

AI 特許紹介(40)
AI 特許を学ぶ！究める！
～AI 設計分子のフィルタリング特許～

2022 年 5 月 10 日
河野特許事務所
所長弁理士 河野英仁

「AI 特許紹介」シリーズは、注目すべき AI 特許のポイントを紹介します。熾烈な競争となっている第 4 次産業革命下では AI 技術がキーとなり、この AI 技術・ソリューションを特許として適切に権利化しておくことが重要であることは言うまでもありません。

AI 技術は Google, Microsoft, Amazon を始めとした IT プラットフォーマ、研究機関及び大学から毎週のように新たな手法が提案されており、また AI 技術を活用した新たなソリューションも次々とリリースされています。

本稿では米国先進 IT 企業を中心に、これらの企業から出願された AI 特許に記載された AI テクノロジー・ソリューションのポイントをわかりやすく解説致します。

1.概要

特許出願人 IBM

出願日 2020 年 5 月 21 日

公開日 2021 年 11 月 25 日

公開番号 US2021/0366580

発明の名称 実験室試験用の人工知能設計分子のフィルタリング

580 特許は、AI が生成した AI 設計分子を標的薬剤の特性を有するか否かを分類する分類器を用いて候補医薬品を選択し、候補医薬品と生物学的標的との間の分子相互作用をシミュレートするシミュレータを用いて、ウェットラボ用の最終候補医薬品を評価する技術に関する。

2.特許内容の説明

機械学習 (ML) と人工知能 (AI) は、新規医薬品の開発に際し、新規分子の設計に活発に利用されている。しかしながら、ML/AI を使用して新規医薬品を発見する場合、多くの問題がある。たとえば、不均衡なクラスとノイズの多いラベルが原因で、多くの

ML/AI 分子設計手法は、ウェットラボ実験を使用して合理的に評価するには非常に多くの候補を生成してしまう。一部の ML/AI 分子設計手法では、数千から数十万の候補を生成する。

現在、ウェットラボ環境で単一の候補を合成およびテストするための最小コストは、3000 ドルから 5000 ドルであり、さらに、ウェットラボで 20 人の候補者が合成してテストする平均時間は、約 1 か月にも及び、効率が良いとは言えないという問題がある。

図 1 は、AI 設計分子候補をフィルタリングするためのパイプライン 100 である。

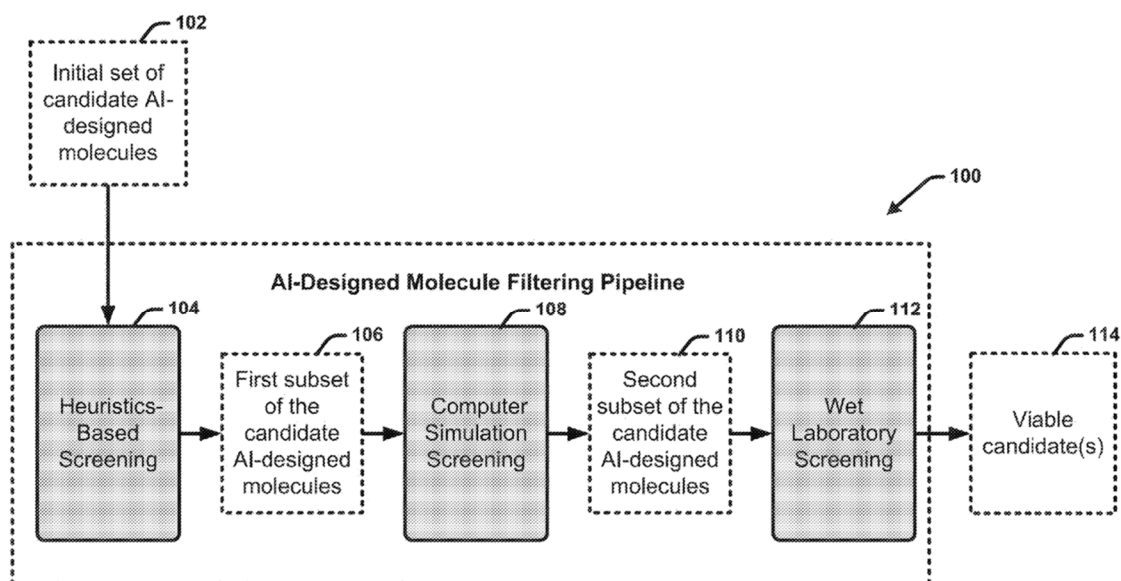


図 1

パイプライン 100 は、3つのスクリーニングを使用し、候補 AI 設計分子の初期セット 102 を実行可能(Viable)な候補 114 にフィルタリングする。3つのスクリーニングフェーズは、ヒューリスティックベースのスクリーニングフェーズ 104、コンピュータシミュレーションスクリーニングフェーズ 108、およびウェットラボスクリーニングフェーズ 112 を含む。パイプライン 100 に従って、ヒューリスティックベースのスクリーニングフェーズ 104 を使用し、分類器を使用する事前定義されたターゲット特徴に基づいて、初期セット 102 から候補の第 1 のサブセット 106 を選択する。

次に、コンピュータシミュレーションスクリーニングフェーズ 108 において、物理駆動型コンピュータシミュレーションを使用して第 1 のサブセット 106 からリード候補 AI 設計分子の第 2 のサブセット 110 を選択し、第 1 のサブセットに含まれるそれぞれの候補の関連する分子動力学を評価する。第 2 のサブセット 110 は、候補がコンピュータシミュレーションにおいて標的行動特性を示すかどうか、及び、どの程度示すかに基

づいて、選択される。

最後に、ウェットラボスクリーニング段階 112 を使用して、第 2 のサブセット 110 に含まれるそれぞれの候補をスクリーニングして、実行可能な候補 114 を特定する。ウェットラボスクリーニングフェーズ 112 は、リード候補を合成し、適切な *in-vitro* および *in-vivo* テストを実行して、リード候補が病原体または別の分子標的に対して実行可能 (viable) か否かを検証する。

図 2 は、AI 設計分子のフィルタリングシステム 200 のブロック図を示す。

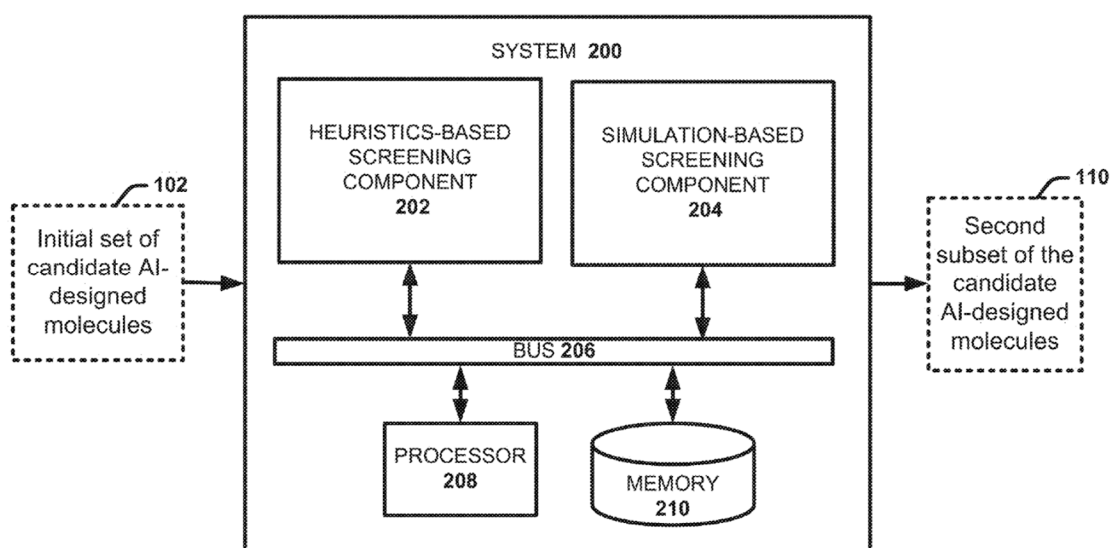


図 2

システム 200 は、AI で設計分子のビッグデータセットを、ウェットラボ実験、新規医薬品のための臨床試験など、ターゲットを絞った有望な候補の非常に小さなデータセットにフィルタリングする。当該目的を容易にするために、システム 200 は、ヒューリスティックベースのスクリーニングコンポーネント 202 およびシミュレーションベースのスクリーニングコンポーネント 204 を含む。

システム 200 は、スクリーニング／フィルタリングのために候補 AI 設計分子の初期セット 102 を受け取る。候補 AI 設計分子の初期セット 102 は、任意の数の候補分子を含むことができる（例えば、数百から数千から数十万以上）。候補 AI 設計分子の初期セット 102 は、診断、治療、治癒、および特定の疾患に関連して特定の生物学的応答を提供するように設計された医薬品を含む。例えば、候補の初期セット 102 は、抗菌剤、抗ウイルス剤、抗癌剤などとして機能するように設計された AI 設計分子を含む。またシステム 200 は、広域スペクトル抗菌ペプチド (AMP: antimicrobial peptides) として

機能するように設計された AI 設計ペプチドをスクリーニングする。

図 3 Aは、ヒューリスティックベースのスクリーニングコンポーネントの例のブロック図を示す。

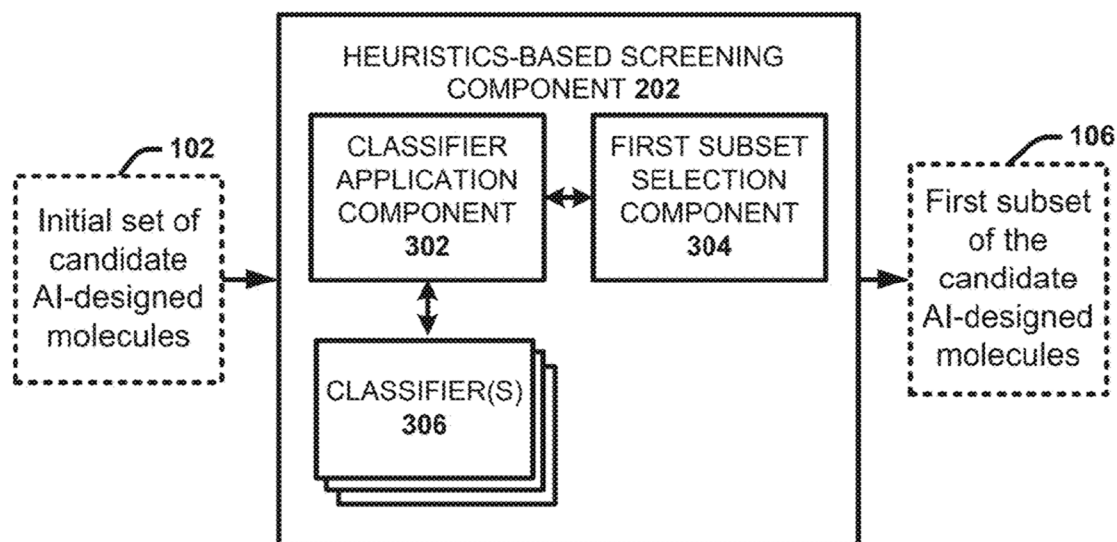


図 3 A

ヒューリスティックベースのスクリーニングコンポーネント 202 は、分類器アプリケーションコンポーネント 302、第 1 のサブセット選択コンポーネント 304、および 1 つまたは複数の分類器 306 を含むことができる。分類器アプリケーションコンポーネント 302 は、分類器を候補 AI 設計分子の初期セット 102 に適用して、初期候補分子のそれぞれが、各分子配列（例えば、タンパク質配列、遺伝子/ヌクレオチド配列、ポリマー配列など）およびそれらの化学構造の分析に基づき、定義された標的特徴を有するかどうかを決定する。例えば、システム 200 を適用して、候補 AI 設計ペプチドをスクリーニングして、最も効果的であり、広域スペクトル抗菌剤を提供する候補 AI 設計ペプチドの小さなサブセットを識別および選択する。

分類器 306 は、定義されたターゲット特徴と、分子配列（例えば、タンパク質配列）または標的の特徴を有する既知の分子の化学構造に反映されたパターンとの間の学習された相関に基づいて、定義されたターゲット特徴を有するかまたは持たないかのいずれかとしてそれぞれの候補を分類するように事前訓練された二項分類モデルを含む。

その他、候補抗菌ペプチドの実装に関して、分類器 306 は、4 つのターゲット機能のそれぞれに 1 つずつ、最大 4 つの別個の分類器を含むことができる（例えば、抗菌機能、広域スペクトルの有効性、毒性が低いかなかったくない、および定義された構造の存在）。

分類器 306 は、長短期記憶（LSTM）ニューラルネットワークベースの分類器などの 1 つまたは複数のディープニューラルネットワークベースの分類器を含む。

図 4 は、候補抗菌ペプチド（AMP）のヒューリスティック分類結果の例である。

Example Heuristics Classification Results for Candidate AMPs

sequence	clfX_amp	clfX_tox	clfX_potency	clfX_broad	clfX_structur
TRGPPPTFRA	0.99999976	0.00046405	1.81E-09	0.82117707	4.74E-05
RRQAQEVRC	0.9997392	1.94E-05	2.92E-10	0.00224675	7.65E-06
IPRPGPFVDP	0.9999405	5.08E-06	3.31E-10	0.07740388	1.25E-06
RPPFGPPFRF	0.99999964	0.02259301	1.15E-09	0.18057713	9.73E-07
KKKKPLTPDF	1	0.00023262	1	0.00870486	8.53E-07
VAKVFRAPK	0.99999917	0.07622346	1.87E-10	0.00184509	1.02E-06
HILRMIRQM	1	1.87E-05	5.04E-10	0.02263985	0.9976706
ILLHAILGVRK	0.99999964	0.10741525	1.85E-09	0.01590915	0.9951361
YRAAMLRRRC	1	0.00013576	3.41E-10	0.17869788	0.9985868
HIRLMRIRQM	1	0.000745	2.82E-10	0.00299489	0.9983133
WEEMDSLRLK	0.9999987	0.00633505	6.95E-10	0.01639144	0.9998312
HIRAMRIRAC	0.99999964	1.15E-05	2.99E-10	0.00109083	0.991021
KTLAQLSAGV	0.99999857	0.00110101	2.73E-10	0.04447917	0.97298425
YLRILIRYMAK	0.99999475	0.00172545	2.89E-10	0.00182426	0.98746395

表 400

図 4

表 400 は、第 1 の列に示されるそれぞれのペプチド配列に基づく複数の候補 AMP 配列への 5 つの異なる分類器の適用を示している。5 つの異なる分類器は、それぞれ「clfX_feature」という表記で特定され、「clf」は頭字語であり、「X」は分類器のトレーニングに使用される特定のトレーニングデータセットを示す。

最初の分類子である clfX_amp（「amp」は「抗菌ペプチド」を表す）は、ペプチド配列が抗菌活性を持っている確率（0.0 から 1.0）である。2 番目の分類子 clfX_tox（「tox」は「毒性」を表す）は、ペプチド配列が毒性である確率（0.0 から 1.0）である。3 番目の分類子である clfX_potency は、ペプチド配列が強力である確率（0.0 から 1.0）である。4 番目の分類子 clfX_broad（「broad」は「広域スペクトル」を表す）は、ペプチド配列が広域スペクトル抗菌剤である確率（0.0 から 1.0）である。5 番目の分類子である clfX_structur（「structur」は「構造」を表す）は、ペプチド配列が二次構造を持つ確率（0.0 から 1.0）である。

図 5 A は、シミュレーションベースのスクリーニングコンポーネントのブロック図を示す。

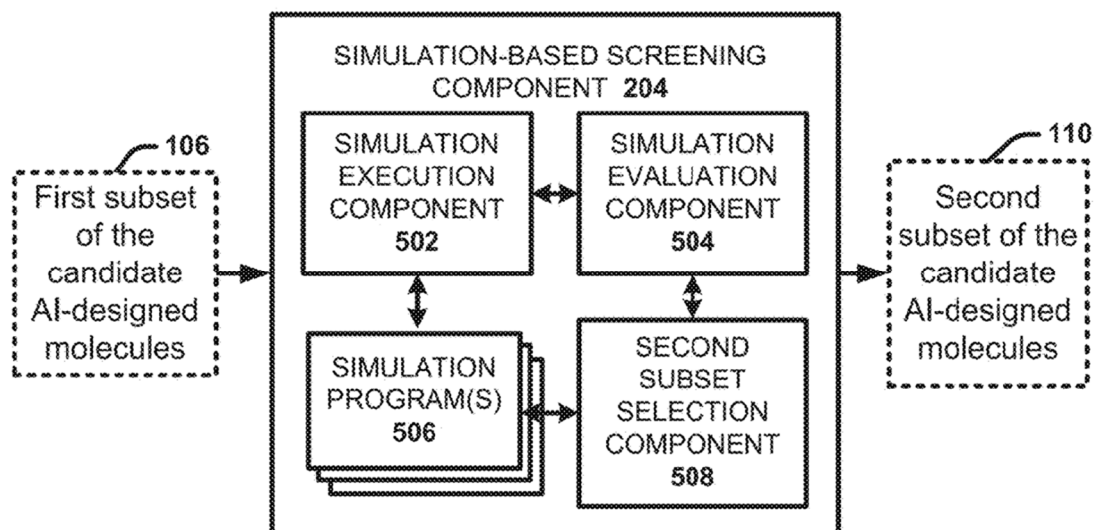


図 5 A

シミュレーションベースのスクリーニングコンポーネント 204 は、第 1 のサブセット 106 に含まれる候補と潜在的な分子および生物学的標的（例えば、病原体の 1 つまたは複数の細胞成分）との間の分子相互作用をシミュレートし、シミュレートされた候補が 1 つまたは複数の所望の相互作用特性を示すかどうか、およびどの程度示すかを決定する。

図 5 A に示すように、シミュレーションベースのスクリーニングコンポーネント 204 は、シミュレーション実行コンポーネント 502、シミュレーション評価コンポーネント 504、1 つまたは複数のシミュレーションプログラム 506、および第 2 のサブセット選択コンポーネント 508 を含む。

シミュレーションプログラム 506 は、物理学に基づく分子相互作用をシミュレートする。シミュレーションプログラム 506 は、それらのモデル化された分子および生物学的構造に基づいて、AI 設計分子と生物学的／分子標的との間の分子相互作用をシミュレートする。例えば、これらのシミュレーションツールは、コースグレイン分子動力学（CGMD: course-grained molecular dynamics）シミュレーションツールなどを含む。

シミュレーション実行コンポーネント 502 は、第 1 のサブセット 106 に含まれるそれぞれの候補に対してシミュレーションを実行する。シミュレーション実行コンポーネント 502 は、第 1 のサブセット 106 に含まれる各候補 AI 設計分子に対して CGMD を実行し、モデル化された分子構造に基づいて、各候補分子と生物学的／分子標的との間

の分子相互作用をシミュレートする。

シミュレーション評価コンポーネント 504 は、それぞれのシミュレーションを評価して、シミュレートされた各候補 AI 設計分子（すなわち、第 1 のサブセット 106 に含まれる各候補分子）が標的分子相互作用/行動特性を示すかどうか、またはどの程度示すかを決定する。たとえば、各シミュレーションの過程でターゲット分子相互作用/動作特性の発生を識別および追跡できる。

第 2 のサブセット選択成分 508 は、シミュレートされた候補分子が標的分子相互作用/挙動特性を示すかどうか、またはその程度に基づいて、候補分子を更に選択して第 2 のサブセット 110 に含める。

候補 AI 設計分子が候補 AMP である場合、候補ペプチドが有望な抗菌剤であるかどうかをスクリーニングするために、シミュレーション実行コンポーネント 502 は、第 1 のサブセット 106 に含まれる候補ペプチドのそれぞれと、モデル脂質二重層または病原体の別の細胞成分との間の相互作用のコンピュータシミュレーションを実行する。

図 6 は、AMP のコースグレイン分子動力学シミュレーションのスナップショットである。

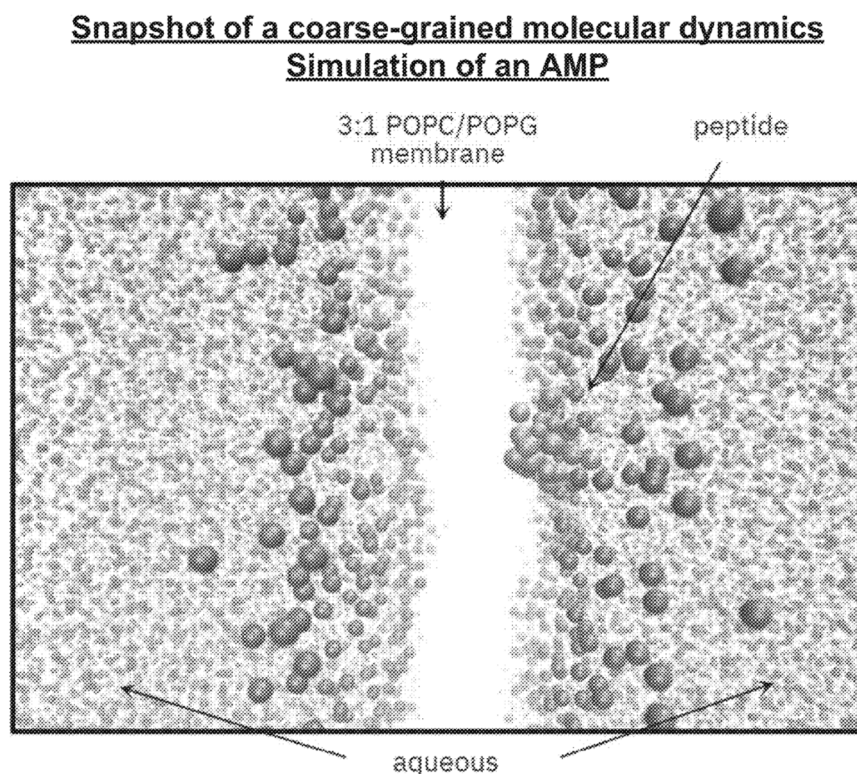


図 6

このシミュレーションでは、モデル化されたペプチドはモデル化された脂質二重層に結合し、ホスファチジルコリン (POPC) とパルミトイルオレオイル PG (POPG) の 3 : 1 混合物である。

図 7 は、候補 AMP のシミュレーション結果を示す。

Example Simulation Results for Candidate AMPs

sequence	length	secondary_str	pos_res	std	mean	binding_time
TRGPPPTFRF	12	coil	3	1.98589547	7.67367257	98
RRQAQEVRC	12	coil	4	1.58225191	6.06144393	351
IPRPGPFVDP	13	coil	3	1.94193763	5.75449735	372
RPPFGPPFER	10	coil	3	1.96882889	7.7845953	236
KKKKPLTPDF	13	coil	4	1.14376825	5.01528662	217
VAKVFRAPKY	14	coil	3	1.83544866	7.39586563	357
HILRMRIQKN	12	helix	3	1.44169014	7.84162437	17
ILLHAILGVRK	13	helix	3	1.19247514	7.15904868	105
YRAAMLRRRC	13	helix	3	1.25045356	8.79043744	19
HILRMRIQKN	12	helix	3	1.49723981	8.38375901	493
WEEMDSLRLK	14	helix	3	1.19813944	6.98079096	117
HIRAMRIRAC	13	helix	3	1.38921803	7.20076151	39
KTLAQLSAGV	14	helix	3	1.45600096	7.61818182	177
YLRIRYMAK	12	helix	3	1.27324818	6.44654882	210

図 7

第 1 の列は候補ペプチド配列であり、第 2 の列はペプチドの長さ、第 3 の列はそれぞれの二次構造、第 4 の列は各配列の正の残基の数、第 5 の列は接点数の標準偏差 (std)、第 6 の列は接点数の平均、第 7 の列は結合時間をナノ秒 (ns) で示す。

3. クレーム

580 特許のクレーム 1 は以下の通りである。

1. システムにおいて、

コンピュータの実行可能コンポーネントを格納するメモリと、

メモリに格納されているコンピュータの実行可能コンポーネントを実行するプロセッサとを備え、前記コンピュータの実行可能コンポーネントは以下を含む：

1 つまたは複数の分類器を使用して人工知能 (AI) で設計された分子のセットを評価し、AI で設計された分子の第 1 のサブセットを候補医薬品として選択する、ヒューリスティックベースのスクリーニングコンポーネントと、

候補医薬品と 1 つまたは複数の生物学的標的との間の分子相互作用の 1 つまたは

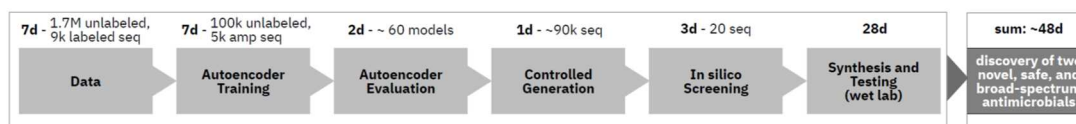
複数のコンピュータシミュレーションを使用して候補医薬品を評価し、ウェットラボ試験用の候補医薬品の第 2 のサブセットを選択するシミュレーションベースのスクリーニングコンポーネントとを備える。

4. 本特許に関連する論文

本特許に関する論文“Accelerating Antimicrobial Discovery with Controllable Deep Generative Models and Molecular Dynamics”¹が、Payel Das 氏らにより公表されている。

論文では CLaSS (Controlled Latent attribute Space Sampling) と称される分子の属性制御生成のための効率的な計算方法を提案している。これは深層学習分類器を原子シミュレーションから派生した新しい機能と組み合わせて使用することにより、生成された分子の重要属性をスクリーニングする。

下記図は、加速抗菌設計のために提案された AI 主導のアプローチの概要とタイムラインである。



論文では、抗生物質耐性に取り組むための新たな薬剤候補である、強力な広域スペクトル効力を有する非毒性抗菌ペプチド (AMP) を特定することを実証した。

わずか 20 の設計された配列の合成とテストにより、膜孔形成を介して、1 つの多剤耐性菌と 1 つの抗生物質耐性 *K. pneumoniae* を含む、多様なグラム陽性菌とグラム陰性菌に対して高い効力を持つ 2 つの新規 AMP が特定された。双方の抗菌剤は、低い invitro および invivo 毒性を示し、薬剤耐性の発症を軽減している。

以上

¹ Payel Das, Tom Sercu, Kahini Wadhawan, Inkit Padhi, Sebastian Gehrmann, Flaviu Cipcigan, Vijil Chenthamarakshan, Hendrik Strobel, Cicero dos Santos, Pin-Yu Chen, Yi Yan Yang, Jeremy P. K. Tan, James Hedrick, Jason Crain, Aleksandra Mojsilovic “Accelerating Antimicrobial Discovery with Controllable Deep Generative Models and Molecular Dynamics” arXiv:2005.11248v2 [cs.LG] 26 Feb 2021

著者紹介

河野英仁

河野特許事務所、所長弁理士。立命館大学情報システム学博士前期課程修了、米国フランクリンピアースローセンター知的財産権法修士修了、中国清華大学法学院知的財産夏季セミナー修了、MIT(マサチューセッツ工科大学)コンピュータ科学・AI 研究所 AI コース修了。

[AI 特許コンサルティング](#)、[医療 AI 特許コンサルティング](#)の他、米国・中国特許の権利化・侵害訴訟を専門としている。著書に「世界のソフトウェア特許(共著)」、「FinTech 特許入門」、「[AI/IoT 特許入門 2.0](#)」、「[ブロックチェーン 3.0](#)(共著)」がある。