

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

2021 年度 年次報告書

岡田 眞里子

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

§ 1. 研究成果の概要

本研究では、計算科学主導型の細胞ネットワーク制御の方法論の確立を目的として、自然言語処理による遺伝子相互作用の抽出、数式フリーの細胞数理モデリング手法の構築、細胞制御化合物デザインの自動化といった要素技術を開発する。テキストからの数理モデルの構築を目的として、自然言語処理とテキストマイニングの2つのアプローチを用い、PubMed および公共データベースから、遺伝子相互作用と制御様式の抽出を進めた。PubTator を用いた自然言語処理手法と蛋白質相互作用データベースの利用で、第一目標とする EGFR ネットワーク周辺の相互作用をある程度抽出することが出来ることがわかった。一方で、今後、カスケードなどのような連続した生化学反応の上下の制御関係を如何に抽出するかが新たな課題として見出された。これまで、研究代表者らが JST の他事業から継続して開発を進めたテキストベースのモデリング手法 text2model を完成させ、論文として発表した。また、このモデリング手法を用い、乳がんの遺伝子発現情報からイン・シリコ分子活性をシミュレーションし、各患者の予後と治療標的を予測するパイプライン、Pasmopy (Patient-Specific Modeling in Python)をパッケージとして公開した。さらに、数理モデル内の機能蛋白質と創薬標的蛋白質の相互作用制御の実現に向け、機械学習による薬物生成モデルと分子シミュレーションを連携させた薬物設計手法開発の調査・検証を行なった。また、数理モデルにおける蛋白質相互作用の反応速度パラメータの推定を目的として、野生型・変異型 EGFR 蛋白質複合体の計算モデルを構築し、MD シミュレーションを実施した。研究はおおむね順調に進展している。

§ 2. 研究実施体制

(1) 大阪大学グループ

- ① 研究代表者: 岡田 眞里子 (大阪大学 蛋白質研究所 教授)
- ② 研究項目
 - ・文献情報・データベースからの遺伝子相互作用とリンクタイプの抽出
 - ・テキストベースモデリングによるがんネットワーク解析

(2) 理化学研究所グループ

- ① 主たる共同研究者: 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)
- ② 研究項目
 - ・薬物生成の機械学習モデル-分子シミュレーションシステムの導入に向けた調査
 - ・数理モデルの反応速度パラメータ推定のための、MD シミュレーションの解析

(3) 京都大学グループ

- ① 主たる共同研究者: 下平 英寿 (京都大学 情報学研究科 教授)
- ② 研究項目
 - ・文献情報から遺伝子・薬剤の特徴抽出と生化学反応モデル言語への翻訳

【代表的な原著論文情報】

- 1) A text-based computational framework for patient-specific modeling for classification of cancers. H Imoto, S Yamashiro, M Okada. *iScience* 25 (3), 103944, 2022.
- 2) $I\kappa B\alpha$ is required for full transcriptional induction of some NF κ B-regulated genes in response to TNF in MCF-7 cells. M Ando, S Magi, M Seki, Y Suzuki, T Kasukawa, D Lefaudeux, A Hoffmann, M Okada. *NPJ Systems Biology and Applications* 7 (1), 1-15, 2021.