Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

🔓 OPEN ACCESS   Check for updates

RNA-dependent proteome solubility maintenance in *Escherichia coli* lysates analysed by quantitative mass spectrometry: Proteomic characterization in terms of isoelectric point, structural disorder, functional hub, and chaperone network

Chan Park[a,b], Bitnara Han[c], Yura Choi[b,d], Yoontae Jin[b,e], Kwang Pyo Kim[c,f], Seong Il Choi[g], and Baik L. Seong[a,b]

[a]Department of Microbiology, College of Medicine, Yonsei University, Seoul, Korea; [b]Vaccine Innovative Technology ALliance (VITAL)-Korea, Yonsei University, Seoul, Korea; [c]Department of Applied Chemistry, Institute of Natural Science, Global Center for Pharmaceutical Ingredient Materials, Kyung Hee University, Yongin, Korea; [d]The Interdisciplinary Graduate Program in Integrative Biotechnology and Translational Medicine, Yonsei University, Incheon, Korea; [e]Department of Microbiology and Immunology, Institute for Immunology and Immunological Diseases, Graduate School of Medical Science, Brain Korea 21 Project, Yonsei University College of Medicine, Seoul, Korea; [f]Department of Biomedical Science and Technology, Kyung Hee Medical Science Research Institute, Kyung Hee University, Seoul, Republic of Korea; [g]Department of Pediatrics, Severance Hospital, Institute of Allergy, Brain Korea 21 PLUS Project for Medical Science, Yonsei University College of Medicine, Seoul, Korea

ABSTRACT

Protein aggregation, a consequence of misfolding and impaired proteostasis, can lead to cellular malfunctions such as various proteinopathies. The mechanisms protecting proteins from aggregation in complex cellular environments have long been investigated, often from a protein-centric viewpoint. However, our study provides insights into a crucial, yet overlooked actor: RNA. We found that depleting RNAs from *Escherichia coli* lysates induces global protein aggregation. Our quantitative mass spectrometry analysis identified over 900 statistically significant proteins from the *Escherichia coli* proteome whose solubility depends on RNAs. Proteome-wide characterization showed that the RNA dependency is particularly enriched among acidic proteins, intrinsically disordered proteins, and structural hub proteins. Moreover, we observed distinct differences in RNA-binding mode and Gene Ontology categories between RNA-dependent acidic and basic proteins. Notably, the solubility of key molecular chaperones [Trigger factor, DnaJ, and GroES] is largely dependent on RNAs, suggesting a yet-to-be-explored hierarchical relationship between RNA-based chaperone (termed as chaperna) and protein-based chaperones, both of which constitute the whole chaperone network. These findings provide new insights into the RNA-centric role in maintaining healthy proteome solubility *in vivo*, where proteins associate with a variety of RNAs, either stably or transiently.

Introduction

To perform their intrinsic biological functions, proteins adopt native forms and maintain solubility against aberrant aggregation in cellular environments [1–4]. Protein misfolding and aggregation are linked with severe proteinopathies, including Alzheimer's disease and Parkinson's disease [5]. How proteins maintain their solubility against aggregation has therefore been the key to understanding protein homeostasis, or proteostasis, and pathogenic consequences [6]. Despite decades of extensive research, the complex conformational landscapes governing protein folding, misfolding, and aggregation in the cellular milieu remain poorly defined. Molecular chaperones, known to facilitate the folding and native assembly of proteins by preventing misfolding and aggregation [7–9], continuously ensure proteostasis due to the inherent instability and misfolding propensity of proteins [10]. To date, proteostasis, encompassing assisted protein folding and inhibition of aggregation, has been largely understood in the context of protein-based molecular chaperones.

However, it is becoming increasingly evident that a variety of RNAs can also act as chaperones. The ribosomes and V-domains of 23S ribosomal RNA (rRNA) are known to promote protein folding *in vitro* [11,12]. Beyond their canonical adaptor function for translation, transfer RNAs (tRNAs) provide robust chaperone functions to their interacting partner proteins in both native and engineered systems [13–15]. It is proposed that RNAs generally act as chaperones – for their physically – whether directly or indirectly – connected polypeptides, partly due to the aggregation inhibition by the intermolecular repulsions of the complexed RNAs with negative surface charges and large excluded volumes [16–18]. These RNA-mediated chaperones are termed as chaperna, uniting chaperone and RNA [19]. *In vitro*, RNAs with diverse sequences exhibit a potent ability to prevent protein aggregation more efficiently than protein-based molecular chaperones like GroEL [20]. G-quadruplexes of RNA inhibit protein aggregation *in vitro* and promote the folding yield of reporter proteins *in vivo* [21]. Moreover, these

G-quadruplexes can also act as foldases that increase protein folding rate [22]. These examples show that RNAs can generally function as chaperones, either alone or in concert with other molecular chaperones.

RNAs can form stable or dynamic complexes with their partner proteins. The native intermolecular interactions of proteins with their ligands can impact protein stability, folding, and conformation [23–27]. A large fraction of the proteome includes intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) [28], which fold and form into native complexes upon association with their ligands [29]. Thus, cellular macromolecules, including RNAs, can play a role in maintaining the solubility of their partner proteins [30,31]. Notably, the M1 RNA ribozyme provides a chaperone function to its interacting C5 protein [32]. RNA-binding proteins, usually enriched with IDRs involved in RNA-binding [33–36], are potentially prone to aggregation in the absence of proper association with their cognate RNAs. Consistently, enzymatic degradation of RNA can lead to widespread protein aggregation in eukaryotic cell and tissue lysates [37].

The impact of RNA on maintaining proteome solubility in prokaryotic systems, such as *E. coli*, remains largely unexplored. Understanding the role of RNA on protein stability and aggregation in prokaryotes is crucial for unravelling the interplay between RNA and proteostasis across different organisms. Therefore, in this study, we investigated the effects of RNase treatment on the solubility of proteome in *E. coli* lysates especially employing quantitative mass spectrometry and conducting proteome-wide characterization. Our findings underscore the central role of RNA in the maintenance of proteome-wide solubility. They highlight its pivotal role in stabilizing functional hub and structural disorder, as well as in orchestrating molecular chaperone network.

Results

RNA-dependent global protein solubility maintenance

To investigate the contribution of RNAs to proteome solubility, RNAs were depleted from *E. coli* lysates using RNase A as illustrated in Figure 1A. The lysates were separated into three fractions through centrifugation: total (T_0), soluble (S_0), and insoluble (P_0). The S_0 fraction then underwent treatment with RNase A and incubation at 37°C for 15 minutes. Subsequently, we obtained fractions T_1, S_1, and P_1 for further analysis. In parallel, an untreated control group was divided into fractions T_2, S_2, and P_2. In the calculation of protein solubility (S/T) throughout this paper, 'S' and 'T' represent the soluble fraction and total fraction of protein, respectively. The solubility of expressed proteins is defined by the ratio of the soluble fraction to the total fraction of the protein [38].

Treatment with RNase A induced the aggregation of proteins across the entire spectrum, as confirmed by SDS-PAGE analysis (Figure 1B, upper panel). In the RNase-treated samples, approximately $60.53 \pm 1.55\%$ (= S_1/T_1) of the total proteins remained soluble, in contrast to $74.9 \pm 0.25\%$ (= S_2/T_2) solubility in the untreated control samples. This indicates that RNA presence is responsible for maintaining the solubility of approximately 14% of the proteome mass. In a related study, around 10% of proteins in eukaryotic lysates require RNA for maintaining solubility [37]. Protein quantification in these fractions for solubility difference was conducted using the BCA assay (Supplementary Figure S1). Additionally, detectable protein aggregates were observed in the control group (P_2), as seen in the upper panel of Figure 1B. To verify RNA degradation, we analysed the same volumes of samples on an agarose gel that were used for the SDS-PAGE protein analysis, ensuring a consistent comparison between the two methods (Figure 1B, down panel). The results demonstrate that most of the RNAs were degraded in our experiments. The amount of protein aggregates positively correlated with the reaction time (Figure 1C). However, it is necessary to consider the possibility that RNase A itself might induce protein aggregation, beyond the expected effect of RNA depletion. To rule out this possibility, we demonstrated that RNase A remains highly soluble under the same experimental conditions (Supplementary Figure S2A). In line with this observation, we found that inactivating RNase A with an excessive amount of RNase inhibitor effectively abolishes the protein aggregation induced by RNase treatment (Supplementary Figure S2B). This result further confirms that the observed global protein aggregation resulted from RNA depletion caused by RNase treatment. Our experiments, including the quantitative mass spectrometry, measured only the total and soluble fractions of proteins. If proteins non-specifically adhere to the tube walls, our measurements could mistakenly interpret this adherence as protein aggregation. To eliminate this possibility, we verified whether proteins were adhering to the tube walls. The results confirm that no detectable proteins were bound to the tube surfaces (Supplementary Figure S3). Thus, the results in Figure 1 demonstrate that RNA depletion (or degradation) by RNase A causes global protein aggregation in *E. coli* lysates. This depletion approach *in vitro* mimics conventional loss-of-function mutations, making it a simple yet suitable method for assessing the physiological role of endogenous RNAs in maintaining protein solubility *in vivo*. Taken together, our findings indicate that endogenous RNAs play a key role in the maintenance of solubility at the proteome level in *E. coli*.

Identification of RNA-dependent proteins using mass spectrometry analysis

To identify RNA-dependent proteins and quantify their solubility changes, we utilized liquid chromatography-tandem mass spectrometry (LC-MS/MS) combined with a Tandem Mass Tag (TMT) labelling method, as depicted in Figure 2A. The TMT method employs isobaric tags to label proteins from different samples, allowing for their simultaneous quantification and comparison within a single experiment, including monitoring changes in solubility, thus facilitating efficient multiplexed protein analysis [39–41]. Detailed procedures for protein identification and the quantitative analysis of solubility are outlined in the Materials & Methods section. In our mass spectrometry analysis, we applied the same
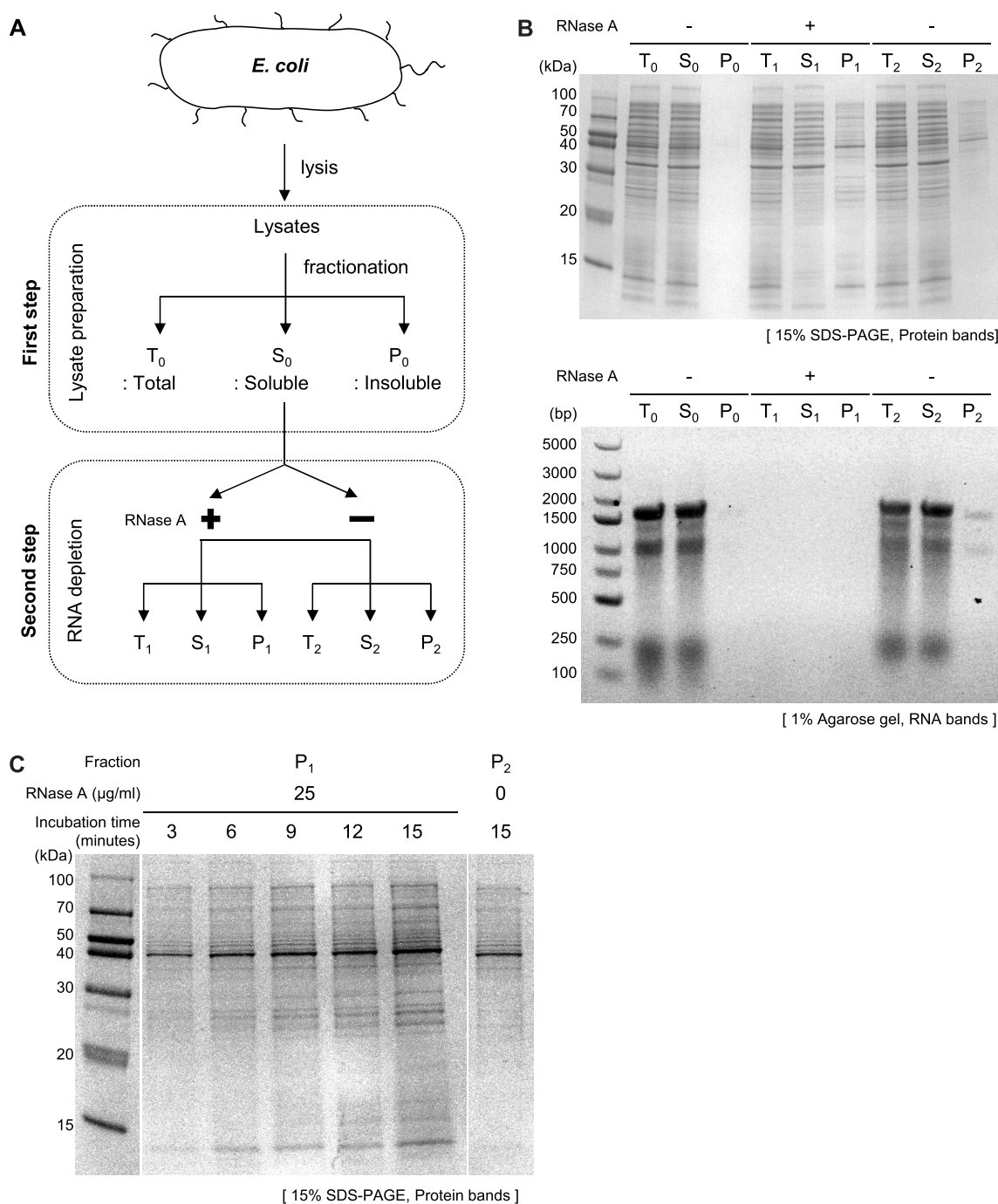
Figure 1. Investigation of RNA's role in maintaining global protein solubility in *E. coli* lysates. (**A**) Diagram illustrating the experimental process, which includes the depletion of RNA from *E. coli* lysates using RNase A and subsequent fractionation into total (T), soluble (S), and insoluble (P) fractions. (**B**) SDS-PAGE analysis comparing protein aggregation in samples treated with RNase A and controls (upper panel) and agarose gel analysis confirming the degradation of RNA following RNase treatment (down panel). RNA degradation was assessed by treating the same volumes of samples used in the SDS-PAGE analysis with proteinase K, with each lane in the agarose gel corresponding to its respective lane in the SDS-PAGE. (**C**) Protein aggregation as a function of reaction time. The insoluble fractions (P_1) from RNase A-treated samples were analysed on SDS-PAGE at various time points (0–15 minute). P_2 represents the insoluble fraction of the control group.

solubility definition as used in the SDS-PAGE analysis, based on the ratio of the soluble to the total protein fraction, thereby allowing for a direct comparison of RNA-dependent protein solubility changes measured by both methods. To assess solubility changes, we calculated the solubility difference ($\Delta S = S_2/T_2 - S_1/T_1$) for each protein by comparing RNase A untreated samples (S_2/T_2) with RNase A treated samples (S_1/T_1).

In our study, as shown in Figure 2A, lysates from two separately cultured cell batches (denoted Sample 1 and Sample 2, respectively) were analysed, representing biological replicates. Each lysate, identically labelled with TMT tags, underwent two rounds of mass spectrometry analysis, acting as technical replicates. This approach provided up to four measurements for each protein. From these measurements, 2,121 proteins were initially detected, among which 1,808 were filtered for further analysis because their solubility values were measured in all four measurements, enabling statistical evaluation. Among these 1,808 proteins, 913 exhibited
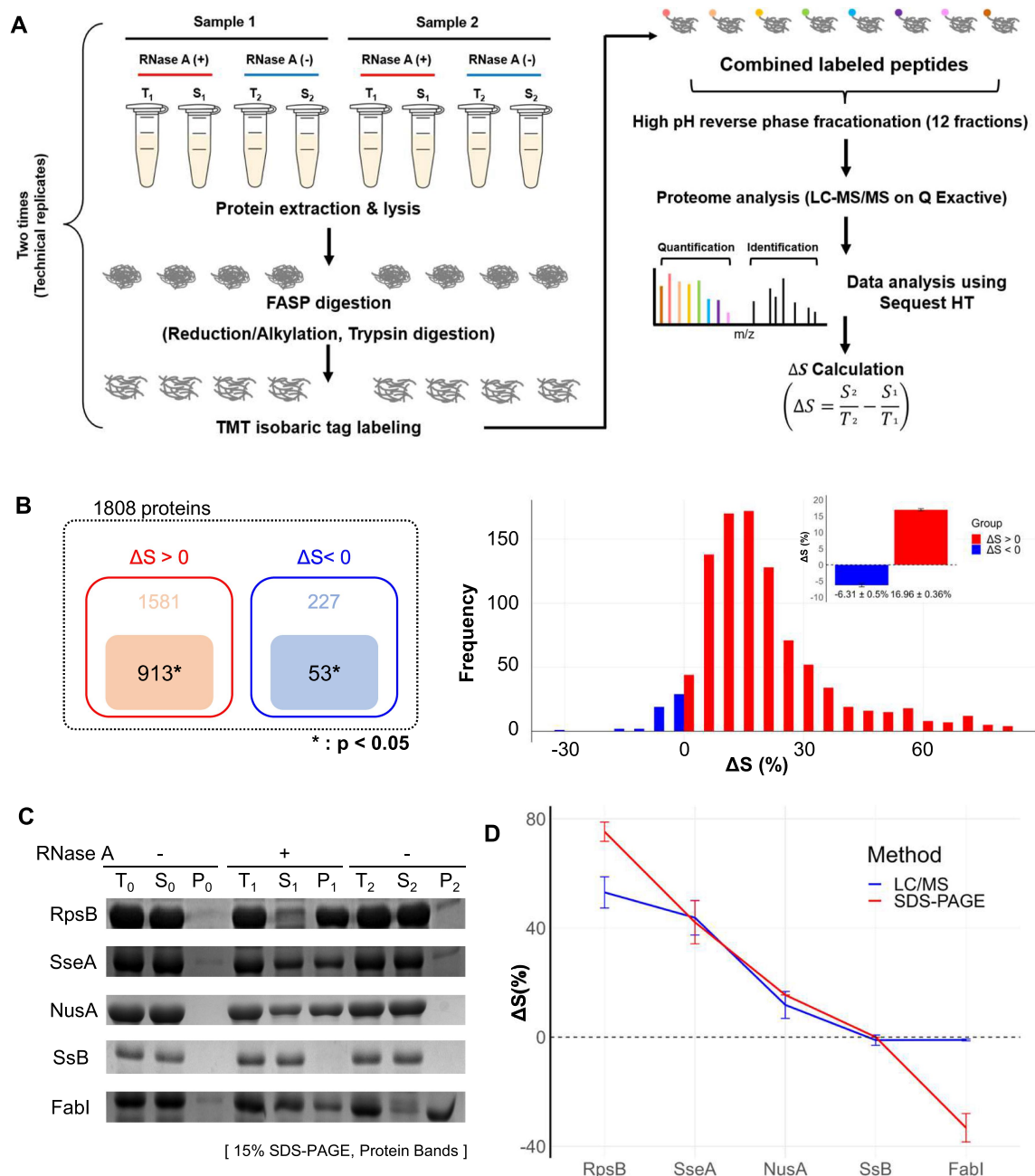
Figure 2. Proteome-wide analysis of proteins responsive to RNA degradation via mass spectrometry. (**A**) Schematic representation of the tandem LC-MS/MS procedure utilized in combination with TMT for quantitative analysis of protein solubility under conditions with and without exogenous RNase treatment. For further details, refer to the materials and methods section. The solubility difference (ΔS) for each protein is defined as the solubility in the control group (S_2/T_2) minus the solubility in the RNase-treated group (S_1/T_1). 'Sample 1' and 'Sample 2' refer to lysates from two distinct cell cultures, underscoring the use of biological replicates in the study. (**B**) The number of identified proteins by mass spectrometry (left panel) and their frequency distribution as a function of ΔS (right panel). A total of 913 proteins with $\Delta S>0$ and 53 proteins with $\Delta S<0$ were identified, all with statistical significance ($p<0.05$). Of the proteins analysed, 913 showed $\Delta S>0$ and 53 $\Delta S<0$, each group identified with statistical significance ($p<0.05$). The histogram of ΔS distribution (bin size: 5%) indicates an average ΔS of 16.96% (SE ±0.36%) for proteins with $\Delta S>0$ and−6.31% (SE ±0.5%) for proteins with $\Delta S<0$. SE is used throughout this paper unless otherwise mentioned. (**C**) SDS-PAGE analysis of ΔS for individually overexpressed proteins RpsB, SseA, NusA, SsB ($\Delta S>0$), and FabI ($\Delta S<0$). The ΔS values obtained for each protein were used to validate the corresponding ΔS values determined by mass spectrometry in *E. coli* lysates. (**D**) Comparative analysis of ΔS between mass spectrometry (red square) and SDS-PAGE (blue circle) analyses. This comparison evaluates the ΔS values in overexpressed proteins from part (in Figure 2C) against those obtained by mass spectrometry in *E. coli* lysates. The ΔS values in overexpressed proteins and mass spectrometry were obtained from three and four independent experiments, respectively.

a significant positive ΔS value, suggesting RNA-dependent solubility, while 53 displayed a significant negative ΔS value ($p<0.05$), indicative of RNA-induced aggregation (Figure 2B, left panel). A histogram plot of frequency versus ΔS revealed mean ΔS values of 16.96% (±0.36% standard error, SE) and −6.31% (±0.5% SE) for these groups, respectively (Figure 2B, right panel). Notably, ΔS values near zero, despite their statistical significance, may be less reliable. All data on the 913 proteins ($\Delta S>0$) and 53 proteins ($\Delta S<0$) used in this study, along with comprehensive information on the entire *E. coli* proteome (based on the November 2017 database, valid as of December 2023), are included in Supplementary Table S1.

To validate the accuracy of the ΔS measurements obtained through our mass spectrometry analysis, we compared these results with the solubility differences of the test proteins, which were directly measured using SDS-PAGE analysis. For this, we individually overexpressed several proteins, including RpsB, SseA, NusA, SsB (where ΔS > 0), and FabI (where ΔS < 0), in *E. coli*. We selected these proteins based on preliminary screening by two-dimensional difference gel electrophoresis (2D-DIGE) (Supplementary Figure S4). However, given the limitations in quantifying protein solubility using 2D-DIGE, we here employed LC-MS/MS coupled with TMT labelling for a more precise, quantitative measurement of solubility difference in a single batch experiment. The solubility of the over-expressed test proteins in *E. coli* lysates, both with and without RNase treatment, was assessed using SDS-PAGE analysis (Figure 2C). The RNase treatment of *E. coli* lysates containing the individually overexpressed proteins resulted in varying degrees of change in protein solubility, as assessed through SDS-PAGE analysis. The observed solubility differences of overexpressed proteins were 75%, 42%, 15%, 0%, and −33% for RpsB, SseA, NusA, SsB, and FabI, respectively. In comparison, the corresponding solubility difference (ΔS) values measured via mass spectrometry analysis of native *E. coli* lysates used for selecting 1808 proteins were 53.1%, 42.1%, 12.0%, 2.1%, and −0.5%, respectively. We observed a strong correlation between the ΔS values derived from mass spectrometry and the solubility differences observed in SDS-PAGE analysis for individual proteins (Figure 2D). These results support the accuracy of our quantitative mass spectrometry analysis.

Characterization of proteins that exhibit RNA-dependent solubility maintenance

We then analysed the proteome-wide characteristics of the identified RNA-dependent proteins using various criteria, including molecular weight, isoelectric point (pI), intrinsic disordered score (IDR score), protein-protein interaction (PPI) score, and Gene Ontology (GO) analysis.

In Figure 3, we present the distribution of three distinct groups of *E. coli* proteins: 913 proteins with ΔS > 0, 53 proteins with ΔS < 0, and the entire proteome. Each group is categorized according to their theoretical pI and molecular weight. For clarity, this figure employs a dual representation approach. Firstly, histograms display the absolute 'count' of proteins within specified pI and molecular weight ranges, with defined bin sizes. Secondly, alongside these histograms, a relative scaling or 'density' graph is used. This approach visualizes the relative distribution of proteins within the same group, showing protein density as a proportion of the total number in each group.

The average pI values for 913, 53, and a whole set of *E. coli* proteins are 6.64, 6.29, and 7.25, respectively. The pI density graphs appear to indicate an enrichment of acidic proteins within the RNA-dependent subset of *E. coli* proteins, based on our analysis of 966 proteins. However, given this dataset represents only a part of the entire proteome that includes both identified and yet-to-be-identified RNA-dependent proteins, it's important to be careful in extending this observation

to the whole proteome. The histogram of 913 proteins with ΔS > 0 in Figure 3 shows that there are many acidic proteins. This result is unexpected because one might assume that RNA-dependent proteins would have a higher average pI, given that RNAs are polyanionic macromolecules. However, our findings challenge this assumption, suggesting a complex interaction between proteins and RNA that is not solely dependent on charge characteristics. Acidic proteins are also prevalent in human neuronal cell lysates, which are listed in Supplementary Table S2, a pattern similar to what we observed in our *E. coli* protein analysis, as further illustrated in Supplementary Figure S5. As for molecular weight, the histogram in the right section of Figure 3 indicates that proteins with ΔS > 0 cover a broad range of molecular weights, aligning with the global protein aggregation observed in the SDS-PAGE analysis shown in Figure 1B. The average molecular weights for 913, 53, and the whole set of *E. coli* proteins are 40.75, 36.90, and 34.47 kDa, respectively. However, the significance of molecular weight in our findings appears less pronounced, suggesting that the size of proteins may not be as crucial in determining their RNA-dependent solubility.

We conducted a linear regression analysis to explore the relationship between the pI value and ΔS in 913 proteins (ΔS > 0) and 53 proteins (ΔS < 0), as shown in Figure 4A. Our findings revealed a significant positive correlation in basic proteins (pI > 7.5) with ΔS > 0, exhibiting a slope of 11.66 (p-value = 1.66E−27, R^2 = 0.397), indicating that higher pI values are associated with ΔS in these proteins. This suggests that basic proteins might interact with RNAs through ionic interactions, which are influenced by pI values. In the field of proteomic data analysis, R^2 value of 0.397, as seen in our study, is relatively substantial. Given the inherent complexity and variability in biological systems, especially in proteomic studies, R^2 values tend to be lower compared to other types of datasets. In contrast, acidic proteins (pI < 7.5) with ΔS > 0, displayed a slope of −0.60 (p-value = 0.254, R^2 = 0.02), suggesting a weaker and less significant trend in the relationship between pI and ΔS for these proteins. The observed trend could possibly be attributed to a predominance of non-ionic interactions between RNAs and proteins. In line with this understanding, many RNA-protein complexes show non-ionic interactions such as hydrogen bonds, van der Waals forces, and hydrophobic interactions [42]. For proteins with ΔS < 0, both basic and acidic proteins displayed a lack of significant difference in their slopes, with basic proteins showing a slope of −3.03 (p-value = 0.162, R^2 = 0.350) and acidic proteins a slope of −2.00 (p-value = 0.021, R^2 = 0.116). The significant slope differences observed between acidic and basic proteins with ΔS > 0 indicate that there might be the distinct RNA-binding modes between these proteins.

To gain a better understanding of the different RNA-protein binding modes, we analysed the electrostatic potentials of selected ribosomal proteins in the ribosome structure (PDB ID: 8B7Y), focusing on both acidic and basic ribosomal proteins. In Figure 4B, we highlighted specific examples: S6 and S2 as acidic ribosomal proteins, and S13, S14, and L20 as basic ribosomal proteins. These proteins were chosen based on their ΔS values, with each group arranged in
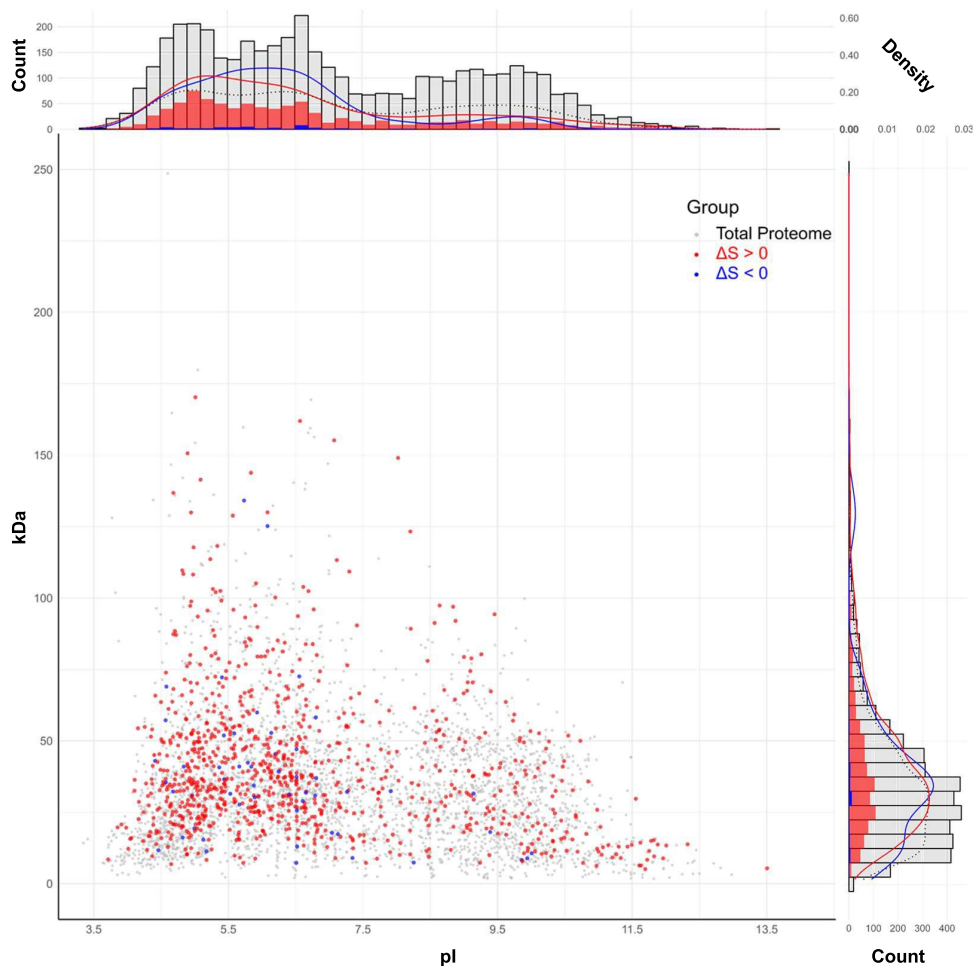
Figure 3. Proteome-wide characterization of RNA-dependent proteins in *E. coli*. The scatter plot displays 913 proteins with ΔS > 0 (red), 53 proteins with ΔS < 0 (blue), and the whole *E. coli* proteome of 4302 (grey) in terms of pI and molecular weight. Note that the red and blue spots are superimposed on the grey spots. Histograms show the absolute count of proteins across different pI values (bin size: 0.2) and molecular weight (bin size: 5 kDa) ranges. Alongside each histogram, density graphs represent the relative frequency distribution of proteins within each bin. These graphs are adjusted so that the total area under each curve sums to 1, allowing for comparison of distribution patterns across groups, irrespective of the total protein number of each group.

descending order. The solubility change and isoelectric point information for these proteins are summarized in Figure 4B, and the acidic and basic ribosomal proteins exhibited distinct binding patterns. The acidic proteins were mainly exposed on the surface of the ribosome, with limited electrostatic interactions at the RNA-protein contact interfaces. In contrast, the basic ribosomal proteins were largely enveloped by rRNAs or located in deep valleys of rRNA structures, with their basic residues playing significant roles in the RNA-protein interfaces. These observed binding modes align with our interpretation derived from the results presented in Figure 4A.

IDRs are known to be enriched in RBPs [33–36]. Therefore, we explored the correlation between IDR scores and pI values (or ΔS) across proteins. To enhance precision and minimize potential bias, we utilized multiple IDR predictors, as outlined in the Materials & Methods section. In our analysis shown in Figure 5A, we focused on two groups: proteins with measured solubility changes (ΔS > 0) and the entire *E. coli* proteome ('Total'). For proteins with ΔS > 0, basic proteins showed a significant positive correlation between pI and IDR scores, with a slope of 6.64 (p-value =

2.91E–22, R^2 = 0.332), while acidic proteins exhibited a less correlation, with a slope of −0.23 (p-value = 0.485, R^2 = 0.001). In the total proteome, basic proteins displayed a correlation with a slope of 3.96 (p-value = 3.22E–44, R^2 = 0.111), while acidic proteins showed a different trend with a slope of −1.08 (p-value = 3.01E–11, R^2 = 0.017). Intriguingly, proteins with ΔS > 0 displayed a correlation between pI and IDR scores (Figure 5A) that resembles the pattern observed between pI and ΔS (Figure 4A). These parallel trends suggest a potential linear correlation between IDR scores and ΔS, a notion supported by the correlation shown in Figure 5B. Thus, we examined the correlation between IDR scores and ΔS. Proteins with ΔS > 0 exhibited a significant positive correlation, with a slope of 0.22 (p-value = 4.51E–29, R^2 = 0.129), while proteins with ΔS < 0 showed a non-significant correlation, with a slope of −0.35 (p-value = 0.174, R^2 = 0.036). IDRs play a crucial role in the formation of multimolecular assemblies and protein-cellular macromolecule networks [29]. Proteins with high IDR scores often serve as functional and structural hubs in PPIs, as evidenced by their high scores in the Search Tool for the Retrieval of Interacting Genes/

A

B
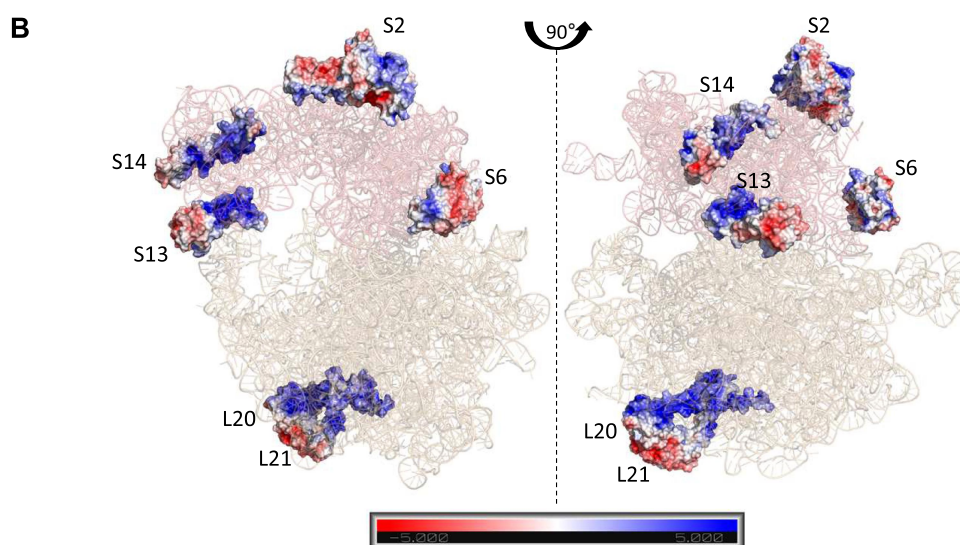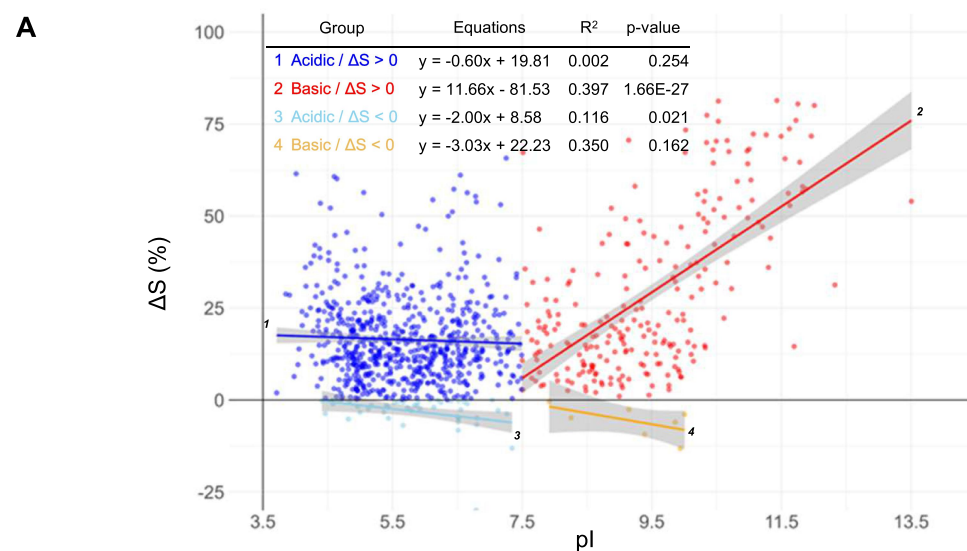
Accession	Entry name	SYMBOL	Protein	ΔS (%)	pI	kDa	IDR (%)	p-value
P02358	RS6_ECOLI	rpsF	S6	60.15	4.64	15.70	31.85	0.037
P0A7V0	RS2_ECOLI	rpsB	S2	53.08	7.17	26.61	14.18	0.000
P0A7S9	RS13_ECOLI	rpsM	S13	81.40	11.43	12.97	30.79	0.000
P0AG48	RL21_ECOLI	rplU	L21	81.24	10.53	11.56	23.62	0.000
P0AG59	RS14_ECOLI	rpsN	S14	80.47	11.75	11.45	44.72	0.000
P0A7L3	RL20_ECOLI	rplT	L20	80.06	12.01	13.37	13.84	0.017

Figure 4. Analysis of the correlation between ΔS and pI across different protein groups. (**A**) A scatter plot comparing ΔS versus pI for the 913 proteins (ΔS > 0) and 53 proteins (ΔS < 0), further divided into acidic (pI < 7.5) and basic (pI > 7.5) groups. The linear regression equation, p-value, and R^2 (coefficient of determination) values for each fitting are provided within the figure data. The plot includes SE to depict the confidence range around each data point. (**B**) Illustration of distinct rRNA-binding modes in acidic (S6, S2) and basic (S13, L21, S14, L20) ribosomal proteins within the ribosome structure, reflecting the different correlations observed in Figure 4A. Positively and negatively charged residues are marked in blue and red, respectively. The corresponding ΔS values and pI for these proteins are provided in the accompanying table.
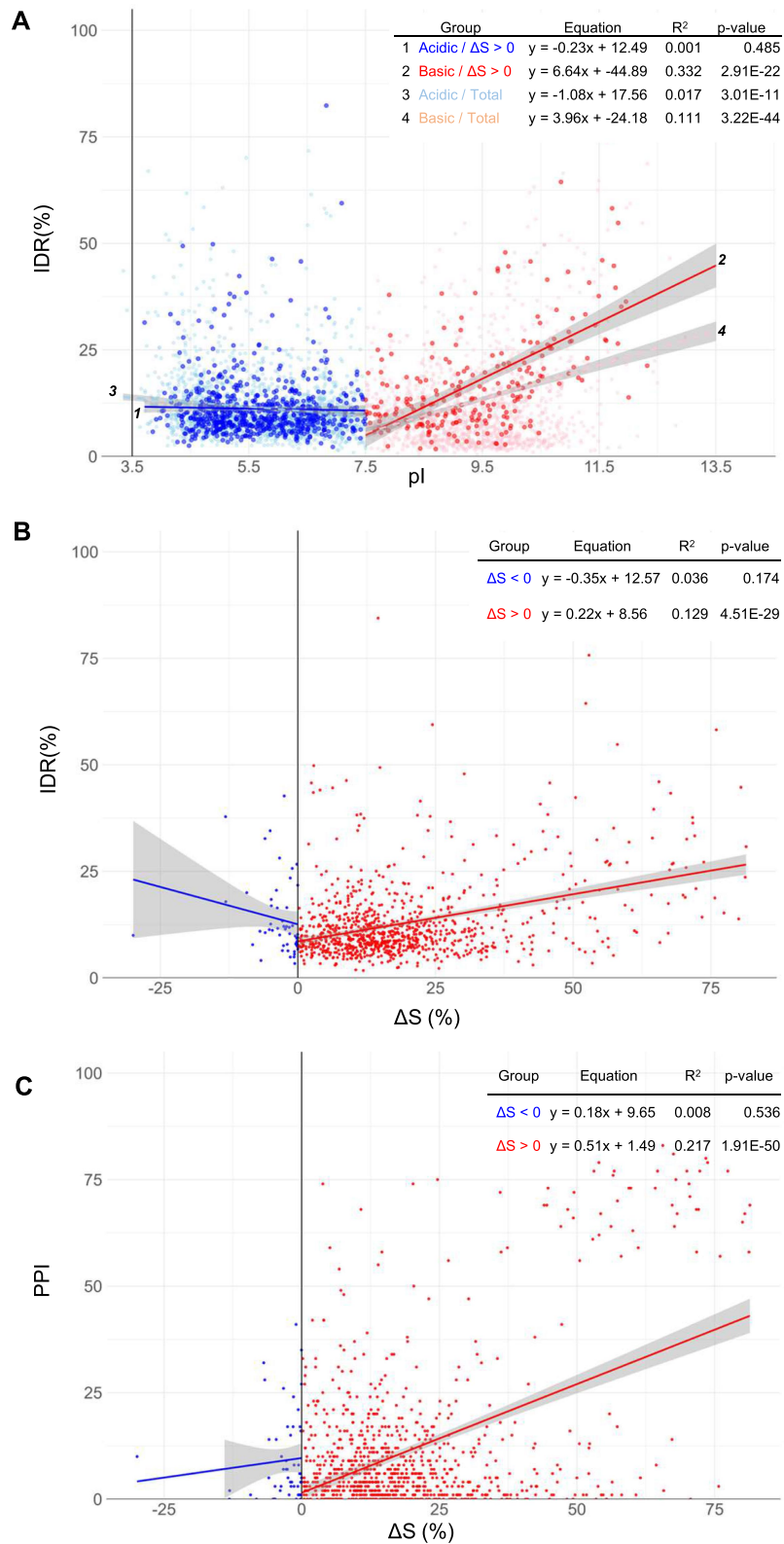
Figure 5. Analysis of the interplay between intrinsic disorder, pI, ΔS, and PPI. (**A**) Plot between IDRs and pI for acidic and basic proteins of proteins (ΔS > 0) and total proteome. 4302 proteins of the total proteome are depicted with grey spots and a black dotted line as a background reference. (**B**) Plot between IDR against protein solubility change (ΔS) for proteins with ΔS > 0 and ΔS < 0. (**C**) A correlation plot between PPI, as determined by a STRING score with a threshold of 0.900, and ΔS for proteins with ΔS > 0 and ΔS < 0. The linear regression equation, p-value, and R^2 (coefficient of determination) values for each fitting are provided within the figure data.

Proteins (STRING), which measures PPI frequency [43,44]. These hub proteins, with their elevated IDR scores, may be more susceptible to aggregation. This led us to investigate the correlation between PPI and ΔS. The results showed a significant positive correlation between PPI frequency and ΔS for proteins with $\Delta S > 0$, characterized by a slope of 0.51 (p-value = 1.91E–50, $R^2 = 0.217$), indicating a significant enrichment of functional and structural hub proteins in RNA associations. Conversely, for proteins with $\Delta S < 0$, the correlation between PPI frequency and ΔS was not significant, with a slope of 0.18 (p-value = 0.536, $R^2 = 0.008$).

Continuing our investigation, we focused on the biological functions of RNA-dependent proteins by conducting GO analysis on the 913 proteins with $\Delta S > 0$, as illustrated in Figure 6. To reflect the distinct RNA-protein binding properties, separate GO analyses were performed for 677 acidic and 236 basic proteins, using the respective groups from the E. coli whole proteome as backgrounds. Interestingly, the acidic group's GO analysis revealed only 3 out of the top 20 enriched terms related to RNA, namely 'translation', 'tRNA metabolic process', and 'ncRNA metabolic process'. This finding contrasts with our earlier observation in Figure 3, which showed a higher proportion of acidic proteins. This disparity could be attributed to the underrepresentation of noncanonical RNA-binding properties in acidic proteins within the GO database.

On the other hand, the basic proteins demonstrated a significant enrichment in RNA-related terms, including 'ribonucleoprotein complex', 'cytosolic ribosome', 'translation', and 'RNA-binding', with 12 of the top 20 terms associated with RNA. This dominance of ribosomal terms led us to perform further GO analysis excluding ribosomal proteins to gain a deeper insight into RNA-related functions. After removing 50 ribosomal proteins, the subsequent analysis showed minimal changes for acidic proteins but revealed only four RNA-related terms among the top 20 for basic proteins. Remarkably, non-RNA-related terms such as 'cell wall synthesis', 'outer membrane-bounded periplasmic space', and 'cell envelope' were also prominent, reflecting reports of tRNAs being involved in cell wall synthesis [45–47].

To broaden our understanding, we applied the same GO analysis approach to RNA-dependent proteins in human neuronal cell lysates (Supplementary Figure S6). For acidic proteins from the sources, only 6 of the top 20 terms were RNA-related. In basic proteins, after excluding 88 ribosomal proteins, 12 out of the top 20 terms remained RNA-related, a higher number than in the E. coli analysis. This suggests a more extensive range of RNA-related functions in basic proteins beyond rRNAs. Intriguingly, terms like 'The citric acid (TCA) cycle' and 'Mitochondrion inner membrane' were also enriched, paralleling the E. coli findings, and indicating a potential conserved pattern of RNA-related functions in membranes across different organisms. Considering that membrane proteins constitute over 30% of the proteome and given the global effect of RNA on
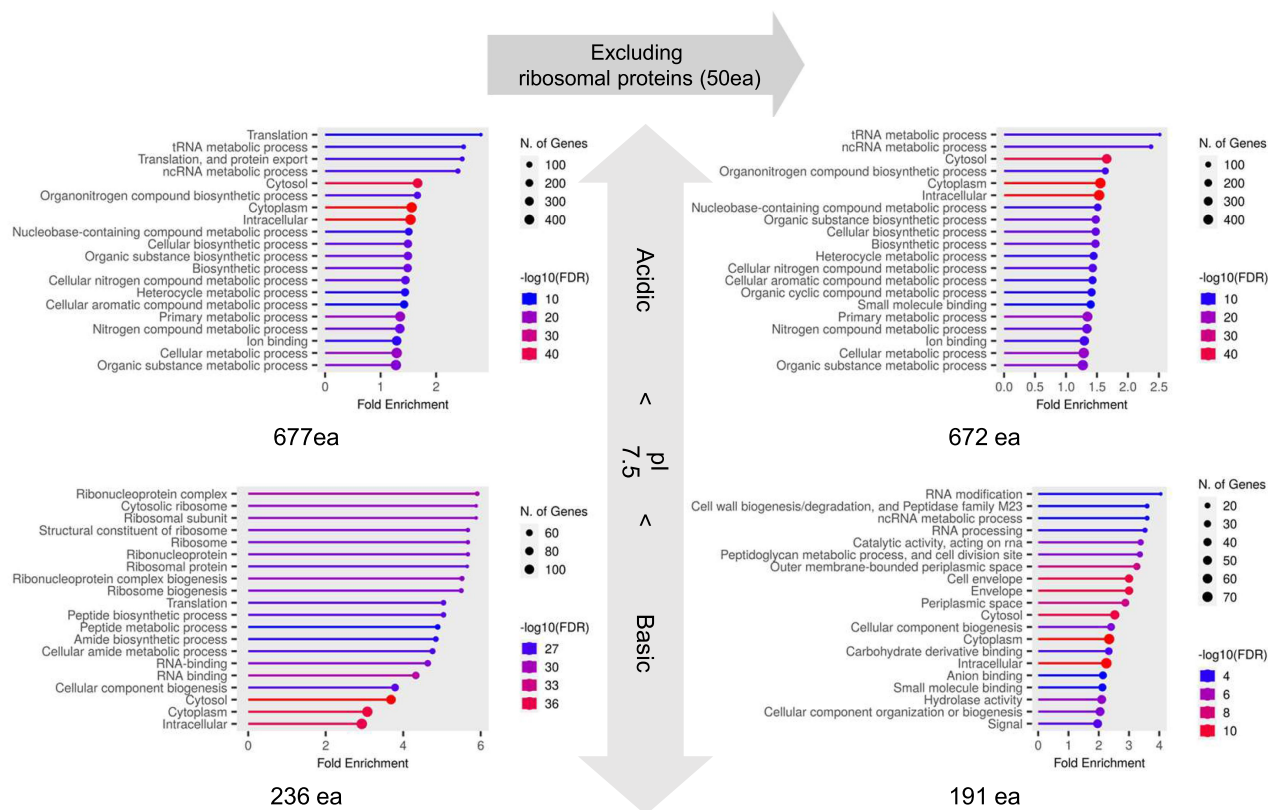
Figure 6. GO analysis between acidic and basic RNA-dependent proteins. This analysis categorizes 913 proteins with $\Delta S > 0$ into 677 acidic and 236 basic proteins, each compared against the background of their respective acidic and basic proteins in the E. coli proteome. Further analysis was conducted on subsets (672 acidic and 191 basic proteins) after excluding 50 ribosomal proteins. The top 20 enriched GO terms were displayed for each group.

proteome solubility maintenance, it is tempting to speculate about RNA's potential role in the functions or biogenesis of membrane proteins.

We specifically investigate the solubility of tRNA synthetases, the primary class of proteins that interact with tRNAs. Around 15% of the total RNA population consists of tRNAs, which interact dynamically with their cognate tRNA synthetases with relatively low affinity. This is in contrast to the tight complexation between ribosomal proteins and rRNAs. This raises the question of whether these dynamic interactions can also affect protein solubility maintenance. As shown in Table 1, the solubility of 16 out of 20 total tRNA synthetases are decreased by the RNA depletion; the solubility differences range from 31% to 0.19%, with the average solubility difference of 10.11%. These findings provide support for the notion that the dynamic interactions with their cognate RNAs can influence protein solubility maintenance. In line with this, the *in vitro* refolding of *E. coli* lysyl-tRNA synthetase was found to rely heavily on its cognate tRNA [13]. The tRNA synthetases are acidic proteins; their average pI value is 5.41. It is plausible that the binding modes between tRNA synthetases and their cognate tRNAs share similarities with those of acidic ribosomal proteins, such as S6 and S2, with rRNAs.

RNA-dependent solubility maintenance of molecular chaperones

Molecular chaperones interact with their clients via hydrophobic interactions [7,10,48,49]. The exposed hydrophobic surfaces of molecular chaperones may make the chaperone proteins prone to aggregation, necessitating the presence of upstream aggregation gatekeepers, such as RNAs. Trigger factor (TF) is associated with ribosomes or RNA-protein complexes [50]. Molecular chaperones, including DnaK and GroEL, associate with messenger RNAs [51,52]. These findings suggest that RNAs may play a role in maintaining the solubility of protein-based molecular chaperones.

To explore this possibility, we employed the same RNA depletion method as used in Figure 1A. As shown in the left part of Figure 7A, the solubility of molecular chaperones TF, DnaK, DnaJ, GroEL, and GroES was investigated through western blot analysis using antibodies specific to each chaperone. Remarkably, the treatment with RNase A caused considerable aggregation of GroES, TF, and DnaJ. In contrast, the solubility of DnaK and GroEL was relatively unaffected by RNA depletion. The solubility of molecular chaperones measured by western blotting (Figure 7A, left panel) appears to correlate with the solubility determined through mass spectrometry analysis (Figure 7A, right panel). In the case of TF, a notable discrepancy was observed in its solubility results when compared between these two methods. Despite not being statistically significant in our mass spectrometry analysis, TF was included in the study due to its critical role as a chaperone. TF, DnaK-DnaJ-GrpE, and GroEL/ES systems are the major chaperone systems in *E. coli* [7,8,53]. Our findings indicate that RNAs can play a pivotal role in modulating the solubility and thus possibly functionality of these systems.

The folding of approximately 52 aggregation-prone proteins has been reported to be highly dependent on GroEL/ES [54]. The GroEL/ES system is also believed to be important for maintaining the solubility of its clients in terms of proteostasis [10]. RNAs may also contribute to the solubility of its clients. Thus, we investigated the identity of the 52 GroEL/ES client proteins within the proteome of $\Delta S > 0$, as shown in Figure 7B. Our study identified twenty-seven clients listed in Supplementary Table S3, which were found to be dependent on RNAs. In conclusion, our findings suggest that RNAs play a role in maintaining the solubility of both molecular chaperones and their clients, as depicted in Figure 7C.

Table 1. Classification and ranking of tRNA synthetase proteins. This table displays tRNA synthetase proteins identified from our LC/MS results, sorted into classes 1 and 2 and ranked by decreasing ΔS values. Statistical significance is indicated by *p* values (*p < 0.05, **p < 0.01, ***p < 0.001).

Accession	Protein Name	Class	kDa	pI	ΔS (%)	IDR (%)	p-value	Significance
P04805	GluRS	1	53.82	5.80	31.08	11.64	0	***
P0A8N3	LysRS	2	57.47	4.88	18.87	12.15	0	***
P00960	GlyRS, α subunit	2	34.77	4.63	18.76	8.69	0	***
P0A8M3	ThrRS	2	74.01	6.12	16.75	8.72	0.001	**
P00961	GlyRS, β subunit	2	76.68	5.08	16.71	10.64	0	***
P60906	HisRS	2	46.90	5.72	16.27	7.27	0.01	*
P11875	ArgRS	1	64.68	5.17	15.19	7.77	0.004	**
P08312	PheRS, α subunit	2	36.83	6.14	13.24	12.18	0.036	*
P0A8N5	LysRS, heat inducible	2	57.70	4.87	12.69	14.55	0.012	*
P00957	AlaRS	2	95.90	5.58	12.25	11.49	0.017	*
P21888	CysRS	1	52.20	5.25	12.01	11.50	0	***
P0AGJ9	TyrRS	1	47.40	5.55	10.47	10.14	0	***
P07813	LeuRS	1	97.23	4.93	9.32	9.26	0	***
P00962	GluRS	1	63.35	6.22	6.84	11.70	0.015	*
P07118	ValRS	1	108.19	4.98	6.38	11.41	0	***
P00954	TrypRS	1	37.44	6.72	5.55	12.28	0.314	n.s
P0A8L1	SerRS	2	48.41	5.18	4.62	18.57	0.165	n.s
P0A8M0	AspRS	2	52.44	4.95	3.69	7.37	0.02	*
P00956	IleRS	1	104.17	5.87	3.33	6.27	0.164	n.s
P21889	AspRS	2	65.91	5.36	2.66	9.58	0	***
P00959	MetRS	1	76.12	5.65	0.19	7.63	0	***
P07395	PheRS, β subunit	2	87.38	4.95	−0.47	7.38	0.215	n.s
P16659	ProRS	2	63.69	4.87	−3.76	8.45	0.226	n.s
	Average		65.33	5.41	10.11	10.29		

Figure 7. RNA's effects on the solubility of molecular chaperones. (**A**) Western blot results (left panel) showing RNA's influence on chaperones' solubility in response to RNA depletion and the corresponding ΔS values (right panel) of these chaperones from our mass spectrometry analysis of *E. coli* lysates. Statistical significance is indicated as follows: * for $p < 0.05$ and ** for $p < 0.01$. Note: TF had a low significance level in our mass spectrometry analysis. (**B**) Analysis of GroEL/ES client proteins impacted by RNA on solubility, categorized by their pI and molecular weight. Proteins are marked in red to indicate their inclusion in the group of 913 proteins with ΔS > 0, while those marked in black represent proteins not included in this group, referred to as 'Others'. (**C**) A schematic illustrates how RNAs maintain the solubility of molecular chaperones and their client proteins by inhibiting protein aggregation.

Discussion

Maintaining protein solubility against aggregation is a fundamental issue in productive protein folding pathway, molecular chaperones, proteostasis, and proteinopathies [1–10,55,56]. Despite notable advancements made over several decades, the issue of maintaining protein solubility remains largely unresolved. Our study reveals that depleting endogenous RNA from *E. coli* lysates leads to widespread protein aggregation, including molecular chaperones (Figures 1B and 7A). By utilizing TMT labelling, we have quantitatively measured the influence of RNA on proteome solubility. Our TMT-based measurements of protein solubility changes align well with the direct solubility measurements obtained from SDS-PAGE (Figure 2D). Consistent with the global aggregation detected through SDS-PAGE analysis (Figure 1B), quantitative mass spectrometry analysis revealed that RNAs impact around 21% of *E. coli* proteome, exhibiting diverse sizes and pI (Figure 3). Solubility is influenced in both acidic and basic proteins, with a more pronounced effect observed in the acidic proteins (Figure 3). Proteome-wide characterization of RNA-dependent proteins shows that their solubility difference is correlated with pI, IDR score, and PPI (Figures 3, 4 and 5). It means that functional and structural hub proteins are

usually associated with RNAs. Even the solubility of molecular chaperones is dependent on RNAs (Figure 7A). Taken together, RNAs play a key role in maintaining proteome solubility, which is crucial for proteins to function effectively, and further extending the chaperna function [13,16,19].

The analysis of RNA's role in proteome solubility in our *E. coli* lysate study is similar to that revealed in human neuronal cell lysates. Approximately 14% of the proteome in *E. coli* lysates and 10% in human neuronal cell lysates, which include a diverse range of proteins, depend on RNA for solubility maintenance in terms of mass. Additionally, a substantial majority of these RNA-dependent proteins are acidic proteins (Figure 3 and Supplementary Figure S5). Interestingly, most of these RNA-dependent proteins are classified as non-canonical RNA-binding proteins, diverging from typical canonical ones. This observation aligns with our GO analysis, where the majority of enriched terms for acidic proteins do not relate to RNA. These findings, coupled with previous studies, suggest that RNA-dependent proteome solubility is a conserved phenomenon across different species, including both prokaryotic and eukaryotic organisms. Considering the evolution of proteins alongside RNA and their diverse interactions within the cell, the proteome-wide

effect of RNA on protein solubility, though unexpected from a traditional protein-centric view, is not entirely surprising.

To comprehend the role of RNAs in maintaining proteome solubility and preventing aggregation, as evidenced in our study, it is necessary to consider the potential fundamental causes. One key aspect is that RNAs physically interact with proteomes through diverse connection types, as illustrated in Figure 8. For example, RNAs form stable or transient RNA-protein complexes through native interactions. As native ligands, RNAs can profoundly prevent the complexed proteins from aggregation [18,31,32,57], which is in line with the changes in protein solubility observed when interacting with rRNAs and tRNAs (Figure 4B and Table 1). Particularly for RNA-binding coupled protein folding as in IDPs and IDRs (Figure 5), their folding, stability, and solubility can be significantly affected by their cognate RNAs. RNAs also serve as macromolecular crowders [58], and transient and non-specific interactions exist between proteins and crowders [59]. In addition, all newly synthesized polypeptides are tightly connected in *cis* to megadalton-sized ribosomes with polyanionic surfaces of rRNAs. RNAs can generally exhibit the intrinsic chaperone ability to inhibit the aggregation of their physically – whether directly or indirectly – connected polypeptides regardless of the connection type, in part due to the intermolecular repulsions resulting from their large excluded volume and surface charges [13,16,17]. RNAs have been shown to inhibit the aggregation of proteins when fused with an RNA-binding module in *cis* [13]. Consistent with our present results (Table 1), tRNAs display a non-canonical

function of chaperones for the folding of cognate tRNA synthetase [13] and assembly of target proteins in an engineered system [14,15], reminiscent of protein-based molecular chaperones [9]. Like proteins, RNAs can adopt three-dimensional structures, exemplified in rRNA and tRNA cloverleaf structures, enabling them to recognize various proteins. RNAs can associate with proteins through diverse non-covalent interactions, including hydrogen bonds, van der Waals interactions, and hydrophobic interactions [42]. Hydrophobic interactions between protein – RNA interfaces, albeit less appreciated, were proposed to be involved in up to 50% of these interactions, depending on the interacting RNA-binding proteins [60]. Thus, RNAs can have the ability to directly interact with the exposed hydrophobic surfaces of aggregation-prone proteins in the same way as protein-based molecular chaperones [16,20,21]. In contrast to proteins, however, polyanionic RNAs can maintain their solubility, even in a denatured state. These considerations, along with our results, illuminate the plausible underlying reasons for the essential role of RNAs in maintaining proteome solubility.

While the main focus of our study is on the role of RNAs in preserving protein solubility, there were instances where RNAs had a negative impact on the protein solubility ($\Delta S < 0$) (Figure 2B,C). Our sensitive quantitative mass spectrometry analysis allowed us to identify this small subset of proteins. Initially, we considered that proteins with $\Delta S < 0$ but close to 0, used as reference for calculating ΔS, may fall within the experimental error range. However, the case of FabI, a protein with $\Delta S < 0$, exhibited a solubility trend contrary to those
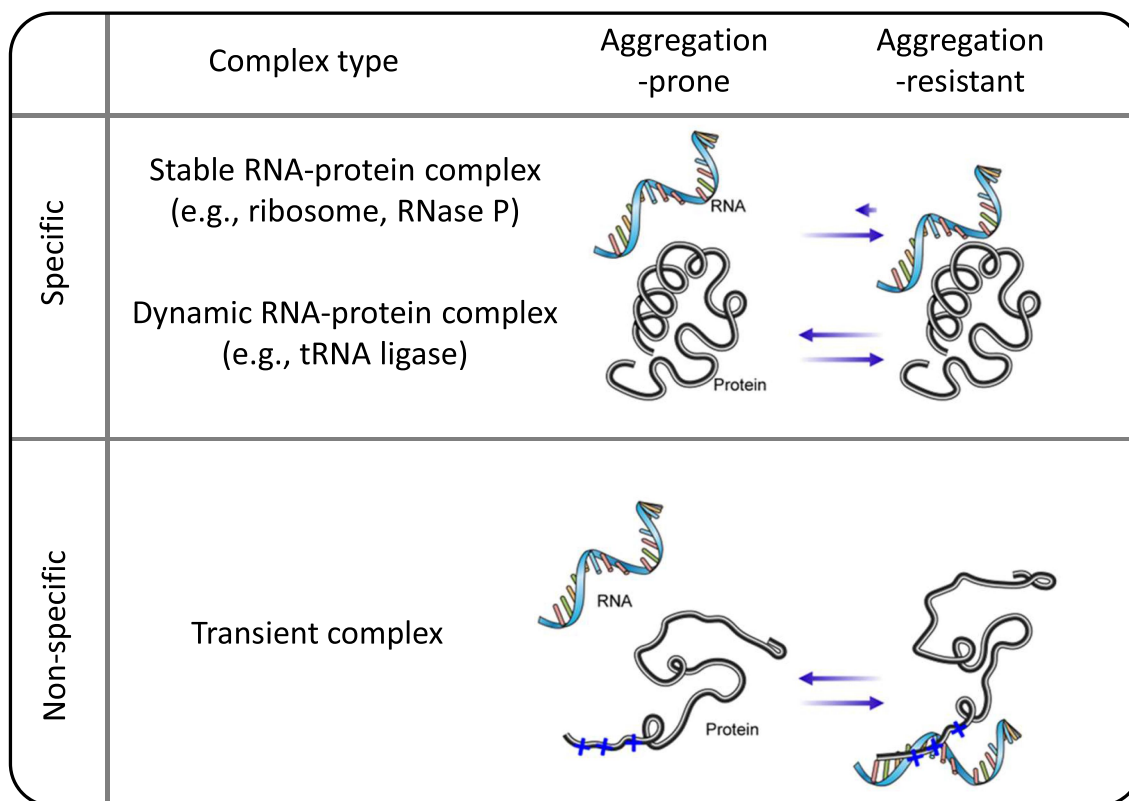

Figure 8. Proteome solubility maintenance through association with RNAs. The diagram illustrates the proteome associated with RNAs. Specific interactions involve the formation of stable and dynamic RNA-protein complexes. Non-specific interactions cover transient associations with RNAs, serving as molecular crowders within the cellular milieu.

proteins with $\Delta S > 0$, suggesting a more complex relationship between RNAs and protein solubility. In line with these findings, RNAs can sometimes promote protein aggregation and modulate it depending on the RNA to protein ratio [61,62]. Moreover, RNAs are important for the formation of membraneless organelles by inducing liquid – liquid phase separation as structural scaffolds for RNA – RNA and RNA – protein associations [63,64]. From a mechanistic standpoint, RNAs could augment protein aggregation via the potential mechanisms such as protein destabilization, an increase in the local protein concentration on the same RNA, and multivalency-mediated bridging between protein molecules. For instance, ribosomes have been shown to thermodynamically destabilize their tethered nascent chains [65], and *in vitro* studies have demonstrated that polyanions can destabilize proteins [66]. Our findings expand the versatility of RNA to modulate protein solubility in a negative manner, although the functional implications require further study.

Our experiments demonstrated that molecular chaperones, such as TF, DnaJ, and GroES, depend heavily on RNA to maintain their solubility (Figure 7). These data address a long-standing question: if molecular chaperones safeguard proteome quality by preventing protein misfolding and aggregation, what mechanism ensures the quality of the chaperones themselves from the outset? Despite traditional belief in molecular chaperones' role in protein folding and preventing aggregation, our data suggest their solubility largely depends on RNAs, implying RNA interactions might influence chaperone functions. Importantly, this dependency suggests that the diverse functions of molecular chaperones may be influenced by their interactions with RNAs. In line with our observations, molecular chaperones were known to exhibit the RNA-binding ability [51,67–70]. Moreover, RNAs were reported to assist the refolding of molecular chaperones *in vitro* [71,72]. However, an intriguing question arises from such dependency: Could the observed effect on proteome solubility upon RNA depletion be an indirect consequence, stemming from a disruption of the molecular chaperone systems due to their reliance on RNA? If this were the case, we would expect to observe the dependency of the whole 52 clients of GroEL/ES; instead, the solubility of only 27 out of 52 clients is affected by the RNA depletion in our experiments (Figure 7B). Moreover, the solubility of most of the ribosomal proteins are affected by the RNA depletion as expected. Considering aforementioned descriptions based on our results, it lends support to the notion that the widespread protein aggregation observed upon RNA depletion (Figure 1) is directly linked to the impaired ability of proteins to associate with RNAs. Future research exploring the collaboration between RNA-based chaperones (chapernas) [19] and traditional protein-based molecular chaperones [7,10] in controlling proteome solubility maintenance will be interesting. The findings guide future research on unravelling the hierarchical relationship of protein-based molecular chaperone and the RNA-based chaperna at molecular and functional level.

We acknowledge certain limitations in our study. While we applied a stringent 1% false discovery rate (FDR) for peptide identification and conducted individual t-tests for each protein's analysis, we did not incorporate multiple hypothesis testing methods like Bonferroni or Benjamini and Hochberg corrections, potentially affecting our interpretation across numerous proteins. Additionally, our experiments included two biological and two technical replicates. Although technical replicates ensure measurement consistency, they might not capture the full scope of biological variability. Although our linear regression analysis provided valuable insights into proteomic tendencies, it may have oversimplified the complexities inherent in proteomic data. Also, our findings highlight the global influence of RNA on proteome solubility, particularly regarding the solubility of several molecular chaperones. However, they do not investigate the specific mechanistic ways of how RNA affects the solubility maintenance-related factors, such as protein folding, stability, and aggregation. This points to the necessity for more detailed research in these areas.

Traditionally, RNAs have been primarily regarded as structural components of RNA-protein complexes and as messengers or adaptors for translation in the paradigm models, such as the central dogma of molecular biology, Anfinsen's thermodynamic hypothesis, and the concept of molecular chaperone. However, proteins have evolved through interaction with RNAs [19,73,74]. RNA-protein interactions play a key role in the coordination of regulatory networks [75]. It is emerging that RNA-binding proteins and RNAs are involved in neurodegenerative diseases [76,77]. Our study highlights the vital role of RNA interactions in maintaining proteome solubility. This opens a broader perspective on the multifaceted influence of RNAs on proteome regulation, extending beyond their traditionally acknowledged roles. These findings invite further exploration in RNA biology, urging us to rethink and expand our understanding of RNA's significance. Embracing this broader viewpoint has the potential to advance our comprehension of critical biological processes and unveil new research avenues.

Materials & Methods

Cell culture and lysis

This study employed competent *E. coli* BL21 (DE3) cells. In addition, we transformed them with plasmids containing RpsB, SseA, NusA, and FabI, respectively. 1 μl of BL21 (DE3) competent cell and transformed cells were inoculated into 3 ml aliquots of LB medium, followed by an overnight incubation. Subsequently, 100 μl of each culture was transferred to fresh LB medium and grown for approximately 2 hours until the cells reached the exponential growth phase. For cells with plasmids, we introduced an inducer to express the proteins for an additional 2 hours, while the competent cells were cultured continuously without inducers. Throughout the culture process, ampicillin, a selective marker, was added to maintain plasmid selection in transformed cells. Following culture, cells were harvested and immediately flash-frozen at −80°C to preserve structural integrity and halt metabolic activity. This step is a crucial step for maintaining the integrity of

cellular components, especially RNAs. Cell lysis was then conducted using B-PER II lysis buffer (Thermo Fisher Scientific), supplemented with DNase I and lysozyme. Following lysis, lysates were centrifuged to separate the supernatants from the cellular debris, preparing them for further experimental analysis.

RNase treatment and fractionation for proteome solubility analysis

RNase A (Invitrogen, EN0531) was added to the supernatants at a concentration of 25 µg/mL to assess the impact of RNA depletion on the solubility of E. coli proteins. RNase A-treated supernatants were incubated for 15 minutes at 37°C. After incubation, the mixtures were centrifuged at 20,800 × g for 10 minutes at 4°C to separate the precipitates. Each fraction was organized and labelled as shown in Figure 1A. To address potential supernatant contamination adhering to the pellets after centrifugation, the pellet fractions P_0, P_1, and P_2 were washed with 100 µl of PBS three times. This washing process effectively dilutes residual supernatant concentration by approximately from 10^{-4} to 10^{-5}-fold from the pellets. Gentle pipetting was employed to preserve pellet integrity during these steps. Following the washes, the pellets were resuspended in the same volume of PBS used in the lysate preparation, excluding the volume used for electrophoresis. This step is to accurately measure the concentration of insoluble proteins in the total lysate. Protein LoBind® tubes (Eppendorf, #022431081) were critical to prevent sample loss and contamination, thereby ensuring accurate solubility assessments. Additional experiments demonstrating the efficacy of this approach are detailed in Supplementary Figure S3.

Protein and RNA analysis via gel electrophoresis

Following the RNase treatment, each fraction was prepared for SDS-PAGE analysis. To facilitate protein solubility analysis, the fractions were combined with 2X SDS-PAGE Loading Buffer (Biosesang, S2002–1) and heated. SDS-PAGE was performed using a 15% polyacrylamide gel. After SDS-PAGE, the polyacrylamide gels were stained with Coomassie Brilliant Blue (CBB), followed by destaining. In parallel, to evaluate the RNA content in each fraction, proteinase K (Thermo Fisher Scientific, EO0491) treatment was used at a concentration of 0.1 mg/ml to remove proteins, allowing for clear RNA visualization on an agarose gel. After a 30-minute incubation at 37°C with RNase A, the RNA in these fractions was visualized by E-Gel™ Agarose Gels with SYBR™ Safe DNA Gel Stain, 1% (Thermo Fisher Scientific, A42100).

Database utilization in proteomic analysis

Initial LC-MS/MS analysis utilized the UniProt E. coli proteome database (November 2017 version), encompassing 4309 entries. This extensive database served as a solid foundation for our comprehensive MS data analysis. Subsequently, for analyses conducted from December 2023 onwards, we transitioned to an updated version of the database, which contains 4302 entries. This refinement was made by eliminating obsolete data, thus ensuring the use of the most current and relevant proteomic information. Supplementary Table S1 provides detailed information on these entries. Notably, despite the constancy in protein amino acid sequences, there is significant variability in methods for calculating pI. Therefore, we opted to apply the EMBOSS method [78] for pI calculation in this study. This approach aligns with the need for consistent and reliable pI values across different protein analyses.

Mass spectrometry analysis

Proteins in the total and supernatant fractions after the second centrifugation were digested using the filter-aided sample preparation method as described previously [79]. Proteins of each sample were denatured and reduced in SDT buffer (4% Sodium dodecyl sulphate (SDS) in 0.1 M Tris-HCl, pH 7.6, and 0.1 M Dithiothreitol (DTT)) at 37°C for 1 hour with shaking and boiling for 10 minutes at 100°C. The buffer was changed to 8 M urea in 0.1 M Tris-HCl, pH 8.5, and the proteins were alkylated with 0.5 M iodoacetamide for 30 minutes at 25°C in the dark. The MS grade trypsin protease (Pierce Biotechnology, IL, USA) was added to the filter at a trypsin to protein ratio of 1:50 (w/w) and incubated for 12 hours at 37°C. The resulting tryptic peptides were desalted using Pierce C-18 spin columns (Pierce Biotechnology, IL, USA), dried using Speed-Vac (Scanvac; LaboGene Aps, Lynge, Denmark), and kept at −80°C until the subsequent TMT labelling. The peptides of each fraction (T, S) were separately labelled using TMT 10-plex reagents (Thermo Fisher Scientific Inc., USA). The total fractions in the presence and absence of RNase A were labelled '126C and 128C' and '127N and 129N', respectively, while their corresponding soluble fractions were labelled '127C and 129C' and '128N and 130N', respectively. This approach ensured the reproducibility of our findings. Labelled peptide separation was performed using high-pH RPLC fractionation based on peptide hydrophobicity as described previously [80].

Labeled peptides were loaded on an analytical column (Xbridge, C18 5 µm, 4.6 mm × 250 mm) and separated into twelve fractions. A gradient was generated using an Agilent 1260 Infinity HPLC system (Agilent, Palo Alto, CA) operated with solvent A (10 mM Ammonium formate in water, pH 10.0) and solvent B (10 mM Ammonium formate in 90% ACN, pH 10.0). The gradient was as follows: 0–10 minute, 5% B; 10–70 minute, 35% B; 70–80 minute, 70% B; 80–85 minute, 70% B; 85–90 minute, 5% B; 90–105 minute, 5% B. The separated peptides were collected and dried in a speed-vac. Each fraction was desalted with a C18 spin column. The HPLC analysis was performed on Easy-nLC 1000 system (Thermo Fisher Scientific Inc., Germany) equipped with a trap column (C18, 75 um x 2 cm, 5 µm, Thermo Scientific., Germany) for cleanup followed by an analytical column (C18, 75 um x 50 cm, 2 µm, Thermo Scientific Inc., Germany). The column temperature was maintained at 60°C. The peptides were separated by using the mobile phase comprising solvent A (0.1% formic acid in water) and solvent

B (0.1% formic acid in acetonitrile) in a gradient elution mode. For LC-MS analysis, Easy-nLC 1000 system was coupled to Q-Exactive mass spectrometer (Thermo Fisher Scientific Inc., Germany). The typical operating source conditions for MS scan in positive ESI mode were optimized as follows: spray voltage, 2.0 KV; heated capillary temperature, 250°C; and nitrogen was used as damping gas. Full MS scans were acquired for the mass range of m/z 400–2000 at the resolution of 70,000 in MS1 level and the MS/MS analysis was performed by data-dependant mode. The MS2 level resolution was set as 35,000 with normalized collision energy of 30 for higher-energy collisional dissociation. The charge states of unassigned, 1, or > 6 were discarded and the dynamic exclusion of 30 seconds. The top ten precursor peaks were selected and isolated for fragmentation. The optimized linear gradient elution program for multiplex quantitation analysis was set as follows: (Tmin/% of solvent B): 0/5, 10/10, 110/35, 118/40, 120/80, 132/80, 134/5, 150/5.

LC-MS/MS data analysis

All tandem spectra data were analysed using Proteome Discoverer version 2.4 (Thermo Fisher Scientific). The analysis was carried out with the Sequest HT search engine against the 2017 version of the same reference database. Strict trypsin specificity was determined for up to two missed cleavages. Carbamidomethylation in cysteine (+57.021 Da), TMT 10-plex modification of lysine, and N-termination (+229.163 Da) were noted as static modifications, and oxidation of methionine (+15.995 Da) was noted as a variable modification. The FDR for peptide level was evaluated as 0.01 for removing as much false-positive data as possible. To quantify each reporter ion, 'peptide and protein quantifier' method in Proteome discoverer 2.4 with TMT 10-plex was used. We calculated the solubility using the reporter ion intensity ratio of the total and supernatant of the RNase A-treated and untreated groups. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD046018[81].

Western blot

Endogenous chaperones in the *E. coli* lysates were prepared and identified by western blot. After electrophoresis, the gel was transferred to PVDF membranes (Thermo Fisher Scientific) by using iBlot2 Transfer Stacks and iBlot2 Dry Blotting System. The membranes were blocked with 5% skim milk in TBST (20 mM Tris, 137 mM NaCl, 2.7 mM KCl, and 0.1% Tween-20) for 1 hour and then washed 3 times with TBST buffer. The blocked membranes were incubated with primary antibodies α-TF (M201, Takara), α-DnaK (ab69617, Abcam), α-DnaJ (ADI-SPA-410, Enzo), α-GroEL (ab82592, Abcam), and α-GroES (ab69823, Abcam) diluted in TBST overnight at 4°C. The membranes were washed 3 times in TBST, then incubated with a secondary antibody—α-mouse or α-rabbit IgG Ab conjugated with horseradish peroxidase (Sigma), depending on the origin of the first antibody – diluted 1:20,000 in TBST, for 40 minutes, and washed

3 times in TBST. The membranes were reacted with an ECL mixture using WEST-ZOL (Intron Biotechnology) and exposed to an X-ray film in a dark room.

Electrostatic surface potential of ribosomal proteins

The structure of the *E. coli* K12 strain ribosome complex (PDB ID: 8B7Y) was obtained from the Protein Data Bank, and all protein structural analyses were performed using PyMOL (The PyMOL Molecular Graphics System, Version 1.3 Schrödinger, LLC). Electrostatic potential maps were generated using PDB2PQR and APBS, which are part of the PyMOL APBS Tools (MGLerner and HA Carlson, 2006, University of Michigan, Ann Arbor).

Gene ontology analysis

In conducting the GO analysis on RNA-responsive proteins using the ShinyGO tool (version 0.77) [82], our study initially encompassed 913 RNA-responsive proteins. However, to refine our analysis, we excluded 50 ribosomal proteins. This exclusion was based on the observation that the top rank list was heavily populated with ribosome-related keywords, as per the insights derived from the EcoCyc database [83]. Thus, our focus shifted to 863 non-ribosomal proteins, further divided into acidic and basic groups based on their pI values. The GO analysis integrated Biological Process (BP), Molecular Function (MF), and Cellular Component (CC) categories into a comprehensive 'All gene set' ranking. This approach was chosen due to the limited GO results for individual categories and for a more articulate presentation of the data. For accurate analysis, specific background lists were employed: 2,648 proteins for the acidic group and 1,654 for the basic group. In the broader context, the entire *E. coli* proteome, consisting of 4,302 proteins, was used as the background for the analysis of all 863 non-ribosomal proteins. The analysis adhered to stringent criteria, applying FDR cut-off of 0.05 to control for multiple testing and to minimize false positives, thereby ensuring the reliability of the results. Furthermore, the 'All gene set' analysis illuminated the top 20 features across BP, MF, and CC categories, providing a detailed view of the functional roles of these proteins within the cellular environment.

Analysis of IDRs using multiple predictors

Protein sequences were analysed for IDRs employing IUPred, VSL2B, and DisEMBL [84–86]. IUPred was used with a threshold of 0.5 to identify both short and long disordered regions. VSL2B, set at a threshold of 0.5, was applied for detecting disordered regions of variable lengths. DisEMBL, utilizing its specific indicators for loops/coils, hot loops, and rem465, was employed with corresponding threshold values for each indicator. Each protein's full amino acid sequence was processed independently through these predictors. The proportion of the sequence classified as disordered by each tool was calculated based on the percentage of amino acids exceeding the respective thresholds. The averaged results from all three predictors were then compiled to determine the final

IDR percentage for each protein, aiming to minimize biases inherent in individual predictive models.

Data analysis methods in using R

In this study, comprehensive data analysis was carried out using the R programming environment (version 4.3.2), primarily utilizing the 'ggplot2' package. For histogram visualizations, density plots were added to represent the distribution of values as a proportion of the total population. Scatter plots were employed to examine different groups, with linear regression analysis conducted using the 'lm' function. This approach was instrumental in identifying significant relationships between various experimental factors. The regression lines in the scatter plots included SE to visually represent the variability around the estimates. Additionally, we calculated and displayed R^2 and p-value for each regression to quantify the strength and significance of the observed relationships.

Disclosure statement

Funding

Data availability statement

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD040618.

References

[1] Dobson CM. Protein folding and misfolding. Nature. 2003;426 (6968):884–890.

[2] Vendruscolo M. Proteome folding and aggregation. Curr Opin Struct Biol. 2012;22(2):138–143. doi: 10.1016/j.sbi.2012.01.005

[3] Varela AE, Lang JF, Wu Y, et al. Kinetic trapping of folded proteins relative to aggregates under physiologically relevant conditions. J Phys Chem B. 2018;122(31):7682–7698. doi: 10.1021/acs.jpcb.8b05360

[4] Choi SI, Seong BL. A social distancing measure governing the whole proteome. Curr Opin Struct Biol. 2021;66:104–111. doi: 10.1016/j.sbi.2020.10.014

[5] Chiti F, Protein Misfolding DC. Amyloid formation, and human disease: a summary of progress over the Last Decade. Annu Rev Biochem. 2017;86(1):27–68. doi: 10.1146/annurev-biochem-061516-045115

[6] Vendruscolo M, Knowles TP, Dobson CM. Protein solubility and protein homeostasis: a generic view of protein misfolding disorders. Cold Spring Harb Perspect Biol. 2011;3(12):a010454–a010454. doi: 10.1101/cshperspect.a010454

[7] Hartl FU, Hayer-Hartl M. Molecular chaperones in the cytosol: from nascent chain to folded protein. Science. 2002;295 (5561):1852–1858.

[8] Bukau B, Horwich AL. The Hsp70 and Hsp60 chaperone machines. Cell. 1998;92(3):351–366.

[9] Liu C, Young AL, Starling-Windhof A, et al. Coupled chaperone action in folding and assembly of hexadecameric rubisco. Nature. 2010;463(7278):197–202.

[10] Hartl FU, Bracher A, Hayer-Hartl M. Molecular chaperones in protein folding and proteostasis. Nature. 2011;475(7356):324–332.

[11] Das B, Chattopadhyay S, Bera AK, et al. In vitro protein folding by ribosomes from Escherichia coli, wheat germ and rat liver: the role of the 50S particle and its 23S rRNA. Eur J Biochem. 1996;235 (3):613–621. doi: 10.1111/j.1432-1033.1996.00613.x

[12] Kudlicki W, Coffman A, Kramer G, et al. Ribosomes and ribosomal RNA as chaperones for folding of proteins. Fold Des. 1997;2 (2):101–108. doi: 10.1016/S1359-0278(97)00014-X

[13] Choi SI, Han KS, Kim CW, et al. Protein solubility and folding enhancement by interaction with RNA. PloS One. 2008;3(7): e2677. doi: 10.1371/journal.pone.0002677

[14] Yang SW, Jang YH, Kwon SB, et al. Harnessing an RNA-mediated chaperone for the assembly of influenza hemagglutinin in an immunologically relevant conformation. FASEB J. 2018;32 (5):2658–2675. doi: 10.1096/fj.201700747RR

[15] Hwang BJ, Jang Y, Kwon SB, et al. RNA-assisted self-assembly of monomeric antigens into virus-like particles as a recombinant vaccine platform. Biomaterials. 2021;269:120650.

[16] Choi SI, Ryu K, Seong BL. RNA-mediated chaperone type for de novo protein folding. RNA Biol. 2009;6(1):21–24. doi: 10.4161/rna.6.1.7441

[17] Choi SI, Son A, Lim KH, et al. Macromolecule-assisted de novo protein folding. Int J Mol Sci. 2012;13(8):10368–10386. doi: 10.3390/ijms130810368

[18] Park C, Jin Y, Kim YJ, et al. RNA-binding as chaperones of DNA binding proteins from starved cells. Biochem Biophys Res Commun. 2020;524(2):484–489. doi: 10.1016/j.bbrc.2020.01.121

[19] Son A, Horowitz S, Seong BL. Chaperna: linking the ancient RNA and protein worlds. RNA Biol. 2021;18(1):16–23. doi: 10.1080/15476286.2020.1801199

[20] Docter BE, Horowitz S, Gray MJ, et al. Do nucleic acids moonlight as molecular chaperones? Nucleic Acids Res. 2016;44 (10):4835–4845. doi: 10.1093/nar/gkw291

[21] Begeman A, Son A, Litberg TJ, et al. G-Quadruplexes act as sequence-dependent protein chaperones. EMBO Rep. 2020;21 (10):e49735. doi: 10.15252/embr.201949735

[22] Son A, Huizar Cabral V, Huang Z, et al. G-quadruplexes rescuing protein folding. Proc Natl Acad Sci U S A. 2023;120(20): e2216308120. doi: 10.1073/pnas.2216308120

[23] Miller DW, Dill KA. Ligand binding to proteins: the binding landscape model. Protein Sci. 1997;6(10):2166–2179. doi: 10.1002/pro.5560061011

[24] Frankel AD, Smith CA. Induced folding in RNA-protein recognition: more than a simple molecular handshake. Cell. 1998;92 (2):149–151. doi: 10.1016/S0092-8674(00)80908-3

[25] Sanchez-Ruiz JM. Ligand effects on protein thermodynamic stability. Biophys Chem. 2007;126(1–3):43–49. doi: 10.1016/j.bpc.2006.05.021

[26] Randles LG, Batey S, Steward A, et al. Distinguishing specific and nonspecific interdomain interactions in multidomain proteins. Biophys J. 2008;94(2):622–628. doi: 10.1529/biophysj.107.119123

[27] Sen S, Udgaonkar JB. Binding-induced folding under unfolding conditions: switching between induced fit and conformational selection mechanisms. J Biol Chem. 2019;294(45):16942–16952. doi: 10.1074/jbc.RA119.009742

[28] Uversky VN, Alghamdi MF, Redwan EM. A bird's-eye view of proteomics. Curr Protein Pept Sci. 2021;22(8):574–583. doi: 10.2174/1389203722666210812120751

[29] Wright PE, Dyson HJ. Linking folding and binding. Curr Opin Struct Biol. 2009;19(1):31–38. doi: 10.1016/j.sbi.2008.12.003

[30] Masino L, Nicastro G, Calder L, et al. Functional interactions as a survival strategy against abnormal aggregation. FASEB J. 2011;25(1):45–54. doi: 10.1096/fj.10-161208

[31] Zacco E, Grana-Montes R, Martin SR, et al. RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43. J Mol Biol. 2019;431(8):1671–1688. doi: 10.1016/j.jmb.2019.01.028

[32] Son A, Choi SI, Han G, et al. M1 RNA is important for the in-cell solubility of its cognate C5 protein: implications for RNA-mediated protein folding. RNA Biol. 2015;12 (11):1198–1208. doi: 10.1080/15476286.2015.1096487

[33] Calabretta S, Richard S. Emerging roles of disordered sequences in RNA-Binding proteins. Trends Biochem Sci. 2015;40 (11):662–672. doi: 10.1016/j.tibs.2015.08.012

[34] Jarvelin AI, Noerenberg M, Davis I, et al. The new (dis)order in RNA regulation. Cell Commun Signal. 2016;14:9. doi: 10.1186/s12964-016-0132-3

[35] Protter DSW, Rao BS, Van Treeck B, et al. Intrinsically disordered regions can contribute promiscuous interactions to RNP granule assembly. Cell Rep. 2018;22(6):1401–1412. doi: 10.1016/j.celrep.2018.01.036

[36] Zhao B, Katuwawala A, Oldfield CJ, et al. Intrinsic Disorder in Human RNA-Binding Proteins. J Mol Biol. 2021;433(21):167229. doi: 10.1016/j.jmb.2021.167229

[37] Aarum J, Cabrera CP, Jones TA, et al. Enzymatic degradation of RNA causes widespread protein aggregation in cell and tissue lysates. EMBO Rep. 2020;21(10):e49585. doi: 10.15252/embr.201949585

[38] Kim CW, Han KS, Ryu KS, et al. N-terminal domains of native multidomain proteins have the potential to assist de novo folding of their downstream domains in vivo by acting as solubility enhancers. Protein Sci. 2007;16(4):635–643. doi: 10.1110/ps.062330907

[39] Thompson A, Schafer J, Kuhn K, et al. Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem. 2003;75(8):1895–1904. doi: 10.1021/ac0262560

[40] Sridharan S, Kurzawa N, Werner T, et al. Proteome-wide solubility and thermal stability profiling reveals distinct regulatory roles for ATP. Nat Commun. 2019;10(1):1155.

[41] Maatta TA, Rettel M, Sridharan S, et al. Aggregation and disaggregation features of the human proteome. Mol Syst Biol. 2020;16 (10):e9500. doi: 10.15252/msb.20209500

[42] Corley M, Burns MC, Yeo GW. How RNA-Binding proteins interact with RNA: molecules and mechanisms. Mol Cell. 2020;78(1):9–29. doi: 10.1016/j.molcel.2020.03.011

[43] Bondos SE, Dunker AK, Uversky VN. On the roles of intrinsically disordered proteins and regions in cell communication and signaling. Cell Commun Signal. 2021;19(1):88.

[44] Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. Nat Rev Mol Cell Biol. 2015;16 (1):18–29.

[45] Dare K, Ibba M. Roles of tRNA in cell wall biosynthesis. Wiley Interdiscip Rev RNA. 2012;3(2):247–264. doi: 10.1002/wrna.1108

[46] Aggarwal SD, Lloyd AJ, Yerneni SS, et al. A molecular link between cell wall biosynthesis, translation fidelity, and stringent response in streptococcus pneumoniae. Proc Natl Acad Sci U S A. 2021;118(14). doi: 10.1073/pnas.2018089118

[47] Grob G, Hemmerle M, Yakobov N, et al. tRNA-dependent addition of amino acids to cell wall and membrane components. Biochimie. 2022;203:93–105.

[48] Rüdiger S, Germeroth L, Schneider-Mergener J, et al. Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. EMBO J. 1997;16 (7):1501–1507. doi: 10.1093/emboj/16.7.1501

[49] Fenton WA, Kashi Y, Furtak K, et al. Residues in chaperonin GroEL required for polypeptide binding and release. Nature. 1994;371(6498):614–619. doi: 10.1038/371614a0

[50] Kramer G, Rauch T, Rist W, et al. L23 protein functions as a chaperone docking site on the ribosome. Nature. 2002;419 (6903):171–174.

[51] Georgellis D, Sohlberg B, Hartl FU, et al. Identification of GroEL as a constituent of an mRNA-protection complex in Escherichia coli. Mol Microbiol. 1995;16(6):1259–1268. doi: 10.1111/j.1365-2958.1995.tb02347.x

[52] Balakrishnan K, De Maio A. Heat shock protein 70 binds its own messenger ribonucleic acid as part of a gene expression self-limiting mechanism. Cell Stress Chaperones. 2006;11 (1):44–50.

[53] Deuerling E, Schulze-Specking A, Tomoyasu T, et al. Trigger factor and DnaK cooperate in folding of newly synthesized proteins. Nature. 1999;400(6745):693–696.

[54] Houry WA, Frishman D, Eckerskorn C, et al. Identification of in vivo substrates of the chaperonin GroEL. Nature. 1999;402 (6758):147–154.

[55] Paraskevopoulou V, Falcone FH. Polyionic tags as enhancers of protein solubility in recombinant protein expression. Microorganisms. 2018;6(2). doi: 10.3390/microorganisms6020047

[56] Qing R, Hao S, Smorodina E, et al. Protein design: from the aspect of water solubility and stability. Chem Rev. 2022;122 (18):14085–14179. doi: 10.1021/acs.chemrev.1c00757

[57] Sun Y, Arslan PE, Won A, et al. Binding of TDP-43 to the 3'UTR of its cognate mRNA enhances its solubility. Biochemistry. 2014;53(37):5885–5894. doi: 10.1021/bi500617x

[58] Zimmerman SB, Trach SO. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. J Mol Biol. 1991;222(3):599–620. doi: 10.1016/0022-2836(91)90499-V

[59] Speer SL, Stewart CJ, Sapir L, et al. Macromolecular Crowding Is More than Hard-Core Repulsions. Annu Rev Biophys. 2022;51 (1):267–300. doi: 10.1146/annurev-biophys-091321-071829

[60] Hu W, Qin L, Li M, et al. A structural dissection of protein-RNA interactions based on different RNA base areas of interfaces. RSC Adv. 2018;8(19):10582–10592. doi: 10.1039/C8RA00598B

[61] Kovachev PS, Banerjee D, Rangel LP, et al. Distinct modulatory role of RNA in the aggregation of the tumor suppressor protein p53 core domain. J Biol Chem. 2017;292(22):9345–9357. doi: 10.1074/jbc.M116.762096

[62] Kovachev PS, Gomes MPB, Cordeiro Y, et al. RNA modulates aggregation of the recombinant mammalian prion protein by direct interaction. Sci Rep. 2019;9(1):12406.

[63] Protter DSW, Parker R. Principles and properties of stress granules. Trends Cell Biol. 2016;26(9):668–679. doi: 10.1016/j.tcb.2016.05.004

[64] Liu M, Li H, Luo X, et al. RPS: a comprehensive database of RNAs involved in liquid-liquid phase separation. Nucleic Acids Res. 2022;50(D1):D347–D55. doi: 10.1093/nar/gkab986

[65] Samelson AJ, Jensen MK, Soto RA, et al. Quantitative determination of ribosome nascent chain stability. Proc Natl Acad Sci U S A. 2016;113(47):13402–13407. doi: 10.1073/pnas.1610272113

[66] Sörensen T, Leeb S, Danielsson J, et al. Polyanions cause protein destabilization similar to that in live cells. Biochemistry. 2021;60 (10):735–746.

[67] Kishor A, White EJF, Matsangos AE, et al. Hsp70's RNA-binding and mRNA-stabilizing activities are independent of its protein chaperone functions. J Biol Chem. 2017;292(34):14122–14133. doi: 10.1074/jbc.M117.785394

[68] Huang YW, Hu CC, Liou MR, et al. Hsp90 interacts specifically with viral RNA and differentially regulates replication initiation of bamboo mosaic virus and associated satellite RNA. PLOS Pathog. 2012;8(5):e1002726. doi: 10.1371/journal.ppat.1002726

[69] Yan W, Schilke B, Pfund C, et al. Zuotin, a ribosome-associated DnaJ molecular chaperone. EMBO J. 1998;17(16):4809–4817. doi: 10.1093/emboj/17.16.4809

[70] Henics T. Extending the 'stressy' edge: molecular chaperones flirting with RNA. Cell Biol Int. 2003;27(1):1–6. doi: 10.1016/S1065-6995(02)00286-X

[71] Ghosh J, Basu A, Pal S, et al. Ribosome-DnaK interactions in relation to protein folding. Mol Microbiol. 2003;48(6):1679–1692. doi: 10.1046/j.1365-2958.2003.03538.x

[72] Kim HK, Choi SI, Seong BL. 5S rRNA-assisted DnaK refolding. Biochem Biophys Res Commun. 2010;391(2):1177–1181. doi: 10.1016/j.bbrc.2009.11.176

[73] Joyce GF. The antiquity of RNA-based evolution. Nature. 2002;418(6894):214–221.

[74] Higgs PG, Lehman N. The RNA World: molecular cooperation at the origins of life. Nat Rev Genet. 2015;16(1):7–17.

[75] Armaos A, Zacco E, Sanchez de Groot N, et al. RNA-protein interactions: central players in coordination of regulatory networks. BioEssays. 2021;43(2):e2000118.

[76] Wiedner HJ, Giudice J. It's not just a phase: function and characteristics of RNA-binding proteins in phase separation. Nat Struct Mol Biol. 2021;28(6):465–473.

[77] Milicevic K, Rankovic B, Andjus PR, et al. Emerging roles for phase separation of RNA-Binding proteins in cellular pathology of ALS. Front Cell Dev Biol. 2022;10:840256. doi: 10.3389/fcell.2022.840256

[78] Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000;16(6):276–277. doi: 10.1016/S0168-9525(00)02024-2

[79] Manza LL, Stamer SL, Ham AJ, et al. Sample preparation and digestion for proteomic analyses using spin filters. Proteomics. 2005;5(7):1742–1745. doi: 10.1002/pmic.200401063

[80] Yang F, Shen Y, Camp DG 2nd, et al. High-pH reversed-phase chromatography with fraction concatenation for 2D proteomic analysis. Expert Rev Proteomics. 2012;9(2):129–134. doi: 10.1586/epr.12.15

[81] Perez-Riverol Y, Bai J, Bandla C, et al. The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. 2022;50(D1):D543–D52. doi: 10.1093/nar/gkab1038

[82] Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics. 2020;36 (8):2628–2629. doi: 10.1093/bioinformatics/btz931

[83] Keseler IM, Gama-Castro S, Mackie A, et al. The EcoCyc database in 2021. Front Microbiol. 2021;12:711077. doi: 10.3389/fmicb.2021.711077

[84] Dosztanyi Z, Csizmok V, Tompa P, et al. The pairwise energy content estimated from amino acid composition discriminates between folded and intrinsically unstructured proteins. J Mol Biol. 2005;347(4):827–839. doi: 10.1016/j.jmb.2005.01.071

[85] Obradovic Z, Peng K, Vucetic S, et al. Exploiting heterogeneous sequence properties improves prediction of protein disorder. Proteins. 2005;61(Suppl 7):176–182. doi: 10.1002/prot.20735

[86] Linding R, Jensen LJ, Diella F, et al. Protein disorder prediction: implications for structural proteomics. Structure. 2003;11 (11):1453–1459.