研 究 終 了 報 告 書

「タンパク質を制御するペプチドのデザイン AI」

研究期間： 2020 年 12 月～2023 年 3 月
研 究 者： 大上 雅史
加速フェーズ期間： 2023 年 4 月～2024 年 3 月

1. 研究のねらい

薬の開発が高難度化している昨今において、従来の低分子/高分子創薬からの脱却が求められている。酵素や GPCR などの「簡単な」創薬標的は開発し尽くされており、低分子医薬品や高分子医薬品が標的とすることが難しかった細胞内のタンパク質間相互作用（protein-protein interaction, PPI）に対する革新的な分子デザイン技術を開発することが、これからの創薬研究・産業においてきわめて重要である。

本研究は、PPI を標的にできる分子としてペプチドに着目した。サイズの大きすぎない（6～15 残基程度の）ペプチドは、化学合成も容易であり、かつ低分子に比べると分子量が大きいため標的の広い領域をカバーし阻害することができる。また抗体などの高分子は細胞膜を透過することができないが、ペプチドであれば可能なものも存在する。したがって PPI を標的にする場合にペプチドを選択することは合理的である。

これまでに機械学習を用いた分子デザイン技術が国内外で盛んに研究されてきたが、そのほとんどは低分子を対象としており、内容も druggability score や docking score 等の計算値を改善する分子生成手法にとどまっていた（De Cao & Kipf. arXiv, 2018; Brown et al. J Chem Inf Model, 2019;等）。ペプチドデザインとしては、抗菌ペプチドや抗がん作用を持つペプチドを機械学習で予測する試み（Cherkasov, et al. ACS Chem Biol, 2009; Tyagi, et al. Sci Rep, 2013;等）が 2000 年代後半からなされているものの、分子標的薬・PPI 標的薬としての設計はデータ数の問題からほぼ不可能であった。一方で、2020 年にベルン大学のグループが報告した PDGA 法（Capecchi, et al. J Chem Inf Model, 2020）は、入力された分子と特徴空間上で類似したペプチド配列を GA によって効率的に探索するという手法であり、大量の学習用データを必要としないペプチドデザイン技術が初めて提案されたとも言える。しかしながら、既知の PPI 標的化合物が必要という点で、汎用的な PPI 阻害ペプチド設計には不十分であった。

本研究は、PPI 阻害ペプチド分子をデザインする AI 手法を開発することをねらいとした。特にこれまで注目されていなかった PPI 構造におけるタンパク質界面の構造情報を利用した AI を開発することで、適切なペプチド配列を提示することを目的とした。ペプチドが自在に設計できるようになることを見据え、PPI 阻害分子の物性・毒性などの予測を行うための AI 開発も行った。

加速フェーズでは、当初の研究期間で得られた構造予測系 AI モデルである AlphaFold2 のペプチド生成への応用の成果をもとに、開発してきたさまざまな関連 AI を結びつけ、創薬への幅広い適用を目指すための手法開発を行うことを目的とした。特に、ペプチド創薬において代謝安定性や細胞膜透過性の面から注目されている環状ペプチドの設計に注目した。これまで扱ってきた直鎖型のペプチドだけでなく、環状ペプチドも扱うことができれば、ペプチド創薬における分子設計の幅が大きく広がることとなる。加速フェーズでは、AlphaFold2 の positional encoding を改変することで、標的タンパク質に結合する環状ペプチドを設計できることを示した。また、併せ

て低分子化合物設計 AI の開発も行い、ペプチドから低分子への変換を AI によって行うことを目指した。ペプチドは合成・改変などが簡単で検証が容易である一方で、最終的には安定性、生産性、代謝、検証過程など様々な理由により、低分子化合物として導出することが重要である。グラフニューラルネットワークを用いた分子の解釈を高める実装により、所望の物性を持つ低分子化合物の獲得支援を行う技術を開発した。

2. 研究成果

（1）概要

本研究成果は、PPI 界面を標的とするペプチドのデザイン AI に関する成果と、その他の化合物に対する弁別・物性・毒性予測技術に関する成果からなる。

PPI 界面を標的とするペプチドのデザイン AI では、AlphaFold2 の公開もあり、当初の方針を変更して構造予測系 AI を利用する形で進めた。AlphaFold2 におけるペプチド複合体構造の推定スコアを用いてペプチド配列をうまくサンプリングすることで、標的結合能の高いペプチドが得られたが、PPI 界面が疎水性領域に偏っていることから、生成されるペプチド配列も難溶解性のものが多く、生化学実験を行う際に支障があることがわかった。損失関数に水溶性残基のパラメータを重み付き和で加えることで、生成されるペプチド配列の水溶性を高めることに成功し、水溶性を調整することによる標的結合能の低下も起きないことが計算機実験によって示された。

その他の化合物に対する弁別・物性・毒性予測技術に関する成果では、PPI 阻害化合物生成モデルの開発と化合物デザインへの応用、化合物選別向けレコメンド指標の開発、血液胎盤関門透過予測と applicability domain の検討を行った。いずれも現存するツールや手法が適用できず、新たに実装・開発を行った。

以上の成果は、Biomedicines 誌をはじめとする 7 件の論文として発表することができた。また、第 35 回安藤博記念学術奨励賞、および情報処理学会 山下記念研究賞の受賞につながった。

加速フェーズではタンパク質と環状ペプチドの複合体予測のための positional encoding を設計し、高精度のタンパク質–環状ペプチド複合体構造予測が可能な改変 AlphaFold2 を提案した。さらに Deep Network Hallucination 法と組合せて、結合能が高い環状ペプチド配列をデザインできることを示した。また、グラフニューラルネットワークを用いた分子の解釈を高める実装により、所望の物性を持つ低分子化合物の獲得支援を行う技術を開発した。加速フェーズの成果は Int J Mol Sci 誌をはじめとする 5 件の論文として発表したほか、CBI 学会若手奨励賞の受賞につながった。

（2）詳細

研究テーマ A「PPI 界面を標的とするペプチドのデザイン AI の開発」

深層学習による高精度なタンパク質立体構造予測プログラムである AlphaFold2（Jumper+ Nature 2021）が発表され、またタンパク質設計問題（特定の立体構造を取りうる配列を予測する問題＝立体構造予測問題の逆問題に相当）において、立体構造予測のために学習されたニューラルネットワークに back propagation して配列をサンプリングしていく Deep

Network Hallucination 法（Anishchenko+ Nature 2021）が提案された。これらを組合せると、AlphaFold2 におけるペプチド複合体構造の推定スコアを用いてペプチド配列をうまくサンプリングすることで標的結合能の高いペプチドが得られると考えた。しかしながら、実際にいくつか試行した結果、PPI 界面が疎水性領域に偏っていることから、生成されるペプチド配列も難溶解性のものが多く、生化学実験を行う際に支障があることがわかった。そこで、損失関数に水溶性残基のパラメータを重み付き和で加えることで、ペプチドの水溶性をコントロールできるか検証した（図 1）。その結果、生成されるペプチド配列の水溶性を高めることに成功し、水溶性を調整することによる標的結合能の低下も起きないことが計算機実験によって示された（図 2）。ただし、損失関数に利用する残基パラメータ（index）の種類も重要であり、特に Hydropathy index または Solubility weight index を利用すると結合能が低下しないこともわかった（図 3）。生成されたペプチドの頻度分布を確認すると、実際に水溶性の高い残基を中心に設計されていたことが確認できた。これらの成果は Biomedicines 誌に発表した。

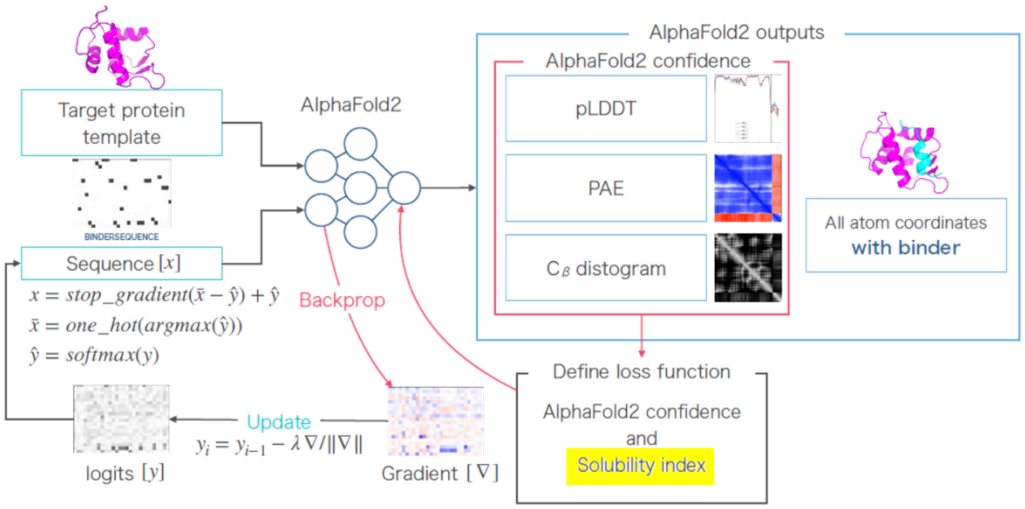


図 1　水溶性を考慮した Hallucination によるペプチド配列設計

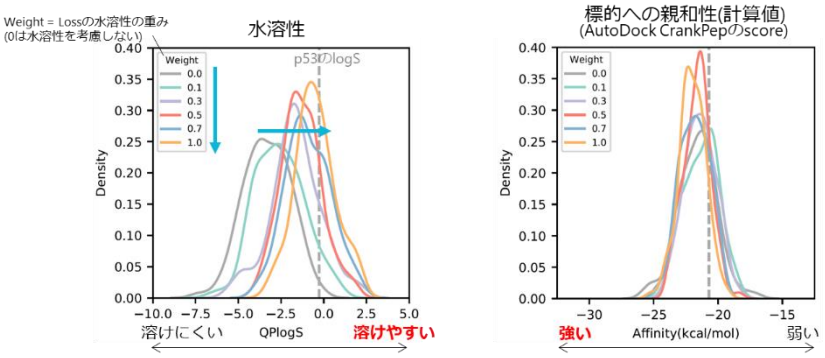


図 2　提案手法によって生成されたペプチド配列の水溶性と標的親和性

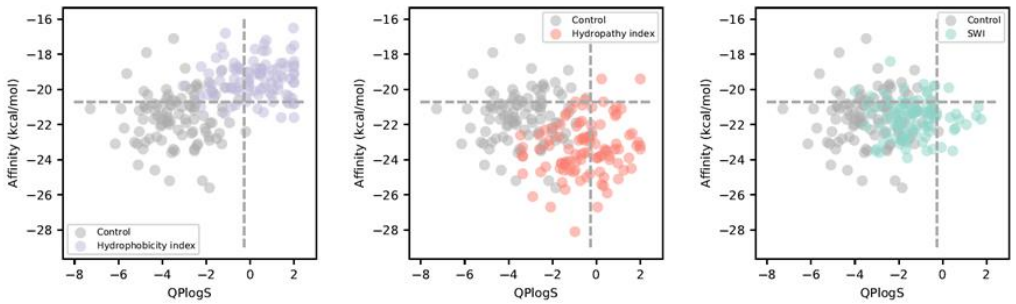


図 3　水溶性を加味した損失関数の設計によるペプチド配列の水溶性と標的親和性の変化（左：Hydrophobicity index 利用、中央:Hydropathy index 利用、右:Solubility weight index 利用）

なお、このテーマでは当初別の方針を取っていたが、ちょうど AlphaFold2 の情報が 2020 年末に一部公開され、2021 年にコードを含めて公開されたことで、チャレンジ支援を活用しながら調査・検証を進めた。結果、これらを活用したペプチドデザイン AI の開発に成功し、またその過程で得た知見も総説として発表できた。

研究テーマ B「PPI 阻害分子の弁別・物性・毒性などの予測を行うための AI 開発」

PPI 阻害分子の弁別・物性・毒性予測等を行う AI を開発し、創薬への展開を加速するための研究を実施した。具体的には、以下の(1)～(3)の研究を実施した。

(1) PPI 阻害化合物生成モデルの開発と化合物デザインへの応用

PPI 阻害剤が持つ特有の性質をモデル化した QEPPI を開発した。薬剤らしさ(Druglikeness)として利用されている QED は PPI 阻害剤のスクリーニングには不適であることを示し、PPI 阻害実験が行われた化合物の情報を収集して PPI 阻害剤スコア QEPPI を新規に開発した。QEPPI は QED よりも PPI 阻害化合物のスクリーニングに適していた（図 4）。以上は 2021 年度に Int J Mol Sci 誌に論文発表した。

（加速フェーズ）QEPPI を報酬とする深層強化学習による分子生成モデルから得られた化合物群を iPPI-REINVENT として公開し、Molecules 誌で発表した。

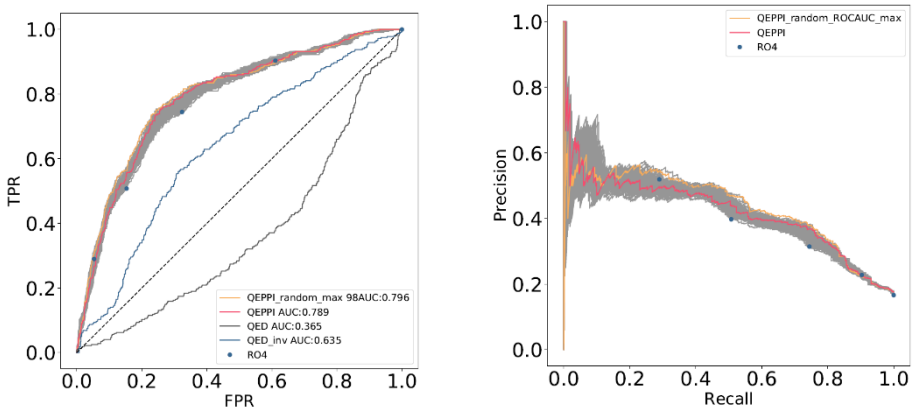


図 4　本研究で新規に開発した PPI 阻害剤様指標 QEPPI の分類性能。
（左：ROC 曲線、右 Precision-Recall 曲線）

(2) 化合物選別向けレコメンド指標の開発

化合物の教師データや実験条件などの種類、標的タンパク質の種類などが混在した状況を想定したデータセットを構築し、ランキング予測問題に特化した予測モデルと評価指標を開発した。この研究では、情報検索分野でよく用いられる NDCG 指標に、創薬分野でよく用いられる enrichment（濃縮）の考え方を組合せて、特定のモデルに対して予測が良いのか悪いのかを正しく判断することを可能にする NEDCG 指標を考案した。以上は査読付国際会議 IEEE CIBCB2022 で論文発表を行った。

(3) 血液胎盤関門透過予測と applicability domain の検討

血液胎盤関門の透過性を学習した PBPredictor を開発した。血液胎盤関門の透過性は胎児毒性を回避するために重要な指標となるが、現状公開されているツールや市販ツールには予測モジュールが存在していなかった。本研究で新たに胎盤灌流実験による血液胎盤関門透過性の実験値を収集し、勾配ブースティング木による予測モデルを構築した。良好な予測性能が得られることをクロスバリデーションによって確認できたため、予測ツールをウェブサービスとして実装して暫定公開した。また、実際に予測が信頼できるかどうかを判断する適用可能領域（applicability domain）の構築技術を提案し、査読付国際会議 IEEE CIBCB2021 で論文発表を行った。

研究テーマ C「PPI 界面を標的とする環状ペプチドのデザイン AI の開発」（加速フェーズ）

AlphaFold2 (Jumper+ Nature 2021) と Deep Network Hallucination 法 (Anishchenko+ Nature 2021)を組合せて直鎖ペプチドを設計する AI を開発してきたが、Rettie らによる AlphaFold2 の Positional encoding の検討により、head-to-tail 型の環状ペプチドを AlphaFold2 で生成できることが判明した (Rettie, bioRxiv, 2023)。この研究では単体の環状ペプチドを生成していたが、これを応用すると目的の標的タンパク質に結合する環状ペプチドを予測できることが予想された。タンパク質との複合体予測のための positional encoding を設計し（図 5）、Deep Network Hallucination 法と組合せて、結合能が高い配列をデザインできることを示した（図 6）。これらの結果は Int J Mol Sci 誌に発表した。

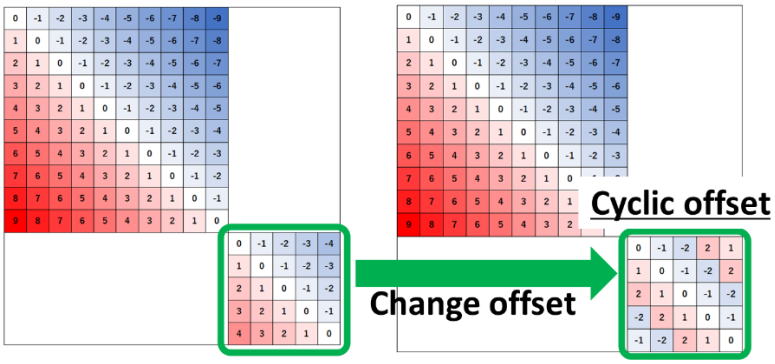
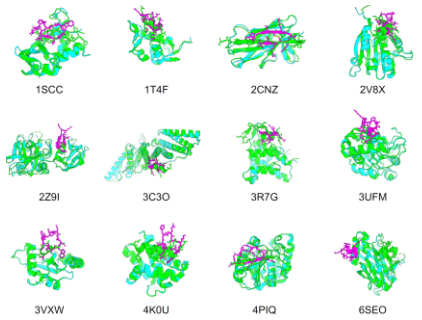


図 5 本研究で新規に開発したタンパク質−環状ペプチド複合体予測のための positional encoding



	pLDDT	Rosetta Interface analyzer score (Design)	RMSD_best	Rosetta Interface analyzer score (Native)
1SSC	88.51	-2.51	2.58	-2.23
1T4F	86.39	-2.49	0.85	-2.40
2CNZ	90.00	-4.05	0.60	-3.98
2V8X	82.89	-1.58	3.06	-2.35
2Z9I	83.79	-0.38	0.52	2.67
3C3O	71.23	-1.10	1.40	-0.21
3R7G	91.53	-1.45	1.34	-1.41
3UFM	77.05	-2.33	4.53	-2.77
3VXW	79.40	-1.90	0.65	4.80
4K0U	95.24	-2.85	0.51	-2.04
4PIQ	84.88	-1.16	0.79	-0.90
6SEO	77.02	-0.96	18.15	2.80

Green : PDB
Cyan : Prediction target protein
Magenta : Prediction cyclic peptide

図 6　改変 AlphaFold2 の Deep Network Hallucination による環状ペプチドデザイン。1 例を除いてネイティブ配列よりも高い結合能が示唆される配列を設計できた。

研究テーマ D「所望の物性を持つ低分子化合物の設計支援技術の開発」（加速フェーズ）

低分子化合物を表現する方法としてグラフ表現があるが、原子とその結合をノードとエッジで示す通常の分子グラフ表現は創薬上重要な部分構造の解釈が難しいという欠点がある。本研究では新たに reduced graph として複数の縮約化された分子グラフ表現を組合せることで、原子単位の細かい解釈から部分構造単位の大まかな解釈を兼ね備えた表現ができるグラフニューラルネットワークモデルを提案した（図 7）。これにより、原子レベルの詳細な表現能力を維持したまま attention weight から重要な部分構造表現を抽出することができるようになり、高い予測精度と解釈性を両立した予測ができるようになった。査読付国際会議 IEEE CIBCB2023 で論文発表を行った。

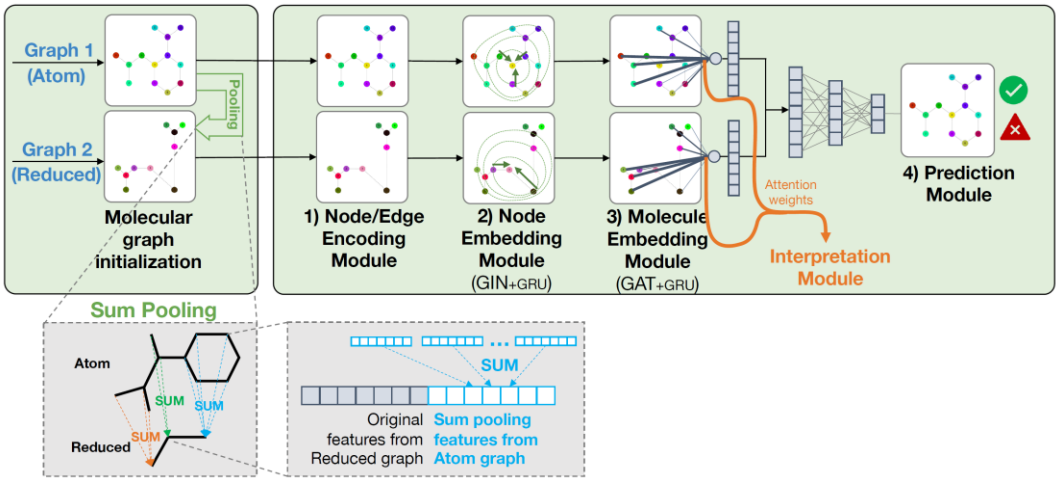


図 7　原子グラフと reduce graph を組合せたグラフニューラルネットワークによる化合物設計 AI

3.　今後の展開

本研究成果により、ペプチドによる標的タンパク質の自在な制御が実現しつつある。今後の展開として、「本成果を利用した共同研究等による展開」、「特殊なペプチドへの手法拡張」、「低分子化による活用領域の拡大」が考えられる。

（加速フェーズ実施後追記）

加速フェーズでは、ペプチドの N 末端と C 末端が結合した環状ペプチドを扱うことができるようになり、「特殊なペプチドへの手法拡張」の一部が達成された。また、低分子に関する新たなグラフニューラルネットワークの成果から、「低分子化による活用領域の拡大」の一部も達成された。

4. 自己評価

研究目的の達成状況

本課題が目的とするタンパク質を制御可能なペプチドのデザイン AI の開発は、試行錯誤や方針転換などがあったものの、結果として目標を十分に達成できた。また、関連した様々な AI 開発は、当初の計画を超えて得られた成果である。

（加速フェーズ実施後追記） 加速フェーズ計画時点では目処が立っていなかった環状ペプチド構造の予測方法を開発することができ、当初の計画を超えて「今後の展開」に記した研究展開が早期に実現できた。

研究の進め方（研究実施体制及び研究費執行状況）

研究補助者として大学院生 RA を雇用し、円滑に研究を進めた。新型コロナウイルス感染症の影響で当初計画していた調査・成果発表の渡航を変更したり、半導体高騰や資材高騰の影響による変更などもあったが、柔軟に対応することで研究費の執行も全体として計画通り執行した。

（加速フェーズ実施後追記） 加速フェーズでも大学院生 RA とともに円滑に研究を進めた。研究費も同様に計画通り執行した。

研究成果の科学技術及び社会・経済への波及効果

すでに共同研究や検討が複数進められており、本研究成果の展開が見込まれている。創薬に AI を活用する「AI 創薬」という単語が叫ばれているが、AI 創薬の主な利用シーンは標的活性の教師あり学習という典型的な問題に留まっていた。本研究成果は、創薬に活用できる AI の新たな側面を提示し、新しいアプローチによる創薬フローを提示する画期的成果である。低分子・中分子創薬はもちろんのこと、基礎生物分野から医学分野まで幅広く応用できる可能性があり、大きな波及効果が期待できる。

（加速フェーズ実施後追記） 加速フェーズを通じて、共同研究等の新規実施を含む新規研究展開が進んだ。

5. 主な研究成果リスト

（1）代表的な論文（原著論文）発表

研究期間累積件数：7件

研究期間累積件数：12件（加速フェーズ実施後更新）

1. Kosugi T, Ohue M. Solubility-aware protein binding peptide design using AlphaFold. Biomedicines, 10(7): 1626, 2022. doi: 10.3390/biomedicines10071626
計算によって標的に結合する可能性のある適切なペプチド配列を設計することは困難であっ

たが、AlphaFold の高精度なタンパク質構造予測の応用により配列設計が可能となった。一方で、AlphaFold を単純に用いて設計したペプチド配列は、相互作用面の疎水性志向を反映し、溶解度が低い傾向となる問題があった。本論文では、アミノ酸の溶解度指標を用いた溶解度損失関数を設計し、溶解度を考慮したペプチド設計法を開発した。設計されたペプチドは、水溶性を保持し、かつドッキングによる結合親和性も高い傾向があった。

2. Kosugi T, Ohue M. Quantitative estimate index for early-stage screening of compounds targeting protein-protein interactions. International Journal of Molecular Sciences, 22(20): 10925, 2021. doi: 10.3390/ijms222010925

定量的な薬効評価には、Drug-likeness スコア QED が一般的に用いられているが、タンパク質間相互作用(PPI)を標的とした化合物のスクリーニングには適さない。本論文では、既知の PPI 阻害剤データを収載した iPPI-DB の化合物を用いて新たに数理モデル化を行い、PPI 標的化合物のスクリーニングに特化した PPI 標的化合物向けスコア QEPPI を開発した。QEPPI は QED よりも PPI 標的化合物のスクリーニングに適していることが示され、特に PPI 標的化合物と FDA 承認薬の識別を F 値 0.501 で行うことができた。

3. Furui K, Ohue M. Compound virtual screening by learning-to-rank with gradient boosting decision tree and enrichment-based cumulative gain. In Proceedings of The 19th IEEE International Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB 2022), 7 pages, 2022. doi: 10.1109/CIBCB55180.2022.9863032

情報検索で広く利用されるランク学習を化合物スクリーニングに応用するため、ランダムスクリーニングに対する改善度を指標とした新しいランキング指標である Normalized Enrichment Discounted Cumulative Gain (NEDCG)を提案した。目的とする標的分子に関する実験データの実験データが複数アッセイによって構成される場合や、標的分子のファミリータンパク質に関するデータがある場合を想定した検証の結果、ランク学習が有効になる状況やランク学習があまり適さない状況を、NEDCG によって定量化できることを示した。

4. Kosugi T, Ohue M. Design of cyclic peptides targeting protein-protein interactions using AlphaFold. International Journal of Molecular Sciences, 24(17): 13257, 2023. doi: 10.3390/ijms241713257

これまで AlphaFold で扱えるペプチドは直鎖型のものであったが、創薬においては環状ペプチドの高い代謝安定性や膜透過性が注目されている。本論文では、AlphaFold の positional encoding を改変してタンパク質-環状ペプチド複合体構造予測を行う方法を新たに提案し、所望の標的に結合する環状ペプチドの配列設計を行う技術を開発した。

5. Kengkanna A, Ohue M. Enhancing Model Learning and Interpretation Using Multiple Molecular Graph Representations for Compound Property and Activity Prediction. In Proceedings of The 20th IEEE International Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB 2023), 8 pages, 2023. doi: 10.1109/CIBCB56990.2023.10264879

低分子のグラフ表現として一般的に用いられる原子グラフは、創薬において重要な部分構造の解釈が難しいという問題があった。本研究では、分子グラフ表現を縮約化した reduced graph を原子グラフとともに用いる新しいグラフニューラルネットワークを提案し、所望の物性

を得るために必要な部分構造を抽出できる解釈性の高い予測手法を開発した。
6. Ohue M, Kojima Y, Kosugi T. Generating potential protein-protein interaction inhibitor molecules based on physicochemical properties. Molecules, 28(15): 5652, 2023. doi: 10.3390/molecules28155652
PPI を阻害できる低分子化合物の予測指標 QEPPI を用いた深層強化学習により、PPI 阻害可能性の高い化合物を仮想的に生成する分子生成モデルを開発した。生成された化合物群は市販化合物と比較してより高い QEPPI を有しており、より適切な化学空間の探索ができていることが示唆された。生成された化合物情報は iPPI-REINVENT として GitHub から公開している。

（2）特許出願

なし

（3）その他の成果（主要な学会発表、受賞、著作物、プレスリリース等）

総説

- 大上雅史. 中分子ペプチド創薬のインフォマティクス. 実験医学, 2023 年 1 月号, 41(1): 20-26, 羊土社, 2023.
- 大上雅史. AlphaFold2 の登場と創薬への影響, 革新的 AI 創薬～医療ビッグデータ、人工知能がもたらす創薬研究の未来像～, 株式会社エヌ・ティー・エス, 2022.
- 大上雅史. AlphaFold のタンパク質立体構造予測の性能, 実験医学, 2022 年 2 月号, Vol.40, No.2, pp. 427-430, 2022.
- 大上雅史. PPI（タンパク質間相互作用）を標的とするドラッグデザイン. タンパク質の構造解析手法と In silico スクリーニングへの応用事例, 402-410, 技術情報協会, 2023. ※加速フェーズ実施の成果
- 大上雅史. AI によって変わる生命科学. 現代化学 2023 年 8 月号, 28-30, 東京化学同人, 2023. ※加速フェーズ実施の成果
- 大上雅史. AlphaFold による高精度なタンパク質立体構造予測と創薬への応用. PHARM STAGE, 23: 8, 24-28, 2023. ※加速フェーズ実施の成果

学会発表

- 大上雅史, タンパク質間相互作用を狙う創薬インフォマティクス, 計算メディカルサイエンスワークショップ 2022, 2022.10.
- 大上雅史, タンパク質の理解と制御のための構造インフォマティクス, 第 65 回日本腎臓学会学術総会, SY2-3, 2022.6.
- 大上雅史, AI で広がる分子設計の可能性, Visionary 農芸化学 100 シンポジウム/第 50 回農芸化学「化学と生物」シンポジウム, 2023.7. ※加速フェーズ実施の成果
- 大上雅史, 生成 AI が加速する創薬・生命科学～AlphaFold2 の衝撃と Transformer アルゴリズム～、日本学術会議公開シンポジウム第 13 回計算力学シンポジウム、2023.12. ※加速フェーズ実施の成果

受賞

- 大上雅史. 一般財団法人 安藤研究所 第 35 回安藤博記念学術奨励賞（2022.6）
  受賞業績:「生体内タンパク質を標的とするバイオナノマテリアル予測技術の先駆的研究」
- 大上雅史. 情報処理学会 山下記念研究賞（2022.3）
  受賞業績:「GBDT による化合物の血液胎盤関門透過性予測」
- 大上雅史. 情報計算化学生物学会(CBI 学会) CBI 学会若手奨励賞（2023.10） ※加速フェーズ実施の成果
- 大上雅史. 情報処理学会 マイクロソフト情報学研究賞（2024.2） ※加速フェーズ実施の成果