FLUOROALCOHOL INDUCED SUPRAMOLECULAR BIPHASIC SYSTEMS OF ZWITTERIONIC SURFACTANTS, MIXED AMPHIPHILES, AND POLAR ORGANIC SOLVENTS IN PROTEOMICS AND ENVIRONMENTAL ANALYSIS.

DISSERTATION

By

Durga Devi Khanal

Submitted to the Faculty of Graduate School of

University of Texas at Arlington, in partial fulfillment

of requirements for the degree of DOCTOR OF PHILOSOPHY

Department of Chemistry and Biochemistry

December 2021, Arlington, TX

Supervising committee:

Dr. Morteza G. Khaledi, Supervising Professor

Dr. Saiful Chowdhury

Dr. Purnendu K Dasgupta

Dr. Krishnan Rajeshwar

I want to dedicate this dissertation to my family

For their endless love, support, and encouragement throughout this journey

ACKNOWLEDGMENTS

I would like to express my great respect and acknowledgment to my supervising professor Dr. Morteza G. Khaledi for providing me this great opportunity to work in his lab under his supervision. He has been always supportive to me throughout my Ph.D. journey with great mentorship to graduate me with my Ph.D. degree. I feel so proud and lucky to pursue my Ph.D. degree under his supervision.

I would also like thank to respected committee members professor Dr. Saiful Chowdhury, professor Dr. Purnendu K. Dasgupta, and professor Dr. Krishnan Rajeshwar for their valuable time and suggestions throughout my Ph.D. journey. I would like to thank the chemistry and biochemistry department of UTA for all the support and time that I have received throughout my graduate study.

I would like to thank all my lab mates especially Dr. Mohammad Mehdi Azizi and Sajad Tasharofi, for their great companions, help, and motivation during this journey. I believe that the lab environment has a great impact on someone's success, and I am grateful for them.

I cannot adequately express my thanks to all my family members for their endless love and support in this long journey. My special thanks to my dear husband Dr. Pawan Thapa for loving and supporting me all the time and encouraging me in my bad days. My respect and love to my dear dad Dilliram Khanal, my dear mom Pabitra Devi Khanal and my dear mother-in-law Tara Thapa for their continuous love, help, and support to get this degree. I would like to remember my daughter Priyasha Thapa and my newborn Pranisha Thapa for their sacrifice and for letting me work in my research lab during the time they needed me. I love you both and I am sure you will be proud of your mom one day

December 2021

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF ABBREVIATIONS

FA*i*C	fluoroalcohol-Induced Coacervate
FA*i*C-BPS	fluoroalcohol-induced coacervate biphasic systems
AOSB	Associated Organic Solvents Biphasic
DMMAPS	dimethylmyristylammoniopropane sulfonate
CHAPS	3-[(3-Cholamidopropyl) dimethyl ammonio]-1-propane sulfonate
CTAB	acetyl trimethyl ammonium bromide
SC	sodium cholate
SDC	sodium deoxycholate
DTAB	dodecyl trimethylammonium bromide
TEAB	tetra ethyl ammonium bromide
TBAB	tetra butyl ammonium bromide
HFIP	hexafluoro isopropanol
FASP	filter assisted sample preparation
GRAVY	grand average of hydropathy
pI	isoelectric point
DTT	dithiothreitol
IAA	iodoacetamide
ABC	ammonium bicarbonate
UTT	urea thiourea
FDR	false discovery rate
ACN	acetonitrile
THF	tetrahydrofuran

ABSTRACT

Fluoroalcohol-Induced Supramolecular Biphasic Systems of Zwitterionic Surfactants, Mixed Amphiphiles, and Polar Organic Solvents in Proteomics and Environmental Analysis.

By

Durga Devi Khanal

The University of Texas at Arlington

Supervising professor: Dr. Morteza G. Khaledi

In this research, we have investigated several HFIP-induced biphasic systems of zwitterionic surfactants like CHAPS and DMMAPS and their mixed amphiphilic systems for the extraction, enrichment, and fractionation of mixture of standard proteins and complex sample mixture of yeast proteins. The biphasic systems of zwitterionic surfactants (CHAPS and DMMAPS) and their mixed amphiphiles with quaternary ammonium salts (QUATS= TBAB, TEAB) provide concomitant extraction of hydrophobic membrane proteins, enrichment of lower abundance proteins to one phase, and fractionation of complex sample mixture into two phases. The results obtained from these biphasic systems were compared with control systems of aqueous solutions of different popular protein solubilizing reagents such as surfactants, amphiphiles, and urea. Among the different controls, 8M urea solution performed best for the total extraction of yeast proteins therefore, it was accepted the main control system throughout this project. Interestingly, the results obtained from biphasic systems of different zwitterionic surfactants and mixed amphiphiles show a significant improvement in extraction of yeast proteins from 11.5% to 18.1% greater as compared to control for the whole proteome. Addition of TBAB to both CHAPS and DMMAPS showed the highest identification improvement of 16.1% and 18.1% respectively. The

FA*i*C-BPS of DMMAPS and mixed amphiphiles outperformed the CHAPS biphasic systems for the solubilization of hydrophobic membrane proteins and alpha helical parts of the membrane proteins with total identification improvement of 18.8% for overall membrane proteins and 26.4% and 555% for the integral component of membrane and proteins with alpha helices respectively as compared to control. The FAiC-BPS of CHAPS and mixed amphiphiles show similar or no identification improvement for the proteins with alpha helices on them.

The improvements in coverage of lower abundance proteins were considerably higher than those for the whole proteome due to significant enrichment effects of the FAiC-BP systems that enabled detection of low abundance proteins that are often undetected in conventional systems. All biphasic systems of CHAPS and DMMAPS and their mixed amphiphiles showed significant identification improvements for the lower abundance proteins below abundance 5000 molecules/cells with respect to the control. We have reported more than 100% identification improvement for low abundance proteins with an abundance range of 0-2000 molecules/cells using HFIP induced biphasic systems of DMMAPS and DMMAPS+QUATS and the best results were obtained from the system containing mixed QUATS and DMMAPS. However, identification improvement of CHAPS and the mixed amphiphilic system is slightly lower than DMMAPS systems but still shows the identification improvement of 60 to 80% higher than control. Most interestingly, in both cases, detection of low abundance proteins (Abd=0-2000 molecules/cells) is higher in mixed amphiphilic systems of CHAPS and DMMAPS. This indicates that the addition of TBAB and TEAB to CHAPS and DMMAPS improves the solubilization and enrichment capability of the system for lower abundance proteins.

Additionally, in this research, we have investigated the fractionation of protein mixtures using biphasic systems. Hydrophobic and long-chain surfactants like CHAPS and DMMAPS mostly

intend to exert the hydrophobic interaction between the coacervate phase and protein molecule, therefore most of the proteins are extracted into the coacervate phase of each system. However, addition of positively charged quaternary ammonium salts (QUATS) with short chains to the zwitterionic surfactant solutions introduce electrostatic interaction to the system. Quaternary ammonium salts are mostly extracted into the coacervate phase of CHAPS and DMMAPS systems and provide positively charged sites of interaction with proteins in the coacervate phase. In HFIP induced supramolecular biphasic systems hydrophilic interaction, hydrophobic interaction, and electrostatic interaction influence protein distribution. As the result, we have observed the fractionation of protein mixture based on their isoelectric point value. In this research, we have investigated fractionation of a standard protein mixture as well as complex sample mixture of yeast proteome. In zwitterionic + QUATS systems more negatively charged acidic proteins are extracted into the coacervate phase and positively charged basic proteins are repelled into the aqueous phase. In the second part of our research, we have investigated Associated Organic Solvents Biphasic (AOSB) for the extraction and enrichment of environmental pollutants like polycyclic aromatic hydrocarbons (PAHs), pesticides, and sex steroid hormones. We have reported the four AOSB systems with acetonitrile, acetone, n-propanol, and tetrahydrofuran as polar organic solvents and observed their phase transition behavior with HFIP and water. The compositional analysis using the GC-MS instrument report that almost 85% of total organic solvents (HFIP + second polar organic solvent) is present in bottom organic phase at their equimolar volume combination. However, the top organic phase contains around 10% of organic solvents of total volume, and around 90% of water. Due to the presence of high concentration of organic solvents in bottom phase is useful for higher solubilization and enrichment of hydrophobic small molecules. This indicates that the top phase is hydrophilic in nature and the bottom phase is hydrophobic in nature

with strong that can solubilize and enrich hydrophobic small molecules. We have selected the composition of 7.5% of polar organic solvents (ACN, acetone, n-propanol, and/ THF) with 7.5% of HFIP in 85% of water as our preliminary systems. These systems have same initial compositions but gave different bottom organic phase volume. We have reported the enrichment factors of 25 to 41 for the PAHs sample in a 30 µL of bottom organic phase volume generated by 7.5% ACN+ 4 % HFIP+88.5% water and 7.5% acetone+ 3.5% HFIP+ 89% water. Similar, results were observed for the organic pesticide samples. However, some of the sex steroid hormones had smaller partition coefficients into the bottom organic phase, that resulted in lower enrichment fact.

CHAPTER 1

1.1.INTRODUCTION

Extraction and enrichment of membrane and lower abundance proteins are important in proteomics since they play a crucial role in various biological activities [1,2]. Membrane proteins play a vital role in cell-cell communication, signal transduction, for the transportation of various types of ions, small molecules, and drug molecules across the cell membrane. It acts as the barrier between a living cell and its external environment. The discovery of several therapeutic methods relies on the proper identification of several membrane proteins and lower abundance proteins involved in the biological activities [1]. Therefore, development of methodologies which are helpful for their proper detection is extremely important. Different types of analytical instruments like HPLC-UV, LC-MS, LC-MS/MS, MALDI-TOF, electrophoresis, SDS-PAGE are used for the detection of proteins based on top-down or bottom-up analysis. Bottom-up proteomics is very common in proteomics since proteins are large molecules and they are hard to detect by top-down approach due to the limitations associated with the MS instrument [3]. In addition, detection of membrane proteins is not so easy since they reside in the cell lipid bilayer and are hydrophobic in nature, and precipitate or aggregate in environments outside of lipid bilayers [4-6]. Also, they are less exposed to enzymatic digestion when they are in the lipid bilayer and less available for detection [7]. Besides, the lower abundance proteins are hard to detect because of their lower concentrations and can be shadowed by higher abundance proteins [8-11]. Better extraction procedures, solubilizing agents, and enrichment methods are helpful for their improved identification.

Surfactants and denaturing agents are very popular for the solubilization of membrane proteins. Solubilizing agents like urea, sodium dodecyl sulfate (SDS), sodium cholate (SC), sodium deoxycholate (SDC), CHAPS, DMMAPS, etc. are commonly used in proteomics research [6-14]. They interrupt the lipid-protein environment and help to solubilize the membrane by mimicking the lipid environment. However, there are several drawbacks associated with these surfactant solubilization. First, their presence leads to low trypsin activity, and they need to be removed before digestion. Second, some of the surfactants like SDS have a strong affinity to proteins and that makes their removal before digestion challenging. There are basically two types of protein extraction techniques: solid-phase extraction (SPE) and liquid-liquid extraction (LLE) [15-17]. SPE is popular for its wide range of surface chemistry and selectivity as compared to LLE, however, there are some problems associated with the sample recovery in SPE. On the other hand, LLE is advantageous for better sample recovery as compared to SPE. Within a decade, our lab has investigated fluoro alcohol or/ fluoro acid-induced supramolecular biphasic systems of different types of surfactants, amphiphiles, mixed surfactants/amphiphiles, polyelectrolytes, sugars, polar organic solvents etc. and their application in proteomics and extraction of small molecules [18-21]. Supramolecular structure refers to the nanostructure of solvents and amphiphilic molecules formed as colloidal particles by the mechanism of self-assembly of solvents and amphiphilic molecules suspended throughout another solvent [22]. They are usually formed as suspended particles into an aqueous solvent and separate into a second phase based on their density difference and polarity. In water, they usually aggregate at the bottom phase and are also called as the coacervate phase. Fluoro alcohol induced supramolecular biphasic systems (FA*i*C-BPS) is a sub-group of liquid-liquid extraction techniques. The term coacervation is first used by scientists Bungenberg de Jong and Kruyt in 1930 [23]. Coacervation occurs based on different physical and

chemical parameters like the molecular structure of amphiphiles, their ionic strength, pH, presence of additives, temperature, pressure, hydrophobic and hydrophilic interaction between molecules, etc. FA*i*C-BP systems include providing a similar selectivity as compared to SPE by adjusting the phase composition using different amphiphiles. At the same time, it provides a higher sample enrichment capability than SPE due to the use of liquid solvent. Additionally, sample loss due to irreversible sample adsorption to the solid phase is lower in the case of the FA*i*C-BP system [24]. Additionally, coacervation is more beneficial because of their ability to concentrate the analytes in one phase and are cost-effective and are environmentally friendly as compared to organic liquid-liquid extractions.

Our laboratory has previously demonstrated fluoroalcohols/fluoroacids induced biphasic supramolecular systems of broad classes of surfactants, amphiphiles, polyelectrolytes, bile salts, and phospholipids, etc and found that fluoroalcohols like HFIP and TFE can greatly facilitate coacervate formation. The ability of fluoroalcohols to induce coacervation and phase separation in an aqueous solution of amphiphiles is based on their strong ability as a proton donor and the presence of hydrophobic fluorocarbon groups cluster around carbon chains of amphiphiles which excludes water molecules and subsequently separate the coacervate phase within the bulk aqueous phase. The presence of high concentrations of amphiphile and the fluoroalcohol into the coacervate phase shows high solubilizing power for hydrophobic compounds and leads to their concomitant extraction and enrichment to the small volumes of the coacervate phase relative to the total solution volume. The enrichment factor of the coacervate phase for analytes is inversely proportional to the volume of the supramolecular (coacervate) phase. Coacervate phases with smaller volumes provide higher enrichment effects. In addition to enrichment, the coacervation approach also provides fractionation of mixtures between two phases. In 2017, our lab published the first report

on the application of FA*iC*-BPS in bottom-up proteomics analysis of yeast proteins by using complex coacervation of anionic sodium dodecyl sulfate (SDS) and cationic cetyl trimethylammonium bromide (CTAB), and simple coacervation of zwitterionic dimyristyl dimethyl ammonium propane sulfate (DMMAPS) [20]. The biphasic FA*iC*-BP systems provided greater coverage than the control experiment (no phase separation). Similarly, another report of TFE and HFIP induced coacervation of tetrabutyl ammonium salts (TBAB) in water showed the fractionation of protein mixtures with hydrophilic, water-soluble proteins residing more into an aqueous phase and hydrophobic and membrane proteins being extracted and enriched in the coacervate phase [25]. Nonetheless, the fluoroalcohol-induced coacervation of TBAB followed similar patterns as those for surfactants. In addition, fractionation patterns in the TBAB-system were altered by the addition of salt or base additives like NaCl and NaOH. The difference between the TBAB and CTAB is based on their molecular structure. The long-chain surfactants interact with proteins through strong hydrophobic effects, while addition of a small catatonically charged TBAB enhances electrostatic interactions with proteins.

The application of fluoroalcohol or fluoroacid indued supramolecular biphasic systems is not limited to proteomics. It has a wide application for the extraction, enrichment, and fractionation of hydrophobic small molecules as well. We have recently investigated the HFIP-induced biphasic systems of polar organic solvents in water for the extraction and enrichment of environmental pollutants like polycyclic aromatic hydrocarbons (PAH), pesticides, and sex steroid hormones. They are the organic pollutants and cause acute and chronic effects on human health and other living organism. These small molecules are produced as byproduct of human activities and some of them are produced naturally. Different industries like food industries, pharmaceutical industries, textile industries, oil industries, etc. produce toxic chemicals: polycyclic aromatic hydrocarbons

(PAH), pesticides, hormones, nitro compounds, synthetic organic dyes, toxic heavy metals, etc. as their waste and affect the aquatic life and land life directly [26-27]. Even at trace levels, many are highly toxic and carcinogenic to a living organism. Several sample preparation techniques like SPE, LLE and analytical methods like HPLC-UV, LC-MS, GC-MS, etc. are used to detect their concentrations in the environment. Before their direct analysis using analytical instruments, they are first extracted and enriched by different sample preparation techniques. Easy, economic, and environmentally friendly sample preparation is always preferred for their better extraction so that they can be detected easily from complex sample mixtures. Therefore, we are introducing the HFIP-induced biphasic systems of polar organic solvents in water for the extraction, enrichment, and fractionation.

The main focus of this research is to investigate the capabilities of new FA*iC*-BPS systems with zwitterionic surfactants and mixed amphiphiles with QUATS in proteomic analysis and different organic solvents like acetonitrile (ACN), acetone, n-propanol, and tetrahydrofuran (THF) for the extraction and enrichment of small molecules like environmental pollutants. These HFIP induced biphasic systems of surfactants, amphiphiles and organic solvents produce strongly hydrophobic bottom phase and aqueous top phase, therefore they can be used to concentrate hydrophobic compounds into the bottom phase of the system. Overall, the HFIP-induced supramolecular biphasic systems of different amphiphiles have a variety of applications in proteomics and environmental analysis.

1.2.REFERENCES

2. Vuckovic D, Dagley LF, Purcell AW, Emili A. Membrane proteomics by high-performance liquid chromatography–tandem mass spectrometry: Analytical approaches and challenges. PROTEOMICS. 2013;13(3-4):404-23.
3. Hedin LE, Illergård K, Elofsson A. An Introduction to Membrane Proteins. Journal of Proteome Research. 2011;10(8):3324-31.
4. Chait BT. Mass spectrometry: bottom-up or top-down? Science. 2006;314(5796):65-6.
5. Bordag N, Keller S. α-Helical transmembrane peptides: A "Divide and Conquer" approach to membrane proteins. Chemistry and Physics of Lipids. 2010;163(1):1-26.
6. Moore SM, Hess SM, Jorgenson JW. Extraction, Enrichment, Solubilization, and Digestion Techniques for Membrane Proteomics. Journal of Proteome Research. 2016;15(4):1243-52.
7. Waas M, Bhattacharya S, Chuppa S, Wu X, Jensen DR, Omasits U, et al. Combine and Conquer: Surfactants, Solvents, and Chaotropes for Robust Mass Spectrometry Based Analyses of Membrane Proteins. Analytical Chemistry. 2014;86(3):1551-9.
8. Masuda T, Tomita M, Ishihama Y. Phase Transfer Surfactant-Aided Trypsin Digestion for Membrane Proteome Analysis. Journal of Proteome Research. 2008;7(2):731-40.
9. Zhang Y, Lin Z, Tan Y, Bu F, Hao P, Zhang K, et al. Exploration of Missing Proteins by a Combination Approach to Enrich the Low-Abundance Hydrophobic Proteins from Four Cancer Cell Lines. Journal of Proteome Research. 2020;19(1):401-8
10. Corthals, G.L., et al., *The dynamic range of protein expression: A challenge for proteomic research.* ELECTROPHORESIS, 2000. **21**(6): p. 1104-1115.

11. Anderson, N.L. and N.G. Anderson, *The human plasma proteome: history, character, and diagnostic prospects.* Molecular & cellular proteomics, 2002. **1**(11): p. 845-867.
12. Pedersen, S.K., et al., *Unseen Proteome: Mining Below the Tip of the Iceberg To Find Low Abundance and Membrane Proteins.* Journal of Proteome Research, 2003. **2**(3): p. 303-311.
13. Helbig, A.O., A.J.R. Heck, and M. Slijper, *Exploring the membrane proteome—Challenges and analytical strategies.* Journal of Proteomics, 2010. **73**(5): p. 868-878.
14. Speers, A.E. and C.C. Wu, *Proteomics of Integral Membrane ProteinsTheory and Application.* Chemical Reviews, 2007. **107**(8): p. 3687-3714.
15. Möglich, A., F. Krieger, and T. Kiefhaber, *Molecular Basis for the Effect of Urea and Guanidinium Chloride on the Dynamics of Unfolded Polypeptide Chains.* Journal of Molecular Biology, 2005. **345**(1): p. 153-162.
16. Martínez-Maqueda, D., et al., *Extraction/Fractionation Techniques for Proteins and Peptides and Protein Digestion*, in *Proteomics in Foods: Principles and Applications*, F. Toldrá and L.M.L. Nollet, Editors. 2013, Springer US: Boston, MA. p. 21-50.
17. Dephoure, N. and S.P. Gygi, *A solid phase extraction-based platform for rapid phosphoproteomic analysis.* Methods, 2011. **54**(4): p. 379-386.
18. Pires, M.J., M.R. Aires-Barros, and J.M.S. Cabral, *Liquid−Liquid Extraction of Proteins with Reversed Micelles.* Biotechnology Progress, 1996. **12**(3): p. 290-301.
19. Khaledi, M.G., S.I. Jenkins, and S. Liang, *Perfluorinated Alcohols and Acids Induce Coacervation in Aqueous Solutions of Amphiphiles.* Langmuir, 2013. **29**(8): p. 2458-2464.
20. Weisner, N. and M.G. Khaledi, *Organic synthesis in fluoroalcohol-water two-phase systems.* Green Chemistry, 2016. **18**(3): p. 681-685.

21. McCord, J.P., D.C. Muddiman, and M.G. Khaledi, *Perfluorinated alcohol induced coacervates as extraction media for proteomic analysis.* Journal of Chromatography A, 2017. **1523**: p. 293-299.
22. Koolivand, A., et al., *Coacervation of Lipid Bilayer in Natural Cell Membranes for Extraction, Fractionation, and Enrichment of Proteins in Proteomics Studies.* Journal of Proteome Research, 2019. **18**(4): p. 1595-1606.
23. Ballesteros-Gómez, A.; Sicilia, M. D.; Rubio, S. Supramolecular Solvents in the Extraction of Organic Compounds. A Review. Anal. Chim. Acta 2010, 677 (2), 108–130. https://doi.org/10.1016/j.aca.2010.07.027.
24. de Jong HGB, Kruyt HR. Koazervation. Kolloid-Zeitschrift. 1930;50(1):39-48.
25. Pires MJ, Aires-Barros MR, Cabral JMS. Liquid−Liquid Extraction of Proteins with Reversed Micelles. Biotechnology Progress. 1996;12(3):290-301.
26. Koolivand, A., et al., *Fluoroalcohol – Induced coacervates for selective enrichment and extraction of hydrophobic proteins.* Journal of Chromatography B, 2018. **1083**: p. 180-188
27. Deng, W., et al., *A density-tunable liquid-phase microextraction system based on deep eutectic solvents for the determination of polycyclic aromatic hydrocarbons in tea, medicinal herbs and liquid foods.* Food Chemistry, 2021. **352**: p. 129331.
28. Shi, Y., et al., *Facile fabrication of composited solid phase microextraction thin membranes for sensitive detections of trace hydroxylated polycyclic aromatic hydrocarbons in human urine.* Analytica Chimica Acta, 2021. **1158**: p. 338422.

CHAPTER 2

Improved Protein Coverage in Bottom-Up Proteomes Analysis Using Fluoroalcohol-Mediated Supramolecular Biphasic Systems with Mixed Amphiphiles for Sample Extraction, Fractionation, and Enrichment

Used with permission from Durga Devi Khanal, Sajad Tasharofi, Mohammadmehdi Azizi, and Morteza G. Khaledi*

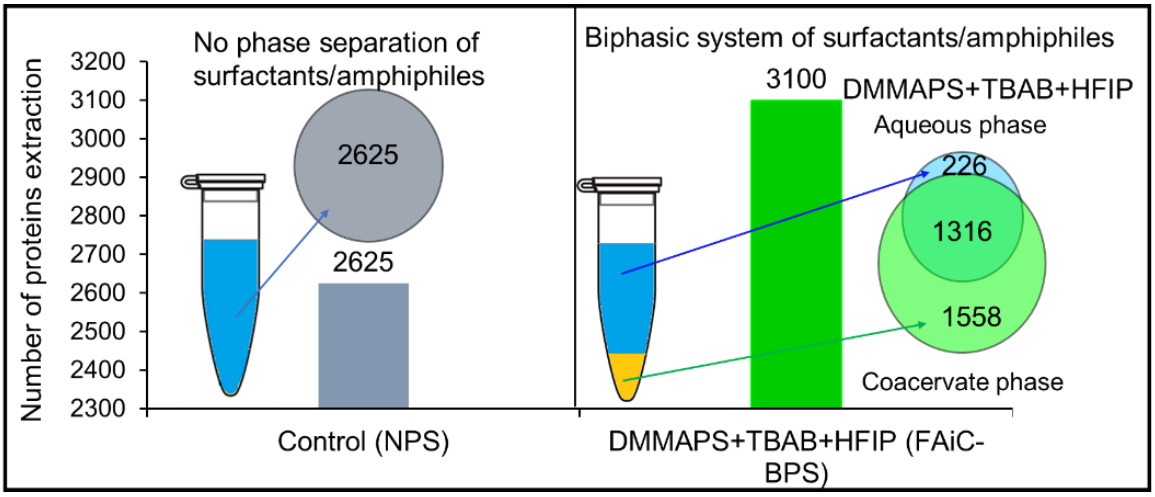
Cite This: Anal. Chem. 2021, 93, 7430−7438

ABSTRACT: A new class of supramolecular biphasic systems containing fluoroalcohol-induced coacervates (FA*i*C) provides concomitant fractionation of complex protein mixtures, high solubilizing power for extraction of various types of proteins, especially those with high hydrophobicity (such as membrane proteins), and enrichment of low-abundance proteins. Subsequently, the use of FA*i*C biphasic systems (BPS) in the bottom-up proteomics workflow resulted in significantly higher coverage for the whole proteome, various sub proteomes, especially those embedded or associated with membranes, post-translationally modified proteins, and low-abundance proteins (LAPs) as compared to the conventional methodologies. In this work, we used a new type of FA*i*C-BPS composed of mixed amphiphiles, a zwitterionic surfactant 3-(N,N-dimethylmyristyl ammonia) propane sulfonate (DMMAPS), a quaternary ammonium salt (QUATS), and hexafluoroisopropanol (HFIP) as the coacervator for extraction, fractionation, and enrichment of yeast proteome in bottom-up proteomics. The coverage of the lower-abundance proteins (abundance below 2000 molecules/cell) improved by more than 100% using DMMAPS and DMMAPS + QUATS systems as compared to the conventional methods using urea or detergent solutions for protein solubilization. Additionally, these coacervate systems show

increased coverage of integral membrane proteins and proteins with α-helices by up to 24 and 555%, respectively.

Keywords: low abundance proteins, integral membrane proteins, bottom-up proteomics, sample enrichment in proteomics, coacervation, fluoroalcohols, preconcentration, subcellular proteomics

GRAPHICAL ABSTRACT



2.1. INTRODUCTION

Biological membranes are composed of phospholipid bilayers, glycolipids, and embedded membrane proteins, which act as essential and selective barriers for cells to their external environment [1−3]. Membrane proteins are linked to various diseases, and thus serve as 50−70% of drug targets [1,4−7]. Similarly, low-abundance proteins (LAP) can provide useful information about diseases due to alteration in structure and concentration in the local environment [8,9]. Unfortunately, characterization of membrane proteins poses significant challenges because of their hydrophobic nature, low aqueous solubility, and tendency to form aggregates [10]. Similarly, identification of low-abundance proteins can be challenging as their detection in the liquid

chromatography−mass spectrometry (LC−MS) analysis is commonly overshadowed by the presence of proteins at higher concentrations [1,11,12]. Sample preparation plays a critical role in the outcome of protein coverage and successful characterization of membrane and low-abundance proteins in proteomics analysis [11−17]. Surfactants and amphiphiles are commonly used chemicals in the sample preparation of proteomics [18]. Several ionic liquid-based protein extraction approaches have been reported [19−21]. For example, the use of 1-dodecyl-3-methylimidazolium chloride (C12Im-Cl) allows the extraction of highly hydrophobic membrane proteins especially the integral component of membrane proteins. However, these approaches only reflect one-phase extraction and are not suitable for the enrichment of lower-abundance proteins. In recent years, we have discovered and investigated the usefulness of aqueous-based supramolecular biphasic systems mediated by polar fluoroalcohols and fluoroacids in sample preparation. The main category of these supramolecular systems is fluoroalcohol-induced coacervate biphasic systems (FA*i*C-BPS) that were first investigated for solubilization, extraction, fractionation, and enrichment of protein samples in bottom-up proteomics in this laboratory [22−25]. Coacervation, first introduced by de Jong and Kruyt in 1930, is the process of self-assembly of long-chain amphiphiles in aqueous media, which allows formation of a separate amphiphile-rich coacervate layer from the bulk aqueous phase [26]. In 2013, we first reported that fluoroalcohols and fluoroacids can induce coacervation in aqueous solutions of various classes of amphiphiles and over a broad range of concentrations and mole fractions in systems composed of mixed amphiphiles [24]. The fluoroalcohol in FA*i*C-BPS is a polar and water-miscible compound, such as hexafluoro isopropanol (HFIP) or trifluoroethanol (TFE) and facilitates formation of coacervates and subsequent phase separation in aqueous solutions. This is in contrast to the conventional coacervates in purely aqueous media where formation of coacervates depends

strongly on the molecular structure of amphiphiles and occurs over a narrow range of concentrations [27,28]. Thus, conventional coacervate systems have limited applications for sample preparation in chemical analysis [29].

In 2017, our group published the first report on the application of FA*i*C-BPS in bottom-up proteomics analysis of yeast proteins [25]. Simple coacervation was induced by HFIP in aqueous solutions of three different surfactants: anionic sodium dodecyl sulfate (SDS), cationic cetyltrimethylammonium bromide (CTAB), and zwitterionic 3-(N,N-dimethylmyristyl ammonio) propane sulfate (DMMAPS). The study also included one complex coacervate system composed of an equimolar mixture of oppositely charged surfactants SDS/ CTAB (1:1). The simple FAiC-BPS provided greater coverage than the control experiment (no phase separation) with commonly used urea for solubilization of the proteome. The increased coverage was particularly higher for membrane and hydrophobic proteins and for proteins that are in lower abundance in yeast cells. The following study reported coacervation of a quaternary ammonium salt (QUATS), tetra butyl ammonium bromide (TBAB), induced by TFE and HFIP. Unlike surfactants with long alkyl chains, TBAB does not self-aggregate to form micelles in aqueous media [23]. Nonetheless, the fluoroalcohol-induced coacervation of TBAB followed similar phase transition patterns as those for surfactants. The results showed that the FA*i*C-BPS composed of TBAB can be used to fractionate mixtures with hydrophilic proteins residing more in the aqueous phase and hydrophobic and membrane proteins being extracted and enriched in the coacervate phase. The long-chain surfactants such as SDS, CTAB, and DMMAPS are known to denature proteins and interact strongly with proteins through hydrophobic interaction, while the TBAB systems appear to enhance the electrostatic effect.

The FA*i*C-BPS, on the other hand, have several unique features that make them suitable for pretreatment of complex samples such as proteomes, as summarized below:

• In the FA*i*C-BPS, the coacervate phase is highly enriched with the constituent amphiphile and the fluoroalcohol, thus offering strong solubilizing power, especially for the very hydrophobic compounds, such as membrane proteins.

• The coacervate phase volume is only a small fraction of the initial solution volume, thus compounds that are extracted into the phase are simultaneously enriched by as much as three orders of magnitude [25]. This feature would be quite useful in trace analysis applications such as detection of low-abundance proteins.

• The FA*i*C-BPS are useful for facile fractionation of mixtures, with more hydrophobic compounds extracted and enriched into the coacervate phase, while more hydrophilic substances residing in the aqueous-rich phase. This characteristic is advantageous in readily reducing the complexity of mixtures such as proteomes through a simple initial step in sample pretreatment.

• Extraction and fractionation selectivity of coacervate phases can be readily altered through judicious selection of the constituent amphiphile(s), type of the fluoroalcohol (or fluoroacid) coacervator, and solution parameters such as pH and ionic strength of the aqueous-rich phase. The FA*i*C-BPS can offer a range of selectivity that nearly matches those in solid-phase extraction (SPE). This is quite remarkable considering that FA*i*C-BPS are liquid-based extraction systems, thus offering a larger sample capacity and are free of sample loss problems due to protein adsorption on solid phases as in SPE.

• As compared to conventional liquid−liquid extraction with organic solvents, the FA*i*C offer a greater range of selectivity, especially for proteins, and are more environmentally friendly as they

are composed of 80− 90% water, 5−100 mM amphiphile(s), and a small concentration of a fluoroalcohol (<20% v/v, typically 5-15% v/v,)[23, 24].

The main focus of this research is to investigate the capabilities of new FA*i*C-BPS systems that utilize mixed amphiphiles, a zwitterionic surfactant, DMMAPS, and a QUATS (TBAB or tetraethylammonium bromide, TEAB) in bottom-up proteomics studies for better solubilization, fractionation, and enrichment of whole and subproteomes, especially membrane proteins and low-abundance proteins in yeast (Saccharomyces cerevisiae) cells. High solubilizing power and enrichment capability of the coacervate phase for low abundance and membrane proteins are key for higher protein coverage. Additionally, the introduction of electrostatic interaction by TEAB and TBAB increases the number of uniquely identified proteins into two phases. As shown below, this strategy was effective in increasing protein coverage in bottom-up proteomic analysis.

2.2. EXPERIMENTAL DESIGN

2.2.1. Materials, Chemicals, and Reagents.

A 3- (N,Ndimethylmyristyl ammonia) propane sulfonate (DMMAPS) zwitterionic surfactant used in this experiment was purchased from Fisher Scientific. 1,1,1,3,3,3-Hexafluoro-2-propanol (HFIP) was obtained from Oakwood Chemical. In addition, two QUATS, tetraethylammonium bromide (TEAB) and tetrabutylammonium bromide (TBAB), were purchased from ACROS Organics. Trifluoroacetic acid (99%) and formic acid (99%) were purchased from Alfa Aesar. Chemicals forpredigestion and digestion of proteins like dithiothreitol (DTT), iodoacetamide (IAA), and sequencing grade trypsin were purchased from Promega Corporation (WI). LC−MS grade organic solvents like acetonitrile (ACN) and isopropanol (IPA) were purchased from Fisher

Chemical. Millipore-deionized (DI) water was used for sample preparation and high-performance liquid chromatography (HPLC) analysis was used from the lab during the experiment.

2.2.2. Cell Lysate and Sample Preparation. S. cerevisiae cells were grown in our lab using YPD broth (Fisher) for 16 to 20 h in a shaker incubator at 30 °C. The detailed procedure of cell lysis and sample preparation using the coacervation approach is given in the Supporting Information (SI) in Section S1. For sample preparation, 50 μL of the yeast cell lysate, which contains 400 μg of proteins, was added to the DMMAPS coacervation. In addition, the control study was done with the same amount of yeast cell lysate in 8 M of urea without phase separation for comparison. (Note: other detergents and amphiphiles like SDS, SC, DMMAPS, CHAPS, TEAB, and TBAB were used as other controls without the phase separation approach and urea was found to be the best among them). The extracted proteins in each phase were washed using an Amicon Ultra 0.5 mL centrifugal filter with a pore size of 10 K by the filter aided sample preparation (FASP) protocol to remove surfactants and salts before digestion. The FASP protocol is included in the Supporting Information SI (Sections S2 and S3) for the control and coacervate systems, respectively. Three replicates of each sample were analyzed, and two of the three common proteins were selected as the reproducible proteins.

2.2.3. Reversed-Phase High-Performance Liquid Chromatography (RP-HPLC).

The yeast sample was analyzed using the bottom-up proteomics approach with Ultimate 3000 RSLC-Nano liquid chromatography systems, Dionex, coupled with an Orbitrap Fusion Lumos MS, Thermo Electron instrument. A C18 nano column was used for the separation with a column length of 75 cm, internal diameter of 75 μm, and particle size of 3 μm. The injection volume of the sample was 1 to 2 μL (based on the protein concentration) and the solvent system used was 0−90 min gradient run with 0−28% of solvent B, and a flow rate of 350 nL/min. Mobile phase A was

2% (v/v) acetonitrile (ACN) and 0.1% (v/v) formic acid (FA) in water, and mobile phase B was 80% (v/v) ACN, 10% (v/v) trifluoroethanol (TFE), and 0.1% FA in water.

2.2.4. Data Analysis of the Yeast Sample. The raw data obtained from sample analysis was analyzed using MaxQuant (Ver. 1.6.2.3) software. The set-up parameters for the MaxQuant are the same as our previously published results [22]. In summary, MaxQuant operating parameters were as follows: FASTA file from UniPort based identification, trypsin digestion, oxidation of methionine, and N-terminal acetylation as variable modifications (maximum 5 modifications per peptide), carbamidomethyl as a fixed modification, 2 missed cleavages, label-free quantification with iBAQ (intensity-based absolute quantification), minimum one unique peptide for protein identification, PSM FDR (peptide-spectrum match false discovery rate) 1%, and protein (false discovery rate) FDR was 1%. Three replicates were carried out. The common proteins in two out of three runs were taken for further data analysis. Further data analysis was done using the UniProt database, Yeast Mine database, and Gene Ontology database to obtain more information for extracted proteins. Gene Ontology (GO) annotation for yeast database is based on gene code from Saccharomyces Genome Database (SGD) (http://www.geneontology.org) [30]. The SGD protein IDs for extracted proteins were retrieved from the UniProt database (http://www.uniprot.org). The information related to the abundance value of proteins and post-translational modification was extracted from the Yeast Mine database.

2.3. RESULTS AND DISCUSSION

2.3.1. Bottom-up Proteomics of Yeast Cells Using FA*i*C-BPS Systems. Three coacervate systems, 50 mM DMMAPS + 10% HFIP, 50 mM DMMAPS + 50 mM TEAB + 10% HFIP, and 50 mM DMMAPS + 50 mM TBAB + 10% HFIP, were investigated for the extraction and

enrichment of proteins from the yeast cell lysate in bottom-up proteomics. The proteins identified by these systems were compared with a control system using 8 M urea without phase separation, which is commonly used for protein solubilization and extraction. Each system was studied in triplicates and two out of three common proteins were taken as the reproducibly identified proteins, as shown in Venn diagrams in Figure 1. The 8 M urea solution provided the best coverage of proteins as compared to various control systems (i.e., no phase separation) composed of aqueous solutions of detergents and amphiphilic compounds like sodium dodecyl sulfate (SDS), sodium cholate (SC), DMMAPS, TEAB, and TBAB. The SDS and SC solutions are common solubilizing reagents in proteomics [19], while aqueous solutions of DMMAPS, TEAB, and TBAB were included because they are used as amphiphiles in the FA*i*C systems.

Table 1 lists the number of proteins that were identified in the aqueous and coacervate phases of the FA*i*C-BPS systems. As shown in Table 1, a larger number of proteins were identified using the FA*i*C systems as compared to all control systems without a phase separation. Protein coverage in the two mixed-amphiphile FA*i*C systems improved by 18.1% (475 proteins for the DMMAPS-TBAB) and 14.4% (377 proteins for DMMAPS-TEAB) as compared to the most effective control system, urea.

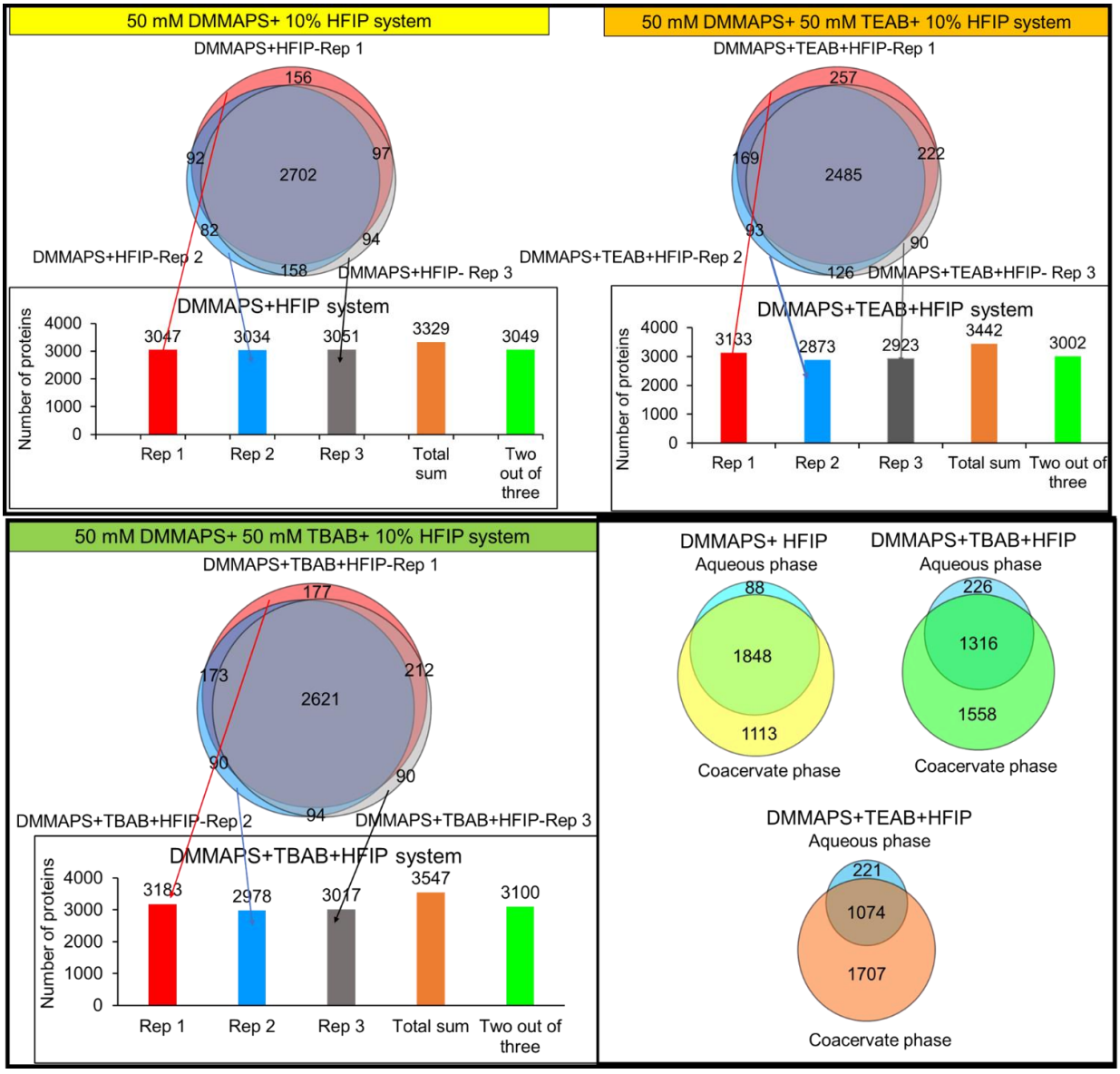


Figure 1: Total number of proteins extracted by three replicates of each system with phase separation approach. Biphasic systems used are: DMMAPS+HFIP, DMMAPS+TEAB+HFIP, DMMAPS+TBAB+HFIP, respectively

Table 1: Identification improvement of yeast proteins by using FA*i*C systems as compared to the conventional approach with 8M Urea without phase separation

FAiC Systems	Proteins identified only in Co (K=∞)	Unique Proteins only in Aq (K=0)	proteins in both Co and Aq (0<K<∞)	Total # Proteins identified	% ID Improve vs. Urea Control
DMMAPS+TBAB+HFIP	1558	226	1316	3100	18.1%
DMMAPS+TEAB+HFIP	1707	221	1074	3002	14.4%
DMMAPS+HFIP	1113	88	1848	3049	16.2%

1B. Control Systems – No Phase Separation

8M Urea (Selected Control)	50 mM SDS (NP)	50 mM SC (NP)	50 mM DMMAPS (NP)	50 mM TBAB (NP)	50 mM TEAB (NP)	50 mM CHAPS (NP)
2625	2410	2307	2293	2392	2362	2393

The list of proteins extracted into the aqueous and coacervate phases of each system and control is provided in Supporting Information Section S4 (see Excel files in SI). The identification improvement can be attributed to higher solubilizing power and enrichment capability of the coacervate systems for the membrane and low-abundance proteins, as discussed below. The first three columns in Table 1 list the number of proteins that were only found in the coacervate or aqueous phases with partition coefficients, K, ∞ or 0, as well as those found in both phases, thus having a finite K value. As shown, a significant majority of proteins are extracted into the coacervate phases. A large majority of those that existed in both phases had a higher affinity for the coacervate phases (K > 1, data not shown).

Note that addition of the QUATS (TBAB or TEAB) to the DMMAPS resulted in the increase in fractionation of proteins into separate phases. This is evident from the larger number of the

uniquely identified proteins in the DMMAPS + QUATS coacervate phases (1558 and 1707, Table 1, Figure 1) vs DMMAPS coacervate (1113) and the corresponding aqueous phases (226 and 221 vs 88, respectively). This can also be seen from the significantly larger number of proteins that can be found in both the coacervate and aqueous phases for the DMMAPS system (1848) as compared to DMMAPS + QUATS (1316 and 1074).

2.3.2. Subcellular Proteomics. The number of proteins identified in various subcellular locations according to Gene Ontology (GO) analysis (membrane proteins, mitochondrial proteins, and other subcellular proteins) in the three FA*i*C-BPS systems and for the 8 M urea as a control system are shown in the tabulated form in the Supporting Information, Section S5 as Table S1. All three FA*i*C-BPS systems performed better as compared to the control system. The last column in Table S1 lists the percentage of increase in the number of proteins identified in DMMAPS + TBAB as compared to the control system. As shown, the percentage of identification improvement was as high as around 80%, with a majority of cases, protein coverage improved by greater than 10%. Interestingly, from Table S1, the overall identification improvement for all membrane proteins is observed greatest in the DMMAPS + TBAB + HFIP system, which is 18.8% higher than control (urea), while the identification improvement of the integral membrane proteins and intrinsic membrane proteins is highest in the DMMAPS + HFIP system, which is 26.4% higher than control. This clearly shows the higher solubilizing power of DMMAPS for hydrophobic membrane proteins. This level of coverage also surpasses that with the natural lipid coacervation system (identification improvement of 13.0% for the IMP) reported previously [22].

Higher coverage of other subcellular proteins was also observed using the FAiC systems than that in the conventional approach. Other Gene Ontology based data analyses of subcellular proteins, proteins involved in biological processes, and molecular components are given in the Supporting

Information in Section S6. The results in Table S1 show that with a few exceptions, the coverage of the subcellular proteins in the three FAiC systems is very similar. However, the Venn diagrams in Figure 2 reveal that the fractionation patterns of the subcellular proteins between the aqueous and coacervate phases vary with the FAiC composition. As was observed for the whole proteome, addition of the QUATS to DMMAPS increases selectivity in protein fractionation between the two phases in FAiC. For example, the shared membrane protein between two phases in DMMAPS FAiC is 514, whereas addition of TBAB and TEAB to DMMAPS reduces the number of shared proteins to 397 and 341, respectively, by increasing the numbers of the uniquely identified proteins in the corresponding aqueous and coacervate phases. This trend is common for all types of yeast proteins. The increasing order of the uniquely identified proteins in the three different systems is as follows: DMMAPS + TEAB > DMMAPS + TBAB > DMMAPS. This is useful for fractionation and enrichment of hydrophobic low-abundance proteins into the coacervate phases as discussed below.

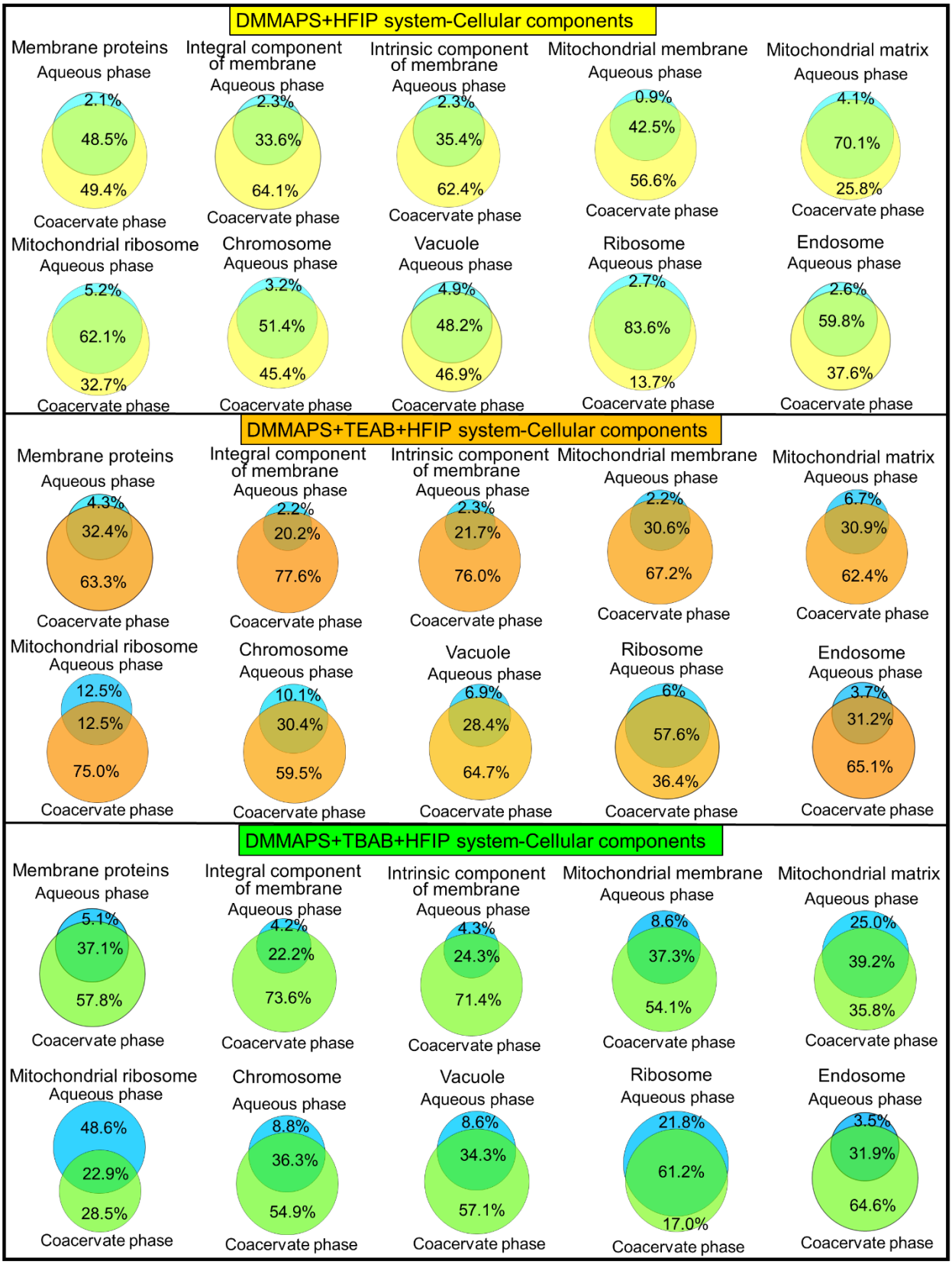

DMMAPS+HFIP system-Cellular components
Membrane proteins
Aqueous phase
2.1%
48.5%
49.4%
Coacervate phase
Integral component of membrane
Aqueous phase
2.3%
33.6%
64.1%
Coacervate phase
Intrinsic component of membrane
Aqueous phase
2.3%
35.4%
62.4%
Coacervate phase
Mitochondrial membrane
Aqueous phase
0.9%
42.5%
56.6%
Coacervate phase
Mitochondrial matrix
Aqueous phase
4.1%
70.1%
25.8%
Coacervate phase
Mitochondrial ribosome
Aqueous phase
5.2%
62.1%
32.7%
Coacervate phase
Chromosome
Aqueous phase
3.2%
51.4%
45.4%
Coacervate phase
Vacuole
Aqueous phase
4.9%
48.2%
46.9%
Coacervate phase
Ribosome
Aqueous phase
2.7%
83.6%
13.7%
Coacervate phase
Endosome
Aqueous phase
2.6%
59.8%
37.6%
Coacervate phase
DMMAPS+TEAB+HFIP system-Cellular components
Membrane proteins
Aqueous phase
4.3%
32.4%
63.3%
Coacervate phase
Integral component of membrane
Aqueous phase
2.2%
20.2%
77.6%
Coacervate phase
Intrinsic component of membrane
Aqueous phase
2.3%
21.7%
76.0%
Coacervate phase
Mitochondrial membrane
Aqueous phase
2.2%
30.6%
67.2%
Coacervate phase
Mitochondrial matrix
Aqueous phase
6.7%
30.9%
62.4%
Coacervate phase
Mitochondrial ribosome
Aqueous phase
12.5%
12.5%
75.0%
Coacervate phase
Chromosome
Aqueous phase
10.1%
30.4%
59.5%
Coacervate phase
Vacuole
Aqueous phase
6.9%
28.4%
64.7%
Coacervate phase
Ribosome
Aqueous phase
6%
57.6%
36.4%
Coacervate phase
Endosome
Aqueous phase
3.7%
31.2%
65.1%
Coacervate phase
DMMAPS+TBAB+HFIP system-Cellular components
Membrane proteins
Aqueous phase
5.1%
37.1%
57.8%
Coacervate phase
Integral component of membrane
Aqueous phase
4.2%
22.2%
73.6%
Coacervate phase
Intrinsic component of membrane
Aqueous phase
4.3%
24.3%
71.4%
Coacervate phase
Mitochondrial membrane
Aqueous phase
8.6%
37.3%
54.1%
Coacervate phase
Mitochondrial matrix
Aqueous phase
25.0%
39.2%
35.8%
Coacervate phase
Mitochondrial ribosome
Aqueous phase
48.6%
22.9%
28.5%
Coacervate phase
Chromosome
Aqueous phase
8.8%
36.3%
54.9%
Coacervate phase
Vacuole
Aqueous phase
8.6%
34.3%
57.1%
Coacervate phase
Ribosome
Aqueous phase
21.8%
61.2%
17.0%
Coacervate phase
Endosome
Aqueous phase
3.5%
31.9%
64.6%
Coacervate phase

Figure 2: Subcellular proteins extracted by three different Biphasic coacervation systems: DMMAPS+HFIP, DMMAPS+TEAB+HFIP, DMMAPS+TBAB+HFIP, respectively and proteins fractionation into aqueous phase and coacervate phase.

2.3.3. Low-Abundance Protein (LAP) Coverage Enhancement.

An advantage of using yeast (S. cerevisiae) as the model organism is the availability of comprehensive protein abundance levels [30,31]. Identification of low-abundance proteins in proteomics samples is extremely challenging due to their low concentrations and the presence of high-abundant proteins in the mixture. Several methodologies have been reported to improve the detection of low-abundance proteins through fractionation of complex mixtures or the sample matrix and analyte enrichment. For example, density-dependent ultra-centrifugation has been used to fractionate the subcellular compartments and enrich low-abundance proteins [14]. Other methods included affinity chromatography, tagging techniques like isotope-coded affinity tags, tandem mass tags, and radioisotope labeling of cells [14,15]. a new approach for the extraction, fractionation, and enrichment of low-abundance proteins. In coacervation, the enrichment factor can be defined as the solute concentration in the coacervate phase with respect to its initial concentration in total solutions [25], which is directly proportional to the volume and partition coefficient of the coacervate phase. Due to the small volume and high concentration of the surfactant and amphiphile in the coacervate phase, it offers high solubilizing power and enrichment of low-abundance proteins.

Different databases exist for the abundance study of yeast proteins expressed either in ppm level or molecules per cell [32−34], but complete proteome coverage is still unavailable. We used the Yeast Mine database for the abundance study of extracted proteins, which are typically expressed in molecules/ cells. Yeast Mine provides more than one abundance value for each protein based

on different experimental approaches. The obtained values were averaged and were considered as the abundance value of each yeast protein, which is attached as a reference in the Supporting Information in Section S7. Figure 3 (top bar graph) represents the abundance chart of the proteins extracted by the three FA*i*C-BPS systems: DMMAPS + HFIP, DMMAPS + TEAB + HFIP, and DMMAPS + TBAB + HFIP as compared to the urea control system. Figure 3 (bottom bar graph) shows the identification improvement expressed as percentages with respect to the control. Herein, we considered abundance below 2000 molecules/cells as the low-abundance proteins. Remarkably, the highest improvement of protein coverage in FA*i*C was observed for the lower abundance levels. There was little or no difference between the FA*i*C and the control systems in protein coverage for abundance levels above 5000 molecules/cells. In other words, the FA*i*C provide an advantage over the conventional methods for the low-abundance proteins. This is attributed to simultaneous fractionation and enrichment that occurs in protein extraction by FA*i*C. [Note: the protein coverage bar graphs as a function of protein abundance for levels above 5000 molecules/cells are included in the Supporting Information, Section S8]. The chart shows the identification improvement of 102%, 112%, and 114% for proteins with abundance with abundance <2000 molecules/cell in DMMAPS+HFIP, DMMAPS+TEAB+HFIP, and DMMAPS+TBAB+HFIP systems, respectively.

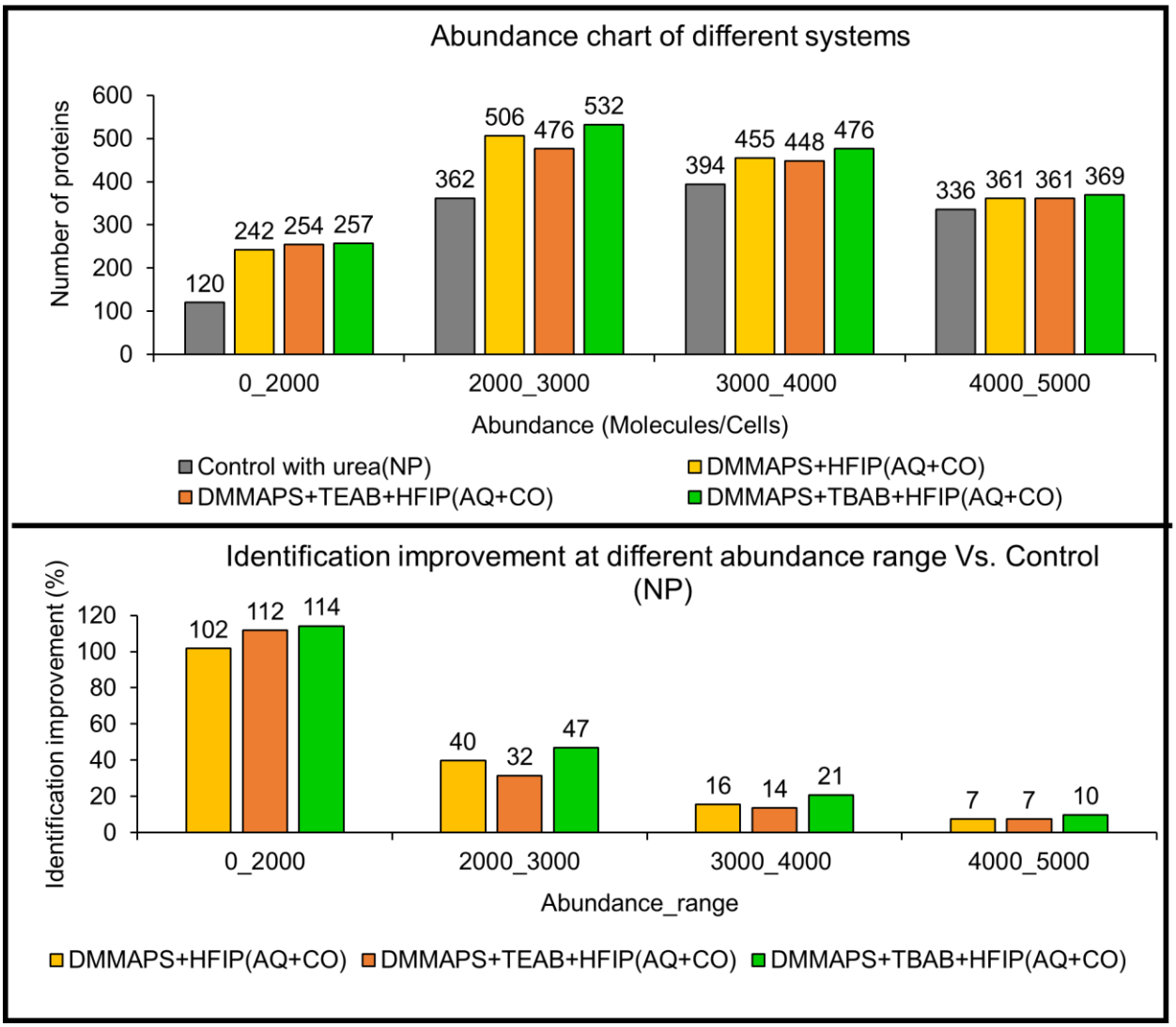


Figure 3: Abundance chart of proteins in molecules/cells extracted by DMMAPS+HFIP and DMMAPS+QUATS+HFIP systems and their identification improvement Vs. Control.

2.3.4. Fractionation Patterns of Low-Abundance Proteins in FA*i*C Systems.

As shown in Figure 4, the distribution of low-abundance proteins between the aqueous and coacervate phases is highest for the lowest-abundance proteins. The number of common proteins found in both the aqueous and coacervate phases increases with the abundance level in all three systems. Addition of TEAB and TBAB to DMMAPS also leads to an increase in fractionation of proteins with fewer common proteins between two phases or a larger number of uniquely identified

proteins in the aqueous and coacervate phases. This could be due to balancing hydrophobic and electrostatic effects in protein distribution in FA*i*C systems with mixed amphiphiles that will be discussed in a future study. In addition, analyses of subcellular proteins in the lower abundance level (abundance values of <2000 and 2000-3000 molecules/cell) are tubulated as Table S11 (see Supporting Information, Section S9), which shows that nearly 50% of low abundance proteins are membrane proteins. In addition, Table S1 shows that between 102 and 116 additional integral membrane proteins were identified in the FA*i*C systems than the control system. The results listed in Table S11 show that 40 to 48 of these proteins have abundance levels of <2000 molecules/cells and 30 to 40 proteins have an abundance of 2000−3000 molecules/cell. This again indicates that the majority of improvement in the identification of integral membrane proteins is due to the detection of proteins with lower-abundance values. In addition, the number of LAP with abundance of <2000 molecules/cell identified in the nucleus, mitochondrion, endoplasmic reticulum, and vacuole were more than double using the FA*i*C systems as compared to the control. Similar results were observed for proteins with abundance levels between 2000 and 3000 molecules/cell. These results indicate that the majority of improvements in protein coverage of the whole and sub proteomes using FA*i*C systems are due to better detection of lower-abundance proteins.

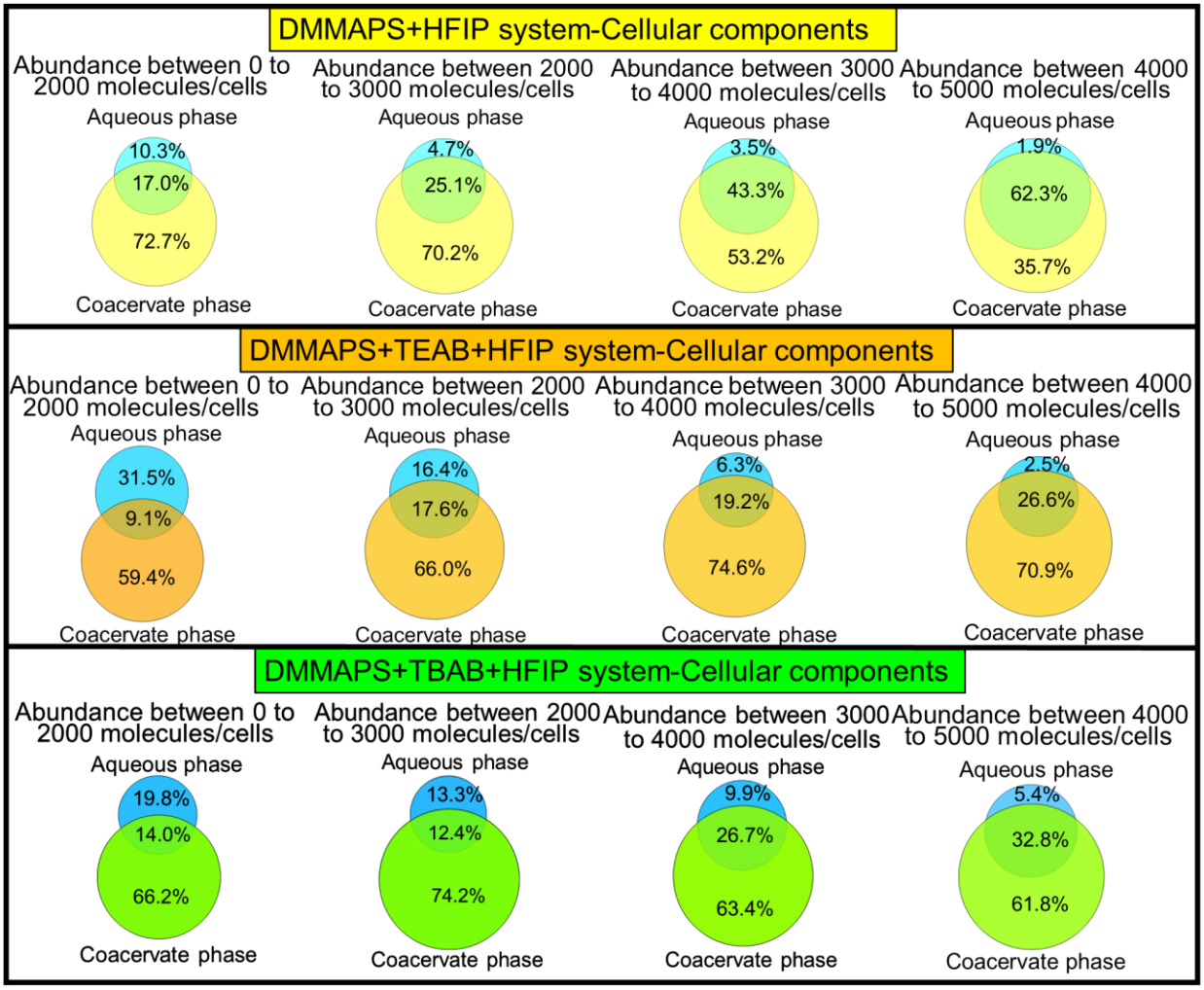


Figure 4: Venn diagrams that show the distribution and selective fractionation of low abundance proteins FA*i*C systems.

2.3.5. Sequence Coverage of α-Helices.

As shown in Figure 5, we identified a significantly larger number of proteins containing α-helices in the FA*i*C systems than the control system, and nearly all of these proteins were found in the coacervate phases. Transmembrane proteins contain α-helical peptide segments that traverse the lipid bilayer membranes and are highly hydrophobic.35,36 Better identification and more sequence coverage of the transmembrane proteins in the FA*i*C systems could be attributed to the higher solubilizing power and enrichment capability of the coacervate phases. We used the hidden

Markov model (http://www.cbs.dtu.dk/services/ TMHMM/) to theoretically determine the group of proteins that potentially contain α-helical transmembrane peptides.37 A list of the α-helical transmembrane proteins of yeast as well as the part of those proteins that have an α-helix structure has been prepared. The list was compared against the identified sequence of those proteins in our sample to calculate how much coverage we see on the α-helix part. We developed the algorithm in R program to identify the α-helical segments in the protein's sequences. The script and instruction for using this script are available in the Supporting Information in Section S10. The results of this analysis show the remarkable power of the FA*i*C (especially the DMMAPS system) in solubilizing α-helical proteins as compared to the control (Figure 5). A comparison between the number of α-helical proteins in the DMMAPS + TBAB + HFIP, DMMAPS + TEAB + HFIP, and DMMAPS + HFIP systems shows, respectively, 372, 555, and 583% identification improvement as compared to the control. In addition, sequence coverages of the α-helical transmembrane proteins are higher in the DMMAPS system. They are shown in Figure S3 (box chart, see the Supporting Information, Section S11), which illustrates the distribution of proteins with their α-helice sequence coverage into four quartiles. The chart shows median sequence coverage values of 21.7% and 20.5% of α-helices in DMMAPS + TEAB and DMMAPS + TBAB, respectively, with the lowest to highest value of 1.4−100%. In contrast, the control system only shows a median sequence coverage value of 6.9% for α-helices with the lowest to highest sequence coverage of 0.8−37.6%. The higher sequence coverage value provides more structural information about proteins and their functions. The list of the proteins with sequence coverage of α-helices for each system is included in the Supporting Information in Section S11.

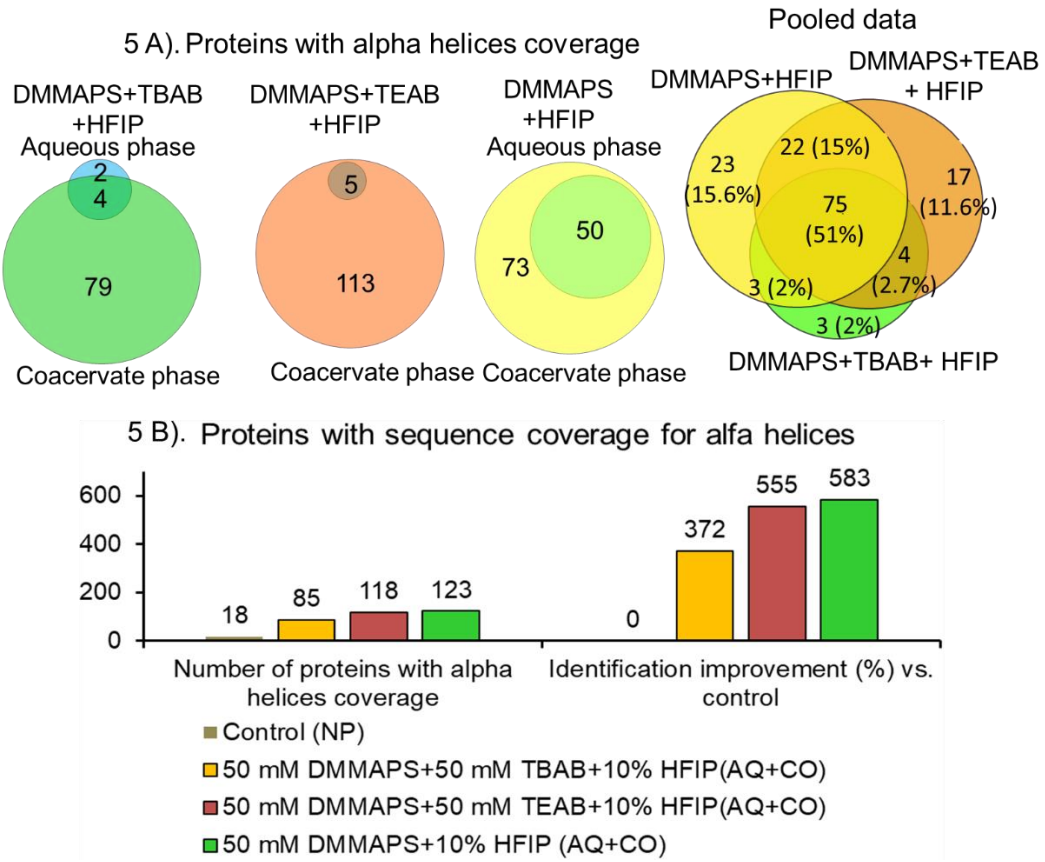


Figure 5 A): Venn diagrams showing the number of proteins with α-helice sequence identified in three FA*i*C systems, Fig 5 B). Identification improvement of proteins with α-helice sequence coverage vs. control.

2.3.6. Extraction of Post-translationally Modified Proteins and Their Sequence Coverage.

Extraction and characterization of post-translationally modified proteins is another important aspect of our study. There are different enrichment techniques used for the enrichment of phosphorylated proteins like metal affinity chromatography, but they are not cost effective [38,39]. The use of the coacervation approach is simple and economic, which leads to higher coverage for the post-translationally modified proteins especially phosphorylated proteins as compared to the control system. The information related to protein modification, their chemical structure, and their sequence coverage help to characterize proteins better. We analyzed the post-translational

modification of proteins based on the Yeast Mine database, as mentioned in the Experimental Section. The reference list of the yeast proteome with phosphorylated residues and ubiquitinylated lyside residues is given in the Supporting Information, Section S12. Figure 6 shows the total number of phosphorylated proteins and ubiquitinylated proteins identified using the FA*i*C and control systems. More than 200 phosphorylated proteins (which is 12.3−14.3% in different systems) and more than 100 proteins (which is 8−8.5% in different systems) with ubiquitinylated lyside residues were identified in the FA*i*C than the control system. The reason behind higher coverage of these proteins could be due to enrichment of lower-abundance proteins and fractionation. Table S12 in the Supporting Information, Section S13, shows that most of the lower abundance proteins identified in the FA*i*C and the control systems are phosphorylated.

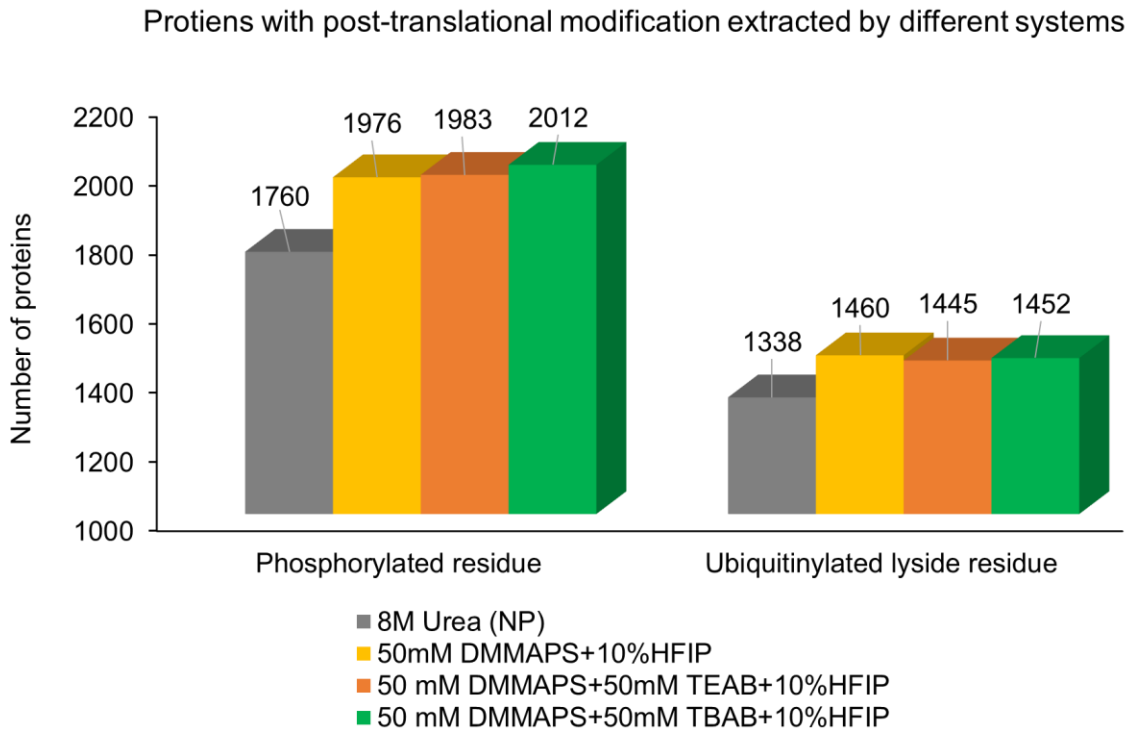


Figure 6. Coverage of post-translationally modified proteins by different FA*i*C and control systems.

2.3.7. Pooling Data from Three FA*i*C Systems.

Figure 7 shows the total number of reproducible proteins extracted by three different systems for the whole proteome, phosphorylated, membrane, and low-abundance proteins. As can be seen in Table S13 (see the Supporting Information, Section 14), the identification improvement increased from 18% (Table 1) to 33% for the whole proteome by pooling the data. Similarly, phosphorylated proteins show an increase in identification improvement from 14 to 26% in the combined systems. For the membrane proteins and low-abundance proteins, identification improvement increased from 19 to 34% and 114 to 245% from the individual to combined system, respectively. As can be seen from the Venn diagrams for the LAP, 58−66 uniquely identified LAP were found in each of the three FA*i*C systems that resulted in an increase in protein coverage from 257 to 414 for proteins with an abundance level of <2000 molecules/cell.

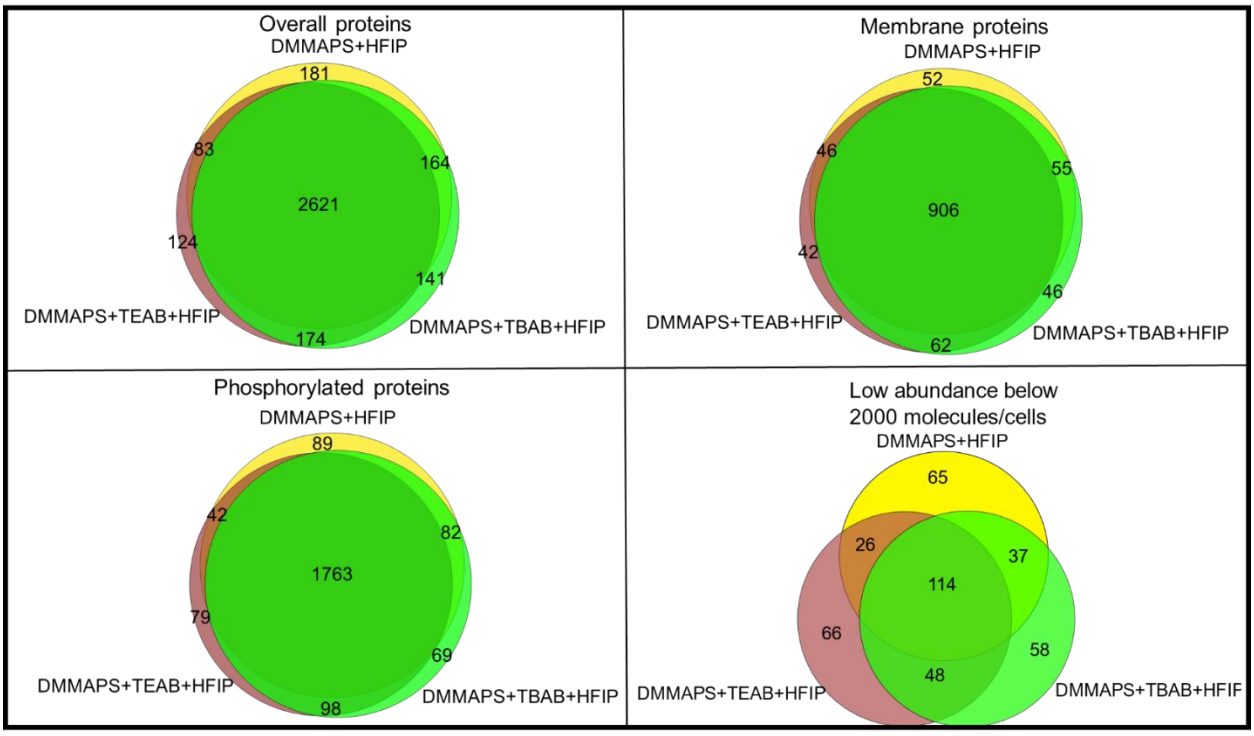


Figure 7. Pooling out of data from three different FA*i*C systems for whole yeast proteins, membrane proteins, phosphorylated proteins, and low abundance proteins.

2.4. CONCLUSIONS

The FA*i*C systems with mixed zwitterionic and cationic amphiphiles for extraction, fractionation, and enrichment of protein samples in bottom-up proteomics provided higher coverage of the whole proteome, subproteomic, and PTM proteins. Most notably were the significantly higher coverages of the membrane and low-abundance proteins. The larger coverage of proteins with α-helices in the FAiC is quite promising for increasing the accuracy of identification of integral membrane proteins based on the direct detection of their transmembrane peptide sequences.

AUTHOR INFORMATION

Corresponding Author Morteza G. Khaledi − Department of Chemistry and Biochemistry, The University of Texas at Arlington, Arlington, Texas 76019, United States; orcid.org/0000- 0001-9248-8178; Email: morteza.khaledi@uta.edu

Authors Durga Devi Khanal − Department of Chemistry and Biochemistry, The University of Texas at Arlington, Arlington, Texas 76019, United States, Sajad Tasharofi − Department of Chemistry and Biochemistry, The University of Texas at Arlington, Arlington, Texas 76019, United States, Mohammadmehdi Azizi − Department of Chemistry and Biochemistry, The University of Texas at Arlington, Arlington, Texas 76019, United States

ACKNOWLEDGMENTS

This research was supported by NSF grant CHE1632221.

2.5. REFERENCES

1. Vuckovic, D., et al., *Membrane proteomics by high performance liquid chromatography–tandem mass spectrometry: Analytical approaches and challenges.* PROTEOMICS, 2013. **13**(3-4): p. 404-423.

2. Vit, O. and J. Petrak, *Integral membrane proteins in proteomics. How to break open the black box?* Journal of Proteomics, 2017. **153**: p. 8-20.

3. Tan, S., H.T. Tan, and M.C. Chung, *Membrane proteins and membrane proteomics.* Proteomics, 2008. **8**(19): p. 3924-3932.

4. Hopkins, A.L. and C.R. Groom, *The druggable genome.* Nature reviews Drug discovery, 2002. **1**(9): p. 727-730.

5. Petschnigg, J., O.W. Moe, and I. Stagljar, *Using yeast as a model to study membrane proteins.* Current Opinion in Nephrology and Hypertension, 2011. **20**(4): p. 425-432.

6. Rucevic, M., D. Hixson, and D. Josic, *Mammalian plasma membrane proteins as potential biomarkers and drug targets.* ELECTROPHORESIS, 2011. **32**(13): p. 1549-1564.

7. Arinaminpathy, Y., et al., *Computational analysis of membrane proteins: the largest class of drug targets.* Drug discovery today, 2009. **14**(23-24): p. 1130-1135.

8. Karlsson, K., et al., *Enrichment of human brain proteins by heparin chromatography.* ELECTROPHORESIS, 1999. **20**(14): p. 2970-2976.

9. Zhang, Y., et al., *Enrichment of Low-Abundance Peptides and Proteins on Zeolite Nanocrystals for Direct MALDI-TOF MS Analysis.* Angewandte Chemie International Edition, 2005. **44**(4): p. 615-617.

10. Rawlings, Andrea E., *Membrane proteins: always an insoluble problem?* Biochemical Society Transactions, 2016. **44**(3): p. 790-795.

11. Moore, S.M., S.M. Hess, and J.W. Jorgenson, *Extraction, Enrichment, Solubilization, and Digestion Techniques for Membrane Proteomics.* Journal of Proteome Research, 2016. **15**(4): p. 1243-1252.

12. Waas, M., et al., *Combine and Conquer: Surfactants, Solvents, and Chaotropes for Robust Mass Spectrometry Based Analyses of Membrane Proteins.* Analytical Chemistry, 2014. **86**(3): p. 1551-1559.

13. Mesquita, R.O., et al., *Method optimization for proteomic analysis of soybean leaf: improvements in identification of new and low-abundance proteins.* Genetics and Molecular Biology, 2012. **35**: p. 353-361.

14. Ahmed, N. and G.E. Rice, *Strategies for revealing lower abundance proteins in two-dimensional protein maps.* Journal of Chromatography B, 2005. **815**(1-2): p. 39-50.

15. Gygi, S.P., et al., *Proteome analysis of low-abundance proteins using multidimensional chromatography and isotope-coded affinity tags.* Journal of proteome research, 2002. **1**(1): p. 47-54.

16. Erde, J., R.R.O. Loo, and J.A. Loo, *Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments.* Journal of proteome research, 2014. **13**(4): p. 1885-1895.

17. Zhang, Y., et al., *Exploration of Missing Proteins by a Combination Approach to Enrich the Low-Abundance Hydrophobic Proteins from Four Cancer Cell Lines.* Journal of Proteome Research, 2020. **19**(1): p. 401-408.

18. Arachea, B.T., et al., *Detergent selection for enhanced extraction of membrane proteins.* Protein Expression and Purification, 2012. **86**(1): p. 12-20.

19. Fang, F., et al., *Molecular Dynamics Simulation-assisted Ionic Liquid Screening for Deep Coverage Proteome Analysis.* Molecular & Cellular Proteomics, 2020. **19**(10): p. 1724-1737.

20. Zhao, Q., et al., *1-Dodecyl-3-methylimidazolium chloride-assisted sample preparation method for efficient integral membrane proteome analysis.* Analytical chemistry, 2014. **86**(15): p. 7544-7550.

21. Sui, Z., et al., *Ionic liquid-based method for direct proteome characterization of velvet antler cartilage.* Talanta, 2016. **161**: p. 541-546.

22. Koolivand, A., et al., *Coacervation of Lipid Bilayer in Natural Cell Membranes for Extraction, Fractionation, and Enrichment of Proteins in Proteomics Studies.* Journal of Proteome Research, 2019. **18**(4): p. 1595-1606.

23. Koolivand, A., et al., *Fluoroalcohol – Induced coacervates for selective enrichment and extraction of hydrophobic proteins.* Journal of Chromatography B, 2018. **1083**: p. 180-188.

24. Khaledi, M.G., S.I. Jenkins, and S. Liang, *Perfluorinated Alcohols and Acids Induce Coacervation in Aqueous Solutions of Amphiphiles.* Langmuir, 2013. **29**(8): p. 2458-2464.

25. McCord, J.P., D.C. Muddiman, and M.G. Khaledi, *Perfluorinated alcohol induced coacervates as extraction media for proteomic analysis.* Journal of Chromatography A, 2017. **1523**: p. 293-299.

26. de Jong, H.G.B. and H.R. Kruyt, *Koazervation.* Kolloid-Zeitschrift, 1930. **50**(1): p. 39-48.

27. Menger, F.M. and B.M. Sykes, *Anatomy of a coacervate.* Langmuir, 1998. **14**(15): p. 4131-4137.

28. Wang, M. and Y. Wang, *Development of surfactant coacervation in aqueous solution.* Soft Matter, 2014. **10**(40): p. 7909-7919.

29. Singh, S.S., et al., *Intermolecular complexation and phase separation in aqueous solutions of oppositely charged biopolymers.* International journal of biological macromolecules, 2007. **41**(2): p. 185-192.

30. Hellerstedt, S.T., et al., *Curated protein information in the Saccharomyces genome database*. Database, 2017. **2017**.

31. Song, G., et al., *Integration of new alternative reference strain genome sequences into the Saccharomyces genome database.* Database, 2016. **2016**.

32. Ghaemmaghami, S., et al., *Global analysis of protein expression in yeast.* Nature, 2003. **425**(6959): p. 737-741.

33. Kulak, N.A., et al., *Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells.* Nature methods, 2014. **11**(3): p. 319.

34. Ho, B., A. Baryshnikova, and G.W. Brown, *Unification of protein abundance datasets yields a quantitative Saccharomyces cerevisiae proteome.* Cell systems, 2018. **6**(2): p. 192-205. e3.

35. Ng, D.P., B.E. Poulsen, and C.M. Deber, *Membrane protein misassembly in disease.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2012. **1818**(4): p. 1115-1122.

36. Cserzö, M., et al., *Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method.* Protein engineering, 1997. **10**(6): p. 673-676.

37. Sonnhammer, E.L., G. Von Heijne, and A. Krogh. *A hidden Markov model for predicting transmembrane helices in protein sequences*. in *Ismb*. 1998.

38. Oda, Y., T. Nagasu, and B.T. Chait, *Enrichment analysis of phosphorylated proteins as a tool for probing the phosphoproteome.* Nature biotechnology, 2001. **19**(4): p. 379-382.

39. Wolschin, F., S. Wienkoop, and W. Weckwerth, *Enrichment of phosphorylated proteins and peptides from complex mixtures using metal oxide/hydroxide affinity chromatography (MOAC).* Proteomics, 2005. **5**(17): p. 4389-4397.

CHAPTER 3

Improved Extraction, Enrichment, and Fractionation of Yeast Proteome in Bottom-Up Proteomics using Fluoroalcohol-Mediated Supramolecular Biphasic Systems with Mixed Amphiphiles for Sample Extraction, Fractionation, and Enrichment.

Used with permission from Durga Devi Khanal, Mohammad Mehdi Azizi, Sajad Tasharofi, Morteza G. Khaledi*

The University of Texas at Arlington

Department of Chemistry and Biochemistry, Arlington, TX, 76019

ABSTRACT

Fluoroalcohol Induced Coacervation Biphasic Systems (FA*iC*-BPS) is a recent and novel approach in sample preparation in proteomics for the purpose of solubilization, fractionation, and enrichment of complex protein mixture. In this report, we introduce a new biphasic supramolecular system that shows increased protein coverage for overall yeast proteomes, especially for membrane proteins and lower abundance proteins, as compared to conventional approach of no phase separation using surfactants and amphiphiles. Hexafluoro isopropanol (HFIP) induced biphasic systems of 3-[(3-Cholamidopropyl) dimethylammonio]-1-propane sulfonate (CHAPS), as a zwitterionic surfactant and its mixed amphiphilic systems with small chain length quaternary ammonium salts (QUATS) is highly effective for the solubilization, fractionation and enrichment of yeast proteins. Using these FA*iC*-BPS in the bottom-up proteomics workflow enabled fractionation of complex sample mixture into two phases and enrichment of low abundance proteins that resulted in improved coverage of low abundance proteins by as much as 80% as compared to using conventional methods for sample solubilization using urea or detergents

solutions (no phase separation). Similarly, higher coverage of phosphorylated and ubiquitinylated proteins was improved as compared to the control systems.

Keywords: Biphasic Systems, Yeast Proteomics, Membrane Proteins, Low Abundance proteins, Fluoroalcohols

3.1. INTRODUCTION

Extraction and purification of membrane proteins from complex protein mixture pose a significant challenge owing to their lower solubility in aqueous media [1-6]. The transmembrane portion of proteins are also less prone to tryptic digestion in bottom-up proteomics; thus, membrane proteins are often identified based on detection of peptides segments residing outside of the lipid bilayer. Due to their innate availability in very low concentrations, identification of lower abundance proteins under the shadow of higher abundance proteins is challenging [7, 8]. Therefore, development of effective sample preparation techniques for solubilization and enrichment of these proteins facilitates their identification and quantitation in proteomic analysis.

Surfactants and amphiphilic reagents are commonly used chemicals for solubilization and extraction of membrane proteins, they mimic the lipid environment and thus, promote solubilization of the membrane proteins. Most detergents trigger denaturation and unfolding of proteins resulting in exposure to enzymatic digestion [9]. Negatively charged surfactants like Sodium Dodecyl Sulfate (SDS), Sodium Cholate (SC), and Sodium Deoxy Cholate (SDC) are very popular for membrane proteins solubilization. Similarly, positively charged surfactants like Cetyl Trimethyl Ammonium Bromide (CTAB), Dodecyl Trimethyl Ammonium Bromide (DTAB), and zwitterionic surfactants like 3-[(3-Cholamidopropyl) dimethylammonio]-1-propane sulfonate (CHAPS), Cholamidopropyl) dimethylammonium)-2-hydroxy-1-propane sulfonate (CHAPSO)),

or 3- (N, N-Dimethyl myristyl ammonia) propane sulfonate (DMMAPS), and chaotropic reagents like urea have also used for solubilization in proteomics [4, 10-13].

Our laboratory first introduced novel fluoroalcohol-induced coacervation biphasic systems (FA*i*C -BPS)[14-16]. Coacervation is a process of assembly of amphiphile molecules in aqueous media and formation of a separate amphiphile-rich phase from the bulk aqueous phase [17]. We showed that water-miscible fluoroalcohols such as hexafluoro isopropanol (HFIP) or trifluoroethanol (TFE) can significantly facilitate coacervation of amphiphilic compounds in aqueous media and subsequently demonstrated the usefulness of FA*i*C -BPS for sample solubilization, fractionation, and enrichment of proteins samples in proteomics applications [18-22]. We have observed that using FA*i*C-BPS results in increased solubilization and enrichment for hydrophobic and lower abundance membrane proteins in one of the two phases in FA*i*C-BPS. Complex protein mixtures are fractionated and enriched in the FA*i*C-BPS based on hydrophobic and electrostatic interactions. The nature of protein fractionation and enrichment patterns depend on the type of amphiphiles in FA*i*C-BPS. In previous studies, we investigated various FA*i*C systems using anionic, cationic, and zwitterionic surfactants with long alkyl chains. We also reported coacervation of lipids in natural cell membranes [20]. In this study, we examined hexafluoro isopropanol induced coacervation of CHAPS and the application of the CHAPS based FA*i*C-BPS in bottom-up proteomics of yeast.

3.2. Experimental section

3.2.1. Chemicals and Reagents: 3-[(3-Cholamidopropyl) dimethylammonio]-1-propane sulfonate (CHAPS), a zwitterionic surfactant was purchased from Fisher Scientific, USA. Small chain amphiphiles like quaternary ammonium bromide salts: tetraethylammonium bromide

(TEAB), tetrabutylammonium bromide (TBAB) were purchased from ACROS Organics, USA. 1,1,1,3,3,3-Hexafluoro-2-propanol (HFIP) was obtained from Oakwood Chemical, USA. Trifluoroacetic acid (99%) and formic acid (99%) were purchased from Alfa Aesar, USA. Chemicals dithiothreitol (DTT), iodoacetamide (IAA), and sequencing grade trypsin required for protein digestion were purchased from Promega Corporation, 2800 Woods Hollow Road, Madison, WI, USA. Organic solvents like acetonitrile (ACN, LC-MS grade) and isopropanol (IPA-LC-MS grade) were purchased from Fisher Chemical, USA. Millipore-DI was water used for sample preparation and HPLC analysis from the lab.

3.2.2. Phase Transition Behavior of CHAPS Coacervation with Hexafluoro Isopropanol: Phase transition behavior of CHAPS was studied with hexafluoro isopropanol at different concentrations of surfactants and HFIP. The total volume of the mixture was kept constant (1000 μL). The phase transition behavior of CHAPS with HFIP and water is shown in **Figure 1**. As the concentration of surfactant (CHAPS) is increased, the volume of the coacervate phase is also increased linearly. An increase in coacervate volume with the linear increase of surfactant concentration is the characteristic of coacervation. Phase transition behavior basically provides the information about concentration range at which surfactant can form two-phase with HFIP. The minimum concentration of CHAPS used in this experiment is 25 mM of CHAPS which forms two phases from 6 to 22% of HFIP. However, as the concentration of CHAPS increases coacervation range is increased with HFIP.

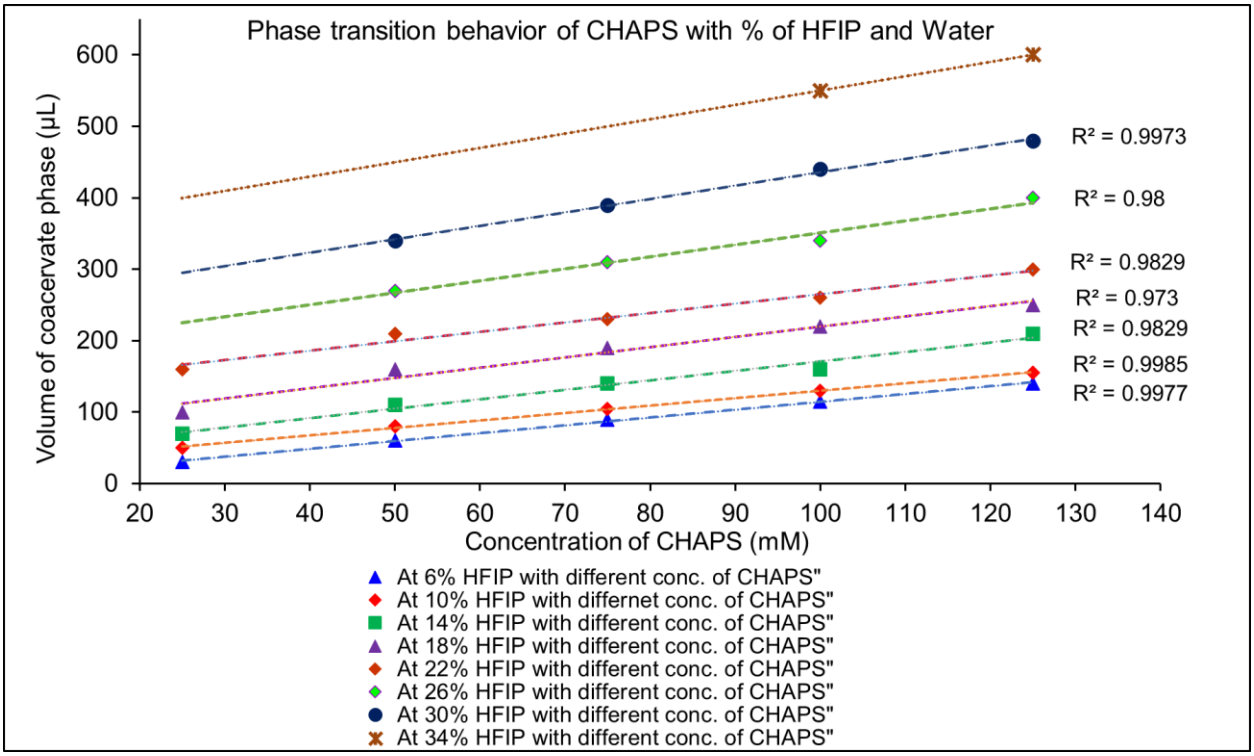


Figure 1: Change in volume of coacervate phase of different concentration of CHAPS at different % of HFIP and water.

3.2.3. Cell Lysis and Coacervation: Yeast cells of species *Saccharomyces cerevisiae* was used throughout the experiment. The detailed procedure of cell growth and cell lysis is given in supplementary information in Section S1. After cell lysis, the lysed cells were kept in -80 ºC freezer for future use. Protein extraction was done with 50 µL of lysed cells (equivalent to 400 µg of protein) and they were mixed with certain concentration of the amphiphile (CHAPS and CHAPS+QUATS), HFIP, and water for coacervation. The total volume of the mixture was kept at 1 mL. The mixture was then vortexed for 2 min, sonication for 3 min and centrifuged for 15 min at the speed of 10,000 g. The two phases were separated using micropipettes. The coacervate phase was evaporated with minimum nitrogen gas flow to remove HFIP. Surfactants and amphiphiles were removed before digestion using filter-aided sample preparation (FASP) approach. The details about FASP protocol are added in the supplementary information in Section S2. Extracted

proteins in each of the two phases (aqueous and coacervate) were digested on the FASP filter using tryptic enzyme for bottom-up proteomic analysis of proteins. Control experiments followed the same procedure using aqueous solutions of different surfactants and urea as solubilizing agents. Each experiment was run in triplicates and two of the three common proteins were selected as the reproducible proteins.

3.2.4. Reserved Phase Chromatography and Data Analysis

The tryptic digest mixtures from the aqueous and coacervate phases were analyzed using Orbitrap Fusion Lumos MS® instrument in a bottom-up approach. Instrument information: Ultimate 3000 RSLC-Nano liquid chromatography systems, Dionex; coupled with Orbitrap Fusion Lumos MS®, Thermo Electron instrument, a reversed-phase C18 nano column with column dimension 75 cm x 75 µm, particle size 3 µm, injection volume of 2 µL, solvent system: 0-90 min gradient run with 0-28% of solvent B with a flow rate of 350 µL/min. Mobile phases: solvent A was 2% (V/V) Acetonitrile (ACN) and 0.1% (V/V) formic acid (FA) in water, and mobile phase B was 80% (V/V) ACN, 10% (V/V) trifluoroethanol (TFE), and 0.1% FA in water.

Raw data obtained from LC-MS analysis were analyzed using MaxQuant software where protein IDs were retrieved as UniProt IDs. The set-up for MaxQuant is same as in our previous published article [21]. They were further analyzed using Perseus software to remove any potential contaminants. Three replicates were analyzed, and the common proteins identified in at least two out of three runs were accepted as reproducible results. Furthermore, detailed data analysis was done based on Saccharomyces Genome Database (SGD) IDs of proteins using the Yeast Mine database (https://yeastmine.yeastgenome.org/yeastmine/begin.do) and Gene Ontology database (http://geneontology.org/). The SGD IDs of proteins were obtained using the UniProt database (https://www.uniprot.org/) based on UniProt IDs of extracted proteins.

3.3. RESULTS AND DISCUSSION

3.3.1. Bottom-Up Proteomics of Yeast Proteomes using FAiC-BPS Systems CHAPS and Mixed Amphiphiles

Three FA*i*C-BPS with compositions of (50 mM CHAPS+ 10%HFIP), (50 mM CHAPS+ 50 mM TEAB+ 10% HFIP), and (50 mM CHAPS+ 50 mM TBAB+10% HFIP) were used for the extraction, fractionation, and enrichment of yeast proteins in bottom-up proteomics workflow. The three replicates were performed for each FA*i*C-BPS and the control systems. The control systems included aqueous solutions of 50 mM CHAPS, 50 mM DMMAPS, 50 mM SC, 50 mM SDS, 50 mM TEAB, 50 mM TBAB and 8M urea. A comparative study between the FA*i*C-BPS and the control systems would reveal the effects of using a two-phase system that enables sample fractionation and enrichment as compared to just using a single-phase solution for protein solubilization. As shown in a previous study, 8M Urea was found to be the most effective that showed the largest number of identified proteins.

The total number of proteins extracted in each FA*i*C-BPS system and that in urea as the best control is shown in Fig. 2. The list of proteins extracted into the aqueous phase and coacervate phase of each system and controls are available in the supplementary information in Section S 3**.** The Venn diagram in Fig. 2 depict the number of unique proteins identified either in the aqueous phase or in the coacervate phase; those that were identified in both phases, the total number of proteins identified in the FA*i*C-BPS, and identification improvement as compared to control. The result shows that CHAPS and mixed amphiphilic coacervation systems extracted more than 300 proteins as compared to control. The highest protein coverage was obtained from mixed CHAPS+TBAB system with the identification improvement of 16.5% (431 proteins) than the control. Additionally, the presence of both TEAB and TBAB to CHAPS resulted in greater protein coverage than the

CHAPS system. Similarly, **Figure 2,** Venn diagram depicts the fractionation of proteins into aqueous phase and coacervate phase for each different system. Interestingly, the addition of TEAB to CHAPS gave the best fractionation of proteins into two phases with highest number of uniquely identified proteins in each phase. However, addition of TBAB to CHAPS extracted a greater number of uniquely identified proteins into an aqueous phase as compared to other system.

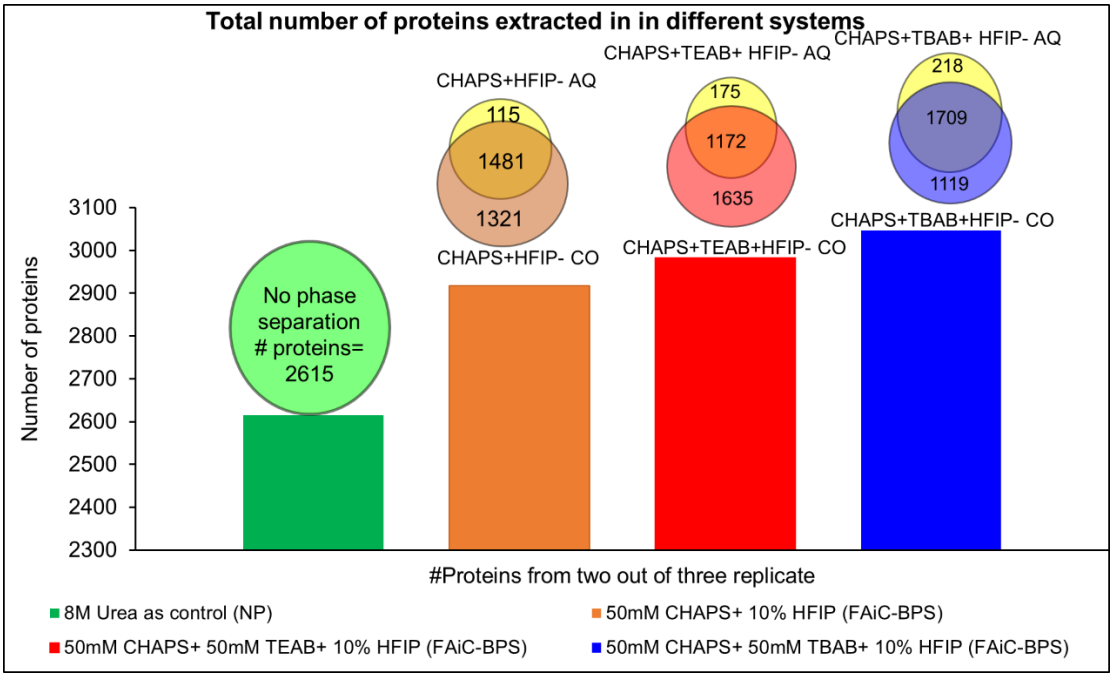


Figure 2: Total number of proteins extracted into three FA*i*C-BPS systems and 8M Urea as a control without phase separation (NP).

3.3.2. Subcellular Proteomics: Gene Ontology was used to identify yeast proteins at the subcellular level. **Table 1** shows the number of proteins in different subcellular locations of yeast proteome by different coacervation systems. The number of proteins extracted from each subcellular locations were compared with control system of 8M urea (NP). Coacervation systems outperformed the control system in identifying most of the protein types, especially those located or associate with cell membranes and mitochondria. The identification improvement of sub-

cellular proteins follows the order of: CHAPS+TBAB+HFIP > CHAPS+TEAB+HFIP ≥ CHAPS+HFIP > 8M Urea (NP). Higher protein coverage using FA*i*C-BPS can be attributed to their capability for sample enrichment and fractionation. These results further indicate the that the mixed amphiphilic coacervation systems provide higher protein coverage than the single amphiphile CHAPS coaceravtes.

Figure 3 illustrates protein fractionation patterns for select subcellular proteins in three FA*i*C-BPS with different amphiphile compositions. Addition of TEAB to CHAPS had a positive effect on fractionating more hydrophobic membrane proteins, while the addition of TBAB to CHAPS enhanced the fractionation of more hydrophilic proteins like mitochondrial ribosome proteins and ribosome proteins into aqueous phase. This observation clearly demonstrates selectivity effects of mixed amphiphile coacervate systems in protein fractionation. Structurally, TEAB and TBAB differ only in hydrophobic chain length, but their presence in coacervate system can significantly change the electrostatic interaction between the phase and the proteins [21].

Table 1: Identification improvement of subcellular yeast proteins (cellular components) by three FAiC-TPS systems: 50 mM CHAPS+ 10% HFIP, 50 mM CHAPS+ 50 mM TEAB+ 10% HFIP, and 50 mM CHAPS+ 50 mM TBAB+ 10% HFIP with respect to control.

GO Cellular components	Control	50 mM CHAPS+ 10% HFIP	Increase/ decrease	50 mM CHAPS+50 mM TEAB+ 10% HFIP	Increase/ decrease	50 mM CHAPS+ 50 mM TBAB+ 10% HFIP	Increase/ decrease
Membrane type proteins							
Membrane	885	1002	13.2%	1012	14.3%	1032	16.6%
Integral component of membrane	427	493	15.5%	493	15.5%	504	18.0%
Intrinsic component of membrane	444	516	16.2%	515	15.9%	522	17.6%
Integral component of	103	118	14.5%	114	10.7%	121	17.5%

organelle membrane							
Integral component of endoplasmic reticulum membrane	43	50	16.3%	50	16.3%	51	18.6%
Integral component of plasma membrane	12	14	16.7%	15	25%	13	8.3%
plasma membrane	240	267	11.2%	273	13.7%	270	12.5%
Mitochondrial protein types							
mitochondrial membrane	185	210	13.5%	211	14.1%	230	24.3%
mitochondrial matrix	121	138	14.0%	139	14.9%	162	33.9%
mitochondrial ribosome	37	52	40.5%	51	37.8%	67	81.1%
mitochondrial envelope	210	237	12.8%	240	14.3%	262	24.8%
Other cellular component types							
ribosome	128	146	14.1%	146	14.1%	161	25.8%
endosome	95	113	18.9%	116	22.1%	118	24.2%
chromosome	215	241	12.%	242	12.6%	245	14.0%
nucleolus	218	232	6.4%	229	5.0%	238	9.2%
vacuole	195	224	14.9%	231	18.5%	239	22.6%
vesicle	176	201	14.2%	206	17.0%	212	20.4%
endoplasmic reticulum	364	408	12.1%	402	10.4%	404	11.0%
Golgi membrane	104	110	5.8%	109	4.8%	110	5.8%
Golgi apparatus	187	204	9.1%	198	4.9%	207	10.7%

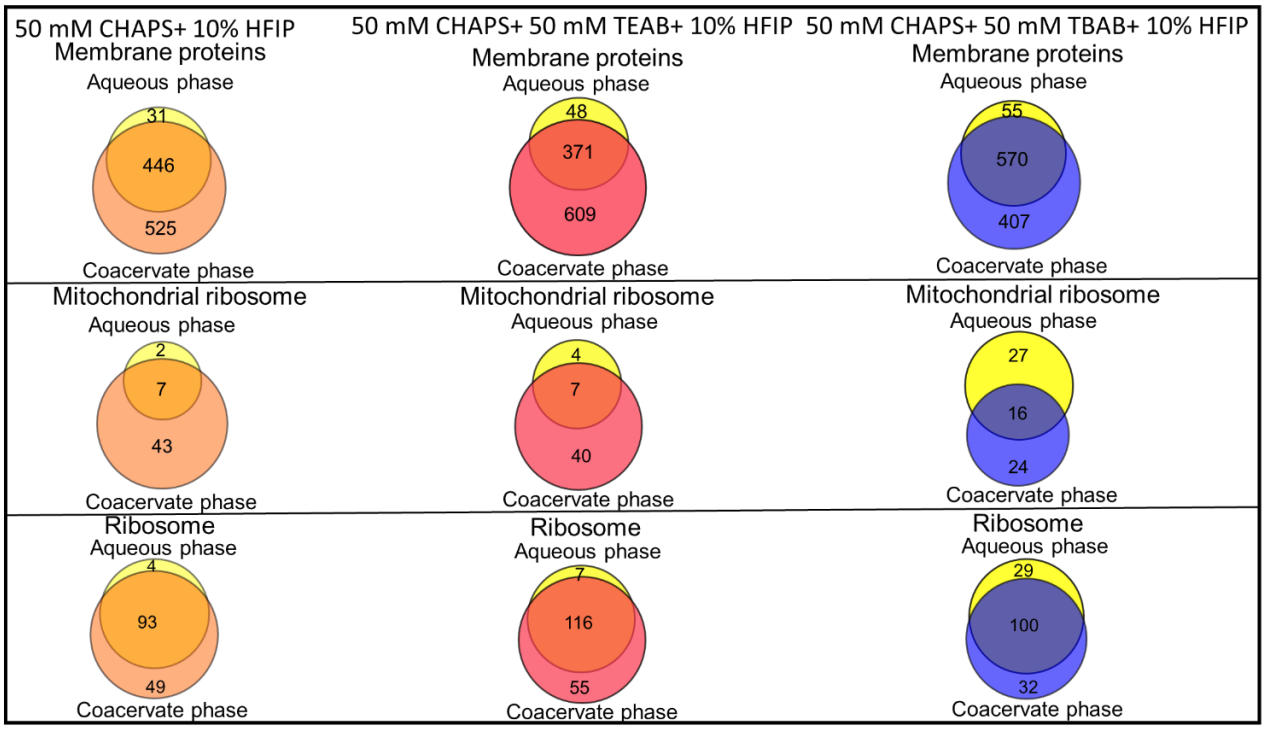


Figure 3: Subcellular proteins showing a change in fractionation by three different coacervation approaches

3.3.3. Extraction of Lower Abundance Proteins (LAP): Yeast proteome comprises majority of lower abundance proteins than higher abundance proteins and therefore they are hard to detect by a single experiment. **Figure 4** is the abundance chart of proteins extracted by three biphasic systems and control with 8M urea (NP). The X-axis in the chart represents the number of proteins extracted and Y-axis represents the abundance range of these proteins in molecules/cells. As can be seen, the extraction of lower abundance proteins is much higher in biphasic systems as compared to the control. At abundance value of 5000 molecules/cells or above, there is little or no difference in the number of proteins identified in FAiC-BPS or the control system. Likewise**,** **Figure 5** represents the plot for identification improvement of proteins in biphasic systems based on their abundance value as compared to the control. X-axis shows the abundance range in molecules/cells, and Y-axis shows identification improvement in percentages. All biphasic

systems show significant identification improvement for lower abundance proteins than control system. Among biphasic system, the mixed amphiphilic systems of CHAPS with TEAB and TBAB have the highest identification improvement value of 78.8% and 80.0% respectively, at abundance level of 0-2000 molecules/cells. [Note: The reference list of the proteins with their abundance value is retrieved from the Yeast mine database].

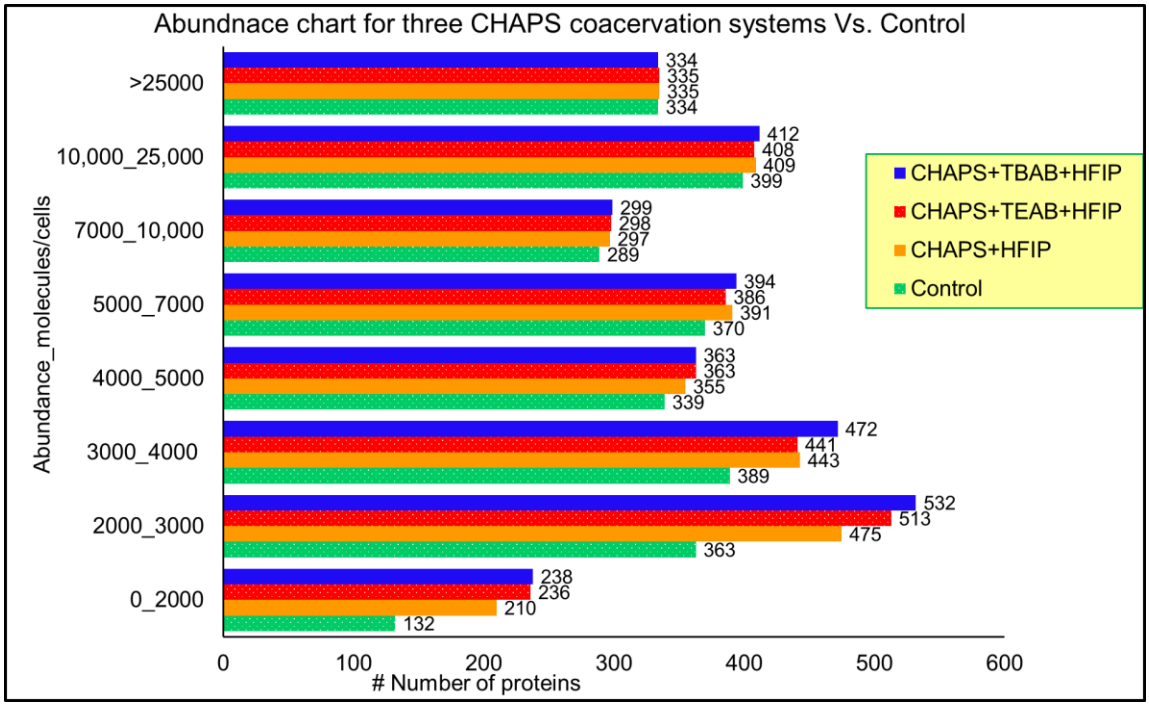


Figure 4: Abundance chart of proteins extracted by CHAPS+HFIP and CHAPS+QUATS+HFIP FAiC systems Vs. 8M Urea as control (NP).

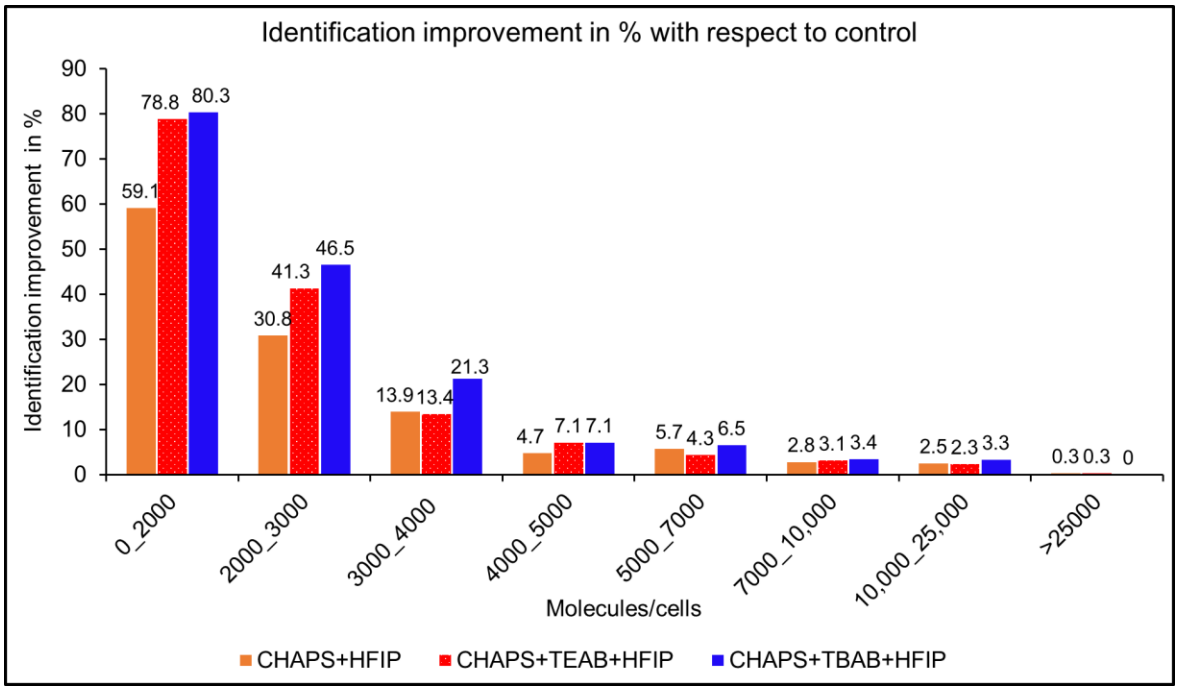


Figure 5: Identification improvement of yeast proteins with a variety of abundance range Vs. 8M Urea as control (NP).

3. Extraction of Proteins with Post-translational Modifications

Detection of phosphorylated proteins has a great significance since they control the major regulatory mechanism of cells [23, 24]. The coacervation approach is helpful for the extractions of higher number of post-translationally modified proteins like phosphorylated proteins and ubiquitinylated proteins. **Figure 6** shows the extraction of higher number of phosphorylated and ubiquitinylated proteins by biphasic systems of CHAPS and mixed amphiphiles (CHAPS+QUATS) compared to control. Table ST1 in the Supplementary Information Section S5 shows the list of phosphorylated proteins and identification improvement as compared to the control. Identification improvements of 11.5%, 14.4%, and 13.7% versus control were observed for phosphorylated proteins using CHAPS+HFIP, CHAPS+TEAB+HFIP, and CHAPS+TBAB+HFIP systems respectively. Similarly, the coverage of the ubiquitinylated

proteins was about 7% higher in FAiC-BPS as compared to the control. The FAiC-BPS with mixed amphiphiles provided better coverage of both the phosphorylated and ubiquitinylated proteins than the single amphiphile CHAPS system.

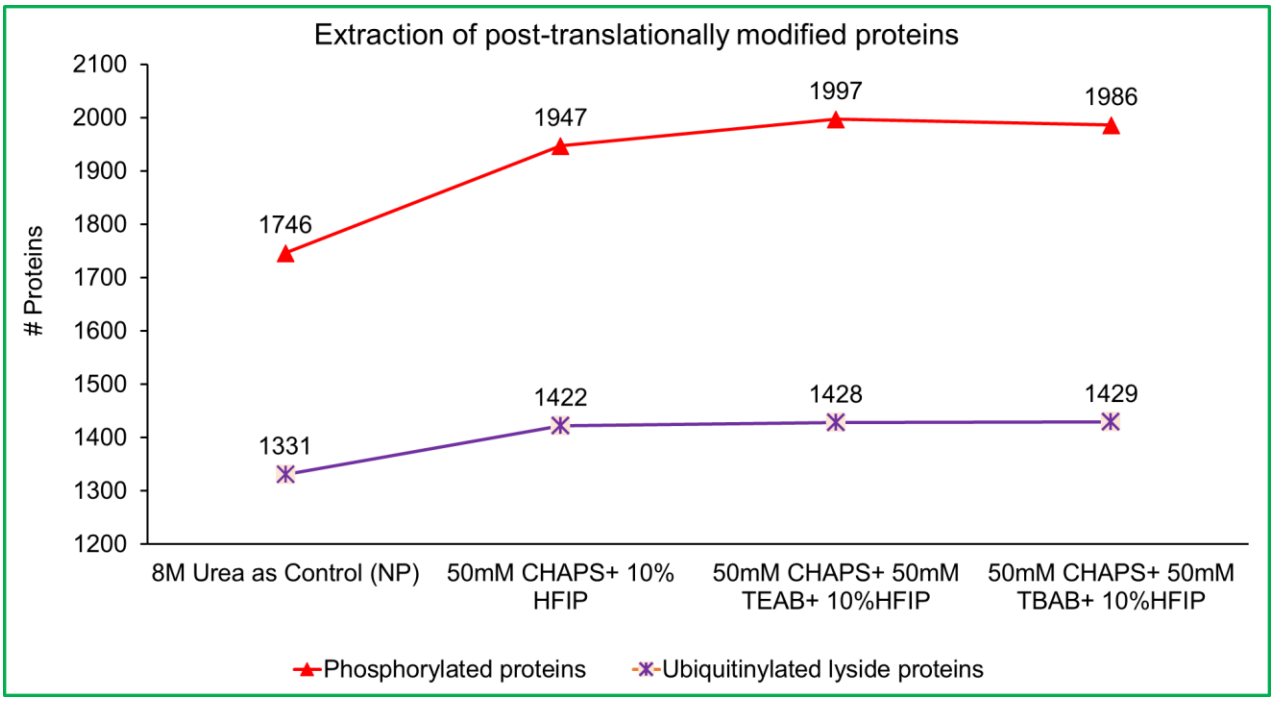


Figure 6: Extraction of total number of phosphorylated proteins and ubiquitinylated proteins in three biphasic systems of CHAPS+HFIP and CHAPS+QUATS+HFIP vs. 8M Urea as control (NP).

4. Pooled Data from three Different Systems

Figure 7 shows the Venn diagrams for the LAP identified in the three FAiC-BPS. The number of unique LAP identified in CHAPS+HFIP, CHAPS+TEAB+HFIP, and CHAPS+TBAB+HFIP were 36, 58, and 58 respectively. The pooled results using the three FAiC-BP systems resulted in identification of 357 extra LAP (<2000 molecules/cell) as compared to the urea control system. In addition, the box chart in Figure 7 illustrates the shows the patterns of the uniquely identified LAP the three biphasic systems in terms of proteins isoelectric points and GRAVY scores. The

shift the pI values from box charts show that more basic lower abundance proteins are extracted into mixed amphiphilic (CHAPS+QUATS) systems than in the CHAPS+HFIP biphasic system. In contrast, slightly more hydrophobic lower abundance proteins are extracted into CHAPS+HFIP system than into CHAPS+QUATS systems (from the median value). However, CHAPS+TBAB system shows the extraction of protein with highest and lowest GRAVY value among all.

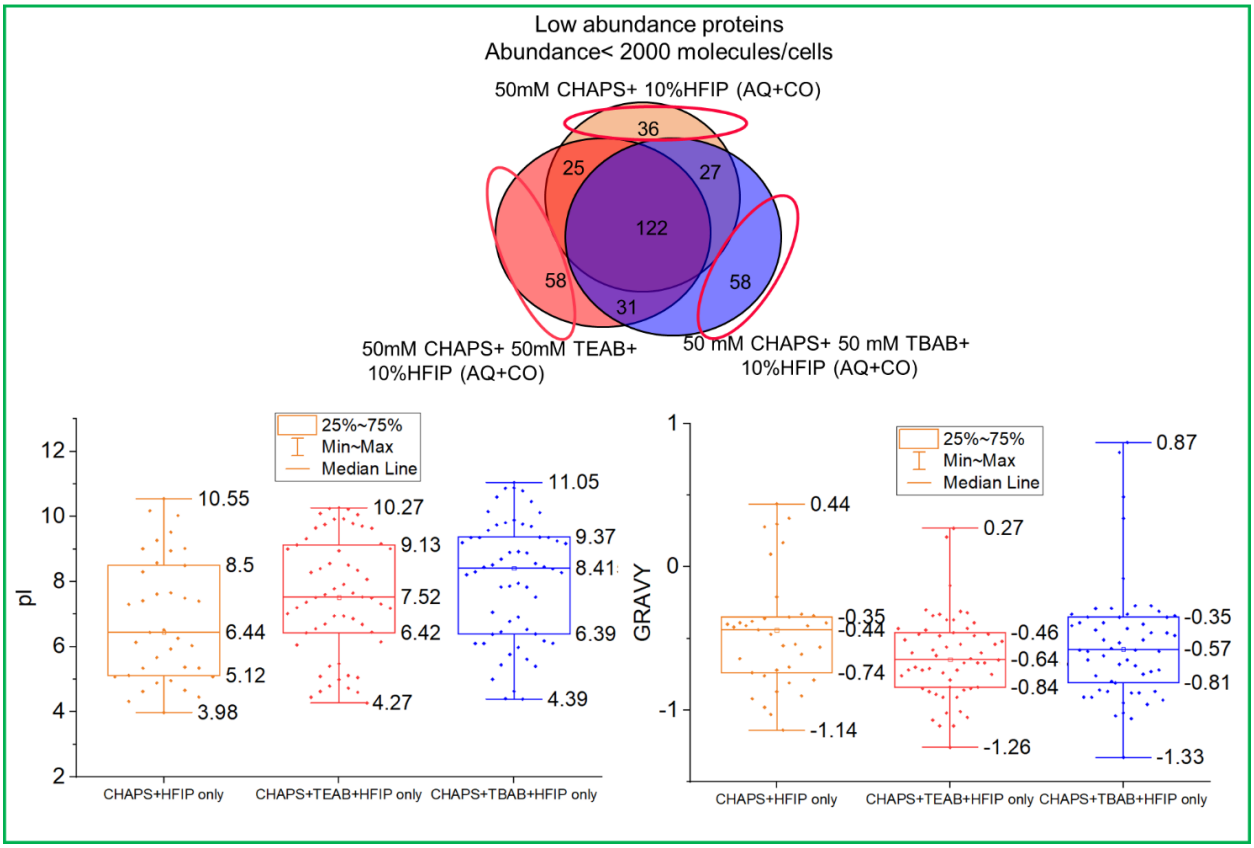


Figure 7: Selectivity difference for low abundance proteins between CHAPS+HFIP, and mixed amphiphilic biphasic systems of CHAPS+QUATS shown based on their isoelectric point (pI) and hydrophobicity (GRAVY) value.

3.4. CONCLUSION

In this research, we have reported the biphasic systems of CHAPS and mixed amphiphiles for the improved coverage, solubilization, enrichment and fractionation of membrane, lower abundance

proteins, and phosphorylated proteins as compared to the conventional approach of no phase separation. This is a simple and economically viable technique and can be used in the proteomic labs for the extraction of different types of proteins from complex sample mixture. Introduction of small chain quaternary ammonium salts to zwitterionic surfactants in biphasic extraction of proteins shows the increased extraction of overall proteins, most importantly membrane and lower abundance proteins is the interesting part of this research.

ACKNOWLEDGMENTS

This research was supported by NSF grant CHE1632221.

3.5. REFERENCES

1. Nfor, B.K., et al., *Design strategies for integrated protein purification processes: challenges, progress and outlook.* Journal of Chemical Technology & Biotechnology, 2008. **83**(2): p. 124-132.
2. Bleakley, S. and M. Hayes, *Algal Proteins: Extraction, Application, and Challenges Concerning Production.* Foods, 2017. **6**(5): p. 33.
3. Morrison, K.A., et al., *Membrane protein extraction and purification using styrene–maleic acid (SMA) copolymer: effect of variations in polymer structure.* Biochemical Journal, 2016. **473**(23): p. 4349-4360.
4. Arachea, B.T., et al., *Detergent selection for enhanced extraction of membrane proteins.* Protein Expression and Purification, 2012. **86**(1): p. 12-20.
5. Vuckovic, D., et al., *Membrane proteomics by high performance liquid chromatography–tandem mass spectrometry: Analytical approaches and challenges.* PROTEOMICS, 2013. **13**(3-4): p. 404-423.

6. Vit, O. and J. Petrak, *Integral membrane proteins in proteomics. How to break open the black box?* Journal of Proteomics, 2017. **153**: p. 8-20.

7. Mesquita, R.O., et al., *Method optimization for proteomic analysis of soybean leaf: improvements in identification of new and low-abundance proteins.* Genetics and Molecular Biology, 2012. **35**: p. 353-361.

8. Isaacson, T., et al., *Sample extraction techniques for enhanced proteomic analysis of plant tissues.* Nature Protocols, 2006. **1**(2): p. 769-774.

9. Waas, M., et al., *Combine and Conquer: Surfactants, Solvents, and Chaotropes for Robust Mass Spectrometry Based Analyses of Membrane Proteins.* Analytical Chemistry, 2014. **86**(3): p. 1551-1559.

10. Salager, J.-L., *Surfactants types and uses.* FIRP booklet, 2002. **300**.

11. Gonenne, A. and R. Ernst, *Solubilization of membrane proteins by sulfobetaines, novel zwitterionic surfactants.* Analytical Biochemistry, 1978. **87**(1): p. 28-38.

12. Möglich, A., F. Krieger, and T. Kiefhaber, *Molecular Basis for the Effect of Urea and Guanidinium Chloride on the Dynamics of Unfolded Polypeptide Chains.* Journal of Molecular Biology, 2005. **345**(1): p. 153-162.

13. Saitoh, T. and W.L. Hinze, *Concentration of hydrophobic organic compounds and extraction of protein using alkylammoniosulfate zwitterionic surfactant mediated phase separations (cloud point extractions).* Analytical Chemistry, 1991. **63**(21): p. 2520-2525.

14. Khaledi, M.G., S.I. Jenkins, and S. Liang, *Perfluorinated Alcohols and Acids Induce Coacervation in Aqueous Solutions of Amphiphiles.* Langmuir, 2013. **29**(8): p. 2458-2464.

15. Nejati, M.M. and M.G. Khaledi, *Perfluoro-Alcohol-Induced Complex Coacervates of Polyelectrolyte–Surfactant Mixtures: Phase Behavior and Analysis.* Langmuir, 2015. **31**(20): p. 5580-5589.

16. Jenkins, S.I., C.M. Collins, and M.G. Khaledi, *Perfluorinated alcohols induce complex coacervation in mixed surfactants.* Langmuir, 2016. **32**(10): p. 2321-2330.

17. Chollakup, R., et al., *Phase Behavior and Coacervation of Aqueous Poly(acrylic acid)−Poly(allylamine) Solutions.* Macromolecules, 2010. **43**(5): p. 2518-2528.

18. McCord, J.P., D.C. Muddiman, and M.G. Khaledi, *Perfluorinated alcohol induced coacervates as extraction media for proteomic analysis.* Journal of Chromatography A, 2017. **1523**: p. 293-299.

19. Koolivand, A., et al., *Fluoroalcohol – Induced coacervates for selective enrichment and extraction of hydrophobic proteins.* Journal of Chromatography B, 2018. **1083**: p. 180-188.

20. Koolivand, A., et al., *Coacervation of Lipid Bilayer in Natural Cell Membranes for Extraction, Fractionation, and Enrichment of Proteins in Proteomics Studies.* Journal of Proteome Research, 2019. **18**(4): p. 1595-1606.

21. Khanal, D.D., et al., *Improved Protein Coverage in Bottom-Up Proteomes Analysis Using Fluoroalcohol-Mediated Supramolecular Biphasic Systems With Mixed Amphiphiles for Sample Extraction, Fractionation, and Enrichment.* Analytical Chemistry, 2021. **93**(20): p. 7430-7438.

22. Azizi, M., et al., *Improving identification of low abundance and hydrophobic proteins using fluoroalcohol mediated supramolecular biphasic systems with quaternary ammonium salts.* Journal of Chromatography A, 2021. **1655**: p. 462483.

23. Ptacek, J., et al., *Global analysis of protein phosphorylation in yeast.* Nature, 2005. **438**(7068): p. 679-684.

24. Vlastaridis, P., et al., *Estimating the total number of phosphoproteins and phosphorylation sites in eukaryotic proteomes.* Gigascience, 2017. **6**(2): p. giw015.

CHAPTER 4

Study on Electrostatic Effect of Small Chain Quaternary Ammonium Salts in Supramolecular Biphasic System of Zwitterionic Surfactants for Proteins Fractionation.

Used with the permission from Durga Devi Khanal, Sajad Tasharofi, Mohammadmehdi Azizi, Morteza G. Khaledi*

The University of Texas at Arlington,

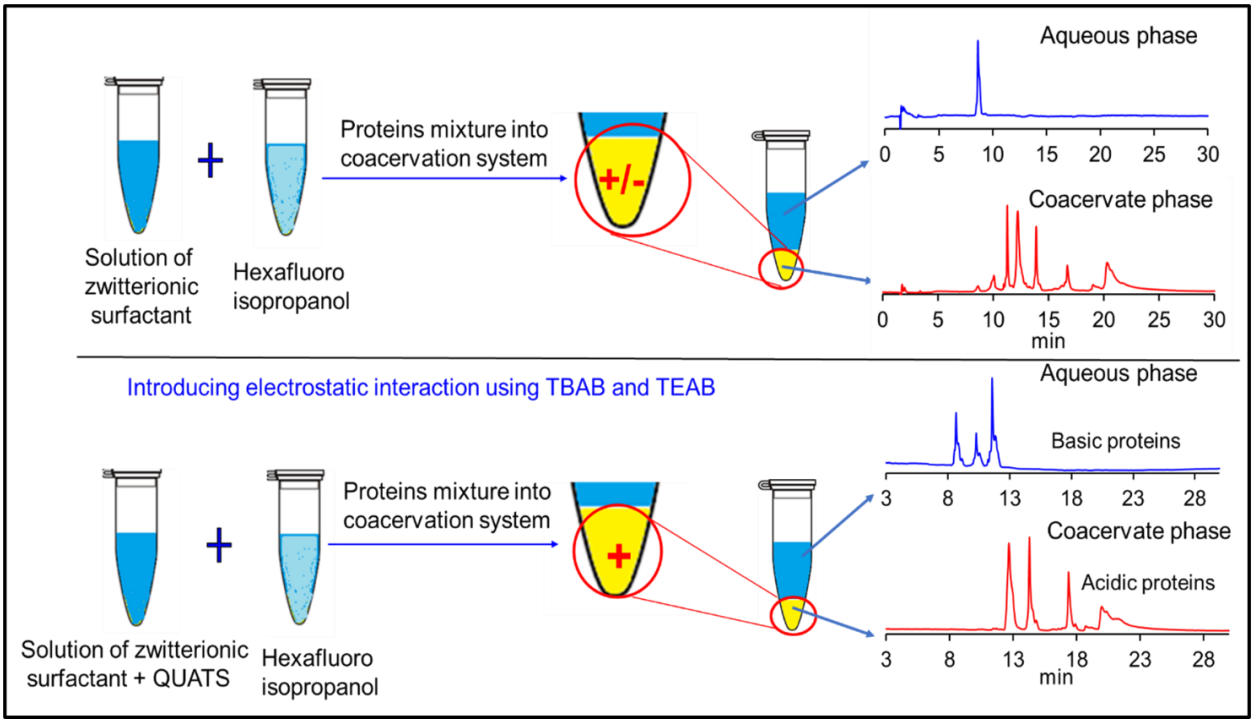
Department of Chemistry and Biochemistry, Arlington, TX, 76019

ABSTRACT: We have previously introduced the fluoroalcohol-induced supramolecular biphasic systems of zwitterionic surfactants and mixed amphiphiles in proteomics for increased protein coverage of overall yeast proteins especially membrane and low abundance proteins with respect to the conventional approach of no phase separation [1]. In that paper, we have also reported increased fractionation of yeast proteins by the addition of quaternary ammonium salts to zwitterionic surfactants during coacervation, however the detail mechanism was unknown. In this paper we are reporting the detail mechanism behind the protein fractionation and introduction of electrostatic effect by the addition of positively charged small chain quaternary salts (QUATS) to strongly hydrophobic zwitterionic surfactants like DMMAPS and CHAPS. Long chain zwitterionic surfactants tend to interact with proteins based on hydrophobic interaction however, addition of TEAB and TBAB to CHAPS and DMMAPS introduce the strong electrostatic effect in addition to hydrophobic effect. For the proof of concept, preliminary work was done with standard proteins mixture with their different physicochemical properties like isoelectric point, hydrophobicity, and molecular weight. Interestingly, selective fractionation of proteins was observed with the addition of QUATS (tetraethylammonium bromide (TEAB) and tetra butyl

ammonium bromide (TBAB)) to DMMAPS and CHAPS based on their isoelectric point and hydrophobicity in which more basic and hydrophilic proteins were extracted into the aqueous phase and more acidic and hydrophobic proteins were extracted into the coacervate phase of mixed amphiphilic systems. However, biphasic systems of DMMAPS and CHAPS without QUATS had extracted almost all proteins into coacervate phase. This is due to that in DMMAPS+HFIP and CHAPS+HFIP systems the fractionation of proteins is primarily due to the hydrophobic interaction between proteins and two phases. Finally, same biphasic systems were used for the extraction of yeast proteins in replicates and found that they follow the similar pattern as observed for standard proteins fractionation. This is due to the introduction of the electrostatic effect using QUATs to the underlying hydrophobic effect of zwitterionic surfactants in biphasic systems with proteins.

Keywords: Electrostatic interaction, Hydrophobic interaction, Liquid-liquid extraction, Fractionation, Membrane, and low abundance proteins

GRAPHICAL ABSTRACT



4.1. INTRODUCTION

Surfactants are very popular as solubilizing agents for membrane proteins [2-4]. They can interrupt protein-lipid interactions and can help in solubilizing hydrophobic proteins by mimicking the lipid membrane environment. In addition, chaotropic reagents such as urea are also useful for the unfolding of proteins. They disrupt protein interactions with the lipids in the membrane and stabilize the unfolded structure [5, 6]. Fluoroalcohol Induced Coacervate (FA*i*C) biphasic systems are a new alternative to using the conventional method of using aqueous solutions of surfactants or urea as solubilizing reagents for protein samples. Addition of a fluoroalcohol such as HFIP to aqueous solutions of amphiphiles facilitates coacervation of the amphiphile that forms a separate phase from the bulk aqueous solution [7-9]. It occurs based on different physical and chemical parameters like molecular compositions, the concentration of amphiphile, temperature, pH, ionic strength, and the presence of additives [10]. Effective solubilization, enrichment, and fractionation benefits offered by the coacervation approach improves the detection of challenging membrane and low abundance proteins [11, 12]. Bottom-up proteomics is a more popular than the top-down approach for characterization of complex proteome samples since detection of peptides by mass spectrometry is easier than large proteins [13, 14]. However, membrane proteins are embedded within the cell membranes lipid bilayer and are less amenable to tryptic digestion and are less available to the mass spectrometric analysis [15-18]. The small volume of coacervate phase concentrated with surfactants/amphiphiles and HFIP help better solubilization of proteins embedded in lipid bilayer so that they are easily available for trypsin digestion [13-18]. Similarly, this approach is helpful for the enrichment of low abundance proteins due to the small volumes of the coacervate phase relative to the total solution volume [19]. Low abundance proteins are too

low in concentrations that fall below the instrument detection limit. Also, the low abundance proteins are often overshadowed by proteins at much higher abundance levels [19-21]. The use of FA*i*C-BPS in proteomics workflow allows simultaneous extraction, fractionation, and enrichment of proteins, thus improving identification of membrane proteins and low abundance proteins [2, 22-24].

Our lab has introduced the Fluoroalcohol induced coacervate biphasic systems (FA*iC*-BPS) in 2013 and demonstrated that the usefulness of these systems in green synthesis and for sample preparation in the chemical analysis [26, 27]. The advantage of using FA*iC*-BPS for sample extraction is that they nearly provide a similar scope of selectivity as SPE through adjustment of the phase composition with different types of amphiphiles, fluoroalcohol, and through controlling pH and ionic strength [28, 29] while sample loss due to irreversible protein adsorption on solid surfaces would be much larger in SPE than FA*iC*-BPS [30]. Similarly, the FA*iC*-BPS are more cost-effective and environmentally friendly than the liquid-liquid extraction systems with volatile organic solvents [8, 25]. Extraction of compounds into the coacervate phase also leads to their concomitant enrichment due to the small volumes of the coacervate phase relative to the total solution volume [30].

The main goal of this work is to better understand the fractionation patterns of proteins in FA*iC*-BPS with different amphiphile compositions. Hydrophobic interaction is the main driving force in protein partitioning into coacervates composed of a zwitterionic amphiphile such as DMMAPS or CHAPS [30]. Addition of a QUATS such as TBAB or TEAB introduces positively charged sites in the coacervate phase, and subsequently electrostatic effects that impact protein distribution between the two phases in FA*i*C-BPS [8, 19][31-33]. Other types of interactions like dispersion, hydrogen bonding, and dipolar play a secondary or minor role.

4.2. EXPERIMENTAL DESIGN

4.2.1. Materials and Reagents: Zwitterionic surfactants: 3- (N, N-Dimethyl myristyl ammonia) propane sulfonate (DMMAPS) and 3-[(3-Cholamidopropyl) dimethyl ammonio]-1-propane sulfonate (CHAPS) were purchased from Fisher Scientific, USA. Positively charged amphiphiles of quaternary ammonium bromide salts like tetraethylammonium bromide (TEAB), tetrabutylammonium bromide (TBAB) were purchased from ACROS Organics, USA. In addition, 1,1,1,3,3,3-Hexafluoro-2-propanol (HFIP) was obtained from Oakwood Chemical, USA Trifluoroacetic acid (99%), formic acid (99%) was purchased from Alfa Aesar, USA. Chemicals required for pre-digestion and digestion of proteins: dithiothreitol (DTT), iodoacetamide (IAA), and sequencing grade trypsin, were purchased from Promega Corporation, 2800 Woods Hollow Road, Madison, WI, USA. Ribonuclease A from the bovine pancreas and beta-lactoglobulin were purchased from Fisher Scientific, USA. The lyophilized powder of lactate dehydrogenase from rabbit muscle, cytochrome c (98%) from equine heart, bovine serum albumin (99%), lysozyme, and Gramicidin from bacillus aneurinilyticus were purchased from Sigma Aldrich, USA. LC-MS grade organic solvents like acetonitrile (ACN) and isopropanol (IPA) were purchased from Fisher Chemical, USA. Millipore-DI water from the lab was used for sample preparation and HPLC analysis.

4.2.2. Cell Lysis and Sample Preparation: Saccharomyces cerevisiae cells were used in our experiments as a complex sample mixture. The detailed procedure of cell growth and cell lysate is reported in the previously published article [1]. In addition, the procedure for coacervation of DMMAPS/ mixed amphiphilic systems is also reported in the previously published article [1]. Sample preparation of standard hydrophilic proteins was done by dissolving them into DI water

and hydrophobic proteins like Lactate dehydrogenase and Gramicidin were dissolved into HFIP. Samples were mixed and fractionated using the coacervation approach.

4.2.3. Reversed-Phase High-Performance Liquid Chromatography (RP-HPLC): The two phases of different coacervation systems with standard proteins were analyzed by reversed-phase chromatography. A method was developed in our lab for the separation of a mixture of seven standard proteins using the reversed-phase C18 column (Zorbax 300SB-C18) with a column dimension of 4.6×150 mm, 3.5µm particle size using Shimadzu HPLC instrument. Nexera X2-LC-30AD and SPD-M30A Diode-array detector). The solvent system used was Solvent A: Water with 0.1% TFA, Solvent B: Acetonitrile with 0.1% TFA, Gradient run with 10-90% of solvent B for 30 min. Chromatograms were collected at a wavelength of 280 and 220 nm. For the complex yeast proteome mixture, sample analysis was done using a bottom-up approach with (Dionex Ultimate 3000 RSLC-Nano liquid chromatography system, coupled with Orbitrap Fusion Lumos MS®, Thermo Electron) instrument. A reversed-phase C18 nano column of column dimension 75 cm x 75 µm, particle size 3 µm was used for the separation. The injection volume of the sample was 2 µL and the solvent system used was a 0-90 min gradient run with: 0-28% of solvent B, the flow rate of 350 nL/min. Mobile phase A was 2% (V/V) Acetonitrile (ACN) and 0.1% (V/V) formic acid (FA) in water, and mobile phase B was 80% (V/V) ACN, 10% (V/V) trifluoroethanol (TFE), and 0.1% FA in water.

4.2.4. Data Analysis of the Real Sample: The raw mass spectrometry data obtained from yeast sample analysis were analyzed using Max Quant (Ver. 1.6.2.3) software. The set-up parameters for the Max Quant were the same as our previously published results [1]. Two out of three proteins from three runs were taken for further data analysis. Further data analysis was done using the UniProt database, Yeast Mine database, and Gene Ontology database to obtain more information

on extracted proteins. The UniProt database was used to convert the UniProt IDs of proteins to SGD ID. Gene Ontology database was used to obtain the information related to protein's cellular locations and function. GRAVY (grand hydropathy), Isoelectric point (pI), and percentages of the amino acid composition of proteins were retrieved by using YeastMine (http://yeastmine. yeastgenome.org) database.

4.3. RESULTS AND DISCUSSION

4.3.1. Effect of QUATS on Proteins Fractionation Patterns in FA*i*C-BPS: A mixture of seven standard proteins with different physicochemical properties like grand hydropathy (GRAVY), isoelectric point (pI), and molecular weight values (Table 1) were selected to examine the distribution patterns in the FA*i*C-BP systems. Among the seven standard proteins in table 1, three of them are basic proteins (pI>7) and four are acidic (pI<7). Additionally, most of the basic proteins happen to have lower GRAVY values than the acidic proteins. **Figure 1** and **Figure 2** show the fractionation pattern of seven protein mixture into the aqueous phase and the coacervate phase of the six FA*iC*-BP systems of DMMAPS/ DMMAPS+QUATS and CHAPS/ CHAPS+QUATS, respectively. Reversed-phase chromatography was used to analyze the proteins extracted into separate phases of each system. It should be noted that due to enrichment effects, proteins concentrations in the coacervate phases were too high and had to be diluted 3-5 times before the HPLC analysis. **Figure 1** (A-F) shows the fractionation of proteins into aqueous and coacervate phases of DMMAPS+HFIP, DMMAPS+TBAB+HFIP, and DMMAPS+TEAB+HFIP systems, respectively. DMMAPS+HFIP system does not show much fractionation of proteins: only RNase A is extracted into aqueous phase and the other six are extracted into coacervate phase. However, the addition of TBAB and TEAB to DMMAPS (**Figure 1** (C, D, E, & F) shows the dramatic changes in the fractionation pattern of proteins between the two phases. The best

fractionation is given by DMMAPS+TEAB+HFIP system in which three hydrophilic proteins with basic pI value are into aqueous phase and four less hydrophilic proteins with acidic pI value are in the coacervate phase. Similar results were observed by the addition of TEAB and TBAB to another zwitterionic surfactant, CHAPS and the chromatograms are shown in **Figure 2** (A-F). In both systems, the changes in protein partitioning and fractionation patterns are due to the addition of electrostatic interactions to the underlying hydrophobic effect. The fractionation patterns in the presence of TEAB and TBAB systems also show different selectivity. Comparing the mixed amphiphilic systems, the DMMAPS+QUATS systems outperform the CHAPS+QUATS for more selective fractionation of these seven proteins. This could be due to structural differences between DMMAPS (with a long chain alkyl attached the zwitterionic head group) and CHAPS (with a planar, steroid-like structure) that would in turn impacts their interactions with the QUATS.

Ultimately, both results indicate that DMMAPS and CHAPS coacervate are strongly hydrophobic in nature and hydrophobic effect plays the predominant role in proteins partitioning. Whereas addition of small positively charged amphiphiles like TEAB and TBAB to the zwitterionic surfactants adds the electrostatic effects that alters fractionation patterns.

Table 1: List of seven standard proteins with different physiochemical properties

Standard proteins	Molecular weight (kDa)	Grand hydropathy (GRAVY)	Isoelectric point (pI)
Ribonuclease A	13.7	-0.213	8.64
Cytochrome C	11.6	-0.875	9.59
Lysozyme	14.3	-0.150	9.32
Bovine Serum Albumin	66.0	-0.429	5.60
β-Lactoglobulin	18.3	-0.010	4.83
Lactate dehydrogenase	144	0.04	6.3
Gramicidin	1.88	1.95	6.0

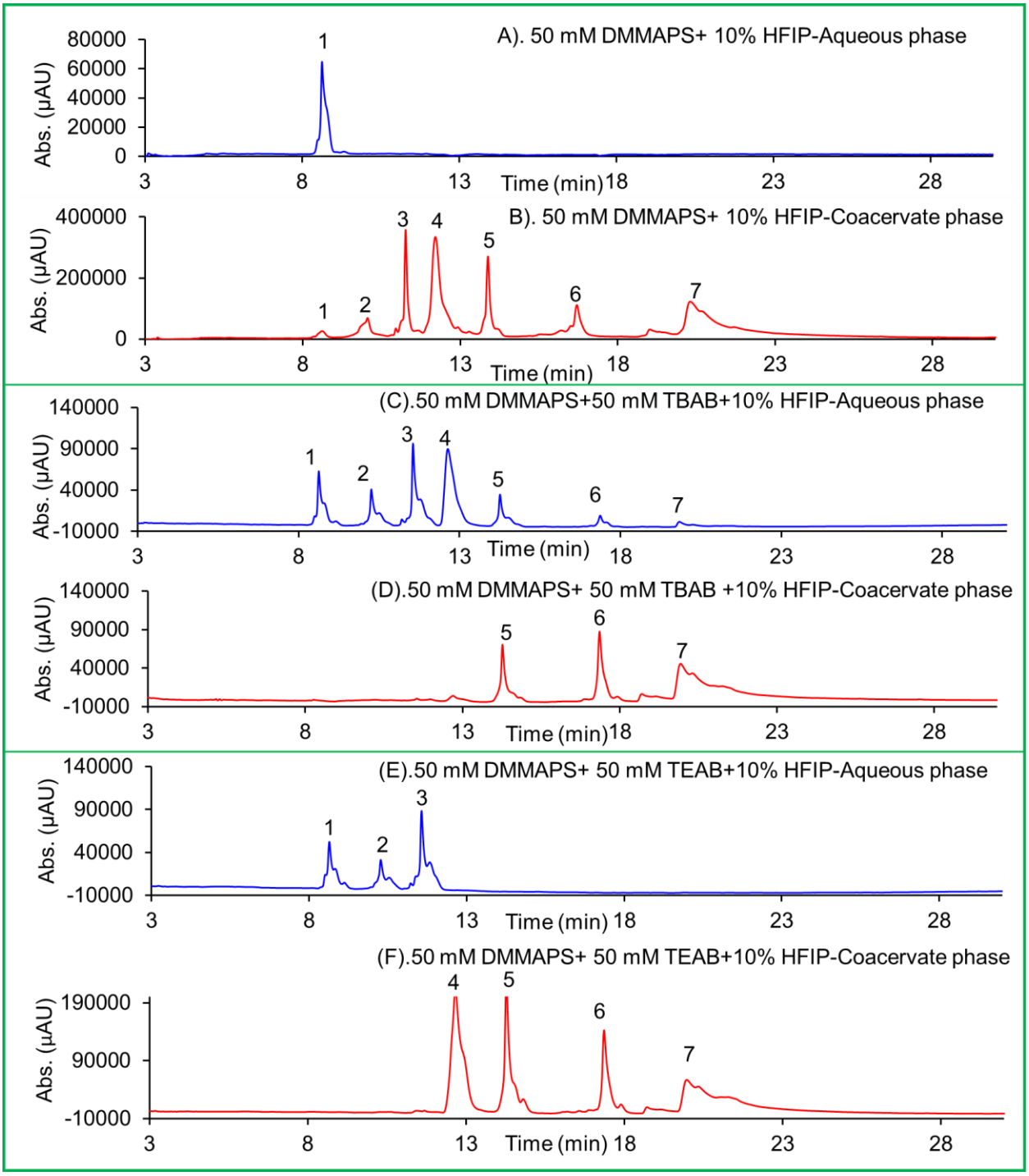


Figure 1: Chromatograms for the proteins extracted into aqueous phase and coacervate phase: (A). 50 mM DMMAPS+ 10% HFIP-Aqueous phase, (B). 50 mM DMMAPS+ 10% HFIP-Coacervate phase, (C). 50 mM DMMAPS+ 50 mM TBAB+ 10% HFIP- Aqueous phase, (D). 50 mM DMMAPS+ 50 mM TBAB+ 10% HFIP- Coacervate phase, (E). 50 mM DMMAPS+ 50 mM

TEAB+ 10% HFIP- Aqueous phase, (F). 50 mM DMMAPS+ 50 mM TEAB+ 10% HFIP- Coacervate phase, Detection at 220 nm. Analytes with elution order: 1- RNase A, 2- Cyto-C, 3- Lysozyme, 4- BSA, 5- Lactoglobulin-B, 6- Lactate dehydrogenase, 7- Gramicidin.

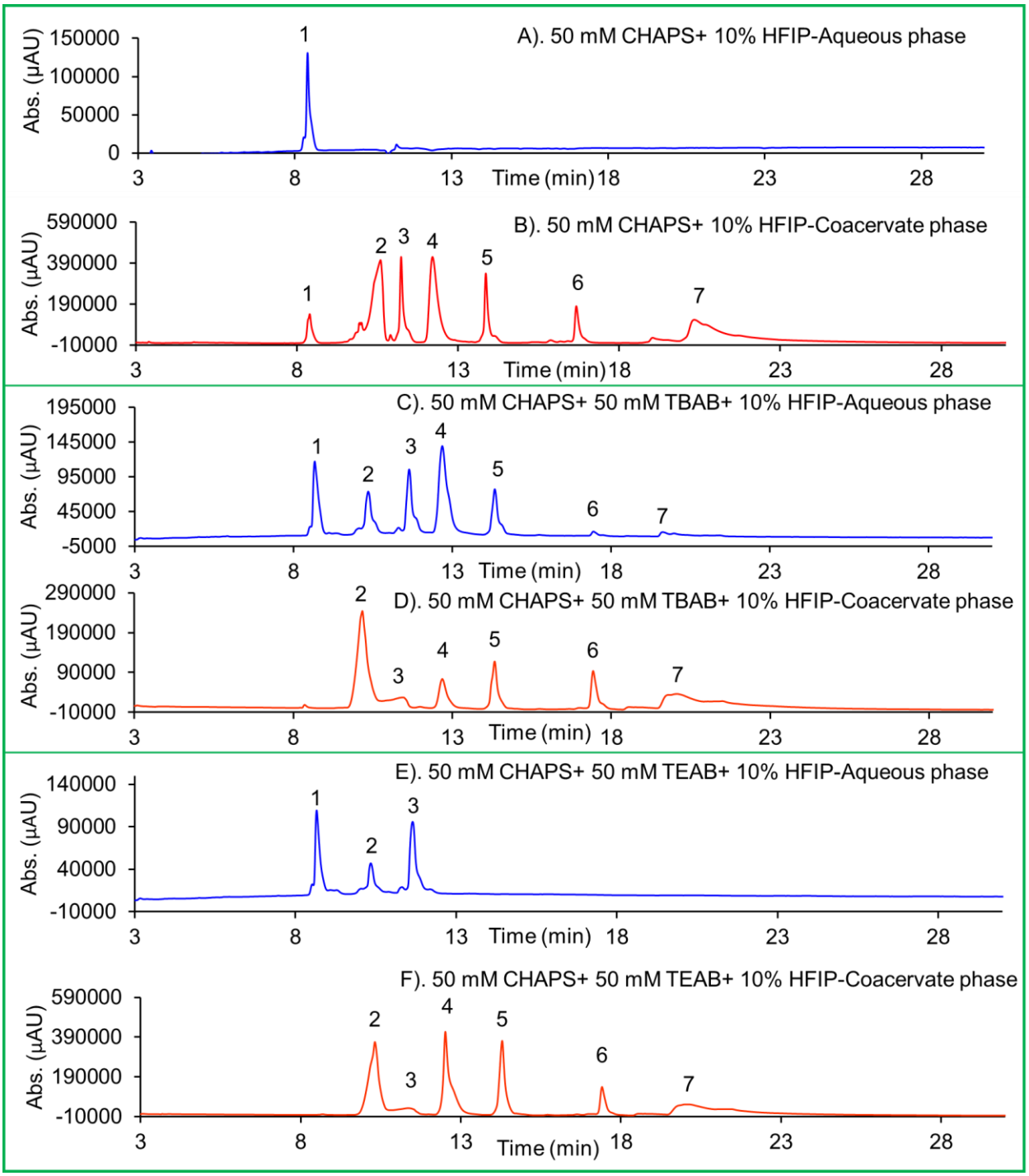


Figure 2: Chromatograms for the proteins extracted into aqueous phase and coacervate phase: (A). 50 mM CHAPS+ 10% HFIP-Aqueous phase, (B). 50 mM CHAPS+ 10% HFIP-Coacervate phase, (C). 50 mM CHAPS+ 50 mM TBAB+ 10% HFIP- Aqueous phase, (D). 50 mM CHAPS+ 50 mM TBAB+ 10% HFIP- Coacervate phase, (E). 50 Mm CHAPS+ 50 mM TEAB+ 10% HFIP- Aqueous phase, (F). 50 mM CHAPS+ 50 mM TEAB+ 10% HFIP-Coacervate phase, Detection at 220 nm. Analytes with elution order: 1- RNase A, 2- Cyto-C, 3- Lysozyme, 4- BSA, 5- Lactoglobulin-B, 6- Lactate dehydrogenase, 7- Gramicidin.

4.3.2. Fractionation of Yeast Proteomes by Mixed Amphiphilic Systems

Table 2 below shows the number of proteins extracted by six FAiC-BP systems in which two are HFIP induced biphasic systems of CHAPS and DMMAPS, and other four are HFIP induced biphasic systems of mixed amphiphiles of CHAPS and DMMAPS with QUATS. Table 2, from left to right list the number of proteins extracted into each system based on their partition coefficient value of $K<0.1$ (90% or higher concentration of proteins into Aq-phase), $10>K<0.1$(fractionation between two phases), and $K>10$ (90% or higher concentration of proteins into Co-phase) respectively. In addition, Venn diagrams in Figure 3 represents Table 2 which illustrates the fractionation patterns of proteins into the aqueous phase and the coacervate phase of six FAiC-BP systems based on their lowest and highest partition coefficient values. All these figures indicate that the addition of TEAB and TBAB to DMMAPS and CHAPS increased the fractionation of proteins into two phases, and the best fractionation is observed by the addition of TEAB to the CHAPS and DMMAPS. This result again supports the results obtained from the fractionation of standard protein mixture as shown in Figures 1 and 2.

Further data analysis was done for the proteins extracted into the aqueous phase (with $K<0.1$ or 0) and coacervate phase (with $K>10$ or ∞) to retrieve their isoelectric point, GRAVY, and molecular

weight (kDa) value and was plotted in a bar chart in Figure 4. As shown in **Figure 4**, distribution pattern of these yeast proteins extracted into the aqueous phase and the coacervate phase of DMMAPS and DMMAPS+QUATS systems are different. Interestingly, the comparison between DMMAPS+HFIP and DMMAPS+QUATS+HFIP coacervation systems shows a great shift in the fractionation pattern of proteins based on their isoelectric point value. In the DMMAPS+HFIP FA*iC* system, both the aqueous phase and the coacervate phase show a similar distribution pattern of proteins with a similar median value of pI ~6.8. This is because DMMAPS is a neutrally charged zwitterionic surfactant and doesn't show any electrostatic interaction towards proteins. However, addition of TEAB or TBAB to DMMAPS changed the protein distribution pattern by extracting proteins with a higher isoelectric point (basic proteins with median pI ~ 9) into the aqueous phase and proteins with a lower isoelectric point (slightly acidic proteins with median pI ~ 6.5) into the coacervate phase. Addition of TBAB provides greater electrostatic interaction than TEAB since grater pI difference is observed for proteins extracted into the aqueous phase and coacervate phase of DMMAPS+TBAB+HFIP system as compared to DMMAPS+TEAB system. As shown earlier with mixture of the seven standard proteins (Fig. 1 and 2), addition of the QUATS to DMMAPS/CHAPS coacervate resulted same fractionation patterns that correlated with their pI.

Additionally, the distribution pattern of proteins based on their GRAVY values differ slightly between three systems. As expected, proteins identified in the aqueous phases have lower GRAVY values than those extracted into the coacervate phases due to hydrophobic interaction. However, addition of TEAB and TBAB to the DMMAPS slightly increases the GRAVY difference between the proteins extracted into two phases. It indicates that TEAB and TBAB not only increases the proteins fractionation based on charge but also increases their fractionation based on their hydrophobicity as shown in Figure 3.

Fractionation of proteins into aqueous phase and coacervate phase based on molecular weight value doesn't follow any specific trend except proteins into the aqueous phase shows the smaller molecular weight value than the proteins into the coacervate phase.

Similar observations were true for CHAPS and mixed amphiphilic systems and are shown in the supplementary information in Figure S1. This explanation was further supported by amino acid composition analysis of proteins extracted into two phases of six different biphasic systems of DMMAPS, CHAPS, and their mixed amphiphiles. The data for compositional analysis of amino acids are shown in the supplementary information in Figure S2 and Figure S3. The results from figure S2 and S3 shows that the proteins into aqueous phase of DMMAPS+QUATS and CHAPS+QUATS systems has higher composition for basic amino acids than proteins into coacervate phase however, without QUATS there is not much difference in their acidic and basic composition.

Table 2: Yeast proteins extracted into two-phase of different FAiC-BP systems based on their partition coefficient value

Systems	Proteins with 90% or more into Aq-phase with $K < 0.1$	Shared proteins with partition coefficient ($0.1 < K < 10$)	Proteins with 90% or more into Co-phase with $K > 10$
DMMAPS+HFIP	160	450	2445
DMMAPS+TEAB+HFIP	221	63	2719
DMMAPS+TBAB+HFIP	228	351	2522
CHAPS+HFIP	116	114	2687
CHAPS+TEAB+HFIP	175	56	2751
CHAPS+TBAB+HFIP	220	393	2433

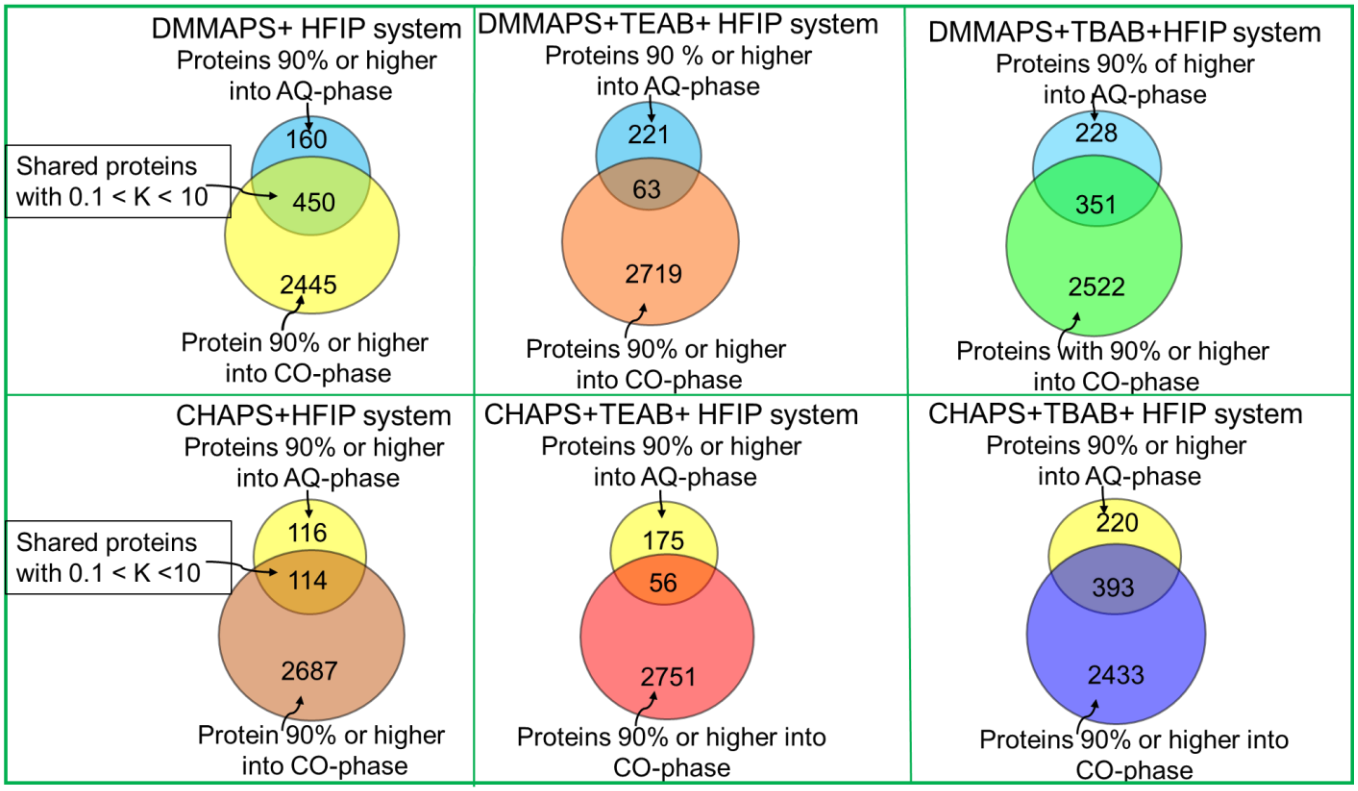


Figure 3: Fractionation of yeast proteins into aqueous and coacervate phase using HFIP induced supramolecular biphasic systems of DMMAPS, CHAPS and their mixed amphiphilic systems based on their partition coefficient value (related to Table 2)

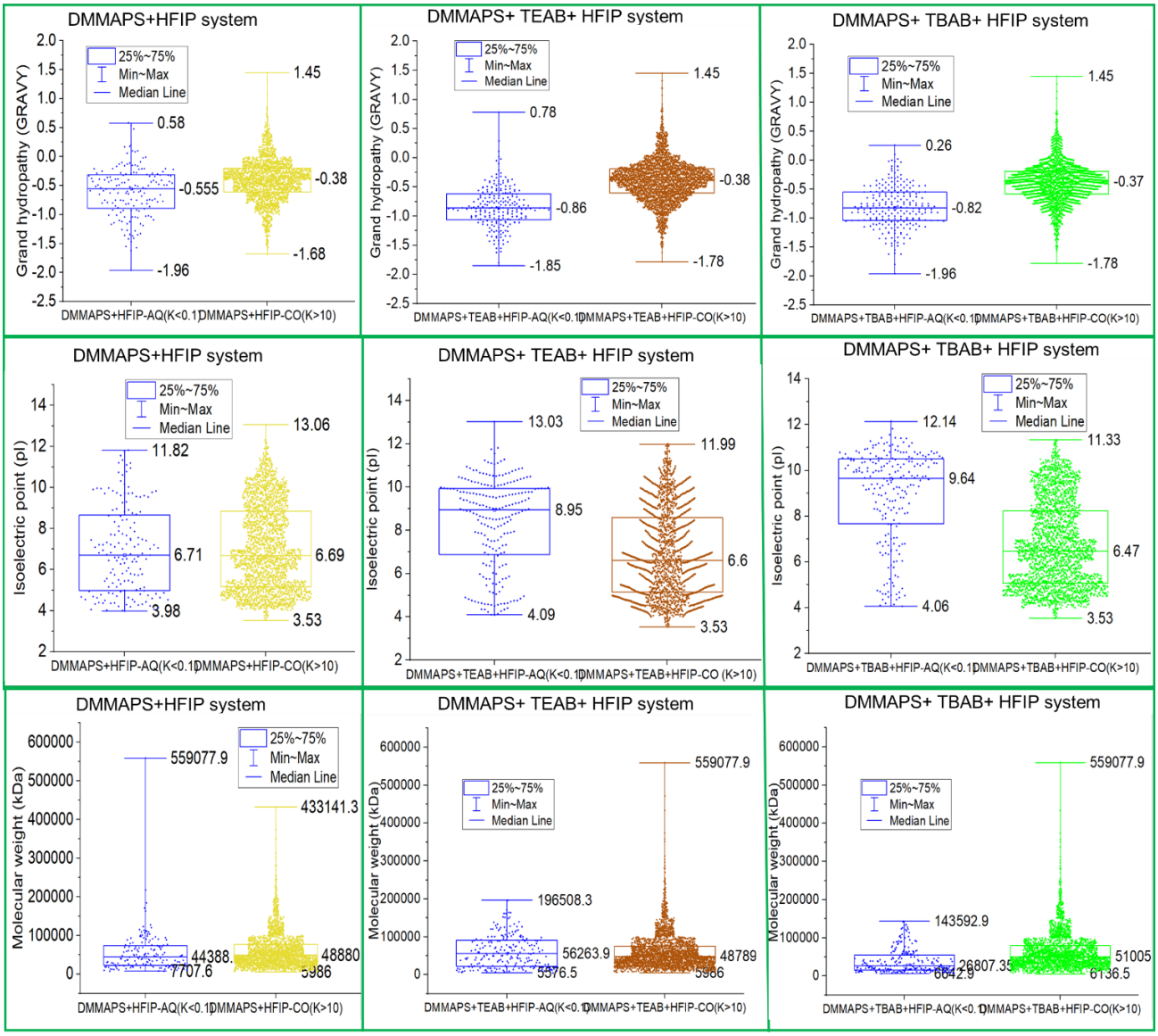


Figure 4: Change in fractionation of yeast proteins in DMMAPS+HFIP, DMMAPS+TEAB+HFIP, and DMMAPS+TBAB+HFIP systems based on their partition coefficient value (proteins with K<0.1 into aqueous phase and proteins with K>10 into coacervate phase, respectively).

4.3.3. Selectivity differences between FA*i*C-BP Systems.

Venn diagrams in **Figure 5** compare the selectivity difference of whole proteins identified in the aqueous and coacervate phases of the three FAiC-BP systems based on their isoelectric point value.

The number of proteins extracted in the aqueous phase (with K<0.1) of three systems: DMMAPS+HFIP, DMMAPS+TEAB+HFIP, and DMMAPS+TBAB+HFIP are 109, 124, and 128 respectively. Similarly, the number of proteins extracted into the three corresponding coacervate phases (with K>10) are 240, 140, and 122. The protein distribution patterns according to their isoelectric points are shown in Figure 5. Basic proteins (pI>7) have larger affinity toward the aqueous phase of the DMMAPS+TBAB+HFIP system, while larger number of acidic proteins are found in the coacervate phase of DMMAPS+TBAB+HFIP. This indicates that the positively charged basic proteins are repelled, while negatively charged acidic proteins are attracted to the positively charged coacervate phases of DMMAPS+TBAB+ HFIP due to electrostatic interaction. However, the effect is reverse with the DMMAPS+HFIP system due to the absence of electrostatic effect. Similar trends were observed for the CHAPS and mixed CHAPS + QUATS system. The Venn diagrams are shown in supplementary information in Figure S4. The selectivity difference is observed for other sub-proteomic type of the yeast proteins are also shown in Figure 6 and they also follow the similar trend.

Figure 6 basically shows fractionation pattern of some specific type of subcellular proteins of yeast cells with and without TEAB and TBAB in zwitterionic surfactant. As shown in figure 6**,** certain sub-cellular proteins such as the mitochondrial ribosome, ribosomal proteins, and chromosomal proteins has greater fractionation into two systems by the addition of TEAB and TBAB to DMMAPS. However, in the presence of TBAB proteins are extracted relatively more proteins into aqueous phase than other two systems. By analyzing the isoelectric point values for the proteins extracted into aqueous phase and coacervate phase only, proteins extracted into aqueous phase only are with higher pI value and proteins extracted into coacervate phase are with lower pI value. Similar results were observed for the CHAPS+TBAB aqueous phase system with

most basic proteins were extracted selectively by its aqueous phase. However, the most acidic proteins are selectively extracted into coacervate phase of CHAPS +TEAB system.

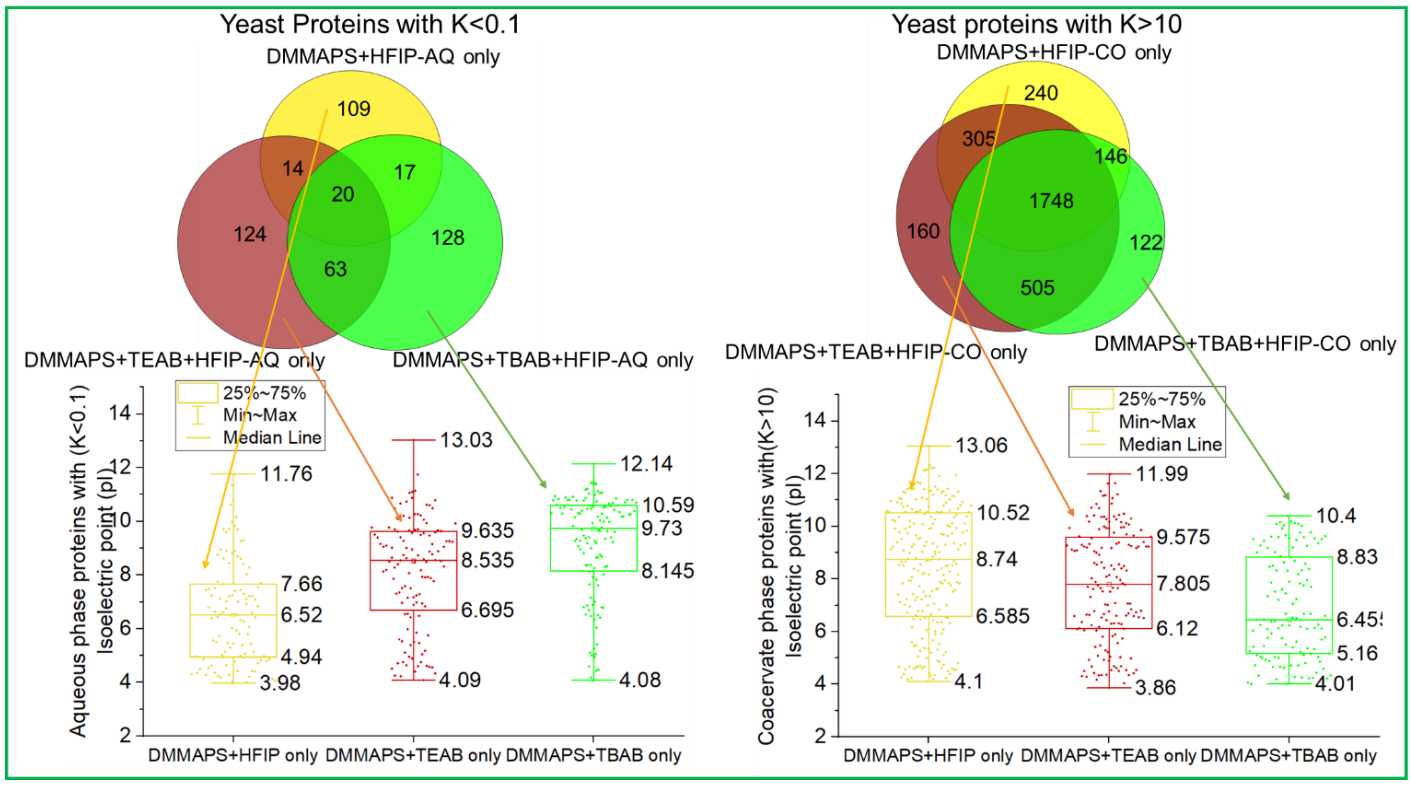


Figure 5: Selectivity difference between three FAiC-BPS of DMMAPS and DMMAPS+QUATS FAiC-BP systems for the extraction of proteins into the aqueous phase (with K<0.1) and coacervate phase (with K>10).

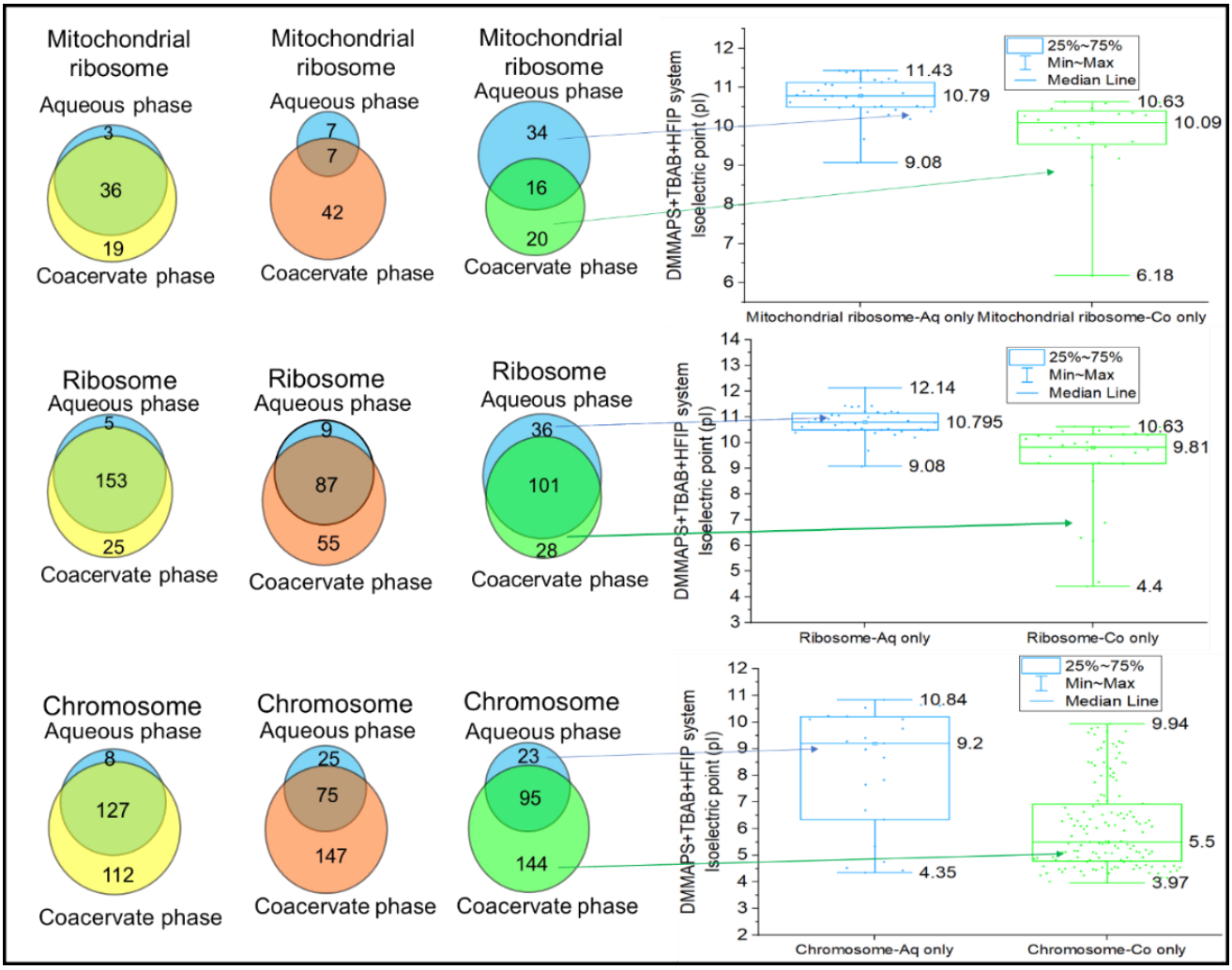


Figure 6: Selectivity difference of three FAiC systems: DMMAPS +HFIP, DMMAPS+TEAB+HFIP, and DMMAPS+TBAB+HFIP system for the extraction of certain sub-cellular proteins.

4.3.4. Similarity and differences between CHAPS and DMMAPS FA*i*C-BP Systems

The overall protein distribution of proteins extracted into six FA*i*C-BP systems and control (NP) based on their GRAVY, pI, and molecular weight value is shown in supplementary information as Figure S5. It shows the similarities and differences between for system for extraction of the yeast proteome. Protein distribution based on protein's isoelectric point and molecular weight value do not show any notable difference between protein extraction of all six systems. However, based on

the GRAVY value of proteins, DMMAPS and mixed amphiphilic system of DMMAPS shows the extraction of more hydrophobic proteins with the extraction of protein with GRAVY=1.45 than biphasic system CHAPS and its mixed amphiphilic systems and control with 8M Urea (NP). This indicates that presence of DMMAPS help better solubilization of hydrophobic proteins than CHAPS systems. This fact is further conformed by data analysis based on the extraction of trans-membrane proteins with alpha helices. Alpha helices are the hydrophobic part of the proteins usually found bounded between lipid bilayers. Data from Figure 7 shows that DMMAPS and mixed amphiphilic systems are extracting up to 555% higher number of transmembrane proteins with alpha helices with respect to control system (NP) [1], however CHAPS and mixed amphiphilic systems of CHAPS shows the extraction of similar or lower number of these proteins with respect to control system (NP).

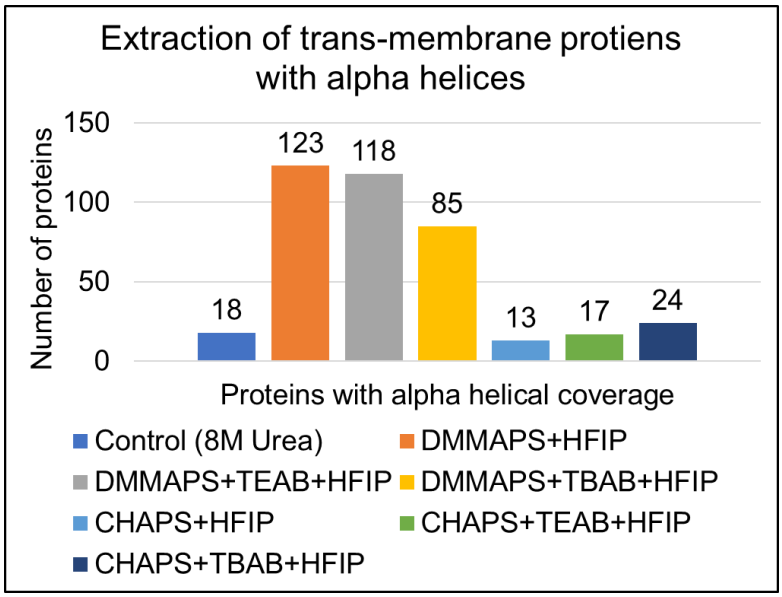


Figure 7: Extraction of trans-membrane proteins with alpha helices by different systems

4.3.5. Selectivity difference for Extraction of Lower Abundance Proteins by Six FA*i*C-BP Systems

The selectivity difference observed for lower abundance proteins in different combinations of coacervation system is another interesting result to study. The beauty of these two phase separation systems is that each system shows the different selectivity for different lower abundance proteins and one example is shown in previous published paper [1]. Figure 8 shows the selectively difference for lower abundance proteins (abundance < 2000 molecules/cells) extracted by different FA*iC*-BP systems of DMMAPS and CHAPS. The first two Venn diagrams in figure shows that each system has significant number of uniquely identified lower abundance proteins. Besides, Venn diagrams in second column shows the comparison between CHAPS and its mixed amphiphilic systems (CHAPS+QUATS+HFIP) with DMMAPS and its mixed amphiphilic systems (DMMAPS+QUATS+HFIP. The total number of lower abundance proteins only into DMMAPS and its mixed amphiphilic systems and CHAPS and its mixed amphiphilic systems are 133 and 76, respectively. Interestingly, these results exemplify that simply modification on one system can have higher identification improvement for lower abundance proteins. This unique properties of biphasic systems does the identification improvement of lower abundance proteins up to 408.3% with respect to 8M urea as control with no phase separation) [1]. In addition, in figure 9, distribution chart of uniquely identified lower abundance proteins of DMMAPS and CHAPS systems based on their GRAVY and pI value shows that proteins uniquely extracted by DMMAPS/DMMAPS + QUATS systems are more hydrophobic and basic in nature than proteins in CHAPS/CHAPS+QUATS systems. This result again indicates that biphasic systems of DMMAPS are better solubilizing agents for lower abundance membrane proteins than CHAPS systems.

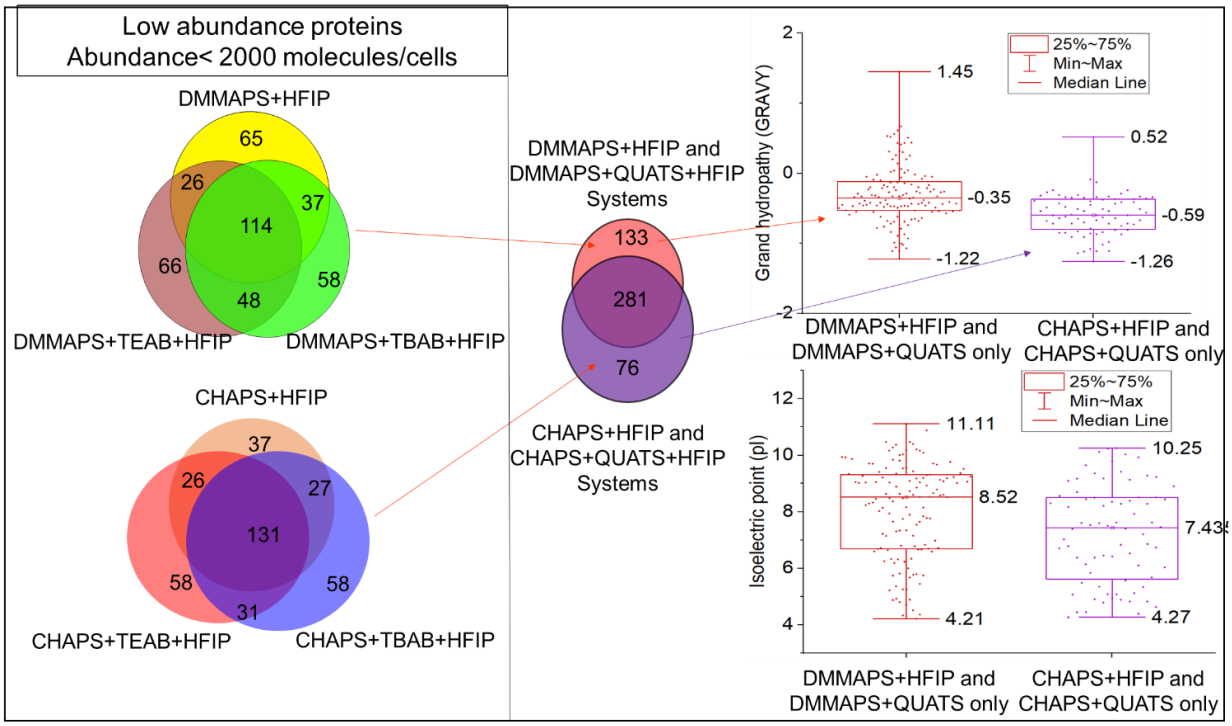


Figure 8: Selectivity difference between CHAPS and DMMAPS systems for the extraction of lower abundance proteins and their distribution based on grand hydropathy (GRAVY) and isoelectric point (pI) value.

4.5. CONCLUSION: The FA*i*C-BP systems composed of mixed zwitterionic surfactant (DMMAPS) with quaternary ammonium salts (QUATS) provide selectivity in fractionation of proteins mixture. This is due to the introduction of electrostatic effect using positively charged small chain quaternary ammonium salts like TEAB and TBAB to long chain zwitterionic surfactants like DMMAPS and CHAPS. During coacervation of zwitterionic surfactant and mixed amphiphiles with HFIP, almost 80-90 % of TEAB and TBAB goes to the coacervate phase along with CHAPS and DMMAPS which makes the interactive part of coacervate phase positively charged and proteins fractionation is observed based on their isoelectric point. However, biphasic systems of long-chain zwitterionic surfactants tend to exert a strong hydrophobic interaction,

therefore fractionation occurs based on hydrophobic interaction between two phases and proteins as observed for biphasic systems of CHAPS+HFIP and DMMAPS+HFIP. Overall, we can alter the fractionation of standard proteins mixture as well as complex sample mixture in biphasic system strong hydrophobic surfactants/amphiphiles by the addition of charged small molecules. We primarily tested the standard protein mixture for their fractionation and finally worked with complex sample mixture of yeast proteins. Overall, the results from both experiments follow the similar trend, and this indicates that fractionation of proteins can be altered by changing the composition of systems using different small, charged molecules.

ACKNOWLEDGMENTS

This research was supported by NSF grant CHE1632221.

4.6. REFERENCES

1. Khanal, D.D., et al., *Improved Protein Coverage in Bottom-Up Proteomes Analysis Using Fluoroalcohol-Mediated Supramolecular Biphasic Systems With Mixed Amphiphiles for Sample Extraction, Fractionation, and Enrichment.* Analytical Chemistry, 2021. **93**(20): p. 7430-7438.
2. Moore, S.M., S.M. Hess, and J.W. Jorgenson, *Extraction, Enrichment, Solubilization, and Digestion Techniques for Membrane Proteomics.* Journal of Proteome Research, 2016. **15**(4): p. 1243-1252.
3. Helbig, A.O., A.J.R. Heck, and M. Slijper, *Exploring the membrane proteome—Challenges and analytical strategies.* Journal of Proteomics, 2010. **73**(5): p. 868-878.
4. Speers, A.E. and C.C. Wu, *Proteomics of Integral Membrane ProteinsTheory and Application.* Chemical Reviews, 2007. **107**(8): p. 3687-3714.

5. Waas, M., et al., *Combine and Conquer: Surfactants, Solvents, and Chaotropes for Robust Mass Spectrometry Based Analyses of Membrane Proteins.* Analytical Chemistry, 2014. **86**(3): p. 1551-1559.

6. Möglich, A., F. Krieger, and T. Kiefhaber, *Molecular Basis for the Effect of Urea and Guanidinium Chloride on the Dynamics of Unfolded Polypeptide Chains.* Journal of Molecular Biology, 2005. **345**(1): p. 153-162.

7. Koolivand, A., et al., *Coacervation of Lipid Bilayer in Natural Cell Membranes for Extraction, Fractionation, and Enrichment of Proteins in Proteomics Studies.* Journal of Proteome Research, 2019. **18**(4): p. 1595-1606.

8. Koolivand, A., et al., *Fluoroalcohol – Induced coacervates for selective enrichment and extraction of hydrophobic proteins.* Journal of Chromatography B, 2018. **1083**: p. 180-188.

9. Nejati, M.M. and M.G. Khaledi, *Perfluoro-Alcohol-Induced Complex Coacervates of Polyelectrolyte–Surfactant Mixtures: Phase Behavior and Analysis.* Langmuir, 2015. **31**(20): p. 5580-5589.

10. Chollakup, R., et al., *Phase Behavior and Coacervation of Aqueous Poly(acrylic acid)−Poly(allylamine) Solutions.* Macromolecules, 2010. **43**(5): p. 2518-2528.

11. Mesquita, R.O., et al., *Method optimization for proteomic analysis of soybean leaf: improvements in identification of new and low-abundance proteins.* Genetics and Molecular Biology, 2012. **35**: p. 353-361.

12. Isaacson, T., et al., *Sample extraction techniques for enhanced proteomic analysis of plant tissues.* Nature Protocols, 2006. **1**(2): p. 769-774.

13. Chait, B.T., *Mass spectrometry: bottom-up or top-down?* Science, 2006. **314**(5796): p. 65-66.

14. Gregorich, Z.R., Y.-H. Chang, and Y. Ge, *Proteomics in heart failure: top-down or bottom-up?* Pflügers Archiv - European Journal of Physiology, 2014. **466**(6): p. 1199-1209.

15. Bordag, N. and S. Keller, *α-Helical transmembrane peptides: A "Divide and Conquer" approach to membrane proteins.* Chemistry and Physics of Lipids, 2010. **163**(1): p. 1-26.

16. Lin, S.-H. and G. Guidotti, *Chapter 35 Purification of Membrane Proteins*, in *Methods in Enzymology*, R.R. Burgess and M.P. Deutscher, Editors. 2009, Academic Press. p. 619-629.

17. Whitelegge, J.P., *Integral Membrane Proteins and Bilayer Proteomics.* Analytical Chemistry, 2013. **85**(5): p. 2558-2568.

18. Yu, Y.Q., M. Gilar, and J.C. Gebler, *A complete peptide mapping of membrane proteins: a novel surfactant aiding the enzymatic digestion of bacteriorhodopsin.* Rapid communications in mass spectrometry, 2004. **18**(6): p. 711-715.

19. McCord, J.P., D.C. Muddiman, and M.G. Khaledi, *Perfluorinated alcohol induced coacervates as extraction media for proteomic analysis.* Journal of Chromatography A, 2017. **1523**: p. 293-299.

20. Corthals, G.L., et al., *The dynamic range of protein expression: A challenge for proteomic research.* ELECTROPHORESIS, 2000. **21**(6): p. 1104-1115.

21. Anderson, N.L. and N.G. Anderson, *The human plasma proteome: history, character, and diagnostic prospects.* Molecular & cellular proteomics, 2002. **1**(11): p. 845-867.

22. Pedersen, S.K., et al., *Unseen Proteome: Mining Below the Tip of the Iceberg To Find Low Abundance and Membrane Proteins.* Journal of Proteome Research, 2003. **2**(3): p. 303-311.

23. Blonder, J., et al., *Enrichment of Integral Membrane Proteins for Proteomic Analysis Using Liquid Chromatography−Tandem Mass Spectrometry.* Journal of Proteome Research, 2002. **1**(4): p. 351-360.

24. Issaq, H.J., et al., *Methods for fractionation, separation and profiling of proteins and peptides.* ELECTROPHORESIS, 2002. **23**(17): p. 3048-3061.

25. Ly, L. and V.C. Wasinger, *Protein and peptide fractionation, enrichment and depletion: Tools for the complex proteome.* PROTEOMICS, 2011. **11**(4): p. 513-534.

26. Khaledi, M.G., S.I. Jenkins, and S. Liang, *Perfluorinated Alcohols and Acids Induce Coacervation in Aqueous Solutions of Amphiphiles.* Langmuir, 2013. **29**(8): p. 2458-2464.

27. Weisner, N. and M.G. Khaledi, *Organic synthesis in fluoroalcohol-water two-phase systems.* Green Chemistry, 2016. **18**(3): p. 681-685.

28. Martínez-Maqueda, D., et al., *Extraction/Fractionation Techniques for Proteins and Peptides and Protein Digestion*, in *Proteomics in Foods: Principles and Applications*, F. Toldrá and L.M.L. Nollet, Editors. 2013, Springer US: Boston, MA. p. 21-50.

29. Dephoure, N. and S.P. Gygi, *A solid phase extraction-based platform for rapid phosphoproteomic analysis.* Methods, 2011. **54**(4): p. 379-386.

30. Pires, M.J., M.R. Aires-Barros, and J.M.S. Cabral, *Liquid−Liquid Extraction of Proteins with Reversed Micelles.* Biotechnology Progress, 1996. **12**(3): p. 290-301.

31. Cladera, J., et al., *Liposome Solubilization and Membrane Protein Reconstitution Using Chaps and Chapso.* European Journal of Biochemistry, 1997. **243**(3): p. 798-804.

32. Arnold, T. and D. Linke, *Phase separation in the isolation and purification of membrane proteins.* Biotechniques, 2007. **43**(4): p. 427-440.

33. Hjelmeland, L.M., *A nondenaturing zwitterionic detergent for membrane biochemistry: design and synthesis.* Proceedings of the National Academy of Sciences, 1980. **77**(11): p. 6368-6370.

CHAPTER 5

Associated Organic Solvents (AOSB) Biphasic Systems for the Extraction and Enrichment of Organic Samples in Trace Analysis.

Used with permission from Durga Devi Khanal, Sajad Tasharofi, Morteza G. Khaledi*

The University of Texas at Arlington

Department of Chemistry and Biochemistry, Arlington, TX, 76019

ABSTRACT: Associated Organic Solvents Biphasic (AOSB) Systems in aqueous media mediated by Hexafluoro isopropanol (HFIP) are introduced as an effective, easy, environmentally friendly, economically viable approach for extraction and enrichment of enrichment of small organic molecules. Addition of HFIP to aqueous mixtures of polar organic solvents leads to formation of two separate phases; one that is highly enriched in HFIP and another polar organic solvent, and an aqueous-rich phase with small concentrations of the two organic solvents. The formation of a separate organic-rich phase is primarily driven by hydrogen bonding interaction between the hydroxyl group of HFIP and a hydrogen-bond acceptor group of the second organic solvent on one hand, and hydrophobic interaction between the fluorocarbon groups of HFIP and the alkyl group of the organic solvent. The Associated Organic Solvents (AOS) phase has a much smaller volume than the total solution volume of water, HFIP, and organic solvent. Thus, a unique feature of the AOSB biphasic systems is that analytes are enriched upon their extraction into the AOS phase due to its small volume fraction. The AOSB systems are considerably safer than the conventional liquid-liquid extraction systems using water immiscible volatile organic solvents. This is due to the small concentrations of HFIP and the second organic solvent that is needed to

form the organic-rich phase, for example, a 15% v/v mixture of 1:1 HFIP and polar organic solvent e.g. acetonitrile results in the formation of an AOSB phase with the total HFIP-organic concentration of ~90% v/v, and a volume fraction of ~ 75 µL. We report four different types of HOAS biphasic systems composed of water, HFIP, and a second organic solvent (acetonitrile, isopropanol, THF, or acetone) for extraction and enrichment of three classes of environmental pollutants polycyclic aromatic hydrocarbons (PAH), pesticides, and hormones at trace levels. The enrichment factors of 25-40 times were observed for the HOAS systems as compared to the conventional liquid-liquid extraction control system. Limit of detection (LOD) of 3.6 – 10.8 ng/mL and limit of quantitation (LOQ) of 10.9 – 32.9 ng/mL were observed for 13 GC-MS analysis of polycyclic aromatic hydrocarbons.

Keywords: Associated organic solvents, Hexafluoro isopropanol, Aqueous two-phase systems, Environmental analysis, Polycyclic aromatic hydrocarbons, Pesticides, Hormones

5.1. INTRODUCTION: Environmental pollution is today's major concern which is affecting human health and aquatic life directly and indirectly**.** Increased industrialization and urbanization are the main causes of environmental pollution. Many industries like food industries, pharmaceutical industries, textile industries, oil industries etc. produce numerous chemicals like polycyclic aromatic hydrocarbons (PAH), pesticides, hormones, nitro compounds, synthetic organic dyes, heavy metals, etc. as their biproducts and waste in a variety of environmental samples like water, aquatic biota, soil, air, etc. [1-3]. These chemicals do not degrade easily in the environment and cause the long-term side effects [2]. Among all, polycyclic aromatic hydrocarbons are a class of organic pollutants that are considered as "priority food and environmental pollutants". They are usually produced from fossil fuel combustion, oil combustion,

different industrial activities, domestic heating, biomass burning etc. and are carcinogenic in nature [4, 5]. Usually, non-smokers get PAH from food and air since PAHs are significantly present in food by different food processing like smoking, grilling, frying, baking etc. [6, 7]. PAHs have low aqueous solubility; therefore, they are more persistent in food and soil. PAHs are considered as a major concern to human health. In 1984 environmental protecting agency (EPA) added 16 PAH to the list of carcinogenic and mutagenic compounds [8]. Similarly, European union also provided the list of 15 PAH samples, eight of which overlap with the EPA list of PAH [9].

In addition to PAH, pesticides and sex steroid are other types of environmental pollutants that have acute and chronic side effects to the human health [10, 11]. Pesticides are among the non-biodegradable pollutants and persist in the environment for several years. Pesticides are used in agricultural lands, gardens, public areas in large amounts to kill different types of insects and pests and increase the agricultural production [12, 13]. Pesticides are either sprayed on plants or directly on the soil to kill fungus, germs, or insect; and as a result, they remain in the soil for a long time and are transported to different places through water. They also get adsorbed on crops like grains, fruits, or leaves and can impact human health through direct contact. They have adverse effects on aquatic life, wildlife, and human health [14]. In 1945, DDT was used widely as a pest control since it can kill wide variety of insects. Unfortunately, insecticides also kill beneficial insects and are very persistent in the environment with long-term side effect. After second world war, people started using other pesticides like aldrin, dieldrin, endrin, BHC etc. Their extensive use made insects resistant to them and showed the negative side effects to the plants and animals. Therefore, these days people are more concerned about their use and risk factor. However, still large number of pesticides are using every day to increase the crop production to fulfill the ongoing demand of

population growth. Environmental protection agencies around the world are limiting their use and have already banned some pesticides with extreme side effect.

Additionally, sex steroid hormones like androgens, estrogens, and progestins are a group of potent endocrine disruptors and they get exposed to environment naturally and synthetically [15, 16]. They can be found in wastewater and soil in trace levels form different sources like synthesized oral contraceptive medications, hormone replacement therapy, and from animal and human urine [17]. However, they have strong side effect on reproductive behavior and gene expression of aquatic life and land life [18-20]. Several studies have been done on the occurrence of estrogens in water resources [21-24]. Natural and synthetic estrogens like estrone, 17b-estradiol, 17a-ehtynylestradiol etc. have wide occurrence in water resources. In addition, concentration of androgens and progestogens have higher occurrence in these resources as compared to estrogens [17].

The presence of these environmental pollutants like PAH, pesticides, and hormones in trace levels in environment make them hard to detect by analytical instruments like GC-MS, HPLC-MS, UPLC-MS, [1, 16]. Effective sample preparation techniques are needed for their enrichment and purification before their analysis by GC-MS or LC-MS instruments. Different types of solid phase extraction, ionic-liquid extraction, liquid-liquid extraction are the common approaches generally used for their sample preparation [3, 25-27]. In this manuscript, we report the usefulness of AOS biphasic systems of for the extraction and enrichment of three classes of environmental pollutants like PAH, pesticides, and hormones. This approach is easy to use, economically viable and most importantly can have the enrichment of hydrophobic small molecules into one phase. Previously, our lab has investigated several fluoroalcohol or fluoroacid induced biphasic systems of surfactants, amphiphiles, polysaccharides and polar organic solvents for their application in the

field of proteomics [28-34]. Use of hexafluoro isopropanol (HFIP) induced biphasic system of polar organic solvents in water for the enrichment of environmental pollutants like polycyclic aromatic hydrocarbons (PAH), pesticides, and sex steroid hormones is a new and an important aspect of our work. We have tested several polar organic solvents like acetonitrile (ACN), acetone, n-propanol and tetrahydrofuran (THF) to understand their phase transition behavior with HFIP and water and their enrichment capability for hydrophobic small molecules.

5.2. EXPERIMENTAL SECTION

5.2.1. Chemicals and Reagents: Standard polycyclic aromatic hydrocarbons like naphthalene (100%), 2-methylnaphthalene (97%), acenaphthene (100%), fluorene (98%), phenanthrene (98%), anthracene (100%), fluoranthene (98%), pyrene (98%), benz(a)anthracene (100%), chrysene (98%), benz(a)pyrene (100%), benzo(b)fluoranthene (100%), perylene (100%) were all purchased from Sigma Aldrich, USA. Similarly, pesticides like pentanochlor pestnatal (99%), chlorpyrifos pestnatal (100%), flutriafol were purchased from Sigma Aldrich, USA, dichlorobiphenyl dichloroethylene (DDE) and dieldrin were purchased from Cerilliant, USA, and dichlorobiphenyl trichloroethane (DDT) was purchased from Restek, USA. Hormones like estrone (100%), β-estradiol (100%), progesterone (100%), and estriol (100%) were purchased from Sigma Aldrich, USA. 1,1,3,3,3-Hexafluoro-2-propanol (HFIP) was obtained from Oakwood Chemical, USA. Additionally, other chemicals like LC/MS grade acetonitrile (ACN), n-propanol, acetone, and tetrahydrofuran (THF) were brought from Thermo Fisher Scientific, USA. Millipore water for the sample preparation was used from the lab throughout the experiment.

5.2.2. Procedure for the Formation of HFIP Induced Polar Organic-Aqueous Biphasic System

Formation of the biphasic system from the miscible polar organic solvent with water in the presence of hexafluoro isopropanol is an easy and economically viable technique. Not all polar organic solvents form biphasic systems with HFIP and water. Their molecular structure and hydrophobicity determine the formation of supramolecular biphasic systems with HFIP. Larger the carbon-hydrogen chain length, higher the possibility of two-phase formation. Herein, we have analyzed biphasic systems of four different polar organic solvents like acetonitrile (ACN), acetone, n-propanol, and tetrahydrofuran (THF) with different concentration HFIP and water. HFIP acts as the phase inducer at room temperature. We had mixed the specific ratio of polar organic solvent, HFIP, and water and then simple step of vertexing for 2 min and centrifugation at the speed of 5 x g for 10 min was provided two distinct phases. The two phases were separated using micropipette and volume of each phase was recorded. Various combination of polar organic solvents (0-30%), HFIP (0-50%) and water (20-90%) were used to check their phase transition behavior. Figure 1 showed the schematic diagram of the formation of the biphasic system. All ACN, acetone, n-propanol, and THF showed a wide range of two-phase formation with HFIP in water and was shown in figure 2. As shown in figure, all four polar organic solvents started formation of two-phase with HFIP and water at 5% of organic solvent with 5% of HFIP and 80 % of water except n-propanol which was giving two-phase at 7.5% of n-propanol with 5% HFIP and 82.5% of water. However, all the organic solvents showed large range of two-phase formation. We have observed the two-phase formation for 30% organic solvent with 30% HFIP and 40% water.

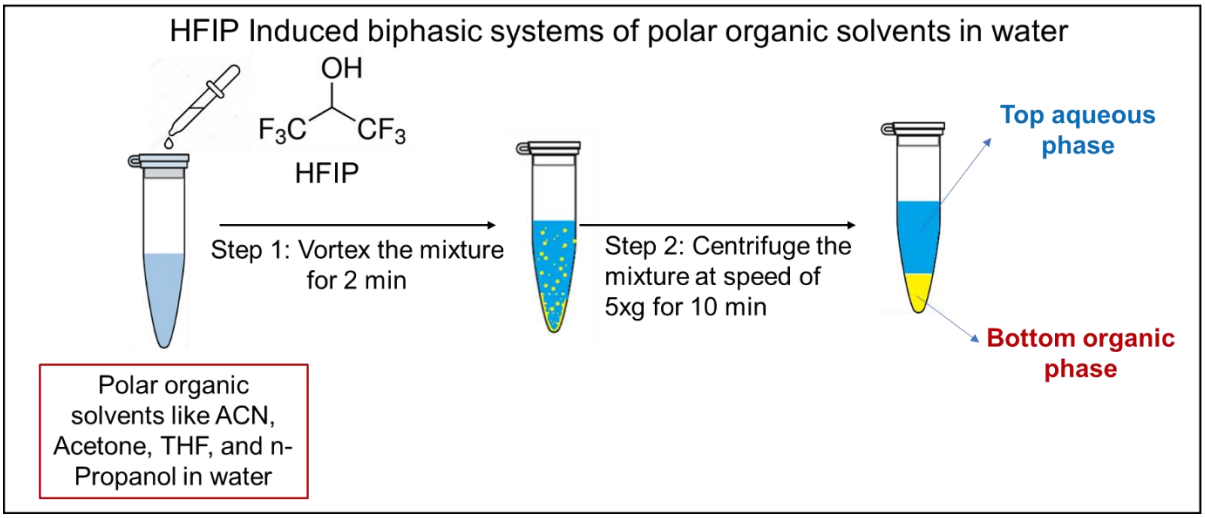


Figure 1: Schematic diagram showing the formation of biphasic system of polar organic solvents in water in the presence of HFIP.

5.2.3. Compositional Analysis of Biphasic System

In this experiment, the GCMS-2010SE (Shimadzu) instrument was used with a capillary column with a column dimension of 30 m x 0.25 mm x 0.25 µm, and stationary phase 5% Phenyl-Arylene 95% Dimethylpolysiloxane to check the compositional analysis of solvents in a biphasic system. The helium gas was used as the carrier gas throughout the experiment with 0.9 flow ratefor solvent analysis. The analysis was done in single ion mode (SIM) to get a better sensitivity for analysis. A method was developed for their characterization and separation and is shown in Table 1. Calibration curves were made and the concentration of solvents into each phase was calculated based on the calibration curve. Before analysis, the top aqueous phase and the bottom organic phase were diluted several times.

Table 1: Method profile for the analysis of solvents using GC-MS 2010E (Shimadzu)

Samples	Sample injection	Carrier gas flow (mL/min)	Oven profile	Ion source Temp (ºC)

Solvent mixture	Inj. temp.(ºC)	Inj. Vol. (µL	Split ratio	Carrier gas	Column flow (mL/min)	Rate (ºC/min)	Oven temp. (ºC)	Hold time (min)	250
	200	1	1:500	Helium	0.9	-	100	0.5	
						25	200	0	

5.2.4. Method Development for Characterization and Separation of Small Molecules by GC-MS Analysis

Shimadzu GCMS-2010SE instrument was used for the characterization of small molecules like polycyclic aromatic hydrocarbons, pesticides, and hormones. It has a capillary column inside with dimension of 30 m x 0.25 mm x 0.25 µm and 5% Phenyl-Arylene 95% Dimethylpolysiloxane coated stationary phase. A method was developed for the separation of mixture of PAHs, mixture of pesticides and mixture of sex steroid hormones separately. Method profile for the separation of these small molecules were in table 2 below and the chromatograms for the separation of 13 PAHs mixture, 6 pesticides mixture and 4 sex steroid hormone mixtures are shown in the figure 2, figure 3 and figure 4 respectively. The sample analysis was done in single ion monitoring (SIM) mode.

Table 2: Method profile for the separation of a mixture of polycyclic aromatic hydrocarbons, a mixture of pesticides, and a mixture of sex steroid hormones, respectively from top to bottom.

Samples	Sample injection			Carrier gas flow (mL/min)		Oven profile			Ion source Temp (ºC)
Mixture of polycyclic aromatic hydrocarbons	Inj. temp.(ºC)	Inj. Vol. (µL	Split ratio	Carrier gas	Column flow (mL/min)	Rate (ºC/min)	Oven temp. (ºC)	Hold time (min)	260
	280	3	1:1	Helium	1.4	-	100	0.5	
						30	260	0	
						25º	325	2	
Mixture of pesticides	250	3	1:10	Helium	1.0	-	200	0	260
						3	230	0	
						15	300	1	
	250	3	1:10	Helium	1.0	-	200	1	
						25	230	0	260

Mixture of sex steroid hormones						1 25	248 300	0 2	

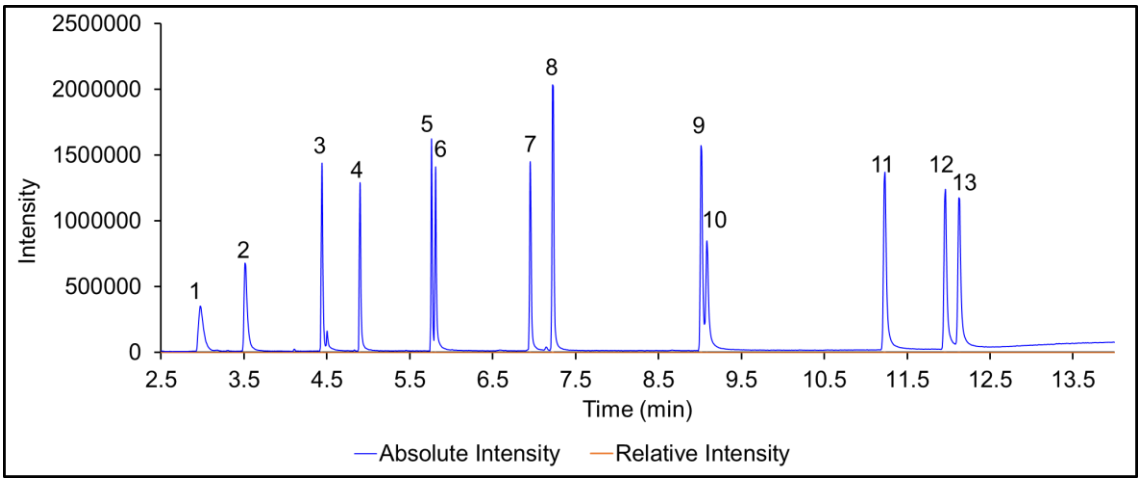


Figure 2: GC-MS separation of 13 polycyclic aromatic hydrocarbons. Analytes with elution order: 1. Naphthalene, 2. 2-Methylnaphthalene, 3. Acenaphthene, 4. Fluorene, 5. Phenanthrene, 6. Anthracene, 7. Fluoranthene, 8. Pyrene, 9. Benz(a)anthracene, 10. Chrysene, 11. Benz(b)fluoranthene, 12. Benz(a)pyrene, 13. Perylene

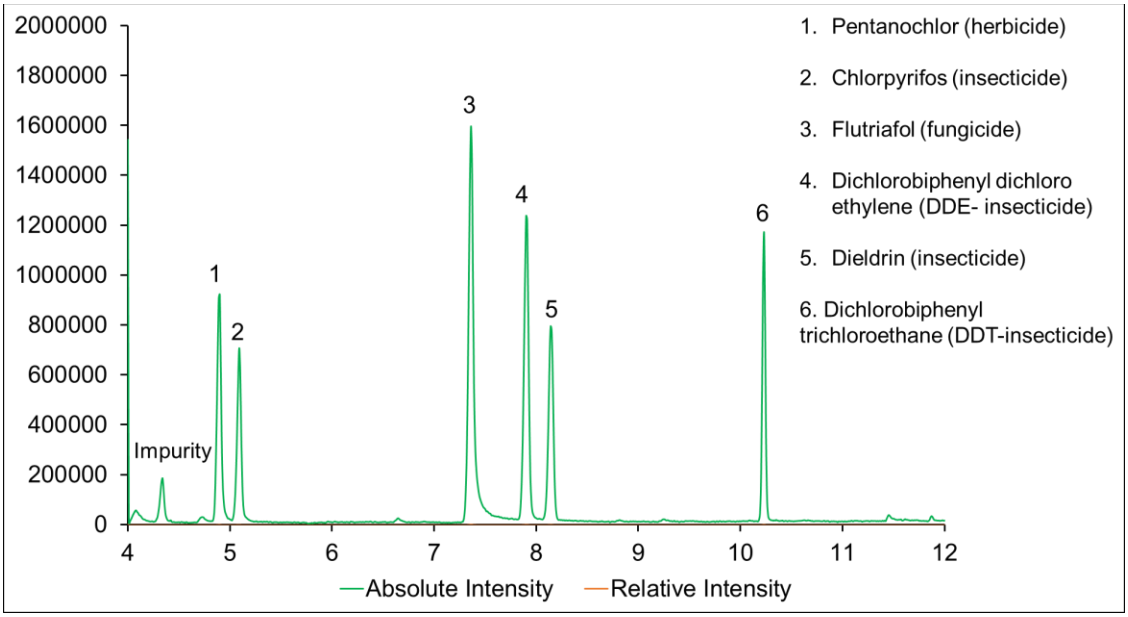


Figure 3: Chromatograms shows the GC-MS separation of six pesticides Analytes with elution order: 1. Pentanochlor, 2. Chlorpyrifos, 3. Flutriafol, 4. Dichlorobiphenyl dichloroethylene (DDE), 5. Dieldrin, 6. Dichlorobiphenyl trichloroethane (DDT)

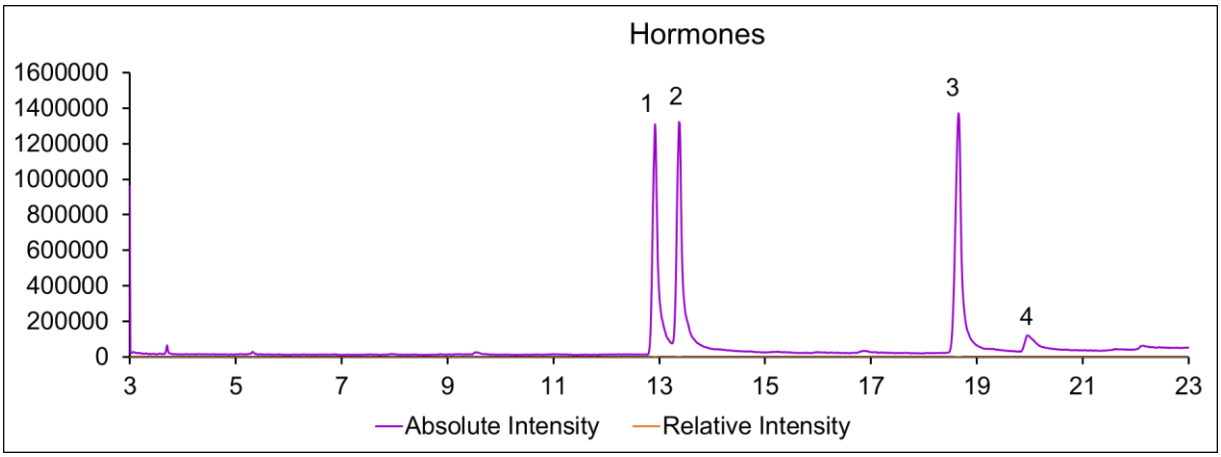


Figure 4: Chromatogram showing the separation of four sex steroid hormones. Analytes with elution order: 1. Estrone, 2. Estradiol, 3. Progesterone, 4. Estriol

5.2.5. Extraction Procedure of Small Molecules into Biphasic Systems of Different Organic Solvents in Water

5.2.5.(A). Extraction Procedure of Polycyclic Aromatic Hydrocarbons (PAH): In this experiment, we have primarily investigated the biphasic systems of four polar organic solvents in water in the presence of HFIP (a phase inducer) and checked their extraction efficiency for PAH samples. These four systems are a. 7.5% ACN+ 7.5% HFIP+ 85% water, b. 7.5% acetone+ 7.5% HFIP+ 85% water, c. 7.5% THF+ 7.5% HFIP+ 85% water, and d. 7.5% n-propanol+ 7.5% HFIP+ 85% water. Although their composition is same, the volume of the bottom organic phase formed is different. In the beginning, 7.5% organic solvent was mixed with 85% of water with samples and 7.5% HFIP was added later. The total volume of mixture was kept 1 mL. Then, the mixture was vortexed for 2 min and centrifuged at the speed of 5 x g for 10 min. Two distinct and clear phases were observed and were separated with the help of a micropipette. Both phases were

analyzed using GC-MS instrument to understand the partition coefficient of sample into two-phases and extraction efficiency of the bottom organic phase for PAH sample.

5.2.5.(B). Extraction Procedure of Pesticides and Sex Steroid Hormones: The same combination of organic aqueous biphasic systems as mentioned in section 2.5.1 was used for the extraction of pesticides and sex steroid hormones from a water sample. The combination of 7.5% organic solvent and 7.5% HFIP was spiked into 85% of water with pesticide sample mixture and then vortexed for 2 min and centrifuged for 10 min at the speed of 5 x g. The two phases were separated and analyzed using GC-MS instrument. The same steps were repeated for the extraction of the hormone sample mixture.

5.3. Extraction Procedure of Polycyclic Aromatic Hydrocarbons from Real Sample (Soil Sample): The soil sample with PAH-CRM 104 was bought from Sigma Aldrich. The PAHs were extracted using 80:20 of water and organic solvent (e.g., ACN, acetone, n-propanol, THF etc.). The 1 gram of sediment sample was weighted and taken into 15 mL centrifuge tube and added 2 mL of 80:20 mixture of water: organic solvent. Four solvent mixtures (80:20 water: ACN, 80:20 water: acetone, 80:20 water: n-propanol, 80:20 water: THF) were used for the extraction of PAHs. The mixture of solvent and soil sample were sonicated for 30 min in water bath. Then, the mixture was kept overnight at 4 degrees Celsius in a freeze. Finally, the sample was centrifuged at 3.3 x g for 30 min and the supernatant was collected with the help of pipette. The supernatant was centrifuged again with the speed of 10 x g for 20 min if there are any soil particles remaining. Then, 378 µL of solvent was taken and 75 of µL HFIP and 547 µL of water was added to make final combination of 7.5% organic solvent+7.5% HFIP and 85% of water in 1 mL. The mixture was centrifuged at 5 x g for 10 min and two phases were separated. The top aqueous phase was discarded, and bottom organic phase was analyzed using GC-MS instrument.

5.3. RESULTS AND DISCUSSION

5.3.1. Associated Organic Solvents Biphasic (AOSB) Systems in Aqueous Media and their Compositional Analysis using GC-MS Instrument:

We investigated the four AOSB systems of ACN, acetone, n-propanol, and THF and they showed the wide range of two-phase formation with HFIP in water. The phase transition behavior of these organic solvents in aqueous media in the presence of HFIP is shown in figure 5. As shown in figure, all four polar organic solvents started formation of two-phases at 5% of organic solvent with 5% of HFIP and 90 % of water except n-propanol which gives two-phase at 7.5% of n-propanol with 5% HFIP and 82.5% of water. However, all the organic solvents showed large range of two-phase formation. We have observed the two-phase formation for 30% organic solvent with 30% HFIP and 40% water. The composition of two phases were then analyzed using the GC-MS-2010 SE (Shimadzu) instrument as mentioned above. Calibration curves were developed for acetonitrile (ACN), n-propanol, HFIP, and water and are shown in the supplementary information in section S1. The concentration of total organic solvents (e.g., ACN and HFIP) and water present in each phase was analyzed using GC-MS analysis based on the calibration curve. Figure 6 shows the compositional analysis of the organic solvents and HFIP in the top aqueous phase and the bottom organic phase of Acetonitrile+ HFIP+ water and n-propanol+ HFIP+ water system at different solvent compositions. Three replicates were done for each analysis. In figure 6, as the concentration of HFIP increases keeping other organic solvents (ACN, n-propanol) constant, the composition of HFIP into both phases of each systems increases, while the concentration of other polar organic solvent (ACN, n-propanol) decreases. On the other hand, an opposite trend is opposite when the concentration of HFIP is kept constant and the concentration of the second organic solvent (ACN, n-propanol) is increased. Table 2 shows the change in concentration of

organic solvent (ACN, n-propanol) and HFIP by changing the composition of biphasic system. It shows that as the ACN kept constant while changing the HFIP concentration, the overall organic solvent concentration (ACN+HFIP) into bottom phase and aqueous phase is not changing. However, with constant HFIP and increased concentration of ACN, the total organic solvent (ACN+HFIP) into bottom phase is slightly decreasing while in aqueous phase is slightly increasing. The similar observation is true for the n-propanol and HFIP combinations except at a constant n-propanol and high concentration of HFIP (i.e., 7.5% n-propanol+ 30% HFIP). Different biphasic systems of ACN+ HFIP+ water have 8-15% total organic solvent into the top phase and 64-87% of total organic solvent into the bottom phase. Similarly, different biphasic systems of n-propanol+ HFIP+ water have 10-15% of total organic solvent into the top aqueous phase and 64-81% of total organic solvents into the bottom phase.

Additionally, figure 7 shows the water contents in the two phases of different biphasic systems. As shown in the figure the total water content in the top aqueous phase of different biphasic system ranges from 85-92% of its total volume and the bottom phase contain 20-38% of its total volume. Interestingly, increasing the total HFIP concentration, results in an increase in water concentration in the bottom organic phase. The HFIP and water molecules are associated through hydrogen bonding. Thus, HFIP molecules are transferred to the bottom phase in the hydrated form, which results in an increase in water concentration and volume of the bottom organic phase. Note that at a constant HFIP%, water concentration in the HFIP-Organic phase remains constant with an increase in the organic solvent concentration.

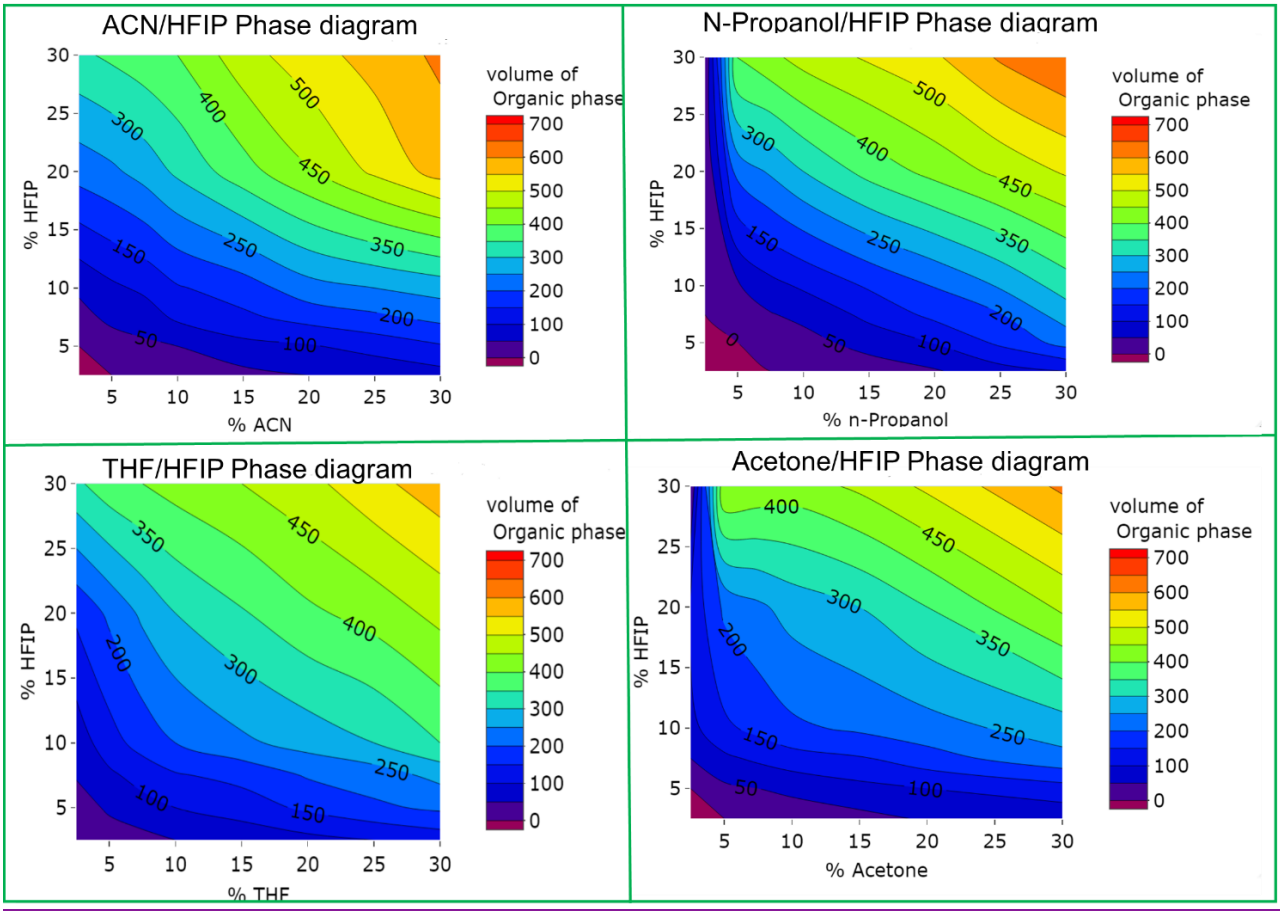


Figure 5: Phase transition behavior of four polar organic solvents ACN, n-propanol, THF, and acetone in aqueous media respectively from left to right induced by the presence of HFIP.

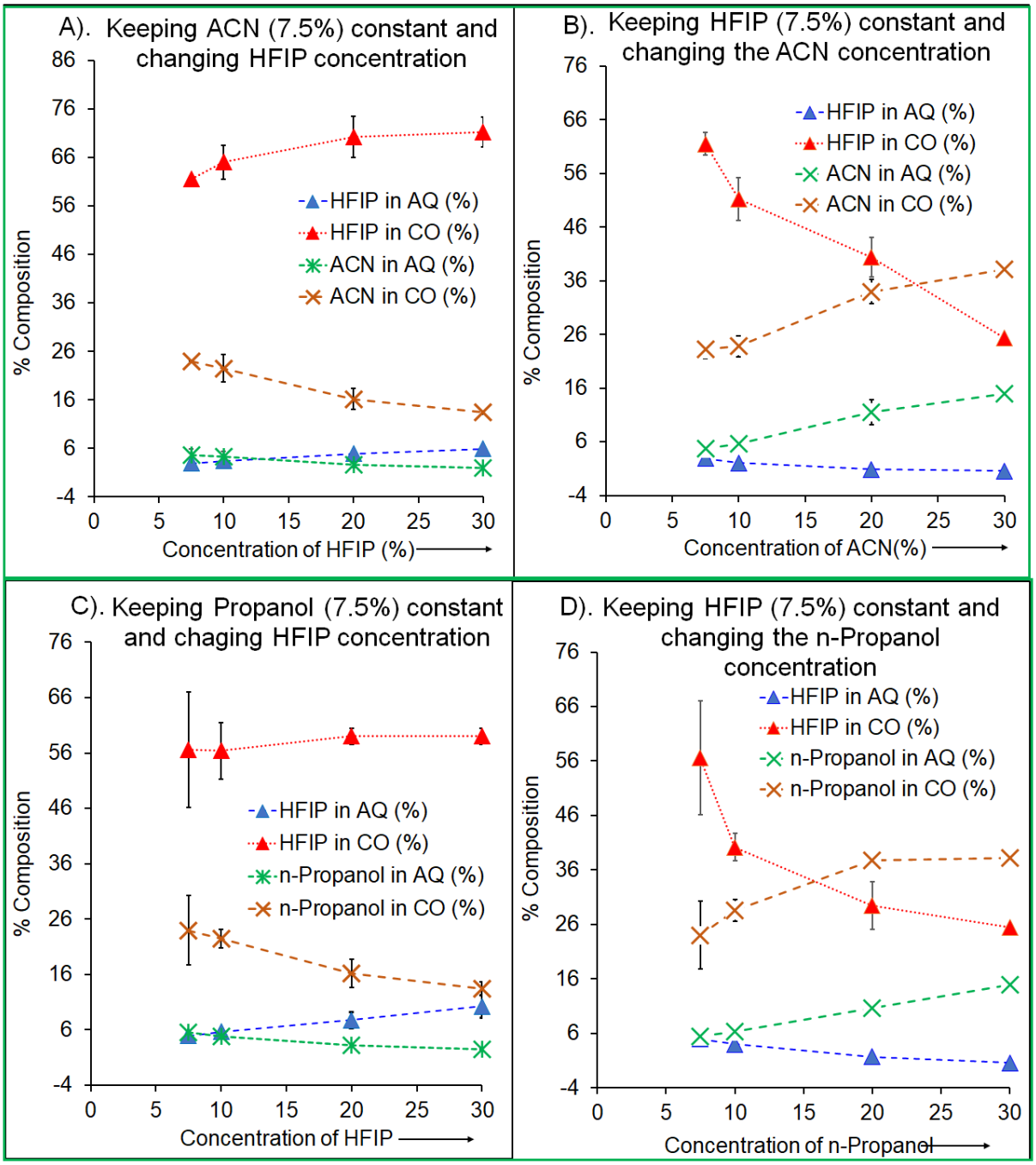


Figure 6: Compositional analysis of organic solvents (ACN and n-propanol) and HFIP into the top aqueous phase and bottom organic phase of different biphasic systems of ACN+ HFIP+ water and n-propanol+ HFIP+ water from top to bottom respectively.

Table 3: The table represents the total organic solvents including HFIP into the top aqueous phase and bottom organic phase of different biphasic systems.

Systems	Total organic solvents in the bottom phase	Standard deviation (%)	Total organic solvents in the top aqueous phase	Standard deviation
7.5% ACN+ 7.5% HFIP+ 85% water	61.5+24.0=85.5%	HFIP=3.5 ACN=2.8	4.7+2.9=7.6 %	HFIP=0.4 ACN=1.1
7.5% ACN+ 10% HFIP+ 82.5% water	65+22.5=87.5%	HFIP=4.2 ACN=2.1	4.3+3.4=7.7%	HFIP=0.3 ACN=1.1
7.5% ACN+ 20% HFIP+ 72.5% water	70.2+16.2=86.4%	HFIP=3.1 ACN=0.3	4.9+2.7=7.6%	HFIP=0.3 ACN=0.7
7.5% ACN+ 30% HFIP+ 62.5% water	71.2+13.4=84.6%	HFIP=1.1 ACN=3.6	5.9+1.9=7.8%	HFIP=0.1 ACN=0.2
10% ACN+ 7.5% HFIP+ 82.5% water	51.2+23.8=75.0%	HFIP=3.9 ACN=1.9	5.6+2.1=7.7%	HFIP=0.3 ACN=0.8
20% ACN+ 7.5% HFIP+ 72.5% water	40.4+34.0=74.4%	HFIP=3.7 ACN=2.2	11.5+0.9=12.4%	HFIP=0.2 ACN=2.3
30% ACN+ 7.5% HFIP+ 62.5% water	38.2+25.4=63.6%	HFIP=0.6 ACN=1.2	15.0+0.6=15.6%	HFIP=0.1 ACN=1.0
7.5% n-propanol+ 7.5% HFIP+ 85% water	56.6+24.0=80.6%	HFIP=10.4 n-propanol=6.2	5.5+4.9=10.4%	HFIP=0.4 n-propanol=0.1
7.5% n-propanol+ 10% HFIP+ 85% water	56.4+22.5=78.9%	HFIP=5.1 n-propanol=1.7	5.7+4.8=10.5%	HFIP=0.4 n-propanol=0.1
7.5% n-propanol+ 20% HFIP+ 85% water	59.0+16.2=75.2%	HFIP=1.4 n-propanol=2.5	7.7+3.2=10.9%	HFIP=1.5 n-propanol=0.1
7.5% n-propanol+ 30% HFIP+ 85% water	59.0+13.4=72.4%	HFIP=1.4 n-propanol=1.3	10.2+2.5=12.7%	HFIP=2.0 n-propanol=0.1
10% n-propanol+ 7.5% HFIP+ 85% water	40.2+28.6=68.8%	HFIP=2.5 n-propanol=1.9	6.4+4.0=10.4%	HFIP=0.7 n-propanol=0.1

20% n-propanol+ 7.5% HFIP+ 85% water	29.5+37.8=67.3%	HFIP=4.3 n-propanol=1.4	10.7+1.7=12.4%	HFIP=0.3 n-propanol=0.4
30% n-propanol+ 7.5% HFIP+ 85% water	25.4+38.2=63.6%	HFIP=0.4 n-propanol=0.3	15.0+0.6=15.6%	HFIP=0.2 n-propanol=0.2

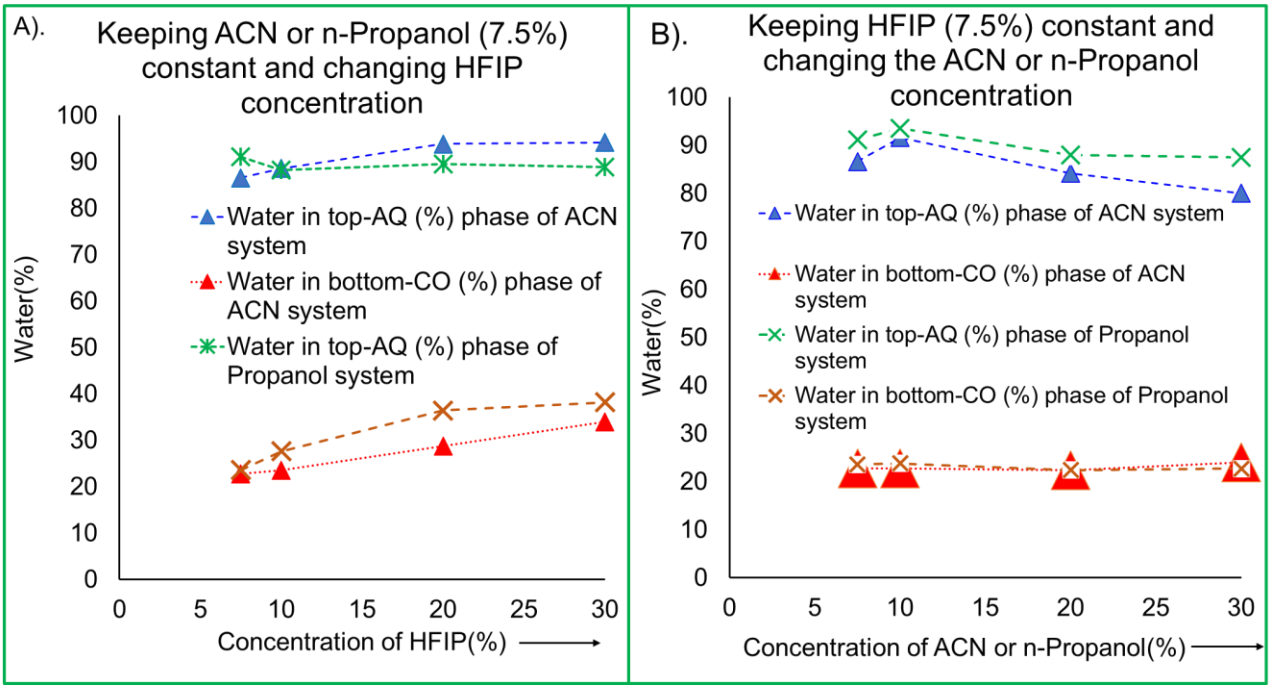


Figure 7: Water analysis into different biphasic systems of ACN and n-propanol with HFIP and water.

5.3.2. Method Development and Validation for the Separation of Polycyclic Aromatic Hydrocarbons

The proposed methods of different aqueous organic biphasic systems were validated for the enrichment of polycyclic aromatic hydrocarbons in terms of limit of detection (LOD), the limit of quantitation (LOQ), linearity, and precision. The calibration curve of each PAH sample in acetonitrile was constructed within the concentration range of 5-100 ng/mL using the GC-MS instrument and LOD and LOQ were calculated based on calibration curves. The calibration curves were made by taking peak area versus concentration of the standard sample. The analysis was done

three times to check the reproducibility of the sample. Figure 8 shows the calibration curve of PAH samples. As shown in the figure, most of the PAH samples follow the linearity range with a regression value of 0.99. Additionally, the list in table 4 represents the limit of detection (LOD) and limit of quantitation (LOQ) of PAH samples from the calibration plots. The LOD is calculated using the formula 3.3* (standard deviation of intercept(response)/ slope) and LOQ is calculated from LOD (LOQ = 3* LOD). The LOD obtained for PAH samples ranges from 3.6 – 10.8 ng/mL and LOQ ranges from 10.9 – 32.9 ng/mL.

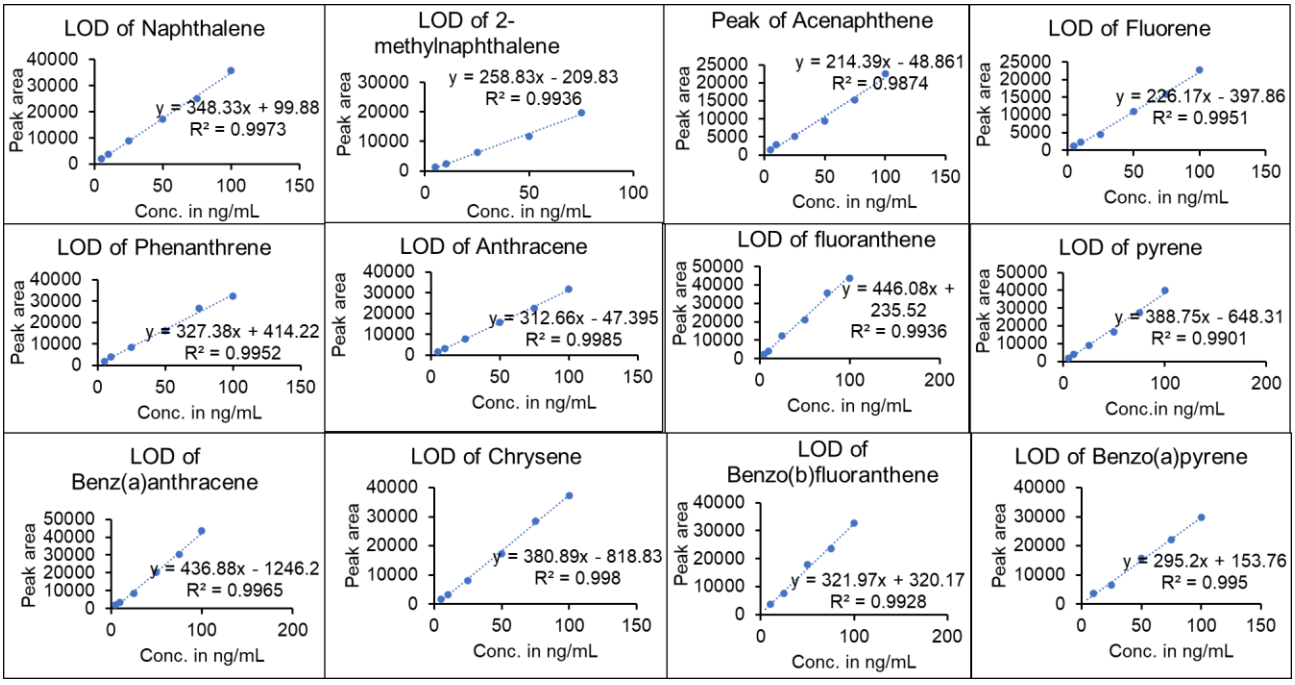


Figure 8: Calibration curves of polycyclic aromatic hydrocarbons for the calculation of limit of detection

Table 4: Calculation of limit of detection (LOD) and limit of quantitation (LOQ) of analytes detected by GC-MS instrument.

Analytes	Regression equation	R^2 value	Conc. range of analytes (ng/mL)	Limit of detection (ng/mL)	Limit of quantitation (ng/mL)
Naphthalene	y=348.33x+99.88	R^2=0.9973	5-100	4.8	14.5
2-Methylnaphthalene	y=258.83-209.83	R^2=0.9936	5-75	6.5	19.5

Acenaphthene	y=214.39x- 48.86	R^2=0.9874	5-100	10.5	31.7
Fluorene	y=226.17x-397.86	R^2=0.9951	5-100	6.5	19.6
Phenanthrene	y=327.38x + 414.22	R^2=0.9952	5-100	6.4	19.5
Anthracene	y= 312.66x - 47.395	R^2=0.9985	5-100	3.6	10.9
Fluoranthene	y=446.08x + 235.52	R^2=0.9936	5-100	7.4	22.4
Pyrene	y= 388.75x - 648.31	R^2=0.9901	5-100	9.3	28.1
Benz(a)anthracene	y= 436.88x - 1246.2	R^2=0.9965	5-100	5.5	16.5
Chrysene	y= 380.89x - 818.83	R^2 =0.998	5-100	4.1	12.5
Benz(b)fluoranthene	y=321.97x + 320.17	R^2=0.9928	10-100	9.9	30.2
Benz(a)pyrene	y = 295.2x + 153.76	R^2=0.995	10-100	8.3	25.1
Perylene	y=352.59x + 56.244	R^2=0.9915	10-100	10.8	32.9

5.3.3. Extraction and Enrichment Study of Polycyclic Aromatic Hydrocarbons (PAH) by four AOSB Systems.

Four associated organic solvents biphasic systems of acetonitrile, n-propanol, acetone, and tetrahydrofuran (THF) with HFIP in water with the composition of 7.5% (organic solvent) + 7.5% HFIP+ 85% water were used for the extraction and enrichment of 13 polycyclic aromatic hydrocarbons. Different volumes of the bottom HFIP-organic phase were observed for the four organic solvent system as THF (~95µL) >acetonitrile (~75µL) >acetone (~55µL) > n-propanol (~45µL). The preliminary work was done with standard PAH samples and the method was optimized. The chromatograms in figure 9 showed the response of polycyclic aromatic hydrocarbons in the control with no phase separation and after their enrichment into the bottom organic-HFIP phase of different biphasic systems is at least 10 times higher. None of the polycyclic aromatic compounds were detected into the aqueous phase of these systems, which is due to their highly hydrophobic nature as well as they might be below the detection limit of instrument. The enrichment factor for each PAH into the bottom HFIP-organic phase of each system was determined. The fold enrichment is directly proportional to the partition coefficient value and inversely proportional to the volume of the bottom phase and is calculated using equation 1 below. (Note: calibration curves were used for the calculation of unknown concentration of PAH samples

which are shown in the supplementary information in section S1) and results were shown in Table 4). Most of the PAH sample showed the comparable enrichment factor to the bottom phase volume of one system however the different systems showed different enrichment factor due to the change in bottom phase volume. As shown in table 4, the rank ordering of the enrichment factor was: n-propanol> acetone> acetonitrile> THF, which is nearly the opposite of the volume fractions of these phases. This result indicates that the lower bottom phase volume results in a higher enrichment. The results obtained were again conformed by taking the two sets of experiments with 7.5% acetonitrile+ 4% HFIP+ 88.5% water and 7.5% acetone+ 3.5% HFIP+ 89% water with the same bottom organic phase volume of around ~30 µL and samples were analyzed to calculate the fold enrichment. The results were given in table 5.

E.F. = Concentration of sample into coacervate phase after enrichment ----------(1)

Initial concentration of sample before enrichment

Table 5: Fold enrichment of 13 polycyclic aromatic hydrocarbons into a bottom organic phase of each system

Fold enrichment into bottom organic-HFIP phase				
Analytes	7.5%n-propanol+7.5% HFIP+85% water system	7.5% acetone+ 7.5%HFIP+85 %water system	7.5% ACN+ 7.5%HFIP+85 %water system	7.5% THF+ 7.5%HFIP+85 %water system
Naphthalene	13	13	11	9
2-Methyl naphthalene	24	17	13	8
Acenaphthene	12	10	10	7
Fluorene	12	10	10	8
Phenanthrene	18	14	13	9
Anthracene	10	7	7	5
Fluoranthene	14	12	10	10
Pyrene	11	9	8	6
Benz(a)anthracene	17	15	10	7
Chrysene	12	8	6	6

Benz(b)fluoranthene	14	9	8	6
Benz(a)pyrene	18	11	9	7
Perylene	12	11	9	9

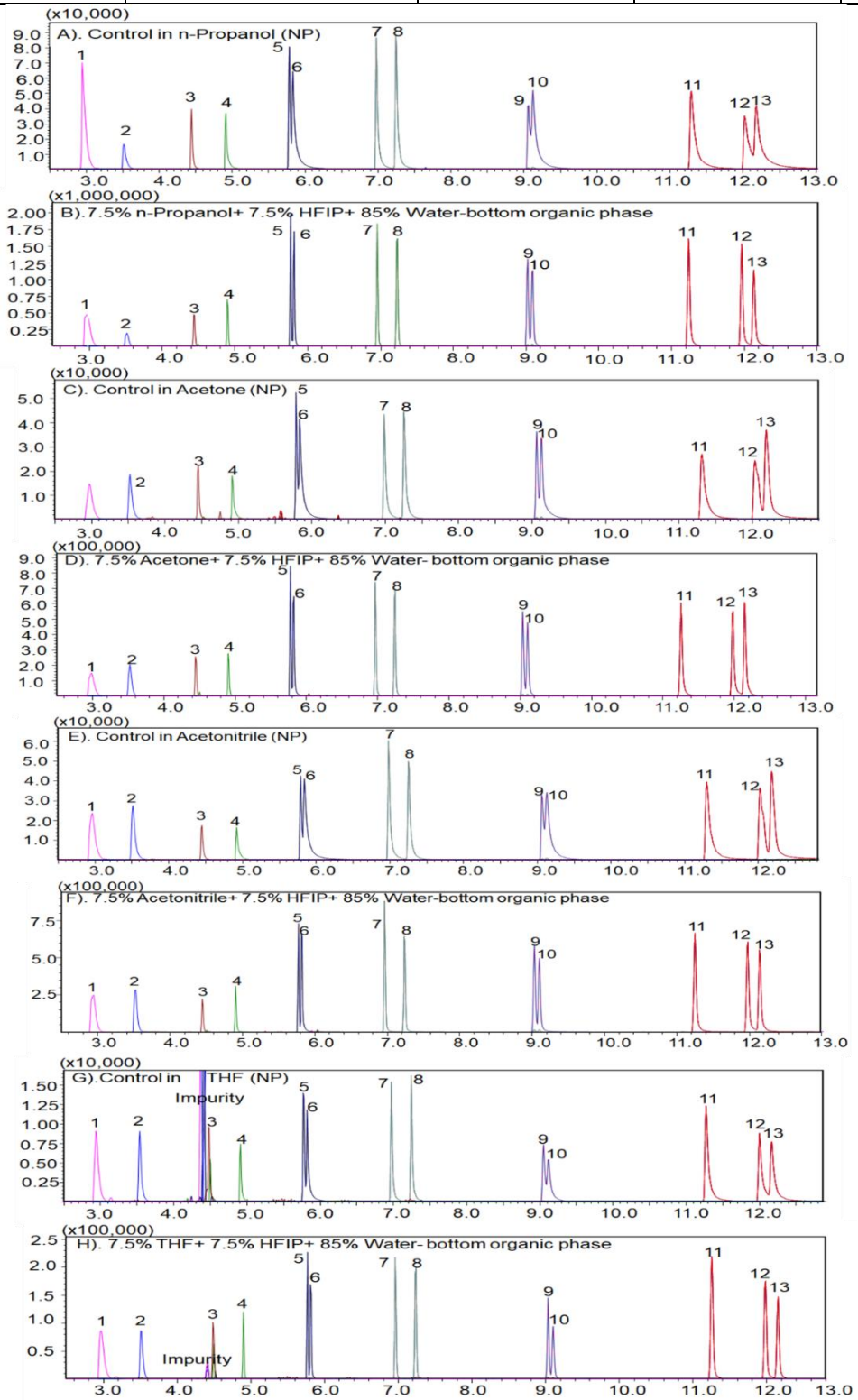


Figure 9: Chromatograms showing the identification of PAH samples into a bottom organic phase of different systems versus control, A). Control in n-Propanol with no phase separation (NP), B). 7.5% n-Propanol+7.5% HFIP+85%Water-bottom organic phase, C). Control in Acetone with no phase separation (NP), D). 7.5% Acetone+7.5%HFIP+85%Water-bottom organic phase, E). Control in Acetonitrile with no phase separation (NP), F).7.5%Acetonitrile+7.5%HFIP+85%Water-bottom organic phase, G). Control in THF with no phase separation (NP), H). 7.5%THF+7.5%HFIP+85%Water-bottom organic phase.

Table 6: Comparative study of enrichment of PAH samples into a bottom organic phase of two systems

Fold enrichment into bottom organic-HFIP phase		
Analytes	7.5% Acetonitrile +4% HFIP+ 88.5% water system	7.5% Acetone+ 3.5%HFIP+89%water system
Naphthalene	22.7	23.5
2-Methlynaphthalene	27.1	28.3
Acenaphthene	25.9	27.0
Fluorene	26.7	26.8
Phenanthrene	35.0	36.5
Anthracene	22.2	24.9
Fluoranthene	30.6	32.4
Pyrene	25.3	28.7
Benz(a)anthracene	29.6	37.7
Chrysene	29.9	31.6
Benz(b)fluoranthene	28.5	34.6
Benz(a)pyrene	36.4	41.4
Perylene	30.4	31.2

Table 6 shows the comparable fold enrichment for the PAH samples into bottom organic phase of these two biphasic systems with the same bottom organic phase volume of 30µL: 7.5% ACN+ 4%HFIP+88.5% water and 7.5% Acetone+ 3.5%HFIP+ 89% water and shows comparable enrichment factor with each other. Also, table shows the increased enrichment factor in both cases

as compared to the one observed in table 5 with larger bottom organic phase. This will again define that enrichment factor is inversely proportional to the volume of enriching phase.

Finally, the extraction of PAH from soil sample was done using procedure as mentioned in experimental section. The extracted samples in 80:20 water: organic solvent mixture was then enriched using different AOSB systems. The chromatograms obtained were shown in supplementary information in figure S2. Initially, the samples in control were not detected by GC-MS analysis, however after their enrichment using associated organic biphasic system bottom phase they have been enriched and detected easily by GC-MS analysis.

5.3.4. Extraction and Enrichment Study of Pesticides

The extraction of pesticides sample was carried out as mentioned in the experimental section. The calibration curves were made for the measurement of concentration of pesticides in each sample and calibration curves are shown in figure 10. They showed the linearity range of 0 to 80 ppm concentration of each sample. The distribution of samples into top aqueous phase and the bottom organic phase was analyzed and partition coefficient and fold enrichment each system into their bottom organic phase was calculated from response factor. The graph in figure 11 shows the partition coefficient and fold enrichment of six different pesticides into the bottom organic phase of two systems. The table in the right side of figure shows the partition coefficient of six pesticides into the bottom phase of each system. In acetone+HFIP system, most of the pesticides were extracted into the bottom phase without fractionation since no pesticides were detected into aqueous phase of system however, acetonitrile+HFIP system showed some fractionation of some pesticides like flutriafol, DDE, and DDT into two phases. Instead, their partition coefficient is more than 100 times to the bottom organic phase indicate that their high affinity for the bottom organic phase rather than the top aqueous phase. Similarly, the left side of the figure (bar graph)

shows the fold enrichment of each pesticide into the bottom phase of the two systems. The values are comparable and shows the fold enrichment of more than 25 times towards bottom organic phase of each system except for DDE and DDT shows slightly lower fold enrichment into bottom organic phase of ACN+HFIP+water. This is because these two samples are partitioning into the two phases of acetonitrile+HFIP system. The fold enrichment is directly proportional to the partition coefficient of sample into tw phases.

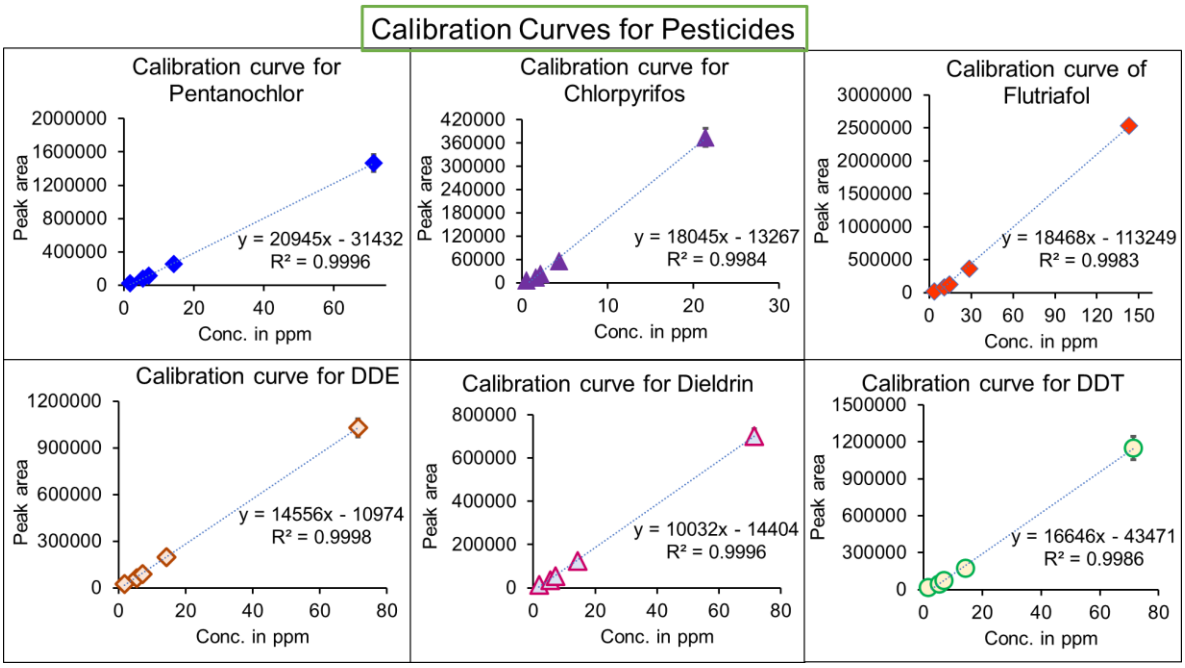
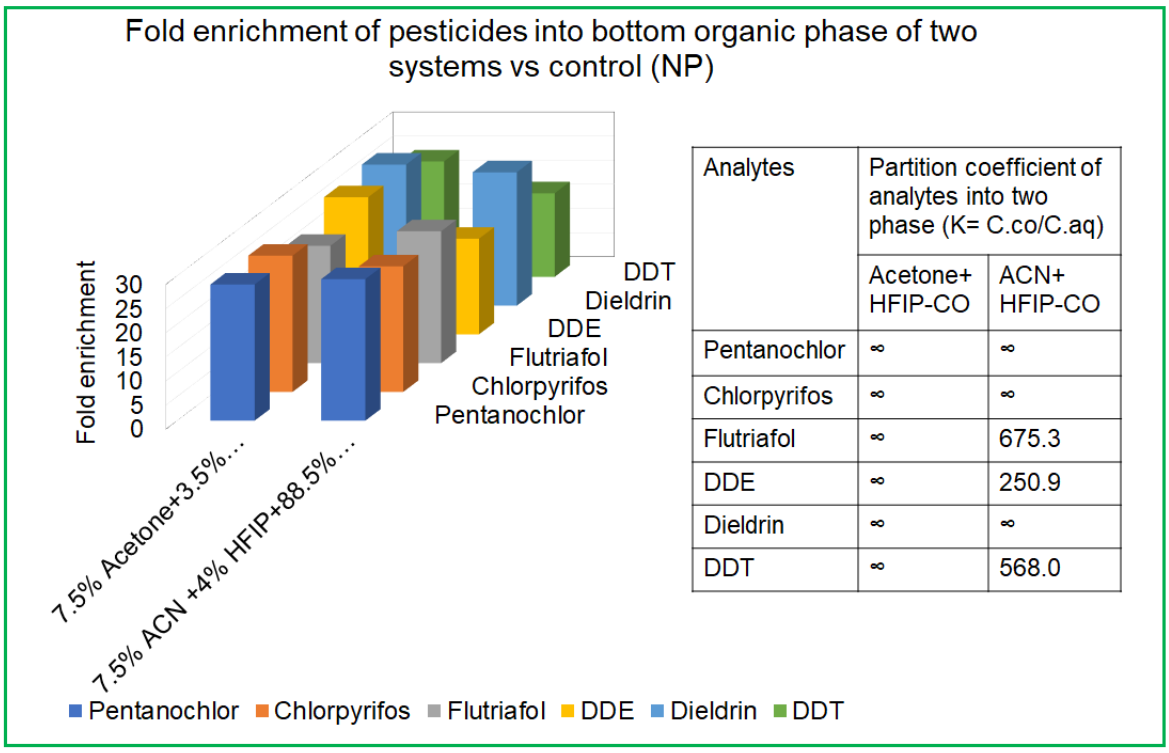


Figure 10: Calibration curves for pesticides in acetone using GC-MS instrument.



Analytes	Partition coefficient of analytes into two phase (K= C.co/C.aq)	
	Acetone+ HFIP-CO	ACN+ HFIP-CO
Pentanochlor	∞	∞
Chlorpyrifos	∞	∞
Flutriafol	∞	675.3
DDE	∞	250.9
Dieldrin	∞	∞
DDT	∞	568.0

Figure 11: Figure shows the fold enrichment and partition coefficient of six pesticide samples into the bottom organic phase of two biphasic systems.

5.3.5. Extraction and Enrichment Study of Sex Steroid Hormones

The extraction of sex steroid hormones was done as mentioned above, and their concentration was analyzed into two phases of each system using the peak area observed in GC-MS analysis. The calibration curves were made for each hormone to check its linearity and to calculate the concentration of unknown concentration of the sample. The calibration curves of each hormone are shown in figure 12. The calibration curve shows the linearity with regression value of 0.99 in the range of 0-180 ppm for most of the hormones except estriol (0-400 ppm). Additionally, the extraction and enrichment study were done with a mixture of a sample into two organic-aqueous biphasic systems: 7.5%ACN+ 4%HFIP+88.5% water and 7.5% acetone+3.5% HFIP+ 89% water. The results are shown in Table 6 below. In table 6, initial concentration of the sample in control

with no phase separation was calculated for each system and then concentration of each sample into the top aqueous phase and the bottom organic phase was detected. The partition coefficient and fold enrichment of each sample into bottom organic phase of each system was then calculated and were shown in table from left to right column. As listed in table, all of four hormones have some fractionation into the top and bottom phase of each system however, they show their higher affinity for bottom organic phase. Partition coefficient value for hormones in Acetone+ HFIP system is slightly greater than ACN+HFIP system and so does the higher fold enrichment. Among four sex steroid hormones, estriol is more hydrophilic in nature due to the presence of 3 OH functional group in its structure therefore its partition coefficient towards bottom phase of two systems is relatively lower than other hormones like estrone, estradiol, and progesterone. The hormones with no polar functional group like progesterone have very high affinity for the hydrophobic bottom organic phase.

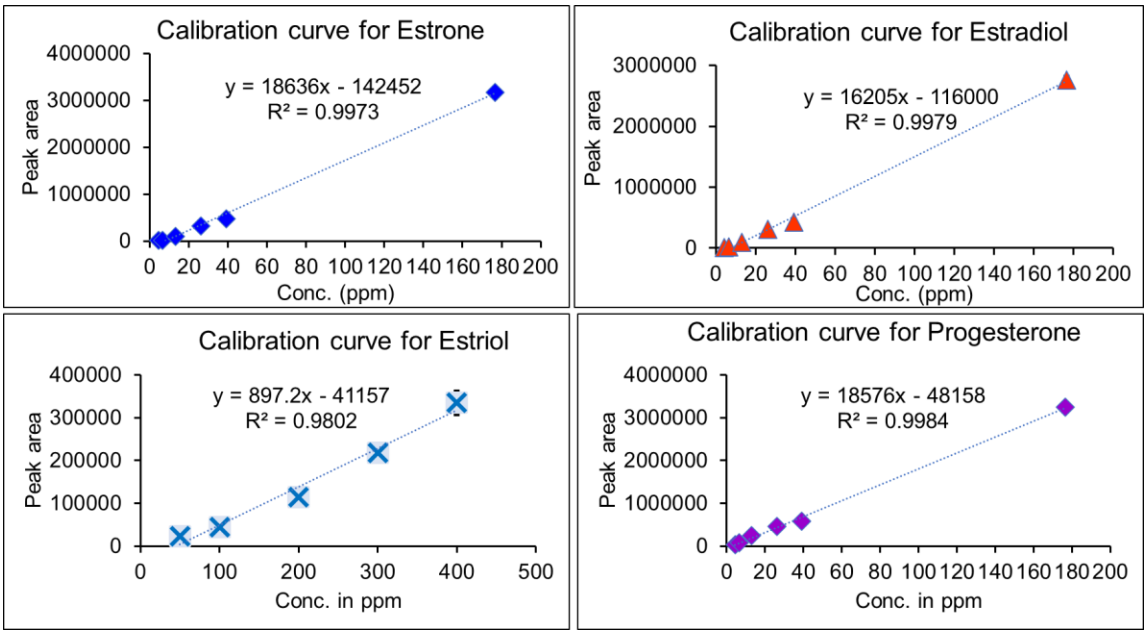


Figure 12: Calibration curves of four sex steroid hormones in acetone using GC-MS analysis.

Table 7: Partition coefficient and fold enrichment of sex steroid hormones into the bottom organic phase of two organic-aqueous biphasic systems.

Standards	Initial conc in 50/50-acetone/water as control (ppm)	Conc into 7.5%acetone+ 3.5%HFIP+ 89%water-top aqueous-phase	Conc into 7.5%acetone+ 3.5%HFIP+ 89%water-bottom organic-phase	Partition coefficient (K=Conc into bottom phase/conc into top phase)	Fold enrichment into bottom-organic-phase (conc into bottom phase/initial conc)
Estrone	24.0	9.6	153.4	63.9	25.6
Estradiol	28.5	10.3	234.4	91.0	32.9
Progesterone	16.0	2.7	146.4	216.9	36.6
Estriol	60.0	163.5	127.9	3.1	8.5
Standards	Initial conc in 50/50-ACN/water as control (ppm)	Conc into 7.5%ACN+ 4%HFIP+ 88.5%water-top aqueous-phase	Conc into 7.5%ACN+ 4%HFIP+ 88.5%water-bottom organic-phase	Partition coefficient (K)	Fold enrichment into bottom-organic-phase
Estrone	18.1	11.9	107.7	36.2	23.8
Estradiol	20.7	13.0	124.1	38.2	23.9
Progesterone	12.5	2.8	83.8	119.7	26.8
Estriol	58.9	97.9	89.1	3.6	6.0
Note: All the coacervate phase are four times diluted					

5.4. CONCLUSION: The HFIP induced biphasic systems of polar organic solvents in water are useful for the extraction and enrichment of small organic molecules including environmental pollutants like PAHs, pesticides, and sex steroid hormones. Detection and enrichment of these environmental pollutants is of great importance since they are found in trace level in environment and have acute and chronic side effects on human health and other lives. The AOSB biphasic systems provide an effective, greener, easy, and economical option for sample extraction and enrichment as compared to the existing LLC and SPE methods.

5.5. REFERENCES:

1. Naing, N.N., K.B. Yeo, and H.K. Lee, *A combined microextraction procedure for isolation of polycyclic aromatic hydrocarbons in ambient fine air particulate matter with determination by gas chromatography-tandem mass spectrometry.* Journal of Chromatography A, 2020. **1612**: p. 460646.
2. Deng, W., et al., *A density-tunable liquid-phase microextraction system based on deep eutectic solvents for the determination of polycyclic aromatic hydrocarbons in tea, medicinal herbs and liquid foods.* Food Chemistry, 2021. **352**: p. 129331.
3. Shi, Y., et al., *Facile fabrication of composited solid phase microextraction thin membranes for sensitive detections of trace hydroxylated polycyclic aromatic hydrocarbons in human urine.* Analytica Chimica Acta, 2021. **1158**: p. 338422.
4. King, A., J. Readman, and J. Zhou, *Determination of polycyclic aromatic hydrocarbons in water by solid-phase microextraction–gas chromatography–mass spectrometry.* Analytica Chimica Acta, 2004. **523**(2): p. 259-267.
5. Yu, C., et al., *Porphyrin-based covalent organic framework coated stainless steel fiber for solid-phase microextraction of polycyclic aromatic hydrocarbons in water and soil samples.* Microchemical Journal, 2021. **168**: p. 106364.
6. Ju, H., et al., *Development of candidate reference method for accurate determination of four polycyclic aromatic hydrocarbons in olive oil via gas chromatography/high-resolution mass spectrometry using 13C-labeled internal standards.* Food Chemistry, 2020. **309**: p. 125639.

7. Li, N., et al., *Effective Enrichment and Detection of Trace Polycyclic Aromatic Hydrocarbons in Food Samples based on Magnetic Covalent Organic Framework Hybrid Microspheres.* Journal of Agricultural and Food Chemistry, 2018. **66**(13): p. 3572-3580.
8. Nukatsuka, I., et al., *A thin-layer solid-phase extraction–liquid film elution technique used for the enrichment of polycyclic aromatic hydrocarbons in water.* Journal of Separation Science, 2021. **44**(9): p. 1989-1997.
9. Veyrand, B., et al., *Innovative method for determination of 19 polycyclic aromatic hydrocarbons in food and oil samples using gas chromatography coupled to tandem mass spectrometry based on an isotope dilution approach.* Journal of Chromatography A, 2007. **1149**(2): p. 333-344.
10. Fantke, P. and O. Jolliet, *Life cycle human health impacts of 875 pesticides.* The International Journal of Life Cycle Assessment, 2016. **21**(5): p. 722-733.
11. Zhang, X., et al., *Work-related pesticide poisoning among farmers in two villages of Southern China: a cross-sectional survey.* BMC Public Health, 2011. **11**(1): p. 429.
12. Liu, Y., et al., *Pesticides in persimmons, jujubes and soil from China: Residue levels, risk assessment and relationship between fruits and soils.* Science of The Total Environment, 2016. **542**: p. 620-628.
13. Özkara, A., D. Akyıl, and M. Konuk, *Pesticides, environmental pollution, and health*, in *Environmental health risk-hazardous factors to living species*. 2016, IntechOpen.
14. Xu, Y., et al., *Zirconium(IV)-based metal-organic framework for determination of imidacloprid and thiamethoxam pesticides from fruits by UPLC-MS/MS.* Food Chemistry, 2021. **344**: p. 128650.

15. Dai, X., et al., *Sorption and desorption of sex hormones in soil-and sediment-water systems: A review.* Soil Ecology Letters, 2021: p. 1-17.

16. Chang, H., et al., *Occurrence of androgens and progestogens in wastewater treatment plants and receiving river waters: comparison to estrogens.* water research, 2011. **45**(2): p. 732-740.

17. Shore, L.S. and M. Shemesh, *Naturally produced steroid hormones and their release into the environment.* Pure and Applied Chemistry, 2003. **75**(11-12): p. 1859-1871.

18. Uchoa, M.F., V.A. Moser, and C.J. Pike, *Interactions between inflammation, sex steroids, and Alzheimer's disease risk factors.* Frontiers in Neuroendocrinology, 2016. **43**: p. 60-82.

19. Varticovski, L., et al., *Endocrine disruptors of sex hormone activities.* Molecular and Cellular Endocrinology, 2021: p. 111415.

20. Ojoghoro, J.O., M.D. Scrimshaw, and J.P. Sumpter, *Steroid hormones in the aquatic environment.* Science of The Total Environment, 2021. **792**: p. 148306.

21. Belfroid, A., et al., *Analysis and occurrence of estrogenic hormones and their glucuronides in surface water and waste water in The Netherlands.* Science of the Total Environment, 1999. **225**(1-2): p. 101-108.

22. Baronti, C., et al., *Monitoring natural and synthetic estrogens at activated sludge sewage treatment plants and in a receiving river water.* Environmental Science & Technology, 2000. **34**(24): p. 5059-5066.

23. Labadie, P. and H. Budzinski, *Development of an analytical procedure for determination of selected estrogens and progestagens in water samples.* Analytical and bioanalytical chemistry, 2005. **381**(6): p. 1199-1205.

24. Kolpin, D.W., et al., *Pharmaceuticals, hormones, and other organic wastewater contaminants in US streams, 1999− 2000: A national reconnaissance.* Environmental science & technology, 2002. **36**(6): p. 1202-1211.

25. Lin, J.-H. and W.-L. Tseng, *Gold nanoparticles for specific extraction and enrichment of biomolecules and environmental pollutants.* Reviews in Analytical Chemistry, 2012. **31**(3-4): p. 153-162.

26. Liu, J.f., Y.g. Chi, and G.b. Jiang, *Screening the extractability of some typical environmental pollutants by ionic liquids in liquid-phase microextraction.* Journal of separation science, 2005. **28**(1): p. 87-91.

27. Lau, E., S. Gan, and H. Ng, *Extraction techniques for polycyclic aromatic hydrocarbons in soils.* International journal of analytical chemistry, 2010. **2010**.

28. Khaledi, M.G., S.I. Jenkins, and S. Liang, *Perfluorinated Alcohols and Acids Induce Coacervation in Aqueous Solutions of Amphiphiles.* Langmuir, 2013. **29**(8): p. 2458-2464.

29. McCord, J.P., D.C. Muddiman, and M.G. Khaledi, *Perfluorinated alcohol induced coacervates as extraction media for proteomic analysis.* Journal of Chromatography A, 2017. **1523**: p. 293-299.

30. Nejati, M.M. and M.G. Khaledi, *Perfluoro-Alcohol-Induced Complex Coacervates of Polyelectrolyte–Surfactant Mixtures: Phase Behavior and Analysis.* Langmuir, 2015. **31**(20): p. 5580-5589.

31. Koolivand, A., et al., *Fluoroalcohol – Induced coacervates for selective enrichment and extraction of hydrophobic proteins.* Journal of Chromatography B, 2018. **1083**: p. 180-188.

32. Koolivand, A., et al., *Coacervation of Lipid Bilayer in Natural Cell Membranes for Extraction, Fractionation, and Enrichment of Proteins in Proteomics Studies.* Journal of Proteome Research, 2019. **18**(4): p. 1595-1606.

33. Khanal, D.D., et al., *Improved Protein Coverage in Bottom-Up Proteomes Analysis Using Fluoroalcohol-Mediated Supramolecular Biphasic Systems With Mixed Amphiphiles for Sample Extraction, Fractionation, and Enrichment.* Analytical Chemistry, 2021. **93**(20): p. 7430-7438.

34. Azizi, M., et al., *Improving identification of low abundance and hydrophobic proteins using fluoroalcohol mediated supramolecular biphasic systems with quaternary ammonium salts.* Journal of Chromatography A, 2021. **1655**: p. 462483.

CHAPTER 6: SUMMARY AND FUTURE WORK

Summary: Fluoroalcohol or fluoroacids induced biphasic systems of different surfactants, amphiphiles, polyelectrolytes, sugar or polar organic solvents in water is a green technique for the extraction and enrichment of variety of samples in proteomics and environmental analysis. In this research, we have investigated different zwitterionic surfactants and mixed amphiphiles for the extraction, enrichment, and fractionation of complex mixture of yeast proteins. We found that HFIP induced supramolecular biphasic systems (FA*i*C-BPS) of zwitterionic surfactants and their mixed amphiphilic systems with quaternary ammonium salts (QUATS) are improving identification of yeast proteins by at least 11% up to 18.1% with respect to the 8M urea (NP) as the control. The control study without phase separation was done with different surfactants like SDS, SC, SDC, DMMAPS, CHAPS, TEAB, TBAB and 8M urea and urea performed best among them. The bottom coacervate phase of biphasic systems are concentrated with surfactants/amphiphiles and HFIP with little water in a small volume and is usually hydrophobic in nature and therefore, are useful for solubilization of hydrophobic membrane proteins and enrichment of lower abundance proteins. In this research we have reported the extraction of membrane proteins by CHAPS systems at least 11.5% higher than control which increased up to 16.5% by the addition of TBAB to CHAPS. However, DMMAPS systems showed the superiority to CHAPS systems for the extraction of membrane proteins by 16.5% without addition of QUATS and 18.8% with the addition of TBAB. Interestingly, identification improvement of integral component of membrane proteins and proteins with alpha helices are increased by 25.5% and 555% by DMMAPS+TBAB+HFIP and DMMAPS+HFIP system respectively. The data analysis showed that most of the membrane proteins are extracted into bottom coacervate phase. Similarly,

biphasic systems of CHAPS, DMMAPS and their mixed amphiphilic systems with QUATS are great for the extraction and enrichment of lower abundance proteins with abundance < 2000 molecules/cells. A single system shows the identification improvement of lower abundance proteins up to 114%.

In addition, these biphasic systems help fractionation of complex sample mixture into two phases. The long chain surfactants like DMMAPS and CHAPS exhibit the fractionation based on strong hydrophobic interaction between coacervate phase and proteins. However, addition of positively charged small chain quaternary ammonium salts alters the fractionation of proteins by introducing electrostatic interaction in addition to hydrophobic interaction. We have observed that addition of TEAB and TBAB to CHAPS and DMMAPS help the extraction of most of the basic proteins into top aqueous phase and acidic proteins into bottom coacervate phase. This shows that we can alter the fractionation of certain types of complex sample mixture by adding different positively or negatively charged salts or amphiphiles to the coacervate systems of long chain surfactants/amphiphiles.

In this dissertation we have also investigated the HFIP induced biphasic systems of polar organic solvents like ACN, acetone, n-propanol and THF and their application for the extraction and enrichment of environmental pollutants like PAHS, pesticides and sex steroid hormones. A comparative study between different systems showed that they can form the different volume of bottom organic phase even with same composition based on their structural difference and has the different fold enrichment for small molecules accordingly. We found that lower the volume of bottom organic higher is the fold enrichment and vice-versa. We have reported the fold enrichment of small molecules from 25 to 41 times higher than control into 30 µL bottom organic phase of acetonitrile+ HFIP and acetone+ HFIP systems if there is little or no fractionation into two phases.

Future work: The use of HFIP induced supramolecular biphasic systems of zwitterionic surfactants and mixed amphiphiles are promising approach for the extraction, enrichment, and fractionation of standard proteins as well as complex sample mixture like yeast proteins. These systems are especially useful for the extraction and enrichment of hydrophobic membrane proteins and lower abundance proteins. Addition of new approaches in sample preparation in proteomics will be of great achievement. We believe that the above optimized techniques are not limited to the standard protein mixtures and yeast proteome only. Therefore, our future plan is to use these systems for the extraction, enrichment and fractionation of proteins from human cells. Extraction of missing membrane and lower abundance proteins from human cell lines would have great impact on discovery of new therapeutics. From our overall research experience, we have observed that change in composition of systems changes the fractionation pattern of the proteins into two phase and helpful to improve the protein coverage by simplifying the complex mixture. We want to explore several chemicals that help the selective and better fractionation of protein sample mixtures into two phases.

Similarly, use of associated organic solvents biphasic systems (AOS-BPS) for the enrichment of environmental sample is another part of work. The previously used systems for the extraction and enrichment of polycyclic aromatic hydrocarbons from real sample needs to be optimized and will be one of our future works.

APPENDIX A: Supporting Information for Chapter 2

Section S 1. Evaluation of coacervate phase of variety of concentration of DMMAPS with of HFIP, Cell lysis and sample preparation.

HFIP induced coacervation of DMMAPS was done using different concentration of DMMAPS and phase transition behavior was observed as shown in figure below. With the increase of concentration of DMMAPS increases the volume of coacervate phase which shows the characteristic of coacervation process. For the protein extraction from yeast sample 50 mM DMMAPS+ 10% was used initially.

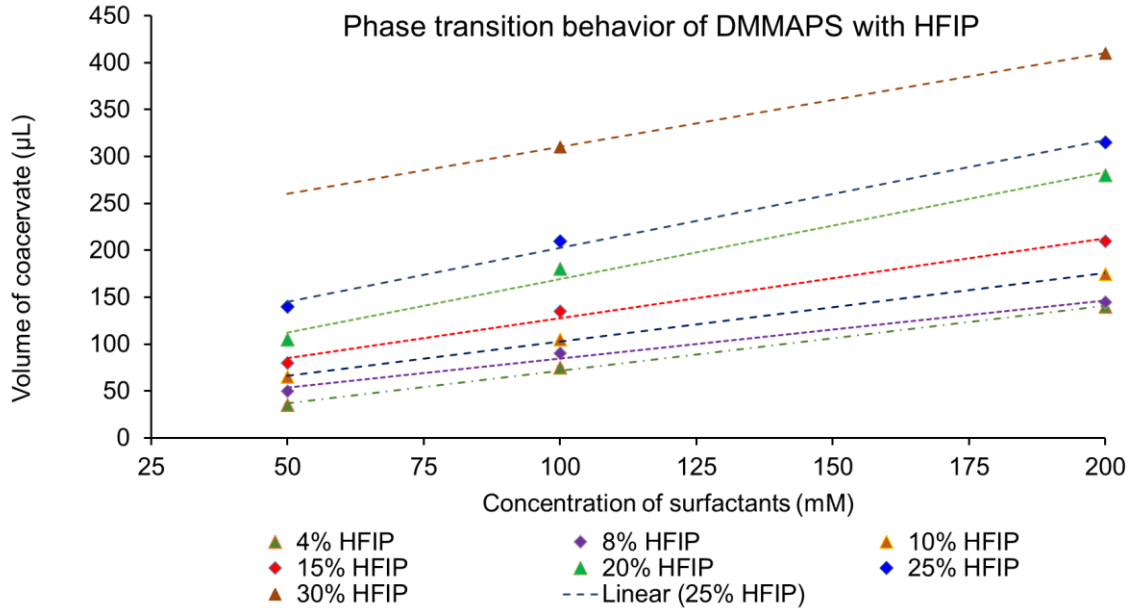


Figure SI-1 Phase transition behavior of DMMAPS with HFIP in water

Cell lysis and sample preparation: Saccharomyces cerevisiae cells were grown in our lab using YPD broth (Fisher, USA) for 16 to 20 h in a shaker incubator at 30 degrees Celsius. Cell pellets were grinded with liquid nitrogen to break the cell wall. Glass beads (diameter 500 µm) were soaked in concentrated hydrochloric acid for 16 hours, and then washed thoroughly with distilled water (until pH reaches 7). After cleaning, they were dried at 150-200 °C for 16 hours and cooled to 4 °C before use. 50 mM ammonium bicarbonate buffer containing protease inhibitor and

pepstatin (lysis buffer) was added to the grinded cells (1:1, V/V), and an equal volumetric amount of glass beads were added. Cells were lysed using cell lyser at the frequency of 30 Hz. This process was repeatedly done 8 times, and cells were cooled between each step. Finally, the optical density of the lysed cell was measured, and the concentration of the lysed cell was adjusted in the range of 2-6 mg/mL by UV-absorbance at 590 nm using Bradford protein assay. The prepared yeast samples were kept at -80 degrees Celsius for future use.

Sample preparation was done taking 50 µL of yeast sample which corresponds to 400 µg of proteins and did coacervation with 50 mM DMMAPS+ 10% HFIP and modified systems. After coacervation, the mixture was vortexed for 2 min, sonicated for 2 min, and finally subjected to centrifuge at the speed of 10xg for 15 min. The clearly observed two phases were then separated into two Eppendorf tubes with the help of syringe. Further, sample preparation was done with filter aided sample preparation (FASP) protocol.

Section S 2: Filter aided sample preparation (FASP) protocol for control

1- Condition the filter by adding 500 µL UTT solution, then centrifuge at 14,000g for 5 min, 1/3 of the UTT buffer should pass the filter. Again, centrifuge at 14,000g until a thin layer of UTT remains in the filter.
2- Dissolve the protein in 5M urea and 2 M thiourea, sonicate for 5 minutes and load the sample to the to the FASP filters
3- Centrifuge at 14,000g for 40 min (If necessary, centrifuge again at 14,000g, until the volume reaches to about 20 µL)
4- If you see cell debris, add 200 µL 70% IPA, centrifuge at 14,000g for 40 min, otherwise, go to step 6

(or more than 40 min until about 20 µL of sample remains in the filter) (each time you add a solution, mix it up and down by pipet: if you want to add 200 µL, first add 100 µL, mix it with pipet, and then wash the same pipet with another 100 µL solution)

5- If you see cell debris, add another 200 µL 70% IPA, centrifuge at 14,000g for 40 min (or more than 40 min until about 20 µL of sample remains in the filter)

6- Add 200 µL UTT solution (UTT solution is 5 M urea, 2 M thiourea, in 0.1 M tris buffer, pH= 8.5), using a pipet break the precipitate until you see a uniform liquid, centrifuge at 14,000g for 40 min

In a 1.6 mL vial, add 39 mg DDT to 1mL of UTT buffer to make 250 mM DTT in UTT buffer.

7- Add 20 µL 250mM DTT to each sample and bring the Vol. to 200 µL to bring the concentration of DTT to 25 mM with UTT solution (for example if the thin layer is 20 µL, add 20 µL 250 mM DTT and then add 160 µL UTT solution). Using a pipet mix it, vortex for 30 sec at 600 rpm, incubate at 37 ℃ for 45 min

8- Cool the sample to room temperature

9- Centrifuge at 14,000g for 40 min (or more than 40 min until about 20 µL of sample remains in the filter)

10- Make a stock solution of 250 mM IAA in UTT buffer (add 44 mg IAA and 1 mL UTT buffer) in darkness. Then dilute it to 54 mM. (Mix 216 µL of 250 mM with 784 µL UTT buffer to make a 54 mM IAA in UTT buffer)

11- Add 200 µL IAA 54 mM, then wash the same pipet with 50 µL IAA 54 mM (concentration of IAA must be 50 mM, so if the final volume is 270 µL, and the concentration would be 50 mM). vortex for 30 sec at 600 rmp and incubate at dark for 45 min. **IAA should be made and added in a dark room.**

12- Centrifuge at 14,000 g for 40 min.

13- Add 200 µL UTT buffer solution and centrifuge 14,000 g for 40 min

14- Add 200 µL ABC (50 mM) and centrifuge 14,000 g for 40 min

15- Add 200 µL ABC (50 mM) and centrifuge 14,000 g for 40 min

16- Add 150 µL **ABC 100 mM** (because trypsin as acidic and we want to bring the pH to 7). Then, add trypsin with the ratio of 1:25

(for Aq: (250/25)*2= 20 µL trypsin; and for coacervate: (150/25)*2= 12 µL trypsin)

17- Check the pH to be around 7. Then seal the vials with parafilm. Shake at 600 rpm for 1 min. Incubate in wet chamber at 37 °C for 16 hrs.

18- Transfer the filters to new collection tubes.

19- centrifuge at 14,000 g for 40 min.

20- Add 200 µL **0.5 M NaCl**, and centrifuge at 14,000 g for 40 min in the collection tube.

21- Add 200 µL **50 mM NaCl**, invert the filter, and centrifuge at 1000 g for 2 min in the collection tube.

22- Acidify the sample with TFA to bring the pH below 2 (about 5 µL TFA is enough, DO NOT over-acidify)

23- Desalt the samples:

Desalting procedure:

1- Precondition the C18 Sep-Pak columns with 3 mL CAN, 1 mL 0.1 TFA in 75% CAN, 1 mL 0.1 TFA in 50% CAN, 3 mL 0.1 TFA in water

2- Centrifuge acidified samples at 8000g for 1 min, then load the samples to the columns

3- Wash the samples with 3 mL 0.1 TFA in water

4- Move the Sep-Pak to a 2-ml microcentrifuge tube. Elute the sample with 0.6 ml of 0.1 % TFA in 50% CAN, followed by 0.6 ml of 0.1% TFA in 75%. This step should be performed by gravity, finally push the samples with pipet.

Section S 3: Filter aided sample preparation (FASP) protocol for zwitterionic surfactants

Solutions preparation:

- 100 mM Trisma buffer, pH= 8.5: add 422 mg Tris HCl and 872 mg Tris base in a 100 mL volumetric flask and bring the volume to 100 mL by adding water.
- 0.5 M NaCl: add 1461 mg NaCl in a 50 mL volumetric flask and bring the volume to 50 mL by adding water.
- 50 mM ABC buffer, pH= 7.8: add 395 mg ammonium bicarbonate in a 100 mL volumetric flask and bring the volume to 100 mL by adding water.
- 100 mM ABC buffer, pH= 7.8: add 395 mg ammonium bicarbonate in a 50 mL volumetric flask and bring the volume to 50 mL by adding water.
- UTT solution: 5 M urea and 2 M thiourea in 100 mM tris buffer (pH=8.5):
  Add 15.015 g urea and 7.612 g thiourea in a 50 mL volumetric flask and bring the volume to 50 mL by adding 100 mM tris buffer (pH=8.5)
- Stock solution of 500 mM DMMAPS, 500 mM TBAB and 500 mM TEAB was prepared in DI water.

FASP protocol for coacervation of surfactants with 10 % HFIP

A) Coacervation

1- Take 400 µg protein (for example if the concentration of cell lysate is 7. 9mg/ml, 50 µl cell lysate is approximately equal to 400 µg)

2- In a 1.6 ml vial, add the following (the total Vol is 1 mL):

- 400 µg proteins
- 100 µL of 500 mM DMMAPS
- 100 µL HFIP
- DI water: 1000 – 100 – 100 –50 µL of cell lysate = 750 µL of water
- Note: the volume of water changes when the composition of coacervation system changes.

3- Centrifuge at 10,000g for 15 min and separate two phase

4- Measure the protein concentration for the aqueous and coacervate phase

B) For the coacervate:

24- Dry the coacervate with nitrogen gas for about 1 min (not completely dried) (adjust the flow of nitrogen to prevent drip)

25- Add 450 µL 70% IPA to the coacervate, then add 76 mg thiourea to dissolve it. vortex 30 sec, sonicate 5 min in (not in ice, because it does not dissolve at low temperatures), and load the dissolved coacervate to the to the FASP filters

26- Centrifuge at 14,000g for 40 min (If necessary, centrifuge again at 14,000g, until the volume reaches to about 20 µL)

27- Add 200 µL 70% IPA, centrifuge at 14,000g for 40 min (or more than 40 min until about 20 µL of sample remains in the filter) (each time you add a solution, mix it up and down by pipet: if you want to add 200 µL, first add 100 µL, mix it with pipet, and then wash the same pipet with another 100 µL solution)

28- Add another 200 µL 70% IPA, centrifuge at 14,000g for 40 min (or more than 40 min until about 20 µL of sample remains in the filter)

29- Add 200 µL UTT solution (UTT solution is 5 M urea, 2 M thiourea, in 0.1 M tris buffer, pH= 8.5), using a pipet break the precipitate until you see a uniform liquid, centrifuge at 14,000g for 40 min

Meanwhile doing B, do the part C

C) For the Aqueous:

30- Put the Aqueous phase in concentrator to evaporate HFIP for about 1.5 hrs, until the volume reaches to about 500 µL

31- Condition the filter by adding 500 µL UTT solution, then centrifuge at 14,000g for 5 min, 1/3 of the UTT should pass the filter. Again, centrifuge at 14,000g until a thin layer of UTT remains in the filter.

32- Load the concentrated aqueous to FASP filter and centrifuge at 14,000g for 40 min (or more than 40 min until about 20 µL of sample remains in the filter)

D) For both Aqueous and coacervate:

33- In a 1.6 mL vial, add 39 mg DDT to 1mL of UTT buffer to make 250 mM DTT in UTT buffer.

34- Add 20 µL 250mM DTT to each sample and bring the Vol. to 200 µL to bring the concentration of DTT to 25 mM with UTT solution (for example if the thin layer is 20 µL, add 20 µL 250 mM DTT and then add 160 µL UTT solution). Using a pipet mix it, vortex for 30 sec at 600 rpm, incubate at 37 ℃ for 45 min

35- Cool the sample to room temperature

36- Centrifuge at 14,000g for 40 min (or more than 40 min until about 20 µL of sample remains in the filter)

37- Make a stock solution of 250 mM IAA in UTT buffer (add 44 mg IAA and 1 mL UTT buffer) in darkness. Then dilute it to 54 mM. (Mix 216 µL of 250 mM with 784 µL UTT buffer to make a 54 mM IAA in UTT buffer)

38- Add 200 µL IAA 54 mM, then wash the same pipet with 50 µL IAA 54 mM (concentration of IAA must be 50 mM, so if the final volume is 270 µL, and the concentration would be 50 mM). vortex for 30 sec at 600 rmp and incubate at dark for 45 min. **IAA should be made and added in a dark room.**

39- Centrifuge at 14,000 g for 40 min.

40- Add 200 µL UTT solution and centrifuge 14,000 g for 40 min

41- Add 200 µL ABC (50 mM) and centrifuge 14,000 g for 40 min

42- Add 200 µL ABC (50 mM) and centrifuge 14,000 g for 40 min

43- Add 150 µL **ABC 100 mM** (because trypsin as acidic and we want to bring the pH to 7). Then, add trypsin with the ratio of 1:25

(for Aq: (250/25) *2= 20 µL trypsin; and for coacervate: (150/25)*2= 12 µL trypsin)

44- Check the pH to be around 7. Then seal the vials with parafilm. Shake at 600 rpm for 1 min. Incubate in wet chamber at 37 °C for 16 hrs.

45- Transfer the filters to new collection tubes.

46- centrifuge at 14,000 g for 40 min.

47- Add 200 µL **0.5 M NaCl**, and centrifuge at 14,000 g for 40 min in the collection tube.

48- Add 200 µL **50 mM NaCl**, invert the filter, and centrifuge at 1000 g for 2 min in the collection tube.

49- Acidify the sample with TFA to bring the pH below 2 (about 5 µL TFA is enough, DO NOT over-acidify)

50- Desalt the samples:

Desalting procedure:

5- Precondition the C18 Sep-Pak columns with 3 mL ACN, 1 mL 0.1 TFA in 75% ACN, 1 mL 0.1 TFA in 50% ACN, 3 mL 0.1 TFA in water

6- Centrifuge acidified samples at 8000g for 1 min, then load the samples to the columns

7- Wash the samples with 3 mL 0.1 TFA in water

Move the Sep-Pak to a 2-ml microcentrifuge tube. Elute the sample with 0.6 ml of 0.1 % TFA in 50% ACN, followed by 0.6 ml of 0.1% TFA in 75%. This step should be performed by gravity, finally push the samples with pipet.

Section S 4. List of the proteins extracted by three coacervation systems and controls.

All the list of the proteins extracted by 50 mM DMMAPS+ 10% HFIP, 50 mM DMMAPS+ 50 mM TEAB+ 10% HFIP and 50 mM DMMAPS+ 50 mM TBAB+ 10% HFIP (Aqueous phase and Coacervate phase) are attached in the excel files **Excel file 1**, **Excel file 2** and **Excel file 3**, respectively along with this supplementary information. Controls are also attached in the **excel file 4.** (Note: These information's are published in analytical chemistry journals)

Section S 5: Gene Ontology based data analysis on Cellular components, Biological process and Molecular function of three coacervation systems

GO_Cellular component	Control: No phase separation	FA*i*C 1 DMMAPS	FA*i*C 2 TEAB-DMMAPS	FA*i*C 3 TBAB-DMMAPS	Id improve FA*i*C3 vs. Control
Types of Membrane proteins					
Membrane proteins	900	1059	1052	1069	18.8%
Integral component of membrane	438	554	540	545	24.4%

Intrinsic component of membrane	454	574	559	563	24.0%
Integral component of organelle membrane	107	132	127	134	25.2%
Integral component of endoplasmic reticulum	44	56	53	58	31.8%
Integral component of mitochondrial membrane	36	43	42	42	16.7%
Integral component of plasma membrane	20	18	25	27	35.0%
Anchored component of membrane	15	17	17	17	13.3%
Mitochondrial types					
Mitochondrial membrane	190	228	229	233	22.6%
Mitochondrial inner membrane	105	130	131	139	32.4%
Mitochondrial outer membrane	71	82	80	78	9.9%
Mitochondrial ribosome	39	58	56	70	79.5%
Mitochondrial matrix	124	147	149	166	33.9%
Other cellular component types					
Golgi membrane	111	121	115	124	11.7%
Golgi apparatus	189	217	208	220	16.4%
Cell wall	42	45	42	46	9.5%
Chromosome	230	247	247	262	13.9%
Nucleus	1171	1310	1335	1367	16.7%
Vacuole	188	243	232	233	23.9%
Vesicle	172	206	197	204	18.6%
Ribosome	133	183	151	165	24.1%
Endosome	93	117	109	113	21.5%
RNA polymerase complex	63	67	62	68	7.9%

Table S1. Subcellular proteins coverage in four FA*i*Cs systems and control (urea) system

Section S 6: Gene Ontology based data analysis on Cellular components, Biological process, and Molecular function.

The identification improvement of different types of proteins based on Gene Ontology data analysis (cellular components, biological process, molecular functions) are given in the table below.

Section S 6: Table S2: GO Cellular components, proteins extracted by 50 mM DMMAPS + 10% HFIP system versus control.

GO_Cellular component	Total proteins in yeast	Control: No phase separation	Coacervate (CO)	Aqueous (AQ)	(AQ+CO)	Identification improvement
Membrane proteins	2054	900	1037	536	1059	17.7%
Integral component of membrane	1229	438	541	199	554	26.5%
Intrinsic component of membrane	1298	454	561	216	574	26.4%
Integral component of plasma membrane	95	20	17	6	18	-10.0%
Integral component of mitochondrial membrane	70	36	43	12	43	19.4%
Integral component of organelle membrane	206	107	132	42	132	23.4%
Integral component of endoplasmic reticulum	80	44	56	15	56	27.3%
Cell wall	141	42	43	39	45	7.1%
RNA polymerase complex	78	63	67	46	67	6.3%
Anchored component of membrane	65	15	17	16	17	13.3%
Vesicle	299	172	203	124	206	19.8%
Vesicle tethering complex	43	24	31	6	31	29.2%
Golgi membrane	154	111	120	73	121	9.0%
Golgi apparatus	298	189	216	115	217	14.8%
Extracellular region	119	36	40	41	44	22.2%
Endosome	180	93	114	73	117	25.8%
Ribosome	256	133	178	158	183	37.6%
Mitochondrial ribosome	90	39	55	39	58	48.7%
Mitochondrial matrix	241	124	141	109	147	18.5%
Mitochondrial membrane	418	190	226	99	228	20.0%
Mitochondrial inner membrane	269	105	129	46	130	23.8%
Mitochondrial outer membrane	107	71	82	44	82	15.5%
Chromosome	509	230	239	135	247	7.4%
Vacuole	505	188	231	129	243	29.3%

Nucleus	2410	1171	1272	854	1310	11.8%

Section S 6: Table S3: GO Molecular function, proteins extracted by 50 mM DMMAPS+ 10%

HFIP system versus control

GO_Molecular function	Total proteins in yeast	Control: No phase separation	Coacervate (CO)	Aqueous (AQ)	(AQ +CO)	Identification improvement
DNA dependent ATPase activity	114	60	60	29	61	1.7%
Histone binding	54	37	39	25	39	5.4%
Structural constituent of ribosome	223	108	153	139	158	46.3%
Structural molecule activity	350	172	218	178	223	29.7%
Transcription regulator activity	271	75	90	49	95	26.7%
Regulatory region nucleic acid binding	162	42	44	31	49	16.7%
DNA binding	618	215	225	157	235	9.3%
Sequence-specific DNA binding	271	74	85	53	91	22.9%
Nucleic acid binding	1281	614	664	504	680	10.7%
Small molecule binding	945	554	616	429	620	11.9%
Nucleotide binding	877	507	567	390	571	12.6%
Ion binding	1610	846	947	672	968	14.4%
Cation binding	846	387	434	327	451	16.5%
Anion binding	970	560	635	431	642	14.6%
Transferase activity	954	466	514	323	519	11.4%
Catalytic activity	2384	1303	1432	984	1456	11.7%
Transferase activity, transferring hexosyl group	82	48	57	26	58	20.8%
Oxidoreductase activity	340	214	231	174	235	9.8%
Unfolded protein binding	89	67	65	51	66	-1.5%
Catalytic activity, acting on DNA	200	68	72	38	72	5.9%

Section S 6: Table S4: GO Biological process, proteins extracted by 50 mM DMMAPS+ 10%

HFIP system versus control

GO_Biological process	Total proteins in yeast	Control: No phase separation	Coacervate (CO)	Aqueous (AQ)	(AQ +CO)	Identification improvement
Metabolic process	3511	1853	2073	1431	2121	14.4%

Biosynthetic process	1544	892	997	721	1014	13.7%
Cellular process	5013	2449	2753	1818	2826	15.4%
Peptide metabolic process	402	242	291	251	298	23.1%
Lipid metabolic process	312	162	192	91	192	18.5%
DNA metabolic process	481	183	196	114	201	9.8%
Sphingolipid metabolic process	45	23	29	11	29	26.1%
Organic acid metabolic process	415	272	285	243	288	5.9%
Peptide biosynthetic process	369	216	264	243	271	25.5%
Lipid biosynthetic process	179	100	120	59	120	20.0%
Membrane lipid biosynthetic process	58	25	36	9	36	44.0%
Amide biosynthetic process	418	248	298	257	306	23.4%
Sphingolipid biosynthetic process	28	14	18	7	18	28.6%
DNA repair	294	124	131	73	135	8.9%
DNA duplex unwinding	63	26	28	16	28	7.7%
Nuclear DNA replication	34	18	18	7	18	0.0%
Mitochondrial gene expression	148	53	69	54	74	39.6%
Translation	365	212	267	231	260	22.6%
Mitochondrial translation	118	48	62	49	67	39.6%
Protein targeting to vacuole	101	41	51	25	51	24.4%
Regulation of autophagy	62	30	36	18	36	20.0%
Negative regulation to RNA metabolic process	289	128	141	79	145	13.3%
Positive regulation to RNA metabolic process	390	176	194	120	201	14.2%
Response to abiotic stimulus	195	99	110	76	114	15.2%
Regulation of transcription by RNA polymerase II	486	208	220	132	228	9.6%

Section S 6: Table S5: GO Cellular components, proteins extracted by 50 mM DMMAPS+50 mM TEAB+ 10% HFIP system versus control.

GO_Cellular component	Total proteins in yeast	Control: No phase separation	Coacervate (CO)	Aqueous (AQ)	(AQ+CO)	Identification improvement
Membrane proteins	2054	900	1007	386	1052	16.89%
Integral component of membrane	1229	438	528	121	540	23.29%

Intrinsic component of membrane	1298	454	546	134	559	23.13%
Integral component of plasma membrane	95	20	24	7	25	25.00%
Integral component of mitochondrial membrane	70	36	42	12	42	16.67%
Integral component of organelle membrane	206	107	127	29	127	18.69%
Integral component of endoplasmic reticulum	80	44	53	11	53	20.45%
Cell wall	141	42	38	36	42	0.00%
RNA polymerase complex	78	63	61	19	62	-1.58%
Anchored component of membrane	65	15	16	13	17	13.33%
Vesicle	299	172	192	70	197	14.53%
Vesicle tethering complex	43	24	26	4	27	12.50%
Golgi membrane	154	111	115	38	115	3.60%
Golgi apparatus	298	189	206	65	208	10.05%
Extracellular region	119	36	35	35	39	8.33%
Endosome	180	93	105	38	109	17.20%
Ribosome	256	133	142	96	151	13.53%
Mitochondrial ribosome	90	39	49	14	56	43.89%
Mitochondrial matrix	241	124	139	56	149	20.16%
Mitochondrial membrane	418	190	224	75	229	20.53%
Mitochondrial inner membrane	269	105	127	42	131	24.76%
Mitochondrial outer membrane	107	71	80	26	80	12.67%
Chromosome	509	230	222	100	247	7.39%
Nucleus	2410	1171	1202	598	1335	14.0%

Section S 6: Table S6: GO Molecular function, proteins extracted by 50 mM DMMAPS+50 mM TEAB+ 10% HFIP system versus control

GO_Molecular function	Total proteins in yeast	Control: No phase separation	Coacervate (CO)	Aqueous (AQ)	(AQ +CO)	Identification improvement

DNA dependent ATPase activity	114	60	57	17	61	1.67%
Histone binding	54	37	34	10	36	-1.67%
Structural constituent of ribosome	223	108	117	76	125	15.74%
Structural molecule activity	350	172	180	106	190	4.65%
Transcription regulator activity	271	75	70	58	99	32.00%
Regulatory region nucleic acid binding	162	42	38	40	62	47.62%
DNA binding	618	215	217	137	262	21.86%
Sequence-specific DNA binding	271	74	73	54	101	36.49%
Nucleic acid binding	1281	614	639	359	698	13.68%
Small molecule binding	945	554	578	281	599	8.12%
Nucleotide binding	877	507	530	265	551	8.68%
Ion binding	1610	846	894	453	955	12.88%
Cation binding	846	387	407	219	444	14.73%
Anion binding	970	560	593	288	618	10.36%
Transferase activity	954	466	510	192	533	14.38%
Catalytic activity	2384	1303	1382	587	1432	9.90%
Transferase activity, transferring hexosyl group	82	48	60	11	60	20.00%
Oxidoreductase activity	340	214	229	134	231	7.94%
Unfolded protein binding	89	67	69	37	69	2.98%
Catalytic activity, acting on DNA	200	68	67	21	72	5.88%

Section S 6: Table S7: GO Biological process, proteins extracted by 50 mM DMMAPS+50 mM TEAB+ 10% HFIP system versus control

GO_Biological process	Total proteins in yeast	Control: No phase separation	Coacervate (CO)	Aqueous (AQ)	(AQ +CO)	Identification improvement
Metabolic process	3511	1853	1963	879	2068	11.60%
Biosynthetic process	1544	892	945	443	979	9.75%
Cellular process	5013	2449	2426	1139	2599	6.12%
Peptide metabolic process	402	242	248	165	259	7.02%
Lipid metabolic process	312	162	188	45	193	19.14%
DNA metabolic process	481	183	181	80	197	7.65%
Sphingolipid metabolic process	45	23	30	6	32	39.13%

Organic acid metabolic process	415	272	278	146	280	2.94%
Peptide biosynthetic process	369	216	223	154	234	8.33%
Lipid biosynthetic process	179	100	117	32	121	21.00%
Membrane lipid biosynthetic process	58	25	35	6	37	48.00%
Amide biosynthetic process	418	248	255	163	266	7.26%
Sphingolipid biosynthetic process	28	14	17	4	19	35.71%
DNA repair	294	124	121	55	135	8.87%
DNA duplex unwinding	63	26	23	5	24	-7.69%
Nuclear DNA replication	34	18	17	4	18	0
Mitochondrial gene expression	148	53	63	19	70	32.07%
Translation	365	212	219	152	230	8.49%
Mitochondrial translation	118	48	57	19	64	33.33%
Protein targeting to vacuole	101	41	49	16	51	24.39%
Regulation of autophagy	62	30	33	17	39	30.00%
Negative regulation to RNA metabolic process	289	128	126	85	157	22.65%
Positive regulation to RNA metabolic process	390	176	170	101	208	18.18%
Response to abiotic stimulus	195	99	100	78	121	22.22%
Regulation of transcription by RNA polymerase II	486	208	205	126	248	19.23%

Section S 6: Table S8: GO Cellular components, proteins extracted by 50 mM DMMAPS+50 mM TBAB+ 10% HFIP system versus control

GO_Cellular components	Total proteins in yeast	Control: No phase separation	Coacervate (CO)	Aqueous (AQ)	(AQ+CO)	Identification improvement
Membrane proteins	2054	900	1015	451	1069	18.78%
Integral component of membrane	1229	438	522	144	545	24.43%
Intrinsic component of membrane	1298	454	539	161	563	24.00%
Integral component of plasma membrane	95	20	25	5	27	35.00%
Integral component of mitochondrial membrane	70	36	38	18	42	16.67%

Integral component of organelle membrane	206	107	129	37	134	25.23%
Integral component of endoplasmic reticulum	80	44	58	13	58	31.82%
Cell wall	141	42	41	40	46	9.52%
RNA polymerase complex	78	63	66	19	68	7.94%
Anchored component of membrane	65	15	16	17	17	13.33%
Vesicle	299	172	199	74	204	18.60%
Vesicle tethering complex	43	24	34	2	34	41.67%
Golgi membrane	154	111	121	45	124	11.71%
Golgi apparatus	298	189	216	79	220	16.40%
Extracellular region	119	36	35	40	42	16.67%
Endosome	180	93	109	40	113	21.50%
Ribosome	256	133	129	137	165	24.06%
Mitochondrial ribosome	90	39	36	50	70	79.48%
Mitochondrial matrix	241	124	122	113	166	33.87%
Mitochondrial membrane	418	190	213	107	233	22.63%
Mitochondrial inner membrane	269	105	121	68	139	32.38%
Mitochondrial outer membrane	107	71	77	36	78	9.85%
Chromosome	509	230	239	118	262	13.91%
Nucleus	2410	1171	1259	678	1367	

Section S 6: Table S9: GO Molecular function, proteins extracted by 50 mM DMMAPS+50 mM TBAB+ 10% HFIP system versus control

GO_Molecular function	Total proteins in yeast	Control: No phase separation	Coacervate (CO)	Aqueous (AQ)	(AQ+CO)	Identification improvement
DNA dependent ATPase activity	114	60	60	22	62	3.33%
Histone binding	54	37	36	12	36	-2.70%
Structural constituent of ribosome	223	108	105	117	140	29.63%
Structural molecule activity	350	172	169	157	206	19.76%
Transcription regulator activity	271	75	82	46	100	33.33%

Regulatory region nucleic acid binding	162	42	48	34	58	38.09%
DNA binding	618	215	230	143	260	20.93%
Sequence-specific DNA binding	271	74	78	53	95	28.38%
Nucleic acid binding	1281	614	650	412	707	15.15%
Small molecule binding	945	554	590	315	605	9.20%
Nucleotide binding	877	507	542	291	556	9.66%
Ion binding	1610	846	917	527	963	13.83%
Cation binding	846	387	423	267	455	17.57%
Anion binding	970	560	606	318	622	11.07%
Transferase activity	954	466	528	224	545	16.95%
Catalytic activity	2384	1303	1421	740	1469	12.74%
Transferase activity, transferring hexosyl group	82	48	53	7	53	10.42%
Oxidoreductase activity	340	214	223	142	231	7.94%
Unfolded protein binding	89	67	69	50	70	4.48%
Catalytic activity, acting on DNA	200	68	68	28	69	1.47%

Section S 6: Table S10: GO_Biological process, proteins extracted by 50 mM DMMAPS+50 mM TEAB+ 10% HFIP system versus control

GO_Biological process	Total proteins in yeast	Control: No phase separation	Coacervate (CO)	Aqueous (AQ)	(AQ +CO)	Identification improvement
Metabolic process	3511	1853	2007	1101	2144	15.70%
Biosynthetic process	1544	892	949	568	1013	13.56%
Cellular process	5013	2449	2495	1369	2684	9.59%
Peptide metabolic process	402	242	242	218	281	16.11%
Lipid metabolic process	312	162	185	60	188	16.05%
DNA metabolic process	481	183	192	89	202	10.44%
Sphingolipid metabolic process	45	23	31	5	31	34.78%
Organic acid metabolic process	415	272	286	192	291	6.98%
Peptide biosynthetic process	369	216	216	203	254	17.59%
Lipid biosynthetic process	179	100	113	37	114	14.00%
Membrane lipid biosynthetic process	58	25	34	3	34	36.00%

Amide biosynthetic process	418	248	253	222	291	17.34%
Sphingolipid biosynthetic process	28	14	18	2	18	28.57%
DNA repair	294	124	129	59	136	9.68%
DNA duplex unwinding	63	26	26	8	26	0
Nuclear DNA replication	34	18	17	3	17	-5.55%
Mitochondrial gene expression	148	53	51	58	87	64.15%
Translation	365	212	212	200	250	17.92%
Mitochondrial translation	118	48	45	57	81	68.75%
Protein targeting to vacuole	101	41	50	15	51	24.39%
Regulation of autophagy	62	30	35	14	37	23.33%
Negative regulation to RNA metabolic process	289	128	139	82	156	21.87%
Positive regulation to RNA metabolic process	390	176	191	102	212	20.45%
Response to abiotic stimulus	195	99	109	74	121	22.22%
Regulation of transcription by RNA polymerase II	486	208	221	122	247	18.75%

Section S 7: List of reference yeast proteins with their abundance value obtained from YeastMine software.

The list of yeast proteins with their abundance values are added in the Excel file **(named as Abundance database R)** along with this supplementary information. (Note: These information's are published in analytical chemistry journals)

Section S 8: Abundance chart of the proteins extracted by different systems with respect to control

The figure below shows the abundance chart of proteins extracted by three different coacervation systems and 8 M urea as control without phase separation. Proteins are ranged into different abundance value in X-axis and their number are represented in Y-axis. The identification improvement for lower abundance proteins is higher and proteins with higher abundance above 5000 molecules/cells is very small. The calculated abundance chart of each systems based on the

reference value are added in the Excel file (**named as abundance chart for different systems**) along with this supplementary information.

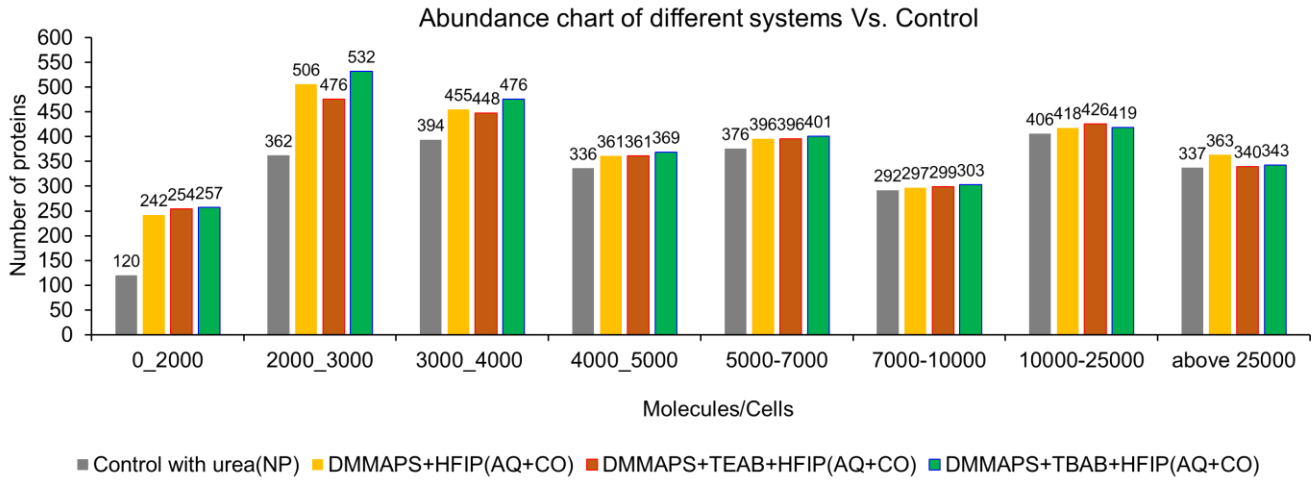


Figure S2: Protein abundance chart extracted by different systems vs. control

Section S 9: Coverage of subcellular low abundance proteins in FA*i*C systems and urea control systems

The table S11 below shows the lower abundance proteins in the subcellular proteome of yeast cells.

GO Cellular locations of lower abundance proteins	#Number of proteins	Nucleus	Membrane	Integral component of membrane	Endoplasmic reticulum	Vacuole	Mitochondrion
LAP (< 2000 molecules/cells)							
Control with 8M urea (NP)	120	41	54	35	22	16	18
50 mM DMMAPS+ 10% HFIP	242	76	118	83	44	37	42
50 mM DMMAPS+ 50 mM TEAB+ 10% HFIP	254	99	109	72	41	36	36

50 mM DMMAPS+ 50 mM TBAB+ 10 % HFIP	257	88	116	75	43	38	39
LAP (2000-3000 molecules/cells)							
Control with 8M urea (NP)	362	171	133	67	37	28	61
50 mM DMMAPS+ 10% HFIP	506	229	188	107	57	45	88
50 mM DMMAPS+ 50 mM TEAB+ 10% HFIP	476	233	180	97	53	37	84
50 mM DMMAPS+ 50 mM TBAB+ 10 % HFIP	532	258	191	103	53	36	105

Table S11: List of number of lower abundance proteins into different subcellular locations of yeast proteome.

Section S 10: The script and instruction for the alpha helices' calculation of trans-membrane proteins.

The software for the alpha helices of trans-membrane proteins calculation was developed in the lab. The script and instruction for the alpha helices' calculation of trans-membrane proteins is available to the link attached https://github.com/Stasharofi/Transmembrane-alpha-helix-calculator The list of the proteins with their transmembrane alpha helices extracted in different systems including control are attached in the excel file along with this supplementary information.

Section S 11: List of the proteins and sequence coverage of alpha helices trans-membrane protiens extracted by different systems and their representation in box chart.

The proteins with their alpha helices coverage of trans-membrane proteins are extracted by different systems in different number. The list of these proteins with their sequence coverage value

are attached in the Excel files (**named as alpha helices coverage for different systems and control**) along with this supplementary information. The sequence coverage of alpha helices of transmembrane proteins extracted by different systems was presented in box chart below. The box shows the distribution of the 25% to 75% of the data, the lines below and above the box each represents the first and the fourth 25% of the data.

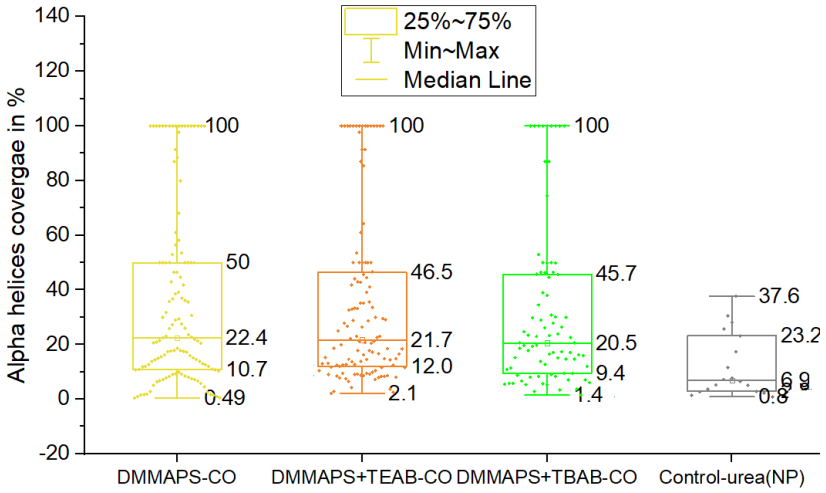


Figure S3 (Box chart): Alpha -helices coverages in FA*i*Cs and control system.

Section S 12: Reference list of yeast proteins with phosphorylated residue and ubiquitinylated lysine residue

These lists are obtained from YeastMine database (https://yeastmine.yeastgenome.org/yeastmine/begin.do) and are compared with the proteins extracted by different systems. List of yeast proteins with their post-translational modifications are attached with this supplementary information in Excel files (named as Yeast proteome with

phosphorylated residue reference and Yeast proteome with ubiquitinylated lysine residue_reference).

Section S 13: Correlation between lower abundance proteins and phosphorylated proteins

GO analysis: Sub-cellular locations	# Total number of lower abundance proteins	#Number of phosphorylated proteins
LAP (<2000 molecules/cells)		
Control with 8M Urea (NP)	120	80
50 mM DMMAPS+ 10% HFIP	242	150
50 mM DMMAPS+ 50 mM TEAB+ 10% HFIP	254	164
50 mM DMMAPS+ 50 mM TBAB+ 10% HFIP	257	170
LAP (2000-3000 molecules/cells)		
Control with 8M Urea (NP)	362	237
50 mM DMMAPS+ 10% HFIP	506	311
50 mM DMMAPS+ 50 mM TEAB+ 10% HFIP	476	309
50 mM DMMAPS+ 50 mM TBAB+ 10% HFIP	532	332

Table S12: Correlation between phosphorylated proteins and lower abundance proteins of yeast proteome.

Section S 14: Total number of proteins pooled for whole, membrane, low abundance and phosphorylated proteins from different coacervation systems.

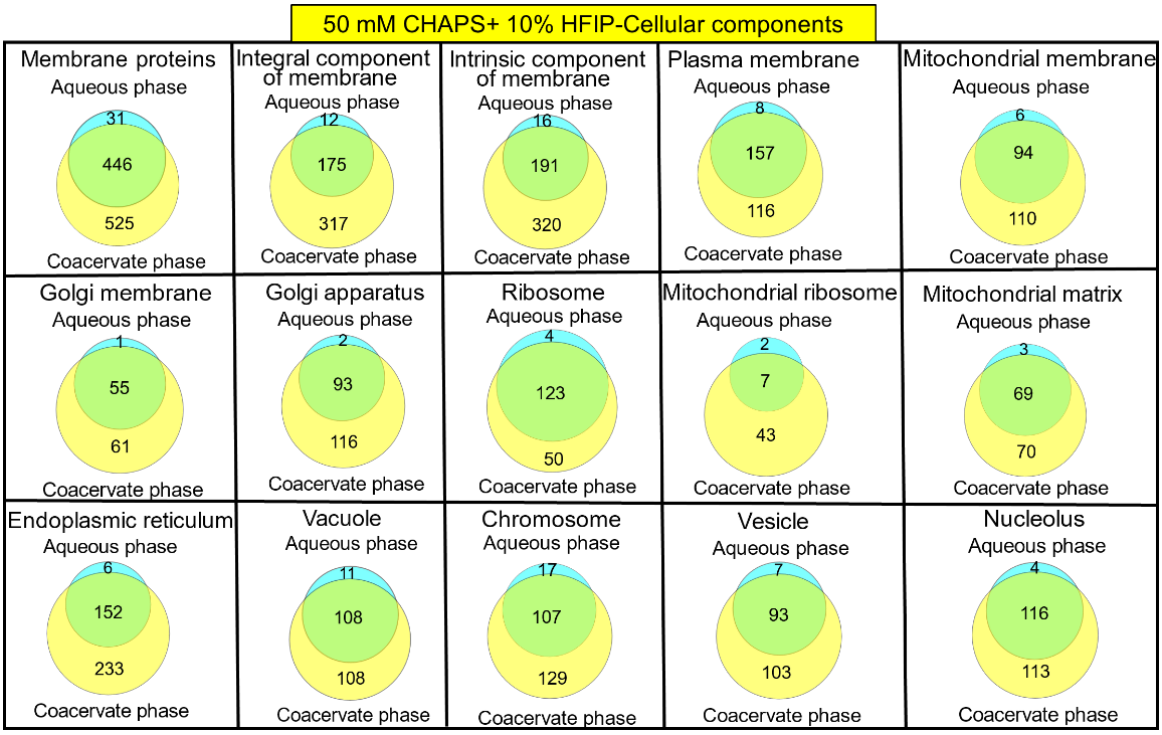
Pooled data from three systems					
Types of proteins	DMMAPS +HFIP	DMMAPS+TEAB +HFIP	DMMAPS+TBAB+ HFIP	Total	# Identification improvement vs. control
Total	3049	3002	3100	3488	863 (32.9%)
Membrane	1059	1052	1069	1209	309 (34.3%)
Low abundance	**242**	**254**	**257**	**414**	**294 (245.0%)**
Phosphorylated	1976	1982	2012	2222	462 (26.3%)

Table S 13: Pooled data for different types of proteins collected from three different DMMAPS coacervation systems.

APPENDIX B: Supporting Information for Chapter 3

Section 1: Fractionation of sub-cellular proteins by CHAPS and CHAPS+QUATS systems

Figure shows that addition of TEAB and TBAB to HFIP biphasic supramolecular systems of CHAPS changes the fractionation of proteins and the best fractionation is observed by the addition of TEAB to CHAPS. This increases the number of uniquely identified proteins into bottom coacervate phase of the system. Similarly, addition of TBAB to CHAPS increases the extraction of uniquely identified proteins more into the top aqueous phase.



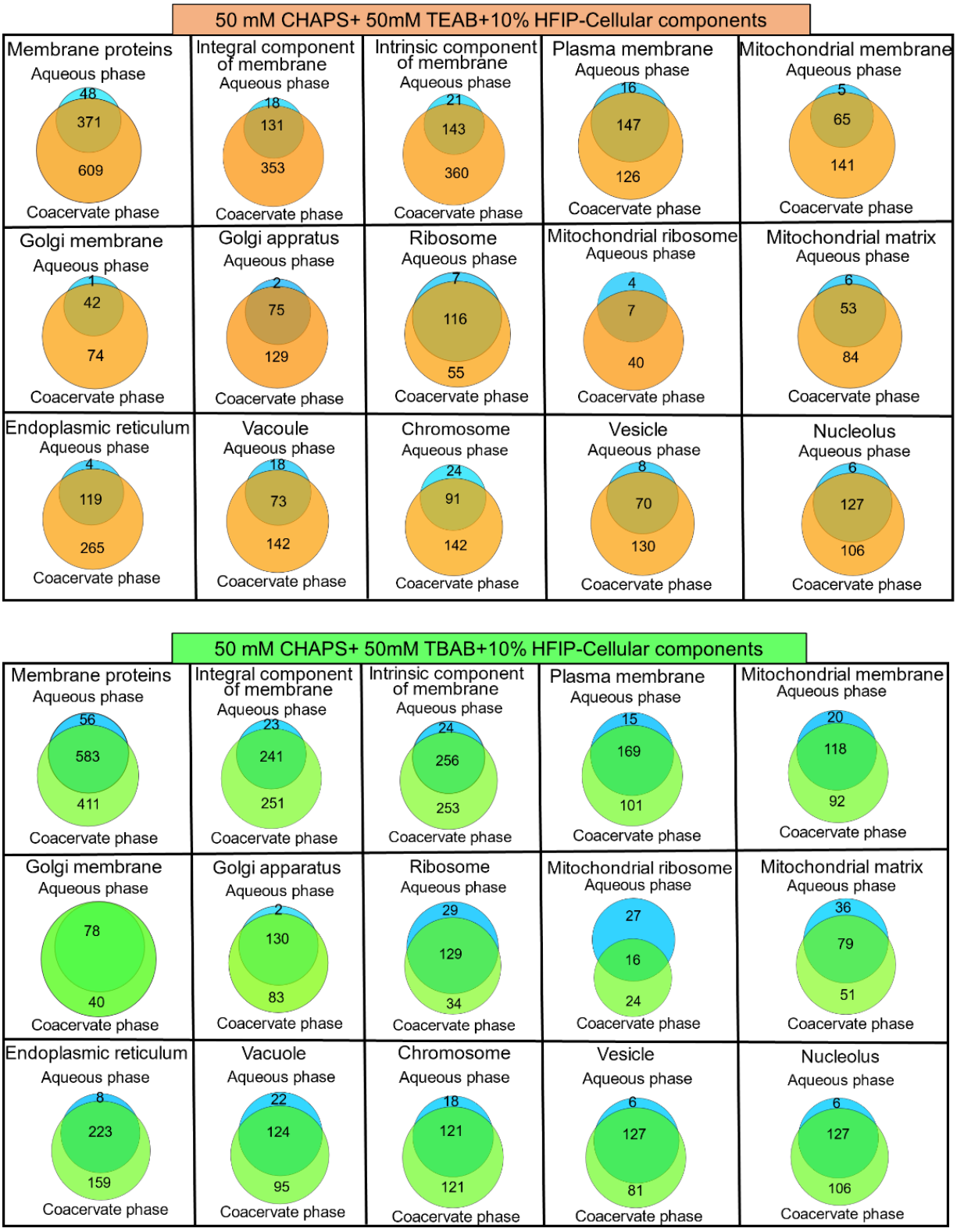


Figure S1: Fractionation of sub-cellular yeast proteins by the HFIP induced biphasic systems of CHAPS and CHAPS+QUATS.

Section 2: Extraction and identification improvement of post-translationally modified proteins using HFIP induced biphasic system of CHAPS and CHAPS+QUATS versus 8M urea as control (NP)

	Control	AQ	CO	Total (AQ+CO)	Identification improvement Vs. Control
Proteins with phosphorylated residue					
CHAPS+HFIP	1746	1200	1898	1947	201 (11.5%)
CHAPS+TEAB+HFIP	1746	1101	1886	1997	251 (14.4%)
CHAPS+TBAB+HFIP	1746	1362	1899	1986	240 (13.7%)
Proteins with ubiquitinylated lyside residue					
CHAPS+HFIP	1331	959	1402	1422	91 (6.8%)
CHAPS+TEAB+HFIP	1331	840	1405	1428	97 (7.3%)
CHAPS+TBAB+HFIP	1331	1059	1410	1429	98 (7.4%)

ST1: Table shows the number of phosphorylated and ubiquitinylated proteins extracted by biphasic system of CHAPS and CHAPS+QUATS and their identification improvement versus control.

APPENDIX C: Supporting Information for Chapter 4

Section1: Fractionation of aqueous phase and coacervate proteins of CHAPS and CHAPS+QUATS systems based on their hydrophobicity and isoelectric point.

The addition of TEAB and TBAB to CHAPS alters the fractionation pattern of proteins based on their isoelectric point. Basically, it follows the similar trend to DMMAPS and DMMAPS+ QUATS systems. Since both DMMAPS and CHAPS are neutral and hydrophobic zwitterionic surfactants, fractionation of proteins was previously triggered by hydrophobic interaction between phase and proteins. Addition of positively charged TEAB and TBAB to them enhances the electrostatic interaction between phase and proteins and fractionation alters.

Same information is provided by their amino acid compositional analysis as well. We analyzed the proteins extracted into aqueous and coacervate phase of each system and plotted them in the box chart diagram as shown in figure S2 and S3. Addition of TEAB and TBAB to the CHAPS and DMMAPS extracted more basic proteins into aqueous phase with higher percentage (%) of basic amino acid into aqueous phase and lower percentage of basic amino acid into coacervate phase and opposite trend for acidic amino acid composition.

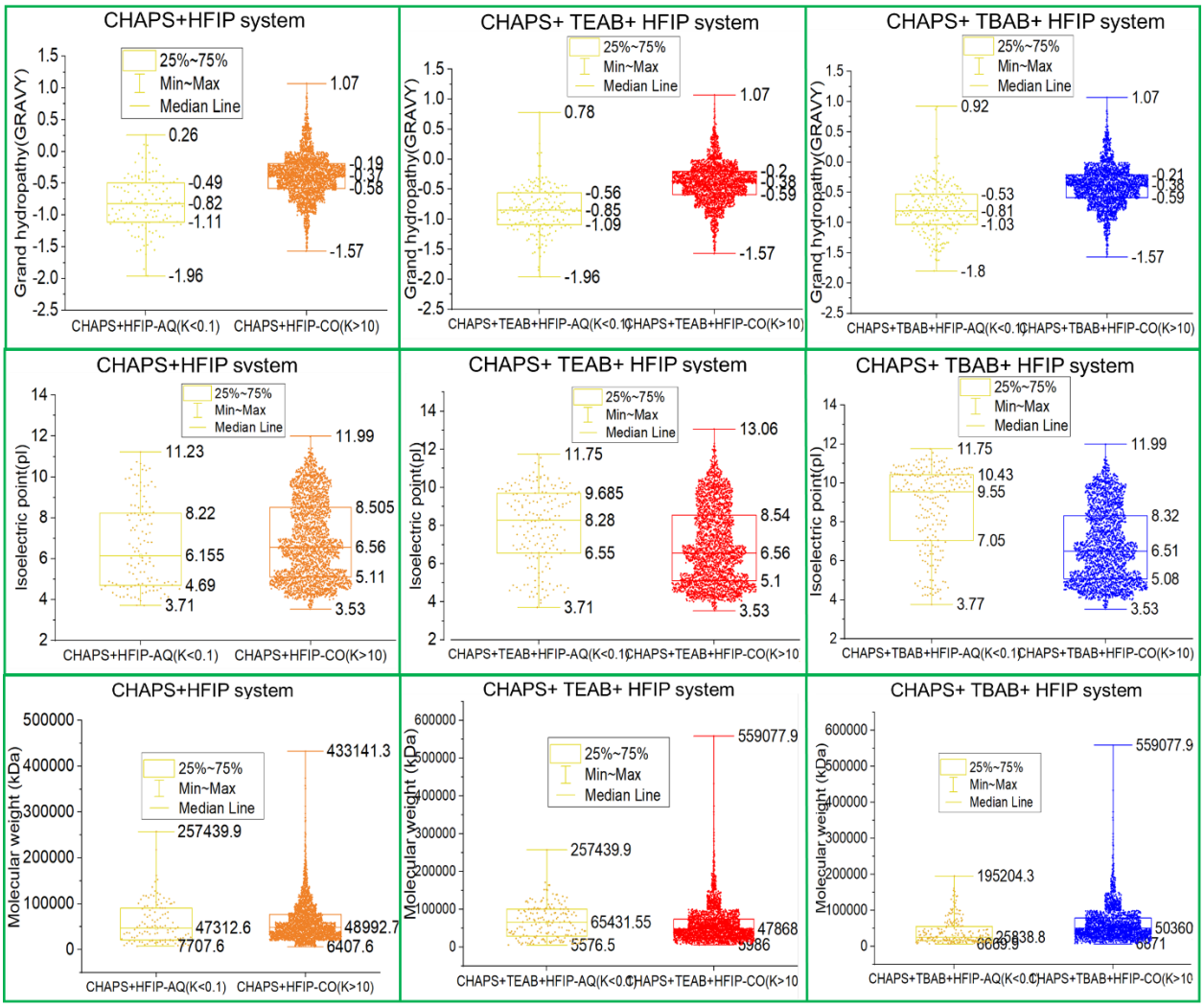


Figure S1: Fractionation of proteins into aqueous phase and coacervate phase of CHAPS and CHAPS+QUATS systems based on their GRAVY value, pI value and molecular weight.

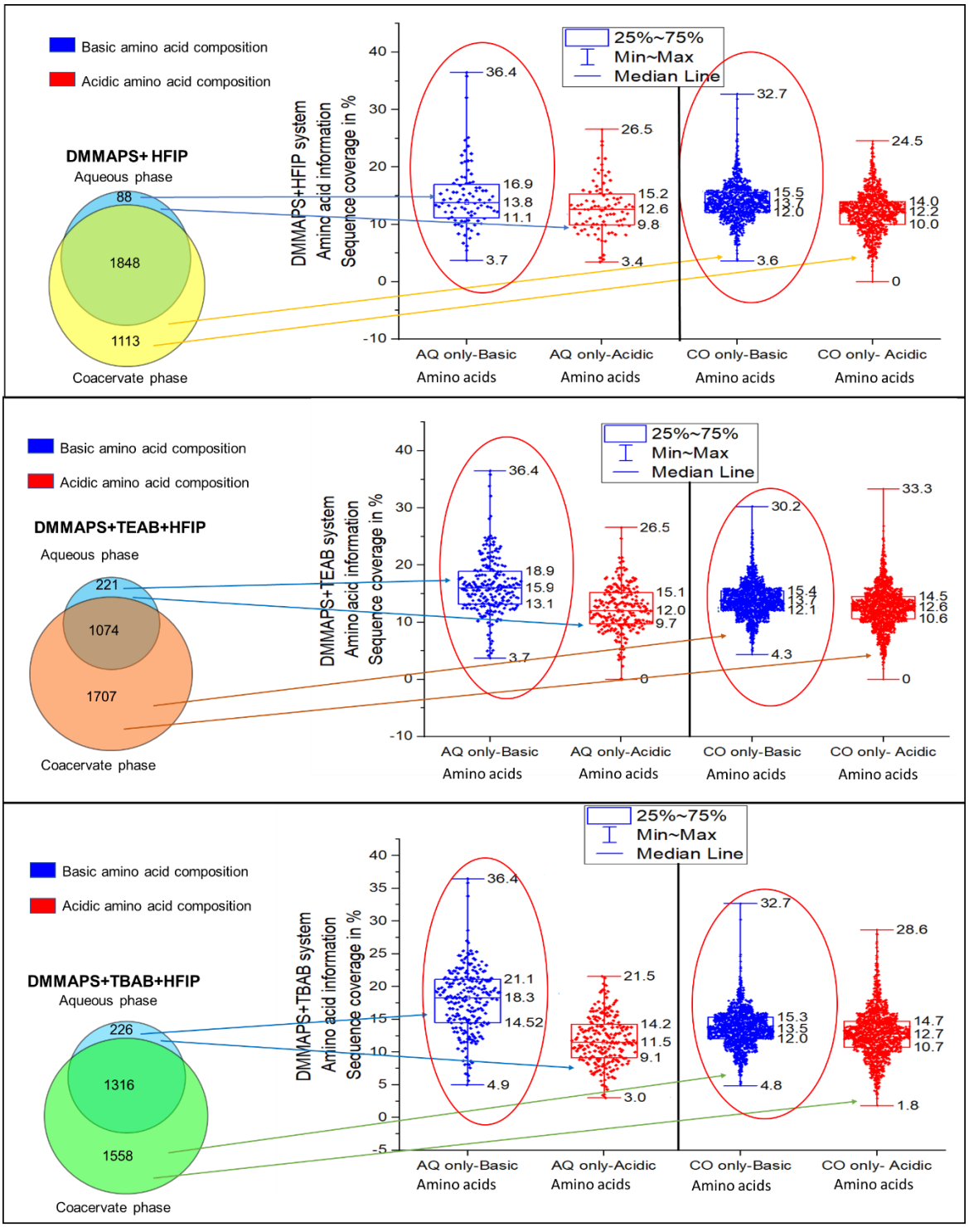


Figure S2: Compositional analysis of acidic and basic amino acids for the proteins extracted into aqueous phase and coacervate phase of DMMAPS and DMMAPS+QUATS systems.

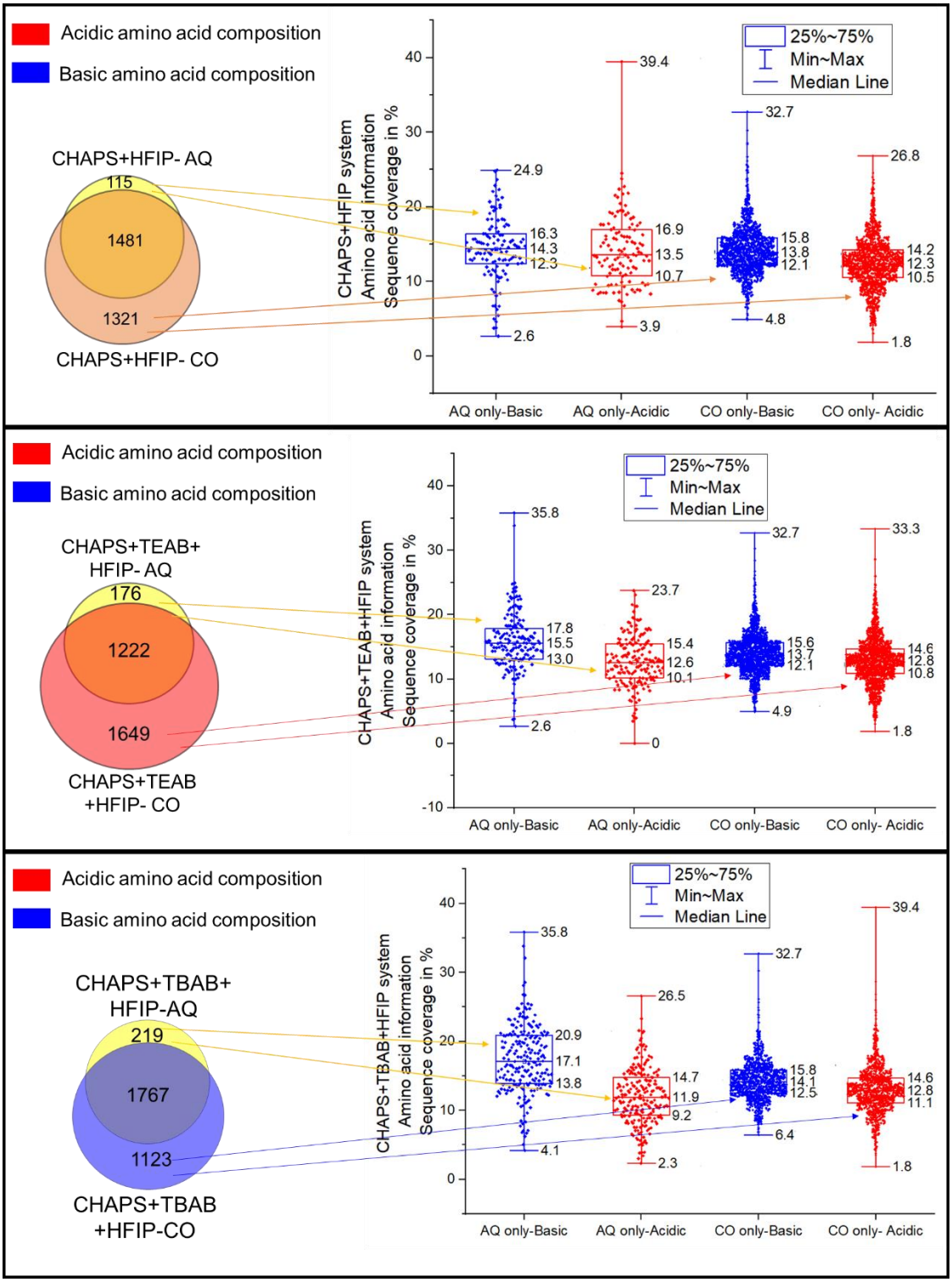


Figure S3: Compositional analysis of acidic and basic amino acids for the proteins extracted into aqueous phase and coacervate phase of CHAPS and CHAPS+QUATS systems.

Section 3: Selectivity difference between three FAiC-BPS of CHAPS and mixed amphiphiles for the extraction of proteins into the aqueous phase and coacervate phase.

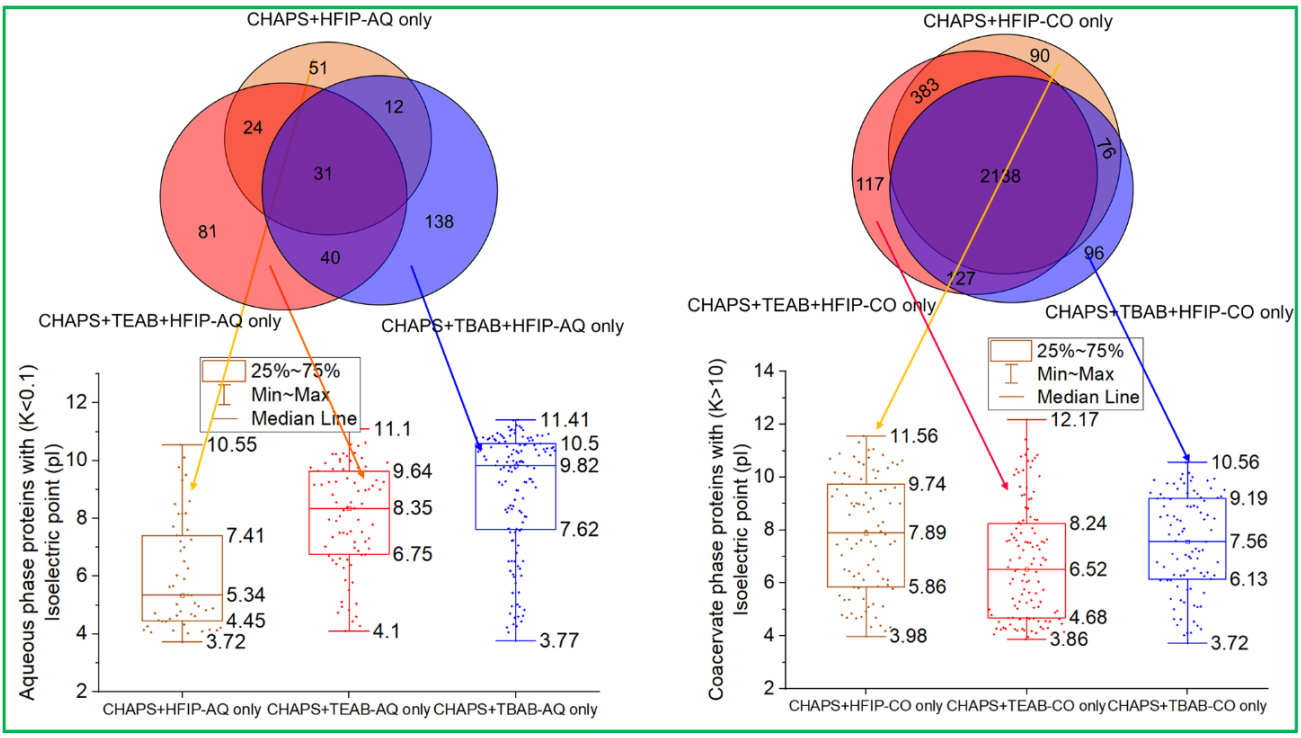


Figure S4: Selectivity difference between the CHAPS and CHAPS+QUATS systems for the extraction of proteins.

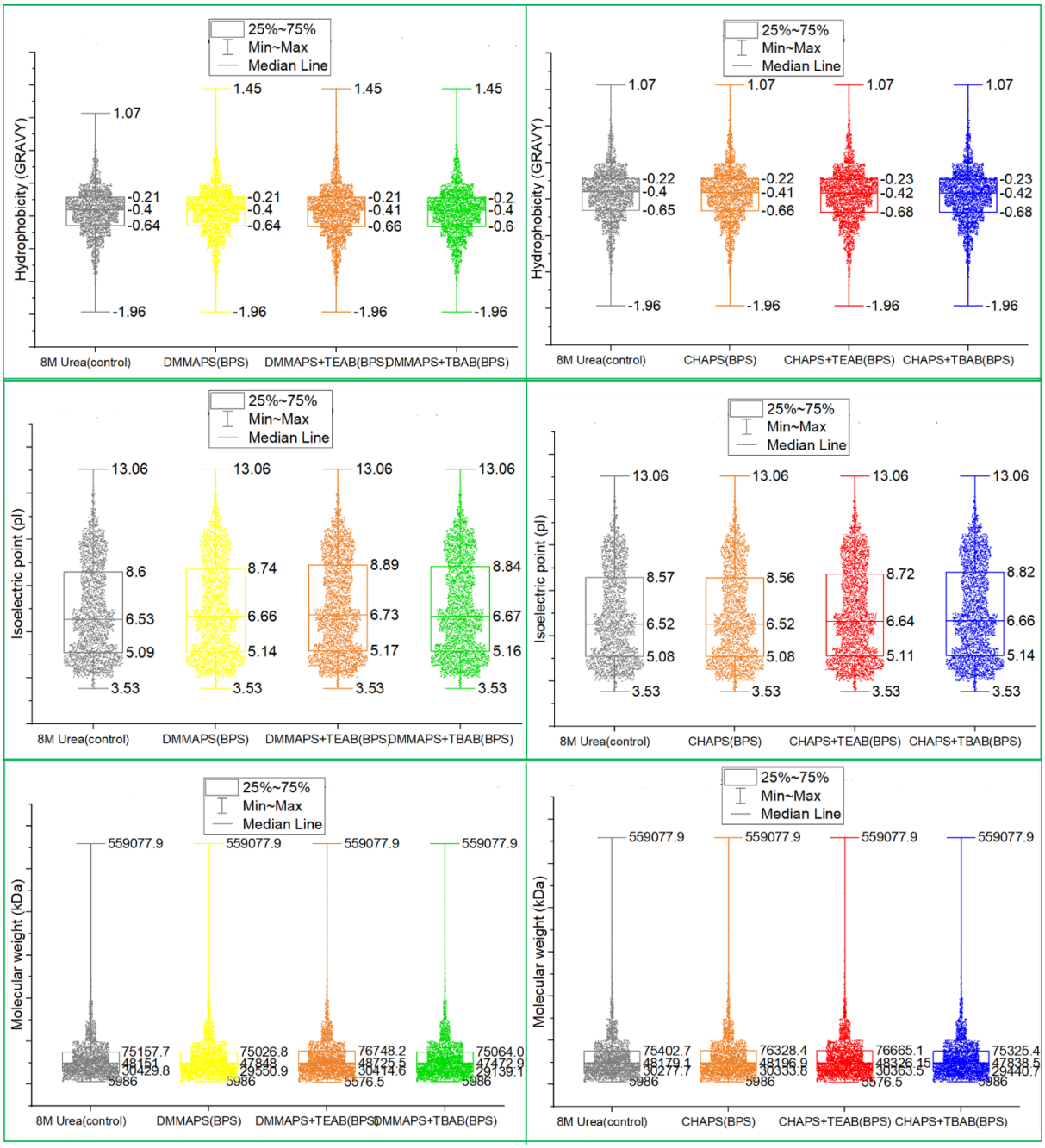


Figure S5: The protein distribution based on their hydrophobicity (GRAVY), isoelectric point (pI), and molecular weight (kDa) extracted by different supramolecular biphasic systems of DMMAPS and mixed amphiphiles and 8M urea as control (NP).

APPENDIX D: Supporting Information for Chapter 5

Section 1: Calibration curves of PAHs used for the calculation of unknown concentration of sample.

The figure shows the calibration curves of different PAHs with the wide range of concentration and shows the linearity from the range of 50 to 5000 ng/mL.

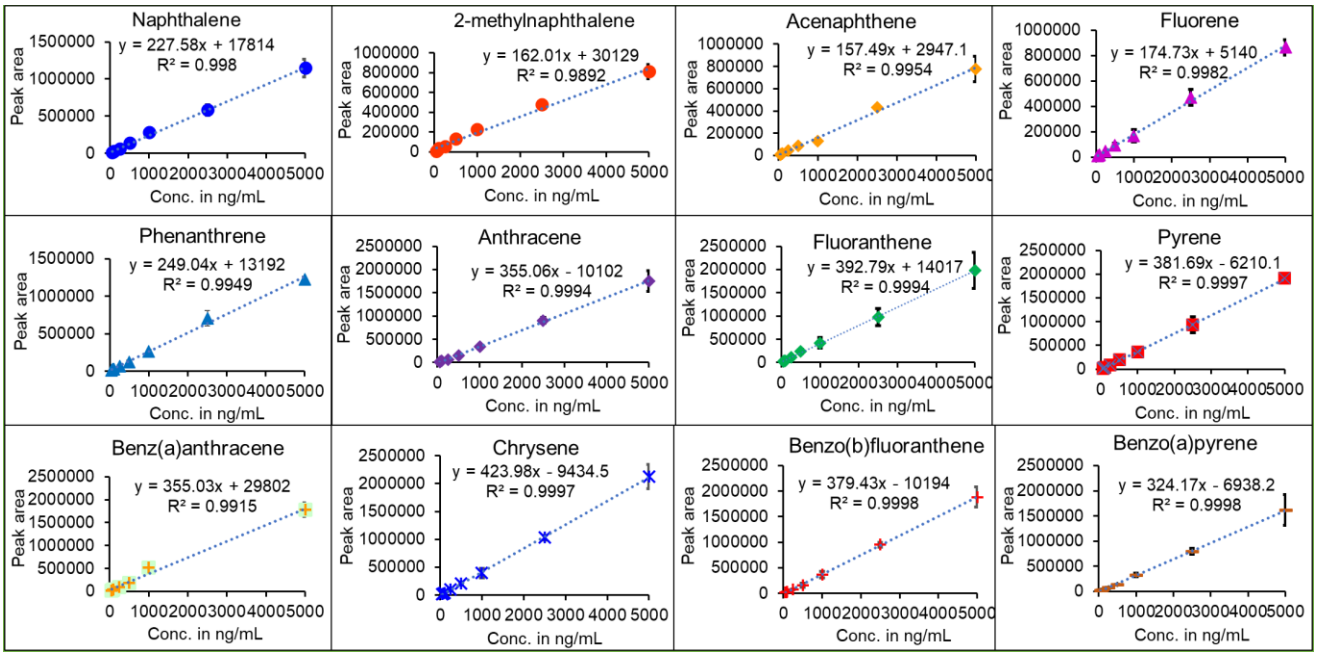
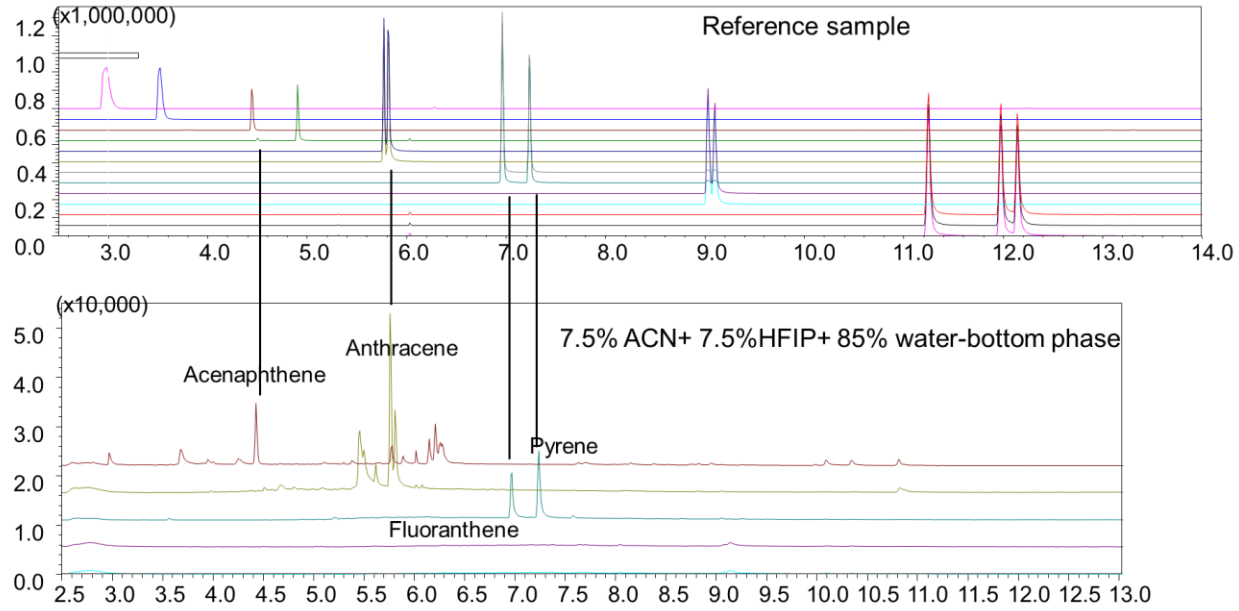
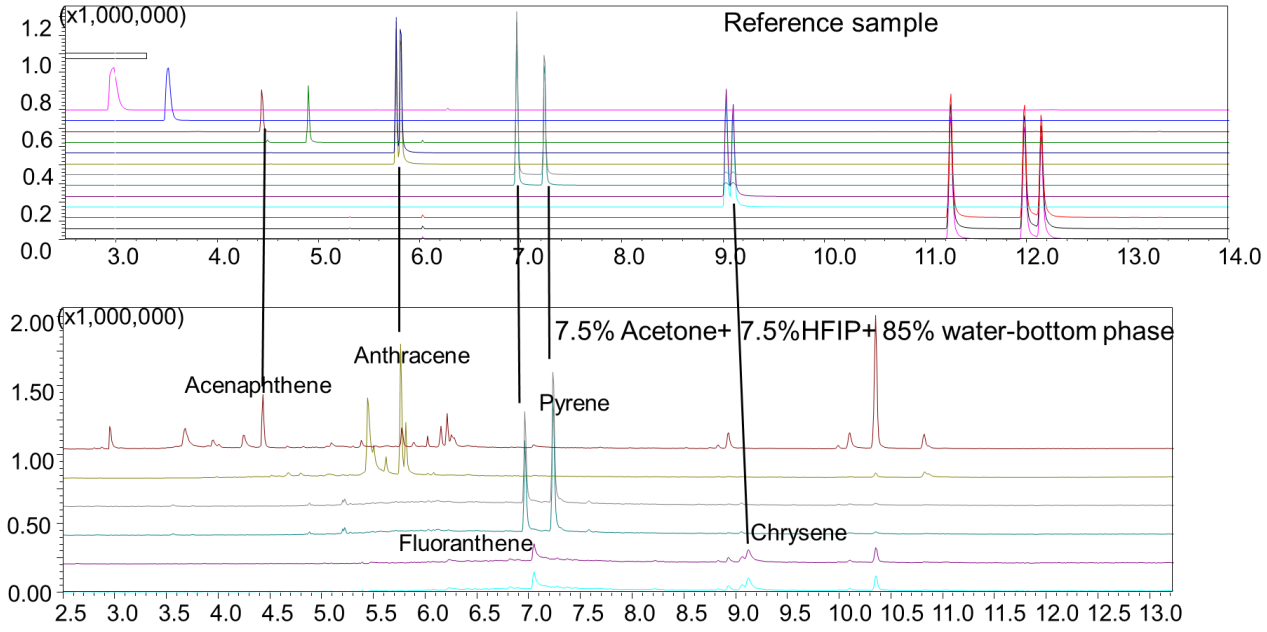


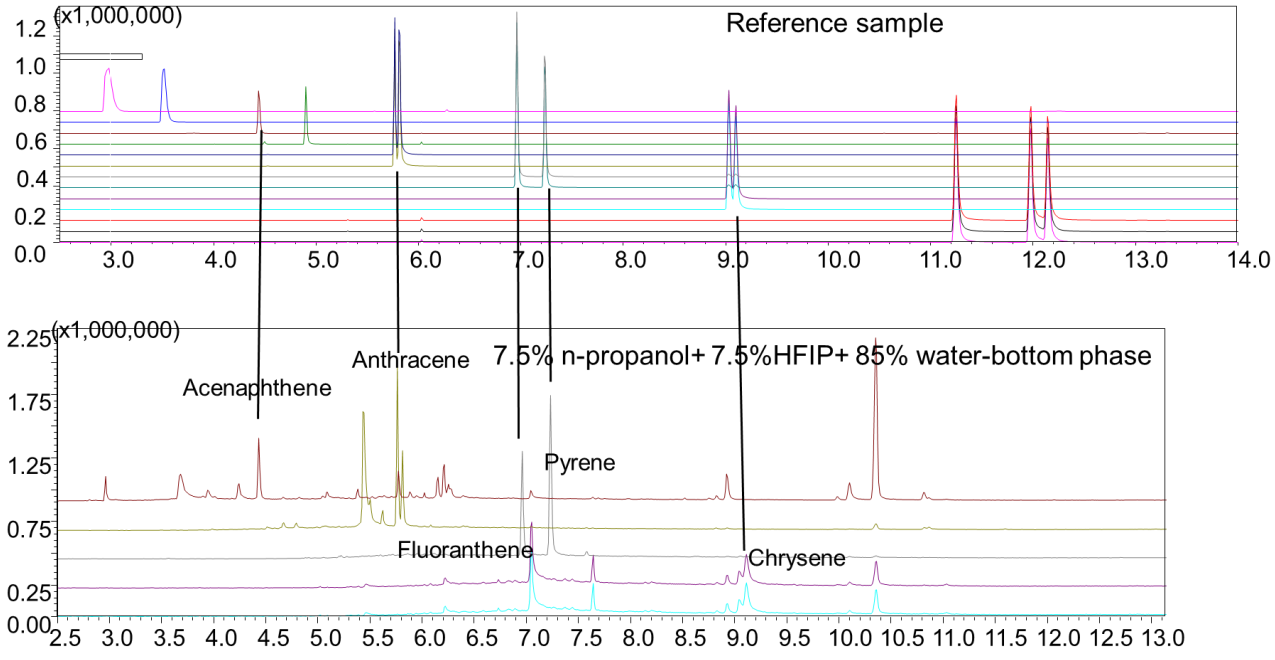
Figure S1: Calibration curves of PAHs samples with wide range of concentration

Section 2: Chromatograms showing the detection of PAHs from soil sample after their enrichment into bottom phase of different associated organic solvents biphasic systems.

The chromatograms shows the extraction and enrichment of PAHs from soil sample using four different associated organic solvents. Chromatograms are compared with chromatogram of reference sample as shown below. Few PAH samples were detected after enrichment with bottom phase of biphasic systems.




(x1,000,000)
Reference sample
(x1,000,000)
7.5% Acetone+ 7.5%HFIP+ 85% water-bottom phase
Acenaphthene
Anthracene
Pyrene
Fluoranthene
Chrysene


(x1,000,000)
Reference sample
(x1,000,000)
7.5% n-propanol+ 7.5%HFIP+ 85% water-bottom phase
Acenaphthene
Anthracene
Pyrene
Fluoranthene
Chrysene

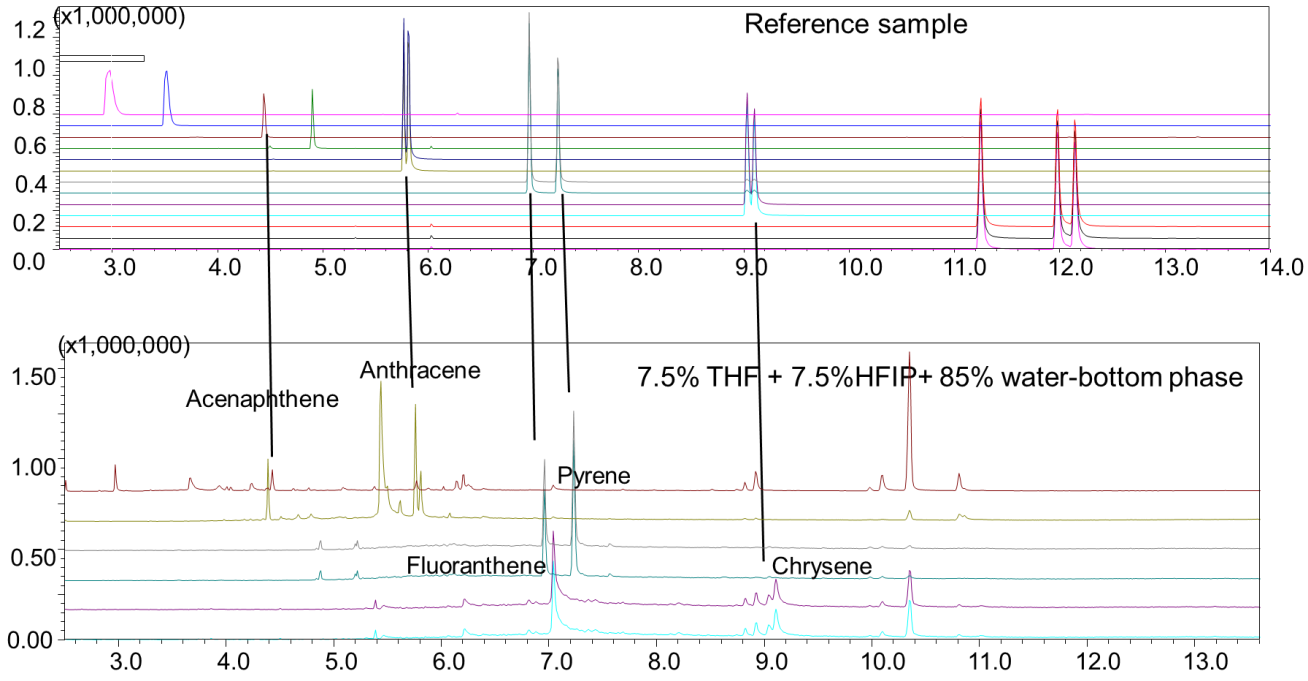


Figure S2: Chromatograms showing the detection of PAHs from soil sample after their enrichment into bottom phase of different associated organic solvents biphasic systems.

BIOGRAPHIC INFORMATION

Durga Devi Khanal is a native of Nepal. She completed her bachelor's degree in Chemistry from Mechi Multiple Campus, Jhapa, Nepal followed by her master's degree in Organic Chemistry from Tribhuvan University, Kathmandu, Nepal. She came to USA in 2013 with her husband and had a family. She graduated with a master's degree from the University of Texas at Arlington in 2018. During that time, she was involved in research work in Dr. Kevin Schug lab in Analytical chemistry division. She continued her PhD program in UTA and joined Dr. Morteza G. Khaledi group to pursue her doctoral degree education. Her Ph.D. research work was focused on the method development for the sample preparation in proteomics using fluoro alcohol induced supramolecular biphasic systems of different surfactants and mixed amphiphiles and sample analysis using LC-MS. In addition, she has also worked on the extraction and enrichment of environmental pollutants and their analysis and separation using GC-MS analysis.