Using cryo-EM methods to uncover structure and function of bacteriophages

Ouyang, R.

1

Chapter 1

General introduction and Thesis outline

General introduction

Bacteriophages, or phages for short, are the most abundant biological entity in nature. They shape bacterial communities and are a major driving force in bacterial evolution. Their ubiquitous nature and their potential use in medical and industrial applications make them attractive targets for fundamental and applied scientific studies. Understanding their structure and function at the molecular level is essential for understanding phage life cycles. In this thesis, I applied different cryo-EM techniques combined with advanced image processing and artificial intelligence methods to gain insight into structure and function of two bacteriophages. In both cases, these phages contain flexible elements which are essential for the infection process. While biologically highly interesting, these flexible components are especially challenging for structural studies.

More specifically, I use cryo-electron microscopy approaches (single-particle analysis and cryo-electron tomography) as two structural research methods to investigate the structure of novel bacteriophages with flexible fibers that are important for their infection process. To obtain a comprehensive understanding of the structure and function of these phages, I adopted a combined research method that integrated sequence prediction, protein structure prediction, computational simulation, machine learning, and artificial intelligence algorithms. Moreover, I developed and tested different protocols and strategies to process single-particle analysis data to balance the quality of results and available computational resources. Additionally, I applied an efficient neural network algorithm to extract complex fiber structures automatically from cryo-electron tomography data.

I present recent studies and discoveries that underscore the importance of investigating tail and capsid fibers in diverse phages, highlighting their potential to provide valuable insights into the intricate structures underlying phage infection initiation and progression. I also studied the structure of a jumbo bacteriophage and its potential use as a therapeutic agent against pan resistant *Klebsiella pneumoniae* strains.

Furthermore, I investigated the process of a flagellotrophic phage infecting *Agrobacterium*, which provided insights into the development of a phage therapy method for the treatment of plant crown gall tumors.

With the advances in computer technology and electron microscopy, researchers can now use various research methods to study different proteins and the structure and function of biological macromolecular machines. This comprehensive research approach enables research targets to span different scales, from dynamic membrane proteins to complex macromolecular biological machines such as giant bacteriophages. The studies presented in this thesis provide valuable insights into phages with flexible components, and provide a useful workflow for researchers with similar research topics.

Thesis outline

This thesis consists of a series of investigations where the various techniques of Cryo-electron microscopy were used to gain insight into the architecture and function of bacteriophages and their involvement in the infection process of their bacterial hosts. Here, we especially focused on bacteriophages that contain flexible fibers in their infection process. These studies include unusual flexible tail fibers as well as head fibers emerging from the phage capsids. Due to their flexible nature, such fibers have been challenging to study despite their importance of virus function.

In Chapter 2, I provide an overview of the cryo-electron microscopy (cryo-EM) techniques and the basic methods involved in preparing, collecting, and processing cryo-EM samples. I also introduced the artificial intelligence technology used in cryo-EM and some applications I used for further research.

In Chapter 3, I present an overview of recent studies and discoveries that underscore the importance of investigating tail and capsid fibers in diverse phages. It provides insights into the current knowledge of the intricate structures of these fibers and their role in phage infection initiation and progression. The rapid advancements in structural methodologies, coupled with the utilization of advanced techniques such as cryo-electron microscopy (cryo-EM) methods, machine learning, and artificial intelligence, significantly enhance our ability to delve into the less explored aspects of tail and capsid fibers. By achieving a comprehensive understanding of these phage fibers, new opportunities emerge for engineering innovative receptor binding proteins (RBPs) or chimeric tail fibers that can precisely recognize and bind to specific host receptors, thereby expanding the host range of phages. These endeavors hold significant promise

for advancing our comprehension of phage infection mechanisms and paving the way for diverse applications in the fields of biotechnology, molecular tools, and ecological indicators.

In Chapter 4, I present the high-resolution reconstructions of the unique jumbo phage ϕKp24, which is capable of infecting the human pathogen *Klebsiella pneumoniae*. The study combines single-particle analysis, protein structure prediction methods, and molecular dynamics simulations to explore the capsid and tail of the phage. The capsid and tail of the phage were reconstructed, and their structures were determined at high resolution, with resolutions of 4.1Å and 3.0 Å for the empty capsid and tail, respectively. Through the use of ISOLDE software, models generated from AlphaFold2 were flexibly fitted into density maps of the capsid and tail. Molecular dynamics flexible fitting (MDFF) was then employed to refine the models, providing insights into the assembly organizations of the capsid and the contractile mechanism of the tail.

In Chapter 5, I employed cryo-electron tomography (cryo-ET) and machine learning techniques to investigate the intricate tail fibers of the jumbo phage ϕKp24. To gain insight into ϕKp24's infection mechanism in a natural environment, I imaged the phage with *Klebsiella pneumoniae* with cryo-ET and reconstructed tomograms of the different conformations of the phage using IMOD. I manually segmented the tail fibers and together with my collaborators, we developed an efficient artificial intelligence network to automatically track the tail fibers and display their complex 3D structures. Our results showed that the branched tail fibers undergo significant rearrangement upon cell surface attachment, involving fourteen putative tail fibers with depolymerase activity that enable ϕKp24 to infect various capsular polysaccharide (CPS) types of *K. pneumoniae*. The study provides valuable structural and functional insights into the adaptability of ϕKp24 to capsulated bacterial pathogens. Overall, **Chapter 4** and **Chapter 5** provide a comprehensive understanding of the structure and function of jumbo phage ϕKp24, which has important implications for the development of new therapy approaches against pan-drug-resistant *K. pneumoniae* strains.

Chapter 6 introduces phage 7-7-1, which is known to be flagellotropic and attach to *Agrobacterium sp. H13-3*'s flagella through fibers emerging from its capsid. The unique features of this phage include the fibers extended from each vertex of the icosahedral

capsid. I studied this phage using a combination of single-particle analysis (SPA), protein structure prediction methods, and molecular dynamics simulations. I was able to reconstruct the capsid of the phage, and the full-DNA capsid structure was determined at a resolution of 3.9 Å. The ISOLDE software was then used to flexibly fit models generated from AlphaFold2 into the density maps of the capsid. The findings provide valuable insights into the assembly organizations of the capsid and the mechanism of the capsid fibers in attaching to the flagella of *Agrobacterium sp. H13-3*.

Chapter 7 of this thesis outlines a study that employed the Cryo-ET technique and machine learning approach to investigate the flagellotropic phage 7-7-1. Previous studies had revealed that the phage 7-7-1 uses the clockwise rotation of flagella to reach the bacterial membrane and complete its infection process. However, little information was available regarding the structure and function of the capsid fibers allowing this interaction with the hosts' flagella. In this study, I co-incubated *Agrobacterium sp. H13-3* and Phage 7-7-1 and reconstructed tomograms using the cryo-ET method to reveal the conformation of Phage 7-7-1 and the flagella in *situ*. The AI network developed in **Chapter 5** was employed to investigate the capsid fibers and flagella, which were auto tracked, extracted, and displayed as 3D objects. The findings of this study demonstrate the effectiveness of the AI network in displaying complicated or unique structures in 3D, which can offer an alternative approach to processing data that may not be suitable for SPA or cryo-ET.

In the final chapter of this thesis, **Chapter 8** provides a comprehensive summary of the investigations carried out in this work. This chapter also discusses the challenges associated with, and the solutions of, different techniques and algorithms to gain a deeper understanding of the structure and function of diverse proteins and macromolecular assemblies. The continuous advancements in computer technology and electron microscopy have facilitated the study of increasingly intricate and diverse biomolecules, ranging from the flexible membrane proteins to the complex macromolecular machines like phages, across varying scales. Overall, the integration of different analytical tools and methods can significantly contribute to our understanding of the underlying mechanisms of these biological entities.