Dynamic Remodeling of Escherichia coli Interactome in Response to Environmental Perturbations

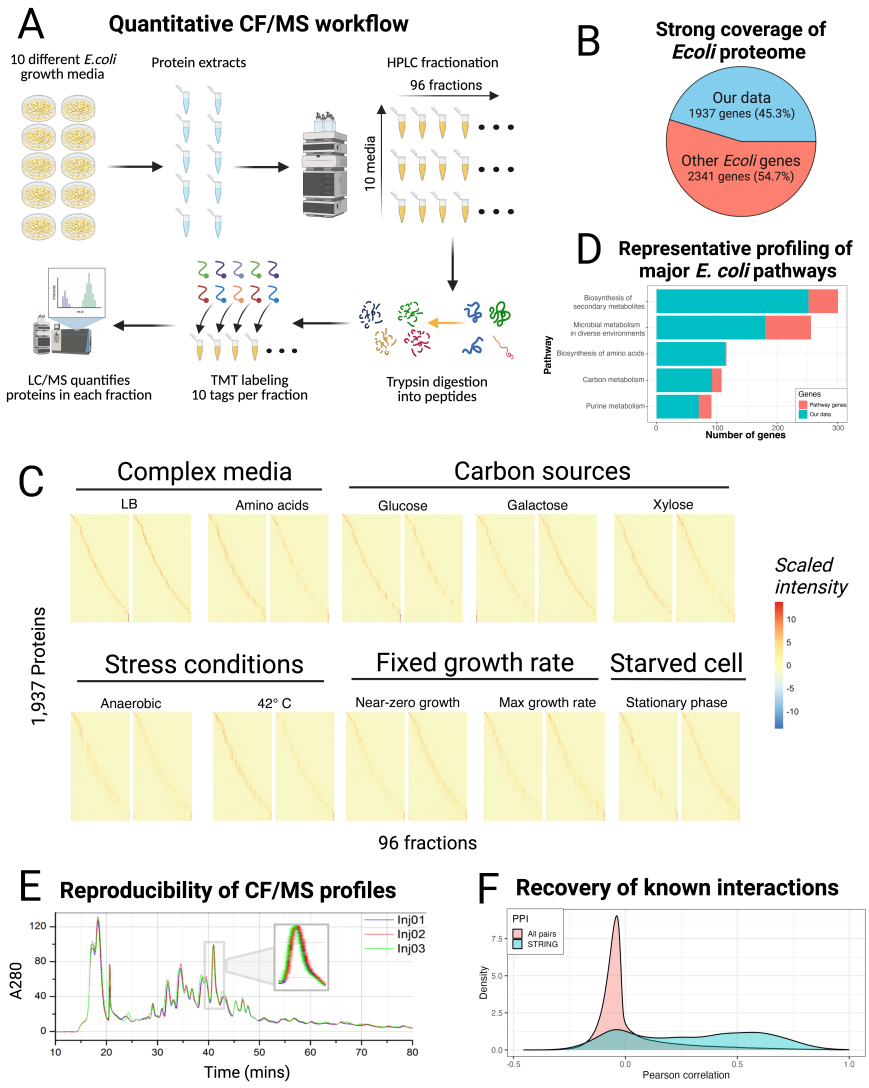
Ahmed Youssef[1], Fei Bian[2], Nicolai Paniikov[3], Mark Crovella[1], and Andrew Emili[4]

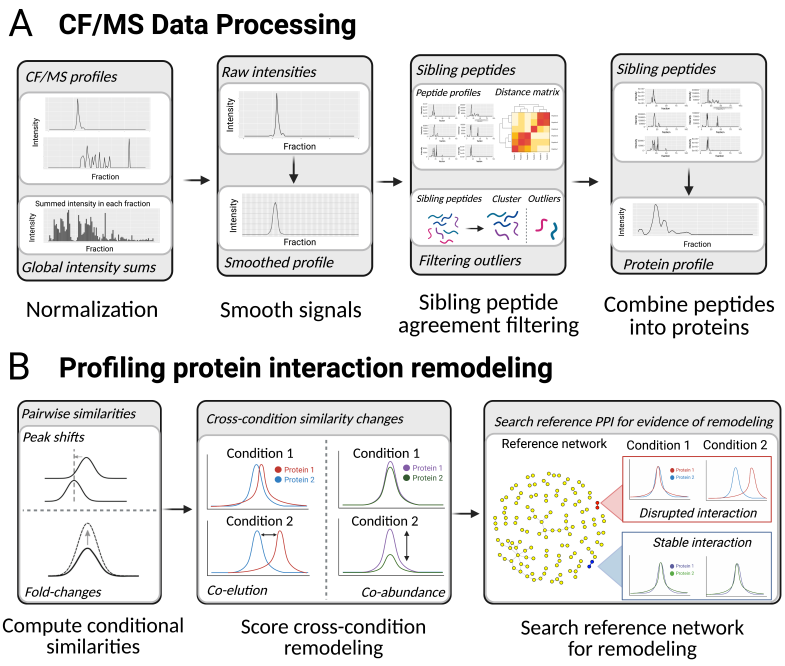
[1]Boston University
[2]Shandong Academy of Agricultural Sciences
[3]Northeastern University - Boston Campus
[4]Oregon Health & Science University

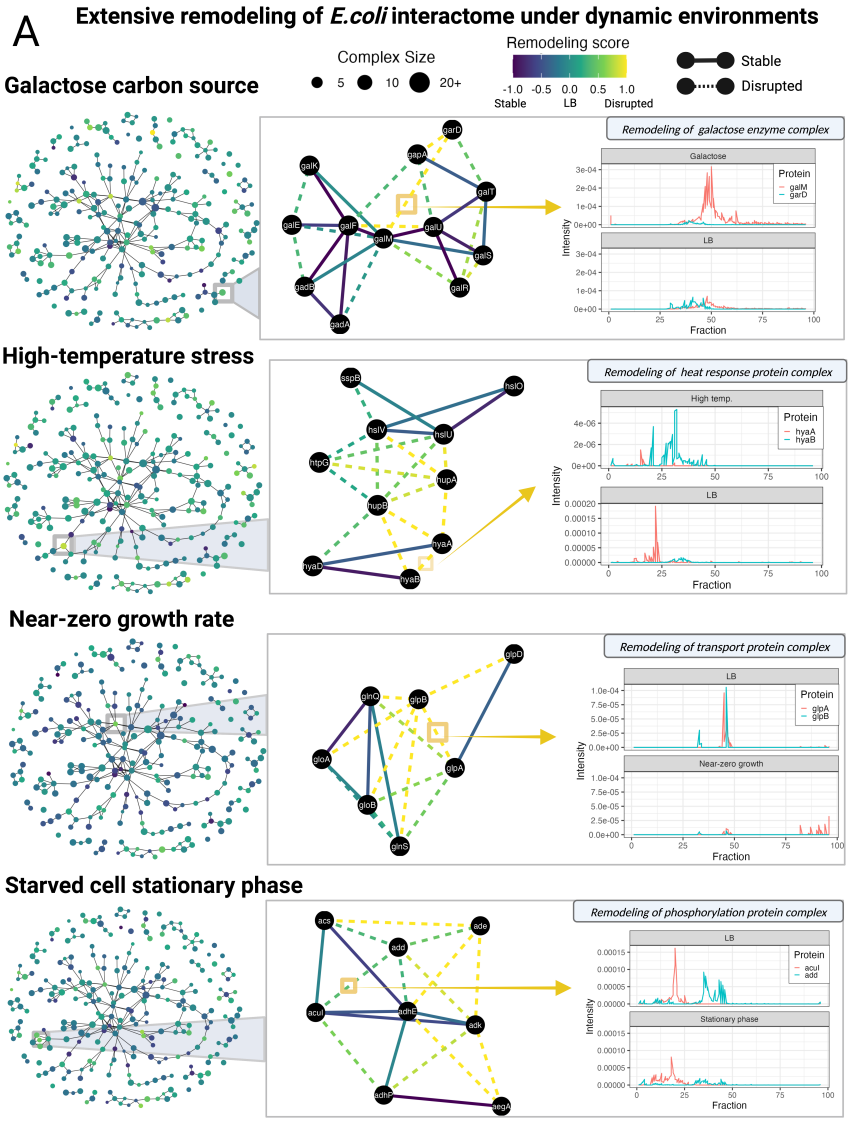
April 3, 2023

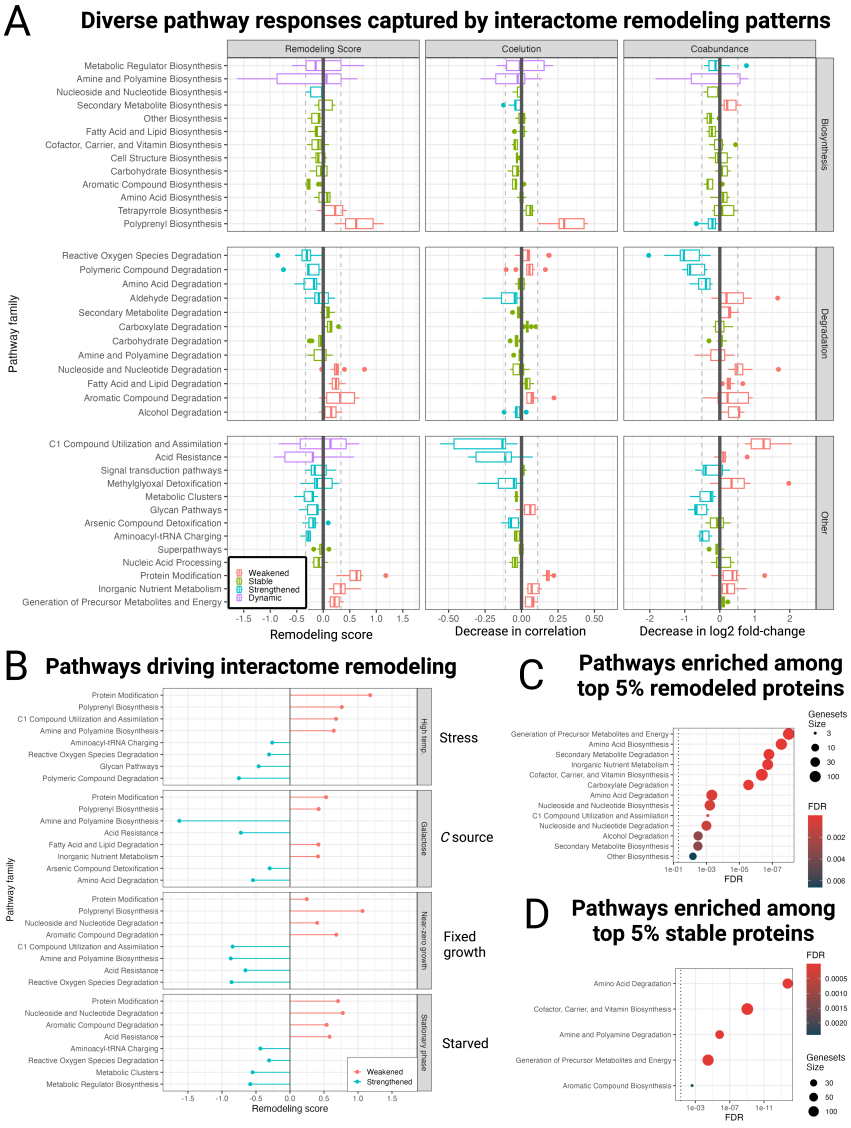
Abstract

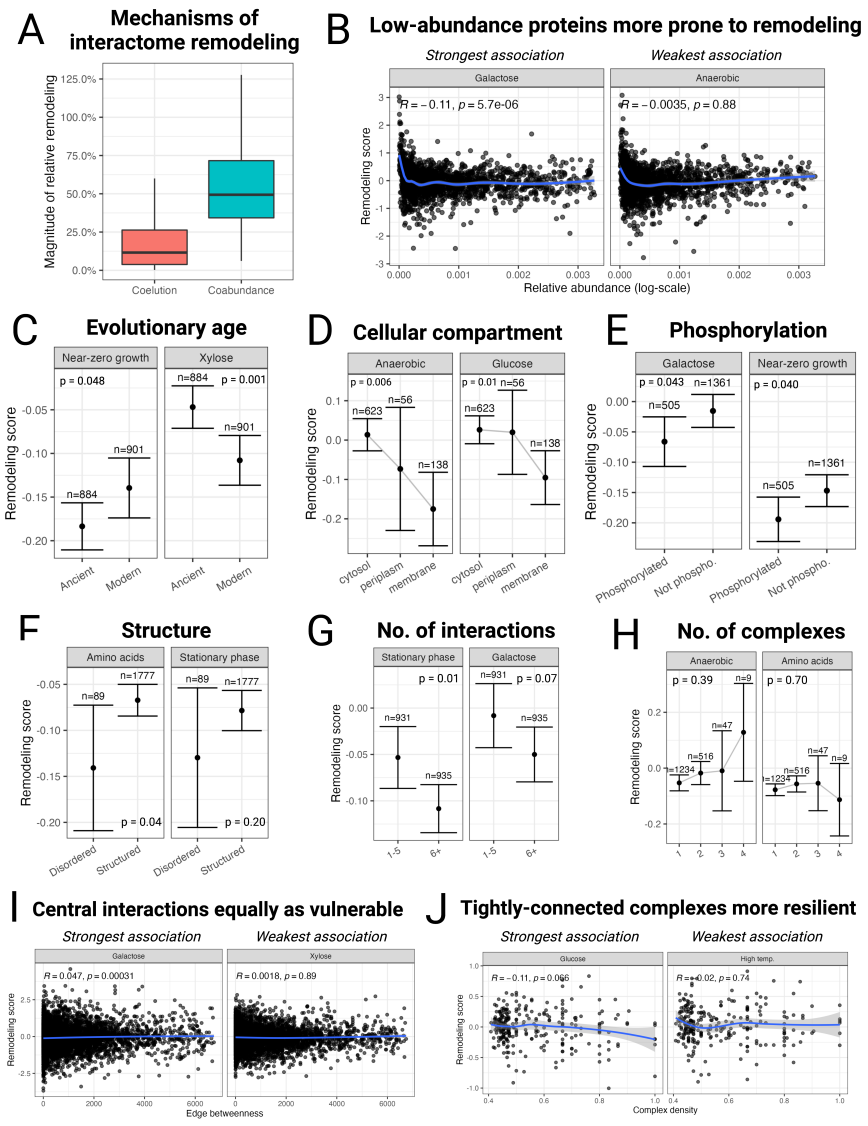
Proteins play an essential role in the vital biological processes governing cellular functions. Most proteins function as members of macromolecular machines, with the network of interacting proteins revealing the molecular mechanisms driving the formation of these complexes. Profiling the physiology-driven remodeling of these interactions within different contexts constitutes a crucial component to achieving a comprehensive systems-level understanding of interactome dynamics. Here, we apply co-fractionation mass spectrometry and computational modeling to quantify and profile the interactions of ˜2,000 proteins in the bacterium Escherichia coli cultured under ten distinct culture conditions. The resulting quantitative co-elution patterns revealed large-scale condition-dependent interaction remodeling among protein complexes involved in diverse biochemical pathways in response to the unique environmental challenges. Network-level analysis highlighted interactome-wide biophysical properties and structural patterns governing interaction remodeling. Our results provide evidence of the local and global plasticity of the E. coli interactome along with a rigorous generalizable framework to define protein interaction specificity. We provide an accompanying interactive web application to facilitate exploration of these rewired networks.











Dynamic Remodeling of *Escherichia coli* Interactome in Response to Environmental Perturbations

Ahmed Youssef[1,2,8], Fei Bian[2,5,8*], Nicolai S. Paniikov[7], Mark Crovella[1,3,4], Andrew Emili[1,2,4,6*]

Affiliations

[1] Graduate Program in Bioinformatics, Boston University

[2] Center for Network Systems Biology, Boston University

[3] Computer Science Department, Boston University

[4] Faculty of Computing and Data Sciences, Boston University

[5] Institute of Crop Germplasm Resources, Shandong Academy of Agricultural Sciences, Jinan, China

[6] Knight Cancer Institute, Oregon Health and Science University

[7] Department of Chemistry and Chemical Biology, Northeastern University

[8] These authors contributed equally to the work.

* Correspondence to: bxf.9@163.com (FB), aemili@bu.edu (AE).

Abstract

Proteins play an essential role in the vital biological processes governing cellular functions. Most proteins function as members of macromolecular machines, with the network of interacting proteins revealing the molecular mechanisms driving the formation of these complexes. Profiling the physiology-driven remodeling of these interactions within different contexts constitutes a crucial component to achieving a comprehensive systems-level understanding of interactome dynamics. Here, we apply co-fractionation mass spectrometry and computational modeling to quantify and profile the interactions of ~2,000 proteins in the bacterium *Escherichia coli* cultured under ten distinct culture conditions. The resulting quantitative co-elution patterns revealed large-scale condition-dependent interaction remodeling among protein complexes involved in diverse biochemical pathways in response to the unique environmental challenges. Network-level analysis highlighted interactome-wide biophysical properties and structural patterns governing interaction remodeling. Our results provide evidence of the local and global plasticity of the *E. coli* interactome along with a rigorous generalizable framework to define protein interaction specificity. We provide an accompanying interactive web application to facilitate exploration of these rewired networks.

Significance Statement

The protein interactome contains the network of physical interactions that enable the functions of most proteins. Protein interactions can be disrupted by many triggers, such as pathogen infection or mutations in protein-coding genes, yet most studies in the field focused on characterizing the interactome in a static manner, with few devoted to investigating the dynamic nature of these interactions. In this study, we profiled the dynamics of the *Escherichia coli* interactome in response to changes in its growth environment. Our results shed light on the mechanisms governing protein interaction remodeling, while also providing a rigorous analytical framework for quantifying interaction dynamics on an interactome-wide scale, representing an important step towards accurate modeling of dynamic biological systems.

Introduction

Microorganisms abundantly exist across all types of ecological environments, but the molecule basis for these adaptive responses is not fully understood. In the laboratory, certain bacterial strains can be cultured in a variety of conditions spanning a range of temperatures, pH, and media compositions, s resulting in different phenotypes and growth rates (μ). For example, *Escherichia coli* can propagate in the Luria-Bertani (LB) medium with a near maximal generation time (μ_{max}) of ~20 minutes at 37°C. However, in its natural environment, which can be nutrient-deficient, the growth rate is significantly reduced. Under nutrient deprivation, microorganisms like *E. coli* can maintain continuous but extremely slow growth rates ($\mu << \mu_{max}$) called "near-zero growth" (NZG)[1,2]. Accurate descriptions of the molecular mechanisms supporting these vastly different physiological states in different environments is crucial to elucidating the fundamental relationship between genotype and phenotype.

While bacteria are known to regulate biochemical responses through transcriptional control of gene operons, post-translational regulation is also thought to mediate their adaptation to changing physiological demands[3,4,5]. Proteomics research therefore plays a particularly important role in elucidating the physiological state of bacteria due to the crucial role of proteins in executing essential cellular functions. Comparative proteome studies have found that the distribution of protein resources in bacteria is related to their growth rate[3]. Peebo et al. used a chemostat to isolate *E. coli* within a range of μ =0.2~0.9 h^{-1}, and found that more proteins in slow-growing cells were used for energy generation, carbohydrate transport and metabolism, whereas most proteins in fast-growing cells functioned in biological processes closely related to protein synthesis pathways[4]. More recently, Schmidt et al. measured the relative abundance of more than 2,300 proteins in *E. coli* under 22 culture conditions, and found that growth rate was positively-correlated with the amount of amino acid transport and ribosomal biogenesis, and negatively-correlated with energy generation pathways[5].

Proteins do not function in isolation in the cell, rather they selectively interacting to form large multi-subunit complexes that collectively are known as the 'interactome'. Elucidating the composition and overall network properties of the interactome is key for revealing the molecular mechanism of cell growth and environmental adaptation on a proteome-wide level. Butland et al and Hu et al used affinity purification mass spectrometry (AP-MS) to define a dense network of protein-protein interactions (PPIs) among the soluble protein complexes of *E. coli* cultured in LB

medium[6,7]. while Babu et al. reported PPIs among membrane-associated proteins[9] under a single static growth condition.

Multiplex co-fractionation mass spectrometry (mCF-MS) is a flexible approach for -detecting and comparing protein complexes and PPI networks under different cellular contexts[10]. In mCF-MS, cellular lysates are biochemically fractionated prior to mass spectrometry-based shotgun sequencing and relative protein quantification. Ion-exchange chromatography (IEX) is a particularly effective method to resolve complex protein mixtures[11,12], while sophisticated computational analysis tools are then used to assign proteins to a given complex based on the similarity of their co-elute profiles. Due to its quantitative, high-throughput nature, mCF-MS technology allows for direct comparison of interactome differences between distinct samples after controlling for spurious variance using biological replicates.

The protein interactome is a dynamic system that changes in response to different stimuli and and environments. Studies on the dynamic response of *S. cerevisiae* PPIs to environmental disturbances showed that more than half of the PPIs only existed under specific culture conditions[13]. Changes of a single protein or a small amount of protein would lead to changes in PPIs, and some proteins modifications would also lead to changes in interactions[13]. Other studies have also revealed that the location and abundance of proteins in *S. cerevisiae* cells influence PPI formation[14,15]. The proteome profile of bacteria similarly varies under different conditions, and so it follows that the bacterial interactome is expected to exhibit dynamic assembly patterns. However, there has been no systematic study yet on the bacterial interactome under different conditions, with the existing E. coli interactome only constructed for LB medium.

In this study, we used mCF/MS along with a customized data analysis pipeline to compare differences in the PPI networks of *E. coli* cells grown under 10 different culture conditions. We defined and investigated extensive condition-dependent remodeling predicted by this dataset. Projecting this PPI remodeling against evolutionary and biochemical traits allowed us to pinpoint key biological factors driving protein interaction dynamics, while statistical assessment of the dynamic networks revealed fundamental mechanistic principles underlying interactome plasticity. Along with a robust computational framework to support future studies of this nature, we developed an interactive web application to facilitate exploration of these results by the broader research community.

Results

Experimental design

We grew *E. coli* BW25113 under 10 alternate growth conditions representing different types of environmental variations (**Table 1**): (i) growth on two distinct types of complex medium, (ii) growth on defined (minimal) medium with one of three different carbon sources, (iii) growth on glucose minimal medium with two different stress (anaerobic growth, heat shock) conditions, (iv) growth in glucose-limited chemostat cultures with varying growth rates (NZG, $\mu \approx 0$; Max, $\mu \approx 0.8$) and (v) stationary phase.. The reasoning behind selecting these particular growth conditions is elaborated upon in the Methods section.

Generation of condition-dependent CF/MS proteome profiles

We used multiplex co-fractionation mass spectrometry (mCF/MS; see Methods) to generate proteomic profiles encompassing 1,937 *E. coli* proteins, in replicate, across 96 IEX-HPLC fractions for each of the 10 growth conditions (**Fig. 1A**). We used isobaric Tandem Mass Tag (TMT) stable isotope chemical labeling to quantify the condition-specific protein elution profiles for just under half of all curated protein-coding genes (**Fig. 1B**). These mCF/MS profiles consisted of quantitative measurements of each protein relative levels in each fraction, using summed precursor (MS2) ion intensity as a proxy for protein abundance (**Fig. 1C**). The proteins profiled in this dataset represent strong coverage (>70%) of the five largest functional annotation pathways in the KEGG[24] database (**Fig. 1D**), with all pathways well-represented on average with >53% coverage. Additionally, chromatograms showed excellent reproducibility between HPLC runs (triplicate technical injections are shown in **Fig. 1E**). Given a key premise of mCF/MS is that physically-interacting proteins display similar elution profiles, we found that pairs of proteins with high-confidence experimental evidence of interacting in the SRING database[16] had significantly higher correlations in our dataset than random protein pairs ($p < 2.2$ e-16; **Fig. 1F**), demonstrating strong recovery of canonical protein interactions.

Computational strategy for profiling interactome remodeling from dynamic CF/MS data

A key challenge of the mCF/MS shotgun sequencing routine is that the multiplexed measurements are made at the peptide-level, after in vitro-digestion of the protein fractions, while quantitative inferences must be assigned at the higher protein-level. Moreover, while our dataset

was expected to be dynamic, existing tools such as *EPIC*[17], *PrInCe*[18], and *CCprofiler*[19] were devised for analyzing static PPI networks using standard non-multiplexed CF/MS data, each proposing different pre-processing strategies. Thus, we developed a pre-processing workflow starting from the peptide-level with the optimal set of steps and parameters for our data that led to the best recapture of literature-reported PPIs (see **Methods**).

Figure 2 outlines the computational analysis pipeline we developed to profile dynamic protein interaction remodeling based on conditional mCF/MS data. The pipeline consists of two modules, with the first dedicated to pre-processing the mCF/MS data (**Fig. 2A**), and the second responsible for quantitatively scoring protein interaction remodeling (**Fig. 2B**).

The second module is concerned with quantitatively profiling protein interaction remodeling from the pre-processed mCF/MS data. Our strategy begins by computing conditional similarity scores for each pair of interacting proteins within a reference interactome predicted from the mCF/MS data using established algorithms (see Methods). We then leveraged the quantitative nature of the ten mCF/MS reporter channel measurements to quantify the nature and extent of remodeling exhibited by each putative binding partner in response to each growth environment based on changes in the similarity of the interactor mCF/MS profiles. This allowed us to profile changes in the interaction patterns at different levels of biological organization, starting from individual pairs of proteins and multi-protein complexes, through to whole interactome, in response to the different growth conditions. Our pipeline enabled the identification of key pathways and molecular mechanism driving interactome remodeling, as described in later sections, while providing a rigorous generalizable framework for interactome remodeling using mCF/MS data. A detailed breakdown of each step in the analysis pipeline can be found in the Methods section.

We first predicted a reference interactome spanning 6,152 high-confidence pairwise interactions among the quantified *E. coli* proteins in the mCF/MS data by combining and scoring all 10 datasets using the established *EPIC* software[16] (see Methods and Supp. Table 1). This unified interactome was shown to be highly modular (Louvain modularity = 0.89), encompassing267 putative multi-component complexes ranging in size from just three subunits to a large ribosomal assembly consisting of 35 polypeptides (**Supp. Table 2**). While most proteins (68.3%) were assigned to a single complex, several 'moonlighting' proteins were predicted to function as members of multiple distinct complexes, with nine proteins operating in as many as four

complexes, including molybdopterin cofactors (*moaE*, *mobA*, *modE*, *moeA*) and rRNA methylases (*rlmA*, *rlmB*, *rlmG*).

Extent of global remodeling of E. coli interactome

We used the standard LB growth medium as a baseline condition to quantify the extent of interactome remodeling occurring in response to the other nine different environmental perturbations (as summarized in **Table 1**). By scoring each constituent interaction's conditional remodeling relative to the LB reference , we placed the predicted protein complexes on a spectrum of stable-to-disrupted complexes detected within each growth medium, with complexes whose underlying mCF/MS profiles showed the largest increase in dissimilarity being assigned higher remodeling scores (see Methods). We found that while individual complexes reacted differently depending on the specific culture conditions, the overall extent of interactome remodeling was remarkably similar across all conditions, with most complexes (57.8% on average) remaining quantitatively unchanged relative to the LB baseline, whereas only a small <5% fraction exhibited high (> 0.5) remodeling scores suggestive of disrupted interactions (**Supp. Table 2**).

Figure 3 shows the interactome-wide patterns of protein complex remodeling seen among select representatives of the four major test conditions evaluated in the comparative experiment. We highlight an example complex from each condition based on biological relevance that are impacted by relatively high levels of PPI remodeling. The quantitative nature of the mCF/MS measurements enabled deciphering the nature of the intra-complex remodeling occurring (e.g. subunit loss versus changes in overall macromolecular abundance relative to LB), and we show the underlying mCF/MS data for a representative remodeled interaction per complex.

For example, in the condition where galactose was the primary carbon source, we found evidence of extensive formation of a protein complex composed mainly of galactose metabolism enzymes, where the assembly extends from one large unit into two distinct subunits connected by the *galM* epimerase protein. Additionally, the dynamic association of *garD*, a galactarate dehydrase, suggests a peripheral role for this protein with the core complex. Together, these remodeling patterns suggest the presence of condition-dependent macromolecules driven by the availability of galactose and simultaneous trimming of less essential interactions.

In response to high temperature (42 °C), we found that one of the top remodeled complexes showed evidence of decomposition into two separate subcomplexes, with one consisting of the

heat shock response proteins from the '*hsl'* family, and the other consisting of hydrogenase proteins from the '*hya'* family of genes. Meanwhile, the dimeric histone-like master transcriptional regulators *hupA* and *hupB* and chaperone protein *htpG* dissociated entirely from a prominent assembly seen in LB. A potential explanation for this behavior could be induction of a heat shock transcriptional response in order for the organism to cope with the high temperature..

A striking example of the influence of highly-connected hub proteins on complex remodeling was observed when *E. coli* was cultured in a chemostat at near-zero growth rate. The complex with the highest remodeling score in this condition involved several transporter proteins, with the respiratory enzyme *glpB* playing a central role by being the only member that physically interacts with all complex members. Both the downregulation and elution shifting of *glpB* under the near-zero growth condition coincided with the destruction of all its intra-complex interactions and the consequent loss of this complex, demonstrating the dynamic nature of complexes with reliance highly-connected subunits for their structural integrity.

Finally, another complex that had a high remodeling score was detected preferentially among starved cells, showing extensive disruption primarily due to the loss of interactions involving the two highly-connected subunits *ade* and *add*. Proteins in this conditional assembly have been associated with multiple types of post-translational protein modifications, including phosphorylation and acetylation[20,21], implying altered activity of an upstream signaling mechanism as a key response to the starvation condition, presumably triggering energy conservation mechanisms.

Biological pathways driving interactome remodeling

Using annotated pathway membership information from the EcoCyc curation database[22] combined with our reference EPIC-derived interactome, we quantified the extent of conditional remodeling detected in each condition among interaction partners mapping to 38 major biochemical pathway families, classifying them according to the spread of remodeling scores around the baseline LB condition (**Fig. 4A**). While most (61.5%) biosynthesis pathway families showed relatively stable behavior across conditions compared to LB, notable exceptions included strengthening of genetic machinery involved in cell replication in conjunction with consistent weakening of interactions related to the synthesis of polyprenyl and tetrapyrrole, which are known to be growth inhibitory. Interestingly, metabolic regulator biosynthesis displayed a wide range of

dynamic remodeling, being strengthened in some conditions and weakened in others. Dissecting the interaction behavior showed that these changes are primarily driven by changes in the elution profiles of the proteins rather than changes in relative abundance, suggesting possible biochemical alterations to the structures of these complexes.

In contrast to the biosynthesis pathways, only one-third of pathways associated with biomolecule degradation were relatively stable across conditions. Reactive oxygen species degradation, a hallmark stress response, was consistently impacted across all changes in the media, alongside changes in the machinery linked to the degradation of amino acids and aldehydes. On the other hand, we observed a relative weakening, i.e. decrease, in assemblies linked to degradation of fatty acids and lipids, aromatic compounds and alcohol, with alterations primarily driven by abundance changes. Additionally, interactions involving signal transduction pathways were elevated concomitant with increased activity of signaling cascades in response to the environmental perturbations (**Figure 4A**). A similar pattern was observed for enzyme assemblies linked to metabolic detoxification and glycan formation. Conversely, persistent global weakening of interactions among components of the protein modification and energy generation pathways was seen under non-conventional environments, reflecting a shift to increased conservation.

We also identified pathways most influential in driving the remodeling within each growth condition. We consequently examined assemblies linked to the top four strengthened and weakened pathways detected within each culture setting (**Fig. 4B**). Consistent with the global trends (**Fig. 4A**), the protein modification machinery was among the most consistently and severely impacted systems across all conditions relative to LB, indicative of the key role played by dynamic post-translational modifications. Interestingly, acid resistance (ability to withstand pH <2.5) was among the most-strengthened pathways detected under near-zero growth yet one of the most weakened in starved cells, despite being generally associated with the stationary phase in the literature[ref].

Hypergeometric enrichment tests revealed that amino acid degradation, cofactor biosynthesis, and precursor metabolite generation pathways are significantly enriched (FDR < 0.01) among both the top 5% remodeled and top 5% stable proteins on average across conditions (**Fig. 5C&D**). Amine degradation and aromatic compound biosynthesis were exclusively strongly represented among the most stable proteins (**Fig. 5D**), while a larger set of 13 pathways spanning diverse biological mechanisms were enriched among the most highly remodeled proteins,

indicative of the multi-faceted nature of interactome disruption across the diverse growth conditions.

Structural and functional properties of interactome remodeling

Our mCF/MS data enabled the capture of two distinct types of data patterns suggestive of interaction remodeling: (1) qualitative: elution (HPLC retention time) shifts, i.e. impacting the overlap of the proteins' cross-fraction co-elution patterns, and (2) quantitative: intensity fold-changes, i.e. the ratio between protein relative expression (co-abundance) between conditions (**Fig. 2B**). While we opted for a single integrative analysis strategy designed to generate a single score combining the changes seen in both patterns (see Methods), we also compared the patterns of each reference interaction and found that the magnitude of changes in interactors' co-abundance tended to be significantly greater than that exhibited by coelution changes, suggesting that expression-level regulation plays a prominent role as the dominant mechanism influencing downstream interaction remodeling (**Fig. 5A**).

To identify biological properties that distinguish proteins based on their levels of interaction remodeling, we computed associations between diverse protein traits and their corresponding averaged integrated remodeling score within each growth condition separately. A negative correlation between remodeling scores and summed protein intensities indicated that low-abundance proteins were more prone to remodeling (**Fig. 5B**), consistent with past findings from mammalian interactome remodeling studies[24]. Classifying the proteins based on their evolutionary age also revealed that certain conditions, including near-zero growth and stationary phase, favored ancient protein interaction stability, implying that adaptation to certain types of environments, such as growth on xylose-rich media, is a modern adaptation (**Fig. 5C**). Strikingly, we found that membrane proteins were relatively more stable than their cytosolic and periplasmic counterparts (**Fig. 5D**), while proteins subject to phosphorylation were likewise generally more stable (**Fig. 5E**). Counter-intuitively, proteins annotated as containing intrinsically disordered structures had a lower median remodeling score (**Fig. 5F**), suggesting they tend to form constitutive assemblies. Similarly, highly-connected hub proteins that participated in many (6 or more) interactions were more stable, likely due to their persistent and essential role in maintaining interactome structure (**Fig. 5G**), while unexpectedly the number of complexes a protein participates in was less influential (**Fig. 5H**).

We explored the overarching structural patterns of interactome remodeling by examining the relationship between the integrated remodeling scores and the mathematical properties at both the pairwise interaction level and the protein complex level. We found no correlation between each interaction's betweenness score within the interactome and its corresponding remodeling propensity (**Fig. 5I**). This indicates that unlike hub proteins, individual central interactions have the same tendency to be remodeled as more peripheral ones. Meanwhile, tightly-connected protein complexes had relatively lower remodeling scores on average than those with sparser intra-complex connections, suggesting higher structural resilience to changes in the surrounding environment (**Fig. 5J**).

Discussion

Here, we present the results of a multi-factorial perturbation experiment investigating the global robustness and localized dynamics of the *E. coli* interactome in response to different types of environmental perturbations. We cultured a K-12 laboratory strain under 10 different growth media and generated high-throughput mCF/MS data to enable comparative protein interaction prediction and quantitative profiling of interactome dynamics. We found rewired protein complexes that were altered preferentially, or even exclusively, in certain physiological contexts that highlight key players in environmental adaptation responses. We also pinpointed macromolecules and their associated biological pathways driving this remodeling in comparison to crucial (housekeeping) assemblies that remain universally unaltered (stable core). Our integrative scoring approach also revealed interactome-wide biological, biophysical, and structural patterns governing the tendency of bacterial interactions to become disrupted or strengthened.

To facilitate exploration of the results, an interactive web application visualizing the dynamic *E.coli* mCF/MS profiles is available at https://bnfweb.bu.edu/EcoliDynamicInteractome/. Finally, we note that the experimental and computational pipelines reported here provide a generalizable workflow for future studies of interactome dynamics in other settings.

Methods

Strain

The wild-type *Escherichia coli* K-12 strain BW25113 [1] was stored in the lab and used to generate the data for all 10 culture conditions.

Media

Chemical reagents for media preparation were purchased from Sigma-Aldrich unless specified otherwise. The LB broth miller was purchased from Fisher BioReagents™. Twenty-five grams of LB broth power, with tryptone 10 g, yeast extract 5 g and NaCl 10 g, was suspended in one liter of Mini-Q water and sterilized by autoclaving. The LB plates were produced by adding 1.5 g agar to 100 mL LB medium mixture before autoclaving.

M9 minimal medium without carbon source was prepared in the following ways: 200 ml of 5×M9 salts (Na_2HPO_4 33.9g/L, KH_2PO_4 15g/L, NH4Cl 5g/L, NaCl 2.5 g/L), 1 ml 1 M $MgSO_4$ solution, 0.1 ml 1 M $CaCl_2$ solution, 1 ml of Trace elements ($ZnSO_4 \cdot 7H_2O$ 0.5 g/L, $CoCl_2 \cdot 6H_2O$ 0.5 g/L, $(NH_4)Mo_7O_{24} \cdot 4H_2O$ 0.5 g/L, $CuSO_4 \cdot 5H_2O$ 0.5 g/L, H_3BO_4 0.1 g/L, $MnCl_2 \cdot 4H_2O$ 0.5 g/L). The resulting solution was filled up to 980 ml with water and then filter sterilized (Nalgene™ Rapid-Flow™ Sterile Disposable Filter Units with PES Membrane, Thermo Fisher Scientific, USA). Different carbon source stocks were prepared with glucose 20%, galactose 10%, and Xylose 20% and filter sterilized. Before use, each carbon source was added to minimal media with a final concentration of 4 g/L to achieve an equal concentration of carbon atoms in each medium. $FeSO_4 \cdot 7H_2O$ solution was prepared with 5g/L (10000 ×), filter sterilized, frozen in -20°C and added 0.1 mL to 1 L M9 minimal medium before use.

The Amino acid medium (AA) was made by supplementing the medium with glucose with a final concentration of 4 g/L, and the amino acids solution (50 ×), which was purchased from Sigma (R7131), was used to replace NH_4Cl in the M9 medium as the nitrogen. The amino acids solution consisted the following individual amino acids: Arginine 10.0 g/L, Asparagine 2.84 g/L, Aspartic Acid 1.0 g/L, Cystine 2.5 g/L, Glutamic Acid 1.0 g/L, Glycine 0.5 g/L, Histidine 0.75 g/L, Hydroxy-L-Proline 1.0 g/L, Isoleucine 2.5 g/L, Leucine 2.5 g/L, Lysine 2.0 g/L, Methionine 0.75 g/L, Phenylalanine 0.75 g/L, Proline 1.0 g/L, Serine 1.5 g/L, Threonine 1.0 g/L, Tryptophan 0.25 g/L, Tyrosine 1.16 g/L, Valine 1.0 g/L. Other supplement salt mixture in AA was the same as M9 minimal medium. For chemostat growth, 3 g/L glucose in M9 minimal medium was used.

Media list

Media list			Growth conditions and why we choose them
Complex medium	1	LB	The most commonly media used for E. coli cultivation. For comparison with published protein complex to illustrate BP method used in this project is good and credible for E.coli protein complex studies.
	2	Glucose+AA	Amino acids replaced ammonium as nitrogen source.
Minimal Medium (MM) with different carbon sources, NH_4Cl as nitrogen source	3	Glucose	Glucose as carbon in MM, as a standard in all 10 conditions.
	4	Galactose	The slowest growth rate in selected batch culture conditions. Showed different proteome data with glucose when analyze reference (1) data
	5	Xylose	Pentose, different with glucose. Special D-xylose metabolic process, significantly upregulated genes were found (1) Highest growth rate in selected MM. (1.18-fold than glucose). In (1) data, E. coli have almost the same growth rate both in glucose and xylose
Stress conditions on glucose	6	42C	42ºC high temperature stress High temperature makes E. coli fragile. When put E.coli culture from 42C to refrigerator and then put back to 42C, E. coli almost can't grow (death cell precipitation appear). It didn't happen in 37C (37C-4C-37C, grow normal). When E.coli grown in 42C, some flasks grow faster and some flasks grow slower, not uniform. Maybe growth heterogeneity happened in high temperature stress.
	7	Anaerobic	E.coli is facultative anaerobic strain. In chemostat anaerobic condition, plan to design growth rate/dilution rate D=0.25 h^{-1} We also have "E. coli grown in chemostat aerobic condition with the same growth rate" sample, can do proteome comparison if necessary.

fixed growth rate on glucose	8	Chemostatμ= 0.8	D=0.8 h^{-1}, the max growth rate in chemostat. In (1) D=0.12 h^{-1}, 0.2 h^{-1}, 0.35 h^{-1}, 0.5 h^{-1} were chosen. Chemostat is different with batch, and commonly used in bacterial cultivation to get high growth rate and high density. Studying E.coli grown in chemostat with max growth rate can help people better understand E. coli growth process and guild people to use E.coli for biotechnology applications.
	9	Chemostat μ=0 (near-zero growth, NZG)	NZG in chemostat continuous culture vessel, simulate E. coli grown in nature environment.
starved cells	10	Stationary 1 day	Starvation condition, different with exponential phase.

Cultivation

For the preculture, a single colony was picked from the LB plate and grown overnight in 50 ml LB medium in a 250-ml Erlenmeyer flask at 37°C, 200 rpm. For the batch cultures, the cells from a preculture were washed twice with sterilized ice-cold phosphate-buffered saline (PBS) and re-inoculated into 100 ml of the appropriate medium in a 500-ml Erlenmeyer flask and grown at 37°C, 200 rpm. The cells were first grown to exponential phase and then transferred into a second shake-flask containing fresh medium under the respective condition and growing to early exponential phase. The cells undergoing temperature stress were grown at 42°C.

A BIOFLO 2000 bioreactor (New Brunswick Sci., USA) was used for batch and chemostat cultivation under a biocontroller of temperature (37°C), pH 7, airflow, pO2 and stirring. The stirring rate varied from 200 to 1200 rpm to keep pO2 above 50% of air saturation. An infrared analyzer LI-800 (LI-COR Biosciences, Lincoln, NE, USA) was used to measure off-gas CO_2.

For cell cultivation, frozen glycerol stocks were inoculated into 5 mL LB medium and grown overnight. The precultures were 1:10 diluted with fresh glucose M9 medium, and allowed to grow in the batch mode to a specific OD before continuous operation initiated. Then cultures were stabilized in chemostat mode at a dilution rate (D) (μ_{min}) =0.00097 h^{-1} (near-zero growth)

until achieving the steady-state. After sample collection, a continuous increase of *D* started until cells could not keep up with the rising *D* (resulting in culture washout) and achieved the maximal specific growth rate (μ_{max}) at ≈ 0.8 h^{-1}.

Starved cells were continuously grown after reaching stationary phase for 1 day.

Protein samples extraction

When cells grow to OD_{600}=0.2~0.3, cells under the respective condition were collected by centrifugation at 3197g at 4°C for 20 min, washed twice with ice-cold PBS buffer, harvested by centrifugation at 10000g and the cell pellet was stored at -80°C until further processing.

For all batch cultures (totally six culture conditions), each culture condition generated three independent culture cells that were subjected to three independent protein samples extraction. The cultured cells under each chemostat culture condition (totally four) had only one biological replicate and each was used to generate protein samples extraction three times.

Cofractionation samples preparation

BP (Biochemical purification) with Ion Exchange Chromatography (HPLC-IEX)

Per 100 μg frozen cells were resuspended in 0.5 ml fresh protein extraction buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.2 mM EDTA, 10% glycerol, protease inhibitor (PI, Roche, Cat. No. 04693159001), 0.25 mM TCEP), and lightly disrupted by sonication on ice. The soluble proteins were obtained by Centrifugation at 15,000 × g for 10 min at 4 °C, a small aliquot of the supernatant was taken to determine proteins concentration using a BCA assay.

An Agilent 1260 Infinity II was equipped with columns PolyWAX LP 204WX0510 (200×4.6mm i.d., 5 μm, 1000-Å) and PolyCAT A 204CT0510 (200×4.6 mm i.d., 5 μm, 1000-Å) (PolyLC INC, MD, USA). Before sample injection, the columns were balanced with 10 bed volumes of buffer A (0.75 mM AmAc). About ~2.5 mg of fresh soluble protein sample was injected to the IEC column and separated by a linear gradient of 4-30% buffer B (2.5 M AmAc) for 90 min and 30-60% buffer B for 30 min, with a flow rate of 0.4 ml/min. Totally 96 fractions were collected using a 96-deep well plate (Thermo Scientific™ Abgene™ AB0564) with 1 min intervals.

Fractionated proteins digestion and peptides desalting

Take a small aliquot (25 µl) of proteins from each fraction of the 96-deep well collection plate, to a new 96-well plate to measure protein concentration using BCA assay in a Bio-tek microplate reader. The rest of proteins of each fraction were dried by a Savant SpeedVac SC210A (Thermo Fisher Scientific, USA), resuspended in 100 µL buffer (8.5 M urea-100 mM Tris-HCl, pH 8.0, 5 mM TCEP) reduced at 37°C for 60 min, alkylated with 15 mM IAA for 60 min, diluted with 50 mM Tris-HCl with urea<1 M, digested by incubation with sequencing-grade modified trypsin (1/20~1/100, w/w) overnight at 37 °C. After digestion, the reaction was stopped by adding FA with a final concentration of 1%. The peptides were desalted by a Sep-Pak tC18 96-well µElution plate (Waters, USA, Product Number 186002318). Before loading samples, the desalting plate was wetted with 0.5 mL methanol twice and washed with 0.5 mL 0.1% FA twice. Peptide fractions were loaded on the plate by centrifugation for 1 min at 100 g. Based on the amount of proteins from each fraction measured with a BCA assay previously, the total peptides for each fraction loaded to the desalting plate should not exceed 1% of the sorbent weight (10 mg). The desalting plate was washed twice with 0.5 mL of 0.1% FA to remove the unbinding materials and eluted with 150 µL 0.1% FA/60% ACN twice, the elution samples were collected with PlateOne™ 96-well 0.5 ml polypropylene plate (USA Scientific, USA, Product number 1896-5110). The fractioned desalted samples in the plate were divided into three low-profile 96-Well PCR Plates (Bio-rad, USA, Product number HSP9601) and dried by Speedvac. Dried peptides in one plate were used for further TMT-labeling, and the other two plates were stored at -80°C as backups.

TMT labeling

We used 20 µl of 50 mM HEPES buffer to resuspend each fraction of peptides in a plate, then transfer 4 µg peptides per fraction into a new low-profile 96-Well PCR Plate and adjust to a final volume of 20 µL with 50mM HEPES buffer. Note that for fractions 1-4, each with a total protein digest of 1 µg, these peptides were all transferred into the new 96-Well PCR Plate. Using this way, cells from 10 culture conditions generated 30 of 96-well plates of peptide samples.

For the 5mg TMT label reagent vials (TMT 10plex Isobaric Label Reagent Set plus TMT11-131C Label Reagent, Thermo Fisher Scientific, catalog number A34808), add 250 µl of ACN to each tube to make the reagents concentration of 20 µg/µl, took half (125 µl) of the reagents to new tubes and diluted to 2 µg/µl with ACN.

When doing TMT labeling, for the 10 plates in a group, took 2 µl from the same fraction of each plate and put them into a new plate. The 11th plate was a mixture of the 10 plates and used

for normalization. Added 10 µl of 2 µg/µl TMT reagent to each fraction of peptides to make the TMT/sample=5:1 (v/v) for labeling. The mixtures were incubated for 1 hour at room temperature. After the reaction, added 2 µl 5% hydroxylamine incubating at room temperature for 15 min to quench the reaction. The labeled peptides in the 11 plates were pooled into a 96-well plate and desalted with a Sep-Pak tC18 96-well µElution plate. The desalted peptides were eluted into a PlateOne^{TM} 96-well 0.5 ml polypropylene plate and divided to three new low-profile 96-Well PCR Plate and dried by SpeedVac. These three plates were stored in -80°C, one waiting for mass spec analysis and the others stored as backups.

LC-MS/MS

The dried peptides were resuspended in 100 µl of solvent A and 15 µl of each sample was taken and loaded on an EASY nLC 1200 system coupled to a Q Exactive HF mass spectrometer equipped with an EASY-Spray ion source (all from Thermo Fisher Scientific, USA). The peptides mixtures were separated by a C18 Acclaim PepMap 100 pre-column (75 µm i.d.×2 cm, 3 µm, 100 Å) hyphenated to a PepMap RSLC C18 analytical column (75 µm i.d.×50 cm, 2 µm, 100 Å) (all from Thermo Fisher Scientific, USA). Each fractionated sample was eluted from the column with a 120-min gradient.

Protein identification with MaxQuant

The MS/MS raw files were searched using MaxQuant Version 1.6.0.16 against the *E. coli* database (Uniprot, download data:2018/06/25). The database consists of 4, 313 *E. coli* proteins as well as known contaminants. Reporter ion MS2 was used for quantification with 10plex TMT and a reporter mass tolerance of 0.003 Da. Peptide search tolerance was set to 4.5ppm for MS1, and MS2 fragment tolerance was set to 10ppm. Match between runs was active with an alignment window of 20 min and a match window of 0.7 min. Other MaxQuant parameters were performed by default.

CF/MS Data Preprocessing

We sought to develop a pre-processing workflow with the optimal set of steps and parameters for our data that would lead to the best recapture of literature-reported protein

interactions. Starting from the peptide-level data, the following four steps represent the pre-processing steps in our final pipeline:

(i) Normalization. To correct for the different number of peptides identified in each of the 96 fractions (850 - 5940 peptides per fraction), MS intensity values were normalized within each fraction, converting the raw intensity of each peptide to its proportion of all peptides within that fraction, followed by log2-transformation.

(ii) Smoothing signals. Each peptide's 96-fraction conditional profile is smoothed by taking a moving average of 4 fractions to smooth out the short-term fluctuations across neighboring fractions.

(iii) Filter outlier sibling peptides. On average, each protein in this dataset has 14 peptides mapped to it (IQR 5-19). It is expected that the peptides that map to the same protein should have similar profiles to each other, and as such any peptide that deviates significantly from its group is likely a faulty measurement or incorrect mapping. To filter out these outlier peptides, we performed average-linkage hierarchical clustering on the sibling peptides based on their similarity to each other, split the resulting dendrogram into two clusters, and retained the peptides belonging to the larger cluster as being representative of the protein.

(iv) Construct protein profiles. Finally, we collapsed the sibling peptide profiles into their corresponding protein profiles by averaging their per-fraction intensities, leading to a final set of 1,937 proteins for downstream analysis.

In developing this pipeline, we benchmarked different strategies for each step as follows:

1) Generate protein profiles from peptide profiles using different sets of processing parameters.

2) Compute distances between resultant protein profiles.

3) Compare distances to literature-curated PPIs using a ROC analysis.

For step 2, the distances between protein profiles were computed using three metrics: Pearson distance, Euclidean distance, and Wasserstein distance, with Pearson distances eventually leading to the best performance. This dataset has 2 replicates, and as such 4 distances were

computed per protein pair, with the average of these 4 distances taken as the representative distance between the two proteins. This distance was then converted into a signal-to-noise ratio by dividing the distance by the average cross-replicate distance, i.e. noise, of the given two proteins.

For step 3, the following datasets were used as reference protein interactions to benchmark against:

1) A set of one million pairwise interactions from the STRING database v11[16]. STRING is a repository of interactions compiled from seven different sources. The dataset was filtered for high-confidence interactions only (STRING score > 0.7). Only pairs for which elution data is present in our data for both members were retained for further analysis, leading to a dataset of 24,912 interactions among 1,821 unique proteins.

2) A set of 184,023 interactions from the BioGRID database[25]. BioGRID is a literature-curated database of genetic and protein interaction data. Only pairs for which elution data is present for both members in our data were retained for further analysis, leading to a dataset of 36,204 interactions among 1,690 unique proteins.

For each set of pre-processing parameters and each reference dataset, the benchmarking ROC analysis was carried out as such:

1) Compute average distance between each pair of proteins.

2) Sort the protein pairs in ascending order of distances.

3) Label each pair as true or false depending on presence within the given reference network.

4) Compute the area under the ROC curve using the pairwise distances as weights.

The pipeline that led to the highest AUC score in this analysis was selected (Fig. 2A).

Predicting E. coli interactome to probe for remodeling

The *EPIC* software[17] was employed to predict protein interactions in the conditional CF/MS data, representing the E. coli interactome that we later searched for evidence of remodeling. The software was run on the pooled CF/MS data across conditions and replicates. The default PPI score cutoff of 0.5 was applied. EPIC used the following metrics for determining co-

elution: mutual information, Euclidean distance, Jaccard index, Pearson correlation, Pearson correlation with Poisson noise coefficient, apex score, and a novel Bayes correlation.

This network contains 6,152 interactions among 1,866 proteins. The average degree is 6.6 (IQR 4-7). EPIC detected 267 complexes among this interactome with an average size of 9.2 proteins (IQR 6-11). A total of 1,806 proteins (96.8% of all proteins in the interactome) were determined to be members of complexes in this interactome. 68.3% of these proteins were a member of just one complex while at the other extreme there are 9 proteins that are members of 4 complexes. 83.6% of edges in the interactome were within complexes as opposed to across complexes.

Computing conditional similarities

We used two different metrics to score the similarity of each pair of protein profiles in each of the growth conditions (Fig. 2B). Each of these metrics captures a different mathematical property of the CF/MS profiles which are potentially informative of different mechanisms of remodeling. To recap, a protein's CF/MS profile in a given condition is its relative abundance in each of the 96 consecutive fractions. The two similarity metrics are:

1) Co-elution. This corresponds to the similarity of the 'shape' of the two protein CF/MS profiles, i.e. the patterns of relative abundance across the fractions, and commonly used for inferring protein interactions from CF/MS data. Computed as the Pearson correlation between the two CF/MS profiles and is independent of the magnitude of the protein's relative abundances. Finally, each protein pair's correlation within each condition was averaged across the two replicates.

2) Co-abundance. This corresponds to the ratio between the two proteins' total relative abundance in each condition. Computed as intensity fold-changes between each pair of proteins in each of the 10 conditions as follows. We first reversed the log2-transformation of the CF/MS profiles. We then summed each protein's MS2 intensity values across fractions. Since protein-protein interactions are inherently undirected in CF/MS methods, the fold-change for a given protein pair was consistently computed as the ratio between the lower summed abundance to the higher one. These ratios were then log2-transformed. Finally, each protein pair's fold-change within each condition was averaged across the two replicates.

We then combined these two scores into one similarity score as follows. We created a table with the combined coelution and coabundance scores for all protein pairs in our predicted E. coli interactome across all conditions. Each row of this table corresponded to one protein-protein interaction (PPI) in one condition, and the table has one column with the corresponding co-elution score and one with the co-abundance score. We then took the first principal component of the PCA decomposition of this table to represent our final PPI similarity scores. This principal component represented 62% of the data variance. This PCA method was used since it computes a score that captures the information in both similarity metrics without being affected by their correlation or the differences in their scales.

Scoring interactome remodeling

We selected the commonly-used LB growth media as a baseline reference condition to compare the other growth conditions against for evidence of protein interaction remodeling. For each interaction in our E. coli interactome, we computed conditional remodeling scores as the difference between its similarity score in the LB condition and its similarity scores in the other conditions. As such, each interaction has one remodeling score per condition. To ensure the remodeling score is higher for PPIs that are considered weakened/disrupted ('more remodeled') and lower for those that are considered strengthened/conserved ('less remodeled'), the remodeling score is always computed as score(LB) - score(condition). This way, the interacting protein profiles that become less similar will have a positive remodeling score and those that become more similar will have a negative remodeling score, while those that remain unchanged compared to the LB condition will have a score of zero.

These scores were summarized at the level of the 267 protein complexes detected in the predicted interactome by averaging the remodeling scores of each complex's intra-complex interactions, where an intra-complex interaction is defined as an interaction between two complex members that was detected in the predicted interactome. The same strategy was used to compute remodeling scores for the E. coli pathways from the EcoCyc database[22] that were examined for remodeling (see section titled 'Compiling E. coli pathways'). Pathways were classified as 'weakened', 'strengthened', 'stable', and 'dynamic' based on whether their minimum cross-condition remodeling score was below 1 standard deviation of all remodeling scores, their maximum score greater than 1 standard variation of all scores, if all the conditional scores were within 1 standard deviation of all pathway scores, or if the maximum and minimum scores both exceeded 1 standard deviation in their corresponding directions, respectively. Finally, the scores were also summarized at the level of the 1,866 individual proteins present in the predicted interactome by averaging the remodeling scores of all the interactions that each protein is involved in.

Compiling E.coli pathways

Information on 445 E. coli pathways including 1,170 protein-coding genes was downloaded from the EcoCyc database[22]. 397 pathways containing at least one protein in our dataset were initially retained prior to downstream interrogation of interactome remodeling.

585 Finally, we quantified the interactome remodeling among the 206 pathways that had at least one
586 intra-pathway interaction in our predicted interactome.
587
588 Pathway enrichment analysis
589 We used the hypeR R package[26] to perform hypergeometric gene set enrichment tests to
590 detect significantly-enriched pathways (FDR < 0.01) among the proteins of interest in our dataset,
591 using the compiled pathways from EcoCyc (see section titled 'Compiling E. coli pathways') as the
592 background gene sets.
593
594 Compiling E.coli protein properties
595 Information on the evolutionary age of 4,140 Ecoli proteins was downloaded from the
596 GenOrigin database[27]. This included the age of 1,785 proteins in our dataset, or 95.7% of all
597 quantified proteins. Protein relative abundance was computed from this dataset by summing the
598 MS2 intensity of each protein across all CF/MS fractions in each growth condition. Cellular
599 compartment information for 817 genes in our dataset was downloaded from EcoCyc[22]. We
600 focused our analysis on the three main cellular compartments: cytosol, membrane, and periplasmic
601 space. Phosphorylation evidence for 535 E.coli proteins were downloaded from the dbPSP
602 database[28]. 505 of these proteins were present in our dataset. Evidence for 101 Ecoli proteins with
603 disordered structures, including 89 in our dataset, was downloaded from the DisProt database[29].
604
605 Interactive web application
606 An interactive web application to explore the results was developed using the R Shiny
607 framework. The web application includes visualizations of individual or grouped protein and
608 peptide CF/MS profiles across the 10 growth conditions and two replicates. It also displays
609 putative conditional protein interactions for any protein of the user's choice based on evidence
610 from our dataset and external databases.
611

612 **Tables**

		Growth conditions (Abbr.)
Complex medium	1	LB (LB)
	2	Glucose + Amino acids (AA)
Carbon sources	3	Glucose (Glc)
	4	Galactose (Gal)
	5	Xylose (Xyl)
Stress conditions on glucose	6	Anaerobic (Ana)
	7	42°C (T42C)
Fixed growth rate on glucose	8	chemostat $\mu \approx 0$ (NZG)
	9	chemostat $\mu \approx 0.8$ (Max)
Starved cell	10	Stationary 1 day (SP)

613 **Table 1. List of 10 different growth conditions of *E. coli* BW25113.**

Figure Captions

Figure 1: Profiling E. coli interactome dynamics

1a: Experimental design. Quantitative CF/MS data was generated from E. coli cultured under 10 different media.

1b: E. coli proteome coverage. Fraction of E. coli protein-coding genes quantified in our dataset.

1c: Coverage of largest E. coli pathways in the KEGG database.

1d: Dynamic CF/MS profiles. Heatmaps visualizing the CF/MS data generated for each of the growth media.

1e: Reproducibility of data. Overlap of replicate HPLC data.

1f: Recovery of known E. coli interactions. Comparison of Pearson correlation between pairs of proteins known to interact in the STRING database and all possible pairs of proteins in our dataset.

Figure 2: Computational pipeline for quantifying interactome remodeling from dynamic CF/MS data

2a: CF/MS data processing. Pipeline for processing peptide-level CF/MS data leading to generation of protein-level profiles.

2b: Profiling protein interaction remodeling. Analysis workflow for quantifying interactome remodeling under different conditions.

Figure 3: Global patterns of interactome remodeling

3a: Leftmost network plots visualize protein complexes predicted from the dynamic CF/MS data. Nodes represent protein complexes color-coded by quantitative extent of remodeling in given condition compared to base growth media (LB) and sized according to number of member proteins. Middle network plots visualize example complexes selected due to high level of remodeling and biological significance. Rightmost plots visualize CF/MS profiles of example intra-complex pairwise protein interactions with high remodeling scores.

Figure 4: Biological pathways driving interactome remodeling

4a: Average remodeling scores of main E. coli pathway families. Boxplots represent distribution of each pathway family's remodeling scores in each of the growth media. Pathway families colored according to spread of remodeling scores around base growth medium (LB).

4b: Top remodeled and stable pathways across conditions. Top four strengthened and weakened pathways for selected growth conditions based on remodeling scores.

4c: Pathways enriched among top remodeled proteins. Results of hypergeometric enrichment test for top 5% most disrupted proteins based on averaged remodeling scores.

4d: Pathways enriched among top stable proteins. Results of hypergeometric enrichment test for top 5% most stable proteins based on averaged remodeling scores.

Figure 5: Structural and functional properties of interaction remodeling

5a: Boxplots comparing distribution of co-elution and co-abundance remodeling scores across all interactions.

5b: Relationship between summed protein MS2 intensities and averaged protein remodeling scores. The downwards trend suggests higher remodeling scores for lower abundance proteins.

5c: Confidence intervals of averaged protein remodeling scores for two categories of proteins based on evolutionary age using information from the GenOrigin database. Some types of environmental changes favored the stability of ancient proteins while others the modern proteins.

5d: Confidence intervals of averaged protein remodeling scores for proteins from different cellular compartments. Membrane proteins tended to be the most stable across conditions.

5e: Confidence intervals of averaged protein remodeling scores for proteins with and without evidence of phosphorylation in the dbPSP database. Phosporylated proteins were marginally more stable across conditions.

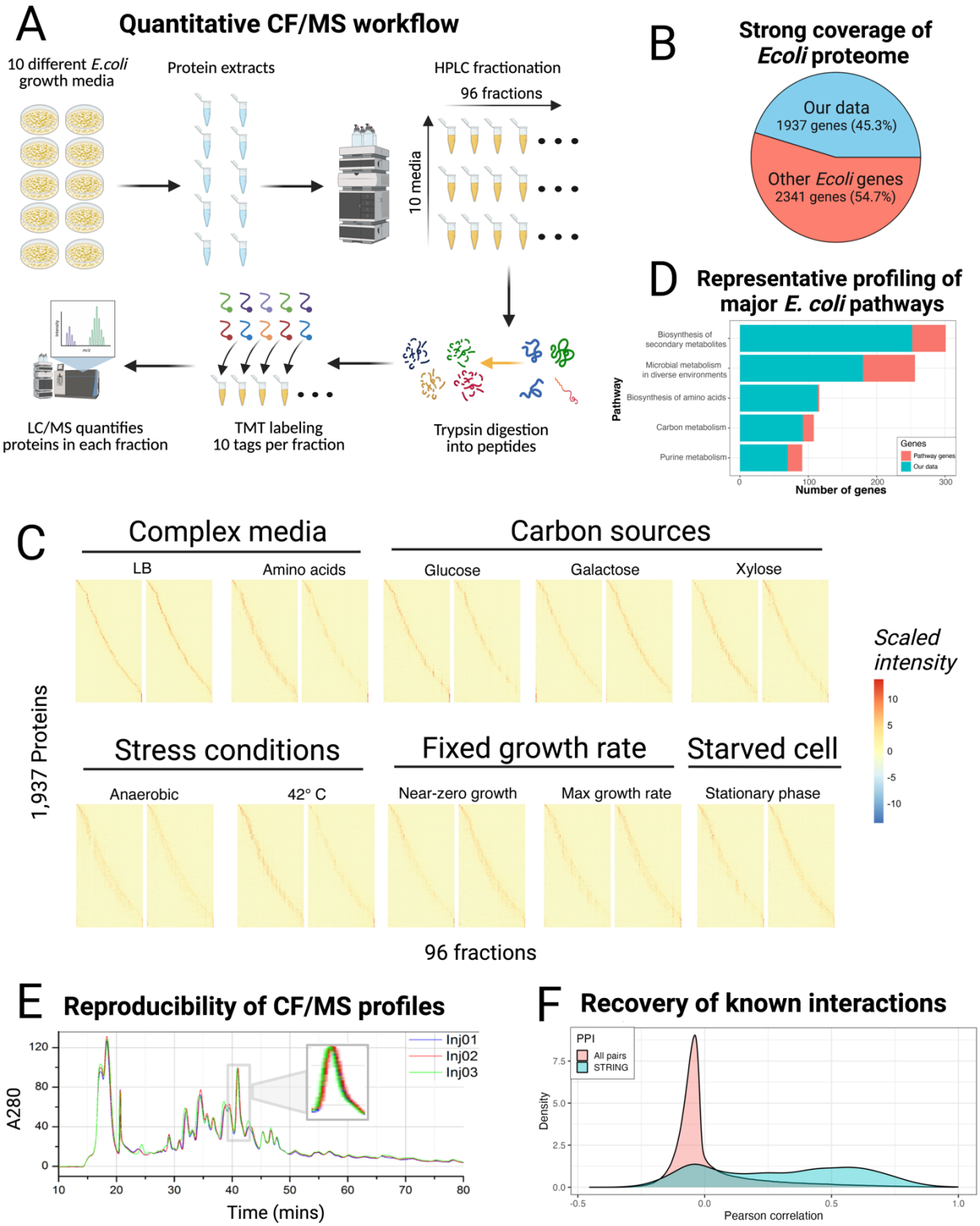
5f: Confidence intervals of averaged protein remodeling scores for proteins with and without evidence of intrinsically disordered structures in the DisProt database. No dominant pattern was observed among intrinsically disordered proteins.

5g: Confidence intervals of averaged protein remodeling scores based on number of interactions. Hub proteins with higher number of interactions were more stable across conditions.

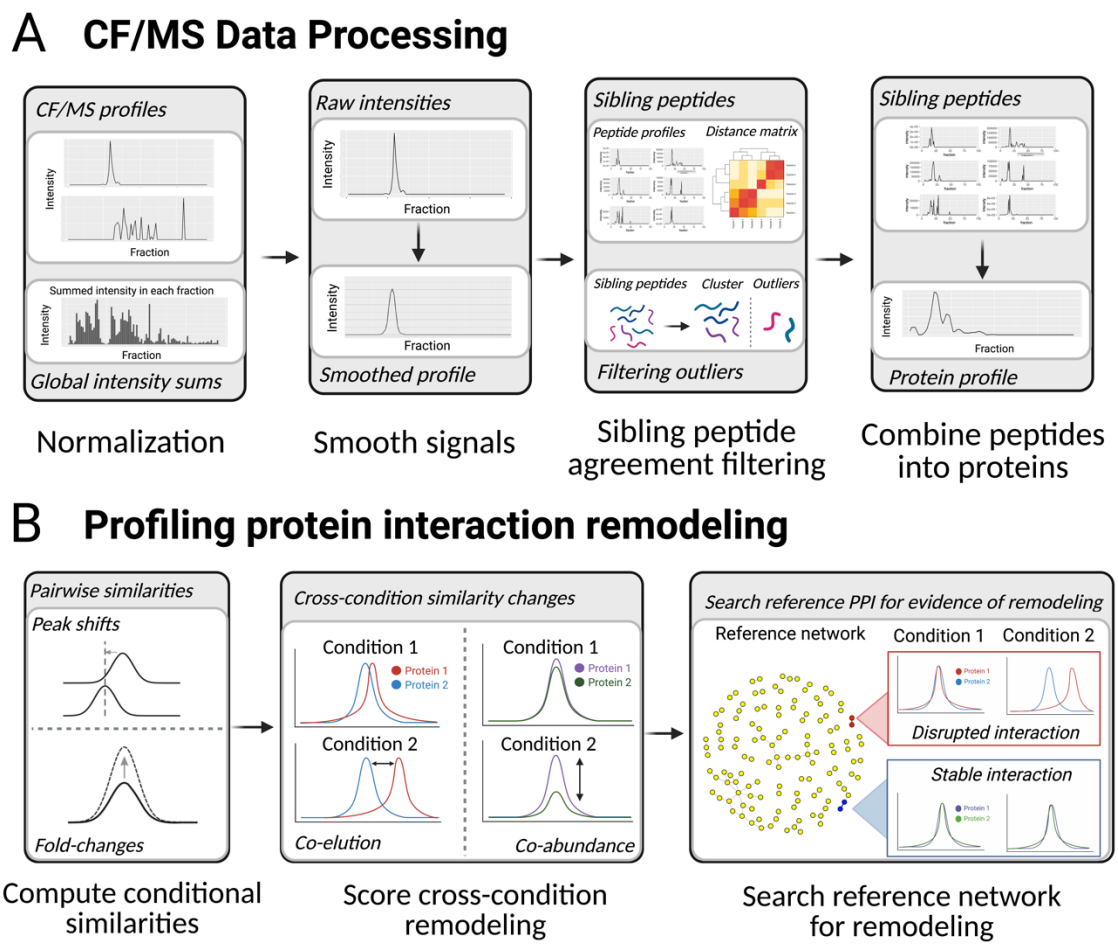
5h: Confidence intervals of averaged protein remodeling scores for proteins based on number of complexes. Despite the higher stability of hub proteins, there was no strong association between protein complex membership and remodeling.

5i: Relationship between edge betweenness score of interactions relative to the interactome structure and remodeling scores. Consistently near-zero correlation suggests that central interactions exhibit the same amount of remodeling as periphery ones.

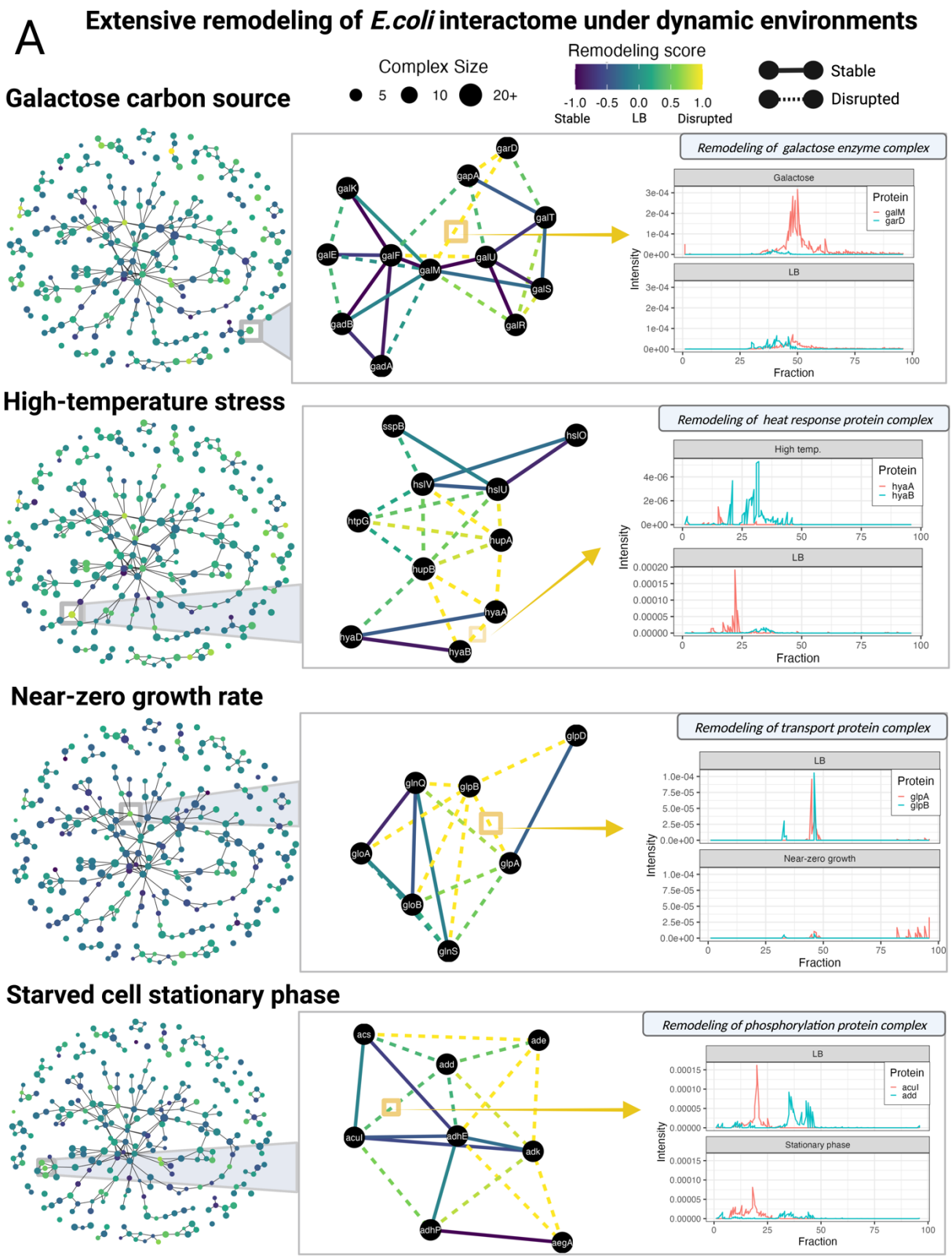
5j: Relationship between protein complex density and averaged protein remodeling scores. The downwards trend suggests tightly-connected complexes are less prone to remodeling.


A
Quantitative CF/MS workflow
10 different E.coli growth media
Protein extracts
HPLC fractionation
96 fractions
10 media
LC/MS quantifies proteins in each fraction
TMT labeling 10 tags per fraction
Trypsin digestion into peptides
B
Strong coverage of Ecoli proteome
Our data
1937 genes (45.3%)
Other Ecoli genes
2341 genes (54.7%)
D
Representative profiling of major E. coli pathways
Biosynthesis of secondary metabolites
Microbial metabolism in diverse environments
Biosynthesis of amino acids
Carbon metabolism
Purine metabolism
Pathway
Genes
Pathway genes
Our data
0
100
200
300
Number of genes
C
Complex media
LB
Amino acids
Carbon sources
Glucose
Galactose
Xylose
Stress conditions
Anaerobic
42° C
Fixed growth rate
Near-zero growth
Max growth rate
Starved cell
Stationary phase
1,937 Proteins
Scaled intensity
10
5
0
-5
-10
96 fractions
E
Reproducibility of CF/MS profiles
A280
Inj01
Inj02
Inj03
Time (mins)
F
Recovery of known interactions
PPI
All pairs
STRING
Density
Pearson correlation

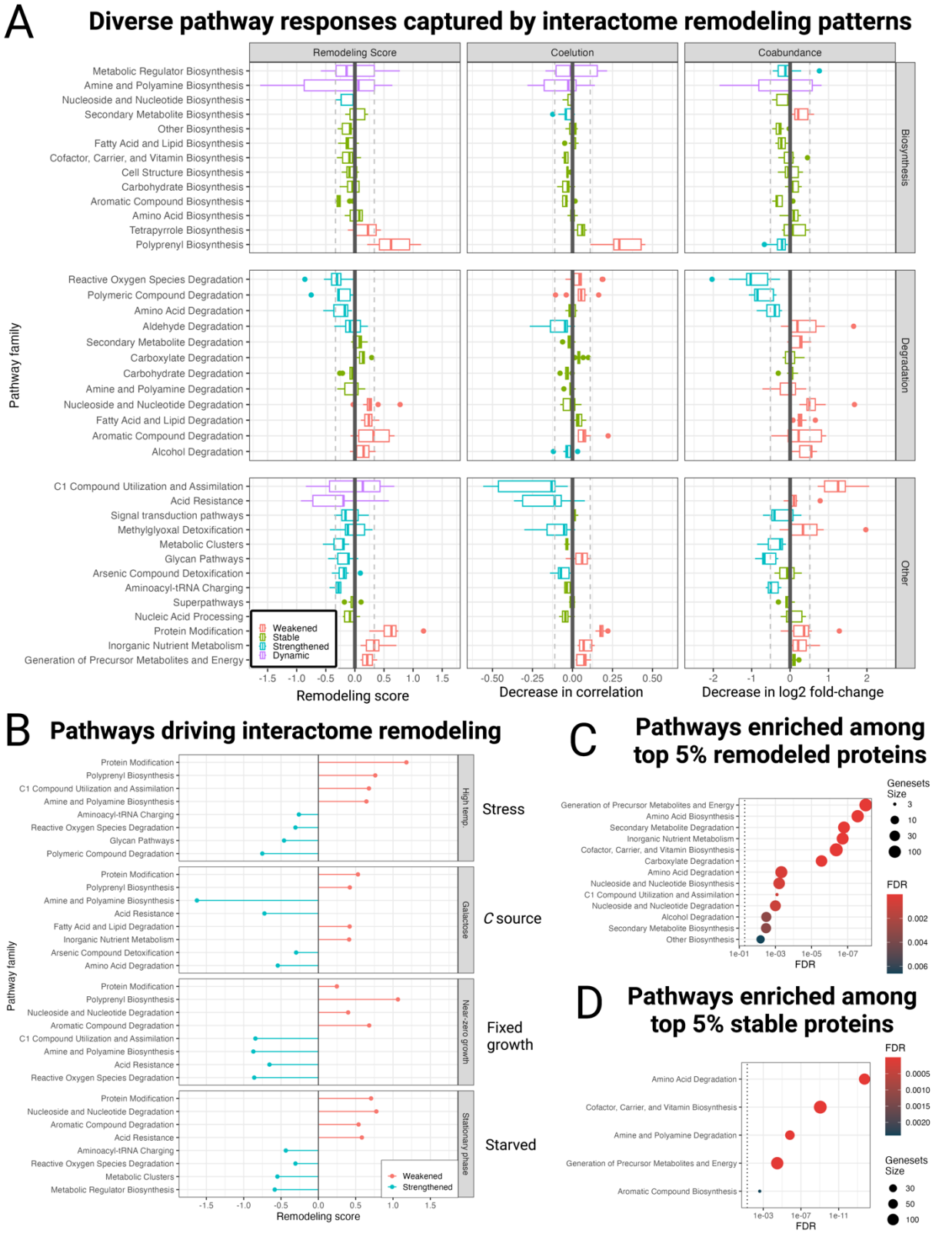
674

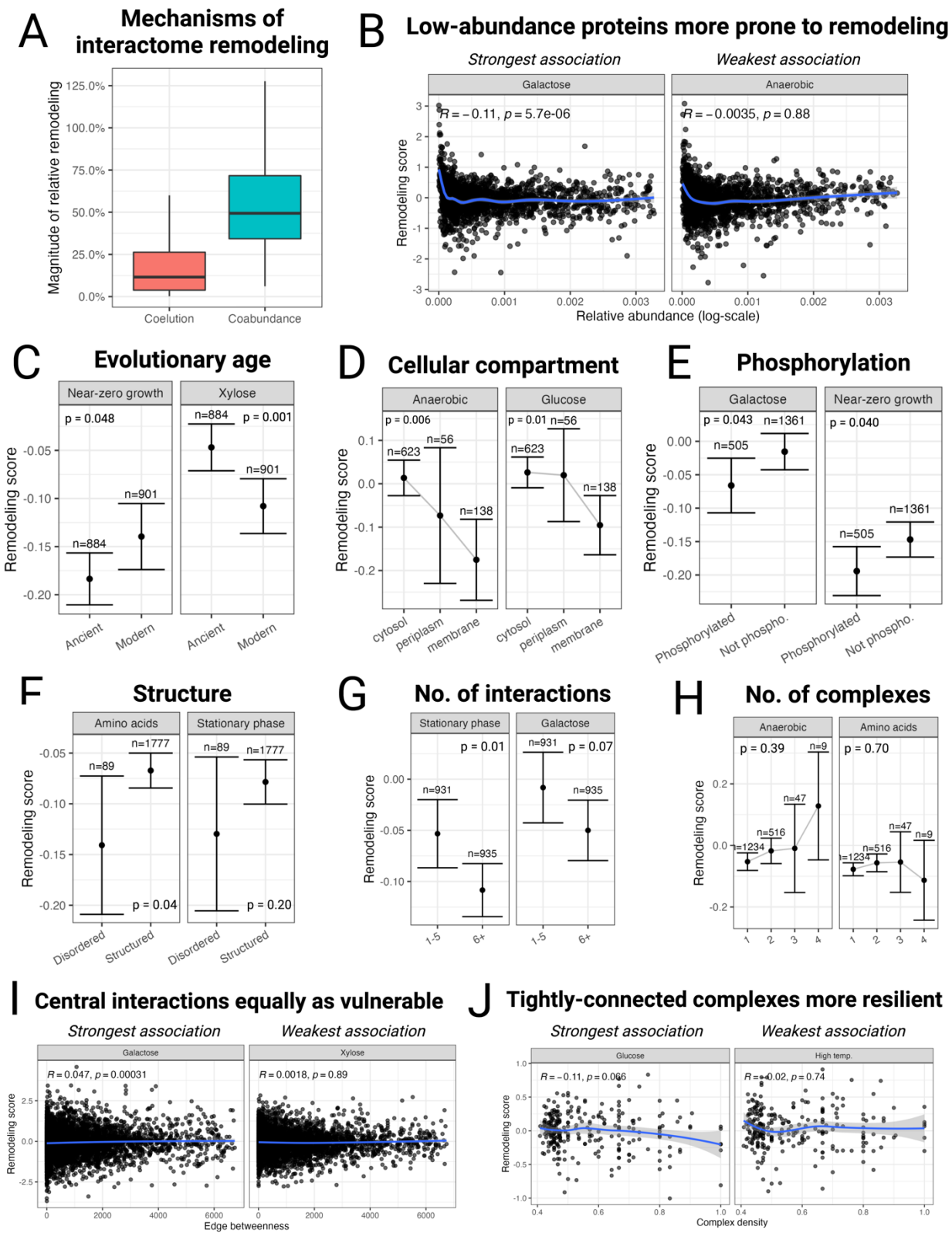

A CF/MS Data Processing
CF/MS profiles
Intensity
Fraction
Summed intensity in each fraction
Intensity
Fraction
Global intensity sums
Raw intensities
Intensity
Fraction
Intensity
Fraction
Smoothed profile
Sibling peptides
Peptide profiles
Distance matrix
Sibling peptides
Cluster
Outliers
Filtering outliers
Sibling peptides
Intensity
Fraction
Protein profile
Normalization
Smooth signals
Sibling peptide agreement filtering
Combine peptides into proteins
B Profiling protein interaction remodeling
Pairwise similarities
Peak shifts
Fold-changes
Cross-condition similarity changes
Condition 1
Protein 1
Protein 2
Condition 1
Protein 1
Protein 2
Condition 2
Condition 2
Co-elution
Co-abundance
Search reference PPI for evidence of remodeling
Reference network
Condition 1
Condition 2
Disrupted interaction
Stable interaction
Compute conditional similarities
Score cross-condition remodeling
Search reference network for remodeling

675


A
Extensive remodeling of E.coli interactome under dynamic environments
Complex Size
5
10
20+
Remodeling score
-1.0 -0.5 0.0 0.5 1.0
Stable
LB
Disrupted
Stable
Disrupted
Galactose carbon source
Remodeling of galactose enzyme complex
garD
gapA
galK
galT
galE
galF
galM
galU
galS
gadB
galR
gadA
Galactose
LB
Protein
galM
garD
Intensity
Fraction
High-temperature stress
Remodeling of heat response protein complex
sspB
hslO
hslV
hslU
htpG
hupA
hupB
hyaA
hyaD
hyaB
High temp.
LB
Protein
hyaA
hyaB
Intensity
Fraction
Near-zero growth rate
Remodeling of transport protein complex
glpD
glnQ
glpB
gloA
glpA
gloB
glnS
LB
Near-zero growth
Protein
glpA
glpB
Intensity
Fraction
Starved cell stationary phase
Remodeling of phosphorylation protein complex
acs
ade
add
acuI
adhE
adk
adhP
aegA
LB
Stationary phase
Protein
acuI
add
Intensity
Fraction

676
677


A
Diverse pathway responses captured by interactome remodeling patterns
Remodeling Score
Coelution
Coabundance
Biosynthesis
Degradation
Other
Pathway family
Metabolic Regulator Biosynthesis
Amine and Polyamine Biosynthesis
Nucleoside and Nucleotide Biosynthesis
Secondary Metabolite Biosynthesis
Other Biosynthesis
Fatty Acid and Lipid Biosynthesis
Cofactor, Carrier, and Vitamin Biosynthesis
Cell Structure Biosynthesis
Carbohydrate Biosynthesis
Aromatic Compound Biosynthesis
Amino Acid Biosynthesis
Tetrapyrrole Biosynthesis
Polyprenyl Biosynthesis
Reactive Oxygen Species Degradation
Polymeric Compound Degradation
Amino Acid Degradation
Aldehyde Degradation
Secondary Metabolite Degradation
Carboxylate Degradation
Carbohydrate Degradation
Amine and Polyamine Degradation
Nucleoside and Nucleotide Degradation
Fatty Acid and Lipid Degradation
Aromatic Compound Degradation
Alcohol Degradation
C1 Compound Utilization and Assimilation
Acid Resistance
Signal transduction pathways
Methylglyoxal Detoxification
Metabolic Clusters
Glycan Pathways
Arsenic Compound Detoxification
Aminoacyl-tRNA Charging
Superpathways
Nucleic Acid Processing
Protein Modification
Inorganic Nutrient Metabolism
Generation of Precursor Metabolites and Energy
Weakened
Stable
Strengthened
Dynamic
-1.5 -1.0 -0.5 0.0 0.5 1.0 1.5
Remodeling score
-0.50 -0.25 0.00 0.25 0.50
Decrease in correlation
-2 -1 0 1 2
Decrease in log2 fold-change
B
Pathways driving interactome remodeling
High temp.
Galactose
Near-zero growth
Stationary phase
Stress
C source
Fixed growth
Starved
Protein Modification
Polyprenyl Biosynthesis
C1 Compound Utilization and Assimilation
Amine and Polyamine Biosynthesis
Aminoacyl-tRNA Charging
Reactive Oxygen Species Degradation
Glycan Pathways
Polymeric Compound Degradation
Acid Resistance
Fatty Acid and Lipid Degradation
Inorganic Nutrient Metabolism
Arsenic Compound Detoxification
Amino Acid Degradation
Nucleoside and Nucleotide Degradation
Aromatic Compound Degradation
Metabolic Clusters
Metabolic Regulator Biosynthesis
Weakened
Strengthened
-1.5 -1.0 -0.5 0.0 0.5 1.0 1.5
Remodeling score
C
Pathways enriched among top 5% remodeled proteins
Genesets Size
3
10
30
100
FDR
0.002
0.004
0.006
1e-01 1e-03 1e-05 1e-07
FDR
D
Pathways enriched among top 5% stable proteins
FDR
0.0005
0.0010
0.0015
0.0020
Genesets Size
30
50
100
1e-03 1e-07 1e-11
FDR


A
Mechanisms of interactome remodeling
Magnitude of relative remodeling
125.0%
100.0%
75.0%
50.0%
25.0%
0.0%
Coelution
Coabundance
B
Low-abundance proteins more prone to remodeling
Strongest association
Weakest association
Galactose
Anaerobic
R = −0.11, p = 5.7e-06
R = −0.0035, p = 0.88
Remodeling score
Relative abundance (log-scale)
C
Evolutionary age
Near-zero growth
Xylose
p = 0.048
p = 0.001
n=884
n=901
n=884
n=901
Remodeling score
Ancient
Modern
Ancient
Modern
D
Cellular compartment
Anaerobic
Glucose
p = 0.006
p = 0.01
n=623
n=56
n=138
n=623
n=56
n=138
Remodeling score
cytosol
periplasm
membrane
cytosol
periplasm
membrane
E
Phosphorylation
Galactose
Near-zero growth
p = 0.043
n=1361
n=505
p = 0.040
n=505
n=1361
Remodeling score
Phosphorylated
Not phospho.
Phosphorylated
Not phospho.
F
Structure
Amino acids
Stationary phase
n=89
n=1777
n=89
n=1777
p = 0.04
p = 0.20
Remodeling score
Disordered
Structured
Disordered
Structured
G
No. of interactions
Stationary phase
Galactose
p = 0.01
p = 0.07
n=931
n=935
n=931
n=935
Remodeling score
1-5
6+
1-5
6+
H
No. of complexes
Anaerobic
Amino acids
p = 0.39
p = 0.70
n=9
n=47
n=516
n=1234
n=47
n=9
n=516
n=1234
Remodeling score
I
Central interactions equally as vulnerable
Strongest association
Weakest association
Galactose
Xylose
R = 0.047, p = 0.00031
R = 0.0018, p = 0.89
Remodeling score
Edge betweenness
J
Tightly-connected complexes more resilient
Strongest association
Weakest association
Glucose
High temp.
R = −0.11, p = 0.066
R = −0.02, p = 0.74
Remodeling score
Complex density

680

Associated data

The raw data has been deposited in the PRIDE database with the identifier PXD041263. The scripts to perform the analysis can be found at https://github.com/AhmedYoussef95/Ecoli-dynamic-interactome. An interactive web application to explore the results was developed and is available at https://bnfweb.bu.edu/EcoliDynamicInteractome/.

Acknowledgements

AE and MC acknowledge a generous pilot award from the Rafik B. Hariri Institute for Computing and Computational Science & Engineering at Boston University. The authors thank Pierre Havugimana for assistance with training and generating the data. The authors thank Alexander Ivanov in Northeastern University for providing lab space for bioreactors. We acknowledge the advice and valuable discussions with Dr. Josh Campbell, Dr. Stefano Monti, and Dr. Trevor Siggers at Boston University. We acknowledge the generous support of the Boston University Graduate Program in Bioinformatics. We acknowledge Mary Ellen Gipson-Fitzpatrick at Boston University for her support with publishing our interactive web application online. We acknowledge Carl White for his assistance uploading the raw data to the PRIDE repository.

Author contributions

AE, FB, AY, and MC conceived the project. FB and PH designed and conducted the experiments. AY and MC designed and performed the computational analysis. NP provided samples of the mass spectrometer cells. AY and FB wrote the manuscript. AE and MC supervised the project.

Conflict of interest statement

The authors have no conflicts of interest to disclose.

References

[1] Babu, M., Bundalovic-Torma, C., Calmettes, C., Phanse, S., Zhang, Q., Jiang, Y., Minic, Z., Kim, S., Mehla, J., Gagarinova, A., Rodionova, I., Kumar, A., Guo, H., Kagan, O., Pogoutse, O., Aoki, H., Deineko, V., Caufield, J. H., Holtzapple, E., … Emili, A. (2018). Global landscape of cell envelope protein complexes in Escherichia coli. *Nature Biotechnology*, *36*(1), 103–112. https://doi.org/10.1038/nbt.4024

[2] Bremer, H., & Dennis, P. P. (2008). Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates. *EcoSal Plus*, *3*(1). https://doi.org/10.1128/ecosal.5.2.3

[3] Butland, G., Peregrín-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V., Starostine, A., Richards, D., Beattie, B., Krogan, N., Davey, M., Parkinson, J., Greenblatt, J., & Emili, A. (2005). Interaction network containing conserved and essential protein complexes in Escherichia coli. *Nature*, *433*(7025), 531–537. https://doi.org/10.1038/nature03239

[4] Caglar, M. U., Houser, J. R., Barnhart, C. S., Boutz, D. R., Carroll, S. M., Dasgupta, A., Lenoir, W. F., Smith, B. L., Sridhara, V., Sydykova, D. K., Vander Wood, D., Marx, C. J., Marcotte, E. M., Barrick, J. E., & Wilke, C. O. (2017). The E. coli molecular phenotype under different growth conditions. *Scientific Reports*, *7*, 45303. https://doi.org/10.1038/srep45303

[5] Celaj, A., Schlecht, U., Smith, J. D., Xu, W., Suresh, S., Miranda, M., Aparicio, A. M., Proctor, M., Davis, R. W., Roth, F. P., & St Onge, R. P. (2017). Quantitative analysis of protein interaction network dynamics in yeast. *Molecular Systems Biology*, *13*(7), 934. https://doi.org/10.15252/msb.20177532

[6] Chong, Y. T., Koh, J. L. Y., Friesen, H., Duffy, S. K., Cox, M. J., Moses, A., Moffat, J., Boone, C., & Andrews, B. J. (2015). Yeast Proteome Dynamics from Single Cell Imaging and Automated Analysis. *Cell*, *161*(6), 1413–1424. https://doi.org/10.1016/j.cell.2015.04.051

[7] Havugimana, P. C., Goel, R. K., Phanse, S., Youssef, A., Padhorny, D., Kotelnikov, S., Kozakov, D., & Emili, A. (2022). Scalable multiplex co-fractionation/mass spectrometry platform for accelerated protein interactome discovery. *Nature Communications*, *13*(1), 4043. https://doi.org/10.1038/s41467-022-31809-z

[8] Havugimana, P. C., Hart, G. T., Nepusz, T., Yang, H., Turinsky, A. L., Li, Z., Wang, P. I., Boutz, D. R., Fong, V., Phanse, S., Babu, M., Craig, S. A., Hu, P., Wan, C., Vlasblom, J., Dar, V.-U.-N., Bezginov, A., Clark, G. W., Wu, G. C., … Emili, A. (2012). A census of human soluble protein complexes. *Cell*, *150*(5), 1068–1081. https://doi.org/10.1016/j.cell.2012.08.011

[9] Hu, P., Janga, S. C., Babu, M., Díaz-Mejía, J. J., Butland, G., Yang, W., Pogoutse, O., Guo, X., Phanse, S., Wong, P., Chandran, S., Christopoulos, C., Nazarians-Armavil, A., Nasseri, N. K., Musso, G., Ali, M., Nazemof, N., Eroukova, V., Golshani, A., … Emili, A. (2009). Global functional atlas of Escherichia coli encompassing previously uncharacterized proteins. *PLoS Biology*, *7*(4), e96. https://doi.org/10.1371/journal.pbio.1000096

[10] Levy, E. D., Kowarzyk, J., & Michnick, S. W. (2014). High-resolution mapping of protein concentration reveals principles of proteome architecture and adaptation. *Cell Reports*, *7*(4), 1333–1340. https://doi.org/10.1016/j.celrep.2014.04.009

[11] Panikov, N. S., Mandalakis, M., Dai, S., Mulcahy, L. R., Fowle, W., Garrett, W. S., & Karger, B. L. (2015). Near-zero growth kinetics of Pseudomonas putida deduced from

proteomic analysis. *Environmental Microbiology*, *17*(1), 215–228. https://doi.org/10.1111/1462-2920.12584

[12] Peebo, K., Valgepea, K., Maser, A., Nahku, R., Adamberg, K., & Vilu, R. (2015). Proteome reallocation in Escherichia coli with increasing specific growth rate. *Molecular bioSystems*, *11*(4), 1184–1193. https://doi.org/10.1039/c4mb00721b

[13] Rajagopala, S. V., Sikorski, P., Kumar, A., Mosca, R., Vlasblom, J., Arnold, R., Franca-Koh, J., Pakala, S. B., Phanse, S., Ceol, A., Häuser, R., Siszler, G., Wuchty, S., Emili, A., Babu, M., Aloy, P., Pieper, R., & Uetz, P. (2014). The binary protein-protein interaction landscape of Escherichia coli. *Nature Biotechnology*, *32*(3), 285–290. https://doi.org/10.1038/nbt.2831

[14] Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrné, E., Volkmer, B., Callipo, L., Knoops, K., Bauer, M., Aebersold, R., & Heinemann, M. (2016). The quantitative and condition-dependent Escherichia coli proteome. *Nature Biotechnology*, *34*(1), 104–110. https://doi.org/10.1038/nbt.3418

[15] Wan, C., Borgeson, B., Phanse, S., Tu, F., Drew, K., Clark, G., Xiong, X., Kagan, O., Kwan, J., Bezginov, A., Chessman, K., Pal, S., Cromar, G., Papoulas, O., Ni, Z., Boutz, D. R., Stoilova, S., Havugimana, P. C., Guo, X., … Emili, A. (2015). Panorama of ancient metazoan macromolecular complexes. *Nature*, *525*(7569), 339–344. https://doi.org/10.1038/nature14877

[16] Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., Doncheva, N. T., Legeay, M., Fang, T., Bork, P., Jensen, L. J., & von Mering, C. (2020). The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Research*, *49*(D1), D605–D612. https://doi.org/10.1093/nar/gkaa1074

[17] Hu, L. Z., Goebels, F., Tan, J. H., Wolf, E., Kuzmanov, U., Wan, C., Phanse, S., Xu, C., Schertzberg, M., Fraser, A. G., Bader, G. D., & Emili, A. (2019). EPIC: software toolkit for elution profile-based inference of protein complexes. *Nature Methods*, *16*(8), 737–742. https://doi.org/10.1038/s41592-019-0461-4

[18] Stacey, R. G., Skinnider, M. A., Scott, N. E., & Foster, L. J. (2017). A rapid and accurate approach for prediction of interactomes from co-elution data (PrInCE). *BMC Bioinformatics*, *18*(1), 457. https://doi.org/10.1186/s12859-017-1865-8

[19] Heusel, M., Bludau, I., Rosenberger, G., Hafen, R., Frank, M., Banaei-Esfahani, A., van Drogen, A., Collins, B. C., Gstaiger, M., & Aebersold, R. (2019). Complex-centric proteome profiling by SEC-SWATH-MS. *Molecular Systems Biology*, *15*(1), e8438. https://doi.org/10.15252/msb.20188438

[20] Barak, R., Prasad, K., Shainskaya, A., Wolfe, A. J., & Eisenbach, M. (2004). Acetylation of the chemotaxis response regulator CheY by acetyl-CoA synthetase purified from Escherichia coli. *Journal of Molecular Biology*, *342*(2), 383–401. https://doi.org/10.1016/j.jmb.2004.07.020

[21] Kim, J., Shen, R., Olcott, M. C., Rajagopal, I., & Mathews, C. K. (2005). Adenylate kinase of Escherichia coli, a component of the phage T4 dNTP synthetase complex. *The Journal of Biological Chemistry*, *280*(31), 28221–28229. https://doi.org/10.1074/jbc.M502201200

[22] Keseler, I. M., Gama-Castro, S., Mackie, A., Billington, R., Bonavides-Martínez, C., Caspi, R., Kothari, A., Krummenacker, M., Midford, P. E., Muñiz-Rascado, L., Ong, W. K., Paley, S., Santos-Zavaleta, A., Subhraveti, P., Tierrafría, V. H., Wolfe, A. J., Collado-

796 Vides, J., Paulsen, I. T., & Karp, P. D. (2021). The EcoCyc Database in 2021. *Frontiers in*
797 *Microbiology*, *12*, 711077. https://doi.org/10.3389/fmicb.2021.711077
798 [23] Skinnider, M. A., Scott, N. E., Prudova, A., Kerr, C. H., Stoynov, N., Stacey, R. G.,
799 Chan, Q. W. T., Rattray, D., Gsponer, J., & Foster, L. J. (2021). An atlas of protein-protein
800 interactions across mouse tissues. *Cell*, *184*(15), 4073–4089.e17.
801 https://doi.org/10.1016/j.cell.2021.06.003
802 [24] Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2016). KEGG as
803 a reference resource for gene and protein annotation. *Nucleic Acids Research*, *44*(D1),
804 D457–D462. https://doi.org/10.1093/nar/gkv1070
805 [25] The BioGRID database: A comprehensive biomedical resource of curated protein, genetic,
806 and chemical interactions
807 [26] Federico, A., & Monti, S. (2020). hypeR: an R package for geneset enrichment
808 workflows. *Bioinformatics* , *36*(4), 1307–1308.
809 https://doi.org/10.1093/bioinformatics/btz700
810 [27] Tong, Y.-B., Shi, M.-W., Qian, S. H., Chen, Y.-J., Luo, Z.-H., Tu, Y.-X., Xiong, Y.-L.,
811 Geng, Y.-J., Chen, C., & Chen, Z.-X. (2021). GenOrigin: A comprehensive protein-coding
812 gene origination database on the evolutionary timescale of life. *Journal of Genetics and*
813 *Genomics = Yi Chuan Xue Bao*, *48*(12), 1122–1129.
814 https://doi.org/10.1016/j.jgg.2021.03.018
815 [28] Shi, Y., Zhang, Y., Lin, S., Wang, C., Zhou, J., Peng, D., & Xue, Y. (2020). dbPSP 2.0,
816 an updated database of protein phosphorylation sites in prokaryotes. *Scientific Data*, *7*(1),
817 164. https://doi.org/10.1038/s41597-020-0506-7
818 [29] Quaglia, F., Mészáros, B., Salladini, E., Hatos, A., Pancsa, R., Chemes, L. B., Pajkos, M.,
819 Lazar, T., Peña-Díaz, S., Santos, J., Ács, V., Farahi, N., Fichó, E., Aspromonte, M. C.,
820 Bassot, C., Chasapi, A., Davey, N. E., Davidović, R., Dobson, L., … Piovesan, D. (2022).
821 DisProt in 2022: improved quality and accessibility of protein intrinsic disorder annotation.
822 *Nucleic Acids Research*, *50*(D1), D480–D487. https://doi.org/10.1093/nar/gkab1082