HIGH SCHOOL EDITION
Journal of Student Research
Volume 13 Issue 1 (2024)

Are Zinc Fingers of *α-proteobacteria* an Important Molecular Mechanism?

Pola Jakubowicz[1] and Nicola Tribe[#]

[1]Adcote School for Girls, United Kingdom
[#]Advisor

ABSTRACT

The zinc-finger proteins in *α-proteobacteria* are considered to be the precursors of zinc fingers in higher Eukaryotes, due to their structural similarities and the crucial role played by *α-proteobacteria* during eukaryogenesis. Since they interact with macromolecules, their emergence in *Bacteria*, and later in the *Eukarya*, provided those organisms with major functional upgrades and evolutionary advantages. Hence, the aim of this research is to prove the important role of α-proteobacterial zinc fingers in this taxon, and the significance of their relics, present in higher Eukaryotes. A part of this investigation involved identifying similar zinc-finger proteins between *α-proteobacteria* and humans using the blastp algorithm, finding their shared domains in the CDD database, and modelling their structures in the ChimeraX programme. The remaining research objectives were reached by analysing data from pre-existing studies, in order to come up with an additional set of conclusions, relevant to this investigation. It was established that the Ros/MucR zinc-finger proteins regulate many pathways, crucial for the survival of *α-proteobacteria* and their interactions with Eukaryotes. Additionally, it was found that many zinc-finger proteins supply *α-proteobacteria* with eukaryotic mechanisms, which differentiate them from other bacterial taxa. It was also concluded that α-proteobacterial zinc fingers may be responsible for the resistance of *α-proteobacteria* to certain heavy metals. This investigation also proposes a new evolutionary hypothesis for the emergence of zinc fingers in *Proteobacteria*, and presents further arguments in favour of the theory that the *Eukaryota* acquired zinc fingers from *α-proteobacteria* during eukaryogenesis.

Introduction

The term zinc finger (ZF) describes a conserved protein domain, which can readily bind to DNA, RNA and other macromolecules, enabling certain proteins to regulate e.g. gene expression, cellular signalling pathways and nucleic acid metabolism (Krishna *et al.*, 2003). In the most common structural type of ZFs, the Cys_2His_2, two cysteine and two histidine residues create four coordinate bonds with a Zn^{2+} ion, which ensures a proper conformation and, thus, the function of the domain. Other types include e.g. the Cys_3His ZFs (Chen *et al.*, 2000), zinc ribbons and zinc-binding loops (Eom *et al.*, 2016).

Initially, ZFs used to be regarded as eukaryotic inventions that got passed on to a small number of plant endosymbionts through horizontal gene transfer (HGT) (Chou *et al.*, 1998). However, since then many more have been discovered in diverse Prokaryotes (D'Abrosca *et al.*, 2020), which provided a basis for the theory that ZFs have a prokaryotic origin and were acquired by Eukaryotes during eukaryogenesis from the *Protomitochondrion* in the *α-proteobacteria* class (Esposti *et al.*, 2018).

Hence, investigating the role of ZFs in *α-proteobacteria* can provide valuable information about the capabilities of molecular regulation systems in one of the most advanced taxa in *Bacteria*. Additionally, it can help explain the wide range of functional improvements seen in *α-proteobacteria*, which act comparably and serve similar purposes to eukaryotic mechanisms. Moreover, this knowledge will clarify the evolutionary history of ZFs, and the proteins they are situated in.

Therefore, the aim of this study is to determine if α-proteobacterial ZFs played an important role in their evolutionary success, execution of crucial environmental processes, and in the advancement of other species. This conclusion will be reached by analysing ZFs in *α-proteobacteria* and their counterparts in other species from the perspective of molecular biology, focusing on their predicted conformation, epigenetic regulation of gene expression, performance under different cellular conditions, and molecular indications of adaptive evolution.

Materials and Methods

Research Papers, Websites and Computer Programmes

The majority of the research papers used in this investigation are deposited in Google Scholar (https://scholar.google.com/), with the rest being found through a Google search (https://www.google.com/). Scientific books used in the Glossary, and those mentioned in the bibliography were accessed from the Main Library of the University of Gdańsk in Poland. All used studies have been peer-reviewed.

This investigation utilised ZF protein sequences found in the Protein database of the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov/protein/). Subsequently, the Conserved Domains Database (CDD) (NCBI *Conserved Domain Search*, 2023) (Marchler-Bauer *et al.*, 2017) was used to further investigate the ZF domains in the chosen proteins. The Protein Data Bank (PDB) (https://www.rcsb.org/) was accessed to import a reference experimental structure of an α-proteobacterial Zn coordination sphere. The Taxonomy Browser feature (NCBI *Taxonomy browser (root)*, 2023) of the NCBI website was used in order to confirm the taxonomy of certain bacterial species.

Then, the blastp version of the BLAST algorithm (Altschul, 1997), available as a search engine on the NCBI website, was utilised to determine the prevalence of certain ZF proteins in *α-proteobacteria* and to find similar sequences between *α-proteobacteria* and humans. Since the blastn version of BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_SPEC=GeoBlast&PAGE_TYPE=BlastSearch) wasn't able to find any similarities between α-proteobacterial and human gene sequences, this study didn't investigate any ZF genes in *α-proteobacteria*.

Next, the freely available ChimeraX 1.4 programme (https://www.cgl.ucsf.edu/chimerax/) from the University of California San Francisco (UCSF) was used to access the ColabFold (Mirdita *et al.*, 2022) software, which accurately predicts and displays the 3D structure of proteins just from their sequence input. The rendered images were then converted into .png files to be analysed and, later, put into this research paper.

Objects of Research

Cadmium, copper and iron were included in this investigation as their high concentrations in the environment pose a big threat to microbial communities (Yu *et al.*, 2021) (Dragone *et al.*, 2022) (Braun, 1997). Therefore, it would be useful to determine if their toxicity in *α-proteobacteria* was linked to the disrupted function of their ZFs coordinating different metal ions. Conversely, if *α-proteobacteria* show resistance to elevated levels of metals in the environment, ZFs can be potentially responsible for protecting *α-proteobacteria* from their harmful effects.

Populations of *α-proteobacteria* in soil and those inhabiting hydrothermal vents are of special interest in this study, because a lot of α-proteobacterial species reside in the root nodules of host plants as nitrogen-fixing endosymbionts (Fatnassi *et al.*, 2015), while others occupy the rhizosphere (Shu *et al.*, 2012). Subsequently, hydrothermal vent micro ecosystems are similar to the first habitable environments on Earth, so they

display well the metabolic pathways and ecological relationships appearing early in the history of life (Martin *et al.*, 2008).

Results and Discussion

Examples of Zinc Finger Diversity in *α-proteobacteria*:

The structure of the Ros/MucR ZF protein family was first fully characterised in an endosymbiont, *Agrobacterium tumefaciens* (D'Abrosca *et al.*, 2020), belonging to the nitrogen-fixing *Rhizobiales* order in *α-proteobacteria* (NCBI *Taxonomy browser (root)*, 2023). This protein family contains a Cys_2His_2 ZF (Fig 1) (Malgieri *et al.*, 2015), similarly to the majority of eukaryotic ZFs.

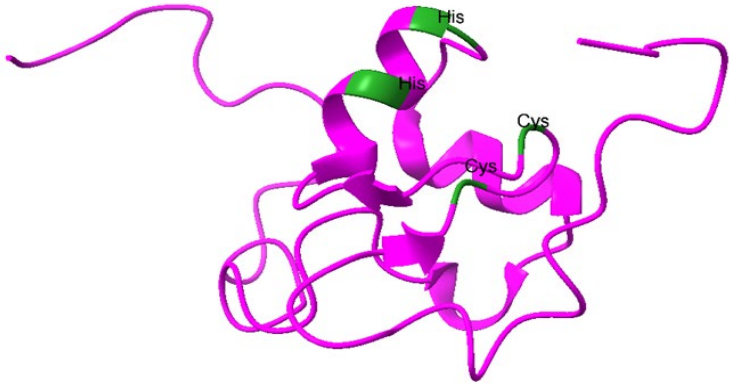

Figure 1. The Cys_2His_2 Zn coordination sphere (coloured green) in Ros87 (2JSP – PDB entry), edited in ChimeraX.

The Ros/MucR proteins also create expansive regulons, including up to 1350 genes (Jiao *et al.*, 2020) (Jiao *et al.*, 2022). Preliminary studies determined that genes under their control are responsible for nutrient sourcing, symbiosis, virulence Cooley *et al.*, 1991), general stress response and cell cycle regulation (Caswell *et al.*, 2013). The ZF domains in Ros/MucR proteins contain highly conserved sequences across different orders in *α-proteobacteria*. This statement is motivated by the fact that the 3D structures of their coordination spheres perfectly overlap in ChimeraX with the corresponding ones from other orders (Fig 2).

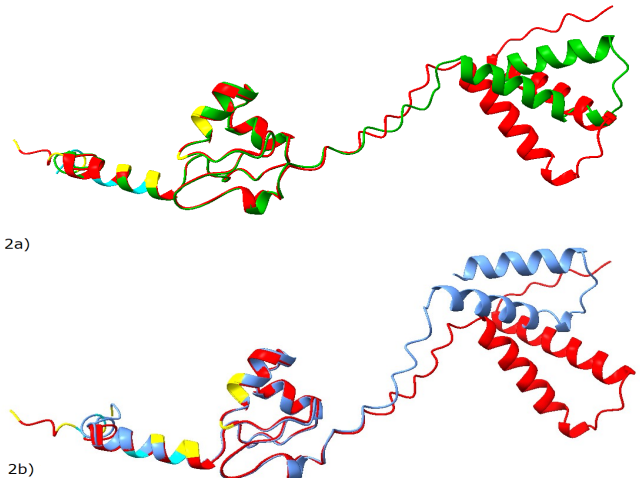

Figure 2. Ros/MucR ZF structures in *Rhizobiales* (NZ_CP033031.1 – NCBI Protein database entry, red) and *Rhodobacterales* (MBA4203399.1, green) in 2a), and *Caulobacterales* (MBV8681375.1, blue) in 2b), predicted and overlapped by AlphaFold (Mirdita *et al.*, 2022) in ChimeraX. Yellow indicates conserved DNA-binding amino acids, while cyan – the ones that differ.

ZitP is a transmembrane ZF protein containing a zinc ribbon structure (Mignolet *et al.*, 2016). Most importantly, it controls the morphogenesis of poles in bacterial cellular membranes, with respect to coordinating a Zn^{2+} ion, which determines its conformation. These intentional structural changes influence its interactions with other co-regulators: CpaM and CpaC, resulting in the initiation of different cellular responses (Mignolet *et al.*, 2016).

ZitP is widely distributed and highly conserved in *α-proteobacteria* (Mignolet *et al.*, 2016) (Fig 3), however, a blastp (Altschul, 1997) search of YP_002517671.2 showed an inconsistent presence of ZitP in other classes of *Proteobacteria*, and its absence in *Archaea*.

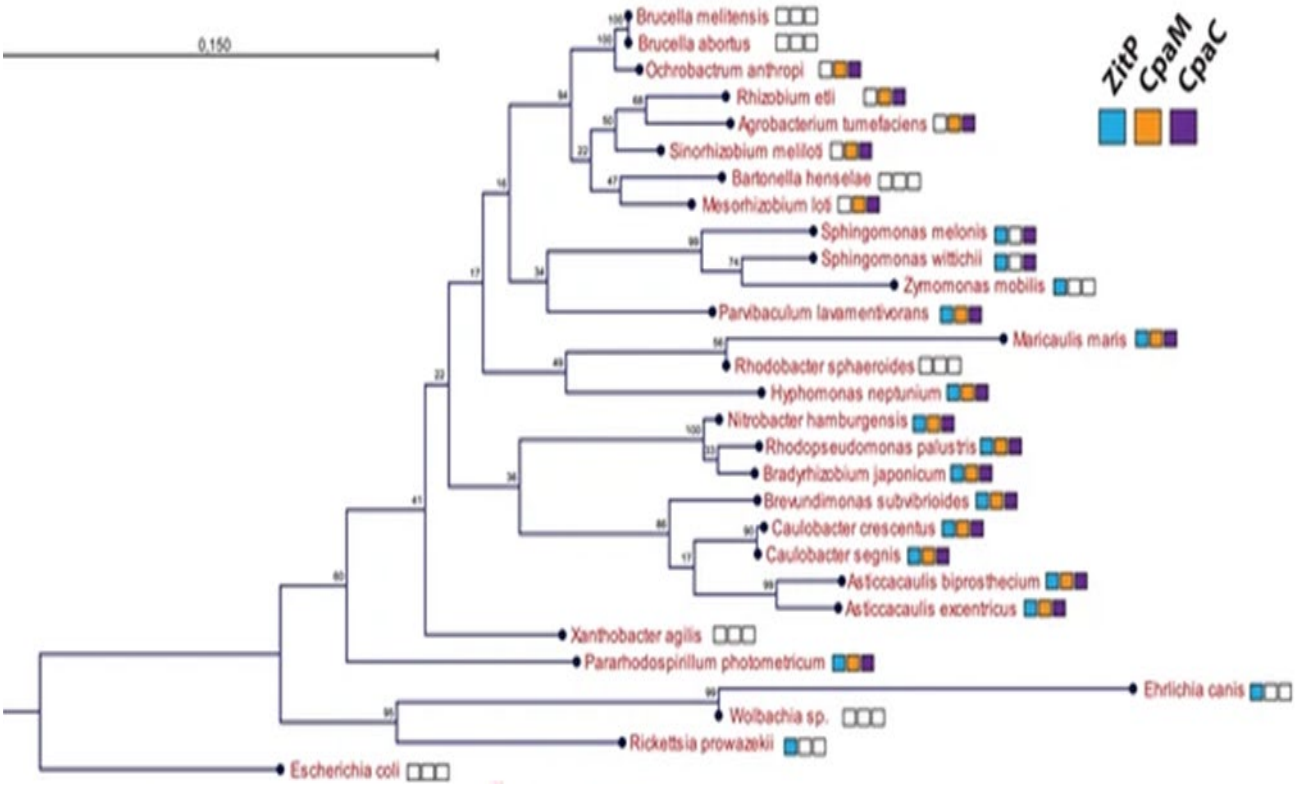

Figure 3. Prevalence of ZitP, CpaM and CpaC in *α-proteobacteria* and in an outside species – *Escherichia coli* (Mignolet *et al.*, 2016).

This unequal ZitP distribution across different taxa implies that the ZitP-based regulation system has a clear prokaryotic origin and is a characteristic feature of *α-proteobacteria*.

Additionally, there are inactivated ZF domains present in *α-proteobacteria*, namely in the Ku proteins, also containing the zinc ribbon structure (Krishna *et al.*, 2010). They bind to DNA during classical non-homologous end-joining (cNHEJ), the assembly process of antibody genes in B and T cells (Roth, 2015), and during the elongation of telomeres (Krishna *et al.*, 2010). *Eukarya* and *α-proteobacteria* are the only taxons in which the Ku proteins have lost their Zn-chelating residues. This, in combination with the heterodimerization of the Ku70 and Ku80 subunits gave rise to a Zn-free bridge-like region present in their Ku proteins (Krishna *et al.*, 2010). Evolving away from needing Zn was an adaptation to end the Zn-dependence of the Ku proteins, while retaining their functionally crucial zinc ribbon structure (Krishna *et al.*, 2010).

Human Paralogs to α-proteobacterial Zinc Fingers

The sequence of the human ZNF184 protein (NP_001305820.1) shows similarity to the ZF proteins in the *Rickettsiales* and *Brucella* families. Those matches were found by the blastp algorithm (Altschul, 1997) in their C-terminal regions, which contain a eukaryotic COG5048 ZF domain (NCBI *Conserved Domain Search*, 2023) (Marchler-Bauer *et al.*, 2017). Additionally, the minimal E-values for the blastp searches were smaller for *Proteobacteria* (1e-105) than for *Archaea* (5e-70) (Altschul, 1997), suggesting that the sequences in *Proteobacteria* are more similar to ZNF184 than the ones in *Archaea* and, thus, the acquisition of ZNF184 from the *Proteobacteria*.

Currently, no experimental structures of ZNF184 are deposited in PDB, so AlphaFold in ChimeraX was used in this investigation to predict the structures of the C-terminus of ZNF184, and its best α-proteobacterial match (RYE13711.1) (Mirdita *et al.*, 2022) (Fig 4). Interestingly, the prokaryotic sequence gave rise to a very similar conformation to a fragment of the human one, which suggests that the gene of the α-proteobacterial protein multiplied over the course of evolution, forming the eukaryotic paralog in the process.

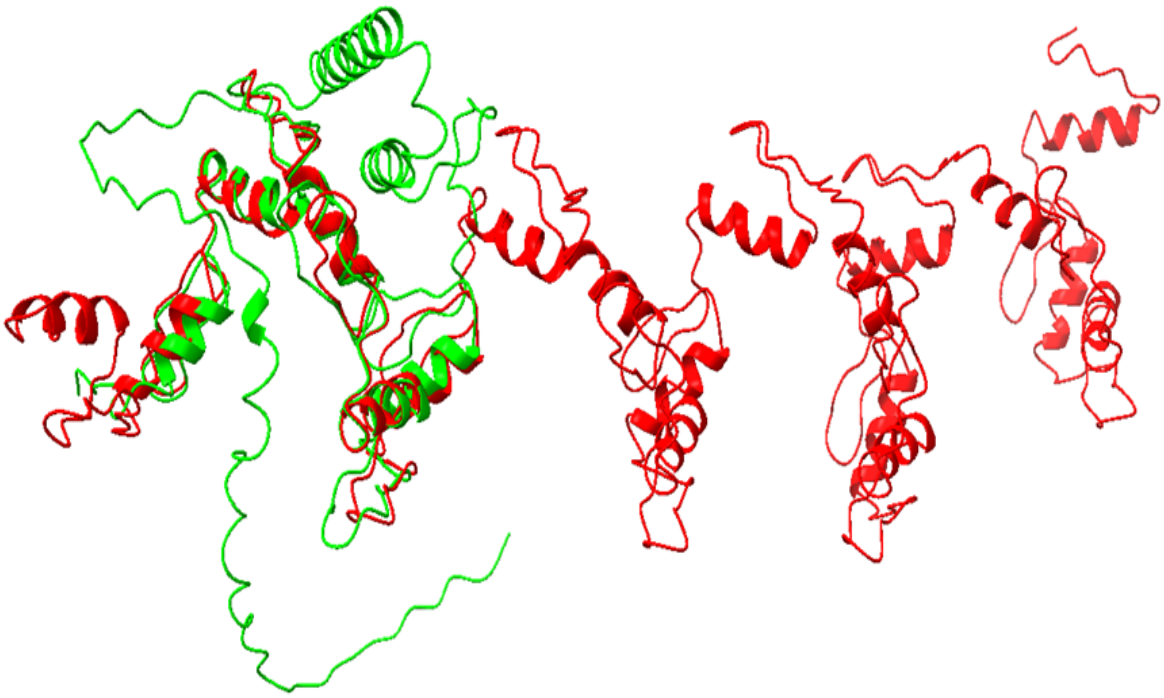

Figure 4. Structure of the human ZNF184 C-terminus (red) and its α-proteobacterial match (green) predicted in ChimeraX (Mirdita *et al.*, 2022).

According to the blastp algorithm, the ZNF226 (XP_047295330.1) C-terminus containing the COG5048 ZF domain (NCBI *Conserved Domain Search*, 2023) (Marchler-Bauer *et al.*, 2017) has similar sequences to the ones in *α-proteobacteria*, while the N-terminal region does not (Altschul, 1997). Favourably, the minimal E-value of the blastp search in *Proteobacteria* (6e-101) is lower than that in *Archaea* (5e-68), suggesting that those proteobacterial sequences are evolutionarily closer to human ZFs than the archaeal ones (Altschul, 1997).

No deposited ZNF226 structures were found in PDB, so AlphaFold in ChimeraX was used to predict the structure of the human ZNF226 C-terminus and its α-proteobacterial match (MBX8803466.1) (Mirdita *et al.*, 2022) (Fig 5). These structure predictions present a similar result to the ZNF184 investigation, which suggests a close evolutionary relationship between α-proteobacterial and eukaryotic ZF proteins. Unfortunately, this AlphaFold prediction is not very accurate, according to the B-factor colouring in ChimeraX.

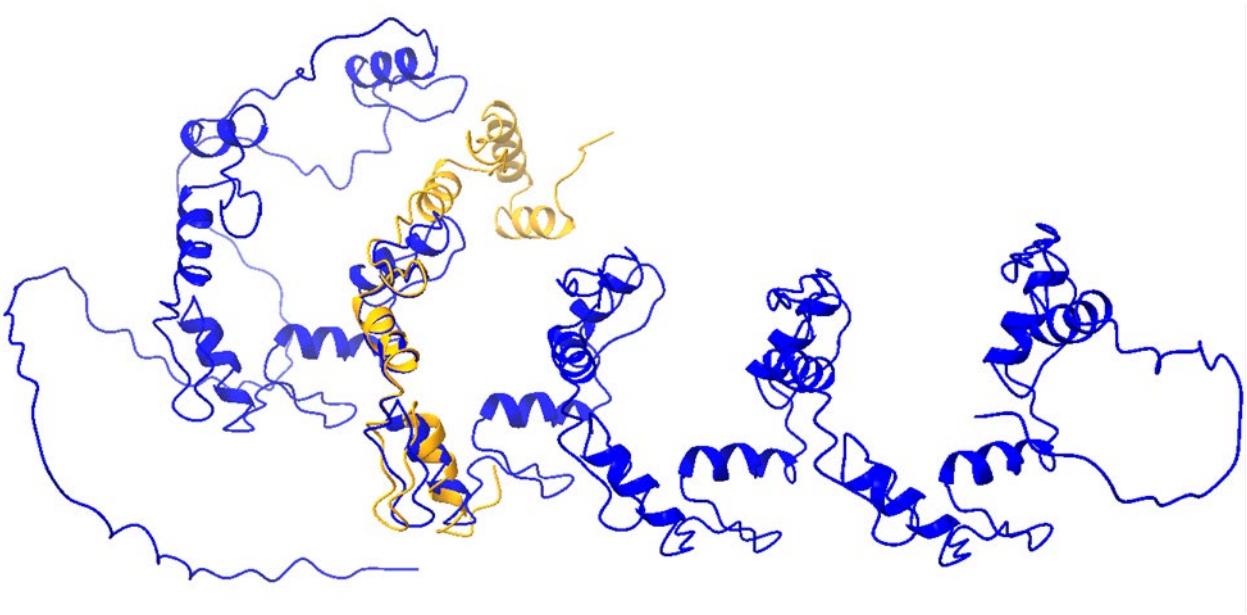

Figure 5. Structure of the human ZNF226 C-terminus (blue) and its α-proteobacterial match (yellow) predicted in ChimeraX (Mirdita *et al.*, 2022).

The ZNF23 (NP_001368913.1), similarly to ZNF184 and ZNF226, contains the COG5048 domain in its C-terminus (NCBI *Conserved Domain Search*, 2023) (Marchler-Bauer *et al.*, 2017).

The prokaryotic protein MBX8819088.1 was used in order to visualise the clear structural overlap of the α-proteobacterial ZF on the ZNF23 C-terminus in ChimeraX (Mirdita *et al.*, 2022)(Fig 6). Once again, this figure provides evidence for the existence of multiplication events in the evolutionary history of eukaryotic ZF proteins.

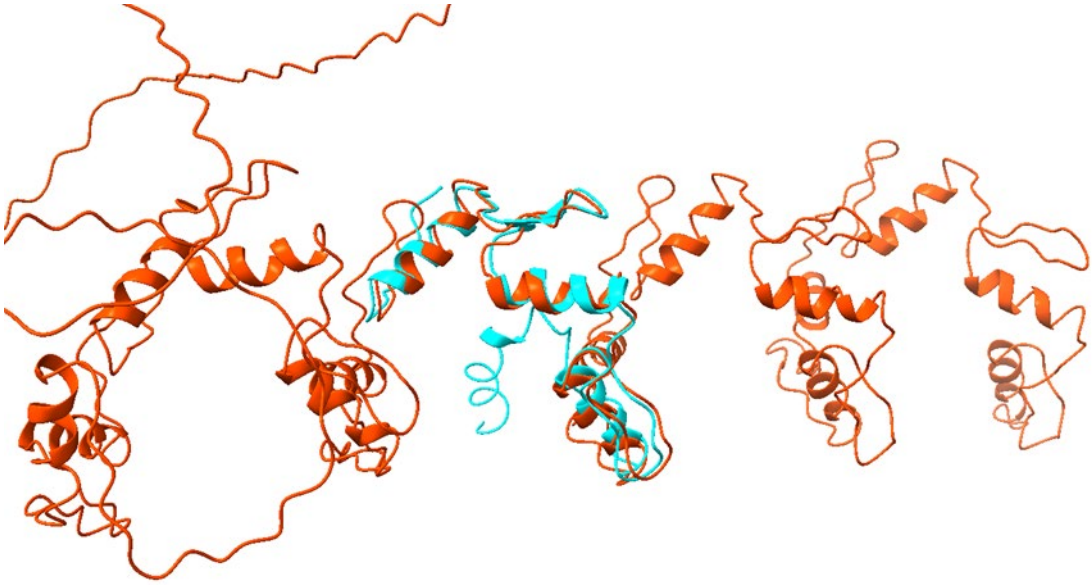

Figure 6. Structure of the human ZNF23 C-terminus (orange) and its α-proteobacterial match (blue) predicted in ChimeraX (Mirdita *et al.*, 2022).

In accordance with the CDD database, the C-terminal section of the ZNFX1 (EAW75661.1) contains domains exhibiting helicase properties (NCBI *Conserved Domain Search*, 2023) (Marchler-Bauer *et al.*, 2017). Interestingly, this is also the region that shows similarity to *α-proteobacteria* (Altschul, 1997).

There are no experimental structures of ZNFX1 deposited in PDB, so the C-terminus of this protein and its closest α-proteobacterial match (MBV35386.1, 200…600 aa) were modelled in AlphaFold via ChimeraX (Mirdita *et al.*, 2022) (Fig 7). The similarities in the structures of these proteins aren't clearly visible, apart from the four Zn-binding sites indicated in the figure. This means that the ZNFX1 protein itself doesn't provide any more evidence directly linking the multiplication of ancient α-proteobacterial ZF sequences with the formation of eukaryotic proteins.

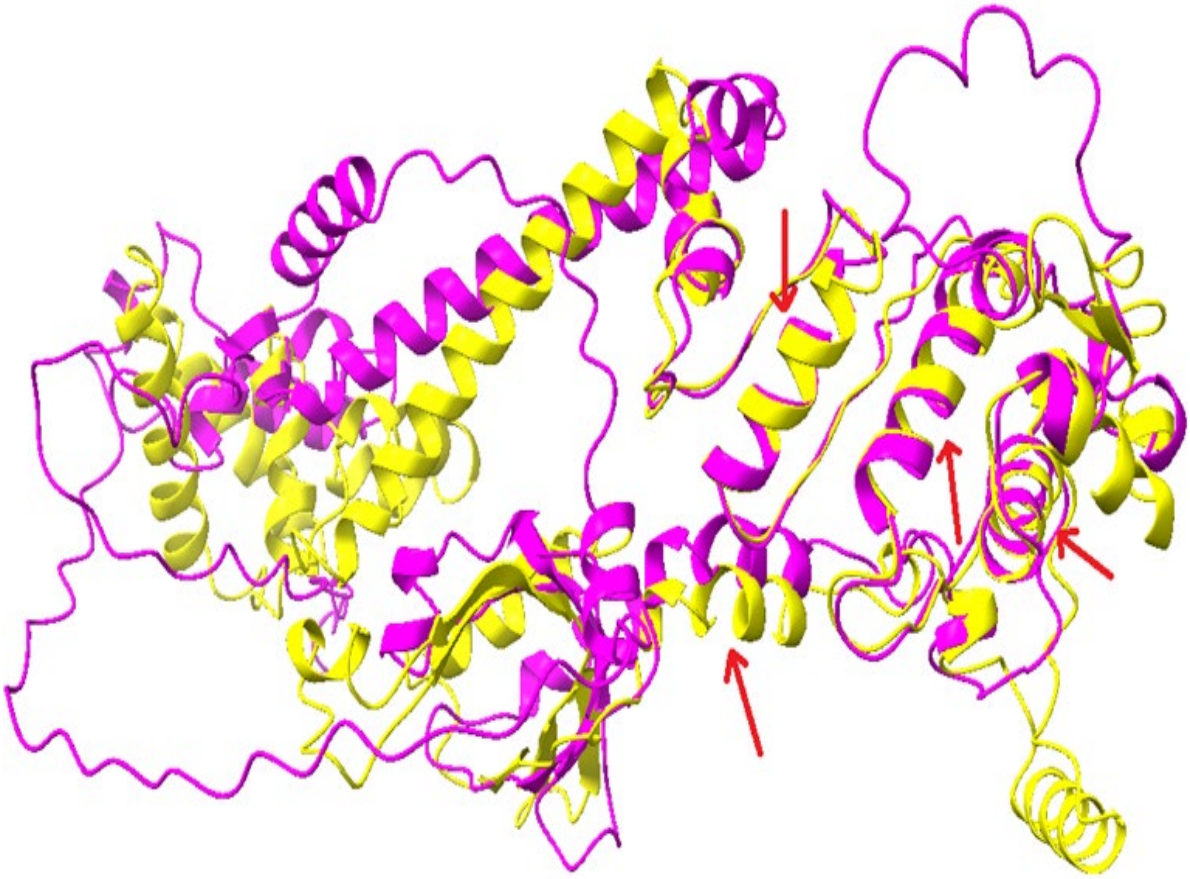

Figure 7. Structure of the human ZNFX1 C-terminus (pink) and its α-proteobacterial match (yellow) predicted in ChimeraX (Mirdita *et al.*, 2022). Four red arrows indicate Zn-binding sites that are located similarly in both proteins.

The ZNF384 (NP_001372675.1), containing the COG5048 domain is widely distributed across some *Proteobacteria* classes (Altschul, 1997). The structures of ZNF384 (200…500 aa) and its α-proteobacterial match (NP_001372675.1) were determined using AlphaFold in ChimeraX (Mirdita *et al.*, 2022) (Fig 8). Similarly to ZNFX1, this figure contains no evidence in favour of the theory that eukaryotic ZF proteins come from the multiplied sequences of α-proteobacterial ZFs.

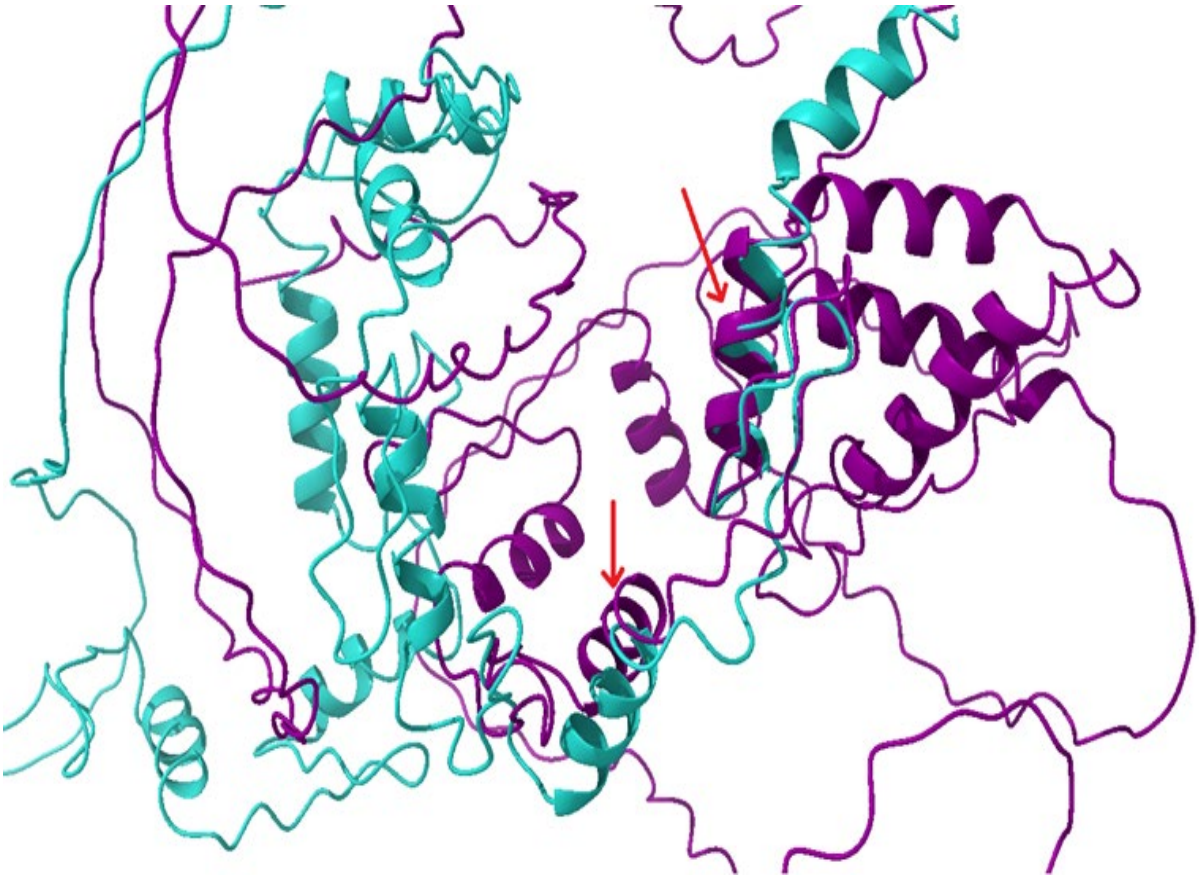

Figure 8. Structure of the human ZNF384 (turquoise) and its α-proteobacterial match (purple) predicted in ChimeraX (Mirdita *et al.*, 2022). Two red arrows indicate Zn-binding sites that are located similarly in both proteins.

Overall, from these structure predictions it is possible to confidently hypothesise that the majority of ZF domains in the C-termini of human ZF proteins are paralogs of ZFs found in *α-proteobacteria*.

Regulatory Functions of Zinc Fingers in *α-proteobacteria*:

The Ros/MucR ZFs are classified as xenogeneic silencers, as the transcription of newly-acquired genes they regulate increases over time, leading to their successful conservation in the genome and, therefore, a repressed expression of novel genes (Jiao *et al.*, 2022).

Additionally, the Ros/MucR proteins in *α-proteobacteria* are referred to as histone-like nucleoid-structuring proteins (H-NS). This is because they can bind to their own promoters, which suggests a negative feedback loop regulation of their own transcription. Ros/MucR also control the magnitude of their effect on gene expression, by adjusting the number of binding sites in their target promoters (Baglivo *et al.*, 2018).

The genes activated by the Ros/MucR regulon (Jiao *et al.*, 2020) (Fig 9) are involved e.g. in the production of exopolysaccharides (Nwodo *et al.*, 2012), which are crucial in the formation of the biofilm, as they facilitate the process of cohesion to adjacent cellular walls, enabling bacteria to recognise each other. Exopolysaccharides also cause biofilms to retain large amounts of water, preventing the colonies from drying out in water-scarce environments (Nwodo *et al.*, 2012).

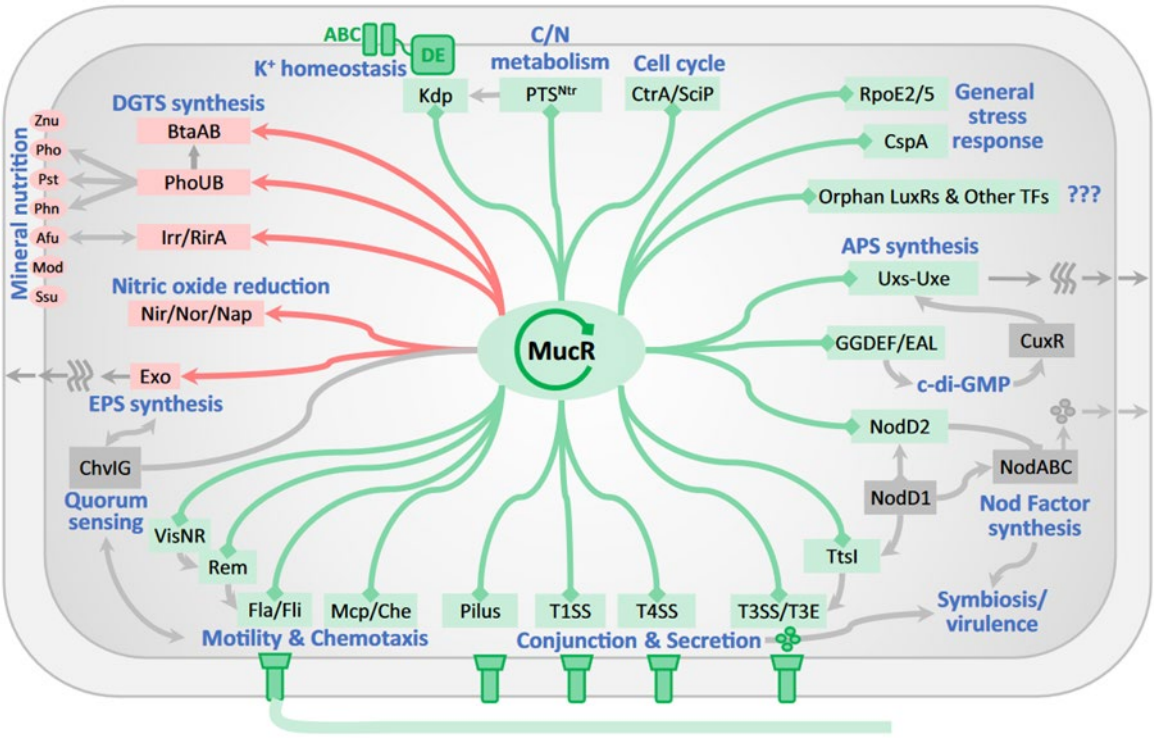

Figure 9. The RosMucR regulon in *α-proteobacteria* (Jiao *et al.*, 2020). Red arrows represent activation and green represent inhibition.

Moreover, Ros/MucR proteins up-regulate the transcription of genes encoding transport proteins for iron, molybdenum and sulphur, which are crucial in creating the catalytic properties of nitrogenase (Jiao *et al.*, 2016). This enzyme facilitates the reaction of reducing atmospheric nitrogen to ammonia – a bioavailable form of nitrogen for plants, enabling them to develop properly in penurious soil (Wagner, 2011). Ros/MucR also activate genes of Zn^{2+} and phosphate transporters, which ensure appropriate levels of these molecules in the bacterial cell and, thus, the proper development of root nodules, necessary for the symbiosis of plants with nitrogen-fixing *α-proteobacteria* (Jiao *et al.*, 2016).

In *Brucella canis* (NCBI *Taxonomy browser (root)*, 2023), the MucR transcription factors also regulate 38 genes involved in the last steps of the synthesis of the murein wall, making bacteria less vulnerable to lactam antibiotics, which disrupt these processes (Sun *et al.*, 2021). It also plays a role in developing its pathogenicity, however, this mechanism is not entirely understood, as the MucR proteins down-regulate some genes encoding lipopolysaccharides (LPS) (Caswell *et al.*, 2013), which are proven to be potent endotoxins (Bertani *et al.*, 2018).

When the ZitP ZF protein chelates a Zn^{2+} ion at one pole of the cell, it influences the swarming motility of bacterial groups and the activity of CtrA, which is a master cell cycle regulator, exclusive to *α-proteobacteria*. At the second pole, where ZitP does not coordinate Zn^{2+}, it is able to bind to CpaM and initiate the transport of CpaC to the outer membrane, which controls the biogenesis of bacterial pili (Mignolet *et al.*, 2016). The N-terminus of ZitP can also interact with PopZ, enabling them both to control the process of cytokinesis (Bergé *et al.*, 2016). It seems that *α-proteobacteria* have found a way of developing 'isoforms' in some of their proteins, by recruiting ZF domains to play that role, instead of performing alternative splicing alike the Eukaryotes (Zhiguo *et al.*, 2013). This fine-tuned use of ZFs in *α-proteobacteria* proves a high structural complexity of their proteomes and the presence of Eukaryote-like regulation mechanisms in their molecular machinery.

Metal Substitution in α-proteobacterial Zinc Fingers

A Cd^{2+} substitution of the Zn^{2+} ion does not change the DNA-binding abilities of the Ros/MucR protein family, because the resulting change in conformation only affects one out of four basic regions involved in DNA interactions (Malgieri *et al.*, 2014). This shows a great structural flexibility of this family, as their functionality isn't impaired by the bigger ionic radius of Cd^{2+} (Shannon, 1976) in the Zn coordination sphere.

In the Ros/MucR family in *α-proteobacteria*, a Zn^{2+} to Cu^{+} substitution prevents proper conformations from forming, which leads to a complete deactivation of these proteins (Dragone *et al.*, 2022). For those ZFs, high concentrations of Cu^{+} in the cell pose a big threat, as Cu^{+} is thermodynamically favoured over Zn^{2+} in Cys_2His_2 ZFs (Doku *et al.*, 2013). This means that those proteins are more likely to bind to Cu^{+} than to Zn^{2+} in the presence of both of these ions in the cytoplasm (Waldron *et al.*, 2009), which results in their lack of functionality in Cu-rich conditions.

Additionally, a 2018 study found that the Cu^{+} ion readily replaces Zn^{2+} in the coordination sphere of a non-classical ZF domain in the copper response regulator 1 transcription factor (CRR1) (Kluska *et al.*, 2018), which has been identified in *α-proteobacteria* (Altschul, 1997). Most importantly, this substitution does not result in significant changes to its structure and the DNA-binding abilities (Kluska *et al.*, 2018), suggesting a functional Zn^{2+} to Cu^{+} substitution in this protein.

A Cu^{2+} substitution of Zn^{2+} does not result in functional Ros/MucR ZF domains, as they are unable to maintain their proper structure, caused by the oxidation of the two cysteines in their coordination sphere and the formation of a stable disulfide bridge between them. In this state, the hydrophobic core of the ZF can not enforce a proper fold, making the protein permanently non-functional (Dragone *et al.*, 2022).

A Zn^{2+} replacement by Fe^{2+} in the Cys_2His_2 ZF of the 5S RNA transcription factor TFIIIA inhibits its DNA recognition (Kluska *et al.*, 2018). Despite it being a eukaryotic transcription factor, the blastp algorithm (Altschul, 1997) found ZF proteins in *α-proteobacteria* with similar sequences to its N-terminal region, abundant in ZF domains (NCBI *Conserved Domain Search*, 2023) (Marchler-Bauer *et al.*, 2017). This suggests that these α-proteobacterial ZF proteins can also be susceptible to defective structural changes after Zn^{2+} to Fe^{2+} substitutions.

Despite the fact that the Fe^{3+} ions are more stable than Fe^{2+} (Kamboch, 2020), there aren't any research papers published on the Zn^{2+} to Fe^{3+} substitution in α-proteobacterial ZFs specifically.

Populations of *α-proteobacteria* Inhabiting Metal-Abundant Environments

The incorporation of a Zn-enriched sewage sludge to soil has no short- and long-term effects on its α-proteobacterial population (Gomes *et al.*, 2010). Additionally, in Zn-polluted soils about 10% of all prokaryotic species represent the nitrogen-fixing order of *Rhizobiales* (Liu *et al.*, 2022) (NCBI *Taxonomy browser (root)*, 2023). It was also found that plant inoculation with nitrogen-fixing *α-proteobacteria* in soil with a Zn concentration of 400 mg kg^{-1}, increased their shoot and root lengths in comparison to the non-inoculated control plants (Fig 10). The biomass of those organs and the Zn uptake from the soil also rose significantly in the presence of *α-proteobacteria* (Jian *et al.*, 2019), which proposes a new method of microbial bioremediation of Zn-polluted soils.

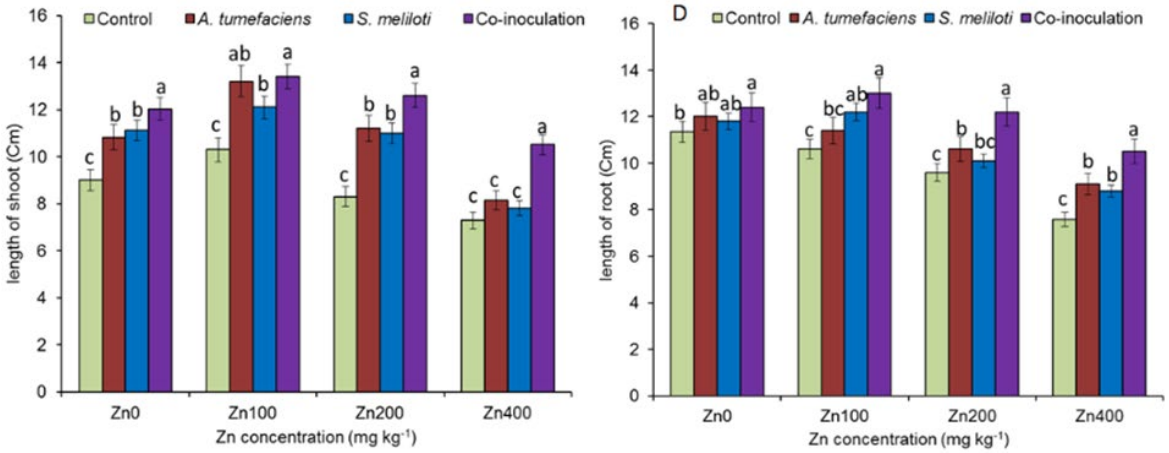

Figure 10. Increase in the shoot and root length in plants inoculated with *α-proteobacteria* in elevated Zn concentrations in soil (Jian *et al.*, 2019).

A-proteobacteria were found to be exceptionally resistant to high Cd levels in soil (Lorenz *et al.*, 2006). Since Cd^{2+} can functionally replace Zn^{2+} in their ZFs (Malgieri *et al.*, 2014), it means that *α-proteobacteria* could act as sinks for the Cd that manages to get to their cytoplasm, lowering its concentration in soil (Waldron *et al.*, 2009). However, due to much higher Zn concentrations in soil than Cd (Wyszkowska *et al.*, 2013), and the miniscule amount of Cd required to assemble ZFs, the effect of this substitution is negligible and, thus, would not influence their Cd resistance.

A-proteobacteria is one of the most abundant prokaryotic taxa in soil with Cu levels naturally elevated to 200 and 500 mg kg^{-1} (Wang *et al.*, 2008). Analogously to raised Zn levels, plant inoculation with nitrogen-fixing *α-proteobacteria* in soil with a Cu concentration of 400
mg kg^{-1}, increased the length of their shoots and roots (Fig 11). Furthermore, it also boosted the biomass of those organs and the Cu uptake from the soil (Jian *et al.*, 2019). Thus, nitrogen-fixing *α-proteobacteria* can potentially be used in the microbial bioremediation of Cu-polluted grounds.

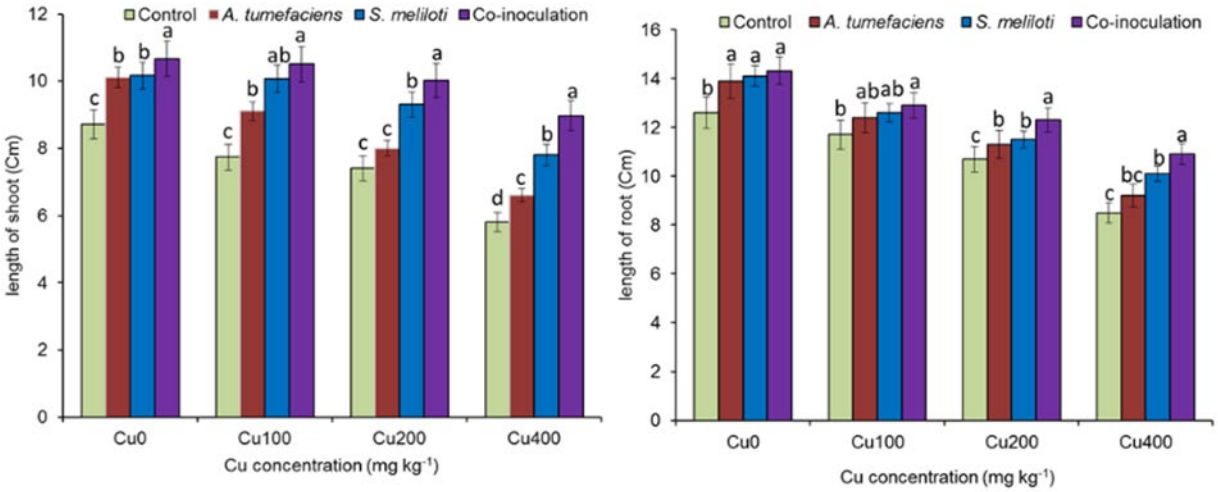

Figure 11. Increase in the shoot and root length in plants inoculated with *α-proteobacteria* in elevated Cu concentrations in soil (Jian *et al.*, 2019).

There are no open-access research papers published on the α-proteobacterial populations in Fe-polluted soils, thus no data about it can be presented here.

Four α-proteobacterial species show resistance to a 65 mg kg^{-1} concentration of Zn^{2+} ions in the environment of the 'Lucky Strike' hydrothermal field (Farias *et al.*, 2015). Additionally, in the Menez Gwen hydrothermal vent system, 13% of the genetic material from the bathyal zone with a Zn concentration of 80 mg kg^{-1} was found to be α-proteobacterial (Cerqueira *et al.*, 2015). However, in vent chimney samples, where Zn is way more abundant (>10,000 mg kg^{-1}), *α-proteobacteria* make up only 1% of all bacterial genetic material (Cerqueira *et al.*, 2015). Therefore, vent *α-proteobacteria* do not display an exceptional Zn resistance in comparison to other extremophilic species inhabiting hydrothermal vents, and other members of this taxon living in Zn-polluted soil.

Four α-proteobacterial species from the 'Lucky Strike' were found to be resistant to a 22.4 mg kg^{-1} concentration of Cd^{2+} (Farias *et al.*, 2015). This means that there are some α-proteobacterial species in hydrothermal vents that are tolerant to elevated Cd concentrations. However, the literature about this resistance is scarce, so no further conclusions can be currently justified.

There are three species of *α-proteobacteria* identified in the 'Lucky Strike' that show resistance to a 63.5 mg kg^{-1} Cu^{2+} concentration (Farias *et al.*, 2015). Similarly, this topic isn't well researched, and the tolerance of some strains to Cu is not an extraordinary characteristic, as it can just be a species trait.

The hydrothermal plumes of the Lau Basin in the south-west Pacific Ocean can be occupied by *α-proteobacteria* that lower the environmental concentrations of Fe^{3+} ions. This is because these plumes show an elevated level of ABC transporters of siderophores that come from the *Rhizobiales* order in *α-proteobacteria* (Cohen *et al.*, 2021) (NCBI *Taxonomy browser (root)*, 2023). After being released into the environment, siderophores chelate Fe^{3+} ions and bind to the membrane receptors of the bacteria. After that, Fe^{3+} is reduced to Fe^{2+}, gets transported into the cell, and as a result, the Fe^{3+} concentration in the environment decreases (Deng, 2021). This means that *α-proteobacteria* can act as sinks for excess Fe^{3+} present in vent environments and, therefore, protect vulnerable species from its harmful effects. However, their ZFs do not seem to be involved in this process.

Evolution of Prokaryotic Zinc Fingers to Binding Zinc

In certain experimental conditions Zn can be favoured over Fe to bind to the IscU protein (Ramelot *et al.*, 2004), necessary during the assembly of Fe/S clusters in other proteins (Py *et al.*, 2010). The Zn-bound form retains the proper structure of this protein and closely resembles the Zn-coordination spheres of Cys_3His ZFs (Chen *et al.*, 2000) (Ramelot *et al.*, 2004). The γ-proteobacterial IscU protein, used in this study shows close sequence similarities to many α-proteobacterial IscU proteins, with the best matches having the E-value of 4e-76 (Altschul, 1997). This means that this favoured Fe to Zn substitution could also happen in *α-proteobacteria*.

Five Cys_3His ZF domains and one Fe/S cluster are confirmed to be present in the CPSF30 factor, found primarily in the *Eukaryota*. It was found that CPSF30 loads Fe before Zn (Shimberg *et al.*, 2016), which leads to the conclusion that it has a higher affinity to Fe than to Zn, and that it is more inclined to creating Fe/S clusters than ZFs in cysteine-rich micro-environments. The blastp algorithm found matches to CPSF30 mostly in *γ-proteobacteria*, with only one sequence in *α-proteobacteria* showing some similarity to CPSF30 (Fatnassi *et al.*, 2015).

Conclusively, this data motivates the formation of a hypothesis that ZFs evolved from Fe/S clusters in *Proteobacteria*. Additionally, no arguments against this novel theory were found.

Acquisition and Evolution of Eukaryotic Zinc Fingers from *α-proteobacteria*

At the time of its discovery, the Ros/MucR protein family was known to be recognised by the transcriptional machinery of the *Prokaryota*, indicating that it was not a novel addition to the prokaryotic genome, and that it most likely evolved in pair with other prokaryotic proteins (Chou *et al.*, 1998). However, during that time, the hypothesis that the Ros/MucR family originated in the *Prokaryota* was not widely accepted, due to the emergence of a theory of its HGT from plants (Chou *et al.*, 1998). Although, after identifying the Ros/MucR protein family in a large quantity of *α-proteobacteria* (D'Abrosca *et al.*, 2020), it has been proposed that the Cys_2His_2 ZF domain is indeed a prokaryotic invention (Moreira *et al.*, 2000) which got distributed to the *Eukaryota* during eukaryogenesis (Esposti *et al.*, 2018).

A blastp (Altschul, 1997) search in *Archaea* found six homologous sequences to Ros/MucR ZFs of *α-proteobacteria*, suggesting that this ZF domain was not present in the common ancestor of *Archaea*. This means that this ZF is not an archaeal invention, but rather originated in *Bacteria*, and that those six species likely acquired it by HGT (Wagner *et al.*. 2017) from their α-proteobacterial endosymbionts. Coincidentally, *α-proteobacteria* have a very effective HGT mechanism – the gene transfer agent (GTA) (Richards *et al.*, 2011). Since not a lot of species of *Archaea* have similar sequences to the α-proteobacterial Ros/MucR ZFs, it suggests that eukaryogenesis and, thus, the transfer of α-proteobacterial ZFs to *Eukaryota* was triggered rapidly, due to the widespread changes done by the α-proteobacterial genes to the Archaeans' metabolism and signalling pathways.

Interestingly, a 2009 study states that the region around the two C-terminal ZFs in the Pol2p subunit of DNA ε-polymerase may have directly come from *α-proteobacteria*, while the ZFs themselves have a distinctive archaeal ancestry from their D- and B-family polymerases (Tahirov *et al.*, 2009). The fact that this region of the polymerase is a fusion between archaeal and α-proteobacterial sequences, provides another piece of evidence in favour of the theory that ZF domains in the *Eukaryota* and the *Archaea* have an *α*-proteobacterial origin, as the genomes of *α-proteobacteria* and the *Archaea* merged together during eukaryogenesis and evolved to form the Last Eukaryotic Common Ancestor (LECA) (Esposti *et al.*, 2018).

By this theory, the ZFs themselves extensively differentiated to better adapt to the needs of Archaeans, and then to the molecular systems of Eukaryotes. Thus, it is unsurprising that those ZF domains are deemed as archaeal, while the region surrounding them still resembles sequences found in *α-proteobacteria*.

Conclusions

The Wide Range of Molecular Approaches and Processes Regulated by ZFs in α-proteobacteria

The ZF protein families in *α-proteobacteria* show a diversified range of morphological approaches, from the Cys_2His_2 of the Ros/MucR family (Malgieri *et al.*, 2015), the classical zinc ribbon of the ZitP (Mignolet *et al.*, 2016) and an atypical one in the Ku proteins (Krishna *et al.*, 2010). The proteins in the Ros/MucR family are the best investigated ones and are found to possess many evolutionarily advanced traits, such as: acting as xenogeneic silencers in the process of conserving novel genes (Jiao *et al.*, 2022), and modulating their own transcription and regulation strength (Baglivo *et al.*, 2018). Additionally, the pathways they control enable *α-proteobacteria* to survive in extreme environmental conditions (Nwodo *et al.*, 2012), initiate the circulation of nitrogen – a crucial element in the ecosystem (Jiao *et al.*, 2016), and to modulate their virulence in order to successfully invade other organisms (Bertani *et al.*, 2018).

The ZitP protein controls entirely different pathways with respect to being bound to a Zn^{2+} ion, which changes its conformation (Mignolet *et al.*, 2016) (Malgieri *et al.*, 2014), producing a similar effect to alternative

splicing in the *Eukaryota* (Zhiguo *et al.*, 2013). This property increases the informational capacity of the bacterial chromosome and provides a major functional advancement in the proteome compared to other *Prokaryota* (Zhiguo *et al.*, 2013).

Crucial DNA metabolism in *α-proteobacteria* is also regulated by proteins with ZF motifs, most notably the Ku proteins (Krishna *et al.*, 2010) (Roth, 2015), which gained independence from environmental Zn by evolving to retain their shape in its absence (Krishna *et al.*, 2010). This proves a critical role of the ZF structure in managing genetic information.

Overall, the properties of α-proteobacterial ZFs enable them to precisely express genes crucial for their survival, the circulation of elements in the biosphere, and their ecological relationships that integrate them into the existence of higher Eukaryotes. Additionally, they provide *α-proteobacteria* with structural and functional advancements in the genome and the proteome, making them more similar to eukaryotic networks. All of these developments result in *α-proteobacteria* being better equipped to survive in hostile environmental conditions.

The Morphological Resemblance of ZF Proteins in α-proteobacteria to the Complex Eukaryotic Ones

Certain human ZF proteins contain COG5048 Zn-binding domains (NCBI *Conserved Domain Search*, 2023) (Marchler-Bauer *et al.*, 2017), which are paralogous to specific α-proteobacterial sequences (Altschul, 1997). These domains are made out of multiple segments with very similar predicted structures to the α-proteobacterial proteins (Mirdita *et al.*, 2022), suggesting that those bacterial sequences multiplied over the course of evolution to create advanced Zn-binding sites in the proteins of higher Eukaryotes. The ZNFX1 protein is especially interesting, as it shows the greatest similarity to the sequences that are recognised as α-proteobacterial (Altschul, 1997) DNA and RNA helicase domains (NCBI *Conserved Domain Search*, 2023) (Marchler-Bauer *et al.*, 2017). This implies that some ZF proteins of higher Eukaryotes originated from α-proteobacterial ones, already involved in the metabolism of nucleic acids.

The type II introns, found abundantly in *α-proteobacteria*, mitochondria and plastids (Casalino, 2017) can be responsible for the lack of similarity between the α-proteobacterial and human genes in this investigation, but a distinctive resemblance in the sequence and the morphology of their proteins. This is because the intron sequence itself can differ dramatically (de Lencastre *et al.*, 2005), depending on the epigenetic regulation and the genome topology present in the cell, while the exons retain a similar, functional sequence. Additionally, type II introns contain ZFs in their structure (Martínez-Abarca *et al.*, 2000), which makes these domains further responsible for increasing the amount of different proteins that can be produced in those organelles and the *α-proteobacteria*.

The Ability of α-proteobacterial Zinc Fingers to Function in Different Environmental Conditions and Its Ecological Implications

The Zn resistance of soil (Dragone *et al.*, 2022) and vent (Cerqueira *et al.*, 2015) *α-proteobacteria* can be explained with their RND Zn excretion mechanisms, which are regulated by the Ros/MucR proteins (Blindauer, 2015). However, RNDs are not able to result in *α-proteobacteria* significantly reducing Zn concentrations in the soil (Liu *et al.*, 2022), which suggests the need for a more comprehensive explanation of the bioremediation properties of *α-proteobacteria*.

Despite functioning after a Zn^{2+} to Cd^{2+} substitution (Malgieri *et al.*, 2014), ZFs are unlikely to play a big part in creating Cd resistance in soil (Lorenz *et al.*, 2006) and vent (Farias *et al.*, 2015) *α-proteobacteria*, as their role would have a negligible effect on the Cd concentration in their cytoplasm. Coincidentally, *α-*

proteobacteria possess a putative Cd efflux system (Salam *et al.*, 2020), which, by itself, can be responsible for their high Cd tolerance.

The Ros/MucR proteins experience irreversible structural degradation when coordinating Cu^{2+} ions (Dragone *et al.*, 2022) and favour Cu^{+} binding over Zn^{2+}, which also leads to their inactivation (Doku *et al.*, 2013). These properties create an opportunity to use Cu^{+}-vulnerable ZFs in the expression of genes involved in stabilising the Cu^{+} concentration in the cytoplasm, which could help explain the Cu resistance of soil *α-proteobacteria* (Lorenz *et al.*, 2006). The potential use of *α-proteobacteria* in bioremediation may be due to the presence of another mechanism, e.g. Cu-chelating exopolysaccharides on their cellular membrane (Llamas *et al.*, 2010).

Vent *α-proteobacteria* can act as Fe^{3+} sinks, as they developed Fe^{3+} to Fe^{2+} reduction mechanisms (Cohen *et al.*, 2021) (Deng, 2021). However, the Fe^{2+} ions are confirmed to negatively affect α-proteobacterial ZFs (Kluska *et al.*, 2018). This can mean that Fe^{3+} ions overall pose a bigger threat to *α-proteobacteria* than Fe^{2+} ions, or that *α-proteobacteria* are able to precisely control the amount of free Fe^{2+} in their cytoplasm, partially in order to protect their ZF proteins.

The Evolutionary History of Zinc Fingers and Its Implications On Other Taxa

There is compelling evidence in favour of the hypothesis that ZFs evolved from Fe/S clusters in *Proteobacteria*. This is mostly because some proteins, e.g. IscU, create Zn-bound clusters more readily than Fe/S clusters (Ramelot *et al.*, 2004) and others, e.g. CPSF30, display an assumed ancestral trait of favouring the creation of Fe/S clusters over ZF domains in cysteine-rich sites (Shimberg *et al.*, 2016). This result was rather unexpected, as it was anticipated to find evidence that ZFs first appeared in *α-proteobacteria*. Nevertheless, this evolution can be an adaptation to unstable environmental conditions during the diversification of *Proteobacteria* from other phylums, most notably a change of Fe concentration in the oceans; or caused by the need for proteins specifically suited for DNA-binding.

Over time, the hypothesis that eukaryotic ZFs have an α-proteobacterial origin through eukaryogenesis has been strengthening (Moreira *et al.*, 2000). This is because the ZFs from the Ros/MucR protein family were identified in a wide range of bacteria (D'Abrosca *et al.*, 2020), with the ZFs from *α-proteobacteria* being the most similar to the eukaryotic ones (Esposti *et al.*, 2018), and requiring minimal point mutations to transform into them (Netti *et al.*, 2013). On top of this structural argument, the Ros/MucR ZFs create a large regulon in proteobacterial cells (Jiao *et al.*, 2020), meaning that the ZFs are precisely interacting with the evolutionarily old and conserved proteins of the bacterial transcription machinery, which implies their prolonged coevolution and, thus, proves the proteobacterial origin of all ZF proteins.

Improvement Areas and Objectives for Future Investigations

Due to the lack of similarity detected by the blastn algorithm between α-proteobacterial and human genes of ZF proteins, a part of this investigation was confined to analysing only protein sequences. Thus, another search algorithm should be used in future studies aiming to compare α-proteobacterial and human genes. It would have also been very beneficial for this study to use a computer programme that can predict the position of coordination spheres in modelled proteins, and which metal ions they can include. Additionally, *α-proteobacteria* occupying a wider variety of ecosystems should be investigated for the presence of abnormalities in their response to heavy-metal pollution, especially caused by iron. Moreover, in order to come to more confident conclusions about the evolutionary history of ZFs in higher Eukaryotes, a greater amount of complex ZFs should be evaluated and compared to their prokaryotic counterparts. There are also more studies needed to confirm or deny the evolutionary hypotheses proposed in this investigation, since recreating a precise evolutionary history of ZF domains is outside the scope of this investigation. Nevertheless, this research proposes

an intriguing evolutionary scenario, well-suited to the possible functional and environmental causes of this adaptation.

Acknowledgments

I would like to thank my advisor for the valuable insight provided to me on this topic.

References

Altschul S.F. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25(17), pp. 3389-3402. doi: 10.1093/nar/25.17.3389

Baglivo I. *et al.* (2018) MucR binds multiple target sites in the promoter of its own gene and is a heat-stable protein. *FEBS Open Bio* 8(4), pp. 711-718. doi: 10.1002/2211-5463.12411

Bergé M. *et al.* (2016) Modularity and determinants of a (bi-)polarization control system from free-living and obligate intracellular bacteria. *eLife* 5. doi: doi.org/10.7554/eLife.20640

Bertani B., Ruiz N. (2018) Function and biogenesis of lipopolysaccharides. *EcoSal Plus* 8(1). doi: 10.1128/ecosalplus.ESP-0001-2018

Blindauer C.A. (2015) Advances in the molecular understanding of biological zinc transport. *Chem. Commun.* 51(22), pp. 4544-4563. doi: 10.1039/C4CC10174J

Braun V. (1997) Avoidance of iron toxicity through regulation of bacterial iron transport. *Biol. Chem.* 378(8), pp. 779-786. PMID: 9377472

Casalino L. (2017) *Pre-mRNA Splicing: An Evolutionary Computational Journey from Ribozymes to Spliceosome.* Ph.D. thesis, Scuola Internazionale Superiore di Studi Avanzati (SISSA). Available at: https://iris.sissa.it/handle/20.500.11767/59221 [accessed: 6 June 2023].

Caswell C.C. *et al.* (2013) Diverse Genetic Regulon of the Virulence-Associated Transcriptional Regulator MucR in *Brucella abortus* 2308. *Infect. Immun.* 81(4), pp. 1040-1051. doi: 10.1128/IAI.01097-12

Cerqueira T. *et al.* (2015) Microbial diversity in deep-sea sediments from the Menez Gwen hydrothermal vent system of the Mid-Atlantic Ridge. *Mar. Genomics* 24(3), pp. 343-55. doi: 10.1016/j.margen.2015.09.001

Chen H. *et al.* (2000) Structure of a (Cys3His) zinc ribbon, a ubiquitous motif in archaeal and eucaryal transcription. *Protein Sci.* 9(9), pp. 1743-1752. doi: 10.1110/ps.9.9.1743

Chou A.Y., Archdeacon J., Kado C.I. (1998) *Agrobacterium* transcriptional regulator Ros is a prokaryotic zinc finger protein that regulates the plant oncogene *ipt*. *Proc. Natl. Acad. Sci. U.S.A.* 95(9), pp. 5293-5298. doi: 10.1073/pnas.95.9.5293

Cohen N. R. *et al.* (2021) Hydrothermal trace metal release and microbial metabolism in the northeastern Lau Basin of the South Pacific Ocean. *Biogeosciences* 18, pp. 5397-5422. doi: 10.5194/bg-18-5397-2021

Cooley M.B., D'Souza M.R., Kado C.I. (1991) The *virC* and *virD* Operons of the *Agrobacterium* Ti Plasmid Are Regulated by the *ros* Chromosomal Gene. *J. Bacteriol.* 173(8), pp. 2608-2616. doi: 10.1128/jb.173.8.2608-2616.1991

D'Abrosca G. *et al.* (2020) Structural Insight of the Full-Length Ros Protein: A Prototype of the Prokaryotic Zinc-Finger Family. *Sci. Rep.* 10. doi: 10.1038/s41598-020-66204-5

de Lencastre A., Hamill S., Pyle A. (2005) A single active-site region for a group II intron. *Nat. Struct. Mol. Biol.* 12, pp. 626-627. doi: 10.1038/nsmb957

Deng S. (2021) Phosphorus and selected metals and metalloids. In: Gentry T., Zauberer D., Fuhrmann J. *Principles and Applications of Soil Microbiology.* 3rd ed. Amsterdam: Elsevier, pp. 523-555. ISBN: 9780323851404

Doku R.T. *et al.* (2013) Spectroscopic characterization of copper(I) binding to apo and metal-reconstituted zinc finger peptides. *J. Biol. Inorg. Chem.* 18, pp. 669-678. doi: 10.1007/s00775-013-1012-6

Dragone M. *et al.* (2022) Copper (I) or (II) Replacement of the Structural Zinc Ion in the Prokaryotic Zinc Finger Ros Does Not Result in a Functional Domain. *Int. J. Mol. Sci.* 23(19). doi: 10.3390/ijms231911010

Eom K., Cheong J., Lee S. (2016) Structural Analyses of Zinc Finger Domains for Specific Interactions with DNA. *J. Microbiol. Biotechnol.* 26(12), pp. 2019-2029. doi: 10.4014/jmb.1609.09021

Esposti M.D. *et al.* (2018) From *Alphaproteobacteria* to Proto-Mitochondria. In: *Phylogeny and Evolution of Bacteria and Mitochondria.* 1st edition. Boca Raton: CRC Press. ISBN: 9781315144498

Farias P. *et al.* (2015) Natural Hot Spots for Gain of Multiple Resistances: Arsenic and Antibiotic Resistances in Heterotrophic, Aerobic Bacteria from Marine Hydrothermal Vent Fields. *Appl. Environ. Microbiol.* 81(7), pp. 2534-2543. doi: 10.1128/AEM.03240-14
Fatnassi I.C. *et al.* (2015) Impact of dual inoculation with *Rhizobium* and PGPR on growth and antioxidant status of *Vicia faba* L. under copper stress. *C. R. Biol.* 338(4), pp. 241-254. doi: 10.1016/j.crvi.2015.02.001

Gomes N.C.M. *et al.* (2010) Effects of Cd- and Zn-enriched sewage sludge on soil bacterial and fungal communities. *Ecotoxicol. Environ. Saf.* 73(6), pp. 1255-1263. doi: 10.1016/j.ecoenv.2010.07.027

Jian L. *et al.* (2019) Promotion of growth and metal accumulation of alfalfa by coinoculation with *Sinorhizobium* and *Agrobacterium* under copper and zinc stress. *PeerJ* 7. doi: 10.7717/peerj.6875

Jiao J. *et al.* (2016) MucR Is Required for Transcriptional Activation of Conserved Ion Transporters to Support Nitrogen Fixation of *Sinorhizobium fredii* in Soybean Nodules. *Mol. Plant Microbe Interact.* 29(5), pp. 352-361. doi: doi.org/10.1094/MPMI-01-16-0019-R

Jiao J. *et al.* (2022) The zinc-finger bearing xenogeneic silencer MucR in *α-proteobacteria* balances adaptation and regulatory integrity. *ISME J.* 16, pp. 738-749. doi: 10.1038/s41396-021-01118-2

Jiao J., Tian C.F. (2020) Ancestral zinc-finger bearing protein MucR in *alpha-proteobacteria*: A novel xenogeneic silencer? *Comput. Struct. Biotechnol. J.* 18, pp. 3623-3631. doi: 10.1016/j.csbj.2020.11.028

Kamboch B. (2020) Why Fe3+ is more stable than Fe2+? *UO Chemists.* Available at: https://www.uochemists.com/fe3-is-more-stable/ [accessed: 20 February 2023].

Kluska K., Adamczyk J., Krężel A. (2018) Metal binding properties, stability and reactivity of zinc fingers. *Coord. Chem. Rev.* 367, pp. 18-64. doi: 10.1016/j.ccr.2018.04.009

Krishna S.S., Aravind L. (2010) The bridge-region of the Ku superfamily is an atypical zinc ribbon domain. *J. Struct. Biol.* 172(3), pp. 294-299. doi: 10.1016/j.jsb.2010.05.011

Krishna S.S., Majumdar I., Grishin N.V. (2003) Structural classification of zinc fingers: SURVEY AND SUMMARY. *Nucleic Acids Res.* 31(2), pp. 532-550. doi: 10.1093/nar/gkg161

Liu N. *et al.* (2022) Specific bacterial communities in the rhizosphere of low-cadmium and high-zinc wheat (*Triticum aestivum* L.). *Sci. Total Environ.* 838(3). doi: 10.1016/j.scitotenv.2022.156484

Llamas I. *et al.* (2010) Characterization of the Exopolysaccharide Produced by *Salipiger mucosus* A3T, a Halophilic Species Belonging to the *Alphaproteobacteria*, Isolated on the Spanish Mediterranean Seaboard. *Mar. Drugs* 8(8), pp. 2240-2251. doi: 10.3390/md8082240

Lorenz L. *et al.* (2006) Response of microbial activity and microbial community composition in soils to long-term arsenic and cadmium exposure. *Soil Biol. Biochem.* 38(6), pp. 1430-1437. doi: 10.1016/j.soilbio.2005.10.020

Malgieri G. *et al.* (2015) The prokaryotic zinc-finger: structure, function and comparison with the eukaryotic counterpart. *FEBS J.* 282(23), pp. 4480-4496. doi: 10.1111/febs.13503

Malgieri G. *et al.* (2014) Zinc to cadmium replacement in the prokaryotic zinc-finger domain. *Metallomics* 6(1), pp. 96-104. doi: 10.1039/c3mt00208j

Marchler-Bauer A. *et al.* (2017) CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* 45(D1), pp. D200-D203. doi: 10.1093/nar/gkw1129

Martin W. *et al.* (2008) Hydrothermal vents and the origin of life. *Nat. Rev. Microbiol.* 6, pp. 805-814. doi: 10.1038/nrmicro1991
Martínez-Abarca F., Toro N. (2000) Group II introns in the bacterial world. *Mol. Microbiol.* 38(5), pp. 917-926. doi: 10.1046/j.1365-2958.2000.02197.x

Mignolet J. *et al.* (2016) Functional dichotomy and distinct nanoscale assemblies of a cell cycle-controlled bipolar zinc-finger regulator. *eLife* 5. doi: 10.7554/eLife.18647

Mirdita M. *et al.* (2022) ColabFold: making protein folding accessible to all. *Nat. Methods* 19, pp. 679-682. doi: 10.1038/s41592-022-01488-1

Moreira D., Rodrıǵuez-Valera F. (2000) A mitochondrial origin for eukaryotic C2H2 zinc finger regulators? *Trends Microbiol.* 8(10), pp. 448-449. doi: 10.1016/s0966-842x(00)01850-3

National Center for Biotechnology Information (2023) *Conserved Domain Search.* Available at: https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi [accessed: 20 February 2023].

National Center for Biotechnology Information (2023) *Taxonomy browser (root).* Available at: https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi [accessed: 20 February 2023].

Netti F. *et al.* (2013) An Experimentally Tested Scenario for the Structural Evolution of Eukaryotic Cys_2His_2 Zinc Fingers from Eubacterial Ros Homologs. *Mol. Biol. Evol.* 30(7), pp. 1504-1513. doi: 10.1093/molbev/mst068

Nwodo U.U., Green E., Okoh A.I. (2012) Bacterial exopolysaccharides: functionality and prospects. *Int. J. Mol. Sci.* 13(11), pp. 14002-14015. doi: 10.3390/ijms131114002

Py B., Barras F. (2010) Building Fe–S proteins: bacterial strategies. *Nat. Rev. Microbiol.* 8. doi: 10.1038/nrmicro2356

Ramelot T. *et al.* (2004) Solution NMR Structure of the Iron–Sulfur Cluster Assembly Protein U (IscU) with Zinc Bound at the Active Site. *J. Mol. Biol.* 344(2), pp. 567-583. doi: 10.1016/j.jmb.2004.08.038

Richards T.A., Archibald J.M. (2011) Cell Evolution: Gene Transfer Agents and the Origin of Mitochondria. *Curr. Biol.* 21(3), pp. 112-114. doi: 10.1016/j.cub.2010.12.036

Roth D.B. (2015) V(D)J Recombination: Mechanism, Errors, and Fidelity. In: Craig N.L. eds. *Mobile DNA III.* 3rd edition. Washington, D.C.: ASM Press, pp. 311-324. doi: 10.1128/microbiolspec.MDNA3-0041-2014

Salam L.B. *et al.* (2020) Effects of cadmium perturbation on the microbial community structure and heavy metal resistome of a tropical agricultural soil. *Bioresour. Bioprocess.* 7(1). doi: 10.1186/s40643-020-00314-w

Shannon R.D. (1976) Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Cryst.* A32(5), pp. 751-767. doi: 10.1107/S0567739476001551

Shimberg G.D. *et al.* (2016) Cleavage and polyadenylation specificity factor 30: AnRNA-binding zinc-finger protein with an unexpected 2Fe–2S cluster. *Proc. Natl. Acad. Sci. U.S.A.* 113(17), pp. 4700-4705. doi: 10.1073/pnas.1517620113

Shu W. *et al.* (2012) Abundance and diversity of nitrogen-fixing bacteria in rhizosphere and bulk paddy soil under different duration of organic management. *World J. Microbiol. Biotechnol.* 28, pp. 493-503. doi: 10.1007/s11274-011-0840-1

Sun J. *et al.* (2021) Deletion of the Transcriptional Regulator MucR in *Brucella canis* Affects Stress Responses and Bacterial Virulence. *Front. Vet. Sci.* 8. doi: 10.3389/fvets.2021.650942

Tahirov T.H. *et al.* (2009) Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol ε and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors. *Biol. Direct.* 4(11). doi: 10.1186/1745-6150-4-11

Wagner A. *et al.* (2017) Mechanisms of gene flow in archaea. *Nat. Rev. Microbiol.* 15, pp. 492-501. doi: 10.1038/nrmicro.2017.41

Wagner S.C. (2011) Biological Nitrogen Fixation. *Nature Education Knowledge* 3(10). Available at: https://www.nature.com/scitable/knowledge/library/biological-nitrogen-fixation-23570419/ [accessed: 20 February 2023].

Waldron K., Robinson N. (2009) How do bacterial cells ensure that metalloproteins get the correct metal?. *Nat. Rev. Microbiol.* 7, pp. 25-35. doi: 10.1038/nrmicro2057

Wang Y. *et al.* (2008) Assessment of microbial activity and bacterial community composition in the rhizosphere of a copper accumulator and a non-accumulator. *Soil Biol. Biochem.* 40(5), pp. 1167-1177. doi: 10.1016/j.soilbio.2007.12.010

Wyszkowska J. *et al.* (2013) Effect of cadmium, copper and zinc on plants, soil microorganisms and soil enzymes. *J. Elem.* 18(4), pp. 769-796. doi: 10.5601/jelem.2013.18.4.455

Yu X. *et al.* (2021) Cadmium Pollution Impact on the Bacterial Community Structure of Arable Soil and the Isolation of the Cadmium Resistant Bacteria. *Front. Microbiol.* 12. doi: 10.3389/fmicb.2021.698834

Zhiguo E., Wang L., Zhou J. (2013) Splicing and alternative splicing in rice and humans. *BMB Rep.* 46(9), pp. 439-47. doi: 0.5483/BMBRep.2013.46.9.161

Fraser C.M., Read T.D., Nelson K.E. (2014) *Microbial Genomes.* 1st edition. Totowa: Humana Press. ISBN: 1588291898

Gargaud M., López-Garcìa P., Martin H. (2011) *Origins and Evolution of Life: An Astrobiological Perspective.* Cambridge: Cambridge University Press. ISBN: 9780521761314

Kahl G. (2001) *The Dictionary of Gene Technology: Genomics, Transcriptomics, Proteomics.* 2nd edition. Weinheim: Wiley-VCH. ISBN: 3527307656

Latchman D.S. (1995) *Eukaryotic Transcription Factors.* 2nd edition. London: Academic Press. doi: 10.1042/bj2700281

Liu J. (2018) *Zinc Finger Proteins: Methods and Protocols.* 1st edition. Totowa: Humana Press. ISBN: 1493993925

Lucchesi J.C. (2019) *Epigenetics, Nuclear Organization & Gene Function: With implications of epigenetic regulation and genetic architecture for human development and health.* 1st edition. Oxford: Oxford University Press. ISBN: 9780198831204

Rédei G.P. (2008) *Encyclopedia of Genetics, Genomics, Proteomics, and Informatics: Proteomics, Volume 6.* 3rd edition. Berlin: Springer Science+Business Media. doi: 10.1007/978-1-4020-6754-9

Rosenberg E., eds. (2014) *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria, Volume 8.* 4th edition. Berlin: Springer. ISBN: 9783642301964

Weiner J., Weiner (3) J. (2018) *Technika pisania i prezentowania przyrodniczych prac naukowych.* 5th edition. Warszawa: Polskie Wydawnictwo Naukowe PWN. EAN: 9788301200404

Węgleński P. (2006) *Genetyka molekularna.* 6th edition. Warszawa: Polskie Wydawnictwo Naukowe PWN. ISBN: 9788301147440