氏　　名　小髙 充弘

学位(専攻分野)　博士(情報学)

学 位 記 番 号　総研大甲第 2497 号

学位授与の日付　2024 年 3 月 22 日

学位授与の要件　複合科学研究科 情報学専攻
学位規則第6条第1項該当

学 位 論 文 題 目　Data-Driven and Knowledge-Based Multiscale Modeling of Viral Dynamics

論 文 審 査 委 員　主　査　宇野 毅明
情報学コース 教授
井上 克巳
情報学コース 教授
杉山 麿人
情報学コース 准教授
Morgan Magnin
エコール・サントラル・ナント 教授
Sophie Limou
エコール・サントラル・ナント 教授
Bertrand Michel
エコール・サントラル・ナント 教授
Loïc Paulevé
LaBRI CNRS 研究部長
Sylvain Soliman
INRIA Sacray 研究員

本学位論文研究は総合研究大学院大学とÉcole Centrale de Nantes（フランス共和国）とのデュアル・ディグリー・プログラム[1]に関する協定に基づく国際共同指導により実施されたものである。

[1] コチュテル（デュアル・ディグリー）
海外の大学との協定に基づき本学に在籍する学生が、同時に相手大学に正規生として在籍し、両大学の教員から共同で学位論文指導を受けるもの。論文完成後、両大学による共同の論文審査に合格した場合は、両大学から単一の学位が授与され、各大学からそれぞれ両大学の共同論文指導によるものである旨を付記した学位記を交付する。

博士論文の要旨

氏　　名　小髙　充弘

論文題目　Data-Driven and Knowledge-Based Multiscale Modeling of Viral Dynamics

We conducted interdisciplinary research combining informatics and biology on multiscale modeling of viral dynamics using data and knowledge. Behind this research, launched in 2019, lies a global issue of the coronavirus disease 2019 (COVID-19) epidemic during the same period. The two primary goals of this research were 1) to understand the mechanism of COVID-19 and 2) to develop a framework that can be used universally to control emerging infectious diseases, not limited to COVID-19.

To achieve the first goal, we focused on the lack of informatics research on viral population dynamics within the host in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection and the need to understand the mechanism as a multiscale system with a network structure of components (Figure 1). By linking these points of view to the closed loop for scientific discovery proposed initially by C. S. Peirce (Figure 2), which consists of three types of reasoning: deduction, induction, and hypothesis finding, we planned two studies (Figure 3).

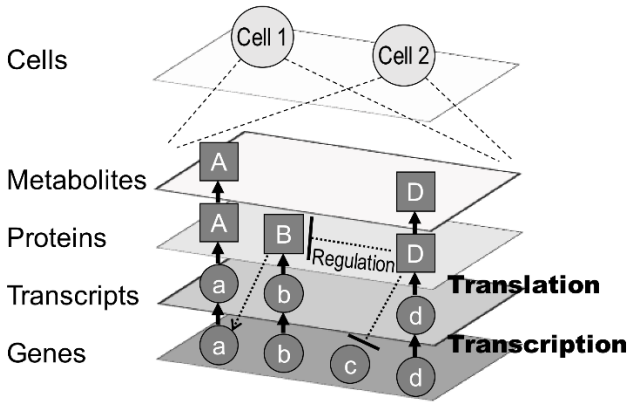

Figure 1　Multiscale model

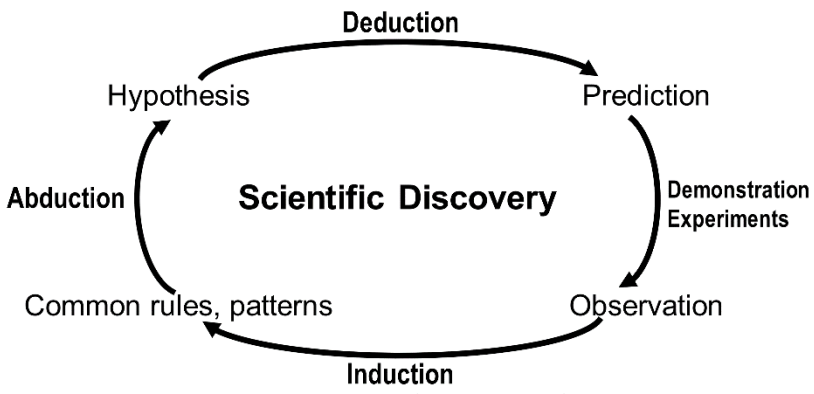

Figure 2　Closed loop for scientific discovery

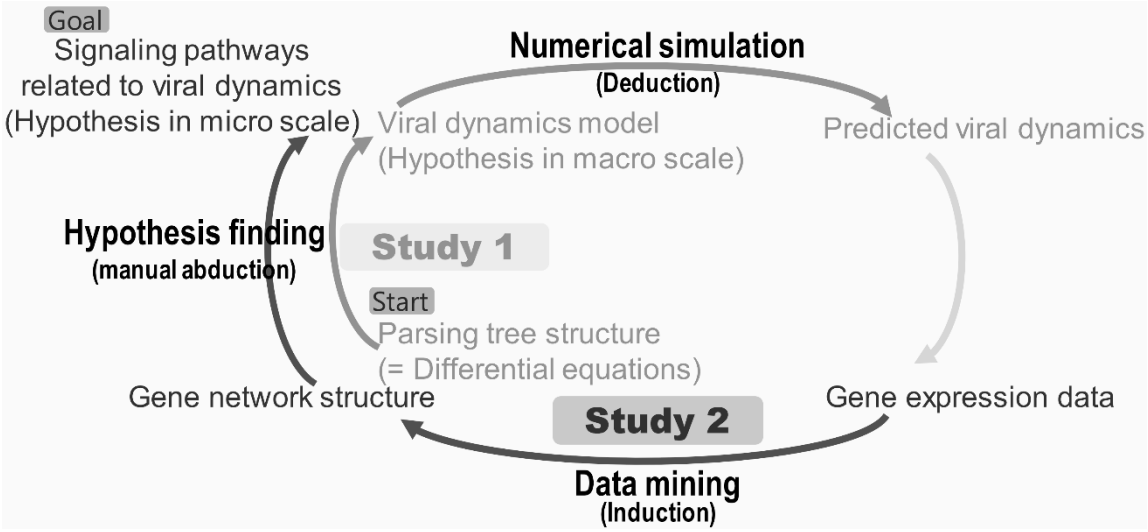

Figure 3 Relationship between two studies in the context of scientific discovery

In Study 1, we constructed multiple SARS-CoV-2 viral dynamics models based on differential equation systems and qualitatively analyzed the models by sensitivity and stability analyses. After model order reduction, we conducted computer simulations and calibration experiments with time-series data of viral antibody titers (viral load) for different severities of illness. The parameter of hypothetical cell-to-cell transmission (direct transmission of virus from one cell to another) in the SARS-CoV-2 infection system was associated with COVID-19 severity, suggesting the cell-to-cell transmission hypothesis for SARS-CoV-2. The results were published in a journal article (Q1, IF=3.8).

In Study 2, we attempted to verify the hypothesis of macro-scale viral dynamics (SARS-CoV-2 cell-to-cell transmission hypothesis) derived from Study 1, based on data on micro-scale molecules, such as genes and transcripts, and domain knowledge. Focusing on the fact that the biological signaling pathways involved in ICAM-1 (Intercellular adhesion molecule 1), a representative molecule responsible for intercellular adhesion involved in cell-to-cell transmission, were not included in the existing knowledge repository, we conducted our research with the aspects of framework development for the second primary goal and the aspect of scientific discovery of unknown signaling pathways. The developed framework includes the following processes.

1) Extracting coexpressed genes with disease-specific variation in gene expression levels by statistical processing of continuous-valued sparse matrices of large-scale single-cell omics data (dimensionality reduction with UMAP, clustering with Louvain algorithm, Wilcoxon rank-sum test) and estimating the coexpressed gene network of *ICAM1* by calculating second-order partial correlation coefficients and removing spurious correlations

2) Model validation based on the background knowledge derived from the seven knowledge bases.

3) Mapping the obtained network to the KEGG pathway database, analyzing the

activated functions, and outputting the results.

We applied this framework to experimental data and compared the results on temporal and spatial axes. The framework was also applied to other genes involved in cell-cell interactions, including *C15orf48* and *ACTB*. Our results suggest the existence of unknown signaling pathways involving ICAM-1, such as the non-classical NF-kB pathway (putative), involved in the regulation of inflammatory responses, activation of molecules involved in microtubule organizing centers, and the integrin pathway (putative) leading to viral synapse formation. In this way, microscale data and knowledge verified our hypothesis on macroscale viral dynamics. To validate our research results from biological perspectives, we contacted experts in infectious diseases at the Department of International Health at the Institute of Tropical Medicine. As an outcome, we published a preprint paper for immediate use and subsequently published it in a journal article (Q2, IF=4.8).

Significant contributions from the above studies are as follows.

1) Contributions to the field of informatics

- Creating a pipeline for acquiring optimum models and parameters from time-series data and literature on state variables

  Novelty:

  - Existing model diversion and original model construction
  - Application of four models to empirical data

- Developing and proposing an original Data-Driven and Knowledge-Based (DD-KB) framework for automatically inferring systems by combining graphical modeling from the large-scale sparse matrix of multivariate (time-series) data and model validation with multiple knowledge bases and successfully forming the basis for further frameworks

  Novelty:

  - Combination of existing data mining and database integration methods
  - Automation of hypothesis discovery from data and knowledge

2) Contributions to the field of biology

- Fitting mathematical models of within-host viral dynamics to the COVID-19 dataset (including existing and original models assuming cell-to-cell transmission effects) and comparing estimated parameters of multiple models
- Demonstrating the DD-KB framework applicability to the COVID-19 gene expression data and biological background knowledge, thereby reproducing existing pathways, constructing novel pathways, and analyzing their spatiotemporal variation for different genes of interest not previously available in the knowledge repositories

(Separate Sheet)

Results of the doctoral thesis defense

Name in Full: ODAKA Mitsuhiro

Title: Data-Driven and Knowledge-Based Multiscale Modeling of Viral Dynamics

The doctoral thesis defense of Mr. ODAKA Mitsuhiro (hereinafter referred to as "the applicant") was held on January 18, 2024, with attendance of all six members of the jury and an external reviewer. The applicant firstly gave oral presentation of the contents of the thesis for 45 minutes, then the examiners asked questions and gave comments for 30 minutes.

The applicant's doctoral thesis has been supervised by Japanese and French sides under the Dual Degree Program (DDP). The main research topic is the dynamics of COVID-19, an infectious disease caused by SARS-CoV-2. Motivated by the social demands undergoing a strategic shift from infection control to prevention, his research was devoted to contributing to two things from a computational aspect: to discover unknown domain knowledge for elucidating a system of signaling pathways involved in COVID-19 and to develop a framework applicable to upcoming diseases, not limited to COVID-19. He placed these purposes on a closed loop of scientific discovery, narrowing down his scope into finding and verifying hypotheses. To realize such scientific discovery, the applicant conducted research from two viewpoints: viral dynamics and multiscale modeling. Specifically, one primary study (called Study 1) is on an attempt to find a hypothesis on viral dynamics from a macroscopic view, and the other main study (called Study 2) attempts to verify the above hypothesis from a microscopic view. These two studies are then summarized in the proposed framework of "Data-Driven and Knowledge-Based Multiscale Modeling". The thesis is written in English, consisting of six chapters.

Chapter 1 is a comprehensive introduction to the social background of COVID-19 as a global issue and the applicant's response to its social demands. The time-dependent change in the global situation and social demands are explained based on statistical data and literature. Subsequently, the definitions of terms, such as biological mechanism, viral dynamics, and multiscale modeling, are given. The chapter ends with the purposes and potential significances throughout the doctoral research, specified as the scientific discovery of viral dynamics with multiscale modeling.

Chapter 2 provides preliminaries about different modeling techniques, modeling of viral dynamics, and a Data-Driven and Knowledge-Based (DD-KB) approach. Formal backgrounds of an equation-based model in fundamental viral dynamics and model validation based on domain knowledge are explained to prepare further chapters.

Chapter 3 corresponds to Study 1, modeling and simulation of viral dynamics based

on differential equations. The applicant proposes a pipeline of methods, including modeling four viral dynamics, sensitivity analysis, stability analysis, and calibration experiments. The results of numerical analyses and model fitting to viral load data in mild and severe cases are shown, which indicates that the assumed cell-to-cell transmission of SARS-CoV-2 would be associated with the COVID-19 severity.

Chapter 4 corresponds to Study 2, which verifies the hypothesis on the cell-to-cell transmission hypothesis from a system of pathways. Focusing on ICAM-1 as a molecule responsible for intercellular interaction, the applicant constructs the pathways of ICAM-1 and the interactions surrounding this molecule. The DD-KB framework, which consists of single-cell omics data analysis, model validation, and pathway construction, is applied to two cases: omics data with different cell types and time points. The results of both case studies showed that the framework can reproduce the existing pathways and newly find the putative pathways, such as upstream non-canonical NF-kB pathway and downstream integrin pathway. The chapter concludes with the DD-KB framework's applicability and future improvement, such as comparison with other gene network inference or causal discovery methods.

Chapter 5 reviews related work for improving Studies 1 and 2, including scientific discovery, causal discovery methods in the general and biological context, and other possible conditions for viral dynamics modeling.

Chapter 6 concludes with a summary of contributions, limitations, and remarks about further perspectives of multiscale modeling and the DD-KB approach.

In the Q&A session, the external reviewer first gave several questions, and the jury members followed next. The applicants adequately answered all the questions.

After the Q&A session, the Defense Committee, i.e., the jury members and external reviewers, evaluated the application. The Defense Committee expressed the importance that the applicant has utilized many diverse methods to analyze real biological and medical data, which is quite impressive, and that the doctoral research has reached a proper level for utilizing, developing, and extending multiple methods from computer/data science and biology. The Defense Committee thus admitted that the applicant has completed the defense with sufficient knowledge and research ability in the frontier field of informatic. As for the publication related to the contents of the thesis, one international journal paper (*Heliyon*) and one international conference paper have been published for Study 1, and one international journal paper (*Frontiers in Genetics*) has come out from Study 2. To sum up, this doctoral thesis is a rationale for the quality of his endeavor to discover the unknown scientific knowledge and demonstrate the high application potential of the DD-KB framework.

The Defense Committee expressed their satisfaction with the quality of the PhD work, and recognized that the submitted thesis fulfilled the requirements for a PhD thesis in the dual degree program. Hence, the Defense Committee unanimously judged a pass of the final defense of Mr. ODAKA Mitsuhiro on his doctoral degree.