

2023 年度年次報告書

データ駆動・AI 駆動を中心としたデジタルトランスフォーメーションによる生命科学研究の革新

2021 年度採択研究代表者

岡田 真里子

大阪大学 蛋白質研究所

教授

自然言語処理とシミュレーションによる細胞制御探索法の構築

主たる共同研究者:

下平 英寿 (京都大学 情報学研究科 教授)

泰地 真弘人 (理化学研究所 生命機能科学研究センター 副センター長・チームリーダー)

## 研究成果の概要

大規模言語モデル(LLM)を利用することで、PubMed1500 万件全件要旨から構築したベクターデータベースを用い、EGFR シグナル伝達系の生化学ネットワークを比較的ノイズ少なく構築することに成功した。また、KEGG に先に開発した Text2Model を適用することで、KEGG から動的な数理モデルの構築が可能になった。また、この研究により、Text2Model がノイズの多い論文情報に対するノーマライゼーション機能を持つことが明らかになった。

遺伝子と薬剤間の相互作用について、事前学習済みの単語ベクトル(BioConceptVec)を用い、KEGG パスウェイ情報のアナロジー計算によって関連性の予測を行った。その結果、遺伝子と薬剤の単語ベクトルから加減算という単純な演算によって関連性の高い遺伝子を比較的高い精度で予測できることが確認できた。また、この手法により、過去の文献に記載された遺伝子と薬剤の関係性から、未来の文献で発見される関係性の予測がある程度可能なことを確認した。

化合物生成 AI-分子ドッキングの連携による薬物設計システムの高精度化のため、化合物生成 AI(MoFlow)をベイズ学習を適用し、シード化合物周辺の丁寧な類似構造探索を可能とした。加えて、transformer をベースとした化合物生成 AI の導入・検証を行い、標的である代謝酵素に対して、効果的な阻害薬候補の生成が可能であることを確認した。また、細胞と分子の二階層を跨ぐシミュレーション統合を目的とし、変異型受容体とその結合ペプチドに対して、分子動力学(MD)シミュレーションを行い、それぞれのシミュレーションにより得られた結合定数の相関を調査し、手法の妥当性を検証した。

熱動力学モデリング手法を用い、2つの Raf 阻害剤の組み合わせによる Raf 過剰発現がん細胞株の増殖阻害効果について論文を発表した。皮膚老化および歯周病の数理モデルの論文をそれぞれ発表した。

### 【代表的な原著論文情報】

- 1) Arakane K, Imoto H, Ormersbach F, Okada M. Extending BioMASS to construct mathematical models from external knowledge. *Bioinformatics Advances*, vbae042, 2024.
- 2) Haga M, Iida K, Okada M. Positive and negative feedback regulation of the TGF- $\beta$ 1-SMAD4 axis explains two equilibrium states in human skin aging. *iScience*, accepted.
- 3) Fujihara C, Murakami K, Magi S, Motooka D, Nantakeeratipat T, Canela A, Tanaka, RJ Okada M, Murakami S. Omics-based mathematical modeling unveils pathogenesis of periodontitis in an experimental murine model. *Journal of Dental Research*, 102 (13), 1468-1477, 2023.