scientific reports

Check for updates

OPEN

Functional flexibility in bacterial hub proteins is driven by proteome expansion

Mladen Paradžik[1], Stefani Prekpalaj[2] & Tina Paradžik[2]✉

Protein-protein interaction networks play a key role in the living cells, with hub proteins (hubs) acting as central nodes in the transmission of biological signals. The features enabling hubs to interact with other proteins in bacteria remain unexplored. We present the first cross-phylum analysis of bacterial hubs using AlphaFold and IUPred. Hubs exhibit higher levels of intrinsic disorder compared to whole proteomes. In contrast to the highly structured proteomes of obligate pathogens, proteomes of free-living bacterial species are more disordered. We show that proteome size is a stronger predictor of disorder accumulation than GC content. Analyses of conserved orthologous hubs show that protein cores are conserved in sequence, 3D structure and length, accumulating disorder in terminal regions in bacteria with large proteomes. Analysis of the *Escherichia coli* interactome shows that 63% of proteins with disorder (> 30 AA) have more than 10 known interaction partners, compared to 19% of fully structured proteins, emphasizing the functional importance of disorder in bacteria. Our results indicate that intrinsic disorder in bacterial proteins is associated with proteome complexity and cannot be explained by GC content alone. Disorder expansion in complex bacteria may contribute to interaction plasticity in different environments similar to those in eukaryotes.

Complex protein-protein interaction (PPI) networks play a essential role in the life processes of all organisms, including bacteria[1–6]. The study of PPIs is crucial to understand the organization and function of biological systems. Hub proteins (hubs) have a large number of interactants and are critical in maintaining the structural and functional integrity of the PPI network[7]. Disruption of hub-encoding genes often leads to severe phenotypic consequences[8]. The characteristics of eukaryotic hubs have been extensively studied for several decades, exhibiting extensive disordered domains[7,9]. Disordered regions facilitate flexible interactions with multiple partners[10,11]. Intrinsic disorder can be found in proteins in all life domains, but is most abundant in eukaryotes[7,12,13]. Disordered proteins or disordered regions within proteins can have structural, functional, and regulatory functions and play a crucial role in PPI networks[14,15].

In contrast, the specific characteristics that enable prokaryotic proteins to function as hubs remain poorly explored. Compared to their eukaryotic counterparts, prokaryotic hubs are generally shorter, have fewer long repetitive regions and less intrinsic disorder[16,17]. Prokaryotic hubs are often associated with essential cellular machinery[18] and their role is mainly studied in processes such as cell division and in model organisms such as *Escherichia coli*[19]. Disorder in prokaryotic proteins has been investigated by several groups[12,13]. The GC content of a genome has a direct effect on the amino acid (AA) composition of the proteins it encodes, as certain AA are encoded by GC-rich codons[13,20]. This correlation suggests that organisms with higher GC content produce proteins with more disordered regions[20]. Contrary, genomes with low GC content may favour structured proteins suitable for certain pathogenic or symbiotic lifestyles[13]. Genome composition influences AA usage and protein disorder, indicating optimization of proteomes to balance structure with environmental and ecological demands.

Previously, disorder in proteins was commonly analysed using sequence-based predictors such as IUPred, which is based on protein AA composition[21]. In *silico* prediction using AlphaFold revolutionised the field of structural biology and biology in general[22]. AlphaFold pLDDT scores have been explored as an indicator of intrinsic disorder, showing that low-confidence regions often correspond to experimentally validated disordered regions, despite some limitations[23–25].

In this study, we used IUPred and AlphaFold to analyse proteomes and hubs from phylogenetically diverse bacteria, including important pathogens, model organisms, industrially significant species, and extremophiles.

[1]Laboratory of Experimental Therapy, Division of Molecular Medicine, Ruđer Bošković Institute, Zagreb, Croatia. [2]Laboratory for chemical and biological crystallography, Division of physical chemistry, Ruđer Bošković Institute, Zagreb, Croatia. ✉email: Tina.Paradzik@irb.hr

We found that bacterial hubs have increased length and more intrinsic disorder compared to the overall proteome, that is associated with GC content and, even more strongly, by proteome size. Free-living bacteria with large, GC-rich genomes display elevated disorder levels in both hubs and total proteomes, while obligate pathogens have very low levels of disorder in their total proteomes. However, their hubs still retain a certain level of disorder, especially as short disorder (SD) stretches, indicating possible role in PPI networks. Despite the evolutionary conservation of hubs, bacteria with large proteomes retain terminal long disordered (LD) regions that are poorly conserved at the sequence level. This pattern suggests selective constraints maintaining flexibility, possibly to accommodate more complex interaction networks. We propose that, in large proteomes, selection for flexible interaction capacity may contribute to elevated GC content, not only in hubs but across the proteome. Conversely, host-dependent bacteria living in specialised ecological niches have reduced proteomes, and their interaction networks may not require such structural plasticity, resulting in decreased disorder accumulation.

Results

Dataset construction and functional annotation of hubs

The species dataset ($n = 40$) comprised a diverse selection of model bacteria from different bacterial taxa (Supplementary File 1 Figure S1). The initial species selection was based on the availability of interactome data from the APID and IntAct databases. This list was expanded to ensure a broader taxonomic spectrum and variation in proteome sizes and GC content despite the unavailability of experimental interactomes. All species included in the dataset have high quality reference proteomes available in the UniProt.

Three datasets were selected and examined for each species: the whole proteome (all annotated proteins), hubs, and non-hubs, as described in the Methods section (Sect. 5.1). A total of 7893 hubs and 5094 non-hubs were analysed in this study (Supplementary file S2). Dataset of hubs was selected individually for each species. Since the selection of hubs depends on the available data in the databases, not all homologues are selected in all species. This contributes to the overall robustness of the analysis as non-homologous hubs contribute to the results. Hubs were found to belong predominantly to essential cellular processes such as translation, nucleic acid metabolism, and signal transduction, consistent with their expected central roles in bacterial PPI networks (Supplementary file 1 Figure S3). For selection of non-hubs, we did not use the entire set of proteins with few interactions but instead applied a stringent filtering strategy resulting in high-confidence subset of non-hubs.

Expansion of hub protein length correlates with proteome complexity in bacteria

The calculated average size of proteins in the proteome is consistent with previous studies[13], showing that hubs are significantly longer than proteins in whole proteome (Fig. 1). A positive correlation of hub length with proteome size can be observed ($R^2 = 0.42$ $p < 0.0001$), with a slope that differs significantly from proteome and non-hub datasets ($p < 0.0001$) (Fig. 1B). This dataset also includes ribosomal proteins, that have highly conserved nature[26] and small size (typically < 300 AAs) that contributes to greater variation in the analysis.

Disorder as a functional feature of bacterial hubs in different contexts

Intrinsic protein disorder plays an indispensable role in the functional versatility and adaptability of proteomes in living organisms[9]. We have investigated the overall presence and distribution of disordered regions in bacterial species with different GC content and proteome sizes, to analyze hubs in comparison to total proteomes. The

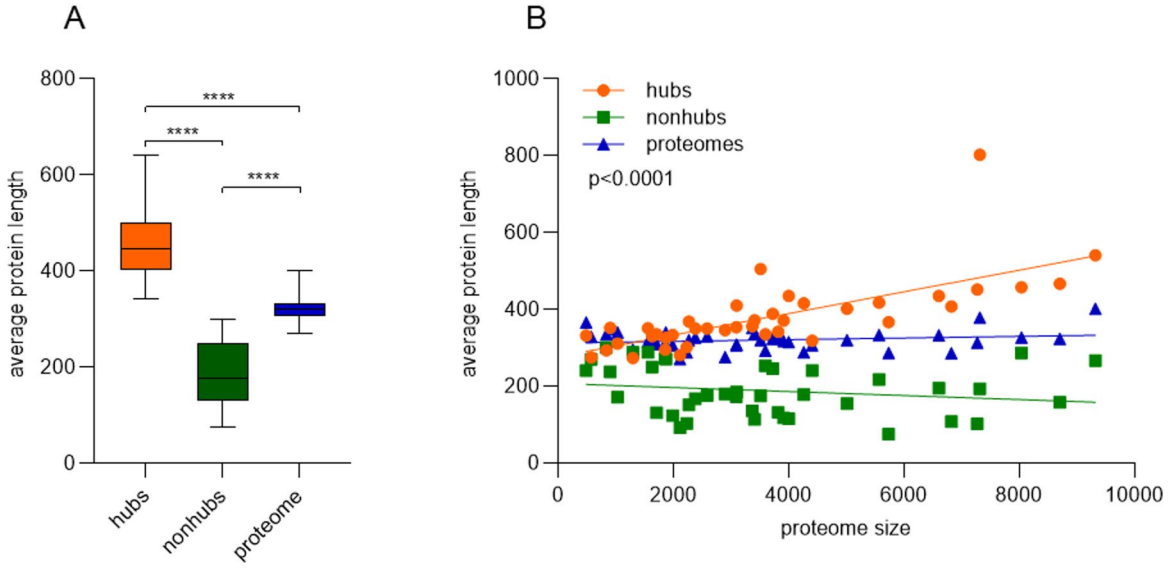

Fig. 1. Comparison of protein length in bacterial proteomes, hubs and non-hubs. (**A**) Hub proteins are significantly longer than the total proteome and non-hub proteins ($p < 0.0001$, ANOVA). (**B**) Hubs, non-hubs and proteome average protein length versus proteome size. The slope of the hubs with ribosomal proteins excluded differs significantly ($p < 0.0001$) from the slopes of proteome and non-hubs.

overall level of disorder in the proteomes is consistent with previous reports[12,13]. In addition, our results show that more hubs contain disordered regions than proteomes or non-hubs (Fig. 2A). The increased prevalence of proteins with disordered regions within the hub datasets probably has a functional significance, as this pattern is observed in different bacterial species.

In Fig. 2B the abundance of disordered proteins is shown for different GC contents in phylogenetically diverse bacteria, being highest for bacteria with high GC content. This group includes species from different taxa, such as Actinomycetota, α-Proteobacteria, Deinococcota, Myxococcota and Planctomycetota. Bacteria with low GC content are also represented by diverse bacterial species (ε-Proteobacteria, Spirochaetota, α-Proteobacteria and γ-Proteobacteria, Mycoplasmatota and Bacillota) that all have a smaller number of disordered proteins. Although phylogenetically distant from each other, these bacteria have similar GC content and proteome size. They also share similar trends in the disorder presence in proteomes and hubs. In line with previous reports[13], bacteria with high GC content have more disordered proteins than bacteria with low GC content (Fig. 2B). However, it can be noticed that hubs always maintain more disorder than general proteome, and this trend is particularly visible in low GC (Fig. 2B and C) and in small proteome bacteria (Fig. 2D). This suggests that low GC organisms have less disordered proteins than high GC bacteria, but their hubs are more disordered with respect to the overall proteome when comparing difference of low and high GC groups. In addition, when proteome size is plotted against disorder, it can be observed that in small proteomes the number of disordered proteins is limited in all datasets, and the hub slope changes when small proteomes (< 2000 proteins) are filtered out (Supplementary File 1 Figure S4, right panel). The correlation of hubs with proteome size becomes non-significant ($p = 0.167$), whereas the correlation for the whole proteome remains strong ($p = 0.001$).

Free-living bacteria generally have a higher GC content and a larger genome size, while obligate pathogens have reduced genome size with low GC content[27]. The causes of this GC content variability remain under scientific debate[20,28,29]. Genomic GC content influences AA composition and thereby the propensity for intrinsic

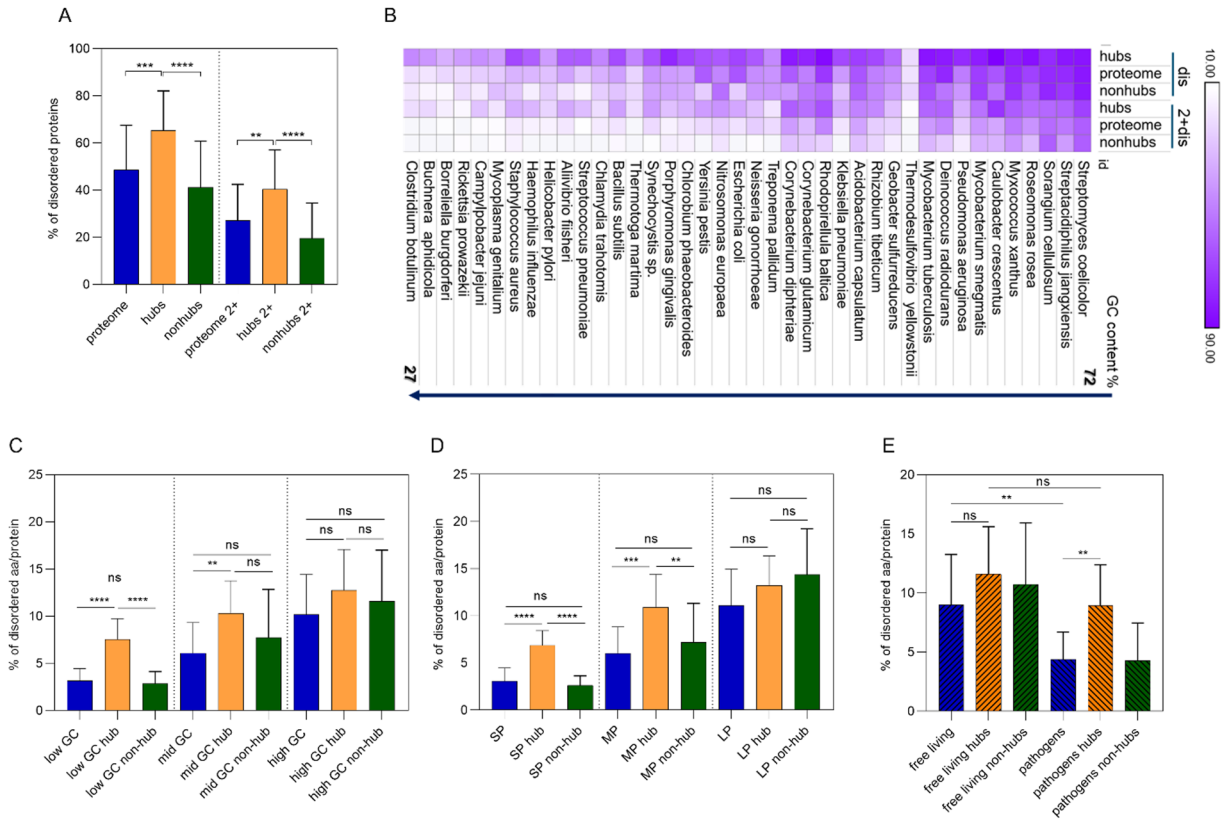

Fig. 2. Analysis of the presence of disordered regions in hub, non-hub and proteome datasets from different bacterial species. (**A**) Comparison of the percentage of proteins with disorder in proteomes, hubs and non-hubs. "2+" means 2 or more disordered stretches in the single protein sequence. (**B**) Heatmap of disordered proteins in different species categorized by GC content. The intensity of purple colour indicates the proportion of disorder. (**C**) Frequency of disordered AAs per protein across GC content categories: low (< 40%), mid (40–60%), and high (> 60%). (**D**) Disordered AAs per protein across proteome size categories: small, SP (< 2000 proteins), medium, MP (2000–5000 proteins), and large, LP (> 5000 proteins). (**E**) Difference in percentage of disordered AAs between hubs and the full proteome in free living and obligate/facultative pathogens. Statistical significance is indicated as $^{*}p < 0.05$, $^{**}p < 0.01$ and $^{***}p < 0.001$, $^{****}p < 0.0001$, "ns" no significance. Details of statistical analysis are presented in Supplementary file 1 Table S1 and S2.

disorder. However, our results indicate that although GC content contributes to variation in disorder, it does not fully explain the observed patterns.

We analysed how the degree of disorder relates to lifestyle of bacteria within our datasets. When comparing disorder in pathogens/specialists' hubs versus free-living bacteria, we have not observed a significant difference (Fig. 2E). However, significant difference in the amount of disordered AA in proteomes has been noticed when these two groups of bacteria are compared, being higher for free living bacteria. Despite the lower level of disorder in proteomes of pathogens (Fig. 2E), the frequency of disorder in hubs is higher compared to the whole proteome, while the difference (hubs vs. proteome) is statistically significant contrary to difference in free-living bacteria. This again shows hubs' tendency to maintain disordered regions and highlights the functional importance of disorder, despite genome reduction in pathogens/specialists.

In conclusion, the level of disorder in bacterial proteins correlates positively with proteome size and GC content, while the difference in the abundance of disorder in hubs and total proteomes is significantly higher in organisms with low GC content.

Expansion of intrinsic disorder in bacterial hubs is constrained by proteome size and modulated by GC content

To go more into the length and abundance of different disorder lengths, the presence and frequency of long (stretches > 30 AA) and SDs (5–15 AA) were analysed for each dataset and species. Previously, it was found that the highest frequency of disordered regions in all phyla of Archaea and Bacteria have segments of length 1–10 AA[13]. Long (> 30 residues) disordered segments were found in 2.0% of archaeal, 4.2% of eubacterial and 33.0% of eukaryotic proteins[13,14,30]. In addition to this, our results indicate that SD is more abundant in hubs compared to proteomes and is dependent on proteome size and GC content (Fig. 3, Supplementary file 1, Figure S5). To emphasise the relative enrichment of disorder in hubs compared to proteomes, we have expressed the presence of SD as the difference (Δ) in percentage of disordered hubs and disordered proteins in the proteome. It appears that hubs retain SD in GC low bacteria even though this type of disorder is rare in general proteome. This indicates that, although low-GC bacteria contain fewer disordered proteins overall than high-GC bacteria, their hubs remain relatively more disordered than the corresponding proteome background, resulting in a larger hub–proteome difference (Fig. 3A). This is consistent with Fig. 2C and indicates that SD stretches may contribute to hub function, potentially by supporting local flexibility and transient interactions in all bacteria. Frequency of short, disordered stretches/protein indicates increased accumulation in hubs of large proteome

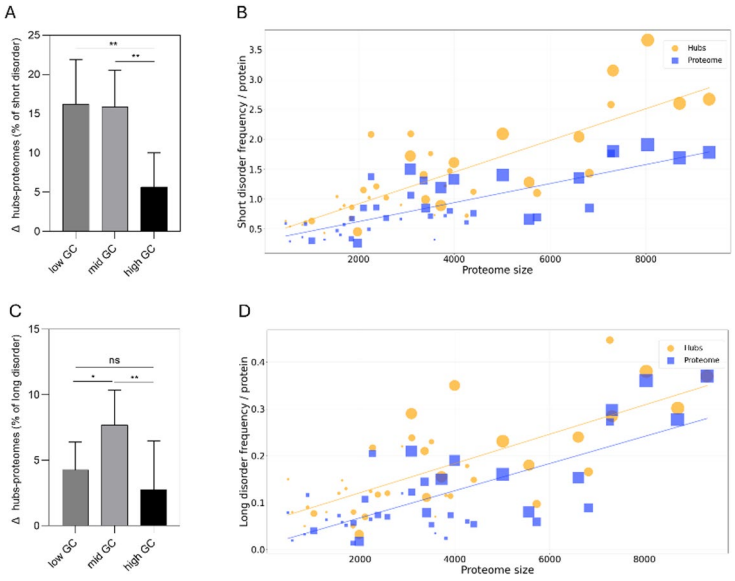

Fig. 3. GC content and proteome size affect short and long disorder accumulation in bacterial hubs and proteomes. (**A**) Difference in the percentage of proteins containing short disordered regions (5–15 aa) between bacterial hubs and their corresponding total proteomes across three GC-content categories. Values are shown as hub percentage minus total proteome percentage for each group (**B**) Relationship between short disorder frequency per protein and proteome size in bacterial hubs and total proteomes. Each point represents one species; orange circles indicate hubs and blue squares indicate total proteomes. Larger symbols correspond to higher GC content. Lines indicate fitted linear trends for each group. (**C**) Difference in the percentage of proteins containing long disordered regions (> 30 aa) between bacterial hubs and their corresponding total proteomes across three GC-content categories. Values are shown as hub percentage minus total proteome percentage for each group. (**D**) Relationship between long disorder frequency per protein and proteome size in bacterial hubs and total proteomes. Each point represents one species; orange circles indicate hubs and blue squares indicate total proteomes. Larger symbols correspond to higher GC content. Lines indicate fitted linear trends for each group. GC categories: low < 40%, mid 40–60%, high > 60%. Statistical significance: $p < 0.05$ (*), $p < 0.01$ (**), ns = not significant. Details of statistical analysis are presented in Supplementary file 1 Table S3.

bacteria (Fig. 3B). The plots show that the accumulation of SDs in eubacterial hubs and proteomes is strongly dependent on both GC content (Supplementary Figure S5) and proteome size (Fig. 3B, Supplementary file 1 Table S3). In addition, hubs are showing different patterns of SD accumulation compared to proteomes with significantly different slopes ($p = 0.0079$) (Fig. 3B).

Figure 3B shows that all analysed bacteria with small proteomes have low amount of SD stretches, despite GC content. The size of bubbles/squares indicates that bacteria with higher GC content in their genomes but smaller proteomes have low amounts of SD. Interestingly, the average frequency of SD stretches in hubs is low in smaller proteomes even if they have high GC content (Fig. 3B), suggesting that a small proteome limits the significant expansion of even small, disordered stretches despite increased GC content. Examples of these are *T.pallidum* (GC 52%, 1027 proteins), *T.yellowstonii* (GC 60.9%, 1982 proteins), *P.gingivalis* (GC 48%, 1863 proteins), all of which have less than 1 SD/protein both in hub and proteome datasets.

LD is less common in bacterial proteomes but still shows a dependence on proteome size and GC content (Fig. 3D, Supplementary file 1 Figure S5 right panel and Table S3). In bacteria with low GC there is a small difference in the percentage of proteins with LD in total proteomes and hubs (Fig. 3C). In these bacteria proteins with LD are rare, indicating evolutionary pressure against the presence LD (even in hubs). The stronger difference in the mid-GC group may reflect an intermediate regime in which LD is enriched in hubs but has not yet expanded to the same extent across the whole proteome; by contrast, LD is rare in low-GC bacteria and more broadly distributed in high-GC proteomes. Although LD is more frequent in hubs than in proteomes, the plot shows a different dynamic to that of SD (Fig. 3D). The accumulation of multiple LD segments is not frequent in bacterial proteins. In fact, many known bacterial hubs have only a single N-terminal or C-terminal disorder region[31–35]. Analysis of disorder position shows that LD is most predominantly found in C-terminal region of bacterial hubs, while SD is uniformly distributed through AA sequence (Supplementary file 1, Figure S6). Statistical analysis indicates that GC content and proteome size together contribute to variations of frequency of LDs by more than 50% in hubs and total proteomes (Supplementary file 1, Table S3). However, proteome size may play a stronger role in predicting LD than GC content since some bacteria with higher GC content, but small proteomes still have low LD frequencies, while all bacteria with small proteomes, especially those with less than 2000 proteins always have low amount of LD (Fig. 3D). This is supported by multiple regression analysis indicating that contribution of GC content is not significant for LD in hubs ($p = 0.430$), and in proteomes this significance is borderline ($p = 0.068$). On the contrary, the effect of proteome size is highly significant both in hubs ($p < 0.001$) and in proteomes ($p = 0.002$). This is probably consistent with the observation that in small proteomes there is a strong selection against LD presence, despite GC content.

Proteome complexity drives terminal disorder expansion in conserved bacterial hubs

To inspect more closely finding that hubs from high GC/large proteomes are in general longer and more disordered than their homologs from low GC/small proteome bacteria, especially at protein termini, we analysed conserved ortholog hubs involved in various essential cellular processes that are present in majority of inspected bacterial species (Supplementary file S2) by using both IUPred and Alphafold pLDDT scores. Both thresholds (< 50 and < 70) showed similar trends, indicating that low pLDDT fraction increase in both protein termini more strongly than in protein core in dependence on proteome size (Fig. 4A and B). While Fig. 4A indicates this increase within last and first 50 AA, Fig. 4B demonstrates the same trend when the first and last 20% of the protein sequence are considered. In addition, analysis of the first and last 20% of protein sequence indicate that this trend increases progressively with proteome size, while the low pLDDT fraction remains similar in central 50% core region, indicating that disorder expansion could be specifically confined to protein termini rather than distributed uniformly across the sequence. However, since pLDDT is not a direct measure of intrinsic disorder, we interpret these results as complementary evidence to IUPred analyses presented throughout the manuscript.

Previously, various proteins involved in DNA metabolism and cell division were found to have long terminal disordered regions, such as FtsZ and SSB[36–38]. Our results show that hubs have LD more frequently at C-terminus (Supplementary file 1 Figure S6). That is further confirmed by the analysis of low pLDDT fractions in different protein regions (Fig. 4A and B). Generally, DNA-binding proteins were found to have increased disorder content[39]. In our ortholog dataset, DNA-binding proteins had the lowest overall pLDDT scores, compared to RNA binding proteins and other involved in a variety of processes (Supplementary file 1 Figure S7). Both DNA and RNA binding proteins are showing more dependence of proteome size than other proteins.

Analysis of disorder presence in five proteins belonging to different cellular pathways in high GC, medium GC and low GC species reveal the presence of extensive C- terminal disordered region in free living bacteria, *S.coelicolor*, and gradual increase from small proteome to large proteome bacteria (Fig. 4C). In the AlphaFold predictions (Fig. 4C), the orange-coloured parts of the AA chains show regions predicted with low structural confidence, often corresponding to intrinsically disordered regions that lack stable secondary or tertiary structure[40]. These regions are typically unresolved in crystal structures and thus represent obscure regions of the proteome that are poorly understood yet functionally crucial in mediating interactions. These findings are further supported by the IUPred analysis (Supplementary file 1 Figure S8) showing the same trends as AlphaFold scores.

These examples illustrate how homologous hubs originating from free living and obligate pathogen or endosymbiont exhibit different disorder levels and hub protein size. Importantly, as visualised by AlphaFold 3D prediction (Fig. 4C) and protein alignments (Supplementary file 1 Figure S9), the protein core region is conserved and similar in length contrary to C-terminal end.

Long disorder and interaction network complexity in bacteria

To further investigate the relationship between proteome size, GC content, LD, and involvement of disordered proteins in interaction networks in bacteria, we analyzed proteins from a model bacterium with a mid-size

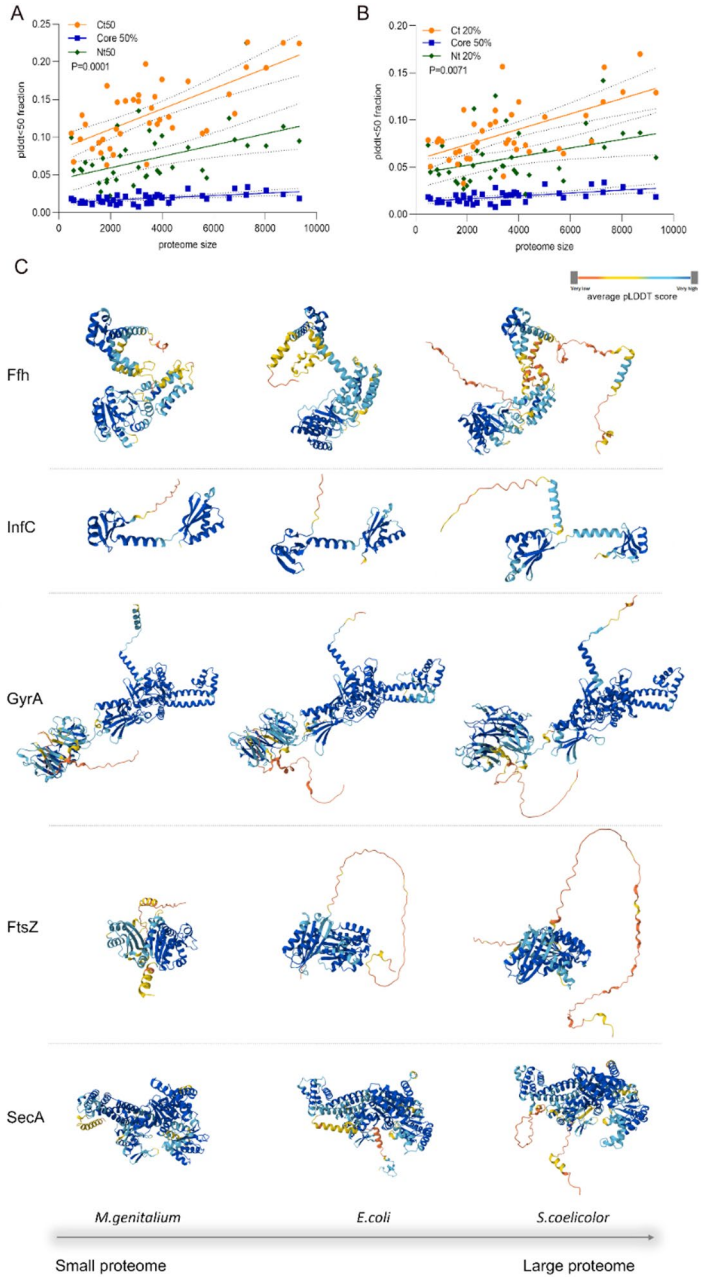

Fig. 4. **A** and **B**. Fractions of residues with low pLDDT (< 50) in terminal and core regions of 49 conserved bacterial orthologs plotted against proteome size. Results of statistical analysis are shown in Supplementary Table S3. Ct50, Nt50- last and first 50 AAs, Ct20, Nt20- last and first 20% of AA in protein sequence. **C**. 3D models of Ffh, InfC, GyrA, FtsZ and SecA proteins in three bacterial species with increasing proteome size. Protein structures highlight ordered (blue) and disordered (yellow/orange) regions, visualized using structural models built by AlphaFold.

proteome (*E. coli*) and a large proteome, high-GC bacterium (*S. coelicolor*). *E. coli* was selected due to the availability of extensive interactome data, while *S. coelicolor* has relatively well-studied interactome compared to other free-living complex bacteria.

E. coli has a 50% GC content in the genome and encodes around 4401 proteins. Within its proteome, 191 proteins were found to contain at least one LD region (> 30 AA in a row). Proteins with LD were inspected using GO Slim analysis. Molecular function (MF) GO terms found for these proteins are shown in Supplementary Table S4, with highest frequency of hydrolase and nucleic acid binding functions. Biological process (BP) GO terms (Fig. 5A), showing the presence of disordered proteins in essential processes, such as transport, nucleic acid metabolic process, response to stress and cell division.

In contrast to *E. coli*, *S. coelicolor* exhibits substantially higher levels of intrinsic disorder. A total of 2890 LD regions were identified in 2071 proteins i.e. 25% of the proteome contains at least one LD region—a much higher

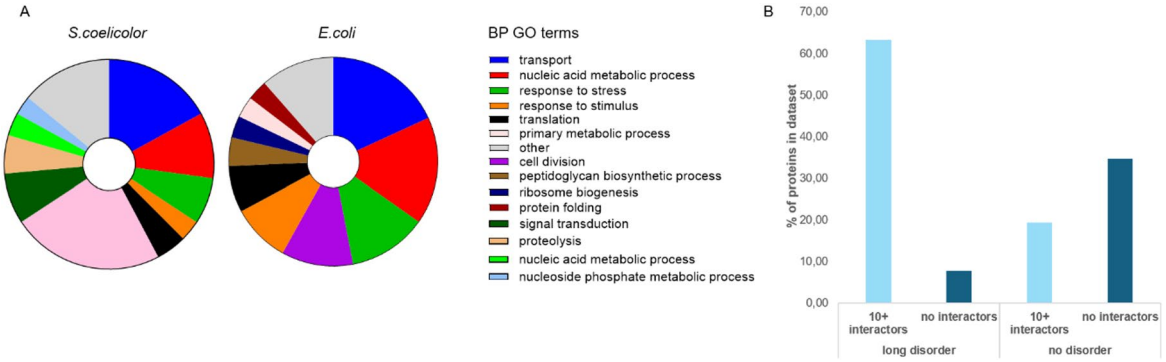

Fig. 5. **A**. Functional classification of proteins with long disordered regions in *Streptomyces coelicolor* and *Escherichia coli* by GO enrichment analysis. **B**. Analysis of interaction network of disordered and non-disordered proteins from *E. coli*, as retrieved from APID database.

proportion than the 4% observed in *E. coli*. Beyond much higher overall abundance of disordered proteins, additional functional categories of GO terms in MF category appear *in S. coelicolor*, such as transcription and regulatory activity (Supplementary Table S5). Two BP GO terms are found with high abundance: primary metabolism and signal transduction (Fig. 5A). To test whether LD is important for protein function, we analysed the *E. coli* interaction network using the experimental databases APID and IntAct. Of the 191 disordered proteins, only 15 proteins had no identified interactors (Fig. 5B). Moreover, most proteins (121) had 10 or more interactors in the APID database, suggesting that proteins with LD regions frequently display hub-like behavior in interaction networks. As a control, we randomly selected 596 proteins without disordered regions and assessed their interaction data in APID and IntAct. Only 19.46% of non-disordered proteins in APID had 10 or more interactors (compared to 63.35% for disordered proteins). Only 12.25% of the non-disordered proteins in IntAct had 10 or more interactors (Fig. 5B). Conversely, the proportion of non-disordered proteins without any interactors was much higher (34.73% in APID and 46.81% in IntAct) compared to disordered proteins (7.85% in APID, 9.26% in IntAct).

This is a strong indicator that LD regions are associated with PPI connectivity in bacteria, supporting the hypothesis that disorder contributes to interaction flexibility and network connectivity. These findings suggest that the presence of LD in large bacterial proteomes, particularly in free-living high-GC organisms, is associated with complex lifestyles and interaction networks rather than being explained as a byproduct of genome growth.

Discussion

Experimental interactome data for bacteria are scarce compared to eukaryotic model species, making network-level analysis challenging. In addition, large bacterial proteomes frequently contain hundreds or even thousands of hypothetical proteins without any interaction data. For example, in the proteome of the planctomycete *Rhodopirellula baltica* (7271 predicted proteins), 2784 proteins are annotated as hypothetical. Only 1% of these proteins have interaction partners in the STRING database under the conditions described in Methods section. The higher proportion of proteins with zero recorded interactions in some large proteomes (Supplementary file 1, Figure S2) likely reflects incomplete interactome coverage and the abundance of hypothetical or poorly characterized proteins, rather than true absence of biological interactions.

Therefore, to maintain data balance, the species dataset was kept relatively small, as high-quality proteomes with well-annotated interactomes are very limited. Consistent with this, a study of the bacterial meta-interactome found only 429 PPIs conserved in 2 or more species indicating that actual differences in genetics and physiology or experimental errors make interaction predictions difficult[41].

Despite these limitations, by combining disorder predictions (IUPred), machine-based structural confidence measures (AlphaFold pLDDT), and regression modelling, we reveal consistent patterns in disorder distribution across both hubs and whole proteomes — providing important insights into bacterial PPI organization.

We found that both hub protein length and disorder content increase with proteome size, particularly in GC-rich, free-living bacteria. (Figures 1B and 2). Compared to eukaryotic hubs that often have multiple forms and domains, as well as increased disorder content[17,42–44], the level of disorder in bacterial hubs is significantly lower. However, hubs still have significantly higher disorder than the species' own total proteome, consistent with the conservation of disorder in contexts where interaction flexibility may be advantageous.

Complex filamentous bacteria such as *Streptomyces* species can produce various secondary metabolites that mediate interactions with other microorganisms in their habitat[45,46]. Thus, an increased level of disorder may contribute to the ability of hubs to interact with a larger number of proteins in complex bacteria. On the other hand, bacteria with small genomes have a low number of genes limited to essential genes[27]. Small genomes have limited amount of disorder in proteins, which is a pattern consistent with constraints associated with genome reduction and structural stability. However, we have observed that hubs maintain SD stretches even in low GC/small genome bacteria, while the presence of disorder is scarce across proteome (Fig. 2). Based on these results, the intrinsic property of hubs to interact with and modulate extensive protein networks could be at least partially supported by the presence of disordered regions in bacteria. Given that hubs play a crucial role in protein

interaction networks, one might expect that free-living bacteria have more proteins with flexible, disordered regions accommodating a larger number of interaction partners. Expansion of proteomes is followed by the emergence of a whole repertoire of new regulatory and signaling proteins, that support the complex lifestyles of such bacteria[45,47].

SD stretches are common in bacterial proteins, particularly in hubs, suggesting possible functional relevance, as their abundance significantly increases with proteome size (Fig. 3A and B, Supplementary file 1 Figure S5). These regions, often associated with flexible linkers and transient interactions are well studied in multicellular organisms[48], could also play an important role in hubs of organisms with large proteomes. For example, the bacterial hub protein PopZ contains an N-terminal intrinsically disordered region with a short MoRF-like segment that can bind to other proteins, such as RcdA[49]. Other examples are FtsZ and SSB, where SD C-terminal stretches mediate protein interactions. In FtsZ, the conserved C-terminal peptide binds division proteins such as FtsA and ZipA[50], whereas in SSB, the disordered acidic tail contains a short terminal motif that recruits numerous SSB-interacting proteins[51].

LD regions are also more frequent in large proteomes, being present in one-third of proteins (Fig. 3B and C). These regions are most commonly located at protein C-termini (Supplementary file 1 Figure S6), which is in accordance with observations that disordered regions are under relaxed structural constraints and can evolve more rapidly while preserving function, especially in hubs involved in multiple transient interactions[52].

Taken together, these observations suggest that both SD and LD regions may contribute to bacterial hub function, with SD likely supporting local flexibility or short interaction elements and LD providing more extended flexible regions associated with high connectivity.

When we analyzed the relationship between the total number of proteins with LD and proteome size, the best-fit model is polynomial ($R^2 = 0.902$), suggesting that disorder abundance increases more sharply in very large proteomes (> 7000 proteins) (Supplementary file 1 Figure S10). Small proteome bacteria consistently have very low fractions of proteins with LD regions. The most extreme example in our dataset, *Buchnera aphidicola*, has only 8 proteins with LD. This pattern is also observed in other low-GC, small-proteome species, confirming that genome reduction is accompanied by a significant loss of disordered proteins.

These findings are further supported by the analysis of conserved bacterial orthologues using AlphaFold models and pLDDT scores, as well as IUPred analysis which show an increase in terminal disorder, especially in C-terminus (Fig. 4, Supplementary figure S8). We restricted the pLDDT analysis to a group of conserved orthologs because AlphaFold confidence scores depend on multiple alignment sequence depth[22], and may give a false indication of disorder in proteins that lack enough close homologs. In addition, whole-proteome comparisons across bacteria with very different genome sizes may partly reflect differences in protein repertoire, including the loss of protein families enriched in disordered regions in reduced genomes. However, the ortholog-based analysis minimizes this effect by comparing homologous hub proteins shared across species, indicating that terminal disorder expansion also occurs within conserved proteins rather than changes in protein repertoire.

The obtained results show an increase in low pLDDT fraction at both protein termini with proteome size, while the protein core (inner 50% AA) remains unchanged. Furthermore, same orthologs have conserved core both in sequence and 3D structure while the length of C-tail increases in complex bacteria (Fig. 4B, Supplementary figures S8, S9). The presence of C-terminal low pLDDT fractions in these proteins aligns with the trends observed in analyses of hub protein length and the distribution of LD regions (Figs. 1 and 3D, Supplementary File 1 Table S3). Disordered terminal extensions have been experimentally characterized in several bacterial hubs. The C-terminal residues of FtsZ from *E. coli* interact with ZapA protein[53], and are conserved in many bacteria with crystal structures showing no interpretable density for the C-terminal regions[54–57]. C-terminus of SSB is intrinsically disordered and mediates interactions with numerous genome maintenance proteins[58–60]. The RecA C-terminus is also intrinsically disordered and plays a key regulatory role in modulating filament dynamics and partner selection[61]. These examples demonstrate that conserved bacterial hubs frequently rely on disordered tails to mediate transient or multi-partner interactions, despite rapid sequence divergence in these regions. Although such experimental evidence is not available for all orthologs considered here, these examples support the interpretation that flexible/disordered terminal extensions are a common and functionally relevant feature of conserved bacterial hubs.

We have further shown that proteins with LD in *E. coli* have significantly more interactors than fully structured proteins (Fig. 5B). If a similar trend applies to large proteome bacteria, such as *S. coelicolor*, this suggests that these bacteria possess more hubs-acting proteins, enabling more complex interaction networks and potentially resembling the disordered eukaryotic interactomes. However, experimental interactome data for *S. coelicolor* and other large-genome, high-GC species remain limited. Many disordered proteins and their interactions in these organisms have yet to be characterized, making it difficult to confirm whether their disorder correlates with increased interaction complexity.

Taken together, our results indicate that disorder presence in bacterial hubs is not a passive consequence of GC content but is associated with proteome expansion, particularly in free-living bacteria. A parallel can be drawn with eukaryotes, where complex interaction networks are regulated by intrinsically disordered proteins to mediate transient interactions and signalling processes[16]. We propose that large proteomes in free-living bacteria evolve with more extensive disordered regions, possibly as a mechanism to accommodate interaction networks in more complex and competitive environments.

Conclusion

This study explores the characteristics of bacterial hubs in terms of genomic GC content, proteome size and lifestyle. Across 40 phylogenetically diverse bacterial species, we show that hubs are longer and enriched in disordered regions than non-hubs and the total proteomes. Our results indicate that intrinsic disorder increases in parallel with proteome size, that involve more regulatory functions and extended interaction networks in free

living bacteria. Our analyses show that both GC content and proteome size are associated with the presence of disorder in bacterial hubs, with proteome size emerging as the stronger predictor of disorder accumulation. Large proteomes tend to accumulate more disorder, with SD being more abundant, especially in hubs. Contrary, SD is less frequent and LD rare in small proteomes of obligate pathogens. Even in such reduced genomes hubs retain SD stretches, highlighting their role in PPI networks. The presence of LD regions in ortholog hubs from large genome bacteria suggests that disorder expansion can occur alongside conservation of core sequence and structure, allowing flexibility with the preserved ability to perform essential functions. Thus, despite sequence conservation of their core function, bacterial hubs are dynamic and adaptive, especially in species with complex lifestyles and regulatory requirements. In conclusion we propose that intrinsic disorder in bacterial hubs is associated with proteome and environmental complexity and may contribute to functional flexibility in bacteria with more complex lifestyles. like mechanisms observed in eukaryotes. GC enrichment in large genomes may be related not only to genome stability, but also to the functional needs of organisms living in a complex and changing environment that may enhance their survival through the increased protein flexibility associated with higher disorder. Future experimental studies will elucidate how this translates into biological functions and to what extent it contributes to bacterial adaptability and evolution. Because our analyses are comparative and correlational, they do not by themselves establish causal relationships between GC content, proteome size, intrinsic disorder, and bacterial lifestyle; rather, they identify robust cross-species associations consistent with these hypotheses.

Methods

Datasets preparation

The following proteomes and interactomes of 40 phylogenetically different bacterial model species (Supplementary file 1 Figure S1, and Supplementary File 2) were analysed in this study. The species tree was generated with the PhyloT tool and modified with iTOL[62].

The reference proteomes were downloaded from the Uniprot database[63]. Three datasets were analysed for each species: Total proteome, hub and non-hub datasets. Interactome data were downloaded from 3 different databases: STRING[64], IntAct[65] and APID[66]. Protein–protein interactions were loaded from the STRING database using stringApp[67] for Cytoscape software (version 3.10.2)[68].

STRING and IntAct interactomes were filtered to only experimentally confirmed data to extract hub proteins. When using whole interactomes with more than 500 identified interactions, APID interactomes were filtered to level two binary interactomes with at least one binary method of direct interaction. Hub proteins were described as having at least 5 interactants in 2 out of 3 databases under the conditions described. For species having few or no interactome data in APID or IntAct databases, hub proteins were selected using STRING (high confidence value 0.8) with at least 20 interactors; for some species with large proteomes, a stricter cut off of at least 30 interactors was applied as indicated in Supplementary file 2., to make hub selection even more stringent in large-proteome species. In two small-proteome species, proteins with more than 10 interactors were used as hubs because too few 20 + interactors were identified. The frequency distribution of interaction-count categories for STRING-based species is provided in Supplementary File 1, and the corresponding raw counts are listed in Supplementary File 2. Intrinsically disordered proteins may be prone to apparent nonspecific interactions in high-throughput assays, only experimentally supported and, where possible, direct/binary interaction datasets were retained to minimize false positive hub assignment.

Non-hub proteins were selected using a stringent filtering strategy. For species with experimentally curated interaction data, non-hubs were defined as proteins without detected interactors in any of the analysed databases. For the other species, proteins lacking interactors at the stricter STRING confidence cut off (0.8) were re-checked using the relaxed confidence threshold of 0.2 to reduce false negatives. Proteins retained after this procedure were used as the non-hub dataset. For several of the smallest proteomes, proteins with one detected interactor were also included because otherwise the dataset would have been too limited. Hub and non-hub protein lists for all analysed organisms are provided in Supplementary File 2.

Data analysis

To obtain an overview of the most important biological processes and molecular functions of the all hub proteins used in this study, we performed a GO Slim analysis within the Quick GO web-based tool[69]. GO Slim analysis was chosen specifically to summarize broad functional categories rather than perform enrichment testing. A curated set of intermediate-level ancestor GO terms was selected to cover major cellular functions relevant to our dataset (e.g. protein synthesis, nucleotide binding, enzyme activity), ensuring broad but interpretable resolution and descriptive functional overview. GO Slim categorization was applied separately to Biological Process (BP) and Molecular Function (MF) domains to summarize the functional distribution of hub proteins (https://www.ebi.ac.uk/QuickGO/slimming). All GO terms used in this study are listed in Supplementary file 1, Figure S3. Gene product identifiers were submitted alongside the GO term lists. GO Slim analysis determines the frequency of proteins assigned to each GO term and calculates the percentage of each GO term within the dataset. The same GO term mapping was applied to proteins with long disorder in *E. coli* and *S. coelicolor*.

The prepared datasets were run with the IUPred3 tool using default parameters for disorder detection[21]. The data from the output files were extracted using custom Python script to obtain only strings of AAs with a score of 0.5 or higher and a length of at least 5 AA (Supplementary file 1). The resulting file contained a single line for each disorder stretch in protein, the number of the AA with which the string ends, the Uniprot ID and the AA string. Predicted disordered regions for the protein from each dataset and organisms are listed in Supplementary File 3, with protein ID, designation of dataset with disorder sequence, disorder length and position of disorder end in protein sequence, in sheets with corresponding species names. The following metrics were calculated for each

species: number of proteins with any type of disorder, percentage of disordered AA/protein and frequency of each length of disorder (short 5–15 AA and long > 30 AA; number of disorder events/total number of proteins).

PDB files of orthologous hub and non-hub proteins for each species in the dataset (corresponding gene names are listed in Supplementary File S2) were downloaded from the AlphaFold Protein Structure Database[70]. AlphaFold pLDDT scores were extracted using custom Python script (Supplementary file 1) and analysed for these proteins to assess the accumulation of disorder in different protein regions: 20% of AA at the N and C termini, the first and last 50 AA, and the 50% core region, using both thresholds of 50 and 70 for consecutive stretches longer than 5 AA (raw data are listed in Supplementary File 3). These thresholds were used because previous studies have shown that pLDDT scores below 50 are often indicative of intrinsically disordered regions, while pLDDT scores between 50 and 70 are typically considered low confidence or ambiguous in structure, indicating flexible or unstructured segments rather than reliably folded domains[25,71].

Heatmap visualisation of disorder within different datasets with respect to GC content was performed by Morpheus online tool[72].

The analysis of disordered regions of all proteins and hubs was performed using custom Python script (Supplementary file 1).

Statistical analyses were performed using GraphPad Prism 8.0 (GraphPad Software Inc.) and Python (3.13.0). Student's t-test or Wilcoxon test was used to assess statistical significance between two datasets, while Tukey's post-hoc test after one-way ANOVA was used for comparisons between multiple datasets. Simple linear regression was used to examine the relationship between proteome size or GC content and average protein length, percentage of disordered proteins, percentage of disordered AAs per protein and, frequency of short/long disorder per protein. Correlation coefficients (R^2 values) and p-values were calculated to determine the strength and statistical significance of these relationships.

Multiple regression analyses were performed in Python using the *statsmodels* package (0.14.4) for regression modelling and *SciPy* for statistical testing (1.15.2)[73,74]. To analyse the relationship between GC content, proteome size, and protein disorder, multiple regression analysis was performed separately for hubs and total proteomes. The contribution of both GC content (%) and proteome size was evaluated in predicting percentage of disordered proteins, percentage of disordered AAs, and frequency of short and long disordered regions per protein. Regression models were constructed using the ordinary least squares (OLS) method. Model fit was assessed using R^2, and statistical significance was determined for each predictor variable. An analysis of covariance (ANCOVA) was performed to test whether the regression slopes significantly differed between hubs and proteomes. Significant differences were determined using the following p-value thresholds (indicated in the figures): $p < 0.05$ (*), $p < 0.01$ (**), $p < 0.001$ (***) and $p < 0.0001$ (****).

Data availability

Data generated or analysed during this study are included in this published article and its supplementary information files (Supplementary Files 1, 2, and 3). Supplementary File 1 contains the Python scripts used for the analyses and data extraction. Supplementary File 2 contains the list of proteomes analysed in this study, the corresponding hub and non-hub protein sets, and the ortholog groups used for the pLDDT analysis. Supplementary File 3 contains the results of the pLDDT analysis across different protein regions and the IUPred-based disorder analyses.

Received: 22 December 2025; Accepted: 8 May 2026
Published online: 15 May 2026

References

1. Venkatesan, K. et al. An empirical framework for binary interactome mapping. *Nat. Methods* **6**, 83–90 (2009).
2. Lu, H. et al. Recent advances in the development of protein–protein interactions modulators: mechanisms and clinical trials. *Signal Transduct. Target. Therapy* **5**(1), 213 (2020).
3. Rajagopala, S. V. et al. The binary protein-protein interaction landscape of *Escherichia coli*. *Nat. Biotechnol.* **32**, 285–290 (2014).
4. Schreiber, G. & Keating, A. E. Protein binding specificity versus promiscuity. *Curr. Opin. Struct. Biol.* **21**, 50–61 (2011).
5. Skawinski, C. L. S. & Shah, P. S. I'm walking into spiderwebs: Making sense of protein–protein interaction data. *J. Proteome Res.* **23**, 2723–2732 (2024).
6. Cao, L. et al. Design of protein-binding proteins from the target structure alone. *Nature* **605**, 551–560 (2022).
7. Patil, A., Kinoshita, K. & Nakamura, H. Hub promiscuity in protein-protein interaction networks. *Int. J. Mol. Sci.* **11**, 1930–1943 (2010).
8. Nithya, C., Kiran, M. & Nagarajaram, H. A. Hubs and bottlenecks in protein-protein interaction networks. *Methods Mol. Biol.* **2719**, 227–248 (2024).
9. Wright, P. E. & Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015).
10. Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M. & Uversky, V. N. Flexible nets: The roles of intrinsic disorder in protein interaction networks. *FEBS J.* **272**, 5129–5148 (2005).
11. Patil, A. & Nakamura, H. Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks. *FEBS Lett.* **580**, 2041–2045 (2006).
12. Peng, Z. et al. Exceptionally abundant exceptions: Comprehensive characterization of intrinsic disorder in all domains of life. *Cell. Mol. Life Sci.* **72**, 137–151 (2014).
13. Pavlovic-Lazetic, G. M. et al. Bioinformatics analysis of disordered proteins in prokaryotes. *BMC Bioinformatics* **12**, 66 (2011).
14. Uversky, V. N. The multifaceted roles of intrinsic disorder in protein complexes. *FEBS Lett.* **589**, 2498–2506 (2015).
15. Uversky, V. N. Functional unfoldomics: Roles of intrinsic disorder in protein (multi)functionality. *Adv. Protein Chem. Struct. Biology*. **138**, 179–210 (2024).
16. Basile, W., Salvatore, M., Bassot, C. & Elofsson, A. Why do eukaryotic proteins contain more intrinsically disordered regions? *PLoS Comput. Biology* **15** (7), (2019).
17. Bertolazzi, P., Bock, M. E. & Guerra, C. On the functional and structural characterization of hubs in protein-protein interaction networks. *Biotechnol. Adv.* **31**, 274–286 (2013).

18. Grazziotin, A. L., Vidal, N. M. & Venancio, T. M. Uncovering major genomic features of essential genes in Bacteria and a methanogenic Archaea. *FEBS J.* **282**, 3395–3411 (2015).
19. Typas, A. & Sourjik, V. Bacterial protein networks: Properties and functions. *Nat. Rev. Microbiol.* **13**, 559–572 (2015).
20. Lamolle, G., Simón, D., Iriarte, A. & Musto, H. Main factors shaping amino acid usage across evolution. *J. Mol. Evol.* **91**, 382–390 (2023).
21. Erdos, G., Pajkos, M. & Dosztányi, Z. IUPred3: Prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res.* **49**, W297–W303 (2021).
22. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
23. Brotzakis, Z. F., Zhang, S., Murtada, M. H. & Vendruscolo, M. AlphaFold prediction of structural ensembles of disordered proteins. *Nature Communications* **16** (1) (2025).
24. Zhao, B., Ghadermarzi, S. & Kurgan, L. Comparative evaluation of AlphaFold2 and disorder predictors for prediction of intrinsic disorder, disorder content and fully disordered proteins. *Comput. Struct. Biotechnol. J.* **21**, 3248–3258 (2023).
25. Reid Alderson, T., Pritišanac, I., Kolaric, D., Moses, A. M. & Forman-Kay, J. D. Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2. *Proceedings Natl. Acad. Sci. United States America* **120** (44), e2304302120 (2023).
26. Nikolaeva, D. D., Gelfand, M. S. & Garushyants, S. K. Simplification of ribosomes in bacteria with tiny genomes. *Mol. Biol. Evol.* **38**, 58–66 (2021).
27. Mann, S. & Chen, Y. P. P. Bacterial genomic G + C composition-eliciting environmental adaptation. *Genomics* **95**, 7–15 (2010).
28. Musto, H. How many factors influence genomic GC content among prokaryotes?. *J. Mol. Evol.* **91**, 6–9 (2023).
29. Hu, E. Z., Lan, X. R., Liu, Z. L., Gao, J. & Niu, D. K. A positive correlation between GC content and growth temperature in prokaryotes. *BMC Genomics* **23** (1) (2022).
30. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **337**, 635–645 (2004).
31. Ghosh, A. et al. Bacterial IF2's N-terminal IDR drives cold-induced phase separation and promotes fitness during cold stress. *bioRxiv: preprint Serv. biology*. https://doi.org/10.1101/2025.02.02.631968 (2025). doi:10.1101/2025.02.02.631968.
32. Paradžik, T., Filić, Ž & Vujaklija, D. Variations in amino acid composition in bacterial single stranded DNA–binding proteins correlate with GC content. *Period. Biol.* **118**, 385–397 (2016).
33. Jamshad, M. et al. The C-terminal tail of the bacterial translocation ATPase SecA modulates its activity. *eLife* **8**, e48385 (2019).
34. Sundararajan, K. & Goley, E. D. The intrinsically disordered C-terminal linker of FtsZ regulates protofilament dynamics and superstructure in vitro. *J. Biol. Chem.* **292**, 20509–20527 (2017).
35. Moro, F., Taneva, S. G., Velázquez-Campoy, A. & Muga, A. GrpE N-terminal domain contributes to the interaction with DnaK and modulates the dynamics of the chaperone substrate binding domain. *J. Mol. Biol.* **374**, 1054–1064 (2007).
36. Paradžik, T., Filić, Ž. & Vujaklija, D. Variations in amino acid composition in bacterial single stranded DNA–binding proteins correlate with GC content. *Periodicum Biologorum* **118**, 385–397 (2017).
37. Pipalović, G. et al. Impact of C-terminal domains of paralogous single-stranded DNA binding proteins from *Streptomyces coelicolor* on their biophysical properties and biological functions. *Int. J. Biol. Macromol.* 131544. https://doi.org/10.1016/j.ijbiomac.2024.131544 (2024).
38. Cohan, M. C., Eddelbuettel, A. M. P., Levin, P. A. & Pappu, R. V. Dissecting the functional contributions of the intrinsically disordered C-terminal tail of *Bacillus subtilis* FtsZ. *J. Mol. Biol.* **432**, 3205–3221 (2020).
39. Wang, C., Uversky, V. N. & Kurgan, L. Disordered nucleiome: Abundance of intrinsic disorder in the DNA- and RNA-binding proteins in 1121 species from Eukaryota, Bacteria and Archaea. *Proteomics* **16**, 1486–1498 (2016).
40. Bret, H., Gao, J., Zea, D. J., Andreani, J. & Guerois, R. From interaction networks to interfaces, scanning intrinsically disordered regions using AlphaFold2. *Nat. Commun.* **15** (1) (2024).
41. Caufield, J. H., Wimble, C., Shary, S., Wuchty, S. & Uetz, P. Bacterial protein meta-interactomes predict cross-species interactions and protein function. *BMC Bioinformatics* **18** (1) (2017).
42. Paradžik, T., Podgorski, I. I., Zeljko, T. V. & Paradžik, M. Ancient origins of cytoskeletal crosstalk: Spectraplakin-like proteins precede the emergence of cortical microtubule stabilization complexes as crosslinkers. *Int. J. Mol. Sci.* **23** (10) (2022).
43. Patil, A., Kinoshita, K. & Nakamura, H. Domain distribution and intrinsic disorder in hubs in the human protein-protein interaction network. *Protein Sci.* **19**, 1461–1468 (2010).
44. Ekman, D., Light, S., Björklund, Å. K. & Elofsson, A. What properties characterize the hub proteins of the protein-protein interaction network of *Saccharomyces cerevisiae*?. *Genome Biol.* **7** (6) (2006).
45. Šarić, E. et al. Phosphoproteome Dynamics of Streptomyces rimosus during Submerged Growth and Antibiotic Production. *mSystems,* **7**, (2022).
46. Lewin, G. R. et al. Evolution and ecology of Actinobacteria and their bioenergy applications. *Annu. Rev. Microbiol.* **70**, 235–254 (2016).
47. Paradzik, T. et al. Structure-function relationships of two paralogous single-stranded DNA-binding proteins from Streptomyces coelicolor: Implication of SsbB in chromosome segregation during sporulation. *Nucleic Acids Res.* **41**, 3659–3672 (2013).
48. Trivedi, R. & Nagarajaram, H. A. Intrinsically disordered proteins: An overview. *Int. J. Mol. Sci.* **23** (22) (2022).
49. Nordyke, C. T., Ahmed, Y. M., Puterbaugh, R. Z., Bowman, G. R. & Varga, K. Intrinsically Disordered Bacterial Polar Organizing Protein Z, PopZ, Interacts with Protein Binding Partners Through an N-terminal Molecular Recognition Feature. *J. Mol. Biol.* **432**, 6092–6107 (2020).
50. Gardner, K. A. J. A., Moore, D. A. & Erickson, H. P. The C-terminal linker of *Escherichia coli* FtsZ functions as an intrinsically disordered peptide. *Mol. Microbiol.* **89**, 264–275 (2013).
51. Bianco, P. R. The mechanism of action of the SSB interactome reveals it is the first OB -fold family of genome guardians in prokaryotes. *Protein Sci.* **30**, 1757–1775 (2021).
52. Brown, C. J., Johnson, A. K., Dunker, A. K. & Daughdrill, G. W. Evolution and disorder. *Curr. Opin. Struct. Biol.* **21**, 441–446 (2011).
53. Mosyak, L. et al. The bacterial cell-division protein ZipA and its interaction with an FtsZ fragment revealed by X-ray crystallography. *EMBO J.* **19**, 3179–3191 (2000).
54. Oliva, M. A., Trambaiolo, D. & Löwe, J. Structural insights into the conformational variability of FtsZ. *J. Mol. Biol.* **373**, 1229–1242 (2007).
55. Leung, A. K. W. et al. Structure of Mycobacterium tuberculosis FtsZ reveals unexpected, G protein-like conformational switches. *J. Mol. Biol.* **342**, 953–970 (2004).
56. Matsui, T. et al. Structural reorganization of the bacterial cell-division protein FtsZ from *Staphylococcus aureus*. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 1175–1188 (2012).
57. Sogues, A. et al. Essential dynamic interdependence of FtsZ and SepF for Z-ring and septum formation in *Corynebacterium glutamicum*. *Nat. Commun.* **11**, 1641 (2020).
58. Bianco, P. R. The mechanism of action of the < scp>SSB</scp> interactome reveals it is the first < scp>OB</scp> -Fold family of genome guardians in prokaryotes. *Protein Sci.* **30**, 1757–1775 (2021).
59. Ding, W. et al. The mechanism of < scp>Single strand binding protein–RecG binding: Implications for < scp>SSB interactome function. *Protein Sci.* **29**, 1211–1227 (2020).
60. Shereda, R. D., Kozlov, A. G., Lohman, T. M., Cox, M. M. & Keck, J. L. SSB as an organizer/mobilizer of genome maintenance complexes. *Crit. Rev. Biochem. Mol. Biol.* **43**, 289–318 (2008).

61. Fan, H. F. & Su, S. The regulation mechanism of the C-terminus of RecA proteins during DNA strand-exchange process. *Biophys. J.* **120**, 3166–3179 (2021).
62. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: Recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* **52**, W78–W82 (2024).
63. Bateman, A. et al. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **51**, D523–D531 (2023).
64. Szklarczyk, D. et al. The STRING database in 2023: Protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).
65. del Toro, N. et al. The IntAct database: efficient access to fine-grained molecular interaction data. *Nucleic Acids Res.* **50**, D648–D653 (2022).
66. Alonso-López, D. et al. APID database: Redefining protein-protein interaction experimental evidences and binary interactomes. *Database* (2019).
67. Doncheva, N. T., Morris, J. H., Gorodkin, J. & Jensen, L. J. Cytoscape StringApp: Network analysis and visualization of proteomics data. *J. Proteome Res.* **18**, 623–632 (2019).
68. Shannon, P. et al. Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
69. Binns, D. et al. QuickGO: A web-based tool for Gene Ontology searching. *Bioinformatics* **25**, 3045–3046 (2009).
70. Varadi, M. et al. AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* **52**, D368–D375 (2024).
71. Piovesan, D., Monzon, A. M. & Tosatto, S. C. E. Intrinsic protein disorder and conditional folding in AlphaFoldDB. *Protein Sci.* **31**(11) (2022).
72. Broad Institute. Morpheus. software.broadinstitute.org/Morpheus.
73. Seabold, S., Perktold, J. & Statsmodels Econometric and Statistical Modeling with Python. *Proceedings of the 9th Python in Science Conference* 92–96 (2010). https://doi.org/10.25080/majora-92bf1922-011
74. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods*. **17**, 261–272 (2020).

Author contributions

M.P., S.P. and T.P performed data analyses, T.P. conceived and designed this study and drafted manuscript, with input and revisions from M.P.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-52986-7.

Correspondence and requests for materials should be addressed to T.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.