Protein domain patterns reveal a functional diversity of the archaeal Cdv system and give insights into the origin of the eukaryotic ESCRT system

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 archaeal cell division machinery Cdv is closely related to the eukaryotic ESCRT system and it is often suggested that Cdv may represent a simplified functional model of ESCRT. However, experimental data suggests that even amongst archaea Cdv-based mechanisms differ, questioning the idea of a common basic principle. Furthermore, both Cdv and ESCRT have had the same time to evolve since their deviation from a putative common ancestor, a fact which is often ignored when archaea are treated as 'simpler versions' of eukaryotes. Here, we use a range of computational methods to elucidate these functional differences and to provide a guide on which Cdv-based mechanisms may or may not be compared to ESCRT. We infer a comprehensive mechanistic theory of Cdv-based cell division based on protein domains that correctly predicts the functional differences found between organisms in experiments and describes the protein evolution that underlies this functional diversity. From these results we infer that there are at least three evolutionary and functionally different Cdv-based systems in archaea, complicating the idea of comparative approaches to ESCRT. However, we describe that the Cdv machinery found in the archaeal super-phylum Asgard probably is functionally highly comparable to the eukaryotic ESCRT system, making it a promising candidate for comparative studies. Taken together, via a novel mechanistic theory of archaeal Cdv-based systems we explain experimental findings of the past and provide a guide for various hypothesis-driven experiments in the future that may lead to a functional model of the highly researched eukaryotic ESCRT system.

Keywords: Cdv, ESCRT, Asgard, Cell division, Archaea, Protein evolution, Protein domains

Introduction

The archaeal cell division machinery Cdv shows close homology to the highly researched eukaryotic membrane bending system ESCRT[1–4]. It has thus been speculated that the relatively simple archaeal machinery may be used as a model system for the eukaryotic machinery, assuming that both are based on the same functional principles[1,2,5,6]. However, due to the great technical difficulties of working with organisms thriving under extreme conditions, progress in Cdv-research is slow. Furthermore, proteins from different archaeal organisms classified as Cdv components seem to give rise to slightly different mechanisms[1,2,7], questioning the idea of a common principle. Therefore, to infer whether Cdv can indeed be used as a model for ESCRT and, if so, to decide which organism to choose best, a deeper understanding of the functional mechanisms of Cdv and their diversity among archaeal organisms is required.

On a molecular level, most biological processes are executed by proteins. However, proteins can exhibit a variety of different functions, dependent on their cellular environment. Therefore, we here propose a new approach, i.e., to focus on the structural domains of the proteins, for elucidating the complex interactions that establish function. From this point of view, multi-protein systems like Cdv or ESCRT can be abstracted as networks of interactions between the functional domains of the involved proteins. Compared with classical protein-based networks, such domain-based networks feature an additional degree of freedom, in the way that they do not only depend on which domains are involved, but also on how they are arranged on the respective proteins. On one hand, the gain or loss of just a few domains in a biological system can massively alter the whole system's function. On the other hand, very different mechanisms can arise from the same set of domains, depending on how they are organised. Thus, the differences between Cdv systems found in archaea might be the result of differences in domain usage and architecture of the involved proteins.

In general, the Cdv system consists of three proteins named CdvA, CdvB and CdvC[1–4]. They were first found in the Crenarchaeon *Sulfolobus acidocaldarius*[3,4], whose system was since treated as prototype machinery for Cdv research[1,2]. However, not all archaeal organisms contain the CdvA protein, and most organisms possess multiple paralogs of CdvB, the number of versions varying. Furthermore, typically some of the CdvB paralogs miss domains[1,2]. Thus, there indeed are variations in protein and domain composition that have the potential to explain functional differences between organisms.

These variations are closely linked to the phylogeny of archaea. There are four archaeal super-phyla, named DPANN archaea, Euryarchaeota, Asgard archaea and TACK archaea[8], and three of them contain organisms that possess Cdv proteins: Euryarchaeota, Asgard archaea and TACK archaea[9]. In Euryarchaeota, most organisms do not possess a Cdv system at all, but some few species show CdvB and/or CdvC proteins encoded in their genome. In Asgard archaea, only little is known because it was just recently possible to cultivate a member of this super-phylum[5,10], but metagenome data and protein prediction indicates that genes encoding for CdvB and CdvC proteins are common within this group. In TACK archaea, all groups except Thermoproteales possess a full system of CdvA, CdvB and CdvC.

As the Cdv systems of Euryarchaeota, Asgard archaea and TACK archaea are composed of different sets of proteins, different functional mechanisms may be assumed. However, up to now, experimental results exist only for organisms of the TACK group[1,4–7,11–15]. While it is mostly suggested that Cdv does not play a role in Euryarchaeota at all and thus experiments do not seem promising, the lack of data for Asgard archaea constitutes a severe problem. Recently, it was suggested that all Eukarya originate from this archaeal group rather than from a common eukaryotic-archaeal ancestor[16]. Thus, if this hypothesis should be correct, a Cdv-based mechanism of Asgard archaea would depict the most promising candidate to be compared to the eukaryotic ESCRT system[1,17]. Results about the Cdv machinery of Asgard archaea are therefore urgently needed.

Surprisingly, despite having the same set of proteins, experiments conducted with TACK archaea showed unequal results in the two main groups of the super-phylum, Crenarchaeota and Thaumarchaeota[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[1,5,6,11,13,15]. 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[1,11,13]. Then, it is assumed that this ring constricts with the help of CdvC, leading to membrane constriction and cell-division[1,5,6,15]. In contrast, in the Thaumarchaeon *Nitrosopumilus maritimus*, only CdvC concentrates at the division site together with CdvA, while the distribution of CdvB is diffuse[1,7,14]. Up to now, there exists no explanation for this deviation and no model of how membrane constriction may be achieved in Thaumarchaeota.

At first glance the difference seems troublesome, as both Cren- and Thaumarchaeota seem to possess the same full set of Cdv proteins. However, all comparative studies of Cdv systems so far have only investigated the existence or absence of Cdv proteins based on sequence homology, but not the domain compositions of these proteins[1,2]. Thus, if two proteins shared very long domains, but possessed different short domains, they were classified as homologs in these studies. But while evolutionary correct, from a functional perspective this may be a misclassification, because biological systems are a result of domain interactions as explained above, not a function of sequence homology. If proteins differ in their domain architecture, they have the potential to give rise to entirely different higher-order systems despite a high sequence similarity. In this case, they would not be functional homologs.

Inspired by this thought, here we analyse the domain composition of archaeal Cdv systems of Euryarchaeota, Asgard and TACK archaea and show that organisms utilise the same domains but arrange them to different architectures. Based on this, we provide a possible explanation of functional differences found between Cren- and Thaumarchaeota in experiments and propose mechanistical models for Cdv-based mechanisms of cell division. Furthermore, we present an in-depth bioinformatic analysis of Cdv-based cell division in Asgard archaea, suggesting that this system might be an intermediate version of Cdv and ESCRT. Finally, we disclose the protein evolution that caused this functional diversity and state to what degree archaeal Cdv systems may or may not be used as models for ESCRT, offering experimentalists a guide which proteins and organisms might be best to focus on in future research.

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 Cdv proteins. From the resulting list we selected 37 organisms by quality and analysed the domain composition of their Cdv proteins. As Figure 1A and B show, eleven different domains occurred 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 Cdv systems developed differently from their common ancestor, providing a hint that this might be an explanation for differences in function. Furthermore, in Asgard archaea, we found two additional candidates for domains. Interestingly, these matched two domains occurring in the eukaryotic ESCRT machinery. First, consistent with the results of Lu et al.[17], we found potential MIM1[18] domains at the C-terminus of most Asgard CdvB homologs. Second, consistent with the results of Caspi et al.[1], we predicted alpha-helices at the N-terminus of one group of Asgard CdvB homologs that match to the ESCRT-III N-terminal ANCHR[19] motif. Thus, while the domains shared between all organisms suggest that all Cdv-based systems have a common ancestor, the domains occurring only in specific phylogenetic groups indicate that the original mechanisms of functional interactions differentiated with evolutionary divergence. To analyse this differentiation, we next investigated the domain compositions of proteins within each super-phylum in detail, starting with Euryarchaeota.

Cdv protein domains in Euryarchaeota do not allow a working 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 were found, making any kind of functional system similar to Cdv-based mechanisms in other archaea very unlikely. Thus, our findings confirm the widely accepted idea that Euryarchaeota do not use a Cdv-based 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 different from other organisms' Cdv protein functions.

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

In contrast to Euryarchaeota, nearly all Cdv proteins in TACK archaea could be deconstructed to specific combinations of conserved domains. The domain architecture of proteins showed patterns that were relatively consistent within phylogenetic orders (Figure 1A, green background), suggesting that mechanisms may be comparable between organisms of the same order, but showed distinct differences at the level of phyla. While in all Thaumarchaeota the proteins previously 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. As BWI and BWH are known to be able to interact, the joint absence or presence of the two domains indicated a functional difference in the machineries of Cren- and Thaumarchaeota. To further investigate 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.

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 we predict CdvA to bind to CdvC, in Crenarchaeota we suggest that CdvA binds to CdvB. This finding matches well to the differences between the two phyla found in fluorescence microscopy studies[1,5–7,11,12,14]. 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 are exactly what our results predict. 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 composition suggests. Reassured by this experimental support, 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 predict two different systems in TACK archaea

First, as Crenarchaeota organisms showed variations in the number of CdvB homologs (Figure 1A), we had to make CdvB homologs comparable between organisms to derive a common basic mechanism. Thus, we first clustered the homologs by phylogeny (Figure 2) and then analysed the domain architectures within the clusters. The result was fully consistent and confirming, meaning that differences in domain architecture fitted perfectly to the phylogenetic clusters (Figure 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). Next, as a basis for our model, we assigned one specific function to each domain (Figure 1B), derived from literature. Then, we used one protein as the initiating protein and predicted the occurring higher order processes resulting from PPIs and other domain functions (Figure 1C – E).

In Crenarchaeota (Figure 1E), the predicted mechanism starts with CdvA, because it is the only protein with a domain able to bind to the membrane, and experiments show that neither CdvB nor CdvC can bind the membrane[1,11–13]. Also, it is validated that CdvA is the first protein to enrich at the division site[1,11–13]. Next, the CdvB class homolog binds to CdvA, based on the BWI-BWH interaction. Then, all CdvB homologs can form a higher order structure, based on the Snf7 domain. This higher-order structure is probably ring-like[1,6]. Thus, initially there is only CdvA at the division site, followed by a mixture of CdvA and CdvB, which is then completed by CdvB1/2 and CdvB3. At this stage, a ring has been assembled at the division site, composed of all CdvB homologs and tethered to the membrane by CdvA.

There are only two ways to interfere with this structure: either by interaction with the MIM2 domain of the CdvB class homolog or with the MIM2 domain of the CdvB1/2 class homolog. Next, the MIT domain of CdvC gets involved. However, as CdvC can bind to both homologs, the question is which version has the higher affinity. Based on sequence and structural data we hypothesised that the CdvB class homolog has a significantly higher affinity to CdvC than the CdvB1/2 class homolog. Interaction is known to take place at a short motif of the MIM2

domain[1,20], governed by interactions with the MIT domain based on multiple prolines. The MIM2 domain of CdvB1/2 class proteins showed consistently less prolines in that motif than the MIM2 domain of CdvB class proteins (Suppl. Figure 1). Therefore, as the CdvB1/2 class proteins have less prolines to interact with CdvC, we expected the affinity of the CdvB class homologs to be higher than the one of the CdvB1/2 class versions. Thus, we concluded that in the next step CdvC binds preferably to the CdvB class homologs in the ring at the division site. Again, this is fits to experimental data[5].

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, taken together, the ring of all 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]. As CdvB3 class homologs have no domains to bind to CdvA or CdvC, they should also dissociate from the division site. How this mechanism results in membrane constriction and cell division can only be speculated, and we will state possible explanations that hopefully are relatively easy to test experimentally in the Discussion.

In Thaumarchaeota, a classification of CdvB homologs as in Crenarchaeota was not possible. Phylogenetic analysis showed no clear clusters (Figure 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 remained unclear. Nonetheless, we were able to deduce a possible functional mechanism based on the domain compositions (Figure 1C,D).

Same as in Crenarchaeota, the inferred 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), or, 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 work as a linker between CdvA and CdvB, and as a disassembler of CdvB at the same time.

Therefore, we suggest the first scenario, which also matches the diffuse distribution of CdvB in fluorescence microscopy experiments[7]. Driven by the high 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*)[10], all CdvB homologs contained either MIM1 or MIM2 domains, both of which 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.[17], whose results suggest that indeed CdvC in Asgard archaea can disassemble CdvB filaments, supporting the idea of a functional Cdv-based mechanism in this super-phylum.

However, to form a cell division system, this potential constriction mechanism must be transferred to the membrane. But there was no CdvA found 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 et al.[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[19] they 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 clustered 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[19]. Thus, it is plausible to assume that the membrane binding of the Vps2/24/46 ESCRT-III group is related to the CdvBa1 group of Asgard archaea. Supporting this idea, the region was furthermore predicted to form alpha helices, while in the CdvBa2 subgroup a conserved proline prohibits this (Figure 3). Therefore, this finding strongly

supports the idea of Caspi et al. 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 than any other Cdv-based mechanism.

Based on these potential interactions, we derived a possible mechanism of cell division in Asgard archaea (Figure 3B): 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 by MIM1-MIT interaction or MIM2-MIT interaction. As differences in affinity between different domains can only be evaluated experimentally, we did not try to hypothesise an order of disassembly like in Crenarchaeota. Furthermore, while in *Cand. P. synthrophicum* the CdvBa1 paralogs seem to possess both MIM1 and MIM2, while 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 highly different Cdv-based mechanisms within the archaeal kingdom, we asked which evolutionary processes might have given rise to the variations and how a simple Cdv machinery in a common ancestor might have looked like. In particular, we wanted to find out to what degree different Cdv-based mechanisms are comparable, and which components of the machineries might be utilised for comparative studies to the eukaryotic ESCRT system.

Cdv-based mechanisms of cell division in TACK archaea are poorly, mechanisms in Asgard archaea highly comparable to ESCRT.

In the calculated phylogeny of CdvB homologs, the included ESCRT-III proteins from human and yeast clustered with Asgard archaea homologs (Figure 2). A similar phylogeny of CdvC/Vps4 homologs gave the same result (Suppl. Figure 2). This is consistent with the

findings of Lu et al.[17] and supports the idea of Zaremba-Niedzwiedzka et al.[16] that all Eukaryotes are descendants of Asgard archaea, rather than having diverged from an early common ancestor of all archaeal organisms. Furthermore, our results confirm the idea of Lu et al.[17] that the divergence of ESCRT-III proteins into the two groups Vps2/24/46 and Vps20/32/60 was already existent in Asgard archaea and thus, that interactions and functional differences between CdvBa1 and CdvBa2 which might be found in future experiments, might be compared to ESCRT-III proteins.

The structure of the phylogeny further indicates that the duplication event(s) which gave rise to these two groups did occur after Asgard archaea diverged from the ancestor shared with TACK archaea (Figure 4). Similarly, the duplication events giving rise to the Crenarchaeota groups CdvB, CdvB1/2 and CdvB3, and the duplications in Thaumarchaeota, probably took place after this speciation event. Thus, the 'original' Cdv machinery, from which today's mechanisms descend, likely consisted of only a single CdvB protein. Given that CdvA proteins were only found in TACK archaea, it is likely that this first version of CdvB possessed a Snf7 domain and a MIM2 domain, but no BWH domain. Therefore, while all domains responsible for interactions between CdvB homologs and CdvC were probably already existent in the common ancestor and thus this interaction should be comparable between organisms of different super-phyla, all interactions between multiple CdvB homologs developed independently. Thus, the formation of higher-order structures based on Snf7 polymerisation is probably not comparable, and possible different functions of CdvBa1 and CdvBa2 have no evolutionary basis for comparison to different functions of CdvB, CdvB1/2 and CdvB3. This finding strongly suggests that Cdv-based mechanisms of cell division in TACK archaea constitute only a very limited model for the eukaryotic ESCRT system. The often-expressed hope that the relatively easy to study Cdv system of *S. acidocaldarius* can eventually be researched as a simpler version of ESCRT is therefore inept. Instead, the much harder to study Cdv-based mechanisms in Asgard archaea promise to provide very useful analogies, as they show a close similarity to ESCRT in protein composition and domain architecture, and their CdvB paralogs have the same evolutionary origin as ESCRT-III paralogs of eukaryotes. Giving respect to the very hard task of studying Asgard archaea *in vivo*, the most promising candidates for the long-wanted Cdv-based simplified model of ESCRT are *in vitro* experiments with Cdv proteins of Asgard archaea.

Discussion

The mechanism we derived for Crenarchaeota fits nicely to a recent study of 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. Their 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 predicts. While 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 lnvolve labeled CdvA and CdvC proteins, it wiould be interesting to see whether our model still holds true in studies investigating the full machinery.

In their paper, 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, the ring would shrink to the smaller preferred diameter of CdvB1/2 class proteins. While this mechanism could explain the initial constriction of the ring, it cannot tell why the CdvB1/2 ring constricts further until the membranes fuse and division is complete. Thus, there must be a different mechanism. Multiple ideas have been proposed for the final fusion/fission step in ESCRT-based processes[1,21,22], however, comparison of these to Crenarchaeota must be done with extreme caution, as we showed that the two systems are only poorly comparable. Based on our model, it may simply be the ring disassembly itself that leads to constriction. Like removing a pearl from a necklace, disassembly of a single CdvB1/2 protein might decrease the diameter by shortening the filament.,

As CdvA was not labelled in their experiment, 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. Again, this can be explained by our model. Initially, CdvC mostly disassembles CdvB class proteins due to the 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 are still some CdvB class proteins left. Thus, even during the final

constriction stage, there are always some remaining membrane anchors, as CdvB class proteins are never entirely removed.

There is, however, one result from another study which cannot be explained by our model. Based on 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.

Up to now, there are no studies investigating the role of CdvB3 class proteins. As we predicted only one possible interaction, namely binding to CdvB homologs, putative functions of the protein are limited. One straightforward idea might be that CdvB3 can temporarily stabilise the ring of CdvB homologs while or after CdvC disassembles a component. However, to lead to constriction, CdvB3 would eventually have to be removed as well. Simple knockout experiments should provide first insights into the role of CdvB3.

In this paper, we stated two possible scenarios for Cdv-based cell division in Thaumarchaeota. While we favoured one of them, only experimental work can reveal which one comes closer to reality. As in scenario 2 CdvB homologs are involved, while in scenario 1 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 many 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. For a start, 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 to verify the central idea of our theory. Second, as we could not quantify differences between CdvBa1 and CdvBa2 paralogs in their affinity to CdvC, this would be an essential information to evolve the theory further.

While the exact mechanism of membrane binding in Asgard archaea remains unknown, it will be clearly different from TACK archaea. However, they both developed from a common ancestor, so the question arises why the mechanisms diverged. This might be related to one of the main questions of eukaryogenesis: One of the key characteristics of Archaea is that their membranes are composed of different lipids than Bacteria and Eukaryotes[23]. Thus, if Eukarya indeed originate from Archaea, the lipid composition of their membranes must have changed at some point. When and how this occurred is not known, and the question is typically referred to as 'lipid transition' or 'lipid divide'[10,24]. The major difference between archaeal and eukaryotic membranes is that the archaeal version is constructed of isoprene-based alkyl chains that are linked to glycerol-1-phosphate by an ether bond, while the eukaryotic version is based on fatty acids that are linked to glycerol-3-phosphate by an ester bond[24]. Interestingly, a study investigating metagenomes of Asgard archaea[24] found that they probably miss a gene (geranylgeranylglyceryl phosphate synthase) needed to synthesise glycerol-1-phosphate. Instead, the study suggested that they might be able to synthetize an intermediate form, which combines isoprene-based alkyl chains with glycerol-3-phosphate via ether bonds[24]. When the paper was published, *Cand. P. syntrophicum* was not cultivated and thus not included in the analysis, but our analysis of the data from this later study fits to the theory. Not only does *Cand. P. syntrophicum* also miss the geranylgeranylglyceryl phosphate synthase gene but possess all genes theorised to be involved in synthesizing the intermediate form, but membrane analysis also indicated that the membrane-lipids indeed possess ether bonds[10]. Thus, it is possible that Asgard archaea possess unique membrane components which are different from other archaea. If this is true, this could be the reason behind the absence of CdvA in Asgard archaea and the occurrence of the ANCHR motif instead. If CdvA can only bind to 'normal' archaeal membranes, an alternative method had to evolve in Asgard archaea.

Furthermore, the changing membrane composition might explain how this function was gained. In some CdvB proteins of TACK archaea a short alpha helix relatively similar to the CdvBa2

version is present at the N-terminus[1], indicating that it might have already been a part of the 'original' protein in the common ancestor. However, in TACK archaea it cannot bind the membrane[13] due to some differences in sequence that brake the helix and make it less amphipathic[1]. However, if there are just a few amino acids different between the non-membrane binding version and the membrane binding version (compare Figure 3), the question arises why the mutations of CdvBa1 proteins allowing membrane binding did only occur in Asgard archaea and not in TACK archaea. The answer might be that membrane binding of the ANCHR motif does not work with classical archaeal membranes but only with eukaryotic-like membranes. Thus, it will be very interesting to test the membrane-binding capacities of Asgard CdvBa1 proteins *in vitro* with different types of membrane.

Taken together, this would allow an evolutionary scenario where CdvA was existent in the 'first' Cdv-based machinery in the common ancestor of TACK and Asgard archaea, and then was lost in Asgard archaea caused by the lipid transition. Instead, in Asgard archaea a short alpha helix at the N-terminus was developed, allowing it to directly attach to the altered membranes. Thus, the question of how the 'original' machinery did bind to the membrane would be solved, resulting in a complete model of Cdv/ESCRT evolution that might answer the fundamental question of how cell-division developed in eukaryotic cells.

In this paper, we described the domain compositions of archaeal Cdv system proteins. Based on this, we derived mechanistic models for Cdv machineries in Crenarchaeota and Thaumarchaeota, which explain their functional differences between organisms as found in experiments. We strongly supported the idea of Caspi et al.[1] and Lu et al.[17] that the Cdv machinery of Asgard archaea is closely related to the eukaryotic ESCRT machinery, making Asgard archaea a good candidate for comparative studies. Our findings also strengthened the hypothesis of Zaremba-Niedzwiedzka et al.[16] that all Eukarya are descendants of Asgard archaea. Finally, we found that Cdv systems of TACK archaea have only a limited potential to be used as model systems for ESCRT, because CdvB/ESCRT-III homologs duplicated after division into Asgard and TACK archaea. Thus, the complex interactions between different versions of CdvB/ESCRT-III proteins and their specific functional roles in the overall process most likely developed independently.

Acknowledgements

We thank Dina Grohmann 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.

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 developed 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[1,11–13], 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[13], so we classified this as domain, named BWH interaction site (BWI). Furthermore, inspired by Caspi et al.[1] and Lu et al.[17], we then compared the conserved regions in Asgard archaea to the N-terminal ANCHR domain[19] and the C-terminal MIM1 domain[18] 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.[20] and Samson et al.[13], 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 provide 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 prediction

Each protein was abstracted as a combination of domains. Based on literature, each domain was assigned a specific function (see Figure 1B and 3C). Then, mechanisms were introduced as a network of PPI states, each state indicating which proteins currently interact. Thus, each node of the network indicated a specific combination of currently interacting proteins. Edges in the network were indicated by the domain functions of the involved proteins that could change the PPI states. This means that each edge was either labelled by two proteins that did not currently interact but might do so, or two proteins that did interact together with a third protein that might dissolve this interaction. To generate a mechanism starting with a specific protein, the node representing no active PPIs was selected and all edges where the label did not include the selected starting protein were deleted. This made sure that the first step of the mechanism included the starting protein. Repeating this process recursively by deleting all edges except the ones labelled by the currently interacting proteins led to a network of possible states,

representing the mechanism caused by the protein domain composition. These where then summarized and simplified to the mechanisms displayed in Figure 1.

Code availability

The code of the analysis scripts is available on request.

Supplemental Information

- SupplementaryFigures.pdf containing Supplementaey Figure 1 and 2.
- Supplementary Table 1.xlsx
- Supplementary Table 2.xlsx

References

1. Caspi, Y., and Dekker, C. (2018). Dividing the Archaeal Way: The Ancient Cdv Cell-Division Machinery. Front. Microbiol. *9*, 174.
2. Makarova, K.S., Yutin, N., Bell, S.D., and Koonin, E. V. (2010). Evolution of diverse cell division and vesicle formation systems in Archaea. Nat. Rev. Microbiol.
3. Lindås, A.-C., Karlsson, E.A., Lindgren, M.T., Ettema, T.J.G., and Bernander, R. (2008). A unique cell division machinery in the Archaea. Proc. Natl. Acad. Sci. U. S. A. *105*, 18942–6.
4. Samson, R.Y., Obita, T., Freund, S.M., Williams, R.L., and Bell, S.D. (2008). A role for the ESCRT system in cell division in archaea. Science (80-.).
5. Pulschen, A.A., Mutavchiev, D.R., Culley, S., Sebastian, K.N., Roubinet, J., Roubinet, M., Risa, G.T., van Wolferen, M., Roubinet, C., Schmidt, U., et al. (2020). Live Imaging of a Hyperthermophilic Archaeon Reveals Distinct Roles for Two ESCRT-III Homologs in Ensuring a Robust and Symmetric Division. Curr. Biol.
6. Risa, G.T., Hurtig, F., Bray, S., Hafner, A.E., Harker-Kirschneck, L., Faull, P., Davis, C., Papatziamou, D., Mutavchiev, D.R., Fan, C., et al. (2019). Proteasome-mediated protein degradation resets the cell division cycle and triggers ESCRT-III-mediated cytokinesis in an archaeon. bioRxiv.
7. Pelve, E.A., Lindås, A.C., Martens-Habbena, W., de la Torre, J.R., Stahl, D.A., and Bernander, R. (2011). Cdv-based cell division and cell cycle organization in the thaumarchaeon Nitrosopumilus maritimus. Mol. Microbiol.
8. Eme, L., Spang, A., Lombard, J., Stairs, C.W., and Ettema, T.J.G. (2017). Archaea and the origin of eukaryotes. Nat. Rev. Microbiol.
9. Adam, P.S., Borrel, G., Brochier-Armanet, C., and Gribaldo, S. (2017). The growing tree of Archaea: new perspectives on their diversity, evolution and ecology. ISME J. *11*, 2407–2425.
10. Imachi, H., Nobu, M.K., Nakahara, N., Morono, Y., Ogawara, M., Takaki, Y., Takano, Y., Uematsu, K., Ikuta, T., Ito, M., et al. (2020). Isolation of an archaeon at the prokaryote–eukaryote interface. Nature.

11. Moriscot, C., Gribaldo, S., Jault, J.M., Krupovic, M., Arnaud, J., Jamin, M., Schoehn, G., Forterre, P., Weissenhorn, W., and Renesto, P. (2011). Crenarchaeal CdvA forms double-helical filaments containing DNA and interacts with ESCRT-III-like CdvB. PLoS One.
12. Dobro, M.J., Samson, R.Y., Yu, Z., McCullough, J., Ding, H.J., Chong, P.L.-G., Bell, S.D., and Jensen, G.J. (2013). Electron cryotomography of ESCRT assemblies and dividing Sulfolobus cells suggests that spiraling filaments are involved in membrane scission. Mol. Biol. Cell *24*, 2319–27.
13. Samson, R.Y., Obita, T., Hodgson, B., Shaw, M.K., Chong, P.L.-G., Williams, R.L., and Bell, S.D. (2011). Molecular and structural basis of ESCRT-III recruitment to membranes during archaeal cell division. Mol. Cell *41*, 186–96.
14. Ng, K.-H., Srinivas, V., Srinivasan, R., and Balasubramanian, M. (2013). The Nitrosopumilus maritimus CdvB, but not FtsZ, assembles into polymers. Archaea *2013*, 104147.
15. Bernander, R., and Ettema, T.J. (2010). FtsZ-less cell division in archaea and bacteria. Curr. Opin. Microbiol. *13*, 747–52.
16. Zaremba-Niedzwiedzka, K., Caceres, E.F., Saw, J.H., Bäckström, D., Juzokaite, L., Vancaester, E., Seitz, K.W., Anantharaman, K., Starnawski, P., Kjeldsen, K.U., et al. (2017). Asgard archaea illuminate the origin of eukaryotic cellular complexity. Nature *541*, 353–358.
17. Lu, Z., Fu, T., Li, T., Liu, Y., Zhang, S., Li, J., Dai, J., Koonin, E. V., Li, G., Chu, H., et al. (2020). Coevolution of eukaryote-like vps4 and ESCRT-III subunits in the asgard archaea. MBio.
18. Sciskala, B., and Kölling, R. (2013). Interaction maps of the Saccharomyces cerevisiae ESCRT-III protein Snf7. Eukaryot. Cell.
19. Buchkovich, N.J., Henne, W.M., Tang, S., and Emr, S.D. (2013). Essential N-Terminal insertion motif anchors the ESCRT-III filament during MVB vesicle formation. Dev. Cell.
20. Kojima, R., Obita, T., Onoue, K., and Mizuguchi, M. (2016). Structural Fine-Tuning of MIT-Interacting Motif 2 (MIM2) and Allosteric Regulation of ESCRT-III by Vps4 in Yeast. J. Mol. Biol.
21. Franquelim, H.G., and Schwille, P. (2017). Revolving around constriction by ESCRT-III. Nat. Cell Biol.
22. Schöneberg, J., Lee, I.-H., Iwasa, J.H., and Hurley, J.H. (2017). Reverse-topology membrane scission by the ESCRT complexes. Nat. Rev. Mol. Cell Biol. *18*, 5.
23. Albers, S.V., and Meyer, B.H. (2011). The archaeal cell envelope. Nat. Rev. Microbiol.
24. Villanueva, L., Schouten, S., and Damsté, J.S.S. (2017). Phylogenomic analysis of lipid biosynthetic genes of Archaea shed light on the 'lipid divide.' Environ. Microbiol.
25. Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res.
26. Makarova, K.S., Wolf, Y.I., and Koonin, E. V. (2015). Archaeal clusters of orthologous genes (arCOGs): An update and application for analysis of shared features between thermococcales, methanococcales, and methanobacteriales. Life.
27. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res.
28. Agarwala, R., Barrett, T., Beck, J., Benson, D.A., Bollin, C., Bolton, E., Bourexis, D.,

Brister, J.R., Bryant, S.H., Canese, K., et al. (2016). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res.

29. Mount, D.W. (2008). Using BLOSUM in Sequence Alignments. Cold Spring Harb. Protoc. *2008*, pdb.top39-pdb.top39.
30. Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., et al. (2014). Pfam: The protein families database. Nucleic Acids Res.
31. Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.Y., El-Gebali, S., Fraser, M.I., et al. (2019). InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res.
32. Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., et al. (2014). InterProScan 5: Genome-scale protein function classification. Bioinformatics.
33. Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., et al. (2015). CDD: NCBI's conserved domain database. Nucleic Acids Res.
34. Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059–66.
35. Drozdetskiy, A., Cole, C., Procter, J., and Barton, G.J. (2015). JPred4: A protein secondary structure prediction server. Nucleic Acids Res.
36. Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., and Huelsenbeck, J.P. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. Syst. Biol. *61*, 539–542.

Figure legends

Figure 1. A: Cdv machinery domain compositions of selected archaea. CdvA homologs in r ed, 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-based system for cell division. In contrast, all investigated Asgard archaea encode CdvB and CdvC homologs, suggesting that the system is conserved and used. Matching to 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

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.

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.

Figure 3: 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 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

Figure 4: Evolutionary development of Cdv/ESCRT machineries based on phylogenetic trees (Figure 2 and Suppl. Figure 2) and maximum parsimony. The ‘original’ Cdv machinery probably consisted of a CdvB protein composed of Snf7 and MIM2 together with an already 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. Taken together, this evolutionary scenario indicates that interactions between CdvB and CdvC in archaea can be compared to the interactions of ESCRT-III and Vps4 in eukaryotes, as they evolved from a common ancestor. Differently, only interactions between CdvB paralogs in Asgard archaea are likely to be comparable to ESCRT-

III paralog interactions, because they show the same evolutionary origin, whilst CdvB paralogs in TACK archaea evolved independently. Thus, the long-held hope that Cdv-based systems of cell division in Crenarchaeota might be simple-to-study models of the complicated ESCRT system is inept, but Asgard archaea promise to provide very useful analogies.

Figure 1

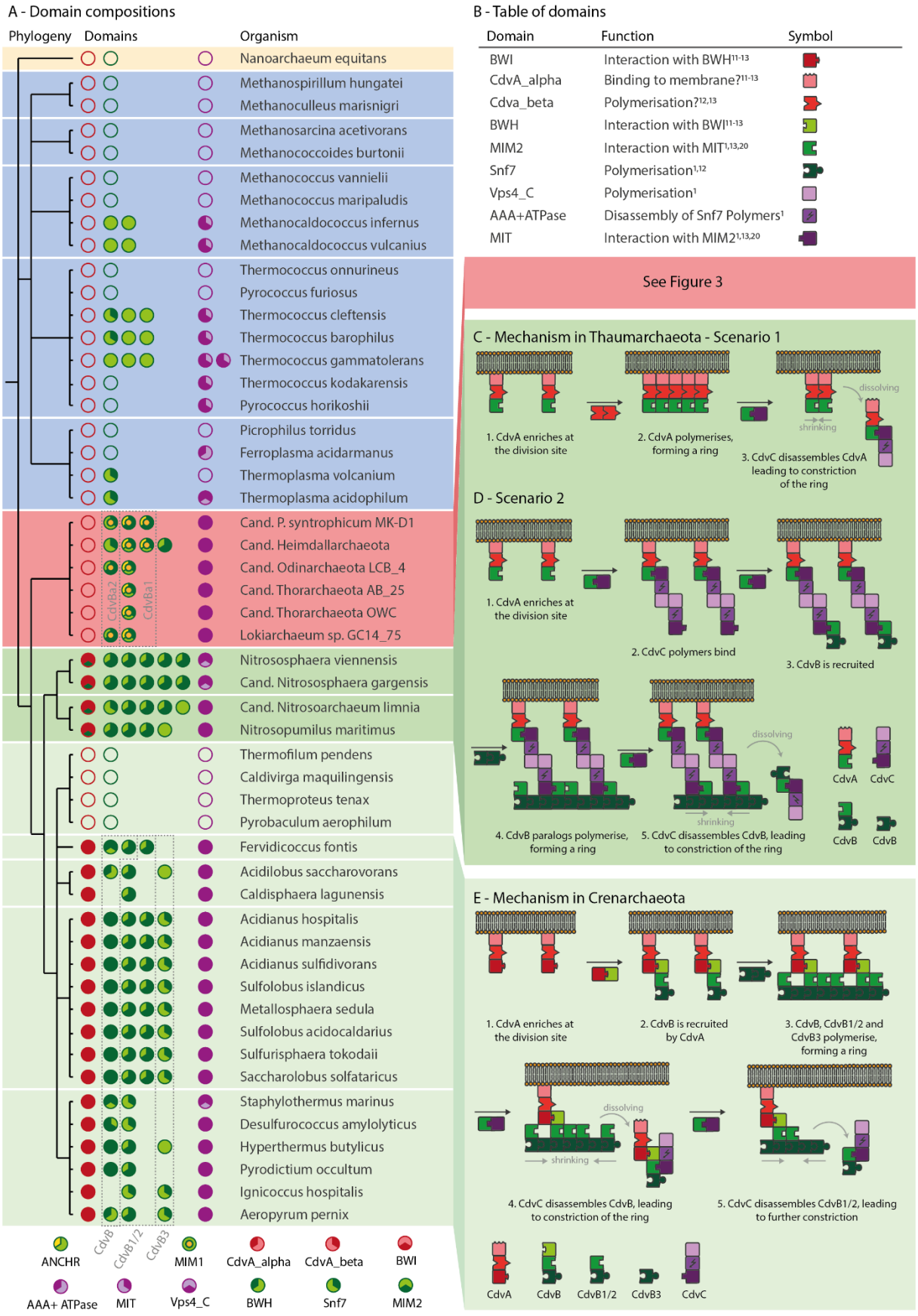

Figure 2

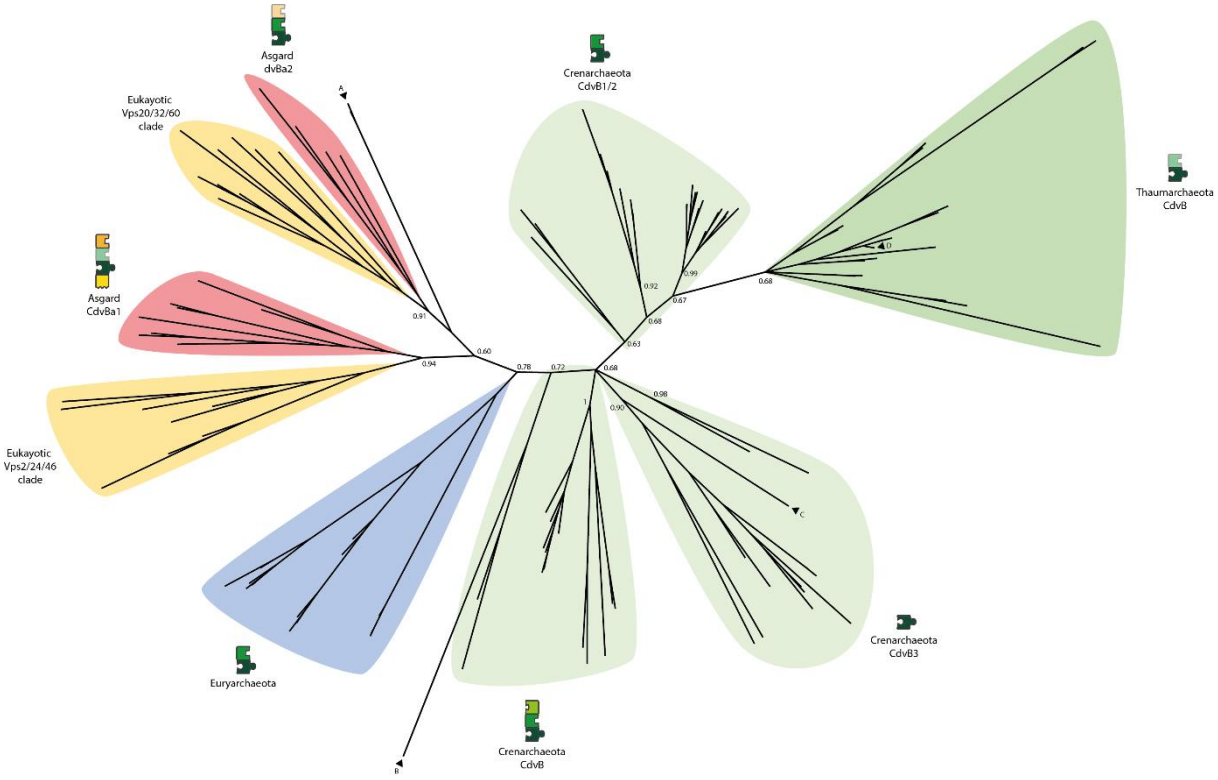

Figure 3

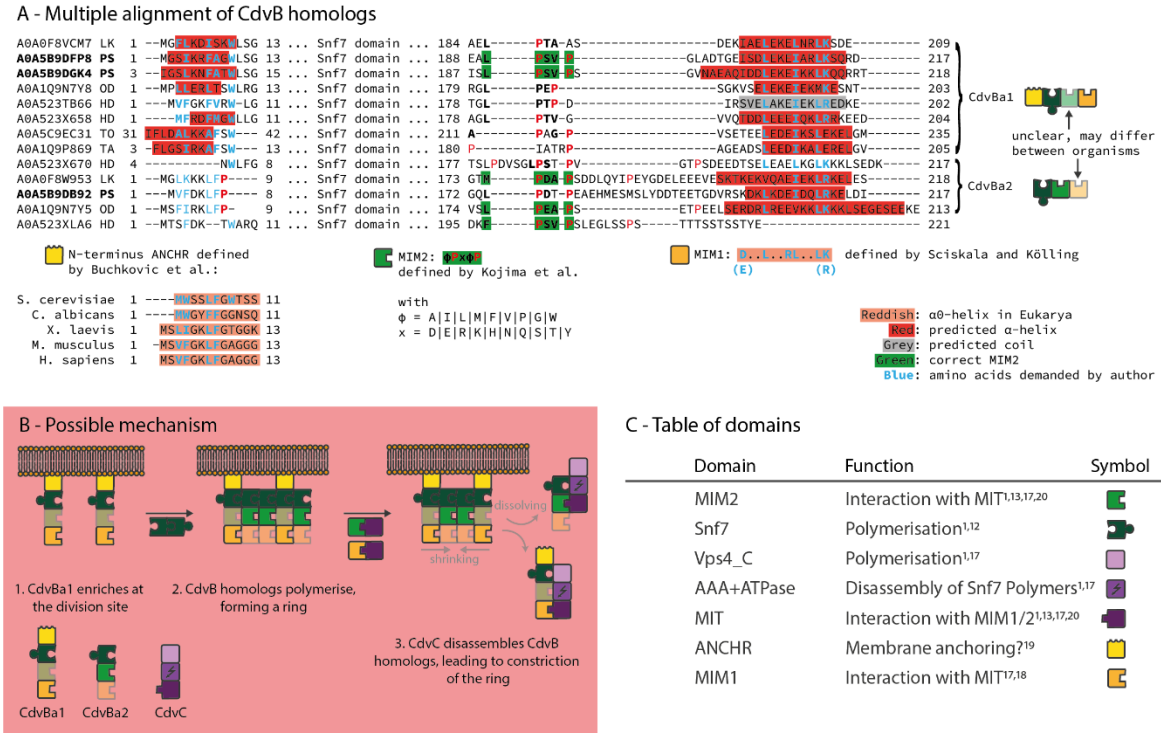

Figure 4