Tracing back variations in archaeal cell division to protein domain architectures

Béla Frohn[1,2], Tobias Härtel[1], Jürgen Cox[2,3,*] and Petra Schwille[1,*]

[1]Department of Cellular and Molecular Biophysics, Max-Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany.

[2]Computational Systems Biochemistry Research Group, Max-Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany.

[3]Department of Biological and Medical Psychology, University of Bergen, Jonas Liesvei 91, 5009 Bergen, Norway.

[*]Correspondence: cox@biochem.mpg.de and schwille@biochem.mpg.de

Summary

The Endosomal Sorting Complex Required for Transport (ESCRT) system is a multi-protein machinery that is involved in cell division of both Eukaryotes and Archaea. This spread across super-phyla suggests that a precursor ESCRT machinery existed already at an evolutionary early stage of life, making it a promising candidate for the (re)construction of a minimal cell division machinery. There are, however, only few experimental data about ESCRT machineries in Archaea, due to high technical challenges in cultivation and microscopy. Here, we analyse the proteins of archaeal ESCRT machineries bioinformatically on a domain level, to enable mechanistical comparison without such challenging experiments. First, we infer that there are at least three different cell division mechanisms utilizing ESCRT proteins in archaea, probably similar in their constriction mechanisms but different in membrane tethering. Second, we show that ESCRT proteins in Asgard archaea are highly similar to eukaryotic ESCRT proteins, strengthening the recently developed idea that all Eukaryotes descended from archaea. Third, we reconstruct a plausible evolutionary development of ESCRT machineries and suggest that a simple ESCRT-based constriction machinery existed in the last archaeal common ancestor. These findings not only give very interesting insights into the evolution of cell division in Archaea and Eukaryotes, but also reveal new research avenues by suggesting hypothesis-driven experiments for both, cell biology and bottom-up synthetic biology.

Introduction

Cell division is one of the most fundamental biological processes and typically regarded as one of the basic properties of life. It must be executed in a highly controlled manner, as errors can quickly have fatal consequences. The emergence of cell division machineries that allowed such a high degree of control was hence a major milestone in early evolution, enabling cells to regulate their reproduction. However, how such an early cell division machinery might have looked like is not yet known, and it has not yet been possible to build an *in vitro* system that can mimic early cell division by fully dividing a membrane vesicle. Thus, the important question of how cell division arose remains unsolved.

One membrane-transforming machinery which is particularly interesting to this open question is the Endosomal Sorting Complex Required for Transport (ESCRT) (1–4). It is involved in cell division of both Eukaryotes and Archaea, so it is likely that ESCRT proteins were already playing a role in cell division in their last common ancestor. Especially, it has been suggested that the mechanistic principles of ESCRT machineries in Eukaryotes and Archaea are comparable (1,2,5,6), indicating that ESCRT-like proteins in the common ancestor already formed a machinery. Comparative studies of the variations of today's ESCRT machineries hence promise to give insights into an evolutionary early mechanism of cell division.

While the eukaryotic ESCRT machinery has become a highly popular research target over the last years, as it is potentially involved in a wide range of human diseases (7–9), progress in researching the archaeal pendant is slow. One reason for this is that most archaea thrive only under extreme conditions, what makes microscopic or *in vitro* studies technically challenging. This lack of information is particularly unfortunate when one is interested in a potential ancient ESCRT machinery, because a comparative approach investigating similarities of machineries

derived from the same ancestor is the more interesting the more different descendants are included. Thus, a deeper understanding of the mechanisms of archaeal ESCRT machineries and their diversity is needed.

The ESCRT machinery in archaea was first found in *Sulfolobus acidocaldarius* (3,4), and it was originally named Cdv system (Cdv for Cell DiVision). In *S. acidocaldarius*, the machinery consists of three proteins, named CdvA, CdvB (having multiple paralogs) and CdvC (1–4). Importantly, CdvB proteins are homologs of the eukaryotic ESCRT-III proteins (also having multiple paralogs), and CdvC is a homolog of the eukaryotic Vps4 protein, which is also a part of the ESCRT machinery (1–4). CdvA, however, has no homologs in Eukaryotes. Thus, there are both similarities and differences in the composition of the machineries, fitting to the hypothesis that they derived from a common ancestral machinery but evolved differently.

After the discovery of CdvA, -B and -C in *S. acidocaldarius*, homologous proteins were identified in many other archaea. Interestingly, for which of the three proteins homologs were found was closely related to the phylogeny of Archaea. There are four archaeal super-phyla, named DPANN, Euryarchaeota, Asgard and TACK (10), and three of them contain species that possess Cdv proteins (11). In Euryarchaeota, some few species show CdvB and/or CdvC proteins encoded in their genome, but most species do not possess Cdv proteins. In Asgard archaea, only little is known, because it was just recently possible to cultivate a member of this super-phylum (5,12), but metagenome data and protein prediction indicates that genes encoding for CdvB and CdvC proteins are common within this group, and first experimental evidence is accumulating (13). In TACK archaea, to which *Sulfolobus acidocaldarius* belongs, species of all groups except Thermoproteales possess all three proteins CdvA, CdvB and CdvC. These variations once more indicate that there was once an ancient Cdv/ESCRT machinery that has been adapted differently. Importantly, it was recently suggested that all Eukaryotes originate

from the Asgard super-phylum (14), so the eukaryotic ESCRT machinery fits nicely into this concept as one highly refined version of the Asgard Cdv system. This relationship emphasizes that it is necessary to investigate archaeal Cdv machineries in addition to eukaryotic ESCRT machineries to draw conclusions about the ancestral machinery.

Surprisingly, experiments conducted with TACK archaea showed unequal results in the two main groups of this super-phylum, Crenarchaeota and Thaumarchaeota, although in both groups all three proteins CdvA, CdvB and CdvC are found (1,2). While in both groups fluorescence microscopy studies showed that initially CdvA proteins enrich at the future division site, probably binding to the membrane, the subsequent steps seem to differ. In Crenarchaeota, CdvB homologs successively join CdvA, together with CdvC (1,5,6,15–18). While the exact mechanism is unknown, it is mostly suggested that CdvB homologs form a ring-like higher order structure, tethered to the membrane by CdvA (1,15,17). Then, it is assumed that this ring constricts with the help of CdvC, leading to membrane fission and cell-division (1,5,6,16). In contrast, in the Thaumarchaeon *Nitrosopumilus maritimus*, only CdvC concentrates at the division site together with CdvA, while the distribution of CdvB is diffuse (1,19,20). Up to now, there exists no explanation for this deviation and no model of how membrane constriction may be achieved in Thaumarchaeota.

Thus, while a comparative analysis of Cdv/ESCRT machineries promises interesting insights into an evolutionary early mechanism of cell division, there are two major challenges. On the one hand, experimental studies with archaeal Cdv machineries are technically highly challenging, due to the extreme conditions archaeal organisms thrive under. On the other hand, comparing presence or absence of homologs between organisms seems insufficient to draw conclusions about differences in the emerging functions: Both Cren- and Thaumarchaeota possess proteins classified as CdvA, CdvB and CdvC, but nonetheless there seem to be

differences in their machineries. Hence, simply investigating which proteins of the Cdv/ESCRT machinery might have existed in the ancient predecessor, as done in previous studies (1,2), cannot lead to a deep mechanistical understanding.

Thus, we here analyse the proteins of archaeal Cdv machineries bioinformatically on a domain level, going beyond the pure existence or absence analysis. This is inspired by the idea that the mechanism of a protein machinery is a result of the interactions between the proteins, and that these interactions are located at specific regions of the proteins' amino acid sequences. We show that Cdv homologs in Cren- and Thaumarchaeota differ in their domain compositions and that the different interactions of the proteins caused by these different compositions have the potential to explain the variations found in experiments. Backed by this finding we then expand the domain analysis to Asgard archaea and find that Cdv machineries in this archaeal super-phylum are highly similar to the eukaryotic ESCRT machinery. Finally, by combining the domain analysis with phylogenetic analysis, we propose an evolutionary scenario of Cdv/ESCRT development, which provides insights into how an ancient machinery based on evolutionary early Cdv/ESCRT proteins might have looked like. Importantly, we suggest that this ancient machinery might have been able to form a contracting ring, making it a strong candidate for a very simple cell division machinery.

Results

Cdv systems are based on eleven domains, seven of them occurring in all organisms, four only in specific phylogenetic groups.

Based on homology to previously studied Cdv machineries, we searched publicly available genomes and proteomes of archaeal organisms for homologs of CdvA, CdvB and CdvC proteins. From the resulting list we selected 37 organisms by quality and analysed the domain

architecture of their Cdv proteins by secondary structure analysis, search against Pfam, InterPro and CDD, and regular expression search (for simplicity, we will call all regions in the amino acid sequences of proteins that are expected to execute a distinct function *domains*, although strictly speaking some are only motifs). As **Figure 1A** and **B** show, we found eleven different domains in Cdv machineries (for the exact locations see **Suppl. Table 1**). Four domains (Snf7, AAA+ATPase, MIT and Vps4_C) were found in organisms of all three super-phyla, indicating that they were already part of evolutionary early Cdv machineries. One domain (MIM2) was found in Asgard and TACK archaea, but not in Euryarchaeota, two domains (CdvA_alpha and CdvA_beta) were found in TACK archaea, but not in Euryarchaeota or Asgard archaea, and two domains (BWI and BWH) were found only in the TACK subgroup called Crenarchaeota. This suggests that by gain or loss of some domains, rather than gain or loss of full proteins, Cdv systems developed differently from their common ancestor, providing a hint that this might be an explanation for differences in their respective mechanisms. Finally, in Asgard archaea, we found two more unique domains. Interestingly, these matched domains occurring in the eukaryotic ESCRT machinery. First, consistent with the results of Lu et al. (21), we found potential MIM1 (22) domains at the C-terminus of most Asgard CdvB homologs. Second, consistent with the results of Caspi and Dekker (1), we predicted alpha-helices at the N-terminus of one group of Asgard CdvB homologs that match the ESCRT-III N-terminal ANCHR (23) motif. Thus, this simple analysis already yielded two interesting results: First, four domains are shared between all organisms, which suggests that the common ancestor of Cdv-including systems might have utilized these domains. Second, some domains occur only in specific phylogenetic groups, thus they are promising candidates to explain differences between the machineries of these groups. To analyse these differences and the machineries they might give rise to, we next investigated the domain architectures of proteins within each super-phylum in detail, starting with Euryarchaeota.

Figure 1. Differences in domain architectures of Cdv machineries in archaea. A: Cdv machinery domain compositions of selected archaea. CdvA homologs in red, CdvB homologs in green and CdvC homologs in purple. Domain compositions are represented by segmentation of circles. Background colours indicate phylogenetic groups. Khaki: DPANN, blue: Euryarchaeota, red: Asgard archaea, dark green: Thaumarchaeota, pale green: Crenarchaeota. Missing protein homologs are displayed as empty circles. Differing domain patterns matching the phylogenetic groups are clearly visible. Only in few Euryarchaeota Cdv protein homologs exist and no domain combinations occur that would allow interaction, supporting the widely accepted view that Euryarchaeota do not utilize a Cdv-including system for cell division. In contrast, all investigated Asgard archaea encode CdvB and CdvC homologs, suggesting that the system is conserved and used. Matching the missing CdvA, no BWH domains occur, whilst the domains necessary for CdvB-CdvC interaction exist. Furthermore, the two domains MIM1 and ANCHR occur, which are also found in eukaryotic ESCRT-III proteins. In Thaumarchaeota, the CdvA homolog possesses a MIM2 but no BWI domain, fitting to the missing BWH domain in CdvB homologs. Domains necessary for interaction with CdvC exist (in *N. viennensis* and *Cand. N. gargensis* no Vps4_C domain is displayed because it was not classified as this domain by the bioinformatic tools used; however, multiple alignment shows that they possess sequences with very high identity and similarity to the Vps4_C domains in *Cand. N. limnia* and *N. maritimus* at the correct position). Two potential mechanisms arising from this domain composition are displayed in **C** and **D** and described in the main text. In Crenarchaeota, CdvB proteins can further be clustered to CdvB, CdvB1/2 and CdvB3 based on their domain composition and phylogenetic analysis (**Figure 2**). Crenarchaeota possess a CdvA homolog differing from Thaumarchaeota by showing a BWI domain instead of a MIM2 domain, a change which is accompanied by the occurrence of a BWH domain in the CdvB class protein that allows interaction with BWI. The differing mechanism of Crenarchaeota that arises from this domain composition is displayed

in **E** and described in the main text (in some of the CdvB1/2 class proteins the MIM2 domain is not displayed because the sequence did not match perfectly to the regular expressions that defined the domain. However, multiple alignment showed that the sequences only differed in very small changes such as one missing amino acid and very likely do in fact constitute a MIM2 motif. Nonetheless, we sticked to the strict cut-offs defined and labelled only domains that were definitely true positives.). **B:** Table of domains and their functions derived from literature.

Abbreviations: BWI: Broken Winged-Helix Interaction Site, CdvA-alpha: alpha-helix rich CdvA domain, CdvA_beta: beta-sheet rich CdvA domain, BWH: Broken Winged Helix domain, MIM1 / -2: MIT-interacting-motif 1 / -2, Snf7: vacuolar-sorting protein SNF7 domain, Vps4_C: Vps4 C terminal oligomerisation domain, MIT: Microtubule Interacting and Trafficking molecule domain, ANCHR: N-terminal membrane binding domain in ESCRT-III proteins

Cdv protein domains in Euryarchaeota do not constitute a common functional system.

Only few Euryarchaeota showed Cdv protein encoding genes, and in half of these no conserved domain sequences could be found (blue background in **Figure 1A**). This indicates that Cdv protein genes in Euryarchaeota mutated beyond domain-recognition, either due to missing evolutionary pressure to conserve them or because they developed a new function. Furthermore, no domains allowing interaction between proteins could be identified, making a system similar to Cdv-including mechanisms in other archaea unlikely. Thus, our findings confirm the widely accepted idea that Euryarchaeota do not use a Cdv-including system for cell division and that Cdv genes found in such organisms are either relicts of a lost machinery or fragments transferred via horizontal gene transfer. However, these proteins might still play physiological roles in machineries different from Cdv/ESCRT in TACK, Asgard and Eukaryotes.

Cdv proteins of Cren- and Thaumarchaeota are composed of different domains.

In contrast to Euryarchaeota, in TACK archaea we identified specific conserved domains in nearly all Cdv proteins. The architecture of these domains showed patterns that were relatively consistent within phylogenetic orders (**Figure 1A**, green background), suggesting that mechanisms may be comparable between organisms of the same phylogenetic order. On the level of phyla, however, architecture showed distinct differences: While in all Thaumarchaeota the proteins classified as CdvA homologs were constructed of the three domains CdvA_alpha, CdvA_beta and MIM2 (or putative MIM2), in all organisms of Crenarchaeota the proteins were composed of CdvA_alpha, CdvA_beta and BWI. Furthermore, while the great majority of Crenarchaeota possessed exactly one protein with a BWH domain, classified as CdvB homolog, this domain did not occur in Thaumarchaeota at all. Importantly, BWI and BWH are known to be able to interact (15,17,18), so the joint absence or presence of the two domains indicated a mechanistical difference in the machineries of Cren- and Thaumarchaeota. To further scrutinize this finding, we next analysed which protein-protein-interactions (PPIs) may be possible based on the domain architecture in TACK archaea.

Protein domain compositions explain experimental findings indicating differences between Cren- and Thaumarchaeota.

To infer PPIs, we searched in the literature which interactions the domains we had found might allow (**Figure 1B**). In Thaumarchaeota, we found four qualitatively different potential PPIs (**Figure 1A,C,D**). First, CdvA can probably interact with CdvC based on the MIM2 domain in CdvA and the MIT domain in CdvC. Second, the same interaction can presumably take place between most of the CdvB homologs and CdvC. Third, most CdvB homologs should be able to polymerise at the Snf7 domain. Fourth, CdvC proteins can probably polymerise as well,

utilising the Vps4_C domain. Similarly, we also found four possible PPIs in Crenarchaeota (**Figure 1A,E**), but these were based on other domains and occurred between different proteins. First, the BWI domain of CdvA can likely bind to the one CdvB homolog possessing a BWH domain. Second, all CdvB homologs can probably bind to each other and polymerise utilising the Snf7 domain. Third, The CdvB homologs which possess a MIM2 domain should be able to interact with the MIT domain of CdvC. Fourth, CdvC can presumably polymerise at the Vps4_C domain. Thus, there is one major difference in the PPIs of Thaumarchaeota to the ones of Crenarchaeota: while in Thaumarchaeota the domain composition suggests that CdvA binds to CdvC, in Crenarchaeota it implies that CdvA binds to CdvB. Interestingly, this finding matches well to the differences between the two phyla found in fluorescence microscopy studies (1,5,6,15,18–20). In the Thaumarchaeon *N. maritimus*, fluorescence bands of CdvA at the division site are followed by bands of CdvC, whilst in Crenarchaeota CdvA bands are followed by CdvB bands.

Thus, these previously unexplained experimental differences fit nicely to our analysis of potential PPIs. Furthermore, it was observed that in Crenarchaeota it is indeed the CdvB homolog with the BWH domain that enriches first, followed by the other CdvB homologs (5). Again, this is what the domain architecture suggests. Reassured by this support from the literature, we next tried to derive mechanistical models based on the domain compositions for membrane constriction and cell division in TACK archaea, beginning with Crenarchaeota.

Mechanistical models of Cdv based cell division suggest two different systems in TACK archaea

Crenarchaeota organisms showed different numbers of CdvB homologs (**Figure 1A**). As we wanted to derive a common mechanism for Crenarchaeota, which might be slightly adapted by

individual organisms, we first had to make homologs comparable to find similarities within the phylum. Thus, we clustered the homologs by phylogeny (**Figure 2**) and then analysed the domain architectures within the clusters. Reassuringly, the differences in domain architecture fitted perfectly to the phylogenetic clusters (**Figure 2**). Additional gene cluster analysis showed similar architectures of Cdv genes within phylogenetic clusters (**Suppl. Fig. 2**). Therefore, we defined three different classes of CdvB homologs, which we named CdvB, CdvB1/2 and CdvB3 with respect to previous work (1). CdvB class proteins are constructed of the three domains Snf7, MIM2 and BWH, CdvB1/2 class proteins of Snf7 and MIM2, and CdvB3 class proteins of Snf7 only (**Figure 1A, E**). Because this result implies that the common ancestor of all Crenarchaeota did possess these three CdvB homologs, we based our mechanistical model on these proteins (together with CdvA and CdvC). Then, to infer a mechanistical model, we first assigned one specific function to each domain (**Figure 1B**), derived from literature, and defined one protein as the initiating one. Second, we analysed possible PPIs step-by-step and thereby deduced possible occurring higher order processes (**Figure 1C–E, Suppl. Figure 1**). Importantly, this is not a mathematical modelling, but a qualitative explanation.

Figure 2: Phylogenetic tree of CdvB homologs calculated via Bayesian phylogeny. The strict separation of the three super-groups (TACK archaea, Asgard archaea, Euryarchaeota) indicates that in their last common ancestor only one CdvB homolog existed. This means that all interactions between different CdvB homologs developed independently and are therefore not comparable. Within the Crenarchaeota group (pale green) CdvB paralogs cluster to three classes, making them comparable between organisms of this group. Each cluster unites proteins constructed of the same domains (they also have similar relative positions on their genomes). Thus, the last common ancestor of Crenarchaeota did probably possess three CdvB homologs. Similarly, two clusters can be found in Asgard archaea, which group together with the two

ESCRT-III groups Vps2/24/46 (CdvBa1) and Vps20/32/60 (CdvBa2). This indicates that CdvB paralogs in Asgard archaea are highly comparable to ESCRT-III paralogs, confirming the results of Lu et al (21).

Background colours as in Figure 1 with Eukaryotes in yellow. Transparent puzzle pieces indicate domains that are not conserved in all homologs of the respective group. Single proteins that do not match the clusters are indicated by arrows. These might be the results of horizontal gene transfer or contamination of metagenome data. A: Two Euryarchaeota CdvB proteins (*Thermoplasa acidophilum*, UniProt ID Q9HIZ5 and *Thermoplasma volcanium*, Q97BR8). B: *Thaumarchaeon Nitrosopumilus maritimus*, A9A4K8. C: *Fervidococcus fontis*, domain architecture of a CdvB protein (Snf7 and MIM2), I0A2N3. D: *Candidatus Heimdallarchaeon*, A0A523XLA6.

In Crenarchaeota (**Figure 1E**), we defined CdvA as initiating protein, because it is the only protein with a potentially membrane-targeting domain, and experiments show that neither CdvB nor CdvC can bind the membrane (1,15,17,18). Also, it has been validated that CdvA is the first protein to enrich at the division site (1,15,17,18). As we had identified only one PPI for CdvA in Crenarchaeota, BWI-BWH, we first suggest that the CdvB class homolog binds to CdvA. This protein then has two more potential PPI options, mediated by the Snf7 and MIM2 domains. Which of these options is utilised first cannot be inferred from our analyses, so we had to consult the literature. It was shown experimentally that CdvB, CdvB1/2 and probably also CdvB3 enrich at the division site before CdvC joins (1,6). Thus, we suggest that the affinity for the Snf7-Snf7 interaction is higher than for the MIM2-MIT interactions, and hence, that a ring-like structure composed of CdvB, CdvB1/2 and probably also CdvB3 is formed. Thus, while we cannot directly infer from our analyses why CdvB polymerisation takes place before

CdvC enrichment, we can explain why CdvB enriches before CdvB1/2 and CdvB3. We strongly encourage experimentalist to compare Snf7-Snf7 and MIM2-MIT affinities to test this idea.

After the ring has been assembled at the division site, composed of CdvB homologs and tethered to the membrane by CdvA, there are two domains left to interfere with: the MIM2 domain of the CdvB class homolog and the MIM2 domain of the CdvB1/2 class homolog. As experimental data suggests that CdvB class homologs enrich before CdvB1/2 class homologs (5), we asked which structural cause this could have. The interaction is reported to take place at the short motif MIM2 (1,24). Importantly, in the MIM2 motif of ESCRT-III there are four prolines which seem to be essential (24). Interestingly, in the CdvB1/2 class homologs we found consistently fewer (2 instead of 4) prolines than in the CdvB class homologs (**Suppl. Figure 3**). Thus, it is a promising idea that this lower number of prolines might cause a lower affinity of the MIM2 domain in CdvB1/2 class homologs to the MIT domain in CdvC than the MIM2 domain in CdvB class homologs. This higher affinity of CdvB class homologs would be a simple mechanistical explanation of the experimental results. Giving respect to the literature, in our hypothesised mechanism we suggest that in the next step CdvC binds preferably to the CdvB class homologs in the ring at the division site.

Based on the AAA+ ATPase and Vps4_C domains of CdvC, this binding presumably results in the disassembly of the bound CdvB class homolog from the higher-order structure (1). As more and more CdvB class homologs get disassembled, CdvC more often binds to CdvB1/2 class proteins, due to the increasing concentration, and starts disassembling them, too. Thus, the ring of homologs first becomes a ring of mostly CdvB1/2 and CdvB3 class homologs, and then gets disassembled. Once again, this is in accordance with experimental data (5). However, we cannot explain how CdvB3 class homologs can dissociate from the division site, as we did not identify

domains mediating interaction with CdvC in these homologs. Finally, how this mechanism results in membrane constriction and cell division can only be speculated, and we will suggest possible explanations in the Discussion.

In Thaumarchaeota, a classification of CdvB homologs as in Crenarchaeota was not possible. Phylogenetic and gene cluster analyses showed no clear clusters (**Figure 2, Suppl. Fig 2**), and the domain composition of most homologs was the same: Snf7 and MIM2 (**Figure 1A**). Thus, the role of different CdvB homologs in Thaumarchaeota remains unclear. Nonetheless, we were able to propose a possible mechanism based on the domain compositions (**Figure 1C,D, Suppl. Fig 1**).

Same as in Crenarchaeota, the suggested process starts with CdvA binding to the membrane and enriching at the division site. Next, CdvC binds to CdvA, utilising the MIM2-MIT interaction. Then, there are two options: either CdvA itself had already formed a ring-like polymer at the division site, which is then disassembled in the same way as CdvB in Crenarchaeota (**Figure 1C**). Alternatively, CdvC functions as a linker to CdvB, which could then result in a mechanism similar to Crenarchaeota (**Figure 1D**). However, in the second scenario it is unclear how the higher-order structure could eventually be disassembled again, as CdvC can hardly function as a linker between CdvA and CdvB, and as a disassembler of CdvB at the same time. Therefore, we strongly prefer the first scenario, which is also supported by the diffuse distribution of CdvB in fluorescence microscopy experiments (19). Driven by the good agreement of our inferred models with experimental data, we next used the same approach to investigate Asgard archaea.

Asgard archaea possess a system closely related to ESCRT.

All investigated genomes of Asgard archaea contained at least one CdvB and one CdvC homolog, indicating their conservation within the super-phylum (**Figure 1A**). Furthermore, in the genome of the only Asgard archaeon that could yet be cultivated and whose genome data is therefore trustworthy (*Candidatus Prometheoarchaeum syntrophiculm*) (12), all CdvB homologs contained either MIM1 or MIM2 domains, both of which could allow interactions with the MIT domain found in the CdvC homolog. Potential interactions between CdvBs and CdvC based on these domains in Asgard archaea were recently investigated in detail by Lu et al. (21), whose results suggest that indeed CdvC in Asgard archaea can disassemble CdvB filaments, supporting the idea of a Cdv-including mechanism comparable to other archaea in this super-phylum. Another recent experimental study further supports the idea of Cdv protein-protein interactions in Asgard archaea similar to Crenarchaeota and Eukaryotes (13).

However, to form a cell division system, this potential constriction machinery must be tethered to the membrane. But there is no CdvA homolog in Asgard archaea to achieve this, so we had to search for different ways of membrane tethering. As in our model every function of a protein is based on domains, for membrane binding there were two options: either some of the CdvB homologs in Asgard archaea possess domains that can bind to the membrane themselves (similar to ESCRT-III), or they possess a domain that can interact with another protein executing this task (similar to Crenarchaeota). Caspi and Dekker (1), who at that time had only metagenome data of a single Asgard archaeon (*Lokiarchaeon sp. GC14_75*) available, hypothesised that the N-terminal alpha helix similar to the ESCRT-III ANCHR motif (23) found in one Lokiarchaeon CdvB homolog might allow direct interaction with the membrane. Our analysis of multiple Asgard archaea genomes and especially the trustworthy *Cand. P. synthrophicum* genome indeed revealed that this region is highly conserved within Asgard

archaea, but only in the CdvBa1 subgroup of Asgard CdvB homologs (**Figure 3A**). Interestingly, in our phylogenetic tree this subgroup clusters with the eukaryotic Vps2/24/46 group of ESCRT-III proteins (**Figure 2**), which are the ESCRT-III versions that can bind to the membrane via the ANCHR motif (23). Thus, it is plausible to assume that the membrane binding of the Vps2/24/46 ESCRT-III group is similar to the CdvBa1 group of Asgard archaea. Supporting this idea, secondary structure prediction suggested that the region forms an alpha helix, while in the CdvBa2 subgroup a conserved proline prohibits this (**Figure 3**). This finding strongly supports the idea of Caspi and Dekker (1) and might point to a solution to the question of membrane tethering in Asgard archaea. Furthermore, the phylogenetic relationship and structural similarity places the Cdv mechanism of Asgard archaea much closer to ESCRT in Eukaryotes than to any other archaeal Cdv-including mechanism.

Figure 3: Cdv system domain architecture in Asgard archaea. A: Multiple alignment and secondary structure prediction of Asgard archeaea CdvB homologs. At the N-terminus, short alpha-helices are predicted in the CdvBa1 homologs that match to the amphipathic helices of the N-terminal ANCHR motif in ESCRT-III proteins. In the CdvBa2 homologs, this region is not predicted to form alpha-helices due to a conserved proline and the conserved hydrophobic tryptophane is missing. At the C-terminus, in most homologs of both CdvBa1 and CdvBa2 alpha-helices are predicted that match to the MIM1 domain occurring in ESCRT-III proteins. However, in the CdvBa2 group the amino acids demanded by Sciskala and Kölling (22) are less conserved. The MIM2 domain is slightly altered in most homologs of the CdvBa1 group, while in the CdvBa2 group it seems to be conserved. However, in the proteins of *Cand. P. synthrophicum*, which is the only Asgard archaeon yet cultivated and thus whose genome data is trustworthy, it is the other way around. Protein domain compositions in Asgard archaea derived from these findings are displayed on the right. **B:** A possible mechanism based on the

derived protein domain composition as described in the main text. **C:** Table of domains and their functions as derived from literature. Abbreviations: LK: Lokiarchaeum sp. GC14_75, PS: Candidatus P. syntrophicum MK-D1, OD: Candidatus Odinarchaeota archaeon LCB_4, HD: Candidatus Heimdallarchaeota archaeon, TO: Candidatus Thorarchaeota archaeon (strain OWC), TA: Candidatus Thorarchaeota archaeon strain AB_25

Based on these potential interactions, we inferred a possible mechanism of cell division in Asgard archaea (**Figure 3B, Suppl. Figure 1**): First, CdvBa1 proteins bind to the membrane, mediated by the ANCHR domain. Second, both CdvB paralog versions polymerise via the Snf7 domain, giving rise to a filamentous ring-like structure. Third, CdvC disassembles CdvB homologs in the same way as in TACK archaea, constricting the ring. In this final step, our model cannot distinguish whether CdvC first disassembles CdvBa1 or CdvBa2 proteins, because interaction of the proteins can take place either via MIM1-MIT interaction or MIM2-MIT interaction. Furthermore, while in *Cand. P. synthrophicum* the CdvBa1 paralogs seem to possess both MIM1 and MIM2 and the CdvBa2 paralog seems only to possess MIM1, in all other Asgard archaea it is the other way round. There, CdvBa1 proteins only possess MIM1 domains, while CdvBa2 proteins possess correct MIM2 motifs but only fragmented MIM1 domains (**Figure 3**). Thus, differences between the two groups of CdvB in Asgard archaea will have to be investigated experimentally until more high-quality data is available. Also, it might be possible that one of the interactions can lead to binding of CdvC to CdvB without subsequent disassembly.

Finally, as our results suggest mechanistically different Cdv-including mechanisms within the archaeal kingdom, we asked which evolutionary processes might have given rise to these

variations and how a simple Cdv machinery in a common ancestor might have looked like. In particular, we wanted to find out if the domain architecture of such an evolutionary early machinery may have been sufficient to give rise to a cell division mechanism.

An ancestral Cdv-including machinery was probably able to perform constriction, but membrane tethering remains unclear.

Based on the structure of the protein phylogenies, the domain architectures and gene cluster analysis (**Figure 2, Suppl. Fig 2, Suppl. Fig. 4**), we inferred an evolutionary scenario of domain reorganisation by maximum parsimony (**Figure 4**). Interestingly, this scenario indicates that all CdvB duplication events took place after speciation of the common ancestor into Euryarchaeota, Asgard and TACK, such that in the 'original' machinery there most likely was only one CdvB protein. This protein was probably composed of two domains, Snf7 and MIM2, such that it might have been able to form filaments based on the Snf7 domain. Furthermore, the scenario suggests that the common ancestor also possessed a complete CdvC protein, such that the early CdvB protein might have been able to interact with it via the MIM2 domain. Thus, it is possible that these two proteins together were already able to form a constricting ring by CdvB polymerisation and CdvC-mediated disassembly.

Figure 4: Potential evolutionary development of Cdv/ESCRT machineries. Scenario based on phylogenetic trees (**Figure 2** and **Suppl. Figure 4**) and maximum parsimony. The 'original' Cdv machinery probably consisted of one CdvB protein, composed of Snf7 and MIM2, together with a complete CdvC protein. In Euryarchaeota, this machinery was mostly lost (or it was developed after the divergence of Euryarchaeota and some Euryarchaeota acquired it by horizontal gene transfer). In Asgard archaea, CdvB duplicated after the MIM1 domain was gained, and then in the CdvBa1 paralog the membrane-binding ANCHR motif was developed.

In TACK archaea, an early version of CdvA was developed, composed of the beta-sheet- and alpha-helix-rich domains. In Thaumarchaeota, a MIM2 domain was added to CdvA, allowing it to interact with the other Cdv proteins. In Crenarchaeota, this was achieved by development of the BWI and BWH in CdvA and CdvB. Additionally, here CdvB duplicated to form the three distinct paralogs CdvB, CdvB1/2 and CdvB3. This scenario indicates that there existed a Cdv machinery in the common ancestor of TACK, Asgard and Euryarchaeota. The domain architecture of this ancient machinery would allow a constriction mechanism comparable to but simpler than today's Cdv/ESCRT machineries. If and how membrane tethering might have been achieved in this machinery remains unclear. It is possible that the membrane-binding ANCHR motif already existed then and was lost in TACK and Euryarchaeota, however, this can only be speculated.

To form a cell division machinery, though, CdvB and CdvC must have been able to attach to the cell membrane. The CdvA protein, which probably fulfils this task in TACK archaea, arose after the speciation of TACK, and the ANCHR domain, which might tether CdvB to the membrane in Asgard archaea, arose only after the speciation of Asgard (**Figure 4**). Thus, it is unclear how the ancient machinery might have fulfilled this task. Interestingly, the region defined as potential ANCHR in Asgard CdvB homologs is also present in TACK CdvB homologs, but it does not form an alpha helix due to a helix-breaking proline. Hence, it is possible that in the evolutionary early CdvB protein this region was functioning as a membrane tether, and that the gain of CdvA made this capability redundant or even obstructive, such that it got lost. If this is true, the ancestral machinery might have been a true cell division machinery.

Discussion

The mechanism we suggested for Crenarchaeota fits nicely to a recent study of Tarrason Risa et al. (6) which investigated the role of CdvB homologs in *S. acidocaldarius*. This organism possesses one protein of the CdvB class, two of the CdvB1/2 class and one of the CdvB3 class. The experiments showed that the CdvB homologs pass four distinct steps: First, the CdvB class proteins assemble at the division site to form a non-constricting ring. Second, CdvB1/2 class proteins are recruited to that ring. Third, CdvB class proteins get disassembled from the division site, leading to a ring consisting of only CdvB1/2 class proteins. Fourth, the CdvB1/2 class ring constricts. Disregarding CdvB3, this is exactly what our model indicates. While Tarrason Risa et al. described these findings but did not detail the mechanism, our study now provides a plausible scenario of how disassembly of CdvB proteins from the ring can take place and why CdvB class homologs are disassembled before CdvB1/2 class homologs. As the fluorescence microscopy experiment did not yet involve labelled CdvA and CdvC proteins, it would be interesting to see whether our model still holds true in studies investigating the full machinery.

In their paper, Tarrason Risa et al. (6) hypothesised that constriction could be the result of different preferential curvatures of CdvB and CdvB1/2 class proteins. Thus, when CdvB class proteins are removed by CdvC, the ring would shrink to the smaller preferred diameter of CdvB1/2 class proteins. Furthermore, a simulation in this study indicated that to achieve full cell division disassembly of CdvB and CdvB1/2 class homologs is needed in addition to the change in preferred curvature (6). This explains why both CdvB and CdvB1/2 class homologs possess a MIM2 domain. Furthermore, as we found only one possible interaction for CdvB3 class homologs, namely binding to other CdvB homologs, this need of disassembly suggests that CdvB3 plays a different role. If it cannot interact with CdvC, it cannot be disassembled like CdvB and CdvB1/2. One straightforward idea for the function of CdvB3 might be that it can

temporarily stabilise the ring of CdvB and CdvB1/2 homologs while or after CdvC disassembles a component. Simple knockout experiments should provide first insights into the role of CdvB3.

As CdvA was not labelled in their experiment, Tarrason Risa et al. (6) did not comment on how the ring of CdvB1/2 proteins could constrict the membrane without being able to bind to CdvA and thus not being tethered to the membrane. This might be explained by our model. Initially, CdvC mostly disassembles CdvB class proteins due to the potential higher affinity of their MIM2 domain. However, when concentrations change due to disassembly, and more and more CdvC binds to CdvB1/2 class proteins, there should be still some CdvB class proteins left. Thus, even during the final constriction stage, there should always be some remaining membrane anchors, as CdvB class proteins are never entirely removed. It will be very interesting to test the idea of different MIM2-MIT affinities.

There is, however, one result from another study which cannot be explained by our model. Based on data taken with a newly developed microscope, Pulschen et al. (5) suggested that the two different class CdvB1/2 proteins of *S. acidocaldarius*, called CdvB1 and CdvB2, may play different roles. While they colocalised in their experiments as expected, cells lacking CdvB1 did occasionally fail cell division, whilst cells lacking CdvB2 showed a wrongly positioned division site. As we classified them as two paralogs belonging to the same group, we cannot explain this behaviour. However, we did not investigate this in detail, as it is a specific modification in one particular subgroup of Crenarchaeota. Here, we tried to infer common shared principles, which might then be slightly adapted individually by single organisms. Furthermore, none of the proteins were essential, as would be expected from our model. Unfortunately, they did not test cells lacking both CdvB1 and CdvB2, which would be very

interesting for validating the theory of membrane constriction due to different preferred curvatures. Such double-knockout experiments would greatly help understanding the mechanism of constriction.

To conclude, in this paper we stated two possible scenarios for Cdv-including cell division in Thaumarchaeota. While we strongly favour one of them, only experimental work can reveal which one comes closer to reality. As in one scenario CdvB homologs are involved, while in the other they are not, the most straightforward way to answer this question might be to knock out all CdvB homologs and see whether cell division is still possible.

While our proposed models of cell division in TACK archaea provide a theoretical basis that should inspire experimentalists to new studies, the results also show that findings in TACK archaea are only poorly transferable to the eukaryotic ESCRT system. However, we demonstrated the high potential of research in Asgard archaea, whose Cdv machinery we found to be closely related to ESCRT. Based on our results we recommend two *in vitro* experiments that might be possible to conduct despite the hard task of Asgard cultivation. First, as we suggested that CdvBa1 proteins can bind to the membrane via their ANCHR domain, whilst CdvBa2 proteins cannot, this should be tested *in vitro* to verify the central idea of our model. Second, as we could not quantify differences between CdvBa1 and CdvBa2 paralogs in their affinity to CdvC, experiments providing insight into this would greatly help to evolve the model further. In addition, a recent study by Hatano et al. (13) further suggests that at least some Asgard archaea also possess functional homologs of the additional proteins involved in the eukaryotic ESCRT machinery, ESCRT-I and ESCRT-II, and that they might be involved in ubiquitin-directed recruitment of CdvB (ESCRT-III). While we did not investigate these additional components here, as our main goal was to find similarities between all archaea that

could be traced back to a very simple division machinery, this study once more emphasises the strong connection between Asgard and the eukaryotic machineries.

The close relationship between Cdv proteins in Asgard archaea and Eukaryotes further allows an interesting hypothesis about membrane binding in the ancient Cdv machinery in the common ancestor of all Cdv machineries. On the one hand, the potential ANCHR domain in Asgard archaea suggests that some CdvB homologs in these organisms can directly bind membranes, just as some ESCRT-III proteins in Eukaryotes. On the other hand, recent analyses of the membrane composition of Asgard archaea suggest that their lipid composition is comparable to other archaea, strongly differing from Bacteria and Eukaryotes. Thus, it is possible that such a direct membrane tethering of CdvB proteins was also possible in the common ancestor, whose membrane probably consisted of similar lipids as archaea today. Furthermore, some CdvB homologs in TACK archaea possess a region similar to and aligning with the region identified as ANCHR domain in Asgard homologs, but it cannot form an alpha helix due to a helix-breaking proline. Thus, as this region is present in diverged organisms, it is likely that it already existed in the common ancestor. The question is then whether it then was or was not able to form an alpha helix. If it was, then the ancient machinery might have been able to function as a true cell division machinery, consisting of only two proteins.

Beyond cell biology and evolution, the reconstruction of this very simple ancient cell division machinery should be of great interest for the synthetic biology community. Bottom-up synthetic biologists try to build life from scratch, and especially to build basic functional building blocks *in vitro*. One such basic functional building block would be a simple system able to divide a lipid vesicle, mimicking true cell division. A synthetic system reconstituting the ancient Cdv machinery inferred by this study should be a very promising candidate for this yet

unaccomplished endeavour. Especially, it will be very interesting to fabricate synthetic ESCRT-III/CdvB proteins consisting of one Snf7 and one MIM2 domain and to combine them with Vps4 or CdvC proteins *in vitro*, possibly even inserting them into giant unilamellar vesicles. If successful, this would depict an extremely simple system of cell division, that might be easy to study and allow interesting insights into the origin of cell division and life.

Methods

Searching Cdv homologs

Archaeal proteins classified in the eggNOG (25) database as belonging to the Cdv system were used as starting points together with the proteins investigated by Makarova et al. (2) in a previous study. These included the entries ENOG4111F6A (CdvA), COG5491 (all CdvBs) and COG0464 (CdvC), corresponding to entries arCOG04054 (interestingly CdvA and B3), arCOG00452 (cdvB1/2), arCOG00453 (CdvB) and arCOG01307 (CdvC) of the archaea-only database arCOG (26). Using the resulting set of proteins as input, a Python script was designed to run PSI-Blast (27) against the NCBI protein (28) database. Default parameters were used except that the scoring matrix was set to BLOSUM45 (29) due to the distant relationships. Then, Python scripts were written to eliminate all hits from organisms with unclear phylogeny, organisms with fragmented genome data, results of marine sediment probes and duplicates from different strains of the same species. Only in Asgard archaea incomplete genomes were accepted and multiple strains included, because all available data is incomplete except for *Cand. P. syntrophicum*. Hits to CdvC homologs were then further analysed because the AAA+ ATPase region is widely spread amongst many different proteins and accounts for most of the sequence of CdvC proteins. Thus, many false positive hits that are no true CdvC homologs could be expected. Only sequences that possessed at least a VPS4_C or a MIT region were further selected (exception: *F. acidarmanus*, because it is part of the arCOG01307 entry and

had a very good PSI-Blast score). Finally, proteins in close genomic neighbourhood of all finally selected proteins were checked whether they may depict overseen homologs. Eventually, 37 organisms remained and were selected for further analysis. The set of proteins can be found in **Suppl. Table 2** (which also includes FtsZ proteins).

Decomposing proteins into domains

To decompose proteins into domains we used two different approaches. On one hand, we used available software tools to scan the proteins' sequences for domains that are described in public databases. For scanning against the Pfam (30) and InterPro (31) databases we used the InterProScan (32) software, while for scanning against the CDD (33) database we used a custom Python script connecting to the online API of the database. The resulting data files were then processed to summarise hits that represented the same domains or were sub-parts of a larger domain. This led to the identification of the domains CdvA_alpha, BWH, Snf7, MIT, AAA+ ATPase and Vps4_C. On the other hand, we had to check for domains and short conserved motifs which are not defined in the three mentioned databases. For this, we generated multiple sequence alignments and predicted the secondary structure of the proteins. We then inspected the results manually for regions with high sequence conservation and consistent predicted secondary structure, which led to the identification of a beta-sheet rich region in TACK archaeota CdvA proteins, a single conserved beta-sheet in Crenarchaeota CdvA proteins, and the conserved alpha helices in Asgard CdvB proteins. As the beta-sheet rich region in CdvA proteins was previously described as a functional region (1,15,17,18), we labelled it as a unique domain, named CdvA_beta. Similarly, the conserved single beta-sheet in Crenarchaeota CdvA proteins was previously described as functional region for interaction with the BWH domain in CdvB (17), so we classified this as domain, named BWH interaction site (BWI). Furthermore, inspired by Caspi and Dekker (1) and Lu et al. (21), we then compared the conserved regions

in Asgard archaea to the N-terminal ANCHR domain (23) and the C-terminal MIM1 domain (22) occurring in some ESCRT-III proteins (**Figure 3**). This led to the identification of the ANCHR and MIM1 domains. Finally, based on the work of Kojima et al. (24) and Samson et al. (17), a python script was generated to search for the MIM2 motif by regular expressions. Three different kinds of MIM2 were defined: MIM2_Core = φPxφP, MIM2_total = MIM2_Core + xxPφP and MIM2_Sulf = xφxxφφPx + MIM2_Core, where φ represents hydrophobic amino acids [AILMFVPGW], x represents charged amino acids [RKDE] and P is proline. In the two Thaumarchaeota *Nitrososphaera viennensis* and *Candidatus Nitrosphaera gargensis* we classified a proline-rich region as 'putative MIM2', although not perfectly fitting to the regular expressions. This is justified because they aligned to the proline rich regions of the Thaumarchaeota possessing a full MIM2 motif, they did not show other proline-rich regions and the structural limitations of MIT-MIM2 interaction are unknown despite being based on proline. Taken together, these analyses led to the eleven basic building blocks that are utilized in the Cdv system. **Suppl. Table 1** provides an overview of the exact location of the domains on the proteins.

Multiple Sequence Alignment and secondary structure prediction

The Multiple Alignment tool MAFFT (34) was used for sequence alignment. For proteins belonging to different super-phyla (different background colours in Figure 1) the substitution matrix BLOSUM45 (29) was used, for alignment of proteins of the same super-phyla BLOSUM62 (29). The exact command-line instruction was

```
mafft --bl 45 (or 62) --localpair --maxiterate 1000 --reorder input.fasta > output.fasta .
```

Secondary structure prediction based on multiple sequence alignment was executed with JPred (35) using default parameters.

Phylogenetic analysis

Based on the multiple sequence alignments, Bayesian phylogenetics was used to generate phylogenetic trees. The software Mr. Bayes (36) was run with a chain length of 1,000,000, a subsample frequency of 1,000, a burn-in length of 100,000, gamma rate variation with 4 categories and a Poisson rate matrix. The resulting effective sample size was 213 resp. 181.

Mechanism inference

Each protein was abstracted as a combination of domains (**Figures 1A and 3A, Suppl. Fig. 1**). Based on literature, each domain was assigned a specific function (**Figures 1B and 3C**). Then, networks of possible interactions were built, based on potential interactions derived from the literature (**Suppl. Fig. 1, Figures 1B and 3C**). To infer a potential mechanism based on these interaction networks, one starting protein was selected for each mechanism. In TACK archaea, CdvA homologs were selected, because fluorescence microscopy studies indicate that they enrich first at the division site (1)(1,6,17). Also, they are the only Cdv proteins able to bind the membrane in TACK archaea, and they possess a CdvA_beta domain, which we did not assume to interact with other Cdv proteins. Thus, this domain has the potential to be involved in the positioning of CdvA. Neither CdvB homologs nor CdvC did show such “unused” domains in TACK archaea, so if instead these proteins would enrich first at the division site, we could not explain this by domain architecture. In Asgard archaea, we selected CdvBa1 as starting protein, because it is potentially able to bind membranes. However, it is of course possible that instead CdvBa2 enriches first. This would change the order of appearance, but the overall assumed mechanism of polymerisation, membrane binding and CdvC mediated depolymerisation would still hold.

The subsequent process of mechanism inference is visualised in **Suppl. Fig. 1**. After selecting a starting protein, we analysed which interactions were possible for this protein. If only one interaction was possible, we assumed this to occur as a next step. If more than one interaction was possible, we tried to infer which interaction might have a higher affinity, and thus would occur more quickly. If this was not possible, we inferred different scenarios. These steps were then executed recursively. The resulting mechanisms were then summarized and simplified to the mechanisms displayed in **Figure 1**.

Code availability

The code of the analysis scripts is available on request

Limitations of the study

The study is purely computational with the goal to generate hypotheses that can be tested experimentally. Thus, the described mechanistical models are suggestions and must be validated experimentally. Furthermore, we based our analysis purely on domain analysis, so any functional changes that result from small sequence or structural alterations that are not reflected in domain architecture are not incorporated in our models. Such small alterations most likely cause small functional differences between highly similar homologs, hence between closely related organisms. Consequentially, our analysis cannot draw detailed conclusions about single particular organisms. We were interested in general principles, whose implementation in distinct organisms might differ slightly. To infer explicit models for single organisms, more detailed structural modelling of the involved proteins is necessary.

Acknowledgements

We thank Dina Grohmann (University of Regensburg) and Henri Franquelim for helpful feedback on the paper.

Author Contributions

B.F. performed the data analysis, T.H. contributed to the data analysis, B.F., J.C. and P.S. wrote the manuscript and J.C and P.S. directed the project.

Declaration of interests

The authors declare no competing interests.

.

References

1. Caspi Y, Dekker C. Dividing the Archaeal Way: The Ancient Cdv Cell-Division Machinery. Front Microbiol. 2018 Mar 2;9:174. Available from: http://journal.frontiersin.org/article/10.3389/fmicb.2018.00174/full

2. Makarova KS, Yutin N, Bell SD, Koonin E V. Evolution of diverse cell division and vesicle formation systems in Archaea. Nat Rev Microbiol. 2010; 8(10):731-741. doi:10.1038/nrmicro2406

3. Lindås A-C, Karlsson EA, Lindgren MT, Ettema TJG, Bernander R. A unique cell division machinery in the Archaea. Proc Natl Acad Sci U S A. 2008 Dec 2;105(48):18942–6. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18987308

4. Samson RY, Obita T, Freund SM, Williams RL, Bell SD. A role for the ESCRT system in cell division in archaea. Science 2008;322(5908):1710-1713. doi:10.1126/science.1165322

5. Pulschen AA, Mutavchiev DR, Culley S, Sebastian KN, Roubinet J, Roubinet M, et al. Live Imaging of a Hyperthermophilic Archaeon Reveals Distinct Roles for Two ESCRT-III Homologs in Ensuring a Robust and Symmetric Division. Curr Biol. 2020; 30(14):2852-2859.e4. doi:10.1016/j.cub.2020.05.021

6. Tarrason Risa G, Hurtig F, Bray S, Hafner AE, Harker-Kirschneck L, Faull P, et al. The proteasome controls ESCRT-III-mediated cell division in an archaeon. Science. 2020; 369(6504):eaaz2532. doi:10.1126/science.aaz2532

7. Sadler JBA, Wenzel DM, Williams LK, Guindo-Martínez M, Alam SL, Mercader JM, et al. A cancer-associated polymorphism in ESCRT-III disrupts the abscission checkpoint and promotes genome instability. Proc Natl Acad Sci U S A. 2018; 115(38):E8900-E8908. doi:10.1073/pnas.1805504115

8. Votteler J, Sundquist WI. Virus budding and the ESCRT pathway. Cell Host and Microbe. 2013; 14(3):232-241. doi:10.1016/j.chom.2013.08.012

9. Lee JA, Beigneux A, Ahmad ST, Young SG, Gao FB. ESCRT-III Dysfunction Causes Autophagosome Accumulation and Neurodegeneration. Curr Biol. 2007 [published correction appears in Curr Biol. 2007 Sep 18;17(18):1622]; 17(18):1561-1567. doi:10.1016/j.cub.2007.07.029

10. Eme L, Spang A, Lombard J, Stairs CW, Ettema TJG. Archaea and the origin of eukaryotes. Nature Reviews Microbiology. 2017; 15(12):711-723. doi:10.1038/nrmicro.2017.133

11. Adam PS, Borrel G, Brochier-Armanet C, Gribaldo S. The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. ISME J. 2017 Nov 4;11(11):2407–25. Available from: http://www.nature.com/articles/ismej2017122

12. Imachi H, Nobu MK, Nakahara N, Morono Y, Ogawara M, Takaki Y, et al. Isolation of an archaeon at the prokaryote–eukaryote interface. Nature. 2020; 577(7791):519-525. doi:10.1038/s41586-019-1916-6

13. Hatano T, Palani S, Papatziamou D, Souza DP, Salzer R, Tamarit D, et al. Characterisation of the Ubiquitin-ESCRT pathway in Asgard archaea sheds new light on origins of membrane trafficking in eukaryotes. bioRxiv [Preprint]. 2021 Aug 17 [cited 2021 Sep 10];2021.08.17.456605. Available from: https://www.biorxiv.org/content/10.1101/2021.08.17.456605v1

14. Zaremba-Niedzwiedzka K, Caceres EF, Saw JH, Bäckström D, Juzokaite L, Vancaester E, et al. Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature. 2017 Jan 19;541(7637):353–8. Available from: http://www.nature.com/articles/nature21031

15. Moriscot C, Gribaldo S, Jault JM, Krupovic M, Arnaud J, Jamin M, et al. Crenarchaeal CdvA forms double-helical filaments containing DNA and interacts with ESCRT-III-like CdvB. PLoS One. 2011;6(7):e21921. doi:10.1371/journal.pone.0021921

16. Bernander R, Ettema TJ. FtsZ-less cell division in archaea and bacteria. Curr Opin Microbiol. 2010 Dec;13(6):747–52. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1369527410001669

17. Samson RY, Obita T, Hodgson B, Shaw MK, Chong PL-G, Williams RL, et al. Molecular and structural basis of ESCRT-III recruitment to membranes during archaeal cell division. Mol Cell. 2011 Jan;41(2):186–96. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1097276510010038

18. Dobro MJ, Samson RY, Yu Z, McCullough J, Ding HJ, Chong PL-G, et al. Electron cryotomography of ESCRT assemblies and dividing Sulfolobus cells suggests that spiraling

filaments are involved in membrane scission. Mol Biol Cell. 2013 Aug;24(15):2319–27. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23761076

19. Pelve EA, Lindås AC, Martens-Habbena W, de la Torre JR, Stahl DA, Bernander R. Cdv-based cell division and cell cycle organization in the thaumarchaeon Nitrosopumilus maritimus. Mol Microbiol. 2011; 82(3):555-566. doi:10.1111/j.1365-2958.2011.07834.x

20. Ng K-H, Srinivas V, Srinivasan R, Balasubramanian M. The Nitrosopumilus maritimus CdvB, but not FtsZ, assembles into polymers. Archaea. 2013;2013:104147. doi:10.1155/2013/104147

21. Lu Z, Fu T, Li T, Liu Y, Zhang S, Li J, et al. Coevolution of eukaryote-like vps4 and ESCRT-III subunits in the asgard archaea. MBio. 2020; 11(3):e00417-20. Published 2020 May 19. doi:10.1128/mBio.00417-20

22. Sciskala B, Kölling R. Interaction maps of the Saccharomyces cerevisiae ESCRT-III protein Snf7. Eukaryot Cell. 2013 ;12(11):1538-1546. doi:10.1128/EC.00241-13

23. Buchkovich NJ, Henne WM, Tang S, Emr SD. Essential N-Terminal insertion motif anchors the ESCRT-III filament during MVB vesicle formation. Dev Cell. 2013; 27(2):201-214. doi:10.1016/j.devcel.2013.09.009

24. Kojima R, Obita T, Onoue K, Mizuguchi M. Structural Fine-Tuning of MIT-Interacting Motif 2 (MIM2) and Allosteric Regulation of ESCRT-III by Vps4 in Yeast. J Mol Biol. 2016; 428(11):2392-2404. doi:10.1016/j.jmb.2016.04.007

25. Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019; 47(D1):D309-D314. doi:10.1093/nar/gky1085

26. Makarova KS, Wolf YI, Koonin E V. Archaeal clusters of orthologous genes (arCOGs): An update and application for analysis of shared features between thermococcales, methanococcales, and methanobacteriales. Life. 2015; 5(1):818-840. Published 2015 Mar 10. doi:10.3390/life5010818

27. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Research. 1997; 25(17):3389-3402. doi:10.1093/nar/25.17.3389

28. Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, Bolton E, et al. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2016; 46(D1):D8-D13. doi:10.1093/nar/gkx1095

29. Mount DW. Using BLOSUM in Sequence Alignments. Cold Spring Harb Protoc. 2008 Jun 1;2008(6):pdb.top39-pdb.top39. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21356855

30. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, et al. Pfam: The protein families database. Nucleic Acids Research. 2014; 42(Database issue):D222-D230. doi:10.1093/nar/gkt1223

31. Mitchell AL, Attwood TK, Babbitt PC, Blum M, Bork P, Bridge A, et al. InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. 2019; 47(D1):D351-D360. doi:10.1093/nar/gky1100

32. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. InterProScan 5: Genome-scale protein function classification. Bioinformatics. 2014; 30(9):1236-1240. doi:10.1093/bioinformatics/btu031

33. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, et al. CDD: NCBI's conserved domain database. Nucleic Acids Res. 2015; 43(Database issue):D222-D226. doi:10.1093/nar/gku1221

34. Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002 Jul 15;30(14):3059–66. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12136088

35. Drozdetskiy A, Cole C, Procter J, Barton GJ. JPred4: A protein secondary structure prediction server. Nucleic Acids Res. 2015; 43(W1):W389-W394. doi:10.1093/nar/gkv332

36. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst Biol. 2012 May 1;61(3):539–42. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22357727

Supplementary Material

Suppl. Table 1: Domain locations of the investigated proteins.

Suppl. Table 2: Protein and gene data of the investigated proteins.

Suppl. Fig. 1: Interactions underlying the suggested mechanisms. In Crenarchaeota, the potential polymerisation capabilities of CdvA and CdvC are not displayed as they do not help explain the suggested mechanism. In Asgard archaea, the CdvC polymerisation capability is likewise not displayed.

Background colours indicate phylogenetic groups. Red: Asgard archaea, dark green: Thaumarchaeota, pale green: Crenarchaeota.

Suppl. Fig. 2: Cluster analysis of Cdv genes in archaea. Organisms are grouped by phylogeny. Genes encoding CdvA, CdvB and CdvC are in direct neighbourhood in Crenarchaeota. In Thaumarchaeota, in three out of four organisms one CdvB homolog gene is in close neighbourhood to the CdvC gene. CdvA homologs are not. In Euryarchaeota, if there are genes encoding both CdvB and CdvC homologs, they are either in direct or in relatively close neighbourhood. In Asgard archaea, except in *Cand. P. syntrophicum* only fragmentary genomes are available. In three out of six Asgard archaea genomes CdvC and CdvB homolog genes are in close neighbourhood. This pattern across all archaeal phyla indicates that indeed CdvC and one CdvB homolog existed in the last common ancestor of TACK, Asgard and Euryarchaeota.

Suppl. Fig. 3: Multiple Alignments of the MIM2 domain in CdvB and CdvB1/2 proteins. In CdvB1/2 proteins the MIM2 domain contains consistently fewer prolines, indicating a weaker affinity to the MIT domain than the MIM2 of CdvB proteins.

Abbreviations: AP: Aeropyrum pernix, AS: Acidilobus saccharovorans, PO: Pyrodictium occultum, HB: Hyperthermus butylicus, AH: Acidianus hospitalis, AM: Acidianus manzaensis, AS: Acidianus sulfidivorans, SI: Sulfolobus islandicus, SS: Saccharolobus solfataricus, SA: Sulfolobus acidocaldarius, ST: Sulfurisphaera tokodaii, MS: Metallosphaera sedula, FF: Fervidicoccus fontis, DA: Desulfurococcus amylolyticus, SM: Staphylothermus marinus, CL: Caldisphaeralagunensis, IH: Ignicoccus hospitalis

Suppl. Fig. 4: Phylogenetic Tree of CdvC proteins. The overall structure shows the same phylogenetic relations as the CdvB tree **(Figure 2**): Asgard archaea are located between Euryarchaeota and Crenarchaeota, while Thaumarchaeota are most distant from Asgard archaea. However, the Euryarchaeota group of Thermoplasma is not located near all other Euryarchaeota but between Asgard archaea and TACK archaea. This might be an artifact because the proteins in Thermoplasma mutated massively and maybe gained an entirely new function (the branch is the longest in the whole phylogeny), or it might be caused by horizontal gene transfer.

Background colours indicate phylogenetic groups. Blue: Euryarchaeota, red: Asgard archaea, dark green: Thaumarchaeota, pale green: Crenarchaeota. Numbers indicate posterior probabilities.

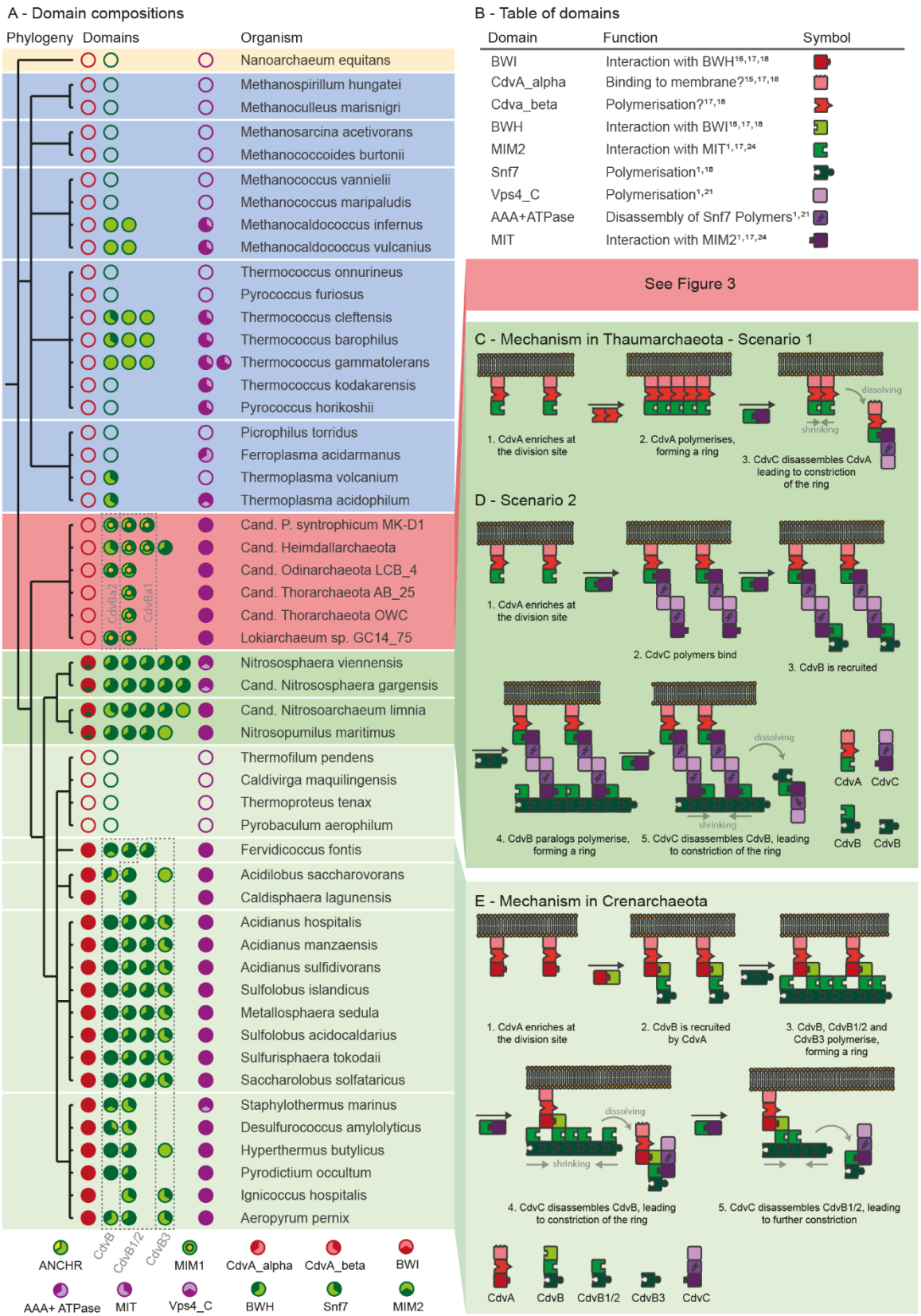
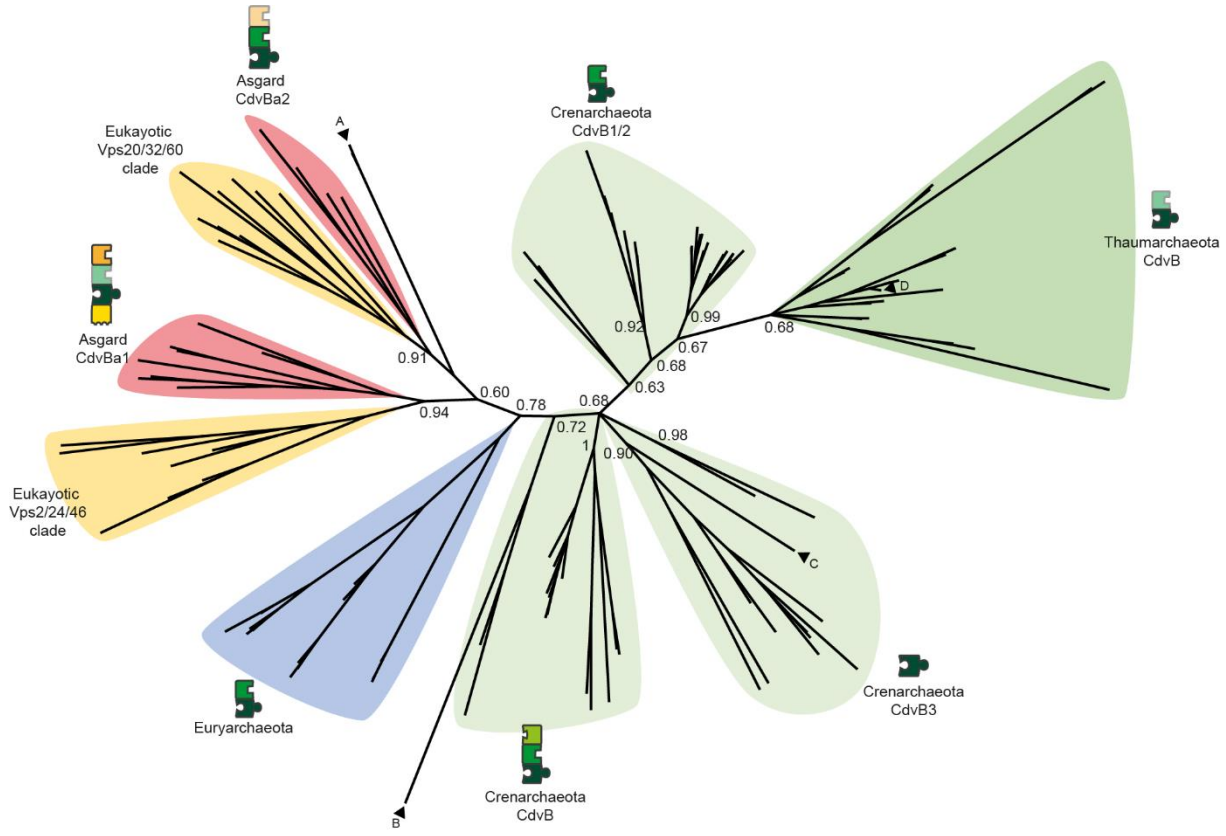

A - Domain compositions
Phylogeny
Domains
Organism
Nanoarchaeum equitans
Methanospirillum hungatei
Methanoculleus marisnigri
Methanosarcina acetivorans
Methanococcoides burtonii
Methanococcus vannielii
Methanococcus maripaludis
Methanocaldococcus infernus
Methanocaldococcus vulcanius
Thermococcus onnurineus
Pyrococcus furiosus
Thermococcus cleftensis
Thermococcus barophilus
Thermococcus gammatolerans
Thermococcus kodakarensis
Pyrococcus horikoshii
Picrophilus torridus
Ferroplasma acidarmanus
Thermoplasma volcanium
Thermoplasma acidophilum
Cand. P. syntrophicum MK-D1
Cand. Heimdallarchaeota
Cand. Odinarchaeota LCB_4
Cand. Thorarchaeota AB_25
Cand. Thorarchaeota OWC
Lokiarchaeum sp. GC14_75
CdvBa2
CdvBa1
Nitrososphaera viennensis
Cand. Nitrososphaera gargensis
Cand. Nitrosoarchaeum limnia
Nitrosopumilus maritimus
Thermofilum pendens
Caldivirga maquilingensis
Thermoproteus tenax
Pyrobaculum aerophilum
Fervidicoccus fontis
Acidilobus saccharovorans
Caldisphaera lagunensis
Acidianus hospitalis
Acidianus manzaensis
Acidianus sulfidivorans
Sulfolobus islandicus
Metallosphaera sedula
Sulfolobus acidocaldarius
Sulfurisphaera tokodaii
Saccharolobus solfataricus
Staphylothermus marinus
Desulfurococcus amylolyticus
Hyperthermus butylicus
Pyrodictium occultum
Ignicoccus hospitalis
Aeropyrum pernix
ANCHR
CdvB
CdvB1/2
CdvB3
MIM1
CdvA_alpha
CdvA_beta
BWI
AAA+ ATPase
MIT
Vps4_C
BWH
Snf7
MIM2
B - Table of domains
Domain
Function
Symbol
BWI
Interaction with BWH[15,17,18]
CdvA_alpha
Binding to membrane?[15,17,18]
Cdva_beta
Polymerisation?[17,18]
BWH
Interaction with BWI[15,17,18]
MIM2
Interaction with MIT[1,17,24]
Snf7
Polymerisation[1,18]
Vps4_C
Polymerisation[1,21]
AAA+ATPase
Disassembly of Snf7 Polymers[1,21]
MIT
Interaction with MIM2[1,17,24]
See Figure 3
C - Mechanism in Thaumarchaeota - Scenario 1
1. CdvA enriches at the division site
2. CdvA polymerises, forming a ring
3. CdvC disassembles CdvA leading to constriction of the ring
dissolving
shrinking
D - Scenario 2
1. CdvA enriches at the division site
2. CdvC polymers bind
3. CdvB is recruited
4. CdvB paralogs polymerise, forming a ring
5. CdvC disassembles CdvB, leading to constriction of the ring
dissolving
shrinking
CdvA
CdvC
CdvB
CdvB
E - Mechanism in Crenarchaeota
1. CdvA enriches at the division site
2. CdvB is recruited by CdvA
3. CdvB, CdvB1/2 and CdvB3 polymerise, forming a ring
4. CdvC disassembles CdvB, leading to constriction of the ring
5. CdvC disassembles CdvB1/2, leading to further constriction
dissolving
shrinking
CdvA
CdvB
CdvB1/2
CdvB3
CdvC

Figure 1



873

874 Figure 2

875

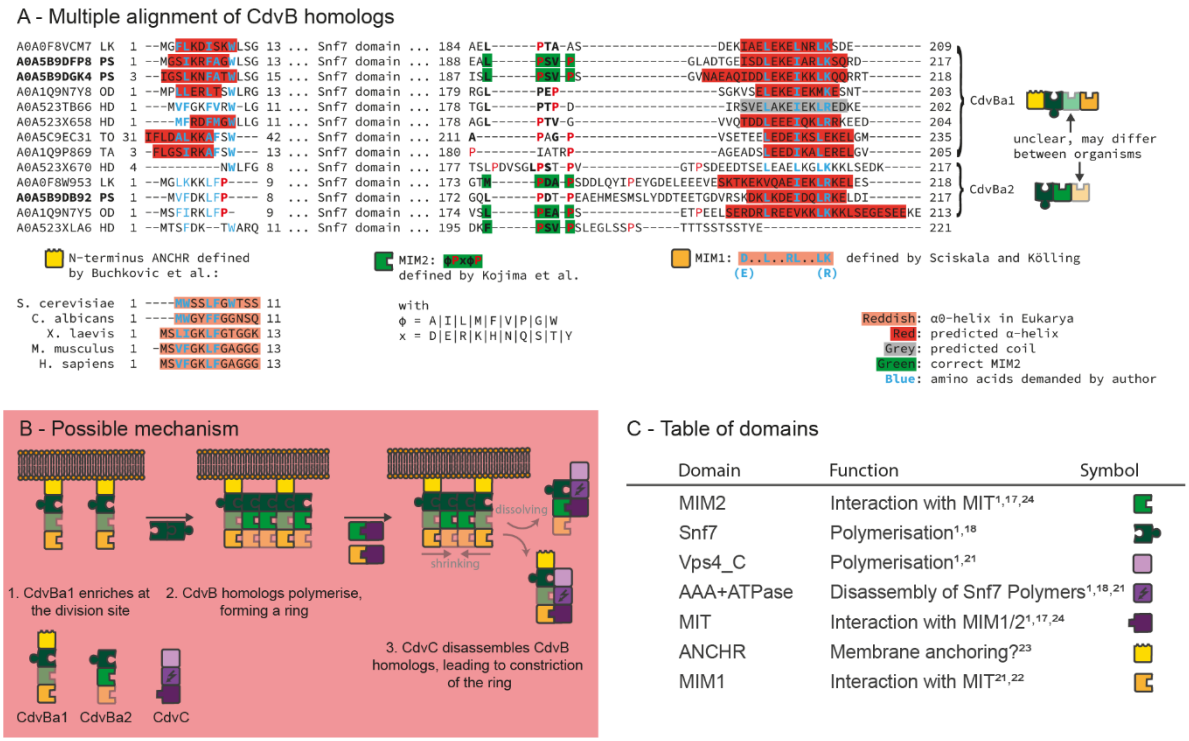

A - Multiple alignment of CdvB homologs
CdvBa1
unclear, may differ between organisms
CdvBa2
N-terminus ANCHR defined by Buchkovic et al.:
MIM2: ΦΦxΦΦ defined by Kojima et al.
with
Φ = A|I|L|M|F|V|P|G|W
x = D|E|R|K|H|N|Q|S|T|Y
MIM1: D..L..RL..LK defined by Sciskala and Kölling
(E) (R)
Reddish: α0-helix in Eukarya
Red: predicted α-helix
Grey: predicted coil
Green: correct MIM2
Blue: amino acids demanded by author
B - Possible mechanism
ATP
dissolving
shrinking
1. CdvBa1 enriches at the division site
2. CdvB homologs polymerise, forming a ring
3. CdvC disassembles CdvB homologs, leading to constriction of the ring
CdvBa1
CdvBa2
CdvC
C - Table of domains
Domain
Function
Symbol
MIM2
Interaction with MIT[1,17,24]
Snf7
Polymerisation[1,18]
Vps4_C
Polymerisation[1,21]
AAA+ATPase
Disassembly of Snf7 Polymers[1,18,21]
MIT
Interaction with MIM1/2[1,17,24]
ANCHR
Membrane anchoring?[23]
MIM1
Interaction with MIT[21,22]

Figure 3

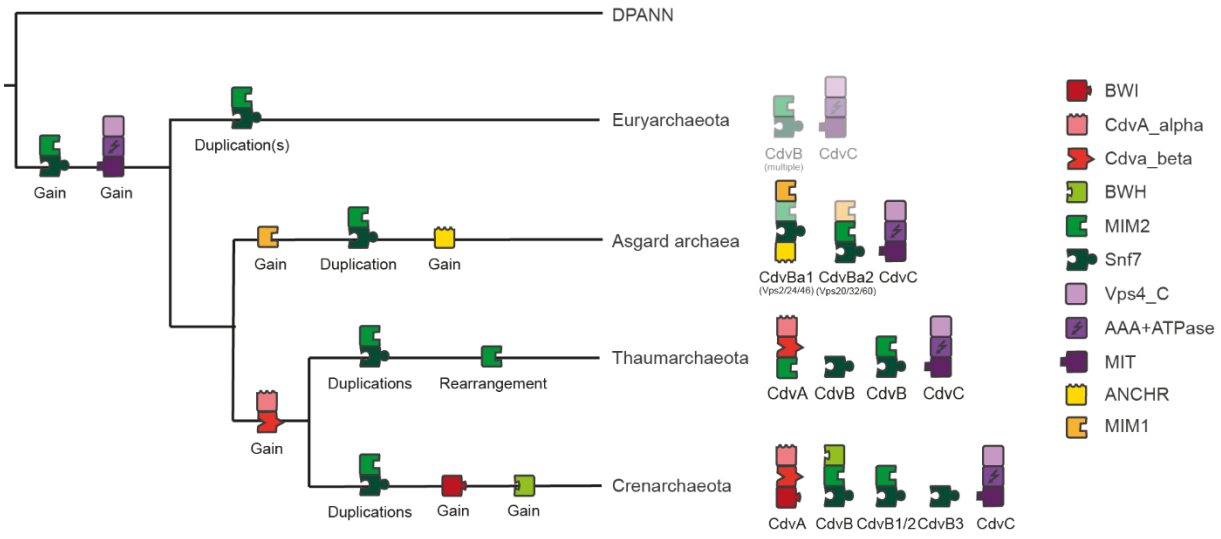
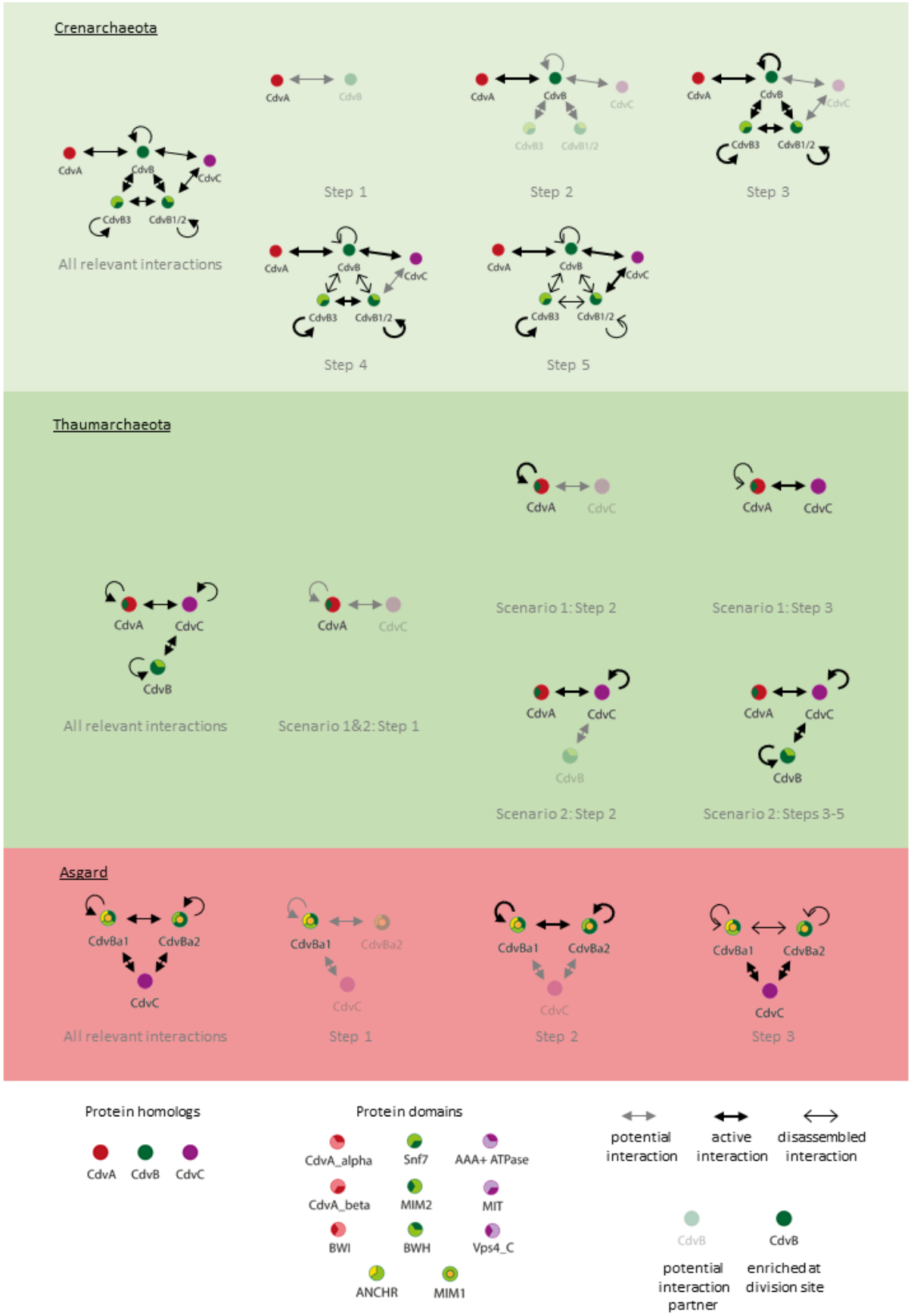
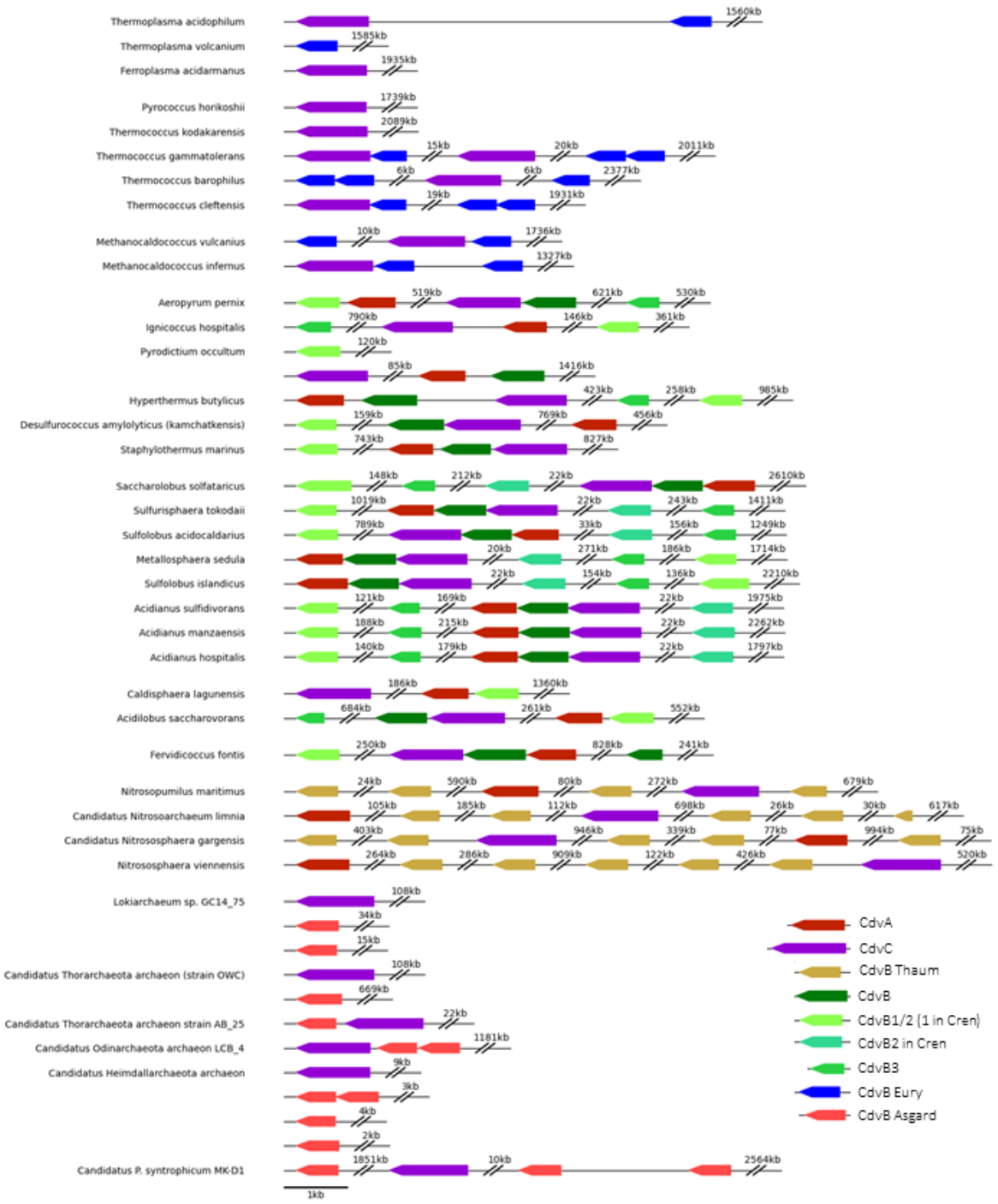

DPANN
Euryarchaeota
Asgard archaea
Thaumarchaeota
Crenarchaeota
Gain
Gain
Duplication(s)
Gain
Duplication
Gain
Duplications
Rearrangement
Gain
Duplications
Gain
Gain
CdvB
(multiple)
CdvC
CdvBa1
(Vps2/24/46)
CdvBa2
(Vps20/32/60)
CdvC
CdvA
CdvB
CdvB
CdvC
CdvA
CdvB
CdvB1/2
CdvB3
CdvC
BWI
CdvA_alpha
Cdva_beta
BWH
MIM2
Snf7
Vps4_C
AAA+ATPase
MIT
ANCHR
MIM1

Figure 4



Suppl. Figure 1



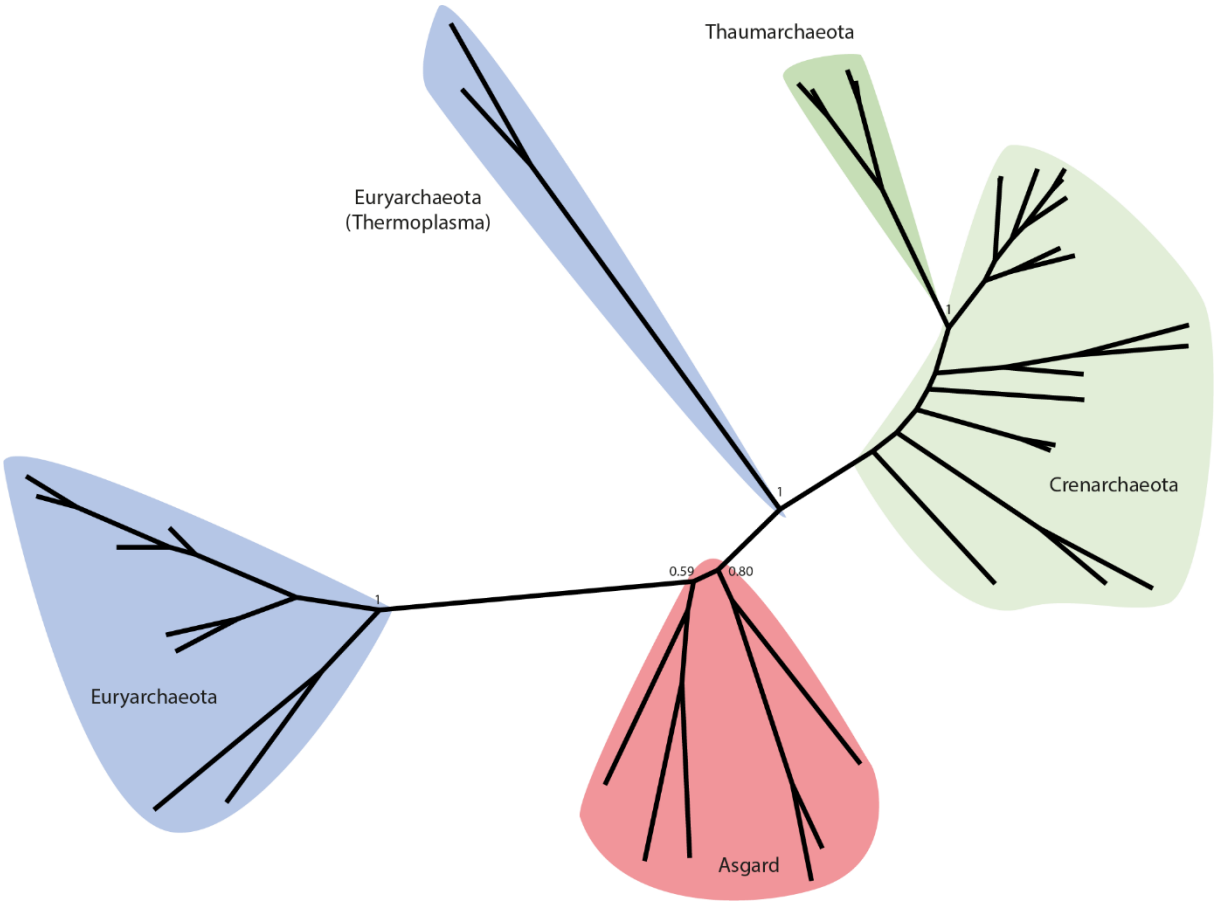
884

885 Suppl. Figure 2

```
                CdvB                                        CdvB1/2

Q9YDF1      AP VP--DPERI----            Q97ZX1      SS FPDLPVGTGLASKSNS-
D9Q219      AS TP--EVPKS----            F0NFA9      SI FPDLPVGAGLASKSNS-
A0A0V8RVW9  PO FP--EPPVEPVIR            F9VMM4      ST FPELPAGGLASSQKS--
A2BKY9      HB FP--EPPVEPIIH            Q4JBG6      SA FPSLPAAGITQQQKS--
F4B4A9      AH LP--DLPHPPSTT            A4YIR6      MS FPDLPVNALAQRA----
A0A1W6K243  AM LP--DLPHPPNTV            A0A2U9IN05  AS FPDLPLSSATINQKA--
A0A2U9IP34  AS LP--DLPHPPSNV            A0A1W6K0Q6  AM FPDLPLSSAAINQKA--
F0NEW1      SI LP--DLPHPPIEQ            F4B970      AH FPELPLSSSTISQKA--
Q97ZJ6      SS LP--DLPHPPVEQ            Q97ZL3      SS FPELPSFVTSTQKVSNQ
Q4J924      SA LP--ELPHPPSEL            F0NEY7      SI FPELPSFVTSTQKVSNQ
Q972B4      ST LP--ELPYPPSDT            Q971Z1      ST FPSLPSFVTSVAQKSNE
A4YHC4      MS MPKLDLPHPPRDV            Q4J8Y4      SA FPSLPSFATSVDQKTNA
I0A072      FF LP--EPKKLIQQT            A0A2U9IP10  AS FPELPSMVTSVQKVNEN
I0A2N3      FF LP--SPP------            A0A1W6K1Z7  AM FPELPSMLTSAQKVNEN
I3XQE4      DA VK--IPPSPPSND            F4B4L2      AH FPELPSMVTTTQKAAPT
A3DP08      SM LP--IPPLPPENQ            A4YHE8      MS FPDLPALVTTTQKAENG
                                        A2BMT1      HB FPELPTAFPTPSGQQDQ
                                        A0A0V8RRG7  PO FPELPSAFPSPTQQGSQ
                                        H9ZZG7      FF FPELPGGSMNLPAPPSA
                                        Q9YFV7      AP FPDLPSFEKVAAQGERT
                                        L0ABT6      CL FPEIPSVEGVPSSANNL
                                        D9Q2V7      AS FPELPSVEGAPTSTGQQ
                                        I3XPX4      DA FAELPKTIVPEGESVKS
                                        A3DLS9      SM YAELPKLLTEAEKSGEV
                                        A8ABN1      IH FPSVSIPNTSRV-----
```

Suppl. Figure 3



Suppl. Figure 4