THE UNIVERSITY OF CHICAGO

STRUCTURAL STUDIES ON PROTEIN COMPLEXES BY SINGLE-PARTICLE CRYOGENIC ELECTRON MICROSCOPY

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY

CHANG LIU

CHICAGO, ILLINOIS

AUGUST 2022

Table of Contents

List of Figures

List of Tables

Acknowledgements

First, I would like to express my sincere thanks to my research advisor Prof. Minglei Zhao. During my Ph.D. career, he has taught me a lot, from experimental techniques to scientific communications, and has always been around to offer me advice and training. I have always been encouraged by his passion for science and his genuine care for students. I was scared when I joined the lab in Jan 2017 because I barely had knowledge of biochemistry and structural biology. I am so grateful that Minglei was very patient with me and trained me to perform cryo-EM, a very sophisticated and complicated technique. I cannot make it without consistent help from Minglei.

I would like to thank all the members of the Zhao lab. The interdisciplinary environment enables me to learn a lot from everyone. Dr. Yuan Xie, our lab manager, managed lab stuff during my Ph.D. career and trained me on molecular cloning and tissue culture. She is very responsible and reliable to work with. Jane Lowick, a graduate student in the same year as me, is very helpful and reliable as a colleague and friend. She is very responsible for her experiments and took care of some lab businesses, and gave me a lot of help in my English communications. Dr. Man Pan and Dr. Yuanyuan Yu, the postdocs in my lab, taught me many basic knowledge and techniques about biochemistry, protein purification, and X-ray crystallography. I also thank Sihao Huang, David Cho, and Ally Cao for assisting my experiments. It has been my great fortune to work in this collaborative group for the past six years of my life.

I also want to thank the Advanced electron microscopy facility staff at the University of Chicago. I would like to thank Dr. Jotham Austin and Dr. Tera Lavoie for maintaining the Talos and Titan Krios and other daily operations in the facility. They responded very fast to all the facility issues and were very patient in training me to use the cryo-EM and Vitrobot. Ms. Yimei

Chen is also very patient in training me on the other two transmission electron microscopes, F30 and spirit.

Also, I want to express my gratitude to my collaborators. Dr. Lulu Hu from Prof. Chuan He's lab at the University of Chicago has provided essential help in the METTL3-METTL14 project in terms of mammalian cell culturing and trained me on protein purification. I have been working closely with Prof. Weixin Tang and Dr. Tong Wu on the project structure-based engineering of the CRISPR-AsCas12f system and have learned from them about CRISPR systems. I also have been studying MsbA transporters with Prof. Arthur Laganowsky, Jixing Lyu, and Samantha Schrecke from Texas A&M University.

I am grateful for the help from Prof. Bozhi Tian and Prof. Chuan He for their time and suggestions as my dissertation committee members.

I cannot finish my research without the support of my parents. They have taken good care of me and love me without any condition, which has been driving me since my childhood. I also greatly thank them for letting me decide my own career path and encouraging me to be the person I want to be.

Abstract

Single-particle cryo-electron microscopy (cryo-EM) analysis plays an essential role in the structural determination of protein complexes that are not suitable for crystallization. Technological and methodological advances have promoted the “resolution revolution”, leading to a boom in structures determined by cryo-EM. In this dissertation, I describe the molecular mechanisms of four proteins by determining the structures of them and their complexes at a near-atomic resolution (3–5 Å). These studies demonstrate the essential roles of single-particle cryo-EM analysis in structural biology and protein sciences. I also describe the future development of cryo-EM and structural biology, with advances in automated blotless sample vitrification, real-time data processing, and more accurate *de novo* protein structure prediction.

List of Publications Based on Work Presented in this Thesis[†]

1. Moghadamchargari, Z., Shirzadeh, M., **Liu, C**., Schrecke, S., Packianathan, C., Russell, D.H., Zhao, M., and Laganowsky, A. (2021) "Molecular assemblies of the catalytic domain of SOS with KRas and oncogenic mutants." *PNAS 118*.
2. Wu, T.*, **Liu, C.***, Zou, S., Zhao, M., and Tang, W. "Engineered hypercompact CRISPR-Cas12f systems with increased gene-editing activity." (in preparation)
3. Lyu, J., **Liu, C.**, Zhang, T., Boone, C. D., Elam, N., Zhao, M., and Laganowsky, A. "A distinct lipid-binding site is associated with lipid-induced stimulation of MsbA ATPase activity." (in preparation)
4. **Liu, C.**, Hu, L., Erramilli, S., He, C., Zhao, M. "Mechanistic studies of methyltransferase METTL3-METTL14 complex." (in preparation)

* Co-authors contributed equally.

[†] The following chapters of the dissertation contain sections and figures adopted from the listed publications with modifications. Chapter 2: publication 4; chapter 3: publication 2; chapter 4: publication 3; chapter 5: publication 1.

Chapter 1 Introduction: Single-Particle Cryogenic Electron Microscopy

Proteins play critical roles in biological processes in organisms. Proteins' structures are significant to understanding their molecular mechanisms in these processes. Scientists have been using various state-of-art techniques for structural determination, including X-ray crystallography [1]–[3], cryogenic electron microscopy (cryo-EM) [4]–[6], and Nuclear Magnetic Resonance (NMR) spectrometry [7]–[9]. X-ray crystallography was pioneered in 1912 and had been developing since then [10]. As of July 6, 2022, 192,489 entries have been deposited in the protein bank (www.pdb.com), with more than 85% of them (166,884) determined by X-ray crystallography, indicating its popularity and advantages. However, since the 2010s, cryo-EM techniques have rapidly progressed with a comparable resolution to X-ray crystallography due to several technical advances, including the development of direct electron detectors [11], [12], automated data collection strategies [13], novel software for image analysis [14]. In 2017, the Nobel Prize in Chemistry was awarded to Jacques Dubochet, Joachim Frank, and Richard Henderson for "developing cryo-EM for the high-resolution structure determination of biomolecules in solution." [15]

Cryo-EM has several advantages over X- ray crystallography [16], [17]: 1) cryo-EM does not require the proteins to crystallize and only needs a small amount of proteins; 2) the proteins are rapidly frozen by liquid ethane, which better maintains the close-to-native state of proteins; 3) various conformations of proteins and their complexes can be detected by cryo-EM; 4) large protein assemblies can be studied by cryo-EM. Due to the advantages mentioned above, there are more structures determined by cryo-EM compared to X-ray crystallography in recent years (Figure 1.1).

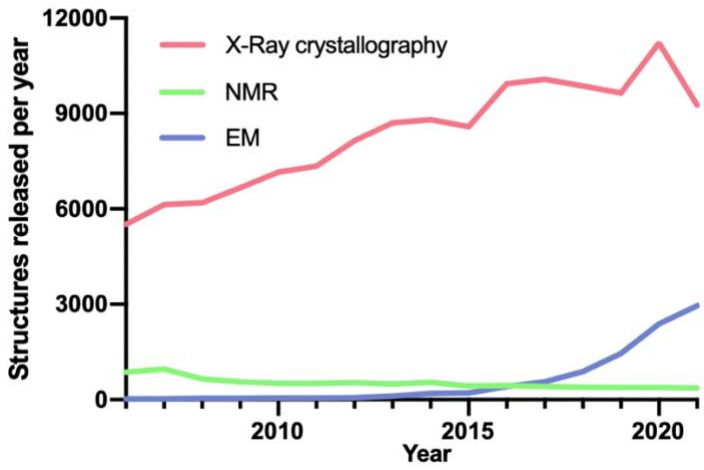

Figure 1.1 Number of structures determined by X-ray crystallography, NMR, and electron microscopy from 2006 to 2021 (www.pdb.com)

1.1 Cryogenic Electron Microscopy Overview: instrument, principle, effect

Transmission electron microscopes (TEM) are to project an accelerated and concentrated electron beam onto a very thin sample, and the electrons collide with atoms in the sample to change direction, thereby producing a three-dimensional angular scattering. The size of the scattering angle is related to the density and thickness of the sample, with more electrons scattered in thick than in thin areas. So, dense areas in samples appear dark in images, and then the image will be displayed on the imaging device (such as phosphor screen, film, and photosensitive coupling components) after magnification and focusing.

Cryo-EM, a type of electron microscopy technique, was invented in the 1930s. It applies to samples, including biological samples [18], [19], and organic and inorganic materials [20], cooled to cryogenic temperatures, preserving the specimens by embedding them into vitreous water. The cryo-EM, as one type of TEM, contains an electron source, lens systems, an energy filter, and a high-resolution camera.

There are two types of electron sources in TEMs [21]: thermionic sources and field-emission sources. Thermionic sources emit electrons when heated, which are not stable and not wavelengths. Field-emission sources produce electrons when a considerable electric potential (100-300 kV) is applied between it and an anode, and the electrons are pulled out from the metal surface and emitted to a vacuum. Compared to thermic sources, field-emission guns produce electrons that are more stable and coherent because they can better control that the electrons come from the same direction (spatial coherent) and have the same speed (temporal coherent) [22]. The field-emission guns are located at the top of cryo-EM, and the electrons are emitted towards lens systems and the sample (Figure 1.2).

The electron lenses have the same functions as the lenses in light microscopes. They can guide the direction of electrons by electromagnetic force. The electron lens systems in cryo-EM are composed of a condenser lens system, an objective lens system, and a projector lens system (Figure 1.2). In each lens system, apertures and stigmators exist right after lenses. Apertures are to reduce the divergence of electron paths through the lenses. Stigmators are to attenuate the problem that electrons focus at a different focal point, which is caused by the fact that lenses are not perfectly round. The condenser lens system is the first lens system right after the electron source. Its function is to gather the electrons emitted from electron sources and focus them on the area that needs to be examined. The objective lens system is the first-stage lens system to magnify the images of the examined area. The quality of the images, including contrast and resolution, are mostly determined by the objective lens system. The projector lens system is the final lens system in the EM and is used to further magnify the images obtained from the objective lens system.

Energy filters are located after the projector lens system. Usually, the electrons that can be used for imaging are unscattered or elastically scattered with the same energy. However, some electrons hit the sample and are inelastically scattered and emerge at lower energy. These electrons are incoherent and can cause low-contrast problems when forming an image. The function of the energy filter is to get rid of those inelastically scattered electrons to ensure the quality of images.

The application of direct electron detection cameras highly promoted cryo-EM in the last ten years [23]. In the past, the cryo-EM was usually equipped with charge-coupled devices, which transform the electrons signals into photon signals first by scintillators, then the photons are counted. The main disadvantage of charge-coupled devices is their low signal-to-noise ratio due to the transformation step. Direct electron detection cameras can directly count the electrons without transforming electron signals, improving detective quantum efficiency when detecting images formed by electrons and enabling images acquired in movie stack mode, which provides comprehensive data for processing.

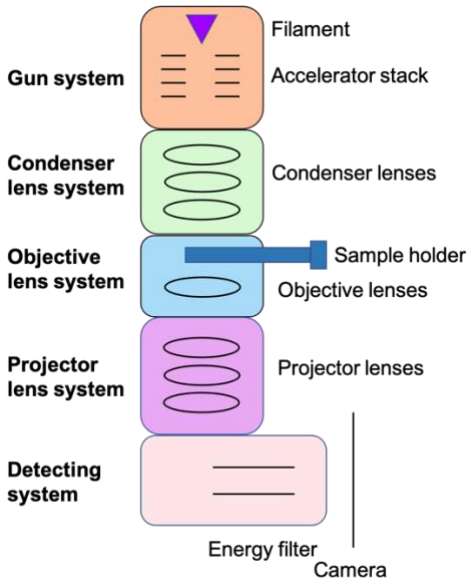

Figure 1.2 Anatomy of transmission electron microscope

1.2 Sample preparation

Cryo-EM sample preparation is the process of applying a pure protein aqueous sample to a grid, reducing its dimension to a ~100–800 Å layer, which is thin and stable in a vacuum, to obtain TEM images at near-atomic resolution, and then flash freezing the layer to prevent the water from crystallizing.

Vitrobot, an automatic plunge-freezer from the FEI company, is the most developed on the market and is used in all my experiments (Figure 1.3). In Vitrobot, the sample is first manually applied to the grid, followed by blotting using blot filter paper, with the blotting time and angle adjusted. After blotting, the grid is plunged into liquid ethane (Figure 1.3).

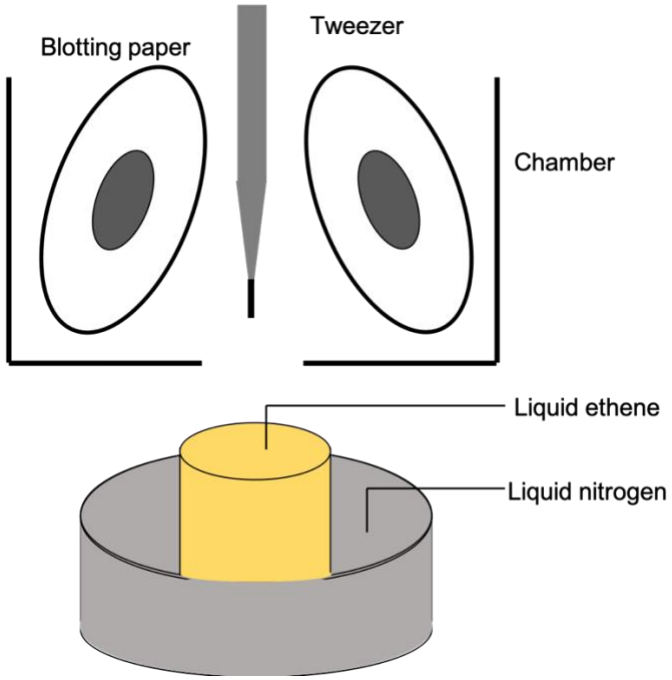

Figure 1.3 Illustration of FEI Vitrobot™

There are three factors in cryo-EM sample preparation [24]–[26]. First, purity is still a prerequisite for cryo-EM, although the proteins are not required to be crystallized and image processing algorithms can identify componential heterogeneity at some level. Second, the

thickness of ice is required to optimize for all samples. It is generally beneficial to produce the ice as thin as possible, but proteins could denature in the extremely thin ice. The concentration of samples is also essential, as the more particles, the more efficient the data collection. However, higher concentration could also result in aggregation and particle distortion. Together, the vitrification conditions should be optimized for each sample before data collection.

1.3 Single-particle analysis and data processing

The single-particle analysis is a computational process to build 3D density maps by combining images of many individual macromolecules in identical or similar conformations obtained from TEM [27]–[30]. In the 2010s, a revolution in structural biology was triggered by single-particle analysis. There are two prerequisites for the success of the analysis method [29], [30]: 1) the frozen molecules are consistent in conformation; 2) the images obtained from low-dose imaging contain enough information to allow the orientation of the underlying particles to be determined .

Briefly, the theory of the single-particle analysis contains several steps [29]–[31]. The macromolecular particles are first identified and extracted from the TEM images to map one image onto another by cross-correlation. Then, these particles in different orientations and/or conformations are classified by image classification algorithms. Meanwhile, these particles are also processed by image filtering and contrast transfer function. Next, the classes of macromolecular particles, which are the projections of the molecules, are reconstructed through the projection-slice theorem.

Many types of software have been developed for single-particle cryo-EM analysis, such as EMAN2 [32], cryoSPARC [33], RELION [34]. CryoSPARC and Relion are the most

developed for processing cryo-EM datasets. CryoSPARC is short for Cryo-EM Single Particle Ab-Initio Reconstruction and Classification. It was developed by Structura Biotechnology Inc., a Canadian software company [33]. RELION is short of REgularized LIkelihood OptimizatioN. It uses a Bayesian approach, in which the reconstruction problem is formulated by identifying the model which is most probably correct with the cryo-EM data [29], [30], [34]. RELION has been developed for years and is one of the most developed software to run single-particle analysis. In conclusion, the single-particle analysis method was developed to extract the information from TEM images of samples, especially for the macromolecules.

Chapter 2 Mechanistic studies of methyltransferase METTL3-METTL14 complex

2.1 Introduction

N^6-methyladenosine (m^{6}A) is methylation that occurs on the adenosine at the N^6-position, which is the most the most prevalent mRNA modification in eukaryotic organisms [35]–[37]. The modification affects the function and localization of mRNA, which in turn, influences protein synthesis and various biomolecular interactions within the cell [35]–[40]. Three classes of proteins are involved in the function of m^{6}A modification: writers (adenosine methyltransferases) [41], erasers (demethylases), and readers (m^{6}A-binding proteins) [42], [43]. m^{6}A modification is installed and removed by writers and erasers, respectively. m^{6}A readers specifically identifies and selectively binds the mRNA sequences with m^{6}A modification, controlling the metabolisms and functions of mRNA. For now, erasers an readers are well characterized, but the composition of m^{6}A complex and the molecular mechanism of how the complex engages mRNA substrate and catalyzes the methyl transfer are still not fully understood.

The m^{6}A writer is a ~900 kDa methyltransferase complex containing two distinct methyltransferase enzymes methyltransferase-like 3 (METTL3) and methyltransferase-like 14 (METTL14) and some auxiliary proteins including WT1 associated protein (WTAP) [44], ZC13H13, RMB15, VIRMA, etc [45] (Figure 2.1a). METTL3 and METTL14 both belong to the β-class of *S*-adenosyl methionine (AdoMet)-dependent methyltransferases and they form a stable complex (METTL3-METTL14 complex) when functioning. Interestingly, the activity of the complex is much higher than the activities of METTL3 or METTL14 alone METTL3-METTL14 complex methylates adenosine in the RRACH consensus (R represents A or G, H represents A,

C, or U), which transfers the methyl group on the AdoMet to the N^6-position of adenosine and eventually produce N^6-methyladenosine and *S*-adenosyl-homocysteine (AdoH).

In 2016, the methyltransferase domains of METTL3-METTL14 in the ligand-free state (Figure 2.1 b,c), AdoMet bound state, and AdoH bound state have been crystallized [46], [47]. However, these structures do not show how METTL3-METTL14 recognizes mRNA, and therefore provide limited information regarding the molecular mechanism. Also, it was shown that the CCCH-typed zinc finger motif adjacent to the N-terminus of the METTL3 methyltransferase domain and the RGG repeats at the C-terminus of METTL14 contribute to the binding of mRNA, respectively [48]. However, controversy existed in different research. Therefore, my research goals are 1) understanding the functions of other important regions in the METTL3-METTL14 complex; 2) drive mechanistic insights into the METTL3-METTL14

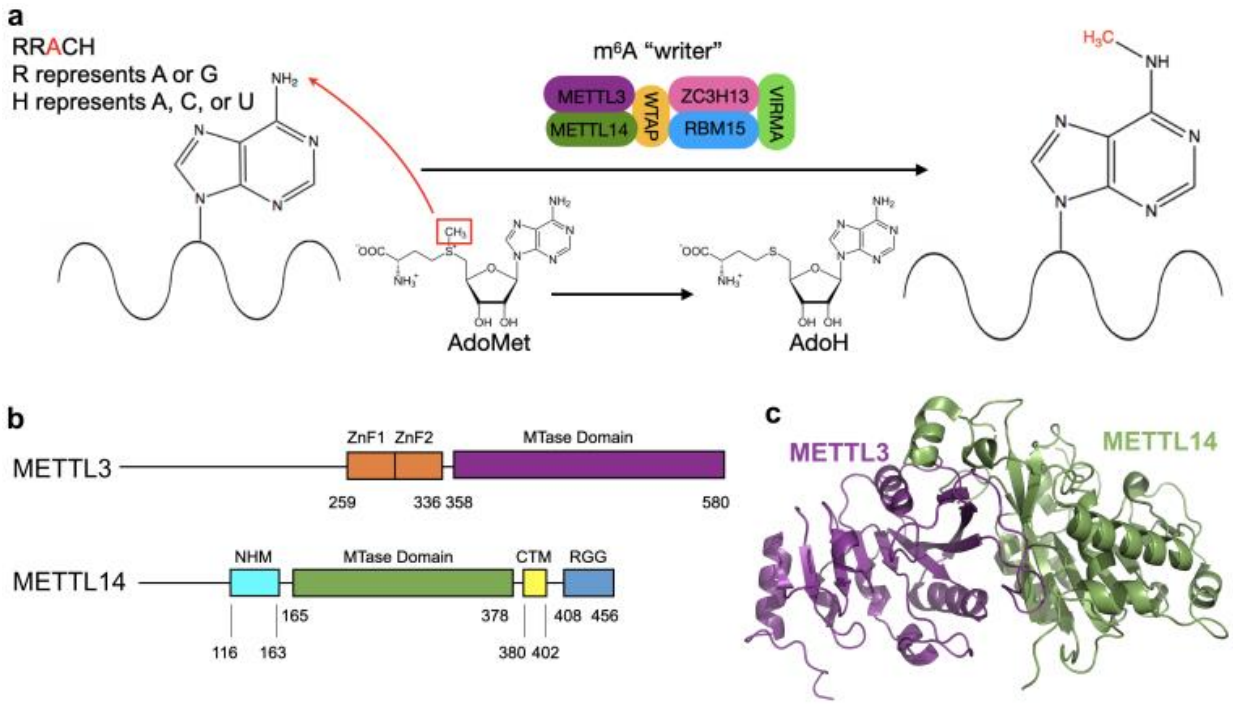

Figure 2.1 Research progresses of N^6-adenosine methylation and m^{6}A writer.
a. The proposed possible *in-vivo* process of N^6-adenosine methylation by m^{6}A writer.
b. Domain diagrams of METTL3 and METTL14
c. Crystal structure of truncated METTL3-METTL14 complex (PDB: 5IL0) [48]

2.2 Results

2.2.1 Assemble and determine the structure of the *in-vivo* full m^{6}A writer complex

To study the full m^{6}A writer complex, my collaborator, Dr. Lulu Hu from Prof. He's lab, cultured ~1,000 plates of mammalian cells, followed by the complex pull-down using flag-tag affinity chromatography. The eluted proteins were crosslinked by GraFix, followed by size exclusion chromatography (SEC). However, the SEC profile revealed that nucleotides were crosslinked to the m^{6}A writer complex (Figure 2.2a). Besides, the representative cryo-EM micrograph (Figure 2.2b) also suggested that the *in-vivo* crosslinked complex aggregated and could not form homogeneous particles. Together, it was difficult, laborious, and expensive to assemble the *in-vivo* full m^{6}A writer complex and determine its structure.

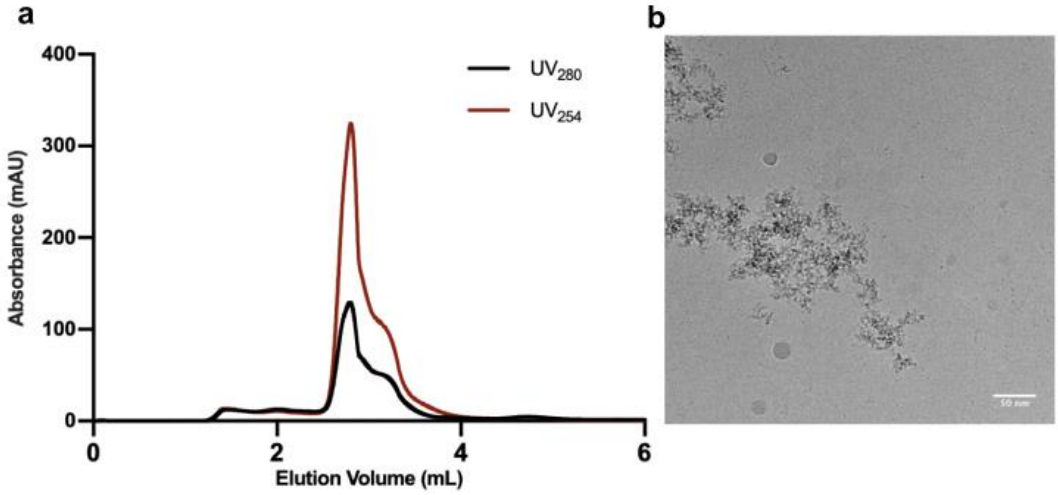

Figure 2.2 *In-vivo* pulled-down m^{6}A writer cannot be purified to absolute homogeneity
a. Size exclusion chromatography profile of the *in-vivo* full m^{6}A writer complex
b. The representative cryo-EM micrograph of the *in-vivo* pulled-down m^{6}A writer is shown along with a 50-nm scale bar.

2.2.2 Assemble m^{6}A writer complex *in vitro*

Since it was difficult to study the m^{6}A writer complex by pulling down the full complex *in vivo*, I tried to assemble the m^{6}A writer complex *in vitro*, starting with WTAP, METTL3, and

METTL14. WTAP, a not well-studied protein as METTL3-METTL14, is vital to both transitional and post-transitional regulation of genes [49], [50]. It is localized on the nucleus, in speckles, and partially colocalizes with splicing factors. Previous research demonstrated that WTAP plays an essential role in the localization of the m^6A writer complex.

WTAP was first overexpressed by *E. coli* or insect cells, followed by Ni-NTA affinity chromatography and size exclusion chromatography. However, the WTAP cannot fold compactly by itself. SEC profile showed that WTAP was eluted close to the void volume (Figure 2.3a), indicating it either is of a non-globular shape or forms oligomers. The negative staining micrograph (Figure 2.3b) and AlphaFold2 prediction (Figure 2.3c) demonstrated that WTAP is of elongated shape instead of folding compactly.

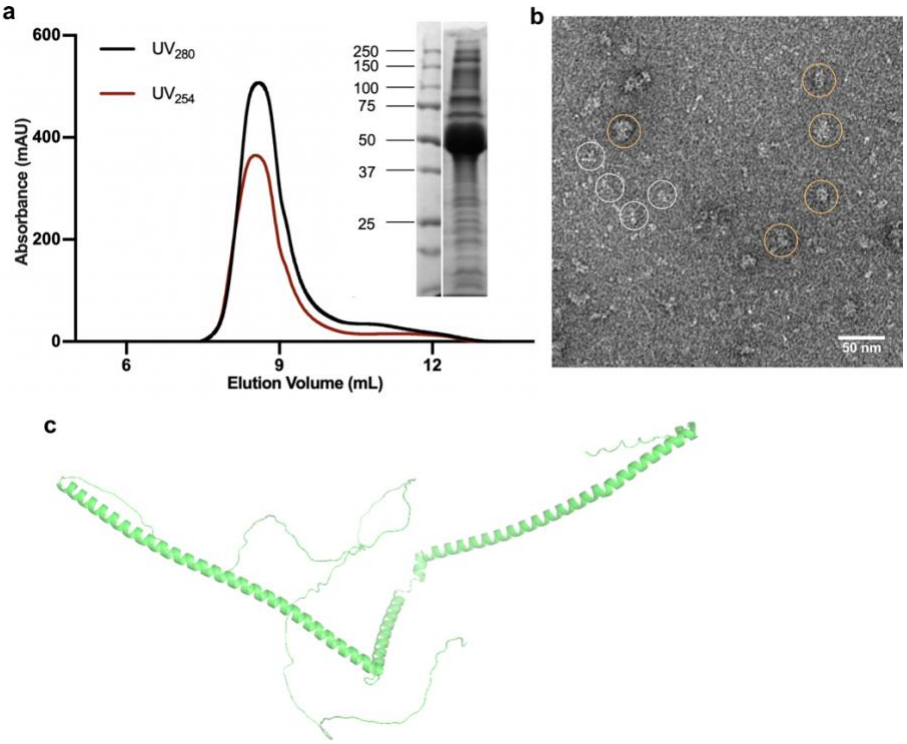

Figure 2.3 Recombinant WTAP cannot be purified to absolute homogeneity.
a. Size exclusion chromatography profile of WTAP
b. The representative negative staining EM micrograph of WTAP is shown along with a 50-nm scale bar
c. Structural prediction of WTAP using AlphaFold2 [51]

I next investigated if WTAP could form a stable and compact ternary complex with METTL3-METTL14 by co-expressing WTAP, METTL3, and METTL14 in *E. coli*, with an N-terminal His-tag on the METTL14, followed by Ni-NTA affinity chromatography. However, the elution only contained METTL3-METTL14 complex, with WTAP in the flowthrough, suggesting that WTAP interacts with METTL3-METTL14 weakly *in vitro* and probably require other factors to form a stable and compact complex. But the critical components in the full m^6A writer are still unknown.

2.2.3 Functions of the zinc finger motif and the RGG repeats in METTL3

In addition to the full m^6A writer complex, I also studied the functions of critical regions in the METTL3-METTL14, considering the truncated construct, whose structure was determined by previous studies, has no catalytic activity.

Previous studies and our sequence analysis suggested that the zinc finger motif at the N-terminus of the MTase domain in METTL3 and the RGG repeats at the C-terminus of the MTase domain in METTL14 are probably critical to RNA binding and recognition. However, it was not clear about their exact functions, and controversy existed in various research. To determine the functions of the zinc finger motif and the RGG repeats, four truncated constructs of METTL3-METTL14 containing or not containing the zinc finger motif and the RGG repeats were cloned and purified (Figure 2.4a), and three 12-nt RNA oligos containing GGAUC or modified consensuses were designed.

RNA gel shift assay for these four truncated constructs towards three RNA oligos revealed that the RGG repeats have a strong binding affinity towards RNAs, while the zinc finger motif shows a weak affinity, as suggested by 1) the METTL3-METTL14 truncated

construct with the RGG repeats but without the zinc finger motif, has a comparable binding affinity with all three RNAs with or without RRACH consensuses; 2) the truncated construct with the zinc finger motif but without the RGG repeats showed a very weak binding affinity.

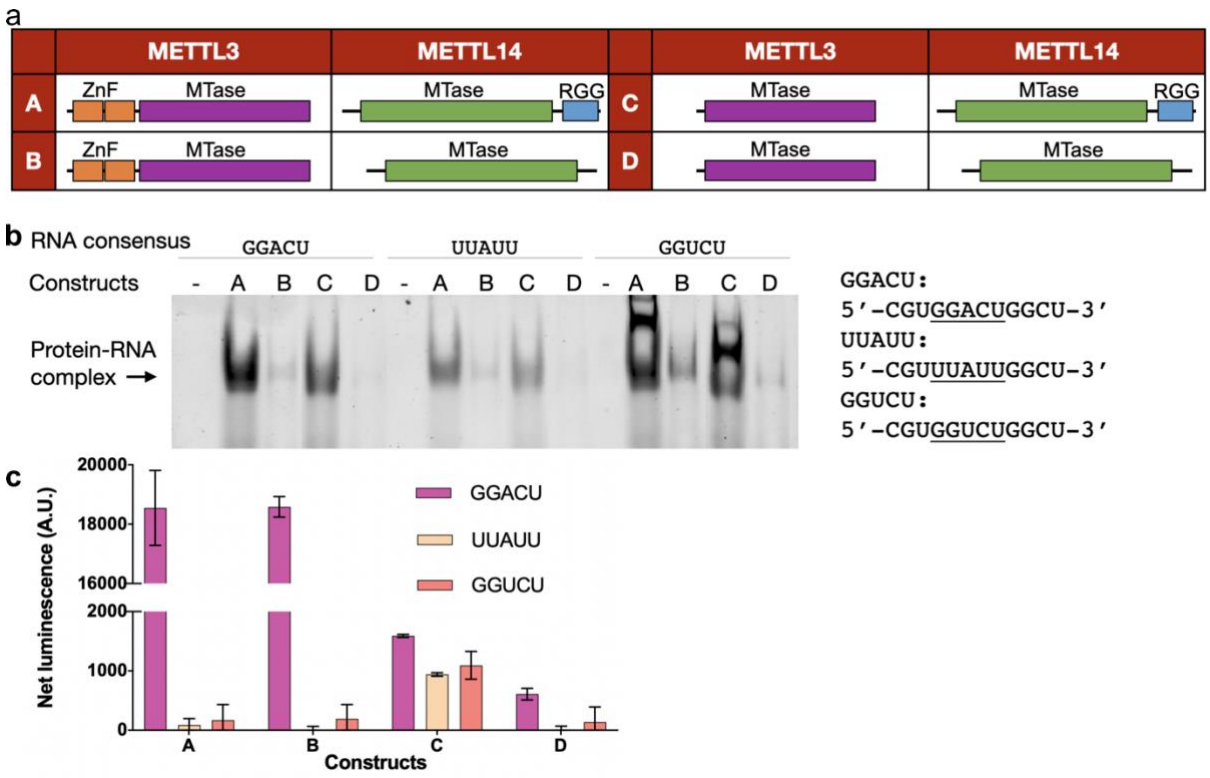

Figure 2.4 RNA-binding affinity and catalytic activity of 4 truncated constructs of METTL3-METTL14.
a. Domain structures of 4 truncated constructs of METTL3-METTL14
b. Gel shift assay of the 4 constructs with 12-nt RNA containing different consensuses
c. Catalytic activity of 4 truncated constructs of METTL3-METTL14 towards 12-nt RNA containing different consensuses

Besides the RNA binding affinity, the methyltransferase activity of these constructs towards different RNAs was also characterized by a methyltransferase activity assay, MTase-Glo™. The methyltransferase activity assay (Figure 2.4c) showed that the zinc finger motif in METTL3 is critical for the specific RNA consensus recognition and methyl group transfer, while RGG repeats are not, as suggested by that the two constructs containing the zinc finger motif,

have catalytic activity towards the RNA containing specific RNA consensus. Meanwhile, these two constructs showed no catalytic activity for the adenosine not in the specific consensus GGACU.

Together, the RNA gel shift assay and methyltransferase activity assay indicated that the RGG repeats could bind RNA strongly without specificity, while the zinc finger motif have a weak affinity with RNA but can recognize the specific RNA consensus.

2.2.4 Crystallization studies of METTL3-METTL14 constructs

To study how METTL3-METTL14 recognizes the specific RNA consensus, I performed crystallization studies for construct B. The reasons that the construct B is optimal for crystallization studies are: 1) the zinc finger motif is important to RNA recognition and catalysis; 2) the RGG repeats is a flexible low-complexity domain and could cause aggregation of constructs; 3) construct B is ~65 kD and suitable for crystallization. After screening a wide range of crystallization states, I obtained some decent crystals of construct B at apo state (Figure 2.5b).

Additionally, to study the interactions between the METTL3-METTL14 complex and RNA oligos with specific consensus, I also co-crystalized the construct B with a 12-nt RNA oligo containing GGACU consensus. However, after many trials, no crystals were observed, probably because RNA oligos degraded at 21 ^{o}C after one week or the complex of construct B bound RNA oligos could not crystalize. To solve the RNA degradation problem, I co-crystalized the construct B with a 12-nt DNA oligo containing GGACT consensus instead, because previous research [52] suggested that METTL3-METTL14 can methylate DNA with specific consensus RRACH (R represents A and C, H represents A, C, or T), similar to that of mRNA. The DNA methylation activity of METTL3-METTL14 was also demonstrated by the methylation assays

(Figure 2.5c). After screening and optimization, suitable crystals for X-ray diffraction data collection were obtained (Figure 2.5b).

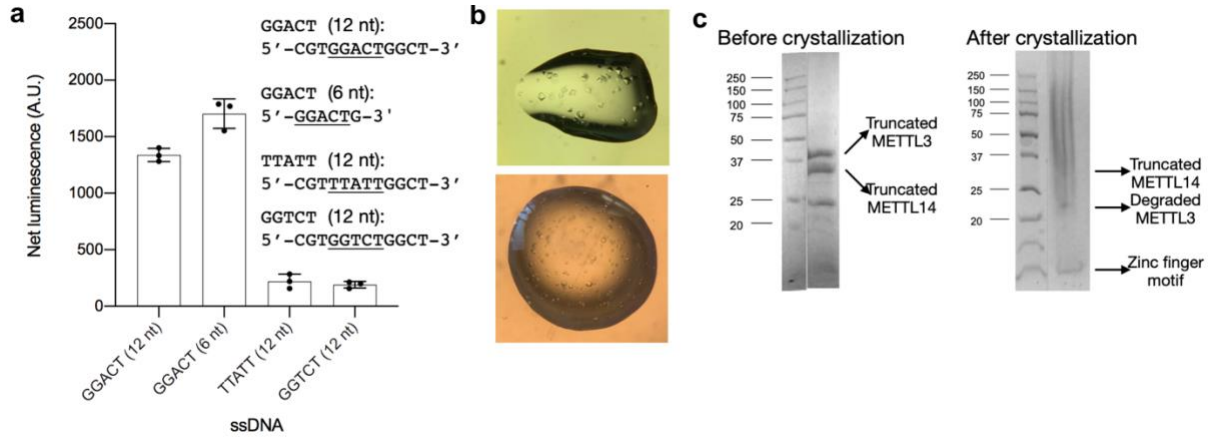

Figure 2.5 DNA methylation activity and crystallization of construct B
a. Catalytic activity of construct B towards DNA containing different consensuses with different lengths
b. Photos of the crystals of apo construct B (upper panel, crystallization buffer: 0.15 M ammonium tartrate (pH 7.0), 24% PEG 3350, 1 µL 14.0 mg/mL construct B + 0.5 µL crystallization buffer) and construct B co-crystallized with 12 nt DNA (5'-CGTGGACTGGCT-3') (lower panel, precipitation buffer: 0.1 M MES (pH 6.0), 0.15 M $(NH_4)_2SO_4$, 20% PEG 8000, 2 µL 10.2 mg/mL construct B binding DNA (molar ratio 1: 5) + 1 µL crystallization buffer)
c. The composition of the proteins before and after crystallization were analyzed by SDS-PAGE.

The structures of apo construct B and construct B binding DNA were determined at resolution of 1.9 Å and 2.1 Å, respectively. However, the electron density of the zinc finger motif could not be observed in the both structures, indicating that the zinc finger motif was either too flexible to be resolved or it was proteolyzed during crystallization. To verify my hypothesis, I performed a SDS-PAGE analysis of the protein before crystallization and the crystals, and it suggested that the zinc finger motif was auto proteolyzed during one week at 21 ^{o}C, with its truncation site probably located at the end of the linker between the zinc finger motif and MTase domain, considering the density could not be resolved right before the MTase domain in METTL3.

To avoid the auto proteolysis problem, some mutations were introduced to stabilize the linker in METTL3. Four constructs were designed (Table 2.1). The possible cleavage site leading to hydrolysis was truncated in construct 1. In construct 2 and construct 3, the linker was mutated to the GGS repeats, a flexible linker commonly used in protein engineering, with different lengths. In construct 4, the linker was mutated to that of METTL3 in zebrafish and tis sequence was verified in UniProt.

Table 2.1: The sequences of the mutated linker between the zinc finger motif and the MTase domain in METTL3

No.	Linker name	Sequence
	Linker between the zinc finger motif and MTase domain in the human wild-type METTL3	YEIDA CMDSE APGSK DHTPS QELAL TQSVG GDSSA DRL
1	Truncated linker	YEIDA CMDSE APGSK DHTPS QELAL TQSVG
2	$(GGS)_5$ linker	GGS GGS GGS GGS GGS
3	$(GGS)_{10}$ linker	GGS GGS GGS GGS GGS GGS GGS GGS GGS GGS
4	Zebrafish linker	YEIDS PPEAE GDALG PQAGA AELGL HSTVG DSNVG KL

To confirm the constructs are active, their RNA methylation activity were characterized and compared to the wild-type construct B and full-length METTL3-METTL14. The assay demonstrated that these mutated constructs have comparable catalytic activity except the construct 3 (Figure 2.6b), the construct B with a $(GGS)_{10}$ linker, probably related to its length and flexibility. The crystallization conditions of constructs 1, 2, 4 were screened and optimized. However, no crystal was obtained after screening and optimizing a wide range of conditions.

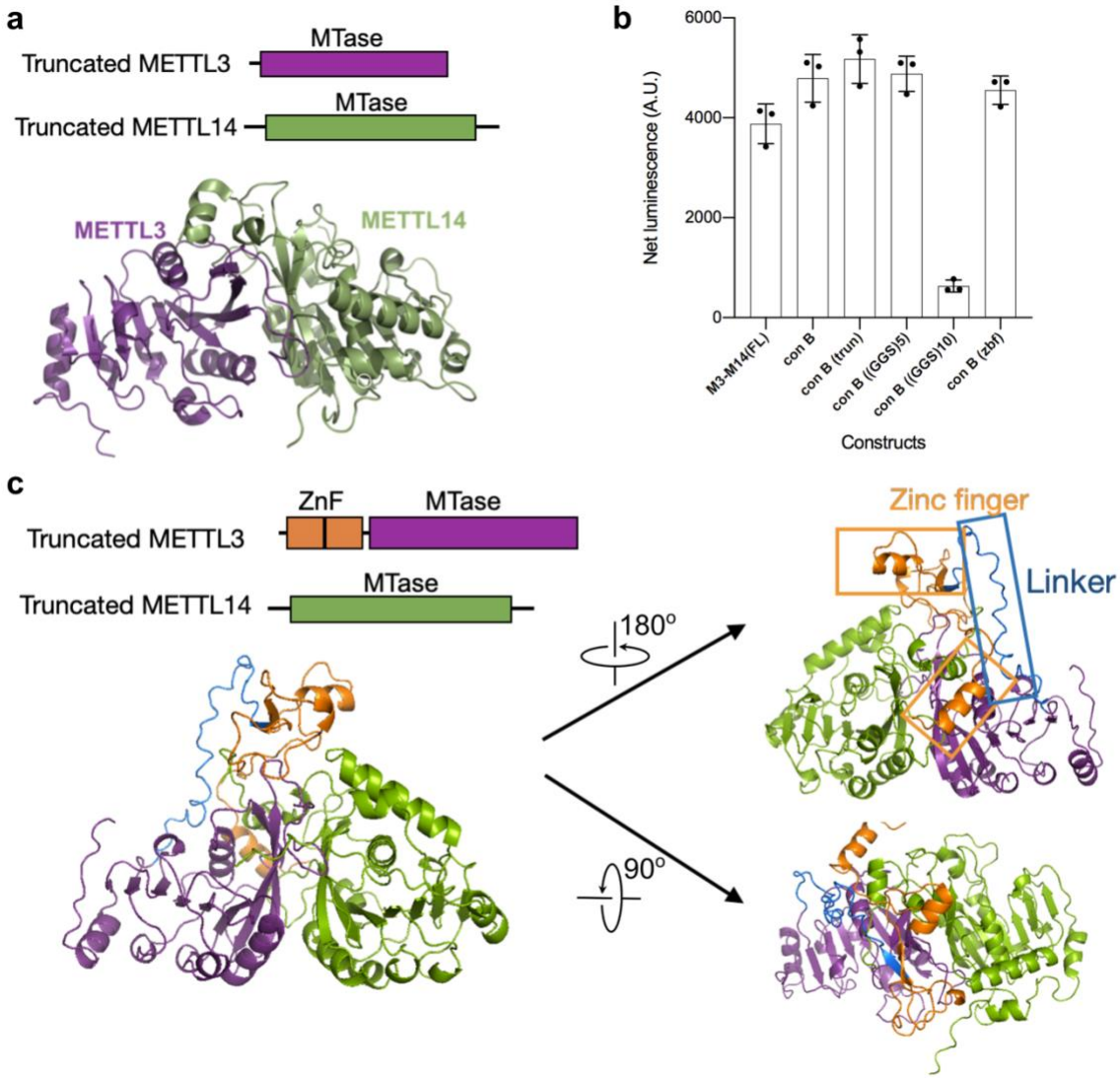

Figure 2.6 METTL3-METTL14 truncated constructs with mutated \ between the zinc finger and the MTase domain in METTL3.
a. Domain diagram and crystal structure of MTase domains of METTL3-METTL14 complex (PDB:5IL0) [48]
(Figure 2.6 continued) b. RNA methylation activity of the four METTL3-METTL14 truncated constructs with mutated linkers
c. Predicted structure of construct B by AlphaFold2 [51]. The color code of the MTase domains is the same as in Figure 2.1. The zinc finger motif and the linker are colored in orange and blue, respectively.

Alphafold2 [51], a powerful tool for protein structure prediction, was used to predict the structure of construct B. The predicted structure was compared to the known structure of the complex of the MTase domains of METTL3-METTL14. The prediction and comparison showed

that the predicted structure of the MTase domains re is very close to the their crystal structure (Figure 2.6c). However, the zinc finger motif and the linker are very flexible in the prediction, in line with the experimental results.

2.2.5 Cryo-EM studies of Fab-construct B

Because the zinc finger motif was auto-proteolyzed during crystallization, I also tried to determine the structure of construct B using cryo-EM, considering that instant protein vitrification during cryo-sample preparation can probably avoid degradation. However, the small size of construct B (~65 kD) led to difficulties in reconstructing the model using single-particle cryo-EM analysis. To address the difficulties, monoclonal antigen-binding fragments (Fabs), ~50 kD antibody fragments that can specially stabilize and enlarge target proteins, were screened using phage display, in collaboration with Dr. Satchal Erramilli from Prof. Tony Kossiakoff's lab. Fifteen Fabs that can specifically form stable complexes with construct B were identified. To characterize their binding affinities with construct B, I performed an Elisa assay, and it showed that all the Fab candidates had a strong binding affinity towards construct B (Figure 2.7, Table 2.2).

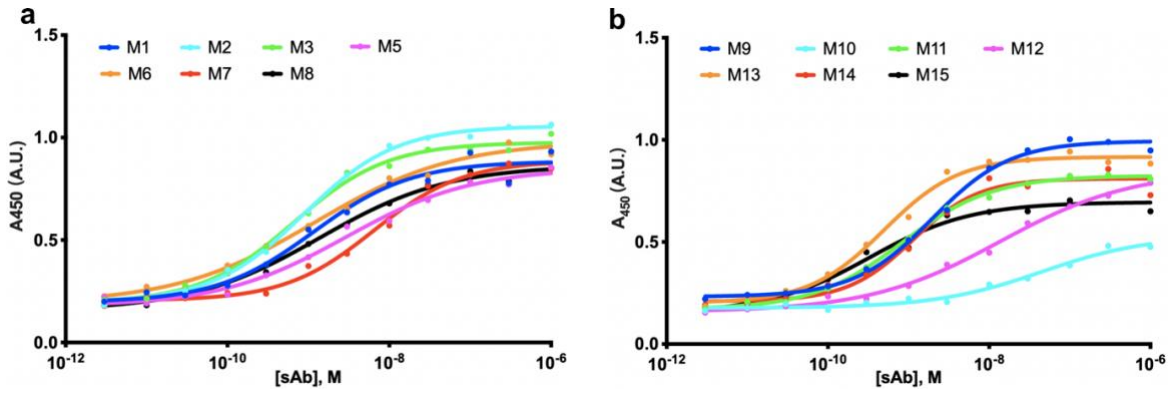

Figure 2.7 Binding affinities of a. FabM1-FabM8 and b. FabM9-FabM15 towards construct B

Table 2.2: IC_{50} value of Fabs towards construct B

Fab	IC_{50}
FabM1	1.163×10^{-9}
FabM2	8.738×10^{-10}
FabM3	6.545×10^{-10}
FabM5	2.885×10^{-9}
FabM6	1.193×10^{-9}
FabM7	6.397×10^{-9}
FabM8	1.373×10^{-9}
FabM9	1.953×10^{-9}
FabM10	4.515×10^{-8}
FabM11	8.353×10^{-10}
FabM12	1.393×10^{-8}
FabM13	4.980×10^{-10}
FabM14	1.130×10^{-9}
FabM15	3.299×10^{-10}

To assemble the complex of Fab and construct B (Fab-construct B), the Fab candidates and construct B were mixed at a 1:1 ratio and then purified by SEC, respectively. SEC profiles showed that all the Fab candidates could form stable complexes with construct B (Figure 2.8), as demonstrated by that Fab-construct B was eluted at a higher volume in SEC. I also visualized the complexes of construct B with all the Fab candidates using cryo-EM, and identified FabM2 as the most suitable candidate because the homogenous particles of FabM2-construct B were observed on its cryo-EM micrographs (Figure 2.12a).

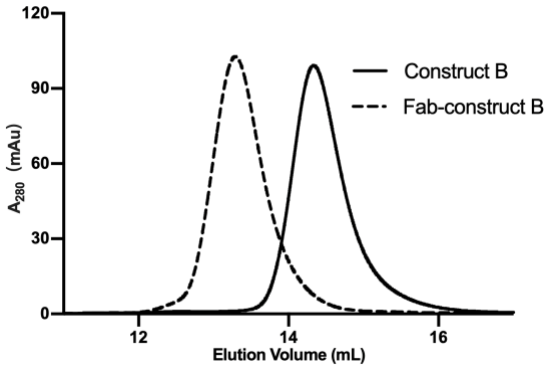

Figure 2.8 A representative size exclusion chromatography profiles of Fab-construct B complex

After screening and optimizing the vitrification conditions, a cryo-EM dataset was collected for the FabM2-construct B, and its structure was determined at a resolution of 2.7 Å (Table 2.3, Figure 2.9). The structure showed that FabM2 and construct B form a complex at 1:1, with FabM2 bound to the truncated METTL3 (Figure 2.9). However, the zinc finger motif was too flexible to be resolved in the cryo-EM map. To stabilize the zinc finger motif, I assembled FabM2-construct B in complexes with a 12-nt RNA oligo or the RNA and the sinefungin, a pan-inhibitor against AdoMet-dependent methyltransferase. Unfortunately, the density of the zinc finger motif was still too weak to resolve in the maps of these complexes.

Table 2.3: Statistics of cryo-EM model refinement and geometry for FabM2-construct B

Model	
Composition (#)	
Chains	4
Atoms	7164 (Hydrogens: 0)
Residues	Protein: 914 Nucleotide: 0
Water	0
Ligands	0
Bonds (RMSD)	
Length (Å) (# > 4σ)	0.003 (0)
Angles (°) (# > 4σ)	0.665 (2)
MolProbity score	1.76
Clash score	11.89
Ramachandran plot (%)	
Outliers	0.11
Allowed	2.89
Favored	97.00
Rama-Z (Ramachandran plot Z-score, RMSD)	
whole (N = 900)	-0.49 (0.28)
helix (N = 126)	-0.83 (0.40)
sheet (N = 273)	-0.00 (0.32)
loop (N = 501)	-0.25 (0.28)
Rotamer outliers (%)	0.00
Cβ outliers (%)	0.00
Peptide plane (%)	
Cis proline/general	8.0/0.0
Twisted proline/general	0.0/0.0
CaBLAM outliers (%)	2.03

Table 2.3 continued: Statistics of cryo-EM model refinement and geometry for FabM2-construct B

ADP (B-factors)	
Iso/Aniso (#)	7164/0
min/max/mean	
Protein	12.33/90.60/43.21
Data	
Box	
Lengths (Å)	65,97, 73.48, 156.98
Angles (°)	90.00, 90.00, 90.00
Supplied Resolution (Å)	2.7
Resolution Estimates (Å)	Masked
d FSC (half maps; 0.143)	2.8
d 99 (full/half1/half2)	2.9/2.4/2.4
d model	2.9
d FSC model (0/0.143/0.5)	2.6/2.7/3.0
Map min/max/mean	-2.37/4.31/0.04
Model vs. Data	
CC (mask)	0.75
CC (box)	0.70
CC (peaks)	0.64
CC (volume)	0.72

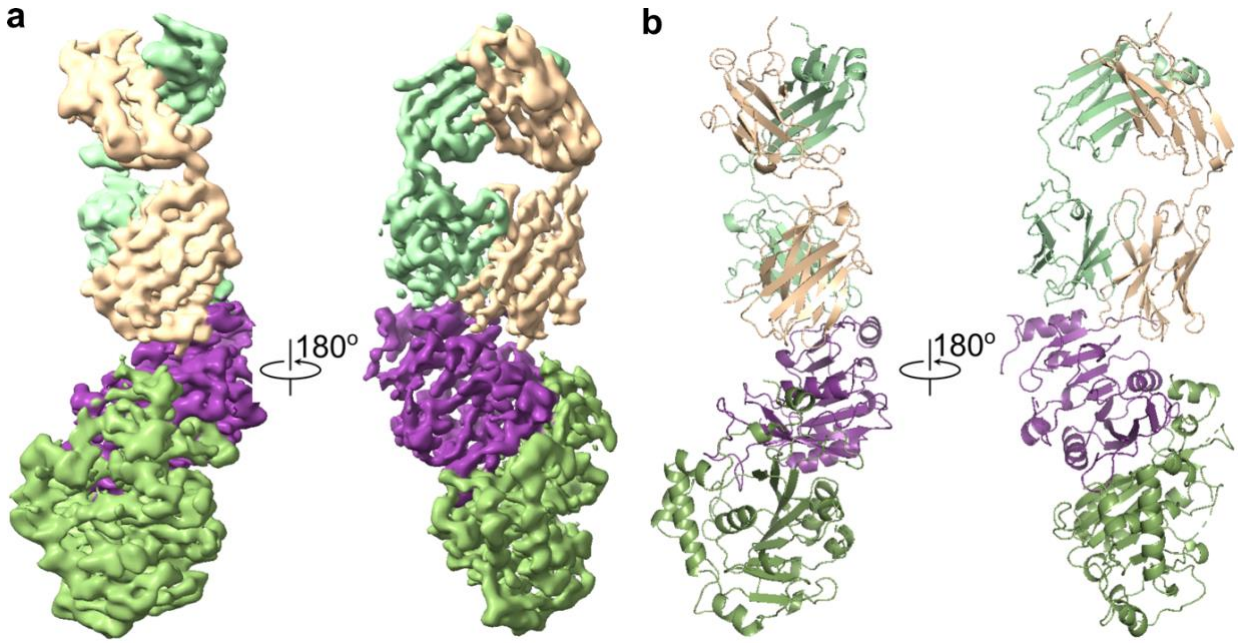

Figure 2.9 Structure of FabM2-construct B
a. Unsharpened cryo-EM map of FabM2-construct B (contoured at a level of 0.3). The color code of the MTase domains is the same as that in Figure 2.1. The heavy chain and light chain of FabM2 are colored in wheat and light green, respectively.

(Figure 2.9 continued) b. The structure of FabM2-construct B complex. The color code is the same as in a.

2.2.6 Dimerization of METTL3-METTL14

Surprisingly, the dimer of FabM2-construct B also existed in solution, with its structure determined from the same cryo-EM dataset (Figure 2.10a,b). The dimer was formed due to the electrostatic interactions between the two METTL3 molecules (Figure 2.10b, Table 2.4). Previous studies also showed the dimerization of METTL3-METTL14 (Figure 2.10c) [53], but it was not studied, and the functions of the dimer are unknown.

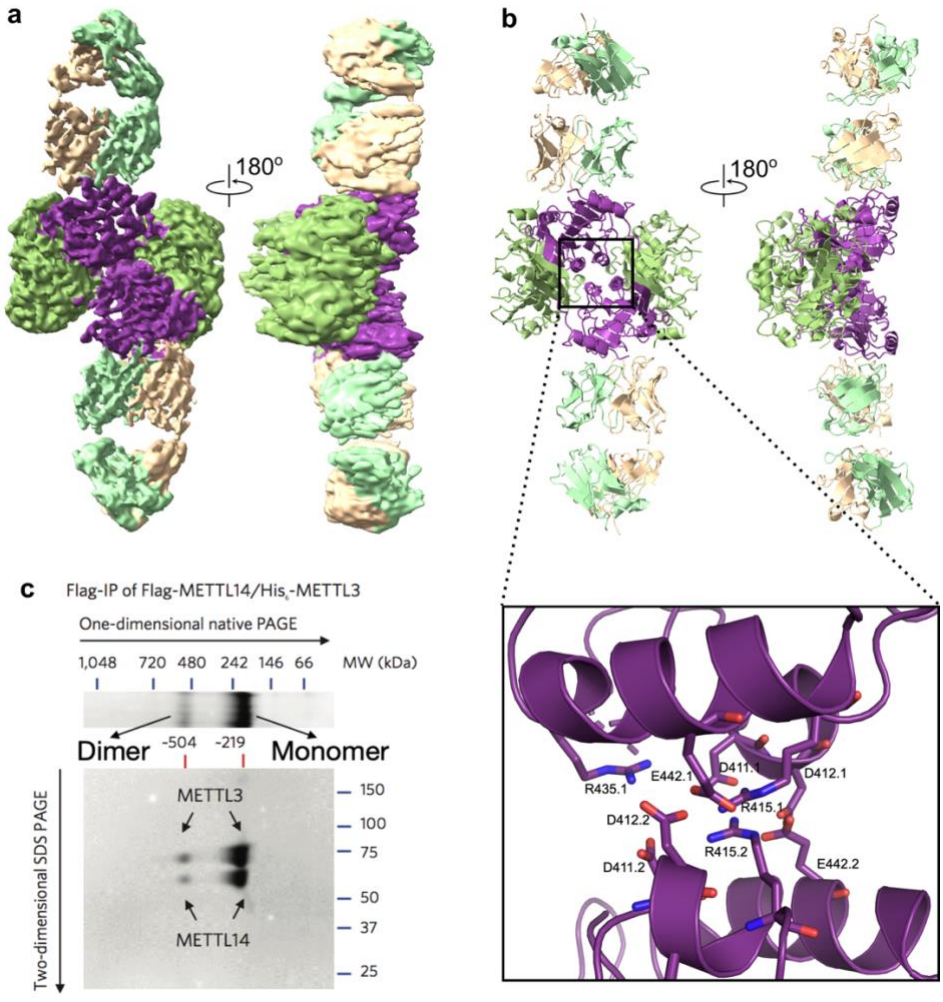

Figure 2.10 Structure of FabM2-construct B dimer

(Figure 2.10 continued) a. Unsharpened cryo-EM map of FabM2-construct B dimer (contoured at a level of 0.15). The color code is the same as in Figure 2.9.
b. Structure of FabM2-construct B dimer. The interactions on the dimer interfaces are labeled.
c. Two-dimensional native gel of METTL3-METTL14 from previous studies. [53]

To study the functions of METTL3-METTL14 dimer, I performed some mutagenesis studies. First, to interfere the formation of the dimer, the arginine at position 415, an essential residue for almost all salt bridges and hydrogen bonds on the dimer interface (Figure 2.10b, Table 2.4), was mutated to aspartic aid, resulting in repulsive electrostatic interactions between the two METTL3 molecules. The SEC profile and the native gel showed that the formation of the construct B dimer was interfered by the R415D mutation (Figure 2.11a,b). Next, the RNA methylation activity of the wild-type construct B and its R415D mutant towards RNAs containing different consensuses were characterized, and found that the R415D mutation led to a ~15% decrease in methylation activity of construct B. Together, these results showed that the R415D mutation disturbs the dimerization of construct B and the methylation activity.

Table 2.4: Interactions between the dimer interface

METTL3.1	METTL3.2
D411.1	R415.2
D412.1	R415.2
R415.1	E442.2
R415.1	D412.2
R435.1	D411.2
E442.1	R415.2

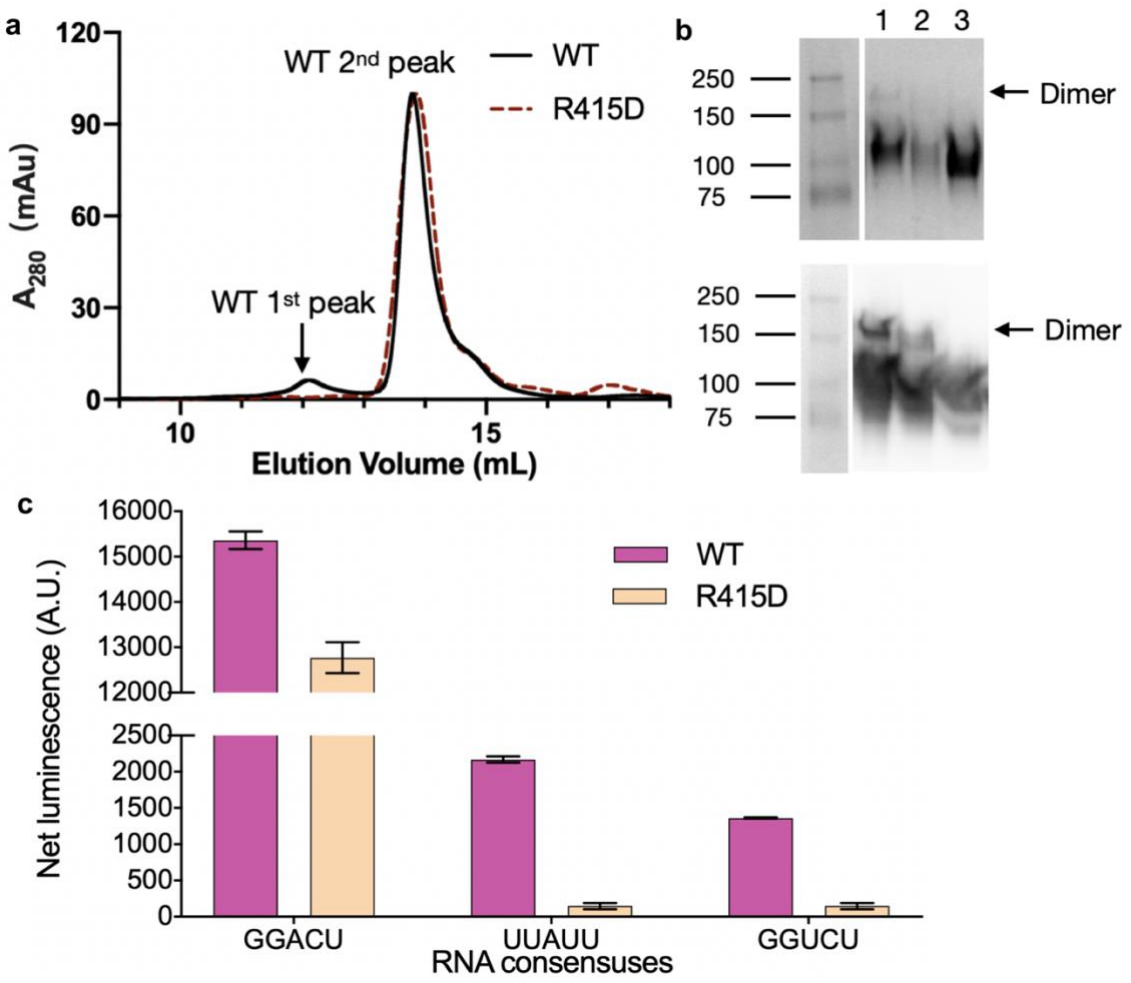

Figure 2.11 R415D mutation can interfere the dimer formation and methylation activity of construct B

a. The size exclusion chromatography profiles of the wild-type construct B and its R415D mutant.

b. Native gel (upper panel) and its western blot of peak fractions (1: WT 1st peak; 2: WT 2nd peak; 3 R415D mutant peak) in size exclusion chromatography in panel a

c. Methylation activity of the wild-type construct B and its R415D mutant towards 12-nt RNA containing different consensuses.

2.3 Discussion and conclusion

In my study, the functions of the zinc finger motif and the RGG repeats have been identified that the RGG repeats bind RNA strongly without specificity, and the zinc finger motif has a weak affinity with RNA but can catalyze methyl transfer from AdoMet to adenosine in a

specific RNA consensus. Besides, cryo-EM studies showed that the monomer and dimer of METTL3-METTL14 heterodimer maintain equilibrium in solution, with the dimer formed due to the electrostatic interactions between the METTL3 interfaces. Mutagenesis studies showed that R415D, a mutation disturbing the METTL3-METTL14 dimer formation, decreased the RNA methylation activity by ~15%. Two major drawbacks of the study are: 1) the cryo-EM map of the dimer was preferentially orientated, resulting in a low confidence of the dimer structure; 2) the biological functions of the METTL3-METTL14 dimer were not identified. To address these two problems, my next steps are 1) stabilizing and crosslinking the dimer by mutating D412 and E442 to cysteines to form disulfide bonds between the two METTL3 molecules; 2) introducing the R415D mutation in mammalian cells to identify the functions of METTL3-METTL14 dimer.

2.4 Methods

2.4.1 Materials

All chemicals used in buffer preparation were purchased from Sigma-Aldrich. RNA and DNA oligos were purchased from Integrated DNA Technologies, Inc. (IDT)

2.4.2 Molecular cloning

The plasmid containing human METTL3-METTL14 gene was obtained from Dr. Lulu Hu. The DNA sequence was cloned into the vector pETDuet1 containing an N-terminal His_6-tag at the N-terminus of METTL14. The mutants of METTL14 was generated by site-directed mutagenesis.

The plasmids of all the Fabs were obtained from Dr. Satchal Erramilli from Prof. Tony Kossiakoff's lab.

2.4.3 Overexpression and purification of METTL3-METTL14 and its truncated constructs

The wild-type His-tagged METTL3-METTL14 and its mutants (constructs A-D) was overexpressed in *E.coli* BL21(DE3). The *E. coli* cells were cultured at 37 ºC until the OD_{600} reached 0.8, and the protein expression was then induced by adding isopropylthio-β-galactoside (IPTG) to a final concentration of 1 mM. The *E. coli* cells were further cultured at 25 ºC for 16 hours before harvest by centrifugation. The cell pellet was resuspended in buffer A (50 mM Tris·HCl, pH 8.0, 300 mM NaCl, 20 mM imidazole, and 0.5 mM TCEP, 1 mM phenylmethylsulfonyl fluoride (PMSF)), and lysed by sonication. The lysate was cleared by centrifugation at 16,500 r.p.m. for 30 min. The supernatant was loaded onto an gravity column containing 2-mL Ni-NTA beads (for 8 L culture) equilibrated by buffer B ((50 mM Tris·HCl, pH 8.0, 300 mM NaCl, 20 mM imidazole, and 0.5 mM TCEP), and the proteins were eluted by buffer C (50 mM Tris·HCl, pH 8.0, 300 mM NaCl, 400 mM imidazole, and 0.5 mM TCEP). The proteins were dialyzed to the 1 L buffer D (50 mM Tris·HCl, pH 8.0, 50 mM NaCl, and 0.5 mM TCEP) overnight. Then, the sample was loaded onto a Mono Q 4.6/100 PE column (GE Healthcare), equilibrated with buffer E (50 mM Tris·HCl, pH 8.0, 50 mM NaCl, and 0.5 mM TCEP) and eluted with a linear gradient of NaCl up to 1 M. The fractions containing purified proteins were collected and concentrated to ~1 mL, and then further purified by size exclusion chromatography using Superdex 200 increase 10/300 column. The purified protein were collected and frozen at -80 ºC for future.

2.4.4 Overexpression and purification of Fabs

All the Fabs were overexpressed in *E.coli* BL21(DE3). The *E. coli* cells were cultured at 37 ºC until the OD_{600} reached 0.8, and the protein expression was then induced by adding IPTG

to a final concentration of 1 mM. The *E. coli* cells were further cultured at 25 ºC for 16 hours before harvest by centrifugation. The cell pellet was resuspended in buffer F (20 mM Tris·HCl, pH 7.4, 300 mM NaCl, 1 mM PMSF), and lysed by sonication. The lysate was then heated to 63 ºC for 30 min and then cleared by centrifugation at 16,500 r.p.m. for 30 min. Next, the supernatant was collected and run through a protein G-sepharose column equilibrated in the buffer G (20 mM Tris·HCl, pH 7.4, 500 mM NaCl). The Fabs were eluted using the elution buffer H (0.1 M glycine, pH 2.7). Then, the Fabs were loaded onto a Mono S column equilibrated in buffer I (50 mM sodium acetate pH 5.0) and eluted using a linear gradient of buffer J (50 mM sodium acetate, pH 5.0, 2 M NaCl).

2.4.5 Electron microscopy sample preparation

Fab-construct B was concentrated to 5.5 mg/mL. Sample vitrification was performed using a Vitrobot Mark IV (Thermo Fisher) operating at 8 ºC and 100% humidity. A total of 3.5 µL of the sample was applied to holey ultra-gold grids (Quantifoil 300 mesh Au 1.2/1.3) glow-discharged for 30 seconds. The grids were blotted for 4 seconds at a blotting force 0 by standard Vitrobot filter paper, and then plunge frozen in liquid ethane.

2.4.6 Data collection

Frozen grids were sent to the Advanced Electron Microscopy Facility at the University of Chicago for data collection. The dataset was acquired as movie stacks using a Titan Krios transmission electron microscope operating at 300 kV and equipped with a K3 direct detector camera (Gatan). Images were recorded at a nominal magnification of 81,000x and super-resolution counting mode by image shift. The total exposure time was set to 4 s with 40 frames

in a single stack and a total exposure of around 50 electrons/Å^2. The defocus range was set at -1.0 to -2.5 μm. See Table 2.4 for the details.

2.4.7 Model Building, Refinement, and Validation

The previously reported structure of MTase domains [48] (PDB: 5IL0) and Fab [54] (PDB: 7kE0) were docked into the cryo-EM map as a rigid body using Chimera [55]. Further model building and refinement was performed in COOT [56], followed by one round of real-space refinement using Phenix [57] with secondary-structure and Ramachandran restraints. Geometry outliers were manually fixed in Coot [56]. The statistics of the final round of model refinement and the model geometry are shown in Table 2.3.

2.4.8 Crystallization of apo construct B and construct B binding 12-nt DNA

Crystals of apo construct B were grown using the hanging drop method and appeared in 1 week at 21 °C from a 2:1 (v/v) mixture of protein (14 mg/mL) and reservoir solution (0.15 M ammonium tartrate pH 7.0, 24% PEG 3350). Crystals were harvested and flash-frozen in liquid nitrogen without any cryo-protectant.

Crystals of construct B binding 12-nt DNA were grown using the hanging drop method and appeared in 1 week at 21 °C from a 2:1 (v/v) mixture of protein-DNA complex (protein concentration: 10.2 mg/mL, protein and DNA were mixed right before crystallization at a molar ration of 1:1.5) and reservoir solution (0.1 M MES pH 6.0, 0.15 M $(NH_4)_2SO_4$, 20% PEG 8000). Crystals were harvested and flash-frozen in liquid nitrogen without any cryo-protectant.

2.4.9 Methyltransferase activity assay

MTase-Glo assays [58] were performed in a 96-well plate with white solid bottom. First, 10 μL of 40 μM enzyme were transferred to the plate, followed by adding10 μL reaction buffer containing 40 μM AdoMet, 2 unit/μL protease inhibitor, 4 μM RNA to each well to trigger the reaction. Then, the plate was incubated at 37 ºC for 1 hr. Upon completion, the conversion of AdoMet to AdoH was detected by adding 5μL 5x MTase-Glo reagent to each well to convert AdoH to ADP then incubating the plate for 30 min at room temperature. Once conversion is complete, 25 μL 2x MTase-Glo Detection Solution was added to each well and then incubated at room temperature for 30 min to convert ADP to ATP, which was then measured by luminescence detection using a using the Synergy HT (Bioteck, Winooski, VT) plate reader.

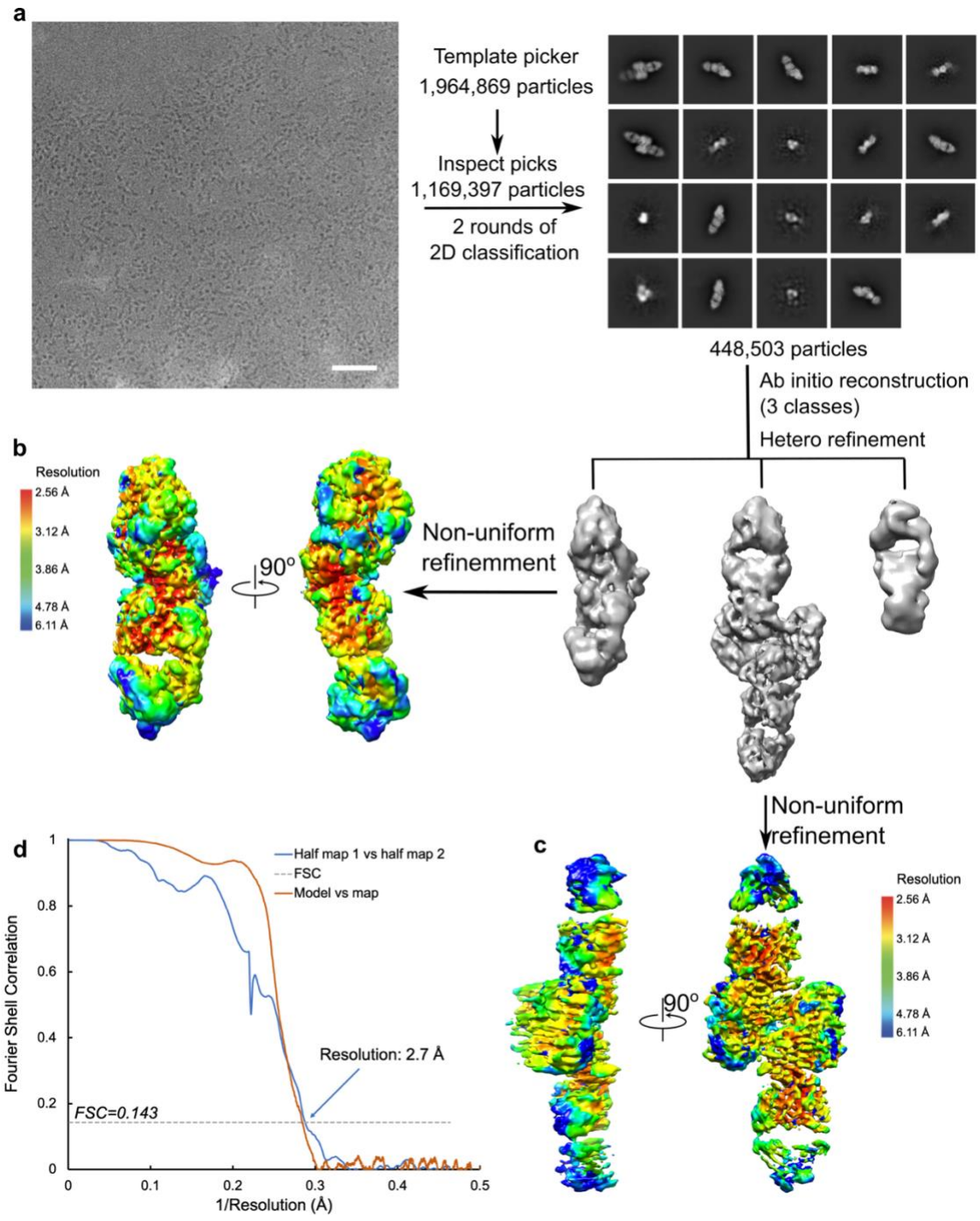

Figure 2.12 Single-particle cryo-EM analysis for FabM2-construct B
a. The workflow of data processing [33]. The representative micrograph is shown along with a 50-nm scale bar. Particle picking was performed using the 2D templates generated by a reconstructed model from earlier runs. Representative 2D class averages are shown, with the box edge corresponding to 334 Å. After disposing contamination and poorly-aligned 2D classes, the particles were classified by hetero refinement with three initial models generated using ab initio reconstruction. A non-uniform refinement was performed on the classes of the monomer and the dimer.

(Figure 2.12 continued) b & c. Local resolution of b. FabM2-construct B monomer and c. FabM2-construct B dimer
d. Fourier shell correlation curves of the final reconstructions and model versus map for Fab-construct B. The resolution of the reconstruction was determined by the FSC=0.143 criterion.

Chapter 3 Structure-based engineering of hypercompact CRISPR-AsCas12f systems

3.1 Introduction

CRISPR (clustered regularly interspaced short palindromic repeats) was first discovered as a family of DNA sequences found in prokaryotes, which are the immunological memories acquired by prokaryotes during subsequent infections [59], [60]. CRISPR, working with Cas (CRISPR-associated) proteins, is now a technology used for genome engineering in eukaryotes [59]–[62]. Based on gene compositions and loci architectures, CRISPR-Cas systems have been classified into two classes, Class 1 and Class 2 [63]. The major difference between the two classes is the number of molecules in effectors: a class 1 effector contains multiple molecules, while a class 2 effector consists of a single molecule. All three types of class II systems – Cas9 (type II) [64], [65], Cas12 (type V) [66], and Cas13 (type VI) [67]–[69] – have family members with confirmed RNA-guided DNA or RNA interference activities in mammalian cells.

The RNA-guided Cas proteins are nucleases that use CRISPR sequences as guidance to recognize and cleave specific DNA strands [70], [71]. To mediate genome alteration in eukaryotic cells, the Cas proteins are usually engineered by adding functional domains to their N- and C-termini, and then the engineered Cas proteins are transfected into the cells through plasmids or viruses [72]. Nowadays, adeno-associated viruses (AAV) are the leading candidate for *in vivo* delivery of CRISPR gene editing systems, because it has a strong track record of transducing multiple tissues in human safely and efficiently [72], [73]. However, the large size of common Cas proteins, such as Cas 9 and Cas12 (950 – 1,400 amino acids), reduces the transfection efficiency and restricts their flexibility in therapeutic applications [72], considering that AAV vectors have a maximum packaging capacity of 4.7 kb [72], [73]. Recently, a group of

Cas proteins, termed Cas12f (consisted of 400 – 700 amino acids), was reported in 2021 [74]. Among this group, Cas12f in *Acidibacillus sulfuroxidans* (AsCas12f) can function as an effective genome-editing tool in both bacteria and human cells by various delivery methods [74], [75].

In this study, we took a structure-based approach to engineering the AsCas12f protein and the single guide RNA (sgRNA) that generates programmed double-stranded breaks (DSBs) in mammalian cells 2- to 10-fold more efficiently than the wild-type protein. I determined a structure of AsCas12f complexed with its gRNA and target DNA at the atomic resolution by cryo-EM. The structure expanded our mechanistic understanding of the hypercompact type V-F effector and guided the engineering of the gRNA and the protein without sacrificing the DNA targeting and cleavage activity.

3.2 Results

3.2.1 AsCas12f purification and AsCas12f-sgRNA-DNA complex reconstruction

AsCas12f was first cloned into the pET47b vector to fuse an N-terminal His-tag. Initial attempts of AsCas12f purification were hindered due to the low yield of AsCas12f in *E. coli* and its high affinity with nucleic acids. After various trials, I obtained a sufficient amount of AsCas12f by overexpressing the proteins at a lower temperature and separated the proteins and nucleic acids by Heparin chromatography (Figure 3.1), which removed the DNA molecules by charge repulsion.

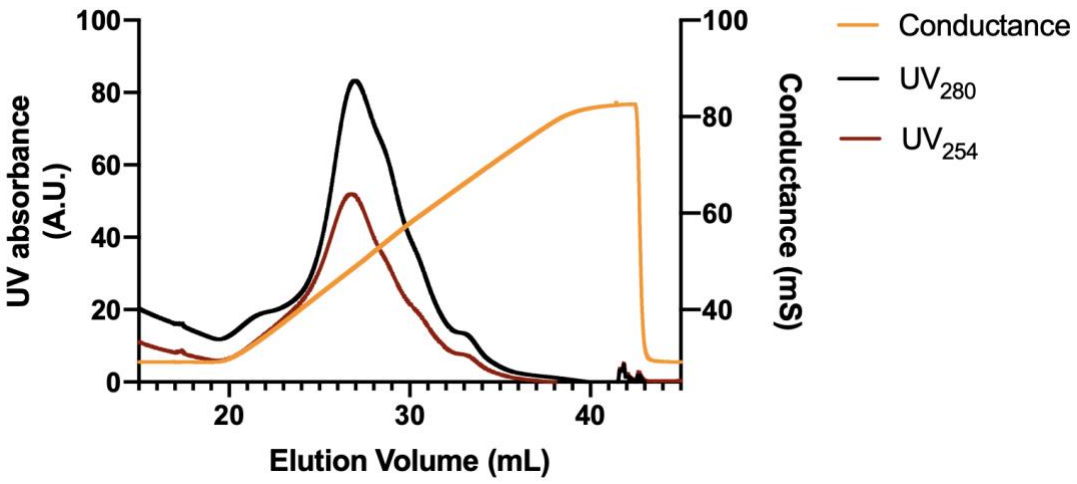

Figure 3.1 A heparin affinity chromatography profile of wild-type AsCas12f

To elucidate the mechanism of how AsCas12f recognizes and cleaves dsDNA guided by the sgRNA, I assembled the AsCas12f-sgRNA-DNA complex by mixing the inactive AsCas12f D225A mutant, the sgRNA (193 nt), and its target dsDNA (42 bp), followed by size-exclusion chromatography (SEC) (Fig 3.2). The inactive AsCas12f mutant ensured the protein would not cleave the DNA during complex assembly and formed a stable complex with sgRNA and its target DNA. The fractions containing the AsCas12f-sgRNA-DNA complex were combined and concentrated for further structural determination.

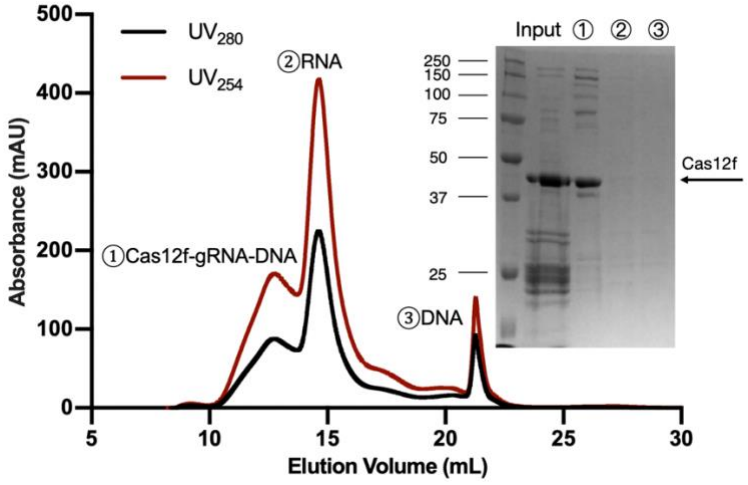

Figure 3.2 Size exclusion chromatography of the AsCas12f-sgRNA-DNA complex

3.2.2 Structural determination of the AsCas12f-sgRNA-DNA complex

The purified complex was analyzed using single-particle cryo-EM. The final reconstructed map had an overall resolution of 2.9 Å (Figure 3.3, Table 3.1). The structure revealed that the AsCas12f-sgRNA-DNA complex is composed of two AsCas12f molecules (referred to as AsCas12f.1 and AsCas12f.2), one sgRNA molecule, and its target dsDNA. An AsCas12f molecule consists of an N-lobe and a C-lobe. The N-lobe contains a WED (wedge) domain and a REC (recognition) domain, linked with the C-lobe comprising a RuvC nuclease domain and a ZF (zinc finger) motif (Figure 3.3).

The two AsCas12f molecules play different roles in the complex, and their domain structures are also slightly different. The REC.1 domain comprises four α-helices, while only three α-helices can be resolved in REC.2 domain. Both of the REC domains are involved in recognition of nucleotides. Similar to other Cas12f proteins, both of the WED domains in the complex adopt an oligonucleotide/oligosaccharide-binding fold [76], which consists of a seven-stranded β-barrel flanked by an α-helix. However, they are located at different positions in the complex and stabilize different regions of sgRNA. In terms of RuvC nuclease domains, only the density of the RuvC.1 domain is observed in the map, probably because the RuvC.2 domain is too flexible to be determined in the complex, or the dataset has an orientation-preference problem. The RuvC domain has an RNase H fold, comprising a five-stranded β-barrel flanked by four α-helices. The CCCC-type zinc finger is inserted into the α-helices of the RuvC domain.

Table 3.1: Statistics of cryo-EM model refinement and geometry for the AsCas12f-sgRNA-DNA complex

Model	
Composition (#)	
Chains	6
Atoms	9030 (Hydrogens: 0)
Residues	Protein: 613 Nucleotide: 190
Ligands	ZN: 1

Table 3.1 continued: Statistics of cryo-EM model refinement and geometry for the AsCas12f-sgRNA-DNA complex

Bonds (RMSD)	
Length (Å) (# > 4σ)	0.003 (0)
Angles (°) (# > 4σ)	0.581 (2)
MolProbity score	1.65
Clash score	13.70
Ramachandran plot (%)	
Outliers	0.00
Allowed	1.49
Favored	98.51
Rama-Z (Ramachandran plot Z-score, RMSD)	
whole (N = 605)	1.07 (0.33)
helix (N = 298)	1.79 (0.29)
sheet (N = 130)	0.20 (0.43)
loop (N = 177)	-0.69 (0.43)
Rotamer outliers (%)	0.00
Cβ outliers (%)	0.00
Peptide plane (%)	
Cis proline/general	0.0/0.0
Twisted proline/general	0.0/0.0
CaBLAM outliers (%)	1.01
ADP (B-factors)	
Iso/Aniso (#)	9030/0
min/max/mean	
Protein	3.03/43.74/18.98
Nucleotide	8.77/109.71/46.30
Ligand	57.05/57.05/57.05
Data	
Box	
Lengths (Å)	85.20, 116.09, 137.39
Angles (°)	90.00, 90.00, 90.00
Supplied Resolution (Å)	2.9
Resolution Estimates (Å)	Masked
d FSC (half maps; 0.143)	2.8
d 99 (full/half1/half2)	2.7/2.4/2.4
d model	2.8
d FSC model (0/0.143/0.5)	2.4/2.6/3.1
Map min/max/mean	-0.16/0.28/0.00
Model vs. Data	
CC (mask)	0.71
CC (box)	0.64
CC (peaks)	0.59
CC (volume)	0.70
Mean CC for ligands	0.66

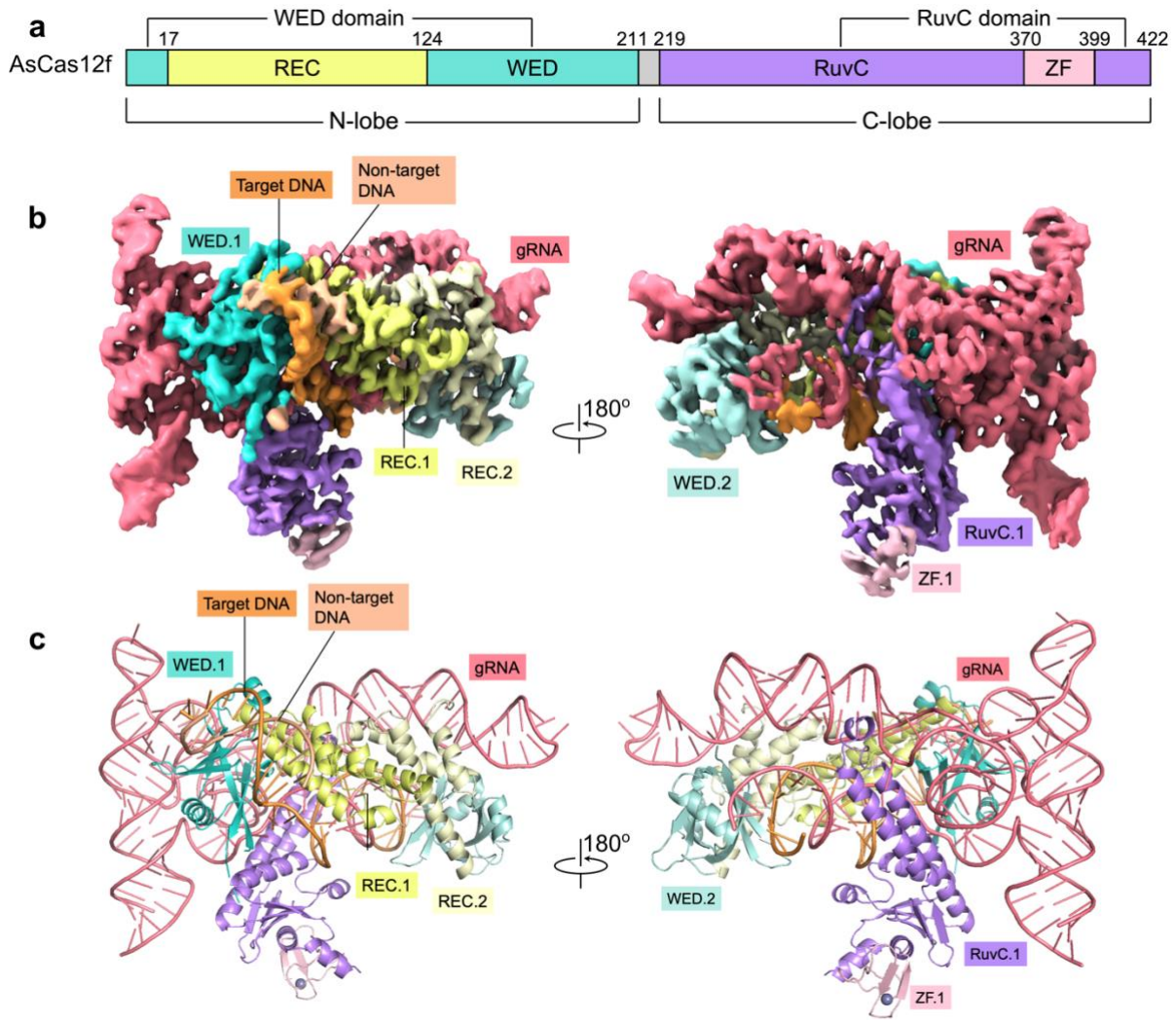

Figure 3.3 The structure of the AsCas12f-sgRNA-DNA complex
a. Domain diagram of AsCas12f.
b. Unsharpened cryo-EM map of the AsCas12f-sgRNA-DNA complex (contoured at a level of 0.020). AsCas12f.1 was colored in the same color as in a. The WED.2 and the REC.1 domains in AsCas12f.3 were colored in a light yellow and light cyan, respectively..
c. Overall structure of the AsCas12f-sgRNA-DNA complex. The color code of the nucleotides is the same as in b.

3.2.3 Dimer interface of the AsCas12f molecules in the AsCas12f-sgRNA-DNA complex

AsCas12f dimerizes through the REC domains of the two copies. The extensive interactions of four interfaces between the REC.1 and REC.2 domains include four salt bridges and thirteen hydrogen bonds (Table 3.2, Figure 3.4a).

On interface 1, hydrogen bonds are formed between R32.2 with I109.1 and L110.1, and R121.2 with D113.1 and G112.1. On interface 2, D51.1 forms a hydrogen bond and a salt bridge with K37.2. On interface 3, E44.1 and Y74.1 form hydrogen bonds with S50.2 and D51.2, respectively. On interface 4, R32.1 interacts with D113.2 through a salt bridge. K37.1 forms hydrogen bonds with I109.2 and L110.2, respectively. Also, hydrogen bonds are formed between T39.1 and S115.2, N36.1 and G112.2, S118.1 and M114.2, and Q40.1 with I109.2 and S115.2, respectively (Figure 3.4a, Table 3.2). Notably, alanine substitutions that interfere these interactions, such as E44A, D51A, and Y52A, led to minimal indel frequency (Figure 3.4b), suggesting the importance of AsCas12f dimerization for its DNA cleavage activity.

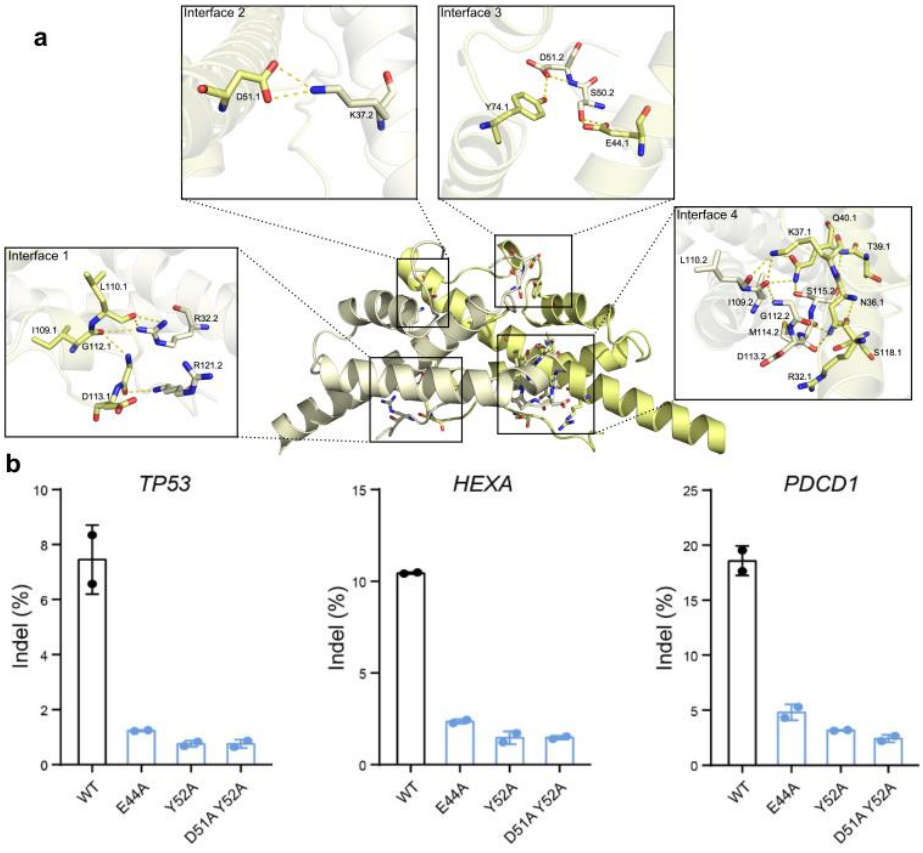

Figure 3.4 AsCas12f dimerization affects its activity.
a, Four dimer interfaces between AsCas12f.1 and AsCas12f.2.
b. Indel frequency generated by the wild-type and mutant AsCas12f variants, with the mutations disrupting the dimer interfaces.

Table 3.2: The polar contacts that maintain the dimer interface of the AsCas12f molecules in the AsCas12f-sgRNA-DNA complex

Interface	AsCas12f.1	AsCas12f.2	Bond type
Interface 1	I109.1	R32.2	Salt bridge
	L110.1	R32.2	Hydrogen bond
	D113.1	R121.2	Hydrogen bond and salt bridge
	G112.2	R121.2	Hydrogen bond
Interface 2	D51.1	K37.2	Hydrogen bond and salt bridge
Interface 3	E44.1	S50.2	Hydrogen bond
	Y74.1	D51.2	Hydrogen bond
Interface 4	R32.1	D113.2	Salt bridge
	K37.1	I109.2	Hydrogen bond
	K37.1	L110.2	Hydrogen bond
	T39.1	S115.2	Hydrogen bond
	N36.1	G112.2	Hydrogen bond
	Q40.1	I109.2	Hydrogen bond
	Q40.1	S115.2	Hydrogen bond
	S118.1	M114.2	Hydrogen bond

3.2.4 PAM and sgRNA recognition by AsCas12f

The TTG PAM sequence of the target dsDNA is recognized by AsCas12f, specifically through the interactions with the REC.1 domain and the WED.1 domain (Figure 3.5). The nucleobases of the non-target DNA strand dT(-3*)-dT(-2*) in the PAM form hydrogen bonds with K80.1 and S92.1 in the REC.1 domain. Besides, these two residues, along with Y68.1, T69.1, and Y76.1 in the REC.1 domain, form hydrogen bonds with the backbone phosphate group of the PAM consensus (Figure 3.5a). The nucleobases of the target DNA strand dC(21) and dA(22), which base pairs with dT(-2*)-dG(-1*), form hydrogen bonds with S92.1 and K96.1 in the REC.1 domain. Besides, K129.1 and A189.1 in the WED.1 domain also form hydrogen bonds with the backbone phosphate group (Figure 3.5b). These interactions guarantee the specificity of AsCas12f to the T-enriched PAMs. Apart from interactions on nucleobases, Y68.1, T69.1, and Y76.1 in the REC.1 domain as well as K129.1 and A189.1 in the WED.1 domain form hydrogen bonds with the backbone phosphate groups within the PAM regions (Figure 3.5).

The Y76A and S92A mutants show markedly reduced DNA cleavage activity (Figure 3.6). In addition to the PAM sequence, the other regions in the dsDNA also form extensive hydrogen bonds and salt bridges with the AsCas12f.1 molecule. Mutating the residues involved in such interactions, including R298 and Y343, into alanine also attenuate the DNA cleavage activity (Figure 3.6).

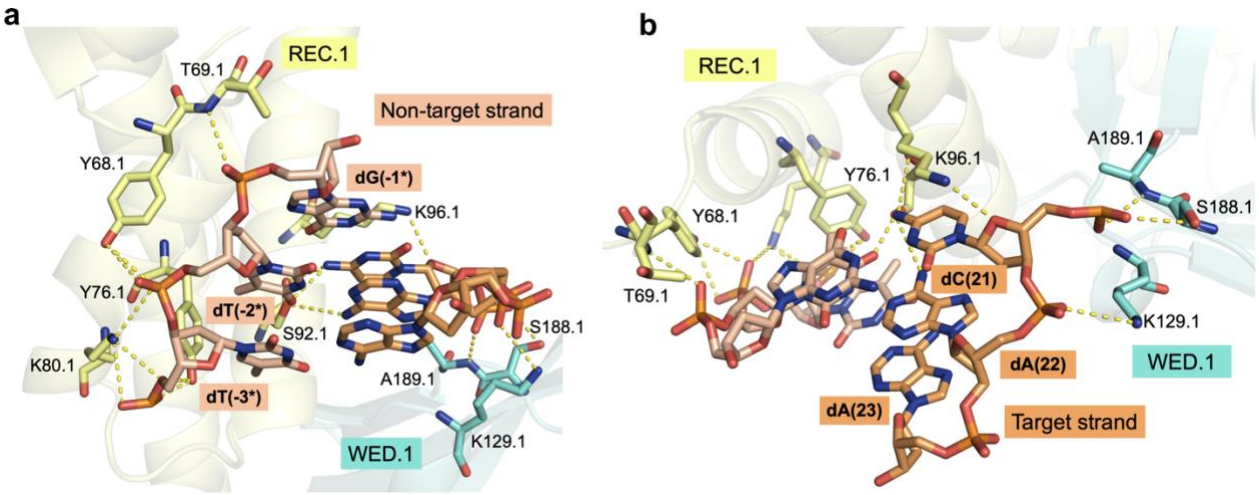

Figure 3.5 Recognition of the PAM motif
a. Recognition of the non-target strand of the PAM motif
b. Recognition of the target strand of the PAM motif

The sgRNA consists of the 49 nt crRNA and a 138 nt tracrRNA, comprising five stem-loop structures (Figure 3.7a). Our structural and functional findings suggested that both AsCas12f molecules in the complex are important to sgRNA recognition. Stem 1 is recognized by RuvC.1 and WED.1 through hydrogen bonds. Stem 2 interacts with almost all the domains in the two AsCas12f copies, and it can be specifically recognized by REC.1, WED.1, RuvC.1, WED.2, and REC.2 domains by hydrogen bonds and salt bridges. Stem 3 does not directly interact with the AsCas12f, but structurally, stem 3 is vital to stabilize the folding and determine the orientation of the gRNA. Stem 4 can be recognized by the WED.1. In stem 5, the density of U(-47)-U(-15) is missing in the cryo-EM map, and the rest of the stem can interact with WED.1,

RuvC.1, and WED.2. Alanine substitutions of residues that interact with the gRNA, such as W17, H72, Y76, R121, and Y351, decrease DNA cleavage activity (Figure 3.6, Figure 3.7b-f). This is because not only do these five residues form hydrogen bonds, salt bridges, and π - π interactions with the sgRNA, but they also play critical roles in dimerization (R121.2, Figure 3.4a and Table 3.1) and PAM recognition (Y76.1, Figure 3.5a)

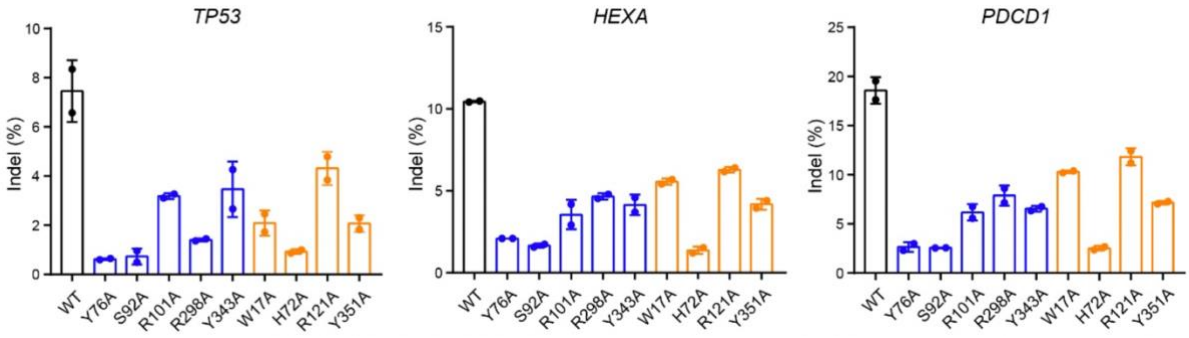

Figure 3.6 Indel frequency on *TP53, HEXA*, and *PDCD1* loci generated by the wild-type and mutant AsCas12f variants, with the mutations disrupting interactions between the protein and DNA (blue) or sgRNA (yellow)

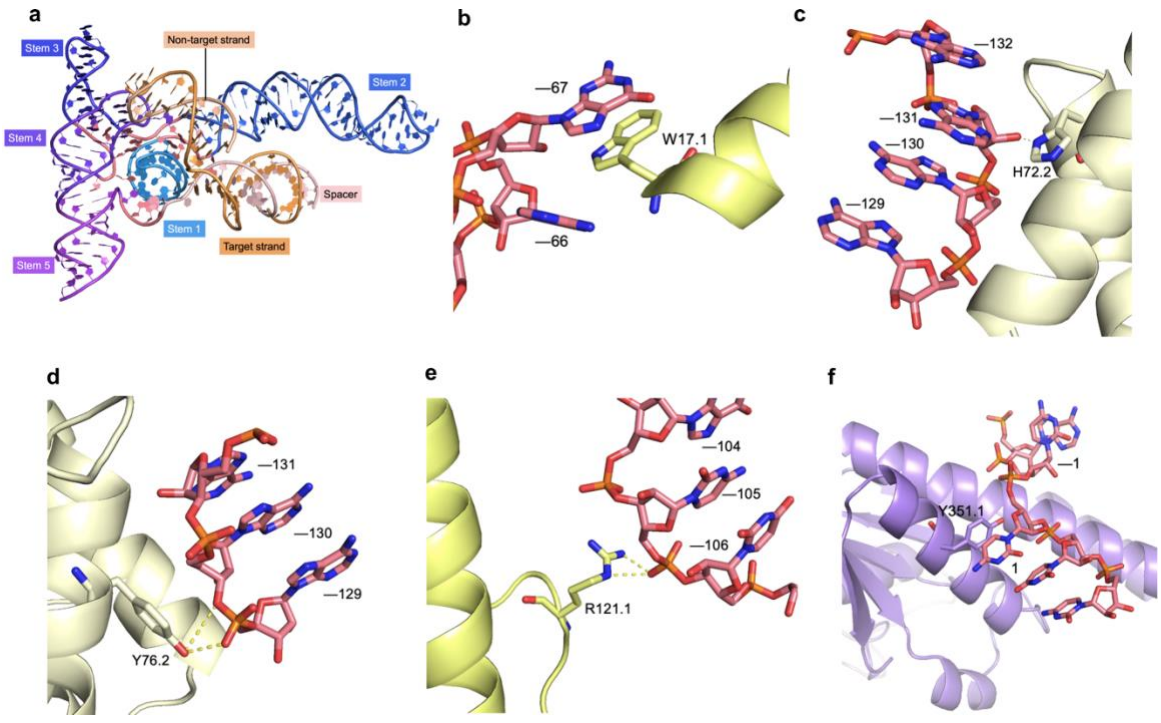

Figure 3.7 Important interactions between sgRNA and AsCas12f

(Figure 3.7 continued) a. Architecture of the sgRNA in complex with the dsDNA. The stems are colored in different colors.
b. W17.1 forms a π - π interaction with the G(-67) on the sgRNA
c. H72.2 forms a hydrogen bond with the A(-131) at O2 position on the sgRNA.
d. Y76.2 forms a hydrogen bond with phosphate backbone of the A(-129) on the sgRNA.
e. R121.1 forms a hydrogen bond with the phosphate backbone of the C(-105) on the sgRNA
f. Y351.1 forms a π - π interaction with the C(1) and on the sgRNA.

3.2.5 AsCas12f protein engineering

Based on previous research, the cleavage activity of Cas proteins can be increased by raising the affinity of proteins to their sgRNA and target DNA [77]. The affinity between nucleic acids and proteins is frequently mediated by salt bridges and hydrogen bonds between the phosphate backbones and positively charged residues on the proteins. So, our strategy for engineering the AsCas12f protein is to introduce some positively charged residues in order to raise their affinity to nucleic acids, boosting the DNA targeting and cleavage activity.

To determine which residues should be mutated into positively charged residues, our collaborator, Dr. Tong Wu, decided to learn from naturally occurring homologous proteins, which are depleted of deleterious mutations due to millions of years of natural evolution, as demonstrated by DNA family shuffling [78]. He searched for the homologies of AsCas12f using the basic local alignment search tool (BLAST) [79] and analyzed the evolutional relationships between Cas12f family proteins. Proteins (including hypothetical proteins) that are closely homologous to AsCas12f were selected for alignment. Positions that align well but host varied amino acids in homologous proteins highlight the “functional diversity” of the protein family and may be particularly amenable to mutation (Figure 3.8a). Then, he constituted a library of AsCas12f variants with one or more mutations to positively charged residues that possibly benefit the targeting and cleavage activity of AsCas12f (Table 3.3). He found that the combination of additional beneficial mutations led to greater levels of indels, with the best

variant (3D: D196K/N199K/G276R/N328G/D364R) exhibiting 2.5- to 3.5-fold gene editing activity on all three tested target sites, resulting in up to 70% indel frequency (Figure 3.8b,c). Notably, three mutated residues in enAsCas12f variants, namely D196, N199, and G276, are physically close to the sgRNA, while D196 and G276 do not interact with the sgRNA, and N199 forms a weak hydrogen bond (3.76 Å) with the N^2-position of the G(-67) (Figure 3.8d,e). This structural evidence supports our hypothesis that substituting these residues with positively charged residues can boost DNA cleavage activity by increasing the binding affinity between AsCas12f and nucleic acids.

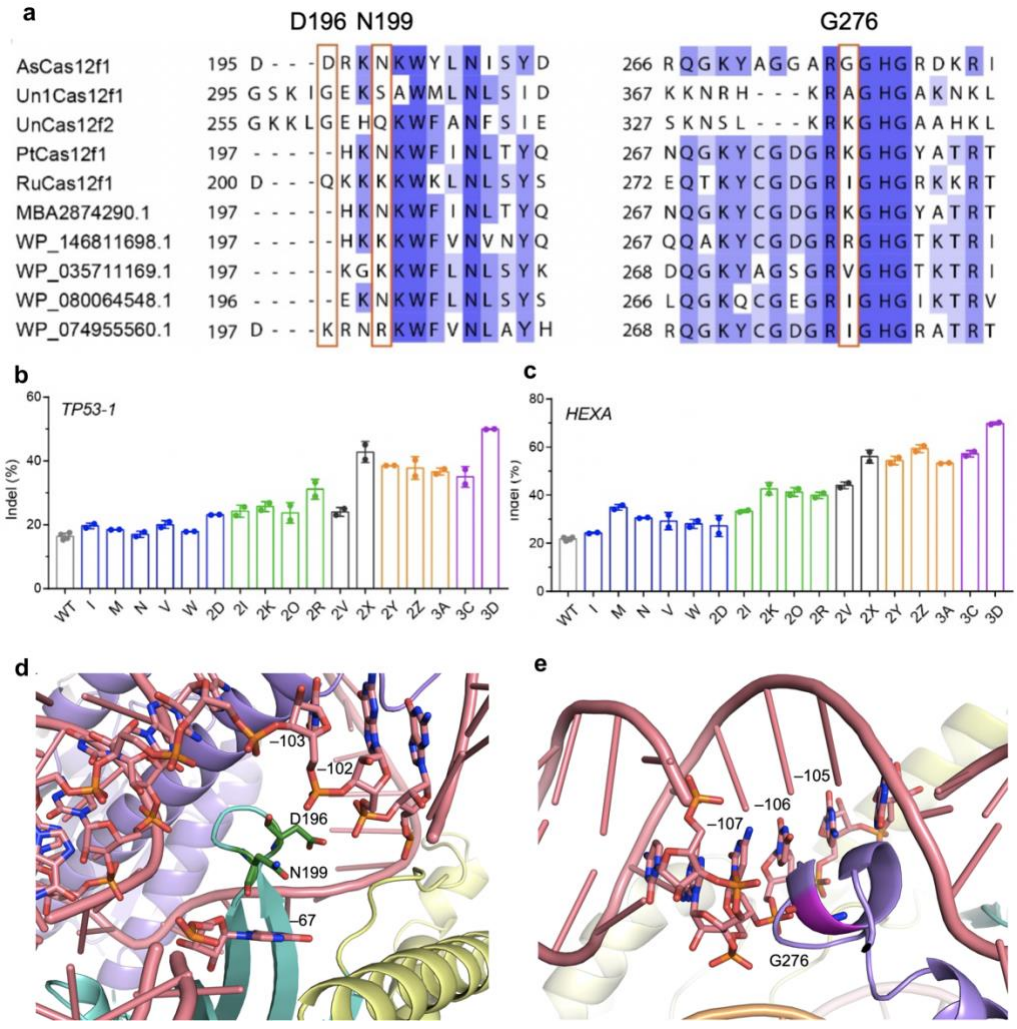

Figure 3.8 Structure-based AsCas12f engineering

(Figure 3.8 continued) a. BLAST alignment [79] of AsCas12f and homologous proteins at the position D196, N199, and G276
b & c. Indel level on b. *TP53-1* and c. *HEXA* loci generated by AsCas12 variants that bear one, two, three, four, or five single-point mutations. A list of mutations included in each AsCas12f variant is included in Table 3.3.
d. D196 and N199 are very close to the phosphate backbone of sgRNA, which D196 does not form any electrostatic interactions with sgRNA and N199 forms a weak hydrogen bond (3.76 Å) with the N^2-position of the G(-67)
e. G276 is also very close to, but does not interact electrostatically with the phosphate backbone of sgRNA.

Table 3.3: enAsCas12fs and their mutations

Construct name	Mutations
I	G276R
M	D364K
N	D364R
V	D196K
W	N199K
2D	N328G
2I	G276R D196K
2K	D364K D196K
2O	D364R D196K
2R	D196K N328G
2V	G276R D196K D364K
2X	D196K N199K N328G
2Y	D196K N199K N328G D364R
2Z	D196K N199K G276R D364K
3A	D196K N199K G276R D364R
3C	D196K N199K G276R N328G D364K
3D	D196K N199K G276R N328G D364R

3.2.6 sgRNA engineering

sgRNA engineering has been demonstrated as another beneficial strategy to improve the efficiency of the CRISPR gene-editing system besides engineering Cas proteins. Based on a high-resolution cryo-EM structure, the RNA regions that are critical to the AsCas12f-sgRNA-DNA complex formation are identified. Our strategy of engineering sgRNA was to truncate the regions that are not directly involved in the complex formation because higher transfection efficiency due to the shortened sgRNA can improve the efficiency of the CRISPR system. The

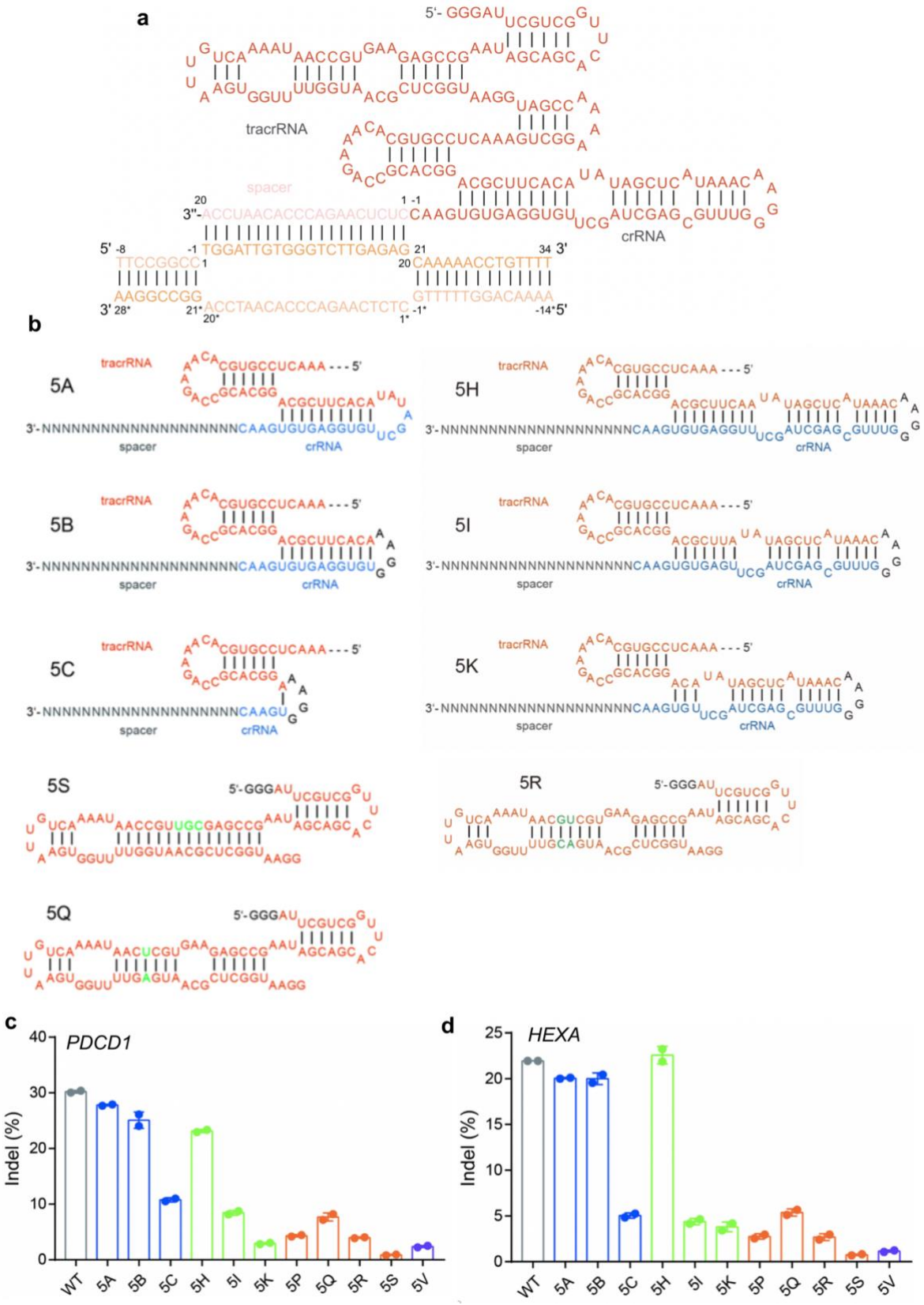

Figure 3.9 Structure-based sgRNA engineering

(Figure 3.9 continued) a. Sequence of sgRNA and target DNA of AsCas12f system
b. Sequences of truncated sgRNA of AsCas12f system
c. & d. Indel level on c. *PDCD1* and d. *HEXA* loci guided by wild-type and truncated sgRNA

cryo-EM map reveals that the density of U(-47)-U(-15) on stem 5 of RNA cannot be resolved and no protein density is observed adjacent to that region. So, I hypothesized that U(-47)-U(-15) does not directly interact with AsCas12f and its truncation would maintain the targeting and cleavage activity of the AsCas12f system. The truncated sgRNAs mediates similar indel formation efficiency (Figure 3.9). Meanwhile, truncations ≥ 3 bp on the other half of stem 5 (A(-57)/U(-5) to A(-48)/U(-14)) that interact with AsCas12f abolishes the activity (Figure 3.9).

3.2.7 Comparison between AsCas12f-sgRNA-DNA and UnCas12f-gsRNA-DNA

A structure of Cas12f from an uncultured archaeon in complex with its sgRNA and target DNA (referred as to UnCas12f-sgRNA-DNA) reported before [76]. The UnCas12f-sgRNA-DNA complex also composed of two UnCas12f molecules, a sgRNA, and its target dsDNA.

The overall structure seems very different due to the differences in sgRNAs (Figure 3.10a), but the folding of AsCas12f.1 and UnCas12f.1 in the complexes are very similar, as suggested by a superimposition of the two complexes (Figure 3.10b). However, the N-lobe cannot be superimposed as well as the C-lobe due to an additional zinc finger motif (78 amino acids) at the N-terminus of the REC domain in the UnCas12f compared to the AsCas12f (Figure 3.10c,d). Another difference between the two complexes is that the density of the Ruv.2 domain is missing from the AsCas12f-sgRNA-DNA complex cryo-EM map, probably due to its flexibility or preferential orientation of the cryo-EM dataset (Figure 3.3b,c, 3.10b). Besides, the dimerization of the UnCas12f and the AsCas12f is also different. UnCas12f dimerizes through the two interfaces of the REC domains and RuvC domains of the two protein copies, while

AsCas12f dimerizes through four interfaces between the REC domains of the two protein copies (Table 3.2, Figure 3.4a).

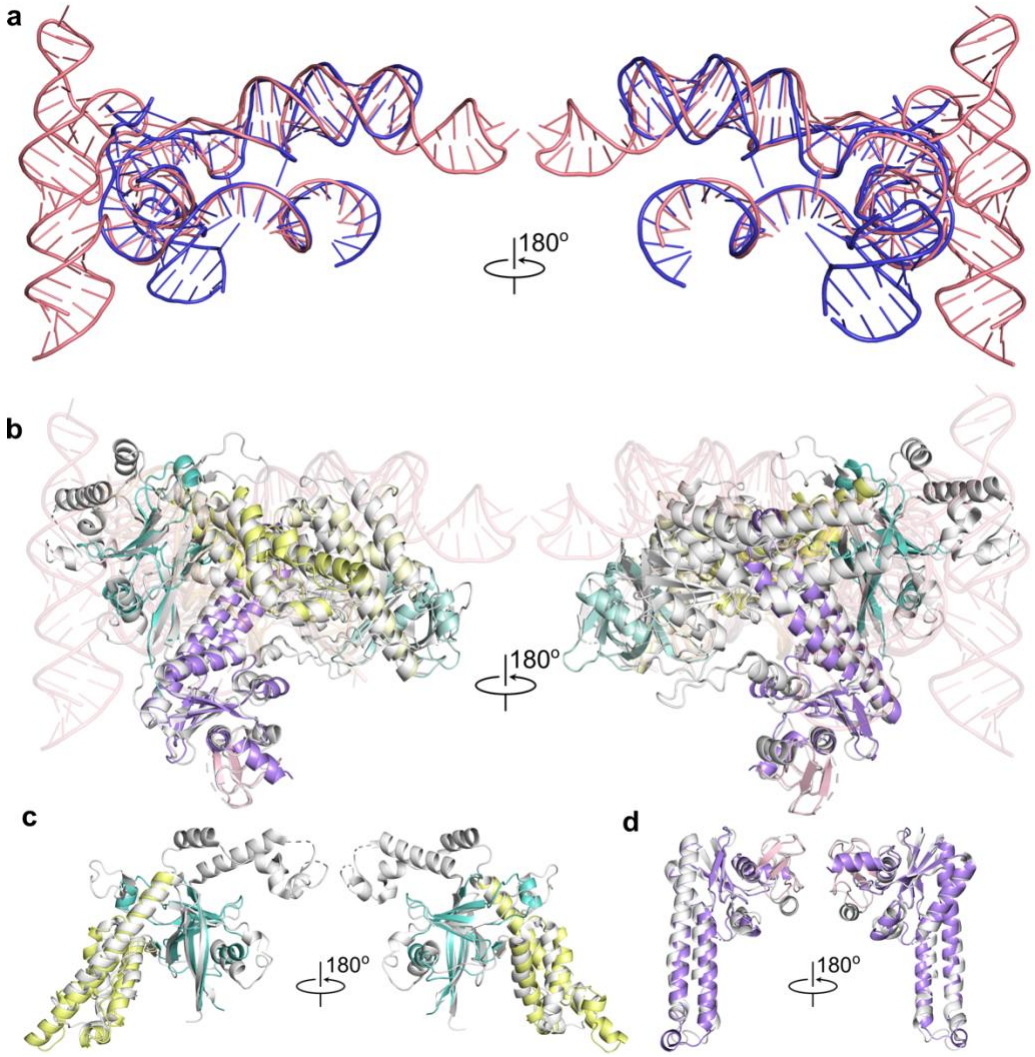

Figure 3.10 Comparison of the complex structures of AsCas12f-sgRNA-DNA and UnCas12f-sgRNA-DNA.
a. Superimposition between sgRNA of AsCas12f system (salmon pink) and sgRNA of UnCas12f system (blue)
b. (Left) The 2 AsCas12f molecules in the AsCas12f-sgRNA-DNA complex are superimposed with the 2 UnCas12f molecules (white) in the UnCas12f-sgRNA-DNA complex. (Right) The superimposed structures have been tilted 180º to see the additional zinc finger motif at the N-terminus of the UnCas12f REC.1 domain. The color code of the AsCas12f domains is the same as in Figure 1**.** RMSD = 6.398 Å (all-atom)
c. The N-lobe of the AsCas12f.1 is superimposed with that of the UnCas12f.1 (white). RMSD = 1.284 Å (all-atom)
d. The C-lobe of the AsCas12f.1 is superimposed with that of the UnCas12f.1 (white). RMSD = 2.961 Å (all-atom)

3.3 Discussion and conclusion

In this study, I determined the structure of the AsCas12f-sgRNA-DNA ternary complex and guided the engineering of AsCas12f and sgRNA. The structure reveals that two AsCas12f molecules form a dimer to recognize the sgRNA and target DNA. We identified important interactions between the AsCas12f molecules, and the mutagenesis studies suggested that AsCas12f dimerization is critical to the cleavage activity of the CRISPR-AsCas12f system. Besides, the important residues to recognize the sgRNA and the specific PAM sequence were also identified by the structural and mutation studies.

Our next step is to improve the CRISPR system by further truncating the sgRNA to improve the transfection efficiency. Our collaborator made a bold hypothesis that they could further engineer the sgRNA of AsCas12f by learning from the sgRNA structure in the UnCas12f-sgRNA-DNA complex, because the AsCas12f and UnCas12f proteins share similar folding patterns (Figure 3.4b). Superimposition of the sgRNAs showed that AsCas12f sgRNA has longer stems than UnCas12f sgRNA (Figure 3.4a) and some regions seem not important to the complex formation. Considering that stem 3 and stem 4 do not form strong interactions with the AsCas12f molecules in the complex, and truncating half of stem 5 does not impair activity, our collaborator further removed stem 3 and stem 4, and modified stem 5 to mimic the folding of the UnCas12f sgRNA. This new sgRNA is 72 nt shorter than the original one, representing a more than 1/3 molecular weight decrease. We are still characterizing the cleavage activity of the new sgRNA.

3.4 Methods

3.4.1 Materials

The holey carbon grids used in this study are Quantifoil 300 mesh Cu 1.2/1.3. **D**NA oligos were purchased from Integrated DNA Technologies, Inc. (IDT)**.** All chemicals used in buffer preparation were purchased from Sigma-Aldrich. RNA oligos were obtained from our collaborator.

3.4.2 Cloning and plasmid construction

The plasmid containing Cas12f gene from *Acidibacillus sulfuroxidans* was purchased from Addgene. The DNA sequence was cloned into the vector pET47b containing an N-terminal His_6-tag. The D225A mutant of AsCas12f was generated by site-directed mutagenesis.

3.4.3 Protein expression and purification

The wild-type His-tagged AsCas12f or its D225A mutant was overexpressed in *E.coli* BL21(DE3). The *E. coli* cells were cultured at 37 ^{o}C until the OD_{600} reached 0.8, and the protein expression was then induced by adding IPTG to a final concentration of 1 mM. The *E. coli* cells were further cultured at 16 ^{o}C for 24 hours before harvest by centrifugation. The cell pellet was resuspended in buffer A (50 mM Tris·HCl, pH 8.0, 300 mM NaCl, 20 mM imidazole, 0.5 mM TCEP, and 1 mM PMSF), and lysed by sonication. The lysate was cleared by centrifugation at 16,500 rpm. For 30 min. The supernatant was loaded onto a gravity column containing 2-mL Ni-NTA beads equilibrated with buffer B (50 mM Tris·HCl, pH 8.0, 300 mM NaCl, 20 mM imidazole, and 0.5 mM TCEP) and the proteins were eluted by buffer B (50 mM Tris·HCl, pH 8.0, 300 mM NaCl, 400 mM imidazole, and 0.5 mM TCEP). The proteins were further loaded

onto a 1 mL HiTrap Heparin HP column (GE Healthcare), equilibrated with buffer C (50 mM Tris·HCl, pH 8.0, 300 mM NaCl, and 0.5 mM TCEP) and eluted with a linear gradient of NaCl up to 1 M. The purified proteins were flash-frozen and stored at -80 ^{o}C until use.

3.4.4 AsCas12f-sgRNA-DNA complex assembly

The AsCas12f-sgRNA-DNA complex was reconstituted by mixing purified AsCas12f inactive D225A mutant, the 193-nucleotide sgRNA (193 nucleotides including 5' GG for in vitro transcription), the 42-nucleotide target DNA strand, and the 42-nucleotide non-target DNA strand, at a molar ratio of 1:0.5:1.2:1.2 on ice for 30 min. The AsCas12f-sgRNA-DNA complex was further purified using a Superdex 200 Increase 10/300 column (GE Healthcare), equilibrated with buffer D (50 mM Tris·HCl, pH 8.0, 50 mM NaCl, 5 mM $MgCl_2$, and 0.5 mM TCEP). The fractions were pooled and concentrated to ~3 mg/mL.

3.4.5 Electron microscopy sample preparation

Sample vitrification was performed using a Vitrobot Mark IV (Thermo Fisher) operating at 8 ^{o}C and 100% humidity. A total of 3.5 μL of the sample was applied to holey carbon grids (Quantifoil 200 mesh Cu 1.2/1.3) glow-discharged for 30 seconds. The grids were blotted for 4 seconds at a blotting force 0 by standard Vitrobot filter paper (Ted Pella, 47000-100), and then plunge frozen in liquid ethane.

3.4.6 Data collection

Frozen grids were sent to the Advanced Electron Microscopy Facility at the University of Chicago for data collection. The dataset was acquired as movie stacks using a Titan Krios

transmission electron microscope operating at 300 kV and equipped with a K3 direct detector camera (Gatan).sImages were recorded at a nominal magnification of 81,000x and super-resolution counting mode by image shift. The total exposure time was set to 4 s with 40 frames in a single stack and a total exposure around 50 electrons/Å^2. The defocus range was set at -1.0 to -2.5 μm (Table 3.4).

Table 3.4: Statistics of cryo-EM data collection and processing for AsCas12f-sgRNA-DNA complex

	AsCas12f-sgRNA-DNA complex
Microscope	Krios (University of Chicago)
Magnification	81,000
Voltage (kV)	300
Spherical aberration (mm)	2.7
Detector	K3
Camera mode	Super resolution counting
Exposure rate (e^-/pixel/s)	15
Total exposure (e^-/Å^2)	50
Defocus range (μm)	-1.0 to -2.5
Pixel size (Å)	0.5325 (1.065 physical)
Mode of data collection	Image shift
Energy filter	20 eV slit
Software for data collection	EPU
Number of micrographs	7,591
Symmetry imposed	C1
Box size (pixel)	256
Initial particle images (no.)	5,285,777
Particle images for 3D (no.)	1,497,300
Final particle images (no.)	490,190
Map resolution, unmasked (Å)	3.0
Map resolution, masked (Å)	2.8

3.4.7 Image processing

Stack images were subjected to motion correction by MotionCor2 [80]. Motion-corrected micrographs were then imported to a cryoSPARC live session for CTF determination and particle picking [33]. Particles were automatic picked by using 2D class averages as templates which were generated from blob picking. The extracted particles were imported to cryoSPARC

for further processing. Contamination and poorly-aligned classes were disposed after 2D classification. The resulting 3,370,441 particles were used to generate three initial models by ab initio reconstruction. A 3D classification was performed in cryoSPARC using the three initial models as starting points. The coordinates of the particles from the best class (1,576,757 particles) were imported to RELION and re-extracted. Another 3D classification was performed using the map from cryoSPARC as the initial model. The best class was subjected to 3D refinement, CTF refinement, Bayesian polishing, and postprocessing. The final map of the AsCas12f-sgRNA-DNA complex was resolved at 2.8 Å based on the criteria of FSC=0.143. The detailed data processing workflow is shown in Figure 3.11.

3.4.8 Model Building, Refinement, and Validation

Model building was perform in COOT [56] using a starting model of AsCas12f predicted by AlphaFold2 [51]. One full copy of AsCas12f and another copy of the N-lobe were identified from the cryo-EM map. Nucleic acids (DNA and sgRNA) were built into the map based on the knowledge of sequence complementarity, secondary structure prediction (Ipknot) [81], and fragment RNA model generated by RNAComposer [82]. The final model was refined in real space, and validated using Phenix [57]. The statistics of model refinement and geometry is shown in Table 3.1.

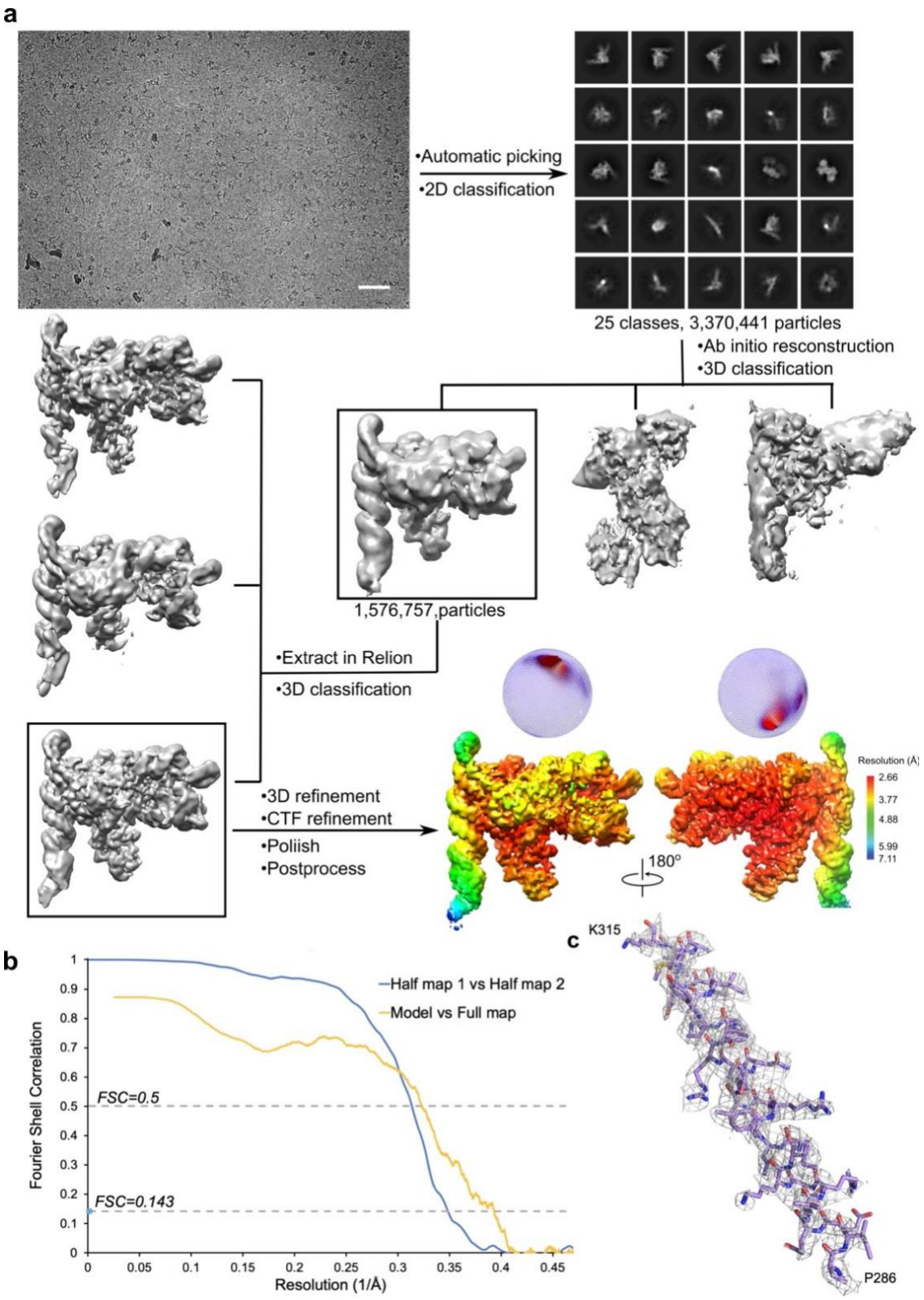

Figure 3.11 Single-particle cryo-EM analysis for the AsCas12f-sgRNA-DNA complex

(Figure 3.11 continued) a. The workflow of data processing. The data was acquired at the Advanced Electron Microscopy Facility at the University of Chicago, using a Titan Krios electron microscope operating at 300 KV. Initial motion correction was carried out using MotionCor2 [80]. A representative micrograph is shown along with a 50-nm scale bar. The following automatic picking and 2D classification were performed in the cryoSPARC live session [33]. Particle selection was performed using the 2D templates generated by blob picker, followed by a 2D classification. Representative 2D class averages are shown, with the box edge corresponding to 273 Å. After disposing contamination and poorly-aligned 2D classes, the particles of the aligned classes were extracted and imported to cryoSPARC for further processing. The particles were classified by 3D classification with 3 initial models generated using ab initio reconstructions. The best class of particles were re-extracted to Relion [34], followed by another 3D classification. Subsequent refinement and polishing were performed on the best 3D class.
b. Fourier shell correlation curves of the half map 1 versus the half map 2 (blue), and the refined model versus the full map (yellow). The resolution of the reconstruction was determined by the FSC = 0.143 criterion.
c. An α-helix of AsCas12f.1 molecule (contoured at a level of 5.0 RMSD)

Chapter 4 Structural basis for lipid-mediated stimulation of MsbA

4.1 Introduction

ATP-binding-cassette (ABC) transporters are critical to transporting small molecules, such as lipids, ions, vitamins, and antibiotics [83], [84]. MsbA, one of the best-studied ABC transporters, is an essential superfamily member in Gram-negative bacteria [83], [85]–[87]. In most Gram-negative bacteria, lipopolysaccharide (LPS), an important bacterial surface molecule, helps bacteria resist antibiotics and environmental stresses and is responsible for an immune response to bacterial infection in animals [88]–[91]. MsbA is essential to export the LPS from the cytoplasmic side, where LPS is synthesized, to the periplasmic leaflet of the inner membrane [91].

Previous studies have demonstrated that MsbA, a 65kD membrane protein, forms a homodimer with two cytosolic nucleotide-binding domains (NBD) and two transmembrane domains (TMD), each TMD consisting of six transmembrane helices. MsbA has two conformations, inward-facing (IF) and outward-facing (OF), and some small molecules have been developed that bind and lock MsbA in an LPS-bound, inward-facing (IF) conformation [86], [87].

The proposed mechanism of the translocation of LPS by MsbA involves several steps [86], [87], [90] (Figure 4.1): 1) the ADP-bound MsbA forms an IF conformation with the NBDs promoting access to LPS; 2) LPS binds to the inner cavity of MsbA and adenosine ATP, in order to promote dimerization of the NDBs; 3) ATP hydrolysis induces the change of MsbA from IF conformation to OF conformation, transporting LPS to the other side of the membrane; 4) the phosphate from the ATP hydrolysis is released, and MsbA recycles back to an IF conformation.

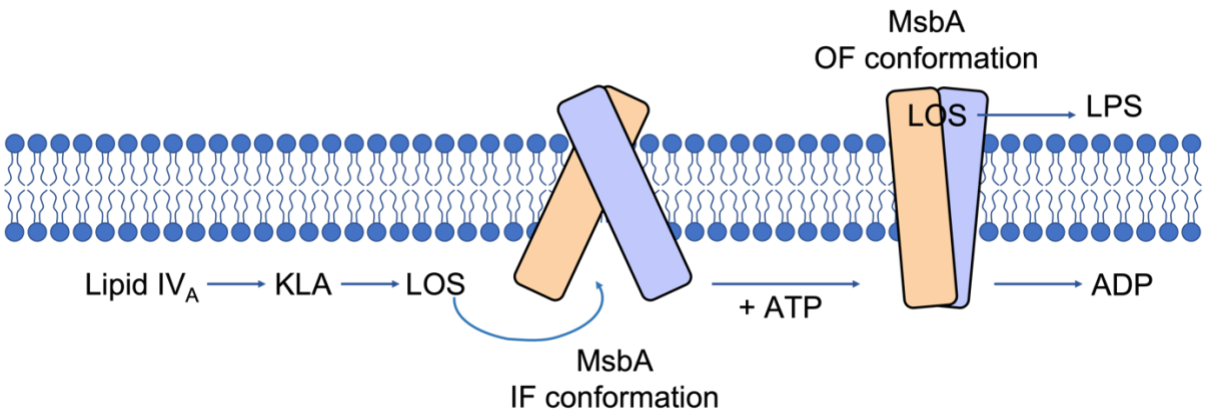

Figure 4.1 Biogenesis of LPS and the process of exporting LPS from the cytoplasmic side to the periplasmic leaflet of the inner membrane using MsbA
LOS: Lipooligosaccharide; KLA: 3-deoxy-D-*manno*-oct-2-ulosonic acid $(Kdo)_2$-lipid A

In this study, our collaborator characterized MsbA-lipid interactions in different conformational states. I determined the structures of the MsbA transporter with an open, IF confirmation and the vanadate-trapped MsbA transporter in complex with 3-deoxy-D-*manno*-oct-2-ulosonic acid $(Kdo)_2$-lipid A (KLA) and $ADP{\cdot}VO_4$, which have not been previously observed in other ABC structures.

4.2 Results

4.2.1 Discovery of copper(II)-bound MsbA and MsbA-lipid binding affinities facilitated by copper(II)-binding

Our collaborator first optimized the purification of MsbA from *E. coli* by screening different detergents, and they found that MsbA samples solubilized in the pentaethylene glycol monododecyl ether ($C_{10}E_5$) detergent exhibited a well-resolved mass spectrum (Figure 4.2a). Four molecular species are found in the mass spectrum, corresponding to dimeric MsbA and the addition of one to three ~65 Da adducts. They further analyzed the MsbA sample using inductively coupled plasma mass spectrometry (ICP-MS) [92] and identified the bound adducts

as copper(II) (Table 4.1). The addition of copper(II) to MsbA saturated the two binding sites (Figure 4.2c,d). The finding was also supported by that the addition of trientine, a copper chelator, to MsbA did not chelate the bound metal.

Table 4.1: Results from ICP-MS analysis of MsbA samples.

Sample	V[ng/mL]	Ni[ng/mL]	Cu[ng/mL]	Zn [ng/mL]
MsbA	38	10	1040	7.4
Vanadate-trapped MsbA	8338	16	1313	11

Our collaborator also determined equilibrium binding constants (K_D) for MsbA binding to different lipids, with 1-palmitoyl-2-oleoyl (PO, 16:0-18:1) as the acyl chain composition and phosphatidic acid (PA), phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylglycerol (PG), and phosphatidylserine (PS) as headgroup. They also included 1,1′,2,2′-tetraoleoyl-cardiolipin (TOCDL, 72:4) and KLA. Surprisingly, they found that copper(II)-bound MsbA increased binding affinities for nearly all the lipids, especially for POPA (Figure 4.2b,e).

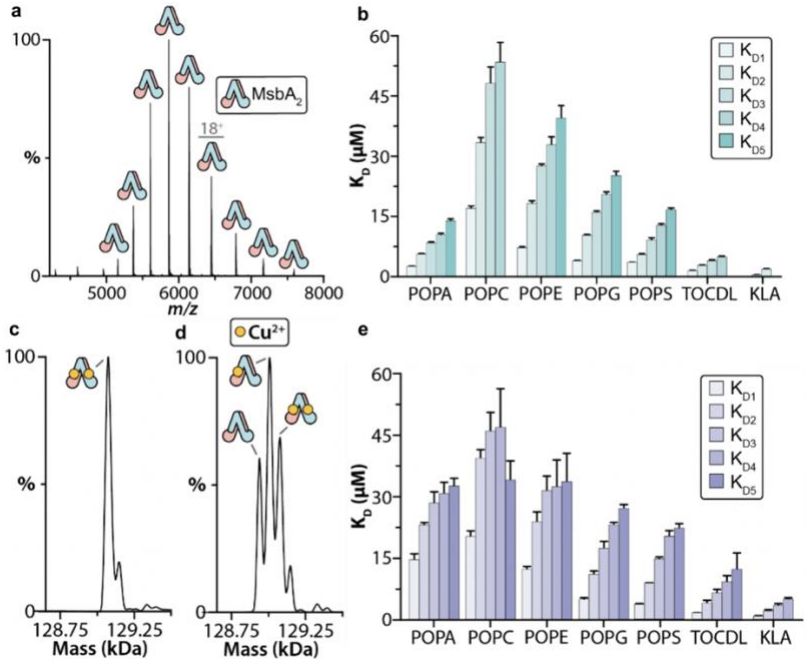

Figure 4.2 Copper(II) binding to MsbA modulates its lipid binding affinity.

(Figure 4.2 continue) a. Native mass spectrum of optimized MsbA samples in $C_{10}E_5$ yields a well-resolved mass spectrum.
b. Equilibrium dissociation constants (K_D) for individual lipid binding events to partially loaded MsbA.
c. Deconvolution of the mass spectrum shown in panel a. The different molecular species correspond to dimeric MsbA with zero to three bound copper ions.
d. Measured mass of MsbA after loading with copper(II) shows saturation of the two-copper(II) bound state.
e. K_{DS} for individual lipid binding events to MsbA fully loaded with copper(II). Reported are mean and standard deviation (n = 3).

4.2.2 Structural determination of the open-state MsbA

To understand how copper(II)-binding facilitates different binding affinities of MsbA, I determined the structure of the open-state MsbA transporter. The map was determined at 3.88 Å (Figure 4.3, Table 4.2). MsbA adopts an open, IF conformation, in which the NDBs are separated by ~57 Å from the Cα of R569 of one subunit to the other. However, the density was not clear enough to place the lipid and the copper(II) binding site.

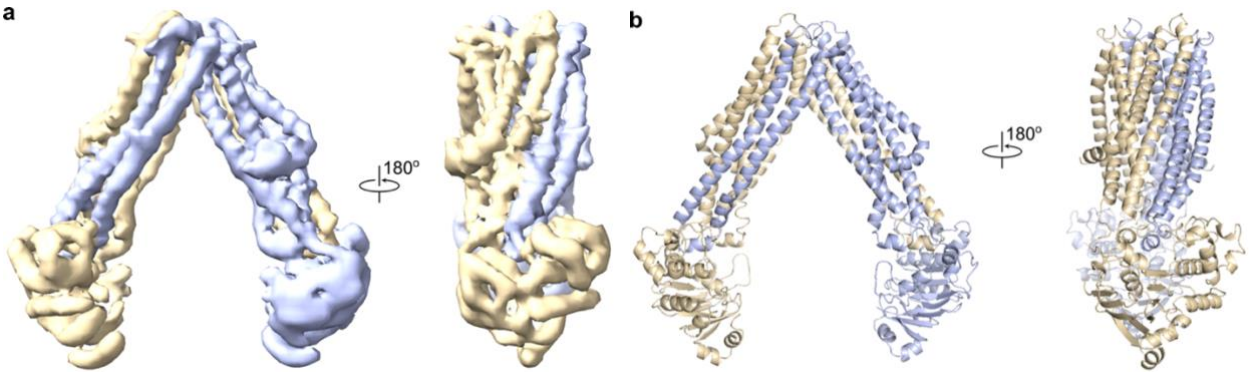

Figure 4.3 The structure of the MsbA transporter with an open, IF confirmation
a. Unsharpened cryo-EM map of the open-state MsbA transporter (contoured at a level of 0.16). Two subunits of MsbA are colored in wheat and purple blue, respectively.
b. Overall structure of the open-state MsbA transporter. The color code is the same as in a.

Table 4.2: Statistics of cryo-EM model refinement and geometry for the open-state MsbA

Model	
Composition (#)	
Chains	2
Atoms	8912 (Hydrogens: 0)
Residues	Protein: 1150 Nucleotide: 0
Water	0
Ligands	0
Bonds (RMSD)	
Length (Å) (# > 4σ)	0.003 (0)
Angles (°) (# > 4σ)	0.626 (1)
MolProbity score	1.66
Clash score	8.91
Ramachandran plot (%)	
Outliers	0.09
Allowed	2.97
Favored	96.95
Rama-Z (Ramachandran plot Z-score, RMSD)	
whole (N = 1146)	1.76 (0.25)
helix (N = 700)	2.03 (0.20)
sheet (N = 74)	1.25 (0.61)
loop (N = 372)	-0.50 (0.32)
Rotamer outliers (%)	0.00
Cβ outliers (%)	0.00
Peptide plane (%)	
Cis proline/general	0.0/0.0
Twisted proline/general	0.0/0.0
CaBLAM outliers (%)	1.58
ADP (B-factors)	
Iso/Aniso (#)	8912/0
min/max/mean	
Protein	0.17/104.70/43.85
Data	
Box	
Lengths (Å)	89.46, 136.32, 136.32
Angles (°)	90.00, 90.00, 90.00
Supplied Resolution (Å)	3.9
Resolution Estimates (Å)	Masked
d FSC (half maps; 0.143)	3.9
d 99 (full/half1/half2)	4.1/4.6/4.7
d model	4.0
d FSC model (0/0.143/0.5)	3.7/3.8/4.0
Map min/max/mean	-1.28/1.91/0.02
Model vs. Data	
CC (mask)	0.77
CC (box)	0.73

Table 4.2 continued: Statistics of cryo-EM model refinement and geometry for the open-state MsbA

CC (peaks)	0.66
CC (volume)	0.74

After inspecting histidine and cysteine residues in the MsbA structures and characterizing the copper(II) binding activity of mutated MsbA, our collaborator found that the copper(II) binding site is at the first four residues of the MsbA. So, they crystallized the N-terminal sequence of MsbA with variable tags in complex with copper(II) and obtained crystals of a green fluorescent protein (GFP) fused construct, leading to the structural determination of the construct at 2.15 Å resolution (Figure 4.4, Table 4.3). The copper(II) is bound to the H2 in the MsbA (Figure 4.4a). Besides, one of the water ligands of copper(II) builds a bridge between copper(II) and the side chain of N3. Interestingly, the structure of copper(II)-bound MSND peptide is similar to that of copper(II)-bound GHK peptide [93]. But the K3 in the GHK peptide does not interact with copper(II) directly or through water ligands (Figure 4.3b), indicating that the third position can be variable. To verify the hypothesis, our collaborator analyzed the sequences of ABC transporter sequences, and they found over 400 proteins contain histidine in the 2nd position, and some of them share a sequence of MHK at the N-terminus, demonstrating the copper(II) binding site may also relate to not only MsbA but also other ABC transporters.

Table 4.3: Summary of X-ray data collection and refinement statistics for MbsA-GFP

Crystal Parameters		
Space Group	$I2_12_12_1$	$I2_12_12_1$
Unit Cell Dimensions	*a*=70.61, *b*=108.54, *c*=140.4	*a*=70.62, *b*=108.96, *c*=140.64
Angles	α=β=γ=90°	α=β=γ=90°
Data Collection		
Synchrotron (Beamline)	APS (24-ID-C)	TAMU (R-Axis IV++)
λ (Å)	1.378	1.54
Resolution (Å)	42.98- 2.15 (2.21- 2.15)	70.32-2.2 (2.28-2.2)
I/σ	12.35 (1.61)	12.17 (2.14)

Table 4.3 continued: Summary of X-ray data collection and refinement statistics for MbsA-GFP

Observed Reflections	60359 (4433)	376860 (27999)
Unique Reflections	17601 (1280)	56321 (4109)
Completeness (%)	99.3 (98.0)	99.2 (98.6)
Refinement		
R_{work} (%)	0.19 (0.29)	0.19 (0.34)
R_{free} (%)	0.21 (0.30)	0.22 (0.34)
Number of Protein Atoms	1968	1955
Nonprotein Atoms	26	23
Number of water molecules	48	52
Number of protein residues	236	236
RMS (bonds)	0.008	0.007
RMS (angles)	0.91	0.88
Ramachandran favored (%)	96.98	97.41
Ramachandran allowed (%)	3.02	2.59
Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	0.96	2.43
Wilson B-factor	52.33	57.39
B factor of protein atoms	62.70	63.71

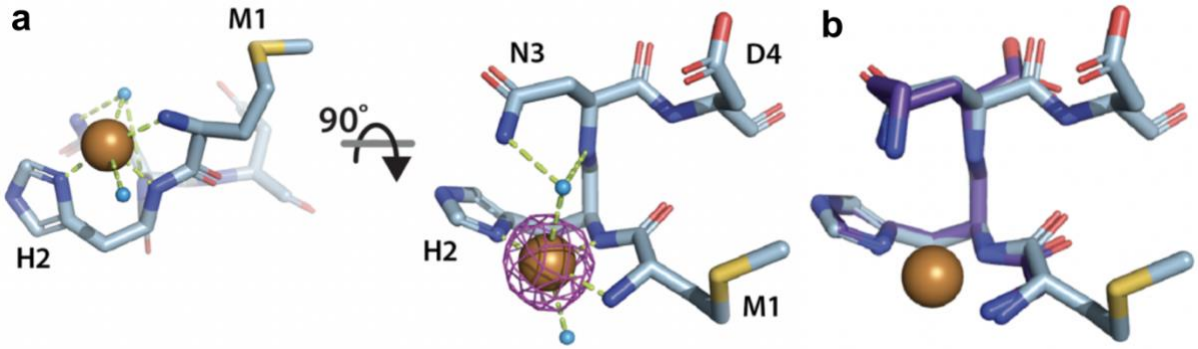

Figure 4.4 The N-terminus of MsbA binds copper(II) and its crystal structure.
a. Structure of the N-terminus of MsbA fused to GFP coordinating copper(II). The peptides are presented in sticks and balls, and water molecules and copper(II) are shown in spheres.
b. Alignment of the N-terminal MsbA peptide bound to copper(II) with GHK-copper(II) complex (CCDC-809108, Cα colored purple).

4.2.3 Lipid binding to vanadate-trapped MsbA

To determine the impact of MsbA conformation on lipid binding affinity, our collaborator performed analogous experiments using MsbA trapped in an OF conformation with

adenosine diphosphate (ADP) and orthovanadate (VO_4). Each subunit of the transporter binding copper(II), ADP, and VO_4 molecules was confirmed by native mass measurement (Figure 4.5a). Compared to the MsbA in the IF conformation, the MsbA in the OF conformation has higher lipid binding affinity for a subset of lipids (Figure 4.5b). For example, each subunit of vanadate-trapped MsbA transporter binds a KLA lipid (Figure 4.5c), while only one of the two subunits binds a KLA lipid (Figure 4.5d) in the presence of 0.4 μM of KLA. Also, the binding affinity for KLA (K_{D1} = 0.3 μM) of the vanadate-trapped MsbA increased by 2-fold compared to the non-trapped MsbA. Taken together, the native mass measurements and dissociation constants showed that MsbA in different conformational states binds lipids with different binding affinities.

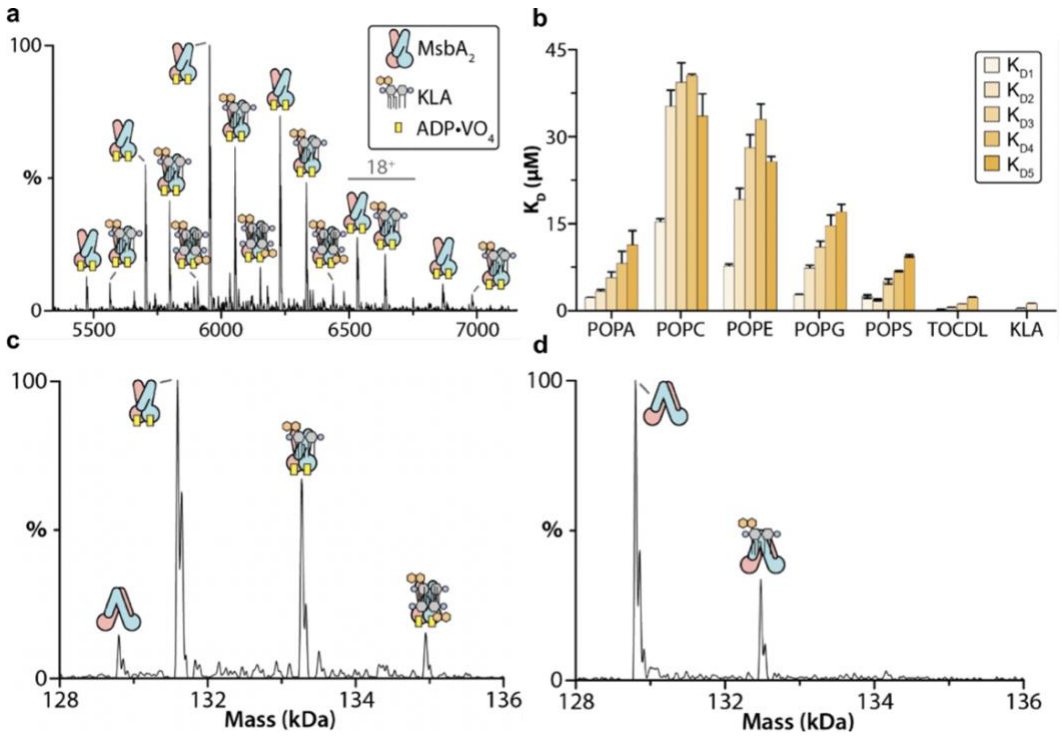

Figure 4.5 Lipid binding affinities of vanadate-trapped MsbA transporter in an OF conformation.
a. Representative MS of MsbA trapped with ADP·VO_4 and in the presence of 0.4 μM KLA.
b. K_D values for individual lipid binding events to vanadate-trapped MsbA. Reported are the mean and standard deviation (n = 3).
c. Deconvolution of the mass spectrum shown in panel a.
d. Deconvoluted mass spectrum of non-trapped MsbA in the presence of the same amount of KLA. A significant reduction in KLA binding to MsbA is observed.

4.2.4 Structural determination of vanadate-trapped MsbA in complex with KLA

Of all the lipids in this study, KLA is the lipid of most interest because it is an LPS precursor known to stimulate MsbA ATPase activity (Figure 4.1), and KLA binds MsbA with an affinity greater than the other lipids (Figure 4.5b). To understand how vanadate-trapped MsbA binds KLA, I determined the structure of vanadate-trapped MsbA binds KLA in complex with ADP·VO_4 by cryo-EM. The resolution of the final map is 3.47 Å (Figure 4.6, Table 4.4)

Table 4.4: Statistics of cryo-EM model refinement and geometry for vanadate-trapped MsbA

Model	
Composition (#)	
Chains	2
Atoms	9314 (Hydrogens: 0)
Residues	Protein: 1152 Nucleotide: 0
Water	0
Ligands	KDL: 2
	AOV: 2
Bonds (RMSD)	
Length (Å) (# > 4σ)	0.003 (0)
Angles (°) (# > 4σ)	0.680 (2)
MolProbity score	1.52
Clash score	9.89
Ramachandran plot (%)	
Outliers	0.00
Allowed	1.13
Favored	98.87
Rama-Z (Ramachandran plot Z-score, RMSD)	
whole (N = 1148)	1.55 (0.26)
helix (N = 728)	1.68 (0.20)
sheet (N = 50)	-0.04 (0.72)
loop (N = 370)	-0.03 (0.35)
Rotamer outliers (%)	0.00
Cβ outliers (%)	0.00
Peptide plane (%)	
Cis proline/general	0.0/0.0
Twisted proline/general	0.0/0.0
CaBLAM outliers (%)	0.52
ADP (B-factors)	
Iso/Aniso (#)	9314/0
min/max/mean	
Protein	189.70/354.03/250.35

Table 4.4 continued: Statistics of cryo-EM model refinement and geometry for vanadate-trapped MsbA

Ligand	236.77/291.74/281.82
Data	
Box	
Lengths (Å)	75.62, 86.27, 137.39
Angles (°)	90.00, 90.00, 90.00
Supplied Resolution (Å)	3.6
Resolution Estimates (Å)	Masked
d FSC (half maps; 0.143)	3.6
d 99 (full/half1/half2)	3.8/4.5/4.6
d model	3.8
d FSC model (0/0.143/0.5)	3.4/3.6/3.8
Map min/max/mean	-1.00/1.42/0.04
Model vs. Data	
CC (mask)	0.56
CC (box)	0.48
CC (peaks)	0.36
CC (volume)	0.58
Mean CC for ligands	0.45

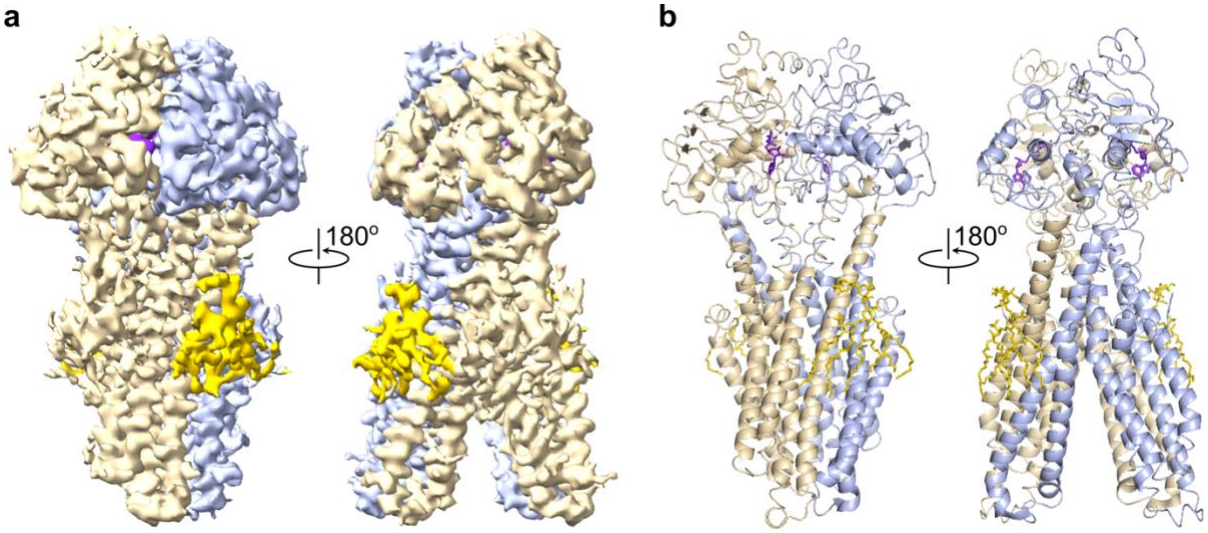

Figure 4.6 The structure of the vanadate-trapped MsbA transporter in complex with KLA and ADP·VO_4

a. Unsharpened cryo-EM map of the vanadate-trapped MsbA **transporter** in complex with KLA and ADP·VO_4 (contoured at a level of 0.11). Two subunits of MsbA are colored in wheat and purple-blue, respectively. The KLA molecule is colored in gold, and the ADP·VO_4 molecule is colored in purple.

b. Overall structure of the vanadate-trapped MsbA transporter in complex with KLA and ADP·VO_4. The color code is the same as in a.

The headgroup of KLA is bound to the basic patch, and the acyl chains of KLA interact with the hydrophobic region of MsbA (Figure 4.6a). The buried surface area between KLA and MsbA is 2,779 $Å^2$. The characteristic phosphoglucosamine (P-GlcN) substituents of KLA form extensive hydrogen bonds and salt bridges with one subunit of MsbA, with R236, Q240, and K243 interacting with one of the Kdo groups of KLA on the other MsbA subunit (Fig 4.6b). Furthermore, the clear electron density of ADP·VO_4 is observed in the NBDs, coordinated by a series of conserved residues (Fig 4.6c)

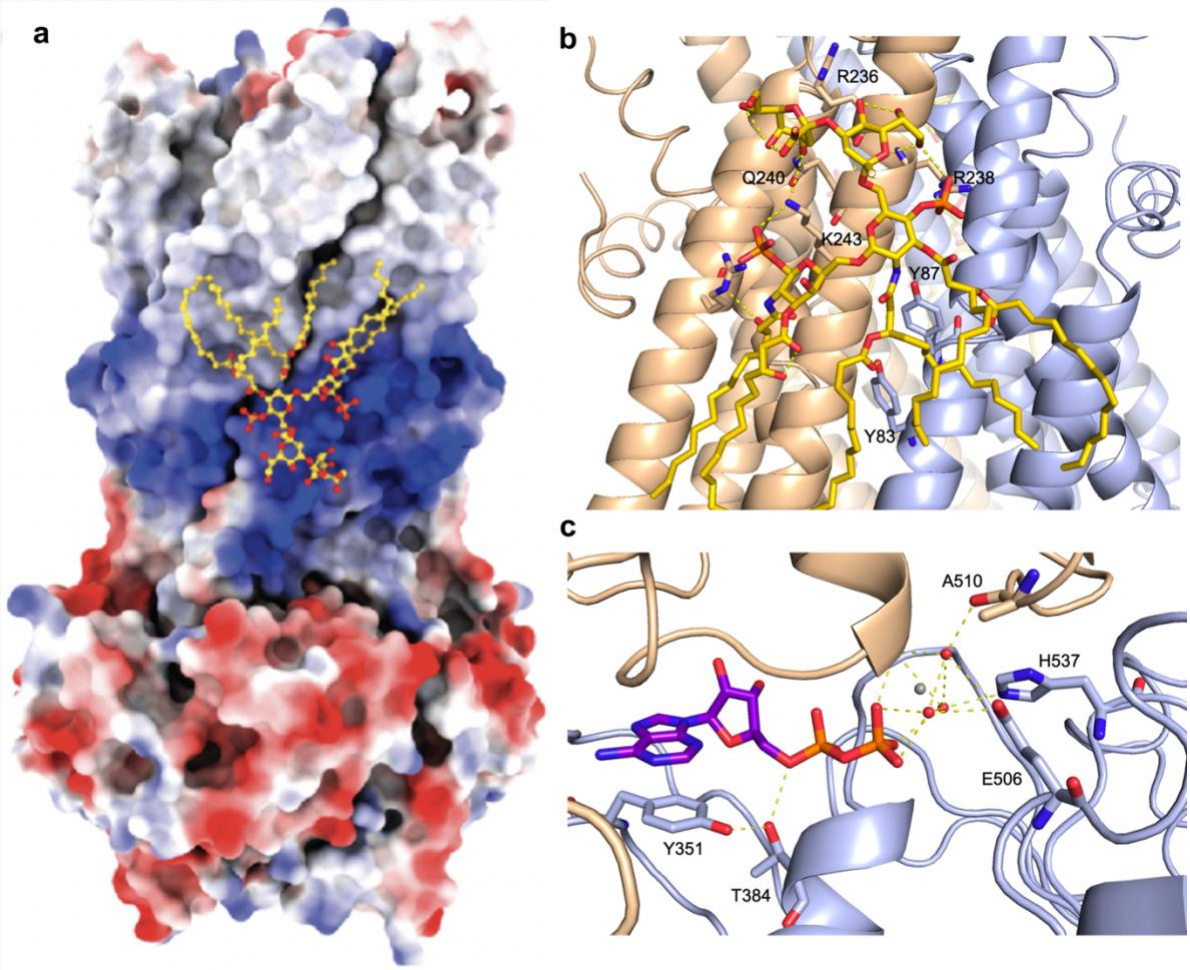

Figure 4.7 Vanadate-trapped MsbA binds KLA and ADP·VO_4
a. Coulombic electrostatic potential (scale bar -10 to +10 units as calculated by PyMOL) is colored red and blue for negative and positive charges, respectively. KLA is shown in ball and sticks representation
b. Different views of KLA bound to MsbA with interacting residues shown in stick representation. Bonds are shown as dashed yellow lines. Residues are labeled.

(Figure 5.6 continued) c. View of the bound ADP·VO_4 and interacting residues. Shown as described in b.

4.3 Discussion and conclusion

In this study, our collaborator found that MsbA co-purified with copper(II), which has not been reported before. Besides, they also characterized the MsbA-lipid interactions in different conformational states by native MS, and they found that protein-lipid interactions are directly affected by copper(II)-binding and the conformational states of MsbA. Structural studies revealed that 1) cooper(II) is bound to the first four residues of MsbA, with a similar folding as the GHK peptide; 2) the binding sites of KLA and ADP·VO_4 are clearly observed in the structure of the vanadate-trapped MsbA transporter. Our findings provide new insights into metal and lipid regulation of MsbA.

However, a major drawback of the study is that the density of the first four residues of MsbA transporter in both conformational states is not clear. In that case, the roles of these four residues and copper(II)-binding in the entire MsbA transporters are not clear.

4.4 Methods

4.4.1 Materials

The protein samples were purified by our collaborator and sent on ice overnight. All the samples were vitrified and screened on the date that the sample arrived. The holey carbon grids used in this study are Quantifoil 300 mesh Cu 1.2/1.3.

4.4.2 Electron microscopy sample preparation for open-state MsbA and vanadate-trapped MsbA

Sample vitrification was performed using a Vitrobot Mark IV (Thermo Fisher) operating at 8 ºC and 100% humidity. A total of 3.5 µL of sample (8 mg/mL open-state MsbA or vanadate-trapped MsbA in 200mM NaCl, 20mM HEPES pH 7.4, and 0.07% $C_{10}E_5$) was applied to 300-mesh holey carbon grids glow-discharged for 30 seconds. The grids were blotted for 5 seconds at a blotting force of 1 using standard Vitrobot filter paper (Ted Pella, 47000-100), and then plunged into liquid ethane.

4.4.3 Data collection for single-particle cryo-EM

The optimized grids were sent to the Advanced Electron Microscopy Facility at the University of Chicago for data collection. The dataset was collected as movie stacks with a Titan Krios electron microscope operating at 300 kV, equipped with a K3 direct detector camera. Images were recorded at a magnification of 81,000x at super-resolution counting mode by image shift. The total exposure time was set to 4 s with a frame recorded every 0.1 s, resulting in 40 frames in a single stack with a total exposure of around 50 electrons/Å^2. The defocus range was set at -1.0 to -2.5 µm (Table 4.5)

Table 4.5: Statistics of cryo-EM data collection and processing for open-state MsbA and vanadate-trapped MsbA

	open-state MsbA	vanadate-trapped MsbA
Microscope	Krios (University of Chicago)	Krios (University of Chicago)
Magnification	81,000	81,000
Voltage (kV)	300	300
Spherical aberration (mm)	2.7	2.7
Detector	K3	K3
Camera mode	Super resolution counting	Super resolution counting
Exposure rate (e^-/pixel/s)	15	15
Total exposure (e^-/Å^2)	50	50

Table 4.5 continued: Statistics of cryo-EM data collection and processing for open-state MsbA and vanadate-trapped MsbA

Defocus range (μm)	-1.0 to -2.5	-1.0 to -2.5
Pixel size (Å)	0.5325 (1.065 physical)	0.5325 (1.065 physical)
Mode of data collection	Image shift	Image shift
Energy filter	20 eV slit	20 eV slit
Software for data collection	EPU	EPU
Number of micrographs	5,530	4,221
Symmetry imposed	C2	C2
Box size (pixel)	280	200
Initial particle images (no.)	3,491,284	1,442,923
Particle images for 3D (no.)	1,633,741	540,820
Final particle images (no.)	421,840	340,615
Map resolution, unmasked (Å)	4.2	4.0
Map resolution, masked (Å)	3.9	3.6
B-factor used for sharpening ($Å^2$)	208.6	213.7

4.4.4 Image Processing.

Collected movies were processed by cryoSPARC [94]. The detailed data processing flow is shown in Figure 4.8 (open-state MsbA) and Figure 4.9 (vanadate-trapped MsbA). Stage drift and anisotropic motion of the stack images were first corrected by patch-based motion correction [94]. CTF parameters for each micrograph were determined by patch-based CTF estimation [94]. To select the particles, the micrographs were picked using the templates generated by the blob picker. Contamination and poorly-aligned classes were disposed of after 2D classification. Then, I used different methods to process the open-state MsbA and the vanadate-trapped MsbA. For the open-state MsbA, the particles were further aligned and cleaned by three-dimensional classification three times with three initial models generated using ab initio reconstructions. Then, the best class of particles was selected for a non-uniform refinement with a C2 symmetry and a C1 symmetry, resulting in final maps resolved at 3.88 Å and 4.06 Å, respectively. For the vanadate-trapped MsbA, the aligned particles were classified by three-dimensional classification once with three initial models generated using ab initio reconstructions, followed by a non-

uniform refinement with a C2 symmetry on the best class. The final maps for the vanadate-trapped MsbA were resolved at 3.63 Å.

4.4.5 Model Building, Refinement, and Validation.

Model building, refinement, and validation for single-particle cryo-EM structures. The previously reported structure of MsbA with ADP-vanadate from *E. coli* (PDB: 5TTP) [90] was docked into the cryo-EM map using Chimera [55]. The model was manually refined using Coot [56]. Phenix [57] was used to generate the coordinates and restraint files for KLA. The final model underwent one round of real-space refinement using Phenix with secondary-structure and Ramachandran restraints. Geometry outliers were manually fixed in Coot [56]. The statistics of the final round of model refinement and the model geometry are reported in Table 4.2 and Table 4.4. Figures were generated using ChimeraX [95] and Pymol [96] (Schrödinger LLC., version 2.1).

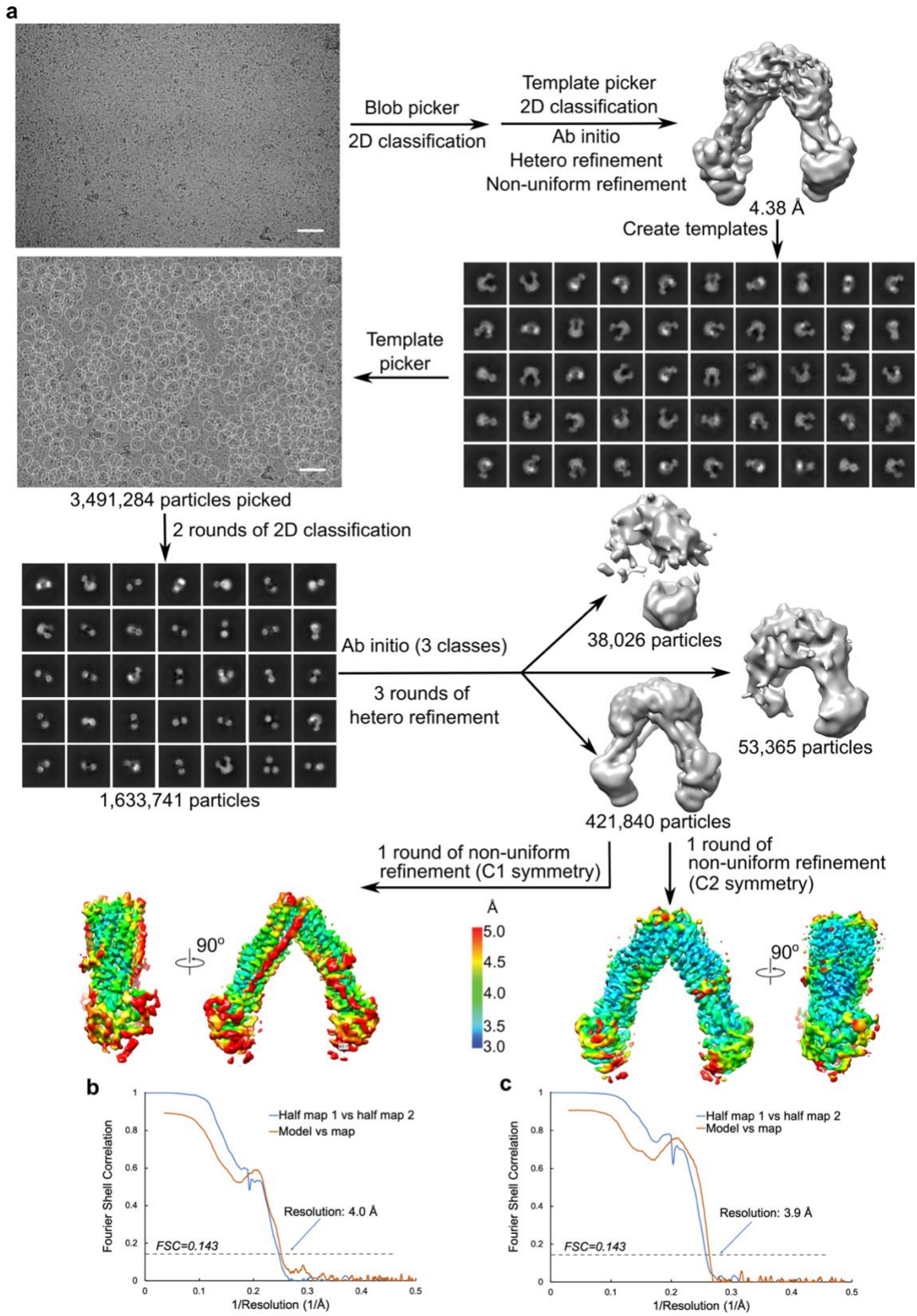

Figure 4.8 Single-particle cryo-EM analysis for the open-state MsbA

(Figure 4.8 continued) a. The workflow of data processing. The representative micrograph is shown along with a 50-nm scale bar. Particle picking was performed using the 2D templates generated by a reconstructed model from earlier runs [33]. Representative 2D class averages are shown, with the box edge corresponding to 298 Å. After disposing contamination and poorly-aligned 2D classes, the particles were classified by hetero refinement with three initial models generated using ab initio reconstructions. A non-uniform refinement was performed on the best class of particles with either C1 or C2 symmetry imposed.
b & c. Fourier shell correlation curves of the final reconstructions and model versus map for b. C1 and c. C2, respectively. The resolution of the reconstruction was determined by the FSC=0.143 criterion.

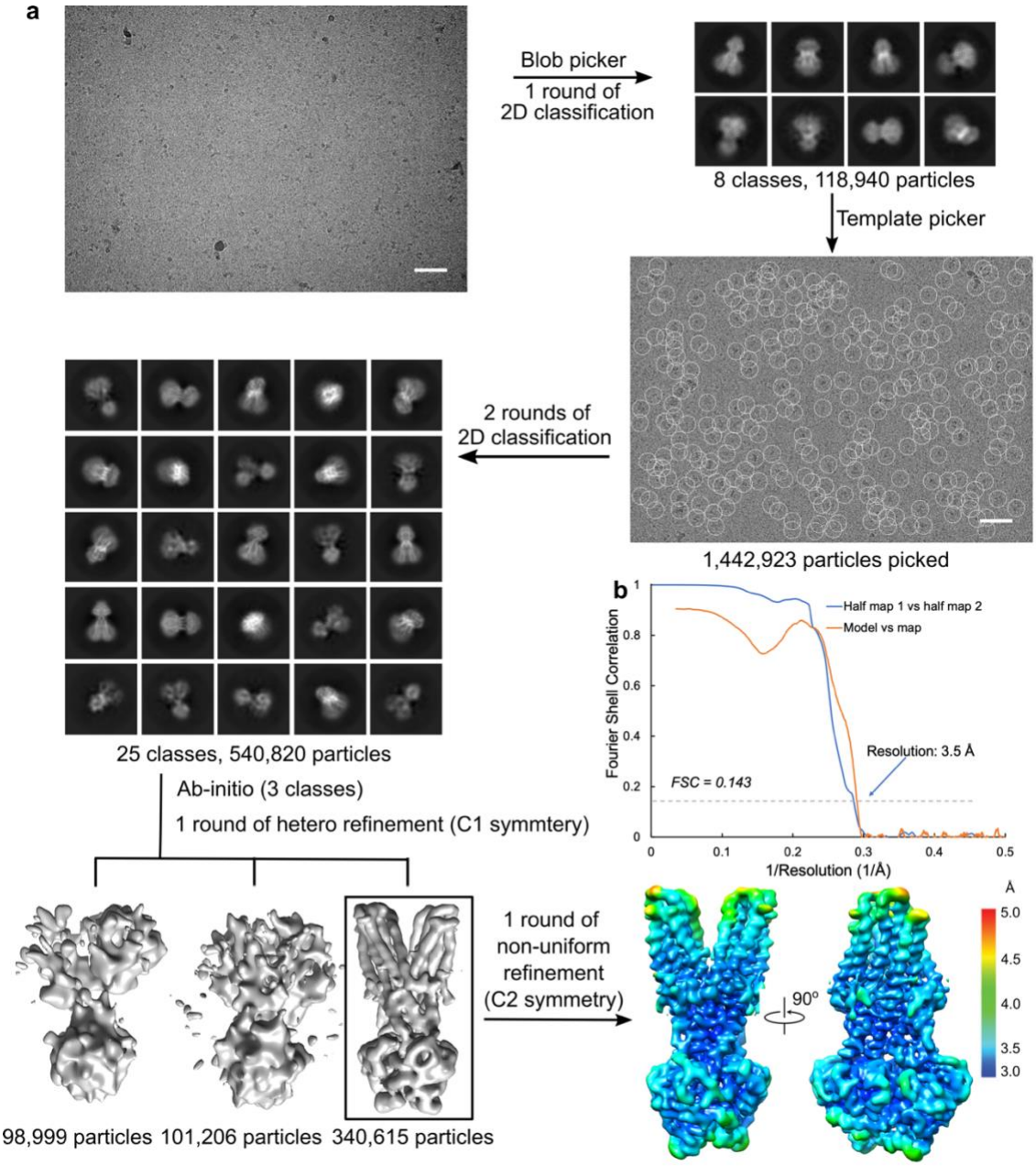

Figure 4.9 Single-particle cryo-EM analysis for the vanadate-trapped MsbA
a. The workflow of data processing. Initial motion correction was carried out using MotionCor2 [80]. A representative motion-corrected micrograph is shown along with a 50-nm scale bar. The following automatic picking and 2D classification were performed in cryoSPARC [33]. Particle selection was performed using the 2D templates generated by blob picker, followed by a 2D classification. Representative 2D class averages are shown, with the box edge corresponding to 213 Å. After disposing contamination and poorly-aligned 2D classes, the particles were classified by hetero refinement with 3 initial models generated using ab initio reconstructions. A non-uniform refinement was performed on the best class of particles with a C2 symmetry.
b. Fourier shell correlation curves of the final reconstruction and model versus map. The resolution of the reconstruction was determined by the FSC = 0.143 criterion.

Chapter 5 Structural determination of KRas genetic mutant in complex with SOS complex

5.1 Introduction

Rat sarcoma virus protein (Ras) [97], [98] is a family of proteins that play an essential role in transmitting signals within cells, ultimately controlling cell growth [99], differentiation [100], and survival [101], [102]. Overactive Ras signaling results in cancers because Ras transmits important signals with diverse cellular roles. Three Ras genes [103] (HRAS, KRAS, and NRAS) are the most common oncogenes in human cancers, and they have high overall sequence identity and are the most commonly mutated of all discovered oncogenes. Among the three oncogenes, KRas is the most frequently mutated isoform in cancers, such as pancreatic cancer (70-90%), colon cancer (30-50%), and lung cancer (20-30%) [98], [104], [105].

Ras regulates cellular signaling by cycling between an inactive GDP-bound state (Ras-GDP) and an active GTP-bound state (Ras-GTP) [106], [107] (Figure 5.1). In normal conditions, the Ras proteins catalyze the guanine nucleotide exchange at a slow rate. The activation of Ras proteins can be triggered by guanine nucleotide exchange factors (GEFs) that reload Ras with GTP. Son of Sevenless (SOS) [108], [109] is a GEF with the cdc25 and Ras exchanger motif domains representing the minimal, functionally-competent unit termed SOS^{cat}. Previous structural studies have discovered that one SOS^{cat} molecule is able to bind two Ras proteins [109]–[113]. One Ras protein is bound to the distal (or allosteric) site, and the other is bound to the catalytic site. The Ras molecule bound at the allosteric site can remarkably increase the guanine nucleotide exchange rate of the Ras at the active site.

In this study, our collaborator investigated the molecular assemblies of the catalytic domain of SOS with KRas and oncogenic mutants by mass spectrometry, and I determined the

structure of $KRas^{G13D}$– $KRas^{G13D}$-SOS^{Cat}- $KRas^{G13D}$·GppNp ternary complex using cryo-EM to understand the interactions between oncogenic KRas mutants and SOS.

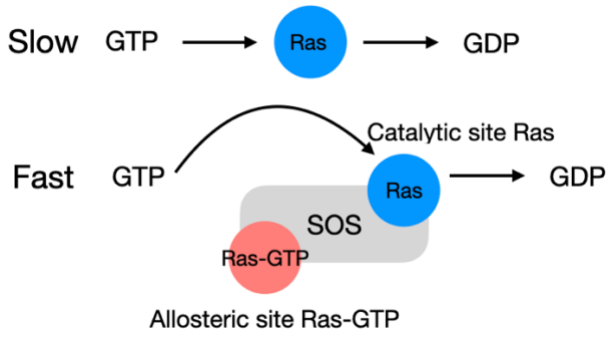

Figure 5.1 The Ras proteins can be activated by GEFs reloading Ras with GTP

5.2 Results

5.2.1 Molecular Assemblies of SOS^{cat} and KRas Oncogenic Mutants.

Our collaborator first determined the conformational dynamics of SOS^{cat} and KRas oncogenic mutants by mass spectrometry. They chose Q61H, G13D, and G12C KRas mutants due to their high occurrence in some cancers. The mutants first loaded with GTP and mixed with SOS^{cat} immediately before MS. From the MS analysis, our collaborator found that ternary complex, $KRas^{Q61H}$-SOS-$KRas^{Q61H}$·GTP, cannot be detected for this mutant. However, abundant $KRas^{Q61H}$·GTP dimer and SOS^{cat}-$KRas^{Q61H}$·GTP complex are found in the sample (Figure 5.2a,d). By contrast, $KRas^{G13D}$ solely forms the $KRas^{G13D}$-SOS-$KRas^{G13D}$·GTP ternary complex with SOS^{cat}, and some $KRas^{G13D}$·GTP dimer can also be detected by MS (Figure 5.2b,d). The third GTP-loaded mutant $KRas^{G12C}$ similarly engaged SOS^{cat} as wild-type KRas except for an increased abundance of the ternary complex containing GDP (Figure 5.2c,d)

Our collaborator next investigated the three oncogenic mutants loaded with GDP and their assembly with SOS^{cat}. $KRas^{Q61H}$-SOS^{cat} and GTP in a similar fashion as that it interacts

with GTP. However, with a 3:1 mixture of $KRas^{Q61H}$ to SOS^{cat}, a weak signal for a GDP-bound binary complex (Figure 5.2e,h). In stark contrast, $KRas^{G13D}$-SOS^{cat} interacts with GDP in a very different way than GTP. Instead of forming a uniform $KRas^{G13D}$-SOS-$KRas^{G13D}$·GTP ternary complex, it predominantly formed ternary complexes composed of $KRas^{G13D}$-SOS^{cat}-$KRas^{G13D}$·$GDP(GDP)_{0–1}$ with ~90% of the signal accounting for the complex bound to only one GDP (Figure 5.2f,h). This unexpected observation of a prevalent ternary complex for $KRas^{G13D}$·GDP suggests it can bind the allosteric site and possibly function as an allosteric modulator of SOS^{cat}. Assembly of SOS^{cat} and $KRas^{G12C}$–GDP led to the formation of complexes reminiscent of wild-type KRas but overall lower in abundance (Figure 5.2g,h).

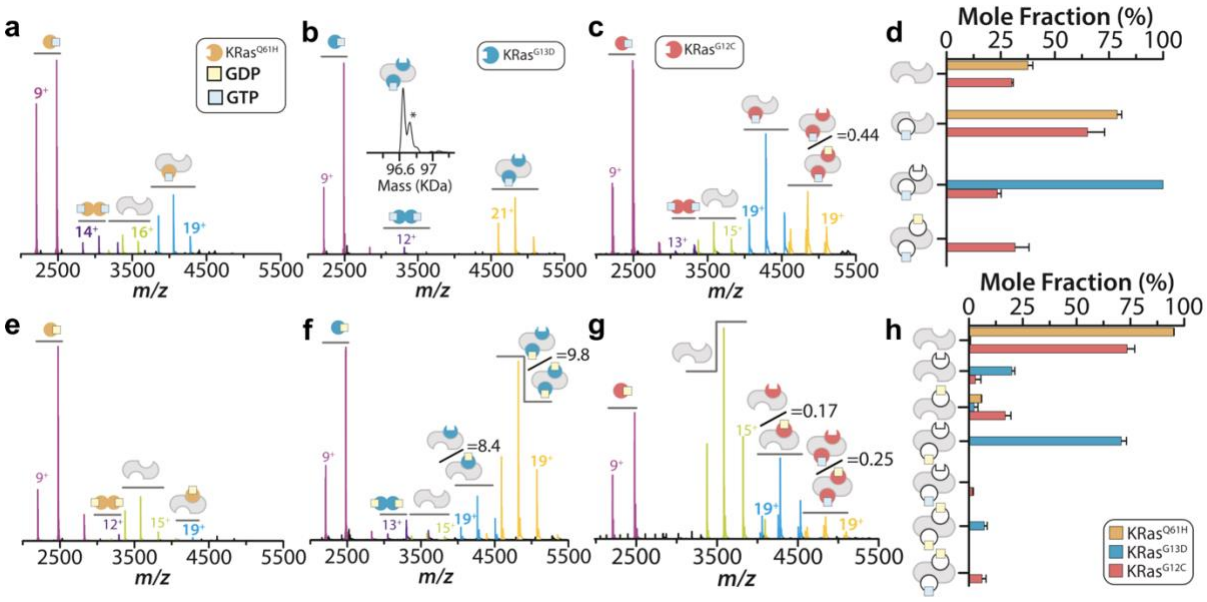

Figure 5.2 Distinct molecular assemblies of SOScat with oncogenic KRas mutants.
a-c. Native mass spectra of 2 μM SOS^{cat} mixed with three equivalents of a. KRasQ61H–GTP, b. $KRas^{G13D}$–GTP, or c. KRasG12C–GTP.
d. Plot of the mole fraction of SOS^{cat} complexes formed with GTP-loaded proteins.
e-g. Mass spectra of 2 μM SOS^{cat} mixed with a threefold molar excess of e. KRasQ61H–GDP, f. $KRas^{G13D}$–GDP, or g. KRasG12C–GDP.
h. Plot of the mole fraction of SOS^{cat} complexes formed with GDP-loaded proteins. and KRas

5.2.2 Structural determination of the $KRas^{G13D}$– $KRas^{G13D}$-SOS^{Cat}- $KRas^{G13D}$·GppNp ternary complex

The unique property that $KRas^{G13D}$ forms predominantly ternary complexes with SOS^{cat} encouraged us to determine the structure of $KRas^{G13D}$-SOS^{Cat} loaded with 5′-guanylyl imidodiphosphate (GppNp), a nonhydrolyzable analog of GTP by cryo-EM. The final resolution of the map is 3.47 Å (Table 5.1, Figure 5.3). In the structure, the $KRas^{G13D}$ and $KRas^{G13D}$–GppNp bound at the active and allosteric sites of SOS^{cat}, respectively, in line with MS results.

Table 5.1: Statistics of cryo-EM model refinement and geometry for $KRas^{G13D}$– $KRas^{G13D}$-SOS^{Cat}- $KRas^{G13D}$·GppNp ternary complex

Model	
Composition (#)	
Chains	3
Atoms	6588 (Hydrogens: 0)
Residues	Protein: 803 Nucleotide: 0
Water	0
Ligands	GNP: 1
	MG: 1
Bonds (RMSD)	
Length (Å) (# > 4σ)	0.005 (0)
Angles (°) (# > 4σ)	0.562 (0)
MolProbity score	1.35
Clash score	6.33
Ramachandran plot (%)	
Outliers	0.00
Allowed	2.02
Favored	97.98
Rama-Z (Ramachandran plot Z-score, RMSD)	
whole (N = 791)	0.05 (0.27)
helix (N = 64)	0.13 (0.68)
sheet (N = 43)	0.71 (0.66)
loop (N = 684)	0.13 (0.22)
Rotamer outliers (%)	0.00
Cβ outliers (%)	0.00
Peptide plane (%)	
Cis proline/general	2.5/0.0
Twisted proline/general	0.0/0.0
CaBLAM outliers (%)	0.13

Table 5.1 continued: Statistics of cryo-EM model refinement and geometry for $KRas^{G13D}$–$KRas^{G13D}$-SOS^{Cat}- $KRas^{G13D}$·GppNp ternary complex

ADP (B-factors)	
Iso/Aniso (#)	6588/0
min/max/mean	
Protein	21.61/141.57/50.14
Ligand	41.29/49.62/49.37
Data	
Box	
Lengths (Å)	73.48, 100.11, 124.61
Angles (°)	90.00, 90.00, 90.00
Supplied Resolution (Å)	3.5
Resolution Estimates (Å)	Masked
d FSC (half maps; 0.143)	3.5
d 99 (full/half1/half2)	6.4/3.3/3.4
d model	4.1
d FSC model (0/0.143/0.5)	3.0/3.1/3.4
Map min/max/mean	-0.14/0.24/0.00
Model vs. Data	
CC (mask)	0.85
CC (box)	0.74
CC (peaks)	0.72
CC (volume)	0.83
Mean CC for ligands	0.92

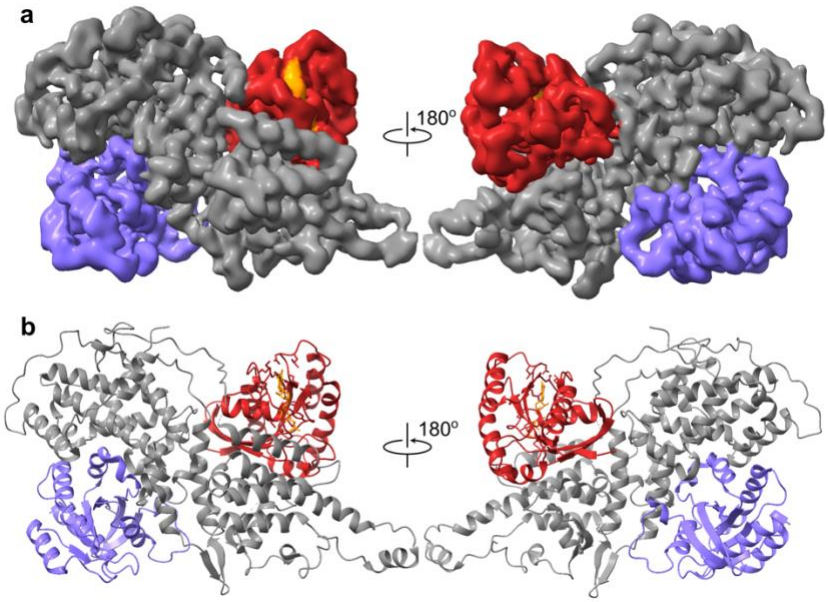

Figure 5.3 The structure of the $KRas^{G13D}$-SOS^{Cat}- $KRas^{G13D}$·GppNp complex

(Figure 5.3 continued) a. Unsharpened cryo-EM map of the KRas^{G13D}-SOS^{Cat}-KRas^{G13D}·GppNp complex
(contoured at a level of 0.02)
b. Overall structure of the open-state KRas^{G13D}-SOS^{Cat}- KRas^{G13D}·GppNp complex

5.2.3 KRas^{G13D}·GppNp complex binds SOS^{cat} at the allosteric site

The structure of KRas^{G13D}-SOS^{cat}-KRas^{G13D}·GppNp complex is very similar to that of the ternary complex HRas-SOS^{cat}-HRas·GppNp complex (PDB: 1NVW), with an RMSD of 0.668 Å (Figure 5.4a). However, the interactions between the SOS^{cat} and the Ras-GppNp molecule are different in these two complexes (Figure 5.4b,c).

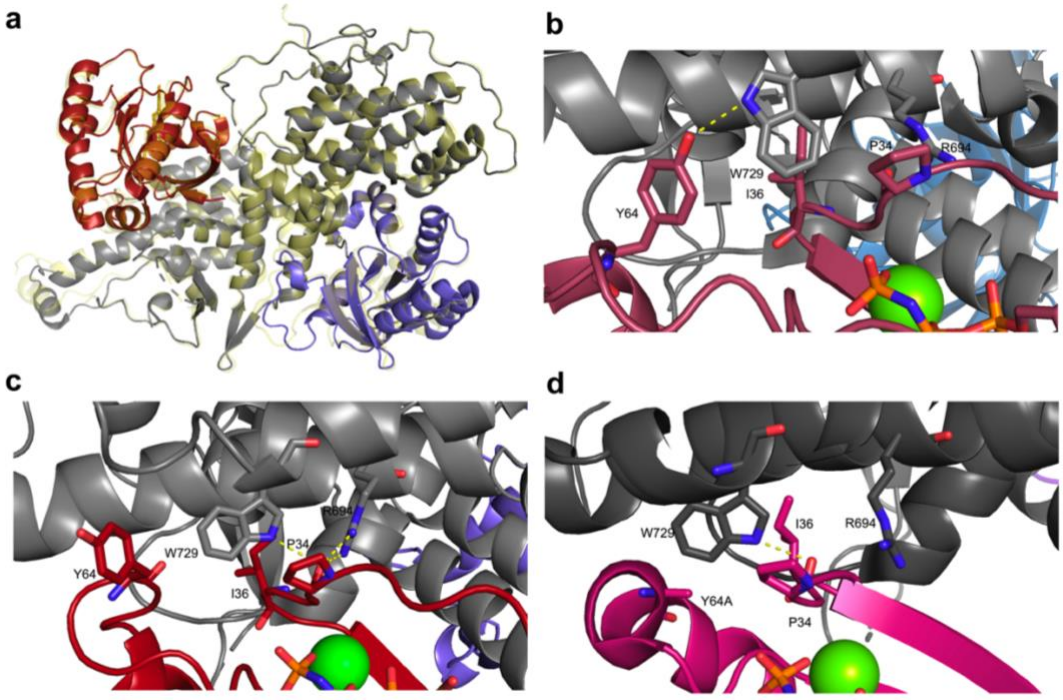

Figure 5.4 Comparison between the HRas-SOS^{cat}-HRas·GppNp complex and KRas^{G13D}-SOS^{cat}-KRas^{G13D}·GppNp complex at the allosteric site
a. The KRas-SOS^{cat}-KRas·GppNp complex is superimposed on the HRas-SOS^{cat}-HRas·GppNp complex (PDB: 1NVW). The HRas-SOS^{cat}-HRas·GppNp complex is colored in yellow
b-d. The interfaces of SOS^{cat} and KRas·GppNp at the allosteric site of b. HRas-SOS^{cat}-HRas·GppNp complex (PDB: 1NVW), c. KRas^{G13D}-SOS^{cat}-KRas^{G13D}·GppNp complex, and d. HRas^{Y64A}-SOS^{cat}-HRas^{Y64A}·GTP complex (PDB: 4NYI)

In the $KRas^{G13D}$-SOS^{cat}-$KRas^{G13D}$·GppNp complex, the side chain of W729 in SOS^{cat} is reoriented and forms hydrogen bound with P34 of $KRas^{G13D}$ instead of Y64. The changed orientation of W729 is similar to that of the $HRas^{Y64A}$-SOS^{cat}-$HRas^{Y64A}$· GppNp complex (PDB: 4NYI) at the allosteric site (Figure 5.4d). Previous studies showed that Y64A mutation in HRas results in a drastic decline in the binding affinity of HRas with SOS^{cat} at the active site [114]. However, MS results showed that the binding affinity of the SOS^{cat} with $KRas^{G13D}$ actually increased at the active site, probably due to the conformational changes of the flexible regions, which are not shown in the structure and the interfaces at the active site. Additionally, the conformation of Ras·GTP allosterically regulates the binding affinity of KRas at the active site. Due to the changes at the allosteric site, the binding affinity of SOS^{cat} and KRas molecule at the active site was expected to alter.

5.2.4 $KRas^{G13D}$ conformation at the active site of $KRas^{G13D}$-SOS^{cat}-$KRas^{G13D}$·GppNp

Overall, the conformation of $KRas^{G13D}$ bound at the active site of SOS^{cat} is also different from that of HRas and $KRas^{G12C}$ in these three regions.

At the active site, G13D mutation forms hydrogen bonds with N86 and K117 at the active site of SOS^{cat}. These interactions orientate its side chain away from the GTP/GDP-binding pocket, resulting in binding GTP to a greater extent. Besides the nucleotide-binding pocket, three key regions play an essential role in cell signaling regulation of Ras proteins: p-loop (residues 10–17), switch I (residues 30–38), and switch II (residues 60–76). G13D mutation also displaced p-loop and switch I in $KRas^{G13D}$ (Figure 5.5 a), compared to the HRas at the active site of SOS^{cat}. Comparison of the $KRas^{G13D}$–GDP structure (PDB: 6E6G) with $KRas^{G13D}$ bound at the active site reveals the side chain of G13D mutation is rotated 180° and pointing away from N86. The

observed conformation of KRas^{G13D} also differs from the structure of a mutated form of KRas^{G12C} bound only at the active site of SOS^{cat} (PDB: 6EPM) (Figure 5.5b,c).

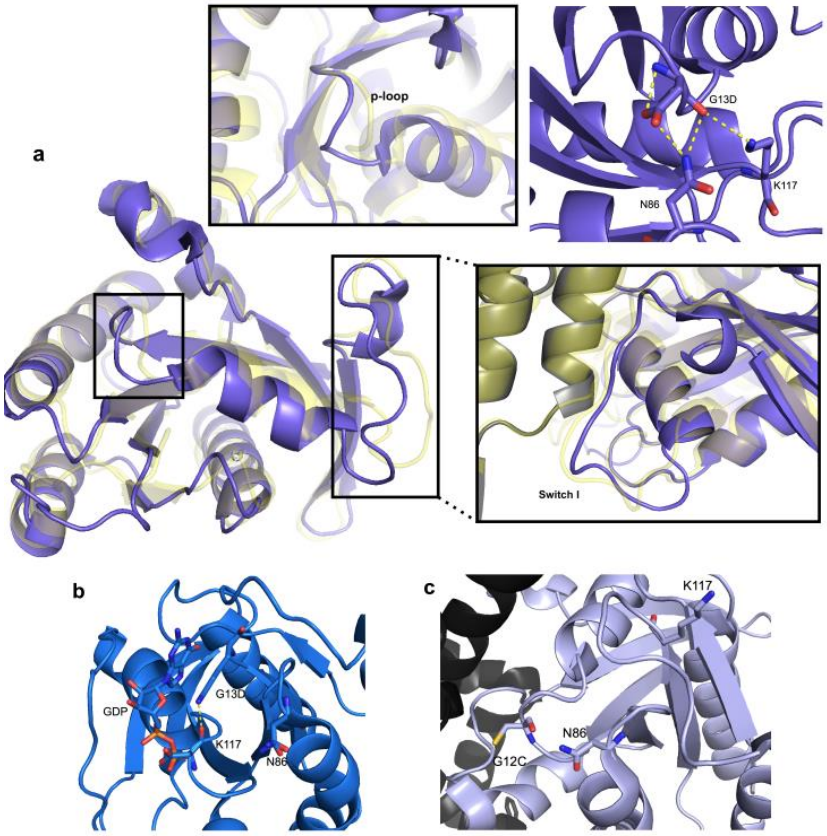

Figure 5.5 G13D mutation enlarges the nucleotide-binding pocket and displaces p-loop and switch I of the Ras molecule at the active site
a. The KRas^{G13D} at the active site of the KRas^{G13D}-SOS^{cat}-KRas^{G13D}·GppNp complex is superimposed on the HRas at the active site of the HRas-SOS^{cat}-HRas·GDP complex
b. Molecular interactions among G13D, K117, and N86 in the KRas^{G13D}·GDP complex structure (PDB: 6E6G)
c. Molecular interactions among G12C, K117, and N86 in the KRas^{G12C} structure (PDB: 6EPM)

5.3 Discussion and conclusion

The structure of the KRas^{G13D}-SOS^{Cat}- KRas^{G13D}·GppNp complex provides mechanistic insights into the higher binding affinity of KRas^{G13D} for SOS^{Cat}. The interactions between the KRas^{G13D} molecule at the active site and the SOS^{cat} molecule are compared with those in complex with HRas and KRas^{G12C}. G13D mutation changes the conformation of the side chains

adjacent to the cavity binding GTP, resulting in binding GTP to a greater extent than the wild-type proteins. The hypothesis is also supported by the MS analysis that $KRas^{G13D}$ solely forms the $KRas^{G13D}$-SOS-$KRas^{G13D}$·GTP ternary complex with SOS^{cat} and can barely observe complex engaging GDP. Besides, considering the structural similarities between the $KRas^{G13D}$-SOS-$KRas^{G13D}$·GTP and HRas-SOS-HRas·GTP, molecular interactions between Ras molecules and SOS^{cat} would be similar for HRas, KRas, and $KRas^{G13D}$. Therefore, the enhanced activity of the $KRas^{G13D}$ is due to the altered dynamics of the mutant protein. With negatively charged residues adjacent to the GTP-binding pocket, the GDP is more readily detached from the $KRas^{G13D}$ protein upon completion of converting GTP to GDP, resulting in a higher exchange rate.

5.4 Methods

5.4.1 Materials

The protein samples were purified by our collaborator and sent on ice overnight. All the samples were vitrified and screened on the date that the sample arrived. The holey carbon grids used in this study are Quantifoil 200 mesh Cu 1.2/1.3.

5.4.2 Cryo-EM grids preparation

The $KRas^{G13D}$-SOS^{cat}-$KRas^{G13D}$·GppNp complex was concentrated to 8.6 mg/mL for cryo-EM grid preparation. To reduce protein aggregation on the grids, 0.5% fluorinated octyl maltoside (FOM) was added to the samples to a final concentration of 0.05% right before freezing. Sample vitrification was performed using a Vitrobot Mark IV (Thermo Fisher) operating at 8 ^{o}C and 100% humidity. A total of 3.5 μL of the sample was applied to holey carbon grids glow-discharged for 30 seconds. The grids were blotted for 1 second at a blotting

force 1 by standard Vitrobot filter paper (Ted Pella, 47000-100), and then plunged frozen in liquid ethane.

5.4.3 Data collection for single-particle cryo-EM

Frozen grids were shipped to the National Center for Cryo-EM Access and Training (NCCAT) for data collection. The dataset was acquired as movie stacks with a Titan Krios electron microscope operating at 300 kV, equipped with a K3 direct detector camera. Images were recorded at a magnification of 105,000x at super-resolution counting mode by image shift. The total exposure time was set to 2.5 s with a frame recorded every 0.05 s, resulting in 50 frames in a single stack with a total exposure of around 65 electrons/Å^2. The defocus range was set at -0.8 to -2.5 μm (Table 5.2).

Table 5.2: Statistics of cryo-EM data collection and processing for KRas^{G13D}-SOS^{Cat}-KRas^{G13D}·GppNp ternary complex

	KRas^{G13D}-SOS^{cat}-KRas^{G13D}·GppNp complex
Microscope	Krios (NCCAT)
Magnification	105,000
Voltage (kV)	300
Spherical aberration (mm)	2.7
Detector	K3
Camera mode	Super resolution counting
Exposure rate (e$^-$/pixel/s)	30
Total exposure (e$^-$/Å^2)	65
Defocus range (μm)	-0.8 to -2.5
Pixel size (Å)	0.5325 (1.065 physical)
Mode of data collection	Image shift
Energy filter	30 eV slit
Software for data collection	Leginon
Number of micrographs	5,202
Symmetry imposed	C1
Box size (pixel)	192
Initial particle images (no.)	5,749,548
Particle images for 3D (no.)	2,294,620
Final particle images (no.)	1,021,288
Map resolution, unmasked (Å)	3.85

Table 5.2 continued: Statistics of cryo-EM data collection and processing for $KRas^{G13D}$-SOS^{Cat}- $KRas^{G13D}$·GppNp ternary complex

Map resolution, masked (Å)	3.47
B-factor used for sharpening ($Å^2$)	-162
EMD accession code	7kfz

5.4.4 Image Processing

Stack images were subjected to beam-induced motion correction by MotionCor2 [80]. CTF parameters for each micrograph were determined by CTFFIND4 [115]. The following particle selection, two- and three-dimensional classifications, three-dimensional refinement, and post-processing were performed in RELION-3.1 [34]. The detailed data processing flow is shown in Figure 5.5. Briefly, particles were selected by automatic picking, using the 2D templates generated from ~2,000 manually picked particles. Contamination and poorly-aligned classes were disposed after 2D classification. 2,294,620 particles were selected for the initial 3D classification using a previously reported HRas-SOS^{cat}-HRas·GDP complex (PDB: 1XD2) as a reference model, followed by 3D refinement, post-processing, CTF refinement, and Bayesian polishing performed on the best class. The final map for the $KRas^{G13D}$-SOS^{cat}-$KRas^{G13D}$·GppNp complex was resolved at 3.47 Å.

5.4.5 Model Building, Refinement, and Validation

Previously reported HRas-SOS^{cat}-HRas·GDP complex structure (PDB: 1XD2) was docked into the cryo-EM map as a rigid body using Chimera. HRas-GDP bound at the allosteric site was replaced by aligning the $KRas^{G13D}$ bound to GppNp (PDB: 6E6F) in Coot. In a similar fashion, the HRas molecule bound at the active site was replaced by aligning $KRas^{G12C}$ (PDB: 6EPL) followed by mutation of residues to match the $KRas^{G13D}$ sequence in Coot. After these operations, the model was manually refined using Coot followed by one round of real-space

refinement using Phenix with secondary-structure and Ramachandran restraints. Geometry outliers were manually fixed in Coot. The statistics of the final round of model refinement and the model geometry are shown in Table 5.1.

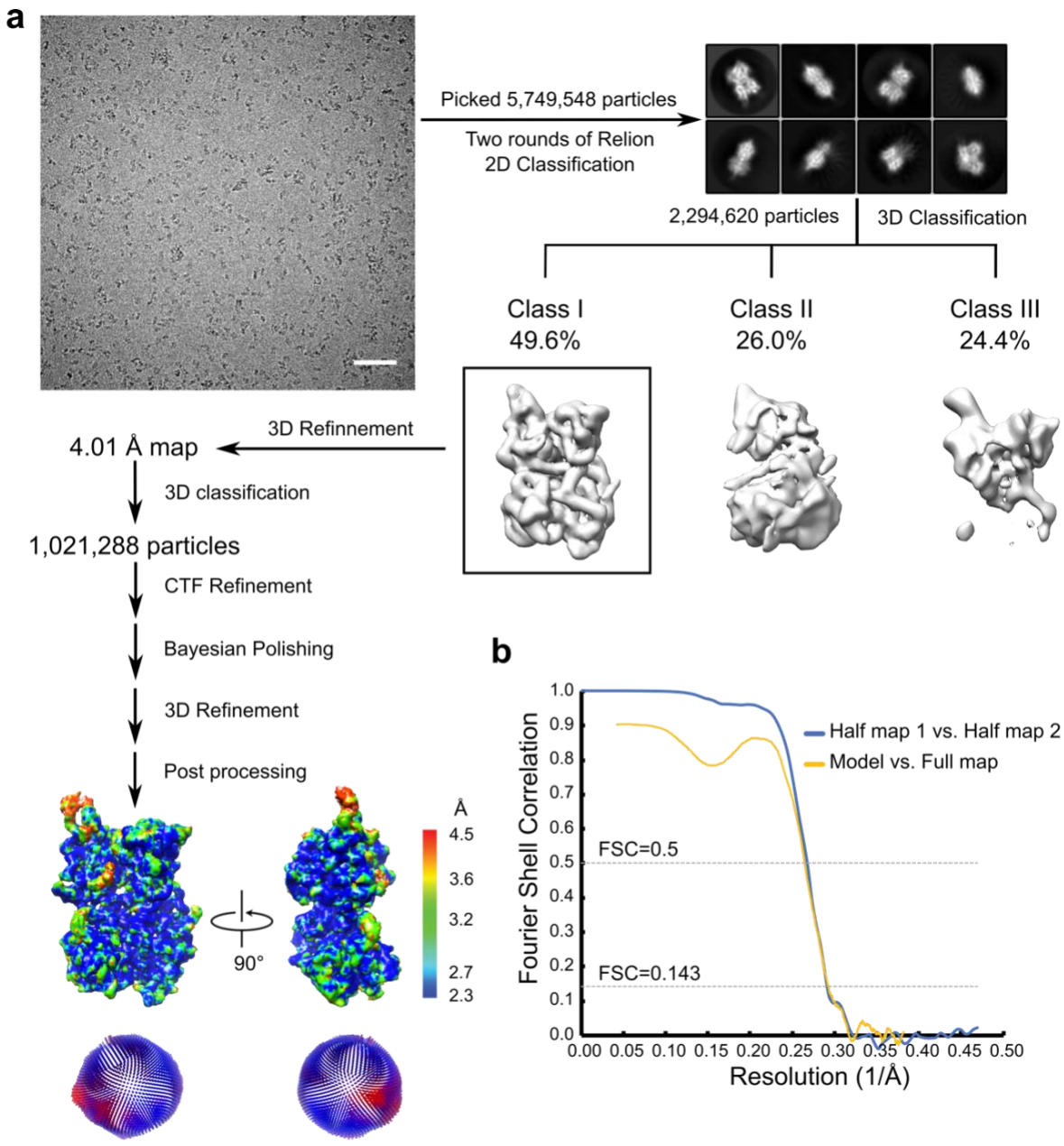

Figure 5.6 Single-particle cryo-EM analysis for the KRas^{G13D}-SOS^{cat}-KRas^{G13D}·GppNp complex

a. The workflow of data processing. The data was acquired at the National Center for cryo-EM Access and Training (NCCAT) using a Titan Krios electron microscope operating at 300 KV.

(Figure 5.6 continued) Initial motion correction was carried out using MotionCor2 . A representative micrograph is shown along with a scale bar corresponding to 50 nm. The following steps were performed in Relion . Representative 2D class averages are shown, with the box edge corresponding to 170 Å. After disposing contamination and poorly-aligned 2D classes, three maps were resolved using 3D classification. Subsequent refinement and polishing were performed on the best 3D class.
b. Fourier shell correlation curves of the half map 1 versus the half map 2 (blue), and the refined model versus the overall map (yellow). The resolution of the reconstruction was determined by the FSC= 0.143 criterion.

Chapter 6 Summary and perspectives

6.1 Technical and methodological improvements of single-particle cryo-EM analysis

Over the past decade, hardware and software developments of single-particle cryo-EM analysis have enabled the determination of macromolecules at high resolution with routine methods, providing many novel molecular insights into biological questions. However, improvements can still be made in various aspects of the single-particle cryo-EM analysis workflow.

First, the improvements of sample vitrification. Commercialization of ample preparation devices has enabled convenient cryo-EM sample preparation. However, the increasing demand for cryo-EM reveals some common problems of sample vitrification, for example, protein degradation and preferential orientation problems caused by the air-water interface [116]. To solve those problems, next-generation cryo-EM sample preparation robots, such as Spotiton [117], VitroJet [118], and CryoWriter [119], have emerged in recent years. Their basic principle is that robots spray proteins onto commercial or self-wicking grids in a computationally controlled system[117]–[119]. Compared to the traditional cryo-EM sample preparation method, the next-generation robots do not need to blot the grids with Vitrobot filter papers and plunge freeze grids at a higher speed [116], [117], [120], [121]. These advances reduce contact of proteins with the air-water interface and assist with protein degradation and preferred orientation problems. Another benefit of these novel models is that fewer protein samples are needed because they only use several nanoliters instead of microliters.

Second, lowering the cost of cryo-EM to promote its further development. Although a boom in cryo-EM structures has been observed in recent years, a high cost of purchase and

maintenance makes cryo-EM still prohibitive to many institutions, especially the cryo-EM operated at 300 kV. Some research groups showed that the two-condenser lens TEMs operating at 200 kV, less expensive cryo-EM models, are also able to produce high-resolution structures even for proteins less than 100 kDa [122], [123]. Theoretically, some benefits are brought by the cryo-EM operated at a lower voltage, including higher image contrast [124], a greater ratio between the elastic and inelastic scattering electrons, and lower radiation damage. However, in reality, the choice of the instrument still depends largely on the biological specimen itself.

Third, improvements in software for data collection and processing [125]. In terms of data collection, image acquisition software packages such as Serial EM [126] and EPU enable automated data collection. However, some bottlenecks, including complicated EM alignments and a significant amount of required user intervention, remain to be solved. In terms of image processing, the workflow has already been standardized over the last decade. But the speed and capabilities of image processing algorithms still have much room to improve. From my perspective, a fast and automated image processing workflow during data collection is the future of cryo-EM data collection and processing software [127], [128].

6.2 The development of structural biology

The main target of structural biology is understanding macromolecules' roles in organisms. While we can understand the roles of proteins, the information we obtain from single-particle cryo-EM analysis lacks spatial and temporal dimensions. I think the future of cryo-EM is to observe protein structures at atomic resolution in real-time inside the cell.

From my perspective, to understand proteins' roles in cellular events, cryogenic electron tomography (cryo-ET) [129], an imaging technique providing three-dimensional views of

samples, has also been developed very fast recently. In a cryo-ET study, a cell, tissue, or organism, flashed frozen and thinned to an appropriate thickness, is imaged using EM. Similar to cryo-EM, the sample is in a close-to-native state. The sample can be tilted in the EM with images captured at different angles. Then, the images are aligned and merged to reconstruct a tomogram using computational techniques.

Static structures of proteins provide limited information on how proteins function. Time-resolved structural biology studies shed light on how the conformation of proteins evolves as it executes their function. Recently, time-resolved X-ray crystallography [130], [131] has provided some undoubtedly impressive results. This technique successfully incorporates time as a fourth dimension to visualize the reactions of proteins. The reaction is first triggered in the protein crystal before X-ray exposure, then the diffraction patterns of the crystals are collected at different time delays.

Besides experimental methods, structural biologists are now better equipped with structural prediction platforms. AlphaFold2 [51], an AI system developed by DeepMind, is a machine learning method incorporating physical and biological knowledge about protein structure. It provides more confident prediction results of proteins based on their sequences than previous structural prediction platforms. AlphaFold2 not only predicts the proteins at apo state but also their complexes with other proteins. Although AlphaFold2 has improved drastically, its applications are still limited [132], [133]. First, AlphaFold2 cannot predict structures of proteins with small molecules, the fundamentals of drug discovery. Second, AlphaFold2 can only predict a single state, but not all conformations of proteins. Third, AlphaFold2 cannot identify the differences between protein variants with mutations. Taken together, AlphaFold2 still has much space to develop and improve.

List of References

[1] A. Srivastava, T. Nagai, A. Srivastava, O. Miyashita, and F. Tama, "Role of computational methods in going beyond x-ray crystallography to explore protein structure and dynamics," *International Journal of Molecular Sciences*, vol. 19, no. 11. MDPI AG, Nov. 01, 2018. doi: 10.3390/ijms19113401.

[2] L. Maveyraud and L. Mourey, "Protein X-ray crystallography and drug discovery," *Molecules*, vol. 25, no. 5. MDPI AG, Feb. 25, 2020. doi: 10.3390/molecules25051030.

[3] M. S. Smyth and J. H. J. Martin, "x Ray crystallography," *Molecular Pathology*, vol. 53, no. 1, p. 8, 2000, doi: 10.1136/MP.53.1.8.

[4] T. Nakane *et al.*, "Single-particle cryo-EM at atomic resolution," *Nature*, vol. 587, no. 7832, pp. 152–156, Nov. 2020, doi: 10.1038/s41586-020-2829-0.

[5] D. Elmlund and H. Elmlund, "Cryogenic electron microscopy and single-particle analysis," *Annual Review of Biochemistry*, vol. 84. Annual Reviews Inc., pp. 499–517, Jun. 02, 2015. doi: 10.1146/annurev-biochem-060614-034226.

[6] M. Carroni and H. R. Saibil, "Cryo electron microscopy to determine the structure of macromolecular complexes," *Methods*, vol. 95. Academic Press Inc., pp. 78–85, Feb. 15, 2016. doi: 10.1016/j.ymeth.2015.11.023.

[7] I. P. Gerothanassis, A. Troganis, V. Exarchou, and K. Barbarossou, "NUCLEAR MAGNETIC RESONANCE (NMR) SPECTROSCOPY: BASIC PRINCIPLES AND PHENOMENA, AND THEIR APPLICATIONS TO CHEMISTRY, BIOLOGY AND MEDICINE," 2002.

[8] Z. Serber, A. T. Keatinge-Clay, R. Ledwidge, A. E. Kelly, S. M. Miller, and V. Dötsch, "High-resolution macromolecular NMR spectroscopy inside living cells," *J. Am. Chem. Soc*, vol. 123, pp. 2446–2447, 2001.

[9] E. Luchinat and L. Banci, "A unique tool for cellular structural biology: In-cell NMR," *Journal of Biological Chemistry*, vol. 291, no. 8, pp. 3776–3784, Feb. 2016, doi: 10.1074/JBC.R115.643247/ATTACHMENT/FBEF8D7A-6A32-42B5-8DEE-BB261961A110/MMC2.PDF.

[10] J. M. Thomas, "The birth of X-ray crystallography," *Nature 2012 491:7423*, vol. 491, no. 7423, pp. 186–187, Nov. 2012, doi: 10.1038/491186a.

[11] G. McMullan, A. R. Faruqi, and R. Henderson, "Direct Electron Detectors," *Methods in Enzymology*, vol. 579, pp. 1–17, Jan. 2016, doi: 10.1016/BS.MIE.2016.05.056.

[12] R. S. Ruskin, Z. Yu, and N. Grigorieff, "Quantitative characterization of electron detectors for transmission electron microscopy," *Journal of Structural Biology*, vol. 184, no. 3, pp. 385–393, Dec. 2013, doi: 10.1016/J.JSB.2013.10.016.

[13] A. Cheng, Y. Z. Tan, V. P. Dandey, C. S. Potter, and B. Carragher, “Strategies for Automated CryoEM Data Collection Using Direct Detectors,” *Methods in Enzymology*, vol. 579, pp. 87–102, Jan. 2016, doi: 10.1016/BS.MIE.2016.04.008.

[14] E. H. Egelman, “The Current Revolution in Cryo-EM,” *Biophysj*, vol. 110, pp. 1008–1012, 2016, doi: 10.1016/j.bpj.2016.02.001.

[15] E. Callaway, “Molecular-imaging pioneers scoop Nobel,” *Nature*, vol. 550, no. 7675, p. 167, Oct. 2017, doi: 10.1038/NATURE.2017.22738.

[16] H. W. Wang and J. W. Wang, “How cryo-electron microscopy and X-ray crystallography complement each other,” *Protein Science*, vol. 26, no. 1, pp. 32–39, Jan. 2017, doi: 10.1002/PRO.3022.

[17] S. C. Shoemaker and N. Ando, “X-rays in the Cryo-EM Era: Structural Biology’s Dynamic Future”, doi: 10.1021/acs.biochem.7b01031.

[18] H. R. Saibil, “Cryo-EM in molecular and cellular biology,” *Molecular Cell*, vol. 82, no. 2, pp. 274–284, Jan. 2022, doi: 10.1016/J.MOLCEL.2021.12.016.

[19] Y. Cheng, “Membrane protein structural biology in the era of single particle cryo-EM,” *Current Opinion in Structural Biology*, vol. 52, pp. 58–63, Oct. 2018, doi: 10.1016/J.SBI.2018.08.008.

[20] S. T. Huber, E. Sarajlic, R. Huijink, F. Weis, W. H. Evers, and A. J. Jakobi, “Nanofluidic chips for cryo-EM structure determination from picoliter sample volumes,” *Elife*, vol. 11, Jan. 2022, doi: 10.7554/ELIFE.72629.

[21] D. B. Williams and C. B. Carter, “Electron Sources,” *Transmission Electron Microscopy*, pp. 73–89, 2009, doi: 10.1007/978-0-387-76501-3_5.

[22] T. Kato *et al.*, “CryoTEM with a Cold Field Emission Gun That Moves Structural Biology into a New Stage,” *Microscopy and Microanalysis*, vol. 25, no. S2, pp. 998–999, Aug. 2019, doi: 10.1017/S1431927619005725.

[23] G. McMullan, A. R. Faruqi, D. Clare, and R. Henderson, “Comparison of optimal performance at 300 keV of three direct electron detectors for use in low dose electron microscopy,” *Ultramicroscopy*, vol. 147, pp. 156–163, Dec. 2014, doi: 10.1016/J.ULTRAMIC.2014.08.002.

[24] R. F. Thompson, M. Walker, C. A. Siebert, S. P. Muench, and N. A. Ranson, “An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology,” *Methods*, vol. 100, pp. 3–15, May 2016, doi: 10.1016/J.YMETH.2016.02.017.

[25] I. Drulyte *et al.*, “Approaches to altering particle distributions in cryo-electron microscopy sample preparation,” *urn:issn:2059-7983*, vol. 74, no. 6, pp. 560–571, May 2018, doi: 10.1107/S2059798318006496.

[26] G. Weissenberger, R. J. M. Henderikx, and P. J. Peters, “Understanding the invisible hands of sample preparation for cryo-EM,” *Nature Methods 2021 18:5*, vol. 18, no. 5, pp. 463–471, May 2021, doi: 10.1038/s41592-021-01130-6.

[27] J. L. Vilas, J. M. Carazo, C. Oscar, and S. Sorzano, “Emerging Themes in CryoEM Single Particle Analysis Image Processing”, doi: 10.1021/acs.chemrev.1c00850.

[28] D. Lyumkis, “Challenges and opportunities in cryo-EM single-particle analysis,” *Journal of Biological Chemistry*, vol. 294, no. 13, pp. 5181–5197, Mar. 2019, doi: 10.1074/JBC.REV118.005602.

[29] M. Carroni and H. R. Saibil, “Cryo electron microscopy to determine the structure of macromolecular complexes,” *Methods*, vol. 95. Academic Press Inc., pp. 78–85, Feb. 15, 2016. doi: 10.1016/j.ymeth.2015.11.023.

[30] D. Elmlund and H. Elmlund, “Cryogenic electron microscopy and single-particle analysis,” *Annual Review of Biochemistry*, vol. 84. Annual Reviews Inc., pp. 499–517, Jun. 02, 2015. doi: 10.1146/annurev-biochem-060614-034226.

[31] T. Nakane *et al.*, “Single-particle cryo-EM at atomic resolution,” *Nature*, vol. 587, no. 7832, pp. 152–156, Nov. 2020, doi: 10.1038/s41586-020-2829-0.

[32] G. Tang *et al.*, “EMAN2: An extensible image processing suite for electron microscopy,” *Journal of Structural Biology*, vol. 157, no. 1, pp. 38–46, Jan. 2007, doi: 10.1016/J.JSB.2006.05.009.

[33] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker, “cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination,” *Nat Methods*, vol. 14, no. 3, pp. 290–296, Feb. 2017, doi: 10.1038/NMETH.4169.

[34] S. H. W. Scheres, “RELION: Implementation of a Bayesian approach to cryo-EM structure determination,” *Journal of Structural Biology*, vol. 180, no. 3, pp. 519–530, Dec. 2012, doi: 10.1016/J.JSB.2012.09.006.

[35] A. Maity and B. Das, “N6-methyladenosine modification in mRNA: machinery, function and implications for health and diseases,” 2015, doi: 10.1111/febs.13614.

[36] Z. M. Zhu, F. C. Huo, and D. S. Pei, “Function and evolution of RNA N6-methyladenosine modification,” *International Journal of Biological Sciences*, vol. 16, no. 11, p. 1929, 2020, doi: 10.7150/IJBS.45231.

[37] L. He, H. Li, A. Wu, Y. Peng, G. Shu, and G. Yin, “Functions of N6-methyladenosine and its role in cancer,” *Molecular Cancer 2019 18:1*, vol. 18, no. 1, pp. 1–15, Dec. 2019, doi: 10.1186/S12943-019-1109-9.

[38] C. Zhang, J. Fu, and Y. Zhou, “A review in research progress concerning m6A methylation and immunoregulation,” *Frontiers in Immunology*, vol. 10, no. APR, p. 922, 2019, doi: 10.3389/FIMMU.2019.00922/BIBTEX.

[39] X. Jiang *et al.*, “The role of m6A modification in the biological functions and diseases,” *Signal Transduction and Targeted Therapy 2021 6:1*, vol. 6, no. 1, pp. 1–16, Feb. 2021, doi: 10.1038/s41392-020-00450-x.

[40] Y. Zhang *et al.*, “m6A modification in RNA: biogenesis, functions and roles in gliomas,” *Journal of Experimental & Clinical Cancer Research 2020 39:1*, vol. 39, no. 1, pp. 1–16, Sep. 2020, doi: 10.1186/S13046-020-01706-8.

[41] S. D. Agarwala, H. G. Blitzblau, A. Hochwagen, and G. R. Fink, “RNA methylation by the MIS complex regulates a cell fate decision in yeast,” *PLoS Genetics*, vol. 8, no. 6, Jun. 2012, doi: 10.1371/JOURNAL.PGEN.1002732.

[42] Y. Xue *et al.*, “m 6 A Reader: Epitranscriptome Target Prediction and Functional Characterization of N 6-Methyladenosine (m 6 A) Readers,” 2020, doi: 10.3389/fcell.2020.00741.

[43] Y. Zhao, Y. Shi, H. Shen, and W. Xie, “M6A-binding proteins: The emerging crucial performers in epigenetics,” *Journal of Hematology and Oncology*, vol. 13, no. 1, Apr. 2020, doi: 10.1186/S13045-020-00872-8.

[44] N. A. Little +, N. D. Hastie, and R. C. Davies, “Identification of WTAP, a novel Wilms’ tumour 1-associating protein,” *Human Molecular Genetics*, vol. 9, no. 15, pp. 2231–2239, 2000.

[45] W. Zhu, J. Z. Wang, J. F. Wei, and C. Lu, “Role of m6A methyltransferase component VIRMA in multiple human cancers (Review),” *Cancer Cell International*, vol. 21, no. 1, pp. 1–13, Dec. 2021, doi: 10.1186/S12935-021-01868-1/FIGURES/4.

[46] P. Śledź and M. Jinek, “Structural insights into the molecular mechanism of the m6A writer complex,” *Elife*, vol. 5, no. September, Sep. 2016, doi: 10.7554/ELIFE.18434.

[47] G. Abstract, “Structural Basis for Cooperative Function of Mettl3 and Mettl14 Methyltransferases,” 2016, doi: 10.1016/j.molcel.2016.05.041.

[48] X. Wang *et al.*, “Structural basis of N 6-adenosine methylation by the METTL3-METTL14 complex Chemical modifications of RNA have essential roles in a vast range of cellular processes 1-3. N 6-methyladenosine (m 6 A) is an abundant internal modification in messenger RNA and long non-coding RNA that can be dynamically added and removed by RNA methyltransferases (MTases) and demethylases, respectively 2-5. An MTase complex comprising methyltransferase-like 3 (METTL3) and methyltransferase-like 14 (METTL14,” 2016, doi: 10.1038/nature18298.

[49] E. Schöller *et al.*, “Interactions, localization and phosphorylation of the m6A generating METTL3-METTL14-WTAP complex,” *RNA*, vol. 24, no. 4, p. rna.064063.117, Jan. 2018, doi: 10.1261/RNA.064063.117.

[50] M. Kobayashi *et al.*, “The RNA Methyltransferase Complex of WTAP, METTL3, and METTL14 Regulates Mitotic Clonal Expansion in Adipogenesis,” *Mol Cell Biol*, vol. 38, no. 16, Aug. 2018, doi: 10.1128/MCB.00116-18.

[51] J. Jumper *et al.*, “Highly accurate protein structure prediction with AlphaFold,” *Nature 2021 596:7873*, vol. 596, no. 7873, pp. 583–589, Jul. 2021, doi: 10.1038/s41586-021-03819-2.

[52] D. Yu *et al.*, “Human MettL3-MettL14 RNA adenine methyltransferase complex is active on double-stranded DNA containing lesions,” *Nucleic Acids Research*, vol. 49, pp. 11629–11642, 2021, doi: 10.1093/nar/gkab460.

[53] J. Liu *et al.*, “A METTL3–METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation,” 2014, doi: 10.1038/nchembio.1432.

[54] Y. Yu, Q. Zheng, S. K. Erramilli, and M. Pan, “K29-linked ubiquitin signaling regulates proteotoxic stress response and cell cycle,” *Nature Chemical Biology*, doi: 10.1038/s41589-021-00823-5.

[55] E. F. Pettersen *et al.*, “UCSF Chimera--a visualization system for exploratory research and analysis,” *J Comput Chem*, vol. 25, no. 13, pp. 1605–1612, Oct. 2004, doi: 10.1002/JCC.20084.

[56] P. Emsley and K. Cowtan, “Coot: model-building tools for molecular graphics,” *Acta Crystallogr D Biol Crystallogr*, vol. 60, no. Pt 12 Pt 1, pp. 2126–2132, Dec. 2004, doi: 10.1107/S0907444904019158.

[57] P. D. Adams *et al.*, “PHENIX: a comprehensive Python-based system for macromolecular structure solution,” *Acta Crystallogr D Biol Crystallogr*, vol. 66, no. Pt 2, pp. 213–221, 2010, doi: 10.1107/S0907444909052925.

[58] K. Hsiao, H. Zegzouti, and S. A. Goueli, “Methyltransferase-Glo: A universal, bioluminescent and homogenous assay for monitoring all classes of methyltransferases,” *Epigenomics*, vol. 8, no. 3, pp. 321–339, Mar. 2016, doi: 10.2217/EPI.15.113.

[59] M. P. Terns and R. M. Terns, “CRISPR-based adaptive immune systems,” *Current Opinion in Microbiology*, vol. 14, no. 3, pp. 321–327, Jun. 2011, doi: 10.1016/J.MIB.2011.03.005.

[60] J. A. Doudna and E. Charpentier, “The new frontier of genome engineering with CRISPR-Cas9,” *Science (1979)*, vol. 346, no. 6213, Nov. 2014, doi: 10.1126/SCIENCE.1258096/ASSET/2313E70A-5C58-4755-A0E6-2E64EE240A09/ASSETS/GRAPHIC/346_1258096_F6.JPEG.

[61] M. M. Kaminski, O. O. Abudayyeh, J. S. Gootenberg, F. Zhang, and J. J. Collins, "CRISPR-based diagnostics," *Nature Biomedical Engineering 2021 5:7*, vol. 5, no. 7, pp. 643–656, Jul. 2021, doi: 10.1038/s41551-021-00760-7.

[62] R. Barrangou and J. A. Doudna, "Applications of CriSPr technologies in research and beyond," 2016, doi: 10.1038/nbt.3659.

[63] "Classy CRISPR: Differences between Class 1 and Class 2 Systems." https://www.synthego.com/blog/crispr-systems (accessed Jul. 09, 2022).

[64] K. Chylinski, K. S. Makarova, E. Charpentier, and E. v. Koonin, "Classification and evolution of type II CRISPR-Cas systems," *Nucleic Acids Research*, vol. 42, no. 10, pp. 6091–6105, Jun. 2014, doi: 10.1093/NAR/GKU241.

[65] F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott, and F. Zhang, "Genome engineering using the CRISPR-Cas9 system," *Nature Protocols 2013 8:11*, vol. 8, no. 11, pp. 2281–2308, Oct. 2013, doi: 10.1038/nprot.2013.143.

[66] W. X. Yan *et al.*, "Functionally diverse type V CRISPR-Cas systems," *Science (1979)*, vol. 363, no. 6422, pp. 88–91, Jan. 2019, doi: 10.1126/SCIENCE.AAV7271/SUPPL_FILE/PAPV2.PDF.

[67] D. B. T. Cox *et al.*, "RNA editing with CRISPR-Cas13," *Science (1979)*, vol. 358, no. 6366, pp. 1019–1027, Nov. 2017, doi: 10.1126/SCIENCE.AAQ0180/SUPPL_FILE/AAQ0180-COX-SM.PDF.

[68] M. R. O'Connell, "Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR–Cas Systems," *Journal of Molecular Biology*, vol. 431, no. 1, pp. 66–87, Jan. 2019, doi: 10.1016/J.JMB.2018.06.029.

[69] S. Xiong, L. Zhang, and Q. Y. He, "Fractionation of proteins by heparin chromatography.," *Methods Mol Biol*, vol. 424, pp. 213–221, 2008, doi: 10.1007/978-1-60327-064-9_18/FIGURES/18_2_978-1-60327-064-9.

[70] B. Wiedenheft, S. H. Sternberg, and J. A. Doudna, "RNA-guided genetic silencing systems in bacteria and archaea," *Nature*, vol. 482, no. 7385, pp. 331–338, Feb. 2012, doi: 10.1038/NATURE10886.

[71] P. Mali *et al.*, "RNA-guided human genome engineering via Cas9," *Science (1979)*, vol. 339, no. 6121, pp. 823–826, Feb. 2013, doi: 10.1126/SCIENCE.1232033/SUPPL_FILE/MALI.SM.PDF.

[72] C. A. Lino, J. C. Harper, J. P. Carney, and J. A. Timlin, "Delivering CRISPR: a review of the challenges and approaches," *Drug Delivery*, vol. 25, no. 1, p. 1234, 2018, doi: 10.1080/10717544.2018.1474964.

[73] A. Li *et al.*, "AAV-CRISPR Gene Editing Is Negated by Pre-existing Immunity to Cas9," *Molecular Therapy*, vol. 28, no. 6, pp. 1432–1441, Jun. 2020, doi:

10.1016/J.YMTHE.2020.04.017/ATTACHMENT/23D1EE50-6EF2-4235-A741-65B63E638011/MMC1.PDF.

[74] Y. Wang, S. Sang, X. Zhang, H. Tao, Q. Guan, and C. Liu, "Efficient Genome Editing by a Miniature CRISPR-AsCas12f1 Nuclease in Bacillus anthracis," *Frontiers in Bioengineering and Biotechnology*, vol. 9, Jan. 2022, doi: 10.3389/fbioe.2021.825493.

[75] Z. Wu *et al.*, "Programmed genome editing by a miniature CRISPR-Cas12f nuclease", doi: 10.1038/s41589-021-00868-6.

[76] S. N. Takeda *et al.*, "Structure of the miniature type V-F CRISPR-Cas effector enzyme," *Molecular Cell*, vol. 81, no. 3, pp. 558-570.e3, Feb. 2021, doi: 10.1016/J.MOLCEL.2020.11.035.

[77] A. A. Sousa and R. T. Walton, "Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing," *Nature Biotechnology*, doi: 10.1038/s41587-018-0011-0.

[78] M. Kikuchi and S. Harayama, "DNA shuffling and family shuffling for in vitro gene evolution.," *Methods Mol Biol*, vol. 182, pp. 243–257, 2002, doi: 10.1385/1-59259-194-9:243/FIGURES/5.

[79] S. McGinnis and T. L. Madden, "BLAST: at the core of a powerful and diverse set of sequence analysis tools," *Nucleic Acids Research*, vol. 32, no. Web Server issue, p. W20, Jul. 2004, doi: 10.1093/NAR/GKH435.

[80] S. Q. Zheng, E. Palovcak, J. P. Armache, K. A. Verba, Y. Cheng, and D. A. Agard, "MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy," *Nat Methods*, vol. 14, no. 4, pp. 331–332, 2017, doi: 10.1038/NMETH.4193.

[81] K. Sato, Y. Kato, M. Hamada, T. Akutsu, and K. Asai, "IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming," *Bioinformatics*, vol. 27, no. 13, Jul. 2011, doi: 10.1093/BIOINFORMATICS/BTR215.

[82] M. Antczak *et al.*, "New functionality of RNAComposer: application to shape the axis of miR160 precursor structure," *Acta Biochimica Polonica*, vol. 63, no. 4, pp. 737–744, 2016, doi: 10.18388/ABP.2016_1329.

[83] A. Ward, C. L. Reyes, J. Yu, C. B. Roth, and G. Chang, "Flexibility in the ABC transporter MsbA: Alternating access with a twist," *Proc Natl Acad Sci U S A*, vol. 104, no. 48, pp. 19005–19010, Nov. 2007, doi: 10.1073/PNAS.0709388104/SUPPL_FILE/09388FIG9.JPG.

[84] A. L. Davidson, E. Dassa, C. Orelle, and J. Chen, "Structure, Function, and Evolution of Bacterial ATP-Binding Cassette Systems," *Microbiology and Molecular Biology Reviews*, vol. 72, no. 2, pp. 317–364, Jun. 2008, doi: 10.1128/MMBR.00031-

07/ASSET/37D5E1BC-EE34-4147-ACCB-FFFFDE399256/ASSETS/GRAPHIC/ZMR0020821840017.JPEG.

[85] S. Molloy, “The flip side of MsbA,” *Nature Reviews Microbiology 2005 3:7*, vol. 3, no. 7, pp. 522–522, Jun. 2005, doi: 10.1038/nrmicro1192.

[86] P. S. Padayatti *et al.*, “Structural Insights into the Lipid A Transport Pathway in MsbA,” *Structure*, vol. 27, no. 7, pp. 1114-1123.e3, Jul. 2019, doi: 10.1016/J.STR.2019.04.007.

[87] C. L. Reyes, A. Ward, J. Yu, and G. Chang, “The structures of MsbA: Insight into ABC transporter-mediated multidrug efflux,” 2005, doi: 10.1016/j.febslet.2005.11.033.

[88] S. Okuda, D. J. Sherman, T. J. Silhavy, N. Ruiz, and D. Kahne, “Lipopolysaccharide transport and assembly at the outer membrane: the PEZ model,” *Nature Publishing Group*, 2016, doi: 10.1038/nrmicro.2016.25.

[89] B. W. Simpson and M. Stephen Trent, “Pushing the envelope: LPS modifications and their consequences”, doi: 10.1038/s41579-019-0201-x.

[90] W. Mi, Y. Li, S. H. Yoon, R. K. Ernst, T. Walz, and M. Liao, “Structural basis of MsbA-mediated lipopolysaccharide transport,” 2017, doi: 10.1038/nature23649.

[91] B. J. Voss and M. Stephen Trent, “LPS Transport: Flipping Out over MsbA,” *Current Biology*, vol. 28, no. 1, pp. R30–R33, Jan. 2018, doi: 10.1016/J.CUB.2017.10.067.

[92] S. C. Wilschefski, M. R. Baxter, S. Wilschefski, and S. C. Au, “Inductively Coupled Plasma Mass Spectrometry: Introduction to Analytical Aspects,” *Clin Biochem Rev*, vol. 40, no. 3, pp. 2019–115, doi: 10.33176/AACB-19-00024.

[93] A. Mehr, F. Henneberg, A. Chari, D. Görlich, and T. Huyton, “The copper(II)-binding tripeptide GHK, a valuable crystallization and phasing tag for macromolecular crystallography,” *urn:issn:2059-7983*, vol. 76, no. 12, pp. 1222–1232, Nov. 2020, doi: 10.1107/S2059798320013741.

[94] A. Punjani, J. L. Rubinstein, D. J. Fleet, and M. A. Brubaker, “cryosPArc: algorithms for rapid unsupervised cryo-em structure determination,” 2017, doi: 10.1038/nmeth.4169.

[95] T. D. Goddard *et al.*, “UCSF ChimeraX: Meeting modern challenges in visualization and analysis,” *Protein Sci*, vol. 27, no. 1, pp. 14–25, Jan. 2018, doi: 10.1002/PRO.3235.

[96] V. Y. Lunin *et al.*, “Theory and Techniques 12. Binary Integer Programming and its Use for Envelope Determination Bulk Solvent Correction for Yet Unsolved Structures Search of the Optimal Strategy for Refinement of Atomic Models Metal Coordination Groups in Proteins: Some Comments on Geometry, Constitution and B-values”, Accessed: Jul. 09, 2022. [Online]. Available: http://www.iucr.org

[97] J. F. Hancock, “Ras proteins: different signals from different locations,” *Nature Reviews Molecular Cell Biology 2003 4:5*, vol. 4, no. 5, pp. 373–385, May 2003, doi: 10.1038/nrm1105.

[98] D. K. Simanshu, D. v. Nissley, and F. McCormick, “RAS Proteins and Their Regulators in Human Disease,” *Cell*, vol. 170, no. 1, p. 17, Jun. 2017, doi: 10.1016/J.CELL.2017.06.009.

[99] G. A. Repasky, A. D. Cox, A. B. Hanker, N. Mitin, and C. J. Der, “Role of r-ras in cell growth,” *Handbook of Cell Signaling, 2/e*, vol. 2, pp. 1753–1762, 2010, doi: 10.1016/B978-0-12-374145-5.00214-X.

[100] P. Crespo and J. León, “Review Ras proteins in the control of the cell cycle and cell differentiation,” 2000.

[101] P. Kurada and K. White, “Ras Promotes Cell Survival in Drosophila by Downregulating hid Expression,” *Cell*, vol. 95, no. 3, pp. 319–329, Oct. 1998, doi: 10.1016/S0092-8674(00)81764-X.

[102] A. Bonni, A. Brunet, A. E. West, S. R. Datta, M. A. Takasu, and M. E. Greenberg, “Cell survival promoted by the Ras-MAPK signaling pathway by transcription-dependent and -independent mechanisms,” *Science (1979)*, vol. 286, no. 5443, pp. 1358–1362, Nov. 1999, doi: 10.1126/SCIENCE.286.5443.1358/ASSET/78664536-7443-40CA-AD94-6037EBB1B248/ASSETS/GRAPHIC/SE449799304A.JPEG.

[103] C. Parikh, R. Subrahmanyam, and R. Ren, “Oncogenic NRAS, KRAS, and HRAS Exhibit Different Leukemogenic Potentials in Mice”, doi: 10.1158/0008-5472.CAN-07-0778.

[104] A. G. Stephen, D. Esposito, R. G. Bagni, and F. McCormick, “Dragging ras back in the ring,” *Cancer Cell*, vol. 25, no. 3, pp. 272–281, Mar. 2014, doi: 10.1016/J.CCR.2014.02.017.

[105] “M Monographs”, doi: 10.1177/1947601911411084.

[106] A. Hennig, R. Markwart, M. A. Esparza-Franco, G. Ladds, and I. Rubio, “Ras activation revisited: Role of GEF and GAP systems,” *Biological Chemistry*, vol. 396, no. 8, pp. 831–848, Aug. 2015, doi: 10.1515/HSZ-2014-0257/ASSET/GRAPHIC/J_HSZ-2014-0257_FIG_004.JPG.

[107] L. A. Quilliam, J. F. Rebhun, and A. F. Castro, “A growing family of guanine nucleotide exchange factors is responsible for activation of ras-family GTPases,” *Progress in Nucleic Acid Research and Molecular Biology*, vol. 71, pp. 391–444, Jan. 2002, doi: 10.1016/S0079-6603(02)71047-7.

[108] D. Bar-Sagi, “The Sos (Son of sevenless) protein,” *Trends in Endocrinology & Metabolism*, vol. 5, no. 4, pp. 165–169, May 1994, doi: 10.1016/1043-2760(94)90014-0.

[109] P. A. Boriack-Sjodin, S. M. Margarit, D. Bar-Sagi, and J. Kuriyan, "The structural basis of the activation of Ras by Sos," *Nature 1998 394:6691*, vol. 394, no. 6691, pp. 337–343, Jul. 1998, doi: 10.1038/28548.

[110] B. E. Hall, D. Bar-Sagi, and N. Nassar, "The structural basis for the transition from Ras-GTP to Ras-GDP," *Proc Natl Acad Sci U S A*, vol. 99, no. 19, p. 12138, Sep. 2002, doi: 10.1073/PNAS.192453199.

[111] H. Sondermann, S. M. Soisson, S. Boykevisch, S. S. Yang, D. Bar-Sagi, and J. Kuriyan, "Structural Analysis of Autoinhibition in the Ras Activator Son of Sevenless," *Cell*, vol. 119, no. 3, pp. 393–405, Oct. 2004, doi: 10.1016/J.CELL.2004.10.005.

[112] S. Xiang, W. Bai, G. Bepler, and X. Zhang, "Activation of Ras by Post-Translational Modifications," *Conquering RAS: From Biology to Cancer Therapy*, pp. 97–118, 2017, doi: 10.1016/B978-0-12-803505-4.00006-0.

[113] Y. C. Li *et al.*, "Analysis of RAS protein interactions in living cells reveals a mechanism for pan-RAS depletion by membrane-targeted RAS binders," *Proc Natl Acad Sci U S A*, vol. 117, no. 22, Jun. 2020, doi: 10.1073/PNAS.2000848117/SUPPL_FILE/PNAS.2000848117.SAPP.PDF.

[114] B. E. Hall, S. S. Yang, P. A. Boriack-Sjodin, J. Kuriyan, and D. Bar-Sagi, "Structure-based mutagenesis reveals distinct functions for Ras switch 1 and switch 2 in Sos-catalyzed guanine nucleotide exchange," *J Biol Chem*, vol. 276, no. 29, pp. 27629–27637, Jul. 2001, doi: 10.1074/JBC.M101727200.

[115] J. A. Mindell and N. Grigorieff, "Accurate determination of local defocus and specimen tilt in electron microscopy," *Journal of Structural Biology*, vol. 142, no. 3, pp. 334–347, Jun. 2003, doi: 10.1016/S1047-8477(03)00069-8.

[116] R. M. Glaeser, "Proteins, interfaces, and cryo-EM grids," *Current Opinion in Colloid and Interface Science*, vol. 34, pp. 1–8, Mar. 2018, doi: 10.1016/J.COCIS.2017.12.009.

[117] V. P. Dandey *et al.*, "Spotiton: New features and applications," *Journal of Structural Biology*, vol. 202, no. 2, pp. 161–169, May 2018, doi: 10.1016/J.JSB.2018.01.002.

[118] R. B. G. Ravelli *et al.*, "Cryo-EM structures from sub-nl volumes using pin-printing and jet vitrification," *Nature Communications 2020 11:1*, vol. 11, no. 1, pp. 1–9, May 2020, doi: 10.1038/s41467-020-16392-5.

[119] C. Schmidli *et al.*, "Miniaturized Sample Preparation for Transmission Electron Microscopy," *JoVE (Journal of Visualized Experiments)*, vol. 2018, no. 137, p. e57310, Jul. 2018, doi: 10.3791/57310.

[120] C. v. Iancu *et al.*, "Electron cryotomography sample preparation using the Vitrobot," *Nature Protocols*, vol. 1, no. 6, pp. 2813–2819, Jan. 2007, doi: 10.1038/nprot.2006.432.

[121] M. Wu and G. C. Lander, “Present and Emerging Methodologies in Cryo-EM Single-Particle Analysis,” *Biophysical Journal*, vol. 119, no. 7, pp. 1281–1289, Oct. 2020, doi: 10.1016/J.BPJ.2020.08.027.

[122] X. Fan *et al.*, “Single particle cryo-EM reconstruction of 52 kDa streptavidin at 3.2 Angstrom resolution”, doi: 10.1038/s41467-019-10368-w.

[123] M. A. Herzik, M. Wu, and G. C. Lander, “High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM,” *Nature Communications 2019 10:1*, vol. 10, no. 1, pp. 1–9, Mar. 2019, doi: 10.1038/s41467-019-08991-8.

[124] M. J. Peet, R. Henderson, and C. J. Russo, “The energy dependence of contrast and damage in electron cryomicroscopy of biological molecules,” *Ultramicroscopy*, vol. 203, pp. 125–131, Aug. 2019, doi: 10.1016/J.ULTRAMIC.2019.02.007.

[125] P. R. Baldwin *et al.*, “Big Data in CryoEM: Automated collection, processing and accessibility of EM data,” *Curr Opin Microbiol*, vol. 43, p. 1, Jun. 2018, doi: 10.1016/J.MIB.2017.10.005.

[126] D. N. Mastronarde, “Automated electron microscope tomography using robust prediction of specimen movements,” 2005, doi: 10.1016/j.jsb.2005.07.007.

[127] K. Maruthi, M. Kopylov, and B. Carragher, “Automating Decision Making in the Cryo-EM Pre-processing Pipeline,” *Structure*, vol. 28, no. 7, pp. 727–729, Jul. 2020, doi: 10.1016/J.STR.2020.06.004.

[128] N. Biyani *et al.*, “Focus: The interface between data collection and data processing in cryo-EM,” *Journal of Structural Biology*, vol. 198, no. 2, pp. 124–133, May 2017, doi: 10.1016/J.JSB.2017.03.007.

[129] A. Doerr, “Cryo-electron tomography,” 2017. doi: 10.1038/nmeth.4115.

[130] H. Cao and J. Skolnick, “Time-resolved x-ray crystallography capture of a slow reaction tetrahydrofolate intermediate,” *Structural Dynamics*, vol. 6, no. 2, p. 024701, Mar. 2019, doi: 10.1063/1.5086436.

[131] “Watching proteins function with time-resolved x-ray crystallography,” 2017, doi: 10.1088/1361-6463/aa7d32.

[132] G.-W. Wei, “Protein structure prediction beyond AlphaFold,” *Nature Machine Intelligence 2019 1:8*, vol. 1, no. 8, pp. 336–337, Aug. 2019, doi: 10.1038/s42256-019-0086-4.

[133] K. M. Ruff and R. v. Pappu, “AlphaFold and Implications for Intrinsically Disordered Proteins,” *Journal of Molecular Biology*, vol. 433, no. 20, p. 167208, Oct. 2021, doi: 10.1016/J.JMB.2021.167208.