The natural history of Get3-like chaperones

Ákos Farkas1 | Evelina Ines De Laurentiis1 | Blanche Schwappach1,2

1Department of Molecular Biology, Göttingen University Medical Center, Göttingen, Germany
2Max-Planck Institute for Biophysical Chemistry, Göttingen, Germany

Correspondence
Blanche Schwappach, Department of Molecular Biology, Göttingen University Medical Center Humboldtallee, 23 D-37073 Göttingen, Germany.
Email: blanche.schwappach@med.uni-goettingen.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: SFB1190 P04

Get3 in yeast or TRC40 in mammals is an ATPase that, in eukaryotes, is a central element of the GET or TRC pathway involved in the targeting of tail-anchored proteins. Get3 has also been shown to possess chaperone holdase activity. A bioinformatic assessment was performed across all domains of life on functionally important regions of Get3 including the TRC40-insert and the hydrophobic groove essential for tail-anchored protein binding. We find that such a hydrophobic groove is much more common in bacterial Get3 homologs than previously appreciated based on a directed comparison of bacterial ArsA and yeast Get3. Furthermore, our analysis shows that the region containing the TRC40-insert varies in length and methionine content to an unexpected extent within eukaryotes and also between different phylogenetic groups. In fact, since the TRC40-insert is present in all domains of life, we suggest that its presence does not automatically predict a tail-anchored protein targeting function. This opens up a new perspective on the function of organellar Get3 homologs in plants which feature the TRC40-insert but have not been demonstrated to function in tail-anchored protein targeting. Our analysis also highlights a large diversity of the ways Get3 homologs dimerize. Thus, based on the structural features of Get3 homologs, these proteins may have an unexplored functional diversity in all domains of life.

KEYWORDS
bacteria, Chlorophyta, Embryophyta, endoplasmic reticulum, Get3p, molecular chaperone, Rhodophyta, tail-anchored protein

1 | INTRODUCTION

Tail-anchored (TA) proteins are a class of membrane proteins that contain a C-terminal hydrophobic transmembrane segment (TMS) and a functional N-terminal cytosolic domain.1,2 TA proteins are a diverse group of eukaryotic membrane proteins found among others in the secretory pathway,3 nuclear envelope,4 peroxisomes,5 mitochondria6 and in chloroplasts.7 They have a wide range of functions, such as assistance in vesicular trafficking,8 protein translocation9 and degradation10 of membrane proteins. The function of TA proteins has been shown to be essential in all domains of life and their transport to the correct biologically membrane, or protein targeting, needs to be efficient and accurate as targeting errors can have detrimental cellular effects. Additionally, TMSs are prone to aggregation and their spontaneous insertion into lipid bilayers may be slow in vivo. Therefore, in order to ensure efficient and organelle-specific insertion of TA proteins and to prevent the aggregation of TMSs in the cytoplasm, most studies to date suggest that the targeting and insertion of TA proteins involves one or more cytosolic factors.

The mechanism through which TA proteins are targeted and inserted is distinct from the co-translational signal recognition particle (SRP)-facilitated process by which most membrane proteins with N-terminal or internal signals are targeted. Indeed, because the C-terminal TMS of a TA protein emerges from the ribosome at the end of translation, TA proteins are targeted and inserted through post-translational processes.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2019 The Authors. Traffic published by John Wiley & Sons Ltd.

Traffic. 2019;20:311–324.
wileyonlinelibrary.com/journal/tra 311
translational mechanisms. One such pathway, the guided entry of TA proteins (GET), identified a little over 10 years ago, has been shown to mediate the proper delivery of several TA proteins in mammals,\textsuperscript{10,11} budding yeast\textsuperscript{12} and more recently in plants.\textsuperscript{13,14}

Extensive biochemical and structural studies performed over the last decade have characterized the targeting of TA proteins utilizing the yeast GET pathway (as reviewed in\textsuperscript{15}). Initially, a pre-targeting complex, consisting of a small glutamine-rich tetratricopeptide repeat containing protein S gt2, and Get4 and Get5 in yeast, or Bag6, SGTA, TRC35 and UBL4A in mammals, captures the TA protein following its release from the ribosome, then transfers it to the ATPase Get3 in budding yeast, or TRC40 in mammals.\textsuperscript{15–17} The TA-bound Get3/ TRC40 protects and delivers the TA protein to the ER membrane, where its receptor complex comprised of Get1 and Get2 in yeast or WRB and CAML in mammals stimulates its subsequent release into the membrane.\textsuperscript{18–21}

Despite the apparent complexity and necessity of the GET pathway to prevent aggregation of hydrophobic proteins, depletion of GET pathway components in budding yeast (Saccharomyces cerevisiae) and Arabidopsis thaliana is not lethal.\textsuperscript{12,14} Yet the functional importance of the GET pathway is highlighted by the fact that GET pathway deletion S. cerevisiae strains show increased heat and oxidative stress sensitivity,\textsuperscript{12,22} and the depletion of TRC40 is embryonically lethal in mice.\textsuperscript{23} Since Get3 was shown to possess chaperone holdase activity upon oxidation\textsuperscript{24} these phenotypes may to some extent reflect a chaperone activity not involved in targeting TA proteins during biogenesis.

## 2 | GET3 HOMOLOGS IN THE DIFFERENT DOMAINS OF LIFE

Some phylogenetic aspects of other GET pathway components have been recently discussed, in particular the evolutionary relationships between components of the pretargeting complex comprising Get4, Get5 and Bag6\textsuperscript{57} and the legacy of membrane protein biogenesis factors similar to bacterial Oxa1 that also include Get1.\textsuperscript{18} Here, we combine a review of the literature on Get3 structure and function with a comprehensive bioinformatic analysis of the structural elements of the protein involved in TA protein binding. This integration focuses on properties of Get3-like proteins in all domains of life that render the hydrophobic cage versatile and should be considered for both functions of these proteins.

A systematic search for Get3- and ArsA-homologous proteins in the KEGG and OrthoDB databases combined with further BLAST analysis yielded 2208 sequences (Supporting Information Table S1), from which 51 representative sequences were chosen to construct a phylogenetic tree (Figure 1). This analysis reveals a functionally unexplored diversity of Get3-like proteins (Table 1). Focusing on structural aspects of different homologs such as domain organization or the presence of sequence motifs and comparing them with known structures and functions of Get3 homologs, we would like to highlight that Get3-like chaperones from different kingdoms are more similar to each other than previously recognized based on a comparison of eukaryotic Get3 or TRC40 with prokaryotic ArsA.\textsuperscript{25} At the same time, they are remarkably diverse with respect to their modes of (pseudo) dimerization and structural features outside the well conserved ATPase domain.

## 3 | STRUCTURAL ORGANIZATION OF ARSA AND GET3 PROTEINS

A bacterial homolog of Get3, ArsA confers resistance to arsenite in Escherichia coli\textsuperscript{26} and shows high structural similarity to Get3 (Figure 2A,B). ArsA folds such that two highly similar domains in tandem form a metal binding site and two nucleotide binding sites (NBS) at their interface.\textsuperscript{27} The NBS is similar to those found in other members of the Signal recognition particle, MinD, BioD (SIMIBI) class of P-loop NTPases and contains conserved structural elements necessary for ATP hydrolysis including the P-loop, Switch I and II and the A-loop.\textsuperscript{28–30} The metal binding site involves three functionally essential cysteine residues, however, these residues are not conserved in eukaryotic Get3 homologs\textsuperscript{27} (Figure 2A, ball-and-stick model residues).

Unlike bacterial ArsA, Get3 in budding yeast (ScGet3) and other fungi and animals has a single Get3-homology domain. Two ScGet3 monomers assemble into rotationally symmetrical homodimers to form a structure analogous to the arrangement of the two domains found in ArsA (Figure 2B). In ArsA, a short helix involved in coordinat- ing the metal ion (orange in Figure 2A) folds into a groove, whereas the same region forms extended helices in Get3 (helix 7, 9; orange in Figure 2B), and also contains an additional stretch of amino acids dubbed the TRC40-insert.\textsuperscript{25} Thus, a large hydrophobic surface is created (Figure 2B, bottom row) allowing Get3 to accommodate the TMSs of TA proteins. At the same time, the helix contained within the TRC40-insert (helix 8, not visible in the structure) is thought to act as a lid that closes on captured TMSs, thus shielding them from the solvent.\textsuperscript{31}

Get3 homologs with the ability to bind TA proteins have also been found in archaea, and one out of the four archaean homologs studied so far could deliver captured substrates to the membrane.\textsuperscript{32,33} In bacteria, the only currently known Get3 homologs with a hydrophobic groove belong to photosynthetic bacteria and they also have an $\alpha$-crystallin domain at the C-terminus (red in Figure 2C).\textsuperscript{34} $\alpha$-crystallin domains are key components of heat shock proteins and are essential for their chaperone function.\textsuperscript{35} Although such Get3 homologs are also found in land plants,\textsuperscript{34} their function remains unknown. Moreover, land plants, Chlorophytes and red algae have been proposed or shown to have several Get3 homologs without an $\alpha$-crystallin domain as well, some of them in chloroplasts and mitochondria.\textsuperscript{14}

## 4 | CONSERVATION OF HELIX 8, THE "LID" CLOSING THE HYDROPHOBIC GROOVE

Get3 has several hydrophobic residues necessary for its interaction with TA proteins, and they mostly converge on the C-terminal portion
Eukaryotic cytosolic Get3 homologs

Organellar Get3 homologs in Archaeplastida (without α-crystallin domain)

Bacterial Get3 homologs with TRC40-insert and two domains in tandem

Archaeal Get3 homologs with TRC40-insert and CxxC motif

Bacterial and archaeal Get3 homologs with TRC40-insert

Bacterial and archaeal ArsA and ArsA-like proteins

FIGURE 1  Maximum likelihood rooted phylogenetic tree of three representative sequences of each group of Get3 homologs as defined in Table 1. Percentage of trees in which the sequences clustered together after applying 1000 bootstraps are indicated at nodes if the value is higher than 70%. Scale bar indicates number of substitutions per site.
| Species                  | Protein               | Identified sequences in group | Localization | Hydrophobic groove/TRC40-insert | CxC motif | CxxC motif | α-crystallin domain | Get3 domains | Get1/2/4 binding residues | Other features/Comments                                                                 |
|-------------------------|-----------------------|------------------------------|--------------|---------------------------------|-----------|------------|---------------------|--------------|--------------------------|-----------------------------------------------------------------------------------------|
| Eukaryotes              | Animals, Fungi        | Get3/TRC40                   | 629          | Cytoplasmic                     | Yes       | Yes        | No                  | Single       | Yes          | In 597 out of 629 sequences                                                                 |
| Land plants, Chlorophytes | Get3/TRC40            | 87                           | Cytoplasmic  | Yes                             | Yes       | No         | No                  | Single       | Yes          | C-terminal charged patch                                                                 |
| Red algae               | Group 1               | 4                            | Cytoplasmic  | Yes                             | Frequent  | No         | No                  | Double       | Yes          | In three out of four sequences                                                                 |
| Red algae, Chlorophytes | Group 2               | 10                           | Organellar   | Yes                             | No        | No         | No                  | Double       | No          |                                                                                           |
| Land plants, Red algae  | Group 3               | 90                           | Organellar   | Yes                             | No        | Yes        | No                  | Single       | No          |                                                                                           |
| Chlorophytes            | Group 4               | 8                            | Organellar   | Yes                             | No        | No         | No                  | Single       | No          |                                                                                           |
| Land plants             | Group 5               | 48                           | Plastidial   | Yes                             | No        | No         | Yes                 | Single       | No          |                                                                                           |
| Bacteria                | Mainly Actinobacteria, Chloroflexi, Chlorobi Cyanobacteria, Firmicutes, some Acidobacteria, Aquificae, Bacteroidetes, Fusobacteria, Spirochaetes, Proteobacteria | Group 6 | 301 | Cytoplasmic                     | Yes       | No         | No                  | Single       | No          | Rarely: nucleotide binding site missing (such homologs excluded from current analysis)                                                           |
| Cyanobacteria           | Group 7               | 29                           | Cytoplasmic  | Yes                             | No        | No         | No                  | Double       | No          |                                                                                           |
| Myxococcales, some Protoeobacteria | Group 8               | 17                           | Cytoplasmic  | Yes                             | No        | Yes        | No                  | Double       | No          |                                                                                           |
| Mainly Actinobacteria, Firmicutes, Proteobacteria, some Aquificae, Spirochaetes, Synergistetes | Group 9               | 131                          | Cytoplasmic  | Yes                             | No        | No         | No                  | Single       | No          |                                                                                           |
| Some Firmicutes         | ArsA-like             | 18                           | Cytoplasmic  | No                              | No        | No         | No                  | Single       | No          | Part of arsenite resistance operon                                                                                                             |
| Mainly Actinobacteria, Bacteroidetes, Proteobacteria, Firmicutes, Planctomycetes | ArsA                  | 432                          | Cytoplasmic  | No                              | No        | No         | No                  | Double       | No          | Part of arsenite resistance operon                                                                                                             |
| Archaea                 | Euryarcheota          | ArsA                         | 27           | Cytoplasmic                     | No        | No         | No                  | Double       | No          | Some of the Get1/2/4 binding residues are                                                                                           |
| Euryarcheota, Crenarcheota | Group 10              | 155                          | Cytoplasmic  | Yes                             | No        | No         | No                  | Single       | No          |                                                                                           |
of helix 7 and the short helix 8 following it. Recently, it has emerged that helix 8 is needed to ensure an efficient transfer of substrates from upstream components to Get3, but it has no major effect on the dissociation of substrates already captured by Get3. In the structures of eukaryotic Get3 homologs, the region around helix 8 is poorly defined because of its high flexibility and is heavily influenced by the overall conformation of the protein. Although helix 8 forms a helix separate from helix 7 in fungal Get3 Structures (Figure 3A), these two helices appear to line up or even merge completely in structures of archael homologs of Get3. The Get3 homolog of the archaeon *Methanocaldococcus jannaschii* (*Mj* Get3) exists either as a dimer, similar to *S. cerevisiae* Get3 (*Sc* Get3) or in a tetrameric form, a dimer of dimers, the assembly of which is mediated by the region corresponding to helix 8 in *Sc* Get3 (Figure 3B). Although *Sc* Get3 and its human homolog TRC40 both form tetramers under specific conditions, these tetrameric structures remain structurally unsolved and the role of helix 8 in their assembly also remains unknown. Nevertheless, it is intriguing that based on secondary structure predictions, the regions corresponding to helix 7 and 8 of *Sc* Get3 would be expected to form a single helix as seen in *Mj* Get3 (Figure 3A), yet whether this region can indeed assume two distinct conformations remains to be seen.

Although the region linking helix 7 and 9 appears to be moderately conserved, especially at helix 8 in eukaryotes, its length varies considerably within and between phylogenetic groups (Figure 3C). Indeed, although the average length of the stretch homologous to the linker between helix 7 and 9 in *Sc* Get3 is approximately 21 to 22 amino acids in eukaryotes, there are notable exceptions as well. For instance, *Sc* Get3 only has 15 amino acids in this region while the archael *Mj* Get3 has 25, showing that from a functional perspective, substantial variation is allowed in this region. Interestingly, unlike in bacterial ArsA (Figure 3A), the length of this region in bacterial Get3 homologs with an *α*-crystallin domain is comparable to that observed in eukaryotes. For example, while the region in Firmicutes is often 21 amino acids long, just like in animals, in Cyanobacteria it is as long as in *Mj* Get3 and in many Chlorobi bacteria almost as short as in *Sc* Get3. However, whether this indicates any functional similarity is not known.

Besides the length of the linker between helix 7 and 9, its amino acid composition also shows variation within and between phyla (Figure 3D). It has been suggested that the methionine-rich nature of the hydrophobic groove is important for the accommodation of the TMS. This hypothesis is further strengthened by the analogy with SRP, where the methionine-rich M domain of Srp54 is essential for binding the signal peptide. In Get3 helix 8, there are two and three methionine residues in *Sc* Get3 and human TRC40, respectively, and their combined loss in *Sc* Get3 leads to decreased substrate binding. Consistent with the idea that the presence of the methionine residues is related to the TA targeting function, in homologs not expected to be involved in TA protein targeting (bacterial and plastidial-mitochondrial Get3 homologs without an *α*-crystallin domain in land plants), there is mostly no or just a single methionine in the corresponding region. However, looking at Get3 homologs known to
bind or target TA proteins, it becomes clear that although eukaryotic and archaeal homologs tend to have at least one or more methionine residues in this stretch, there are several species without any as well (Figure 3D). Taken together, although helix 8 may have become enriched in methionine in certain species to support TA protein targeting, the presence of methionine residues does not seem to be a requirement for helix 8 to fulfill its function.

5 | HELICES LINING THE HYDROPHOBIC GROOVE

One of the defining features of Get3 with respect to bacterial ArsA is the presence of the TRC40-insert, which corresponds to helix 8 in ScGet3 and the amino acids linking it to helix 9 (Figure 4A). The TRC40 insert with an extended helix 7 and 9, together with helices 4, 5 and 6 creates a hydrophobic area so that TMSs can be accommodated and shielded from solvents in the resulting groove. Mutational studies have revealed that some of the hydrophobic residues of helix 7 and 8 are important for substrate binding by Get3. Interestingly, while the residues that show the strongest effect in mutational studies of ScGet3 are not conserved in bacterial ArsA and Get3, and mutations of hydrophobic residues in this helix mostly affect the ATPase activity of Get3 but not substrate binding. Since one of the cysteines involved in coordinating the metal ion is close to the N-terminal part of this helix, its presence, coupled with the lack of the Get3-/TRC40-insert, is expected to be a strong indicative feature of ArsA homologs (Figure 4B). Indeed, most such ArsA homologs in our analysis are highly similar to E. coli ArsA and have two domains in tandem. However, some Firmicutes bacteria seem to be unique in that they possess two copies of such ArsA homologs, but each with only a single domain instead of two (Table 1). In this case, one of them is similar to the first domain of E. coli ArsA, and the other is similar to the second. However, because there is no obvious feature that would mediate dimerization, it is unknown whether they actually do form dimers and function as a bona fide ArsA in vivo.

Another special feature in helix 6 is found in plastidial and mitochondrial Get3 homologs without an α-crystallin domain in land plants, Chlorophyta and red algae, that is, the Archaeplastida clade. Besides the Get3 homologs already shown to localize to the chloroplast and mitochondria, similar organelar Get3 homologs are predicted to exist in other groups within the Archaeplastida clade as well (Table S2). In spite of overall sequence similarity to ScGet3, many of these homologs have several proline residues at the N-terminus of helix 6 (Figure 4B), the relevance of which is currently unknown. Furthermore, such homologs uniformly lack the CxC motif found on the beta strand following helix 9, a feature strongly, although not

![Figure 2](image_url)
Prediction

D. gibsoniae (Bacteria)
M. jannaschii (Archaea)
S. cerevisiae (Fungi)
K. dejecticola (Fungi)
H. sapiens (Mammals)
C. elegans (Nematoda)
A. thaliana, cytoplasmic (Plants)
A. thaliana, mitochondrial (Plants)

FIGURE 3  Legend on next column.
universally conserved among eukaryotic cytoplasmic Get3 homologs (Figure 4C). Considering that these homologs also lack key residues required for binding Get1, Get2 and Get4 (Figure 4C, Table 1), it is clear that such organellar Get3 homologs should fulfill a related yet distinct function compared to cytoplasmic Get3 homologs.

Looking at the hydrophobic groove as a whole, its methionine-rich nature has been thought to be a feature related to the TA protein targeting function of Get3.25 As stated above, the presence of methionine residues in helix 8 is probably not a prerequisite for TA protein targeting. However, counting all the methionine residues that could potentially flank the hydrophobic groove (from helix 4 to helix 9), it becomes clear that despite considerable variety, all fungi have at least four methionine residues in this region (ScGet3 has six), and most vertebrates have three times as many (Figure 4D). On the other hand, many bacterial Get3 homologs with an α-crystallin domain also have multiple methionine residues (Table S3), with several Firmicutes homologs having as many as eight (Figure 4D). As an exception, homologs in Actinobacteria tend to have fewer or no methionine residues at all (Figure 4D). Taken together, the fact that many bacterial homologs have as many methionine residues as some fungi do, and that there has been no indication so far that these homologs target TA proteins in bacteria, it is likely that the methionine-rich nature of the hydrophobic groove had already been present before the TA protein targeting function of Get3 was acquired. Then, as eukaryotic Get3 became more and more specialized to target TA proteins, it may have acquired further methionine residues in the groove to facilitate the binding of TMSs.

6 | GET3 HOMOLOGS IN THE EUKARYOTIC GROUP ARCHAEPLASTIDA

In yeast and most other eukaryotes, Get3 functions as a rotationally symmetrical homodimer because dimerization is necessary for both its ATPase activity and the formation of the TMS binding hydrophobic groove.22,25 In fungi and mammals, a conserved CxxC motif in each subunit aligns to coordinate a zinc ion (Figure 5A), which is necessary for dimer formation.22,25 Bacterial ArsA homologs are very similar structurally, but the two halves of the dimer are produced as two domains in tandem, just like bacterial ArsA, and they similarly lack the CxxC motif as well (Figure 5A and C).

Compared to cytoplasmic homologs, predicted organellar Get3 variants without an α-crystallin domain in Archaeplastida display an even greater diversity. Namely, these proteins lack the CxxC motif and key Get1, Get2, Get4 binding residues and they often contain extra prolines in helix 6. Even so, they are still hypothesized to dimerize and use several different ways to achieve this (Figure 5C). In organellar homologs in land plants and red algae, a CxxC motif is present, and is likely used to form a dimer. However, red algae have other homologs as well, as do Chlorophytes, that contain two domains in a single protein. Intriguingly, additional homologs of Get3 can be found in Chlorophytes that have no apparent dimerization motif, which does not exclude the possibility that they still form dimers in unexpected ways.

It has to be noted that an organellar homolog from Chlamydomonas reinhardtii predicted here to be organellar has been previously proposed to be cytoplasmic.43 However, the protein is highly similar to other homologs in land plants that have been shown to be organellar,14 and homologous proteins from other Chlorophytes are consistently predicted to be organellar as well (Table S2). Therefore, the localization of these homologs in Chlorophytes remains uncertain for the moment. The picture is further complicated by the fact that some of the Get3 homologs in Archaeplastida are highly similar in sequence to other organellar homologs, yet are predicted to be

FIGURE 3 Comparison of the TRC40-insert between species. A, Known secondary structure of ScGet3 (top) compared with the predicted structure of the same region in different species (bottom, predicted helices marked with black frame). Hydrophobic residues shown in peach, aromatic residues in ochre, basic residues in blue, acidic residues in red, hydrophilic residues in green, proline and glycine in mauve, cysteine in yellow. B, Structure of M. jannaschii Get3 (PDB ID: 3UG6). Subunits are marked with cyan, magenta, orange and blue. The region homologous to the region between helix 7 and 9 in ScGet3 is shown in red. C, Distribution of the length of the region homologous to the sequence between helix 7 and 9 in ScGet3 among the sequences used for the current analysis. All bins containing at least 1% of the sequences are shown in the chart. D, Number of analyzed sequences: Bacteria—299; Archaea—376; Fungi—489; Animals—140; Land plants (cytoplasmic)—78; Land plants (organellar, excluding α-crystallin domain Get3 homologs)—87. D, Distribution of the number of methionine residues in the region homologous to the sequence between helix 7 and 9 in ScGet3 among the sequences used for the current analysis. All bins containing at least 1% of the sequences are shown in the chart. The number of sequences analyzed are as in C.
FIGURE 4

Legend on next column.
cytoplasmic because of a lack of a transit peptide (Table S2). This could either indicate a further cytoplasmic group of such homologs or simply reflect an inaccurate bioinformatic prediction of the N-terminus of the proteins based on genomic sequences.

Besides the above-mentioned homologs, land plants also have a plastidial Get3 homolog that is closer in similarity to cyanobacterial homologs than it is to eukaryotic ones (Figures 5C and 1 and Table 1).\(^\text{34}\) Accordingly, this is the only eukaryotic Get3 homolog currently known to have an α-crystallin domain at its C-terminus like the one seen in the structure of NostocGet3 (Figure 2C).

The fact that land plants have Get3 homologs in the chloroplast with hypothetically two different ways to dimerize (one with an α-crystallin domain, one with a CxxC-motif) raises the question of what advantage having two such close homologs may bring. A possibility would be that the different modes of dimerization allow the organism to regulate the activity or various functions of the protein. Nonetheless, as no study has been carried out on these proteins to date, their function remains elusive as of now.

### 7 | PREVIOUSLY UNNOTICED BACTERIAL GET3 HOMOLOGS

As mentioned above, several major groups of bacteria have a Get3 homolog with the TRC40-insert similar to NostocGet3 (Figure 2C), with an α-crystallin domain attached to the C-terminus and a nucleotide binding site present, which is missing in NostocGet3 (Figure 5D, Table 1). Although it is known that α-crystallin domains can mediate dimerization and act as a chaperone,\(^\text{35}\) it is not clear from the available structure of NostocGet3 whether or how it contributes to the stabilization of the Get3 dimer. The possibility that it may have a different function is supported by the fact that several groups of bacteria have Get3 homologs with the TRC40-insert but no α-crystallin domain (Figure 5D and Table 1). These are highly similar to archaeal Get3 homologs lacking a CxxC motif, at least one of which has been demonstrated to be able to form dimers and bind TA proteins.\(^\text{33}\) Therefore, it is highly likely that they can also dimerize and bind hydrophobic sequences.

Furthermore, uniquely among bacteria, Cyanobacteria, Myxococcales species and some further Proteobacteria contain Get3 homologs with two domains in tandem where both domains contain a TRC40-insert (Figure 5D). Except for Cyanobacteria, they also have the CxxC motif, which makes these homologs unique not only among bacteria but in all domains of life.

It is currently unclear what the functions of these Get3 homologs are. Taking into consideration that all of the above-mentioned bacterial homologs have the TRC40-insert, current theory would predict their involvement in TA protein biogenesis. Since they have not been characterized yet, we can only rely on predictions. If they indeed insert TA proteins into the membrane, one would expect that bacterial species with more TA proteins would be more likely to have a TRC40-insert containing Get3 homolog than species that have fewer TA proteins. Indeed, Proteobacteria and Cyanobacteria with Get3 homologs arranged as two domains in tandem tend to have more predicted TA proteins than other bacteria (empty circles in Figure 5E and Table S4). However, comparing the abundance of predicted TA proteins between different bacterial species and the presence or absence of other Get3 homologs with a TRC40-insert reveals no correlation (Figure 5E and Table S4). Furthermore, it has been shown that other chaperones are responsible for TA protein targeting in bacteria, at least in E. coli.\(^\text{44}\)

Considering that ScGet3 can act as a more general chaperone under specific conditions,\(^\text{38}\) that a large group of bacterial Get3 homologs have an α-crystallin domain with expected chaperone activity, and that all the TRC40-insert-containing homologs mentioned above are predicted to have a hydrophobic groove, it is possible that these homologs act more as general chaperones than TA protein targeting factors. From this perspective, the TA protein targeting activity of cytoplasmic eukaryotic Get3 homologs could represent an adaptation of an ancient, more general chaperoning function. On the same note, it would be interesting to know whether the above-mentioned Get3 homologs in chloroplasts and mitochondria have a similar function to those found in bacteria or represent a third group of Get3-like chaperones with unexpected functions. As summarized in Table 1, it is clear that Get3-like chaperones are widespread and structurally diverse and much remains to be discovered about the dynamic structure and function of these proteins.

### 8 | MATERIALS AND METHODS

#### 8.1 | Retrieval and processing of sequences and structures

Identifiers of Get3 and ArsA homologs were retrieved from KEGG Database (https://www.genome.jp/kegg/) and OrthoDB\(^\text{45}\) [https://}
Identifiers of Get3 homologs with an α-crystallin domain in land plants were retrieved using a blast search in land plants using the sequence of Nostoc Get3. Sequences were retrieved from Uniprot (www.uniprot.org) based on the identifiers collected from the databases and the blast search and aligned using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/). Alignment was visualized and manually adjusted using Jalview.46 Incomplete sequences were filtered out based on missing major regions compared to other homologs.

**FIGURE 5** Get3 homologs use various strategies to form dimers. A, Comparison of the sequence adjacent to the CxxC motif in ScGet3 and homologs from other organisms. B, Consensus sequence and secondary structure prediction of the charged C-terminal helix found in cytoplasmic Get3 homologs in land plants. C, Graphical representation of main structural features of land plant (LP), chlorophyte (C) and red algal (R) Get3 homologs. D, Graphical representation of main structural features of bacterial Get3 homologs. E, Comparison of the presence or absence of a TRC40-insert containing Get3 homolog in bacterial species with the number of predicted TA proteins in the given species. Empty circles represent Proteobacteria and Cyanobacteria with TRC40-insert containing Get3 homologs arranged as two domains in a single polypeptide.
Secondary structure predictions were carried out using JPred4 (http://www.compbio.dundee.ac.uk/jpred4/). Logos of consensus sequences were visualized using WebLogo 3. Structures were retrieved from RCSB (https://www.rcsb.org/) and visualized using VMD (http://www.ks.uiuc.edu/Research/vmd/).

8.2 | Construction of the phylogenetic tree

All analyses related to the phylogenetic tree were carried out in Mega X. Sequences were aligned using the MUSCLE algorithm with default settings and manually adjusted when necessary. The evolutionary history was inferred by using the Maximum Likelihood method and Whelan and Goldman + Freq. model. The tree with the highest log likelihood (~46,225.64) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches if they clustered together in more than 70% of the trees. Initial trees for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. A discrete Gamma distribution was used to model evolutionary rate differences among sites (five categories [+G, parameter = 1.3658]). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This analysis involved 51 amino acid sequences. There were a total of 1055 positions in the final dataset.

8.3 | Prediction of TA proteins in bacteria

The proteome of each species was downloaded from Uniprot (www.uniprot.org). Transmembrane domains in the whole proteome were predicted using TMHMM 2.0 (http://www.cbs.dtu.dk/services/TMHMM/). Proteins with a single TMS and less than 30 amino acids between the TMS and the C-terminus were considered candidates. These were tested for the presence of an N-terminal signal sequence using SignalP 4.1 (http://www.cbs.dtu.dk/services/SignalP/) with a cutoff value set to the value recommended by the software for the given bacterial phylum.

ACKNOWLEDGMENTS

We thank Nica Borgese, Ursula Jakob and Kathrin Ulrich as well as members of the Schwappach lab for extremely useful comments and discussion. The work was funded by the Deutsche Forschungsgemeinschaft (SFB1190 P04 to B.S. and A.F.).

REFERENCES

1. Kutay U, Hartmann E, Rapoport TA. A class of membrane proteins with a C-terminal anchor. Trends Cell Biol. 1993;3:72-75.
2. Borgese N, Fasana E. Targeting pathways of C-tail-anchored proteins. Biochim Biophys Acta Membr. 2011;1808(3):937-946.
3. Hardwick KG, Pelham HRB. SED5 encodes a 39-kD integral membrane protein required for vesicular transport between the ER and the Golgi complex. J Cell Biol. 1992;119(3):513-521.
4. Manial S, Man NT, Sewry CA, Morris GE. The Emery-Dreifuss muscular dystrophy protein, emerin, is a nuclear membrane protein. Hum Mol Genet. 1996;5(6):801-808. https://doi.org/10.1093/hmg/5.6.801.
5. Elgersma Y, Kwast L, van den Berg M, et al. Overexpression of Pex15p, a phosphorylated peroxisomal integral membrane protein required for peroxisome assembly in S. cerevisiae, causes proliferation of the endoplasmic reticulum membrane. EMBO J. 1997;16(24):7326-7341. https://doi.org/10.1093/emboj/16.24.7326.
6. Mozdy AD, McCaffery JM, Shaw JM. Dnm1p GTPase-mediated mitochondrial fission is a multi-step process requiring the novel integral membrane component Fis1p. J Cell Biol. 2000;151(2):367-380. https://doi.org/10.1083/jcb.151.2.367.
7. von Loeffelholz O, Kriechbaumer V, Ewan RA, et al. OEP61 is a chaperone receptor at the plastid outer envelope. Biochem J. 2011;438(1):143-153. https://doi.org/10.1042/BJ20110444.
8. Finke K, Plath K, Panzner S, et al. A second trimeric complex containing homologs of the Sec61p complex functions in protein transport across the ER membrane of S. cerevisiae. 1996;157(4):1482-1494.
9. Sommer T, Jentsch S. A protein translocation defect linked to ubiquitin conjugation at the endoplasmic reticulum. Nature. 1993;365(6442):176-179. https://doi.org/10.1038/365176a0.
10. Stefanovic S, Hegde RS. Identification of a targeting factor for post-translational membrane protein insertion into the ER. Cell. 2007;128(6):1147-1159. https://doi.org/10.1016/j.cell.2007.01.036.
11. Favaloro V, Spasic M, Schwappach B, Dobberstein B. Distinct targeting pathways for the membrane insertion of tail-anchored (TA) proteins. J Cell Sci. 2008;121(11):1832-1840. https://doi.org/10.1242/jcs.020321.
12. Schuldiner M, Metz J, Schmid V, et al. The GET complex mediates insertion of tail-anchored proteins into the ER membrane. Cell. 2008;134(4):634-645. https://doi.org/10.1016/j.cell.2008.06.025.
13. Srivistava R, Zalisko BE, Keenan RJ, Howell SH. The GET system inserts the tail-anchored SYP72 protein into endoplasmic reticulum membranes. Plant Physiol. December 2016;2016:00928-01145. https://doi.org/10.1104/pp.16.00928.
14. Xing S, Mehlhorn DG, Wallneroth N, et al. Loss of GET pathway orthologs in Arabidopsis thaliana causes root hair growth defects and affects SNARE abundance. Proc Natl Acad Sci. 2011;114(8):E1544-E1553. https://doi.org/10.1073/pnas.1619525114.
15. Mateja A, Keenan RJ. A structural perspective on tail-anchored protein biogenesis by the GET pathway. Curr Opin Struct Biol. 2018;51:195-202. https://doi.org/10.1016/j.sbi.2018.07.009.
16. Mock J-Y, Chartron JW, Zaslaver M, Xu Y, Ye Y, Clemons WM. Bag6 complex contains a minimal tail-anchor–targeting module and a mock BAG domain. Proc Natl Acad Sci. 2015;112(1):106-111. https://doi.org/10.1073/pnas.1402745112.
17. Mock J-Y, Xu Y, Ye Y, Clemons WM. Structural basis for regulation of the nucleo-cytoplasmic distribution of Bag6 by TRC35. Proc Natl Acad Sci. 2011;114(44):11679-11684. https://doi.org/10.1073/pnas.1702940114.
18. Anghel SA, McGillivray PT, Hegde RS, Keenan RJ. Identification of Oxal homologs operating in the eukaryotic endoplasmic reticulum. Cell Rep. 2017;21(13):3708-3716. https://doi.org/10.1016/j.celrep.2017.12.006.

ORCID

Ákos Farkas https://orcid.org/0000-0002-7295-3391
Evelina Ines De Laurentis https://orcid.org/0000-0002-6469-1627
Blanche Schwappach https://orcid.org/0000-0003-0225-6432
SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.