングでは手遅れとなるため、初期の設計段階に安全フィルターを組み込むことが不可欠です。

規制の枠組みも適応してきました。2026年1月、米国食品医薬品局（FDA）と欧州医薬品庁（EMA）は、規制当局への申請におけるAI生成データの利用に関する共同ガイダンスを発表し、どのロボット機器がどのようなパラメータのもとでデータを記録したかを明記する追跡可能な監査ログを義務付けました。プロトコルの標準化は、臨床応用に向けた規制上の要件となっています。

自律化の効率は分野によって異なります。乾燥した前駆体の加熱を伴う固相無機合成はロボットによる自動化が容易ですが、哺乳類細胞生物学は取り扱いのばらつきや継代数に敏感な生体系を扱います。生物学的アッセイの自動化は技術的に難易度が高く、この点こそが抗体設計におけるLatent-Yの実験的検証が重要である理由を際立たせています。

経済的パラメータも変化しています。カナダのAcceleration Consortiumは2億カナダドル（CAD）の資金を確保し、発見までの期間とコストを10年・1,000万ドルから1年・100万ドルへと圧縮することを目指しています。モジュール式のロボットワークセルはハードウェア単体で1万〜5万ドルの費用がかかり、ソフトウェアの統合にはさらに10万〜50万ドルが加算されます。初期のクラウドラボプロバイダーに見られる事業方針転換は、運用の持続可能性には技術力と並んで実行可能なビジネスモデルが必要であることを示しています。

人間の専門知識は依然として中心にあります。Latent-Yに提供されたプロンプトは人間の科学者が作成したものであり、どの標的タンパク質を対象とすべきか、そしてその理由は何かを特定していました。AIは実験的探索の速度を変えますが、どのような科学的問いを立てるべきかを決定するのは人間の判断です。

参考文献

[1] Schmon, S. M., Pretorius, D., Mathis, S., et al. (2026). Latent-Y: A lab-validated autonomous agent for de novo drug design. arXiv preprint arXiv:2603.29727. https://arxiv.org/abs/2603.29727

[2] Latent Labs. (March 23, 2026). Latent-Y: The autonomous AI agent for drug design at scale. https://www.latentlabs.com/press-release/latent-y/

[3] Burger, B., Maffettone, P. M., Gusev, V. V., et al. (2020). A mobile robotic chemist. Nature, 583(7815), 237-241. https://doi.org/10.1038/s41586-020-2442-2

[4] Szymanski, N. J., Rendy, B., Fei, Y., et al. (2023). An autonomous laboratory for the accelerated synthesis of novel materials. Nature, 624(7990), 86-91. https://doi.org/10.1038/s41586-023-06734-w

[5] Boiko, D. A., MacKnight, R., Kline, B., & Gomes, G. (2023). Autonomous chemical research with large language models. Nature, 624(7992), 570-578. https://doi.org/10.1038/s41586-023-06792-0

[6] Benchmarking self-driving labs. (2026). Digital Discovery, 5(1), 14. https://doi.org/10.1039/D5DD00337G

[7] Rao, J., et al. (2026). BOAT: Navigating the Sea of In Silico Predictors for Antibody Design via Multi-Objective Bayesian Optimization. arXiv preprint arXiv:2604.13980. https://arxiv.org/abs/2604.13980

[8] Baker, C., Ren, T., Rafferty, K., Wang, H., & McDade, S. (2026). Autonomous mechanistic discovery of colorectal cancer vulnerabilities via multi-scale AI swarms. arXiv preprint arXiv:2607.16262. https://arxiv.org/abs/2607.16262

[9] Xu, T., et al. (2026). MolSafeEval: A Benchmark for Uncovering Safety Risks in AI-Generated Molecules. arXiv preprint arXiv:2607.00464. https://arxiv.org/abs/2607.00464

[10] Abolhasani, M., & Kumacheva, E. (2023). The rise of self-driving labs in chemical and materials sciences. Nature Synthesis, 2(6), 483-492. https://doi.org/10.1038/s44160-022-00231-0

第13章：バイオセキュリティ、デュアルユースガバナンス、および生成バイオAIの安全ガイドライン

13.1 AI設計病原体のリスクとウイルス合成シナリオにおけるスクリーニングの極めて重要な役割

2021年の秋、ノースカロライナ州にある小さな計算創薬企業の研究者たちが、国際会議の発表スライドを準備していました。そのテーマは、分子薬理学における生成AIの悪用の可能性でした。具体的な実証を試みるため、彼らは分子設計アルゴリズムにおける1つの動作パラメータを反転させました。毒性を最小化する目的関数を、予測される致死性を最大化する関数に置き換えたのです。コードの調整は、わずか1行を変更するだけでした。彼らは夕方に実行を開始し、オフィスを後にしました。

6時間のうちに、モデルは4万個の候補分子を生成しました[1]。出力されたリストには、既知の化学神経剤とともに、公表されている化学兵器よりも致死性が高いと予測される新規の構造類似体が並んでいました。研究者たちは出力されたデータセットを外部からアクセスできないよう封印し、この出来事を詳述した2ページの短い解説記事を執筆しました[1]。簡潔な長さではありましたが、科学界に与えた影響は即座に現れました。

この実験は、計算薬理学が持つデュアルユースの性質を実証しました。病気を治す治療分子を探索するために使われるアルゴリズムは、生理学的な害をもたらす分子を探索するために使われるアルゴリズムと数学的に同一であり、目的関数の符号が異なるだけです。料理を作るためにも危害を加えるためにも使われる包丁の比喩が、そのまま当てはまります。科学界ではこれを「デュアルユースのジレンマ」と呼んでいます。これは、1つの技術が有益な能力と危険な能力を同時に内包しており、ツールそのものを無効化することなく危険性だけを排除することが不可能な状態を指します。

コンピュータ上の設計と物理的な生物学的実体の間には、大きな隔たりが存在します。モニター上に描画された配列そのものは、直接的な物理的危害をもたらしません。それが現実の脅威となるには、DNA遺伝子合成、宿主細胞への形質転換、培養、精製、そして生物学的拡散が必要です。この物理的なパイプラインの中で最も顕著で狭いボトルネック（チョークポイント）となるのが、商業的な遺伝子合成の注文です。この合成のボトルネックにおいて防御的なスクリーニングを確立することは、効果的な介入ポイントとなります。その結果、各国政府や業界コンソーシアムは、モデルの入出力フィルターと商業用DNA合成注文のスクリーニングという2つのインターフェースに検証チェックポイントを配置しました。

合成コストの低下は、このバイオセキュリティの状況を一変させました。かつてカスタム遺伝子断片の注文は、研究室の予算において大きな支出項目でしたが、今日では日常的な消耗品とほとんど変わらない費用で済みます。低コスト化は注文数の増加をもたらし、審査を要する取引ログ

の数を何倍にも増やしています。合成における安全性の維持は、もはや人手による目視確認だけに依存することはできなくなっています。

多層的なバイオセキュリティスクリーニングは、研究ライフサイクル全体にわたって相互に連動する複数のチェックポイントを確立します。具体的には、事前学習前の学習データセットのフィルタリング、潜在表現内での有害な生成の抑制、推論時におけるユーザープロンプトと出力配列のモニタリング、商業的なDNA合成注文のスクリーニング、そしてクラウドベースの自律ロボットラボの監査などです。それぞれの層は異なる障害モードに対応しています。データセットのフィルタリングだけでは配備後の危険な配列に対するファインチューニングを防ぐことができず、出力テキストフィルターは攻撃者が複数の注文に分散して要求を断片化させた場合に有効性が低下します。単一の完璧なフィルターに頼るよりも、複数の独立した防衛層を維持するほうが効果的です。

2025年10月2日、『Science』誌に掲載された研究によって、標準的な合成スクリーニングプロトコルの脆弱性が明らかになりました[2]。Microsoftの研究者や国際遺伝子合成コンソーシアム（IGSC）の専門家による共著チームは、オープンソースのタンパク質設計ツールを用いて、規制対象となっている毒素や病原体の機能的な変異体を生成しました。ヌクレオチド配列は大幅に変更されていたものの、三次構造と生物学的な病原性は保持されていました。商業用の合成スクリーニングアルゴリズムは、野生型の参照病原体を容易に検出して警告を発した一方で、AIによって再設計された変異体の検出には頻繁に失敗しました。

このリスク軽減プロセスは、計算生物学における協調的な脆弱性開示の先例を打ち立てました。著者らは2023年後半から2年間にわたり、極秘裏に調査を進めました。脆弱性を発見したのち、彼らは10か月をかけて遺伝子合成事業者やバイオセキュリティ機関、政策立案者への説明を行い、研究結果を発表する前に更新されたスクリーニングソフトウェアのパッチを配布しました。産業界の合成事業者に防御用パッチが広く配備された後にのみ論文が一般公開されたことは、サイバーセキュリティ分野におけるゼロデイ脆弱性の修正手順を生物学的ガバナンスへと応用したものです。

合成スクリーニングの進化は、なぜ防衛に複数の層が必要なのかを示しています。初期のスクリーニングプロトコルは、規制病原体のデータベースに対する直接的な配列相同性比較に依存していました。ヌクレオチド配列が同義置換や新規の構造バックボーンを用いて改変されている場合、単純な配列アライメントアルゴリズムでは検出できません。現在、スクリーニングアルゴリズムは、生の配列入力から3次元構造や生物学的機能を直接予測する方向へとシフトしています。これと並行してモデル内部の防衛策も進歩しており、2026年3月の研究では、特定の分類群データセットでタンパク質言語モデルをファインチューニングすると有害な出力が引き起こされ

ることが示され、危険な生成を抑制するための推論時アクティベーションステアリングが提案されました[3]。2026年7月に公開されたMolSafeEvalは、AIが生成した分子の安全リスクを評価するための専用ベンチマーク指標を確立しました[4]。

実証的なペネトレーションテストは、防御フィルターの有効性を測定します。2025年5月、アフリカの研究者らは安全ガードレールを無効化したMoremi Bioエージェントを評価し、制約のないシステムが危険な化学・生物学的設計の要求をそのまま処理してしまうことを確認しました[7]。その目的は危険物を設計することではなく、消防訓練における統制された発煙テストのように、防衛境界線の厚みをベンチマーク評価することでした。研究者らは具体的な危険な出力を伏せつつ、総合的なバイパス率を報告し、責任ある情報開示を維持しました。

事後的な推論フィルターのみに依存することには、構造的な脆弱性が存在します。2025年8月の研究では、既存のバイオセキュリティのガバナンスが推論時の出力フィルターに過度に依存していることが指摘されました[8]。推論フィルターは外部のソフトウェアラッパーであるため、オープンなモデルの重みに直接アクセスできる悪意ある攻撃者は、ガードレールを速やかに解除できてしまいます。したがって、防御策は内部のアーキテクチャ上の制約と、物理的なサプライチェーンの管理の双方に及ばなければなりません。2025年10月には、国際的な研究機関を横断する19名の研究者が、生物学的設計における技術的障壁の低下に対処するためのガバナンスロードマップを発表し、標準化された能力評価、管理されたアクセス制御、および国際的な規制調整の必要性を提唱しました[9]。

こうした状況に合わせて、規制の基準値も改定されてきました。2024年9月、ホワイトハウス科学技術政策局（OSTP）は核酸合成スクリーニングの枠組みを発表し、200塩基対（bp）を超える注文に対する審査を義務付けました[5]。2025年10月13日には、手作業での連結を目的として断片化されたオリゴヌクレオチドを注文するという抜け穴を塞ぐため、この基準値が50塩基対（bp）へと引き下げられました。IGSCは統一スクリーニングプロトコルを更新し、加盟プロバイダーに対して2026年10月24日までにこの50塩基対（bp）の基準値を導入するよう求めました[6]。2025年5月5日の大統領令は各機関に2024年の枠組みの更新を指示しましたが、2026年8月時点でも正式なガイダンスは保留されたままであり、業界コンソーシアムの合意が実務上の標準として機能しています。

スクリーニングアルゴリズムが注文にフラグを立てた場合でも、その取引が自動的に拒否されるわけではありません。人間のバイオセキュリティ担当者が申請内容を審査し、研究者の適格性、所属機関、および物理的な配送先住所を確認して、配送先の施設がバイオセーフティ基準を満たしているかを確かめます。大学のワクチン研究所からの注文と、身元が確認されていない個人からの注文では、取り扱いが異なります。機械のアルゴリズムが配列の相同性にフラグを立て、人

間の担当者が組織的な文脈を解釈します。規制基準値の厳格化に伴い、十分な専門能力を備えたスクリーニング担当者の育成と、コンプライアンス人員の拡充が求められています。

国や地域による規制の差異は、規制上の課題をもたらしています。遺伝子合成スクリーニングが義務付けられている法域がある一方で、任意にとどまる法域もあり、国境を越えて合成注文が迂回される事態が生じています。ベンチトップ型DNA合成機の普及は、ガバナンスにおけるさらなる空白を生み出しています。米国の枠組みでは、機器メーカーがベンチトップ型のハードウェアに内部スクリーニング検証機能を組み込むことを推奨していますが[5]、その遵守は依然として勧告にとどまっています。軍備管理のアナリストたちは、2025年11月にこの規制の空白を指摘しました[10]。分散型のベンチトップ型合成機が普及するにつれて、バイオセキュリティの適用範囲は、中央集権的な注文ポータルから、販売時点におけるハードウェアの流通管理へと拡大されなければなりません。

バイオセキュリティスクリーニングは、科学の進歩速度に対する事務的な摩擦として描かれることがあります。しかし、歴史的な記録はその逆を示しています。4万種の分子を対象とした神経剤の実験が公表された後も計算創薬が進展し続けたのは、研究者たちが危険なデータを封印し、安全手順を確立したためでした。また、タンパク質工学の研究が継続できているのも、『Science』誌の研究著者たちが論文の公開前に合成上の脆弱性に対処したためです。スクリーニングフィルターは科学的発見を妨げるものではなく、オープンな研究を持続させるために必要な安全基準（ベースライン）を確立するものなのです。

参考文献

[1] Urbina, F., Lentzos, F., Invernizzi, C., & Ekins, S. (2022). Dual use of artificial-intelligence-powered drug discovery. Nature Machine Intelligence, 4(3), 189-191. https://doi.org/10.1038/s42256-022-00465-9

[2] Wittmann, B. J., Alexanian, T., Bartling, C., Beal, J., Clore, A., Diggans, J., Flyangolts, K., Gemler, B. T., Mitchell, T., Murphy, S. T., Wheeler, N. E., & Horvitz, E. (2025). Strengthening nucleic acid biosecurity screening against generative protein design tools. Science, 390(6768), 82-87. https://doi.org/10.1126/science.adu8578

[3] Fernández Burda, M., Aranguri, S., Arcuschin Moreno, I., & Ferrante, E. (2026). Inference-Time Toxicity Mitigation in Protein Language Models. arXiv preprint arXiv:2603.04045. https://arxiv.org/abs/2603.04045

[4] Xu, T., Cao, X., Zhu, Z., Ding, K., & Chen, H. (2026). MolSafeEval: A Benchmark for Uncovering Safety Risks in AI-Generated Molecules. arXiv preprint arXiv:2607.00464. https://arxiv.org/abs/2607.00464

[5] Office of Science and Technology Policy. (September 2024). Framework for Nucleic Acid Synthesis Screening. https://aspr.hhs.gov/S3/Documents/OSTP-Nucleic-Acid-Synthesis-Screening-Framework-Sep2024.pdf

[6] International Gene Synthesis Consortium. Harmonized Screening Protocol v3.0. https://genesynthesisconsortium.org/wp-content/uploads/IGSC-Harmonized-Screening-Protocol-v3.0-1.pdf

[7] Hattoh, G., Ayensu, J., Ofori, N. P., Eshun, S., & Akogo, D. (2025). Can Large Language Models Design Biological Weapons? Evaluating Moremi Bio. arXiv preprint arXiv:2505.17154. https://arxiv.org/abs/2505.17154

[8] Feldman, J., & Feldman, T. (2025). Resilient Biosecurity in the Era of AI-Enabled Bioweapons. arXiv preprint arXiv:2509.02610. https://arxiv.org/abs/2509.02610

[9] Zhang, Z., Chakraborty, S., Bedi, A. S., Mathew, E., Saravanan, V., Cong, L., Velasquez, A., Lin-Gibson, S., Blewett, M., Hendrycks, D., London, A. J., Zhong, E., Raphael, B., Dieng, A. B., Ma, J., Xing, E., Altman, R., Church, G., & Wang, M. (2025). Generative AI for Biosciences: Emerging Threats and Roadmap to Biosecurity. arXiv preprint arXiv:2510.15975. https://arxiv.org/abs/2510.15975

[10] Arms Control Association. (November 24, 2025). Regulatory Gaps in Benchtop Nucleic Acid Synthesis Create Biosecurity Vulnerabilities. https://www.armscontrol.org/blog/2025-11-24/regulatory-gaps-benchtop-nucleic-acid-synthesis-create-biosecurity-vulnerabilities

13.2 生物学基盤モデルのための初期安全性専門家統合フレームワーク

建物の安全検査を建設完了後にしか行わないことは、構造的なリスクを生み出します。すでに主要な配管設備が通っている場所に検査官が非常口の増設を命じれば、多額の改修費用が発生することになります。その結果、施設管理者は補助的な消火器を設置し、避難口の誘導標識を掲示して、形式的に規制遵守の承認を得るにとどまります。AIの安全ガバナンスも歴史的にこれと同様の道をたどってきました。一般公開の直前に、学習済みのモデルに対して事後的なレッドチーミングを行って脆弱性を探り、最適化が完了した後に漏れを塞ぐという手法に依存してきたのです。

生物学における事後評価の有効性は、2026年6月のABC-Bench研究によって検証されました[1]。研究者たちは、リキッドハンドリングロボットの制御スクリプト生成、DNAアセンブリ構築物の設計、および遺伝子合成ソフトウェアにおけるスクリーニング回避経路の特定という3つの運用タスクにわたって言語モデルエージェントを評価しました。評価されたエージェントは、3つのタスクすべてにおいて人間の専門家の中央値と同等か、それを上回る能力を示しました。実際のウェットラボでの検証試験では、エージェントが生成したスクリプトが自動分注機上で直接実行されました[1]。生物学的リスクは、単に事実に関する質問へ回答することから、物理的な実験手順を実行することへと拡大しています。

数百億のパラメータを持つ生物学基盤モデルに対して、事後的な安全パッチを適用することは脆弱性を伴います。これらのモデルは線形なヌクレオチド文字列を単に記憶しているのではなく、熱力学的および生化学的な折りたたみ文法を内部の潜在多様体へと圧縮しています。学習テキストから特定の病原性配列の識別子を消去したとしても、その根底にある生物物理学的文法までを取り除くことはできません。このことは、前述の2026年3月の研究でも実証されています。汎用タンパク質モデルを特定の生物分類群でファインチューニングしたところ、明示的な毒性最適化の目標が与えられていなかったにもかかわらず、システムが毒性タンパク質を生成する結果となったのです。

その結果、開発手法は「Security-by-Design（設計段階からのセキュリティ確保）」へと移行しつつあります。これは、学習完了後の配備レビューではなく、初期のデータキュレーションやアーキテクチャ設計の段階からバイオセキュリティの専門家を参画させるアプローチです。専門領域のエキスパートが学習コーパスを精査し、プロジェクトの開始時から病原体配列に対する事前学習フィルターを構築し、評価ベンチマークを調整します。機械学習のアンラーニング技術は、学習済み重みから危険な知識を取り除くことを目指していますが、潜在空間全体にわたって完全

な除去を検証することは依然として困難です。事後的に表現を解きほぐそうと試みるよりも、最初の段階で学習データをフィルタリングする方がはるかに確実です。

生物学におけるデータキュレーションは技術的に極めて複雑です。病原体の病原性因子と相同な配列は、無害な環境微生物にも広く存在しています。過度に厳格なフィルタリングを行うと、ワクチン設計に必要な機能的生物学の知識までモデルから削ぎ落とされてしまいますが、一方で寛容すぎるフィルタリングは危険な合成経路を残してしまいます。こうしたフィルタリングの閾値を適切に調整するには、進化学的な配列コンテクストを理解する微生物学者の存在が必要であり、監査可能性を維持するための文書化されたキュレーションログも欠かせません。

リスクプロファイルはモデルのアーキテクチャによって異なります。2023年6月の分析では、対話型の大規模言語モデル（LLM）と特化型の生物学的設計ツールの運用上の違いが明確に示されました[6]。対話型モデルは公開された科学文献を統合し、学習への参入障壁を下げます。生物学的設計ツールは新規の機能性分子をde novoで生成します。前者は知識を持つユーザー層を拡大し、後者は機能性生体分子の領域そのものを拡大します。防御的な緩和策はこの違いを反映しなければなりません。生のアミノ酸テンソルを処理するツールに対して、対話型テキストのモニタリングは効果がないためです。

適切な能力ベンチマークを設計することは容易ではありません。選択式の生物学試験で高得点を獲得することが、物理的な実験プロトコルを適切に実行できることと直接相関するわけではないからです。2026年4月の評価では、4つの最先端モデルに対して73件の初級レベルの生物学に関する質問が提示されましたが[7]、モデルは過度な拒絶（正当な教育目的の問い合わせの拒絶）を示すか、根拠のない回答を返すかのいずれかでした。有用性と安全性の境界のバランスをとるには、生物学者と安全性研究者の共同による調整作業が必要です。

モデルの重みの公開ポリシーは、運用上に永続的な影響をもたらします。クラウドホスト型のAPIモデルは危険な悪用が確認された時点で停止できますが、オープンに公開されたモデルの重みは回収することができません。外部からのファインチューニングによって、オープンな重みから安全ラッパーが数時間で除去される可能性があります。100人以上の研究者がリリース前のバイオセキュリティ評価を支持する宣言に署名したものの、一般公開されたモデルのうち文書化された評価を受けていたものは2.5%未満でした。2025年12月にはタンパク質設計のアップデートが前世代の10倍の速度で動作するようになり、リリース前の評価プロトコルが一貫性を欠いたままである一方で、設計のスループットが加速していることが浮き彫りになりました。

モデルレベルのコンテンツフィルタリングだけでこれらの課題を解決することはできません。2026年2月に発表された研究では、キーワードブロックや出力フィルタリングが生物学的設計ツ

ールには不適切であることが強調されました[2]。同じペプチド配列であっても、ワクチン研究においては正当であるのに対し、別の運用コンテキストでは危険となり得るからです。著者らは、バイオセキュリティのインフラとして本人確認（KYC: Know Your Customer）を制度化することを提案しました。これは、高度な設計機能へのアクセスを許可する前に、所属機関、倫理審査委員会の承認、物理的な実験室の所在地を確認する仕組みです。

この運用モデルはマネージドアクセス（Managed Access）として具体化されました。これは、オープンな重みの直接ダウンロードを制限しつつ、認証付きAPIやセキュアなクラウドインターフェースを通じて機能へのアクセスを提供するものです。核脅威イニシアティブ（NTI）は、生物学的設計ツールのリスク階層を分類し、段階的なアクセス検証とユーザー監査プロトコルを確立するマネージドアクセスフレームワークを発表しました[9]。一般的な構造予測ツールには引き続きアクセス可能ですが、de novo生成機能については、事後的な監査証跡を維持するために所属機関の本人確認と改ざん不可能なクエリログ記録が求められます。NTIのホライズンスキャンでは、その後の18か月間が業界の慣行と国際的なガバナンスを整合させるための決定的な期間であると特定されました[3]。

制度的な枠組みの構築が進んでいます。2026年5月5日、米国国立標準技術研究所（NIST）傘下のAI標準・イノベーションセンター（CAISI）は、Google DeepMind、Microsoft、xAIとの間で配備前評価に関する正式な合意を締結しました。これにより、すでに合意していたOpenAIやAnthropicに加えて40件のモデル評価が完了し[8]、機密環境下でデュアルユースのバイオセキュリティ審査が実施されました。産業界では、Anthropicが2025年5月にASL-3安全保護策を発効させ、生物学的リスクを軽減するためのモデル重みセキュリティ対策や配備制限を導入しました[10]。欧州では、大規模な生物学的モデルがEU AI法のシステミックリスク規定に該当するかどうかについて法的分析が行われました[4]。米国では、2026年2月に上院でバイオセキュリティ近代化法案が提出され、あわせて29か国の技術評価をまとめた国際AI安全性報告書が公表されました[5]。

基盤データのキュレーション段階でバイオセキュリティの専門家を参画させることで、混乱を伴う事後的な介入の必要性を減らすことができます。2025年の『Science』誌の論文で特定された合成の脆弱性を是正する際には、ツールがすでに配備されていたため、10か月に及ぶ調整が必要でした。アーキテクチャの最初期段階から安全の専門家を関与させることで、セキュリティ統制が基盤そのものに確実に組み込まれるようになります。

参考文献

[1] Liu, A. B., Nedungadi, S., Cai, B., Kleinman, A., Bhasin, H., & Donoughe, S. (2026). ABC-Bench: An Agentic Bio-Capabilities Benchmark for Biosecurity. arXiv preprint arXiv:2606.11150. https://arxiv.org/abs/2606.11150

[2] Feldman, J., Feldman, T., & Anton, A. I. (2026). Know Your Scientist: KYC as Biosecurity Infrastructure. arXiv preprint arXiv:2602.06172. https://arxiv.org/abs/2602.06172

[3] Nuclear Threat Initiative. (Spring 2026). AIxBio Horizon Scan: Spring 2026. https://www.nti.org/analysis/articles/aixbio-horizon-scan-spring-2026/

[4] Qureshi, T., Griffith, J., Holtman, K., Mir Teijeiro, M., Chin, Z. S., & Gipiškis, R. (2026). The Case for ESM3 as a General-Purpose AI Model with Systemic Risk Under the EU AI Act. arXiv preprint arXiv:2605.01611. https://arxiv.org/abs/2605.01611

[5] Bengio, Y. et al. (2026). International AI Safety Report 2026. arXiv preprint arXiv:2602.21012. https://arxiv.org/abs/2602.21012

[6] Sandbrink, J. B. (2023). Artificial intelligence and biological misuse: Differentiating risks of language models and biological design tools. arXiv preprint arXiv:2306.13952. https://arxiv.org/abs/2306.13952

[7] Richter, M. (2026). Model Capability Assessment and Safeguards for Biological Weaponization. arXiv preprint arXiv:2604.19811. https://arxiv.org/abs/2604.19811

[8] OpenAI. (2026). Working with US CAISI and UK AISI to build more secure AI systems. https://openai.com/index/us-caisi-uk-aisi-ai-update/

[9] Nuclear Threat Initiative. A Framework for Managed Access to Biological AI Tools. https://www.nti.org/analysis/articles/a-framework-for-managed-access-to-biological-ai-tools/

[10] Anthropic. (May 22, 2025). Activating ASL-3 protections. https://www.anthropic.com/news/activating-asl3-protections

第14章：バイオAIのグローバルガバナンス、レギュラトリーサイエンス、人間の説明責任

14.1 EU AI法におけるバイオAIの診断安全性とシステミックリスク管理

2026年7月24日、欧州連合官報（Official Journal of the European Union）に新たな規制法令が公布され、3日後の7月27日に法的効力が発生しました。「デジタルオムニバス（Digital Omnibus）」と名付けられたこの規則は、EU AI法に基づく高リスク人工知能条項の適用スケジュールを延期するものでした[1]。この法制化は、欧州のバイオテクノロジー企業の規制コンプライアンス担当者たちが技術文書を取りまとめていた最中、当初の期限であった8月2日の6日前に行われました。

延期は廃止を意味するものではありません。付属書III（Annex III）に記載されたスタンドアロン型の高リスクAIシステムの遵守義務は2027年12月2日へと移行し、既存のEU調和立法によって規制される製品の安全構成要素として組み込まれたAIシステムは2028年8月2日へと移行しました[1][2]。医療機器ソフトウェアはこの後者の経路に該当します。EU医療機器規則（MDR）の下でクラスIIa以上に分類される診断用AIは、AI法第6条第1項により高リスクに分類され、遵守義務の期限は2028年8月2日となっています[3]。重複した認証手続きを新たに設けるのではなく、AI法の要件が公認機関（Notified Body）による適合性評価に直接統合され、確立された規制枠組みの中で監査チェックリストが拡張される形をとっています。

その他の規制条項は遅滞なく進められました。AIによって生成された合成画像、音声、動画に対する明示的なラベル付けを義務付ける第50条の透明性義務は、予定通り2026年8月2日に発効しました[1]。汎用AI（GPAI）モデルに関するガバナンス義務はその1年前の2025年8月2日に発効しており、欧州委員会による制裁金執行権限は2026年8月2日に発動されました[4]。2025年7月10日、欧州委員会は透明性、著作権ガバナンス、安全性リスク評価を網羅した自主的コンプライアンス枠組みである「汎用AI実践規範（General-Purpose AI Code of Practice）」を公表しました[4]。署名は任意であるものの、署名企業は法令上の義務に対する適合性の法的推定を享受できます。

では、生物学的モデルはこの分類のどこに位置づけられるのでしょうか。2026年5月の分析では、ESM3のような最先端の生物学的基盤モデルを生物学的危険の伝達経路と照らし合わせ、これらのシステムがシステミックリスクをもたらす汎用AI（GPAI）モデルの基準を満たしていると結論付けました[5]。法的な推定閾値は、累積学習計算量が10の25乗浮動小数点演算（FLOPs）を超えることと定義されています。コンパクトなアーキテクチャであっても生物学的デュアルユースのリスクをもたらしうる一方で、より大規模なモデルが無害である場合もあるため、FLOPs閾値には精査や批判的な視線が向けられているものの、法的な閾値は明確な管轄上の境界を定めています。

高リスクに分類されたシステムには、記録保持の義務が課されます。第12条は運用ライフサイクル全体を通じたログの自動記録を義務付けており、第19条は最低6か月間のログ保存を義務付けています。明示的なログ項目には、運用のタイムスタンプ、照会された参照用生物学的データベース、陽性の診断判定を導いた入力データ、そして計算出力を確認・検証した人間の専門家の検証可能な身元情報が含まれます[6]。附属書IVは、より広範な技術文書を10年間保管することを義務付けています。ゲノム変異解釈ソフトウェアの場合、臨床的病原性を判定するために患者のどの変異がどの参照データベースと照合・評価されたのか、ならびに報告書を認定した特定の遺伝学者が誰であるかを記録に残さなければなりません。

規制上の罰則体系は、頻繁に誤って引用されます。全世界年間売上高の7%という罰則の上限は、第5条に基づく禁止対象のAI慣行への違反にのみ適用されます（3,500万ユーロまたは全世界売上高の7%のいずれか高い方を上限とします）。ハイリスクに関する義務や技術文書の要件への違反に対しては、より低い法定上限が定められており、最大1,500万ユーロまたは全世界年間売上高の3%となっています[6]。監査ログの記載漏れによって7%の罰金が科されることはありません。正確な法的階層を理解することが、規制リスク管理における第一歩となります。

コンプライアンス（法令遵守）の範囲はログ記録にとどまりません。第10条は学習データ、検証データ、テストデータセットに対するデータガバナンス基準を義務付けており、地理的起源、人口統計学的代表性、コホートバイアスを技術文書に記録することを求めています[2]。これはゲノミクスにおいて極めて重要な意味を持ちます。公共のゲノムリポジトリは主にヨーロッパ系の祖先を持つ個人をサンプリングしているためです。こうした偏った分布で学習された分類器は、非ヨーロッパ系の患者コホートにおける臨床変異を解釈する際に高いエラー率を示します。さらに第9条は継続的なライフサイクルリスク管理プロトコルを義務付け、第15条は技術的正確性とサイバーセキュリティの堅牢性に関する基準を定めています。ラボにおける実践的な教訓は、データセットの出所、キュレーションパイプライン、生物学的境界条件を記録した検証可能な記録を確立することです。

急速な開発サイクルのなかで技術文書を手作業で管理するのは非効率的です。モデルが年に何度も更新される場合、静的なワープロ文書はすぐに時代遅れになってしまいます。その結果、機械可読な文書化フレームワークが登場しました。2025年に導入されたAI Model Passportは、ヘルスケアAIのデータセットの出所、学習ハイパーパラメータ、検証指標、責任者を標準化された機械可読マニフェストにパッケージ化し、モデルアーティファクトとともにバンドルします[7]。これは、バッチの内容物、製造日、認定生産者を記載した食品の栄養成分表示ラベルに似ています。

規制遵守には運用上の大きな負担が伴います。大手製薬企業は、常勤の法務顧問や品質保証担当者を擁する専任の薬事規制チームを維持しています。一方、研究者5名で構成される大学発スピンオフには、こうした専任部署が存在しません。研究者が技術文書を起草し、ソフトウェアエンジニアが監査ログの保持ポリシーを設定することになります。こうした運用面での格差に加え、欧州整合規格が依然として策定中であるという事実が、コンプライアンス期限を2027年12月および2028年8月へと延長する決定に影響を与えました。技術標準が正式に承認される前に法律上の遵守を強制することは、法的な摩擦を生じさせるからです。

韓国は、欧州のスケジュールに先駆けて法定のAIガバナンスを制定しました。「人工知能基本法」およびその施行令は、2026年1月22日に施行されました [8]。医療診断や生体認証など、市民の生命、健康、基本的人権に直接影響を及ぼす領域は「高影響AI」に分類されています。欧州が高リスク規制の執行を2027年後半へと延期するなか、韓国はいち早く包括的な法定AI監督体制を導入した国となりました。ただし政府は少なくとも1年間の猶予期間を設け、罰則の執行を2027年1月まで延長しており、行政過料の上限も3,000万ウォンに設定されています。

医薬品開発において、規制の文言はリスクに応じたアプローチを踏襲しています。2016年以降、AIコンポーネントを含む500件以上の医薬品承認申請が行われたことを受け、米国FDAは2025年1月、医薬品の安全性、有効性、製造品質にわたる規制上の意思決定を支援するAIモデルの信頼性確立に関するガイダンス案を発表しました。2026年1月14日には、米国FDAと欧州医薬品庁（EMA）が共同で「医薬品開発における適正機械学習基準（GMLP）のための10の指導原則」を発表しました [9]。人間中心の設計、リスクに基づく検証、使用コンテキストの定義、データガバナンス、ライフサイクル監査を重視するこれらの原則は、AIモデルを禁止するものではなく、治験依頼者（スポンサー）に対して指定された使用コンテキスト内でのモデルの信頼性を実証することを義務付けています。

最も高度な対応が求められる規制の交差点は、バイオセキュリティにあります。前述した2025年10月の『Science』誌の研究は、その脆弱性を浮き彫りにしました。オープンソースのタンパク質設計モデルが、規制対象となっている72種の毒素や病原体にわたって76,080個の機能的配列変異体を生成し、それらが標準的な商用DNA合成スクリーニングデータベースをすり抜けたのです [10]。著者らは論文の公開前に脆弱性の修正を調整しました。モデルレベルの出力フィルターだけでは敵対的ファインチューニングに対して脆弱であることを認識し、政策の専門家たちは、顧客身元確認（Know Your Scientist / KYC）を基礎的なバイオセキュリティインフラとして確立することを提唱しています [11]。さらに、生成された分子の化学反応性や毒性リスクを評価するMolSafeEvalのような標準化された安全性ベンチマークによって、これを補完することが求められています [12]。

バイオセキュリティ上の危険は、行政上のコンプライアンス文書とは根本的に異なります。書類上の誤記載は紙の上で訂正できますが、物理的に合成されて発送されたDNAを敵対勢力から回収することはできないからです。そのため、バイオセキュリティは多層防御に依存しています。すなわち、有害な出力が生じないようモデルの学習選好を調整すること、商用DNA合成の注文をスクリーニングすること、そして顧客の所属機関の適格性を確認することです。単一の層で完全に防ぎきれるものはありませんが、相互に補完し合う3つの層を重ねることで攻撃経路を削減できます。これは、外扉を3つの異なる機械式錠前で施錠するのと同様の仕組みです。

法的枠組みの進化は緩やかであり、立法の遅れに直面することも多く、国や地域によっても異なります。しかし、規制が向かう方向は明確です。性能のベンチマークには、今や監査可能性と人間による説明責任が対になって求められています。バイオテクノロジー研究所のサーバーラックには、改ざん不能なログ保存用ストレージが併設されるようになりました。モデルのデプロイはもはや最終段階ではありません。デプロイこそが、規制上の監査証跡が始まる起点となるのです。規制当局の検査官から、6か月前の臨床分類がどのような入力データに基づいて下されたのかを詳細に示す監査証跡を求められた際、システムは改ざん不能な記録を提示できなければなりません。そうした記録を用意しておくことは、システムを設計した科学者や技術者の責務なのです。

参考文献

[1] Gibson Dunn. (July 2026). EU AI Act Omnibus Agreement: Postponed High-Risk Deadlines and Other Key Changes. https://www.gibsondunn.com/eu-ai-act-omnibus-agreement-postponed-high-risk-deadlines-and-other-key-changes/

[2] European Parliament and Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

[3] European Commission. Article 6: Classification Rules for High-Risk AI Systems. EU Artificial Intelligence Act. https://artificialintelligenceact.eu/article/6/

[4] European Commission. (July 10, 2025). The General-Purpose AI Code of Practice. Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai

[5] Qureshi, T., et al. (2026). The Case for ESM3 as a General-Purpose AI Model with Systemic Risk Under the EU AI Act. arXiv preprint arXiv:2605.01611. https://arxiv.org/abs/2605.01611

[6] European Commission. Article 19: Automatically Generated Logs. EU Artificial Intelligence Act. https://artificialintelligenceact.eu/article/19/

[7] Kalokyri, V., et al. (2025). AI Model Passport: Data and System Traceability Framework for Transparent AI in Health. arXiv preprint arXiv:2506.22358. https://arxiv.org/abs/2506.22358

[8] National Information Society Agency (NIA). (January 22, 2026). Enforcement of the AI Framework Act and Release of Subordinate Decrees. https://nia.or.kr/site/nia_kor/ex/bbs/View.do?cbIdx=99835&bcIdx=28987&parentSeq=28987

[9] U.S. Food and Drug Administration. Artificial Intelligence for Drug Development. Center for Drug Evaluation and Research. https://www.fda.gov/about-fda/center-drug-evaluation-and-research-cder/artificial-intelligence-drug-development

[10] Wittmann, B. J., Alexanian, T., et al. (2025). Strengthening nucleic acid biosecurity screening against generative protein design tools. Science. https://doi.org/10.1126/science.adu8578

[11] Feldman, J., et al. (2026). Know Your Scientist: KYC as Biosecurity Infrastructure. arXiv preprint arXiv:2602.06172. https://arxiv.org/abs/2602.06172

[12] Xu, T., et al. (2026). MolSafeEval: A Benchmark for Uncovering Safety Risks in AI-Generated Molecules. arXiv preprint arXiv:2607.00464. https://arxiv.org/abs/2607.00464

14.2 臨床的信頼性、再現性の危機の克服、そして人間の知恵と責任の不可欠な役割

96ウェルマイクロプレートがインキュベーターに入れられます。モニター上では、モデルが99%の予測結合確率を表示しています。3日後、プレートがマイクロプレートリーダーにセットされますが、アッセイは何の生物学的応答も示さず、96ウェルすべての吸光度値は陰性対照と一致します。試薬のロットを変え、細胞株を変更し、濃度勾配を変化させてアッセイを繰り返しても、実験の測定結果は陰性のままです。

このような結果は、ウェットラボの研究では日常茶飯事です。2016年、Nature誌が1,576人の科学者を対象に実施した調査では、70%が他の研究者が発表した知見を再現できなかったと回答し、半数以上が自分自身の実験さえ再現できなかったと報告しました[1]。その11年前、ジョン・イオアニディス（John Ioannidis）は、発表された研究成果の大部分がなぜ誤りであるのかを統計的に証明しました[2]。再現性の危機は人工知能の登場以前から存在していました。機械学習（ML）はこの危機を生み出したのではなく、仮説生成の速度を加速させたのです。

機械学習（ML）は、実験台での手作業のミスではなく、評価設計の欠陥に根ざした独自の方法論的失敗モードをもたらします。RNA二次構造予測を例に考えてみましょう。2026年3月の研究では、学習データと評価データの分割を配列相同性に基づいて厳密に分けたところ、性能のリーダーボードが逆転し、従来の深層学習（DL）モデルやRNA基盤モデルが汎化能力を過大評価していたことが明らかになりました[3]。同様の知見は1細胞トランスクリプトミクスでも得られており、パラメータフリーな表現手法が、下流のベンチマークにおいて複雑な基盤モデルと同等以上の性能を示しました[4]。またゲノム言語モデルにおいては、位置的な摂動テストによって、アーキテクチャが位置に基づく制御ロジックを学習できていないことが判明しました[5]。すなわち、根本的な生物物理学的メカニズムを学習することなく、配列の相関関係を単に暗記していたのです。

構造生物学の画像処理においても、同様の脆弱性が見られます。単粒子クライオ電子顕微鏡（cryo-EM）において、最初の計算ステップは顕微鏡画像の視野から個々のタンパク質粒子の座標を抽出することです。2025年の研究では、純粋な合成ノイズに対してテンプレートマッチングやニューラル粒子ピッカーを適用すると、ノイズからもっともらしい3次元のタンパク質構造が生成されてしまうことが実証されました[6]。これは、モデルが見つけ出すようにあらかじめ設定された特徴を再構成してしまうという、アルゴリズム的確証バイアスを示しています。

データ前処理のパイプラインには、目に見えにくい分析上の分岐が存在します。非標的LC-MSメタボロミクスにおいて、生の質量分析データを整理された特徴量テーブルへと処理する際には、

ベースラインノイズ除去、ピークアライメントアルゴリズム、正規化手法、欠損値補完ルールなど、何十もの分析上の選択が伴います。個々の前処理の選択自体は妥当なものであっても、任意のパイプラインを選択することによって、下流の統計的有意性が変化してしまいます。2026年7月の研究では、考えられるすべての前処理経路を評価し、すべての実行ルートにわたって統計的に堅牢であり続ける生物学的特徴のみを保持する「マルチバース・コンセンサス・パイプライン」が導入されました[7]。これは、最終的な結果を確定する前に、前処理のバリエーションによる出力の感度を測定する手法です。

標準化されたベンチマークインフラの再構築が進められています。JUMP Cell Paintingデータセットは、115テラバイトにも及ぶハイコンテント細胞イメージングデータを網羅しています。しかしそのデータ量の膨大さから、各研究室がデータセットを独自に切り出してモデルを評価することになり、研究間での比較可能性が損なわれていました。2026年8月に登場したJUMP-liteは、同一条件下で公平なモデル評価を行えるようにするため、このコーパスを標準化されたベンチマークパーティションへと圧縮しました[8]。標準化されたベンチマークサブセットを整備することは、有意義なモデル比較を行うための基盤となります。

現代の予測アーキテクチャは、点予測とともに明示的な不確実性の指標を出力します。ゲノミクス全般における不確実性推定手法を評価した2026年8月の実証研究では、その信頼性に大きなばらつきが見られ、特定の手法では実験アッセイの誤差に対する適切なキャリブレーションに失敗していることが報告されました[9]。計算毒性学の分野では、再現性のあるリーダーボードによって、過去のTox21チャレンジの提出モデルが同一の標準化ルールの下で再評価されました[10]。評価の軸は、単なる正解率スコアの最大化から、予測の確信度のキャリブレーションへと移行しつつあります。0.99の確率を出力するモデルと、0.51を出力するモデルは明確に区別されなければなりません。

臨床応用への橋渡しにおける乖離は、依然として課題となっています。学習ラベルは、真の臨床エンドポイントではなく代替指標を反映していることが少なくありません。例えば、in vitroでの細胞生存率を生体内（in vivo）での腫瘍縮小の代替指標としたり、結合親和性を臨床症状の緩和の代替指標としたりする場合です。アッセイ環境、試薬の供給元、あるいは患者コホートが変化すると、代替指標と臨床エンドポイントの間の統計的関係性にズレが生じます。これは「Supervision Drift（教師データのドリフト）」と呼ばれる現象です[11]。データの分布シフトが生じても安定して機能する予測モデルを構築することは、トランスレーショナル情報学における中心的な研究課題となっています。

臨床試験の指標は、このトランスレーショナルなギャップを浮き彫りにしています。AIが見出した分子の第1相臨床試験の成功率は80〜90%に達し、従来の製薬業界のベンチマークを大幅に上

回っています。しかし、第2相臨床試験の成功率は約40%にとどまり、従来の基準値と同水準のままです[12]。第1相はヒトにおける薬物動態と安全性を評価するものであり、これらは構造的な分子設計を通じて最適化される特性です。一方、第2相は治療効果を評価するものであり、選択した標的を制御することがヒト患者の疾患転帰を改善するかどうかを検証します。AIは低分子化合物の生成や化学的最適化を加速させますが、その基礎となる生体標的が治療上有効であるかどうかを判断することは、依然として実証的な臨床の課題です。AIが見出した創薬プログラムのうち173件以上が臨床開発段階に入っていますが、最終的なFDA承認を取得したものはまだ存在しません。探索期間は何年も短縮されましたが、第2相および第3相臨床試験の期間は依然としてヒトの生物学的反応速度に縛られています。ヒト患者における臨床的な薬物応答の評価は、計算アルゴリズムによって短縮できるものではありません。

このことは、統計的相関と生物学的因果関係との違いを明確に浮き彫りにしています。人工知能は既存の学習データセット全体から統計的相関を見出しますが、因果的な事前知識がなければ、未観測の生物学的介入がもたらす結果を推論することはできません。遺伝子制御生物学において確固たる基盤モデルを開発するには、既知の生物物理学的法則をモデルの事前情報として組み込む半メカニスティックなアーキテクチャが必要です[13]。観察データを蓄積するだけでは、因果の方向性は自動的には解決されません。因果推論を行うには、対照群を定義し、交絡変数を固定し、生物学的ノードに系統的な摂動を加えるといった、周到な実験計画が不可欠です。

ある創薬プロジェクトのミーティングを考えてみてください。ホワイトボードには10種類の候補分子がリストアップされています。ある研究者が、そのうちの3つの候補が異なる溶媒マトリックスで調製された外部の化合物ライブラリに由来していることに気づきます。そして、マイクロプレートの隅に追加のビヒクル対照ウェルが配置されます。3日後、そのビヒクル対照ウェルに光学シグナルが現れました。観察された生物活性は、候補の低分子化合物ではなく溶媒マトリックスに由来するものだったのです。溶媒に関するメタデータが学習コーパスから抜け落ちていたため、AIモデルはこの交絡変数を特定できませんでした。

生命科学における永続的な競争優位性は、モデルの重みやコードの中にあるわけではありません。アルゴリズムはオープンに公開され、モデルパラメータは数か月以内に再現されます。永続的な基盤となるのは、体系的に記録されたネガティブデータを含む高品質な縦断的実験データと、適切な生物学的問いを立てる科学的見識という2つの資産です。疾患の病態生理学的メカニズムを理解している研究者は、計算によって得られた1万件の予測結果の中から有望な10個の標的を選び出すことができます。その人間の選択こそが、研究の進展速度を左右するのです。

AIシステムは日々、何万もの科学的仮説を生成します。どの仮説にラボのリソースを投資すべきか、見込みのないプログラムをいつ中止すべきか、そして治療薬候補がいつ臨床投与に対して安

全であるかを判断することは、依然として人間の責任です。この原則は法的なガバナンスにも反映されています。欧州AI法の監査ログに関する規定では、自動化された意思決定を検証した自然人の身元を記録することが明確に義務付けられています。法的な要件は、機械の予測の傍らに人間の署名を求めているのです。

実験ノートの最後の行には、研究者の確認署名欄があります。世界中の研究室で、研究者たちはその欄に署名し、日付を記録しています。「3回目の生物学的反復実験、陰性、アッセイ再設計」。インクの筆圧は次のページにまで届きます。その署名欄に書かれるのは、ニューラルネットワークの名前ではありません。

参考文献

[1] Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. Nature, 533(7604), 452-454. https://doi.org/10.1038/533452a

[2] Ioannidis, J. P. A. (2005). Why most published research findings are false. PLoS Medicine, 2(8), e124. https://doi.org/10.1371/journal.pmed.0020124

[3] Chen, Z., et al. (2026). Fair splits flip the leaderboard: CHANRG reveals limited generalization in RNA secondary-structure prediction. arXiv preprint arXiv:2603.22330. https://arxiv.org/abs/2603.22330

[4] Souza, H., & Mehta, P. (2026). Parameter-free representations outperform single-cell foundation models on downstream benchmarks. arXiv preprint arXiv:2602.16696. https://arxiv.org/abs/2602.16696

[5] Cheng, B., & Zhang, J. (2026). The Mechanistic Invariance Test: Genomic Language Models Fail to Learn Positional Regulatory Logic. arXiv preprint arXiv:2604.06549. https://arxiv.org/abs/2604.06549

[6] Balanov, A., et al. (2025). Structure from Noise: Confirmation Bias in Particle Picking in Structural Biology. arXiv preprint arXiv:2507.03951. https://arxiv.org/abs/2507.03951

[7] Al-Huraibi, M. S., et al. (2026). A multiverse-consensus pipeline for reproducible feature selection in untargeted LC-MS metabolomics. arXiv preprint arXiv:2607.17345. https://arxiv.org/abs/2607.17345

[8] Muñoz, A. F., et al. (2026). JUMP-lite: Compact, reproducible benchmarking of cell representations. arXiv preprint arXiv:2608.07632. https://arxiv.org/abs/2608.07632

[9] Saran, S., et al. (2026). Uncertainty-Aware Deep Learning for Genomics Applications: Insights from an Empirical Study. arXiv preprint arXiv:2608.11054. https://arxiv.org/abs/2608.11054

[10] Ebner, A., et al. (2025). Measuring AI Progress in Drug Discovery: A Reproducible Leaderboard for the Tox21 Challenge. arXiv preprint arXiv:2511.14744. https://arxiv.org/abs/2511.14744

[11] Shoeibi, M., et al. (2026). Learning Stable Predictors from Weak Supervision under Distribution Shift. arXiv preprint arXiv:2604.05002. https://arxiv.org/abs/2604.05002

[12] Jayatunga, M. K. P., Ayers, M., Bruens, L., Jayanth, D., & Meier, C. (2024). How successful are AI-discovered drugs in clinical trials? A first analysis and emerging lessons. Drug Discovery Today, 29(6), 104009. https://doi.org/10.1016/j.drudis.2024.104009

[13] Kovacevic, L., et al. (2025). No Foundations without Foundations -- Why semi-mechanistic models are essential for regulatory biology. arXiv preprint arXiv:2501.19178. https://arxiv.org/abs/2501.19178

人工知能が拓く生命科学の新時代

ISBN |

著者 | 金景珍（キム・ギョンジン）

発行人 | 金景珍（キム・ギョンジン）

発行所 | 金景珍出版

登録番号 | 2025年3月10日 (No. 2025-000015)

所在地 | 大韓民国 ソウル特別市 トンデムング チョンノンロ 91 ペギルビル 304号室 (Seoul, Republic of Korea)

電話番号 | +82-2-6338-1905

メールアドレス | kimkj008@gmail.com

定価 | 20,000ウォン

© Kyung-jin Kim 2026

本書の権利はすべて留保されています。書面による事前の許可なく、本書の全部または一部を複製・送信することを禁じます。

注：本書の内容の一部および編集の統合には、AIツールの支援を受けています。

この本を応援する

この本が役に立ったなら、今後の翻訳・公開版・
オープンAI知識プロジェクトをPayPalで応援できます。

PayPal

https://paypal.me/kimkjkj

QRコードを読み取るか、上のリンクを開いてください。