Mechanistic Studies of ArsA ATPase and ArsB Transporter of the Bacterial Arsenite Efflux System

Thesis by
Shivansh Mahajan

In Partial Fulfillment of the Requirements for the
Degree of
Doctorate of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2026
Defended August 28, 2025

Shivansh Mahajan
ORCID: [0000-0002-3041-7988]

I dedicate this thesis in the memory of my friend, Rutuj Gavankar and my grandfather (Nanu), Arjun Kumar Mahajan

ACKNOWLEDGMENTS

It was sometime in 7th grade that I started to get intrigued by the mysteries of science and nature, particularly chemistry. It wasn't very long until I new that if I ever had to pick a profession for a career, it wouldn't be a doctor or an engineer but a *chemist*. While chemistry was fun, I realized in high school that the chemistry of life was even more intricate and complex, and that biology is best studied, and is even more fun, through the lens of chemistry. An enthusiasm for fundamental scientific research in chemistry/biology prompted my decision to go the United States in hope of benefiting from its widespread scientific opportunities for basic research, which were still scarce in India. While studying chemistry at IUPUI, I was introduced to enzyme chemistry in my undergraduate research lab. As I was applying for graduate schools, I had a general idea that during my PhD I would like to learn and combine diverse chemical and biophysical approaches to understand the chemistry of proteins and enzymes. Although much to my surprise, when I got admitted to Caltech Biochemistry and Molecular Biophysics graduate program, I was thrilled to obtain my scientific training under the mentorship of some of the most brilliant and pioneering scientists in the world.

Five years later, I feel very fortunate to have had the support and companionship of a village of amazingly talented, supportive and wonderful people during the course of graduate school. First of all, I am immensely grateful for my PhD advisors, Bil Clemons and Doug Rees, and that I had the opportunity to pursue my graduate studies under the mentorship of two excellent biochemists. I also want to extend my gratitude to Dianne Newman, who played an instrumental role in helping me navigate labs during my first year at Caltech. Wanting to study enzyme mechanisms using the tools of structural biology while also wanting to explore physiology and bioenergetics in bacteria by probing molecular mechanisms of those enzymes, I was offered an opportunity by Bil, Doug and Dianne to write a pilot grant to start a new PhD project to study bacterial arsenite efflux pump ArsAB, that would combine the interests of the three labs — enzyme mechanisms, membrane protein structural biology and microbiology. Although the major portion of my thesis work was primarily focused on structural biology and mechanisms of the ArsA ATPase and the ArsB transporter, this project gave me a platform to think synergistically about understanding and elucidating the molecular mechanisms of bacterial arsenite resistance. I am very grateful to Bil and Doug for showing trust in me, and for their

patience and support as I navigated through this project while also learning and developing skills in structural biology. Although both Bil and Doug are structural biologists at heart, their guidance provided me the space to explore my own scientific interests and ideas. Moreover, having two PhD advisors often exposed me to distinct perspectives for navigating through a scientific problem.

I credit Bil for identifying and guiding me through the ArsA project. Although he always urged me early on to keep trying cryo-EM to determine ArsA structures, he supported me when I insisted on doing crystallography while I still struggled through the hurdles of learning cryo-EM. Whether it was my first diffraction pattern or my first low-resolution cryo-EM map of ArsA, Bil's excitement was worth all the wait. I have learned from Bil to question traditional approaches to address a problem, and identify new approaches to solve a problem more efficiently either by learning a new skill or through collaborations. Moreover, I'm very grateful to Bil for giving me the opportunity to work on the 'IWA ATPases' review article as this exercise enabled me to familiarize with the P-loop NTPase literature that served as the framework for understanding the mechanism of ArsA later on.

I am grateful for all the invaluable advice I have received from Doug, both about science and life/career. One very powerful advice that Doug gave me when I had just started to screen for crystallization conditions for ArsA was to "fail faster", which, in fact, seems to apply to any challenging project. As my crystallography phase in the lab was nothing short of challenging, Doug's profound expertise in X-ray diffraction and symmetry was indispensable as I worked toward understanding translational pseudosymmetry in my ArsA crystals. I have learned from Doug to critically evaluate a result, and be skeptical and cautious when drawing inferences, a strategy that enhances the quality of science. I often found it very helpful to discuss with Doug about important experiments that I wasn't confident about or critical decisions such as picking a postdoc lab. His advice and positive outlook has helped me make some important decisions over the last few years.

I want to especially acknowledge Jim Howard, with whom I never had the pleasure of meeting unfortunately. When I was new in the group, both Bil and Doug used to talk about how much Jim enjoyed talking about the mechanisms of P-loop NTPases and the so-called 'IWA ATPases'. This made me excited about meeting and learning from Jim. However, my first and only interaction with Jim was an email exchange after my first group meeting when he asked me what I meant by 'OG' when I jokingly referred to the 'signature lysine residue' as the 'OG lysine'. He showed interest in

the mechanism of ArsA and I am sure he would have been very excited to see our findings. I am also very thankful to my other committee members, Shu-ou Shan and Steve Mayo, for asking challenging and insightful questions during my candidacy exam and committee meeting. I find both of them to be very inspirational scientists.

I will be forever grateful to my undergraduate advisor from IUPUI — Mike McLeish — for mentoring me and shaping my research interests early on. I basically owe my interest in enzyme catalysis and mechanisms to Mike. Letting me join his group as a freshman with little idea about what biochemistry was, he practically taught me *how to do biochemistry*. Mike is a great medicinal chemist, enzymologist, kineticist and research mentor. When you ask him a question in lab, after first making a sarcastic comment, he would basically ask you more questions in return and eventually help you answer your question yourself. This ensured critical thinking and thoughtful scientific inquiry. Mike exposed me to numerous opportunities including taking me to enzymology conferences and introducing me to potential graduate school mentors. Mike understood the challenges that international student face, and often went above and beyond to vouch for me. I couldn't have asked for a better undergraduate advisor who introduced me to *scientific research* and *enzymes*, and continues to mentor me to date. From my time in the McLeish group, I would like to thank Alex Latta as well, the PhD student in the group at the time, who introduced me to X-ray crystallography. I learned a lot from our discussions on crystallography and reaction mechanisms. I am grateful for my SURF mentors at Mayo Clinic from Summer 2018 — Evette Radisky and Maryam Raeeszadeh-Sarmazdeh; my crystallization efforts from this opportunity resulted in my first publication on engineering selective protein inhibitors of matrix metalloproteinases in *JBC*. I also want to acknowledge the mentorship of my high school biology teacher — Savita Sharma — whose enthusiasm for teaching and understanding nature inspires me to date.

In addition to my mentors during graduate school, I have been very lucky to know and learn from many talented folks in the Clemons and Rees groups. One of the non-scientific lessons I have learned in graduate school is the often understated significance of having good, helpful and enabling people around you. Science will always be hard, but what makes the pursuit of hard scientific problems enthralling and fun is the relationships you foster along the way. I would first like to thank Evelyn; using her expertise in *picking squares with good ice thickness* and data processing, Evelyn assisted me in solving my first cryo-EM structure of ArsA in December 2023. I will remember her as a very competent and efficient structural

biologist, and a good friend with whom talking science and ranting about random stuff was always fun. I am grateful for Victor for his friendship, mentorship and his expertise in working with challenging cryo-EM data. His hard work, persistence, and the consequent success in solving challenging cryo-EM structures by learning and adopting diverse data processing approaches has benefited many members of the group, including myself. In addition, Victor is a very jolly person to be around, who has an ability to always uplift your mood if you're having a bad day. I am glad to have had him as a labmate, friend, roommate and the occasional workout buddy. Jacob's arrival in the lab in 2024 was particularly exciting. With his background in physical chemistry, Jacob brought a new perspective of thinking about many structural and biological problems in the group. It was nice to have a chemist in the group and someone with whom I shared the enthusiasm for enzyme mechanisms. Our prolonged discussions on mechanisms, structural biology, science and academia in general, have enabled me to become a more thoughtful scientist over the past few months. One advice that Jacob once gave me that I find to be very powerful for exercising effective scientific inquiry is to ask "what do you really want to know?". I find his arrival in the lab very timely as he played an important role in advising me as I reflected on my academic goals and applied for postdocs over the last few months. Jacob is a true academic, a great mentor and friend, and I'm excited to witness and learn from his scientific endeavors. I want to thank Xiang for not only being an awesome and a very resourceful benchmate, but also for his pearls of wisdom on everything ranging from metalloproteins to cryo-EM. I also want to extend my gratitude to Ailiena for being an amazing mentor and friend. Her knowledge and passion for X-ray crystallography were instrumental in helping me collect high quality data at the beamline and solve the ArsA crystal structure. I feel fortunate I had the opportunity to collaborate on writing the IWA ATPases review article with her, as our discussions on the mechanistic features of these ATPases set the foundation for my thesis. I want to thank Yusrah for her friendship during the early years of graduate school. I am thankful for Karen, who in addition to being my rotation mentor in the group, always ensured with her lively attitude that the lab was a fun and inclusive place for everyone. I'd also like to thank other members of the Clemons and Rees groups, both past and present, with whom I overlapped and learned from, including Juliet, Beccah, Alex, Grace, Roujon, Angham, Conner, Isabel, Yuanbo, Storm, Vishal and Jessica. I got a lot to learn from the wisdom of Jens while I was still doing crystallography. Due to his support and invaluable knowledge, I have been able to grow as a crystallographer over the last few years.

I want to thank Allen and Ann for their hard work keeping the lab up and running. Also, I'm thankful for Welison for all the support with our technological and data management needs. Moreover, I feel fortunate to have the opportunity to be a SURF mentor to Lily, who has demonstrated scientific excellence at and beyond Caltech. I would also like to thank my collaborators — Tim, Ashley and Kemal. Their technical and intellectual contributions have been invaluable for this thesis.

My time at Caltech and in Pasadena was filled with many memorable experiences and friendships that will last forever. I am thankful for Tony, my friend and roommate of two years, and our exploration of coffee, Mexican food, and hikes around LA. I am thankful for RJ, my first friend from Delhi at Caltech. I am inspired by his dedication to stimulated Raman spectroscopy, Peet's coffee, and Erewhon. I also appreciate Alec, Nik, Pratyush, Eva, Andres, Suraj, Pranav and Aditya for their friendship and camaraderie. I will miss the fun times during the year I stayed at the *Blue Cordova House* along with Victor, Tom, and Adjeoda. I will also miss the surfing trips with the Caltech Surf Club that gave me a chance to experience the surf culture of Southern California, despite my injured back. Furthermore, I am thankful for my friends from IUPUI and India, majority of whom I had to struggle explaining what I do and why I do so as a scientist. Despite the distance and time zones, these friends have been there through thick and thin. I'm thankful for my IUPUI friends — Krishay, Yavat, Rutuj, Anna and Pranesh. I want to express my gratitude for Rutuj, who was like a brother to me, and whose scientific curiosity and inquisitiveness will always inspire me to question everything and aim high. I am also grateful for my friends from back home — Dinesh, Kshitij, Dikshita, Sushrut, Abhishek and Varun. These people have helped me over the last few years to maintain my sanity, which in turn helped me keep things in perspective.

Most importantly, I wouldn't be on the path I am today without the hard work, love and faith of my parents. With absolutely no frame of reference for what a career in scientific research might look like, they supported me for my interests and choices, and always encouraged me to pursue my dreams. My mother always emphasized the value of education and her remarkable teaching skills likely influenced my interest in learning and doing science. My father exemplified hard work and dedication toward one's work, virtues that I always strive to adopt in my life. I am very grateful for my sister Saumya for her love and support. I feel lucky to have a very loving family that has always been by my side in my scientific journey.

TABLE OF CONTENTS

ABSTRACT

Arsenic is a notorious metalloid that contaminates the groundwater in several regions worldwide. The trivalent state of arsenic — arsenite (As^{III}) — is the abundant species of arsenic under reducing conditions of subsurface waters, and is readily mobilized in aqueous environments. This exposes organisms to toxic concentrations of the metalloid. As^{III} is particularly toxic to living systems due to its ability to form stable polar covalent bonds with exposed thiol groups, thus disrupting protein structure and function. Arsenic detoxification systems such as efflux pumps, exist in most organisms that confer tolerance to toxic concentrations of arsenicals found in their environment. The *ars* operon in many bacteria and some archaea confers resistance to As^{III} via ArsB, an integral membrane transporter and ArsA, a cytoplasmic P-loop ATPase. These proteins, collectively referred to as the 'ArsAB efflux pump', facilitate toxic As^{III} export in an ATP-dependent manner. In addition, ArsB can operate by itself as a proton-coupled secondary transport and confer intermediate levels of As^{III} resistance. The mechanisms of this dual mode of As^{III} efflux are poorly understood, particularly the molecular events associated with the capture of As^{III} from the cytoplasm by ArsA, its transfer to ArsB and subsequent vectorial transport across the membrane. Apart from understanding fundamental mechanisms of toxic metalloid detoxification in living systems, molecular-level investigations of As^{III} efflux systems are of broad biotechnological interest for their potential to inform robust and sustainable bioremediation strategies. In this thesis, we elucidate the mechanism of ArsA ATPase and ArsB transporter using structural approaches. We characterized the nucleotide hydrolysis mechanism of ArsA by single particle cryogenic electron microscopy (cryo-EM), outlining various conformational states of the ATPase that modulate the nucleotide-dependent capture and delivery of As^{III} for efflux. We show that this mechanism is consistent with the general mechanistic framework of the Intradimeric Walker A (IWA) family of ATPases. Furthermore, overexpression and purification of the membrane transporter ArsB enabled characterization of the first structure of ArsB by cryo-EM, in both apo and As^{III}-bound states. Lastly, we show that ArsB enhances steady-state ATPase activity of ArsA, indicating a direct interaction between the two components of the efflux pump. Computational modeling gives some insights into a putative ArsAB interaction interface. While several mechanistic questions remain, the findings reported in this thesis together constitute a foundation for future mechanistic elucidation of the ArsAB efflux system.

PUBLISHED CONTENT AND CONTRIBUTIONS

(1) Mahajan, S.; Pall, A.E.; Li, Y.E.; Stemmler, T. L.; Rees, D.C.; Clemons Jr., W.M. Nucleotide- and metalloid-driven conformational changes in the arsenite efflux ATPase ArsA. *Proceedings of the National Academy of Sciences*. 122 (35) e2506440122. (2025). https://doi.org/10.1073/pnas.2506440122.
S.M. conceived the study, designed and performed experiments, analyzed the data, and wrote and edited the manuscript.

(2) Maggiolo, A. O.*; Mahajan, S.*; Rees, D. C.; Clemons, W. M. Intradimeric Walker A ATPases: Conserved features of a functionally diverse family. *Journal of Molecular Biology.*, 434-11, 167965. (2023). https://doi.org/10.1016/j.jmb.2023.167965
*Indicates equal contributions.
S.M. performed structural analysis, wrote and edited the review article.

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

Metals play many essential roles in biology in the form of enzyme cofactors, redox mediators, signaling factors, gene regulators, and so on [1, 2]. Despite their biological indispensability, metal homeostasis is critical for all living systems, as these metals themselves can become toxic to cells beyond a concentration threshold. Homeostasis of metals such as Fe, Ca, Cu, Zn and Mn that perform some of these essential roles is accomplished through membrane efflux pumps [3–5]. However, other heavy metals such as Cd, Pb and Hg that are frequently encountered by organisms in their external environment are highly toxic. Nature has evolved a myriad of detoxification mechanisms to counter the effects of their toxicity [6–8] . Naturally occurring metalloids of group 15 of the periodic table (pnictogens) — including arsenic (As) and antimony (Sb) — are part of the latter group and are toxic to almost all living systems while rarely serving an essential biological role [9]. Therefore, detoxification of these toxic metalloids is critical for organisms, and elucidation of the molecular mechanisms underlying biological detoxification promises broad applications for heavy metal bioremediation and drug design. Moreover, the unusual chemical properties of arsenic and antimony, intermediate between those of metals and non-metals, lead to unique modes of interactions with biological molecules, including proteins, that are still poorly understood.

1.1 Arsenic chemistry and toxicity

Arsenic ($[Ar]3d^{10}4s^{2}4p^{3}$) is a metalloid, so it exhibits chemical properties of both metals and non-metals. It has an electronegativity similar to the lighter pnictogen, phosphorus. Common oxidation states of arsenic are +3, +5 and -3. Trivalent (As^{III}) and pentavalent (As^{V}) oxidation states commonly exist as oxo-acids under aqueous conditions, known as arsenite/arsenous acid and arsenate/arsenic acid respectively. Arsenite has three pK_a values of 9.2, 12.1 and 13.4 [10, 11]. Arsenate as well has three pK_a values of 2.3, 6.8 and 11.5. As a result, under physiological conditions, arsenite exists as an 'uncharged' species, $As(OH)_3$ and arsenate exists as a 'dibasic anion', $AsO_3^{2-}(OH)$. As^{III} is sometimes referred to as a 'soft metal' due to its relatively low charge density and high polarizability that make it a 'soft Lewis acid' [12]. This property makes it susceptible to form stable covalent bonds with 'soft

Lewis bases' such as thiolates, that can spontaneously exchange with the hydroxyl ligands in $As(OH)_3$ to form As-S bonds.

Arsenic is widespread in the Earth's crust with a natural abundance of 1.5-2 ppm, and is found in over 200 minerals [13]. Inorganic arsenic is primarily found as a constituent of sulfide minerals, most commonly arsenopyrite (FeAsS). Biogeochemical cycling facilitates the release of inorganic arsenic from these minerals into surface waters and soil. Under aqueous conditions in nature, inorganic arsenic mostly exists as arsenate ($AsO_3^{2-}(OH)$) or arsenite ($As(OH)_3$). As a consequence, organisms in these environments are constantly exposed to toxic arsenic. Anthropogenic factors such as industrial activities and mining, further contribute to the mobilization of arsenic compounds (arsenicals) to distant locations, which contaminates groundwater and soil. As^V confers its toxic effects by mimicking inorganic phosphate ($PO_3^{2-}OH$). Under oxic conditions, due to lower solubility and tendency to adsorb to sulfide minerals [14], As^V remains out of solution. Under anoxic conditions of subsurface waters, reductive dissolution of AsV leads to the release of As^{III} into the water. In fact, arsenic toxicity is typically attributed to As^{III} due to relatively high solubility and mobility in aqueous environments [15]. The trihydroxyl form of As^{III} enters the cell via aquaglyceroporins such as GlpF in *E. coli*, which also imports small solutes like glycerol [16]. Subsequently, As^{III} exerts its toxic effects by its soft metallic character by tightly binding to exposed thiol groups (cysteine residues) in proteins; consequently, disrupting enzyme function, protein folding, respiration and redox homeostasis. Hazardous effects of arsenic exposure in humans, both acute and chronic, have been widely recognized. Groundwater concentrations of arsenic in eastern parts of India and Bangladesh have been reported to exceed 200 µg/L [17], which far exceeds the permissive limit of 10 µg/L of arsenic suggested by the World Health Organization (WHO). As groundwater serves and sustains several populations around the globe, its contamination with arsenic poses a major public health concern. Therefore, conceiving efficient arsenic bioremediation strategies for decontamination of such environments is of widespread interest.

1.2 The *ars* operon — arsenic detoxification system in bacteria

Owing to the widespread occurrence and detrimental impact of arsenic, organisms have evolved diverse and complex strategies for arsenic resistance. Two major modes of detoxification are common — chemical biotransformation of relatively more toxic arsenite to less toxic derivatives such as arsenate or organoarsenicals, and physical extrusion of arsenite from the cell via efflux pumps [18]. In humans,

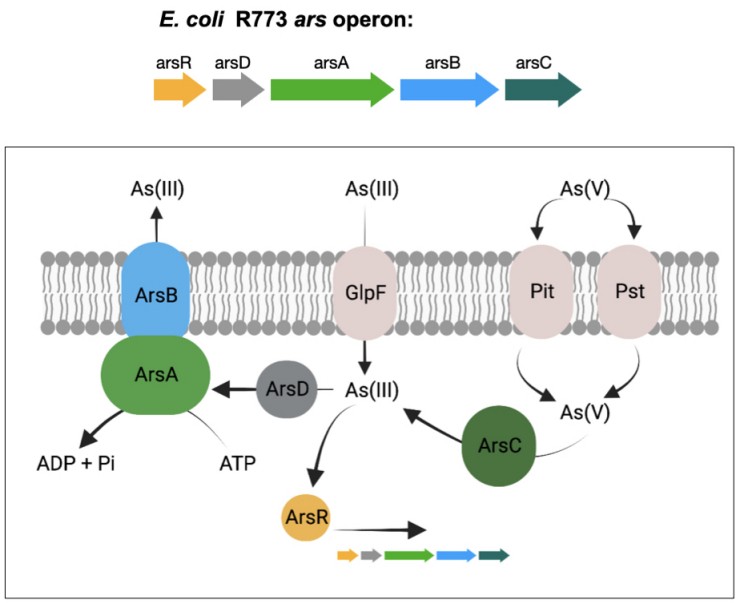


Figure 1.1: **Arsenic uptake and detoxification system encoded by the *ars* operon of *E. coli* R773 plasmid.** Uptake of As^{III} and AsV is facilitated by GlpF and phosphate transporters (Pit and Pst), respectively. Detoxification of arsenic from the cytoplasm is facilitates by proteins encoded by the *arsRDABC* operon.

arsenic exposure is managed by efflux via multidrug resistance proteins (MRP), and in some individuals, by methylation of As^{III} via a methyltransferase encoded by the AS3MT gene [19]. In many bacteria and some archeae, arsenic resistance is encoded on the *ars* operon (Fig. 1.1) [18]. Although resistance to As^{III}, As^{V} and antimonite (Sb^{III}) via efflux was first discovered through plasmid-encoded genes in *Staphylococcus aureus* in 1968 and in *E. coli* in 1973 [20], the sequences of the genes in the *ars* operon of the *E. coli* plasmid R773 were not reported until 1986 [21]. The chromosomal *ars* operon of *E. coli* was identified much later in 1994 [22].

The *ars* operon of the R773 plasmid constitutes five genes — arsR, a As(III)-responsive DNA repressor that regulates the operon; arsB, an integral membrane protein that transports As^{III}; arsA, a cytoplasmic ATPase that couples ATP hydrolysis to As^{III} efflux by ArsB; arsD, a As^{III}-metallochaperone that delivers As^{III} to ArsA; and ArsC, a cytoplasmic reductase that reduces As^{V} to As^{III} for efflux. As such, the R773 *ars* operon (*arsRDABC*) confers resistance to both oxidation states of arsenic. In contrast, the chromosomal *ars* operon of *E. coli* is much simpler as it only comprises the arsR, arsB and arsC genes (*arsRBC*), as do the *S. aureus* operons [23]. A strong preference for resistance to trivalent arsenic is suggested in these

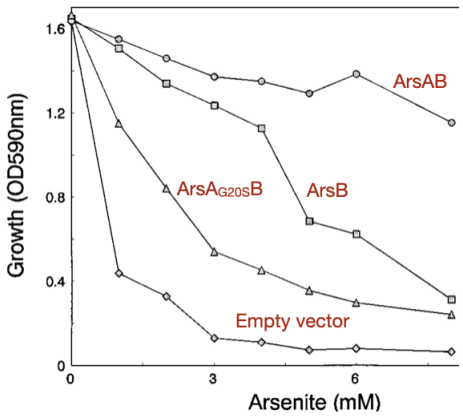


Figure 1.2: **Effect of arsenic resistance genes on the growth of *E. coli* in presence of arsenite.** [Figure adapted from Dey et al. *J. Bacteriol.* (1995)].

systems, indicating ancestral origins of the operon dating back to before the *Great Oxygen Event* when Earth had a reducing atmosphere. The chromosomal *arsRBC* operon alone confers intermediate levels of arsenic resistance, in contrast to the *arsRDABC* operon that significantly enhances the efficiency of As^{III} removal from the cell by coupling As^{III} efflux by ArsB to ATP hydrolysis by ArsA [24, 25]. Adding further complexity, the *ars* operons of some prokaryotes have an unrelated As^{III} transporter, Acr3, instead of ArsB [26]. However, despite the variability observed in the genes of the operon, the *arsRDABC* and *arsRBC* operon architectures are the most well-studied to date.

The current model for arsenite efflux by the *arsRDABC* operon is well-understood at the cellular level (Fig. 1.1). As^{III} and As^{V} enter the cell via GlpF and phosphate transporters, respectively. In the presence of As^{III}, repressor ArsR is triggered to release from the *ars* operon and enable the transcription of the *ars* genes [27]. ArsC reduces any As^{V} to As^{III} [28]. At low or intermediate As^{III} concentrations in the cell, ArsD sequesters As^{III} from the cytoplasm with high affinity and transfers the metalloid to ArsA in an ATP-dependent manner [29]. Subsequently, As^{III}-bound ArsA delivers the metalloid to ArsB in an ATP-dependent manner, and the latter then transports it across the plasma membrane [16]. Under high arsenic stress, the so-called 'ArsAB efflux pump' or ArsB alone can capture and remove As^{III}, and both modes of efflux operate synergistically to combat arsenic toxicity at varying As^{III} concentrations.

Although the biochemical roles of the individual protein of the *arsRDABC* operon are well-characterized, the details of key molecular-level events underlying the capture and transport of toxic As^{III} out of the cell remain poorly understood. At the heart of this mechanism is the 'ArsAB efflux pump'. Initial studies on this system have highlighted mechanistic similarities between ArsAB and ABC transporters [30, 31]. However, the nucleotide-binding domains of ABC transporters constitutes a distinct fold comapred to the ATPase subunit in ArsAB, and the latter is beneficial but not obligatory for ArsAB to remove As^{III}. As ArsB is sufficient to carry out As^{III} efflux, how ArsAB captures and facilitates As^{III} transport to confer higher As^{III} tolerance to bacteria, is largely unexplored. Importantly, ArsB has not been biochemically or structurally characterized, precluding an understanding of the mechanism of the transporter itself. In the case of ArsA, while its nucleotide hydrolysis mechanism has been studied quite extensively and X-ray crystal structures are available [32–34], these data together do not explain how the ATPase relays As^{III} between ArsD and ArsB for efflux as suggested by biochemical studies [29, 35]. Moreover, the interaction between ArsA and ArsB remains elusive, raising the question whether a stable 'ATPase-coupled efflux pump' even exists. Investigation of these key mechanistic questions will not only contribute to our understanding of the arsenic detoxification machinery, but also lay the foundation for understanding, more generally, the mechanisms for toxic metal detoxification and homeostasis in biological systems.

In addition to the fundamental questions concerning the biochemistry of ArsA and ArsB, mechanistic elucidation of arsenic efflux pumps is broadly applicable from a bioengineering perspective toward designing robust and sustainable strategies for bioremediation of toxic arsenic from groundwater and contaminated environments. While chemical modifications of arsenite can alleviate its toxic effects [18], an efficient bioremediation strategy would ensure physical exclusion of these derivatives from contaminated environments. Therefore, investigations of arsenite efflux mechanisms by biological pumps such as ArsAB hold broad biotechnological potential.

1.3 Single particle cryogenic-electron microscopy

To address some of the key mechanistic questions outlined above, we adopted structural approaches, namely X-ray crystallography and cryogenic electron microscopy (cryo-EM) to investigate the ArsAB efflux pump. By the end of the 20th century, X-ray crystallography had revolutionized structural biology, broadening our understanding of biological mechanisms at atomic resolution. However, crystallization of

the molecule of interest is necessary, and this step is often the bottleneck in determining the structures using X-ray diffraction. Electron microscopy is an alternative technique which enables direct visualization of the molecule in relatively more native-like state, requires minimal sample quantity, and does not require crystallization. However, it was not until the last decade that cryogenic electron microscopy started to push the boundaries of high-resolution structural biology. Technical advancements such as the direct electron detector and particle processing algorithms dramatically accelerated cryo-EM single particle analysis [36–39], thus enabling atomic-resolution structures approaching those determined using X-ray crystallography [40, 41]. As the state-of-the-art cryo-EM is constantly evolving structural biology, in this section we discuss some fundamental concepts underlying single particle cryo-EM, related advancements and some challenges.

1.3.1 Fundamentals of cryogenic electron microscopy

While optical microscopy utilizes visible light to visualize microscopic specimens, electron microscopy exploits the energy of high-speed electrons to visualize specimens approaching atomic resolution (Angstroms, Å). The underlying principle of electron microscopy is based on the concept that electrons propagating at very high energies have a de Broglie wavelength, λ, which is related to the operating voltage, V of the microscope by the relationship:

$$\lambda = \frac{hc}{\sqrt{(eV)^2 + 2m_e c^2 eV}} \tag{1.1}$$

where h is the Planck's constant, c is the speed of light, e is the charge of the electron and m_e is the resting mass of the electron. This relationship accounts for relativistic effects as the speed of the electron approaches that of light. As the operating voltage accelerates electrons to energies on the order of hundreds of keV, electron microscopes are capable of achieving very high resolution images of the specimen. Transmission electron microscopy (TEM) is capable of visualizing subcellular structures and biomacromolecular specimens such as proteins. While most electrons transmit through the sample and are detected by the detector, a fraction of the electrons are elastically scattered by the specimen. Electron scattering is a result of interaction between the incident electron and the Coulomb potential of the individual atoms in the specimen [42]. As a result, a Coulomb or electrostatic potential (ESP) map of the specimen is obtained that reveals the positions of the scattering atoms.

The amplitude of electron scattering of an atom is influenced by the atomic number of the atom (Z) and the charge state, that together influence the Coulomb potential of individual atoms. Heavier and more dense atoms scatter electrons more strongly due to larger scattering cross-sections. Interestingly, the charge of the atom in question affects electron scattering quite differently from X-ray scattering, since X-ray scattering factor depends on the number of electrons in an atom while electron scattering factor depends on Coulomb potentials; hence, cryo-EM maps are more sensitive to atomic charges. Positively charged ions have very high scattering amplitudes at low scattering angles, whereas negatively charged ions can even have negative scattering amplitude at low scattering angles as for O^- [43]. Neutral atoms tend to have low dependence of electron scattering factors on the scattering angle, similar to X-ray scattering [43]. The relationship between electron and X-ray scattering factors is described by the Mott-Bethe equation:

$$f_e(s) = K \cdot \frac{Z - f_x(s)}{s^2} \tag{1.2}$$

where $f_e(s)$ and $f_x(s)$ are electron and X-ray scattering factor, respectively, Z is the atom number, s is the scattering vector magnitude that is proportional to the scattering angle, and K is a proportionality constant. The more positive the charge on an atom, the larger is the value of the Coulomb density for that atom at lower resolutions. Therefore, an EM density map can be exploited for assessment of electrostatic effects inside proteins, especially identification of metal ions that play crucial roles in protein structure and function [44, 45].

Other factors such as occupancy and temperature factors (B-factor) also contribute to the appearance of the Coulomb potential maps. In addition, the signal from scattered electrons can be dampened from inelastic scattering due to factors such as susceptibility to radiation damage and sample thickness.

1.3.2 Single particle analysis

For cryo-EM single particle analysis of proteins and protein complexes, preparation of the sample is a very important consideration to ensure high-quality reconstruction. The sample needs to be embedded in a frozen yet hydrated state to visualize its high-resolution features that resemble the native state of the protein. This is achieved by capturing the sample in a thin layer of vitreous ice by rapid plunge-freezing [46, 47]. The goal is to trap molecules of the specimen in several random orientations under the vitreous ice layer that is just thick enough to support the particles and ensure sufficient electron scattering. An ideal sample is the one that is not limited

by image contrast and maximizes the signal-to-noise ratio (SNR) of several random orientations of the particle, thus affecting the resolution of reconstruction. Traditional defocus-based cryo-EM involves collecting images over a range of defocus positions to include images with good contrast while ensuring sufficient SNR from high-resolution features of the target. After data collection, several thousand images (micrographs) are subjected to motion correction to correct for any beam-induced motion of the particles, and dose-weighted to account for radiation damage from the electron beam. Several 2D projections of the target particles are extracted from the micrographs, averaged into 2D projections spanning various random orientations, and then reconstructed into a 3D Coulomb potential map. As single particle analysis is inherently based on particle signal averaging, sample heterogeneity, both from multiple conformational states and local dynamics, can be resolved from high quality data.

1.3.3 The 'resolution revolution' and some challenges

The 'resolution revolution' has been described as the rapid increase in structures solved by cryo-EM at near-atomic resolution over the last decade [48, 49]. Structures in the range of 2-4 Å are now routinely reported in single particle cryo-EM studies. In fact, a 1.2 Å structure of the model protein, apoferritin, has been reported as well [50]. Membrane protein structural biology has immensely benefited from single particle cryo-EM due to their prevalence in large biologically-relevant heterocomplexes and also low sample requirement of the technique [47]. Structures of G-protein coupled receptors can be routinely solved at sub-2.5 Å by cryo-EM [51]. One of the challenges that remains is the high-resolution structure determination of small proteins that have molecular weights below 100 kDa. 'Small proteins' can be challenging to support on thin vitreous ice, difficult to align due to lack of distinct features and result in low SNR due to fewer scattering atoms [52]. Strategies to obtain high-resolution cryo-EM analysis of small proteins are intended to maximize the SNR and facilitate particle alignment, often by augmenting with fiducial markers. However, this landscape is continuously evolving. Over the past few years, traditional cryo-EM methods have been shown to enable sub-4 Å structure determination of small proteins. For instance, high resolution structures of streptavidin (52 kDa) and hemoglobin (64 kDa) have been solved [53, 54]. An increasing number of high-resolution cryo-EM structure of small proteins promises immense potential to provide access to understanding biological mechanisms relating to enzyme catalysis, allosteric regulation and drug-target interactions, at the atomic resolution. Notably,

both arsenite resistance proteins, ArsA (64 kDa) and ArsB (45 kDa), fall into the category of 'small proteins'.

1.4 Chapter overview

In this thesis, we investigate the mechanisms of arsenic efflux in bacteria, with a focus on ArsA ATPase and ArsB transporter, which together can confer resistance to toxic metalloids, As^{III} and Sb^{III}, at a wide range of concentrations.

In Chapter 2, we present a *perspective* on the Intradimeric Walker A (IWA) family of ATPases, members of which, including ArsA, are implicated in diverse and complex biological functions. We discuss conserved structural and functional features across IWA ATPases that have been revealed from nearly three decades of structural characterization of nucleotide-bound states of various proteins in this family. This analysis sheds light on some unifying mechanistic themes in these enzymes and how they support the coupling of nucleotide hydrolysis to their respective substrate translocation functions. This chapter lays the foundation for the study of the nucleotide hydrolysis mechanism of ArsA, which is the theme of Chapter 3.

In Chapter 3, we probe the nucleotide hydrolysis mechanism of the IWA ATPase ArsA to elucidate how the ATPase modulates the transfer of arsenite between its partners, ArsD and ArsB, to eventually remove the metalloid from the cell. To that end, we report several high-resolution cryo-EM structures of ArsA that reveal key nucleotide- and arsenite-dependent conformational changes in the enzyme. We characterized three distinct nucleotide-bound conformations — *ATP-open*, *ADP-open* and As^{III}-bound *ATP-closed* — presenting a structural mechanism for metalloid-dependent ATP hydrolysis by ArsA. This mechanism outlines the nucleotide-dependent global conformational changes throughout the catalytic cycle of ArsA, highlighting the central role of the ATPase as it modulates capture of As^{III} from the cytoplasm via ArsD and subsequent transfer to ArsB for efflux. These structures now place ArsA in the mechanistic framework of IWA ATPases.

In Chapter 4, we examine the mechanism of arsenite transport by the membrane pump, ArsB. We discuss the optimization of expression and purification strategy that enabled structural and biochemical studies of ArsB. We report the first cryo-EM structural characterization of ArsB in apo and as well as As^{III}- and Sb^{III}-bound states, and discuss these in light of metalloid (substrate) binding and transport mechanism of ArsB. Toward the end, we discuss our efforts to probe the interaction between ArsB and ArsA to elucidate the interaction interface of a putative complex

and draw insights into the role of ArsA in arsenite efflux via ArsB. We show that ArsB stimulates the steady-state ATPase activity of ArsA *in vitro*. Computational modeling of a complex using machine learning-based models offers some insight into the interaction interface but with poor confidence. We developed a growth complementation assay in *E. coli* to map the interaction between ArsA and ArsB using site-directed mutagenesis.

Lastly, in Chapter 5, we summarize these studies to propose mechanisms of As^{III} efflux via ArsB and ArsAB. In addition, some outstanding questions are highlighted. Future studies may be designed to address these questions on the basis of the proposed mechanistic models.

REFERENCES

[1] D. C. Crans and K. Kostenkova. Open questions on the biological roles of first-row transition metals. In: *Communications Chemistry* 3.1 (Aug. 7, 2020), p. 104. DOI: 10.1038/s42004-020-00341-w.

[2] S. K. Aulakh et al. The molecular landscape of cellular metal ion biology. In: *Cell Systems* 16.7 (July 16, 2025). DOI: 10.1016/j.cels.2025.101319.

[3] D. A. Capdevila et al. Bacterial Metallostasis: Metal Sensing, Metalloproteome Remodeling, and Metal Trafficking. In: *Chemical Reviews* 124.24 (Dec. 25, 2024), pp. 13574–13659. DOI: 10.1021/acs.chemrev.4c00264.

[4] J. A. Delmar, C.-C. Su, and E. W. Yu. Heavy metal transport by the CusCFBA efflux system. In: *Protein Science : A Publication of the Protein Society* 24.11 (Nov. 2015), pp. 1720–1736. DOI: 10.1002/pro.2764.

[5] A. G. Turner et al. The PerR-Regulated P1B-4-Type ATPase (PmtA) Acts as a Ferrous Iron Efflux Pump in Streptococcus pyogenes. In: *Infection and Immunity* 85.6 (May 23, 2017). DOI: 10.1128/iai.00140-17.

[6] K. J. Tsai, K. P. Yoon, and A. R. Lynn. ATP-dependent cadmium transport by the cadA cadmium resistance determinant in everted membrane vesicles of Bacillus subtilis. In: *Journal of Bacteriology* 174.1 (Jan. 1992), pp. 116–121. DOI: 10.1128/jb.174.1.116-121.1992.

[7] S. Liu et al. Uptake, efflux, and toxicity of inorganic and methyl mercury in the endothelial cells (EA.hy926). In: *Scientific Reports* 10.1 (June 2, 2020), p. 9023. DOI: 10.1038/s41598-020-66444-5.

[8] Priyanka and S. K. Dwivedi. Fungi mediated detoxification of heavy metals: Insights on mechanisms, influencing factors and recent developments. In: *Journal of Water Process Engineering* 53 (July 1, 2023), p. 103800. DOI: 10.1016/j.jwpe.2023.103800.

[9] C. W. Saltikov, R. A. Wildman, and D. K. Newman. Expression Dynamics of Arsenic Respiration and Detoxification in Shewanella sp. Strain ANA-3. In: *Journal of Bacteriology* 187.21 (Nov. 2005), pp. 7390–7396. DOI: 10.1128/jb.187.21.7390-7396.2005.

[10] A. J. Bard, R. Parsons, and J. Jordan. *Standard potentials in aqueous solution*. Marcel Dekker, Inc.,New York, NY, Dec. 31, 1984.

[11] H. Yamazaki, R. P. Sperline, and H. Freiser. Spectrophotometric determination of the dissociation constant (p*K*a) of arsenous acid. In: *Analytica Chimica Acta* 284.2 (Dec. 30, 1993), pp. 379–384. DOI: 10.1016/0003-2670(93)85323-C.

[12] C. Rensing, M. Ghosh, and B. P. Rosen. Families of Soft-Metal-Ion-Transporting ATPases. In: *Journal of Bacteriology* 181.19 (Oct. 1999), pp. 5891–5897. DOI: 10.1128/jb.181.19.5891-5897.1999.

[13] H. Garelick et al. Arsenic pollution sources. In: *Reviews of Environmental Contamination Volume 197* (2009), pp. 17–60. DOI: 10.1007/978-0-387-79284-2_2.

[14] P. A. O'Day et al. The influence of sulfur and iron on dissolved arsenic concentrations in the shallow subsurface under changing redox conditions. In: *Proceedings of the National Academy of Sciences* 101.38 (Sept. 21, 2004), pp. 13703–13708. DOI: 10.1073/pnas.0402775101.

[15] C. F. Harvey et al. Arsenic Mobility and Groundwater Extraction in Bangladesh. In: *Science* 298.5598 (Nov. 22, 2002), pp. 1602–1606. DOI: 10.1126/science.1076978.

[16] Y.-L. Meng, Z. Liu, and B. P. Rosen. As(III) and Sb(III) Uptake by GlpF and Efflux by ArsB in Escherichia coli*. In: *Journal of Biological Chemistry* 279.18 (Apr. 30, 2004), pp. 18334–18341. DOI: 10.1074/jbc.M400037200.

[17] M. S. Rahaman, N. Mise, and S. Ichihara. Arsenic contamination in food chain in Bangladesh: A review on health hazards, socioeconomic impacts and implications. In: *Hygiene and Environmental Health Advances* 2 (June 1, 2022), p. 100004. DOI: 10.1016/j.heha.2022.100004.

[18] I. Ben Fekih et al. Distribution of arsenic resistance genes in prokaryotes. In: *Frontiers in microbiology* 9 (2018), p. 2473. DOI: 10.3389/fmicb.2018.02473.

[19] T. C. Wood et al. Human Arsenic Methyltransferase (AS3MT) Pharmacogenetics: GENE RESEQUENCING AND FUNCTIONAL GENOMICS STUDIES*. In: *Journal of Biological Chemistry* 281.11 (Mar. 17, 2006), pp. 7364–7373. DOI: 10.1074/jbc.M512227200.

[20] S. Silver et al. Inducible plasmid-determined resistance to arsenate, arsenite, and antimony (III) in escherichia coli and Staphylococcus aureus. In: *Journal of Bacteriology* 146.3 (June 1981), pp. 983–996. DOI: 10.1128/jb.146.3.983-996.1981.

[21] C. M. Chen et al. Nucleotide sequence of the structural genes for an anion pump. The plasmid-encoded arsenical resistance operon. In: *Journal of Biological Chemistry* 261.32 (Nov. 1986), pp. 15030–15038. DOI: 10.1016/S0021-9258(18)66824-3.

[22] H. J. Sofia et al. Analysis of the Escherichia coli genome. V. DNA sequence of the region from 76.0 to 81.5 minutes. In: *Nucleic Acids Research* 22.13 (July 11, 1994), pp. 2576–2586. DOI: 10.1093/nar/22.13.2576.

[23] G. Ji and S. Silver. Regulation and expression of the arsenic resistance operon from Staphylococcus aureus plasmid pI258. In: *Journal of Bacteriology* 174.11 (June 1992), pp. 3684–3694. DOI: 10.1128/jb.174.11.3684-3694.1992.

[24] S. Dey, D. Dou, and B. P. Rosen. ATP-dependent arsenite transport in everted membrane vesicles of Escherichia coli. In: *Journal of Biological Chemistry* 269.41 (1994), pp. 25442–6. DOI: 10.1016/S0021-9258(18)47270-5.

[25] S. Dey and B. P. Rosen. Dual mode of energy coupling by the oxyanion-translocating ArsB protein. In: *Journal of Bacteriology* 177.2 (1995), pp. 385–9. DOI: 10.1128/jb.177.2.385-389.1995.

[26] H.-L. Fu et al. Properties of Arsenite Efflux Permeases (Acr3) from Alkaliphilus metalliredigens and Corynebacterium glutamicum. In: *The Journal of Biological Chemistry* 284.30 (July 24, 2009), pp. 19887–19895. DOI: 10.1074/jbc.M109.011882.

[27] J. Wu and B. P. Rosen. The ArsR protein is a trans-acting regulatory protein. In: *Mol Microbiol* 5.6 (1991), pp. 1331–6. DOI: 10.1111/j.1365-2958.1991.tb00779.x.

[28] T. B. Gladysheva, K. L. Oden, and B. P. Rosen. Properties of the arsenate reductase of plasmid R773. In: *Biochemistry* 33.23 (1994), pp. 7288–7293.

[29] J. Yang et al. Arsenic Binding and Transfer by the ArsD As(III) Metallochaperone. In: *Biochemistry* 49.17 (May 4, 2010), pp. 3658–3666. DOI: 10.1021/bi100026a.

[30] B. P. Rosen et al. Mechanism of the ArsA ATPase. In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1461.2 (Dec. 1999), pp. 207–215. DOI: 10.1016/S0005-2736(99)00159-5.

[31] A. R. Walmsley et al. Antimonite regulation of the ATPase activity of ArsA, the catalytic subunit of the arsenical pump. In: *Biochemical Journal* 360.3 (Dec. 2001), pp. 589–597. DOI: 10.1042/bj3600589.

[32] T. Zhou. Structure of the ArsA ATPase: the catalytic subunit of a heavy metal resistance pump. In: *The EMBO Journal* 19.17 (Sept. 2000), pp. 4838–4845. DOI: 10.1093/emboj/19.17.4838.

[33] T. Zhou et al. Conformational changes in four regions of the Escherichia coli ArsA ATPase link ATP hydrolysis to ion translocation. In: *Journal of Biological Chemistry* 276.32 (2001), pp. 30414–22. DOI: 10.1074/jbc.M103671200.

[34] A. R. Walmsley et al. A Kinetic Model for the Action of a Resistance Efflux Pump*. In: *Journal of Biological Chemistry* 276.9 (Mar. 2, 2001), pp. 6378–6391. DOI: 10.1074/jbc.M008105200.

[35] Y.-F. Lin, A. R. Walmsley, and B. P. Rosen. An arsenic metallochaperone for an arsenic detoxification pump. In: *Proceedings of the National Academy of Sciences* 103.42 (Oct. 2006), pp. 15617–15622. DOI: 10.1073/pnas.0603974103.

[36] A.-C. Milazzo et al. Active pixel sensor array as a detector for electron microscopy. In: *Ultramicroscopy* 104.2 (Sept. 1, 2005), pp. 152–159. DOI: 10.1016/j.ultramic.2005.03.006.

[37] X. Li et al. Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. In: *Nature Methods* 10.6 (June 2013), pp. 584–590. DOI: 10.1038/nmeth.2472.

[38] S. H. Scheres. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. In: *Journal of Structural Biology* 180.3 (Dec. 2012), pp. 519–530. DOI: 10.1016/j.jsb.2012.09.006.

[39] A. Punjani et al. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. In: *Nature Methods* 14.3 (Mar. 2017), pp. 290–296. DOI: 10.1038/nmeth.4169.

[40] M. Liao et al. Structure of the TRPV1 ion channel determined by electron cryo-microscopy. In: *Nature* 504.7478 (Dec. 2013), pp. 107–112. DOI: 10.1038/nature12822.

[41] X.-c. Bai et al. Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. In: *eLife* 2 (Feb. 19, 2013). Ed. by W. Kühlbrandt, e00461. DOI: 10.7554/eLife.00461.

[42] N. F. Mott. The scattering of electrons by atoms. In: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character* 127.806 (Jan. 1997), pp. 658–665. DOI: 10.1098/rspa.1930.0082.

[43] J. Wang and P. B. Moore. On the interpretation of electron microscopic maps of biological macromolecules. In: *Protein Science* 26.1 (2017), pp. 122–129. DOI: 10.1002/pro.3060.

[44] J. Wang et al. Identification of ions in experimental electrostatic potential maps. In: *IUCrJ* 5 (Pt 4 July 1, 2018), pp. 375–381. DOI: 10.1107/S2052252518006292.

[45] T. Bick, P. M. Dominiak, and P. Wendler. Exploiting the full potential of cryo-EM maps. In: *BBA Advances* 5 (Jan. 1, 2024), p. 100113. DOI: 10.1016/j.bbadva.2024.100113.

[46] E. Nogales. The development of cryo-EM into a mainstream structural biology technique. In: *Nature Methods* 13.1 (Jan. 2016), pp. 24–27. DOI: 10.1038/nmeth.3694.

[47] M. Wu and G. C. Lander. How low can we go? Structure determination of small biological complexes using single-particle cryo-EM. In: *Current Opinion in Structural Biology*. Biophysical and Computational Methods Cryo EM 64 (Oct. 1, 2020), pp. 9–16. DOI: 10.1016/j.sbi.2020.05.007.

[48] R. Henderson. Overview and future of single particle electron cryomicroscopy. In: *Archives of Biochemistry and Biophysics*. Electron Microscopy in Structural Biology 581 (Sept. 1, 2015), pp. 19–24. DOI: 10.1016/j.abb.2015.02.036.

[49] S. Subramaniam et al. Resolution advances in cryo-EM enable application to drug discovery. In: *Current Opinion in Structural Biology*. Multi-protein assemblies in signaling • Catalysis and regulation 41 (Dec. 1, 2016), pp. 194–202. DOI: 10.1016/j.sbi.2016.07.009.

[50] T. Nakane et al. Single-particle cryo-EM at atomic resolution. In: *Nature* 587.7832 (Nov. 2020), pp. 152–156. DOI: 10.1038/s41586-020-2829-0.

[51] R. Danev et al. Routine sub-2.5 Å cryo-EM structure determination of GPCRs. In: *Nature Communications* 12.1 (July 15, 2021), p. 4333. DOI: 10.1038/s41467-021-24650-3.

[52] R. Nygaard, J. Kim, and F. Mancia. Cryo-electron microscopy analysis of small membrane proteins. In: *Current opinion in structural biology* 64 (2020), pp. 26–33.

[53] M. A. Herzik, M. Wu, and G. C. Lander. High-resolution structure determination of sub-100 kDa complexes using conventional cryo-EM. In: *Nature Communications* 10.1 (Mar. 4, 2019), p. 1032. DOI: 10.1038/s41467-019-08991-8.

[54] L. Zheng et al. Uniform thin ice on ultraflat graphene for high-resolution cryo-EM. In: *Nature Methods* 20.1 (Jan. 2023), pp. 123–130. DOI: 10.1038/s41592-022-01693-y.

Chapter 2

INTRADIMERIC WALKER A ATPASES - CONSERVED MECHANISMS OF A FUNCTIONALLY DIVERSE FAMILY

This chapter is adapted and reprinted from:
Maggiolo, A. O.*; Mahajan, S.*; Rees, D. C.; Clemons, W. M. Intradimeric Walker A ATPases: Conserved features of a functionally diverse family. *Journal of Molecular Biology*, 435-11, 167965. (2023). Doi: 10.1016/j.jmb.2023.167965
*Indicates equal contributions.
S.M. and A.O.M. performed structural analysis and co-wrote the review article. All authors co-edited the review article.

Abstract

Nucleoside-triphosphate hydrolases (NTPases) are a diverse, but essential group of enzymes found in all living organisms. NTPases that have a G-X-X-X-X-G-K-[S/T] consensus sequence (where X is any amino acid), known as the Walker A or P-loop motif, constitute a superfamily of P-loop NTPases. A subset of ATPases within this superfamily contains a modified Walker A motif, X-K-G-G-X-G-K-[S/T], wherein the first invariant lysine residue is essential to stimulate nucleotide hydrolysis. Although the proteins in this subset have vastly differing functions, ranging from electron transport during nitrogen fixation to targeting of integral membrane proteins to their correct membranes, they have evolved from a shared ancestor and have thus retained common structural features that affect their functions. These commonalities have only been disparately characterized in the context of their individual proteins systems, but have not been generally annotated as features that unite the members of this family. In this chapter, we report an analysis based on the sequences, structures, and functions of several members in this family that highlight their remarkable similarities. A principal feature of these proteins is their dependence on homodimerization. Since their functionalities are heavily influenced by changes that happen in conserved elements at the dimer interface, we refer to the members of this subclass as intradimeric Walker A ATPases.

2.1 Introduction

The hydrolysis of nucleoside triphosphates (NTPs) to nucleoside diphosphates (NDPs) supports the progression of a multitude of cellular processes. P-loop NTPases, named after the characteristic functional loop within the fold that coordinates the phosphate moiety of the bound NTP, are a large superfamily that contain a distinctive three-layered $\alpha\beta\alpha$-sandwich fold structure [1–3]. Members of the P-loop NTPases perform a wide range of cellular functions, including transcription, translation, cellular regulation, DNA repair, cell division, and membrane transport [1, 4].

The P-loop, also known as the Walker A motif, was first identified by Walker and coworkers in several ATP-hydrolyzing proteins (ATPases), including the α and β subunits of ATP synthase, myosin, and adenylate kinase [5]. The Walker A motif contains the consensus sequence G-X-X-X-X-G-K-[T/S] (where X is any amino acid) that directly follows an N-terminal β-strand (Fig. 2.1A). The invariant Lys and Thr/Ser residues of this motif fix the position of the nucleotide by stabilizing the negative charge on the NTP γ-phosphate and coordinating the associated Mg^{2+} ion, respectively [1]. The structured glycine-rich P-loop plays an essential role in orienting phosphates for hydrolysis [1, 6–8]. Walker A proteins also contain a conserved Walker B motif that harbors an aspartate residue that coordinates the associated Mg^{2+} ion through a bridging water molecule. After the original Walker A characterization in ATPases, several small GTP-binding proteins were identified to contain both Walker A and B motifs, including the elongation factor Tu and the H-Ras GTPase [9, 10]. In these systems, the Walker A motif interacts with the phosphate groups of GTP, extending the P-loop terminology to both ATP and GTP hydrolyzing proteins. In recent years, the 'Walker A motif' and 'P-loop' terminology have been used interchangeably in the literature.

P-loop NTPases form a distinctive three-layer nucleotide binding fold that consists of a parallel β-sheet that is flanked on both sides by a layer of α-helices. This architecture is proposed to have emerged before the last universal common ancestor [3, 11]. The P-loop fold is structurally similar to another ancient fold, the Rossmann fold, differing primarily in the strand order of the β-sheet, with P-loop folds most commonly having a $\beta2$-$\beta3$-$\beta1$-$\beta4$-$\beta5$ topology [3, 12]. Both folds serve as a scaffold for functional loops that succeed the β-strands. Residues in these loops either directly interact with nucleotide (as in the case of the P-loop), or interact with a water or proteinaceous network. In the latter case, these loops are commonly termed 'Switch

I' and 'Switch II,' as they adopt discrete conformations that can be correlated to nucleotide state [13]. Several comprehensive works provide a framework for the classification and mechanism of P-loop NTPases (for further reading, see [8, 14]).

While P-loop NTPases are ubiquitous among the tree of life, several evolutionary clades have been identified with variations in the Walker A sequence [15]. Based on sequence and structural analysis, Leipe and coworkers broadly distributed the GTPase superclass of the P-loop NTPase superfamily into two classes: TRAFAC, which includes several well-characterized translation factor-related GTPases, and SIMIBI, which includes the signal recognition GTPases, MinD- and BioD-related NTPases [2]. It is known that SIMIBI proteins, that include both GTPases and ATPases, predominantly function as homo- or heterodimers.

A unique subset of proteins, later classified within the SIMIBI proteins, was first identified in the bacterial plasmid partitioning systems, having an uncommon alteration in the first and last glycine of the Walker A motif [16]. This family of proteins with a distinct Walker A sequence, X-K-G-G-X-G-K-S/T , was discussed in detail by E. Koonin in 1993 and expanded to include a handful of evolutionarily related ATPases, which have disparate functions and are found in all three domains of life [17]. Importantly, Koonin identified a second invariant lysine positioned near the beginning of the Walker A motif, which has since been regarded as the 'signature lysine' for members of this subset (Fig. 2.1A, bottom). This subset of ATPases, due to their divergence from other Walker A sequences, was termed the 'deviant Walker A' proteins; however, the question of ancestral origin is unresolved.

In this chapter, we discuss the commonalities among this distinct subset of P-loop NTPases. We coin the term 'intradimeric Walker A' (IWA) proteins as they are functional homodimers with important catalytic residues that span across the dimer interface. Some of the well characterized members of this group include the nitrogenase iron protein (NifH), the enzyme that supplies electrons during biological nitrogen reduction; the guided entry of tail-anchored proteins factor 3 (Get3), which shepherds tail-anchored integral membrane proteins to their destined membrane for insertion; the cytosolic subunit of the bacterial arsenite efflux pump (ArsA), which transfers arsenite to its partner membrane transporter for cellular detoxification; and the protein involved in the regulation of bacterial cell division (MinD). Other members of this family include DNA partitioning proteins such as ParA and Soj, cell division regulator, MipZ, and flagellar regulation protein, FleN. Due to the plethora of structural information, particularly for NifH and Get3, we use NifH,

Get3, and ArsA as representative examples to examine the conserved features of the IWA proteins at the sequence, structural, and functional levels. We find that the proteins in this special subset, with a common evolutionary origin but diverse functionalities, retain remarkable similarity in their ATP-dependent function and allostery as a product of their conserved structural features.

2.2 Sequence conservation

In comparison to the classical Walker A motif (Fig. 2.1A, top), the IWA proteins differ in the positions of the conserved glycine residues in the P-loop region and have a second invariant lysine after the first β-strand, making the consensus sequence for this subset of Walker A proteins β-strand 1- X-K-G-G-X-G-K-S/T (Fig. 2.1A, bottom) [17]. Although there are several hallmarks that differentiate the IWA proteins from the larger SIMIBI family, the initial lysine has generally been considered the defining feature of this subset. Similar to other SIMIBI NTPases, the IWA proteins contain a conserved motif located ~20 residues downstream of the Walker A motif on the second β-strand of the core β-sheet, designated as the Walker A' motif. This motif overlaps with Switch I and is characterized by a conserved aspartate, implicated in general base catalysis to activate the nucleophilic water for ATP hydrolysis in these proteins [14].

The general topology of P-loop NTPases in the IWA and greater SIMIBI family consists of a 5-stranded parallel β-sheet, where the first three β-strands make up the Walker A, Walker A', and the Walker B motifs. Each of these β-strands is followed by functional loops, which are designated the P-loop, Switch I, and Switch II, respectively (Fig. 2.1B). The IWA topology differs from the most common P-loop NTPase topology in that the IWA proteins have three additional β-strands in the core 5-stranded β-sheet from the insertion of two β-strands between β2 and β3 and a third β-strand between β4 and β5 (Fig. 2.1B, bold outlined β-strands). Nonetheless, the P-loop, Switch I, and Switch II come together on the same side of the β-sheet, such that the loops directly connect the nucleotide binding site to a second functional site that is specific to each protein (Fig. 2.1C).

2.3 Structural conservation

The IWA proteins have high structural conservation, which is dominated by the common P-loop NTPase fold. Each IWA protein has auxiliary helices and loops, which decorate the core αβα-sandwich and are specific to the function of each protein. For example, Get3 furnishes a series of flexible loops and helices, which

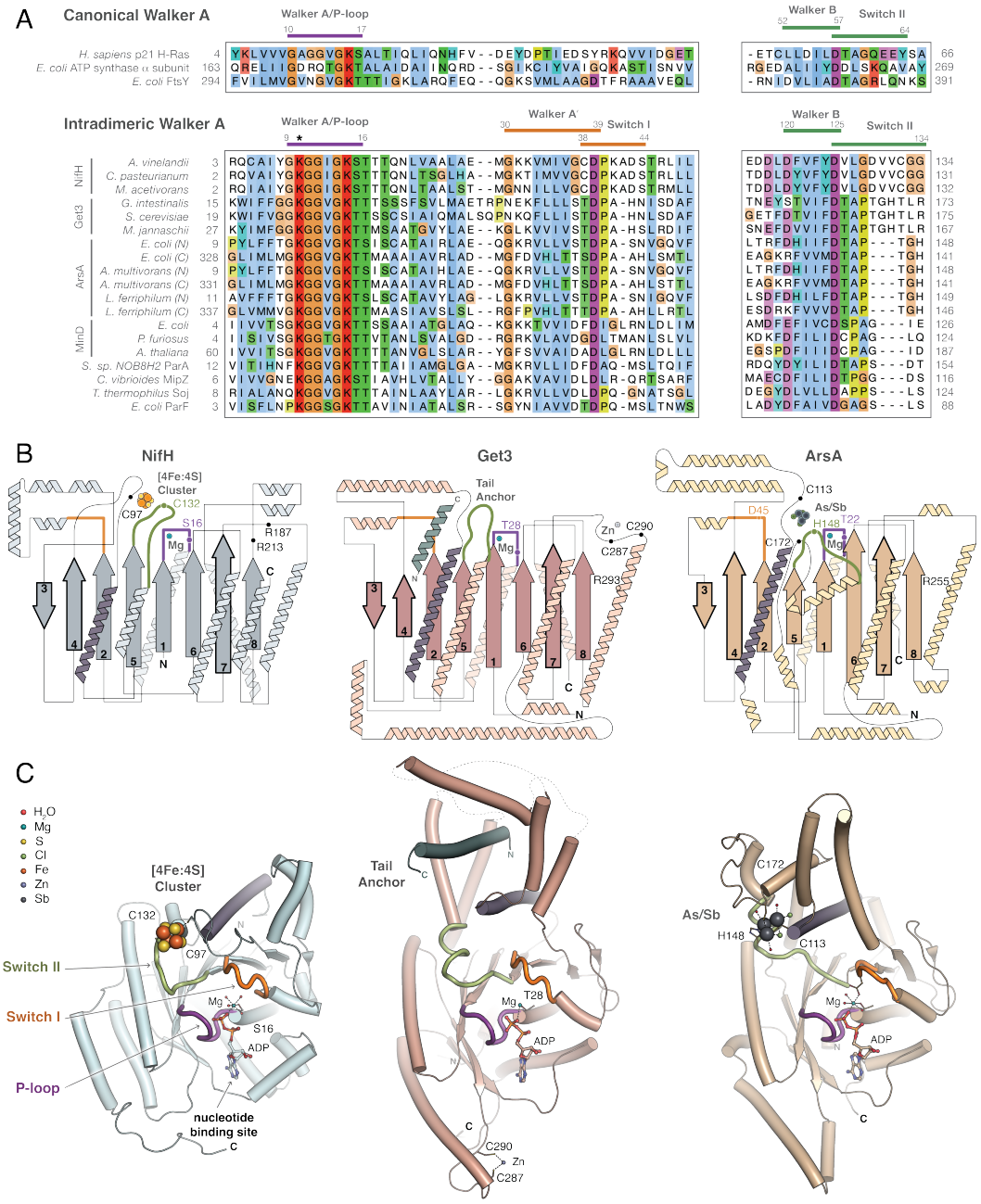

A
Canonical Walker A
Walker A/P-loop
Walker B
Switch II
H. sapiens p21 H-Ras
E. coli ATP synthase α subunit
E. coli FtsY
Intradimeric Walker A
Walker A'
Switch I
NifH
Get3
ArsA
MinD
A. vinelandii
C. pasteurianum
M. acetivorans
G. intestinalis
S. cerevisiae
M. jannaschii
E. coli (N)
E. coli (C)
A. multivorans (N)
A. multivorans (C)
L. ferriphilum (N)
L. ferriphilum (C)
E. coli
P. furiosus
A. thaliana
S. sp. NOB8H2 ParA
C. vibrioides MipZ
T. thermophilus Soj
E. coli ParF
B
NifH
[4Fe:4S] Cluster
C132
C97
S16
Mg
R187
R213
Get3
Tail Anchor
T28
Zn
C290
C287
R293
ArsA
C113
As/Sb
D45
H148
T22
C172
R255
C
N
C
Switch II
Switch I
P-loop
nucleotide binding site
ADP
H2O
Mg
S
Cl
Fe
Zn
Sb

Figure 2.1: **Sequence and structural conservation of Intradimeric Walker A (IWA) ATPases.** (A) (Top) Sequence alignment of representative members of the canonical Walker A NTPase superfamily and (bottom) of the IWA family, highlighting the conserved motifs. Positions of the motifs are numbered based on p21 H-Ras and *Av*NifH for canonical Walker A and IWA sequences, respectively. Asterisk indicates the location of the IWA signature lysine. Full names of the abbreviated organisms can be found in Table S2.1. Alignments were constructed using a structure-based alignment method with PROMALS3D [18], and residues are colored based on ClustalX [19]. (B) Topology diagrams of NifH (based on PDB 6N4L), Get3 (PDB 7SPY and PDB 7SQ0 for tail-anchor positioning), and ArsA (PDB 1F48) show the IWA proteins adopt a three-layer αβα-sandwich fold, commonly found in P-loop NTPases [20–22]. The relative β-sheet and α-helix lengths are shown to scale. The connecting loops are not to scale. Select residues are indicated using single letter codes. Bold outlined β-strands indicate insertions to the core β-sheet in the canonical P-loop NTPase fold. Purple-colored core α-helix indicates the 'substrate-helix'. (C) Cartoon representation of one monomer of *Av*NifH (left, PDB 6N4L), *Gi*Get3 (middle, PDB 7SQ0), and the N-terminal subunit of the *Ec*ArsA (right, PDB 1F48) pseudodimer. The IWA proteins contain a P-loop (purple), Switch I loop (orange), and Switch II loop (green) which couple the nucleotide binding site to the substrate/cargo binding site.

form a hydrophobic groove used to capture tail-anchored protein cargo, while NifH utilizes helices to form the interaction surface with its partner protein, NifDK.

Aside from their common P-loop NTPase fold, a key conserved feature of IWA proteins is the requisite homodimeric state (Fig. 2.2A). In the case of NifH, the homodimer is covalently linked by coordination of the active site cluster, containing four iron and four sulfur atoms ([4Fe:4S]). Similarly, Get3 is a homodimer that is stabilized by a zinc ion. In both NifH and Get3, a number of X-ray crystal structures are available that span multiple functional states and delineate changes at the dimer interface, that are critical for function. ATP-dependent dimerization is an essential feature for the function of MinD, which exists as a monomer prior to ATP binding. [23]. In these systems, large dimer conformational changes occur in response to the bound nucleotide state. Thus, there appears to be communication from one monomer to the other throughout the nucleotide hydrolysis cycle. ArsA differs in this respect as it is expressed as a single chain that effectively forms a covalent dimer, often referred to as a pseudodimer, with two nucleotide binding domains connected by a flexible linker. This has been suggested to be a result of gene duplication and fusion of an ancestor [24]. The pseudosubunits have ∼30% sequence identity and together form a dimer interface that mimics the conserved features found in NifH

and Get3.

A critical interaction across the dimer interface is provided by the IWA signature lysine. This was first recognized in the X-ray crystal structure of NifH complexed to an ATP-transition state analogue, ADP·AlF_4^- [25]. In this structure, the amino group of Lys10 in each NifH monomer interacts with the negatively charged phosphate group of the nucleotide in the adjacent monomer (Fig. 2.2B, left). A similar interaction is seen in the ATP-bound Get3 dimer (Fig. 2.2B, center). Thus, the lysine was proposed to act as both a charge neutralizer for the nucleotide phosphate moieties and to lower the activation barrier of ATP hydrolysis by stabilizing the negative charge build-up on the phosphate leaving group in the transition state. This interaction is structurally equivalent to the 'arginine finger' of GTPase activating proteins that stimulate GTP hydrolysis in Ras proteins [26]. The location of the catalytic lysines on the opposing subunits serves to facilitate communication about the nucleotide phosphorylation state across a dimer interface and regulate hydrolysis, which can only proceed once the dimer is in the appropriate conformational state. A number of non-nucleotide bound structures of NifH and Get3 reveal that the dimer splays apart in preparation for nucleotide uptake. In these structures, the dimer adopts an 'open' conformation and the residues that make up the dimer interface are oriented such that the IWA lysines are too far apart to interact with the opposing subunit. After nucleotide binding, the dimer interface restructures as the monomers twist together to enclose the nucleotides and form a 'closed' dimer conformation (Fig. 2.3A-B). This mechanism ensures that nucleotide hydrolysis is directly coupled to the conformational state of the dimer.

In contrast to the structural commonalities at the nucleotide binding domains among the IWA proteins, the auxiliary domain outside the nucleotide binding domain have evolved to adopt functions unique to each protein. Thus, the IWA proteins can be thought of as having two functional sites: (i) the site of ATP hydrolysis and (ii) the site of substrate/cargo binding that is specific to each protein (simplified as substrate for the rest of the text) (Fig. 2.2A). The two sites communicate to one another by the intervening loops (Switch I and Switch II) and the overall dimer conformation, which is how substrate binding and nucleotide hydrolysis are coupled. For example, NifH has a [4Fe:4S] cluster that enables inter-protein electron transfer between NifH and its partner protein, NifDK. To prevent futile redox cycles, electron transfer only occurs when NifH is properly coordinated to its partner protein and poised to deliver electrons. Structures of NifH complexed to NifDK and the ATP analog, ADP·AlF_4^-,

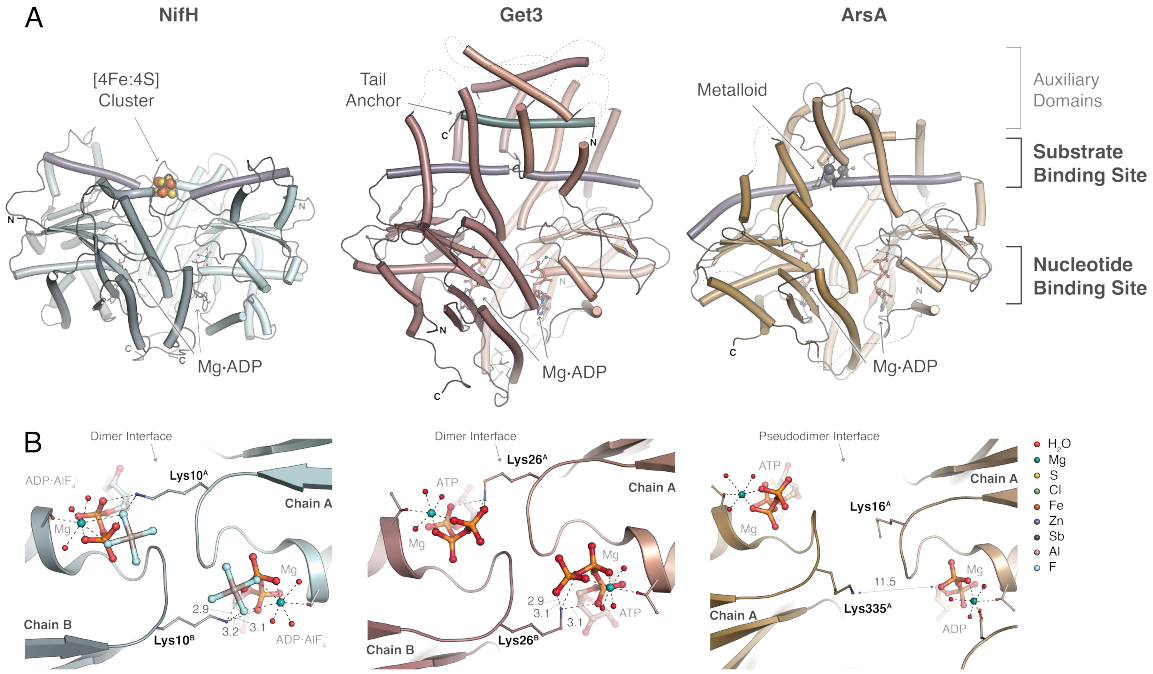


Figure 2.2: **Overall architecture and dimer interface of representative IWA ATPases — NifH, Get3 and ArsA.** (A) IWA proteins have disparate functions, but commonly act as 'shuttle proteins' that are driven by ATP hydrolysis. In this regard, they have two functional sites: one for nucleotide binding and one for substrate binding. The substrate binding site is tailored to accommodate the cargo specific to each protein, in some cases by the addition of an auxiliary domain. Cartoon representation of three IWA proteins with their nucleotide and substrate binding sites indicated. NifH (PDB 6N4L) carries electrons to its partner protein complex, NifDK, for the reduction of dinitrogen; Get3 (PDB 7SQ0) delivers tail-anchored proteins to the Get1 and Get2 complex for targeted insertion into the membrane; and ArsA (PDB 1F48) scavenges and transports toxic metalloids to the ArsB efflux pump for extrusion out of the cell. The substrate-helix is indicated in purple. (B) The signature N-terminal lysine in IWA proteins reaches across the dimer interface and interacts with the phosphate moiety of the nucleotide in its partner monomer. Representative examples shown are NifH (left, PDB 1M34), Get3 (middle, PDB 7SPY), and ArsA (right, PDB 1II0) [21, 27, 28]. The lysine provides the positive charge necessary to initiate ATP hydrolysis. In addition, it plays a role in neutralizing the build-up of negative charge after ATP hydrolysis and communicates the nucleotide phosphorylation state across the dimer interface. Distances are indicated in angstroms.

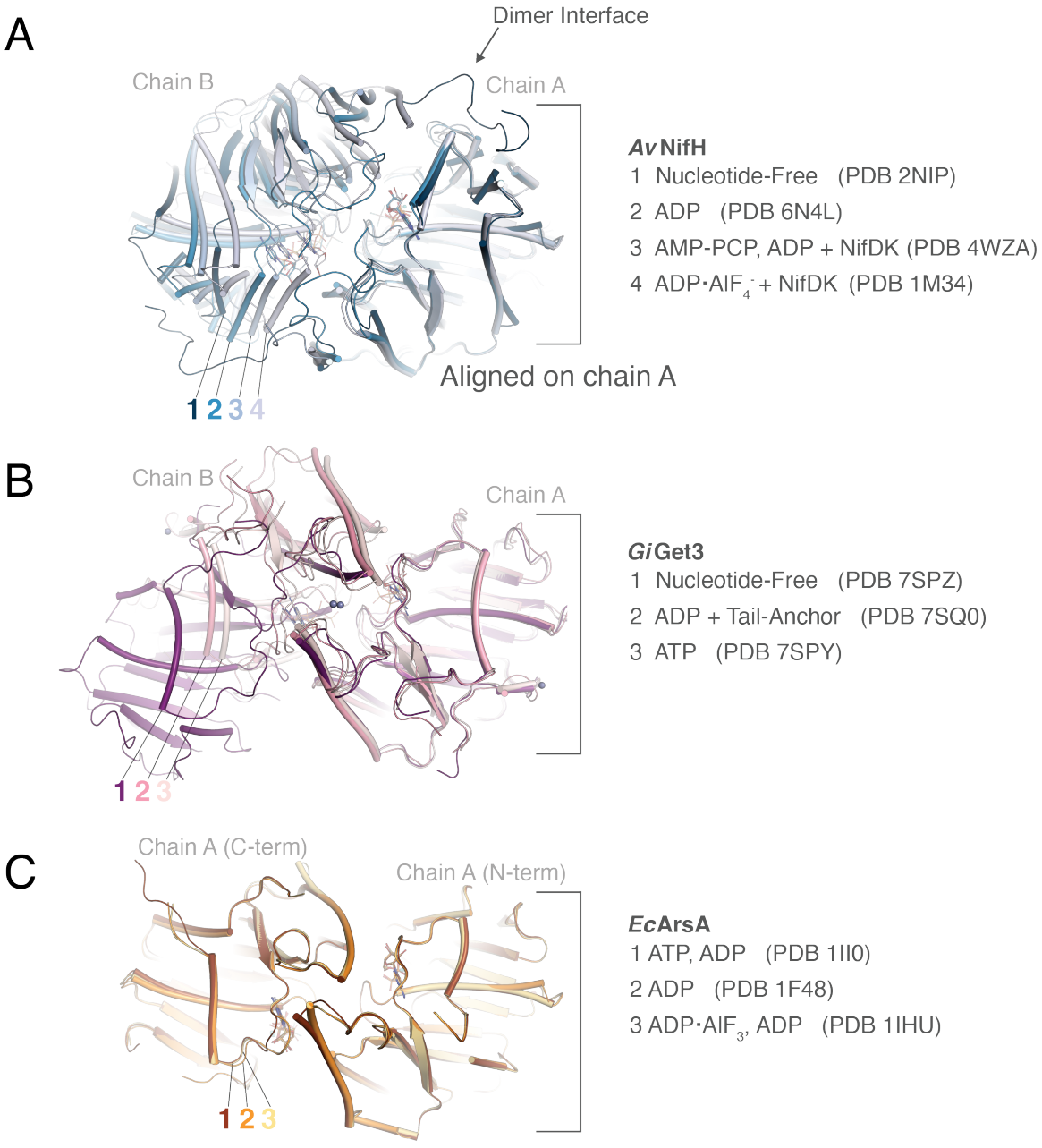


Figure 2.3: **Nucleotide-dependent dimer dynamics of IWA ATPases.**The dimer dynamics of each protein can be visualized by alignment on chain A of several nucleotide-dependent dimer states of (A) *Av*NifH, (B) *Gi*Get3, and (C) *Ec*ArsA. The line of sight for the dimer is from the bottom relative to the dimer conformation as viewed in Fig. 2.2B.

reveal that ATP binding pushes the cluster 4 Å towards the surface of the protein [25]. This shift of the cluster enables electron transfer by minimizing the distance between the [4Fe:4S] cluster of NifH and the receptor iron sulfur cluster in NifDK. The movement of the [4Fe:4S] cluster in NifH is facilitated by a concerted shift of the Switch I and II loops in response to ATP binding (Fig. 2.4A-D). Thus, the two active sites are coupled by the components of the IWA motif that relay the nucleotide binding signal to the [4Fe:4S] cluster site and enable the overall activity to be modulated by ATP hydrolysis.

Likewise in Get3, signal is propagated from the ATP binding site to the tail-anchored protein binding site, predominantly through changes that happen in the Switch II loop. Upon binding ATP, the Switch II loop shifts, affecting changes in residues that will subsequently form the hydrophobic cleft at the dimer interface (Fig. 4E-H) [21]. Accessory proteins, such as Get4, and the tail-anchored protein likely facilitate structural rearrangements that form the groove for tail-anchored protein binding, which then stimulates ATP hydrolysis [29, 30]. This is evident in the post-hydrolysis tail-anchored protein-bound state of *Giardia Intestinalis* Get3 (*Gi*Get3), wherein the C-terminal end of the Switch II loop has restructured into an α-helix (Fig. 4G-H) [21]. Furthermore, in both NifH and Get3, the helix that precedes the β-strand harboring the Walker B motif (residues 96-112 in *Av*NifH and 139-157 in *Gi*Get3), herein referred to as the 'substrate-helix', moves as part of the conformational changes that occur at the substrate binding site (Fig. 2.1B-C, purple helices). This is most apparent in structures of Get3, where this helix shifts down towards the nucleotide binding site upon tail-anchored protein binding, forming the base of the hydrophobic cleft [21, 30]. In NifH, the change in conformation is more subtle, as rearrangements in this helix are coupled to repositioning of the [4Fe:4S] cluster to facilitate intermolecular electron transfer when complexed to NifDK. Switch II likely coordinates with the substrate-helix to mediate communication between the nucleotide binding site and the substrate binding site.

ArsA is structurally more similar to Get3 than the other well-characterized IWA proteins. Like Get3, ArsA contains the auxiliary helical domain that extends above the nucleotide binding sites (Fig. 2.1C, center and right). In Get3 this domain makes up the hydrophobic cleft for tail-anchored protein binding. In ArsA, it encloses a cysteine-rich site located at the dimer interface that forms the binding site for its physiological metalloid substrates, arsenite and antimonite [22, 31]. The presence of three conserved cysteines at this site and the absence of the features necessary for

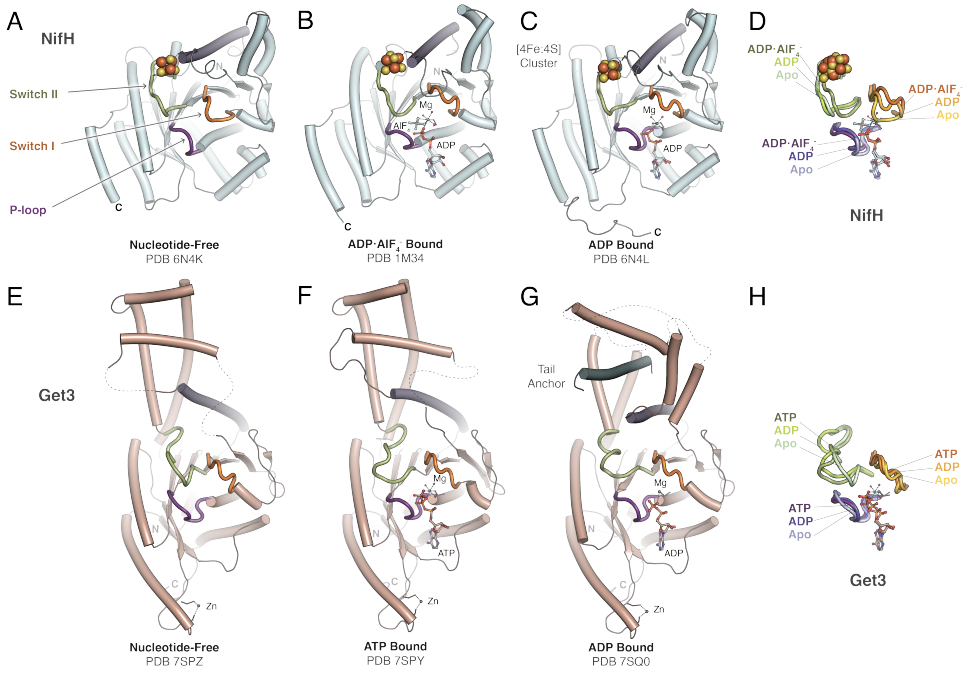


Figure 2.4: **Conformations of the Switch I and Switch II in various nucleotide-bound states of NifH and Get3.** (Top) Conformations of the nucleotide-dependent functional loops (P-loop (purple), Switch I (orange), and Switch II (green)) in one monomer of NifH in the (A) nucleotide-free, (B) ADP·AlF_4^-, and (C) ADP bound states. (D) Direct comparison of the three functional loops to demonstrate their nucleotide-dependent conformational changes. (Bottom) Conformations of the nucleotide-dependent functional loops in one monomer of Get3 in the (E) nucleotide-free, (F) ATP, and (G) ADP and tail-anchor substrate bound states. (H) Comparison of the Get3 functional loops.

tail-anchored protein binding are the hallmarks that distinguish ArsA from Get3, all of which were indiscriminately annotated as ArsA in databases. Mutagenesis and tryptophan fluorescence kinetic studies revealed that the metalloid binding site of ArsA is allosterically coupled to ATP binding and hydrolysis [31, 32]. Analogous to NifH and Get3, the substrate binding site of ArsA is structurally linked to the ATP binding site via the Switch loops (Fig. 2.1B-C, right). X-ray crystal structures of *Escherichia coli (Ec)*ArsA bound to various nucleotides have been reported [22, 28]. Surprisingly, there are no significant conformational changes between these structures, unlike the dramatic changes seen in the nucleotide-free versus nucleotide-bound states of Get3 and NifH (Fig. 2.3C). Likewise, the signature lysines in the structures of ArsA bound to ADP·AlF_3 or ATP are >10 Å apart and are oriented away from the nucleotide, resembling the 'open' dimer conformation (Fig. 2.2B), right). It is plausible that the crystal packing was not conducive to explore the

full conformational landscape of ArsA. Nevertheless, the presence of all the IWA structural motifs in ArsA strongly indicate that the ATP-dependent conformation rearrangements are likely preserved and required to catalyze the transfer of the metalloid to the membrane efflux pump, ArsB.

2.4 Functional conservation

In IWA proteins, ATP binding and hydrolysis impose changes at the dimer interface that span a range of nucleotide-dependent dimer conformational states (Fig. 2.3A-B). An abundance of structural data, particularly among the NifH and Get3 proteins, supports that the members of the IWA family exhibit common nucleotide-dependent dimer dynamics. In the nucleotide-free structures of NifH and Get3, the dimers adopt the open conformation with minimal protein interactions across the dimer interface [20, 21]. In the structures of NifH and Get3 in the pre-hydrolytic ATP-bound state, the dimers are stabilized in the 'closed' conformation, wherein the dimer interface constricts, bringing the IWA signature lysine in proximity to the nucleotides in preparation for hydrolysis [21, 33]. In these systems, docking of a partner protein or substrate stabilizes the 'closed' dimer conformation and serves as a trigger for ATP hydrolysis. In NifH, this trigger is provided by complexation to the partner protein, NifDK [27]. In Get3, tail-anchored protein binding and subsequent release from a binding partner initiates hydrolysis [21, 34]. In MinD, binding of the partner protein, MinE, activates hydrolysis [35].

At the hydrolysis step, a water molecule positioned adjacent to the γ-phosphate and trans to the β-phosphate, serves as the nucleophile for ATP hydrolysis. In NifH, several aspartate residues within the nucleotide binding site are candidates for activation of the catalytic water. The structure of the ADP·AlF_4^--bound state of NifH, contains a hydrogen bonding network involving the conserved aspartate residues on the Switch I and II loops, the water molecules coordinating the Mg^{2+} ion, and the conserved Ser/Thr residue on the P-loop (Fig. 2.5A-B). It is difficult to tell from structures alone whether a single aspartate is responsible for water activation. Based on both structural interpretation and mutagenesis studies, it is more likely that several apartates are essential and an intricate hydrogen bonding network surrounding the catalytic water facilitates its activation via a proton relay mechanism [14, 36–38]. Similar to the IWA signature lysine, Asp129 reaches across the dimer interface to stabilize the catalytic water in the ADP·AlF_4^--bound state, but not in the ADP-bound state (Fig. 2.5C), further affirming the importance of nucleotide-dependent intradimeric interactions. A similar arrangement of aspartates

forming a hydrogen bonding network is also present in the ATP-bound Get3 structure [21]. Following hydrolysis and phosphate release, the dimer expands, returning to its open conformation, as represented in the ADP-bound structures of NifH and Get3 [20, 21].

The ATPase cycle mediates discrete conformational rearrangements of the Switch loops. This allosterically induces structural changes at the substrate binding site involved in the primary biological function of the respective proteins. This typically involves transfer of the substrate to its downstream binding partner as characterized for NifH, Get3, and ArsA. On the other hand, MinD facilitates localization of an entire protein, MinC, to the site on the plasma membrane where non-productive cell divison must be prevented [23, 39]. The combined oscillation of MinD and MinC between the poles of a dividing cell is required for cell division regulation. Based on several structures, the substrate binding site of MinD differs from other IWA members in that MinC binds at the dimer interface near the nucleotide binding sites of MinD [**szewczakharris_cryo_2019**, 40, 41], which is a distinct region from the substrate binding site of NifH, Get3, and ArsA. Nonetheless, the presence of the IWA features indicates that functionally relevant interactions at the analogous IWA substrate binding site on MinD must be conserved and the corresponding details await future characterization. As the ATPase cycle regulates transfer of a substrate from an IWA proteins to either a partner protein or a different location in the cell, these proteins, in principle, can be thought of as 'shuttle proteins' that have evolved to catalyze the transport (shuttling) of their respective substrates. Although a similar description has previously been made with reference to DNA segregation proteins (MinD and ParA) [42], we identify that this functional role is conserved across the IWA family.

The catalytic steps of the ATPase cycle regulating the shuttle function of the IWA proteins are mediated by partner proteins that are functionally analogous to effectors. NifH is an example where the catalytically competent conformation required for both ATP hydrolysis and electron transfer is stabilized only when it is bound to NifDK; this property effectively prevents futile ATPase cycles [25]. After ATP hydrolysis and electron transfer, the NifH dimer dissociates from NifDK and transitions to the open dimer conformation. In this state, ATP exchanges for ADP and the [4Fe:4S] cluster can be reduced by the appropriate reductant [43]. Although both Get3 and ArsA exhibit basal ATPase activity in the absence of substrate, effector binding in both cases has been shown to regulate the ATPase cycle [34, 44]. Binding of the

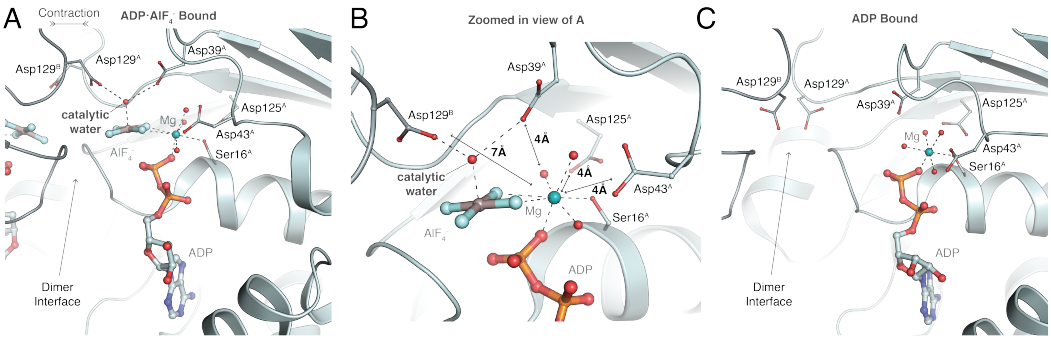


Figure 2.5: **Hydrogen-bonding network composed of Switch I and Switch II aspartates at the nucleotide binding site of NifH.** (A) Asp39 and 43 (Switch I), and Asp125 and 129 (Switch II) participate in an intricate hydrogen bonding network around coordinated ADP·AlF_4^-. Asp43, 125, and 129 are essential for the Mg·ATP hydrolysis activity of *Av*NifH (PDB 1M34) [8, 36–38]. Mutation of Asp39 has been shown to affect the FeMo-cofactor synthesis and maturation activities of NifH [38]. (B) Zoom-in of panel A indicating the distances from the aspartates to the Mg^{2+} ion (PDB 1M34). (C) Aspartate residues around the nucleotide binding site of *Av*NifH when ADP is bound (PDB 6N4L).

effector protein, Get4, to Get3 stabilizes its ATP-bound 'closed' dimer conformation, exposing the substrate binding site for tail-anchored protein binding. Hydrolysis is inhibited until the tail-anchor is successfully bound [29, 30]. Subsequently, Get3 binds to the Get1 and Get2 endoplasmic reticulum membrane complex, which stabilizes the open dimer conformation, facilitating the transfer of the tail-anchored protein into the membrane [**mariappan_mechanism_nodate**]. In the case of ArsA, a metallochaperone, ArsD, putatively binds ArsA to facilitate transfer of arsenite to the substrate binding site of ArsA [45]. Additionally, by analogy to NifH-NifDK, interactions between ArsA and its membrane partner ArsB may also be critical to regulate the nucleotide state of ArsA suitable for arsenite transfer. Yet another mechanism of regulation of IWA proteins, especially notable in Get3, MinD, and chromosomal segregation proteins such as ParA and Soj, has been suggested to be conferred by membrane association of these proteins. NifH has been found to associate with membranes in nitrogen-fixing bacteria under low oxygen tension [46]. Indeed, it has been proposed that lipids can trigger conformational changes in these systems [47].

The presence of such regulatory features in IWA proteins that control their switching between the open and 'closed' dimer conformations might serve as a critical mechanistic feature that ensures precise timing of the substrate shuttling function.

That these proteins act as 'molecular clocks' is well established for MinD, which oscillates between the poles of the cell and prevents non-productive division septa formation necessary for cell division. The ATP-dependent dimerization linked to binding of MinC and MinE proteins helps regulate the periodicity of MinD function [23]. A similar mechanistic role of ATP hydrolysis has been discussed previously in light of the nitrogenase mechanism, suggesting that ATP hydrolysis is thermodynamically non-essential for electron transfer and instead acts as a conformational gate to maintain the unidirectionality of electron transfer from NifH to NifDK [48]. We propose that the well regulated ATPase cycle is critical for controlling the flux of the substrate being shuttled, therefore, ensuring precise timing in other IWA proteins, as well.

2.5 Structurally conserved arginine residues

Beyond the residues in the conserved sequence motifs defining this family, IWA proteins have additional conserved interactions across the dimer interface that correlate with nucleotide-dependent conformational states. Of particular interest are interactions involving arginine residues that are frequently found at the dimer interface of IWA proteins. One striking example, described in *Av*NifH by Tezcan, Rees, and coworkers, involves a pair of arginine residues (Arg187 and Arg213) from each NifH subunit in a complex of NifH with NifDK, stabilized by the asymmetric binding of AMP-PCP and ADP [49]. Arg187 and Arg213 are positioned on loops following β-strand 7 and β-strand 8, respectively (Fig. 2.1B). The intriguing aspect of these arginines was the prominent chain of π-stacking interactions involving the adenine bases and the guanidinium groups of the arginines to form the π-stacked chain adenine - Arg213 - Arg187 - Arg187 - Arg213 - adenine linking the two nucleotide binding sites across the homodimer interface (Fig. 2.6A). This observation prompted examination of these residues in a series of NifH structures to evaluate the relationship between the locations of these arginines and the nucleotide dependent NifH dimer conformation. Arg187 and Arg213 were found to interact via discrete hydrogen bonding and salt bridge interactions with each other and with a neighboring glutamate (Glu154 in *Av*NifH) in structures of NifH complexed to NifDK with ADP·AlF_4^- [25, 27], ADP [33], and AMP-PCP [49]. Structures of nucleotide-free NifH do not exhibit direct arginine-arginine interactions across the dimer interface as the dimer is splayed too far apart; additionally, the adenine appears to provide an important handle for arginine stacking (Fig. 2.6B) [20, 50]. We note that mutagenesis of Arg213 to cysteine renders the protein inactive as measured by *in*

vivo acetylene reduction assays [51]. The interactions between these four arginine residues and the nucleotide could serve to control nucleotide positioning across an interface coupling the dimer conformation to the nucleotide state. Although the interaction between two positively charged side chains has an unfavorable electrostatic component, arginine-arginine interactions across protein interfaces have been found to be thermodynamically favorable [52–54], thereby facilitating the formation of the 'closed' dimer conformation.

In addition to NifH, a survey of IWA motif-containing proteins in the Protein Data Bank (PDB) revealed the presence of an arginine residue (sometimes a pair of arginine residues) that is structurally analogous to NifH Arg187 and 213. Similar to the IWA signature lysine, these arginines are found to participate in extensive inter-subunit interactions with both the protein and the bound nucleotides. Arginine-arginine interactions are maintained in MinD and more broadly in the Par family of proteins (Fig. 2.6C-D). Upon examination of the available MinD structures from different organisms, we find inter-subunit hydrogen bonds made by the two dimer-related arginines that also interact with nucleotides. In *Pyrococcus furiosus(Pf)*MinD, a pair of arginines (residues Arg172 and Arg175) stack over the nucleotide in the monomeric ADP-bound structure (Fig. 2.6C). As there are no dimeric ATP-bound structures of *Pf*MinD available, we examined the ATP-bound, dimeric *Ec*MinD that has arginines homologous to those of *Pf*MinD (Arg182 and Arg187, respectively). In this structure, Arg187 is dissociated from the stack upon dimerization and Arg182 is hydrogen bonded to Arg187 from the adjacent monomer (Fig. 2.6D).

In Get3, arginine interactions at the dimer interface are most clear in *Saccharomyces cerevisiae (Sc)*Get3. There is a single arginine residue (Arg291 in *Sc*Get3) that participates in nucleotide-dependent intradimeric contacts. In structures of *Sc*Get3 with ATP bound, Arg291 from each monomer interacts with the adenine base of the ATP in the opposing subunit (Fig. 2.6E). In the structures with no nucleotide bound, Arg291 from each monomer interacts across the dimer interface via direct hydrogen bonding or π-π stacking (Fig. 2.6F). An investigation of 69 mutations in *Sc*Get3 by Suloway, Clemons, and coworkers found that mutation of Arg291 resulted in a strong loss of function phenotype [55]. Incubation of R291A with wildtype protein could not recover function, indicating mutation of Arg291 had a dominant negative effect (unpublished results). The corresponding arginines in *Ec*ArsA interact with the adenine in the opposing monomer in the single dimer conformation seen for

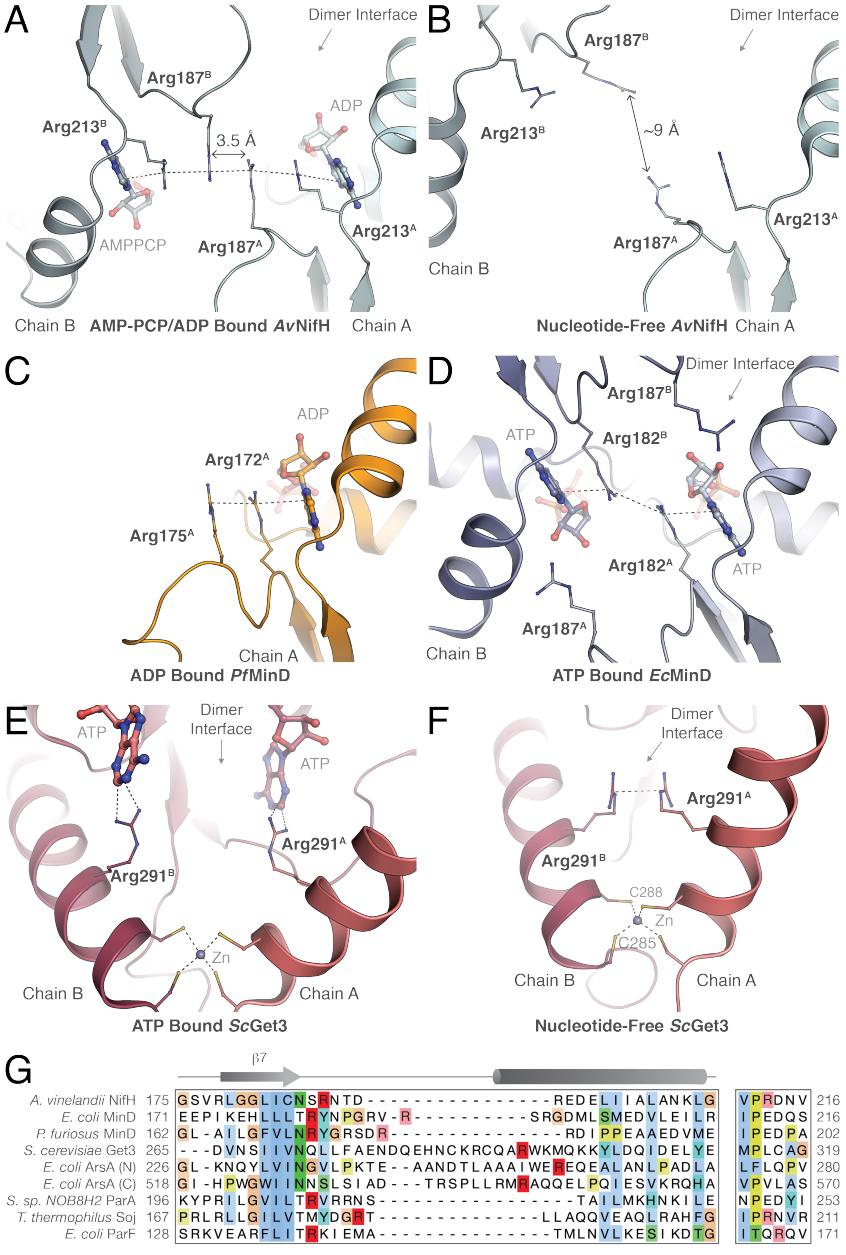

A
Dimer Interface
Arg187B
Arg213B
3.5 Å
ADP
AMPPCP
Arg187A
Arg213A
Chain B
AMP-PCP/ADP Bound AvNifH
Chain A
B
Dimer Interface
Arg187B
Arg213B
~9 Å
Arg213A
Arg187A
Chain B
Nucleotide-Free AvNifH Chain A
C
ADP
Arg172A
Arg175A
Chain A
ADP Bound PfMinD
D
Dimer Interface
Arg187B
ATP
Arg182B
Arg182A
ATP
Arg187A
Chain B
ATP Bound EcMinD
E
ATP
Dimer Interface
ATP
Arg291A
Arg291B
Zn
Chain B
Chain A
ATP Bound ScGet3
F
Dimer Interface
Arg291A
Arg291B
C288
Zn
C285
Chain B
Chain A
Nucleotide-Free ScGet3
G
β7
A. vinelandii NifH 175 GSVRLGGLICNSRNTD----------REDELIIALANKLG 216
E. coli MinD 171 EEPIKEHLLLTRYNPGRV-R----------SRGDMLSMEDVLEILR 216
P. furiosus MinD 162 GL-AILGFVLNRYGRSDR----------RDIPPEAAEDVME 202
S. cerevisiae Get3 265 ---DVNSIIVNQLLFAENDQEHNCKRCQAPWKMQKKYLDQIDELYE 319
E. coli ArsA (N) 226 GL-KNQYLVINGVLPKTE--AANDTLAAAIWEHEQEALANLPADLA 280
E. coli ArsA (C) 518 GI-HPWGWIINNSLSIAD---TRSPLLRMRAQQELPQIESVKRQHA 570
S. sp. NOB8H2 ParA 196 KYPRILGVILTRVRRNS----------TAILMKHNKILE 253
T. thermophilus Soj 167 PRLRLLGILVTMYDGRT----------LLAQQVEAQLRAHFG 211
E. coli ParF 128 SRKVEARFLITRKIEMA----------TMLNVLKESIKDTG 171
VPRDNV
IPEDQS
IPEDPA
MPLCAG
LFLQPV
VPVLAS
NPEDYI
IPRNVR
ITQRQV

Figure 2.6: **Structurally conserved arginine stacking motif at the dimer interface of IWA ATPases.** (A) Four arginine residues (two from each monomer) reside at the dimer interface of *Av*NifH. In the absence of nucleotide (PDB 6N4K), the dimer adopts an 'open' conformation with the arginines too widely separated to interact. Dimer interface distances are indicated as solid arrows. (B) Upon nucleotide binding, the NifH dimer constricts and the four arginines π-π stack over the nucleotides and across the dimer interface (PDB 4WZA). (C) Two arginine residues from the same monomer that structurally align with the IWA arginine residues in *Av*NifH, adopt a similar stacking motif in the monomeric *Pf*MinD. (D) In the dimeric ATP-bound *Ec*MinD structure, one pair of arginines stack over the bound nucleotides and extend interactions across the dimer interface. (E) Arg291 in *Sc*Get3 interacts with the nucleotide in the neighboring monomer when ATP is bound (PDB 4XVU) and (F) interact at the dimer interface when either no nucleotide or ADP is bound (PDB 3SJA). (G) The IWA arginine residues (highlighted in red) are structurally conserved but are not discernible from the sequence alignment. Full names of the abbreviated organisms can be found in Table S2.1.

ArsA; however, a π-π stacking interaction between the arginines is not observed [22, 28]. As discussed previously, the conformational states of ArsA do not appear to have been fully explored and it is likely that these types of interactions will be observed in conformational states of ArsA yet to be characterized.

The absence of a conserved sequence motif surrounding these arginines has precluded recognition of the significance of these interactions (Fig. 2.6G, dark red). Nevertheless, the arginines are structurally conserved with respect to the nucleotide binding site and are present on the loop or the α-helix following β-strand 7. In some IWA proteins, a second arginine is present at the dimer interface and typically resides on a different structural element (Fig. 2.6G, light red). Given the striking conformation of these arginine residues, which interact across a dimer interface at the nucleotide binding site in several members of the IWA subclass, and that the intra-subunit interactions correlate with multiple structural intermediate states in NifH, the arginine residues further support the importance of cross-dimeric interactions in the IWA family.

2.6 Conclusions

The IWA proteins had previously been ascribed to a subset of Walker A proteins that have an additional, fully conserved N-terminal lysine residue that is essential in ATP hydrolysis in these enzymes that function as obligate homodimers. Due to their common P-loop NTPase fold, they share a conserved nucleotide binding site.

The IWA proteins have a second functional site, also at the dimer interface and typically at the opposing end of the dimer from the nucleotide binding site. This second functional site serves as the location for substrate binding and is tailored with auxiliary structural components to accommodate substrates that are specific to each protein system. Within a monomer, a series of loops that straddle the nucleotide and substrate binding sites directly couple the substrate transport activity to the ATPase cycle. As illustrated by studies of MinD, regulation of the ATPase cycle of IWA proteins by interactions with partner proteins and/or the cell membrane ensures precise timing and unidirectionality of substrate transport. This timing mechanism might be a critical feature that enables IWA proteins to carry out their respective biological functions efficiently.

Structural analysis reveals that the nucleotide-dependent conformations of the homodimer are an important feature in the progression of the ATPase cycle in IWA proteins. Several residues that make intradimeric interactions, including the IWA signature lysine, are essential in facilitating communication about the phosphorylation state of the coordinated nucleotides across the dimer interface to induce concerted dimer motions. Herein, we identify arginine residues that are present at the dimer interface in the IWA proteins, but not conserved in the IWA sequence motif. It is tempting to speculate on their potential role in the IWA proteins, which may not necessarily be the same for all IWA members. Although these arginines appear to be unique to IWA proteins, a highly conserved lysine residue in the related NTPase, p21 Ras, (Lys117 in *Homo sapiens*) is structurally homologous to Arg213 in NifH [48, 56]. Lys117 in Ras positions directly over the base of the bound nucleotide and plays an important role in guanine base recognition [56, 57]. The connection between Arg213 in NifH and Lys117 in Ras has been made before as one of the residues in the P-loop NTPase fold responsible for adenine versus guanine base specificity [58]. Whether these arginines partake in either base specificity or dimer dynamics, or both, in NifH and other IWA proteins is worth further investigation.

The commonalities described here between the members in this special subset of Walker A proteins is seemingly at odds with their divergent functions. On some level, it is challenging to see the similarity between electron, metalloid, and hydrophobic tail-anchored protein transfer. Moreover, the sequence of steps of the ATPase cycle-mediated substrate transfer may not necessarily be conserved and may be specific to the mechanism of each IWA protein. Nonetheless, the ability of these proteins to transport various substrates seems less surprising considering

the core of the protein remains virtually unchanged and only the regions of the protein that harbor various cargo appeared to undergo considerable evolutionary alterations. Thus, the IWA proteins retain their most integral feature, which is substrate transport driven by ATP-dependent structural changes. We predict that further structural characterization of the substrate and nucleotide-bound states of other IWA proteins will reveal conservation of these mechanistic features. An interesting note is that Get3 is the only IWA protein known to date found in humans and yeast, unlike the many different IWA proteins found in bacteria and archeae. An outstanding question concerns the evolution of the IWA family in the P-loop NTPase evolutionary timeline and its functionally critical homodimeric architecture that has evolved to adapt remarkably diverse biological functions.

2.7 Supplementary information

Table S2.1: **Full names of organisms mentioned throughout the text.**

H. sapiens	*Homo sapiens*
E. coli	*Escherichia coli*
A. vinelandii	*Azotobacter vinelandii*
C. pasteurianum	*Clostridium pasteurianum*
M. acetivorans	*Methanosarcina acetivorans*
G. intestinalis	*Giardia intestinalis*
S. cerevisiae	*Saccharomyces cerevisiae*
M. jannaschii	*Methanococcus jannaschii*
A. multivorans	*Acidiphilium multivorans*
L. ferrooxydans	*Leptospirillum ferrooxydans*
P. furiosus	*Pyrococcus furiosus*
A. thaliana	*Arabidopsis thaliana*
S. sp. NOB8H2	*Sulfolobus species NOB8H2*
C. vibrioides	*Caulobacter vibrioides*
T. thermophilus	*Thermus thermophilus*

REFERENCES

[1] M. Saraste, P. R. Sibbald, and A. Wittinghofer. The P-loop — a common motif in ATP- and GTP-binding proteins. In: *Trends in Biochemical Sciences* 15.11 (Nov. 1990), pp. 430–434. DOI: 10.1016/0968-0004(90)90281-F.

[2] D. D. Leipe et al. Classification and evolution of P-loop GTPases and related ATPases. In: *Journal of Molecular Biology* 317.1 (Mar. 2002), pp. 41–72. DOI: 10.1006/jmbi.2001.5378.

[3] L. M. Longo et al. On the emergence of P-Loop NTPase and Rossmann enzymes from a Beta-Alpha-Beta ancestral fragment. In: *eLife* 9 (Dec. 2020), e64415. DOI: 10.7554/eLife.64415.

[4] I. R. Vetter and A. Wittinghofer. Nucleoside triphosphate-binding proteins: different scaffolds to achieve phosphoryl transfer. In: *Quarterly Reviews of Biophysics* 32.1 (Feb. 1999), pp. 1–56. DOI: 10.1017/s0033583599003480.

[5] J. Walker et al. Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. In: *The EMBO Journal* 1.8 (Aug. 1982), pp. 945–951. DOI: 10.1002/j.1460-2075.1982.tb01276.x.

[6] C. S. Mimura, S. R. Holbrook, and G. F. Ames. Structural model of the nucleotide-binding conserved component of periplasmic permeases. In: *Proceedings of the National Academy of Sciences* 88.1 (Jan. 1991), pp. 84–88. DOI: 10.1073/pnas.88.1.84.

[7] M. M. Georgiadis et al. Crystallographic Structure of the Nitrogenase Iron Protein from *Azotobacter* vinelandii. In: *Science* 257.5077 (Sept. 1992), pp. 1653–1659. DOI: 10.1126/science.1529353.

[8] M. I. Kozlova et al. Common Mechanism of Activated Catalysis in P-loop Fold Nucleoside Triphosphatases—United in Diversity. In: *Biomolecules* 12.10 (Sept. 2022), p. 1346. DOI: 10.3390/biom12101346.

[9] F. Jurnak. Structure of the GDP Domain of EF-Tu and Location of the Amino Acids Homologous to *ras* Oncogene Proteins. In: *Science* 230.4721 (Oct. 1985), pp. 32–36. DOI: 10.1126/science.3898365.

[10] S.-H. Kim et al. ras Oncogene Proteins: Three-dimensional Structures, Functional Implications, and a Model for Signal Transducer. In: *Cold Spring Harbor Symposia on Quantitative Biology* 53.0 (Jan. 1988), pp. 273–281. DOI: 10.1101/SQB.1988.053.01.034.

[11] L. Aravind et al. A novel family of P-loop NTPases with an unusual phyletic distribution and transmembrane segments inserted within the NTPase domain. In: *Genome Biology* 5.5 (Apr. 2004), R30. DOI: 10.1186/gb-2004-5-5-r30.

[12] M. G. Rossmann, D. Moras, and K. W. Olsen. Chemical and biological evolution of a nucleotide-binding protein. In: *Nature* 250.5463 (July 1974), pp. 194–199. DOI: 10.1038/250194a0.

[13] M. V. Milburn et al. Molecular Switch for Signal Transduction: Structural Differences Between Active and Inactive Forms of Protooncogenic *ras* Proteins. In: *Science* 247.4945 (Feb. 1990), pp. 939–945. DOI: 10.1126/science.2406906.

[14] M. I. Kozlova et al. Common Patterns of Hydrolysis Initiation in P-loop Fold Nucleoside Triphosphatases. In: *Biomolecules* 12.10 (Sept. 2022), p. 1345. DOI: 10.3390/biom12101345.

[15] I. Hanukoglu. Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites: Rossmann Fold in FAD, NAD and NADP Binding Domains. In: *Biochemistry and Molecular Biology Education* 43.3 (May 2015), pp. 206–209. DOI: 10.1002/bmb.20849.

[16] M. Motallebi-Veshareh, D. A. Rouch, and C. M. Thomas. A family of ATPases involved in active partitioning of diverse bacterial plasmids. In: *Molecular Microbiology* 4.9 (Sept. 1990), pp. 1455–1463. DOI: 10.1111/j.1365-2958.1990.tb02056.x.

[17] E. V. Koonin. A Superfamily of ATPases with Diverse Functions Containing Either Classical or Deviant ATP-binding Motif. In: *Journal of Molecular Biology* 229.4 (Feb. 1993), pp. 1165–1174. DOI: 10.1006/jmbi.1993.1115.

[18] J. Pei, B.-H. Kim, and N. V. Grishin. PROMALS3D: a tool for multiple protein sequence and structure alignments. In: *Nucleic Acids Research* 36.7 (Apr. 2008), pp. 2295–2300. DOI: 10.1093/nar/gkn072.

[19] J. D. Thompson et al. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. In: *Nucleic Acids Research* 25.24 (Dec. 1997), pp. 4876–4882. DOI: 10.1093/nar/25.24.4876.

[20] B. B. Wenke, T. Spatzal, and D. C. Rees. Site-Specific Oxidation State Assignments of the Iron Atoms in the [4Fe:4S] $^{2+/1+/0}$ States of the Nitrogenase Fe-Protein. In: *Angewandte Chemie International Edition* 58.12 (Mar. 2019), pp. 3894–3897. DOI: 10.1002/anie.201813966.

[21] M. Y. Fry et al. Structurally derived universal mechanism for the catalytic cycle of the tail-anchored targeting factor Get3. In: *Nature Structural & Molecular Biology* 29.8 (Aug. 2022), pp. 820–830. DOI: 10.1038/s41594-022-00798-4.

[22] T. Zhou. Structure of the ArsA ATPase: the catalytic subunit of a heavy metal resistance pump. In: *The EMBO Journal* 19.17 (Sept. 2000), pp. 4838–4845. DOI: 10.1093/emboj/19.17.4838.

[23] J. Lutkenhaus and M. Sundaramoorthy. MinD and role of the deviant Walker A motif, dimerization and membrane binding in oscillation. In: *Molecular Microbiology* 48.2 (2003), pp. 295–303. DOI: 10.1046/j.1365-2958.2003.03427.x.

[24] C. M. Chen et al. Nucleotide sequence of the structural genes for an anion pump. The plasmid-encoded arsenical resistance operon. In: *Journal of Biological Chemistry* 261.32 (Nov. 1986), pp. 15030–15038. DOI: 10.1016/S0021-9258(18)66824-3.

[25] H. Schindelin et al. Structure of ADP·AIF4—stabilized nitrogenase complex and its implications for signal transduction. In: *Nature* 387.6631 (May 1997), pp. 370–376. DOI: 10.1038/387370a0.

[26] M. R. Ahmadian et al. Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP-hydrolysis reaction of Ras. In: *Nature Structural Biology* 4.9 (Sept. 1997), pp. 686–689. DOI: 10.1038/nsb0997-686.

[27] B. Schmid et al. Biochemical and Structural Characterization of the Cross-Linked Complex of Nitrogenase: Comparison to the ADP-AlF -Stabilized Structure . In: *Biochemistry* 41.52 (Dec. 2002), pp. 15557–15565. DOI: 10.1021/bi026642b.

[28] T. Zhou et al. Conformational changes in four regions of the Escherichia coli ArsA ATPase link ATP hydrolysis to ion translocation. In: *Journal of Biological Chemistry* 276.32 (2001), pp. 30414–22. DOI: 10.1074/jbc.M103671200.

[29] H. B. Gristick et al. Crystal structure of ATP-bound Get3–Get4–Get5 complex reveals regulation of Get3 by Get4. In: *Nature Structural & Molecular Biology* 21.5 (May 2014), pp. 437–442. DOI: 10.1038/nsmb.2813.

[30] A. F. A. Keszei et al. Structural insights into metazoan pretargeting GET complexes. In: *Nature Structural & Molecular Biology* 28.12 (Dec. 2021), pp. 1029–1037. DOI: 10.1038/s41594-021-00690-7.

[31] X. Ruan, H. Bhattacharjee, and B. P. Rosen. Cys-113 and Cys-422 form a high affinity metalloid binding site in the ArsA ATPase. In: *Journal of Biological Chemistry* 281.15 (2006), pp. 9925–34. DOI: 10.1074/jbc.M600125200.

[32] A. R. Walmsley et al. Antimonite regulation of the ATPase activity of ArsA, the catalytic subunit of the arsenical pump. In: *Biochemical Journal* 360.3 (Dec. 2001), pp. 589–597. DOI: 10.1042/bj3600589.

[33] F. A. Tezcan et al. Nitrogenase Complexes: Multiple Docking Sites for a Nucleotide Switch Protein. In: *Science* 309.5739 (Aug. 2005), pp. 1377–1380. DOI: 10.1126/science.1115653.

[34] U. S. Chio et al. A protean clamp guides membrane targeting of tail-anchored proteins. In: *Proceedings of the National Academy of Sciences* 114.41 (Oct. 2017). DOI: 10.1073/pnas.1708731114.

[35] Z. Hu and J. Lutkenhaus. Topological Regulation of Cell Division in E. coli: Spatiotemporal Oscillation of MinD Requires Stimulation of Its ATPase by MinE and Phospholipid. In: *Molecular Cell* 7.6 (June 2001), pp. 1337–1343. DOI: 10.1016/S1097-2765(01)00273-8.

[36] W. N. Lanzilotta, M. J. Ryle, and L. C. Seefeldt. Nucleotide Hydrolysis and Protein Conformational Changes in Azotobacter vinelandii Nitrogenase Iron Protein: Defining the Function of Aspartate 129. In: *Biochemistry* 34.34 (Aug. 1995), pp. 10713–10723. DOI: 10.1021/bi00034a003.

[37] D. Wolle, D. R. Dean, and J. B. Howard. Nucleotide-Iron-Sulfur Cluster Signal Transduction in the Nitrogenase Iron-Protein: the Role of Asp ^{125}. In: *Science* 258.5084 (Nov. 1992), pp. 992–995. DOI: 10.1126/science.1359643.

[38] P. Rangaraj et al. In Vitro Biosynthesis of Iron-Molybdenum Cofactor and Maturation of the nif-encoded Apodinitrogenase. In: *Journal of Biological Chemistry* 274.28 (July 1999), pp. 19778–19784. DOI: 10.1074/jbc.274.28.19778.

[39] Z. Hu, C. Saez, and J. Lutkenhaus. Recruitment of MinC, an Inhibitor of Z-Ring Formation, to the Membrane in *Escherichia* coli : Role of MinD and MinE. In: *Journal of Bacteriology* 185.1 (Jan. 2003), pp. 196–203. DOI: 10.1128/JB.185.1.196-203.2003.

[40] W. Wu et al. Determination of the structure of the MinD-ATP complex reveals the orientation of MinD on the membrane and the relative location of the binding sites for MinE and MinC: Structure of MinD-ATP complex. In: *Molecular Microbiology* 79.6 (Mar. 2011), pp. 1515–1528. DOI: 10.1111/j.1365-2958.2010.07536.x.

[41] D. Ghosal et al. MinCD cell division proteins form alternating copolymeric cytomotive filaments. In: *Nature Communications* 5.1 (Dec. 2014), p. 5341. DOI: 10.1038/ncomms6341.

[42] J. Lutkenhaus. The ParA/MinD family puts things in their place. In: *Trends in Microbiology* 20.9 (Sept. 2012), pp. 411–418. DOI: 10.1016/j.tim.2012.05.002.

[43] H. M. Segal et al. Electrochemical and structural characterization of Azotobacter vinelandii flavodoxin II. In: *Protein Science: A Publication of the Protein Society* 26.10 (Oct. 2017), pp. 1984–1993. DOI: 10.1002/pro.3236.

[44] B. P. Rosen et al. Mechanism of the ArsA ATPase. In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1461.2 (Dec. 1999), pp. 207–215. DOI: 10.1016/S0005-2736(99)00159-5.

[45] Y.-F. Lin, A. R. Walmsley, and B. P. Rosen. An arsenic metallochaperone for an arsenic detoxification pump. In: *Proceedings of the National Academy of Sciences* 103.42 (Oct. 2006), pp. 15617–15622. DOI: 10.1073/pnas.0603974103.

[46] T. Hurek et al. Induction of complex intracytoplasmic membranes related to nitrogen fixation in *Azoarcus* sp. BH72. In: *Molecular Microbiology* 18.2 (Oct. 1995), pp. 225–236. DOI: 10.1111/j.1365-2958.1995.mmi_18020225.x.

[47] G. Bange and I. Sinning. SIMIBI twins in protein targeting and localization. In: *Nature Structural & Molecular Biology* 20.7 (July 2013), pp. 776–780. DOI: 10.1038/nsmb.2605.

[48] J. B. Howard and D. C. Rees. NITROGENASE: A Nucleotide-Dependent Molecular Switch. In: *Annual Review of Biochemistry* 63.1 (June 1994), pp. 235–264. DOI: 10.1146/annurev.bi.63.070194.001315.

[49] F. A. Tezcan et al. Structural Evidence for Asymmetrical Nucleotide Interactions in Nitrogenase. In: *Journal of the American Chemical Society* 137.1 (Jan. 2015), pp. 146–149. DOI: 10.1021/ja511945e.

[50] J. L. Schlessman et al. Conformational variability in structures of the nitrogenase iron proteins from Azotobacter vinelandii and Clostridium pasteurianum. In: *Journal of Molecular Biology* 280.4 (July 1998), pp. 669–685. DOI: 10.1006/jmbi.1998.1898.

[51] C. L. Chang et al. Characterization of nifH mutations of Klebsiella pneumoniae. In: *Journal of Bacteriology* 170.9 (Sept. 1988), pp. 4015–4022. DOI: 10.1128/jb.170.9.4015-4022.1988.

[52] D. Pednekar, A. Tendulkar, and S. Durani. Electrostatics-defying interaction between arginine termini as a thermodynamic driving force in protein-protein interaction. In: *Proteins: Structure, Function, and Bioinformatics* 74.1 (Jan. 2009), pp. 155–163. DOI: 10.1002/prot.22142.

[53] D. Lee, J. Lee, and C. Seok. What stabilizes close arginine pairing in proteins? In: *Physical Chemistry Chemical Physics* 15.16 (2013), p. 5844. DOI: 10.1039/c3cp00160a.

[54] R. M. Vernon et al. Pi-Pi contacts are an overlooked protein feature relevant to phase separation. In: *eLife* 7 (Feb. 2018), e31486. DOI: 10.7554/eLife.31486.

[55] C. J. M. Suloway et al. Model for eukaryotic tail-anchored protein binding based on the structure of Get3. In: *Proceedings of the National Academy of Sciences* 106.35 (Sept. 2009), pp. 14849–14854. DOI: 10.1073/pnas.0907522106.

[56] N. Osaka et al. Divergent Mechanisms Activating RAS and Small GTPases Through Post-translational Modification. In: *Frontiers in Molecular Biosciences* 8 (July 2021), p. 707439. DOI: 10.3389/fmolb.2021.707439.

[57] R. Baker et al. Differences in the Regulation of K-Ras and H-Ras Isoforms by Monoubiquitination. In: *Journal of Biological Chemistry* 288.52 (Dec. 2013), pp. 36856–36862. DOI: 10.1074/jbc.C113.525691.

[58] K. A. Denessiouk and M. S. Johnson. When fold is not important: A common structural framework for adenine and AMP binding in 12 unrelated protein families. In: *Proteins: Structure, Function, and Genetics* 38.3 (Feb. 2000), pp. 310–326.

Chapter 3

NUCLEOTIDE- AND METALLOID-DRIVEN CONFORMATIONAL CHANGES IN THE ARSENITE EFFLUX ATPASE ARSA

This chapter is adapted and reprinted from:
Mahajan, S.; Pall, A.E.; Li, Y.E.; Stemmler, T. L.; Rees, D.C.; Clemons Jr., W.M. Nucleotide and metalloid-driven conformational changes in the arsenite efflux ATPase ArsA. *Proceedings of the National Academy of Sciences* 122 (35) e2506440122. (2025). Doi: 10.1073/pnas.2506440122
S.M., D.C.R., and W.M.C. designed research; S.M. and Y.E.L. performed structural characterization; S.M. performed biochemical assays; S.M., A.E.P., and T.L.S., performed X-ray absorption experiments and analysis; S.M., D.C.R., and W.M.C. co-wrote and co-edited the paper.

Abstract

Arsenite (As^{III}) is toxic to all organisms due to its ability to tightly bind exposed thiols within cells. An important As^{III} resistance mechanism in prokaryotes involves proteins encoded by the ars operon. A central component of the *ars* operon in many bacteria is the cytoplasmic ATPase, ArsA, which orchestrates a series of nucleotide-dependent handoffs, starting with the capture of As^{III} by the ArsD metallochaperone and culminating in its removal from the cell by the ArsB efflux pump. Although the mechanism of ArsA has been widely studied, the molecular details of how nucleotide hydrolysis modulates these events remain unclear. ArsA is an archetypal member of the intradimeric Walker A (IWA) family of ATPases, implicated in a diversity of complex biological functions. Conformational changes typical of IWA ATPases have been postulated to drive these molecular events but have not been demonstrated. We report cryogenic electron microscopy (cryo-EM) structures of ArsA in MgADP-bound and MgATP-bound 'open' states, as well as a distinct 'closed' MgATP-bound state liganded to As^{III}. X-ray absorption spectroscopy confirmed three-coordinate binding of As^{III} to the conserved cysteines at the metalloid-binding site of the 'closed' state. Coupled with biochemical characterization, our cryo-EM structures reveal key conformational changes in the ArsA catalytic cycle consistent with other IWA ATPases and provide the structural basis for allosteric activation of nucleotide hydrolysis by As^{III}. This work establishes how the nucleotide state of ArsA transiently creates a high-affinity binding site that can sequester metalloid within the cell, followed by a nucleotide-driven handoff to ArsB for efflux.

3.1 Introduction

P-loop NTPases, representing one of the most prevalent nucleotide-binding folds in nature, catalyze nucleotide triphosphate (NTP) hydrolysis to facilitate a broad range of biological functions, ranging from transcription regulation to nitrogen fixation [1, 2]. One distinct lineage in the superfamily is the 'intradimeric Walker A' (IWA) ATPases, which are characterized by a unique consensus P-loop sequence (X-K-X-X-X-G-K-[T/S]), formerly referred to as the 'deviant Walker A' ATPase family [3, 4]. These homodimeric ATPases couple ATP binding and hydrolysis to the transfer of a cargo (or substrate) to a partner protein involved in a distinct biological function. The conserved Lys residue at the N-terminus of the Walker A motif, referred to as the 'IWA lysine', is critical for triggering nucleotide hydrolysis in each nucleotide binding site across the dimer interface [5, 6], much in the same way as the 'arginine finger' motif of the G-protein-specific GTPase-activating protein [7]. IWA ATPases mediate diverse cellular functions and have been mechanistically studied to understand their respective biological processes, including, but not limited to, the tail-anchored (TA) membrane protein targeting factor Get3, the reductase component of nitrogenase NifH, and the spatial regulation of bacterial cell division protein MinD [4].

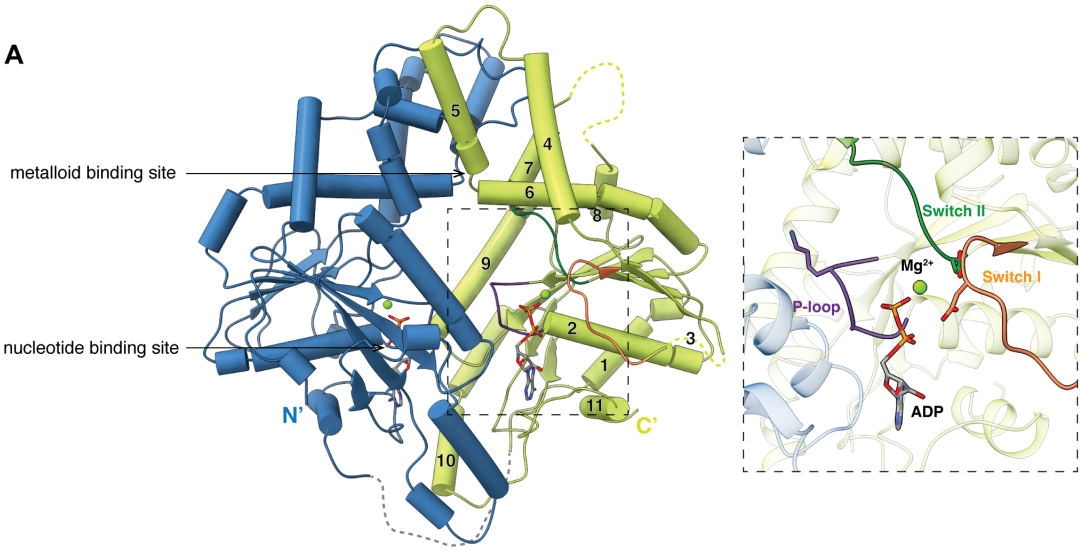
A distinct member of the IWA family is the cytoplasmic ATPase ArsA, which is involved in the bacterial arsenic detoxification machinery. Trivalent arsenite (As^{III}) induces toxicity in organisms across domains of life by tightly binding to thiols of cysteine residues, resulting in inactivation of thiol-containing molecules, including lipoic acid and enzymes, and inducing protein misfolding and aggregation [8, 9]. As such, arsenic toxicity is a widespread global health concern, and understanding the molecular mechanisms of detoxification has garnered broad interest due to their potential for arsenic bioremediation of contaminated environments [10, 11]. In bacteria and some archaea, resistance to As^{III} is encoded in the ars operon, which involves the membrane efflux of As^{III} by the integral membrane protein ArsB [12, 13]. While diverse combinations of ars genes are found across species, ArsB is most commonly found in association with genes for a DNA repressor, ArsR, which regulates the operon, and an aerobic reductase, ArsC, that reduces pentavalent AsV to As^{III}. In many species, the operon additionally encodes the ArsA ATPase and the metallochaperone ArsD. Under arsenic stress, ArsD sequesters As^{III} from the cytoplasm and transfers the toxic metalloid to ArsA in a nucleotide-dependent manner [14, 15]. Subsequently, ArsA transfers As^{III} to ArsB by coupling metalloid transport to ATP hydrolysis. Notably, in the presence of ArsA, ArsB is considerably

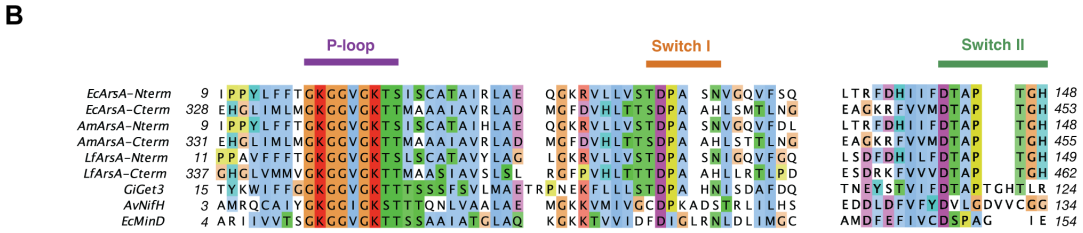
more efficient at exporting As^{III} [16] and has been collectively referred to as the 'ArsAB pump' [17, 18]. This mode of resistance enables the bacteria to survive a wide range of environmentally contaminating As^{III} concentrations [14, 16].

In contrast to other homodimeric IWA ATPases, ArsA is the only known member of the family that is pseudodimeric, comprising two homologous nucleotide-binding domains covalently linked by a flexible 25-residue linker [18, 19] (Fig. 3.1A). The metalloid-binding site, composed of three conserved cysteine residues [20], is located at the pseudodimer interface and can bind As^{III} or antimonite (Sb^{III}). Both domains of ArsA, referred to as the N- and C-domains, can support the binding and hydrolysis of ATP; however, the N-domain does so at a significantly higher rate than the C-domain [21]. Despite considerable work on understanding the catalytic mechanism of ArsA, the molecular details of its nucleotide-dependent mechanism, especially in the context of the broader IWA family, remain to be understood.

Nucleotide-dependent conformational changes around the dimer interface are a conserved mechanistic feature among IWA ATPases [4]. The conformational landscapes of Get3 and NifH have been extensively characterized. The catalytic cycles of both ATPases involve conformational switching of the dimer between a 'closed' state that is catalytically competent for ATP hydrolysis and an 'open' state that is ATPase inactive [5, 6, 24–27]. The 'open' conformation is reported for either the apo or post-hydrolysis MgADP-bound states and is characterized by P-loops that are far apart [26, 28]. In contrast, the 'closed' state is reported for the pre- or mid-hydrolysis states where the two P-loops are adjacent [5, 6], mediating the position of the IWA lysines across the dimer interface. Two additional loop motifs, Switch I and II (Fig. 3.1A and B), are conserved across IWA ATPases and adopt discrete conformations depending on the nucleotide state. These nucleotide-dependent conformational transitions are critical for the function of IWA ATPases, as ATP binding and hydrolysis are functionally coupled to the targeting of a substrate specific to each enzyme [4].

An unresolved question concerns how nucleotide-dependent conformational transitions during the ArsA catalytic cycle facilitate As^{III} capture and efflux. Biochemical and spectroscopic characterizations of ArsA suggest that the ATPase undergoes a series of conformational changes in the presence of nucleotide and metalloid [29, 30]. X-ray crystal structures of *E. coli* ArsA (*Ec*ArsA) in multiple nucleotide-bound states reveal only subtle conformational changes between the various structures with a root mean square deviation (r.m.s.d.) of 0.38 Å over 536 residues (Fig. 3.1C) (14,





C

***E. coli* ArsA X-ray crystal structures**

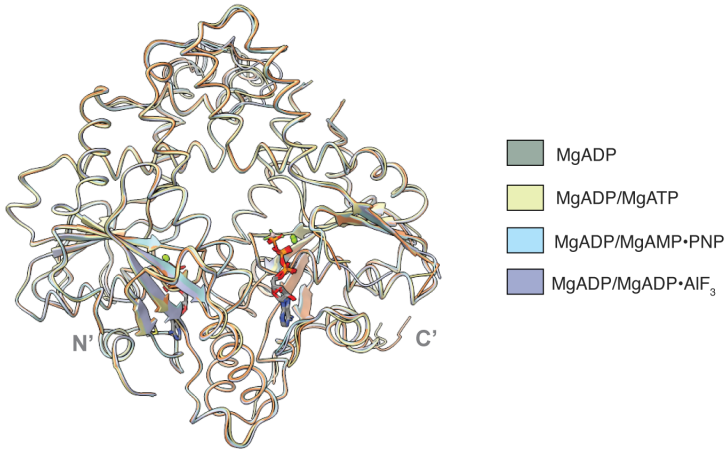


Figure 3.1: **Arsenite efflux ATPase ArsA from *E. coli*.** (A) *Ec*ArsA structure bound to MgADP (PDB: 1F48) highlighting the N- (blue) and C- (green) domains, the nucleotide-binding sites and the metalloid-binding site. The inset highlights the P-loop (violet), Switch I (orange) and Switch II (green) motifs of the C-domain. Helices are numbered 1 through 11 on the C-domain, and have homologous helices on the N-domain. Helix 3 is not modeled in the C-domain of this structure, but the N-domain helix 3 is shown. (B) Multiple sequence alignment of N- and C-domains of ArsA homologues from *E. coli* (*Ec*ArsA), *Acidiphilium multivorum* (*Am*ArsA), and *Leptospirillum ferriphilum* (*Lf* ArsA) with other IWA ATPases — Get3 from *Giardia intestinalis* (*Gi*Get3), Nitrogenase Fe protein from A*zotobacter vinelandii* (*Av*NifH), and bacterial cell division regulatory ATPase from *E. coli* (*Ec*MinD) highlighting the conserved IWA motifs as shown in A. Alignment was constructed using Promals3D [22] and residues were colored using the ClustalX scheme [23]. (C) Global alignment of all X-ray crystal structures of *Ec*ArsA bound to various nucleotide — MgADP (PDB: 1F48), MgADP/MgATP (PDB: 1II0), MgADP/MgAMP•PNP (PDB: 1II9), and MgADP/MgADP•AlF_3 (PDB: 1IHU). MgADP is bound at the N-domain and a different nucleotide is bound at the C-domain in each structure. Sb^{III}/Cd2+ and Cl- ions are modeled at the metalloid-binding sites in each of these structures but were omitted from this representation.

29), supporting that crystallization has likely restricted the range of motion in the protein. Kinetic assays have shown that the metalloid substrate activates ArsA by stimulating ATP hydrolysis, which is the proposed rate-limiting step of the enzyme [30]. Based on biochemical studies with Sb^{III}, the metalloid substrate binds at a single high-affinity site in ArsA [31]. Of the three conserved cysteines, Cys172, located on a mobile loop between helices 7 and 8 of the N-domain of *Ec*ArsA, has been proposed to control the metalloid affinity of this site [31, 32]. As the crystal structures are not consistent with the results above, the mechanism of metalloid binding and ArsA activation requires further investigation.

In this chapter, we report single-particle cryogenic electron microscopy (cryo-EM) structures of ArsA from the thermotolerant and acidophilic bacterium *Leptospirillum ferriphilum*, in nucleotide-bound 'open' and 'closed' conformations. In the MgADP-bound 'open' state, the ArsA pseudodimer adopts the conformation of the previously reported ArsA crystal structures, all representing an ATPase inactive 'open' state with low affinity for As^{III}. MgATP alone is unable to stabilize a major conformational change from this 'open' state. In the presence of MgATP and As^{III}, a distinct 'closed' conformation with high As^{III} affinity is adopted that is analogous to the 'closed' conformations observed for other IWA ATPases. These structures provide views of the large-scale conformational transitions in the ArsA catalytic

cycle that were previously not revealed in the crystal structures. Combined with X-ray absorption spectroscopy (XAS) analysis and supporting biochemical studies, our cryo-EM structures reveal the As^{III} activation mechanism of ATP hydrolysis in ArsA. Altogether, this work elucidates the nucleotide- and metalloid-dependent mechanism that enables a series of handoffs of toxic metalloids between ArsA and its partner proteins ArsD and ArsB by modulating metalloid affinity to efficiently carry out efflux, thus placing the protein within the mechanistic landscape of the biologically important IWA ATPase family.

3.2 Results

3.2.1 ArsA ATPase from *Leptospirillum ferriphilum* strain ML-04

Following a bioinformatic analysis of ars operons resembling the well-studied *E. coli* R773 plasmid ars operon [13], we identified an ArsA homolog from *Leptospirillum ferriphilum* strain ML-04 (*Lf*ArsA), an acidophilic and moderately thermophilic bacterial species, where some strains are found in arsenic-rich acid mine drainage ecosystems (Fig. S3.1A) [33]. The *arsA* gene has been shown to express under arsenic stress conditions in this strain, and the operon confers arsenic resistance in closely related *L. ferriphilum* strains [33–35]. *Lf*ArsA has 69% sequence identity to *Ec*ArsA, retaining all the characteristic features of ArsA (Fig. 3.1B). We cloned, overexpressed, and purified *Lf*ArsA with a C-terminal 6xHis tag. Purified *Lf*ArsA hydrolyzes ATP *in vitro* with a pseudo first-order rate constant of 0.5 ± 0.1 min^{-1}, and saturating concentrations of As^{III} stimulate the ATPase activity ~15-fold above the basal activity (Fig. S3.1B, black plot). To elucidate the conformational landscape of the ArsA catalytic cycle, we sought to determine nucleotide-bound *Lf*ArsA structures both in the presence and absence of its substrate As^{III} using cryo-EM.

3.2.2 Structure of the 'open' conformation of ArsA bound to MgADP

To determine the conformation of ArsA in an MgADP- and As^{III}-bound state, ArsA was incubated with 2 mM $MgCl_2$, 2 mM ADP, and 5 mM sodium arsenite, and grids were prepared for single-particle analysis. From a dataset of 4,732 movies collected on a 300 kV Titan Krios Transmission Electron Microscope (TEM) and several rounds of 2D and 3D classification (Fig. S3.2), we obtained a 103,258-particle reconstruction of the ArsA·MgADP state (Fig. 3.2A). The final Coulomb potential map was refined to a 3.4 Å gold-standard Fourier shell correlation (FSC) resolution and sharpened using a uniform B-factor of -165 $Å^2$ to better visualize high-resolution features. AlphaFold2 [36] predictions for the N- and C-domains

of *Lf* ArsA were docked into the sharpened map, followed by model building and refinement to obtain the ArsA·MgADP structure (Fig. 3.2B). The average Q-score for the structure was 0.64 (Fig. S3.3A), indicating good resolvability of the cryo-EM map [37]. About 95% of the protein chain was modeled into the map, with missing residues including part of the linker region between the N- and C-domains (residues 295–307), residues 167–168, and residues 464–476 (Fig. 3.2B). The N (residues 1–298) and C (residues 315–587) domains are pseudo-symmetric with an r.m.s.d. of 1.0 Å over 139 residues (Fig. 3.2C).

At both nucleotide-binding sites, ADP is coordinated to an Mg^{2+} ion (Fig. 3.2D). The catalytic Switch I aspartates — Asp46 (N) and Asp364 (C) — are oriented away from the Mg^{2+} ion (>4.5 Å). The Switch II aspartates — Asp143 (N) and Asp447 (C) — interact with the Mg^{2+} ion within 4 Å and form hydrogen bonds with Thr23 (N) and Thr341 (C), respectively. Notably, the P-loops are sufficiently separated such that the catalytic IWA lysines (Lys17 and Lys335) are oriented away from the ADP bound to the opposing domain, positioning the terminal amino groups of the lysines over 8 Å away from the nucleotide phosphate groups (Fig. 3.2E).

We propose that our ArsA•MgADP structure represents an 'open' form of ArsA, designated as the *ADP-open* state for the rest of the discussion. Based on the general mechanism of the IWA ATPase family, this conformation represents an inactive form of ArsA unable to hydrolyze ATP. Comparison of the cryo-EM structure to the X-ray crystal structures of *Ec*ArsA reveals the structures adopt similar conformations (Fig. 3.1C); for example, the cryo-EM structure and the *Ec*ArsA·MgADP crystal structure (PDB: 1F48) have an r.m.s.d. of 0.95 Å over 475 residues. Despite the presence of millimolar concentrations of As^{III} in the samples, there was no evidence of As^{III} at the metalloid-binding site in either our EM or the X-ray structures (Fig. S3.4) [18, 38].

3.2.3 Structures of nucleotide-bound ArsA in the absence of As^{III}

To clarify the absence of As^{III} at the metalloid-binding site in the ArsA•MgADP structure, we solved a cryo-EM structure using identical experimental conditions but in the absence of As^{III}. The resulting 3.8 Å resolution map of the ArsA•MgADP state is in the 'open' conformation (Figs. 3.2F and S3.3B and S3.5), similar to the structure solved in the presence of As^{III}, with an r.m.s.d. of 0.57 over 545 residues. Importantly, the configuration of the metalloid-binding site remains unchanged (Fig. 3.2F; inset on right), confirming that both states lack As^{III}. This suggests that the

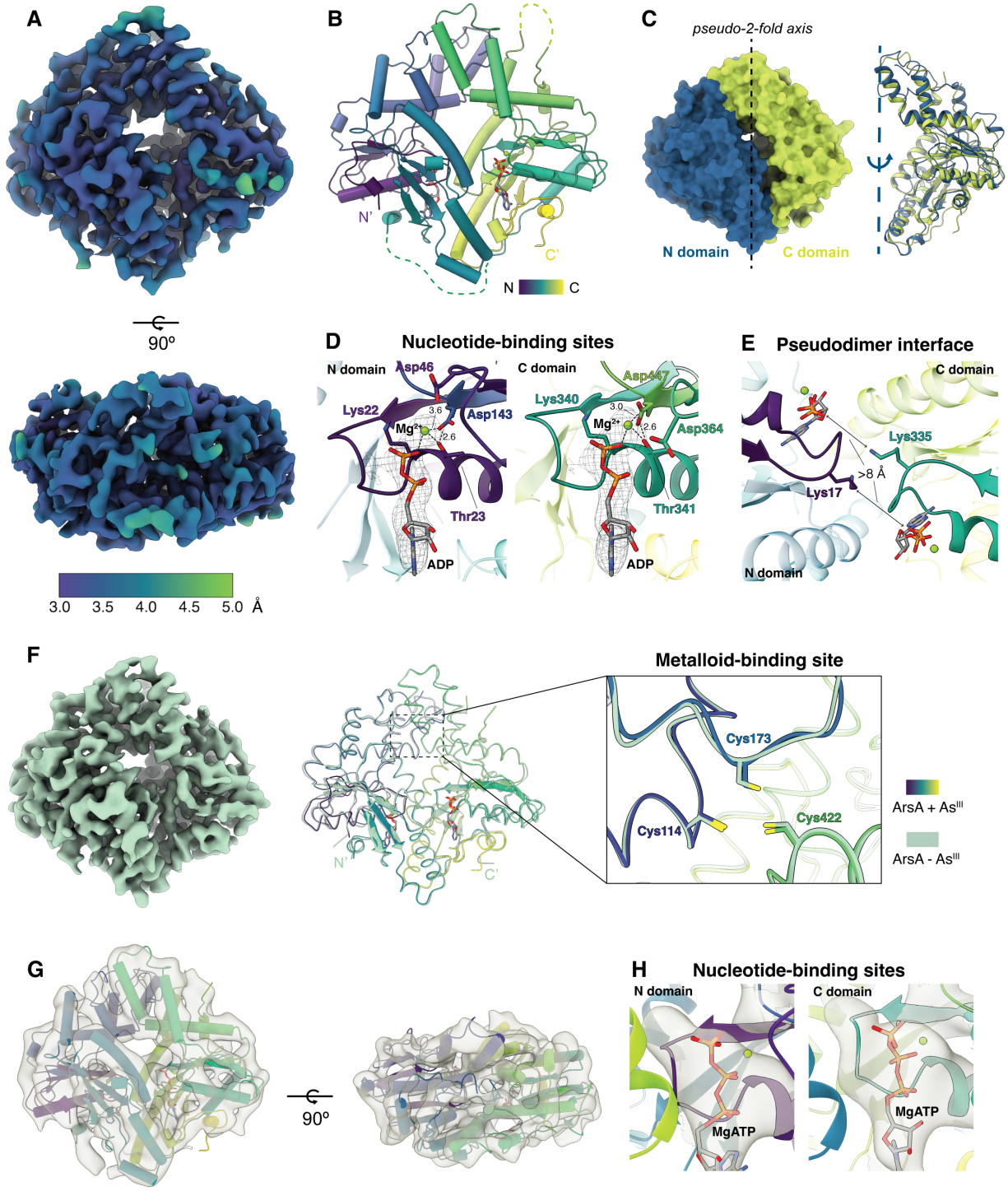

A
90°
3.0 3.5 4.0 4.5 5.0 Å
B
N'
C'
N C
C
pseudo-2-fold axis
N domain
C domain
D Nucleotide-binding sites
N domain
Asp46
Lys22
Asp143
Mg2+
Thr23
ADP
C domain
Asp447
Lys340
Asp364
Thr341
E Pseudodimer interface
C domain
Lys335
>8 Å
Lys17
N domain
F
Metalloid-binding site
Cys173
Cys114
Cys422
ArsA + AsIII
ArsA - AsIII
G
90°
H Nucleotide-binding sites
N domain
MgATP
C domain
MgATP

Figure 3.2: **Cryo-EM structures of *Lf*ArsA in the nucleotide-bound 'open' conformation.** (A) Two views of the sharpened cryo-EM map colored based on local resolution (3.4 Åoverall) for *Lf*ArsA•MgADP 'open' state (ADP-open) in the presence of As^{III}. (B) Cartoon representation of the ADP-open state colored in viridis. Unmodeled regions are indicated as dotted lines. (C) Left, surface representation of the ArsA pseudodimer highlighting the N- (blue) and C- (green) domains. Right, cartoon representation of the overlay of the two domains based on the pseudo-2-fold symmetry. (D) Nucleotide-binding sites of the ADP-open state showing the Coulomb potential map around the MgADP shown as sticks. Switch I II Asp residues of both N- and C-domains are shown as sticks. (E) The pseudodimer interface of the ADP-open state highlighting the distance between the IWA lysines. (F) Structure of *Lf*ArsA•MgADP 'open' state in the absence of As^{III}. Left, 3.8 Åcryo-EM map; middle, ribbons model (pale green) overlayed with the ADP-open state model from B; right, the metalloid-binding sites from both structures. (G) Cryo-EM map (6-7 Åoverall resolution) of the *Lf*ArsA non-hydrolyzing variant (D46N/D364N) solved in the presence of MgATP with the ADP-open state model from B docked in the map. (H) Nucleotide-binding sites of the *Lf*ArsA•MgATP state showing Coulomb potential map likely corresponding to ATP. MgATP was manually fit using Coot.

ADP-open state does not effectively bind the metalloid.

The presence of As^{III} stimulates steady-state ATP hydrolysis by ArsA above the basal activity (Fig. S3.1B) [29, 31]. To characterize the ATP-bound ArsA conformation in the absence of metalloid, we generated a non-hydrolyzing variant of *Lf*ArsA by mutating the catalytic Switch I aspartates — Asp46 and Asp364 of the N- and C-domains, respectively — to asparagines (Fig. S3.1B, red plot). We used this variant to capture the MgATP-bound state, similar to the approach used to obtain the MgATP-bound Get3 structure (PDB: 7SPY) [5]. For the rest of the discussion, structures solved using the D46N/D364N variant will be simply referred to as ArsA, and the Asn residue will be simply referred to as Asp residue. Under experimental conditions similar to those used to solve the ADP-open structures, we could only obtain a low-resolution (6–7 Å) EM map that revealed an 'open' state of ArsA (Fig. S3.6). We could confidently fit the model of the ArsA•MgADP structure into the map (Fig. 3.2G). While the density observed at the nucleotide-binding sites could not be unambiguously distinguished at this resolution, it is consistent with MgATP binding (Fig. 3.2H). This is reminiscent of NifH that is refractory to crystallization in the presence of ATP but readily crystallizes in the presence of both ATP and its partner protein, NifDK (MoFe protein) [28]. It is plausible that binding of ATP makes ArsA conformationally flexible, preventing high-resolution cryo-

EM reconstruction, and that a partner protein or As^{III} is required for stabilization. Importantly, this structure is consistent with the low basal ATPase activity of ArsA. Together, these As^{III}-free structures support that nucleotide binding alone is unable to stabilize a conformation competent for ATP hydrolysis.

3.2.4 Structure of the 'closed' conformation of ArsA bound to MgATP and As^{III}

To understand the structural basis of As^{III} activation, we sought to characterize the conformation of ArsA in the presence of both MgATP and As^{III}. We incubated the non-hydrolyzing ArsA variant with 2 mM $MgCl_2$, 2 mM ATP, and 2 mM sodium arsenite and then performed single-particle cryo-EM (Fig. S3.7). The resulting Coulomb potential map was resolved to an overall gold-standard FSC resolution of 3.0 Å and sharpened using a uniform B-factor of -124 $Å^2$ (Fig. 3.3A). AlphaFold2 models of N- and C-domains of *Lf*ArsA were docked into the sharpened density map. The model was subsequently rebuilt and refined to obtain the structure of ArsA•MgATP•As^{III} with clear density for MgATP at both nucleotide-binding sites (Fig 3.3B and C). Significantly, at the metalloid-binding site, we observed density enclosed between the three conserved cysteines — Cys114, Cys173, and Cys422 — into which we modeled an arsenic atom (Fig. 3.3D). The average Q-score for the structure was 0.71 (Fig. S3.3C), indicating high resolvability of the cryo-EM map. About 97% of the protein was modeled into the map, while the regions with poor density, such as most of the linker region (residues 299-305) and residues 474-479, were omitted from the final model.

Within both nucleotide-binding sites, ATP is stabilized by electrostatic interactions of the γ-phosphate group with Mg^{2+} and Lys (22 and 340, respectively), and by hydrogen bonds involving the α- and β-phosphates with the P-loop backbone (Figs. 3.3C and Fig. S3.8A). The ribose sugar and the adenine base are stabilized by hydrogen bonds contributed by sidechains from both domains, and by backbone hydrogen bonds from the loop region corresponding to the adenine-binding loop (A-loop), as described in Get3 structures (Fig. S3.8A) [26]. Additionally, ATP bound to the N-domain is stabilized by potential cation-π interactions formed between the adenine base and both Arg207 (N) and Arg544 (C) (Fig. S3.8A, left). However, in the C-domain, the corresponding interaction is only observed between the adenine and Arg256 (N). In this structure, the Switch I aspartates (mutated to asparagines here) — Asp46 (N) and Asp364 (C) — are positioned adjacent to each Mg^{2+} ion and anchor the catalytic water molecule (W_{cat}) near the γ-phosphate of ATP via a 3.0

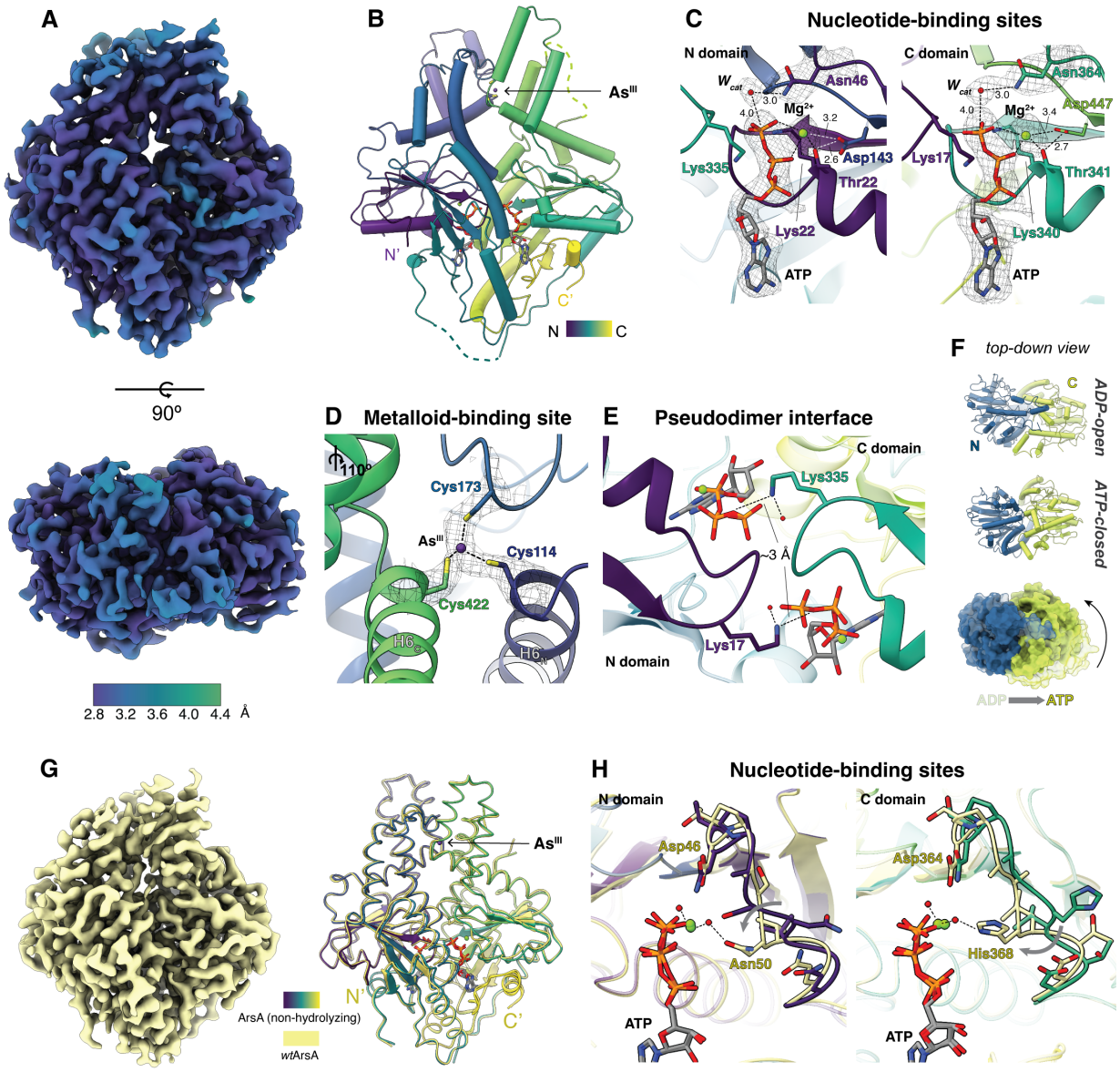

A
90°
2.8 3.2 3.6 4.0 4.4 Å
B
As^III
N'
C'
N
C
C Nucleotide-binding sites
N domain
C domain
W cat
Mg2+
Asn46
Asp143
Lys335
Thr22
Lys22
ATP
Asn364
Asp447
Lys17
Thr341
Lys340
D Metalloid-binding site
110°
Cys173
As^III
Cys114
Cys422
H6c
H6N
E Pseudodimer interface
C domain
Lys335
~3 Å
Lys17
N domain
F top-down view
C
N
ADP-open
ATP-closed
ADP
ATP
G
As^III
N'
C'
ArsA (non-hydrolyzing)
wtArsA
H Nucleotide-binding sites
N domain
Asp46
Asn50
ATP
C domain
Asp364
His368
ATP

Figure 3.3: **Cryo-EM structure of *Lf*ArsA 'closed' conformation bound to MgATP and As^{III}.** (A) Two views of the sharpened cryo-EM map colored based on local resolution (3.0 Å overall) for *Lf*ArsA•MgATP•As^{III} 'closed' state (ATP-closed) of the non-hydrolyzing variant (D46N/D364N). (B) Cartoon representation of the ATP-closed state colored in viridis. Unmodeled regions are indicated as dotted lines. (C) Nucleotide-binding sites of the ATP-closed state showing the Coulomb potential map around MgATP and Switch I catalytic residues (Asn46 and Asn364) shown as sticks. Switch II Asp residues of both N- and C-sites are shown as sticks. The nucleophilic water molecule (W_{cat}) is highlighted. (D) Metalloid-binding site of ATP-closed state showing Coulomb potential map for As^{III} (purple atom) coordinated by Cys114, Cys173 and Cys422. The view is rotated 110º clockwise relative to B. (E) The pseudodimer interface of the ATP-closed state highlighting the distance between the IWA lysines (Lys17 and Lys335). (F) Top-down view, relative to A, comparing the ADP-open and the ATP-closed states of *Lf*ArsA colored as in Fig. 3.2C. showing both states in cartoon and then overlayed in surface representation. The overlay is aligned to the N-domain P-loop (residues 16-23) showing the relative change of the C-domain. (G) Structure of *Lf*ArsA•MgATP•As^{III} 'closed' state solved using wild-type *Lf*ArsA. Left, 3.0 Åmap cryo-EM map; right, ribbons model (yellow) overlayed with the model from B. (H) Comparison of Switch I conformations between the two ATP-closed structures. Differences in interactions are highlighted.

Å hydrogen bond for nucleophilic attack, priming the enzyme for hydrolysis (Fig. 3.3C). Switch II aspartates — Asp143 (N) and Asp447 (C) — are positioned at 3.2 and 3.4 Å from the Mg^{2+} ion, respectively and form short hydrogen bonds with the Thr residue coordinated to the respective Mg^{2+} ions, similar to the MgADP-bound structure.

This structure adopts a 'closed' pseudodimer conformation previously unobserved for ArsA that we define as the 'ATP-closed' state. Notably, the P-loops of the two domains are in close proximity (Fig. 3.3E). The terminal amino group of the conserved IWA lysine of each domain — Lys17 (N) and Lys335 (C) — is positioned ~3 Å from the bridging oxygen atom between the β- and γ-phosphate groups of the ATP bound to the opposing domain. This enables stabilization of negative charge build-up on the nucleotides in the transition state. Together with the Switch I aspartates, which are positioned to activate W_{cat} for nucleophilic attack on ATP, this 'closed' pseudodimer conformation represents a catalytically competent state of ArsA at both nucleotide-binding sites. The ATP-closed state is analogous to the 'closed' dimer states previously reported for Get3 (PDB: 7SPY) and NifH (PDB: 1M34) (Fig. S3.9) [5, 39]. Notably, in our structure, similar to the Get3 'closed'

structure, both IWA lysines appear to form a hydrogen bond with a water molecule found at the pseudodimer interface; these waters may assist with anchoring the terminal amino groups close to the nucleotides in preparation for ATP hydrolysis (Fig. 3.3E).

Relative to the ADP-open state, the two domains rotate towards the pseudodimer interface, forming a tighter interface in the ATP-closed state (Fig. 3.3F). This generates new inter-domain contacts not observed in the ADP-open state (Fig. S3.8B). Glu215 and Glu507, from the N- and C-domains, respectively, form hydrogen bonds with the catalytic water molecule and Switch I residues of the opposing nucleotide-binding site. Electrostatic contacts are formed between Arg214 of the N-domain and Asp371 of the C-domain. This results in a buried surface area of 2,380 $Å^2$ between the domains, ~500 $Å^2$ more than that of the 'open' conformation.

We next sought to establish the conformation of ArsA under turnover conditions. Wild-type ArsA (wtArsA) was incubated with 2 mM $MgCl_2$, 2 mM ATP, and 2 mM sodium arsenite at room temperature for ~1.5 min before cryo-EM grid preparation. This resulted in a Coulomb potential map at 3.0 Å overall gold-standard FSC resolution (Fig. 3.3G, left and Figs. S3.3D and S3.10). The final refined structure revealed a 'closed' conformation similar to the ArsA non-hydrolyzing variant (r.m.s.d = 0.53 Å over 565 residues) with MgATP present at both nucleotide-binding sites and arsenic modeled at the metalloid-binding site (Fig. 3.3G, right). The IWA lysines in this structure are oriented towards the nucleotides (Fig. S3.11A), supporting a catalytically competent pre-hydrolysis state consistent with the variant structure. While the Switch I aspartate in each nucleotide-binding site is positioned next to the Mg^{2+} ion anchoring the W_{cat} for nucleophilic attack, subtle differences are observed in the conformations of the Switch I loop between the wtArsA and variant 'closed' structures. In the N-domain, the Switch I loop moves towards the nucleotide in wtArsA such that Asn50 now interacts with the Mg^{2+} ion via an intervening water molecule (Fig. 3.3H, left). Likewise, in the C-domain, Switch I loop shifts towards the nucleotide, causing His368 to interact with Mg^{2+} via an intervening water molecule (Fig. 3.3H, right). Additionally, we see evidence for a hydrogen-bonding network composed of ordered water molecules including W_{cat}, that connects the Switch I aspartates at the N- (Asp46) and C- (Asp364) sites, potentially coupling the two active sites (Fig. S3.11B). Together, these subtle structural changes observed in the pre-hydrolysis state prime the enzyme to initiate hydrolysis in both N- and C-domains.

3.2.5 The As^{III}-binding site of ArsA

We observed density for arsenic at the metalloid-binding site in the ATP-closed state, coordinated by the thiolate groups of Cys114, Cys173, and Cys422 in a pyramidal geometry (Fig. 3.3D and Fig. S3.11C). Cys114 and Cys422 are homologous residues within the pseudodimer, located at the N-terminal end of helix 6, denoted as $H6_N$ and $H6_C$, respectively, while Cys173 is located on the intervening loop (residues 155-183) between helix 7 and helix 8 in the N-domain.

To confirm the As^{III} coordination environment, we performed solution X-ray absorption spectroscopy (XAS). The ArsA non-hydrolyzing variant (1.95 mM) supplemented with 15 mM $MgCl_2$, 15 mM ATP, and 1.5 mM sodium arsenite was subjected to XAS analysis. X-ray absorption near-edge spectra (XANES) confirmed the presence of trivalent arsenic (Fig. 3.4A, black spectra), as the K-edge at 11865.9 eV was consistent with the first inflection edge energies of As^{III} model compounds (11867 eV). Best-fit simulations of the extended X-ray absorption fine structure (EXAFS) spectra supported the presence of As^{III} coordinated by three sulfur ligands (AsS_3) at an average As–S bond length of 2.27 Å (Fig. 3.4B, top panel and Table S3.2), consistent with cysteine coordination in the cryo-EM structure. Long-range carbon scattering was also observed in the sample, likely from secondary and tertiary sphere carbon atoms associated with the β- and α-carbon atoms of the cysteines, respectively. The oxidation state and coordination environment of the ArsA sample are consistent with that of a control sample consisting of As^{III} and L-cysteine (Fig. 3.4A, purple spectra; 3.4B, middle panel; and Table S3.2). Similar As^{III} coordination has been reported for the ArsR repressor and ArsD metallochaperone, based on XAS analysis [40–42]. We quantified the amount of As^{III} bound to ArsA under conditions similar to those used in cryo-EM sample preparation using inductively coupled plasma mass spectrometry (ICP-MS). In the presence of MgATP, ArsA binds As^{III} with 1:1 stoichiometry (Fig. 3.4C, black plot). Taken together, cryo-EM, XAS, and ICP-MS analysis support the binding of a single three-coordinate As^{III} at the ArsA pseudodimer interface in the presence of MgATP.

The As^{III}-coordinating cysteines are conserved among ArsA homologs (Fig. S3.13). Alanine mutations of these residues ablate As^{III} activation of steady-state ATP hydrolysis (Fig. 3.4D, yellow plot). By contrast, mutation of Cys173 located on the loop region of N-domain alone appears to only decrease the binding of As^{III} to ArsA, as the C173A variant requires higher As^{III} concentrations than the wild-type

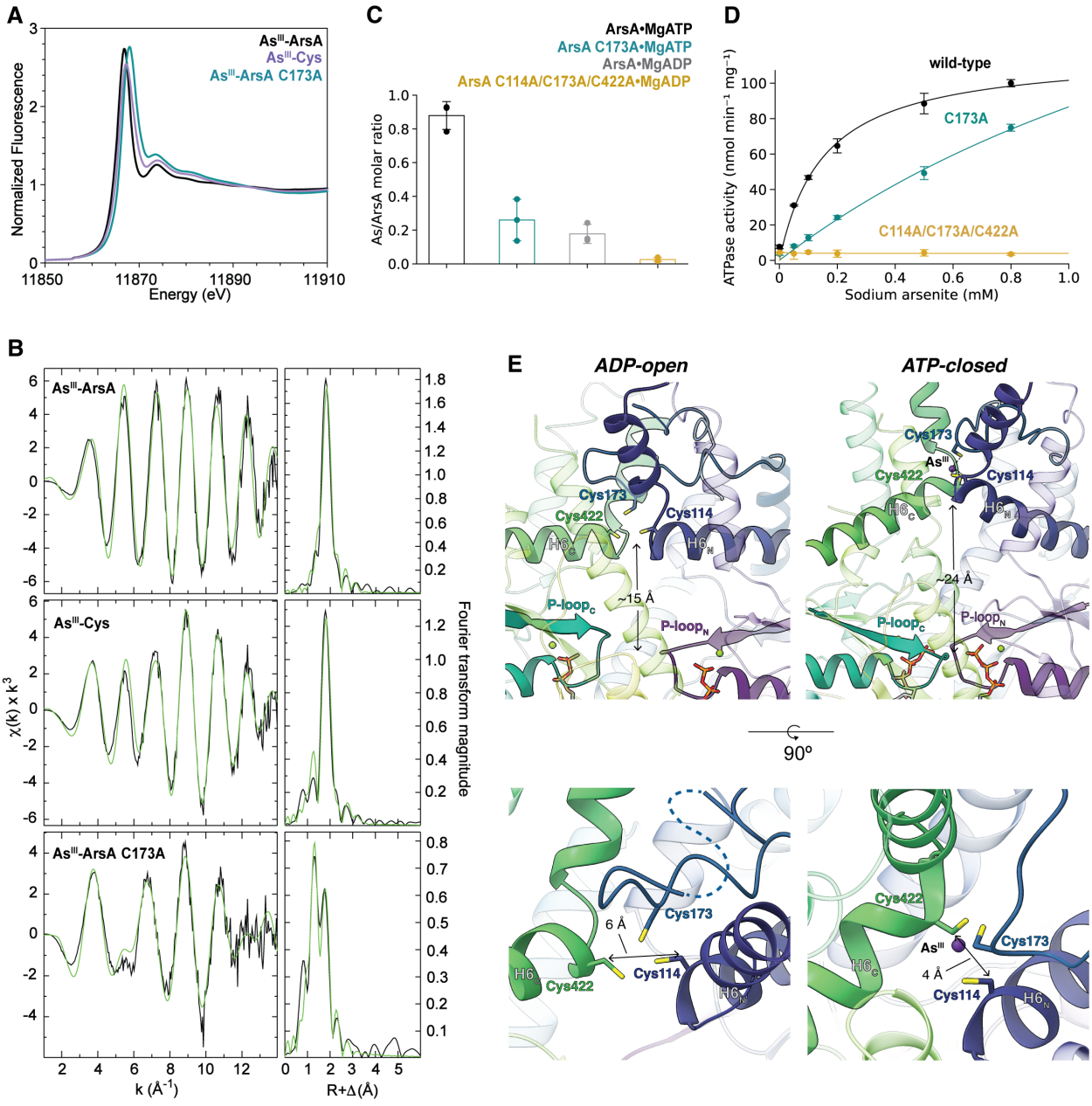

A
As^{III}-ArsA
As^{III}-Cys
As^{III}-ArsA C173A
Normalized Fluorescence
Energy (eV)
C
ArsA•MgATP
ArsA C173A•MgATP
ArsA•MgADP
ArsA C114A/C173A/C422A•MgADP
As/ArsA molar ratio
D
wild-type
C173A
C114A/C173A/C422A
ATPase activity (nmol min^{-1} mg^{-1})
Sodium arsenite (mM)
B
As^{III}-ArsA
As^{III}-Cys
As^{III}-ArsA C173A
χ(k) x k^3
k (Å^{-1})
R+Δ(Å)
Fourier transform magnitude
E
ADP-open
ATP-closed
Cys173
Cys422
Cys114
H6$_C$
H6$_N$
~15 Å
~24 Å
P-loop$_C$
P-loop$_N$
90°
6 Å
4 Å
As^{III}

Figure 3.4: **Characterization of the As^{III} binding site of ArsA.** (A) X-ray absorption near edge spectra (XANES) for *Lf*ArsA•MgATP•As^{III} ('As^{III}-ArsA'; black), As^{III}-L-cysteine control ('As^{III}-Cys'; purple), and *Lf*ArsA(C173A)•MgATP•As^{III} ('As^{III}-ArsA C173A'; teal). (B) Left column, extended X-ray absorption fine structure (EXAFS) spectra (black) with best-fit simulations (green) for each complex. Right column, corresponding Fourier-transforms. A replicate of each XAS sample was individually prepared and analyzed (Fig. S3.12). (C) ICP-MS analysis for As^{III} quantification in nucleotide-bound *Lf*ArsA samples (ArsA•MgATP, black; ArsA C173A•MgATP, teal; ArsA•MgADP, gray; ArsA C114A/C173A/C422A•MgADP, yellow). Bars represent mean As/ArsA molar ratio for n = 3 individually prepared samples with error bars representing standard deviation. (D) Steady-state ATPase activity of *Lf*ArsA (nmol min^{-1} mg^{-1}) plotted against varying As^{III} concentrations (mM) in the presence of 5 mM $MgCl_2$ and 5 mM ATP at 37ºC (wild-type, black; C173A, teal; C114A/C173A/C422A, yellow). Data points represent mean of n = 3 and error bars represent standard deviation. The data was fit to the Michaelis-Menten equation. (E) Comparison of metalloid-binding sites of ADP-open (left) and ATP-closed (right) highlighting the relative positions of Cys114, Cys173 and Cys422 that enable As^{III} (purple) binding.

enzyme to approach half-maximal activation of ATPase activity (Fig. 3.4D, teal plot). XANES and EXAFS spectra for the C173A variant incubated with MgATP and As^{III} suggest AsO_2S coordination, indicating that As^{III} is no longer coordinated to three cysteines (Fig. 3.4A, teal spectra; 3.4B, bottom panel; and Table S3.2). Moreover, ICP-MS analysis of this variant shows over 50% decrease in As^{III}–ArsA molar ratio (Fig. 3.4C, teal plot). As previously reported [31], Cys173 regulates the binding affinity of As^{III} and is not critical for the stimulation of nucleotide hydrolysis.

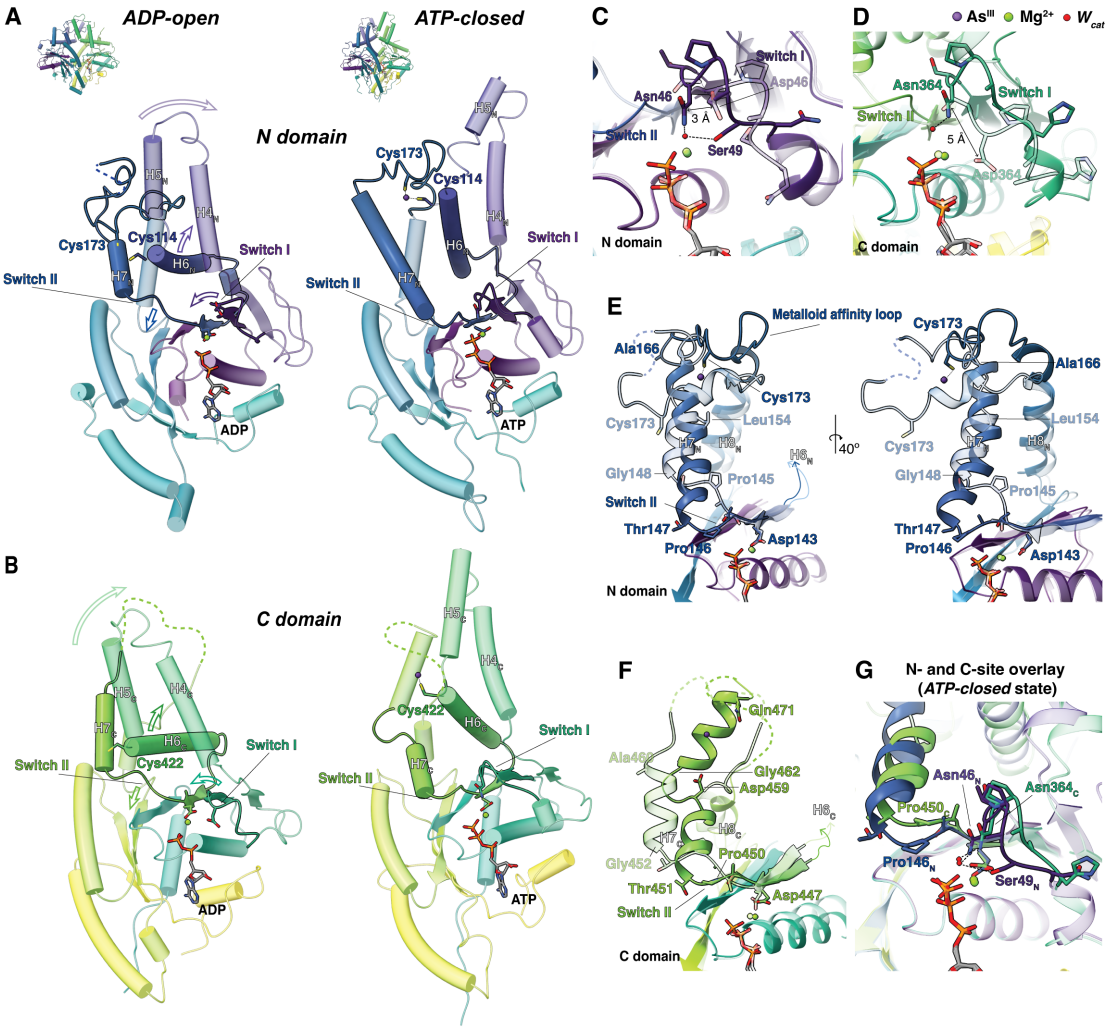
Based on our ICP-MS analysis, ArsA•MgADP binds As^{III} far less tightly than ArsA•MgATP (Fig. 3.4C, black and gray plots), corroborating the absence of As^{III} density in the ADP-open structure. Mutating the As^{III}-coordinating cysteines completely disrupts As^{III} binding to ArsA•MgADP (Fig. 3.4C, yellow plot). How do the conformational changes between 'open' and 'closed' states modulate the affinity for As^{III}? Our cryo-EM structures reveal that switching from 'open' to 'closed' state causes reorientation of helix 6 of both domains such that Cys114 on $H6_N$ and Cys422 on $H6_C$ shift ~9 Å away from the P-loops (Fig. 3.4E, top panels). This makes the metalloid-binding site more accessible to the solvent for As^{III} binding. Additionally, a change in the relative positions of the sidechains of Cys114 and Cys442 accompanies this conformational rearrangement. As shown in Fig. 3.4E (bottom panels), the distance between Cβ atoms of Cys114 and Cys422

reduces from 6 Å in the 'open' state to 4 Å in the 'closed' state, facilitating As^{III} binding. Therefore, by controlling the relative positions of the three cysteines at the metalloid-binding site, nucleotide-dependent conformational changes in ArsA modulate the binding affinity for As^{III}.

3.2.6 Intra-domain conformational rearrangements between 'open' and 'closed' states

In addition to the nucleotide-dependent conformational changes at the pseudodimer interface, large-scale structural rearrangements are observed within each domain between the ADP-open and the ATP-closed states (Fig. 3.5A and B). These changes are primarily linked to the Switch I and II motifs. Compared to the ADP-open state, the Switch I loop of both the N- and C-domains rearranges to shift towards the nucleotide in the ATP-closed state (Fig. 3.5C and D). This enables the positioning of Asp46 and Asp364 to activate W_{cat} and initiate ATP hydrolysis. Differences in the Switch I conformation between the two domains allow Ser49 to additionally stabilize the catalytic water in the N-domain (Fig. 3.5C and G). While the carboxylate group of Asp364 shifts about 5 Å towards the active site from 'open' to 'closed' state, the carboxylate of Asp46 shifts only 3 Å (Fig. 3.5C and D).

Switch II, acting as a link between the nucleotide-binding site and the metalloid-binding site, modulates a series of dramatic conformational changes within each domain between the 'open' and 'closed' states. Upon As^{III} binding to the metalloid-binding site of the ATP-closed state, both $H6_N$ and $H6_C$ undergo rigid body rearrangement such that the N-terminal end of each helix, bearing the As^{III} ligands — Cys114 and Cys422, respectively — rotates away from the nucleotide-binding sites (Fig. 3.5A and B). Reorientation of helix 6 results in the restructuring of the downstream Switch II motif in the ATP-closed state. Although the Switch II aspartates (Asp143 and Asp447) of the respective domains remain anchored near the Mg^{2+} ion in both states, downstream residues 144–147 of the N-domain (Switch II_N), residues 448–451 of the C-domain (Switch II_C), and both helix 7s shift towards the nucleotide in the ATP-closed state (Fig. 3.5E and F). Switch II_N shifts towards the nucleotide to a greater extent than Switch II_C (Fig. 3.5G). Hydrogen-bonding interactions between Thr147 (Switch II_N) and Glu117 ($H6_N$) and between Thr451 (Switch II_C) and Glu425 ($H6_C$), in the ADP-open state, are destabilized in the ATP-closed state to facilitate the conformational change coordinated by helix 6 and Switch II (Fig. S3.14). In the wtArsA ATP-closed state, Switch II_C further shifts towards the nucleotide, adopting the same conformation as Switch II_N (Fig.



S3.11D). This symmetric conformation of Switch II results in an additional contact between the two domains where His149 ($H7_N$) and His453 ($H7_C$) form a π–π stacking interaction (Fig. S3.11E).

A striking difference is seen between the N- and C-domains in the conformation of helix 7 following the Switch II loop. In the ADP-open state, $H7_N$ is composed of residues Gly148–Leu154. In the ATP-closed state, this helix is extended by incorporation of formerly loop residues Gln155–Ala166 into the C-terminal end of $H7_N$ transition, resulting in a twisted helix composed of residues Thr147–Ala166 (Fig. 3.5E). By contrast, $H7_C$ transitions from residues Gly452–Ala460 in the ADP-open state to a discontinuous helical segment between Thr451 and Gln471, connected by a loop formed by residues Asp459–Gly462 in the ATP-closed state

Figure 3.5: **Intra-domain conformational changes between ADP-open and ATP-closed states.** (A, B) Cartoon representations of overall structural rearrangements within N- and C-domains, respectively. Arrows indicate direction of motion of the corresponding motif from the ADP-open (left) to the ATP-closed state (right). All models in these and subsequent panels are colored in viridis. Highlighted residues are shown as sticks. All overlayed structures are aligned on the respective P-loops (residues 16-23 for N-domain and 334-341 for C-domain). (C, D) Conformational changes in Switch I motif between 'open' (50% opacity) and 'closed' (100% opacity) states at both N- and C-nucleotide binding sites, respectively. Shift in the carboxylate group of catalytic Asp from 'open' to 'closed' states is indicated in Å(E) Conformational changes in Switch II and $H7_N$ of the N-domain between 'open' (50% opacity) and 'closed' (100% opacity) states, shown in two views rotated 40º apart. The 'metalloid affinity loop' connecting $H7_N$ and H8N is highlighted. (F) Conformational changes in Switch II and $H7_C$ of the C-domain between 'open' (50% opacity) and 'closed' (100% opacity) states. In both panels E and F, helix 6 has been omitted for visual clarity. (G) Comparison of Switch I and II conformations at the N- and C-nucleotide binding sites of the ATP-closed state.

(Fig. 3.5F). Complete helical rearrangement of $H7_N$ arises by virtue of Cys173 on the loop region between $H7_N$ and H8N, that serves as the third ligand for As^{III} binding in the ATP-closed state. This loop, designated as the 'metalloid affinity loop', undergoes a considerable rearrangement between the 'open' and 'closed' states to stabilize three-coordinate As^{III} binding (Fig. 3.5E). An analogous residue is not found on the C-domain; in fact, the corresponding loop region in the C-domain is poorly resolved in both states. Notably, such helical transition of helix 7 is also observed upon substrate (TA protein) binding at the homologous site in Get3 (PDB: 7SQ0) [5]. Taken together, the conformational changes associated with As^{III} binding in the ATP-closed state demonstrate how three-coordinate As^{III} binding at the metalloid-binding site of ArsA is allosterically coupled to the nucleotide-binding sites via Switch II.

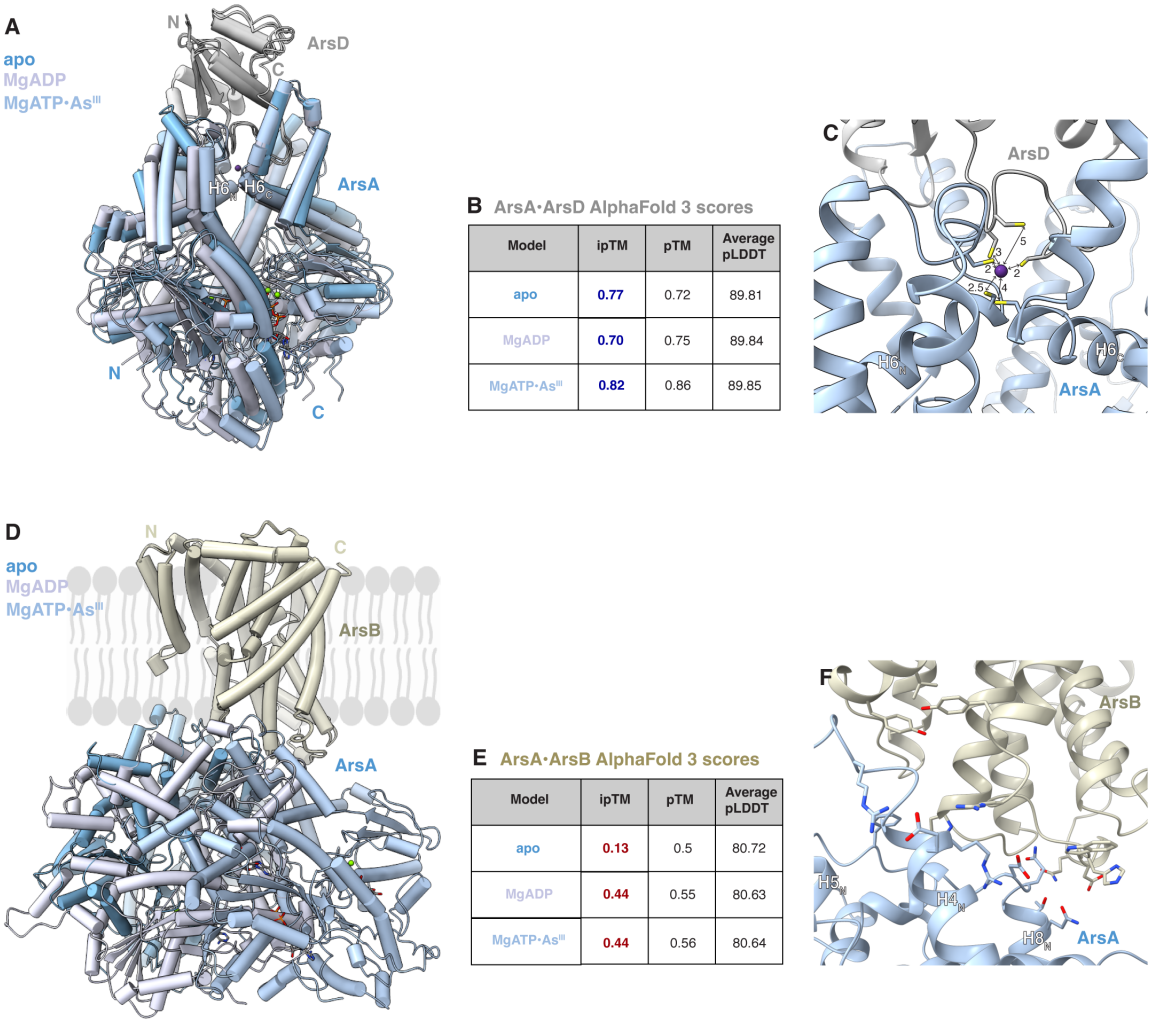
3.2.7 Interaction of ArsA with partner proteins, ArsD and ArsB

We next inquired how the nucleotide- and As^{III}-driven conformational changes of ArsA described here modulate its interaction with its partner proteins, ArsD and ArsB. Genetic and biochemical studies suggest that ArsA and ArsD interact independent of metalloid and nucleotide, and that ATP hydrolysis facilitates metalloid transfer from ArsD to ArsA [42]. In fact, ArsD stimulates ArsA ATPase activity by increasing its affinity for As^{III} from millimolar to micromolar range [14]. On the other hand, there is no biochemical evidence to suggest a direct interaction between

ArsA and ArsB, primarily due to challenges in overexpressing ArsB, except for one study that reported saturable binding of ArsA to membrane vesicles isolated from *E. coli* cells that expressed ArsB [17]. In either case, a complex structure has been elusive. To understand whether the 'open' or 'closed' state of ArsA preferentially interacts with its partner proteins, we performed AlphaFold 3 modeling of the *Lf* ArsA•ArsD and *Lf* ArsA•ArsB complexes in apo, MgADP-bound, and MgATP•As^{III}-bound states [43]. The latter two predictions were intended to mimic the ADP-open and ATP-closed states, respectively. For predictions with ArsD, all three cases revealed ArsA•ArsD complexes with good 'predicted template modeling' (pTM) scores 0.6 and reasonable 'interface pTM' (ipTM) scores > 0.6 (Fig. 3.6A-B). In each case, ArsD is docked proximal to the metalloid binding site of ArsA (Fig. 3.6A). The As^{III}-binding residues of ArsD — Cys12, Cys13 and Cys18 — interface with the As^{III}-binding residues of ArsA — Cys114, Cys173 and Cys422 (Fig. 3.6C). Notably, ArsA approaches a 'closed' conformation in each case, supporting the notion that ArsD transfers As^{III} to ArsA by preferentially interacting with the 'closed' state of ArsA. This is consistent with our experimental results, indicating that the 'closed' state has a higher affinity for the metalloid than the 'open' state. The predictions for the ArsA•ArsB complex were relatively less informative due to poor ipTM scores (< 0.6) for all three predictions (Fig. 3.6D-E). Although ArsA consistently docks near the predicted cytoplasmic face of ArsB, and some interactions between the ArsA N-domain residues and ArsB residues can be observed in the MgATP•As^{III} prediction (Fig. 3.6F), a nucleotide or metalloid-dependent complex is not implied.

3.3 Discussion

Among the IWA ATPases implicated in various complex biological functions, ArsA was one of the first proteins in the family to be studied biochemically and structurally. ArsA plays a central role in ensuring the survival of many bacterial species under toxic concentrations of As^{III} by coupling ATP binding and hydrolysis, not only to capture As^{III} from the cytoplasm via ArsD but also to facilitate its transport across the membrane via ArsB [16, 42]. The underlying nucleotide hydrolysis mechanism of ArsA, which regulates these events, is not fully understood. Based on pre-steady state kinetic analysis, the pseudodimeric ATPase undergoes nucleotide-dependent conformational changes throughout its catalytic cycle [29, 30]. MgATP-bound ArsA has been reported to alternate between two conformations that differ in their affinities for the metalloid substrate [30]. Metalloid binding stabilizes one of these



Model	ipTM	pTM	Average pLDDT
apo	0.77	0.72	89.81
MgADP	0.70	0.75	89.84
MgATP·As^{III}	0.82	0.86	89.85

Model	ipTM	pTM	Average pLDDT
apo	0.13	0.5	80.72
MgADP	0.44	0.55	80.63
MgATP·As^{III}	0.44	0.56	80.64

Figure 3.6: **AlphaFold 3 modeling of interactions of *Lf* ArsA with its partner proteins, ArsD and ArsB.** (A) Cartoon representations of AlphaFold 3 models of *Lf* ArsA•ArsD complex in apo, MgADP and MgATP•As^{III} nucleotide states of ArsA, aligned on ArsD. A truncated form of *Lf* ArsD sequence was used as input, corresponding to the truncated ArsD1-109 from *E. coli* having a C-terminal truncation that was used for biochemical and structural studies in the literature [44]. (B) ArsA•ArsD model confidence metrics indicating scores for interaction (ipTM), overall model (pTM) and per residue confidence (pLDDT). ipTM scores indicating good confidence predictions are highlighted in blue. (C) Interaction between Cys-rich loop of ArsD with the Cys-rich metalloid binding site of ArsA in the ArsA•ArsD model with MgATP•As^{III}. (D) Cartoon representations of AlphaFold 3 models of *Lf* ArsA•ArsB complex in apo, MgADP and MgATP•As^{III} nucleotide states of ArsA, aligned on ArsB. Topology of ArsB in the membrane was predicted using the PPM web server of the Orientation of Proteins in Membranes (OPM) database [45]. (E) ArsA•ArsB model confidence metrics as shown in B. (F) Potential residue interactions at the ArsA-ArsB interface in the model with MgATP•As^{III}. ipTM scores indicating poor confidence predictions are highlighted in red.

conformations, followed by hydrolysis and phosphate release. Previous X-ray crystal structures were broadly inconsistent with this mechanism. Our high-resolution cryo-EM efforts provide the structural basis for the conformational states proposed in this mechanism. We have characterized 'open' and 'closed' conformations of ArsA. The 'open' state can exist bound to either ADP or ATP (Fig. 3.2B, F-G), has low metalloid affinity, and is ATPase inactive. The 'closed' state is stabilized as a ternary complex when As^{III} binds ArsA•MgATP (Fig. 3.3B) and is competent for ATP hydrolysis. The structures enable us to propose a catalytic mechanism for ArsA (Fig. 3.7). ATP binds both N- and C-sites of ArsA where the enzyme exists in equilibrium between 'open' and 'closed' states (states 1 and 2). ATP alone is unable to drive the equilibrium to favor the 'closed' state; as a result, ArsA hydrolyzes nucleotide at a basal rate. When As^{III} is coordinated by three cysteines at the metalloid-binding site (Fig. 3.3D), the 'open' conformation transitions into the catalytically competent 'closed' conformation (state 3). Coordination of As^{III} to Cys114 and Cys422 allosterically triggers the transition of the pseudodimer from 'open' to 'closed' state via helix 6 and Switch II (Fig. 3.5A and B). Cys173 on the 'metalloid affinity loop' acts as a switch controlling the binding affinity of As^{III} for ArsA and stabilizes the ternary complex in the 'closed' state. As^{III} binding by this motif is coupled to the nucleotide-binding site of the N-domain via $H7_N$ and Switch II_N (Fig. 3.5E). This mechanism reveals how three-coordinate As^{III} binding allosterically stimulates ATP hydrolysis, which is the proposed rate-limiting step of ArsA [30]. Following ATP hydrolysis, the enzyme loses affinity for As^{III}, releasing the metalloid and phosphate, and returns to the 'open' state (state 4). Nucleotide exchange can then reset the enzyme for another catalytic cycle. The conformational cycling of ArsA between high-affinity 'closed' and low-affinity 'open' states facilitates coordination of a series of handoffs to capture the metalloid from ArsD, followed by transfer to ArsB at the membrane for export.

The conformational changes of ArsA reported here are consistent with the general structural and mechanistic framework of IWA ATPases (Fig. 3.8). Despite each IWA ATPase binding a unique substrate at the dimer interface, the structural changes corresponding to the substrate, e.g., As^{III} binding to ArsA, remain conserved. In the case of Get3, TA protein binding induces a conformational change in helix 6 (substrate helix) and Switch II, thus regulating the dimer dynamics [5, 25]. Likewise in NifH, association with NifDK affects the positioning of the [4Fe:4S] cluster for electron transport, which then induces conformational changes in the substrate helix and Switch II, stabilizing the 'closed' dimer conformation [6]. Notably, the pseu-

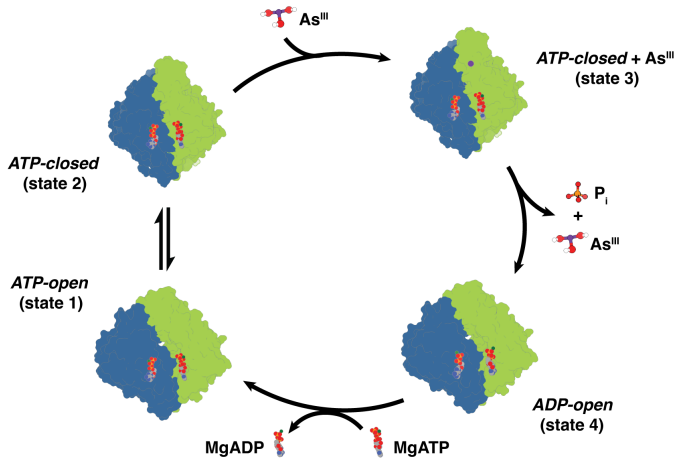


Figure 3.7: **Mechanistic model for ArsA catalytic cycle.** ATP-bound ArsA alternates between 'open' (state 1) and 'closed' (state 2). As^{III} binding stabilizes the catalytically competent 'closed' conformation (state 3). Following ATP hydrolysis and release of As^{III} and phosphate (Pi), ArsA switches to the ADP-bound 'open' conformation (state 4). Exchange of the nucleotide resets the enzyme for another catalytic cycle. States 1, 3 and 4 were characterized in this study. ArsA is colored by N (blue) and C (green) domains. Nucleotides are shown as space-filling models. As^{III} ($As(OH)_3$ at physiological pH) and Pi are shown as ball & sticks.

dodimeric architecture of ArsA is distinct from other ATPases in the family. While both N- and C-domains have been shown to be competent for ATP hydrolysis [21], structural differences between the domains in the ATP-closed structures highlight the functional asymmetry within ArsA. Both Switch I and II loops adopt distinct conformations relative to the nucleotide in the N- and C-domains in the 'closed' state (Figs. 3.5G and Fig. S3.15A), reflecting the difference in ATPase activities, where the N-site is more competent for hydrolysis than the C-site (Fig. S3.15B). These structural differences become less obvious in the wtArsA 'closed' state under turnover conditions where a potential hydrogen-bonding network linking the two active sites may coordinate hydrolysis at the two sites (Fig. S3.11B and D). Further evaluation of the role of pseudodimer would be insightful for understanding functional asymmetry as a generalized mechanistic feature across homodimeric IWA ATPases.

How ArsA enhances the efficiency of As^{III} removal by ArsB is an outstanding question. Properties of the 'ArsAB pump' are distinct from other ATP-dependent transporters, including ABC transporters and P-type ATPases, both in terms of the protein fold of the ATPase domain and that the ATPase activity is only required to

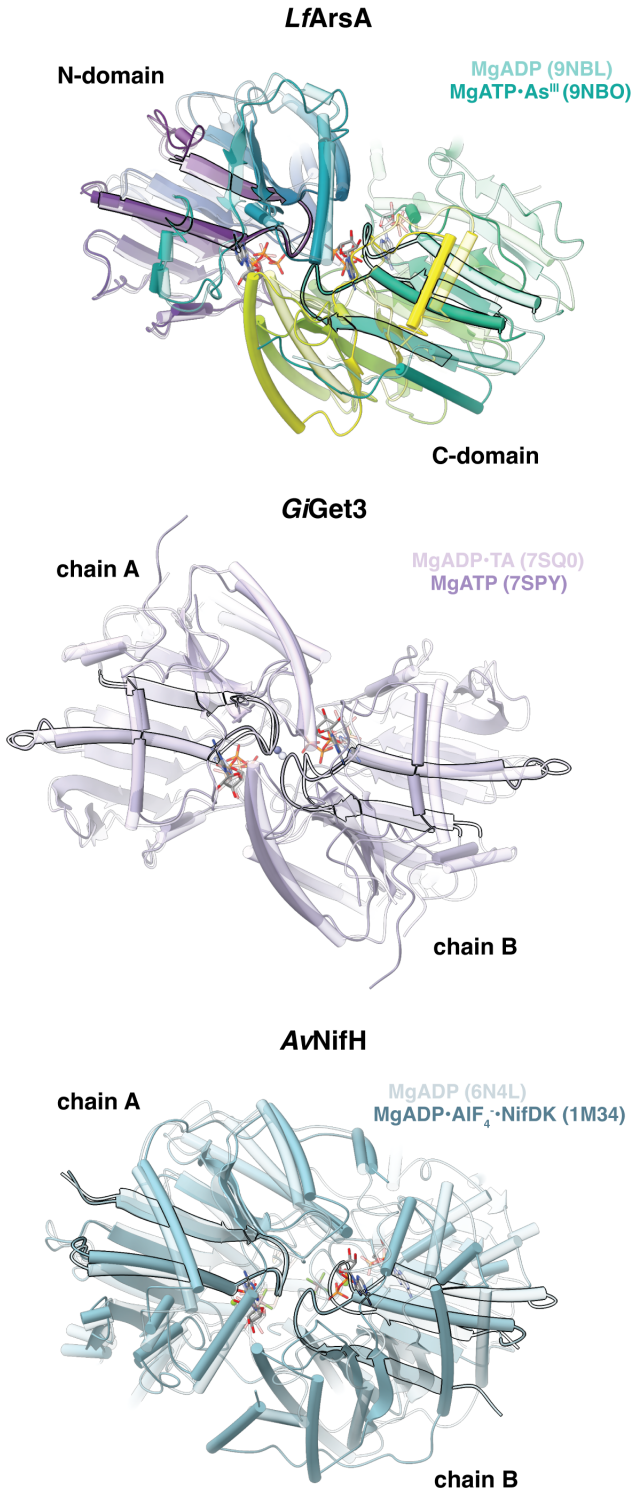


Figure 3.8: **Conformational changes at the dimer or pseudodimer interface of Intradimeric Walker A ATPases.** *Lf* ArsA pseudodimer uses viridis color scheme from N- to C-terminus, whereas GiGet3 and AvNifH dimers are each shown in a single distinct color. *Lf* ArsA is aligned on the P-loop of N-domain, whereas both GiGet3 and AvNifH structures are aligned on the respective P-loops of chain A. The P-loop, the preceding β-sheet and the proceeding helix in each model are highlighted with a black outline.

improve the efficiency of transport [1, 46, 47]. ArsB can transport the metalloid by itself as a secondary carrier [16]. An N-domain P-loop mutant of *Ec*ArsA (G20S) is hydrolysis-deficient and inhibits the ability of ArsB to remove As^{III} [48], implying that the inactive ATPase may trap the transporter in a non-functional state. ArsA may enhance efflux either by triggering a transient conformational change in ArsB to facilitate metalloid transport, or by increasing the local concentration of As^{III} at the membrane, thereby reducing the entropic penalty of delivering the metalloid to ArsB by diffusion. In either case, the nucleotide-dependent conformational landscape of ArsA enables the formation of a transient high-affinity metalloid binding site to direct the toxic metalloid for removal, and therefore, ensures survival of the bacteria under arsenic stress.

3.4 Methods

The *Lf* ArsA gene was designed and purchased from Twist Biosciences (San Francisco, CA). Primers for amplification were designed using the New England Biolabs (NEB) (Ipswich, MA) primer design tool and obtained from Integrated DNA Technologies (Newark, NJ). Cloned constructs were sequenced by whole plasmid sequencing through Plasmidsaurus (Eugene, OR).

3.4.1 Cloning, Expression, and Purification of *Lf* ArsA and its variants

Lf ArsA (Uniprot Accession Number: J9ZFA3) was cloned into the multiple cloning site-1 of pETDuet-1 expression vector encoding a 6x-His tag on the 3'-end, using Hi-Fi DNA assembly protocol (NEB). This resulted in an *Lf* ArsA construct with a C-terminal 6x-His tag. The vector backbone and gene fragment were amplified using the Q5 High-Fidelity 2X Master Mix protocol (NEB). ArsA variants were prepared by point mutagenesis, using the Q5 High-Fidelity 2X Master Mix protocol as well.

Lf ArsA plasmid was transformed into BL21 Gold (DE3) competent cells (Invitrogen). Starter overnight cultures (5-mL) were prepared by inoculated single colony in Luria broth media in the presence of 100 µg/mL ampicillin and grown for 18 h. Sterile 2xYT media (1 L) supplemented with 100 µg/mL ampicillin was inoculated with the overnight culture and grown at 37ºC until OD reached 0.6, when the cells were induced with 0.4 mM isopropyl β-D-1-thiogalactopyranoside (IPTG), and grown at 18 ºC for 16-18 h. Cells were harvested, flash frozen in liquid nitrogen and stored at -80ºC. For purification, thawed cells were resuspended into 'buffer A' (50 mM HEPES pH 7.5, 300 mM NaCl, 10% glycerol, 20 mM imidazole, 5 mM β-

mercaptoethanol (β-ME) and 0.2 mM phenylmethylsulfonyl fluoride (PMSF)), and lysed using an M-110L microfluidizer (Microfluidics). The debris was spun down at 24,000 xg for 30 min and the supernatant was incubated with Ni-NTA affinity resin (Qiagen) at 4ºC for 1 h. Unbound material was removed by batch method using a bench-top centrifuge at 700 xg for 5 min, followed by 2x washes with 50 column-volumes of buffer A. At this point, the resin was transferred to an Econo-Pac chromatography column (Bio-Rad), and four fractions (1 column-volume each) were eluted with 'buffer B' (buffer A + 200 mM imidazole). After SDS-PAGE analysis, fractions containing ArsA were pooled, concentrated to 5 mL in an Amicon 30k concentration filter, and loaded onto a 120-mL HiLoad Superdex-200 column pre-equilibrated with 50 mM HEPES pH 7.5, 100 mM NaCl, 10% glycerol and 5 mM dithiothreitol (DTT). Alternatively, 5 mM tris(2-carboxyethyl)phosphine (TCEP) was used as reducing agent for storage buffer. After SDS-PAGE analysis, pure fractions were stored at -80ºC. Protein concentration was determined by measuring the absorbance at 280 nm with a NanoDrop 2000 spectrophotometer (Thermo Fisher), using molar extinction coefficient of 25,400 $M^{-1}cm^{-1}$ and molecular weight of 63.8 kDa.

3.4.2 X-ray absorption spectroscopy of ArsA

'As^{III}-ArsA' sample was prepared by incubating 1.95 mM of *Lf* ArsA D46N/D364N (non-hydrolyzing variant) with 15 mM $MgCl_2$, 15 mM ATP (Sigma), and 1.5 mM of sodium arsenite (Sigma) and then buffer exchanged into 50 mM HEPES pH 7.5, 100 mM NaCl, 30% glycerol and 5 mM TCEP using Micro Bio-Spin P-6 desalting column (Bio-Rad) to remove excess unbound arsenite. 'As^{III}-ArsA C173A' sample was prepared similarly with 2.5 mM of *Lf* ArsA (non-hydrolyzing variant) C173A incubated with 20 mM $MgCl_2$, 20 mM ATP, and 20 mM sodium arsenite. The 'As^{III}-cysteine' sample (control) was prepared by adding 1.2 mM sodium arsenite to 6 mM L-cysteine in 50 mM HEPES pH 7.5, 100 mM NaCl, 30% glycerol, and 5 mM TCEP. A replicate was prepared and analyzed for each sample. As^{III}-ArsA replicate was supplemented with 15 mM sodium arsenite instead. Samples (150 µL each) were injected into leucite cells using a Hamilton syringe, avoiding air bubbles, and immediately frozen in liquid nitrogen.

XAS data were collected at the Stanford Synchrotron Radiation Laboratory (SSRL) on beamline 7-3, equipped with Si[220] double-crystal monochromator with a harmonic rejection mirror. Samples were maintained at 10 K using an Oxford Instruments continuous-flow liquid helium cryostat. Protein fluorescence excitation

spectra were recorded using a 30-element Ge solid-state array detector. A germanium filter (0.6 mm in width) and solar slits were placed between the cryostat and detector to filter scattering fluorescence not associated with protein-bound arsenic signals. XAS spectra were recorded in 5 eV steps in the pre-edge region (11625–11825 eV), 0.25 eV steps in the edge region (11850–11900 eV), and 0.05 $Å^{-1}$ increments in the extended X-ray absorption fine structure (EXAFS) region out to a k range of 14 $Å^{-1}$. The data were integrated from 2 to 25 s in a k-weighted manner in the EXAFS region for a total scan length of 45 min. X-ray energies were calibrated using an arsenic foil absorption spectrum collected simultaneously with the protein data. The first inflection point for the arsenic foil edge was assigned to 11867 eV. Each fluorescence channel of each scan was examined for spectral anomalies prior to averaging. The data represent an average of five to six scans for each sample. Data were processed using the Macintosh OS X version of the EXAFSPAK software suite integrated with F_{eff} version 7 for theoretical model generation. Data, collected out to k = 14.0 $Å^{-1}$, corresponds to a spectral resolution of 0.121 $Å^{-1}$ for all metal–ligand interactions; therefore, only independent scattering environments at distances >0.121 Åwere considered resolvable in the EXAFS fitting analysis. The final EXAFS fitting analysis was performed on raw/unfiltered data. Protein EXAFS data were fit using single-scattering Feff theoretical models calculated for carbon, oxygen, and sulfur coordination to simulate arsenic-ligand environments, with values for the scale factors (0.98) and E_0 (-10) following a previously published fitting protocol. All spectra were fit using identical protocols, first by distinguishing the best single-shell fit to the data and then by progressively adding extra scattering environments to the fit. Best fit selection criteria were identified by having the lowest mean square deviation between experimental data and the theoretical fit (F' value), along with an acceptable absorber-scatterer bond disorder value (Debye-Waller factor) of < 6.0 x 10^3 $Å^2$. XAS parameters for each sample can be seen in Table S3.2. XAS spectra and parameters for replicate samples can be seen in Fig. S3.12 and Table S3.3).

3.4.3 Cryo-EM sample preparation

For the ADP-open structure, the *Lf* ArsA sample was buffer-exchanged into 50 mM HEPES pH 7.5, 100 mM NaCl, and 4 mM TCEP using Micro Bio-Spin P-6 desalting columns (Bio-Rad). The resulting sample was diluted to 10 mg/mL and incubated with 2 mM each of $MgCl_2$, ADP (Sigma), and 5 mM sodium arsenite for 3 hours. Samples for As^{III}-free structures were prepared similarly. For the

ATP-closed structure, *Lf* ArsA D46N/D364N was buffer-exchanged into 50 mM HEPES pH 7.5, 100 mM NaCl, and 5 mM DTT. The resulting sample was diluted to 10 mg/mL and incubated with 2 mM each of $MgCl_2$, ATP, and sodium arsenite for 30 min. The sample for wtArsA ATP-closed structure was prepared similarly but incubated with ligands for 1.5 min prior to grid preparation. For cryo-EM grid preparation, 3µL of the sample supplemented with 0.05% CHAPSO was applied to glow-discharged Quantifoil holey carbon R1.2/1.3 300 Mesh, Copper (Quantifoil, Micro Tools GmbH) grids using a Vitrobot (FEI Vitrobot Mark v4 x2, Mark v3). Grids were blotted at 100% humidity and 4ºC using a blot time of 4-5 seconds and blot force of 7, and immediately followed by plunge-freezing into liquid ethane.

3.4.4 Cryo-EM data acquisition and processing

The grids were screened for ice thickness and sample quality using a 200 kV Talos Arctica TEM equipped with a Gatan K3 detector. Data collection was performed using a 300 kV Titan Krios TEM equipped with a Gatan K3 direct electron detector and Gatan Energy Filter (slit width 20eV) in super-resolution mode using SerialEM [49]. Each dataset was acquired at a nominal magnification of 130,000x with a raw pixel size of 0.325 Å/pixel, electron exposure of 70 e-/$Å^2$ over 40 frames (exposure rate of 1.75 e-/$Å^2$/frame), and a defocus range of -0.5 to -2.5 µm. Correlated double sampling (CDS) mode was enabled to improve the signal-to-noise ratio of the images [50].

All datasets were processed in cryoSPARC v4.4.1-4.5.1 using the same overall processing workflow [51]. Movie frames were motion corrected using 'patch motion correction' with 0.5 F-cropping, resulting in a pixel size of 0.65 Å/pixel. The contrast transfer function (CTF) for the motion-corrected micrographs was estimated using 'patch CTF estimation'. Micrographs were manually curated, and particles were picked using 'blob-picking', extracted with a 2x bin (1.3 Å/pixel), and subjected to multiple rounds of 2D classification to retain good-quality particle picks. Ab-initio reconstruction was performed on this particle set to result in four 3D volumes, followed by 'heterogenous refinement' that yielded one good volume revealing secondary structural features of ArsA. Particles were re-extracted from the micrographs with no binning (0.65 Å/pixel) and subjected to multiple rounds of ab-initio reconstruction and heterogeneous refinement followed by 'reference-based motion correction', 'global CTF refinement', and 'non-uniform refinement' [52], to obtain a final set of particles resulting in a good-quality map. The overall resolution was estimated from the gold-standard Fourier shell correlation (FSC) curve at a

cut-off of 0.143 in cryoSPARC. B-factor sharpening values were determined using 'sharpening tools' in cryoSPARC. Sharpened maps were exported from cryoSPARC for model building and refinement. The local resolution of each map was calculated using 'local resolution estimation' job in cryoSPARC. For the ArsA•MgATP map, iterative rounds of ab-initio reconstruction were performed instead to obtain the final set of particles. The image processing pipeline for each cryo-EM map reported in this work is presented in the SI — ArsA•MgADP in the presence of As^{III}, Fig. S3.2; ArsA•MgADP in the absence of As^{III}, Fig. S3.5; ArsA•MgATP, Fig. S3.6; ArsA•MgATP•As^{III}, Fig. S3.7; and wtArsA•MgATP•As^{III}, Fig. S3.10. Data collection and processing parameters can be seen in Table S3.1.

3.4.5 Model building and refinement

All models were built and refined similarly unless stated otherwise. Initial models were obtained by docking the N (residues 1-294) and C (residues 316-587) domains of *Lf* ArsA AlphaFold2 model prediction determined using ColabFold v1.5.2 [36], as distinct model entries into the sharpened EM maps using 'Dock in map' in Phenix v1.21.2 [53]. The model was adjusted by manual model building into good-quality regions of the map in Coot v0.9.8.7 [54]. Ligands (Mg^{2+}, nucleotides and/or As^{III}) and any ordered water molecules in the nucleotide-binding sites were docked in Coot. Models were refined using 'Real-space refinement' in Phenix and 'Real-space refine zone' in Coot. Model refinement and validation statistics can be seen in Table S3.1. Custom geometry restraints for As-S bond length were obtained from the XAS data and applied to the As^{III} binding site during refinement cycles in Phenix for the ArsA•MgATP•As^{III} structure. Each As-S bond length was restrained to 2.27 Åwith a sigma-value of 0.06 Åderived from the Debye-Waller factor σ^2 of 3.49 x 103 $Å^2$ corresponding to As-S bond distance of 'As^{III}-ArsA' sample (Table S3.2). Average and per-residue Q-scores for each refined model were calculated using the Qscore plugin in ChimeraX v1.7.1.

3.4.6 ATPase assays

Steady-state ATPase activity of ArsA was measured spectrophotometrically using an NADH-linked coupled assay with an ATP regeneration system that couples ATP hydrolysis to oxidation of NADH [55]. The reaction mixture (100 µL) consisted of 50 mM HEPES pH 7.5, containing 5 mM $MgCl_2$, 2 mM phosphoenol pyruvate (Roche), 20 U/mL pyruvate kinase from rabbit muscle (MP Biomedicals), 20 U/mL L-lactate dehydrogenase from rabbit muscle (Sigma), 0.2 mM NADH (Roche), 2-5

μM ArsA or its variant and varying concentrations of sodium arsenite (0, 0.05, 0.1, 0.2, 0.5 and 0.8 mM). The mixture was incubated for 5-10 min at 37 ºC. The assay was conducted in a 96-well plate setup using SpectraMax M3 plate reader (Molecular Devices). The reaction was initiated by adding 5 mM ATP into the wells with gentle mixing. This was immediately followed by the measurement of a steady-state decrease in the NADH absorbance at 340 nm for 20 min on the plate reader at 37ºC. Slopes were calculated for the steady region of the progress curves in the units of A340 min^{-1}. Rates were calculated as nanomoles of ATP hydrolyzed per min per mg of ArsA (nmol min^{-1} mg^{-1}) using the extinction coefficient of NADH at 340 nm (6,220 M^{-1} cm^{-1}). ATPase activity was fitted to the Michaelis-Menten equation as a function of arsenite concentration and plotted using Python 3 in Jupyter notebook with assistance from ChatGPT, OpenAI in writing the code.

3.4.7 Inductively coupled plasma mass spectroscopy (ICP-MS)

ArsA-bound arsenic was quantified for ArsA•MgATP, ArsA C173A•MgATP, ArsA•MgADP and ArsA C114A/C173A/C422A•MgADP samples. The non-hydrolyzing variant was utilized when trapping complexes with MgATP. Each sample was prepared similar to the cryo-EM samples by incubating 10 mg/mL of *Lf* ArsA or its variant, with 2 mM $MgCl_2$, 2 mM ATP/ADP, and 5 mM sodium arsenite at 4ºC for 3 hours. Free arsenite was removed from the samples by exchanging the buffer with 50 mM HEPES pH 7.5 using Micro Bio-Spin P-6 desalting column. Subsequently, arsenic was extracted from the protein by incubating 30 μL of the sample with 1.1 mL of 70% (v/v) HNO_3 (ACS grade) in borosilicate glass tubes at 50ºC for 30 min. Each sample was then diluted to 15 mL with double distilled water in a 50-mL flat bottom tube, resulting in maximum arsenic concentration equivalent to 23.5 ppb or 0.313 μM of ArsA in 5% (v/v) HNO_3. Arsenic concentrations were determined by ICP-MS using an Agilent 8800. The sample introduction system consisted of a Miramist nebulizer, Scott-type spray chamber, and 2.0 mm fixed injector quartz torch. A guard electrode was used, and the plasma was operated at 1500 W. Arsenic analysis was performed in He mode using MS/MS scan mode. Arsenic standards were prepared in 5% (v/v) HNO_3 from a 10 μg/mL arsenic standard solution (Inorganic Ventures) in the range of 0 to 40 ppb of As. The probe was subjected to 4 rinses (1 flowing, 3 static) between samples to minimize any possibility of cross-contamination. Sample to sample carryover has been found to be less than 1% between the highest standard (40 ppb) and the blank for similar analyses. Results were analyzed using ICP Masshunter 4.5 (Agilent Technologies).

3.4.8 AlphaFold 3 modeling

All ArsA•ArsD and ArsA•ArsB complex models were predicted using AlphaFold 3 [43], as installed on the Resnick High Performance Computing (HPC) cluster at Caltech. AlphaFold 3 model parameters were directly received from Google DeepMind upon request using a Google form available at AlphaFold 3 GitHub repository (https://github.com/google-deepmind/alphafold3). *Lf* ArsD and *Lf* ArsB sequences were retrieved from Uniprot. For each complex prediction, a single 'JavaScript Object Notation' (JSON) input was prepared that contained the protein chain sequences and the Chemical Component Dictionary (CCD) codes for the ligands. A smiles string was used to model As^{III}. The Predicted Local Distance Difference Test (pLDDT) was used to estimate the per-residue confidence for each model. Predicted Template Modeling (pTM) and interaction pTM (ipTM) scores were used as confidence metrics for the full structure and the interaction interface, respectively.

3.5 Supplementary information

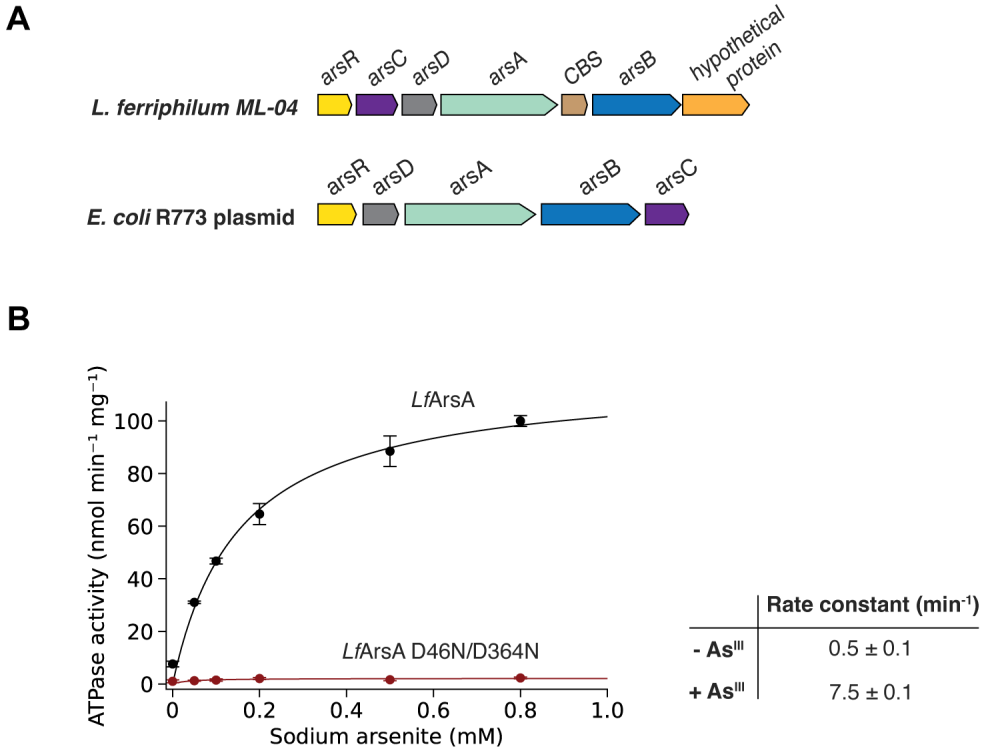


Figure S3.1: **Arsenite efflux pump ATPase ArsA from *Leptospirillum ferriphilum* strain ML-04.** (A) Architecture of the *ars* operon in *L. ferriphilum* ML-04 in comparison with that of the *E. coli* plasmid R773. Both operons constitute genes for an As^{III}-responsive repressor, arsR; As^{III}-metallochaperone, arsD; As^{III}-translocating ATPase, arsA; the membrane transporter, arsB; and arsenate (AsV) reductase, arsC. The *L. ferriphilum ars* operon consists of two additional genes — a cystathione-β-synthase (CBS) domain regulatory protein, and a hypothetical protein annotated as a putative periplasmic component of an ABC-type phosphate transporter [34]. (B) Steady-state ATPase activity of *Lf* ArsA (black plot) and the Switch I aspartate double mutant, D46N/D364N (red plot) as a function of As^{III} concentration with 5 mM each of $MgCl_2$ and ATP at 37ºC. While the mutant is non-hydrolyzing, wild-type ArsA shows As^{III}-dependent activation of ATPase activity. In the absence of As^{III} (0 mM data point), the calculated pseudo first-order rate constant is 0.5 ± 0.1 min^{-1} (- As^{III}) and saturating As^{III} concentrations enhance the rate constant to 7.5 ± 0.2 min^{-1} (+ As^{III}). Data points represent mean of n = 3 and error bars represent standard deviations. The data was fit to the Michaelis-Menten equation. Rate constants were calculated using ArsA molecular weight of 63.8 kDa.

Table S3.1: **Cryo-EM data collection, refinement and validation statistics.**

	***Lf*ArsA·MgADP (+ As^{III})** (EMD-492321, PDB: 9NBL)	***Lf*ArsA·MgADP (– As^{III})** (EMD-49232, PDB: 9NBM)	***Lf*ArsA·MgATP** (EMD-49233)	***Lf*ArsA·MgATP·As^{III} (D46N/D364N)** (EMD-49234, PDB: 9NBO)	***Lf*ArsA·MgATP·As^{III} (WT)** (EMD-49237, PDB: 9NBW)
Data collection and processing					
Magnification	130,000	130,000	130,000	130,000	130,000
Voltage (kV)	300	300	300	300	300
Electron exposure (e^{-}/Å^2)	70	70	70	70	70
Defocus range (μm)	–0.5 to –2.5	–0.5 to –2.5	–0.5 to –2.5	–0.5 to –2.5	–0.5 to –2.5
Pixel size (Å) (super-resolution)	0.325	0.325	0.325	0.325	0.325
Movies	4,732	2,122	3,994	4,031	4,509
Total extracted particles	2,643,084	1,824,091	1,673,570	4,265,167	1,449,927
Final particles	103,258	34,966	48,382	91,645	124,599
Symmetry imposed	C1	C1	C1	C1	C1
Map resolution (Å, FSC 0.143)	3.4	3.8	6.6	3.0	3.0
Refinement					
Initial model used	*Lf*ArsA AlphaFold model	*Lf*ArsA AlphaFold model	*Lf*ArsA AlphaFold model	*Lf*ArsA AlphaFold model	*Lf*ArsA AlphaFold model
Protein residues	558	552	571	570	570
Ligands	ADP:2, Mg:2	ADP:2, Mg:2	ATP:2, Mg:2, ARS:1, HOH:5	ATP:2, Mg:2, ARS:1, HOH:13	ATP:2, Mg:2, ARS:1, HOH:13
Map sharpening B factor (Å^2)	–165.4	–163.0	–124.1	–133.0	–133.0
R.m.s. bond lengths (Å)	0.003	0.003	0.004	0.002	0.002
R.m.s. bond angles (°)	0.616	0.661	1.024	0.561	0.561
Validation					
MolProbity score	1.63	1.41	1.34	1.21	1.21
Clashscore	13.30	7.42	6.25	4.32	4.32
Poor rotamers (%)	0	0	0	0	0
Ramachandran favored (%)	98.18	98.16	98.05	98.23	98.23
Ramachandran allowed (%)	1.82	1.84	1.95	1.77	1.77
Ramachandran outliers (%)	0	0	0	0	0

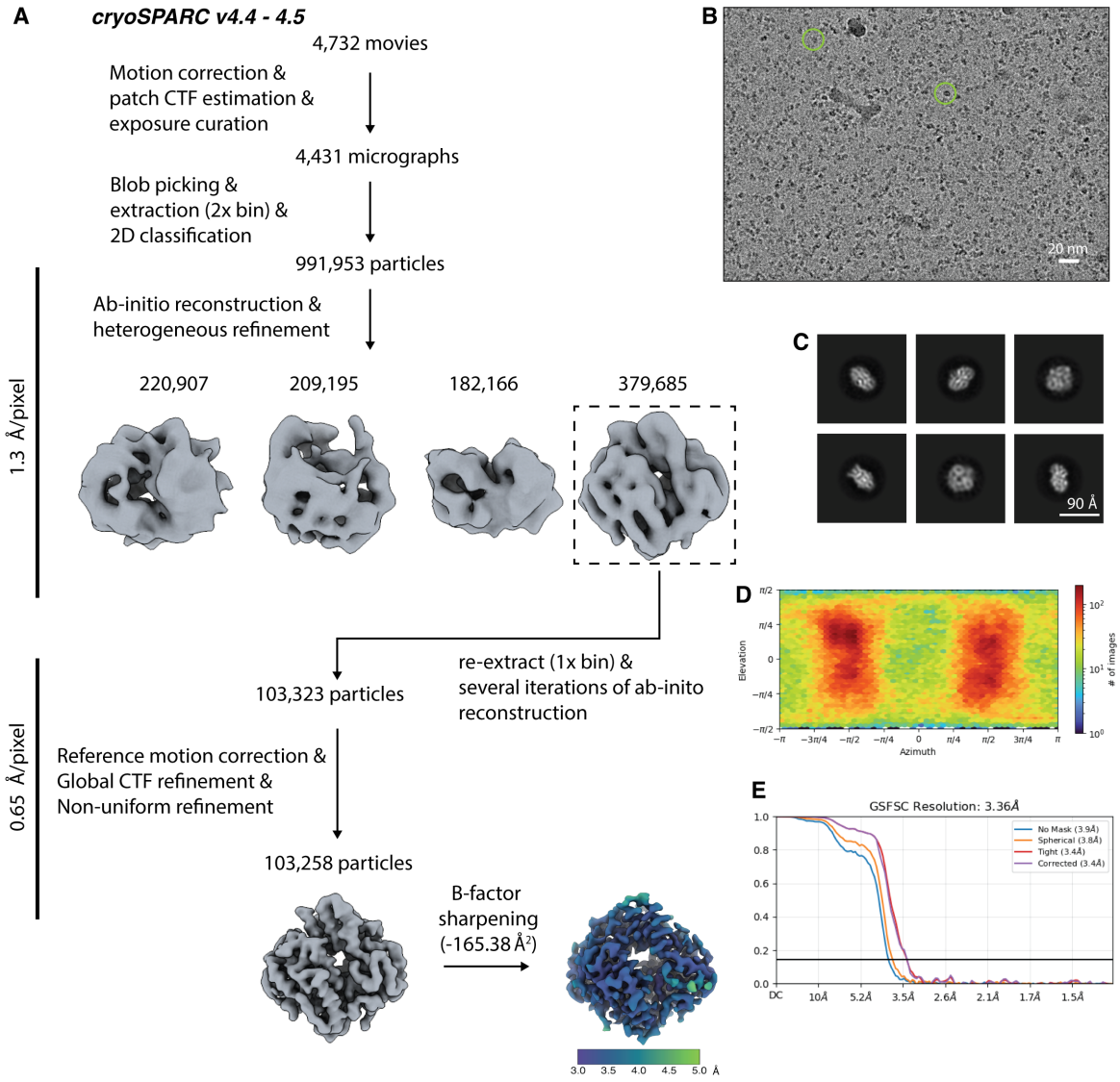


Figure S3.2: **Cryo-EM processing and validation for *Lf*ArsA•MgADP structure determined in presence of As^{III}.** (A) Cryo-EM data processing workflow in cryoSPARC. (B) Representative micrograph. (C) 2D class averages showing different views of ArsA in the 'open' state. (D) Angular distribution heatmap plot. (E) Gold-standard Fourier shell correlation (GSFSC) curve.

Table S3.2: **Arsenic EXAFS best fit simulation parameters.**

Sample	Nearest Neighbor Ligand Environment[a]				Long-Range Ligand Environment[b]			F'[g]
	Atom[c]	R (Å)[d]	C.N.[e]	σ^{2}[f]	Atom[c]	R (Å)[d]	C.N.[e]	
As^{III}-ArsA	S	2.27	3	3.49	C	2.74	1	0.66
					C	3.19	2	
					C	3.36	3	
As^{III}-Cys	S	2.27	2	3.06	C	3.21	1.5	0.30
					C	3.67	1	
As^{III}-ArsA C173A	O/N	1.78	2	3.80	C	2.72	2	0.34
	S	2.28	1	3.71	C	2.86	1.5	

[a,b] Independent metal-ligand scattering environments. [c] Scattering atoms: N (nitrogen), O (oxygen), C (carbon), S (sulfur). [d] Average metal-ligand bond length. [e] Average metal-ligand coordination number. [f] Average Debye-Waller factor ($Å^2 \times 10^3$). [g] Degrees of freedom weighted mean square deviation between data and fit.

Table S3.3: **Arsenic EXAFS best fit simulation parameters for replicate samples.**

Sample	Nearest Neighbor Ligand Environment[a]				Long-Range Ligand Environment[b]			F'[g]
	Atom[c]	R (Å)[d]	C.N.[e]	σ^{2}[f]	Atom[c]	R (Å)[d]	C.N.[e]	
As^{III}-ArsA	S	2.27	2.5	3.46	C	2.72	1	0.60
					C	3.20	2	
					C	3.36	3	
As^{III}-Cys	S	2.27	2	3.12	C	3.31	2	1.28
					C	3.63	1.5	
As^{III}-ArsA C173A	O/N	1.78	2	4.06	C	2.71	2	0.61
	S	2.28	1	3.20	C	2.89	1	
					C	3.33	2	

[a,b] Independent metal-ligand scattering environments. [c] Scattering atoms: N (nitrogen), O (oxygen), C (carbon), S (sulfur). [d] Average metal-ligand bond length. [e] Average metal-ligand coordination number. [f] Average Debye-Waller factor ($Å^2 \times 10^3$). [g] Degrees of freedom weighted mean square deviation between data and fit.

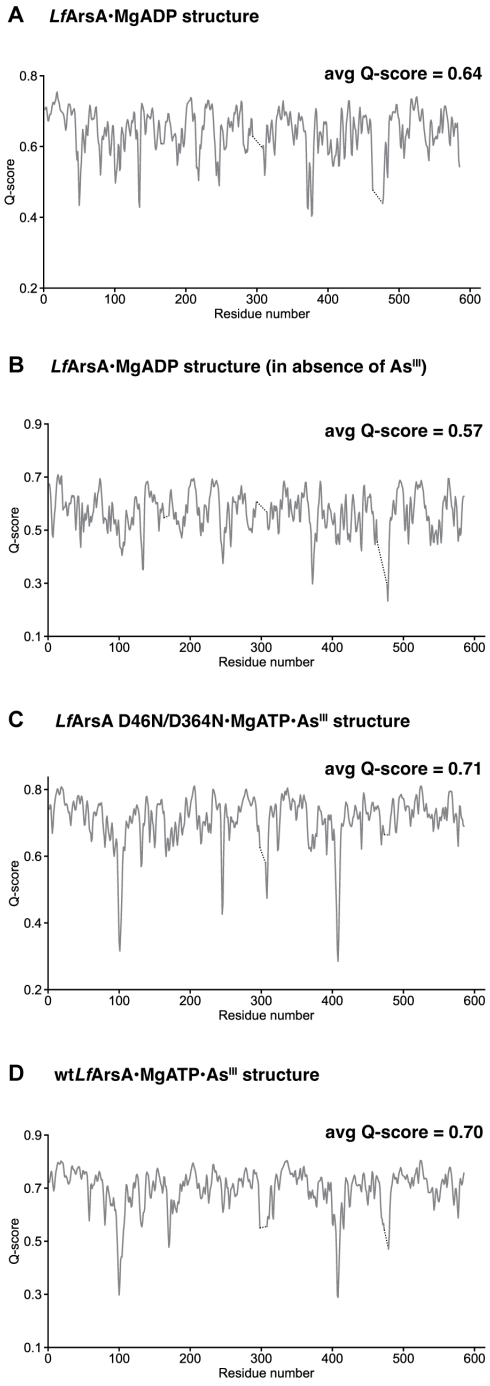


Figure S3.3: **Per-residue average Q-scores for ArsA cryo-EM structures.** (A) *Lf* ArsA•MgADP (in presence of As^{III}), (B) *Lf* ArsA•MgADP (in absence of As^{III}), (C) *Lf* ArsA D46N/D364N•MgATP•As^{III} and (D) wild-type *Lf* ArsA•MgATP•As^{III}. Average Q-scores are reported for residues 2-585 and were calculated using the Qscore plugin in ChimeraX v1.7.1. Unmodeled residues are shown as dotted lines.

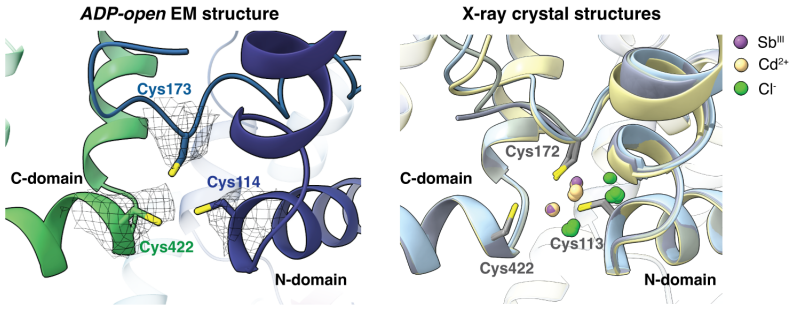


Figure S3.4: **Comparison of the metalloid-binding sites of *Lf*ArsA ADP-open cryo-EM structure (left) and *Ec*ArsA X-ray crystal structures (right).** While the latter includes Sb^{III}, Cd^{2+} and Cl^- ions modeled at the metalloid-binding site, the relative positions of the Cys residues are similar in both EM and X-ray structures. Likely, the ions present in the crystal structures were non-specifically acquired from the crystallization conditions, and therefore, do not represent a functionally relevant metalloid-bound state of ArsA. Crystal structures shown on the right include MgADP-bound (gray, PDB: 1F48), MgADP/MgATP-bound (yellow, PDB: 1II0), MgADP/MgAMP•PNP (blue, PDB: 1II9) and MgADP/MgADP•AlF_3 (purple, PDB: 1IHU) states.

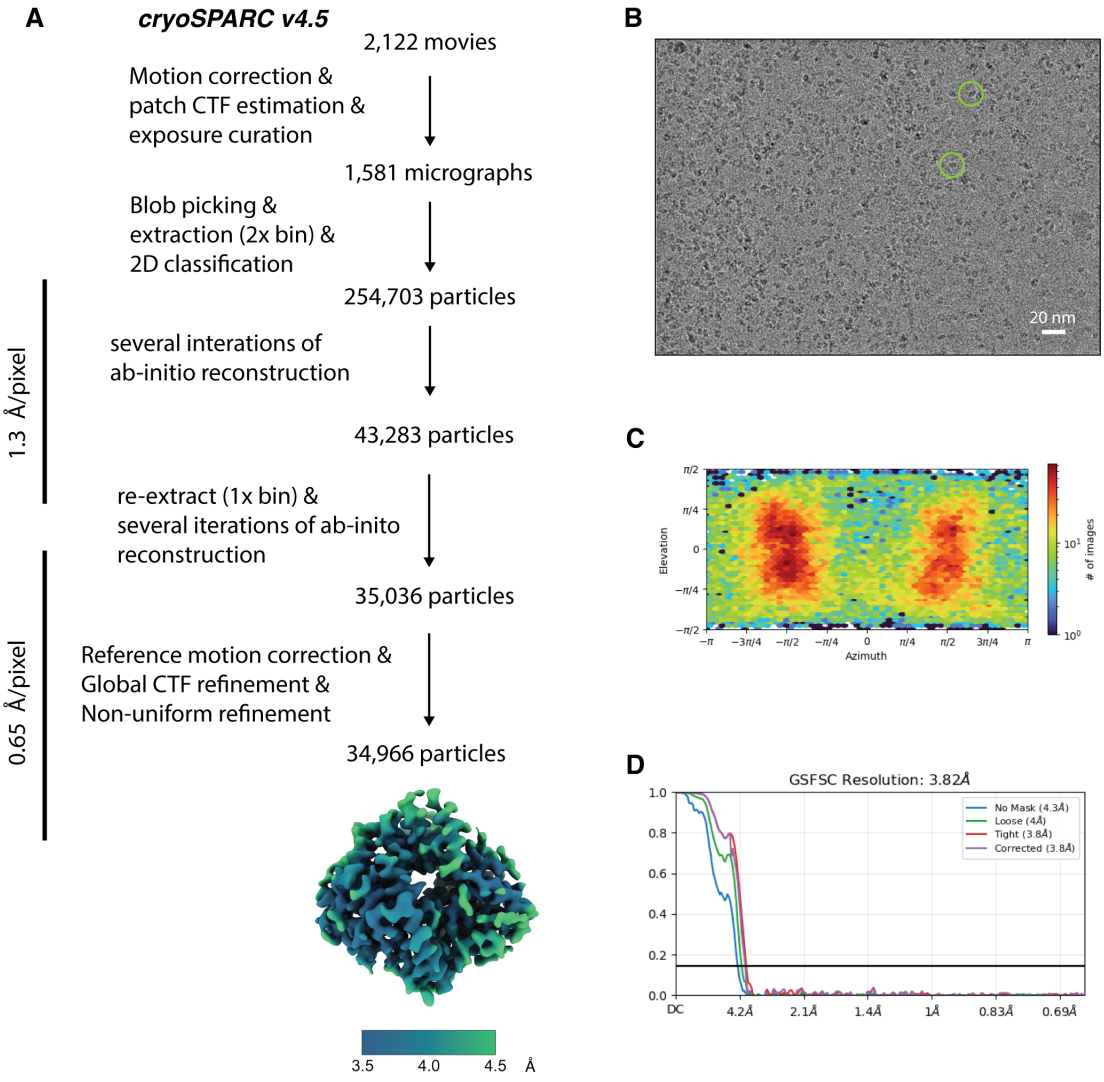


Figure S3.5: **Cryo-EM processing and validation for *Lf*ArsA•MgADP structure determined in absence of As^{III}.** (A) Cryo-EM data processing workflow in cryoSPARC. (B) Representative micrograph. (C) Angular distribution heatmap plot. (D) Gold-standard Fourier shell correlation (GSFSC) curve.

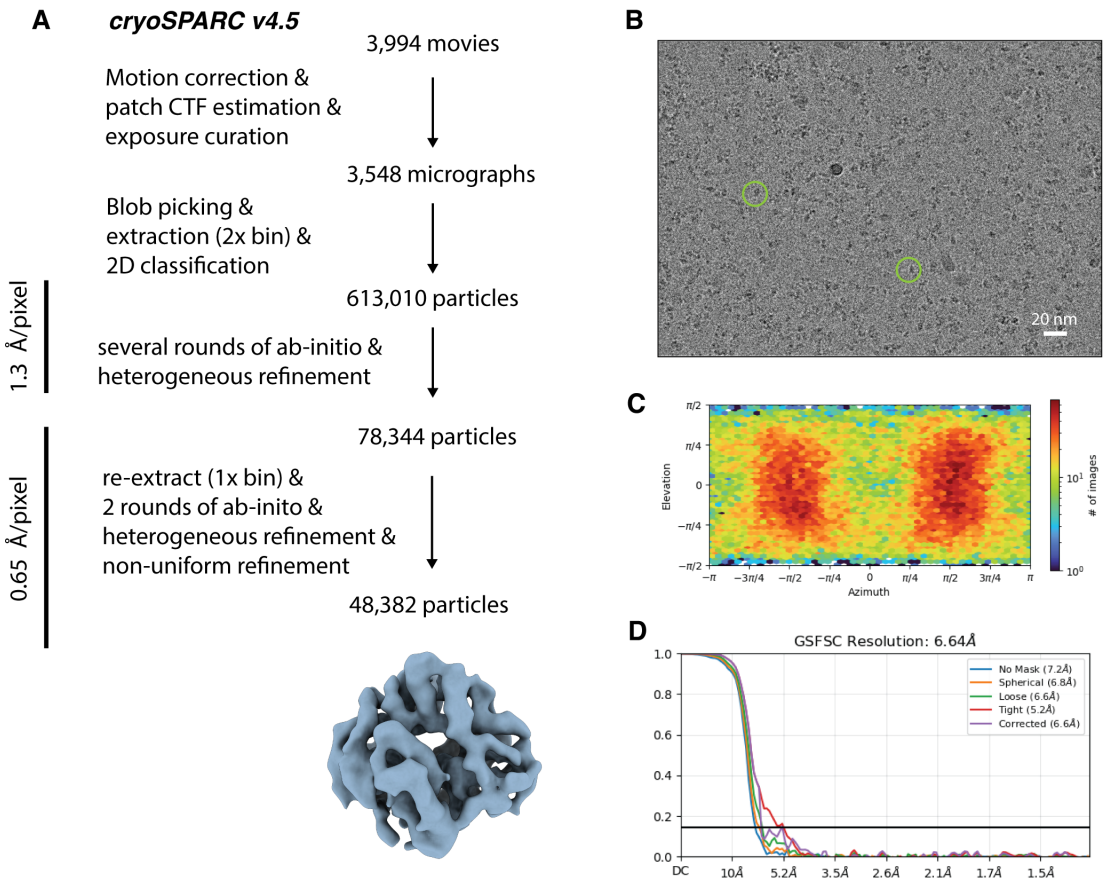


Figure S3.6: **Cryo-EM processing and validation for *Lf*ArsA•MgATP structure.** (A) Cryo-EM data processing workflow in cryoSPARC. (B) Representative micrograph. (C) Angular distribution heatmap plot. (D) Gold-standard Fourier shell correlation (GSFSC) curve.

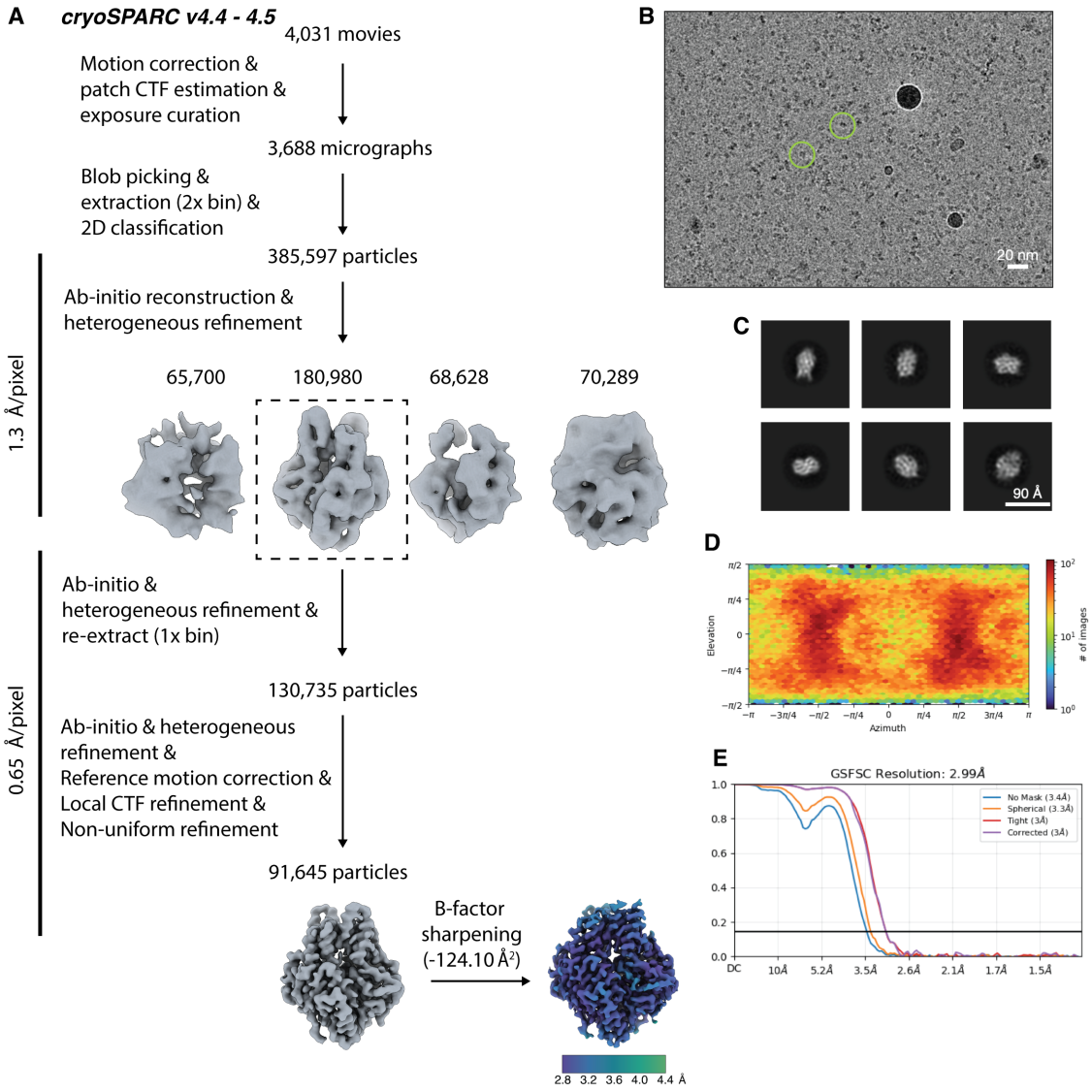


Figure S3.7: **Cryo-EM processing and validation for *Lf*ArsA•MgATP•As^{III} (non-hydrolyzing variant) structure.** (A) Cryo-EM data processing workflow in cryoSPARC. (B) Representative micrograph. (C) 2D class averages showing different views of ArsA in the 'closed' state. (D) Angular distribution heatmap plot. (E) Gold-standard Fourier shell correlation (GSFSC) curve.

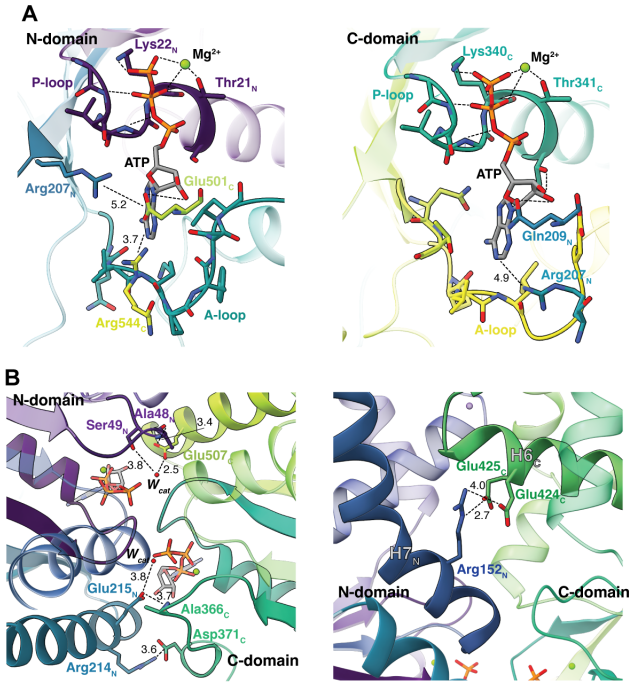


Figure S3.8: **Residue interactions observed in the ATP-closed state of *Lf* ArsA.** (A) Residues from both domains interacting with ATP bound at N- (left) and C-sites (right). (B) Inter-domain contacts formed at the pseudodimer interface. Left, top-down view and right, front view of the pseudodimer showing interactions between N- and C-domains. All interacting residues are shown as sticks. The structure referred to here is from the non-hydrolyzing ArsA variant colored using viridis.

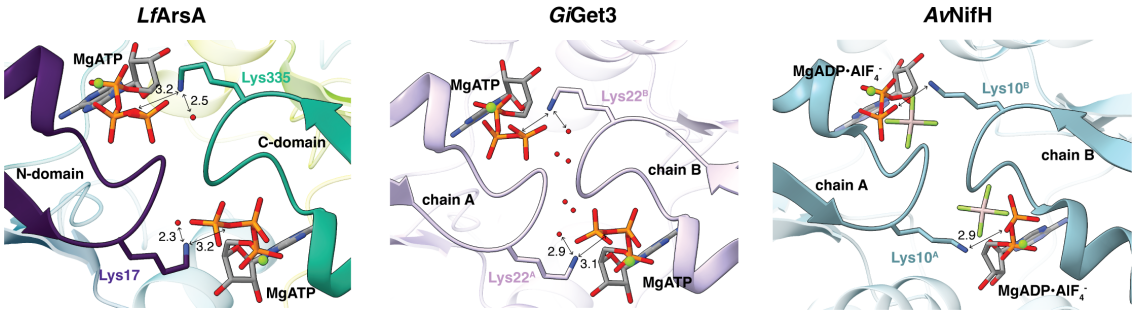


Figure S3.9: **Comparison of P-loop interactions at the pseudodimer interface of the ATP-closed state of *Lf* ArsA (left) with that of GiGet3 (center) and AvNifH (right) 'closed' dimer state structures.** Waters at the interface are shown as red spheres (Gi – *Giardia Intestinalis* and Av – *Azotobacter vinelandii*).

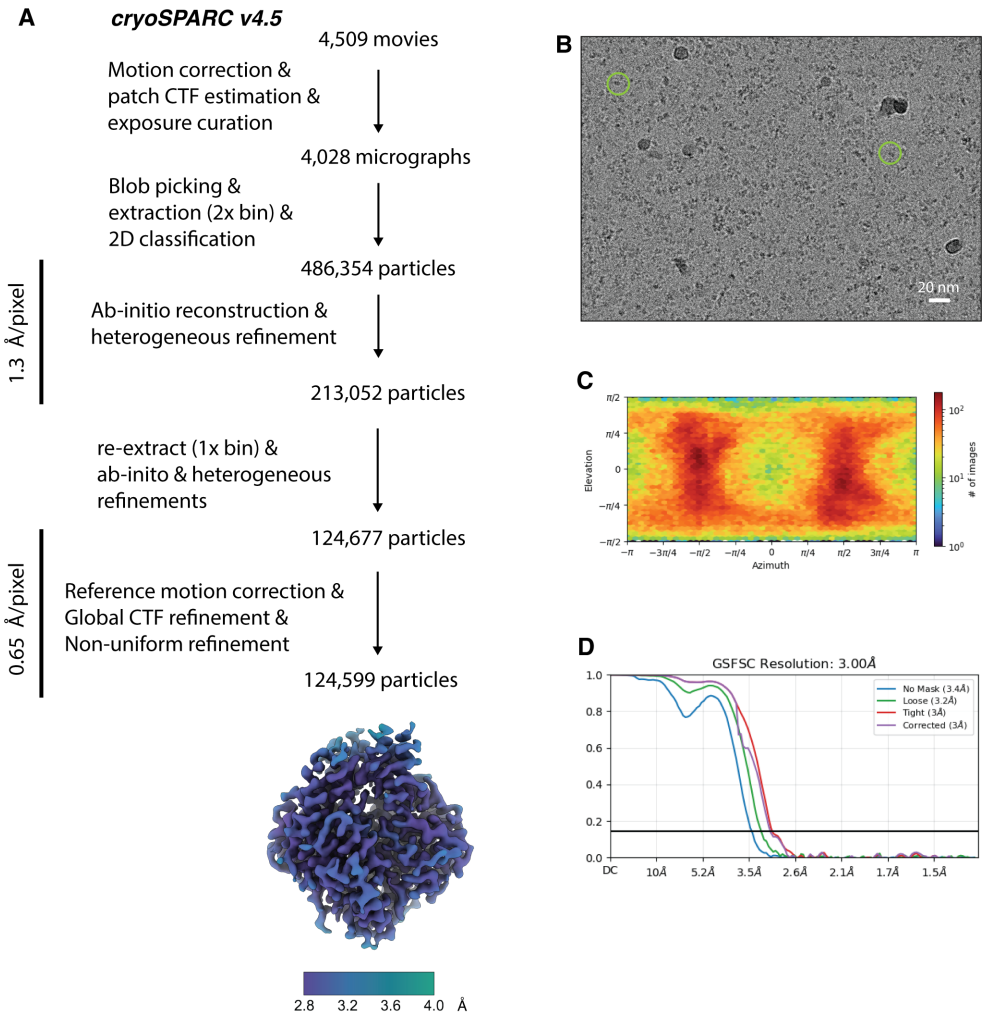


Figure S3.10: **Cryo-EM processing and validation for *Lf*ArsA•MgATP•As^{III} (wild-type) structure.** (A) Cryo-EM data processing workflow in cryoSPARC. (B) Representative micrograph. (C) Angular distribution heatmap plot. (D) Gold-standard Fourier shell correlation (GSFSC) curve.

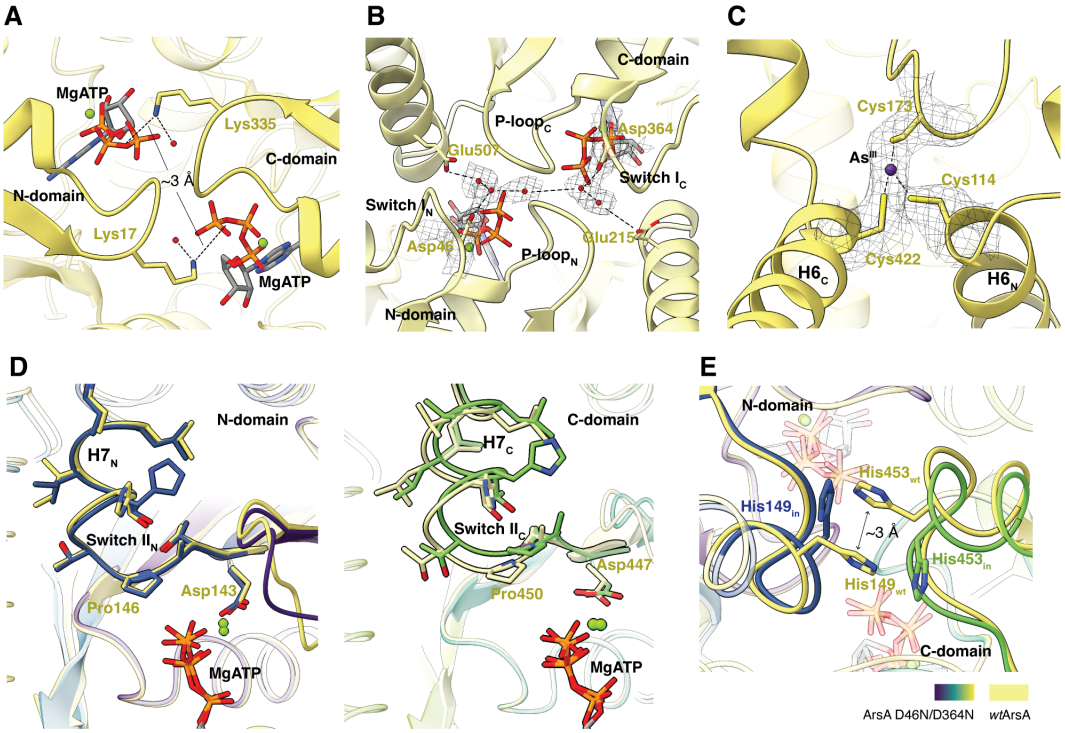


Figure S3.11: **Structural features observed in the ATP-closed state of wild-type *Lf* ArsA.** (A) Pseudodimer interface showing that the P-loops of the two domains are in proximity and the IWA lysines (Lys17 and Lys335) are both oriented towards the ATP bound to the opposite domain. (B) Hydrogen-bonding network of water molecules (red) and Asp/Glu residues connecting the N- and C-sites. (C) Metalloid-binding site showing Coulomb potential map for As^{III} (purple) enclosed between Cys114, Cys173 and Cys422. (D) Comparison of Switch II conformation between wtArsA (yellow) and non-hydrolyzing variant (viridis) 'closed' structures in the N- (left) and C- (right) domains. (E) His149 and His453 on helix 7 of N- and C-domains, respectively, form a stacking interaction in the wtArsA structure that is not observed in the non-hydrolyzing variant structure.

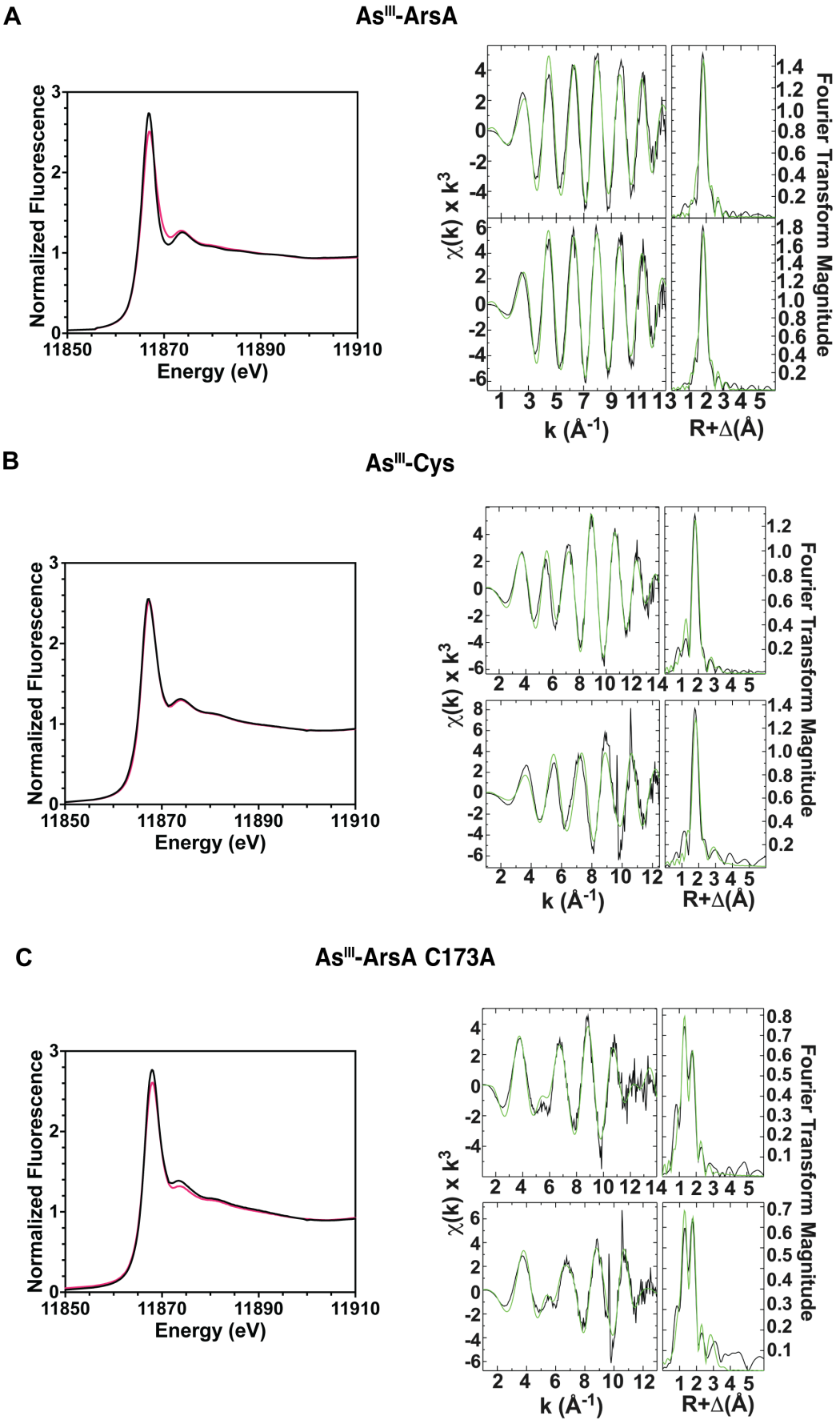

A
As^{III}-ArsA
Normalized Fluorescence
Energy (eV)
χ(k) x k^3
k (Å^{-1})
R+Δ(Å)
Fourier Transform Magnitude
B
As^{III}-Cys
Normalized Fluorescence
Energy (eV)
χ(k) x k^3
k (Å^{-1})
R+Δ(Å)
Fourier Transform Magnitude
C
As^{III}-ArsA C173A
Normalized Fluorescence
Energy (eV)
χ(k) x k^3
k (Å^{-1})
R+Δ(Å)
Fourier Transform Magnitude

Figure S3.12: **Duplicate XAS data for ArsA samples.** XANES (left) and EXAFS (right) spectra for (A) *Lf* ArsA•MgATP•As^{III} (As^{III}-ArsA), (B) As^{III}-L-cysteine control (As^{III}-Cys) and (C) *Lf* ArsA(C173A)•MgATP•As^{III} (As^{III}-ArsA C173A). Duplicate XANES spectra are shown in black and red.

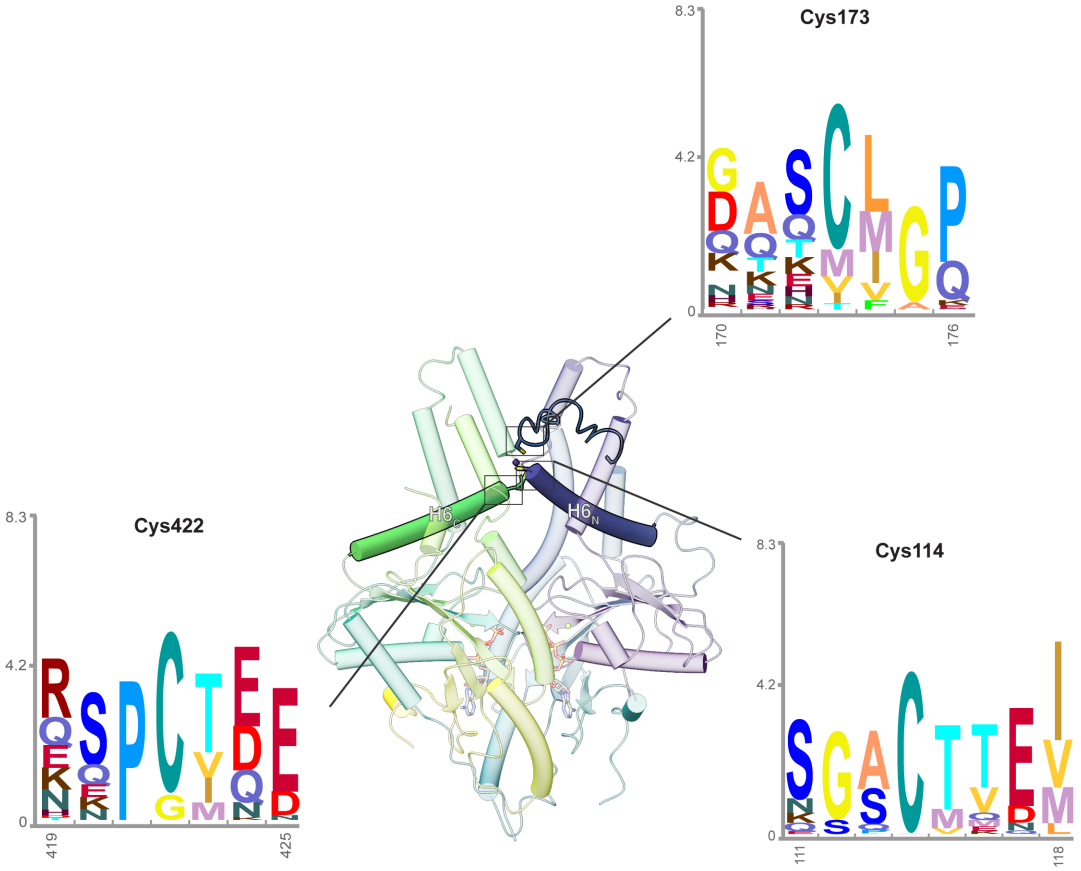


Figure S3.13: **Conservation of Cys residues of the metalloid-binding site across ArsA homologs.** Hidden Markov Model (HMM) logos depict conservation of Cys114, Cys173 and Cys422 necessary for As^{III} binding. Logos were constructed using 'Skylign' web tool [56] from a multiple sequence alignment of sequences assigned to the ArsA InterPro family (IPR027541).

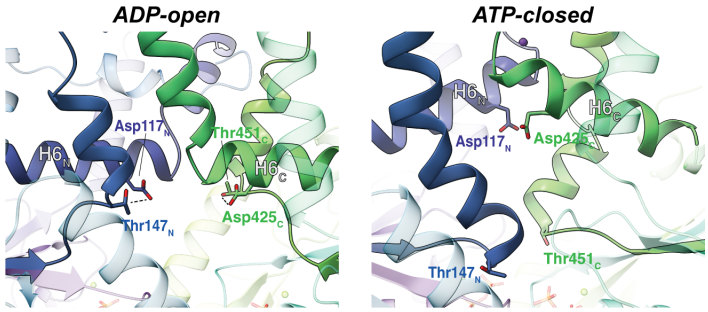


Figure S3.14: **Interaction between Thr residue (Switch II) and Asp residue (helix 6) observed in the ADP-open (left) state is destabilized in the ATP-closed (right) state as Switch II and helix 6 undergo conformational changes in both N- and C-domains.**

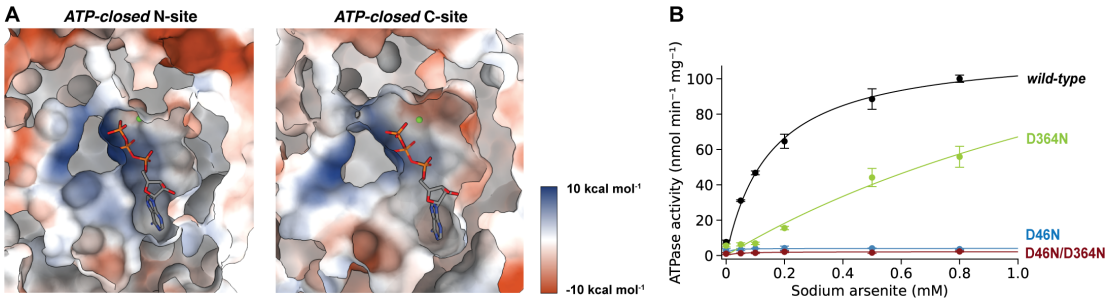


Figure S3.15: **Non-equivalence of N- and C-domains of ArsA.** (A) Electrostatic potential surface representation of N- and C-sites of *Lf* ArsA ATP-closed state, showing that ATP binds in a tighter active site pocket in the N-site than in the C-site. Surfaces were generated in ChimeraX using the 'coulombic' command. (B) Steady-state ATPase activity of ArsA (nmol min^{-1} mg^{-1}) plotted against varying As^{III} concentrations (mM), with 5 mM each of $MgCl_2$ and ATP at 37ºC (wild-type, black; D46N, blue; D364N, green and D46N/D364N, red). Data points represent mean of n = 3 and error bars represent standard deviations. The data was fit to the Michaelis-Menten equation.

REFERENCES

[1] D. D. Leipe et al. Classification and evolution of P-loop GTPases and related ATPases. In: *Journal of Molecular Biology* 317.1 (Mar. 2002), pp. 41–72. DOI: 10.1006/jmbi.2001.5378.

[2] I. R. Vetter and A. Wittinghofer. Nucleoside triphosphate-binding proteins: different scaffolds to achieve phosphoryl transfer. In: *Quarterly Reviews of Biophysics* 32.1 (Feb. 1999), pp. 1–56. DOI: 10.1017/s0033583599003480.

[3] E. V. Koonin. A Superfamily of ATPases with Diverse Functions Containing Either Classical or Deviant ATP-binding Motif. In: *Journal of Molecular Biology* 229.4 (Feb. 1993), pp. 1165–1174. DOI: 10.1006/jmbi.1993.1115.

[4] A. O. Maggiolo et al. Intradimeric Walker A ATPases: Conserved Features of A Functionally Diverse Family. In: *Journal of Molecular Biology*. Changing Shape of Molecules and Changing Landscape of Molecular Biology 435.11 (June 1, 2023), p. 167965. DOI: 10.1016/j.jmb.2023.167965.

[5] M. Y. Fry et al. Structurally derived universal mechanism for the catalytic cycle of the tail-anchored targeting factor Get3. In: *Nature Structural & Molecular Biology* 29.8 (Aug. 2022), pp. 820–830. DOI: 10.1038/s41594-022-00798-4.

[6] H. Schindelin et al. Structure of ADP·AIF4—stabilized nitrogenase complex and its implications for signal transduction. In: *Nature* 387.6631 (May 1997), pp. 370–376. DOI: 10.1038/387370a0.

[7] M. R. Ahmadian et al. Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP-hydrolysis reaction of Ras. In: *Nature Structural Biology* 4.9 (Sept. 1997), pp. 686–689. DOI: 10.1038/nsb0997-686.

[8] T. Jacobson et al. Arsenite interferes with protein folding and triggers formation of protein aggregates in yeast. In: *Journal of Cell Science* 125 (Pt 21 Nov. 1, 2012), pp. 5073–5083. DOI: 10.1242/jcs.107029.

[9] S. Shen et al. Arsenic binding to proteins. In: *Chemical reviews* 113.10 (2013), pp. 7769–7792.

[10] L. M. Mateos et al. Corynebacterium glutamicum as a model bacterium for the bioremediation of arsenic. In: *International Microbiology: The Official Journal of the Spanish Society for Microbiology* 9.3 (Sept. 2006), pp. 207–215.

[11] J. S. Y. Preetha et al. Biotechnology Advances in Bioremediation of Arsenic: A Review. In: *Molecules* 28.3 (Feb. 3, 2023), p. 1474. DOI: 10.3390/molecules28031474.

[12] I. Ben Fekih et al. Distribution of arsenic resistance genes in prokaryotes. In: *Frontiers in microbiology* 9 (2018), p. 2473. DOI: 10.3389/fmicb.2018.02473.

[13] C. M. Chen et al. Nucleotide sequence of the structural genes for an anion pump. The plasmid-encoded arsenical resistance operon. In: *Journal of Biological Chemistry* 261.32 (Nov. 1986), pp. 15030–15038. DOI: 10.1016/S0021-9258(18)66824-3.

[14] Y.-F. Lin, A. R. Walmsley, and B. P. Rosen. An arsenic metallochaperone for an arsenic detoxification pump. In: *Proceedings of the National Academy of Sciences* 103.42 (Oct. 2006), pp. 15617–15622. DOI: 10.1073/pnas.0603974103.

[15] J. Ye et al. The 1.4 A crystal structure of the ArsD arsenic metallochaperone provides insights into its interaction with the ArsA ATPase. In: *Biochemistry* 49.25 (2010), pp. 5206–12. DOI: 10.1021/bi100571r.

[16] S. Dey and B. P. Rosen. Dual mode of energy coupling by the oxyanion-translocating ArsB protein. In: *Journal of Bacteriology* 177.2 (1995), pp. 385–9. DOI: 10.1128/jb.177.2.385-389.1995.

[17] S. Dey, D. Dou, and B. P. Rosen. ATP-dependent arsenite transport in everted membrane vesicles of Escherichia coli. In: *Journal of Biological Chemistry* 269.41 (1994), pp. 25442–6. DOI: 10.1016/S0021-9258(18)47270-5.

[18] T. Zhou. Structure of the ArsA ATPase: the catalytic subunit of a heavy metal resistance pump. In: *The EMBO Journal* 19.17 (Sept. 2000), pp. 4838–4845. DOI: 10.1093/emboj/19.17.4838.

[19] H. Jia and P. Kaur. Role of the Linker Region of the Anion-stimulated ATPase ArsA: EFFECT OF DELETION AND POINT MUTATIONS IN THE LINKER REGION*. In: *Journal of Biological Chemistry* 276.31 (Aug. 1, 2001), pp. 29582–29587. DOI: 10.1074/jbc.M103042200.

[20] H. Bhattacharjee et al. Role of cysteinyl residues in metalloactivation of the oxyanion-translocating ArsA ATPase. In: *Journal of Biological Chemistry* 270.19 (1995), pp. 11245–50. DOI: 10.1074/jbc.270.19.11245.

[21] Y. Jiang et al. Nonequivalence of the nucleotide binding domains of the ArsA ATPase. In: *Journal of Biological Chemistry* 280.11 (2005), pp. 9921–6. DOI: 10.1074/jbc.M413391200.

[22] J. Pei, B.-H. Kim, and N. V. Grishin. PROMALS3D: a tool for multiple protein sequence and structure alignments. In: *Nucleic Acids Research* 36.7 (Apr. 2008), pp. 2295–2300. DOI: 10.1093/nar/gkn072.

[23] J. D. Thompson et al. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. In: *Nucleic Acids Research* 25.24 (Dec. 1997), pp. 4876–4882. DOI: 10.1093/nar/25.24.4876.

[24] M. M. Georgiadis et al. Crystallographic Structure of the Nitrogenase Iron Protein from *Azotobacter* vinelandii. In: *Science* 257.5077 (Sept. 1992), pp. 1653–1659. DOI: 10.1126/science.1529353.

[25] A. Mateja et al. Structure of the Get3 targeting factor in complex with its membrane protein cargo. In: *Science* 347.6226 (2015), pp. 1152–1155.

[26] C. J. M. Suloway et al. Model for eukaryotic tail-anchored protein binding based on the structure of Get3. In: *Proceedings of the National Academy of Sciences* 106.35 (Sept. 2009), pp. 14849–14854. DOI: 10.1073/pnas.0907522106.

[27] F. A. Tezcan et al. Structural Evidence for Asymmetrical Nucleotide Interactions in Nitrogenase. In: *Journal of the American Chemical Society* 137.1 (Jan. 2015), pp. 146–149. DOI: 10.1021/ja511945e.

[28] B. B. Wenke, T. Spatzal, and D. C. Rees. Site-Specific Oxidation State Assignments of the Iron Atoms in the [4Fe:4S]2+/1+/0 States of the Nitrogenase Fe-Protein. In: *Angewandte Chemie International Edition* 58.12 (2019), pp. 3894–3897. DOI: 10.1002/anie.201813966.

[29] A. R. Walmsley et al. The ATPase mechanism of ArsA, the catalytic subunit of the arsenite pump. In: *Journal of Biological Chemistry* 274.23 (1999), pp. 16153–61. DOI: 10.1074/jbc.274.23.16153.

[30] A. R. Walmsley et al. Antimonite regulation of the ATPase activity of ArsA, the catalytic subunit of the arsenical pump. In: *Biochemical Journal* 360.3 (Dec. 2001), pp. 589–597. DOI: 10.1042/bj3600589.

[31] X. Ruan, H. Bhattacharjee, and B. P. Rosen. Characterization of the metalloactivation domain of an arsenite/antimonite resistance pump. In: *Molecular microbiology* 67.2 (2008), pp. 392–402.

[32] X. Ruan, H. Bhattacharjee, and B. P. Rosen. Cys-113 and Cys-422 form a high affinity metalloid binding site in the ArsA ATPase. In: *Journal of Biological Chemistry* 281.15 (2006), pp. 9925–34. DOI: 10.1074/jbc.M600125200.

[33] S. Mi et al. Complete genome of Leptospirillum ferriphilum ML-04 provides insight into its physiology and environmental adaptation. In: *Journal of Microbiology* 49.6 (Dec. 1, 2011), pp. 890–901. DOI: 10.1007/s12275-011-1099-9.

[34] B. Li et al. Arsenic resistance operon structure in *Leptospirillum ferriphilum* and proteomic response to arsenic stress. In: *Bioresource Technology* 101.24 (Dec. 1, 2010), pp. 9811–9814. DOI: 10.1016/j.biortech.2010.07.043.

[35] I. M. Tuffin et al. Resistance Determinants of a Highly Arsenic-Resistant Strain of Leptospirillum ferriphilum Isolated from a Commercial Biooxidation Tank. In: *Applied and Environmental Microbiology* 72.3 (Mar. 2006), pp. 2247–2253. DOI: 10.1128/AEM.72.3.2247-2253.2006.

[36] J. Jumper et al. Highly accurate protein structure prediction with AlphaFold. In: *Nature* 596.7873 (2021), pp. 583–589. DOI: 10.1038/s41586-021-03819-2.

[37] G. Pintilie et al. Measurement of atom resolvability in cryo-EM maps with Q-scores. In: *Nature Methods* 17.3 (Mar. 2020), pp. 328–334. DOI: 10.1038/s41592-020-0731-1.

[38] T. Zhou et al. Conformational changes in four regions of the Escherichia coli ArsA ATPase link ATP hydrolysis to ion translocation. In: *Journal of Biological Chemistry* 276.32 (2001), pp. 30414–22. DOI: 10.1074/jbc.M103671200.

[39] B. Schmid et al. Biochemical and Structural Characterization of the Cross-Linked Complex of Nitrogenase: Comparison to the ADP-AlF -Stabilized Structure . In: *Biochemistry* 41.52 (Dec. 2002), pp. 15557–15565. DOI: 10.1021/bi026642b.

[40] J. Qin et al. Convergent evolution of a new arsenic binding site in the ArsR/SmtB family of metalloregulators. In: *Journal of Biological Chemistry* 282.47 (2007), pp. 34346–55. DOI: 10.1074/jbc.M706565200.

[41] W. Shi et al. The Role of Arsenic-Thiol Interactions in Metalloregulation of the *ars* Operon (∗). In: *Journal of Biological Chemistry* 271.16 (Apr. 19, 1996), pp. 9291–9297. DOI: 10.1074/jbc.271.16.9291.

[42] J. Yang et al. Arsenic Binding and Transfer by the ArsD As(III) Metallochaperone. In: *Biochemistry* 49.17 (May 4, 2010), pp. 3658–3666. DOI: 10.1021/bi100026a.

[43] J. Abramson et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. In: *Nature* 630.8016 (June 2024), pp. 493–500. DOI: 10.1038/s41586-024-07487-w.

[44] Y.-F. Lin, J. Yang, and B. P. Rosen. ArsD Residues Cys12, Cys13, and Cys18 Form an As(III)-binding Site Required for Arsenic Metallochaperone Activity*. In: *Journal of Biological Chemistry* 282.23 (June 8, 2007), pp. 16783–16791. DOI: 10.1074/jbc.M700886200.

[45] M. A. Lomize et al. OPM database and PPM web server: resources for positioning of proteins in membranes. In: *Nucleic Acids Research* 40 (Database issue Jan. 2012), pp. D370–376. DOI: 10.1093/nar/gkr703.

[46] L. Aravind, M. Y. Galperin, and E. V. Koonin. The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold. In: *Trends in Biochemical Sciences* 23.4 (Apr. 1, 1998), pp. 127–129. DOI: 10.1016/S0968-0004(98)01189-X.

[47] K. P. Locher, A. T. Lee, and D. C. Rees. The E. coli BtuCD Structure: A Framework for ABC Transporter Architecture and Mechanism. In: *Science* 296.5570 (May 10, 2002), pp. 1091–1098. DOI: 10.1126/science.1071142.

[48] C. E. Karkaria, C. M. Chen, and B. P. Rosen. Mutagenesis of a nucleotide-binding site of an anion-translocating ATPase. In: *Journal of Biological Chemistry* 265.14 (May 15, 1990), pp. 7832–7836.

[49] D. N. Mastronarde. Automated electron microscope tomography using robust prediction of specimen movements. In: *Journal of Structural Biology* 152.1 (Oct. 1, 2005), pp. 36–51. DOI: 10.1016/j.jsb.2005.07.007.

[50] M. Sun et al. Practical considerations for using K3 cameras in CDS mode for high-resolution and high-throughput single particle cryo-EM. In: *Journal of Structural Biology* 213.3 (Sept. 2021), p. 107745. DOI: 10.1016/j.jsb.2021.107745.

[51] A. Punjani et al. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. In: *Nature Methods* 14.3 (Mar. 2017), pp. 290–296. DOI: 10.1038/nmeth.4169.

[52] A. Punjani, H. Zhang, and D. J. Fleet. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. In: *Nature Methods* 17.12 (Dec. 2020), pp. 1214–1221. DOI: 10.1038/s41592-020-00990-8.

[53] P. D. Adams et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. In: *Acta Crystallographica Section D: Biological Crystallography* 66.2 (Feb. 1, 2010), pp. 213–221. DOI: 10.1107/S0907444909052925.

[54] P. Emsley and K. Cowtan. Coot: model-building tools for molecular graphics. In: *Acta Crystallographica Section D: Biological Crystallography* 60.12 (Dec. 1, 2004), pp. 2126–2132. DOI: 10.1107/S0907444904019158.

[55] G. Vogel and R. Steinhart. ATPase of Escherichia coli: purification, dissociation, and reconstitution of the active complex from the isolated subunits. In: *Biochemistry* 15.1 (Jan. 13, 1976), pp. 208–216. DOI: 10.1021/bi00646a032.

[56] T. J. Wheeler, J. Clements, and R. D. Finn. Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. In: *BMC Bioinformatics* 15.1 (Jan. 13, 2014), p. 7. DOI: 10.1186/1471-2105-15-7.

Chapter 4

STRUCTURAL AND MECHANISTIC STUDIES OF THE ARSENITE TRANSPORTER ARSB AND ITS INTERACTION WITH ARSA

This chapter is based on a manuscript in preparation:
Mahajan, S.; Demirer K.; Clemons Jr., W.M.; Rees, D.C. Structure and transport mechanism of the bacterial arsenite efflux transporter ArsB.

ArsB construct optimization discussed in section 4.2.2 was performed in collaboration with Yen Ju (Luke) Lu.
Development of arsenite resistance growth assays discussed in section 4.2.9 were performed in collaboration with Kemal Demirer.

ABSTRACT

ArsB, a member of the ion transporter (IT) superfamily, is the arsenite efflux transporter of the *ars* operon that confers arsenite resistance to bacteria. Arsenite transport across the membrane may either be driven by coupling to the proton-gradient across the membrane (secondary active antiport) or to ATP hydrolysis (primary active transport). The dearth of biochemical and structural characterization of the transporter in the literature has precluded understanding of the molecular mechanism of metalloid transport via ArsB. Both the mechanism of metalloid binding and translocation are unclear. In addition, a molecular interaction between ArsB and ArsA ATPase, that has implications for the ATP-coupled metalloid transport, is poorly understood. In this chapter, we report the first structure of ArsB (45 kDa) solved in detergent micelles by cryo-EM, without the use of any fiducial markers. The structure reveals an internal two-fold inverted repeat architecture resembling that of other well-characterized transporters of the IT superfamily. We solved structures of two distinct dimeric states of ArsB — a parallel dimer where the termini of both subunits are oriented to the same side of the micelle, and an antiparallel dimer where both subunits are oriented in an inverted topology. Structures determined in the presence of metalloid substrates, As^{III} and Sb^{III}, reveal the basis for substrate recognition of ArsB. Based on the 'elevator' model of transport, each of these structures reflect the 'inward-open' conformation of ArsB. Furthermore, we demonstrate that ArsB activates ATP hydrolysis activity of ArsA *in vitro*, indicating an interaction between the two proteins. Bioinformatic analysis and computational modeling do not support a conserved ArsAB heterodimer interface. We have developed an arsenite resistance growth complementation assay in *E. coli* to screen for variants of the ArsA and ArsB that will be leveraged to map an interaction interface of the complex. Together, these results constitute the groundwork for understanding the arsenite transport mechanism of ArsB, both as a secondary and primary active transporter.

4.1 Introduction

Integral membrane protein transporters span a wide range of biological functions from essential metabolite uptake to drug resistance. Active transporters carry out movement of their substrates against the electrochemical gradient across the biological membranes. A significant fraction of these transporters are 'secondary carriers' i.e. they couple the translocation of a substrate against its electrochemical gradient with the movement of a secondary molecule, typically a Na+ ion or proton (H^+), along its electrochemical gradient. In the case of H^+-coupled secondary transport, the electrochemical potential is a combination of the membrane potential and the pH gradient across the membrane, and is called the 'proton motive force' (PMF). The membrane potential, pH gradient, or both could drive the substrate transport.

ArsB is an integral membrane protein involved in the removal of toxic arsenite or antimonite in bacteria. Originally discovered to confer arsenite resistance in *S.aureus* and *E. coli* via the respective plasmid-encoded *ars* operons [1], *E. coli* expressing ArsB are unable to remove As^{III} when the PMF is disrupted using carbonyl cyanide m-chlorophenylhydrazone (CCCP), a proton uncoupler, thus indicating that ArsB functions as a H^+-coupled secondary transporter [2]. As arsenite exists in a neutral trihydroxylated form ($AsOH)_3$) at physiological pH [3], ArsB carries out 'electrogenic transport' of a neutral arsenite substrate using the PMF across the membrane [4]. As a model system for heavy metal(loid) transport in bacteria, ArsB is of interest from both mechanistic and bioremediation perspectives. No structural or biochemical study investigating the molecular mechanism of ArsB has been reported to date, primarily due to challenges with overexpression and purification of the membrane protein. The mechanism of arsenite export by ArsB remains unclear.

ArsB is classified under the Ion Transporter (IT) superfamily that includes highly versatile secondary transporters that move charged substrates across the cell membrane [5]. The transporter classification database (TCDB) categorizes ArsB under the 'arsenite-antimonite efflux family' (TCDB 2.A.45). Some of the families under the IT superfamily that have been structurally and mechanistic studied include the divalent anion-sodium symporters (DASS family, TCDB 2.A.47), p-aminobenzoyl glutamate transporters (AbgT family, TCDB 2.A.68) and the tripartite ATP-independent periplasmic transporters (TRAP family, TCDB 2.A.56) [6–8]. These proteins generally adopt a two-domain architecture — a scaffold domain that includes helices involved in dimerization in the case of DASS proteins, and a transport domain that carries out transport of substrate (Fig. 4.1). The membrane topology of ArsB

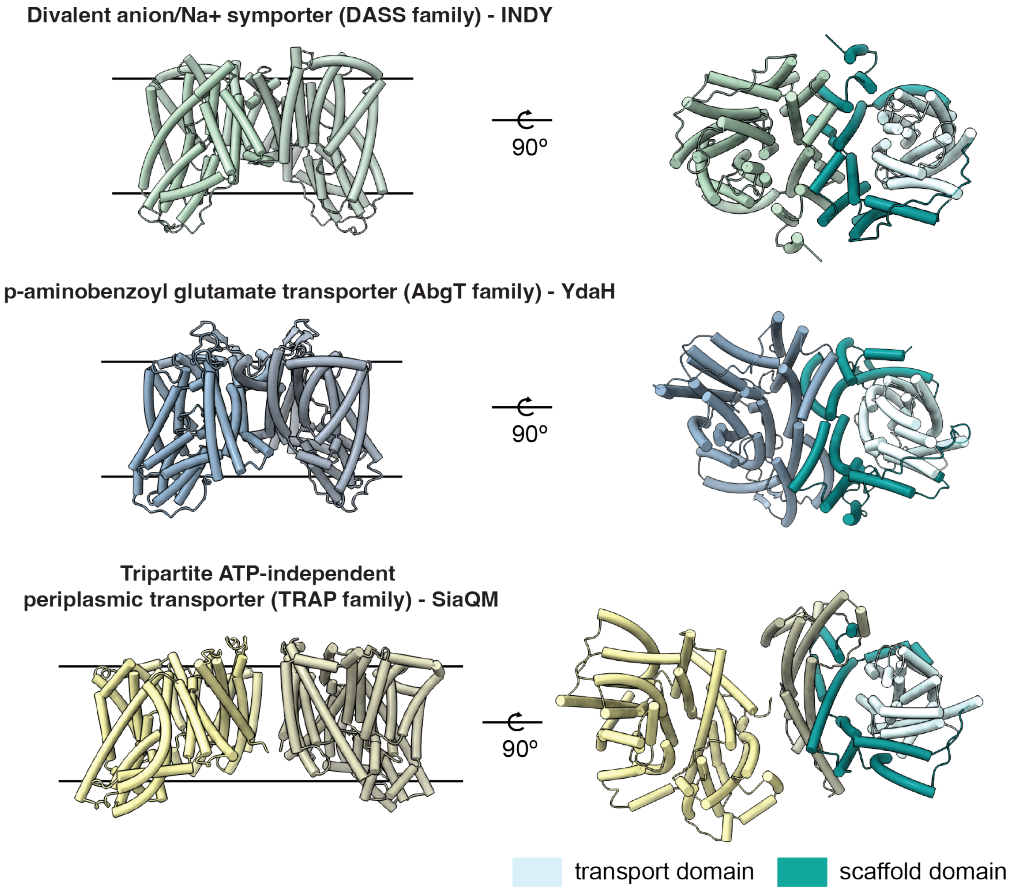


Figure 4.1: **Structure and domain architecture of some well-characterized transporters of the ion transporter (IT) superfamily.** Each transporter forms a dimer and the subunits are colored by chain. On the right, domain organization into transport domain (cyan) and scaffold domain (teal) is shown for one subunit of the dimer. PDB IDs – INDY, 7T9F; YdaH, 4R0C; SiaQM, 8THI.

has been studied only indirectly using invasive methods [9], suggesting that the membrane transporter is composed of 12 transmembrane α-helices arranged in two groups of six. AlphaFold2 prediction of ArsB revealed a high confidence model of the efflux pump (Fig. 4.2, left) [10]. Despite low sequence identity, this prediction revealed a structural fold that resembles other IT superfamily proteins (Fig 5.2). Proteins of the DASS, AbgT and TRAP families all bind and transport oxyanionic substrates. Many DASS proteins transport Krebb's cycle intermediates such as succinate, fumarate and malate [11], or other anionic substrates such as phosphate and sulfate [12, 13]. AbgT transporter, YdaH imports p-aminobenzoyl-glutamate for de novo folic acid biosynthesis [7]. TRAP transporters import carboxylates including C4-dicarboxylates, α-keto acids, amino acids and sialic acids [14]. DASS transporters typically function by 'Na^+-coupled electrogenic symport' of the substrate [15]. Additionally, in the Na^+-dependent dicarboxylate transporter from *Vibrio cholerae* (*Vc*INDY), substrate binding is facilitated by Na^+ binding first at conserved motifs in the core of the TMDs [15]. In contrast, arsenite is known to exist only in its charge-free $As(OH)_3$ form under physiological conditions, indicating that substrate recognition in ArsB may be divergent from other IT members, unless the binding site on ArsB triggers formation of a deprotonated state of arsenite. Binding of Na^+ (or other secondary ions) to ArsB has not been discussed or examined to date, likely due to the notion that ArsB operates via an 'H^+-coupled antiport' mechanism. Furthermore, while the AbgT and DASS transporters have been reported to exist only as dimers [7, 11], both monomer and dimer structures have been solved for TRAP transporters [8, 16], suggesting that oligomerization may not be critical for transport and likely has a regulatory role. Such oligomerization has not been investigated for ArsB. While domain architectures reveal conserved mechanistic features across the IT superfamily, differences relating to substrate recognition, secondary ion specificity and oligomerization might still exist, provoking the question whether the metalloid transport mechanism of ArsB is distinct from other proteins of the IT superfamily.

Many well-characterized secondary transporters facilitate substrate translocation by the 'alternating-access' mechanism where the membrane protein undergoes conformational changes between outward and inward states to alternate access of the substrate and the carrier ion between the two sides of the membrane. While several secondary transporters characterized to date operate by the 'rocking-bundle' or 'rocker-switch' mechanisms, alternating between cytoplasmic and periplasmic facing conformations [17], a distinct mechanism of transport, known as the 'elevator'

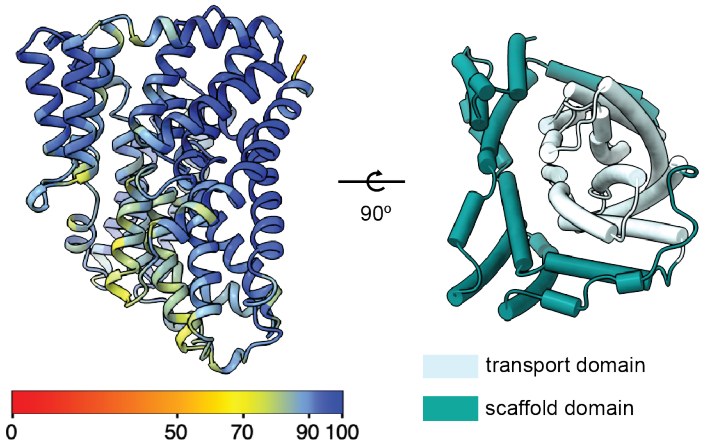


Figure 4.2: **AlphaFold2 model of *E. coli.* R773 ArsB colored by per-residue pLDDT scores (left), and highlighting conserved domain architecture of IT superfamily (right).**

mechanism, has been described for some transporter families [6, 18]. In contrast to the 'rocking-bundle' or 'rocker-switch' mechanisms where the substrate binding site remains stationary during transport, the 'elevator' mechanism is characterized by a vertical 'elevator-type' movement of the substrate binding site along the width of the membrane to facilitate transport. This mechanism was first described for the glutamate transporter, Glt_{Ph} [19]. Recently, several members of the divalent anion-Na^+ symporter family have been elucidated to adopt the 'elevator' mechanism of transport [6, 12, 13, 20]. Additionally, the TRAP transporter, SiaQM, that imports sialic acid in bacteria, functions by an 'elevator with an operator' mechanism [21], where an extracytoplasmic substrate-binding subunit, SiaP, facilitates conformational changes during the transport cycle. Although these studies argue that the 'elevator' mechanism is a conserved feature across the IT superfamily, whether ArsB fits in this mechanistic framework remains to be elucidated. Importantly, how the proton-gradient across the membrane can drive these elevator-type conformational changes in ArsB to facilitate vectorial transport of arsenite is of key interest.

What makes ArsB unique as a member of the IT superfamily is that in addition to an 'H^+-coupled antiporter', it can also function as an ATP-coupled primary active transporter. The ATP-coupled arsenite transport is facilitated by the IWA ATPase,

ArsA (see Chapter 3). ArsB is sufficient to confer resistance to arsenite. However, in presence of ArsA, ArsB removes As^{III} from the cells much more efficiently than just ArsB alone (Fig. 1.2) [2]. These observations suggested a dual mode of energy coupling for arsenite efflux via ArsB. Based on this model, ArsB functions as a secondary transporter to confer resistance to intermediate concentrations of arsenite, whereas under high arsenic stress, ATP-dependent primary active transport is facilitated by ArsA. Two questions in regard to the mechanism of efflux by the ArsAB pump are relevant — how do ArsB and ArsA interact at the molecular level? and how does this interaction substitute for H^+-coupling to remove the toxic metalloid more efficiently than just ArsB alone? Recombinantly expressed ArsA was shown to localize to the membrane only when ArsB was present, suggesting that ArsB is the membrane anchor for the cytosolic ArsA [22]. This result argues that a stable complex exists between ArsA and ArsB. Apart from this result, it is unclear whether ArsA actually forms a stable interface with ArsB.

In the studies reported in this chapter, we sought to answer some of the questions concerning the metalloid transport mechanism of ArsB primarily using a structural approach. We first investigate the distribution of arsB genes across species using bioinformatics, in an attempt to distinguish ArsB sequences that are associated with ArsA from the ones that are not. Next, we demonstrate how a GFP-fusion-construct-based strategy for overexpression and purification of ArsB enabled us to perform single particle cryo-EM studies, resulting in the first structures of ArsB from *L. ferriphilum* (*Lf* ArsB) in apo and metalloid (As^{III} and Sb^{III})-bound states. These structures, supported by site-directed mutagenesis, provide insights into metalloid recognition and transport mechanism of ArsB. Finally, we discuss efforts toward mapping the interaction between ArsB and ArsA for the characterization of an ArsAB complex. Computational modeling results in a few insights into the nature of interactions at the interface. We discuss the development of a growth complementation assay in *E. coli* for rational screening of putative interaction variants for mapping an ArsAB interface. Collectively, this work lays the foundation for a mechanistic elucidation of toxic metalloid efflux by ArsB and more broadly, the investigation of mechanisms of energetic coupling in heavy metal transporters across biological systems.

4.2 Results

4.2.1 Bioinformatic analysis of arsB gene

The arsB gene is found to exist in broadly two forms of *ars* operon architectures in prokaryotes. The first form is a standalone secondary active transporter gene with the arsR repressor gene, along with, in many cases, an arsC gene encoding an arsenate reductase. The second form is in association with the arsA gene, and thus likely functions as a primary active transporter. We examined the sequence diversity of arsB genes in genome databases, particularly interested in the distribution of arsB genes that were found in association with or isolation of arsA gene, in the *ars* operon context. Additionally, we were interested to learn about any sequence-level hallmarks that differentiated the two types of ArsB sequences. To accomplish this, we constructed a sequence similarity network (SSN) of ArsB using the Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST) [23]. SSNs are a very useful tool to visualize and examine sequence-level diversity in protein families and resolve novel functional features. Initially, to retrieve ArsB protein sequences, we performed a BLAST search using *E. coli* ArsB sequences with a sequences length filter of 400-500 residues, and then constructed an SSN. The resulting SSN included a total of 5,194 non-redundant sequences, organized into nodes, each of which was a 'representative node' with a sequence identity (ID) threshold of 90% (Fig. 4.3A). Broadly, ArsB sequences cluster into two large cluster of gram-positive (blue nodes) and gram-negative (red nodes) bacteria. Only 1,260 nodes of these, corresponding to 3,681 non-redundant sequences, were present in *ars* operon context, filtered by looking for arsR as a neighboring gene (Pfam IDs PF01022 and PF12840) (Fig. 4.3B, beige nodes). Of this filtered subset of ArsB sequences, only 148 nodes, corresponding to 813 non-redundant sequences, represented those associated with an arsA gene within the operon (Fig. 4.3B, pink nodes). Interestingly, the ArsA-associated ArsB sequences were scattered throughout the SSN and depicted no clustering based on global sequence-level relationships alone, whatsoever. This analysis not only explores the distribution of ArsB across species, but also opens up questions about the dual mode of active transport in ArsB.

4.2.2 Expression and purification of ArsB

Toward the goal of biochemical and structural characterization of ArsB, we attempted to overexpress and purify this membrane protein. ArsB has been reported to be a challenging candidate to express and purify for biochemical studies [24]. Owing to these difficulties, we decided to first exploit recent developments in *E.*

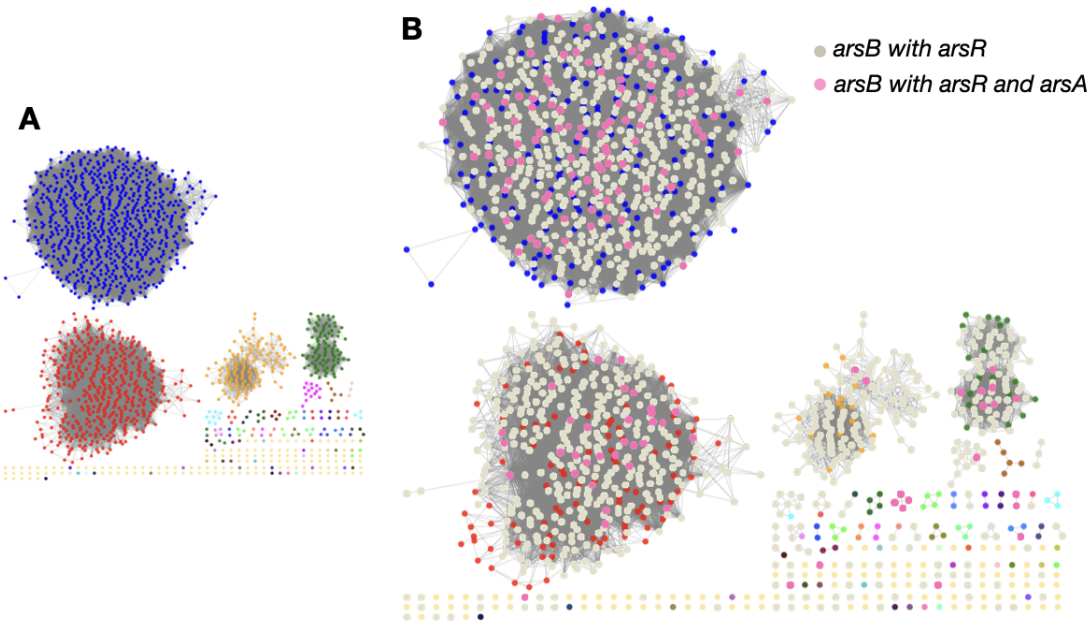


Figure 4.3: **Sequence Similarity Network (SSN) of ArsB sequences.** (A) Full SSN where each node represents sequences with over 90% identity. Nodes are clustered using an alignment score threshold of 160 and are colored arbitrarily by cluster number. (B) SSN in panel A with 'arsB genes with ArsR' (beige) and 'arsB genes with arsR and arsA' (pink) highlighted.

coli expression strains engineered for toxic membrane protein expression. As an initial screen for expression, we used a hybrid strain of the *E. coli* Lemo21 (DE3) and *E. coli* Nico21 (DE3) strains purchased from New England Biolabs (referred to as Nimo21 (DE3)), that enables improved membrane protein expression using a tunable T7 expression system and allows for removal of common *E. coli* contaminants at the affinity chromatography step. Initially, *Ec*ArsB in a pETDuet backbone with either an N-terminal or C-terminal 6x-His affinity tag was screened for overexpression in this strain. While overexpression of a protein was not evident, in-gel trypsin digest and mass spectrometry analysis of the C-terminal 6x-His construct purification revealed that the subunits of 'cytochrome bo(3) ubiquinol oxidase', not ArsB, were expressed instead (Table S4.1). Overexpression of the ArsB ortholog from *L. ferriphilum* (*Lf* ArsB with a C-terminal 6x-His tag) was not successful either. These expression trials reaffirmed the challenges associated with recombinant overexpression of ArsB.

Subsequently, we identified an alternative strategy to overexpress ArsB. Hsieh and

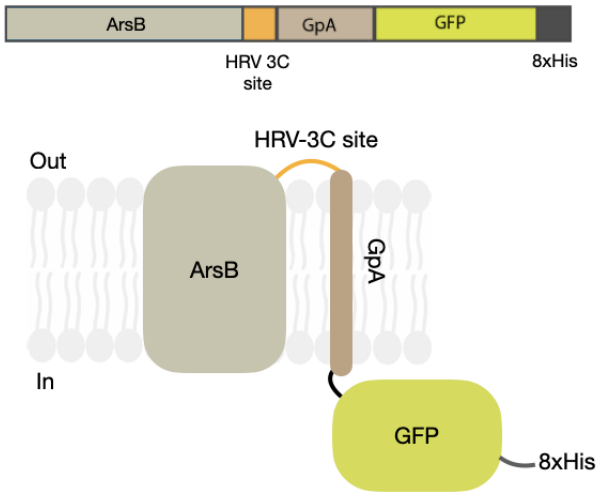


Figure 4.4: **ArsB expression construct with a C-terminal glcophorin A (GpA) segment and superfolder GFP-fusion.** This construct is derived from the pWarf(+) construct in Hsieh *et al.* [25].

coworkers used a GFP-based fusion construct to assist with the expression of difficult membrane proteins [25]. Interestingly, *Ec*ArsB was one of their candidates. Membrane topology analysis of ArsB predicts that both its termini must be localized to the periplasm [26]. Based on this assumption and considering that GFP can only fold properly in the cytoplasm, they used a pWarf(+) expression construct of ArsB including a C-terminal single-pass transmembrane domain, glycophorin A (GpA) followed by GFP, which was followed by a His tag. Adding GpA ensured that GFP localized to the cytoplasm, as translocation of the GFP to the periplasm would not have been feasible without a signal sequence. This strategy not only dramatically improved the expression yields due to solubility improvement of the target membrane protein owing to the GFP fusion, but also allowed for tracking the fusion construct during the course of purification, after which the fusion proteins can be cleaved off using a protease, resulting in stable membrane protein.

We designed a similar construct for *Ec*ArsB and *Lf*ArsB and inserted them into pETDuet backbone to test for expression (Fig. 4.4). Human rhinovirus (HRV)-3C cleavage site with a 5'-2xGS linker was inserted between ArsB and GpA to enable GpA-GFP cleavage after purification. *Ec*ArsB could be overexpressed, consistent with the results of Hsieh and coworkers (Fig 4.5A); however, in our hands, the protein aggregated overnight indicating issues with stability of cleaved *Ec*ArsB. We repeated the expression of the GFP-fusion construct with *Lf*ArsB and obtained

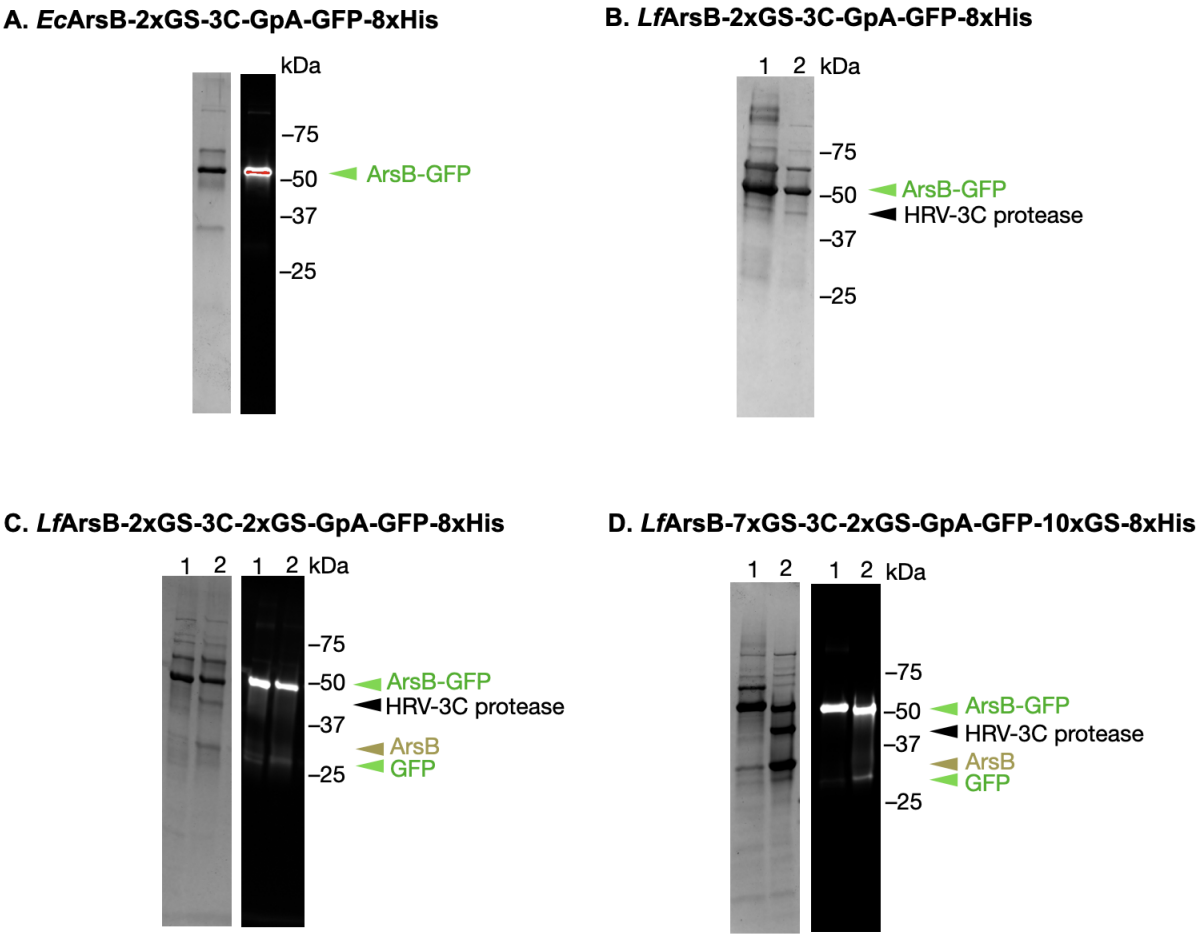


Figure 4.5: **Purification of C-terminal 8xHis-tagged constructs of *Ec*ArsB-GpA-GFP and *Lf* ArsB-GpA-GFP.** (A-D) SDS-PAGE for various constructs after Ni-affinity chromatography, showing Coomassie Blue-stained image (left) and Pro-Q emerald 488 nm image (right). When labeled, 1 and 2 indicate sample before and after cleavage by HRV-3C protease.

overexpressed protein (Fig 4.5B), that appeared to be more stable than the *Ec*ArsB construct. Unfortunately, this construct could not be successfully cleaved with commercial HRV-3C-protease to obtain cleaved *Lf*ArsB (Fig 4.5B). Speculating that the protease cleavage site could be sterically inaccessible to the protease on the *Lf*ArsB construct, in addition to the 2xGS linker on the 5'-end of the cleavage site, we attached a 2xGS linker on the 3'-end as well. Although the fusion protein overexpressed and purified well, only partial cleavage was observed (Fig. 4.5C). To achieve high cleavage efficiency and to potentially enhance purification yields, we designed variants of this construct by replacing the HRV-3C cleavage site with SUMOEu1 or TEV cleavage sites [27], flanked by GS linkers. Neither of the two constructs resulted in sufficient yields of *Lf*ArsB to test for cleavage efficiency (Fig. S4.1). Additionally, yield remained low when the His tag was moved to the N-terminus of the construct (Fig. S4.2). Eventually, we leveraged the HRV-3C construct for optimization by inserting a 5'-7xGS linker and a 3'-2xGS linker around the 3C cleavage site and used homemade, instead of commercial, protease to maximize cleavage efficiency. A 10xGS linker was also inserted between GFP and the 8xHis tag to further enrich the protein during purification. These modifications resulted in a drastically improved cleavage efficiency of the LfArsB fusion protein compared to previous trials (Fig. 4.5D). In-gel trypsin digest and mass spectrometric analysis of the SDS-PAGE bands identified as uncleaved and cleaved ArsB confirmed that *Lf*ArsB was the most abundant protein in the samples with a coverage of 29%, albeit with some ubiquinol oxidase contamination also present in the samples (Table S4.2). As *Lf*ArsB resulted in higher purification yield and stability compared to *Ec*ArsB, all subsequent biochemical and structural characterization was performed using *Lf*ArsB. Moreover, while initial constructs of ArsB were extracted and purified in DDM, other detergents that are frequently noted in the literature to yield high-resolution structures were also screened for protein stability and suitability with downstream cryo-EM studies, including n-decyl-β-D-maltopyranoside (DM) and lauryl maltose neopentyl glycol/cholesteryl hemisuccinate (LMNG/CHS).

4.2.3 Cryo-EM characterization of ArsB

When purified in DDM or LMNG/CHS, size-exclusion chromatography (SEC) of *Lf*ArsB resulted in two distinct peaks — one eluting approximately at 11 mL annotated as Peak 1 (P1) and another eluting approximately at 12 mL annotated as Peak 2 (P2). Additionally, a shoulder was consistently observed before P1 (Fig. 4.6A and B). In the case of DDM, both P1 and P2 eluted at similar intensities,

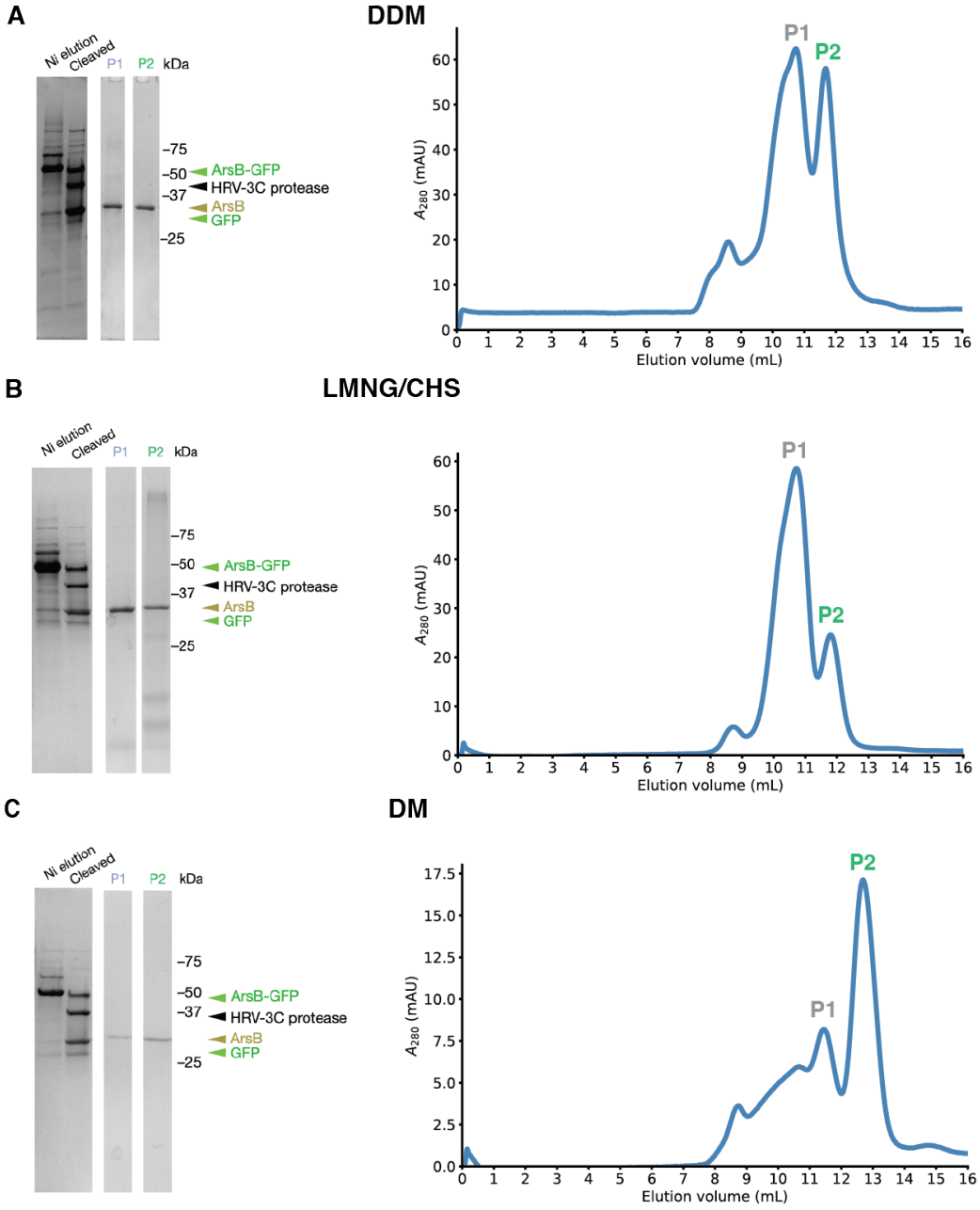


Figure 4.6: **Purification of *Lf* ArsB-7xGS-3C-2xGS-GpA-GFP-10xGS-8xHis in three different detergents.** (A-C) SDS-PAGE (left) and SEC profile (right) for ArsB purified in DDM, LMNG/CHS and DM, respectively. P1 and P2 fractions correspond to peaks 1 and 2, respectively on the SEC.

whereas in the case of LMNG/CHS, P1 was the predominant peak. When purified in DM, P2 became the predominant peak (Fig. 4.6C). Both peaks had similar purity across detergents, based on SDS-PAGE analysis (Fig. 4.6). We hypothesized that the multiple peaks corresponded to a dimer-monomer equilibrium, as both dimers and monomers have previously been characterized for IT superfamily members [6, 8]. When P1 fractions were subjected to a second round of gel filtration, P2 re-emerged, supporting a putative ArsB dimer dissociating into monomers (Fig.S4.3). We performed cryo-EM analysis on both P1 and P2 samples of DDM-solubilized ArsB to examine their compositions. Two-dimensional (2D) class averages for sample P2 revealed detergent micelles with transmembrane domain (TMD)-like features enclosed; however, a 3D reconstruction only resulted in a low-resolution map indicating a monomer within the micelle (Fig. S4.4). Consistent with our hypothesis, for sample P1, both 2D and 3D classes revealed two distinct sets of TMDs suggesting a dimer, albeit also at low resolution (Fig. S4.5).

4.2.4 Structure of apo ArsB in 'inward-open' conformation

High-resolution cryo-EM structures of the apo form of an ArsB dimer was obtained in LMNG/CHS. From a cumulative dataset of 10,625 movies and utilizing a deep learning-based model, Topaz, for particle picking [28], 616,321 particles were extracted from the micrographs. 2D class averages were obtained in various orientations of the micelle-enclosed ArsB (Fig. S4.6A and B). From the first round of *ab-initio* 3D reconstruction, both dimer and monomer were revealed. While particles of the monomer could not be refined to an atomic resolution map, the two dimeric 3D classes could be refined to atomic resolution and resulted in two dimers each having a distinct two-fold axis (Fig. S4.6C and D). After several rounds of 3D reconstructions and refinements, the two dimers were resolved to overall resolutions of 3.6 Å and 3.3 Å respectively (Fig. S4.6C-D and S4.7). One is a 'parallel dimer' with a two-fold rotation axis perpendicular to the plane of the micelle with the monomers oriented so that the respective termini are localized on the same side of the micelle (Fig. 4.7A and C). In contrast, the second is an 'antiparallel dimer' with a two-fold axis in the plane of the micelle and the respective termini localized to opposite sides of the micelle(Fig. 4.7B and D). Despite two distinct dimeric states, all ArsB subunits are in the same conformation. ArsB is composed of a total of 10 TMDs and 2 sets of hairpin helices, arranged into 'scaffold' and 'transport' domains (Fig. 4.8A & B). As for related proteins, ArsB has an inverted repeat fold. TM1-5 are related to TM6-10 by pseudo-two-fold symmetry (Fig. 4.8C & D). As part

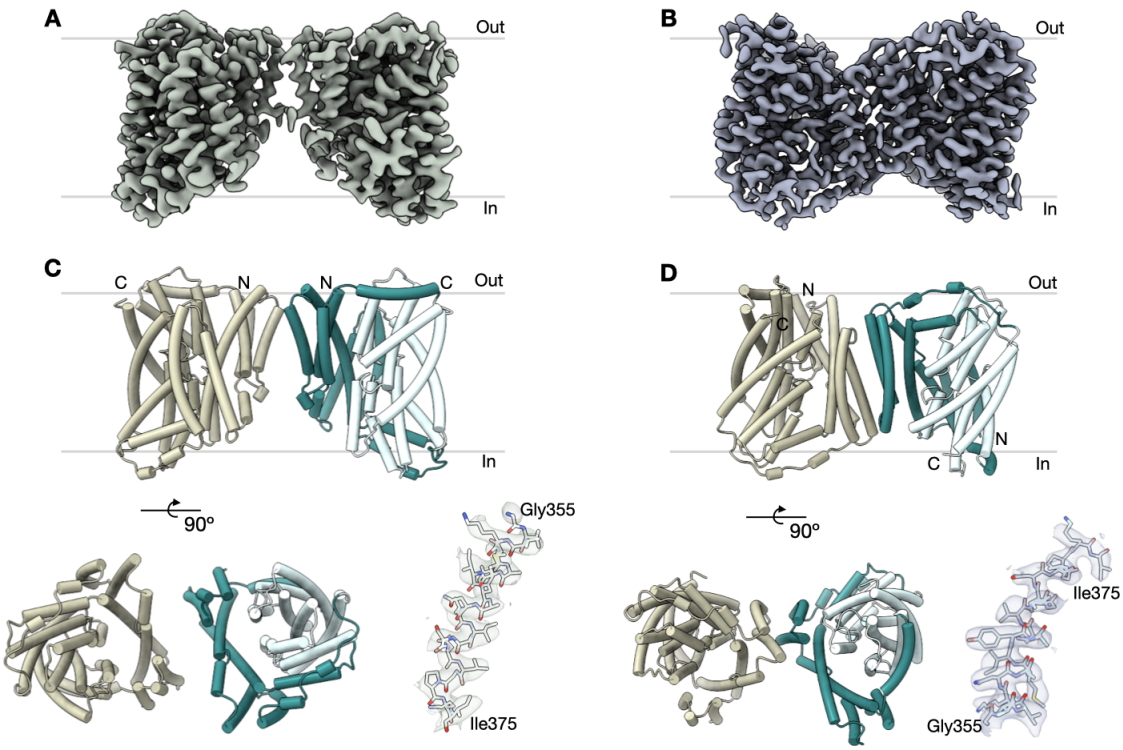


Figure 4.7: **Cryo-EM structures of *Lf*ArsB with parallel and antiparallel dimers.** (A) Sharpened 3.6 Å map of a parallel dimer. (B) Sharpened 3.3 Å resolution map of a antiparallel dimer. (C) Cartoon model of the parallel dimer. (D) Cartoon model of the antiparallel dimer. For both dimer models, one chain is colored by domain (cyan - transport; teal - scaffold). Model fit to map is shown for a representative transmembrane helical segment Gly355-Ile375 for both dimers.

of this symmetry, a helix-turn-helix hairpin (HP1) is formed by helices HP1a and HP1b and the discontinuous helix 4 is formed by TM4a and TM4b, with a break at loop L1, in the N-terminal half (Fig. 4.8B). Likewise, hairpin (HP2) is formed by helices HP2a and HP2b and the discontinuous helix 9 is formed by TM9a and TM9b, with a break at loop L2, in the C-terminal half (Fig. 4.8B). These two motifs form the core of the transport domain. The interface of the scaffold and transport domains is largely composed of hydrophobic residues, consistent with the DASS family transporters [6, 12].

The top homologs from a DALI server [29] search for ArsB reveals the DASS transporters Na^+-dicarboxylate symporter from *V. cholerae* (*Vc*INDY) and the phosphate transporter 90 from *S. cerevisiae* (*Sc*Pho90) as the closest in structure. Comparison to these homologs revealed a conserved secondary structural arrangement characteristic of the IT superfamily fold (Fig. 4.9A). Apo ArsB structure aligns very well

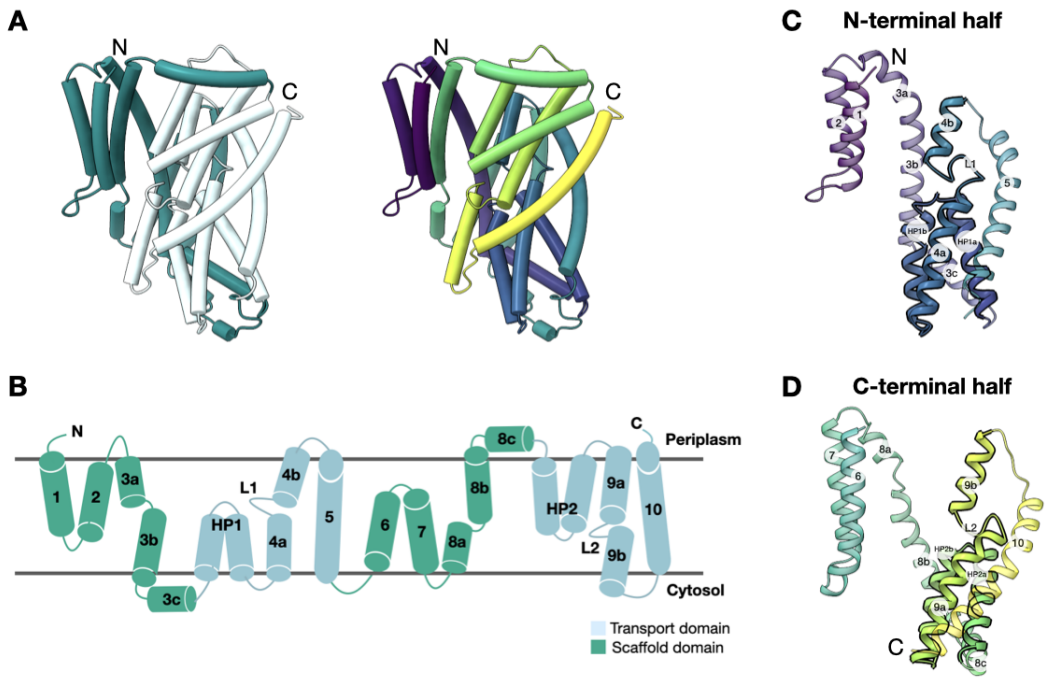


Figure 4.8: **Topology and transmembrane domain (TMD) architecture of ArsB structure.** (A) Cartoon models of ArsB colored by domain (left) and by TMDs using viridis (right). (B) Membrane topology diagram of ArsB. (C) The architecture of the N-terminal half of ArsB oriented the same way as in panel A, showing labeled helices and loops. (D) Architecture of the C-terminal half of ArsB after 180º rotation along the axis in the membrane plane and then along the axis going through the membrane plane, showing labeled helices and loops.

with the apo VcINDY and ScPho90 structures, indicating that ArsB may function via the conserved 'elevator mechanism' of transport of substrate as postulated for VcINDY, Pho90, and other IT superfamily transporters. By comparison to both VcINDY and Pho90, this ArsB conformation is described to be the 'inward-facing' or 'inward-open' conformation with the substrate binding pocket open to the cytoplasmic side. Structural motifs of the transport domain of our ArsB structure in the 'inward-open' conformation form minimal contact with the scaffold domain on the cytoplasmic side, opening a pocket into the core of the TMDs mostly lined with hydrophobic residues and a small patch of negative potential (Fig. 4.9B). This pocket could be accessed by arsenite to reach the substrate binding site at the core of the membrane. Unlike VcINDY and Pho90 which are substrate importers, ArsB exports its substrate. We predict that this 'inward-facing' conformation of ArsB binds arsenite to facilitate export out of the cell.

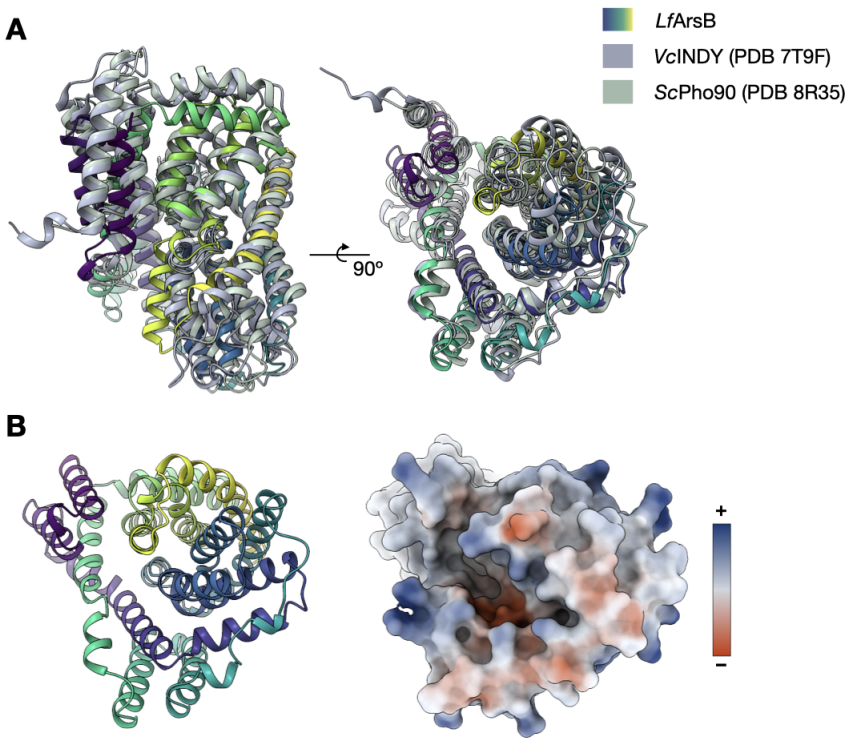


Figure 4.9: **Inward-open conformation of *Lf*ArsB.** (A) Comparison of *Lf*ArsB with 'inward-open' structures of *Vc*INDY (green) and *Sc*Pho90 (gray). (B) Cytoplasmic view of *Lf*ArsB showing the relative positions of the TMDs that result in an opening of the transporter on the cytoplasmic side. Electrostatic surface representation (right) reveals a hydrophobic cavity with a negative electrostatic patch that may facilitate metalloid binding. The electrostatic surface was calculated using *coulombic* command in *ChimeraX*.

4.2.5 The arsenite binding pocket of ArsB

To confirm the binding of arsenite to the 'inward-open' state, we incubated ArsB purified in LMNG, with 2 mM arsenite and determined a cryo-EM reconstruction at 3.4 Å (Fig. S4.8). We see particles consistent with parallel and antiparallel dimers and additionally observe a tetrameric complex that appears to be a dimer of dimers (Fig. S4.8). Of these, the parallel dimer gave the best resolved structure. Overall, the monomer in this dimer was identical to the apo ArsB structure (r.m.s.d. = 0.4 Å) (Fig. 4.10A). To observe high-resolution features within the parallel dimer, we masked out the poorly resolved inverted subunits from the map and performed a focused refinement on just the parallel dimer. This improved the quality of the map with a marginal improvement in the overall resolution to 3.3 Å (Fig. S4.8). Closer inspection of the core of the TMDs in this map revealed additional density that was consistent with a substrate binding site. We modeled this density with an

$As(OH)_3$ moiety (Fig. 4.10 A, right and B). The arsenite binding site is nestled in the core of the transport domain between the HP1-L1 (N-terminal half) and HP2-L2 (C-terminal halves) motifs (Fig. 4.10B).

The arsenite binding site is composed of polar residues located on the helix turns of HP1 and HP2, and L1 and L2 loops, including Asn111, Asp112, Asn158, Asn337 and Ser380, all of which are within 3-4 Å from As^{III} (Fig. 4.10C). The polar pocket is surrounded by hydrophobic residues, including Phe60, Leu159, Val160, Leu381, and Met338. Although $As(OH)_3$ is considered to be the aqueous form of As^{III} due to its high first pK_a of 9.2, we did not rule out the possibility of modeling a positively-charged As^{III} ion or a deprotonated $As(OH)_3$ into the density. However, based on the chemical properties of residues surrounding this site, neither of the two charged species seems likely from a chemical standpoint. Although the resolution of the map precludes unambiguous modeling of the metalloid, putative hydrogen-bonding residues such as Asn and Asp surrounding the binding site support binding of a trihydroxylated form of As^{III}. As^{III} binding to ArsB is likely facilitated by hydrogen-bonding interactions with polar side-chains at the core of the protein. To further validate metalloid binding at this site, we determined the cryo-EM structure of ArsB in the presence of 2 mM antimonite (Sb^{III}) and observed density corresponding to the metalloid at the same site with side-chain interactions preserved (Fig. 4.10D and S4.9). Notably, this binding mode is strikingly different from As^{III} binding observed in ArsA, where three conserved Cys residues coordinate the metalloid in a trigonal fashion (Refer to Chapter 3).

The As^{III} binding site in ArsB is consistent with the substrate binding site of other IT proteins, such as the divalent anion symporter *Vc*INDY (Fig. 4.10E). A distinctive feature of ArsB is the absence of density corresponding to Na^+ ions in our EM maps that have been consistently observed in the structures of other IT superfamily members [6, 30]. Two Na^+ ions typically bind at the interfacial pocket formed by the hairpin and the loop region of the discontinuous helix of the transport domain. Comparison of these structural motifs in ArsB with that of *Vc*INDY reveals a smaller binding cleft formed by HP1-L1 and HP2-L2 motifs in ArsB (Fig. 4.10E). It is unlikely that Na^+ binding is compatible with this conformation, even in the presence of its metalloid substrate. From a functional standpoint, the Na^+ ions contribute to charge compensation of the anionic substrates in DASS transporters, as well as providing the electrochemical gradient to drive substrate translocation across the membrane [15]. For ArsB, charge compensation is not necessary due to

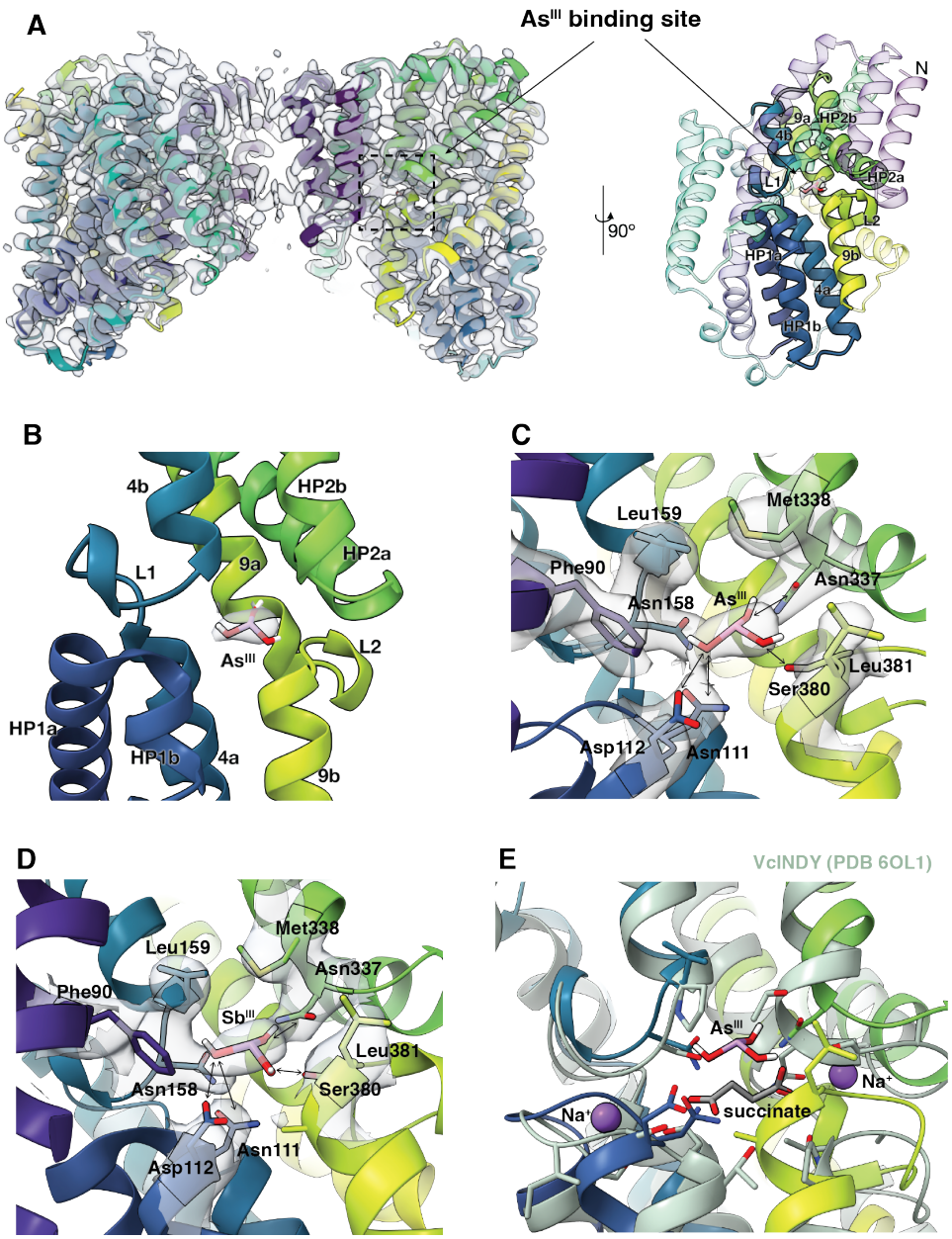


Figure 4.10: **Metalloid recognition site of *Lf*ArsB.** (A) Sharpened 3.3 Å cryo-EM map of *Lf*ArsB in presence of As^{III}, docked with the parallel dimer colored in viridis. Model rotated 90º shows TMDs enclosing the As^{III} binding site. (B) Close-up view of the secondary-structure motifs in the protein interior, including the helix-turn-helix motifs HP1 and HP2, and the helix-loop-helix motifs, L1 and L2, that enclose the As^{III} binding site. (C) Metalloid binding site of *Lf*ArsB, showing density for $As(OH)_3$ and the surrounding residues. Double-headed arrows indicate potential hydrogen-bonding interactions with residues within 3-4 Åof the metalloid density. (D) Metalloid binding site of *Lf*ArsB structure solved with Sb^{III}, showing density for $Sb(OH)_3$ and the surrounding residues. Double-headed arrows indicate potential hydrogen-bonding interactions with residues within 3-4 Åof the metalloid density. (E) Overlay of *Lf*ArsB•As^{III} structure with VcINDY bound to Na^+ and succinate, showing overlapping substrate binding sites and differences in the Na^+ binding cleft.

uncharged nature $As(OH)_3$. Similarly, evidence supports that transport is driven by the proton-gradient across the membrane instead of a metal-ion gradient [4]. The sequence motifs implicated in substrate and Na^+ binding in DASS transporters are not conserved in ArsB (Fig. S4.10 and S4.11). These motifs, located on the HP1-L1 and HP2-L2 regions, are replaced by conserved polar residues that are located near the As^{III} species bound to ArsB.

4.2.6 The dimer interface of ArsB structures

Structures of ArsB revealed two distinct dimer interfaces (Fig. 4.8). In the parallel dimer, the two subunits are separated by ~20 Å and have a buried surface area of merely 40 $Å^2$, making even Van der Waal interactions unlikely to stabilize a dimer interface (Fig. 4.11A). The interface is composed of hydrophobic residues of the scaffold domain in the center and arginines near the periphery of the micelle (Fig. 4.11B). A similar dimer has been observed previously for the TRAP transporter, SiaQM where the two subunits are ~16 Å apart, resulting in a small interface (Fig. 4.11C, left). Intriguingly, this dimer was also obtained by purification in LMNG, similar to ArsB in our case. Monomer structures solved in nanodiscs have also been reported for SiaQM, in addition to the dimer [8]. Unlike ArsB, the dimer interface in SiaQM is composed of residues from the Q-subunit at the N-terminus, which is absent in other IT proteins including ArsB. Both ArsB and SiaQM parallel dimers contrast with the *Vc*INDY dimer that has a more extensive interface (Fig. 4.11C, right). Structures of DASS family proteins are predominantly dimeric with the dimer interface composed of scaffold domain residues. Although the relative orientation of the two subunits in the ArsB parallel dimer are similar to that of the *Vc*INDY dimer, the distance between them is too large to form interfacial contacts. We examined sequence conservation of the scaffold domain at the putative dimer interface between ArsB and structurally-characterized dimeric DASS transporters, including *Vc*INDY (Fig. S4.12). The interfacial aromatic residues that are thought to stabilize the *Vc*INDY dimer by stacking interactions are conserved across DASS members including ArsB (Fig. S4.12), but do not form a stacking interaction in the ArsB parallel dimer structure. We observed that the parallel dimer solved from the As^{III} sample formed a marginally different dimer interface than the apo parallel dimer (C-α r.m.s.d. of 4.3 Å), albeit with a similar spacing between the monomers (Fig. S4.13).

Upon examining the Coulomb potential map at the dimer interface of the ArsB parallel dimer, we identified extensive density that did not correspond to the ArsB

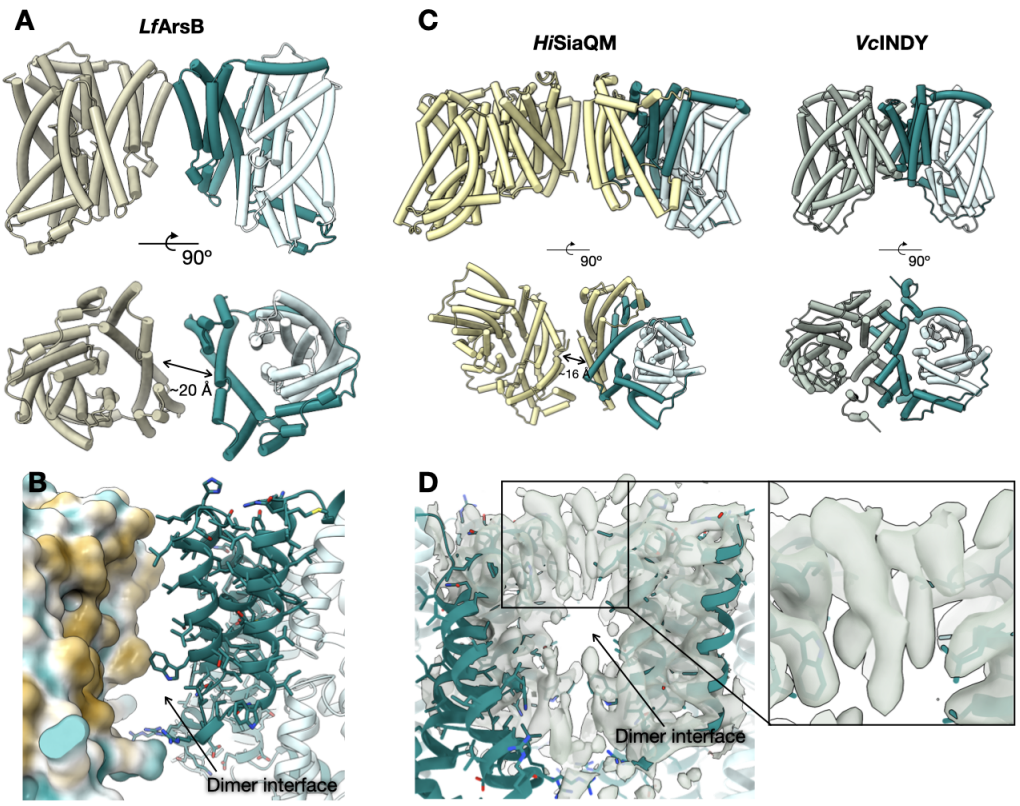


Figure 4.11: **The dimer interface of *Lf* ArsB.** (A) Parallel dimer in two orthogonal orientations highlighting the 20 Åseparation between the scaffold domains of the two subunits. (B) Close-up of the dimer interface showing hydrophobic residues lining the interface and a few polar/charged residues at the periphery. (C) Cryo-EM dimer structures of SiaQM from *H. influenzae* (left) and INDY from *V. cholerae*. (D) Unmodeled lipid density at the dimer interface of *Lf* ArsB, resembling hairpin-shaped phospholipids.

protein chains (Fig. 4.11D). Hairpin-like density in this region, likely corresponding to either lipids or detergent was observed. We conducted lipid analysis to detect the presence of native *E. coli* lipids at this site by extraction and liquid chromatography-mass spectrometry (LC-MS) on detergent-purified ArsB (P1 from SEC trace in Fig. 4.7B). Phosphotidylglycerol and phosophotidylethanolamine (PE) derivatives of acyl chain-lengths ranging from C14 to C18, including saturated and monoun-saturated chains, were detected (Table S4.3). This experiment indicates that *E. coli* lipids may occupy the interface between the two ArsB protomers.

The ArsB antiparallel dimer forms a more extensive dimer interface composed of Van der Waal interaction between the hydrophobic residues of the scaffold domain with a buried surface area of ~812 $Å^2$(Fig. S4.14A). Density for putative interfacial

lipids or detergent is observed in the antiparallel dimer as well (Fig. S4.14B). An antiparallel dimer has also been observed for SiaQM, which was also solubilized in LMNG-containing buffer, and from the same dataset as the parallel SiaQM dimer. In either case, the antiparallel (inverted) membrane topology is unlikely to be biologically relevant. For ArsB to exist in a reverse topology in the bilayer where the termini are on the cytoplasmic side, the GFP tag used in our expression construct must translocate to the periplasm due to the intervening GpA domain. This is unlikely to occur due to the absence of a signal peptide in our construct. Similarity between the parallel and antiparallel dimers of two distinct IT proteins purified in the same detergent indicates that dimer formation is likely an artifact of detergent solubilization. Notably, reconstitution of SiaQM into lipid nanodiscs resulted in both dimers and monomers [8, 16]. It is plausible that the two proteins may exist as weak dimers in the membrane, possibly resembling the parallel dimer state stabilized by interfacial lipids. When extracted from the membranes, the protein enters a monomer-dimer equilibrium in micelles with the thinner scaffold domain facilitating the formation of a dimer interface in the micelle. Both monomeric and dimeric SiaQM are active for transport [16]. While DASS transporters are predominantly dimeric in membrane extracted structures, the subunits function in a non-cooperative manner [31]. Together, these observations suggest that a monomer-dimer equilibrium is common in the IT superfamily. Functionally, whether ArsB favors a monomer or a dimer requires further study.

4.2.7 Proton-coupled As^{III} transport mechanism of ArsB

In light of the 'elevator' mechanism, both apo and As^{III}-bound structures of ArsB represent the 'inward-open' conformation that is accessible to the metalloid substrates, As^{III} and Sb^{III}, from the cytoplasmic side. The translocation steps following metalloid binding and the corresponding conformational changes remain to be understood. Based on the elevator model, the As^{III}-bound transport domain must translate along the membrane, transitioning from 'inward-open' to 'outward-open' conformation, thus translocating As^{III} across the membrane. The 'outward-open' conformation remains to be characterized for ArsB. The free energy for the conformational change from inward (elevator-down) to outward (elevator-up) state would be supplied by the proton gradient across the membrane, in line with the bioenergetics of ArsB [2, 4]. The molecular basis of this proton-coupling is critical to understand the proton-arsenite antiport mechanism of ArsB. Recently, an ortholog of INDY protein from *Drosophila melanogaster* (*Ds*INDY) has also been proposed

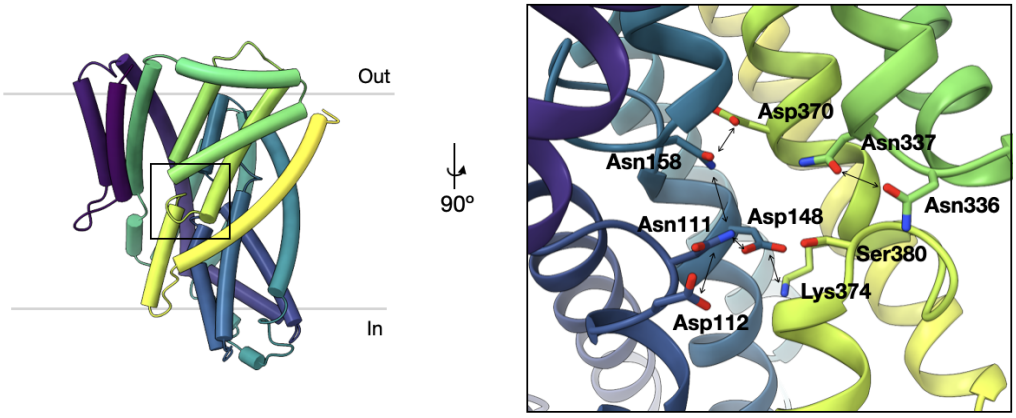


Figure 4.12: **Conserved polar and general acid/base residues in the metalloid binding pocket of *Lf* ArsB indicating a potential hydrogen-bonding network.**

to function by a H^+-coupled transport of citrate, as opposed to a Na^+-coupling mechanism [20]. However, the proton shuttling events that drive the transport are unclear in this system as well.

We inspected our ArsB structures for conserved general acid/base residues such as aspartate, glutamate, and histidine, the corresponding sidechain pK_as of which can be modulated by their environment to facilitate protonation/deprotonation at physiological pH. We identified a network of such residues in the interior of ArsB that may facilitate influx of H^+ to drive the pump. The protonation state of these residues could then be leveraged to stabilize an alternative conformation of ArsB. Such a strategy has been adopted to trap alternative conformations for other H^+-gradient dependent transporters such as *Ds*INDY and the *S. aureus* drug efflux protein, NorA [20, 32]. In the core of the transport domain surrounding the As^{III} binding site are three highly conserved aspartates — Asp112 (HP1), Asp148 (TM4a) and Asp370 (TM9a) (Fig. 4.12 and S4.11). Asp112 and Asp370 flank Asn111 and Asn158 within 3-4 Å, forming a potential hydrogen-bonding network (Fig. 4.12). Additionally, Asp148, which is located at a hydrogen-bonding distance from Asn111, forms an ion-pair with Lys374. We hypothesize that a network of proton transfers orchestrated by these residues may play a key role in facilitating H^+-coupled metalloid transfer and associated conformational changes in the transport cycle of ArsB.

4.2.8 Interaction of ArsB with ArsA ATPase

In addition to operating as an H^+-coupled secondary transporter, ArsB can also function as a primary active transporter by associating with ArsA to couple ATP hydrolysis with the arsenite efflux (Refer to Chapter 3 for the nucleotide hydrolysis mechanism of ArsA). Growth assays in *E. coli* indicate that ArsB removes arsenite more efficiently in the presence of ArsA than H^+coupled efflux [2]. As such, the association with ArsA is beneficial, but not obligatory for ArsB to transport arsenite.

To understand the interaction of ArsA and ArsB, we first looked at the bioinformatic analysis of arsB genes across species from section 4.2.1. While only 22% of arsBs were found to be associated with arsA, no sequence-level variations were apparent between the two subsets of arsBs, *i.e.*, those that have a corresponding arsA gene in the same *ars* operon and those that do not, leaving the nature of the ArsA-ArsB interaction elusive. This observation is supported by conservation analysis [33], which revealed that conservation near the cytoplasmic side of ArsB is similar between the two subsets (Fig. S4.15). Two notable exceptions to this trend are Glu77 and Arg199, located on the scaffold domain, which emerge as highly conserved for ArsBs with corresponding ArsAs. Previously, arsenite resistance growth assays in the presence of As^{III} have shown that an ATPase inactive ArsA expressed with ArsB has reduced As^{III} resistance relative to ArsB alone [2]. This suggests that a catalytically-dead ArsA must physically interact with ArsB to block ArsB export. *In vitro* or direct evidence for an interaction has not been observed. As we could purify *Lf*ArsB in detergent micelles, we sought to test the effect of *Lf*ArsB on the steady-state ATPase activity of *Lf*ArsA. At saturating concentrations of As^{III}, we observed an enhancement in the ATPase activity of ArsA when ArsB was present in the reaction mixture. This effect was concentration-dependent and at a 2:1 molar ratio, ArsB resulted in ~2-fold enhancement of the ATPase activity at saturating concentrations of MgATP and As^{III} (Fig. 4.13). Both P1 and P2 samples of ArsB showed similar rate enhancement, suggesting that ArsB, independent of its oligomeric state, undergoes a functional interaction with ArsA by stimulating its ATPase activity.

We turned to machine learning-based protein structure prediction models to compute a complex between ArsA and ArsB. We compared *Ec*ArsAB complex predictions using various models such as AlphaFold2 Multimer, AlphaFold3, Chai-1 and Boltz-1. The predictions of ArsA and ArsB proteins individually were accurate, but the complex interface for each of the 4 predictions was different (Fig. 4.14). While

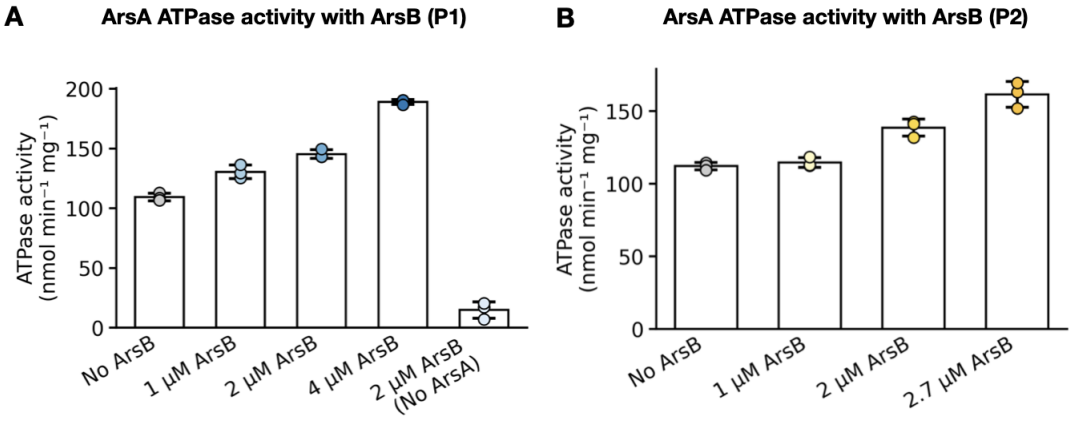


Figure 4.13: **Activation of steady-state ATPase activity of ArsA by detergent-solubilized ArsB.** (A) ATPase activity of *Lf* ArsA in presence of 5 mM $MgCl_2$, ATP and 2 mM As^{III}, at varying concentrations of *Lf* ArsB P1 fraction. (B) ATPase activity of *Lf* ArsA in presence of 5 mM $MgCl_2$, ATP and 2 mM As^{III}, at varying concentrations of *Lf* ArsB P2 fraction. Assays were performed in 50 mM HEPES pH 7.5 containing 0.005%/0.0005% LMNG/CHS. Data are plotted as mean of triplicates with error bars representing standard deviation.

the per-residue confidence metric (pLDDT) was high for each prediction, the ipTM score that is indicative of interface prediction confidence was consistently low across models (Fig. S4.16). Low ipTM scores suggest that either the predicted interfaces are spurious or a low-affinity complex with a small interface formed between the two proteins. This observation corroborates the fact that low sequence-level differences exist between the two subsets of ArsB. However, in each prediction, ArsA was consistently docked near the cytoplasm-facing side of ArsB, consistent with a complex suitable for arsenite transfer (Fig. 4.14). Particularly in AlphaFold2 Multimer and Boltz-1 predictions, the metalloid binding site of ArsA forms an interface, albeit weak, with ArsB in its 'inward-open' conformation (Fig. 4.14A and D), suggesting a physically reasonable mechanism for arsenite transfer from ArsA to ArsB. The ArsA prediction from Boltz-1 is distinct from other models as it is the only ArsA conformation of the 4 predictions that closely resembles the ATP- and As^{III}-bound 'catalytically-competent closed' conformation of ArsA determined by cryo-EM (Refer to Chapter 3) (Fig. 4.14, right). Other predictions are more consistent with an ATPase-inactive 'open' conformation. In the Boltz-1 prediction, contacts are seen between both domains of ArsA and the scaffold domain of ArsB. Together with our observation that ArsB stimulates ArsA ATPase activity,

the Boltz-1 prediction supports an interaction of the 'inward-open' state of ArsB with the 'closed' state of ArsA.

4.2.9 Arsenite resistance growth complementation assays

To support our cryo-EM structural analysis of ArsB and functionally map the interaction between ArsA and ArsB, we performed growth assays in *E. coli* to measure resistance to arsenite. Mutants of *Ec*ArsA and *Ec*ArsB prepared by site-directed mutagenesis were screened for growth in the presence of toxic arsenite concentrations. We leveraged this assay to investigate functional motifs critical for the transport function of ArsB, and to map the interaction interface between ArsA and ArsB.

These growth assays were based on the methods reported in the works of Rosen and coworkers [34]. These assays were typically conducted in *E. coli* strain JM109 to measure arsenite resistance phenotype from complementation of plasmids bearing *E. coli* R773 arsAB genes or their mutants. As an initial screen, we measured growths in this strain under varying As^{III} concentrations using pETDuet, pACYCDuet and pRSFDuet vectors with *Ec*ArsA cloned into MCS-1 and *Ec*ArsB cloned into MCS-2. Consistent with the copy number of these plasmid vectors, we could measure rescue of growth only with the pRSFDuet vector that has the highest copy number out of the three vectors (Fig. S4.17A), and hence was chosen as the candidate for further screening. We identified that the JM109 strain which still bears its chromosomal ars operon demonstrated a small background arsenite resistance, so we decided to switch to the *ars* knockout strain - *E. coli* AW3110 (Δ *ars::cam*) for complementation of mutants. This was especially critical for studying ArsA and ArsB interaction to avoid any possible interference from the background expression of chromosomal ArsB. Growths of pRSFDuet-ArsAB plasmid in AW3110 strain were significantly improved over the background growth from the empty vector which essentially did not support growth at any As^{III} concentration (Fig. S4.17B). Furthermore, we noticed that cells with ArsB alone on a pRSFDuet backbone, showed similar growth profile to ArsAB even at As^{III} concentrations above 3 mM (Fig. S4.17B). We decided to address this by first screening alternative promoter systems. We replaced the T7 promoter upstream of the MCS-2 in the pRSFDuet vector with lacUV5 promoter (pLac) and deleted the second T7 promoter upstream of MCS-2. Alternatively, we cloned *Ec*ArsAB into a pTrc99a plasmid bearing an inducible trc promoter. Unfortunately, the pLac construct did not seem to consistently support growth in As^{III} with ArsB, whereas the pTrc99a construct did not support growth

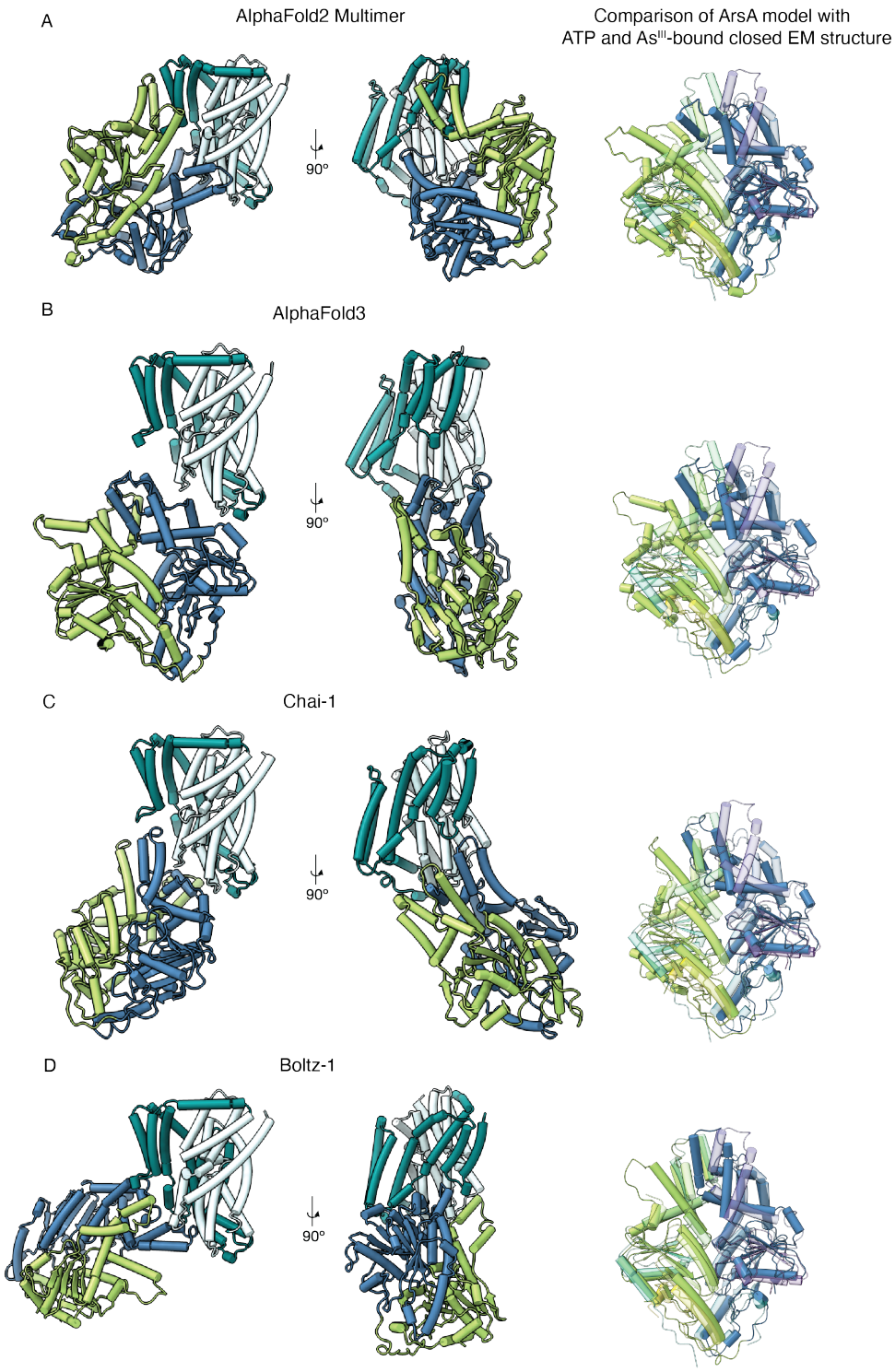


Figure 4.14: **Computational modeling of *Ec*ArsAB complex using (A) Alphafold2 Multimer, (B) AlphaFold3, (C) Chai-1, and (D) Boltz-1.** Both ArsA and ArsB are colored by domains. ArsB is colored in teal (scaffold) and cyan (transport), and ArsA is colored in blue (N-domain) and green (C-domain). On the right, superposition of the ArsA predictions on *ATP-closed* EM structure (viridis and 50% opacity).

in As^{III} at all with either ArsAB or ArsB (Fig. S4.17C). Eventually, deletion of the T7 promoter along with the lac operator upstream of MCS-2 from the original pRSFDuet construct was found to be sufficient to support A^{III} growth with both ArsAB and ArsB. ArsAB rescued growth better than ArsB alone at As concentrations above 3 mM (Fig. 4.15A). Importantly, these growth levels provide sufficient dynamic range to screen both ArsA and ArsB variants. Therefore, the pRSFDuet backbone containing the ArsAB or ArsB insert with deletion of the second T7 promoter and lac operator, named 'pRSFdelT72', was used for subsequent As^{III} growth complementation assays.

We first screened for growth of the As^{III} binding site variants of *Ec*ArsB. Based on the As^{III}-bound structure, we prepared point mutants for highly conserved residues located within a 5 Å radius of the modeled As^{III} molecule. In the presence of 3 mM As^{III}, AW3110 cells complemented with alanine variants of Asn111, Asp112, Asn337 and Met338 were far less resistant to As^{III} than *Ec*ArsB (Fig. 4.15B). Based on the cryo-EM map, except for Met338, all other residues appear to make direct contacts with As^{III}. To our surprise, mutation of Asn158 that is located within hydrogen-bonding distance from As^{III} in the structure, did not affect As^{III} resistance at all. While the mutagenesis and growth screening largely supports our As^{III}-bound cryo-EM structure, further screening of more variants will enable characterization of As^{III} interaction with ArsB. Furthermore, asparagine mutation of a conserved aspartate (D148N) located ~5 Å from As^{III} resulted in decreased As^{III} resistance as well, implying a functional role of Asp148 in As^{III} transport. Both Asp112 and Asp148 are proposed to be involved in a H^+-coupling mechanism in ArsB (Section 4.2.7).

In addition to screening for functionally critical ArsB residues, this assay will be leveraged in follow-up work to screen for residues of ArsA and ArsB at the putative interface based on the predictions reported in Section 4.2.8, enabling us to rationally identify interacting residues between ArsA and ArsB.

4.3 Discussion

ArsB is a key component of the ars operon as it carries out the membrane efflux of toxic arsenite from the cell. Until now, a structural understanding of the metalloid transporter was lacking, precluding elucidation of its transport mechanism. Unlike most transporters of the IT superfamily, ArsB is indicated to function both as a secondary ion-coupled active transporter similar to DASS transporters, and as an

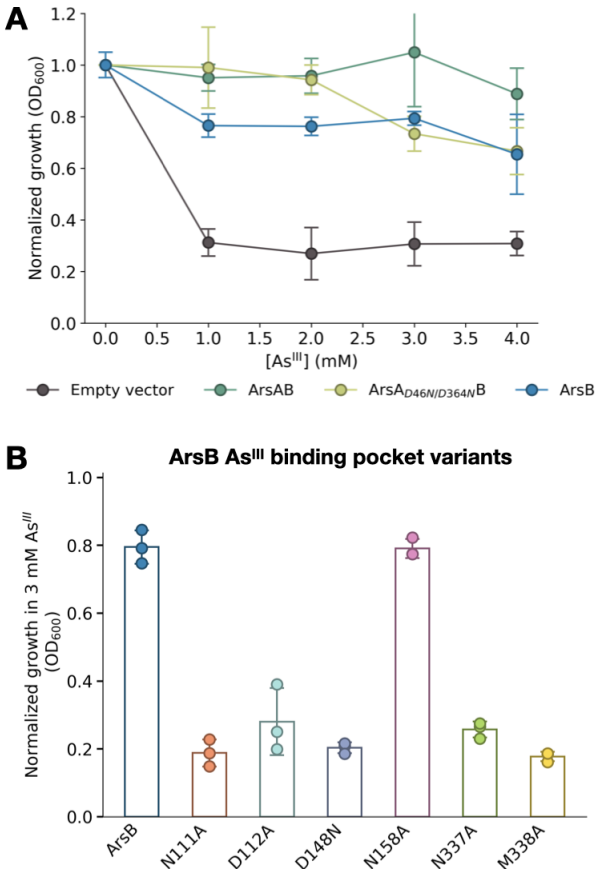


Figure 4.15: **Resistance to As^{III} in *E. coli* strain AW3110 complemented with *Ec*ArsB, *Ec*ArsAB or their variants.** (A) Growth in the presence of As^{III}, measured for empty pRSFdelT72 vector, *Ec*ArsB, *Ec*ArsAB and *Ec*$ArsA_{D46N/D364N}B$, measured after 5 hours at 37ºC. (B) Growth of As^{III} binding pocket residues of *Ec*ArsB in the presence of 3 mM As^{III}, measured after 10 hours at 37ºC. Data are plotted as mean of biological triplicates with error bars representing standard deviation.

ATP-coupled active transporter in the presence of ArsA. The underlying mechanisms of these two individual modes of transport and how ArsB selects for one mechanism over the other are some questions of interest. The work reported in this chapter forms the foundation for further mechanistic elucidation of metalloid transport via ArsB.

We characterized the structures of ArsB in an 'inward-open' conformation in detergent micelles, without the use of any fiducial markers. These structures reveal both parallel and antiparallel dimeric architectures. While a parallel dimer with a weak interface stabilized by intervening phospholipids could exist in the membrane, the stability and functional sufficiency of a monomer suggests that the dimers are

perhaps artefactual. This conformation of ArsB is also compatible with metalloid binding as demonstrated by our metalloid-bound structures. Although the resolution of these structures prevents us from determining the exact speciation of As^{III} or Sb^{III} present in the protein interior, chemical intuition and hydrogen-bonding residues surrounding the binding pocket indicate the presence of a neutral trihydroxylated species, $As(OH)_3$ or $Sb(OH)_3$. A non-covalent binding mode of the metalloid for ArsB, unlike other cysteine-rich proteins such as ArsA that can form As-S covalent bonds, seems mechanistically reasonable to ensure that the metalloid is effectively translocated across the membrane.

Based on the secondary active transport mechanism, ArsB functions as an 'antiporter' by coupling uphill metalloid efflux to the downhill influx of H^+dissipating the PMF across the membrane. As the nature of transport is 'electrogenic', uptake of H^+ from the periplasm must drive the transport cycle. This raises the question whether the metalloid and H^+, in turn, both occupy the same sites on the protein in the transport cycle, a common phenomenon across antiporters, or an allosteric site protonation site exists within the TMDs [17]. A hydrogen-bonding network identified adjacent to the metalloid binding site in our 'inward-open' structures could facilitate a potential H^+-shuttling cascade to drive transport. We hypothesize that modulation of the protonation states of some of the conserved aspartates at this site, including Asp112, Asp148 and Asp370, triggers conformational rearrangements through the transport cycle of ArsB.

Alternatively, ArsB can associate with ArsA to drive metalloid transport by coupling to ATP hydrolysis. In fact, transport via this mode results in As^{III} removal more efficiently from cells (Fig. 4.15A). Our computational modeling results suggest that the two proteins of efflux pump, perhaps do not associate to form a stable complex with a large interface. A weak and/or a transient interaction seems to be more plausible. While this explains how interaction between the two proteins would facilitate metalloid transfer to ArsB, it remains unclear how a weak interaction would trigger rearrangements of the transport domain for efflux. The arsenite resistance growth complementation assays would allow us to identify candidate residues at the putative interface of ArsAB complex that form the structural basis for primary active transport via ArsB.

4.4 Methods

4.4.1 Sequence Similarity Network Analysis

Initial sequence similarity network of ArsB was constructed from a BLAST search of *Ec*ArsB sequence quering a maximum of 6000 sequences with a default e-value of 5, using the EFI-Enzyme Similarity Tool (https://efi.igb.illinois.edu/efi-est/). The SSN was constructed using a sequence length filter of 400-500 and alignment score of 160. Colored SSN was generated using the Genome Neighborhood Network (GNN) tool to incorporate a window of 5 neighbors both upstream and downstream of ArsB in the analysis. The resulting colored SSN, with representative nodes of 90% sequence identity, was visualized using Cytoscape in 'organic layout'. Non-redundant ArsB sequences found in association with ArsR and/or ArsA were extracted from an exported csv file that contained the neighborhood pfam families in a separate column. Sequences of interest were extracted and compiled in FASTA format using a Python script, written with assistance from ChatGPT (OpenAI).

4.4.2 Cloning and construct optimization of ArsB

*Ec*ArsB gene with C-terminal or N-terminal 6x-His tags was obtained from GenScript pre-cloned into the MCS-2 of pETDuet-1 backbone. *Lf* ArsB gene with C-terminal 6x-His tag was obtained as a gene fragment from Twist Bioscience (South San Francisco, CA) and cloned into the MCS-2 of pETDuet-1 backbone. All DNA cloning was performed using the HiFi-DNA assembly protocol (NEB).

Expression of ArsB was successfully achieved using a GFP-fusion protein-based strategy reported by Hsieh and coworkers [25]. Based on Orientation of Proteins in Membranes (OPM) server for membrane protein topology prediction [26], ArsB is predicted to have both N and C termini localized to the periplasm. To construct a GFP-fusion ArsB, a single-pass transmembrane helix, Glycophorin A (GpA) followed by GFP was fused to the 3'-end of ArsB gene. An 8xHis tag was added on the 3'-end of the GFP sequence, separated by a 10xGS linker. An HRV-3C protease cleavage site flanked by a 7xGS linker on the 5'-end and a 2xGS linker on the 3'-end, was added between ArsB and GpA sequences to produce a cleavable construct. Alternatively, constructs with TEV and SUMO cleavage sites, and N-terminal His tag were tested for expression, but none of these resulted in sufficient protein yields for purification and downstream studies.

4.4.3 Expression and purification of ArsB GFP-fusion construct

pETDuet-1 *Lf* ArsB-7xGS-3C-2xGS-GpA-GFP-10xGS-8xHis was transformed into *E. coli* Nico21 (DE3) competent cells (NEB). Colonies from overnight grown LB agar plate were inoculated into a starter culture containing ampicillin and grown at 37ºC for 2 hours before inoculating into 1 L 2xYT media. Cells were induced with 0.4 mM IPTG at OD 0.6-0.8 and grown at 18ºC for 16-20 hours. Successful expression was indicated by a cell mass of 4 g per liter of media and a pale-yellowish appearance of the pellet. Cells were harvested by spinning in a Beckman Coulter centrifuge at 6000 rpm at 4ºC for 10 min and lysed using 3 passes through a microfluidizer. The lysate was clarified by centrifugation at 12,500 rpm at 4ºC for 30 min, followed by isolation of membrane fraction by centrifugation at 38,000 rpm at 4ºC for 40 min using an Optima XE-90 Ultracentrifuge (Beckman Coulter). The resulting membrane pellet was resuspended in extraction buffer (50 mM HEPES pH 7.5, 300 mM NaCl, 5% glycerol, 1%/0.1% LMNG/CHS, 10 mM imidazole, 5 mM β-ME) and incubated at 4ºC for 1 hour. The supernatant after an ultracentrifuge run was incubated with 1-mL Ni-NTA resin (Qiagen) pre-equilibrated with extraction buffer at 4ºC for 1 hour. Protein-bound resin with a bright green appearance due to GFP, was spun down at 700xg at 4ºC for 5 min, and then washed with 50 column volumes (CV) of wash buffer (50 mM HEPES pH 7.5, 300 mM NaCl, 5% glycerol, 0.005%/0.0005% LMNG/HS, 30 mM imidazole, 5 mM β-ME) and 1-CV fractions were collected with elution buffer (wash buffer + 200 mM imidazole). Purity of the fractions was assessed by SDS-PAGE analysis. All protein usually eluted in the first 4 fractions with high purity. These fractions were pooled and concentrated to 3 mL using a 100k MWCO Amicon filter (Milli-pore) and then exchanged into low-salt and imidazole free buffer or the SEC buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 5% glycerol, 0.005%/0.0005% LMNG/CHS and 5 mM β-ME) using a 10-mL desalting column (Bio-Rad) in preparation for protease cleavage. Purified HRV-3C protease (section 4.4.2) was added at 4 mg per mL of Ni-eluant and incubated at 4ºC for 16 h. Fully cleaved ArsB was separated from cleaved-off GpA-GFP-His fusion tag, uncleaved ArsB and His-tagged 3C protease by rebinding to the Ni-NTA resin. Cleaved ArsB sample was applied by 4 passes on Ni-NTA resin pre-equilibrated with SEC buffer. Two additional fractions were collected using SEC buffer supplemented with 30 mM imidazole. After purity of fractions was analyzed by SDS-PAGE, fractions containing cleaved ArsB only were pooled and concentrated to 500 µL using a 100k MWCO Amicon filter (Milli-pore). Sample was applied to a Superdex 200 Increase 10/300 GL column (Cytiva) pre-equilibrated

with filtered and degassed SEC buffer and run at 0.4 mL/min with 250 µL fractions collected. ArsB typically eluted as two distinct peaks at 11-mL and 12-mL; a shoulder usually appeared before the 11-mL peak. Extraction and purification in DDM was performed with 1.5% and 0.03% detergent, respectively, whereas in DM was performed using 1.5% and 0.1% detergent, respectively.

4.4.4 Expression and purification of HRV-3C protease

The plasmid bearing HRV-3C protease with an N'-GST tag was a kind gift from Dr. Vy Nguyen (Voorhees Group, Caltech). A 6x-His tag was engineering at the C'-terminus of the 3C protease sequence to enable purification using Ni^{2+}-affinity chromatography. The resulting plasmid wwas transformed into BL21 Gold(DE3) chemically competent cells (Invitrogen), and grown from an overnight culture in 2xYT media. Cells were induced with IPTG at OD600 of 0.6 at 37ºC shaking at 225 rpm, and further grown for 4 hours at 37ºC before harvesting cells and freezing the pellet at -80ºC. For purification of HRV-3C protease, cell lysis and Ni^{2+} affinity purification were performed as described in section 3.4.1 (Chapter 3). The elution fractions from the Ni^{2+} affinity were pooled and exchanged into ArsA storage buffer with 5 mM TCEP as described in section 3.4.1 (Chapter 3). The protease was saved as 50 uL aliquots and stored at -80ºC until required.

4.4.5 In-gel trypsin digest and mass spectrometry analysis

Proteins bands excised from the Coomassie-stained SDS-PAGE gel. Trypsin digest of the excised bands was performed using the Pierce In-gel Trypsin Digest Kit (Thermo). MS of the digested samples and data analysis was performed by the Protein Exploration Laboratory (PEL) at Caltech.

4.4.6 Cryo-EM sample preparation

Fractions of interest were pooled and concentrated using Amicon 100k concentration filter spun at 7500xg at 4ºC until 4-6 mg/mL of protein was achieved. The molecular weight and extinction coefficient at 280 nm corresponding to cleaved ArsB were 47.2 kDa and 73,900 $M^{-1}cm^{-1}$, respectively. For apo ArsB, 3 µL of the sample supplemented with 0.05% CHAPSO was applied to glow-discharged Quantifoil holey carbon R1.2/1.3 300 Mesh, Copper (Quantifoil, Micro Tools GmbH) grids using a Vitrobot (FEI Vitrobot Mark v4 x2, Mark v3). For ArsB with metalloids, sample was mixed with 2 mM sodium arsenite or potassium antimonyl tartrate before adding CHAPSO. In each case, grids were blotted at 100% humidity and 4ºC

and immediately followed by plunge-freezing into liquid ethane. Structures in the presence of As^{III} and Sb^{III} were solved by incubating ArsB for a few seconds before grid application with 2 mM sodium meta-arsenite and potassium antimonyl tartrate, respectively.

4.4.7 Cryo-EM data acquisition and processing

Cryo-EM data was collected on a 300 kV Titan Krios TEM equipped with a Gatan K3 direct electron detector and Gatan Energy Filter (slit width 20eV) in super-resolution mode using SerialEM (42). Datasets were acquired at 130,000x magnification with a raw pixel size of 0.325 Å/pixel, electron exposure of 70 e-/Å^2 over 40 frames (exposure rate of 1.75 e-/Å^2/frame), and a defocus range of -0.8 to -2.8 µm in CDS mode. All datasets were processed in cryoSPARC v4.7.1 [35] using the same general processing workflow as described here. Data collection and processing details for individual datasets are listed in Table S4.4. During image collection, cryoSPARC LIVE was used for simultaneous motion correction, CTF-estimation and manual curation of the collected movies. Micrographs with a CTF outside the range of 2.5-5 Å were filtered out. Particles were picked using a Topaz model [28] trained on particle-set from the ArsB datasets from DDM detergent that resulted in reasonable reconstructions of the TMDs but could not achieve high resolution. These particles were extracted with a 2x bin (1.3 Å /pixel), and subjected to one round of 2D classification to remove obvious junk particles or apparently empty micelles. Ab-initio reconstruction was performed on this particle set to obtain four 3D volumes. Volumes that showed TMD density were subjected to multiple rounds of ab-initio reconstruction. This curated set of particles was then used to train a new Topaz model and extract particles again from the micrographs, followed by iterative ab-initio reconstructions. Finally, reference-based motion correction, global or local CTF refinement and Non-uniform refinements were performed to result in the final set of particles resulting a high-resolution map. The overall resolution was estimated from the gold-standard Fourier shell correlation (FSC) curve at a cut-off of 0.143 in cryoSPARC.

4.4.8 Model building and refinement

We docked the AlphaFold3 model of LfArsB into the sharpened EM maps in ChimeraX v1.10 to assemble the models for two distinct dimer states. The dimer models were then fitted into the cryo-EM maps using ‘Dock in map’ using Phenix v1.21.2. Docked models were refined to fix backbone geometries and fit sidechain

rotamers to the density using 'Real-space refinement' in Phenix and 'Real-space refine zone' in Coot v0.9.8.7. For the As^{III} and Sb^{III}-bound structures, geometry restraints were generated for trigonal $As(OH)_3$ and $Sb(OH)_3$ using 'elbow' in Phenix. Ligand was docked into the density in Coot and refined with the restraints using Phenix.

4.4.9 Lipid analysis

Analysis of putative phospholipids from *E. coli* in the purified *Lf* ArsB samples was performed by lipid extraction and liquid chromatography-mass spectrometry (LC-MS). The experiment was performed in collaboration with Dr. Virginia K. James and Prof. Arthur D. Laganowsky at the Department of Chemistry, Texas A&M University.

4.4.10 ArsAB complex computational modeling

All complex predictions were made using the sequences of *Ec*ArsA and *Ec*ArsB of the R773 plasmid. AlphaFold2 Multimer prediction was made using ColabFold notebook [36]. AlphaFold3 and Chai-1 predictions were made using the respective online servers [37, 38]. Boltz-1 prediction was made using the ChimeraX 1.10 plugin [39]. Default parameters were used for each prediction.

4.4.11 ArsA ATPase activation assays

Lf ArsA ATPase assays were performed using the NADH-coupled spectrophotometric assay as described in section 3.4.5 (Chapter 3). LMNG/CHS-solubilized *Lf* ArsB (P1 or P2 samples) were added to the reaction mixture at desired concentrations along with other components before the addition of ATP. Assay buffer was supplemented with LMNG/CHS to result in a final concentration of 0.005%/0.0005%. Reaction containing ArsB and no ArsA was run as a control to measure background ATPase activity.

4.4.12 Arsenite resistance growth assays

Growth assays of *E. coli* strain JM109 were performed in 2 mL cultures, whereas all subsequent assays were performed with a 96-well plate microplate reader. For initial assays using strain JM109, *Ec*ArsA and *Ec*ArsB genes both bearing a C-terminal 6x-His tag were cloned into MCS-1 and MCS-2, respectively of each of pETDuet, pRSFDuet and pACYCDuet backbones. These plasmids bearing *Ec*ArsAB were then transformed into JM109 chemically competent cells (Promega) on LB agar plates containing 100 μg/mL carbenicillin or 50 μg/mL kanamycin. Overnight cul-

tured were grown with respective antibiotics from single colonies for ~16 hours. The overnight cultures were diluted 100-fold into 2 mL LB media containing respective antibiotic, 0.1 mM IPTG and sodium arsenite ranging from 0 mM to 5 mM in 1 mM increments. Cells were grown for more than 12 hours at 37ºC shaking at 225 rpm. Cell density at 600 nm was measured manually using a bench-top specotrophotometer in 1-cm path length cuvettes.

E. coli strain AW3110 (*E. coli* strain W3110 Δ*ars::cam*) was obtained as a gift from Prof. Chad Satlikov (University of California Santa Cruz). In preparation for growth assays, AW3110 cells were first made electrocompetent using standard protocols (https://www.embl.org/groups/protein-expression-purification/protocols/preparation-of-electrocompetent-e-coli-cells/). Plasmids (pRSFDuet, pRSFdelT72, pRSFlac or pTrc99a) bearing *Ec*ArsAB or *Ec*ArsB genes, cloned from the pRSFDuet construct, were transformed into AW3110 cells by electroporation. Transformed cells were plated onto LB agar plates containing antibiotic corresponding to the vector and 25 μg/mL chloramphenicol. Overnight cultures were grown from single colonies for ~16-18 hours before inoculation. For growth in 96-well plates, 150 uL of LB media supplemented with 50 μg/mL ampicillin or kanamycin and 25 μg/mL chloramphenicol, 0.1 mM IPTG, and sodium arsenite at 0, 1, 2, 3, ,4 and 5 mM were added to the wells in triplicates. Overnight cultures were added to each well at 100-fold dilution. Cell growth was recorded by measuring cell density at 600 nm in a Tecan Infinite 200 Pro microplate reader every 10 min at 37ºC shaking at 218 rpm (orbital shaking) for at least 10 hours. For the final optimized pRSFdelT72 constructs of *Ec*ArsAB, *Ec*ArsB and their respective variants, growths were recorded using a BioTek 96-well plate reader with same settings as above. Normalized growth data were plotted using a python script written in Jupyter notebook, using assistance from ChatGPT.

4.5 Supplementary information

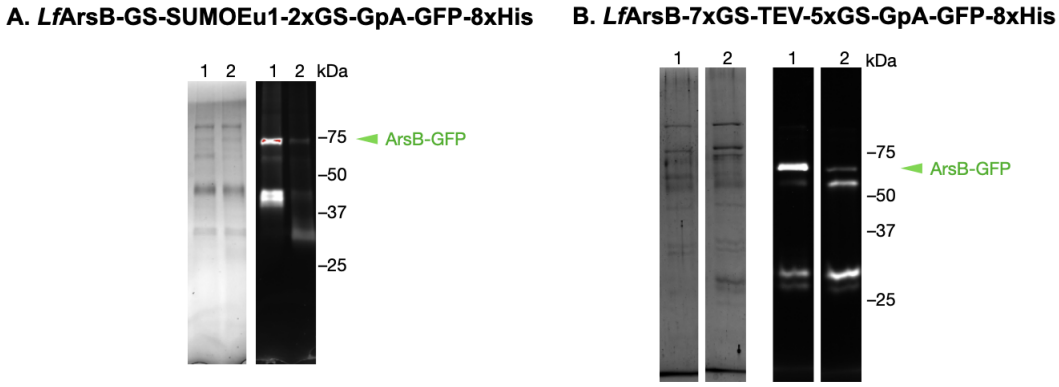


Figure S4.1: **Purification of *Lf* ArsB-GpA-GFP constructs with (A) SUMOEu1 and (B) TEV protease cleavage sites.** SDS-PAGE after Ni-affinity chromatography, showing Coomassie Blue-stained image (left) and Pro-Q emerald 488 nm image (right). Labels 1 and 2 indicate sample before and after cleavage with respective protease.

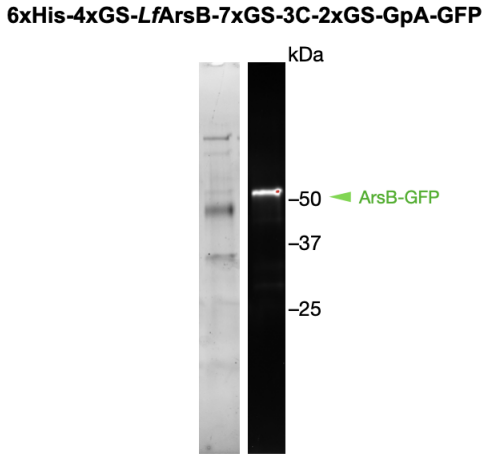


Figure S4.2: **Purification of *Lf* ArsB-GpA-GFP construct with a N-terminal 6xHis tag.**

Table S4.1: **Top 10 protein hits identified in *Ec*ArsB gel fragment by mass spectrometry.**

Protein Name	**Coverage (%)**	**Unique Peptides**	**PSMs**	**MW (kDa)**	**Abundance *Ec*ArsA***	**Abundance *Ec*ArsB**
Arsenical pump-driving ATPase OS=*Escherichia coli* OX=562 GN=arsA PE=1 SV=1	80	47	1132	63.1	8.69×10^{10}	9.19×10^{7}
Maltose/maltodextrin-binding periplasmic protein OS=*Escherichia coli* (strain K12) OX=83333 GN=malE PE=1 SV=1	47	15	32	43.4	6.49×10^{9}	1.52×10^{7}
Keratin, type II cytoskeletal 1 OS=*Homo sapiens* OX=9606 GN=KRT1 PE=1 SV=6	60	46	412	66.0	3.57×10^{9}	5.52×10^{9}
Keratin, type I cytoskeletal 10 OS=*Homo sapiens* OX=9606 GN=KRT10 PE=1 SV=6	48	31	294	58.8	1.84×10^{9}	2.88×10^{9}
Keratin, type I cytoskeletal 9 OS=*Homo sapiens* OX=9606 GN=KRT9 PE=1 SV=6	73	46	432	62.0	1.40×10^{9}	3.81×10^{9}
Keratin, type I cytoskeletal 14 OS=*Homo sapiens* OX=9606 GN=KRT14 PE=1 SV=3	51	24	238	51.6	1.14×10^{9}	2.12×10^{9}
Keratin, type II cytoskeletal 5 OS=*Homo sapiens* OX=9606 GN=KRT5 PE=1 SV=6	46	21	198	62.4	9.12×10^{8}	1.89×10^{9}
Elongation factor Tu OS=*Escherichia coli* (strain K12) OX=83333 GN=tufA PE=1 SV=1	62	28	341	43.3	7.25×10^{8}	2.13×10^{9}
Keratin, type I cytoskeletal 16 OS=*Homo sapiens* OX=9606 GN=KRT16 PE=1 SV=6	40	16	156	51.1	6.21×10^{8}	1.04×10^{9}
Keratin, type I cytoskeletal 17 OS=*Homo sapiens* OX=9606 GN=KRT17 PE=1 SV=5	44	18	170	48.4	5.88×10^{8}	9.61×10^{8}

**Ec*ArsA was used as a positive control.

Table S4.2: **Top 10 protein hits identified in *Lf* ArsB gel fragments by mass spectrometry.**

Protein Name	**Coverage (%)**	**Unique Peptides**	**PSMs**	**MW (kDa)**	**Abundance *Lf* ArsB cleaved**	**Abundance *Lf* ArsB uncleaved**
L. ferriphilum (strain ML-04) ArsB OX=1048260 GN=*Lf* ML04_2457 PE=3 SV=1	29	11	72	45.2	2.30×10^9	1.55×10^7
Ubiquinol oxidase subunit 2 OS=*Escherichia coli* (strain B / BL21-DE3) OX=469008 GN=ECBD_3226 PE=3 SV=1	59	13	138	34.9	5.23×10^8	4.22×10^5
Undecaprenyl-phosphate 4-deoxy-4-formamido-L-arabinose transferase OS=*E. coli* (strain B / BL21-DE3) OX=469008 GN=arnC PE=3 SV=1	59	18	70	36.3	2.41×10^8	0.00×10^0
50S ribosomal protein L17 OS=*E. coli* (strain B / BL21-DE3) OX=469008 GN=rplQ PE=3 SV=1	28	5	10	14.4	4.27×10^7	3.95×10^7
Chaperone protein DnaJ OS=*E. coli* (strain B / BL21-DE3) OX=469008 GN=dnaJ PE=3 SV=1	31	11	23	41.0	3.94×10^7	1.53×10^7
DNA-directed RNA polymerase subunit beta' OS=*E. coli* (strain B / BL21-DE3) OX=469008 GN=rpoC PE=3 SV=1	21	47	133	155.1	2.49×10^7	0.00×10^0
ATP synthase subunit alpha OS=*E. coli* (strain B / BL21-DE3) OX=469008 GN=atpA PE=3 SV=1	43	28	61	55.0	2.34×10^7	0.00×10^0
ATP-dependent RNA helicase DeaD OS=*E. coli* (strain B / BL21-DE3) OX=469008 GN=deaD PE=3 SV=1	28	19	42	69.8	2.12×10^7	0.00×10^0
ATP synthase subunit beta OS=*E. coli* (strain B / BL21-DE3) OX=469008 GN=atpD PE=3 SV=1	48	26	60	50.1	1.93×10^7	0.00×10^0
Elongation factor G OS=*E. coli* (strain B / BL21-DE3) OX=469008 GN=fusA PE=3 SV=1	38	19	47	77.0	1.78×10^7	0.00×10^0

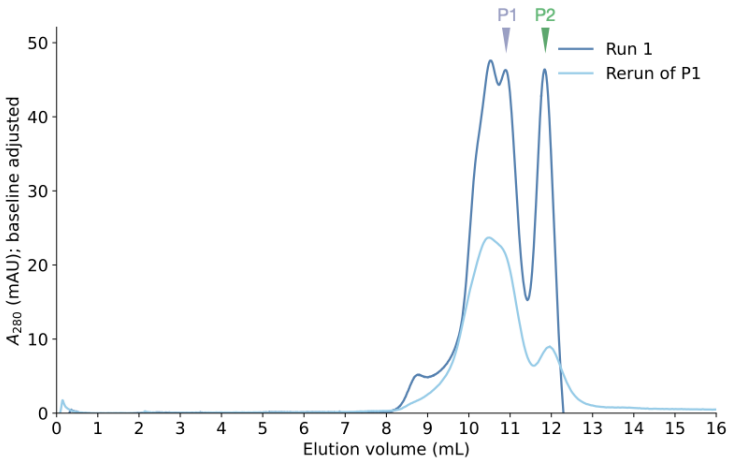


Figure S4.3: **SEC profile of *Lf* ArsB showing that peak 1 (P1) fractions dissociate into peak 2 (P2) during a second SEC run.**

Table S4.3: **Lipidomic analysis of purified *Lf* ArsB P1 fraction.**

Lipid ID	Precursor m/z (theo)	Precursor m/z (exp)	ppm error	retention time
PE(16:0/17:1)	702.5079	702.5112	-4.70	21.993
PE(16:1/14:0)	660.4610	660.4654	-6.66	16.237
PE(16:0/14:0)	662.4766	662.4806	-6.04	19.043
PE(16:1/16:1)	686.4766	686.4799	-4.81	16.351
PE(16:0/16:1)	688.4923	688.4946	-3.34	19.453
PG(16:1/14:0)	691.4556	691.4582	-3.76	14.157
PE(16:1/17:1)	700.4950	700.4957	-1.00	18.326
PE(16:1/18:1)	714.5079	714.5103	-3.36	19.536
PE(16:0/18:1)	716.5236	716.5267	-4.33	23.650
PG(16:0/16:1)	719.4869	719.4901	-4.45	16.595
PE(18:1/17:1)	728.5236	728.5270	-4.67	22.056
PG(16:0/17:1)	733.5025	733.5055	-4.09	18.644
PE(18:1/18:1)	742.5392	742.5426	-4.58	23.618
PG(16:1/18:1)	745.5025	745.5032	-0.94	16.601
PG(16:0/18:1)	747.5182	747.5215	-4.41	19.796
PG(18:1/17:1)	759.5182	759.5205	-3.03	18.696
PG(18:1/18:1)	773.5338	773.5367	-3.75	19.862

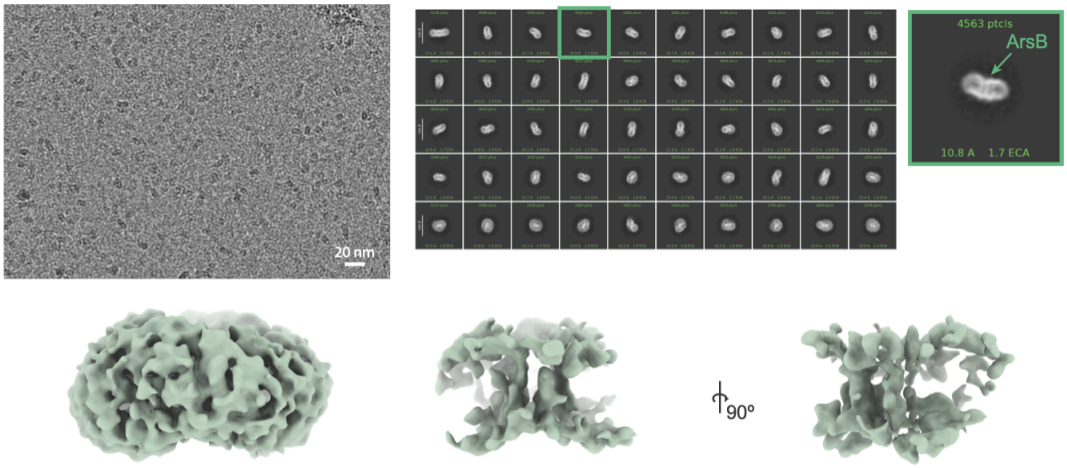


Figure S4.4: **Preliminary cryo-EM analysis of *Lf*ArsB P2 fraction purified in DDM, suggesting that a monomer is predominant.**

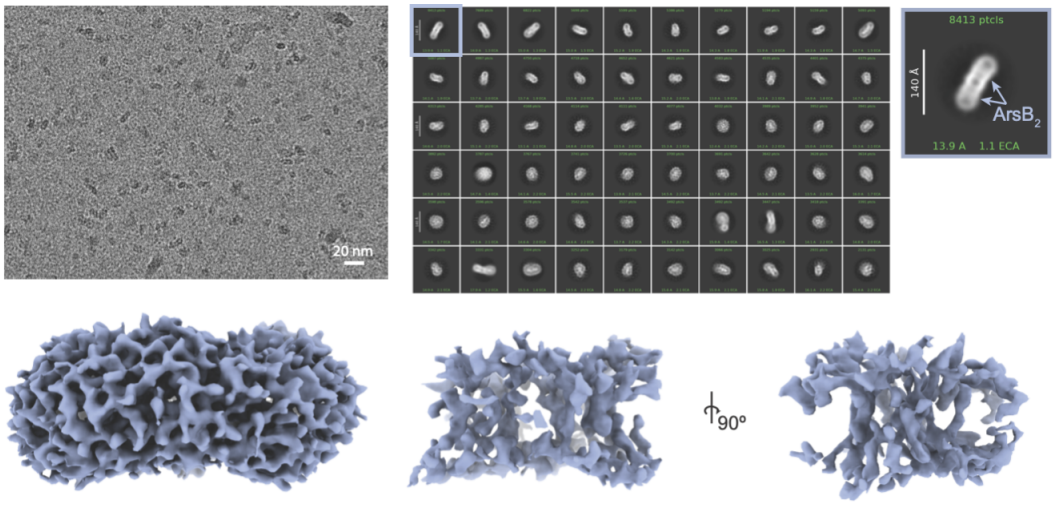


Figure S4.5: **Preliminary cryo-EM analysis of *Lf*ArsB P1 fraction purified in DDM, suggesting that a dimer is predominant.**

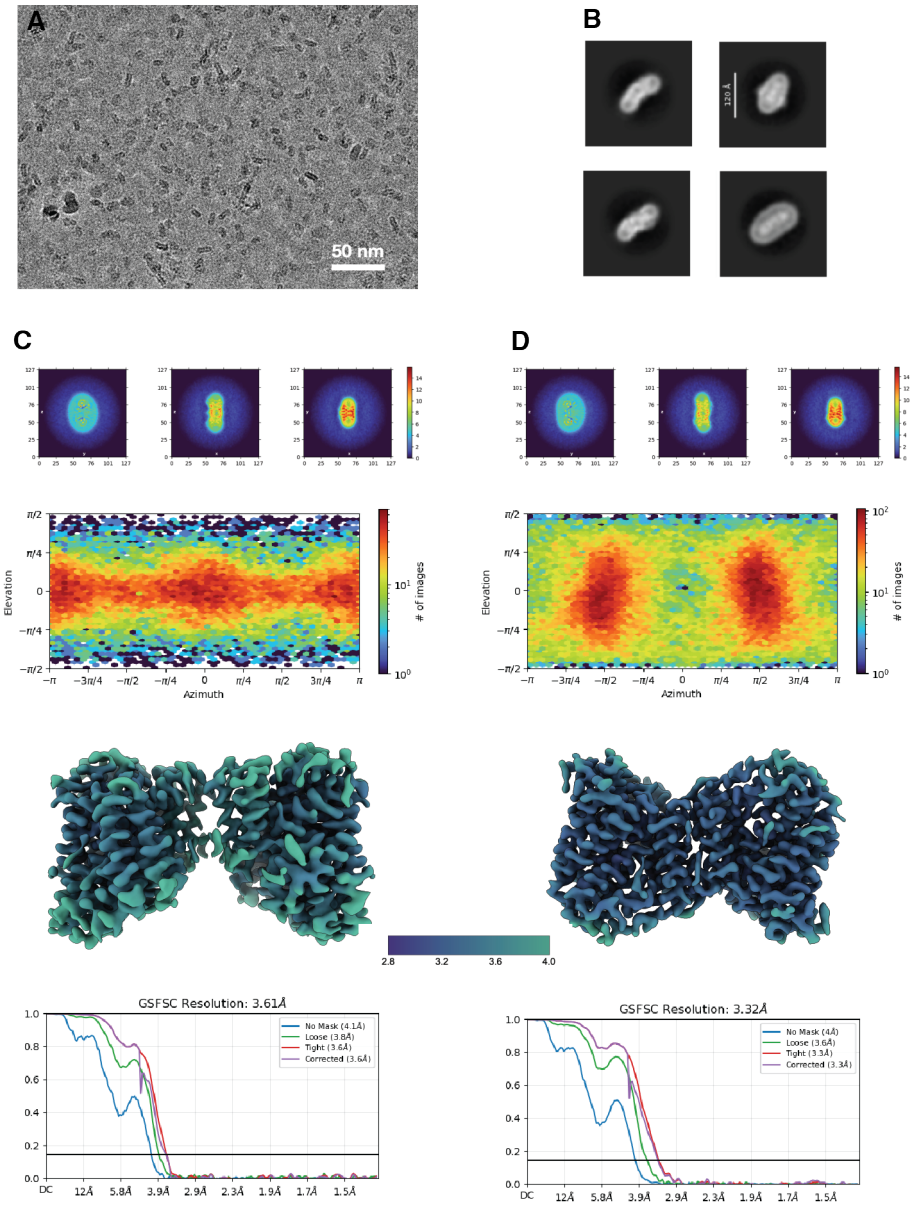


Figure S4.6: **Cryo-EM structural characterization of *Lf*ArsB purified in LMNG/CHS.** (A) Representative micrograph showing distribution of particles at 130,000x magnification. (B) Representtive 2D class averages showing 4 different orientations of the detergent micelle enclosing ArsB. (C-D) 3D reconstruction and orientation distribution of parallel and antiparallel dimers. Maps are colored by local resolution per scale in Å.

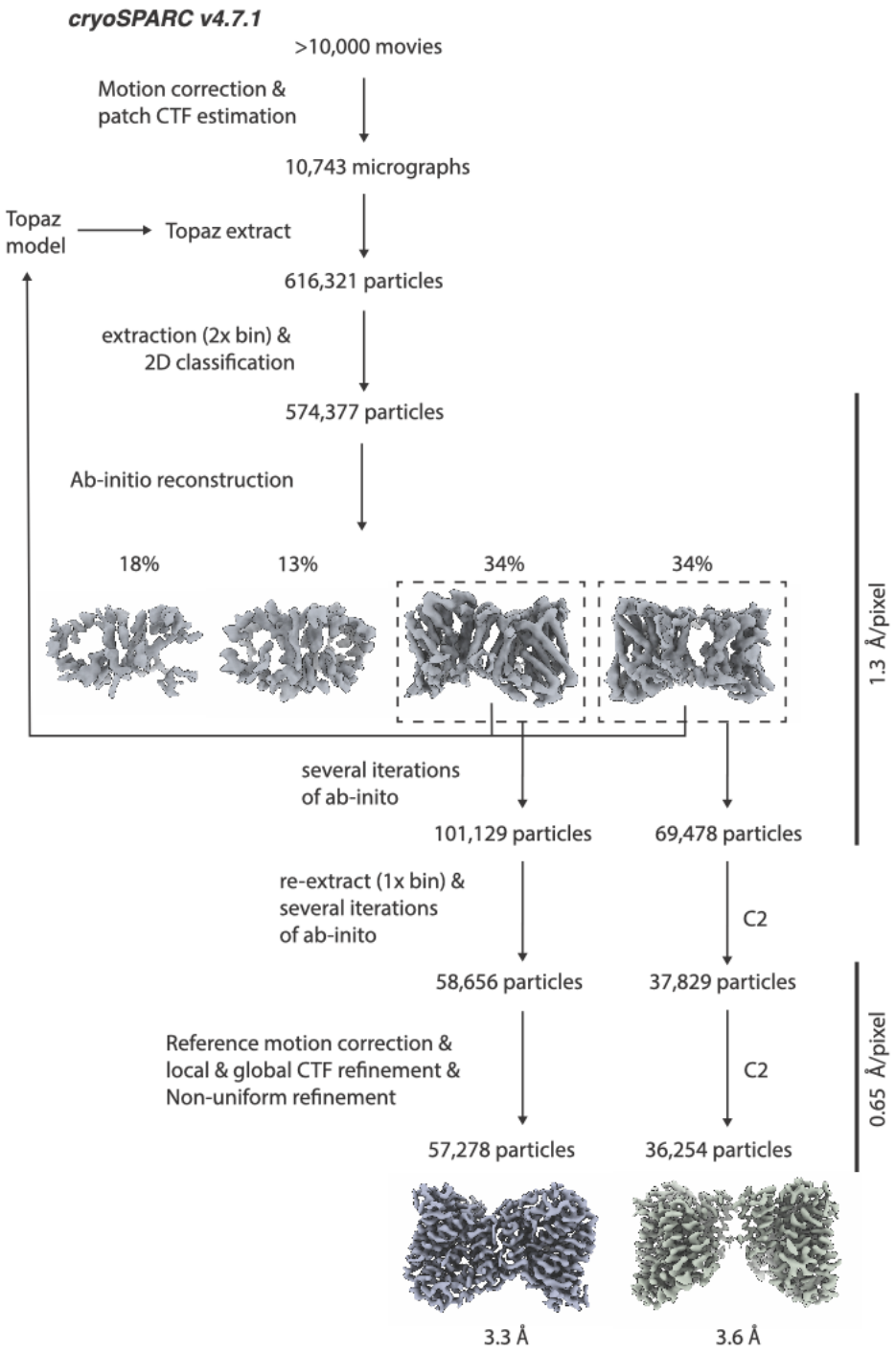


Figure S4.7: **Cryo-EM processing workflow for apo *Lf* ArsB structures.**

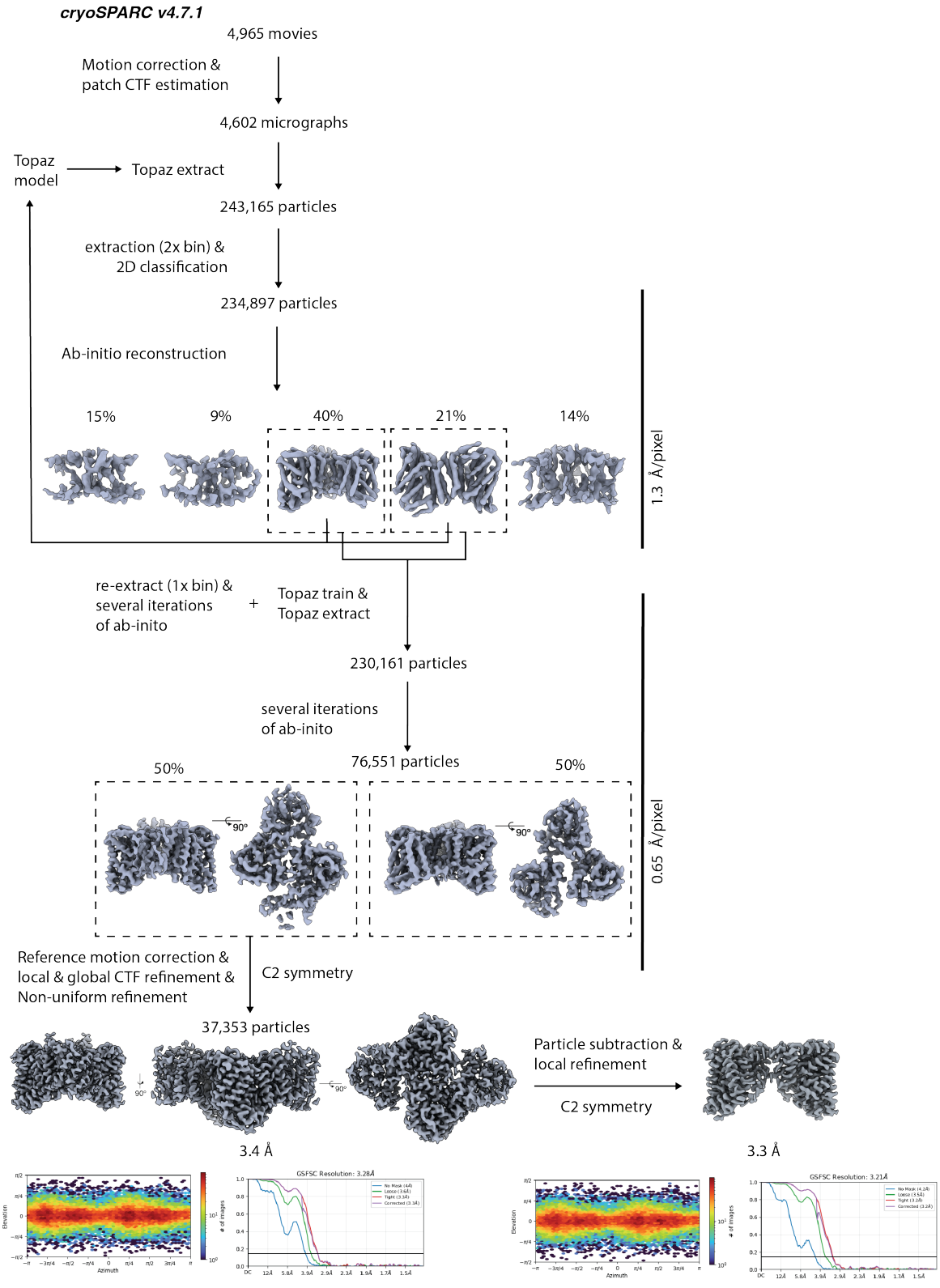


Figure S4.8: **Cryo-EM processing workflow for *Lf* ArsB structure in presence of As^{III}.**

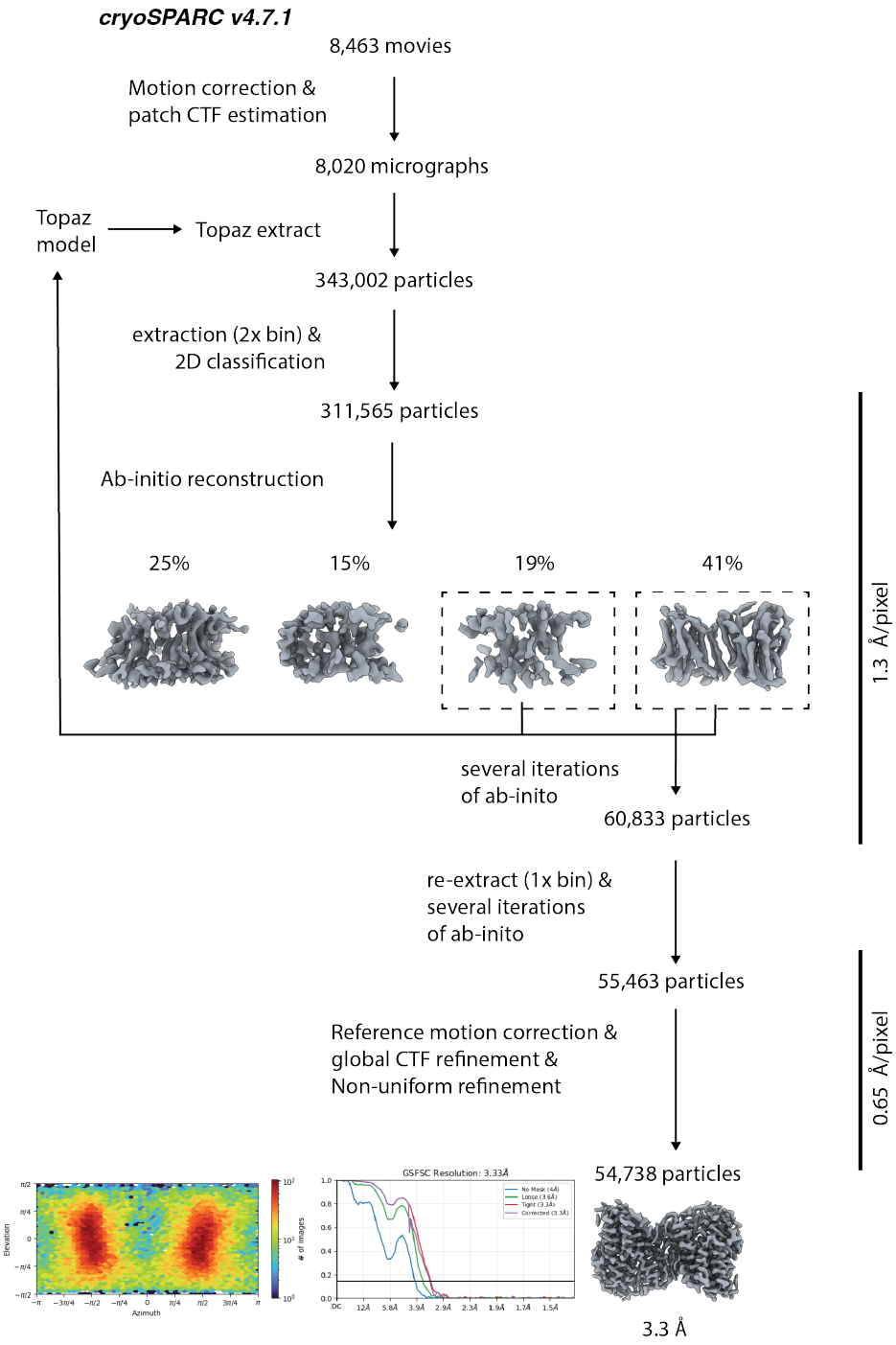


Figure S4.9: **Cryo-EM processing workflow for *Lf* ArsB structure in presence of Sb^{III}.**

Figure S4.10: **Multiple sequence alignment of ArsB and DASS family transporters highlighting substrate and Na^+ binding motifs in the N- and C-halves of the transport domain.** Promals3D [40] was used to create structure-based sequence alignment, utilizing the structure of *Lf* ArsB, and one structure each of the other DASS proteins.

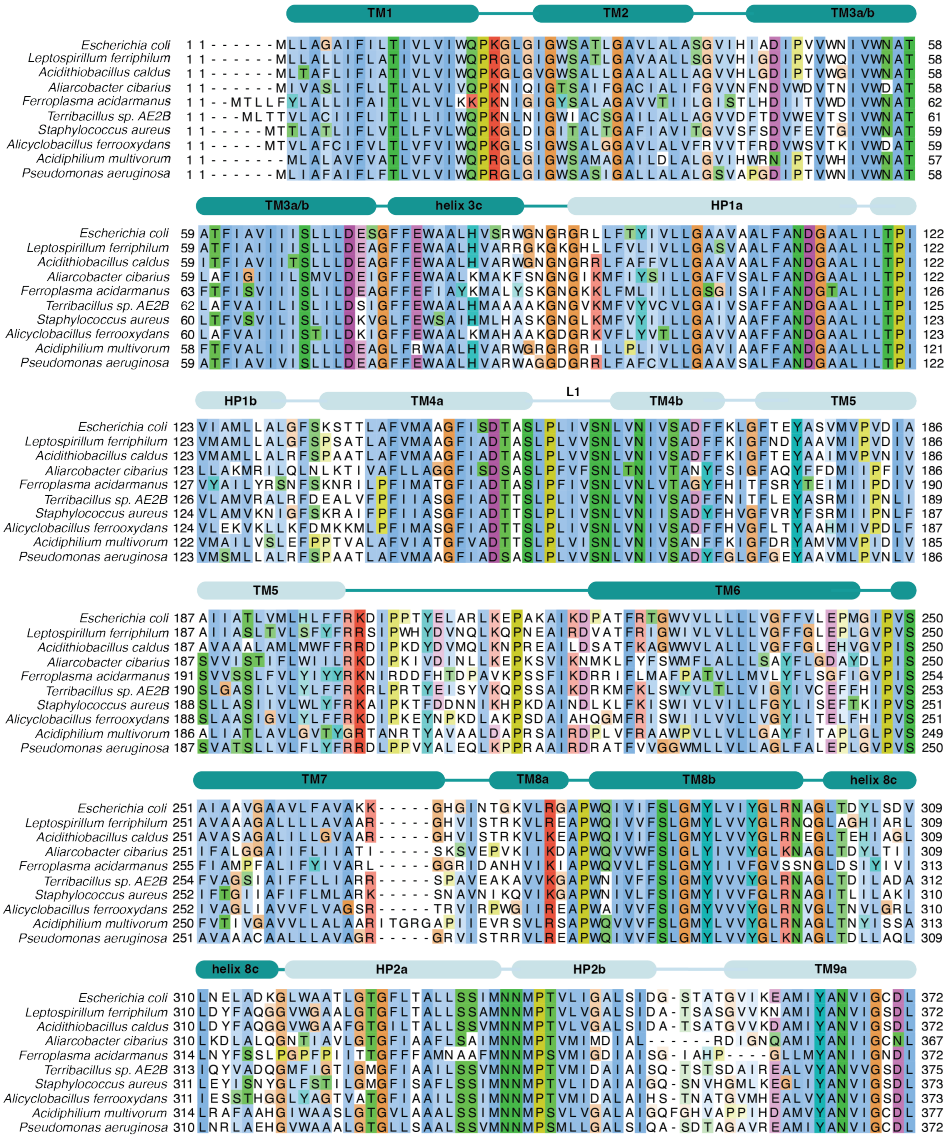


Figure S4.11: **Multiple sequence alignment of ArsB family showing conserved transmembrane domain architecture.** Sequences were randomly curated from ArsR-associated ArsB sequences only, based on the SSN illustrated in Fig. 4.3.

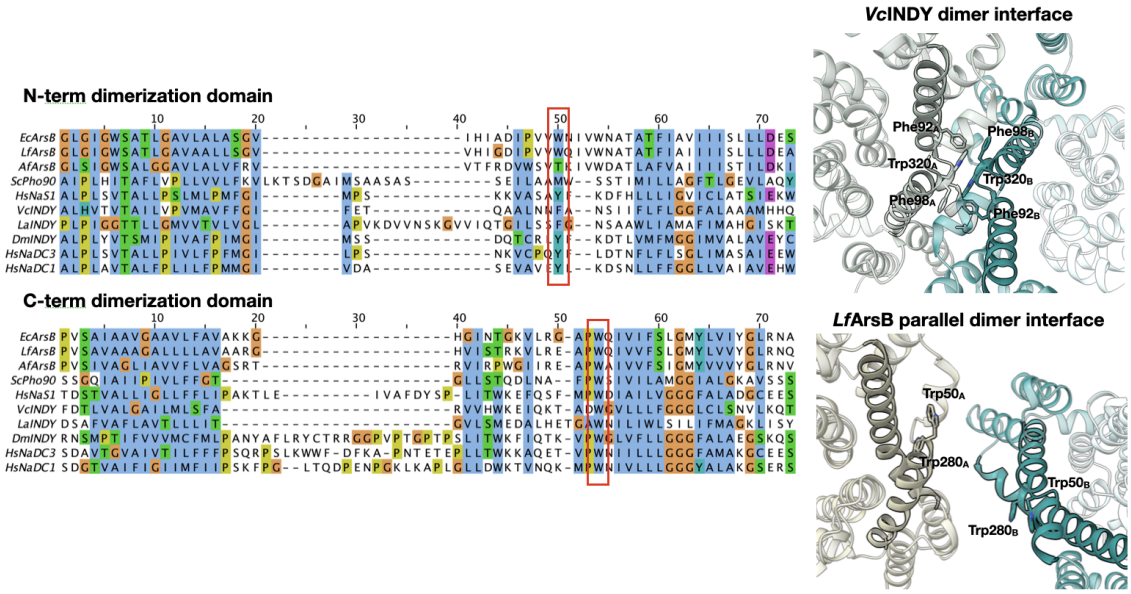


Figure S4.12: **Multiple sequence alignment of ArsB and DASS family transporters highlighting conserved aromatic residues that form contacts at the dimer interface in DASS proteins.** Promals3D [40] was used to create structure-based sequence alignment, utilizing the structure of *Lf* ArsB, and one structure each of the other DASS proteins. On the left, comparison of VcINDY and *Lf* ArsB dimer structures suggests that despite the conserved aromatic residues present in both proteins, extensive dimer contacts are only observed in VcINDY structure.

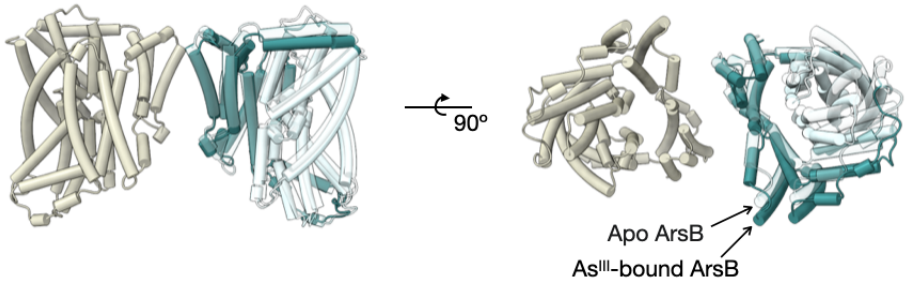


Figure S4.13: **Overlay of the *Lf* ArsB parallel dimers from apo (50% opacity) and As^{III}-bound (100% opacity) structures showing a shift in the relative orientation of the subunits between the two dimers.**

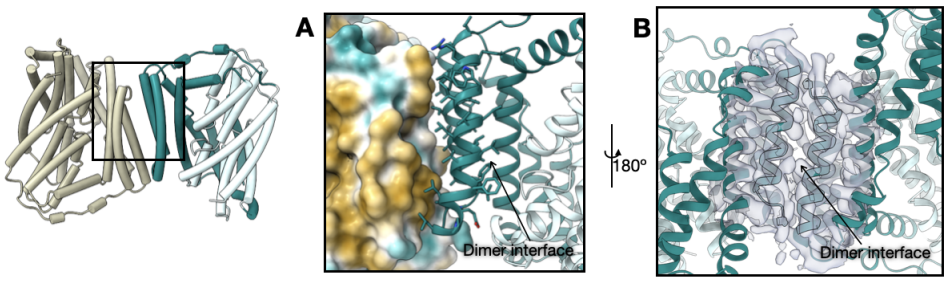


Figure S4.14: **Dimer interface of *Lf* ArsB antiparallel dimer.** (A) Hydrophobic residues of the scaffold domain forming inter-subunit contacts at the dimer interface. (B) Density for putative lipids trapped in the hydrophobic dimer interface.

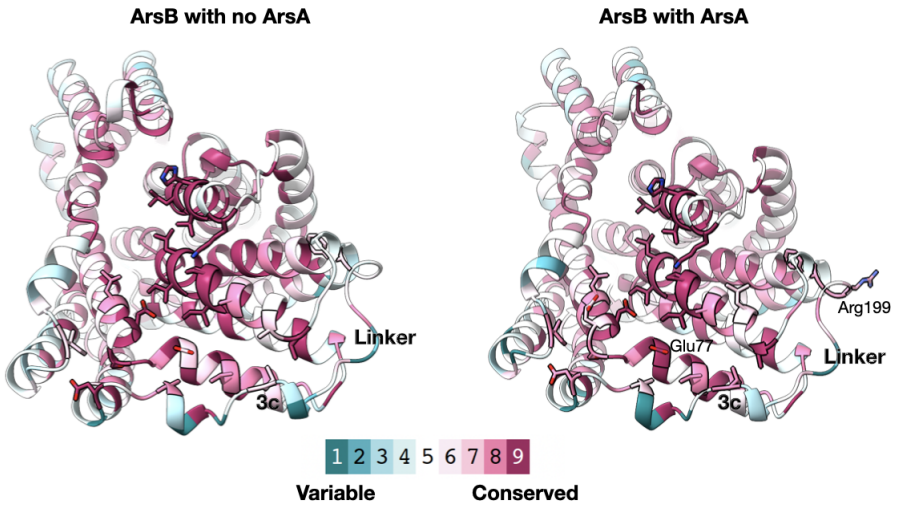


Figure S4.15: **Sequence conservation mapped onto *Ec*ArsB AlphaFold models for ArsB sequences found alone (left) and in association with ArsA (right).** Analysis was performed using Consurf server [33], using custom MSAs corresponding to the two subsets of ArsB sequences.

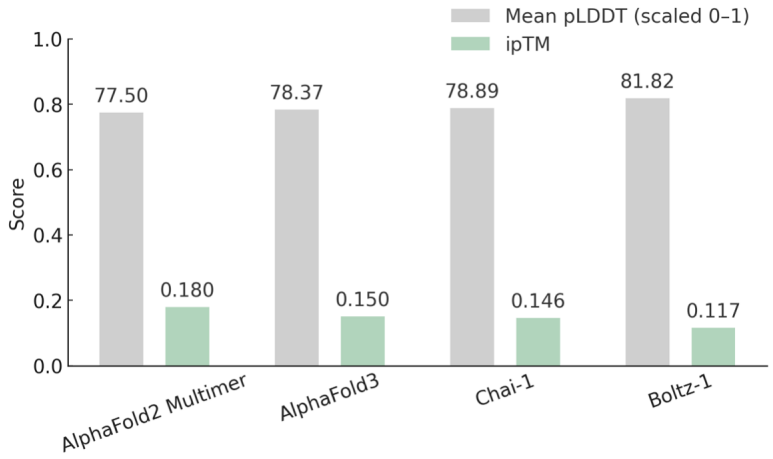


Figure S4.16: **Comparison of prediction confidence scores (pLDDT and ipTM) across various models for ArsAB complex modeling.**

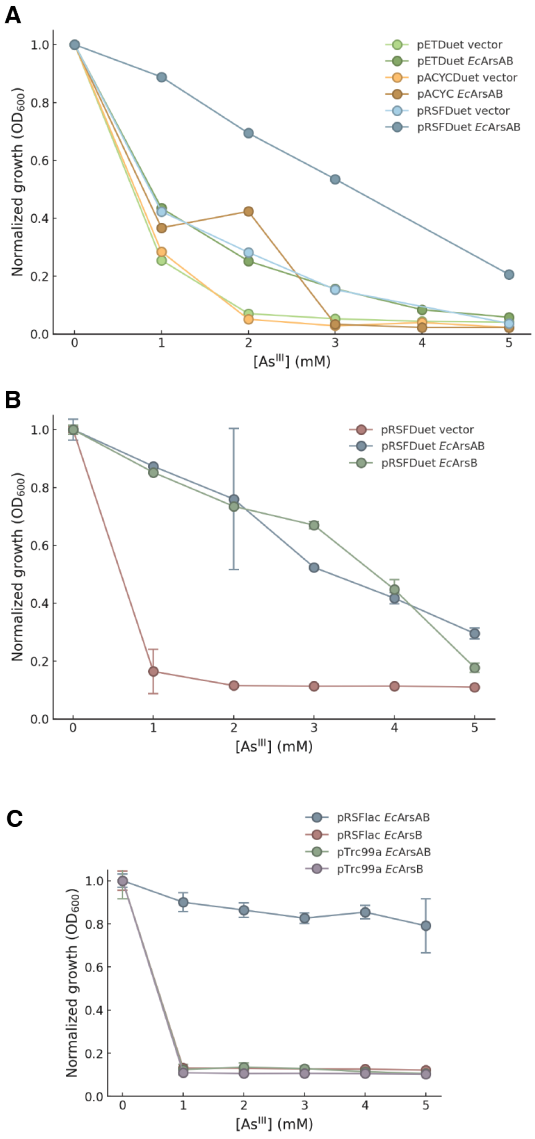


Figure S4.17: **Optimization of As^{III} resistance growth assays in *E. coli*.** (A) Comparison of growth between pETDuet, pRSFDuet and pACYCDuet vectors bearing *Ec*ArsAB in *E. coli* strain JM109, measured after 12-18 h at 37ºC. (B) Comparison of growth between *Ec*ArsAB and *Ec*ArsB cloned into pRSFDuet vector in *E. coli* strain AW3110, measured after 10 h at 37ºC. (C) Comparison of growth between pRSFlac (lac promoter) and pTrc99 (trc promoter) vectors bearing *Ec*ArsAB or *Ec*ArsB in *E. coli* strain AW3110, measured after 10 h at 37ºC.

Table S4.4: **Cryo-EM data collection and refinement statistics.**

	***Lf*ArsB apo parallel dimer**	***Lf*ArsB apo antiparallel dimer**	***Lf*ArsB + As^{III} parallel dimer**	***Lf*ArsB + Sb^{III} antiparallel dimer**
Data collection and processing				
Magnification	130,000	130,000	130,000	130,000
Voltage (kV)	300	300	300	300
Electron exposure (e^{-}/Å^2)	70	70	70	70
Defocus range (μm)	−0.8 to −2.8	−0.8 to −2.8	−0.8 to −2.8	−0.8 to −2.8
Pixel size (Å) (super-resolution)	0.325	0.325	0.325	0.325
Movies	10,743	10,743	4,602	8,020
Total extracted particles	616,321	616,321	316,729	343,002
Final particles	57,278	36,254	37,353	54,738
Symmetry imposed	C2	C1	C2	C1
Map resolution (Å) (FSC 0.143)	3.6	3.3	3.4 (3.3)	3.3
Refinement				
Initial model used	AlphaFold model	AlphaFold model	AlphaFold model	AlphaFold model
Protein residues	856	856	856	856
Ligands	–	–	$As(OH)_3$	$Sb(OH)_3$
Map sharpening *B* factor (Å^2)	−129	−116	−102 (−92)	−116
R.m.s. bond lengths (Å)	0.003	0.003	0.004 (0.003)	0.003
R.m.s. bond angles (°)	0.503	0.534	0.499 (0.426)	0.465
Validation				
MolProbity score	1.36	1.45	1.13 (1.13)	1.95
Clashscore	4.47	5.08	1.71 (1.87)	18.56
Poor rotamers (%)	0	0	0 (0)	0
Ramachandran favored (%)	97.3	96.95	96.71 (96.95)	96.95
Ramachandran allowed (%)	2.7	3.05	3.29 (3.05)	3.05
Ramachandran outliers (%)	0	0	0	0

*Values in parentheses correspond to the local refinement map.

REFERENCES

[1] S. Silver et al. Inducible plasmid-determined resistance to arsenate, arsenite, and antimony (III) in escherichia coli and Staphylococcus aureus. In: *Journal of Bacteriology* 146.3 (June 1981), pp. 983–996. DOI: 10.1128/jb.146.3.983-996.1981.

[2] S. Dey and B. P. Rosen. Dual mode of energy coupling by the oxyanion-translocating ArsB protein. In: *Journal of Bacteriology* 177.2 (1995), pp. 385–9. DOI: 10.1128/jb.177.2.385-389.1995.

[3] A. Ramírez-Solís et al. Experimental and theoretical characterization of arsenite in water: insights into the coordination environment of As- O. In: *Inorganic chemistry* 43.9 (2004), pp. 2954–2959.

[4] Y.-L. Meng, Z. Liu, and B. P. Rosen. As(III) and Sb(III) Uptake by GlpF and Efflux by ArsB in Escherichia coli*. In: *Journal of Biological Chemistry* 279.18 (Apr. 30, 2004), pp. 18334–18341. DOI: 10.1074/jbc.M400037200.

[5] S. Prakash et al. The ion transporter superfamily. In: *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1618.1 (2003), pp. 79–92.

[6] C. Mulligan et al. The bacterial dicarboxylate transporter VcINDY uses a two-domain elevator-type mechanism. In: *Nature Structural & Molecular Biology* 23.3 (Mar. 2016), pp. 256–263. DOI: 10.1038/nsmb.3166.

[7] J. A. Delmar and E. W. Yu. The AbgT family: A novel class of antimetabolite transporters. In: *Protein Science : A Publication of the Protein Society* 25.2 (Feb. 2016), pp. 322–337. DOI: 10.1002/pro.2820.

[8] J. S. Davies et al. Structure and mechanism of a tripartite ATP-independent periplasmic TRAP transporter. In: *Nature Communications* 14.1 (Feb. 27, 2023), p. 1120. DOI: 10.1038/s41467-023-36590-1.

[9] J. Wu, L. S. Tisa, and B. P. Rosen. Membrane topology of the ArsB protein, the membrane subunit of an anion-translocating ATPase. In: *Journal of Biological Chemistry* 267.18 (1992), pp. 12570–6.

[10] J. Jumper et al. Highly accurate protein structure prediction with AlphaFold. In: *Nature* 596.7873 (2021), pp. 583–589. DOI: 10.1038/s41586-021-03819-2.

[11] D. B. Sauer et al. Structural basis for the reaction cycle of DASS dicarboxylate transporters. In: *eLife* 9 (Sept. 1, 2020), e61350. DOI: 10.7554/eLife.61350.

[12] S. Schneider, W. Kühlbrandt, and Ö. Yildiz. Complementary structures of the yeast phosphate transporter Pho90 provide insights into its transport mechanism. In: *Structure* 32.7 (July 11, 2024), 979–988.e4. DOI: 10.1016/j.str.2024.04.005.

[13] X. Chi et al. Cryo-EM structures of the human NaS1 and NaDC1 transporters revealed the elevator transport and allosteric regulation mechanism. In: *Science Advances* 10.13 (Mar. 29, 2024), eadl3685. DOI: 10.1126/sciadv.adl3685.

[14] M. F. Peter et al. Conformational coupling of the sialic acid TRAP transporter HiSiaQM with its substrate binding protein HiSiaP. In: *Nature Communications* 15.1 (Jan. 8, 2024), p. 217. DOI: 10.1038/s41467-023-44327-3.

[15] D. B. Sauer et al. Structural basis of ion - substrate coupling in the Na+-dependent dicarboxylate transporter VcINDY. In: *Nature Communications* 13.1 (May 12, 2022), p. 2644. DOI: 10.1038/s41467-022-30406-4.

[16] M. J. Currie et al. Structural and biophysical analysis of a Haemophilus influenzae tripartite ATP-independent periplasmic (TRAP) transporter. In: *eLife* 12 (Feb. 13, 2024). Ed. by R. B. Stockbridge and M. Maduke, RP92307. DOI: 10.7554/eLife.92307.

[17] D. Drew and O. Boudker. Ion and lipid orchestration of secondary active transport. In: *Nature* 626.8001 (Feb. 2024), pp. 963–974. DOI: 10.1038/s41586-024-07062-3.

[18] C. Lee et al. A two-domain elevator mechanism for sodium/proton antiport. In: *Nature* 501.7468 (Sept. 2013), pp. 573–577. DOI: 10.1038/nature12484.

[19] Y. Ji et al. Transport mechanism of a glutamate transporter homologue GltPh. In: *Biochemical Society Transactions* 44.3 (June 9, 2016), pp. 898–904. DOI: 10.1042/BST20160055.

[20] S. Kim et al. Cryo-EM structures reveal the H+/citrate symport mechanism of Drosophila INDY. In: *Life Science Alliance* 8.4 (Apr. 1, 2025). DOI: 10.26508/lsa.202402992.

[21] J. S. Davies et al. TRAPs: the 'elevator-with-an-operator' mechanism. In: *Trends in Biochemical Sciences* 49.2 (Feb. 1, 2024), pp. 134–144. DOI: 10.1016/j.tibs.2023.11.006.

[22] L. S. Tisa and B. P. Rosen. Molecular characterization of an anion pump. The ArsB protein is the membrane anchor for the ArsA protein. In: *Journal of Biological Chemistry* 265.1 (1990), pp. 190–4.

[23] J. A. Gerlt et al. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. In: *Biochimica Et Biophysica Acta* 1854.8 (Aug. 2015), pp. 1019–1037. DOI: 10.1016/j.bbapap.2015.04.015.

[24] D. Dou et al. Construction of a chimeric ArsA-ArsB protein for overexpression of the oxyanion-translocating ATPase. In: *Journal of Biological Chemistry* 267.36 (1992), pp. 25768–75.

[25] J. M. Hsieh et al. Bridging the gap: A GFP-based strategy for overexpression and purification of membrane proteins with intra and extracellular C-termini. In: *Protein Science : A Publication of the Protein Society* 19.4 (Apr. 2010), pp. 868–880. DOI: 10.1002/pro.365.

[26] M. A. Lomize et al. OPM database and PPM web server: resources for positioning of proteins in membranes. In: *Nucleic Acids Research* 40 (Database issue Jan. 2012), pp. D370–376. DOI: 10.1093/nar/gkr703.

[27] A. Vera Rodriguez, S. Frey, and D. Görlich. Engineered SUMO/protease system identifies Pdr6 as a bidirectional nuclear transport receptor. In: *The Journal of Cell Biology* 218.6 (June 28, 2019), pp. 2006–2020. DOI: 10.1083/jcb.201812091.

[28] T. Bepler et al. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. In: *Nature Methods* 16.11 (Nov. 2019), pp. 1153–1160. DOI: 10.1038/s41592-019-0575-8.

[29] L. Holm et al. DALI shines a light on remote homologs: One hundred discoveries. In: *Protein Science: A Publication of the Protein Society* 32.1 (Jan. 2023), e4519. DOI: 10.1002/pro.4519.

[30] D. B. Sauer et al. Structure and inhibition mechanism of the human citrate transporter NaCT. In: *Nature* 591.7848 (Mar. 2021), pp. 157–161. DOI: 10.1038/s41586-021-03230-x.

[31] C. D. Kinz-Thompson et al. *Elevator mechanism dynamics in a sodium-coupled dicarboxylate transporter*. Jan. 26, 2025. DOI: 10.1101/2022.05.01.490196.

[32] J. Li et al. Proton-coupled transport mechanism of the efflux pump NorA. In: *Nature Communications* 15.1 (May 27, 2024), p. 4494. DOI: 10.1038/s41467-024-48759-3.

[33] H. Ashkenazy et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. In: *Nucleic acids research* 44 (W1 2016), W344–W350. DOI: 10.1093/nar/gkw408.

[34] H. Jia and P. Kaur. Role of the Linker Region of the Anion-stimulated ATPase ArsA: EFFECT OF DELETION AND POINT MUTATIONS IN THE LINKER REGION*. In: *Journal of Biological Chemistry* 276.31 (Aug. 1, 2001), pp. 29582–29587. DOI: 10.1074/jbc.M103042200.

[35] A. Punjani et al. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. In: *Nature Methods* 14.3 (Mar. 2017), pp. 290–296. DOI: 10.1038/nmeth.4169.

[36] M. Mirdita et al. ColabFold: making protein folding accessible to all. In: *Nature Methods* 19.6 (June 2022), pp. 679–682. DOI: 10.1038/s41592-022-01488-1.

[37] J. Abramson et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. In: *Nature* 630.8016 (June 2024), pp. 493–500. DOI: 10.1038/s41586-024-07487-w.

[38] C. D. Team et al. *Chai-1: Decoding the molecular interactions of life*. Oct. 11, 2024. DOI: 10.1101/2024.10.10.615955.

[39] J. Wohlwend et al. *Boltz-1 Democratizing Biomolecular Interaction Modeling*. Nov. 20, 2024. DOI: 10.1101/2024.11.19.624167.

[40] J. Pei, B.-H. Kim, and N. V. Grishin. PROMALS3D: a tool for multiple protein sequence and structure alignments. In: *Nucleic Acids Research* 36.7 (Apr. 2008), pp. 2295–2300. DOI: 10.1093/nar/gkn072.

Chapter 5

CONCLUSIONS

Arsenic toxicity is a widespread environmental and health concern across the globe. Robust and sustainable strategies for arsenic bioremediation are widely sought after to mitigate the toxic effects of the metalloid. Arsenite (As^{III}) efflux pumps offer a simple yet efficient strategy for arsenic detoxification and is adopted by organisms spanning all domains of life. Understanding the molecular mechanism of As^{III} efflux systems holds broad potential to inform biotechnological strategies for bioremediation. In addition to remediation from contaminated groundwater, arsenic bioaccumulation in rice crops, that serves as the staple for more than half of the world's population, can be mitigated by engineering robust efflux pumps into the plant membranes. Design of such strategies can be guided by efflux pumps found in nature. In this thesis, we investigated the molecular mechanisms of a widely-recognized bacterial As^{III} efflux pump, ArsB and its soluble ATPase partner, ArsA, using structural approaches.

ArsB is a unique efflux system due to its mechanistic versatility, as it can remove As^{III} either as a secondary H^+-coupled antiporter or a primary ATP-coupled transporter by associating with ArsA. While some ABC transporters and P-type ATPases are capable of removing toxic metals in ATP-dependent manner and Major Facilitator Superfamily (MFS) efflux pumps depend on a secondary ion gradient for transport, no major transporter family other than ArsB has been demonstrated to operate by both mechanisms of transport. We combined structural characterization using single particle cryo-EM with biochemistry, spectroscopy and *in vivo* assays to elucidate the mechanisms of these two modes of As^{III} efflux.

In Chapter 4, we solved cryo-EM structures of ArsB that revealed an 'inward-open conformation'. In addition, we determined structures in the presence of two metalloid substrates, As^{III} and Sb^{III}, both of which adopted the same "inward-open" conformation, indicating this conformation is perhaps the resting state of ArsB. The metalloid-bound structures reveal that both metalloids bind in a highly conserved site enclosed within the TMDs lined by polar residues (Asn, Asp and Ser), suggesting that ArsB likely transports the metalloid in its trihydroxylated form via labile hydrogen-bonding interactions. As ArsB is described to operate

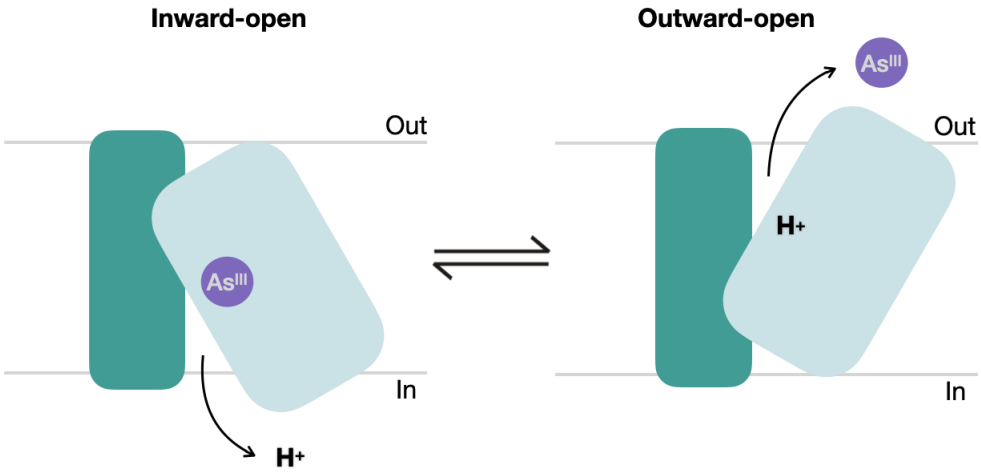


Figure 5.1: **Proposed elevator mechanism for H^+-coupled As^{III} antiport by ArsB.**

as an H^+-coupled antiporter, the associated conformational changes of its transport cycle, following the metalloid-bound state, remain to be characterized. In the framework of the elevator mechanism, ArsB must undergo conformational changes between broadly two states — a cytoplasmic (inward-open) state where the transport domain is accessible to the cytoplasm facilitating metalloid binding/capture, and a periplasmic (outward-open) state where the transport domain is accessible to the periplasm facilitating metalloid release (Fig. 5.1). In addition, metalloid-bound occluded intermediate states must exist as well. For H^+-coupled antiport mechanism, the elevator movements of the transporter must enable reciprocal influx of H^+ and efflux of metalloid (Fig. 5.1). Whether a conserved hydrogen-bonding network observed at the substrate binding site of ArsB is directly involved in a proton-relay mechanism to drive ArsB transport cycle, remains to investigated.

When ArsB functions in association with ArsA, it confers higher resistance to metalloid. ArsA not only couples ATP hydrolysis to delivery of the metalloid to ArsB, but also to its capture from the metallochaperone, ArsD in the cytoplasm. As discussed in Chapter 3, our cryo-EM structures of ArsA highlight the nucleotide- and As^{III}-dependent global conformational changes throughout its ATPase cycle. These structures illustrate that the nucleotide state of ArsA allosterically regulates its affinity for As^{III} and hence, modulates the capture and release of the metalloid for efflux. This property makes ArsA a central component of the *arsRDABC* arsenite efflux system. The high affinity metalloid binding site in the ATP-closed state of

ArsA is formed by major structural rearrangements that bring the three conserved cysteines in proximit to stabilize tricoordinate As^{III} binding via three As-S polar covalent bonds (2.7 Å). Compared to the hydrogen-bonding interactions between $As(OH)_3$ and the surrounding residues (3-4 Å) as seen in the ArsB structure, the tricoordinate As-S binding in ArsA is roughly an order of magnitude stronger. The difference in the nature of As^{III} binding between ArsA and ArsB is mechanistically essential, ensuring capture of As^{III} by ArsA and subsequent delivery to ArsB for transport.

However, some questions regarding the mechanism of ArsA itself remain unclear such as the functional role of the pseudodimer, which is a distinctive feature of ArsA in context of the broader IWA family. While an asymmetric metalloid binding site seems necessary to facilitate tricoordinate As^{III} binding, the non-equivalence of the nucleotide binding sites is difficult to rationalize. Despite sequence-level differences, the Switch I and Switch II loops adopt similar conformations in the two domains in the *ATP-closed* state from *wild-type* ArsA, unlike the non-hydrolyzing variant structure. Additionally, a network of water molecules ('water wire') bridging the catalytic aspartates of the two domains is observed only in the wild-type structure. Hydrogen-bonding networks mediated by ordered water molecules are widely implicated in allosteric regulation in proteins [1]. These observations collectively point toward coordination between the two domains of the pseudodimer, supporting the idea that due to allosteric coupling of nucleotide hydrolysis, the N- and C-sites of ArsA are equivalent in near- and mid-hydrolysis states.

While the structures of ArsA elucidate the mechanistic details of ATP-coupled transfer of the metalloid to ArsB, how the ATPase associates with ArsB and facilitates metalloid transport remains a major outstanding question. Our bioinformatic analysis revealed poor sequence-level differences near the cytoplasmic side between the ArsA-associated and ArsA-free ArsBs. This finding is consistent with the observation that ArsA of *E. coli* R773 plasmid not only improves arsenic resistance levels of R773 ArsB and but also of *S. aureus* ArsB, while the two ArsBs are only related by 58% sequence identity [2]. This prompted us to re-evaluated the requirement of sequence-level specificity for ArsAB interaction. Owing to the dual mode of efflux demonstrated by ArsB, we argued that the two proteins perhaps undergo a weak interaction which is tunable depending on the mode of transport. In Chapter 5, we showed activation of ArsA ATPase activity *in vitro* by purified ArsB, as the first evidence for a direct interaction between the two proteins. Moreover, computational

modeling suggests a weak low-affinity complex, likely with the 'closed' state of ArsA. These findings collectively suggest a mechanism of conformational coupling between ArsB and ArsA, meaning that a specific conformation of ArsB selectively interacts with the 'closed' state of ArsA. Upon ATP hydrolysis and phosphate release, metalloid is transferred to ArsB, transitioning the 'closed' state of ArsA into the open state, which results in the dissociation of the complex and transport of the metalloid. Whether it is the inward or the outward state of ArsB that interacts with the 'closed' state of ArsA is unclear. Although not ATP-dependent, the ArsAB interaction can be mechanistically compared with the tripartite complex of the TRAP transporter, SiaPQM that requires the substrate binding protein, SiaP for substrate delivery to the transporter and facilitate the 'elevator-type' conformational changes [3]. Similar to SiaP, ArsA facilitates metalloid delivery to ArsB; however, while SiaQM still depends on Na^+-gradient for transport, ArsB must depend on the ATP hydrolysis cycle of ArsA to drive metalloid transport, the molecular details of which are currently elusive. Both SiaQM and ArsB are related under the IT superfamily along with the DASS transporters. However, the duality of substrate transport mechanism of ArsB highlights its versatility as a transporter, condensing properties of both secondary and primary active transport in one membrane protein fold. The studies presented in this thesis provide a foundation for further elucidation of the mechanistic questions surrounding arsenite efflux outlined above.

REFERENCES

[1] M. I. Kozlova et al. Common Patterns of Hydrolysis Initiation in P-loop Fold Nucleoside Triphosphatases. In: *Biomolecules* 12.10 (Sept. 2022), p. 1345. DOI: 10.3390/biom12101345.

[2] S. Dey and B. P. Rosen. Dual mode of energy coupling by the oxyanion-translocating ArsB protein. In: *Journal of Bacteriology* 177.2 (1995), pp. 385–9. DOI: 10.1128/jb.177.2.385-389.1995.

[3] J. S. Davies et al. TRAPs: the 'elevator-with-an-operator' mechanism. In: *Trends in Biochemical Sciences* 49.2 (Feb. 1, 2024), pp. 134–144. DOI: 10.1016/j.tibs.2023.11.006.

Appendix A

X-RAY CRYSTAL STRUCTURE OF ARSA FROM *L. FERRIPHILUM* IN A *SPLAYED* CONFORMATION

A.1 Introduction

The ArsA ATPase, first functionally characterized in the 1980s, has garnered much interest for its nucleotide hydrolysis mechanism and the role thereof, in facilitating As^{III} efflux via ArsB. By the early 2000s, X-ray crystal structures of the nitrogenase Fe protein (NifH), a distant relative of ArsA under the IWA family, had revealed key nucleotide-dependent conformational changes elucidating a critical mechanism of electron shuttling in nitrogenase [1, 2]. At the time, crystal structures of ArsA had already been determined in four distinct nucleotide-bound states. However, the nucleotide-dependent conformational changes described based on these structures did not reflect the global conformational rearrangements that were outlined for NifH, thus making ArsA an outlier in the mechanistic landscape of the broader IWA ATPase family.

Four structures of *E. coli* ArsA have been reported in the presence of various nucleotides or nucleotide analogs. In each of these structures, MgADP is bound at the nucleotide binding site on the N-domain, whereas MgATP, MgAMPPNP or MgADP•AlF_4^- replace MgADP at the analogous site on the C domain in three of these structures (Fig. 3.1) [3, 4]. Apart from subtle structural changes in the coordination environment of the catalytic Mg^{2+} ion, these structures represented the same pseudodimer conformation. Our cryo-EM structures discussed in Chapter 3 have now demonstrated that the crystal structures represent an 'open' conformation of ArsA. Based on our analysis of these structures, there are several plausible reasons that restricted the conformational changes in the ArsA crystal structures. First, various nucleotide-bound states were crystallized using the same crystallization condition, and likely restrained by the crystal lattice packing. This is supported by the observation that despite ADP- and ADP•AlF_3-bound structures were solved in space group I222, and ATP- and AMP•PNP-bound structures were solved in space group $P2_12_12$, the lattice packing was conserved between the two crystal forms. Second, some of the structures were obtained by soaking the ADP•ArsA crystals in other nucleotides, thus restricting any corresponding large-scale conformational

changes that could disrupt the crystal packing. Finally, only the ADP bound structure (PDB 1F48) shows strong electron density at the metalloid binding site that was modeled as Sb^{III}, whereas the other three structures have Cd^{II} modeled at this site instead. Upon closer inspection of the metalloid binding site in these structures, we inferred that the species in the ADP-bound structure is likely Cd^{II} as well, instead of Sb^{III}, due to similar atomic radii (0.4 Å) of the two ions. As Cd^{II} is not a substrate of ArsA, the divalent ion likely restricted conformational changes. Together these factors support that the crystallization conditions were not conducive to explore the complete conformational landscape of ArsA.

In addition to our cryo-EM efforts discussed in Chapter 3, we attempted to trap distinct catalytic conformations of ArsA using X-ray crystallography. Here, we report the X-ray crystal structure of ArsA from *L. ferriphilum* in a conformation distinct from previously reported structures. This X-ray crystal structure represents a 'splayed' conformation of ArsA, wherein ADP is bound only at the N-domain and the C-domain adopts a flanked position relative to the N-domain. While a similar conformation, to our knowledge, has not been noted to date for an IWA ATPase, the relative orientations of N- and C-domains, covalently connected by a peptide linker, dramatically highlights the intrinsic conformational flexibility ArsA. Moreover, elucidation of this structure serves as a lesson in handling complex pathologies such as translational non-crystallographic symmetry (tNCS) that initially complicated refinement, but was solved by estimation of phases by a combination of molecular replacement and single anomalous diffraction (MR-SAD) strategy.

A.2 Results and discussion

A.2.1 Initial crystallization screening

Initial crystallization hits for *Lf*ArsA were obtained from a condition containing 100 mM Na citrate pH 6.5, 0.5 M ammonium sulfate and 1 M lithium sulfate. Crystals resulting from initial optimizations of this condition diffracted to about 6 Å; however, with poor diffraction quality (Fig. A.1). Severely mosaic diffraction spots resulted from multiple crystal forms stacked together in one. Optimizing this condition using aliquot or streak seeding and using additive screens did not seem to improve to the quality of the diffraction (Fig. A.1). Moreover, a wide cryoprotectant screening was performed to account for any detrimental effects on the diffraction quality of the crystals, albeit this did not improve diffraction. Next, we attempted to obtain new hits by aliquot seeding from a seed-stock dilution prepared using the crystals obtained above, into new crystallization screens in 96-well plates. One of

the hits obtained from these screens, containing 100 mM KSCN and 30% PEG MME-2000 resulted crystals in presence of $MgCl_2$, ADP and arsenite, with a more geometrical morphology than the crystals used to obtain the seed stock (Fig. A.2). Some of these crystals diffracted beyond 4 Å but importantly, the diffraction quality was considerably better compared to the previous datasets (Fig. A.2).

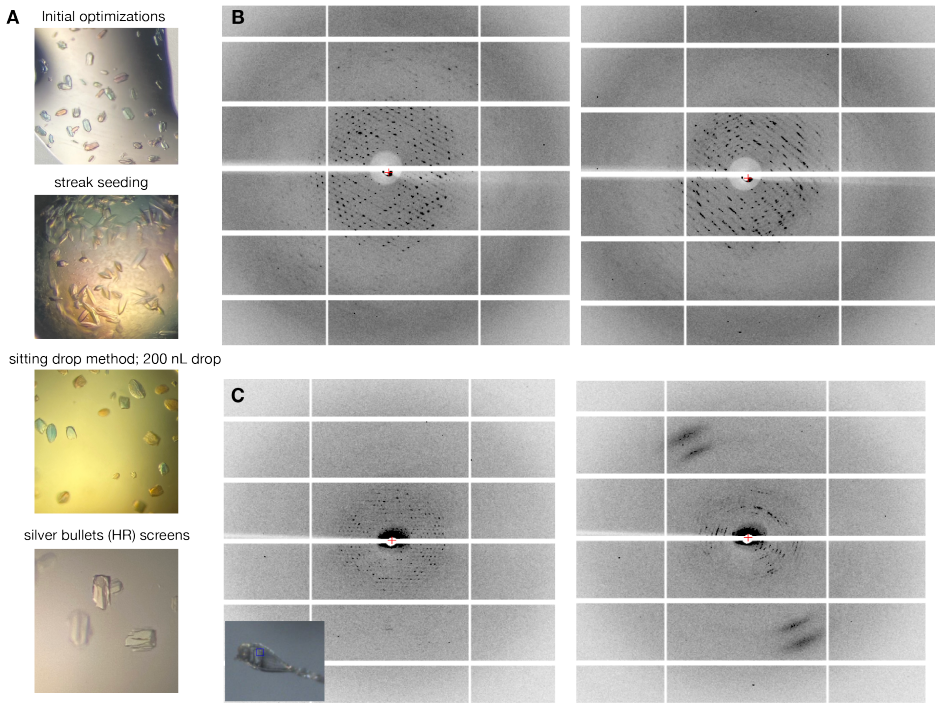


Figure A.1: **Crystallization and X-ray diffraction of *Lf* ArsA in presence of MgADP.** (A) Crystals of *Lf* ArsA obtained from crystal optimizations. (B) Diffraction images of original *Lf* ArsA crystals taken 90º apart. 15% glycerol was used as a cryoprotectant. (C.) Diffraction images of *Lf* ArsA crystals looped without cryo-protectant, taken 90º apart.

A.2.2 Molecular replacement solution from native dataset

Integration and scaling of the native *Lf* ArsA datasets in P1 space group and evaluation of the data reduction statistics from *Xtriage* strongly indicated the presence of 2-fold symmetry with a screw axis, so the subsequent steps of structure solution were carried out using $C222_1$ space group. Using a resolution cut-off corresponding to mean $I/\sigma I$ value of 1, the dataset was cut at 2.8 Å , having a corresponding R_{merg} of 10.6% and over 99% completeness (Table A.1). Moreover, no anomalous signal was observed as indicated by *SHELXC* [5], suggesting that arsenite present in the crystallization condition either did not bind to the protein or the anomalous scattering is just weak. Analysis of the scaled and merged reflection data strongly indicated the

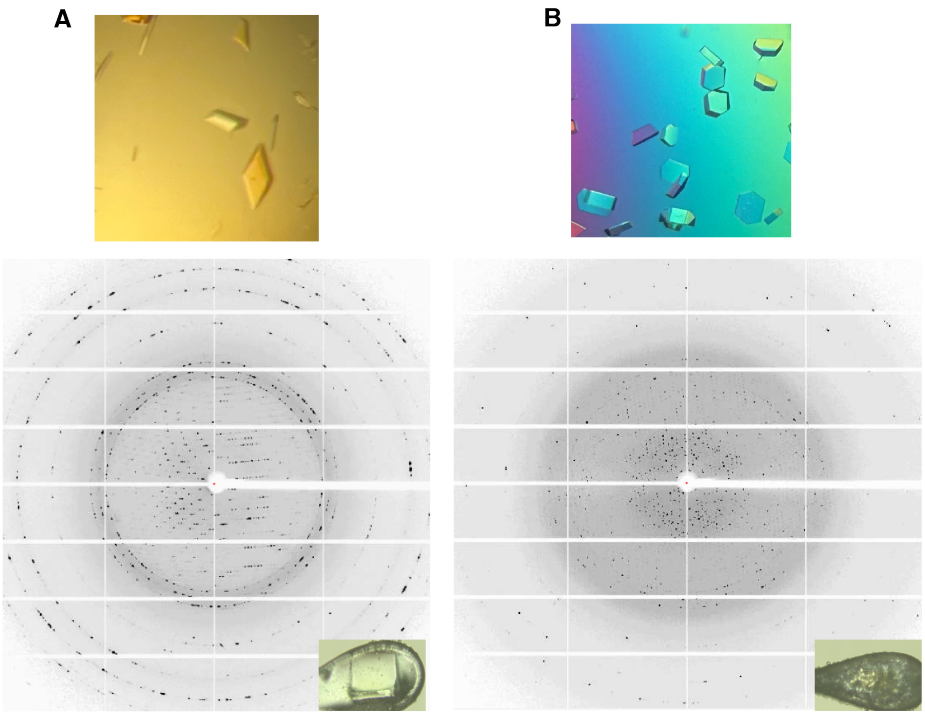


Figure A.2: **Crystallization and X-ray diffraction of *Lf*ArsA obtained from seeding into new condition in presence of MgADP and arsenite.** (A) Native crystals, top; and corresponding diffraction image from crystal soaked in mother liquor for a few seconds before cryo-cooling, bottom; inset shows the crystal within the loop on the goniometer. (B) Crystals used for $HgCl_2$ soak, top; and corresponding diffraction image from crystal soaked in mother liquor supplemented with $HgCl_2$ before cryo-cooling, bottom; inset shows the crystal within the loop on the goniometer.

presence of translational non-crystallographic symmetry (tNCS). A native Patterson map was calculated from the reflection data, which revealed a strong non-origin Patterson peak of 45% relative to the origin, with the translation vector located on the face c of the unit cell (Fig. A.3A). Based on solvent content calculations, the asymmetric unit was indicated to have at least 4 copies of ArsA. Both of these observations strongly supported the presence of tNCS [6]. The presence of a 2-fold NCS axis in the ASU that is parallel to the crystallographic 2-fold axis likely generated tNCS in this crystal lattice.

Table A.1: **X-ray crystallographic data collection and refinement statistics.**

	***Lf*ArsA native**	***Lf*ArsA Hg-1**	***Lf*ArsA Hg-2**	***Lf*ArsA Hg merged 1,2**
Data collection				
Space group	C 2 2 2_1	C 2 2 2_1	C 2 2 2_1	
Wavelength (Å)	0.98	1.00	1.00	
Unit cell a, b, c (Å)	108.9, 238.6, 257.8	108.7, 235.1, 256.2	108.8, 235.4, 256.9	
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	
Resolution (Å)*	49.0–2.80 (2.85–2.80)	49.34–2.80 (2.86–2.80)	49.38–3.20 (3.30–3.20)	
R_{merge} (%)	10.6 (170.6)	24.4 (266.8)	31.4 (191.3)	
R_{pim} (%)	4.4 (67.5)	6.6 (71.8)	12.0 (72.5)	
Mean $I/\sigma(I)$	10.3 (1.0)	9.5 (1.3)	8.3 (1.2)	
$CC_{1/2}$ (%)	99.8 (66.3)	99.7 (56.9)	98.5 (61.4)	
$CC_{1/2}$ (anomalous)	-	0.11 (-0.01)	0.11 (0.02)	0.19 (0.02)
Completeness (%)	97.8 (99.5)	99.6 (97.5)	97.2 (98.2)	
Redundancy	6.7 (7.2)	13.5 (13.4)	6.9 (6.8)	
Refinement				
Resolution (Å)	47.38–2.80			
No. reflections	80,621			
R_{work} / R_{free}	0.236 / 0.271			
No. atoms	15,375			
Protein	15,267			
Ligand/ion	108			
B-factors (Å^2)				
Protein	92.0			
Ligand	58.9			
R.m.s. deviations				
Bond lengths (Å)	0.009			
Bond angles (°)	1.789			

*Values in parentheses correspond to the highest-resolution shell.

Molecular replacement (MR) was performed using the AlphaFold prediction of *Lf* ArsA as search model using Phaser. Initially, Phaser could not find a solution using the intact pseudodimer as search model. Subsequently, N and C domains of the ArsA pseudodimer were used individually, without the linker, to enable rigid body fitting of each domain using Phaser. This MR strategy accounted for any major conformational changes between the two domains. The results from Phaser revealed 4 chains each of N- and C-domains, equivalent to 4 molecules/pseudodimers per asymmetric unit (Fig. A.3B). On inspection of the solution, none of the chains resembled the previous *Ec*ArsA structures or even the AlphaFold model, nor there was an interface formed between N- and C-domains in this solution. Intriguingly, we could see F_o-F_c positive difference density peaks near the P-loops (residues 16-22) in all N-domains corresponding to the nucleotide (Fig. A.3C), indicating that the placement of the N-domains at least was correct. However, we considered the possibility of incorrect placement of the chains owing to the high structural homology between the N- and C-domains, in addition to a large ASU. The refinement R-factors of this model could not get better than 40%. Despite the indication that the N-domains were correctly modeled, it was clear at this stage that the presence of strong tNCS was hindering phasing by MR alone.

A.2.3 Molecular Replacement-Single Anomalous Diffraction (MR-SAD) solution from Hg-derivative datasets

To determine phases for this intricate dataset, we turned to experimental phasing. A variety of heavy metal soaks of the native *Lf* ArsA crystals using Pt(II), Hg(II), I^-, Br^-, Cd(II) and Tantalum Bromide cluster (Ta_6Br_{12}) were performed to obtain heavy atom derivatized crystals by briefly washing the crystals in reservoir solutions supplementing with each heavy atom salt. During data collection, X-ray absorption scans in fluorescence mode were performed near a suitable absorption edge of each metal corresponding to their respective maximum theoretical X-ray scattering coefficient, *f"*. Data were collected at a single energy (single anomalous diffraction or SAD) slightly above the energy maxima determined from the scan (Fig. A.4A). The experimental scattering coefficients, *f"* and *f'* were calculated from the scans (Fig. A.4B), and later used for the estimation of experimental phases.

Anomalous signal was detected in Hg derivative datasets; however, the weak anomalous correlation (Table A.1) suggested that the determination of metal sites would be difficult using the anomalous differences alone. Therefore, we performed MR-SAD to determine substructure of the Hg atom sites and eventually estimate phases for the

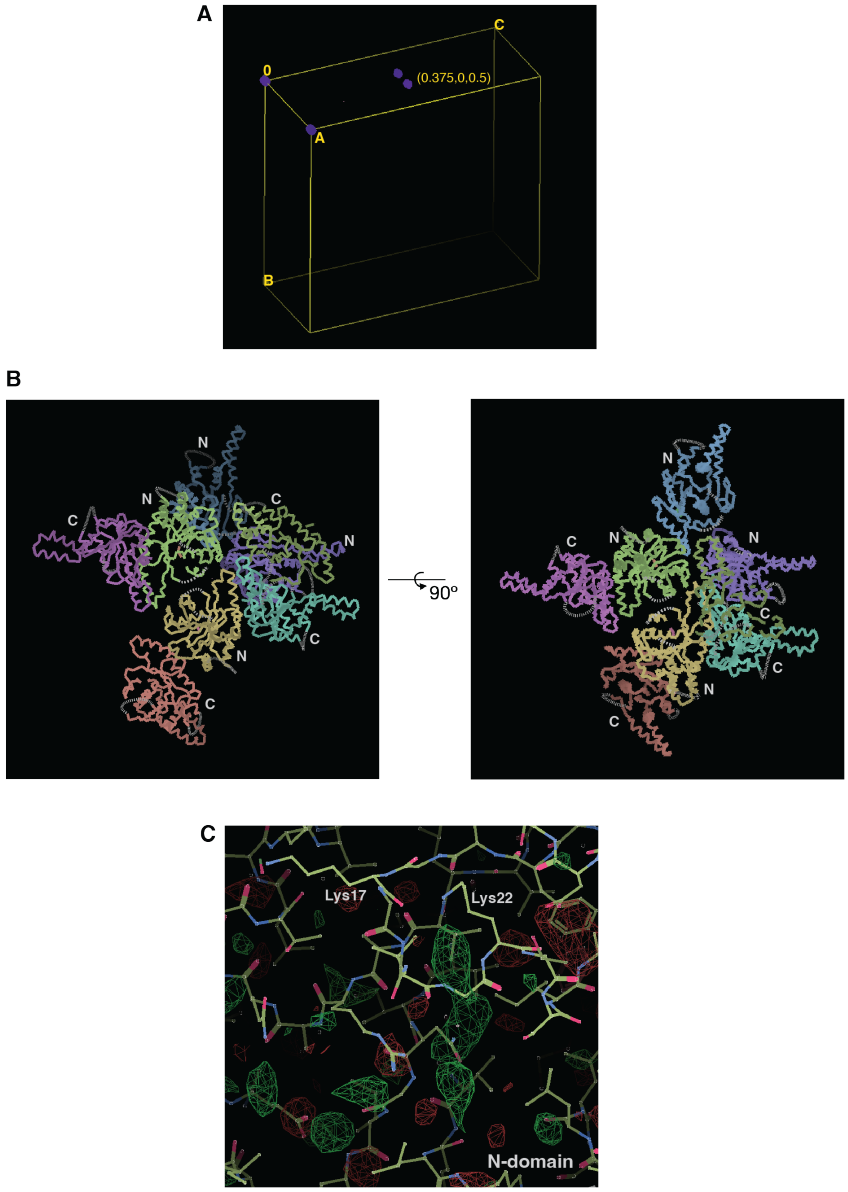


Figure A.3: **Structure solution of *Lf*ArsA (native) by Molecular Replacement.** (A) Native patterson map calculated from Intensities in the scaled and merged data. (B) Initial MR solution consisting of 4 copies of N- and C-domains each. (C) F_o-F_c difference density corresponding to the ADP bound at the P-loop of the N-domain in the MR solution from panel B.

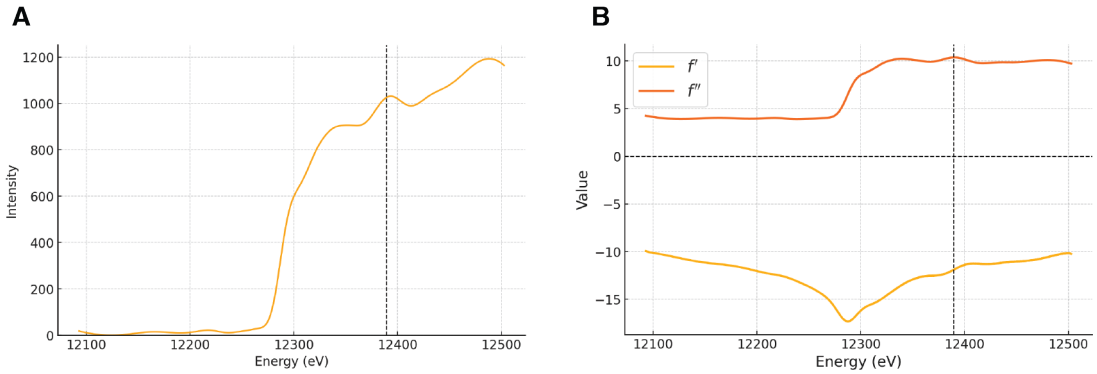


Figure A.4: **X-ray absorption edge scan of Hg-soaked *Lf*ArsA crystals.** (A) Fluorescence scan; the vertical dotted line indicates the energy chosen for data collection. (B) Experimental scattering factor coefficients for Hg, calculated from the edge scan.

overall dataset. We were fairly confident that 4 ArsA N-domains were positioned correctly in the initial MR solution, as the difference density of ADP was observed in each of the 4 chains (Fig. A.3C). To calculate the positions of the Hg atom sites by MR-SAD, we used the partial solution of 4 N chains with ADP (Fig. A.5A), and two merged datasets of Hg derivative collected at distinct spots on the same crystal (Table A.1). Anomalous difference Fourier map revealed a strong peak indicative of an Hg atom site near Cys27 residue in the N-domain of chain D of the partial model (Fig. A.5B). Subsequently, upon density modification of this substructure and automated model building, the electron density map revealed the positions of the Hg atom on the N-domains of each of the 4 chains, and showed unambiguous density for β-sheets and helices corresponding to the nucleotide binding fold of the C-domains. Initial rounds of refinements of this model with partially built C-domains revealed density for the peptide linker emerging from residue 294 of the N-domain all the way into the partially built C-domain (Fig. A.5C). Most residues forming the core of the N- and C-nucleotide binding domains and as well as the linker region were built into the model. However, few regions having high B-factors could not be built due to weak electron density, including residues 158-183 in the N-domain and residues 451-485 of the C-domain. Density for either Mg^{2+} or As^{III} was not observed, so these ligands were not modeled in. Overall, this model is similar to the one obtained originally by MR alone, consisting of 4 ArsA pseudodimers in the ASU (Fig. A.5D). Although complicated by tNCS at first, the MR-SAD strategy validates this solution. A summary of the structure determination workflow is described in Fig. A.6.

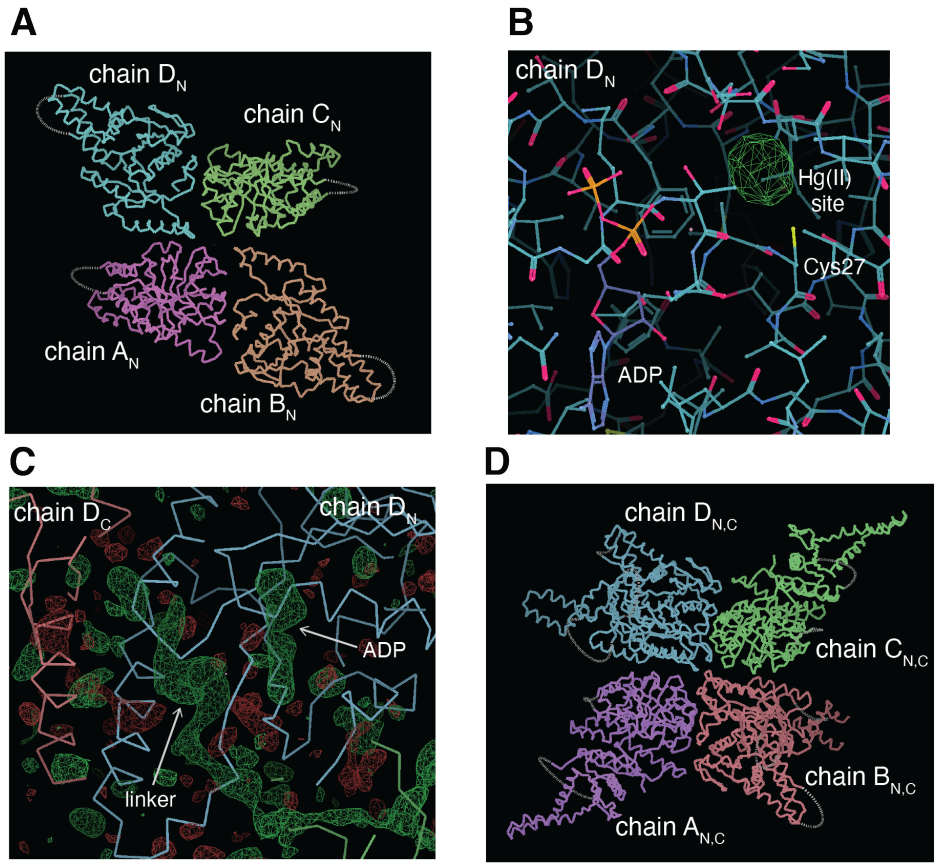


Figure A.5: **Various stages of structure solution of *Lf* ArsA by MR-SAD.** (A) Partial model of 4 N-domains in the ASU colored by chain, that was used to obtain initial phases for MR-SAD. (B) Anomalous difference Fourier map (green) contoured at 8.0σ, showing the Hg atom site next to Cys27 of the N-domain of chain D of the partial model. (C) Difference (F_o-F_c) map contoured at 2.3σ showing electron density for ADP and the linker peptide between N- and C-domains of chain D. (D) Refined model showing of the ASU composition showing chains A-D of *Lf* ArsA built to completion, except for regions of weak electron density.

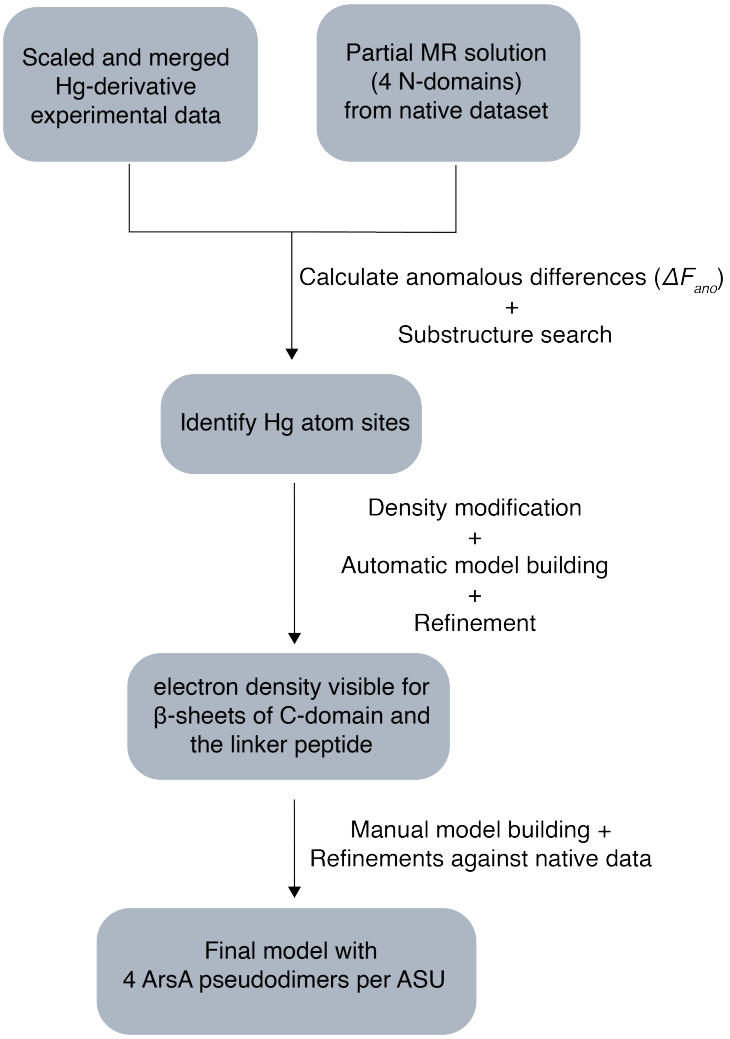


Figure A.6: **Structure solution workflow for *Lf* ArsA•ADP by MR-SAD.**

A.2.4 The 'splayed' conformation of ArsA

The resulting structure is an unprecedented conformation of ArsA (Fig. A.7A). Notably, in this conformation, the C-domain, instead of forming an interface with the N-domain at their respective nucleotide binding sites, rotates away from the pseudo-two-fold axis. As the P-loop of the C-domain is oriented about 34 Å from the N-domain P-loop, we refer to this state of ArsA as the 'splayed' conformation. Difference density for ADP was only found on the N-domain, whereas the C-domain appears to be flanking out. Comparison of this state with the N-domain of the ADP-bound *Ec*ArsA crystal structure (PDB: 1F48), suggests some rearrangements in the N-domain (Fig. A.7B). In particular, the helix hairpin formed by helices 4 and 5 seem to have re-oriented away from the pseudo-two-fold axis in the *Lf* ArsA structure (Fig. A.7B). Moreover, the mobile region between helix 7 and helix 8, that forms

the 'metalloid affinity loop' as defined in Chapter 3, is highly flexible in this state and therefore, could not be modeled. This observation, along with the large distance between the two domains bearing the conserved cysteines for metalloid binding, is consistent with the fact that this conformation is unable to bind As^{III}.

Another notable feature of this structure is the linker region of ArsA in the structure. Unlike the previous crystal structures of *Ec*ArsA, the linker in the *Lf* ArsA structure is in an ordered conformation and could be modeled completely by manually building into the difference density. Closer inspection of the linker region near its C-terminus reveals its proximity to the nucleotide binding site of the N-domain (Fig. A.7C). The backbone of the liner is located ~8 Å from the backbone of the P-loop; however, no specific interactions can be noted between the two motifs. Although far from a mechanistic implication, nucleotide binding near the linker peptide of ArsA has previously been observed [7], supporting the proximity of the linker to the nucleotide binding site.

Furthermore, we considered the possibility that this conformation of ArsA may form a dimer interface between two N-domains as visualized in chains A and B (Fig. A.8). ArsA has previously been proposed to dimerize in the nucleotide-bound state [8]. However, this model is not supported by the buried interface of ~750 sq. Å between the N-domains of chain A and B in our structure. Apart from π-π stacking interaction between $Phe281_A$ and $Phe281_B$ of the N-domains and Van der Waal contacts between $Pro297_A$ and $Pro297_B$ of the linker, no major interactions are formed between the two chains in this structure. The interactions observed in this structure are likely due to the crystal contacts.

In summary, the 'splayed' conformation of ArsA discussed here, though unprecedented, supports the notion that the ArsA pseudodimer is a conformationally flexible protein. Plausibly, in the absence of the metalloid, the two pseudo-subunits of ArsA can explore a large conformational space while still connected and restrained at the linker. Nucleotide and metalloid binding is necessary to bring the two domains closer in order to facilitate hydrolysis and transfer of the metalloid. In light of the nucleotide hydrolysis mechanism, cryo-EM structures of *Lf* ArsA, as discussed in Chapter 3 in detail, are more insightful for understanding the catalytic and metalloid transfer events.

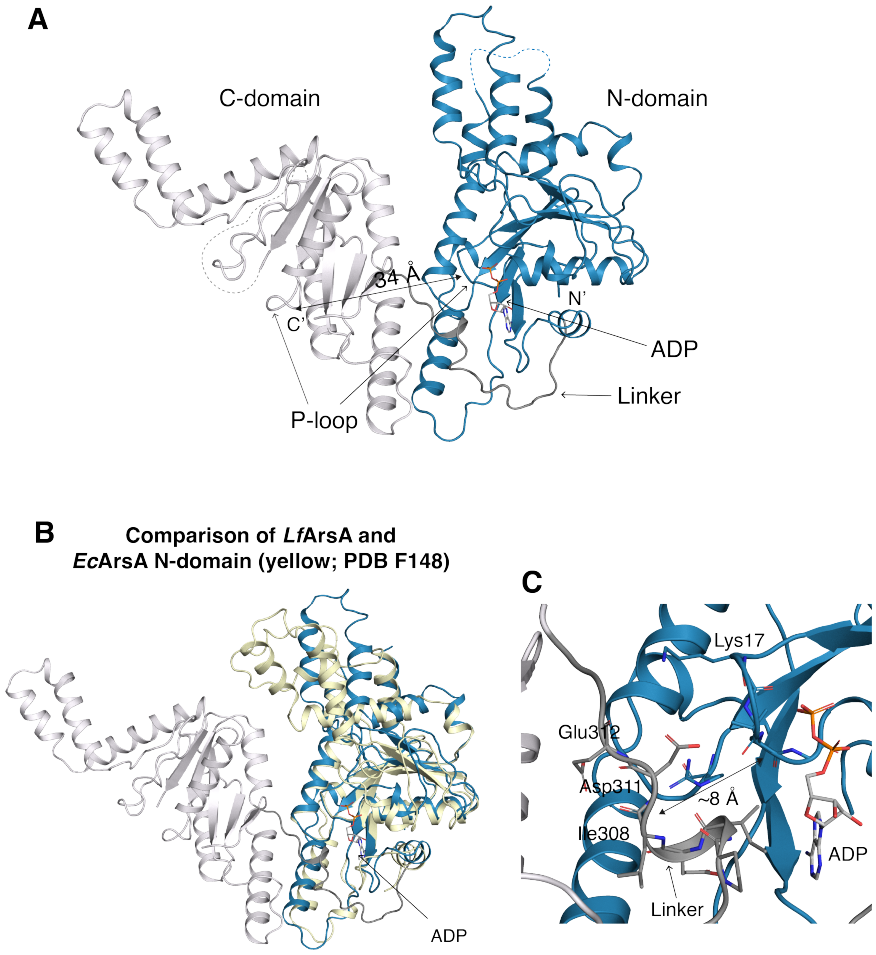


Figure A.7: **X-ray crystal structure of the 'splayed' conformation of *Lf*ArsA•ADP.** (A) Single chain of *Lf*ArsA pseudodimer showing the N-domain, C-domain and the linker peptide in blue, silver, and gray, respectively. (B) Comparison of *Lf*ArsA and *Ec*ArsA (PDB: 1F48) N-domains. (C) Close-up view of the linker peptide next to the N-domain nucleotide binding site.

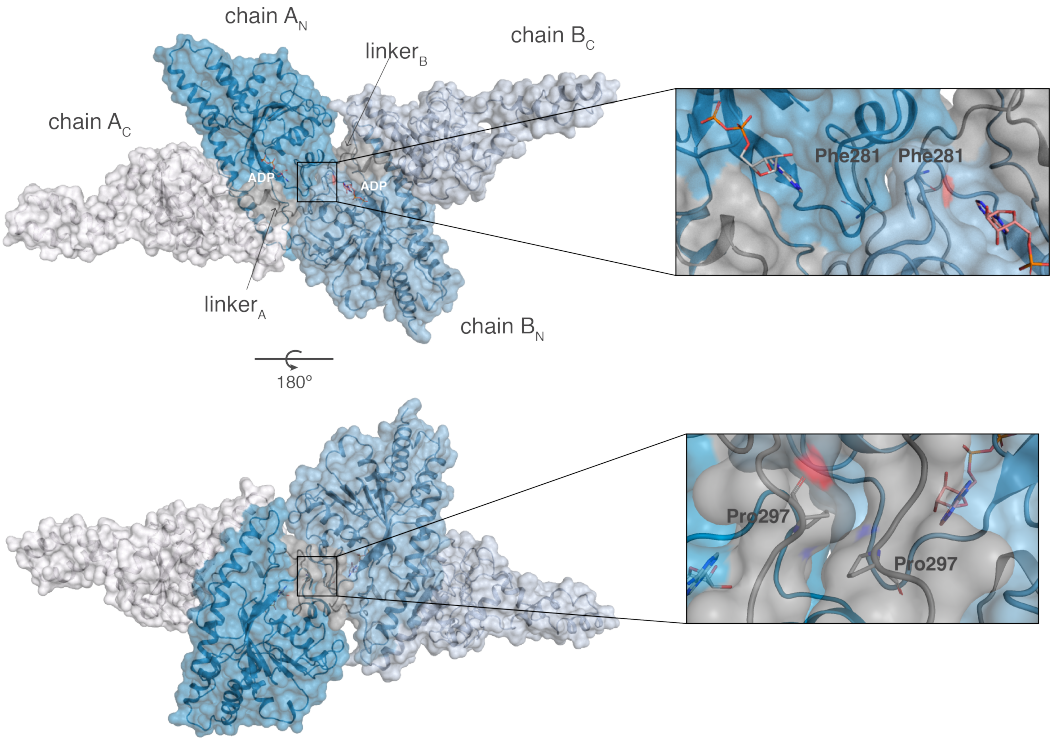


Figure A.8: **Assessment of a putative dimer interface between adjacent chains A and B of the *Lf* ArsA crystal structure.** The N-domains are shown in two shades of blue and are proposed to form the dimer interface. The C-domains are shown in two shades of gray and are flanking away from the N-domains. The linker peptide is shown in dark gray. Interfacial interactions between the N-domains (Phe281) and the linker peptides (Pro297) are shown in the zoomed-in panels on right.

A.3 Methods

A.3.1 Crystallization of *Lf* ArsA

Diffraction-quality crystals of *Lf* ArsA were obtained by the aliquot seeding method. Seed stock was prepared by homogenizing 3-4 chunky crystals obtained by growth in 100 mM Na citrate pH 6.5, 0.5 M ammonium sulfate and 1 M lithium sulfate and dilution to 1:500 and 1:1000 seed stocks using the reservoir solution. The crystallization drop contained 1:3 ratio of seed stock to reservoir solution (200 mM KSCN and 20% (w/v) PEG MME-2000), and then 5-10 mg/mL protein was added to the drop at a 1:1 ratio with seed-reservoir solution. The protein was supplemented with 2 mM $MgCl_2$, 2 mM ADP and 2 mM sodium arsenite (in that order), followed by incubation on ice for ~30 min. Any debris was separated by spinning the protein samples containing MgADP at 10,000 xg for 10 min. Crystal trays were set up in either 24-well trays by hanging-drop method or 96-well trays by sitting-drop

method. After 2-4 weeks, crystals were harvested by quickly washing them in the well solution, followed by flash-cooling in liquid nitrogen. Although initial attempts to replicate this condition using home-made reagents were unsuccessful, the commercial reagent from Hampton Research resulted in several crystals using the aliquot seeding method. For Hg soaks, crystals were washed in the mother liquor supplemented with 10 mM $HgCl_2$ for a few seconds before flash-cooling.

A.3.2 X-ray data collection and processing

All data collection was performed at the Stanford Synchrotron Radiation Lightsource (SSRL) beamline 12-2 equipped with an Eiger 16M detector. Datasets were typically collected at 20% transmittance using a 50 x 30 μm beam at an energy of 12658.43 eV. Initially, 360º of data were collected for native crystals, but the automated data processing strongly indicated the presence of 2-fold rotational symmetry and a high multiplicity of 13, so for subsequent data collections, 180º of data were collected at 0.2º φ-slicing with an exposure time of 0.2 sec at a detector distance of 300-350 mm. For Hg-soaked crystals, datasets were collected by first running an absorption edge scan, measuring the fluorescence readout, to determine presence of Hg in the crystal. Datasets were collected at an energy of 12389.28 eV, slightly higher than the edge scan maxima energy corresponding to the L-III edge of Hg. Reflections were indexed, integrated and scaled using XDS [9]. Laue group and point group symmetry were determined using POINTLESS [10]. Scaled reflection data were merged using AIMLESS [10].

A.3.3 Model building and refinement

Molecular replacement was performed using Phaser in Phenix [11, 12]. To prepare the search model, the AlphaFold model was pre-processed to trim low pLDDT regions [13], and convert the pLDDT scores into B factors, necessary to perform molecular replacement. Refinements were performed using Refmac5 [14] in CCP4i2 suite [15] or in Phenix. For the Hg derivative datasets, since the anomalous correlations for individual datasets were not enough to calculate the substructure by SAD or MR-SAD, we merged two datasets of comparable R-values collected on the same crystal. In addition to the merged reflection data, partial model containing 4 N-domains of ArsA was used for MR-SAD. Substructure determination (using Hg f' and f'' values of -12 and 10, respectively), density modification and automated model building were performed using Autosol [16] and Autobuild in Phenix. After all N- and C-domains of the ASU were built, refinements were performed against

the native dataset, which resulted in better R-factors as opposed to the Hg datasets. Model building and refinements were performed using Coot [17].

REFERENCES

[1] H. Schindelin et al. Structure of ADP·AIF4—stabilized nitrogenase complex and its implications for signal transduction. In: *Nature* 387.6631 (May 1997), pp. 370–376. DOI: 10.1038/387370a0.

[2] B. Schmid et al. Biochemical and Structural Characterization of the Cross-Linked Complex of Nitrogenase: Comparison to the ADP-AlF -Stabilized Structure . In: *Biochemistry* 41.52 (Dec. 2002), pp. 15557–15565. DOI: 10.1021/bi026642b.

[3] T. Zhou. Structure of the ArsA ATPase: the catalytic subunit of a heavy metal resistance pump. In: *The EMBO Journal* 19.17 (Sept. 2000), pp. 4838–4845. DOI: 10.1093/emboj/19.17.4838.

[4] T. Zhou et al. Conformational changes in four regions of the Escherichia coli ArsA ATPase link ATP hydrolysis to ion translocation. In: *Journal of Biological Chemistry* 276.32 (2001), pp. 30414–22. DOI: 10.1074/jbc.M103671200.

[5] G. M. Sheldrick. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. In: *Acta Crystallographica Section D: Biological Crystallography* 66 (Pt 4 Apr. 1, 2010), pp. 479–485. DOI: 10.1107/S0907444909038360.

[6] M. Jamshidiha et al. Coping with strong translational noncrystallographic symmetry and extreme anisotropy in molecular replacement with Phaser: human Rab27a. In: *Acta Crystallographica. Section D, Structural Biology* 75 (Pt 3 Feb. 28, 2019), pp. 342–353. DOI: 10.1107/S2059798318017825.

[7] J. Li and B. P. Rosen. The linker peptide of the ArsA ATPase. In: *Mol Microbiol* 35.2 (2000), pp. 361–7. DOI: 10.1046/j.1365-2958.2000.01696.x.

[8] M. H. Ching et al. Substrate-induced dimerization of the ArsA protein, the catalytic component of an anion-translocating ATPase. In: *Journal of Biological Chemistry* 266.4 (1991), pp. 2327–32.

[9] W. Kabsch. XDS. In: *Acta Crystallographica Section D: Biological Crystallography* 66.2 (Feb. 1, 2010), pp. 125–132. DOI: 10.1107/S0907444909047337.

[10] P. R. Evans and G. N. Murshudov. How good are my data and what is the resolution? In: *Acta Crystallographica. Section D, Biological Crystallography* 69 (Pt 7 July 2013), pp. 1204–1214. DOI: 10.1107/S0907444913000061.

[11] A. J. McCoy et al. Phaser crystallographic software. In: *Journal of Applied Crystallography* 40 (Pt 4 Aug. 1, 2007), pp. 658–674. DOI: 10.1107/S0021889807021206.

[12] P. D. Adams et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. In: *Acta Crystallographica Section D: Biological Crystallography* 66.2 (Feb. 1, 2010), pp. 213–221. DOI: 10.1107/S0907444909052925.

[13] J. Jumper et al. Highly accurate protein structure prediction with AlphaFold. In: *Nature* 596.7873 (2021), pp. 583–589. DOI: 10.1038/s41586-021-03819-2.

[14] G. N. Murshudov, A. A. Vagin, and E. J. Dodson. Refinement of Macromolecular Structures by the Maximum-Likelihood Method. In: *Acta Crystallographica Section D: Biological Crystallography* 53.3 (May 1, 1997), pp. 240–255. DOI: 10.1107/S0907444996012255.

[15] L. Potterton et al. CCP4i2: the new graphical user interface to the CCP4 program suite. In: *Acta Crystallographica. Section D, Structural Biology* 74 (Pt 2 Feb. 1, 2018), pp. 68–84. DOI: 10.1107/S2059798317016035.

[16] T. C. Terwilliger et al. Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. In: *Acta Crystallographica Section D: Biological Crystallography* 65.6 (June 1, 2009), pp. 582–601. DOI: 10.1107/S0907444909012098.

[17] P. Emsley and K. Cowtan. Coot: model-building tools for molecular graphics. In: *Acta Crystallographica Section D: Biological Crystallography* 60.12 (Dec. 1, 2004), pp. 2126–2132. DOI: 10.1107/S0907444904019158.