Structure-based function assignment for archaellum regulatory network proteins of *Sulfolobus acidocaldarius*

Dissertation

kumulativ

zur Erlangung des Grades eines

Doktor der Naturwissenschaften

(Dr. rer. nat.)

des Fachbereichs Chemie der Philipps-Universität Marburg

vorgelegt von

Lukas Korf

aus Mannheim

Marburg, 2024

Die vorliegende Dissertation wurde von 05/2020 bis 02/2024 am Fachbereich Chemie der Philipps-Universität Marburg unter Leitung von Prof. Dr. Lars-Oliver Essen angefertigt.

Vom Fachbereich Chemie der Philipps-Universität Marburg
(Hochschulkennziffer 1180) als Dissertation angenommen am

_____

Erstgutachter(in): Prof. Dr. Lars-Oliver Essen

Zweitgutachter(in): Prof. Dr. Peter Graumann

Tag der Disputation: _____

Erklärung

Ich erkläre, dass eine Promotion noch an keiner anderen Hochschule als der Philipps-Universität Marburg, Fachbereich Chemie, versucht wurde.

Hiermit versichere ich, dass ich die vorliegende Dissertation

Structure-based function assignment for archaellum regulatory network proteins of *Sulfolobus acidocaldarius*

selbstständig, ohne unerlaubte Hilfe Dritter angefertigt und andere als die in der Dissertation angegebenen Hilfsmittel nicht benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten oder unveröffentlichten Schriften entnommen sind, habe ich als solche kenntlich gemacht. Dritte waren an der inhaltlich-materiellen Erstellung der Dissertation nicht beteiligt; insbesondere habe ich hierfür nicht die Hilfe eines Promotionsberaters in Anspruch genommen. Kein Teil dieser Arbeit ist in einem anderen Promotions- oder Habilitationsverfahren verwendet worden. Mit dem Einsatz von Software zur Erkennung von Plagiaten bin ich einverstanden.

_____ _____

Ort/Datum Unterschrift (Vor- und Nachname)

Erklärung zur kumulativen Dissertation

Ich versichere, dass die im kumulativen Teil der Dissertation mit dem Titel:

Structure-based function assignment for archaellum regulatory network proteins of *Sulfolobus acidocaldarius*

Aufgeführten Anteile der Autoren an den verfassten Publikationen und Manuskripten korrekt und vollständig dargelegt sind.

Der kumulative Teil umfasst die Kapitel: 3.1 und 3.2.

Marburg, den

_____	_____
Unterschrift Doktorand/in	Unterschrift Betreuer/in
Name: Lukas Korf	Name: Prof. Dr. Lars-Oliver Essen

„Success is a journey, not a destination. The doing is often more important than the outcome."

- Arthur Ashe

Danksagung - Acknowledgement

Jede Abschlussarbeit beginnt mit der Zuteilung eines Projekts und Prof. Dr. Lars-Oliver Essen hat mir nicht ein oder zwei, sondern drei unterschiedliche und interessante Projekte ermöglicht, welche ich unter größten Freiheiten bei ihm im Labor bearbeiten durfte. Angefangen von der Bachelor- über die Master- bis hin zur Promotionsarbeit hat mich Prof. Essen bei meinen wissenschaftlichen Arbeiten begleitet, unterstützt, und stets meine wissenschaftlichen Ambitionen gefördert. Daher möchte ich mich für nun fast sieben Jahre Zusammenarbeit, Wachstum und außergewöhnliche Möglichkeiten von ganzem Herzen bedanken.

Bei Prof. Dr. Peter Graumann, der mir nicht nur wiederholt als Zweitgutachter zur Verfügung steht, sondern auch für eine stets interessante Biochemie-Vorlesung und die damit verbundene Förderung meiner biochemischen Ambitionen gesorgt hat, möchte ich mich ebenfalls herzlich bedanken.

I extend my heartfelt thanks to the members of the audit committee, Prof. Olalla Vázquez and Ulrich Tallarek, for their generous and prompt readiness to be part of the board.

I also want to thank all collaboration partners for their great work and support.

Dr. Marian Vogt möchte ich einen besonderen Dank aussprechen für die immense Rolle die er bei der Einarbeitung in das Promotionsprojekt sowie bei meiner persönlichen Weiterentwicklung gespielt hat. Insbesondere möchte ich mich dafür bedanken, dass er mir über die Jahre immer als Berater und auch jetzt noch als Korrekturleser dieser Arbeit zur Verfügung gestanden hat. Eine solch intensive Betreuung und Zusammenarbeit wünsche ich jedem neuem PhD Kandidaten.

Dr. Laura Werel, die auch heute noch zum Erfolg meiner Karriere beiträgt, möchte ich für die Unterstützung bei komplexeren Problemen der Proteinkristallisation, Einweisung in die Remote Control von Synchrotron Messungen, wissenschaftlichen Austausch und nicht so wissenschaftlichen Kaffeepausen herzlich danken.

Petra Gnau, Julia Witsch und Kirsten Bubenheim möchte ich für die Hilfsbereitschaft und die großartige Arbeit und Zuverlässigkeit die sie tagtäglich bereitgestellt haben besonders danken. Ihr seid einfach super!

Dem Rest der Arbeitsgruppe Essen, Assoziierte und Alumni möchte ich für eine aufregende Zeit, gute Kollaboration, feucht-fröhliche (Weihnachts-)Feiern und regen Austausch danken. Die lieben Kollegen Anke Werne, Bastian Pook, Christin Schulz, Christof Taxis, Elisa Seiler, Filipp Bezold, Hagen Nußhär, Johannes Scheffer, Jonathan Trauth, Joshua Rosner, Leonie Schneider, Maisa Alkheder Alahmad, Martin Saft, Max Biermeier, Mohamed Watad, Roberta Spadaccini, Sepp Hepp, Silvana Ruppenthal, und Stefan Petschak werden mir immer eine Erinnerung an diese aufregende Zeit sein.

Vielen Dank an die Mitglieder verschiedener anderer AGs: Alicia Fischer, Christin Schulz, Hannah Schmitz und Alex Kilb, für eine wunderschöne Promotionszeit.

I would also like to thank Po Hsun Wang, aka Nigel, for the incredible trip we shared to Taiwan and Japan. His warm hospitality and friendly nature made the experience truly unforgettable. I will forever cherish the wonderful memories. Next time we officially break the dumpling-eating record, I promise!

Meinen Freunden möchte ich für die wundervollen Momente, aufregenden Abenteuer, absolut blödsinnigen Vorhaben und alle erinnerungswürdigen, und auch die nicht ganz glanzvollen, Momente der gesamten Studienzeit aus tiefsten Herzen danken. Ein ganz besonderer Dank an dieser Stelle an Luca Jung, der einfach meine bessere Hälfte ist und ohne den ich weder eine so erfolgreiche noch so amüsante Zeit gehabt hätte.

Ich möchte mich aus tiefsten Herzen bei meiner Familie bedanken, insbesondere bei den Familienmitgliedern die sich meiner angenommen und unterstützt haben als ich es am meisten gebraucht habe. Opa, Claudia, Rudolf, Sabine, Tobias, Sabrina, Milena, Kai, Kerstin, Dominik und die kleinen Pürzel Maxi, Amelie und Charlotte ich danke euch für all die lieben Worte, die unglaublich schönen Momente, euren Rat und eure ständige Motivation.

Das Beste kommt bekanntlich zum Schluss: Bei meiner Partnerin Tina—und Partnerin ist hier in allen möglichen Auslegungsformen gemeint—möchte ich mich von ganzem Herzen bedanken dafür, dass du dem Wort Partnerin jeden Tag aufs Neue Gewicht gibst, dafür, dass du mich immer unterstützt hast, es nie versäumst mich aufzumuntern, die vielen Sonntagmorgen die wir verquatscht haben und für jedes einzelne Lachen das ich mit dir teilen durfte. Du bist einfach in jeglicher Hinsicht großartig!

Wissenschaftlicher Werdegang

Nicht Teil der Bibliotheksversion

List of Publications

Publications with contribution discussed in this thesis

1.) L. Korf et al., Archaeal GPN-loop GTPases involve a lock-switch-rock mechanism for GTP hydrolysis, *MBio* **14**, e0085923 (2023).

2.) L. Korf et al., Sequential conformational transition of ArnB, an archaeal ortholog with Sec23/Sec24 core motif*,* (2024) (under review in *Nat. commun.*)

Publications and manuscripts with contribution not discussed in this thesis

3.) F. F. Brescia, L. Korf, L.-O. Essen, H. Zorn, M. Rühl, A novel O- and S-methyltransferase from *Pleurotus sapidus* is involved in flavor formation, *J. Agr. Food Chem.*, (2024) (in revision)

4.) M. Maestre-Reyna et al, Visualizing the DNA repair process by a photolyase at atomic resolution, *Science* **382**, eadd7795 (2023)

5.) M. Moldenhauer et al., Parameterization of a single H-bond in Orange Carotenoid Protein by atomic mutation reveals principles of evolutionary design of complex chemical photosystems, *Front. Mol. Biosci*. **10**, (2023)

6.) D. Karrer, E. Wedler, C. Mewe, M. Gand, M. Vogt, L. Korf L.-O. Essen, M. Rühl, A flexible loop of an ene/yne-reductase shapes the substrate tunnel through an open-closed mechanism. Loop design study opens up a pathway for tailored biocatalysts, (manuscript)

7.) P. Schöppner, V. Kalugin, L. Korf, M. Veelders, L.-O. Essen, H.-U. Mösch, Structural basis for cell wall remodeling and cell separation by yeast SUN family proteins, (manuscript)

Zusammenfassung

Der Ursprung des Lebens ist ein komplexes Thema, das weltweit wissenschaftliche Diskussionen befeuert. Im Allgemeinen ist es akzeptiert, dass die Endosymbiose eine zentrale Rolle bei der Eukaryogenese spielt. Für diese Arbeit wurde die *inside-out* Hypothese, welche sich auf die Asgard Archaeen stützt und ein frühes Endomembransystem voraussetzt, als Ausgangslage angenommen. Da Asgard Archaeen noch nicht kultiviert wurden, stellen die TACK Archaeen mit ihrem archaealen Modellorganismus *Sulfolobus acidocaldarius* einen der nächsten, verfügbaren Verwandten der Eukaryoten dar. Interessanterweise wurde das Protein ArnB des *archaellum regulatory networks* (Arn), welches strukturelle Homologien mit dem Sec23/24 Protein des COPII-systems aufweist, in *S. acidocaldarius* Exosomen gefunden. Entsprechend erweist sich ArnB als vielversprechendes Protein, um den prä-endosymbiotischen Transport in Archaeen zu untersuchen.

Diese Arbeit umfasst eine Zusammenfassung der Forschungsergebnisse zu ArnB und seinen assoziierten Proteinen. Hervorzuheben ist dabei die Untersuchung der Interaktion von ArnB mit dem ZnF- und FHA-Domänen enthaltenden ArnA, welche eine sequenzielle, phosphorylierungsabhängige strukturelle Transition mit globalen Änderungen offenbart. Die Untersuchung des ArnAB-Komplexes ergab zum einen neuartigen sequenziellen Phosphorylierungsmechanismus und zeigte die Gegenwart des Sec23/24-Kernmotivs in allen Domänen des Lebens auf. In Zuge dieser Analysen, konnten dem ArnAB-Komplex bzw. dem Sec23/24-Kernmotiv zahlreiche potenzielle Funktionen zugeschrieben werden, die von der Beteiligung am Ubiquitin-Netzwerk bis zur Bildung von COPII-Vesikeln und Assemblierungsrollen reichen. Zudem beinhaltet diese Arbeit die Charakterisierung einer GPN-Loop-GTPase aus *S. acidocaldarius*, *Sa*GPN, welche ebenfalls mit dem Arn-System assoziiert ist und potentiell mit dem ArnAB-Komplex interagiert. GPN-Loop-GTPasen stellen einen eigenen interessanten GTPase-Typ dar, der durch nukleotidunabhängige Homodimere gekennzeichnet ist. Durch die Kristallstrukturen die in dieser Arbeit gelöst wurden, lassen sich mechanistische Unklarheiten im Zusammenhang mit GPN-Loop-GTPasen lösen und ein Lock-Switch-Rock-Mechanismus für archaeale GPN-Loop-GTPasen ableiten.

Zusammenfassend trägt diese Arbeit zum allgemeinen Verständnis der GPN-Loop-GTPasen bei und hebt das ArnAB-Sec23/24-Kernmotiv als eine der häufigsten und funktional vielfältigsten Protein-Faltungen in allen Bereichen des Lebens hervor.

Abstract

The origin of Life is an enigmatic topic that fuels scientific discussion all over the world. In general, it is an accepted concept that endosymbiosis plays a central role within eukaryogenesis. In this context the inside-out theory, which centers on the emerging Asgard Archaea superphylum necessitating an early endomembrane system prior to any endosymbiotic events, holds significant promise as a hypothesis. Since Asgard Archaea had not been cultivatable the sister branch of TACK Archaea with its genetically tractable *Sulfolobus acidocaldarius* is one of the closest, available model organisms preceding the first eukaryotic common ancestor (FECA). Interestingly, the archaellum regulatory network protein ArnB, which shares structural homology with the membrane curvature protein Sec23/24 of COPII vesicles, was abundantly found in *S. acidocaldarius* exosomes. Accordingly, ArnB emerges as a promising research target to study pre-endosymbiotic trafficking in Archaea.

This work encapsulates a culmination of research concerning ArnB and associated proteins. Of particular note is the exploration of ArnB's interaction with its ZnF and FHA domain-containing counterpart, ArnA, revealing a sequential, phosphorylation-dependent structural transition with far-reaching implications. Investigation of the ArnAB complex not only unveiled a novel sequential phosphorylation mechanism but also underscores the ubiquity of the Sec23/24 core motif across all domains of life. Accordingly, numerous potential functions can be attributed to the ArnAB complex or the Sec23/24 core motif, spanning from involvement in the ubiquitin network to COPII vesicle formation and even to other assembly roles. Additionally, this research delves into the characterization of a GPN-loop GTPase from *S. acidocaldarius*, known as *Sa*GPN, which is implicated in the archaellum regulatory network as well and potentially interacts with the ArnAB complex. Despite the relative neglect of GPN-loop GTPases within the P-loop NTPases family, they represent a distinct GTPase type characterized by nucleotide-state independent homodimerization. The *Sa*GPN crystal structures resolved mechanistic ambiguities surrounding GPN-loop GTPases, proposing a lock-switch-rock mechanism for *Sa*GPN and its archaeal counterparts.

In summary, this work advances the field of GPN-loop GTPases and positions the ArnAB-Sec23/24 core motif as one of the most prevalent and functionally diverse protein folds across all domains of life. Consequently, the presented research opens avenues for further exploration and discoveries in the Archaea-to-Eukarya transition.

Content

1.Introduction

1.1 The origin of life

The origin of life on Earth is an enduring scientific debate that has captivated researchers for centuries (1). While there is obviously no doubt about that life existed in early stages of the Earth, there seems to be a huge discrepancy about how vital molecules and later life itself came to existence on earth. Despite the inherent uncertainty, a few broadly agreed-upon points seem to arise from the debate. Earth, is about 4.5 billion years old and was exposed to a tumultuous initial half-billion years with celestial impacts capable of evaporating oceans and sterilizing the surface (2–5). Moreover, well-preserved microfossils resembling modern blue-green algae, dating back approximately 3.5 billion years, suggest early life's emergence (6). Additionally, persuasive evidence even hints at life's presence 3.8 billion years ago right after the big bombardment of the Earth-moon system stopped (7). While there is the possibility that life was transported from an extraterrestrial source to Earth after it became habitable, scientists usually assume a *de novo* beginning of life on our planet (8). In conclusion, these foundational insights into a primordial Earth suggest a myriad of potential pathways for the emergence of life. While certain initially prominent theories, such as those based on the Miller-Urey experiment, are now vigorously debated, others appear more plausible as of today (9–11). Nevertheless, there is one point where all theories about the origin of life converge: regardless of how life originated on Earth, once it appeared, evolution shaped and molded early life into the diverse array of species we recognize today.

1.2 Evolution of phylogenetic trees

Evolution is a crucial process for the persistence of any organism, which is why human curiosity naturally strives to unravel its mysteries not only for solely academic reasons but also to harness it as a biotechnical tool for addressing modern problems (12). However, our understanding of evolution is heavily based on CHARLES DARWIN approach of a universal ancestor in combination with a cell being the smallest biological unit. These theories led to the conclusion, that all life is related to a last universal common ancestor known as LUCA (13, 14). From this primordial organism life

diversified into many different branches, which are classified under the three kingdoms of life: Bacteria, Archaea and Eukarya. Since the classification of Archaea in the 1970s it was assumed that the three kingdoms emerged from one universal cell and therefore there are three primary domains of life (15). Moreover, it was found that Archaea share features with Bacteria, e.g. prokaryotic ribosomes, circular chromosomes, lack of membrane-enclosed organelles, as well as Eukaryotes, e.g. DNA associated with histones, various RNA polymerases, use of methionine in protein synthesis etc. (14). With all these shared features it was discussed, weather Archaea actually present a new domain of life with all three of them sharing the same LUCA which would promote the assumed three-domain phylogenetic tree (16). The other possible option that was early discussed that Eukarya evolved from the TACK (Thaumarchaeota, Aigarchaeota, Crenarchaeota, Korarchaeota) superphylum which would led to a two-domain phylogenetic tree with the oldest LUCA only being direct ancestors to Bacteria and Archaea (17, 18). According to this, in recent studies more and more evidence was presented, that Eukarya actually form a sister branch in the archaeal domain, hence, the current understanding, although a non-unanimous one, of LUCA is that only two primary domains of life directly branched from it (refer to Figure 1) (19–23).

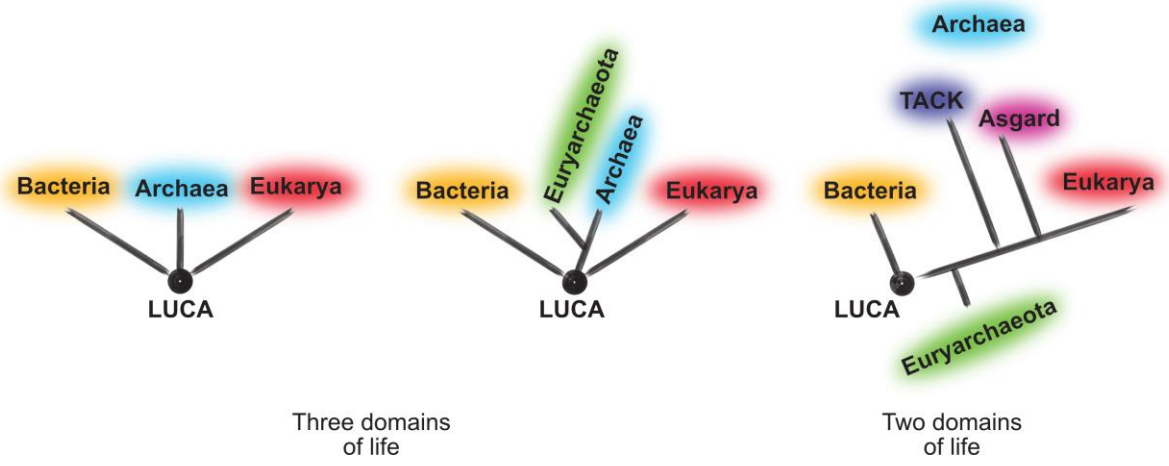

Figure 1: Simplified evolution of phylogenetic trees from three to two domains. The tree of life, as proposed by WOESE, was initially built upon three primary domains of life, incorporating the newly discovered Archaea as a distinct lineage from the Last Universal Common Ancestor (LUCA). As phylogenetics advanced, so did the tree of life. Despite ongoing debates throughout the years regarding the primary domains of life, the widely accepted understanding now presents the tree of life as a two-domain structure, with Archaea forming a sister branch within the eukaryotic lineage. Figure was inspired by Eme et al. 2017.

The ongoing discourse regarding the intricate composition of the tree of life has been spurred in 2015 by the revelation of a newfound class of Archaea originating from within the TACK superphylum, now recognized as Asgard Archaea, which is arguably

the most significant discovery in evolutionary biology since WOESE'S identification of Archaea in 1977 (15, 24). Through diverse genome analyses, Asgard Archaea exhibit numerous components that serve as the foundation for eukaryote-specific genes already inherent in these novel organisms (22, 25, 26). These homologues encompass a range of eukaryotic signature proteins, including ribosomal proteins (e.g. S25e, L14e, S30e), informational proteins (e.g. RNA polymerase A, ELF1, RpoG), cytoskeleton elements (e.g. Actin, Gelsolin, Tubulin), the ubiquitin system (e.g. URM1, E1-E3 proteins), and components of the trafficking machinery (e.g. ESCRT I-III and Sec23/24 orthologs, arrestin-like proteins) (24, 26). Consequently, these distinctive Archaea emerge as formidable candidates for the 'missing link' between Archaea and Eukaryotes, constituting a rapidly expanding superphylum comprising Loki-, Thor-, Odin-, Baldr-, Hel-, Borr-, Heimdallarchaeota, among others (27–31). Noteworthy is the observation that Asgard Archaea seemingly possess a somewhat ‘sluggish’ metabolism, posing challenges for cultivation and further *in vivo* exploration, with cell division occurring at prolonged intervals, spanning weeks (32). Nevertheless, a breakthrough has been recently achieved in cultivating two Asgard Archaea, namely *Prometheoarchaeum synthrophicum* and *Lokiarchaeum ossiferum*, providing unprecedented insights into this superphylum, which was limited before to (meta)genome data (32, 33). Intriguingly, the structural composition of these Asgard cells exhibits a lack of uniformity, presenting a complex and variable surface marked by numerous engulfments. Despite the limited data available on these organisms, it introduces compelling new perspectives to the ongoing discourse on eukaryogenesis—the evolution of eukaryotes from the first to the last eukaryotic common ancestor (FECA to LECA) (26).

1.3 Eukaryogenesis: From FECA to LECA

Eukaryotes are the most intricate and diverse cellular entities, the origins of which can be traced back billions of years to the base of the phylogenetic tree. Although it is now generally accepted that Archaea and Eukaryotes are evolutionarily closely related and that Eukaryotes have probably descended from Archaea, the adoption of bacterial traits in Eukaryotes is a more intricate process (26). Phylogenetic trees generally show that Bacteria split off into a separate branch early in evolutionary history (Figure 1); however, Eukaryotes share up to 67 % of their genome with Bacteria (34). These

attributes include the ability to thrive even on suboptimal or harmful nutrition, parasitize other eukaryotic organisms, be resilient under various environmental conditions characterized by extreme temperatures, acidity, anaerobiosis or toxicity, and employ defense mechanisms against potential threats from other organisms (35). Some of these features may be explained by the process of horizontal gene transfer (HGT), which involves the transfer of genetic information without the necessity of parentage (non-sexual), making it difficult to understand the phylogenetics of eukaryotic organisms (36, 37). Nevertheless, HGT is a well-documented process today and involves many different mechanisms including conjugation, transduction, transformation and others. This leads to an array of genes being transferred, not autochthonously duplicated and evolving in new hosts without productional gene transfer over the course of evolutionary history (38, 39). Therefore, sometimes the two-domain tree of life is, arguably more fittingly, referred to as a web of life (40).

However, HGT alone does not account for all the bacterial similarities Eukaryotes share with their prokaryotic distant relatives. It is universally accepted that the emergence of eukaryotes is related to a process called endosymbiosis, where a prokaryotic host cell took up another one without digesting it and forming a symbiotic bond, pathing the way for the last eukaryotic common ancestor (LECA) (41). The most prominent examples of endosymbiosis are surely the emergence of mitochondria and chloroplasts where one of the endosymbionts was an α-proteobacterium or a cyanobacterium, respectively. While endosymbiosis is recognized as a key process in eukaryogenesis, aspects like the timing, the cellular nature of endosymbionts, the complexity of cells prior to endosymbiosis and the quantity of these events are controversial (42–44). Naturally, this leads to many hypotheses for endosymbiosis or eukaryogenesis in general with each of them focusing on different observations.

For example, a model based on the first cultured Asgard Archaeon called the Entangle-Engulf-Endogenize (E^3) model involves the interesting membrane shapes and engulfments observed to obtain an endosymbiont (33). Moreover, the inside-out theory focuses on the autonomy of nuclei and cell organelles shaping from the host around the future nucleus in so called blebs from a nucleus starting point to full eukaryotic cell (inside-out) (45). Others focus more on the bacterial similarities of Eukaryotes like the syntrophy hypothesis with a δ-proteobacterial host and two endosymbiotic events for the formation of a nucleus and mitochondria respectively (43). Then there is the model of serial endosymbiosis involving multiple events explaining the acquisition of big

amounts of heterologous gene material (46, 47). While there are more and less convincing theories with supportive data, it is difficult to compare and find similarities, which might be the first step in discovery of an ultimate eukaryogenesis theory. In an attempt to analyze and extract common ground between them, a recently published framework based on phylogenetics and the phylogenetic definition of Eukaryotes categorizes three groups of endosymbiotic theories (42). Based on this framework the hydrogen hypothesis (48), the phagocytizing archaeon (49), the reverse flow (50), the inside-out and the E^3 model represent group one with the same two archaeal and α-proteobacterium stem-linages (42) (Figure 2, left). In contrast, the syntrophy hypothesis involves an additional lineage derived from δ-proteobacteria, hence, is classified as group two (Figure 2, middle). Group three involves the serial endosymbiosis model with a multitude of endosymbiotic events, which have yet to be determined (Figure 2, right). Notably, the E^3 and inside-out hypotheses are phylogenetically indistinguishable while also sharing the same prerequisite of an

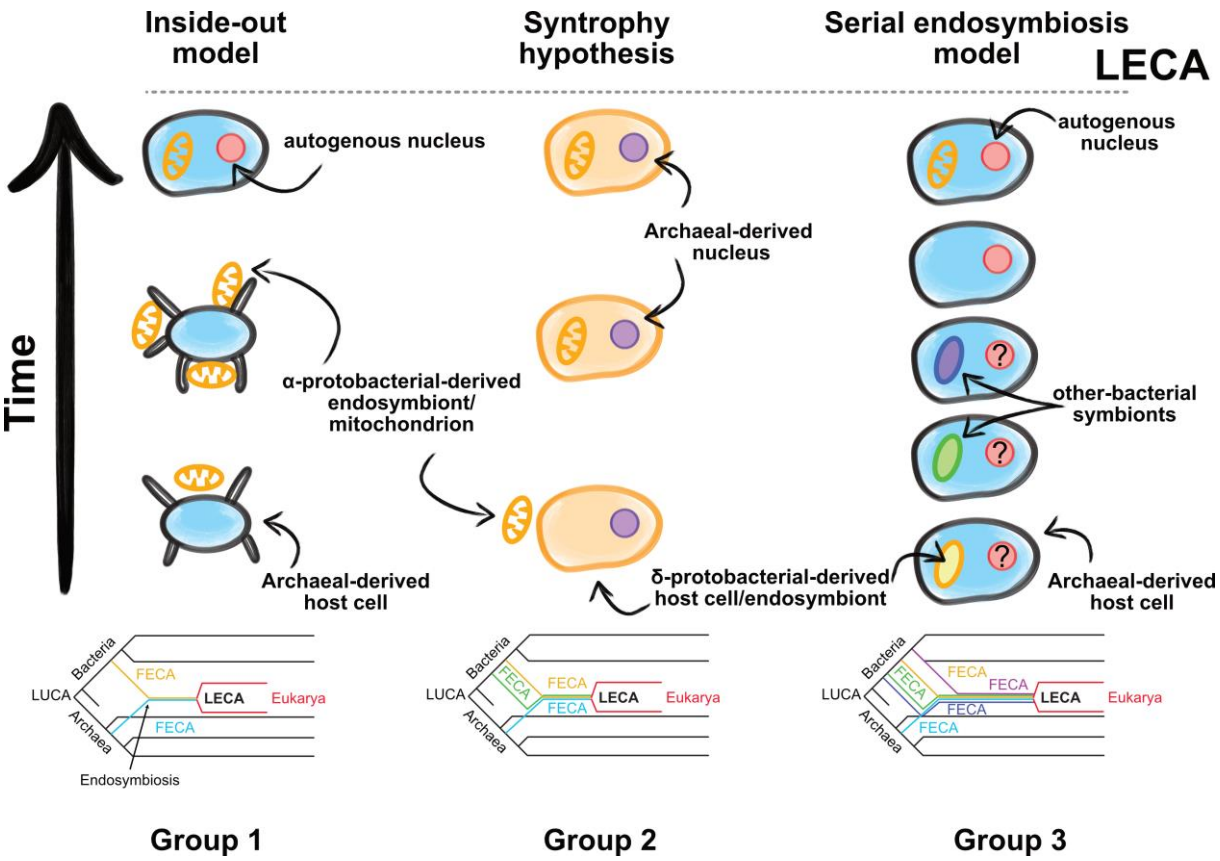

Figure 2: Eukaryogenesis hypotheses grouped based on phylogenetically distinguishability. Group one involves one endosymbiotic event until the emergence of LECA, as displayed in the corresponding phylogenetic tree. This group consists of the hydrogen hypotheses, reverse flow, phagozytose-archaeon E^3 and inside-out model. The second group involves an additional stem lineage and therefore an additional endosymbiotic event. Group three consists of the serial endosymbiosis model which involves multiple endosymbiotic events before the emergence of LECA. The figure is inspired by Donoghue et al. 2023.

intricate cytoskeleton and early endomembrane system predating the endosymbiotic event (42).
Overall, the emergence of an endomembrane system is quite relevant in the context of eukaryogenesis and particularly for this work, which is why the next section focuses on some fundamentals of archaeal and eukaryotic lipids.

1.4 Archaeal lipids and the lipid divide

As elaborated above, the emergence of eukaryotes involves in most cases an (Asgard) archaeal host, with the exception of the syntrophy model, which leads to the intrinsic problem known as the lipid divide (51). Lipids present essential building blocks in all organisms with their most abundant occurrence in cell membranes, which are the barriers that separate cells from their environment. While Eukaryotes share a lipid-double layer that is similar to the bacterial membrane, Archaea have lipids that are more suited for their extreme habitats including high temperature (>140°C), extreme pHs (0-11), high pressure (>500 atm) and high salinity (<300 g/L NaCl) (52, 53). The main differences between archaeal and bacterial lipids involve the different stereochemistry (*sn*-1 instead of *sn*-3) of their glycerol backbones and the preference of the archaea for ether-linked isoprenoid-based alkyl chains over ester-linked fatty acyl chains in the formation of their hydrophobic components (52).
Given spatial separation membrane functions also include signaling, immunology, metabolism and various trafficking mechanisms (54–56). Especially the latter requires eukaryotic and bacterial organisms to be able to adjust lipid fluidity against environmental factors to stay both rigid and flexible. Interestingly, no hints for membrane manipulating factors known from Eukaryotes (e.g. cholesterol) were found in Archaea, although other substances like carotenoids, polyprenols, quinones and apolar polyisoprenoids are discussed to exert these functions in Archaea (57–60). In consideration of all these factors it seems only natural that archaeal lipids differ quite heavily from their eukaryotic descendants with comparably moderate habitats. In detail this means that the backbone of archaeal membrane lipids is usually an *sn*-glycerol 1-phosphat (G1P) in contrast to eukaryotic lipids having an *sn*-glycerol 3-phosphat (G3P) scaffold. Furthermore, onto these glycerol phosphate backbones Archaea link hydrophobic isoprenoid lipid tails over ether-bonds, whereas Eukaryotes link fatty acyl chains via ester-bonds (61). Archaeal isoprene lipids, also called archaeol, are also

known to form macrocycles as well as tetraethers as tail-to-tail linked glycerol-dialkyl-glyceroldiether lipids with an additional glycerol known as caldarchaeol (62, 63). Therefore, Archaea are even able to have membranes that are actually monolayers instead of the usually known bilayer from Bacteria or Eukarya. Moreover, some archaeols also include multiple cyclic features like cyclopentane to further stabilize against high temperature (52). In addition, archaeal lipids are also more stable against enzymatic digestion by lipases and have larger surface area per lipid than their eukaryotic counterparts (62). As a result of all these specific lipids, archaeal membranes feature a higher viscosity density of packing, overall stabilities as well as lower chemical reactiveness and permeability to water and ions, matching the conditions of an archaeal lifestyle (62, 64–67).

Considering all these differences between archaeal and eukaryotic/bacterial lipids it is a rather intricate problem that eukaryotes came to have bacterial-like membranes, given that most endosymbiosis models involve an archaeal host. This topic is still part of an engaging discussion with theories ranging from hybrid membranes, proton flow advantages of the ester carbonyl leading to energy efficiency or regioselectivity advantage models (68–70). Interestingly, an archaeal-bacterial hybrid membrane theory seems to be promising due to recent findings of bacterial lipids and their corresponding synthase in archaeal organisms and vice versa (51, 61, 70–72). While these discoveries challenge the previously held notion that Archaea exclusively employ isoprenoid lipids, they provide only a subtle indication toward establishing well-founded theories addressing the questions of why, how, and when the lipid divide occurred. The capacity of archaea to synthesize both ester and ether lipids could lend support to an inside-out theory, suggesting that the engulfment of the endosymbiont culminated in a bypass of the bacterial membrane, ultimately giving rise to LECA. However, it's important to note that for this scenario to unfold, a vesicle or rudimentary trafficking system would be imperative, heavily relying on membrane curvature proteins (45).

1.5 Trafficking machinery in Eukaryotes

To comprehend the trafficking machinery in Archaea, a prudent approach might involve an examination of the repertoire of eukaryotic transporting systems. Such a retrospective view could unravel early endomembrane features. Therefore, this segment focuses on eukaryotic vesicle formation systems and their key characteristics.

The endomembrane system encompasses both coat protein complexes I and II (COPI & COPII), clathrin-coated vesicles, and the ESCRT system, capable of cleaving vesicles from the membrane without enveloping them in a protein coat. These vesicle formation systems play pivotal roles in a myriad of cellular processes, including communication between the Golgi apparatus and endoplasmic reticulum, exo-/endocytosis, viral budding, trans-nuclear transportation, and others (73, 74). Consequently, the formation, organization, and cargo vary among the distinct types of vesicles.

1.5.1 Coat protein vesicle assembly (clathrin, COP I&II)

Protein-coated vesicles play a diverse role in various trafficking processes within cellular organization. As implied by the name 'coat protein complex,' these vesicles are enveloped by an assembly of proteins. Accordingly, these complexes consist of multiple subunits with variations in both their composition and the process through which they recruit to the membrane. The discovery of coated vesicles began with clathrin-coated vesicles (CCVs), characterized by their distinctive football-like shape composed of hexagons and pentagons. These lattices are constructed from three clathrin subunits forming a triskelion-like shape (Figure 3, left). Each triskelion includes multiple repeats, constituting the C-terminal 'leg' or heavy chain, which is bifurcated into a proximal and a distal leg by the 'knee.' The middle section features an all-helical 'ankle,' and the N-terminal region forms a seven-blade β-propeller 'foot.' (75).

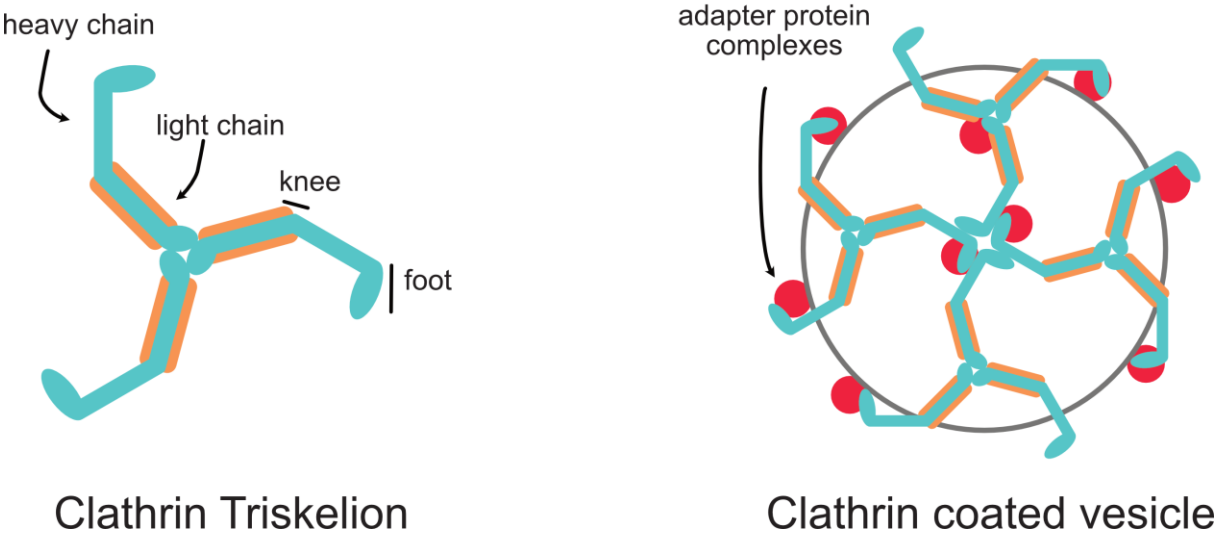

Figure 3: Clathrin coat structure and assembled vesicle structure. The left shows the clathrin triskelion with labels of the different regions assembling as a trimer. The right shows an assembled coatomer with adaptor proteins responsible for membrane association and tetrameric clathrin forming conjunctions. The Figure was inspired by https://www.mbi.nus.edu.sg/mbinfo/what-is-clathrin-mediated-endocytosis, accessed at 31.01.2024.

Moreover, the C-termini of the three heavy chains engage in interaction and are encircled by three light chains arranged in a triangular orientation (76, 77). To construct the clathrin coat, four clathrin triskelions interact, forming one edge of the fullerene-like structure connected through an adapter protein complex, AP2 (Figure 3, right) (78). However, clathrin itself is incapable of directly binding to the membrane and requires various adapter protein complexes, such as the aforementioned AP2 complex, to associate with the membrane and effectively form a cargo-loaded vesicle (79). Alongside the membrane recruitment of clathrin subunits and the linking of clathrin tetramers, the AP2 complex, comprising multiple subunits (α, β, γ, σ, and μ), also plays a crucial role in cargo recognition (77, 80). Given the diversity of cargoes for various vesicle types, specific recognition adaptors are essential for clathrin-coated vesicles (CCVs). Examples include β-arrestins 1 and 2, epsins, or CALM (clathrin assembly lymphoid myeloid leukemia protein) (81, 82). Furthermore, more than 30 auxiliary proteins associated with CCVs are known, many of which are presumed to aid in cargo identification (80). This intricate network results in numerous possibilities for cargo recognition in CCVs, aligning seamlessly with their primary function in endocytosis.

COPII vesicles, in contrast, consist of smaller protein subunits that are considerably smaller than clathrin and lack a filamentous nature. The coat protein II (COPII) complex comprises only a few proteins, including the two heterodimeric complexes Sec23/24p and Sec13/31p, a Sar1 GTPase, and the guanosine exchange factor (GEF) Sec12, which is also responsible for recruiting Sar1 to the membrane (83, 84). Subsequent to Sar1's recruitment to the membrane, Sec23/24 complexes are recruited themselves by Sar1, where Sec24 serves as the cargo anchor, and Sec23 acts as a guanosine activation protein (GAP) (85). The recruitment of Sec13/31 tetrameric subunits, forming an outer layer, follows the acquisition of this first inner coat layer, and this process is also mediated by Sar1, although the precise nature of the GTPase activity in all these processes remains a subject of debate (83, 86). Due to the GTPase-dependent nature of vesicle formation, with both the GAP and the GEF in close proximity, COPII vesicles are considered relatively unstable, as the nucleotide-bound state of Sar1 affects both assembly and disassembly (87). However, COPII vesicles exhibit flexibility, allowing for different shapes and sizes, which is ideal for a vesicle transportation system required to accommodate diverse and potentially bulky cargo, ranging up to 300 nm, from the endoplasmic reticulum (ER) to the Golgi apparatus (85, 86).

COPI vesicles are involved in both inter-Golgi transport and Golgi apparatus-to-ER

trafficking. The coatomer is composed of various subunits that form a heteroheptameric complex, including the α-, β-, β′-, γ-, δ-, ϵ-, and ζ-COP subunits (88). In addition to these coat proteins, a small GTPase is necessary to manage all the subunits. However, in this instance, it is not a Ras-like GTPase like Sar1 but rather an Arf1-like GTPase. Arf1 is recruited to the membrane with the assistance of its corresponding guanine nucleotide exchange factor (GEF) and subsequently recruits the entire COPI complex *en bloc*, leading to membrane scission followed by uncoating (89, 90). Similar to the role of the Sar1 GTPase in COPII vesicles, Arf1 plays a pivotal role in various assembly and disassembly processes of the COPI system, with the latter being particularly important, as cargo can only be released from an uncoated vesicle (89). Intriguingly, after the complex is assembled, it remains stable for many hours in the cytosol, distinguishing COPI vesicles from the COPII complex vesicles (91).

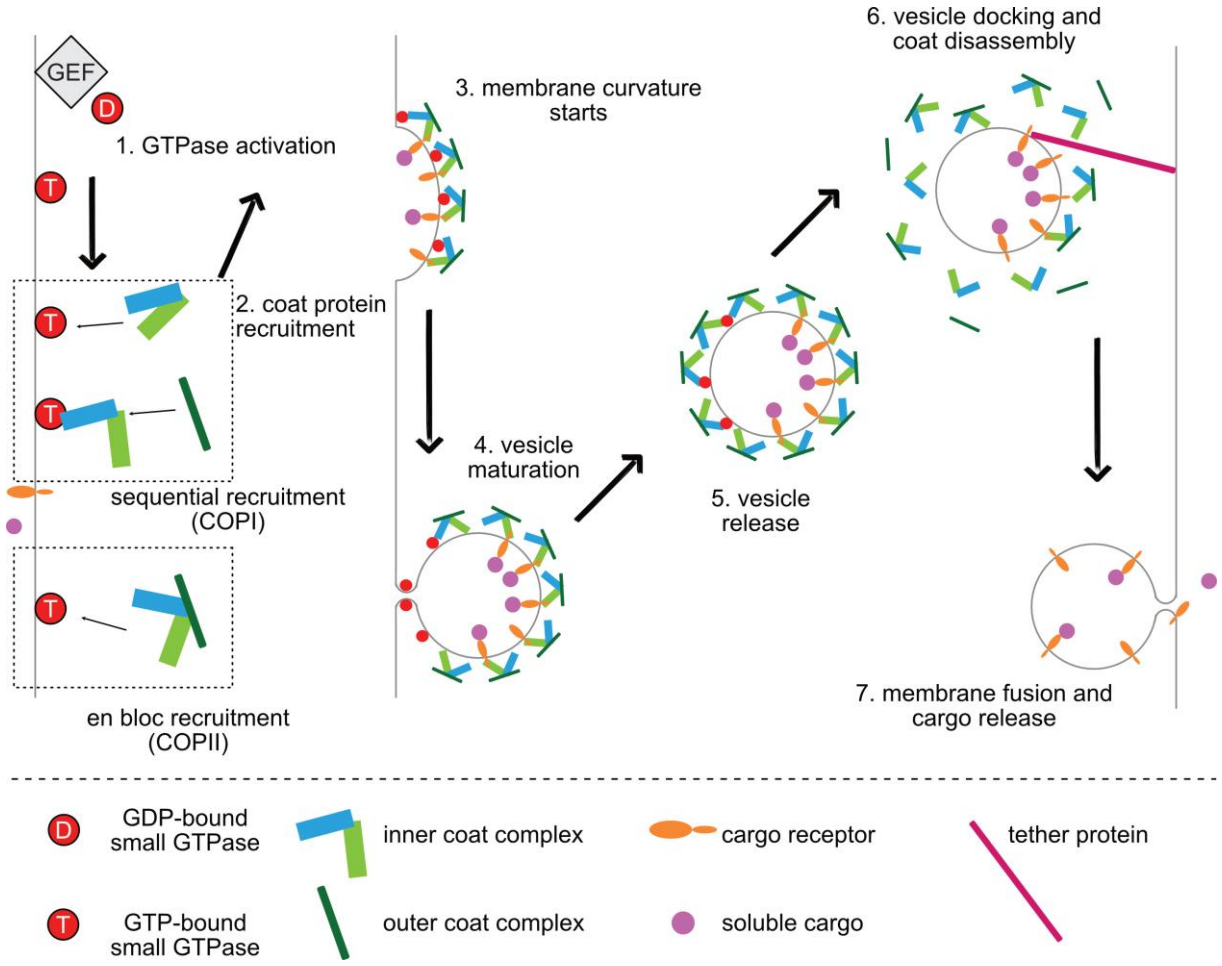

Figure 4: General coat protein assembly. After the GTPase is recruited to the membrane and activated (1.) the coat complexes are recruited in a COPI/COPII specific way (2.) following membrane curvature around loaded cargo receptors (3.) until full maturation (step 4.) and subsequent vesicle release from the membrane (5.). A tether protein mediates coat disassembly (6.) so the vesicle can fuse with the membrane to release its cargo (7.). The Figure was inspired by Béthune et al. 2018.

In general, the assembly of coats can be summarized as follows: a GTPase is recruited to the membrane by its guanine nucleotide exchange factor (GEF), leading to a nucleotide switch, followed by the recruitment of the inner coat layer and subsequently the outer coat, which can also occur simultaneously in the case of COPI (Figure 4). Cargo sorting is already facilitated by the respective receptors (such as AP2, Sec23, etc.), and the membrane begins to curve until complete scission is achieved. Subsequently, coats are disassembled through GTP hydrolysis, according to the current understanding of the GTPase function, followed by cargo unloading at the destination due to membrane refusion (88). This implies that despite differences in appearance, cargo, routes, and destination of coated vesicles, the overall mechanism is similar, potentially indicating a universal origin. Furthermore, it is known that COPI and clathrin adapter proteins share structural homology, further supporting an evolutionary relationship (92). However, COPII Sec23/24 does not share this homology, suggesting an independent origin. This difference is also reflected in the distinct recruiting processes (*en bloc* vs. sequential) and has sparked an open discussion about the origin of coated trafficking machineries (93).

1.5.2 The ESCRT system

As mentioned earlier, coat proteins constitute a diverse and plentiful group of proteins crucial for vesicle formation. However, an alternative trafficking machinery is capable of generating vesicles without the need for coating proteins. This machinery is known as ESCRT (endosomal sorting complex required for transport), and it can generate lipid-only vesicles without the support of a structural scaffold. The ESCRT pathway plays a key role in the biogenesis of multi vesicular bodies (MVBs), transitioning from early endosomes to eventually mature into lysosomes or vacuoles (94). Furthermore, ESCRT proteins play crucial roles in viral budding and cytokinetic abscission, showcasing the versatility of this cargo-recognition and membrane-sculpting system (95). Comprising five main complexes, ESCRT-0, -I, -II, -III, and the ESCRT-associated Alix/Bro1 complex, this system is responsible for transporting and sorting ubiquitinylated cargo, involving over 30 proteins in total (96–99). The AAA ATPase Vps4 contributes to regulating membrane scission, effectively separating the assembled ESCRT complexes and releasing the vesicle (99). To initiate complex assembly, the different ESCRT complexes are sequentially recruited to the membrane. ESCRT-0, a hetero complex featuring two anti-parallel coiled coil-interacting GAT

domains (Vps27/Hse1), is the first to be recruited upon detection of ubiquitin or phosphatidylinositol 3-phosphate (100). Subsequently, the weakly membrane-binding ESCRT-I, a heterotetrameric complex (Vps23, -28, -37, Mvb12) with a 20 nm rod-like structure, associates with the ESCRT-0 complex, forming a 12-membered ring structure (95, 101, 102). At the other end, ESCRT-I connects with the Y-shaped ESCRT-II heterotetramer (Vps36, -22, -25), where two subunits (Vps25) form a homodimer (95). Thereby, the GLUE domain (GRAM-like ubiquitin-binding in EAP45) of Vps36 in ESCRT-II enables recognition of phosphatidylinositol 3-phosphate and ubiquitinylated cargo (103, 104). Additionally, ESCRT-II plays a crucial role in recruiting the heterotetrameric ESCRT-III (Vps2, -20, -24, Snf7), which, unlike the other ESCRT complexes, does not form a stable complex (95). Subsequently, the binding of Vps25 to Vps20 triggers the full assembly of ESCRT-III filaments, forming spirals that reshape the membrane to create the vesicle structure (95). Once assembled, the ESCRT-III complex cannot dissociate from the membrane, necessitating the action of Vps4. Working in conjunction with the ESCRT-III complex, Vps4 exerts ATP-dependent force on the premature vesicle, leading to reverse-topology membrane scission (Figure 5) (105). Vps4 is also responsible in the disassembly of ESCRT-III proteins as well as their recycling (106). These vesicles are subsequently released into the lysosome lumen, forming intralumenal vesicles (ILVs) or, in the case of MVBs membrane refusion, into the extracellular space, essentially becoming exosomes or extracellular vesicles (EVs).

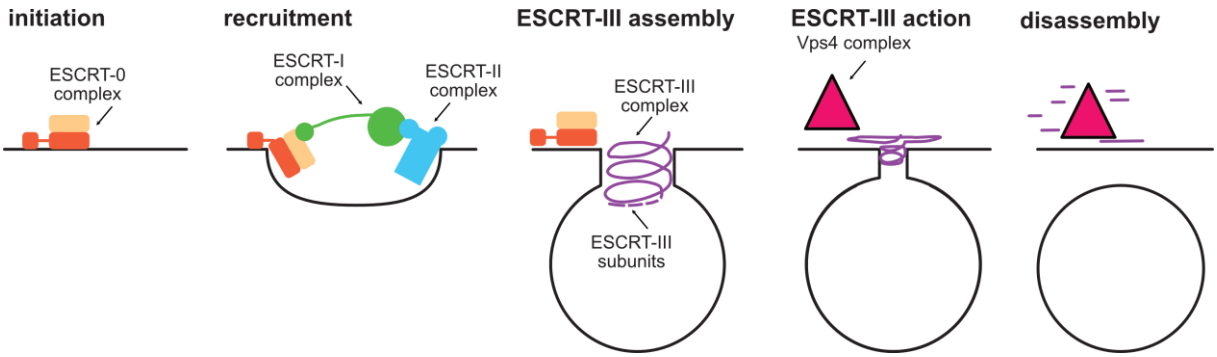

Figure 5: ESCRT system vesicle assembly. Vesicle formation is initiated by the ESCRT-0 hetero complex, followed by ESCRT-I, which in turn binds to the ESCRT-II complex at its other end. ESCRT-III subunits are then assembled into a filamentous complex until membrane scission is achieved, mediated by ESCRT-II. Release from the membrane and disassembly of the stable ESCRT-III filament is orchestrated by Vps4. Figure is inspired by Henne et al. 2011.

The ESCRT system and coat protein complexes equip eukaryotic cells with potent mechanisms to shape and remodel membranes, enabling precise trafficking and maintaining cellular homeostasis. Nonetheless, in the context of an early endomembrane system, whether pre- or post-endosymbiotic, it is plausible that these

systems were not as sophisticated. This raises the intriguing question of how LECA managed to facilitate trafficking through vesicular cargo transport.

1.6 What do we know about archaeal trafficking?

LECA could have originated in various ways, as discussed in the preceding sections. According to the inside-out theory, it must have had a mechanism for facilitating trafficking over membranes, possibly involving an archaeal host assumed to belong to the Asgard superphylum (45, 107). Therefore, investigating trafficking machineries from LACA (last archaeal common ancestor) to FECA (first eukaryotic common ancestor) is of interest. Metagenome analysis of Asgard Archaea revealed many trafficking-related or homologous proteins known from eukaryotes, such as ESCRT-I-III complexes, Sec23/24 homologs, WD40 domains, notably the clathrin terminal 'foot' domain consists of WD40 repeats, TRAPP, and arrestin domains (25, 75). Among these, the ESCRT machinery seems to be commonly spread in the Asgard superphylum, while coat-related homologs (Sec23/24) and clathrin homologs (WD40 repeats) were found only in Thorarchaeota (24, 25). Though members of the Asgard Archaea may possess coat-like systems, unraveling their function and pathways is challenging, especially since few to none Asgard Archaea, particularly no Thorarchaeota, have been cultivated (108).

However, the TACK superphylum is the closest relative of the Archaea domain to the Asgard sister phylum (25). Therefore, a TACK Archaeon could be a suitable candidate to explore an early endomembrane system, if it existed before endosymbiosis. Although there are no sequence-based results suggesting the origin of coat systems in TACK Archaea, other membrane modeling proteins, including the cytokinesis-involved GTPase FtsZ and ESCRT-III complex homologs, have been identified (109). Interestingly, research has demonstrated that ESCRT-III homologs play a substantial role in cell division, even in a species lacking FtsZ or related genes (109, 110). This observation raises the possibility of two distinct executions of cytokinesis. Considering their structural similarity, despite differences in activation mechanisms, it underscores the presence of related systems in Archaea (111). Notably, the TACK Archaeon genus *Sulfolobus* is known to form exogenic vesicles containing ESCRT-III proteins, Vps4, and a solid proteinaceous surface layer (S-layer) (112). Recent findings indicate that extracellular vesicle (EV) formation in these *Sulfolobus* species is ESCRT-dependent,

laying the foundation for an early endomembrane system (113). Additionally, some Asgard Archaea use the ESCRT system not only for EV formation but also for sorting based on ubiquitin recognition, similar to eukaryotes (114). However, despite annotation for Thorarchaeota, the presence of coat protein structures or related proteins could not be validated (24, 115, 116). Hence, the existence of coat complexes as trafficking systems in Archaea remains uncertain, whether due to challenges in identifying them from sequence data, potential annotation inaccuracies—as suggested by Ferrelli et al.'s extensive structural analysis—or the possibility that coats are simply not integral to archaeal trafficking. Regardless, delving into the in-depth exploration of Asgard archaeal trafficking machinery (24) holds the promise of illuminating this intriguing area. Unfortunately, current limitations in cultivating Asgard Archaea restrict studies to laborious *in vitro* approaches. On the other hand, TACK Archaea, being close relatives to the Asgard superphylum, boast ESCRT-related proteins and exhibit extracellular vesicle (EV) formation capabilities. Among TACK archaea, *Sulfolobus acidocaldarius*, a member of the Crenarchaeon phylum, emerges as a prime candidate for investigating a potential early endomembrane system, regardless of its specific characteristics.

1.7 *Sulfolobus acidocaldarius* - a model organism

Sulfolobus acidocaldarius is a hyperthermophile, acidophile (pH 2-3) Crenarchaeon, now also known as Thermoproteota, and was found in hot springs like the Locomotive Spring (Yellowstone National Park, *USA*) (117). The *Sulfolobus acidocaldarius* genome is fully sequenced, with a redundancy of ~6-fold sequence coverage, comprising 2,225,959 bp resulting in around 2300 gene products (118). It can thrive on various carbon sources including a wide array of sugars and amino acids and is also capable of aerobically oxidizing sulfur species to sulfuric acid (118). Moreover, *S. acidocaldarius* stands out for its relatively high division rates among Archaea, with duplication times ranging from 6.5-8 hours (with yeast extract) to 37-55 hours (elemental sulfur), depending on the available nutrition (119). The species also exhibits an intriguing UV-induced DNA damage repair mechanism. This process involves cellular aggregation facilitated by pili structures, leading to DNA exchange for homologous recombination repair between distinct cells (120). Additionally, *Sulfolobales* species, including *S. acidocaldarius*, demonstrate the ability to produce

toxins that inhibit the growth of other *Sulfolobus* species in their perimeter. This mechanism serves to eliminate unwanted competition among different Sulfolobales during DNA damage responses (121, 122). The release of toxins in *S. acidocaldarius* is thought to be facilitated through vesicular release (121). Consequently, this highlights the remarkable capacity of this archaeal organism to utilize extracellular vesicles (EVs) for purposes extending beyond the mere disposal of unwanted materials. Notably, *S. acidocaldarius* has been found to harbor proteins associated with the ESCRT machinery, strongly indicating their role in the formation of extracellular vesicles (112, 113, 123). The organism's close relationship to Asgard Archaea, its thermophilic nature, proficiency in EV formation, rapid growth rates, robust response to UV-induced DNA damage, comprehensive genome sequencing, diverse metabolic pathways, and the presence of ESCRT-related proteins collectively position it as an ideal model organism. *S. acidocaldarius* is not only valuable for studies on DNA damage response and DNA replication but also holds promise for exploring potential remnants of an early endomembrane system. The availability of both *in vitro* and *in vivo* experimentation further enhances its suitability for detailed analyses in these research areas. (118).

Interestingly, exosomes derived from *S. acidocaldarius* were discovered not only to contain ESCRT-related proteins but also to exhibit a high abundance of a protein known as ArnB (Saci_1211) (112). ArnB serves as a regulatory protein intricately involved in the archaellum regulatory network (Arn) (124). The archaellum, functioning as the motility structure in Archaea, bears a resemblance to the flagellum in Bacteria. Moreover, the regulatory mechanism governing the archaellum is notably complex, with certain pathways remaining not fully understood to date.

1.8 The Archaellum Regulatory Network

Many organisms utilize motility structures to facilitate movement, responding to various signals such as chemical cues (chemotaxis), adhesion sites (haptotaxis), and temperature (thermotaxis) or for general relocation (125). While Bacteria typically have a flagellum for this purpose, Archaea possess a similar structure known as the archaellum, resembling a type IV pilus (126). *Sulfolobus acidocaldarius*, like many other archaeal species, can assemble the archaellum. However, the archaellum's role in motility for Crenarchaeota is undirected, unlike chemotactic processes, and it is

activated in response to stress, often linked to nutrient depletion, helping the archaeon seek more favorable living conditions when necessary (127). This stress usually involves nutrient depletion and allows the Archaeon to find more suitable living conditions when necessary (128). Regardless, the assembly and disassembly of the archaellum must be carefully regulated to ensure its use only when required. The proteins responsible for this regulatory mechanism are collectively known as Arn (128). Genes governing the formation of the archaellum fall under the regulation of an operon known as *arl* (formerly *fla*), encompassing seven genes in the case of S. acidocaldarius (arlB, arlX, arlG, arlF, arlH, arlI, and arlJ) (129, 130). The archaellum primarily consists of archaellin, a major structural filamentous protein represented by ArlB, featuring a signal peptide cleaved by PibD before its integration into the filamentous structure. Additionally, a motor complex is formed by ArlI, ArlJ, ArlX, and ArlH (130). ArlI acts as a crucial ATPase involved in both archaellum assembly and rotation. It interacts with the transmembrane platform protein ArlJ and the RecA family ATP-binding protein ArlH, surrounded by ArlX, thought to function as a scaffold guiding the assembly of the archaellum motor complex (127, 128, 130). The pseudo-periplasm-spanning proteins ArlF and ArlG together create a stator complex against which ArlB rotates, driven by ATP and facilitated by ArlI (131, 132). The *arl* operon, commencing with the archaellin gene *arlB*, is flanked by two regulatory paralogs, *arnR* positioned upstream of the operon, and *arnR1* situated downstream of the archaellum genes at the operon's end (128). Collectively, these proteins form various heteromultimeric complexes, each featuring an identical domain arrangement. This configuration includes an N-terminal helix-turn-helix motif, a HAMP (histidine kinases, adenyl cyclases, methyl-accepting proteins, and phosphatase) domain, a sensor domain believed to be associated with starvation detection, and two C-terminal membrane helices (128, 133). Notably, these regulators have been demonstrated to target the promoter of the archaellum genes, robustly inducing the expression of *arlB* in response to nutrient depletion (133). Furthermore, both ArnR and ArnR1 are thought to undergo regulation through phosphorylation by the Ser/Thr kinase ArnC (134). ArnC assumes a pivotal role in the archaellum regulatory network, orchestrating numerous phosphorylation-dependent regulations (133, 134). Its regulatory function extends to the archaeal biofilm regulator AbfR1, a confirmed positive regulator of *arlB*. Additionally, under nutrient-rich conditions, ArnC can phosphorylate ArnB in the presence of its partner ArnA.

ArnB is a van Willebrand factor type A (vWA) containing protein that interacts with

ArnA, a fork-head associated (FHA) domain containing protein via the zinc-finger (ZnF) domain of ArnA (124). Notably, FHA domains are recognized for their interaction with phosphorylated threonine residues, a feature abundantly present in ArnB's C-terminal helix domain (124). Consequently, upon phosphorylation-dependent interaction, ArnB and ArnA form a robust heterodimeric complex, resulting in a suppressed expression of *arlB* (124, 134, 135). Conversely, dephosphorylation of the ArnAB complex, facilitated by the phosphatase PP2A, induces *arlB* expression, leading to the assembly of archaellum structures (124, 134). Furthermore, the regulatory role of the complex is underscored by the hypermotile phenotype resulting from the deletion of either *arnA* or *arnB*. Another kinase, ArnD, has also the capability to phosphorylate ArnB, although the specific context of this relationship remains unclear, except for the observation that an *arnD* knockout induces hypermotility, whereas an *arnC* knockout results in diminished motility in *S. acidocaldarius* (136). To decipher the intricacies of this phosphorylation-dependent regulatory pathway involving the kinases ArnC, and potentially other kinases, along with the PP2A phosphatase, a co-immunoprecipitation (co-IP) assay was undertaken (134). This assay unveiled two additional targets of PP2A, the universal stress protein with the gene name Saci_0887 and the GPN-loop

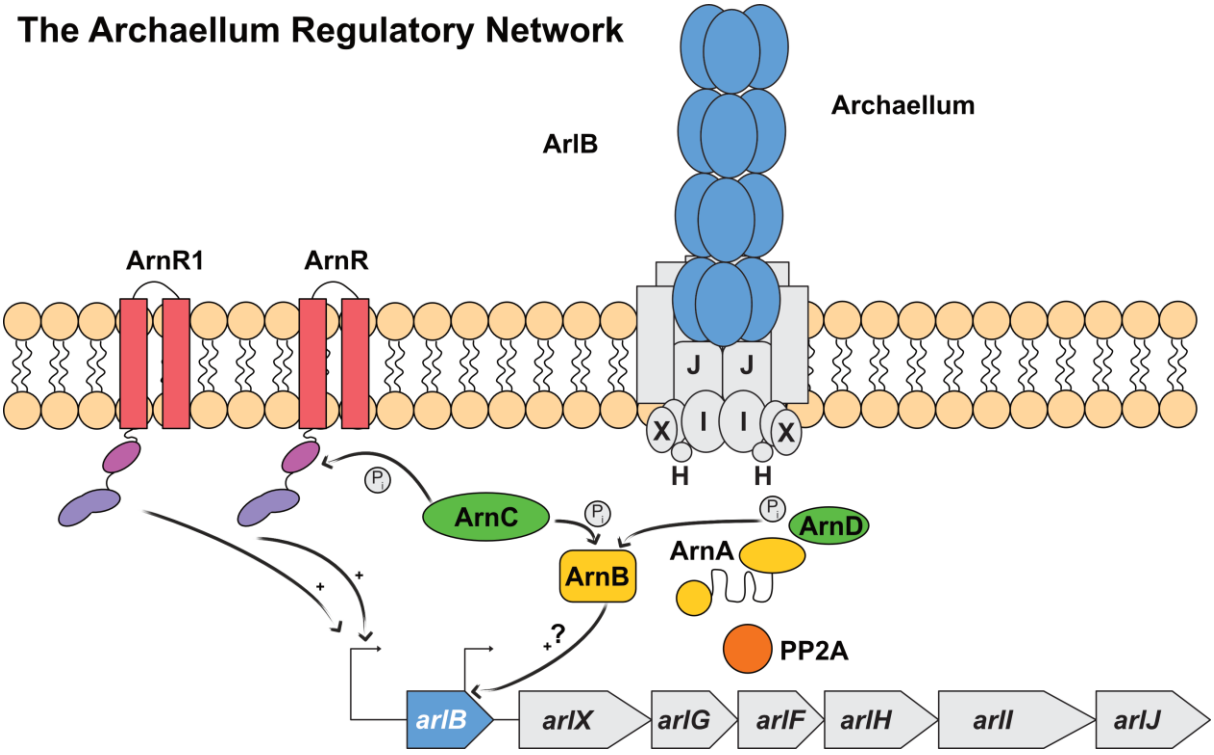

Figure 6: Simplified summary of the Archaellum regulatory Network with focus on ArnB. ArnA and ArnB form a complex mediated by ArnC and possibly also by ArnD. Upon nutrient starvation the complex dissociates, probably by the intervention of PP2A, which triggers *arlB* expression, although it is not fully clarified how this regulation is executed. The regulators ArnR and ArnR1, which are known positive regulators of *arlB*, are also regulated by the phosphorylation of ArnC. Figure derived from Bischof et al. 2019.

GTPase Saci_1281, annotated as ATP/GTP binding protein, alongside ArnA and ArnB, as primary targets. The pulldown assay leaves room for interpretation—whether the interaction partners directly engage with PP2A or if they are secondary hits due to their interactions with each other or the ArnAB complex. Nonetheless, these findings introduce intriguing research avenues for investigating the archaellum regulatory network and signal transduction in archaea. Although much of the archaellum regulatory network's regulation has been elucidated, certain aspects, especially the redundancy of regulatory functions, remain unclear. The formation and regulation of the archaellum, particularly centered around the interaction of ArnA and ArnB, continues to be a captivating subject. Moreover, given some structural features of ArnB, it holds potential significance in discussions about an early endomembrane system in Archaea.

2. Motivation

The emergence of Eukaryotes through endosymbiosis is accepted in eukaryogenesis, but the circumstances and processes leading to this event remain a subject of ongoing debate. Despite numerous hypotheses for explaining eukaryogenesis by focusing on different scientific findings, none have fully traced the origin of eukaryotes within an endosymbiotic framework. One promising approach is the inside-out theory. However, a crucial prerequisite for this theory is the involvement of some form of early endomembrane system in the host, likely belonging to the Asgard archaeal superphylum. Metagenome studies suggest the presence of several trafficking genes found in eukaryotes within these Asgard archaea, making them to a promising target for investigation, although the challenging cultivation of Asgard Archaea hinders their suitability as model organisms. Notably, the TACK superphylum, a close relative of the Asgard family, has been shown to contain homologs of the ESCRT trafficking system, e.g. in the thermophile organism *Sulfolobus acidocaldarius*. This makes *S. acidocaldarius* a prime target for studying pre-endosymbiosis trafficking.

Here, the exosomes of *S. acidocaldarius* exhibit a high content of ESCRT-III proteins and other proteins involved in the archaellum regulatory network, such as ArnA and ArnB. While ArnA and ArnB are known regulators for ArlB, subject to phosphorylation by ArnC and dephosphorylation by PP2A, the exact mechanism of this regulation remains unclear. It could be shown by ESSEN that ArnB corresponds to the core motif of Sec23/24, the membrane curvature protein of COPII vesicles. Accordingly, the initial working hypothesis of this thesis proposed that ArnB plays a role as an early Sec23/24 analog, functionally linked to vesicle biogenesis, i.e. exosome formation. Consequently, a structure-focused approach was employed to characterize the ArnB-Sec23/24 core motif, aiming to investigate this intriguing structural homology. Additionally, the intricate regulation of (de)phosphorylation in ArnA and ArnB, lacking both structural and functional assignments, became research subjects in this thesis. Furthermore, previous data hinted towards possible interaction partners of ArnA and ArnB potential contributing to their assigned function within the archaellum regulatory network, prompting structural and functional characterization efforts for these potential interaction partners. To explore different functional contexts of the ArnAB complex as well as its interaction partners experiments involving proteome analyses, bioinformatic analyses and enzymatic characterization were conducted.

3. Publications

3.1 Archaeal GPN-loop GTPases involve a lock-switch-rock mechanism for GTP hydrolysis

This paper was part of the collaboration of the Volkswagen Foundation Life? Program involving the Albers group from Freiburg University and the Essen group from Marburg University. The goal was to provide structural and functional support of the potential ArnAB interactors as well as elucidating the archaeal class of GPN-loop GTPases.

Citation

L. Korf, X. Ye, M. S. Vogt, W. Steinchen, M. Watad, C. Does, M. Tourte, S. Sivabalasarma, S. V. Albers, L.-O. Essen, Archaeal GPN-loop GTPases involve a lock-switch-rock mechanism for GTP hydrolysis*, MBio* **14**, e0085923 (2023).

Contribution

L. Korf played a pivotal role in orchestrating and executing a comprehensive array of biochemical, bioinformatic, and biophysical experiments throughout this project. This encompassed the planning and execution of isothermal titration calorimetry (ITC), enzyme kinetics analysis, sequence similarity network (SSN) generation and analysis, as well as the processes of crystal growing, structure determination, and refinement for both nucleotide states of *Sa*GPN. L. Korf also established recombinant purification of the GTPase in collaboration with Dr. M.S. Vogt and was responsible for providing any purified *Sa*GPN samples. Moreover, L. Korf assumed responsibility for the thorough analysis of resulting data. Furthermore, L. Korf took the lead in designing all figures showcased within the manuscript, with input data of the HDX, proteomics and *in vivo* experiments contributed by the other coauthors. In addition, L. Korf wrote the first draft of the paper with the exception of the HDX data section, introduction and initial version of the “Biological effect of *saGPN* deletion on the archaeon S. acidocaldarius” paragraph. Beyond the research endeavors, L. Korf undertook the task of organizing and managing the administrative aspects of manuscript publication This included overseeing revisions and crafting the revised manuscript in collaboration with L.-O. Essen and inputs from the coauthors, thereby contributing significantly to the final polished outcome and ultimately justifying the first author-ship.

Summary

GPN-loop GTPases (GPNs) constitute a unique class of GTPases that form dimers irrespective of the nucleotide-bound state and can even create heterodimers in Eukaryotes, enabled by the presence of three paralogous GPNs in most eukaryotic organisms. In Eukaryotes, they are recognized for their involvement in RNA polymerase II trafficking, DNA repair, and methylation-dependent transcription. However, in Archaea, GPNs usually exist as singlets, as analyzed in the sequence similarity network, with their function only to be speculated about. To explore this understudied class of GTPases, the GPN-loop GTPase *Sa*GPN from *Sulfolobus acidocaldarius* was focused in this study, investigating its proteome. This work provides fundamental structural support for *Sa*GPN and GPNs in general, based on its two distinct nucleotide states. Notably, the discovery of a closed state when bound to GTP led to the formulation of a mechanism aptly named the "lock-switch-rock mechanism" due to its illustrative depiction of the allosteric changes observed in this GPN member. In this scenario, binding to GTP locks *Sa*GPN in a state of high tension, which is released during the switching event—either the hydrolysis of GTP to GDP or the release of free phosphate—in a rocking-like motion. While the nucleotide affinities of *Sa*GPN support high binding efficiency, the low intrinsic GTPase activity appears to be a hallmark of archaeal GPNs. Nevertheless, especially the latter suggests the potential involvement of yet unobserved GAPs and/or GEFs in the mechanism of GPNs. Proteomic data from the *saGPN* deletion strain suggests the involvement of the GPN-loop GTPase in global homeostasis due to the significant impact on protein levels, including ArlB, the major Archaellin. However, unaffected growth rates indicate that the role of *Sa*GPN might correspond to different lifestyles, as seen in the observed loss of motility.

In comparison, eukaryotic GPNs have diverse roles, including participation in sister chromatid cohesion (human) and acting as a chaperone for the assembly of the 12-subunit RNA polymerase II (*S. cerevisiae*). They are also implicated in processes like mitochondrial homeostasis and ribosome biogenesis. Given the broad spectrum of functions exhibited by eukaryotic GPNs, it is plausible that archaeal GPNs, such as *Sa*GPN, which undergo conformational switching between GTP- and GDP-bound states, may play a similar role, potentially acting as a chaperone during multi-protein assembly.

AMERICAN SOCIETY FOR MICROBIOLOGY | mBio

Check for updates

| Structural Biology | Research Article

Archaeal GPN-loop GTPases involve a lock-switch-rock mechanism for GTP hydrolysis

Lukas Korf,[1] Xing Ye,[2] Marian S. Vogt,[1] Wieland Steinchen,[1,3] Mohamed Watad,[1] Chris van der Does,[2] Maxime Tourte,[2] Shamphavi Sivabalasarma,[2,4] Sonja-Verena Albers,[2] Lars-Oliver Essen[1]

AUTHOR AFFILIATIONS See affiliation list on p. 20.

ABSTRACT Three GPN-loop GTPases, GPN1–GPN3, are central to the maturation and trafficking of eukaryotic RNA polymerase II. This GTPase family is widely represented in archaea but typically occurs as single paralogs. Structural analysis of the GTP- and GDP-bound states of the *Sulfolobus acidocaldarius* GPN enzyme (*Sa*GPN) showed that this central GPN-loop GTPase adopts two distinct quaternary structures. In the GTP-bound form, the γ-phosphate induces a closed dimeric arrangement by interacting with the GPN region that is relaxed upon hydrolysis to GDP. Consequently, a rocking-like motion of the two protomers causes a major allosteric structural change toward the roof-like helices. Using a lock-switch-rock mechanism, homo- and heterodimeric GPN-like GTPases are locked in the GTP-bound state and undergo large conformational changes upon GTP hydrolysis. A Δ*saGPN* strain of *S. acidocaldarius* was characterized by impaired motility and major changes in the proteome underscoring its functional relevance for *S. acidocaldarius in vivo*.

IMPORTANCE GPN-loop GTPases have been found to be crucial for eukaryotic RNA polymerase II assembly and nuclear trafficking. Despite their ubiquitous occurrence in eukaryotes and archaea, the mechanism by which these GTPases mediate their function is unknown. Our study on an archaeal representative from *Sulfolobus acidocaldarius* showed that these dimeric GTPases undergo large-scale conformational changes upon GTP hydrolysis, which can be summarized as a lock-switch-rock mechanism. The observed requirement of *Sa*GPN for motility appears to be due to its large footprint on the archaeal proteome.

KEYWORDS GPN-loop, GTPase, sulfolobus, RNA polymerase II, allostery

GTPases are a large family of GTP-binding and -hydrolyzing enzymes that are widely distributed across all three domains of life (1). They contain a highly conserved GTPase domain (G domain) housing five fingerprint motifs responsible for coordination of GTP and catalysis: G1 (also P-loop or Walker A motif) interacts with the 5′ phosphate moieties of GTP, the G2 and G3 motifs are required for coordination of a magnesium ion essential for catalysis, the latter of which specifically accommodates the 5′ γ-phosphate of GTP, and G4 and G5 establish specific binding of the nucleotides' guanine base (2, 3). As GTPases cycle between their GTP and GDP-bound states via their intrinsic GTP hydrolytic activity, they often function as molecular switches differentially regulating a plethora of downstream effector proteins involved in crucial cellular processes (2, 4, 5).

Most GTPases belong either to the TRAFAC (translation factor association) class involved in translation, signal transduction and intracellular transport, or the SIMIBI (signal recognition particle, MinD, and BioD) class; members of the latter engage in protein localization and trafficking, membrane transport, and chromosome partitioning (1, 4, 6). Many well-studied GTPases, such as Ras and heterotrimeric G proteins (7–9),

Invited Editor Yilin Hu, University of California Irvine, Irvine, California, USA

Editor Markus W. Ribbe, University of California Irvine, Irvine, California, USA

Address correspondence to Sonja-Verena Albers, sonja.albers@biologie.uni-freiburg.de, or Lars-Oliver Essen, essen@chemie.uni-marburg.de.

Lukas Korf and Xing Ye contributed equally to this article. Author order was determined alphabetically and by mutual agreement of both first authors.

The authors declare no conflict of interest.

See the funding table on p. 20.

Received 5 April 2023
Accepted 5 October 2023
Published 14 November 2023

belong to the TRAFAC class, whose members generally do not form nucleotide-dependent dimers (1). In contrast, homo- and heterodimerization of SIMIBI class proteins like the signal recognition particle (SRP) and SRP receptor GTPases depends on ATP- or GTP-binding (4, 10) and is hence relevant to regulation of their intrinsic ATP/GTP hydrolysis activity, specific interactions with effector proteins, and their biological functions (6, 11, 12).

A novel type of SIMIBI GTPase, the GPN-loop GTPase, was discovered by structural analysis of the GPN-loop GTPase PAB0955 (*Pa*GPN, Uniprot: Q9UYR9) from the euryarchaeon *Pyrococcus abyssi* (13). *Pa*GPN adopts a homodimeric topology, whose quaternary assembly was found to be independent of GTP binding (13). This characteristic of GPN-loop GTPases sets them aside from other SIMIBI GTPases, which require switch-dependent changes of their quaternary state to fulfill their biological function (6). Another feature is the eponymous GPN motif in the G domain that is highly conserved in archaeal *Pa*GPN and eukaryotic orthologs. This motif (Gly-Pro-Asn) is inserted in between the SIMIBI class motifs G2 and G3. GPN-loop GTPases were described to occur only in archaea and eukaryotes, but not in bacteria (14). Eukaryotes typically feature three GPN-loop GTPase paralogs: GPN1 (annotated before as Npa3, XAB1, or MBD*in*), GPN2 and GPN3. These paralogs play essential roles in nuclear localization and biogenesis of the RNA polymerase II (14–17). In yeast, deletion of genes encoding the XAB1 homologs GPN1, GPN2 or GPN3 was found to be lethal (18). Furthermore, yeast GPNs assemble into heterodimers with different yGPN combinations, which appear to be irrespective of the nucleotide-bound state of the GPNs as pull-down assays, FRET, and molecular modeling studies indicate (14, 19).

The biological function of archaeal GPN-loop GTPases remains unknown, although they exhibit the highest sequence similarities with eukaryotic GPN1 orthologs (1, 19). Nevertheless, the crystal structures of *P. abyssi* GPN (PDB: 1YR6 [apo form], 1YRA [GDP complex], 1YR7 [GTP-γ-S complex], 1YR8 [GTP complex], 1YR9 [GDP+HPO_4^{2-} complex], 1YRB [GDP+Mg^{2+} complex], 2OXR [GDP+Mg^{2+} complex, after GTP hydrolysis]) and yeast GPN1 (also called Npa3, PDB: 5HCI [GDP complex], 5HCN [GMPPCP complex]) revealed a similar homodimeric topology, nucleotide coordination, and a mode of catalysis, which involves the asparagine residue of the GPN motif (13, 16).

For long time, the cellular function of GPN-loop GTPases remained enigmatic. *Homo sapiens* GPN1 was described first as an XPA-binding protein 1 (XAB1) or MBD2-interacting protein (MBD*in*) because of its interaction with the nucleotide excision repair protein Xeroderma pigmentosum group A protein (XPA) and methyl-CpG-binding protein MBD2, thus putting forward a role in DNA repair and methylation-dependent transcription (20, 21). Elucidation of the protein interaction network of all three GPN proteins in yeast revealed their essential role in the regulation of nuclear import of multiple RNA polymerase II (RNAPII) subunits (22, 23). The interaction of yeast GPN1/Npa3 with RNAPII subunit Rpb1 *in vitro* was stronger in the presence of GTP than GDP (24). Furthermore, mutations in the G1, G2, G3, and GPN motifs of GPN1/Npa3 were either lethal or resulted in a slow growth phenotype in yeast (23), collectively suggesting a link between the nucleotide-bound state of GPN GTPases and their cellular function.

In the archaeon *Sulfolobus acidocaldarius*, the GTPase *Sa*ci1281 (from hereon: *Sa*GPN) represents the solitary GPN-loop GTPase. We previously identified *Sa*GPN as an interaction partner of the Ser/Thr specific protein phosphatase PP2A (25), which we anticipate to play an important role in intercellular signaling pathways because *S. acidocaldarius* possesses a diverse phosphoproteome (26). Interestingly, PP2A interacts with ArnA and ArnB, which are described as negative regulators of the assembly of the archaellum (25, 27, 28). PP2A deletion strains are therefore hypermotile and defective in cell size regulation and metabolism (26).

We investigated the *in vivo* function of *Sa*GPN in *S. acidocaldarius*, focusing on its potential role in the regulation of archaella synthesis. We also analyzed the protein structure and conformational changes of *Sa*GPN by crystallization and HDX analysis. The

results from this study shed light on the *in vivo* function of GPN-loop GTPase in archaea, and we will also present a novel activation mechanism of GPN-loop GTPases.

RESULTS

GPNs occur in all domains of life

To provide an overview of GPN-loop GTPases (IPR004130), we analyzed and visualized this protein family using a sequence similarity network (SSN), which provides insights into protein relationships based on different sequence alignment scores (Fig. 1). Thereby, nodes correspond to protein sequences, and each edge represents one relation, which scales in stringency with decreasing *E*-values, allowing clustering based on this alignment threshold (29). Cluster analysis of IPR004130 with an *E*-value of 10^{-40} reveals that GPN-loop GTPases (GPNs) from archaea share a closer relation to GPN1 than to GPN2 and GPN3, which is consistent with the essential diversity of GPNs in eukaryotes (15). Moreover, GPN1 does not cluster with its sister paralogs GPN2/3, although the latter shares the same cluster, suggesting a greater functional discrepancy between GPN1 and GPN2/3 than between GPN2 and GPN3. Additionally, it appears that GPNs of TACK archaea, which include *S. acidocaldarius*, have a closer relationship with GPN1 than with other archaeal GPN-loop GTPases such as *P. abyssi* (Fig. 1). Furthermore, *Pyrococcus* and *Sulfolobus* GPNs begin to separate from each other at alignment scores with *E*-values of $<10^{-40}$. Segregation became even more evident at a slightly increased *E*-value cutoff of $<10^{-45}$, with the eukaryotic GPN1 clade being completely dissociated from archaeal GPNs. Likewise, *Pyrococcus*-like GPNs are almost completely clustered separate from the *Sa*GPN-loop GTPases (Fig. 1). This indicates, that despite sharing identity to some extent, these archaeal GTPases have some differences that are yet to be clarified. In our analysis of 1,541 unique archaeal sequences, belonging to 469 different organisms, we found that GPN-loop GTPases in archaea mostly occurred as singlets, with a total of 412 organisms. However, 57 archaeal organisms had two or more GPN paralogs, refuting the established assumption that archaea generally harbor only a single ortholog of GPN-loop GTPases (16, 19, 30).

Interestingly, 4,331 bacterial sequences are found in the SSN of IPR004130, which can be assigned to 1,859 different bacterial strains. Here, singlets also form the largest group in the bacterial domain (762 organisms); 410 organisms were identified with two paralogs, and 386 with three. In addition, some bacterial species even appear to have four or more paralogs, exceeding the number of three, GPN1–GPN3, as known in eukaryotes. However, all these bacterial members of IPR004130 share few relationships with the clusters dominated by either archaeal or eukaryotic members. Structure prediction and sequence alignments actually show that these bacterial GTPases miss the GPN motif and are structurally more related to ARF-type small GTPases (Fig. S10). This indicates that these bacterial members have been falsely annotated as GPN-loop GTPases and hence play a different and maybe more diverse role in bacteria when compared to eukaryotes or archaea. Furthermore, at an *E*-value of 10^{-40} archaeal orthologs assemble almost exclusively within a single, large cluster and thereby share links with almost every other archaeal GPN, whereas bacteria split into several small clusters that have only minor relationships.

Biological effect of *saGPN* deletion on the archaeon *S. acidocaldarius*

*Sa*GPN (Saci_1281, Uniprot: Q4J9A7) was identified as a specific interactor of the serine/threonine protein phosphatase PP2A (25). In *S. acidocaldarius,* PP2A regulates growth, cell size, and swimming motility (26). To investigate the physiological role of *Sa*GPN in *S. acidocaldarius*, a markerless in-frame deletion mutant was constructed. Deletion of *saGPN* did not affect cell growth (Fig. 2A), whereas a pronounced decrease in swimming motility on semi-solid gelrite plates was observed (Fig. 2B). Interestingly, electron microscopy (EM) analysis showed that Δ*saGPN* cells still assembled archaella. However, these samples have to be taken after 4 h of starvation as otherwise archaella cannot be

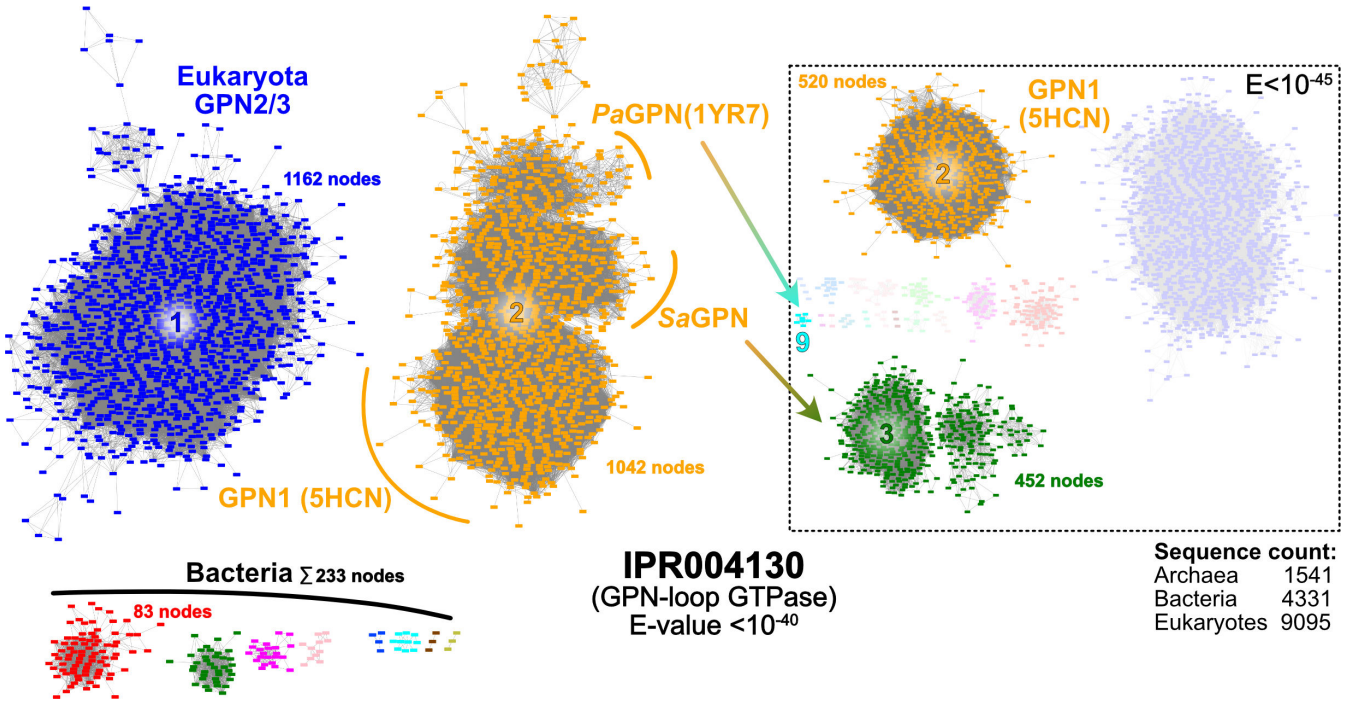

FIG 1 Sequence similarity network (SSN) of InterPro protein familiy IPR004130 (GPN-loop GTPases). The SSN shows connection between related proteins of the input protein familiy at the given *E*-value cutoff of 10^{-40}. Cluster analysis numbers clusters based on member count, giving insights in overall cluster size. While the archaeal GPNs are clustered with GPN1 in cluster 2, other eukaryotic GPNs (GPN2/3) are located in cluster 1, revealing differences between GPN1 and its sister proteins. Additionally, cluster separation of cluster 2 is observable, where PaGPN, SaGPN, and GPN1 start to form their own subclusters. This is even more pronounced at a higher stringency of 10^{-45} (dashed box), where SaGPN separates from cluster 2 alongside most other archaeal GPNs into cluster 3 (452 nodes) and PaGPN is clustered away into its own cluster 9 (cyan). Nodes not found in clusters 2 and 3 clustered in smaller/singleton subclusters, which are not displayed for clarity. Notably, bacterial members of IPR004130 (lower left) miss the GPN-loop motif and are assignable to ARF-like GTPases (Fig. S10).

observed even in the wild type (WT) (Fig. 2C). To test whether deletion of *saGPN* affected *arlB* expression (encoding the archaellum filament protein, archaellin), Δ*saGPN* cells were starved for nutrients for 0–4 h, and *arlB* expression was analyzed by Western blot analysis and qRT-PCR (Fig. 2D). *arlB* expression generally increased during nutrient starvation in both the wild-type strain MW001 and Δ*saGPN* mutant. However, ArlB protein levels were significantly reduced in the Δ*saGPN* mutant compared with those in MW001. Similarly, decreased *arlB* expression was observed on the RNA level when comparing the Δ*saGPN* mutant with MW001 at 1, 1.5, 2, and 4 h after starvation (Fig. 2E). Thus, *Sa*GPN affects the archaellum regulation network and apparently acts as a positive regulator *in S. acidocaldarius*.

To explore whether *Sa*GPN affects other functions in *S. acidocaldarius*, we performed a proteome-scale analysis of wild-type and Δ*saGPN S. acidocaldarius* strains. These were grown as biological triplicates and samples were taken from nutrient-rich and nutrient-starved conditions. Each sample was measured as technical duplicates using a timsTOF (trapped ion mobility spectrometry, time of flight) mass spectrometer before averaging and further downstream processing. Measurement of these proteomes led to the identification of 1,422–1,512 proteins per sample, which corresponded to a total of 1,627 different proteins of the 2,222 proteins encoded in the *S. acidocaldarius* genome. Comparing the two nutritional states of the WT, it is apparent that most of the proteome is nutrient-independent with only 37 proteins not identified in both states simultaneously (Fig. 3A). For a quantitative assessment of the measured protein abundances, data were further filtered using a *t* test to analyze only hits that showed statistically significant differences in abundance for data in their respective comparison. A total of 242 of the 1,492 (16.2%) overlapping proteins found in both nutritional states of the WT passed this filtering, indicating that the expression levels of at least 16% of members of the

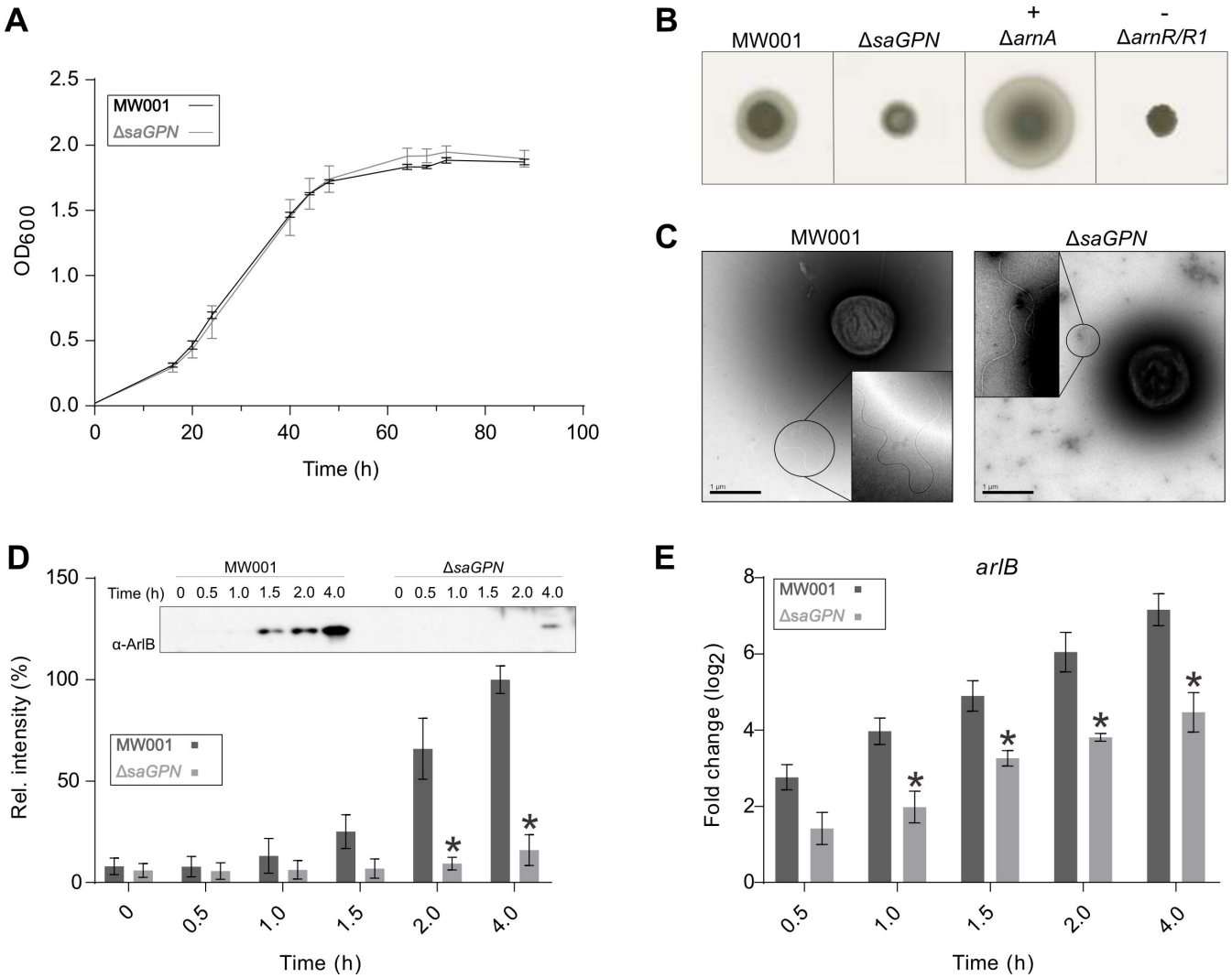

FIG 2 Deletion of saGPN impedes motility and arlB (flaB) expression in *S. acidocaldarius*. (A) Growth curves of *S. acidocaldarius* MW001 (black line) and Δ*saGPN* mutant (gray line) in nutrient-rich medium. (B) Motility assays. Same amounts of cells from tested *S. acidocaldarius* strains were spotted on semi-solid gelrite plates and incubated at 75°C for 5 days, then the plates were scanned and recorded. Δ*arnA* and Δ*arnR/R1* strains were used as hyper-motile and non-motile control, respectively. (C) Analysis of the archaellum formation in *S. acidocaldarius* strains. *S. acidocaldarius* MW001 and Δ*saGPN* knockout were cultivated in nutrient-depleted medium. After 4 h growth, cell samples were collected and applied to EM analysis. (D) ArlB expression on protein level and (E) *arlB* transcription level in *S. acidocaldarius*. *S. acidocaldarius* MW001 and Δ*saGPN* mutant were cultivated in nutrient-depleted medium for 4 h. Cell samples were taken at different time points, and analyzed by Western blotting with α-ArlB (left) and qRT-PCR (right). A representative Western blot is shown (**D**) and Western blots from biological triplicates were quantified. Relative transcription levels of arlB were normalized to secY. The values represent fold changes compared with the control from biological triplicates. Significant differences between MW001 (dark boxes) and Δ*saGPN* mutant (light boxes; *P* value < 0.05) were indicated by an asterisk.

proteome depend on nutrient availability. These data resemble previous label-based iTRAQ quantification analyses, which also indicated nutrient-dependent proteome changes of up to 12% (31). However, comparison of the Δ*saGPN* proteomes with their corresponding WT strains indicated that the amount of significantly altered proteins was approximately two times as high for the knockout strains, with 546/1,413 (38.6%) for nutrient-rich and 445/1,467 (30.3%) nutrient-starved conditions, respectively (Fig. 3A). While this analysis reflected statistically significant changes, we also verified that the changes in protein levels were appropriately high by additionally filtering for proteins that showed a change of at least 50% in their respective comparison. This further filtering for largely altered expression levels (>50%) yielded 83 proteins for the $WT_{starved}/WT_{rich}$, 134 for the $WT_{rich}/\Delta saGPN_{rich}$ and 127 for the $WT_{starved}/\Delta saGPN_{starved}$ comparison (Fig.

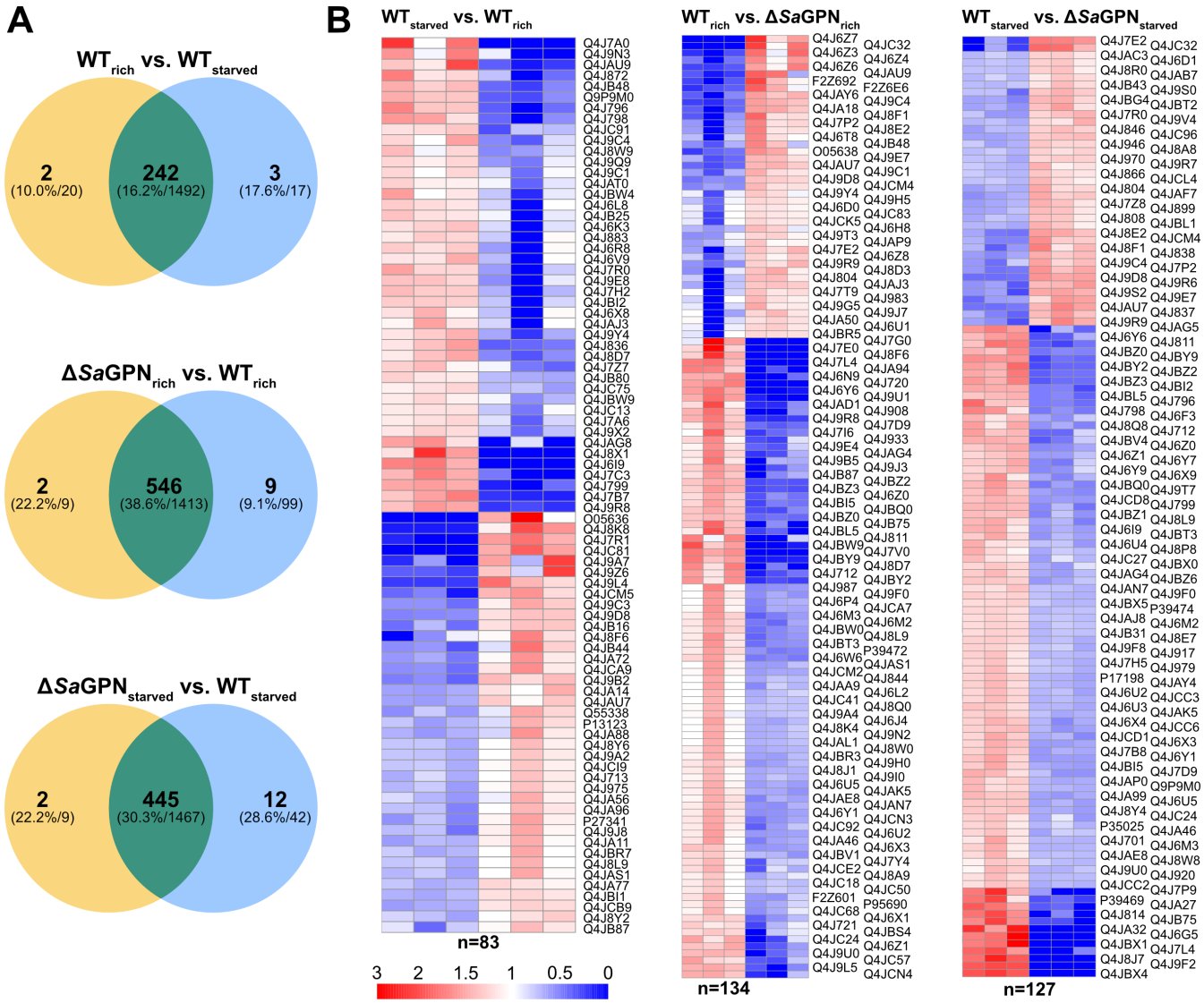

FIG 3 Knockout of *saGPN* has extensive ramifications on the *S. acidocaldarius* proteome. (A) Venn diagrams of proteomic data comparing different nutritional states of the wild type (WT) and *saGPN* deletion mutant. Diagrams show the number of overlapping and non-overlapping proteins that passed a one-paired *t*-test ≤0.01, to check for variation of the biological triplicates and the technical duplicates of their measurement, in bold. The proportion of these proteins filtered by the *t*-test to the total overlapping and non-overlapping proteins found is given in parentheses. The amounts of total proteins as identified by timsTOF, i.e. without *t*-test filtering, are given in absolute numbers next to them. (B) Heatmaps of proteins that exhibited highly affected expression levels (>50%) in their respective comparisons. The number of proteins that were severely affected is indicated below the maps. The scale bar represents the fold change of over-regulation (red) to under-regulation (blue). Uniprot identifiers of the respective proteins are indicated next to the rows.

3B). Heatmap analysis of this filtering shows that the distribution of over- and underregulated proteins is similar between WT states, whereas the knockout seems to have a more pronounced effect on downregulation. Nevertheless, the absence of the GPN-loop GTPase has profound effects on protein expression, even when only highly significant and major changes are considered.

Evaluation of the gene ontology (GO) terms of the proteins shown in the heatmap analysis (Fig. 3B) revealed that a large variety of different protein functions are affected by the *saGPN* knockout (e.g., ArlB, Uniprot: Q4J9K5), pointing to a global role of *Sa*GPN in protein homeostasis (Table S4). However, given the unaffected growth rate of the *saGPN* deletion strain, these changes in protein levels may correspond more to different lifestyles of this archaeal organism, e.g., due to the observed loss of motility (Fig. 2).

Biochemical and biophysical characterization of *Sa*GPN

In size-exclusion chromatography, purified *Sa*GPN (30.5 kDa) eluted as a monodisperse peak at a volume corresponding to a molecular weight of ~60 kDa, suggesting *Sa*GPN forms a dimer in solution (Fig. S1A and B). GTPase activity of *Sa*GPN was highest in the presence of Mg^{2+} at the optimal growth temperature (75°C) of *S. acidocaldarius* (Fig. S1C). However, this was not applicable for all experiments, as mechanical or chemical stress could result in an onset of protein aggregation at 75°C, which is why the temperature was adjusted to 65°C for kinetic experiments if not stated otherwise. *Sa*GPN preferred Mg^{2+} as the cofactor over other tested metal ions (Fig. 4B). To determine the GTP hydrolysis kinetics, *Sa*GPN was incubated with 0–500 µM GTP at 65°C in the presence of Mg^{2+}. The calculated GTP hydrolysis was plotted against the GTP concentration and data were fitted with the Michaelis-Menten equation. The GTPase activity of *Sa*GPN had a K_M of 40.48 µM and a V_{max} of 6.4 nmol·mg^{-1}·min^{-1} (36.4 nmol·L^{-1}·s^{-1}), respectively (Fig. 4A).

Nucleotide binding efficiency of *Sa*GPN was investigated by employing isothermal titration calorimetry (ITC), yielding dissociation constants (K_D) in the low nanomolar range for all nucleotides investigated. To avoid protein degradation throughout the measurements, nucleotide affinity was determined at 65°C, revealing K_D values of 6.15 and 22.0 nM for GDP and the non-hydrolyzable GTP derivative GppNHp, respectively (Fig. 4C and D). At 25°C, the nucleotide affinity is even higher without a significant disparity for triphosphate nucleotides, indicating that GppNHp is a suitable substitute for GTP (Fig. S1D, E and F). Notably, dimerization of *Sa*GPN is very stable against high dilutions of ≤15 nmol as shown by massphotometry, as no dissociation was observed during measurements (Fig. S1G).

*Sa*GPN overall structure

The structure of *Sa*GPN was determined from a monoclinic crystal form at a resolution of 1.8 Å in its GppNHp-bound state (PDB: 7ZHF) and solved by molecular replacement, covering residues Y2-A240, with one molecule in the asymmetric symmetry unit (a.s.u.) forming a physiological dimer when crystal symmetry is applied. In addition to the GppNHp structure, we have succeeded in solving the *Sa*GPN structure in its GDP-bound state from an orthorhombic crystal form (PDB: 7ZHK, Y2-A240) at a resolution of 2.4 Å. *Sa*GPN for both crystal forms was purified before in a nucleotide-free state and co-crystallized with the respective nucleotide (Fig. S9C). The GDP-bound structure contains three molecules per a.s.u., where monomers A and C form together a physiological dimer. Likewise, monomer B forms a dimer with its crystal mate B′. The difference between both *Sa*GPN dimers in this crystal form is minor, as a superposition of the dimers yields an r.m.s.d value of only 0.24 Å (445 Cα atoms). Interestingly, the *Sa*GPN dimer interface shows local heterogeneity near the GTP binding site (Fig. S12B through D). In monomers A and B, a double hydrogen bond between N70 of the GPN motif with Q106′ is formed, whereas altered χ_2 angles of N70′ and Q106 of monomers C and A, respectively, are consistent only with the formation of a single H-bond.

The general structure of SaGPN involves 12 α-helices and a five-stranded, parallel β-sheet that is sandwiched by α6/α8 and α1/α12, respectively (Fig. 5A). Moreover, several small helices and helical turns consisting only of a few amino acids can be found throughout the whole structure, including α5, α7, and α10. The crystal structures of both crystal forms show homodimers of *Sa*GPN, concurring with the SEC elution volume (Fig. S1). The dimer interface emerges alongside the surface of helices α3, α5, α8, and α11, whereby α8 contributes substantially to the dimeric assembly due to its central position allowing the formation of a hydrophobic pocket. This pocket is covered by a roof-like scaffold formed by α9–α11/α11′–α9′ (residues E178-N212) laying orthogonally over α8 helices (residues P143-R159) and parallel to the α2 (residues V58-Y65), the so-called skid region, located at the opposite end. The eponymous GPN motif (residues G68-N70) is found at the base of α3 (residues N70-L81), which is involved in activation of the inline attacking water molecule (N70) and nucleotide coordination at the substrate pocket of

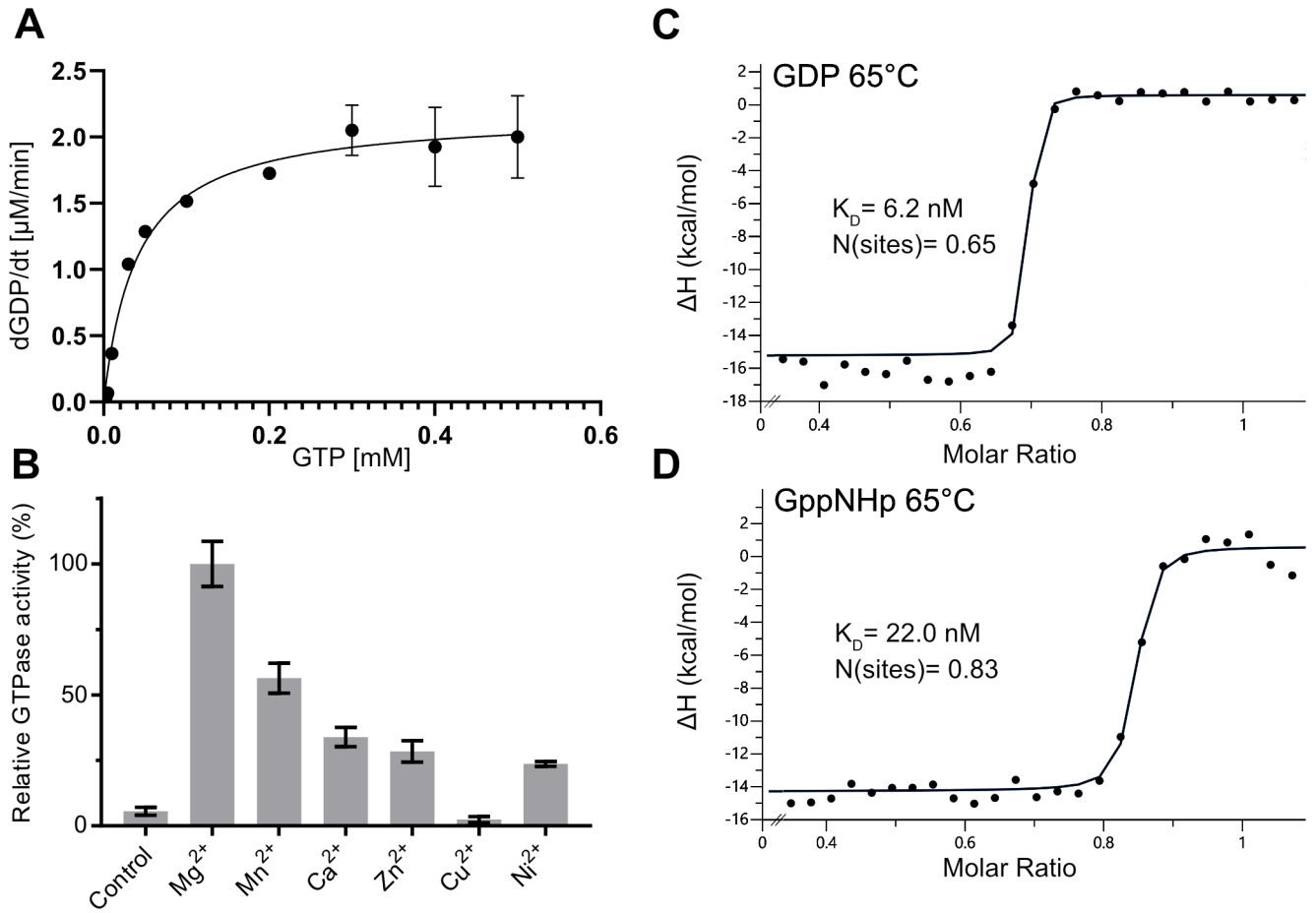

FIG 4 Despite showing low GTPase activity *Sa*GPNs nucleotide binding is very efficient. (A) Michaelis-Menten plot of *Sa*GPN activity at 65°C, revealing an activity of 6.4 nmol hydrolyzed GTP per min and per mg of protein. Data represent mean ± s.d. of $n = 3$ replicants. Error bars (standard deviation) of some data points are too small to be visualized. (B) Effect of different bivalent metal ions on GTPase activity, which is highest with Mg^{2+}. Data represent mean ± s.d. of $n = 3$ replicants. (C) ITC measurement of *Sa*GPN at 65°C (to avoid premature protein degradation) with titration against GDP, showing a substrate affinity of K_D = 6.15 nM for GDP. (D) Same measurement with titration against GppNHp, showing a substrate affinity of K_D = 22.0 nM for GppNHp. Values are calculated from triplicate measurements.

the second protomer. Notably, the water that is suitably positioned for inline attack has still a distance of 3.4 Å from the γ-phosphate of the GppNHp substrate mimic. Accordingly, additional conformational changes for the GPN motif may be feasible during hydrolysis of GTP itself. Nucleotide binding is mainly guided by interactions with the backbone; however, side chain interaction of D172 and K170 seems mandatory for coordination of the base and the ribose, respectively. Additionally, the Walker A motif (GxxxxGK[T/S], x = any amino acid) takes part in nucleotide binding, whereas K14 is responsible for the coordination and orientation of the β/γ phosphates of the nucleotide; T15 is the only proteinaceous part of the octahedral Mg^{2+} coordination sphere (Fig. 5B) (32). Mg^{2+} coordination is also supported by the Walker B motif (hhhhD/E, h = hydrophobic), which interacts indirectly with the Mg^{2+} ion over a water bridge to D102; the latter residue also forms a hydrogen bond with the hydroxyl moiety of T15.

As mentioned above, we have succeeded in solving the *Sa*GPN structure in its GDP-bound state, allowing comparison between both states as well as with the only other known archaeal GPN-loop GTPase structure, *Pa*GPN from *P. abyssi* (see Discussion). Apart from the missing γ-phosphate the binding mode of Mg^{2+}·GDP to *Sa*GPN is comparable to that of Mg^{2+}·GppNHp (Fig. 5C). Interestingly, we observe a water molecule coordinated between D38 and N70′ (primed residues mark the second protomer) close to the position observed before for the inline attacking water molecule of the GppNHp-bound

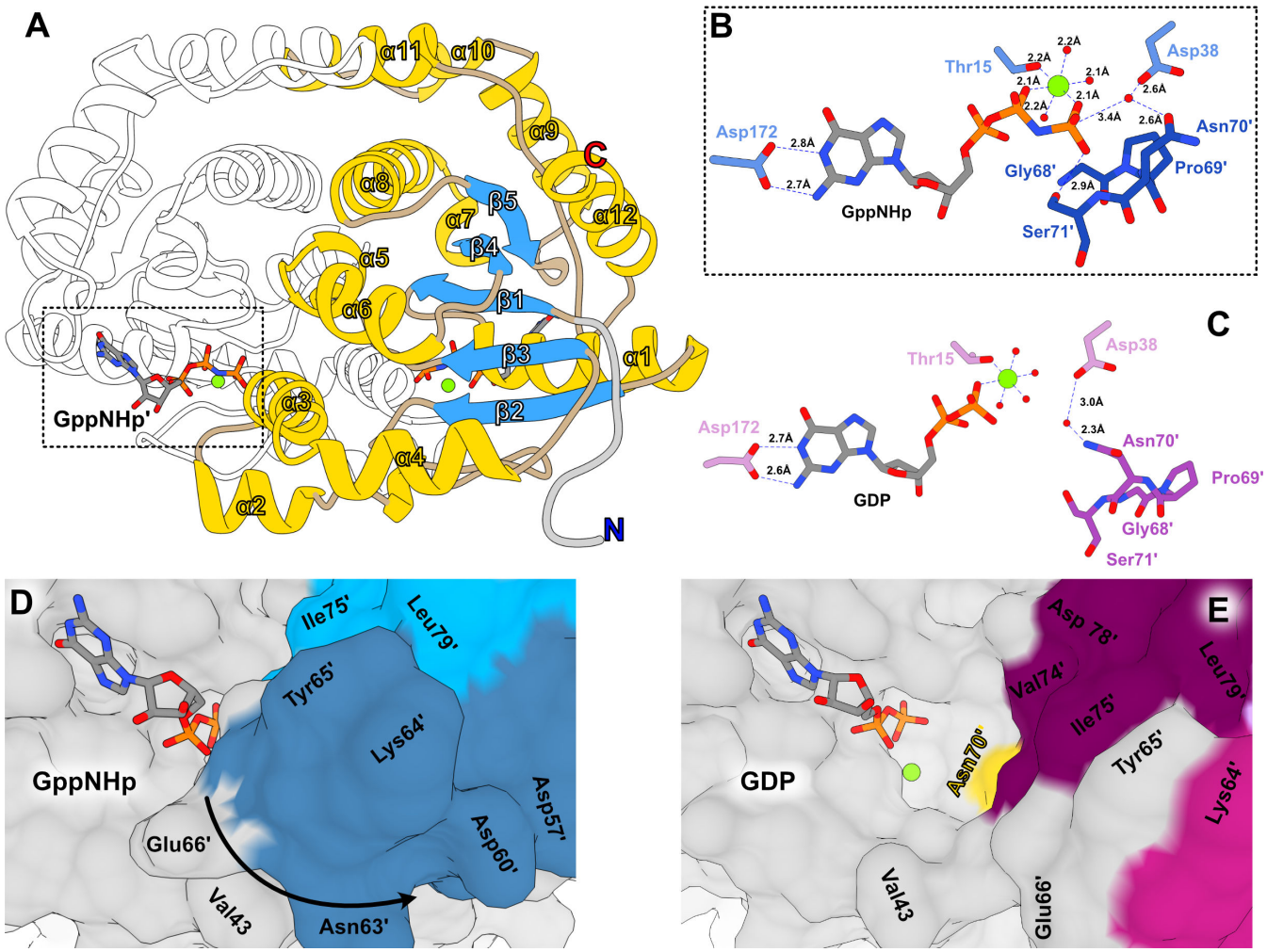

FIG 5 Overall structure of *Sa*GPN and its substrate pocket. (A) Structure of *Sa*GPN(GppNHp) with numbering of secondary structure elements on one protomer. Helices and helical turns are colored in yellow (although helical turns of less than four amino acids are not numbered), β-sheets in blue, non-secondary regions in wheat, and the His-tag in gray. (B) Coordination of GppNHp in the substrate pocket with most important side chain interactions, including the near perfect Mg^{2+} octahedron and the catalytic active water coordination, with Mg^{2+} in green and waters in red. Primed residues refer to the second protomer. (C) Analogous coordination of GDP in the substrate pocket, revealing the difference in Mg^{2+} coordination as well as the increased distance toward the catalytic active water. (D **and** E) Surface view of the substrate pocket in its GppNHp (closed, PDB: 7ZHF) and its GDP state (open, PDB: 7ZHK) with coloring analogous to Fig. 6A for better orientation.

state thus corroborating the role of N70′ of the GPN motif in positioning the water molecule for catalysis. However, due to the loss of the γ-phosphate interaction, the GPN loop detaches from the nucleotide, resulting in an opened substrate pocket around the phosphates (Fig. 5D and E). Interestingly, attempts to derive a GDP·AlF_4^-- or GDP·P_i-bound form that mimics the transition state or product of hydrolysis from the GDP-bound form by soaking or co-crystallization failed, as we only got structures of the GDP-bound state. Accordingly, the GDP-bound state of *Sa*GPN apparently represents a conformation, whose stability exceeds those of the non-covalently linked GDP·AlF_4^-- or GDP·P_i-bound states.

Allosteric triggering of *Sa*GPN by bound nucleotides

While the quaternary structures of the GppNHp and GDP dimer states are clearly different for *Sa*GPN (Fig. 6, r.m.s.d values of 2.3 Å over 442 Cα atoms after superposition), structural changes of the different binding states for *Pa*GPN are only subtle, as shown by an r.m.s.d value of 1.0 Å (453 Cα atoms) between GTP- and GDP-bound states (1YR8 vs

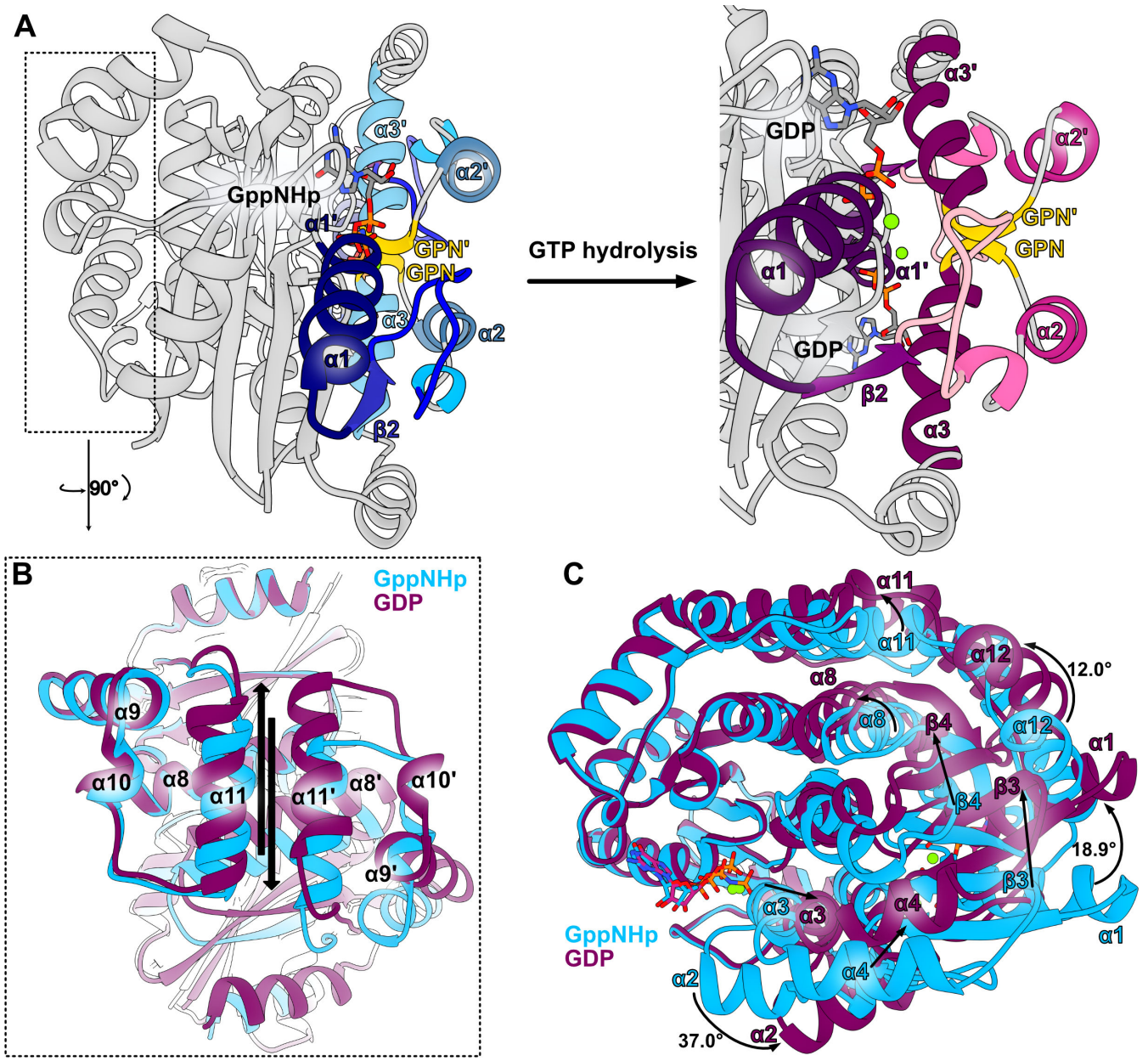

FIG 6 Allosteric changes of *Sa*GPN upon nucleotide hydrolysis observed in crystal structure. (A) *Sa*GPN skid region with corresponding allosteric changes upon nucleotide hydrolysis, revealing the push-out movement of the α2 skids. Secondary structure elements are colored in a blue gradient (GppNHp state) or purple gradient (GDP state) for better orientation, and GPN motif in yellow. (B) Top-down view on the roof helices with arrows indicating displacement between both states upon nucleotide hydrolysis. (C) Overall structure alignment of both nucleotide states, revealing major allosteric changes between states on outside helices as well as inside β-sheets. Most prominent helix shifts as calculated by Chimera are given in degrees.

1YRB). Considering the monomer itself, comparisons between *Sa*GPN and *Pa*GPN reveal relatively small differences for the GDP-bound state, 1.3 Å for 179 Cα atoms, which are increased to 1.6 Å (181 Cα atoms) for the GTP/GppNHp-bound states. Accordingly, the GDP states of *Sa*GPN and *Pa*GPN are closely related, whereas the GTP-bound form of *Sa*GPN differs from its GDP state as well as the *Pa*GPN states described before. Accordingly, these large allosteric changes undergone by *Sa*GPN upon nucleotide hydrolysis are mostly rigid body-like motions, which are driven by a scissors-like movement of the α2 (residues V58-Y65) skid region (Fig. 6A). Achieving this scissors-like motion seems to involve many local changes throughout the whole GTPase assembly: the roof helices

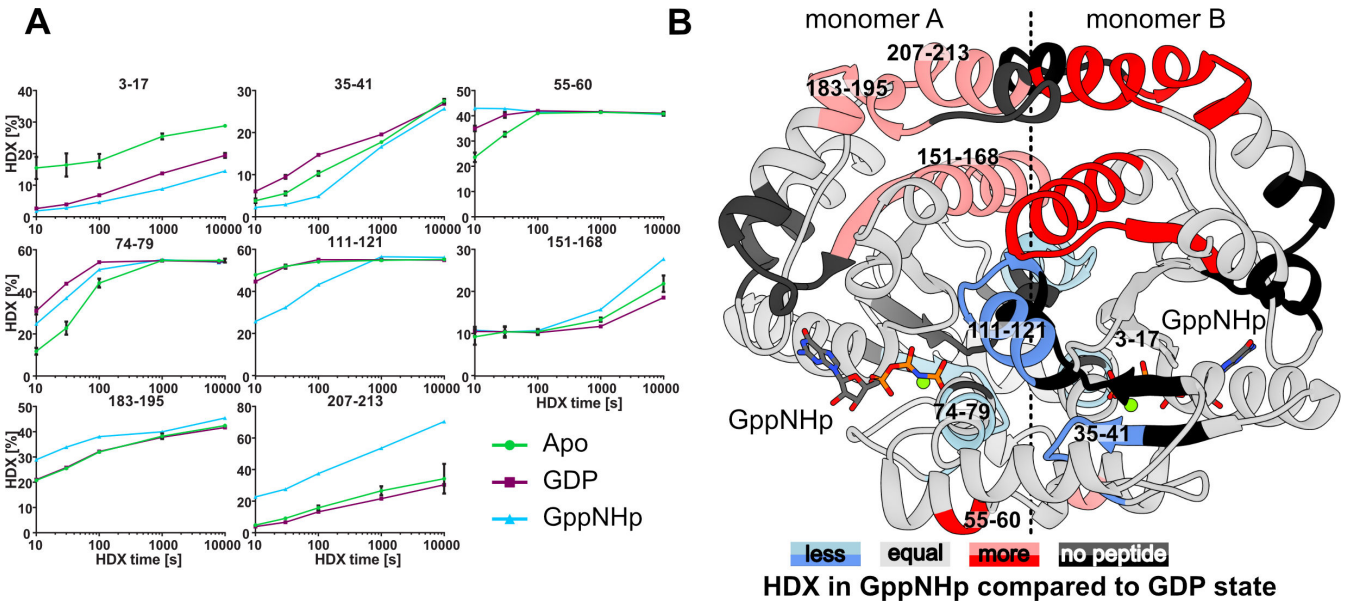

FIG 7 Allosteric changes of SaGPN upon nucleotide hydrolysis observed in solution. (A) HDX of representative SaGPN peptides covering regions of difference between the apo state (green circles), GDP-bound state (purple squares) and GppNHp-bound state (blue triangles). Data represent mean ± s.d. of $n = 3$ measurements. (B) Difference in HDX between GppNHp and GDP state mapped on SaGPNs GppNHp structure (cumulative meaning that any HDX that happened of the respective region was mapped onto the structure with its highest difference). HDX data show that differences between both states are most prominent in roof region, core elements, and substrate coordinating/hydrolyzing regions, which goes along with most allosteric changes observed in the crystal structure. Numbers refer to amino acid position.

shift against each other (Fig. 6B) with the C-terminal α10 (residues S190-E194) tails being pushed upwards, while α8 helices (residues P143-R159) move toward and α9/α10 helices (residues E178-D188/S190-E194) drift away from each other. Moreover, α1 (residues T15-N27) is stretched away from the skids and α3 (residues N70-L81) bases move closer to each other, resulting in a convergence of the GPN motifs. Overall, the allosteric movement of both *Sa*GPN states takes primarily place alongside the dimer interface. This also includes α1 (residues T15-N27) and α12 (residues L228-L239), which are not part of the dimer interface yet contribute significantly to the scissors motion, with displacements of 18.9° and 12.0°, respectively (Fig. 6C). Additionally, the γ-phosphate-GPN interaction is lost after hydrolysis, resulting in a swivel-off motion of the GPN-loop including the preceding α2 skids (residues V58-Y65), thus relaxing the whole assembly.

In solution analysis of *Sa*GPN allosteric movements

We furthermore probed the nucleotide-dependent changes of *Sa*GPN in solution by hydrogen-deuterium exchange (HDX) coupled to mass spectrometry. Coordination of the GDP and GppNHp nucleotides by *Sa*GPN is evidenced by a reduction in HDX in the G1 motif (peptide F3-L17), compared to the apo state (Fig. 7A and B; Fig. S4 and S5). Further conformational changes induced by GDP encompass helices α2 and α3 (peptides Y55-D60 and V74-L79, Fig. 7B; Fig. S4), both of which exhibit elevated HDX, and helix 6 and the subsequent strand β5 displaying reduced HDX (Fig. S4). Binding of GppNHp to *Sa*GPN evokes even more pronounced perturbations in HDX (Fig. S5). In addition to helices α3 and α4 incorporating more deuterium similar to the GDP-bound state, the roof-constituting helices α9-α11 (residues E178-N212) specifically for GppNHp/*Sa*GPN incorporate more deuterium (peptides L151-L168, R183-L195, and K207-L213; Fig. 7A and B). The changed incorporation rates go along with the significant movement of the roof helices (residues E178-N212) observed in the crystal structure. Most notably, upon GTP hydrolysis, helix α11 (E200-N212) moves with its center-of-mass (COM) by 3.0 Å

relative to its core *Sa*GPN domain. This and the changed packing of the core domains in the dimer cause an overall displacement of the α11 helices by 6.3 Å (COM) relative to each other with chain A as point of reference (Fig. 6B). Interestingly, the roof top helices of *Sa*GPN are not involved in crystal contacts, neither in the GDP- nor in the GppNHp-bound form. On the contrary, parts of helices α6 (peptide A111-S121) close to the subunit interface incorporate less deuterium, suggesting an altered topology of *Sa*GPN in presence of GppNHp. The large conformational changes between GDP and GppNHp-bound *Sa*GPN are also reflected in a direct comparison of their HDX behaviors (Fig. S6). Notably, while changes in HDX can be interpreted as changes in the structural assembly or movement of the region affected, the converse is no confirmation for a lack of conformational changes as exemplified by the skid region.

GTPase activity and phosphorylation of *Sa*GPN are not required for motility

In order to investigate the impact of nucleotide binding by *Sa*GPN *in vivo*, we conceived of mutants that, based on our crystal structure should be defective in GTP binding or hydrolysis (i.e., K14, D38A, D102A, and GPN-AAA). All of these mutants were expressed in *Escherichia coli*, purified, and tested for their GTPase activity, which was almost completely lost in all mutant proteins (Fig. S2A). When complementing the Δ*saGPN* strain with these *Sa*GPN mutants, all mutant proteins could be expressed in *S. acidocaldarius* (Fig. S2C). Surprisingly, compared with the control, the swimming defect could partly be restored by these *Sa*GPN mutants, although they were catalytically inactive (Fig. S7A). MANT-GTP binding assays indicated that these mutants still retained 20–50% of the wild-type GTP-binding capacity (Fig. S7C), suggesting that nucleotide binding is unusually strong for a GTPase, since nucleotides generally fail to bind to G1/Walker A mutants (33). Apparently, *Sa*GPN functions *in vivo* independently from its GTPase activity.

Notably, a phosphoproteomics study of *S. acidocaldarius* demonstrated that *Sa*GPN can be apparently phosphorylated at Y59 (26). A regulatory function for Y59 by (de)phosphorylation can be ruled out as shown by the Y59F mutants swarming assay (Fig. S7B). The *in vivo* and *in vitro* function of *Sa*GPN is unaffected by the removal of this putative site of phosphorylation (Fig. S7B).

DISCUSSION

GPNs are present in most organisms, in archaea occurring mostly as a single paralog and in eukaryotes in triple paralogs (GPN1–GPN3), where they perform non-redundant essential functions (14). However, little is known about this class of GTPases compared to other guanosine nucleoside phosphate hydrolyzing enzymes like small GTPases or G proteins. In 2007, GPNs were introduced by Gras et al. (13) as a self-activating, homodimeric GTPase family alongside the first GPN crystal structure from *P. abyssi*. Our data now provide new insights into this understudied protein family of GPN loop GTPases.

We were able to solve crystal structures of *Sa*GPN loaded with either an analog of GTP or the hydrolysis product GDP. These structures reveal major changes of the GPN quaternary assembly upon nucleotide hydrolysis, which have not been observed before. After GTP hydrolysis and loss of the γ-phosphate, the substrate pocket switches from a closed into an open state by losing the interaction to the second protomer's GPN motif. When making a comparison between *Sa*GPN and the only other characterized archaeal GPN, *Pa*GPN (sequence identity 33%) from the hyperthermophile *P. abyssi*, we found that one of the most prominent differences is found in the region of the roof helices. These helices differ in length and, most importantly, adopt alternative orientations when comparing the *Pa*GPN and *Sa*GPN homodimers (Fig. S11). Additionally, the so-called skid region comprising the α2-helix undergoes a prominent conformational change upon transition from the GTP- to the GDP-bound state of *Sa*GPN, but not of *Pa*GPN (Movie S1; Fig. S11). This lack of observable conformational changes for *Pa*GPN may be caused by the enhanced hyperthermophilic nature of *P. abyssi* (growth optimum at 96°C compared to 80°C for *S. acidocaldarius*), which necessitates thermal activation. For example, conformational differences between heat-activated and non-activated enzymes have

been found for the dimeric homoserine dehydrogenase from *Sulfurisphaera tokodaii* (34). Interestingly, the inactive homoserine dehydrogenase bound a non-cognate $NADP^+$ cofactor and underwent conformational changes near its active site with release of this cofactor only upon heat activation. Although in the *Pa*GPN structures, a partial closing of the catalytic cavity like in *Sa*GPN was not observable, minor changes in the quaternary structure could still reflect the onset of allosteric changes. Nevertheless, other factors like crystal packing or harboring a second Mg^{2+} ion in the *Pa*GPN GDP structure (PDB: 1YRB) may alternatively cause arrest of *Pa*GPN in a single quaternary state.

Additionally, GPNs, including *Sa*GPN, are annotated as GTPases activated by dimerization (GAD), which are G proteins known for exhibiting GTPase activity in the absence of GEFs or guanosine activation protein (GAP) (3). However, while dissociation constants of GADs to guanosine nucleotides are reported to be within the low µM range, *Sa*GPN features an exceedingly high affinity for guanosine nucleotides in the low nM range, matching the values of GTPases employing GEFs (3, 10, 35, 36). Furthermore, GPNs are dimeric regardless of their nucleotide-bound state, distinguishing them from GADs (10, 13). Accordingly, GPNs are different from most other G proteins and form a distinct class of their own as defined by the major part of the IPR004130 family. For example, Q106 of *Sa*GPN is the structural pendant of Q61 from the switch II region of the small G-protein Ras that contributes to catalysis by interacting with the γ-phosphate via a bridging water molecule. In the *Sa*GPN dimer, Q106 is placed in its interface region. Here, the side chain of Q106 H-bonds to that of N70′ in the GDP-bound state, whereas it flips to form a staggering interaction in the GTP-bound state. Despite conservation, Q61 in Ras exerts a different function than Q106, as in the *Sa*GPN dimer, this residue entangles the two nucleotide binding states via the catalytically relevant N70 residues. Notably, the comparably low GTPase activity appears to be a hallmark of archaeal GPN-loop GTPases as shown by turnover rates of 6.4 nmol/min/mg (65°C) for *Sa*GPN and 12 nmol/min/mg (80°C) for the orthologous *Pa*GPN (3, 13). These turnover rates may still allow a biological function for archaeal GPN-loop GTPases in the absence of canonic GAPs and/or GEFs, although the complete absence of at least an allosteric partner enhancing archaeal GPNs GTPase function cannot be ruled out. The latter implies that the term "self-activated" as originally introduced by Gras et al. might not be appropriate. Another issue is given by the fact that kinetic experiments for *Sa*GPN had to be performed at 65°C for technical reasons, i.e., below the optimal growth temperature of 75–80°C of *S. acidocaldarius*. Accordingly, nucleotide exchange rates, affinities, and the Michaelis-Menten-like kinetic analysis of GTP hydrolysis might be affected. However, due to the apparently low intrinsic GTPase activity and high guanosine nucleotide affinities, both nucleotide-bound states of *Sa*GPN have considerable half-lives. This may be particularly relevant for the GTP-bound state to exert a potential function in regulation and protein-protein interaction. For comparison, eukaryotic GPNs are not only involved in sister chromatid cohesion (human) and, like a chaperone, the assembly of the 12-subunit RNA polymerase II (*S. cerevisiae*), but also in other biological processes such as mitochondrial homeostasis and ribosome biogenesis (14, 16, 37, 38). Given this broad range of functions for eukaryotic GPNs, archaeal GPNs like *Sa*GPN with their conformational switching between GTP- and GDP-bound states may exert a similar type of function, e.g., by acting as a chaperone during multi-protein assembly.

Deletion of GPNs is known to be lethal in eukaryotes (13, 14, 18), but interestingly not in the case of the crenarchaeote *S. acidocaldarius*. This implies that the biological context of archaeal GPNs differs from that of their eukaryotic counterparts. The deletion of *saGPN* in *S. acidocaldarius* yields a phenotype with a highly reduced motility despite lacking any apparent growth defect. The motility phenotype is not strictly dependent on the catalytic capability of *Sa*GPN to hydrolyze GTP, as different mutations in the G-box, which cause impaired nucleotide binding and/or lack the ability to hydrolyze GTP, resulted only in gradually diminished motility (Fig. 2; Fig. S7). These observations suggest that the intrinsic GTPase activity of *Sa*GPN is dispensable for motility of *S. acidocaldarius*. However, it is unclear whether GTPase activity is of relevance to other processes

that were not investigated in this study. Moreover, we showed that a mutation within the α2-helix, Y59D, affects swarming similar to the G-box mutants besides decreasing catalytic activity by ~70%. Nevertheless, this correlation between GTPase activity and motility may be serendipitous as G-box mutants lack GTPase activity while still exhibiting motility. Thus, the decreased motility of the surface-exposed Y59D mutant may be due to distorted interactions with unknown downstream partners, which control synthesis, assembly or, activity of the *S. acidocaldarius* archaellum.

A broader biological function of *Sa*GPN than its requirement for motility is indicated by our proteomics analysis of the *saGPN* knockout in *S. acidocaldarius*. Relative protein quantification by timsTOF of the Δ*saGPN* mutant compared to the wild-type strain revealed an unexpectedly large number of proteins being affected in their levels. Notably, these proteins are assigned to a highly diverse set of biological functions, which are mostly not directly linked to motility. Compared to the pronounced metabolic adaptation of the *S. acidocaldarius* proteome upon starvation (Fig. 3A), the even larger impact of the *saGPN* deletion suggests a key role of *Sa*GPN in protein homeostasis of *S. acidocaldarius*. Interestingly, the *saGPN* deletion also resulted in a complete loss of an ortholog of the universal stress protein family, UspD (Uniprot: Q4JA32). UspD was undetectable in the proteome of the *saGPN* deletion strain, although it was highly abundant in all WT samples (Table S4). Interestingly, UspD has only very limited sequence identity (8–11%) to the known Usp proteins of *S. acidocaldarius*, UspA-UspC, although it shares the typical Usp fold with the latter (Fig. S8). Given that *Sa*GPN was found together with UspA in a previous co-IP assay (25), it is notable that UspA levels are almost unaffected in the *saGPN* deletion strain.

Since we were able to obtain snapshots of *Sa*GPNs catalytic cycle, we propose a mechanism, which we call "lock-switch-rock" mechanism (Fig. 8) that assigns a catalytic function to the conserved Gly-Pro-Asn motif of GPN-loop GTPases. The GTP-bound state as visualized by our structure of the *Sa*GPN·GppNHp complex (7ZHF) shows a partial closure of the nucleotide-binding site by the embracing interaction between the GPN motif from the opposing *Sa*GPN protomer and the γ-phosphate (Fig. 5). As a consequence, a water molecule that is coordinated between D38 from the G2 motif and N70′ of the GPN motif is now found in an inline attack position for initiating the S_N2 reaction at the γ-phosphate. When deprotonated, D38 acts here as a general base for transient formation of the attacking hydroxide nucleophile like the corresponding residue in the G2 box of other SIMIBI GTPases (13). Overall, the GTP-bound state of GPN-loop GTPases with its closed GPN-lid corresponds to a 'locked' state. The quaternary structure changes in a 'rock'-like motion upon the 'switch' event, i.e., nucleotide hydrolysis and phosphate release. Its trigger may be other protein interaction partners as exemplified by eukaryotic GPN-loop GTPases during RNA polymerase II assembly, but these factors, if they exist, are yet not known for *Sa*GPN. Unfortunately, efforts to crystallize a transition mimic of the *Sa*GPN·GTP complex, i.e., the "switch" state, failed. Accordingly, we do not yet know the exact nature of the "switch" state of GPN-loop GTPases as switching may be triggered by either the hydrolysis event or the release of phosphate from the *Sa*GPN·GDP·P_i state. In any case, the "rocking" motion causes an increased distance between the nucleotide and the GPN loop besides changes at the distal side of the GPN dimer along the roof helices. Interestingly, the interaction between D38/N70′ with a water molecule is even maintained in the GDP-bound form. This corroborates the notion that N70 of the GPN loop is essential to position the attacking water in the GTP-bound state, while the Gly-Pro motif caps helix α3′ at its N-terminus, so that it points toward the γ-phosphate and D38 for stabilizing dipole-ion interactions.

Overall, hydrolysis to GDP with phosphate release removes these intimate interactions between the GPN motif/helix α3′ and the γ-phosphate and subsequently drives the switching from a locked state toward a relaxed state with a concurrent rocking motion that affects the whole GPN dimer (Fig. 8; Movie S1).

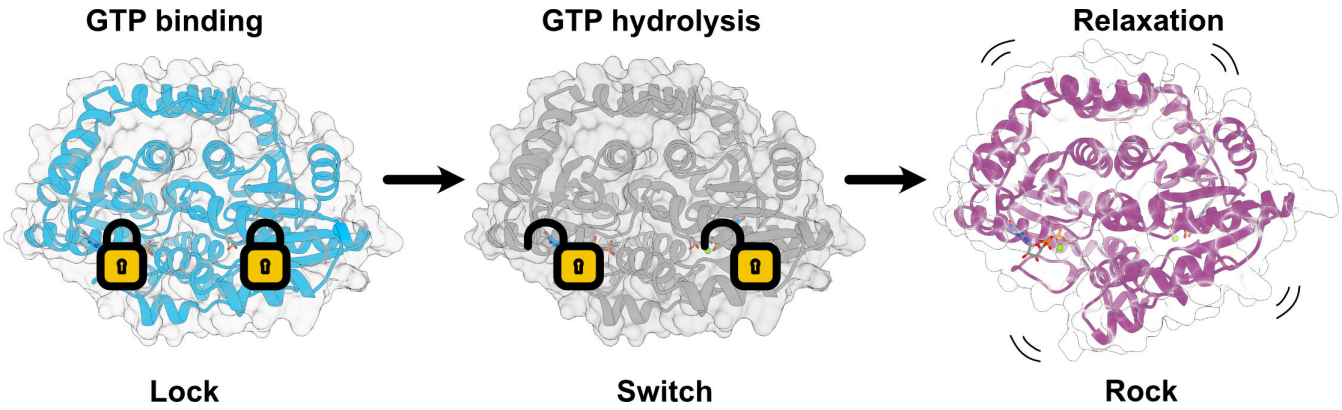

FIG 8 Lock-switch-rock (LSR) mechanism of *Sa*GPN. After GTP binding and before nucleotide hydrolysis, *Sa*GPN enters a locked state and stores tension throughout the structure. GTP hydrolysis leads to a loss of interaction in the GPN loop, which allows the transition from the locked state to a relaxed state in which a rocking motion is performed, stretching the GTPase assembly. Since these three steps represent the main catalytic states, we developed the lock-switch-rock mechanism, or LSR mechanism, for easy visualization and reference.

MATERIALS AND METHODS

Strains and growth conditions

S. acidocaldarius wild-type strain MW001 and all derived mutants in this study (Table S1) were cultivated at 75°C in Brock basal medium (pH 3.0–3.5, all salts for the media are from Roth) supplemented with 0.1% (wt/vol) NZ-amine (Sigma-Aldrich), 0.2% (wt/vol) dextrin (Roth), and 10 µg/mL uracil (Sigma-Aldrich) (39, 40). For *S. acidocaldarius* strains containing complementation plasmids, uracil was not needed.

Construction of *S. acidocaldarius* Δ*saGPN* mutant

The plasmid pSVA5117 for the markerless in-frame deletion mutant Δ*saGPN* (Table S1) was constructed as described previously (40). After transformation into *S. acidocaldarius* MW001, positive transformants were screened on first selection gelrite (Roth) plates lacking uracil, then grown on second selection gelrite plates with uracil and 5-FOA (100 µg/mL). Finally, Δ*saGPN* mutants were screened by colony PCR with checking primers (Table S2), and further sequencing.

Nutrient starvation assays and Western blots

Nutrient starvation assays and Western blots were performed as described (41). Briefly, at an OD_{600} of 0.4–0.5, overnight *S. acidocaldarius* cultures were collected at 75°C and re-suspended in Brock medium in the absence of NZ-amine and dextrin, followed by cultivation at 75°C. Samples for the Western blot analysis were taken at the indicated times.

RNA isolation and qRT-PCR

Total RNA samples were prepared with *S. acidocaldarius* cultures from nutrient starvation assays and qRT-PCR analysis was performed as described before (25). *secY* is a house-keeping gene, which was used as the reference gene for data normalization.

Transmission electron microscopy

Δ*saGPN* and MW001 strains were grown in 50 mL Brock's medium supplemented with 0.1% (wt/vol) NZ-amine, 0.2% (wt/vol) dextrin, and 10 µg/mL uracil. At an OD_{600} = 0.2–0.3, nutrient starvation was done for 4 h to induce archaellation. Afterward, cells were applied on freshly glow-discharged carbon/formvar coated copper grids (300 mesh,

Plano GmbH) and incubated for 30 s. This was repeated three times and excess liquid was blotted away. Cells were negatively stained with 2% (wt/vol) uranyl acetate (Ted Pella). Imaging was done with Hitachi HT7800 operated at 100 kV, equipped with an EMSIS XAROSA 20 Megapixel CMOS camera.

Motility assays

Motility assays were performed as described (40). *S. acidocaldarius* strain Δ*arnA* and Δ*arnR/R1* were used as a hypermotile and non-motile control, respectively. All strains originate from previous work (27, 42).

MANT-GTP binding assays

Binding of fluorescent MANT-GTP [(2′-(or-3′)-O-(N-methylanthraniloyl) guanosine 5′-triphosphate] (Jena-Bioscience) to *Sa*GPN was measured by monitoring fluorescence increase upon binding to the protein utilizing a Fluoromax-4 spectrofluorometer (Horiba). The excitation wavelength was set to 285 nm, and the emission wavelength was set to 450 nm. Slit widths for both excitation and emission were set to 10 nm. Binding was measured by incubation of 2.5 µM of *Sa*GPN/variants with MANT-GTP (5 µM) in buffer 50 mM MES pH 6.5, 150 mM NaCl, 5 mM $MgCl_2$ (Roth) at 25°C. Fluorescence was corrected for MANT-GTP fluorescence in the absence of proteins. Fluorescence of *Sa*GPN wild-type protein was set as 100% for relative MANT-GTP binding analysis.

Sample preparation for proteomic mass spectrometry analysis

Strains were grown and subjected to 30 min starvation conditions as described above. Subsequently, 200 µL cell pellet equivalents was lysed in 900 µL urea buffer (8 M urea dissolved in 0.1 M NH_4HCO_3). After that, the samples were disrupted using glass beads and a FastPrep-24 (MP Biomedicals) homogenizer (6.5 m/s, three cycles of 30 s, 5 min resting on ice between cycles) with subsequent centrifugation (50 min, 14,000 rpm). Supernatant concentration was determined using Bradford assay and afterwards resuspended with the cell debris to yield full cell samples again. Protein concentrations of full cell samples were estimated to be double the amount of the supernatant concentration. Samples were adjusted to 40 µL (8 M urea dissolved in 0.1 M NH_4HCO_3) containing 100–200 mg protein based on the concentration estimation and mixed with 1 µL TCEP (0.2 M) with subsequent incubation (1 h, 37°C, 1,000 rpm). About 1 µL iodoacetamide (0.4 M) was added and the solutions were incubated in the dark (30 min, 25°C, 500 rpm), followed by addition of 1 µL *N*-acetyl-cysteine (0.5 M). Samples were incubated (10 min, 25°C, 500 rpm) and diluted with 10.3 µL urea buffer (6 M urea dissolved in 0.1 M NH_4HCO_3), followed by the addition of 1.25 µL Lys C (0.2 mg/mL, 1:400 wt/wt) and digestion (4 h, 37°C). Solutions were diluted with 145.3 µL urea buffer (1.6 M urea dissolved in 0.1 M NH_4HCO_3), mixed with 2 µL trypsin (1:100 [wt/wt\), and digested (overnight, 37°C). pH was adjusted to <2 with 2.5 µL TFA (0.1% [vol/vol]). Samples were centrifuged (1 min, 14,000 rpm) and transferred to an equilibrated (0.1% [vol/vol] TFA) Chromabond C_{18} spin column (30 s, RT, 2,000 rpm). Peptides were eluted with 2 × 150 µL elution solution (50% [vol/vol] ACN, 0.1% [vol/vol] TFA), dried in a vacuum centrifuge (45°C, 4,000 rpm) and resuspended in 30 µL acetonitrile-TFA solution (10%[vol/vol], 0.1% [vol/vol]). Samples were then measured using a timsTOF mass spectrometer in collaboration with the MarMass facility of Philipps University Marburg.

Proteomics data evaluation

Data analysis of the timsTOF data was performed with MaxQuant 2.1.3 (43), sequence data for *S. acidocaldarius* DSM 639 (2,222 entries) were downloaded from Uniprot. MaxQuant parameters for label-free quantification (LFQ) analysis were set to a maximal

peptide mass of 4,000 Da, up to three modifications per peptide, up to three missed cleavages and the requirement of MS/MS for LFQ comparisons; all other parameters are set to default. Protein abundances of biological replicates of the Δ*saGPN* strain and the wild type strain were filtered by a two-tailed, homoscedastic Student's *t* test ($P < 0.05$) using a mean abundance difference of >10% as second criterion. The resulting lists of proteins were used for protein heat maps after filtering with a mean abundance difference of >50% and visualization by the pheatmap R package. Gene ontology enrichment analysis was carried out using the GOATOOLS library (44), version 1.2.3, in python. In this gene ontology enrichment analysis, terms are represented for molecular function (MF), cellular component (CC), and biological processes (BP) of the identified proteins. To implement the GOATOOLS library, gene ontology terms (GO-Terms) for *S. acidocaldarius* DSM639 were downloaded using the QuickGO annotation online tool (45). A total of 9,487 annotations were available for *S. acidocaldarius* DSM639 proteins covering 1,548 gene products (70%); the remaining 674 are mostly still assigned as "conserved archaeal proteins."

Sequence similarity network and *in silico* analysis

Generation of primary SSN data was done with the EFI-Enzyme Similarity Tool (46) web service (Uniprot Version: 2021_03; InterPro Version 87) based on InterPro Family IPR004130 (GPN-loop GTPases) with an UniRef90 restraint due to the size of the IPR (29). SSNs were generated for *E*-values 10^{-40} and 10^{-45} with subsequent cluster analysis (standard options) and colorization. Analysis and visualization of the generated network were performed with Cytoscape 3.8.2 using the yfiles organic layout (47).

Expression and purification of *Sa*GPN

Recombinant overexpression of N-terminal His_{10} tagged *Sa*GPN was performed in an expression media containing 10 g/L tryptone (Sigma-Aldrich), 10 g/L NaCl, 5 g/L yeast extract (Sigma-Aldrich) and 12.5 g/L lactose (Roth), employing BL21(DE3) Rosetta cells (37°C, 150 rpm, 18 h). Harvesting (20°C, 5,000 rpm, 20 min) was followed by resuspension in lysis buffer (150 mM NaCl, 50 mM Tris, 15 mM EDTA, pH = 8.0) and cell lysis via French Press. The lysate was centrifuged (18,000 rpm, 20°C, 20 min), the supernatant heat-treated (55°C, 15 min) and centrifuged again (18,000 rpm, 20°C, 20 min) before being loaded to a Ni-NTA column (5 mL). After sample application, the column was washed with nine column volumes wash buffer (150 mM NaCl, 50 mM Tris, 25 mM imidazole (Roth), 15 mM EDTA, pH = 8.0) and eluted with five column volumes elution buffer (150 mM NaCl, 50 mM Tris, 500 mM imidazole, pH = 8.0). The elution was concentrated (<2.5 mL) and applied to a HiLoad 16/60 Superdex 200 pg (GE Healthcare) size-exclusion column, which had been equilibrated with running buffer (150 mM NaCl, 50 mM Tris, 10 mM $MgCl_2$, pH = 8.0). Fractions containing *Sa*GPN with a 260/280 nM ratio ≤0.55 were collected, concentrated, frozen in liquid nitrogen and stored at −80°C. Protein concentrations were determined by NanoDrop spectrophotometry for the apo-state; Bradford assays were used to confirm concentrations of the nucleotide-bound *Sa*GPN.

GTPase assays

GTPase activity of *Sa*GPN was measured using the Malachite green assay (48). For a 600-µL reaction mix at 65°C 60 µL 10× *Sa*GPN (final concentration 10 µM) was mixed with 60 µL 10× GTP stocks and adjusted to 600 µL with a buffer containing 50 mM Tris pH 8.0, 150 mM NaCl, and 10 mM $MgCl_2$. For each time point, 60 µL of that mixture was taken and quenched with the Malachite green reaction solution. After incubation for 30 min, the Malachite green solution developed its color and was measured. The absorption at 620 nm was measured and a phosphate standard was used for quantification.

PP2A phosphatase activity assays

Serine/threonine phosphatase activity assay of PP2A was performed using the artificial p-peptide RRA(pT)VA substrate (Promega) in lysis buffer containing 1 mM $MnCl_2$ at 70°C (26). The release of free phosphate from artificial p-peptides was measured using the Malachite green assay as described above (48).

Crystallization of *Sa*GPN

Crystallization screens were performed as sitting drop experiments with 24.5 mg/mL *Sa*GPN (Saci_1281, Uniprot: Q4J9A7) that includes the N-terminal His_{10}-tag (MHHHHHHHHHHLEVLFQGPS) and 3 mM nucleotide in running buffer mixed in a 1:1 ration with the respective crystallization condition, using 0.3 µL of each. Crystallization was observed in different conditions of all JCSG Core suits (NeXtal); however, the crystal used for the structure determination of *Sa*GPN(GppNHp) crystallized in the presence of 200 mM NaCl, 100 mM NaOAc (pH = 4.6) and 30% (vol/vol) MPD (JCSG Core III/G11) after 24 h. *Sa*GPN(GDP) was crystallized in the presence of 200 mM $MgCl_2$, 100 mM MES (pH 5.5), and 40% (vol/vol) PEG 400 (JCSG Core II/G3) after 24 h. *Sa*GPN crystallization in the apo-state was also attempted but resulted only in weakly diffracting crystals (resolution > 7 Å) and was therefore not suitable for structure determination. Crystals of the apo state grew in many different conditions of the NeXtal JCSG Core suits I–IV.

Structure determination of *Sa*GPN

Data collection was performed with the Swiss Light Source at Paul Scherrer Institute in Switzerland. Phasing was done employing molecular replacement using BALBES (GppNHp state) and PHASER (GDP state) of the ccp4 pipeline, followed by model building with PDB-REDO and ARP/wARP (49–53), revealing one molecule per asymmetric symmetry unit for the GppNHp-bound and three for the GDP-bound state. The physiological dimeric state for the *Sa*GPN·GppNHp complex can be generated by applying crystal symmetry. For the *Sa*GPN·GDP complex, the dimer is represented by chains A/C and chains B/B′. Structure refinements were done by multiple rounds of manual model building with Coot followed by phenix.refine (54, 55). For both nucleotide-bound states, we were able to build residues Y2-A240 and some residues of the His_{10}-tag, resulting in only 14 disordered amino acids at the C-terminus. For the GDP state, structural imposition of the *Sa*GPN chains reveals an r.m.s.d. of 0.23 Å for chains A vs C and chains B vs C that is only slightly increased to 0.24 Å for chains A vs B. Structural analysis was carried out with PyMOL and visualization with Chimera 1.15 (56, 57). F_o-F_c omit maps showing the density around the nucleotides are shown in Fig. S9. Structural morphing between the two nucleotide-bound states was done by Chimera 1.15, rendering of the resulting poses (Movie S1) by Blender 3.6 (blender.org) and Molecular Nodes (bradyajohnston.github.io/MolecularNodes).

Isothermal titration calorimetry

ITC experiments were carried out with purified *Sa*GPN concentrated to a 100 µM stock solution, corresponding to monomeric protein, and stored at −80°C as 300 µL aliquots. Sample cell had a volume of 200 µL. For each run, *Sa*GPN aliquots and nucleotide stocks (800 µM) were thawed just prior to measurement. The sample cell of the MicroCal PEAQ-ITC, (Malvern) was heated to 65°C or 25°C, before in total 40 µL of nucleotides was added from the injection syringe using constant stirring (52 injections). Evaluation and visualization of ITC data were performed with the Malvern evaluation software. A 1:1 model was used for fitting and calculating K_D/*N*-values. It should be mentioned that the high affinity for guanosine nucleotides caused low data resolution. Accordingly, fittings showed only a moderate significance for calculated *N*-values.

Hydrogen-deuterium exchange mass spectrometry

HDX-MS experiments on *Sa*GPN were carried out similarly as described previously (58). The samples contained 50 µM *Sa*GPN and 5 mM nucleotides (GDP, GppNHp) in a buffer containing 50 mM Tris-Cl pH 8.0, 150 mM NaCl, and 10 mM $MgCl_2$. Preparation of the HDX reactions was aided by a two-arm robotic autosampler (LEAP technologies). For deuterated samples, 7.5 µL of *Sa*GPN solution (with or without nucleotides) was supplemented with 67.5 µL of D_2O - containing buffer to start the exchange reaction. After 10, 30, 100, 1,000 or 10,000 s at 25°C, samples (55 µL) were taken from the reaction and mixed with 55 µL of quench buffer (400 mM KH_2PO_4/H_3PO_4, 2 M guanidine-HCl, pH 2.2) kept at 1°C. About 95 µL of the resulting mixture was injected into an ACQUITY UPLC M-Class System with HDX Technology (Waters [59]). Undeuterated samples of *Sa*GPN were prepared similarly by 10 - fold dilution of the protein solution with H_2O - containing buffer. Samples were flushed out of the loop (50 µL) with H_2O + 0.1% (vol/vol) formic acid (flow rate of 100 µL/min) and guided to a column (2 mm × 2 cm) packed with immobilized porcine pepsin and kept at 12°C for proteolytic digestion. The peptic peptides thus generated were collected on a trap column (2 mm × 2 cm) filled with POROS 20 R2 material (Thermo Scientific) kept at 0.5°C. After 3 min, the trap column was placed in line with an ACQUITY UPLC BEH C18 1.7 µM 1.0 × 100 mm column (Waters), and the peptides were eluted at 0.5°C column temperature using a gradient of H_2O + 0.1% (vol/vol) formic acid (A) and acetonitrile + 0.1% (vol/vol) formic acid (B) at a flow rate of 30 µL/min as follows: 0–7 min/95–65% A, 7–8 min/65–15% A, 8–10 min/15% A, 10–11 min/5% A, and 11–16 min/95% A. The peptides were ionized with an electrospray ionization source (250°C capillary temperature, 3.0 kV spray voltage) and mass spectra acquired in positive ion mode over a range of 50 to 2,000 *m/z* on a G2-Si HDMS mass spectrometer with ion mobility separation (Waters), using Enhanced High Definition MS ($HDMS^E$) or High Definition MS (HDMS) mode for undeuterated and deuterated samples, respectively (60, 61). Lock mass correction was implemented with [Glu1]-Fibrinopeptide B standard (Waters). During each chromatographic run, the pepsin column was washed three times with 80 µL of 4% (vol/vol) acetonitrile and 0.5 M guanidinium chloride, and blank injections were performed between each sample. All measurements were carried out in triplicate.

*Sa*GPN peptides were identified from the undeuterated samples with the software ProteinLynx Global SERVER 3.0.1 (PLGS, Waters), using the amino acid sequence of *Sa*GPN, porcine pepsin and their reverted sequences as database, and their deuterium incorporation determined with the software DynamX 3.0 (Waters) as described previously (58).

ACKNOWLEDGMENTS

We thank the beamline staff of the Swiss Light Source (SLS), PSI, Villigen, Switzerland, for support, the staff of Marburg crystallization facility (MarXtal) for technical support, and Marta Rodriguez-Franco for assistance during EM imaging. We acknowledge support by the DFG through the DFG-core facility for interactions, dynamics, and macromolecular assembly (WS, project 324652314 to Prof. Gert Bange, Marburg). We also thank Franziska Sendker and Dr. Georg Hochberg (Max-Planck Institute for Terrestrial Microbiology, Marburg) for providing access and support to massphotometry.

This work was supported by the Chinese Scholarship Council (PhD scholarship to X.Y.). L.-O.E. and S.-V.A. thank the Life program of the Volkswagen Foundation for funding. We would also like to thank the EM facility at the Faculty of Biology, University of Freiburg, for access to the TEM for generation of data. The TEM (Hitachi HT7800) was funded by the DFG grant (project number 426849454) and is operated by the University of Freiburg, Faculty of Biology, as a partner unit within the Microscopy and Image Analysis Platform (MIAP) and the Life Imaging Center (LIC), Freiburg.

L.K., X.Y., M.S.V., S.V.A., and L.-O.E. designed research; L.K. and X.Y. performed experiments; W.S. performed and analyzed HDX-MS experiments; L.K., M.W., and L.-O.E.

analyzed proteome data; L.K., X.Y., M.S.V., and L.-O.E. analyzed data; and L.K., X.Y., M.W., S.-V.A., and L.-O.E. wrote the paper.

AUTHOR AFFILIATIONS

[1]Department of Chemistry, Philipps University, Marburg, Germany
[2]University of Freiburg, Institute of Biology, Molecular Biology of Archaea, Freiburg, Germany
[3]Center for Synthetic Microbiology (SYNMIKRO), Karl-von-Frisch-Strasse, Marburg, Germany
[4]Spemann Graduate School of Biology and Medicine, University of Freiburg, Freiburg, Germany

AUTHOR ORCIDs

Lukas Korf http://orcid.org/0000-0003-4882-1177
Sonja-Verena Albers http://orcid.org/0000-0003-2459-2226
Lars-Oliver Essen http://orcid.org/0000-0003-4272-4026

FUNDING

Funder	Grant(s)	Author(s)
Volkswagen Foundation (VolkswagenStiftung)	Az-96727	Lukas Korf Xing Ye Marian S. Vogt Mohamed Watad Sonja-Verena Albers Lars-Oliver Essen
Deutsche Forschungsgemeinschaft (DFG)	324652314	Wieland Steinchen

AUTHOR CONTRIBUTIONS

Lukas Korf, Formal analysis, Investigation, Methodology, Project administration, Validation, Visualization, Writing – original draft, Writing – review and editing | Xing Ye, Investigation, Methodology, Writing – original draft | Marian S. Vogt, Project administration, Supervision, Writing – review and editing | Wieland Steinchen, Formal analysis, Methodology, Writing – original draft, Writing – review and editing | Mohamed Watad, Methodology, software | Chris van der Does, Validation | Maxime Tourte, Methodology | Shamphavi Sivabalasarma, Visualization | Sonja-Verena Albers, Conceptualization, Data curation, Funding acquisition, Project administration, Supervision, Writing – review and editing | Lars-Oliver Essen, Conceptualization, Data curation, Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Writing – review and editing

DATA AVAILABILITY

Protein structure data are deposited in the PDB and can be accessed via the codes 7ZHF (*Sa*GPN GppNHp structure) and 7ZHK (*Sa*GPN GDP structure). Remaining data are found in the supplementary material.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Supplemental material (mBio00859-23-s0001.docx). Fig. S1 to S12, Tables S1 to S3, and supplemental legends.

Table S4 (mBio00859-23-s0002.xlsx). timsTOF data and GO term analysis of S. acidocaldarius WT and saGPN knockout strains.
Table S5 (mBio00859-23-s0003.xlsx). HDX data and its analysis.
Movie S1 (mBio00859-23-s0004.mp4). Morph of the SaGPNs•GppNHp state towards its GDP-bound state, visualizing the major allosteric changes upon nucleotide hydrolysis.

REFERENCES

1. Leipe DD, Wolf YI, Koonin EV, Aravind L. 2002. Classification and evolution of P-loop GTPases and related ATPases. J Mol Biol 317:41–72. https://doi.org/10.1006/jmbi.2001.5378
2. Bourne HR, Sanders DA, McCormick F. 1991. The GTPase superfamily: conserved structure and molecular mechanism. Nature 349:117–127. https://doi.org/10.1038/349117a0
3. Wittinghofer A, Vetter IR. 2011. Structure-function relationships of the G domain, a canonical switch motif. Annu Rev Biochem 80:943–971. https://doi.org/10.1146/annurev-biochem-062708-134043
4. Verstraeten N, Fauvart M, Versées W, Michiels J. 2011. The universally conserved prokaryotic GTPases. Microbiol Mol Biol Rev 75:507–542, https://doi.org/10.1128/MMBR.00009-11
5. Neves SR, Ram PT, Iyengar R. 2002. G protein pathways. Science 296:1636–1639. https://doi.org/10.1126/science.1071550
6. Bange G, Sinning I. 2013. SIMIBI twins in protein targeting and localization. Nat Struct Mol Biol 20:776–780. https://doi.org/10.1038/nsmb.2605
7. Simanshu DK, Nissley DV, McCormick F. 2017. RAS proteins and their regulators in human disease. Cell 170:17–33. https://doi.org/10.1016/j.cell.2017.06.009
8. Gray JL, von Delft F, Brennan PE. 2020. Targeting the small GTPase Superfamily through their regulatory proteins. Angew Chem Int Ed 59:6342–6366. https://doi.org/10.1002/anie.201900585
9. Kostenis E, Pfeil EM, Annala S. 2020. Heterotrimeric GQ proteins as therapeutic targets. J Biol Chem 295:5206–5215. https://doi.org/10.1074/jbc.REV119.007061
10. Gasper R, Meyer S, Gotthardt K, Sirajuddin M, Wittinghofer A. 2009. It takes two to tango: regulation of G proteins by dimerization. Nat Rev Mol Cell Biol 10:423–429. https://doi.org/10.1038/nrm2689
11. Bange G, Kümmerer N, Grudnik P, Lindner R, Petzold G, Kressler D, Hurt E, Wild K, Sinning I. 2011. Structural basis for the molecular evolution of SRP-GTPase activation by protein. Nat Struct Mol Biol 18:1376–1380. https://doi.org/10.1038/nsmb.2141
12. Jagath JR, Rodnina MV, Wintermeyer W. 2000. Conformational changes in the bacterial SRP receptor FtsY upon binding of guanine nucleotides and SRP. J Mol Biol 295:745–753. https://doi.org/10.1006/jmbi.1999.3427
13. Gras S, Chaumont V, Fernandez B, Carpentier P, Charrier-Savournin F, Schmitt S, Pineau C, Flament D, Hecker A, Forterre P, Armengaud J, Housset D. 2007. Structural insights into a new homodimeric self-activated GTPase family. EMBO Rep 8:569–575. https://doi.org/10.1038/sj.embor.7400958
14. Alonso B, Beraud C, Meguellati S, Chen SW, Pellequer JL, Armengaud J, Godon C. 2013. Eukaryotic GPN-loop GTPases paralogs use a dimeric assembly reminiscent of archeal GPN. Cell Cycle 12:463–472. https://doi.org/10.4161/cc.23367
15. Guerrero-Serrano G, Castanedo L, Cristóbal-Mondragón GR, Montalvo-Arredondo J, Riego-Ruíz L, DeLuna A, De Las Peñas A, Castaño I, Calera MR, Sánchez-Olea R. 2017. Npa3/ScGpn1 carboxy-terminal tail is dispensable for cell viability and RNA polymerase II nuclear targeting but critical for microtubule stability and function. Biochim Biophys Acta Mol Cell Res 1864:451–462. https://doi.org/10.1016/j.bbamcr.2016.12.010
16. Niesser J, Wagner FR, Kostrewa D, Mühlbacher W, Cramer P. 2016. Structure of GPN-loop GTPase Npa3 and implications for RNA polymerase II assembly. Molecular and Cellular Biology 36:820–831. https://doi.org/10.1128/MCB.01009-15
17. Wittinghofer A, Vetter IR. 2011. Structure-function relationships of the G domain, a canonical switch motif. Annu Rev Biochem 80:943–971. https://doi.org/10.1146/annurev-biochem-062708-134043
18. Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, Dow S, Lucau-Danila A, Anderson K, André B, et al. 2002. Functional profiling of the *Saccharomyces Cerevisiae* genome. Nature 418:387–391. https://doi.org/10.1038/nature00935
19. Cristóbal-Mondragón GR, Lara-Chacón B, Santiago Á, De-la-Rosa V, González-González R, Muñiz-Luna JA, Ladrón-de-Guevara E, Romero-Romero S, Rangel-Yescas GE, Fernández Velasco DA, Islas LD, Pastor N, Sánchez-Olea R, Calera MR. 2019. FRET-based analysis and molecular modeling of the human GPN-loop GTPases 1 and 3 heterodimer unveils a dominant-negative protein complex. FEBS J 286:4797–4818. https://doi.org/10.1111/febs.14996
20. Lembo F, Pero R, Angrisano T, Vitiello C, Iuliano R, Bruni CB, Chiariotti L. 2003. MBD in, a novel MBD2-interacting protein, relieves MBD2 repression potential and reactivates transcription from methylated promoters. Mol Cell Biol 23:1656–1665. https://doi.org/10.1128/MCB.23.5.1656-1665.2003
21. Nitta M, Saijo M, Kodo N, Matsuda T, Nakatsu Y, Tamai H, Tanaka K. 2000. A novel cytoplasmic GTPase XAB1 interacts with DNA repair protein XPA. Nucleic Acids Res 28:4212–4218. https://doi.org/10.1093/nar/28.21.4212
22. Boulon S, Pradet-Balade B, Verheggen C, Molle D, Boireau S, Georgieva M, Azzag K, Robert M-C, Ahmad Y, Neel H, Lamond AI, Bertrand E. 2010. HSP90 and its R2TP/Prefoldin-like cochaperone are involved in the cytoplasmic assembly of RNA polymerase II. Mol Cell 39:912–924. https://doi.org/10.1016/j.molcel.2010.08.023
23. Forget D, Lacombe A-A, Cloutier P, Al-Khoury R, Bouchard A, Lavallée-Adam M, Faubert D, Jeronimo C, Blanchette M, Coulombe B. 2010. The protein interaction network of the human transcription machinery reveals a role for the conserved GTPase RPAP4/GPN1 and microtubule assembly in nuclear import and biogenesis of RNA polymerase II. Mol Cell Proteomics 9:2827–2839. https://doi.org/10.1074/mcp.M110.003616
24. Staresincic L, Walker J, Dirac-Svejstrup AB, Mitter R, Svejstrup JQ. 2011. GTP-dependent binding and nuclear transport of RNA polymerase II by NPA3 protein. J Biol Chem 286:35553–35561. https://doi.org/10.1074/jbc.M111.286161
25. Ye X, Vogt MS, van der Does C, Bildl W, Schulte U, Essen L-O, Albers S-V. 2020. The phosphatase PP2A interacts with ArnA and ArnB to regulate the oligomeric state and the stability of the ArnA/B complex. Front Microbiol 11:1849. https://doi.org/10.3389/fmicb.2020.01849
26. Reimann J, Esser D, Orell A, Amman F, Pham TK, Noirel J, Lindås A-C, Bernander R, Wright PC, Siebers B, Albers S-V. 2013. Archaeal signal transduction: impact of protein phosphatase deletions on cell size, motility, and energy metabolism in *Sulfolobus acidocaldarius*. Mol Cell Proteomics 12:3908–3923. https://doi.org/10.1074/mcp.M113.027375
27. Reimann J, Lassak K, Khadouma S, Ettema TJG, Yang N, Driessen AJM, Klingl A, Albers S-V. 2012. Regulation of archaella expression by the FHA and von willebrand domain-containing proteins ArnA and ArnB in *Sulfolobus acidocaldarius*. Mol Microbiol 86:24–36. https://doi.org/10.1111/j.1365-2958.2012.08186.x
28. Hoffmann L, Anders K, Bischof LF, Ye X, Reimann J, Khadouma S, Pham TK, van der Does C, Wright PC, Essen L-O, Albers S-V. 2019. Structure and interactions of the archaeal motility repression module ArnA-ArnB that modulates archaellum gene expression in *Sulfolobus acidocaldarius* J Biol Chem 294:7460–7471. https://doi.org/10.1074/jbc.RA119.007709
29. Gerlt JA, Bouvier JT, Davidson DB, Imker HJ, Sadkhin B, Slater DR, Whalen KL. 2015. Enzyme function initiative-enzyme similarity tool (EFI-EST): a web tool for generating protein sequence similarity networks. Biochim Biophys Acta 1854:1019–1037. https://doi.org/10.1016/j.bbapap.2015.04.015
30. Minaker SW, Filiatrault MC, Ben-Aroya S, Hieter P, Stirling PC. 2013. Biogenesis of RNA polymerases II and III requires the conserved GPN small GTPases in *Saccharomyces cerevisiae*. Genetics 193:853–864. https://doi.org/10.1534/genetics.112.148726
31. Bischof LF, Haurat MF, Hoffmann L, Albersmeier A, Wolf J, Neu A, Pham TK, Albaum SP, Jakobi T, Schouten S, Neumann-Schaal M, Wright PC,

Kalinowski J, Siebers B, Albers S-V. 2018. Early response of *Sulfolobus acidocaldarius* to nutrient limitation. Front Microbiol 9:3201. https://doi.org/10.3389/fmicb.2018.03201
32. Kanade M, Chakraborty S, Shelke SS, Gayathri P. 2020. A distinct motif in a Prokaryotic small Ras-like GTPase highlights unifying features of Walker B motifs in P-loop NTPases. J Mol Biol 432:5544–5564. https://doi.org/10.1016/j.jmb.2020.07.024
33. Hanson PI, Whiteheart SW. 2005. AAA+ proteins: have engine, will work. Nat Rev Mol Cell Biol 6:519–529. https://doi.org/10.1038/nrm1684
34. Kubota T, Kurihara E, Watanabe K, Ogata K, Kaneko R, Goto M, Ohshima T, Yoshimune K. 2022. Conformational changes in the catalytic region are responsible for heat-induced activation of hyperthermophilic homoserine dehydrogenase. Commun Biol 5:704. https://doi.org/10.1038/s42003-022-03656-7
35. Moser C, Mol O, Goody RS, Sinning I. 1997. The signal recognition particle receptor of *Escherichia coli* (FtsY) has a nucleotide exchange factor built into the GTPase domain. Proc Natl Acad Sci U S A 94:11339–11344. https://doi.org/10.1073/pnas.94.21.11339
36. Gasper R, Scrima A, Wittinghofer A. 2006. Structural insights into HypB, a GTP-binding protein that regulates metal binding. J Biol Chem 281:27492–27502. https://doi.org/10.1074/jbc.M600809200
37. Ma L, Xie D, Zhao X, Wang L, Hou L, Liu X, Li Z, Cheng H, Zhang J, Gao M, Zeng F. 2022. Npa3-Gpn3 cooperate to assemble RNA polymerase II and prevent clump of its subunits in the cytoplasm. Int J Biol Macromol 206:837–848. https://doi.org/10.1016/j.ijbiomac.2022.03.081
38. Mora-García M, Ascencio D, Félix-Pérez T, Ulloa-Calzonzin J, Juárez-Reyes A, Robledo-Márquez K, Rebolloso-Gómez Y, Riego-Ruiz L, DeLuna A, Calera MR, Sánchez-Olea R. 2022. Synthetic negative genome screen of the GPN-loop GTPase Npa3 in *Saccharomyces cerevisiae*. Curr Genet 68:343–360. https://doi.org/10.1007/s00294-022-01243-1
39. Brock TD, Brock KM, Belly RT, Weiss RL. 1972. Sulfolobus: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. Archiv Mikrobiol 84:54–68. https://doi.org/10.1007/BF00408082
40. Wagner M, van Wolferen M, Wagner A, Lassak K, Meyer BH, Reimann J, Albers S-V. 2012. Versatile genetic tool box for the crenarchaeote *Sulfolobus acidocaldarius*. Front Microbio 3:1–12. https://doi.org/10.3389/fmicb.2012.00214
41. Ye X, van der Does C, Albers S-V. 2020. SaUspA, the universal stress protein of *Sulfolobus acidocaldarius* stimulates the activity of the PP2A phosphatase and is involved in growth at high salinity. Front Microbiol 11:598821. https://doi.org/10.3389/fmicb.2020.598821
42. Lassak K, Peeters E, Wróbel S, Albers SV. 2013. The one-component system ArnR: a membrane-bound activator of the crenarchaeal archaellum. Mol Microbiol 88:125–139. https://doi.org/10.1111/mmi.12173
43. Cox J, Mann M. 2008. MaxQuant enables high peptide identification rates, individualized p. p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 26:1367–1372. https://doi.org/10.1038/nbt.1511
44. Klopfenstein DV, Zhang L, Pedersen BS, Ramírez F, Warwick Vesztrocy A, Naldi A, Mungall CJ, Yunes JM, Botvinnik O, Weigel M, Dampier W, Dessimoz C, Flick P, Tang H. 2018. GOATOOLS: a python library for gene ontology analyses. Sci Rep 8:1–17. https://doi.org/10.1038/s41598-018-28948-z
45. Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R. 2009. Quickgo: a web-based tool for gene ontology searching. Bioinformatics 25:3045–3046. https://doi.org/10.1093/bioinformatics/btp536
46. Zallot R, Oberg N, Gerlt JA. 2019. The EFI web resource for genomic enzymology tools: leveraging protein, genome, and metagenome databases to discover novel enzymes and metabolic pathways. Biochemistry 58:4169–4182. https://doi.org/10.1021/acs.biochem.9b00735
47. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504. https://doi.org/10.1101/gr.1239303
48. Van Veldhoven PP, Mannaerts GP. 1987. Inorganic and organic phosphate measurements in the nanomolar range. Analytical Biochemistry 161:45–48. https://doi.org/10.1016/0003-2697(87)90649-X
49. Joosten RP, Long F, Murshudov GN, Perrakis A. 2014. The PDB_REDO server for macromolecular structure model optimization . IUCrJ 1:213–220. https://doi.org/10.1107/S2052252514009324
50. Chojnowski G, Choudhury K, Heuser P, Sobolev E, Pereira J, Oezugurel U, Lamzin VS. 2020. The use of local structural similarity of distant homologues for crystallographic model building from a molecular-replacement solution. Acta Crystallogr D Struct Biol 76:248–260. https://doi.org/10.1107/S2059798320000455
51. Long F, Vagin AA, Young P, Murshudov GN. 2008. BALBES: a molecular-replacement pipeline. Acta Crystallogr D Biol Crystallogr 64:125–132. https://doi.org/10.1107/S0907444907050172
52. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. 2007. Phaser crystallographic software. J Appl Crystallogr 40:658–674. https://doi.org/10.1107/S0021889807021206
53. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. 2011. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 67:235–242. https://doi.org/10.1107/-S0907444910045749
54. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of coot. Acta Crystallogr D Biol Crystallogr 66:486–501. https://doi.org/10.1107/S0907444910007493
55. Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD. 2012. Towards automated crystallographic structure refinement with phenix.refine . Acta Crystallogr D Biol Crystallogr 68:352–367. https://doi.org/10.1107/S0907444912001308
56. The PyMol molecular graphics system, version 1.2r3pre. 2010. Schrödinger, LLC
57. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF chimera - a visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612. https://doi.org/10.1002/jcc.20084
58. Osorio-Valeriano M, Altegoer F, Steinchen W, Urban S, Liu Y, Bange G, Thanbichler M. 2019. ParB-type DNA segregation proteins are CTP-dependent molecular switches. Cell 179:1512–1524. https://doi.org/10.1016/j.cell.2019.11.015
59. Wales TE, Fadgen KE, Gerhardt GC, Engen JR. 2008. High-speed and high-resolution UPLC separation at zero degrees celsius. Anal Chem 80:6815–6820. https://doi.org/10.1021/ac8008862
60. Geromanos SJ, Vissers JPC, Silva JC, Dorschel CA, Li G-Z, Gorenstein MV, Bateman RH, Langridge JI. 2009. The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. Proteomics 9:1683–1695. https://doi.org/10.1002/pmic.200800562
61. Li G-Z, Vissers JPC, Silva JC, Golick D, Gorenstein MV, Geromanos SJ. 2009. Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. Proteomics 9:1696–1719. https://doi.org/10.1002/pmic.200800564

Supplementary Information for

Archaeal GPN-loop GTPases involve a lock-switch-rock mechanism for GTP hydrolysis

Lukas Korf[†,1], Xing Ye[†,2], Marian S. Vogt[1], Wieland Steinchen[1,3], Mohamed Watad[1], Chris van der Does[2], Maxime Tourte[2], Shamphavi Sivabalasarma[2], Sonja-Verena Albers[2], Lars-Oliver Essen[1]

[1]Philipps University, Department of Chemistry, Hans-Meerwein-Strasse 4, 35032 Marburg, Germany

[2]University of Freiburg, Institute for Biology II, Molecular Biology of Archaea, Schaenzlestrasse 1, 79104 Freiburg, Germany

[3]Center for Synthetic Microbiology (SYNMIKRO), Karl-von-Frisch-Strasse 14, 35043 Marburg, Germany

[†]Contribution of both authors is considered equal

Email: essen@chemie.uni-marburg.de
sonja.albers@biologie.uni-freiburg.de

This PDF file includes:

Figures S1 to S12

Tables S1 to S3

Legends for Movies S1

SI References

Other supplementary materials for this manuscript include the following:

Movies S1

Table S4 (proteomics data)

Table S5 (HDX data)

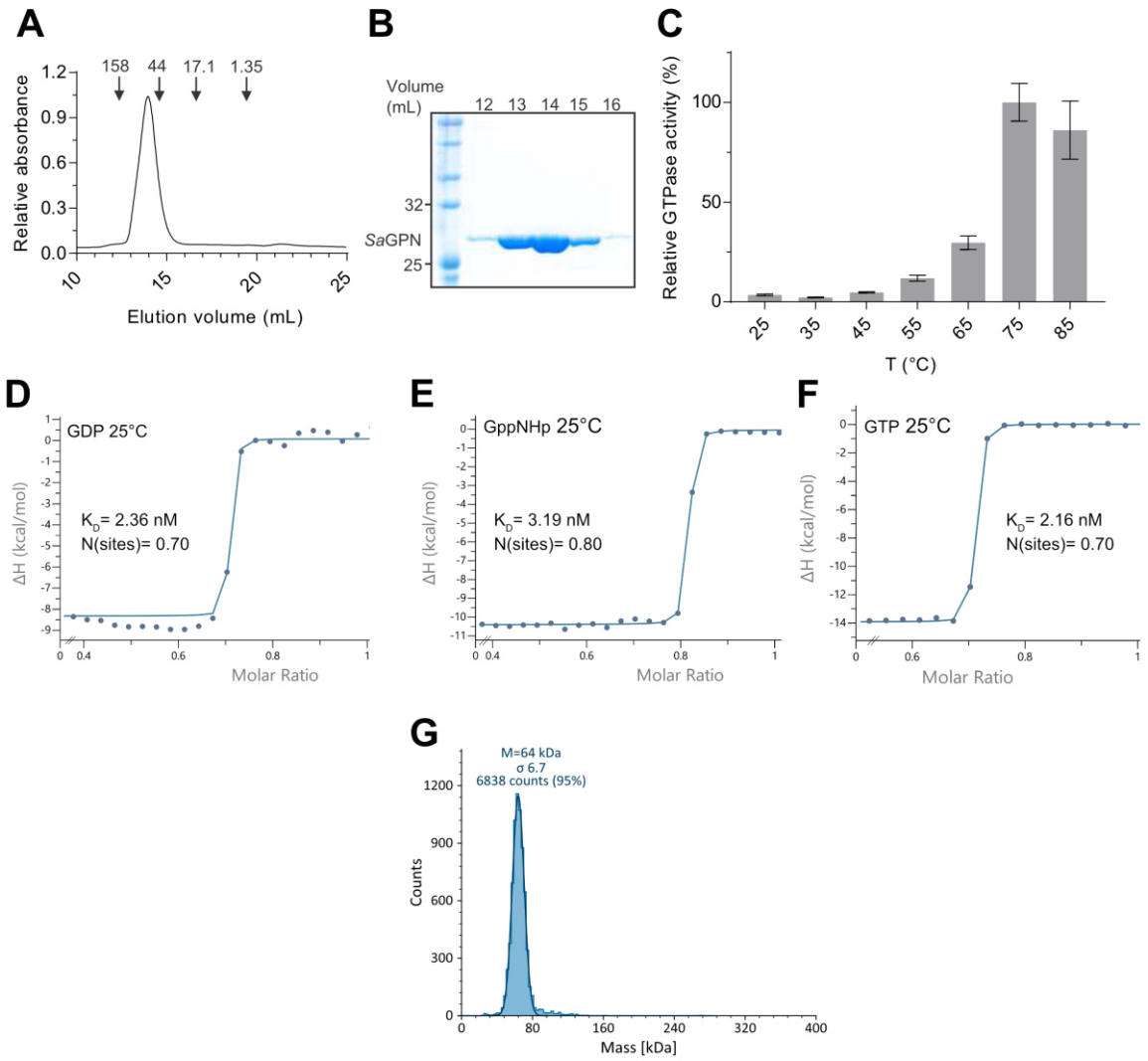

Fig. S1. Biochemical support data to *Sa*GPN. (A) Purified *Sa*GPN was analyzed on a Superdex 200 10/300 GL column. Molecular weight standard of 158, 44, 17, and 1.35 kDa are indicated. **(B)** Fractions from (A) were analyzed by SDS-PAGE. **(C)** Temperature dependency of *Sa*GPN hydrolysis activity. Data represent mean +- s.d. of n=3 replicates **(D)** ITC measurement of *Sa*GPN at 25 °C with titration against GDP, showing a substrate affinity of K_D = 2.4 nM for GDP. **(E)** Same measurement with titration against GppNHp, showing a substrate affinity of K_D = 3.2 nM for GppNHp. **(F)** ITC measurement of *Sa*GPN at 25 °C with titration against GTP, showing a substrate affinity of K_D = 2.2 nM for GTP. This measurement was performed as a single run for check-up only. **(G)** Massphotometer measurement of *Sa*GPN with a final concentration of 50 nmol to check the potential dissociation of *Sa*GPN dimer against high dilution. No dissociation of the dimeric *Sa*GPN could be observed.

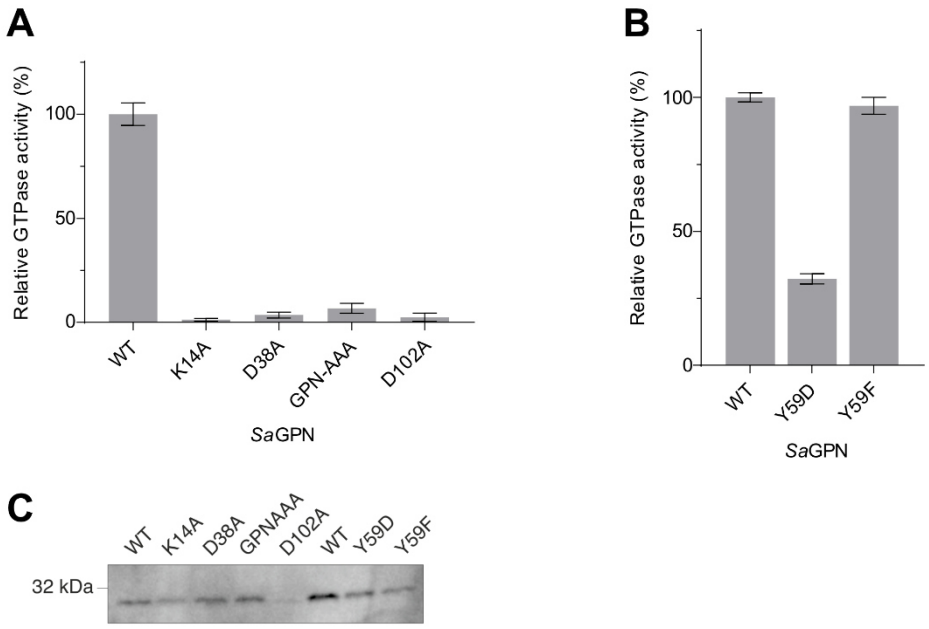

Fig. S2. GTPase activity of *Sa*GPN mutants. (A) GTPase activity of *Sa*GPN K14A, D38A, GPN-AAA and D102A mutants (GTPase activity defect mutants) and **(B)** phosphorylation mutants (*Sa*GPN Y59D and Y59F), determined at 65 °C. Data represent mean +- s.d. of n=3 replicates. **(C)** Confirmation of plasmid-based expression of HA-tagged *Sa*GPN (GTPase activity defect mutants and phosphorylation mutants) in the Δ*saGPN* strain confirmed by western blot analysis. *S. acidocaldarius* cells were grown in nutrient-rich medium without uracil and cell samples were taken at OD_{600} of 0.4, which was analyzed by Western blot with α-HA antibody.

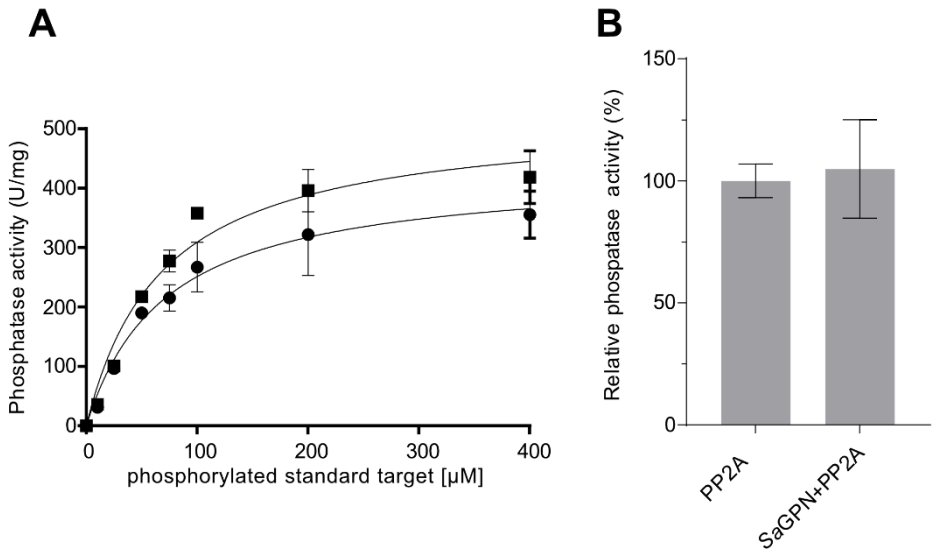

Fig. S3. Phosphatase activity dependency of PP2A. (A) Phosphatase activity of PP2A in absence (squares) or presence (circles) of *Sa*GPN. All experiments were performed in biological triplicates. **(B)** Phosphatase activity of PP2A in absence or presence of *Sa*GPN. All experiments were performed in biological triplicates. Data represent mean +- s.d. of n=3 replicates.

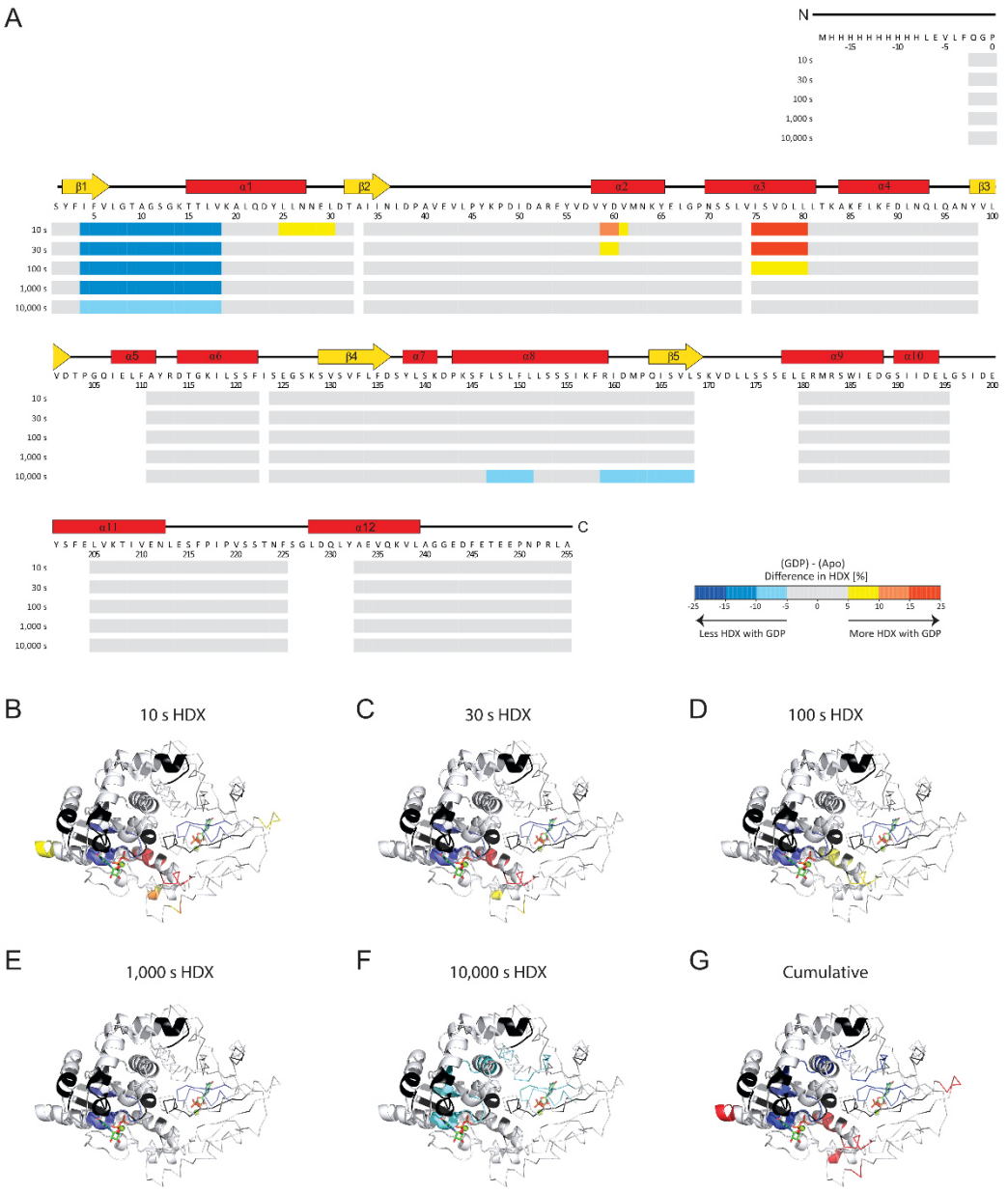

Fig. S4. HDX difference of *Sa*GPN GDP state – Apo state. (A) HDX of time points 10-10,000s mapped on sequence with secondary structure elements. Color code is explained by legend down right. **(B-F)** HDX of respective time points mapped on structure with one protomer in ribbon and one in cartoon view for orientation. **(G)** Cumulated HDX indicating if any HDX occurred during any time point with the highest magnitude in difference displayed on the structure.

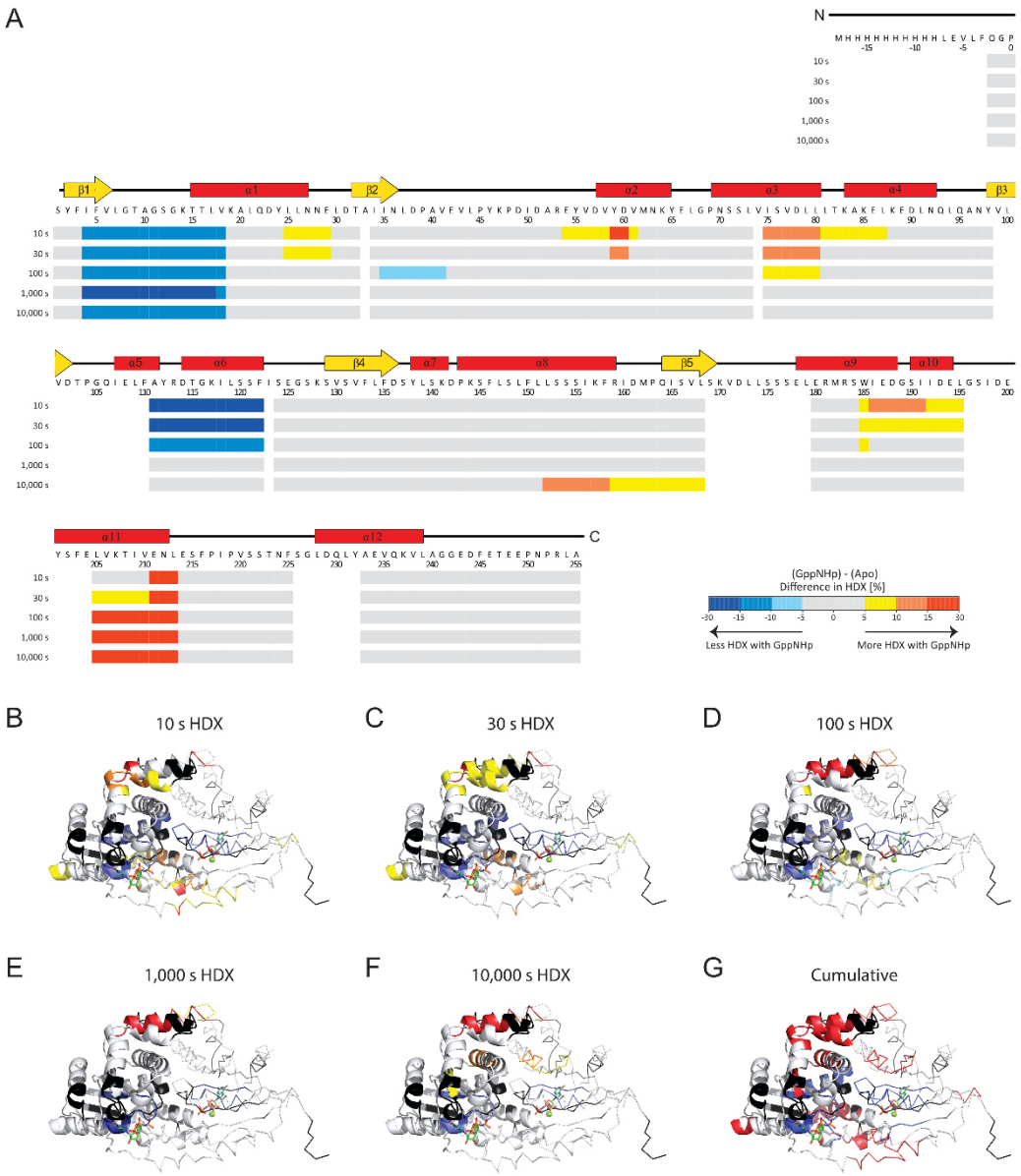

Fig. S5. HDX difference of *Sa*GPN GppNHp state – Apo state. (A) HDX of time points 10-10,000s mapped on sequence with secondary structure elements. Color code is explained by legend down right. **(B-F)** HDX of respective time points mapped on structure with one protomer in ribbon and one in cartoon view for orientation. **(G)** Cumulated HDX indicating if any HDX occurred during any time point with the highest magnitude in difference displayed on the structure.

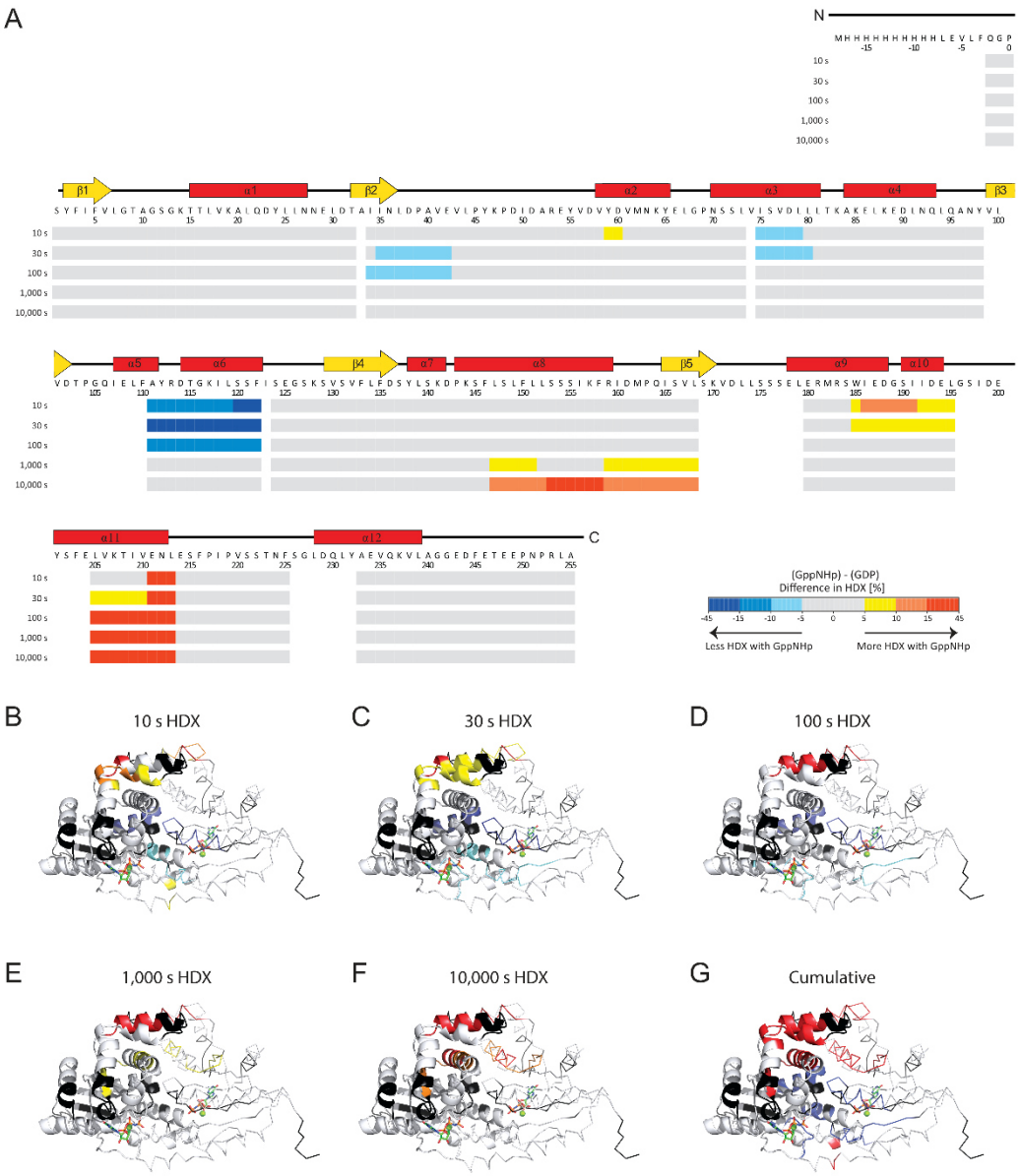

Fig. S6. HDX difference of *Sa*GPN GppNHp state – GDP state. (A) HDX of time points 10-10,000s mapped on sequence with secondary structure elements. Color code is explained by legend down right. **(B-F)** HDX of respective time points mapped on structure with one protomer in ribbon and one in cartoon view for orientation. **(G)** Cumulated HDX indicating if any HDX occurred during any time point with the highest magnitude in difference displayed on the structure.

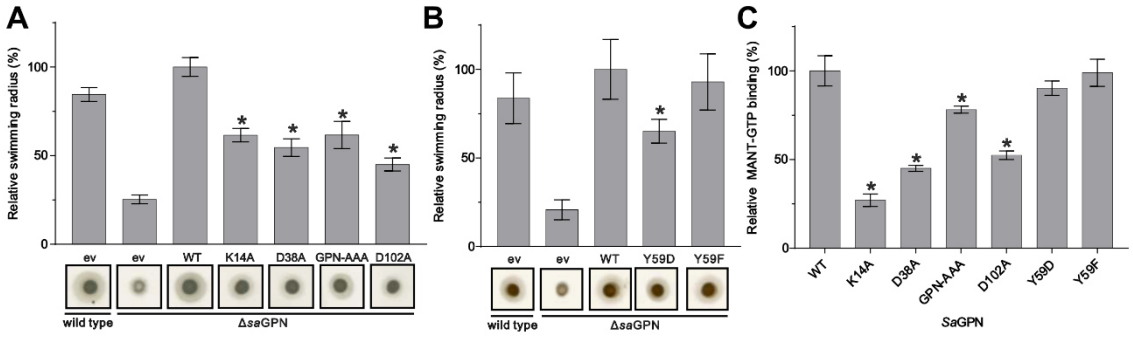

Fig. S7. GTPase activity of *Sa*GPN is dispensable for motility of *S. acidocaldarius*. (A) Δ*saGPN* strains were complemented with GTPase activity defect mutants varied in motifs G1 (K14A), G2 (D38A), GPN (AAA) and G3 (D102A), and their swimming motility was determined. 'ev' represents empty vector samples, i.e., ev Δ*saGPN* is the control sample. **(B)** Analysis of the potential phosphorylation target residue Y59 by complementation of Δ*saGPN* with non-phosphorylatable Y59F and non-homologous Y59D mutants. Swimming assays were carried out as in (A). Y59F could fully restore the swimming motility defect in contrast to Y59D, showing that phosphorylation at Y59 is not relevant for *Sa*GPNs biological role. Significant differences of swimming motility between complementation of Δ*saGPN* mutant with wild type *sa*GPN and with variants (p-value < 0.05) were indicated by an asterisk. **(C)** MANT-GTP binding assays of *Sa*GPN and its mutants show that GTP-binding is impaired by GTPase defect and Y59 mutants. Significant differences of MANT-GTP binding between wild type *Sa*GPN and variants (p-value < 0.05) were indicated by an asterisk. Data represent mean +- s.d. of n=3 replicates.

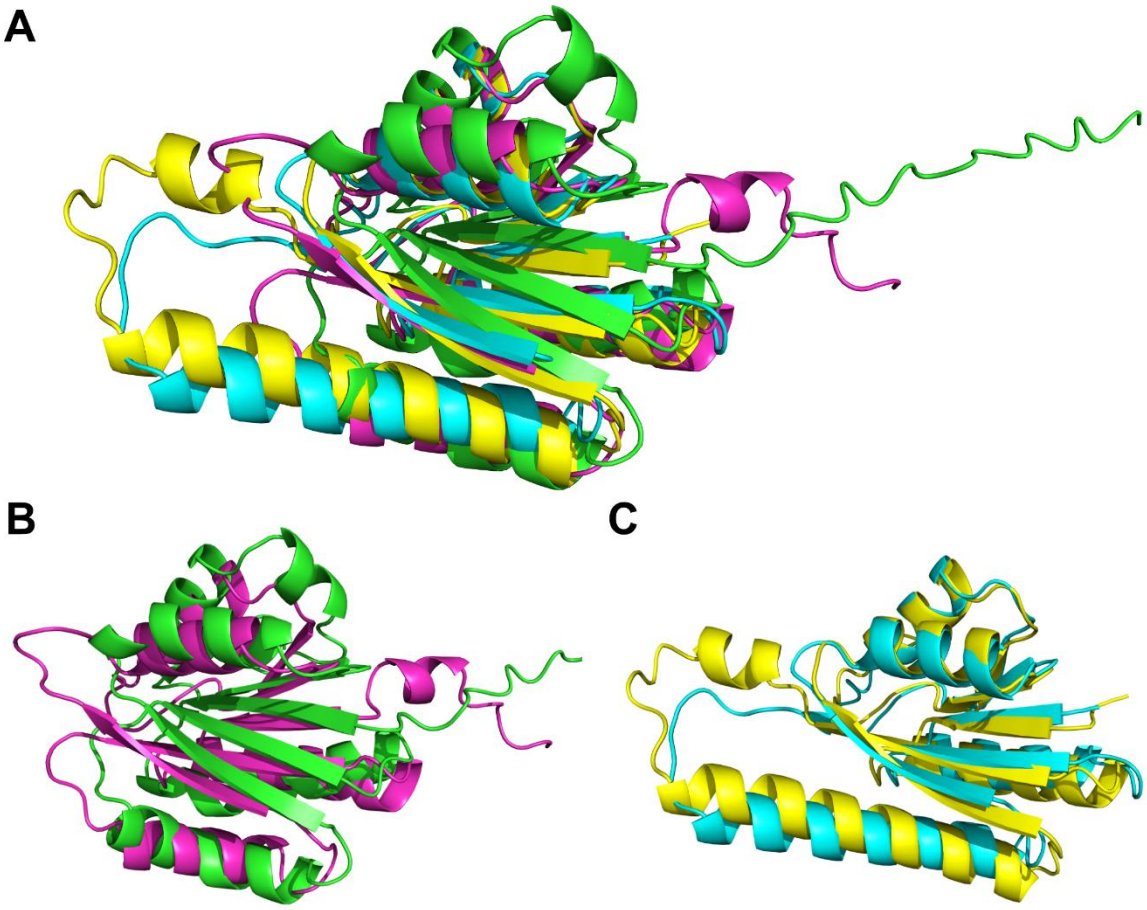

Fig. S8. Superposition of *S. acidocaldarius* Usp structure predictions. (A) Superposition of *Sa*UspA (cyan), *Sa*UspB (magenta), *Sa*UspC (yellow), and *Sa*UspD (green). AlphaFold models (1) show good alignment of core Usp domains, despite low sequence identities between UspA, UspB and UspC (19-29%), and even lower of UspA/UspB/UspC to UspD (8-11%). **(B)** Superposition of *Sa*UspD and *Sa*UspB (seq. id. 11%, r.m.s.d. 3.504 Å for 64 Cα-positions) highlights their structural similarity. **(C)** Superposition of the pairwise related *Sa*UspA and *Sa*UspC (seq. id. 29%, r.m.s.d. 1.845 Å for 100 Cα-positions). AlphaFold models were created using google colab running AlphaFold v2.3.2. with standard settings.

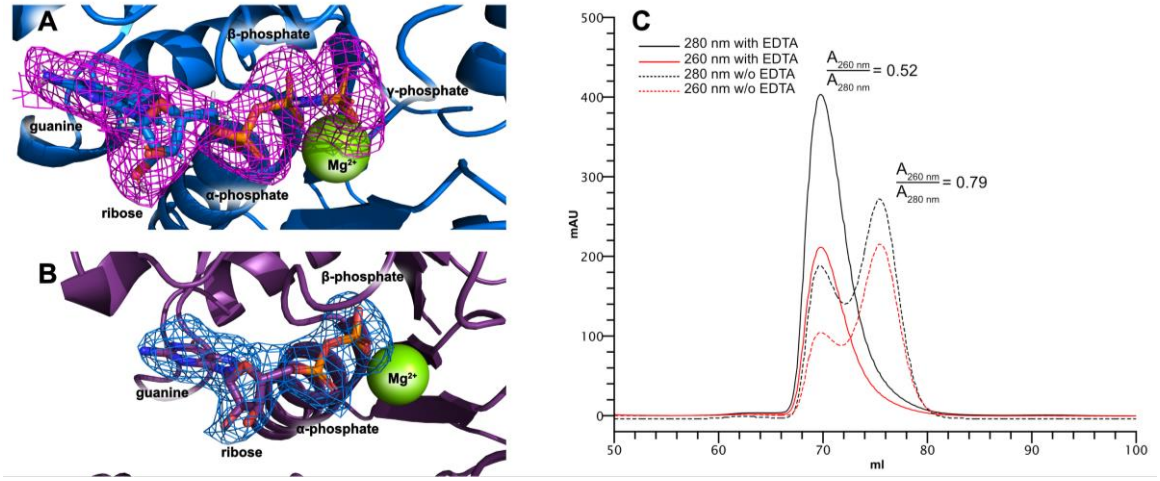

Fig. S9. F_o-F_c omit maps displayed around ligands of *Sa*GPN structures confirming nucleotide states. (A) F_o-F_c map around GppNHp reveals very well defined density including the γ-phosphate position. **(B)** F_o-F_c map around GDP showing no excessive density for a potential γ-phosphate. **(C)** SEC chromatogram showing a representative *Sa*GPN purification with EDTA added during cell lysis and subsequent purification steps (continues lines) and without EDTA (dashed lines). Corresponding peak 260/280 nm ratios are given next to them. The 0.52 ratio peak refers to a nucleotide free extraction of *Sa*GPN while the 0.79 ratio refers to a nucleotide bound state. Black graphs display absorption at 280 nm while red graphs refer to absorption at 260 nm wavelength.

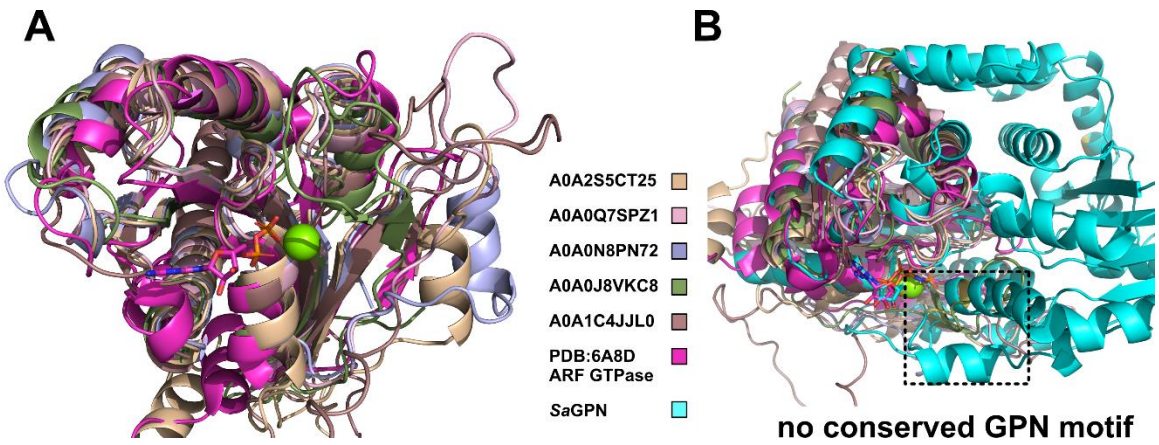

Fig. S10. Superposition of selected bacterial IPR004130 (GPN-loop GTPase) AlphaFold models superposed on *Chlamydomonas reinhardtii* ARF small GTPase. (A) Superposition of *C. reinhardtii* ARF small GTPase (PDB: 6A8D) revealing high similarity of bacterial annotated GPN-loop GTPase with ARF like small GTPases. **(B)** Annotated bacterial GPNs also lack the fundamental GPN-loop as visualized by the alignment with *Sa*GPN (PDB: 7ZHF). *C. reinhardtii* ARF small GTPase is colored in pink, *Sa*GPN in cyan and bacterial GPNs are colored in different pale colors.

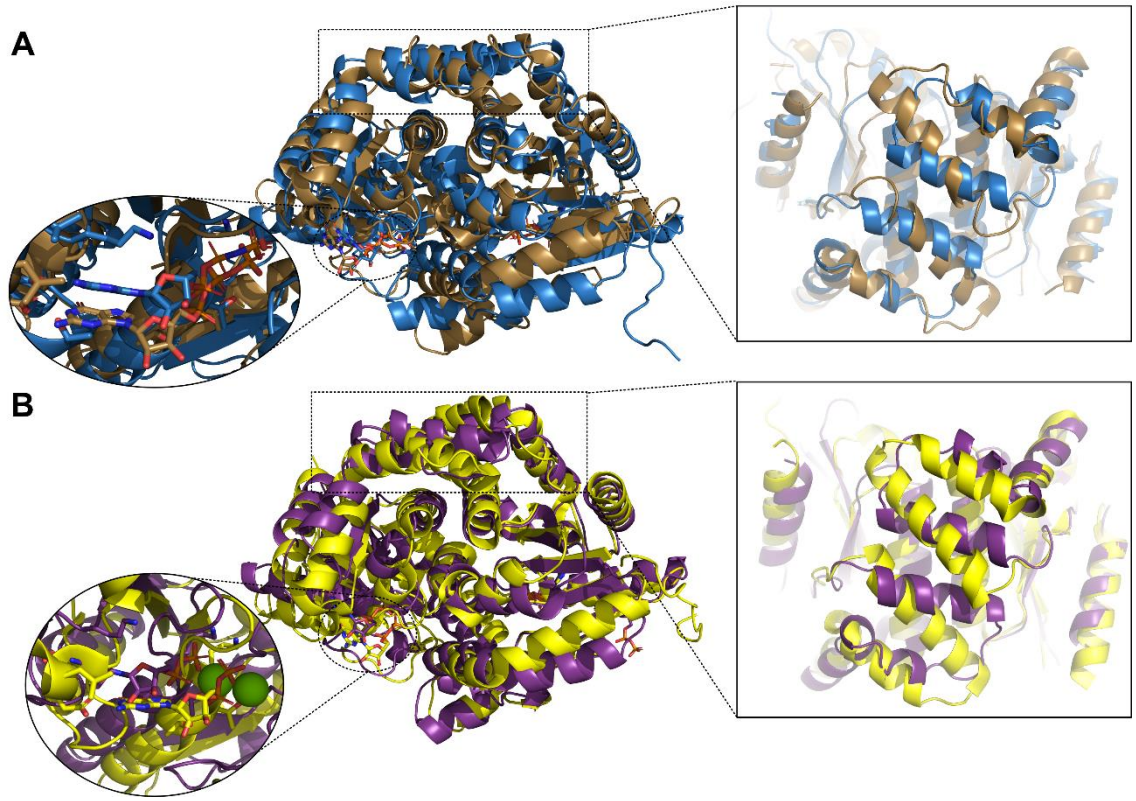

Fig. S11. Structural comparison of *Sa*GPN and *Pa*GPN in both states. (A) Structural comparison of SaGPN (GppNHp, blue) and *Pa*GPN (GTP, brown, PDB: 1YR8 (2)) after superposition on monomer A. The roof helix region is highlighted as inlay in a top down view (right), the nucleotide cavity is shown as inlay (left). **(B)** Structural comparison of *Sa*GPN (GDP, purple) and *Pa*GPN (GDP, yellow, PDB: 1YRB(2)) focusing both the roof helix region in a top down view (right) and the nucleotide cavity (left).

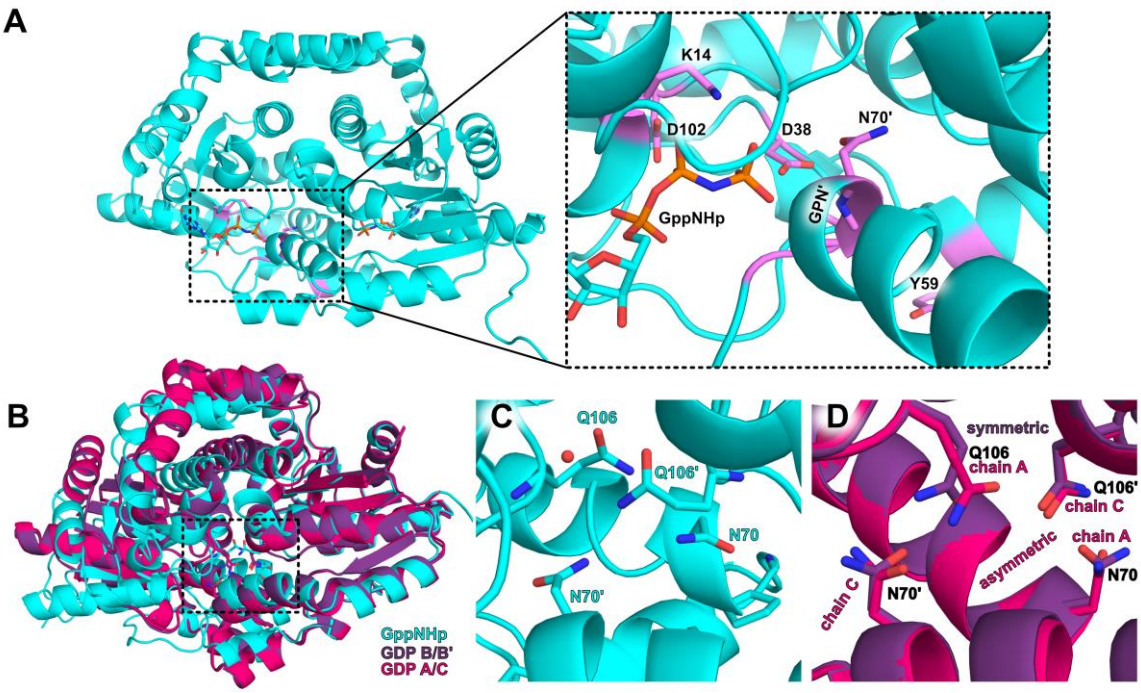

Fig. S12. Structural localization of SaGPN mutants and other structural highlights. (A) Highlighting of all residues that were mutated in this study for reference. **(B)** Structural comparison of different *Sa*GPN dimers (GDP chains A/C pink, GDP chains B/B' purple, GppNHp cyan). **(C)** Focus on the dimer of *Sa*GPN GppNHp state showing the Q106/N70 interface**. (D)** Focus on the Q106/N70 interface of the two different *Sa*GPN·GDP dimers revealing the local asymmetry of the chain A/C dimer.

Table S1. Strains and plasmids in this study.

Strains/plasmids	Genotype	Source/Reference
Escherichia coli Top10	F- *mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 nupG recA1 araD139 Δ(ara-leu)7697 galE15 galK16 rpsL(Str^R) endA1 λ$^-$*	Invitrogen
ER1821	λ$^-$ F$^-$ *glnX44 e14$^-$(McrA$^-$) rfbD1 endA1 thi-1 Δ(yjiT-opgB)114::IS10 (EcoKI R$^-$ M$^-$ McrBC$^-$ Mrr$^-$) + rpoS393(am) creC510 lrhA::IS3 ydeN::IS10*	New England Biolabs
Rosetta (DE3) pLysS	F$^-$ *ompT hsdS*$_B$(r$_B^-$ m$_B^-$) *gal dcm* (DE3) pLysSRARE (Cam^R)	Novagen
Sulfolobus acidocaldarius		
MW001	*Sulfolobus acidocaldarius* DSM639 ∆*pyrE*	(3)
MW351	MW001 Δ*saci1210* (Δ*arnA*)	(4)
MW332	MW001 Δ*saci_1171Δsaci_1180 (ΔarnR ΔarnR1)*	(5)
MW829	MW001 Δ*saci1281* (Δ*saGPN*)	This study
Plasmids		
pSVA407	Gene targeting plasmid, pGEM-T Easy backbone, *pyrEFSSO* and *lacSSSO* cassette; single crossover method	(3)
pSVA5142	In-frame deletion of *saGPN*, cloned into pSVA407 with *Nco*I*, BamH*I	This study
pSVAaraFX-HA	pRN-1 based shuttle vector with *lacSSSO* reporter gene	(6)
pSVA5173	Complementation of Δ*saGPN* with *saGPN* using its native promoter, cloned into pSVAaraFX-HA instead of *lacS* with *Sac*II and *Xho*I	This study
pSVA5174	Complementation of Δ*saGPN* with *saGPN^{K14A}* using its native promoter, cloned into pSVAaraFX-HA instead of *lacS* with *Sac*II and *Xho*I	This study
pSVA5175	Complementation of Δ*saGPN* with *saGPN^{D38A}* using its native promoter, cloned into pSVAaraFX-HA instead of *lacS* with *Sac*II and *Xho*I	This study
pSVA5176	Complementation of Δ*saGPN* with *saGPN$^{GPN-AAA}$* using its native promoter, cloned into pSVAaraFX-HA instead of *lacS* with *Sac*II and *Xho*I	This study
pSVA5177	Complementation of Δ*saGPN* with *saGPN^{D102A}* using its native promoter, cloned into pSVAaraFX-HA instead of *lacS* with *Sac*II and *Xho*I	This study
pSVA5193	Complementation of Δ*saGPN* with *saGPN^{Y59D}* using its native promoter, cloned into pSVAaraFX-HA instead of *lacS* with *Sac*II and *Xho*I	This study
pSVA13433	Complementation of Δ*saGPN* with *saGPN^{Y59F}* using its native promoter, cloned into pSVAaraFX-HA instead of *lacS* with *Sac*II and *Xho*I	This study
pSVA1037	Saci0884 (*pp2a*) with C-terminal His-tag cloned into pETDuet-1 with *NcoI*, *BamHI* in MCSI	(4)
p7XC3H	FX cloning expression plasmid with C-term His tag	(7)
p7XNH3	FX cloning expression plasmid with N-term His tag	(7)
p7XNS3	FX cloning expression plasmid N-term Strep tag	(8)
pSVA5145	*saGPN* cloned into p7XNS3 by FX cloning method	This study
pSVA5146	*saGPN^{K14A}* cloned into p7XNS3 by FX cloning method	This study
pSVA5147	*saGPN^{D38A}* cloned into p7XNS3 by FX cloning method	This study
pSVA5152	*saGPN$^{GPN-AAA}$* cloned into p7XC3H by FX cloning method	This study
pSVA5153	*saGPN^{D102A}* cloned into p7XC3H by FX cloning method	This study
pSVA5195	*saGPN^{Y59D}* cloned into p7XNH3 by FX cloning method	This study
pSVA13434	*saGPN^{Y59F}* cloned into p7XC3H by FX cloning method	This study

Table S2. Primers used in this study.

Primer name	Sequence (5´- 3´)	Purpose
	primers for pSVA5142	
9143	AACTGCTCAAACCTAGGTCAGGATCCCTCGCAATATCCGGTATAG	Δ*saGPN* downstr rev
9144	ACGTCGCATGCTCCCGGCCGCCATGTCCCATATCGCCCATCACAG	Δ*saGPN* upstr fw
9169	GAACTTCGGCAGCCTTAACTAAAGTGGTCTTAC	Δ*saGPN* upstr rev ol
9170	AGTTAAGGCTGCCGAAGTTCAAAAGGTCTT	Δ*saGPN* downstr fw ol
9108	GCTCAAGACCGTACAAACTC	Δ*saGPN* check primer fw
9109	GTCACACCCAAGTTCTTGAG	Δ*saGPN* check primer rev
	primers for pSVA5173	
9149	GTCAGTGAGCGAGGAAGCCCACCGCCCTGCGGTGGATATAATGAC	*saGPN* compl fw
9150	CCGGAACGTCATACGGGTACTCGAGAAGCCTGGGGTTCGGTTCCT	*saGPN* compl rev
	primers for pSVA5145	
9141	ATATATGCTCTTCTAGTTACTTTATATTTGTACTAGGAACTGCA	*saGPN* expr fw
9142	TATATAGCTCTTCATGCAAGCCTGGGGTTCGGTTCCTCTGTTTC	*saGPN* expr rev
	Site-directed mutagenesis	
9153	TCAGGTGCAACCACTTTAGTTAAGGCTTTACAAGATTATTTG	$saGPN^{K14A}$ fw
9154	CCTTAACTAAAGTGGTTGCACCTGAACCTGCAGTTCCTAG	$saGPN^{K14A}$ rev
9155	TTAACCTAGCACCAGCAGTGGAAGTATTGCC	$SaGPN^{D38A\ A}$ fw
9156	CTGCTGGTGCTAGGTTAATTATTGCTGTGTCTAACTCATTG	$saGPN^{D38A}$ rev
9157	GAATTAGCAGCTGCATCCTCACTAGTGATTTCCGTAGATC	$saGPN^{GPN\text{-}AAA}$ fw
9158	TGAGGATGCAGCTGCTAATTCGTACTTATTCATCACATCATATAC	$saGPN^{GPN\text{-}AAA}$ rev
9159	CGTTTTAGTTGCAACTCCTGGTCAAATAGAACTTTTTGC	$saGPN^{D102A}$ fw
9160	CAGGAGTTGCAACTAAAACGTAATTGGCTTGAAGCTG	$saGPN^{D102A}$ rev
11266	GAGAGTACGTTGATGTAGATGATGTGATGAATAAGTACGA	$saGPN^{Y59D}$ fw
11267	ATCTACATCAACGTACTCTCTGGCATCG	$saGPN^{Y59D}$ rev
11978	GAGAGTACGTTGATGTATTTGATGTG	$saGPN^{Y59F}$ fw
11979	AAATACATCAACGTACTCTCTGGC	$saGPN^{Y59F}$ rev
	primers qRT-PCR	
1480	CCTGCAACATCTATCCATAACATACCGA	*secY*-qRT-PCR-fw
1481	CCTCATAGTGTATATGCTTTAGTAGTAG	*secY* -qRT-PCR-rev
1424	ACTGCGTCTACTGCGTTATCTTTATC	*flaB*-qRT-PCR-fw
1425	GGAGATAAGTCTACACTAGATACACCAGAA	*flaB*-qRT-PCR-rev

Table S3. Crystallographic table for *Sa*GPN structures.

	*Sa*GPN•GppNHp (**7ZHF**)	*Sa*GPN•GDP (**7ZHK**)†
Data collection and processing		
X-ray source, beamline	SLS, Beamline X06SA	SLS, Beamline X06SA
Detector	DECTRIS EIGER X 16M	DECTRIS EIGER X 16M
Wavelength (Å)	1.000009	1.000031
Space group	I 1 2 1	C 2 2 2_1
Cell dimensions (*a*, *b*, *c*, *α*, *β*, *γ*)	46.30Å, 84.95Å, 75.32Å 90.00°, 95.62°, 90.00°	117.69Å, 141.88Å, 100.24Å 90.00°, 90.00°, 90.00°
Resolution (Å)	41.09 – 1.80	45.29 – 2.40
Total reflections	85,536	133,417
Multiplicity	3.5	4.3
Unique reflections	24,163	30,856
Completeness (%)#	89.6 (64.1)	76.8 (19.7)
R_{merge}	0.048 (0.098)	0.066 (1.22)
CC1/2	0.999 (0.999)	0.998 (0.325)
I/σ(I)	9.1 (2.2)	12.9 (1.5)
Wilson *B*-factor (Å^2)	41.2	52.8
Refinement		
Resolution (Å)	29.5 – 1.80 (1.9-1.8)	43.9 – 2.40 (2.5-2.4)
R_{work}, R_{free}	0.172, 0.204 (0.289)	0.188 (0.256), 0.247 (0.332)
Reflections (working, test set)	24,806 (2757), 796 (96)	25,464 (716), 1283 (36)
r.m.s.d. from ideal:		
Bond lengths (Å)	0.006	0.002
Bond angles (°)	0.82	0.55
Total number of atoms	4,320	6,114
Water	196	351
Chlorid Ions	1	0
Average *B*-factor (Å^2)	70.3	70.8
Ligands	55.6	51.4
Ramachandran favored (%)	98.8	99.6
Ramachandran allowed (%)	1.2	0.4
Ramachandran outliers (%)	0.0	0.0

Statistics for the highest-resolution shell are shown in parentheses.
† Dataset was corrected by STARANISO (9) for anisotropic diffraction. Cut-offs used direction 0.808 a* + 0.255 b* + 0.531 c* for best diffraction (2.40 Å), 0.023 a* + 0.993 b* + 0.115 c* for worst diffraction (3.25 Å). # Spherical completeness.

Table S4 (separate file). timsTOF data and GO term analysis of *S. acidocaldarius* WT and *saGPN* knockout strains.

Table S5 (separate file). *Sa*GPN HDX data and evaluation.

Movie S1 (separate file). The movie shows a morph of the *Sa*GPNs•GppNHp state towards its GDP-bound state, visualizing the major allosteric changes upon nucleotide hydrolysis.

SI References

1. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

2. Gras, S. *et al.* Structural insights into a new homodimeric self-activated GTPase family. *EMBO Rep.* **8**, 569–575 (2007).

3. Wagner, M. *et al.* Versatile genetic tool box for the crenarchaeote Sulfolobus acidocaldarius. *Front. Microbiol.* **3**, 1–12 (2012).

4. Reimann, J. *et al.* Regulation of archaella expression by the FHA and von Willebrand domain-containing proteins ArnA and ArnB in Sulfolobus acidocaldarius. *Mol. Microbiol.* **86**, 24–36 (2012).

5. Lassak, K., Peeters, E., Wróbel, S. & Albers, S. V. The one-component system ArnR: A membrane-bound activator of the crenarchaeal archaellum. *Mol. Microbiol.* **88**, 125-139 (2013).

6. Van Der Kolk, N. *et al.* Identification of XylR, the activator of arabinose/xylose inducible regulon in Sulfolobus acidocaldarius and its application for homologous protein expression. *Front. Microbiol.* **11**, 1066 (2020).

7. Geertsma, E. R. & Dutzler, R. A versatile and efficient high-throughput cloning tool for structural biology. *Biochemistry* **50**, 3272–3278 (2011).

8. Ye, X. *et al.* The phosphatase PP2A interacts with ArnA and ArnB to regulate the oligomeric state and the stability of the ArnA/B complex. *Front. Microbiol.* **4**, 1849 (2020).

9. Vonrhein, C. *et al.* Advances in automated data analysis and processing within autoPROC, combined with improved characterisation, mitigation and visualisation of the anisotropy of diffraction limits using STARANISO. *Acta Crystallogr.* **74**, a360–a360 (2018).

3.2 Sequential conformational transition of ArnB, an archaeal ortholog with Sec23/Sec24 core motif

This paper was part of the collaboration of the Volkswagen Foundation Life? Program involving the Albers group from Freiburg University and the Essen group from Marburg University. The paper provides a comprehensive summary of the data concerning the ArnAB complex accumulated over the years since the last publication on this topic by Ye et al. in 2020. Please note that the manuscripts layout was slightly altered to better fit into this thesis.

Citation

L. Korf, W. Steinchen, M. Watad, F. Bezold, M. S. Vogt, L. Selbach, A. Penner, M. Tourte, S. Hepp, S. V. Albers, L.-O. Essen, Sequential conformational transition of ArnB, an archaeal ortholog with Sec23/Sec24 core motif, 2024 (under review in Nat. commun.)

Contribution

The ArnAB complex structure discussed in this paper originates from a crystal grown by A. Penner under supervision of F. Bezold. This protein structure, solved by L.-O. Essen, was analyzed by M. S. Vogt and L. Korf. Follow-up experiments of phosphorylation studies, proteomics sample preparation, bioinformatic research and all necessary protein preparations were conducted by L. Korf. Data analysis, with the exception of proteomics and HDX data, and interpretation of the data was done by L. Korf in collaboration with L.-O. Essen. In addition, initial writing of the manuscript was performed by L. Korf in collaboration with L.-O. Essen and W. Steinchen. Administrative tasks surrounding the publication and organization of the data and files was done by L. Korf.

Summary

The regulatory network governing the archaellum, the motility structure of Archaea, involves an intricate interplay of proteins in various *Sulfolobales* species and other Archaea. In *Sulfolobus acidocaldarius*, two key regulators, ArnA and ArnB, form a heterocomplex. The interaction is mediated by the ZnF-domain of ArnA, predominantly engaging with the β-barrel domain of ArnB, and by the phospho-Threonine (pThr) interaction of the ArnA FHA-domain with the C-terminal helix bundle domain of ArnB.

Notably, the sequential phosphorylation of ArnB by the kinase ArnC induces partial unfolding and/or partial disordered structures due to excess negative charge. Despite the challenges posed by the structural disorder, the FHA-ArnB interaction was traced to the tip region of the two most C-terminal helices of ArnB. Multiple assembly possibilities of the phosphorylated ArnAB complex, primarily based on the disordered linker region of ArnA, include a normal 1:1 heterodimeric complex, a 1:1:X heterotrimer with an additional partner, or polymerization in general leading to multiple higher mass complexes. Beyond the regulation of the Archaellum formation during nutrient limitation, proteomic data indicates that ArnA and ArnB play roles in metabolic pathways of amino acids with additional nitrogen atoms. Knockouts of ArnA or ArnB also impact central metabolic pathways, including the citrate cycle and carbohydrate metabolism. Intriguingly, ArnB exhibits structural homology to Sec23/24, the primary membrane curvature protein of the COPII vesicle system, suggesting a potential role in an early endomembrane system in Archaea. However, while tempting, the evolutionary analysis reveals diverse structural orthologs of ArnB, especially in plants fungi, Asgard Archaea, Eukaryotes, and also Bacteria, raising questions about the precise evolutionary connections and the timing of the potential trafficking function.
In summary, ArnA and ArnB are mainly associated with the archaellum regulatory network by forming a heterocomplex featuring an intricate assembly. In addition, association with exosomes of *S. acidocaldarius* and structurally similarities to the Sec23/24 complex, suggests a potential regulatory function of archaeal exosomes, even if their direct involvement remains uncertain.

Main Manuscript for

Sequential conformational transition of ArnB, an archaeal ortholog with Sec23/Sec24 core motif

Lukas Korf[1], Wieland Steinchen[1,2], Mohamed Watad[1], Filipp Bezold[1], Marian S. Vogt[1], Leon Selbach[1], Alexander Penner[1], Maxime Tourte[3], Sebastian Hepp[1], Sonja-Verena Albers[3,4], Lars-Oliver Essen[1,*]

[1]Department of Chemistry, Philipps University, Hans Meerwein-Str. 4, 35043 Marburg, Germany
[2]Center for Synthetic Microbiology, Philipps University Marburg, Karl-von-Frisch-Str. 14, 35043 Marburg, Germany
[3]University of Freiburg, Institute for Biology, Molecular Biology of Archaea, Schaenzlestrasse 1, 79104 Freiburg, Germany
[4]Signalling Research Centres BIOSS and CIBSS, University of Freiburg, 79104 Freiburg, Germany
*corresponding author

Email: essen@chemie.uni-marburg.de
sonja.albers@biologie.uni-freiburg.de

Author Contributions: L.K.: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Validation, Visualization, Supervision of L.S., Writing – original draft, and Writing – review and editing; **W.S.:** Investigation of HDX experiments, Formal analysis of HDX data, Writing – original draft, and Writing – review and editing; **M.W.:** Formal analysis of MS data and Visualization; **F.B.:** Investigation, and Supervision of A.P.; **M.S.V.:** Investigation, and Project administration; **L.S.:** Investigation; **A.P.:** Investigation; **S.H.:** Investigation of SAXS experiments, and Formal analysis of SAXS data; **S.V.A.:** Funding acquisition, Project administration, Supervision, Writing – original draft, and Writing – review and editing; **L.O.E.:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Writing – original draft, and Writing – review and editing.

Competing Interest Statement: The authors declare no competing interests.

Classification: Biological Sciences / Biochemistry.

Keywords: ArnB, Sec23/24, sequential phosphorylation, order to disorder, bidirectional transition

This Manuscript file includes:

- Abstract
- Main Text
- Methods
- Data availability statement
- Acknowledgements
- Figures 1-6 and figure legends

Abstract

ArnA and ArnB serve as regulators within the archaellum regulatory network by affecting the levels of archaellum components ArlB and ArlX in response to nutrient cues. Together, they form either a loose or a tight complex, whose transition is directed by phosphorylation via the kinase ArnC. For a structure-based analysis of this transition we solved a cocrystal structure of the ArnA/ArnB complex revealing that the zinc finger domain of ArnA interacts with the β-sandwich and C-terminal domain of ArnB. HDX data corroborate the phosphorylation-dependent transition from loose to tight ArnAB complexes. This transition depends on a structural transformation of ArnB by sequential phosphorylation, exposing the interaction surface of the C-terminal domain of ArnB for the forkhead-associated domain of ArnA. Furthermore, we found a striking structural similarity between ArnB and the membrane-curving proteins of the COPII vesicle system, Sec23/Sec24. The common Sec23/Sec24 core motif can be found in all domains of life, where it can apparently adopt a multitude of different functions. Overall, this implies that Sec23/Sec24 orthologs with a function in vesicle formation arose in Lokiarchaeota from related, but not necessarily functionally linked relatives as found in TACK Archaea.

Main Text

Introduction

The archaellum serves as the primary motility apparatus in Archaea, facilitating self-locomotion and enabling adaptive habitat transitions [1]. Its responsiveness to diverse environmental stimuli is governed by a regulatory network known as the archaellum regulatory network (Arn) [2]. Within this intricate system, which regulates transcriptional and translational adaptations, the proteins ArnA and ArnB play pivotal roles in the Crenarchaeon *Sulfolobus acidocaldarius*, particularly in the regulation of *arlB* (formerly *flaB*), the gene encoding the archaellin [3–5]. In the context of nutrient availability in *S. acidocaldarius*, ArnA and ArnB orchestrate the regulation of *arlB* [6]. Notably, the von Willebrand factor domain-containing ArnB forms a complex with the forkhead-associated domain (FHA) protein ArnA [5]. Deprivation of either ArnA or ArnB results in an upregulation of *arlB* expression, leading to hypermotility in the organism [6]. Furthermore, the interplay between nutrient levels and the formation of the ArnAB complex affecting the formation of archaella has been elucidated [5]. Starvation conditions induce a loss of interaction within the complex, subsequently promoting the expression of *arlB*. Consequently, ArnA and ArnB are identified as negative regulators of the archaellum when complex formation is established. Phosphorylation-dependent interactions between ArnA and ArnB are facilitated by the phosphatase PP2A and the kinase ArnC, maintaining a delicate balance between complex formation (phosphorylation) and dissociation (dephosphorylation) [7]. Despite in-depth *in vitro* and *in vivo* studies, the exact interaction mode of ArnA and ArnB, purportedly located in the C-terminal region of ArnB, remains elusive. Moreover, the mechanism by which the ArnAB complex controls the ArlB levels remains enigmatic. The formation of ArnAB oligomers that depends on phosphorylation is another feature [7]. Intriguingly, we found that the structural characteristics of ArnB with its intriguingly associated β-sandwich, von Willebrand factor type A (vWFA) and α-helical domains form a structural blueprint widely distributed among all domains of life, which also includes the eukaryotic COPII-vesicle assembly factors Sec23/Sec24. However, functional divergence as observed in archaeal proteins like the histones [8], which are functionally distinct in eukaryotes despite sharing a common core fold, illustrates the limitation of relying solely on structure-based function inference. Given this caveat we performed a biochemical and

structural analysis of ArnAB interaction that revealed a novel mode of sequential conformational change of its Sec23/Sec24 core motif by phosphorylation.

Results

Overall structure of the ArnA-ArnB complex and structural comparison with *Sa*vWA2

We were able to solve the crystal structure of the vWA2 paralog [5] and archaellum regulating factor ArnB (UniProt: Q4J9H3, saci_1211) in its complex with ArnA (UniProt: Q4J9H4, saci_1210). The ArnAB cocrystals comprise two complexes per asymmetric symmetry unit, whose structures were solved by molecular replacement and refined at 2.5 Å resolution. The ArnAB complexes are defined by electron density for residues T2-S380 of ArnB and the zinc-finger (ZnF) domain of ArnA (P16-K42, Figure 1A, PDB: 8S05). The latter implies a loss of the FHA domain and the linker region of ArnA, possibly by unspecific proteolysis as observed before when solving the structure of the ArnA FHA domain [5]. The overall architecture of ArnB corresponds mostly to its paralog, the van Willebrandt Factor A (vWFA)-containing protein *Sa*vWA2, including the vWFA domain, an eight-stranded β-sandwich whose topology is split by the vWFA domain, as well as the C-terminal, four helix bundle motif, that has been classified as ArnB_C domain (InterPro entry IPR040929) (Figure 1A) [2,5]. In contrast to *Sa*vWA2, the ArnB_C domain of ArnB harbors an elongated helix pair at the terminal region of the motif, revealing one of the most substantial structural differences between ArnB and *Sa*vWA2 (Figure 1D). This extension presents additional threonine residues that allow for potential interaction of the forkhead-associated (FHA) domain of ArnA, known for its ability to bind phospho-threonines, with ArnB (Figure 1D right panel) [9]. Moreover, ArnB features an additional helix (P95-Q103) in the vWFA domain between β5 and α2 of *Sa*vWA2 increasing the total helix count to ten (Figure 1D). The vWFA domain of ArnB also harbors a Na^+ ion coordinated by D46, S50, T110, T135 and D136 in the metal ion-dependent adhesion site (MIDAS, Figure 1C).

While ArnA interacts strongly with its FHA domain to phosphorylated ArnB (see below), the zinc finger domain (ZnF domain) itself is sufficient to promote an interaction without a post-translational modification of ArnB. This ZnF domain belongs to the RanBP2-type (IPR001876) and is characterized as a ZnF ribbon domains by two consecutive, distorted β-hairpin motifs, which together a zinc ion via C21, C24, C35 and C38. The interface of the ZnF domain of ArnA with ArnB has a rather moderate size of 504/574 $Å^2$ for ArnAB chains A/C and B/D, respectively. The interactions are mostly of

hydrophobic nature and include the C-terminal β-hairpin motif (D31-Q41) of the ZnF domain as well as the N-terminal β1-β2 loop (H12-K21) of the β-sandwich domain, it's α-helical linker to the ArnB_C domain (V286-I293) and adjacent residues of the ArnB_C domain facing the ZnF domain. Accordingly, the ArnA-ArnB interaction based on the ZnF domain appears to be rather weak as indicated by pulldown assays (Figure S1A), but traceable by mass photometry [7]. A further point for the uniqueness of the ZnF-mediated ArnA-ArnB interactions are AF2-multimodels, which were unbiased of the ArnAB structure and show almost an identical interaction as in the ArnA-ArnB cocrystal structure for 4 of the 5 predicted ArnA-ArnB models (Figure S2) with displacement r.m.s.d. values of 1.57-1.65 Å for M1-Q25 of the ZnF domain. This indicates that the intrinsic sequence covariation for ArnA and ArnB domains is already significant enough to provide a robust indicator for the relatively small ArnB/ZnF domain interface. Moreover, an ArnA-ArnB interaction is also displayed in solution by SAXS data (Figure 1E) as the *ab initio* envelope as derived from the SAXS data is fittable to the ArnA-FHA and ArnAB crystal structures. Here, the pair distance distribution function P(r) suggests an overall elongated shape (Figure S3) and thereby supports an ArnA-ArnB interaction based on the ZnF domain, as in the crystal structure, with a flexible region followed by an unbound FHA domain due to a lack of pThr anchor points.

Promiscuity of ArnB phosphorylation-dependent interaction sites

In the search for the entire interaction site of the strong ArnA-ArnB interaction based on phospho-threonine interaction, we performed a comprehensive mass spectrometry-based analysis. Notably, there are many potential interaction sites found in ArnB, especially in the C-terminal HTH motif (Figure 2A) where interaction appears to be most likely, based on structural analysis. Accordingly, we first opted to investigate the phosphorylation pattern by the kinase ArnC (UniProt: Q4J9J0, Saci_1193) with the previously reported phosphorylation conditions [7]. Interestingly, we found that many threonines in the ArnB_C domain were phosphorylated during our *in vitro* phosphorylation, raising some interesting questions about the native phosphorylation conditions. However, as 60 minutes of incubation time at 55 °C yielded many different potential interaction sites we opted to find the ones that are phosphorylated first. Hence, we investigated the effect of different incubation times on the phosphorylation pattern of ArnB. In time points of 5 min, 15 min, 30 min and 60 min (resembling our

reference sample) the phosphorylation pattern of ArnB was analyzed in a tryptic digest MS experiment. Interestingly, ArnA could still be efficiently pulled down along with ArnB during purification with phosphorylation time of little as 5 min. Moreover, phosphorylation of threonines located closer to the N-terminal side of the ArnB_C domain takes place only after T353, T354, T359, T363 and T375 are phosphorylated and only until T322, which aligns with the known phenomenon that multiple phosphorylation usually occur as clusters in a protein [10]. Accordingly, these main threonines are phosphorylated after 5 minutes already, while other phosphorylated threonines that have potential counterparts in vWA2 are found phosphorylated only after 15 minutes or more incubation time. This may indicate ordered processivity of hyperphosphorylation, like in the cyclin–Cdk1–Cks1 system [11] and goes along with the threonines being located in the extension of the C-terminal HTH motif of ArnB, besides T375, which are therefore not present in vWA2. Additionally, it is reported that vWA2 does not interact with ArnA [6], supporting the assumption that the C-terminal HTH extension is the main interaction side for the ArnA FHA domain.

As this narrowed down the possible interaction sites, we attempted to identify the exact position of the ArnB-FHA interaction by a comprehensive alanine mutagenesis study. Inducing the mutation of either T353A+T354A or T359A+T363A it was not possible for the FHA domain to interact with ArnB anymore, if phosphorylated for 5 min or less. However, increasing the phosphorylation time to 60 min, ArnA could be pulled down with ArnB without issues again. Additionally, even when threonines that were found to be phosphorylated only after 5 min were mutated as well, ArnA could still be pulled down along with ArnB, represented by the ArnB-T343A-T344A-T353A-T354A-T359A-T363A-T371A-T375A mutant (Figure S1B). In general, this gives an idea of the more likely interaction side of the FHA domains location at the C-terminal elongation of the ArnB_C domain, but leaves it open if multiple FHA interaction sites play a role in the interaction of ArnA and ArnB. However, the hyper-phosphorylation we observed leads to multiple structural rearrangements, as discussed in the following, and hence are not necessarily interaction sites but structural factors important for a correct ArnB-FHA interaction.

HDX-MS reveals phosphorylation-dependent structural relieves of ArnB

In order to get insights into the interaction site(s) between ArnA and phosphorylated ArnB in solution, we made use of the property of amide protons to exchange for protons

from the aqueous solvent, the extent of which was traced upon incubation of the proteins in deuterated buffer. This hydrogen/deuterium exchange (HDX) was then, after digestion of the proteins into peptides, quantified by mass spectrometry (MS).

We subjected ArnA, ArnB, phosphorylated ArnB, and the ArnA/phosphorylated ArnB complex to HDX-MS experiments, allowing us to resolve, *i)* the conformational changes that ArnB undergoes upon phosphorylation, and *ii)* the regions of ArnA and phosphorylated ArnB establishing the interaction interface in their complex.

For ArnA and ArnB, we could identify 80 and 145 peptides, respectively, that covered more than 90% of their corresponding amino acid sequences (Figures. S1A and S2A, Supplementary Dataset 1). The HDX profile of ArnA corroborated its predicted domain topology, in particular the disordered nature of the linker (S28-N99, where maximal HDX was reached after 10 s of deuteration) joining the N-terminal ZnF domain to the C-terminal FHA domain (Figure S4B). Upon complex formation of ArnA with phosphorylated ArnB, HDX reduction became apparent in both the ZnF domain and FHA domains thus marking the major sites of interaction for ArnA (S1C-D). Specifically, the phosphate recognition module of the FHA domain constituted by R132 and R147 exhibited the strongest HDX reduction consistent with their role in FHA-mediated phosphate recognition for strengthening the ArnA/ArnB interaction [5].

The HDX profile of ArnB itself primarily showed regions of higher-order structure (low HDX at 10 s of deuteration and progression in HDX over the time-course) and only small disordered areas coinciding with short linkers of the crystal structure (Figures S2B, Figure 1A). However, phosphorylation of ArnB induces widespread conformational changes, according to HDX increases of the ArnB domain, in parts of the β-sandwich and vWFA domains vicinity (Figure 3B). These changes may reflect either a partial unfolding event or a disentanglement of the three associated domains. Furthermore, binding of ArnA to phosphorylated ArnB, in turn, reduced the observed HDX rates in proximity of the ArnA-ZnF domain binding site (Figure 3A). Notably, ArnB residues D295-A310, constituting the N-terminal portion of helix α7, incorporate more deuterium upon phosphorylation, whereas a reduction was apparent upon ArnA binding (Figure S6). It may be hence speculated that the phosphorylation-induced conformational change could be a prerequisite for tight ArnA binding. Likely due to the phosphorylation of the threonine side chains, no peptides covering the C-terminus' residues T353 onwards could be retrieved thus precluding further conclusions on this presumed ArnA-FHA domain interaction site by HDX-MS of ArnB (Figure S5A).

Overall, HDX-MS corroborates the binding site of the ArnA-ZnF domain on ArnB observed in the complex structure (Figure 1) and provides evidence for a secondary interaction site established with the ArnA-FHA phosphate recognition module.

tims-ToF proteomics data for ArnA and ArnB deletion strains

Although the roles of ArnA and ArnB in the archaellum regulatory network are established [5,6], the mechanism by which regulation occurs under nutrient limitation remains undisclosed. Consequently, we analyzed the *Sulfolobus acidocaldarius* proteome of the *ΔarnA* and *ΔarnB* strains in two nutritional states in comparison with the respective wild type (WT). For this analysis biological triplicates were grown of each strain, and samples were taken under nutrient rich and starved conditions. Samples were measured in a timsTOF (trapped ion mobility spectrometer) mass spectrometer and quantified via label free quantification before averaging the technical duplicates of each sample and further analysis. These measurements led to the identification of 1,699-1,710 proteins per sample leading to the identification of 1,723 overall proteins of the 2,222 gene products known for the respective *Sulfolobus acidocaldarius* strain (Supplementary Dataset 2). The comparison of the knockout strains with their respective WT sample reveals 1,713-1,716 identified proteins per comparison with overlapping count of 1,694-1,704 (98.7-99.4%). As the overall proteome is basically unaffected by the deletions, we investigated the effects of the deletion strains under higher stringency i.e. a two-tailed *t*-test with a p-value cutoff of <0.05. This analysis revealed 325(rich)/384(starved) statistically changed proteins compared to the WT proteome for the ArnA knockout and 398(rich)/510(starved) for the ArnB knockout, respectively. In addition, we also included the stringency factor of a fold change of at least 50% to not only check for statistical relevance but also for biological effects (Figure 4A/B). As displayed in the volcano plot analysis (Figure 4A/B, S8-9), this revealed 115(rich) / 243(starved) proteins for the *ΔarnA* and 185(rich) / 279(starved) proteins for the *ΔarnB* strain passing that high stringency test. This shows that under starvation conditions the effect on the proteome level is more prominent under nutrient starved conditions for both knockouts. Moreover, the effect of the ArnB knockout is also slightly more impactful on overall altered proteins both after t-testing and after employing the additional fold change cutoff of 50%. To see if these statistically relevant changes on the proteome level also reflect on the biology of *S. acidocaldarius*, we conducted an intensive gene ontology (GO) term analysis (Figure 4C/D). Interestingly,

this analysis showed that both knockouts have a preference for metabolic enzymes, especially for additional nitrogen containing amino acids pathways. Nevertheless, purine/nucleotide, acetyl-CoA and carbohydrate metabolism enzymes were also significantly enriched after knockout of either ArnA or ArnB. Despite a small difference in statistical significance and actual enrichment score the GO term analysis delivered quite similar results for both knockouts. Notably, when comparing both knockouts with each other, only 43-61% of the significantly altered proteins can be found in both deletion strains simultaneously (Figure 4G). However, those significantly altered proteins that can be found in both knockouts do share a very high correlation of 82-86% (Figure 4E/F). Together this leads to the conclusion, that around half of the impact of knocking out either ArnA or ArnB is apparently based on the interaction of them with each other to some degree. *Vice versa* this means that the other half of the effected protein levels are apparently independent from the ArnA-ArnB interaction or at least without direct correlation. A comparison of these proteomics data with a deletion variant of *S. acidocaldarius* that missed the GPN-loop GTPase *Sa*GPN and exhibited diminished motility [12] shows that the proteome changes of the hypermotile ArnA and/or ArnB deletion variants [13] are more modest. This suggests a more intimate involvement of ArnA and ArnB in the regulation of motility than of *Sa*GPN, which causes large-scale changes in the proteome network. Accordingly, there is no correlation for significantly altered protein levels between the *Sa*GPN knockout and the ArnA or ArnB knockouts (Figure S10). This suggests that the role of ArnA and ArnB in the archaeal regulatory network is independent of *Sa*GPN.

Evolutionary context of ArnB, which contains the Sec23/Sec24-core motif

In the current state of literature, the function of ArnB, a homolog of the vWA2 protein in *S. acidocaldarius* [5], has been associated with the archaellum regulatory network. It has been reported that ArnB is a negative regulator of the archaellin arlB, regulated by the interaction with its partner ArnA. However, during investigation of ArnB we found structural similarities with the membrane curvature proteins Sec23/24, which are central components of the assembly machinery for eukaryotic COPII vesicles. Sec23/24 harbor a domain arrangement highly reminiscent of ArnB and even include an N-terminal ZnF domain like ArnA. Accordingly, Sec23/24 contain as a core motif the vWFA, β-sandwich and C-terminal helical domain, the latter being similar to the ArnB_C domain, besides an additional C-terminal domain (Figure 5A). Superposition

of the Sec24 core motif (PDB: 1m2v, P301-I749 of chain B) and ArnB reveals a structural deviation of 5.7 Å for 302 Cα atoms. Here, differences between the Sec24 core motif and ArnB are mostly found for the length of helices in the helix pair α9/α10 of the ArnB_C domain and the relative orientation of the β-sandwich. Notably, the ZnF domain of Sec23/Sec24 and that of ArnA in the ArnAB complex pack to different sites of the β-sandwich of the core motif. In the ArnAB complex structure, the ZnF domain occupies a site made up by an edge between the β-sandwich and the ArnB_C domains, whereas in Sec23/Sec24 the ZnF-domain associates to the opposite site of the β-sandwich domain (Figure 5B). Another feature in this context is the packing between the Sec23/Sec24 core motif and the C-terminal domain. Although ArnB lacks the gelsolin-type C-terminal domain of Sec23/24, the FHA domain of ArnA is predicted to interact in a phosphorylation-dependent manner with the α-helix domain that resembles the packing of the Sec23/Sec24 motif and the C-terminal domain.

The close relationship between the ArnAB assembly and the eukaryotic Sec23/24 core motif prompt for a wider distribution of the Sec23/Sec24 core motif within the domains of life. Using foldseek and the ArnB crystal structure we found a wider occurrence for the ArnB/Sec23 core motif (Supplementary Dataset 3) than previously suggested by the structural relationship between vWA2 from *S. acidocaldarius* and a vWA protein from the actinobacterium *Catenulispora acidiphila* [5]. Structural orthologs are found in other bacterial phyla than the actinobacteria. For example, *Escherichia coli* has a predicted structural ortholog, YfbK, whose domain topology is found in ~5900 gene products (status 01/24, Figure S11A). Besides matching Alphafold2 models from Archaea including Euryarchaeota like *Halorubrum* we also find Sec23/Sec24 core motifs in other *Korarchaeota* and *Heimdallarchaeota* (Figure S11B). More surprisingly, the closest structural hits outside the archaeal domain are found in the plant and fungal kingdoms despite marginal sequence identities. Orthologs in plants like *Arabidopsis thaliana* and *Zea mays* (Figure S11C) currently lack an assigned function. Notably, foldseek hits in the animal kingdom, apart from the expected Sec23/Sec24 orthologs, lack a direct structural relationship to the Sec23/Sec24 core motif and ArnB (Figure S11D).

Discussion

This work presents much needed structural information, supported in both solution and crystal structure, of the ArnAB complex and features more information about its *in vivo*

role. The role of ArnA, ArnB and its hetero complex has so far been associated with the archaellum regulatory network, hence the name Arn [5,7]. Nevertheless, their biological role is not yet fully understood leaving open questions of why they are found in exosomes and why they share a structural homology with the Sec23/24 complex. Recently, it was reported that the ArnA homolog from *Saccharolobus islandicus* (*Sis*) interacts with promotor sequences supporting a transcription factor role for ArnA [14], although DNA binding was rather unspecific. However, during our work, we could not find any support that ArnA from *S. acidocaldarius* (*Sa*ArnA) itself is capable of binding DNA. Moreover, there are no structural indications that either ArnB or ArnA have DNA binding sites, especially since the ZnF domain of ArnA, which would be an obvious target, is part of a protein-protein interface. As *Sis*ArnA was purified from the native organism, its DNA binding activity may depend on additional interaction partners than ArnB or post-translational modifications, which require further elucidation.

Furthermore, our comprehensive proteomic data suggests involvement of ArnA and ArnB in anabolic metabolism pathways. Accordingly, knockout of either of them lead to a gene enrichment in enzymes concerning amino acid metabolism, especially N-rich amino acids, carbohydrates and acetyl-CoA metabolism. This is particularly resembled by changes in protein levels of the Gln-synthase (Q9HH09), succincyl-CoA ligase (Q4J9C0) or an MFS sugar transporter protein (Q4J7I0). Obviously, the known overproduction of the archaellum, an indicator of hypermotility, caused by the ArnA/ArnB knockouts [6], indicates *per se* changes of the metabolic repertoire in *S. acidocaldarius*, when switching from a an immobile, sessile form to a planktonic lifestyle. Compared to other proteomic data regarding proteins that are involved in the Arn-system like the GPN-loop GTPase from *S. acidocaldarius Sa*GPN, these changes are completely different [12]. While only half of the statistically relevant proteome changes is common between ArnA and ArnB deletion strains, the intrinsic connection between both is also mirrored by the nearly symmetrical alterations observed in proteins that play pivotal roles in both strains. Therefore, we are confident to say that the metabolic impacts we observed are directly related to the specific knockouts. Accordingly, the large-scale proteome impacts observed for the *Sa*GPN knockout appear to be particularly more random. Notably, as *Sa*GPN is being considered part of the Arn system, physical association of the ArnAB complex with *Sa*GPN might link to a collaborative function of all three of them. This could also explain why the previously reported PP2A co-IP assay [7] pulls down not only ArnA and ArnB but *Sa*GPN

as well. However, it may also be of coincidental nature and no physical association connects the ArnAB complex with *Sa*GPN. Nevertheless, the aspect of the ArnA-ArnB interaction corresponds to a ‘flip switch’ depending on the phosphorylation state of ArnB and its respective oligomeric state. These states are regulated by the PP2A and ArnC phosphatase-kinase system and induce heterodimerization as well as higher oligomerization upon threonine phosphorylation based on the FHA-pThr interaction. Interestingly, the oligomerization processes possibly involve a (partially) unwinding or loosening of the ArnB motifs. Accordingly, the stand-alone domain of PP2A exhibits a more restricted array of control modes compared to ArnC, which exhibits an elongated N-terminal region containing a TPR-like helical domain in addition to its C-terminal Ser/Thr kinase domain and thereby offering a more diverse range of regulatory mechanisms.

The Sec23/24-like domain architecture of ArnB with its β-sandwich, vWFA and C-terminal helical bundle domain is conserved among all domains of life. Captivatingly, the evolutionary connections between different kinds of proteins within all three domains of life is apparent in their structure with only moderate changes. However, due to a weak phylogenetic connection between them the functions of the eukaryotic Sec23/24 exert completely different functions than their archaeal predecessors. Therefore, the structural homology between ArnB features with Sec23/24 does not necessarily place ArnB with the COPII vesicle system, as structural homologs of ArnB can even be found in *Escherichia coli* (Figure S11A). Nevertheless, two main scenarios appear likely: function in vesicle transport either arose late, i.e. in Asgard Archaea with their symbiotic lifestyles with bacteria, or upon the transition to Eukaryotes, i.e. during the manifestation of endosymbiosis. Moreover, a prerequisite for a functional switch may be the association with exosomes as observed before [15], even when ArnA/ArnB are not directly involved in exosome formation.

Interestingly, ArnB demonstrates a phosphorylation-dependent alteration of its conformational flexibility as revealed by HDX-MS. Phosphorylation proceeds at multiple, but specific sites and follows a pattern consistent with a model in which ArnC-mediated phosphorylation starts at threonines located in the tip regions of the C-terminal helical bundle (T353-T363). Subsequently, it propagates to threonine sites, which are mostly buried within the helix bundle/vWFA domain interface (T371-T375), and finally extends to sites closer to the N-terminus of the helix bundle domain. It is worth noting that such a sequential phosphorylation and associated partial unfolding

of the protein is not without precedence [16,17]. For example, for $p19^{INK4d}$, an ankyrin-repeat protein, it is reported that the sequential phosphorylation within the helical repeat domain progressively destabilizes the N-terminal half of the protein, ultimately leading to the unfolding of the corresponding section [16]. Moreover, the casein kinase 1a (CK1a) of *Neurospora* progressively hyper-phosphorylates the clock protein FREQUENCY (FRQ), which triggers a conformational change driven by clustered phosphorylation [17]. In general, multi-phosphorylation exerts regulatory influence over protein structures, facilitating bidirectional transitions between ordered and disordered states [18]. Notably, the sequential part of the phosphorylation patterns apparently does not necessarily affect ArnB-FHA interaction, as different threonine mutants of ArnB were still able to bind the FHA domain of ArnA, as long as alternative phospho-sites were available (Figure 2C/D). However, these large-scale conformational changes observed in ArnB align with a dual-point attachment of ArnA facilitated by both its ZnF and FHA domains. This interaction could serve as the catalyst for the assembly of larger molecular complexes involving ArnA and ArnB, resembling but not identical to the established Sec23/Sec24 COPII coat formation.

Many of the intricacies surrounding the functionality and assembly of the ArnAB complex stem from uncertainties related to the oligomerization behavior, as well as the precise interaction point of the ArnA FHA domain and the role of the ArnA linker domain. The absence of structural support has led to various plausible scenarios. Consequently, we posit three primary operational modes governing the heterocomplex formation of ArnA and ArnB, based on our structurally supported findings. Firstly, these scenarios encompass the possibility of simple heterodimerization, where observed oligomers may lack biological relevance. This could be attributed to their potentially artificial nature *in vitro*, stemming from a deficiency in regulated phosphorylation, underscored by the varying incubation times with ArnC, as illustrated in this and previous studies [7] (Figure 6 top right). Secondly, it's conceivable that another element plays a role in ArnAB function. Both DNA and other proteins, such as *Sa*GPN and *Sa*UspA, identified in the previously reported co-IP assay on PP2A interaction partners, emerge as potential targets (Figure 6 mid right). Lastly, polymerization based on the FHA domain appears probable. Numerous phosphorylated states of the ArnAB complex [7], especially with extended ArnC incubation times, hint at a multifaceted interplay facilitated by the FHA domain-pThr interaction, thereby mimicking a Sec23/24-like behavior (Figure 6 bottom right).

In conclusion, the intriguing phosphorylation-dependent interaction between the archaellum regulatory network proteins, ArnA and ArnB, not only adds to our understanding of archaellum-related functions but also positions them as potential precursors to COPII homologs. This suggests a broader significance, possibly making them among the earliest COPII-like proteins, which were already present in TACK Archaea.

Methods

Materials

If not stated otherwise chemicals and materials were obtained from Carl Roth. Primers were ordered from Microsynth.

Protein expression

Overexpression of recombinant proteins of this study was performed in LB medium (10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract) supplemented with 34 µg/mL chloramphenicol, using BL21(DE3) Rosetta cells. After induction of gene expression (0.5 mM IPTG at OD_{600nm}=0.6-0.8) cultures were incubated overnight (16 °C, 150 rpm, 18 h). Harvesting (20 °C, 5,000 rpm, 20 min) was followed by resuspension in lysis buffer (150 mM NaCl, 50 mM Tris, pH=8.0). Cell pellets were either used freshly for protein purification or frozen in liquid N_2 and stored at -80 °C for later use.

Protein purification without phosphorylation

Cell pellets were lysed using a micro-fluidizer (2 min/2L cell pellet equivalent) after harvesting. The lysate was centrifuged (18,000 rpm, 20 °C, 20 min), the supernatant heat-treated (70 °C, 20 min) and centrifuged again (18,000 rpm, 20 °C, 20 min) before being loaded to a Ni-NTA column (5 mL, Cytiva). After sample application, the column was washed with 9 column volumes wash buffer (150 mM NaCl, 50 mM Tris, 25 mM imidazole, pH=8.0) and eluted with 5 column volumes elution buffer (150 mM NaCl, 50 mM Tris, 500 mM imidazole, pH=8.0). The elution was concentrated (<2.5 mL) and applied to a HiLoad® 16/60 Superdex® 200 pg (Cytiva) size-exclusion column, which had been equilibrated with running buffer (150 mM NaCl, 50 mM Tris, pH=8.0). Fractions containing the proteins, validated by SDS-PAGE, were collected, concentrated, frozen in liquid N_2 and stored at -80 °C.

Purification of the ArnAB complex - Phosphorylation of ArnB

For the generation of the phosphorylated ArnAB complex a 1L cell pellet of untagged ArnC, which was expressed and harvested as described above, was added to the ArnA-ArnB coexpression cell pellets before lysis. In addition, after the heat step ATP (Sigma Aldrich) and $MnCl_2$ (2.5 mM final concentration each) were added to the supernatant before incubation in a water bath (55 °C, 5-60 min). Purification of the phosphorylated complex was performed with Ni-NTA affinity chromatography and SEC as described above.

Crystallization and structure determination of the ArnAB complex

Freshly purified, equimolar protein of ArnA and ArnB (4.85 mg/mL) was filtered with a centrifuge reaction tube micro filter (Merck) prior to crystallization experiments. Initial crystal hits of the ArnAB complex grew in well F11 of the JCSG Core III Suite (NeXtal) but were unsuitable for structure determination. Accordingly, a fine screen hanging drop vapor diffusion experiment on 24-well plates by mixing 1 µL of crystallization condition with 1 µL protein solution was conducted. The final monoclinic crystals grew in the presence of 0.17 mol/L ammonium acetate, 0.085 mol/L sodium citrate pH 6.0, 25.5% (w/v) PEG 4000, 16.5% (v/v) glycerin after 5 months.

ArnAB data set was collected at the Swiss Light Source at Paul Scherrer Institute (*Switzerland*). Structure determination of the ArnAB complex was done by molecular replacement with the vWA2 structure (PDB: 5A8J) from *Sulfolobus acidocaldarius* as search model. Afterwards, the structure was refined by multiple rounds of manual structure building with Coot [19] followed by phenix.refine [20], which also led to electron density defining the ArnA ZnF-domain. Data collection and refinement statistics can be found in Table S2. Protein structure was deposited to the Protein Data Base and can be accessed via the code: 8SO5.

Small angle X-ray scattering (SAX)

SAXS datasets were collected at the synchrotron Bio-SAXS beamline BM29 (ESRF) Grenoble, France [21]. The wavelength λ= 1.0 Å and the sample-to-detector distance of 2.43 m resulted in scattering vectors, q, ranging from 0.0025 $Å^{-1}$ to 0.50 $Å^{-1}$. The scattering vector is defined as $q = 4\pi \sin\theta/\lambda$, where 2θ is the scattering angle. All experiments were performed at 20 °C and the data were initially processed by the ATSAS software package [22]. SAXS data collection was performed in 150 mM NaCl,

50 mM Tris (pH 8) buffer. One-dimensional datasets were subtracted from the buffer-only spectrum then merged and analyzed in Primus [23]. 1D-scattering of different concentrations were merged, where appropriate, given that the lower concentrations better represent low *q* data points. The radius of gyration was calculated by ScÅtter [24]. The I_0-values were calculated considering bovine serum albumin (BSA) as standard, where the R_g was 33.6 ±0.4 Å and I_0 93.13 ±0.1. P(r) distance distribution functions were calculated and refined by the program ScÅtter. The distance *r*, where the *P*(*r*) functions approach zero probability, identifies the maximal dimension (D_{max}) of the macromolecule. Using the refined data from ScÅtter in GNOM-format *ab initio* models for the corresponding SAXS-derived electron densities were calculated by averaging 20 rounds of DENSS [25]. These electron densities were fitted with structures and finally visualized by ChimeraX. The SAXS data sets are deposited in SASBDB [26] with accession codes SASDUW2.

Hydrogen/deuterium exchange mass spectrometry (HDX-MS).

Investigated proteins (ArnA, ArnB, ArnB~P, ArnA/ArnB~P) were employed as 50 µM concentrated stocks solutions in 20 mM Tris-Cl pH 8.0, 150 mM NaCl. From these, HDX reactions were prepared by an autosampler (LEAP Technologies) as follows: 6.5 µL of protein stock solution were pre-dispensed in a 96-well plate, 58.5 µL of a buffer (20 mM Tris-Cl pH 8.0, 150 mM NaCl) prepared with 99.9% D_2O added and incubated at 25 °C for 10/30/100/1,000/10,000 s. The HDX reaction was stopped by transferring 55 µL of the reaction to 55 µL of quench solution (400 mM KH_2PO_4/H_3PO_4, 2 M guanidine-HCl (pH 2.2) pre-dispensed in another 96-well plate cooled down to 1 °C. 95 µL of the quenched reaction were injected into an ACQUITY UPLC M-Class system with HDX Technology (Waters) operating at 0.5 °C through a 50-µL injection loop. Non-deuterated samples were generated analogously with an H_2O-based buffer.

Samples were washed out of the loop with water + 0.1% (v/v) formic acid (100 µL/min) and digested at 12 °C in a cartridge (2 cm x 2 mm) filled with either bead-immobilized porcine pepsin or a 1:1 mixture of bead-immobilized protease type XVIII from *Rhizopus* sp. and protease type XIII from *Aspergillus saitoi*. The resulting peptides were collected on a trap column (2 mm x 2 cm; 0.5 °C) filled with POROS 20 R2 reversed phase resin (ThermoFisher Scientific). After 3 min of digestion and trapping, the trap was placed in line with an ACQUITY UPLC BEH C18 1.7 µm 1.0 x 100 mm column (Waters) operating at 0.5 °C and the peptides eluted with a gradient of water + 0.1% (v/v) formic

acid (eluent A) and acetonitrile + 0.1% (v/v) formic acid (eluent B) at 30 µL/min, as follows: 0-7 min/95-65% A, 7-8 min/65-15% A, 8-10 min/15% A. The peptides were guided to a Synapt G2-Si mass spectrometer (Waters) and ionized by electrospray ionization (capillary temperature: 250 °C; spray voltage: 3.0 kV). Mass spectra were acquired with MassLynX MS 4.1 (Waters) over 50 to 2,000 *m/z* in enhanced high-definition MS ($HDMS^E$)[27,28] or high-definition MS (HDMS) mode for non-deuterated and deuterated samples, respectively. A short spray of [Glu1]-fibrinopeptide B standard (Waters) every 45 s was employed for lock-mass correction. During peptide separation, the protease column was washed three times with 80 µL of 0.5 M guanidine-HCl in 4% (v/v) acetonitrile, and blank runs (double-distilled H_2O) were performed between each sample. All measurements were performed in triplicates for each protease column type (separate HDX reactions).

Peptides were identified ProteinLynx Global SERVER 3.0.1 (PLGS, Waters) and DynamX 3.0 (both Waters) from the non-deuterated samples acquired with $HDMS^E$ as described previously[29] employing low-energy, elevated-energy and intensity thresholds of 300, 100 and 1,000 counts, respectively and matched using a database containing the amino acid sequences of ArnA, ArnB, porcine pepsin, and their reversed sequences (peptide tolerance = automatic; fragment tolerance = automatic; min fragment ion matches per peptide = 1; min fragment ion matches per protein = 7; min peptide matches per protein = 3; maximum hits to return = 20; maximum protein mass = 250,000; primary digest reagent = non-specific; missed cleavages = 0; false discovery rate = 100). For quantification of deuterium incorporation with DynamX, peptides had to fulfil the following criteria: identification in two-thirds of the non-deuterated samples; minimum intensity of 30,000 counts; maximum length of 30 residues; minimum number of products of three and 0.1 products per residue; maximum mass error of 25 ppm; retention time tolerance of 0.5 minutes. Hereby, the datasets generated with porcine pepsin or after digestion with proteases type XIII and XVIII were pooled, and all spectra manually inspected. Residue-specific deuterium uptake of peptides was calculated with DynamX. If a residue was covered by a single peptide, the residue-specific deuterium uptake was equal to that of the whole peptide. In the case of overlapping peptides for a given residue, the residue-specific deuterium uptake was determined by the shortest peptide covering that residue. When multiple peptides were of the shortest length, the peptide with the residue closest to the C-terminus was utilized.

Proteomics data evaluation

Sample preparation and measuring of the *S. acidocaldarius* proteomes was performed as previously described [12] in collaboration with the bioanalytics MarMass facility of the Philipps-University of Marburg.

The timsTOF data was analyzed utilizing MaxQuant 2.1.3. Sequencing information for *S. acidocaldarius* DSM 639, comprising 2,222 entries, was retrieved from the UniProt database. For label-free quantification (LFQ) analysis, MaxQuant parameters were configured as follows: maximum peptide mass of 4,000 Da, allowance of up to three modifications per peptide, up to three missed cleavages, and the necessity of MS/MS for LFQ comparisons; all other parameters remained at default settings. Protein abundances from biological triplicates of the ΔArnA, ΔArnB and the wild type strain were filtered using a two-tailed student's t-test ($p<0.05$). The remaining proteins were visualized in protein heat maps, with further filtering based on a mean abundance difference of > 0.5 $\log_2$ fold-change. Visualization was done using the pheatmap R package [30] in addition to the gplots package [31]. The volcano plots were constructed by plotting the $\log_2$ fold-change in protein abundance on the x-axis against the negative logarithm (base 10) of the p-values (calculated from the student's t-test) on the ordinate. Plots were created using the ggplot2 [32] and ggrepel [33] packages in R, providing a comprehensive visualization of the proteomic changes in the context of significance levels. Gene Ontology (GO) enrichment analysis was performed on the sets of significant proteins ($p<0.05$) from each comparison using GOATOOLS library version 1.2.3 [34]. The significance of the GO enrichment analysis was determined by applying the Benjamini-Hochberg procedure with a false discovery rate (FDR) correction at a significance threshold of $p=0.05$. The analysis was carried out with gene ontology terms specific to *S. acidocaldarius* DSM639, obtained via the QuickGO [35] annotation tool.

Mutagenesis of ArnB threonines

For point mutation of the different ArnB mutants 10-60 ng of starting vector was mixed with 150 ng of corresponding primers (Table S1), 1.5 µL DMSO, 1 µL dNTP mix (NEB), 0.5 µL of Phusion DNA-polymerase stock (~6 mg/mL) and adjusted to 50 µL with H_2O. The PCR was performed for 18 cycles, primer suitable annealing temperatures (50-56 °C) and an elongation temperature of 72 °C. PCR was followed by a DpnI (2 µL, NEB) digest (37 °C, 2h) and transformation into *E. coli* DH5α cells. Afterwards plasmids

were extracted from overnight cultures using the QIAprep Spin Miniprep Kit (Quiagen) and checked by Sanger sequencing.

Data availability

All data can be found in the supplementary information text or excel files. Protein structures have been uploaded to the PDB.

Acknowledgements

We thank the beamline staff of the Swiss Light Source (SLS), PSI, Villigen, Switzerland, as well as the beamline staff of the ESRF, Grenoble, France, for their support. We also want to thank the staff of Marburg Mass spectrometry and crystallization facility (MarXtal) for technical support. We acknowledge support by the German Research Council (DFG) through the core facility for HDX-MS (project numbers 260989694 and 324652314 to Gert Bange, Marburg). We want to thank the Life? program of the Volkswagen Foundation for funding (Grant: Az-96727).

620 **Bibliography**

621 1. Jarrell, K. F. & Albers, S. V. The archaellum: an old motility structure with a new
622 name. *Trends Microbiol.* **20**, 307–312 (2012).

623 2. Hoffmann, L. *et al.* Expanding the archaellum regulatory network – the eukaryotic
624 protein kinases ArnC and ArnD influence motility of Sulfolobus acidocaldarius.
625 *Microbiologyopen* **6**, 1–14 (2016).

626 3. Lassak, K. *et al.* Molecular analysis of the crenarchaeal flagellum. *Mol. Microbiol.*
627 **83**, 110–124 (2012).

628 4. Lassak, K., Peeters, E., Wróbel, S. & Albers, S. V. The one-component system
629 ArnR: a membrane-bound activator of the crenarchaeal archaellum. *Mol.*
630 *Microbiol.* **88**, 125–139 (2013).

631 5. Hoffmann, L. *et al.* Structure and interactions of the archaeal motility repression
632 module ArnA-ArnB that modulates archaellum gene expression in Sulfolobus
633 acidocaldarius. *J. Biol. Chem.* **294**, 7460–7471 (2019).

634 6. Reimann, J. *et al.* Regulation of archaella expression by the FHA and von
635 Willebrand domain-containing proteins ArnA and ArnB in Sulfolobus
636 acidocaldarius. *Mol. Microbiol.* **86**, 24–36 (2012).

637 7. Ye, X. *et al.* The Phosphatase PP2A Interacts With ArnA and ArnB to Regulate
638 the Oligomeric State and the Stability of the ArnA/B Complex. *Front. Microbiol.*
639 **11**, 1849 (2020).

640 8. Henneman, B., van Emmerik, C., van Ingen, H. & Dame, R. T. Structure and
641 function of archaeal histones. *PLoS Genet.* **14**, e1007582 (2018).

642 9. Mahajan, A. *et al.* Structure and function of the phosphothreonine-specific FHA
643 domain. *Sci. Signal.* **1**, re12 (2008).

644 10. Nishi, H., Shaytan, A. & Panchenko, A. R. Physicochemical mechanisms of
645 protein regulation by phosphorylation. *Front. Genet.* **5**, 1–10 (2014).

646 11. Kõivomägi, M. *et al.* Multisite phosphorylation networks as signal processors for
647 Cdk1. *Nat. Struct. Mol. Biol.* **20**, 1415–1424 (2013).

648 12. Korf, L. *et al.* Archaeal GPN-loop GTPases involve a lock-switch-rock

mechanism for GTP hydrolysis. *MBio* **14**, e0085923 (2023).

13. Reimann, J. *et al.* Archaeal signal transduction: Impact of protein phosphatase deletions on cell size, motility, and energy metabolism in sulfolobus acidocaldarius. *Mol. Cell. Proteomics* **12**, 3908–3923 (2013).

14. Jiang, Z. *et al.* The FHA domain protein ArnA functions as a global DNA damage response repressor in the hyperthermophilic archaeon Saccharolobus islandicus. *MBio* **14**, e0094223 (2023).

15. Ellen, A. F. *et al.* Proteomic analysis of secreted membrane vesicles of archaeal Sulfolobus species reveals the presence of endosome sorting complex components. *Extremophiles* **13**, 67–79 (2009).

16. Kumar, A. *et al.* Phosphorylation-induced unfolding regulates p19INK4d during the human cell cycle. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 3344–3349 (2018).

17. Querfurth, C. *et al.* Circadian conformational change of the Neurospora clock protein FREQUENCY triggered by clustered hyperphosphorylation of a basic domain. *Mol. Cell* **43**, 713–722 (2011).

18. Thapar, R. Structural Basis for Regulation of RNA-Binding Proteinsby Phosphorylation. *ACS Chem. Biol.* **10**, 652-666 (2015).

19. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).

20. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Cryst.* **68**, 352–367 (2012).

21. Pernot, P. *et al.* Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution. *J. Synchrotron Radiat.* **20**, 660–664 (2013).

22. Petoukhov, M. V. *et al.* New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45**, 342–350 (2012).

23. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).

24. Tully, M. D., Tarbouriech, N., Rambo, R. P. & Hutin, S. Analysis of SEC-SAXS

data via EFA deconvolution and Scatter. *J. Vis. Exp.* **167**, e61578 (2021).

25. Grant, T. D. Ab initio electron density determination directly from solution scattering data. *Nat. Methods* **15**, 191–193 (2018).

26. Valentini, E., Kikhney, A. G., Previtali, G., Jeffries, C. M. & Svergun, D. I. SASBDB, a repository for biological small-angle scattering data. *Nucleic Acids Res.* **43**, D357–D363 (2015).

27. Geromanos, S. J. *et al.* The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* **9**, 1683–1695 (2009).

28. Li, G. Z. *et al.* Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **9**, 1696–1719 (2009).

29. Joiner, J. D. *et al.* HilE represses the activity of the Salmonella virulence regulator HilD via a mechanism distinct from that of intestinal long-chain fatty acids. *J. Biol. Chem.* **299**, 105387 (2023).

30. Kolde, R. CRAN - Package pheatmap. (2019).

31. Warnes, G. R. *et al.* Various R Programming Tools for Plotting Data [R package gplots version 3.1.3]. (2022).

32. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis. Springer* (2016).

33. Slowikowski, K. Automatically Position Non-Overlapping Text Labels with 'ggplot2' [R package ggrepel version 0.9.4]. (2023).

34. Klopfenstein, D. V. *et al.* GOATOOLS: A Python library for Gene Ontology analyses. *Sci. Rep.* **8**, 1–17 (2018).

35. Binns, D. *et al.* QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045 (2009).

Figures

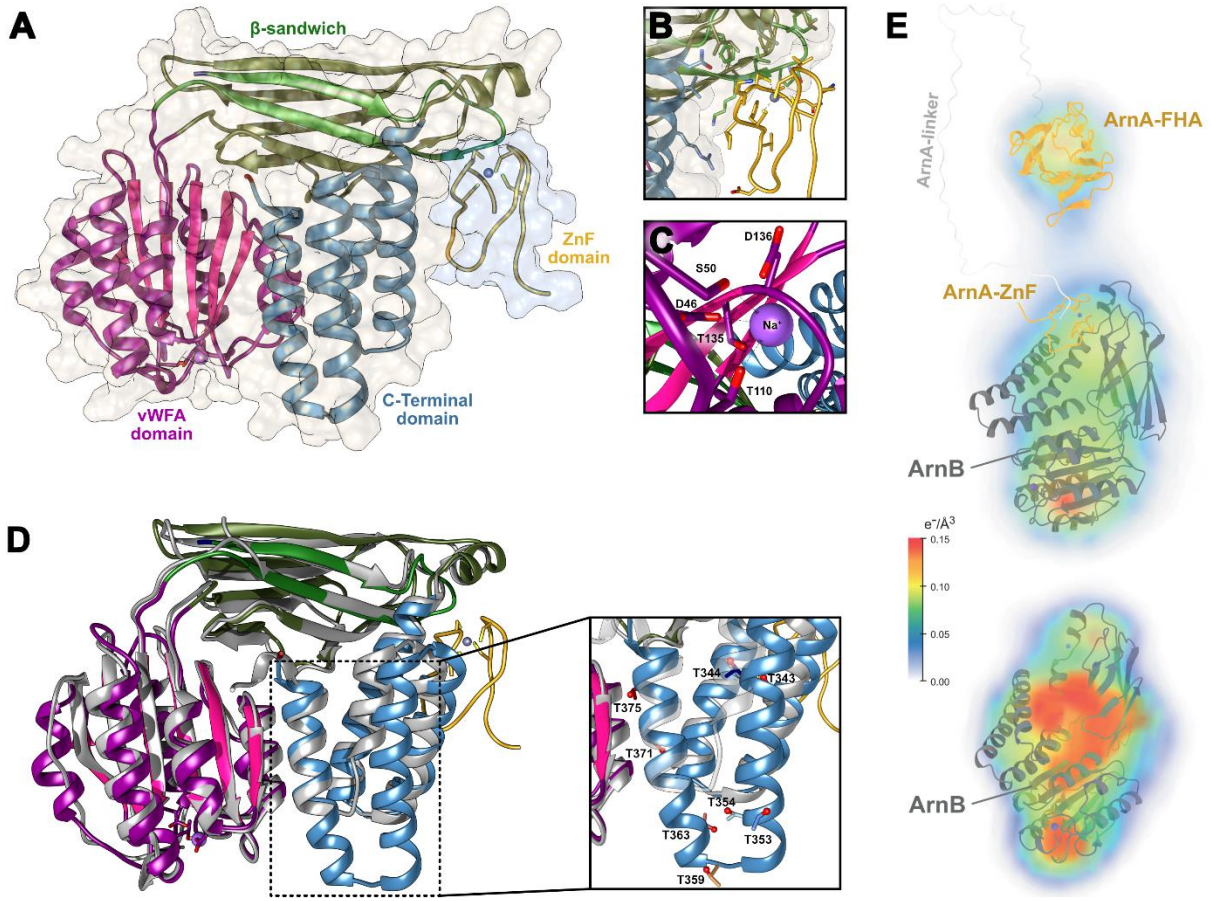

Figure 1: Overall Structure of the ArnAB complex and comparison with *Sa*vWA2. A: Full structure of ArnAB complex and domain overview showing the ArnA ZnF domain interaction site. Domains are color-coded as follows: β-sandwich (green), vWFA (pink/purple), C-terminal domain (blue), ZnF domain (yellow). **B:** Focus view on the ZnF interaction site with important residues showing as sticks. **C:** MIDAS site of the vWFA domain of ArnB reveals Na^+ interacting residues. **D:** ArnB and vWA2 (grey) superposition with secondary structure elements numbering. The magnification shows the C-terminal domain of ArnB with threonine residues shown as sticks and OH-groups as balls, structurally highlighting potential FHA interaction sites. **E:** *Ab initio* envelope as calculated from SAXS-derived scattering vectors using 20 iterations of DENSS[25] for electron density modeling and ScÅtter[24] for refining GNOM-processed SAXS data. For the ArnAB complex (top) a D_{max} value of 15.7 nm was derived (R_g 4.16 nm), the DENSS model density has a resolution of 33 Å; the values for ArnB (bottom) alone are D_{max}=7.5 nm, Rg=2.41 nm, resolution=30 Å. Structural models for the FHA domain of ArnA (orange), ArnB (grey) and the ArnB/ArnA-ZnF complex were fitted by ChimeraX into SAXS densities.

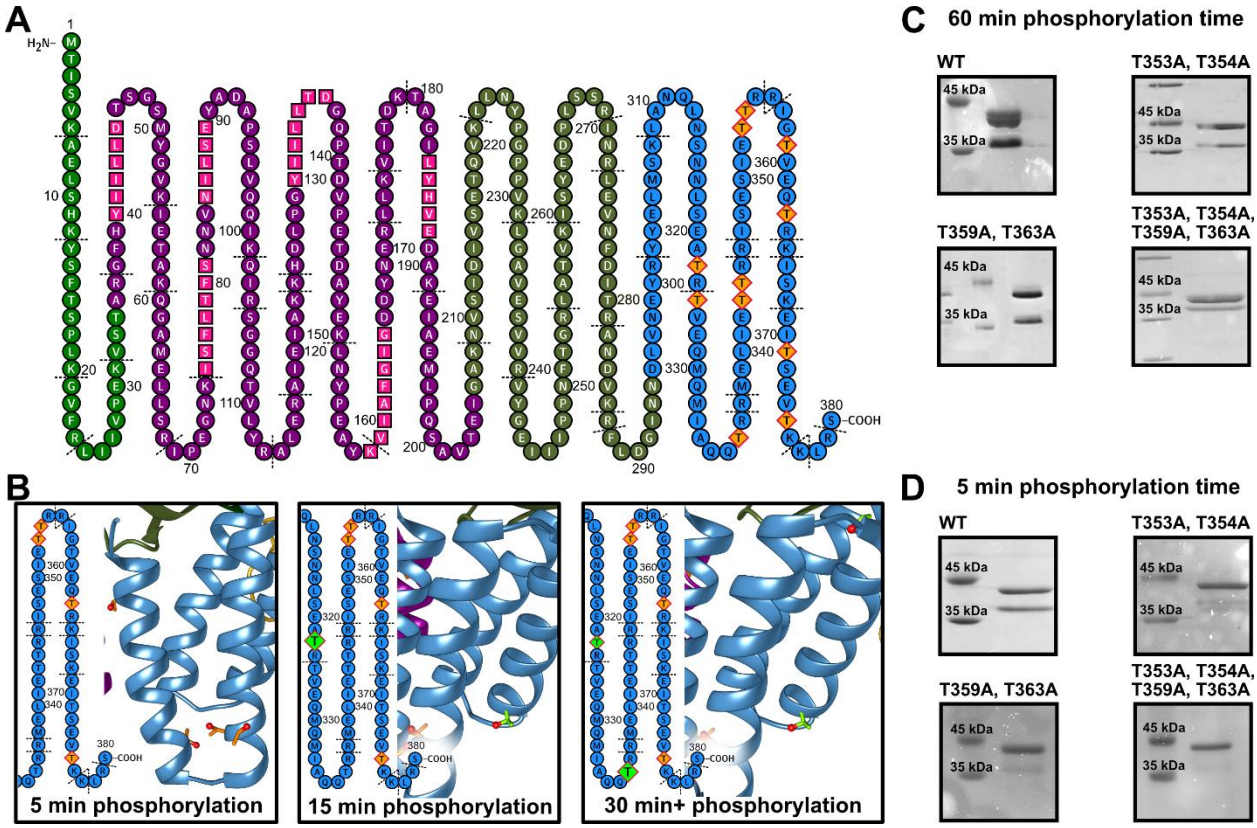

Figure 2: Promiscuity of ArnB phosphorylation sites. A: Sequence view of ArnB with trypsin cleavage sites, amino acid count and highlighted all potential phosphorylation sites found during this study. Domain coloring is analogous to Figure 1. **B:** Impact of phosphorylation time on threonine phosphorylation by MS. Green highlights newly found threonines at the respective time point. **C/D:** SDS gel pulldown assays showing that phosphorylation time has an impact on ArnB mutants, as different ArnB mutants (upper bands) are not able to sufficiently interact with ArnA (lower bands) when phosphorylation time is limited and, thus, cannot be pulled down anymore.

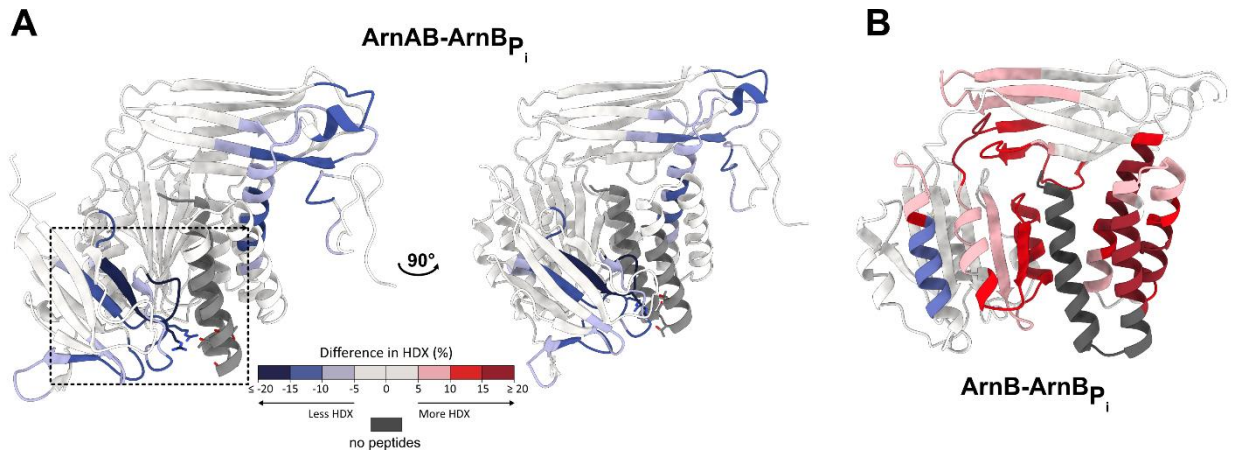

Figure 3: HDX reveals ArnA ArnB interaction sites and impact of phosphorylation. A: Cumulative HDX, i.e. highest scoring change per any given time point, of the phosphorylated ArnAB where the phosphorylated ArnB HDX is subtracted from, revealing the interaction sites of ArnB with ArnA. The dashed box focuses on the FHA interaction with the C-terminal helix as a possible interaction site. The legend is representative for the whole figure. B: Impact of phosphorylation on ArnB as shown by cumulative HDX of ArnB-phosphorylated ArnB. The changes in HDX suggest a structure wide impact of the phosphorylation.

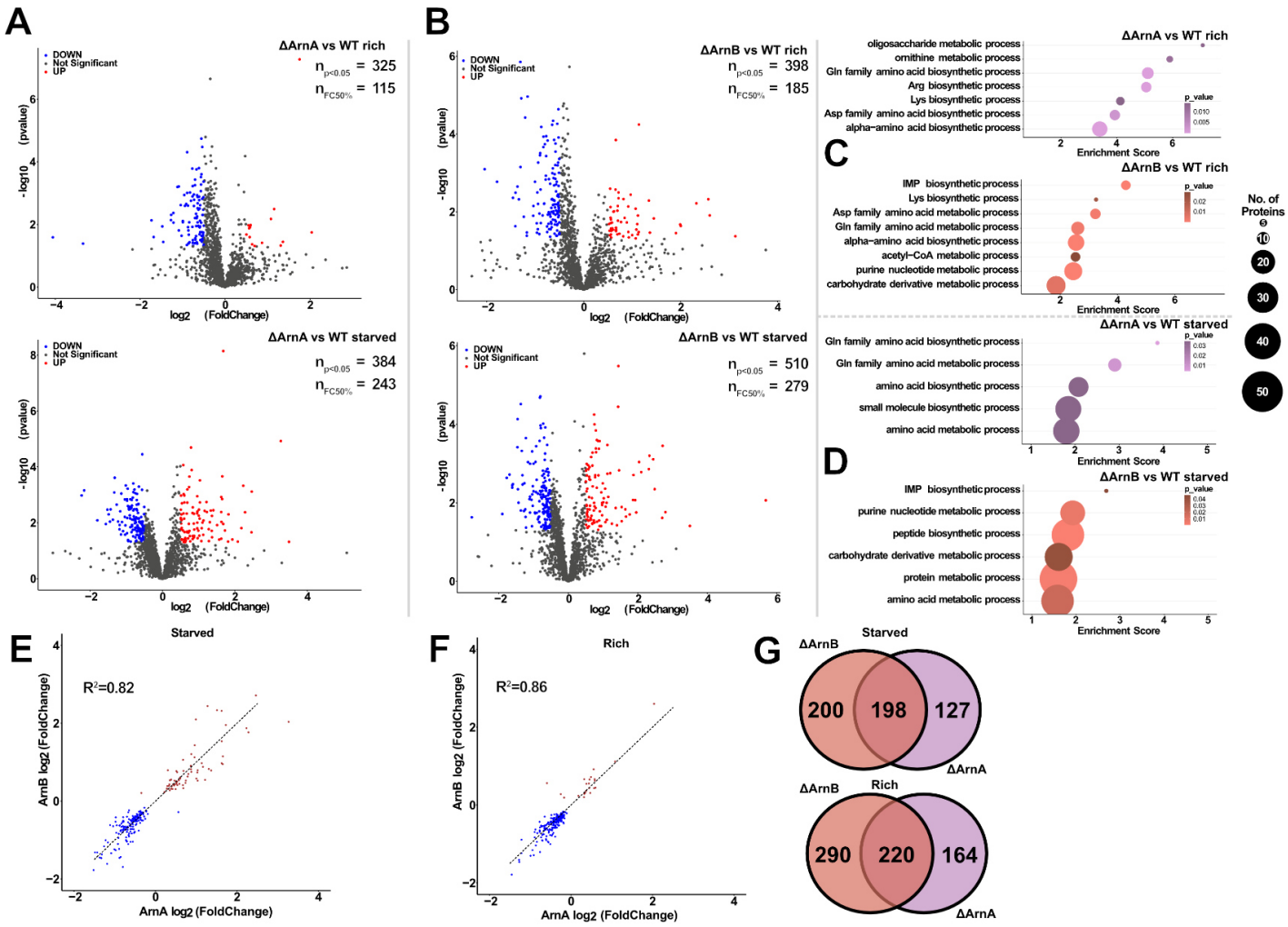

Figure 4: Proteomics analysis of *Sulfolobus acidocaldarius* Δ*arnA* and Δ*arnB* strains in nutrient rich and starved conditions compared to the wild type. A/B: ArnA/B knockout proteomic analysis visualized as volcano plots in respect to the corresponding WT nutritional state. The top plots show nutrient rich conditions and the bottom plots nutrient starved conditions. Sample size comprises n=3 biological triplicates per sample, including respective WT samples, as well as two technical duplicates per measurements. Protein levels that have p-value < 0.05 and a $\log_2$ Fold Change of >0.5 (i.e. >50%) are colored blue, i.e. down regulated in the respective knockout and nutritional state, or red, i.e. up regulated in the respective knockout and nutritional state. **C/D:** Gene enrichment analysis of respective protein hits, clustering them into GO terms revealing mostly metabolic involvement. **E/F:** Scatter plots of proteins found in both ArnA and ArnB knockout reveal high cross-effected protein rate, highlighting their cooperative role in *S. acidocaldarius*. **G:** Venn diagrams of ArnA and ArnB knockouts show that only ~half of the affected proteins (p<0.05) can be found in both knockouts, leaving room for individual roles.

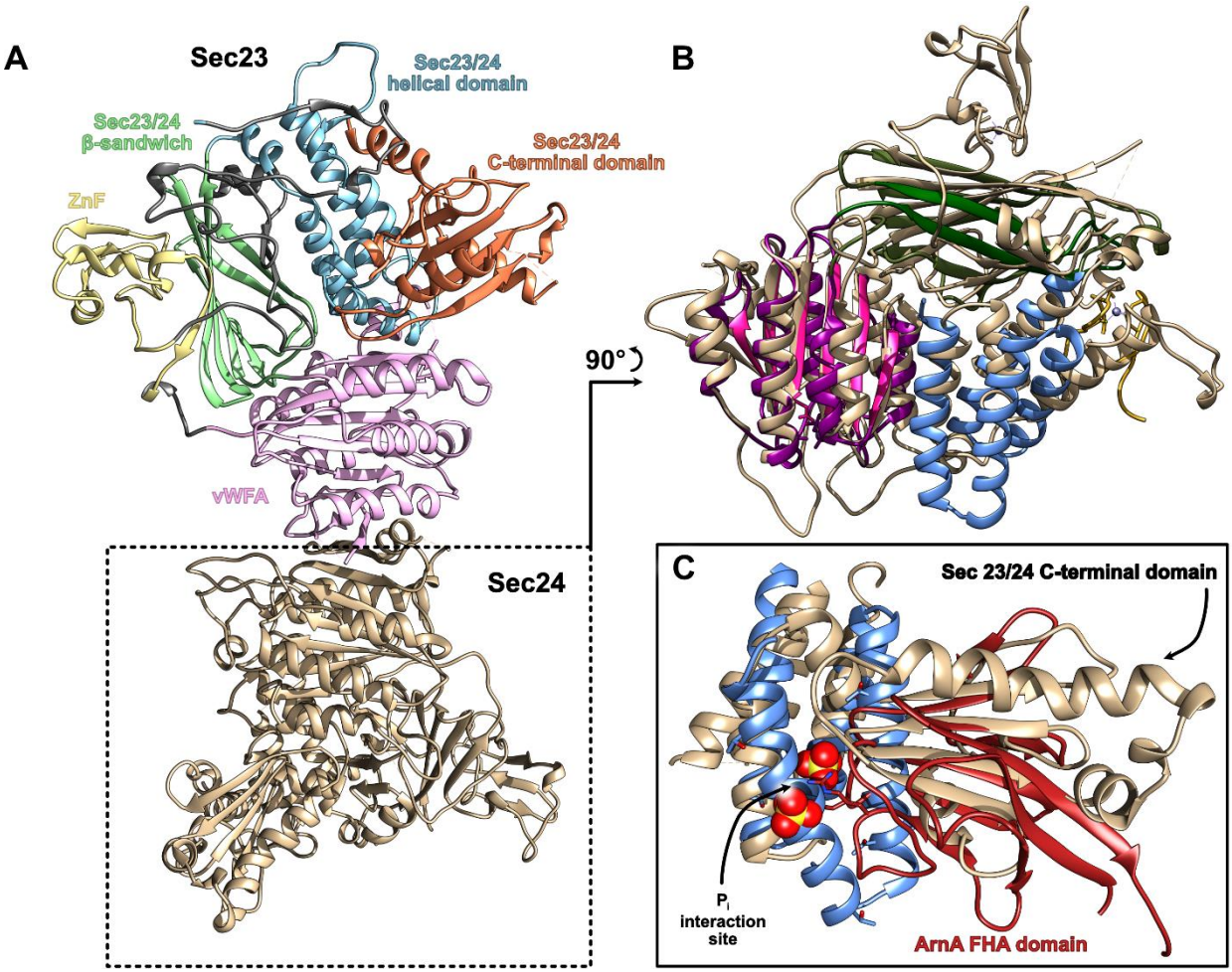

Figure 5: ArnB is a structural homolog of the COPII main membrane curvature factor Sec23/24. **A:** Sec23/24 complex with analogous domain coloring to ArnB but in paler colors. Accordingly, the Sec23/24 helical domain represents the C-terminal domain of ArnB and the C-terminal Sec23/24 domain is colored orange. **B:** Superposition of Sec23/24 with ArnB without the C-terminal domain of Sec23/24 for clarification reasons, showing high structural alignment. **C:** Superposition of ArnA FHA domain onto the C-terminal domain of Sec23/24 also reveals structural similarities and aligns well with a potential interacting position in a respective ArnAB complex.

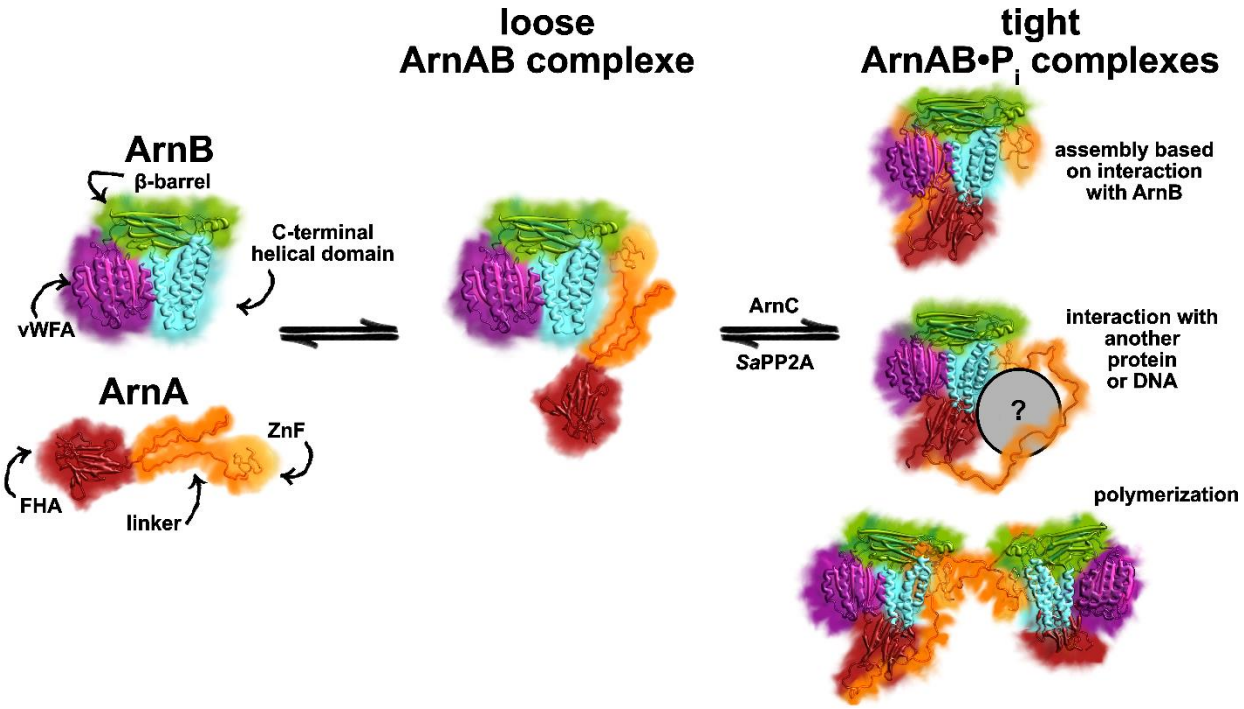

Figure 6: Role of heterodimerization to the ArnAB complex in different functional scenarios. The phosphorylation-dependent formation of tight ArnAB complexes due to tethering via ZnF domains and FHA-pThr interaction apparently plays a role in the archaellum regulatory network and several metabolic pathways. A 'loose' ArnAB complex is formed by the 1:1 interaction between the ArnA-ZnF domain and ArnB. This single tether is augmented by ArnC-dependent phosphorylation of the ArnB_C domain and the ArnA-FHA domain (tight ArnAB complex). Double-tethered ArnAB complexes may exist as stoichiometric 1:1 assemblies (top, right) or multimeric higher complexes prone to polymerization (bottom, right). In the tight ArnAB complex and the ArnA's long linker region may enable binding of further interaction partners (mid, right). In this context, transition to loose ArnAB complexes and disassembly is driven by dephosphorylation due to PP2A triggering further arlB production.

Supplementary Information for

Sequential conformational transition of ArnB, an archaeal ortholog with Sec23/Sec24 core motif

Lukas Korf[1], Wieland Steinchen[1,2], Mohamed Watad[1], Filipp Bezold[1], Marian S. Vogt[1], Leon Selbach[1], Alexander Penner[1], Maxime Tourte[3], Sebastian Hepp[1], Sonja-Verena Albers[3], Lars-Oliver Essen[1,*]

[1]Department of Chemistry, Philipps University, Hans Meerwein-Str. 4, 35043 Marburg, Germany
[2]Center for Synthetic Microbiology, Philipps University Marburg, Karl-von-Frisch-Str. 14, 35043 Marburg, Germany
[3]University of Freiburg, Institute for Biology II, Molecular Biology of Archaea, Schaenzlestrasse 1, 79104 Freiburg, Germany
[4]Signalling Research Centres BIOSS and CIBSS, University of Freiburg, 79104 Freiburg, Germany
*corresponding author

Email: essen@chemie.uni-marburg.de
sonja.albers@biologie.uni-freiburg.de

This PDF file includes:

Supplementary text
Figures S1 to S11
Table S1 to S2

Other supplementary materials for this manuscript include the following:

Supplementary Dataset 1 (HDX data)
Supplementary Dataset 2 (proteomics data)

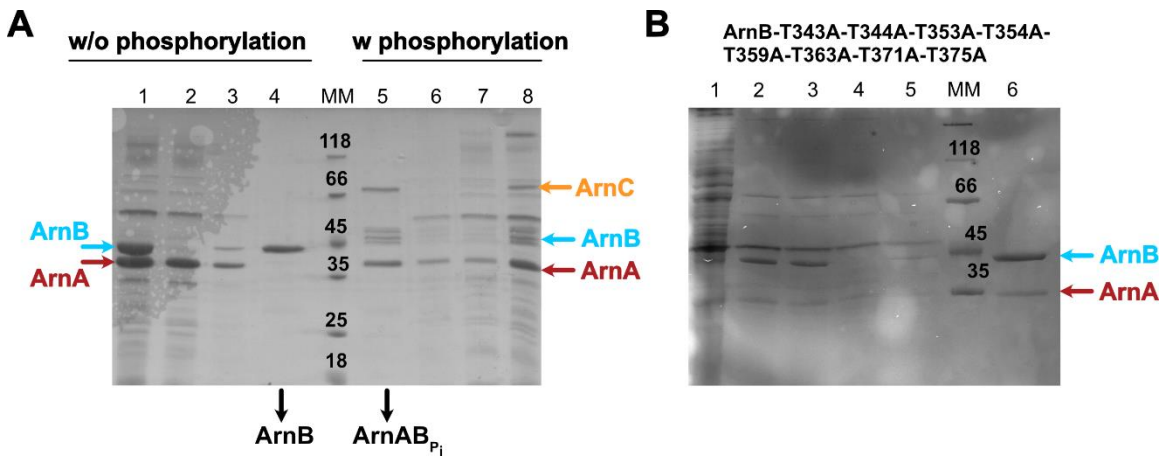

Figure S1: ArnA/ArnB pulldown assays. A: SDS-PAGE with a comparison of Ni-NTA column purifications of ArnA and ArnB without phosphorylation (left part, lanes 1-4) and with phosphorylation by ArnC (right side, lanes 5-8). ArnB is only able to pulldown ArnA when they were subjected to prior phosphorylation. 1: heat step, 2: flow through, 3: wash, 4: elution, MM: molecular marker, 5: elution, 6: wash, 7: flow through, 8: heat step/ P_i assay. **B:** Ni-NTA purification of the multiple Thr→Ala mutant of ArnB after 60 min phosphorylation by ArnC. Under these prolonged phosphorylation times this mutant of ArnB depleted of threonines in the α-helical domain is still able to pulldown ArnA. 1: lysate, 2: heat step, 3: P_i assay, 4: flow through, 5: wash, 6: elution.

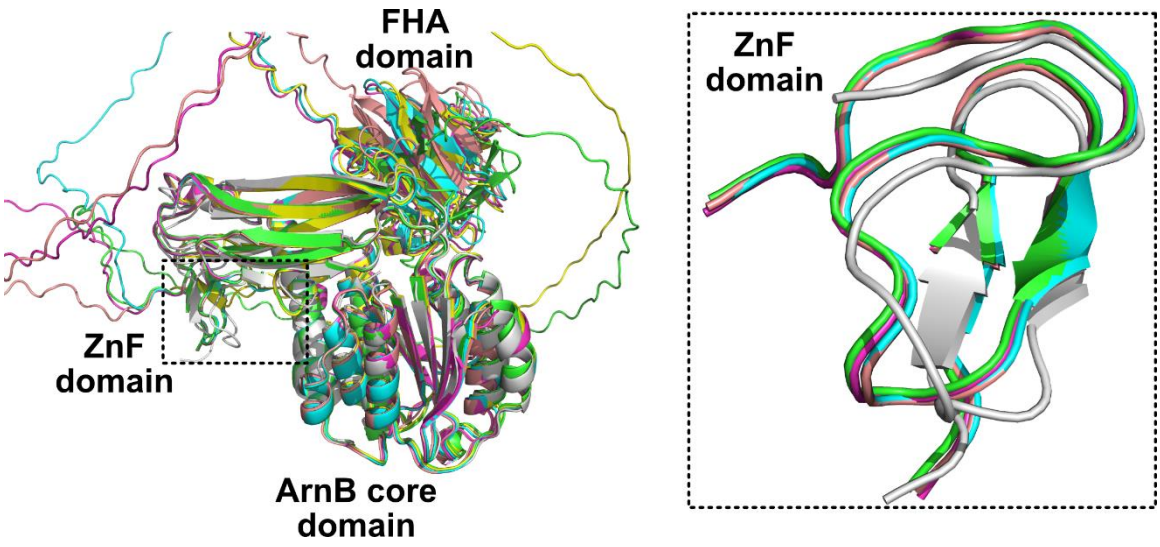

Figure S2: Superposition of experimental and predicted ArnAB complex. Theoretical complexes for ArnAB were predicted by the AlphaFold2-Multimer pipeline and superimposed to the ArnAB crystal structure. Left, complete alignment showing the different models. Disordered regions of the AF2 models correspond to the linker region between the ZnF and FHA domains of ArnA. Right, detail view on the ZnF domains in the ArnAB complex showing only minor deviations from the ArnAB cocrystal structure (gray). The five AF2 models of the ArnAB complex are colored according to their rank in green, cyan, pink, yellow and salmon.

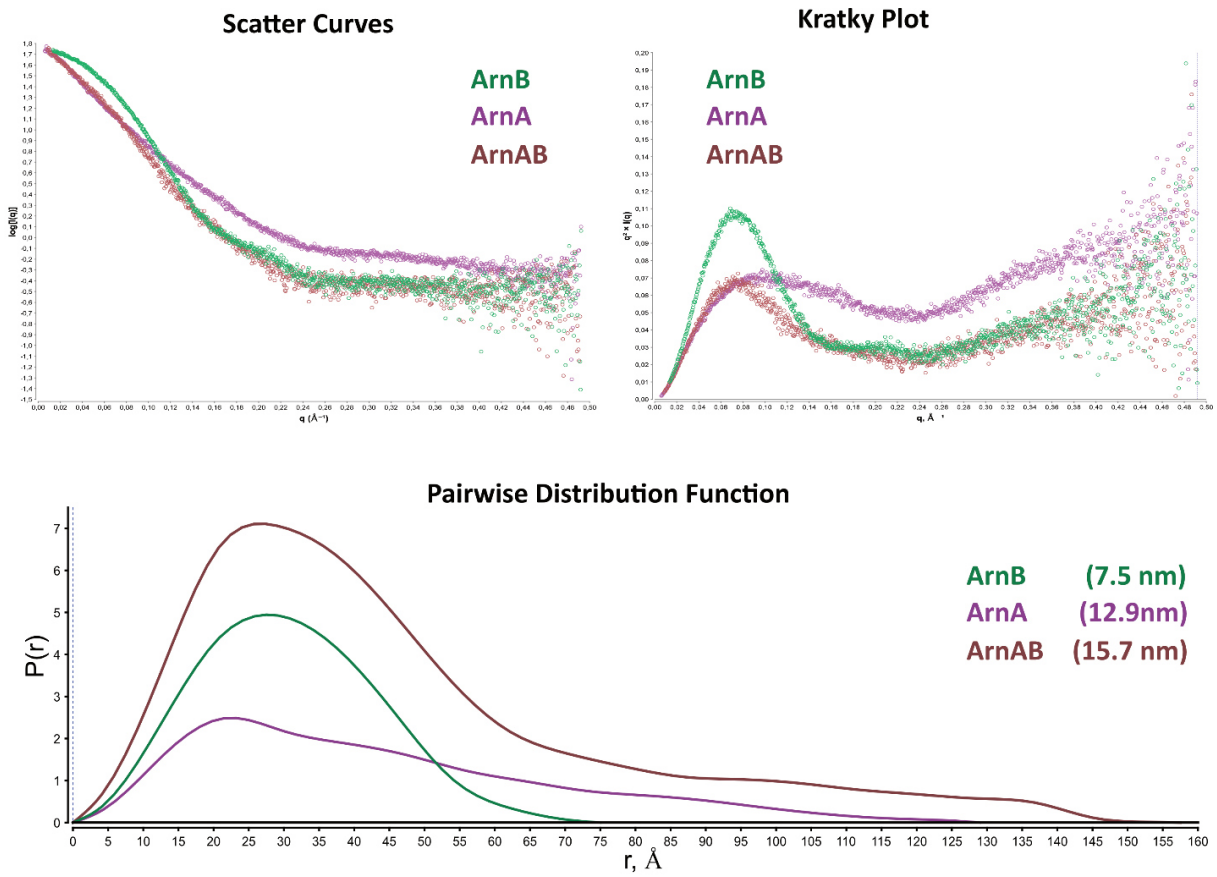

Figure S3: SAXS data analysis with scatter curves, Kratky plot and pairwise distribution function. The plots show the average of technical triplicates for each of the three SAXS samples: loose ArnAB complex (brown), ArnA (purple) and ArnB (green) with D_{max} shown in parentheses.

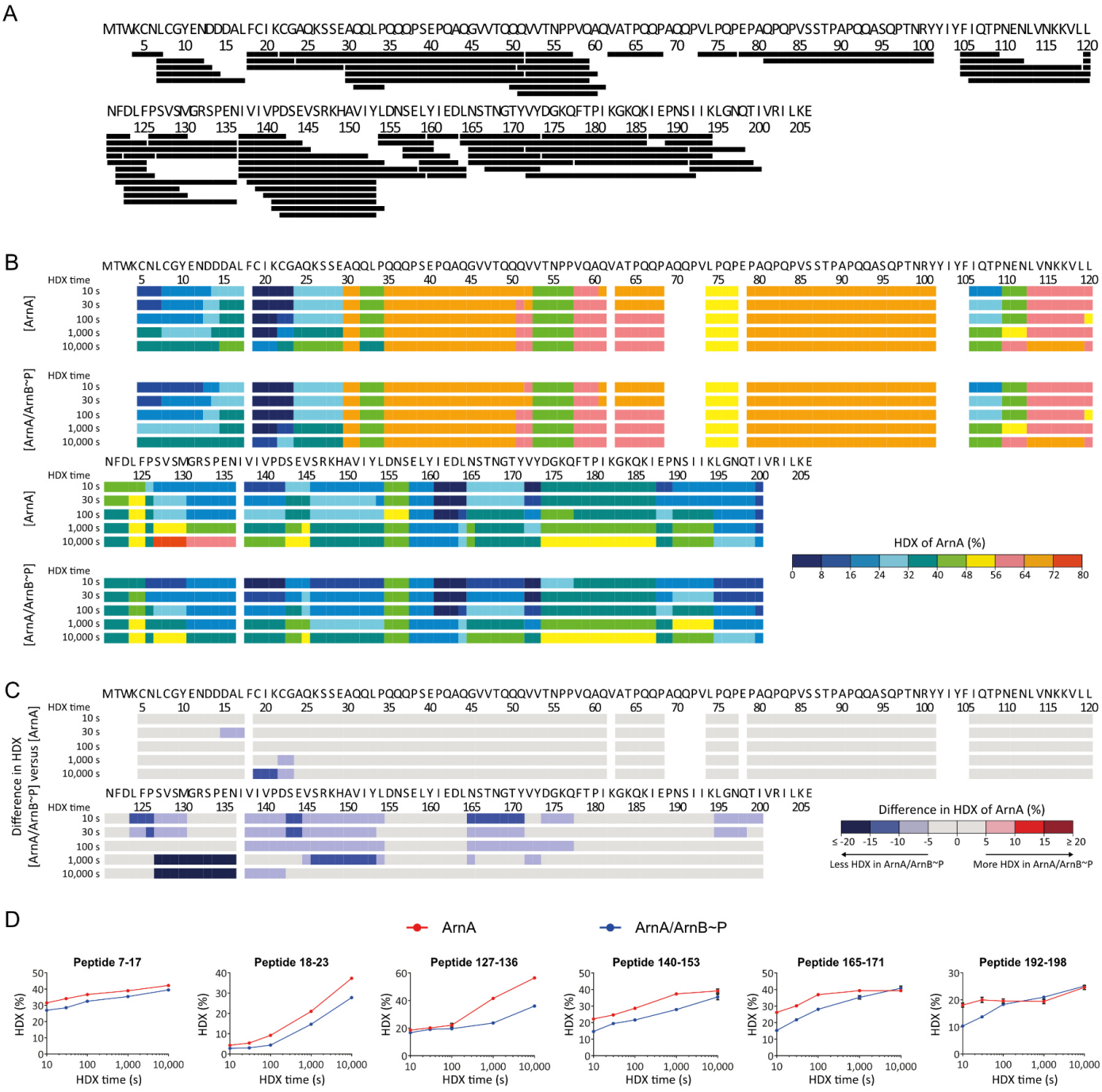

Figure S4: HDX supplementary data evaluation of ArnA. A: Each black bar represents a peptide of ArnA identified in HDX-MS. **B:** The residue-specific HDX of individual ArnA and that of ArnA in complex with ArnB~P is projected onto the ArnA amino acid sequence. **C**: The difference in residue-specific HDX between ArnA in complex with ArnB~P and individual ArnA is projected onto the ArnA amino acid sequence. Blue color denotes reduced HDX of ArnA when in complex with ArnB~P. **D**: Progression of HDX over time for selected representative ArnA peptides derived from ArnA (red) and ArnA/ArnB~P (blue) samples. Data represent mean ± s.d. of n = 3 technical replicates (individual HDX reactions).

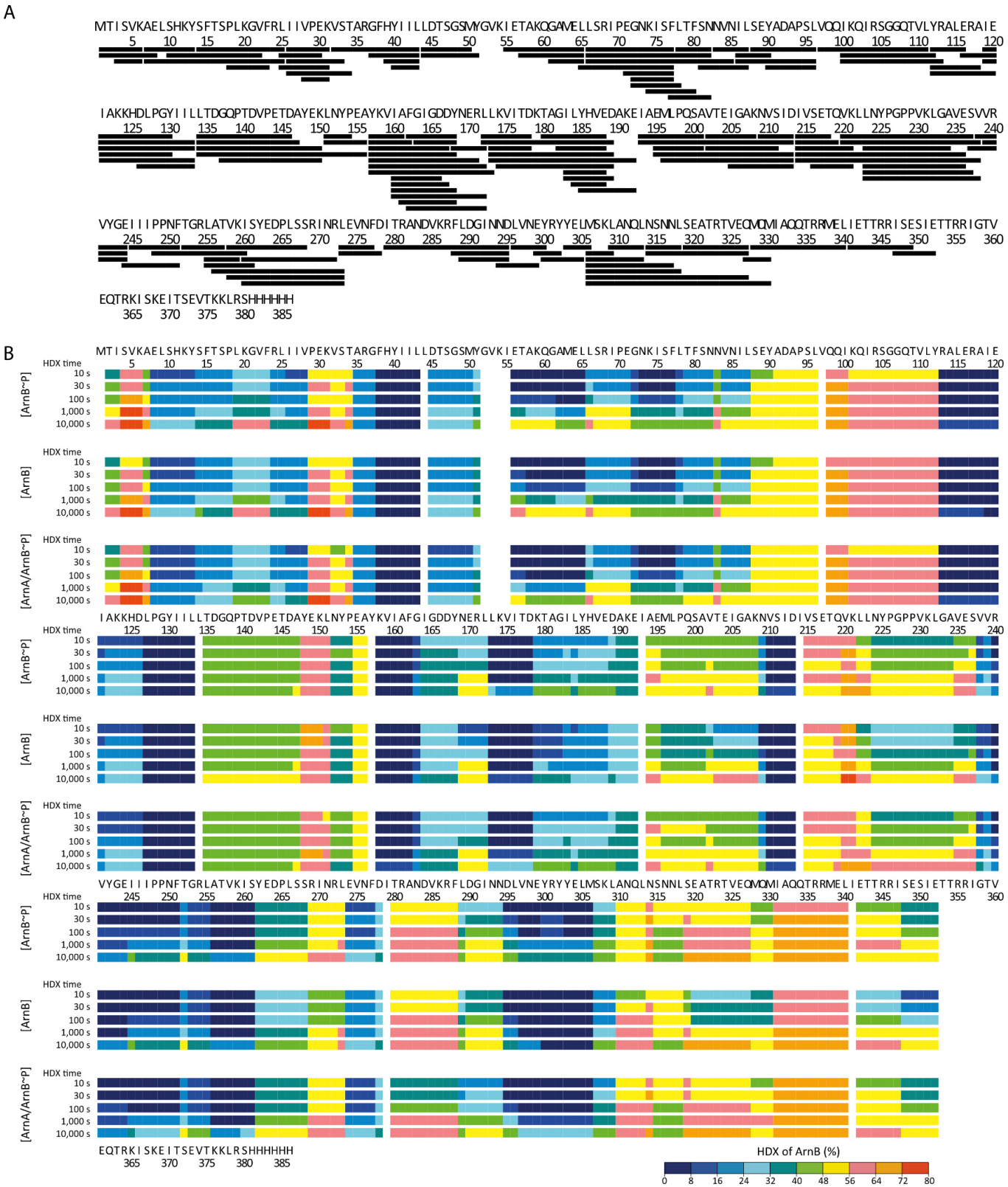

Figure S5: HDX supplementary data evaluation of ArnB. A: Each black bar represents a peptide of ArnB identified in HDX-MS. **B:** The residue-specific HDX of ArnB, ArnB~P and that of ArnB~P in complex with ArnA is projected onto the ArnB amino acid sequence.

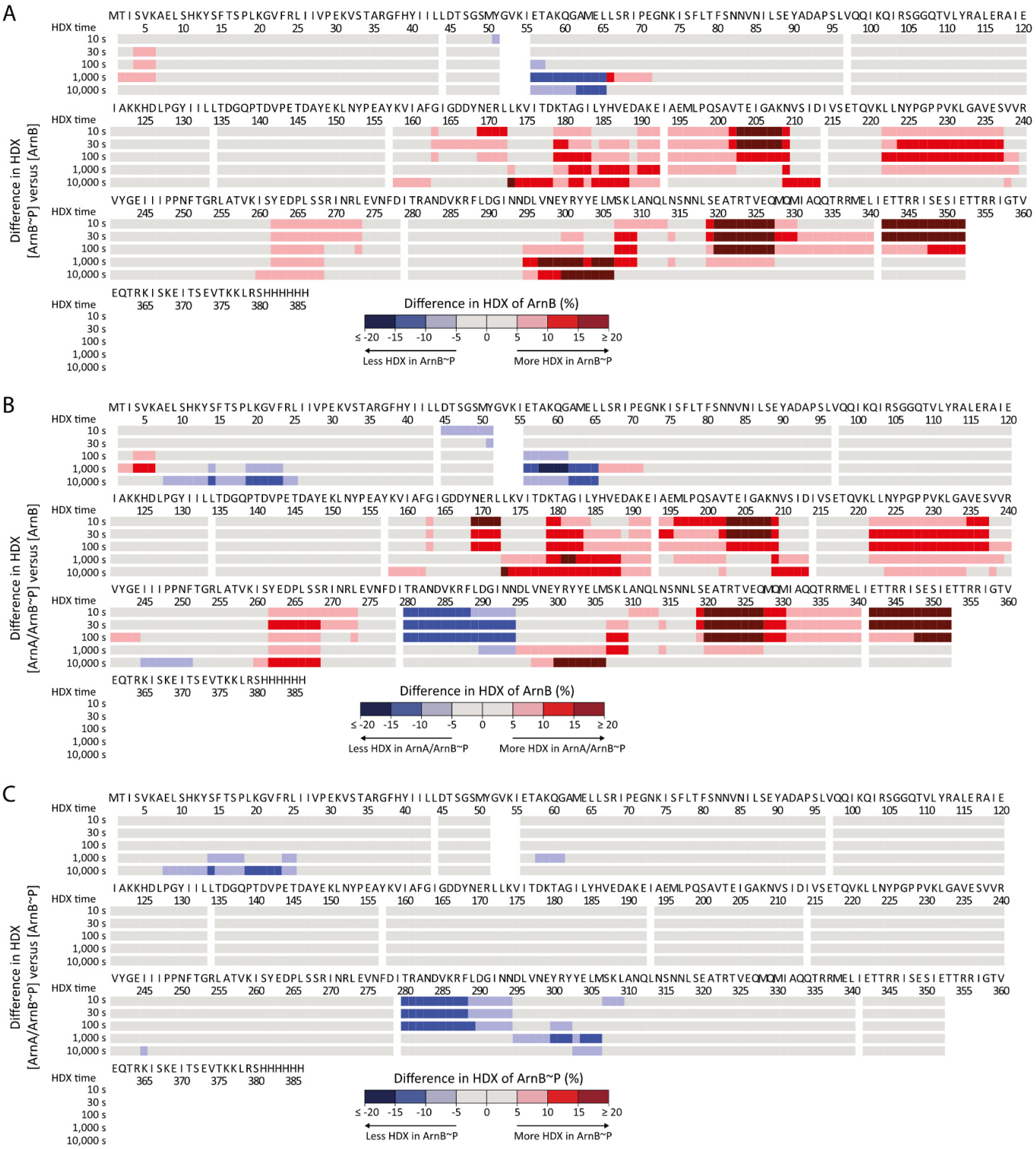

Figure S6: HDX supplementary data evaluation of the impact of phosphorylation and interaction of ArnA and ArnB. A-C: The difference in residue-specific HDX between **A:** ArnB~P and ArnB, **B:** ArnA/ArnB~P and ArnB, and **C.** ArnA/ArnB~P and ArnB~P is projected onto the ArnB amino acid sequence. Red color reflects elevated HDX of ArnB upon phosphorylation (A and B), and blue color reflects reduced HDX upon phosphorylation (A and B) and ArnA binding (B and C).

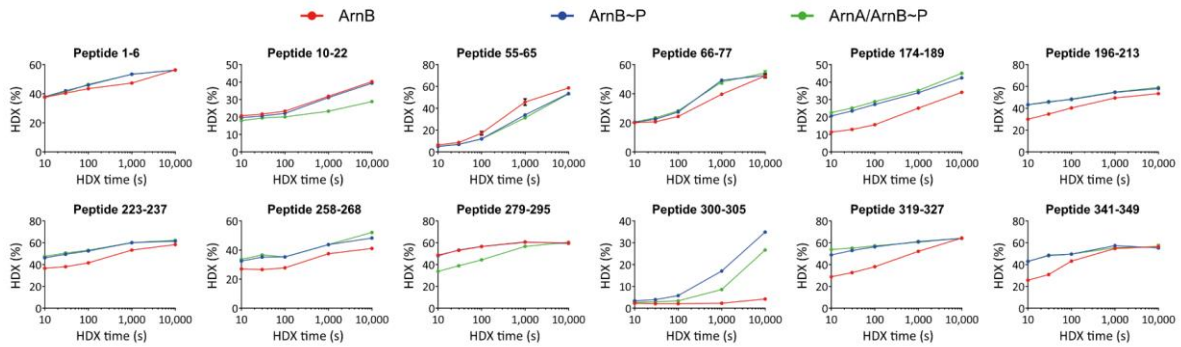

Figure S7: HDX MS of ArnA and ArnB peptides. Progression of HDX over time for selected representative ArnB peptides derived from ArnB (red), ArnB~P (blue) and ArnA/ArnB~P (green) samples. Data represent mean ± s.d. of n = 3 technical replicates (individual HDX reactions).

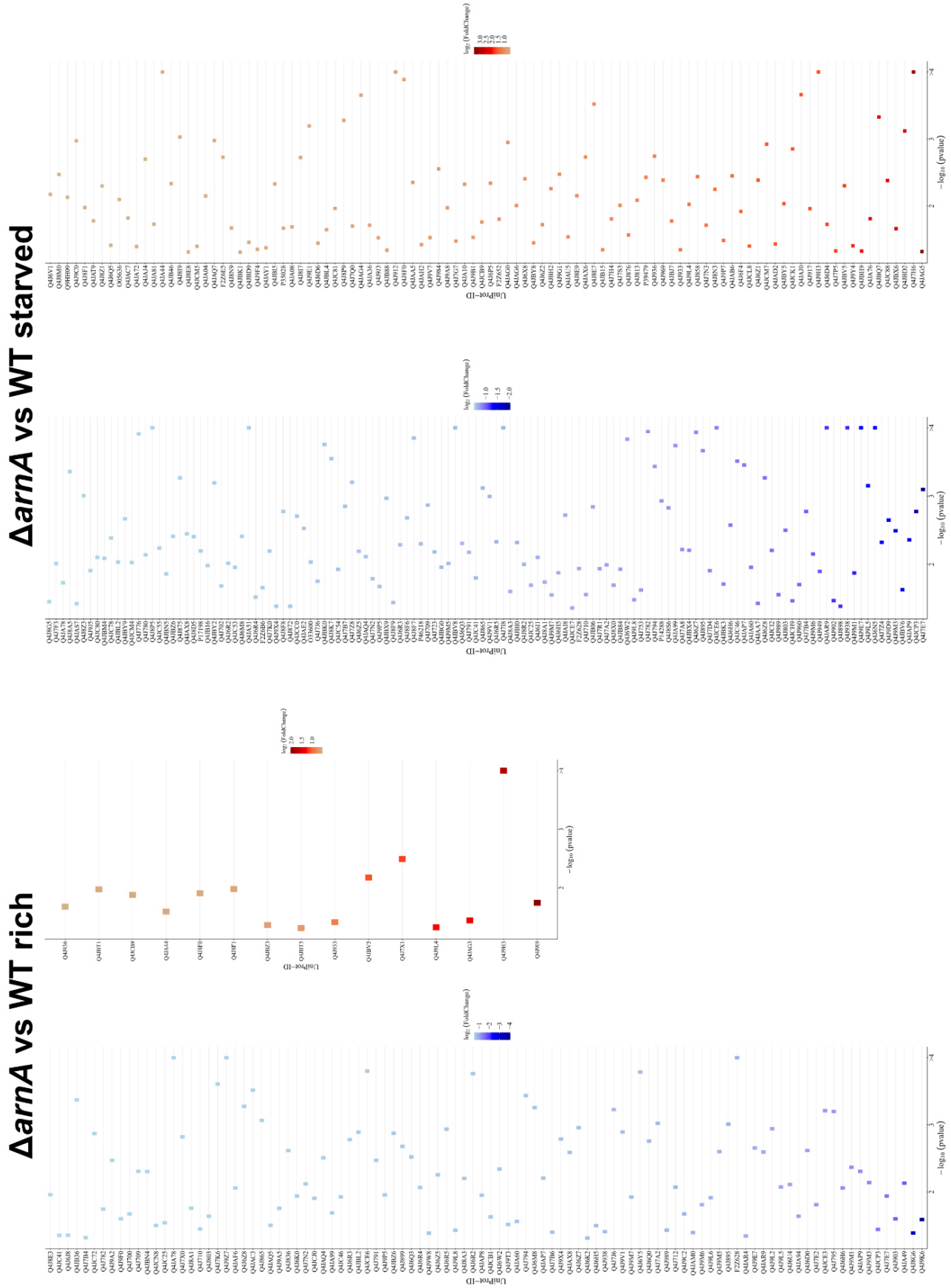

Figure S8: Volcano plot analysis of Δ*arnA* vs. WT comparison showing Uniprot IDs of over-(red) or under-regulated (blue) proteins in the respective comparison. Only proteins with a significance level of at least p<0.05 and a twofold change were considered. The analysis charts do refer to the indicated knockout strain, i.e. which proteins were up- or down-regulated in the knockout strain in comparison to WT. Higher significant levels are shown towards the right, whereas the degree of changes are ranked from the top to the bottom.

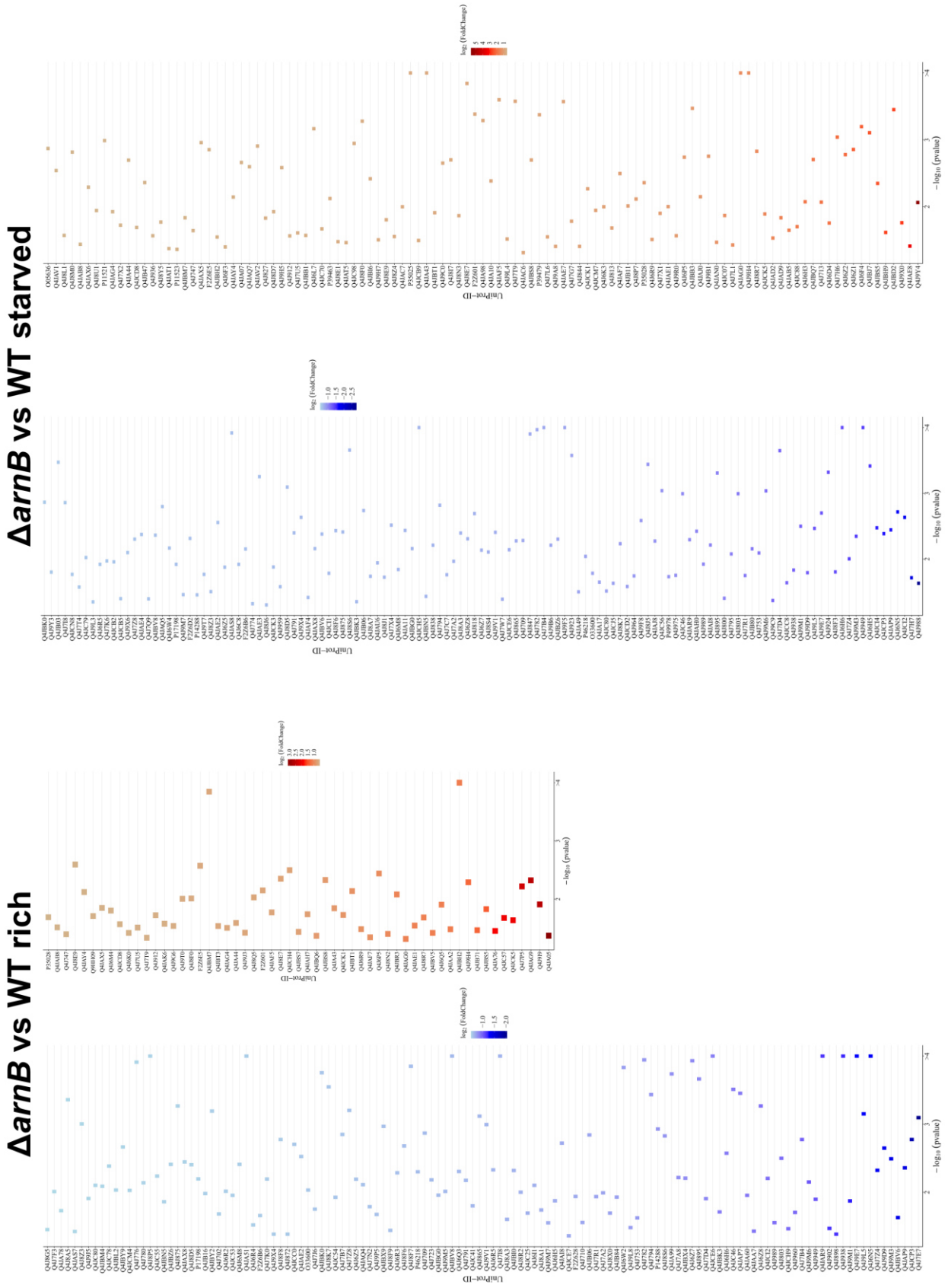

Figure S9: Volcano plot analysis of Δ*arnB* vs. WT comparison showing Uniprot IDs of over- (red) or under-regulated (blue) proteins in the respective comparison. Only proteins with a significance level of at least p<0.05 and a change of at least 50% were considered. The analysis charts do always refer to the knockout strain, i.e. which proteins were up- or down-regulated relative to the WT; significance levels increase from left to right, level changes from top to bottom. Some of the proteins with minor down-regulated changes under starved conditions were cut off for simplicity but are listed in the proteomics Table S3.

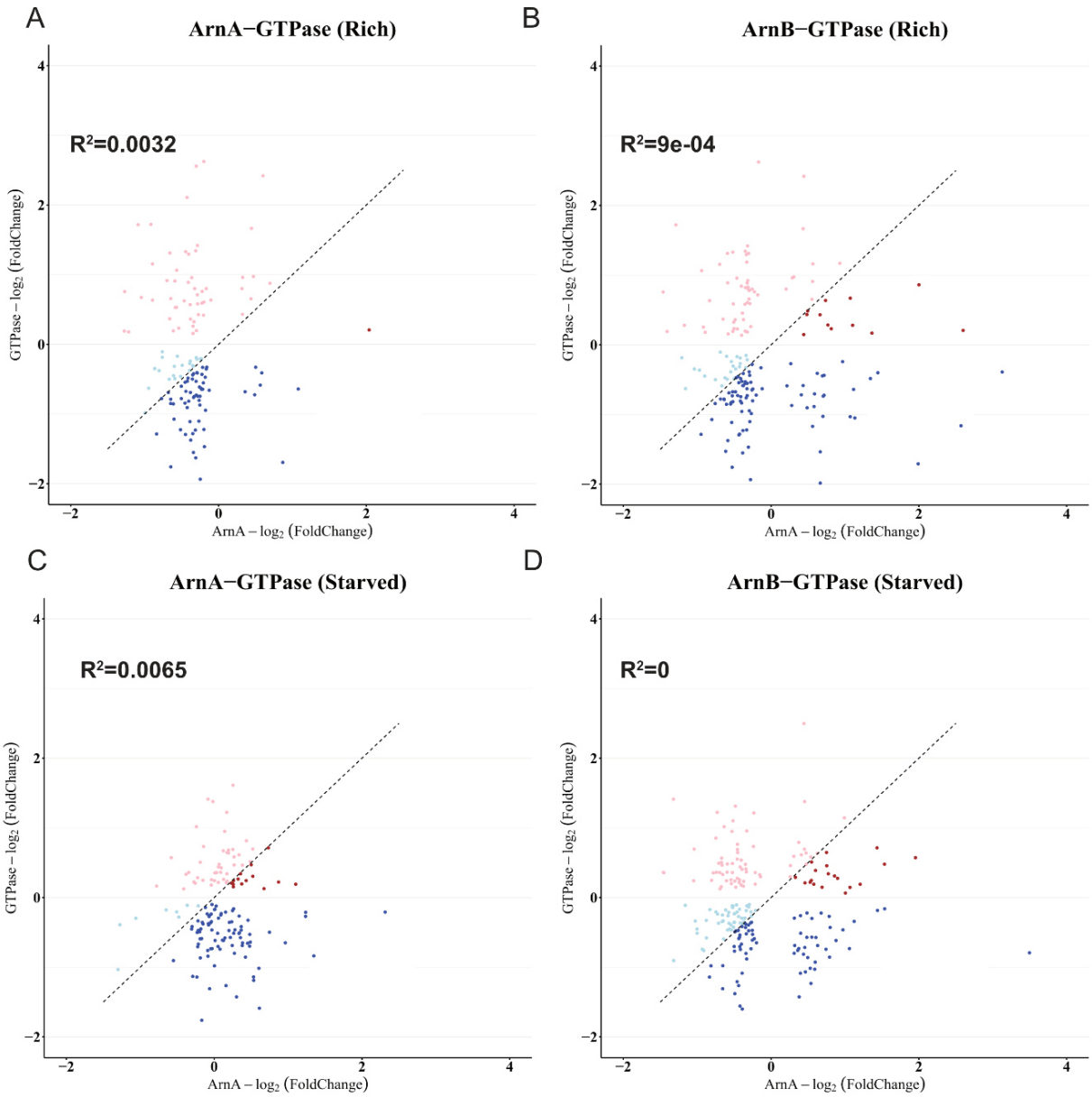

Figure S10: Proteomic changes are independent in *Sa*GPN and ArnA/ArnB knockout strains. A: Correlation plot of the *S. acidocaldarius* ArnA vs. *Sa*GPN knockout under nutrient-rich conditions, revealing very weak correlation. **B:** Correlation plot of the *S. acidocaldarius* ArnB and *Sa*GPN knockout under nutrient-rich conditions. **C:** *S. acidocaldarius* ArnA vs. *Sa*GPN knockout under nutrient starvation. **D:** *S. acidocaldarius* ArnB vs. *Sa*GPN knockout under nutrient starvation.

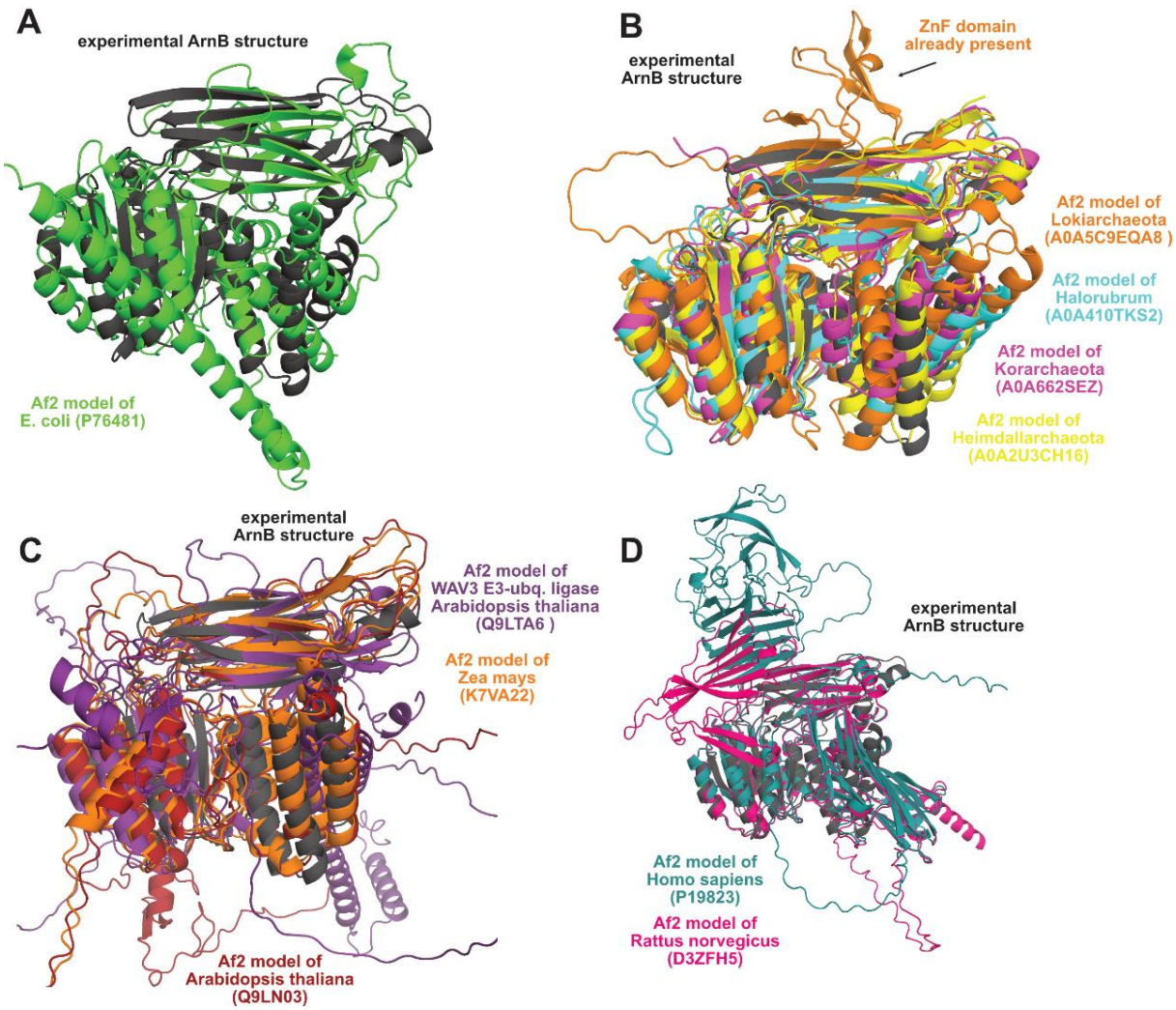

Figure S11: Structural conservation of the Sec23/Sec24 core motif and the ArnB structure. The ArnB crystal structure was subjected to foldseek search; selected hits as predicted by AlphaFold2 were superimposed to the ArnB structure that is representative of the Sec23/24 core motif. **A:** For the bacterial kingdom, the superposition of YfbK (Uniprot entry P76481) is shown. **B:** Archaeal orthologs from Lokiarchaeota, Haloarchaea, Korarchaeota and Asgard archaea. Interestingly, the member of the Lokiarchaeota is predicted to include a Sec23/24-like ZnF domain at an analogous position. **C:** Members of the plant family not belonging to the Sec23/Sec24-family show a high degree of structural conservation of the Sec23/Sec24 core motif, whereas animals hits (**D**) are predicted to be unrelated in terms of their domain packing. Uniprot identifiers are given in parentheses.

Table S1: Primers used for the side directed mutagenesis. Primers were used on the pETDuet™-1 coexpression vector encoding both *arnB-His$_6$* and *Strep-arnA*.

Primer Sequence	Introduced Mutation
CAGTAGAGCAAGCTAGAAAGATATCTAAGG	T363A, fw
CCTTAGATATCTTTCTAGCTTGCTCTACTG	T363A, rev
GAAGAATAGGTGCAGTAGAGCAAAC	T359A, fw
GTTTGCTCTACTGCACCTATTCTTC	T359A, rev
CGAAAGTATAGAGGCAGCTAGAAGAATAGG	T353A_T354A, fw
CCTATTCTTCTAGCTGCCTCTATACTTTCG	T353A_T354A, rev
GCTAATAGAGGCAGCTAGAAGAATAAGCG	T343A_T344A,fw
CGCTTATTCTTCTAGCTGCCTCTATTAGC	T343A_T344A, rev
CTAGTGAAGTTGCTAAGAAGTTGAGG	T375A, fw
CCTCAACTTCTTAGCAACTTCACTAG	T375A, rev
CTAAGGAAATAGCTAGTGAAGTTACTAAG	T371A, fw
CTTAGTAACTTCACTAGCTATTTCCTTAG	T371A, rev

Table S2: Crystallographic table for *Sa*GPN structures.

	ArnAB complex (8SO5)
Data collection and processing	
X-ray source, beamline	SLS, Beamline X06SA
Detector	DECTRIS EIGER X 16M
Wavelength (Å)	1.00000
Space group	I 1 2 1
Cell dimensions (*a*, *b*, *c*, *α*, *β*, *γ*)	50.1 Å, 147.4 Å, 62.6 Å 90.00°, 107.16°, 90.00°
Resolution (Å)	49.14 - 2.11 (2.36 - 2.11)
Total reflections	93365 (3307)
Multiplicity	3.4 (2.4)
Unique reflections	27478 (1375)
Completeness, spherical (%)†	55.1 (9.7)
Completeness, ellipsoidal (%)†	89.9 (52.3)
R_{merge}	0.079 (0.505)
CC1/2	0.996 (0.638)
Mean I/σ(I)	9.2 (1.8)
Wilson *B*-factor (Å^2)	36.73
Refinement	
Resolution (Å)	49.14 - 2.11 (2.36 - 2.11)
R_{work}, R_{free}	0.2087 (0.3488), 0.2634 (0.2591)
Reflections (test set)	27435 (1363)
r.m.s.d. from ideal:	
Bond lengths (Å)	0.009
Bond angles (°)	1.10
Total number of atoms	6473
solvent	29
ligands	4
Average *B*-factor (Å^2)	48.29
Ligands	59.11
Ramachandran favored (%)	97.51
Ramachandran allowed (%)	2.11
Ramachandran outliers (%)	0.37

Statistics for the highest-resolution shell are shown in parentheses. † Dataset was corrected by STARANISO# for anisotropic diffraction. Cut-offs used direction -0.058 a* + 0.994 b* + 0.088 c* for best diffraction (2.11 Å), 0.588 a* + 0.809 c* for worst diffraction (3.54 Å). # Tickle, I.J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., Bricogne, G. (2016). STARANISO, http://staraniso.globalphasing.org/cgi-bin/staraniso.cgi. Cambridge, United Kingdom: Global Phasing Ltd.

4. Unpublished data

Most of the important data that originates from this work can be found in sections 3.1 and 3.2. However, some interesting data is not yet published and this section therefore deals with this data.

4.1 Structure of PP2A from *Sulfolobus acidocaldarius*

The phosphatase PP2A from *S. acidocaldarius* stands as a pivotal element in the dissociation of the ArnAB complex, thereby exerting a significant influence on the functionality of both ArnA and ArnB. Consequently, it emerges as a primary focal point for structural investigation, holding the potential to enhance the understanding of the regulatory mechanisms governing ArnA and ArnB. Additionally, exploring PP2A's structure may shed light on the intricate multi-phosphorylation pathway of ArnB, potentially revealing a competitive phosphorylation-dephosphorylation mechanism at play. Although the crystallization of the mega complex involving phosphorylated ArnAB with PP2A remains elusive, successful attempts have been made to crystallize isolated PP2A. The crystal structure of PP2A was determined from a monoclinic crystal form at a resolution of 2.3 Å for the complete residue range of N2-V293. The crystal structure revealed four protomers per asymmetric symmetry unit of two biological dimers formed by chains A and C as well as chains B and D. Figure 7 summarizes the crystallographic elucidation of the *S. acidocaldarius* PP2A structure.

Notably, the structure of PP2A in *S. acidocaldarius* closely mirrors the core structure of eukaryotic PP2A, as visualized by the superposition with PP2A from *Arabidopsis thaliana* (*At*PP2A, PDB: 7OBE) in Figure 7D. Accordingly, despite the missing extension that is harbored by *At*PP2A, the phosphatase core domain is quite similar given an r.m.s.d. value of 1.15 Å for 177 Cα atoms. Interestingly, despite PP2A strong dimer character during other experiments the interface of a *Sa*PP2A dimer is only 833.4 $Å^2$, as calculated by PDBePISA web server (137). Moreover, the catalytic cavity apparently features only one not two manganese ions compared to the eukaryotic structure. Nevertheless, this arguably explained by a lack of manganese during crystallization of SaPP2A, as no manganese was supplemented during purification or crystallization, resulting in a more ambiguous electron density in the area surrounding

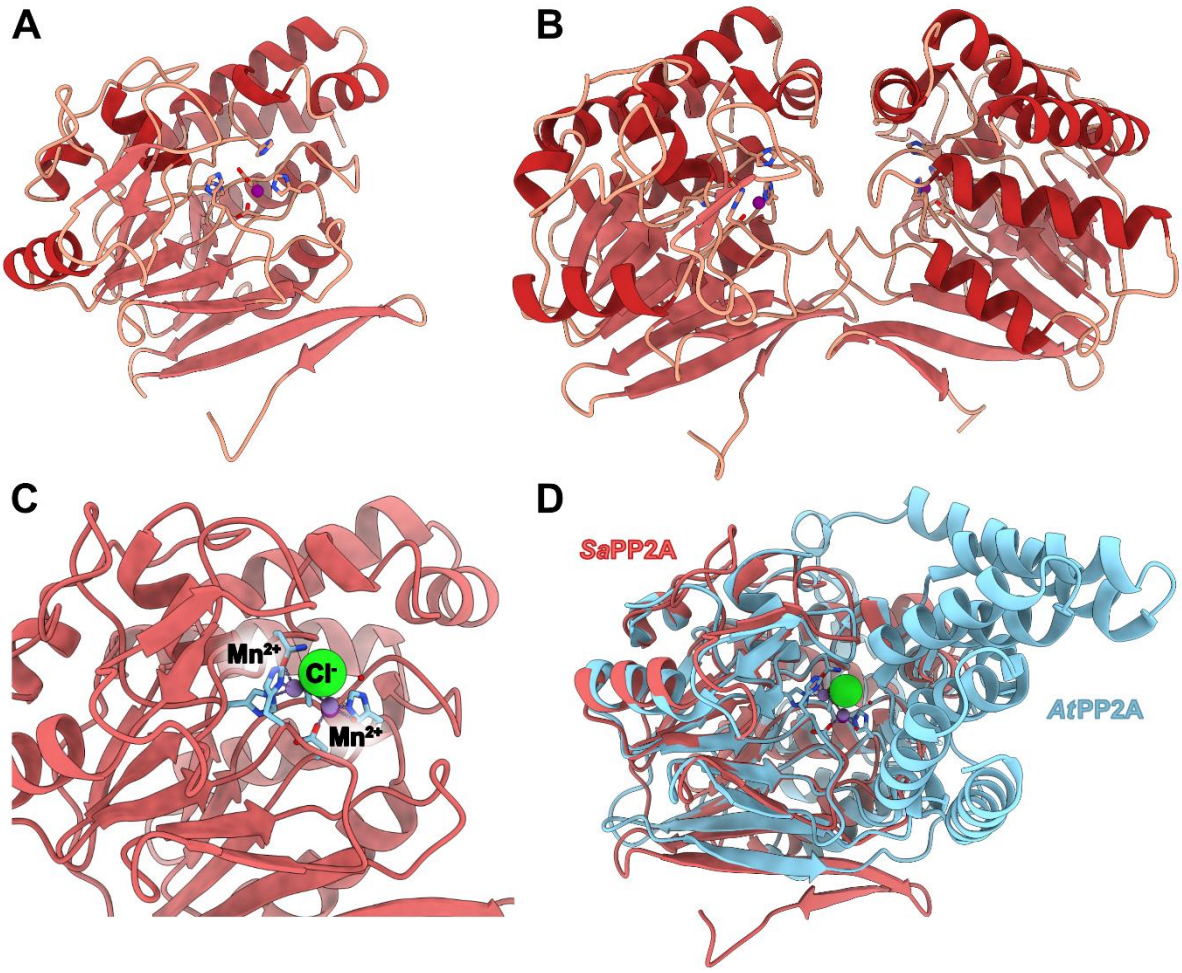

Figure 7: *S. acidocaldarius* PP2A crystal structure and comparison with PP2A from *Arabidopsis thaliana* (PDB:7OBE). A: Monomeric crystal structure with focus on the active cavity. **B:** Biological assembly of *Sa*PP2A as dimer reveals a small dimer interface. **C:** Overlay of active amino acids of *Sa*PP2A and *At*PP2A reveal high similarities in the substrate pocket despite additional Mn^{2+} and Cl^- ions. **D:** Superposition of *Sa*PP2A and *At*PP2A reveals high structural similarity of the phosphatase core domain.

the metal ions. However, the enzymatic cavity is quite similar to its eukaryotic ortholog, as all the histidine and aspartates involved in the manganese coordination are apparently conserved, both in sequence and in structural orientation. This observation suggests the conservation of enzymatic mechanics across different evolutionary stages. Furthermore, it unveils that in *S. acidocaldarius*, PP2A exists as a stand-alone domain, hinting at the possibility that additional adaptors may be necessary for it to fulfill its universal phosphatase role (138).

In conclusion, the PP2A from *S. acidocaldarius* does exhibit similarities to eukaryotic PP2A, suggesting that its enzymatic properties may be derived from its eukaryotic counterparts. Nevertheless, the interaction between phosphorylated ArnB and PP2A remains elusive, due to the absence of obvious interaction sites or adaptors. Consequently, further experimentation to unravel the intricacies of their interaction is required.

4.2 Methods

C-terminal His_6 tagged PP2A was freshly purified (refer to protein production and purification of *Korf et al. 2024*) and filtered before crystallization using a centrifuge micro filter. Crystals of PP2A grew in the presence of 0.5 M Ammonium sulfate, 0.1 M Tris pH 7.0, 30% (v/v) PEG 600, 10% (v/v) Glycerol (JCSG Core suite II E8, *NeXtal*). Diffraction data was collected at the Swiss Light Source at Paul Scherrer Institute (*Switzerland*). Phasing was done by molecular replacement with PHASER (139) and an AlphaFold2 (140) generated search model based on the native sequence of PP2A using standard settings. Afterwards, the structure was refined by multiple rounds of manual structure building with Coot (141) followed by phenix.refine (142). Refinement and data collection parameters can be found in the crystallographic

Table 1: Crystallographic table of SaPP2A data collection parameters and refinement statistics. Values in parenthesis show statistics for highest resolution shell.

Data collection and processing	
X-ray source, beamline	SLS, Beamline X06SA
Detector	DECTRIS EIGER X 16M
Wavelength (Å)	1.000031
Space group	P 1 2_1 1
Cell dimensions (*a*, *b*, *c*, *α*, *β*, *γ*)	72.35Å, 117.04Å, 87.48Å 90.00°, 110.00°, 90.00°
Resolution (Å)	47.67 - 2.4 (2.36 - 2.3)
Total reflections	232704
Multiplicity	3.9
Unique reflections	59792 (2690)
Completeness (%)	98.40 (98.03)
R_{merge}	0.098 (0.532)
CC1/2	0.995 (0.723)
I/σ(I)	9.3 (2.8)
Wilson *B*-factor (Å^2)	33.21
Refinement	
Resolution (Å)	47.67 - 2.3 (2.34 - 2.3)
R_{work}, R_{free}	0.168 (0.222), 0.213 (0.300)
Reflections (working, test set)	59792 (2690), 3109 (147)
r.m.s.d. from ideal:	
Bond lengths (Å)	0.002
Bond angles (°)	0.55
Total number of atoms	9,656
Water	499
Average *B*-factor (Å^2)	37.2
Ligands	43.6
Ramachandran favored (%)	96.08
Ramachandran allowed (%)	3.84
Ramachandran outliers (%)	0.09

5. Discussion

Biochemists all over the world commonly are taught that the shape of a protein is directly linked to its corresponding function (143) —often simplified as ‘form follows function’. While this might be a generic concept—analogous to BOHR’S electron shell or PAULING’S hybridization model—for understanding protein structures and their functional relations it has limitations (144). Prior to deep-learning approaches most of structural biology relied on comparing experimental and/or homology based model structures for deriving functions (145). In this ‘form follows function’ scenario respective analyses postulate that similar protein structures inherently share similar functions, although this is not universally true. However, applying this concept is a valid approach in many cases when trying to derive biological functions or enzymatic mechanisms from structural data (144, 146). Nevertheless, discrepancies can arise due to distinct evolutionary origins, concealed secondary functions, or incorrect annotations.

In the context of Archaea, these complexities pose significant hurdles due to a lack of reliable functional and genetic data (147). There are several cases reported for Archaea, where functional relations cannot be traced by structural similarity; one prominent example of this is arguably represented by the archaeal histones of the *Halobacteria* class (148). These histones are structurally highly related to eukaryotic histones with the latter forming octameric nucleosomes. However, archaeal histones mainly involve dimers to bind DNA and assemble into nucleosomes of variable size, i.e. filaments, also referred to as hypernucleosomes (148). Given these features it is hypothesized that the regulatory function of archaeal histones corresponds more to bacterial histones (148). Other structure motifs that involve multiple functions include the ROSSMANN fold (e.g. nucleotide cofactor binding domain (149) vs. ribonuclease activity (150)), TIM barrels (catalyzing at least five different enzyme class reactions (151, 152)), helical bundles (e.g. DNA binding (153) vs actin binding (154)) or the Greek key motif (155). In conclusion, the discussion about novel archaeal gene products should not only rely on structure-to-function inferences but also structure-independent functional data. Consequently, the following discussion of the gene products playing a claimed role in the archaellum regulatory network examined in this study, *Sa*GPN and the ArnAB complex, relies not only on straightforward, but error-prone structure-to-function delineations, but also on large-scale proteomic analyses, comparative structural bioinformatics and mechanistic studies.

5.1 What is the function of archaeal GPN-loop GTPases?

G proteins constitute a pervasive family of proteins distributed across all domains of life, showcasing diverse functions that surpasses many other protein families. This family of proteins, also called GTPases, can be broadly categorized into two main groups: the translation factors or TRAFAC group and the signal recognition particle MinD and BioD or SIMIBI group (156). In the context of GTPases, attention often centers on small Ras-like members, which include Sar1, of the TRAFAC subfamily. However, they represent only one facet in the diverse world of GTPases, highlighting their versatile nature. TRAFAC GTPases play pivotal roles in translation, signal transduction, cell motility, and intracellular transport, supporting the array of functions observed for GTPases (157). Another notable subgroup within the TRAFAC category is comprised of the G protein-coupled receptors, distinguished by their unique assembly feature of forming heterotrimers. In contrast, SIMIBI GTPases, representing an evolutionarily older type of GTPase, exhibit a distinct assembly characteristics as they are commonly activated by dimerization, setting them apart from the TRAFAC group (158).

Notably, GPN-loop GTPases stand out as a captivating subset within the SIMIBI group, as they do not undergo dimerization upon nucleotide acquisition. Instead, they maintain their dimeric quaternary structure irrespective of the nucleotide-bound state, at least within the context of Archaea (159). Although there are precedents for all-dimeric NTPases in the SIMIBI class, encompassing proteins binding guanosine phosphate nucleotides as well as adenosine phosphate nucleotides—exemplified by GET3 and NifH—these differ from GPNs in that both GET3 and NifH depend on ATP as opposed to the GPNs (158, 160, 161). Consequently, the GPN-loop GTPases present a unique type within the SIMIBI group, featuring their distinctive structure and assembly. In Eukaryotes, the GPN family encompasses three homologs with the ability to form heterodimers—a characteristic limited in most Archaea, where GPN paralogs are only observed in select species. Notably, in TACK Archaea like *Sulfolobus*, GPNs exhibit functional proximity to Eukaryotic GPN1 rather than its paralogs GPN2 and GPN3, as illustrated in the GPN sequence similarity network (SSN, Figure 1, *Korf et al. 2023*). Intriguingly, Eukaryotic GPN1 is primarily associated with a chaperone or assembly-assisting role for RNA polymerase II subunits and their nuclear transport (162, 163). This suggests potential alternative or additional functions for archaeal GPNs, particularly in the context of eukaryotic nuclear transportation.

However, in contrast to Eukaryotes, Archaea possess only one, not three, RNA polymerase complexes, intriguingly homologous to the eukaryotic RNA polymerase II complex. This suggests a potential role for *Sa*GPN and other archaeal GPNs in the assembly of archaeal RNA polymerase. The multifaceted changes observed in the proteome of the *saGPN* knockout support such a role due to the significant impact on protein levels. On the other hand, the knockout is non-lethal for *S. acidocaldarius*, unlike eukaryotic GPN knockouts where the loss of either GPN1-3 is known to be lethal (164). Moreover, no growth changes were observed in the Δ*saGPN* knockout strain compared to the wild type, strongly suggesting that *Sa*GPN is not involved in critical sustainability processes, which would include the assembly of RNA polymerase.
Nevertheless, an assembly function in general—a characteristic commonly known for GTPases, e.g. in association with ribosome and tubulin assembly (165–167)—which later evolved towards RNA polymerase assembly, could explain the apparent functional disparity. This is particularly supported by the distinct roof-helices of eukaryotic GPNs, one of the two regions exhibiting the most structural changes during nucleotide hydrolysis. Additionally, AlphaFold2 models of eukaryotic (human) GPN1, as referenced in Uniprot database entries, display low confidence values for the roof-helices and often for the skid-region as well, these being the most likely interacting regions of (archaeal) GPNs. Therefore, the acquisition of an additional RNA polymerase assembly function may be attributed to the evolution of these specific structural regions. Interestingly, a STRING analysis of *Sa*GPN (Saci_1281) gives predictions of associated functions derived from various bioinformatic and

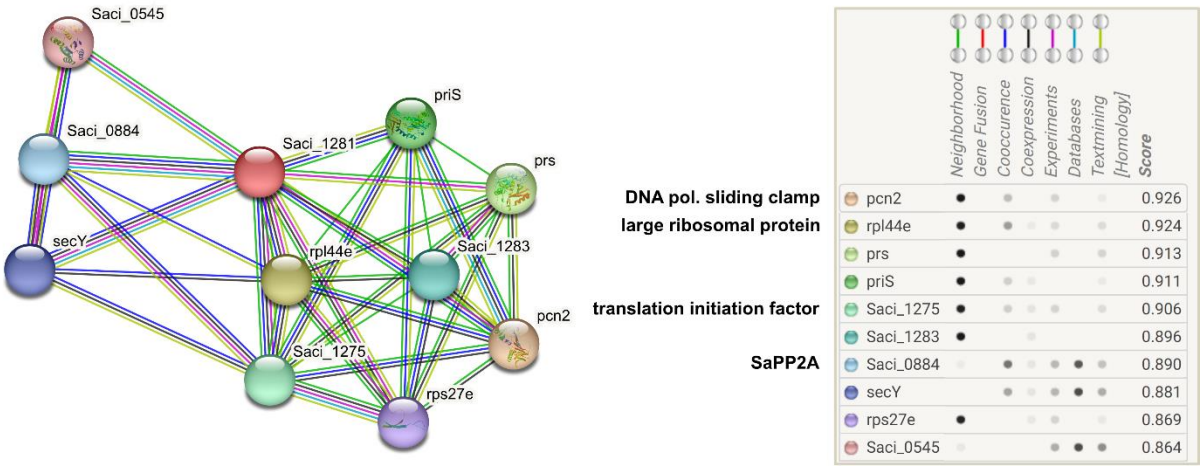

Figure 8: STRING Analysis of *Sa*GPN (Saci1281) reveals interesting associated functions. Left side shows the functional association of proteins within a network and the right side shows the corresponding legend. While this STRING analysis shows potential functional association it does not actually predict physical interaction. STRING version 11.5 with standard settings was used to generate the analysis (202).

experimental data Figure 8. In this network, the association between *Sa*GPN and *Sa*PP2A is notable, possibly linked to the identification in Ye et al. 2020. Additionally, other nodes suggest more functional partners; it's important to note that this kind of analysis indicates functional association and not necessarily physical interaction. Saci_1275, a translation initiation factor, Saci_1277, a large subunit ribosomal protein l44e, and Saci_1280, a DNA polymerase sliding clamp responsible for anchoring the catalytic subunit of DNA polymerase and other proteins to DNA during rapid replication, are particularly intriguing and rank high in scores. This analysis primarily links *Sa*GPN with nucleic acid metabolism-related proteins, akin to eukaryotic GPNs. While this is a prediction, the association of these proteins based on cooccurrence, neighborhood, coexpression, and different databases adds an interesting layer to their potential functional linkage.

In essence, it is important to highlight the limited availability of information regarding archaeal GPNs. Currently, there are only two comprehensive reports exploring this field of study—one presented above (*Korf et al. in 2023*) and the other introducing the class of GPN-loop GTPases in 2007 without assigning any biological role (159). Consequently, functional discussion has been initiated in *Korf et al. in 2023*, which are of high relevance for other archaeal phyla as well. Moreover, beyond the structural unveiling of the *Sa*GPN dimer, which aligns with its universal assembly role based on the described lock-switch-rock mechanism, the report introduces a nuanced challenge in comprehending the functionality of GPNs. Specifically, the knockout of *Sa*GPN in *S. acidocaldarius* induced a surprising large-scale alteration of protein levels under planktonic lifestyle conditions, yet it did not impact growth rates. Notably, the sessile lifestyle exhibited motility-deficient phenotypes. Nevertheless, the observation of diminished motility is also reflected by a reduction in ArlB levels in the planktonic state as described by proteomic data.

Given these finding, a pivotal question arises in the distinction between observation and correlation, raising the query of whether the observed changes are (1) direct consequences of the *Sa*GPN knockout or (2) whether there exists a concealed, indirect connection. In the event of an active effect resulting from the knockout, *Sa*GPN would instigate a plethora of proteomic alterations, being directly implicated in various processes, e.g. as a universal assembly factor. Despite the immense proteomic effects this notion apparently conflicts with a lack of being essential for cellular homeostasis. This scenario implies that the primary role of *Sa*GPN is mostly the regulation of

archaellum turnover, i.e. the loss of *Sa*GPN is directly correlated to the loss of motility. This proposition gains some support from the detection of *Sa*GPN in the PP2A co-IP assay alongside the already described archaellum regulators ArnA and ArnB (134). Conversely, considering PP2A as a potential interaction partner, it appears feasible that the extensive proteomic changes observed in the Δ*saGPN* strain may be associated with altered levels of PP2A. In this perspective, the observed changes would be of an indirect nature, more likely linked to an effect on PP2A, known to exert widespread proteomic effects (138). Nevertheless, the deletion of PP2A results in a diminished growth rate, a phenomenon not in line with the observations from the *Sa*GPN deletion strain. Furthermore, *Sa*GPN did not exhibit significant effects on the activity of PP2A, as detailed in the supplementary material of *Korf et al. 2023*, rendering this a somewhat improbable scenario.

Notably, the impact of another phosphatase, PTP (protein tyrosine phosphatase), on proteome changes was investigated (138). Intriguingly, unlike the PP2A knockout, the deletion of PTP does not influence growth rates. Additionally, the majority of affected protein levels following PTP knockout were down-regulated, aligning seamlessly with the observations from the *ΔsaGPN* strain. Consequently, the extensive proteomic changes observed in *Sa*GPN deletion strains might be linked to an indirect effect on PTP function or levels. Furthermore, unlike the PP2A knockout, the deletion of PTP did not seem to affect archaellum genes. This implies that the effects observed on diminished motility in the *ΔsaGPN* strain might be directly associated with the absence of *Sa*GPN. This hints at a complex interplay involving both indirect and direct effects of the *Sa*GPN-induced proteomic changes. However, it has to be noted that the knockout of *Sa*GPN did not result in a significant reduction of PTP levels; on the contrary, PTP levels even increased, albeit without the additional >50% change restriction (refer to Table S4 in *Korf et al. 2023*, PTP=Q4JB88). While other observed changes in protein levels, such as small ribosomal subunit protein uS11 (P39469) or DNA polymerase 1 (P95690), could potentially contribute to large-scale proteome changes, no clear correlation between proteome changes and unaffected growth rates can be drawn.

Consequently, an alternative scenario of greater relevance may be involved, especially considering the two distinct lifestyles adopted by *S. acidocaldarius*. The proteomic data, confined to the planktonic lifestyle, presents challenges, as the sessile lifestyle produces a unique phenotype that may not align entirely with the characteristics of the

liquid culture lifestyle. In this context, the extensive proteomic effects observed might be attributed to factors such as nutrient availability and lifestyle, as evidenced by noteworthy proteome changes even in the wild type under various nutritional conditions (refer to Figure 3 in *Korf et al. 2023*). In this proposed scenario, the direct effects of the *Sa*GPN knockout could be influenced or obscured by the lifestyle assumed in cultural media. For instance, the loss of motility may not impact growth rates when nutrients are abundant, yet proteomic differences can still manifest, a general concept known to be involved in proteomic analyses (168). This could imply that the observed effects related to archaellum changes are indeed associated with *Sa*GPN, considering the diminished motility observed in both lifestyles, i.e., loss of motility on solid media and reduced ArlB levels under nutrient limitation. Furthermore, *Sa*GPN, facilitated by the lock-switch-rock mechanism, is inherently equipped for an assembly function of any kind, aligning well with a potential role within the archaellum regulatory network. Given the disparities in lifestyles and the magnitude of proteomic changes, confidently extracting either a direct or indirect effect of the *Sa*GPN knockout from the available data is challenging. Nevertheless, a more universal assembly role for GPNs in Archaea gains credibility based on the comprehensive data involving the lock-switch-rock mechanism and its associated assembly function. Consequently, exploring other potential roles for GPNs in Archaea based on the LSR assembly mechanism may be a worthwhile avenue to pursue.

A universal assembly role might be one the most likely scenarios for archaeal GPNs even if the exact targets remain elusive, however, an involvement in filamentous protein assembly, which is a known concept for GTPases (167), could present an interesting lead. For example, GPN1 from *S. cerevisiae* extends its functional repertoire beyond its primary role in RNA polymerase assembly, namely an involvement in microtubules assembly and maintenance of their integrity (169). While archaeal origin of tubulins is debated it finds support in the notion that tubulins were discovered in Thaumarchaeota (170), which is in line with current assumption of eukaryogenesis involving Asgard archaea, close relatives of the TACK superphylum (24). Furthermore, a potential affiliation with tubulins, or structural proteins in a general sense, may offer insights into the intriguing phenotype observed on solid media plates for *S. acidocaldarius* following the *Sa*GPN knockout (refer to Figure 2, *Korf et al. 2023*). Notably, Bacteria lack genuine GPNs; although annotated as such, bacterial 'GPNs' share closer relations with small GTPases. This observation suggests a potential link

between the kind of motility structure and the presence of GPN-loop GTPases as eukaryotic cilia structures include microtubules. Furthermore, the evolution of bacterial and archaeal motility structures appears to have followed distinct paths, despite sharing mechanistic similarities (171), which could support a potential archaeal linage of microtubule cilia assembled by GPNs.

Interestingly, the knockout of *Sa*GPN revealed significant reduction of universal stress proteins (Usp) proteins. This might add to the functional repertoire of archaeal GPNs and not only enhances their potential roles but more likely lends further support to their assembly function. Stress proteins are acknowledged for their chaperone-like functions and their associations with motility (172, 173), this could indicate that GPNs may play a role in an assembly function linked to stress response. In this regard, the Archaellum, regulated by nutrient availability, stands out as a key target for these stress-responsive assembly functions.

Moreover, a widely recognized prokaryotic tubulin, FtsZ, renowned for its pivotal involvement in cell division, is identified in numerous archaeal species, thereby supporting the potential existence of FtsZ-like cell division mechanisms in archaea (123, 174). Contrarily, FtsZ is absent in Crenarchaeota, thereby excluding *Sulfolobales* species from participating in such a division mechanism (123). However, it was proposed that ESCRT-III orthologs mediate this function in thermophilic Crenarchaeota (123) and they have been identified in exosomes of *S. acidocaldarius* as well (112). Accordingly, it is plausible that archaeal GPNs may be intricately linked to the ESCRT-like cellular divisional pathway, potentially assuming chaperone-like or regulatory functions, particularly within *Sulfolobales* and similar organisms. This proposal may rationalize the distinctive clustering pattern discerned in the SSN among archaeal GPNs between *Pyrococcus abyssi* and *Sulfolobus*-like GPNs. Given that *P. abyssi*, classified as an Euryarchaeon, houses FtsZ and governs division through the FtsZ division machinery, it follows that GPNs from different archaeal phyla may exhibit functional disparities (175). Consequently, the multifaceted array of functions within the GPN-loop GTPases, as observed in Eukaryotes, adds further nuance to their roles in different archaeal contexts, even in the presence of only a single paralog. Nevertheless, both the ESCRT and FtsZ cell division pathways could potentially incorporate archaeal GPNs as assembly factors, facilitating the assembly of single subunits or multi-filaments. However, no noticeable changes in growth were observed in liquid media for the *Sa*GPN knockout, hinting that any potential role of *Sa*GPN in cell

division must be indirect, if indeed existent.

Moreover, eukaryotic GPNs are recognized for undertaking seemingly unrelated functions, e.g. GPN1 also contributes to ribosome biogenesis; where a C-terminal extension of GPN1 may play a role —a structural feature not found in archaeal GPNs (162, 169). Consequently, this suggests a similar potential in Archaea where GPNs, in general, may contribute to various functions, either directly or indirectly. This notion finds support in the large-scale structural changes of the lock-switch-rock mechanism describing *Sa*GPNs mode of action as well as the substantial proteome-level changes observed in *S. acidocaldarius* following the knockout of *Sa*GPN. Hence, archaeal GPNs encompass both the mechanistic toolkit and the associated functional network, positioning them as crucial and versatile regulators/assembly factors capable of contributing to diverse functional roles.

In conclusion, tracing possible archaeal GPN functions from eukaryotic GPN1 suggests an association with filamentous proteins, e.g. tubulin orthologs, leading to a possible connection with cell skeleton and motility structures. The divergence to bacterial motility, coupled with the respective absence of GPNs in this domain of life, as well as the association with Usps might corroborate the association of archaeal GPNs with motility. Moreover, the STRING analysis suggest association with

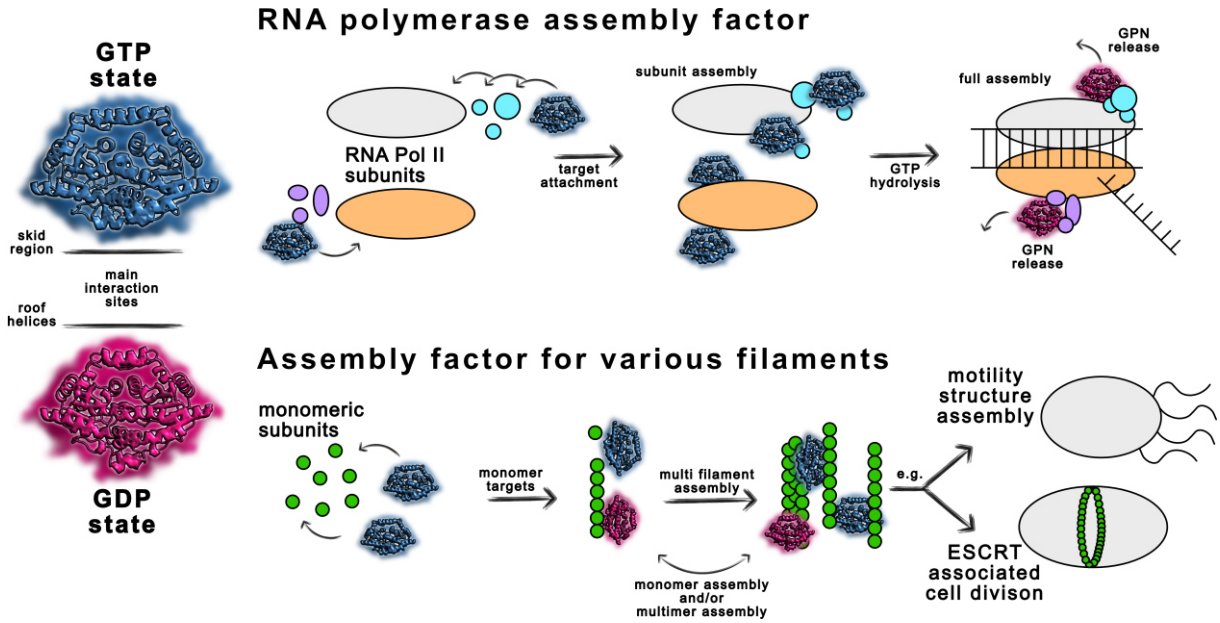

Figure 9: Graphical summary of potential GPN functions in Archaea. Various potential functions might be attributed to archaeal GPNs, all of which revolve around their assembly factor role. Given the association with the assembly of RNA polymerase II in Eukaryotes, it is probable that they assume a similar role in Archaea (top), particularly since eukaryotic RNA polymerase II bears a striking resemblance to the archaeal RNA polymerase. Moreover, the assembly role of archaeal GPN loop GTPases could potentially encompass filamentous proteins (bottom), including those associated with tubulin-like structures, the assembly of ESCRT-related cell division proteins, or motility structures like the archaellum.

ribosomal as well as translational proteins, which in turn further supports a eukaryotic-like function, although some aspects remain ambiguous due to the discrepancies observed in our studies. Nevertheless, the potential repertoire of archaeal GPNs might extends beyond assembly/integrity management of early tubulin orthologs, assembly of motility structures as well as association with ESCRT-III-mediated cell division, to encompassing the eukaryotic-like association with ribosomal proteins and assembly of the RNA polymerase complex. While for many proteins, this would typically involve an either/or scenario, the diversity of GPN functions observed in Eukaryotes, along with the data obtained during this work, suggests a potential incorporation of all these functions for archaeal GPNs.
This concludes the discussion on the potential roles of GPNs in archaea, with Figure 9 presenting a graphical summary of the discussed hypotheses for easy reference.

5.2 ArnB, the archaeal blueprint for eukaryotic Sec23/24?

ArnA and ArnB intricately involved in the regulatory network of the Archaellum, orchestrating the regulation of the main archaellin in response to nutrient starvation (135). The surprising structural resemblance to eukaryotic Sec23/24 COPII membrane curvature proteins (refer to Figure 5, *Korf et al 2024*) prompts consideration, apparently challenging the conventional notion that "form follows function"—a phrase initially employed for man-made structures by LOUIS SULLIVAN (176) but finding relevance in biology as well (177, 178). This project was initiated with the emerging aim of exploring archaeal proteins that share structural affinities with eukaryotic trafficking proteins, predating any endosymbiotic events. The findings did not only provide structural, biochemical and biophysical data to elucidate the mechanistic mode of operation but also unveil intriguing scenarios regarding why ArnB is linked to Archaellum regulation while sharing structural similarities with trafficking proteins that most likely emerged later. In the following discussion, I will delve into the two most plausible scenarios, exploring the intriguing elements associated with each.

5.2.1 Scenario I: ArnB, a function analogous to Sec23/24 proteins

Given the structural homology of ArnB to the principal membrane curvature proteins of the COPII machinery in eukaryotes, it appears reasonable that ArnB may serve as a precursor protein with a likewise function. Furthermore, the presence of ArnB and ArnA

in secreted vesicles of *S. acidocaldarius* (112) not only supports a potential role in vesicle formation but is also consistent with the inside-out theory for the formation of FECA, i.e. fulfilling the prerequisites for an endosymbiotic event. In this context, ArnB is capable to form multimeric complexes upon phosphorylation, surpassing the primary heterodimeric species and potentially indicating polymerization (134). While none of the higher multimers have been isolated and characterized beyond their multimeric status to date, this proneness to auto-assembly is crucial for a coating or vesicle formation protein. Moreover, *Sa*GPN, identified as a potential interaction partner by a co-IP assay (134), might serve as an assembly factor akin to a Sar1-Sec12 unit, responsible for recruiting protein coats. Despite *Sa*GPN not being a small Ras-like GTPase, its potential role as an assembly factor could facilitate the formation of a COPII-like coatomer in *S. acidocaldarius*. Thus, the interactions observed in the co-immunoprecipitation assay may indicate a PP2A-regulated, ArnAB-*Sa*GPN based

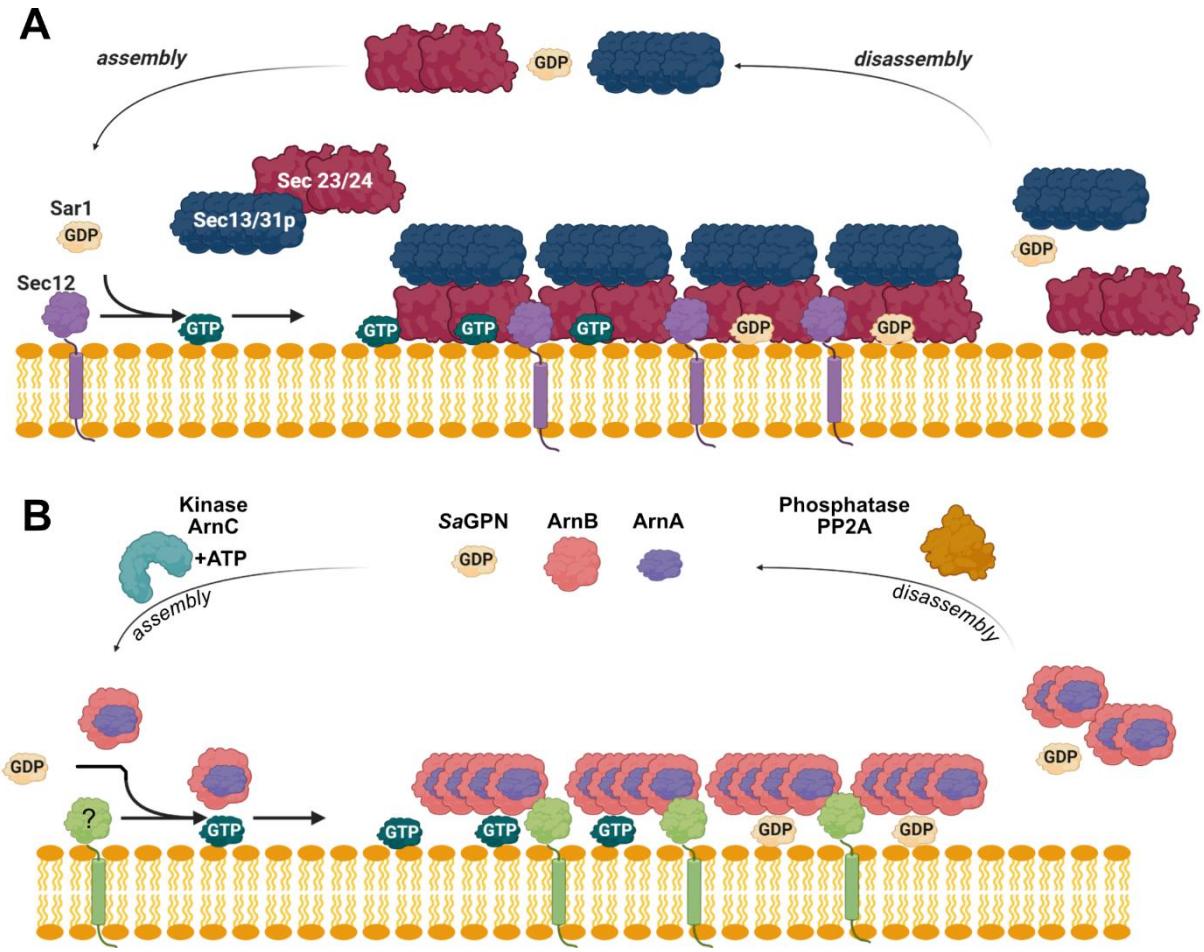

Figure 10: COPII vesicle formation process and potential adaption in *S. acidocaldarius*. A: Assembly of COPII vesicles showing Sec23/24 being recruited to the membrane by the membrane anchored Sec12 and polymerization is facilitated by the GTPase Sar1. **B:** Adoption of the COPII vesicle formation mechanism in *S. acidocaldarius* featuring the ArnAB complex and *Sa*GPN in the role of Sar1. The membrane anchor in this scenario is not clear. The phosphates PP2A and the kinase regulate complex dissociation of the subunits. However, the hydrolysis of GTP triggered by the ArnAB complex would be responsible for the coat protein dissociation from the membrane in this hypothetical scenario.

vesicle formation system. In this context, ArnB could function as a direct functional ancestor to Sec23/24. Within this hypothetical scenario, *Sa*GPN would be recruited to the membrane by an as-yet-unknown factor, stimulating the assembly of ArnAB complexes and mediating a correct polymerization pattern that results in a curved membrane and, ultimately, vesicle formation. Consequently, the ArnAB complex would act as a GTPase-activating protein (GAP), similar to Sec23, facilitating the dissociation of *Sa*GPN and promoting vesicle formation (Figure 10) (179). Notably, a significant challenge confronts this theory, besides *Sa*GPN being no Ras-like GTPase: ArnA and ArnB are intricately linked to the regulation of the Archaellum, a connection supported by numerous experimental findings.

The question arises: How could a protein fulfill entirely different functions, even if it possesses the necessary structural prerequisites of a correct fold? The phenomenon of proteins serving multiple functions is a recognized concept, often explained by various processes such as alternative RNA splicing, DNA rearrangement, or posttranslational processing, some of which are relatively sophisticated concepts for archaeal organisms (180). However, there exists another concept, prevalent in virtually all domains of life, allowing a protein to have multiple functions—gene sharing or protein moonlighting (181–183). In this paradigm, a protein performs two usually unrelated functions without requiring posttranslational modifications, gene fusion, or the use of multiple domains (180). This concept is known to apply to archaeal organisms as well (184) , suggesting that ArnA and ArnB might exemplify cases where they are not solely responsible for regulating the Archaellum based on nutrient availability but also play a role in an early form of vesicular trafficking. It's worth noting that the structure of eukaryotic Sec23/24 supports at least partial gene fusion with an ArnA-like gene due to the direct presence of the ZnF domain in the Sec23/24 fold, hence, this scenario may not qualify as a classic example of moonlighting. However, it is plausible that the multifunctionality of ArnA and ArnB could have been established prior to the occurrence of the gene fusion event. This temporal sequence could indeed support a moonlighting scenario. Consequently, ArnB might possess the capacity to fulfill both the role of an Archaellum regulator and a potential function in early COPII-like trafficking.

In summary, the proposition of an early endomembrane system, rooted in the structural homology between ArnB and Sec23/24, appears generally plausible, albeit subject to debate regarding its probability. For instance, while the *Sa*GPN GTPase might serve

as a more universal assembly factor, ascribing a Sar1-like function to it remains speculative, given the lack of evidence indicating similar functionality beyond a shared GTP hydrolysis function. Additionally, the ArnA-ArnB polymerization, reliant on a kinase sequentially phosphorylating ArnB, differs significantly from the COPII system. Moreover, the presence of ArnB homologs in all domains of life suggests either a universal role for these proteins—which is unlikely, considering they would probably have been identified as trafficking proteins by now—or differentiation throughout evolution. The latter seems more plausible based on the comprehensive bioinformatic results, although it cannot be entirely ruled out that ArnB might possess functions associated with Sec23/24.

5.2.2 Scenario II: ArnAB complex is unrelated to vesicle trafficking

Undoubtedly, ArnB emerges as a structural homolog to the Sec23/24 complex, suggesting a potential shared function. This natural conclusion aligns with a common approach often utilized when handling similar protein structures, as evidenced by numerous tools (such as homology models) employed in predicting protein functions and structures, particularly in the pre-deep learning era (185). Hence, investigation of structural similarities frequently yields functional analogies, even when the sequences of two proteins are not conserved.

However, with ArnB, the structural homology to the Sec23/24 core motif might potentially mislead in identifying functional similarities. It is not unprecedented for proteins to adopt similar structures without apparent functional relations, indicating individual evolutionary origins—a phenomenon, albeit less common, that has been reported (186, 187). Instances of proteins sharing folds with unrelated proteins of different functions abound, encompassing examples such as ferredoxin, ribosomal protein L7/L12, crambin, eglin C, *Eco*-RV endonuclease, various toxin families, and others. (186). Moreover, these proteins do not share similarities on a sequence level either, which is an additional factor that proteins tend to be of different origin (188). While proteins sharing a similar fold as well as a functional relation without sequence conservation is not a rare phenomenon (185), this aspect contributes to the disparity of ArnB and Sec23/24. Consequently, considering the established phenomenon that proteins from various origins can evolve toward similar structures (188), the structural association between ArnB and the Sec23/24 core motif might be coincidental, although unlikely given the spatial arrangement of three domains in a common core motif and

the wide array of organisms harboring this motif.

In addition, it is assumed that new functions arising through mutation are incorporated into the primary function before subsequent diversification and gene duplication occur (188, 189). Therefore, if ArnB and Sec23/24 do not share a functional relation in *S. acidocaldarius*, it could be attributed to a later evolutionary event where an ArnB-like precursor evolved towards vesicular trafficking functions. For example, in Lokiarchaeota Sec23/24-like proteins, summarized in the IPR006896 family, can be found that already share a Sec23/Sec24-trunc domain, i.e. including the ZnF domain near the β-sandwich domain (Figure S11B, *Korf et al. 2024*). Interestingly, there are also other, more ArnB-like proteins in Asgard Archaea with a length of ~400 amino acid potentially presenting a closer relative (Figure S11B, *Korf et al. 2024*).

Accordingly, the discovery that the Sec23/24 core motif is present in all domains of life, including Bacteria (refer to *Korf et al. 2024*), supports a divergent evolution scenario. Here, the ancestor of the Sec23/24 is probably to find within the Asgard Archaea phylum, whereas the TACK Archaea Sec23/24 core motif protein evolved towards ArnB/vWA2. As a result, ArnB could potentially be a sister protein to Sec23/24, lacking a direct functional connection due to an early divergence in evolutionary pathways. Moreover, it was proposed that protein structures are not as diverse as initially thought, with the basic kit of protein folds likely numbering only in the thousands (190). Hence, the Sec23/24 core motif may represent a remarkably versatile and universal fold with significant potential for functional diversity, giving notion towards an array of function in all domains of life.

However, the widespread presence of the ArnB fold does not rationalize why ArnA/B were discovered in *S. acidocaldarius* exosomes (112). Interestingly, orthologs of the Sec23/24 core motif from the plant domain are associated with the ubiquitin network, such as the WAV3 E3-ubiquitin ligase from *Arabidopsis thaliana* (Figure S11C *Korf et al. 2024*, Uniprot: Q9LTA6). The ubiquitin network is known to be involved in a myriad of cellular processes, including cargo recognition and the ESCRT system (191). Therefore, if ArnB also shares functional relations with the WAV3 E3-ubiquitin ligase, it is plausible that the ArnAB complex plays a role in general assembly functions or cargo sorting of exosomes in *S. acidocaldarius*. Notably, ESCRTIII subunits are unable to directly bind to ubiquitin (191), a prerequisite for ubiquitin ligases, and only ESCRTIII orthologs have been identified in Crenarchaeota thus far (106, 113). Nevertheless, ESCRTIII subunits prevent cargo from diffusing from the nascent vesicle (191) and are

therefore relevant for ubiquitin related cargo sorting and trafficking.

Other plausible explanations for the presence of ArnA/B in exosomes include the notion that ArnA/B may serve as cargo within exosomes or are membrane-associated to regulate Archaellum formation by enhanced localization. On the other hand, there is a possibility that ArnB exhibits functional multiplicity, such as moonlighting, as discussed earlier, although this remains more speculative in nature, despite the elevated likelihood attributed to the apparent omnipresence of the ArnB fold (183). It is worth noting that proteins associated with sugar metabolism and the tricarboxylic acid (TCA) cycle are often suggested to exert moonlighting functions (183). Interestingly, this observation aligns with the highly enriched carbohydrate, amino acid, and TCA cycle affected proteins identified in proteome knockout studies of ArnA and ArnB in *S. acidocaldarius* (refer to *Korf et al. 2024*). While this does not necessarily elucidate the presence of these proteins in exosomes, it further heightens the likelihood of their involvement in moonlighting functions.

However, the intriguing discovery that Sec23 can regulate the autophagosome in eukaryotes based on nutritional availability through a phosphorylation-dephosphorylation cycle indicates at least a potential connection on a regulatory level (179). Considering that Archaea lack an autophagy pathway, the idea might seem somewhat speculative. However, it has been discovered that Archaea harbor homologs of the autophagy pathway (192). Therefore, it is plausible that the mode of regulation has been conserved over the course of evolutionary diversification. The heteromerization of the ArnAB complex and its regulation, however, relies on a sequential phosphorylation pattern, as described in *Korf et al. 2024*. Consequently, the phosphorylation behavior of ArnB appears to differ from the phosphorylation of Sec23/24, for which sequential phosphorylation has not been reported. Yet, given the differences in habitats, primary functions, and the relatively unknown mechanisms of archaeal signal transduction, it is conceivable that the phosphorylation pattern has evolved over time.

Regardless, ArnA and ArnB serve apparently as regulators of the Archaellum as their primary function (135) and they are notably enriched in exosomes (112). It's worth noting that the structure of the Archaellum is a type IV Pili-like structure (128), whose structural analogs are also expressed in *S. acidocaldarius* as part of the UV-induced DNA damage response (120). In this context, *S. acidocaldarius* cells aggregate over these pilins, facilitating DNA repair through the transfer of DNA from one cell to

another, followed by subsequent homologous recombination (193). Accordingly, ArnA and ArnB could potentially serve as additional regulators for the UV-induced DNA damage upregulation of pilins. Two plausible reasons emerge: either the original cell aims to efficiently produce more pilins and discards ArnA/B regulators by swiftly extracting them in exosomes, or ArnA/B as cargo of these exosomes facilitate the (up)regulation of pilin formation in the exosome recipient, thereby increasing the chances of surviving UV-induced DNA damage. Here, ArnA and ArnB function as cargo within the exosomes, which appears to be feasible, given their presence in exosomes and that no direct DNA- or membrane binding site could be identified (*Korf et al. 2024*). Notably, the recent discovery of the ArnA homolog in *Saccharolobus islandicus*, *Sis*ArnA, revealed its association with the UV-induced damage response. Interestingly, the knockout of *Sis*arnA and *Sis*arnB resulted in a more DNA damage-resistant phenotype, accompanied by elevated transcription levels of respective DNA response genes (194). Overall, this implies that the ArnAB complex in *S. acidocaldarius* might be similarly involved, potentially elucidating the presence of ArnA/B in exosomes, supporting a rapid 'getting-rid' scenario to initiate pilin formation. Although direct DNA binding of ArnA and ArnB, as observed in the homologs from *S. islandicus*, has not been reported for *S. acidocaldarius*, the potential association with (UV-induced) DNA damage response seems plausible, offering a potential explanation for their occurrence in exosomes.

In conclusion, ArnB emerges as a structural homolog of Sec23/24, albeit with low sequence identity. However, beyond the structural homology, these proteins do not appear to share apparent functions. Given the ubiquity of the Sec23/24 core motif across all domains of life, a more universal role with a wide array of function seems to arise as a prominent scenario. Moreover, a divergent evolutionary pathway for the TACK/Asgard Sec23/24 ancestor would explain the structural similarities of ArnB and Sec23/24 without any functional relation. Consequently, the presence of ArnA and ArnB in *S. acidocaldarius* exosomes may be indicative of a role in UV-induced DNA damage response, as observed in other *Sulfolobales* species. Therefore, ArnB could either represent a sister version of Sec23/24 evolving differently or exemplify proteins which share only structural homology. In any case, a functional relation to Sec23/24 or a COPII-like behavior of ArnB in *S. acidocaldarius* appears less likely, especially when considering archaeal homologs of ArnA/B and their potential cargo role in respective exosomes.

5.3 ArnB, a potential chaperone-like assembly platform?

Considering that the structural homology between ArnB and Sec23/24 might be a misleading avenue for exploring potential roles and interacting factors, it may be more productive to examine what is already known about the ArnAB complex. The primary function attributed to ArnA and ArnB revolves around heterodimerization or dissociation of the complex, subsequently regulating the Archaellum based on nutrient availability. In this process, ArnB undergoes sequential phosphorylation by the kinase ArnC, facilitating a robust interaction with the FHA domain of ArnA (refer to *Korf et al. 2024*). Furthermore, this phosphorylation induces a stepwise unfolding or structure-loosening effect throughout the entire ArnB fold, involving all three main domains. Thus, ArnB represents one of the rare cases where sequential phosphorylation introduces charge repulsion into the structure, leading to subsequent (partial) unfolding—similar to the known mechanisms observed in the ankyrin repeats protein p19INK4d or somewhat akin to the Cdc37-Hsp90 complex (195, 196). Interestingly, Hsp90 and the cochaperone Cdc37 are integral components of a chaperone system that undergoes regulation not only by a PP2A family phosphatase, PP5, and multi-phosphorylation but also involves the recruitment of multiple client kinases (197). Moreover, PP5 comprises a PP2A-like Mn^{2+} dependent phosphatase domain as well as a tetratricopeptide repeat domain (TPR). Within this system, phosphorylation of Cdc37 triggers partial unfolding of its C-terminal domain, leading to further phosphorylation of Hsp90. The subsequent exposed amphipathic region of Hsp90 then serves as a docking site for protein kinase clients (196, 198). Dephosphorylation of client kinases and their dissociation from the Hsp90-Cdc37 system relies on the Ser/Thr phosphatase PP5 (197).

Notably, the kinase ArnC, responsible for the heterocomplexation of ArnA and ArnB, also features a TPR domain in addition to its kinase effector domain. Additionally, ArnB is known to associate with multiple kinases (134), and the PP2A bait co-IP assay revealed a potential interaction partner of PP2A and ArnA/B to be a universal stress protein, *Sa*UspA, proteins known to be involved in phosphorylation-dependent regulation (172, 199). The sequential phosphorylation mechanism observed in ArnB bears some resemblance to the Hsp90-Cdc37 system, particularly when considering features such as TPR-mediated docking, chaperone activities, and the association with multiple kinases—all characteristics and functions already linked to the ArnAB complex. Furthermore, Hsp90 is recognized for mediating the membrane association

of target proteins and binding to phospholipids (200, 201). Considering these aspects, it becomes conceivable that the ArnAB complex may participate in a chaperone-like system regulated by ArnC-*Sa*PP2A-mediated phosphorylation and dephosphorylation, respectively. However, in this hypothetical scenario, ArnC involves the TPR domain,

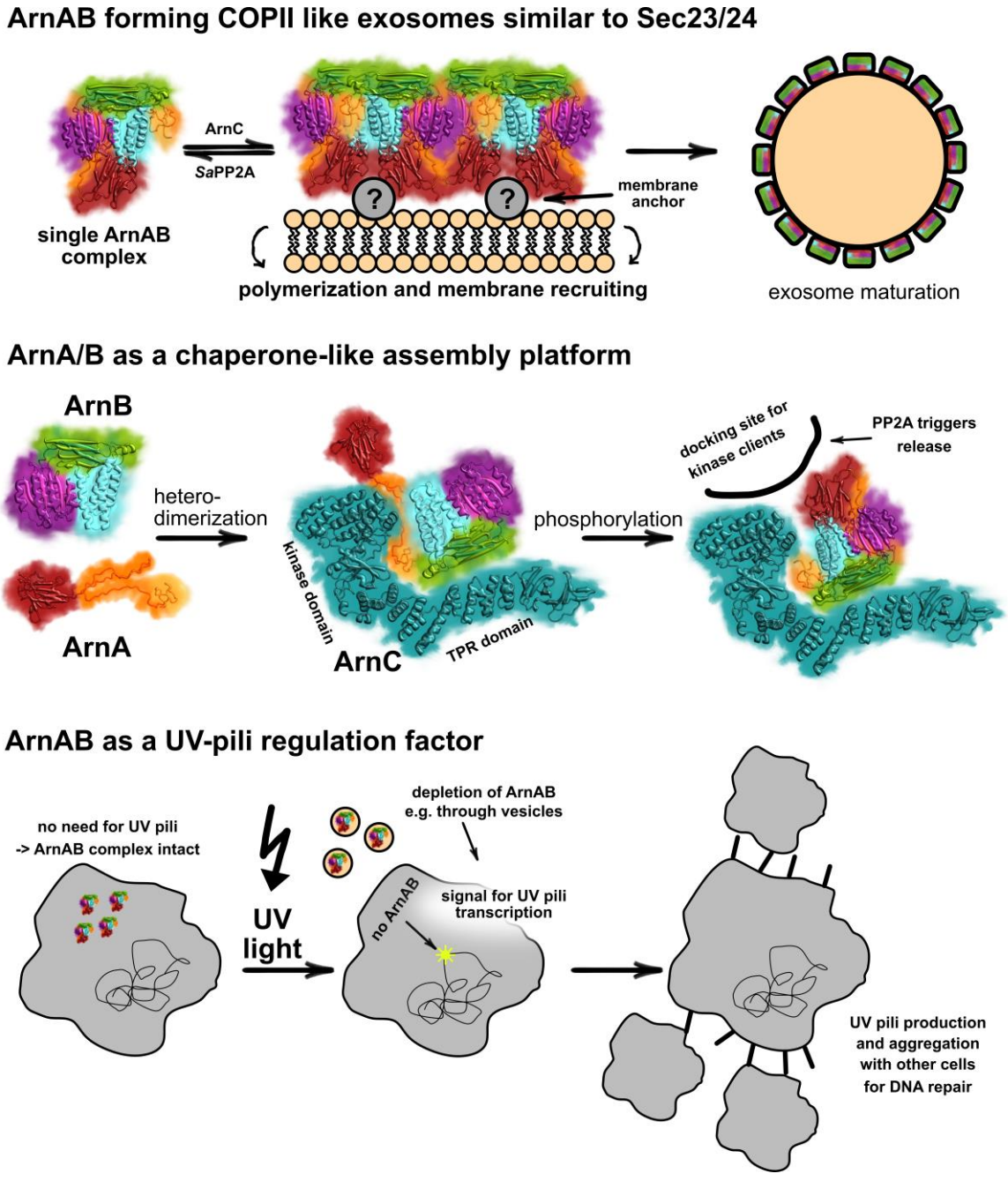

Figure 11: Graphic summary of hypothetic ArnAB functions and associations. Top: A potential function, possibly rooted in the structural homology of ArnB with Sec23/24, is the role in COPII-like exosome formation. The complete clarification of how the ArnAB complex associates with the membrane is yet to be achieved, even if *Sa*GPN might present such a target, contributing to its classification as one of the less probable scenarios. **Mid:** ArnA and ArnB may be linked in a signal transduction pathway, with ArnC potentially serving as the primary kinase that regulates other kinases. The release from the complex in this scenario would be triggered by PP2A. **Bottom:** ArnAB might play a role in the UV-induced DNA damage response, influencing the regulation of UV pili formation analogous to S. islandicus. A critical aspect of this hypothesis involves the depletion of the ArnAB complex from the cytosol via exosomes.

while *Sa*PP2A has only an effector domain, which is recruited to the complex. Additionally, the unfolding of ArnB could potentially serve as a platform analogous to the phosphorylation-dependent platform of Hsp90, facilitating the assembly of the chaperone macro complex. Furthermore, the kinase ArnD, associated with ArnB (134), may contribute to further regulation or act as a client kinase factor. Conversely, ArnB might simply function as a client protein recruited by the PP2A-ArnC system and be subservient to their regulation. Nonetheless, this system could represent a chaperone-like signal transduction system similar to Hsp90, already present in Archaea, and involved in the regulation of nutrient-dependent Archaellum production. Even considering that Hsp90 proteins are not conserved among Archaea, a regulation system similar to it might be available, especially considering the nuanced nature of protein structure evolution, as extensively discussed earlier. While this proposition remains a hypothesis, the seemingly coincidental association of numerous similar factors, determined in various independent experiments, raises questions about the likelihood of mere chance. Therefore, I propose that the observed regulation of the Archaellum, as demonstrated by ArnA and ArnB, might involve an Hsp90-like chaperone system, presenting a universal assembly factor and/or phosphorylation platform.

For a convenient summary and visualization, Figure 11 provides a graphical abstract of the predominantly discussed scenarios, arranged in order of increasing relevance from top to bottom.

6. Conclusion and Outlook

This study represents groundbreaking research with a structural focus on proteins involved in the Archaellum regulatory network, aiming to elucidate their origins and relationships with eukaryotic functionalities. The proteins ArnA, ArnB, PP2A, ArnC, *Sa*UspA, and the GPN-loop GTPase *Sa*GPN, identified through the *Sulfolobus acidocaldarius* PP2A co-IP assay, were extensively investigated using structural, biochemical, biophysical, and bioinformatic approaches.

The study delved into the intricacies of the ArnA-ArnB complex formation, revealing a sequential phosphorylation and subsequent unfolding process involving global changes. This not only advanced the mechanistic understanding of these proteins but also shed light on the general signal transduction mechanisms in Archaea. Bioinformatic analyses provided insights into the functional diversity, ubiquity, and structural intricacies of the ArnAB complex. Notably, while structural homologies between ArnB and Sec23/24 are probably unrelated towards a function in trafficking, the Sec23/24 core motif was found to be inherent in all domains of life. This clearly adds a higher importance to the Sec23/24 core motif, the observed functional diversity of the motif and the phosphorylation intricacies surrounding the ArnAB complex.

Moreover, the potential role of ArnA/B in the UV-induced DNA damage response as observed in other *Sulfolobales* species is proposed for further investigation, offering insights into archaeal-specific functions of the Sec23/24 core motif. Accordingly, further exploring ArnB-like proteins identified by bioinformatic tools during this study could unveil even more functional potential. Consequently, further research fo the sequential phosphorylation and the search for analogous archaeal proteins may illuminate archaeal signal transduction pathways. Additionally, the existence of a chaperone-like assembly platform or signal transduction pathway in Archaea is proposedly an intriguing avenue for exploration, with comprehensive structural analysis of the ArnAB complex and associated partners considered crucial.

In addition to all that, the findings of the present work significantly contributed to the understanding of GPN-loop GTPases, unraveling the novel lock-switch-rock mechanism that is crucial for comprehending all GPNs. While attempts to crystallize the 'switch' state of *Sa*GPN faced challenges, alternative methods, such as using not yet tested metal ion complexes or additives, is suggested. Interaction studies of associated proteins based on STRING analysis could serve as a starting point to delve

deeper into GPNs in the archaeal clade. The absence of GPNs in bacteria might offer interesting research regarding the link between different motility structures and the presence of GPNs, thereby potentially also enhancing understanding of eukaryogenesis based on motility structures.

In summary, this work involves structural investigations of understudied archaeal proteins, advancing knowledge of their eukaryotic counterparts and contributing to the understanding of potential eukaryogenesis pathways. Of particular note are the proposed lock-switch-rock mechanism for archaeal GPNs as well as the found omnipresence and functional diversity of the Sec23/24 core motif. Furthermore, the project yielded at least four novel protein structures, resulting in two first author publications and opens avenues for future studies to comprehensively elucidate the emergence of Eukaryotes and the origin of life.

7. Bibliography

1. H. J. Cleaves, *The Origins of Life: A Review of Scientific Inquiry* (John Templleton Foundation, 2020).

2. C. F. Chyba, P. J. Thomas, L. Brookshaw, C. Sagan, Cometary Delivery of Organic Molecules to the Early Earth. *Science* **249**, 366–373 (1990).

3. J. F. Kasting, Earth' s Early Atmosphere. *Science* **259**, 920–926 (1993).

4. N. H. Sleep, K. J. Zahnle, J. F. Kasting, H. J. Morowitz, Annihilation of ecosystems by large asteroid impacts on the early Earth. *Nature* **342**, 139–142 (1989).

5. G. B. Dalrymple, The age of the Earth in the twentieth century: A problem (mostly) solved. *Geol. Soc. Spec. Publ.* **190**, 205–221 (2001).

6. J. W. Schopf, "Earth's earliest biosphere: Its origin and evolution" (Princeton University Press, 1984).

7. S. J. Mojzsis, *et al.*, Evidence for life on Earth before 3,800 million years ago. *Nature* **384**, 55–59 (1996).

8. L. E. Orgel, The origin of life-a review of facts and speculations. *Trends Biochem. Sci.* **23**, 491–495 (1998).

9. S. L. Miller, A production of amino acids under possible primitive earth conditions. *Science* **117**, 528–529 (1953).

10. G. Horneck, D. M. Klaus, R. L. Mancinelli, Space Microbiology. *Microbiol. Mol. Biol. Rev.* **74**, 121–156 (2010).

11. A. P. Johnson, *et al.*, The Miller volcanic spark discharge experiment. *Science* **322**, 404 (2008).

12. S. D. Castle, C. S. Grierson, T. E. Gorochowski, Towards an engineering theory of evolution. *Nat. Commun.* **12**, 3326 (2021).

13. C. Darwin, *On the origin of species: By means of natural selection, or the preservation of favoured races in the struggle for life* (John Murray, 1859)

https:/doi.org/10.4324/9781912281244.

14. F. Prosdocimi, M. V. José, S. T. de Farias, “The First Universal Common Ancestor (FUCA) as the Earliest Ancestor of LUCA’s (Last UCA) Lineage” in *Evolution, Origin of Life, Concepts and Methods*, 1st Ed., P. Pontarotti, Ed. (Springer, 2019), pp. 43–54.

15. C. R. Woese, G. E. Fox, Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5088–5090 (1977).

16. C. R. Woese, O. Kandler, M. L. Wheelis, Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 4576–4579 (1990).

17. L. Guy, J. H. Saw, T. J. G. Ettema, The Archaeal Legacy of Eukaryotes: A Phylogenomic Perspective. *Cold Spring Harb. Perspect. Biol.* **6**, a016022 (2014).

18. J. A. Lake, E. Henderson, M. Oakes, M. W. Clark, Eocytes: A new ribosome structure indicates a kingdom with a close relationship to eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 3786–3790 (1984).

19. T. A. Williams, P. G. Foster, C. J. Cox, T. M. Embley, An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* **504**, 231–236 (2013).

20. K. Raymann, C. Brochier-Armanet, S. Gribaldo, The two-domain tree of life is linked to a new root for the Archaea. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6670–6675 (2015).

21. V. Da Cunha, M. Gaia, D. Gadelle, A. Nasir, P. Forterre, Lokiarchaea are close relatives of Euryarchaeota, not bridging the gap between prokaryotes and eukaryotes. *PLoS Genet.* **13**, 1–38 (2017).

22. A. Spang, *et al.*, Asgard archaea are the closest prokaryotic relatives of eukaryotes. *PLoS Genet.* **14**, e1007080 (2018).

23. V. Da Cunha, M. Gaia, A. Nasir, P. Forterre, Asgard archaea do not close the debate about the universal tree of life topology. *PLoS Genet.* **14**, e1007215 (2018).

24. A. Spang, *et al.*, Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* **521**, 173 (2015).

25. K. Zaremba-Niedzwiedzka, *et al.*, Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* **541**, 353–358 (2017).

26. L. Eme, A. Spang, J. Lombard, C. W. Stairs, T. J. G. Ettema, Archaea and the origin of eukaryotes. *Nat. Rev. Microbiol.* **15**, 711–723 (2017).

27. R. Amils, “Asgard, Archaea” in *Encyclopedia of Astrobiology*, M. Gargaud, *et al.*, Eds. (Springer Berlin Heidelberg, 2023), pp. 215–216.

28. R. Lu, *et al.*, Asgard archaea in the haima cold seep: Spatial distribution and genomic insights. *Deep Sea Res. Part I Oceanogr. Res. Pap.* **170**, 103489 (2021).

29. M. Cai, *et al.*, Diverse Asgard archaea including the novel phylum Gerdarchaeota participate in organic matter degradation. *Sci. China. Life Sci.* **63**, 886–897 (2020).

30. F. Wu, *et al.*, Unique mobile elements and scalable gene flow at the prokaryote–eukaryote boundary revealed by circularized Asgard archaea genomes. *Nat. Microbiol.* **7**, 200–212 (2022).

31. Y. Liu, *et al.*, Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature* **593**, 553–557 (2021).

32. T. Rodrigues-Oliveira, *et al.*, Actin cytoskeleton and complex cell architecture in an Asgard archaeon. *Nature* **613**, 332–339 (2022).

33. H. Imachi, *et al.*, Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature* **577**, 519–525 (2020).

34. J. Brueckner, W. F. Martin, Bacterial Genes Outnumber Archaeal Genes in Eukaryotic Genomes. *Genome Biol. Evol.* **12**, 282–292 (2020).

35. F. Husnik, J. P. McCutcheon, Functional horizontal gene transfer from bacteria to eukaryotes. *Nat. Rev. Microbiol.* **16**, 67–79 (2017).

36. M. Syvanen, Evolutionary Implications of Horizontal Gene Transfer. *Annu. Rev. Genet.* **46**, 341–358 (2012).

37. P. J. Keeling, J. D. Palmer, Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.* **9**, 605–618 (2008).

38. T. J. Treangen, E. P. C. Rocha, Horizontal transfer, not duplication, drives the expansion of protein families in prokaryotes. *PLoS Genet.* **7**, e1001284 (2011).

39. K. S. Swithers, S. M. Soucy, J. P. Gogarten, The role of reticulate evolution in creating innovation and complexity. *Int. J. Evol. Biol.* **2012**, 1–10 (2012).

40. S. M. Soucy, J. Huang, J. P. Gogarten, Horizontal gene transfer: building the web of life. *Nat. Rev. Genet.* **16**, 472–482 (2015).

41. J. M. Archibald, Endosymbiosis and Eukaryotic Cell Evolution. *Curr. Biol.* **25**, R911–R921 (2015).

42. P. C. J. Donoghue, *et al.*, Defining eukaryotes to dissect eukaryogenesis. *Curr. Biol.* **33**, R919–R929 (2023).

43. P. López-García, D. Moreira, The Syntrophy hypothesis for the origin of eukaryotes revisited. *Nat. Microbiol.* **5**, 655–667 (2020).

44. P. López-García, D. Moreira, Cultured Asgard Archaea Shed Light on Eukaryogenesis. *Cell* **181**, 232–235 (2020).

45. D. A. Baum, B. Baum, An inside-out origin for the eukaryotic cell. *BMC Biol.* **12** (2014).

46. A. A. Pittis, T. Gabaldón, Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature* **531**, 101–104 (2016).

47. T. Gabaldón, Relative timing of mitochondrial endosymbiosis and the “pre-mitochondrial symbioses” hypothesis. *IUBMB Life* **70**, 1188–1196 (2018).

48. W. Martin, M. Müller, The hydrogen hypothesis for the first eukaryote. *Nature* **392**, 37–41 (1998).

49. J. Martijn, T. J. G. Ettema, From archaeon to eukaryote: the evolutionary dark ages of the eukaryotic cell. *Biochem. Soc. Trans.* **41**, 451–457 (2013).

50. A. Spang, *et al.*, Proposal of the reverse flow model for the origin of the eukaryotic cell based on comparative analyses of Asgard archaeal metabolism.

Nat. Microbiol. **4**, 1138–1148 (2019).

51. L. Villanueva, S. Schouten, J. S. S. Damsté, Phylogenomic analysis of lipid biosynthetic genes of Archaea shed light on the 'lipid divide.' *Environ. Microbiol.* **19**, 54–69 (2017).

52. T. Řezanka, L. Kyselová, D. J. Murphy, Archaeal lipids. *Prog. Lipid Res.* **91**, 101237 (2023).

53. E. DeLong, Archaeal Means and Extremes. *Science* **280**, 542–543 (1998).

54. B. Davletov, C. Montecucco, Lipid function at synapses. *Curr. Opin. Neurobiol.* **20**, 543–549 (2010).

55. M. P. Lechevalier, C. W. Moss, Lipids in bacterial taxonomy - A taxonomist's view. *Crit. Rev. Microbiol.* **5**, 109–210 (1977).

56. S. Tan, T. Wu, D. Zhang, Z. Zhang, Cell or Cell Membrane-Based Drug Delivery Systems. *Theranostics* **5**, 863–881 (2015).

57. M. Salvador-Castell, M. Tourte, P. M. Oger, In Search for the Membrane Regulators of Archaea. *Int. J. Mol. Sci.* **20**, 4434 (2019).

58. S. Horbach, B. Neuss, H. Sahm, Effect of azasqualene on hopanoid biosynthesis and ethanol tolerance of Zymomonas mobilis. *FEMS Microbiol. Lett.* **79**, 347–350 (1991).

59. M. A. McCloskey, F. A. Troy, Paramagnetic isoprenoid carrier lipids. 2. Dispersion and dynamics in lipid membranes. *Biochemistry* **19**, 2061–2066 (1980).

60. C. F. Clarke, A. C. Rowat, J. W. Gober, Osmotic stress: Is CoQ a membrane stabilizer? *Nat. Chem. Biol.* **10**, 242–243 (2014).

61. C. Sohlenkamp, Crossing the lipid divide. *J. Biol. Chem.* **297**, 100859 (2021).

62. A. O. Chugunov, P. E. Volynsky, N. A. Krylov, I. A. Boldyrev, R. G. Efremov, Liquid but Durable: Molecular Dynamics Simulations Explain the Unique Properties of Archaeal-Like Membranes. *Sci. Rep.* **4**, 1–8 (2014).

63. C. T. Lloyd, *et al.*, Discovery, structure and mechanism of a tetraether lipid

synthase. *Nature* **609**, 197–203 (2022).

64. A. Rasouli, Y. Jamali, E. Tajkhorshid, O. Bavi, H. N. Pishkenari, Mechanical properties of ester- and ether-DPhPC bilayers: A molecular dynamics study. *J. Mech. Behav. Biomed. Mater.* **117**, 104386 (2021).

65. A. Yasmann, S. Sukharev, Properties of diphytanoyl phospholipids at the air-water interface. *Langmuir* **31**, 350–357 (2015).

66. C. G. Choquet, G. B. Patel, G. D. Sprott, T. J. Beveridge, Stability of pressure-extruded liposomes made from archaeobacterial ether lipids. *Appl. Microbiol. Biotechnol.* **42**, 375–384 (1994).

67. M. De Rosa, Archaeal lipids: structural features and supramolecular organization. *Thin Solid Films* **284**–**285**, 13–17 (1996).

68. Y. Koga, Early evolution of membrane lipids: How did the lipid divide occur? *J. Mol. Evol.* **72**, 274–282 (2011).

69. M. Mencía, The archaeal-bacterial lipid divide, could a distinct lateral proton route hold the answer? *Biol. Direct* **15**, 1–14 (2020).

70. L. Villanueva, *et al.*, Bridging the membrane lipid divide: bacteria of the FCB group superphylum have the potential to synthesize archaeal ether lipids. *ISME J.* **15**, 168–182 (2020).

71. D. X. Sahonero-Canavesi, *et al.*, Disentangling the lipid divide: Identification of key enzymes for the biosynthesis of membrane-spanning and ether lipids in Bacteria. *Sci. Adv.* **8**, eabq8652 (2022).

72. M. Exterkate, *et al.*, A promiscuous archaeal cardiolipin synthase enables construction of diverse natural and unnatural phospholipids. *J. Biol. Chem.* **296**, 100691 (2021).

73. R. Duden, ER-to-Golgi transport: COP I and COP II function. *Mol. Membr. Biol.* **20**, 197–207 (2003).

74. Y. Liu, P. Li, L. Fan, M. Wu, The nuclear transportation routes of membrane-bound transcription factors. *Cell Commun. Signal.* **16**, 1–9 (2018).

75. E. Ter Haar, A. Musacchio, S. C. Harrison, T. Kirchhausen, Atomic structure of

clathrin: a beta propeller terminal domain joins an alpha zigzag linker. *Cell* **95**, 563–573 (1998).

76. F. M. Brodsky, C. Y. Chen, C. Knuehl, M. C. Towler, D. E. Wakeharn, Biological basket weaving: formation and function of clathrin-coated vesicles. *Annu. Rev. Cell Dev. Biol.* **17**, 517–568 (2001).

77. M. S. Robinson, Forty Years of Clathrin-coated Vesicles. *Traffic* **16**, 1210–1238 (2015).

78. D. E. Wakeham, C. Y. Chen, B. Greene, P. K. Hwang, F. M. Brodsky, Clathrin self-assembly involves coordinated weak interactions favorable for cellular regulation. *EMBO J.* **22**, 4980 (2003).

79. T. Böcking, *et al.*, Key Interactions for Clathrin Coat Stability. *Structure* **22**, 819–829 (2014).

80. R. Puertollano, Clathrin-mediated transport: assembly required. *EMBO Rep.* **5**, 942–946 (2004).

81. L. M. Traub, J. S. Bonifacino, Cargo Recognition in Clathrin-Mediated Endocytosis. *Cold Spring Harb. Perspect. Biol.* **5**, a016790 (2013).

82. N. M. Willy, *et al.*, CALM supports clathrin-coated vesicle completion upon membrane tension increase. *Proc. Natl. Acad. Sci. U. S. A.* **118**, e2010438118 (2021).

83. M. G. Hanna, *et al.*, Sar1 GTPase activity is regulated by membrane curvature. *J. Biol. Chem.* **291**, 1014–1027 (2016).

84. E. Futai, S. Hamamoto, L. Orci, R. Schekman, GTP/GDP exchange by Sec12p enables COPII vesicle bud formation on synthetic liposomes. *EMBO J.* **23**, 4286–4296 (2004).

85. G. Zanetti, K. B. Pahuja, S. Studer, S. Shim, R. Schekman, COPII and the regulation of protein sorting in mammals. *Nat. Cell Biol.* **14**, 20–28 (2011).

86. J. Hutchings, V. Stancheva, E. A. Miller, G. Zanetti, Subtomogram averaging of COPII assemblies reveals how coat organization dictates membrane shape. *Nat. Commun.* **9**, 1–8 (2018).

87. B. Antonny, D. Madden, S. Hamamoto, L. Orci, R. Schekman, Dynamics of the COPII coat with GTP and stable analogues. *Nat. Cell Biol.* **3**, 531–537 (2001).

88. J. Béthune, F. T. Wieland, Assembly of COPI and COPII Vesicular Coat Proteins on Membranes. *Annu. Rev. Biophys* **47**, 63–83 (2018).

89. R. J. Taylor, G. Tagiltsev, J. A. G. Briggs, The structure of COPI vesicles and regulation of vesicle turnover. *FEBS Lett.* **597**, 819–835 (2023).

90. S. Hara-Kuge, *et al.*, En bloc incorporation of coatomer subunits during the assembly of COP-coated vesicles. *J. Cell Biol.* **124**, 883–892 (1994).

91. A. Eugster, G. Frigerio, M. Dale, R. Duden, COP I domains required for coatomer integrity, and novel interactions with ARF and ARF-GAP. *EMBO J.* **19**, 3905–3917 (2000).

92. T. Serafini, *et al.*, A coat subunit of Golgi-derived non-clathrin-coated vesicles with homology to the clathrin-coated vesicle coat protein β-adaptin. *Nature* **349**, 215–220 (1991).

93. M. Faini, R. Beck, F. T. Wieland, J. A. G. Briggs, Vesicle coats: structure, function, and general principles of assembly. *Trends Cell Biol.* **23**, 279–288 (2013).

94. R. C. Piper, D. J. Katzmann, Biogenesis and Function of Multivesicular Bodies. *Annu. Rev. Cell Dev. Biol.* **23**, 519–547 (2007).

95. W. M. Henne, N. J. Buchkovich, S. D. Emr, The ESCRT Pathway. *Dev. Cell* **21**, 77–91 (2011).

96. D. J. Katzmann, M. Babst, S. D. Emr, Ubiquitin-dependent sorting into the multivesicular body pathway requires the function of a conserved endosomal protein sorting complex, ESCRT-I. *Cell* **106**, 145–155 (2001).

97. M. Babst, D. J. Katzmann, E. J. Estepa-Sabal, T. Meerloo, S. D. Emr, ESCRT-III: An endosome-associated heterooligomeric protein complex required for MVB sorting. *Dev. Cell* **3**, 271–282 (2002).

98. M. Babst, D. J. Katzmann, W. B. Snyder, B. Wendland, S. D. Emr, Endosome-associated complex, ESCRT-II, recruits transport machinery for protein sorting

at the multivesicular body. *Dev. Cell* **3**, 283–289 (2002).

99. A. C. Camacho, D. Alfandari, E. Kozela, N. Regev-Rudzki, Biogenesis of extracellular vesicles in protozoan parasites: The ESCRT complex in the trafficking fast lane? *PLOS Pathog.* **19**, e1011140 (2023).

100. G. Prag, *et al.*, The Vps27/Hse1 Complex Is a GAT Domain-Based Scaffold for Ubiquitin-Dependent Sorting. *Dev. Cell* **12**, 973–986 (2007).

101. M. S. Kostelansky, *et al.*, Molecular architecture and functional model of the complete yeast ESCRT-I heterotetramer. *Cell* **129**, 485–498 (2007).

102. T. G. Flower, *et al.*, A helical assembly of human ESCRT-I scaffolds reverse-topology membrane scission. *Nat. Struct. Mol. Biol.* **27**, 570–580 (2020).

103. H. Teo, *et al.*, ESCRT-I core and ESCRT-II GLUE domain structures reveal role for GLUE in linking to ESCRT-I and membranes. *Cell* **125**, 99–111 (2006).

104. T. Slagsvold, *et al.*, Eap45 in mammalian ESCRT-II binds ubiquitin via a phosphoinositide-interacting GLUE domain. *J. Biol. Chem.* **280**, 19600–19606 (2005).

105. J. Schöneberg, *et al.*, ATP-dependent force generation and membrane scission by ESCRT-III and Vps4. *Science* **362**, 1423–1428 (2018).

106. J. G. Carlton, B. Baum, Roles of ESCRT-III polymers in cell division across the tree of life. *Curr. Opin. Cell Biol.* **85**, 102274 (2023).

107. B. Baum, A. Spang, On the origin of the nucleus: a hypothesis. *Microbiol. Mol. Biol. Rev.* **87**, e00186-21 (2023).

108. J. Löwe, Mysterious Asgard archaea microbes reveal their inner secrets. *Nature* **613**, 246–248 (2022).

109. K. S. Makarova, N. Yutin, S. D. Bell, E. V. Koonin, Evolution of diverse cell division and vesicle formation systems in Archaea. *Nat. Rev. Microbiol.* **8**, 731–741 (2010).

110. R. Y. Samson, T. Obita, S. M. Freund, R. L. Williams, S. D. Bell, A role for the ESCRT system in cell division in archaea. *Science* **322**, 1710–1713 (2008).

111. S. Tang, *et al.*, Structural basis for activation, assembly and membrane binding of ESCRT-III Snf7 filaments. *Elife* **4**, 1–22 (2015).

112. A. F. Ellen, *et al.*, Proteomic analysis of secreted membrane vesicles of archaeal Sulfolobus species reveals the presence of endosome sorting complex components. *Extremophiles* **13**, 67–79 (2009).

113. J. Liu, *et al.*, Archaeal extracellular vesicles are produced in an ESCRT-dependent manner and promote gene transfer and nutrient cycling in extreme environments. *ISME J.* **15**, 2892–2905 (2021).

114. T. Hatano, *et al.*, Asgard archaea shed light on the evolutionary origins of the eukaryotic ubiquitin-ESCRT machinery. *Nat. Commun.* **13**, 1–16 (2022).

115. C. M. Klinger, A. Spang, J. B. Dacks, T. J. G. Ettema, Tracing the Archaeal Origins of Eukaryotic Membrane-Trafficking System Building Blocks. *Mol. Biol. Evol.* **33**, 1528–1541 (2016).

116. M. L. Ferrelli, *et al.*, Prokaryotic membrane coat - like proteins: An update. *J. Struct. Biol.* **215**, 107987 (2023).

117. T. D. Brock, K. M. Brock, R. T. Belly, R. L. Weiss, Sulfolobus: A new genus of sulfur-oxidizing bacteria living at low pH and high temperature. *Arch. Mikrobiol.* **84**, 54–68 (1972).

118. L. Chen, *et al.*, The genome of Sulfolobus acidocaldarius, a model organism of the Crenarchaeota. *J. Bacteriol.* **187**, 4992–4999 (2005).

119. D. W. Shivvers' And, T. D. Brock, Oxidation of Elemental Sulfur by Sulfolobus ac idocaldarius. *J. Bacteriol.* **114**, 706–710 (1973).

120. M. van Wolferen, M. Ajon, A. J. M. Driessen, S. V. Albers, Molecular analysis of the UV-inducible pili operon from Sulfolobus acidocaldarius. *Microbiologyopen* **2**, 928–937 (2013).

121. A. F. Ellen, *et al.*, The Sulfolobicin Genes of Sulfolobus acidocaldariusEncode Novel Antimicrobial Proteins. *J. Bacteriol.* **193**, 4380 (2011).

122. D. Prangishvili, *et al.*, Sulfolobicins, specific proteinaceous toxins produced by strains of the extremely thermophilic archaeal genus Sulfolobus. *J. Bacteriol.*

182, 2985–2988 (2000).

123. T. Härtel, P. Schwille, ESCRT-III mediated cell division in Sulfolobus acidocaldarius – a reconstitution perspective. *Front. Microbiol.* **5**, 257 (2014).

124. L. Hoffmann, *et al.*, Structure and interactions of the archaeal motility repression module ArnA-ArnB that modulates archaellum gene expression in Sulfolobus acidocaldarius. *J. Biol. Chem.* **294**, 7460–7471 (2019).

125. A. Briegel, *et al.*, Universal architecture of bacterial chemoreceptor arrays. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 17181–17186 (2009).

126. K. F. Jarrell, S. V. Albers, The archaellum: an old motility structure with a new name. *Trends Microbiol.* **20**, 307–312 (2012).

127. S. V. Albers, K. F. Jarrell, The Archaellum: An Update on the Unique Archaeal Motility Structure. *Trends Microbiol.* **26**, 351–362 (2018).

128. S. V. Albers, K. F. Jarrell, The archaellum: How Archaea swim. *Front. Microbiol.* **6**, 23 (2015).

129. F. Altegoer, *et al.*, Structural insights into the mechanism of archaellar rotational switching. *Nat. Commun.* **13**, 1–12 (2022).

130. C. Paushali, P. Tripp, S.-V. Albers, “Expression, Purification, and Assembly of Archaellum Subcomplexes of Sulfolobus acidocaldarius” in *Protein Complex Assembly: Methods and Protocols, Methods in Molecular Biology*, A. J. Marsh, Ed. (Springer Science+Business Media, LLC, part of Springer Nature, 2018), pp. 307–314.

131. A. Banerjee, *et al.*, FlaF Is a β-Sandwich Protein that Anchors the Archaellum in the Archaeal Cell Envelope by Binding the S-Layer Protein. *Structure* **23**, 863–872 (2015).

132. D. Ortega, M. Beeby, How Did the Archaellum Get Its Rotation? *Front. Microbiol.* **12**, 803720 (2022).

133. L. F. Bischof, M. F. Haurat, S. V. Albers, Two membrane-bound transcription factors regulate expression of various type-IV-pili surface structures in Sulfolobus acidocaldarius. *PeerJ* **7**, e6459 (2019).

134. X. Ye, *et al.*, The Phosphatase PP2A Interacts With ArnA and ArnB to Regulate the Oligomeric State and the Stability of the ArnA/B Complex. *Front. Microbiol.* **11**, 1849 (2020).

135. J. Reimann, *et al.*, Regulation of archaella expression by the FHA and von Willebrand domain-containing proteins ArnA and ArnB in Sulfolobus acidocaldarius. *Mol. Microbiol.* **86**, 24–36 (2012).

136. L. Hoffmann, *et al.*, Expanding the archaellum regulatory network - the eukaryotic protein kinases ArnC and ArnD influence motility of Sulfolobus acidocaldarius. *Microbiologyopen* **6**, e00414 (2017).

137. E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).

138. J. Reimann, *et al.*, Archaeal signal transduction: Impact of protein phosphatase deletions on cell size, motility, and energy metabolism in sulfolobus acidocaldarius. *Mol. Cell. Proteomics* **12**, 3908–3923 (2013).

139. A. J. McCoy, *et al.*, Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).

140. J. Jumper, *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

141. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010).

142. P. V. Afonine, *et al.*, Towards automated crystallographic structure refinement with phenix.refine. *Acta Cryst.* **68**, 352–367 (2012).

143. J. M. Berg, J. L. Tymoczko, G. J. Gatto jr., L. Stryer, *Stryer - Biochemie*, 8th Ed. (Springer Spektrum, 2018) https:/doi.org/10.1007/978-3-662-54620-8.

144. R. D. Sleator, Proteins: Form and function. *Bioeng. Bugs* **3**, 80–85 (2012).

145. J. Yang, Y. Zhang, I-TASSER server: new development for protein structure and function predictions. *Nucleic Acids Res.* **43**, W174–W181 (2015).

146. A. Goncearenco, I. N. Berezovsky, Exploring the evolution of protein function in Archaea. *BMC Evol. Biol.* **12**, 1–14 (2012).

147. K. S. Makarova, Y. I. Wolf, E. V. Koonin, Towards functional characterization of archaeal genomic dark matter. *Biochem. Soc. Trans.* **47**, 389–398 (2019).

148. B. Henneman, C. van Emmerik, H. van Ingen, R. T. Dame, Structure and function of archaeal histones. *PLoS Genet.* **14**, e1007582 (2018).

149. I. Hanukoglu, Proteopedia: Rossmann fold: A beta-alpha-beta fold at dinucleotide binding sites. *Biochem. Mol. Biol. Educ.* **43**, 206–209 (2015).

150. A. Barbas, A. Popescu, C. Frazão, C. M. Arraiano, A. M. Fialho, Rossmann-fold motifs can confer multiple functions to metabolic enzymes: RNA binding and ribonuclease activity of a UDP-glucose dehydrogenase. *Biochem. Biophys. Res. Commun.* **430**, 218–224 (2013).

151. N. Wymer, *et al.*, Directed evolution of a new catalytic site in 2-keto-3-deoxy-6-phosphogluconate aldolase from Escherichia coli. *Structure* **9**, 1–9 (2001).

152. A. D. Goldman, J. T. Beatty, L. F. Landweber, The TIM Barrel Architecture Facilitated the Early Evolution of Protein-Mediated Metabolism. *J. Mol. Evol.* **82**, 17–26 (2016).

153. S. Banchenko, C. Weise, E. Lanka, W. Saenger, S. Geibel, Helix Bundle Domain of Primase RepB′ Is Required for Dinucleotide Formation and Extension. *ACS Omega* **6**, 28903–28911 (2021).

154. J. W. Howarth, S. Ramisetti, K. Nolan, S. Sadayappan, P. R. Rosevear, Structural Insight into Unique Cardiac Myosin-binding Protein-C Motif: A PARTIALLY FOLDED DOMAIN. *J. Biol. Chem.* **287**, 8254–8262 (2012).

155. A. V. Finkelstein, A. M. Gutun, A. Y. Badretdinov, Why are the same protein folds used to perform different functions? *FEBS Lett.* **325**, 23–28 (1993).

156. R. Gasper, S. Meyer, K. Gotthardt, M. Sirajuddin, A. Wittinghofer, It takes two to tango: regulation of G proteins by dimerization. *Nat. Rev. Mol. Cell Biol.* **10**, 423–429 (2009).

157. D. D. Leipe, Y. I. Wolf, E. V. Koonin, L. Aravind, Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol.* **317**, 41–72 (2002).

158. G. Bange, I. Sinning, SIMIBI twins in protein targeting and localization. *Nat.*

Struct. Mol. Biol. **20**, 776–780 (2013).

159. S. Gras, *et al.*, Structural insights into a new homodimeric self-activated GTPase family. *EMBO Rep.* **8**, 569–575 (2007).

160. M. M. Georgiadis, *et al.*, Crystallographic structure of the nitrogenase iron protein from Azotobacter vinelandii. *Science* **257**, 1653–1659 (1992).

161. C. J. M. Suloway, J. W. Chartron, M. Zaslaver, W. M. Clemons, Model for eukaryotic tail-anchored protein binding based on the structure of Get3. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 14849–14854 (2009).

162. D. Forget, *et al.*, The protein interaction network of the human transcription machinery reveals a role for the conserved GTPase RPAP4/GPN1 and microtubule assembly in nuclear import and biogenesis of RNA polymerase II. *Mol. Cell. Proteomics* **9**, 2827–2839 (2010).

163. C. Carré, R. Shiekhattar, Human GTPases Associate with RNA Polymerase II To Mediate Its Nuclear Import. *Mol. Cell. Biol.* **31**, 3953–3962 (2011).

164. B. Alonso, *et al.*, Eukaryotic GPN-loop GTPases paralogs use a dimeric assembly reminiscent of archeal GPN. *Cell Cycle* **12**, 463–472 (2013).

165. S. Paul, A. Audhya, Q. Cui, Molecular mechanism of GTP binding- and dimerization-induced enhancement of Sar1-mediated membrane remodeling. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2212513120 (2023).

166. P. Kaláb, A. Pralle, E. Y. Isacoff, R. Heald, K. Weis, Analysis of a RanGTP-regulated gradient in mitotic somatic cells. *Nature* **440**, 697–701 (2006).

167. L. Romberg, P. A. Levin, Assembly Dynamics of the Bacterial Cell Division Protein FtsZ: Poised at the Edge of Stability. *Annu. Rev. Microbiol.* **57**, 125–154 (2003).

168. M. Öztürk, *et al.*, Proteome effects of genome-wide single gene perturbations. *Nat. Commun.* **13**, 1–10 (2022).

169. M. Mora-García, *et al.*, Synthetic negative genome screen of the GPN-loop GTPase NPA3 in Saccharomyces cerevisiae. *Curr. Genet.* **68**, 343–360 (2022).

170. N. Yutin, E. V. Koonin, Archaeal origin of tubulin. *Biol. Direct* **7** (2012).

171. S. Khan, J. M. Scholey, Assembly, Functions and Evolution of Archaella, Flagella and Cilia. *Curr. Biol.* **28**, R278–R292 (2018).

172. L. Nachin, U. Nannmark, T. Nyström, Differential Roles of the Universal Stress Proteins of Escherichia coli in Oxidative Stress Resistance, Adhesion, and Motility. *J. Bacteriol.* **187**, 6265 (2005).

173. D. Luo, *et al.*, Universal Stress Proteins: From Gene to Function. *Int. J. Mol. Sci.* **24**, 4725 (2023).

174. S. Ithurbide, S. Gribaldo, S. V. Albers, N. Pende, Spotlight on FtsZ-based cell division in Archaea. *Trends Microbiol.* **30**, 665–678 (2022).

175. A. Blanch Jover, C. Dekker, The archaeal Cdv cell division system. *Trends Microbiol.* **31**, 601–615 (2023).

176. K. Hwang, Form Follows Function, Function Follows Form. *J. Craniofac. Surg.* **31**, 335 (2020).

177. J. Sykes, B. R. Holland, M. A. Charleston, A review of visualisations of protein fold networks and their relationship with sequence and function. *Biol. Rev.* **98**, 243–262 (2023).

178. K. Crozier-Reabe, G. R. Moran, Form Follows Function: Structural and Catalytic Variation in the Class A Flavoprotein Monooxygenases. *Int. J. Mol. Sci.* **13**, 15601–15639 (2012).

179. J. Jing, B. Wang, P. Liu, The Functional Role of SEC23 in Vesicle Transportation, Autophagy and Cancer. *Int. J. Biol. Sci.* **15**, 2419 (2019).

180. J. Piatigorsky, G. J. Wistow, Enzyme/Crystallins: Gene Sharing as an Evolutionary Strategy Minireview. *Cell Press* **57**, 197–199 (1989).

181. B. Su, Z. Qian, T. Li, Y. Zhou, A. Wong, PlantMP: a database for moonlighting plant proteins. *Database* **2019**, baz050 (2019).

182. V. J. Arvizu-Rubio, L. C. García-Carnero, H. M. Mora-Montes, Moonlighting proteins in medically relevant fungi. *PeerJ* **10**, e14001 (2022).

183. D. H. E. W. Huberts, I. J. van der Klei, Moonlighting proteins: An intriguing mode of multitasking. *Biochim. Biophys. Acta - Mol. Cell Res.* **1803**, 520–525 (2010).

184. B. Jia, G. W. Cheong, S. Zhang, Multifunctional enzymes in archaea: promiscuity and moonlight. *Extremophiles* **17**, 193–203 (2013).

185. J. Koehler Leman, *et al.*, Sequence-structure-function relationships in the microbial protein universe. *Nat. Commun.* **14**, 1–11 (2023).

186. C. A. Orengo, T. P. Flores, D. T. Jones, W. R. Taylor, J. M. Thornton, Recurring structural motifs in proteins with different functions. *Curr. Biol.* **3**, 131–139 (1993).

187. R. B. Russell, G. J. Barton, Structural Features can be Unconserved in Proteins with Similar Folds: An Analysis of Side-chain to Side-chain Contacts Secondary Structure and Accessibility. *J. Mol. Biol.* **244**, 332–350 (1994).

188. J. Catazaro, A. Caprez, D. Swanson, R. Powers, Functional Evolution of Proteins. *Proteins* **87**, 492–501 (2019).

189. A. Aharoni, *et al.*, The “evolvability” of promiscuous protein functions. *Nat. Genet.* **37**, 73–76 (2004).

190. A. Grant, D. Lee, C. Orengo, Progress towards mapping the universe of protein folds. *Genome Biol.* **5**, 107 (2004).

191. B. Korbei, Ubiquitination of the ubiquitin-binding machinery: how early ESCRT components are controlled. *Essays Biochem.* **66**, 169–177 (2022).

192. S. Zhang, Y. Hama, N. Mizushima, The evolution of autophagy proteins – Diversification in eukaryotes and potential ancestors in prokaryotes. *J. Cell Sci.* **134**, jcs233742 (2021).

193. F. Schult, *et al.*, Effect of UV irradiation on Sulfolobus acidocaldarius and involvement of the general transcription factor TFB3 in the early UV response. *Nucleic Acids Res.* **46**, 7179 (2018).

194. Z. Jiang, *et al.*, The FHA domain protein ArnA functions as a global DNA damage response repressor in the hyperthermophilic archaeon Saccharolobus islandicus. *MBio* **14**, e0094223 (2023).

195. A. Kumar, *et al.*, Phosphorylation-induced unfolding regulates p19INK4d during the human cell cycle. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 3344–3349 (2018).

196. A. B. Bachman, *et al.*, Phosphorylation induced cochaperone unfolding promotes kinase recruitment and client class-specific Hsp90 phosphorylation. *Nat. Commun.* **9**, 1–14 (2018).

197. J. Oberoi, *et al.*, HSP90-CDC37-PP5 forms a structural platform for kinase dephosphorylation. *Nat. Commun.* **13**, 1–13 (2022).

198. M. Jaime-Garza, *et al.*, Hsp90 provides a platform for kinase dephosphorylation by PP5. *Nat. Commun.* **14**, 2197 (2023).

199. X. Ye, C. van der Does, S. V. Albers, SaUspA, the Universal Stress Protein of Sulfolobus acidocaldarius Stimulates the Activity of the PP2A Phosphatase and Is Involved in Growth at High Salinity. *Front. Microbiol.* **11**, 598821 (2020).

200. M. J. E. Bijlmakers, M. Marsh, Hsp90 Is Essential for the Synthesis and Subsequent Membrane Association, But Not the Maintenance, of the Src-Kinase p56lck. *Mol. Biol. Cell* **11**, 1585–1595 (2000).

201. M. Zhang, *et al.*, Interaction of Hsp90 with phospholipid model membranes. *Biochim. Biophys. Acta - Biomembr.* **1860**, 611–616 (2018).

202. D. Szklarczyk, *et al.*, STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **47**, D607–D613 (2019).

8. Auxiliary Tools

Software	Purpose
Word	writing the manuscript
DeepL	grammatical corrections and improvements
ChatGPT v3.5	linguistic and grammatical improvements
Affinity Designer v1.2	figure design
BioRender	some figure elements
PDF24	generating print version