Elucidating the dynamic remodelling of Escherichia Coli interactome in different growth conditions using multiplex co-fractionation MS (mCF-MS)

Teck Yew Low[1]

[1]UKM Medical Molecular Biology Institute (UMBI)

May 23, 2023

Abstract

Most proteins function by forming complexes within a dynamic interconnected network that underlies various biological mechanisms. To systematically investigate such interactomes, high-throughput techniques including CF-MS have been developed to capture, identify, and quantify protein-protein interactions (PPIs) in large-scale. Compared to other techniques, CF-MS allows the global identification and quantification of native protein complexes in one setting, without genetic manipulation and overexpression. Furthermore, quantitative CF-MS can potentially elucidate the distribution of a protein in multiple co-elution features, informing the stoichiometries and dynamics of a target protein complex. In this issue, Youssef et al. (Proteomics 2023, XX, XXXX-XXXX) combined multiplex CF-MS and an in-house algorithm to study the dynamics of the PPI network for Escherichia coli grown under ten different conditions. While the results demonstrated that while most proteins remained stable, the authors were able to detect disrupted interactions that were growth condition-specific. Further bioinformatics analyses also revealed biophysical properties and structural patterns that govern such a response.

Elucidating the dynamic remodelling of *Escherichia Coli*interactome in different growth conditions using multiplex co-fractionation MS (mCF-MS)

Teck Yew Low

UKM Medical Molecular Biology Institute (UMBI), Universiti Kebangsaan Malaysia, Kuala Lumpur, Malaysia

Correspondence to :

Teck Yew Low

UKM Medical Molecular Biology Institute (UMBI)

Universiti Kebangsaan Malaysia (UKM)

UKM Medical Centre

Jalan Yaacob Latiff

Bandar Tun Razak

56000 Kuala Lumpur

Malaysia

lowteckyew@ppukm.ukm.edu.my

Phone: +60 (3) 9145 9081

Fax: +60 (3) 9171 7185

743 Words

Keywords : Co-fractionation-Mass Spectrometry (CF-MS), protein-protein interaction, co-elution, co-migration, interactome

Abstract

Most proteins function by forming complexes within a dynamic interconnected network that underlies various biological mechanisms. To systematically investigate such interactomes, high-throughput techniques including CF-MS have been developed to capture, identify, and quantify protein-protein interactions (PPIs) in large-scale. Compared to other techniques, CF-MS allows the global identification and quantification of native protein complexes in one setting, without genetic manipulation and overexpression. Furthermore, quantitative CF-MS can potentially elucidate the distribution of a protein in multiple co-elution features, informing the stoichiometries and dynamics of a target protein complex. In this issue, Youssef et al. (Proteomics 2023, XX, XXXX-XXXX) combined multiplex CF-MS and an in-house algorithm to study the dynamics of the PPI network for *Escherichia coli* grown under ten different conditions. While the results demonstrated that while most proteins remained stable, the authors were able to detect disrupted interactions that were growth condition-specific. Further bioinformatics analyses also revealed biophysical properties and structural patterns that govern such a response.

(**156 Words**)

Youssef et al. conducted a study using multiplex co-fractionation mass spectrometry (mCF-MS) and a custom software pipeline to investigate protein-protein interactions (PPI) in*Escherichia coli* under ten different growth conditions [1]. The results revealed dynamic rewiring of protein networks, and subsequent bioinformatics analyses shed light on the biophysical properties and structural patterns that govern the response to environmental stimuli.

Bacteria such as *E. coli* possess mechanisms to adapt to changing physiological demands through genetic and biochemical processes [2]. These mechanisms involve regulating protein abundance and the dynamic formation of multi-subunit assemblies. Currently, comprehensive studies of these complex "interactomes" mainly rely on high-throughput strategies that combine biochemical purification/fractionation with MS-based techniques [3]. A related concept for capturing protein-protein interactions (PPIs) arises from the spatiotemporal co-behaviour of proteins, including co-expression, co-localization, co-aggregation, or co-fractionation. Co-fractionation occurs when proteins from the same assembly comigrate from an analytical column under native conditions, suggesting potential co-localisation [4].

In this study, *E. coli* K-12 was subjected to ten different growth conditions, including various culture media and stressors. The authors analysed the *E. coli* interactome under each condition using mCF-MS, which involved fractionation of native protein complexes by high-performance liquid chromatography (IEX-HPLC). Subsequently, a bottom-up proteomics workflow with isobaric TMT labelling facilitated the identification and quantification of condition-specific protein complex elution patterns.

A critical aspect of this work is a two-module data analysis pipeline developed by Youssef et al. The first module performs pre-processing steps, including data normalisation, smoothing, and filtering, to construct protein profiles and correlate them with reference PPI databases, including STRING and BioGRID [5,7]. This pre-processing workflow addresses the disparity between peptide-based TMT measurements and the quantification of proteins inferred from tryptic peptides. This workflow also dynamically captures PPIs dynamically, distinguishing it from existing software tools [6,8,9]. Meanwhile, the second module computes a conditional similarity score for each predicted PPI pair from the mCF-MS data compared to a reference

interactome. To estimate the degree of interactome remodelling, the authors introduced a similarity score derived from two distinct data metrics: co-elution (qualitative measure of HPLC retention time shifts) and co-abundance (quantitative measure of intensity fold changes).

Using this pipeline, the authors generated a reference interactome comprising 6,152 high-confidence pairwise interactions from all ten mCF/MS datasets. Most proteins (68.3%) formed a single complex, while the rest participated in multiple complexes. The overall extent of interactome remodelling remained relatively stable in all conditions, with less than 5% showing high remodelling scores indicative of disrupted interactions. Notably, under galactose as the primary carbon source, a protein complex mainly composed of galactose metabolism enzymes was extensively formed. The high temperature (42 °C) resulted in the decomposition of a remodelled top-ranking complex into two subcomplexes—one consisting of heat shock response proteins from the 'hsl' family and the other consisting of hydrogenase proteins from the 'hya' gene family. The authors also identified pathways that were influential in driving remodelling within each growth condition. The protein modification machinery consistently showed a significant impact across all conditions, highlighting the role of dynamic post-translational modifications.

The authors conducted an analysis to explore the associations between different protein traits and integrated remodelling scores within each growth condition. First, they found a negative correlation between remodelling scores and protein intensities, indicating that proteins with lower abundance were more prone to remodelling. Furthermore, at near-zero growth and stationary phase, the stability of ancient protein interactions (based on evolution age) was found to be favoured. Furthermore, membrane proteins exhibited relatively higher stability compared to their cytosolic and periplasmic counterparts, and proteins undergoing phosphorylation tended to be more stable. On the other hand, proteins with intrinsically disordered structures had lower median remodelling scores, indicating their tendency to form constitutive assemblies. Additionally, highly connected hub proteins involved in multiple interactions demonstrated greater stability, while the number of complexes in which a protein participated had a lesser impact. To facilitate the exploration of results, the authors also developed an interactive web application that visualises the dynamic profiles of*E. coli* mCF-MS.

At the end of the manuscript, the authors discussed the limitations of their workflows. One drawback pertained to the freeze-thawing and sonication of the *E. coli* cells prior to co-fractionation. They also acknowledged their sheer focus on large-scale interactome remodelling patterns, while the ten different *E. coli* culture conditions would have provided ample material for additional biological experiments. Notwithstanding, the experimental and computational pipelines presented here have contributed to proof-of-concept and a ready-to-use workflow for future investigations into interactome dynamics in different biological systems.

The author declared no conflict of interest.

References

[1] Ahmed Youssef, Bian, F., Paniikov, N.S., Crovella, M., Emili, A., Dynamic Remodeling of Escherichia coli Interactome in Response to Environmental Perturbations. *Proteomics* n.d.

[2] Caglar, M.U., Houser, J.R., Barnhart, C.S., Boutz, D.R., et al., The E. coli molecular phenotype under different growth conditions.*Sci. Rep.* 2017, 7.

[3] Low, T.Y., Syafruddin, S.E., Mohtar, M.A., Vellaichamy, A., et al., Recent progress in mass spectrometry-based strategies for elucidating protein–protein interactions. *Cell. Mol. Life Sci.*2021, 78, 5325–5339.

[4] Salas, D., Stacey, R.G., Akinlaja, M., Foster, L.J., Next-generation Interactomics: Considerations for the Use of Co-elution to Measure Protein Interaction Networks. *Mol. Cell. Proteomics*2020, 19, 1–10.

[5] Szklarczyk, D., Gable, A.L., Nastou, K.C., Lyon, D., et al., The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets.*Nucleic Acids Res.* 2021, 49, D605–D612.

[6] Hu, L.Z.M., Goebels, F., Tan, J.H., Wolf, E., et al., EPIC: software toolkit for elution profile-based inference of protein complexes. *Nat. Methods* 2019, 16, 737–742.

[7] Oughtred, R., Rust, J., Chang, C., Breitkreutz, B.J., et al., The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* 2021, 30, 187–200.

[8] Stacey, R.G., Skinnider, M.A., Scott, N.E., Foster, L.J., A rapid and accurate approach for prediction of interactomes from co-elution data (PrInCE). *BMC Bioinformatics* 2017, 18.

[9] Heusel, M., Bludau, I., Rosenberger, G., Hafen, R., et al., Complex-centric proteome profiling by SEC-SWATH-MS. *Mol. Syst. Biol.* 2019, 15.