

Systematic analysis of structural disorder in the minimal proteome of *Mycoplasma pneumoniae*

Uberto Pozzoli,¹ Diego Forni,¹ Federica Arrigoni,² Alessandra Mozzi,¹ Rachele Cagliani,¹ Luca De Gioia,² Manuela Sironi³

AUTHOR AFFILIATIONS See affiliation list on p. 15.

ABSTRACT *Mycoplasma pneumoniae* (Mpn, class *Mollicutes*) is both an important human pathogen and a model organism. We performed a proteome-wide investigation of intrinsically disordered regions (IDRs) in Mpn. Compared to other bacteria, a considerable fraction of the Mpn proteome (17%) is embedded in IDRs, which are abundant in membrane, non-essential proteins, as well as in proteins that mediate cytoadherence and virulence. Notably, proteins that form the attachment organelle, a specialized structure, are particularly rich in IDRs. Likewise, analysis of protein architectures indicated that some Mollicute-specific domains are preferentially associated with IDRs. Perusal of proteome-wide data also revealed that, as in eukaryotes, structural disorder associates with higher protein degradation rates and that Mpn IDRs are preferential targets of phosphorylation. When we investigated the ensemble features for Mpn IDRs, we used two predictors and benchmarked the results using coarse-grained simulations. We found that ensemble properties are mediated by similar sequence features as in eukaryotes, so that compact IDRs tend to have high residue stickiness, high hydrophathy decoration, and few charged residues. We also found that IDRs in attachment organelle proteins are particularly extended and display high conformational entropy. We suggest that these features are exploited for motility through the generation of an entropic force. In summary, our results suggest that structural disorder contributes to very specialized functions in Mpn. Our data also highlight the functional relevance of IDRs, as the minimal proteome of this model organism displays a considerable level of structural disorder.

IMPORTANCE We performed a proteome-wide investigation of intrinsically disordered regions (IDRs) in *Mycoplasma pneumoniae* (Mpn, class *Mollicutes*). A considerable fraction of the Mpn proteome (17%) is embedded in IDRs, which tend to be associated with Mollicute-specific domains and are abundant in membrane, non-essential proteins, as well as in proteins that mediate cytoadherence and virulence. As in eukaryotes, structural disorder associates with higher protein degradation rates, and Mpn IDRs are preferential targets of phosphorylation. The ensemble properties of Mpn IDRs are mediated by similar sequence features as in eukaryotes, and IDRs in attachment organelle proteins display high conformational entropy. We suggest that this feature is exploited for motility through the generation of an entropic force. In summary, we show that structural disorder contributes to very specialized functions in Mpn. Our data highlight the functional relevance of IDRs, as the minimal proteome of this model organism displays a considerable level of structural disorder.

KEYWORDS *Mycoplasma pneumoniae*, intrinsically disordered region, AlphaFold, structural disorder, ensemble features

Intrinsically disordered proteins (IDPs) and protein regions (IDRs) do not adopt a stable three-dimensional structure (3D), but rather exist in conformational ensembles—that is, collections of energetically accessible, rapidly interconverting structures (1). Molecular

Editor John M. Atack, Griffith University-Gold Coast Campus, Gold Coast, Australia

Address correspondence to Uberto Pozzoli, uberto.pozzoli@lanostrafamiglia.it.

The authors declare no conflict of interest.

See the funding table on p. 16.

Received 31 March 2025

Accepted 26 May 2025

Published 18 June 2025

Copyright © 2025 Pozzoli et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

interactions within an IDR determine the overall properties of the conformational ensemble, such as its compaction or extension. In turn, conformational features are often linked to functional properties (2, 3). Thus, IDRs are generally described by “sequence-ensemble-function” relationships. Structural disorder is common in the proteomes of living organisms and viruses. In eukaryotes, IDRs cover up to 40% of the proteome and have been implicated in a variety of cellular processes (1). In viruses, large variations in the IDR fraction are observed among, but also within, viral families (4–7). In viral proteomes, IDR-containing proteins often contribute to host-virus interactions (4, 5, 8–10). In bacteria and archaea, IDRs are less abundant, but they were reported to be enriched in proteins that contribute specific functions such as sporulation, septum formation, cell wall biosynthesis, and flagellum-dependent motility (1, 5, 11–13). Analysis of individual IDRs in bacterial proteins also revealed that they may be important for mechanisms related to adaptation and pathogenesis such as toxin-antitoxin systems, protection against antimicrobial peptides, and biofilm formation (14–18). Despite these insights, systematic analyses of IDRs in bacterial systems are missing, and the functions of these regions remain largely unexplored. We thus aimed to fill this knowledge gap by focusing on the proteome of *Mycoplasma pneumoniae* (Mpn, order Mycoplasmales, class Mollicutes), a relevant human pathogen and a model organism (19). Among self-replicating prokaryotes, Mpn has one of the smallest known genomes, which encodes fewer than 700 proteins (19, 20). The small genome of Mpn is a consequence of its adaptation to an obligate parasitic lifestyle, and the bacterium lacks a cell wall (20). Infection with Mpn can cause chronic human respiratory tract disease and pneumonia (21). Across age groups, Mpn is estimated to be responsible for up to 40% of cases of community-acquired pneumonia (21). In this study, we integrated the predictive power afforded by AlphaFold2, coarse-grained simulations, and deep learning tools to analyze conformational ensembles, as well as large-scale *in vitro* data to provide an overview of the functional features of IDRs in the Mpn proteome.

RESULTS

IDR prediction in Mpn proteins

We aimed to predict the overall IDR content of the Mpn proteome. To benchmark the prediction method, we retrieved information on all ($n = 18$) Mpn proteins with an experimentally solved structure from the PDB database (22) (Table S1). From these structures, we annotated gap regions—that is, regions that were not solved in the crystal and are thus likely to be disordered. Gaps longer than 29 amino acids were considered IDRs (3). We identified five IDRs in three proteins (15 proteins were fully folded). We next used two approaches for IDR prediction. Specifically, we applied Metapredict V2, a deep-learning-based method that combines different predictors to generate consensus disorder scores (23, 24), and we used an IDR prediction approach based on per-residue predicted local difference test (pLDDT) scores from AlphaFold2 models (3) (see Materials and Methods). In both cases, disordered regions shorter than 30 residues were not considered IDRs (3). We next calculated the sensitivity, specificity, and accuracy for the two methods based on the fractions of residues that were correctly or incorrectly (according to the structural data) classified as IDR/folded. Results indicated that the AlphaFold2-based method (sensitivity = 0.996, specificity = 0.941, accuracy = 0.943) performed better than Metapredict V2 (sensitivity = 0.577, specificity = 0.871, accuracy = 0.859).

IDRs are enriched in membrane proteins that mediate Mpn cytoadherence

Given the results above, we used the AlphaFold2-based approach to investigate the IDR content in the reference proteome of Mpn strain M129 (686 proteins). We estimated that ~17.4% of residues in the Mpn proteome are embedded within IDRs and that 260 proteins (37.9%) have at least one IDR. These proportions are higher than those reported for most bacteria (25–27). We first asked whether IDRs are more common in

bacterial essential proteins or in proteins that are dispensable for optimal growth. We thus retrieved data from a transposon insertion mutagenesis study that categorized Mpn proteins as essential, non-essential, or fitness (which have an impact on growth, but do not affect cell viability) (28). We found that 23.7% of essential proteins have at least one IDR, whereas this percentage amounts to 58.9% and 33.3% in non-essential proteins and proteins that affect fitness (Table S2). These differences are statistically significant (Fisher’s Exact Test, P -value $< 2 \times 10^{-16}$) (Fig. 1A). In fact, when we compared the fraction of protein sequence in IDRs, this was significantly higher in non-essential proteins compared to the essential ones (Brunner-Munzel test, P -value $< 2 \times 10^{-16}$) and to the ones that have an impact on growth (Brunner-Munzel test, P -value = 4.56×10^{-5}) (Fig. 1B).

We next investigated whether IDRs are particularly abundant depending on Mpn protein function or cellular localization. We evaluated the significance of whether

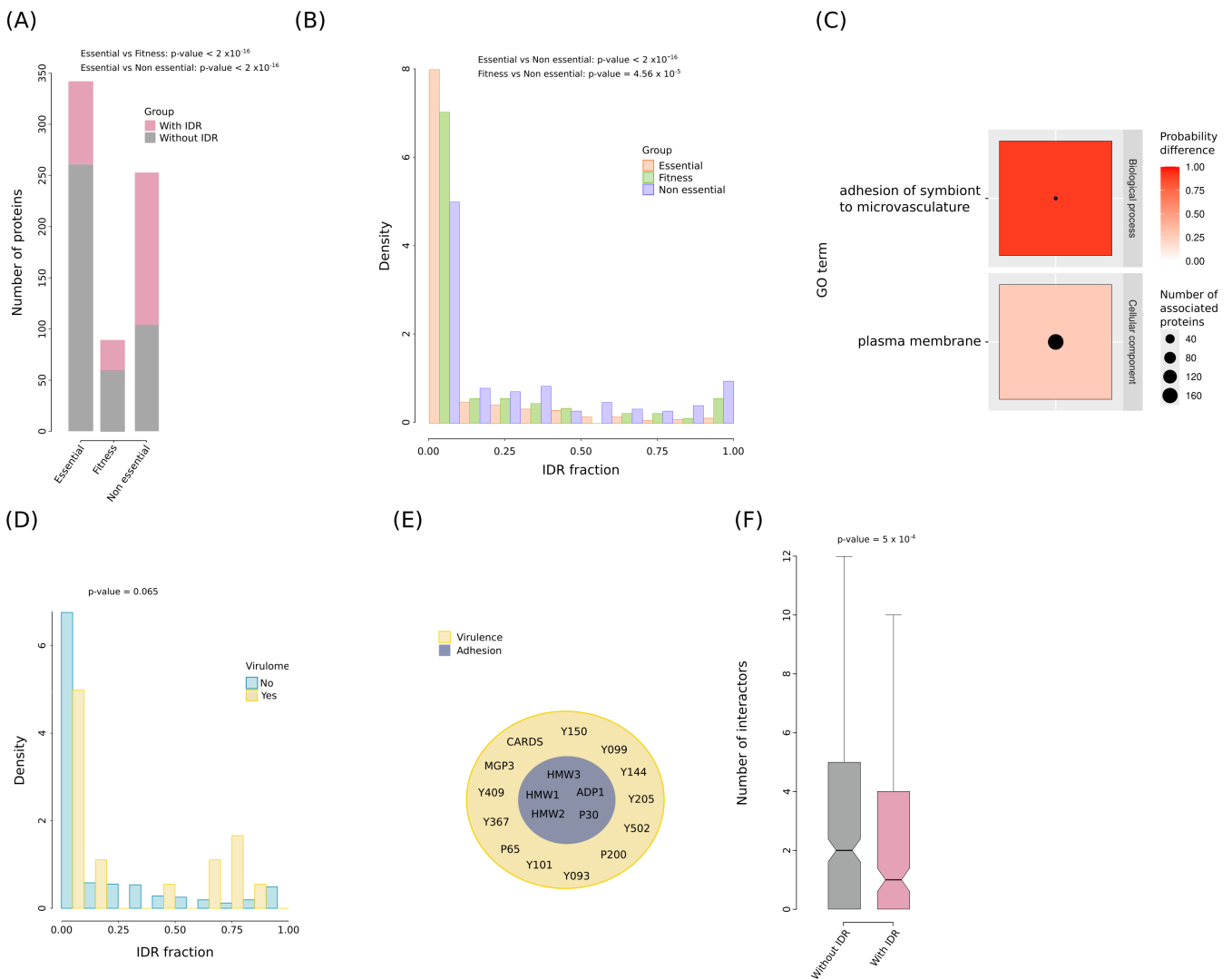


FIG 1 Function and interactions of IDRs in the Mpn proteome. (A) Comparison of the fraction of proteins with IDR in different essentiality categories. Fitness refers to proteins that have some impact on growth, but do not affect viability. P -values derive from Fisher’s exact tests. (B) Density histogram of IDR fraction (per protein) in the same classes. P -values are from Brunner-Munzel tests. (C) GO analysis of IDR fraction. Only significant terms are shown (P -value < 0.01 after FDR correction). The red scale represents the probability difference from Brunner-Munzel tests. The number of proteins associated with any term is represented by circle diameters. (D) Density histogram of IDR fraction in proteins that influence virulence and in those that do not (P -value from Brunner-Munzel test). (E) Venn diagram of proteins involved in virulence and of those associated with the GO term “adhesion of symbiont to microvasculature.” (F) Number of protein-protein interactions (PPIs) for Mpn proteins with or without IDRs (P -value from Brunner-Munzel test).

proteins associated with a given gene ontology (GO) term have a higher IDR fraction than proteins not associated with that term. Results indicated that proteins involved in adhesion to microvasculature and those located at the plasma membrane have a significantly higher IDR fraction than proteins that are not associated with these terms (Fig. 1C) (Table S2). Inspection of proteins associated with the GO term “adhesion of symbiont to microvasculature” indicated that all of them are components of the Mpn attachment organelle, a specialized structure that mediates adherence and gliding motility (29).

We next sought to explore IDR content in relation to virulence. We thus retrieved a list of 18 Mpn virulence factors using VirulenceFinder through the MicroScope website (30) (Table S2). Results indicated that virulence factors have a higher IDR fraction than proteins that do not contribute to virulence, although the statistical significance threshold was not reached, possibly because of the small sample size (Brunner-Munzel test, P -value = 0.065) (Fig. 1D). Because adherence is linked to Mpn pathogenic features, we found a considerable overlap among proteins identified in the GO analysis under the term “adhesion of symbiont to microvasculature” and those involved in virulence (Fig. 1E).

Finally, because IDRs are common in eukaryotic proteins that function as signaling hubs, we asked whether this was also the case in Mpn. We thus retrieved information about Mpn protein complexes from a previous work (31), and we counted the number of interactions that each protein establishes with other proteins (Table S2). We observed that proteins with IDRs have significantly fewer protein-protein interactions (PPIs) than proteins without IDRs (Brunner-Munzel test, P -value = 5×10^{-4}) (Fig. 1F). Overall, these results indicate that, in Mpn, IDRs may mediate attachment to host cells and contribute to virulence. Conversely, they have a limited role in controlling housekeeping cellular functions and are unlikely to participate in protein complexes.

Domains specific to Mollicutes are enriched in IDR-containing proteins

We next wished to investigate the architecture of IDR-containing proteins in terms of domain representation. We thus used SMART (Simple Modular Architecture Research Tool) to identify and annotate protein domains in the Mpn proteome. We identified a total of 563 different domains, but only 125 of these occurred in more than one protein. We thus limited the analysis to domains that occurred at least twice, and we evaluated the significance of whether proteins with a given domain have a higher IDR fraction than proteins without the same domain. We found statistically significant associations with IDR fraction for 10 domains (Fig. 2A). Of these, four (AAA_19, UvrD-helicase, UvrD_C_2, and UvrD_C) appear in very few proteins that have a role in DNA binding and repair. Two additional domains (lipoprotein X and 10) are characteristics of prokaryotic lipoproteins (32) and often occur in tandem in the same protein (Fig. 2B) (Table S3). In Mpn, these proteins were classified as family 2 lipoproteins (33). According to InterPro annotations, the remaining four domains are only found in Mollicutes. One of them, the EAGR box domain, has been implicated in gliding motility (34). In fact, it appears in three proteins of the attachment organelle (topJ, p200, and high molecular weight protein 1 [HMW1]) (Fig. 2B). The three other domains are classified as DUF (domains of unknown function). However, DUF31 is a putative serine protease, and in fact, it often overlaps with a peptidase S7 domain (Fig. 2B). Interestingly, the DUF31 domain was detected in an IgG protease encoded by *Mycoplasma mycoides* (but not Mpn), and analysis of Mollicute genomes showed that it is more common in species that are pathogenic to humans or other animals compared to plant pathogens (35). DUF3713 and DUF5396 have no known function. Whereas the former is relatively short, DUF5396 domains cover a substantial portion of two Mpn proteins (P75324 and P75327, both annotated in UniProt as uncharacterized lipoproteins), also overlapping with IDRs (Fig. 2B). To gain insight into the possible function of DUF5396, we used methods that search protein structure databases for structural homologs (36, 37). The top-scoring hits of both Foldseek and DALI were bacterial proteins that function as oligopeptide transporters

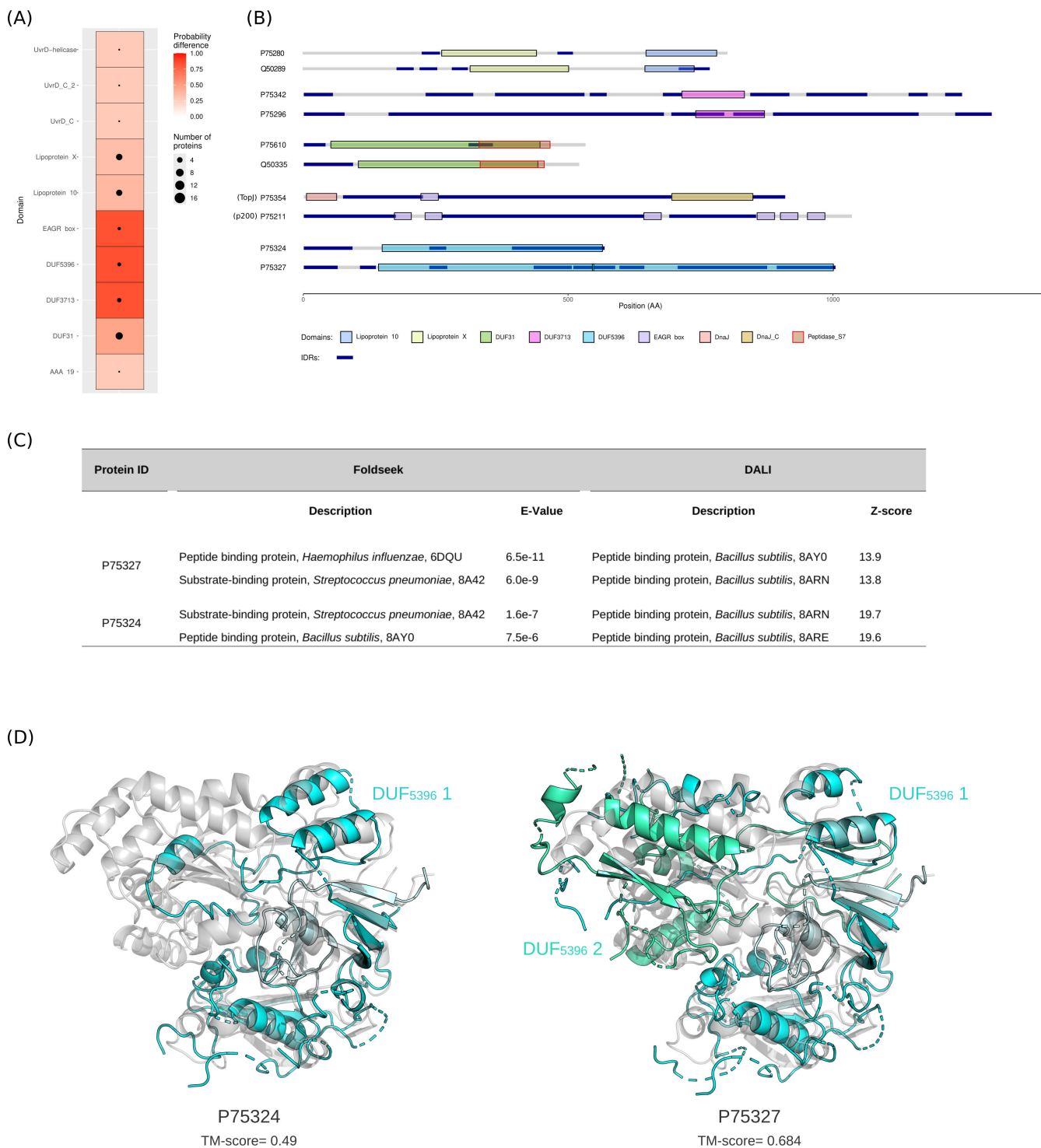


FIG 2 Analysis of protein domains associated with IDR fraction. (A) Protein domains associated with a high IDR fraction. Only significant terms are shown (P -value < 0.01 after FDR correction). The red scale represents the probability difference from Brunner-Munzel tests. (B) Domain structures of representative proteins carrying domains identified in the enrichment analysis. (C) Results of the structural homology searches. The AlphaFold-predicted structures of P75324 (AF-P75324-F1-v4) and P75327 (AF-P75327-F1-v4) were used to search the protein structure database using Foldseek and DALI. (D) Structural alignment (TM-align) of AF-P75324-F1-v4 and AF-P75327-F1-v4 with the solved structure of the OppA protein from *Bacillus subtilis* (PDB ID: 8arn). TM scores are also reported.

(Fig. 2C). Pairwise structural alignments using TM-align of the DUF5396 domains in P75324 and P75327 with the OppA protein of *Bacillus subtilis* showed very high scores and good structural similarity (Fig. 2D). This suggests that DUF5396-containing proteins function in oligopeptide binding.

IDR content correlates with protein degradation rates

In eukaryotes and *Escherichia coli*, disordered regions affect protein degradation rates (38–40). To assess whether this was also the case in Mpn, we exploited data from two recent works that measured protein abundance, half-life, degradation rate, and translational efficiency for ~470 proteins that were also included in our data set (41, 42) (Table S2). When proteins with IDRs were compared to those without IDRs, no differences were observed in protein abundance or translational efficiency (not shown). However, proteins with IDRs were found to have significantly shorter half-lives than proteins with no IDRs (Brunner-Munzel test, P -value = 7.6×10^{-4}) (Fig. 3A). Because protein half-life and degradation rates are inversely related (Pearson correlation coefficient = 0.49, P -value < 2×10^{-16}), IDR presence was also associated with significantly faster degradation (Brunner-Munzel test, P -value = 9.0×10^{-4}) (Fig. 3A). To determine whether these effects were secondary to the functional or biochemical characteristics of IDR-containing proteins, we restricted the analysis to proteins with IDRs (Fig. 3B). A significant negative correlation between half-life and IDR fraction was observed, while the correlation with degradation rates was significant and positive (Fig. 3B). It should, however, be noted that the correlations are weak and should be treated with caution. In any case, these

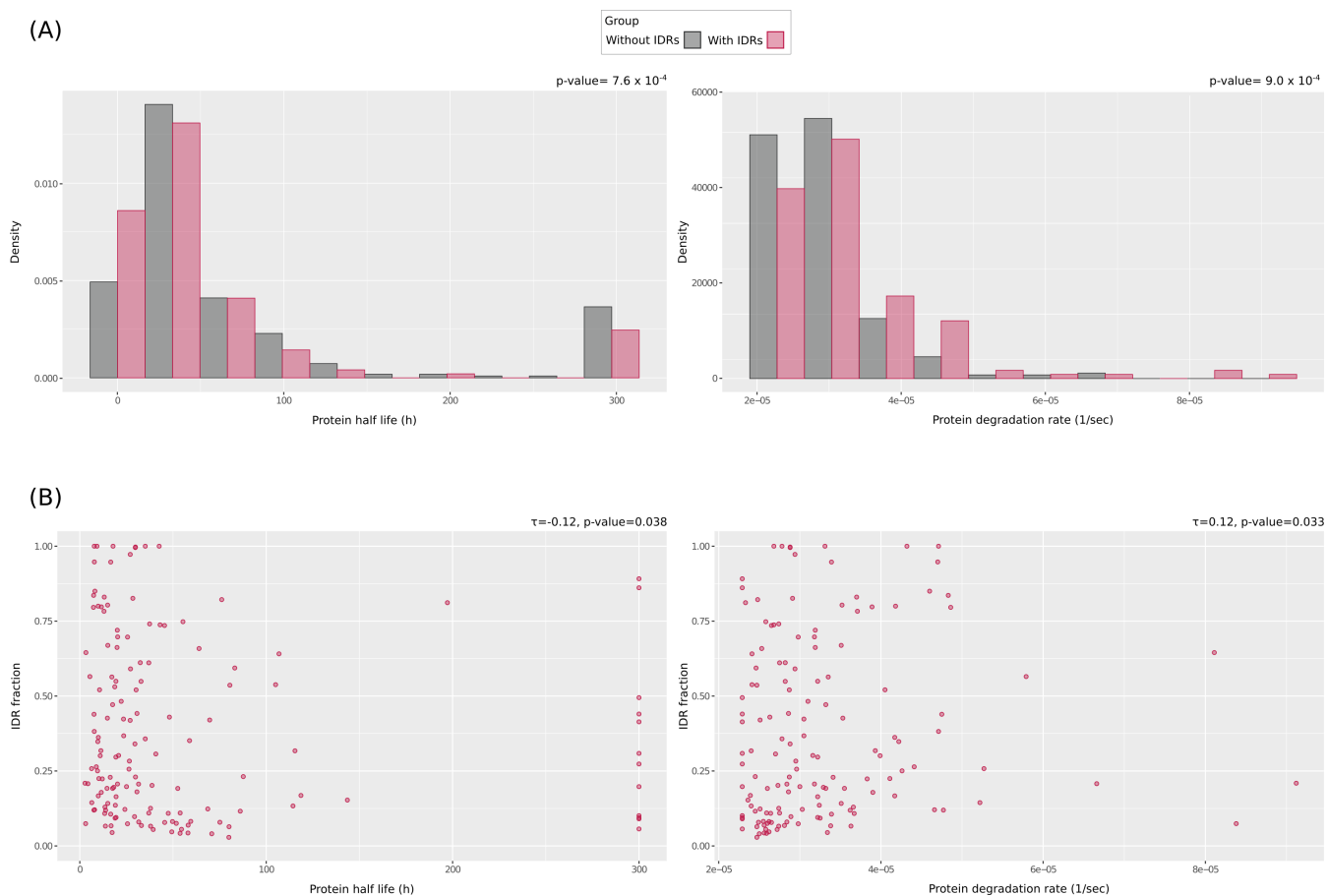


FIG 3 Analysis of protein half-life and degradation rates in relation to structural disorder. (A) Density histograms of protein half-life (left) and degradation rates (right) for Mpn proteins that have or do not have IDRs. P -values are from Brunner-Munzel tests. (B) Kendall's correlation between IDR fraction and protein half-life (left) or degradation rate (right). Only IDR-containing proteins were analyzed.

results suggest that structural disorder affects protein stability in Mpn by increasing degradation rates.

IDRs are preferential sites of phosphorylation, but not of acetylation or proteolytic cleavage

In the human proteome, IDRs were shown to be common targets of post-translational modifications (PTMs) (43). It is unknown whether this is also the case for IDRs in prokaryotes. To explore this issue, we exploited data from a study that profiled phosphorylation and lysine acetylation in 460 Mpn proteins, 459 of which were present in our data set (44). We first asked whether, among these proteins, those with IDRs had a different number of PTMs than those with no IDRs. We found that IDR-containing proteins carry more phosphorylation sites (Exact binomial test, P -value = 0.020) and significantly fewer acetylated lysines (Exact binomial test, P -value = 1.01×10^{-6}) (Fig. 4A) (Table S2). We next moved to the site level and investigated the localization of PTMs with respect to IDRs and structured domains. Results indicated that, out of 79 phosphorylation sites, 22 are located in IDRs, a number significantly higher than expected given the proportion of sequence embedded in IDRs (Exact binomial test, P -value = 0.004) (Fig. 4B). The opposite was true in the case of acetylated lysines, with fewer than expected being located in IDRs (54 in IDRs out of 634, Exact binomial test, P -value = 3.6×10^{-7}) (Fig. 4B).

Mpn was shown to use a mechanism of controlled post-translational cleavage as a strategy to diversify its proteome (45). Recently, about half of Mpn proteins were found to undergo proteolytic processing, and the cleavage sites were identified (46). We thus hypothesized that IDR-containing proteins are preferentially cleaved, but we found no difference in the number of cleavage sites between IDR-containing and IDR-lacking proteins (average number of cleavage sites = 2.7 and 2.6, Exact binomial test, P -value = 0.58) (Fig. 4A) (Table S2). Surprisingly, at the site level, we found that proteolytic cleavage is less likely to occur in IDRs than in structured domains (281 in IDRs out of 1835, Exact binomial test, P -value = 0.015) (Fig. 4B).

In the human proteome, PTMs were shown to be particularly abundant in short IDRs (fewer than 20 AA in length) (43). We thus hypothesized that the same might happen in Mpn and that our cutoff for the definition of IDRs at 30 AA affected the conclusion that acetylation and proteolysis preferentially occur in structured domains. We thus examined the pLDDT scores of all sites involved in PTMs that are not located in IDRs. We found that the overwhelming majority of such sites have pLDDT scores higher than 80, indicating that they are not located in short IDRs or flexible loops (Fig. 4C).

Conformational properties of Mpn IDRs

We next aimed to investigate the biophysical properties of IDRs in the Mpn proteome. Although they do not adopt fixed structures, IDRs can be described through their ensemble features. These include the Flory scaling exponent (ν), a length-independent measure of chain compactness, as well as other parameters such as the conformational entropy per residue (S_{conf}/N) (1). Recently, two predictors of IDR conformational properties were developed. One of them is based on a support vector regression (SVR) model and was trained on coarse-grained (CG) simulations of all the IDRs in the human proteome (3). The other, ALBATROSS, is a deep-learning model trained on GC simulations of naturally occurring IDRs mostly from eukaryotic proteomes, but also from prokaryotic ones, plus a library of synthetic sequences (47). Because none of the predictors was specifically trained on bacterial IDRs, we used both to estimate ν for all IDRs in the Mpn proteome (Table S4). Comparison of the estimates revealed a generally good correlation, although some IDRs deviated sensibly (Fig. 5A). We thus selected 44 IDRs that were at the extremes (5th and 95th percentiles) in the distribution of the difference between predictors to run CG simulations using the CALVADOS 2 force field (3, 48). Results showed that the estimates of ν from the SVR predictor were more similar to the ones from the CG simulations compared to those obtained with ALBATROSS (Fig. 5A; Table S4). A strong correlation was also detected for S_{conf}/N between the SVR predictor and the CG

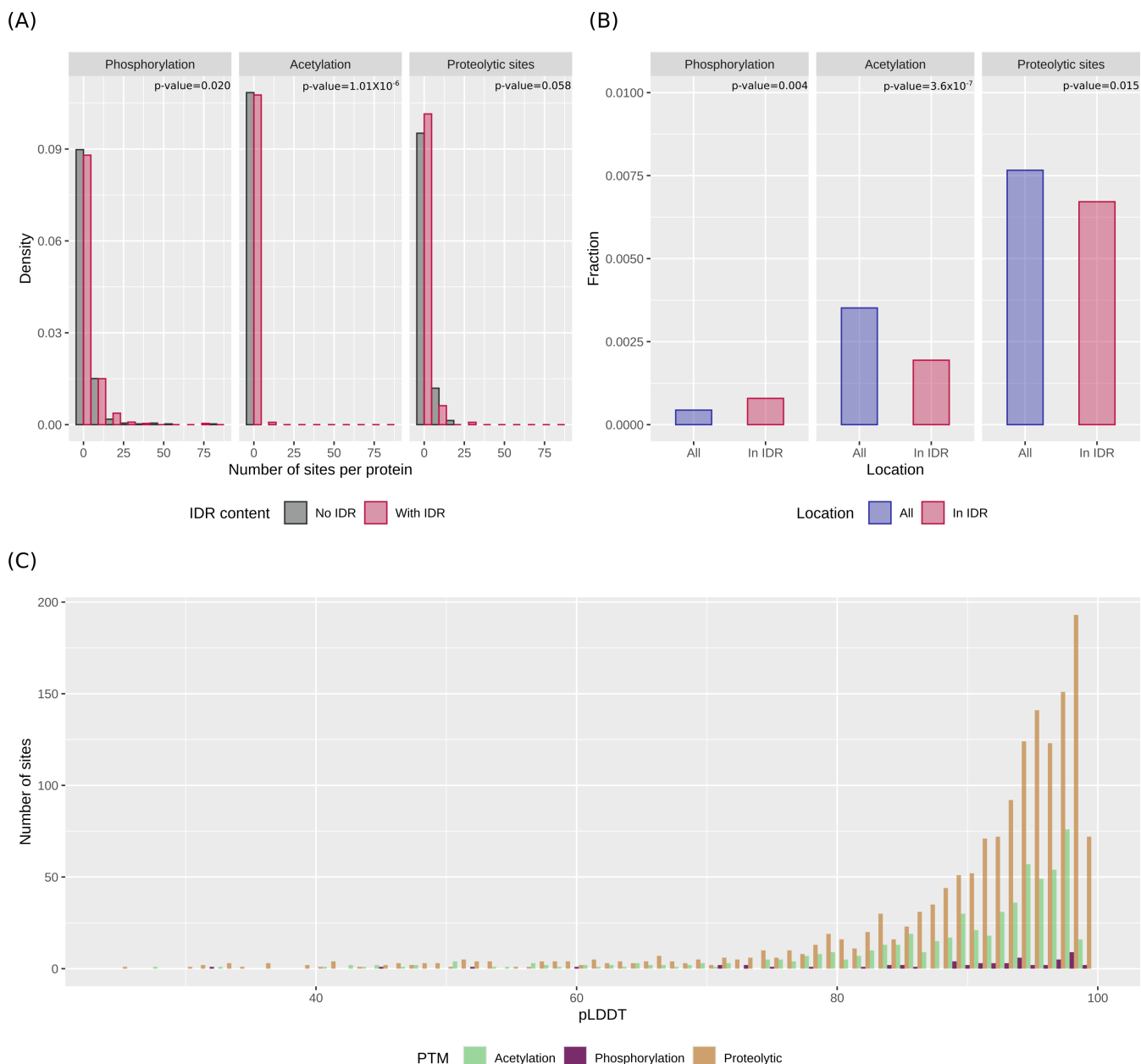


FIG 4 Analysis of PTMs in relation to structural features. (A) Density histograms of the number of phosphorylation, acetylation, and proteolytic sites in Mpn proteins that have or do not have IDRs. *P*-values are from Exact binomial tests. (B) Fraction of sites in IDRs compared to all PTMs. *P*-values are from Exact binomial tests. (C) pLDDT scores for all phosphorylation, acetylation, and proteolytic sites that are not located within IDRs.

simulations (this parameter is not implemented in ALBATROSS) (Fig. 5A). It is worth mentioning that the SVR predictor was trained on simulations based on the CALVADOS 2 force field, possibly explaining its better fit with our CG simulation results. However, for IDRs in the human proteome, comparison of CALVADOS2 simulations with ALBATROSS predictions showed a very high correlation (R^2 coefficient = 0.98) (47). Thus, these data suggest that the investigation of IDR conformational properties benefits from the comparison of different methods, especially in organisms whose IDRs are poorly studied and not included in predictor training. Also, it is worth noting that even the agreement between the SVR predictor and the CG simulations is good, but far from perfect, further supporting the need for more extended training data.

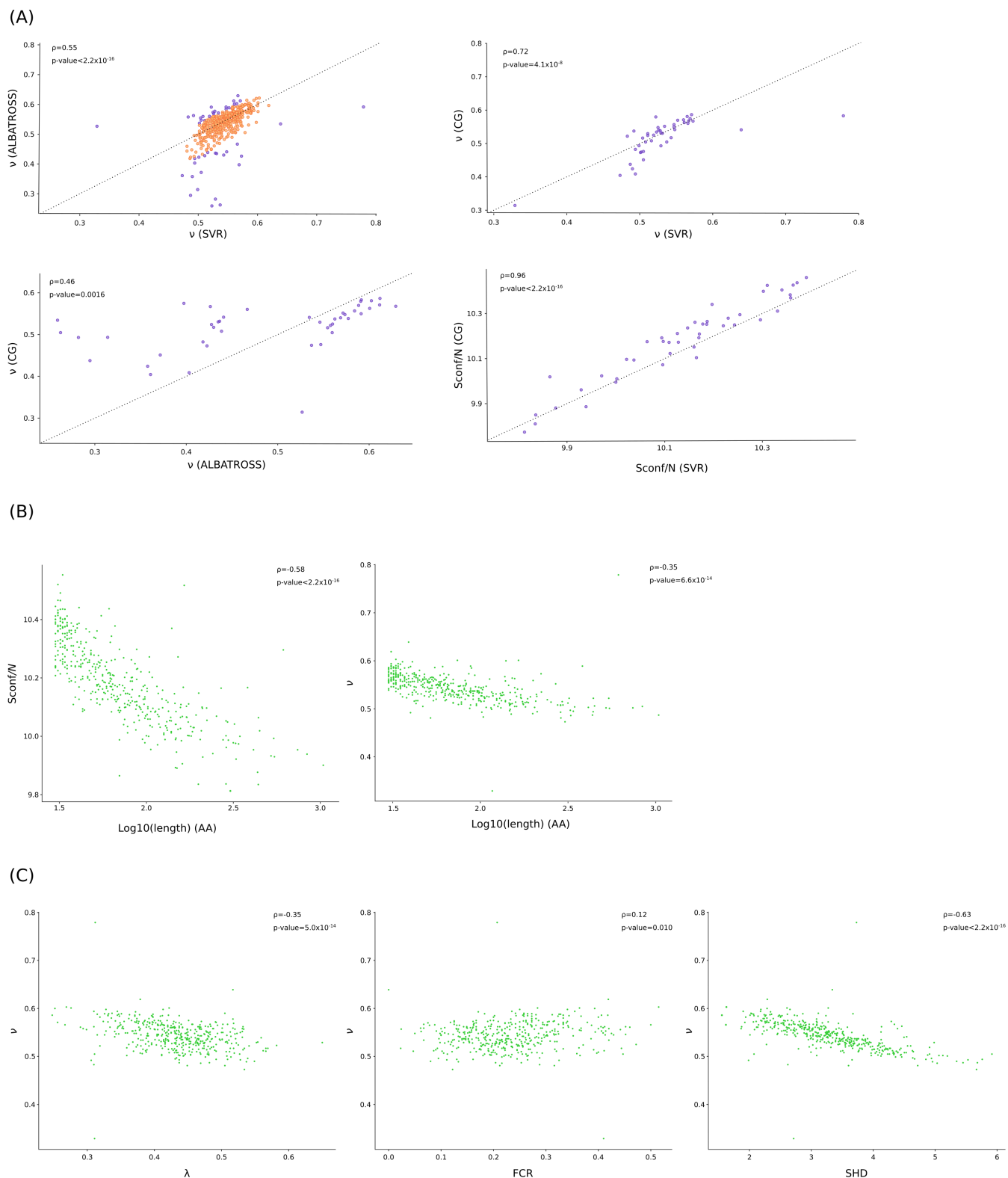


FIG 5 Conformational properties of Mpn IDRs. (A) Pearson’s correlation of estimates of the Flory scaling exponent (v) obtained with ALBATROSS and with the SVR predictors. 44 IDRs that are at the extremes (5th and 95th percentiles) in the distribution of the difference between predictors are shown in purple. These IDRs were analyzed using coarse-grained (CG) simulations, and results are shown in comparison with those obtained from the SVR predictor or from ALBATROSS. Comparison between conformational entropy per residue ($Sconf/N$) estimates from CG simulations and those obtained from the SVR predictor is also shown. (B) Pearson’s correlation between $Sconf/N$ or v values (from the SVR predictor) with IDR length. (C) Pearson’s correlation between $Sconf/N$ or v values (from the SVR predictor) with different sequence features: average residue stickiness (λ), sequence hydropathy decoration (SHD), and fraction of charged residues (FCR).

Given the results above, we performed all the following analyses using the estimates obtained with the SVR model. As is the case in the human proteome, we observed that long IDRs tend to be more compact and have lower conformational entropy (Fig. 5B) (3). To assess the sequence features associated with IDR compaction, we next used the SVR predictor to calculate the following parameters: average residue stickiness (λ , a measurement of the strength of attractive intra-chain interactions [48]), sequence hydrophobic decoration (SHD, a measure of the patterning of hydrophobic residues [49]), and fraction of charged residues (FCR) (Table S4). Again, in line with data from the human proteome, we found that compact IDRs tend to have high values of λ and more clustered hydrophobic residues (high SHD), as well as fewer charged residues (Fig. 5C) (3). Overall, these results suggest that even if the representation of IDRs in the Mpn proteome is much lower than in human proteins, sequence-ensemble relationships tend to be conserved irrespective of whether proteins have a eukaryotic or prokaryotic origin. We finally sought to determine whether IDRs that are targets of PTMs are more or less compact compared to those that are not. To test this, we fitted linear regression models with v as the dependent variable and IDR size and presence/absence of individual PTMs as independent variables. This approach was motivated by the fact that, although v is a length-independent measure, we observed a relation with IDR size, and the probability of a PTM occurring in a given IDR will also depend on its length. Beyond the effect of length, we found a significant contribution of phosphorylation (estimate = 0.027, P -value = 0.0002), whereas the coefficients associated with acetylation and proteolysis were not significant (not shown). This indicates that phosphorylation preferentially occurs in extended IDRs.

IDRs in attachment organelle proteins are extended and have high conformational entropy

The GO term “adhesion of symbiont to microvasculature” includes five contributing proteins (Fig. 1E), all localizing to the attachment organelle, which is, however, formed by at least 14 proteins, organized in different structures. We thus investigated whether the organelle proteins not assigned to the GO term related to adhesion also had IDRs (Table S2). We found this to be the case, as all of them, excluding CpsG, have at least one IDR. In fact, these additional nine proteins have a significantly higher IDR fraction than all other Mpn proteins (Brunner-Munzel Test, P -value = 0.002).

We next sought to determine whether IDRs in proteins that constitute the attachment organelle have distinctive conformational features. We again used linear regression models that account for IDR size and found that IDRs in attachment organelle proteins are more extended (estimate = 0.018, P -value = 0.0003) and have higher conformational entropy (estimate = 0.064, P -value = 0.0013) compared to those in other Mpn proteins. In line with the results above, compared to other IDRs in the Mpn proteome, IDRs in attachment organelle proteins have lower λ (Brunner-Munzel Test, P -value = 1.7×10^{-9})

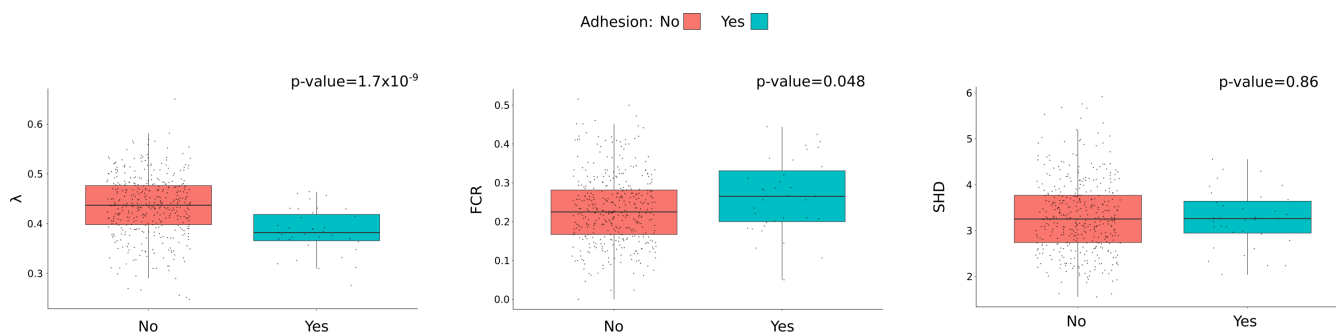


FIG 6 Analysis of IDRs located in attachment organelle proteins. Comparison of sequence features between IDRs that are located in attachment organelle proteins or in all other proteins. Data are shown in classic box and whisker plots. The P -values derive from Brunner-Munzel tests.

and higher FCR (Brunner-Munzel Test, P -value = 0.048). No difference in SHD was instead observed (Brunner-Munzel Test, P -value = 0.86) (Fig. 6).

DISCUSSION

In this study, we performed a proteome-wide investigation of IDRs in Mpn, which leverages multiple levels of biological information. We found that, compared to other bacteria (25–27), a considerable fraction of the Mpn proteome is embedded in IDRs, which are particularly abundant in membrane, non-essential proteins, as well as in proteins that mediate cytoadherence and, hence, virulence. Notably, proteins that form the attachment organelle, a structure that is only found in Mpn and a few closely related mycoplasmas (*M. genitalium* and *M. gallisepticum*) (29), have an extremely high IDR content. Gliding motility in Mpn is not chemotactic, but contributes to the infection process, enabling the bacterium to translocate from the tips of bronchial cilia to the host cell surface (50, 51). Thus, these data suggest that structural disorder contributes to very specific functions in Mpn and, likely, in other mycoplasmas. In line with this view, we show that, in terms of domain architecture, IDR content is not particularly high in proteins that display nucleic-acid binding modules, such as transcription factors (as is instead the case in eukaryotes [13, 52]). Rather, IDRs tend to be preferentially associated with prokaryote- or Mollicute-specific domains. Among these, the DUF31 domain has been suggested to represent a determinant of pathogenicity (35), whereas the EAGR box, being located in attachment organelle proteins, is involved in gliding mobility (34). For another domain, DUF5396, despite its specificity for Mollicutes, we detected significant structural homology with the oligopeptide-binding proteins of other bacteria. These membrane-bound proteins are responsible for the uptake of peptides and, thus, for providing essential metabolites, with some also showing the ability to bind heme (53). It is thus possible that Mpn proteins carrying DUF5396 domains function in nutrient acquisition. Whether IDRs modulate essential features of these proteins, including binding specificity, or whether they contribute to nutrient sensing remains an open question for future investigation. Likewise, proteins carrying lipoprotein domains may act as transporters (33). However, lipoproteins in different *Mycoplasma* species have also been shown to regulate host cell immune response (54). Interestingly, some Mpn lipoproteins in family 2, which carry the lipoprotein X and lipoprotein 10 domains (33), and that we show here to have particularly high fractions of residues in IDRs, are differentially expressed when the bacterium gets in contact with human lung epithelial cells or in oxidative/acidic stress conditions (55). Thus, these proteins may play a role in adaptation to different environmental conditions and, possibly, contribute to virulence. Overall, these observations suggest that IDRs are common in Mpn proteins that either directly affect adherence, motility, and virulence or that play a role in the broader ability of the bacterium to survive and thrive within its host. We were also interested in determining whether, despite the fact that they are often located in proteins with mycoplasma-specific functions, IDRs in Mpn proteins share similar functional, biochemical, and conformational features with IDRs in other organisms. We found mixed evidence, depending on the analyzed features. For instance, we found that, as is the case for *E. coli*, yeast, and mammals, IDR content affects protein degradation rates and half-life (38–40). Protein degradation is mediated by different mechanisms in eukaryotes and prokaryotes. In the former, the covalent binding of ubiquitin marks proteins for degradation by the proteasome, whereas bacteria use ATP-dependent proteases for substrate degradation (56), and Mpn encodes only two such proteases (Lon and FtsH [41]). In eukaryotes, the lengths of the disordered regions that affect protein half-life were shown to be compatible with the structure of the proteasome (40). In bacteria, exposure of aromatic or hydrophobic residues in IDRs may promote degradation (41, 57). Thus, the shorter half-life of proteins with IDRs seems to be mediated by independent mechanisms in eukaryotes and prokaryotes. The two Mpn proteases belong to the AAA+ (ATPases associated with various cellular activities) family, hexameric enzymes that recognize specific substrates and use ATP hydrolysis to mechanically unfold the target protein and

translocate it into the degradation chamber (58, 59). Initiation of degradation requires the recognition of a specific peptide signal (a degron) and the engagement of an unstructured segment of the substrate. The unstructured segment can be N-terminal, C-terminal, or internal (58–60). It is thus possible that the intrinsic lack of structure of IDRs facilitates engagement by the proteases and thus promotes protein degradation. In this respect, it is worth noting that we found a significant association between IDR presence and degradation rates, whereas the correlation between IDR fraction and protein degradation was weak. This suggests that the presence of unstructured segments rather than their length modulates protein turnover. An interesting possibility is that evolution has developed mechanisms to increase the degradation rate of IDR-containing proteins, possibly to control their aggregate-forming propensity. Just like IDRs, PTMs are less common in prokaryotes compared to eukaryotes. Our analysis of proteome-wide studies showed that, as in the case of eukaryotes, phosphorylation is common within Mpn IDRs, particularly those that assume an extended conformation, which may make them even more accessible to kinases (43, 61). This is interesting because phosphorylation, which equates to the gain of a negative charge, can impact IDR ensemble features and, as a consequence, properties such as binding affinity and protein interactions (1). Conversely, we did not find an increased frequency of acetylation or proteolytic processing in IDRs. Especially the latter finding was surprising, as IDRs are known to be intrinsically more susceptible to degradation, in line with the results we obtained for protein half-life (62). However, the controlled proteolysis of Mpn proteins, rather than serving purposes related to cellular homeostasis, is thought to function as a strategy to diversify the minimal proteome of the bacterium so that cleaved fragments can moonlight at the cell surface (45, 46). The products resulting from proteolysis were suggested to be more disordered than the original proteins, as it may be expected when cleavage disrupts the architectures of folded domains. Such products also engage in more protein-protein and protein-nucleic acid interactions, possibly due to the exposure of short linear motifs that are inaccessible within folded domains (46). Thus, if disorder is functional to the adaptive strategy of Mpn, it may be advantageous to use proteolytic sites within structured regions rather than in already disordered segments.

Recently, methods based on CG simulations and deep learning have allowed the large-scale analysis of IDR conformational properties, which are, in turn, related to function (1). To obtain ensemble features for Mpn IDRs, we used two methods, based on different force fields, which were shown to have very high concordance when human IDRs were analyzed (3, 47). Consistently, we found an overall good agreement between the two methods, although the correlation coefficient was not particularly high. We also observed that the prediction for some IDRs was quite divergent. Among these, the largest discrepancies were observed for IDRs that were predicted to be much more compact by ALBATROSS than by SVR. We thus used CG simulations to determine which predictions better fit the simulated ensembles. Specifically, we performed CG simulations under the CALVADOS model, and it is thus probably expected that we observed a better fit with the cognate predictor. However, some differences between the simulations and the predictor were still observable, and two IDRs that were predicted to have extremely extended ensembles ($\nu > 0.6$) were instead found to be more compact with the simulations. Overall, these data indicate that, whereas the deep learning predictors provide an overall good picture of the ensemble properties of Mpn IDRs, some of the latter might fail to be accurately described using these approaches. More generally, our data call for a wider benchmarking of computational methods to predict IDR ensemble properties, especially when these are applied to prokaryotes or other organisms whose IDRs were not used for predictor training. These observations notwithstanding, our results consistently show that ensemble properties in Mpn IDRs are mediated by similar sequence features as in eukaryotes and that compact IDRs tend to have high residue stickiness, high hydropathy decoration, and few charged residues (3, 47).

Remarkably, when we analyzed IDRs in attachment organelle proteins, we found that they do not only cover a considerable portion of the protein sequence, but are

also particularly extended and display high conformational entropy. The attachment organelle forms at a cell pole as a protrusion comprising surface nap-like structures (the P1/P90/P40 adhesin complexes) and an internal rod-like cytoskeletal core surrounded by a translucent area, which is thought to be occupied by stiff material. The nap-like structures interact with sialic acids at the target cell surface, and the source of energy is thought to derive from ATP hydrolysis, although the underlying mechanisms are unknown (29, 63, 64). Very recent evidence showed that, when the conformational entropy of an IDR is constrained, for instance by tethering to a surface or by confinement in periplasmic space, an effective entropic force is generated (65, 66). This mechanism is, for instance, exploited by Gram-positive bacteria to drive protein translocation through the cell wall in the absence of chaperones or energy sources (65). An interesting possibility is that a similar force is generated in the attachment organelle of Mpn by the IDRs, which we show here to have high conformational entropy. Their confinement by the stiff translucent area or other structures might generate an entropic force that contributes to movement. This hypothesis is supported by the observation that phosphorylation of proteins in the attachment organelle, which is, in turn, expected to modulate ensemble properties, affects gliding motility (67–69). It is also worth noting that IDRs were previously shown to contribute to flagellum-mediated swimming and swarming in *gamma*proteobacteria (12). Compared to the attachment organelle, the flagellum has a completely different structure, and the small, disordered proteins that mediate motility in this system play unknown roles (12). However, structural disorder seems to have been exploited independently in distinct bacteria to promote different types of motility. Future studies will be necessary to investigate the role of IDRs and the overall prevalence of these sequences in proteins that mediate motility in bacteria.

In summary, our results indicate that structural disorder contributes to very specialized functions in Mpn and most likely affects virulence, motility, and nutrient scavenging. Our data also highlight the functional relevance of IDRs: despite genome reduction, whereby a number of pathways were lost or simplified, the minimal proteome of Mpn displays a considerable level of structural disorder.

MATERIALS AND METHODS

Proteome and disorder prediction

The reference proteome of Mpn strain ATCC 29,342/M129 (UP000000808) was downloaded from UniProt. Of 686 proteins, 685 were available in the AlphaFold structure database and had the same length as in UniProt (Table S2).

To identify IDRs, we used two methods: an AlphaFold2-based method, as recently reported in a study that analyzed the human proteome (3) and Metapredict V2 (23, 24). Briefly, we used the protti R package (70) to derive pLDDT scores from the AlphaFold structure database (<https://alphafold.ebi.ac.uk/>) (71). We next generated window-averaged pLDDT scores using a window size of 15 AA (72). Residues with $\langle \text{pLDDT} \rangle > 0.8$ were considered as folded, and those with $\langle \text{pLDDT} \rangle < 0.7$ were labeled as disordered. Residues with $0.7 \leq \langle \text{pLDDT} \rangle \leq 0.8$ were initially defined as gap regions. Next, folded and disordered regions shorter than 10 residues were reclassified as gaps. Gap regions were then reassigned to the disordered fraction if they were flanked by disordered regions on both edges or on one single edge when N- or C-terminal. All other gap regions were instead relabeled as folded. Disordered regions shorter than 30 residues were not considered IDRs (3) (Table S4).

The Metapredict tool defines IDRs by applying a deep-learning algorithm based on a consensus score calculated from eight different disorder predictors (23). Metapredict V2 was run using default parameters, and IDRs were defined as consecutive disordered stretches longer than 30 residues.

Comparison with experimentally solved protein structures

To evaluate the performance of the AlphaFold-based method and of Metapredict, we compared the IDRs detected by the two approaches with experimentally solved structures, where regions that fail to be resolved (missing residues) are likely to represent IDRs (73). The structures of all available Mpn solved structures were downloaded from PDB (22) and are reported in Table S1. The PDB files were parsed to annotate the protein region that was used for structure definition and the missing residues (gaps). Gaps longer than 30 were considered IDRs. We next compared the gap locations from the PDB files with the AlphaFold and Metapredict IDR predictions. Specifically, we calculated specificity ($TN/(TN + FP)$), sensitivity ($TP/(TP + FN)$), and accuracy ($(TP + TN)/(TP + TN + FP + FN)$), where TP = true positive, TN = true negative, FP = false positive, and FN = false negative.

Functional characterization of Mpn proteins

Data of protein essentiality were retrieved from a transposon insertion mutagenesis study (28). A list of virulence factors was derived from the MicroScope website using the VirulenceFinder tool with default parameters (<https://mage.genoscope.cns.fr/>) (30). Only experimentally validated hits were retained (Table S2). Information about Mpn protein complexes was obtained from a previous work (31). GO terms were derived from QuickGO (<https://www.ebi.ac.uk/QuickGO/>) using the protti R package (70) (Table S2).

Protein domains, structure similarity search, and structural alignment

Domains were identified and annotated for all proteins in the Mpn proteome using the SMART batch access script (<http://smart.embl-heidelberg.de/smart/batch.pl>) with default parameters and inclusion of Pfam domains (74) (Table S3).

Structural models of P75324 and P75327 were obtained from the AlphaFold database (alphafold.ebi.ac.uk) (71). Foldseek (<https://search.foldseek.com/>) and DALI (<http://ekhidna2.biocenter.helsinki.fi/dali/>) were run using default parameters (36, 37, 75). For pairwise structure alignment, TM-align (76) was run from the dedicated website (<https://zhanggroup.org/TM-align/>) using the models of P75324 and P75327 and the resolved structure of the OppA protein from *Bacillus subtilis* (PDB ID: 8arn).

Protein degradation rates, half-life, and PTMs

Measures of translational efficiency, protein abundance, degradation rates, and protein half-lives were derived from recent works (41, 42).

Data on phosphorylation and acetylation sites were retrieved from the study by van Noort and coworkers (44). Proteolytic processing data were derived from Reference (46).

Analysis of IDR conformational properties and sequence patterns

The Flory scaling exponent (ν) was calculated for all Mpn IDRs identified using the AlphaFold2-based approach described above (77). ν was estimated using two methods: (i) a Colab notebook (https://colab.research.google.com/github/KULL-Centre/_2023_Tesei_IDRome/blob/main/IDR_SVR_predictor.ipynb), which uses a support vector regression model trained on simulations performed using the CALVADOS model (48, 78), and (ii) a second Colab notebook (<https://colab.research.google.com/github/holehouse-lab/>

[ALBATROSS-colab/blob/main/example_notebooks/polymer_property_predictors.ipynb](https://colab.research.google.com/github/holehouse-lab/ALBATROSS-colab/blob/main/example_notebooks/polymer_property_predictors.ipynb)) that implements the ALBATROSS (A deep-Learning Based Approach for predicting pRoperties Of disordered proteinS) predictor (47). Both models predict conformational properties starting from IDR amino acid sequences. The SVR predictor was also used to derive other measures: conformational entropy per residue (S_{conf}/N), the fraction of charged residues (FCR), the average residue stickiness (λ , a measurement of the strength of attractive intra-chain interactions), and the sequence hydrophobicity decoration (SHD, a measure of the patterning of hydrophobic residues) (Table S4).

Molecular simulations

The conformational ensembles of IDRs were generated using the CALVADOS 2 coarse-grained model (3, 48). MD simulations were conducted under NVT conditions, employing the Langevin integrator with a 10 fs time step, and setting temperature and ionic strength at 310 K and 0.150 M, respectively. Both termini were modeled as charged, while histidine residues were kept in neutral form. MD simulations were performed for a total of 44 IDRs of lengths ranging from 32 to 840 residues, and depending on sequence length, the simulation time was calculated according to the guidelines of (3). More in detail, the total simulation steps are set as $1010 \times (\text{number of saved frames})$, and converted to nanoseconds as $ns = (\text{number of steps}) \times 0.01 / 1000$. The number of saved frames is 7,000 for IDRs shorter than 150 residues, while for longer ones, it is calculated using the quadratic formula $3e^{-4} \times (\text{number of residues})^2 \times 1,000$. To improve sampling, four independent trajectories were simulated for each sequence. The properties of interest, v and S_{conf}/N , were calculated as described in Reference (3), with their values averaged across the four replicas to ensure robust statistics.

Statistical analysis

All statistical analyses were performed in the R v.4.0.5 environment. The Fisher's exact test was used when comparing categorical data that can be structured as a contingency table (e.g., counts of proteins with or without IDRs in different essentiality classes). The test is robust to small samples. Pearson's correlation was used to measure the degree of the relationship between linearly related variables. Kendall's correlation is a non-parametric test that measures the strength of dependence between two variables based on the ranks of the data. Kendall's correlation is reliable even in the presence of ties (79). The Exact binomial test compares the observed frequencies of the two categories of a dichotomous variable (e.g., being or not a PTM site) to the frequencies that are expected under a binomial distribution. The test is appropriate when the sample size is small compared to the population subject to inference. Linear models were used to analyze IDR features as a function of different parameters (e.g., belonging to the attachment organelle or not) accounting for size. The Brunner-Munzel test was performed using the `brunnermunzel` R package. This test is used to compare the distribution of a continuous variable (e.g., IDR fraction) between two groups. It was chosen because it is not influenced by tied values, and it does not require the assumption of equal variances between groups (80). Statistical details about each comparison can be found in the figure legends.

ACKNOWLEDGMENTS

The work was supported by the Italian Ministry of Health—"Ricerca Corrente" program (to R.C.) and by the European Union—NextGenerationEU through the Italian Ministry of University and Research under PNRR-M4C2-I1.3 Project PE_00000019 "HEAL ITALIA" (to L.D.G.).

AUTHOR AFFILIATIONS

¹Scientific Institute IRCCS E. MEDEA, Computational Biology Unit, Bosisio Parini, Italy

²Department of Biotechnology and Biosciences, University of Milan-Bicocca, Milan, Italy

³School of Medicine and Surgery, University of Milano-Bicocca, Monza, Italy

AUTHOR ORCIDs

Uberto Pozzoli  <http://orcid.org/0000-0003-0670-7106>

Diego Forni  <http://orcid.org/0000-0001-9291-5352>

Alessandra Mozzi  <http://orcid.org/0000-0003-3911-1042>

Rachele Cagliani  <https://orcid.org/0000-0003-2670-3532>

FUNDING

Funder	Grant(s)	Author(s)
Ministero della Salute	Ricerca Corrente	Rachele Cagliani
Ministero dell'Università e della Ricerca	PE_00000019 "HEAL ITALIA	Luca De Gioia

DATA AVAILABILITY

Data supporting the findings of this study are available within the article and its supplemental material.

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Table S1 (Spectrum00968-25-s0001.xlsx). Mpn proteins with an experimentally solved structure and their not solved regions.

Table S2 (Spectrum00968-25-s0002.xlsx). List of proteins in the Mpn proteome.

Table S3 (Spectrum00968-25-s0003.xlsx). List of domains associated with IDRs.

Table S4 (Spectrum00968-25-s0004.xlsx). List of Mpn IDRs.

REFERENCES

- Holehouse AS, Kragelund BB. 2024. The molecular basis for cellular function of intrinsically disordered protein regions. *Nat Rev Mol Cell Biol* 25:187–211. <https://doi.org/10.1038/s41580-023-00673-0>
- Pritišanac I, Alderson TR, Kolarić Đ, Zarin T, Xie S, Lu A, Alam A, Maqsood A, Youn J-Y, Forman-Kay JD, Moses AM. 2024. A functional map of the human intrinsically disordered proteome. *Bioinformatics*. <https://doi.org/10.1101/2024.03.15.585291>
- Tesei G, Trolle AI, Jonsson N, Betz J, Knudsen FE, Pesce F, Johansson KE, Lindorff-Larsen K. 2024. Conformational ensembles of the human intrinsically disordered proteome. *Nature* 626:897–904. <https://doi.org/10.1038/s41586-023-07004-5>
- Mozzi A, Forni D, Cagliani R, Clerici M, Pozzoli U, Sironi M. 2020. Intrinsically disordered regions are abundant in simplexvirus proteomes and display signatures of positive selection. *Virus Evol* 6:veaa028. <https://doi.org/10.1093/ve/veaa028>
- Peng Z, Yan J, Fan X, Mizianty MJ, Xue B, Wang K, Hu G, Uversky VN, Kurgan L. 2015. Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell Mol Life Sci* 72:137–151. <https://doi.org/10.1007/s00018-014-1661-9>
- Pushker R, Mooney C, Davey NE, Jacqué J-M, Shields DC. 2013. Marked variability in the extent of protein disorder within and between viral families. *PLoS One* 8:e60724. <https://doi.org/10.1371/journal.pone.0060724>
- Xue B, Blocquel D, Habchi J, Uversky AV, Kurgan L, Uversky VN, Longhi S. 2014. Structural disorder in viral proteins. *Chem Rev* 114:6880–6911. <https://doi.org/10.1021/cr4005692>
- Charon J, Barra A, Walter J, Millot P, Hébrard E, Moury B, Michon T. 2018. First experimental assessment of protein intrinsic disorder involvement in an RNA virus natural adaptive process. *Mol Biol Evol* 35:38–49. <https://doi.org/10.1093/molbev/msx249>
- Lauber C, Kazem S, Kravchenko AA, Feltkamp MCW, Gorbalenya AE. 2015. Interspecific adaptation by binary choice at *de novo* polyomavirus T antigen site through accelerated codon-constrained Val-Ala toggling within an intrinsically disordered region. *Nucleic Acids Res* 43:4800–4813. <https://doi.org/10.1093/nar/gkv378>
- Ortiz JF, MacDonald ML, Masterson P, Uversky VN, Siltberg-Liberles J. 2013. Rapid evolutionary dynamics of structural disorder as a potential driving force for biological divergence in flaviviruses. *Genome Biol Evol* 5:504–513. <https://doi.org/10.1093/gbe/evt026>
- Brunet YR, Habib C, Brogan AP, Artzi L, Rudner DZ. 2022. Intrinsically disordered protein regions are required for cell wall homeostasis in *Bacillus subtilis*. *Genes Dev* 36:970–984. <https://doi.org/10.1101/gad.349895.122>
- Oguri T, Kwon Y, Woo JKK, Pehna G, Lee H, Ning M, Won K-J, Lee J, Mei S, Shi Y, Jeong H, Lee H. 2019. A family of small intrinsically disordered proteins involved in flagellum-dependent motility in *Salmonella enterica*. *J Bacteriol* 201:e00415-18. <https://doi.org/10.1128/JB.00415-18>
- van der Lee R, Buljan M, Lang B, Weatheritt RJ, Daughdrill GW, Dunker AK, Fuxreiter M, Gough J, Gsponer J, Jones DT, Kim PM, Kriwacki RW, Oldfield CJ, Pappu RV, Tompa P, Uversky VN, Wright PE, Babu MM. 2014. Classification of intrinsically disordered regions and proteins. *Chem Rev* 114:6589–6631. <https://doi.org/10.1021/cr400525m>
- Busby JN, Trevelyan S, Pegg CL, Kerr ED, Schulz BL, Chassagnon I, Landsberg MJ, Weston MK, Hurst MRH, Lott JS. 2024. The ABC toxin complex from *Yersinia entomophaga* can package three different cytotoxic components expressed from distinct genetic loci in an unfolded state: the structures of both shell and cargo. *IUCrJ* 11:299–308. <https://doi.org/10.1107/S2052252524001969>
- Byeon C-H, Hansen KH, Jeffrey J, Saricayir H, Andreasen M, Akbey Ü. 2024. Intrinsically disordered *Pseudomonas chaperone* FapA slows down the fibrillation of major biofilm-forming functional amyloid FapC. *FEBS J* 291:1925–1943. <https://doi.org/10.1111/febs.17084>
- Hadži S, Živič Z, Kovačić M, Zavrtnik U, Haesaerts S, Charlier D, Plavec J, Volkov AN, Lah J, Loris R. 2024. Fuzzy recognition by the prokaryotic transcription factor HigaA2 from *Vibrio cholerae*. *Nat Commun* 15:3105. <https://doi.org/10.1038/s41467-024-47296-3>
- Mets T, Kurata T, Ernits K, Johansson MJO, Craig SZ, Evora GM, Buttress JA, Odai R, Wallant KC, Nakamoto JA, Shyrokovna L, Egorov AA, Doering CR, Brodiazhenko T, Laub MT, Tenson T, Strahl H, Martens C, Harms A, Garcia-Pino A, Atkinson GC, Haurlyuk V. 2024. Mechanism of phage sensing and restriction by toxin-antitoxin-chaperone systems. *Cell Host Microbe* 32:1059–1073. <https://doi.org/10.1016/j.chom.2024.05.003>
- Ostan NKH, Cole GB, Wang FZ, Reichheld SE, Moore G, Pan C, Yu R, Lai CC-L, Sharpe S, Lee HO, Schryvers AB, Moraes TF. 2024. A secreted bacterial protein protects bacteria from cationic antimicrobial peptides by entrapment in phase-separated droplets. *PNAS Nexus* 3:pgae139. <https://doi.org/10.1093/pnasnexus/pgae139>
- Hutchison CA, Chuang R-Y, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L, Pelletier JF, Qi Z-Q, Richter RA, Strychalski EA, Sun L, Suzuki Y, Tsvetanova B, Wise KS, Smith HO, Glass JL, Merryman C, Gibson DG, Venter JC. 2016. Design and synthesis of a minimal bacterial genome. *Science* 351:aad6253. <https://doi.org/10.1126/science.aad6253>
- Sirand-Pugnet P, Citti C, Barré A, Blanchard A. 2007. Evolution of molluscs: down a bumpy road with twists and turns. *Res Microbiol* 158:754–766. <https://doi.org/10.1016/j.resmic.2007.09.007>

21. Hu J, Ye Y, Chen X, Xiong L, Xie W, Liu P. 2023. Insight into the pathogenic mechanism of *Mycoplasma pneumoniae*. *Curr Microbiol* 80:14. <https://doi.org/10.1007/s00284-022-03103-0>
22. Burley SK, Bhatt R, Bhikadiya C, Bi C, Biester A, Biswas P, Bittrich S, Blaumann S, Brown R, Chao H, et al. 2025. Updated resources for exploring experimentally-determined PDB structures and Computed Structure Models at the RCSB Protein Data Bank. *Nucleic Acids Res* 53:D564–D574. <https://doi.org/10.1093/nar/gkae1091>
23. Emenecker RJ, Griffith D, Holehouse AS. 2021. Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure. *Biophys J* 120:4312–4319. <https://doi.org/10.1016/j.bpj.2021.08.039>
24. Emenecker RJ, Griffith D, Holehouse AS. 2022. Metapredict V2: an update to metapredict, a fast, accurate, and easy-to-use predictor of consensus disorder and structure. *Bioinformatics*. <https://doi.org/10.1101/2022.06.06.494887>
25. Basile W, Salvatore M, Bassot C, Elofsson A. 2019. Why do eukaryotic proteins contain more intrinsically disordered regions? *PLoS Comput Biol* 15:e1007186. <https://doi.org/10.1371/journal.pcbi.1007186>
26. Uversky VN. 2019. Intrinsically disordered proteins and their “Mysterious” (Meta)Physics. *Front Phys* 7:10. <https://doi.org/10.3389/fphy.2019.00010>
27. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT. 2004. Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 337:635–645. <https://doi.org/10.1016/j.jmb.2004.02.002>
28. Lluch-Senar M, Delgado J, Chen W-H, Lloréns-Rico V, O'Reilly FJ, Wodke JA, Unal EB, Yus E, Martínez S, Nichols RJ, Ferrar T, Vivancos A, Schmeisky A, Stülke J, van Noort V, Gavin A-C, Bork P, Serrano L. 2015. Defining a minimal cell: essentiality of small ORFs and ncRNAs in a genome-reduced bacterium. *Mol Syst Biol* 11:780. <https://doi.org/10.15252/msb.20145558>
29. Mizutani M, Sasajima Y, Miyata M. 2021. Force and stepwise movements of gliding motility in human pathogenic bacterium *Mycoplasma pneumoniae*. *Front Microbiol* 12:747905. <https://doi.org/10.3389/fmicb.2021.747905>
30. Vallend D, Calteau A, Dubois M, Amours P, Bazin A, Beuvin M, Burlot L, Bussell X, Fouteau S, Gautreau G, Lajus A, Langlois J, Planel R, Roche D, Rollin J, Rouy Z, Sabatet V, Médigue C. 2019. MicroScope: an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Res* gkz2926. <https://doi.org/10.1093/nar/gkz2926>
31. Kühner S, van Noort V, Betts MJ, Leo-Macias A, Batisse C, Rode M, Yamada T, Maier T, Bader S, Beltran-Alvarez P, et al. 2009. Proteome organization in a genome-reduced bacterium. *Science* 326:1235–1240. <https://doi.org/10.1126/science.1176343>
32. Hayashi S, Wu HC. 1990. Lipoproteins in bacteria. *J Bioenerg Biomembr* 22:451–471. <https://doi.org/10.1007/BF00763177>
33. Hallamaa KM, Browning GF, Tang SL. 2006. Lipoprotein multigene families in *Mycoplasma pneumoniae*. *J Bacteriol* 188:5393–5399. <https://doi.org/10.1128/JB.01819-05>
34. Burgos R, Pich OQ, Querol E, Piñol J. 2007. Functional analysis of the *Mycoplasma genitalium* MG312 protein reveals a specific requirement of the MG312 N-terminal domain for gliding motility. *J Bacteriol* 189:7014–7023. <https://doi.org/10.1128/JB.00975-07>
35. Arfi Y, Minder L, Di Primo C, Le Roy A, Ebel C, Coquet L, Claverol S, Vashee S, Jores J, Blanchard A, Sirand-Pugnet P. 2016. MIB-MIP is a mycoplasma system that captures and cleaves immunoglobulin G. *Proc Natl Acad Sci USA* 113:5406–5411. <https://doi.org/10.1073/pnas.1600546113>
36. Holm L. 2020. DALI and the persistence of protein shape. *Protein Sci* 29:128–140. <https://doi.org/10.1002/pro.3749>
37. van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, Söding J, Steinegger M. 2024. Fast and accurate protein structure search with Foldseek. *Nat Biotechnol* 42:243–246. <https://doi.org/10.1038/s41587-023-01773-0>
38. Nagar N, Ecker N, Loewenthal G, Avram O, Ben-Meir D, Biran D, Ron E, Pupko T. 2021. Harnessing machine learning to unravel protein degradation in *Escherichia coli*. *mSystems* 6:e01296-20. <https://doi.org/10.1128/mSystems.01296-20>
39. Tompa P, Prilusky J, Silman I, Sussman JL. 2008. Structural disorder serves as a weak signal for intracellular protein degradation. *Proteins* 71:903–909. <https://doi.org/10.1002/prot.21773>
40. van der Lee R, Lang B, Kruse K, Gspöner J, Sánchez de Groot N, Huynen MA, Matouschek A, Fuxreiter M, Babu MM. 2014. Intrinsically disordered segments affect protein half-life in the cell and during evolution. *Cell Rep* 8:1832–1844. <https://doi.org/10.1016/j.celrep.2014.07.055>
41. Burgos R, Weber M, Martínez S, Lluch-Senar M, Serrano L. 2020. Protein quality control and regulated proteolysis in the genome-reduced organism *Mycoplasma pneumoniae*. *Mol Syst Biol* 16:e9530. <https://doi.org/10.15252/msb.20209530>
42. Weber M, Sogues A, Yus E, Burgos R, Gallo C, Martínez S, Lluch-Senar M, Serrano L. 2023. Comprehensive quantitative modeling of translation efficiency in a genome-reduced bacterium. *Mol Syst Biol* 19:e11301. <https://doi.org/10.15252/msb.202211301>
43. Bludau I, Willems S, Zeng W-F, Strauss MT, Hansen FM, Tanzer MC, Karayel O, Schulman BA, Mann M. 2022. The structural context of posttranslational modifications at a proteome-wide scale. *PLoS Biol* 20:e3001636. <https://doi.org/10.1371/journal.pbio.3001636>
44. van Noort V, Seebacher J, Bader S, Mohammed S, Vonkova I, Betts MJ, Kühner S, Kumar R, Maier T, O'Flaherty M, Rybin V, Schmeisky A, Yus E, Stülke J, Serrano L, Russell RB, Heck AJR, Bork P, Gavin A-C. 2012. Cross-talk between phosphorylation and lysine acetylation in a genome-reduced bacterium. *Mol Syst Biol* 8:571. <https://doi.org/10.1038/msb.2012.4>
45. Arfi Y, Lartigue C, Sirand-Pugnet P, Blanchard A. 2021. Beware of *Mycoplasma* anti-immunoglobulin strategies. *MBio* 12:e0197421. <https://doi.org/10.1128/mBio.01974-21>
46. Berry IJ, Widjaja M, Jarocki W, Steele JR, Padula MP, Djordjevic SP. 2021. Protein cleavage influences surface protein presentation in *Mycoplasma pneumoniae*. *Sci Rep* 11:6743. <https://doi.org/10.1038/s41598-021-86217-y>
47. Lotthammer JM, Ginell GM, Griffith D, Emenecker RJ, Holehouse AS. 2024. Direct prediction of intrinsically disordered protein conformational properties from sequence. *Nat Methods* 21:465–476. <https://doi.org/10.1038/s41592-023-02159-5>
48. Tesei G, Lindorff-Larsen K. 2023. Improved predictions of phase behaviour of intrinsically disordered proteins by tuning the interaction range. *Open Res Europe* 2:94. <https://doi.org/10.12688/openreseurope.14967.2>
49. Zheng W, Dignon G, Brown M, Kim YC, Mittal J. 2020. Hydrophathy patterning complements charge patterning to describe conformational preferences of disordered proteins. *J Phys Chem Lett* 11:3408–3415. <https://doi.org/10.1021/acs.jpclett.0c00288>
50. Krunkosky TM, Jordan JL, Chambers E, Krause DC. 2007. *Mycoplasma pneumoniae* host-pathogen studies in an air-liquid culture of differentiated human airway epithelial cells. *Microb Pathog* 42:98–103. <https://doi.org/10.1016/j.micpath.2006.11.003>
51. Prince OA, Krunkosky TM, Krause DC. 2014. *In vitro* spatial and temporal analysis of *Mycoplasma pneumoniae* colonization of human airway epithelium. *Infect Immun* 82:579–586. <https://doi.org/10.1128/IAI.01036-13>
52. Liu J, Perumal NB, Oldfield CJ, Su EW, Uversky VN, Dunker AK. 2006. Intrinsic disorder in transcription factors. *Biochemistry* 45:6873–6888. <https://doi.org/10.1021/bi0602718>
53. Tanaka KJ, Pinkett HW. 2019. Oligopeptide-binding protein from nontypeable *Haemophilus influenzae* has ligand-specific sites to accommodate peptides and heme in the binding pocket. *J Biol Chem* 294:1070–1082. <https://doi.org/10.1074/jbc.RA118.004479>
54. Christodoulides A, Gupta N, Yacoubian V, Maithel N, Parker J, Kelesidis T. 2018. The role of lipoproteins in *Mycoplasma*-mediated immunomodulation. *Front Microbiol* 9:1682. <https://doi.org/10.3389/fmicb.2018.01682>
55. Hallamaa KM, Tang S-L, Ficorilli N, Browning GF. 2008. Differential expression of lipoprotein genes in *Mycoplasma pneumoniae* after contact with human lung epithelial cells, and under oxidative and acidic stress. *BMC Microbiol* 8:124. <https://doi.org/10.1186/1471-2180-8-124>
56. Mahmoud SA, Chien P. 2018. Regulated proteolysis in bacteria. *Annu Rev Biochem* 87:677–696. <https://doi.org/10.1146/annurev-biochem-062917-012848>
57. Gur E, Vishkautzan M, Sauer RT. 2012. Protein unfolding and degradation by the AAA+ Lon protease. *Protein Sci* 21:268–278. <https://doi.org/10.1002/pro.2013>
58. Olivares AO, Baker TA, Sauer RT. 2016. Mechanistic insights into bacterial AAA+ proteases and protein-remodelling machines. *Nat Rev Microbiol* 14:33–44. <https://doi.org/10.1038/nrmicro.2015.4>
59. Sauer RT, Baker TA. 2011. AAA+ proteases: ATP-fueled machines of protein destruction. *Annu Rev Biochem* 80:587–612. <https://doi.org/10.1146/annurev-biochem-060408-172623>

60. Morehouse JP, Baker TA, Sauer RT. 2022. FtsH degrades dihydrofolate reductase by recognizing a partially folded species. *Protein Sci* 31:e4410. <https://doi.org/10.1002/pro.4410>
61. lakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK. 2004. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* 32:1037–1049. <https://doi.org/10.1093/nar/gkh253>
62. Suskiewicz MJ, Sussman JL, Silman I, Shaul Y. 2011. Context-dependent resistance to proteolysis of intrinsically disordered proteins. *Protein Sci* 20:1285–1297. <https://doi.org/10.1002/pro.657>
63. Aparicio D, Scheffer MP, Marcos-Silva M, Vizarraga D, Sprankel L, Ratera M, Weber MS, Seybert A, Torres-Puig S, Gonzalez-Gonzalez L, Reitz J, Querol E, Piñol J, Pich OQ, Fita I, Frangakis AS. 2020. Structure and mechanism of the Nap adhesion complex from the human pathogen *Mycoplasma genitalium*. *Nat Commun* 11:2877. <https://doi.org/10.1038/s41467-020-16511-2>
64. Vizarraga D, Kawamoto A, Matsumoto U, Illanes R, Pérez-Luque R, Martín J, Mazzolini R, Bierge P, Pich OQ, Espasa M, Sanfeliu I, Esperalba J, Fernández-Huerta M, Scheffer MP, Pinyol J, Frangakis AS, Lluch-Senar M, Mori S, Shibayama K, Kenri T, Kato T, Namba K, Fita I, Miyata M, Aparicio D. 2020. Immunodominant proteins P1 and P40/P90 from human pathogen *Mycoplasma pneumoniae*. *Nat Commun* 11:5188. <https://doi.org/10.1038/s41467-020-18777-y>
65. Halladin DK, Ortega FE, Ng KM, Footer MJ, Mitić NS, Malkov SN, Gopinathan A, Huang KC, Theriot JA. 2021. Entropy-driven translocation of disordered proteins through the Gram-positive bacterial cell wall. *Nat Microbiol* 6:1055–1065. <https://doi.org/10.1038/s41564-021-00942-8>
66. Yu F, Sukenik S. 2023. Structural preferences shape the entropic force of disordered protein ensembles. *J Phys Chem B* 127:4235–4244. <https://doi.org/10.1021/acs.jpcc.3c00698>
67. Miyata M, Hamaguchi T. 2016. Integrated information and prospects for gliding mechanism of the pathogenic bacterium *Mycoplasma pneumoniae*. *Front Microbiol* 7:960. <https://doi.org/10.3389/fmicb.2016.00960>
68. Page CA, Krause DC. 2013. Protein kinase/phosphatase function correlates with gliding motility in *Mycoplasma pneumoniae*. *J Bacteriol* 195:1750–1757. <https://doi.org/10.1128/JB.02277-12>
69. Schmidl SR, Gronau K, Hames C, Busse J, Becher D, Hecker M, Stülke J. 2010. The stability of cytoadherence proteins in *Mycoplasma pneumoniae* requires activity of the protein kinase PrkC. *Infect Immun* 78:184–192. <https://doi.org/10.1128/IAI.00958-09>
70. Quast J-P, Schuster D, Picotti P. 2022. protti: an R package for comprehensive data analysis of peptide- and protein-centric bottom-up proteomics data. *Bioinform Adv* 2:vbab041. <https://doi.org/10.1093/bioadv/vbab041>
71. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Židek A, Potapenko A, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596:583–589. <https://doi.org/10.1038/s41586-021-03819-2>
72. Akdel M, Pires DEV, Pardo EP, Jänes J, Zalevsky AO, Mészáros B, Bryant P, Good LL, Laskowski RA, Pozzati G, et al. 2022. A structural biology community assessment of AlphaFold2 applications. *Nat Struct Mol Biol* 29:1056–1067. <https://doi.org/10.1038/s41594-022-00849-w>
73. Maiti S, Singh A, Maji T, Saibo NV, De S. 2024. Experimental methods to study the structure and dynamics of intrinsically disordered regions in proteins. *Curr Res Struct Biol* 7:100138. <https://doi.org/10.1016/j.crsbtbi.2024.100138>
74. Letunic I, Khedkar S, Bork P. 2021. SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res* 49:D458–D460. <https://doi.org/10.1093/nar/gkaa937>
75. Holm L. 2022. Dali server: structural unification of protein families. *Nucleic Acids Res* 50:W210–W215. <https://doi.org/10.1093/nar/gkac387>
76. Zhang Y, Skolnick J. 2005. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res* 33:2302–2309. <https://doi.org/10.1093/nar/gki524>
77. Flory P, Volkenstein M. 1969. Statistical mechanics of chain molecules. *Biopolymers* 8:699–700. <https://doi.org/10.1002/bip.1969.360080514>
78. Tesei G, Schulze TK, Crehuet R, Lindorff-Larsen K. 2021. Accurate model of liquid-liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *Proc Natl Acad Sci USA* 118:e2111696118. <https://doi.org/10.1073/pnas.2111696118>
79. Salkind N. 2007. Encyclopedia of measurement and statistics. Sage Publications, Inc., Thousand Oaks California, USA. Available from: <https://methods.sagepub.com/reference/encyclopedia-of-measurement-and-statistics>
80. Brunner E, Munzel U. 2000. The nonparametric behrens-fisher problem: asymptotic theory and a small-sample approximation. *Biom J* 42:17–25. [https://doi.org/10.1002/\(SICI\)1521-4036\(200001\)42:1<17::AID-BIMJ17>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1521-4036(200001)42:1<17::AID-BIMJ17>3.0.CO;2-U)